Stochastic gene expression, phenotypic variability and adaptation of budding yeast to environmental stresses

Jian Liu

Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)

Présentée et soutenue par :
LIU Jian
le vendredi 19 juin 2015

Titre :

Stochastic Gene Expression, Phenotypic Variability and Adaptation of Budding Yeast to Environmental Stresses

---

École doctorale et discipline ou spécialité :

ED SEVAB : Ingénieries microbienne et enzymatique

Unité de recherche :

Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, UMR INSA/CNRS 5504/INRA 792

Directeur(s) de Thèse :

Jean-Marie FRANCOIS et Jean-Pascal CAPP

Jury :

Bruno BLONDIN	Professeur des Universités, SupAgro, Montpellier	Président
Gaël YVERT	Directeur de recherche, ENS de Lyon	Rapporteurs
Gianni LITI	Chargé de recherche, IRCAN, Nice	Rapporteurs
Olivier GADAL	Directeur de recherche, LBME, Toulouse	Examinateur
Jean-Marie FRANCOIS	Professeur des Universités, INSA de Toulouse	Directeur
Jean-Pascal CAPP	Maître de Conférences, INSA de Toulouse	Directeur

Table of contents

Remercîment

Tous d'abord, je voudrais remercier mon directeur de thèse Jean-Marie FRANCOIS qui m'a donné l'opportunité de faire ma thèse au sein de la Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés. Egalement, je voudrais remercier mon co-directeur de thèse Jean-Pascal CAPP. Merci pour tes confiances de moi au cours de ma thèse. J'étais très touché par ta passion sur la science. Tu m'as enseigné beaucoup dans les domaines de la recherche. Tu as toujours favorisé mon autonomie et m'encouragé découvrir à des nouveaux aspects. Et surtout merci beaucoup pour tes gros travails sur mon papier et ma thèse.

Et je voudrais remercier Hélène MARTIN-YKEN, Sylvie DEQUIN et Olivier GADAL de participer mon comité de thèse. Les discussions pendant les deux réunions et vos conseils sur les expériences m'aident beaucoup à améliorer ce projet.

Ensuite, je voudrais remercier Adilya Timmers pour toutes les discussions et les conseils sur ce projet. C'est une grande chance de travail avec toi. Merci aussi pour tes résultats, je les ai piqués dans ma thèse.

Je voudrais remercier les personnes de l'équipe EAD5: Marie-Ange TEST, Jean-luc PARROU, Thomas WALTHER, Laurent BENBADIS et Gustavo de BILLERBECK pour toutes vos discussions et conseils sur les manips et les résultats. Egalement aussi remercie des collègues de l'équipe EAD5: Ran LIU, Ceren AIKIM, Debora TRICHEZ, Agustina LLANOS, Luce LOZANO-HUGUET, Marion SCHIAVONE, Marjorie PETITJEAN, Cléa LACHAUX, Yvan CAM. Le temps était très bien passé avec vous tous.

En plus, je voudrais remercier le Conseil des bourses chinoises (CSC) pour financer ma thèse en France.

Enfin, énorme merci à ma famille pour me soutenir au cours de ma thèse, vous me manquez beaucoup !

Summary

The increase in phenotypic variability through gene expression noise is proposed to be an evolutionary strategy in selective environments. Differences in promoter-mediated noise between *Saccharomyces cerevisiae* strains could have been selected for thanks to the benefit conferred by gene expression heterogeneity in the stressful conditions, for instance, those experienced by industrial strains. In the first part of this thesis, we used a genome-wide approach to identify promoters conferring high noise levels in the industrial wine strain EC1118. Many promoters of genes related to environmental factors were identified, some of them containing genetic variations compared to their counterpart in the laboratory strain S288c. Each variant of eight promoters has been fused to *yEGFP* and integrated in the genome of both strains. Some industrial variants conferred higher expression associated, as expected, to lower noise, but other variants either increased or decreased expression without modifying variability, so that they might exhibit different levels of transcriptional-mediated noise at equal mean. At different induction conditions giving similar expression for both variants of the *CUP1* promoter (*pCUP1*), we indeed observed higher noise with the industrial variant. Nevertheless, this difference was only observed in the industrial strain, revealing epistasis in the generation of promoter-mediated noise. Moreover, the increased expression variability conferred by this natural yeast promoter variant provided a clear benefit in the face of an environmental stress. Thus modulation of gene expression noise by a combination of promoter modifications and *trans*-influences might be a possible adaptation mechanism in yeast.

During the screening of high noise promoters, we noticed that some fragments conferred bimodal expression profiles. This expression pattern is generally considered as a bet-hedging strategy to generate distinct subpopulations in the population to cope with unpredictable environments. In the second part, we tried to find the genetic basis of these expression patterns. Unfortunately, genomic integration of the fragments generally eliminated this expression pattern so as expression from centromeric plasmid was the main factor leading to bimodal expression. Among the eight chosen promoters, only *pCUP1* in BY4720 (an auxotrophic derivative of S288c) exhibited bimodal expression at the genomic level when it was induced by copper. But the *YFP*-fusion at the original position of *CUP1* showed unimodal expression whether induced or not. Thus bimodal expression patterns are controlled at several levels and further studies are needed to understand the underling mechanisms.

The third part of this thesis was focused on the phenotypic effects of expression noise in DNA-

repair or maintenance genes. The main hypothesis underlying this work is that expression noise of these genes might produce sub-populations which exhibit higher genome instability and therefore can more rapidly adapt to new environments through genome modification. Thus high noise in these genes might increase the global fitness in varying environments. To test this hypothesis, *RAD52* and *RAD27* were fused to *YFP* and expressed under the dependence of promoters conferring similar mean expression but different noise levels. These genes were chosen because their deletion affects the homologous recombination (HR) frequency in opposite ways. Systems to measure the HR frequency and perform experiments in fluctuating environments have been constructed. Some promoter variants described in the literature were tested to find pairs conferring similar expression levels but different noise with the YFP-fused genes. But additional experiments are needed to get a system allowing testing the initial hypothesis.

Résumé

Les variations des niveaux d'expression des gènes entre les cellules individuelles existent dans les populations isogéniques même dans des environnements constants et peuvent avoir des conséquences phénotypiques profondes (Raser & O'Shea 2004; Raj & van Oudenaarden 2008; Balazsi *et al*., 2011). Ce phénomène qui s'appelé le bruit d'expression des gènes permettrait les cellules génétiquement identiques ayant des phénotypes hétérogènes et des comportements différents. Ainsi, il serait favorable à la prolifération des cellules pré-adapté lorsqu'une contrainte apparaît (Blake *et al*., 2006; Acar *et al*., 2008;. Fraser & Kaern, 2009; Ito *et al*., 2009; Ackermann, 2013).

Les gènes répondant aux changements environnementaux sont plus bruyants que ceux qui sont impliqués dans les processus d'entretien ménager (Bar-Même *et al.* 2006;. Newman *et al*., 2006). Les niveaux de bruit semblent avoir été sélectionnés selon les coûts et les avantages potentiels de cette variation. Le bruit est également minimisée pour des gènes essentielles (Fraser *et al*., 2004;. Lehner, 2008). Une étude chez *Escherichia coli* montre que les mutants avec le bruit d'expression plus élevé ainsi que les mutants avec le moyen d'expression plus élevé sont sélectionné sous une pression forte (Ito *et al*., 2009). L'augmentation de l'hétérogénéité phénotypique est aussi observée chez *Saccharomyces cerevisiae* dans les expériences évolutionnaires (New *et al*., 2014; Holland *et al*., 2014), qui a probablement eu lieu par le bruit de la transcription médiée par l'expression du gène.

Partie I Connexion le bruit d'expression et l'adaptation du stress

Comme le bruit d'expression est évolutif et héréditaire (Ito *et al.*, 2009) et le bruit des gènes liés au stress est plus élevés (Newman *et al.*, 2006), les souches industrielles pourraient avoir évolué vers des niveaux de bruit plus élevés pour les gènes impliqués l'adaptation des environnements stressants. Notre objectif dans cette partie était de déterminer si les différences génétiques observées entre la souche industrielle (EC1118) et la souche laboratoire (S288c) génèrent des différences de bruit entre elles et si ces différences sont suffisamment importantes pour générer une différence de croissance dans des environnements sélectifs spécifiques.

L'approche génomique utilisée pour identifier les promoteurs qui confèrent un haut niveau de stimulation du promoteur dans la souche EC1118 (Novo *et al.*, 2009) était basée sur le procédé développé initialement avec *Salmonella typhimurium* (Freed *et al.*, 2008). Des fragments d'ADN génomique de l'haploïde 59A (dérivé de la souche EC1118 diploïde) ont été insérés avant le yEGFP dans des vecteurs centromériques sans promoteur et ensuite transformée dans la souche de

laboratoire CEN.PK. La méthode de la sélection fluctuante décrite par Freed *et al.* (2008) permettant l'enrichissement des fragments produisant une expression de yEGFP plus variable a ensuite été appliquée. Sept cycles de tri ont été effectués alternativement avec les 5% plus fluorescent cellules ou les 5% moins fluorescent cellules. On a mesuré l'expression de yEGFP des cellules individuelles dans la banque après sélection. Les fragments conduisant l'expression de yEGFP plus bruyant ont été séquencés avec succès dans 97 clones. Cinquante promoteurs distincts ont été identifiés. On a choisi et étudié 8 paires de variantes de promoteur au niveau génomique dans BY4720 et 59A par rapport leurs fonctionnalités. Au contraire dans la littérature, la variante avec une moyenne plus forte n'a pas toujours été associée à un bruit plus faible. Néanmoins, cela dépend souvent du contexte génétique parce que certaines variantes donnent différentes moyennes en 59A mais pas dans BY4720. Ainsi, les éléments *cis*- (variations dans la séquence du promoteur) et les facteurs *trans* (facteurs cellulaires impliqués dans l'expression des gènes) sont associés pour permettre un profil d'expression spécifique, révélant l'épitasis dans la génération du bruit d'expression.

Nous avons confirmé la différence de bruit entre les variantes de *pCUP1* en les induisant avec différentes concentrations de sulfate de cuivre qui confèrent la même moyenne d'expression. Mais le bruit plus élevé est une caractéristique de la variante industrielle seulement révélée dans la souche 59A. Nous montrons ici l'influence combinée et corrélée des facteurs cis et *trans* qui contribuent à augmenter le bruit d'expression d'un gène. Par contre, les conséquences des SNPs et du Indel entre les variantes de *pCUP1* ne sont pas dramatiques et peuvent révéler que les systèmes naturels évoluent par des modifications de promoteur générant de petits effets sur le bruit mais n'influant pas fortement la variabilité entre les cellules.

En fusionnant les variantes naturelles de *pCUP1* avec ZeoR, nous montrons ici que leurs différences de bruit à une moyenne d'expression égale offrent des capacités distinctes pour survivre dans un environnement sélectif, même si les différences sont faibles. Mais ce bénéfice conféré par $pCUP1_{EC1118}$ applique seulement aux concentrations intermédiaires de phléomycine. L'identification d'un *pCUP1* plus fort dans une souche de vin industrielle est logique car les levures de vins sont fréquemment exposées aux concentrations fortes de cuivre pendants la fermentation. Certains gènes impliqués dans la réponse au stress, comme *CUP1*, ont été corrélés positivement avec la durée de la fermentation (Ambroset *et al*, 2011). Ainsi, un bruit d'expression plus élevé serait avantageux en ce sens qu'il rendrait la population optimale entre l'adaptabilité aux environnements riches en cuivre et la capacité de fermentation.

Partie II Etude du profil d'expression bimodal

Au cours du ciblage des clones avec un bruit d'expression plus élevé dans la banque après la sélection FACS, nous avons identifié des clones intéressants conférant des profils d'expression bimodaux, ce qui signifie qu'ils avaient deux pics dans la distribution alors que les autres clones n'avaient qu'un seul pic. L'un des avantages est qu'il confère deux sous-populations très différentes qui peuvent être adaptées à des situations contraintes. Mais les promoteurs correspondants fusionnés à GFP intégrés dans le locus *LEU2* ont tous présenté un profil d'expression uni-modal dans YPD. En échangeant le promoteur de *CAN1* et le fragment correspondant entre les plasmides intégratifs et centromériques, nous avons montré que les deux ont conféré un profile bimodal lorsqu'ils sont situés dans le plasmide centromérique alors qu'ils présentent un profile uni-modal après l'intégration génomique. Ainsi, le profil d'expression bimodal est une caractéristique de l'expression des plasmides centromériques. Ces effets sont supposés être dus aux différentes caractéristiques structurelles entre les plasmides et les chromosomes. Les chromosomes sont fortement structurés par des histones et d'autres protéines associées à la chromatine et leurs arrangements sont précisément contrôlé dans le noyau alors que les plasmides sont moins régulés par tels événements.

Le profil bimodal a également été observé avec des variantes de *CUP1* dans la souche BY4720 lorsqu'ils sont induits par $CuSO_4$, mais pas dans 59A. Ce phénomène pourrait être lié au différent nombre de copie de *CUP1* entre BY4720 et 59A. On a aussi étudié la cinétique de la bimodalité de *pCUP1.* Sous une concentration de cuivre précise, une proportion constante de la population est induite et le niveau d'induction de cette sous-population a d'abord augmenté puis diminué pendant le temps d'induction, alors qu'avec le même temps d'induction, des concentrations supérieures de $CuSO_4$ ont induit plus de cellules à des niveaux d'expression plus élevés. Nous proposons que la dynamique d'induction de *CUP2* puisse jouer un rôle important dans ces phénomènes.

Enfin, lorsque *CUP1* a été fusionné à YFP à son propre locus chromosomique dans 59A et BY4720, il a conféré un niveau d'expression plus élevé et un profil d'expression uni-modal avec ou sans induction par $CuSO_4$, ce qui indique que le locus chromosomique joue un rôle important dans la régulation de *CUP1* et ses effets sont indépendants de l'induction du cuivre. Le niveau d'expression au locus d'origine est beaucoup plus élevé qu'au locus *LEU2* indiquant que les régulateurs de transcription pourraient de préférence se lier au locus d'origine.

Partie III Connexion le bruit d'expression et l'instabilité du génome

Dans cette partie, nous avons essayé de tester notre hypothèse : le bruit d'expression plus élevé dans certains gènes de maintenance du génome peut donner une meilleure adaptabilité globale aux différents stress en augmentant le taux de génération de variantes génétiques (RGVG, Capp, 2010).

Ainsi, nous avons fusionné des gènes de réparation et de maintenance de l'ADN avec deux promoteurs modifiés de *PHO5*, qui confèrent des moyenne d'expression similaires mais différents niveaux de bruit (Raser & O'Shea, 2004,. $pPHO5_{TATA\text{-}A1T6}$ et $pPHO5_{UASm1}$). $pPHO5_{TATA\text{-}A1T6}$ a un niveau de bruit inférieur que $pPHO5_{UASm1}$. Ces constructions ont été transformées dans la souches JA0200 qui a *URA3* inséré dans les répétitions en tandem (TR) qui présentent dans le gène *FLO1* (Verstrepen *et al*, 2005). La recombinaison homologue (HR) entre les TRs peut supprimer le gène *URA3* qui permet la survie dans les boites contenant 5-FOA. Par conséquent, nous pouvons utiliser JA0200 pour mesurer la fréquence de HR.

En variant les conditions d'induction, nous avons réussi à obtenir des moyennes d'expression similaire, mais des bruits différents pour les deux variantes de *pPHO5* ($pPHO5_{UASm1}$ et $pPHO5_{TATA\text{-}A1T6}$) fusionnées avec *RAD52-YFP* ou *RAD27-YFP*. Nous avons testé si des bruits d'expression différents des gènes de maintenance du génome peuvent conduire à différents niveaux d'instabilité du génome en mesurant la fréquence de HR. Mais malheureusement, nous n'avons pas observé de différence significative. Étant donné que les cellules ne se sont poussées que 1,5 fois pendant la période d'induction dans le milieu sans phosphate, la fréquence était principalement contrôlée par la saltus initiale de la population. Par conséquent, la différence phénotypique potentielle générée par les variantes de *pPHO5* induites n'était pas détectable.

Introduction

1 Introduction

1.1 Variability of gene expression

1.1.1 Expression of protein-coding genes

How genetic information is stored in the DNA was elucidated after the discovery of its structure (Watson & Crick, 1953). Gene expression is the process by which information from DNA is used in the synthesis of a functional gene product. The products could be proteins or functional RNAs (such as ribosomal RNA (rRNA) or transfer RNA (tRNA)).

For protein-coding genes, which are the purpose of this dissertation, gene expression includes two main steps: transcription and translation (Figure 1). Transcription is the process by which genetic information is transferred into RNA. This process produces messager RNAs (mRNA) used to produce proteins by ribosomes, in the process called translation.

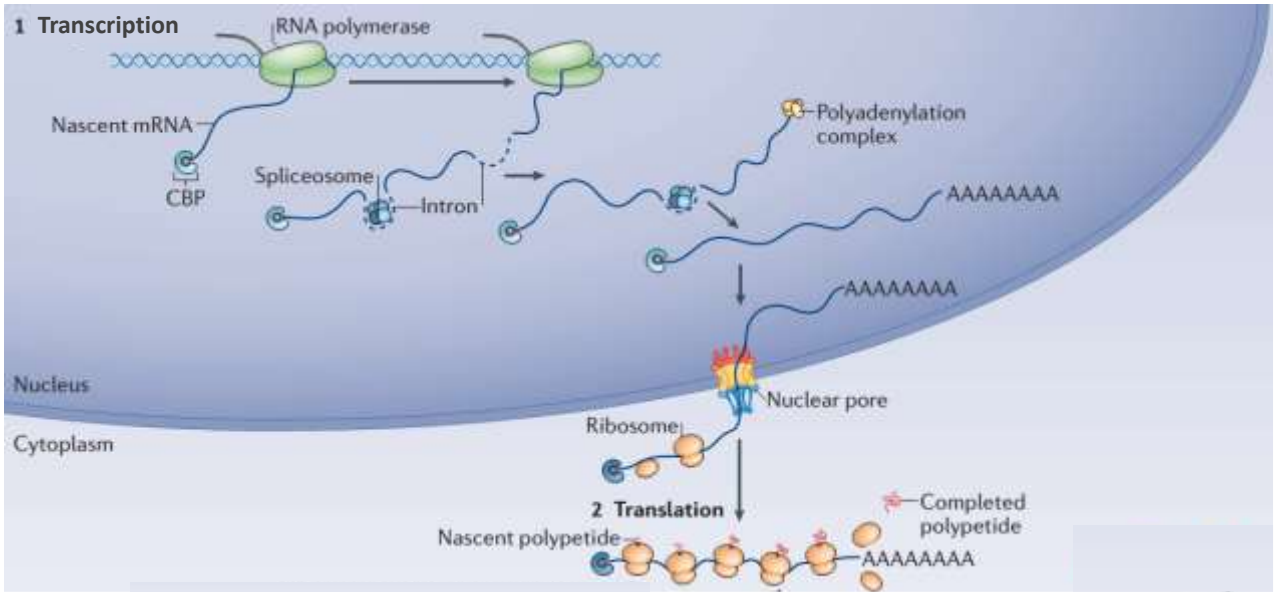


Figure 1: Expression of protein-coding genes (modified from Kervestin & Jacobson (2012))**.**

1.1.1.1 Transcription

The transcription process is performed by RNA polymerases (Hurwitz, 2005). RNA polymerases use DNA as a template to synthesize single strand RNA from 5' to 3' end. In eukaryotic cells, there are three RNA polymerases (I, II, III). Here we will focus on RNA polymerase II (RNAPII) which is involved in the transcription of protein-coding genes (Roeder & Rutter, 1969).

The RNAPII-dependent transcription process generally includes 4 steps: pre-initiation, initiation,

elongation and termination (Hirose & Manley, 2000). The whole process is finely controlled in the cell by transcription factors and *cis*-regulatory elements. Transcription factors (*trans*-regulatory factors) are proteins that can recognize and interact with specific sequences in gene promoters (through *cis*-elements) to regulate transcription. They can be General Transcription Factors (GTF), such as the ones involved in all RNAPII-dependent transcription initiation events (TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH etc.) or transcription regulators (activator or repressor). Co-regulators can also be involved to connect transcription factors to RNAPII and to modify the chromatin structure that plays an important role in the process of transcription (Shandilya & Roberts, 2012). Before describing in details the different *trans*-acting factors, *cis*-elements important in transcription are described.

1.1.1.1.1 Cis-elements

A gene is usually defined as the DNA region which is transcribed from the transcription starting site (TSS) to the terminator into a pre-RNA. The region upstream of a gene is the promoter: it is used to initiate the transcription, while the terminator is the region where the transcription is terminated. In yeast, the *cis*-regulatory elements are usually located in the promoter region.

The length of promoters varies from 100 to 1000bp. It contains the indispensable elements for transcription initiation (core promoter) and some regulatory elements. The core promoter region contains sequences around the TSS that are recognized by GTF, while regulatory elements are gene specific and vary in terms of number and position. Sometimes, especially in higher eukaryotes, regulatory elements, such as enhancers, can be several kilobases away from the TSS.

i) Core promoter

The structure of the core promoter is shown in Figure 2 (Butler & Kadonaga, 2002). The TSS is numbered as the +1 position, and the position of the *cis*-elements are shown by the distance between the element and the TSS. Around the TSS, a motif called initiator (Inr) is present in metazoans (e.g. human, *Drosophila*) (Smale & Baltimore, 1989; Juven-Gershon *et al*, 2008), whereas Inr-like sequences are found in some yeast core promoters (Yang *et al*. (2007)). Recently, Lubliner *et al.* (2013) found that the sequences around TSS are predictive of the maximal promoter activity in yeast. There are T-rich regions and A-rich regions located upstream and downstream of the TSS, respectively, that are indicative of high promoter activity.

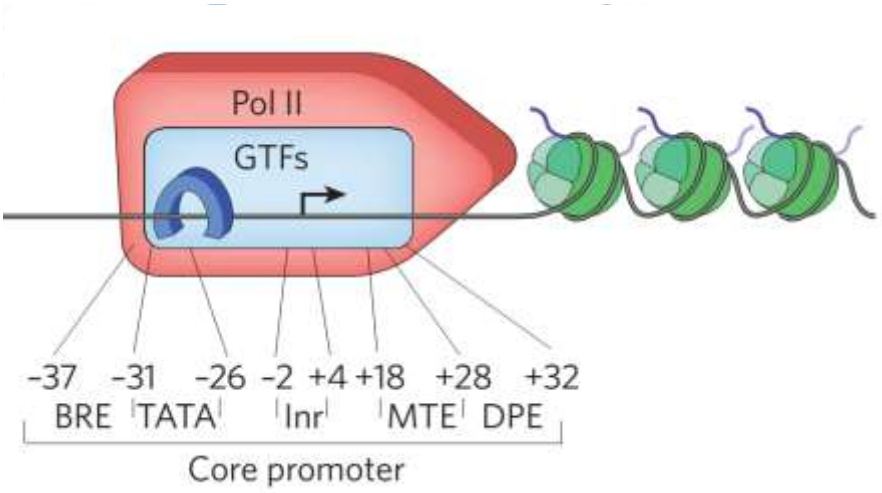


Figure 2: Structure of the core promoter of a RNAPII transcribed gene (modified from Fuda *et al*. (2009))

Upstream of the Inr, the first identified eukaryotic core promoter motif can be observed: the TATA box (Breathnach & Chambon, 1981). The consensus sequence for TATA box is TATAWAWR. TATA box is not a very common motif. In yeast, only about 20% of genes contain a TATA box (Shandilya & Roberts, 2012) and its position varies from about -40 to -100 (Butler & Kadonaga, 2002). But if sequence elements with up to two mismatches with the TATA box consensus are searched, almost all yeast promoters contain "TATA-like" sequences (Rhee & Pugh, 2012). TATA box-containing genes are associated with stress-response and are stringently regulated whereas TATA-less promoters are generally engaged in basic housekeeping functions (Basehoar *et al*, 2004).

The TFIIB recognition element (BRE) can be upstream or downstream of the TATA box (Deng & Roberts, 2007). Depending on the promoter context, it can act in either a positive or negative manner. The downstream promoter element (DPE) is a downstream TFIID recognition sequence first found in *Drosophila* (Burke & Kadonaga, 1996). In yeast, BRE and DPE are very rare (Yang *et al*, 2007).

ii) Regulatory elements

The core promoter is found upstream of all protein-coding genes, while regulatory elements are gene-specific, such as enhancers that are *cis*-acting regulatory sequences that markedly increase expression level and can be either upstream or downstream of the gene, and silencers, *cis*-acting regulatory sequences that decrease expression level (Gaszner & Felsenfeld, 2006).

In yeast, a well-studied *cis*-acting regulatory sequence is the upstream activating sequence (UAS) which is located upstream of promoters and increases expression level (Gaszner & Felsenfeld, 2006). Many different UAS exist such as the Gal4 binding site (Brand & Perrimon, 1993) and the Gcn4 binding

site (Hope & Struhl, 1987) and several UAS can be found in the same promoter. Transcriptional activators are able to bind UAS and help recruitment of the core transcriptional machinery. Detailed examples will be given in the next section.

iii) Nucleosome-influencing sequences

Genomic DNA is wrapped into nucleosomes in the nucleus. These nucleosomes are an obstacle for the recruiting of the Pre-Initiation Complex (PIC). *In vivo*, nucleosome positioning takes place at preferred sequences which have periodic AA/TT/AT dinucleotide sequences every ~10 bp (Segal *et al*, 2006), while poly(dT:dA) acts as nucleosome disfavoring sequence (Sharon *et al*, 2014). When these sequences are present on promoters, they affect gene expression through their influence on nucleosome positioning (Carey *et al*, 2013; Choi & Kim, 2009; Sharon *et al*, 2014).

1.1.1.1.2 Trans-acting factors

Eukaryotic transcription is a precisely timed event and, at every step, there are different *trans*-acting factors used to regulate the process (Figure 3). It starts with the recruitment of GTFs and RNAPII to the target gene promoters. They form the PIC. Then the transcription is initiated and elongated along the gene. Finally, RNAPII reaches the termination sites and is released (Thomas & Chiang, 2008; Shandilya & Roberts, 2012).

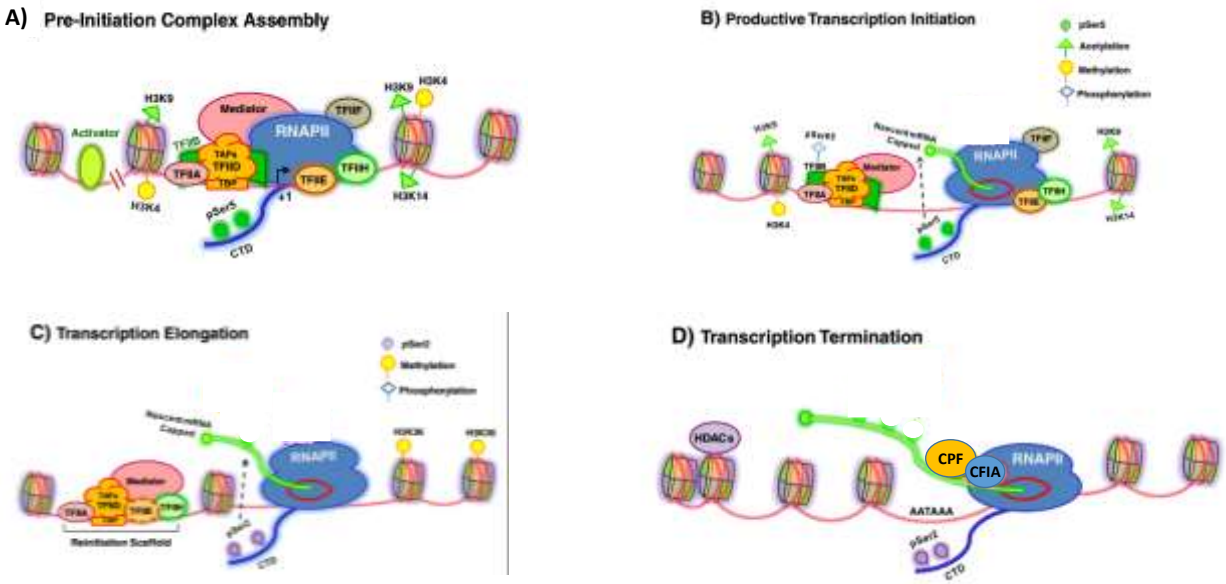


Figure 3: RNAPII transcription cycle (modified from Shandilya & Roberts, 2012)

i) *Pre-initiation*

a) General transcription factors

The recognition of the TATA box sequence by the TATA Binding Protein (TBP) is the beginning of the assembly of the PIC. TBP along with the TBP-associated factors (TAFs) forms the TFIID complex. The TAFs can recognize the Inr and DPE sites. Thus TFIID is a core promoter-binding factor (Thomas & Chiang, 2008). Meanwhile, the TFIIB binding to the BRE can stabilize the interaction between TFIID and the core promoter. Then the RNAPII-TFIIF complex binding to the TSS, followed by TFIIE, and TFIIH, forming the PIC (Butler & Kadonaga, 2002). The recruitment of TBP is a critical step for the PIC formation and can be regulated both positively and negatively. In yeast, transcription activators such as Hsf1 and Msn2 can stimulate the interaction between TBP and the TATA-box (Kuras & Struhl, 1999) while the repressors such as Mot1 and Taf1 suppress this binding (Cang *et al*, 1999). Another TAF containing co-activator is the SAGA (Spt-Ada-Gcn5 acetyltransferase) complex that can contact with the TBP and activate transcription. About half of the TATA-containing promoters, which are stress-inducible, appear to be SAGA-dependent. The TBP on these promoters preferably contacts SAGA rather than forms the TFIID complex (Basehoar *et al*, 2004). Zanton et Pugh (2004) also showed that, during heat stress, yeast can induce transcription of several genes through the recruitment of both TFIID and SAGA.

b) Sequence specific regulators

Sequence specific regulators and components affect the transcription efficiency by promoting or preventing the formation of the PIC (Venters & Pugh, 2009; Venters *et al*, 2011). The signals from these factors are transmitted to the general transcription machinery by Mediator. Mediator can enhance RNAPII entry to the PIC as well as the TBP binding to the TATA-box. Mediator can also facilitate the reinitiation of transcription from the same promoter (Thomas & Chiang, 2008).

Two examples of sequence specific regulators can be cited here: Gal4 and Pho4. Gal4 has two domains: the DNA-binding domain (DB) and the activation domain (AD). The BD recognizes the UAS sequence, and the AD binds to mediator and activates the transcription of Gal1 by facilitating the formation of PIC (Figure 4, Ang *et al*, 2012).

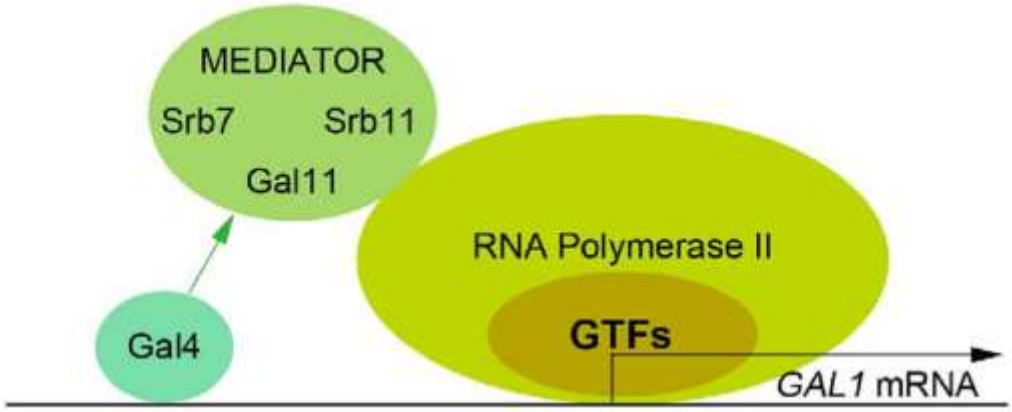


Figure 4: The interaction between Gal4 and UAS activate GAL1 transcription (Ang *et al*, 2012)

Pho4 activates the expression of *PHO5* (Nakao *et al*, 1986; Venter *et al*, 1994; Barbaric *et al*, 1998). There are two Pho4 binding sites on the *PHO5* promoter: UASp1 and UASp2. During phosphate (Pi) starvation, Pho4 is accumulated in the nucleus and binds to the first site UASp1. This binding leads to the hyperacetylation of the nucleosomes located on the promoter. The hyperacetylated nucleosomes are then evicted. Thus the promoter becomes open and the UASp2 can be bound by Pho4 too. With the help of other factors, the binding of Pho4 to the UAS sites finally activates the transcription of *PHO5* (Figure 5).

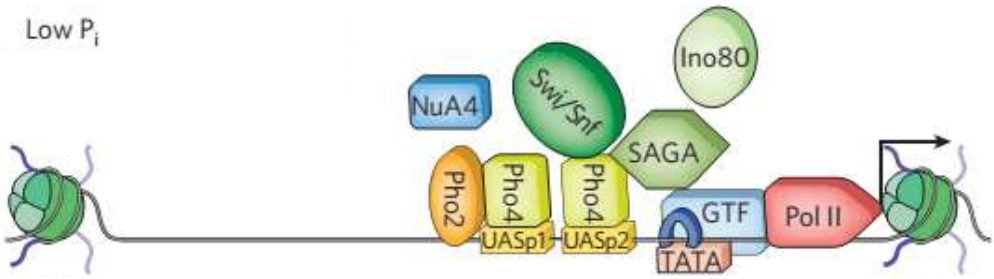


Figure 5: Activation of the *PHO5* promoter (modified from Fuda *et al*, 2009)

ii) *Initiation and elongation*

The formation of a functional and stable PIC is indispensable for transcription initiation. The DNA–GTF interactions in the PIC link RNAPII to the promoter, whereas TFIIH promotes ATP-dependent unwinding of approximately 10 base pairs in the promoter, leading to the open complex state (Grünberg & Hahn, 2013). This progress is favored by the presence of a TATA box because of its high AT

proportion. Otherwise, the negative supercoiling introduced by histones can also facilitate the unwinding step (Thomas & Chiang, 2008). RNAPII then starts mRNA synthesis from the TSS. Transcripts of less than 10nt are unstable, resulting in a high frequency of abortive initiation because the RNA molecule is too short to form a stable DNA:RNA duplex. At about 10 nt, this duplex is stabilized and promoter escape is favored over abortive initiation. At transcript length of around 25 nt, productive initiation is achieved and transcription elongation begins (Saunders *et al*, 2006).

Then RNAPII is dissociated from the initiation complex and starts moving along the gene. The synthesis occurs at a varying speed, especially because secondary structures in the mRNA can slow down or even stop RNAPII. The duration of pausing depends on the rate of recruitment of factors that trigger pause release. Universally conserved transcription elongation factors are the Spt5/NusG factors, which reduce the frequency of transcription pausing or arrest in cooperation with Spt4. TFIIS is also beneficial to the elongation: it can help RNAPII to bypass specific blocks (Wind & Reines, 2000). Another elongation factor is elongin: it activates elongation by suppressing transient pausing of RNAPII (Aso *et al*, 1995). Otherwise, the maintenance of the RNA:DNA hybrid is also a critical determinant of elongation (Shandilya & Roberts, 2012).

The mRNA maturation events such as 5'-capping and intron splicing are performed during transcription elongation. The 5'-cap is a guanine nucleotide connected to the mRNA via an unusual 5' to 5'-triphosphate linkage. This guanosine is methylated on the 7 position (m7G) (Shatkin, 1976). The cap is formed shortly after transcription initiation, when nascent RNA chains are about 25–30 nucleotides in length (Coppola *et al*, 1983). The formation of the cap needs the RNA methyltransferase, RNA guanylyltransferase and RNA triphosphatase. The RNAPII subunit IIO and its Carboxy-Terminal Domain (CTD) are involved in this process too (Hirose & Manley, 2000). The most important role of 5'-cap is protecting mRNA from degradation, but is also involved in all the transcription, translation, even degradation processes (Cowing, 2010).

Moreover when introns are present in genes, they are spliced to form the mature mRNA. In *S. cerevisiae*, only 5% of genes have introns and more than 95% of them contain only one single intron (Nash *et al*, 2007). Splicing of mRNA precursors takes place in a large macromolecular complex called spliceosome, which is composed of small nuclear ribonucleoprotein particles (snRNPs) and non-snRNP proteins (Hirose & Manley, 2000). The CTD of RNAPII and the IIO subunit also play an important role during splicing (Hirose *et al*, 1999). Thus the transcription and splicing processes are coupled (Hirose & Manley, 2000). An interesting aspect is the alternative splicing process which can produce variable mature mRNAs (Kalsotra & Cooper, 2011) but it is very rare in *S. cerevisiae*. For example, in *S.*

cerevisiae, *PTC7* can be alternatively spliced to create two mRNAs that code for distinct proteins. The protein translated from the spliced mRNA localizes to the mitochondria, while the protein translated from the unspliced mRNA localizes to the nuclear envelope (Juneau *et al*, 2009).

iii) Termination

Termination is the last step of transcription, but may also serve as a junction for recycling RNAPII at the promoter for subsequent rounds of transcription (Hirose & Manley, 2000). There are two well-studied pathways of transcription termination, the poly (A)-dependent pathway and the Nrd1–Nab3–Sen1-dependent pathway (Shandilya & Roberts, 2012). The Nrd1–Nab3–Sen1-dependent pathway is found in the termination of non-coding RNA transcription by RNAPII in yeast (Vasiljeva & Buratowski, 2006). For protein-coding genes, the poly (A)-dependent pathway is used to terminate transcription.

In eukaryotes, protein-coding genes have a highly conserved poly(A) signal (AATAAA) which is followed by a GC-rich sequence. This signal leads to the polyadenylation of precursors and the release of RNAPII. There are several protein complexes that involved in transcription termination. In yeast, the most important complexes is the cleavage and polyadenylation factor (CPF) and cleavage factor IA (CFIA) (Kuehner *et al*, 2011). They cut the pre-RNA at the signal site what is followed by poly(A) addition to the 3'end of the upstream cleavage product by the poly(A) polymerase (Shandilya & Roberts, 2012), and by the RNAPII release.

As for capping and splicing, the CTD of RNAPII and its subunit IIO play an important role during polyadenylation (Shatkin & Manley, 2000). The 3'-tail can stabilize mRNA and facilitate translation. But recent researches have shown that lots of genes contain more than one polyadenylation site, a phenomenon called alternative polyadenylation (APA). APA contributes to the complexity of the transcriptome (Elkon *et al*, 2013). Ozsolak *et al*. (2010) found that almost seventy percent of the yeast genes had more than one polyadenylation site.

1.1.1.1.3 Chromatin architecture

The organization of eukaryotic DNA into the chromatin is a significant obstacle for the binding of transcription factors to their cognate sequences and hence in most scenarios negatively influences all steps of transcription (Shandilya & Roberts, 2012). The basic unit of chromatin is the nucleosome (

Figure *6*). In a nucleosome, the DNA is wrapped twice around a core histone octamer (a structure that consists of two copies each of the core histones, H2A, H2B, H3, and H4) in a left-handed super helix. Then a linker histone H1 binds to the nucleosome to stabilize it. The diameter of nucleosome is

around 10 nm. The 10 nm fibre can be further compacted to form the 30 nm fibre (secondary structure,

Figure *6*). The 30 nm structure can be finally compacted to form higher-order chromatin fibre as shown in

Figure *6* (Luger *et al*, 2012).

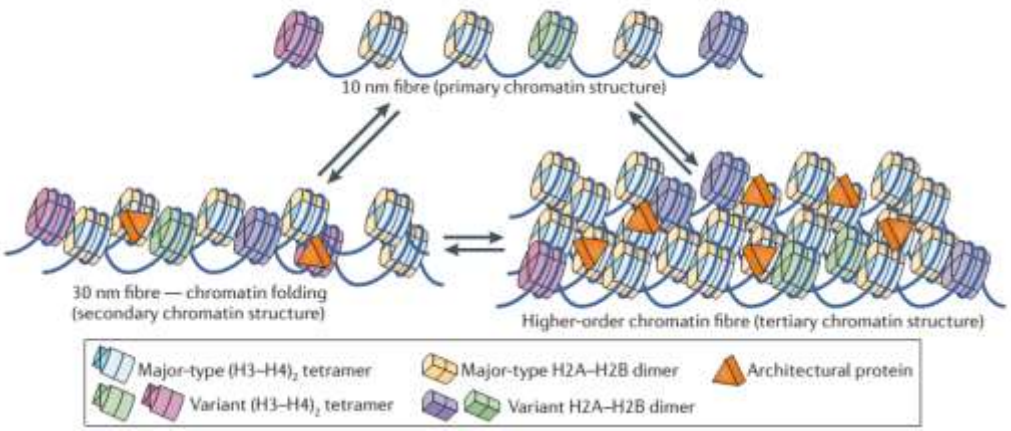


Figure 6: Chromatin structure (Luger *et al*, 2012)

i) *Nucleosome positioning*

Modifications of the nucleosome positioning correlate with the conversion of genes from a repressed state to a transcriptionally competent state (Paranjape *et al*, 1994). During the induction of *PHO5*, nucleosome repositioning is indispensable. Mutations in the *PHO5* promoter which promote expression under non-permissive conditions lead to shifts of positioned nucleosomes similar to the ones observed during the induction of the *PHO5* promoter. By contrast, mutations that reduce gene expression upon induction stabilize nucleosomes on the promoter region (Small *et al*, 2014).

ii) Histone modifications

Every histone has several acetylation and methylation sites (Paranjape *et al*, 1994). The reversible modifications of these sites have important effect on the regulation of transcription. The most common example in yeast is the gene silencing of the silent mating-type cassettes and telomeres. The silenced gene have hypoacetylated histones on the chromatin, compared to active genes (Braunstein *et al*, 1993). More generally, histone acetylation creates an open chromatin that favors transcription whereas deacetylation is correlated to compacted and repressive chromatin. The effects of histone methylation on the chromatin compaction level depend on the modified histone and amino acid.

Histone modifications during the transcription process play an important role. Gene-specific activators mediate the recruitment of histone modifying enzymes. Indeed, TFs recruit coactivators that can remove nucleosomes from the promoter or modify histones. Then these modified histones constitute an open and permissive chromatin environment competent for transcription (Ansari, 2009). For example, the binding of activated Pho4 (unphosphorylated) to the UASp1 site on the *PHO5* promoter, when yeast cells encounter phosphate (Pi) starvation, triggers histone hyperacetylation on the promoter which leads to nucleosome disassociation. Thus TFs can more easily bind the promoter and finally activate *PHO5* transcription (Fuda *et al*, 2009).

iii) *Gene and chromosome looping*

Gene looping is a phenomenon by which the terminal and promoter regions of active genes interact by chromatin bending (Figure 7). The formation of a gene loop needs CPSF (cleavage and polyadenylation specific factor) and CstF (cleavage stimulatory factor), the initiation factors TFIIB and TFIIH, as well as the 3'-end processing complex (Singh & Hampsey, 2007). Gene loops have been found in yeast, in HIV and human cancer cells. Gene loops facilitate transcriptional memory, both for activation and repression (Lainé *et al*, 2009). Gene loops also enforces transcriptional directionality on bidirectional promoters to produce mRNA rather than noncoding RNA (Tan-Wong *et al*, 2012).

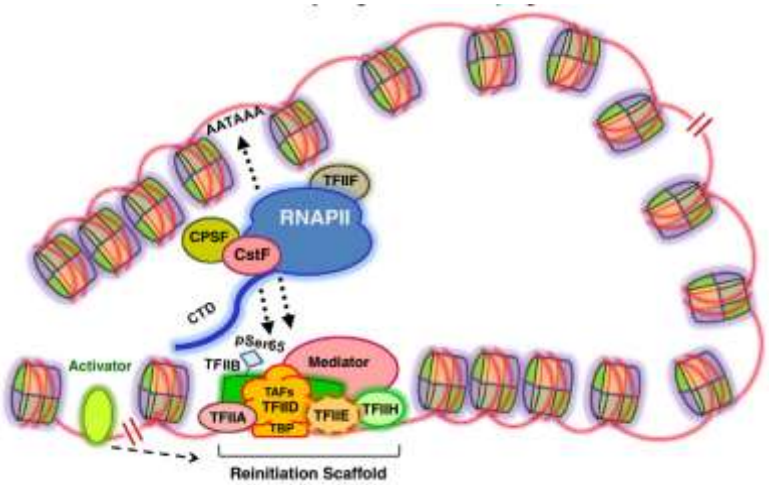


Figure 7: RNAPII recycling via gene looping (Shandilya & Roberts, 2012)

Active alleles of genes that are separated by several megabases, or even located on different chromosomes, are found to co-localize in the nucleus under certain conditions (Deng *et al*, 2013). These transcriptional foci are called "transcription factories", which are discrete entities composed of active RNAPII and several transcription associated factors. It was found that multiple genes might share

the same transcription factory (Shandilya & Roberts, 2012).

1.1.1.1.4 Transcriptional diversity

In yeast, non-coding RNAs (ncRNAs) which also have a 5'-cap and a 3'-poly(A) tail are produced: they are called cryptic unstable transcripts (CUTs) and the stable uncharacterized transcripts (SUTs). They are transcripts, which do not overlap with existing annotations, found in RNA-arrays or RNA-seq experiments. These ncRNAs mainly come from the anti-sense transcription of bidirectional promoters (Xu *et al*, 2009). It was already known that pervasive transcription exists in yeast: 85% of the genome can be transcribed (David *et al*, 2006). Pelechano *et al*. (2013) have characterized of the extensive diversity in the transcriptional output from the yeast genome. Through transcript isoform sequencing (TIF-seq), they have found large isoform diversity in yeast such as transcripts covering 2 open reading frames (ORFs) (bicistronic transcripts), overlapping 2 ORFs, overlapping 3' or 5' of one ORF, untranslated region (UTR) ends variations, etc.

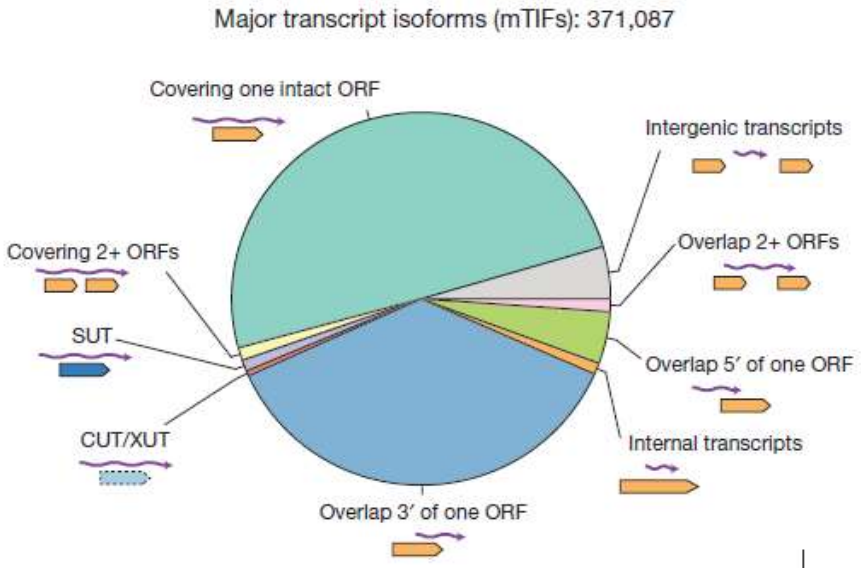


Figure 8: Major transcript isoforms in yeast (Pelechano *et al*, 2013)

Over 26 major transcript isoforms per protein-coding gene are expressed in yeast and 70% of genes express alternative isoforms. These variations in transcripts production can affect mRNA stability, localization and translation, or produce truncated proteins that differ in localization or function. N-ter truncation can be due to skipping of the first start codon and C-ter truncation due to early polyadenylation creating new stop codon.

Moreover the mean RNA molecule number per gene in yeast is one copie per cell (Miura *et al*,

2008). Thus, if 26 isoforms are possible and if only one copie is present per cell in mean, one can expect a high cell-to-cell variability in mRNA structure. Finally, isoform diversity generates diversity in RNA stability because of variations in binding of RNA binding proteins (Gupta *et al*, 2014).

1.1.1.1.5 mRNA degradation

Indeed the stability of a particular mRNA is controlled by specific interactions between its structural elements and RNA-binding proteins that can be general or mRNA-specific. The sequence elements important for stability are throughout the whole mRNA (Tourrière *et al*, 2002). The most common elements are the 5'-cap and the 3'-poly(A) tail. The process of 3'-poly(A) tails shortening which is performed by deadenylases can trigger mRNA decay in eukaryotic cells. After deadenylation, the mRNAs can be degraded by 3' → 5' exonucleases or through the decapping pathway (Chen & Shyu, 2010). The cap structure can be removed by a complex consisting of two proteins, Dcp1 and Dcp2 (Dunckley & Parker, 1999), then the mRNA body will be degraded by 5' → 3' exonucleases. Moreover, eukaryotic mRNAs that have abnormalities in translation will be degraded through specific pathways such as Nonsense-Mediated Decay (NMD) for mRNAs with premature translation termination codons, Nonstop Decay (NSD) for mRNAs lacking translation termination codons and No-Go decay (NGD) for mRNAs having strong pauses in elongation (Doma & Parker, 2007).

1.1.1.2 Translation and post-translational processes

In eukaryotes, mature mRNAs are exported from the nucleus to the cytoplasm, and then ribosomes are assembled on mRNAs to translate them into proteins. Like for transcription, there are lots of translation factors used to ensure productive translation.

1.1.1.2.1 Translation factors

i) Ribosome

The ribosome consists of two subunits in all species. In eukayotes, the subunits are designated as 40S and 60S, and together make up the 80S ribosome. In *S. cerevisiae*, the small 40S subunit contains a 18s rRNA and 33 ribosomal proteins whereas the large 60S subunit contains the 5.8S, 25S and 5S rRNAs and 46 ribosomal proteins (Melnikov *et al*, 2012). There are three binding sites for tRNA, designated the A (aminoacyl) site, which accepts the incoming aminoacylated tRNA, the P (peptidyl) site, which holds the tRNA with the nascent peptide chain, and E (exit) site, which holds the deacylated tRNA before it leaves the ribosome. The small subunit binds the mRNA and the anticodon stem-loops of tRNAs, and contributes to the fidelity of translation by monitoring base pairing between codon and

anticodon in the decoding process. The large subunit binds the acceptor arms of tRNA and catalyzes peptide bond formation between the incoming amino acid and the nascent peptide chain (Ramakrishnan, 2002).

ii) Initiation

In eukaryote, the 40S subunit carrying the eIF2-GTP-Met-tRNA$_i$ ternary complex first interacts with the 5'-cap of a mRNA (Figure 9), and then goes through the 5'-Untranslated Region (5'UTR) until the start codon where it is assembled with the 60S subunit. Many eukaryotic initiation factors are involved translation initiation and can simulate translation (Figure 9).

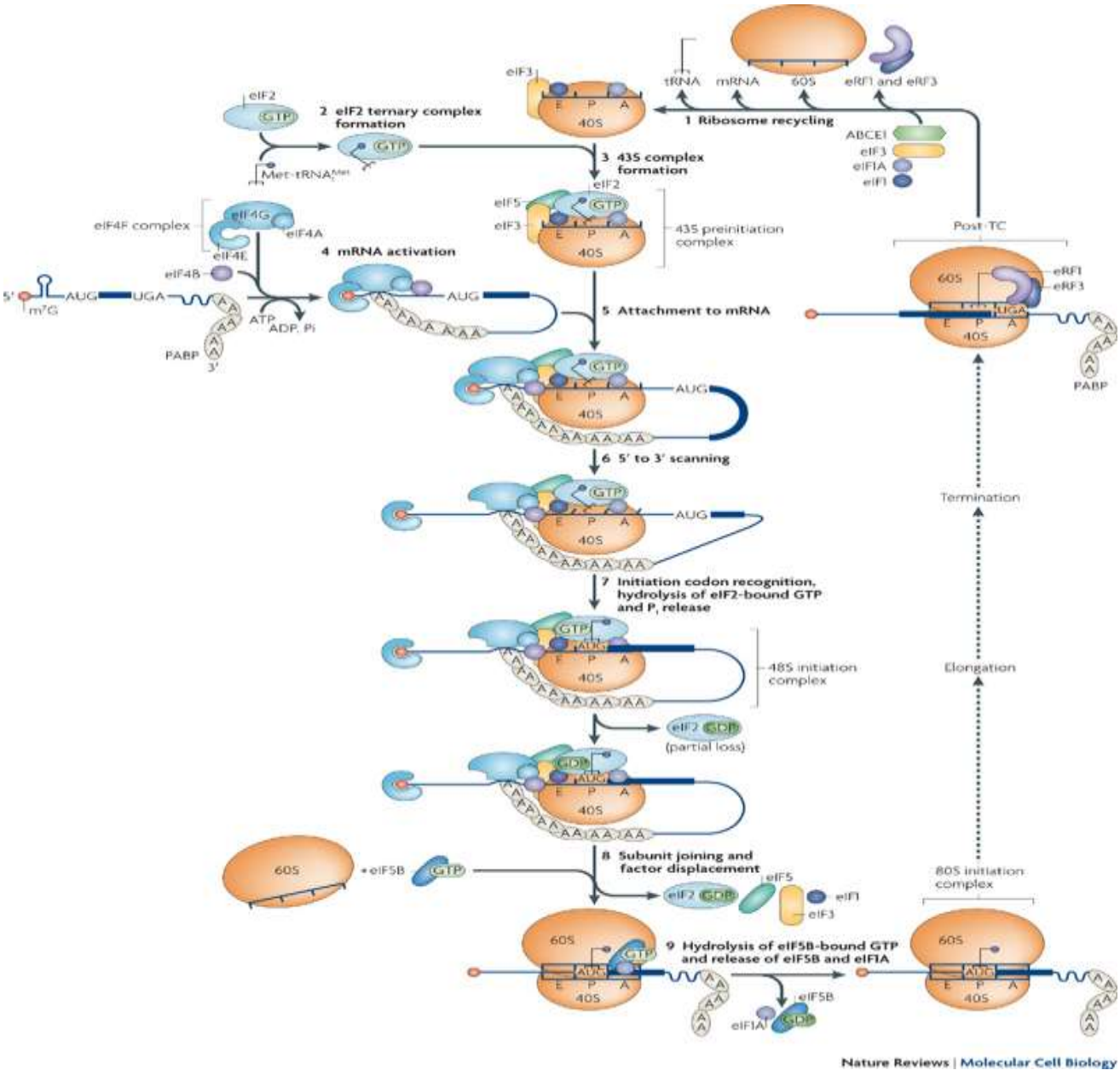


Figure 9: Translation initiation cycle (Jackson *et al*, 2010)

iii) Elongation

The elongation cycle is shown in Figure 10 (Schneider-Poetsch *et al*, 2010). An aminoacylated tRNA bound to GTP and the eukaryotic elongation factor 1A (eEF1A) enters the A site. If its anticodon matches the next codon in the mRNA (with full complementarity or only partial complementarity thanks to base wobble), the ribosome catalyzes the peptide bond formation. After ribosome translocation, the released deacetylated tRNA enters the E site before ejection, and the tRNA with the nascent amino acid chain is in the P site awaiting the next cycle.

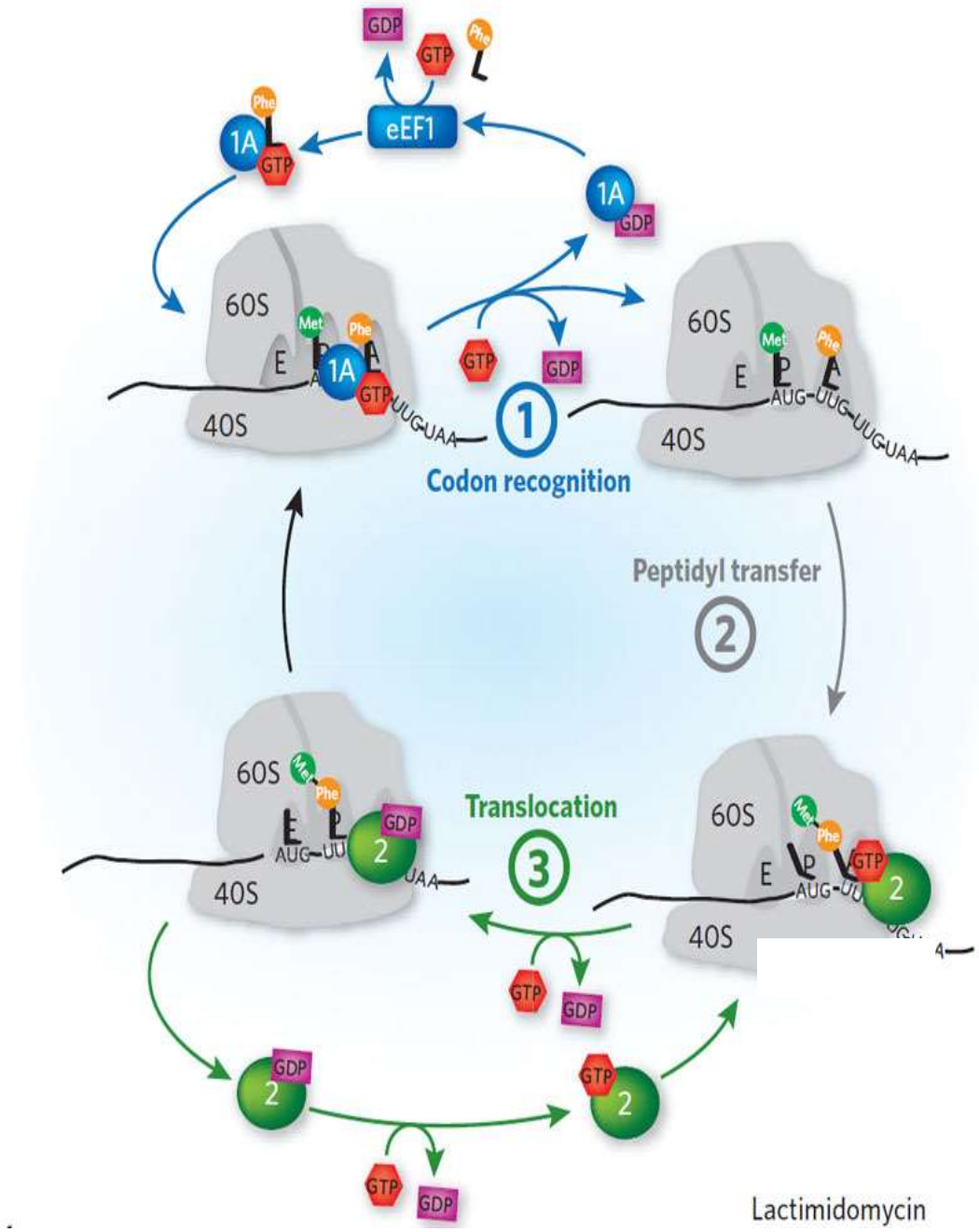


Figure 10: Translation elongation cycle (Schneider-Poetsch *et al*, 2010)

iv) Termination

The presence of a stop codon in the A site stimulates the termination process. Because no tRNA is able to recognize the stop codons, the eRF1-eRF3-GTP ternary complex can enter the A site and recognize the stop codon. After GTP hydrolysis, eRF3 is released (Dever & Green, 2012). Then the binding of ABCE1/Rli1 to eRF1 facilitate hydrolysis of the peptidyl-tRNA linkage. Thus the polypeptide and the mRNA are released and the ribosome disassembles through several steps (Kapp & Lorsch,

2004).

1.1.1.2.2 Post-translational process

i) Protein folding

Polypeptides resulting from translation have to fold into a proper 3-dimensional structure which is indispensable for their functions. In the cell, there are alternative pathways to help and control polypeptides folding (Wickner, 1999). Chaperones are proteins that catalyze protein folding. By binding exposed hydrophobic patches on proteins, chaperones prevent proteins from aggregating into insoluble, nonfunctional aggregates and help them to reach their stable native state (Ben-Zvi & Goloubinoff, 2001). Some proteins are co-translationally folded whereas others are released in an unfolded form. Then they spontaneously fold or are submitted to chaperone-mediated folding. The nascent chains first bind to the nascent-chain-associated complex (NAC), Hsp70 proteins and Hsp40 proteins. Then with the help of nucleotide exchange factors (NEFs), these chaperons mediate co- and posttranslational folding. Hsp70s can also transmit their substrate to Hsp90 proteins to mediate folding with additional cofactors. Partially folded substrates are transferred to the tailless complex polypeptide-1 (TCP-1) ring complex (TRiC)/chaperonin-containing TCP-1 (CCT) to be further folded (Figure 11).

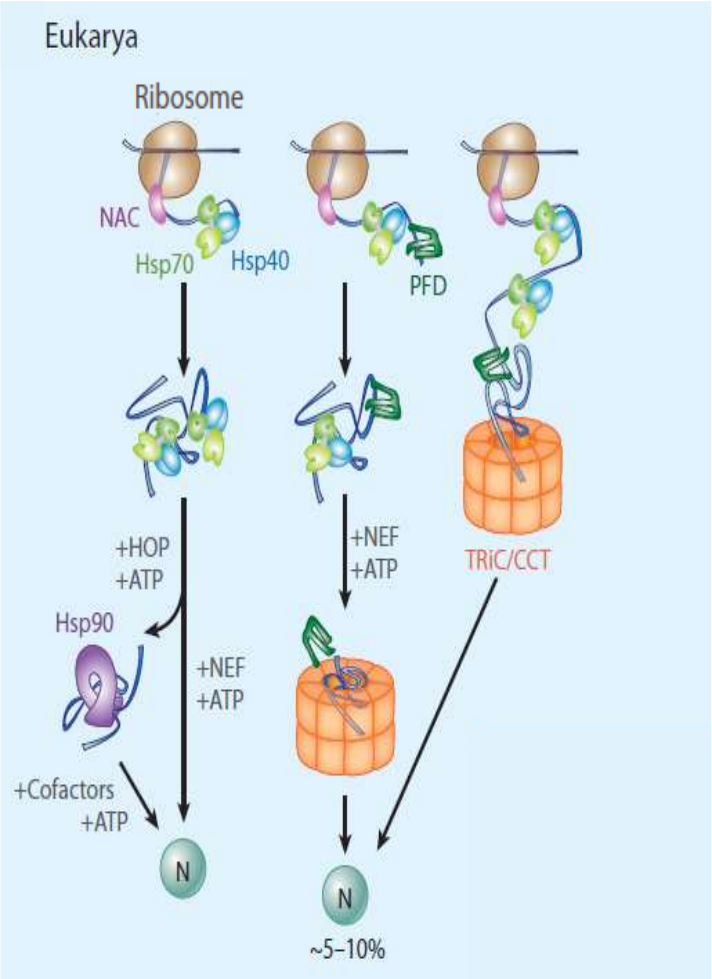


Figure 11: Eukaryotic chaperone pathways (Kim *et al*, 2013)

ii) *Protein degradation*

When polypeptides are irreversible misfolded or when proteins are damaged because of severe environmental stress, they have to be degraded. The ubiquitin-dependent proteasome pathway plays an important role in the misfolding protein degradation. An ubiquitin chain (ubiquitin is a 76 amino-acid peptide) is added onto lateral lysine of damaged proteins and serves as a signal to bring these proteins to the proteasome (Kravtsova-Ivantsiv & Ciechanover, 2012). The proteasome recognizes ubiquitin tags in substrates through its receptors and then initiates degradation at an unstructured region in the substrate. The substrate is then pulled into the degradation channel, the ubiquitin chain is cleaved off and the substrate unfolded and finally cleaved into peptides (Figure 12).

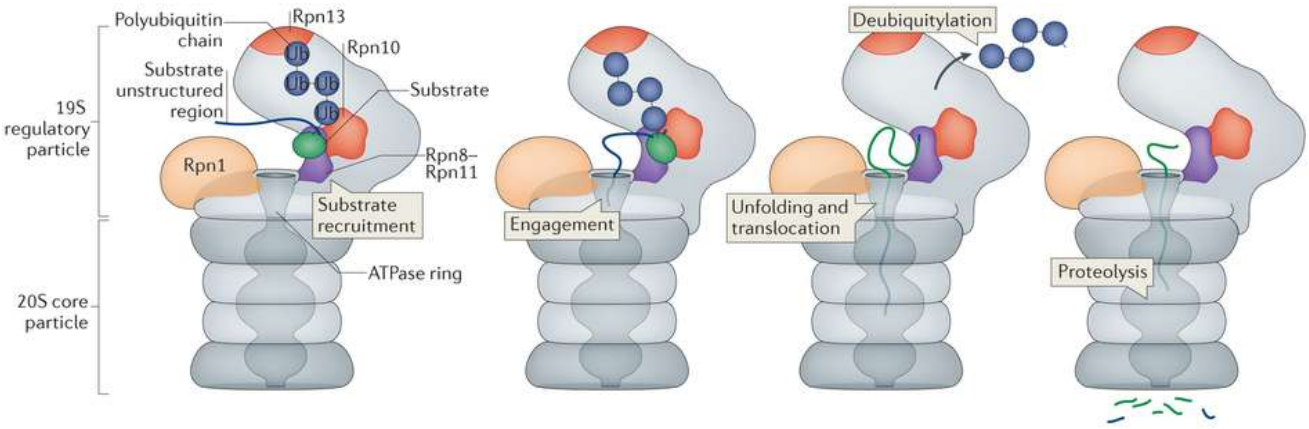


Figure 12: Steps of proteasomal degradation (Bhattacharyya *et al*, 2014)

This pathway also needs cooperation of chaperons such as Hsp70 and Hsp90 as well as other cofactors. The proteasome pathway can only degrade proteins in a non-aggregated state. Aggregated proteins are removed by autophagy and lysosomal/vacuolar degradation (Kim *et al*, 2013).

The gene expression process is well controlled at every step so that mean protein levels in the population are stable under constant environment. Nevertheless the protein level varies at the single cell level even in isogenic population and in fixed environment (Raser & O'Shea, 2005). This gene expression variability from cell-to-cell, also called "gene expression noise", confers phenotypic heterogeneity in the population, which might play a role in adaptation to changing environments and cell fate decisions (Raj & van Oudenaarden, 2008). Next we discuss origins and consequences of gene expression noise.

1.1.2 Gene expression noise

In an isogenic population, the expression level of every gene is variable from cell to cell even under constant environments. The term 'noise' is used to indicate this heterogeneity of gene expression levels. The stochastic phenomena underlying many cellular processes have attracted interests for many years because of their implications for cellular regulation and non-genetic heterogeneity, but the limits of the technologies allowing single cell analysis were the main obstacle for researchers to study them. Thanks to advances in techniques which enable precise and quantitative measurements of gene expression levels in single cells, the basic mechanisms underlying these phenomena and their consequences can be deciphered.

Gene expression noise refers to the gene expression variability in an isogenic population. To

quantify the level of noise, it is important to take into account the effect of the mean expression value. Several values are used to quantify gene expression noise among a population (Chalancon *et al*, 2012):

a) Coefficient of variation (CV): standard deviation (σ) divided by mean value (μ);

b) CV^2: variance (σ^2) divided by the squared mean (μ^2);

c) Noise strength (Fano factor): variance (σ^2) divided by mean (μ) (Thattai & van Oudenaarden, 2001).

The relationships between noise and mean value will be discussed later. In this dissertation, we will use CV or CV^2 to quantify gene expression noise. CV is the most direct and unambiguous measurement of noise (Kaern *et al*, 2005), while CV^2 better reflects than the CV how large the standard deviation is compared to the mean expression level (Chalancon *et al*. 2012). The noise strength could reveal some trends obscured by the dependence of noise to mean level (Kaern *et al*, 2005). For example, from the simulation of Kaern *et al*. (2005), noise decreases when mean level is increased by varying either transcription rate or translation efficiency, but noise strength increases when mean level is increased by varying translation efficiency.

1.1.2.1 Sources of noise

Gene expression noise can be divided into two components: intrinsic noise which is due to stochastic events during the gene expression process and affect the expression of every gene independently and differently, and extrinsic noise which is due to any existing cell-to-cell heterogeneity in cellular components involved in the global gene expression process in the cell and affect the expression of all genes in a given cell (Swain *et al*, 2002). Details about the sources of these two types of noise are given in the next section.

Using double fluorescent reporter systems, the contribution of each noise component can be deciphered (Elowitz *et al*, 2002). In yeast, Raser and O'Shea (2004) constructed diploid yeast strains that express both Cyan and Yellow Fluorescent Proteins (CFP and YFP) from identical promoters, integrated at the same locus on homologous chromosomes, to distinguish intrinsic noise and extrinsic noise (Figure 13).

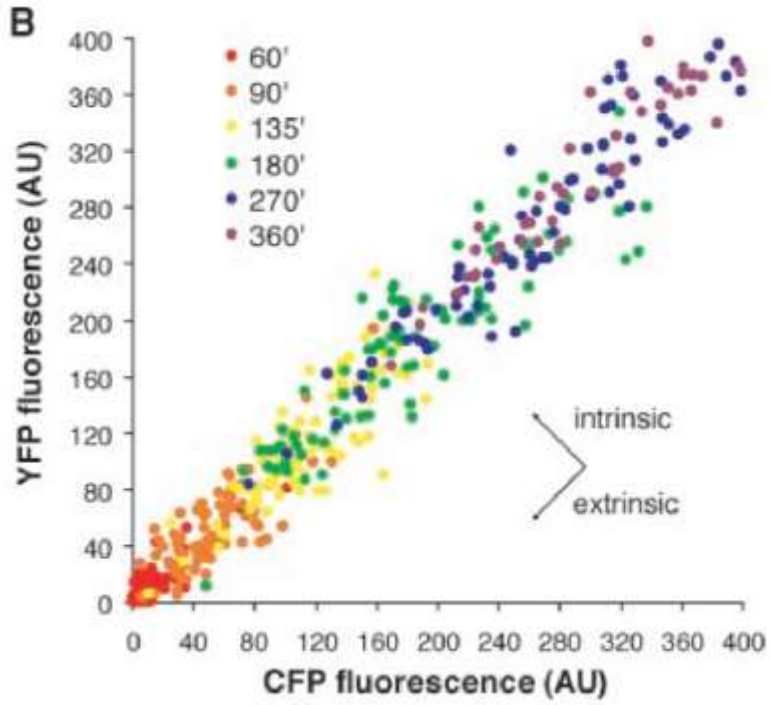


Figure 13: Deciphering of intrinsic and extrinsic noises. Extrinsic noise is manifested as scatter along the diagonal and intrinsic noise as scatter perpendicular to the diagonal (Raser & O'Shea, 2004).

In *S. cerevisiae*, the contribution of intrinsic noise and extrinsic noise to the total noise is variable and depends on the gene. The *GAL1* and *PHO84* promoters have small and constant intrinsic noise, while the *PHO5* promoter has larger basal intrinsic noise that decreases when gene expression level increases (Raser & O'Shea (2004)). Colman-Lerner *et al*. (2005) showed that the expression variability of the α-factor response gene *PRM1* mainly comes from differences among cells in the signal transition pathway which affect most genes (extrinsic noise). Moreover extrinsic noise is the main component for abundant protein while extrinsic and intrinsic noise are equal for proteins of intermediate abundance (Bar-Even *et al*. (2006)). Finally, Zhang *et al*. (2013) showed that intrinsic noise is dominant for the *HO* promoter. Thus, the level of intrinsic noise is gene-specific or promoter-specific and it is a heritable trait.

Recently, Rhee *et al*. (2014) proposed a method to decompose noise by nonequivalent reporters. They used two different reporters in the same signaling pathway. The signaling pathway can be simplified as: signal → common upstream L → different reporters (X and Y). Then the total noise can be decomposed as trunk noise and branch noise. Trunk noise is the noise introduced to L, whereas branch noise is the noise introduced to X and Y respectively. When equivalent reporters are used in this method, the trunk noise is the same as extrinsic noise, and the average of two branch noise is the same

as intrinsic noise.

1.1.2.1.1 Origins of intrinsic noise

Intrinsic noise is due to the inherent randomness of biochemical processes such as transcription and translation. These processes involve proteins that can be at low cellular concentrations, therefore exhibiting stochastic variations from cell-to-cell. More generally, stochastic fluctuations in production and degradation of mRNA and proteins generate intrinsic noise. Nevertheless transcription is by far the most studied process for its contribution to intrinsic noise.

i) Promoter sequence

The transcription machinery assembly at a promoter site can be perturbed in many ways, therefore causing intermittent and random bursts in mRNA production, especially for highly expressed genes. These successive bursts of mRNA production can largely influence gene expression noise (Hebenstreit, 2013; Chalancon *et al*, 2012). The statistics of mRNA expression can be described by two parameters: burst frequency (the frequency at which bursts occur) and burst size (the average number of mRNA produced within each burst). In theory, increasing burst size would increase mean expression without changing noise, whereas increasing burst frequency would increase mean expression and decrease noise (Hornung *et al*, 2012). Promoter structure affects transcriptional bursting and noise.

Recently, researches in *E. coli* showed that bacterial promoter structure dictates intrinsic transcriptional noise (Jones *et al*, 2014). The authors have systematically varied promoter strength, transcription factor binding strength, and transcription factor copy numbers in a set of promoters to observe how these changes affected variability in gene expression. In bacteria, it also appears that transcriptional bursting is strongly linked to the DNA supercoiled state (Chong *et al*, 2014). Positive supercoiling buildup on a DNA segment by transcription slows down transcription elongation and eventually stops transcription initiation. Nevertheless many studies on transcriptional-mediated noise also exist on eukaryotic cells, especially on yeast promoters.

a) TATA box

Systemic analysis of the yeast GFP-tagged library showed that TATA-box containing genes are noisier than the others (Newman *et al*, 2006). Also, mutations in the TATA-box of the *PHO5* promoter that weaken its strength decrease the noise level (Raser & O'Shea, 2004). Same results are observed with the *GAL1* promoter (Blake *et al*, 2006). It is suggested that the relationships between TATA box strength and noise (stronger TATA box confers higher noise) is due to its effects on the stability of the

pre-initiation complex (PIC) that influences the burst size (Sanchez *et al*, 2013). Strong TATA box increases expression level through stabilizing PIC what increases burst size without changing noise. Thus promoters containing strong TATA box exhibit higher noise than promoters conferring lower burst size and higher burst frequency that decreases noise.

By mutating a series of yeast promoters, Hornung *et al*. (2012) also found that decreasing the strength of TATA-box leads to reduction of gene expression noise. But it is only observed in the promoters that contain both a TATA box and a proximal nucleosome-occupied site. The effect of the TATA-box on noise is not significant in promoters lacking a proximal nucleosome. Thus it is assumed that the effect of the TATA box on noise is due to a combined effect of nucleosome positioning and TATA sequences (Sanchez *et al*, 2013).

b) ***Nucleosome positioning***

A promoter with nucleosome binding sites has two alternative states: OFF (silenced) and open (initiated) states. The transient open state leads to transcriptional bursting which generates transcriptional noise and cell-to-cell variability in gene expression (Sanchez *et al*, 2013). The deletion of any component of chromatin remodeling complexes and mutations in UAS sites that affect the recruitment of these complexes both increase the intrinsic noise of the *PHO5* promoter, because they prevent the removal of nucleosomes in the promoter region during activation (Raser & O'Shea, 2004). Using electron microscopy, Brown *et al*. (2013) have studied the different structural nucleosome configurations of these two states of the *PHO5* promoter, indicating that gene expression fluctuations are related to promoter nucleosome dynamics. Furthermore, they have observed the nucleosome configuration of two conjugated *PHO5* promoters, showing that this configuration is intrinsically variable and can be a cause of expression noise (Brown & Boeger, 2014). Single-cell nucleosome mapping found significant cell-to-cell variation in nucleosome positions during the induction of *PHO5* (Small *et al*. (2014)), suggesting that an underlying complexity of nucleosome positioning may contribute to the flexibility and heterogeneity of gene expression.

Recently, Sharon *et al*. (2014), using thousands of designed promoters, showed that poly nucleosome-disfavouring sequences (specifically, poly(dA:dT) tracts that are highly prevalent in eukaryotic promoters) result in lower noise through affecting the transcription burst frequency, what is consistent with the results of Dadiani *et al*. (2013). Also, computational analyses of genomic nucleosome occupation patterns found that promoters containing nucleosome binding sites close to the TSS have large cell-to-cell variability in gene expression (Tirosh *et al*, 2009). Otherwise, essential

genes which should be less noisy tend to be clustered in nucleosome-depleted open-chromatin regions on the chromosome (Batada & Hurst, 2007). Thus genes in open chromatin regions where nucleosomes are depleted have lower noise whereas genes covered by nucleosomes have higher noise (Sanchez *et al*, 2013).

Moreover, the nucleosome remodeling dynamics can affect the transition from TF signal input to expression output, with effects on expression noise. Hansen and O'shea (2013) studied the response of the Msn2 target genes to dynamical Msn2 inputs. They find two different classes of promoters: promoters having high amplitude threshold (threshold of stimulus for the activation of expression) and slow activation process (HS), and promoters having low amplitude threshold and fast activation process (LF). The activation process is related to the nucleosome remodeling process. Thus HS promoters need strong input signal to be activated and the activation process is slow because of slow dynamics of nucleosome remodeling. Meanwhile, LF promoters can be activated by weak input signals and the activation process is fast because of fast dynamics of nucleosome remodeling. In consequence, HS promoters can transmit strong signals to heterogeneous expression responses (noisy expression) and filter out weak signals, whereas LF promoters confer stable expression within both conditions.

c) ***Transcription factor binding sites***

Transcription factor (TF) binding site strength, number and position also have an influence on promoter-mediated noise. To & Maheshri (2010) inserted different number (1 or 7) of the tet-transcriptional activator (tTA) binding site (tetO) into a synthetic yeast promoter and found that the 7XtetO promoter exhibited higher noise than the 1XtetO promoter. Recently, same results were obtained by Sharon *et al*. (2014). They tested thousands of artificial promoters with different numbers of nucleosome disfavoring sequences and different numbers of Gcn4 binding sites and found that, for a given expression level, promoters that contained more Gcn4 binding sites were noisier. This phenomenon might be due to stochasticity of TF binding and falling off the promoter (Sanchez *et al*, 2013).

Moreover, TF binding site position can also affect gene expression noise. Octavio *et al*. (2009) put tetO at different positions within the *FLO11* promoter and showed that promoters with tetO close to the TATA box have lower noise than the others which is consistent with results on the *GAL1* promoter (Murphy *et al*, 2007).

One promoter can also have different types of binding sites for the same TF. Carey *et al*. (2013) showed an interesting expression profile of *ZRT2* expression which is regulated by Zap1. As Zap1

expression level increases, *ZRT2* expression level first increases and then decreases, while its noise level first decreases and then remains constant. Thus at the same mean expression level, the distribution at high induction is less noisy than at low induction. It can be explained by the fact that the *ZRT2* promoter has two different Zap1 binding sites, an activating binding site and a repressive binding site. At low induction, the activation site increases the expression level by increasing burst frequency and thus decreases noise whereas at high induction, the repression site near the TATA box decreases *ZRT2* expression through burst size reduction that does not change noise level and leads to lower expression with lower noise.

ii) Chromosomal structure

a) Chromatin epigenetics

Chromatin epigenetics which includes histone modifications and DNA methylation that can be removed and added to change the state of a promoter (on/off) contributes to gene expression noise (Chalancon *et al*, 2012).

In diploid yeast cells, the two copies of *FLO11* have random epigenetic states that change independently and lead to expression heterogeneity in the population (Octavio *et al*, 2009). Meanwhile, in theory, Miller-Jensen *et al.* (2011) proposed a two-state model that explains how chromatin remodeling can affect noise level. More recently, Weinberger *et al.* (2012) measured expression of *ADH3-GFP* in yeast strains with various deletion of chromatin-remodelling factors and showed that histone deacetylation complexes (HDACs) have different effects: Rpd3(L)C and Set3C both repress expression but Rpd3(L)C only also increases noise. Vinuelas *et al.* (2013) used 5-azacytidine (inhibitor of DNA methyltransferase) and trichostatin A (inhibitor of histone deacetylase) to modify the chromatin state of chicken cells and showed that chromatin dynamics has remarkable effects on gene expression noise (but direction and magnitude are different among different cell lines).

Finally, expression heterogeneity linked to chromatin epigenetics can be inherited during several cell cycles (a phenomenon called memory) and then become stochastic in the progeny cells (Kaufmann *et al*, 2007). Zhang *et al.* (2013) showed that *HO* promoter states in a mutated yeast strain (*swi5*) where its expression profile is bimodal are inherited during one cell generation. By mutation analysis, they suggested that this memory is related to histone acetylation in the promoter region. Recently, researches on human cells showed that this memory also depends on mRNA and protein stability (Corre *et al*, 2014).

b) Chromosomal position

Gene position on chromosomes also has important effects on noise. The position can affect gene expression activation rate of low abundance mRNA and produce remarkable variability in the population (Becskei *et al*, 2005). Anderson *et al.* (2014) studied the expression profile of *TLO* genes at different chromosome positions in *Candida albicans*. They found that *TLO* genes in sub-telomeric regions where genes are highly silenced by chromatin modifications exhibit higher noise levels than in chromosome internal regions and that the deletion of the key silencing gene *SIR2* can reduce the noise level from sub-telomeric regions.

c) Gene loop

Gene loop makes its promoter and terminator spatially close to each other and form a 'bridge' for the polymerase to join the PIC immediately after transcription termination. This phenomena can cause transcription bursting which is a major source of expression noise in eukaryotes (Hebenstreit, 2013).

iii) Others

In theory, all the steps of the gene expression process can have effects on intrinsic noise, such as nuclear architecture, translation rates, mRNA degradation rates and protein degradation rates (Chalancon *et al*, 2012). Blake *et al.* (2003) have shown that, under full transcription efficiency, translation efficiency has little effects on noise while with partial transcription, high translation efficiency leads to high noise. Noise arising from transcription contributes more than noise generated at the translational level in eukaryotes, while translational level is the dominant source of expression noise in prokaryote because of the translational burst from individual transcripts (Ozbudak *et al*, 2002).

1.1.2.1.2 Origins of extrinsic noise

i) Variable availability of machineries

The gene expression process depends on the abundance of expression machineries in the cell, such as RNA polymerases or ribosomes, and this abundance varies a lot in the population, having great effects on global noise (Raj & van Oudenaarden, 2008). The number of ribosomes can affect translation efficiency, and some researchers showed that translation efficiency has remarkable influences on gene expression noise (Guimaraes *et al*, 2014; Salari *et al*, 2012). Yang *et al*. (2014) constructed an interesting system in *E. coli* where they can follow the abundance of T7 RNAP by fusing YFP to the protein and also

can follow gene expression controlled by T7 RNAP through fusion of CFP and mCherry to T7 promoters. The results showed that extrinsic noise is correlated to the variation of RNAP abundance whereas intrinsic noise is independent to it. Moreover Johnston *et al*. (2012) proposed a model explaining that differences of volume and functionality of mitochondria among cells strongly affect expression variability.

ii) *Noise propagation*

The intrinsic noise of upstream regulators can propagate to downstream targets as a source of extrinsic noise (Chalancon *et al*, 2012). Increasing expression noise of a TF raises expression variability of its target genes in yeast (Blake *et al*, 2003). Hooshangi *et al*. (2005) constructed artificial transcriptional cascades in *E. coli* showing that noise can be amplified through transmission from regulators to targets. Deletion of the MAP kinase Fus3 increases variations of the α-factor response gene *PRM1* expression in yeast at high concentration while the deletion of MAP kinase Kss1 decreases variations at low concentrations, indicating the importance of signal transmission to expression noise (Colman-Lerner *et al*, 2005). Moreover high noise in TF expression can lead to bimodal expression of downstream genes in yeast (To & Maheshri, 2010). Stewart-Ornstein *et al.* (2012) performed systematic pair-wise correlation analysis demonstrating noise transmission from upstream factors to several downstream genes along their regulation pathway. But researches on HeLa cells showed that propagation of noise from TFs to their targets is blocked by formation of heterodimers and chromatin compaction (Shah & Tyagi, 2013).

Otherwise, fluctuations of metabolism genes can lead to growth fluctuations in the population, and furthermore can cause fluctuations of other genes which have no functional interactions between them (Kiviet *et al*, 2014). The authors studied fluctuations of the *lac* operon genes induced by different concentrations of IPTG in the medium using lactulose as carbon source in *E. coli* which was metabolized by *lac* genes but didn't induce this operon. They found that fluctuations of the *lac* genes can cause growth fluctuations which leads to increased expression variability of unrelated genes.

iii) Asymmetric partitioning during cell division

During cell division, segregation of mRNAs and proteins is not equal (asymmetric partitioning), and thus causes heterogeneity between mother and daughter cells. Few studies focused on this aspect and most of stochastic gene expression models consider cell division with symmetric partitioning, while Huh and Paulsson (2011) proposed a stochastic model including asymmetric partitioning during cell

division. Their *in silico* simulation shows that random segregation is a great source of extrinsic noise in gene expression.

iv) Cell cycle

In a population, cells are in different cell cycle phases in which global transcription and translation activities are different, thus cell cycle is an important extrinsic source of expression noise. Colman-Lerner *et al.* (2005) constructed a yeast strain that can be arrested at the G2/M transition and showed that arrested cells reduced total noise by 45%. Moreover, a synthetic noisy promoter (7XtetO) in yeast showed that expression noise is driven by differences in transcription rates between G1 and S/G2/M (Zopf *et al.* (2013)).

v) Fluctuations micro-environment

The micro-environment encountered by microorganisms in their culture medium is not totally homogeneous and might lead to different global expression noises (Chalancon *et al*, 2012). Interestingly, in multicellular organisms, the micro-environment can be changed via paracrine signal, which further regulates gene expression noise. Shalek *et al.* (2014) used single-cell RNA-seq to show that paracrine signaling can affect cellular expression heterogeneity through cell-to-cell communication in mouse bone-marrow-derived dendritic cells.

vi) *Cell volume*

Recently, Kempe *et al.* (2014) took into account the effects of cell volume to expression noise in human cells. They found that the variability of mRNA concentration (mRNA number divided by cell volume) is due to different gene expression activities among cells, while the noise of mRNA number is much larger because of the variance in cell volume.

1.1.2.2 Expression noise is a quantitative trait

As for the mean expression level of genes, quantitative trait studies can be used to identify the genetic determinants of noise between different yeast strains. Quantitative trait studies combine the genetic information and phenotypic information from different strains to identify the regions or loci that might contribute to a quantitative strait (referred as Quantitative Trait Loci (QTL)) through statistic tools. Ansel *et al.* (2008) first mapped 3 QTLs affecting *pMET17-GFP* expression noise in yeast using BY×RM F1 segregants and introgression. They found that the defective uracil metabolism increases

expression noise of *pMET17-GFP* by impairing transcription elongation. Furthermore, successive backcrosses have been performed by selecting clones with higher noise and unaffected mean value (Fehrmann *et al*, 2013). After 7 generations, 3 high noise segregants were generated. By genome sequencing, four loci which contribute to high noise without changing mean expression level were identified. The authors defined these loci as expression Probabilistic Trait Loci (ePTL) that can change the probability that an individual displays a given expression level without necessarily changing the mean expression value of all individuals carrying the same genotype. Finally, they finely mapped 2 loci: one with a shifted mutation in the Erc1p transmembrane transporter in BY which reduces expression variability of *pMET17-GFP* and one with *cis*-regulatory polymorphisms in the *MUP1* methionine permease gene in RM which increase expression of *MUP1* in *cis* and expression variability of *pMET17-GFP* in *trans*.

Thanks to the advances in high throughput next-generation technologies, QTL analysis provides a powerful tool to study the effects of genetic variants on expression noise. But these studies can only identify effects in *trans* which come from upstream elements in pathways related to the expression of the gene studied. They cannot identify *cis*-effects coming from promoter sequences.

1.1.2.3 Relationship between noise and mean expression levels

In general, high expression level is associated with low noise. This correlation has been verified with different promoters and in several conditions in yeast. Figure 14 gives some examples of noise versus mean plots. The first graph (A) shows the yeast GFP-tagged library (Newman *et al*, 2006). It contains thousands of proteins fused to GFP in C-terminal and under the control of their native promoter. The second graph (B) represents the natural *ZRT1* promoter fused to GFP (Carey *et al*, 2013) induced by different concentration of Zn^{2+}. The third graph (C) shows mutated *ERG6* promoters fused to GFP (Hornung *et al*, 2012). Finally, the fourth graph (D) represents the synthetic promoter p4XUPRE (the *CYC1* promoter with four UPRE (unfolded protein response element)) motifs fused to GFP (Zuleta *et al*, 2014) and induced by different concentration of tunicamycin. The noise / mean relationship can be defined as $\eta^2 = \frac{b}{m} + \eta^2_{ext}$, where η^2 is the noise, m the mean expression value, b the bust size, η_{ext} the extrinsic noise (Hornung *et al*, 2012). But this formula is only correct when the expression level is low or medium. For high expression levels, extrinsic noise is dominant and the total noise is more stable when the mean increases (Zuleta *et al*, 2014), as shown in graph A and D.

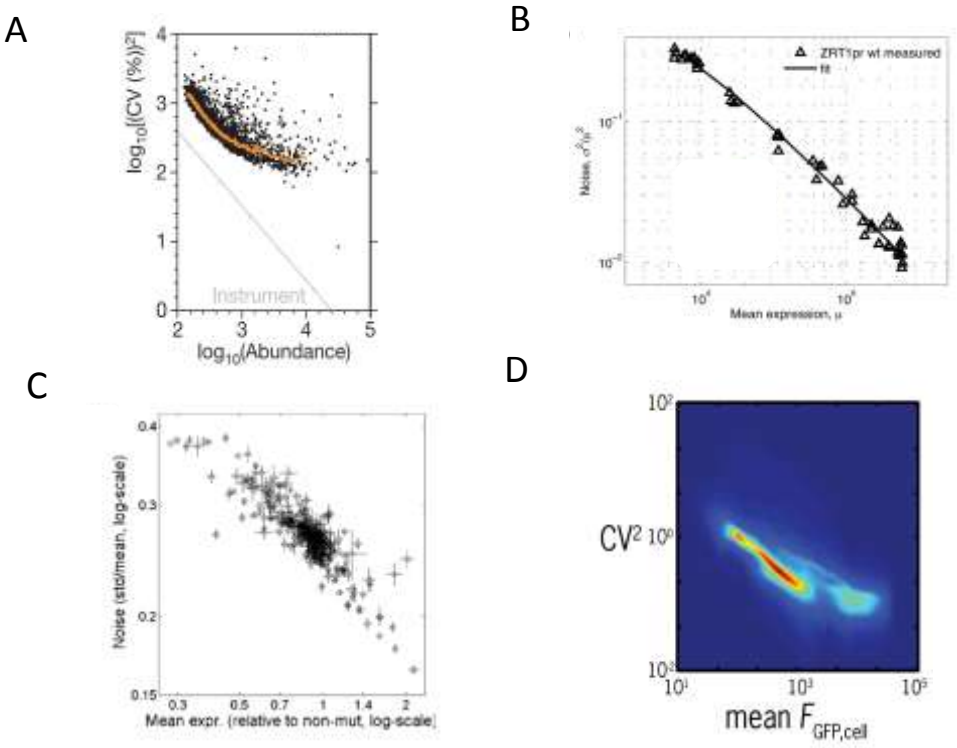


Figure 14: Relationship between noise and mean expression levels in various experiments in yeast.

This is the global tendency between expression noise and mean expression level. We can decouple these two elements through genetic modifications in the promoter. The modifications that only increase burst frequency increase expression level and decrease noise level whereas the modifications that only increase burst size will increase expression level without changing noise level.

1.1.2.4 Expression noise versus expression plasticity

Expression noise is the expression variability of one gene among different cells in an isogenic population, while expression plasticity is the expression variability of one gene among different conditions. These two phenomena have common correlations as the ones with strong TATA box containing promoters and high nucleosome occupancy promoters. Thus it is proposed that they might be highly coupled (Tirosh *et al*, 2009). Some genes have been shown to exhibit coupling between expression noise and plasticity (Blake *et al*, 2006; Anderson *et al*, 2014).

In contrast, large analyses of yeast expression data showed that the coupling of noise and plasticity is not a general rule (Lehner, 2010; Bajić & Poyatos, 2012). It is linked to the property of particular promoter architectures, and also under evolutionary constraints. Noise and plasticity are uncoupled for essential genes and gene duplication favours the coupling (Lehner, 2010). The

relationship between noise and plasticity is more complicated than expected, depending on the regulation strategy of the genes. As shown in Figure 15, TATA box containing or SAGA regulated promoters together with nucleosome-covered promoters have high noise and plasticity, whereas poor translational regulated genes with low translational efficiency have low noise and plasticity. Short genes with high translational efficiency exhibit high noise but low plasticity. Some growth-related genes that are dependent on TAF1 exhibit low noise but high plasticity (Bajić & Poyatos, 2012). Recently, similar results have obtained in *E.coli* (Singh, 2013), showing that the evolution and selection for coupling of noise and plasticity are common in different organisms.

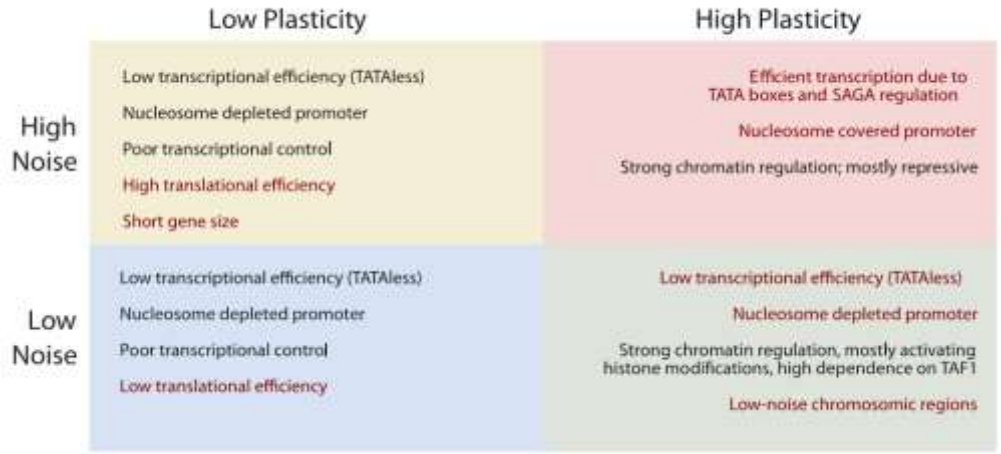


Figure 15: Relationship between expression noise and plasticity (Bajić & Poyatos, 2012)

1.1.2.5 Bimodal expression

In general the distribution of gene expression levels in a population has only one peak (Figure 16, A). It means that the majority of the population is around the mean value. But some genes can have a bimodal expression profile (Figure 16B) where the distribution has two peaks. Thus bimodal expression confers high noise. This expression pattern makes the population having two highly different sub-populations that can have very different behaviours. This phenomenon has been found in bacteria (Silva-Rocha & de Lorenzo, 2012), yeast (Pelet *et al*, 2011) and mammalian (Shalek *et al*, 2014), showing that it might be a common evolutionary strategy.

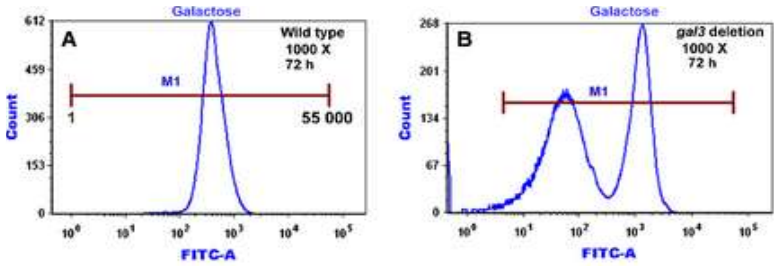


Figure 16: Unimodal and bimodal expression profiles in *GAL1* expression in yeast (Kar *et al*, 2014)

The underlying mechanism of this pattern is still not clear. First, high noise in an upstream regulator can lead to bimodal expression of downstream outputs. Using an artificial regulation system in yeast, Blake *et al.* (2003) first observed bimodal expression in early induction stages, but the population returned to unimodal in later stages. Then by increasing the expression noise of the input regulator, they observed stable bimodal expression during long-term induction. Similarly, To and Maheshri (2010) achieved bimodal output by increasing input noise through adding more TF sites in the promoter of a regulator. Recently, it was found that *GAL1* has a bimodal expression pattern in the yeast *gal3Δ* strain when induced by galactose (Kar *et al*, 2014) (Figure 16). This pattern is caused by the lognormal distribution of Gal4p. Second, intermediate stimulus can induce bimodal expression in yeast. Under the stimulus of 0.1 M NaCl, the *STL1* promoter exhibits bimodal pattern expression in yeast (Pelet *et al*, 2011). Moreover, being pretreated with 8 μM ATc, an artificial regulation network exhibits bimodal output when the carbon source was changed from glucose to galactose, while without pretreatment or with 250 μM ATc, it exhibits unimodal output (Wu *et al*, 2013). In mammalian cells, key immune genes are already expressed in bimodal pattern under normal stress (Shalek *et al*, 2014). Other bimodal expression patterns are found in specific conditions. For example, Silva-Rocha and de Lorenzo (2012) showed that mixed carbon sources medium (succinate and m-xylene) induces the bimodal expression of *Pu* and *Pm* (promoters induced by m-xylene) in *Pseudomonas putida* mt-2. Also, antisense transcription of *PHO84* can lead to bimodal expression of its sense transcription in yeast (Castelnuovo *et al*, 2013).

1.1.2.6 Evolution of gene expression noise

Gene expression noise is a gene special traits under natural selection and evolution. Newman et al. (2006) systemically measured the expression noise of the yeast GFP tagged genes library and showed that house-keeping genes have lower noise while stress related genes have higher noise. Further analysis of their results showed that genes which is essential for the proliferation tend to keep

low expression noise (Lehner, 2008). Recently Metzger et al.(2015) compared the expression profile of 85 nature variants of TDH3 promoter and 236 point mutations in the same promoter and showed that the TDH3 promoter is undergoing natural selection to minimize noise. These researches suggested that essential genes are subjected to purifying selection to maintain low expression noise.

In contrast, positive selection for high noise are also observed for some genes. Zhang et al. (2009) reanalyzed the data of Newman et al. (2006) and showed that plasma-membrane transporter genes exhibit elevated expression noise which might confer some advantages under certain conditions such as fluctuant and unpredictable environments where higher noise can confer better growth with less cost. Artificial system in E. coli showed the evolution towards higher noise under stressful conditions while more evidence should be found in yeast.

1.1.3 Measurement of noise

The measurement of noise strongly relies on advances in single-cell technologies allowing simple and accurate expression measurement in individual cells. Gene expression noise studies need tools to detect the amount of molecules (mRNA or protein) in single cells. But considering the limited number of these molecules and the necessary sampling volume to get accurate statistical results, these methods must have capacities of high-resolution and high-throughput (Ohno *et al*, 2014). Recent advances have made possible the detection of single molecules in single cells, and the analysis of thousands of individual cells during one experiment (Golding *et al*, 2005; Huang *et al*, 2007; Castelnuovo *et al*, 2013). Besides high sensitivity, reproducibility among biological and technical repeats is also critical (Lidstrom & Konopka, 2010). Combining requirements of high-throughput and minimization of experimental handling errors, automated manipulation systems are preferred (Newman *et al*, 2006; Little *et al*, 2013; Zuleta *et al*, 2014). Ultimately, single cell methods are preferably non-destructive so that they can follow the dynamics of living cell to dissect the temporal property of gene expression profiles (Rosenfeld *et al*, 2005; Huang *et al*, 2007). Besides, simultaneous multiparameter analysis is the trend of single cell methods (Lidstrom & Konopka, 2010): it can be used to understand the regulation network of expression heterogeneity or be used to combine phenotypic variability analysis with gene expression noise measurement.

1.1.3.1 Single-cell manipulation

1.1.3.1.1 Flow cytometry and microscopy

In the research field on gene expression noise, flow cytometry and microscopy that can detect fluorescence signals in individual cells play a very important role that is hard to overstate. Both of them need to target specific molecules by fluorescence-tagging. Also, they allow multi-parameter analysis through tagging of several targets with different fluorescence markers (Spiller *et al*, 2010).

i) Flow cytometry

The main advantage of flow cytometry is its high-throughput capacity. Moreover, sample preparation is generally easy. The traditional flow cytometry consists of 3 key systems: the fluidic system, the optic system and the electronic system. The fluidic system drives cells to the narrow "core" of a capillary, where the fluorophores can be excited, by hydrodynamic forces produced by flowing sheath fluid surrounding the injected samples. The optical system allows excitement of the fluorophores by lasers at precise wavelengths and collects emission fluorescence (or scatter lights). The electronic system converts optical signals into electric signals that can be treated by the computer (Shapiro, 2003). Thus flow cytometry can be used to measure the level of target fluorescences (reflecting the amount of target molecules), the forward scatter (reflecting cell size) and the side scatter (reflecting cell complexity). New technologies have been applied to improve the capacity, sensitivity and stability of flow cytometry, such as the Attune® acoustic focusing cytometer from Life Technologies which uses ultrasonic waves to position cells into a single, focused line along the central axis of the capillary and thus reduces the variability due to cell positioning in the capillary (Picot *et al*, 2012). Other new systems have been developed by combining flow cytometry with other technologies. For instance, the Amnis® Corporation developed the imaging flow cytometry (George *et al*, 2004) combining high-speed multi-spectral imaging with flow cytometry. This system can detect the fluorescence level and take a picture of the cell at the same time, thus combining physiological and morphological data. Ehrlich *et al.* (2011) used a parallel multi-channel microfluidic system with flow cytometry to largely increase the detection capacity of rare events. Bendall *et al.* (2011) developed a new technology called mass cytometry which uses time-of-flight (TOF) mass spectrometry to detect and quantify target molecules labeled by antibodies with heavy metals (isotopes). This machine has lower sensitivity compare to fluorescence-based flow cytometry, but largely increases the multi-parameter analysis capacity. Finally, Zuleta *et al.* (2014) combined a robotic manipulating system with flow cytometry to enable a simultaneous high-throughput and time-resolved measurement of gene expression profiles,

cellular growth rates and protein degradation rates.

ii) Microscopy

Using fluorescent microscopy, the abundance of the labeled target molecules can be measured by the optical intensity of fluorescence spots (Elowitz *et al*, 2002; Ozbudak *et al*, 2002). The advantage of microscopy is that it can provide the spatial distribution of the target molecule in the cell. Also, it can provide cell shape information as well as the cell cycle state (in yeast) (Zopf *et al*, 2013). An important application is the time-lapse microscopy which regularly records pictures of selected clones along the experimental procedure, thus providing the temporal dynamics of gene expression (Sanchez & Golding, 2013). The recent advances in microscopy technologies are mainly in the improvement of their resolution, for instance, spatial resolution has been beyond the optical diffraction limit (Davis, 2009). It is especially important to study expression noise at the mRNA level (see below).

1.1.3.1.2 Microfluidic system

Microfluidic systems have attracted more and more attention in the study of expression noise because they can combine multi-parameter measurement and single cell manipulation (de Vargas Roditi & Claassen, 2015). Figure 17 gives two examples of microfluidic systems. On the left is a microfluidic system for imaging (Cai *et al*, 2008). Cells were kept in chambers to be recorded by microscopy for the fluorescence. On the right is the C1^{TM} Single Cell Auto Prep chip from Fluidigm that can capture only one cell in each chamber of the chip (Shalek *et al*, 2014). Then the chip can be used for DNA-Seq or RNA seq for instance (see below for these methods).

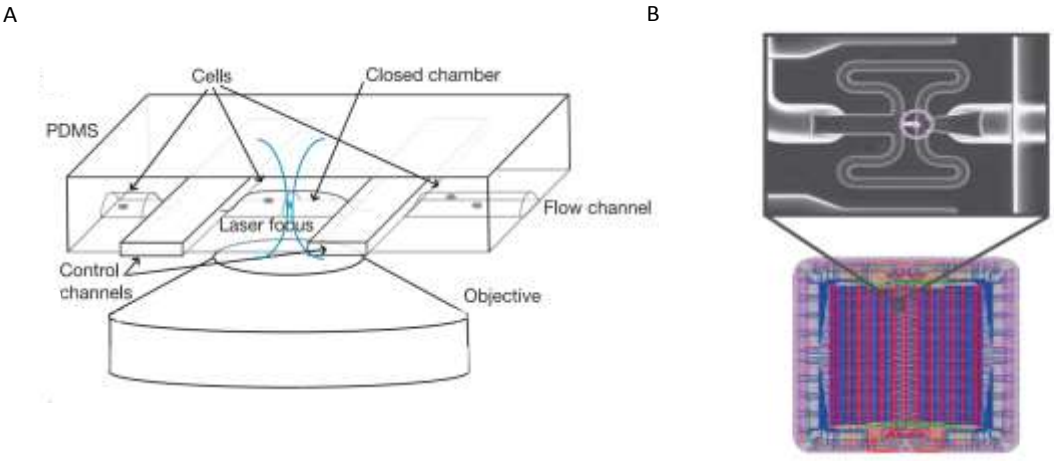


Figure 17: Examples of microfluidic systems

Microfluidic systems are not only used to capture single cell, they can also be used to perform long-term live cell imaging (Raj & van Oudenaarden, 2009) that allow studying the response dynamics

of the target by adding different stimuli in the system. An important advantage of microfluidic systems to study expression noise is that they can reduce environmental fluctuations (Spiller *et al*, 2010), because the chambers are very small and each chamber carries out the same stimulation at the nanoliter or even picoliter scale. Hansen and O'shea (2013) used a microfluidic system to study the transition from environmental signals to expression outputs and found that slowly activated promoters exhibit high noise response expression.

1.1.3.2 RNA level

In eukaryotic cells, expression heterogeneity mainly comes from the transcription process (Blake *et al*, 2003; Ochiai *et al*, 2014). Thus, it is interesting to carry out research at the mRNA level. mRNA expression noise analysis uses methods for studying mRNA at the single cell level. Here two types of methods are discussed: visualizing methods (based on fluorescence tagging and microscopy) and non-visualizing methods (based on single cell cDNA synthesis and sequencing).

1.1.3.2.1 Visualizing methods

Visualizing methods are indirect methods where the number of target mRNA molecules is measured by fluorescent reporter molecules binding to the target. There are two major tagging systems to visualize mRNA: the MS2 system and the fluorescent *in situ* hybridization system (FISH).

i) MS2 system

MS2 is a bacteriophage whose coat proteins strongly bind to a particular RNA hairpin (Raj & van Oudenaarden, 2009). This system needs two steps of recombination. First, the coat protein gene fused to the *GFP* gene is integrated into the genome and expressed under a constitutive promoter. Second, several repeated RNA hairpins are integrated into the 5'-untranslated region of the target gene. Finally, the target mRNA produced by RNAPII is bound by the GFP-tagged protein through the hairpin structures, and thereby can be visualized by fluorescence microscopy. Golding *et al.* (2005) using this method in *E. coli* demonstrated that the artificial promoter $P_{lac/ara}$ (which is induced by IPTG and arabinose) stochastically switched between an active and an inactive state. Recently, Ochiai *et al.* (2014) constructed this system in mouse embryonic stem cells to study the expression profile of the important transcription factor Nanog. They found that the *Nanog* gene is stochastically activated to be transcribed and that this 'intrinsic noise' contributed to 45% of the total expression variability.

ii) RNA-FISH

This method visualizes target mRNA molecules through fluorescent labeled probes that specifically hybridize with the target *in situ* (Junker & van Oudenaarden, 2012). This method is widely applied to study expression noise. In yeast, RNA-FISH has been used to reveal the expression profile of several genes (Zenklusen *et al*, 2008). Constitutively expressed genes exhibit smaller than expected noise, whereas a SAGA-regulated gene, *PDR5*, has larger expression variability through transcription bursting. More recently, Zopf *et al.* (2013) studied expression from an artificial promoter 7texO in yeast at different cell cycle stages. They found that the different transcription rates among the different cell cycle stages largely contribute to expression variability.

1.1.3.2.2 Non-visualizing method

Non-visualizing methods are methods where absolute mRNA concentrations in the cell can be measured. Here we discuss two major methods: mRNA-sequencing and RT-qPCR.

i) mRNA-sequencing

Thanks to advances in high-throughput next generation sequencing, single cell mRNA-seq can be carried out to study gene expression noise in a population. The first challenge of single cell RNA-seq is how to capture single cells. Shalek *et al.* (2013) first proposed fluorescence activated cell sorting (FACS) by flow cytometry. Single cells were sorted into 96 well-plates, and the plates were then subjected to regular mRNA extraction, cDNA synthesis and sequencing. This method was employed to study the response of mice bone-marrow-derived dendritic cells (BMDCs) to lipopolysaccharides. They found that immune system related genes exhibit bimodal expression after stimulation. Recently, the same group (Shalek *et al*, 2014) proposed a new method using a microfluidic system to capture single cells. They studied how mice BMDCs respond to several conditions, showing that cell-to-cell communication contributes to expression variability. Obviously, the main advantage of single cell mRNA-seq is that it can simultaneously measure the concentration of all mRNA in the same experiment.

ii) RT-qPCR

Even if the cost of next generation sequencing is much cheaper than few years ago, considering the minimum sampling volume to get statistical reliable results of single cell RNA-seq, it is still very expensive. Thus, if only few genes have to be measured, single cell RT-qPCR is more appropriate. For instance, Bengtsson *et al.* (2005) used a micromanipulator to isolate single mice islet cells and carry out RT-qPCR on 5 genes for each cell. They found that gene expression levels in the population are lognormally distributed. Warren *et al.* (2006) combined single cell sorting by FACS and digital RT-qPCR

that uses a microfluidic device to achieve high throughput RT-qPCR analysis to study the expression profile of *PU.1* and *GAPDH* in five hematopoietic precursors, showing different expression patterns in different cell types.

1.1.3.3 Protein level

1.1.3.3.1 Fluorescent proteins

The use of fluorescent proteins in the researches on gene expression noise is quite common. For instance, to study the effects of different promoter architectures on expression noise in yeast, a gene coding for a fluorescent protein can be directly fused to the promoter sequence (Raser & O'Shea, 2005; Sharon *et al*, 2014). Of course, the fluorescence protein can be fused to the studied protein too (Newman *et al*, 2006; Anderson *et al*, 2014). One advantage of this method is that the protein amount can be followed in real-time experiments. New fluorescence proteins are still developed to enlarge the spectrum to achieve multi-parameter analysis or to increase detection capacity and sensitivity (Ohno *et al*, 2014).

1.1.3.3.2 β-galactosidase

β-galactosidase can hydrolyze synthetic fluorogenic substrates to produce various fluorescent products. A microfluidic system has been developed to determine the β-galactosidase expression pattern in *E. coli*, yeast and mouse cells showing its potential applications to study gene expression noise (Cai *et al*, 2006). The fluorescent molecules produced by β-galactosidase are rapidly expelled from the cell and the fluorescence is measured inside a microfluidic chamber where single cells are captured. Thus, this method can reduce the risks of phototoxicity and autofluorescence, and be used to study low protein numbers.

1.1.3.3.3 Antibodies

The expression level of a protein can also be measured through a fluorescent labeled antibody. This method does not need any gene recombination but it is highly affected by the recognition ability of the antibody. Moreover, cells have to be permeabilized to give the antibody access to the protein if it is not on the cell surface, thus this method cannot be used to perform real-time analysis. Huang *et al.* (2007) proposed this method to study the distribution of β2 adrenergic receptors expressed in insect cells (SF9), showing diverging expression levels in the population. Shi *et al.* (2012) also used protein antibody chips to study the expression pattern of genes involved in the PI3K signaling pathway in tumor

cells showing notable cell-cell heterogeneity. Chang *et al.* (2008) used antibodies to label the stem cell marker Sca-1 in mouse haematopoietic progenitor cells. They found marketable differences of Sca-1 level in the population that lead to different cell states that have different differentiation tendencies.

1.1.3.4 Cell free system

A cell free system is a cell-sized container where are placed all the necessary elements to achieve the studied biological process. Recently, two articles have been published using cell-free systems to study gene expression noise (Karig *et al*, 2013; Nishimura *et al*, 2014). They both showed that expression noise is an inherent property of the expression process which is determined by the structure of the genes. The advantage of cell-free systems is that they can eliminate extrinsic noise and study different aspects of intrinsic noise.

1.1.4 Consequences of noise

Expression variability (noise) enables an isogenic population to contain non-genetic phenotypic heterogeneity (Raser & O'Shea, 2005) that can be important in stressful environments and for cell fate decisions. On one hand, non-genetic heterogeneity can make the population have different responses to extrinsic signals and further have different fates. On the other hand, this heterogeneity can give the population more chances of adaptation to variable conditions thanks to subpopulations that might be already adapted to certain stressful conditions (a phenomenon also called bed-hedging (Beaumont *et al*, 2009)). Moreover, noise dependent heterogeneity could be tunable under selection and has effects on long timescale evolutionary transition because it represents an evolutionarily accessible phenotypic parameter (Eldar & Elowitz, 2010). The relationships between noise and adaptation will be discussed in depth in section 1.2, here we only focus on cell fate decisions.

1.1.4.1 Cell state of unicellular organisms

Unicellular organisms can exhibit different cell states or types depending on the environmental conditions. Moreover microorganism populations can contain heterogeneous subpopulations with different behaviours and properties, and the transition from one cell type to another can be a stochastic process. Noise in the expression of key genes can play an important role during these cell fate transitions. An example is the transition to the competent state in *Bacillus subtilis*, which gives the

cells the ability (competence) to uptake exogenous extracellular DNA in stationary phase. Only around 15% of cells in the population become competent. This transition depends on the expression level of *ComK*. When the number of ComK proteins exceeds a threshold, the competent state is acquired. By increasing the rate of transcription and decreasing the rate of translation, Maamar *et al.* (2007) decreased the expression noise of *ComK* with similar mean expression level compared to wild type cells. They found that low *ComK* noise highly decreased the transition rate to competence. Another example is yeast sporulation. Nachman *et al.* (2007) followed the sporulation dynamics in yeast (*S. cerevisiae*) population during nutritional starvation. They found that the entry in sporulation and the duration of sporulation are both variable among identical cells and this variability is determined by the variable expression level of *Ime1* which is an early regulator of sporulation.

1.1.4.2 Cell differentiation in multicellular organisms

Multicellular organisms consist of different types of differentiated cells with numerous differences in size, components, structure, functions, etc. Cell differentiation is the basic developmental process in multicellular organisms. Moreover the differentiation process occurs continuously in some tissues from adult stem cells during the whole lifetime. But whether one stem cell in the population will further develop to a differentiated progeny type and which one can be stochastic (Eldar & Elowitz, 2010). Some genes are responsible for switches in the differentiation process. Their expression pattern in space and time influences the direction of differentiation and their expression noise might largely affect the differentiation decision (Chalancon *et al*, 2012). An example is the transcription factor Nanog in mouse, which is a key regulator of the pluripotent state of embryonic stem cells (ESCs). The expression level of Nanog is highly variable in the population (Kalmar *et al*, 2009; Ochiai *et al*, 2014). Furthermore, cells with a low Nanog level are more susceptible to enter differentiation whereas a high Nanog level stabilizes the pluripotent state (Abranches *et al*, 2014). Another example is provided by Raj *et al.* (2010) who found that embryonic cells of *Caenorhabditis elegans skn-1* mutants exhibit variable expression of *end-1* whose expression level determines the differentiation decision. Thus only cells where its expression level exceeds a precise threshold undergo the differentiation process whereas the others keep their original state, resulting in heterogeneous cell fates in the population. In this network, the feedback loops might act to decrease expression noise and ensure reproducible cell fate decisions.

1.1.4.3 Diseases

1.1.4.3.1 Cancer

Cancer cells can show large heterogeneity in their response to drugs. Some of them can rapidly develop drug resistance, whereas others die. Non-genetic heterogeneity, especially due to gene expression noise, can be responsible for these differences. Indeed noise in the expression of key genes plays an important role because it can produce a subpopulation pre-resistant to some drugs (Miller-Jensen *et al*, 2011). Cohen *et al.* (2008) showed that, treated with chemotherapy drugs such as camptothecin, individual human H1299 lung carcinoma cells with high expression level of certain proteins such as DDX5 and RFC1 can survive when cells with low levels die. Similar results are also observed during the process of apoptosis induced by TRAIL (tumor necrosis factor (TNF)-related apoptosis-inducing ligand) in Hela cells (Spencer *et al*, 2009). Moreover, the heterogeneity in signaling pathways among the population is also well correlated to drug resistance (Singh *et al*, 2010). Levin *et al.* (2011) showed that the response to type-I interferons of cancer cells is heterogeneous and depends on the expression level of its receptors. Only cells with a high number of receptors carry out the anti-proliferative response. Finally, Sharma *et al.* (2010) showed that the *de novo* drug resistance of lung cancer cells is reversible. The appearance of resistance is due to altered chromatin configurations arising trough noise in the expression of chromatin modifying genes.

Noise-mediated heterogeneity is also observed during cancer cell proliferation. For instance, a small proportion of melanoma cells harboring high expression level of H3K4 demethylase JARID1B have long cycling time but highly proliferating progeny (Roesch *et al*, 2010). JARID1B is important for continuous growth but is not a prerequisite for the tumor initiation *in vivo*. The sorted subpopulations with different levels of JARID1B recover the heterogeneous distribution after several days of growth in conventional medium, indicating that switches between alternative states of cancer cells might be a basic procedure for tumor maintenance.

1.1.4.3.2 HIV latency

The human immunodeficiency virus (HIV) has a latent period after infection. During this period, transcription of the viral genes is silenced and there is no viremia in patients. But these proviruses are also unaffected by normal antiretroviral therapies which only target active viruses. Latent viruses can be reactivated in individual cells what leads to viremia rebound. Thus, the latent state of these viruses forms the main barrier for HIV cure (Miller-Jensen *et al*, 2011). During HIV infection, the transcriptional activator Tat is essential for the decision between productive infection or latency. Tat exhibits a very

noisy expression pattern that leads to bifurcation phenotypes in a clonal population. These heterogeneous subpopulations might enter different stages (Weinberger *et al*, 2005, 2008). Moreover, Dar *et al.* (2014) discovered chemicals that affect the expression noise of HIV genes by screening thousands of compounds and proposed a new effective therapy approach that combines conventional drugs and drugs acting on noise.

1.2 Gene expression noise and adaptation

In this thesis, we mainly focus on the relationships between gene expression noise and stress adaptation. Moreover, a *Saccharomyces cerevisiae* industrial strain was chosen as a model to study some of our hypotheses. Thus in this section, we will first discuss the general mechanisms of stress adaptation in industrial yeasts, then we will review the known or possible relationships between gene expression noise and adaptation in yeast, and finally we will give our working hypothesis.

1.2.1 Stress adaptation in industrial yeasts

Industrial *S. cerevisiae* strains provide a good model to study molecular adaptation to changing environments. They have been selected for rapid fermentations and are specifically adapted to the stressful conditions of fermentation, characterized by high sugar content, high alcohol content, low pH, presence of sulfites, limiting amount of nitrogen, lipid and vitamins, anaerobiosis, and other environmental stresses. While they are genetically highly related to their laboratory counterpart, the genetic basis of their technological properties as compared to laboratory yeast strains that are inefficient under these fermentation conditions are still largely unknown. Genome-wide approaches have received a strong interest in the recent years to address the question of the adaptation of industrial yeasts to these specific conditions (Dunn *et al*, 2012, 2005; Novo *et al*, 2009; Ambroset *et al*, 2011; Dequin & Casaregola, 2011; Salinas *et al*, 2012).

1.2.1.1 Polyploidy and aneuploidy

In general, yeast have a stable karyotype, either haploid or diploid, under optimal conditions (Nishant *et al*, 2010). During the fermentation process, yeast strains are predominantly diploid (Carreto

et al, 2008), while polyploid or aneuploid isolates are also common (Storchova, 2014). Legras *et al.* (2007) have shown the remarkable existence of tetraploidy and aneuploidy in industrial strains by exploring the biodiversity of 12 microsatellite marker sites in a collection of 651 strains from 56 different geographical origins, worldwide. Tetraploid strains can arise by duplication of identical genomes or by fusion of two or more cells of same or closely related species (Albertin *et al*, 2009). Aneuploidy can improve adaptation. Selmecki *et al.* (2015) showed that polyploid cells exhibit high rate of beneficial mutations, and therefore can rapidly adapt to new environments. These cells also exhibit high level of chromosomal instability (CIN) that can rapidly lead to aneuploidy by losing certain chromosomes. Aneuploid cells can also arise from unequal segregation during cell division (Storchova, 2014).

Aneuploid cells exhibit higher growth rate than euploid cells with the same genetic background (Pavelka *et al*, 2010). Chen *et al.* (2012) found that some *S. cerevisiae* clones surviving when Hsp90 is inhibited have gained additional copies of chromosome XV. Nevertheless, Yona *et al* (2012) showed that aneuploidy is just a transient step during adaptation in long evolutionary experiments: the gained chromosomes were further eliminated because some *de novo* mutations with the same adaptive effects were then fixed. The adaptive advantages of aneuploidy is probably due to the increased gene expression levels with gained chromosomes or the decreased expression levels upon chromosome loss (Storchova, 2014).

1.2.1.2 Chromosomal rearrangement

Yeast chromosomes can be rearranged through translocations, deletions and amplifications, the so called gross chromosomal rearrangements (GCR), that can lead to high level of chromosomal length polymorphism (Dequin & Casaregola, 2011). The GCR events are mainly mediated by ectopic recombination between repeated Ty sequences (retrotransposons) (Carreto *et al*, 2008), duplicated genes or repeated telomeric sequences (Aksenova *et al*, 2013). Translocations and insertions of chromosomal regions are found in many yeast strains, such as the wine strain EC1118 (Novo *et al*, 2009), the clinical isolate YJM789 (Wei *et al*, 2007) and flor yeasts (Infante *et al*, 2003), showing the potential advantages of GCR in natural adaptation while the underlying mechanisms are still unclear. One possible assumption is that GCR can modulate the expression pattern or the number of genes on the modified regions and thereby confer certain adaptive advantages. For example, the translocation between chromosomes VIII and XV in wine yeasts, which increases the expression of *SSU1*, can confer sulfite resistance (Pérez-Ortín *et al*, 2002). Also, the segmental duplication on chromosome VII and VIII,

which increases the number of *CUP1* and *CUP2* copies, can increase copper resistance in EC strains (Chang *et al*, 2013). Otherwise chromosomal rearrangements play an important role during the evolution process of interspecific hybrid cells to gain adaptive advantages (Zheng *et al*, 2014) (see below).

1.2.1.3 Copy number variations

Copy number variations (CNV) generated by gene amplification or gene deletion also play an important role during the yeast adaptation process to stressful environments (Dequin & Casaregola, 2011). Studies on the genetic biodiversity of large yeast collections by different methods have shown the widespread existence of CNV in industrial yeasts. Most variations are found in transporter genes or genes located in telomeric chromosomal regions (Legras *et al*, 2007; Carreto *et al*, 2008; Dunn *et al*, 2005, 2012). Recent evolutionary experiments demonstrated the advantages of CNV. Zhang *et al.* (2013a) showed that CNV of *SFA1* or *CUP1* mediated by ectopic recombination between Ty elements on chromosome V confer notable resistance to formaldehyde or copper, respectively. Moreover, CNV of nitrogen transporters including *PUT4*, *DUR3* and *DAL4* are the predominant evolutionary solution during the long-term nitrogen limitation (Hong and Gresham (2014)). A well-documented example is that different numbers of *CUP1* copies that vary from 1 to 18 in different strains confer different copper resistance levels (Zhao *et al*, 2014). This number can be increased under copper treatment (Adamo *et al*, 2012). Finally, CNV can arise from chromosome rearrangement or unequal crossover between tandem repeats during cell division (Zhao *et al*, 2014; Chang *et al*, 2013).

1.2.1.4 Gene introgression

Gene introgression is the phenomenon by which an organism gains exogenous DNA from other species by horizontal gene transfer (HGT) (Dequin & Casaregola, 2011). Many researchers have shown the existence of gene introgression in various yeast strains, indicating its probable selective advantages during natural and domestic evolution (Novo *et al*, 2009; Zhang *et al*, 2010; Muller & McCusker, 2009; Dunn *et al*, 2012). For example, the sequencing of the industrial wine yeast EC1118 (Novo *et al*, 2009) revealed large chromosomal insertions containing novel genes related to nitrogen and carbon metabolism coming from different species, even non-*Saccharomyces* species. The function of some of these genes has been documented, such as *FSY1*, encoding a high-affinity fructose symporter (Galeote *et al*, 2010), and the fungal oligopeptide transporters (FOT) genes (Damon *et al*, 2011). Thus these

genes could enlarge the range of potential carbon and nitrogen resources for wine yeast and thereby ensure adaptive advantages during fermentative process, where nutrient resources are limited. Finally, interspecific hybrid might contribute to the gene introgression (Dunn *et al*, 2013; Zheng *et al*, 2014).

1.2.1.5 Sequence polymorphism

Large number of sequence polymorphisms (Single-Nucleotide Polymorphisms (SNPs) or small Insertions/deletions (Indels)) have been observed among different yeast collections (Liti *et al*, 2009; Schacherer *et al*, 2009). The consequences of some polymorphisms have also been characterized. For example, the two mutations observed in *FLO11* in flor yeast strains located in the promoter and the ORF can increase the expression level and the adhesion ability among cells respectively (Fidalgo *et al*, 2006). Otherwise, QTL analyses that use statistical methods to link phenotypic information (trait measurements) and genetic information (molecular markers like SNPs) have been applied to study the effects of sequence polymorphisms on fermentation ability for instance. Large numbers of QTLs and eQTLs have been identified in different positions through several combination of different strains (Ambroset *et al*, 2011; Salinas *et al*, 2012; Brion *et al*, 2013). For example, the *ABZ1* allele of some wine yeasts can increase the fermentation rate by facilitating nitrogen utilization.

In this work, we used the oenological strain Lalvin EC1118 as a model to study our hypotheses. This strain, also known as ‘‘Prise de mousse’’, is a *S. cerevisiae* wine strain isolated in Champagne (France) and manufactured by Lallemand Inc. This strain has been sequenced by Novo *et al.* (2009). The map is 11.7 Mb long with 31 scaffolds, corresponding to 96.7% of the S288c genome with 46,825 SNPs. 5,728 ORFs have been predicted, of which 5,685 were mapped to S288c. Three introgression regions have been identified with 34 novel genes which are not present in the S288C genome. They are involved in key wine fermentation functions. EC1118 is a diploid wine strain, whereas the haploid derivatives are named as 59A or 59α. The sequencing has been performed on the derivative 59A.

1.2.1.6 Interspecific hybrid

As shown in Figure 18, interspecific hybridization is an common breeding process during the domestication of industrial *Saccharomyces* yeasts (Dequin & Casaregola, 2011). Hybridization can highly increase the genome complexity and by this way, cells can rapidly acquire new genes with essential functions, thus conferring notable advantages under selective pressure (Morales & Dujon, 2012). Moreover, the heterosis, which is the improved or increased function of any biological feature in

a hybrid offspring gained from the combination of heterogeneous chromosome sets, may also provide certain advantages (Comai, 2005). At the same time, hybridization can serve as an important source of chromosomal rearrangements and gene introgression (Storchova, 2014). Recently, the feasibility of yeast breeding by artificial interspecific hybridization has been demonstrated to improve the fermentation ability of some yeast strains (Dunn *et al*, 2013; Zheng *et al*, 2014).

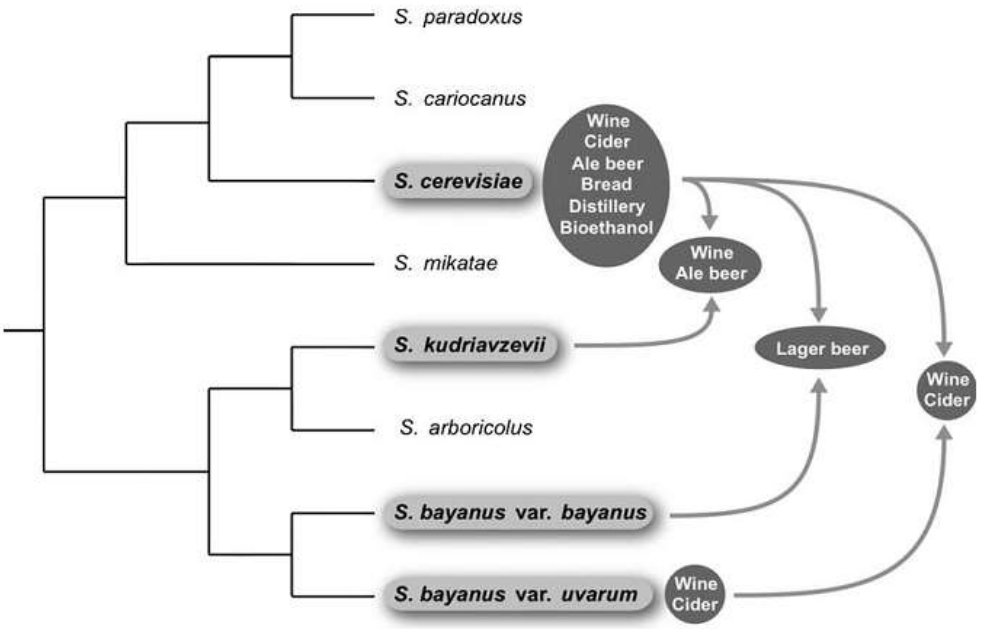


Figure 18: Schematic diagram of the phylogenetic relationships between the *Saccharomyces* species and their industrial specialization. The species involved in industrial processes and/or in hybrids are boxed in light grey. The products of industrial processes involving the hybrids and non-hybrids are boxed in dark grey. The arrows correspond to hybrids (Dequin & Casaregola, 2011).

1.2.2 Noise and stress adaptation

The consequences of noise depend on the gene function. Some researchers showed that housekeeping genes exhibit less variability (noise minimization) (Lehner, 2008; Fraser *et al*, 2004; Monteoliva *et al*, 2013; Newman *et al*, 2006) whereas stress-related genes exhibit higher noise in yeast (Newman *et al*, 2006; Zhang *et al*, 2009; Venancio *et al*, 2010). Thus expression noise seems to be a special property of a gene and might be an evolvable trait, especially in stressful environments.

1.2.2.1 Constant stress

Under constant stress, cell survival mainly depends on the mean expression level of some key genes. For example, cells expressing enough antibiotic resistance proteins are selected in the presence of this antibiotic. Thus the selective advantage of expression noise in constant stress depends on the

position of the expression threshold allowing survival. As shown in Figure 19, if the threshold is lower than the mean value, low-noise genotypes are favoured, and vice versa, high noise is favourable when the threshold is higher than the mean value (Eldar & Elowitz, 2010).

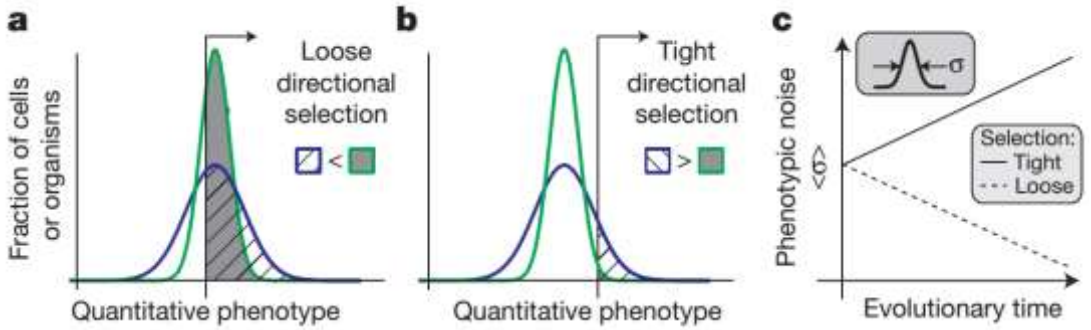


Figure 19: Selection pressure on noise (Eldar & Elowitz, 2010) **a**: Directional selection for values of the trait above a threshold (black line and arrow) can lead to reduced noise when the threshold is low. Thus, the noisier distribution (blue line) has less area above the threshold (cross-hatch) than a less noisy distribution (green line, grey shading) with the same mean. **b**: By contrast, when selection is tighter, the noisier distribution is favoured, as shown by the larger above-threshold area under the blue distribution compared to the green distribution. **c**: Over evolutionary timescales, noise (σ, defined as the standard variation of the distribution) would thus be expected to increase under tight selection and decrease under weak selection.

This phenomenon has been clarified and experimentally tested with an artificial yeast regulation network using the *GAL1* promoter and *zeoR* (conferring zeocin resistance) as an indicator gene expressed with different levels of noise but the same mean in different strains (Blake *et al*, 2006). In weak zeocin concentrations, the clone exhibiting low noise grows better, wheheas the clone exhibiting high noise is the only one surviving in high zeocin concentrations (Figure 20).

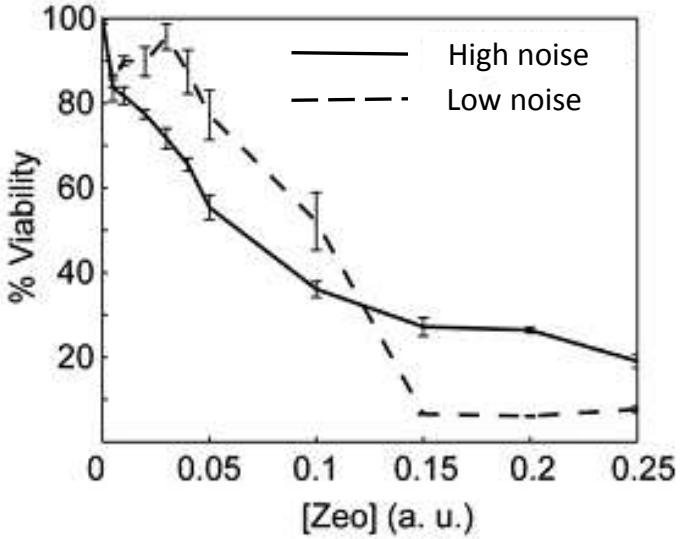


Figure 20: Viability (percentage of growth with zeocin compared to growth without zeocin) of clones with similar mean expression level but different noise of *zeoR* (Blake *et al*, 2006).

1.2.2.2 Fluctuating environments

On one hand, increasing the expression level of key genes is beneficial in specific constant stressful conditions, but it can be harmful when the environmental conditions vary. On the other hand, when key genes are noisy, an isogenic population can contain cells with heterogeneous phenotypes and some 'persister' cells (representing a small part of the population that resist to certain stress levels) can emerge, thereby increasing the global adaptation to fluctuating environments (Chalancon *et al*, 2012). Several researchers have demonstrated the fact that variable expression of certain genes can confer resistance and can be an advantage in fluctuating conditions. For example, the heterogeneous expression of *GTS1* is responsible for the oscillations of glutathione (GSH) content in yeast and confers heterogeneous resistance to cadmium (Cd) and H_2O_2 (Smith *et al*, 2007). Furthermore, the clone with higher *GTS1* expression noise showed a significant selective advantage over the more stable one at same mean expression level during competition in fluctuating experiments (medium successively changed between YPD and YPD + Cd).

Phenotypic heterogeneity has been shown to predominantly exist in wild yeast isolates, indicating the large selection pressure acting on it (Holland *et al*, 2014). Strategies increasing heterogeneity are also called bet-hedging strategies where geometric mean fitness is maximized at the expense of arithmetic mean fitness. These strategies increase the potential adaptation to certain environments in subpopulations whereas decreasing the mean adaptive effects in the current environment (Viney & Reece, 2013). Several researchers have shown that noisy or variable expression of certain genes play a role in bet-hedging strategies (Veening *et al*, 2008; Levy *et al*, 2012; Silva-Rocha & de Lorenzo, 2012).

Moreover, evolutionary experiments can selected for clones with increased noise. Ito *et al.* (2009) first confirmed the possibility to make noise evolve in *E. coli*. After several rounds of constant selection, they found both clones with elevated mean expression level and clones with elevated expression noise but a similar mean. Recently, similar results were observed in yeast (New *et al*, 2014). They grow yeast cells in recurrent YPD and YPGal media with different interval time for a long period. The authors found two types of mutated clones during the evolutionary experiments: specialist clones with strict catabolite repression (low expression noise of genes related to alternative carbon sources), high growth rates in glucose but long lag phases when cells are placed in a different carbon source; and generalist clones with loose catabolite repression (high expression noise of genes related to alternative carbon sources), slow growth in glucose but short lag phases when cells are placed in a different carbon source. It is notable that selection with long interval periods between different carbon sources favours

the specialist phenotype whereas selection with short interval periods favours the generalist phenotype.

1.2.2.3 Instability switches

As we mentioned before, bimodal expression is an extreme example of high noise expression and is widely observed in different organisms (Acar *et al*, 2005; Blake *et al*, 2006; Shalek *et al*, 2013; Silva-Rocha & de Lorenzo, 2012). This expression pattern often results in bistable cell states in the population, each of them conferring advantages in certain conditions. The switching rate between different states can affect stress adaptation. Acar *et al.* (2008) created an artificial gene network with the *GAL1* promoter whose expression pattern (its switching rate between on and off states) can be modified by different inducing conditions. Using *URA3* as indicator gene, they measured the growth rate of clones with different switching rates under different fluctuating environments (successively changed between medium without uracil and medium with uracil and 5-FOA). They found that the fast switching clone is favoured in rapidly fluctuating environments whereas the slow switching clone grows better in environments that rarely change. Moreover, an evolutionary experiment with *Pseudomonas fluorescens* selecting for clones switching their morphology identified isolates with rapid phenotypic switching (Beaumont *et al*, 2009), indicating that modification of the switching rate between bistable or even multistable states is an important evolutionary strategy for stress adaptation.

1.2.3 Our hypothesis

Our hypothesis is based on two fundamental assumptions. First, industrial yeast strains are challenged by various inconstant stresses and at the same time they are well adapted to these fluctuating conditions with good global fitness compared to laboratory yeasts. This adaptation is gained through various molecular mechanisms during domestication. Second, gene expression noise of stress-related genes leads to phenotypic heterogeneity that increases the fitness of the population and is evolvable. Thus, we assume that fermentative strains might exhibit higher level of noise in key genes related to stress adaptation than laboratory strains.

The identification of a panel of genes harboring different noise levels in different *S. cerevisiae* strains would constitute an original result and represent an innovative strategy for the study of genetic determinants of stress adaptation and tolerance. More generally, our main aim is to determine the role of stochastic variation of gene expression in yeast adaptation through its impact on genetic and non-

genetic variability. It will improve our knowledge of domesticated yeasts strains, along with our basic knowledge of eukaryotic cell adaptation to fluctuating environments.

1.3 Gene expression noise and genome stability

As we mentioned in the previous chapter on the adaptation of industrial yeasts (section 1.2.1), their genomes display high plasticity, especially in stressful fermentation conditions. Some works showed that acquisition of chromosome rearrangements in yeast genomes can be influenced by exposure to environmental stress. Cells exposed to stress conditions like fermentation in high specific gravity wort or at higher than normal temperatures, undergo gross chromosomal rearrangements, small deletions and regional amplifications whereas 'standard' fermentation conditions did not generate significant changes (James *et al*, 2008). High genetic instability and loss of heterozygosity (LOH) were often observed in natural wine yeasts (Ambrona *et al*, 2005; Carreto *et al*, 2008; Dunn *et al*, 2005, 2012, 2013), indicating that the maintenance of an unbalanced chromosome set is advantageous in such environments. Their complex genomes allow loss, duplication, inversion or translocation of genetic material which form at a faster rate than other types of mutations (Hastings *et al*, 2009). Here we will describe our assumption that expression noise of genes related to genome stability could favour rapid genome modifications and therefore could increase the fitness to varying environments.

The genome can be modified through different mechanisms. For example, chromosome rearrangements can arise from homologous recombination (HR) or non-homologous end joining (NHEJ) induced by double-strand breaks (DSB) repair after DNA damage. Otherwise, mutations and Indels can arise from DNA replication errors. These modifications can be repaired by the mismatch repair pathway. In this work, we mainly focused on the homologous recombination pathway.

1.3.1 Homologous recombination

1.3.1.1 HR pathways

There are several possible pathways in HR. Here we will focus on the Rad51-dependent pathway. The HR process begins from a DSB occasionally occurring in a cell (Figure 21A). Several replication proteins A (RPA) bind to the damaged DNA ends which is resected as single-stranded DNA (ssDNA)

(Figure 21B-C) (Holthausen *et al*, 2010). RPA can protect the unstable ssDNA from further damage. Then, with the help of Rad52, several recombinase proteins Rad51 replace RPA on the ssDNA (Sung, 1997). Rad51 helps the search for a homologous fragment and facilitate the invasion of the ssDNA overhang into the homologous double-stranded DNA (dsDNA) sequence (Figure 21E). The specific DNA structure after invasion is referred to as a displacement-loop (D-loop) (Arai *et al*, 2011). Once the formation of D-loop is formed, a DNA polymerase can extend the ssDNA end using the homologous DNA as a template, thereby repairing the DSB (Karpenshif & Bernstein, 2012).

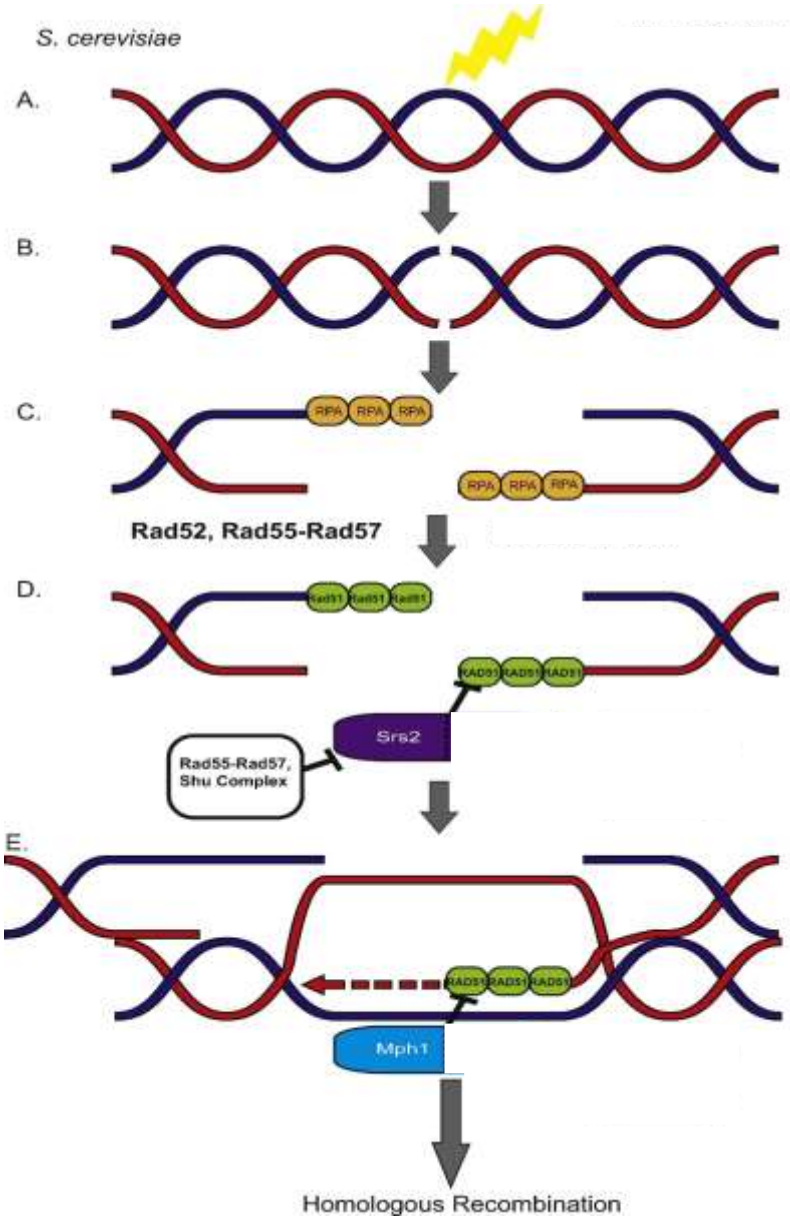


Figure 21: Rad51-dependent homologous recombination (Karpenshif & Bernstein, 2012)

After synthesis, there are two different pathways to create the final repaired products: synthesis dependent strand annealing (SDSA, Figure 22A) and double-Holliday junction (dHJ, Figure 22B)

(Karpenshif & Bernstein, 2012). SDSA creates non-crossover products whereas dHJs can result in crossover or non-crossover products (Figure 22). Non-crossover products repair the DSB without modifying the chromosome structure. Otherwise crossover products repair the DSB and introduce chromosome exchanges at the same time and create genome rearrangements.

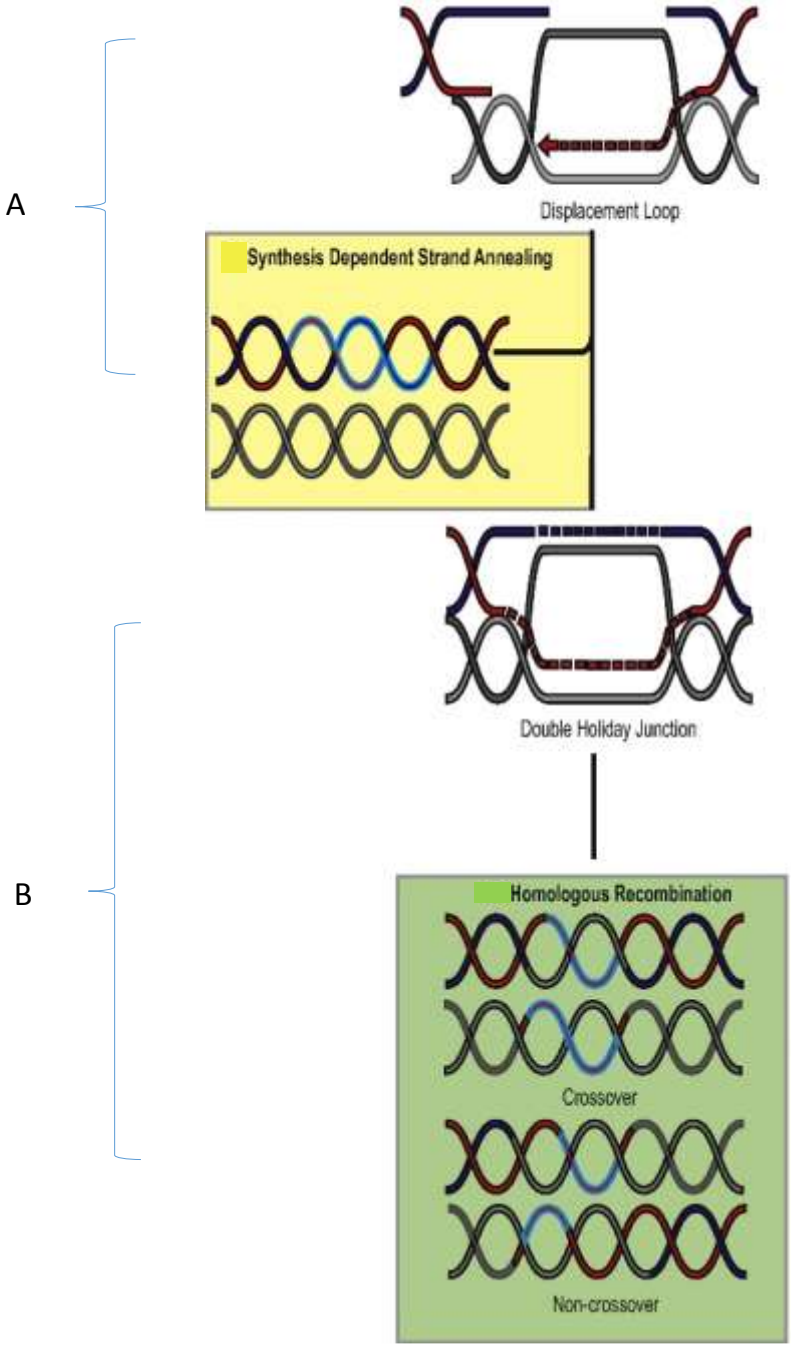


Figure 22: Synthesis dependent strand annealing and double-Holliday junction (Karpenshif & Bernstein, 2012) **A**: Synthesis dependent strand annealing (SDSA) is a form of non-crossover homologous recombination, in that it uses a homologous sequence as a template for DNA repair. However, this process does not involve second end capture of the DNA end. After end processing, one strand is synthesized using a homologous template and then is re-ligated to the broken end. The newly synthesized strand is then used as a template and base pairs to the complimentary sequence, consequently resolving the break. **B**: Homologous recombination uses a homologous template for repair. After the DNA ends are resected, the

ssDNA 3' overhang invades a homologous sequence and restores any missing information at the break site. The second end of the DNA is captured resulting in a double-Holliday junction that can be resolved into a crossover or non-crossover product depending upon where the junctions are cut. **Legend**: Red–Blue double-helix is the broken molecule of DNA. Grey DNA is the homologous sequence. Blue highlighted segments of DNA are newly synthesized regions.

1.3.1.2 Rad52 and HR

Rad52 has two major functions during the HR process: it harbors a recombination mediator activity that helps Rad51 to displace RPA, and binds to the ssDNA ends to use its ssDNA annealing activity that helps the annealing process either in the SDSA pathway or in the dHJ pathway (Xiong *et al*, 2014). The *rad52* deletion in *S.cerevisiae* partially stabilizes the karyotype by suppressing recombination at telomeric and subtelomeric regions and reducing the rate of changes in chromosomal size by 30% (Carro *et al*, 2003). It is found that defect of *RAD52* can also induce deletion rate in the genome (Fritsch *et al*, 2009). By contrast, overexpression of *RAD52* do not increase the HR frequency comparing to wild type (Dornfeld & Livingston, 1991; Paffett *et al*, 2005).

1.3.1.3 Rad27 and HR

RAD27 encodes a flap endonuclease acting with Dna2 to cleave flap structures between Okazaki fragments during DNA replication (Xiong *et al*, 2014). The *RAD27* deletion causes the formation of double-stranded breaks during replication resulting in enhanced spontaneous recombination, and repetitive DNA instability to bypass the replication defect (Debrauwère *et al*, 2001). It is shown that the *rad27* null strains exhibit high recombination rate (Sun *et al*, 2003; Ayyagari *et al*, 2002; Xie *et al*, 2001). Especially the *RAD27* deletion favours the recombination between di- and trinucleotide repeats (Xie *et al*, 2001).

1.3.1.4 Measurement of HR frequency

There are several methods to measure the HR frequency. For example we can use the recombination reporter plasmid pBYA819 (Dornfeld & Livingston, 1991) which contains two HIS3 gene with different mutations that can be recovered by HR. The rate of HIS3 restoration indicates the HR frequency. Another method take advantages of the strain FW1259 (Selva *et al*, 1995). The defection of *SPT15* in FW1259 creates LYS autotrophy. Two copy of *SPT15* with different mutations are inserted to the genome, and then the recombination between them will generate intact *SPT15* and thus LYS

prototrophy. The rate of LYS prototrophy indicates the HR frequency. Verstrepen *et al.* (2005) created a series of strains allowing measurement of the HR frequency. They took advantage of the tandem repeats (TRs) present inside the *FLO1* gene encoding a cell-surface adhesin. In the wild type gene (in the laboratory strain S288c), there are 18 TRs of around 100 nt, separated by less conserved 45 nt sequences. A single copy of the *URA3* gene was inserted among these repeats at several locations. These strains were first grown in URA$^-$ medium to avoid proliferation of cells lacking *URA3*, and then a precise number of cells were spread on 5-FOA plate. Finally, the number of clones appearing on the plate divided by the total number spread indicated the HR frequency. In this study we will use the strain KV133, where *URA3* is in the middle of the repeats, which is the location conferring the highest HR frequency (Figure 23). It has been shown that the deletion of *RAD52* in this strain results in decreased HR frequency whereas the deletion of *RAD27* increases the frequency.

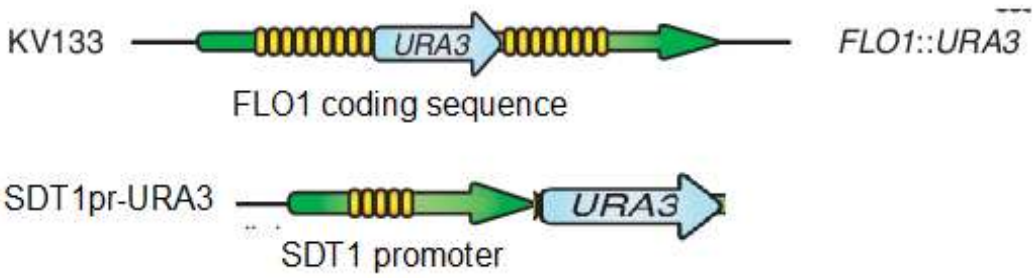


Figure 23: Structure of the genomic reporter used to measure HR frequency. *FLO1* has tandem repeats inside its coding sequence. In the strain KV133, the *URA3* gene was inserted in the middle of the repeats (modified from Verstrepen *et al.* (2005)). *SDT1* has tandem repeats in its promoter. In the SDT1pr-URA strain, we inserted *URA3* just downstream of the *SDT1* promoter (modified from Vinces et al. (2009)).

Another strain we will use takes advantage of the TRs in the promoter region of *SDT1* (Figure 23) (Vinces *et al*, 2009). The number of TRs in the promoter can affect the expression level of *URA3* (Vinces *et al*, 2009) (the promoter with 13 repeats exhibited highest expression level). Moreover, the number of TRs in this promoter evolves during evolutionary experiments towards the number giving the maximal expression level in medium without uracil (Vinces *et al*, 2009). Considering that the expression level can be modified by changing TR numbers through HR in the promoter while the intact ORF was kept in the genome, we infer that it can be subjected to fluctuating experiments. In our work, we replaced *SDT1* by *URA3* (see Methods) because we expect an increase of the *URA3* expression in URA$^-$ medium and a decrease of its expression level in 5-FOA containing medium by modification of the number of TRs in the promoter.

1.3.2 Our hypothesis

The expression level of DNA repair and maintenance genes (DRMGs) affects the genome stability by changing the rate of genetic variant generation (RGVG). Thus we can infer that high variability in the expression of DRMGs could favour the acquisition of genetic modifications favourable to population survival and enable fine tuning of the RGVG (Capp, 2010). No study has explored the expression variability of these genes at the single cell level and its potential consequences on the RGVG.

Moreover, industrial yeast populations are exposed to stressful and fluctuating environments and can probably take advantage of a high tunability of this rate. This tuning could be easier with a higher noise level in the expression of DNA repair and maintenance genes. A boarder range of their expression levels could optimize the emergence of cells harboring the 'just right' level of genetic variability. So we assume that these strains might have evolved to exhibit higher variability in the expression of DRMGs than laboratory strains do, because they are exposed to more variable growth conditions. Higher heterogeneity in the expression of these genes in industrial yeasts could explain their ability to rapidly acquire genetic modifications in stressful environments and their genomic instability in non-selective conditions.

Results

2 Connecting Noise to Stress Adaptation

As we mentioned in the introduction, gene expression noise can increase the phenotypic heterogeneity in an isogenic population, and therefore can increase fitness in selective and fluctuating environments. Moreover stress-related genes have been shown to exhibit higher noise than housekeeping genes. Thus, considering that industrial strains are selected through complex fermentation conditions, we assume that the evolutionary strategy that modifies the expression noise of certain genes to increase the global fitness of the population might have occurred in the domestication of these strains. To study this hypothesis, we screened for promoters conferring high expression variability in the sequenced industrial wine *S. cerevisiae* strain EC1118, to compare their sequence with the ones of their counterpart of the laboratory strain S288c. Our aim was to determine if the observed genetic differences generate noise differences between the variants and if these differences are important enough to generate a difference of growth in specific selective environments.

2.1 Results

2.1.1 Screening for high noise promoters

2.1.1.1 Construction of the EC1118 genomic library

The genomic approach used to screen promoters conferring high promoter-mediated noise in the sequenced *S. cerevisiae* strain EC1118 (Novo *et al*, 2009) was based on the method developed initially with *Salmonella typhimurium* (Freed *et al*, 2008). In this modified protocol (see Methods), genomic DNA fragments from the haploid 59A strain (derivative of the winemaking diploid EC1118 strain) were inserted before the start codon of *yEGFP* in a series of three distinct promoterless *yEGFP*-coding vectors (the different plasmids contain zero, one or two additional base(s) between the insertion site and the *yEGFP* start codon). These centromeric plasmids minimized problems of copy number that would contribute to variations in fluorescence levels. The resulting library was transformed in the laboratory strain CEN.PK as we were looking for *cis*-effects on noise in *yEGFP* expression.

While no fluorescence was detected without genomic library, around 4% of the cells transformed with the library were fluorescent before any cell sorting (Figure 24). Fluorescence levels above the auto-fluorescence threshold were spread on at least two logs, showing strong promoter activity of some fragments.

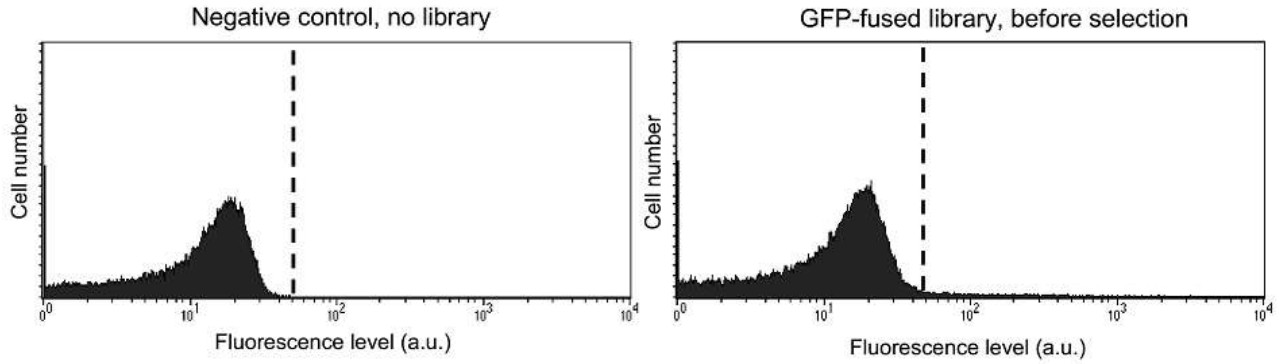


Figure 24: Fluorescence distribution in the control population and the population transformed with the *yEGFP*-fused genomic library

2.1.1.2 Fluctuating selection by FACS

The fluctuating selection method described by Freed *et al.* (2008) enabling enrichment of fragments producing highly variable *yEGFP* expression was then applied. Briefly, seven rounds of cell sorting were performed with alternatively the highest 5% or the lowest 5% fluorescence levels conserved (Figure 25, except for the first round where only the fluorescent cells were sorted). 10^5 cells were sorted at each round, cultured overnight, diluted in the next morning to sort cells again in exponential phase in the afternoon at the same OD at each round.

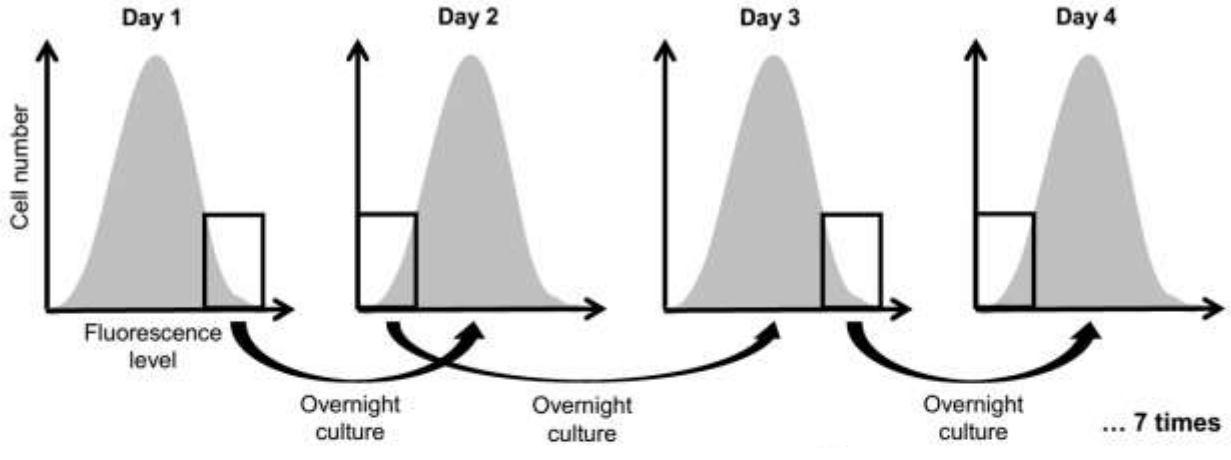


Figure 25: Procedure of fluctuating FACS selection

During the selection process, we chose for convenience to characterize cell-cell variability by the coefficient of variation (CV) (standard deviation divided by the mean). As expected, the selected population exhibited higher CV: it approximately doubled after the selection procedure (Figure 26). The mean expression level was also increased, showing that the library was enriched in fragments with

promoter activity.

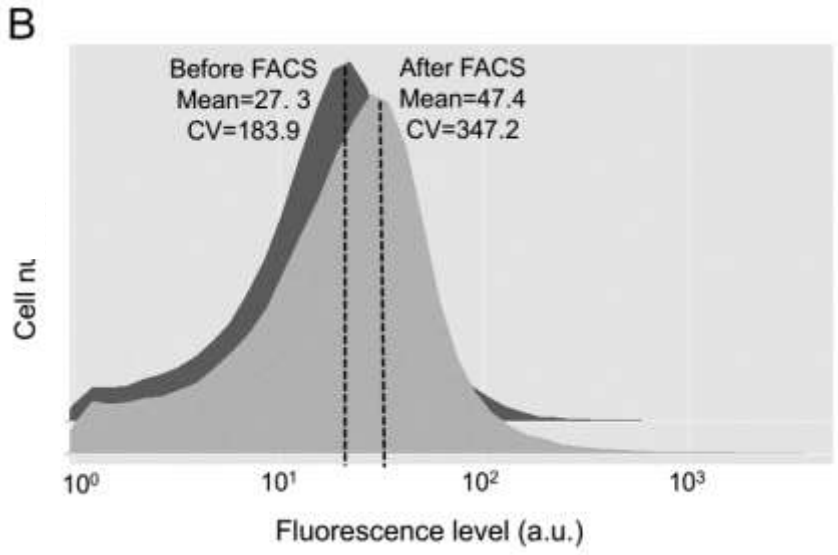


Figure 26: Distribution of the fluorescence levels in the library before and after selection

We also measured the CV of individual clones randomly isolated from the library before or after selection. Among the selected population, more clones showed high CV (p = 0.047, Figure 27). Thus the fluctuating selection efficiently enriched the population in fragments giving high variability in *yEGFP* expression. Nevertheless, clones with noise levels similar to control clones were still present, as was previously observed with the original protocol (Freed *et al*, 2008) and might be explained by aggregation with non-fluorescent cells in the sorting process. Even if this method might not be the ideal way to strongly enrich for increased noise, selecting a smaller percentage of cells at each round could make the enrichment process more efficient.

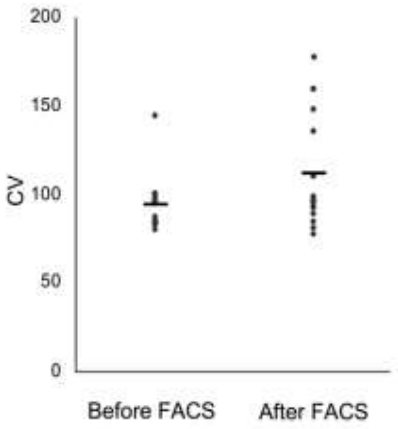


Figure 27: CV of *yEGFP* expression in clones from selected and control populations

2.1.1.3 Isolation of high noise clones

In order to avoid sequencing of "non-noisy" clones, we screened for individual clones with high CV in the enriched library. By setting threshold values on mean and noise, only clones exhibiting highly variable *yEGFP* expression in the population were selected (Figure 28). Expression profiles among these single clones were highly heterogeneous, but they all possessed CV among the highest CV values that are observed in the yeast genome (Newman *et al*, 2006), confirming the efficiency of the screening.

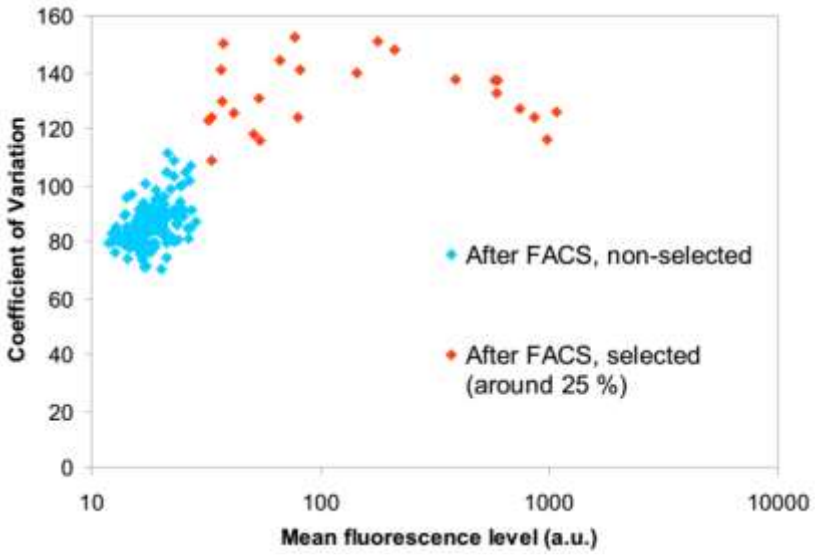


Figure 28: Screening for clones with high CV among the selected population

To confirm that phenotypic noise is a stable property of a clone, we re-isolated on plates and analyzed 2 sub-clones from each selected clone. Ninety-nine clones with high CV have been chosen for further investigations because mean and CV values of their sub-clones were highly reproducible compared to the initial clone (R>0.97 for mean values and R>0.8 for noise values, Appendix I). We also wanted to verify that the clones were mainly dominated by intrinsic noise originating from transcription. The contribution of extrinsic noise in total noise can be decreased by reducing the FSC and SSC gates (Newman *et al*, 2006). Indeed, whereas ungated populations are dominated by extrinsic noise, analysis on a more homogeneous part of the population decreases extrinsic noise either to levels comparable to intrinsic noise or to a level below that of intrinsic noise because the coefficient of variation in *GFP* expression is calculated from a subset of cells similar in size, shape, and cellular complexity. Thus, to reduce in our CV measurements the extrinsic noise linked to cell-to-cell variations in global physiological factors, we extracted a subset of cells that were very homogeneous in size and granularity. By measuring fluorescence only on a more homogeneous part of the population (around

50% of the cells), CV were reduced by around 28.5% (Table 1). We concluded that promoter-mediated noise was the main contributor to the elevated CV observed in these clones.

Table 1: Reduction of noise by sub-gating on a more homogeneous part of the population in selected clones

Clone	Number of cells	Mean	CV	Number of cells-small gate	Mean-small gate	CV-small gate	CV reduction (%)
28-G4-01	99060	110,15	193,33	39475	120,2	128,64	0,335
28-H3-02	98935	224,6	175,75	54986	151,47	108,87	0,381
30-BE11-01	99036	1030,76	150,58	55414	718,37	117,55	0,219
30-BE11-02	99238	1036,96	148,86	56247	739,07	118,23	0,206

2.1.2 Genetic differences between noisy promoters from EC1118 and their counterpart in S288c

The fragments driving *yEGFP* expression were successfully sequenced in 97 clones (Table 2) and 95 were localized by mapping reads to the S288c reference genome (Appendix II). First, around 33% of the inserted fragments were found at least 2 times, sometimes with different end points. These fragments with different ends were independently selected and reinforced the validation of the fluctuating selection. Second, the mean length of the fragments was around 650bp. Most of them were fully sequenced by a single round of Sanger sequencing starting from 75 bp downstream the start codon of *yEGFP*. This mean length seemed low compared to the range of size selected to construct the genomic library (500 to 3000bp) and probably reflected the preference for smaller fragments in the cloning process. (Nevertheless this mean was slightly under-estimated because a small minority of longer fragments has not been fully sequenced.) Third, the majority of fragments corresponded to promoter sequences (Appendix II). A total of 50 distinct promoters were found (Table 2) (a fragment is considered as a promoter if its last base is at less than 350bp from the start codon of the downstream ORF). Among these fragments with known promoter sequences, 27 contained the promoter only, with 4 to 350 bp lacking before the start codon of the corresponding ORF, and 23 contained the promoter and a part of the corresponding ORF fused to *yEGFP* (Appendix II).

Table 2: Summary of sequencing results and mapping of genomic fragments from clones exhibiting noisy *yEGFP* expression

Reads and genetic differences	Number
Reads	97
Empty vectors	0
Mapped reads (to S288c)	95

Independent loci	65
Fragments with known promoters	50
Known promoters with SNPs or INDELs compared to S288c	37
Total SNPs and INDELs in known promoters compared to S288c	170

GO processes analysis on these 50 promoters revealed over-representation of genes involved in nitrogen compound transport and anion transport (Table 3). Plasma-membrane transporters are known to show significantly elevated expression noise (Zhang *et al*, 2009). Moreover ion transport and nitrogen compound metabolic processes are among the few GOs identified to have greater-than-expected expression noise in yeast (Zhang *et al*, 2009).

Table 3: Results of GO analysis

GO_ID	GO_term	Cluster frequency	Background frequency	*p*-value	Gene(s) annotated to the term
6820	anion transport	7 out of 50 genes, 14.0%	139 out of 7167 background genes, 1.9%	0.02	*PHO88*/*AGP2*/*GNP1*/*CAN1*/*DNF1* /*PHO86*/*MCH5*/*YOR306C*
71705	nitrogen compound transport	8 out of 50 genes, 16.0%	214 out of 7167 background genes, 3.0%	0.04	*AGP2*/*GNP1*/*CAN1*/*VRG4*/*YSSL2*/ /*MCH5*/*NEW1*/*YPL226W*

We compared these EC1118 genomic fragments with their counterpart in S288c. A total of 170 genetic variations were detected in 37 of the 50 fragments containing known promoters (Table 2, details in Appendix III). The remaining 13 fragments did not show any difference between the strains. The variations were mostly SNPs and small Indels ranging from 1 to 7 bp. Only one longer insertion of 21pb was present in a promoter from EC1118. The number of variations greatly varied from promoter to promoter, ranging from one SNP to 15 different variations including SNPs and Indels. Interestingly promoters driving expression of genes involved in stress response (e.g. *HAC1*, *CUP1*) or in diverse transports (e.g. *CAN1*, *GNP1*, *AGP2*) were present in this list. We hypothesized that some of these natural genetic variations could generate differences in noise level in the expression of these genes, and thus could confer an adaptive advantage in specific challenging environments. Eight promoters were particularly interesting because they are related to environmental factors (Table 4). We choose these genes on the basis of their function and not depending on either the nature or the localization of the genetic variations because even the effects of mutations in regions well-known to modify noise strongly depends on promoter context (Hornung *et al*, 2012) and because of the high number of potential binding sites for transcription factors in these promoters.

Table 4 Genes whose promoter variants are studied at the genomic level

Gene	Function
BMH1/YER177W	14-3-3 protein, major isoform, regulates many processes
BMH2/YDR099W	14-3-3 protein, minor isoform, regulates many processes
CUP1/YHR055C	Binds copper, mediates resistance to high concentrations of copper
CAN1/YEL063C	Plasma membrane arginine permease
YCK2/YNL154C	Palmitoylated plasma membrane-bound casein kinase I isoform
HAC1/YFL031W	Transcription factor, regulates the unfolded protein response
GNP1/YDR508C	High-affinity glutamine permease; also transports Leu,Ser,Thr,Cys,Met,Asn
AGP2/YBR132C	Plasma membrane regulator of polyamine and carnitine transport

2.1.3 Noise levels conferred by each variant of selected promoters at the genomic level

We decided to finely compare at the genomic level the expression and noise levels conferred by both variants of these promoters of interest. Indeed, plasmids do not fully recapitulate chromosomal organization and might generate experimental bias. It was also necessary to compare expression profiles from each variant in both strains (BY4720, an auxotrophic derivative of S228c, and 59A, the haploid derivative of EC1118) to distinguish *cis*-effects (promoter variations) on noise from *trans*-effects linked to the genetic background. Indeed, strain effects on noise are known (Ansel *et al*, 2008) and require determining if epistasis is observed in the generation of promoter-mediated noise. Comparing promoter pairs in both strains also required insertion of the variants in strictly the same chromosomal context. We took advantage from an insertion plasmid (pJRL2) previously used to compare expression variability from mutated *PHO5* promoters inserted in the *LEU2* locus (Raser & O'Shea, 2004). By replacing the *PHO5* promoter by our 16 variants, we were able to insert them in the same locus. We chose to clone 1000bp before the start codon, and not only the fragment inserted in the library plasmids. These longer loci sometimes contained more genetic variations between S228c and EC1118 compared to the cloned fragments but did not contain additional gene or promoter.

To be as accurate as possible, we wanted to reduce extrinsic noise even if both variants of the same promoter were always compared in a given genetic background. Thus fluorescence levels were measured in small gates where cells are homogenous in terms of size and granularity (strictly the same gate was used for the different variants of a promoter in a given background). Also, noise was calculated in the next steps by dividing the variance by the squared mean because this value reflects better than the CV how large the standard deviation is compared to the mean expression level (Chalancon *et al*, 2012). Mean expression and noise levels have been measured in optimal growth

conditions and exponential phase at the same OD. Finally, the 59A strain possesses a high aggregation tendency making flow cytometry analysis difficult. Therefore the *amn1Δ* 59A strain was used because Amn1p plays the leading role in this cell aggregation (Li *et al*, 2013).

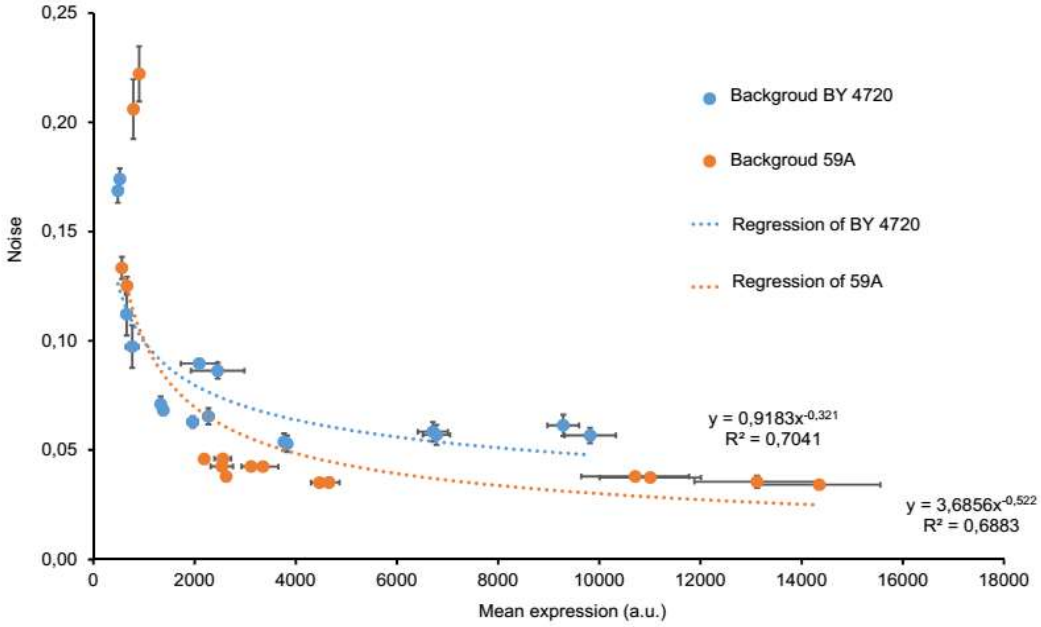


Figure 29: Relationship between noise and mean expression. Blue dots show the data from BY4720 back ground and orange dots show the data from 59A background. Blue dash line is the noise-mean regression curve for BY4720. Orange dash line is the noise-mean regression curve for 59A. The formulae and coefficients of determination are shown beside the curves.

The results obtained were first plot on a graph (Figure 29) with mean expression as x-axis and noise as y-axis. In both background noise decreases as mean increases, which is overserved in many researches. Comparing the two stains, we found that with low expression level (below 1000) 59A shows higher noise level than BY4720 but with high expression level 59A shows lower noise than BY4720. These results indicate the complexity of background effects on noise.

Then the results for the same promoter pair in both strains are presented on the same histogram in Figure 30 but it is worth noting that comparing results for a given variant in the different backgrounds is not possible because the gates where fluorescence levels were measured were not strictly identical for both strains (59A had higher cell size and granularity). Nevertheless, as we compared the consequences on noise of promoter sequences variations and their dependence on the *trans*-background, we chose to show them on the same plot.

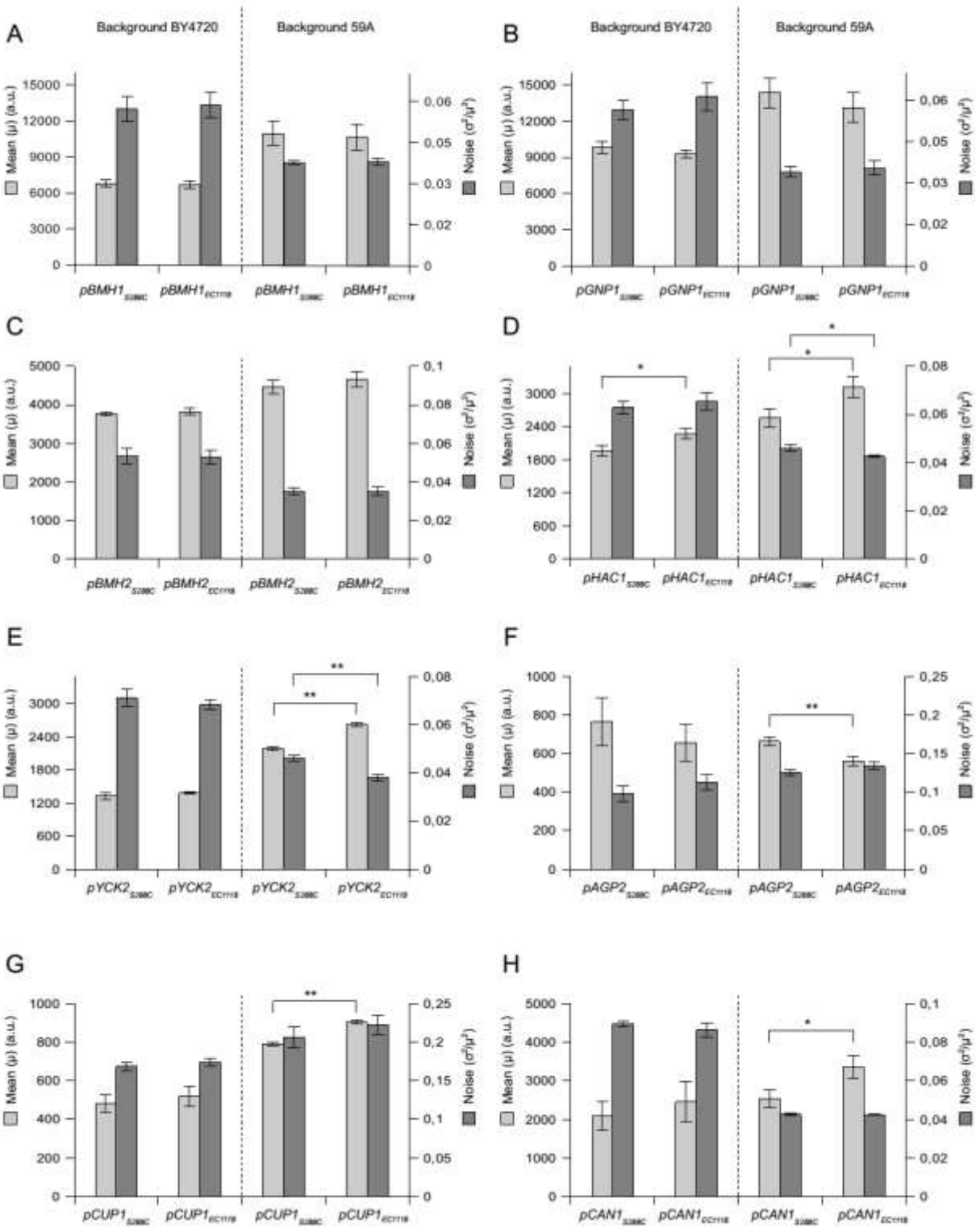


Figure 30: Different behaviours of promoter variants at the genomic level. Dashed lines separate the results obtained in the different backgrounds for a given promoter pair (BY4720 on the left, 59A on the right). (A-H) Results for the *pBMH1*, *pGNP1*, *pBMH2*, *pHAC1*, *pYCK2*, *pAGP2*, *pCUP1* and *pCAN1* promoter variants respectively. Scale for the mean is on the right and scale for the noise is on the left of the histograms. The same nomenclature is used for each histogram: each variant is named by the promoter name with the name of the strain where it comes from in subscript (S288c or EC1118). Results are means of three independent cultures, and error bars are standard deviations. A significant statistical difference between mean or noise levels conferred by the promoter variants in a given genetic background is represented by (*) when $p<0.05$ in T test or (**) when $p<0.01$.

On one hand, both variants of *pBMH1*, *pGNP1* and *pBMH2* gave the same mean expression and noise levels in each strain (Figure 30A-C). The genetic differences between these variants did not generate any effect on fluorescence profiles. On the other hand, the five other promoter pairs exhibited differences in mean and/or noise in at least one backgrounds. $pHAC1_{EC1118}$ conferred higher mean expression compared to the lab variant in both backgrounds (p=0.02 and p=0.018) (Figure 30D) while $pYCK2_{EC1118}$, $pCUP1_{EC1118}$ and $pCAN1_{EC1118}$ gave increased expression only in 59A (p=10^{-4}, p=10^{-4} and p=0.02 respectively) (Figure 30E, G and H), indicating a strain-effect that contributed to reveal the consequences of the genetic variations between the variants. This was also observed with *pAGP2* but here expression was lower with the industrial variant only in 59A (p=0.005) (Figure 30F). A higher expression for one version was associated to lower noise only in 2/6 cases among these promoter pairs while higher mean expression is generally accompanied by decreased expression variability (Bar-Even *et al*, 2006; Newman *et al*, 2006; Hornung *et al*, 2012; Carey *et al*, 2013). Thus laboratory and industrial variants of these promoters differed by their mean expression in at least one background, but enhanced expression did not necessarily produce lower noise so that the genetic variations might generate a different level of noise at similar mean expression level. This hypothesis has been tested by inducing the *pCUP1* variants with different copper concentrations.

2.1.4 Noise levels conferred by the *pCUP1* variants during copper induction

CUP1 is involved in copper detoxification. The *CUP1* copy number is highly correlated to copper resistance (Zhao *et al*, 2014) and strains evolving in copper-rich environments amplify *CUP1* and contain many copies (Adamo *et al*, 2012; Chang *et al*, 2013). Moreover its promoter has been studied in details for instance in terms of transcription factors binding kinetics (Karpova *et al*, 2008) or nucleosome repositioning (Shen *et al*, 2001) during copper induction. We induced the *pCUP1* variants by copper to determine in which conditions mean expression levels were similar with both variants, and if increased noise was conferred by $pCUP1_{EC1118}$ in this case.

2.1.4.1 Induction of pCUP1

2.1.4.1.1 Induction of pCUP1 in YPD medium

We first measured mean expression levels after 1h induction when cells were exposed to

different copper sulfate ($CuSO_4$) concentrations in YPD medium (Figure 31A-B). In both strains and with both variants, induction increased with copper sulfate concentration to reach a plateau in concentrations higher than 20µM $CuSO_4$ in BY4720, and 6µM in 59A. While the variants behaved very similarly in BY4720 (Figure 31A), $pCUP1_{EC1118}$ was more strongly induced than $pCUP1_{S288c}$ in 59A at each concentration (Figure 31B). We chose 5 µM $CuSO_4$ for both variants in BY4720, and 20 µM and 1.5 µM $CuSO_4$ for $pCUP1_{S288c}$ and $pCUP1_{EC1118}$ respectively in 59A to compare noise at similar mean levels and thus independently of the mean. These concentrations avoided experimental bias linked to heterogeneous copper concentrations in the basis medium. We found no differences in BY4720 (p=0.25) (Figure 31C) but $pCUP1_{EC1118}$ was clearly noisier than $pCUP1_{S288c}$ in 59A (p=0.015) (Figure 31D).

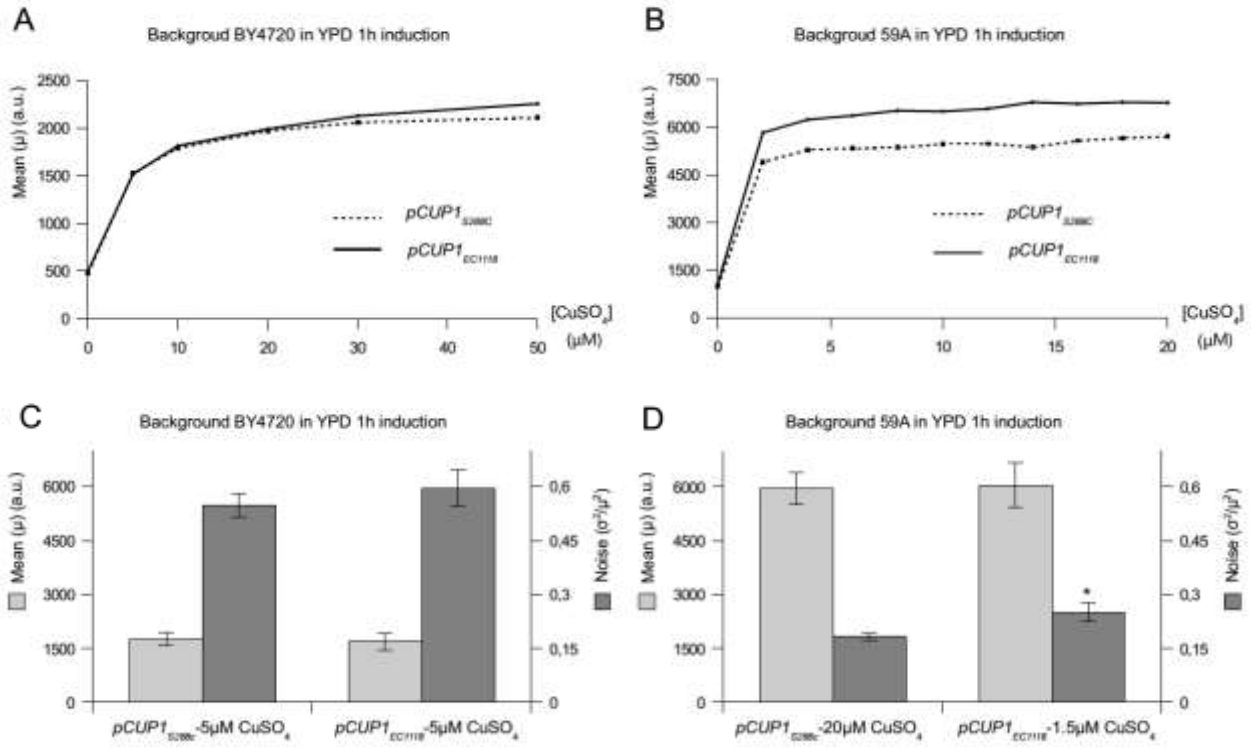


Figure 31: One hour induction of *pCUP1* in BY4720 and 59A in YPD medium. A and B: Induction curves of $pCUP1_{S288c}$ and $pCUP1_{EC1118}$ by $CuSO_4$. C and D: Different induction conditions of $pCUP1_{S288c}$ and $pCUP1_{EC1118}$ giving similar expression levels. The asterisk mark indicates a difference in T-test ($p < 0.05$).

2.1.4.1.2 Induction of pCUP1 in YNB medium

To confirm the differences observed in YPD, one hour induction of the two *pCUP1* variants in 59A was performed in YNB medium. We observed similar induction curve and significant difference in expression noise between the two variants at similar mean expression level (10 µM and 5 µM for $pCUP1_{S288c}$ and $pCUP1_{EC1118}$ respectively, p=0.045, Figure 32A and C). Of note, this difference was also observed after overnight induction (p=0.016, Figure 32B and D). Finally, the difference in noise was still

observed in the same $CuSO_4$ concentration (10µM) in YNB medium after 1h induction (p=0.02, Figure 32E) (mean values are not significantly different in this case while *pCUP1*$_{EC1118}$ conferred a slightly higher mean (*p*=0.12, Figure 32E)). Therefore this natural variant of *pCUP1* exhibited higher promoter-mediated noise in gene expression.

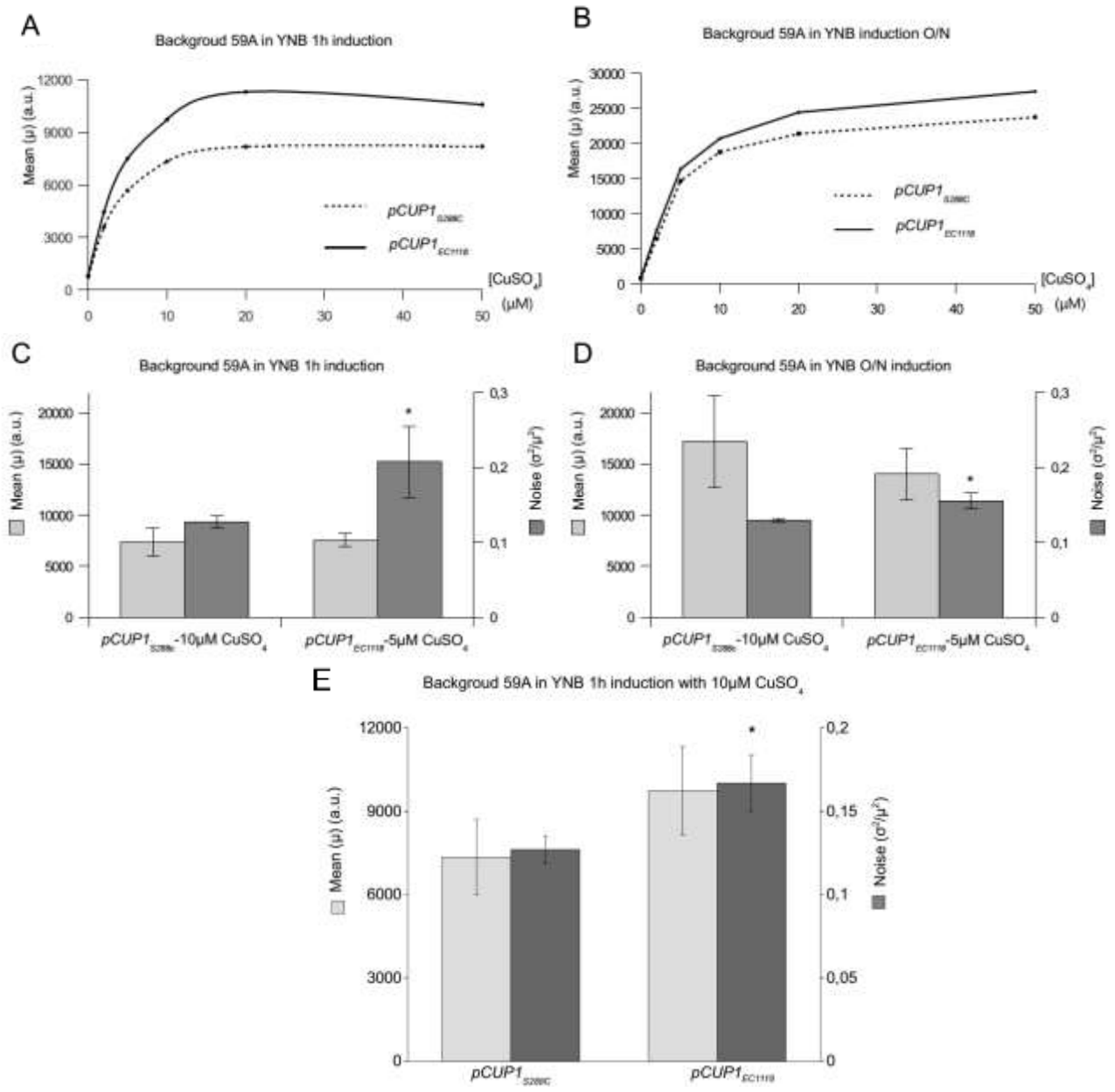


Figure 32: One hour and overnight induction of *pCUP1* in 59A in YNB medium. A and B: Induction curves of *pCUP1*$_{S288c}$ and *pCUP1*$_{EC1118}$ by $CuSO_4$. C , D and E: Different induction conditions of *pCUP1*$_{S288c}$ and *pCUP1*$_{EC1118}$ giving similar expression levels. The asterisk mark indicates a difference in T-test ($p < 0.05$).

2.1.4.1.3 Induction dynamics of pCUP1 in YPD medium

We also searched for induction times giving the same mean expression for both variants in each strain when cells were exposed to 20µM $CuSO_4$ in YPD medium. On one hand, induction curves were

similar for both variants in BY4720 (Figure 33A). It increased in the first hour of induction and then decreased in the next 1.5h. One the other hand, the promoters' behaviour is different in 59A: induction was clearly stronger with $pCUP1_{EC1118}$ (Figure 33B). At similar mean expression levels (t=30 min for $pCUP1_{EC1118}$ and t=90 min for $pCUP1_{S288c}$), $pCUP1_{EC1118}$ was clearly noisier than $pCUP1_{S288c}$ (Figure 33C) (p=0.014). In spite of the slightly lower mean expression from the industrial variant that could favour the observed difference in noise in Figure 32D and Figure 33C, these results were in the same tendency as results in Figure 32D and Figure 32C giving a statistically significant difference in noise with very similar mean levels. Also, increased noise is still observed with $pCUP1_{EC1118}$ at the same copper concentration in YNB medium (Figure 32E), even if its mean value is slightly higher (while mean values are not significantly different), so that $pCUP1_{EC1118}$ appears to be clearly noisier in all conditions. Finally, it is worth noting that the induction factor is far higher in 59A than in BY4720 for both variants (about 5 to 7 and 1.5 respectively in 20µM CuSO4).

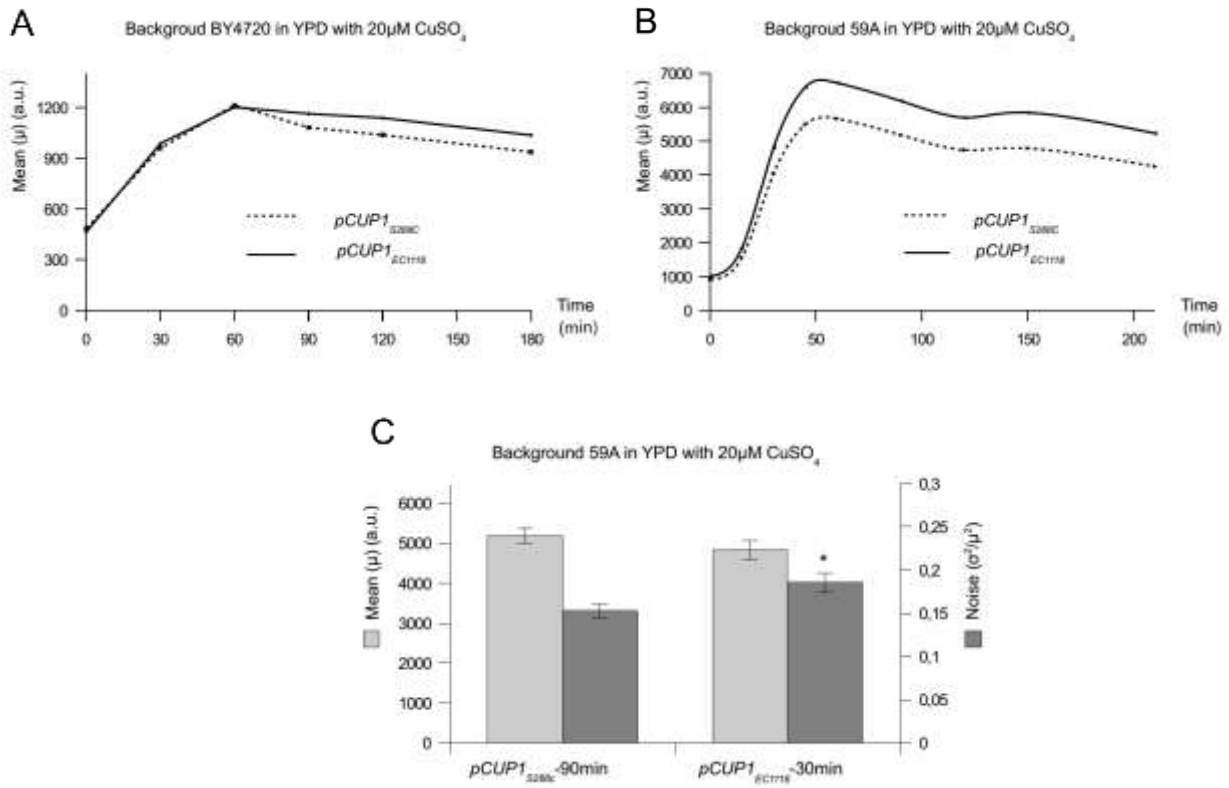


Figure 33: Induction dynamics of *pCUP1* in YPD with 20 µM $CuSO_4$

2.1.4.2 Genetic differences between $pCUP1_{EC1118}$ and $pCUP1_{S288c}$

Various promoter elements contribute to noise mainly by modulating mRNA production burst size and thus noise (Sanchez *et al*, 2013; Sanchez & Golding, 2013). Here three SNPs and one 4-bases

deletion exist in *pCUP1*$_{EC1118}$ compared *pCUP1*$_{S288c}$ (Appendix IV). Of note, these variations are common in 10 other wine strains while they are generally not present in laboratory strains (except in sigma1278b, Appendix IV). Other SNPs and Indels are also commonly found in *pCUP1* in wine strains while the coding sequence is always identical either in lab or in wine strains (the SNP in T73 is synonymous), showing that the *CUP1* transcription kinetics might be subject to many changes due to *cis*-modifications of the *CUP1* promoter. The SNPs between *pCUP1*$_{EC1118}$ and *pCUP1*$_{S288c}$ are upstream the transcription starting site, and the deletion is in the 5'-UTR. Several transcription factor binding sites are suppressed in *pCUP1*$_{EC1118}$, but only in the reverse orientation (Appendix V). The first SNP is in an HSF1p binding site, but modify a position where any base can be found.

2.1.4.3 Directed mutagenesis of pCUP1$_{S288c}$

We performed directed mutagenesis on each SNP or INDEL position in *pCUP1*$_{S288c}$ and measured mean and noise in 59A with induction at 5 µM $CuSO_4$. Only the second SNP conferred significantly higher mean expression compared to *pCUP1*$_{S288c}$ (Figure 34). Nevertheless no significant change in noise level was observed so that this SNP might contribute to increase cell-cell-variability at equal mean expression levels.

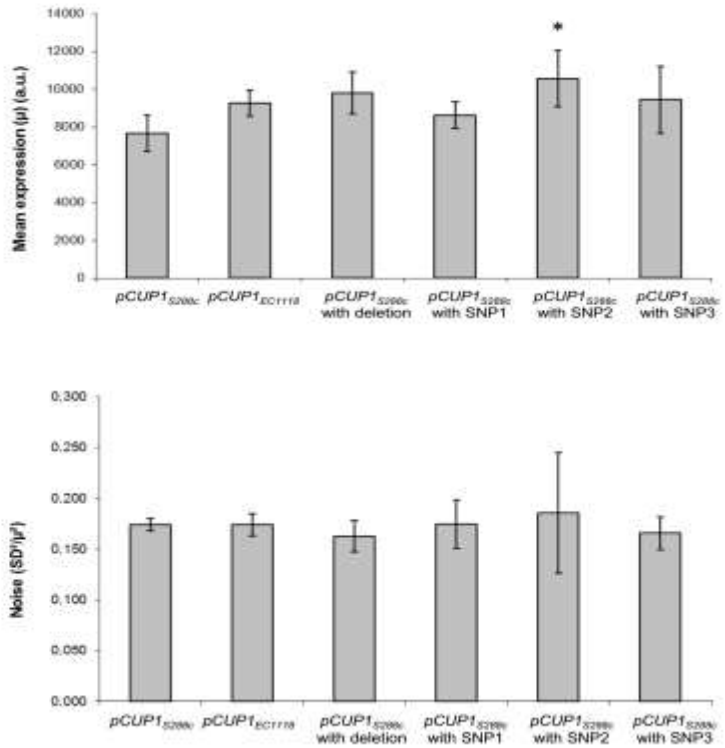


Figure 34: Measurement of mean and noise levels conferred by mutated *pCUP1*$_{S288c}$

2.1.5 Selective advantage conferred by the noisiest *pCUP1* promoter variant

It remained to consider whether this higher noise observed with $pCUP1_{EC1118}$ might confer a benefit and increase population survival upon exposure to constant selective conditions. To test for selective advantages, each variant was fused to the She *ble* gene (*ZeoR*) (conferring resistance to phleomycin) in the pJRL2 plasmid and integrated in the *LEU2* locus in 59A. The main problem with using copper as a selective agent in these growth experiments would have been the need to use the same copper concentration for both variants. Indeed, as they are not induced in the same manner by copper (less copper is needed to get the same mean expression with the industrial variant), we would have had differences of mean expression levels between the variants by using copper as a selective agent, making interpretations about the impact of noise impossible. This led us to choose *ZeoR* and phleomycin as an adequate system to test our hypothesis. Experiments were performed at steady state induction levels after overnight induction (see Methods and (Blake *et al*, 2006)). Briefly, each strain was induced in adequate copper concentrations (10 µM for $pCUP1_{S288c}$ and 5 µM $CuSO_4$ for $pCUP1_{EC1118}$) to obtain similar mean expression levels and differences of noise in YNB medium. The mean expression levels of *ZeoR* have been verified by RT-qPCR (

Figure *35*).

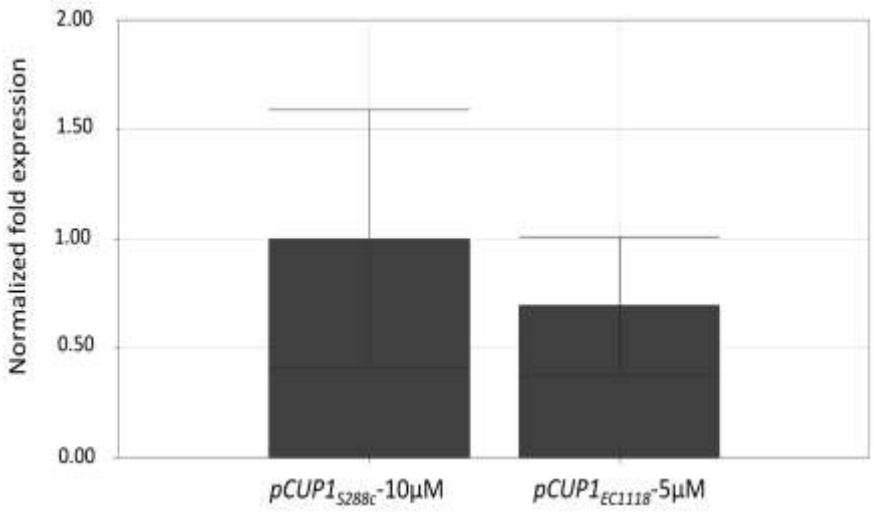


Figure 35: RT-qPCR of *ZeoR* expression

Then in exponential phase, each strain was or was not exposed to different concentrations of phleomycin to determine the residual growth at each concentration (Figure 36). While the 59A control without *ZeoR* was highly sensitive, induced strains had the highest residual growth and non-induced

strains had intermediate phenotypes. *pCUP1$_{EC1118}$* conferred a slightly higher residual growth in 30 µg/mL phleomycin without copper, probably linked to the higher basal mean expression (Figure 30G). In inducing conditions, *pCUP1$_{EC1118}$* significantly improved growth in 50 µg/mL phleomycin compared to *pCUP1$_{S288c}$* (Figure 36). In lower and higher concentrations, both strains exhibited the same behaviour. Thus the effects of the increased cell-cell variability conferred by this variant on growth in selective environment are observed only in a specific range of phleomycin concentrations.

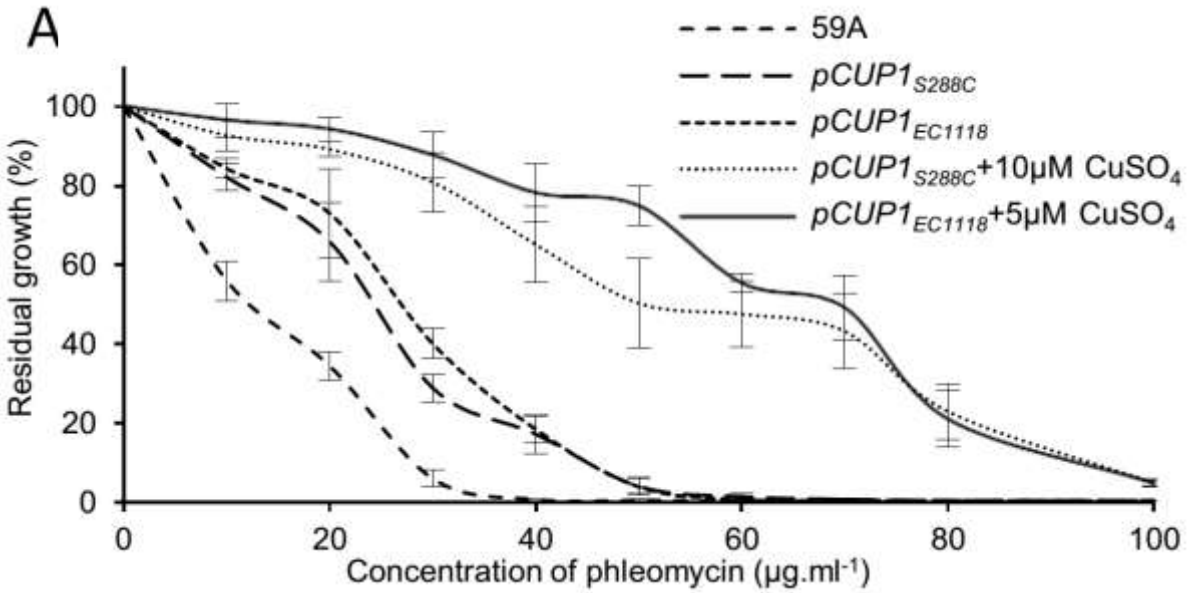


Figure 36: Residual growth after 24h treatment with various phleomycin concentrations

To finely determine growth kinetics, we followed the growth curves in 40, 50 and 60 µg/mL phleomycin in proper copper concentrations for both strains to have identical mean expression levels, as well as the growth curves with copper only (Figure 37). While growth did not show any difference in 5 µM and 10 µM $CuSO_4$ without phleomycin, *pCUP1$_{EC1118}$-ZeoR* induced in 5µM $CuSO_4$ gave a better growth than *pCUP1$_{S288c}$-ZeoR* induced in 10µM $CuSO_4$ at all measurement points in 50 µg/mL phleomycin (Figure 37B). We also observed a significant difference in 40 and 60 µg/mL phleomycin (Figure 37A and C) but it was less important, as one could have expected by looking at the residual growth curves (Figure 36). Taken together, this result confirmed that the difference of promoter-mediated noise between the natural promoter variants *pCUP1$_{EC1118}$* and *pCUP1$_{S288c}$* is sufficient to confer different abilities to survive, but only in a given range of selective pressure.

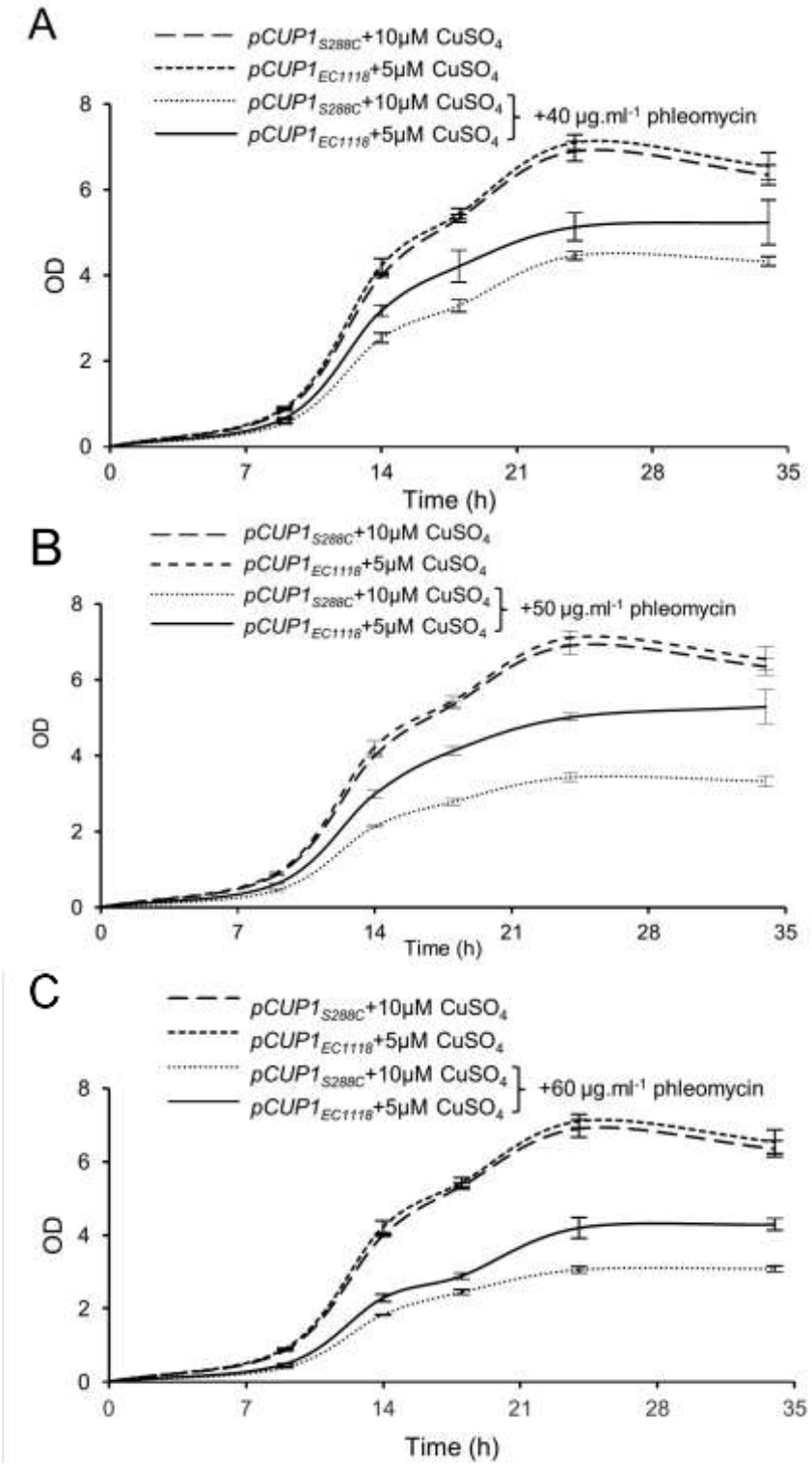


Figure 37: The *pCUP1*$_{EC1118}$ and *pCUP1*$_{S288c}$ promoter variants confer distinct growth abilities in various phleomycin concentrations.

2.2 Discussion

The genetic determinants of transcriptional-mediated noise have been characterized by rationally manipulating or randomly mutating gene promoters (Sanchez & Golding, 2013). All these studies unravel the origins that underlie gene-specific expression variability by showing that intrinsic noise is at least in part generated by *cis*-acting regulatory elements embedded within the DNA sequence of each promoter. As far as we are aware, there is no studies focused on natural promoter variants that may confer different noise levels and consequently different benefits in stressful environments, so that the relevance to adaptation in natural systems remains to be determined (Ackermann, 2013; Viney & Reece, 2013; Holland *et al*, 2014).

Here we screened for promoters conferring high noise in the genome of the haploid 59A yeast strain derived from the industrial wine EC1118 strain to search for natural variants conferring adaptive advantage in stressful conditions through enhanced cell-cell variability. As expected, sequenced EC1118 genomic fragments conferring noisy expression were enriched in GO categories possessing significantly greater-than-expected expression noise (Zhang *et al*, 2009). Especially, promoters of genes involved in nitrogen compounds transport were over-represented. Genes implicated in nitrogen metabolism are also among the genes showing significant variation in expression among natural isolates of *S. cerevisiae* (Carreto *et al*, 2011). These data should make sense because a positive correlation is known between gene expression noise and gene expression divergence in yeast (Lehner, 2008; Zhang *et al*, 2009). Moreover, nitrogen assimilation is highly variable among wine strains and correlates to fermentation efficiency (Treu *et al*, 2014).

Our study of 8 promoter variant pairs at the genomic level reveals that higher mean expression with one variant was not always associated to lower noise while this correlation has been reported many times (Newman *et al*, 2006; Hornung *et al*, 2012; Carey *et al*, 2013). When differences in mean expression exist, higher expression is generally observed with promoters from EC1118 (except for *pAGP2*). Nevertheless it often depends on the genetic background because the variants of some promoters give different mean expression levels in 59A while they do not show any difference in BY4720. Therefore *cis*- (promoter sequence variations) and *trans*- (cellular factors involved in gene expression) factors are associated to enable this enhanced expression in the industrial strain. Interestingly, the difference of noise observed between variants of several promoter pairs also depends on the genetic background, revealing epitasis in the generation of promoter-mediated noise.

We confirmed the difference in terms of noise between the *pCUP1* promoter variants by inducing them in different copper sulfate concentrations conferring the same mean expression level. One

problem could be that changes in gene expression level could be from gene expression burst frequency or/and burst size and that adding more copper into a media might increase burst frequency and thus produce less gene expression noise for the laboratory variant. Nevertheless the difference in noise observed between the two variants in these induction conditions seems not to be due to changes in burst frequency generated by different copper concentrations because at the basal level, we already observed a non-expected result between the two variants of the *CUP1* promoter in 59A (while it is not observed in S288c): the higher mean conferred by the industrial variant was not associated with lower noise. Moreover, $pCUP1_{EC1118}$ is still noisier than $pCUP1_{S288c}$ in 59A at the same copper concentration, even if the mean level conferred by $pCUP1_{EC1118}$ is higher (while mean values are not significantly different in Figure 32E). Thus, increased noise is a feature of the industrial variant only revealed in the 59A background, independently of copper concentration. This result indicates that the genetic variations probably increase burst size and decrease burst frequency at the same copper concentration in 59A, explaining the increased mean and noise levels.

Differences in global constraints on noise have already been reported between yeast strains (Ansel *et al*, 2008; Fehrmann *et al*, 2013) but most studies showed that stochastic transcriptional kinetics in yeast is mainly determined by gene-specific effects (Sanchez *et al*, 2013). Here we show the combined and correlated influence of *cis*- and *trans*-acting factors which contribute to enhance expression noise for a given gene. Thus, as suggested recently, single-cell transcriptional kinetics is affected by both promoter architecture and genome-wide processes in yeast (Sanchez & Golding, 2013). The hypothesis of noisier expression of regulatory factors (noise propagates in regulatory pathways (Blake *et al*, 2003)) is not relevant because both versions would be identically affected. As $pCUP1_{EC1118}$ enhances noise only in 59A, the fact that proteins involved in *CUP1* transcription or in the global transcriptional process in this strain might be more sensitive to genetic variations in *pCUP1* in terms of promoter binding. This hypothesis might provide an explanation for this epistatic interaction in the consequences of *pCUP1* sequence modifications. In any case, the consequences of the SNP or the deletion between the *pCUP1* variants are not dramatic and might reveal that natural systems evolve through promoter modifications generating small effects on noise, and not strongly affecting cell-cell variability such as the ones produced by rational manipulation of yeast promoters.

Phenotypic consequences of noise in gene expression in terms of survival in selective environments are little studied (Viney & Reece, 2013). Only artificial systems using rationally modified promoters have been employed to test noise-mediated fitness differences (Blake *et al*, 2006; Smith *et al*, 2007). For instance, by introducing mutations within the TATA region of an engineered *S. cerevisiae GAL1* promoter, Blake et al have shown that increased cell-cell variability in the expression of *ZeoR*

confer a clear benefit in Zeocin containing medium (Blake *et al*, 2006). Nevertheless differences of noise between the mutated promoters were very high and this proof of principle did not imply that such an adaptation mechanism through noise modulation may naturally occur. By fusing natural yeast variants of *pCUP1* with *ZeoR*, we show here that their different noise levels at equal mean expression indeed provide distinct abilities to survive in a constant environmental stress, even if the noise difference is far less important than in Blake's study. Growth curves show that growth reduction is only observable in a specific range of phleomycin concentrations. This result is quite different from Blake's study where enhanced noise was always either disadvantageous or beneficial in all tested concentrations. Here a benefit is conferred by $pCUP1_{EC1118}$ at intermediate concentrations of phleomycin and no growth difference with $pCUP1_{S288c}$ exists at lower or higher antibiotics concentrations. The slightly more heterogeneous expression distribution with $pCUP1_{EC1118}$ might provide a possible explanation. Indeed the difference between the distributions might be too weak to reveal a selective advantage at high phleomycin concentrations because very few cells express more ZeoRp with $pCUP1_{EC1118}$ in the extreme subpopulation. On the contrary, at intermediate concentrations a larger proportion of the population is above the expression threshold necessary to grow in these concentrations. Therefore the difference between the distributions might be sufficient in this case to generate a benefit because more cells express more ZeoRp with $pCUP1_{EC1118}$.

Identification of a stronger *pCUP1* in an industrial wine strain makes sense because wine yeasts are frequently exposed to high copper content in fermentation must, especially because the Bordeaux mixture containing copper sulfate is widely used as a fungicide in vineyards. *CUP1* is one of the best examples of high correlation between evolution in a stressful environment, the expression level of the gene conferring resistance to this environment, and the resistance itself (Adamo *et al*, 2012; Chang *et al*, 2013). One can also expect that the genome of wine strains would contain more *CUP1* copies than laboratory strains. which is actually not the case as for EC1118 as well as for other wine strains (Dunn *et al*, 2005). This might be explained by a selective disadvantage of having lots of *CUP1* copies counterbalancing the benefits of copper resistance. Interestingly, various genes involved in stress response, especially *CUP1*, have been positively correlated with the fermentation duration (Ambroset *et al*, 2011). Thus a high stress response was associated with a low fermentation capacity. Evolution of *CUP1* towards higher noise in its expression is conceivable: high mean expression is both advantageous in high copper concentrations and disadvantageous for fermentation. These traits exert opposing selective pressures on *CUP1* expression. Noisier expression would be advantageous in that it would make the population harbour an optimum between fermentation capacity and adaptability to copper

rich environments when necessary. In bacteria, a constant selective pressure select for mutants harbouring either higher mean with no noise increase or higher noise with similar mean (Ito *et al*, 2009). But opposing selective pressures would most likely select for increased expression variability only. Thus, *pCUP1* might have evolved towards higher noise because it might increases fitness in fermentation-associated copper-rich environments. More direct evidence could be provided by evolving *S. cerevisiae* in controlled fermentative copper-rich environments, where high *CUP1* expression is advantageous regarding copper, and disadvantageous regarding fermentation efficiency. It would likely confirm the selection for noisy *CUP1* expression. It has recently been shown that environmental stress selects for organisms with increased phenotypic heterogeneity in yeast populations, but no links with increased variability in the expression of key genes have been established yet (Holland *et al*, 2014). Finally, numerous genomic studies are in progress and new yeast genomes will be available soon to determine if positive selection seems to have recently occurred on the *pCUP1* of wine strains. Genetic studies would also determine if the observed genetic variations are under positive selection or if alternative explanations (neutral evolution, relaxed selective constrain, or fixation of slightly deleterious mutation) should be favoured.

Collectively, our results provide evidences that natural yeast promoter variants can exhibit different levels of transcriptional-mediated noise but also that epistasis exists in the generation of this noise: the combined influence of promoter sequence modifications and the trans-background acting contribute to modify it. Finally, we show that natural yeast promoter variants conferring distinct abilities to survive in a stressful environment through noise modulation can be found among *S. cerevisiae* strains, showing that this possible adaptation mechanism has to be considered when studying yeast evolution and when exploring natural and artificial genetic diversity to improve industrial yeast strains (Steensels *et al*, 2014).

2.3 Perspectives

2.3.1 Further exploration of the library

Newman *et al* (2006) have shown that the yeast genome contains at least one thousand high noise genes. As the library we have constructed confidently covers the whole yeast genome, there should be much more “noisy clones” in the library after the FACS selection than the number that we have identified. Therefore, it should be possible to identify more high noise clones in the library by large scale screening using high performance flow cytometry and robotics. Single cells can be isolated

into 96-well plates by a sorting system and their expression profile analyzed as described before. By this way, more high noise promoters would be identified after sequencing of the fragments conferring the most variable yEGFP expression. Moreover, analysis of these new data by bioinformatic tools could reveal interesting information such as particular sequence motifs or *cis*-elements in the high noise promoters in the industrial strain which might be more likely to evolve.

The study of 2 variants of 8 yeast promoters at the genomic level in different strains indicated that *cis*-elements and *trans*- factors both play an important role in the control of expression noise. Here we only focused on the effects of *cis*-elements (in promoters), but the effects of *trans*- factors are also worth of exploiting. QTL analysis is an useful tool to find how elements of the genetic background affects noise levels *in trans* (Ansel *et al*, 2008; Fehrmann *et al*, 2013). If we compare the same variant of a promoter in different strains in Figure 30, we can see for instance that $pCUP1_{EC1118}$ exhibited higher mean level and higher noise in 59A compared with BY4720, whereas $pCAN1_{S288C}$ exhibited lower noise in 59A with the same mean expression level as in BY4720. Thus these 2 promoters in different strains could be used for independent QTL analysis to find the *trans*-factors in 59A which might increase or decrease the expression noise.

Moreover, we have demonstrated that the higher noise conferred by $pCUP1_{EC1118}$ in 59A can increase the phleomycin resistance at certain concentrations when driving *ZeoR* expression. But whether this difference can confer better resistance to copper is still to be tested. We have previously explained that these two promoters have different mean expression levels when the media where they are induced contain the same copper concentration, making impossible any interpretation of the effects of noise. An alternative method would be using promoters independent of the copper concentration, which confer similar mean expression levels and different noise levels, either when induced or at the basal expression level. By expressing the *CUP1* gene under the control of these promoters, it would be possible to test if a promoter giving higher noise indeed gives better resistance to high concentrations of copper. Nevertheless, we need *cup1Δ* strains for this experiment but this deletion is very difficult to obtain because of gene duplication. Alternatively, its regulator *CUP2* could be deleted (the *cup2Δ* strains are highly sensitive to copper) what would affect all the *pCUP1* copies.

Promoters such as *pAGP2* and *pCAN1* are also interesting to be further studied. Like $pCUP1_{EC1118}$, $pCAN1_{EC1118}$ confers a higher expression level in 59A without decreasing expression noise, suggesting that it probably provides higher noise at equal mean. Unfortunately, *pCAN1* is not inducible, thus we cannot directly compare the noise level from different variants to confirm this hypothesis because modulation of the mean expression level is not possible. Nevertheless, opposing selection pressures selecting for noisy *CAN1* expression might be provided by its positive effect in nitrogen-poor musts

(favouring enhanced *CAN1* expression to transport arginine) and its negative consequence in musts where toxic agents transported by the permease Can1p are present (such as the arginine analogue canavanine which is synthesized by many plants). Again, it would be interesting to directly test this hypothesis by placing *S.cerevisiae* in fluctuating environments containing alternatively high canavanine or low arginine concentrations.

In contrast to $pCAN1_{EC1118}$, $pAGP2_{EC1118}$ exhibits lower expression level in 59A with similar expression noise as $pAGP2_{S288c}$, indicating that its expression noise is more constrained in the industrial strain. *pAGP2* cannot be induced neither but *AGP2* is a sensor for environmental changes (Schreve & Garrett, 2004) and affects the regulation of many transporter genes (Maraganore *et al*, 2005). Further analysis of *AGP2* to explore hypotheses explaining why its noise level is lower in 59A could give some insights on the different aspects of noise evolution in industrial strains.

2.3.2 Strain improvement through increasing expression noise

Increased expression noise of certain genes confers selective advantages in both fluctuating and constant environments through increasing the global phenotypic heterogeneity in the population. Thus it should be possible to improve the ability of industrial stains to resist to certain stresses through increasing the expression noise of the genes related to these stresses. For instance, we can consider the osmoadaptation pathway in yeast: the membrane proteins Sln1, Hkr1, or Msb2 act as osmosensors that react to environmental information and simulate the MAPK signaling pathway to allow resistance to osmotic stress (Tanaka *et al*, 2014). The transmembrane transporter genes are shown to exhibit high expression noise in wild strains (Fehrmann *et al*, 2013), thus by increasing their expression noise, resistance to osmotic stress might be increased. Their expression noise could be modified through mutating their promoter so that only the genes we are interested in would be affected, and that the industrial properties of the strain we want to improve will remain the same. As it is difficult to expect precise effects on noise by targeting specific promoter elements, and because it should be interesting to find mutated promoters with increased noise but similar mean (what is improbable because increasing noise often decreases the mean), the promoter region of these genes could be randomly mutagenized by PCR and then cloned and fused to yEGFP into an integrative plasmid. These mutated promoters fused to yEGFP could then be integrated in an industrial strain and the mutated promoter library be subjected to the same procedure as described previously to select high noise promoters. Then resistance to osmotic stress of the strain where the original promoter have been replaced by a selected mutated variant would be tested by parallel experiments or competition experiments.

2.3.3 Evolution of high noise promoter

In this work, we used the fluctuating selection method developed by Freed *et al.* (2008) to enrich high noise clones in the library. But this method can also serve as a selection pressure and be used in evolutionary experiments. Wildenberg and Murray (2014) subjected a strain harboring a high mutation rate and P_{FLO1}-*YFP* in the genome to fluctuating selection for high and low fluorescence levels during 30 cycles (30 days). But they didn't apply any gate on cells homogeneous in terms of size and complexity during the experiment, so that the highest fluorescent part of the population they selected mostly was cell clumps, whereas the lowest part was individual cells. Finally, they selected for cells harboring a 24h oscillation between a single cell state and multicellular clumps. But what if we apply a gate to select for on the most homogeneous subpopulation containing only single cells with similar cell size and status? We assume that the population would evolve towards higher noise expression if we subject a strain with YFP fused to a gene in its original locus, for example *HOG1* which is a key kinase in the osmotic response pathway, to this long-term selection procedure.

3 Study of bimodal expression pattern

During the screening for high noise clones in the genomic library after FACS selection, we identified some interesting clones harbouring bimodal expression patterns, meaning that they had two peaks in the distribution whereas the other clones had only one peak. This expression pattern has been observed in different organisms from *E. coli* to humans (Figure 16) (Silva-Rocha & de Lorenzo, 2012; Pelet *et al*, 2011; Shalek *et al*, 2013). One of the advantages of this pattern is that it confers two very different subpopulations that can be adapted to contrasted situations in an isogenic population (referred as bet-hedging, (Viney & Reece, 2013). The underlying mechanisms of a bimodal expression pattern are not clear. In this part, we worked on several bimodal clones to further analyze their behavior at the genomic level.

3.1 Results

3.1.1 Bimodal expression clones after FACS selection

During the isolation of high noise clones (section 2.1.1.3), we found around 30% of the clones we selected exhibiting bimodal expression (Figure 38A). This expression pattern was very stable and reproducible. First, we isolated 2 single sub-clones from the clones we selected and both of them exhibited bimodal expression as the original ones (Appendix I). Second, we applied a very small gate to select the most homogenous cells in the population and it did not change the expression profile (Figure 38B). Third, we extracted the plasmids from 12 of these clones and re-transformed them to the CEN.PK and BY4720 strains. The *de novo* transformants all exhibited bimodal expression (Figure 38C). Thus we inferred that this expression pattern is determined by the genomic sequences inserted before *yEGFP* in the plasmids.

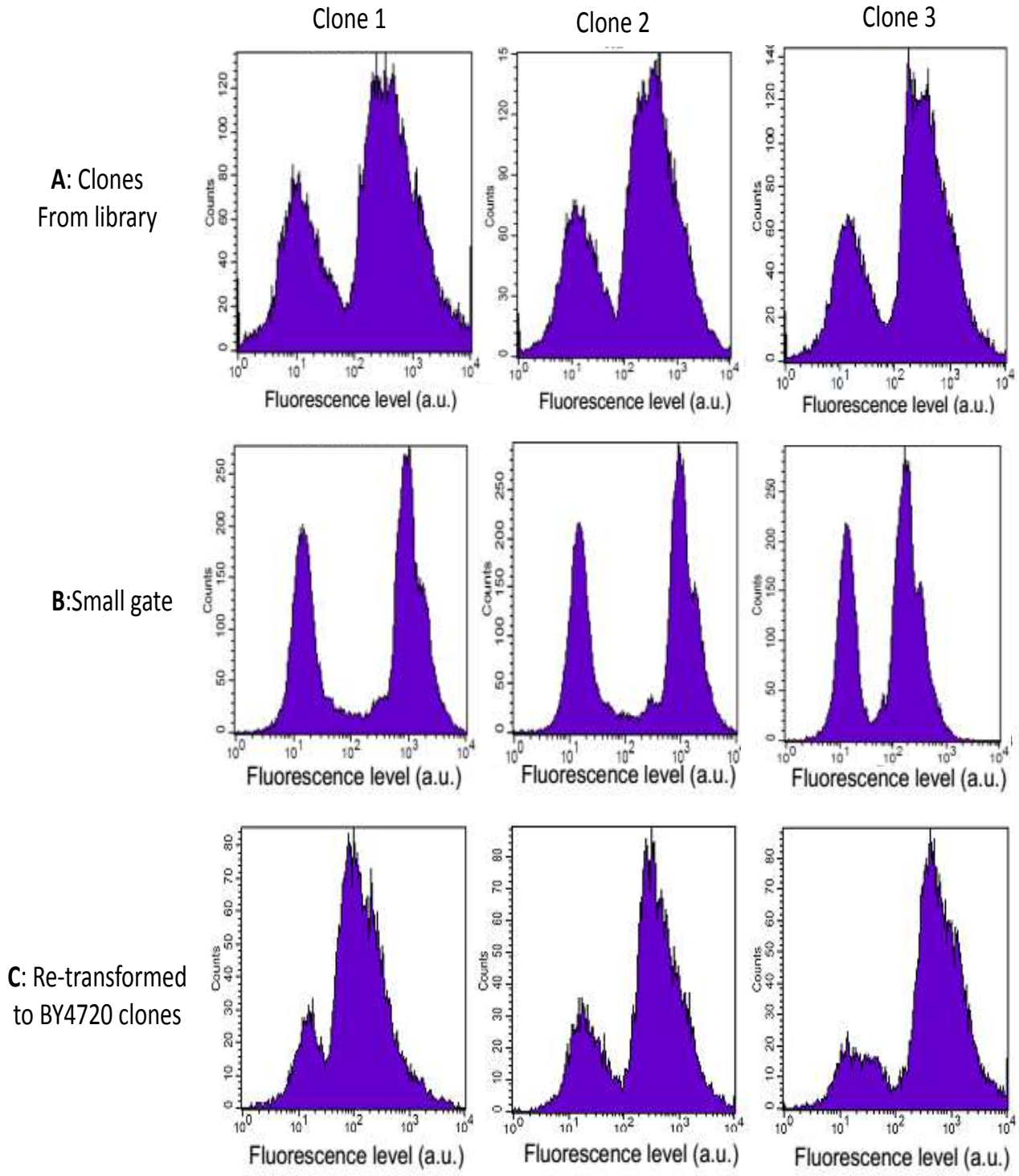


Figure 38: Fluorescence distribution in some selected clones showing bimodal expression

Finally, the plasmids from the 31 bimodal clones were sequenced, and 19 different fragments were mapped to S288c genome. All of them corresponded to promoter sequences (see section 2.1.2), seventeen for protein coding genes and two for non-coding RNA genes has been identified (Table 5).

Table 5: Bimodal promoters sequenced

Genes	Functions
BMH1/YER177W	14-3-3 protein, involved in regulation of many processes
BMH2/YDR099W	14-3-3 protein, involved in regulation of many processes
CUP1/YHR055C	Binds copper, resistance to high concentrations of copper and cadmium
CAN1/YEL063C	Plasma membrane arginine permease
YCK2/YNL154C	Palmitoylated plasma membrane-bound casein kinase I isoform
HAC1/YFL031W	bZIP transcription factor, regulates the unfolded protein response
GNP1/YDR508C	High-affinity glutamine permease
AGP2/YBR132C	High affinity polyamine permease, osmotic stress
RPS29B/YDL061C	Protein component of the small (40S) ribosomal subunit
RPS4B/YHR203C	Protein component of the small (40S) ribosomal subunit
RPS0B/YLR048W	Protein component of the small (40S) ribosomal subunit
NEW1/YPL226W	ATP binding cassette protein
YOP1/YPR028W	Membrane protein that interacts with Yip1p to mediate membrane traffic
VRG4/YGL225W	Golgi GDP-mannose transporter
TPI1/YDR050C	Triose phosphate isomerase, abundant glycolytic enzyme;
MFA1/YDR461W	Mating pheromone a-factor
CCT7/YJL111W	Subunit of the cytosolic chaperonin Cct ring complex
SNR30/snR30	Small nuclear RNA
SNR50/snR50	Small nuclear RNA

All the eight genes that we have selected for further analysis at the genomic level in the first part of this work were included in these bimodal expression promoters (see section 2.1.2). Unfortunately, after integration into the locus *LEU2*, all these promoters exhibited a unimodal expression profile in both backgrounds (59A and BY4720). Considering that we used the full promoter region instead of the sequenced fragment for the integration, we exchanged the corresponding fragment and the promoter of *CAN1* in the centromeric plasmid (pUG35, see methods) and the integrative plasmid (pJRL2, see methods). The new plasmids were transformed to BY4720. We found that only centromeric plasmids can exhibit bimodal expression (Figure 39).

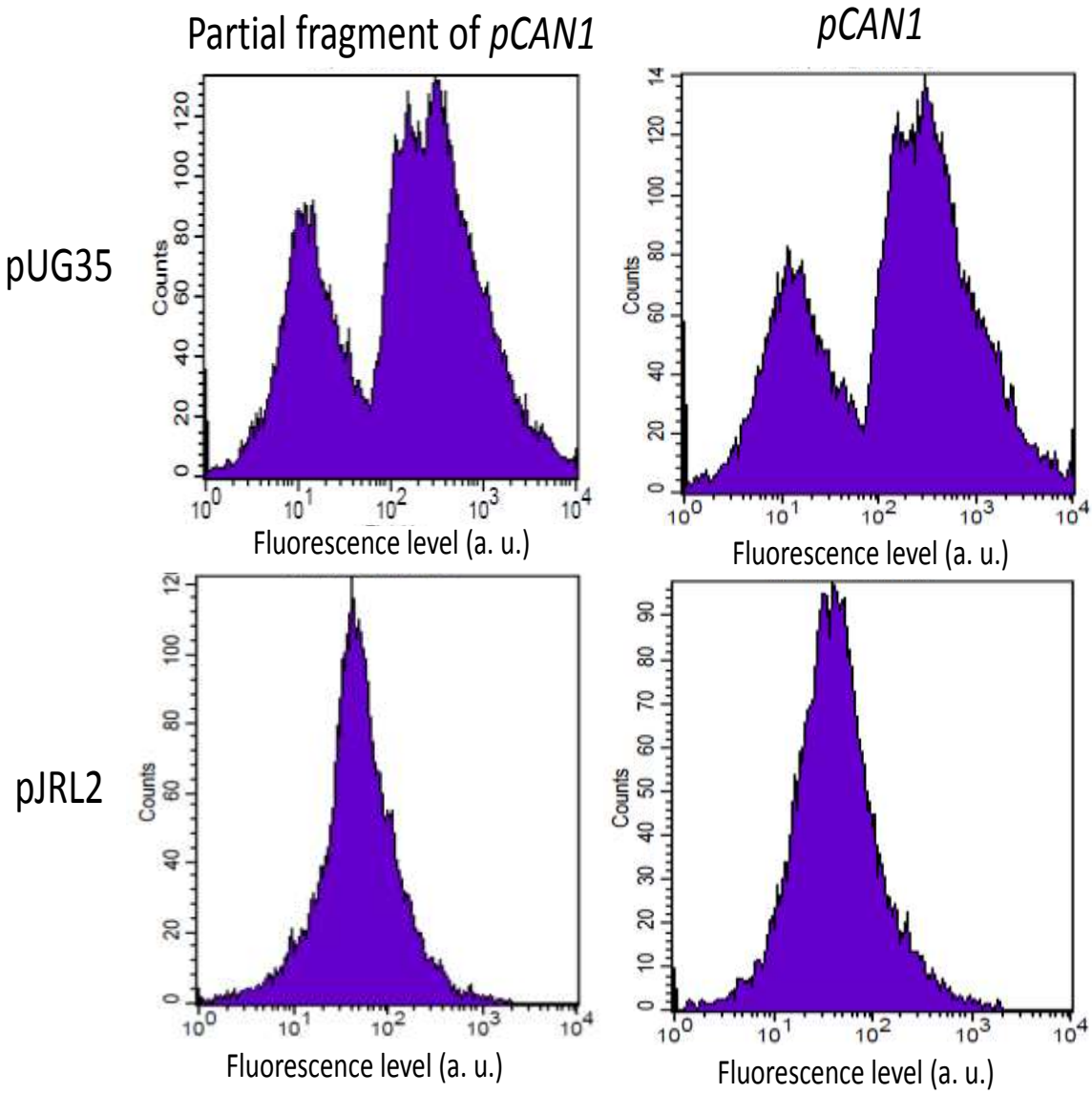


Figure 39: Expression profile of *pCAN1* and its partial fragment isolated from the enriched library in different plasmids in BY4720

3.1.2 Bimodal expression of *CUP1*

When we induced the *CUP1* promoter variants at genomic level in BY4720 in YPD medium, both variants showed bimodal expression (Figure 40). There was no differences and we did not observe this expression profile in 59A (Figure 40), indicating that this expression pattern was mainly controlled by *trans*-factors.

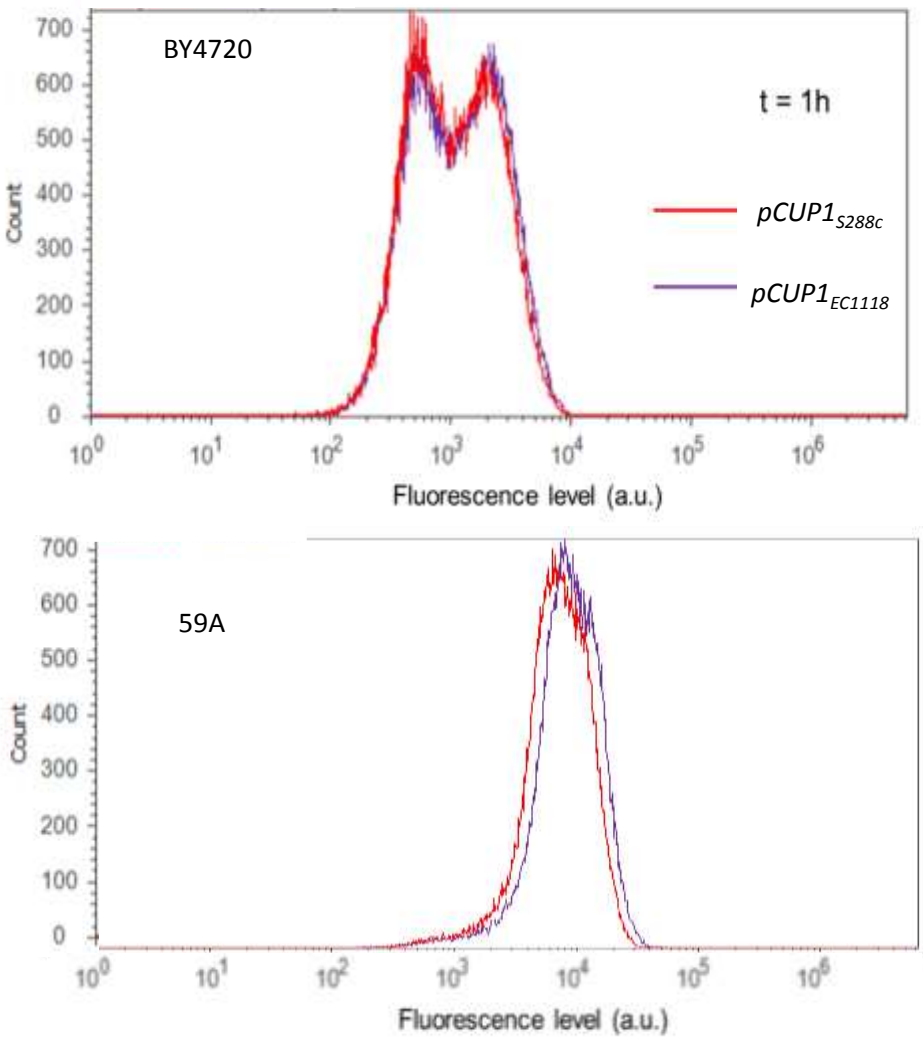


Figure 40: Expression profile of the *CUP1* promoter variants in BY4720 and 59A. The figure shows the expression profile of $pCUP1_{S288c}$ and $pCUP1_{EC1118}$ when they are induced by 20µM copper sulfate for 1h.

We studied the induction dynamics of $pCUP1_{S288c}$ in BY4720 (Figure 41). For the induced bimodal profile, the first peak was at the same expression level as the basal non-induced expression and the second peak moved as time increased. After the first 60 min, the positive peak moved to the highest levels but the cell proportions in the 2 peaks didn't change, so the mean expression level increased. After 1h the second peak moved to lower levels, and the proportion still did not change. Finally, the 2 peak fused to one peak (t=120min).

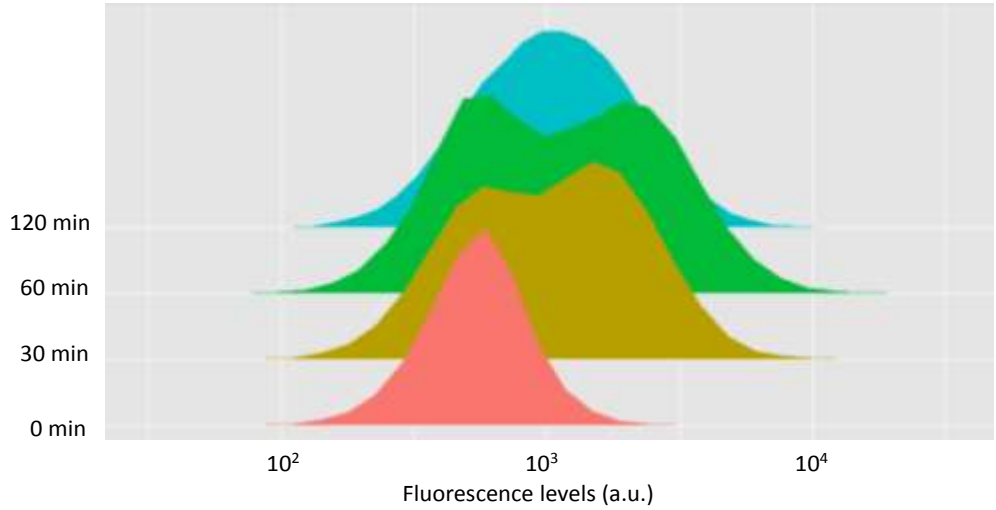


Figure 41: Expression dynamics of *pCUP1*$_{S288c}$ at different time of induction by 20µM copper sulfate in BY4720

Moreover, we also tried different concentrations of copper sulfate to induce *pCUP1*$_{S288c}$ (Figure 42). Again, the first peak was at the same position as for the basal expression. The second peak moved to higher levels as copper increased, and the cell proportion in this peak also increased.

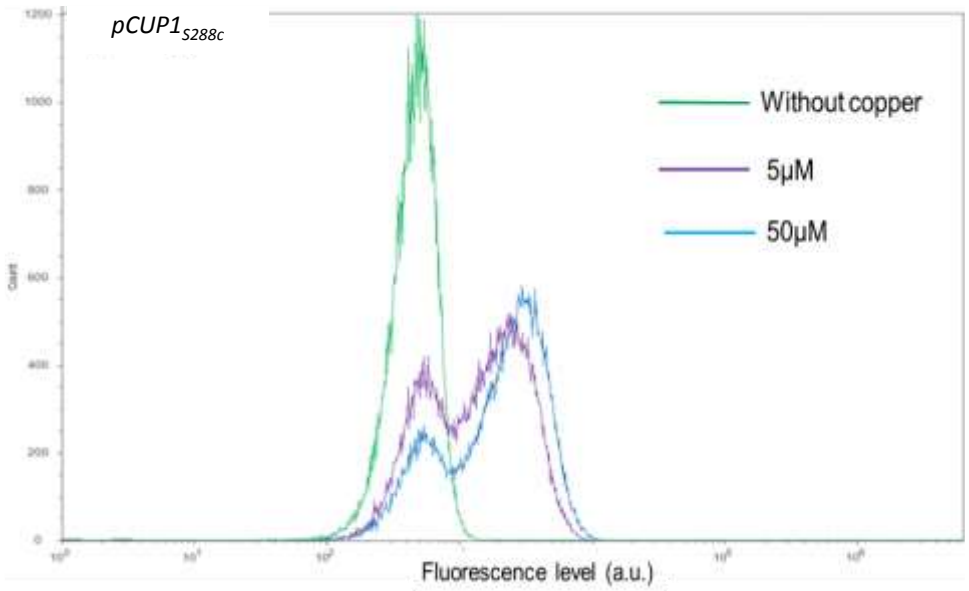


Figure 42 Expression profile of *pCUP1*$_{S288c}$ induced by different concentrations of copper sulfate for 1h in BY4720

Finally we fused YFP to *CUP1* at the original locus in BY4720 and 59A, and then measured the expression profiles in different conditions. Unfortunately, these strains did not exhibit bimodal

expression profile neither in normal conditions nor in inducing conditions (Figure 43).

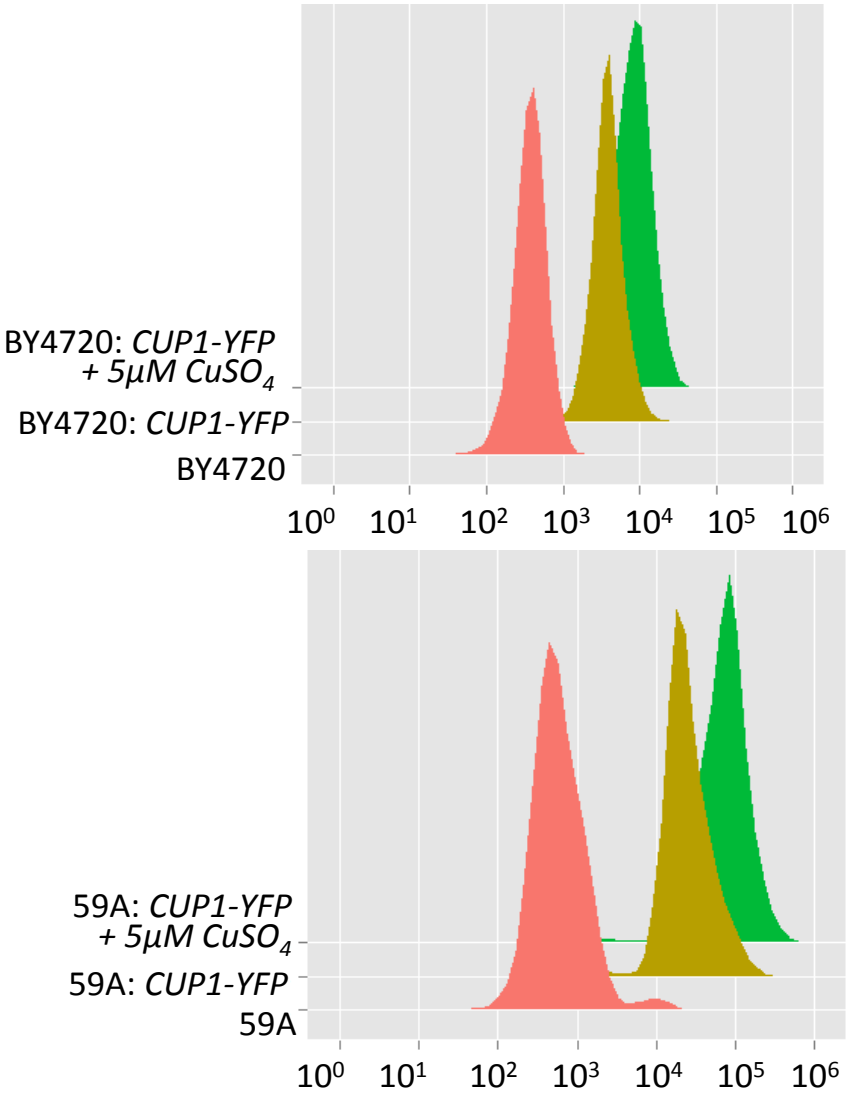


Figure 43: Expression profile of YFP fused to *CUP1* in BY4720 and 59A

3.2 Discussion

We identified 19 different fragments conferring a bimodal expression pattern in a centromeric plasmid. But their corresponding promoters fused to YFP integrated in the *LEU2* locus all exhibited a unimodal expression pattern in an optimal environment (YPD medium). By exchanging the whole promoter of *CAN1* and its corresponding fragment between the centromeric and the integrative plasmids, we showed that both conferred bimodal pattern when they are located in the centromeric plasmid whereas they both exhibited unimodal pattern after genomic

integration. Thus, in this case the bimodal expression pattern is a feature of the expression from centromeric plasmids. Different expression profiles of the same gene located on chromosome and centromeric plasmid have been observed for long time. In general, promoters giving low and moderate expression levels are sensitive to this context effect (Marczynski & Jaehning, 1985). We observed this effect for *pCUP1* and *pAGP2*: they conferred low unimodal expression at the genomic level, while a higher expression level and a bimodal profile were observed in the centromeric plasmid. On the contrary, the expression level from promoters giving higher expression level such as *pBMH1, pBMH2 and pGNP2* did not change a lot, but the expression profile changed. These effects are supposed to be due to the different structural features between plasmids and chromosomes. Chromosomes are highly structured by histones and other proteins in the chromatin and their arrangement is precisely controlled in the nucleus whereas plasmids are less regulated by such events.

Bimodal pattern was also observed from the *CUP1* promoter variants in BY4720 when they are induced by copper sulfate, but not in 59A. One hypothesis to explain this phenomenon might be the different copy number of *CUP1* in BY4720 and 59A. Recent researches have shown that the laboratory strain S288c can contain up to 14 copies of *CUP1* (Zhao *et al*, 2014), while the sequencing of EC1118 indicated that it contains much less number of *copies* than S288c (Novo *et al*, 2009). Thus, to compensate its lower *CUP1* copy number, 59A might exhibit higher induction efficiency of *CUP1* than BY4720 what might lead to unimodal expression pattern. Otherwise, *CUP1* is mainly activated by the transcription factor Cup2 (Shen *et al*, 2001). The Cup2 binding site in the *CUP1* promoter is identical in BY4720 and 59A. Thus the different induction efficiencies in these strains might be due to different *CUP2* expression levels.

In BY4720, *pCUP1* exhibited different behaviors during induction, either depending on time or copper concentration. Under certain copper concentrations, a constant proportion of the population is induced and the induction level of this subpopulation first increased then decreased along induction time, whereas after the same induction time, higher copper sulfate concentrations induced more cells at higher expression levels. But the underlying mechanism still has to be explored. We think the induction dynamics of *CUP2* might play an important role in these phenomena.

Finally, when *CUP1* was fused to *YFP* at its own chromosome locus in 59A and BY4720, it showed higher expression level and unimodal expression pattern in both cases (Figure 43) with or without induction by copper sulfate, indicating that the chromosome locus and environment play

an important role in the regulation of *CUP1* and their effects are independent of copper induction. The expression level from the original locus is much higher than from the *LEU2* locus indicating that the transcriptional regulators might preferably bind to the original locus.

When we fused YFP to *CUP1* by transformation, we obtained several clones. We tested their expression profile and found all of them exhibited the unimodal expression pattern. Considering that there are several copies of *CUP1* in the genome, YFP might be fused to different copy of *CUP1* in different clones. But they all expressed *CUP1* similarly, indicating that every copy of *CUP1* seems to be similarly controlled.

3.3 Perspectives

Our results showed that the regulation of bimodal expression is very complex. The main question is why the centromeric plasmid can confer bimodal expression pattern. The deciphering of regulation differences between plasmids and chromosomes can help us to better understand the control of gene expression in yeast. The differences between plasmid and chromosome that affect gene expression might be their topological structure and/or their nucleosome configuration. On one hand, in the population transformed with the centromeric plasmid, cells could be divided to two parts according to their expression level. Then the plasmids from these two subpopulations could be isolated to see if there are differences between them in terms of nucleosome configuration. On the other hand, for the bimodality at *LEU2* locus, using inhibitors of chromatin modifications such as trichostatin A (inhibits histones deacetylation) and 5-azacytidine (inhibit DNA methylation) could change the chromatin status and help us to understand the possible relationship between chromosome structure and bimodality.

As mentioned before, Cup2 plays an important role in the *CUP1* regulation. Exploring the *CUP2* expression dynamics could help us to understand the complex expression pattern of *CUP1*. Thus we could fuse tdTtomato to *CUP2* in its original position in both backgrounds. YFP and tdTtomato can be detected simultaneously and can be used to couple the studies of the *CUP1* and *CUP2* expression dynamics what might uncover the possible mechanism of bimodal expression in BY4720. We could also try to eliminate the copy number effects of *CUP1*. All *CUP1* repeats could be deleted and replaced by a single copy of *pCUP1-CUP1-YFP*, what would help us to better

understand the regulation of *CUP1* without copy number variation in the genome.

Otherwise we can see that the *CUP1-YFP* fusion in 59A exhibits much higher expression levels than in BY4720, but its noise level is similar. Thus *CUP1-YFP* probably exhibit higher noise in 59A. We could perform gene introgression analysis to find the possible background effects that affect expression noise. In BY4720, *YFP* was integrated with *kanR* whereas in 59A it was with *natR*. Thus we could just mix these two strains to create diploid cells and select the diploid cells on plates with both G418 and nourseothricin. Then these diploids would be put on KAc plates to promote sporulation and the haploid cells be tested for *CUP1-YFP* expression level and noise. One haploid strain conferring high expression noise and neocin resistant would be backcrossed with BY4720. This procedure could be repeated for at least 10 times to finally sequence the last strain to find QTLs affecting the expression noise of *CUP1-YFP*.

4 Connecting Noise to Genome Instability

In this part, we have tried to test our hypothesis that higher noise in some genome maintenance genes can give better global adaptability to different stresses by increasing the rate of genetic-variant generation (RGVG) (Capp, 2010). Thus we fused DNA repair and maintenance genes with two different modified *PHO5* promoters, which confer similar mean expression levels with different levels of noise (Raser & O'Shea, 2004). One of them had mutations in the TATA box region (TATATA were converted to AATATT, referred as $pPHO5_{TATA\text{-}A1T6}$), the other one had mutations in the UAS (upstream activating sequence) region (CACG were converted to AAGC, referred as $pPHO5_{UASm1}$). $pPHO5_{TATA\text{-}A1T6}$ has a lower noise level than $pPHO5_{UASm1}$. These promoters both need a medium without phosphate to induce the expression of the genes of interest, and the expression level can be modified by adding phytic acid as source of phosphate. These constructions were transformed to two different strains: JA0200 which has *URA3* inserted inside the tandem repeats (TRs) present in the *FLO1* gene (Verstrepen *et al*, 2005) and JA0300 where *URA3* is placed under the dependence of the *SDT1* promoter (Vinces *et al*, 2009). In the first strain, homologous recombination between the TRs in *FLO1* can delete the *URA3* gene which allows survival on 5-FOA plates. Therefore we can use JA0200 to measure the homologous recombination frequency (HRF). In the second strain, the number of TRs in the promoter of *SDT1* can affect the mean expression level of *URA3* which therefore can evolve towards higher or lower expression. Thus JA0300 can be used to perform fluctuating selection experiments.

4.1 Results

4.1.1 Effects of noise on the homologous recombination frequency

4.1.1.1 Induction conditions of the PHO5 promoter variants conferring similar expression levels and different noise levels

We first considered the HR pathway for studying the impact of noise on genome stability. *RAD52* and *RAD27* are proteins that can impact the HRF. The *RAD52* deletion can decrease the HR frequency whereas the *RAD27* deletion increases the HRF. pJRL2 plasmids containing the different *PHO5* promoter variants fused to *RAD52-YFP* or *RAD27-YFP* were transformed to be integrated into the *LEU*2 locus in the JA0200 strain. This strain is a derivative of the original KV133 strain (where

URA3 has been inserted in *FLO1*) where *LEU2* was reinserted into the genome(Figure 23). Then the native *RAD52* or *RAD27* genes were deleted using *LYS2*. Under full induction (Pi-free medium without any phytic acid), $pPHO5_{TATA\text{-}A1T6}$ exhibited lower noise level but higher mean expression level than $pPHO5_{UASm1}$ (Figure 44, $p<0.01$ for both mean values and noise levels and both *RAD52-YFP* and *RAD27-YFP*). We tried to decrease the expression level of $pPHO5_{TATA\text{-}A1T6}$ by adding phytic acid into the phosphate free medium (Raser & O'Shea, 2004). Finally, we found that $pPHO5_{TATA\text{-}A1T6}$ induced in Pi-free medium containing 60µM phytic acid exhibited the same mean expression level as $pPHO5_{UASm1}$ under full induction (Figure 44, $p=0.56$ for *RAD52-YFP* and $p=0.99$ for *RAD27-YFP*), whereas they had different noise levels (Figure 44, $p=0.03$ for *RAD52-YFP* and $p=0.02$ for *RAD27-YFP*). We also measured the native expression level of *RAD52* and *RAD27* through fusing YFP to the N-terminal domain at their original chromosomal positions. We found that, under the induction conditions we have chosen, the mean expression levels from the *PHO5* promoters are about the half of their native levels (Figure 44).

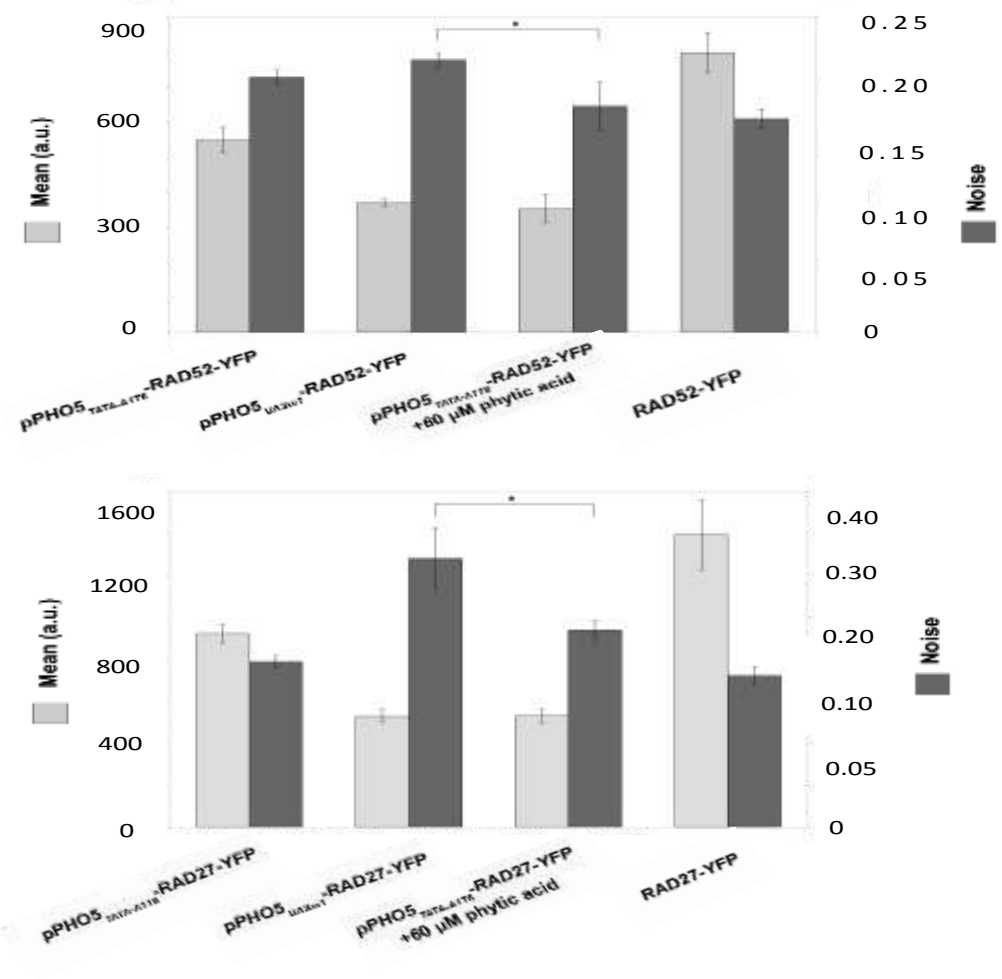


Figure 44: Induction of the *pPHO5* promoter variants fused to *RAD52-YFP* or *RAD27-YFP*

4.1.1.2 Measurement of the homologous recombination frequency

First, we measured the HRF of all the JA0200 derivative strains incubated in induction media for 7h (Figure 45). The addition of 60µM phytic acid in the medium did not affect the frequency in the control strain JA0200. Thus, phytic acid has no effects on homologous recombination. The *RAD52* deletion leads to lower HRF whereas the *RAD27* deletion leads to higher HRF, what is consistent with previous works (Verstrepen *et al*, 2005). The strains with the YFP fusions at the genomic level (*RAD52-YFP* and *RAD27-YFP*) exhibited similar HRF as the control strain JA0200, indicating that Rad52-YFP and Rad27-YFP are functional.

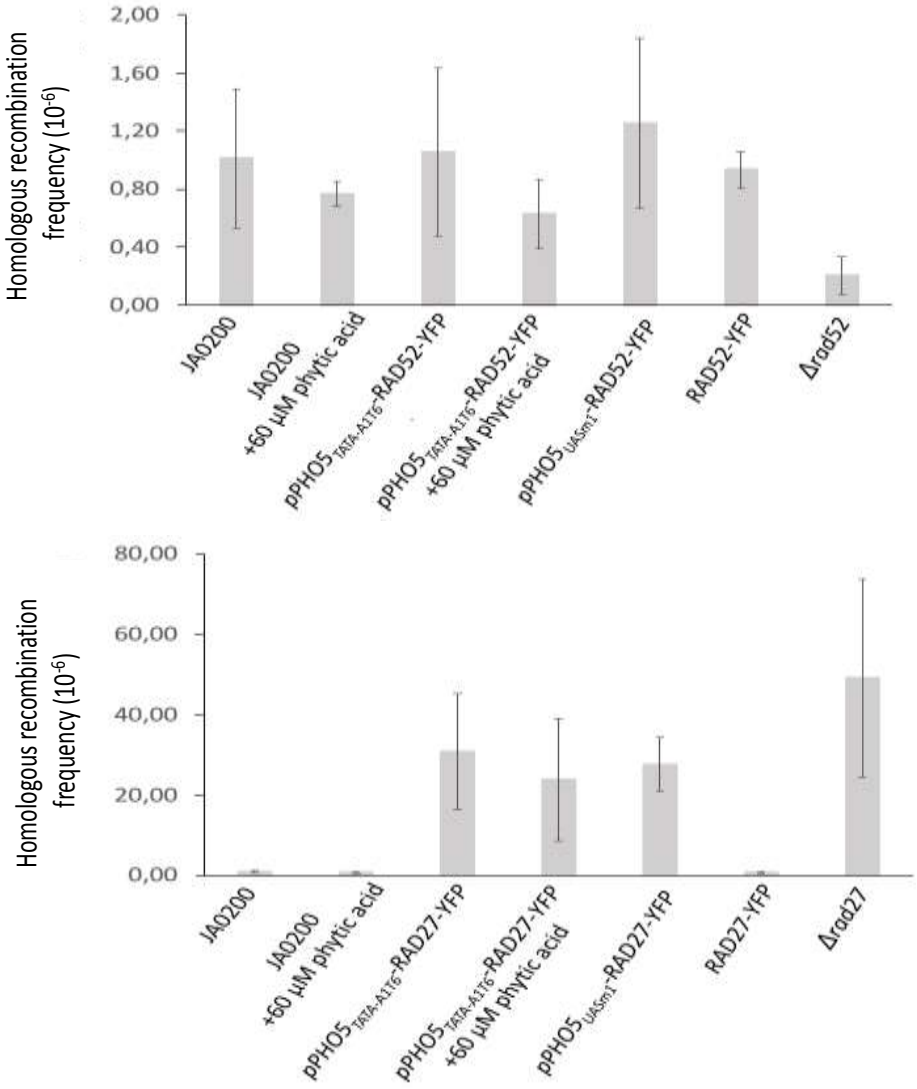


Figure 45: Homologous recombination frequency in the JA0200 derivative strains expressing different DNA recombination genes with different noise levels and similar mean expression

The strains harbouring *RAD52-YFP* under the control of the *PHO5* promoter variants exhibited the same HFR as JA0200, indicating that the cassette *pPHO5-RAD52-YFP* can compensate the *RAD52* deletion. On the contrary, the HRF of the strains harboring *RAD27-YFP* under the control of the *PHO5* promoter variants were between those of Δ*rad27* and JA0200, indicating that the cassette *pPHO5-RAD27-YFP* could only partially compensate the *RAD27* deletion. Unfortunately, we did not detect statistically phenotypic differences between the promoters confer similar expression level but different noise level.

We also transformed the plasmids containing only *YFP* under the control of the *PHO5* promoter variants into JA0200. We found that, without induction, $pPHO5_{UASm1}$ already exhibited higher expression than $pPHO5_{TATA\text{-}A1T6}$ (Figure 46). Even though this difference was undetectable when *YFP* was fused to *RAD52* or *RAD27*, we cannot ignore this effect which might generate different HRF in these strains before induction by Pi-free medium. Thus to be more accurate, we should measure the HRF before induction and subtract it from the final HRF. But this makes the experiments very complex. Thus some other interesting promoters were tried to simply the system.

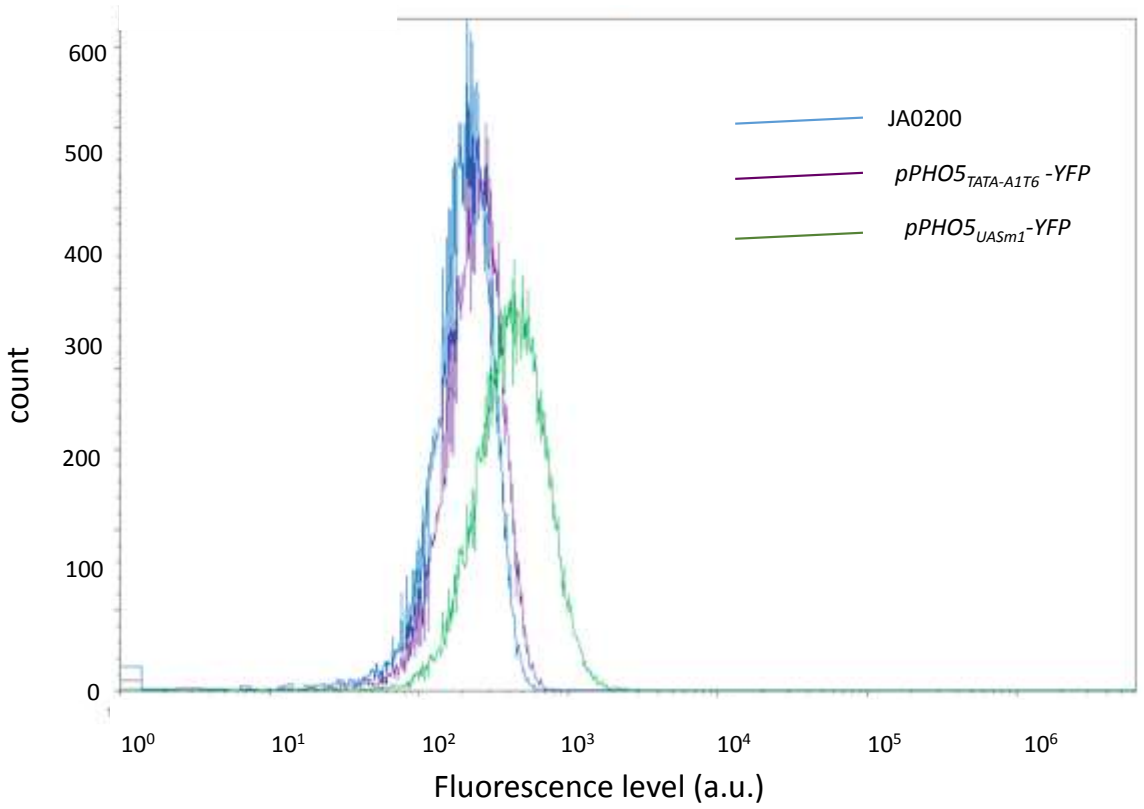


Figure 46: Expression of $pPHO5_{TATA\text{-}A1T6}$-YFP and $pPHO5_{UASm1}$-YFP without induction

4.1.1.3 Searching for new promoters conferring similar expression levels with different noise

We selected some other interesting promoters or promoter pairs from the literature which confer the same mean expression level but very different noise levels. We replaced the *PHO5* promoter variants by these new promoter pairs in the plasmids containing *RAD27-YFP* and *RAD52-YFP*. The first candidate was *pZRT2* which exhibit different noise (but similar mean) when induced by different concentrations of zinc sulfate (Carey *et al*, 2013). We selected 4 wild-type promoters (p*RNR2, pCDC19, pCCW12* and p*TDH2*) from the results of Newman *et al.* (2006) because they constitute two promoter pairs with similar mean and different noise. Finally, we selected six promoters from the library of Sharon *et al.* (2014). These promoters consist of the *HIS3* core promoter and a variable sequence containing different combinations of Gcn4 binding sites and poly (A:T) sequences.

All of these plasmids we mentioned above were transformed into JA0200 to integrate the constructions into the *LEU2* locus, and their expression level in YNB URA^- were compared in exponential phase. Finally, we found two promoters that we named P1 and P4 from Sharon *et al.* (2014) which confer similar expression levels but different noise when driving *RAD27-YFP* and *RAD52-YFP* expression (Table 6).

Table 6: Expression and noise levels of *RAD27-YFP* or *RAD52-YFP* conferred by promoter variants chosen from Sharon *et al.* (2014)

Gene	Expression level (a.u.)	Noise
*P1-GFP**	13	0. 037
*P4-GFP**	9	0.0056
P1-RAD52-YFP	1189±120	0.363±0.031**
P4-RAD52-YFP	1073±143	0.493±0.032
P1-RAD27-YFP	4872±573	0.254±0.021**
P4-RAD27-YFP	4372±620	0.363±0.018

* Data from (Sharon *et al*, 2014)

** Difference between two promoters in T-test (P<0.01)

4.1.2 Variations of the tandem repeats number in the *SDT1* promoter

The JA0300 strain was constructed by replacing the ORF of *SDT1* by *URA3* at its original locus (Figure 23). With this strain, we can perform opposite selection on the expression of *pSDT1-URA.* Using URA$^-$ medium, high expression level of *pSDT1-URA* can be selected for through modifications of the tandem repeats (TRs) number in the *SDT1* promoter (Vinces *et al*, 2009), whereas in the opposite way we expect selection for low expression level of *pSDT1-URA* in 5-FOA containing medium. Thus, this strain could be used to create fluctuating selection and see if higher noise in the expression of *RAD52* or *RAD27* increases adaptability in this fluctuating context through increasing the ability to change the number of TRs.

First, we tried to verify that the *SDT1* promoter evolves in 5-FOA towards different numbers of TRs reducing *URA3* expression. Cells grown in liquid YNB URA$^-$ medium were spread on YNB plates containing 0.5 g/L 5-FOA (which is half of the concentration used for the anti-selection of *URA3* to avoid severe mutations under strong selection). After 3 days, some clones appeared, but there were two types of clones (Figure 47): we called them big size clones (B) and small size clones (P).

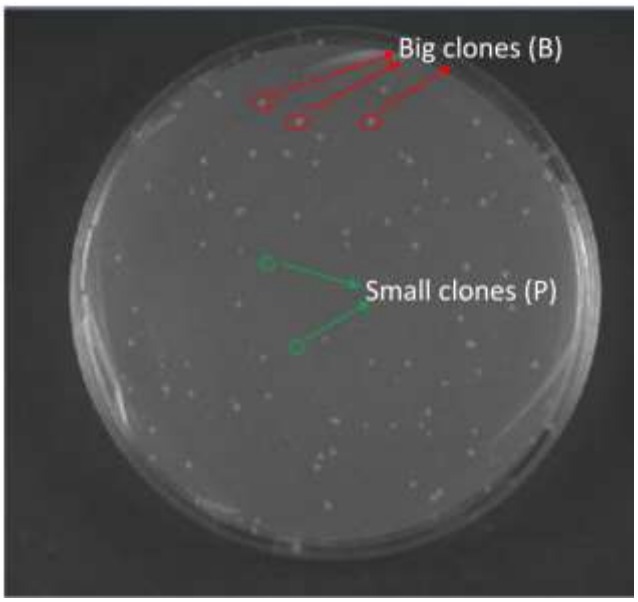


Figure 47: Two types of clones on 5-FOA plate (0.5 g/L) among the population harbouring the *pSDT1-URA3* cassette

To know if the growth difference on plate was due to differences in the TRs number in the

SDT1 promoter controlling *URA3* expression, we randomly selected 10 big and 10 small clones. The TRs region in the *SDT1* promoter of these clones was PCR-amplified and sequenced. We obtained 11 reliable sequences (Figure 48). In some clones, a decrease in the TRs number indeed occurred (up lines) compared to the initial length (C1), what is expected to decrease expression (Vinces *et al*, 2009). But some of other clones still kept the original length (B4, P1). Moreover, there were no relationships between the clone size and the change of TRs number.

Figure 48: Different TRs numbers in the *SDT1* promoter driving *URA3* expression in clones appearing on 5-FOA plate (0.5 g/L). C1 is the original strain which contains 22 TRs. The other clones are numbered randomly, while B means "big-size clone" and P means "small-size clone".

Then we tested if the promoters with short TRs can increase again their length to recover high expression of *URA3* under URA^- condition. Thus we selected clones P2 and B5 and spread them on URA^- plate. After 3 days, some clones appeared. We randomly selected 10 clones to sequence the TRs-containing region of the *SDT1* promoter and found that all the 10 clones had exactly the same TRs number as the original one (22 TRs).

4.2 Discussion

Through varying the induction conditions, we succeeded in getting similar mean expression but different noise levels for the two promoter variants $pPHO5_{UASm1}$ and $pPHO5_{TATA\text{-}A1T6}$ fused to *RAD52-YFP* or *RAD27-YFP*. But the fluorescence levels given by these fusion proteins and the difference between them are much smaller than when the promoter are fused to YFP only, indicating important post-transcriptional effects on noise regulation. This phenomenon was also observed with the promoters chosen from Sharon *et al.* (2014). Thus post-transcriptional regulation of genome maintenance genes seems to minimize their expression noise when they are produced from these promoters. These genes might be subjected to negative selection of

phenotypic heterogeneity buffering molecular variations (Richard & Yvert, 2014).

We tried to test if different expression noises in the expression of genome maintenance genes can lead to different levels of genome instability by measuring the HRF. But unfortunately we did not observe any significant difference. Considering that the cells divided only 1.5 times during the induction period in Pi-free medium, the HRF we measured were mainly controlled by the initial frequency in the population. Therefore any potential phenotypic difference generated by the induced *pPHO5* variants was not easily detectable.

We also observed different expression levels from $pPHO5_{UASm1}$ and $pPHO5_{TATA\text{-}A1T6}$ before induction. This makes the precise measurement of the effect of noise on HRF very difficult. Thus, we tried different other promoters or promoter pairs from other published researches to find suitable candidates. Only a pair of synthetic promoters created by Sharon *et al.* (2014) gave us non-induced expression levels with similar means but different noises.

We also constructed a strain (JA0300) for fluctuating experiments where the *URA3* expression is driven by *pSDT1.* The expression level conferred by *pSDT1* can be varied by changing the TRs number in the promoter through HR. The sequencing results of clones growing either on URA- or 5-FOA medium showed that the TRs number varied between these 2 environments, suggesting that we were able to select for higher *URA3* expression levels in URA^- medium and select for lower *URA3* expression level in 5-FOA medium obtained by HR between the TRs in *pSDT1*. Thus this system could be applied to test our hypothesis that high noise in the expression of genome-stability related genes (*RAD52* or *RAD27* in this case) could increase the population fitness in fluctuating environments (URA^- and 5-FOA in our system) by increasing the tunability of the RGVG.

4.3 Perspective

Among the promoters conferring similar expression levels but different noise levels that several researches have identified (Newman *et al*, 2006; Hornung *et al*, 2012; Sharon *et al*, 2012; Carey *et al*, 2013), we have already tested several pairs, but they often exhibited different behaviours comparing to the reference, mainly because we expressed proteins fused to YFP and not YFP only as most of these articles did. This shows the complexity of noise regulation. On one hand, we will continue to test some interesting promoters from literature. On the other hand, we can also perform artificial mutations on native promoters (like in section 2.3.2) to create suitable

variations.

Once we will have confirmed that we have suitable promoters, the HRF will be tested in the whole populations, but we could also sort subpopulations according to their fluorescence levels and measure the corresponding HRF, especially to test if the high noise clone has more cells exhibiting higher or lower genome instability. Moreover, these data could be used to establish precise relationships between the expression level of our genes of interest and the HRF.

Moreover, other genome-stability relating genes involved in other genome maintenance processes could also be used to test our hypothesis, such as *MSH2* and *MLH1* which are important for mismatch repair. If yeast cells lack any of these genes, the mutation rate is highly increased. In yeast, the mutation rate can be measured through the *URA3* gene: the number of clones appearing on 5-FOA plates indicates the mutation rate in *URA3*. Accurate procedures such as the method developed by Lang and Murry (Lang & Murray, 2008) allow reliable evaluation of the mutation rate. Finally, if we can prove our hypothesis, we could search in collections of natural yeast isolates if these genes do exhibit higher noise in specific environments, and if this adaptation strategy might naturally occur.

5 General conclusion

Gene expression heterogeneity, the so-called noise, in an isogenic population has been observed for decades, but finely and quantitatively measured since the beginning of the 2000s. It is an inherent property of the gene expression process. Recent researches have demonstrated that gene expression noise can be tunable according to the gene function and the cell state. It might be subjected to selection: it is shown that high expression noise is a favoured under tight selections. Expression noise also increases the phenotypic heterogeneity which is an advantage under unstable environments.

Compared to lab strains, wild and industrial yeast strains face to more complicated and stressful environments. Thus one can suppose that these strains exhibit higher expression noise in stress-related genes. In the first part of this thesis, we try to verify this hypothesis in an oenological strain EC1118. We first constructed a genomic library driving the expression of yEGFP and adapted a fluctuating selection procedure by FACS to select the genomic fragments conferring higher expression noise of yEGFP. We found that this procedure indeed enriched for high noise fragments in the library but there were still large number of negative clones. Thus we isolated single clones in the library after selection and tested their expression profiles. Finally we identify 98 clones exhibiting high expression noise of yEGFP. So combined FACS selection and single clone verification can efficiently select for the high noise promoters in yeast.

The fragments in the 98 clones were sequenced and mapped to the reference genome S288c. We found 50 potential promoters in these sequences. According to the function of the corresponding genes, we selected 8 promoters to compare the yEGFP expression profile driven by different variant in different background. The expression profile is different between different backgrounds as well as different variants, indicating the interactions between background and promoter sequence are very common and also play an important role in the generation of gene expression noise.

There is a general tendency between the mean expression level and expression noise: high expression level tends to be associated with low expression noise. But, in 59A, $pCUP1_{EC1118}$ exhibited high expression level but similar noise level as $pCUP1_{S288}$. Thus we assumed that if we can make their expression level identical, $pCUP1_{EC1118}$ would have higher noise than $pCUP1_{S288}$. So we induced the expression of *pCUP1* by different concentration of copper sulfate in different media. We showed that indeed $pCUP1_{EC1118}$ exhibits higher noise level than $pCUP1_{S288}$ in the 59A background while these two promoter variants did not show any difference in the BY4720 background. So we demonstrated that industrial strain harbour high expression noise in the

expression of certain genes (*CUP1*), but this effect needs the interaction between the promoter sequence (*cis*-effects) and the genetic background (*trans*-effects), revealing epistasis in the generation of promoter-mediated noise.

As *CUP1* confers copper resistance in yeast, one crucial question comes: does high noise expression of *CUP1* indeed increase copper resistance? Unfortunately, we cannot answer it directly because the two promoter variants are not induced at the same level by a precise copper concentration. So, we used the *zeoR* gene conferring resistance to phleomycin to study the phenotypic consequences of different noises of *pCUP1*. We demonstrated that the slightly higher noise in the expression of *zeoR* conferred by $pCUP1_{EC1118}$ compared with $pCUP1_{S288}$ is sufficient to provide a better resistance in a range of phleomycin concentrations. Moreover there are often many *CUP1* copies in the yeast genome. In general, increasing the *CUP1* copy numbers can increase copper resistance, but large number of *CUP1* might decrease the fermentation ability. Thus increasing copper resistance through higher noise expression would be a compromise between these two aspects.

During the selection of high noise clones in the library, we found some clones which exhibited a bimodal expression profile, thus the second part of this thesis we tried to understand the underlying mechanisms of bimodal expression. We found that centromeric plasmid allow bimodal expression whereas genomic integration in the *LEU2* locus generally suppress this property. One possible explanation is the difference between the genome and centromeric plasmids in terms of nucleosome binding and modifications. We also found that *pCUP1-YFP* in BY4720 could harbour a bimodal expression profile at the genomic level when it is induced by copper sulfate. Because Cup2 is the major transcription factor that regulates the expression of *CUP1*, we suppose that *CUP2* might play a role in this bimodality.

The third part of this thesis tried to verify the hypothesis that high noise in genome stability related genes can increase the rate of genetic variant generation, what would be an advantage in stressful and fluctuating environments, and thus beneficial for rapid adaptation. We first focused on the homologous recombination pathway and two genes which impact this pathway (RAD52 and RAD27). Unfortunately no consequences on genome stability of a difference of noise in the expression of these genes have been found yet.

Material and methods

6 Material and methods

6.1 Strains, primers, plasmids and media

6.1.1 Strains

6.1.1.1 Bacteria

The *E. coli* strain DH 5-alpha (High efficiency, New England Biolabs) was used in this study to amplify plasmids by chemical transformation or electroporation transformation (see below).

6.1.1.2 Yeast

All the *Saccharomyces cerevisiae* strains we used and constructed in this study are listed in the Appendix VI.

6.1.1.2.1 Strains used in section 2 and 3

The genomic library was constructed in CEN.PK (MATa *ura3-52*, see below). The pJRL2 derivative plasmids (see below 5.1.2) containing selected promoters fused to *yEGFP* were integrated into the *LEU2* locus of BY4720 (S288c auxotrophic derivative, *MATα lys2Δ0 trp1Δ63 ura3Δ0*) and 59A *MATα Δamn1-loxP*. 59A is a haploid derivative of EC1118 which exhibits high level flocculation that do not allow analysis by flow cytometry. So we used the *Δamn1* 59A (provided by V. Galeote, INRA SupAgro) that no longer flocculates because of the deletion of *AMN1* encoding a cell wall protein responsible for the flocculation.

6.1.1.2.2 Strains used in section 4

The JA0200 strain was created by adding the intact *LEU2* gene into the KV133 strain (BY4742 MATα *his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0, FLO1::URA3*, see Verstrepen *et al*, 2005) at its original locus to then use the pJRL2 plasmids recombining in *LEU2* (Raser & O'Shea, 2005). The JA0300 strain was created by replacing *SDT1* by *URA3* in BY4720. Then the pJRL2 derivative plasmids containing *RAD52-YFP* and *RAD27-YFP* under the control of two mutated *PHO5* promoters conferring similar mean expression level with different noise were integrated into the *LEU2* locus in JA0200 and JA0300. The original *RAD52* and *RAD27* genes were deleted afterwards by using the PCR-amplified *LYS2* gene with short sequences of homology with *RAD52* or *RAD27* at the end of the primers (the *RAD52* and *RAD27* deletions affect the transformation efficiency, thus we first integrated plasmids

and then deleted the original gene).

6.1.2 Plasmids

The pJRL2 plasmids from Raser and O'shea (2005) were used for genomic integration in the *LEU2* locus. Nevertheless the selection cassette was modified: the *his-URA3-kanR-his* sequence was replaced by *kanMX4* only isolated from the pFA6a-GFP-kanMX4 vector (using BglII and KpnI). Moreover the *YFP* gene where a Kozak sequence had been introduced has been replaced by *yEGFP* without Kozak sequence after PCR amplification from the pUG35 vector and insertion within pJRL2 using EcoRI and NotI (Figure 49). The 2 variants of all the 8 promoters studied at the genomic level in part 2 were PCR-amplified with SalI and EcoRI sites in the primers to replace the original promoter sequence driving GFP expression in the pJRL2 plasmid. *ZeoR*, *RAD52-YFP* and *RAD27-YFP* were PCR-amplified with EcoRI and NotI sites in the primers to replace the *yEGFP* gene in the pJRL2 plasmid. *ZeoR* was amplified from plasmid pZE2 (a gift from LBME (Laboratory of Eukaryotic Molecular Biology, Toulouse)). *RAD52-YFP* and *RAD27-YFP* was amplified from the genome DNA of their N-terminal YFP fusion strain (see below).

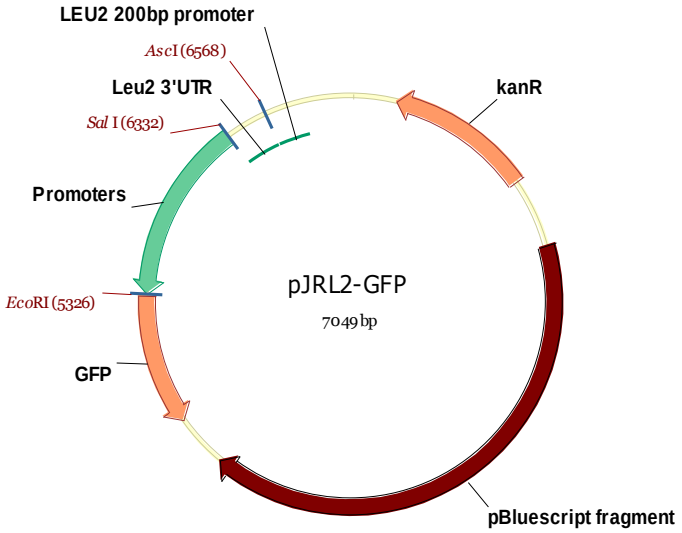


Figure 49: Structure of the pJRL2 plasmids

The plasmids pFA6a-*yEGFP-kanMX6* and pFA6a-*YFP-kanMX6* were used to create N-terminal fusion of *RAD52*, *RAD27* or *CUP1* with *GFP* or *YFP* at their original chromosomal locus. All plasmids are listed in Appendix VII.

6.1.3 Primers

The primers are listed in Appendix VIII.

6.1.4 Media and culture conditions

6.1.4.1 E. coli

E. coli were grown in LB medium. It contains: 1% tryptone, 1% yeast extract, 0,5% NaCl. 50 μg.ml^{-1} ampicillin (150 mg.ml^{-1} as stock) was added to select for transformants with the *ampR* gene. Solid media contain 2% agar. Ampicillin was added when the media were below 60 °C after autoclaving.

6.1.4.2 S. cerevisiae

Yeast strains were grown in YPD or YNB medium. YPD medium contains: 2% glucose, 1% peptone, 1% yeast extract. 200 μg.ml^{-1} G418 (100 mg.ml^{-1} as stock) was added to select for transformants with the *kanR* gene. 100 μg.ml^{-1} nourseothricin (100 mg.ml^{-1} as stock) was added to select for transformants with the *natR* gene. YNB medium contains: 2% glucose, 0.17% yeast nitrogen base, 0.5% ammonium sulfate. 0.077% complete supplement mixture (CSM) minus URA was added to select for *URA3* transformants. 0.069% CSM minus LEU was added to select for *LEU2* transformants. 0.074% CSM minus LYS was added to select for *LYS2* transformants. Solid media contain 2% agar. All the antibiotics were added when the media were below 60 °C after autoclaving.

Pi-free medium contains five parts: (1) carbon source: 2% glucose; (2) nitrogen source: 0.5% ammonium sulfate; (3) minerals: 0.055% potassium chloride, 0.05% magnesium sulfate, 0.01% sodium chloride and 0.1% calcium chloride; (4) vitamins: 2 μg.L^{-1} biotin, 400 μg.L^{-1} calcium pantothenate, 2 μg.L^{-1} folic acid, 2000 μg.L^{-1} inositol, 400 μg.L^{-1} niacin, 200 μg.L^{-1} p-Aminobenzoic acid, 400 μg.L^{-1} Pyridoxine hydrochloride, 200 μg.L^{-1} riboflavin and 400 μg.L^{-1} thiamine hydrochloride; (5) trace elements: 500 μg.L^{-1} boric acid, 40 μg.L^{-1} copper sulfate, 100 μg.L^{-1} potsssium Iodide, 200 μg.L^{-1} ferric chloride, 400 μg.L^{-1} manganese sulfate, 200 μg.L^{-1} sodium molybdate and 400 μg.L^{-1} zinc sulfate. Vitamins and trace elements were prepared as 1000× stock and sterilized through 0.22 μm filtering. The carbon source, nitrogen source and minerals were mixed and autoclaved. 0.077% CSM minus URA was added to select for the *URA3* gene.

Plates containing 5-FOA were prepared as follows: Solutions I containing 4% glucose and 4% agar was autoclaved, solution II containing 0.34% yeast nitrogen base, 1% ammonium sulfate, 0.158% CSM and 0.2% 5-FOA was sterilized through 0.22 μm filter. Same volumes of these two solutions were mixed when solution I was below 60 °C.

6.1.4.3 Culture conditions

E. coli cells were grown in liquid media at 37 °C with shaking speed of 170 rpm/min, or on solid plate at 37 °C in an incubator. Yeast cells were grown in liquid medium at 30 °C with shaking speed of 200 rpm/min, or on solid plate at 30 °C in an incubator.

6.2 Basic molecular biology methods

6.2.1 DNA manipulation

6.2.1.1 DNA extraction

Yeast genomic DAN was extracted using the MasterPure™ Yeast DNA Purification Kit (Epicentre), following its standard protocol in the manual. Plasmids were extracted from *E.coli* using the GeneJET Plasmid Miniprep Kit (Thermo), following its standard protocol in the manual.

Extraction of plasmids from yeast was performed as follow: 1.5 ml overnight culture were centrifuged before the pullets were resuspended with 0.2 ml extraction mixture (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-Cl (pH 8), 1 mM Na2-EDTA). Then 0.3g glass beads and 0.2 ml of phenol:chloroform:isoamyl alcohol (25:24:1) were added to the tube. The tube was vortex for 2 min and centrifuged for 5 min afterwards. The aqueous layer was collected for further usage.

6.2.1.2 PCR

PCR was carried out on the My cycler^{TM} thermal cycler (Bio-Rad). PCR reaction tubes contained (50 μl): 23.5 μl H_2O, 10 μl Phusion High-Fidelity (HF) buffer (NEB), 5 μl dNTP (2.5 mM), 5 μl forward primer (2.5 μM), 5 μl reverse primer (2.5 μM) and 0.5 μl Phusion HF DNA Polymerase (NEB).

The PCR cycles were: 1, initial denaturation at 98 °C for 30 seconds; 2, denaturation at 98 °C for 20 seconds; 3, annealing for 30 seconds; 4, extension at 72 °C for a certain time according to the length of the target (30 seconds per kb); steps 2 to 4 were repeated for 30 cycles; 5, final extension at 72 °C for

5 min.

6.2.1.3 DNA cloning

The plasmids and PCR products were digested by the enzymes we have choosen with their proper buffer at their optimal temperature. The target fragments were separated by electrophoresis and purified through the GeneJET Gel Extraction Kit (Thermo), following its standard protocol in the manual. The digested plasmids and PCR products were ligated by T4 ligase (NEB) at room temperature (RT) overnight. The final plasmids were transformed to *E.coli* and verified by sequencing. For the fragments containing several construction sites inside them, we used the In-Fusion® HD Cloning Kit (Clontech) instead, following its stand protocol in the manual.

6.2.2 Transformation

6.2.2.1 Bacteria

6.2.2.1.1 Chemical transformation

i) Competent cells preparation

An overnight culture was diluted to OD 0.1. The fresh culture was incubated at RT with shaking for 6h until the OD reach 0.6 to 0.8. The final culture was placed on ice for 10 min and centrifuged at 3000 rpm for 10 min at 4 °C. The pullets were suspended with 1⁄8 volume ice-cold TB buffer (10 mM pipes, 55 mM MnCl2, 15 mM CaCl2, 250 mM KCl, pH = 6.7). The suspension were kept on ice for 10 min and centrifuged for 10 min. The pullets were re-suspended with 1⁄12 volume TB buffer containing 7% DMSO. 100 µl aliquots were taken and thrown in liquid nitrogen immediately. The aliquots were stored at -80 °C.

ii) Transformation

The competent cells were taken out from -80 °C and thawed on ice. 2µl plasmid DNA were added and mixed gently. The suspension was kept on ice for 5 min. Then it was put into 42 °C bath for 45 seconds. Then the suspension was put on ice for another 5 min. 1ml LB medium was added to the suspension and was incubated for 1 h at 37 °C with shaking before it was spread on plates with ampicillin. The plates were incubated at 37 °C for one night.

6.2.2.1.1 Electroporation transformation

i) *Competent cells preparation*

An overnight culture was diluted 100 times. The fresh culture was incubated at RT with shaking for 8h until the OD reach 0.5 to 0.8. The final culture was placed on ice for 15 min and centrifuged at 3000 rpm for 15 min at 4 °C. The pullets were suspended with the same volume of water. The suspension was centrifuged for another 15 min. The pullets were re-suspended with half volume water and centrifuged as above. Then 1/50 volume 10% glycerol solution was used to suspend the pullets. This suspension was also centrifuged as above. Finally, 1/500 volume 10% glycerol solution was used to suspend the pullets. 40 µl aliquots were taken and thrown in liquid nitrogen immediately. The aliquots were stored at -80 °C.

ii) Transformation

The competent cells were taken out from -80 °C and thawed on ice. The suspension was mixed with 2µl plasmid DNA and transferred to a cold 1.5mm cuvette. The electroporation was performed on a Bio-Rad Gene Pulser (25 µF, 2.5 kV and 200 W). 1 ml LB was added immediately with gentle mixing. The suspension was incubated for 1 h at 37 °C with shaking before it was spread on plates with ampicillin. The plates were incubated at 37 °C for one night.

6.2.2.2 Yeast

An overnight culture was diluted to OD 0.5. The fresh culture was incubated at 30 °C with shaking for about 4h (≈2 divisions, OD≈2). The final culture was centrifuged at 5000 rpm for 5 min (5ml culture per transformation). The pullets were washed by water and centrifuged again as above. 1.0 ml 100 mM LiAc were used to re-suspend the pullets. The suspension was centrifuged at 13000 rpm for 30 sec. The following mixture was added to the pullets: 240 µl PEG (50% w/v); 36 µl 1.0 M LiAc; 50 µl ss-DNA (2.0 mg/ml, denatured at 95 °C for 5 min and kept on ice); X µl Plasmid DNA or PCR products (0.1 - 10 µg, the pJRL2 plasmids have to be digested by AscI before transformation) and 34-X µl sterile H_2O (respecting the order above). The cells had to be well mixed with these solutions. The suspension was first incubated at 30 °C for 30 min and then at 42 °C for another 30 min. Finally the suspension was centrifuged at 8000 rpm for 30 sec. If auxotrophic marker was applied, the pullets were re-suspended with 200 µl water and spread on corresponding plates. If an antibiotic agent was applied, the pullets were re-suspended with 500 µl YPD and incubated at RT for one night. They were spread on corresponding plates the next day. The plates were incubated at 30 °C for at least 2 days. The

integration of integrative plasmids was verified by PCR.

6.2.3 Directed mutagenesis of the *CUP1* promoter

6.2.3.1 Point mutations

We used directed mutagenesis to insert three SNPs to the $pCUP1_{S288c}$ variant. Primers with one of these SNPs (Appendix VIII) were used to amplify the plasmid with $pCUP1_{S288c}$ by PCR. The original plasmids were then digested by DpnI. The final plasmids with one point mutation were transformed to *E. coli* and verified by sequencing.

6.2.3.2 Deletions

The pJRL2 plasmids containing $pCUP1_{S288c}$-*yEGFP* and $pCUP1_{EC1118}$-*yEGFP* were digested with XbaI (cutting between SNP3 and the deletion) and SacI (cutting downstream *yEGFP*) and the deletion was introduced into $pCUP1_{S288c}$ by exchanging the fragments.

6.2.4 RT-qPCR

6.2.4.1 RNA extraction

5ml yeast culture at OD ≈ 1 was centrifuged at 3000 rpm for 5 min. The cells were washed by water and centrifuged as above. The pullets were thrown in liquid nitrogen immediately and stored at -80 °C. The total RNAs were extracted through SV Total RNA Isolation Kits (Promega), following its standard protocol in the manual. The concentrations were measured by Nano-drop (Thermo). The quality of RNA was controlled by RNA 6000 Lab-on-Chip bioanalyzer (Agilent).

6.2.4.2 RT-qPCR

1µg total RNA was mixed with iScript™ Reverse Transcription Supermix (Bio-Rad) in 20 µl final. The reverse transcription was performed as describing in the manual. This system was diluted 10 times for qPCR analysis.

Real-time PCR were carried out on the iCycler iQ5 thermal cycler (Bio-Rad). The qPCR system contains: 4 µl diluted RT system, 2 µl forward primer (2 µM), 2 µl reverse primer (2 µM), 2 µl water and 10 µl SsoAdvanced™ Universal Supermixes (Bio-Rad). The PCR cycles: 1, 98 °C 30 sec; 2, 98 °C 10 sec; 3, 54 °C 30 sec; steps 2-3 were repeated for 40 cycles; 4, 98 °C 5 min; 5, increasing from 55

°C to 90 °C by 0.5 °C per sec. The data were analyzed by CFX Manager™ software (Bio-Rad). *TAF10, UBC6* and *ALG9* were used as reference genes to normalize the gene expression level.

6.3 Flow cytometry

6.3.1 FACS selection of high noise promoters

6.3.1.1 Construction of the yEGFP-fused genomic DNA library

6.3.1.1.1 Construction of the promoterless yEGFP-coding vectors

Three different promoterless *yEGFP*-coding vectors (Figure 50) were constructed using the *yEGFP*-coding pUG35 centromeric plasmid as a backbone vector. The *MET25* promoter and the multiple cloning site upstream yEGFP of this vector have been replaced by the kanMX4 gene and a unique SnaBI restriction site (generating blunt ends). The *kanMX4* gene was PCR amplified from the pFA6 vector to recombine within pUG35. The forward primer contained homology to the beginning of the *MET25* promoter of pUG35. The reverse primers contained homology to the end of the multiple cloning site of pUG35. Three different reverse primers have been used to construct three different plasmids containing 0, one or two additional base(s) between SnaBI and the start codon of *yEGFP*. The resulting plasmids were promoterless *yEGFP*-coding vectors containing a SnaBI restriction site before the start codon of *yEGFP* to fuse genomic fragments to *yEGFP*.

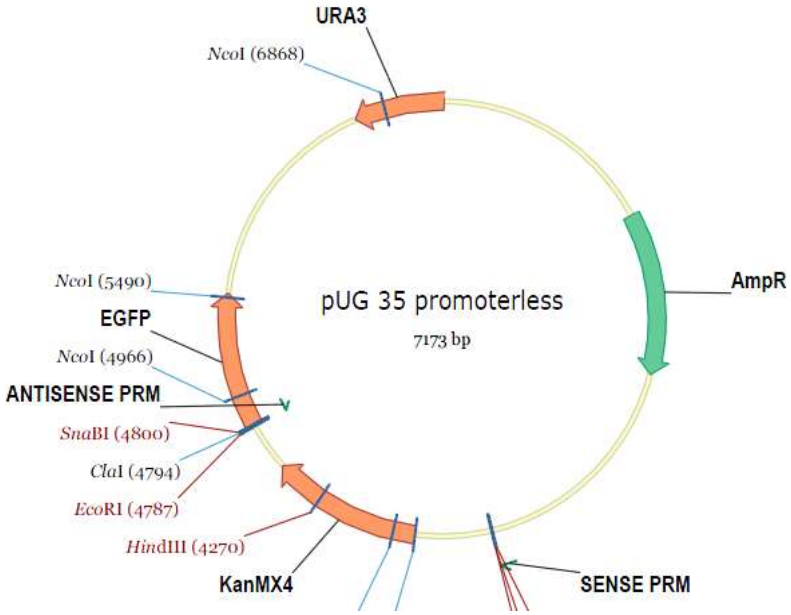


Figure 50: Structure of the promoterless yEGFP-coding vectors

6.3.1.1.2 Construction of the genomic library

The 59A genomic DNA has been fragmented independently by the two 4-cutter restriction enzymes RsaI and AluI generating blunt ends compatible with the ends generated by SnaBI in the promoterless yEGFP-coding vectors. Reaction times and enzyme concentrations have been optimized to produce DNA fragments ranging from 500pb to 3kb. For RsaI, 1,5µg DNA has been digested during 15min by 1 U enzyme. For AluI, 3µg DNA have been digested during 30min by 0,5 U enzyme.

Fragments from 500pb to 3kb generated by each enzyme have been extracted from gel (QIAquick Gel Extraction Kit, Qiagen) and ligated independently by the DNA ligase T4 (overnight, 16°C, Quick Ligation Kit, New England Biolabs) with each of the three promoterless yEGFP-coding vectors previously digested by SnaBI, dephosphorylated (Antarctic Phosphatase, New England Biolabs), and purified. The ratio (vector:inserts) used for the 6 ligation reactions (2 enzymes, 3 vectors) was 1:2,6 because it gave the higher number of transformants. The number of transformants required for each (enzyme:vector) pair to give a 99% confidence level that all sequences of the genome are represented with a mean insert size of 2kb was 30000. This number has been multiplied by 2 because any fragment can be inserted in both senses. Thus 60000 transformants for each (enzyme:vector) pair have been independently obtained after ligation, transformation of competent *E. coli* cells using standard methodology, and growth on selective LB medium. Then cells were harvested and pooled to isolate the plasmids from the 6 bulk cultures (GenElute™ HP Plasmid Midiprep Kit, Sigma-Aldrich). Redigestion with *SnaBI* has been performed to linearize empty promoterless *yEGFP*-coding plasmids. Plasmids from each (enzyme:vector) pair have then been retransformed in the laboratory yeast strain CEN-PK using classical lithium acetate method. Again 60000 transformants for each (enzyme:vector) pair have been independently obtained after growth in selective medium (YNB URA^-). Finally the transformants originated from the 6 (enzyme:vector) pairs have been pooled together at similar OD and equal volume to form the final library used for fluctuating selection.

6.3.1.2 Fluctuating selection using cell sorting

The method described by Freed *et al.* (2008) has been adapted to *Saccharomyces cerevisiae*. An overnight culture of the population containing the genomic library was diluted to OD=0.5 and cells were grown for around 5 hours to reach exponential growth (OD=2) (YNB URA^-). Cultures were spun down at 3000g for five minutes at 4°C. Growth media was removed and cultures were

re-suspended in ice cold PBS. Cells were then kept on ice until cell sorting. The *yEGFP*-fused genomic library was subjected to fluctuating selection on fluorescence intensity, where selection for bright cells alternated with selection for dim cells using fluorescence-activated cell sorting (FACS) performed by the FACS Calibur associated to the Cellquest™ sorting software (Becton Dickinson). On the first day, a gate was drawn to include the highest 5% of *yEGFP* expression. 1×10^5 cells were collected into a sterile Falcon tube. Cells were collected at medium flow rate and sorted on the basis of "single cell" and "purity". After sorting, cells were spun at 3000g for ten minutes and any FACS buffer was removed. Cells were re-suspended in 1ml YNB URA^- medium and grown overnight. The following day the process was repeated but the gate included only the lowest 5% of cells expressing *yEGFP*. This process was repeated for a total of seven rounds of selection, with gates being drawn for selected populations in a fluctuating manner with alternatively the highest or the lowest 5% of *yEGFP* expression in the gate. After the 4th round of selection cells were placed at 4°C for 48 hours. After this time, selection was resumed as normal until the 7th round.

6.3.2 Flow cytometry analysis

6.3.1.1 Enrichment of high noise clones

The library after FACS selection was plated on YNB URA^- agar plates, and single colonies were randomly selected to confirm the enrichment in clones with high noise in *yEGFP* expression. Because not all the clones harbored noisy *yEGFP* expression, a screening was needed to sequence plasmids from the noisiest clones only. Single clones were randomly selected, grown in 96-well plates overnight in YNB URA^- medium. 10^5 cells of each clone were analyzed for *yEGFP* expression on FACS Calibur (Becton Dickinson). Analysis of cytometry data was performed by the Cellquest™ software (Becton Dickinson). Calculation of variation in *yEGFP* expression was performed as followed to limit the influence of cellular aggregates, cell detritus, and undefined values: for each clone, a gate was created on the forward scatter (FSC) and side scatter (SSC) dot plot to exclude extreme or zero values from total counts and to include only a population of cells homogeneous in terms of size, shape, and cellular complexity. A single gate size was chosen for all analyses in order to maintain a conservative estimate of noise. The coefficient of variation (CV) was calculated for fluorescence in this gate. For some clones, a smaller gate was also applied on the densest subset of cells using the FSC/SSC density plot. This gating lowered average CV values because it minimizes "extrinsic" noise due to physiological differences between cells. This allowed verifying that CV was

mainly due to “intrinsic” noise. Clones with mean expression level greater than 30 or noise level (CV) greater than 120% were chosen. Two individual sub-clones were further re-isolated from each selected clone to analyze if the noise conferred by the genomic fragment was a stable property of the plasmid.

6.3.1.2 Measurement of fluorescence expression at genomic level

Analysis of mean and noise levels at the genomic level after integration to the *LEU2* locus (all the pJRL2 derivatives) was performed on the Attune™ Acoustic Flow Cytomoter (Life Technologies). Cells at exponential stage were collected and washed by PBS before analysis. For each strain, 10^5 cells were analyzed for *yEGFP* expression. Analysis of cytometry data was performed by the Attune™ software (Life Technologies). A gate containing at least 10^4 cells for robust analysis was applied on the densest subset of cells using the FSC/SSC density plot. The same gate has been used to measure mean and noise levels conferred by the variants of a given promoter in a given strain in order to maintain a conservative estimate of noise.

6.4 Sequencing and bioinformatics analysis

Plasmids from 97 individual clones harboring noisy *yEGFP* expression have been extracted using standard phenol-chloroform extraction method and inserted fragments have been sequenced using a primer hybridizing in the yEGFP gene 75 bp upstream the start codon. Sequencing was performed using an Applied Biosystems 3730xl DNA Analyzer. Base calling was performed using TraceTuner 3.0.4 beta (Denisov *et al*, 2004) to obtain fasta and quality values. Vector and quality trimming were performed using Lucy 1.19p (Chou & Holmes, 2001). Only 96 reads were retained after trimming. Mapping reads was performed using SMALT 0.7.3 (http://www.sanger.ac.uk/resources/software/smalt/) resulting in 95 reads correctly mapped to the S288c genome. Variants were obtained using the mpileup command of SAMtools 0.1.18 (Li *et al*, 2009) and further filtered to keep those found upstream of ORF using a custom perl script.

GO analysis has been performed using SGD Gene Ontology Term Finder (http://www.yeastgenome.org/cgi-bin/GO/goTermFinder.pl). Sequence alignments have been performed using the MultAlin on-line software (http://multalin.toulouse.inra.fr/multalin/).

6.5 Study of the *CUP1* promoter variants

6.5.1 Induction of *pCUP1-yEGFP* by copper

An overnight culture was diluted to OD=0.3 and cells were grown in YPD or YNB medium until exponential phase (4~5 h) before adding $CuSO_4$ (ProLabo). Time-dependent induction was measured in 20µM $CuSO_4$ during up to 3h and concentration-dependent induction studies was measured after 1h in concentrations up to 50µM $CUSO_4$. Strains are then analyzed by flow cytometry as described above. When induction conditions giving similar mean expression levels for the different *CUP1* promoter variants have been determined, experiments were reproduced at fixed time and concentration for a given variant. Overnight induction was measured after dilution of a copper-induced overnight culture to OD=0.3 and growth in the same $CuSO_4$ concentration during 5h.

6.5.2 Growth in phleomycin-containing medium

Individual colonies of *ZeoR*-expressing strains were used to inoculate YNB medium, and strains were grown overnight either with appropriate $CuSO_4$ concentrations to get steady-state induction at the same mean level (10µM $CuSO_4$ for $pCUP1_{S288c}$-*ZeoR* and 5µM for $pCUP1_{S288c}$-*ZeoR*) or in the absence of $CuSO_4$. The expression levels of *ZeoR* were verified by RT-qPCR described as above.

After dilution to OD=0.2, cultures were grown 5h in the same culture conditions as overnight (either with the same $CuSO_4$ concentrations or in the absence of $CuSO_4$) prior to phleomycin exposure. Then these cultures was diluted 100 times with the same media with the same $CuSO_4$ concentrations and divided into 11 aliquots. Appropriate volume of phleomycin solution (Invivogen) was added to generate a series of cultures containing 0 to 100 $\mu g.ml^{-1}$ phleomycin. This series was inoculated at 30°C with 200 rpm shaking and the OD of each tube was measured after 24h. The residual of growth was calculated as the percentage of OD with certain phleomycin comparing to OD without any phleomycin. For experimental growth time course at 40, 50 and 60 µg.ml-1 phleomycin or without phleomycin, OD were followed during 35h to draw the growth curve. All these experiments were repeated at least 3 times.

6.6 Measurement of the homologous recombination frequency

6.6.1 Induction of the *PHO5* promoters

The induction of *PHO5* promoters was modified from Raser and O'Shea (2004). An overnight cultures (YNB URA^-) was diluted to OD=0.5 by phosphate-free medium with different concentrations of phytic acid (cells were washed at least 4 times by water before inoculating). These fresh cultures were incubated at 30 °C with 200 rpm shaking. The expression level of YFP at different time points were measured as described above. The conditions conferring similar expression level with different expression noises for the two *PHO5* promoter variants were chosen for further analysis (60 μM phytic acid for TATA-A1T6 and without any phytic acid for UASm1, both at 7h induction).

6.6.2 Measurement of homologous recombination frequency

The cell density of the cultures was measured by flow cytometry. 100 cells were spread on YPD plate to measure the viability of each strain. 10^6, 10^7 and 10^8 (Nt) cells were spread on 5-FOA plate (the plate which had around 100 clones on the plate was taken to calculate the homologous recombination frequency (HRF)). The number of clones appeared on the YPD plates (N1) and 5-FOA plates (N2) after 2 days were used to calculate the frequency of homologue recombination (FHR): f=(N2/Nt)*(N1/100). The HRF before and after induction were both measured.

Bibliography

Abranches E, Guedes AM V, Moravec M, Maamar H, Svoboda P, Raj A & Henrique D (2014) Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. *Development* **141:** 2770–9

Acar M, Becskei A & van Oudenaarden A (2005) Enhancement of cellular memory by reducing stochastic transitions. *Nature* **435:** 228–32

Acar M, Mettetal JT & Van Oudenaarden A (2008) Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* **40:** 471–475

Ackermann M (2013) Microbial individuality in the natural environment. *ISME J.* **7:** 465–7

Adamo GM, Lotti M, Tamás MJ & Brocca S (2012) Amplification of the CUP1 gene is associated with evolution of copper tolerance in Saccharomyces cerevisiae. *Microbiology* **158:** 2325–35

Aksenova AY, Greenwell PW, Dominska M, Shishkin AA, Kim JC, Petes TD & Mirkin SM (2013) Genome rearrangements caused by interstitial telomeric sequences in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **110:** 19866–71

Albertin W, Marullo P, Aigle M, Bourgais A, Bely M, Dillmann C, DE Vienne D & Sicard D (2009) Evidence for autotetraploidy associated with reproductive isolation in Saccharomyces cerevisiae: towards a new domesticated species. *J. Evol. Biol.* **22:** 2157–70

Ambrona J, Vinagre A & Ramírez M (2005) Rapid asymmetrical evolution of Saccharomyces cerevisiae wine yeasts. *Yeast* **22:** 1299–306

Ambroset C, Petit M, Brion C, Sanchez I, Delobel P, Guérin C, Chiapello H, Nicolas P, Bigey F, Dequin S & Blondin B (2011) Deciphering the molecular basis of wine yeast fermentation traits using a combined genetic and genomic approach. *G3 (Bethesda).* **1:** 263–81

Anderson MZ, Gerstein AC, Wigen L, Baller JA & Berman J (2014) Silencing Is Noisy: Population and Cell Level Noise in Telomere-Adjacent Genes Is Dependent on Telomere Position and Sir2. *PLoS Genet.* **10:** e1004436

Ang K, Ee G, Ang E, Koh E, Siew WL, Chan YM, Nur S, Tan YS & Lehming N (2012) Mediator acts upstream of the transcriptional activator Gal4. *PLoS Biol.* **10:** e1001290

Ansari KI (2009) MLL histone methylases in gene expression, hormone signaling and cell cycle. *Front. Biosci.* **Volume:** 3483

Ansel J, Bottin H, Rodriguez-Beltran C, Damon C, Nagarajan M, Fehrmann S, François J & Yvert G (2008) Cell-to-cell stochastic variation in gene expression is a complex genetic trait. *PLoS Genet.* **4:** e1000049

Arai N, Kagawa W, Saito K, Shingu Y, Mikawa T, Kurumizaka H & Shibata T (2011) Vital roles of the second DNA-binding site of Rad52 protein in yeast homologous recombination. *J. Biol. Chem.* **286:** 17607–17

Ardehali MB & Lis JT (2009) Tracking rates of transcription and splicing in vivo. *Nat. Struct. Mol. Biol.* **16:** 1123–4

Aso T, Lane W, Conaway J & Conaway R (1995) Elongin (SIII): a multisubunit regulator of elongation by RNA polymerase II. *Science (80-.).* **269:** 1439–1443

Ayyagari R, Gomes X V., Gordenin DA & Burgers PMJ (2002) Okazaki Fragment Maturation in Yeast: I. DISTRIBUTION OF FUNCTIONS BETWEEN FEN1 AND DNA2. *J. Biol. Chem.* **278:** 1618–1625

Bajić D & Poyatos JF (2012) Balancing noise and plasticity in eukaryotic gene expression. *BMC Genomics* **13:** 343

Barbaric S, Münsterkötter M, Goding C & Hörz W (1998) Cooperative Pho2-Pho4 interactions at the PHO5 promoter are critical for binding of Pho4 to UASp1 and for efficient transactivation by Pho4 at UASp2. *Mol. Cell. Biol.* **18:** 2629–39

Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y & Barkai N (2006) Noise in protein expression scales with natural protein abundance. *Nat. Genet.* **38:** 636–43

Basehoar AD, Zanton SJ & Pugh BFF (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116:** 699–709

Batada NN & Hurst LD (2007) Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat. Genet.* **39:** 945–9

Beaumont HJE, Gallie J, Kost C, Ferguson GC & Rainey PB (2009) Experimental evolution of bet hedging. *Nature* **462:** 90–3

Becskei A, Kaufmann BB & van Oudenaarden A (2005) Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat. Genet.* **37:** 937–44

Bendall SC, Simonds EF, Qiu P, Amir ED, Krutzik PO, Finck R, Bruggner R V, Melamed R, Trejo A, Ornatsky OI, Balderas RS, Plevritis SK, Sachs K, Pe'er D, Tanner SD & Nolan GP (2011) Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332:** 687–96

Bengtsson M, Ståhlberg A, Rorsman P & Kubista M (2005) Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Res.* **15:** 1388–92

Ben-Zvi AP & Goloubinoff P (2001) Review: mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. *J. Struct. Biol.* **135:** 84–93

Bhattacharyya S, Yu H, Mim C & Matouschek A (2014) Regulated protein turnover: snapshots of the proteasome in action. *Nat. Rev. Mol. Cell Biol.* **15:** 122–33

Blake WJ, Balázsi G, Kohanski M a, Isaacs FJ, Murphy KF, Kuang Y, Cantor CR, Walt DR & Collins JJ (2006) Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* **24:** 853–65

Blake WJ, KAErn M, Cantor CR & Collins JJ (2003) Noise in eukaryotic gene expression. *Nature* **422:** 633–7

Brand AH & Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118:** 401–15

Braunstein M, Rose AB, Holmes SG, Allis CD & Broach JR (1993) Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes Dev.* **7:** 592–604

Breathnach R & Chambon P (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50:** 349–83

Brion C, Ambroset C, Sanchez I, Legras J-L & Blondin B (2013) Differential adaptation to multi-stressed conditions of wine fermentation revealed by variations in yeast regulatory networks. *BMC Genomics* **14:** 681

Brown CR & Boeger H (2014) Nucleosomal promoter variation generates gene expression noise. *Proc. Natl. Acad. Sci. U. S. A.*

Brown CR, Mao C, Falkovskaia E, Jurica MS & Boeger H (2013) Linking Stochastic Fluctuations in Chromatin Structure and Gene Expression. *PLoS Biol.* **11:** e1001621

Burke TW & Kadonaga JT (1996) Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* **10:** 711–24

Butler JEF & Kadonaga JT (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* **16:** 2583–92

Cai L, Dalal CK & Elowitz MB (2008) Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455:** 485–90

Cai L, Friedman N & Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* **440:** 358–62

Cang Y, Auble DT & Prelich G (1999) A new regulatory domain on the TATA-binding protein. *EMBO J.* **18:** 6662–71

Capp J-P (2010) Noise-driven heterogeneity in the rate of genetic-variant generation as a basis for evolvability. *Genetics* **185:** 395–404

Carey LB, van Dijk D, Sloot PMA, Kaandorp JA & Segal E (2013) Promoter sequence determines the relationship between expression level and noise. *PLoS Biol.* **11:** e1001528

Carreto L, Eiriz MF, Domingues I, Schuller D, Moura GR & Santos MAS (2011) Expression variability of co-regulated genes differentiates Saccharomyces cerevisiae strains. *BMC Genomics* **12:** 201

Carreto L, Eiriz MF, Gomes AC, Pereira PM, Schuller D & Santos MAS (2008) Comparative genomics of wild type yeast strains unveils important genome diversity. *BMC Genomics* **9:** 524

Carro D, Bartra E & Pin B (2003) Karyotype Rearrangements in a Wine Yeast Strain by rad52-Dependent and rad52-Independent Mechanisms. *Appl. Environ. Microbiol.* **69:** 2161–2165

Castelnuovo M, Rahman S, Guffanti E, Infantino V, Stutz F & Zenklusen D (2013) Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat. Struct. Mol. Biol.*

Chalancon G, Ravarani CNJ, Balaji S, Martinez-Arias A, Aravind L, Jothi R & Babu MM (2012) Interplay between gene expression noise and regulatory network architecture. *Trends Genet.* **28:** 221–32

Chang HH, Hemberg M, Barahona M, Ingber DE & Huang S (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453:** 544–7

Chang S-L, Lai H-Y, Tung S-Y & Leu J-Y (2013) Dynamic large-scale chromosomal rearrangements fuel rapid adaptation in yeast populations. *PLoS Genet.* **9:** e1003232

Chen C-YA & Shyu A-B (2010) Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip. Rev. RNA* **2:** 167–83

Chen G, Bradford WD, Seidel CW & Li R (2012) Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy. *Nature* **482:** 246–50

Choi JK & Kim Y-J (2009) Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nat. Genet.* **41:** 498–503

Chong S, Chen C, Ge H & Xie XS (2014) Mechanism of Transcriptional Bursting in Bacteria. *Cell* **158:** 314–326

Chou HH & Holmes MH (2001) DNA sequence quality trimming and vector removal. *Bioinformatics* **17:** 1093–104

Cohen AA, Geva-Zatorsky N, Eden E, Frenkel-Morgenstern M, Issaeva I, Sigal A, Milo R, Cohen-Saidon C, Liron Y, Kam Z, Cohen L, Danon T, Perzov N & Alon U (2008) Dynamic proteomics of individual cancer cells in response to a drug. *Science* **322:** 1511–6

Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Pesce CG & Brent R (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437:** 699–706

Comai L (2005) The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6:** 836–46

Coppola JA, Field AS & Luse DS (1983) Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped. *Proc. Natl. Acad. Sci.* **80:** 1251–1255

Corre G, Stockholm D, Arnaud O, Kaneko G, Viñuelas J, Yamagata Y, Neildez-Nguyen TMA, Kupiec J-J, Beslon G, Gandrillon O & Paldi A (2014) Stochastic fluctuations and distributed control of gene expression impact cellular memory. *PLoS One* **9:** e115574

Cowing V (2010) Regulation of mRNA cap methylation. *Biochem. J.* **425:** 295–302

Dadiani M, van Dijk D, Segal B, Field Y, Ben-Artzi G, Raveh-Sadka T, Levo M, Kaplow I, Weinberger A & Segal E (2013) Two DNA-encoded strategies for increasing expression with opposing effects on promoter dynamics and transcriptional noise. *Genome Res.*

Damon C, Vallon L, Zimmermann S, Haider MZ, Galeote V, Dequin S, Luis P, Fraissinet-Tachet L & Marmeisse R (2011) A novel fungal family of oligopeptide transporters identified by functional metatranscriptomics of soil eukaryotes. *ISME J.* **5:** 1871–80

Dar RD, Hosmane NN, Arkin MR, Siliciano RF & Weinberger LS (2014) Screening for noise in gene expression identifies drug synergies. *Science (80-.).* **344:** 1392–1396

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW & Steinmetz LM (2006) A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. U. S. A.* **103:** 5320–5

Davis I (2009) The 'super-resolution' revolution. *Biochem. Soc. Trans.* **37:** 1042–4

Debrauwère H, Loeillet S, Lin W, Lopes J & Nicolas a (2001) Links between replication and recombination in Saccharomyces cerevisiae: a hypersensitive requirement for homologous recombination in the absence of Rad27 activity. *Proc. Natl. Acad. Sci. U. S. A.* **98:** 8263–9

Deng B, Melnik S & Cook PR (2013) Transcription factories, chromatin loops, and the dysregulation of gene expression in malignancy. *Semin. Cancer Biol.* **23:** 65–71

Deng W & Roberts SGE (2007) TFIIB and the regulation of transcription by RNA polymerase II. *Chromosoma* **116:** 417–29

Denisov G, Arehart A & Curtin M (2004) System and method for improving the accuracy of DNA sequencing and error probability estimation through application of a mathematical model to the analysis of electropherograms.

Dequin S & Casaregola S (2011) The genomes of fermentative Saccharomyces. *C. R. Biol.* **334:** 687–93

Dever TE & Green R (2012) The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **4:** a013706

Doma MK & Parker R (2007) RNA Quality Control in Eukaryotes. *Cell* **131:** 660–668

Dornfeld KJ & Livingston DM (1991) Effects of controlled RAD52 expression on repair and recombination in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **11:** 2013–7

Dunckley T & Parker R (1999) The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif. *EMBO J.* **18:** 5411–22

Dunn B, Levine RP & Sherlock G (2005) Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures. *BMC Genomics* **6:** 53

Dunn B, Paulish T, Stanbery A, Piotrowski J, Koniges G, Kroll E, Louis EJ, Liti G, Sherlock G & Rosenzweig F (2013) Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression. *PLoS Genet.* **9:** e1003366

Dunn B, Richter C, Kvitek DJ, Pugh T & Sherlock G (2012) Analysis of the Saccharomyces cerevisiae pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res.* **22:** 908–24

Ehrlich DJ, McKenna BK, Evans JG, Belkina AC, Denis G V, Sherr DH & Cheung MC (2011) Parallel imaging microfluidic cytometer. *Methods Cell Biol.* **102:** 49–75

Eldar A & Elowitz MB (2010) Functional roles for noise in genetic circuits. *Nature* **467:** 167–73

Elkon R, Ugalde AP & Agami R (2013) Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.* **14:** 496–506

Elowitz MB, Levine AJ, Siggia ED & Swain PS (2002) Stochastic gene expression in a single cell. *Science* **297:** 1183–6

Fehrmann S, Bottin-Duplus H, Leonidou A, Mollereau E, Barthelaix A, Wei W, Steinmetz LM & Yvert G (2013) Natural sequence variants of yeast environmental sensors confer cell-to-cell expression variability. *Mol. Syst. Biol.* **9:** 695

Fidalgo M, Barrales RR, Ibeas JI & Jimenez J (2006) Adaptive evolution by mutations in the FLO11 gene. *Proc. Natl. Acad. Sci. U. S. A.* **103:** 11228–33

Fraser HB, Hirsh AE, Giaever G, Kumm J & Eisen MB (2004) Noise minimization in eukaryotic gene expression. *PLoS Biol.* **2:** e137

Freed NE, Silander OK, Stecher B, Böhm A, Hardt W-D & Ackermann M (2008) A simple screen to identify promoters conferring high levels of phenotypic noise. *PLoS Genet.* **4:** e1000307

Fritsch ES, Schacherer J, Bleykasten-Grosshans C, Souciet J-L, Potier S & de Montigny J (2009) Influence of genetic background on the occurrence of chromosomal rearrangements in Saccharomyces cerevisiae. *BMC Genomics* **10:** 99

Fuda NJ, Ardehali MB & Lis JT (2009) Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461:** 186–92

Galeote V, Novo M, Salema-Oom M, Brion C, Valério E, Gonçalves P & Dequin S (2010) FSY1, a horizontally transferred gene in the Saccharomyces cerevisiae EC1118 wine yeast strain, encodes a high-affinity fructose/H+ symporter. *Microbiology* **156:** 3754–61

Gaszner M & Felsenfeld G (2006) Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* **7:** 703–13

George TC, Basiji DA, Hall BE, Lynch DH, Ortyn WE, Perry DJ, Seo MJ, Zimmerman CA & Morrissey PJ (2004) Distinguishing modes of cell death using the ImageStream multispectral imaging flow cytometer. *Cytometry. A* **59:** 237–45

Golding I, Paulsson J, Zawilski SM & Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. *Cell* **123:** 1025–36

Grünberg S & Hahn S (2013) Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem. Sci.* **38:** 603–11

Guimaraes JC, Rocha M & Arkin AP (2014) Transcript level and sequence determinants of protein abundance and noise in Escherichia coli. *Nucleic Acids Res.* **42:** 4791–9

Gupta I, Clauder-Münster S, Klaus B, Järvelin AI, Aiyar RS, Benes V, Wilkening S, Huber W, Pelechano V & Steinmetz LM (2014) Alternative polyadenylation diversifies post-transcriptional regulation by selective RNA-protein interactions. *Mol. Syst. Biol.* **10:** 719

Hansen AS & O'Shea EK (2013) Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. *Mol. Syst. Biol.* **9:** 704

Hastings PJ, Lupski JR, Rosenberg SM & Ira G (2009) Mechanisms of change in gene copy number. *Nat. Rev. Genet.* **10:** 551–64

Hebenstreit D (2013) Are gene loops the cause of transcriptional noise? *Trends Genet.* **null:**

Hirose Y & Manley JL (2000) RNA polymerase II and the integration of nuclear events. *Genes & Dev.* **14:** 1415–1429

Hirose Y, Tacke R & Manley JL (1999) Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev.* **13:** 1234–1239

Holland SL, Reader T, Dyer PS & Avery S V (2014) Phenotypic heterogeneity is a selected trait in natural yeast populations subject to environmental stress. *Environ. Microbiol.* **16:** 1729–40

Holthausen JT, Wyman C & Kanaar R (2010) Regulation of DNA strand exchange in homologous recombination. *DNA Repair (Amst).* **9:** 1264–72

Hong J & Gresham D (2014) Molecular specificity, convergence and constraint shape adaptive evolution in nutrient-poor environments. *PLoS Genet.* **10:** e1004041

Hooshangi S, Thiberge S & Weiss R (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl. Acad. Sci. U. S. A.* **102:** 3581–6

Hope IA & Struhl K (1987) GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *EMBO J.* **6:** 2781–4

Hornung G, Bar-ziv R, Rosin D, Tokuriki N, Tawfik DS, Oren M & Barkai N (2012) Noise – mean relationship in mutated promoters. *Genome Res.* **22:** 2409–2417

Huang B, Wu H, Bhaya D, Grossman A, Granier S, Kobilka BK & Zare RN (2007) Counting low-copy number proteins in a single cell. *Science* **315:** 81–4

Huh D & Paulsson J (2011) Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.* **43:** 95–100

Hurwitz J (2005) The discovery of RNA polymerase. *J. Biol. Chem.* **280:** 42477–85

Infante JJ, Dombek KM, Rebordinos L, Cantoral JM & Young ET (2003) Genome-wide amplifications caused by chromosomal rearrangements play a major role in the adaptive evolution of natural yeast. *Genetics* **165:** 1745–59

Ito Y, Toyota H, Kaneko K & Yomo T (2009) How selection affects phenotypic fluctuation. *Mol. Syst. Biol.* **5:** 264

Jackson RJ, Hellen CUT & Pestova T V (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11:** 113–27

James TC, Usher J, Campbell S & Bond U (2008) Lager yeasts possess dynamic genomes that undergo rearrangements and gene amplification in response to stress. *Curr. Genet.* **53:** 139–52

Johnston IG, Gaal B, Neves RP das, Enver T, Iborra FJ & Jones NS (2012) Mitochondrial variability as a source of extrinsic cellular noise. *PLoS Comput. Biol.* **8:** e1002416

Jones DL, Brewster RC & Phillips R (2014) Promoter architecture dictates cell-to-cell variability in gene expression. *Science (80-.).* **346:** 1533–1536

Juneau K, Nislow C & Davis RW (2009) Alternative splicing of PTC7 in Saccharomyces cerevisiae determines protein localization. *Genetics* **183:** 185–94

Junker JP & van Oudenaarden A (2012) When noisy neighbors are a blessing: analysis of gene expression noise identifies coregulated genes. *Mol. Cell* **45:** 437–8

Juven-Gershon T, Hsu J-Y, Theisen JW & Kadonaga JT (2008) The RNA polymerase II core promoter - the gateway to transcription. *Curr. Opin. Cell Biol.* **20:** 253–9

Kaern M, Elston TC, Blake WJ & Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6:** 451–64

Kalmar T, Lim C, Hayward P, Muñoz-Descalzo S, Nichols J, Garcia-Ojalvo J & Martinez Arias A (2009) Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* **7:** e1000149

Kalsotra A & Cooper TA (2011) Functional consequences of developmentally regulated alternative splicing. *Nat. Rev. Genet.* **12:** 715–29

Kapp LD & Lorsch JR (2004) The molecular mechanics of eukaryotic translation. *Annu. Rev. Biochem.* **73:** 657–704

Kar RK, Qureshi MT, DasAdhikari AK, Zahir T, Venkatesh K V & Bhat PJ (2014) Stochastic galactokinase expression underlies GAL gene induction in a GAL3 mutant of Saccharomyces cerevisiae. *FEBS J.* **281:** 1798–817

Karig DK, Jung S-Y, Srijanto B, Collier CP & Simpson ML (2013) Probing cell-free gene expression noise in femtoliter volumes. *ACS Synth. Biol.*

Karpenshif Y & Bernstein KA (2012) From yeast to mammals: recent advances in genetic control of homologous recombination. *DNA Repair (Amst).* **11:** 781–8

Karpova TS, Kim MJ, Spriet C, Nalley K, Stasevich TJ, Kherrouche Z, Heliot L & McNally JG (2008) Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. *Science* **319:** 466–9

Kaufmann BB, Yang Q, Mettetal JT & van Oudenaarden A (2007) Heritable stochastic switching revealed by single-cell genealogy. *PLoS Biol.* **5:** e239

Kempe H, Schwabe A, Crémazy F, Verschure PJ & Bruggeman FJ (2014) The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. *Mol. Biol. Cell*

Kervestin S & Jacobson A (2012) NMD: a multifaceted response to premature translational termination. *Nat. Rev. Mol. Cell Biol.* **13:** 700–12

Kim YE, Hipp MS, Bracher A, Hayer-Hartl M & Hartl FU (2013) Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82:** 323–55

Kiviet DJ, Nghe P, Walker N, Boulineau S, Sunderlikova V & Tans SJ (2014) Stochasticity of metabolism and growth at the single-cell level. *Nature*

Kravtsova-Ivantsiv Y & Ciechanover A (2012) Non-canonical ubiquitin-based signals for proteasomal degradation. *J. Cell Sci.* **125:** 539–48

Kuehner JN, Pearson EL & Moore C (2011) Unravelling the means to an end: RNA polymerase II transcription termination. *Nat. Rev. Mol. Cell Biol.* **12:** 283–94

Kuras L & Struhl K (1999) Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. *Nature* **399:** 609–13

Lainé J-P, Singh BN, Krishnamurthy S & Hampsey M (2009) A physiological role for gene loops in yeast. *Genes Dev.* **23:** 2604–9

Lang GI & Murray AW (2008) Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae. *Genetics* **178:** 67–82

Legras J-L, Merdinoglu D, Cornuet J-M & Karst F (2007) Bread, beer and wine: Saccharomyces cerevisiae diversity reflects human history. *Mol. Ecol.* **16:** 2091–102

Lehner B (2008) Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol. Syst. Biol.* **4:** 170

Lehner B (2010) Conflict between noise and plasticity in yeast. *PLoS Genet.* **6:** e1001185

Levin D, Harari D & Schreiber G (2011) Stochastic receptor expression determines cell fate upon interferon treatment. *Mol. Cell. Biol.* **31:** 3252–66

Levy SF, Ziv N & Siegal ML (2012) Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLoS Biol.* **10:** e1001325

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G & Durbin R (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–9

Li J, Wang L, Wu X, Fang O, Wang L, Lu C, Yang S, Hu X & Luo Z (2013) Polygenic molecular architecture underlying non-sexual cell aggregation in budding yeast. *DNA Res.* **20:** 55–66

Lidstrom ME & Konopka MC (2010) The role of physiological heterogeneity in microbial population behavior. *Nat. Chem. Biol.* **6:** 705–12

Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, Tsai IJ, Bergman CM, Bensasson D, O'Kelly MJT, van Oudenaarden A, Barton DBH, Bailes E, Nguyen AN, Jones M, Quail MA, et al (2009) Population genomics of domestic and wild yeasts. *Nature* **458:** 337–41

Little SC, Tikhonov M & Gregor T (2013) Precise Developmental Gene Expression Arises from Globally Stochastic Transcriptional Activity. *Cell* **154:** 789–800

Lubliner S, Keren L & Segal E (2013) Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. *Nucleic Acids Res.* **41:** 5569–81

Luger K, Dechassa ML & Tremethick DJ (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* **13:** 436–47

Maamar H, Raj A & Dubnau D (2007) Noise in gene expression determines cell fate in Bacillus subtilis. *Science* **317:** 526–9

Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant PVK, Frazer KA, Cox DR & Ballinger DG (2005) High-resolution whole-genome association study of Parkinson disease. *Am. J. Hum. Genet.* **77:** 685–93

Marczynski GT & Jaehning JA (1985) A transcription map of a yeast centromere plasmid: unexpected transcripts and altered gene expression. *Nucleic Acids Res.* **13:** 8487–506

Melnikov S, Ben-Shem A, Garreau de Loubresse N, Jenner L, Yusupova G & Yusupov M (2012) One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.* **19:** 560–7

Miller-Jensen K, Dey SS, Schaffer D V & Arkin AP (2011) Varying virulence: epigenetic control of expression noise and disease processes. *Trends Biotechnol.* **29:** 517–25

Miura F, Kawaguchi N, Yoshida M, Uematsu C, Kito K, Sakaki Y & Ito T (2008) Absolute quantification of the budding yeast transcriptome by means of competitive PCR between genomic and complementary DNAs. *BMC Genomics* **9:** 574

Monteoliva D, McCarthy CB & Diambra L (2013) Noise Minimisation in Gene Expression Switches. *PLoS One* **8:** e84020

Morales L & Dujon B (2012) Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. *Microbiol. Mol. Biol. Rev.* **76:** 721–39

Muller LAH & McCusker JH (2009) A multispecies-based taxonomic microarray reveals interspecies hybridization and introgression in Saccharomyces cerevisiae. *FEMS Yeast Res.* **9:** 143–52

Murphy KF, Balázsi G & Collins JJ (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **104:** 12726–31

Nachman I, Regev A & Ramanathan S (2007) Dissecting timing variability in yeast meiosis. *Cell* **131:** 544–56

Nakao J, Miyanohara A, Toh-e A & Matsubara K (1986) Saccharomyces cerevisiae PHO5 promoter region: location and function of the upstream activation site. *Mol. Cell. Biol.* **6:** 2613–23

Nash R, Weng S, Hitz B, Balakrishnan R, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hong EL, Livstone MS, Oughtred R, Park J, Skrzypek M, Theesfeld CL, Binkley G, Dong Q, Lane C, Miyasato S, et al (2007) Expanded protein information at SGD: new pages and proteome browser. *Nucleic Acids Res.* **35:** D468–71

New AM, Cerulus B, Govers SK, Perez-Samper G, Zhu B, Boogmans S, Xavier JB & Verstrepen KJ (2014) Different Levels of Catabolite Repression Optimize Growth in Stable and Variable Environments. *PLoS Biol.* **12:** e1001764

Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL & Weissman JS (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441:** 840–6

Nishant KT, Wei W, Mancera E, Argueso JL, Schlattl A, Delhomme N, Ma X, Bustamante CD, Korbel JO, Gu Z, Steinmetz LM & Alani E (2010) The baker's yeast diploid genome is remarkably stable in vegetative growth and meiosis. *PLoS Genet.* **6:** e1001109

Nishimura K, Tsuru S, Suzuki H & Yomo T (2014) Stochasticity in gene expression in a cell-sized compartment. *ACS Synth. Biol.*

Novo M, Bigey F, Beyne E, Galeote V, Gavory F, Mallet S, Cambon B, Legras J-L, Wincker P, Casaregola S & Dequin S (2009) Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. *Proc. Natl. Acad. Sci. U. S. A.* **106:** 16333–8

Ochiai H, Sugawara T, Sakuma T & Yamamoto T (2014) Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. *Sci. Rep.* **4:** 7125

Octavio LM, Gedeon K & Maheshri N (2009) Epigenetic and conventional regulation is distributed among activators of FLO11 allowing tuning of population-level heterogeneity in its expression. *PLoS Genet.* **5:** e1000673

Ohno M, Karagiannis P & Taniguchi Y (2014) Protein Expression Analyses at the Single Cell Level. *Molecules* **19:** 13932–13947

Ozbudak EM, Thattai M, Kurtser I, Grossman AD & van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. *Nat. Genet.* **31:** 69–73

Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B & Milos PM (2010) Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* **143:** 1018–29

Paffett KS, Clikeman JA, Palmer S & Nickoloff JA (2005) Overexpression of Rad51 inhibits double-strand break-induced homologous recombination but does not affect gene conversion tract lengths. *DNA Repair (Amst).* **4:** 687–698

Paranjape SM, Kamakaka RT & Kadonaga JT (1994) Role of chromatin structure in the regulation of transcription by RNA polymerase II. *Annu. Rev. Biochem.* **63:** 265–97

Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, Sanderson BW, Hattem GL & Li R (2010) Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* **468:** 321–5

Pelechano V, Wei W & Steinmetz LM (2013) Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497:** 127–31

Pelet S, Rudolf F, Nadal-Ribelles M, de Nadal E, Posas F & Peter M (2011) Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science* **332:** 732–5

Pérez-Ortín JE, Querol A, Puig S & Barrio E (2002) Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. *Genome Res.* **12:** 1533–9

Picot J, Guerin CL, Le Van Kim C & Boulanger CM (2012) Flow cytometry: retrospective, fundamentals and recent instrumentation. *Cytotechnology* **64:** 109–30

Raj A & van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135:** 216–26

Raj A & van Oudenaarden A (2009) Single-molecule approaches to stochastic gene expression. *Annu. Rev. Biophys.* **38:** 255–70

Raj A, Rifkin SA, Andersen E & van Oudenaarden A (2010) Variability in gene expression underlies incomplete penetrance. *Nature* **463:** 913–8

Ramakrishnan V (2002) Ribosome Structure and the Mechanism of Translation. *Cell* **108:** 557–572

Raser JM & O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. *Science* **304:** 1811–4

Raser JM & O'Shea EK (2005) Noise in gene expression: origins, consequences, and control. *Science* **309:** 2010–3

Rhee A, Cheong R & Levchenko A (2014) Noise decomposition of intracellular biochemical signaling networks using nonequivalent reporters. *Proc. Natl. Acad. Sci. U. S. A.***:** 1411932111–

Rhee HS & Pugh BF (2012) Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483:** 295–301

Richard M & Yvert G (2014) How does evolution tune biological noise? *Front. Genet.* **5:** 374

Roeder RG & Rutter WJ (1969) Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms. *Nature* **224:** 234–237

Roesch A, Fukunaga-Kalabis M, Schmidt EC, Zabierowski SE, Brafford PA, Vultur A, Basu D, Gimotty P, Vogt T & Herlyn M (2010) A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. *Cell* **141:** 583–94

Rosenfeld N, Young JW, Alon U, Swain PS & Elowitz MB (2005) Gene regulation at the single-cell level. *Science* **307:** 1962–5

Salari R, Wojtowicz D, Zheng J, Levens D, Pilpel Y & Przytycka TM (2012) Teasing apart translational and transcriptional components of stochastic variations in eukaryotic gene expression. *PLoS Comput. Biol.* **8:** e1002644

Salinas F, Cubillos FA, Soto D, Garcia V, Bergström A, Warringer J, Ganga MA, Louis EJ, Liti G & Martinez C (2012) The genetic basis of natural variation in oenological traits in Saccharomyces cerevisiae. *PLoS One* **7:** e49640

Sanchez A, Choubey S & Kondev J (2013) Regulation of Noise in Gene Expression. *Annu. Rev. Biophys.*

Sanchez A & Golding I (2013) Genetic Determinants and Cellular Constraints in Noisy Gene Expression. *Science (80-.).* **342:** 1188–1193

Saunders A, Core LJ & Lis JT (2006) Breaking barriers to transcription elongation. *Nat. Rev. Mol. Cell Biol.* **7:** 557–67

Schacherer J, Shapiro JA, Ruderfer DM & Kruglyak L (2009) Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae. *Nature* **458:** 342–5

Schneider-Poetsch T, Usui T, Kaida D & Yoshida M (2010) Garbled messages and corrupted translations. *Nat. Chem. Biol.* **6:** 189–198

Schreve JL & Garrett JM (2004) Yeast Agp2p and Agp3p function as amino acid permeases in poor nutrient conditions. *Biochem. Biophys. Res. Commun.* **313:** 745–51

Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK, Wang J-PZ & Widom J (2006) A genomic code for nucleosome positioning. *Nature* **442:** 772–8

Selmecki AM, Maruvka YE, Richmond PA, Guillet M, Shoresh N, Sorenson AL, De S, Kishony R, Michor F, Dowell R & Pellman D (2015) Polyploidy can drive rapid adaptation in yeast. *Nature* **advance on:**

Selva EM, New L, Crouse GF & Lahue RS (1995) Mismatch Correction Acts as a Barrier to Homeologous Recombination in Saccharomyces cerevisiae. *Genetics* **139:** 1175–1188

Shah K & Tyagi S (2013) Barriers to transmission of transcriptional noise in a c-fos c-jun pathway. *Mol. Syst. Biol.* **9:** 687

Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D, Trombetta JT, Gennert D, Gnirke A, Goren A, Hacohen N, Levin JZ, Park H & Regev A (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*

Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, Chen P, Gertner RS, Gaublomme JT, Yosef N, Schwartz S, Fowler B, Weaver S, Wang J, Wang X, Ding R, Raychowdhury R, Friedman N, Hacohen N, Park H, et al (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **advance on:**

Shandilya J & Roberts SGE (2012) The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim. Biophys. Acta* **1819:** 391–400

Shapiro HM (2003) Practical Flow Cytometry 4th ed. Hoboken: John Wiley & Sons, Inc.

Sharma S V, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, McDermott U, Azizian N, Zou L, Fischbach MA, Wong K-K, Brandstetter K, Wittner B, Ramaswamy S, Classon M & Settleman J (2010) A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141:** 69–80

Sharon E, van Dijk D, Kalma Y, Keren L, Manor O, Yakhini Z & Segal E (2014) Probing the effect of promoters on noise in gene expression using thousands of designed sequences. *Genome Res.***:** gr.168773.113–

Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, Zeevi D, Keren L, Yakhini Z, Weinberger A & Segal E (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **30:** 521–30

Shatkin A (1976) Capping of eucaryotic mRNAs. *Cell* **9:** 645–653

Shatkin AJ & Manley JL (2000) The ends of the affair: capping and polyadenylation. *Nat. Struct. Biol.* **7:** 838–42

Shen CH, Leblanc BP, Alfieri JA & Clark DJ (2001) Remodeling of yeast CUP1 chromatin involves activator-dependent repositioning of nucleosomes over the entire gene and flanking sequences. *Mol. Cell. Biol.* **21:** 534–47

Shi Q, Qin L, Wei W, Geng F, Fan R, Shin YS, Guo D, Hood L, Mischel PS & Heath JR (2012) Single-cell proteomic chip for profiling intracellular signaling pathways in single tumor cells. *Proc. Natl. Acad. Sci. U. S. A.* **109:** 419–24

Silva-Rocha R & de Lorenzo V (2012) Stochasticity of TOL plasmid catabolic promoters sets a bimodal expression regime in Pseudomonas putida mt-2 exposed to m-xylene. *Mol. Microbiol.* **86:** 199–211

Singh BN & Hampsey M (2007) A transcription-independent role for TFIIB in gene looping. *Mol. Cell* **27:** 806–16

Singh DK, Ku C-J, Wichaidit C, Steininger RJ, Wu LF & Altschuler SJ (2010) Patterns of basal signaling heterogeneity can distinguish cellular populations with different drug sensitivities. *Mol. Syst. Biol.* **6:** 369

Singh GP (2013) Coupling Between Noise and Plasticity in Escherichia coli. *G3 (Bethesda).***:** g3.113.008540–

Smale ST & Baltimore D (1989) The 'initiator' as a transcription control element. *Cell* **57:** 103–113

Small EC, Xi L, Wang J-P, Widom J & Licht JD (2014) Single-cell nucleosome mapping reveals the molecular basis of gene expression heterogeneity. *Proc. Natl. Acad. Sci. U. S. A.* **111:** E2462–71

Smith MCA, Sumner ER & Avery S V (2007) Glutathione and Gts1p drive beneficial variability in the cadmium resistances of individual yeast cells. *Mol. Microbiol.* **66:** 699–712

Spencer SL, Gaudet S, Albeck JG, Burke JM & Sorger PK (2009) Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* **459:** 428–32

Spiller DG, Wood CD, Rand D a & White MRH (2010) Measurement of single-cell dynamics. *Nature* **465:** 736–45

Steensels J, Snoek T, Meersman E, Nicolino MP, Voordeckers K & Verstrepen KJ (2014) Improving industrial yeast strains: exploiting natural and artificial diversity. *FEMS Microbiol. Rev.* **38:** 947–95

Stewart-Ornstein J, Weissman JS & El-Samad H (2012) Cellular noise regulons underlie fluctuations in Saccharomyces cerevisiae. *Mol. Cell* **45:** 483–93

Storchova Z (2014) Ploidy changes and genome stability in yeast. *Yeast* **31:** 421–30

Sun X, Thrower D, Qiu J, Wu P, Zheng L, Zhou M, Bachant J, Wilson DM & Shen B (2003) Complementary functions of the Saccharomyces cerevisiae Rad2 family nucleases in Okazaki

fragment maturation, mutation avoidance, and chromosome stability. *DNA Repair (Amst).* **2:** 925–940

Sung P (1997) Function of yeast Rad52 protein as a mediator between replication protein A and the Rad51 recombinase. *J. Biol. Chem.* **272:** 28194–7

Swain PS, Elowitz MB & Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **99:** 12795–800

Tanaka K, Tatebayashi K, Nishimura A, Yamamoto K, Yang H-Y & Saito H (2014) Yeast osmosensors Hkr1 and Msb2 activate the Hog1 MAPK cascade by different mechanisms. *Sci. Signal.* **7:** ra21

Tan-Wong SM, Zaugg JB, Camblong J, Xu Z, Zhang DW, Mischo HE, Ansari AZ, Luscombe NM, Steinmetz LM & Proudfoot NJ (2012) Gene loops enhance transcriptional directionality. *Science* **338:** 671–5

Thattai M & van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U. S. A.* **98:** 8614–9

Thomas MC & Chiang C-M (2008) The General Transcription Machinery and General Cofactors. *Crit. Rev. Biochem. Mol. Biol.*

Tirosh I, Barkai N & Verstrepen KJ (2009) Promoter architecture and the evolvability of gene expression. *J. Biol.* **8:** 95

To T-L & Maheshri N (2010) Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science* **327:** 1142–5

Tourrière H, Chebli K & Tazi J (2002) mRNA degradation machines in eukaryotic cells. *Biochimie* **84:** 821–837

Treu L, Campanaro S, Nadai C, Toniolo C, Nardi T, Giacomini A, Valle G, Blondin B & Corich V (2014) Oxidative stress response and nitrogen utilization are strongly variable in Saccharomyces cerevisiae wine strains with different fermentation performances. *Appl. Microbiol. Biotechnol.* **98:** 4119–35

De Vargas Roditi L & Claassen M (2015) Computational and experimental single cell biology techniques for the definition of cell type heterogeneity, interplay and intracellular dynamics. *Curr. Opin. Biotechnol.* **34:** 9–15

Vasiljeva L & Buratowski S (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol. Cell* **21:** 239–48

Veening J-W, Stewart EJ, Berngruber TW, Taddei F, Kuipers OP & Hamoen LW (2008) Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc. Natl. Acad. Sci. U. S. A.* **105:** 4393–8

Venancio TM, Balaji S, Geetha S & Aravind L (2010) Robustness and evolvability in natural chemical resistance: identification of novel systems properties, biochemical mechanisms and regulatory interactions. *Mol. Biosyst.* **6:** 1475–91

Venter U, Svaren J, Schmitz J, Schmid A & Hörz W (1994) A nucleosome precludes binding of the transcription factor Pho4 in vivo to a critical target site in the PHO5 promoter. *EMBO J.* **13:** 4848–55

Venters BJ & Pugh BF (2009) A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome Res.* **19:** 360–71

Venters BJ, Wachi S, Mavrich TN, Andersen BE, Jena P, Sinnamon AJ, Jain P, Rolleri NS, Jiang C, Hemeryck-Walsh C & Pugh BF (2011) A comprehensive genomic binding map of gene and chromatin regulatory proteins in Saccharomyces. *Mol. Cell* **41:** 480–92

Verstrepen KJ, Jansen A, Lewitter F & Fink GR (2005) Intragenic tandem repeats generate functional variability. *Nat. Genet.* **37:** 986–90

Vinces MD, Legendre M, Caldara M, Hagihara M & Verstrepen KJ (2009) Unstable tandem repeats in promoters confer transcriptional evolvability. *Science* **324:** 1213–6

Viney M & Reece SE (2013) Adaptive noise. *Proc. Biol. Sci.* **280:** 20131104

Vinuelas J, Kaneko G, Coulon A, Vallin E, Morin V, Mejia-Pous C, Kupiec J-J, Beslon G & Gandrillon O (2013) Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. *BMC Biol.* **11:** 15

Warren L, Bryder D, Weissman IL & Quake SR (2006) Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proc. Natl. Acad. Sci. U. S. A.* **103:** 17807–12

Watson JD & Crick FHC (1953) Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171:** 737–738

Wei W, McCusker JH, Hyman RW, Jones T, Ning Y, Cao Z, Gu Z, Bruno D, Miranda M, Nguyen M, Wilhelmy J, Komp C, Tamse R, Wang X, Jia P, Luedi P, Oefner PJ, David L, Dietrich FS, Li Y, et al (2007) Genome sequencing and comparative analysis of Saccharomyces cerevisiae strain YJM789. *Proc. Natl. Acad. Sci. U. S. A.* **104:** 12825–30

Weinberger L, Voichek Y, Tirosh I, Hornung G, Amit I & Barkai N (2012) Expression noise and acetylation profiles distinguish HDAC functions. *Mol. Cell* **47:** 193–202

Weinberger LS, Burnett JC, Toettcher JE, Arkin AP & Schaffer D V (2005) Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* **122:** 169–82

Weinberger LS, Dar RD & Simpson ML (2008) Transient-mediated fate determination in a transcriptional circuit of HIV. *Nat. Genet.* **40:** 466–70

Wickner S (1999) Posttranslational Quality Control: Folding, Refolding, and Degrading Proteins. *Science (80-.).* **286:** 1888–1893

Wildenberg GA & Murray AW (2014) Evolving a 24-hr oscillator in budding yeast. *Elife* **3:** e04875

Wind M & Reines D (2000) Transcription elongation factor SII. *Bioessays* **22:** 327–36

Wu M, Su R-Q, Li X, Ellis T, Lai Y-C & Wang X (2013) Engineering of regulated stochastic cell fate determination. *Proc. Natl. Acad. Sci.***:** 1305423110–

Xie Y, Liu Y, Argueso JL, Henricksen LA, Kao H-I, Bambara RA & Alani E (2001) Identification of rad27 Mutations That Confer Differential Defects in Mutation Avoidance, Repeat Tract Instability, and Flap Cleavage. *Mol. Cell. Biol.* **21:** 4889–4899

Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RKC, Hua Y, Gueroussov S, Najafabadi HS, Hughes TR, Morris Q, Barash Y, Krainer AR, Jojic N, Scherer SW, Blencowe BJ & Frey BJ (2014) The human splicing code reveals new insights into the genetic determinants of disease. *Science (80-.).* **347:** 1254806–

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W & Steinmetz LM (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457:** 1033–7

Yang C, Bolotin E, Jiang T, Sladek FM & Martinez E (2007) Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **389:** 52–65

Yang S, Kim S, Rim Lim Y, Kim C, An HJ, Kim J-H, Sung J & Lee NK (2014) Contribution of RNA polymerase concentration variation to protein expression noise. *Nat. Commun.* **5:** 4761

Yona AH, Manor YS, Herbst RH, Romano GH, Mitchell A, Kupiec M, Pilpel Y & Dahan O (2012) Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl. Acad. Sci. U. S. A.* **109:** 21010–5

Zanton SJ & Pugh BF (2004) Changes in genomewide occupancy of core transcriptional regulators during heat stress. *Proc. Natl. Acad. Sci. U. S. A.* **101:** 16843–8

Zenklusen D, Larson DR & Singer RH (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* **15:** 1263–71

Zhang H, Skelton A, Gardner RC & Goddard MR (2010) Saccharomyces paradoxus and Saccharomyces cerevisiae reside on oak trees in New Zealand: evidence for migration from Europe and interspecies hybrids. *FEMS Yeast Res.* **10:** 941–7

Zhang H, Zeidler AFB, Song W, Puccia CM, Malc E, Greenwell PW, Mieczkowski PA, Petes TD & Argueso JL (2013a) Gene copy-number variation in haploid and diploid strains of the yeast Saccharomyces cerevisiae. *Genetics* **193:** 785–801

Zhang Q, Yoon Y, Yu Y, Parnell EJ, Garay JAR, Mwangi MM, Cross FR, Stillman DJ & Bai L (2013b) Stochastic expression and epigenetic memory at the yeast HO promoter. *Proc. Natl. Acad. Sci.***:** 1306113110–

Zhang Z, Qian W & Zhang J (2009) Positive selection for elevated gene expression noise in yeast. *Mol. Syst. Biol.* **5:** 299

Zhao Y, Strope PK, Kozmin SG, McCusker JH, Dietrich FS, Kokoska RJ & Petes TD (2014) Structures of Naturally Evolved CUP1 Tandem Arrays in Yeast Indicate That These Arrays Are Generated by Unequal Nonhomologous Recombination. *G3 (Bethesda).* **4:** 2259–69

Zheng D-Q, Chen J, Zhang K, Gao K-H, Li O, Wang P-M, Zhang X-Y, Du F-G, Sun P-Y, Qu A-M, Wu S & Wu X-C (2014) Genomic structural variations contribute to trait improvement during whole-genome shuffling of yeast. *Appl. Microbiol. Biotechnol.* **98:** 3059–70

Zopf CJ, Quinn K, Zeidman J & Maheshri N (2013) Cell-Cycle Dependence of Transcription Dominates Noise in Gene Expression. *PLoS Comput. Biol.* **9:** e1003161

Zuleta IA, Aranda-Díaz A, Li H & El-Samad H (2014) Dynamic characterization of growth and gene expression using high-throughput automated flow cytometry. *Nat. Methods* **11:** 443–8

Lists

List of abbreviations

5'UTR	5' untranslated region
5-FOA	5-Fluoroorotic Acid
AD	Activation domain
BMDC	Bone-marrow-derived dendritic cells
BRE	TFIIB recognition element
CDK	Cyclin-dependent kinase
CFIA	Cleavage factor IA
CFP	Cyan fluorescent protein
CGH	Comparative genome hybridization
CIN	Chromosomal instability
CNV	Copy number variations
CPF	Polyadenylation factor
CPSF	Cleavage and polyadenylation specific factor
CstF	Cleavage stimulatory factor
CTD	C-termninal domain
CUT	Cryptic unstable transcripts
CV	Coefficient of variation
DB	DNA-binding domain
dHJ	Double-Holliday junction
D-loop	Displacement-loop
DNA	Deoxyribonucleic acid
DPE	Downstream promoter element
DRMG	DNA repair and maintenance gene
DSB	Double-strand broken
dsDNA	Double-stranded DNA
ePTL	Expression probabilistic trait loci
ESC	Embryonic stem cells
FACS	The fluorescence activated cell sorting
FACS	Fluorescence activated cell sorting
FISH	Fluorescent in situ hybridization system
FOT	Fungal oligopeptide transporters

FHR	Frequency of homologue recombination
GCR	Gross chromosomal rearrangement
GFP	Green fluorescent protein
GRN	Gene regulation network
GSH	Glutathione
GTF	General transcription factor
HDAC	Histone deacetylation complexes
HGT	Horizontal gene transfer
HIV	Human immunodeficiency virus
HR	Homologous recombination
Indel	Insertion or deletion of bases in the DNA
LOH	Loss of heterozygosity
MMR	Mismatch repair
mRNA	Messager RNA
ncRNA	Non-coding RNA
NHEJ	Non-homologous end joint
ORF	Open reading frame
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
QTL	Quantitative trait locus
RGVG	Rate of genetic-variant generation
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
RPA	Replication protein A
rRNA	Ribosomal RNA
RT-qPCR	Reverse transcription-real time PCR
SAGA	Spt-Ada-Gcn5 acetyltransferase
SDSA	Synthesis dependent strand annealing
SNP	Single nucleotide polymorphism
ssDNA	Single-stranded DNA
TAF	TBP-associated factor
TBP	TATA-binding-protein
tetO	tTA binding site

TIF-seq	Transcript isoform sequencing
TNF	Tumor necrosis factor
TOF	Time-of-flight
TRAIL	TNF-related apoptosis-inducing ligand
tRNA	Transfer RNA
TSS	Transcription starting site
tTA	tet-transcriptional activator
UAS	Upstream activating sequence
UPRE	Unfolded protein response element
YFP	Yellow fluorescent protein

List of figures

List of tables

Appendixes

Appendix I: Isolated clones of the selected clones by screening

Number	Origina clone		1st subclone		2nd subclone	
	Mean	CV	Mean	CV	Mean	CV
1	197,56	131,73	136,77	133,04	157,28	136,89
2	547,87	126,08	477,54	122,03	500,78	128,15
3	220,09	127,08	156,03	125,04	200,24	129,07
4	1076,81	127,59	996,35	125,99	1000,54	128,14
5	163,55	130,91	139,44	127,73	150,25	124,05
6	333,57	125,58	542,11	120,19	500,41	119,41
7	590,38	127,29	541,46	121,41	551,24	122,2
8	776,26	124,24	756	124,21	761,2	121,96
9	1219,14	124,18	1028	124,55	1147,23	129,44
10	737,52	129,9	861,7	128,4	812,47	127,75
11	251,22	124,81	242,59	118,79	224,23	121,79
12	424,14	124,81	404,44	123,59	415,47	120,28
13	238,95	137,21	216,06	140,75	227,13	135,37
14	232,69	130,66	345,12	128,58	300,14	125,05
15	390,62	127,36	254,92	122,69	354,32	127,42
16	230,11	132,05	224,6	131,36	231,21	127,17
17	574,99	128,5	557,57	121,12	521,36	129,83
18	667,87	126,27	606,06	120,9	603,12	122,77
19	855,96	129,06	794,33	126,42	800,47	128,39
20	211,03	127,27	209,08	126,08	221,33	123,75
21	790,47	132,04	1030,76	125,18	853,13	131,19
22	191,43	126,69	197,84	126,1	199,78	121,01
23	947,8	125,76	886,84	123,85	900,85	126,91
24	178,32	127,18	181,52	120,32	178,58	122,77
25	710,52	137,43	705,61	133,33	712,33	135,04
26	584,97	135,66	498,29	135,09	556,32	130,56
27	143,72	129,07	107,27	126,27	133,24	127,36
28	507,99	131,62	603,89	130,74	557,36	137,62
29	978,35	129,81	560,2	124,77	887,12	129,37
30	676,2	130,56	588,01	122,89	624,33	123,18

31	248,13	126,06	243,94	120,08	257,14	123,58
32	34,4	121,22	34,63	123,85	35,47	125,55
33	77,42	117,45	76,48	116,24	75,44	113,46
34	27,88	113,75	27,91	112,51	28,11	110,16
35	30,32	120,76	31,74	128,46	32,21	123,91
36	21,53	122,67	24,4	124,6	23,37	126,55
37	67,01	116,44	65,99	116,94	66,89	114,38
38	28,47	112,87	27,93	114,38	27,97	115,45
39	54,48	118,65	57,42	115,05	58,34	111,73
40	62,3	119,07	67,32	120,76	66,27	125,91
41	37,01	116,95	32,36	119,72	33,24	112,17
42	55,03	122,16	58,31	124,42	53,21	126,31
43	32,42	123,41	28,47	129,5	31,36	120,64
44	47,18	119,94	35,17	120,61	45,24	121,69
45	82,44	114,89	85,63	113,94	84,53	119,52
46	82,25	114,34	84,95	118,8	83,65	114,59
47	46,11	120,2	42,5	115,78	75,37	119,63
48	55,97	114,06	53	104,77	55,38	110,61
49	85,53	122,54	89,8	129,66	88,15	124,9
50	40,36	113,98	38,8	115,57	40,17	117,96
51	32,26	117,49	29,97	112,52	30,25	118,33
52	28,35	115,45	25,68	116,84	27,11	121,48
53	54	118,11	55,88	114,97	53,47	117,2
54	45,84	120,3	38,15	120,08	43,26	127,06
55	67,4	116,82	68,92	115,94	65,17	113,19
56	40,49	121,98	43,84	122,44	44,71	121,81
57	95,31	115,18	90,65	117,53	93,31	113,24
58	79,21	119,4	71,51	116,26	77,84	116,95
59	96,32	120,31	90,21	121,52	93,86	119,44
60	66,28	117,68	66,13	118,93	65,14	113,58
61	28,35	123,4	25,05	122,44	27,03	121,99
62	69,02	114,03	61,55	114,69	68,61	113,5
63	50,01	112,15	52,33	115,68	53,14	115,8
64	66,84	113,08	69,2	115,47	67	115,54
65	55,45	111,27	60,12	118,67	57,15	115,04

66	43,12	123,91	45,28	125,38	47	123,4
67	67,87	112,2	69,64	113,8	68,9	106,1
68	26,95	122,82	23,42	124,38	25,29	122,12
69	44,39	123,09	43,07	115,64	45,99	118,76
70	33,52	124,04	35,37	124,98	36,18	127,17
71	41,58	113,72	44,38	119,81	43,35	115,32
72	41,84	114,55	42,52	112,5	43,17	118,86
73	56,04	119,29	51,22	115,3	55,31	113,98
74	33,31	117,31	31,6	115,46	35,87	116,04
75	63,67	119,6	58,66	118,44	66,47	118,02
76	33,99	116,36	38,07	116,38	35,69	115,98
77	40,36	111,99	42,63	117,35	43,5	110,75
78	32,05	122,05	31,98	123,34	35,36	124
79	30,48	115,3	38,44	117,08	35,22	112,24
80	81,34	118,32	87,92	111,06	85,99	116,27
81	32,42	115,57	35,32	119,16	34,82	113,92
82	81,58	122,46	80,43	128,27	83,26	124,37
83	30,19	119,35	33,39	115,78	31,28	113,32
84	40,55	122,55	35,7	124,29	41,73	121,73
85	43,12	115,73	45,64	119,69	46,17	119,87
86	51,25	119,4	56,65	116,69	55,42	116,09
87	37,58	117,03	30,43	119,23	35,4	118,95
88	41,65	117,89	39,75	117,18	41,26	116,2
89	39,32	114,55	35,37	114,8	37,76	117,81
90	33,52	115,87	36,43	111,56	34,58	114,46
91	50,01	116,71	55,59	113,03	52,47	118,14
92	42,94	120,73	37,04	122,73	41,82	128,47
93	69,02	112,05	61,64	110,6	65,77	115,64
94	55,45	113,89	53,02	115,64	54,78	110,38
95	32,28	121,69	27,62	115,35	30,54	121,26
96	56,73	122,18	51,71	128,45	57,17	124,96
97	33,99	124,14	36,68	121,66	34,32	127,36

Appendix II: Results of mapping of the sequenced fragments to S288c

#Read_name	chrom	begin	end	strand	length	gene	name	unimodal/ bimodal	note
NewNoisyClones_A01-PrimSeqpUG35promless.ab1	chr04	341226	341926	1	700	YDL061C	RPS29B	bi	40 bp before 5'UTR intron
NewNoisyClones_A02-PrimSeqpUG35promless.ab1	chr14	345487	345916	1	429	YNL154C	YCK2	bi	200bp before ATG
NewNoisyClones_A03-PrimSeqpUG35promless.ab1	chr05	140362	140965	1	603	YEL008C-A		bi	ORF in the fragment
NewNoisyClones_A04-PrimSeqpUG35promless.ab1	chr16	121465	121750	-1	285	YPL226W	NEW1	bi	17bp before ATG
NewNoisyClones_A05-PrimSeqpUG35promless.ab1	chr12	197902	198871	-1	969		SNR30	bi	promoter of snRNA
NewNoisyClones_A06-PrimSeqpUG35promless.ab1	chr04	652724	653587	-1	863	YDR099W	BMH2	bi	contains 5'UTR intron
NewNoisyClones_A07-PrimSeqpUG35promless.ab1	chr04	341226	341926	1	700	YDL061C	RPS29B	bi	40 before ATG 5'UTR intron
NewNoisyClones_A08-PrimSeqpUG35promless.ab1	chr16	623162	623714	-1	552	YPR028W	YOP1	bi	contains part of the ORF
NewNoisyClones_A09-PrimSeqpUG35promless.ab1	chr15	258740	259351	-1	611		SNR50	bi	promoter of snRNA
NewNoisyClones_A10-PrimSeqpUG35promless.ab1	chr04	1468472	1469302	1	830	YDR508C	GNP1	bi	30bp before ATG
NewNoisyClones_A11-PrimSeqpUG35promless.ab1	chr04	341226	341926	1	700	YDL061C	RPS29B	bi	40bp before ATG 5'UTR intron
NewNoisyClones_A12-PrimSeqpUG35promless.ab1	chr08	505222	506123	1	901	YHR203C	RPS4B	bi	contains part of the ORF
NewNoisyClones_B01-PrimSeqpUG35promless.ab1	chr05	33421	34369	1	948	YEL063C	CAN1	bi	contains part of the ORF
NewNoisyClones_B02-PrimSeqpUG35promless.ab1	chr12	197950	198871	-1	921		SNR30	bi	promoter of snRNA
NewNoisyClones_B03-PrimSeqpUG35promless.ab1	chr04	556343	556846	1	503	YDR050C	TPI1	bi	contains part of the ORF
NewNoisyClones_B04-PrimSeqpUG35promless.ab1	chr12	198313	198808	-1	495		SNR30	bi	promoter of snRNA
NewNoisyClones_B05-PrimSeqpUG35promless.ab1	chr06	74506	75332	-1	826	YFL031W	HAC1	bi	contains part of the ORF
NewNoisyClones_B06-PrimSeqpUG35promless.ab1	chr08	212749	213681	1	932	YHR055C	CUP1-2	bi	30bp before ATG
NewNoisyClones_B07-PrimSeqpUG35promless.ab1	chr04	1384840	1385156	-1	316	YDR461W	MFA1	bi	20bp before ATG
NewNoisyClones_B08-PrimSeqpUG35promless.ab1	chr04	652605	653587	-1	982	YDR099W	BMH2	bi	contain 5'UTR intron

NewNoisyClones_B09-PrimSeqpUG35promless.ab1	chr07	76192	76880	-1	688	YGL225W	VRG4	bi	14bp before ATG
NewNoisyClones_B10-PrimSeqpUG35promless.ab1	chr10	207378	207845	-1	467	YJL111W	CCT7	bi	32bp before ATG
NewNoisyClones_B11-PrimSeqpUG35promless.ab1	chr02	501552	502527	1	975	YBR132C	AGP2	bi	110bp before ATG
NewNoisyClones_B12-PrimSeqpUG35promless.ab1	chr04	341226	341926	1	700	YDL061C	RPS29B	bi	40bp before ATG 5'UTR intron
NewNoisyClones_C01-PrimSeqpUG35promless.ab1	chr15	258740	259351	-1	611		SNR50	bi	promoter of snRNA
NewNoisyClones_C02-PrimSeqpUG35promless.ab1	chr12	197895	198871	-1	976		SNR30	bi	promoter of snRNA
NewNoisyClones_C03-PrimSeqpUG35promless.ab1	chr05	544698	545638	-1	940	YER177W	BMH1	bi	contains part of the ORF
NewNoisyClones_C04-PrimSeqpUG35promless.ab1	chr07	76192	76880	-1	688	YGL225W	VRG4	bi	14bp before ATG
NewNoisyClones_C05-PrimSeqpUG35promless.ab1	chr05	544727	545638	-1	911	YER177W	BMH1	bi	contains part of the ORF
NewNoisyClones_C06-PrimSeqpUG35promless.ab1	chr04	1468472	1469317	1	845	YDR508C	GNP1	bi	30bp before ATG
NewNoisyClones_C07-PrimSeqpUG35promless.ab1	chr12	241765	242244	-1	479	YLR048W	RPS0B	uni	contains part of the ORF
NewNoisyClones_C08-PrimSeqpUG35promless.ab1	chr10	745187	745396	1	209	YJR162C		uni	contains part of the ORF
NewNoisyClones_C09-PrimSeqpUG35promless.ab1	chr16	794224	795035	-1	811	YPR132W	RPS23B	uni	contains part of the ORF
NewNoisyClones_C10-PrimSeqpUG35promless.ab1	chr16	794314	795035	-1	721	YPR132W	RPS23B	uni	contains part of the ORF
NewNoisyClones_C11-PrimSeqpUG35promless.ab2								uni	not sequenced
NewNoisyClones_C12-PrimSeqpUG35promless.ab1	chr15	101154	101566	1	412			uni	no known promoter found
NewNoisyClones_D01-PrimSeqpUG35promless.ab1	chr04	974175	974596	-1	421			uni	no known promoter found
NewNoisyClones_D02-PrimSeqpUG35promless.ab1	chr13	291059	291720	1	661			uni	no known promoter found
NewNoisyClones_D03-PrimSeqpUG35promless.ab2								uni	not sequenced
NewNoisyClones_D04-PrimSeqpUG35promless.ab1	chr13	558857	559144	-1	287	YMR147W		uni	55bp before ATG
NewNoisyClones_D05-PrimSeqpUG35promless.ab1	chr08	198307	199084	1	777	YHR046C	INM1	uni	22bp before ATG
NewNoisyClones_D06-PrimSeqpUG35promless.ab1	chr08	198307	199084	1	777	YHR046C	INM1	uni	22bp before ATG
NewNoisyClones_D07-PrimSeqpUG35promless.ab1	chr02	429048	429658	-1	700			uni	no known promoter found
NewNoisyClones_D08-	chr08	508244	509009	-1	765	YHR205W	SCH9	uni	350bp before ATG

PrimSeqpUG35promless.ab1									
NewNoisyClones_D09-PrimSeqpUG35promless.ab1	chr02	451878	452748	-1	870	YBR106W	PHO88	uni	contains part of the ORF
NewNoisyClones_D10-PrimSeqpUG35promless.ab1	chr03	127501	128269	-1	768	YCR008W	SAT4	uni	201bp before ATG
NewNoisyClones_D11-PrimSeqpUG35promless.ab1	chr02	590863	591660	-1	797			uni	no known promoter found
NewNoisyClones_D12-PrimSeqpUG35promless.ab1	chr10	83124	83469	-1	345	YJL184W	GON7	uni	contains part of the ORF
NewNoisyClones_E01-PrimSeqpUG35promless.ab1	chr15	891588	892152	1	564	YOR306C	MCH5	uni	156bp before ATG
NewNoisyClones_E02-PrimSeqpUG35promless.ab1	chr02	451900	452748	-1	848			uni	no known promoter found
NewNoisyClones_E03-PrimSeqpUG35promless.ab1	chr07	208700	208920	-1	220	YGL157W	ARI1	uni	78bp before ATG
NewNoisyClones_E04-PrimSeqpUG35promless.ab1	chr02	562929	563345	1	416			uni	no known promoter found
NewNoisyClones_E05-PrimSeqpUG35promless.ab1	chr02	562929	563346	1	417			uni	no known promoter found
NewNoisyClones_E06-PrimSeqpUG35promless.ab1	chr05	511634	512449	-1	815	YER166W	DNF1	uni	300bp before ATG
NewNoisyClones_E07-PrimSeqpUG35promless.ab1	chr11	526259	526848	1	589	YKR048C	NAP1	uni	contains part of the ORF
NewNoisyClones_E08-PrimSeqpUG35promless.ab1	chr09	282274	283107	1	833	YIL038C	NOT3	uni	contains part of the ORF
NewNoisyClones_E09-PrimSeqpUG35promless.ab1	chr10	702391	703569	-1	1178	YJR146W		uni	316bp before ATG
NewNoisyClones_E10-PrimSeqpUG35promless.ab1	chr04	1164914	1165625	-1	711			uni	no known promoter found
NewNoisyClones_E11-PrimSeqpUG35promless.ab1	chr15	948866	949834	1	968	YOR335C	ALA1	uni	243bp before ATG
NewNoisyClones_E12-PrimSeqpUG35promless.ab1	chr13	252855	253267	1	412	YML008C	ERG6	uni	contains part of the ORF
NewNoisyClones_F01-PrimSeqpUG35promless.ab1	chr09	282274	283123	1	849	YIL038C	NOT3	uni	contains part of the ORF
NewNoisyClones_F02-PrimSeqpUG35promless.ab1	chr09	82948	83285	1	337	YIL143C	SSL2	uni	contains part of the ORF
NewNoisyClones_F03-PrimSeqpUG35promless.ab1	chr12	837517	837938	-1	421			uni	no known promoter found
NewNoisyClones_F04-PrimSeqpUG35promless.ab1	chr08	150121	150556	-1	435			uni	no known promoter found
NewNoisyClones_F05-PrimSeqpUG35promless.ab1	chr09	82948	83285	1	337	YIL143C	SSL2	uni	contains part of the ORF
NewNoisyClones_F06-PrimSeqpUG35promless.ab1	chr16	876197	877017	1	820	YPR166C	MRP2	uni	ORF in the fragment
NewNoisyClones_F07-PrimSeqpUG35promless.ab1	chr16	408770	409140	1	370	YPL078C	ATP4	uni	26bp before ATG

NewNoisyClones_F08-PrimSeqpUG35promless.ab1	chr16	876197	877185	1	988	YPR166C	MRP2	uni	ORF in the fragment
NewNoisyClones_F09-PrimSeqpUG35promless.ab1	chr13	556608	557065	1	457	YMR145C	NDE1	uni	133bp avant
NewNoisyClones_F10-PrimSeqpUG35promless.ab1	chr02	562929	563346	1	417			uni	no known promoter found
NewNoisyClones_F11-PrimSeqpUG35promless.ab1	chr13	451404	452214	1	810	YMR091C	NPL6	uni	39bp before ATG
NewNoisyClones_F12-PrimSeqpUG35promless.ab1	chr05	491978	492695	1	717	YER159C	BUR6	uni	contains part of the ORF
NewNoisyClones_G01-PrimSeqpUG35promless.ab1	chr10	192138	192507	-1	369	YJL117W	PHO86	uni	24bp before ATG
NewNoisyClones_G02-PrimSeqpUG35promless.ab1	chr16	411695	412607	-1	912	YPL075W	GCR1	uni	contains part of the ORF
NewNoisyClones_G03-PrimSeqpUG35promless.ab1	chr09	82948	83285	1	337	YIL143C	SSL2	uni	contains part of the ORF
NewNoisyClones_G04-PrimSeqpUG35promless.ab1	chr15	510811	511846	-1	1035	YOR099W	KTR1	uni	contains part of the ORF
NewNoisyClones_G05-PrimSeqpUG35promless.ab1	chr07	949385	950259	1	874			uni	no known promoter found
NewNoisyClones_G06-PrimSeqpUG35promless.ab1	chr04	1164914	1165625	-1	711			uni	no known promoter found
NewNoisyClones_G07-PrimSeqpUG35promless.ab1	chr05	490821	491359	1	538	YER158C		uni	260bp before ATG
NewNoisyClones_G08-PrimSeqpUG35promless.ab1	chr15	256031	256456	-1	425	YOL036W		uni	289bp before ATG
NewNoisyClones_G09-PrimSeqpUG35promless.ab1	chr10	192138	192507	-1	369	YJL117W	PHO86	uni	24bp before ATG
NewNoisyClones_G10-PrimSeqpUG35promless.ab1	chr15	101154	101566	1	412			uni	no known promoter found
NewNoisyClones_G11-PrimSeqpUG35promless.ab1	chr14	491233	491519	-1	286	YNL071W	LAT1	uni	4 bp before ATG
NewNoisyClones_G12-PrimSeqpUG35promless.ab1	chr14	491233	491519	-1	286	YNL071W	LAT1	uni	4 bp before ATG
NewNoisyClones_H01-PrimSeqpUG35promless.ab1	chr09	82948	83285	1	337	YIL143C	SSL2	uni	contains part of the ORF
NewNoisyClones_H02-PrimSeqpUG35promless.ab1	chr02	716392	716854	1	462	YBR248C	HIS7	uni	contains part of the ORF
NewNoisyClones_H03-PrimSeqpUG35promless.ab1	chr05	511443	512449	-1	1006	YER166W	DNF1	uni	300bp before ATG
NewNoisyClones_H04-PrimSeqpUG35promless.ab1	chr02	562929	563346	1	417			uni	no known promoter found
NewNoisyClones_H05-PrimSeqpUG35promless.ab1	chr12	845812	846281	1	469			uni	no known promoter found
NewNoisyClones_H06-PrimSeqpUG35promless.ab1	chr04	872147	872910	-1	763			uni	no known promoter found
NewNoisyClones_H07-	chr12	902921	903799	-1	878	YLR390W-A	CCW14	uni	contains part of the ORF

PrimSeqpUG35promless.ab1									
NewNoisyClones_H08-PrimSeqpUG35promless.ab1	chr13	668119	668624	-1	505	YMR203W	TOM40	uni	contains part of the ORF
NewNoisyClones_H09-PrimSeqpUG35promless.ab1	chr10	130803	131553	1	750	YJL155C	FBP26	uni	150bp before ATG
NewNoisyClones_H10-PrimSeqpUG35promless.ab1	chr09	82948	83285	1	337	YIL143C	SSL2	uni	contains part of the ORF
NewNoisyClones_H11-PrimSeqpUG35promless.ab1	chr13	558857	559144	-1	287	YMR147W		uni	55bp before ATG
NewNoisyClones_H12-PrimSeqpUG35promless.ab1								uni	No alignment found
AP9	chr14	51899	52725	1	826	YNL309W	STB1	uni	contains part of the ORF
AP22	chr15	778858	779788	-1	930			uni	no known promoter found
AP35		507925	509009	1	1084			uni	no known promoter found

Appendix III: List of the polymorphisms

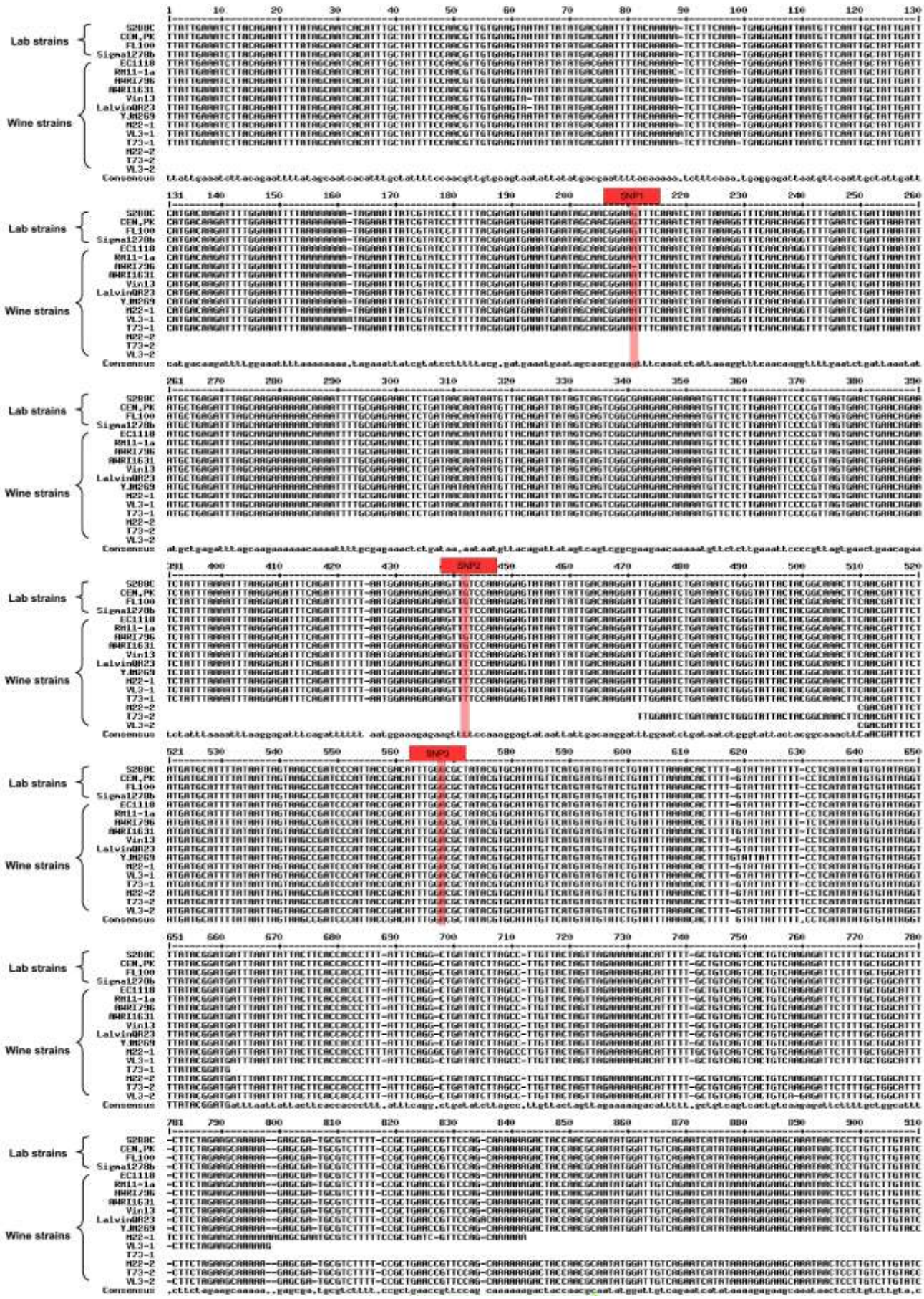
#Chr	Pos	Ref S228c	Alt EC1118	Gene	Name	Number of fragments containing the promoter
chr02	501576	A	T	YBR132C	AGP2	1
chr02	501990	C	T			
chr02	502058	G	A			
chr02	502111	T	C			
chr02	502224	G	A			
chr02	502292	C	T			
chr02	502305	A	T			
chr02	502313	G	A			
chr02	716467	C	T	YBR248C	HIS7	1
chr02	716483	G	A			
chr03	128224	T	C	YCR008W	SAT4	1
chr04	652634	C	CC	YDR099W	BMH2	2
chr04	652695	T	G			
chr04	653246	T	C			
chr04	653336	C	T			
chr04	653403	T	C			
chr04	1468519	A	AA	YDR508C	GNP1	2
chr04	1468562	G	A			
chr04	1468662	AA	A			
chr04	1468720	C	A			
chr04	1468817	C	A			
chr04	1468887	TT	T			
chr04	1468900	C	T			
chr04	1468935	A	G			
chr04	1469066	C	T			
chr04	1469161	A	G			
chr04	1469246	G	A			
chr04	1469247	A	AA			
chr04	1469271	A	G			
chr04	1469289	T	A			
chr04	1469311	T	A			
chr05	33551	A	G	YEL063C	CAN1	1
chr05	33621	A	AA			
chr05	33782	C	A			
chr05	34040	A	G			
chr05	34050	G	A			
chr05	34052	T	A			
chr05	34310	G	A			
chr05	34320	G	GG			
chr05	140700	C	T	YEL009C	GCN4	1
chr05	492129	C	T	YER159C	BUR6	1
chr05	511800	T	C	YER166W	DNF1	2

chr05	511824	C	T			
chr05	545025	G	T	YER177W	BMH1	2
chr05	545199	GGGG	G			
chr06	74522	TT	T	YFL031W	HAC1	1
chr06	74531	A	AA			
chr06	74548	C	T			
chr06	74566	G	GG			
chr06	74604	A	ATATATA			
chr06	74709	A	AAA			
chr06	74817	G	C			
chr06	74848	A	T			
chr06	74849	G	C			
chr06	74850	AA	A			
chr06	74858	A	G			
chr06	74862	A	G			
chr06	74864	A	G			
chr07	76247	A	G	YGL225W	VRG4	2
chr07	76310	A	AA			
chr07	76490	C	T			
chr07	76512	G	A			
chr07	76519	G	A			
chr07	76663	G	C			
chr07	208776	A	ATTTCTATAAA TTCGGCCGAA	YGL157W	ARI1	1
chr07	208835	G	T			
chr08	198394	G	C	YHR046C	INM1	2
chr08	198485	TTTT	T			
chr08	198549	T	C			
chr08	198648	A	G			
chr08	215155	C	T	YHR055C	CUP1	1
chr08	215281	C	A			
chr08	215511	C	T			
chr08	508824	AAAA	A	YHR205W	SCH9	1
chr09	83061	C	T	YIL143C	SSL2	5
chr09	83067	T	C			
chr09	83076	G	A			
chr09	83076	G	GT			
chr09	83081	T	G			
chr09	83083	TT	T			
chr09	83087	T	C			
chr09	83104	A	C			
chr10	83220	T	C	YJL184W	GON7	1
chr10	83271	G	C			
chr10	83298	G	A			
chr10	83388	G	A			
chr10	131062	TT	T	YJL155C	FBP26	1
chr10	131168	C	T			

chr10	131236	T	C			
chr10	131240	C	T			
chr10	131358	C	T			
chr10	192487	A	G	YJL117W	PHO86	2
chr10	207460	ATTGCA	A			
chr10	207695	T	A			
chr10	207758	AAA	A			
chr10	703339	A	G	YJR146W		1
chr10	703431	C	T			
chr10	703435	T	G			
chr10	703440	AA	A			
chr10	703446	T	A			
chr10	703466	C	A			
chr10	745270	A	G	YJR162C		1
chr10	745275	C	A			
chr10	745277	A	C			
chr10	745278	T	G			
chr10	745284	T	A			
chr10	745310	G	T			
chr10	745315	G	GAG			
chr10	745319	G	GT			
chr10	745321	A	G			
chr10	745322	T	TA			
chr10	745385	G	A			
chr10	745390	G	C			
chr12	902948	C	A	YLR390W-A	CCW14	1
chr12	902960	T	A			
chr12	902967	C	T			
chr12	902969	C	T			
chr12	902970	T	C			
chr12	902979	T	TT			
chr12	902989	T	TT			
chr12	903013	CT	C			
chr13	253031	T	A	YML008C	ERG6	1
chr13	451463	T	C	YMR091C	NPL6	1
chr13	451494	T	C			
chr13	451502	T	C			
chr13	451590	A	G			
chr13	451743	C	T			
chr13	452064	G	A			
chr13	452181	C	T			
chr13	556842	G	A	YMR145C	NDE1	1
chr13	558877	G	A	YMR147W		1
chr13	558884	A	G			
chr13	558885	GA	G			
chr13	558919	T	C			

chr13	558987	A	G			
chr13	559059	AAA	A			
chr13	559107	A	AA			
chr13	559132	G	T			
chr13	668220	C	A	YMR203W		
chr14	345712	A	G	YNL154C	YCK2	1
chr14	491381	T	C	YNL071W YNL071W	LAT1	2
chr14	491385	A	C			
chr15	259230	T	TT	SNR50/snR50		2
chr15	511142	G	A	YOR099W	KTR1	1
chr15	511243	T	A			
chr15	511379	T	C			
chr15	511443	T	C			
chr15	511578	A	AAA			
chr15	511736	A	C			
chr15	511740	C	T			
chr15	511741	T	TC			
chr15	511746	C	T			
chr15	891617	G	A	YOR306C	MCH5	1
chr15	891758	C	T			
chr15	891827	G	C			
chr15	892024	G	C			
chr15	949796	G	GG	YOR335C	ALA1	1
chr16	121484	A	G	YPL226W	NEW1	1
chr16	121557	G	A			
chr16	121585	TT	T			
chr16	408842	C	T	YPL078C	ATP4	1
chr16	408889	G	A			
chr16	876647	T	C	YPR166C	MRP2	2
chr16	876799	A	AA			
chr16	876811	T	C			
chr16	876853	A	G			
chr16	876931	T	C			
chr16	877132	G	A			
chr16	877133	C	CC			
chr16	877177	T	TT			

Appendix IV: Alignment of the *CUP1* promoter and ORF in 17 strains




Transcription initial site
Deletion
Start codon
SNP synonymous
Lab strains
Wine strains
S288C
CEN.PK
FL100
Sigma1278b
EC1118
RM11-1a
AWRI796
AWRI1631
Vin13
LalvinQR23
YJM269
M22-1
VL3-1
T73-1
M22-2
T73-2
VL3-2
Consensus

Appendix V: Transcription factor binding sites in *pCUP1$_{S288c}$* and *pCUP1$_{EC1118}$*

Transcription Factor	Consensus	Strand	Position on *pCUP1$_{S288c}$* (size 1000)	Position on *pCUP1$_{EC1118}$* (size 996)
Ash1p	YTGAT	F	-878	-874
Ash1p	YTGAT	F	-755	-751
Ash1p	YTGAT	F	-700	-696
Ash1p	YTGAT	F	-526	-522
Ash1p	YTGAT	F	-311	-307
Ash1p	YTGAT	R	-12	-12
Ash1p	YTGAT	R	-16	-16
Ash1p	YTGAT	R	-20	
Ash1p	YTGAT	R	-104	-100
Azf1p	AAGAAAAA	F	-728	-724
Azf1p	AAGAAAAA	F	-44	-40
Cat8p, Sip4p	YCCNYTNRRCCGN	F	-199	-195
Cup2p	HTHNNGCTGD	F	-317	-313
Cup2p	HTHNNGCTGD	F	-274	-270
Cup2p	HTHNNGCTGD	F	-245	-241
Cup2p	HTHNNGCTGD	F	-201	-197
Cup2p	HTHNNGCTGD	R	-175	-171
Cup2p	GCGTCTTTTCCGCTGA	F	-207	-203
Cup2p	TCTTTTGCTG	F	-246	-242
Cup2p	TCTTTTTTGCTG	R	-172	-168
Fkh1p, Fkh2p	RYMAAYA	F	-130	-126
Fkh1p, Fkh2p	RYMAAYA	R	-539	-535
Fkh1p, Fkh2p	RYMAAYA	R	-873	-869
Gcn4p	TGATTCA	F	-877	-873
Gcn4p	TGACTGA	R	-258	-254
Gcn4p	TGACTMT	R	-669	-665
Gcr1p	CTTCC	R	-792	
Gcr1p	CWTCC	R	-345	-341
Gcr1p	CWTCC	R	-792	
Hac1p	CCAGC	F	-185	-181
Hac1p	CCAGC	R	-235	-231
Hsf1p binding site	NGAANNTTCN	F	-797	-793
Hsf1p	NGAANNTTCN	R	-787	-783
Hsf1p	NTTCNNGAAN	F	-230	-226
Hsf1p	NTTCNNGAAN	R	-220	-216
Mcm1p	DCCYWWWNNRG	R	-773	-769
Mcm1p	CCYWWWNNRG	R	-774	-770
Mot3p	ATGGAT	F	-159	-155
Mot3p	AAGGWT	F	-778	-774
Mot3p	AAGGWT	F	-767	-763
Mot3p	AAGGWT	F	-539	-535
Mot3p	AAGGWT	R	-822	-818
Rgt1p	CGGANNA	F	-351	-347
Rgt1p	CGGANNA	R	-195	-191
Rtg1p, Rtg3p	GTCAC	F	-262	-258
Stb5p	CGGNS	F	-667	-663
Stb5p	CGGNS	R	-631	-627
Tec1p	RMATTCYY	F	-641	-637
Yap1p, Cad1p, Yap3p, Cin5p, Yap5p	TTACTAA	R	-461	-457
Yap1p	TKACAAA	F	-921	-917
Yrr1p	WCCGYKKWW	F	-199	-195
Gsm1p	CGGNNNNNNNNNCGG	R	-445	-441

Appendix VI: List of strains

Number	Gene type	Box of conservation
JF1094	CEN.PK *MATa ura3-52*	stock of the team
JA0001	library before FACS selection	stock for strain JA
JA0002	library after FACS selection	stock for strain JA
JA-S	Isolated clones sequenced	sequenced-JA
JF1581	BY4720 *MATα lys2Δ0 trp1Δ63 ura3Δ0*	stock of the team
JA0111	JA0100 *leu2:: pBMH1$_{S288C}$-yEGFP*	JA -7 gene
JA0112	JA0100 *leu2:: pBMH1$_{EC1118}$-yEGFP*	JA -7 gene
JA0113	JA0100 *leu2:: pBMH2$_{S288C}$-yEGFP*	JA -7 gene
JA0114	JA0100 *leu2:: pBMH2$_{EC1118}$-yEGFP*	JA -7 gene
JA0115	JA0100 *leu2:: pCAN1$_{S288C}$-yEGFP*	JA -7 gene
JA0116	JA0100 *leu2:: pCAN1$_{EC1118}$-yEGFP*	JA -7 gene
JA0117	JA0100 *leu2:: pCUP1$_{S288C}$-yEGFP*	JA -7 gene
JA0118	JA0100 *leu2:: pCUP1$_{EC1118}$-yEGFP*	JA -7 gene
JA0119	JA0100 *leu2:: pHAC1$_{S288C}$-yEGFP*	JA -7 gene
JA0120	JA0100 *leu2:: pHAC1$_{EC1118}$-yEGFP*	JA -7 gene
JA0121	JA0100 *leu2:: pGNP1$_{S288C}$-yEGFP*	JA -7 gene
JA0122	JA0100 *leu2:: pGNP1$_{EC1118}$-yEGFP*	JA -7 gene
JA0123	JA0100 *leu2:: pYCK2$_{S288C}$-yEGFP*	JA -7 gene
JA0124	JA0100 *leu2:: pYCK2$_{EC1118}$-yEGFP*	JA -7 gene
JA0125	JA0100 *leu2:: pAGP2$_{S288C}$-yEGFP*	JA -7 gene
JA0126	JA0100 *leu2:: pAGP2$_{EC1118}$-yEGFP*	JA -7 gene
JA0127	S288c *CUP1-YFP-Kan*	JA -7 gene
JA0128	S288c *Δcup1::amdS*	JA -7 gene
JA0129	S288c: *Δcup2*	JA -7 gene
JA0130	BY4720: *Δcup2*	JA -7 gene
JA0131	JA0129 *leu2:: pPHO5$_{TATA}$-CUP1-kan*	JA -7 gene
JA0132	JA0129 *leu2:: pPHO5$_{UAS}$-CUP1-kan*	JA -7 gene
JA0134	JA0129 *leu2:: pPHO5$_{TATA}$-YFP-kan*	JA -7 gene
JA0135	JA0129 *leu2:: pPHO5$_{UAS}$-YFP-kan*	JA -7 gene
JA0200	KV133 *(*BY4742 *MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0, FLO1::URA3) LEU2*	stock for strain JA
JA0201	JA0200 *leu2:: pPHO5$_{TATA}$-CFP-kan*	stock for strain JA
JA0202	JA0200 *leu2:: pPHO5$_{UAS}$-CFP-kan*	stock for strain JA
JA0203	JA0200 *leu2:: pPHO5$_{TATA}$-RAD27-kan*	stock for strain JA
JA0204	JA0200 *leu2:: pPHO5$_{UAS}$-RAD27-kan*	stock for strain JA
JA0205	JA0200 *leu2:: pPHO5$_{TATA}$-RAD52-kan*	stock for strain JA
JA0206	JA0200 *leu2:: pPHO5$_{UAS}$-RAD52-kan*	stock for strain JA
JA0207	JA0200 *leu2:: pPHO5$_{TATA}$-RAD52-YFP-kan*	stock for strain JA
JA0208	JA0200 *leu2:: pPHO5$_{UAS}$-RAD52-YFP-kan*	stock for strain JA
JA0209	JA0200 *leu2:: pPHO5$_{TATA}$-YFP-kan*	stock for strain JA
JA0210	JA0200 *leu2:: pPHO5$_{UAS}$-YFP-kan*	stock for strain JA
JA0211	JA0200 *leu2:: pPHO5$_{TATA}$-RAD27-kan Δrad27*	stock for strain JA
JA0212	JA0200 *leu2:: pPHO5$_{UAS}$-RAD27-kan Δrad27*	stock for strain JA
JA0213	JA0200 *leu2:: pPHO5$_{TATA}$-RAD52-kan Δrad52*	stock for strain JA
JA0214	JA0200 *leu2:: pPHO5$_{UAS}$-RAD52-kan Δrad52*	stock for strain JA
JA0215	JA0200 *leu2:: pPHO5$_{TATA}$-RAD52-YFP-kan Δrad52*	stock for strain JA
JA0216	JA0200 *leu2:: pPHO5$_{UAS}$-RAD52-YFP-kan Δrad52*	stock for strain JA
JA0217	JA0200 *Δrad27*	stock for strain JA
JA0218	JA0200 *Δrad52*	stock for strain JA
JA0219	JA0200 *RAD27-YFP-kanR*	stock for strain JA
JA0220	JA0200 *RAD52-YFP-kanR*	stock for strain JA
JA0221	JA0200 *leu2:: pPHO5$_{UAS}$-YFP-nat*	stock for strain JA

JA0222	JA0200 *leu2:: pPHO5$_{UAS}$-RAD27-nat*	stock for strain JA
JA0223	JA0200 *leu2:: pPHO5$_{UAS}$-RAD52-nat*	stock for strain JA
JA0224	JA0200 *leu2:: pPHO5$_{UAS}$-RAD52-YFP-nat*	stock for strain JA
JA0225	JA0200 *leu2:: pPHO5$_{UAS}$-RAD27-nat Δrad27*	stock for strain JA
JA0226	JA0200 *leu2:: pPHO5$_{UAS}$-RAD52-nat Δrad52*	JA -7 gene
JA0227	JA0200 *leu2:: pPHO5$_{UAS}$-RAD52-YFP-nat Δrad52*	stock for strain JA
JA0228	JA0200 *leu2:: pPHO5$_{TATA}$-RAD27-YFP-kan*	stock for strain JA
JA0229	JA0200 *leu2:: pPHO5$_{UAS}$-RAD27-YFP-kan*	stock for strain JA
JA0230	JA0200 *leu2:: pPHO5$_{TATA}$-RAD27-YFP-kan Δrad27*	stock for strain JA
JA0231	JA0200 *leu2:: pPHO5$_{UAS}$-RAD27-YFP-kan Δrad27*	stock for strain JA
JA0232	JA0200 *leu2:: pPHO5$_{UAS}$-RAD27-YFP-nat*	stock for strain JA
JA0233	JA0200 *leu2:: pPHO5$_{UAS}$-RAD27-YFP-nat Δrad27*	stock for strain JA
JA0234	JA0200 *leu2:: pHIS3$_{M1}$-RAD52-YFP-nat Δrad52*	JA -7 gene
JA0235	JA0200 *leu2:: pHIS3$_{M4}$-RAD52-YFP-nat Δrad52*	JA -7 gene
JA0236	JA0200 *leu2:: pHIS3$_{M1}$-RAD27-YFP-kan Δrad27*	JA -7 gene
JA0237	JA0200 *leu2:: pHIS3$_{M4}$-RAD27-YFP-kan Δrad27*	JA -7 gene
JA0300	BY4720 *pSDT1-URA3*	stock for strain JA
JA0301	JA0300 *leu2:: pPHO5$_{TATA}$-CFP-kan*	stock for strain JA
JA0302	JA0300 *leu2:: pPHO5$_{UAS}$-CFP-kan*	stock for strain JA
JA0303	JA0300 *leu2:: pPHO5$_{TATA}$-RAD27-kan*	stock for strain JA
JA0304	JA0300 *leu2:: pPHO5$_{UAS}$-RAD27-kan*	stock for strain JA
JA0305	JA0300 *leu2:: pPHO5$_{TATA}$-RAD52-kan*	stock for strain JA
JA0306	JA0300 *leu2:: pPHO5$_{UAS}$-RAD52-kan*	stock for strain JA
JA0307	JA0300 *leu2:: pPHO5$_{TATA}$-RAD52-YFP-kan*	stock for strain JA
JA0308	JA0300 *leu2:: pPHO5$_{UAS}$-RAD52-YFP-kan*	stock for strain JA
JA0309	JA0300 *leu2:: pPHO5$_{TATA}$-YFP-kan*	stock for strain JA
JA0310	JA0300 *leu2:: pPHO5$_{UAS}$-YFP-kan*	stock for strain JA
JA0311	JA0300 *leu2:: pPHO5$_{TATA}$-RAD27-kan Δrad27*	stock for strain JA
JA0312	JA0300 *leu2:: pPHO5$_{UAS}$-RAD27-kan Δrad27*	stock for strain JA
JA0313	JA0300 *leu2:: pPHO5$_{TATA}$-RAD52-kan Δrad52*	stock for strain JA
JA0314	JA0300 *leu2:: pPHO5$_{UAS}$-RAD52-kan Δrad52*	stock for strain JA
JA0315	JA0300 *leu2:: pPHO5$_{TATA}$-RAD52-YFP-kan Δrad52*	stock for strain JA
JA0316	JA0300 *leu2:: pPHO5$_{UAS}$-RAD52-YFP-kan Δrad52*	stock for strain JA
JA0317	JA0300 *leu2:: pPHO5$_{UAS}$-YFP-nat*	stock for strain JA
JA0318	JA0300 *leu2:: pPHO5$_{UAS}$-RAD27-nat*	stock for strain JA
JA0319	JA0300 *leu2:: pPHO5$_{UAS}$-RAD52-nat*	stock for strain JA
JA0320	JA0300 *leu2:: pPHO5$_{UAS}$-RAD52-YFP-nat*	stock for strain JA
JA0321	JA0300 *leu2:: pPHO5$_{UAS}$-RAD27-nat Δrad27*	stock for strain JA
JA0322	JA0300 *leu2:: pPHO5$_{UAS}$-RAD52-nat Δrad52*	stock for strain JA
JA0323	JA0300 *leu2:: pPHO5$_{UAS}$-RAD52-YFP-nat Δrad52*	stock for strain JA
JA0324	JA0300 *leu2:: pPHO5$_{TATA}$-RAD27-YFP-kan*	stock for strain JA
JA0325	JA0300 *leu2:: pPHO5$_{UAS}$-RAD27-YFP-kan*	stock for strain JA
JA0326	JA0300 *leu2:: pPHO5$_{TATA}$-RAD27-YFP-kan Δrad27*	stock for strain JA
JA0327	JA0300 *leu2:: pPHO5$_{UAS}$-RAD27-YFP-kan Δrad27*	stock for strain JA
JA0328	JA0300 *leu2:: pPHO5$_{UAS}$-RAD27-YFP-nat*	stock for strain JA
JA0329	JA0300 *leu2:: pPHO5$_{UAS}$-RAD27-YFP-nat Δrad27*	JA -7 gene
JA0330	JA0300 after 5-FOA P1	stock for strain JA
JA0331	JA0300 after 5-FOA P2	stock for strain JA
JA0332	JA0300 after 5-FOA B5	stock for strain JA
JF2350	59A *MATα Δamn1-loxP*	stock of the team
JA0511	JA0500 *leu2:: pBMH1$_{S288C}$-yEGFP*	JA -7 gene
JA0512	JA0500 *leu2:: pBMH1$_{EC1118}$-yEGFP*	JA -7 gene
JA0513	JA0500 *leu2:: pBMH2$_{S288C}$-yEGFP*	JA -7 gene
JA0514	JA0500 *leu2:: pBMH2$_{EC1118}$-yEGFP*	JA -7 gene

JA0515	JA0500 *leu2:: pCAN1$_{S288c}$-yEGFP*	JA -7 gene
JA0516	JA0500 *leu2:: pCAN1$_{EC1118}$-yEGFP*	JA -7 gene
JA0517	JA0500 *leu2:: pCUP1$_{S288c}$-yEGFP*	JA -7 gene
JA0518	JA0500 *leu2:: pCUP1$_{EC1118}$-yEGFP*	JA -7 gene
JA0519	JA0500 *leu2:: pHAC1$_{S288c}$-yEGFP*	JA -7 gene
JA0520	JA0500 *leu2:: pHAC1$_{EC1118}$-yEGFP*	JA -7 gene
JA0521	JA0500 *leu2:: pGNP1$_{S288c}$-yEGFP*	JA -7 gene
JA0522	JA0500 *leu2:: pGNP1$_{EC1118}$-yEGFP*	JA -7 gene
JA0523	JA0500 *leu2:: pYCK2$_{S288c}$-yEGFP*	JA -7 gene
JA0524	JA0500 *leu2:: pYCK2$_{EC1118}$-yEGFP*	JA -7 gene
JA0525	JA0500 *leu2:: pAGP2$_{S288c}$-yEGFP*	JA -7 gene
JA0526	JA0500 *leu2:: pAGP2$_{EC1118}$-yEGFP*	JA -7 gene
JA0527	JA0500 *CUP1-YFP-Kan*	JA -7 gene
JA0528	JA0500 *CUP1-YFP-Nat*	JA -7 gene
JA0529	JA0500 *pCUP1::amdSYM-CUP1-YFP-kan*	JA -7 gene
JA0530	JA0500 *Δcup1::amdS*	JA -7 gene
JA0531	JA0500 *CAN1-YFP-Kan*	stock for strain JA
JA0532	JA0500 Δcup2	JA -7 gene
JA0533	JA0532: *pPHO5$_{TATA}$-CUP1-kan*	JA -7 gene
JA0534	JA0532: *pPHO5$_{UAS}$-CUP1-kan*	JA -7 gene
JA0536	JA0532: *pPHO5$_{TATA}$-YFP-kan*	JA -7 gene
JA0537	JA0532: *pPHO5$_{UAS}$-YFP-kan*	JA -7 gene
JA0538	JA0500 *pCUP1$_{Del}$- YFP-kan*	JA -7 gene
JA0539	JA0500 *pCUP1$_{SNP793}$-YFP-kan*	JA -7 gene
JA0540	JA0500 *pCUP1$_{SNP563}$-YFP-kan*	JA -7 gene
JA0541	JA0500 *pCUP1$_{SNP437}$-YFP-kan*	JA -7 gene
JA0542	JA0500 *pCUP1$_{S288C}$-zeoR-kan*	JA -7 gene
JA0543	JA0500 *pCUP1$_{EC1118}$-zeoR-kan*	JA -7 gene
JA0545	JA0500 *pCUP1$_{SNP-combine}$-YFP-kan*	JA -7 gene
JA0546	JA0500 *pCUP1$_{S288C}$-zeoR-YFP-kan*	JA -7 gene
JA0547	JA0500 *pCUP1$_{EC1118}$-zeoR-YFP-kan*	JA -7 gene

Appendix VII: List of plasmids

Number	plasmid
pJA001	pJRL2: pPHO5$_{TATA(A1T6)}$-CFP-his-KanMX4-URA3-his
pJA002	pJRL2: pPHO5$_{UASm1}$-YFP-his-KanMX4-URA3-his
pJA003	pJRL2: pPHO5$_{TATA(A1T6)}$-YFP-KanMX4
pJA004	pJRL2: pPHO5$_{UASm1}$-YFP-KanMX4
pJA005	pJRL2: pPHO5$_{TATA(A1T6)}$-yEGFP-KanMX4
pJA006	pJRL2: pBMH1$_{S288C}$-yEGFP-KanMX
pJA007	pJRL2: pBMH1$_{EC1118}$-yEGFP-KanMX
pJA008	pJRL2: pBMH2$_{S288C}$-yEGFP-KanMX
pJA009	pJRL2: pBMH2$_{EC1118}$-yEGFP-KanMX
pJA010	pJRL2:pCAN1$_{S288C}$-yEGFP-KanMX
pJA011	pJRL2:pCAN1$_{EC1118}$-yEGFP-KanMX
pJA012	pJRL2:pCUP1$_{S288C}$-yEGFP-KanMX
pJA013	pJRL2:pCUP1$_{EC1118}$-yEGFP-KanMX
pJA014	pJRL2:pYCK2$_{S288C}$-yEGFP-KanMX
pJA015	pJRL2:pYCK2$_{EC1118}$-yEGFP-KanMX
pJA016	pJRL2:pHAC1$_{S288c}$-yEGFP-KanMX
pJA017	pJRL2:pHAC1$_{EC1118}$-yEGFP-KanMX
pJA018	pJRL2:pAGP2$_{S288c}$-yEGFP-KanMX
pJA019	pJRL2:pAGP2$_{EC1118}$-yEGFP-KanMX
pJA020	pJRL2:pGNP1$_{S288c}$-yEGFP-KanMX
pJA021	pJRL2:pGNP1$_{EC1118}$-yEGFP-KanMX
pJA012	pJRL2:pCUP1$_{S288C}$-ZeoR-KanMX
pJA013	pJRL2:pCUP1$_{EC1118}$-ZeoR-KanMX
pJA022	pJRL2: pPHO5$_{TATA(A1T6)}$-RAD52-KanMX4
pJA023	pJRL2: pPHO5$_{UASm1}$-RAD52-KanMX4
pJA024	pJRL2: pPHO5$_{TATA(A1T6)}$-RAD52-YFP-KanMX4
pJA025	pJRL2: pPHO5$_{UASm1}$-RAD52-YFP-KanMX4
pJA026	pJRL2: pPHO5$_{TATA(A1T6)}$-RAD27-KanMX4
pJA027	pJRL2: pPHO5$_{UASm1}$-RAD27-KanMX4
pJA028	pJRL2: pPHO5$_{TATA(A1T6)}$-RAD27-YFP-KanMX4
pJA029	pJRL2: pPHO5$_{UASm1}$-RAD27-YFP-KanMX4
pJA030	pJRL2: pPHO5$_{UASm1}$-RAD52-Nat
pJA031	pJRL2: pPHO5$_{UASm1}$-RAD52-YFP-Nat
pJA032	pJRL2: pPHO5$_{UASm1}$-RAD27-Nat
pJA033	pJRL2: pPHO5$_{UASm1}$-RAD27-YFP-Nat
pJA034	pFA6a-yEGFP-KanMX4
pJA035	pFA6a-YFP-KanMX4
pJA036	pUG35
pJA037	pUG35 promoterless (+0)
pJA038	pUG35 promoterless (+1)
pJA039	pUG35 promoterless (+2)
pJA-S	98 sequenced plasmids from the library

Appendix VIII: List of Primers

CHECK-4gene-for	CTCAACATAACGAGAACACACA
CHECK-4gene-rev	TCCAGTGAAAAGTTCTTCTCC
LEU2-ADE2-for	TATAAATTGGTGCGTAAAATCGTTGGATCTCTCTTCTAAGTACATATGACAAAACCTCTTCCGAT
LEU2-ADE2-rev	CTGTAAGCGTTGATTTCTATGTATGAAGTCCACATTTGATGTAATCCCTCCTCCTTGTCAATATT
CHECK-Leu2Int-for	GCTACGAACCGGGTAATACTAA
CHECK-Leu2Int-rev	GGCAGAATCAATCAATTGATGT
Rad52-EcoR1-For	GCCGCGGAATTCAACAAAATGAATGAAATTATGGATATGGAT
Rad52-Not1-Rev	TATATAATTACTTTGCGGCCGCTCAAGTAGGCTTGCGTGC
Rad52-YFP-Not1-Rev	GGCATATGCGGCCGCTTATTTGTACAATTCATCCATACC
Rad27-EcoR1-For	CGCCGGAATTCAACAAAATGGGTATTAAAGGTTTGAATG
Rad27-Not1-Rev	TAAATAAGCGGCCGCTCATCTTCTTCCCTTTGTGA
Lys2-Rad52-For	CGTTTTTAAGCTATTTTGCCACTGAGAATCAACAAATGCAAACAACTTCTGCTGCGTATTATTCTGC
Lys2-Rad52-Rev	ATAAATAATGATGCAAATTTTTTATTTGTTTCGGCCAGGAAGCGTTTTAAGCTGCTGCGGAGCTT
Lys2-Rad27-For	CGTTGACAGCATACATTGGAAAGAAATAGGAAACGGACACCGGAAG CTTCTGCTGCGTATTATTCTGC
Lys2-Rad27-Rev	TGCCAAGGTGAAGGACCAAAAGAAGAAAGTGGAAAAAGAACCCCCTTAAGCTGCTGCGGAGCTT
CHECK-Lys2-For-rad52	TGGATTCAACAACTCCCTTG
CHECK-Lys2-For-rad27	CCGGCTGGTAAGTTATGATAGA
CHECK-Lys2-Rev	GCATTGTCCTGGAAAATGTC
CHECK-URA-for	GGAGGGTACCCGCCTCGCAA
CHECK-URA-rev	TTAGTTTTGCTGGCCGCATC
CHECK-PJL2-for	ACATACATAAACATACGCGC
CHECK-PJL2-Rev	TTATCACGTTGAGCCATTAG
Rad52-EcoR1-For	GCCGCGGAATTCAACAAAATGAATGAAATTATGGATATGGAT
Rad52-Not1-Rev	TATATAATTACTTTGCGGCCGCTCAAGTAGGCTTGCGTGC
Rad52-YFP-Not1-Rev	GGCATATGCGGCCGCTTATTTGTACAATTCATCCATACC
Rad27-EcoR1-For	CGCCGGAATTCAACAAAATGGGTATTAAAGGTTTGAATG
Rad27-Not1-Rev	TAAATAAGCGGCCGCTCATCTTCTTCCCTTTGTGA
Lys2-Rad52-For	CGTTTTTAAGCTATTTTGCCACTGAGAATCAACAAATGCAAACAACTTCTGCTGCGTATTATTCTGC
Lys2-Rad52-Rev	ATAAATAATGATGCAAATTTTTTATTTGTTTCGGCCAGGAAGCGTTTTAAGCTGCTGCGGAGCTT
Lys2-Rad27-For	CGTTGACAGCATACATTGGAAAGAAATAGGAAACGGACACCGGAAG CTTCTGCTGCGTATTATTCTGC
Lys2-Rad27-Rev	TGCCAAGGTGAAGGACCAAAAGAAGAAAGTGGAAAAAGAACCCCCTTAAGCTGCTGCGGAGCTT
CHECK-Lys2-For-rad52	TGGATTCAACAACTCCCTTG
CHECK-Lys2-For-rad27	CCGGCTGGTAAGTTATGATAGA
CHECK-Lys2-Rev	GCATTGTCCTGGAAAATGTC
CHECK-URA-for	GGAGGGTACCCGCCTCGCAA
CHECK-URA-rev	TTAGTTTTGCTGGCCGCATC
CHECK-PJL2-for	ACATACATAAACATACGCGC
CHECK-PJL2-Rev	TTATCACGTTGAGCCATTAG
CHECK-4gene-plus-rev	AATGGTCAGGTCATTGAGTG
CHECK-Leu2Int-plus-rev	GGACACCTGTAAGCGTTGAT
CHECK-4gene-add-for	GATGCAAGAGTTCGAATCTCT
Rad52-YFP-kan-for	GAGAAGTTGGAAGACCAAAGATCAATCCCCTGCATGCACGCAAGCCTACTAGTAAAGGAGAAGAACTTTTC

Rad52-YFP-kan-rev	AGTAATAAATAATGATGCAAATTTTTTATTTGTTTCGGCCAGGAAGCGTTCTATTTGTATAGTTCATCCAT
Rad27-YFP-kan-for	AAAATAAAAAATTGAACAAAAATAAGAATAAAGTCACAAAGGGAAGAAGAAGTAAAGGAGAAGAACTTTTC
Rad27-YFP-kan-rev	ATATATGCCAAGGTGAAGGACCAAAAGAAGAAAGTGGAAAAAGAACCCCCCTATTTGTATAGTTCATCCAT
Check-Rad52-YFP –for	CGCGAGGGATTCTGTCTATGAA
Check-Rad27-YFP-for	GCCACCAAAGGAGAAGGAACTT
Check-Rad-YFP-rev	TTGGGATCTTTCGAAAGGGC
FLO-URA-for	TCTAACTGTACTGTCCCTGACC
FLO-URA-rev	TGGTGATTTGTCCTGAAGAT
SDT1-TRsz-For	CAGTAATATAATAGCACGAGGG
692-SDT1-TRsz-F	CCCGCCTCGCAAACTTATTGAT
693-SDT1-TRsz-R	GTAATATGATATCAAGAAGGCG
BMH1-SalI-For	ATTTGTCGACATTTCTATGCAACAAGAATA
BMH1-EcoRI-Rev	CCCGGAATTCTTTTATCTTTAGTTTATCTTTAAC
BMH2-SalI-For(infu)	GCATCTCGAGGTCGAAATAGGGAATCGGTATTTCTG
BMH2-EcoRI-Rev(infu)	CTTTAGACATGAATTTTTTTTTGTTGTAACGGGTAC
CUP1-2-SalI-For	ACGGTCGACACAGAATTTTATAGCAATCAC
CUP1-2-EcoRI-Rev	TTCGGAATTCTTTATGTGATGATTGATTGA
CAN1-SalI-For	ATTGTCGACTGTGTGTATGGGCACAAACC
CAN1-EcoRI-Rev	CCGGAATTCTGCTATGCCTTTTTTTTTTTTG
YCK2-SalI-For	CATTGTCGACTCTCTCGTACTATTAACATACC
YCK2-EcoRI-Rev	CTCGGAATTCTTTTGGAAAACTATTTTCTT
HAC1-SalI-For	ATCTGTCGACCGTGTTCCACTGTGGAGAGC
HAC1-MefI-Rev	CCGCAATTGAGTGGCGGTTGTTGTCGTAG
GNP1-SalI-For	ATCTGTCGACCGTGTATTTATTTGTAAAACT
GNP1-EcoRI-Rev	ACGGAATTCAATGTGCAATATTTGATATT
GFP-EcoRI-for	GCCCCGGAATTCATGTCTAAAGGTGAAGAATTA
GFP-NotI-Rev	ATTGCGGCCGCTTATTTGTACAATTCATCCA
Check-Rad52-YFP-plus-rev	GTTGGGAATGGATAGGTCCGAT
Check-Rad27-YFP-plus-rev	TGGCAAACGAATTACAGCCAG
Rad52-YFP-Add-kan-rev	CTTGTAAAATAATAAGAATTTTTTATTCGATTTAAAGTAAATATTAATACTACTTTCTGCGCACTTAACT
Rad27-YFP-Add-kan-rev	AGGTAAGAATGAAAAATTCCACGTTCAAGTTCCCAGAAAAACTGGCAAAATACTTTCTGCGCACTTAACT
YFP-PacI-for	CGGCCCCGTTAATTAACATGTCTAAAGGTGAAGAATTA
YFP-AscI-Rev	TATTGGCGCGCCTTATTTGTACAATTCATCCA
Rad52-add-YFP-kan-for	GAGAAGTTGGAAGACCAAAGATCAATCCCCTGCATGCACGCAAGCCTACTCGGATCCCCGGGTTAATTAA
Rad27-add-YFP-kan-for	AAAATAAAAAATTGAACAAAAATAAGAATAAAGTCACAAAGGGAAGAAGACGGATCCCCGGGTTAATTAA
FLO-URA-plus-rev	CAACACTACTGCTGTGACCA
FLO-URA-add-for	CAGACGGGTTCTTACACATT
AGP2-SalI-for	CCGGTCGACTCAAGAAATGTGACCATACACC
AGP2-MfeI -Rev	CCGCAATTGGCTTGGCAAAAGTTATGGAA
Lys2-Sir2-For	TTAGTGAAGAGATGTAAAGCCCATTCTCACGTATTTCAAGAAATTCTTCTGCTGCGTATTATTCTGC
Lys2-Sir52-Rev	ATTGATATTAATTTGGCACTTTTAAATTATTAAATTGCCTTCTACTTAAGCTGCTGCGGAGCTT
Chk-Lys2-For-Sir2	CTTGTCTTTTTCACCACCCA
Chk-fusYFP-SIR2-For	GCACAGATAAACTGGTGCAGTG

Chk-fusYFP-MSH2-For	GCAGGAGATTCCCAATTGAA
CAN1-EcoR1-For-in	GAGATTACCAGAATTCAACAAAATGACAAATTCAAAAGAAGAC
CAN1-Not1-Rev-in	AAGCAGAGAGCGGCCGCCTATGCTACAACATTCCAAA
CAN1-add-YFP-kan-for	CATGAACCAAAGACTTTTTGGGACAAATTTTGGAATGTTGTAGCACGGATCCCCGGGTTAATTAA
CAN1-YFP-Add-kan-rev	GATAAAAATAAATATACTGAGATTATAGTAAGCTCATTGATCCCTTACTTTCTGCGCACTTAACT
Chk-fusYFP-CAN1-For	CAACCATTATTTCTGCCGCA
Chk-fusionYFP-rev	ATGCATCATCAGGAGTACGGAT
Ble-CAN1-For	GTTTTCAATCTGTCGTCAATCGAAAGTTTATTTCAGAGTTCTTCAGACATGGAGGCCCAGAATAC
Ble-CAN1-rev	GCGAAATGGCGTGGAAATGTGATCAAAGGTAATAAAACGTCATATTCAGTCCTGCTCCTCGGCCA
Chk-del-CAN1-For	CGGGTGAGTCATACGGCTTT
Chk-del-ble-rev	GATGAACAGGGTCACGTCGT
Lys2-MSH2-For	TCTCTTTATCTGCTGACCTAACATCAAAATCCTCAGATTAAAAGTCTTCTGCTGCGTATTATTCTGC
Lys2-MSH2-Rev	TATATTATCTATCGATTCTCACTTAAGATGTCGTTGTAATATTAATTAAGCTGCTGCGGAGCTT
Chk-Lys2-For-MSH2	CAAGTGAACCTCAACAGCTACA
CUP1-add-YFP-for	TGCGGTAACAAGTCTGAAGAAACCAAGAAGTCATGCTGCTCTGGGAAACGGATCCCCGGGTTAATTAA
CUP-add-YFP -rev	ATCCGCTTCAAATAAATAGATCATTGAAAGTGACGGGGATAACAGCATTTTACTTTCTGCGCACTTAACT
Chk-fuYFP-CUP1-For	TCATGAGTGCCAATGCCAAT
Chk-fuY-CUP1-Nat-R	ACTGATTAGGGGCAGGGCAT
Del-pro-amdS-For	GCAAATCATTTTATTGAAATCTTACAGAATTTTATAGCAATCACATTTGCGACATGGAGGCCCAGAATAC
Del-pro-amdS-rev	GCACTCATGACCTTCATTTTGGAAGTTAATTAATTCGCTGAACATAGAGCTCCAGTATAGCGACCA
Pro-BY-F	CAAGCGGCAATAATCGCTTTC
Pro-BY-R	ACCACATTGGCATTGGCACT
amdS-CAN1-For	GTTTTCAATCTGTCGTCAATCGAAAGTTTATTTCAGAGTTCTTCAGACATGGAGGCCCAGAATAC
amdS -CAN1-rev	GCGAAATGGCGTGGAAATGTGATCAAAGGTAATAAAACGTCATATAGAGCTCCAGTATAGCGACCA
Chk-amdSYM-rev	TGGAGTAACAGTGTGACCAGCC
Chk-fuY-CUP1-plus-R	AACCCACGAAGATGACATGGT
De-CU-hph-F	GCAAATCATTTTATTGAAATCTTACAGAATTTTATAGCAATCACATTTGCAGATTGTACTGAGAGTGCAC
De-CU-hph-R	AAAGACTATTCGTTTCATTTCCCAGAGCAGCATGACTTCTTGGTTTCTTCCTGTGCGGTATTTCACACCG
Chk -del-CU-R	GCCATTGTCCGTCAGGACAT
CUP1-EcoR1-For	GGGGGCGGAATTCAACAAAATGTTCAGCGAATTAATTAACTT
Chk-amdS-For	GGAAGAATTGGCTGCTGACA
De-CU-amdS-F	CTGAATATTGAAGAATTTTTGCCAATCTTTGACAAAACCCTCTTACTAGAGACATGGAGGCCCAGAATAC
De-CU-amdS-R	ATCCGCTTCAAATAAATAGATCATTGAAAGTGACGGGGATAACAGCATTTAGAGCTCCAGTATAGCGACCA
Chk-CU-del-aS-F	GTTTGCAAAGTTAACTTCCCACC
CU-del-aS-Ad-R	TCTAGCGAGTCAGAAGCTGTCAAG
Mut-del-For	CTGTACAATCAATCAATCATCACATAAAATGTTCAGCG
Mut-del-Rev	CGCTGAACATTTTATGTGATGATTGATTGATTGTACAG
Mut-793-For	GATGAAATGAATAGCAACGGAAATTTCAAATCTATTAAAGGTTTC
Mut-793-Rev	GAAACCTTTAATAGATTTGAAATTTCCGTTGCTATTCATTTCATC
Mut-563-For	GATTTTTTAATGGAAAGAGAAGTTTTCCAAAGGAGTATAATTATTGAC
Mut-563-Rev	GTCAATAATTATACTCCTTTGGAAAACTTCTCTTTCCATTAAAAAATC
Mut-437-For	CCCATTACCGACATTTGGACGCTATACGTGCATATGT
Mut-437-Rev	ACATATGCACGTATAGCGTCCAAATGTCGGTAATGGG
CUP1-YFP-Int-For	CTGAATATTGAAGAATTTTTGCCAATCTTTGACAAAACCCTCTTACTAGAATGTTCAGCGAATTAATTAACTT

CUP1-YFP-Int-Rev	ATCCGCTTCAAATAAATAGATCATTGAAAGTGACGGGGATAACAGCATTTGAGGCAAGCTAAACAGATCTAT
CUP1-Not1-Rev	TAAATAAGCGGCCGCTCATTTCCCAGAGCAGCA
Mut-del-For-New	TCACATAAAATGTTCAGCGAA
Mut-del-Rev-New	TGATTGATTGATTGTACAGTTTG
Del-CUP2-hph-for	ATTAGACGGCGGCTTGATAAAAGAGGACTGATAATCAGTGTATTCACAGAAGATTGTACTGAGAGTGCAC
Del-CUP2-hph-rev	CTGCCAGCTTGCCGGGAGAACAAACAACCGCCAATATATGTATATGTATACTGTGCGGTATTTCACACCG
Chk-Del-CUP2-for	GAAAATCGCCAAAACGAGGC
ZeoR-EcoR1-For	GGGGGCGGAATTCATGGCCAAGTTGACCAGTGC
ZeoR-Not1-rev	TAAATAAGCGGCCGCTCAGTCCTGCTCCTCGGCCACG
URA-tdTomato-for	TCTTAACCCAACTGCACAGAACAAAAACCTGCAGGAAACGAAGATAAATCatggtgagcaagggcgagga
URA-tdTomato-rev	TAATTTGTGAGTTTAGTATACATGCATTTACTTATAATACAGTTTTttacttgtacagctcgtcca
URA-GFP-for	TCTTAACCCAACTGCACAGAACAAAAACCTGCAGGAAACGAAGATAAATCATGTCTAAAGGTGAAGAATTA
URA-GFP-rev	TAATTTGTGAGTTTAGTATACATGCATTTACTTATAATACAGTTTTTTATTTGTACAATTCATCCA
URA-tdTomato-dup-rev	TAATTTGTGAGTTTAGTATACATGCATTTACTTATAATACAGTTTTTTCTGGGCCTCCATGTCGCT
ZeoR-Not1/ApaI-rev	TAAATAAGCGGCCGCGGGCCCGTCCTGCTCCTCGGCCACG
GFP-ApaI-for	GATTAGGGGCCCATGTCTAAAGGTGAAGAATTA
CUP1-Not1/ApaI-rev	AATAAGCGGCCGCGGGCCCTTTCCCAGAGCAGCATGAC
q-CUP1-for	TGAAGGTCATGAGTGCCAATGC
q-CUP1-rev	GGGCATTTGTCGTCGCTGTT
q-zeoR-for	GCCAAGTTGACCAGTGCCGT
q-zeoR-rev	TGATGAACAGGGTCACGTCGTC
2/q-CUP1-for	TCAGACTTGTTACCGCAGGGGC
2/q-CUP1-rev	TGAGTGCCAATGCCAATGTGG
1/q-zeoR-for	CGGAAGTTCGTGGACACGACCT
1/q-zeoR-rev	TGTTCATCAGCGCGGTCCAG
2/q-zeoR-for	GCCACGAAGTGCACGCAGTT
2/q-zeoR-rev	TCGTGTCCACGAACTTCCGG
1/q-GFP-for	CCATACCATGGGTAATACCAGCAGC
1/q-GFP-rev	CCAATTGGTGATGGTCCAGTCTTGT
2/q-GFP-for	TCTTGAACATAACCTTCTGGCATGG
2/q-GFP-rev	CCAGTTCCATGGCCAACCTTAGT