The impact of chromatin factors in gene expression noise in a dynamic system

Dissertation

zum Erwerb des Doktorgrades der Naturwissenschaften
an der Fakultät für Biologie der
Ludwig-Maximilians-Universität München

vorgelegt von
Marlet Morales Franco
München, 25.03.2024

Diese Dissertation wurde am Helmholtz Zentrum München,

dem Deutschen Forschungszentrum für Gesundheit und Umwelt,

im Institut der funktionellen Epigenetik verfasst

1. Gutachter:	Prof. Dr. Robert Schneider
2. Gutachter:	Prof. Dr. Wolfgang Enard
Tag der Abgabe:	25.03.2024
Tag der mündlichen Prüfung:	07.10.2024

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorliegende Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt ist.

München, 25.03.2024 Marlet Morales Franco

Ort, Datum Marlet Morales Franco

Erklärung

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist. Ich erkläre weiter, dass ich mich anderweitig einer Doktorprüfung ohne Erfolg nicht unterzogen habe.

München, 25.03.2024 Marlet Morales Franco

Ort, Datum Marlet Morales Franco

TABLE OF CONTENTS

ABSTRACT

Gene expression consists of multiple interlinked steps. For a gene to be expressed, chromatin needs to be in an “active” conformation, allowing the binding of the transcriptional machinery. Once a gene is transcribed, the resulting mRNA is processed and exported to the cytoplasm, where it is bound by ribosomes and translated. Every one of these steps is subject to stochastic fluctuations, and thus, cells will exhibit gene expression variability even under the most homogeneous conditions. This cell-to-cell variability has been termed “noise” and exists at the core of processes such as embryonic development, cancer emergence, and antibiotic resistance.

In this project, I addressed noise from an environmental adaptation perspective and asked how chromatin-related factors can alter the noise of a gene during induction. I used *Saccharomyces cerevisiae* undergoing a switch in carbohydrate source as a model system and adopted the endogenous gene Gal1 fused to GFP to quantify noise during gene induction.

The project started by analyzing data from a high-throughput screening of Gal1-GFP induction performed in strains lacking a single chromatin-related gene. The screening identified several chromatin factors that affected population heterogeneity of Gal1 while maintaining the same average protein expression level. Subsequent experiments validated that the absence of these factors created higher or lower noise levels during induction, but not when steady-state Gal1 expression was reached.

I further addressed gene expression noise at the transcript level. The conclusions were similar when observing Gal1-GFP mRNA, suggesting that transcriptional regulation contributes to the higher or lower noise observed on the Gal1-GFP protein. Mathematical models of the transcript and protein data suggested that some of the identified chromatin factors affected the DNA activation/inactivation of Gal1 but not its transcription or translation, suggesting a potential mechanism for noise emergence. Finally, our findings point towards the anchor domain of Ies2 as indispensable for noise control.

My results explore how chromatin regulation, individual cell behavior, and population responses to environmental cues relate to each other. In addition, I highlighted the importance of examining non-steady-states when describing the noise of gene expression.

ZUSAMMENFASSUNG

Die Genexpression besteht aus mehreren miteinander verknüpften Schritten. Damit ein Gen exprimiert werden kann, muss sich das Chromatin in einer "aktiven" Konformation befinden, die die Bindung der Transkriptionsmaschinerie ermöglicht. Nach der Transkription eines Gens wird die resultierende mRNA verarbeitet und in das Zytoplasma exportiert, wo sie von Ribosomen gebunden und übersetzt wird. Jeder dieser Schritte ist stochastischen Fluktuationen unterworfen, so dass Zellen selbst unter den homogensten Bedingungen eine Variabilität der Genexpression aufweisen. Diese Variabilität von Zelle zu Zelle wird als "Rauschen" bezeichnet und steht im Mittelpunkt von Prozessen wie der Embryonalentwicklung, der Krebsentstehung und der Antibiotikaresistenz.

In diesem Projekt untersuchte ich das Rauschen aus der Perspektive der Umweltanpassung und fragte, wie chromatinbezogene Faktoren das Rauschen eines Gens während der Induktion verändern können. Als Modellsystem verwendete ich *Saccharomyces cerevisiae*, das eine Umstellung der Kohlenhydratquelle durchläuft, und nahm das endogene Gen Gal1, das mit GFP fusioniert ist, zur Quantifizierung des Rauschens während der Geninduktion.

Das Projekt begann mit der Analyse von Daten aus einem Hochdurchsatz-Screening der Gal1-GFP-Induktion in Stämmen, denen ein einzelnes chromatinbezogenes Gen fehlt. Bei diesem Screening wurden mehrere Chromatinfaktoren identifiziert, die die Heterogenität der Gal1-Population bei gleichbleibender durchschnittlicher Proteinexpression beeinflussten. Anschließende Experimente bestätigten, dass das Fehlen dieser Faktoren zu einem höheren oder niedrigeren Rauschen während der Induktion führte, jedoch nicht, wenn der Steady-State der Gal1-Expression erreicht war.

Ich untersuchte das Rauschen der Genexpression auch auf der Transkriptionsebene. Die Schlussfolgerungen waren ähnlich, wenn man die Gal1-GFP-mRNA betrachtete, was darauf hindeutet, dass die Transkriptionsregulierung zu dem höheren oder niedrigeren Rauschen beiträgt, das beim Gal1-GFP-Protein beobachtet wird. Mathematische Modelle der Transkript- und Proteindaten legten nahe, dass einige der identifizierten Chromatinfaktoren die DNA-Aktivierung/Inaktivierung von Gal1, nicht aber seine Transkription oder Translation beeinflussen, was auf einen möglichen Mechanismus

für die Entstehung von Rauschen hindeutet. Schließlich deuten unsere Ergebnisse darauf hin, dass die Ankerdomäne von Ies2 für die Rauschkontrolle unerlässlich ist.

Meine Ergebnisse zeigen, wie die Chromatinregulation, das Verhalten einzelner Zellen und die Reaktionen von Populationen auf Umweltreize miteinander zusammenhängen. Außerdem habe ich deutlich gemacht, wie wichtig es ist, bei der Beschreibung des Rauschens der Genexpression nicht-stationäre Zustände zu untersuchen.

1. INTRODUCTION

1.1. DNA, nucleosomes, and chromatin organization

DNA contains the genetic information necessary for a cell to function. In eukaryotes, DNA is bound by proteins forming a complex known as chromatin. The first level of chromatin organization is orchestrated by nucleosomes (Fig. 1.1). Nucleosomes are protein octamers composed of two copies each of core histones H2A, H2B, H3, and H4 (Fig. 1.1a,b). DNA is wrapped around nucleosomes [1, 2], and the DNA found in between nucleosomes is often bound by the linker histone H1 [3].

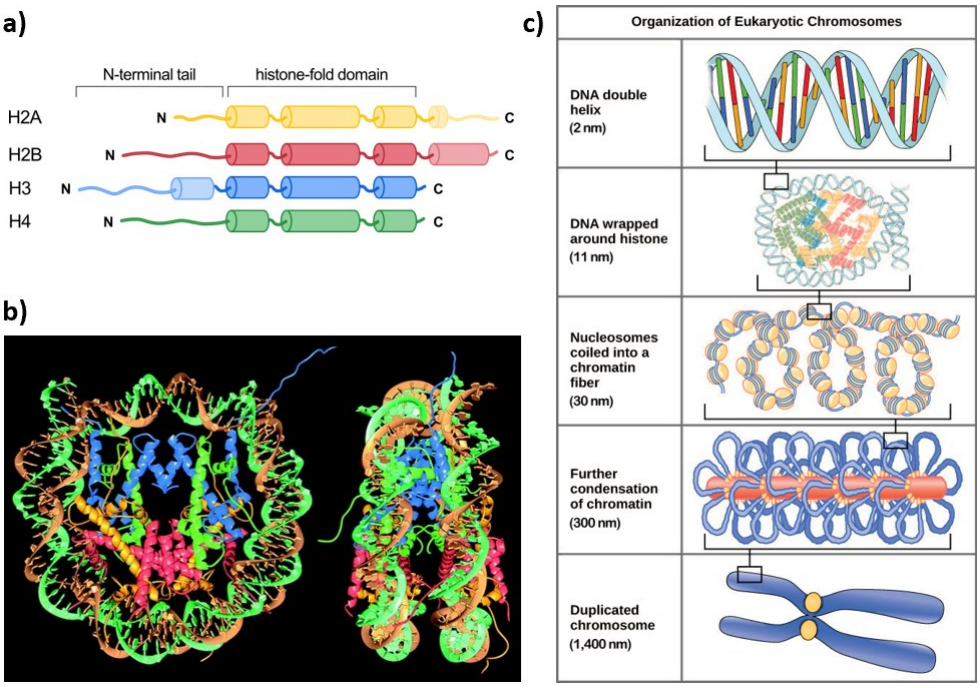


Figure 1.1: Nucleosome composition and chromatin organization. a) Four types of histones form the core nucleosome. Each histone has an N-terminal domain, also referred to as "tail", and a histone fold domain that serves for histones to interact with DNA and with each other. **b)** Front and side view of a nucleosome bound by DNA (orange and green ribbons). A nucleosome core particle is composed of one H3-H4 tetramer and two H2A-H2B dimers. Histone color is shared with panel a. **c)** Different levels of chromatin organization. DNA is wrapped around nucleosomes to form a first level of compaction nicknamed "beads on a string." A second level of condensation is accomplished by the 30 nm chromatin fiber. Ultimately, entire chromosomes achieve different levels of compaction to fit inside the cell's nucleus. Panel a adapted from [4]. Panel b extracted from [1]. Panel c adapted from [5] with information from [6].

Histones are small, positively charged proteins that are highly conserved across eukaryotes. Core histones are composed of a "histone fold domain" that mediates interactions among the histones and contacts the DNA, as well as an N-terminal tail that is subject to extensive post-translational

modifications (Fig. 1.1a). This nucleosomal array forms a 10 nm fiber (Fig. 1.1c), also known as “beads on a string” [4]. Chromatin can compact further to form a 30 nm fiber [7, 8], although the exact molecular mechanism remains still elusive [9]. Further chromatin organization levels include chromatin loops, Topologically Associating Domains, chromatin domains and, finally, chromosome territories [10].

Chromatin can be separated into two classes, heterochromatin and euchromatin, originally discovered as areas of the chromosome with different compaction and sensitivity to cytological stains [11, 12]. Heterochromatin is highly compact and transcriptionally silent, whereas euchromatin’s conformation is more open and is associated with transcriptionally-active regions [13].

1.2. Gene expression

In eukaryotes, transcription is carried out by three different RNA polymerases (RNA Pol I, II, and III), with RNA Pol II being responsible for transcribing most protein-coding genes [4]. In this section I will focus on RNA Pol II, although the molecular mechanism of the other polymerases is similar.

Transcription is divided into three phases: initiation, elongation, and termination. Initiation starts with the recruitment and binding of RNA Pol II to the gene promoter, a DNA region upstream of the transcription start site. RNA Pol II binds the promoter together with general transcription factors, forming an assembly known as the preinitiation complex (PIC) [4]. PIC is enough to transcribe genes *in vitro*, while efficient transcription *in vivo* requires the cooperation of the Mediator complex and gene-specific transcription factors (TFs) [4]. Additional steps during the initiation process, such as chromatin remodeling, are explained below.

In order to enter elongation phase, the carboxy-terminal domain of RNA Pol II must be phosphorylated at Serine 2. This allows for elongation factors and RNA-processing enzymes to be recruited, after which RNA Pol II escapes the promoter and continues the synthesis of messenger RNA (mRNA). The emerging mRNA is capped and, in most cases, spliced [14].

Once RNA Pol II transcribes the poly-A signals located at the end of the gene, the mRNA is cleaved and polyadenylated at its 3’ end during the transcription termination phase [15]. At some genes, RNA Pol

II is not dissociated from the DNA template and is “recycled” for transcribing a second RNA molecule. Once the mRNA has been fully synthesized and processed, it gets exported out of the nucleus, bound by ribosomes, and translated into protein [16]. The process of translation is out of the scope of this introduction.

1.2.1 Transcriptional bursting

A specific feature of eukaryotic gene expression is transcriptional bursting. Bursting refers to the phenomenon in which a gene switches between active and inactive conformations. When in the active or “on” state, the gene is transcribed in a fast and intense fashion, resulting in several molecules of RNA. What follows is a longer period of transcriptional inactivity, also called an “off” state [17, 18]. Two metrics help describe transcriptional bursts: size and frequency. Burst size refers to the number of transcripts synthesized during an active period. Burst frequency quantifies how many active periods occurred in a given amount of time [19].

Bursting patterns are highly gene-specific, and multiple factors are known to contribute to their dynamics. They include cis-regulatory regions and promoter composition, enhancer regions, nucleosome occupancy, and chromatin modifications. Each of these elements may affect burst size, burst frequency, or both [18].

The number and affinity of DNA regulatory elements were shown to correlate positively with burst size [20, 21]. A separate study observed that the presences of TATA boxes at the promoter was also indicative of genes with increased burst size [22, 23]. On the other hand, enhancers act by increasing the bursting frequency of their target genes [24, 25]. The same effect was observed by the absence of nucleosomes at the promoter region [21]. Histone acetylation contributes to increased burst frequency [26, 27], although increased burst size has also been reported [20, 28].

1.3. Chromatin regulation of gene expression

As mentioned above, the eukaryotic genome is wrapped around nucleosomes, which can act as physical barriers to transcription. In this way chromatin imposes a new layer of regulation onto gene expression. Two major strategies exist for chromatin-based regulation: remodeling and covalent modifications.

1.3.1 Chromatin remodeling

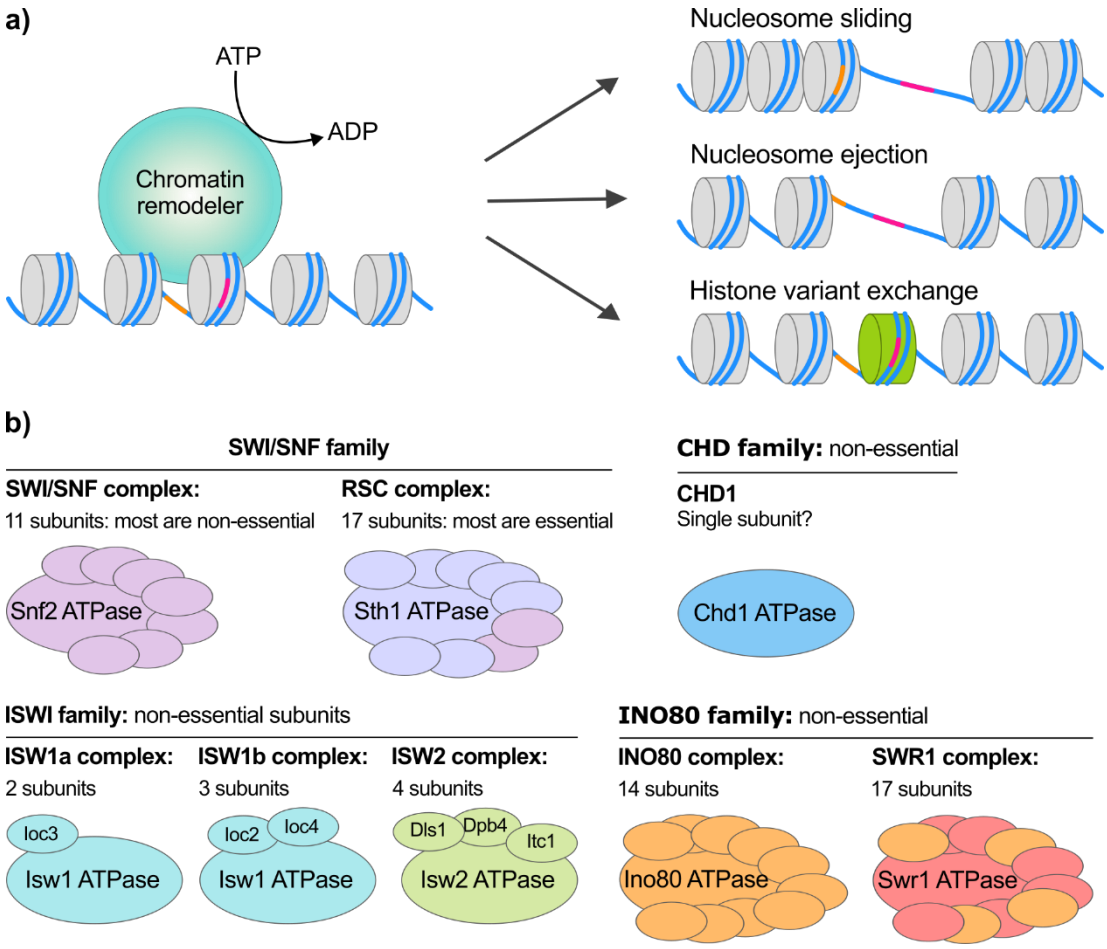


Figure 1.2: Chromatin remodeling and its effectors. a) Using ATP hydrolysis, chromatin remodelers can slide nucleosomes, exchange histones for histone variants, and/or evict nucleosomes. Adapted from [29]. **b)** Four families of chromatin remodelers. Each complex contains several subunits, one of which is an ATPase. Subunits might be shared among different complexes (symbolized by the same color). Adapted from [30].

Chromatin remodelers are large multi-subunit complexes that can move or displace nucleosomes. In doing so, they expose or hide specific DNA regions that act as binding platforms for TFs and/or the transcriptional machinery [31].

There are three mechanisms of chromatin remodeling: nucleosome sliding, nucleosome ejection, and histone variant exchange (Fig. 1.2a). Nucleosome sliding shifts the DNA along the surface of the octamer by hydrolyzing ATP. Nucleosome ejection takes place when an entire nucleosome is fully removed or “kicked out” of the DNA template. Variant exchange occurs when a histone dimer is removed from the nucleosome and replaced with a variant-containing dimer [31, 32].

Chromatin remodeling is extensive at the promoter of eukaryotic genes. For instance, yeast promoters are characterized by a Nucleosome Depleted Region (NDR) followed by precise positioning of nucleosomes upstream and downstream of the promoter [33]. In addition, promoter nucleosomes are enriched in histone variant H2A.Z [32].

Four families of chromatin remodelers can be distinguished based on specific protein domains [31]: SWI/SNF [34], ISWI [35], CHD [36], and INO80 [37]. Each family is divided into multi-subunit complexes of variable composition (Fig. 1.2b). While some complex subunits are indispensable for the remodeler function, other subunits are accessory and instead confer specificity for particular genes or genomic regions [30].

1.3.2 Chromatin modifications

Histones can be post-translationally modified by the addition of (generally small) chemical groups. The most studied modifications are acetylation, methylation, phosphorylation, and ubiquitination [38] (Fig. 1.3). However, recent studies have uncovered dozens of new modifications [39, 40].

Specific histone marks are linked to active transcription. The association of acetylation with increased RNA synthesis was first reported sixty years ago [41]. Since then, the molecular mechanisms of how acetylation exerts its function have been described in more detail. Acetylated chromatin acts mainly as a binding platform for proteins involved with active transcription [42], such as those containing bromo-domains [43]. Acetylation has a minor second role due to its negative charge. When added to histone tails, the acetyl groups neutralize the positive charge of lysine residues, hence reducing the nucleosome interactions with the negatively-charged DNA [44]. Acetyl groups are added by histone acetyltransferases (HATs) and removed by histone deacetylases (HDACs) [42].

The case of methylation is more complex, as methylation has been linked with both active and inactive chromatin [45, 46]. The effect of methylation depends on which amino acid residue is methylated and its position within the histone. For example, methylation of lysine 9 (K9) of histone H3 is linked with heterochromatin [46], whereas H3K4 methylation is associated with active promoters [47]. Methylations are added onto histones by histone methyltransferases (HMTs) and removed by histone demethylases (HDMs) [48].

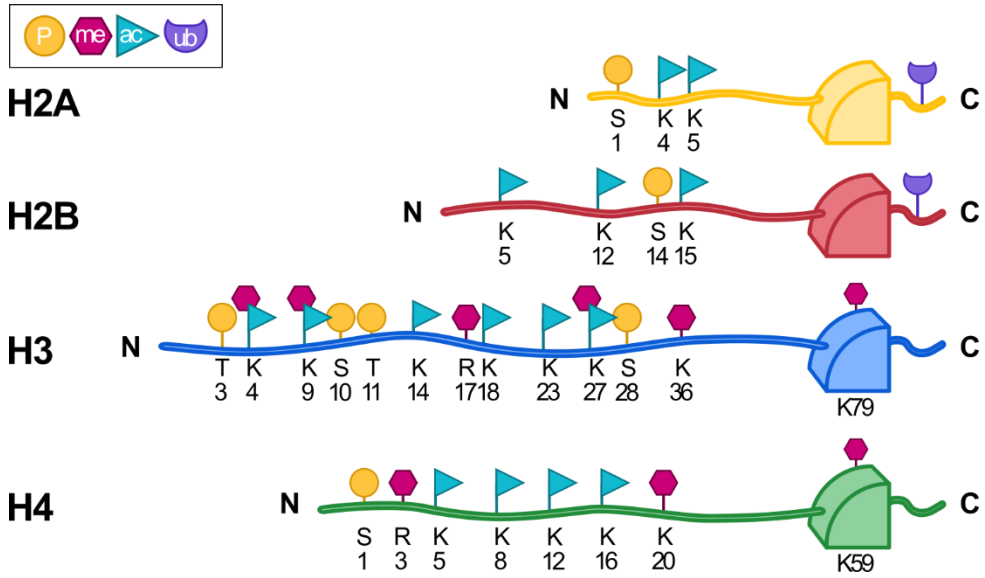


Figure 1.3: Histones and histone modifications. A diagram of selected histone modifications. Histones are mostly modified at the N-terminus tail; however, a few modifications in the histone-fold domain have been characterized. P: phosphorylation; me: methylation; ac: acetylation; ub: ubiquitination. Adapted from [4].

Chromatin modifications and chromatin remodelers may act in concert, as some histone modifications can increase or decrease the remodeling activity of the complexes as well as provide a binding platform for chromatin remodelers themselves [49].

1.4. Transcriptional noise

1.4.1. A brief overview of noise research

As described above, gene expression is regulated by multiple steps, forming a complex and intricated process [50]. It was noted early on that gene expression exhibited variation between cells in homogeneous populations [51, 52]. Ko and colleagues [53] used a system of mouse cells expressing the bacterial β-galactosidase (*lacZ*) under the control of glucocorticoid-response elements. They detected expression of *lacZ* by addition of X-gal [54] and noted that individual cells showed different levels of gene induction (Fig. 1.4a).

A similar observation was made by the Herzenberg lab [57]. They used T-cells expressing *lacZ* under the control of IL2 and NF-AT, two transcription factors indicative of T-cell activation. They measured the expression of *lacZ* by fluorescein-di-β-D-galactopyranoside (FDG) [58] and noted a bimodal response in the population, which did not coalesce even after prolonged hours of induction.

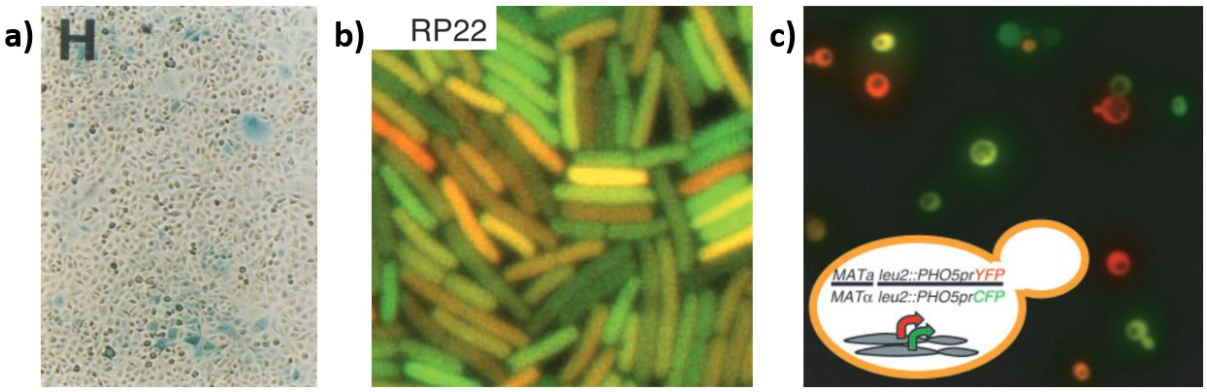


Figure 1.4: Gene expression variation in homogeneous populations. a) Mouse Ltk- cells expressing the bacterial β-galactosidase (*lacZ*) were stained by enzyme histochemistry. Heterogeneous expression of *lacZ* can be seen in blue coloration. Extracted from [53]. **b)** *E. coli* cells expressing CFP (shown in green) and YFP (shown in red) under identical regulatory sequences. Extracted from [55]. **c)** Diploid *S. cerevisiae* cells expressing YFP (shown in red) and CFP (shown in green). Both reporters were inserted in identical loci in homologous chromosomes and expressed by an ectopic PHO5 promoter. Inset shows diagram of the constructs. Extracted from [56]. Figure in (b) from Elowitz, Levine, Siggia, & Swain, 2002. Reprinted with permission from AAAS. Figure in (c) from Raser & O'Shea, 2004. Reprinted with permission from AAAS.

The variability of cells among a population, also referred as "noise", is the consequence of gene expression being governed by molecules present in low concentrations, which makes them subject to stochastic fluctuations [59, 60]. Although the term "noise" has been applied in many fields and different contexts, I will limit its use to populations expressing heterogeneous behaviors under identical conditions.

Investigations on the source and effect of gene expression noise came initially from the fields of network and synthetic biology [60]. Becskei and Serrano [61] studied the ability of negative feedback loops to stabilize genetic circuits. They compared three circuits under different regulation dynamics, all of which had the repressor TetR fused to GFP as a readout. The autoregulated TetR-GFP was expressed under the control of the tetracyclin repressor, while the unregulated circuits had TetR mutated in the DNA binding domain, preventing the binding to its own promoter, or wild-type TetR-GFP expressed under an unrelated promoter (lac operon). Using various biological parameters, the authors concluded that the autoregulated circuit exhibited the least variation, pointing towards negative feedback as a major contributor for noise suppression.

This agrees with conclusions obtained by Elowitz and Leibler in their repressilator [62]. The authors created a synthetic gene circuit by combining three transcriptional repressors (lacI, tetR, and cI) that regulated the expression of each other, creating a cycle of negative feedback. As a reporter, they used GFP under a tet-repressible promoter. The system exhibited regular oscillatory GFP fluorescence, showing that although there is a stochastic component in gene expression, a predictable pattern can be achieved using negative feedback loops.

Noise research saw an upsurge in studies during the early XXI century. Elowitz and colleagues [55] aimed to distinguish between the intrinsic and extrinsic noise of genes. Intrinsic noise refers to that originated by the stochastic nature of biochemical reactions, whereas extrinsic noise is caused by differences in the metabolic state of cells (e.g. cell-cycle stage, cell size, number of ribosomes) [63, 64]. The authors constructed a strain of *Escherichia coli* that contained two different fluorescent reporters (CFP and YFP) under identical regulatory regions. Variation of the reporters inside the cell was a reflection of intrinsic noise, whereas variation among cells of the population was a consequence of extrinsic noise (Fig. 1.4b). The scientists observed low noise levels when the reporters' expression was driven by strong constitutive promoters, and, correspondingly, increased noise when transcription rates were reduced. Under all conditions tested, they also noted that most of the noise was extrinsic in nature.

A different system was implemented by Ozbudak et al. [65]. They created different *Bacillus subtilis* strains with an IPTG-inducible GFP reporter. The gene contained point mutations that altered its transcription (mutations at promoter) or translation (mutations at the Ribosome Binding Site and start codon). The noise of GFP expression was measured in the WT strain under different IPTG concentrations, or in mutant strains with altered translation. The authors concluded that most of the phenotypic noise was a result of variations in translational efficiency, and that gene expression noise was dominated by translational bursts. This was due to the fact that at low transcription levels, noise was increased by high translational rates that created large and infrequent protein bursts. In contrast, the noise of high transcription was buffered at low translational rates, creating only small bursts with minimal fluctuations in protein concentration. The importance of translational bursts became a paradigm for noise in bacteria.

Further studies expanded the research of noise into eukaryotes. Of note are the studies of Blake [66] and Raser [56]. They both used *Saccharomyces cerevisiae* and a fluorescent reporter under the control of an inducible promoter to investigate general properties of noise. The study of Blake [66] used a similar methodology to Ozbudak, as they likewise tested the contributions of transcriptional and translational efficiencies to noise. The authors achieved different transcriptional levels by using an artificial system in which GFP was regulated by a GAL1 promoter (P_{GAL1}) coupled to two *tet* operons. Hence, GFP could be induced by varying amounts of galactose and ATc (anhydrotetracycline; relieves TetR-mediated repression). Contrarily to the observations made in *B. subtilis* [65], the authors found the effect of transcriptional efficiency being prominent in eukaryotes. Furthermore, noise did not scale linearly with increased transcription, but instead was minimal at low transcription, increased when transcription was at 20-40% efficiency, and gradually decreased when the gene was at full induction. Translational efficiency played a role in noise as well, although its contribution was via enhancing the already-present transcriptional noise. The authors deepened their comprehension of eukaryotic noise by performing stochastic simulations that included eukaryote-specific transcriptional features, such as intermediate promoter steps representative of slow chromatin remodeling, binding of TBP (TATA-box-binding protein), promoter reinitiation, among others. Their results indicated that transcriptional reinitiation was a fundamental property of noise modulation.

Raser et al. [56], implemented the dual reporter system developed by Elowitz [55] using the diploid form of *S. cerevisiae*. The scientists integrated CFP and YFP in homologous chromosomes and expressed the fluorescent reporters under three different promoters: P_{PHO5}, P_{PHO84}, or P_{GAL1} (Fig. 1.4c). Using genes controlled by identical regulatory sequences allowed them to distinguish between intrinsic and extrinsic noise. They observed that total noise was dominated by its extrinsic component and that the contributions of intrinsic noise were promoter-specific. The authors further investigated the sources of transcriptional noise by mutating P_{PHO5} or by deleting individual components of chromatin remodeling complexes. Disrupting the UAS (Upstream Activating Sequences) of P_{PHO5} interfered with the removal of positioned nucleosomes [67], generating higher intrinsic noise. In contrast, mutating the upstream TATA box resulted in decreased intrinsic noise. Deletion of chromatin remodeling factors Snf6 (SWI/SNF), Arp8 (INO80), and Gcn5 (SAGA) each resulted in increased intrinsic

noise, although to different extents. Compared to prokaryotes, in which noise seemed to be mostly due to translational effects, the studies of Blake [66] and Raser [56] highlighted how transcriptional noise seemed to be essential in eukaryotes.

Some of the studies mentioned above [55, 56] noticed a linear correlation between the levels of gene expression and noise. This observation could be, however, linked to the specific promoter and experimental setups tested. To investigate that hypothesis, two studies addressed noise on a global scale. Bar-Even et al. [68] examined the noise of 43 *Saccharomyces cerevisiae* proteins under 11 environmental conditions. They confirmed the relationship between mean abundance of a protein and its noise. This was also the case when the genes of interest were measured while undergoing short environmental perturbations. Newman et al. [69] obtained similar conclusions, this time measuring >2,500 proteins of *S. cerevisiae* growing in rich or poor media. They observed that intrinsic noise was inversely proportional to mean protein abundance, particularly for genes of low and medium levels of expression. The relationship between noise and mean gene expression was subsequently confirmed by multiple studies [22, 70–74].

The functional role of a gene was also indicative of its noise level. Genes with essential functions, those involved in protein synthesis and degradation, and those belonging to multi-subunit complexes exhibited low noise, while genes involved in stress-response and environmental adaptation displayed high levels of variation [68, 69, 75].

1.4.2. Tools to measure noise

Because noise is a metric of how variable gene expression is in a population, bulk techniques are not suited for the study of noise. Instead, single-cell methods must be used to dissect cellular heterogeneity [76].

Noise can be measured at the RNA or at the protein level. Early studies of noise, including those mentioned in the previous section, used protein reporters detectable by enzyme histochemistry [53, 57] or by fluorescence microscopy [55, 65, 66, 56]. Nowadays, the use of fluorescent protein reporters is more ubiquitous [77], with detection either by microscopy or by flow cytometry.

The study of RNA at the single-cell level proved to be more challenging. However, the application of Fluorescence *In Situ* Hybridization (FISH) [78–80] to visualize single mRNAs revolutionized the study of transcript variability. Other techniques to monitor RNA expression include the MS2 and PP7 labeling systems, in addition to fluorophore-conjugated aptamers [81]. Finally, single-cell RNA sequencing (scRNA-seq) has also been used in noise studies to address genome-wide trends of transcriptional noise [82–87].

1.4.3. Statistics to measure noise

Multiple metrics exist to quantify noise [88]. They all rely on characterizing the mean abundance of the gene, as well as its standard deviation. The Coefficient of Variation (CV) is defined as the standard deviation divided by the mean of a distribution ($CV = \sigma/\mu$) [89], and has been used extensively to characterize gene expression noise [60, 61, 55, 69, 90].

A widely used alternative metric is CV^2, which is equivalent of dividing the variance of a gene (that is, standard deviation squared) by the mean squared ($CV^2 = \sigma^2/\mu^2$) [56, 65, 66, 68].

Although less used, the Fano factor can also be found in noise research [90–92]. The Fano factor is defined as the variance of a distribution divided by the mean ($F = \sigma^2/\mu$). In that sense, it is equivalent to CV^2 multiplied by the mean, or CV multiplied by the standard deviation.

During this thesis, I have used CV as the metric to quantify heterogeneity.

1.4.4. Noise and transcriptional bursting

The initial noise studies in eukaryotes of Blake [66] and Raser [56] hinted at transcriptional bursting as a key element of gene expression noise. This conclusion was drawn by applying stochastic modeling to their protein-level data. Further research confirmed this hypothesis by looking directly at the cell-to-cell variability of mRNA [79]. The mRNA distributions quantified by Raj et al. [79] did not follow Poisson distributions, implying a coordinated transcriptional activity interspersed with silent periods.

Transcriptional burst dynamics are quantified by burst size and burst frequency [19]. Researchers have shown that increasing the burst size of a gene (for example, by using stronger TF binding sites) results

in higher noise, whereas increasing the burst frequency (by adding nucleosome-disfavoring sequences) reduces the population variability [21].

1.4.5. Noise and chromatin

As discussed above, chromatin regulation is an essential feature for understanding eukaryotic gene expression. As such, it is sensible to hypothesize that chromatin has a substantial impact on noise. Several groups showed that genes exhibiting high noise have heterogeneous accessibility in the chromatin at their *cis*- regulatory sequences [83–86, 93, 94]. In regard to chromatin modifications and noise, simultaneous measuring of gene expression and histone modifications has been more challenging. Nevertheless, it was shown that increasing the chromatin accessibility of a locus via Cas9-targeted histone acetylation decreased the noise levels of a gene without significantly affecting its average level of expression [95]. Variability in other histone modifications has been probed as well. Heterogeneous patterns of H3K4me3 at gene promoters were found to correlate significantly with variability in the expression of their target genes [96]. A separate study identified similar correlations for H3K4me1, H3K27ac, and H3K27me3, and, to a lesser extent, H3K4me3 and H3K9me3 [97].

1.5. The galactose network

Galactose is a sugar that is metabolized via genes belonging to the galactose metabolic pathway (commonly referred to as *GAL* genes). Members of this pathway interact in a complex network in order to import, sense, and convert the galactose to glucose-6-phosphate [98]. The expression of *GAL* genes is dependent on the carbohydrates available in the environment. The presence of galactose is a pre-requisite for the genes to be transcribed, while other sugars – such as glucose – inhibit their expression [99]. Briefly, the galactose-induced genes include GAL1, GAL2, GAL3, GAL7, GAL10, and GAL80 [99]. A description of their roles, as well as their transcriptional regulation, can be found below.

1.5.1. Enzymatic activity of the galactose metabolic pathway

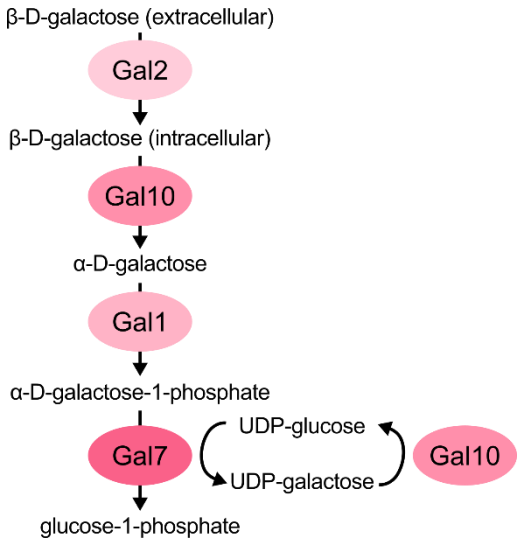


Figure 1.5: Conversion of galactose into glucose by the galactose metabolic pathway. Extracellular β-D-galactose is imported into cells by the membrane-bound permease Gal2. The conversion of intracellular galactose into glucose-1-phosphate is mediated by the enzymes Gal10, Gal1, and Gal7 (details in text). UDP: Uridine diphosphate. Adapted from [100].

First, β-D-galactose is transported by the galactose permease Gal2 [101]. Once in the cytoplasm, the mutarotase activity of Gal10 converts β-D-galactose into α-D-galactose [102], which is subsequently phosphorylated by Gal1 [103]. The phosphorylated galactose is recognized by the uridylyltransferase Gal7, catalyzing its conversion into uridine diphosphate (UDP)-galactose. The final step of the galactose pathway is the conversion of UDP-galactose into UDP-glucose by the epimerase activity of Gal10 (Fig. 1.5) [102].

1.5.2. Transcriptional regulation of the galactose metabolic pathway

The proteins responsible for regulating the expression of *GAL* genes are the activator Gal4, the inhibitor Gal80, and the signal transducer Gal3.

Gal4 is a transcription factor that binds the promoter of *GAL* genes at a sequence called UAS (Upstream Activating Sequence) [104]. Under repressive sugars such as glucose, Gal4 gets bound by Gal80, which blocks the domain of Gal4 that recruits the transcriptional machinery [105], therefore preventing the expression of *GAL* genes (Fig. 1.6a). Besides Gal4 expression being diminished under glucose exposure [106], an additional mechanism mediated by Mig1 ensures full repression when glucose is present (Fig. 1.6a). Mig1 binds the promoter of *GAL* genes and recruits the corepressor complex Tup1-Cyc8 [107],

which in turn interacts with histone deacetylases and promotes a repressive chromatin environment [108].

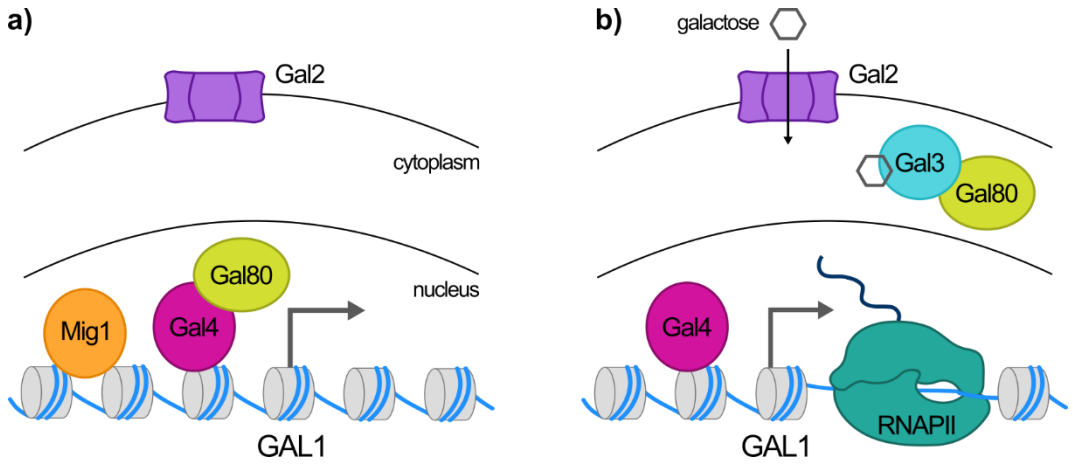


Figure 1.6: Schematics of the GAL1 locus in repressing and inducing conditions. a) In repressing sugars such as glucose, Gal80 blocks the activating domain of Gal4, thereby impeding the recruitment of the transcriptional machinery. A second repressive mechanism is carried out by Mig1 (see text for details). **b)** Galactose is transported inside the cells by Gal2. Intracellular galactose triggers the hijacking of Gal80 by Gal3, releasing the repression on Gal4. In consequence, Gal4 recruits RNAPII and associated transcriptional activators. RNAPII: RNA polymerase II. Adapted from [109].

In the presence of galactose, the Gal4-Gal80 complex is disrupted by Gal3, which sequesters Gal80 and allows Gal4 to exert its activating role (Fig. 1.6b) [110]. The interaction between Gal3 and Gal80 is galactose- and ATP-dependent [110, 111], although the location and nature of their interaction is controversial. Potential mechanisms include i) sequestering of cytoplasmic Gal80 [112–114], ii) sequestering of nuclear Gal80 [114, 115], iii) sequestering of nuclear Gal80 and subsequent export of Gal3-Gal80 to the cytoplasm [116], and iv) Gal3 inhibiting the self-association of Gal80 necessary to bind Gal4 [117]. The presence of a short-lived tripartite complex between Gal4, Gal80, and Gal3 has likewise been reported [110, 118].

While glucose represses the transcription of *GAL* genes, other carbon sources such as raffinose and glycerol maintain the *GAL* genes in a non-induced state [98, 119].

1.5.3. Positive and negative feedback loops regulate transcription of *GAL* genes

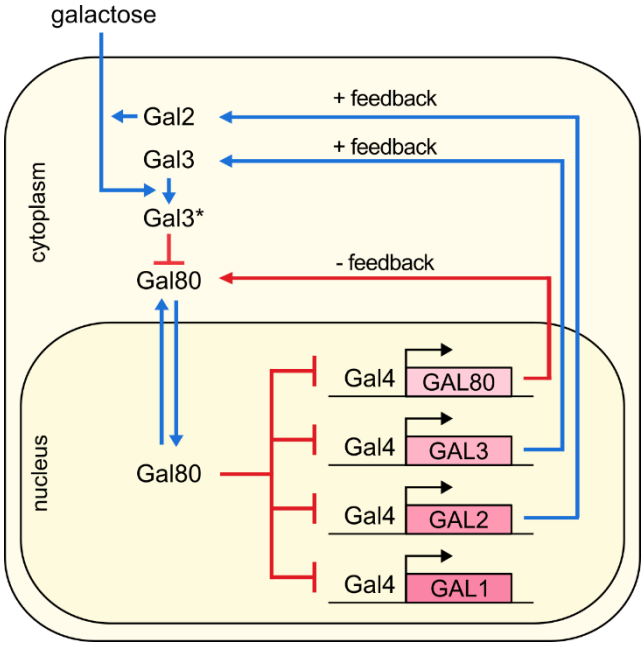


Figure 1.7: Regulation of the galactose metabolic network. Gal4 is the master activator of the network, and it induces the transcription of GAL80, GAL3, GAL2, and GAL1. Gal80 inhibits the activity of Gal4, symbolized by blunt arrows. Galactose-associated Gal3 (represented as Gal3*) prevents the inhibitory effect of Gal80. See text for details. Adapted from [120].

The *GAL* genes are regulated under multiple positive and negative feedback loops (Fig. 1.7). Positive loops are achieved by Gal3, Gal1, and Gal2, whereas Gal80 mediates the negative feedback [120].

The gene GAL3 contains a UAS in its promoter, which makes it subject to the transcriptional control of Gal4. Higher concentrations of Gal3 counteract the Gal80-mediated repression on Gal4, leading to an increased transcription of *GAL* genes [121]. In this way Gal3 stimulates its own expression.

The positive loop mediated by Gal1 follows the same rationale. Gal1 and Gal3 are paralogs that arose during the whole genome duplication of *Saccharomyces cerevisiae* [122]. They share 73% amino acid identity and 92% similarity [123]. GAL1 is also regulated by Gal4, and it has been shown that once the concentration of Gal1 is high enough, it can substitute the activity of Gal3 [121]. In fact, Gal1 overexpression restores the transcription of *GAL* genes in yeast that lack Gal3 [124].

As shown in Fig. 1.6, Gal2 is a membrane protein that transports galactose inside the cell. Intracellular galactose is necessary for the formation of the Gal3-Gal80 complex [110, 111]. This interaction relieves

the repression on Gal4, which simulates the transcription of *GAL* genes, including GAL2. Higher levels of Gal2 increase the levels of intracellular galactose, creating a positive feedback [120].

The negative loop of the galactose metabolic pathway is orchestrated by Gal80. In inducing conditions, Gal4 stimulates the transcription of GAL80. An excess of Gal80 counteracts the sequestering by Gal3 [120]. This leads to the formation of Gal4-Gal80 complexes, which repress the expression of GAL80.

1.6. Previous results on which my doctoral project was build

My doctoral dissertation is a follow-up of the work published in Bheda et al [125]. This work included a Gal1 induction screen, in which the authors monitored Gal1-GFP induction kinetics in single *Saccharomyces cerevisiae* cells over time in wild-type (WT) cells and in 535 strains with single-gene deletions of non-essential chromatin-related factors (Fig. 1.8a, Table S1). To identify chromatin factors that are implicated in transcriptional noise during gene induction, one of the authors (Dr. Johannes Becker) re-analyzed the Gal1 induction data for strains with altered gene expression noise. To measure gene expression heterogeneity, he calculated the Coefficient of Variation (CV = standard deviation/mean) of Gal1-GFP fluorescence intensities and selected strains with an outlier CV as described below.

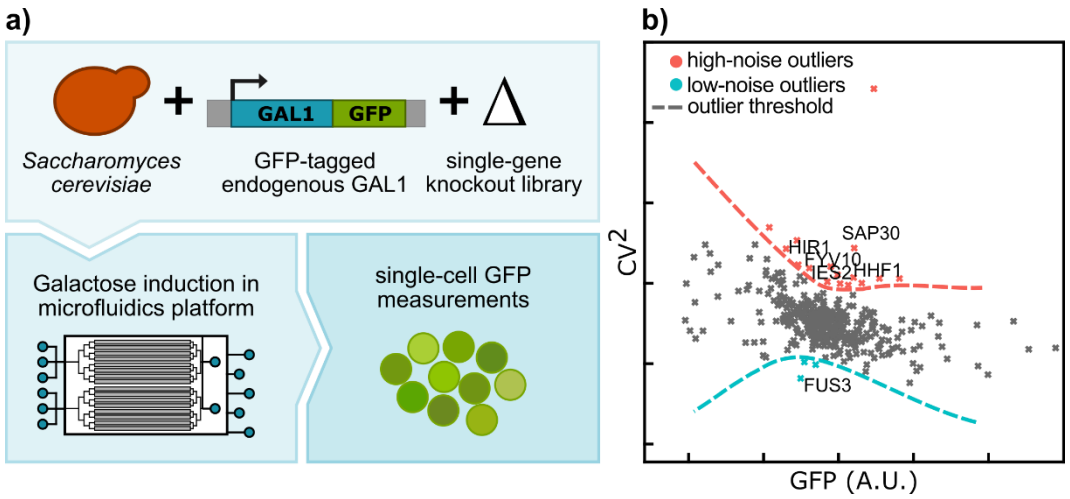


Figure 1.8: High-throughput screening reveals chromatin factors with altered gene induction noise. **a)** Schematics of Gal1 induction screening by Bheda et al. [125]. The endogenous Gal1 protein was fused to superfolder GFP. This reporter strain was crossed with a yeast strain library of 535 single-gene deletions of non-essential chromatin-related factors. Cells were grown in a microfluidics platform and Gal1-GFP expression was monitored by live-cell imaging over two rounds of galactose induction and repression totaling 11 hours. Cells were segmented, tracked, and their fluorescence was quantified to calculate the Coefficient of Variation (CV = standard deviation/mean) of Gal1-GFP signal intensities. **b)**

Example of outlier identification for deletion strains with increased or decreased Gal1 noise, quantified by the CV of the GFP signal. The data depicted in this plot corresponds to frame #37 of the time course performed by Bheda et al. [125]. Plotted in y-axis is the CV^2 at a specific time point during galactose induction against the mean GFP fluorescence signal of the population (x-axis). Every data point corresponds to one strain. Upper and lower dashed lines represent the outlier thresholds calculated based on the interquartile range of a moving window of 11 data points (see Materials and Methods). Outlier strains with CV^2 higher or lower than the outlier threshold are depicted in red or blue, respectively. A strain would be considered outlier if it was classified as such in 4 or more consecutive frames. Analysis performed by Dr. Johannes Becker.

Instead of doing hundreds of mutant-WT comparisons, the outlier detection method relied on the assumption that most of the mutant strains would have a modest effect on gene expression noise, if at all. He therefore focused on those mutants whose CV fell consistently outside of the behavior displayed by the majority of the strains. For every time point of the galactose induction, upper and lower thresholds of noise would be established based on the interquartile range of the CV^2 distribution (Fig. 1.8b; for details see Methods). This revealed 28 strains with an outlier CV^2 for at least 4 consecutive frames (22 higher noise, 6 lower noise) of 10 minutes each. From the identified outliers, I selected 9 chromatin deletions (6 higher, 3 lower noise) to investigate in my doctoral dissertation (Table S2).

2. AIMS OF THE STUDY

Although gene expression is a highly regulated process, it is subject to cell-to-cell variation due to the stochastic nature of biochemical reactions. This variability is referred to as “noise” and, although it has been a topic widely studied, research so far has mostly focused on the noise of genes at steady-state.

There is a lack of understanding of how noise can evolve over time, particularly in genes undergoing transcriptional activation. The aim of my dissertation was to study gene expression noise in dynamic systems and analyze the impact of chromatin-related factors on noise during gene induction.

Specifically, the goals were:

- Quantify the gene expression noise under dynamic environmental conditions, specifically gene induction due to changes in carbohydrate source using single-cell techniques.
- Identify chromatin-related factors whose absence led to altered induction noise.
- Determine and quantify gene noise at the transcript and protein levels.
- Characterize the molecular mechanisms by which one of the chromatin-related factors affects noise.

3. RESULTS

3.1. Gal1 noise during induction is dependent on carbohydrate source

Different induction patterns have been characterized for *GAL* genes, in particular Gal1. Induction is unimodal when cells are grown in a non-repressing, non-inducing sugar (such as raffinose or glycerol) prior to switching to galactose [126, 127]. In contrast, induction is bimodal when transitioned from glucose into galactose [128, 127]. Thus, Gal1 induction depends on pre-growth media, induction media, and time of induction [120, 129, 130], exemplifying how finely-tuned the *GAL* network is to the environment [131].

To select the most appropriate growth conditions for the experiments and study the potential gene expression noise mutants, I first measured Gal1 noise under different combinations of sugars for pre-growth and induction (Fig. 3.1). For pre-growth I tested glucose, which represses the GAL1 promoter, and raffinose, which allows for basal GAL1 expression [99, 132]. For GAL1 induction, I tested galactose and a mixture of raffinose plus galactose (Fig. 3.1a). The four conditions gave rise to different induction kinetics (mean Gal1 expression vs. time) (Fig. 3.1b) and noise trajectories (CV vs. mean expression) (Fig. 3.1c).

Galactose induction of cells pre-grown in glucose resulted in a bimodal Gal1-GFP intensity distribution, in which a subpopulation of cells induced faster than the others, until ultimately the whole population reached an induced state (Fig. 3.1a). The bimodality occurred irrespective of the induction sugar (galactose or raffinose + galactose) but only after pre-growth in glucose. In contrast, the induction of cells pre-grown in raffinose is rather unimodal, giving rise to a population with lower heterogeneity. Populations that were induced with raffinose + galactose reached full Gal1 expression levels around 6h after media switch. In contrast, cells induced only with galactose expressed Gal1 slower and only reached steady-state after 8h of induction (Fig. 3.1b). The time required to reach steady-state was independent of the pre-growth sugar.

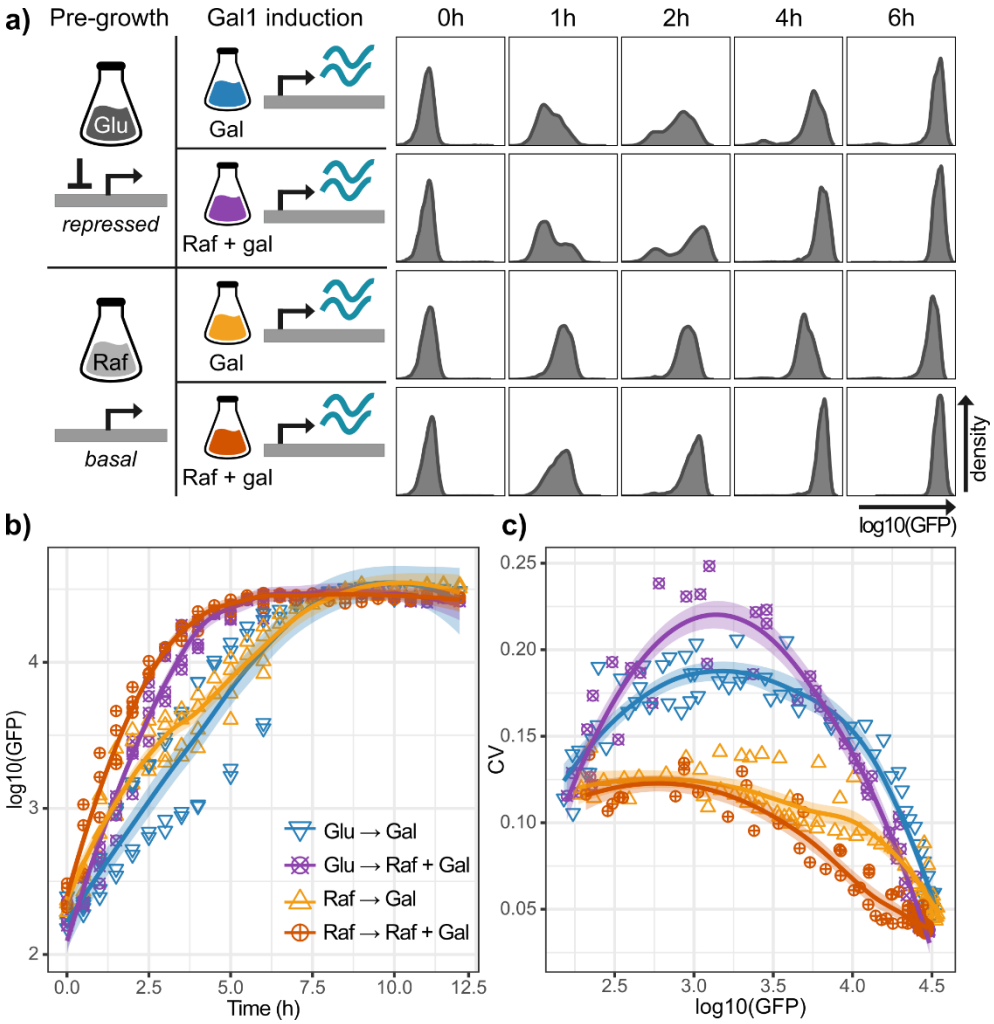


Figure 3.1: Analysis of Gal1 induction dynamics under different carbohydrate sources. WT cells containing a Gal1-GFP fusion at the endogenous GAL1 locus were grown and induced with multiple carbohydrate combinations. Cells were switched from glucose pre-growth to galactose induction (blue), from glucose to raffinose + galactose (purple), from raffinose to galactose (yellow), or from raffinose to raffinose + galactose (orange) in 12-hours induction with sample acquisition every 30 minutes. Glu: glucose; Raf: raffinose; Gal: galactose. Statistics were calculated on merged data from four biological replicates. The average number of cells per data point was 8,838 ± 347 (mean ± sd). Line and ribbon represent the fit and 95% confidence intervals of a curve fitting with *loess*, respectively. **a)** Left: experimental setup. Right: Histograms of a representative experiment showing population distributions of Gal1-GFP at indicated induction time points as measured by flow cytometry. **b)** Mean expression of Gal1-GFP during galactose induction. Plotted is mean log10(GFP) signal vs. time for the indicated sugars. **c)** Noise of Gal1-GFP over different levels of expression, quantified by the Coefficient of Variation (CV = standard deviation/mean). Plotted is CV vs. mean log10(GFP).

The noise trajectory (CV vs. mean expression) of Gal1 during induction of glucose pre-grown cells is semi-circular, starting from a homogeneous non-expressing population (low noise), passing through a bimodal state with a wide range of Gal1-GFP levels (high noise), and ending with a homogeneous, all-induced population (low noise) (Fig. 3.1a,c). The semi-circular shape of the noise trajectory was

irrespective of induction media. In contrast, the raffinose pre-grown cultures exhibit a decreasing noise trajectory, in which the CV is highest at low levels of induction, and continuously decreases as the induction progresses (Fig. 3.1c).

Given that Gal1 exhibited the highest noise when cells were pre-grown in glucose (Fig. 3.1c), I decided to further characterize high and low noise outliers during galactose induction starting from a glucose-repressed state. As for Gal1 induction, I chose galactose alone as a simplified version compared to the raffinose + galactose mixture.

3.2. Identification of chromatin factors that affect Gal1 noise during induction

For the 6 higher and 3 lower noise outliers selected for further analysis (see Table S2), Dr. Poonam Bheda generated de-novo knockout strains to avoid possible off-target mutations from the yeast deletion library used in the screen. I then pre-grew these cells in glucose, followed by a 12-hour galactose induction during which I monitored Gal1-GFP expression by flow cytometry. As shown in Fig. 3.2, all strains exhibited a semicircular noise trajectory.

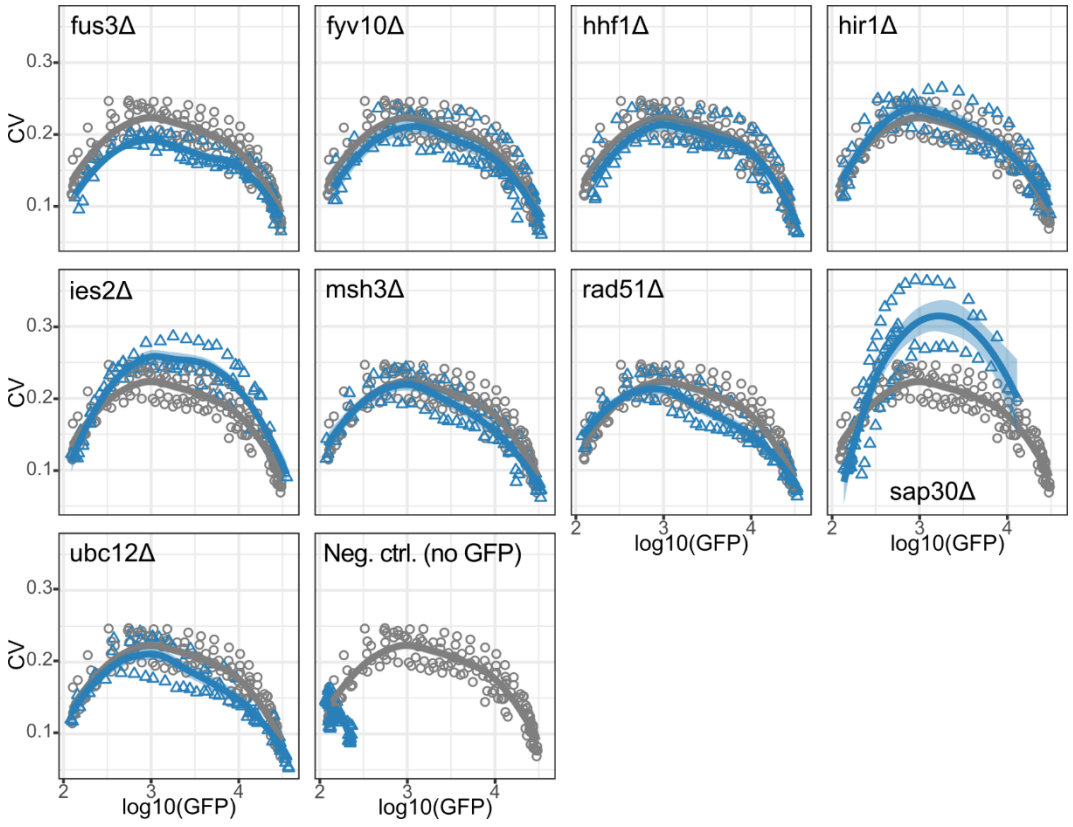


Figure 3.2: Noise trajectories of selected mutants during 12 hours of galactose induction. Noise of Gal1-GFP over different levels of expression, quantified by the Coefficient of Variation (CV = standard

deviation/mean). Plotted is CV vs. mean log10(GFP). Gray: WT strain; blue: indicated knock-out strain. Line and ribbon represent the fit and 95% confidence intervals of a curve fitting with *loess*, respectively. Statistics were calculated on merged data from three independent biological replicates. WT strain was processed in duplicate within each replicate. The average number of cells per data point was 8,171 ± 950 (mean ± sd). The time course spanned 12 hours of induction with sample acquisition by flow cytometry every 30 minutes. Negative control is WT strain without the GFP fusion.

These experiments confirmed the higher Gal1 noise in two strains (ies2Δ and sap30Δ) and lower noise in two strains (fus3Δ and msh3Δ) during the 12-hour galactose induction (Fig. 3.3). Interestingly, the differences in noise between mutants and WT were only detectable during induction, but disappeared once steady-state Gal1 expression was reached. Thus, with the exception of Sap30 (see Discussion), these mutants had not been identified as noise outliers before. Similar to WT, the mutant strains achieved a minimal level of noise once Gal1 reached maximum expression.

Comparing populations of cells with similar mean expression levels (dotted line in Fig. 3.3b) showed that changes in CV were related to changes in the shape of the bimodal distributions (Fig. 3.3c). Mutants with CV lower than WT (fus3Δ and msh3Δ) had simultaneously a smaller fraction of uninduced cells and induced cells with lower GFP expression. The contributions of both subpopulations generated a mean Gal1 expression similar to WT, although with a smaller distance between uninduced and induced peaks. This reduced gap between subpopulations lead to a smaller standard deviation compared to WT and, consequently, a lower CV.

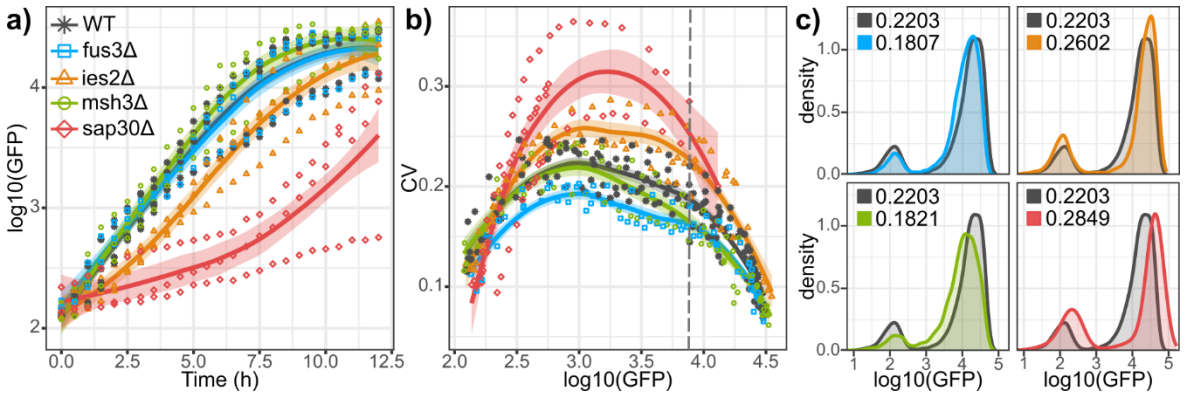


Figure 3.3: Gal1 expression noise in fus3Δ, ies2Δ, msh3Δ, and sap30Δ strains. Panels a) and b) represent statistics calculated on merged data from three independent biological replicates. WT strain was processed in duplicate within each replicate. The average number of cells per data point was 8,145 ± 1,020 (mean ± sd). Line and ribbon represent the fit and 95% confidence intervals of a curve fitting with *loess*, respectively. Strains depicted are WT (gray), fus3Δ (blue), ies2Δ (orange), msh3Δ (green), and sap30Δ (red). **a)** Mean expression of Gal1-GFP over time for WT and selected mutant strains. Plotted is mean log10(GFP) signal vs. time. **b)** Noise trajectories of Gal1-GFP over different levels of

induction. Plotted is CV vs. mean log10(GFP). Dotted line indicates samples with similar mean expression. **c)** Overlay of log10(GFP) distributions of populations at similar Gal1-GFP levels (corresponding to dotted line in b). Numbers denote the CV of each population.

The opposite observation was true for high-noise outliers. Both ies2Δ and sap30Δ strains had a higher proportion of cells in the uninduced fraction, while the expressing population exhibited higher Gal1-GFP levels than WT. The depletion of cells with intermediate levels of expression and increase in cells with either null or high Gal1 levels created a population of high heterogeneity, resulting in a higher CV compared to WT (Fig. 3.3c).

3.3. Different Gal1 noise trajectories also occur at the mRNA level

Since so far I focused on Gal1-GFP protein levels, I decided to study the corresponding mRNA next. For this, I measured Gal1-GFP mRNA levels of individual cells using a Single Molecule Fluorescence *In-Situ* Hybridization (smFISH) protocol with Quasar 670-tagged probes complementary to the Gal1 mRNA sequence (Fig. 3.4). I simultaneously imaged GFP and Quasar 670 signals at different time points to directly compare protein and transcript expression during galactose induction (Fig. 3.4a). Consistent with results shown in previous sections, Gal1 induction in WT cells was bimodal both at the mRNA (Fig. 3.4b) and protein (Fig. 3.4c) levels. This was also the case for the mutants I studied (Fig. S1).

The relationship between Gal1 transcript and protein expression showed not to be linear, but instead followed an L-shaped pattern over time (Fig. 3.4d). At early time points, measured Quasar 670 and GFP intensities corresponded to background fluorescence. Transcription of Gal1 mRNA initiated as the induction progressed, observed by a higher Quasar 670 signal after 30 minutes (Fig. 3.4b,e). GFP fluorescence only increased after 1.5 hours, indicating translation of the Gal1-GFP mRNA (Fig. 3.4c,f). The coupling of both events can be visualized as a two-step process in which mRNA transcription precedes translation (Fig. 3.4d). Initially, Quasar 670 signal increased while GFP levels remained constant, represented by a horizontal shift in the early time points of Fig. 3.4d. It is only after Gal1 translation initiated that cells increased their GFP intensity, reflected by a vertical shift during the last few hours of the galactose induction.

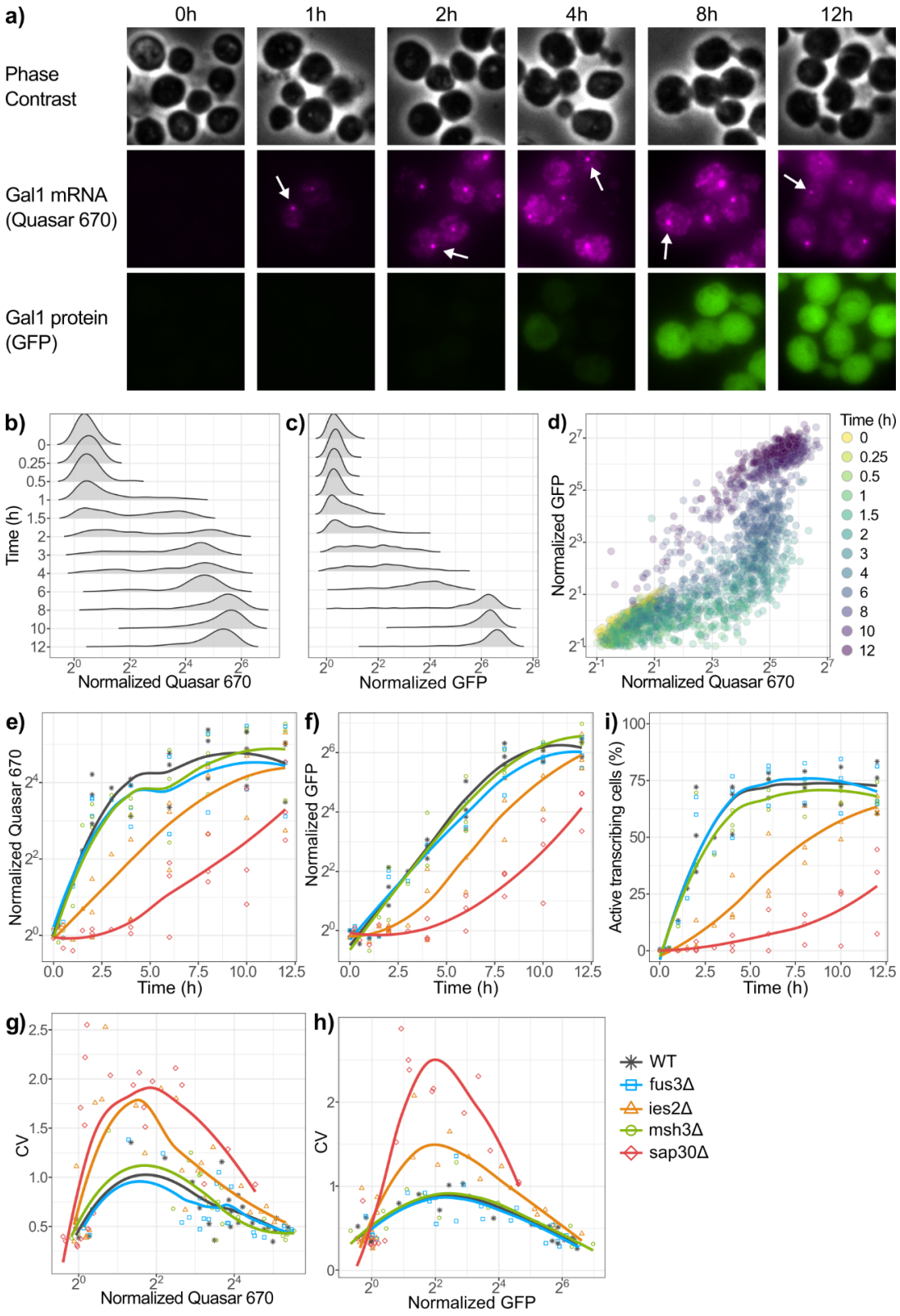


Figure 3.4: Quantification of Gal1 mRNA by FISH and Gal1 protein by fluorescent reporter. Cells undergoing galactose induction were subjected to a Single Molecule Fluorescence *In-Situ* Hybridization (smFISH) protocol and subsequently imaged by epifluorescence microscopy. The endogenous GFP-tagged Gal1 transcripts were labeled with Quasar 670-conjugated probes complementary to Gal1 mRNA sequence. **a)** Representative images of WT cells during Gal1 induction. Time indicates hours after switching to galactose. Top panel: phase contrast images. Middle panel:

maximum projection of z-stacks in Quasar 670 channel. White arrows point to examples of active GAL1 transcription sites. Bottom panel: images in GFP channel. **b,c,d)** Single-cell measurements of WT at different induction time points. Plotted are (b) normalized Quasar 670 intensities corresponding to Gal1-GFP transcripts, (c) normalized GFP intensities corresponding to Gal1-GFP proteins, and (d) relationship between both fluorescent signals over time from one representative replicate. In (d), each data point represents an individual cell, and the color indicates the number of hours after galactose induction. **e,f)** Mean population fluorescence of (e) Quasar 670 and (f) GFP intensities during Gal1 induction for indicated strains. **g,h)** Noise trajectories across different levels of expression for Gal1-GFP (g) mRNA and (h) protein. **i)** Average percentage of cells exhibiting an active GAL1 transcription site. Normalized fluorescence intensities are quantified as the fold-change vs. WT at 0h of induction (see Materials and Methods). Statistics (panels e-i) were calculated on merged data from three biological replicates (one replicate 0-12h induction, two replicates 2-12h induction). Line represents curve fitting with *loess*. The average number of cells per data point was 311 ± 154 (mean ± sd). Panels e-i share color scheme: WT (gray), fus3Δ (blue), ies2Δ (orange), msh3Δ (green), and sap30Δ (red).

Similar to the flow cytometry measurements (Fig. 3.3), fluorescent microscopy detected different Gal1-GFP protein noise trajectories between WT and some of the mutants (Fig. 3.4h). Interestingly, this was also observed for Gal1-GFP mRNA (Fig. 3.4g), suggesting that regulation at the transcript level contributes to the higher or lower noise observed during gene induction (see Discussion).

Our smFISH protocol allowed to detect actively transcribed genomic sites. These loci are distinguishable from individual mRNAs due to their increased brightness (Fig. 3.4a), a consequence of the smFISH probes being bound to several simultaneously transcribed RNA molecules. I identified different transcription activation dynamics between WT and mutant strains (Fig. 3.4i). WT, fus3Δ, and msh3Δ populations exhibited active GAL1 transcription sites in up to 84.5% of cells. This percentage of transcribing cells stabilized after 4 hours of induction. In contrast, ies2Δ and sap30Δ populations displayed a slower activation of the GAL1 locus. The dynamics of GAL1 locus transcription resembled the increase in signal fluorescence for Gal1-GFP mRNAs during induction (Fig. 3.4e).

Overall, the chromatin factor mutants display altered mRNA and protein heterogeneity in a dynamic gene expression system. I also noted that differences in CV disappeared once the population reached steady-state expression of Gal1, pointing toward differences in the contribution of chromatin factors to transcriptional noise during gene induction.

3.4. Mathematical modelling suggests altered DNA activation and inactivation rates for mutants with higher gene expression noise

Using the data from the smFISH experiments, Dr. Lea Schuh and Dr. Dantong Wang (Institute of AI for Health, Helmholtz Munich) developed a mathematical model to identify which steps during gene induction might be responsible for the different noise trajectories observed in mutant and WT populations (Fig. 3.5). Dr. Schuh and Dr. Wang established, troubleshooted, and refined the model, while I performed the final run of results featured in this thesis.

The model architecture considered DNA_{off} and DNA_{on} states, in which transcription is repressed and permissible, respectively. They included multiple DNA intermediate steps to represent regulatory events of chromatin opening up for gene expression [17]. Additionally, the model allows transcriptional bursting, in which cells can turn "on" and "off" a gene by alternating between DNA_{on} and DNA_{off} states. The model consists of six different parameters grouped in three pairs: DNA activation/inactivation (γ_{on}/γ_{off}), mRNA production/degradation ($\lambda_{prod}/\lambda_{deg}$), and protein production/degradation ($\kappa_{prod}/\kappa_{deg}$) (Fig 3.5a).

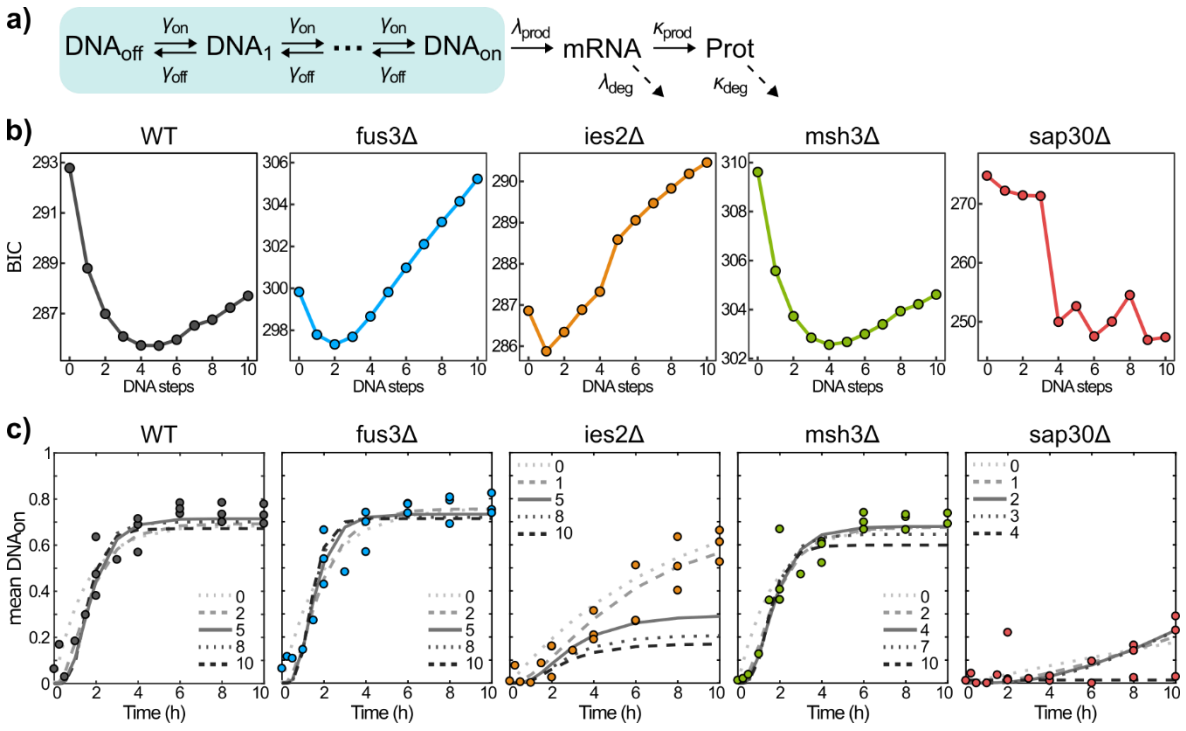


Figure 3.5: Mathematical modelling identifies deletion-specific gene expression parameters. a) Depiction of gene expression model. Cells alternate between DNA_{off} and DNA_{on} states, symbolized by arrows. The model evaluates potential intermediate steps between DNA_{off} and DNA_{on}. DNA_{on} allows

expression of mRNAs, which are subsequently translated into proteins. Dashed arrows represent molecule degradation. See text for parameter details. Prot: protein. **b)** Bayes Information Criterion (BIC; y-axis) calculated for models with 0-10 intermediate DNA steps (x-axis). The best model corresponds to the lowest BIC. **c)** Fits for the fraction of cells in the DNA_{on} state (y-axis) as a function of time spent in galactose (x-axis). The number in the legend specifies the number of intermediate steps used in each fit. Every data point corresponds to the statistic calculated for an individual biological replicate.

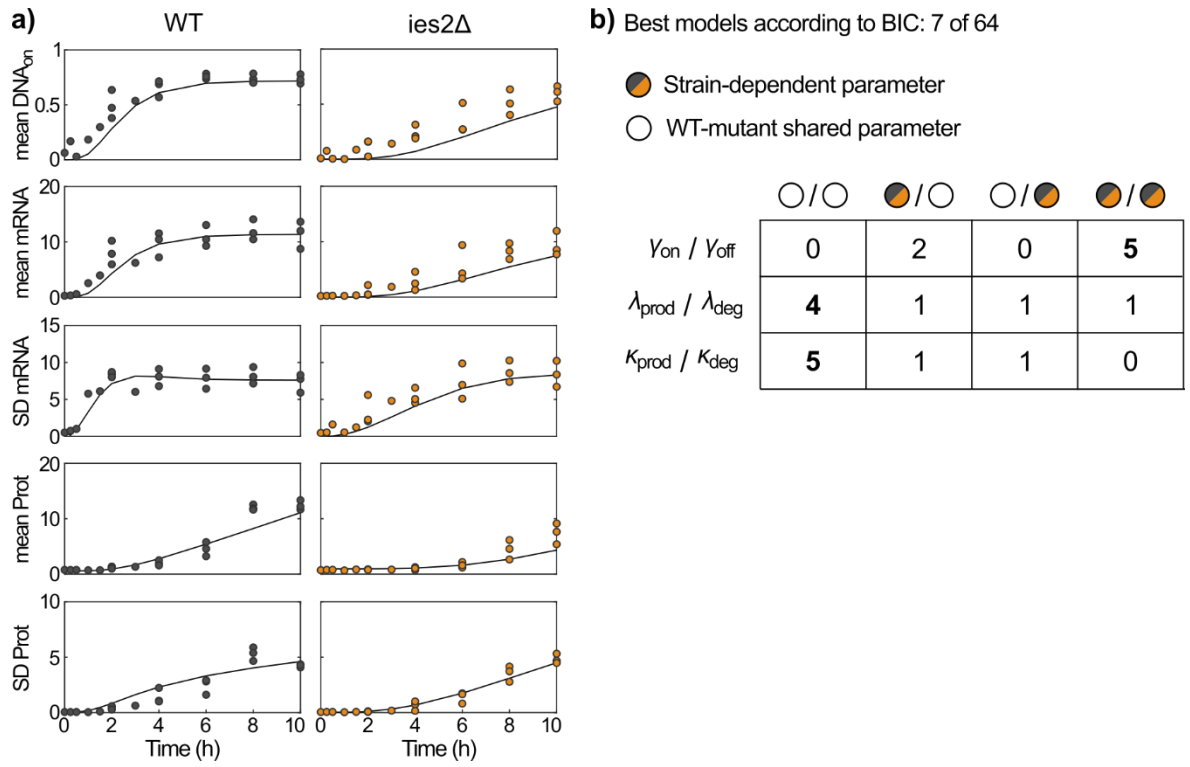
The mathematical model was fitted using the moments equations (mean and standard deviation) of Gal1-GFP mRNA and protein, as well as the percentage of the population in the DNA_{on} state, calculated as the fraction of cells with an active transcription site (see Fig 3.4a,i). The best models were determined as those with a Bayesian Information Criterion (BIC) [133] 10 units or less away from the lowest calculated BIC. A threshold of 10 units was chosen as a commonly-adopted limit for model selection [134–137].

They first determined the number of intermediate steps between DNA_{off} and DNA_{on} (Fig. 3.5b,c). They considered between 0 and 10 steps and calculated the BIC for each model. The best fits were obtained when considering 1-5 intermediate steps for WT, fus3Δ, ies2Δ, and msh3Δ. The optimal number of intermediate steps for sap30Δ was inconclusive, as the BIC displayed a non-continuous behavior (see Discussion).

We selected 3 intermediate DNA steps as a denominator common to all strains, determined by calculating the average of the best number of steps for WT, fus3Δ, ies2Δ, and msh3Δ (5, 2, 1, and 4 steps, respectively). Although 3 steps were not the optimal for any individual strain, the BIC of the model was less than 10 units away from the lowest BICs, indicating no strong evidence to reject the model. Furthermore, the percentage of sap30Δ cells in the DNA_{on} state could be adequately modeled with 3 intermediate steps (Fig. 3.5c).

To identify differences in expression dynamics between strains, we estimated the value of the six parameters that form the model (γ_{on}, γ_{off}, λ_{prod}, λ_{deg}, κ_{prod}, κ_{deg}; see Fig. 3.5a) by fitting WT and each mutant strain in pairs (Fig. 3.6a and Fig. S2). Because every rate can either be identical or different between WT and mutant, $2^6 = 64$ models are to be considered (Fig. S2a). Any model with a BIC equal

to or inferior to the lowest BIC plus 10 units was considered a god fit and became part of the set of "best models" (Fig. 3.6b and Fig. S2b).



	○/○	◐/○	○/◐	◐/◐
γ_{on} / γ_{off}	0	2	0	**5**
λ_{prod} / λ_{deg}	**4**	1	1	1
κ_{prod} / κ_{deg}	**5**	1	1	0

Figure 3.6: Model comparison for WT and ies2Δ populations. a) Model fit for WT and ies2Δ populations. From top to bottom, plots show: fraction of cells in the DNA_{on} state; mean Gal1-GFP mRNA; standard deviation (SD) of mRNA; mean Gal1-GFP protein; SD of protein. **b)** 7 models were identified as having the best fit for the WT-ies2Δ pair, which were classified into different categories according to whether the strains shared parameter rates or not. In bold is the category with the highest value.

Fig. 3.6 shows a summary of the results for WT-ies2Δ comparison. Out of the 64 models tested, 7 were considered as the "best models". The majority of them supported a scenario in which λ_{prod}, λ_{deg}, κ_{prod}, and κ_{deg} rates were identical between WT and mutant. However, none of them supported shared γ_{on}, γ_{off} rates between the WT and ies2Δ strain (Fig. 3.6b). This is a strong indication that WT and ies2Δ diverge in their DNA activation/inactivation dynamics, an observation already suggested in Fig. 3.4i. A similar conclusion was obtained for sap30Δ, in which 13 out of 21 best models suggested strain-specific γ_{on} and γ_{off} parameters (Fig. S2b). These results point towards potential mechanisms for these high-noise outliers.

The model comparison for fus3Δ and msh3Δ revealed the best model for both strains had all six rates identical between WT and mutant (Fig. S2b). These results could in part be attributed to the limitations of the FISH data (see Discussion) (Fig. 3.4). Although fus3Δ and msh3Δ were shown to be low-noise outliers by flow cytometry, FISH and microscopy imaging were unable to capture our previous observations (see Discussion).

3.5. Gal1 noise during induction is dependent on growth phase and metabolic state

While the validation of outliers was performed in cultures undergoing a direct switch to galactose after glucose pre-growth, later experiments showed reduced Gal1 induction and high inter-experiment variability. In order to improve reproducibility, I took a step back and optimized the galactose induction protocol.

Glucose is the preferred carbon source of *S. cerevisiae*. Yeast metabolizes this sugar by glycolysis and produces ethanol during the process. Once glucose becomes depleted, cells undergo a diauxic shift in which cell growth is reduced and metabolism changes to aerobic respiration [138]. Instead of performing galactose induction in cultures that underwent diauxic shift and reached stationary state, I added an extra growth step in fresh glucose (Fig. 3.7). I refer to this step as “intermediate growth”, and it comprises diluting cells from the pre-growth culture into fresh glucose prior to galactose switching. In this way, cells are allowed to exit stationary phase, transition back to aerobic fermentation, and resume division before being subjected to the metabolic reprogramming associated with galactose induction.

To test how long this intermediate growth should be, I grew WT cells in glucose overnight, diluted them in fresh glucose for 2, 4, or 6 hours, and proceeded to galactose induction as previously described (Fig. 3.7a). The OD_{600} of cultures was kept below 0.5 to ensure stationary phase was not reached. A control sample in which cells did not go through the intermediate growth was included.

Irrespective of the presence or absence of an intermediate growth step, switching from glucose to galactose resulted in a bimodal Gal1-GFP induction (Fig. 3.7b). However, there were stark differences in induction dynamics. Cells with no or 2h intermediate growth induced the earliest, while cells from 4h and 6h took longer to have detectable GFP signal. Similarly, different amounts of time were required

for cells to reach full Gal1 expression levels. It took 9 hours for cells coming out of 2h intermediate growth, 11 hours for cells coming out of 4h, and more than 12 hours for cells coming out of 6h (Fig. 3.7c). Interestingly, cells with no intermediate growth in fresh glucose displayed a slow Gal1 induction. The mean GFP levels of these cells were less than 10-fold of that measured for other conditions after 12 hours of induction. The noise trajectories of all conditions were semi-circular (Fig. 3.7d). Cells with no intermediate growth exhibited the highest CV, followed by 6h and 4h intermediate steps. Cells with only 2h growth displayed the lowest population variability.

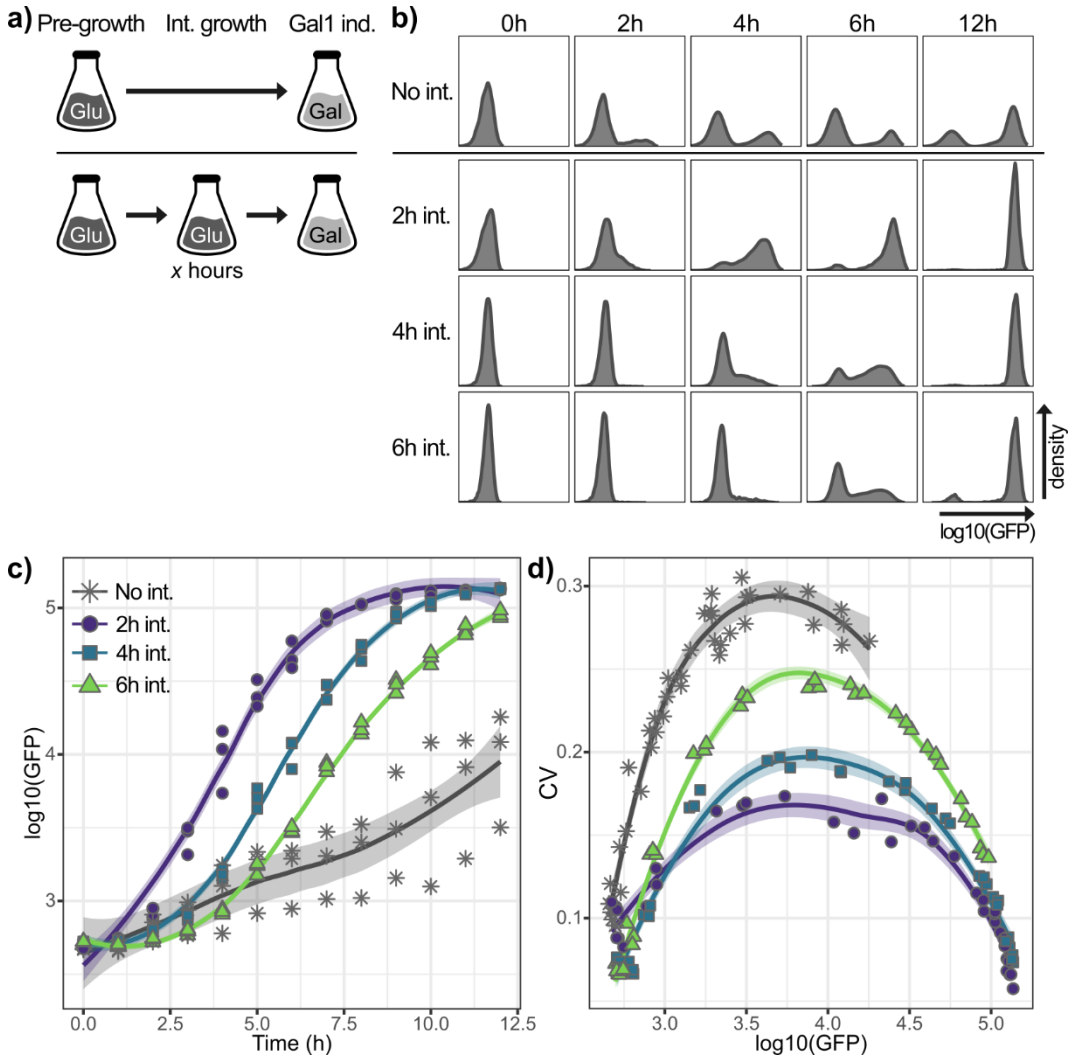


Figure 3.7: Analysis of Gal1 induction dynamics at different metabolic states. WT cells were grown until saturation phase (pre-growth), diluted in fresh glucose media for different times (intermediate growth), and induced with galactose. Sample acquisition was every hour for a total of 12 hours. Statistics were calculated on merged data from three biological replicates. The average number of cells per data point was 9,764 ± 857 (mean ± sd). Line and ribbon represent the fit and 95% confidence intervals of a curve fitting with *loess*, respectively. Cells were switched from glucose pre-growth directly into galactose (gray) or underwent an intermediate step of glucose growth for 2 (purple), 4

(blue), or 6 hours (green). Int: intermediate; Ind: induction; Glu: glucose; Gal: galactose. **a)** Experimental setup. Top: culture with no intermediate growth step. Bottom: culture with intermediate growth in glucose; *x* corresponds to 2, 4, or 6 hours. **b)** Histograms showing population distributions of Gal1-GFP signal at indicated induction time points as measured by flow cytometry. Each row corresponds to a different intermediate growth length. **c)** Mean Gal1-GFP expression during induction. Plotted is mean log10(GFP) signal vs. time. **d)** Noise of Gal1-GFP over different levels of expression. Plotted is CV vs. mean log10(GFP).

Cells transitioning directly from glucose pre-growth into galactose showed the greatest variability between replicates, as can be appreciated in the spread of mean Gal1-GFP signal in Fig. 3.7c. Although the samples with no intermediate growth are identical in nature, the variability across replicates highlighted the difficulties of obtaining reproducible inductions when cells had to simultaneously exit stationary phase and adapt to galactose. I selected an intermediate step of 4h for future experiments, as it balanced the hours required to achieve maximum Gal1 expression with the levels of noise exhibited by the population.

3.6. Validation of noise outliers under a different induction protocol

Once I decided to add an intermediate growth of 4 hours in fresh glucose, I re-validated the noise outliers using the new protocol (Fig. 3.8).

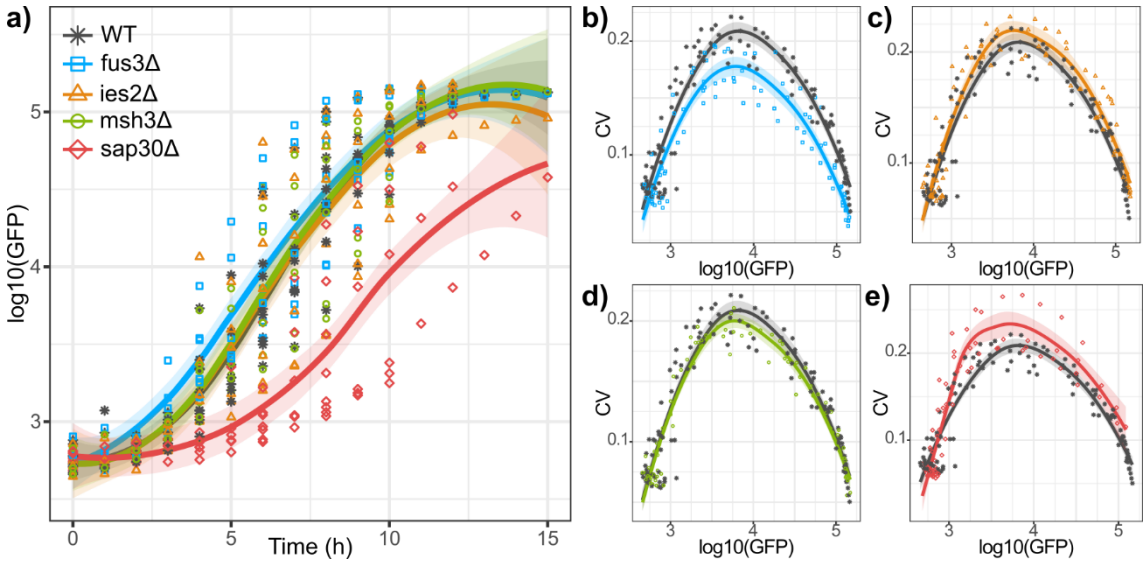


Figure 3.8: Validation of chromatin noise outliers after adding 4 hours of intermediate growth. Strains shown correspond to WT (gray), fus3Δ (blue), ies2Δ (orange), msh3Δ (green), and sap30Δ (red). Statistics were calculated on merged data from eight biological replicates for mutants and 10 biological replicates for WT. The average number of cells per data point was 10,012 ± 509 (mean ± sd). Line and ribbon represent the fit and 95% confidence intervals of a curve fitting with *loess*. **a)** Mean expression of Gal1-GFP over time for WT and selected mutant strains. Plotted is mean log10(GFP) signal vs. time.

b-e) Noise trajectories of Gal1-GFP over different levels of induction. Plotted is CV vs. mean log10(GFP) for WT and mutant strains: (b) fus3Δ, (c) ies2Δ, (d) msh3Δ, (e) sap30Δ.

Gal1 expression patterns followed a similar trend when comparing no intermediate growth (Fig. 3.3a) versus a 4-hour intermediate growth step (Fig. 3.8a). After 4 hours of growth in glucose and upon induction with galactose, cells showed an increase in GFP levels, with most strains reaching maximum Gal1 expression around 11h. sap30Δ showed a slow-inducer phenotype in both conditions, whereas ies2Δ induced slower than WT in the absence of intermediate growth but induced at the same rate as WT when the 4-hour extra step was included.

The noise trajectories in the presence or absence of an intermediate growth step are similar as well, although the outlierness of the mutant strains decreased when adding intermediate growth (Fig. 3.8b-e). In this case, only fus3Δ and sap30Δ exhibited different noise from WT. ies2Δ and msh3Δ showed only modest differences, with 95% confidence intervals overlapping.

Although the differences in noise were partially lost using the new protocol, I decided to continue its use as it was the approach that resulted in reproducible Gal1 induction. Furthermore, I reasoned that observations obtained from this protocol were a better reflection of intrinsic variability in Gal1 expression. A direct transition from pre-growth to galactose had the potential of mixing mutant-specific Gal1 noise with differential strain fitness while exiting stationary phase.

3.7. Gal1 noise differences are reduced after additional growth step

Once I decided it was necessary to include the intermediate growth step, I repeated the smFISH experiments to see if the differences in noise were conserved or lost at the mRNA level (Fig. 3.9).

Induction dynamics of Gal1 (Fig. 3.9a,b) showed a similar pattern as the one recorded by flow cytometry (Fig. 3.8a), including the observation that ies2Δ Gal1 expression behaves more similar to WT after the addition of the intermediate growth step.

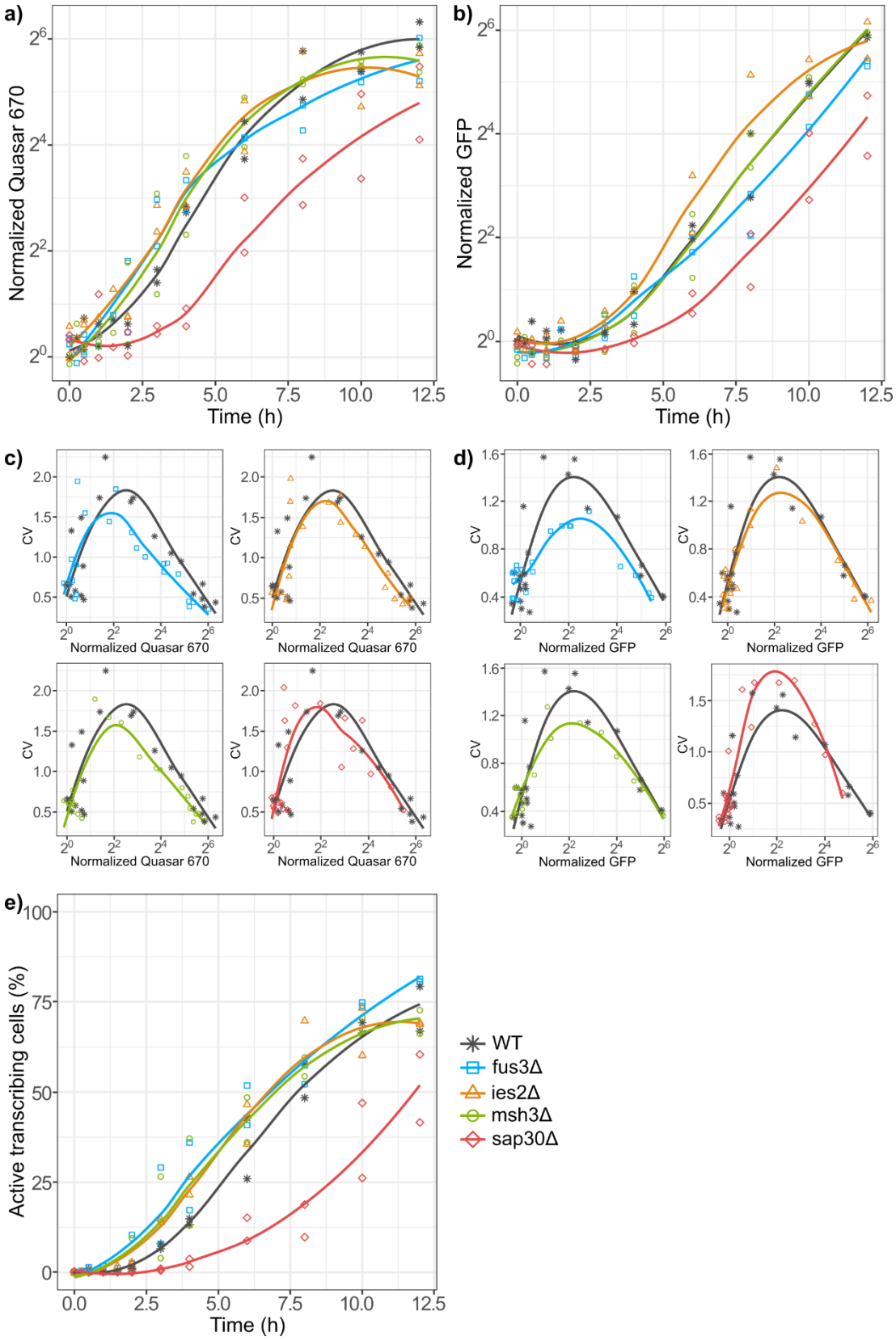


Figure 3.9: Quantification of Gal1 mRNA by FISH and Gal1 protein by fluorescent reporter after adding an intermediate growth step. Cells were pre-grown in glucose until saturation, diluted into fresh glucose for 4-5.5h of intermediate growth, transferred into galactose for 12h, and subjected to a Single Molecule Fluorescence *In-Situ* Hybridization (smFISH) protocol with subsequent imaging by

epifluorescence microscopy. The Gal1-GFP transcripts were labeled with Quasar 670-conjugated probes complementary to Gal1 mRNA sequence. **a,b)** Mean population fluorescence of (a) Quasar 670 and (b) GFP intensities during Gal1 induction for indicated strains. **c,d)** Noise trajectories across different levels of expression of Gal1-GFP (c) mRNA and (d) protein, plotted as CV (y-axis) vs. normalized fluorescence (x-axis). Each panel corresponds to a mutant plotted against WT. **e)** Average percentage of cells exhibiting an active GAL1 transcription site. Normalized fluorescence intensities are quantified as the fold-change vs. WT at 0h of induction (see Materials and Methods). Statistics were calculated on merged data from two biological replicates (one with 4h and one with 5.5h intermediate growth). Line represents the curve fitting with *loess*, intended as a visual aid. The average number of cells per data point was 517 ± 211 (mean ± sd). WT (gray), fus3Δ (blue), ies2Δ (orange), msh3Δ (green), and sap30Δ (red).

In these experiments, the noise trajectories of Gal1-GFP showed differences in noise only for some of the mutants (Fig. 3.9c,d). fus3Δ displayed reduced noise at the mRNA (Fig. 3.9c) and protein (Fig. 3.9d) levels. A comparable trend was observed for msh3Δ, although the decreased noise of mRNA was not as strong. Regarding the two high-noise outliers, neither of them displayed increased noise at the mRNA level (Fig. 3.9c). Surprisingly, ies2Δ noise trajectory formed a curve below WT, an observation that is not in line with the results obtained when no intermediate growth was performed (Fig. 3.4g). At the protein level (Fig. 3.9d), the noise of WT and ies2Δ behave similarly. sap30Δ displayed higher noise at the protein level, resembling flow cytometry results (Fig. 3.8e). Of note, *loess* fits should be interpreted with caution and are plotted as a visual aid for the reader. *Loess* fits are sensitive to plotting parameters and may suggest differences even in the absence of real variation. Instead, conclusions should be based on the scatter of the data points.

Finally, the dynamics of GAL1 activation were also altered by the addition of the intermediate growth (Fig. 3.9e). Under these growth conditions, no strain reached steady-state GAL1 transcription by 12 hours, although the locus induction of WT, ies2Δ, and msh3Δ appeared to be plateauing at 10-12 hours.

Overall, addition of intermediate glucose growth showed reduced differences in noise between WT and chromatin mutants. fus3Δ proved to be an outlier at the mRNA and protein levels after the addition of this step, even though it had displayed the same noise as WT when undergoing galactose exposure directly from stationary phase (Fig. 3.4g,h).

3.8. Ies2 anchor domain is involved in Gal1 noise trajectory

My project aimed to understand the mechanistic bases of chromatin regulation and its effect on gene induction noise. Based on the results I had at the moment (induction directly after glucose pre-growth), we chose one mutant for further mechanistic studies. We decided to focus on ies2Δ which, in the initial experiments, displayed increased noise at the mRNA and protein levels (Fig. 3.4g,h), and the mathematical modeling had highlighted altered DNA activation/inactivation rates (Fig. 3.6b). Although sap30Δ was the most extreme noise outlier, ies2Δ Gal1 induction levels were more similar to WT compared to sap30Δ cells. In addition, sap30Δ had a growth defect in galactose media which could potentially affect Gal1 noise in an indirect manner.

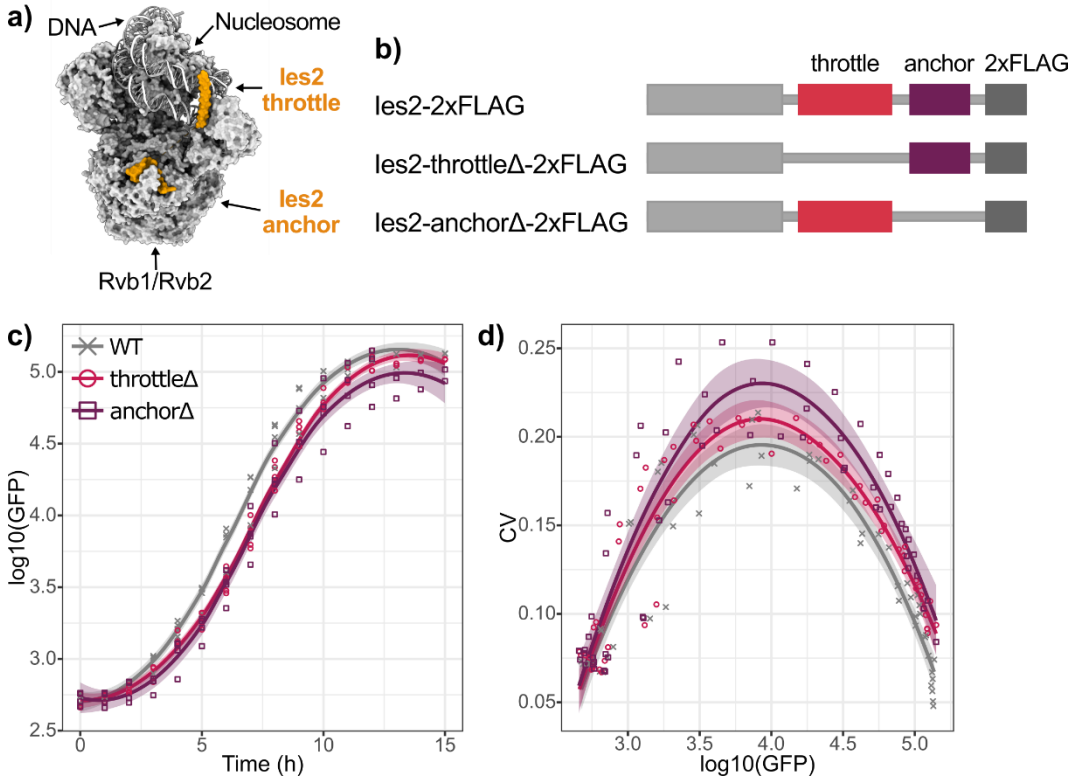


Figure 3.10: Lack of anchor domain of Ies2 results in higher Gal1 noise during induction. a) Depiction of Ies2 domains (orange) in the context of the nucleosome and DNA-bound INO80 complex (gray). Adapted from [139]. **b)** Ies2 domain deletions used in this study. All strains had the endogenous Ies2 tagged with two FLAG peptides at the C terminus. Strains had either no further modifications (top; WT), throttle domain deleted (middle; throttleΔ), or anchor domain deleted (bottom; anchorΔ). **c,d)** Strains shown in (b) were pre-grown in glucose, diluted into a 4h intermediate growth step, and induced with galactose for 12 or 15 hours. **c)** Mean population expression of Gal1-GFP over time for different Ies2 mutants. Plotted is mean log10(GFP) signal vs. time. **d)** Noise trajectories of Gal1-GFP over different levels of induction. Plotted is CV vs. mean log10(GFP). Statistics were calculated on merged data from four biological replicates (two 12-hour inductions, two 15-hour inductions). The average number of cells per data point was 10,335 ± 408 (mean ± sd). Line and ribbon represent the

fit and 95% confidence intervals of a curve fitting with loess, respectively. Strains depicted are WT (gray), throttleΔ (maroon), and anchorΔ (purple).

Ies2 is a member of the INO80 chromatin remodeler complex. Structural studies of the INO80 complex have highlighted two important regions of Ies2: the throttle domain and the anchor domain (Fig. 3.10a) [139, 140]. The throttle is responsible for holding the nucleosome in place while DNA is being slid around it, whereas the anchor attaches Ies2 to the Rvb1/Rvb2 structural module of INO80 [139]. I thus generated single deletion strains in which the throttle or anchor domains of the endogenous Ies2 were removed (throttleΔ, anchorΔ) (Fig. 3.10b) and measured their mean Gal1 expression and noise trajectories during galactose induction after 4 hours pre-growth (Fig. 3.10c,d).

Tagging the endogenous Ies2 with 2xFLAG did not have an effect on Gal1 induction or noise trajectories (Fig. S3). However, deleting the throttle or anchor domains reduced Gal1 expression kinetics slightly (Fig. 3.10c). Interestingly, while the deletion of the throttle only modestly affected the noise trajectory compared to WT cells, deletion of the anchor domain resulted in higher noise (Fig. 3.10d). This observation suggests the anchor domain of Ies2 is involved in the heterogeneous Gal1 induction.

4. DISCUSSION

Gene expression is stochastic in nature. Noise creates population variability even under the most homogeneous of conditions. While noise in gene expression has been subject to multiple studies over the last two decades, most efforts have been dedicated to the analysis during steady state gene expression. In this project, I focused on different chromatin-related factors and investigated how they alter the noise of a model gene during its induction. As an experimental system, I chose *Saccharomyces cerevisiae* undergoing galactose induction. To measure gene expression, I employed the galactose-responding gene Gal1 fused to GFP in its C-terminus (Gal1-GFP). I then applied different single-cell techniques to measure population variability.

Throughout the discussion I mention the terms “induction kinetics” and “noise trajectory.” Induction kinetics refers to the mean population expression over time. The term “noise trajectory” describes the pattern formed by the Coefficient of Variation (CV; standard deviation/mean) when plotted against different levels of gene expression.

My studies validated four chromatin-related factors as outliers for Gal1 noise during gene induction. fus3Δ and msh3Δ exhibited lower noise than WT, whereas higher noise was observed for ies2Δ and sap30Δ. Differences in noise were observed at the Gal1 protein level and, although to a lesser extent, at the mRNA level. Finally, I focused on Ies2 for dissecting the molecular mechanisms of altered noise. My results suggest that the interaction between Ies2 and the INO80 chromatin remodeling complex (INO80c) regulates gene induction noise at the Gal1 locus (Fig. 4.1).

This project commenced by re-analyzing the data produced by Bheda et al. [125]. In this paper, the authors used a microfluidic platform to follow the induction kinetics of Gal1-GFP in individual cells during two rounds of galactose induction and repression. They tested a library of 535 strains with single gene deletions for chromatin-related factors. One of the authors, Dr. Johannes Becker, re-analyzed the data to identify strains that consistently displayed altered noise. From the 28 identified strains, I selected 9 for further studies, from which only 4 were validated as outliers in subsequent experiments (Fig. 3.2 and 3.3).

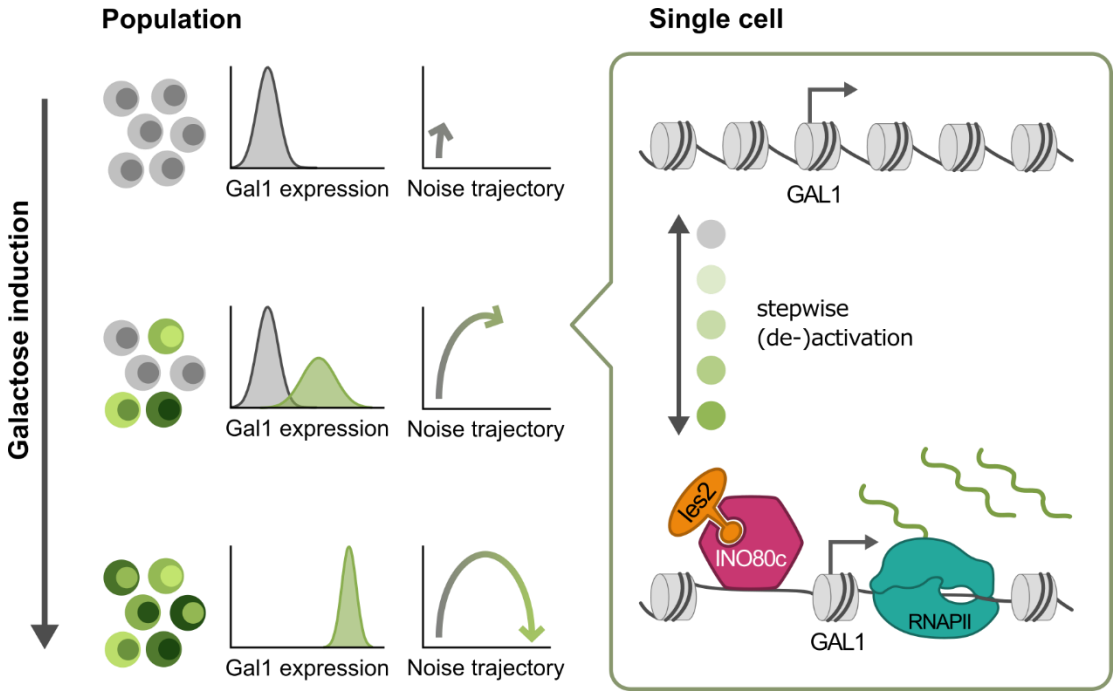


Figure 4.1. Proposed mechanism of Gal1 noise regulation through Ies2. Upon galactose exposure, cells induce Gal1 in a bimodal fashion. The noise trajectory of Gal1 peaks at intermediate levels of expression, and returns to a minimum once cells reach steady-state. At the single cell level, GAL1 fluctuates between inactive and active conformations in a stepwise manner. The transitions between steps are modulated by chromatin factors such as Ies2, which interacts with the INO80 complex through its anchor domain (represented by a protruding section). INO80c: INO80 complex; RNAPII: RNA polymerase II.

The discrepancy might be due to differences in experimental set up. The authors in [125] grew cells in glucose for 4 hours, performed a first induction of 1.5 hours using raffinose + galactose, repressed the cells in glucose by further 4 hours, and performed a second induction of 1.5 hours. The outliers identified by Dr. Becker considered all frames from the beginning of the first induction to the end of the second induction (7 hours total).

Factors showing altered noise in two consecutive short inductions separated by a repression could have a different effect during an individual long induction. In addition, the choice of sugars has an impact on Gal1 expression and noise (Fig. 3.1 and [131, 127, 126, 128]). We were aware of the experimental and biological differences between the article published by Bheda and my doctoral project, so I used the high-throughput screening data as a starting point to guide the outlier selection process, rather than working under the assumption my results had to be identical. The sugar combination the authors employed was considered during the protocol establishment but, based on

the results from Fig. 3.1, I selected glucose as the sugar in which to pre-grow the cells and galactose as the sugar to induce Gal1 expression.

After analyzing Gal1 noise in the mutants at the protein level, I further focused my attention on mRNA. I detected Gal1 transcripts using a Single Molecule Fluorescence *In-Situ* Hybridization (smFISH) protocol (Fig. 3.4). Because the readout of smFISH is microscopy images, I simultaneously acquired data on the Gal1-GFP transcript and protein expression level. Noise trajectories of Gal1 transcript and protein (Figs. 3.4g,h) were similar to those obtained for protein by flow cytometry (Fig. 3.3b), although the differences between WT and some of the mutants were not as stark. There are multiple explanations behind this.

Flow cytometry has the advantage of acquiring thousands of cells in the span of seconds, whereas microscopy is limited by the amount of time necessary to acquire images. Indeed, I encompassed x20 more cells per data point in the flow cytometry experiments than in the smFISH, which allowed to distinguish subtle differences in fluorescence—and thus noise—by flow cytometry. Due to the technical complexity of the protocols, it was also simpler to obtain better time resolution with flow cytometry than acquiring samples for smFISH. In Fig. 3.3, 12 hours of induction are divided into 25 time points per strain. In the same induction time span, smFISH experiments only had between 6-12 time points (Figs. 3.4 and 3.9).

Another aspect that could have an impact is sample preparation. Cells in flow cytometry required minimal preparation and were measured immediately or on the following day after induction. In contrast, cells subjected to smFISH require at least three days of work before samples can be imaged. The fluorescent signal is subject to bleaching between sample preparation and imaging.

After analyzing Gal1 mRNA, Dr. Lea Schuh and Dr. Dantong Wang developed a mathematical model to identify the steps of gene expression with rates differing between WT and mutant strains. First, the model evaluated the number of steps between a DNA_{on} and a DNA_{off} state (Fig. 3.5). Such multistep models have been used extensively [17]. For example, the expression of a yeast gene induced under osmotic shock (STL1) was shown to require 4 activation steps to explain its mRNA distribution [141].

The Bayesian Information Criterion (BIC) was used to select the model with the best fit, and a threshold of 10 units was used to reject less suitable models.

The BIC is defined as $\log(n)\,k - 2logL$, where n is the number of data points, k is the number of parameters and $logL$ is the log-likelihood value for the maximum likelihood estimate of the model parameters. The best model is defined as the one with the lowest BIC and, given that BIC increases with the inclusion of additional parameters, BIC is a metric against overfitting. Results showed that any number of steps could adequately account for all strains, except for sap30Δ (Fig. 3.5b).

10 random starts were enough to converge at the optimal parameters for WT, fus3Δ, ies2Δ, and msh3Δ. However, even using 100 random starts in sap30Δ only showed convergence when considering 0-3 DNA steps. Using 4 steps or more, although lowering the BIC, resulted in an inferior number of estimations converging and showed poor fits to the data. A likely explanation is that sap30Δ behavior diverges considerably from any other strain, as shown by its noise trajectory. It is plausible that our model architecture does not properly account for the difference in sap30Δ expression dynamics and thus results in poor fitting.

We selected a common number of DNA steps as a feature of all strains to simplify result comparisons. The number of steps chosen was 3, as it corresponded to the mean of best steps for WT, fus3Δ, ies2Δ, and msh3Δ. We proceeded to fit the model in pairs of WT-mutant and defined a subset of "best models," chosen as those that whose BIC was less than 10 units away from the lowest BIC.

The best models for fus3Δ and msh3Δ supported a scenario in which all rates of gene expression (γ_{on}, γ_{off}, λ_{prod}, λ_{deg}, κ_{prod}, κ_{deg}; see Fig. 3.5a) were identical to those of WT cells (Fig. S2b). This conclusion is not entirely surprising given that the smFISH results used to feed the models showed minimal or no differences in Gal1 expression dynamics among the three strains (Fig. 3.4e-i).

The opposite conclusion was achieved for high noise outliers, as the set of best models required mutant-specific DNA activation/inactivation rates (γ_{on}/γ_{off}). No additional differences in model parameters were necessary to explain the data for ies2Δ and sap30Δ (Fig. 3.6b and Fig. S2b). These observations align with the known roles of these factors. Ies2 is part of the INO80 chromatin

remodeler, whereas Sap30 belongs to the Rpd3L histone deacetylase complex. Based on their biological functions, we would expect Ies2 and Sap30 to act at the step of DNA regulation (γ_{on}/γ_{off}) and not, for example, at the level of protein production and degradation (κ_{prod}, κ_{deg}).

Subsequent experiments showed difficulties reproducing my original results, particularly in obtaining a successful galactose response. Although I was never able to determine the cause of the slow Gal1 induction, I found an alternative protocol that reestablished Gal1 expression. It consisted in the addition of an "intermediate growth" step in which cells were resuspended into fresh glucose and resumed cell growth and division before being challenged with galactose (Fig. 3.7).

My results revealed that not only the intermediate growth step altered Gal1 induction and noise kinetics, but also its duration had an effect. 2 hours in fresh glucose was enough to change the phenotype displayed by the samples with no intermediate growth. Instead of being slow inducers with high noise, samples that underwent 2h of intermediate growth showed the fastest induction and the lowest noise of all samples tested (Fig. 3.7c,d). I attributed this drastic difference to the fact that cells transitioning directly into galactose must exit stationary phase before adjusting to the new carbohydrate source. The presence of a subpopulation of cells showing distinctive Gal1 expression by 2 hours (Fig. 3.7a) signaled how heterogeneous is a population coming out of stationary phase.

Another aspect that this experiment highlighted is that the number of hours spent in glucose affects how easily cells induce Gal1. Cells reaching saturation in the pre-growth state enter a so-called diauxic shift as glucose becomes depleted. I selected 4 hours of intermediate growth, as 2 hours was not enough time to ensure the population had undergone a full cell division as monitored by OD_{600} (data not shown).

Once experimental conditions had been defined, I re-tested the outlier phenotype of the four mutants of interest (fus3Δ, ies2Δ, msh3Δ, and sap30Δ). When the intermediate growth was included, only fus3Δ and sap30Δ displayed an outlier phenotype (Fig. 3.8). This could be because, as observed in Fig. 3.7c,d and described in the Results section, cells with no intermediate growth display more variability within the population as well as between replicates. If cell populations become more uniform during the

intermediate growth, it is likely that differences between mutants and WT diminish as well, and only those mutants with the stronger phenotype (fus3Δ and sap30Δ) remained outliers after the "homogenization" of the intermediate growth step. A second possible source of discrepancies is technical. Different flow cytometers were used for the two conditions, which is exemplified by the range of GFP fluorescence in each experimental setup (compare x-axes of Fig. 3.3a and Fig. 3.8a).

Once we noted that adding an intermediate growth step reduced the differences in noise between WT and mutants, and considering the technical difficulties of smFISH explained above, it was not surprising to find only subtle differences between WT and mutant mRNA noise (Fig. 3.9c). The big variability of Gal1 expression and noise between replicates, both from the biological point of view and from the technical aspects of smFISH sample preparation, made it harder to confirm the significance of these subtle noise differences.

Lastly, as part of this project, I aimed to characterize the molecular mechanisms by which one of the chromatin-related factors altered gene induction noise. With the information I had at the time, namely results without an intermediate growth step, Ies2 was selected for further testing. I deleted two domains of Ies2: the throttle domain, which stabilizes the nucleosome during DNA sliding, and the anchor domain, which attaches Ies2 to the rest of INO80c [139].

Although I had expected the throttle domain to participate—at least to some extent—in Ies2-mediated noise, its deletion generated a noise trajectory similar to WT. On the other hand, the anchorΔ mutant exhibited higher noise (Fig. 3.10d), resembling the effect observed when deleting Ies2 entirely. Although these results cannot address whether Ies2 can bind INO80c in the absence of the anchor domain, they suggest the interaction between these two proteins is implicated in regulating population variability of Gal1 induction.

4.1. The shape of noise trajectories

The inverse linear relationship between the mean abundance of a gene and its noise was observed by the early studies of Bar-Even [68] and Newman [69], and has been confirmed by multiple studies since then [22, 70–74]. It was surprising then when I found the semi-circular noise trajectory formed by Gal1 during induction.

Such trajectory is logical when we consider the process of gene induction. The population transitioned from a state in which all cells are "off" to one in which all cells are "on." Those two scenarios at the edges of the Gal1 expression spectrum are the lowest in noise as all cells share the same expression status. The Gal1 induction that happens in between these two extremes is subject to stochastic fluctuations. Because the populations induced Gal1 in a bimodal fashion when transitioning from glucose to galactose, the heterogeneity is higher in those intermediate states of expression (Fig. 3.1).

In contrast, cells transitioning from raffinose induced as a single unit (unimodal induction; Fig. 3.1a). Their noise trajectories described a horizontal line at the beginning of the induction, after which the CV only decreased as cells approached steady state (Fig. 3.1c). The CV of all conditions tested resembled the linear relationship mentioned above only when Gal1 reached intermediate-to-high levels of expression. This highlights the relevance of studying dynamic systems and not only final system states.

In this project, I only compared the noise of WT and mutants when the populations exhibited similar levels of expressions. Comparing CV values matched by induction time could lead to the wrong assumption that two samples exhibited different noise due to their genetic background, and not because they were in different stages of induction.

While reviewing literature in noise dynamics, I found that the semi-circular noise shape observed during Gal1 induction had been described before [23, 55, 56, 66], although its existence had been overshadowed by the multiple studies showing a linear relationship between noise and mean gene expression [22, 68–74]. The authors describing semi-circular noise shapes compared noise (quantified by the CV, CV^2, or Fano factor) at different levels of gene expression, although the methods for achieving a range of expression levels were different than considering level of induction over time.

The ground-breaking paper of Elowitz et al. deciphered the contributions of internal and external noise by using a dual-color assay [55]. The reporter used for their study was under the control of a lac-repressible promoter. Instead of analyzing the noise levels of the reporter genes during induction, they achieved different expression levels by using a gradient of IPTG, a molecule that binds and inactivates the lac repressor, therefore inducing the expression of their reporter. Their results showed that cells

were uniformly repressed or uniformly induced at low and high IPTG concentrations, respectively. Consequently, they exhibited low noise. In contrast, fluctuations in LacI at intermediate concentrations of IPTG caused variations in the transcription of the reporter gene, increasing the total noise of the population [55]. It is important to mention that the measurements of noise were made on steady-state conditions and there was no evidence that their population was bimodal.

Raser and O'Shea employed the dual-color assay developed by Elowitz, this time under the control of the GAL1 promoter in *S. cerevisiae* [56]. They reported a similar pattern to the one observed in my experiments, although the authors did not test noise during GAL1 induction. Instead, they monitored the noise of GFP under the control of a GAL1 promoter after 8 hours of induction, but achieved different levels of expression by changing the concentration of galactose in the media.

Their observation was similar to the one obtained by Blake an colleagues [66]. The authors also considered GFP under the control of a GAL1 promoter, but this time coupled to two *tet* operons, and measured the noise of GFP at different transcriptional efficiencies, achieved by varying amounts of galactose and ATc (anhydrotetracycline; relieves TetR-mediated repression). They measured noise after more than 30 hours of cell growth. They observed the same pattern when GFP was driven by the promoter of ADH1, exemplifying that the semi-circular pattern is not strictly specific for the GAL1 promoter.

A follow-up study by the same lab explored the role of TATA boxes in cell-to-cell variability [23]. Their induction system was still dependent on galactose and ATc concentrations and they characterized the induction dynamics of a GFP containing the endogenous TATA box (TA-WT), or a mutated TATA box with severe induction impairment (TA-sev). They noticed a rapid increase in noise as soon as GFP was induced, after which the population heterogeneity decreased as cells approached steady state. Similar to my results, their induction of GFP was bimodal and the shape of the distribution was correlated with the gene's noise. Not only did the authors compared noise in TA-WT and TA-sev at matched time points, but they also tested the noise by matching both populations by mean expression, although this expression was at steady state and achieved using different concentrations of ATc. They concluded that even at similar mean population expression, the heterogeneity of TA-WT was higher than TA-sev due

to a bigger transcriptional bursting in WT cells. This heterogeneity had implications for the survival of cells when exposed to an environmental stress. The study of Blake et al. [23] showed the importance of the promoter region for cell-to-cell variability. It also showed differences in noise could exist even in populations with the same mean expression, and these differences in noise could be functionally relevant.

4.2. The role of chromatin mutants in noise

From the high-throughput microfluidic screening of Bheda et al. [125] I did a first selection of nine chromatin-related factors to test whether they exhibited differences in Gal1 noise during gene induction. From those nine candidates I validated four, namely fus3Δ, ies2Δ, msh3Δ, and sap30Δ (Figs. 3.2 and 3.3). In this section, I will briefly describe the known roles of each factor, as well as their reported contributions in cell-to-cell variability.

4.2.1. Fus3

Fus3 is a mitogen-activated protein (MAP)-kinase involved in pheromone signal transduction during mating [142]. Fus3 was shown to be involved in cell-to-cell variability during pheromone response, although its deletion was linked to a higher variability of the reporter gene when cells were exposed to high doses of mating pheromone α-factor [143]. In a separate study, although the deletion of Fus3 alone did not result in increased cell variation, Fus3 was shown to be involved in a signaling cascade that created population variability via microtubule perturbations [144].

Fus3 has no DNA binding activity. Instead, it stimulates gene transcription indirectly by phosphorylation of transcription factors such as Ste12, Far1, Bni1, and Sst2 [145–147]. None of them are known regulators of GAL1, so the link between Fus3 and galactose induction is not evident.

It is possible that Fus3 has an uncharacterized target that mediates signal transduction between Fus3 and the GAL1 promoter. Another scenario is that, in the absence of Fus3, there is crosstalk between other MAPK pathways, as has been shown for osmostress. Hog1 is a MAPK involved in osmoregulation [148]. In hog1Δ cells, the osmostress signal is diverted and gets transmitted through Fus3 instead [149].

4.2.2. Ies2

Ies2 is member of the INO80 chromatin remodeler complex (INO80c). Its role is to maintain the nucleosome in place while DNA is being slid around it [139]. Despite being a subunit of an essential chromatin regulator, the role of Ies2 has not been described until recently.

Ies2 has been investigated in quiescence in *Schizosaccharomyces pombe* [150]. Cells bearing different deletions of INO80c subunits, including Ies2, had higher mortality rates and failed to induce quiescence genes. It is difficult to compare the conclusions of the study to my experimental setup; the earliest time point the authors tested was 24 hours, whereas the pre-growth I employed lasted no more than 16 hours. Nevertheless, it is possible that a quiescence-associated behavior is involved in the noise differences of ies2Δ cells with and without intermediate growth.

There are reports of the Arp5-Ies6 module being unable to bind INO80c in the absence of Ies2 [151]. Arp5 couples ATP hydrolysis to nucleosome sliding by functioning as a counter grip to the Ino80 ATPase domain [139]. It is possible that Ies2 is not directly responsible for the Gal1 noise trajectories we observed, and the phenotype is mediated by Arp5 instead. Neither Arp5 nor Ies6 deletions were considered in the initial high-throughput screening published by Bheda et al. [125].

Other subunits of INO80c might be involved in noise regulation as well. Arp8 can work as a “ruler” to position nucleosomes [33]. The deletion of Arp8 resulted in a slightly diminished expression and higher noise of a constitutively-expressed YFP reporter driven by the promoter of GAL1 [152]. The authors only tested noise at steady state, and it is unclear whether the increase in noise was due to the Arp8 deletion or due to the expected noise of YFP when its expression is reduced. Removal of Arp8 has also been shown to decrease gene expression and increase intrinsic noise of a reporter expressed from a maximally-induced PHO5 promoter [56]. Deletion of Arp8 was not assessed by Bheda et al. [125].

Deletions of other INO80c components were tested in the high-throughput screening of Bheda et al. [125], namely Ies4, Nhp10, Ies1, and Ies5. Only nhp10Δ and ies1Δ were detected as outliers, resulting in higher noise in 1 and 3 frames respectively. For comparison, ies2Δ was outlier for 10 frames of the time lapse.

4.2.3. Msh3

Msh3 is a mismatch repair protein that, together with Msh2, mediates the repair of small DNA insertions and deletions during DNA replication [153]. The Msh2-Msh3 heterodimer also participates in heteroduplex formation during mitotic and meiotic recombination [154] and in the removal of non-homologous DNA ends during Double Strand Break Repair [155].

The only reported link between Msh3 and the GAL1-GAL10 locus was in the context of transcription-stimulated mitotic recombination [156]. The authors analyzed the recombination rates at the GAL1-GAL10 promoter, whose activity was regulated by the presence or absence of Gal80. When the donor or the recipient allele were highly transcribed, msh3Δ showed an increase in recombination compared to WT, whereas the recombination rate was similar to WT under low-transcription conditions. No assessment of cell heterogeneity was carried out in this paper.

The yeast employed in my experiments is haploid, but mitotic recombination could occur once cell division has resumed and cells are in S or G2 phase, roughly 7-9 hours after glucose-galactose switching (data not shown). If that was the case, msh3Δ would only be an outlier for noise during the latest hours of the induction, which is indeed the observed phenotype (Fig. 3.3b), although these two events might not be linked.

It is important to mention that Rad51, a protein involved in homology recombination and known interactor of Msh2 [157], was one of the nine outliers selected from the microfluidics screen. Although the phenotype of rad51Δ was not further investigated, it showed a slight reduction in Gal1 noise compared to WT (Fig. 3.2).

4.2.4. Sap30

Sap30 is a component of the Rpd3L histone deacetylase complex (HDAC) [158], which has been linked to gene expression noise previously. In the paper of Weinberger et al. [159], the authors searched for regulators of transcriptional bursting. An initial screen of 137 non-essential chromatin factors deletions suggested the Rpd3L complex was involved in transcriptional burst repression. Follow-up quantification of 200 GFP-fusion reporters in steady state revealed that sap30Δ had an overall increase

in burst frequency compared to WT. Because the authors defined burst frequency as the inverse of CV, it meant sap30Δ displayed a lower CV than WT in their study.

This observation is opposite to my results, as deleting Sap30 resulted in higher CV during Gal1 induction. A possible explanation is that the authors in [159] averaged the burst frequency of multiple genes, whereas the higher noise of Gal1 might be gene-specific. A second consideration is that the authors only measured burst frequency in steady state. As described in Results, the noise outliers I characterized had a different noise than WT during gene induction, but the same noise once Gal1 had reached maximum expression. I did not report on steady-state of Gal1 for sap30Δ as the time it took to fully induce was longer than the maximum time tested (15 hours). However, given the noise trajectory displayed by all mutants, it is likely that sap30Δ and WT populations exhibit the same noise at Gal1 steady-state.

The Rpd3L complex has been further linked to noise, albeit in the context of aging cells [160]. The authors expressed YFP from a GAL1 promoter and analyzed YFP expression and noise over multiple generations. The absence of the catalytic subunit of Rpd3L caused a reduction in noise. Similar to [159], the authors only measured noise at steady state.

All other components of the Rpd3L complex, with the exception of Ash1, were tested in the high throughput screening of Bheda et al. [125]. Only sin3Δ, sap30Δ, and sds3Δ were proved to be outliers, each for 5, 16, and 24 frames respectively. sin3Δ was discarded as junction PCR demonstrated the strain employed in the screen contained both WT and mutant cells. Although sds3Δ was outlier for more frames than sap30Δ, it was not considered for further analysis because it had a deficient induction during validation experiments (data not shown).

4.3. Other chromatin factors that regulate Gal1 expression and noise

Moreno and Acar [152] investigated chromatin-related factors that impact expression and noise. YFP fluorescence expressed from a GAL1 promoter was measured in six mutant strains, each one bearing a deletion of a protein belonging to a chromatin complex known to participate in GAL1 expression.

The authors tested set1Δ and set2Δ (histone methyl-transferases), jhd2Δ (histone demethylase), snf6Δ (SWI/SNF remodeling complex), arp8Δ (INO80 remodeling complex), and gcn5Δ (SAGA histone acetyl-transferase complex). YFP was induced not by addition of galactose, but by growing the cells in mannose and deleting the negative regulator GAL80. All the mutants investigated exhibited lower expression and higher noise of GAL1 promoter-driven YFP. In the following paragraphs, I'll discuss each of the deletions, the complexes they belong to, and I will make a parallel with the outlier detection analysis performed by Dr. Johannes Becker in the original screen (see Fig. 1.8).

Set1 belongs to the histone methyl-transferase complex COMPASS [161]. Although set1Δ was not included in the microfluidics screening [125], six other members of COMPASS were tested, two of which were identified as outliers. sdc1Δ and swd1Δ were detected as low-noise outliers for 6 and 5 frames, respectively. Although we cannot draw direct comparisons given the experimental differences of both studies, it is interesting that Moreno and Acar found set1Δ to be linked to higher noise, whereas our results point towards COMPASS having the opposite effect.

Set2 was included in the deletions tested in [125]. The outlier detection algorithm identified it as outlier in only 2 frames, potentially indicating that while set2Δ might be linked to higher noise in GAL1 steady state [152], its role in GAL1 induction is not as prominent. The histone demethylase Jhd2 was also part of the high-throughput screening, although it was not identified as a noise outlier in any condition.

Snf6 is a component of the chromatin-remodeling complex SWI/SNF [162]. Snf6 was not part of the high-throughput screening. However, three other members of SWI/SNF were included, one of which (snf11Δ) was a high-noise outlier for 3 frames.

Lastly, while Gcn5 was not one of the proteins tested in [125], five other members of the SAGA histone acetyl-transferase complex were identified as noise outliers. Individual deletions of Chd1, Sgf11, Spt3, and Ubp8 caused higher noise. Ubp8 showed the strongest phenotype, as its deletion resulted in higher noise for 28 frames (4h 40m). Ubp8 is the ubiquitin protease catalytic subunit of SAGA [163]. It is part of the DUB (deubiquitinase) module, whose role is to cleave monoubiquitination from H2B and

stimulate transcriptional elongation [164]. The detection of five members of the complex as noise outliers strongly suggests SAGA is linked to GAL1 induction variability.

4.4. Conclusion

In this project, I have addressed the role of chromatin regulation in the noise of a gene undergoing induction. I focused on four genes whose deletion was associated with differential Gal1-GFP noise. fus3Δ and msh3Δ exhibited lower noise than WT, whereas ies2Δ and sap30Δ displayed higher noise. Differences in Gal1 noise were detected at the protein level and, for some of these factors, also at the mRNA level. These observations were specific for gene induction, as mutant and WT strains showed equal levels of variation once Gal1 had reached maximum expression. Overall, the results of my dissertation highlight the importance of studying noise in dynamic, non-steady-state systems.

Part of my work was dedicated to identifying the molecular mechanisms that regulate gene induction variability. I showed that the anchor domain of Ies2 is implicated in Gal1 induction noise. However, additional molecular characterization will be necessary to understand how the INO80 complex acts at the GAL1 locus and which factors regulate its stochastic activation.

The improvement of single-cell techniques will facilitate our understanding of noise in dynamic systems. In particular, it will be interesting to elucidate how the nucleosome landscape, histone modifications, and histone variants change during Gal1 induction in individual cells over time.

5. MATERIALS

5.1. Buffers and media

Name	Composition
Synthetic complete (SC) media	2.75 g/L Drop-out medium
	8.375 g/L Yeast Nitrogen Base w/o Amino Acids
	1.5-2% sugar
Phosphate Buffer	0.1 M potassium phosphate
	pH 7.5
FACS Buffer	1x PBS
	1% Fetal Bovine Serum
	2mM EDTA
	25 mM HEPES
	pH 7.5
Buffer B	1.2 M sorbitol
	0.1 M potassium phosphate
	pH 7.5
Spheroplast Buffer	Buffer B
	20 mM VRC
	20 mM β-mercaptoethanol
Resuspension Buffer	Buffer B
	20 mM VRC
Hybridization Buffer	100 mg/ml dextran sulphate
	1 mg/ml *E. coli* tRNA
	2 mM VRC
	20 ug/ml Bovine Serum Albumin
	2x SSC
	100 uL/mL Formamide
Saline-sodium citrate (SSC)	17.53 mg/mL sodium chloride (NaCl)
	8.82 mg/mL sodium citrate
Wash Buffer	10% formamide
	2x SSC

5.2. Chemicals and reagents

Name	Supplier	Catalog ID
Bovine Serum Albumin (BSA)	New England Biolabs	B9000S
Coverslips	VWR	ECN 631-1567
D-(+)-galactose	Sigma-Aldrich	G6404
D-(+)-glucose	Sigma-Aldrich	G8270
D-(+)-raffinose	Sigma-Aldrich	83400
DAPI	Sigma-Aldrich	D9564
Dextran sulphate	Sigma-Aldrich	D8906

Di-potassium hydrogen phosphate (K_2HPO_4)	Roth	7758-11-4
Drop-out Mix Complete w/o Yeast Nitrogen Base	US Biological	D9515
Dulbecco's Phosphate Buffered Saline (DPBS)	Thermo Fisher Scientific	14190094
E. coli tRNA	Roche	10109541001
Ethanol	Merck Millipore	64-17-5
Ethylenediaminetetraacetic acid (EDTA)	Roth	6381-92-6
Fetal Bovine Serum (FBS)	Thermo Fisher Scientific	12103c
Formaldehyde	Sigma-Aldrich	F8775
Formamide	Sigma-Aldrich	F9037
HEPES	Roth	7365-45-9
Lyticase	Sigma-Aldrich	L2524
Potassium di-hydrogen phosphate (KH_2PO_4)	Roth	7778-77-0
ProLong Gold	Life Technologies	P36934
Sodium chloride (NaCl)	Roth	7647-14-5
Sorbitol	Sigma-Aldrich	S7547
Superfrost microscope slides	Thermo Fisher Scientific	12372098
SYTOX Blue	Life Technologies	S34857
Tri-sodium citrate dihydrate	Roth	4088.1
Vanadyl Ribonucleoside Complex (VRC)	New England Biolabs	S1402S
Yeast Nitrogen Base without Amino Acids	Becton Dickinson	291940
β-mercaptoethanol	Sigma-Aldrich	M3148

5.3. Equipment

Type	Name	Supplier
Flow Cytometer	LSRFortessa	BD Biosciences
Flow Cytometer	CytoFlex S Flow Cytometer	Beckman Coulter
Microscope	Eclipse Ti-E with SPECTRA X light engine illumination	Nikon
Camera	iXon Ultra 888	Andor

5.4. Software

Name	Version
R	v4.3.1
FlowJo	v10.8.1
Cell-ACDC	v1.4.8
spotMAX	NA
MATLAB	R2017a
gcc	4.8.5
g++	4.8.5

6. METHODS

6.1. Yeast growth and culture

Yeast cells were grown in SC media supplemented with 2% (weight/volume) of a single sugar, or 1.5% of each sugar in case of raffinose + galactose mixtures. Cell pre-growth was performed overnight. In cases when cultures had no intermediate growth, cells were washed twice with miliQ water and diluted at 0.4 OD in inducing media. Cultures subjected to intermediate growth were diluted in fresh media at an OD_{600} of 0.2 and let grow for the corresponding amount of time before being washed twice with miliQ water and transferred to inducing media. During intermediate growth and induction, all cultures were diluted frequently to ensure cells were in exponential phase.

6.2. Yeast reporters and constructs

The wild type (WT) strain used in this study is the same employed in [125]. It is a derivative of parent strain Y7092 (SGA WT query strain, MATα *can1Δ*::STE2prSp_his5 *lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*) with GFP fused to the C-terminus of Gal1 by transformation of a PCR product.

The gene deletions selected from the high-throughput screening were confirmed by junction PCR. Subsequently, deletions were re-done in the WT strain to ensure the phenotype observed was not caused by secondary mutations occurred during SGA. Gene deletions were performed by PCR-mediated homologous recombination.

Ies2 mutant constructs were generated in a plasmid and amplified by PCR. WT yeast was transformed with the purified PCR product and clones with a successful integration were selected by a LEU2 metabolic marker. The marker was later removed by recombination of flanking pLox sites. Constructs were validated by Sanger sequencing.

The anchor and throttle domains of *S. cerevisiae* Ies2 were determined based on (i) the INO80 structure of *Chaetomium thermophilum* [139], (ii) AlphaFold [165, 166] predictions of *S. cerevisiae* Ies2, and (iii) peptide sequence similarity among both organisms.

Structural analysis of *C. thermophilum* Ies2 showed the throttle domain as a single α helix, while the anchor domain is composed of two short β sheets followed by two short α helices. The analogous

structures were identified in the AlphaFold prediction of *S. cerevisiae* Ies2 (a long α helix for the predicted throttle, and two short β sheets for the predicted anchor). Lastly, the Ies2 peptide sequences of both organisms were aligned with Protein BLAST (https://blast.ncbi.nlm.nih.gov/). Results showed amino acid similarities between the characterized *C. thermophilum* throttle domain and the predicted *S. cerevisiae* throttle domain.

The amino acid sequences of the predicted *S. cerevisiae* domains are the following:

Ies2 throttle (a.a. 235-273):
DEEIQLRRAENARKRKNLSEKRLEEEKQDTINKLLKKRA

Ies2 anchor (a.a. 306-319):
TRILRRYEEDLFCT

Both constructs, as well as the endogenous Ies2, were tagged with two FLAG peptides at the C-terminus.

6.3. High-throughput screening of mutants

Extended details of the method can be found in Bheda et al. [125]. The collection of mutants was generated by crossing the Gal1-GFP reporter strain with the Synthetic Genetic Array (SGA) single-deletion collection of non-essential genes [167]. The library tested contained 567 yeast strains, out of which 535 deletions corresponded to chromatin-related factors, 31 to randomly-selected non-chromatin associated factors, and 1 control strain for the antibiotic marker.

Strains were grown in individual chambers of a microfluidic platform that allowed for automated media changes [168]. Cells were initially grown for 4 hours in glucose, followed by a first raffinose + galactose induction for 1.5 hours (1.5% of each sugar), a repression with glucose for 4 hours, and a second raffinose + galactose induction for another 1.5 hours.

Images of the yeast were captured at 60x magnification with time points every 10 minutes. Cells were segmented in PhyloCell (available on GitHub https://github.com/gcharvin) and analyzed with custom MATLAB scripts. The screening was performed in three biological replicates, each with two technical replicates per strain within the same chip.

6.3.1. Outlier detection of high-throughput screening

Outlier detection was based on the assumption that most of the gene deletions would have minimal or no effect on the noise of Gal1-GFP. I was therefore interested in deletions that repeatedly behaved outside the norm.

For every frame of the time lapse, strains were plotted based on their mean GFP abundance (x-axis) and their CV^2 (coefficient of variation; CV = sd/mean) (Fig. 3.1b). A moving window of 11 data points was used to calculate upper and lower thresholds of "outlierness". The thresholds were based on the interquartile range (IQR) of the 11 data points. IQR is a measure of dispersion and is calculated by subtracting the 25% quartile (Q1) from the 75% quartile (Q3). The upper threshold was Q3 + IQR; the lower threshold was Q1 – IQR. Once the threshold values were calculated for all moving windows, cubic smoothing splines were applied to get smooth curves for the threshold values (red and blue dashed lines in Fig. 3.1b).

This process was repeated for every frame, for a total of 43 frames (7 hours of time lapse). Outliers were explored in the individual datasets (unnormalized data), as well as in the condensed dataset formed by merging the three replicates by loess (normalized data). Strains were ranked based on the number of frames in which they were outliers. The final candidates corresponded to those that were outliers for at least 4 frames in the normalized data.

6.4. Flow Cytometry

6.4.1 Sample acquisition

For experiments depicted in Figs. 3.2-3.4, cells were collected at different time points during the induction, washed once with Phosphate Buffer to remove residual galactose, resuspended in 500 uL of FACS Buffer and stored overnight at 4 °C. The following morning, cells were incubated for approximately 10 minutes with SYTOX Blue to identify dead cells, and samples were analyzed using a BD LSRFortessa flow cytometer with a high-throughput sampler of 96-well plates.

For experiments depicted in Figs. 3.8, 3.9, 3.10, and S3, cells were collected at different time points during the induction, centrifuged, and resuspended directly in SYTOX Blue for 10 minutes. Immediately after, samples were recorded in a CytoFlex S Flow Cytometer.

6.4.2. Sample gating

Events were gated by forward and side scatter (FSC-A and SSC-A) to remove debris. Further gating of FSC-A and FSC-W removed doublets and budding yeast. The SYTOX Blue-positive population (dead cells) was excluded from the analysis. Gating was performed with FlowJo v10.8.1. Resulting cell measurements were exported as a .csv and analyzed with custom R scripts.

6.4.3. Analysis of flow cytometry data

Cells with GFP intensities below 1 A.U. were removed before calculating statistics. This accounted for an average of 0.36% of cells discarded per sample (0.96 quantile = 2% of removed cells). GFP fluorescence intensities were subsequently log10-transformed and the mean, standard deviation (SD), and coefficient of variation (CV = sd/mean) of each population were calculated.

Curve fitting of data points was performed using the 'loess' function in ggplot:

```
stat_smooth(method = 'loess', span = 0.75, level = 0.95)
```

6.5. Fluorescence in Situ Hybridization

6.5.1 Sample processing and acquisition

I adapted the Stellaris RNA FISH protocol for *S. cerevisiae*. Cells at OD_{600} 0.4 were fixed with formaldehyde to a final concentration of 3.7% and incubated at room temperature for 45 min. Cells were washed twice with cold Buffer B and resuspended in Spheroplast Buffer pre-warmed at 30 °C. Cells were lysed for 30 min at 30 °C by adding lyticase at a final concentration of 375 U/mL. Cells were washed twice with cold Resuspension Buffer and permeabilized in 70% ethanol at 4 °C overnight.

The following day, cells were resuspended in Hybridization Buffer containing 125 nM of Gal1-targeted fluorescent probes. The probes were designed by Stellaris Probe Designer, were conjugated to Quasar 670 fluorophore, and targeted the gene body of the endogenous Gal1 sequence. After resuspension in Hybridization Buffer, cells were incubated in the dark at 30 °C overnight. The following day, excess probe was washed away with Wash Buffer. Cells were incubated in DAPI (10 ng/mL in PBS) for 1 hour at 30 °C in the dark, washed once with PBS, and mounted in microscope slides using ProLong Gold. Slides were left to polymerize overnight and imaged on the following day. Depending on the number of samples per experiment, it took between 1-7 days for samples to be imaged, during which the slides were stored at -20 °C in the dark.

A Nikon Eclipse Ti-E with SPECTRA X light engine illumination and an Andor iXon Ultra 888 camera were used for epifluorescence microscopy. A plan-apo λ 100x/1.45 NA Ph3 oil immersion objective was used to take phase contrast and fluorescence images. 21 z-planes (0.25 um each) were acquired for each position. Images were captured at 150x magnification, and enough positions were imaged to encompass at least 100 cells per sample. Mean number of cells per sample was 409 cells; 80% of the samples had 233 cells or more).

6.5.2. Analysis of FISH images

6.5.2.1 Cell segmentation

Images were processed with Cell-ACDC v1.4.8 [169] for cell segmentation. The automatic, 2D segmentation was performed on the phase contrast images by YeaZ using parameters:

```
Threshold value     0
Min distance        10 pixels
Minimum area        100 pixels
Minimum solidity    0.90
Max elongation      3
```

Predicted cells were filtered based on (i) maximum size to discard multiple cells detected as a single object (1-2%), and (ii) based on phase contrast intensity to discard dead cells (1-3%).

6.5.2.2 Fluorescence quantification and statistics

Cell-ACDC was used for quantification of Quasar 670 and GFP fluorescent signals of the focal plane of the z-stack. A square of ~128 x 128 pixels containing no cells (Background ROI) was used for calculating the background fluorescence of each image. Quasar 670 and GFP fluorescent signal was background-corrected by subtracting the median fluorescence of the Background ROI to the cell fluorescence. Total fluorescence of the cell was divided by the cell area (sum of fluor. of all pixels / number of pixels). Results were exported from Cell-ACDC and analyzed with custom R scripts.

Fluorescence was normalized in each replicate as the fold enrichment compared to mean WT signal at 0h. For the two replicates that did not include a 0h timepoint, a “pseudo 0h” fluorescence was calculated to estimate the autofluorescence of uninduced cells. “Pseudo 0h” was determined by pulling together all cells at the earliest time point acquired (2 hours after galactose switch). Since some strains were slow inducers, pulling all samples together ensured that a negative population was identifiable, compared to using WT cells at 2h, which were mostly induced. The “pseudo 0h” value

corresponded to the negative peak of the linear fluorescence distribution of all cells at 2h post-induction.

Mean, standard deviation (sd), and Coefficient of Variation (CV = sd/mean) were calculated on the normalized fluorescence for Quasar 670 and GFP signals. Curve fitting of data points was performed using the 'loess' function in ggplot:

```
stat_smooth(method = 'loess', span = 0.85, level = 0.95)
```

6.5.2.3 Transcription Site and mRNA detection

For detecting Transcription Sites (TS), I applied spotMAX (Padovani et al., in preparation) paired to the segmentation mask obtained from Cell-ACDC/YeaZ. SpotMAX was run on the Quasar 670 channel using the following parameters:

Calculate ref. channel network length?	FALSE
Compute spots size?	TRUE
Cy5 emission wavelength (nm):	675
Detection method:	peak_local_max
Effect size limit:	0.5
Effect size used:	effsize_glass_s
Filter by segmenting ref. channel?	FALSE
Filter good peaks method:	effect size
Filter spots by reference channel?	FALSE
Filter spots too close to z-boundaries?	FALSE
Fit 3D Gaussians?	FALSE
Gaussian filter sigma:	0.75
Invert the filtering of good peaks?	FALSE
Is ref. channel a single object per cell?	FALSE
Load a reference channel?	FALSE
Local or global threshold for spot detection?	Global
Numerical aperture:	1.45
Peak finder threshold function:	threshold_li
Reference channel threshold function:	threshold_li
Segmentation info (ignore if not present):	2D
Sharpen image prior spot detection?	TRUE
Spots channel name:	Cy5
Spots file name:	Cy5
Spotsize limits (pxl)	1.0, 9.6
YX resolution multiplier:	1
Z resolution limit (um):	1
ZYX voxel size (um):	[0.25, 0.0862, 0.0862]

The biggest challenge was to find a set of parameters that worked across biological replicates and across different levels of Gal1 induction. A lenient threshold of '`Effect Size [Glass] ≥ 0.5`' was used for a first round of detection, after which false positives were removed.

The identified spots were divided into four clusters based on their values for '`Effect Size [Glass]`' and '`voxspot`' (highest voxel fluorescence of the spot). Both metrics showed bimodal distributions, and the thresholds were chosen as the valley that separated the peaks of the distributions.

Manual inspection of the detected spots showed that the active TS corresponded to the cluster of spots that had higher '`Effect Size [Glass]`' and '`voxspot`' than the thresholds. Spots with lower values in both thresholds were considered false positives and were removed prior to any statistical calculation. Any other spot was determined to be an mRNA, and the estimated mRNA counts were used to fit the mathematical model of Figs. 3.6, 3.7, and S2 (see details below).

The combined outputs of Cell-ACDC and spotMAX were exported as a single dataset and the results were analyzed with custom R scripts.

6.6. Mathematical Modelling

The mathematical models were written and run in MATLAB R2017a. The pipeline utilized AMICI [170] for simulation of the ordinary differential equations. AMCI was run with the following parameters:

```
options.ami.sensi_meth    'forward'
options.ami.atol          1e-15
options.ami.rtol          1e-8
options.ami.sensi         0
```

The compilation of `.mex` simulation files by AMICI was done using gcc v4.8.5 and g++ v4.8.5 in Linux.

Model parameters were estimated with PESTO [171] by performing multi-start local optimization of the negative loglikelihood. PESTO was run with default options and only the number of random starts was modified. To determine the appropriate number of DNA steps, 10 random starts were used for all strains except sap30Δ, in which the number had to be increased to 100 starts to improve convergence. For model comparisons between WT and mutants, the number of starts was set to 20.

The mathematical model consisted of six rates (γ_{on}, γ_{off}, λ_{prod}, λ_{deg}, κ_{prod}, κ_{deg}) explained in the main text, plus five additional parameters ($mRNA_{01}$, v_{01}, P_{01}, vP_{01}, $vmRNAP_{01}$) representing the mean and variance of the mRNA, the mean and variance of the protein, and the covariance between mRNA and protein, respectively. This created a set of 11 parameters to be estimated for each strain. The five last parameters were assumed to be different between WT and mutant. Parameters were calculated in the log10 space to improve the convexity of the optimization problem [172]. The parameter space had lower and upper boundaries set at -4 and +4, respectively, except for λ_{deg}, $mRNA_{01}$, vP_{01}, and $vmRNAP_{01}$, in which the lower boundaries were shifted to -10.

The model was fitted to data from the smFISH experiments (Fig. 3.5). The proportion of cells in the DNA_{on} state was calculated as the fraction of cells exhibiting an active transcription site, identifiable as a single bright spot in the Quasar 670 channel (see Fig. 3.5a). The mean and standard deviation (sd) of mRNA were fitted using the estimated mRNA counts from spotMAX. The mean and sd of protein were fitted using the GFP fluorescence normalized by cell area and scaled down by 10^3 to match the range of the rest of the parameters.

Model selection was performed using the Bayes Information Criterion (BIC). BIC is defined as:

$$\mathrm{BIC} = \log(n)\, k - 2logL$$

where n is the number of data points, k is the number of parameters and $logL$ is the log-likelihood value for the maximum likelihood estimate of the model parameters. A threshold of 10 units above the lowest BIC was chosen to define the set of best models.

To estimate the number of intermediate DNA steps, 11 models were compared for each strain (0-10 steps). The best number of steps for each strain was selected as the model with the lowest BIC. The average of the resulting numbers was selected as the common number of DNA steps for all strains. sap30Δ was not considered in the average as the best number of steps could not be reliably determined (Fig. 3.6b).

To identify the differences in expression dynamics between WT and mutant strains, the model was fitted in pairs of WT-mutant. Out of the 11 parameters estimated for each strain (see above), 5 were

strain-specific and 6 could be shared between WT and mutant. This generated 2^6 = 64 models (Fig. S2a) for each WT-mutant pair. The total number of different parameters to be estimated ranged from 22 (model #1: all 11 parameters different between strains) to 16 (model #64: the 6 rates are shared among strains, but each strain has its own mean, sd, and covariance for mRNA and protein). Because BIC penalizes against the use of additional parameters, it guarantees that the rates of WT and mutant are identical unless strain-specific rates significantly improve the maximum likelihood of the model.

After determining the set of best models for each WT-mutant pair, we categorized the best models depending on whether, for a given pair of rates (x_i, x_j), the model had:

i) Same x_i, same x_j

ii) Different x_i, same x_j

iii) Same x_i, different x_j

iv) Different x_i, different x_j

The (x_i, x_j) pairs corresponded to DNA activation/inactivation (γ_{on}, γ_{off}), mRNA production/degradation (λ_{prod}, λ_{deg}), and protein production/degradation (κ_{prod}, κ_{deg}) (Fig. 3.7b and S2b). Depending on the number of models in each category, we determined for each mutant whether the differences to WT occurred at the DNA, mRNA, and/or protein level.

7. SUPPLEMENTARY MATERIAL

Table S1: Deletion mutants used in the high-throughput screening.

Gene name	Systematic name	Gene name	Systematic name	Gene name	Systematic name	Gene name	Systematic name	Gene name	Systematic name
ABP140	YOR239W	ELP6	YMR312W	MBP1	YDL056W	RCR1	YBR005W	SPT3	YDR392W
ACH1	YBL015W	ERG10	YPL028W	MCK1	YNL307C	RDH54	YBR073W	SPT8	YLR055C
ACS1	YAL054C	ERG13	YML126C	MCM16	YPR046W	REC104	YHR157W	SQS1	YNL224C
ADE1	YAR015W	ERG3	YLR056W	MCM21	YDR318W	REC107	YJR021C	SRB2	YHR041C
AHA1	YDR214W	ERS1	YCR075C	MCM22	YJR135C	REC114	YMR133W	SRB8	YCR081W
AHC1	YOR023C	ESC1	YMR219W	MED1	YPR070W	REC8	YPR007C	SRC1	YML034W
AHC2	YCR082W	ESC2	YDR363W	MEK1	YOR351C	RED1	YLR263W	SRO9	YCL037C
AIM2	YAL049C	ESC8	YOL017W	MEP1	YGR121C	REV1	YOR346W	SRS2	YJL092W
AIM4	YBR194W	EST1	YLR233C	MET1	YKR069W	REV3	YPL167C	SSE2	YBR169C
AIR1	YIL079C	EXO1	YOR033C	MET18	YIL128W	RFA3	YJL173C	STB5	YHR178W
AIR2	YDL175C	FBP1	YLR377C	MFT1	YML062C	RFM1	YOR279C	STE20	YHL007C
APC9	YLR102C	FKH1	YIL131C	MGS1	YNL218W	RIC1	YLR039C	STE50	YCL032W
APL3	YBL037W	FKH2	YNL068C	MHF1	YOL086W-A	RIF1	YBR275C	SUB1	YMR039C
APQ12	YIL040W	FLO1	YAR050W	MHF2	YDL160C-A	RIF2	YLR453C	SUC2	YIL162W
ARP1	YHR129C	FPR1	YNL135C	MHT1	YLL062C	RIM1	YCR028C-A	SUM1	YDR310C
ARP6	YLR085C	FPR2	YDR519W	MIG1	YGL035C	RKM1	YPL208W	SWC3	YAL011W
ASF1	YJL115W	FPR3	YML074C	MIH1	YMR036C	RKM2	YDR198C	SWC5	YBR231C
ASK10	YGR097W	FPR4	YLR449W	MLH1	YMR167W	RKM3	YBR030W	SWC7	YLR385C
ASR1	YPR093C	FUB1	YCR076C	MMS1	YPR164W	RKM4	YDR257C	SWD1	YAR003W
AVT5	YBL089W	FUN30	YAL019W	MMS2	YGL087C	RKM5	YLR137W	SWD3	YBR175W
BCK2	YER167W	FUS3	YBL016W	MMS22	YLR320W	RKR1	YMR247C	SWE1	YJL187C
BDF2	YDL070W	FYV10	YIL097W	MMS4	YBR098W	RLF2	YPR018W	SWI3	YJL176C
BLM10	YFL007W	GAL2	YLR081W	MND2	YIR025W	RMD5	YDR255C	SWI4	YER111C
BMH1	YER177W	GAL3	YDR009W	MPH1	YIR002C	RMI1	YPL024W	SWI5	YDR146C
BMT2	YBR141C	GAL4	YPL248C	MPP6	YNR024W	RMT2	YDR465C	SWI6	YLR182W
BMT5	YIL096C	GAL80	YML051W	MPT5	YGL178W	RNR3	YIL066C	SWM1	YDR260C
BMT6	YLR063W	GAL83	YER027C	MRC1	YCL061C	RPA14	YDR156W	SWP82	YFL049W
BRE1	YDL074C	GBP2	YCL011C	MRE11	YMR224C	RPA34	YJL148W	SWR1	YDR334W
BRE2	YLR015W	GCR2	YNL199C	MRM2	YGL136C	RPD3	YNL330C	TAE1	YBR261C
BRE5	YNR051C	GID8	YMR135C	MRN1	YPL184C	RPL4A	YBR031W	TAH1	YCR060W
BUB1	YGR188C	GIM3	YNL153C	MSA2	YKR077W	RPN10	YHR200W	TEC1	YBR083W
BUD13	YGL174W	GIM4	YEL003W	MSH2	YOL090W	RPN13	YLR421C	TEL1	YBL088C
BUD27	YFL023W	GIM5	YML094W	MSH3	YCR092C	RPN14	YGL004C	TEX1	YNL253W
BUL1	YMR275C	GIS1	YDR096W	MSH6	YDR097C	RPN4	YDL020C	THO1	YER063W
BUL2	YML111W	GRR1	YJR090C	MSI1	YBR195C	RRD1	YIL153W	THP2	YHR167W
CAC2	YML102W	GRX7	YBR014C	MSL1	YIR009W	RRD2	YPL152W	TOD6	YBL054W

CAD1	YDR423C	GTR1	YML121W	MSN2	YMR037C	RRM3	YHR031C	TOF1	YNL273W
CAF130	YGR134W	GTR2	YGR163W	MSN4	YKL062W	RRP6	YOR001W	TOF2	YKR010C
CAF40	YNL288W	HAT1	YPL001W	MTQ1	YNL063W	RRP8	YDR083W	TOP1	YOL006C
CAT2	YML042W	HAT2	YEL056W	MUB1	YMR100W	RSM22	YKL155C	TOP3	YLR234W
CDA1	YLR307W	HCM1	YCR065W	MUD1	YBR119W	RTF1	YGL244W	TOS4	YLR183C
CDA2	YLR308W	HDA1	YNL021W	MUM2	YBR057C	RTG1	YOL067C	TOS8	YGL096W
CDC73	YLR418C	HDA3	YPR179C	MUS81	YDR386W	RTG2	YGL252C	TRF5	YNL299W
CHD1	YER164W	HEL1	YKR017C	MYO4	YAL029C	RTG3	YBL103C	TRM12	YML005W
CHK1	YBR274W	HEL2	YDR266C	NAM7	YMR080C	RTS1	YOR014W	TRM2	YKR056W
CHL1	YPL008W	HFA1	YMR207C	NAP1	YKR048C	RTS2	YOR077W	TRM3	YDL112W
CHL4	YDR254W	HHF1	YBR009C	NDJ1	YOL104C	RTT101	YJL047C	TRM7	YBR061C
CHZ1	YER030W	HHF2	YNL030W	NEW1	YPL226W	RTT102	YGR275W	TSA1	YML028W
CIN1	YOR349W	HHO1	YPL127C	NFI1	YOR156C	RTT103	YDR289C	TTI1	YKL033W
CIN4	YMR138W	HHT1	YBR010W	NHP10	YDL002C	RTT106	YNL206C	UBA3	YPR066W
CIT1	YNR001C	HHT2	YNL031C	NHP6A	YPR052C	RTT107	YHR154W	UBC11	YOR339C
CIT2	YCR005C	HIR1	YBL008W	NNT1	YLR285W	RTT109	YLL002W	UBC12	YLR306W
CIT3	YPR001W	HIR2	YOR038C	NOP6	YDL213C	RUB1	YDR139C	UBC13	YDR092W
CKA1	YIL035C	HIR3	YJR140C	NTO1	YPR031W	RXT2	YBR095C	UBC4	YBR082C
CKA2	YOR061W	HIS3	YOR202W	NUP133	YKR082W	RXT3	YDL076C	UBC5	YDR059C
CKB1	YGL019W	HMO1	YDR174W	NUP2	YLR335W	SAC3	YDR159W	UBC7	YMR022W
CKB2	YOR039W	HMT1	YBR034C	NUP60	YAR002W	SAE2	YGL175C	UBC8	YEL012W
CLA4	YNL298W	HO	YDL227C	NUT1	YGL151W	SAM1	YLR180W	UBI4	YLL039C
CLB2	YPR119W	HOS1	YPR068C	NVJ1	YHR195W	SAM4	YPL273W	UBP1	YDL122W
CPR1	YDR155C	HOS2	YGL194C	OAF1	YAL051W	SAN1	YDR143C	UBP11	YKR098C
CPR6	YLR216C	HOS3	YPL116W	OGG1	YML060W	SAP190	YKR028W	UBP12	YJL197W
CPR7	YJR032W	HOS4	YIL112W	OMS1	YDR316W	SAP30	YMR263W	UBP13	YBL067C
CRC1	YOR100C	HPA2	YPR193C	PAC10	YGR078C	SAS2	YMR127C	UBP14	YBR058C
CRG1	YHR209W	HPA3	YEL066W	PAC2	YER007W	SAS3	YBL052C	UBP15	YMR304W
CSE2	YNR010W	HPC2	YBR215W	PAT1	YCR077C	SAS4	YDR181C	UBP2	YOR124C
CSM1	YCR086W	HPM1	YIL110W	PAU7	YAR020C	SAS5	YOR213C	UBP3	YER151C
CSM3	YMR048W	HRB1	YNL004W	PBY1	YBR094W	SBA1	YKL117W	UBP5	YER144C
CST6	YIL036W	HSL1	YKL101W	PDA1	YER178W	SCS22	YBL091C-A	UBP6	YFR010W
CST9	YLR394W	HSP26	YBR072W	PDC1	YLR044C	SCS7	YMR272C	UBP7	YIL156W
CTF18	YMR078C	HST1	YOL068C	PHO23	YNL097C	SDC1	YDR469W	UBP8	YMR223W
CTF19	YPL018W	HST2	YPL015C	PHO4	YFR034C	SDS3	YIL084C	UBP9	YER098W
CTF3	YLR381W	HST3	YOR025W	PHO80	YOL001W	SEM1	YDR363W-A	UBR1	YGR184C
CTF4	YPR135W	HST4	YDR191W	PIH1	YHR034C	SER2	YGR208W	UBR2	YLR024C
CTF8	YHR191C	HTA1	YDR225W	PIL1	YGR086C	SET2	YJL168C	UFD2	YDL190C
CTI6	YPL181W	HTA2	YBL003C	PIP2	YOR363C	SET3	YKR029C	UFD4	YKL010C
CTM1	YHR109W	HTZ1	YOL012C	PLM2	YDR501W	SET4	YJL105W	UGA2	YBR006W

CWC21	YDR482C	HUL4	YJR036C	POG1	YIL122W	SET5	YHR207C	UGA3	YDL170W
DAL81	YIR023W	HUR1	YGL168W	POL3	YDL102W	SET6	YPL165C	ULS1	YOR191W
DBR1	YKL149C	HXT17	YNR072W	POL32	YJR043C	SFM1	YOR021C	UME1	YPL139C
DCC1	YCL016C	IES1	YFL013C	POT1	YIL160C	SGF11	YPL047W	UME6	YDR207C
DEG1	YFL001W	IES2	YNL215W	PPH21	YDL134C	SGF29	YCL010C	VID28	YIL017C
DEP1	YAL013W	IES4	YOR189W	PPH22	YDL188C	SGF73	YGL066W	VID30	YGL227W
DLS1	YJL065C	IES5	YER092W	PPH3	YDR075W	SGN1	YIR001C	VMA1	YDL185W
DMA1	YHR115C	IKI3	YLR384C	PPM1	YDR435C	SGS1	YMR190C	VPS71	YML041C
DMC1	YER179W	IML3	YBR107C	PRE9	YGR135W	SHE1	YBL031W	VPS72	YDR485C
DNL4	YOR005C	INO4	YOL108C	PSH1	YOL054W	SHG1	YBR258C	VPS75	YNL246W
DOA1	YKL213C	IOC2	YLR095C	PSK1	YAL017W	SIF2	YBR103W	WHI4	YDL224C
DOT1	YDR440W	IOC3	YFR013W	PSY2	YNL201C	SIN3	YOL004W	XRN1	YGL173C
DOT5	YIL010W	IOC4	YMR044W	PSY4	YBL046W	SIR1	YKR101W	XRS2	YDR369C
DOT6	YER088C	IRC20	YLR247C	RAD1	YPL022W	SIS2	YKR072C	YAF9	YNL107W
DPB3	YBR278W	IRC5	YFR038W	RAD10	YML095C	SIZ1	YDR409W	YBP2	YGL060W
DPB4	YDR121W	IST3	YIR005W	RAD16	YBR114W	SKA1	YKL023W	YCK3	YER123W
DPH5	YLR172C	ISW1	YBR245C	RAD23	YEL037C	SKI8	YGL213C	YHP1	YDR451C
DSF2	YBR007C	ISW2	YOR304W	RAD26	YJR035W	SKY1	YMR216C	YKE2	YLR200W
DST1	YGL043W	ITC1	YGL133W	RAD27	YKL113C	SLI1	YGR212W	YKU70	YMR284W
DUR12	YBR208C	IWR1	YDL115C	RAD30	YDR419W	SLK19	YOR195W	YKU80	YMR106C
EAF1	YDR359C	JHD1	YER051W	RAD4	YER162C	SLX1	YBR228W	YLR278C	YLR278C
EAF3	YPR023C	JHD2	YJR119C	RAD5	YLR032W	SLX4	YLR135W	YMR209C	YMR209C
EAF6	YJR082C	KTI12	YKL110C	RAD51	YER095W	SLX5	YDL013W	YNG1	YOR064C
EAF7	YNL136W	LAT1	YNL071W	RAD52	YML032C	SLX8	YER116C	YNL092W	YNL092W
ECM15	YBL001C	LEO1	YOR123C	RAD54	YGL163C	SNF1	YDR477W	YOR338W	YOR338W
ECM5	YMR176W	LGE1	YPL055C	RAD55	YDR076W	SNF11	YDR073W	YRM1	YOR172W
EDC1	YGL222C	LHP1	YDL051W	RAD57	YDR004W	SNF3	YDL194W	YRR1	YOR162C
EDC2	YER035W	LIF1	YGL090W	RAD59	YDL059C	SNF4	YGL115W	YSA1	YBR111C
EFM1	YHL039W	LRP1	YHR081W	RAD6	YGL058W	SNT1	YCR033W	YSY6	YBR162W-A
EFM2	YBR271W	LSM1	YJL124C	RAD61	YDR014W	SNT2	YGL131C	YTA7	YGR270W
EFM4	YIL064W	LSM6	YDR378C	RAD7	YJR052W	SNU66	YOR308C	YUH1	YJR099W
ELC1	YPL046C	LSM7	YNL147W	RAD9	YDR217C	SOH1	YGL127C	ZDS1	YMR273C
ELF1	YKL160W	LYS20	YDL182W	RAM1	YDL090C	SPO11	YHL022C	ZDS2	YML109W
ELG1	YOR144C	LYS21	YDL131W	RAS2	YNL098C	SPP1	YPL138C	ZIP2	YGL249W
ELP2	YGR200C	MAC1	YMR021C	RBG1	YAL036C	SPT2	YER161C	ZTA1	YBR046C
ELP3	YPL086C	MAD2	YJL030W	RCM1	YNL022C	SPT21	YMR179W		
ELP4	YPL101W	MAG1	YER142C	RCO1	YMR075W	SPT23	YKL020C		

Table S2: Deletion strains identified as high- or low-noise outliers in high-throughput screening. Marked in bold are strains considered for further testing. *Consecutive frames or with 1 frame in between consecutives.

High noise outliers	Total frames in which it is outlier	Max. consecutive frames*
BRE5	34	34
SDS3	24	18
STB5	18	18
SAP30	**16**	**14**
UGA3	15	13
FYV10	**22**	**10**
RPN10	11	10
HHF1	**9**	**9**
CPR6	14	8
MSH3	**9**	**8**
RRD1	8	8
HIR1	**11**	**7**
RRP8	9	5
SIN3	5	5
IES2	**10**	**4**
ISW2	8	4
UBP14	7	4
ERG3	6	4
PPH3	6	4
PSY2	5	4
NUP60	4	4
TOD6	4	4

Low noise outliers	Total frames in which it is outlier	Max. consecutive frames*
LGE1	12	12
UBC12	**11**	**8**
RAD51	**10**	**5**
UBA3	5	5
FUS3	**10**	**4**
SWD1	5	4

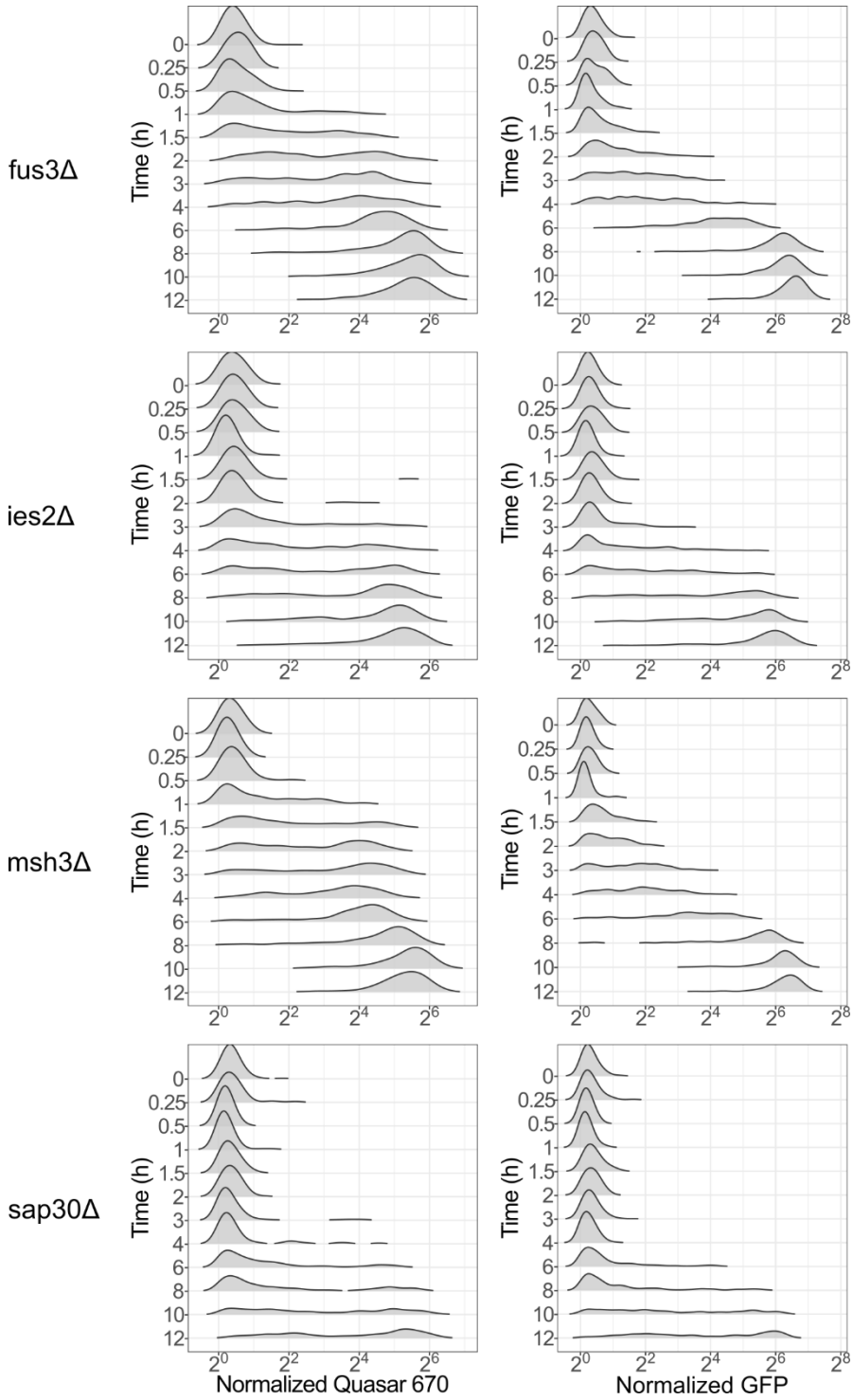


Figure S1: Bimodal expression of Gal1 under galactose induction. Plotted are normalized Quasar 670 intensities corresponding to Gal1-GFP transcripts (left) and normalized GFP intensities corresponding to Gal1-GFP proteins (right) at different induction time points (y-axis). Related to Fig. 5.

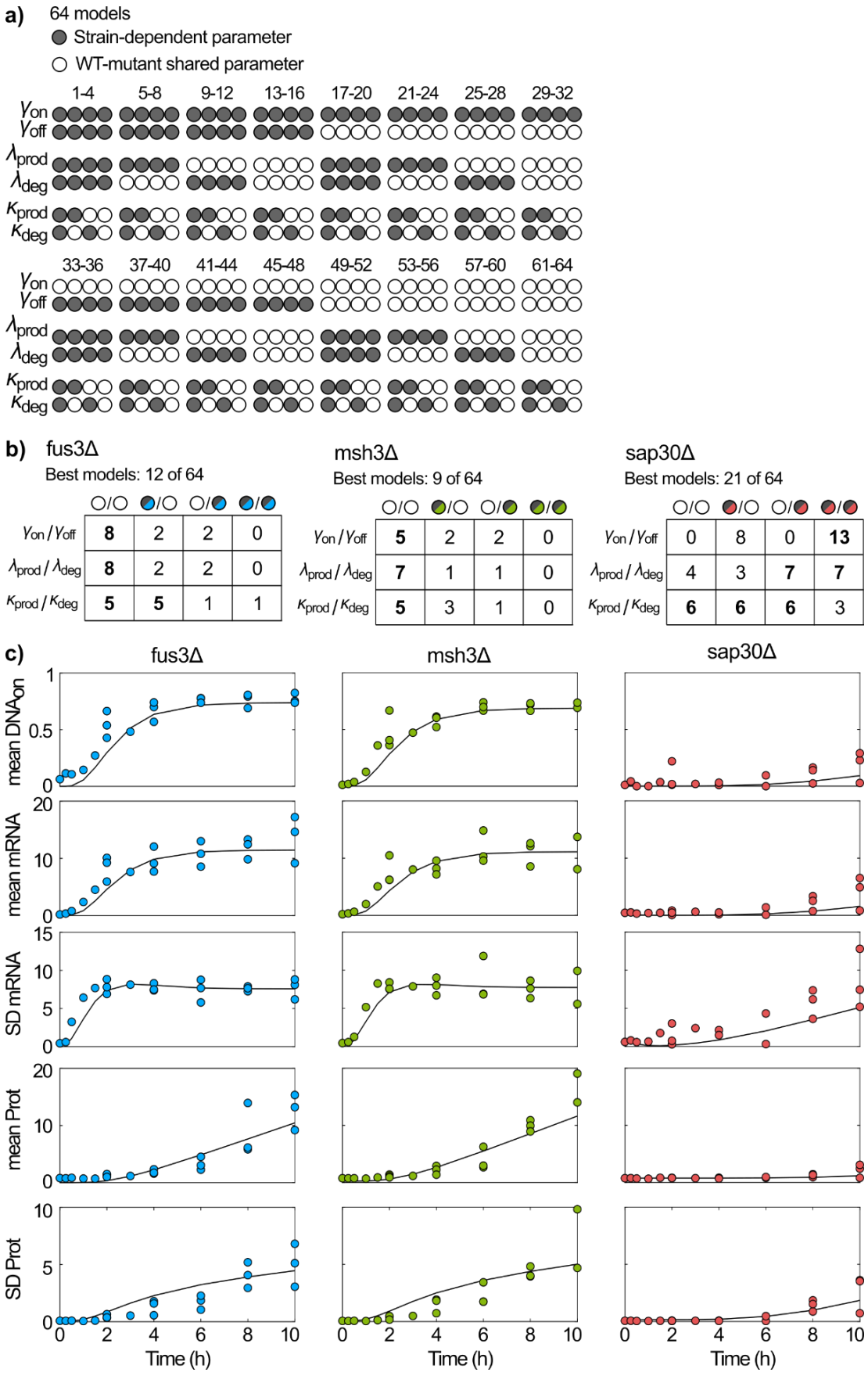


Figure S2: Mathematical modeling for mutant strains using FISH data. a) Depiction of the 64 models compared between WT and mutants. The six rates had the possibility of being equal or different among

tested strains, giving rise to 2^6 = 64 possible models. Filled circles represent the models in which WT and mutant parameters had different values; empty circles represent a shared value between WT and mutant. **b)** Categorization of the best models for each mutant when compared against WT. In bold is the category with the highest value. **c)** Fits for the best model of each strain. From top to bottom, plots show: fraction of cells in the DNA_{on} state; mean Gal1-GFP mRNA; standard deviation (SD) of mRNA; mean Gal1-GFP protein; SD of protein. Every data point corresponds to the statistic calculated for an individual biological replicate. Related to Figs. 6 and 7.

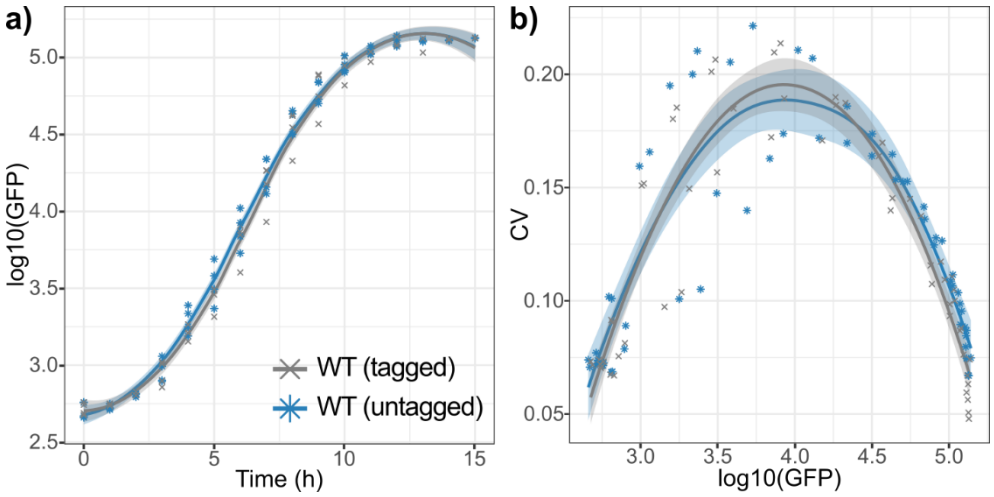


Figure S3: Addition of 2xFLAG tags to endogenous Ies2 does not alter Gal1 induction dynamics or noise trajectory. Strains were pre-grown in glucose, diluted into a 4h intermediate growth step, and induced with galactose for 12 or 15 hours. **a)** Mean population expression of Gal1-GFP over time. Plotted is mean log10(GFP) signal vs. time. **b)** Noise trajectories of Gal1-GFP over different levels of induction. Plotted is CV vs. mean log10(GFP). Statistics were calculated on merged data from four biological replicates (two 12-hour inductions, two 15-hour inductions). The average number of cells per data point was 10,224 ± 272 (mean ± sd). Line and ribbon represent the fit and 95% confidence intervals of a curve fitting with *loess*, respectively. Strains depicted are FLAG-tagged WT (gray) and untagged WT (blue). Related to Fig. 11.

8. ACKNOWLEDGEMENTS

There have been innumerable people who made this thesis possible. To everyone involved in the project: you have my warmest and most sincere thanks.

To Rob Schneider, for hosting me in his lab when I was a bachelor student. For inviting me to come back and do a PhD under his supervision. For replying to my midnight calls when the smFISH buffer turned blue and I didn't know what to do. For every scientific conversation we have had over the last couple of years. Thank you.

To Poonam Bheda, for starting this project with me and for all the help and supervision in the experimental aspects of my PhD. For being a great mentor, inside and outside of the lab. For keeping in touch when she left the lab and for all the times we've spent together since then. Thank you.

To Lea Schuh, Dantong Wang, and Carsten Marr, for their contributions to this project. For opening the doors of the ICB/AIH for me and making me feel welcome. For answering all my last-minute emails and taking the time to explain the ins and outs of modeling. Thank you.

To María Rocha Acevedo, for her incredible and vital support in the smFISH experiments. For staying all those weekend nights in the lab with me collecting samples. For helping me with yeast cultures. For being the best ahijada I could have asked for. For being a friend. Thank you. Figs. 3.5 and 3.10 go for you!

To Francesco Padovani, for all his support with the Cell-ACDC and spotMAX analyses. For his prompt replies to my weekend messages and for modifying the pipelines to make my analyses easier. For being "the best IT support I have ever seen". Thank you.

To Igor Kukhtevich, for all his aid with imaging and yeast handling. For teaching me to use the microscope via Zoom during COVID-19 times. For all his experimental guidance. Thank you.

To Scarlett Dornauer, for making the Ies2 constructs for the project. For taking the time to answer all my "Scarlett, how do you do (insert laboratory technique)?" as a non-experimental person trying not to set the lab on fire. Thank you.

To every member of the Institute of Functional Epigenetics, for their support throughout my doctoral journey. Special shoutouts to Thomas Gerling, for helping me navigate the sea of paperwork inherent to German bureaucracy, and the Northlab, for dealing with me hoarding the elephant foot and running around like a headless chicken in experiment days. Thank you.

To my friends and family on both sides of the pond. For listening to my rants about life and science. For their infinite moral and emotional support. Thank you.

And finally, the biggest thanks go to my partner Scott Atwell. For keeping me level-headed and supporting me throughout my entire PhD. For all his science and life advice. For his immeasurable patience. For being the shoulders on which I stand. Thank you; this wouldn't have been possible without you.

9. APPENDIX

9.1. Abbreviations

ac	Acetyl
ADP	Adenosine diphosphate
ATc	Anhydrotetracycline
ATP	Adenosine triphosphate
A.U.	Arbitrary units
B. subtilis	*Bacillus subtilis*
BIC	Bayesian Information Criterion
°C	Degree Celsius
C. thermophilum	*Chaetomium thermophilum*
CV	Coefficient of Variation
DAPI	4',6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FACS	Fluorescence-activated cell sorting
FDG	fluorescein-di-β-D-galactopyranoside
Fig.	Figure
FISH	Fluorescence *in situ* hybridization
fluor	fluorescence
FSC-A, -W	Forward scatter – area, - width
h	Hour
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HDM	Histone demethylase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMT	Histone methyltransferase
INO80c	INO80 complex
IPTG	Isopropyl β- d-1-thiogalactopyranoside
IQR	Interquartile range
K	Lysine
loess	Locally estimated scatterplot smoothing
MAP	Mitogen-activated protein
MAPK	MAP kinase
me	Methyl
min	Minute
mL	Milliliters
mRNA	Messenger RNA
NA	Numerical aperture
NDR	Nucleosome depleted region
nm	Nanometer
nM	Nanomolar

OD	Optical density
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
R	Arginine
RNA	Ribonucleic acid
RNA Pol	RNA polymerase
RNAPII	RNA polymerase II
ROI	Region of interest
S	Serine
S. cerevisiae	*Saccharomyces cerevisiae*
scRNA-seq	single-cell RNA sequencing
sd	Standard deviation
SGA	Synthetic genetic array
smFISH	Single molecule fluorescence *in situ* hybridization
SSC	Saline-sodium citrate
SSC-A	Side scatter - area
T	Threonine
TA-sev	TATA box severely impaired
TA-WT	Wild type TATA box
TF	Transcription factor
tRNA	Transfer RNA
TS	Transcription site
U	Units
UAS	Upstream activating sequence
UDP	Uridine diphosphate
um	micrometer
VRC	Ribonucleoside Vanadyl Complex
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
WT	Wild type
μ	Mean
σ	Standard deviation

9.2. Figure licenses

Fig. 1.1b: "Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), 251-260." Shared under Springer Nature license #5715541296974

Fig. 1.4a: "Ko, M. S., Nakauchi, H., & Takahashi, N. (1990). The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *The EMBO journal*, *9*(9), 2835-2842." Shared under Springer Nature license RP2024074.

Fig. 1.4b: "Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science*, *297*(5584), 1183-1186." Reprinted with permission from AAAS.

Fig. 1.4c: “Raser, J. M., & O'Shea, E. K. (2004). Control of stochasticity in eukaryotic gene expression. *science*, *304*(5678), 1811-1814.”. Reprinted with permission from AAAS.

10. BIBLIOGRAPHY

1. Luger K, Maeder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A˚resolution. Nature. 1997;389.

2. Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution. Journal of Molecular Biology. 2002;319:1097–113.

3. Woodcock CL, Skoultchi AI, Fan Y. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. Chromosome Res. 2006;14:17–25.

4. Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R. Molecular Biology of the Gene. Pearson Education; 2013.

5. CNX OpenStax. Biology. Wikimedia Commons. 2016.

6. Jansen A, Verstrepen KJ. Nucleosome Positioning in Saccharomyces cerevisiae. Microbiology and Molecular Biology Reviews. 2011;75:301–20.

7. Thoma F, Koller T, Klug A. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J Cell Biol. 1979;83 2 Pt 1:403–27.

8. Zhu P, Li G. Structural insights of nucleosome and the 30-nm chromatin fiber. Current Opinion in Structural Biology. 2016;36:106–15.

9. Tremethick DJ. Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. Cell. 2007;128:651–4.

10. Bonev B, Cavalli G. Organization and function of the 3D genome. Nat Rev Genet. 2016;17:661–78.

11. Heitz E. Das Heterochromatin der Moose. Jahrbücher für Wissenschaftliche Botanik. 1928;69:762–818.

12. Berger F. Emil Heitz, a true epigenetics pioneer. Nat Rev Mol Cell Biol. 2019;20:572–572.

13. Morrison O, Thakur J. Molecular Complexes at Euchromatin, Heterochromatin and Centromeric Chromatin. International Journal of Molecular Sciences. 2021;22:6922.

14. Proudfoot NJ, Furger A, Dye MJ. Integrating mRNA Processing with Transcription. Cell. 2002;108:501–12.

15. Rosonina E, Kaneko S, Manley JL. Terminating the transcript: breaking up is hard to do. Genes Dev. 2006;20:1050–6.

16. Reed R, Hurt E. A Conserved mRNA Export Machinery Coupled to pre-mRNA Splicing. Cell. 2002;108:523–31.

17. Tunnacliffe E, Chubb JR. What Is a Transcriptional Burst? Trends in Genetics. 2020;36:288–97.

18. Nicolas D, Phillips NE, Naef F. What shapes eukaryotic transcriptional bursting? Molecular BioSystems. 2017;13:1280–90.

19. Lenstra TL, Rodriguez J, Chen H, Larson DR. Transcription Dynamics in Living Cells. Annual Review of Biophysics. 2016;45:25–47.

20. Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. Science. 2011;332:472–4.

21. Dadiani M, Dijk D van, Segal B, Field Y, Ben-Artzi G, Raveh-Sadka T, et al. Two DNA-encoded strategies for increasing expression with opposing effects on promoter dynamics and transcriptional noise. Genome Res. 2013;23:966–76.

22. Hornung G, Bar-Ziv R, Rosin D, Tokuriki N, Tawfik DS, Oren M, et al. Noise–mean relationship in mutated promoters. Genome Res. 2012;22:2409–17.

23. Blake WJ, Balázsi G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, et al. Phenotypic Consequences of Promoter-Mediated Transcriptional Noise. Molecular Cell. 2006;24:853–65.

24. Walters MC, Fiering S, Eidemiller J, Magis W, Groudine M, Martin DI. Enhancers increase the probability but not the level of gene expression. Proceedings of the National Academy of Sciences. 1995;92:7125–9.

25. Bartman CR, Hsu SC, Hsiung CC-S, Raj A, Blobel GA. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. Molecular Cell. 2016;62:237–47.

26. Nicolas D, Zoller B, Suter DM, Naef F. Modulation of transcriptional burst frequency by histone acetylation. Proceedings of the National Academy of Sciences of the United States of America. 2018;115:7153–8.

27. Chen L-F, Lin YT, Gallegos DA, Hazlett MF, Gómez-Schiavon M, Yang MG, et al. Enhancer Histone Acetylation Modulates Transcriptional Bursting Dynamics of Neuronal Activity-Inducible Genes. Cell Reports. 2019;26:1174-1188.e5.

28. Harper CV, Finkenstädt B, Woodcock DJ, Friedrichsen S, Semprini S, Ashall L, et al. Dynamic Analysis of Stochastic Transcription Cycles. PLoS Biol. 2011;9:e1000607.

29. Becker PB, Workman JL. Nucleosome Remodeling and Epigenetics. Cold Spring Harb Perspect Biol. 2013;5:a017905.

30. Prajapati HK, Ocampo J, Clark DJ. Interplay among ATP-Dependent Chromatin Remodelers Determines Chromatin Organisation in Yeast. Biology. 2020;9:190.

31. Clapier CR, Cairns BR. The Biology of Chromatin Remodeling Complexes. Annu Rev Biochem. 2009;78:273–304.

32. Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. Nat Rev Mol Cell Biol. 2015;16:178–89.

33. Oberbeckmann E, Niebauer V, Watanabe S, Farnung L, Moldt M, Schmid A, et al. Ruler elements in chromatin remodelers set nucleosome array spacing and phasing. Nature Communications. 2021;12:3232.

34. Singh A, Modak SB, Chaturvedi MM, Purohit JS. SWI/SNF Chromatin Remodelers: Structural, Functional and Mechanistic Implications. Cell Biochem Biophys. 2023;81:167–87.

35. Mellor J, Morillon A. ISWI complexes in Saccharomyces cerevisiae. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 2004;1677:100–12.

36. Murawska M, Brehm A. CHD chromatin remodelers and the transcription cycle. Transcription. 2011;2:244–53.

37. Willhoft O, Wigley DB. INO80 and SWR1 complexes: the non-identical twins of chromatin remodelling. Current Opinion in Structural Biology. 2020;61:50–8.

38. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. 2011;21:381–95.

39. Chan JC, Maze I. Nothing Is Yet Set in (Hi)stone: Novel Post-Translational Modifications Regulating Chromatin Function. Trends in Biochemical Sciences. 2020;45:829–44.

40. Chou KY, Lee J-Y, Kim K-B, Kim E, Lee H-S, Ryu H-Y. Histone modification in Saccharomyces cerevisiae: A review of the current status. Computational and Structural Biotechnology Journal. 2023;21:1843–50.

41. Allfrey VG, Faulkner R, Mirsky AE. Acetylation and methylation of histones and their possible role in the regulation of rna synthesis*. Proceedings of the National Academy of Sciences. 1964;51:786–94.

42. Shahbazian MD, Grunstein M. Functions of Site-Specific Histone Acetylation and Deacetylation. Annu Rev Biochem. 2007;76:75–100.

43. Zaware N, Zhou M-M. Bromodomain biology and drug discovery. Nat Struct Mol Biol. 2019;26:870–9.

44. Grunstein M. Histone acetylation in chromatin structure and transcription. Nature. 1997;389:349–52.

45. Wagner EJ, Carpenter PB. Understanding the language of Lys36 methylation at histone H3. Nat Rev Mol Cell Biol. 2012;13:115–26.

46. Allshire RC, Madhani HD. Ten principles of heterochromatin formation and function. Nat Rev Mol Cell Biol. 2018;19:229–44.

47. Park K, Kim J-A, Kim J. Transcriptional regulation by the KMT2 histone H3K4 methyltransferases. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2020;1863:194545.

48. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. Exp Mol Med. 2017;49:e324–e324.

49. Swygert SG, Peterson CL. Chromatin dynamics: Interplay between remodeling enzymes and histone modifications. Biochim Biophys Acta. 2014;1839:728–36.

50. Orphanides G, Reinberg D. A Unified Theory of Gene Expression. Cell. 2002;108:439–51.

51. Spudich JL, Koshland DE. Non-genetic individuality: chance in the single cell. Nature. 1976;262:467–71.

52. Ko MSH. Induction mechanism of a single gene molecule: Stochastic or deterministic? BioEssays. 1992;14:341–6.

53. Ko MS, Nakauchi H, Takahashi N. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. The EMBO Journal. 1990;9:2835–42.

54. Burn SF. Detection of β-Galactosidase Activity: X-gal Staining. In: Michos O, editor. Kidney Development. Totowa, NJ: Humana Press; 2012. p. 241–50.

55. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic Gene Expression in a Single Cell. Science. 2002;297:1183–6.

56. Raser JM, O'Shea EK. Control of Stochasticity in Eukaryotic Gene Expression. Science. 2004;304:1811–4.

57. Fiering S, Northrop JP, Nolan GP, Mattila PS, Crabtree GR, Herzenberg LA. Single cell assay of a transcription factor reveals a threshold in transcription activated by signals emanating from the T-cell antigen receptor. Genes Dev. 1990;4:1823–34.

58. Rotman B. MEASUREMENT OF ACTIVITY OF SINGLE MOLECULES OF β-D-GALACTOSIDASE*. Proceedings of the National Academy of Sciences. 1961;47:1981–91.

59. McAdams HH, Arkin A. It's a noisy business! Genetic regulation at the nanomolar scale. Trends in Genetics. 1999;15:65–9.

60. Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. Proceedings of the National Academy of Sciences. 2001;98:8614–9.

61. Becskei A, Serrano L. Engineering stability in gene networks by autoregulation. Nature. 2000;405:590–3.

62. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. Nature. 2000;403:335–8.

63. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. Proceedings of the National Academy of Sciences of the United States of America. 2002;99:12795–800.

64. Eling N, Morgan MD, Marioni JC. Challenges in measuring and understanding biological noise. Nature Reviews Genetics. 2019;20:536–48.

65. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. Nature Genetics. 2002;31:69–73.

66. Blake WJ, Kaern M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. Nature. 2003;422:633–7.

67. Fascher K-D, Schmitz J, Hörz W. Structural and Functional Requirements for the Chromatin Transition at the PHO5 Promoter in Saccharomyces cerevisiae upon PHO5 Activation. Journal of Molecular Biology. 1993;231:658–67.

68. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, et al. Noise in protein expression scales with natural protein abundance. Nature Genetics. 2006;38:636–43.

69. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature. 2006;441:840–6.

70. Carey LB, Dijk D van, Sloot PMA, Kaandorp JA, Segal E. Promoter Sequence Determines the Relationship between Expression Level and Noise. PLOS Biology. 2013;11:e1001528.

71. Dey SS, Foley JE, Limsirichai P, Schaffer DV, Arkin AP. Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. Molecular Systems Biology. 2015;11:806.

72. Dar RD, Shaffer SM, Singh A, Razooky BS, Simpson ML, Raj A, et al. Transcriptional bursting explains the noise-versus-mean relationship in mRNA and protein levels. PLoS ONE. 2016;11.

73. Schmiedel JM, Carey LB, Lehner B. Empirical mean-noise fitness landscapes reveal the fitness impact of gene expression noise. Nature Communications. 2019;10:3180.

74. Stewart-Ornstein J, Weissman JS, El-Samad H. Cellular Noise Regulons Underlie Fluctuations in Saccharomyces cerevisiae. Molecular Cell. 2012;45:483–93.

75. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. Noise Minimization in Eukaryotic Gene Expression. PLOS Biology. 2004;2:e137.

76. Carter B, Zhao K. The epigenetic basis of cellular heterogeneity. Nature Reviews Genetics. 2021;22:235–50.

77. Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluorescent proteins. Nat Methods. 2005;2:905–9.

78. Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. Science. 1998;280:585–90.

79. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA Synthesis in Mammalian Cells. PLOS Biology. 2006;4:e309.

80. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nature Methods. 2008;5:877–9.

81. Braselmann E, Rathbun C, Richards EM, Palmer AE. Illuminating RNA Biology: Tools for Imaging RNA in Live Mammalian Cells. Cell Chemical Biology. 2020;27:891–903.

82. Grün D, Kester L, Van Oudenaarden A. Validation of noise models for single-cell transcriptomics. Nature Methods. 2014;11:637–40.

83. Litzenburger UM, Buenrostro JD, Wu B, Shen Y, Sheffield NC, Kathiria A, et al. Single-cell epigenomic variability reveals functional cancer heterogeneity. Genome Biology. 2017;18:15.

84. Liu L, Liu C, Quintero A, Wu L, Yuan Y, Wang M, et al. Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. Nat Commun. 2019;10:470.

85. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, et al. Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell. 2020;183:1103-1116.e20.

86. Muto Y, Wilson PC, Ledru N, Wu H, Dimke H, Waikar SS, et al. Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. Nat Commun. 2021;12:2190.

87. Faure AJ, Schmiedel JM, Lehner B. Systematic Analysis of the Determinants of Gene Expression Noise in Embryonic Stem Cells. Cell systems. 2017;5:471-484.e4.

88. Chalancon G, Ravarani CNJ, Balaji S, Martinez-Arias A, Aravind L, Jothi R, et al. Interplay between gene expression noise and regulatory network architecture. Trends in Genetics. 2012;28:221–32.

89. Lewontin RC. On the measurement of relative variability. Systematic Zoology. 1966;15:141–2.

90. Phillips R, Belliveau NM, Chure G, Garcia HG, Razo-Mejia M, Scholes C. Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression. Annual Review of Biophysics. 2019;48:121–63.

91. Feng Y-L, Sun J, Liu Y-F, Ren J-G, Dong J-M. Effects of restrained degradation on gene expression and regulation. arXiv. 2019.

92. Kumar N, Kulkarni RV. Constraining the complexity of promoter dynamics using fluctuations in gene expression. Phys Biol. 2019;17:015001.

93. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature. 2015;523:486–90.

94. Jin W, Tang Q, Wan M, Cui K, Zhang Y, Ren G, et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. Nature. 2015;528:142–6.

95. Fraser LTCR, Dikdan RJ, Dey S, Singh A, Tyagi S. Reduction in gene expression noise by targeted increase in accessibility at gene loci. Proceedings of the National Academy of Sciences of the United States of America. 2021;118.

96. Ku WL, Nakamura K, Gao W, Cui K, Hu G, Tang Q, et al. Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. Nat Methods. 2019;16:323–5.

97. Zhu C, Zhang Y, Li YE, Lucero J, Behrens MM, Ren B. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. Nat Methods. 2021;18:283–92.

98. Sellick CA, Campbell RN, Reece RJ. Galactose Metabolism in Yeast-Structure and Regulation of the Leloir Pathway Enzymes and the Genes Encoding Them. International Review of Cell and Molecular Biology. 2008;269:111–50.

99. Johnston M. A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. Microbiol Rev. 1987;51:458–76.

100. Harrison M-C, LaBella AL, Hittinger CT, Rokas A. The evolution of the GALactose utilization pathway in budding yeasts. Trends Genet. 2022;38:97–106.

101. Tschopp JF, Emr SD, Field C, Schekman R. GAL2 codes for a membrane-bound subunit of the galactose permease in Saccharomyces cerevisiae. Journal of Bacteriology. 1986;166:313–8.

102. Thoden JB, Holden HM. The Molecular Architecture of Galactose Mutarotase/UDP-Galactose 4-Epimerase from Saccharomyces cerevisiae. Journal of Biological Chemistry. 2005;280:21900–7.

103. Timson DJ, Reece RJ. Kinetic analysis of yeast galactokinase: implications for transcriptional activation of the GAL genes. Biochimie. 2002;84:265–72.

104. Bram RJ, Lue NF, Kornberg RD. A GAL family of upstream activating sequences in yeast: roles in both induction and repression of transcription. The EMBO Journal. 1986;5:603–8.

105. Ma J, Ptashne M. The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80. Cell. 1987;50:137–42.

106. Griggs DW, Johnston M. Regulated expression of the GAL4 activator gene in yeast provides a sensitive genetic switch for glucose repression. Proceedings of the National Academy of Sciences. 1991;88:8597–601.

107. Nehlin JO, Carlberg M, Ronne H. Control of yeast GAL genes by MIG1 repressor: a transcriptional cascade in the glucose response. The EMBO Journal. 1991;10:3373–7.

108. Lee B, Church M, Hokamp K, Alhussain MM, Bamagoos AA, Fleming AB. Systematic analysis of tup1 and cyc8 mutants reveals distinct roles for TUP1 and CYC8 and offers new insight into the regulation of gene transcription by the yeast Tup1-Cyc8 complex. PLOS Genetics. 2023;19:e1010876.

109. Bheda P, Kirmizis A, Schneider R. The past determines the future: sugar source history and transcriptional memory. Curr Genet. 2020;66:1029–35.

110. Platt A, Reece RJ. The yeast galactose genetic switch is mediated by the formation of a Gal4p–Gal80p–Gal3p complex. The EMBO Journal. 1998;17:4086–91.

111. Yano K, Fukasawa T. Galactose-dependent reversible interaction of Gal3p with Gal80p in the induction pathway of Gal4p-activated genes of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 1997;94:1721–6.

112. Peng G, Hopper JE. Evidence for Gal3p's Cytoplasmic Location and Gal80p's Dual Cytoplasmic-Nuclear Location Implicates New Mechanisms for Controlling Gal4p Activity in Saccharomyces cerevisiae. Mol Cell Biol. 2000;20:5140–8.

113. Peng G, Hopper JE. Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein. Proceedings of the National Academy of Sciences. 2002;99:8548–53.

114. Jiang F, Frey BR, Evans ML, Friel JC, Hopper JE. Gene Activation by Dissociation of an Inhibitor from a Transcriptional Activation Domain. Molecular and Cellular Biology. 2009;29:5604–10.

115. Egriboz O, Jiang F, Hopper JE. Rapid GAL Gene Switch of Saccharomyces cerevisiae Depends on Nuclear Gal3, Not Nucleocytoplasmic Trafficking of Gal3 and Gal80. Genetics. 2011;189:825–36.

116. Wightman R, Bell R, Reece RJ. Localization and Interaction of the Proteins Constituting the GAL Genetic Switch in Saccharomyces cerevisiae. Eukaryotic Cell. 2008;7:2061–8.

117. Egriboz O, Goswami S, Tao X, Dotts K, Schaeffer C, Pilauri V, et al. Self-Association of the Gal4 Inhibitor Protein Gal80 Is Impaired by Gal3: Evidence for a New Mechanism in the GAL Gene Switch. Molecular and Cellular Biology. 2013;33:3667–74.

118. Abramczyk D, Holden S, Page CJ, Reece RJ. Interplay of a Ligand Sensor and an Enzyme in Controlling Expression of the Saccharomyces cerevisiae GAL Genes. Eukaryotic Cell. 2012;11:334–42.

119. Johnston M, Flick JS, Pexton T. Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression in Saccharomyces cerevisiae. Mol Cell Biol. 1994;14:3834–41.

120. Acar M, Becskei A, van Oudenaarden A. Enhancement of cellular memory by reducing stochastic transitions. Nature. 2005;435:228–32.

121. Bhat PJ, Hopper JE. Overproduction of the GAL1 or GAL3 protein causes galactose-independent activation of the GAL4 protein: evidence for a new model of induction for the yeast GAL/MEL regulon. Mol Cell Biol. 1992;12:2701–7.

122. Kellis M, Birren BW, Lander ES. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature. 2004;428:617–24.

123. Platt A, Ross HC, Hankin S, Reece RJ. The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase. Proceedings of the National Academy of Sciences. 2000;97:3154–9.

124. Bhat PJ, Hopper JE. The mechanism of inducer formation in gal3 mutants of the yeast galactose system is independent of normal galactose metabolism and mitochondrial respiratory function. Genetics. 1991;128:233–9.

125. Bheda P, Aguilar-Gómez D, Becker NB, Becker J, Stavrou E, Kukhtevich I, et al. Single-Cell Tracing Dissects Regulation of Maintenance and Inheritance of Transcriptional Reinduction Memory. Molecular Cell. 2020;78:915-925.e7.

126. Ramsey SA, Smith JJ, Orrell D, Marelli M, Petersen TW, de Atauri P, et al. Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. Nat Genet. 2006;38:1082–7.

127. Stockwell SR, Rifkin SA. A living vector field reveals constraints on galactose network induction in yeast. Molecular Systems Biology. 2017;13:908.

128. Zacharioudakis I, Gligoris T, Tzamarias D. A Yeast Catabolic Enzyme Controls Transcriptional Memory. Current Biology. 2007;17:2041–6.

129. Biggar SR, Crabtree GR. Cell signaling can direct either binary or graded transcriptional responses. EMBO J. 2001;20:3167–76.

130. Hawkins KM, Smolke CD. The Regulatory Roles of the Galactose Permease and Kinase in the Induction Response of the GAL Network in Saccharomyces cerevisiae. Journal of Biological Chemistry. 2006;281:13485–92.

131. Stockwell SR, Landry CR, Rifkin SA. The yeast galactose network as a quantitative model for cellular memory. Molecular BioSystems. 2015;11:28–37.

132. Lohr D, Venkov P, Zlatanova J. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. The FASEB Journal. 1995;9:777–87.

133. Schwarz G. Estimating the Dimension of a Model. The Annals of Statistics. 1978;6:461–4.

134. Schuh L, Kukhtevich I, Bheda P, Schulz M, Bordukova M, Schneider R, et al. Altered expression response upon repeated gene repression in single yeast cells. PLoS Computational Biology. 2022;18:1–17.

135. Schuh L, Loos C, Pokrovsky D, Imhof A, Rupp RAW, Marr C. H4K20 Methylation Is Differently Regulated by Dilution and Demethylation in Proliferating and Cell-Cycle-Arrested Xenopus Embryos. cels. 2020;11:653-662.e8.

136. Alabert C, Loos C, Voelker-Albert M, Graziano S, Forné I, Reveron-Gomez N, et al. Domain Model Explains Propagation Dynamics and Stability of Histone H3K27 and H3K36 Methylation Landscapes. Cell Reports. 2020;30:1223-1234.e8.

137. Bast L, Buck MC, Hecker JS, Oostendorp RAJ, Götze KS, Marr C. Computational modeling of stem and progenitor cell kinetics identifies plausible hematopoietic lineage hierarchies. iScience. 2021;24.

138. Galdieri L, Mehrotra S, Yu S, Vancura A. Transcriptional Regulation in Yeast during Diauxic Shift and Stationary Phase. OMICS: A Journal of Integrative Biology. 2010;14:629–38.

139. Eustermann S, Schall K, Kostrewa D, Lakomek K, Strauss M, Moldt M, et al. Structural basis for ATP-dependent chromatin remodelling by the INO80 complex. Nature. 2018;556:386–90.

140. Tosi A, Haas C, Herzog F, Gilmozzi A, Berninghausen O, Ungewickell C, et al. Structure and Subunit Topology of the INO80 Chromatin Remodeler and Its Nucleosome Complex. Cell. 2013;154:1207–19.

141. Neuert G, Munsky B, Tan RZ, Teytelman L, Khammash M, Van Oudenaarden A. Systematic Identification of Signal-Activated Stochastic Gene Regulation. Science. 2013;339:584–7.

142. Tedford K, Kim S, Sa D, Stevens K, Tyers M. Regulation of the mating pheromone and invasive growth responses in yeast by two MAP kinase substrates. Curr Biol. 1997;7:228–38.

143. Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, et al. Regulated cell-to-cell variation in a cell-fate decision system. Nature. 2005;437:699–706.

144. Pesce CG, Zdraljevic S, Peria WJ, Bush A, Repetto MV, Rockwell D, et al. Single-cell profiling screen identifies microtubule-dependent reduction of variability in signaling. Molecular Systems Biology. 2018;14:e7390.

145. Elion EA, Satterberg B, Kranz JE. FUS3 phosphorylates multiple components of the mating signal transduction cascade: evidence for STE12 and FAR1. MBoC. 1993;4:495–510.

146. Matheos D, Metodiev M, Muller E, Stone D, Rose MD. Pheromone-induced polarization is dependent on the Fus3p MAPK acting through the formin Bni1p. Journal of Cell Biology. 2004;165:99–109.

147. Garrison TR, Zhang Y, Pausch M, Apanovitch D, Aebersold R, Dohlman HG. Feedback Phosphorylation of an RGS Protein by MAP Kinase in Yeast *. Journal of Biological Chemistry. 1999;274:36387–91.

148. Hohmann S. An integrated view on a eukaryotic osmoregulation system. Curr Genet. 2015;61:373–82.

149. O'Rourke SM, Herskowitz I. The Hog1 MAPK prevents cross talk between the HOG and pheromone response MAPK pathways in Saccharomyces cerevisiae. Genes Dev. 1998;12:2874–86.

150. Zahedi Y, Zeng S, Ekwall K. An essential role for the Ino80 chromatin remodeling complex in regulation of gene expression during cellular quiescence. Chromosome Res. 2023;31:14.

151. Yao W, Beckwith SL, Zheng T, Young T, Dinh VT, Ranjan A, et al. Assembly of the Arp5 (Actin-related Protein) Subunit Involved in Distinct INO80 Chromatin Remodeling Activities *. Journal of Biological Chemistry. 2015;290:25700–9.

152. Moreno DF, Acar M. Chromatin regulatory genes differentially interact in networks to facilitate distinct GAL1 activity and noise profiles. Current Genetics. 2021;67:267–81.

153. Habraken Y, Sung P, Prakash L, Prakash S. Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. Current Biology. 1996;6:1185–7.

154. Stone JE, Petes TD. Analysis of the Proteins Involved in the in Vivo Repair of Base–Base Mismatches and Four-Base Loops Formed During Meiotic Recombination in the Yeast Saccharomyces cerevisiae. Genetics. 2006;173:1223–39.

155. Sugawara N, Pâques F, Colaiácovo M, Haber JE. Role of Saccharomyces cerevisiae Msh2 and Msh3 repair proteins in double-strand break-induced recombination. Proceedings of the National Academy of Sciences. 1997;94:9214–9.

156. Freedman JA, Jinks-Robertson S. Effects of mismatch repair and Hpr1 on transcription-stimulated mitotic recombination in the yeast Saccharomyces cerevisiae. DNA Repair. 2004;3:1437–46.

157. Spell RM, Jinks-Robertson S. Role of Mismatch Repair in the Fidelity of RAD51- and RAD59-Dependent Recombination in Saccharomyces cerevisiae. Genetics. 2003;165:1733–44.

158. Zhang Y, Sun Z-W, Iratni R, Erdjument-Bromage H, Tempst P, Hampsey M, et al. SAP30, a Novel Protein Conserved between Human and Yeast, Is a Component of a Histone Deacetylase Complex. Molecular Cell. 1998;1:1021–31.

159. Weinberger L, Voichek Y, Tirosh I, Hornung G, Amit I, Barkai N. Expression Noise and Acetylation Profiles Distinguish HDAC Functions. Molecular Cell. 2012;47:193–202.

160. Liu P, Song R, Elison GL, Peng W, Acar M. Noise reduction as an emergent property of single-cell aging. Nature Communications. 2017;8:1–13.

161. Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, et al. COMPASS: A complex of proteins associated with a trithorax-related SET domain protein. Proceedings of the National Academy of Sciences. 2001;98:12902–7.

162. Peterson CL, Tamkun JW. The SWI-SNF complex: a chromatin remodeling machine? Trends in Biochemical Sciences. 1995;20:143–6.

163. Henry KW, Wyce A, Lo W-S, Duggan LJ, Emre NCT, Kao C-F, et al. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes Dev. 2003;17:2648–63.

164. Cheon Y, Kim H, Park K, Kim M, Lee D. Dynamic modules of the coactivator SAGA in eukaryotic transcription. Exp Mol Med. 2020;52:991–1003.

165. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9.

166. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research. 2022;50:D439–44.

167. Tong AHY, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, et al. Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. Science. 2001;294:2364–8.

168. Dénervaud N, Becker J, Delgado-Gonzalo R, Damay P, Rajkumar AS, Unser M, et al. A chemostat array enables the spatio-temporal analysis of the yeast proteome. Proceedings of the National Academy of Sciences of the United States of America. 2013;110:15842–7.

169. Padovani F, Mairhörmann B, Falter-Braun P, Lengefeld J, Schmoller KM. Segmentation, tracking and cell cycle analysis of live-cell imaging data with Cell-ACDC. BMC Biology. 2022;20:174.

170. Fröhlich F, Weindl D, Schälte Y, Pathirana D, Paszkowski Ł, Lines GT, et al. AMICI: high-performance sensitivity analysis for large ordinary differential equation models. Bioinformatics. 2021;37:3676–7.

171. Stapor P, Weindl D, Ballnus B, Hug S, Loos C, Fiedler A, et al. PESTO: Parameter EStimation TOolbox. Bioinformatics. 2018;34:705–7.

172. Hass H, Loos C, Raimúndez-Álvarez E, Timmer J, Hasenauer J, Kreutz C. Benchmark problems for dynamic modeling of intracellular processes. Bioinformatics. 2019;35:3073–82.