Investigating the effect of H3K4 and H3K79 methylation on gene expression noise in *Saccharomyces cerevisiae*

MSc by Research in Biochemistry, University of Oxford



Walaa Khushaim

Oriel College, Trinity Term 2018

Approximate word count: 25, 000

Investigating the effect of H3K4 and H3K79 methylation on gene expression noise in *Saccharomyces cerevisiae*

A thesis submitted to the University of Oxford for the degree of Master by Research
Walaa Khushaim, Oriel College, Trinity Term 2018.

Abstract

Gene expression noise, the stochastic variation in a gene's protein level in a population of genetically identical cells, has been shown to play an important role in several biological processes. However, it is not yet understood how gene expression noise is controlled. Following Newman et al. (2006)'s work, where a library of GFP-tagged genes at COOH termini and FACS was used to analyse noise levels in *S. cerevisiae* for 2,204 genes, we discovered at the genome-wide level a strong correlation between H3K4 methylation, H3K79 methylation, H3K123 ubiquitylation and high levels of noise, out of 1,180 gene features. Histone methyltransferases, Set1 and Dot1 which are responsible for H3K4 and H3K79 methylation, respectively, both depend on ubiquitination of H2B on lysine 123. H3K4 and H3K79 methylation are found on the majority of genes, therefore, the aim of this work was to examine whether the correlation between H3K4 and H3K79 methylation, and noise was causal or not. The hypothesis that removing H3K4 tri-methylation would affect noise was first tested by deleting *SPP1*, which is encoding a subunit of the Set1 complex (Set1C). Deletion of *SPP1*, which primarily affected H3K4me3, developed noise and reduced gene expression in certain genes. However, deletion of *SET1* had the opposite effect on noise and gene expression, at genes that were sensitive to changes in H3K4 methylation. Both of these results were consistent with the observed bioinformatic correlation. Despite the strong correlation between noise and H3K4 methylation, only ~20% of the 18 genes tested showed changes in gene expression noise with a change in levels of H3K4 methylation, suggesting that the effect is limited to certain genes. Since H3K79 methylation also shows a strong correlation with noise, further work was conducted to investigate the possibility that H3K79 methylation functions redundantly with H3K4 methylation. Interestingly, Dot1 redundantly functions with Spp1 and only at genes where the loss of H3K4me3 increases noise. The increase in noise that was observed on the loss of H3K4me3 was reduced to the wild-type level in the double knock-out of *DOT1* and *SPP1*. In conclusion, this study provides evidence that gene expression noise may be regulated by histone methylation, particularly H3K4 and H3K79, and that Dot1 (H3K79me3) could redundantly function with Spp1 (H3K4me3).

Acknowledgements

I would like to acknowledge many people who contribute towards this thesis. My sincere gratitude to my advisor, Jane Mellor, who gave me the opportunity to work in her laboratory. She has been very supportive during my project work, and I am really grateful for her friendly advices and constructive criticism that have taught me a lot. Special thanks to Anitha Nair, the laboratory technician who has been very kind and helpful when first I arrived at the lab. I would like to thank Françoise Howe who taught me everything I needed to know during my first year, and she provided me with her ChIP data analysis that are presented in Table 6. I also would like to thank Struan Murray for the analysis of the Newman's noise data and other genome-wide analyses featured in this thesis. Philip Lorenz and Joseph Rayappu constructed some of the strains and performed FACS experiments that present in this study in Table 7. I am also grateful to the members of Mellor lab. They made me feel immediately welcome when I first arrived at the lab, and they made coming to the lab an enjoyable experience. Finally, I would like to thank my family for being supportive throughout the last two years.

Abbreviations

Abbreviation	Term in full
FACS	fluorescence-activated cell sorting
SAGA	Spt-Ada-Gcn5 acetyltransferase
TBP	TATA-binding protein
TFIID	RNA-polymerase II transcription factor D
BSA	bovine serum albumin
ChIP	chromatin immunoprecipitation
ChIP-qPCR	chromatin immunoprecipitation-quantitative real-time PCR
ChIP-seq	chromatin immunoprecipitation-sequencing
COMPASS	complex of proteins associated with Set1
CTD	C-terminal domain
COV	Standard deviation-to-mean ratio
EDTA	ethylene diamine tetra-acetic acid
ac	acetylation
H2Bub	H2B monoubiquitination
HAT	histone acetyltransferase
HDAC	histone deacetylase
KMT	lysine methyltransferase
me	methylation
me1	monomethylation
me2	dimethylation
me3	trimethylation
mRNA	messenger RNA
nt	nucleotide
OD	optical density
PHD	plant homeodomain
PIC	pre-initiation complex
qPCR	quantitative PCR
RNA FISH	RNA fluorescence in situ hybridisation
ub	ubiquitination
WT	wildtype
YPD	yeast extract – bactopeptone – glucose
YPG	yeast extract – bactopeptone – galactose

In this thesis, 'yeast' will be used interchangeably with *S. cerevisiae*, 'noise' will be used interchangeably with gene expression noise.

Table of contents

Abstract I
Acknowledgements II
Abbreviations III
Table of contents IV
Table of Figures VI
Index of Tables VII
1. Introduction 2
1.1. Gene expression is a highly regulated process 2
1.2. Gene expression noise 3
1.2.1. Noise plays an important role in all organisms 4
1.2.2. Using FACS and GFP-tagged genes to determine noise 4
1.3. Origins of noise 6
1.3.1. Extrinsic and intrinsic noise 6
1.3.2. Transcriptional and translational bursting contributions to noise 7
1.4. What factors can influence noise? 11
1.5. Histone modifications 14
1.6. The role of H3 methylation and acetylation in gene expression 15
1.6.1. H3K4 methylation 16
1.6.2. H3K79 methylation 19
1.6.3. H3 acetylation and H3K9ac 20
1.7. Crosstalk between histone modifications 21
1.8. Aims: testing causal links between (epi)genetic features and noise 22
2. Materials and Methods 24
2.1. Cloning and yeast transformation 24
2.2. Cell Cultures and GAL10 induction 24
2.3. Preparing cells for FACS 25
2.4. Measuring noise 25
2.5. Statistical analysis 25
2.6. Chromatin Immunoprecipitation (ChIP) - qPCR 26
2.7. Yeast strains 28
2.8. Tested Genes 32
3. Investigating the role of H3K4 methylation in noise 34
3.1. Introduction 34
3.1.1. H3K4 methylation strongly correlates with noise 34
3.1.2. Genes with different noise levels show a reduction in H3K4me3 upon the loss of SPP1 36
3.1.3. Aims 37
3.2. Results 38
3.2.1. GAL10 induction in galactose in WT strain 38
3.2.2. Tagging genes of interest with GFP in WT and mutant strains 40
3.2.3. Deletion of SPP1 decreases gene expression and increases noise at URA4 41
3.2.4. SML1 shows changes in noise but not expression levels 43

3.2.5. SET1 affects both gene expression levels and noise at GAL10, URA4, ARG1 and SWD1 .. 45
3.2.6. SML1 shows changes in noise but not expression levels in loss of SET1 46
3.2.7. Effect of H3K4me is limited to certain genes 47
3.3. Discussion..... 50
4. Investigating the link between H3K4me3 with H3K79me and its role in noise 54
4.1. Introduction..... 54
4.1.1. Crosstalk between H3K79me and H3K4me3 55
4.1.2. H3K4me and H3K79me intensities corelate with transcriptional bursting 56
4.1.3. Set1 and Dot1 function together..... 58
4.1.4. Aims 60
4.2. Results 61
4.2.1. Loss of DOT1 affect gene expression and noise at SML1 61
4.2.2. The double knock-out of spp1Δ-dot1Δ increases noise and decreases expression at ARG1. 63
4.2.3. The loss of DOT1 prevents the effect that was observed in spp1Δ at URA4 and SML1..... 64
4.2.4. Reduction of H3K79me2/me3 levels upon the loss of Dot1 67
4.2.5. Increase in H3K79me3 levels upon the loss of SPP1 68
4.2.6. Increase in H3K79me3 levels upon the loss of SET1 at promoter region 71
4.3. Discussion..... 73
5. Investigating the role of H3K9ac in noise..... 79
5.1. Introduction..... 79
5.2. Aims 80
5.3. Results 81
5.3.1. Loss of SGF29 decreases expression and increases noise 81
5.3.2. Decrease in H3K9ac upon the loss of Spp1 84
5.4. Discussion..... 88
6. Discussion 90
6.1. Does H3K36me3 redundantly function with H3K4me3 and H3K79me3? 91
6.2. Future work..... 92
6.3. Concluding remarks 93
7. Appendix..... 94
8. Bibliography..... 95

Table of Figures

Figure 1. Noise representation. 3
Figure 2. The use of FACS to determine distributions of protein levels in individual cells..... 5
Figure 3. YFP and CFP are expressed from two identical copies of a specific gene. 7
Figure 4. Schematic shows how gene expression can occur continuously (Mellor, unpublished, 2013). 9
Figure 5. Correlation between transcription burst size (Fano factor) and mean expression levels from three different studies for each species. 10
Figure 6. Factors that corrrelate with transcriptional bursting parameters. 11
Figure 7. Pearson correlation analysis between 1,180 different genomic features and noise..... 13
Figure 8. Set1 interactions within COMPASS and its domain structure (P. Lorenz, unpublished).. 16
Figure 9. H3K4me3 peaks before (in orange), simultaneously (dark blue) or after their transcripts peaks (light blue) in a number of genes..... 18
Figure 10. Set1 and Dot1 are relevant enzymes that recruits H3K4 and H3K79 methylation respectively. Adapted from Howe, unpublished, 2016. 22
Figure 11. The seven sub-units of *S. cerevisiae's* Set1 complex. 35
Figure 12. Relative fluorescence intensities of GAL10 and GAL10_GFP in WT strain. 38
Figure 13. Histograms show relative fluorescence intensities of GAL10_GFP in WT strain. 39
Figure 14. Relative fluorescence intensities of GFP-tagged strains..... 40
Figure 15. Deletion of SPP1 decreases expression levels and increases noise. 42
Figure 16. spp1Δ have no effects on gene expression levels for the 8 target genes in WT and spp1Δ strains. 43
Figure 17. spp1Δ affects SML1 *noise.* 44
Figure 18. set1Δ affects gene expression and noise levels. 45
Figure 19. set1Δ affects SML1 noise. Note: *, $P < 0.05$. Error bars= standard deviation between experiments (N=3). 46
Figure 20. Crosstalk between H2BK123ub1 and H3K4/H3K79 methylation..... 56
Figure 21. Correlations of the intensities of histone modification with transcriptional bursting parameters..... 57
Figure 23. The link between Spp1 and Dot1 is still unknown. ..... **Error! Bookmark not defined.**
Figure 24. The loss of DOT1 affects only SML1 expression and noise levels..... 62
Figure 25. Dot1 redundantly functions with Spp1 in ARG1. 63
Figure 26. The loss of both DOT1-SPP1 inhibit the observed increase in noise in spp1Δ. 64
Figure 27. No change in both gene expression and noise levels in dot1Δ-spp1Δ. 65
Figure 28. The loss of both DOT1-SET1 does not alter the effect of the loss SET1 at URA4. 66
Figure 29. Reduction in H3K79me3 and H3K79me2 in the loss of DOT1..... 67
Figure 30. H3K79me3 levels increase at the promoter region of the tested genes upon the loss of SPP1. 69
Figure 31. Decrease in H3K79me2 levels upon the loss of SPP1 at promoter region except for SML1. 70
Figure 32. Levels of H3K79me3 and H3K79me2 upon the loss of SET1. 71
Figure 33. No change in either expression or noise levels in loss of sgf29Δ. 81
Figure 34. A decrease in expression level and an increase in noise upon the loss of SGF29 in four genes..... 83
Figure 35. Reduction in H3K9ac upon the loss of Sfg29. 84
Figure 36. Decrease in H3K9ac upon the loss of SPP1 in certain genes..... 86
Figure 37. Increase in H3K9ac upon the loss of Set1. 87

Index of Tables

Table 1. Sequences for primers used in real-time PCR. 27
Table 2. Details of the antibodies used during this study. 27
Table 3. Ratios of H3K4me2 over H3K4me3 in the +1 nucleosomes of the 5 target genes. 28
Table 4. The description of each gene tagged with GFP and analysed with FACS. 32
Table 5. Enzyme levels correlation with noise. 35
Table 6. Genome-wide ChIP data generated from the lab shows values of H3K4me2 and H3K4me3 at the promoter regions (0:300) of the target genes analysed in this study and by past laboratory members in *spp1*Δ (S. Murray, unpublished). 37
Table 7. The effects of set1Δ and spp1Δ on individual gene expression levels and expression noise (genes that were analysed in this study and by past laboratory members). Note: ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$. (N=3). 48
Table 8. Enzyme levels correlation with noise 58
Table 9. Yeast genes analysed with FACS experiment to test the link between SPP1 and DOT1 and their function in noise (SGD). 59
Table 10. Plasmids used in this study 94

Chapter 1.

Introduction

1. Introduction

1.1. Gene expression is a highly regulated process

Gene expression is an essential cellular process by which information from a gene is used in the formation of a functional gene product, using stored genetic information from DNA. In eukaryotes, the process of gene expression, is complex, tightly regulated and dependant on a wide range of mechanisms that the cells use at any step of the gene expression pathway. Gene expression can be up-regulated or down-regulated depending on many circumstances such as cellular structure, cellular function under development, response to stress, metabolism, and alterations in the environment. For example, yeast cells can adapt to environment changes by switching nearly 10% of their genes 'on' or 'off' in order to increase their chances of survival (Gasch *et al.* 2000; Causton *et al.* 2001). During gene expression, an RNA is formed from an initial DNA template (transcription) to produce either a functional gene product or a messenger RNA (mRNA), which is exported from the nucleus. In the cytosol, mRNA is translated by the ribosomes to produce a polypeptide (translation), and then post-translational modifications modify folded polypeptides. Although regulation of gene expression can occur at each of these stages it mainly takes place at the transcription level. Failure to correctly manage this process at any given stage could lead to the development of different diseases including cancer, thus emphasising the importance of gene expression and the necessity to understand it, especially with respect to transcription mechanisms and the circuitry of regulation.

1.2. Gene expression noise

Gene expression noise, the stochastic variation in a gene's protein level in a population of genetically identical cells, has been a topic of interest over the past decade (Fig.1) (Blake *et al*., 2003; Kaern *et al.,* 2005; Raser & O'Shea, 2005). The concept of stochasticity in gene expression was first introduced in 1957 by Novick and Weiner. In single bacteria, they studied the lac-operon expression level in response to thio-methylgalactoside (TMG), which is an artificial non-metabolizable inducer. TMG cannot be degraded by the induced enzyme (β-galactosidase), therefore, it was expected an average increase of expression level of the enzyme. Surprisingly, they found that some cells were expressing a large amount of β-galactosidase, while other cells were expressing a small amount under the up-regulation of the production of β-galactosidase (Novick & Weiner, 1957). This was the start of noise studies.

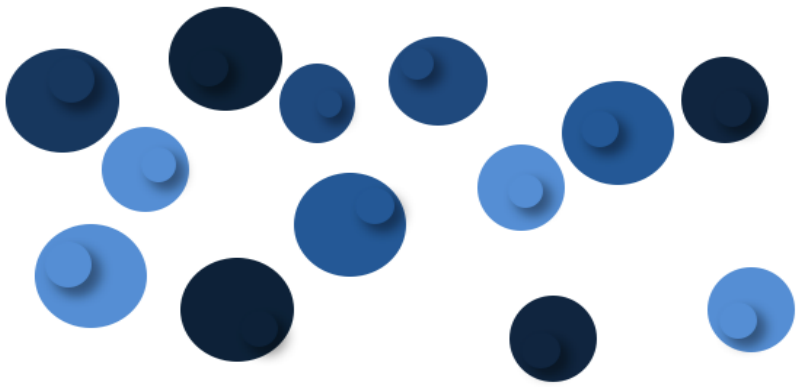


Figure 1. Noise representation.

Variation in gene expression is shown in a clonal population. The intensity of the colour is representing level of expression (S. Murray, 2016, unpublished)**.**

1.2.1. Noise plays an important role in all organisms

Observed in all organisms from prokaryotes to eukaryotes, gene expression noise is thought to affect an organism's stress response, pathogenesis, metabolism, development, cell cycles, aging and circadian rhythms (Raj & Van Oudenaarden, 2008). Gene expression noise has been implicated in several biological processes including: (a) controlling the determination between lysis and latency in HIV (Weinberger *et al*., 2005; Singh and Weinberger, 2009); (b) playing a role in cell fate decision during organismal development (Fraser and Kaern, 2009); (c) contributing to a heterogeneous population of tumour cells (Brock *et al.,* 2015). Indeed, the distinctive phenotypic equilibria that appears in breast cancer stem-like cells is showed to be based on gene expression noise and that over time, equilibrium fractions of subpopulation are reached in the purified cancer cells (Gupta *et al*., 2011). Furthermore, gene expression noise could lead to a delay in DNA repair in Escherichia coli. Uphoff et al., (2016), found that undamaged cells produce an average of one Ada, a protein that is responsible for repairing DNA damage in E. coli. Because of gene expression noise, some of the damaged cells are not producing any of the Ada. Therefore, a delay in response for repairing DNA can occur resulting in a subpopulation that have high rate of mutations (Uphoff et al., 2016). However, the mechanism by which the cell regulates and maintains variation is not yet understood.

1.2.2. Using FACS and GFP-tagged genes to determine noise

Expression levels can be analysed in single cells using Fluorescence-activated cell sorting (FACS) which measures the fluorescence intensity of GFP tagged strains in single cells across a population of 10,000 cells. As depicted in Fig. 2, a common measure of gene expression noise is the Coefficient of Variation (COV = standard deviation in protein expression level divided by the mean expression level) (Munsky *et al*., 2012; Thattai & Van Oudenaarden, 2001).

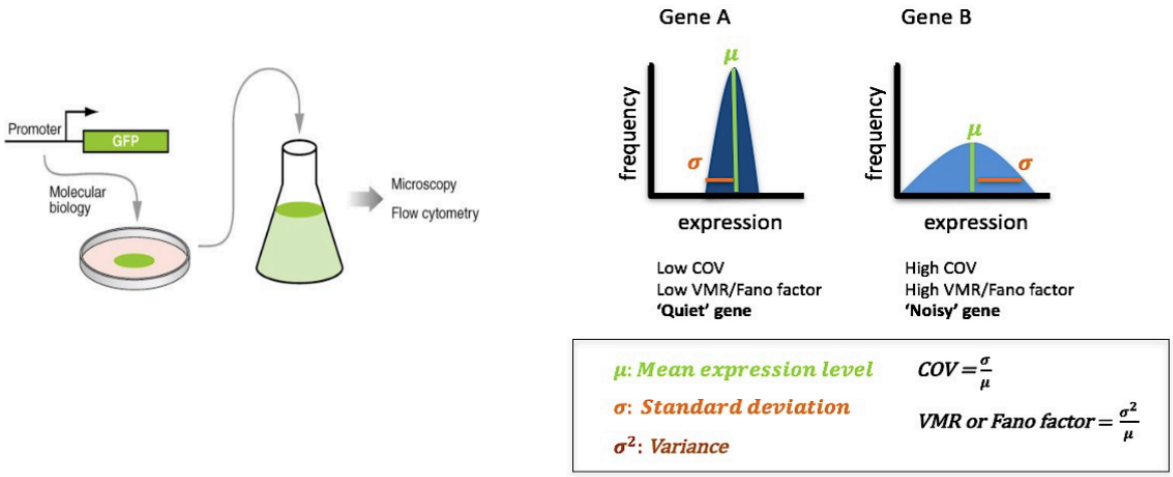


Figure 2. The use of FACS to determine distributions of protein levels in individual cells.

FACS histograms show intrinsic noise. "Noisy" genes have high COV or VMR, while "quiet" genes have low COV or VMR. Modified from (P. Lorenz, unpublished; Mellor, unpublished, 2013).

In living organisms, several recombinant protein tags, including green fluorescent protein (GFP), are widely used to determine gene expression, noise and protein localisation and dynamics (Cubitt *et al.,* 1995; Ludin & Matus, 1998; Newaman *et al.,* 2006). The GFP-tag, which is made by the jellyfish *Aequorea victoria*, was discovered in 1955 (Davenport *et al.,* 1955). GFP-tag can be transferred to and expressed in organisms such as bacteria, yeasts, plants, insects, fishes and mammals (Chalfie *et al.,* 1994; Niedenthal *et al.,* 1996; Casper & Holt, 1996; Wang & Hazelrigg, 1994; Moss *et al.,* 1996; Ludin *et al.,* 1996). Unlike many proteins, fixation or permeabilisation of cells is not required to detect the GFP signal (Wu *et al.,* 2017). The GFP-tag is generally non-toxic to cells, has a long sequence (27-kDalton monomer protein) and a long half-life. Because GFP relatively is large as a tag, it could affect the degradation rate of mRNA or protein of tagged genes. Moreover, domains involved in interactions with mRNA binding proteins or a steric hindrance around a catalytic domain of the tagged genes could be affected by the GFP-tag (Chalfie *et al.,* 1994). Some problems with the background autofluorescence have been also linked with GFP-tag (Hermann *et al.,* 2005; Coburn *et al.,* 2013). Therefore, FACS results for GFP-tagged genes should be interpreted with caution.

1.3. Origins of noise

1.3.1. Extrinsic and intrinsic noise

A large dynamic range of expression noise across the genome makes some genes noisier than others (Lehner, 2008; Newman *et al.,* 2006). The mechanisms and factors by which the cell controls specific genes' variation are not yet fully identified. However, all molecular interactions and reactions could inequitably contribute to cell-to-cell variation. In particular, there are two types of expression noise based on the origin of stochastic variation in a gene's protein levels: extrinsic noise and intrinsic noise (Swain *et al*., 2002; Maheshri & O'Shea, 2007).

Extrinsic noise affects multiple genes' expression and originates from extrinsic global cellular factors and machineries. Variation of the numbers of RNA polymerase, ribosome availability, or the proteasome could affect cellular capacity in terms of synthesis and degradation rates leading to cell-to-cell variation (Elowitz *et al.*, 2002; Raj & Van Oudenaarden, 2008). Moreover, the amino acid pool used for protein synthesis can be randomly and irregularly diluted due to variation in cell volumes caused by uneven division of cells (Hooshangi *et al.,* 2005). Cell-to-cell variation in metabolic intake and signalling can also be caused by fluctuations resultant from nutrient availability (Huh & Paulsson, 2011). The cell cycle is also major contributor of extrinsic noise (Becskei *et al.,* 2005).

Intrinsic noise on the other hand, affects each gene individually and originates from intrinsic sources and gene-specific factors such as transcription and translation. Intrinsic noise features or factors can take place at two distinct stages: *cis* and *trans*. *Cis*-acting regulatory factors are the DNA, RNA or protein's sequence or structural aspect that can expose a specific gene to a set of regulation events. For example, a gene can be a potential target of a specific transcriptional factor (TF) because of the gene's special motif that can be specifically recognised by the active or expressed TF. *Trans*-acting regulators such as transcription factors, RNA-binding proteins, post-translational and modulating enzymes, for example, are molecules that can target specific genes. Noise can result from the

combination of both *cis* and *trans* factors because this composite can impact the steady-state expression levels of the gene (reviewed by Chalancon *et al.,* 2012).

Intrinsic noise is the primary cause of noise in eukaryotes but in prokaryotes on the other hand, extrinsic noise is the main cause (Ko *et al.,* 1990; Golding *et al.,* 2005) and this difference may result from chromatin remodelling in eukaryotes (Raj & Van Oudenaarden, 2008). In order to exclude extrinsic noise, targeted genes can also be tagged with a red florescence protein (RFP). The two fluorescent proteins (GFP, RFP, YFP or CFP) can be incorporated into two alleles of the same endogenous gene (Fig. 3). Controlled by identical promoters, this can be used to determine intrinsic noise. The fluorescence difference between the two fluorescent proteins can only contribute to intrinsic noise because, in the same intracellular environment, two fluorescent proteins are expressed from two identical copies of a specific gene (Elowitz *et al*., 2002; Wu *et al.,* 2017).

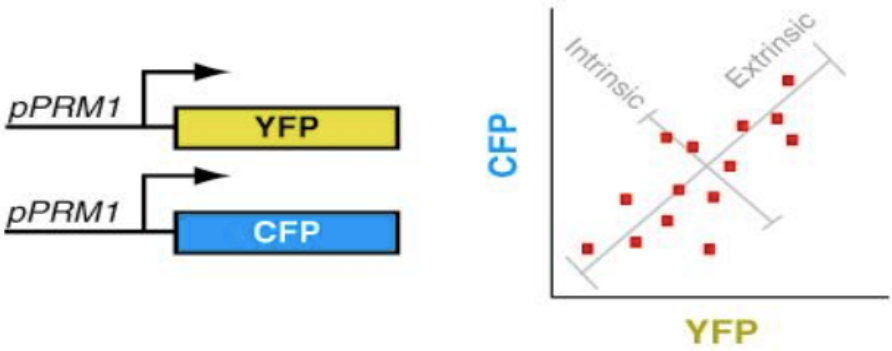


Figure 3. YFP and CFP are expressed from two identical copies of a specific gene.

Intrinsic noise cause variation in YFP and CFP levels while extrinsic noise makes YFP and CFP brighter or dimmer (Mellor, unpublished, 2013).

1.3.2. Transcriptional and translational bursting contributions to noise

A strong association between gene regulation mechanisms and noise has been observed in many studies that measured the variability of mRNA and protein abundance. Therefore, gene expression models on mRNA, protein production, processing and degradation have been used to describe noise.

Previously, Poisson distributions were used to characterise mRNA and protein levels in cells (Brenner *et al.,* 2015). Recent studies have shown that gene expression can occur in a burst-like mode. Random bursts can take place at both transcriptional and translational levels. Furthermore, some studies on noise have shown that fluctuations in transcription could impact both mRNA and protein variation (Fraser *et al.,* 2004; Munsky *et al.,* 2012; Wu *et al.,* 2017).

Transcription is a fundamental cellular process and a key point in controlling the gene expression pathway. In eukaryotes, three RNA polymerase (Pol) complexes, RNA polymerase I, II, and III, are responsible for transcription of nuclear DNA. RNA polymerase II is required to transcribe protein-coding genes. Gene transcription involves three stages: initiation, elongation and termination (reviewed by Hahn & Young, 2011) During initiation, gene-specific transcription factors recruit the subunits in the pre-initiation complex (PIC) to a gene promoter, the freely dissociated polymerase then binds to DNA and synthesises the initial base pairs of RNA. Transcription elongation begins when the polymerase progresses along the DNA and RNA forms. In transcription termination, the polymerase is released from the DNA, and the mRNA molecule is produced by the processing of the RNA 3' end. As transcriptional bursting, where a number of mRNA are being produced in short pulses, is the dominant transcription mode in eukaryotes, it serves as a major explanation of the overall gene expression noise (Fig. 4) (Raj *et al.,* 2006; Blake *et al.,* 2003). Gene expression noise changes inversely with burst frequency (repetition of burst events), whereas it changes proportionally with burst size (average number of mRNA molecules made per burst event) (Paulsson, 2004). Transcriptional initiation is considered to be the most influential step for generating noise. At transcriptional initiation, the assembly of the pre-initiation complex, RNA polymerase II processivity and availability of transcription factors can affect transcriptional bursting parameters, which are unlikely to be determined by transcriptional elongation (Gerber & Shilatifard, 2003; Nguyen & Zhang, 2011; reviewed by Nicolas *et al.,* 2017).

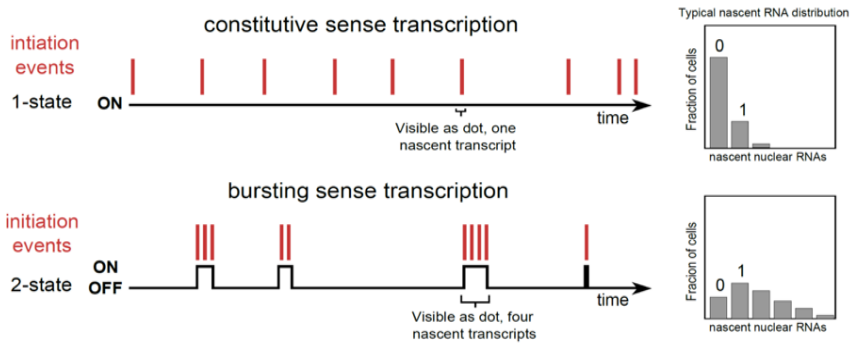


Figure 4. Schematic shows how gene expression can occur continuously (Mellor, unpublished, 2013).

Gene expression occurs in a burst-like mode. This figure shows the distribution of nascent transcription in the nucleus of a cell (detected by RNA-FISH (right)).

Although changes in mRNA levels describe the variations that are observed at protein levels, in eukaryotes, mRNA and protein levels are only weakly correlated in a population of genetically identical cells. The ratio of levels of protein to mRNA in a single cell remain inconsistent (Pradet-Balade *et al.,* 2001; de Sousa Abreu *et al.,* 2009). For example, Pradet-Balade et al., (2001) shows that, for given expression levels in yeast, mRNAs were weakly correlated with their corresponding proteins (r = 0.57). This low correlation was also observed in 9 various tissues derived from 57 human cell lines, where it showed an even weaker correlation between mRNA and protein abundance for glutathione-S-transferase (r = 0.43) (Balade *et al.,* 2001).

Translational burst, which is a brief period during which RNAs are translated to proteins (Blake *et al*., 2003), depends on the transcript and ribosome availability (Raj *et al.,* 2006; Weinberger *et al.*, 2012). Several studies in mammalian cells (Raj *et al.,* 2006) and budding yeast *S. cerevisiae* (Zenklusen *et al,* 2008), have reported that both mRNA and protein levels can be affected by translational bursts. Dacheux *et al.,* (2017) has explained the poor correlation between transcriptional bursting and protein noise by demonstrating a significant contribution of translation and ribosomal scanning process to protein noise. The insertion of a poly(G) motif and stem-loop, types of structural elements, into the 5' untranslated location of mRNA could generate noise as well as inhibition of translational initiation (Dacheux *et al.,* 2017). Furthermore, translation efficiency has been shown to

be positively correlated with increased levels of noise in *Bacillus* (Ozbudak *et al.,* 2002) and a similar correlation was observed in *Saccharomyces cerevisiae* (Blake *et al.,* 2003). Higher noise levels were shown to have a correlation with higher adaption of the tRNA index (a measure of intracellular tRNA usage for gene codons) (Salari *et al.,* 2012). Therefore, generation of a high level of noise can be a result of both weak transcription and efficient translation (Ozbudak *et al.,* 2002; Blake *et al.,* 2003).

Indeed, because the rates of transcriptional and translational bursts regulate the level of RNAs and proteins produced, it has been suggested that transcriptional and translational bursts are inversely linked to noise (Paulsson, 2004; Cai *et al.,* 2006). In fact, highly expressed genes have a larger length of burst size and are considered as "noisier" genes. Although gene expression noise has been studied in different organisms including *E. coli*, *S. cerevisiae* and mammalian cells, they have shown different bursting behavior (Sanchez *et al.,* 2013) (Fig. 5).

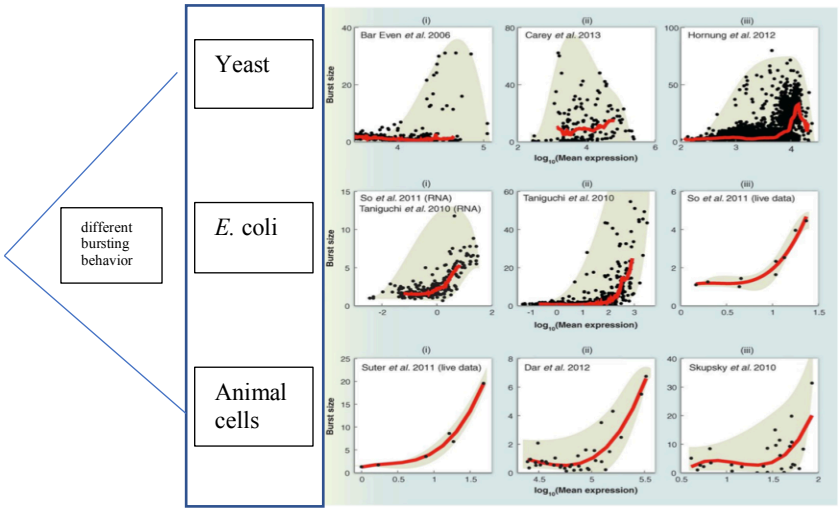


Figure 5. Correlation between transcription burst size (Fano factor) and mean expression levels from three different studies for each species.

High mean expression levels in yeast bursty behavior was not significant when compared to the rise in burst size correlating with a rise in mean expression levels in both *E. coli* and mammalian cells. Adapted from Sanchez *et al.,* 2013.

Higher eukaryotes have large variations in gene expression due to the global constraints, e.g. a more regulated state of chromatin and because eukaryotic cells such as mammalian cells, show very bursty mRNA and protein distributions (Fig. 5). This makes gene expression noise in higher eukaryotes particularly important (Golding *et al.,* 2005; Sanchez *et al.,* 2013) and therefore, studying the causal impact of many features that have been shown to correlate with noise in a simple eukaryotic model such as yeast appears to be a sensible starting point.

1.4. What factors can influence noise?

Higher levels of noise can result from a number of specific molecular features have been identified in the last few decades. These features arise from intrinsic factors such as the presence of a TATA box sequence at the promoter region, high turnover rates of the nucleosome, high number of TF binding sites, promoter architecture and transcriptional control all of which have been shown to contribute to increases levels of gene expression noise (Choi & Kim, 2009; Boeger *et al*., 2008; Blake *et al.,* 2006; Newman *et al.,* 2006; Ravarani *et al.,* 2016). Other factors including local chromatin environment, nucleosome occupancy, histone modifications and a number of cis-regulatory elements, that could affect transcriptional bursting parameters may also influence noise (Fig. 6) (reviewed by Nicolas *et al.,* 2017; reviewed by Chalancon *et al.,* 2012; Weinberger *et al.* 2012).

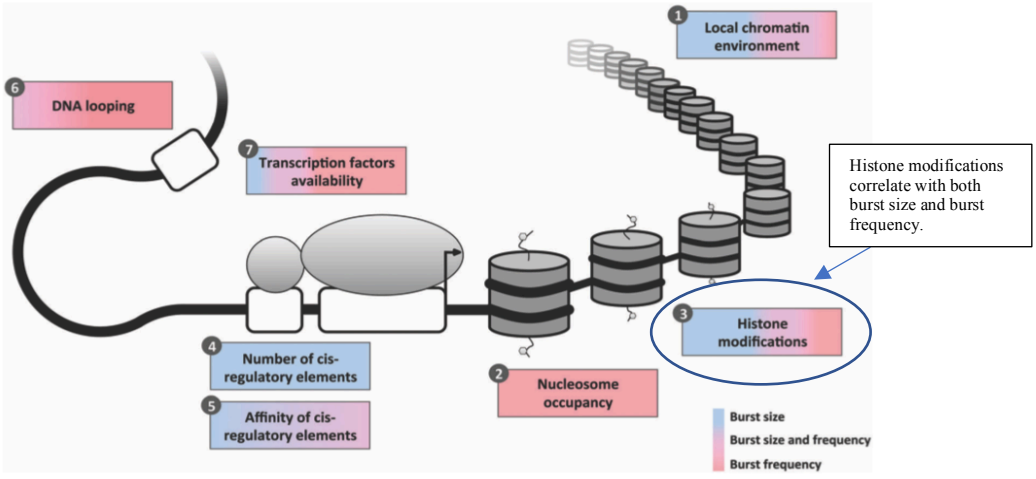


Figure 6. Factors that corrrelate with transcriptional bursting parameters.

Adapted from Nicolas *et al.,* 2017.

Recent studies have investigated the effect of some epigenetic and chromatin remodeling factors, such as the SAGA, Spt-Ada-Gcn5-acetyltransferase complex, which can promote preinitiation complex formation by its histone acetyltransferase function, on noise (Golding *et al.,* 2005; Ravarani *et al.,* 2016). For example, studies in yeast have shown higher levels of noise are observed in the presence of a TATA-box motif at the promoter region due to the variation of nucleosome occupancy and pre-initiation complex dynamics that resulted from the presence of the TATA-box motif (Raser & O'Shea, 2004; Blake *et al.,* 2006; Newman *et al.,* 2006; Brown & Boeger, 2014). Ravarani et al., (2016) showed that noise associated with TATA-box genes is due to the preferential utilisation of TATA-box genes to bind to SAGA rather than TFIID (first protein that binds to DNA during the formation of the PIC). This study also found that, in TATA-box genes, higher recycling of monomeric TBP by Mot1p (protein regulates TATA-binding protein (TBP)) resulted from the high competition between SAGA (Spt3) and Mot1p to bind to TBP. High noise level is generated due to the increase in waiting time between the transcription "on" and "off" states resultant from the high recycling of monomeric TBP. Therefore, the loss of *SPT3* in SAGA complex decreases the competition between SAGA and Mot1p, as well as decreasing noise. Low abundance of Mot1p or high abundance of SAGA also decreases noise (Ravarani *et al.,* 2016). However, comparing to another study, the presence of a TATA-box motif in the promoter region has not been shown to be correlated with transcriptional bursting parameters that could affect protein noise (Wu *et al.,* 2017).

Post-translational modifications – such as methylation – which can be added to histones' tails that protrude from the nucleosome, can play important roles in regulating gene expression by the control of DNA accessibility (Kouzarides, 2007; Venkatesh and Workman, 2015). As histone modifications have been shown to be correlated with transcriptional bursting parameters (Fig. 6, reviewed by Nicolas *et al.,* 2017), previous studies have shown that gene expression noise could be regulated by histone modifications. Weinberger *et al.*, (2012) found that Rpd3(L)C, a histone deacetylase complex, could indirectly regulate gene expression noise. Another study, Wu et al., (2017), reported that deleting *DOT1* – required for the addition of all types of methylation at lysine 79 on histone 3 (H3K79) in *S. cerevisiae* – increases gene expression noise for highly expressed genes by increasing

their transcriptional busting frequency. It is worth noting however that out of the seven genes studied in Wu *et al.,* (2017), this change in noise was only observed in two genes.

Protein levels have been shown to be inversely linked with noise (Elowitz *et al.,* 2002; Bar-Even *et al.,* 2006; Newman *et al.,* 2006. For instance, Newman et al. (2006) analysed noise levels in *S. cerevisiae* for 2,204 genes using a library of GFP-tagged genes along with FACS, they reported a strong negative correlation between gene expression and noise. However, at lower protein abundance, this relation is not always observed in mammalian cells at all time (Raj *et al.,* 2006). Following Newman's et al., (2006) work, we found at the genome-wide level some strong correlations in 1,180 gene features, including histone modifications, and noise (Fig. 7).

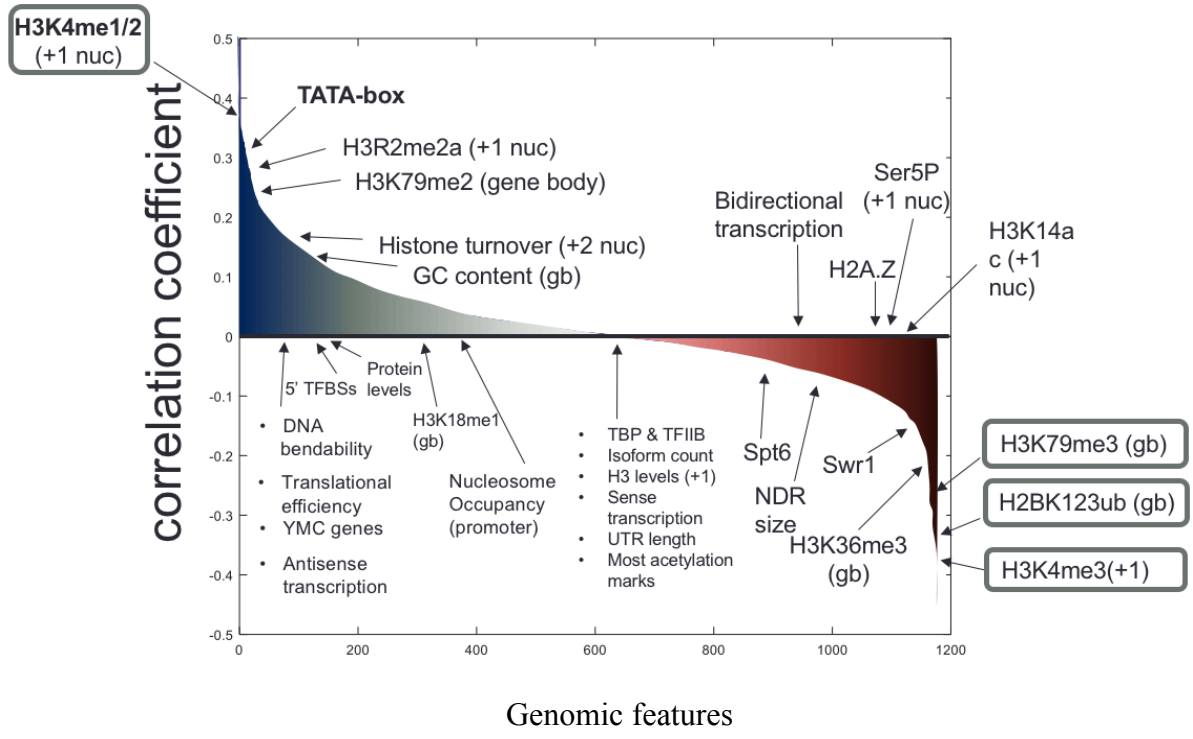


Figure 7. Pearson correlation analysis between 1,180 different genomic features and noise.

Note: correlated histone modifications are located in either in gene body (gb) or +1 nucleosome (+1 nuc) (Murray, unpublished, 2016).

The histone modifications that most strongly correlate with protein noise are the three types of methylation marks: mono-, di-, and tri-methylation at lysine 4 on histone 3 (H3K4), often referred to as H3K4me1, H3K4me2, and H3K4me3, respectively. H3K4me1 and H3K4me2 at the +1 nucleosome show a positive correlation with noise while H3K4me3 shows a negative correlation with noise. Furthermore, tri-methylation of lysine79, lysine36 and ubiquitylation of lysine K123 of histone 3 in the gene body (H3K79me3, H3K36me3 and H2BK123ub, respectively), show a positive correlation with noise. Other features such as histone turnover, strong TATA boxes, or GC content also correlate with noise as shown in Fig. 7 (S. Murray, unpublished). However, no causal relationship between H3K4 methylation and noise has been reported. Thus, the aim of this work is to examine this relationship.

1.5. Histone modifications

In eukaryotes, a nucleosome, the basic unit of chromatin is comprised of 147 base-pairs of DNA wrapped 1.65 times around a core of histone proteins consisting of two histone, H3 and H4 homodimers and two histone H2A-H2B heterodimers (Luger *et al.,* 1997). In the nucleosome, the negatively charged phosphate groups in the DNA's backbone are neutralised due to the strong binding of DNA with basic histone proteins. Higher order chromatin structures can result from different compactions of chromatin or gene loops (reviewed in Jansen and Verstrepen 2011). The two forms of the higher order chromatin structures are euchromatin, a lightly compacted chromatin, and heterochromatin, a highly compact chromatin. The accessibility of the DNA-processing factors to the DNA are regulated as a consequence of this packing. Indeed, transcription, DNA replication, DNA repair and other DNA-based processes can be influenced due to the higher order chromatin. The underlying DNA sequence can determine the position of nucleosomes along the DNA, and also a number of *trans*-acting factors such as chromatin remodellers, RNA polymerase and transcription factors can play a role in this arrangement (reviewed in Jansen and Verstrepen 2011). In yeast, heterochromatin loci, including telomeres, silent mating-type (HM), and ribosomal DNA (rDNA) (Grunstein, 1998; Moazed, 2001), have deacetylated histones in their associated chromatin (Braunstein *et al.,* 1996; Suka *et al.,* 2001).

The N- and C-terminal flexible tails of histone proteins play an important role in all chromatin-mediated processes and the assembly of nucleosomes into higher order chromatin forms (Luger and Richmond 1998). The flexible tails of the histone proteins are accessible for extensive post-translational modifications because these tails protrude from channels formed from the aligned DNA minor grooves (Luger et al., 1997a).

Post-translational modifications – including acetylation, methylation, phosphorylation, ubiquitination and sumoylation– which can be added to histones, can play important roles in regulating gene expression by the control of DNA accessibility. These modifications can have a direct impact on chromatin structure by affecting histone-DNA interactions or an indirect effect by interacting with effector proteins (Strahl and Allis 2000; Turner 2000). The histone code hypothesis proposed that the chemical modifications that can be added to histone proteins have unique functions which can regulate the accessibility to DNA and control formation of an active or silent chromatin. Nevertheless, depending on the genomic context, the same histone modification can have various functions which suggests that histone modifications are a more complex chromatin language than proposed by the histone code hypothesis (Berger 2007; Howe *et al.,* 2014; Howe *et al.,* 2016).

1.6. The role of H3 methylation and acetylation in gene expression

Methyltransferases (KMTs) are enzymes that are responsible for the addition of methyl groups to histone lysine residues. Except for Dot1, which is responsible for methylation at lysine 79 in H3 (Feng et al., 2002), each KMTs contains a catalytic SET domain and can only add methyl groups (S-adenosyl-L- methionine is used as the methyl donor by KMTs) with the catalytic SET domain. Furthermore, only one type of modification can be added to a lysine residue. For example, K methylation can be prevented if the residues are acetylated, and vice versa. The formation of acetyl-lysine is catalysed by histone acetyltransferases (HATs), and removal of the acetyl group is catalysed by histone deacetylases (HDACs).

1.6.1. H3K4 methylation

Methylation of histones is one of the most common histone modifications. All eukaryotes, including, yeast, *C. elegans, A. thaliana, D. melanogaster,* and mammals, have H3K4me3 at their 5' end region of their active genes (Santos-Rosa *et al.,* 2002; Howe *et al.,* 2016). All those species have lysine methyltransferases (KMTs) that catalyse the transfer of a methyl group to lysine residues on histones. In *S. cerevisiae*, the Set1 complex (Set1C) – a methyltransferase complex (Fig. 8A) – is the only enzyme that is responsible for the addition of all types of methylation on H3K4 (Briggs et al., 2001; Roguev *et al.,* 2001; Nagy *et al.,* 2002; Krogan *et al.,* 2002; Santos-Rosa et al. 2002). In mammals, there are six enzymes that catalyse this modification, therefore yeast seems to be a simple and appropriate model organism to study the patterns of H3K4 methylation (Howe *et al.,* 2016). For full methylation activity of the three types of methylation marks (H3K4me1, H3K4me2, and H3K4me3), Set1 requires a number of other proteins that are associated with the Set1 complex (Miller *et al*., 2001; Fig. 8). It is worth noting that Set1 also methylates non-histone proteins such as Dam1, a kinetochore protein (Zhang, K *et al.,* 2005). Furthermore, Set1 has been shown to promote telomere silencing and maintain telomere structure (Nislow et al., 1997; Corda *et al.,* 1999; Leung *et al.,* 2011) and the loss of *SET1* can create defect telomeric silencing (Margaritis *et al.,* 2012; Ng *et al.,* 2003).

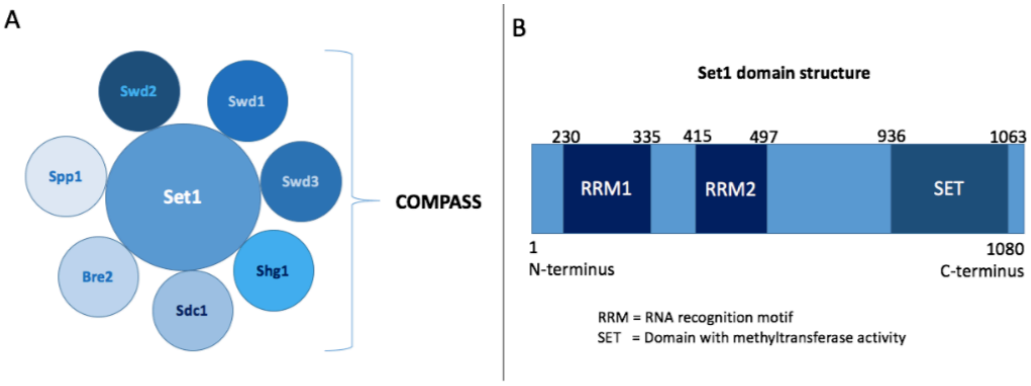


Figure 8. Set1 interactions within COMPASS and its domain structure (P. Lorenz, unpublished).

RRM1 and RRM2 are two predicted RNA-recognition motifs in Set1. Deposition of H3K4me3 requires RRM1 H3K4me2 does not (Schlichter and Cairns, 2005). On deletion of the RRM1 domain of Set1, H3K4me3 levels are significantly decreased. Mutation of the SET catalytic domain (G990E) of Set1 produces a hyperactive version of Set1, which is able to fully restore the level of H3K4me3 observed in the deletion of the RRM1 (Schlichter and Cairns, 2005). A schematic of the Set1 domains is shown in Fig. 8B. Moreover, ubiquitination of H2B lysine 123 is also required for Set1C catalytic activity. Rad6 and Bre1, ubiquitin-conjugating enzymes, catalyse H2BK123Ub. Ubiquitination of H2B lysine 123 promotes Set1C recruitment to chromatin by controlling Swd2 (Lee *et al*., 2011). Furthermore, Jhd2, a monomeric enzyme, is a H3K4 demethylase in yeast. It demethylates all three types of methylation marks at lysine 4 on H3 (Ingvarsdottir *et al*., 2007; Liang *et al.,* 2007; Huang et al., 2010). H3K4 methylation can be also inhibited by Jhd2 by reducing the amount of Set1C on the chromatin (Ingvarsdottir *et al*. 2007). Levels of H3K4me2 have been shown to be globally increased on the loss of Jhd2 (Liang *et al*., 2007).

Set1 directly interacts with Spp1 – a subunit of Set1C (Fig. 8A). Spp1 can influence the catalytic activity of Set1 (Takahashi *et al*., 2011), and is also primarily required for the addition of tri-methylation on H3K4 (Morillon *et al*., 2005). *SPP1* deletion decreases H3K4me3 globally (Morillon *et al*., 2005), although its effect on individual genes varies. For example, ribosomal protein genes maintain H3K4me3 independently of Spp1 (Howe *et al*., 2014). However, *SPP1* deletion decreases H3K4me3 globally, but around 20% of H3K4me3 levels remain, and this is found primarily at the promoter of ribosomal protein gene (Sommermeyer *et al.,* 2013; Howe *et al.,* 2014). At other promoter regions, loss of Spp1 is also associated with a build-up of H3K4me2. A recent study found that Spp1 not only alters H3K4me3, but also affects the overall activity of Set1C. In *SPP1* deletion, some reduction in H3K4me2 levels was also observed and there is an upstream shift of the H3K4 methylation gradient (H3K4me3 located in proximal promoter regions, followed by H3K4me2 and then H3K4me1 located in distal promoter sites) (Soares *et al.,* 2017). Shi et al., (2007) suggests that the Set1 complex might be stabilised by Spp1 through its binding to H3K4me2 via its PHD domain pushing the equilibrium further towards H3K4me3. Furthermore, Howe et al., (2014) found that levels of H3K4me3 are reduced in H3K14A mutants, whereas H3K14R, which was used as a control,

does not affect H3K4me3 levels. H3K4me3 levels were reduced to 40% of the WT levels in H3K14Q mutants and levels were reduced to 20% of the WT levels in H3K14A mutants. H3K14A mutants do not affect H3K4me1 or H3K4me2 levels, thus these could be used to test the effect of gradually change in H3K4 tri-methylation levels on gene expression noise. These strains would provide insight on whether H3K4me3 acts gradually on gene expression noise levels or whether there is a threshold effect.

Active promoters have been mainly associated with H3K4me3 while both active and inactive genes show an association with H3K4me1 and H3K4me2. In yeast, many studies on H3K4 methylation have shown a positive correlation of H3K4me3 levels with transcription levels according to their ChIP-seq, NET-seq and PRO-seq analysis (Newman *et al.,* 2006; Howe et al., 2014; Howe et al., 2016). On the other hand, during the Yeast Metabolic Cycle (YMC), Howe et al., (2016) shows that transcription levels are not significantly affected upon the loss of most of H3K4me3 mark. Fig. 9 reveals that the H3K4me3 peaks as a result of transcription because its peak appears at the promoter after mRNA level peaks (Howe et al., 2016). Therefore, other gene regulatory mechanisms including splicing mechanism, transcription termination, transcription elongation and transcription consistency are believed to be affected by H3K4me3 (Howe *et al.,* 2016).

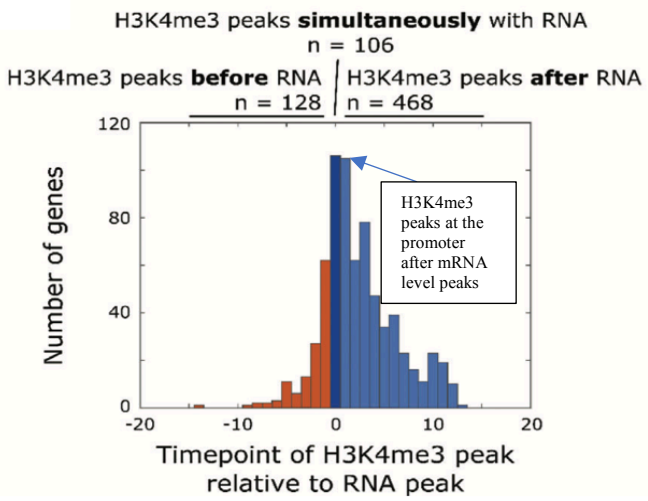


Figure 9. H3K4me3 peaks before (in orange), simultaneously (dark blue) or after their transcripts peaks (light blue) in a number of genes.

Adapted from Howe *et al.,* 2016.

The effects of H3K4 methylation and its role in gene expression is not fully understood. In genome-wide studies, for example, levels of expression are increased for individual *S. cerevisiae* genes by the loss of all type of methylation at lysine 4 on H3 on deletion of *SET1* (Briggs *et al*., 2001; Carvin and Kladde, 2004; Lenstra et al., 2011). Moreover, H3K4me2 has been shown to suppress nucleosome acetylation, which is considered to activate transcription by increasing the accessibility of DNA (Anderson et al., 2001; Kurdistani et al., 2004; Liu et al., 2005; Pokholov et al., 2005; Guillemette et al., 2011), on recruiting the Set3 histone deacetylase (Kim and Buratowski, 2009).

1.6.2. H3K79 methylation

Dot1 is the only known enzyme that is responsible for histone H3 lysine 79 (H3K79) methylation (Ng et al. 2002; van Leeuwen et al. 2002). Dot1 and hDOT1L, the human homologue, have similar protein structures and catalytic activity (Feng et al. 2002). It has been found that almost 90% of H3K79 is methylated *in vivo* (Van Leeuwen et al. 2002), and only a small correlation between gene activity and H3K79me3 have been observed (Pokholok et al. 2005). Sir2 and Sir3, silencing proteins, are restricted to telomeric regions by H3K79me. It has been shown that loss of H3K79me results in Sir proteins being delocalised over the genome thereby adjusting silencing at the telomeres (Van Leeuwen et al. 2002; Ng et al. 2002).

Unlike the other KMTs, Dot1 does not contain the catalytic SET domain (Feng et al. 2002). The arginine methyltransferases catalytic domain is similar to the yeast and human Dot1 (Min et al. 2003; Sawada et al. 2004). Additionally, Dot1 is also different to other KMTs because it is unable to control the generation of specific H3K79 methylation states due to the non-processivity of Dot1 resulting in overlapping gene distributions of the three H3K79 methylation states (Frederiks et al. 2008; Shahbazian et al. 2005). Furthermore, since the location of H3K79 in the nucleosome body is on an uncovered surface (unlike for other H3 modifications which are on the histone tails), H3K79 does not contact any other histones or DNA (Luger et al. 1997a).

1.6.3. H3 acetylation and H3K9ac

High transcription levels have been generally associated with histone acetylation levels. Levels of H3 acetylation including H3K9, H3K14, H3K18, H4K5 and H4K12 have been shown to be positively correlated with active transcription (Liu et al. 2005; Pokholok et al. 2005; Kurdistani et al. 2004; Pokholok et al. 2005; Guillemette et al. 2011), except for H4K8ac and H4K16ac which are shown to be negatively correlated with transcription (Wang et al. 2002). Moreover, independent impacts on transcription can also result from histone acetylation, for example, the prevention of increasing heterochromatin domains by H4K16ac (reviewed in Millar and Grunstein 2006), and the nucleosome assembly coupled with replication resulting from H3K56ac (Li et al. 2008; Burgess et al. 2010). A direct impact of histone acetylation on chromatin compaction might result from the neutralisation of the lysine residue by acetylation making the interaction of the histone tail-DNA and internucleosomal interactions weak. In fact, DNA nucleosome accessibility is decreased in the absent of acetylation (Anderson et al., 2001; Widlund et al., 2000). This would allow more factors related to DNA-processing to bind to DNA such as Gal4 (Vettese-Dadey et al. 1996) or nucleosome remodellers.

Acetylation on histone H3 at lysine 9 has been shown to be associated with active promoters, however, its function is not fully understood. As for other acetylation marks, H3K9ac marks may influence chromatin structure (Kouzarides, 2017) and promote progression of transcription into elongation (Gates et al., 2017). For example, in HeLa cells, two HAT complexes, SAGA (Spt-Ada-Gcn5 acetyltransferase) and ATAC (Ada2a-containing) catalyse acetylation through their subunits GCN5 and PCAF (Brownell et al., 1997; Spedale et al., 2012). It has been found that paused Pol II occupancy is increased in the loss of H3K9ac by knocking-out GCN5 (HATs generally control nonderepressible) and PCAF (p300/CBP-associated factor) (Jin et al., 2010; Foulds et al., 2013) for a number of genes, and that maximal Pol II pause release through the super elongation complex (SEC) has been shown to require H3K9ac (Gates et al., 2017; Jin et al., 2010; Foulds et al., 2013).

1.7. Crosstalk between histone modifications

The impact of one residue or histone modification on the modification of residue that is nearby or distant can define the crosstalk among histone modifications. There are a number of possible mechanisms of the crosstalk such as: (a) more than one type of modification can target the same residue such as lysines that can be acetylated, methylated or ubiquitinated; (b) one modification can depend on a second modification. For instance, H3K4 and H3K79 di- and trimethylation require the requirement of H2B monoubiquitination (Bannister and Kouzarides 2011). Set1 and Dot1 both depend on ubiquitination of H2B on lysine 123 (Dover *et al.,* 2002; Wood *et al.,* 2005; Briggs *et al.,* 2002). H2Bub has been shown to direct di- and tri-methylation of H3K4 and H3K79 through the recruitment of relevant enzymes, Set1 and Dot1, respectively (Fig. 10), facilitating histone crosstalk in both yeast and mammals. Mono-methylation of H3K4 and H3K79 was found to be H2Bub independent (Vethantham et al., 2012). Furthermore, AF9, a subunit of YEATS domain, is strongly preferred to bind to H3K9ac (Le Masson *et al.,* 2003; Gates *et al.,* 2017; Li *et al.*, 2014). On active chromatin, this interaction is important for the activation of Dot1 to deposit H3K79 methylation (Li *et al.*, 2014). Furthermore, both H3K4 methylation and H3K9ac are also associated with active transcription (Vermeulen *et al.,* 2010; Bian *et al.,* 2011). For certain genes, the loss of H3K4me3 by *SPP1* deletion leads to the loss of acetylation at lysine 9 on histone 3 (H3K9ac). H3K4me3 can promote H3K9ac through binding to Sgf29 (a SAGA subunit) and this interaction is required for maximal acetylation of H3K9ac. It has also been found that *SGF29* deletion leads to a larger loss of H3K9ac than in *SPP1* deletion (Bian *et al.,* 2011; Vermeulen *et al.,* 2010). Therefore, H3K79 and H3K4 methylation may indirectly stimulate each other.



Figure 10. Set1 and Dot1 are relevant enzymes that recruits H3K4 and H3K79 methylation respectively. Adapted from Howe, unpublished, 2016.

1.8. Aims: testing causal links between (epi)genetic features and noise

The work of this thesis aims to provide some insights into the role of H3K4, H3K79 methylation and H3K9ac in gene expression and gene expression noise. The causal relationship of those modifications will be investigated using GFP-tagged genes with FACS, and a mechanism of their effects will be proposed. Additionally, the crosstalk between H3K4me3 and H3K79me will be determined using gene-specific ChIP experiments and its involvement in the regulation of noise will be explored.

Chapter 2.

Materials and Methods

2. Materials and Methods

2.1. Cloning and yeast transformation

Homologous recombination-based method was used to create all the strains (Longtine et al. 1998). An appropriate selectable marker encoded by plasmid templates was used to perform PCR using Phusion DNA polymerase. PCR primers containing 40 nucleotides of sequence homologous to the regions flanking the desired site were used for either deleting or tagging a gene. Using a LiAc/heat shock transformation method, the PCR product was transformed into yeast. Appropriate selection plates were used to plate the transformed yeast. Genomic DNA was extracted from transformant colonies and used as the templates for PCR to identify the positive transformants.

2.2. Cell Cultures and GAL10 induction

Pre-cultures were inoculated in YPD (1% bactopeptone, 1% yeast extract, 2 % glucose) the strains were grown overnight at 30°C with shaking at 180 rpm then used to inoculate cultures at a starting O.D. of 0.2. Cultures were grown at 30°C until they reached exponential growth phase (OD ≈ 0.6) after which these were re-diluted to obtain a starting OD of 0.2. Samples of 2 ml were collected for FACS-sample preparation. For *GAL10*, after the cells reached the exponential growth phase in YPD, 2 ml samples were collected for time point zero. To start *GAL10* induction, the cells were washed with water then transferred to YPG (1% bactopeptone, 1% yeast extract, 2 % galactose) and samples were collected after 4, 6, 8 and 24 hours.

2.3. Preparing cells for FACS

Cells at the mid-log phase were harvested by centrifugation, resuspended in 4% paraformaldehyde/PBS then incubated for 45-min at room temperature in the dark with shaking at 80 rpm. Cells were again harvested by centrifugation and resuspended in 400 µl PBS then stored at 4°C. For FACS measurement, 50 µl of a paraformaldehyde fixed sample was diluted in 1 ml PBS in a 5ml polystyrene FACS tube and stored in the absence of light before vortexing and sampling on the BD FACSCalibur instrument.

2.4. Measuring noise

Genes of interest were tagged with GFP and transformed with mutations of interest. Five genes were randomly chosen and analysed after tagging them in three strains: WT, *set1*Δ and *spp1*Δ. Then, their expression levels were analysed in single cells using Fluorescence-activated cell sorting (FACS) which measured the fluorescence intensity of GFP tagged strains in single cells across a population of 10,000 cells. The collected data was expressed as mean level (µ), standard deviation from mean (σ) and coefficient of Variance (σ/µ) (CoV) which is used as a measure of gene expression noise. High CoV represents "noisy" genes, and low CoV represents "quiet" genes.

2.5. Statistical analysis

Because the collected data of the fluorescence intensity (mean) vs. cell counts using FACS is normally distributed (Fig.2), a t-test was used to analyse and compare two populations' means through calculating P values. Then, significant variances between the two normal distributions were determined. Because the mean, standard deviation and coefficient variance of the untagged wild-type (BY4741) and all mutant strains (*set1*Δ*, spp1*Δ, *dot11*Δ, *spp1*Δ- *dot11*Δ and *sgf291*Δ) showed no statistical differences in their autofluorescence, the average autofluorescence of untagged WT strain (BY4741) was compared to all wild-type GFP-tagged strains' fluorescence levels to determine if the

tagging strategy was successful. Moreover, the average dataset of the tagged genes in WT strain (BY4741) was compared to all GFP-tagged strains' fluorescence levels.

2.6. Chromatin Immunoprecipitation (ChIP) - qPCR

ChIP experiments were performed in this study using the method described in Morillon et al. (2005). Cells were grown to OD_{600} 0.5 in 50 ml of appropriate media. Following which cells were harvested by centrifugation then fixed with 45mls PBS containing 1 % formaldehyde for 30 min at 22°C. Fixation was halted by the addition of 125mM glycine and incubated for 5 min. Cells were washed twice with 10ml cold PBS and fixed cell pellets collected by centrifugation at 3000 rpm for 5 min. Cell pellets were resuspended in 500µl cold FA-150 buffer (10mM HEPES pH 7.9, 150mM NaCl,

0.1 % SDS, 0.1% sodium deoxycholate, 1% Triton X-100) along with 1ml of glass beads and the cells disrupted using a MagnaLyser instrument. The sample volume was increased to 2mls by the addition of FA-150. Shearing of the fixed chromatin was carried out by sonication in a Biorupter (Diagenode, 30 min, 1 min on, 20s off, medium setting. The sonicated chromatin sample was cleared by centrifugation (10,000 rpm, 15 min, 4°C). Chromatin was incubated with antibody in 1.5 ml siliconised Eppendorf tubes on a rotating wheel for 15-20 hours at 4°C. 50µl of protein A-Sepharose (pre-blocked with sonicated salmon sperm DNA) was added to the tubes and immunoprecipitation of bound chromatin was performed for 90 min at 22°C on a rotating wheel. The chromatin bound sepaharose beads were collected by centrifugation (2600 rpm, 1 min) and washed with TSE-150 buffer (20 mM Tris-Cl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1 % SDS, 1 % Triton X-100) for 3 min, TSE-500 buffer (20 mM Tris-Cl pH 8.0, 500 mM NaCl, 2 mM EDTA, 0.1 % SDS, 1 % Triton X-100) for 3 min, LiCl buffer (0.25 M LiCl, 10 mM 52 Tris-Cl pH 8.0, 1 mM EDTA, 1 % dioxycholate, 1 % NP-40) for 15 min and twice with TE, all at 22°C. Chromatin bound beads were incubated in elution buffer for 30 mins at 65°C following which the chromatin was eluted from the beads with elution buffer (0.1 M NaHCO3, 1 % SDS) at 65°C. Reversal of cross links was carried out by the addition of 350mM NaCl to the samples which were then incubated for 3 hours at 65°C. Samples were treated with RNase A for 1 hour at 37°C followed by Proteinase K overnight at 65°C. DNA was purified using a PCR-purification kit (Qiagen) and eluted in 400µl 1 mM Tris-Cl pH 8.0.

Real-time quantitative PCR was carried out using a Corbett Rotorgene and Sybr green mix (Bioline). The collected data ([IP - no antibody control]/input) were expressed as a percentage of the input and normalised to levels of H3 where appropriate. The standard error of three real-time PCR reactions represents the error bars. The antibodies and primers used are listed in Table 1 and Table 2, respectively.

Table 1. Sequences for primers used in real-time PCR.

Primer name	**Sequence**
*URA4*_F_Promoter	ATCGGCATCGTCAGAGAAGC
*URA4*_R_Promoter	AGCTGCCCAACAACATGCCA
*SML1*_Promoter_F	TTTTTTCCTGGCTGTCCCTT
*SML1*_Promoter_R	ATTGAGCAGTTGAAGAGAGG
*GCV3*_Promoter_F	CGGCGGAGTCACTTTCGCTA
*GCV3*_Promoter_R	TATGCTTCATGTGGAGATTC
*BAT2*_Promoter_F	GGGAAAAAATTCATTCCTTT
*BAT2*_Promoter_R	AACTCGTGGAGATGCTTTCC

Table 2. Details of the antibodies used during this study.

Antibody	**Catalogue number**	**Species reactivity**	**Amount for ChIP (µl)**
H3	Mi 07-690	Rabbit	5
H3K4me3	Mi 05-745R	Rabbit	5
H3K9ac	Mi 07-352	Rabbit	5
H3K79me2	ab3594	Rabbit	5
H3K79me3	ab2621	Rabbit	5

2.7. Yeast strains

In this study, BY4741-derived strains were used to perform all experiments and the strains used are detailed in Table 3.

Table 3. Ratios of H3K4me2 over H3K4me3 in the +1 nucleosomes of the 5 target genes.

Strain	Parent	Genotype	Origin
BY4741 (WT)	S288C	*MAT* α *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0*	Euroscarf
***set1*Δ**	BY4741	*MAT* α *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SET1::KanMX6*	Courtesy of A. Morillon
***spp1*Δ**	BY4741	*MAT* α *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::His3MX6*	Courtesy of F. Howe
BY *QCR8-GFP His*	BY4741	*MAT* α *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 QCR8-GFP His3MX6*	This study
BY *set1*Δ *kanmx QCR8-GFP His*	BY4741 *Δset1 kanmx*	*MAT* α *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SET1::KanMX6 QCR8-GFP His3MX6*	This study
BY *spp1*Δ *Leu QCR8-GFP His*	BY4741 *Δspp1 Leu*	*MAT* α *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu QCR8-GFP His3MX6*	This study
BY *GPP1-GFP His*	BY4741	*MAT* α *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 GPP1-GFP His3MX6*	This study
BY *set1*Δ *kanmx GPP1-GFP His*	BY4741 *Δset1 kanmx*	*MAT* α *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SET1::KanMX6 GPP1-GFP His3MX6*	This study
BY *spp1*Δ *Leu GPP1-GFP His*	BY4741 *Δspp1 Leu*	*MAT* α *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu GPP1-GFP His3MX6*	This study
BY *STM1-GFP His*	BY4741	*MAT* α *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 STM1-GFP His3MX6*	This study
BY *set1*Δ *kanmx STM1-GFP His*	BY4741 *Δset1 kanmx*	*MAT* α *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SET1::KanMX6 STM1-GFP His3MX6*	This study
BY *spp1*Δ *Leu STM1-GFP His*	BY4741 *Δspp1 Leu*	*MAT* α *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu STM1-GFP His3MX6*	This study
BY *GAL10-GFP His*	BY4741	*MAT* α *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 GAL10-GFP His3MX6*	Courtesy of A. Nair

BY *set1Δ kanmx GAL10-GFP His*	BY4741 *Δset1 kanmx*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SET1::KanMX6 GAL10-GFP His3MX6*	This study
BY *spp1Δ Leu GAL10-GFP His*	BY4741 *Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu GAL10-GFP His3MX6*	This study
BY *TEF1-GFP His*	BY4741	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 TEF1-GFP His3MX6*	This study
BY *set1Δ kanmx TEF1-GFP His*	BY4741 *Δset1 kanmx*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SET1::KanMX6 TEF1-GFP His3MX6*	This study
BY *spp1Δ Leu TEF1-GFP His*	BY4741 *Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu TEF1-GFP His3MX6*	This study
BY *dot11Δ kanmx TEF1-GFP His*	BY4741 *Δdot1 kanmx*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 DOT1::KanMX6 TEF1-GFP His3MX6*	This study
BY *dot11Δ kanmx spp1Δ Leu TEF1-GFP His*	BY4741 *Δdot1 kanmx Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 DOT1::KanMX6 SPP1::Leu TEF1-GFP His3MX6*	This study
BY *spp1Δ Leu TEF1-GFP His*	BY4741 *Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu TEF1-GFP His3MX6*	This study
BY *URA4-GFP His*	BY4741	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 URA4-GFP His3MX6*	Courtesy of J. Rayappu
BY *set1Δ kanmx URA4-GFP His*	BY4741 *Δset1 kanmx*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SET1::KanMX6 URA4-GFP His3MX6*	This study
BY *spp1Δ Leu URA4-GFP His*	BY4741 *Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu URA4-GFP His3MX6*	Courtesy of J. Rayappu
BY *dot11Δ kanmx* URA4-*GFP His*	BY4741 *Δdot1 kanmx*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 DOT1::KanMX6 URA4-GFP His3MX6*	This study
BY *dot11Δ kanmx spp1Δ Leu URA4-GFP His*	BY4741 *Δdot1 kanmx Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 DOT1::KanMX6 SPP1::Leu URA4-GFP His3MX6*	This study
BY *spp1Δ Leu URA4-GFP His*	BY4741 *Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu URA4-GFP His3MX6*	This study
BY SML1-*GFP His*	BY4741	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SML1-GFP His3MX6*	This study
BY *set1Δ kanmx SML1-GFP His*	BY4741 *Δset1 kanmx*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SET1::KanMX6 SML1-GFP His3MX6*	This study

BY *spp1Δ Leu SML1-GFP His*	BY4741 *Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu SML1-GFP His3MX6*	This study
BY *dot11Δ kanmx SML1-GFP His*	BY4741 *Δdot1 kanmx*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 DOT1::KanMX6 SML1-GFP His3MX6*	This study
BY *dot11Δ kanmx spp1Δ Leu SML1-GFP His*	BY4741 *Δdot1 kanmx Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 DOT1::KanMX6 SPP1::Leu SML1-GFP His3MX6*	This study
BY *spp1Δ Leu SML1-GFP His*	BY4741 *Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu SML1-GFP His3MX6*	This study
BY *GCV3-GFP His*	BY4741	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 GCV3-GFP His3MX6*	Courtesy of J. Rayappu
BY *set1Δ kanmx GCV3-GFP His*	BY4741 *Δset1 kanmx*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SET1::KanMX6 GCV3-GFP His3MX6*	Courtesy of J. Rayappu
BY *spp1Δ Leu GCV3-GFP His*	BY4741 *Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu GCV3-GFP His3MX6*	Courtesy of J. Rayappu
BY *dot11Δ kanmx GCV3-GFP His*	BY4741 *Δdot1 kanmx*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 DOT1::KanMX6 GCV3-GFP His3MX6*	This study
BY *dot11Δ kanmx spp1Δ Leu GCV3-GFP His*	BY4741 *Δdot1 kanmx Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 DOT1::KanMX6 SPP1::Leu GCV3-GFP His3MX6*	This study
BY *spp1Δ Leu GCV3-GFP His*	BY4741 *Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu GCV3-GFP His3MX6*	This study
BY *ARG1-GFP His*	BY4741	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 ARG1-GFP His3MX6*	This study
BY *set1Δ kanmx ARG1-GFP His*	BY4741 *Δset1 kanmx*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SET1::KanMX6 ARG1-GFP His3MX6*	This study
BY *spp1Δ Leu ARG1-GFP His*	BY4741 *Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu ARG1-GFP His3MX6*	This study
BY *dot11Δ kanmx ARG1-GFP His*	BY4741 *Δdot1 kanmx*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 DOT1::KanMX6 ARG1-GFP His3MX6*	This study
BY *dot11Δ kanmx spp1Δ Leu ARG1-GFP His*	BY4741 *Δdot1 kanmx Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 DOT1::KanMX6 SPP1::Leu ARG1-GFP His3MX6*	This study
BY *spp1Δ Leu ARG1-GFP His*	BY4741 *Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu ARG1-GFP His3MX6*	This study

BY *SWD1-GFP His*	BY4741	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SWD1-GFP His3MX6*	This study
BY *set1Δ kanmx SWD1-GFP His*	BY4741 *Δset1 kanmx*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SET1::KanMX6 SWD1-GFP His3MX6*	This study
BY *spp1Δ Leu SWD1-GFP His*	BY4741 *Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu SWD1-GFP His3MX6*	This study
BY *dot11Δ kanmx SWD1-GFP His*	BY4741 *Δdot1 kanmx*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 DOT1::KanMX6 SWD1-GFP His3MX6*	This study
BY *dot11Δ kanmx spp1Δ Leu SWD1-GFP His*	BY4741 *Δdot1 kanmx Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 DOT1::KanMX6 SPP1::Leu SWD1-GFP His3MX6*	This study
BY *spp1Δ Leu SWD1-GFP His*	BY4741 *Δspp1 Leu*	*MAT α his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 SPP1::Leu SWD1-GFP His3MX6*	This study

2.8. Tested Genes

Description of the GFP-tagged genes that were analysed and discussed in this study (Table 4).

Table 4. The description of each gene tagged with GFP and analysed with FACS.

Standard Name	**Systematic Name**	**Name Description**
QCR8	YJL166W	ubiQuinol-cytochrome C oxidoReductase
SML1	YML058W	Suppressor of Mec1 Lethality
GPP1	YIL053W	Glycerol-3-Phosphate Phosphatase
STM1	YLR150W	Suppressor of ToM1
GAL10	YBR019C	GALactose metabolism
TEF1	YPR080W	Translation Elongation Factor
ARG1	YHR018C	ARGinine requiring
SWD1	YAR003W	Set1c, WD40 repeat protein
URA4	YLR420W	Uracil requiring

Chapter 3.

Investigating the role of H3K4 methylation in noise in *Saccharomyces cerevisiae*

3. Investigating the role of H3K4 methylation in noise

3.1. Introduction

3.1.1. H3K4 methylation strongly correlates with noise

In *S. cerevisiae*, Set1 complex (Set1C) – a methyltransferase complex (Fig. 11) – is responsible for the addition of all types of methylation on H3K4, while Spp1 – a subunit of Set1C (Takahashi *et al*., 2011) – is primarily required for the addition of tri-methylation on H3K4 (Morillon *et al*., 2005). H3K4me3 is removed by Jhd2 demethylase (Ingvarsdottir et al. 2007; Liang et al. 2007; Seward et al. 2007). In active genes, H3K4me3 is generally found at the 5' end. H3K4me3 has a positive correlation with transcripts levels (Santos-Rosa et al. 2002; Bernstein et al. 2002; Liu et al. 2005; Pokholok et al. 2005). *SET1* deletion results in the loss of all methylation at K4 on H3 (Briggs *et al*., 2001). On the other hand, *SPP1* deletion decreases H3K4me3 globally (Morillon *et al*., 2005), but its effect on individual genes varies (Howe *et al*., 2014). It is worth noting however that although *SPP1* deletion decreases H3K4me3 globally, around 20% of H3K4me3 levels remain, and this is found primarily at the promoter of ribosomal protein gene (Sommermeyer *et al.,* 2013; Howe *et al.,* 2014). Interestingly, H3K4 methylation show a strong correlation with noise but the enzymes (Set1 and Spp1) show a weaker correlation with noise (Table 5). This suggests that the modifications themselves, rather than the proteins, may affect noise.

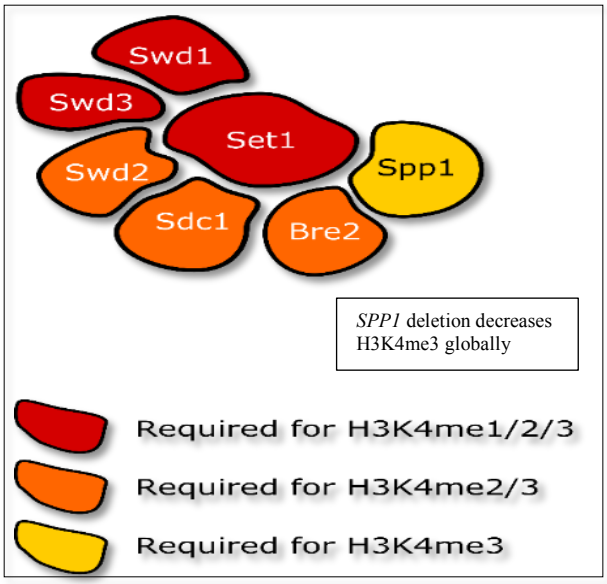


Figure 11. The seven sub-units of *S. cerevisiae's* Set1 complex.
Adapted from Howe *et al.,* 2016.

Table 5. Enzyme levels correlation with noise.

levels at promoter	Correlation with noise (spearman correlation coefficient)
Spp1	-0.197
Set1	-0.027
H3K4me2	0.380
H3K4me3	-0.318

Previous studies have shown that gene expression noise could be regulated by histone modifications. Weinberger et al. (2012) found that Rpd3(L)C, a histone deacetylase complex, could indirectly regulate gene expression noise. Another study, Furthermore, protein levels have been shown to be inversely linked with noise (Elowitz et al., 2002; Bar-Even et al., 2006; Newman et al., 2006; Hurnung et al., 2012). Newman et al. (2006) analysed noise levels in *S. cerevisiae* for 2,204 genes using a library of GFP-tagged genes along with FACS, they reported a strong negative correlation between gene expression and noise. However, at lower protein abundance, this relation has not been observed in mammalian cells (Raj et al., 2006). Following Newman's et al. (2006) work, we found at the

genome-wide level some strong correlations in 1,180 gene features, including histone modifications, and noise (see Fig. 7).

The histone modifications that most strongly correlate with protein noise are the three types of methylation marks: mono-, di-, and tri-methylation at lysine 4 on histone 3 (H3K4), often referred to as H3K4me1, H3K4me2, and H3K4me3, respectively. H3K4me1 and H3K4me2 at the +1 nucleosome show a positive correlation with noise while H3K4me3 shows a negative correlation with noise. This could explain the different correlation between noise, Set1 and H3K4me3 (Table 5).

3.1.2. Genes with different noise levels show a reduction in H3K4me3 upon the loss of SPP1

In this work, 18 genes are discussed. Nine genes were analysed as part of this study and 9 genes by other laboratory members (P. Lorenz and J. Rayappu, unpublished; note: *URA4* in *spp1Δ* strain was first analysed by J. Rayappu and several repeats were completed in this study). Tested genes have different noise levels as depicted in Table 6. At all genes, loss of *SPP1* caused a reduction in levels of H3K4me3 in the region occupied by nucleosome +1, located downstream from the Transcription Start Site (TSS) at 0. By contrast, the effects of the *spp1*Δ on levels of H3K4me2 in the same region were variable. At 10 genes, levels of H3K4me2 were slightly increased at the promoter region of the tested genes, and at 8 genes, levels remained similar to levels on the wild-type (WT) strain (Table 6).

Table 6. Genome-wide ChIP data generated from the lab shows values of H3K4me2 and H3K4me3 at the promoter regions (0:300) of the target genes analysed in this study and by past laboratory members in *spp1*Δ (S. Murray, unpublished).

Gene Name	WT H3K4me3 0:300*	spp1 H3K4me3 0:300	spp1/BY ratio H3K4me3 0:300	WT H3K4me2 0:300	spp1 H3K4me2 0:300	spp1/BY ratio H3K4me2 0:300	Noise level (Newman et al., 2006; Wu et al., 2017)
GAL10	0.65	0.23	0.35	1.37	1.36	0.99	-
QCR8	1.41	0.52	0.37	1.38	1.77	1.28↑	High
SML1	0.67	0.26	0.39	1.42	1.42	1.00	-
GPP1	1.52	0.52	0.34	1.52	1.64	1.08↑	High
STM1	1.41	0.71	0.50	1.20	1.60	1.33↑	High
TEF1	1.61	0.54	0.34	1.48	1.60	1.08↑	High
ARG1	0.74	0.23	0.31	1.37	1.33	0.97	Low
SWD1	1.47	0.37	0.25	1.44	1.30	0.90	-
URA4	1.54	0.55	0.36	1.32	1.50	1.14↑	High
BAT2	1.39	0.36	0.26	1.26	1.37	1.09↑	High
GCV3	1.36	0.45	0.33	1.22	1.19	0.97	High
ATP4	1.40	0.63	0.45	1.26	1.47	1.17↑	High
SED1	1.32	0.45	0.34	1.45	1.68	1.16↑	High
GDH1	1.58	0.44	0.28	1.31	1.29	0.98	High
ARG4	1.38	0.36	0.26	1.28	1.26	0.99	-
HMS2	1.40	0.47	0.34	1.21	1.48	1.22↑	Low
KAP104	1.60	0.53	0.33	1.41	1.49	1.06↑	Low
HOM3	1.51	0.52	0.34	1.38	1.33	0.97	Low

** Note: 0 is the TSS and the region studied is over the first 300nt (0:300).*

Based on the genome-wide correlation, H3K4me1/2 may enhance gene expression noise while H3K4me3 may suppress gene expression noise (Fig. 7). Therefore, it was hypothesised that the loss of H3K4me3, by removing *SPP1*, would increase gene expression noise, whereas the loss of all types of methylation, by removing *SET1*, would decrease gene expression noise.

3.1.3. Aims

The aim of this chapter is to characterise the relationship between H3K4 methylation and noise. In particular, it will be investigated whether the loss of H3K4me3 or/and all types of methylation will have any gene-specific noise effects if so, whether these genes fall into any functionally related classes.

3.2. Results

3.2.1. GAL10 induction in galactose in WT strain

After glucose-based catabolite repression is relieved, *GAL10* and other galactose metabolic genes including *GAL1, GAL7, GAL2* are induced to a significant degree. In response to galactose in yeast, activator Gal4p induces all GAL genes (Douglas and Hawthorne, 1964). However, when glucose is present, Mig1p represses the Gal gene expression (Nehlin et al., 1991). In this study, *GAL10* was tagged with the open-reading frame encoding COOH-fused GFP and a selective marker, and samples were collected and analysed after 0, 4, 6, 8 and 24 hours of induction in galactose. *GAL10* starts to express after 4 hours of the induction in galactose (Fig. 12). After 8 hours induction in galactose, *GAL10* reaches its maximum expression level, as there was no significant difference between expression levels after 8 and 24 hours in galactose.

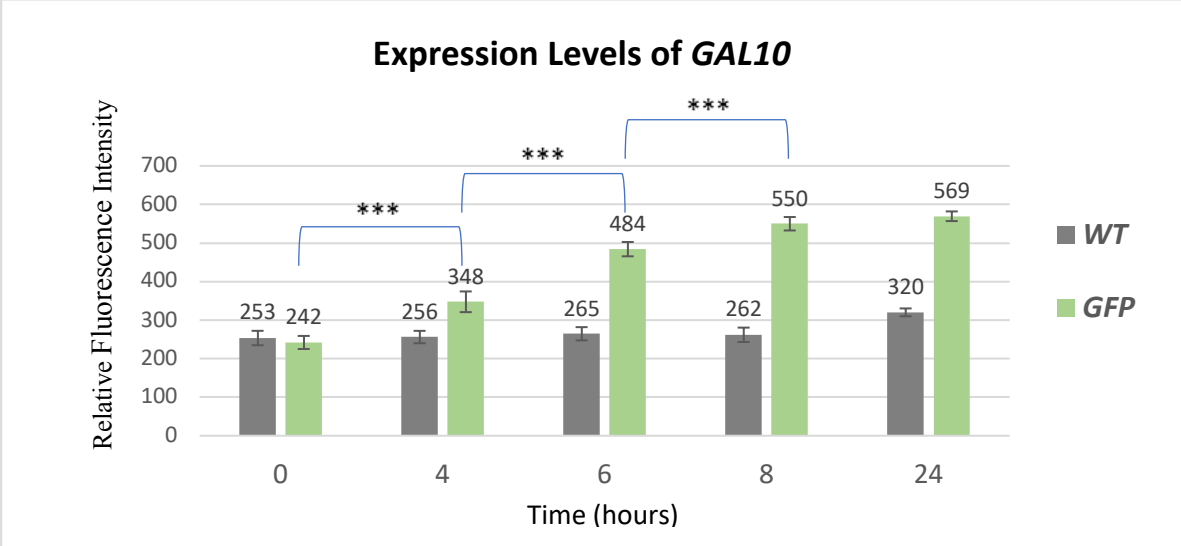


Figure 12. Relative fluorescence intensities of *GAL10* and *GAL10*_GFP in WT strain.

Galactose induction for *GAL10* was performed and samples were collected and analysed after 0, 4, 6, 8 and 24 hours of the induction. ***= $P < 0.001$; *= $P < 0.05$. Error bars= standard deviation between experiments (N=3).

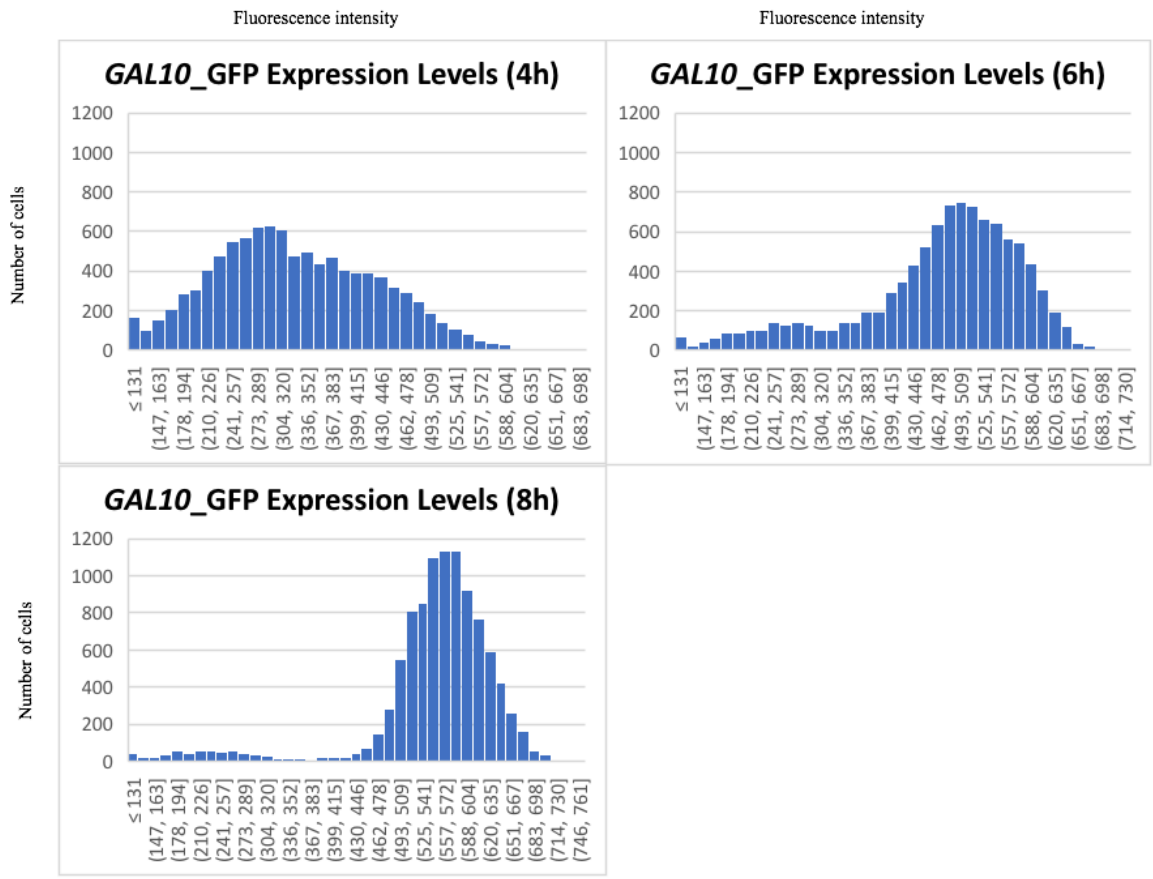


Figure 13. Histograms show relative fluorescence intensities of *GAL10*_GFP in WT strain.

Histograms show fluorescence-intensity distribution for *GAL10*-GFP among 10,000 cells analysed by the FACS machine in WT. Galactose induction for *GAL10* was performed and samples were collected and analysed after 4, 6 and 8 hours of the induction.

After 4 and 6 hours of *GAL10* induction in galactose, *GAL10* is not fully expressed as the histograms in Fig. 13 also show unsymmetrical distribution of fluorescence across a population of 10,000 cells. There was no significant change in number of cells that are expressing fluorescence after 4 and 6 hours of *GAL10* induction in galactose although there was an increase in the levels of fluorescence in the cells. However, after 8 hours of *GAL10* induction, the number of cells that are expressing GFP is increased as is the levels of fluorescence in each cell (Fig. 13). Therefore, the best time to assess *GAL10* expression level for efficient measurement of noise is after 8 hours of induction in galactose.

3.2.2. Tagging genes of interest with GFP in WT and mutant strains

Along with *GAL10* strains, six strains were constructed by tagging another 6 genes (*QCR8, SML1, GPP1, STM1, TEF1, ARG1, SWD1, URA4*) with the open-reading frame encoding COOH-fused GFP and a selective marker. Using FACS, all GFP-tagged strains show a significant increase in fluorescence levels in comparison to the autofluorescence of untagged WT strain (BY4741) (Fig. 14).

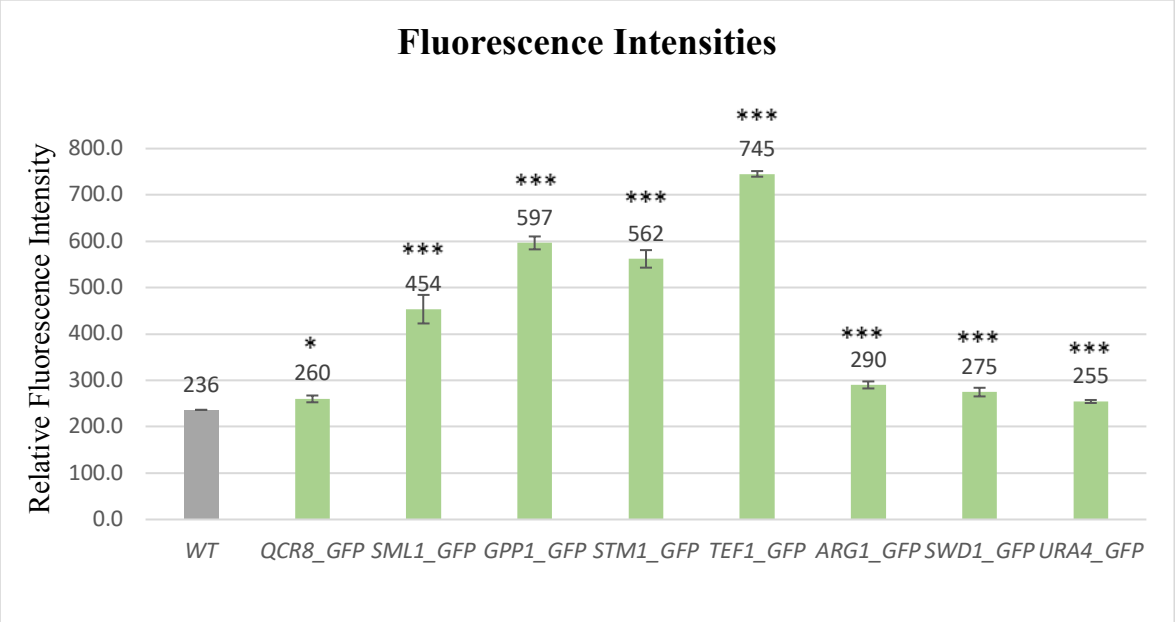


Figure 14. Relative fluorescence intensities of GFP-tagged strains.

***= P < 0.001; *= P < 0.05. Error bars= standard deviation between experiments (N=3).

The above shows that the tagging strategy was successful. Because there is a significant difference between the background fluorescence and the fluorescence of the GFP-tagged genes, there was no concerns of not subtracting the autofluorescence. GFP tags were also introduced into *set1*Δ and *spp1*Δ strains, which remove all methylation or primarily tri-methylation at K4 on H3, respectively. Then, the effect of deletions of *SPP1* and *SET1* on mean gene expression levels (μ) was assessed (Fig. 11).

3.2.3. Deletion of SPP1 decreases gene expression and increases noise at URA4

The hypothesis discussed above states that H3K4me3 supresses noise. As a first approach to test this hypothesis, strains were created that lack H3K4me3 via deletion of *SPP1*. Knowing the effects *spp1*Δ on H3K4me2 and H3K4me3 for the 9 genes individually (Table 6), gene expression noise can then be measured to see if this modification has any effect on noise. *spp1*Δ does not have effects on gene expression at these tested genes with the exception of *URA4* (Fig. 15 & 16). Fig. 15 shows the results of the FACS readings for the *URA4* in wild-type and *spp1*Δ strains. The first striking observation is that deletion of *SPP1* leads to a lower mean expression level.

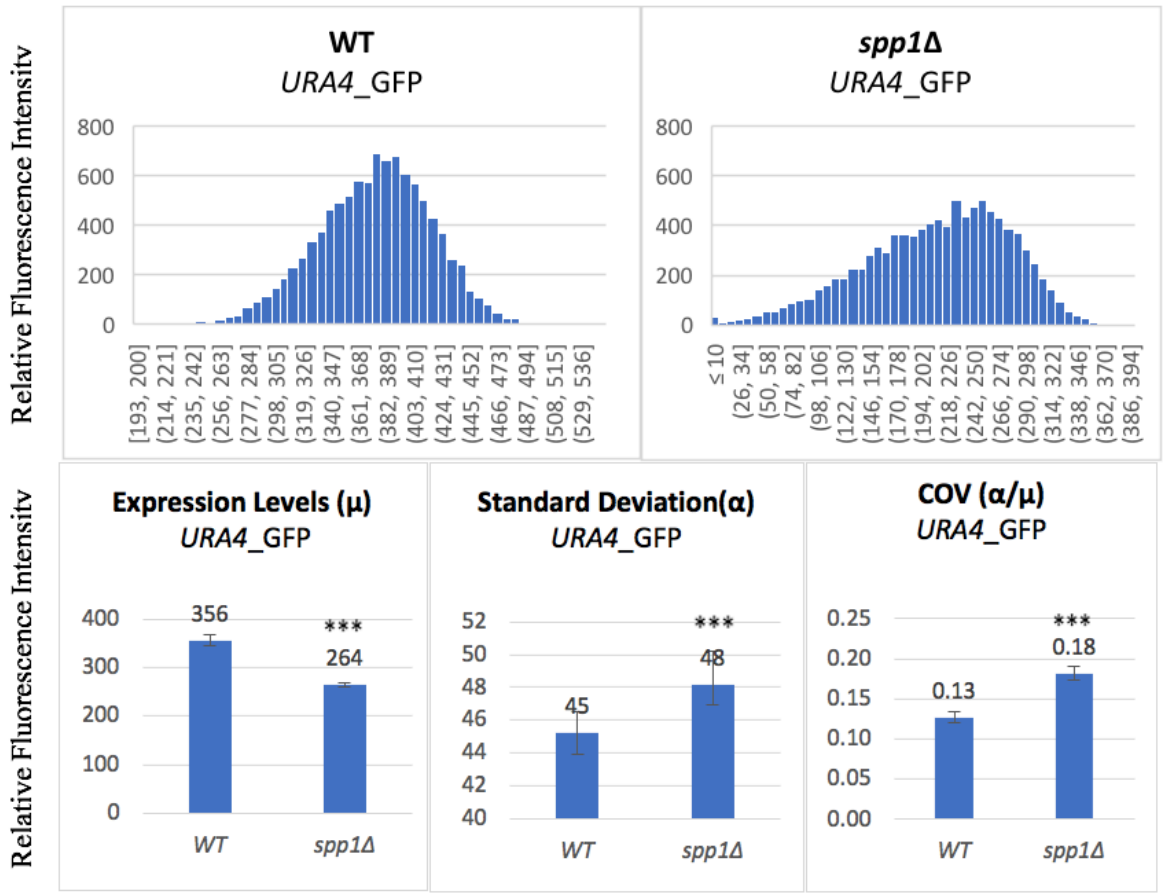


Figure 15. Deletion of *SPP1* decreases expression levels and increases noise.

Histograms show fluorescence-intensity distribution for *URA4*_GFP among 10,000 cells analysed by the FACS machine in WT, *spp1*Δ.

In addition, deletion of *SPP1* leads to an increase in standard deviation or variance for *URA4*. Therefore, the observed increase in the COV (as a measure of noise) may not simply due to just a lower mean expression but in fact due to a higher variation of gene expression across the population of 10,000 cells. The *URA4* locus shows an approximately 28% increase in noise, in the absence of Spp1. These observations strongly suggest that Spp1 has an effect on gene expression noise.

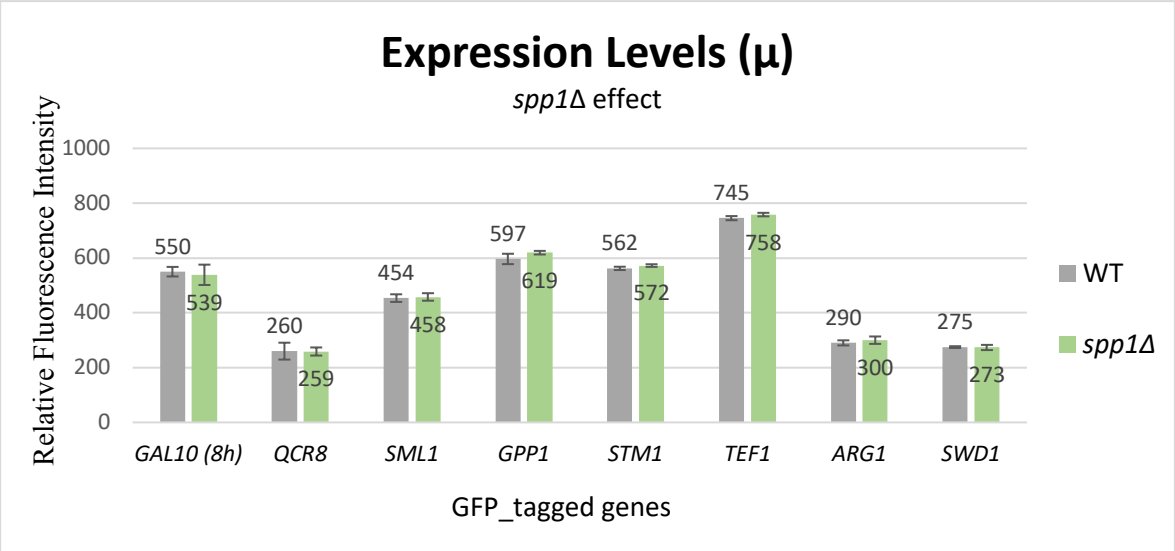


Figure 16. *spp1*Δ have no effects on gene expression levels for the 8 target genes in WT and *spp1*Δ strains.

Of the 9 genes tested, only *URA4* showed an increase in expression levels in the *spp1*Δ strain (Fig. 16). However, as Newman (2006) showed a strong negative correlation between expression levels and noise (i.e. reduced levels of gene expression are associated with increased levels of noise), the change in overall expression levels is the most likely explanation for the observed changes in standard deviation and COV (noise) at *URA4*. This is why the genes where *spp1*Δ does not affect expression levels – but does alter noise – are of most interest in the present study. Next the effect of Spp1 on noise for the remaining GFP-tagged proteins was examined. At these genes, loss of *SPP1* does not change mean expression levels (Fig. 16).

3.2.4. SML1 shows changes in noise but not expression levels

Of the 8 genes, only *SML1* shows a significant change in both noise (COV) and standard deviation in *spp1*Δ (by 3.7% and 4.9%, respectively). This indicates that the change in noise for *SML1* is the result of the loss of specific components of the Set1C rather than of a change in expression levels (Fig. 17).

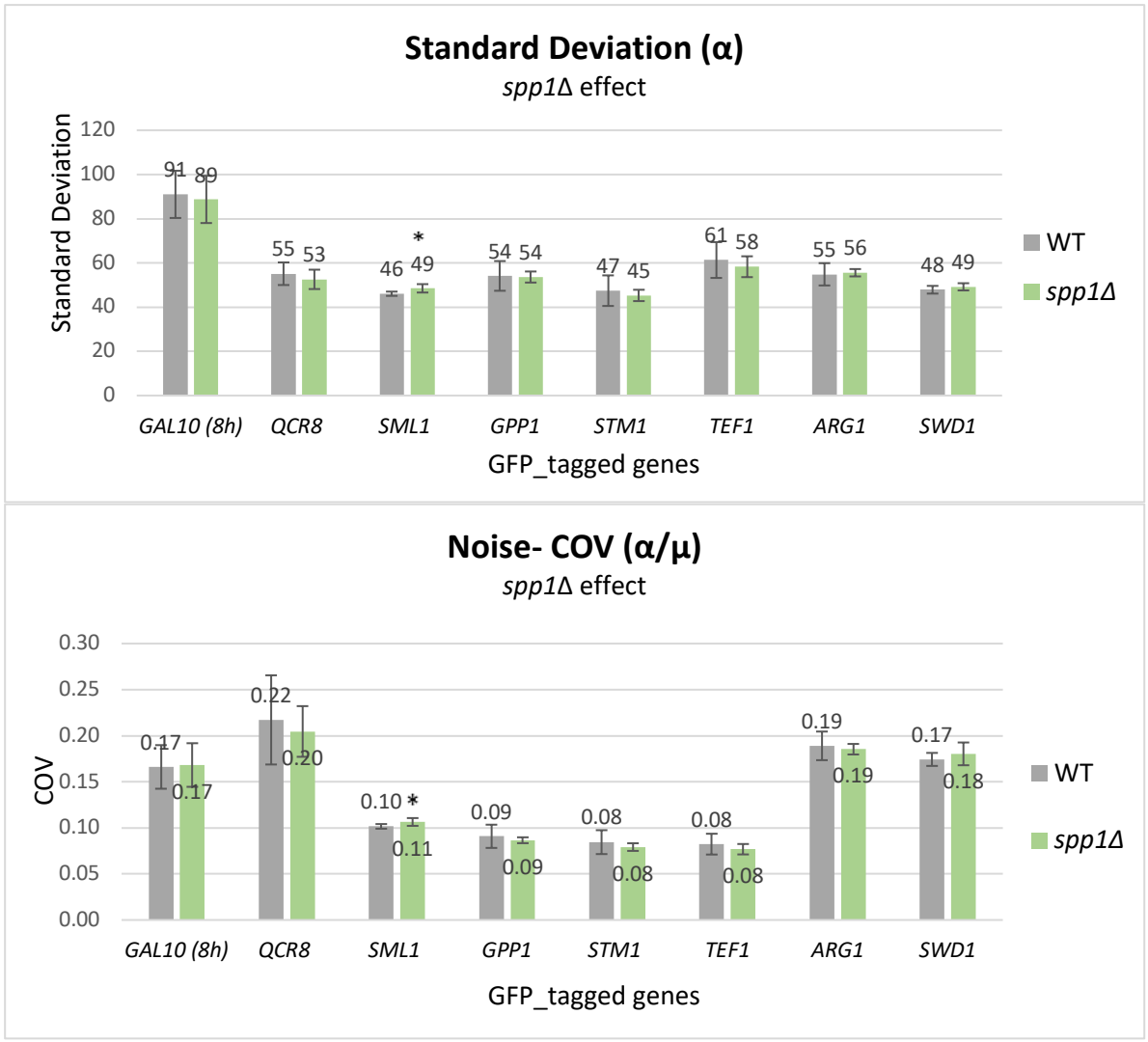


Figure 17. spp1Δ affects* SML1 *noise.

Note: ***, P < 0.001; **, P < 0.01; *, P < 0.05. Error bars= standard deviation between experiments (N=3).

In conclusion, for *SML1*, the data suggests that H3K4me3 supresses noise while H3K4me2 enhances noise. This is in line with the hypothesis put forward in this study. Interestingly, however, this effect was observed at only one gene, out of the 8 genes studied. Next, the effect of loss of *SET1* on noise was tested.

3.2.5. SET1 affects both gene expression levels and noise at GAL10, URA4, ARG1 and SWD1

The hypothesis discussed above states that the loss of all types of methylations (*set1*Δ) would decrease noise. Of the 9 genes tested, only *GAL10, ARG1*, *SWD1* and *URA4* showed an increase in expression levels in the *set1*Δ strain (Fig. 19 & 18). At these four genes, deletion of *SET1* also leads to a decrease in noise. This decrease in noise would be explained by the strong negative correlation between expression levels and noise (Newman et al., 2006). The change in overall expression levels is the most likely explanation for the observed changes in standard deviation and COV (noise) at those genes.

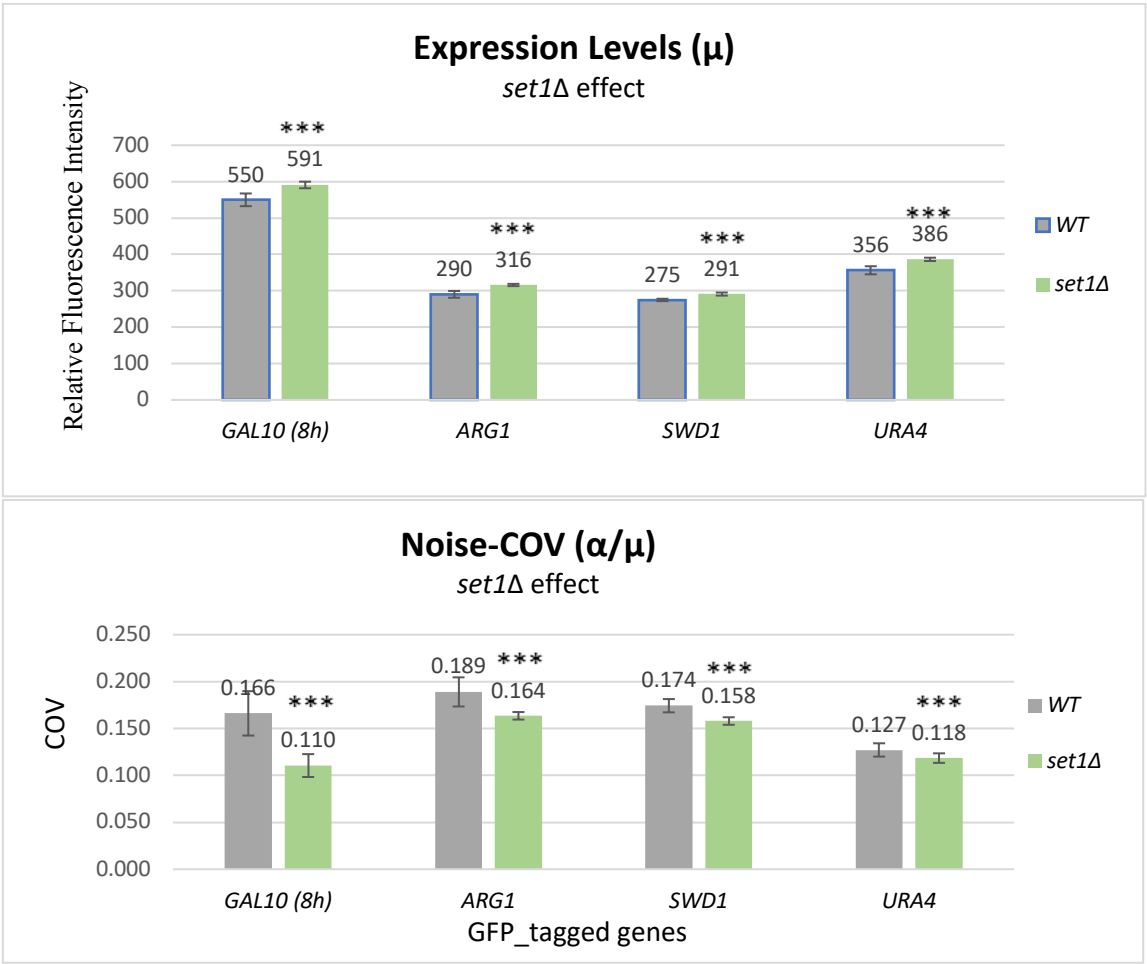


Figure 18. *set1Δ* affects gene expression and noise levels.

***, P < 0.001. Error bars= standard deviation between experiments (N=3).

3.2.6. SML1 shows changes in noise but not expression levels in loss of SET1

As for the remaining genes, no change in gene expression was observed in *set1*Δ (Fig. 19). Interestingly, only *SML1* shows a significant change in noise (COV) and standard deviation in *set1*Δ (by 3.9% and 2.4% respectively). This indicates that the change in noise for *SML1* is the result of the loss of specific components of the Set1C rather than of a change in expression levels (Fig. 19).

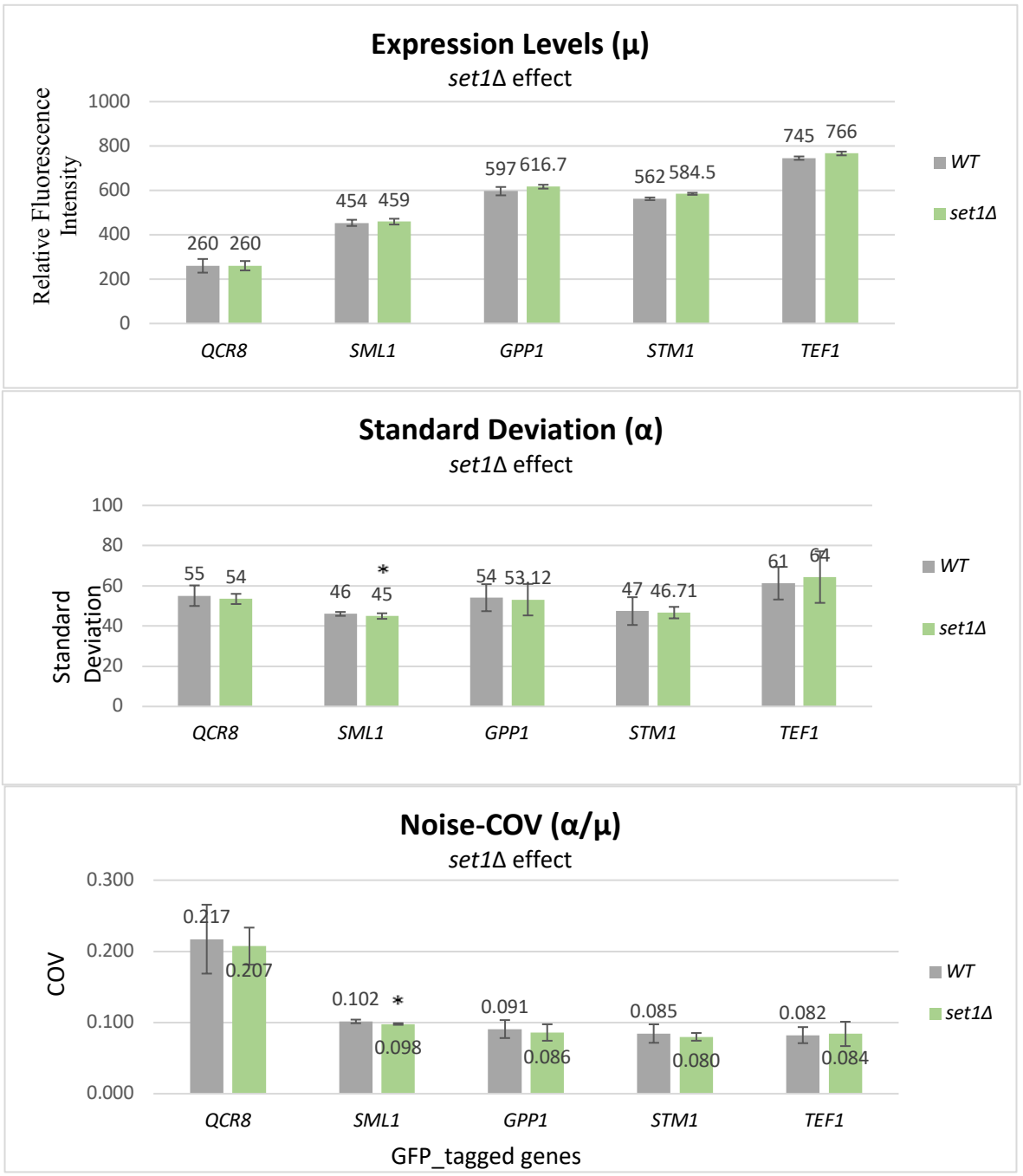


Figure 19. *set1*Δ affects *SML1* noise. Note: *, P < 0.05. Error bars= standard deviation between experiments (N=3).

In conclusion, for *SML1*, the data suggests that H3K4me3 supresses noise while H3K4me2 enhances noise. This is in line with the hypothesis put forward in this study. Interestingly, however, this effect was observed at only one gene, out of the 9 genes studied.

3.2.7. Effect of H3K4me is limited to certain genes

To examine how widespread the effect of H3K4me is on noise, the data in this work was compared to results obtained by other lab members at 9 additional genes. These genes have different noise levels based on the Newman's data (see Table 6). Similar to the observation at *GAL10*, *ARG1*, *SWD1* and *URA4,* on deletion of *SET1,* four of the 9 tested genes showed increase in gene expression levels as well as concomitant decrease in gene expression noise. This suggests that a great deal of "noise" results from changes in gene expression and extrinsic noise. Thus, at these genes, *SET1* suppresses "noise" indirectly and primarily by affecting levels of gene expression.

Table 7. The effects of set1Δ and spp1Δ on individual gene expression levels and expression noise (genes that were analysed in this study and by past laboratory members). Note: ***, P < 0.001; **, P < 0.01; *, P < 0.05. (N=3).

Gene Name	WT Mean	set1Δ Mean	spp1Δ Mean	WT COV	set1Δ COV	spp1Δ COV	Origin
***GAL10* (8h)**	550	591*** ↑	539	0.166	0.110***. ↓	0.170	This study
QCR8	260	260	259	0.217	0.207	0.205	This study
SML1	454	459	458	0.102	0.0980* ↓	0.1064* ↑	This study
GPP1	597	617	619	0.091	0.086	0.087	This study
STM1	566	583	574	0.081	0.08	0.078	This study
TEF1	745	766	758	0.082	0.08404	0.076818	This study
ARG1	290	316*** ↑	300	0.190	0.164*** ↓	0.185	This study
SWD1	275	291*** ↑	273	0.174	0.158*** ↓	0.180	This study
URA4	356	386*** ↑	264*** ↓	0.127	0.118*** ↓	0.182*** ↑	This study & Courtesy of J. Rayappu
BAT2	409	430.0*** ↑	409	0.112	0.0971***. ↓	0.1168* ↑	Courtesy of J. Rayappu
GCV3	333	334	336	0.157	0.147* ↓	0.16	Courtesy of J. Rayappu
ATP4	404	433.3** ↑	413	0.142	0.1063* ↓	0.135	Courtesy of J. Rayappu
SED1	272	279	277	0.192	0.185	0.184	Courtesy of J. Rayappu
GDH1	551	_	559	0.097	_	0.090	Courtesy of J. Rayappu
ARG4	269	279	281	0.202	0.183	0.183	Courtesy of P. Lorenz
HMS2	252	279	259	0.208	0.158	0.188	Courtesy of P. Lorenz
KAP104	302	310.8** ↑	301	0.167	0.1599* ↓	0.171	Courtesy of P. Lorenz
HOM3	343	362.5** ↑	354	0.123	0.1013***. ↓	0.113	Courtesy of P. Lorenz

Similar to the observation at the 9 tested genes in this study, level of H3K4me3 is also decreased in *spp1*Δ at the 9 genes tested by other lab members (Table 6). However, the *spp1*Δ effect on H3K4me2 levels is variable in individual genes (Table 5). At these five genes, levels of H3K4me2 increase on deletion of *SPP1*, similar to the five gene in this study. Thus, at all these genes the combined effect of loss of H3K4me3 and the increase in H3K4me2 might be expected to increase noise in the *spp1*Δ. However, the expression of only 3 genes showed increased noise in the *spp1*Δ and one of these genes, *SML1*, showed no increase in levels of H3K4me2.

At one of the 9 tested genes, *URA4*, the increase in noise can be explained by the decrease in mean expression. Furthermore, only 2 genes of the 18 genes studied (*BAT2* and *SML1*) showed a change in noise without a change in gene expression level in *spp1*Δ. Similarly, two genes (*GCV3* and *SML1*) showed a change in noise without a change in gene expression in *set1*Δ. Finally, *SML1* is the only gene to show both increased noise in *spp1*Δ and decreased noise in *set1*Δ without a change in gene expression noise. These results are consistent with the hypothesis put forward in the present study but suggest that the effect of loss of *spp1*Δ or *set1*Δ are limited to selected genes (Table 7).

3.3. Discussion

Previous work has demonstrated a correlation between the H3K4 methylation and gene expression noise (Newman *et al.,* 2006; S. Murray, unpublished). The aim of the current project is to further investigate the causal relationship between H3K4 methylation and gene expression noise. This study postulates that the loss of all types of methylations (*set1*Δ) would decrease gene expression noise, whereas the loss of H3K4me3 (*spp1*Δ) would increase noise. This hypothesis holds but is limited to certain genes.

The loss of Set1 or Spp1 primary changes gene expression levels. This might be caused by nucleosome stability or better access to transcriptional factors in *set1*Δ and the opposite in *spp1*Δ (Chandrasekharan, Huang and Sun, 2010; Vermeulen *et al.,* 2010). Thus, the observed change in noise in *set1*Δ or *spp1*Δ, for some genes, might be a result of extrinsic noise. Cell sizes and cell cycle status are also major contributors of extrinsic noise (Becskei *et al.,* 2005). To exclude extrinsic noise, tested genes can be also tagged with a red florescent protein (RFP). The two fluorescent proteins (GFP and RFP/YFP) can be incorporated to two alleles of the same endogenous gene. Controlled by identical promoters, this can be used to determine intrinsic noise. The fluorescence difference between GFP and RFP can only contribute to intrinsic noise because, in the same intracellular environment, GFP and RFP are expressed from two identical copies of a specific gene (Elowits et *al*., 2002; Wu *et al.,* 2017). In this study the autofluorescence values of the untagged stains was not subtracted from the expression data of the tagged genes. Subtracting the background fluorescence might alter the data in which it would give more accurate measurements of the change in fluorescence intensities between WT and mutant strains of the tagged genes without concerning about how the change of autofluorescence in each repeat would affect the results.

Only ~20% of the tested genes, loss of Set1 or Spp1 does not affect gene expression levels but does alter noise (*SML1* and *BAT2* in *spp11*Δ, *SML1* and *GCV3 set1*Δ). This rarer event might not be observed in other genes due to neutral effect resulted from the association of decreasing of H3K4me3

with increasing in H3K4me2 in *spp1*Δ. At the promoter region, loss of Spp1 would not only decrease H3K4me3 as it is also associated with a build-up of H3K4me2, which is hypothesized to influence gene expression noise. The *spp1*Δ strain shows lower levels of H3K4me3 in all genes in comparison to the WT strain, whereas it shows an increase in H3K4me2 at the promoter region of some tested genes and no effect for the other genes such as *SML1* (Table 6). Furthermore, a recent study found that Spp1 is not only alter H3K4me3, but also it affects the overall activity of Set1C. In *SPP1* deletion, loss of H3K4me2 in downstream and upstream shift of H3K4 methylation gradient (H3K4me3 located in proximal promoter regions, followed by H3K4me2 and then H3K4me1 located in distal promoter sites) was observed (Soares *et al.,* 2017). Shi et al., (2007) suggest that the Set1 complex might be stabilised by Spp1 through its binding to H3K4me2 via its PHD domain pushing the equilibrium further towards H3K4me3. This could explain why *spp1*Δ also shows higher di-methylation levels beside the lower tri-methylation levels.

As mentioned before, Howe et al., (2014) found that levels of H3K4me3 are reduced in H3K14A mutants, whereas H3K14R, which was used as a control, does not affect H3K4me3 levels. H3K4me3 levels were reduced to 40% of the WT levels in H3K14Q mutants, whereas levels were reduced to 20% of the WT levels in H3K14A mutants. H3K14A mutants do not affect H3K4me1 or H3K4me2 levels, thus these could be used to test the effect of gradually changing H3K4 tri-methylation levels on gene expression noise. These strains would provide insight on whether H3K4me3 acts gradually in effecting gene expression noise or not.

H3K4 methylation might work in redundancy with H3K79 methylation. A recent study found that H3K79 methylation can regulate gene expression noise. *DOT1* deletion increases noise for the highly expressed genes, and it does not change the expression levels of those genes (Wu *et al.,* 2017). Set1 and Dot1 both depend on ubiquitination of H2B on lysine 123 (Dover *et al.,* 2002; Wood *et al.,* 2005; Briggs *et al.,* 2002), and both H3K4 and H3K79 methylations are strongly correlated with gene expression noise (Fig. 8). An indirect link between H3K79me and H3K4me may together affect gene

expression noise. Therefore, H3K79 and H3K4 methylation may indirectly stimulate each other in highly expressed genes affected by that protein noise.

AF9, a subunit of YEATS domain, is strongly preferred to bind to H3K9ac (Le Masson *et al.,* 2003; Gates *et al.,* 2017; Li *et al.*, 2014). On active chromatin, this interaction is important for the activation of Dot1 to deposit H3K79 methylation (Li *et al.*, 2014). Furthermore, both H3K4 methylation and H3K9ac are also associated with active transcription (Vermeulen *et al.,* 2010; Bian *et al.,* 2011). For certain genes, the loss of H3K4me3 by *SPP1* deletion leads to the loss of acetylation at lysine 9 on histone 3 (H3K9ac). H3K4me3 can promote H3K9ac through binding to Sgf29 (a SAGA subunit) and this interaction is required for maximal acetylation of H3K9ac. It has also been found that *SGF29* deletion leads to a larger loss of H3K9ac than in *SPP1* deletion (Bian *et al.,* 2011; Vermeulen *et al.,* 2010). Therefore, H3K79 and H3K4 methylation may indirectly stimulate each other in highly expressed genes affected by that protein noise. Accordingly, a greater effect on protein noise would be observed with the loss of both H3K4 and H3K79 methylation.

In conclusion, this result has given some insights into the role of H3K4 methylation in gene expression noise. Further work is needed to determine why only certain genes are sensitive to the loss of H3K4 methylation.

Chapter 4.

Investigating the link between H3K4me3 and H3K79me and its role in noise

4. Investigating the link between H3K4me3 with H3K79me and its role in noise

4.1. Introduction

Dot1 is the only known enzyme that is responsible for histone H3 lysine 79 (H3K79) methylation (Ng et al. 2002; van Leeuwen et al. 2002). Dot1 and hDOT1L, the human homologue, have similar protein structures and catalytic activity (Feng et al. 2002). It has been found that almost 90% of H3K79 is methylated *in vivo* (Van Leeuwen et al. 2002), and only a small correlation between gene activity and H3K79me3 have been observed (Pokholok et al. 2005). Sir2 and Sir3, silencing proteins, are restricted to telomeric regions by H3K79me. It has been shown that loss of H3K79me results in Sir proteins being delocalised over the genome thereby adjusting silencing at the telomeres (Van Leeuwen et al. 2002; Ng et al. 2002). Unlike the other KMTs, Dot1 does not contain the catalytic SET domain (Feng et al. 2002). The arginine methyltransferases catalytic domain is similar to the yeast and human Dot1 (Min et al. 2003; Sawada et al. 2004). Additionally, Dot1 is also different to other KMTs because it is unable to control the generation of specific H3K79 methylation states due to the non-processivity of Dot1 resulting in overlapping gene distributions of the three H3K79 methylation states (Frederiks et al. 2008; Shahbazian et al. 2005). Furthermore, since the location of H3K79 in the nucleosome body is on an uncovered surface (unlike for many other H3 modifications which are on the histone tails), H3K79 does not contact any other histones or DNA (Luger et al. 1997a).

It is becoming increasingly clear that histone modifications could play a role in gene expression noise. For instance, Weinberger et al. (2012) found that Rpd3(L)C, a histone deacetylase complex, could indirectly regulate gene expression noise. Another study, Wu et al. (2017), reported that deleting *DOT1* – required for the addition of all types of methylation at lysine 79 on histone 3 (H3K79) in *S. cerevisiae* – increases noise for highly expressed genes. It is worth noting however that this change in noise was only observed in two genes out of the 7 tested genes (Wu et al. 2017).

As mentioned before, it has been shown that histone modifications participate in a crosstalk rather that acting independently, and one modification can depend on a second modification. As histone modifications are associated with chromatin remodelling, the crosstalk between those histone modifications could also affect their processes and functions on chromatin. Therefore, it is significant to explore more about how a particular modification such as H3K4me functions with other modifications to understand more about the role of this modification in gene expression noise. Therefore, as the previous results in this study show that the loss of H3K4me3 in *spp1*Δ can only affect the noise of certain genes, H3K79 methylation is a possible candidate for creating this limitation of H3K4me3 effect in gene expression noise.

4.1.1. Crosstalk between H3K79me and H3K4me3

One of the possible mechanisms of the crosstalk is that one modification can depend on a second modification. For instance, H3K4 and H3K79 di- and trimethylation require the requirement of H2B monoubiquitination (Bannister and Kouzarides 2011). In yeast, H2Bub has been shown to direct di- and tri-methylation of H3K4 and H3K79 through the recruitment of relevant enzymes, Set1 and Dot1, respectively, facilitating histone crosstalk in both budding yeast and mammals (Dover *et al.,* 2002; Wood *et al.,* 2005; Briggs *et al.,* 2002). In H2B, ubiquitin is attached into K123 by ubiquitin-conjugating enzymes, which are Rad6 and Bre1 (the E3 ligase) (Osley et al., 2004; Shilatifard, 2006; Weake and Workman, 2008). On the other hand, Ubp8 and Ubp10 are H2BK123ub1 deubiquitinases, and they specifically remove ubiquitin in H2BK123. The trans-histone crosstalk between H2BK123ub and H3K4/-K79 has been shown when levels of H3K4me2, H3K4me3 and H3K79me3 are removed when *RAD6* or *BRE1* is lost (Osley et al., 2004; Shilatifard, 2006; Weake and Workman, 2008) while levels of H3K4 and H3K79 methylation are increased in the loss of *UBP8* and *UBP10* (Henry et l., 2003; Gardner et al., 2005) (Fig. 20).

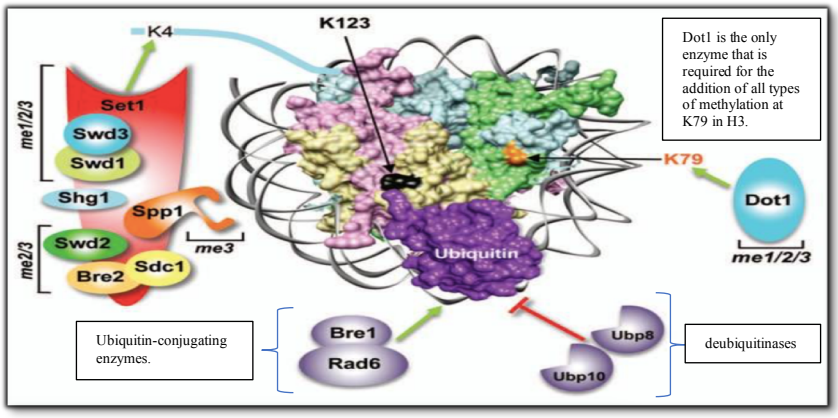


Figure 20. Crosstalk between H2BK123ub1 and H3K4/H3K79 methylation

Adapted from Chandrasekharan *et al.,* 2010.

Schematic figure shows an ubiquitinated nucleosome structure in yeast (H3, pale blue; H4, pale green; H2A, pale pink; H2B, pale yellow).

4.1.2. H3K4me and H3K79me intensities corelate with transcriptional bursting

Gene expression noise changes inversely with burst frequency (repetition of burst events), whereas it changes proportionally with burst size (average number of mRNA molecules made per burst event) (Paulsson, 2004). Transcriptional initiation is considered to be the most influential step for generating noise. At transcriptional initiation, the assembly of the pre-initiation complex, RNA polymerase II processivity and availability of transcription factors can affect transcriptional bursting parameters, which are unlikely to be determined by transcriptional elongation (Gerber & Shilatifard, 2003; Nguyen & Zhang, 2011; reviewed by Nicolas *et al.,* 2017).

As mentioned before, Newman et al. (2006) analysed noise levels in *S. cerevisiae* for 2,204 genes using a library of GFP-tagged genes along with FACS. Following Newman's et al. (2006) work, we found at the genome-wide level some strong correlations in 1,180 gene features, including histone modifications, and noise (S. Murray, unpublished). The histone modifications that most correlate with protein noise are the three types of methylation marks: H3K4me1, H3K4me2, and H3K4me3. Furthermore, tri-methylation of lysine79, lysine36 and ubiquitylation of lysine K123 in the gene body of histone 3 (H3K79me3, H3K36me3 and H2BK123ub, respectively), show a positive correlation with noise (S. Murray, unpublished). Furthermore, Wu et al., 2017 also followed Newman's et al.

(2006) work and investigated the association of histone modification with noise by calculating transcriptional burst frequency (r_{BF}) and size (r_{BS}) of these genes. They also calculated the correlation coefficients between intensity of histone modification and r_{BF} and r_{BS} (Fig. 21). H3K79me3 (gene-body-localized histone modification) has been shown to be correlated with r_{BF} more than it does with r_{BS} while H3K4me3 (Promoter-localized histone modification) correlates with both r_{BF} and r_{BS} (Wu et al., 2017). Since transcriptional initiation can determine transcriptional bursting parameters, which are unlikely to be affected by transcriptional elongation (Gerber & Shilatifard, 2003; Nguyen & Zhang, 2011; reviewed by Nicolas *et al.,* 2017), promoter-localized histone modification are more likely to play a role in both transcriptional bursting and noise more than gene-body-localized histone modification. Therefore, a possible crosstalk between H3K4me3 and H3K79me, where the function of one modification can depend on a second modification, may affect noise.

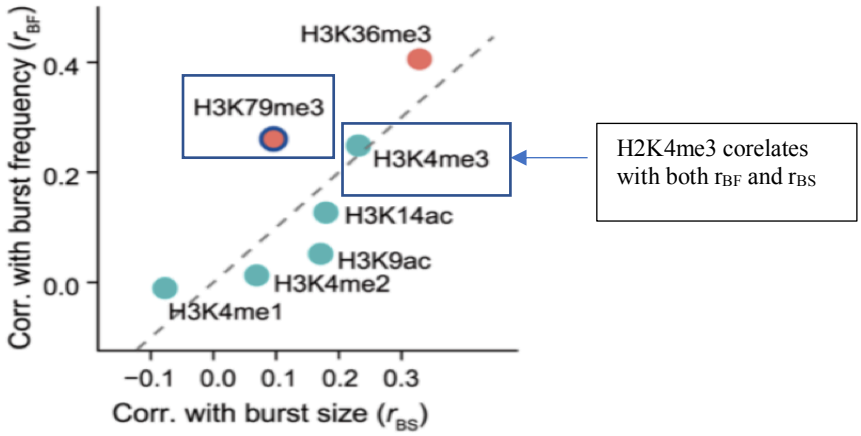


Figure 21. Correlations of the intensities of histone modification with transcriptional bursting parameters.

Adapted from Wu et al., 2017.

Schematic shows the correlations of the intensities of histone modifications with transcriptional burst frequency (Y-axis) and transcriptional burst size (X-axis).

4.1.3. Set1 and Dot1 function together

In meiosis, the loss of Set1 has been found to play a role in the formation of double-strands breaks (DSBs). Interestingly, it has been shown that in absence of Set1, Dot1 is involved in the formation of DSBs (Bani Ismail et al., 2014). Furthermore, *SWD2*, a subunit of Set1C (Miller et al., 2001; Roguey et al., 2001), is required for H3K4me3 (Cheng et al., 2004), and has been also found to interact with Dot1. *SWD2* is also required for H3K79me3, and the loss of *SWD2* results in a reduction of H3K79me3 (Lee et al., 2007). Swd2 might be a factor that translate crosstalk between H2BK123ub and H3K4 and H3K79 methylation (Lee et al., 2007). However, the mechanism in which *SWD2* can affect both H3K4me and H3K79me is not fully known.

While both H3K4me3 and H3K79me3 show a strong negative correlation with noise, the levels of their enzymes (Set1C (Spp1) and Dot1, respectively) that are responsible for depositing those modification show a weak correlation with noise (Table 8). Therefore, the change in noise on the loss of *SPP1* or *DOT1* is likely to result from the loss of the histone modification. Alternatively because the tri-methylation marks and di-methylations marks have opposite effects on noise, this could explain that luck of correlation between the enzymes and noise.

Table 8. Enzyme levels correlation with noise

Enzymes/ modification levels at promoter	**Correlation with noise (spearman correlation coefficient)**
Dot1	0.01
Set1	-0.03
Spp1	-0.20
H3K79me3	-0.25
H3K79me2	0.25
H3K4me2	0.38
H3K4me3	-0.32

The genes that were used to test the link between Spp1 and Dot1 and its role in gene expression noise are listed in Table 9. Previous results in this study show that at only ~20% of the 18 tested genes, loss of Spp1 does not affect gene expression levels but alters noise (*SML1* and *BAT2 spp1*Δ). Therefore, one of these genes will be first tested to examine the proposed relationship between Spp1 and Dot1. Different genes with different expression and noise levels will be also tested.

Table 9. Yeast genes analysed with FACS experiment to test the link between *SPP1* and *DOT1* and their function in noise (SGD).

Gene Name	**Function of Protein (SGD)**	**Expression** (Wu et al., 2017)	**Noise** (Wu et al., 2017)	**Noise affected upon SPP1 deletion**
SML1	ribonucleotide reductase inhibiting protein SML1	High	High	Yes
URA4	dihydroorotase	High	High	Yes
GCV3	glycine decarboxylase subunit H	High	High	Yes
TEF1	translational elongation factor	High	Low	No
ARG1	glutamate dehydrogenase (NADP(+)) GDH1	Low	High	No
SWD1	Argininosuccinate lyase	Low	Low	No

4.1.4. Aims

The work in this chapter aims to investigate the link between H3K4me3 and H3K79 methylation and its role in gene expression noise. The relative activities of Spp1 and Dot1 in noise will be examined by monitoring noise levels in the wild-type and *dot1Δ-spp1Δ* double knock-out mutant strains. The potential for crosstalk between H3K79me and H3K4me3 will be explored to understand what the contribution of Dot1 and Spp1 are to gene expression noise (Fig. 22).

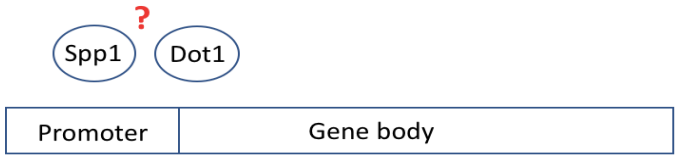


Figure 22. The link between Spp1 and Dot1 is still unknown.

4.2. Results

4.2.1. Loss of DOT1 affect gene expression and noise at SML1

Six strains were constructed by tagging 6 genes (*URA4, SML1, GCV3, TEF1, ARG1* and *SWD1*) with the open-reading frame encoding COOH-fused GFP and a selective marker as shown in previous chapter (Fig. 14). As a first approach to test the hypothesis, strains that lack H3K79me were created via deletion of *DOT1*, and noise level was monitored using FACS.

The loss of *DOT1* (H3K79me) does not affect gene expression except at *SML1* (Fig. 23). In *dot1Δ,* gene expression is decreased by 9.7% while standard deviation and noise levels increase by 4.2% and 14.3% respectively at *SML1* (Fig. 23). As mentioned before, Newman (2006) showed a strong negative correlation between expression levels and noise, the change in overall expression levels is the most likely explanation for the observed changes in standard deviation and COV (noise) at *SML1.* However, the observed results in *SML1* is consistent with the reported correlation that H3K79me3 negatively correlated with gene expression noise. The results are also consistent with previous finding, where *DOT1* deletion increases noise for the highly expressed genes (Wu *et al.,* 2017). However, the other 5 genes show no significant effect either in gene expression or noise in absence of *DOT1* (Fig. 23). This observation is also expected since Wu et al., 2017 showed only 2 out of the 7 tested genes were affected in absence of *DOT1*. This mean the effect of Dot1 on noise is also limited for certain genes.

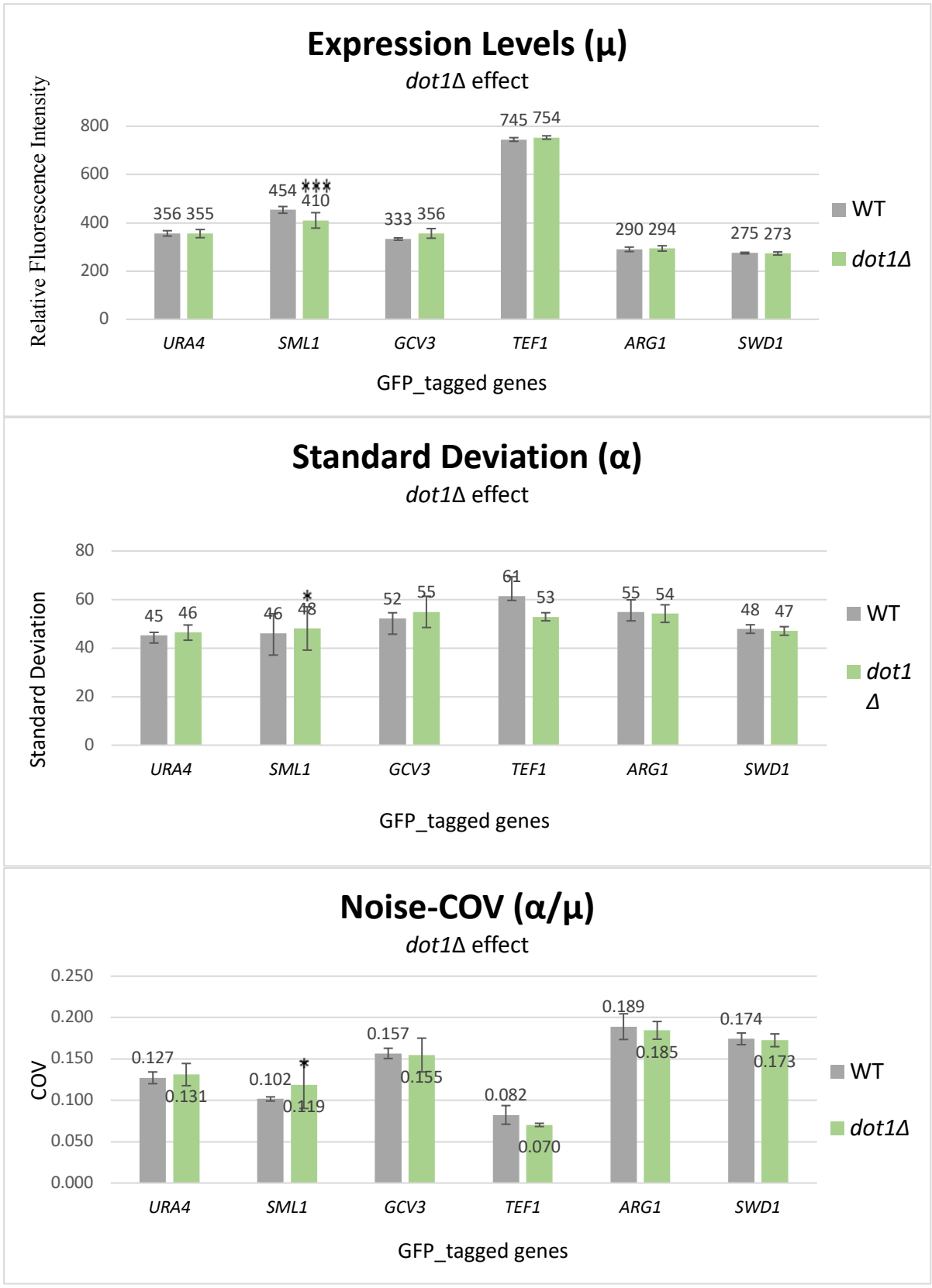


Figure 23. The loss of *DOT1* affects only *SML1* expression and noise levels.

***, P < 0.001; *, P < 0.05. Error bars= standard deviation between experiments (N=3).

4.2.2. The double knock-out of spp1Δ-dot1Δ increases noise and decreases expression at ARG1

Since Spp1 affects the overall activity of Set1C (Soares *et al.,* 2017), the role of H3K4 and H3K79 methylation could be investigated by studying the role of both Spp1 and Dot1 in noise. *spp1Δ* and *dot1Δ* double knock-out strains were created to lack both H3K4me3 and H3K79me, respectively. Gene expression noise is then measured to investigate if the loss of both *SPP1* and *DOT1* has any effect in noise.

The first striking observation that was observed in *spp1Δ* and *dot1Δ* double knock-out strains is that in the loss of both H3K4me3 and H3K79me (*spp1*Δ −*dot1*Δ), only *ARG1* shows a significant decrease in expression level (by ~13%) and a significant increase in standard deviation and noise (by ~8.3% and 22%, respectively) (Fig. 24). The observed results are consistent the reported correlation, where both H3K79me3 and H3K4me3 negatively correlate with noise. Therefore, the loss of H3K79me3 (*dot1Δ*) and H3K4me3 (*spp1Δ*) would result in the increase of noise as was observed at *ARG1*. Interestingly, in single deletion of *SPP1* or *DOT1*, there was no observed effect in both expression and noise levels. This observation is consistent with the hypothesis and suggests the redundant function of Dot1 with Spp1. However, as gene expression and noise have a strong negative correlation (Newman et al., 2006), the change in overall expression levels is the most likely explanation for the observed changes in standard deviation and COV (noise) at *ARG1*.

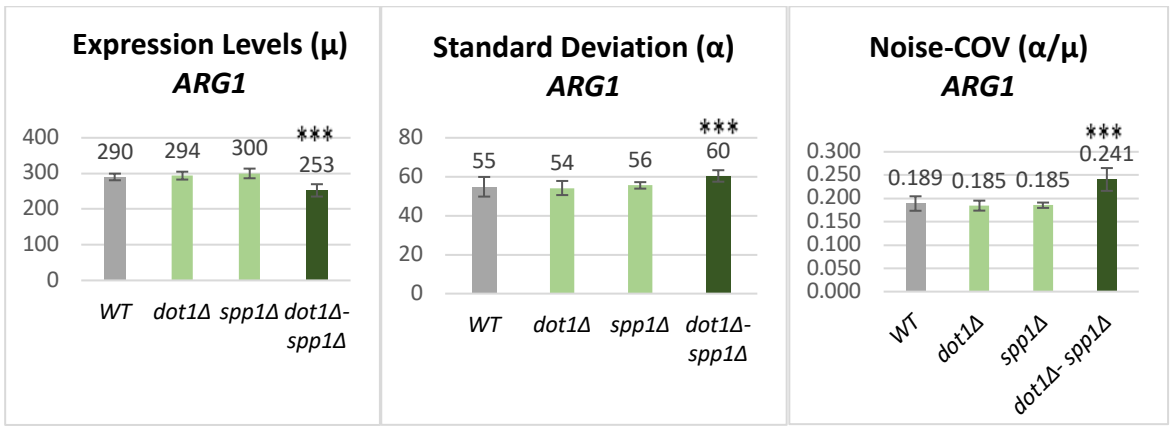


Figure 24. Dot1 redundantly functions with Spp1 in *ARG1*.

The double knock-out of spp1Δ-dot1Δ decreases expression and increases noise levels at *ARG1*. Note: ***, P < 0.001. Error bars= standard deviation between experiments (N=3).

4.2.3. The loss of DOT1 prevents the effect that was observed in spp1Δ at URA4 and SML1

Previous results in this study show that the loss of *SPP1* does not have effects on the gene expression at all genes except *URA4* (Fig. 15 and Fig. 16). Deletion of *SPP1* leads to a lower mean expression level (by ~26%) and increase in noise (by ~28%) at *URA4*. By contrast, the loss of *DOT1*, neither gene expression nor noise were not affected at *URA4*. Interestingly, the loss of both H3K4me3 and H3K79me (*spp1*Δ −*dot1*Δ) brings back both the noise and expression level to the WT at *URA4* (Fig. 25).

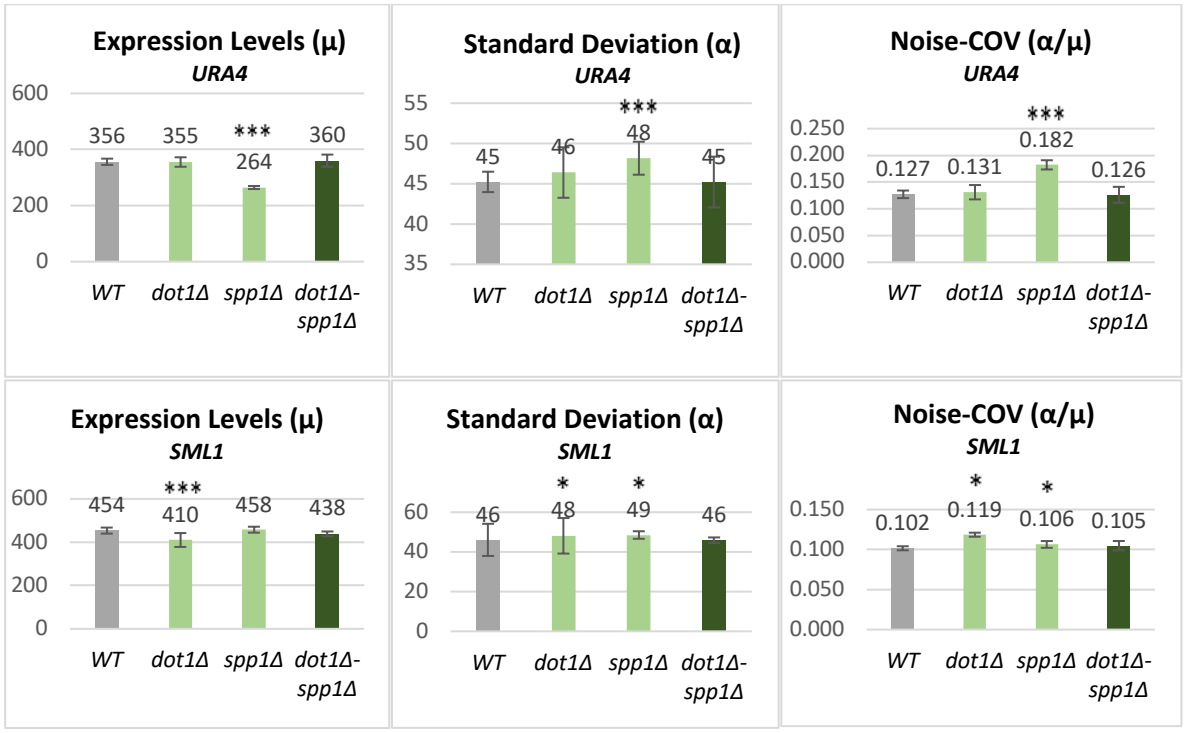


Figure 25. The loss of both *DOT1-SPP1* inhibit the observed increase in noise in *spp1Δ*.

Note: ***, P < 0.001; **, P < 0.01; *, P < 0.05. Error bars= standard deviation between experiments (N=3).

Furthermore, previous results in this study show that the loss of *SPP1* do not affect gene expression levels but alters noise at *SML1* and *GCV3*. *SML1* show an increase in noise level without an effect in gene expression in *spp1*Δ. Like *URA4*, similar result was also observed at *SML1*, the loss of both H3K4me3 and H3K79me brings back both the noise and expression level to the WT. The loss of H3K79me prevents the observed increased noise and decreased expression that was observed in the loss of H3K4me3 in *spp1Δ* at *URA4*. The presence of the Dot1 (or H3K79me2 that is positively correlated with noise) when *SPP1* is removed may create the observed noise. However, other genes that do not show any effect in the absence of *SPP1, DOT1* or both *SPP1-DOT1* do not show any effect in either expression or noise (Fig. 26).

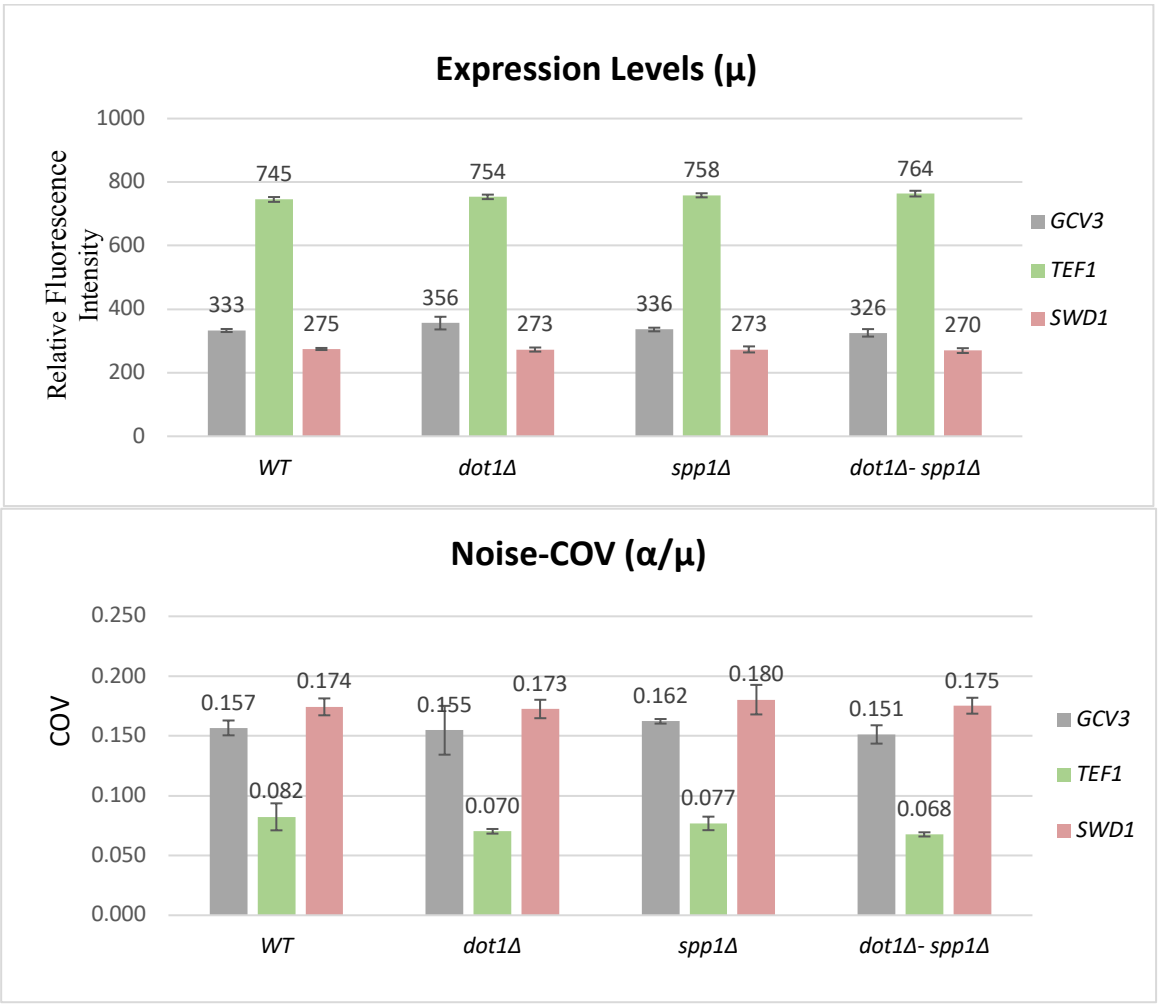


Figure 26. No change in both gene expression and noise levels in *dot1Δ-spp1Δ*.

Error bars= standard deviation between experiments (N=3).

Next, we wanted to test if similar observation would be observed in absence of both *SET1* and *DOT1* at *URA4.* While loss of H3K4me (*set1*Δ) decreases noise and increases expression at *URA4,* the loss of both H3K4me and H3K79me in *set1*Δ −*dot1*Δ keeps both the noise and expression level values similar to loss of only H3K4me in *set1*Δ (Fig. 27). This means the observed results when the double knock-out of *spp1Δ-dot1Δ* prevents the increase in noise in the loss of *SPP1* is only associated to the redundant function of Dot1 (H3K79me) with Spp1 (H3K4me3).

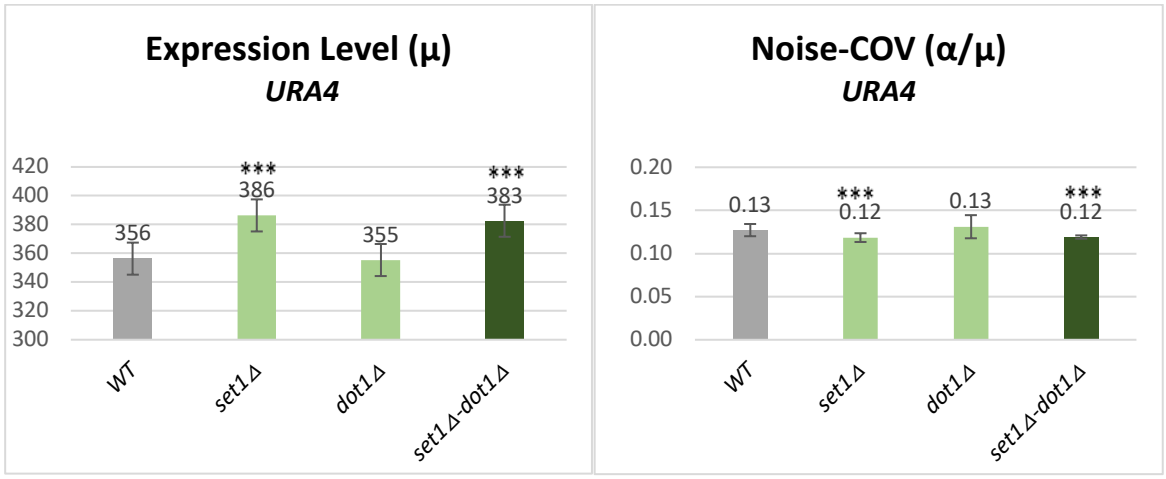


Figure 27. The loss of both *DOT1-SET1* does not alter the effect of the loss *SET1* at URA4.

Note: ***, P < 0.001. Error bars= standard deviation between experiments (N=3).

In conclusion, for *ARG1,* the data may suggest that both Dot1 (H3K79me3) and Spp1 (H3K4me3) suppress noise, while at both *ARG1*, *URA4* and *SML1* the data suggests that Dot1 redundantly functions with Spp1. This is in line with the hypothesis put forward in this study. Interestingly, however, the effect of both Dot1 and Spp1 was observed at only one gene (*ARG1*), out of the 6 genes studied. Next, the effect of loss of *SPP1* or *SET1* on H3K79me3 (positively correlated with noise) and H3K79me2 (negatively correlated with noise) was tested.

4.2.4. Reduction of H3K79me2/me3 levels upon the loss of Dot1

Further investigation in the different effects of the loss of *SPP1* or *SET1* on H3K79me3 were validated by chromatin immunoprecipitation (ChIP)-qPCR at the promoter region of 4 genes (*URA4, SML1, GCV3* and *BAT2*). First, since Dot1 is the only methyltransferase that is responsible for the addition of all types of methylation at K79 in H3, the loss of *DOT1* would remove H3K79me3 and H3K79me2 in individual genes. As depicted in Fig.28, a significant reduction in H3K79me3 and H3K79me2 levels were observed in absence of *DOT1*. The absence of H3K79me3 and H3K79me2 in the *dot1Δ* strain indicates that the antibodies raised against H3K79me3 and H3K79me2 are specific for a modification on H3K79.

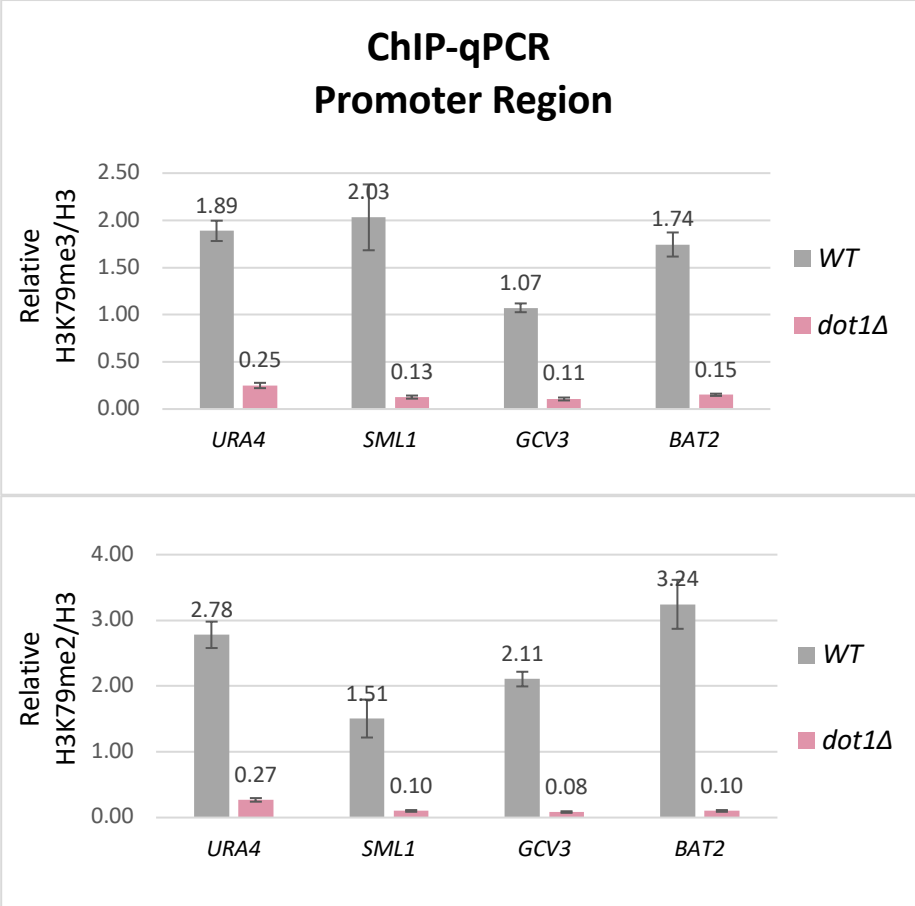


Figure 28. Reduction in H3K79me3 and H3K79me2 in the loss of *DOT1*.

Levels of H3K79me3 assessed by chromatin immunoprecipitation (ChIP)-qPCR in the promoter region of *URA4, SML1, GCV3* and *BAT2*. Signals are normalised to levels of H3. Error bars show standard errors of the real- time PCR reaction.

4.2.5. Increase in H3K79me3 levels upon the loss of SPP1

To investigate the H3K79me3-H3K4me3 crosstalk further, whole cell extracts were prepared from yeast strains in which H3K4me3 was reduced by the loss of *SPP1* and analysed by ChIP-qPCR for specific levels of H3K79me3. In the loss of *SPP1,* the reduction levels of H3K4me3 at the promoter region for *URA4, SML1, GCV3* and *BAT2* are 64%, 61%, 67% and 67%, respectively (Table 6). As depicted in Fig. 29 and Fig.30, levels of H3K79me3 at the promoter region of those genes was determined in *spp1*Δ strain.

Interestingly, levels of H3K79me3 increase upon the loss of *SPP1* (H3K4me3) at the promoter of *URA4*, *SML1*, *GCV3* and *BAT2* (by 41%, 28%, 64%, 31%, respectively) (Fig. 29). As has been shown before, both H3K79me3 and H3K4me3 negatively correlate with noise, both may suppress gene expression noise. This means when the loss of H3K4me3 in *spp1*Δ strain is associated with an increase on H3K79me3, a neutral effect in noise would be observed. For example, *SML1* and *BAT2* show the least increase in H3K79me3 levels in absence of *SPP1*. As shown in previous chapter, *SML1* and *BAT2* were the only genes that were affected in their noise without changing in their expression levels by the loss of *SPP1.* While *GCV3* is the only tested gene that shows no change in both gene expression and noise in *spp1*Δ*,* it shows a huge increase in H3K79me3 by 64% and a decrease in H3K4me3 by 67% upon the loss of *SPP1*. This could also suggest the neutral effect in noise that raises from the loss of *SPP1*. However, in the double knock-out of both *DOT1-SPP1*, no change in noise was observed in *GCV3*. This could be due to further redundant function, possibly change position or level of H3K36 methylation that also correlates with noise. Furthermore, the fact that there is a robust level of gene-specific H3K79me3 in the *spp1*Δ strain indicates that Dot1 (H3K79me3) redundantly functions with Spp1 (H3K4me3), which is consistent with the proposed hypothesis in this study.

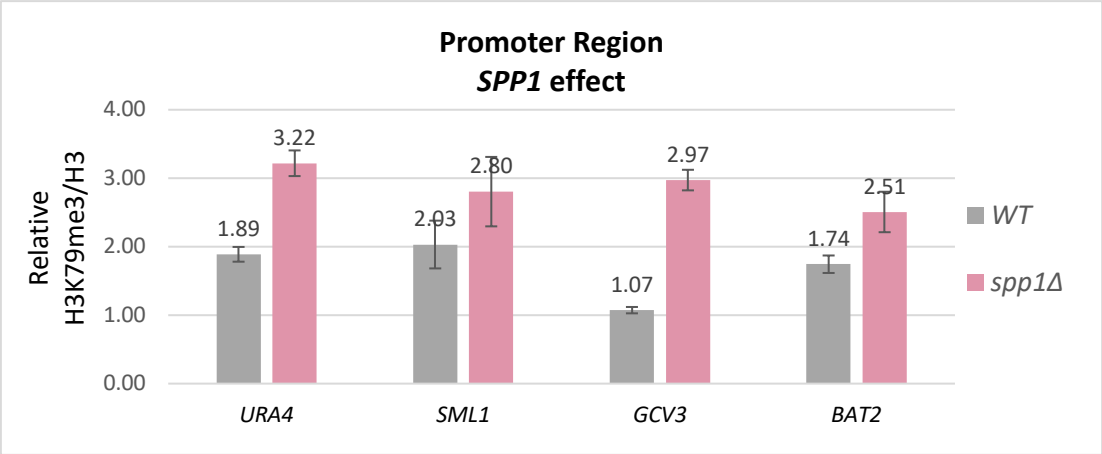


Figure 29. H3K79me3 levels increase at the promoter region of the tested genes upon the loss of *SPP1*.

Levels of H3K79me3 assessed by ChIP-qPCR. Signals are normalised to levels of H3. Error bars show standard errors of the real- time PCR reaction.

While H3K79me3 and H3K4me3 negatively correlate with noise and may suppress gene expression noise, H3K79me2 positively correlates with noise. At many genes, levels of H3K79me2 were decreased in absence of *SPP1* except for *SML1* (Fig. 30). Interestingly, *SML1* is the only gene that shows an increase in noise (Fig. 24) and an increase on H3K79me2 levels (Fig. 30) in the loss of *SPP1*. Furthermore, the observed increase in noise in *SPP1* deletion was reduced to the wild-type on the loss of both *SPP1* and *DOT1* together. This may indicate the observed increase in noise on the loss of *SPP1* was due to the increased levels of H3K79me2 at *SML1*.

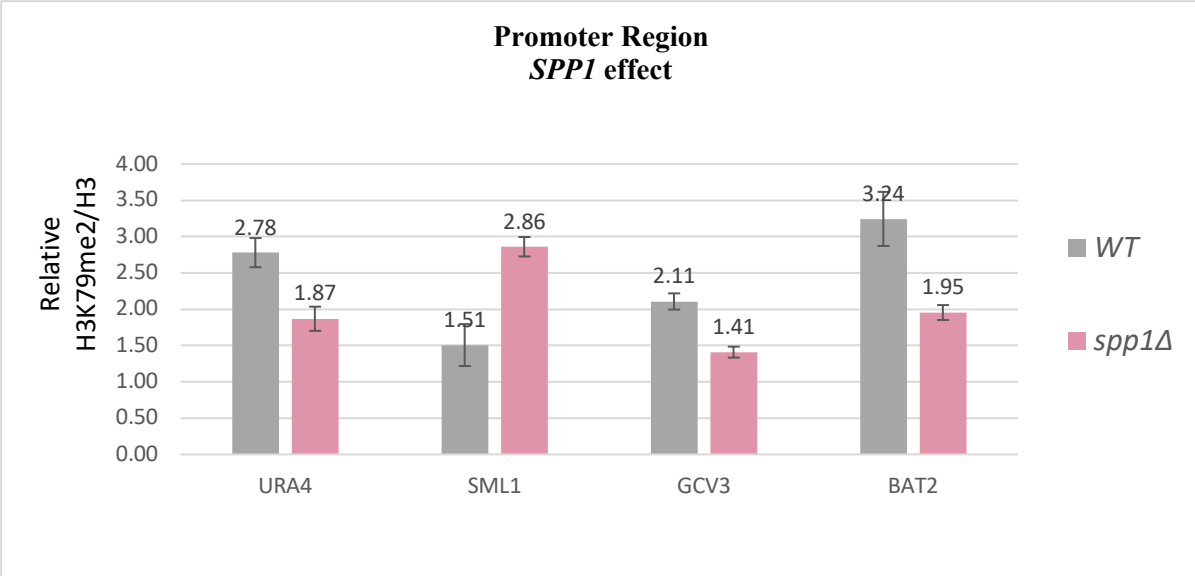


Figure 30. Decrease in H3K79me2 levels upon the loss of *SPP1* at promoter region except for *SML1*.

SML1 shows an increase in H3K79me2 in absence of *SPP1*. Levels of H3K79me3 assessed by chromatin immunoprecipitation (ChIP)-qPCR in the promoter region of URA4, SML1, GCV3 and BAT2. Signals are normalised to levels of H3. Error bars show standard errors of the real- time PCR reaction.

In *spp1Δ,* the loss of H3K4me3 (negatively correlated with noise) can be associated with both an increase in H3K79me3 (negatively correlated with noise) and a decrease in H3K79me2 (positively correlated with noise). This might explain why no effect in noise would be observed in certain genes such as for *GCV3* in the loss of *SPP1*. However, in *spp1Δ,* levels of H3K79me2 and H3K4me2 (which are positively correlated with noise) can be raised, which may create the noise as was observed in *SML1*.

4.2.6. Increase in H3K79me3 levels upon the loss of SET1 at promoter region

Similar to the observation that level of H3K79me3 was increased upon the loss of *SPP1*, also H3K79me3 levels were also increased on the loss of *SET1* in *URA4*, *SML1*, *GCV3* and *BAT2* (by 25%, 53%, 49%, 14%, respectively) (Fig. 31). Furthermore, there was no observed change in level of H3K79me2 upon the loss of all types of methylation by the deletion of *SET1* (Fig. 31B).

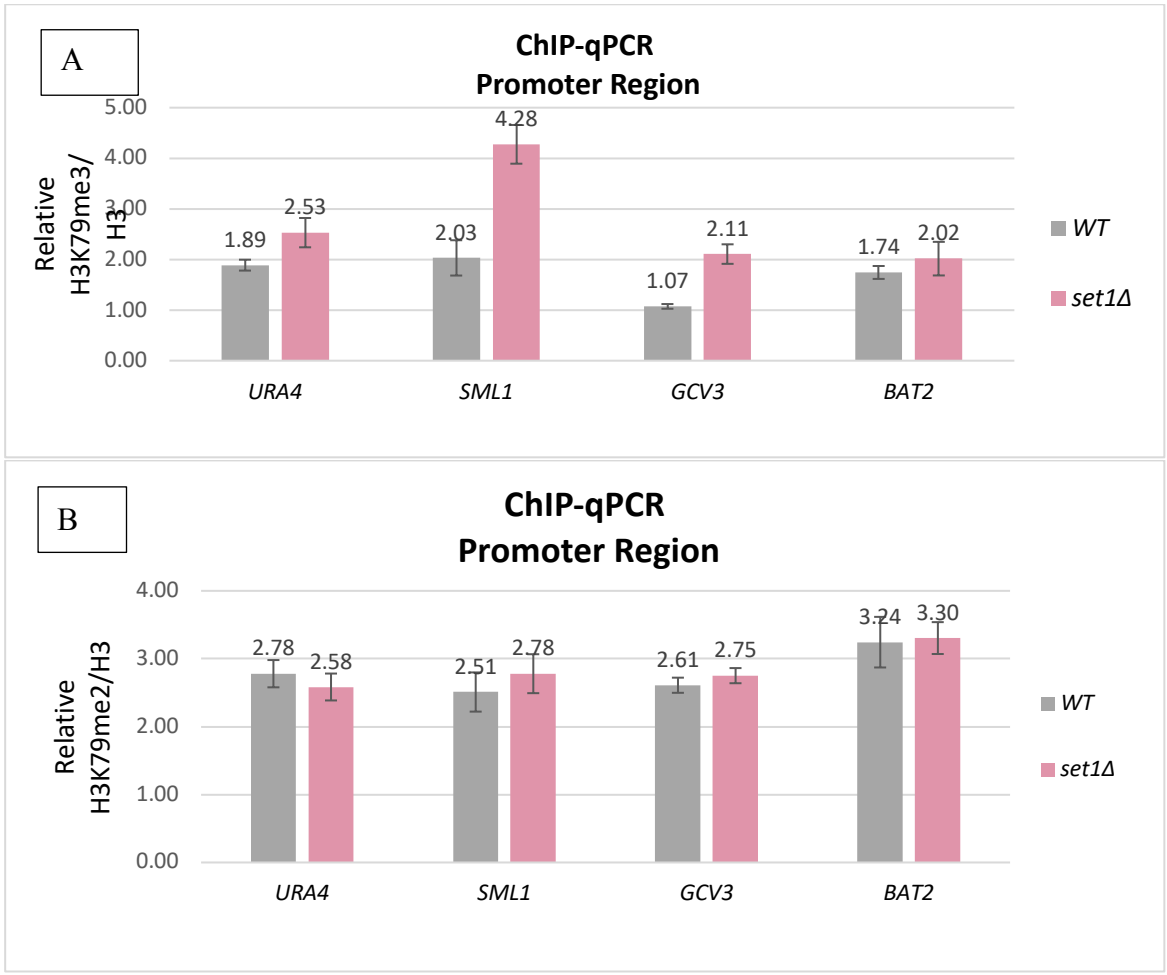


Figure 31. Levels of H3K79me3 and H3K79me2 upon the loss of *SET1*.

"A" increase in H3K79me3 levels upon the loss of *SET1*. "B" No change in H3K79me2 levels upon the loss of *SET1*. Levels of H3K79me2 assessed by chromatin immunoprecipitation (ChIP)-qPCR in the promoter region of *URA4, SML1, GCV3* and *BAT2*. Signals are normalised to levels of H3. Error bars show standard errors of the real- time PCR reaction.

In conclusion, at *ARG1,* the observed increase in noise and decrease in expression levels in the loss of both H3K79me and H3K4me3 in the *dot1Δ-spp1Δ* double knock-out were not observed in the single loss of either *DOT1* or *SPP1.* Furthermore, the loss of H3K4me3 in *spp1Δ* was also associated with an increase in H3K79me3. These observations suggest that Dot1 (H3K79me3) redundantly functions with Spp1 (H3K4me3), and that Dot1 and Spp1 may act together to suppress noise. This is in line with the hypothesis put forward in this study. Interestingly, however, this effect is also limited to certain genes.

4.3. Discussion

Previous work has demonstrated a strong correlation between both H3K79 and H3K4 methylation and noise (Newman *et al.,* 2006; S. Murray, unpublished). Furthermore, H2Bub has been shown to direct di- and tri-methylation of H3K4 and H3K79 through the recruitment of relevant enzymes, Set1 and Dot1, respectively, facilitating histone crosstalk in both budding yeast and mammals (Dover *et al.,* 2002; Wood *et al.,* 2005; Briggs *et al.,* 2002). The aim of this chapter was to further investigate the link between H3K79me and H3K4me3 and its effect in noise. This study postulates that H3K79me (Dot1) redundantly functions with H3K4me3 (Spp1) and may affect protein noise. This hypothesis holds but is limited to certain genes.

In this study, loss of *DOT1* (H3K79me) decreases gene expression at one out of 6 tested genes (*SML1*; Fig. 24). This might be due to some changes in transcriptional regulation related to the switch between heterochromatin and euchromatin as a result of the loss of H3K79 methylation (reviewed by Nguyen and Zhang, 2011). The loss of *DOT1* was also associated with an increase in noise at *SML1*. This observation was consistent with previous studies. For example, Wu et al., (2017) reported an increase in noise in 2 out of the 7 tested genes in absence of *DOT1* in yeast. Another study in yeast also shows that a lower noise level in genes that were inserted into high H3K79me3 genomic regions (Chen and Zhang, 2016). These results along with the reported strong negative correlation of H3K79me3 with noise (Newman *et al.,* 2006; S. Murray, unpublished) suggest that H3K79me3 may play a role in supressing noise. However, in this study, the observed change in noise in *dot1*Δ for *SML1* might be a result of the change in overall expression levels or extrinsic noise. As mentioned before, cell sizes and cell cycle status are also major contributors of extrinsic noise (Becskei *et al.,* 2005). To exclude extrinsic noise, tested genes can be also tagged with a red/yellow florescent protein. The fluorescence difference between florescent proteins can only contribute to intrinsic noise because, in the same

intracellular environment, florescent proteins are expressed from two identical copies of a specific gene (Elowits et *al*., 2002; Wu *et al.,* 2017). However, as the suggested redundant function of H3K79me3 with H3K4me3 is previously proposed to limit the effect of H3K4me3 in noise, it may also explain the limited effect of the loss of H3K79me3 on noise. It is worth examining if the loss of Dot1 can change levels of H3K4me3.

As indicated in the noise measurements in Fig. 26, loss of both Dot1 and Spp1 significantly decreases expression and increases noise to a significant extent in one of the 5 genes analysed (*ARG1*) while single deletion of either *DOT1* or *SPP1* does not affect either noise or expression levels. This observation is a novel finding presented in this study. Therefore, one can conclude that Dot1 (H3K79me) could redundantly functions with Spp1 (H3K4me3). Furthermore, a particularly noticeable increase in noise of more than a fifth was observed in *ARG1*. This redundant function may also impact gene expression noise and is, more specifically, likely to act as a suppressor of noise. Along with paying attention to the reported strong negative correlation of noise with gene expression (i.e. reduced levels of gene expression is associated with increased levels of noise) (Newman et al., 2006), the change in overall expression levels may also explain the observed changes in noise at *ARG1*. Therefore, the observed change in noise in *dot1*Δ- *spp1*Δ for *ARG1* might be a result of extrinsic noise.

The reason that the loss of Spp1 only changes noise at a limited number of genes might be due to the neutral effect of a concomitant increase in H3K79me3 when H3K4me3 levels decrease in *spp1*Δ. At the promoter region, loss of Spp1 would not only decrease H3K4me3 as it is also associated with a build-up of H3K79me3, which is also hypothesized to suppress gene expression noise along with H3K4me3, and a decrease in H3K79me2, which is hypothesized to enhance noise. Thus, the overall

response might reflect the sum of the effect of these modifications on noise. The *spp1*Δ strain shows lower levels of H3K4me3 in all genes in comparison to the WT strain (Table 6), whereas it shows an increase in H3K79me3 and a decrease in H3K79me2, at the promoter region of the four genes (*URA4, SML1, GCV3* and *BAT2*) analysed by ChIP-qPCR (Fig. 32). For instance, while *GCV3* is the only tested gene that show no effect in both gene expression and noise in *spp1*Δ*,* it shows a huge increase in H3K79me3 by 64% and a decrease in H3K4me3 by 67% upon the loss of *SPP1*. This result suggests the neutral effect in noise that has been raised in the loss of *SPP1* and emphasize the redundant function of Dot1 with Spp1 (H3K4me3). However, in the double knock-out of both *DOT1-SPP1*, no change in noise was observed in *GCV3*. Therefore, we proposed another factor such as H3K36me3, which also negatively correlated with noise and hypothesized to suppress noise, may have a redundant function with H3K79me3 or H3K4me3.

In yeast, *SET2*, which provides methylation of H3K36, has been shown to play a role in transcription elongation by modulating chromatin at transcribed DNA region (Lee et al., 2005). H3K36me3 has been also found to be associated with active transcription (Chantalat et al., 2011). Moreover, H3K36me3 has been shown to be negatively correlated with noise (Newman *et al.,* 2006; S. Murray, unpublished). Therefore, H3K36me3 is hypothesized to suppress noise as H3K4me3 and H3K79me3 do. Furthermore, Wu et al., 2017 reported that, like H3K79me3, H3K36me3 exhibits a stronger correlation with transcriptional burst frequency than it does with transcriptional burst size. As mentioned before, transcriptional burst frequency is shown to inversely correlate with noise (i.e. increase in transcriptional frequency is associated with decrease in noise level). Moreover, in several studies, positive feedback loop between transcription initiation and elongation has been proposed. For example, Chen and Zhang reported that, in GFP tagged genes, the intensity of H3K79me3 is still associated with low noise levels in the complete knock-out of the open reading frame of these genes. Along with the fact that transcriptional initiation is most likely to determine transcriptional burst

parameters (Gerber & Shilatifard, 2003; Nguyen & Zhang, 2011; reviewed by Nicolas *et al.,* 2017), possible presence of a positive feedback loop between transcription initiation and elongation may control both transcriptional bursting parameters and noise. Furthermore, Kozhina et al., (2016) has found that both Dot1 and Set2 have been involved in controlling similar pathways of DNA repair. Therefore, we propose a possible redundant function between Dot1 and Set2 in noise similar to the one observed between Spp1 and Dot1. *SET2* knock-out and double knock-out of *SET2* with either *DOT1* or *SPP1* would be required to test this hypothesis.

As mentioned before, in different genes, Spp1 can affect both H3K4me2 and H3K4me3 in different ways. For instance, the *spp1*Δ strain shows lower levels of H3K4me3 in all genes in comparison to the WT strain, whereas it shows an increase in H3K4me2 at the promoter region of some tested genes such as *HOM3, ARG4* and *KAP104* and no effect for the other genes such as *SML1* (Table 1). On the other hand, the loss of Dot1 would result in a reduction level of all H3K79me and not just H3K79me3, which is most in interest. As has been shown in previous chapter, out of the 18 tested genes, *URA4* was the only gene that show both a decrease in expression levels and increase in noise (Fig. 26). Interestingly, at *URA4*, the loss of both Spp1 and Dot1 removes the observed effect resulted from the deletion of only *SPP1*. This observation might result due to the presence of H3K79me2 (hypothesized to enhance noise), which could generate the noise at *URA4* in *spp1*Δ strain. Therefore, the loss of Dot1, which removes all types of methylation at K79 in H3, returns back both the expression and noise levels to WT. This result could also provide an evidence that Dot1 works with Spp1. Unfortunately, this observation along with the fact that loss of Spp1 would not only decrease H3K4me3 as it is also associated with a build-up of H3K4me2, which is hypothesized to influence gene expression noise, show that using either Dot1 or Spp1 to test the redundant function between H3K79me3 and H3K4me3 and its effect in noise is limited. That is because other factors such as H3K79me2 and H3K4me2 that may have opposite effect in noise could change upon the loss of Dot1 or Spp1.

In conclusion, this result has given some insights into the redundant function of Dot1 (H3K79me3) with Spp1 (H3K4me3) and its effect in gene expression noise. Further work is needed to determine why only certain genes are sensitive to this redundant function and what other factor could be involved with the redundant function of Dot1 with Spp1.

Chapter 5.

Investigating the role of H3K9ac in noise

5. Investigating the role of H3K9ac in noise

5.1. Introduction

Levels of H3 acetylation have been shown to be positively correlated with active transcription (Liu et al. 2005; Pokholok et al. 2005; Kurdistani et al. 2004; Pokholok et al. 2005; Guillemette et al. 2011). H3K9ac has been shown to be associated with active promoters, however, its function is not fully understood. As for other acetylation marks, H3K9ac marks may influence chromatin structure (Kouzarides, 2017) and promote progression of transcription into elongation (Gates et al., 2017). For example, in HeLa cells, two HAT complexes, SAGA (Spt-Ada-Gcn5 acetyltransferase) and ATAC (Ada2a-containing) catalyse acetylation through their subunits GCN5 and PCAF (Brownell et al., 1997; Spedale et al., 2012). It has been found that paused Pol II occupancy is increased in the loss of H3K9ac by knocking-out GCN5 (generally controls nonderepressible) and PCAF (p300/CBP-associated factor) (Jin et al., 2010; Foulds et al., 2013) for a number of genes, and that maximal Pol II pause release through the super elongation complex (SEC) has been shown to require H3K9ac (Gates et al., 2017; Jin et al., 2010; Foulds et al., 2013).

Transcriptional bursting, where a number of mRNA are being produced in short pulses, is the dominant transcription mode in eukaryotes, it serves as a major explanation of the overall gene expression noise (Raj *et al.,* 2006; Blake *et al.,* 2003). Gene expression noise changes inversely with burst frequency (repetition of burst events), whereas it changes proportionally with burst size (average number of mRNA molecules made per burst event) (Paulsson, 2004). It is becoming increasingly clear that histone acetylation could play a role in noise. For example, Weinberger et al. (2012) found that Rpd3(L)C, a histone deacetylase complex, could indirectly regulate gene expression noise. They found that the increase in acetylation level on the loss of Rpd3 leads to a higher noise level (Weinberger et al., 2012). Consistent with this, Wu et al., (2017) shows that H3K9ac strongly correlates with transcriptional burst size. This means high level of H3K9ac is associated with higher transcriptional burst size and noise level.

AF9, a subunit of YEATS domain, strongly prefers to bind to H3K9ac (Le Masson *et al.,* 2003; Gates *et al.,* 2017; Li *et al.*, 2014). On active chromatin, this interaction is important for the activation of Dot1 to deposit H3K79 methylation (Li *et al.*, 2014). Furthermore, both H3K4 methylation and H3K9ac are also associated with active transcription (Vermeulen *et al.,* 2010; Bian *et al.,* 2011). For certain genes, the loss of H3K4me3 by *SPP1* deletion leads to the loss of acetylation at lysine 9 on histone 3 (H3K9ac). H3K4me3 can promote H3K9ac through binding to Sgf29 (a SAGA subunit) and this interaction is required for maximal acetylation of H3K9ac. It has also been found that *SGF29* deletion leads to a larger loss of H3K9ac than in *SPP1* deletion (Bian *et al.,* 2011; Vermeulen *et al.,* 2010).

Since H3K9ac strongly correlates with transcriptional burst size, it may inhibit or limit H3K4me3 effect in noise. Therefore, it could be other explanation for limiting the effect of H3K4me3 in gene expression noise in *spp1*Δ. To investigate how H3K9ac could be involved in gene expression noise, the effect of the loss of Sgf29 in noise can be tested. We hypothesized that the loss of H3K9ac by the loss of Sgf29 would decrease noise. Similar genes with different noise levels that were examined in the previous chapter (Table 9) will be used to test this hypothesis.

5.2. Aims

The aim of this chapter is to characterise the relationship between H3K9ac and noise. In particular, it will be investigated whether the loss of H3K9ac by the loss of Sgf29, which prefers to bind to H3K4me3, will have any gene-specific noise effects.

5.3. Results

5.3.1. Loss of SGF29 decreases expression and increases noise

Six strains were constructed by tagging 6 genes (*URA4, SML1, GCV3, TEF1, ARG1* and *SWD1*) with the open-reading frame encoding COOH-fused GFP and a selective marker as has been shown in Chapter 3 (Fig. 16). The hypothesis discussed above states that H3K9ac enhances noise. To test the hypothesis, strains that lack H3K9ac were created via deletion of *SGF29*, and noise level was monitored using FACS.

Fig. 32 and Fig. 33 show the results of the FACS readings for the 6 tested genes in wild-type and *sgf29*Δ strains. The first striking observation is that deletion of *SGF29* primarily changes gene expression levels. Loss of Sgf29 significantly leads to a lower mean expression level at *URA4, SML1, SWD1* and *ARG1* (by 5.1%, 6.6%, 10.0% and 4.8%, respectively) (Fig. 33), while it has no effect in the expression level in the remaining two genes, *GCV3* and *TEF1* (Fig. 32).

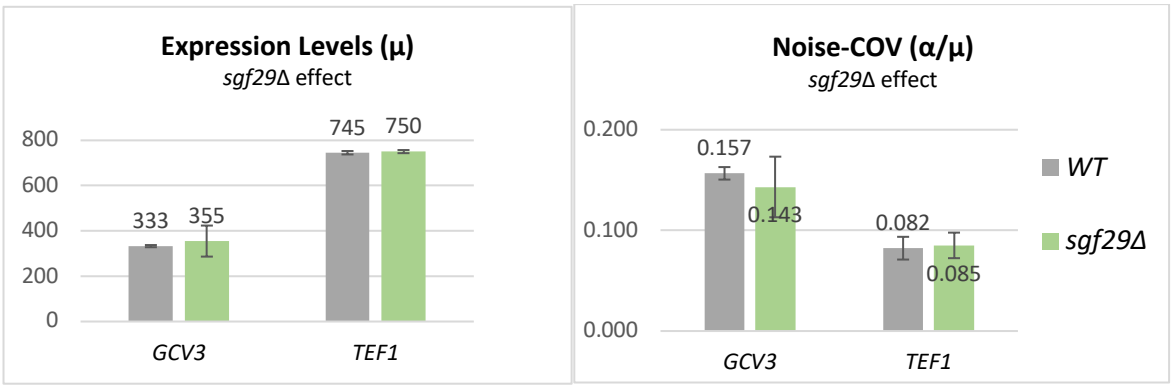


Figure 32. No change in either expression or noise levels in loss of *sgf29Δ*.

Error bars= standard deviation between experiments (N=3).

At genes where expression levels were decreased, standard deviation is affected *ARG1,* although COV (noise) is increased for all 4 genes (Fig. 33). Noise level was increased by 7.3%, 8.1%, 16.3% and 19.2% at *URA4, SML1, SWD1* and *ARG1*, respectively (Fig. 33). The increase in noise disagrees with the hypothesis put forward in this study. The observed increase in the COV (as a measure of noise) may simply due to just a lower mean expression. That is because there is no change in variation (standard deviation or distance from the mean expression) of gene expression across the population of 10,000 cells except *ARG1* (Fig. 33). Therefore, this result might be due to the strong negative correlation between expression levels and noise (i.e. reduced levels of gene expression is associated with increased levels of noise), and that the change in overall expression levels is the most likely explanation for the observed changes in standard deviation and COV (noise) at 4 affected genes.

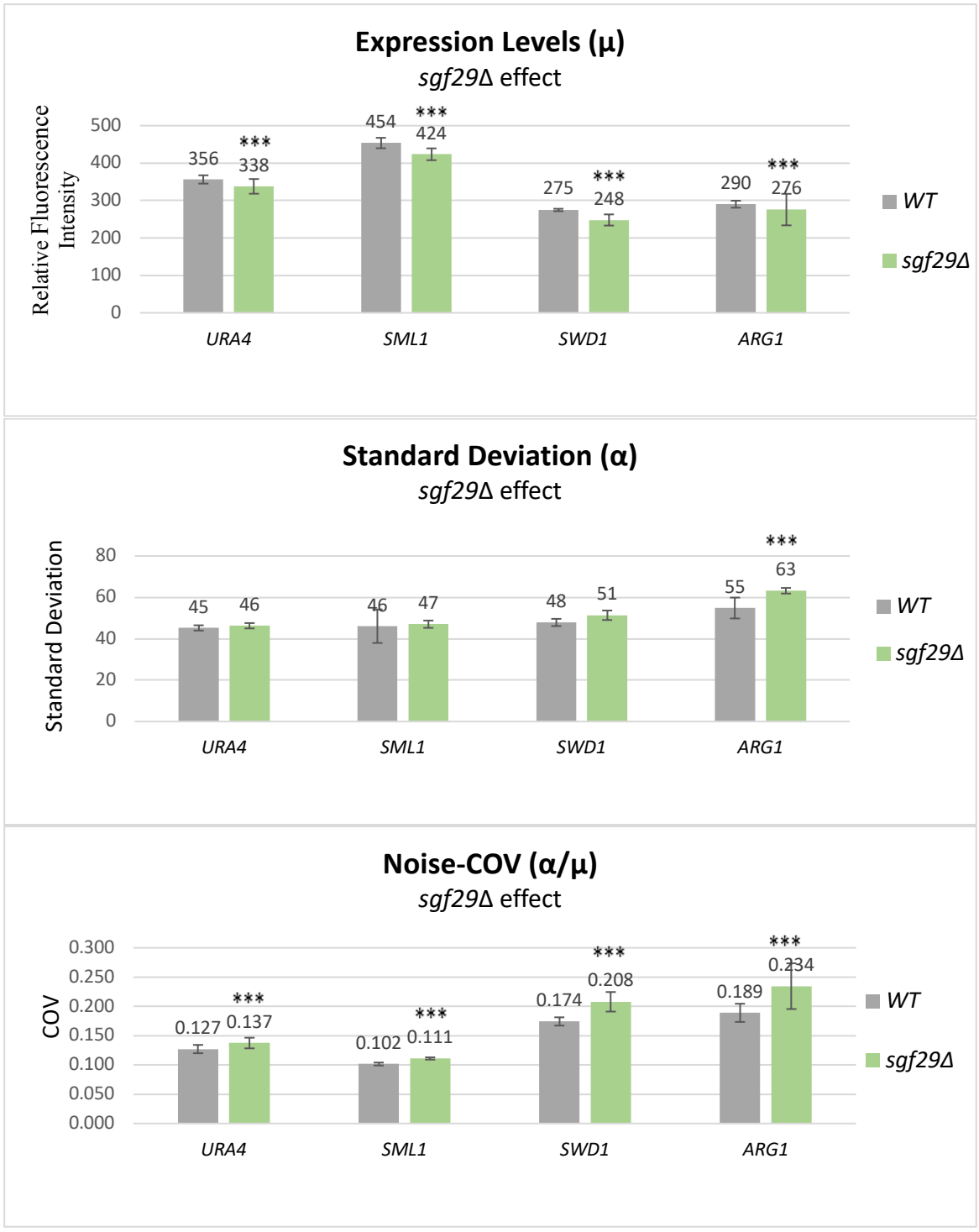


Figure 33. A decrease in expression level and an increase in noise upon the loss of *SGF29* in four genes.

***, P < 0.001. Error bars= standard deviation between experiments (N=3).

5.3.2. Decrease in H3K9ac upon the loss of Spp1

H3K4me3 (which is previously hypothesized to suppress noise) can promote H3K9ac through binding to Sgf29 (a SAGA subunit) as this interaction is required for maximal acetylation of H3K9ac. We tested how the loss of H3K4me3 could affect H3K9ac. The different effects of the loss of Spp1 or Set1 on H3K9ac were validated by chromatin immunoprecipitation (ChIP)-qPCR at the promoter region of 4 genes (*URA4, SML1, GCV3* and *BAT2*). First, since Sgf29 can stimulate H3K9ac levels, the loss of *SFG29* would remove H3K9ac in individual genes. As depicted in Fig. 34, a significant reduction in H3K9ac levels were observed in the loss of Sgf29. The absence of H3K9ac in the *sgf29Δ* strain indicates that the antibodies raised against H3K9ac are specific for a modification on H3K9.

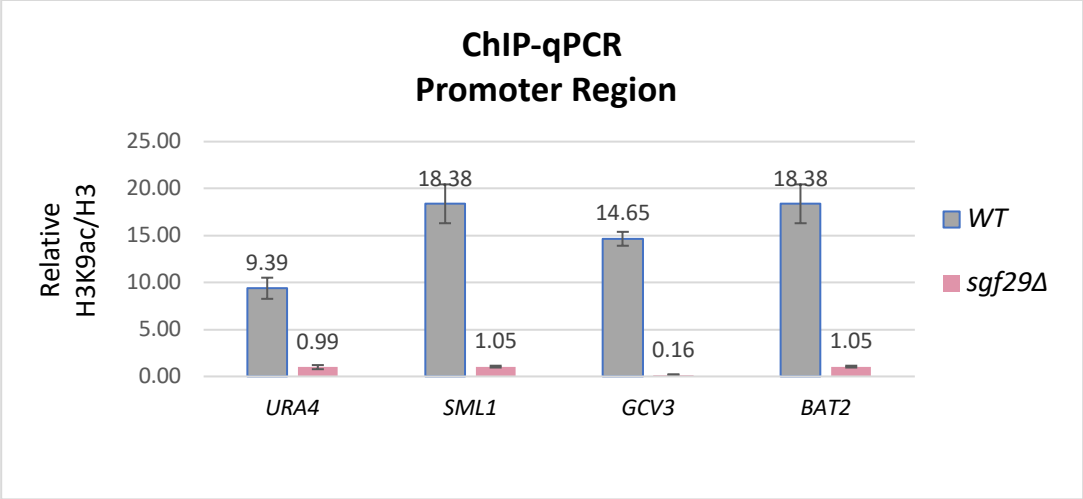


Figure 34. Reduction in H3K9ac upon the loss of Sfg29.

Levels of H3K9ac assessed by chromatin immunoprecipitation (ChIP)-qPCR in the promoter region of *URA4, SML1, GCV3* and *BAT2*. Signals are normalised to levels of H3. Error bars show standard errors of the real-time PCR reaction.

To investigate the H3K9ac-H3K4me3 crosstalk further, whole cell extracts were prepared from yeast strains in which H3K4me3 was reduced by the loss of Spp1 and analysed by ChIP-qPCR for specific levels of H3K9ac. Upon the loss of Spp1, the reduction levels of H3K4me3 at the promoter region for *URA4, SML1* and *GCV3* are 64%, 61% and 67%, respectively as has been shown before (Table 6). As depicted in Fig. 37, levels of H3K9ac at the promoter region of these genes was determined in *spp1*Δ strain, but the effect was not consistent.

Levels of H3K9ac decreases upon the loss of Spp1 (H3K4me3) at the promoter of *SML1* and *GCV3* (51% and 64%, respectively) (Fig. 35). However, *URA4* shows an increase in H3K9ac by 10% in the loss of Spp1 (Fig. 35). H3K4me3 negatively correlates with noise while H3K9ac strongly correlates with transcriptional burst size, which is proportionally correlates with noise. Therefore, H3K4me3 and H3K9ac are proposed to have an opposite effect in noise. This means when the loss of H3K4me3 in *spp1*Δ strain is also associated with loss of H3K9ac a neutral or small effect in noise could be observed. For example, *SML1* and *GCV3* shows a reduction in H3K4me3 by 61% and 67%, respectively, and a reduction in H3K9ac by 51% and 64%, respectively, and they are the only two genes that show a small increase in noise in the loss of Spp1 without change in expression. This small effect in noise might be resulted from the different effects of both H3K9ac and H3K4me3 on noise.

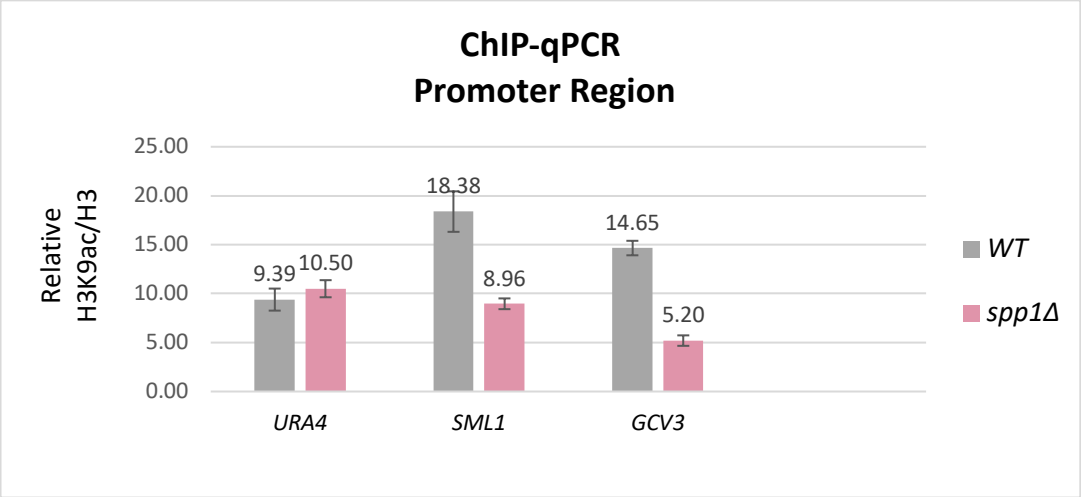


Figure 35. Decrease in H3K9ac upon the loss of *SPP1* in certain genes.

Levels of H3K9ac assessed by chromatin immunoprecipitation (ChIP)-qPCR in the promoter region of *URA4, SML1* and *GCV3*. Signals are normalised to levels of H3. Error bars show standard errors of the real- time PCR reaction.

Next, we wanted to examine the effect of Set1 on H3K9ac. Loss of Set1 remove all type of methylation on H3K4. Interestingly, unlike Spp1, the loss of Set1 increases levels of H3K9ac at the 4 tested genes (*URA4, SML1, GCV3* and *BAT2* by 19%, 36%, 40% and 36%) (Fig. 36). H3K4me2 is absent in the loss of Set1, thus the increase in H3K9ac might be due to the absent of Set3, histone deacetylase, that can be recruited by H3K4me2 (Kim and Buratowski, 2009).

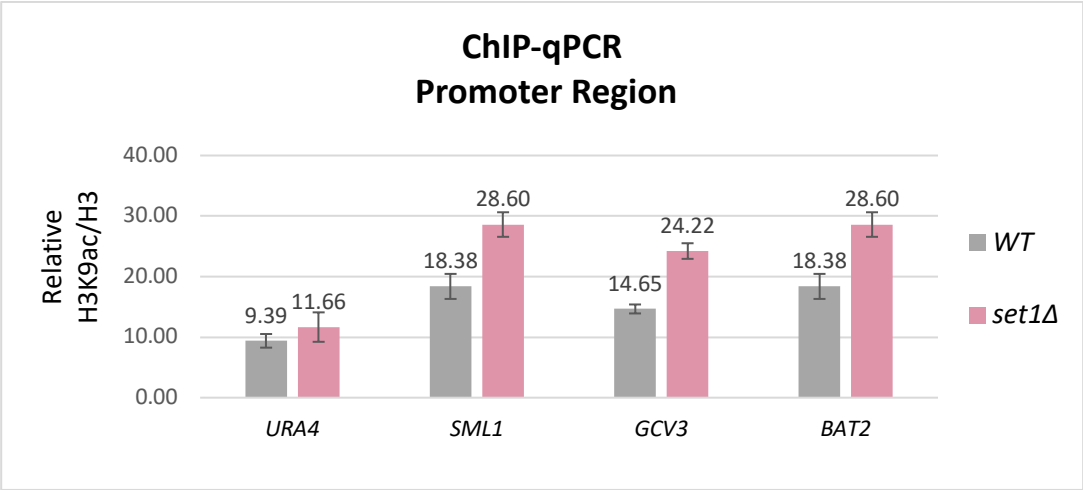


Figure 36. Increase in H3K9ac upon the loss of Set1.

Levels of H3K9ac assessed by chromatin immunoprecipitation (ChIP)-qPCR in the promoter region of *URA4, SML1, GCV3* and *BAT2*. Signals are normalised to levels of H3. Error bars show standard errors of the real-time PCR reaction.

In conclusion, loss of Sgf29 decreases levels of H3K9ac, which is hypothesized to enhance noise. At 4 out of 6 tested genes, loss of Sgf29 mainly affects gene expression due to the absence of acetylation. Because there was no observed change in the standard deviation of these 4 genes, the change in noise is mainly resulted from the change in expression levels. Furthermore, genes that were slightly affected by the loss of Spp1 show a reduction in H3K9ac, which is hypothesized to enhance noise.

5.4. Discussion

Previous work has demonstrated a strong correlation between the intensities of H3K9ac and transcriptional burst size, which is proportionally corelated with noise (Wu et al., 2017). The aim of the current project is to further investigate the causal relationship between H3K9ac (promoted by Sgf29 that binds to H3K4me3) and noise. This study postulates that H3K9ac (Sgf29) could affect both expression and noise levels but not the standard deviation.

Loss of Sgf29 (H3K9ac) affects gene expression at four out of the six genes analysed in this chapter. It decreases expression levels at *URA4*, *SML1*, *ARG1* and *SWD1*. This result can be expected since H3K9ac at the +1 nucleosome has been reported to have a strong correlation with gene expression (Lie et al., 2005; Robyr et al., 2002; Weinberger et al., 2012). In the loss of Sgf29 (H3K9ac), the change in expression level might be caused by the fact that DNA nucleosome accessibility, nucleosome instability and paused Pol II occupancy are affected in the loss of acetylation, specifically H3K9ac (Anderson et al., 2001; Widlund et al., 2000; Jin et al., 2010; Foulds et al., 2013). Because there is a strong negative correlation between noise and gene expression (i.e. a decrease in expression level is associated with an increase in noise level), therefore, the observed change in noise in *sgf29*Δ, for the four tested genes, in this study is more likely to result from the change in expression level and extrinsic noise. As mentioned before, cell sizes and cell cycle status are also major contributors of extrinsic noise (Becskei *et al.,* 2005).

Loss of Spp1 not only decreases H3K4me3 but it is associated with a build-up of H3K4me2 and H3K79me3 for some genes (as has been shown in previous two chapters). Furthermore, H3K9ac and H3K4me3 are present at active promoters (Lie et al., 2005; Robyr et al., 2002; Weinberger et al., 2012; Santos-Rosa et al. 2002; Bernstein et al. 2002; Liu et al. 2005; Pokholok et al. 2005). At the promoter region, loss of Spp1 would also alter H3K9ac levels at the promoter region. For example, the loss of H3K4me3 by *SPP1* deletion leads to the loss of H3K9ac at *SML1* and *GCV3* at the

promoter region. Similar result has been also reported at *GAL1* (Bian *et al.,* 2011). However, at *URA4*, loss of Spp1 increases H3K9ac by 10%. H3K4me3 is hypothesized to suppress noise whilst H3K9ac is hypothesized to enhance noise. Therefore, a neutral effect in noise would be raised in the loss of Spp1. Interestingly, unlike the loss of Spp1, levels of H3K9ac are increased in absent of Set1 (H3K4me) for the 4 analysed gene in this study. As the loss of Set1 is associated with loss of H3K4me2. The increase in H3K9ac could be due to the absent of Set3, histone deacetylase, that can be recruited by H3K4me2 (Kim and Buratowski, 2009).

In conclusion, this result has given some insights into the effect of the loss of Spp1 or Set1 in H3K9ac. In their loss, the observed changes in either noise or expression is not only associated with H3K4me, but it could be also associated with H3K9ac. Further work is needed to determine possible crosstalk between histone modifications that can occur in the certain genes that are sensitive to the loss of Spp1 or Set1.

6. Discussion

The work in this thesis has aimed to increase the knowledge about the factors, specifically histone modification, that can play a role in gene expression noise. In this work we investigated the causal relationship of some of the histone modifications including H3K4me, H3K79me and H3K9ac with noise. Furthermore, the effect of the crosstalk, which is the impact of one histone modification on the modification that is nearby or distant, was also investigated. In particular, the crosstalk that could affect trimethylation of H3K4 function in noise has been investigated because this modification strongly correlates with noise. It was hoped that by increasing the understanding about how noise is regulated by histone modification, it would shed some light on the function of these modifications in gene expression.

The work in Chapter 3 characterised correlation between H3K4me and noise. Building on this, in Chapter 4 the link between H3K79me (Dot1) and H3K4me3 (Spp1) and its function in noise was investigated, and a novel redundant function between Dot1 and Spp1 was identified. Interestingly, the level of H3K79me3 at the promoter of some individual genes with different noise values were differentially increased by the loss of Spp1 (H3K4me3). Furthermore, the increase in noise and variation in the deletion of both *DOT1* (H3K79me) and *SPP1* (H3K4me3) together but not in their single deletion suggests that these two modifications may supress noise by an overlapping mechanism along with their redundant function. Lastly, in Chapter 5, the effect of H4K9ac (Sgf29) in noise was characterised, and the link between H3K9ac and H3K4me3 (Spp1) was probed.

H3K4 methylation at promoters may serve to prevent spreading of marks such on H3K79 methylation, normally continued to the gene-body. Spreading into the promoter may serve to modulate levels of noise both positively and negatively, as observed. It is clear the determination of noise is complex and highly interrelated. Methylation rather than a particular methylated residue may suppress/enhance noise.

6.1. Does H3K36me3 redundantly function with H3K4me3 and H3K79me3?

As discussed in Chapter 4, during their initial characterisation of the correlation between H3K4me3 and transcriptional bursting parameters, Wu et al., (2017) did not detect any stronger correlation of H3K4me3 intensities with either transcriptional burst frequency or transcriptional burst size. This result would explain why most of the genes are not affected in the loss of Spp1 (H3K4me3). On the other hand, two genes (out of the 18 genes) that showed an increase in noise in the loss of Spp1 are in line with the reported strong negative correlation between H3K4me3 and noise (Murray, unpublished) and with the hypothesis that states H3K4me3 supresses noise. These two genes, *SML1* and *GCV3*, show an increase in noise in absence of H3K4me3 (Spp1) without changing expression. Along with this, we found a redundant function between Dot1 (H3K79me) and Spp1 (H3K4me3), which could also explain the limitation of H3K4me3 (Spp1) effect in noise. Interestingly, the effect in noise of this redundant function was also limited to certain genes. Therefore, it is possible that other examples of the redundant function between histone modifications would also play a role in noise and exhibit gene-specific effects.

H3K36me3 has been shown to be negatively correlated with noise as H3K79me3 and H3K4me3 do (Newman et al., 2006; S. Murray, unpublished). Therefore, all these three modifications are hypothesized to suppress noise. Furthermore, Wu et al., 2017 reported that, like H3K79me3, H3K36me3 exhibits a stronger correlation with transcriptional burst frequency than it does with transcriptional burst size. As mentioned before, transcriptional burst frequency is shown to inversely correlate with noise (i.e. an increase in transcriptional burst frequency is associated with a decrease in noise level).

As mentioned before, H3K36me3 plays a role in transcription elongation by modulating chromatin at transcribed DNA region (Lee et al., 2005), and it is also found to be associated with active transcription (Chantalat et al., 2011). In several studies, positive feedback loop between transcription

initiation and elongation has been proposed. For example, Chen and Zhang reported that, in GFP tagged genes, the intensity of H3K79me3 is still associated with low noise levels in the complete knock-out of the open reading frame of these genes. Along with the fact that transcriptional initiation is most likely to determine transcriptional bursting parameters (Gerber & Shilatifard, 2003; Nguyen & Zhang, 2011; reviewed by Nicolas et al., 2017), the possible presence of a positive feedback loop between transcription initiation and elongation may control both transcriptional bursting parameters and noise. Furthermore, Kozhina et al., (2016) has found that both Dot1 and Set2 has been involved in controlling similar pathways of DNA repair. Based on all these facts, we proposed that Set2 (H3K36me3), could be involved in the redundant function of Dot1 (H3K79me) and Spp1 (H3K4me3).

6.2. Future work

The hypotheses proposed in this study could be strengthened by further experiments. For example, in this work, the noise measurements could be mainly generated from extrinsic noise. As mentioned in the introduction, extrinsic noise affects multiple genes' expression and originates from extrinsic global cellular factors and machineries (Elowitz et al., 2002; Raj & Van Oudenaarden, 2008). Intrinsic noise is the primary cause of noise in eukaryotes but in prokaryotes on the other hand, extrinsic noise is the main cause (Ko et al., 1990; Golding et al., 2005) and this difference may result from chromatin remodelling in eukaryotes (Raj & Van Oudenaarden, 2008). Further work can be focused in excluding extrinsic noise, by tagging targeted genes with two fluorescence proteins. The two fluorescent proteins can be incorporated into two alleles of the same endogenous gene. Controlled by identical promoters, this can be used to determine intrinsic noise. The fluorescence difference between the two fluorescent proteins can only contribute to intrinsic noise because, in the same intracellular environment, two fluorescent proteins are expressed from two identical copies of a specific gene (Elowits et al., 2002; Wu et al., 2017).

Moreover, Howe et al., (2014) found that levels of H3K4me3 are reduced in H3K14A mutants, whereas H3K14R, which was used as a control, does not affect H3K4me3 levels. H3K4me3 levels were reduced to 40% of the WT levels in H3K14Q mutants, whereas levels were reduced to 20% of the WT levels in H3K14A mutants. H3K14A mutants do not affect H3K4me1 or H3K4me2 levels, thus these could be used to test the effect of gradually change H3K4 tri-methylation levels on noise. These strains would provide insight on whether H3K4me3 acts gradually in effecting gene expression noise or not.

Furthermore, by using RNA-FISH (Fluorescence *in situ* hybridisation), future experiments can also focus in measuring noise on the RNA-level after the loss of both Dot1-Spp1. This could give an insight into how histone modifications are controlling transcriptional bursting parameters affected by that protein noise and determine whether the observed changes in protein noise is a result of transcriptional regulation. If similar observation would be observed at the RNA-level in the loss of both Dot1-Spp1, this would support the current study and indicates that transcription is an important step in regulating noise (Sherman et al., 2015).

6.3. Concluding remarks

In summary, the work in this thesis has characterised several known correlations between histone modifications and noise. This has increased the understanding about the regulation of noise by histone modification. Since the redundant function of H3K79me3 (Dot1) and H3K4me3 (Spp1) has been a novel finding in this project, it is unknown whether other examples between histone modifications would also play a role in noise and exhibit gene-specific effects. This would add a layer of complexity onto the role of histone modification in gene expression and noise if they do.

7. Appendix

Table 10. Plasmids used in this study

Plasmid Name	Features / Purpose of Use	Citation
pFA6a-GFP(S65T)-kanMX6	Source of GFP-cassette with KAN selective marker.	pFA6a-GFP(S65T)-kanMX6 was a gift from Jurg Bahler & John Pringle (Addgene plasmid # 39292 ; http://n2t.net/addgene:39292 ; RRID:Addgene_39292).
pFA6a-GFP(S65T)-HIS3MX6	Source of GFP-cassette with HIS selective marker.	pFA6a-GFP(S65T)-His3MX6 was a gift from John Pringle (Addgene plasmid # 41598 ; http://n2t.net/addgene:41598 ; RRID:Addgene_41598)

8. Bibliography

Bahler J, Wu JQ, Longtine MS, Shah NG, McKenzie A 3rd, Steever AB, Wach, A, Philippsen P, Pringle JR. 1998. Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast,* 14(10):943-51.

Berger SL. 2007. The complex language of chromatin regulation during transcription. *Nature* **447**: 407–12.

Blake, W.J., Kærn, M., Cantor, C.R. and Collins, J.J., 2003. Noise in eukaryotic gene expression. *Nature*, *422*(6932), p.633.

Brenner, N., Newman, C.M., Osmanović, D., Rabin, Y., Salman, H. and Stein, D.L., 2015. Universal protein distributions in a model of cell growth and division. *Physical Review E*, *92*(4), p.042713.

Brenner, N., Braun, E., Yoney, A., Susman, L., Rotella, J. and Salman, H., 2015. Single-cell protein dynamics reproduce universal fluctuations in cell populations. *The European Physical Journal E*, *38*(9), p.102.

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y. and Barkai, N., 2006. Noise in protein expression scales with natural protein abundance. *Nature genetics*, *38*(6), p.636.

Becskei, A., Kaufmann, B.B. and van Oudenaarden, A., 2005. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nature genetics*, *37*(9), pp.937-944.

Bian, C., Xu, C., Ruan, J., Lee, K.K., Burke, T.L., Tempel, W., Barsyte, D., Li, J., Wu, M., Zhou, B.O. and Fleharty, B.E., 2011. Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *The EMBO journal*, *30*(14), pp.2829-2842.

Blake, W.J., Kærn, M., Cantor, C.R. and Collins, J.J., 2003. Noise in eukaryotic gene expression. *Nature*, *422*(6932), pp.633-637.

Boeger, H., Griesenbeck, J. and Kornberg, R.D., 2008. Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. *Cell*, *133*(4), pp.716-726.

Braunstein, M., Sobel, R.E., Allis, C.D., Turner, B.M. and Broach, J.R., 1996. Efficient transcriptional silencing in Saccharomyces cerevisiae requires a heterochromatin histone acetylation pattern. *Molecular and Cellular Biology*, *16*(8), pp.4349-4356.

Brock, A., Krause, S. and Ingber, D.E., 2015. Control of cancer formation by intrinsic genetic noise and microenvironmental cues. *Nature Reviews Cancer*, *15*(8), pp.499-509.

Brown, C.R. and Boeger, H., 2014. Nucleosomal promoter variation generates gene expression noise. *Proceedings of the National Academy of Sciences*, *111*(50), pp.17893-17898.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y. and Allis, C.D., 1996. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*, *84*(6), pp.843-851.

Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F. and Allis, C.D., 2001. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes & development*, *15*(24), pp.3286-3295.

Briggs, S.D., Xiao, T., Sun, Z.W., Caldwell, J.A., Shabanowitz, J., Hunt, D.F., Allis, C.D. and Strahl, B.D., 2002. Gene silencing: trans-histone regulatory pathway in chromatin. *Nature*, *418*(6897), pp.498-498.

Cai, L., Friedman, N. and Xie, X.S., 2006. Stochastic protein expression in individual cells at the single molecule level. *Nature*, *440*(7082), p.358.

Casper, S.J. and Holt, C.A., 1996. Expression of the green fluorescent protein-encoding gene from a tobacco mosaic virus-based vector. *Gene*, *173*(1), pp.69-73.

Causton, H.C., Ren, B., Koh, S.S., Harbison, C.T., Kanin, E., Jennings, E.G., Lee, T.I., True, H.L., Lander, E.S. and Young, R.A., 2001. Remodeling of yeast genome expression in response to environmental changes. *Molecular biology of the cell*, *12*(2), pp.323-337.

Chalancon, G., Ravarani, C.N., Balaji, S., Martinez-Arias, A., Aravind, L., Jothi, R. and Babu, M.M., 2012. Interplay between gene expression noise and regulatory network architecture. *Trends in genetics*, *28*(5), pp.221-232.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. and Prasher, D.C., 1994. Green fluorescent protein as a marker for gene expression. *Science*, *263*(5148), pp.802-805.

Chandrasekharan, M.B., Huang, F. and Sun, Z.W., 2010. Histone H2B ubiquitination and beyond: Regulation of nucleosome stability, chromatin dynamics and the trans-histone H3 methylation. *Epigenetics*, *5*(6), pp.460-468.

Chantalat, S., Depaux, A., Héry, P., Barral, S., Thuret, J.Y., Dimitrov, S. and Gérard, M., 2011. Histone H3 trimethylation at lysine 36 is associated with constitutive and facultative heterochromatin. *Genome research*.

Choi, J.K. and Kim, Y.J., 2009. Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nature genetics*, *41*(4), p.498.

Coburn, C., Allman, E., Mahanti, P., Benedetto, A., Cabreiro, F., Pincus, Z., Matthijssens, F., Araiz, C., Mandel, A., Vlachos, M. and Edwards, S.A., 2013. Anthranilate fluorescence marks a calcium-propagated necrotic wave that promotes organismal death in C. elegans. *PLoS biology*, *11*(7), p.e1001613.

Corda, Y., Schramke, V., Longhese, M.P., Smokvina, T., Paciotti, V., Brevet, V., Gilson, E. and Géli, V., 1999. Interaction between Set1p and checkpoint protein Mec3p in DNA repair and telomere functions. *Nature genetics*, *21*(2), p.204.

Cubitt, A.B., Heim, R., Adams, S.R., Boyd, A.E., Gross, L.A. and Tsien, R.Y., 1995. Understanding, improving and using green fluorescent proteins. *Trends in biochemical sciences*, *20*(11), pp.448-455.

Dacheux, E., Malys, N., Meng, X., Ramachandran, V., Mendes, P. and McCarthy, J.E., 2017. Translation initiation events on structured eukaryotic mRNAs generate gene expression noise. *Nucleic acids research*, *45*(11), pp.6981-6992.

Davenport, D. and Nicol, J.C., 1955. Luminescence in hydromedusae. *Proc. R. Soc. Lond. B*, *144*(916), pp.399-411.

de Sousa Abreu, R., Penalva, L.O., Marcotte, E.M. and Vogel, C., 2009. Global signatures of protein and mRNA expression levels. *Molecular BioSystems*, *5*(12), pp.1512-1526.

Dover, J., Schneider, J., Tawiah-Boateng, M.A., Wood, A., Dean, K., Johnston, M. and Shilatifard, A., 2002. Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. *Journal of Biological Chemistry*, *277*(32), pp.28368-28371.

Duveau, F., Toubiana, W. and Wittkopp, P.J., 2017. Fitness effects of cis-regulatory variants in the Saccharomyces cerevisiae TDH3 promoter. *Molecular biology and evolution*, *34*(11), pp.2908-2912.

Elowitz, M.B., Levine, A.J., Siggia, E.D. and Swain, P.S., 2002. Stochastic gene expression in a single cell. *Science*, *297*(5584), pp.1183-1186.

Fraser, D. and Kaern, M., 2009. A chance at survival: gene expression noise and phenotypic diversification strategies. *Molecular microbiology*, *71*(6), pp.1333-1340.

Foulds, C.E., Feng, Q., Ding, C., Bailey, S., Hunsaker, T.L., Malovannaya, A., Hamilton, R.A., Gates, L.A., Zhang, Z., Li, C. and Chan, D., 2013. Proteomic analysis of coregulators bound to ERα on DNA and nucleosomes reveals coregulator dynamics. *Molecular cell*, *51*(2), pp.185-199.

Fraser, H.B., Hirsh, A.E., Giaever, G., Kumm, J. and Eisen, M.B., 2004. Noise minimization in eukaryotic gene expression. *PLoS biology*, *2*(6), p.e137.

Gardner, R.G., Nelson, Z.W. and Gottschling, D.E., 2005. Ubp10/Dot4p regulates the persistence of ubiquitinated histone H2B: distinct roles in telomeric silencing and general chromatin. *Molecular and cellular biology*, *25*(14), pp.6123-6139.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D. and Brown, P.O., 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Molecular biology of the cell*, *11*(12), pp.4241-4257.

Gates, L.A., Shi, J., Rohira, A.D., Feng, Q., Zhu, B., Bedford, M.T., Sagum, C.A., Jung, S.Y., Qin, J., Tsai, M.J. and Tsai, S.Y., 2017. Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *Journal of Biological Chemistry*, *292*(35), pp.14456-14472.

Gerber, M. and Shilatifard, A., 2003. Transcriptional elongation control and histone methylation. *Journal of Biological Chemistry*.

Golding, I., Paulsson, J., Zawilski, S.M. and Cox, E.C., 2005. Real-time kinetics of gene activity in individual bacteria. *Cell*, *123*(6), pp.1025-1036.

Grunstein, M., 1998. Yeast heterochromatin: regulation of its assembly and inheritance by histones. *Cell*, *93*(3), pp.325-328.

Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C. and Lander, E.S., 2011. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*, *146*(4), pp.633-644.

Hahn, S. and Young, E.T., 2011. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics*, *189*(3), pp.705-736.

Henry, K.W., Wyce, A., Lo, W.S., Duggan, L.J., Emre, N.T., Kao, C.F., Pillus, L., Shilatifard, A., Osley, M.A. and Berger, S.L., 2003. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes & development*, *17*(21), pp.2648-2663.

Hermann, G.J., Schroeder, L.K., Hieb, C.A., Kershner, A.M., Rabbitts, B.M., Fonarev, P., Grant, B.D. and Priess, J.R., 2005. Genetic analysis of lysosomal trafficking in Caenorhabditis elegans. *Molecular biology of the cell*, *16*(7), pp.3273-3288.

Hooshangi, S., Thiberge, S. and Weiss, R., 2005. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proceedings of the National Academy of Sciences*, *102*(10), pp.3581-3586.

Howe, F.S., Boubriak, I., Sale, M.J., Nair, A., Clynes, D., Grijzenhout, A., Murray, S.C., Woloszczuk, R. and Mellor, J., 2014. Lysine acetylation controls local protein conformation by influencing proline isomerization. *Molecular cell*, *55*(5), pp.733-744.

Howe, F. S., Fischl, H., Murray, S. C., & Mellor, J. 2016. Is H3K4me3 instructive for transcription activation? *BioEssays*, *1600095*(m), 1–12.

Huh, D. and Paulsson, J., 2011. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nature genetics*, *43*(2), p.95.

Ismail, M.B., Shinohara, M. and Shinohara, A., 2014. Dot1-dependent histone H3K79 methylation promotes the formation of meiotic double-strand breaks in the absence of histone H3K4 methylation in budding yeast. *PLoS One*, *9*(5), p.e96648.

Jansen, A. and Verstrepen, K.J., 2011. Nucleosome positioning in Saccharomyces cerevisiae. *Microbiology and Molecular Biology Reviews*, *75*(2), pp.301-320.

Jin, Q., Yu, L.R., Wang, L., Zhang, Z., Kasper, L.H., Lee, J.E., Wang, C., Brindle, P.K., Dent, S.Y. and Ge, K., 2011. Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *The EMBO journal*, *30*(2), pp.249-262.

Kærn, M., Elston, T.C., Blake, W.J. and Collins, J.J., 2005. Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, *6*(6), pp.451-464.

Kim, T. and Buratowski, S., 2009. Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5′ transcribed regions. *Cell*, *137*(2), pp.259-272.

Ko, M.S., Nakauchi, H. and Takahashi, N., 1990. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *The EMBO journal*, *9*(9), p.2835.

Kouzarides, T., 2007. Chromatin modifications and their function. *Cell*, *128*(4), pp.693-705.

Kozhina, T.N., Evstiukhina, T.A., Peshekhonov, V.T., Chernenkov, A.Y. and Korolev, V.G., 2016. Dot1 and Set2 histone methylases control the spontaneous and UV-induced mutagenesis levels in the Saccharomyces cerevisiae yeasts. *Russian journal of genetics*, *52*(3), pp.263-272.

Le Masson, I., David, Y.Y., Jensen, K., Chevalier, A., Courbeyrette, R., Boulard, Y., Smith, M.M. and Mann, C., 2003. Yaf9, a novel NuA4 histone acetyltransferase subunit, is required for the cellular response to spindle stress in yeast. *Molecular and cellular biology*, *23*(17), pp.6086-6102.

Lehner, B., 2008. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Molecular systems biology*, *4*(1), p.170.

Leung, A., Cajigas, I., Jia, P., Ezhkova, E., Brickner, J.H., Zhao, Z., Geng, F. and Tansey, W.P., 2011. Histone H2B ubiquitylation and H3 lysine 4 methylation prevent ectopic silencing of euchromatic loci important for the cellular response to heat. *Molecular biology of the cell*, *22*(15), pp.2741-2753.

Li, Y., Wen, H., Xi, Y., Tanaka, K., Wang, H., Peng, D., Ren, Y., Jin, Q., Dent, S.Y., Li, W. and Li, H., 2014. AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell*, *159*(3), pp.558-571.

Longtine, M.S., McKenzie III, A., Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P. and Pringle, J.R., 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast*, *14*(10), pp.953-961.

Ludin, B. and Matus, A., 1998. GFP illuminates the cytoskeleton. *Trends in cell biology*, *8*(2), pp.72-77.

Ludin, B., Doll, T., Meili, R., Kaech, S. and Matus, A., 1996. Application of novel vectors for GFP-tagging of proteins to study microtubule-associated proteins. *Gene*, *173*(1), pp.107-111.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), pp.251-260.

Raser, J.M. and O'shea, E.K., 2004. Control of stochasticity in eukaryotic gene expression. *Science*, *304*(5678), pp.1811-1814.

Ly, T., Ahmad, Y., Shlien, A., Soroka, D., Mills, A., Emanuele, M.J., Stratton, M.R. and Lamond, A.I., 2014. A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *Elife*, *3*, p.e01630.

Maheshri, N. and O'Shea, E.K., 2007. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annual review of biophysics and biomolecular structure*, *36*.

Margaritis, T., Oreal, V., Brabers, N., Maestroni, L., Vitaliano-Prunier, A., Benschop, J.J., Van Hooff, S., Van Leenen, D., Dargemont, C., Geli, V. and Holstege, F.C., 2012. Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3′-end antisense transcription. *PLoS genetics*, *8*(9), p.e1002952.

Moazed, D., 2001. Common themes in mechanisms of gene silencing. *Molecular cell*, *8*(3), pp.489-498.

Morillon, A., Karabetsou, N., Nair, A. and Mellor, J., 2005. Dynamic lysine methylation on histone H3 defines the regulatory phase of gene transcription. *Molecular cell*, *18*(6), pp.723-734.

Moss, J.B., Price, A.L., Raz, E., Driever, W. and Rosenthal, N., 1996. Green fluorescent protein marks skeletal muscle in murine cell lines and zebrafish. *Gene*, *173*(1), pp.89-98.

Munsky, B., Neuert, G. and Van Oudenaarden, A., 2012. Using gene expression noise to understand gene regulation. *Science*, *336*(6078), pp.183-187.

Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L. and Weissman, J.S., 2006. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, *441*(7095), p.840.

Ng, H.H., Dole, S. and Struhl, K., 2003. The Rtf1 component of the Paf1 transcriptional elongation complex is required for ubiquitination of histone H2B. *Journal of Biological Chemistry*, *278*(36), pp.33625-33628.

Nicolas, D., Phillips, N.E. and Naef, F., 2017. What shapes eukaryotic transcriptional bursting?. *Molecular BioSystems*, *13*(7), pp.1280-1290.

Nislow, C., Ray, E. and Pillus, L., 1997. SET1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes. *Molecular biology of the cell*, *8*(12), pp.2421-2436.
Niedenthal, R.K., Riles, L., Johnston, M. and Hegemann, J.H., 1996. Green fluorescent protein as a marker for gene expression and subcellular localization in budding yeast. *Yeast*, *12*(8), pp.773-786.

Novick, A. and Weiner, M., 1957. Enzyme induction as an all-or-none phenomenon. *Proceedings of the National Academy of Sciences*, *43*(7), pp.553-566.

Nguyen, A.T. and Zhang, Y., 2011. The diverse functions of Dot1 and H3K79 methylation. *Genes & development*, *25*(13), pp.1345-1358.

Osley, M.A., 2004. H2B ubiquitylation: the end is in sight. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, *1677*(1-3), pp.74-78.

Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D. and Van Oudenaarden, A., 2002. Regulation of noise in the expression of a single gene. *Nature genetics*, *31*(1), p.69.

Paulsson, J., 2004. Summing up the noise in gene networks. *Nature*, *427*(6973), p.415.

Pradet-Balade, B., Boulmé, F., Beug, H., Müllner, E.W. and Garcia-Sanz, J.A., 2001. Translation control: bridging the gap between genomics and proteomics?. *Trends in biochemical sciences*, *26*(4), pp.225-229.

Raj, A. and van Oudenaarden, A., 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, *135*(2), pp.216-226.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y. and Tyagi, S., 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS biology*, *4*(10), p.e309.

Raser, J.M. and O'shea, E.K., 2005. Noise in gene expression: origins, consequences, and control. *Science*, *309*(5743), pp.2010-2013.

Ravarani, C.N., Chalancon, G., Breker, M., De Groot, N.S. and Babu, M.M., 2016. Affinity and competition for TBP are molecular determinants of gene expression noise. *Nature communications*, *7*, p.10417.

Rundlett, S.E., Carmen, A.A., Kobayashi, R., Bavykin, S., Turner, B.M. and Grunstein, M., 1996. HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. *Proceedings of the National Academy of Sciences*, *93*(25), pp.14503-14508.

Salari, R., Wojtowicz, D., Zheng, J., Levens, D., Pilpel, Y. and Przytycka, T.M., 2012. Teasing apart translational and transcriptional components of stochastic variations in eukaryotic gene expression. *PLOS computational biology*, *8*(8), p.e1002644.

Sanchez, A. and Golding, I., 2013. Genetic determinants and cellular constraints in noisy gene expression. *Science*, *342*(6163), pp.1188-1193.

Sherman, M.S., Lorenz, K., Lanier, M.H. and Cohen, B.A., 2015. Cell-to-cell variability in the propensity to transcribe explains correlated fluctuations in gene expression. *Cell systems*, *1*(5), pp.315-325.

Shi, X., Kachirskaia, I., Walter, K.L., Kuo, J.H.A., Lake, A., Davrazou, F., Chan, S.M., Martin, D.G., Fingerman, I.M., Briggs, S.D. and Howe, L., 2007. Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *Journal of Biological Chemistry*, *282*(4), pp.2450-2455.

Shilatifard, A., 2006. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu. Rev. Biochem.*, *75*, pp.243-269.

Singh, A. and Weinberger, L.S., 2009. Stochastic gene expression as a molecular switch for viral latency. *Current opinion in microbiology*, *12*(4), pp.460-466.

Soares, L.M., He, P.C., Chun, Y., Suh, H., Kim, T. and Buratowski, S., 2017. Determinants of Histone H3K4 Methylation Patterns. *Molecular cell*, *68*(4), pp.773-785.

Sommermeyer, V., Béneut, C., Chaplais, E., Serrentino, M.E. and Borde, V., 2013. Spp1, a member of the Set1 Complex, promotes meiotic DSB formation in promoters by tethering histone H3K4 methylation sites to chromosome axes. *Molecular cell*, *49*(1), pp.43-54.

Spedale, G., Timmers, H.T.M. and Pijnappel, W.P., 2012. ATAC-king the complexity of SAGA during evolution. *Genes & development*, *26*(6), pp.527-541.

Strahl BD, and Allis CD. 2000. The language of covalent histone modifications. *Nature* **403**: 41–45.

Suka, N., Suka, Y., Carmen, A.A., Wu, J. and Grunstein, M., 2001. Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. *Molecular cell*, *8*(2), pp.473-479.

Swain, P.S., Elowitz, M.B. and Siggia, E.D., 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, *99*(20), pp.12795-12800.

Takahashi, Y.H., Westfield, G.H., Oleskie, A.N., Trievel, R.C., Shilatifard, A. and Skiniotis, G., 2011. Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. *Proceedings of the National Academy of Sciences*, *108*(51), pp.20526-20531.

Thattai, M. and Van Oudenaarden, A., 2001. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, *98*(15), pp.8614-8619.

Turner BM. 2000. Histone acetylation and an epigenetic code. *BioEssays* **22**: 836–45.

Uphoff, S., Lord, N.D., Okumus, B., Potvin-Trottier, L., Sherratt, D.J. and Paulsson, J., 2016. Stochastic activation of a DNA damage response causes cell-to-cell mutation rate variation. *Science*, *351*(6277), pp.1094-1097.

Venkatesh, S. and Workman, J.L., 2015. Histone exchange, chromatin structure and the regulation of transcription. *Nature reviews Molecular cell biology*, *16*(3), pp.178-189.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G. and Mann, M., 2010. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell*, *142*(6), pp.967-980.

Wang, S. and Hazelrigg, T., 1994. Implications for bcd mRNA localization from spatial distribution of exu protein in Drosophila oogenesis. *Nature*, *369*(6479), p.400.

Weake, V.M. and Workman, J.L., 2008. Histone ubiquitination: triggering gene activity. *Molecular cell*, *29*(6), pp.653-663.

Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P. and Schaffer, D.V., 2005. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell*, *122*(2), pp.169-182.

Weinberger, L., Voichek, Y., Tirosh, I., Hornung, G., Amit, I. and Barkai, N., 2012. Expression noise and acetylation profiles distinguish HDAC functions. *Molecular cell*, *47*(2), pp.193-202.

Wood, A., Krogan, N.J., Dover, J., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Golshani, A., Zhang, Y., Greenblatt, J.F. and Johnston, M., 2003. Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. *Molecular cell*, *11*(1), pp.267-274.

Wu, S., Li, K., Li, Y., Zhao, T., Li, T., Yang, Y.F. and Qian, W., 2017. Independent regulation of gene expression level and noise by histone modifications. *PLoS computational biology*, *13*(6), p.e1005585.

Zenklusen, D., Larson, D.R. and Singer, R.H., 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature structural & molecular biology*, *15*(12), p.1263.

Zhang, K., Lin, W., Latham, J.A., Riefler, G.M., Schumacher, J.M., Chan, C., Tatchell, K., Hawke, D.H., Kobayashi, R. and Dent, S.Y., 2005. The Set1 methyltransferase opposes Ipl1 aurora kinase functions in chromosome segregation. *Cell*, *122*(5), pp.72.