



COMPASS Mediates Transgenerational Epigenetic Inheritance in *Caenorhabditis elegans*

By:

Rosamund Clifford

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Leeds

Faculty of Biological Sciences

School of Biology

Submission Date

February 2021

Acknowledgements

There are a great many people who have contributed to the development of this thesis, who I would like to thank here.

Professor Ian Hope, who took on the role of my primary supervisor midway through my PhD and guided me through to the end with constant support and advice during the bulk of the experiments and entire writing process.

My first PhD supervisor, Dr Ron Chen, whose ideas and support in the first two years gave me a solid foundation on which to build this thesis.

Professor Mark Dickman and the Dickman group at the University of Sheffield, in particular Dr Eleanor Hanson, Dr Joby Cole and Dr Caroline Evans, for their technical expertise in LC-MS/MS and support with performing the experiments described here.

Professor Julie Ahringer and the Ahringer group at the University of Cambridge, for their support with the immunofluorescence microscopy experiments, which I learned to do under the supervision of Yan Dong on a visit to the Ahringer lab. Thanks also to the imaging team at the Gurdon Institute, particularly Dr Nicola Lawrence, for training me to use a confocal microscope, and Dr Richard Butler, for writing the code that enabled analysis of the images.

At the University of Leeds, my thanks go to Dr Ruth Hughes at the Bio-imaging facility for her advice on image analysis, and Dr Brittany Graham for helping me settle into the Hope lab.

The MRC DiMeN DTP for funding and training opportunities, in particular Dr Emily Goodall, the DTP manager, for facilitating my extra funding applications that enabled me to learn about commercial mass spectrometry applications on a summer placement, and to visit the Ahringer lab in Cambridge.

My friends and family, for their moral support through the trials and tribulations of the PhD process. Dheemanth, for sharing an office with me while writing this thesis, putting up with frequent interruptions to read sections and cheering me on to the finish line.

Table of Contents

Chapter 1.	Introduction.....	16
1.1	Epigenetics.....	16
1.1.1	Interest and controversy	16
1.1.2	Definition	16
1.1.3	Mechanisms.....	16
1.2	The Histone Code	17
1.2.1	Overview.....	17
1.2.2	The nucleosome core particle	17
1.2.3	Nomenclature.....	18
1.2.4	How histone modifications regulate gene expression	19
1.2.5	Cross-talk	20
1.2.6	Transgenerational epigenetic inheritance.....	21
1.2.7	Methods and limitations of study	24
1.3	Mass spectrometry technologies	24
1.3.1	Principle of mass spectrometry.....	24
1.3.2	Ionisation methods.....	25
1.3.3	Mass analysers and detectors	25
1.3.4	Tandem mass spectrometry	26
1.3.5	The Q-Exactive mass spectrometer	27
1.3.6	Data acquisition methods.....	28
1.4	Histone Proteomics	28
1.4.1	Proteomic approaches.....	28
1.4.2	Preparation of histones for bottom-up MS.....	30
1.4.3	Reverse-phase high-pressure liquid chromatography	33
1.4.4	Data analysis.....	34

1.5	COMPASS.....	36
1.5.1	The complex	36
1.5.2	Set1/SET-2	38
1.5.3	Cfp1/CFP-1.....	42
1.6	Thesis aims	45
Chapter 2.	Materials and Methods	46
2.1	<i>C. elegans</i> maintenance.....	48
2.1.1	Standard growth conditions	48
2.1.2	Synchronization and/or decontamination of strains by bleaching	49
2.2	Single worm PCR genotyping.....	49
2.3	Mass spectrometry.....	51
2.3.1	Histone extraction	51
2.3.2	Propionylation of Histone Samples	57
2.3.3	Loading	58
2.3.4	Data acquisition.....	59
2.3.5	Raw data analysis	59
2.4	RNAi	64
2.4.1	Experimental design	64
2.4.2	Slide preparation	66
2.4.3	Scoring protocol.....	66
2.4.4	Data analysis.....	66
2.4.5	Statistical testing	66
2.5	Lifespan assays	67
2.5.1	Set up.....	67
2.5.2	Maintenance.....	67
2.5.3	Exclusions	67
2.5.4	Data analysis.....	68

2.6	Developmental progression assay.....	69
2.7	Immunofluorescence.....	70
2.7.1	Worm preparation/selection.....	70
2.7.2	Slide preparation	70
2.7.3	Antibody staining.....	71
2.7.4	Antibody validation by dot blot.....	72
2.7.5	Confocal Imaging	72
2.7.6	Image analysis.....	72
Chapter 3.	Developing an LC-MS/MS approach to characterise two <i>C. elegans</i> COMPASS mutants	75
3.1	Abstract	75
3.2	Introduction.....	75
3.3	Experimental design	76
3.4	Results	77
3.4.1	Use of heptafluorobutyric acid instead of conventional trifluoroacetic acid as ion-pairing reagent enables detection of H3K4me2- and H3K4me3-modified peptides	77
3.4.2	H3K4me3 levels are dramatically reduced in <i>cfp-1</i> and <i>set-2</i> mutants.....	78
3.4.3	H3K4me2 levels are reduced by approximately half in <i>cfp-1</i> and <i>set-2</i> mutants.....	79
3.4.4	H3K14ac and H3K27ac levels are significantly elevated in <i>cfp-1</i> and <i>set-2</i> mutants	80
3.4.5	H3K23ac levels are significantly elevated in <i>cfp-1</i> but not <i>set-2</i> mutants	81
3.5	Discussion	83
Chapter 4.	Somatic gene derepression in the <i>cfp-1</i> germline is enhanced by RNAi knockdown of <i>hda-3</i> , and possibly <i>hda-1</i> and <i>sin-3</i>	84
4.1	Abstract	84
4.2	Introduction.....	84
4.2.1	HDAC complexes in <i>C. elegans</i>	84
4.2.2	The <i>unc-119::GFP</i> reporter	85
4.3	Experimental Design.....	86
4.4	Results	87

4.4.1	<i>hda-3</i> RNAi significantly enhanced <i>cfp-1</i> germline gene derepression.....	88
4.4.2	<i>sin-3</i> RNAi may enhance <i>cfp-1</i> germline gene derepression.....	90
4.4.3	<i>hda-1</i> RNAi may enhance <i>unc-119::GFP</i> expression in <i>cfp-1</i> but not <i>set-2</i> mutants	92
4.4.4	The proportion of <i>unc-119::GFP</i> -expressing germlines increases over generations in <i>cfp-1</i> mutants	93
4.5	Discussion	96
Chapter 5. Wild-type descendants of COMPASS mutants inherit the mutant long-lived phenotype, and this is associated with a loss of H3K4me2 and gain of H3K9me3 in the long-lived descendants ⁹⁷		
5.1	Abstract	97
5.2	Introduction.....	97
5.3	Experimental design	99
5.4	Results	101
5.4.1	WT descendants of <i>set-2(bn129)</i> also have increased longevity in the F3 and F4 generations	101
5.4.2	WT descendants of <i>set-2(bn129)</i> are not quite back to WT lifespan by the F5 generation	102
5.4.3	The extent of the longevity extension of WT descendants of <i>set-2(bn129)</i> mutants depends on the generation of the mutant ancestor.....	103
5.4.4	Later generation <i>set-2(bn129)</i> mutants develop slowly compared to an earlier generation	104
5.4.5	Increased longevity in F3 and F4 WT descendants of <i>set-2(bn129)</i> mutants is accompanied by a decrease in H3K4me2 and an increase in H3K9me3	105
5.5	Discussion	106
Chapter 6. Assessing the recovery of histone modifications in the germlines of wild-type descendants of COMPASS mutants using immunofluorescence microscopy		
6.1	Abstract	108
6.2	Introduction.....	108
6.3	Experimental design	108
6.4	Results	112
6.4.1	H3K4me2	112

6.4.2	H3K4me3	117
6.4.3	H3K9me2	119
6.4.4	H3K9me3	121
6.5	Discussion	124
Chapter 7.	Discussion	125
7.1	Summary of results.....	125
7.1.1	Model.....	127
7.1.2	The role of histone modifications in the model	129
7.2	Further work.....	130
Chapter 8.	Supplementary data	132
8.1	Germline mass spectrometry data for <i>set-2(bn129)</i> descendants.....	132
Chapter 9.	Bibliography.....	133

List of figures

Figure number	Figure name	Page number
Figure 1.1	Structure of a nucleosome	18
Figure 1.2	The histone modification naming convention	19
Figure 1.3	Inheritance of parental histones	22
Figure 1.4	The restriction of H3K27me3 to the gamete-of-origin chromosome set	23
Figure 1.5	The universal processes of mass spectrometry	25
Figure 1.6	Comparing the process of standard MS vs MS/MS	26
Figure 1.7	Four of the ways the TKQTAR peptide might fragment in a collision chamber	27
Figure 1.8	The components of the Thermo Scientific Q Exactive Benchtop LC-MS/MS	28
Figure 1.9	Propionylation of a lysine	31
Figure 1.10	Peptide-generation from the H3 N-terminal tail	32
Figure 1.11	Screenshot from the histone proteomic DIA data software package Skyline	35
Figure 1.12	Conservation of COMPASS in eukaryotes	37
Figure 1.13	The reaction mechanisms of KMTs and KDMS	39
Figure 1.14	The gene structure of <i>C. elegans set-2</i>	40
Figure 2.1	The expansion of <i>C. elegans</i> stock	52
Figure 2.2	Coomassie stained SDS-PAGE gel	54
Figure 2.3	The stages of development of a <i>C. elegans</i> embryo	55

Figure 2.4	The crossing and genotyping procedure used to isolate WT and mutant descendants of <i>set-2</i>	56
Figure 2.5	The cuts made to obtain an intact gonad for later histone extraction	57
Figure 2.6	How mixed embryo-derived relative abundance data was organised in Prism	64
Figure 2.7	Experimental design for RNAi experiments with <i>hda-3</i> and <i>sin-3</i>	65
Figure 2.8	Formula for testing the null hypothesis that two genes will not interact in an RNAi experiment	67
Figure 2.9	Germline development from L3 to young adult in the <i>C. elegans</i> hermaphrodite	70
Figure 2.10	The Macro code used to quantify fluorescence in gonad imaging experiments	73
Figure 2.11	The data output of the fluorescence quantifying code	74
Figure 3.1	HFBA but not TFA allows detection of di- and tri-methylated H3K4-modified peptides.	78
Figure 3.2	Relative abundances of H3K4-modified peptides	79
Figure 3.3	Relative abundances of H3K9-, K14-, K27- and K36-modified peptides.	81
Figure 3.4	Relative abundances of H3K18- and K23-modified peptides	82
Figure 4.1	<i>unc-119::GFP</i> -expression in a dissected gonad	87
Figure 4.2	<i>unc-119::GFP</i> expression increases over generations in <i>cfp-1</i> mutants.	95
Figure 5.1	Lifespan extension of <i>set-2(ok952)</i> descendants	98
Figure 5.2	The crossing and genotyping procedure used to isolate WT and mutant descendants of <i>set-2(bn129)</i>	100
Figure 5.3	Example genotyping results for the F2 progeny	101
Figure 5.4	Lifespan extension of WT descendants of <i>set-2(bn129)</i> mutants depends on the generation of the mutant ancestor	104
Figure 5.5	Later generation <i>set-2(bn129)</i> mutants develop more slowly than the early generation	105
Figure 5.6	LC-MS/MS profiling of H3 PTMs in germlines of WT descendants of <i>set-2(bn129)</i>	106

Figure 6.1	Slides as prepared for imaging.	109
Figure 6.2	A comparison of the excitation and emission wavelengths of the Alexafluor 405, 488 and 594 dyes.	110
Figure 6.3	The organisation of the hermaphrodite <i>C. elegans</i> gonad	111
Figure 6.4	Selection of nuclei for fluorescence quantification	111
Figure 6.5	Anti-H3K4me2 immunofluorescence intensity in the germline of descendants of early and late generation <i>set-2(bn129)</i> mutants	116
Figure 6.6	Anti-H3K4me3 immunofluorescence intensity in the germline of descendants of early and late generation <i>set-2(bn129)</i> mutants.	118
Figure 6.7	Anti H3K9me2 immunofluorescence intensity in the germline of descendants of early and late generation <i>set-2(bn129)</i> mutants	120
Figure 6.8	Anti-H3K9me3 immunofluorescence intensity in germline of descendants of early and late generation <i>set-2(bn129)</i> mutants	123
Figure 7.1	In the absence of COMPASS function, H3K4me3 levels fall and H3K14ac and H3K27ac levels rise	125
Figure 7.2	A model for the heritability of transgenerationally accumulated epigenetic damage	127
Figure 7.3	The role of histone modifications in the model for heritability of transgenerationally accumulated chromatin changes	128
Figure 8.1	Relative abundance of di- and trimethylated H3K4 and H3K9 in WT and mutant descendants of <i>set-2(bn129)</i> , and WT controls	131

List of tables

Table number	Table name	Page number
Table 1.1	Types of Proteomics	30
Table 1.2	The increase in mass associated with propionylation, methylation and acetylation modifications	32
Table 1.3	A summary of the <i>ok952</i> and <i>bn129</i> alleles and their effects on the SET-2 protein.	41
Table 2.1	Chemical list with suppliers	46
Table 2.2	Buffer recipes	48
Table 2.3	Cycling conditions used in PCR genotyping	50
Table 2.4	Primers used in PCR genotyping	51
Table 2.5	List of proteoforms searched for in DIA analysis	60
Table 2.6	Correction factors which peptides are divided by to correct for isotopic masking and detection efficiency	62
Table 2.7	Charge states factored into relative abundance calculations for each peptide	63
Table 2.8	Details of clones used in RNAi experiments.	65
Table 2.9	Exclusions from lifespan assays	68
Table 2.10	Antibodies used to detect histone modifications	71
Table 3.1	Mean relative abundance and significance for H3K4-modified peptides	79
Table 3.2	Mean relative abundance for H3K14 and K27 modifications that significantly differed between WT, <i>cfp-1</i> and <i>set-2</i> .	80
Table 3.3	Mean relative abundance for H3K23 modifications that significantly differed between WT, <i>cfp-1</i> and <i>set-2</i> .	82
Table 4.1	<i>unc119::GFP</i> expression scoring data for <i>hda-3</i> RNAi in <i>cfp-1</i> and WT germlines	89

Table 4.2	<i>unc119::GFP</i> expression scoring data for <i>sin-3</i> RNAi in <i>cfp-1</i> and WT germlines	91
Table 4.3	<i>unc119::GFP</i> expression scoring data for <i>hda-1</i> RNAi in <i>cfp-1</i> , <i>set-2</i> and WT germlines	93
Table 4.4	<i>unc-119::GFP</i> expression increases over generations in <i>cfp-1</i> mutants on EV negative control RNAi plates	94
Table 4.5	<i>unc-119::GFP</i> expression is significantly increased in late vs early generation <i>cfp-1</i> mutants grown on NGM plates.	95
Table 5.1	Median lifespan and significance of WT descendant's longevity compared to WT controls	103
Table 5.2	Developmental stages reached by WT, <i>set-2</i> generation 4 and generation 60 mutants after 60 hours at 20°C	105
Table 6.1	Summary of results for all IF experiments for H3K4me2	115
Table 6.2	Summary of results for all IF experiments for H3K4me3	118
Table 6.3	Summary of results for all IF experiments for H3K9me2	120
Table 6.4	Summary of results for all IF experiments for H3K9me3	122

Table of abbreviations

ABC	Ammonium bicarbonate
ac	acetylation
AGC	Automatic gain control
ACN	Acetonitrile
ANOVA	Analysis of variance
C18	Alkyl chain with 18 carbons
CaCl ₂	Calcium chloride
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation followed by sequencing of the DNA fragments
CHO cells	Chinese hamster ovary cells
COMPASS	Complex of Proteins Associated with Set1
CpG	Cytosine preceding guanine
DAPI	4',6-diamidino-2-phenylindole
DDA	Data-dependent acquisition
DIA	Data independent acquisition
DTT	Dithiothreitol (reducing agent)
E(z)	Enhancer of Zeste
ESC	Extra sex combs
ESI	Electrospray ionisation
EV	Empty vector
F1	First generation progeny of a cross
GFP	Green fluorescent protein
H1	Histone 1 (linker)
H2A	Histone 2 isoform A
H2B	Histone 2 isoform B
H ₂ SO ₄	Sulphuric acid
H3	Histone 3
H4	Histone 4
HAT	Histone acetyltransferase
HCl	Hydrochloric acid
HDAC	Histone deacetylase
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HFBA	Heptafluorobutyric acid
HMT	Histone methyltransferase
HOT region	Highly occupied target region
HPLC	High performance liquid chromatography
KCl	Potassium chloride

KD	Knock-down
KQLA	Histone 3-derived peptide containing lysines 18 and 23
L4	<i>C. elegans</i> larval stage 4
LB	Lysogeny broth
LC-MS/MS	Liquid chromatography-coupled tandem mass spectrometry
LF	Loss-of-function
M	Molar
m/z	Mass to charge ratio
M9	Media 9
MALDI	Matrix-assisted laser desorption ionisation
me	methylation
MES	Maternal effect sterile
MgCl ₂	Magnesium chloride
MLL	Mixed lineage leukaemia
Mrt	Mortal germline phenotype
MS/MS	Tandem mass spectrometry
MS1	Mass spectrometer 1
MS2	Mass spectrometer 2
Muv	Multivulval phenotype
NaCl	Sodium chloride
NGM	Nematode growth medium
NH ₄ OH	Ammonium hydroxide
NP-40	Nonyl phenoxyethoxyethanol (detergent)
OP50	<i>E. coli</i> strain grown on NGM plates as food for <i>C. elegans</i>
P	(as a suffix) phosphorylation
PBST	Phosphate buffered saline with 0.1% Tween
PCR	Polymerase chain reaction
PFA	Pentafluoropropionic acid
PHD	Plant homeodomain
PTM	Post-translational modification
RNAi	Ribonucleic acid interference
RP-HPLC	Reverse-phase high performance liquid chromatography
rpm	revolutions per minute
RRM	RNA recognition motif
SDS-PAGE	Sodium dodecyl sulphate-coupled polyacrylamide gel electrophoresis
TCA	Trichloroacetic acid
TFA	Trifluoroacetic acid
TKQTAR	Histone 3-derived peptide containing lysine 4
TOF	Time of flight
Trr	Trithorax-related protein

Trx	Trithorax
UV/TMP	Ultraviolet with trimethyl psoralen mutagenesis
WT	Wild type
YA	Young adult

Chapter 1. Introduction

1.1 Epigenetics

1.1.1 Interest and controversy

In recent years, research in the field of Epigenetics, especially the branch of transgenerational epigenetic inheritance, has boomed. The idea that stress and its effects may be something that you inherit has really captured the public imagination. There is now even discussion in the media of ‘transgenerational justice’ for victims of decisions by ancestors, companies or governments that negatively affected their ‘epigenomes’ and potentially those of their descendants also (Rothstein et al., 2017, Kabasenche and Skinner, 2014).

There has also been a great deal of interest in the development of therapeutics targeting epigenetic mechanisms. For example, inhibitors of histone acetyltransferases and histone deacetylases have shown promise as anti-cancer drugs.

However, there is some controversy in the scientific community about how to accurately define Epigenetics, and when the use of the word may be inappropriate.

1.1.2 Definition

Literally meaning ‘on top of genetics’ the word epigenetics originated from Aristotle’s discussion of ‘epigenesis’, a term he coined to describe differentiation of distinct parts from an undifferentiated egg. Some scientists feel the ‘on top of genetics’ understanding of epigenetics is too all-encompassing, it doesn’t exclude generic transcriptional regulatory mechanisms from those which truly constitute a mechanism of cellular memory (Henikoff and Grealley, 2016). Cellular memory can be defined as the maintenance of an activity state of a gene once it is triggered, for example, an environmental cue such as extreme temperature inducing gene desilencing over multiple generations (Klosin et al., 2017).

1.1.3 Mechanisms

There are various biological mechanisms that could be considered vehicles for cellular memory. Transcription factors are attractive candidates for transmitting cellular memory because their sequence specificity provides a plausible link between particular genetic loci and epigenetic phenomena. However, while their presence is often necessary for cellular memory, it is not sufficient. For example, in the orange and black patches of calico cats, both the active and inactive X chromosome are exposed to the same

transcription factor concentration, but only one stays active (Henikoff and Greally, 2016). The memory mechanism actually involves a long non-coding RNA (Xist) which recruits the repressive PRC2 H3K27 methyltransferase complex.

DNA methylation is an obvious mechanism of cellular memory because covalent modifications remain on the DNA strands during replication and are thus conserved in the daughter chromatids. Moreover, in mammals, hemi-methylated (meCG/GC) dinucleotides are restored to symmetrically modified (meCG/GCme) by the Dnmt1 methyltransferase post-replication. In other model organisms, such as *Drosophila*, which lacks this maintenance activity, and *C. elegans* and *S. cerevisiae*, which lack DNA methylation entirely, there must be other mechanisms accounting for cellular memory.

The final mechanism of cellular memory to discuss is the inheritance of histone modifications, but as this mechanism is the most pertaining to the thesis, it will be the topic of the next section.

1.2 The Histone Code

1.2.1 Overview

Within the field of Epigenetics, a major area of study is the heritable modifications to the histone proteins that package DNA into chromatin, the so-called “histone code” (Strahl and Allis, 2000) (Jenuwein and Allis, 2001). Histone modifications regulate gene expression by controlling the accessibility of genomic regions to the transcriptional machinery (Kornberg and Lorch, 1992), and by providing binding sites for chromatin remodelling proteins and/or transcription factors (Sanchez and Zhou, 2011, Clayton et al., 2000, Zippo et al., 2009). Research in recent years has led to greater knowledge and understanding of the “writers”, “readers” and “erasers” responsible for the deposition, function and removal of histone modifications, and the ways in which these processes influence each other, termed, “cross-talk”.

1.2.2 The nucleosome core particle

In chromatin, DNA is wound around histone protein cores in repeating structures called nucleosomes (Fig. 1.1) (Richmond and Davey, 2003), which are arranged into higher order structures. The histone core particle comprises two of each of these four histone types; H2A, H2B, H3 and H4, each of which has a globular domain and an N-terminal tail (Kornberg, 1977). The globular domains aggregate to form the octameric core, combining one H3-H4 tetramer with two H2A-H2B dimers, while the N-terminal tails protrude out from the nucleosome. Residues targeted for post-translational modifications (PTMs) in these tails include positively charged lysine and arginine, for methylation (me) and acetylation (ac), and hydroxyl group-containing serine,

threonine and tyrosine, for phosphorylation (P) (Suganuma and Workman, 2011). There are other types of modifications, such as sumoylation or ubiquitination (Gill, 2004), but these go beyond the scope of this thesis. The focus will be acetylation and methylation of lysines on histone 3.

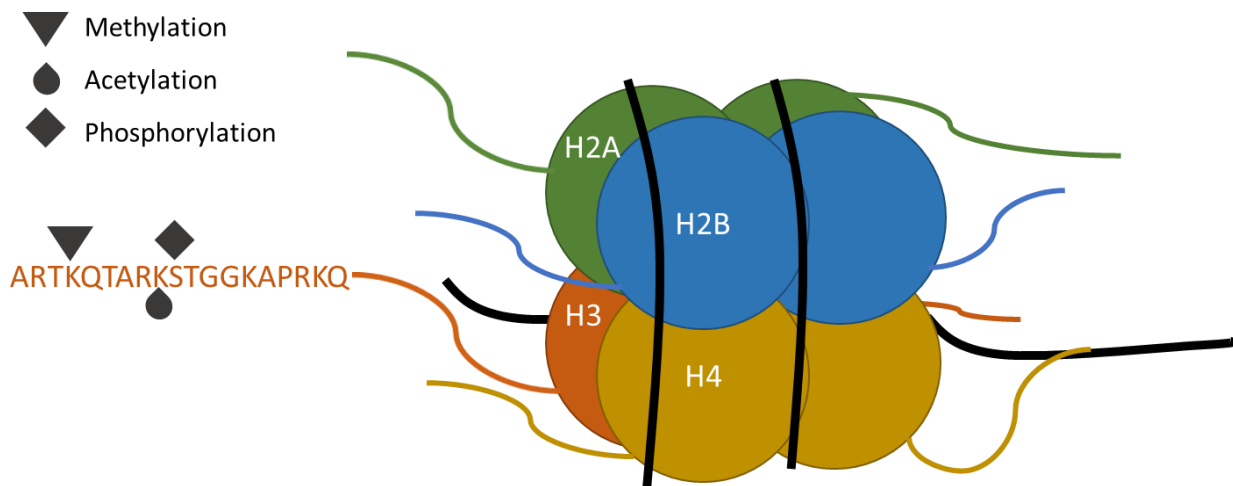


Figure 1.1 Structure of a nucleosome – DNA (black) winds around the octameric core of H2A, H2B, H3 and H4 subunits. The N-terminal tails, in corresponding colours, are free to interact with the DNA. Part of the H3 amino acid sequence is shown, to illustrate amino acids (K4, K9 and S10) which are targets for modification and relevant here. The grey shapes correspond to common types of chemical modification typically found on these residues.

1.2.3 Nomenclature

The nomenclature for histone PTMs, which will be referred to frequently throughout this thesis, follows the rules illustrated (Fig. 1.2). The standard abbreviation starts with the histone the post translational modification (PTM) is found on (e.g. H3 for histone 3), followed by the single amino acid code of the modified residue with its numerical position from the N-terminus (e.g. K9 for lysine 9) and the standard abbreviation for the chemical modification (e.g. me for methylation, ac for acetylation, P for phosphorylation) and where appropriate, the number of chemical moieties attached e.g. (me3 for three methyl groups).

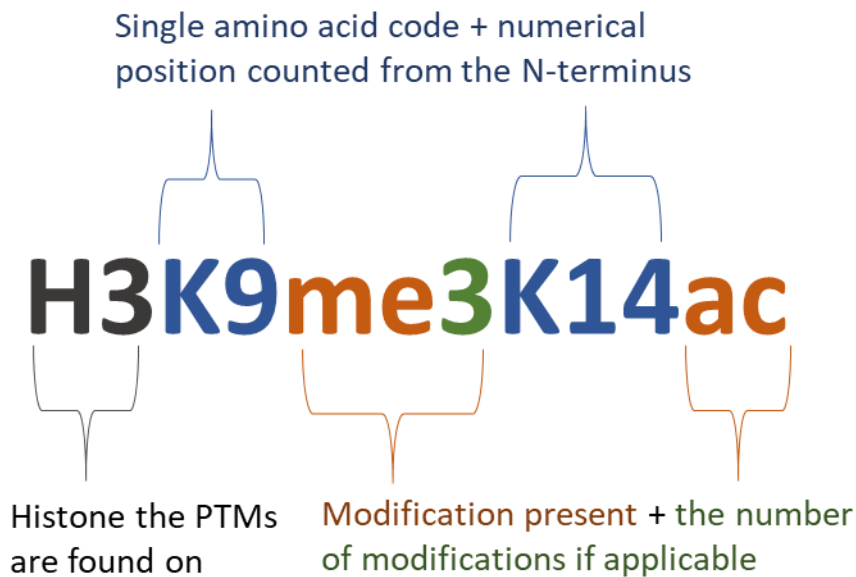


Figure 1.2 The histone modification naming convention - Each individual component is colour-coded to: the histone protein in grey, the amino acid in blue, the modification in orange and the number of modifications (where there are multiple) in green.

1.2.4 How histone modifications regulate gene expression

1.2.4.1 Demarcating chromatin domains

Support for the role of histone modifications in the regulation of gene expression comes firstly from their selective association with different chromatin domains (Gelato and Fischle, 2008). In the simplest classification, chromatin can be divided into euchromatin, which is characterised by low compaction and enrichment for actively expressed genes, and heterochromatin (Passarge, 1979), which is compacted and minimally transcriptionally active. The latter is further subdivided into constitutive or facultative heterochromatin (Metzler-Guillemain et al., 2008), depending on whether the compaction persists throughout cellular differentiation or is altered in responses to stimuli at cell cycle or developmental stages.

A protein localisation study in *Drosophila* linked histone modification enrichment to five principal chromatin types (Filion et al., 2010). Of the three heterochromatin types, one was associated with Polycomb PcG proteins and enriched for H3K27me3, and another associated with HP1 protein and enriched for H3K9me2. The remaining type, accounting for 48% of genome coverage, was categorised rather by a lack of H3K4me2 and H3K79me3, as detected by ChIP, and was the most repressive to transgenes. Of the euchromatin types, one was enriched for H3K36me3 and genes with a broad expression pattern over many embryonic stages

and tissues, described as “housekeeping genes”. The other type was characterised by a lack of H3K36me3 and genes linked to more specific processes. Thus, histone modifications can provide an epigenetic signature for different functional chromatin subsets.

1.2.4.2 *Chromatin accessibility*

Histone modifications may impact chromatin structure by directly changing interactions within or between nucleosomes. For example, neutralising a positive charge on lysine or arginine with acetylation will reduce the attraction between an N-terminal tail and the underlying negatively charged DNA. On a large scale, this might lead to a looser and more accessible local chromatin structure (Winter and Fischle, 2010). Biochemical fractionation experiments have provided some evidence for this, by demonstrating the association of histone acetylation with active chromatin. Nuclei from hepatoma tissue culture (HTC) cells treated with sodium butyrate, to induce hyperacetylation of nucleosomal histones, showed a 2-3 fold increase in the initial rate of digestion by DNase I (Nelson et al., 1978). This indicates that the hyperacetylation of the histones made the associated DNA more accessible to the DNase enzyme. Moreover, a later study showed that acetylation may interfere with the formation of a higher order chromatin structure. Polynucleosomal chromatin fragments enriched for acetylated species of histones H2B, H2A.Z (a variant of H2A) and H4 were found to be resistant to linker histone-induced precipitation in salt solution, compared to fragments from transcriptionally repressed chromatin (Ridsdale et al., 1990).

1.2.4.3 *Specific protein recruitment*

Another way that histone modifications may influence gene expression is by recognition by proteins that alter the structure and function of chromatin, or that recruit additional protein complexes that do. There are well-characterised protein binding domains known to interact with particular histone marks, for example chromo- and bromodomains, which bind methylated and acetylated lysines, respectively. Plant homeodomain (PHD) finger domains in different proteins have been shown to recognise both methylated and acetylated lysines, in particular H3K4me2/3, H3K9me3 and H3K14ac, but those that recognise methylated peptides are the largest group (Sanchez and Zhou, 2011) .

1.2.5 *Cross-talk*

The term ‘cross-talk’, in this context, refers to the interaction between different histone modifications. There is evidence for these interactions happening on different scales. For example, in the same N-terminal tail, histone modifications may influence the deposition of other modifications on neighbouring residues, usually by their action on the enzymatic activity of a “writer”. A good example occurs in gene activation, where

H3S10P stimulates the histone acetyltransferase Gcn5 to acetylate nearby H3K9 and H3K14. These three modifications act synergistically to provide a more favourable binding site for 14-3-3 proteins than H3S10P would alone, while excluding heterochromatin HP1 proteins from binding to their target, H3K9me2/3. The 14-3-3 proteins recruit chromatin remodelling complexes to promoter regions, enabling transcription (Clayton et al., 2000). This illustrates the way in which histone modification patterns may form cooperative binding sites for some protein factors, while excluding others.

Cross-talk may also take place on a larger scale, between nucleosomes. This type of interaction is mediated by modification binding proteins, which, depending on the spacing of the binding domains, can be very long-range. Brd4, for example, can simultaneously bind acetylated lysines on H3 and H4 via its two bromodomains. It is thought to play a mitotic marking function, since despite global hypoacetylation of mitotic chromatin, Brd4 remains bound throughout mitosis to genes that are immediately expressed afterwards (Zippo et al., 2009).

1.2.6 Transgenerational epigenetic inheritance

Transgenerational epigenetic inheritance is an umbrella term encompassing various examples of inheritance by means other than the genetic code. Despite a huge amount of public interest, it is very challenging to find true human examples because, firstly, it is only truly transgenerational epigenetic inheritance when the phenotype is inherited in the F3 generation. In a gestating mother, for example, it is possible for the F2 gametes in developing foetus to have been exposed to the environmental stress which caused the phenotype. This is a key flaw in examples cited of transgenerational epigenetic inheritance in humans, such as the Dutch Hunger Winter. There is also the confounding factor in any study relating to diet, that parents also pass on behaviours surrounding food and exercise to their children (Heard and Martienssen, 2014).

Nevertheless, in model organisms such as *C. elegans*, which are sufficiently short-lived to make the continued observation of a long series of generations experimentally feasible, there is a large body of evidence pointing to the existence of transgenerational epigenetic inheritance mechanisms. Some papers cite small RNA-based mechanisms (Guerin et al., 2014), but this subsequent section will discuss the inheritance of histone modifications as an important mechanism of transgenerational epigenetic inheritance.

1.2.6.1 Inheritance of parental histones

There is a convincing body of evidence supporting a model for the inheritance of parental histone marks by the direct incorporation of modified parental histones into the newly synthesised daughter chromatids

during DNA replication (Fig. 1.3). Such a model depends on the local and temporal coupling of the passage of the replication fork with nucleosome assembly.

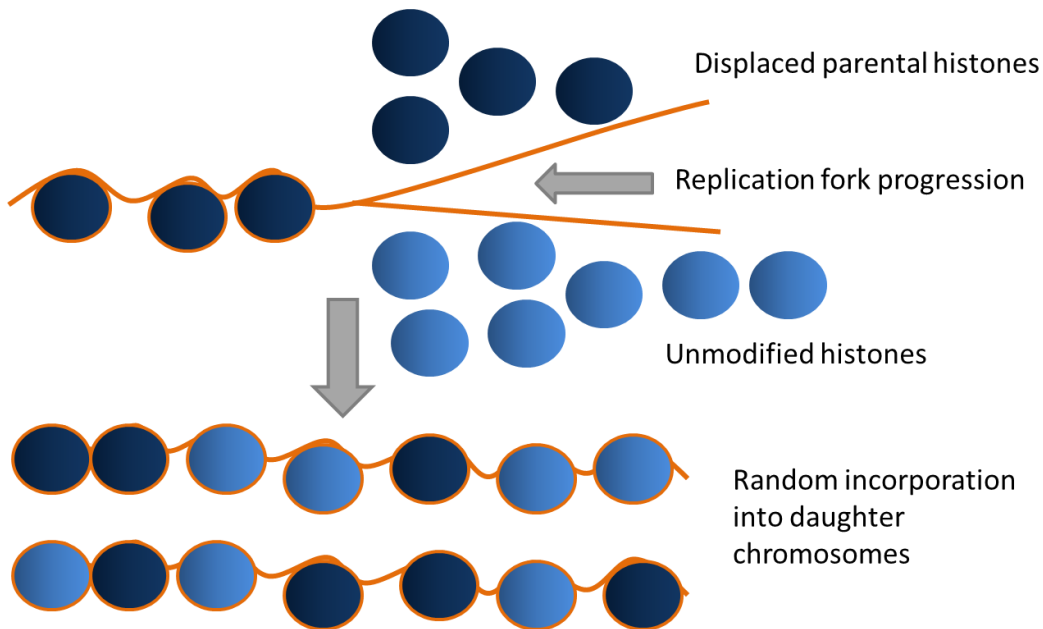


Figure 1.3 Inheritance of parental histones- passage of the replication fork liberates a pool of “parental” histones carrying modifications characteristic of the local chromatin environment. Parental and unmodified histones are randomly incorporated into the daughter chromosomes.

Evidence for this coupling comes from the observation that manipulating the location of a replication fork stall results in local changes to histone modification patterns (Schiavone et al., 2014). During DNA replication, the formation of G4 quadruplex structures can halt the progress of DNA polymerase so that it becomes uncoupled from DNA helicase, resulting in gaps that are replicated after the passage of the replication fork, in the absence of the pool of parental histones. Changing the position of a G4 motif in relation to the transcription start site at the BU-1 locus altered the pattern of H3K4me3 and H3K9/14ac within the gene body, without producing any alteration in expression, suggesting that the altered modification pattern occurs as a result of the changed location of the replication fork stall (Schiavone et al., 2014).

Furthermore, in vivo experiments in *C. elegans* embryos are consistent with a model whereby modified histones are passed to product DNA strands in-cis and locally during DNA replication. Mutants in the *mes-3* gene, encoding an essential component of Polycomb Repressive Complex 2 (PRC2), lack H3K27me3. By

creating embryos with either a *mes-3* oocyte and WT sperm (Fig. 1.4A), or a WT oocyte and *mes-3* sperm (Fig. 1.4C), it was possible to ensure that all H3K27me3-marked chromosomes were exclusively paternally or maternally contributed. In both types of embryo, it was shown that H3K27me3 remained restricted to the ‘gamete-of-origin’ chromosome set (Fig. 1.4B & D) until the de novo establishment of repressed chromatin domains during larval germline development (Gaydos et al., 2014). The paternally inherited H3K27me3 was only transmitted through four rounds of DNA replication, in the absence of a functional PRC2 H3K27 methyltransferase. However, embryos with maternally contributed H3K27me3 also inherited maternally supplied PRC2 with the oocyte, which maintained high H3K27me3 levels throughout embryonic development. This observation supports the idea that the dilution of histone marks with each round of DNA replication must be compensated. In this example, PRC2 could be perpetuating inherited patterns by binding pre-existing H3K27me3, which then stimulates its methyltransferase activity to locally methylate neighbouring unmethylated H3K27 residues.

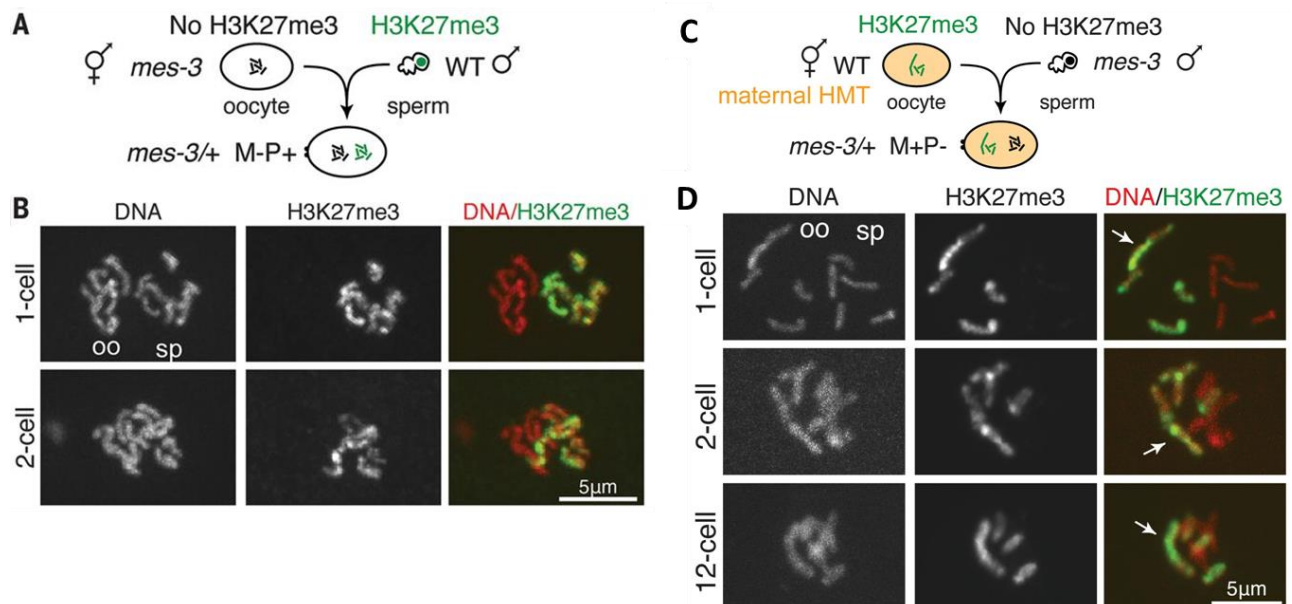


Figure 1.4 The restriction of H3K27me3 to the gamete-of-origin chromosome set. Figure 1.4A shows the formation of an embryo with H3K27me3 inherited solely from the sperm-derived chromosomes. The oocyte is *mes-3* mutant and therefore has no active PRC2 and no H3K27me3. Figure 1.4B shows images of two pronuclei in the one-cell embryo (upper row) and a diploid nucleus in a two-cell embryo (lower row). Merge panels show DNA in red and H3K27me3 in green. In the diploid nucleus, the H3K27me3-marked, sperm-derived chromosomes remain distinct. Figure 1.4C shows the formation of an embryo with H3K27me3 inherited solely from oocyte-derived chromosomes. This time, the oocyte has PRC2 activity, but the sperm is *mes-3* mutant. Figure 1.4D shows images of two pronuclei (upper row) and diploid nuclei (lower two rows) as in 1.4B. The arrows point to a III-X-IV fusion chromosome in the oocyte-derived chromosome set. As in 1.4B, the H3K27me3-marked, oocyte-derived chromosomes, remain distinct. Adapted from (Gaydos et al., 2014).

1.2.7 Methods and limitations of study

One of the challenges levelled by those that dispute the existence of a histone code is the lack of evidence that histone modifications act combinatorially in influencing downstream developmental events (Henikoff and Grealley, 2016). Moreover, across the different model organisms, evidence for universality in the histone code is lacking. It appears that histone modification cross-talk depends very much on context. In other words, it may vary from cell to cell, tissue to tissue, organism to organism (Winter and Fischle, 2010).

Furthermore, how we “see” the histone code is powerfully influenced by the experimental method used to probe it. Chromatin immunoprecipitation (ChIP), for example, uses antibodies to target specific histone modifications, and pull-down DNA fragments in close proximity to these modifications that can then be mapped back to the genome. Thus, ChIP data will reveal genome-wide patterns of histone mark distribution. Separation and analysis of peptide fragments by mass spectrometry, on the other hand, will provide information about associations within a single N-terminal tail. The advantage of mass spectrometry is that it is more quantitative than any antibody-based approach. However, while it may quantify the different modified peptides in a sample accurately, it cannot tell you where those modifications came from. Thus, mass spectrometry and ChIP approaches are very complementary. As mass spectrometry is the major technique employed in obtaining the data presented in this thesis, the next section will explain the principles of the technique, and the specific method used in these experiments.

1.3 Mass spectrometry technologies

1.3.1 Principle of mass spectrometry

The objective of mass spectrometry is to accurately determine masses of chemicals present in complex mixtures. This is achieved by a common sequence of processes (Fig. 1.5). Briefly, molecules in a sample need to be ionised in order for them to travel in an electric or magnetic field towards a detector. The speed at which ions are able to travel through the mass analyser is directly proportional to their charge and indirectly proportional to their mass. The data output is a mass to charge ratio (m/z) that is easily converted to mass by looking at isotope series. The m/z of the +1 ion will correspond to the mass of that ion. The +1 ion will be separated from the +2 ion by 0.5 m/z , and from the +3 ion by 0.33 m/z (Yergey et al., 1983). So, to work out the mass from a +2 ion, it is necessary to add the mass of two protons and then divide by 2.

Mass spectrometry can be used to quantify the amount of a substance by calculating the intensity of the ions it detects (Urban, 2016). In proteomic mass spectrometry, it is the intensity of peptide ions that is measured.

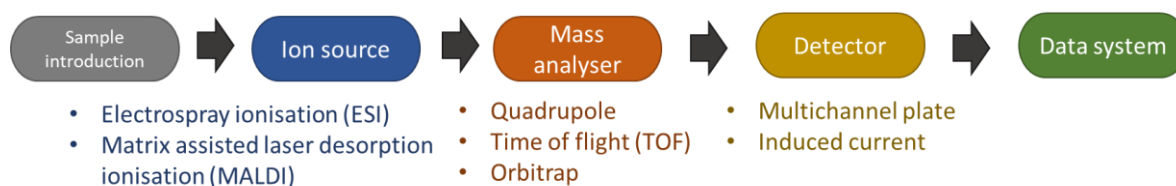


Figure 1.5 The universal processes of mass spectrometry - the various technologies listed in bullet points are colour coded to the stage they correspond to and will be described in detail later in this chapter.

1.3.2 Ionisation methods

The most popular methods of ionisation for proteomic mass spectrometry, electrospray ionisation (ESI) and matrix-assisted laser desorption ionisation (MALDI), both add charge to peptide ions and convert them to gaseous form. This enables them to travel in the mass analyser. In ESI, a sample must be acidified before being passed through a capillary needle with a high electric charge applied (Sarg et al., 2013). Protons from the acid bind to the peptides in the sample, which become gaseous. In MALDI, an analyte molecule is co-crystallised with a matrix, usually 2,5-dihydroxybenzoic acid or α -Cyano-4-hydroxycinnamic acid. The aromatic structures within the matrix molecule enable absorption of laser energy at a particular wavelength, which induces a complex ion-molecule interaction resulting in proton transfer from the matrix to the analyte molecule (Schurenberg et al., 1999). This change is accompanied by a solid to gas phase transfer from the crystallised analyte-matrix mixture to a gaseous, charged analyte ion, which can travel into the mass analyser. One helpful feature of MALDI is that it results in predictably charged ions, usually +1. ESI produces a wider range of charges. MALDI has also facilitated the development of mass spectrometry imaging, which works by spraying matrix over a fixed tissue section and scanning with a laser, providing spatial information about the distribution of compounds (Aichler and Walch, 2015).

1.3.3 Mass analysers and detectors

The purpose of a mass analyser is to separate ions according to their m/z . They can be separated with magnetic fields or electric potentials. Quadrupoles, so called because they comprise four rods arranged around a beam axis, create an oscillating electric field which selectively stabilises or destabilises the motion of ions (Miller and Denton, 1986). The destabilised ions are lost, however, which limits sensitivity. Time-of-flight (TOF) gives all ions the same energy at the source, but then separates them by the time they take to travel down the flight path to the detector, which is proportional to their m/z (Boesl, 2017). Fewer ions are lost with TOF than when using a quadrupole. However, the Orbitrap mass analyser has the best resolution

and mass accuracy. Orbitrap detectors trap ions in a cell, in which they oscillate along a central spindle electrode, which is held at a positive or negative potential. The frequency at which the ions oscillate is proportional to their m/z , and the higher that frequency, the faster the ions can escape the Orbitrap and be detected (Hu et al., 2005). Different mass spectrometers can use one or more of these three technologies. The Q-Exactive, on which all experiments discussed in this thesis were performed, uses two types of quadrupole to isolate different m/z windows, and an Orbitrap final detector.

1.3.4 Tandem mass spectrometry

Tandem mass spectrometry, often abbreviated MS/MS, involves a fragmentation step between two mass analysers (Fig. 1.6). Fragmentation is usually achieved by collision-induced dissociation (CID) with a background gas. The first mass analyser (MS1) selects ions of a particular m/z for fragmentation, and the second mass analyser (MS2) determines the m/z of the fragment ions. Because the fragment ions can be traced back to their parent ions during analysis, MS/MS can provide valuable structural information about a molecule.

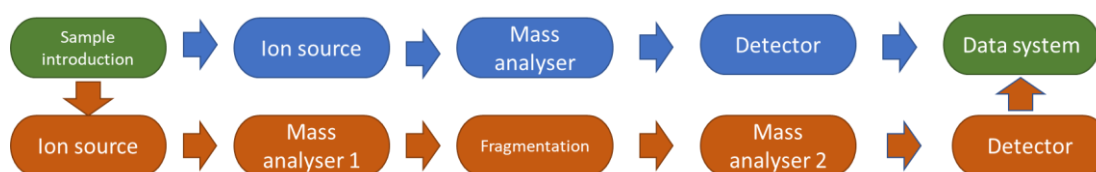


Figure 1.6 Comparing the process of standard MS vs MS/MS - Green steps indicate the common beginning and end; the blue trajectory is standard MS and the orange is MS/MS.

For peptides, fragmentation usually occurs at amide bonds, producing a predictable range of ions (Fig. 1.7). Depending on whether the proton released by hydrolysis of the amide bond associates with the C- or N-terminal end of the peptide, the ion is referred to as 'b' or 'y'. The number of amino acids in the ion is suffixed in subscript after the letter, for example 'b₃'.

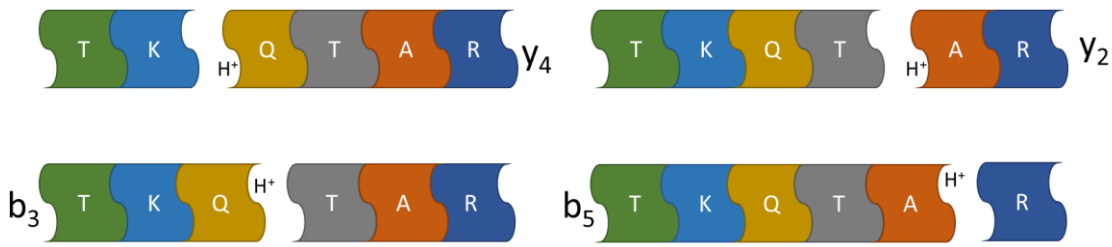


Figure 1.7 – Four of the ways the TKQ TAR peptide might fragment in a collision chamber, at a variety of amide bonds. The first row shows y ions and the second, b ions, next to the unprotonated fragment that will not proceed to the mass analyser.

1.3.5 The Q-Exactive mass spectrometer

An overview of the workings of the Q-Exactive (Fig. 1.8) will help illustrate the use of some of the technologies discussed previously and is important background for later considerations. The Q-Exactive performs alternate MS1 and MS2 scans. Upon entering the instrument, ions are filtered and focused by an S-lens, a type of focusing quadrupole that removes any molecules that were unsuccessfully ionised at the source. For an MS1 scan, the selective quadrupole is not engaged, and all ions pass into the C-trap (Perry et al., 2008), an ion guide with a curved central axis. The C-trap is required to ensure that ions' kinetic energy is suitably matched to the voltage on the centre electrode of the Orbitrap. Too low and ions will be sucked into the centre electrode, but too high and they will collide with the outer electrode. Nitrogen within the C-trap facilitates collisional cooling of ions. Once cooled sufficiently, they are released and pass into the Orbitrap detector. Under these conditions, m/z from all intact ions are detected.

For an MS2 scan, an electrostatic field is applied which allows only ions within a specified isolation window to enter the C-trap. Moreover, loss of energy in the C-trap can be controlled such that when ions are sent into the collision chamber to fragment, predictable bonds will break. Peptides, for example, break at amide bonds. The fragment ions are then also directed to the Orbitrap detector.

Because the Q-Exactive can only perform either an MS1 or MS2 scan at one time, it is necessary to control the entry of the analytes from a complex sample into the instrument so that it is not overwhelmed.

Normally, a HPLC column is linked to the mass spectrometer, to separate analytes according to their relative hydrophobicity/hydrophilicity.

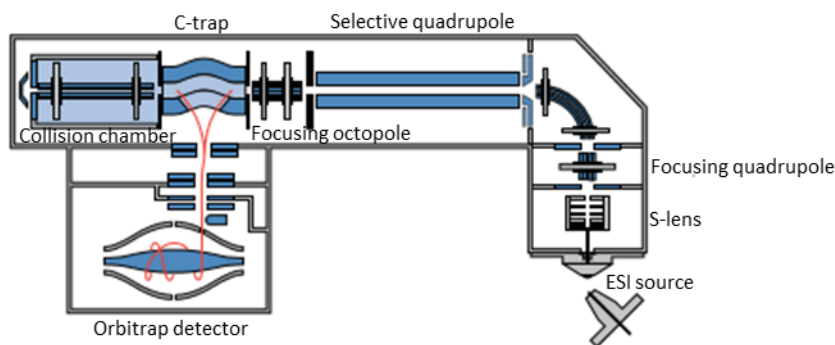


Figure 1.8 – The components of the Thermo Scientific Q Exactive Benchtop LC-MS/MS. Briefly, ions enter the instrument at the ESI source, pass through the quadrupoles before reaching the C-trap, from where they proceed to the collision chamber (for an MS2 scan) and/or finally the Orbitrap detector. Adapted from tools.thermofisher.com.

1.3.6 Data acquisition methods

Whether to choose data dependent (DDA) or data-independent acquisition (DIA) is an important consideration before embarking on an MS/MS experiment. DDA selects the most abundant ions in an MS1 spectrum for fragmentation. The number of ions selected can be set by the user. This is often the method used when there is no prior knowledge about what is present in a sample. An advantage of DDA is that it tends to produce a very high quality MS2 spectrum. However, this approach risks missing low-abundance analytes that may be of interest.

DIA, by contrast, selects all ions in a pre-specified isolation window (between 10 m/z and 50 m/z) for fragmentation, which is repeated over the full m/z range in the MS1 spectrum (Doerr, 2015). This approach produces a much more complex MS2 spectra, since more than one MS1 analyte is fragmented in each MS2 scan, but the advantage is that no data is lost (Canterbury et al., 2014). Moreover, there are analysis techniques to help sort the relevant fragment ions from the noise in the MS2 spectra. DIA data analysis relies on spectral libraries made from DDA data.

1.4 Histone Proteomics

1.4.1 Proteomic approaches



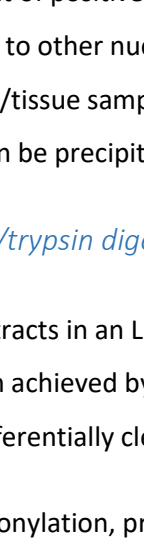
The field of proteomics is expansive and diverse in its methods, but the main approaches can be grouped into top-down, middle-down and bottom-up classifications (Table 1.1).

The top-down approach runs samples containing intact proteins, with extensive fragmentation once in the instrument. The advantage of this method is that sample preparation is very simple, but sensitivity is poor (Janssen et al., 2017). It can provide useful information about protein structure and is an especially helpful approach for the analysis of complex mixtures of proteins.

The middle-down approach is helpful for researchers interested in the combinations of modifications found on a particular histone tail. Histones are minimally digested with proteinases such as Glu-C, so called because it preferentially cleaves at the C-terminal of glutamic acid residues, and Asp-N (N-terminal of aspartic acid). These enzymes typically cut infrequently enough that the entire histone N-terminal tail can be detected as one peptide. MS analysis of these peptides can provide useful information about the coexistence of modifications. For example, digestion with the enzyme Glu-C produced a 50 amino acid-long peptide containing lysines 4, 9, 23, 27 and 36 of histone 3. Analysis of these peptides revealed that H3K4me3 is found in combination with K9me3, K23ac and K36me2 on the same peptide, but K27me3 never is, indicating bivalency of K4me3- and K27me3-marked domains (Sidoli et al., 2014).

The experiments discussed here used bottom-up proteomics, so-called because the proteins are extensively digested prior to mass spectrometry, rather than fragmenting mostly in the instrument. A bottom-up approach was the one most likely to provide information specifically about H3K4me3, the focus of this study, as the TKQTAR peptide containing the K4 residue can be specifically targeted. It is a challenging modification to detect, owing in part to its low natural abundance. A prior study looking at changes in histone H3 marks during the *C. elegans* lifecycle, by middle-down proteomics, does not even mention H3K4me3, suggesting uncertainty regarding its detection (Sidoli et al., 2016).

Table 1.1 – Types of Proteomics; a comparison of top-down, middle-down and bottom-up approaches

Bottom-up	Top-down	Middle-down
<ul style="list-style-type: none"> • Involves extensive digestion of proteins into peptides (with e.g. trypsin) which can then be easily separated by reversed-phase HPLC • Sensitive • Limited information available about the coexistence of combinatorial marks • Sample preparation requires special precaution due to the generation of hydrophilic peptides 	<ul style="list-style-type: none"> • Analysis of intact proteins, with extensive fragmentation • Low sensitivity • Simple sample prep 	<ul style="list-style-type: none"> • Proteins digested into large peptides (50-60aa), keeping the N-terminal tail intact. • Less sensitive than bottom-up • Low abundance PTMs are frequently missed • Provides information about the coexistence of combinatorial marks 

1.4.2 Preparation of histones for bottom-up MS

1.4.2.1 Histone extraction

Owing to their high content of positively charged lysine and arginine residues, histones are uniquely acid- and salt-soluble compared to other nuclear proteins. They are, therefore, straightforwardly extracted by first purifying nuclei from a cell/tissue sample, and then resuspending nuclei in acid or salt to solubilise the histones. Histones can then be precipitated out of solution, resolubilised in water and stored frozen.

1.4.2.2 Propionylation/trypsin digestion

Before running histone extracts in an LC-MS/MS instrument, they must be digested into smaller peptides and stabilised. This is often achieved by a combination of propionylation and trypsin digestion. Trypsin is a serine protease which preferentially cleaves at the C-terminus of lysine and arginine residues.

In the first round of propionylation, propionic anhydride reacts with the amine group on the lysine sidechain, leaving a propionyl moiety (Fig. 1.9). During subsequent trypsin digestion, these moieties provide steric interference around lysine residues, limiting trypsin cutting to only arginine residues. On H3, this yields peptides of between 6 and 14 amino acids, an appropriate size for running in the instrument (Fig. 1.10). A second round of propionylation is performed to stabilise the N-terminal amines on each peptide.

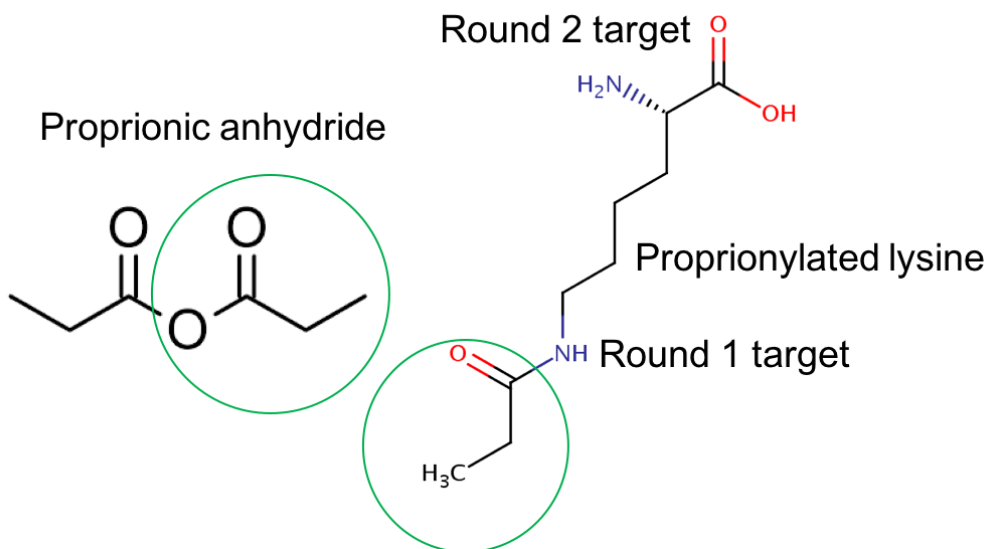


Figure 1.9 – Propionylation of a lysine. The green circle highlights the reactive group on propionic anhydride which ends up attached to the amine group on the lysine sidechain. After a second round of propionylation, the same chemical group will be attached to the free amine group at the N-terminus of each peptide generated by enzyme digestion, labelled 'Round 2 target'.

It is crucial that propionylation is performed within a pH range of 7-9, because the propionic acid waste product will lower the pH over the course of the reaction. At low pH, propionic acid may auto-hydrolyse, reducing the amount of reagent available to react with amine groups of lysine residues or N-termini. Above pH 10, though, and unintended amino acids may become propionylated, causing problems with downstream analysis (Janssen et al., 2017).

Two rounds of propionylation as described here (Garcia et al., 2007) has many advantages. As mentioned previously, it prevents cleavage by trypsin after every lysine residue, ensuring that peptides are a suitable size for detection by LC-MS/MS. It also helps ensure that peptides are consistent, since some lysine modifications could interfere with trypsin cleavage. Trimethylation (me3), for example, would prevent trypsin cleavage, whereas monomethylation (me1) would not. This is essential for identification because the mass increase conferred by all the modification states of a peptide needs to be predictable (Table 1.2). Note how the mass increase for me1 (70.09) is larger than that for me2 (28.05). This is because monomethylated lysines are able to accept a propionyl group in the first round of propionylation, but dimethylated and trimethylated lysines are not. Moreover, the propionylmethyl group ensures that me1-modified peptides elute later than their me2/me3-modified counterparts, by conferring additional hydrophobicity. This is

helpful because it spaces out the entry of similar sized peptides into the Q Exactive and thus allows for a more complete analysis.

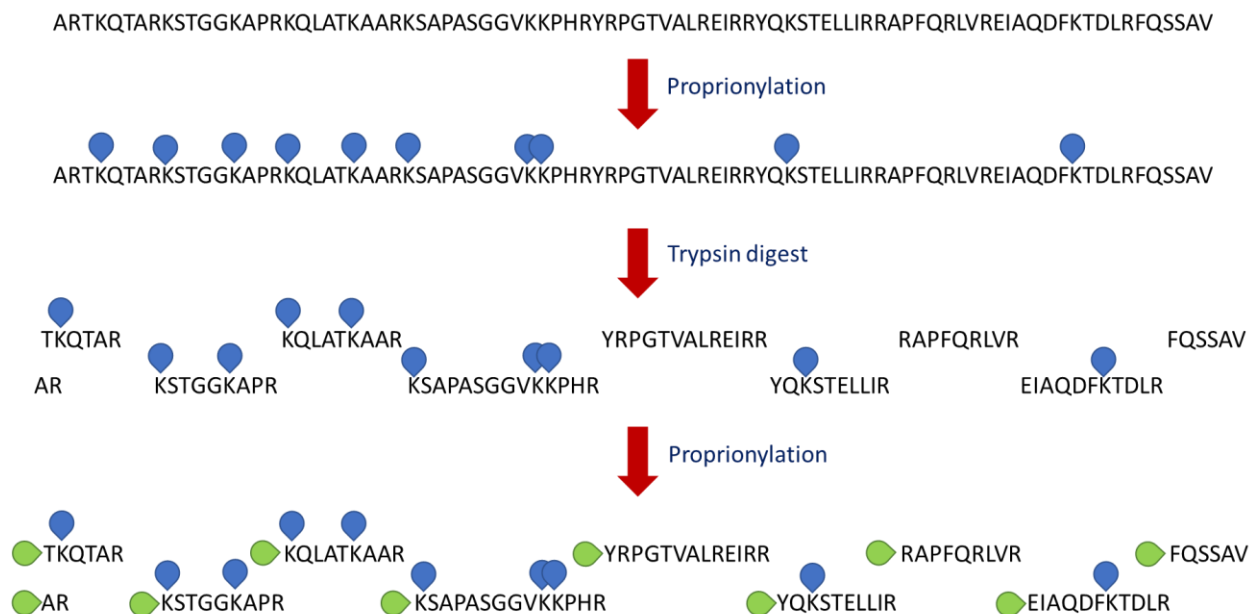


Figure 1.10 Peptide-generation from the H3 N-terminal tail. Blue shapes indicate the position of carbonyl moieties on lysine sidechains after the first round of propionylation. Green shapes indicate the same but on N-terminal amines of all types of amino acid after the second round. Trypsin cleavage occurs at all arginine residues; thus all peptide sequences end in ‘R’. Peptide sizes range between 3 and 14 amino acids, but anything smaller than the TKQTAR peptide will not be detected.

Table 1.2 – The increase in mass associated with propionylation, methylation and acetylation modifications, in their common arrangements. Data analysis software uses these expected increases to search for histone peptides.

Modification	Increase in Mass
Propionyl (Pr)	56.06
Methylpropionyl (me1)	70.09
Dimethyl (me2)	28.05
Trimethyl (me3)	42.08
Acetyl (ac)	42.03
Double Propionyl (N-term lysine)	112.11
Methylpropionyl Propionyl (N-term lysine)	126.11
Dimethyl Propionyl (N-term lysine)	84.11
Trimethyl Propionyl (N-term lysine)	98.14
Acetyl Propionyl (N-term lysine)	98.1

The N-terminal propionylation in the second round makes all peptides more hydrophobic and thus improves their retention on the HPLC column, as well as reducing the likelihood that they will be removed in a wash step. Nevertheless, one group has suggested further caution with respect to the most hydrophilic peptides, recommending that in the N-terminal modification step, propionic anhydride is replaced with phenyl isocyanate, which confers even more hydrophobicity (Maile et al., 2015).

1.4.2.3 *Desalting and ion-pairing*

Desalting or 'tip clean-up' happens after propionylation. It is done to remove salts and/or other contaminants which could interfere with downstream processes. Briefly, peptides are resuspended in 0.1% trifluoroacetic acid (TFA), the 'binding solvent'. This solution is pipetted up and down within a special pipette tip packed with a graphite carbon-based material, known as a Hypercarb tip. The tip interior selectively binds highly polar compounds, so should in theory only bind the peptides. These can then be eluted in a different solvent comprising 90% acetonitrile (ACN) and 0.1% TFA.

Before the purified peptide sample can be injected into the LC-MS/MS, the elute must be dried down to remove the ACN, and then resuspended in an appropriate anionic ion-pairing reagent, usually a perfluorinated acid. TFA has already been mentioned and is the most commonly used, however pentafluoropropionic acid (PFPA) and heptafluorobutyric acid (HFBA) are more hydrophobic and may be a more suitable choice for hydrophilic peptides such as H3K4me3-modified TKQTAR (Shibue et al., 2005). Whichever acid is selected, the TFA⁻, PFPA⁻ or HFBA⁻ anions will interact with basic sidechains or N-terminal amines, neutralise their positive charge and reduce their hydrophilicity, thus increasing their affinity for the stationary phase of the HPLC column.

1.4.3 *Reverse-phase high-pressure liquid chromatography*

Reverse-phase high performance liquid chromatography (RP-HPLC) separates analytes according to their hydrophobicity and is the most common form of chromatography used in LC-MS/MS applications (Josic, Kovak 2010). It uses a column, packed with silica beads, to which alkyl chains are attached. These chains bind analytes when the solvent running through the column is aqueous, the so-called stationary phase. However, as more organic solvent is added, the analytes begin to elute, with the most hydrophilic eluting first. The column dimensions, particle size and length of the alkyl chains are all variables which can be adjusted for different experiments. For peptides, a particle size of 5µm is favoured, and a longer alkyl chain, usually 8-18 carbons (C8-C18). The rationale for this is that while whole proteins are likely to have many hydrophobic moieties that will bind very short alkyl chains, peptides and small molecules need longer chain lengths to be captured.

HPLC machines need to be able to pump and mix two solvents, one aqueous and conventionally referred to as solvent A, the other organic, called solvent B. For LC-MS applications, both solvents normally contain 0.1% formic acid, to provide a source of protons at the ESI stage. Solvent A is mainly HPLC grade water, and solvent B is mainly a water-miscible organic solvent such as ACN or methanol. The solvent A used in these experiments contained 0.1% formic acid and 3% ACN, because a small amount of organic solvent prevents long C18 alkyl chains from matting down in a hydrophobic layer, and thus being ineffective at capturing peptides.

In addition to the column, particle and solvent parameters discussed, it is also necessary to establish an effective gradient for the replacement of solvent A with solvent B. This controls the speed at which the analytes elute and needs to be optimised to allow enough of a time interval between elution peaks, but not so slow as to be experimentally impractical. In these experiments, a multi-step gradient was used to improve retention and separation of the more hydrophilic peptides (Sidoli et al., 2015). Solvent B was increased from 3% to 8% over 5 mins, 8% to 25% over 55 mins and 25% to 60% over 26 mins. The 55 min step is when most peptides are expected to elute, which is why the gradient is reduced and, consequently, the elution slowed down.

1.4.4 Data analysis

1.4.4.1 Skyline

Skyline is an open-access software for the processing of histone proteomic DIA data (MacLean et al., 2010). It allows the user to manually identify peptides, displayed in the software as peaks on a graph with retention time (the time at which the peptide leaves RP-HPLC column) on the x axis and intensity (of the peptide-derived ions hitting the detector) on the y axis (Fig. 1.10).

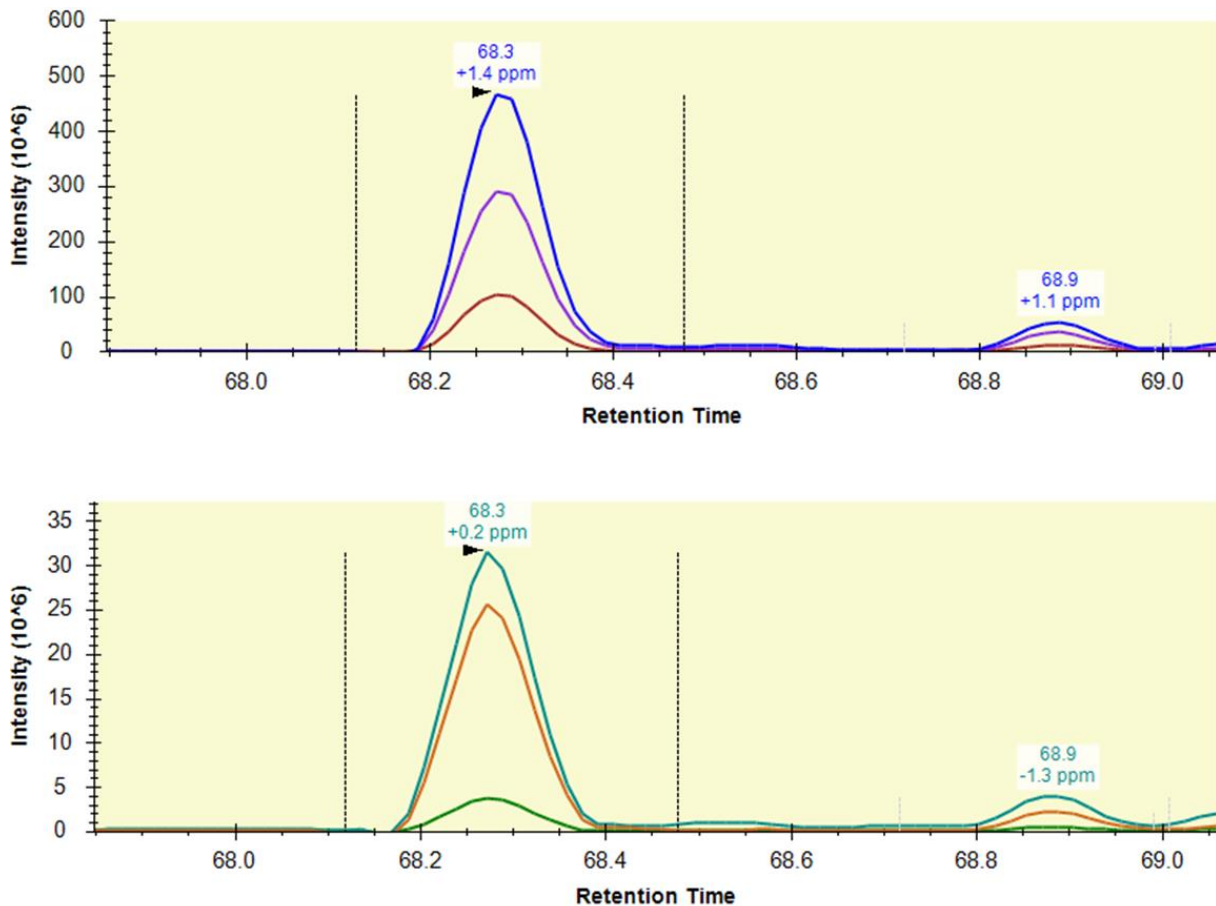


Figure 1.11 – A screenshot from the histone proteomic DIA data software package Skyline, showing the representation of the MS1 and MS2 spectra.

In the top panel (Fig 1.11), the MS1 spectrum, the different colours represent the different charge states of the ions. The blue peak represents the +1 ion, the purple peak the +2 ion and the red peak the +3 ion. The software compares the observed ratio of '+1: +2: +3' with an expected ratio, based on the likelihood of each of these ions forming at the ionisation stage. The output of this comparison is an 'idotp' number. The closer to 1 the 'idotp' is, the more likely that the MS1 peaks correspond to the peptide of interest.

In the bottom panel (Fig 1.11), the MS2 spectrum, the four different coloured peaks represent the four most abundant MS2 ions. A 'dotp' number compares the observed ratio of these ions with an expected ratio, determined from a spectral library, a predetermined reference spectrum of the fragmentation pattern of the peptide in question. It should also be close to 1.

Because the MS1 and MS2 peaks are lined up on the x axis i.e. have the same retention time, it is likely that the MS2 ions are derived from the MS1 ions. The lining up of the peaks, 'idotp' and 'dotp', similarity of retention times between replicates and a low ppm (measure of error), are all parameters that must be considered, to be sure of correct peptide identification.

1.4.4.2 EpiProfile

EpiProfile is a software run as code in MATLAB which ascertains retention times for each modified peptide of interest and their relative abundance (Yuan et al., 2015). The output can be helpful on its own or alongside the manual Skyline analysis, to cross-check retention times and later, relative abundances.

1.4.4.3 Correction Factors

Different ions behave differently in the MS/MS instrument. Not all ions that enter the instrument are detected and some are more likely to reach the detector than others. Some peptides have very similar masses and elution times and may mask one another. For this reason, correction factors are usually applied during the analysis to make the calculated relative abundances more accurately reflect the true relative abundances. In this analysis, two correction factors were applied, for ionisation efficiency and isotopic abundance (2.3.5.2), as published in (Lin et al., 2014).

1.5 COMPASS

1.5.1 The complex

In this thesis, LC-MS/MS was used to investigate histone modification changes due to actions of the COMPASS complex. COMPASS is a contraction of Complex of Proteins Associated with Set1. It is a conserved eukaryotic histone methyltransferase, targeting H3K4 specifically. In yeast, it is the only H3K4 methyltransferase. In *C. elegans*, there are two H3K4 KMTs, SET-2 and SET-16. *Drosophila* has two additional H3K4 methyltransferases, Trithorax (Trx) and Trithorax-related protein (Trr), which are related to the mammalian Mixed Lineage Leukaemia (MLL) group (Eissenberg and Shilatifard, 2010). Mammals have four MLL methyltransferase complexes (MLL 1-4) and two Set1 homologues, Set1A and Set1B (Lee and Skalnik, 2005), which makes functional COMPASS study in mammalian systems challenging.

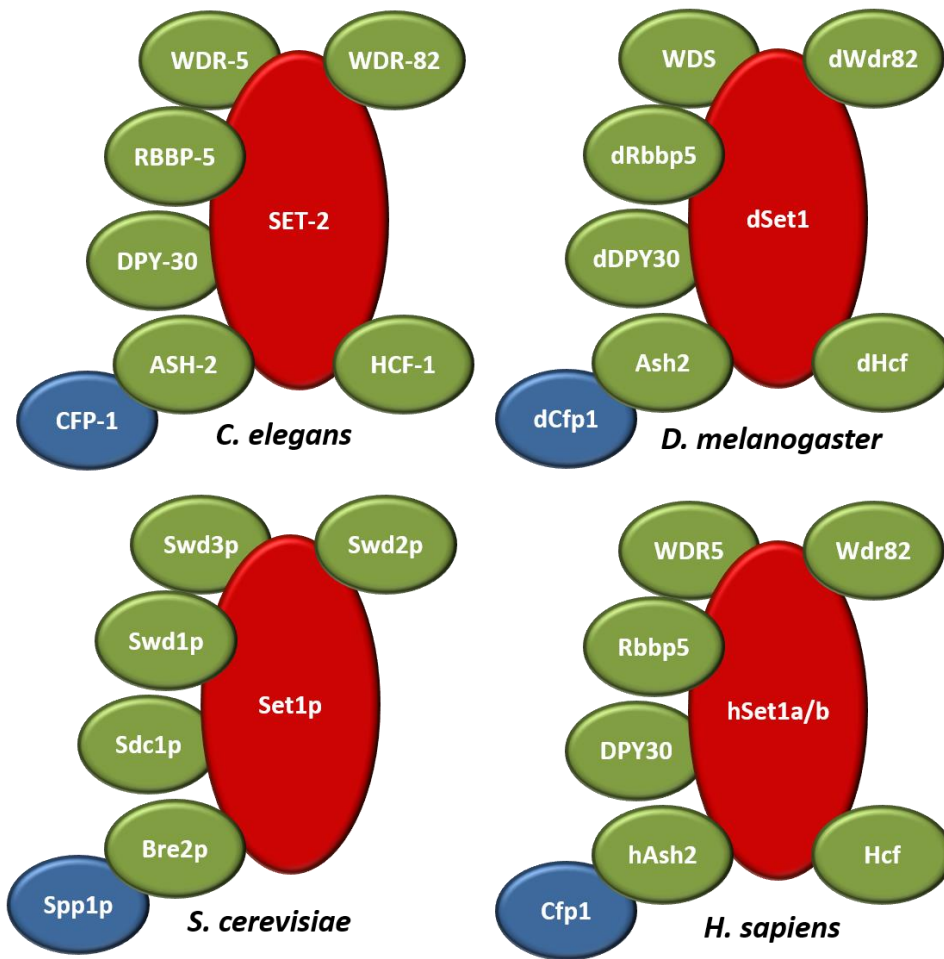


Figure 1.12– Conservation of COMPASS in eukaryotes. Worm, fly, yeast and human complexes are illustrated for comparison. Protein names are written as summarised in (Ardehali et al., 2011).

Yeast Set1p, *C. elegans* SET-2, Drosophila dSet1 and human Set1A/B all fall into the same subfamily of H3K4 methyltransferases according to phylogenetic comparison (Ardehali et al., 2011). They share an RNA recognition motif that is missing in Trx/MLL-type proteins. The Set1 homologues and the proteins they associate with are highly conserved (Fig 1.12). The various functions of the individual subunits have been researched extensively, but this thesis will focus on SET-2 and CFP-1, the catalytic and targeting subunits, respectively.

1.5.2 Set1/SET-2

1.5.2.1 Discovery

Set1/SET-2 is the catalytic subunit of the COMPASS complex. Its catalytic activity is conferred by its SET domain, a conserved domain common, but not exclusive, to all histone lysine methyltransferases (KMTs). SET is a contraction of Su(var)3-9, Enhancer of zeste and Trithorax.

The first Set1, and the associated H3K4 methylation machinery that made up COMPASS, was identified in *S. cerevisiae* (Miller et al., 2001). Homologues of Set1 were then identified in *S. pombe* (Roguev et al., 2003), humans (Lee and Skalnik, 2005), *C. elegans* (Simonet et al., 2007) and *Drosophila* (Ardehali et al., 2011).

1.5.2.1.1 Lysine methyltransferase activity

KMTs targeting other lysines had been previously characterised before the discovery of Set1. Su(var)3-9 homologues, mammalian SUV39H1 and *S. pombe* Clr4, were the first to demonstrate in vitro KMT activity (Rea et al., 2000). Interestingly, in the same study, Trx tested negative in KMT assays. The authors proposed that this might be due to Trx lacking a Cysteine-rich preSET region, a similarity it shares with Set1. Because deletion of the preSET region in an SUV39H1 fusion protein abolished KMT activity, it was thought to be essential. Moreover, when Set1/COMPASS was first discovered it failed to show methyltransferase activity with free histones or nucleosomes (Miller et al., 2001). However, it was shown that same year by purifying COMPASS from Bre2p or Shg1p proteins, that it did show KMT activity specific to H3K4 (Roguev et al., 2001). The authors noted that purifying COMPASS by tagging Set1 had an inhibitory effect and could have been why previous studies failed to find KMT activity in the Trx/Set1 branch of SET domain proteins

1.5.2.1.2 Lysine methylation dynamics

After the discovery of SUV39H1, dozens more KMTs were identified through homology searches with the SET domain (Dillon et al., 2005). Characterisation of these revealed that KMTs are specific both for the lysine within the substrate, and the degree of methylation. KMT2A or MLL1, for example, preferentially dimethylates unmodified H3K4 (Nakamura et al., 2002), but can trimethylate when associated with its endogenous interacting proteins (Schneider et al., 2005). This is because trimethylation requires optimal enzymatic configuration to achieve. All known KMTs use S-adenosyl-methionine (SAM) as the methyl group donor (Fig. 1.13) (Bannister et al., 2002).

Lysine demethylases (KDMs) were discovered later, first LSD1, a component of the C-terminal binding protein corepressor complex which catalyses the demethylation of H3K4me2 and H3K4me1 through its

flavin adenine dinucleotide (FAD)-dependent amine oxidase domain (Shi et al., 2004). Another KDM enzyme class was later discovered that use the Jumanji C (JmjC) domain to catalyse demethylation through the oxidation of methyl groups. They rely on α -ketoglutarate, molecular oxygen and Fe²⁺ cofactors (Shi and Whetstine, 2007).

KMTs and KDMs rely on cofactors that are key intermediates from metabolic processes, suggesting that these enzymes may be responsive to metabolic states. S-adenosylhomocysteine (SAH) is a by-product of methylation reactions using SAM (Fig. 1.13) and can act as a competitive inhibitor of methyltransferases, potentially creating a negative feedback loop mechanism for regulation of methylation (Huang, 2002).

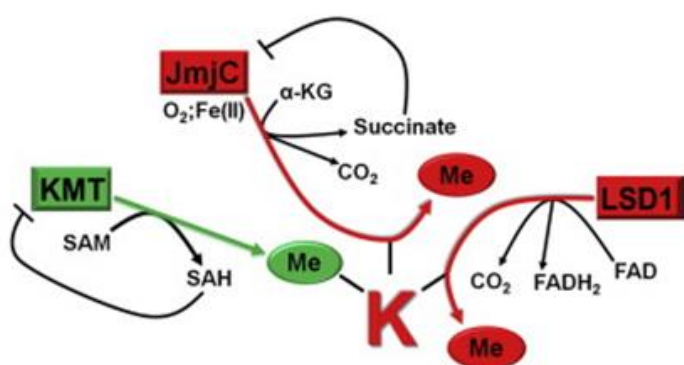


Figure 1.13 The reaction mechanisms of KMTs and KDMs. Adapted from (Black et al., 2012).

1.5.2.2 Discovery of the *C. elegans* homologue of yeast Set1, SET-2

The first Set1 protein in *C. elegans* was identified in a RNAi screen for enhancement of Mes sterility, where Mes stands for ‘maternal effect sterile’ i.e. progeny of Mes mutant mothers are sterile (Xu and Strome, 2001). The Mes phenotype was thought to be due to the derepression of gene expression in the germline, supported by the observation that high copy transgenic arrays, that are normally silenced in the WT *C. elegans* germline, were shown to be desilenced in the germlines of sterile *mes* mutants (Kelly and Fire, 1998). In repressing inappropriate gene expression, MES proteins in the *C. elegans* germline play a similar role to PcG proteins in the *Drosophila* soma. Indeed, MES-2 and MES-6 are the *C. elegans* orthologs of PcG proteins Enhancer of Zeste [E(Z)] and Extra Sex Combs (ESC). MES-3 is the remaining protein in the *C. elegans* Polycomb complex, responsible for H3K27 methylation. MES-4 is a SET-domain protein responsible for H3K36 methylation, a mark of actively expressed genes.

Progeny of homozygous *mes* mutant mothers are completely sterile but M+Z- *mes* mutants, homozygous zygotic mutants from a heterozygous mother, have a milder phenotype of reduced brood size, indicative of some germline defect and suitable to test for worsening with candidate enhancers.

The RNAi screen found a novel enhancer of *Mes* sterility, which was selected as a candidate for the screen due to its sequence similarity to *trithorax*. It was named *set-2* on account of the SET domain found at the C termini of its two isoforms. RNAi of *set-2* induced sterility in M+Z- *mes-3* and *mes-4* mutants (from 11-92% and 0-95% respectively in the F1 offspring of injected P0 mothers), but not *mes-2* or *mes-6* mutants, or WT worms (Xu and Strome, 2001).

1.5.2.3 *C. elegans set-2* gene structure

The *set-2* locus encodes two overlapping transcripts (Figure 1.14), one 6.5kb in length (*set-2_l*) and one 3.4 kb in length (*set-2_s*) (Xu and Strome, 2001). The three C-terminal exons of both transcripts encode the SET domain, but the RRM RNA binding domain is unique to the long transcript. The short transcript has a unique 5' exon within intron 9 of *set-2_l* (Xu and Strome, 2001).

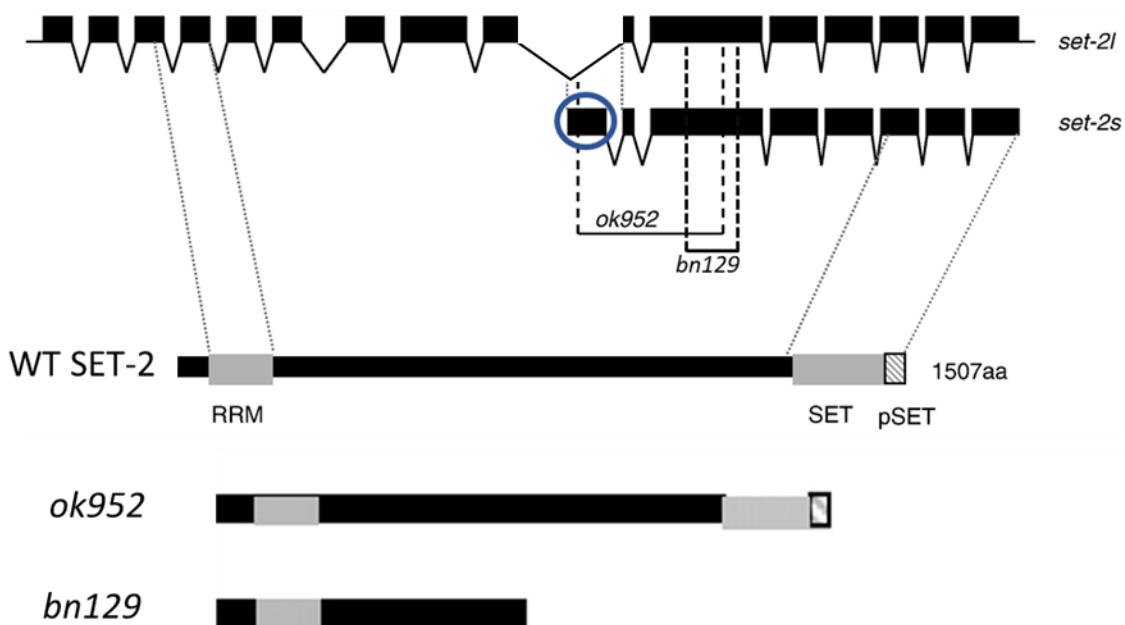


Figure 1.14 The gene structure of *C. elegans set-2*, showing the long and short transcripts, the exons which contribute to the key RRM and SET domains, the position of the *ok952* and *bn129* deletion alleles, and the effect of these deletions on the long form of the SET-2 protein. The unique N-terminal exon is encircled in blue. Adapted from (Simonet et al., 2007) and (Xiao et al., 2011).

1.5.2.3.1 *ok952*

The *set-2(ok952)* is a deletion allele, generated by Robert Barstead (strain name: RB1025) using UV/TMP mutagenesis (Barstead et al., 2012). It is a 1269bp deletion spanning 3 exons in *set-2_s*, 2 in *set-2_L* (Figure 1.14, Table 1.3). It is an out of frame deletion of SET-2_s, resulting in a stop codon at amino acid 38. In SET-2_L it removes 208 amino acids between the RRM and the SET domain (Simonet et al., 2007). Crucially, the SET domain in SET-2L is left intact, and potentially functional. Some functionality of the SET domain in SET-2L (*ok952*) is supported by the observation of no significant decrease in H3K4me3 (by quantitative western blot) in *C. elegans* embryos and L1s, and only a 20-40% reduction in L3s and L4s (Xiao et al., 2011). Therefore, it is thought to be a hypomorphic allele.

1.5.2.3.2 *bn129*

Also a deletion allele, *set-2(bn129)* was identified by the Strome laboratory in a PCR screen of a deletion library generated by the Koelle method (Xiao et al., 2011). It removes 745bp from exon 11 of *set-2_L*, and from exon 3 of *set-2_s* (Figure 1.14, Table 1.3). This results in a frameshift after 885 amino acids of SET-2L and after 117 amino acids of SET-2s, leading to a premature stop codon four amino acids later in both proteins, before the SET domain can be translated. This effect is predictive of a null allele and is supported by observations of a more drastic reduction of H3K4me3 and H3K4me2 in *set-2(bn129)* mutants, 70-80% and 50-70%, respectively, at all stages of development (Xiao et al., 2011).

Table 1.3 – Summary of the *ok952* and *bn129* alleles and their effects on the SET-2 protein. Information from (Simonet et al., 2007) and (Xiao et al., 2011).

	<i>ok952</i>	<i>bn129</i>
Size (bp)	1269	745
Effect on SET-2S	Out of frame, premature stop codon at aa38	Out of frame, premature stop codon at aa121
Effect on SET-2L	In frame, truncated protein with 208 aa missing between RRM and SET domain	Out of frame, premature stop codon at aa889

1.5.3 Cfp1/CFP-1

1.5.3.1 Discovery

Cfp1/CFP-1 plays a targeting role in the COMPASS complex. The first cloning of Cfp-1 protein was achieved by a ligand screening approach, in which it was isolated as a human cDNA encoding a novel CpG (cytosine preceding guanine) binding protein, which was found to function as a transcriptional activator (Voo et al., 2000). Human CFP1 contains two PHD domains, acidic and basic regions, a coiled-coil domain and a CXXC domain. The CXXC domain is common to a number of DNA binding proteins including DNA methyltransferase 1, methyl-CpG binding domain protein 1, and human trithorax. An alanine-scanning study found that the two conserved cysteine residues in the CXXC motif are essential for DNA binding, as is the presence of Zn^{2+} (Lee et al., 2001). The consensus binding site is (A/C)CpG(A/C) and while mutation of the flanking residues to thymine does not abolish binding to the CXXC domain, it reduces the efficiency. CpG alone is necessary and sufficient for CFP1 binding.

Native human CFP-1 was shown to *trans*-activate promoters that contain CpG motifs, pointing to a role in modulating the expression of CpG island genes.

1.5.3.2 CpG islands

CpG, shorthand for cytosine preceding a guanine residue in DNA, is rare in vertebrate DNA. In humans, it occurs with a frequency almost 10 times less than would be expected from base composition. There appears to be an inverse relationship between the extent of methylation in the genome and CpG content, with vertebrates in the 'heavy methylation' category, echinoderms with partial methylation and insects with undetectable methylation (Bird, 1980). The reduction of CpG content in heavily methylated genomes is thought to be due partly to the heightened mutability of 5-methylcytosine, which can be deaminated to give thymine, and thus over time converting C-G to A-T pairs (Tykocinski and Max, 1984).

Non-methylated CpG is rarer still, accounting for 10-40% of total CpG. First described as '*HpaII* Tiny Fragments (HTF) islands' on account of the methylation-sensitive restriction enzyme, *HpaII*, which first identified them (Cooper et al., 1983), non-methylated CpG sequences were thought to mark 'housekeeping' genes, that need to be constantly accessible to transcriptional machinery (Bird, 1986).

CpG islands are usually but not always associated with the 5' end of the coding region, often overlapping with annotated transcription start sites, which suggested potential involvement in promoter function. An early model proposed that CpG islands keep neighbouring genes available by binding proteins which sterically prevent methylase from accessing and methylating the underlying CpGs (Bird, 1986). More recent

research has supported this with, for example, the characterisation of highly-occupied-target (HOT) regions in *C. elegans* (Chen et al., 2014). These are CpG-rich promoters in *C. elegans* and human genomes, characterised by the binding of a large number of transcription factors and enriched for binding of CFP-1/Cfp1. Therefore, non-methylated CpG sequence appears to be a conserved genomic signal, targeted for H3K4me3 by a CXXC1 orthologue, that promotes an open chromatin state.

1.5.3.3 Function

After its discovery in humans, homologues of CFP-1 were identified in other systems including *D. melanogaster*, *C. elegans*, *S. cerevisiae* and *S. pombe*. Sequence alignment revealed that the CXXC domain is not present in yeast or *C. elegans*, organisms which do not have cytosine methylation. Without the CXXC domain, can CFP-1 still be binding DNA in these organisms? Perhaps it is binding methylated and/or acetylated lysines with its PHD domain instead. Research is still needed to establish the role of CFP-1 in organisms which lack DNA methylation and if it has a conserved mechanism of action in all eukaryotes.

In the *S. cerevisiae* study that first discovered COMPASS (Miller et al., 2001), the authors describe a protein of 40kDa with homology to mammalian CFP1, containing a conserved region found in human trithorax (MLL). They called it Cps40. Mutants for this protein showed the same phenotypes as Set1 mutants, slowed growth, defective telomere silencing and sensitivity to hydroxyurea, indicative of a role in DNA replication, transcription or chromatin remodelling. Importantly, the authors were unable to detect binding by purified *S. cerevisiae* COMPASS to the DNA sequence reported by (Voo et al., 2000) in their ligand screening study. This supports Cps40 and its mammalian homologue, Cfp1, having different sequence specificities for DNA binding.

A study published that same year named this protein Spp1p, and found that it has a PHD finger domain (Roguev et al., 2001). PHD finger domains are able to recognise methylated and acetylated lysines and thus may serve as 'readers' of the histone code. The PHD domain in Spp1p is closely related to the same domain in *S. pombe* and human homologues, which would suggest a common and conserved function.

Perhaps the PHD domain in mammalian Cfp1 was overlooked due to the more obvious role of the CXXC domain in non-methylated CpG binding. However, experimental evidence from mammalian systems suggests not all functionality of Cfp1 is linked to the CXXC domain. Severe phenotypic consequences of CFP1 knockout in mammalian systems such as embryonic lethality (Carlone and Skalnik, 2001), cannot be accounted for by the reduction in cytosine methylation alone (Carlone et al., 2005). Inducing this artificially, for example by Dnmt1 knockout, does not produce as severe effects (Lei et al., 1996).

There is some evidence in the literature for CFP1 providing a restrictive role in the genomic distribution of H3K4me3. CXXC^{-/-} mouse ES cells showed a 4-fold increase in H3K4me3 following the induction of differentiation with growth factor removal, whereas wild-type controls showed a slight decline (Lee and Skalnik, 2005). In a later study, Cfp1-deficient ES cells had ectopic H3K4me3 accumulation at sites enriched for CTCF and cohesin binding, purported chromatin loops. A DNA-binding domain deficient CFP1 could rescue the loss of H3K4me3 at promoters in these cells, but not the ectopic accumulation (Clouaire et al., 2012).

Taken together, these observations support an ancestral role for CFP-1, independent of cytosine methylation but perhaps regulating chromatin organisation by some other manner.

1.5.3.4 CFP-1 in *C. elegans*

The *C. elegans* CFP-1 protein was discovered in an RNAi screen for suppressors of *hpl-1* and *hpl-2* mutant phenotypes (Simonet et al., 2007). HPL-1 and HPL-2 are the *C. elegans* HP1 proteins, whose conserved function is to interact with methylated H3K9 to facilitate the formation of heterochromatin (Couteau et al., 2002). H3K9 methylation is performed by the Su(var)3-9 histone methyltransferase, the first demonstrated to have histone methyltransferase activity (Rea et al., 2000).

In *C. elegans*, HPL-1 and HPL-2 play partially redundant roles in post-embryonic development. Neither is essential, but *hpl-2* mutants are defective in growth, somatic gonad development and germline function (Coustham et al., 2006), and the *hpl-1; hpl-2* double mutant is larval lethal (Schott et al., 2006).

Initially, RNAi inactivation of *set-2*, *mes-2* and *mes-4* was found to restore normal body size and growth in *hpl-2* mutants (Simonet et al., 2007). Of these, *mes-2* and *mes-4* are involved in H3K27 and H3K36 methylation, respectively. Little was known at the time about the role of H3K4 histone methyltransferase complexes in animal development, and thus *set-2* was selected as the candidate for further study.

Subsequent testing of *C. elegans* homologues of yeast and mammalian SET1/MLL complex subunits found that RNAi inactivation of the Spp1/CFP1 homologue, initially named *cfpl-1*, could also suppress *hpl-2* growth defects and *hpl-1;hpl-2* larval arrest like *set-2* (Simonet et al., 2007).

Interestingly, the study found RNAi inactivation of *cfp-1*, but not *set-2*, could suppress the Muv phenotype of *hpl-1;hpl-2* mutants, suggesting that CFP-1 may antagonise the function of SynMuv genes in vulval cell fate determination by a SET-2/COMPASS-independent mechanism.

1.6 Thesis aims

This thesis aims to better characterise the *set-2(bn129)* and *cfp-1(tm6369)* strains, using LC-MS/MS to assess the impact of these deletion mutations on the whole histone H3 modification landscape. Any differences between the results obtained for the two strains will be informative in the context of recent research suggesting novel roles for CFP-1 independent of the COMPASS complex (Pokhrel et al., 2019, Beurton et al., 2019).

This analysis will also confirm whether the *set-2(bn129)* allele truly is loss of function, as indicated previously by less quantitative antibody-based approaches (Xiao et al., 2011). This might help settle debate in the *C. elegans* COMPASS research community over whether *ok952* or *bn129* is the better allele to use in loss of function studies.

Leading on from this, the thesis will explore the transgenerational epigenetic inheritance of extended longevity reported in WT descendants of *set-2* mutants (Greer et al., 2011). If the findings from this study, using the *set-2(ok952)* (putative hypomorph) mutant, can be reproduced using the *set-2(bn129)* strain, it would give greater confidence in the earlier findings.

Moreover, after establishing a successful LC-MS/MS pipeline for the quantification of histone PTMs in mixed embryo-derived histones, it will be possible to modify the protocol for the analysis of germline-derived histones. This will facilitate the comparison of long-lived WT descendants of *set-2(bn129)* mutants with homozygous *set-2(bn129)* offspring and WT controls, to identify any perturbations in histone modifications which coincide with the increased longevity and return to basal WT control levels when the phenotype ceases to be inherited.

Chapter 2. Materials and Methods

Table 2.1 - Chemical list with suppliers

Chemical	Supplier
Acetone	Thermo Fischer Scientific
Acetonitrile (ACN)	Thermo Fischer Scientific
Acrylamide	Thermo Fischer Scientific
Ammonium bicarbonate (ABC)	Sigma-Aldrich
Ammonium hydroxide	Sigma-Aldrich
Ammonium persulfate (APS)	Sigma
Arginine	Sigma
Arginine 10	CK Isotopes
Blue pre-stained standard Broad range protein ladder	New England BioLabs
Bovine serum albumin (BSA)	Sigma-Aldrich
Bradford reagent	Thermo scientific
Bromophenol blue	Sigma-Aldrich
Calf histone standard	Sigma-Aldrich.
Chitinase lyophilised powder	Sigma-Aldrich
Colloidal Coomassie brilliant blue dye	Sigma-Aldrich
CoverGrip coverslip sealant	Biotium
DAPI	Thermo Fischer Scientific
Dimethyl sulfoxide (DMSO)	Sigma - Aldrich
Dithiothreitol (DTT)	Sigma-Aldrich
Enhanced Chemiluminescent (ECL) detection reagent	Sigma-Aldrich
Ethanol	Thermo Fischer Scientific
Ethidium Bromide	Sigma-Aldrich
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich
Formic acid	Sigma-Aldrich
Glacial acetic acid	Thermo Fischer Scientific
Glycerol	Sigma-Aldrich
Glycine	Thermo Fischer Scientific
HPLC grade water	Thermo Fischer Scientific
HT supplement	Thermo Fischer Scientific
Iodoacetic acid (IAA)	Sigma-Aldrich
Isopropanol	Thermo Fischer Scientific
L-Glutamine	Gibco, Life technology
Lysine	Sigma
Magnesium chloride (MgCl ₂)	Sigma-Aldrich

Methanol	Sigma-Aldrich
NP-40	Sigma-Aldrich
Nuclease-free water	Thermo Fischer Scientific
Phosphate buffered saline (PBS)	Sigma-Aldrich
Phusion Hot Start II High-Fidelity DNA polymerase	Thermo Fischer Scientific
Pierce protease and phosphatase inhibitor mini tablets	Thermo Fischer Scientific
Porcine pancreas trypsin	Sigma-Aldrich
Potassium chloride (KCl)	Sigma-Aldrich
Potassium phosphate (KPO ₄)	Sigma-Aldrich
ProlongGold anti-fade mounting medium	Thermo Fischer Scientific
Propionic anhydride	Sigma-Aldrich
Puromycin	Life technologies
Roche Complete EDTA free , Protease inhibitor	Roche
Sodium Butyrate (NaBu)	Sigma
Sodium chloride (NaCl)	Thermo Fischer Scientific
Sodium dodecyl sulphate (SDS)	Sigma-Aldrich
Sodium dodecyl sulphate running buffer	National Diagnostics
Sodium phosphate dibasic	Thermo Fischer Scientific
Sodium phosphate monobasic	Sigma
Sucrose	Sigma-Aldrich
Sulphuric acid (H ₂ SO ₄)	Thermo Fischer Scientific
TEMED	Sigma
Trichloroacetic acid (TCA)	Thermo Fischer Scientific
Trifluoroacetic acid (TFA)	Thermo Fischer Scientific
Tris-Cl pH 8.0,	Sigma-Aldrich
Triton	Thermo Fischer Scientific
Tween-20	Sigma-Aldrich
β-mercaptoethanol	Sigma-Aldrich

2.1 *C. elegans* maintenance

2.1.1 Standard growth conditions

Unless stated otherwise, *C. elegans* strains used in these experiments were maintained at 20°C on 5 cm plates containing nematode growth medium (NGM) (Table 2.2). The plates were seeded with an overnight LB culture of OP50, a uracil-requiring *E. coli* strain suitable as a food source for *C. elegans* due to its limited growth on NGM, thereby allowing easy observation (Brenner, 1974).

Table 2.2 – Buffer recipes

Buffer	Recipe
Egg buffer	HEPES pH 7.3 (5mM), NaCl (118mM), KCl (48mM), CaCl ₂ (3mM), MgCl ₂ (3mM)
Freezing buffer	KPO ₄ (50mM), NaCl (100mM), Glycerol (30%)
Hypotonic lysis buffer (Buffer A)	Tris. Cl (15mM), MgCl ₂ (2mM), Sucrose (0.34M)
Liquid culture bleach	Sodium hypochlorite (10%), NaOH (1M)
M9	Na ₂ HPO ₄ (33.7mM) KH ₂ PO ₄ (22mM) NaCl (8.55mM) MgSO ₄ (1mM)
Nematode growth medium (NGM)	Agar (1.7% w/v), NaCl (50mM), Peptone (0.25% w/v), CaCl ₂ (1mM), Cholesterol (5µg/mL), KH ₂ PO ₄ (25mM), MgSO ₄ (1mM)
Proteinase K/worm lysis reaction buffer	Proteinase K (20ng), 20µL 5x HF Phusion PCR buffer, 75µL autoclaved water (for 8x 10µL reactions)
SDS-PAGE sample loading buffer	Tris-Cl (62.5mM), SDS (2.5%), Bromophenol Blue (0.1%), glycerol (10%), DTT (10mM)
Phusion Hot Start PCR master mix	Phusion Hot Start II High-Fidelity DNA polymerase (0.2U/µL), 5X Phusion HF buffer (1X), dNTPs (200µM), forward primer (0.5µM), reverse primer (0.5µM)

Phosphate buffered saline with 0.1% Tween (PBST)	NaCl (137mM), KCl (2.7mM), Na ₂ HPO ₄ (10mM), KH ₂ PO ₄ (1.8mM) and Tween-20 (0.1%)
Tris buffered saline with 0.1% Tween (TBST)	Tris-HCl (20mM), NaCl (150mM) and Tween-20 (0.1%)

2.1.2 Synchronization and/or decontamination of strains by bleaching

A standard NGM/OP50 5cm plate with many gravid YAs was flushed with 1ml sterile water to facilitate the transfer of worms, via pipette, into a 1ml Eppendorf tube. Worms were resuspended in a 1:1 ratio of water to liquid culture bleach (Table 2.2) and vortexed continuously for at least 5 mins, until a colour change to yellow indicated that most embryos had been released. The bleach was then diluted out by two washes in sterile water and one in M9 (Table 2.2). Embryos were left to hatch in 7ml M9 overnight, being repeatedly tilted in a 15ml falcon tube. After hatching, L1s were centrifuged (2 mins at 2000 rpm) and resuspended in 1 ml M9, before being pipetted into fresh NGM/OP50 plates.

2.2 Single worm PCR genotyping

Individual YA worms were picked into PCR tubes containing 10µL of Proteinase K/worm lysis reaction buffer (Table 2.2). Worms were lysed in a thermocycler, under the conditions specified (Table 2.3). The lysates were refrigerated until 2.5µL from each were added to 12.5µL Phusion Hot Start PCR master mix (Table 2.2). Thermocycling conditions for genotyping each strain (Table 2.4) and the primers used (Table 2.4) are specified below.

Table 2.3 – Cycling conditions used in PCR genotyping

Protocol	Cycling conditions
Proteinase K/ worm lysis	50°C 90 mins 98°C 15 mins 4°C hold
<i>set-2(bn129)</i> genotyping	98°C 5 mins 98°C 10 seconds 56°C 20 seconds 72°C 1 min 30 seconds 72°C 5 mins 4°C hold
<i>cfp-1(tm6369)</i> genotyping	98°C 5 mins 98°C 10 seconds 58°C 20 seconds 72°C 1 min 30 seconds 72°C 5 mins 4°C hold

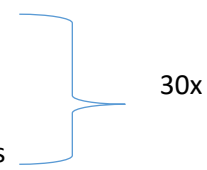


Table 2.4 – Primers used in PCR genotyping

Allele	Forward Primer sequence (5'-3')	Reverse Primer sequence (5'-3')	Expected PCR product sizes
<i>set-2(bn129)</i>	AGAGCACCATCATCATGCGAA	TTGGTTGGTGGTGGTTCATAAT	WT 969 bp Mutant 220 bp
<i>cfp-1(tm6369)</i>	AACACCAAGAGCACCATCATC	CGTAGAAAGCATCTGGCAGTC	WT 690 bp Mutant 440 bp

2.3 Mass spectrometry

2.3.1 Histone extraction

2.3.1.1 Mixed embryos

2.3.1.1.1 Stock Expansion

A chunk of agar from a crowded 5 cm diameter plate (Fig. 2.1) was transferred to 3 x 9 cm diameter plates and the worms grown to adulthood. Young adults (YAs) from all 3 plates were collected in sterile water, centrifuged (2 mins at 2000 rpm) and resuspended in a 1:1 ratio of sterile water to liquid culture bleach (Table 2.2). Bleaching took place as described in 2.1.2, after which the hatched L1s were transferred in equal volumes to 6 x 15 cm diameter plates. L1s were left to grow to young adulthood (60 hours for N2 and 72 hours for *set-2* and *cfp-1* mutant strains on account of their slower development).

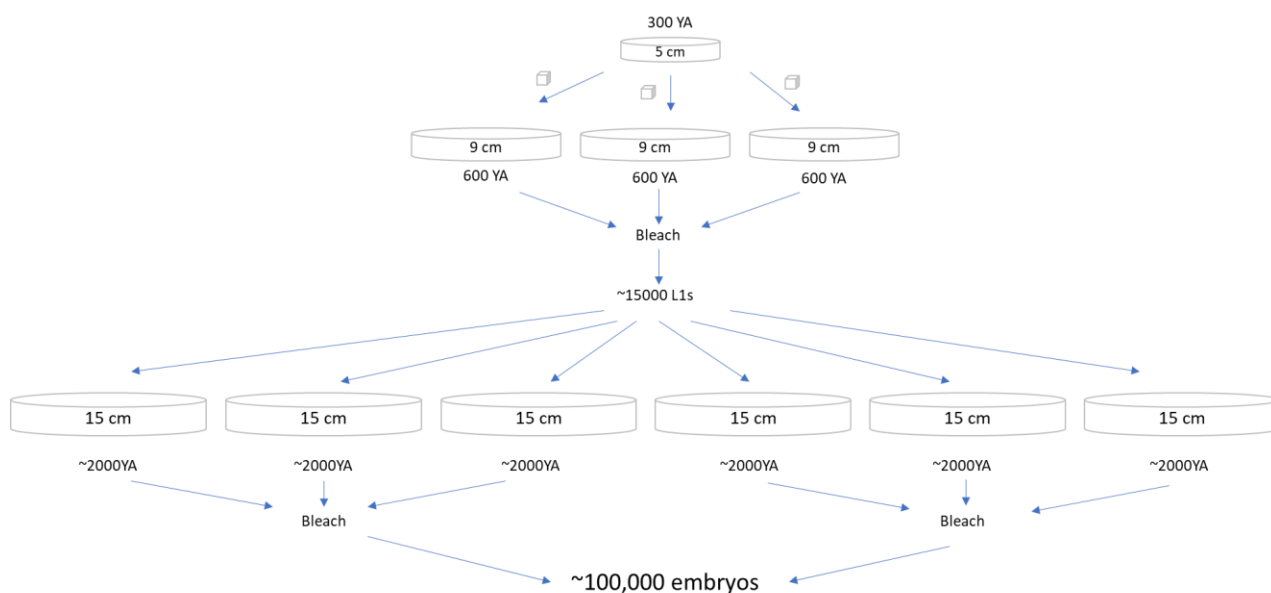


Figure 2.1 - The expansion of *C. elegans* stock to yield sufficient embryos prior to extraction. The estimated numbers assume a yield of ~80% with each collection and bleaching.

2.3.1.1.2 Final Embryo Collection

YAs were collected in sterile water from the 15 cm plates, washed once in M9 + 0.1% Triton and bleached as described previously. The pack volume of YAs was recorded to give a rough measure of the size of each collection. Recording the pack volume of embryos was more challenging owing to their tendency to stick to the centrifuge tube walls, the *cfp-1* mutant in particular.

2.3.1.1.3 Sucrose Floating

After the final M9 wash post-bleaching, embryos were resuspended in a 1:1 ratio of M9 to 60% sucrose and centrifuged at 1000 rpm for 5 mins. Live embryos were retrieved in 1 ml of M9 + 0.1% Triton by adding this gradually to the floating layer of embryos and then drawing it up again while gently swirling the pipette tip at the liquid interface. The retrieved embryos were washed twice in M9 + 0.1% Triton and once in M9. After the final wash, embryos were resuspended in 0.5 ml M9, of which 5 μ L were taken to fix in 1 ml methanol (for later embryo staging 2.3.1.1.9), and another 5 μ L were spotted to an unseeded 6 cm NGM plate to perform a hatching assay (2.3.1.1.8).

2.3.1.1.4 Chitinase Treatment

Sucrose-floated embryos were resuspended in a 1:1 mixture of M9 with 1 unit/ml Chitinase (lyophilised powder resuspended in M9) and shaken at regular intervals over a period of 10 minutes, while checking for a change in embryo shape indicating the breakdown of the chitin shell.

2.3.1.1.5 Nuclei Extraction

Chitinase-treated embryos were washed twice in 10 ml of hypotonic lysis buffer (Tris-Cl 15mM, MgCl₂ 2mM, Sucrose 0.34M) with added protease and phosphatase inhibitors (Pierce mini tablets, Thermo Fisher), as described (Ooi et al., 2010). After the second wash, cells were left to swell on ice in 10 ml hypotonic lysis buffer, then pelleted in a refrigerated centrifuge (2000 g, 10 mins, 4°C). The pellet was resuspended in 7 ml hypotonic lysis buffer containing 0.1% NP-40, a mild detergent. The suspension was ground in a chilled 15 ml glass douncer (Sigma-Aldrich) for 20 strokes, and then centrifuged (100 g, 5 mins, 4°C) to pellet nuclear debris. The supernatant was transferred to a new centrifuge tube and centrifuged (3000 g, 10 mins, 4°C) to pellet the nuclei.

2.3.1.1.6 Histone Recovery by Acid Extraction

Following the protocol provided by the Dickman Group, University of Sheffield, the pelleted nuclei were resuspended in 400 µL 0.2 M sulphuric acid, transferred to a lo-bind Eppendorf tube and placed horizontally on ice on a shaker for at least 4 hours (to solubilise the histones). The contents were centrifuged (16000 g, 10 mins, 4°C) to pellet any nuclear debris. The supernatant was transferred to a new tube and 142 µL of trichloroacetic acid (TCA) was added dropwise, with inversion to mix, before incubating overnight at 4°C (to precipitate the histones). The following morning, histones were pelleted (17000 g, 20 mins, 4°C). The pellet was washed twice in 1 ml acetone, then left to dry completely before resuspension in 100 µL nuclease-free water. The histones were then frozen and stored at -80°C.

2.3.1.1.7 SDS-PAGE quality check and concentration determination

Premade SDS-PAGE gels (NuPage Bis-Tris 4-12%) were run to assess quality, purity and concentration of histone extracts (Fig. 2.2). A 10µL aliquot from each 100µL sample was transferred to a new Eppendorf tube, denatured for 5 mins at 100°C in SDS-PAGE sample loading buffer (Tris-Cl 62.5mM, SDS 2.5%, Bromophenol Blue 0.1%, glycerol 10%, DTT 10mM), and electrophoresed at 180 V in NuPage MES SDS running buffer for 30-40 mins. Gels were stained with Coomassie brilliant blue stain (Coomassie R-250 0.1%, ethanol 40%, acetic acid 10%) for 1 hour and de-stained overnight in de-stain solution (ethanol 10%, acetic acid 7.5%).

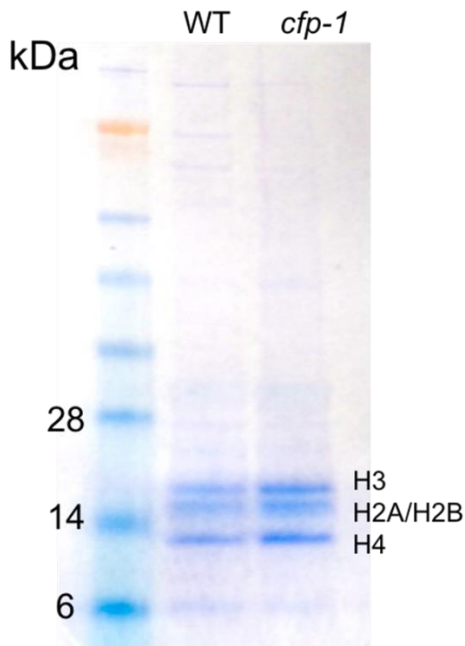


Figure 2.2 –Coomassie stained SDS-PAGE gel showing bands for histones H3, H2A/H2B and H4 from two mixed-embryo-derived nuclear histone extracts. Absence of any other bands indicates purity of the sample. The strength of staining gives a rough indication of the concentration of histones in the sample, which is necessary to decide how large an aliquot to take forward to the propionylation step.

2.3.1.1.8 Hatching Assay

After sucrose floating, a 5 μ L aliquot of embryos was spotted to an unseeded 6 cm diameter NGM plate and left overnight at 20°C. The next day the proportion of eggs that had hatched was estimated. An acceptable rate was set at >90%.

2.3.1.1.9 Embryo Staging

Embryos stored in the -20°C freezer in methanol were washed once in M9, then resuspended in 1mL DAPI (1 μ g/mL) for 10 minutes. The stained embryos were washed in M9 and viewed under the microscope. Each embryo observed was assigned to one of six categories (Fig. 2.3): '4 cells and under', '8-26 cells', '26-200 cells', '200-500 cells', '500 cells-bean stage', 'comma stage +'. Later, the 'comma stage +' category was subdivided to include the 1.5-fold, 2-fold and 3-fold stages, to better represent those collections with a greater proportion of later stage embryos.

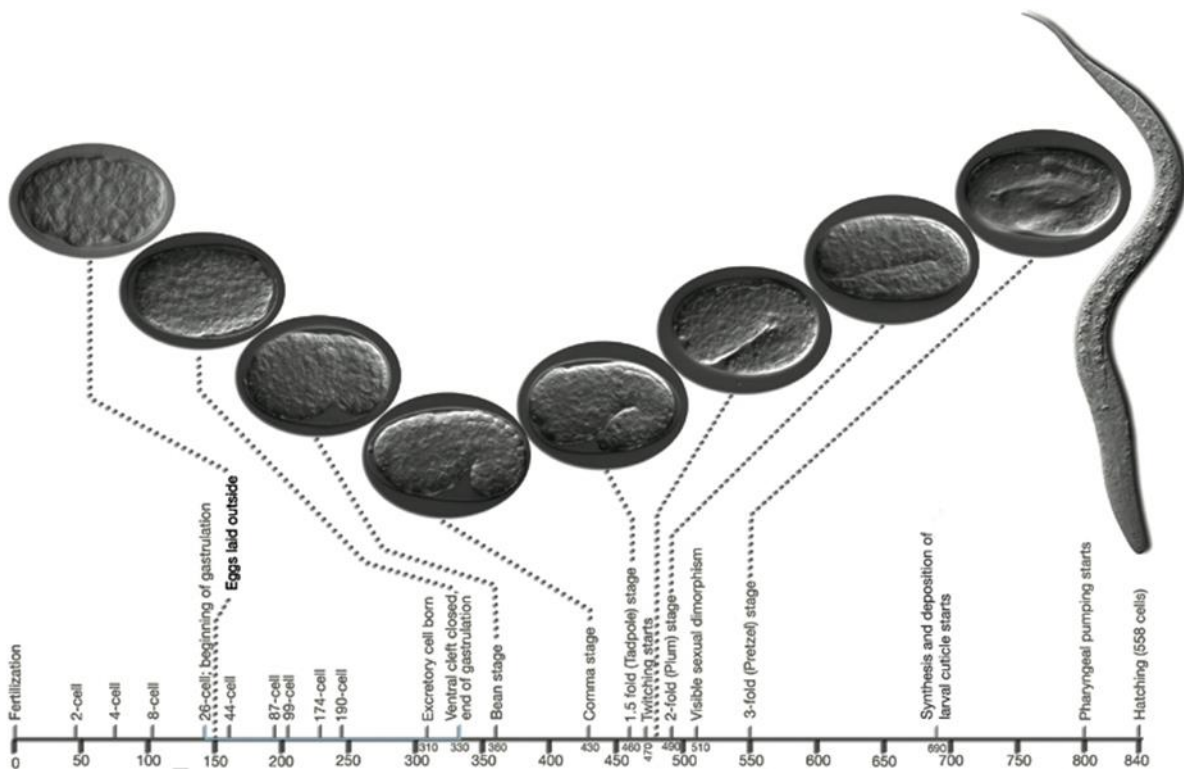


Figure 2.3 – The stages of development of a *C. elegans* embryo, taken from wormatlas.org

2.3.1.2 Germlines

2.3.1.2.1 Transgenerational collection.

Two weeks in advance of the planned F3 gonad collection, a *set-2* YA hermaphrodite, of between 10 and 15 generations post-outcrossing, was crossed with an N2 male (Fig. 2.4). A simultaneous cross between WT hermaphrodites and males was performed to prepare the control WT F3. Of the *set-2* cross progeny, the F1, 6-8 L4 hermaphrodites were picked to individual NGM plates, a process known as singling. They were allowed to lay eggs for 24 hours and genotyped by single-worm PCR (2.2). After the PCR results confirmed which individuals were heterozygous for *set-2*, one heterozygote was selected. Its progeny, the F2, were randomly singled to 16 plates (Fig. 2.4), allowed to lay eggs for 24 hours as with the F1, and genotyped. Of these 16, it was likely, according to Mendelian segregation, that 4 would be WT, 4 homozygous mutant and 8 heterozygous. Even with the variation expected by chance, there should be enough F3 progeny from the WT and homozygous mutant plates for dissecting. Of the F3, 4 WT and 4 mutant L4s were picked to a new plate to lay the F4 and genotyped afterwards for extra verification. The same occurred for each generation thereafter.

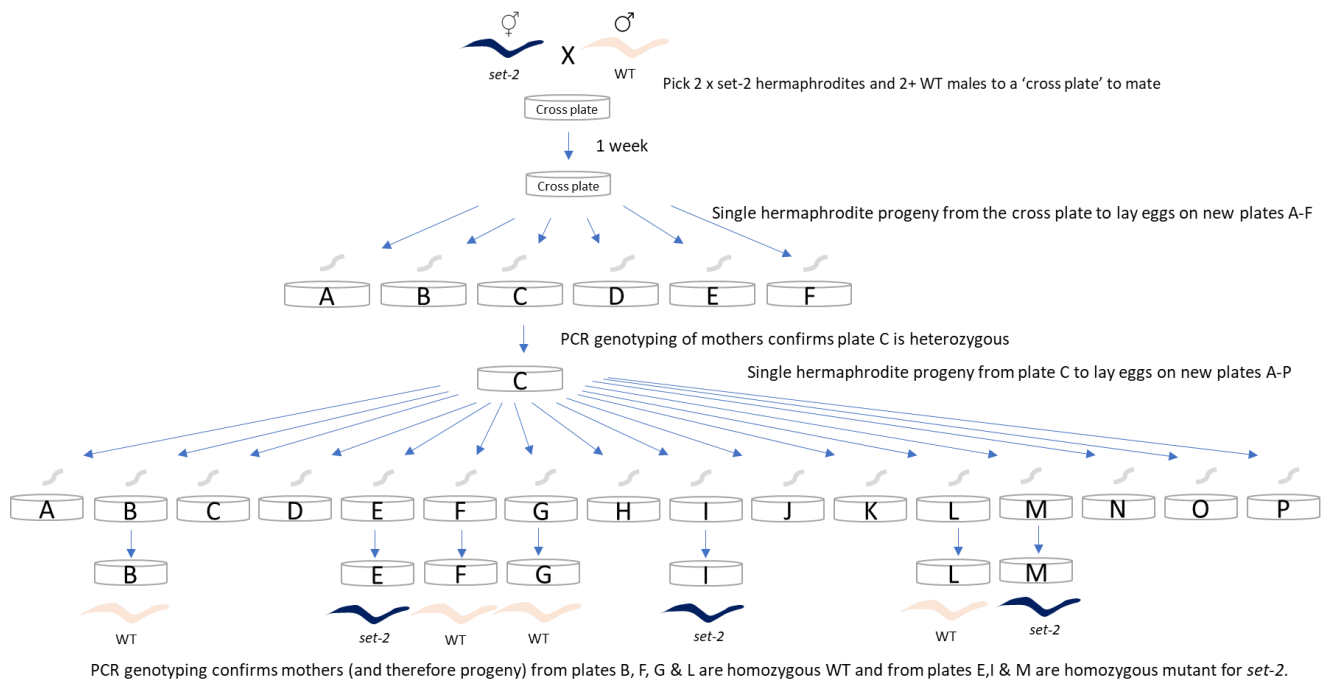


Figure 2.4 – The crossing and genotyping procedure used to isolate WT and mutant descendants of *set-2*, used for the transgenerational lifespan and mass spectrometry studies

2.3.1.2.2 Dissection

YA hermaphrodites were picked from plates into a 5µL of Egg Buffer (Table 2.2). Cuts were made as illustrated (Fig. 2.5) using a needle, first at the head end of the worm to release the gonad from the main body, and second at the spermatheca to detach the germline completely. Intact germlines were retrieved using an eyelash picker and transferred to 100µL Buffer A (Tris-Cl 15mM, MgCl₂ 2mM, Sucrose 0.34M) + 0.1% NP-40 (Pierce protease and phosphatase inhibitor tablets were added just before use, 1 tablet per 10ml buffer) in the cap of an Eppendorf tube placed on ice. After retrieving between 30 and 50 germlines, the cap was closed and the germlines centrifuged down in the buffer, before freezing at -80 °C. When >100 germlines had been obtained for each condition, over multiple dissection sessions, the collections were thawed, pooled together and briefly sonicated in a chilled water bath sonicator (30 second pulse, 60% amplitude) before resuspending in 400 µL 0.2 M H₂SO₄, and proceeding with histone extraction as described in 2.3.1.1.6.

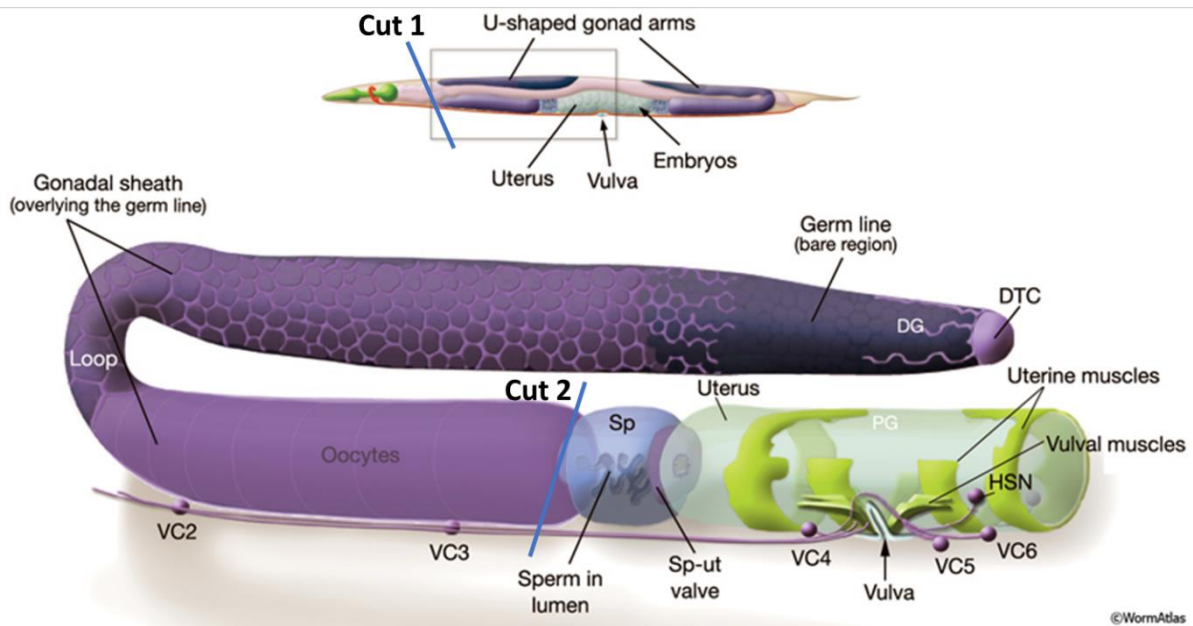


Figure 2.5 – The cuts made to obtain an intact gonad for later histone extraction. Upper image. *C. elegans* YA, with a thick blue line labelled ‘Cut 1’ indicating where the first cut was made to release the gonad from the body. Lower image. *C. elegans* gonad, with a second blue line labelled ‘Cut 2’ indicating where a second cut is made, just before the spermatheca, to separate the germline from the body and enable its retrieval as an intact structure.

2.3.2 Proprienylation of Histone Samples

2.3.2.1 First round proprienylation of nuclear histone extracts

As described by (Garcia et al., 2007), histone samples (containing 7-10 μg of histone protein) were diluted in a 1:1 ratio with 100mM ammonium bicarbonate (ABC) buffer (pH 8), vortexed, then treated with 4 μL ammonium hydroxide (NH_4OH). 10 μL of proprienylation reagent (1:3 ratio proprienic anhydride to isopropanol) was added to each histone sample, and the pH checked using universal indicator paper. If too acidic, additional $\text{NH}_3(\text{aq})$ was added to adjust the pH to 8 (see 1.3.2.2 for explanation of the importance of pH control in proprienylation). The reaction was incubated at 37°C for 15 minutes, after which the sample was dried down to 5 μL in a SpeedVac concentrator at room temperature. This process was repeated a second time to ensure at least a 95% conversion of amino groups to proprienyl amides.

2.3.2.2 Trypsin Digest of Propionylated Histone Samples

20 mg Trypsin was reconstituted in 20 μ L 0.1M HCl, vortexed and centrifuged down to mix. Dried, twice-propionylated histone samples were diluted in 40 μ L of 100mM ABC and 1.5 μ L of reconstituted Trypsin was added to each. The samples were then vortexed and centrifuged to mix and incubated at 37°C overnight. The reaction was stopped by adding 4 μ L of glacial acetic acid to each sample and leaving them on ice for 1 hour.

2.3.2.3 Propionylation of Trypsin-digested Histone peptides

After quenching of the trypsin digest, samples were dried in the SpeedVac before repeating the derivatisation with propionic anhydride as described in 2.3.2.1. After two rounds of propionylation, samples were dried in the Speed Vac and stored at -80°C.

2.3.3 Loading

2.3.3.1 Purification of histone peptides/tip clean-up

Propionylated peptides were resuspended in 30 μ L 0.1% TFA. C18 Hypercarb tips (Minshull et al., 2016) were primed for binding peptides by drawing up and expelling 5 x 20 μ L elution solvent (90% ACN, 0.1% TFA), then 5 x 20 μ L binding solvent (0.1% TFA). Resuspended peptides were continuously aspirated, with a volume of 20 μ L, for 2 minutes (approx. 100 times), to absorb the peptides onto the Hypercarb tip. The tip was washed with 2 x 20 μ L binding solvent, and then peptides were eluted into a fresh lo-bind tube, in 20 μ L increments of elution solvent, aspirating 10 times before taking a new 20 μ L of solvent, reaching a total volume of 200 μ L. The ACN concentration was then reduced by adding 80 μ L binding solvent, then drying in the Speed Vac.

2.3.3.2 HPLC gradient

Desalted/purified peptides were resuspended in 0.05% HFBA (see 1.3.2.3 for a discussion of why TFA was unsuitable) and injected onto an Ultimate 3000 online capillary liquid chromatography system with a PepMap300 C18 trapping column (Thermo Fischer), coupled to a 50 cm x 75 μ m Easy-Spray PepMap C18 analytical column. Flow rate was set at 300 nL/min and column temperature was maintained at 40 °C (see 1.3.3). Following injection onto the trapping column, samples were washed with solvent A (3% ACN, 0.1% formic acid) for 1 minute. Peptides were eluted using a stepped gradient on the analytical column, with solvent B (80% ACN, 0.1% formic acid) rising from 3% to 8% in 5 mins, then from 8% to 25% over 55 mins, then 25% to 60% over 26 mins, then washed with 90% for 5 mins before resetting to 1%.

2.3.4 Data acquisition

Data acquisition was performed on the Orbitrap Q Exactive HF (Thermo Fischer, 1.2.5), in data independent mode. MS1 scans had a range of 300-1100 m/z with 60,000 resolution, an automatic gain control (AGC) target of 3e6 (number of ions to be in the detector Orbitrap) and a maximum fill time of 55 milliseconds. An MS1 scan was performed every 10 MS2 scans. MS2 resolution was 30,000, with an AGC target of 1e6. Isolation windows spanned the range 300-900 m/z, using 20m/z quadrupole isolation windows (1.2.6).

2.3.5 Raw data analysis

2.3.5.1 Skyline

Peptide identification and quantification was performed in Skyline (MacLean et al., 2010), with the aid of a spectral library (1.3.4.1) based on data-dependent (DDA) analysis of CHO histones, created by Eleanor Hanson of the Dickman group, University of Sheffield. Peptides derived from H3 were searched, using a list supplied to the Skyline software, provided by the Dickman group (Table 2.5). The peptides included are based on previous data from this group and the literature on common H3 modifications.

Identification takes into account MS2 data (b and y ions), retention time consistency between samples, and distributions of different charge states (1.3.4.1). When all peptide identifications had been checked manually, the area under the MS1 peak was extracted using the MS stats package.

Table 2.5 – List of proteoforms searched for in data-independent analysis

H3K4me0	H3K18ac0K23ac0
H3K4me1	H3K18ac1K23ac0
H3K4me2	H3K18me0k23me1
H3K4me3	H3K18me1K23me0
	H3K18ac0K23ac1
H3K9me0K14ac0	H3K18ac1K23ac1
H3K9me1k14me0	
H3K9me2K14ac0	H3K27me0K36me0
H3K9me3K14ac0	H3K27me0K36me1
H3K9ac1K14ac0	H3K27me0K36me2
H3K9me0K14ac1	H3K27me0K36me3
H3K9me1K14ac1	H3K27me1K36me0
H3K9me2K14ac1	H3K27me1K36me1
H3K9me3K14ac1	H3K27me1K36me2
H3K9ac1K14ac1	H3K27me1K36me3
	H3K27me2K36me0
H3K56ac	H3K27me2K36me1
H3K56me0	H3K27me2K36me2
	H3K27me3K36me0
H3K79me0	H3K27me3K36me1
H3K79me1	H3K27me3K36me2
H3K79me2	H3K27ac1K36me0

2.3.5.2 Calculation of relative abundance

MS1 peak areas (corresponding to peptide intensity) extracted by the MS Stats package were imported into Excel and divided by the appropriate correction factors (Table 2.6). The isotope correction factor corrects for the effects of isotope masking, when an isotopic peak of one peptide overlaps with the monoisotopic peak of another and could be misleadingly increasing its perceived amount (Lin et al., 2014). The relative abundance correction factor takes into account detection efficiencies of each peptide.

Depending on how peptides have their intensity split, different combinations of charge states are summed to calculate relative abundance. For example, the KQLATKAAR peptide (containing K18 and K23) has over 90% of its intensity found on one charge state, +2. Therefore, only +2 charge state intensities for each of the H3K18K23 proteoforms are taken into account in the relative abundance calculation, in which the intensity of each proteoform is divided by the sum of intensities of all proteoforms for the peptide. The TKQTAR peptide (containing K4) on the other hand, has its intensity split more evenly between the +1 and +2 charge

states. So, to calculate relative abundance of the different proteoforms for this peptide, the combined intensity of +1 and +2 for one proteoform is divided by the total intensity of all the +1 and +2 charge states for all proteoforms. See Table 2.7 for details of all the peptides.

Table 2.6 – Correction factors which peptides are divided by to correct for isotopic masking and detection efficiency. Published in (Lin et al., 2014). The numbers in square brackets next to the single letter amino acid residues refer to the mass increase added to that residue by propionylation, methylation, acetylation or some combination of these.

Proteoform	Sequence	Isotope correction factor	Relative abundance correction factors
H3K4me0	T[+56]K[+56]QTAR	1.14	0.0326
H3K4me1	T[+56]K[+70]QTAR	1.15	0.178
H3K4me2	T[+56]K[+28]QTAR	1.12	0.0018
H3K4me3	T[+56]K[+42]QTAR	1.13	0.0019
H3K9ac1K14ac0	K[+98]STGGK[+56]APR	1.25	0.378
H3K9ac1K14ac1	K[+98]STGGK[+42]APR	1.27	0.276
H3K9me0K14ac0	K[+112.1]STGGK[+56]APR	1.3	0.571
H3K9me0K14ac1	K[+112.1]STGGK[+42]APR	1.25	0.378
H3K9me1K14ac1	K[+126.1]STGGK[+42]APR	1.3	1.014
H3K9me1k14me0	K[+126.1]STGGK[+56]APR	1.31	1.295
H3K9me2K14ac0	K[+84.1]STGGK[+56]APR	1.28	0.303
H3K9me2K14ac1	K[+84.1]STGGK[+42]APR	1.26	0.177
H3K9me3K14ac0	K[+98.1]STGGK[+56]APR	1.23	0.252
H3K9me3K14ac1	K[+98.1]STGGK[+42]APR	1.23	0.156
H3K18ac0K23ac1	K[+112.1]QLATK[+42]AAR	1.3	1.43
H3K18ac1K23ac0	K[+98]QLATK[+56]AAR	1.3	1.43
H3K18ac1K23ac1	K[+98]QLATK[+42]AAR	1.33	1.711
H3K18me0K23me	K[+112.1]QLATK[+70]AAR	1.28	1.591
H3K18me1K23ac0	K[+126.1]QLATK[+56]AAR	1.38	1.21
H3K27ac1K36me0	K[+98]SAPATGGVK[+56]K[+56]PHR	1.84	2.305
H3K27me0K36me0	K[+112.1]SAPATGGVK[+56]K[+56]PHR	1.86	1.885
H3K27me0K36me1	K[+112.1]SAPATGGVK[+70]K[+56]PHR	1.83	1.87
H3K27me0K36me2	K[+112.1]SAPATGGVK[+28]K[+56]PHR	1.79	1.678
H3K27me0K36me3	K[+112.1]SAPATGGVK[+42]K[+56]PHR	1.81	1.762
H3K27me1K36me0	K[+126.1]SAPATGGVK[+56]K[+56]PHR	1.88	2.081
H3K27me1K36me1	K[+126.1]SAPATGGVK[+70]K[+56]PHR	1.79	1.377
H3K27me1K36me2	K[+126.1]SAPATGGVK[+28]K[+56]PHR	1.75	1.377
H3K27me1K36me3	K[+126.1]SAPATGGVK[+42]K[+56]PHR	1.88	1.67
H3K27me2K36me0	K[+84.1]SAPATGGVK[+56]K[+56]PHR	1.84	3.201
H3K27me2K36me1	K[+84.1]SAPATGGVK[+70]K[+56]PHR	1.71	2.288
H3K27me2K36me2	K[+84.1]SAPATGGVK[+28]K[+56]PHR	1.77	1.619
H3K27me3K36me0	K[+98.1]SAPATGGVK[+56]K[+56]PHR	1.79	3.176
H3K27me3K36me1	K[+98.1]SAPATGGVK[+70]K[+56]PHR	1.77	2.84
H3K56ac	Y[+56]QK[+42]STELLIR	1.54	0.512
H3K56me0	Y[+56]QK[+56]STELLIR	1.56	0.217
H3K79me0	E[+56]IAQDFK[+56]TDLR	1.61	0.125
H3K79me1	E[+56]IAQDFK[+70]TDLR	1.63	0.069
H3K79me2	E[+56]IAQDFK[+28]TDLR	1.59	0.548

Table 2.7 – Charge states factored into relative abundance calculations for each peptide

Peptide sequence	Proteoforms	Charge states included
TKQTAR	H3K4me0 H3K4me1 H3K4me2 H3K4me3	+1, +2
KSTGGKAPR	H3K9ac1K14ac0 H3K9ac1K14ac1 H3K9me0K14ac0 H3K9me0K14ac1 H3K9me1K14ac1 H3K9me1K14me0 H3K9me2K14ac0 H3K9me2K14ac1 H3K9me3K14ac0 H3K9me3K14ac1	+2, +3
KQLATKAAR	H3K18ac0K23ac1 H3K18ac1K23ac0 H3K18ac1K23ac1 H3K18me0K23me H3K18me1K23ac0	+2
KSAPASGGVKKPHR	H3K27ac1K36me0 H3K27me0K36me0 H3K27me0K36me1 H3K27me0K36me2 H3K27me0K36me3 H3K27me1K36me0 H3K27me1K36me1 H3K27me1K36me2 H3K27me1K36me3 H3K27me2K36me0 H3K27me2K36me1 H3K27me2K36me2 H3K27me3K36me0 H3K27me3K36me1	+2, +3, +4

2.3.5.3 Statistical tests

2.3.5.3.1 Mixed embryos

Relative abundance data were imported into GraphPad/Prism, using the Grouped table format, which allows replicate data to be placed in side-by-side sub-columns. WT, *set-2* and *cfp-1* results were assigned as groups, the columns' main titles. Each detected proteoform was defined as a family, with one family per row (Fig. 2.6).

Table format: Grouped		Group A				Group B				Group C			
		WT				<i>cfp-1</i>				<i>set-2</i>			
		A:1	A:2	A:3	A:4	B:1	B:2	B:3	B:4	C:1	C:2	C:3	C:4
1	H3K4un	42.93	63.40	40.69	50.27	72.30	90.61	69.48	76.21	77.42	91.69	78.55	77.33
2	H3K4me1	1.30	2.20	1.22	1.25	1.91	3.15	2.30	2.37	1.99	4.30	1.96	2.50
3	H3K4me2	39.56	22.29	35.61	30.17	24.91	6.09	25.76	20.45	19.99	4.01	18.66	19.27
4	H3K4me3	16.19	12.11	22.48	18.30	0.85	0.16	2.46	0.98	0.59	1.23e-003	0.82	0.90
5	Title												

Figure 2.6 - Example of how mixed embryo-derived relative abundance data was organised in Prism

A 2-way repeated measures ANOVA ($\alpha=0.05$), based on the generalised linear model, was selected as the most appropriate test. The data is balanced, which allows values in each sub-column to be matched for each replicate (e.g. A:4 to B:4 to C:4). Multiple comparisons, Tukey-corrected, were set up to compare *cfp-1* vs *set-2*, *cfp-1* vs WT and *set-2* vs WT, for each family/proteoform. Each p-value reported is a multiplicity-adjusted value.

2.4 RNAi

2.4.1 Experimental design

2.4.1.1 RNAi feeding plate preparation

RNAi clones (Table 2.8) were streaked onto plates containing ampicillin (50 μ g/ml) and tetracycline (10 μ g/ml) antibiotics. Plates were incubated at 37°C overnight, after which single colonies were inoculated in 2ml LB containing ampicillin (50 μ g/ml). The culture was placed in a shaking incubator at 37°C for 6h, then seeded onto NGM plates containing IPTG (1mM) and ampicillin (50 μ g/ml). After drying, plates were incubated at 37°C for 24h, and stored at room temperature thereafter.

Table 2.8 – Details of clones used in RNAi experiments. All are from the Ahringer library.

Gene KD	GenePairs name	Source Bioscience location
<i>hda-1</i>	C53A5.3	V-9F11
<i>hda-3</i>	RO6C1.1	I-6A01
<i>sin-3</i>	F02E9.4	1-4K03

2.4.1.2 *hda-3* and *sin-3*

Ten late L4-stage hermaphrodites from unsynchronized mixed plates maintained at 20 °C were picked to RNAi plates and left for 24hrs at 25 °C to lay eggs. After 24hrs they were transferred to fresh RNAi plates to lay eggs for another 24hrs, then picked off. The eggs laid on both sets of plates, hereafter referred to as the '24 hour plates' and '48 hour plates' respectively (Fig. 2.7), were left to hatch and grow to adulthood at 25 °C.

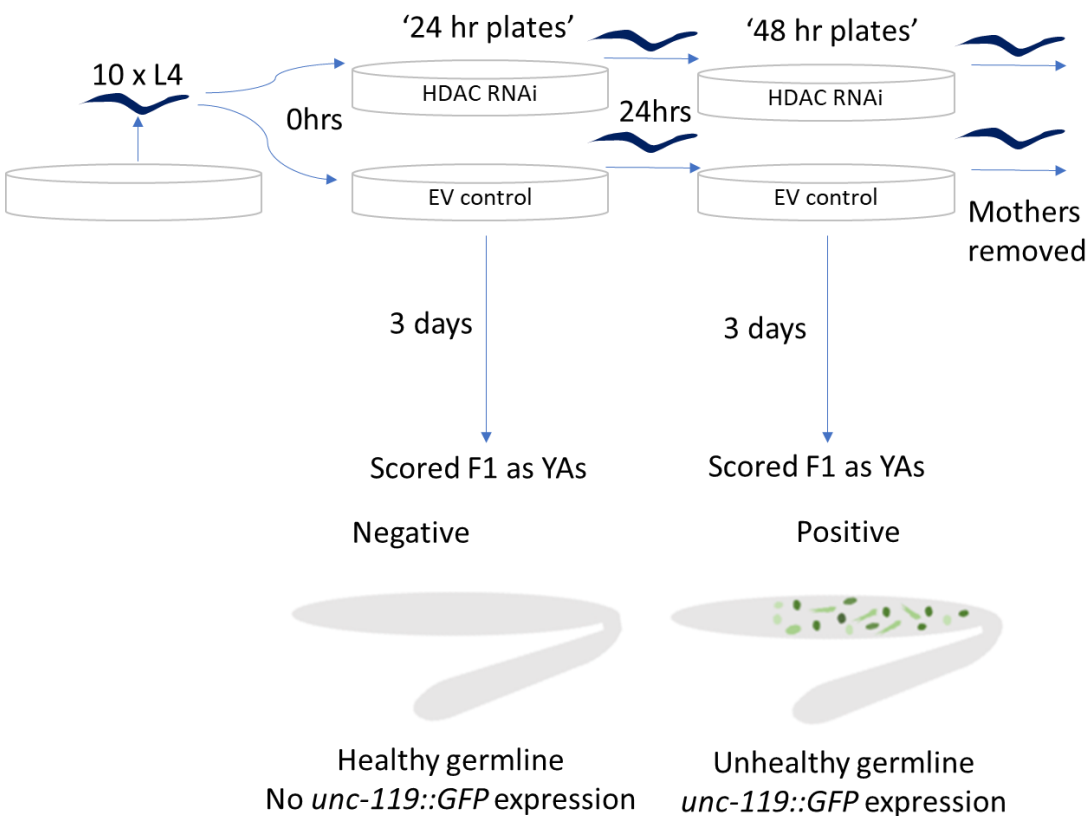


Figure 2.7 – Experimental design for *hda-3* and *sin-3* RNAi experiments. EV is an abbreviation for empty vector or negative control RNAi.

2.4.1.3 *hda-1*

Synchronised plates of YA were bleached as described in 2.1.2 to release embryos, which were left to hatch in M9 overnight, then placed as approximately 100 L1s onto RNAi plates. They were left to grow to young adulthood for two days at 25°C.

2.4.2 Slide preparation

Molten agar was flattened on a microscope slide to form pads to protect YAs during viewing. 10 YAs were picked into a 5µL drop of M9 + 1mM levamisole on a coverslip, which was then placed on an agar pad. The immobilised YAs were then viewed on a fluorescence microscope.

2.4.3 Scoring protocol

Both gonad arms in a YA were scored independently for GFP expression, unless so occluded by the gut or other structures that it could not be clearly observed. Gonads were scored as either positive or negative for GFP expression, defined as either the presence or total absence of GFP fluorescence in the distal arm of the gonad, prior to the turn. The 24 hour plates were scored first, then the 48 hour plates at the same time the next day.

2.4.4 Data analysis

The data, first recorded manually as the number of positive and negative gonads (2 per worm), were imported into Prism in a contingency table format, the two outcomes being 'GFP expressing' and 'Non-GFP expressing'. These raw numbers were converted into a 'Fraction of Total' table, representing the proportion of germlines in each category rather than an absolute number.

2.4.5 Statistical testing

Testing for significance was performed using a Chi-square test. The expected number/proportion of GFP-expressing germlines was calculated using a formula (Fig. 2.8) that assumes the null hypothesis that the two genes do not interact to affect germline *unc-119::GFP* expression. Under the null hypothesis, the expected germline *unc-119::GFP* expression of RNAi knockdown in a mutant background is the product of germline *unc-119::GFP* expression of the mutant on control empty vector (EV) RNAi and the WT on RNAi, divided by that of the WT on EV. In practice, the calculation was performed using the non-GFP expressing fraction, in order to avoid dividing by zero.

$$\frac{\text{WT on RNAi} \quad \times \quad \text{M on EV}}{\text{WT on EV}} = \text{M on RNAi}$$

Figure 2.8 - Formula for testing the null hypothesis that two genes will not interact in an RNAi experiment.

2.5 Lifespan assays

2.5.1 Set up

Crosses between *set-2(bn129)* and WT worms were set up, and F3 homozygous WT and *set-2(bn129)* mutant descendants were obtained as described in 2.3.1.2.1 and illustrated in Figure 2.4. A simultaneous cross between WT hermaphrodites and males was set-up, and one F1 and one F2 singled, to prepare the control WT F3. After PCR genotyping, the lifespan assay was set up by picking 12 worms to 9 NGM/OP50 plates for each condition: control WT F3, WT descendant F3 (F3+/+) and *set-2(bn129)* descendant F3 (F3-/-). The rationale for this set-up was that 12 worms per plate would be easy to track and that 108 worms per condition would comfortably give an n of at least 90, the n reported in (Greer et al., 2011), allowing for the small fraction (1/6) of the worms expected to be excluded from the final death count as a result of early death by vulval extrusion or the ‘bag of worms’ phenotype, when eggs are not released and hatch inside the mother.

2.5.2 Maintenance

Day 0 was set as the day of hatching, and worms were checked at the same time each day on day 2,4,6 and all even numbers thereafter until all worms had died. Worms were transferred to fresh plates every two days while laying eggs in order to avoid confusion of experimental animals with their progeny, and every four days thereafter to maintain adequate nutrition and reduce the likelihood of contamination of plates.

2.5.3 Exclusions

Worms were excluded from the final n for the following reasons: early death due to vulval extrusion or ‘bag of worms’, suspected injury sustained during transfer and loss due to either crawling up the side of the plate and drying out or burrowing into the agar. The numbers and percentages excluded for these reasons are presented in Table 2.8.

Table 2.8 – Exclusions from lifespan assays for reasons of vulval extrusion, bag of worms, loss or accidental death by injury sustained during transfer. The final n is the number of true deaths that were inputted into the final survival analysis.

Generation and genotype	First replicate					
	Vulval extrusion	Bag of worms	Lost	Killed in transfer	Total excluded	Final n
WT F3	32% (36/112)	1% (1/112)	11% (12/112)	0% (0/94)	44% (49/112)	63
WT F4	29% (27/94)	2% (2/94)	6% (6/94)	1% (1/94)	38% (36/94)	58
WT F5	29% (30/104)	1% (1/104)	14% (15/104)	1% (1/104)	45% (47/104)	57
F3 +/+	22% (22/102)	1% (1/102)	8% (8/102)	0% (0/102)	30% (31/102)	71
F3 -/-	23% (24/103)	1% (1/103)	4% (4/103)	4% (4/103)	32% (33/103)	70
F4 +/+	15% (15/101)	2% (2/101)	6% (6/101)	2% (2/101)	25% (25/101)	76
F4 -/-	13% (13/97)	3% (3/97)	15% (15/97)	0% (0/97)	32% (31/97)	66
F5 +/+	17% (19/115)	1% (1/115)	10% (12/115)	0% (0/115)	28% (32/115)	83
F5 -/-	23% (24/105)	3% (3/105)	7% (7/105)	3% (3/105)	35% (37/105)	68

Generation and genotype	Second replicate					
	Vulval extrusion	Bag of worms	Lost	Killed in transfer	Total excluded	Final n
WT F3	11% (10/95)	2% (2/95)	16% (15/95)	1% (1/95)	29% (28/95)	67
WT F4	20% (19/94)	1% (1/94)	7% (7/94)	2% (2/94)	31% (29/94)	65
WT F5	7% (7/104)	1% (1/104)	2% (2/104)	0% (0/104)	10% (10/104)	94
F3 +/+	2% (2/98)	1% (1/98)	0% (0/98)	2% (2/98)	5% (5/98)	93
F3 -/-	9% (9/100)	2% (2/100)	7% (7/100)	2% (2/100)	20% (20/100)	80
F4 +/+	5% (5/94)	1% (1/94)	10% (9/94)	2% (2/94)	18% (17/94)	77
F4 -/-	6% (6/97)	1% (1/97)	2% (2/97)	4% (4/97)	13% (13/97)	84
F5 +/+	4% (4/94)	1% (1/94)	6% (6/94)	1% (1/94)	13% (12/94)	82
F5 -/-	0% (0/98)	1% (1/98)	2% (2/98)	0% (0/98)	3% (3/98)	95

2.5.4 Data analysis

Deaths of non-excluded worms were treated as true deaths and included in the survival analysis in Prism, where each death is recorded as a '1' next to the day on which the death was recorded. When all deaths have been recorded, Prism computes percent survival at each time point e.g. day 12 (post hatching), plots a Kaplan-Meier survival plot and compares survival using the log-rank and Gehan-Breslow-Wilcoxon (GBW) tests. The log-rank statistics were chosen to be presented in the main results because this test gives equal weight to deaths at all time points, whereas the GBW test gives more weight to deaths at earlier time points, which was not considered appropriate for this analysis. In addition, the log-rank test was used by the studies these experiments were modelled on (Greer et al., 2010, Greer et al., 2011).

The WT F3 and WT F4 replicates were combined in the final survival analysis but the second WT F5 replicate was excluded due to its longer median lifespan (20 days) compared to the 18 day median lifespan of the other WT F3, F4 and F5 replicates. Moreover, the slightly larger initial and final n (104 and 94) of the second WT F5 replicate compared to the other WT controls led to suspicion that some F6 progeny were mistakenly

carried over during transfer and counted as F5s, thereby artificially extending the lifespan. Therefore, only the first replicate WT F5 data was carried forward to the final analysis.

Pairwise log-rank comparisons were performed comparing each WT control with its generation-matched WT *set-2* descendants e.g. WT F3 vs F3 +/+, WT F4 vs F4 +/+.

2.6 Developmental progression assay

Three 5cm NGM/OP50 plates with gravid YAs, one plate each for N2/WT, *set-2(bn129)* early (G3) and *set-2(bn129)* late (G59), were simultaneously bleached as described in 2.1.2. Synchronised L1s were pipetted onto fresh NGM/OP50 plates and left to develop for 60 hours at 20°C.

After 60 hours had passed, worms were washed off from each plate and resuspended in methanol to fix, then frozen at -20°C. Immediately prior to scoring, worms were thawed and washed in M9, then resuspended in 1ml DAPI (1µg/ml) for 10 mins at room temperature. Stained worms were washed in M9, pipetted onto slides and viewed under the microscope.

Approximately 100 worms were scored per condition. Appearance of the gonad was used to assign developmental stages (Figure 2.9). DAPI staining of the germline nuclei helped visualise this. At the L3 stage, the gonad is small and has not yet started to bend, but by mid-L4 there is a visible bend and by late L4 distal tip migration is complete and the gonad is completely extended. The presence of embryos indicates the YA stage has been reached.

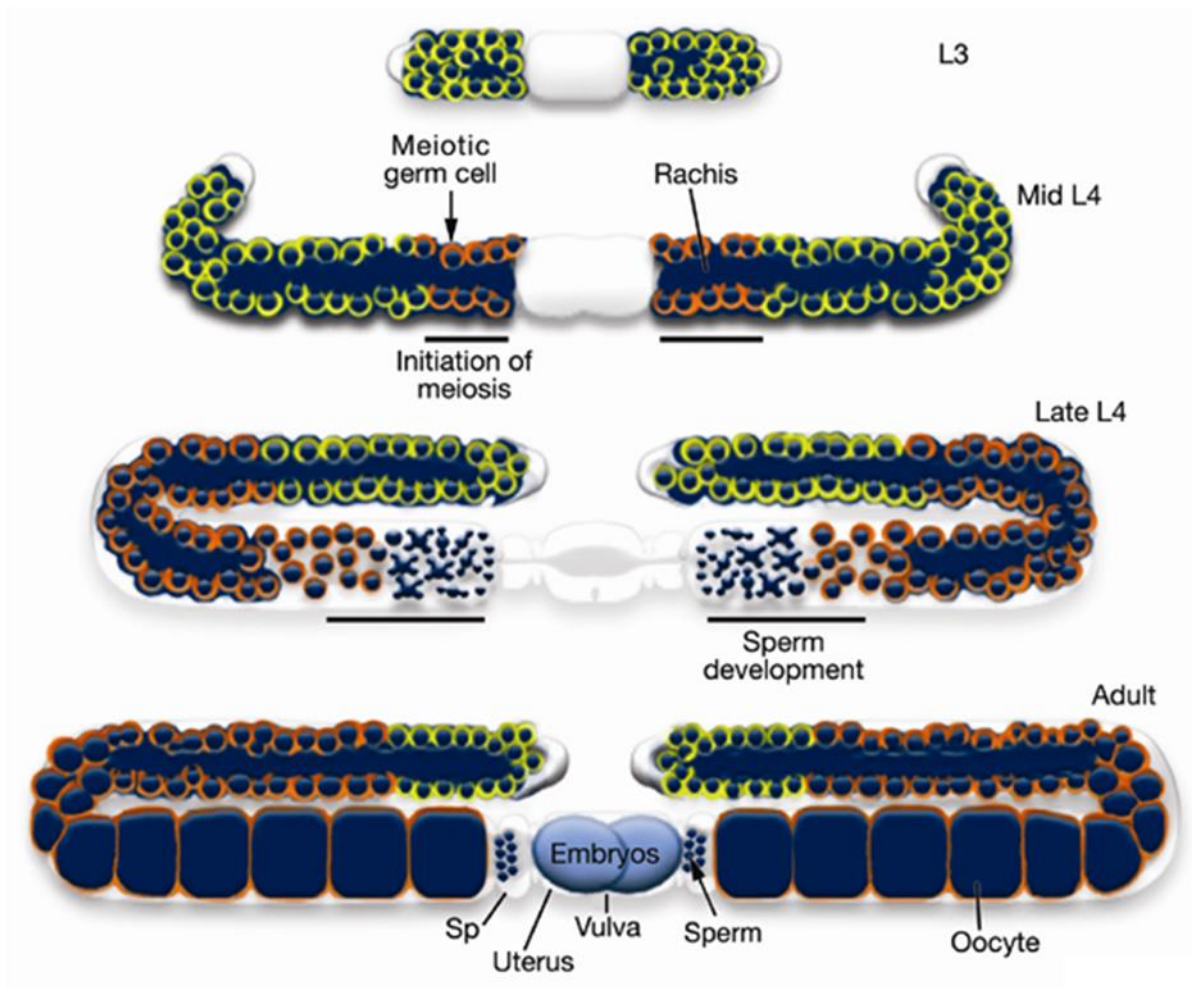


Figure 2.9 – Germline development from L3 to young adult in the *C. elegans* hermaphrodite. The distal tip of the gonad migrates throughout the L4 stage and thus the appearance of the gonad is a useful indicator for developmental stage. Adapted from wormatlas.org.

2.7 Immunofluorescence

2.7.1 Worm preparation/selection

Crosses between *set-2(bn129)* and WT worms were set up, and F3 homozygous WT and *set-2(bn129)* mutant descendants were obtained as described in 2.3.1.2.1 and illustrated in Figure 2.4. Worms were dissected for use in the immunofluorescence (IF) assays as YAs.

2.7.2 Slide preparation

Slides with three 14x14mm wells (Thermo Scientific 30-2066a-brown-ce24) were coated in 0.1% poly-L-lysine. Approximately 20 young adult worms were transferred to a 7 μ L drop of M9 in the centre of a well

and dissected to release the gonad, as illustrated in Figure 2.5. Wells were then covered with 22 x 40 mm coverslips and snap-frozen by placing on a pre-chilled metal plate on dry ice. After 5 mins, coverslips were flicked off and the slides were placed in Coplin jars filled with room temperature methanol. After fixing in methanol for 20 mins, slides were washed twice in PBS + 0.1% Tween (PBST)(Table 2.2).

2.7.3 Antibody staining

Primary antibodies in combination were mixed in PBST and diluted to the working concentrations specified in Table 2.9, and 50 µL of antibody mix was pipetted into each well. Slides were transferred to a humid chamber and the primary incubation was left overnight at 4°C. After two PBST washes, 50 µL of mixed anti-mouse, anti-rabbit and anti-sheep secondary antibodies (concentrations in Table 2.9) was pipetted into the wells and incubated for one hour at 37°C, followed by another two PBST washes. Slides were mounted in 7 µL ProlongGold anti-fade mounting medium, covered with 22x22mm coverslips and sealed after 24 hours in Biotium CoverGrip coverslip sealant. Prior to imaging slides were kept in the dark at 4°C.

Table 2.9 - Antibodies used to detect histone modifications

Antigen/Modification	Manufacturer	Manufacturer ID	Concentration used
H3	Abcam	ab128012	1:1000
H3K4me2	Diagenode	C15200151	1:5000
H3K4me3	Abcam	ab8580	1:20,000
H3K9me2	Wako	MABI0317	1:500
H3K9me3	Abcam	ab176916	1:100,000
anti-sheep Alexa Fluor 405	Abcam	ab175676	1:2500
anti-mouse Alexa Fluor 488	Abcam	ab150117	1:2500
anti-rabbit Alexa Fluor 594	Abcam	ab150088	1:2500

2.7.4 Antibody validation by dot blot

Peptides with one of each of the following modifications:

H3K4me2, H3K4me3, H3K9me2, H3K9me3, H3K27me2, H3K27me3, H3K36me2, H3K36me3, H3K79me2, H3K79me3, H4K20me2 and H4K20me3

were isolated from a peptide library, diluted in HEPES buffer (10mM HEPES, 0.5mM EDTA) and pipetted in 5µL volumes onto nitrocellulose membrane in serial dilutions of 100pM, 25pM and 10pM. Four identical sections of membrane were prepared in this way to test each of the four primary antibodies. H3K4me2, H3K4me3, H3K9me2 and H3K9me3 antibodies listed above (Table 2.9) were diluted in 20ml TBST (Table 2.2) + 5% skimmed milk powder to the working concentration used in the IF experiments (Table 2.9). Primary antibody incubations were left overnight, after which membranes were washed three times (5 mins each) in TBST and incubated with either anti-rabbit (for H3K4me3 and H3K9me3) or anti-mouse (for H3K4me2 or H3K9me2) HRP-conjugated secondary antibodies (50ml 1:10,000 dilution in TBST) for 1 hour at room temperature. Membranes were washed again in TBST and incubated with enhanced chemiluminescent (ECL) reagent for 1 min before exposing to X-ray film in a dark room to develop the signal.

2.7.5 Confocal Imaging

Imaging was performed on a Zeiss LSM880 + Airyscan Upright Microscope, using 16-bit depth with 512 x 512 pixel dimension and line averaging of 4. 12 gonads were imaged for each experimental condition, that is, generation (F3, F4, F5), genotype (WT +/+ or *set-2(bn129)* mutant -/-) and antibody combination (2.6.3).

2.7.6 Image analysis

Imaging analysis was performed in FIJI software. The following code (Figure 2.10), written in the Macro language, was used to quantify fluorescence from a subset of manually selected nuclei in the pachytene zone of the gonad. Briefly, the code minimises background and measures fluorescence from the three channels imaged: channel 1 at 488nm (green), channel 2 at 405nm (blue) and channel 3 at 594nm (red).


```

title = getTitle();
Stack.getPosition(c,z,t);
getSelectionBounds(roix, roiy, _ _);
run("Duplicate...", "duplicate slices="+z);
setBackground(0, 0, 0);
if(selectionType() != -1){
    run("Clear Outside", "stack");
}
run("Gaussian Blur...", "sigma=0.50 scaled slice");
setAutoThreshold("Huang dark");
run("Create Selection");

run("Set Measurements...", "area mean standard min median stack redirect=None decimal=2");
Stack.setChannel(1);
run("Measure");
Stack.setChannel(2);
run("Measure");
Stack.setChannel(3);
run("Measure");
close();
selectImage(title);
run("Restore Selection");
getSelectionBounds(offsetx, offsety, _ _);
setSelectionLocation(roix+offsetx, roiy+offsety);

```

Figure 2.10 – The Macro code used to quantify fluorescence in gonad imaging experiments

The data output from running the Macro code is illustrated in the screenshot below (Figure 2.11). It automatically comes in an Excel file and gives a mean and median value for the fluorescence measured in the area highlighted, for each of the three channels. These are arbitrary raw values without units, but the data is normalised by dividing *the fluorescence signal of histone modifications (channel 1 and 3)* by the fluorescence signal of the anti-H3 control (channel 2). The area measure was kept between 150 and 200 square pixels for the pachytene zone, to improve consistency between the size of the area highlighted for each image.

	A	B	C	D	E	F	G	H
1		Area	Mean	StdDev	Min	Max	Median	Ch
2	1	104.44	5324.11	765.96	4273	8447	5188	1
3	2	104.44	8900.92	4180.25	2140	31167	8061	2
4	3	104.44	3924.28	1711.64	901	11106	3573	3

Figure 2.11 – The data output of the fluorescence quantifying code

The median, rather than mean, values from the data output as illustrated above (Table 2.11) were taken forward to the main analysis, as these would be less influenced by outlying data points. After normalisation, these medians were then inputted to a Column table in GraphPad/Prism. Outliers were removed using the ROUT method (Q=10%). Data were plotted as scatter plots with the mean and standard deviation indicated. Statistical testing was performed using a 1-way ANOVA with Tukey’s multiple comparison.

Chapter 3. Developing an LC-MS/MS approach to characterise two *C. elegans* COMPASS mutants

3.1 Abstract

A liquid chromatography-tandem mass spectrometry (LC-MS/MS) approach was developed, to quantitatively determine and compare the global levels of a wide-range of histone post-translational modifications (PTMs) in *C. elegans*. This study assessed the relative abundances of all histone modifications in nuclei isolated from mixed stage embryos of wild-type and two COMPASS mutant strains, *set-2(bn129)* and *cfp-1(tm6369)*. In addition to the expected drop in abundance of H3K4me3-modified peptides in the mutants, there was also a notable increase in the relative abundance of H3-derived peptides with acetylated lysine residues. Lysine acetylation is a feature of actively expressed euchromatin, so this increase may underlie the inappropriate somatic reporter gene expression reported in germlines of COMPASS mutants (Robert et al., 2014).

3.2 Introduction

COMPASS is a conserved histone methyltransferase (HMT) complex, and the only H3K4 HMT to have been characterised in *C. elegans*. While there is another MLL-like SET-domain containing protein in *C. elegans*, SET-16, study of this has been challenging because the *set-16* mutation is lethal and therefore mutant strains have to be maintained with a balancer (Fisher et al., 2010). In mammals and flies there are six COMPASS complexes and three H3K4 trimethylating complexes (Ardehali et al., 2011, Mohan et al., 2011), which complicate study with the potential for redundancy. Yeast, despite having one H3K4 HMT, is limited as a model because H3K4 methylation is exclusively associated with transcribing genes (Ng et al., 2003). In *C. elegans*, like in other metazoans, Set1/MLL activities can operate independently of transcription (Li and Kelly, 2011). This has allowed a diversification of H3K4me3 function, which is more likely to be conserved between *C. elegans* and other metazoans, especially given the conservation of the COMPASS complex composition and structure between different organisms.

In the *C. elegans* COMPASS complex, SET-2 is the catalytic subunit whereas CFP-1 plays a role in targeting the complex to actively expressed promoters (Chen et al., 2014). Of the two genes, *set-2* is the better characterised. Mutants with the null *set-2(bn129)* allele had smaller brood sizes compared to wild type and showed temperature-sensitive sterility, slow growth (Xiao et al., 2011) and somatic gene derepression in the germline (Robert et al., 2014). Notably the *set-2(ok952)* mutants, with a likely hypomorphic allele, did not

show any germline defects (Simonet et al., 2007, Robert et al., 2014). Some similar phenotypes were noted, including slow growth and temperature sensitive sterility, following RNAi against *cfp-1* (Kamath et al., 2003).

Other studies, however, have reported differences upon *set-2* and *cfp-1* function disruption. For example, in the suppressor screen in which the *cfp-1* gene was first reported in *C. elegans*, RNAi inactivation of *cfp-1* suppressed the multivulval (Muv) phenotypes of *hpl-1;hpl-2* double mutants, whereas *set-2* RNAi did not (Simonet et al., 2007). Moreover, CFP-1 has been reported to play a role in other complexes besides COMPASS (Beurton et al., 2019).

It would be informative, therefore, to compare the histone modification profiles of *cfp-1* and *set-2* loss-of-function (LOF) mutants and see whether H3K4me3 and other modifications are impacted similarly or differently. The main advantage of the LC-MS/MS approach is that it offers a means of appreciating the global impact of the *set-2* and *cfp-1* mutations on the histone modification landscape. Antibody-based studies of histone modifications, such as western blotting or immunostaining, require the experimenter to have some idea of what they expect to find, and there can be issues with binding specificity and epitope occlusion that may cast doubt on the meaning of the results, especially where quantification is needed.

This particular bottom-up proteomic method is novel for *C. elegans* research. Other groups have used the middle-down method with success, discovering interesting combinatorial histone modification patterns (Sidoli et al., 2014) and even the novel H3K23me2 heterochromatic mark (Vandamme et al., 2015). However, the challenge of detecting H3K4 has limited study of this modification by mass spectrometry up until now.

3.3 Experimental design

Mixed stage embryos were selected as the tissue from which to isolate a nuclear histone extract. They are easy to isolate from the adult worm by bleaching, represent many different tissue types and have been used for numerous ChIP studies in *C. elegans* (Chen et al., 2014, Beurton et al., 2019).

Prior to embryo collection, mixed populations of *cfp-1*, *set-2* and WT strains were successively bleached and the hatched L1s transferred to larger plates in order to support an increasingly large, synchronised population. When a population of approximately 2000 synchronized young adults per plate was reached, bleaching was performed to retrieve the final collection of embryos, from which the live ones were isolated by sucrose floating. These embryos were treated with chitinase to destroy the protective shell around the embryos and make nuclei retrievable by douncing in hypotonic buffer. Nuclei were resuspended in acid to solubilise the histones within, which were then precipitated out of solution and resuspended in water.

Histone extracts were checked for purity and concentration by SDS-PAGE, to assess suitability for mass spectrometry.

Before running the histone extracts in the LC-MS/MS instrument, they were digested into smaller peptides and stabilised. This was achieved by a combination of propionylation and trypsin digestion, after which samples were desalted using C18 Hypercarb tips. The object of this step, known informally as 'tip clean-up', is to purify peptides from the original sample solution, that may still contain traces of reagents from the histone extraction procedure. It is essential that as pure a sample as possible is run in the instrument. A potential problem posed by this step is the selective loss of H3K4me2 and H3K4me3-modified peptides, these being very hydrophilic. Therefore, in this experiment, the loss was partly compensated for by careful choice of ion-pairing reagent. This is usually a perfluorinated acid whose dissociated anions neutralise positive charges on basic amino acid sidechains and N-terminal amines, increasing hydrophobicity of the peptides and thereby prolonging their retention on the stationary phase of the HPLC column.

3.4 Results

3.4.1 Use of heptafluorobutyric acid instead of conventional trifluoroacetic acid as ion-pairing reagent enables detection of H3K4me2- and H3K4me3-modified peptides

Trifluoroacetic acid (TFA) is the most common perfluorinated acid used for ion-pairing in LC-MS/MS experiments, but there are others which confer greater hydrophobicity onto peptides and may therefore be a more suitable choice when very hydrophilic peptides need to be detected. TFA was the first ion-pairing reagent tried, but as suspected it did not confer sufficient hydrophobicity to enable the detection of very hydrophilic H3K4me2- and H3K4me3-modified peptides (Figure 3.1). Use of heptafluorobutyric acid (HFBA) allowed the detection of these di- and tri-methylated peptides, but not acetylated H3K4. This was acceptable because the methylated peptides were of greater interest in these experiments, and so thereafter HFBA was the ion-pairing reagent used for all LC-MS/MS experiments reported in this thesis. Nevertheless, in future it would be informative to try the phenyl-isocyanate approach (Maile et al., 2015) to see if that can improve the detection of H3K4ac and other very hydrophilic peptides.

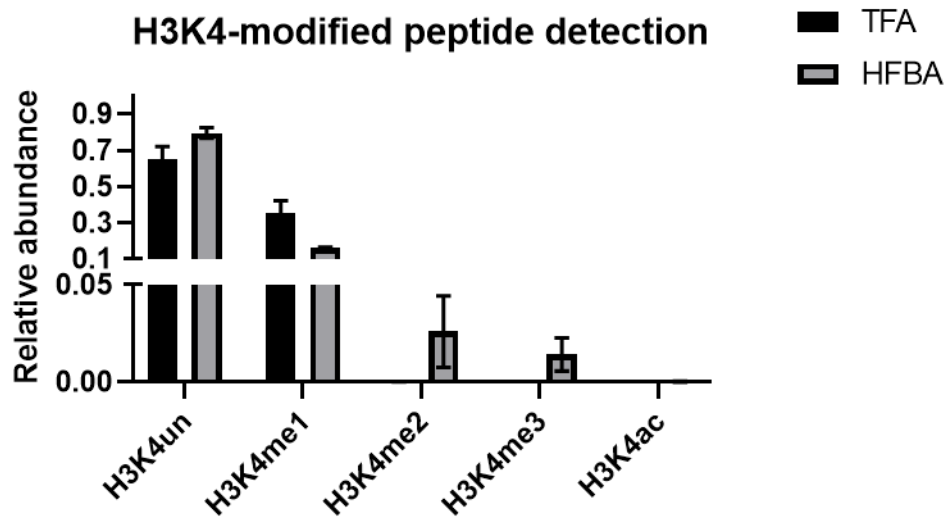


Figure 3.1 – HFBA but not TFA allows detection of di- and tri-methylated H3K4-modified peptides.

The height of the bars represents the mean of three experiments, the error bars represent the standard deviation. The y-axis has been split to improve the resolution of the bars representing the very low abundance H3K4me2- and H3K4me3-modified peptides, compared with the more abundant unmodified and H3K4me1-modified peptides. Relative abundance is the percentage intensity of each modified peptide (e.g. H3K4me3) detected, out of the total intensity detected for all TKQTAR peptides. This is uncorrected data, prior to application of correction factors (2.3.5.2).

3.4.2 H3K4me3 levels are dramatically reduced in *cfp-1* and *set-2* mutants.

As would be expected for these predicted loss-of-function mutants, relative abundance of H3K4me3 was reduced 16-fold in *cfp-1* (mean 0.58), and 30-fold in *set-2* mutants (mean 1.1) compared with WT (mean 17.27) (Table 3.1, Figure 3.2). These means were highly significantly different between each of the mutants and WT ($p < 0.0001$), but the difference in relative abundance of H3K4me3 between *cfp-1* and *set-2* was not significant. These results are consistent with *set-2(bn129)* and *cfp-1(tm6369)* being loss-of-function alleles. They suggest that the COMPASS complex is indeed responsible for the vast majority of H3K4me3 in *C. elegans* embryos.

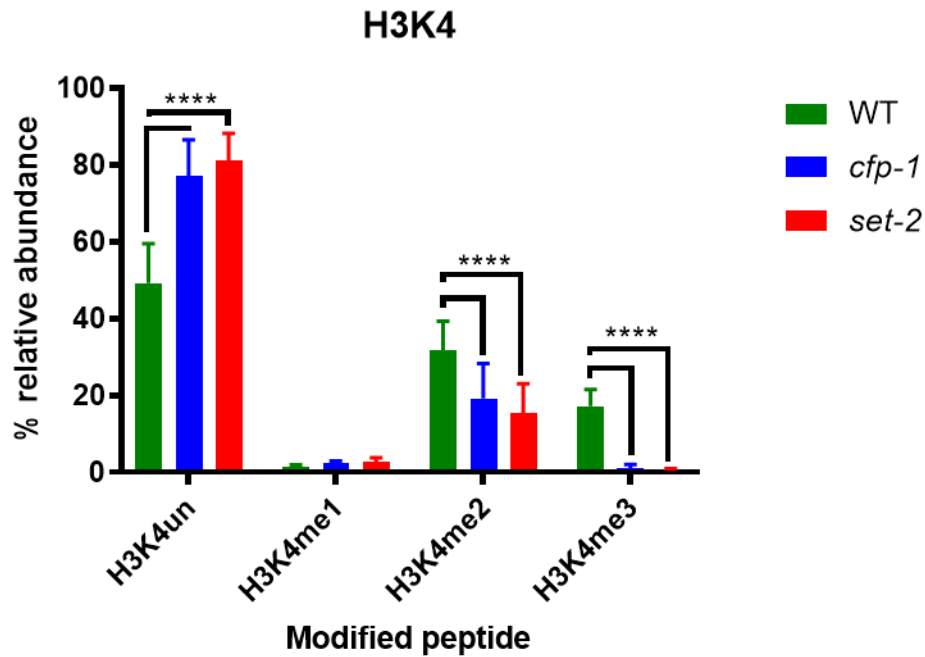


Figure 3.2 – Relative abundances of H3K4-modified peptides. The height of the bars represents the mean of four experiments, the error bars represent the standard deviation. The graph shows the relative abundances of all proteoforms detected for the TKQTAR peptide, containing K4.

Modification	Mean relative abundance			Significance		
	WT	<i>cfp-1</i>	<i>set-2</i>	WT vs <i>cfp-1</i>	WT vs <i>set-2</i>	<i>cfp-1</i> vs <i>set-2</i>
H3K4un	49.32	77.15	81.25	****p<0.0001	****p<0.0001	*p = 0.0277
H3K4me1	1.494	2.431	2.687	ns p = 0.8036	ns p = 0.7026	ns p = 0.9835
H3K4me2	31.91	19.3	15.48	****p<0.0001	****p<0.0001	*p = 0.0418
H3K4me3	17.27	1.112	0.5779	****p<0.0001	****p<0.0001	ns p = 0.9308

Table 3.1 - Mean relative abundance and significance between means for H3K4-modified peptides. Means are the average of four replicates, they were statistically analysed by a 2-way ANOVA with Tukey's multiple comparisons test.

3.4.3 H3K4me2 levels are reduced by approximately half in *cfp-1* and *set-2* mutants.

Relative abundance of H3K4me2 was reduced 1.7-fold in *cfp-1* (mean 19.3), and 2-fold in *set-2* mutants (mean 15.48) compared with WT (mean 31.91) (Table 3.1, Figure 3.2). This is a much less dramatic reduction than for H3K4me3 but is still very significant. These results suggest that COMPASS is a major but not the only HMT responsible for H3K4me2 deposition. Perhaps SET-16 is also involved. Moreover, there is a smaller but significant (p<0.05) reduction in H3K4me2 levels in *set-2* compared to *cfp-1*. Correspondingly, *set-2* mutants

have significantly ($p < 0.05$) increased levels of the unmodified peptide, H3K4un. This suggests that lysine demethylases (KDMs) may be more active in the absence of COMPASS. H3K4me1 levels do not differ significantly between any of the strains. This mark has been previously shown to be enriched at enhancers and distal gene regulatory elements (Hon et al., 2009), and so it is surprising that in this experiment it appears so much less abundant (in the WT condition) than H3K4me3, which has a much more restricted distribution at transcription start sites (Barski et al., 2007). That H3K4me1 abundance does not change in the absence of COMPASS suggests that it is regulated by other enzymes, perhaps SET-16.

3.4.4 H3K14ac and H3K27ac levels are significantly elevated in *cfp-1* and *set-2* mutants

In addition to changes on the H3K4 lysine, the target of COMPASS, *cfp-1* and *set-2* mutants had significantly elevated acetylation modifications of H3K14 and H3K27, compared to WT. The mean relative abundance of the H3K9unK14ac-modified peptide was significantly higher in *cfp-1* (29.3) and *set-2* (29.79) compared to WT (23.48) (Table 3.2, Figure 3.3A). Correspondingly, the relative abundance of the unmodified H3K9unK14un peptide was reduced in the mutants compared to WT, significantly less for *cfp-1* ($p = 0.003$) but not for *set-2* ($p = 0.29$).

While there are no significant differences in relative abundance of other proteoforms of the H3K9/K14-containing peptide, sequence KSTGGKAPR, there are slightly more reduced levels of H3K9me2K14un and H3K9me3K14un in the *set-2* mutant compared with WT and *cfp-1* (Figure 3.3A). The differences in the levels of peptides with unmodified K14 are all small and not significant statistically, and therefore may not be real or biologically meaningful, but may be indicative of differences between *cfp-1* and *set-2* mutants. Such a difference is not unexpected given the proteins' different roles within the COMPASS complex, and potential differences in interaction partners outside COMPASS too.

Table 3.2 - Mean relative abundance for H3K14 and K27 modifications that significantly differed between WT, *cfp-1* and *set-2*. Means are the average of four replicates, they were statistically analysed by a 2-way ANOVA with Tukey's multiple comparisons test.

Modification	Mean relative abundance			Significance		
	WT	<i>cfp-1</i>	<i>set-2</i>	WT vs <i>cfp-1</i>	WT vs <i>set-2</i>	<i>cfp-1</i> vs <i>set-2</i>
H3K9unK14un	44.6	39.36	42.3	** $p = 0.003$	ns $p = 0.2948$	ns $p = 0.1389$
H3K9unK14ac	23.48	29.3	29.79	*** $p = 0.0009$	*** $p = 0.0003$	ns $p = 0.9445$
H327me2K36un	58.36	52.76	49.41	**** $p < 0.0001$	**** $p < 0.0001$	* $p = 0.0143$
H3K27acK36un	12.06	17.85	17.11	**** $p < 0.0001$	*** $p = 0.0001$	ns $p = 0.8013$

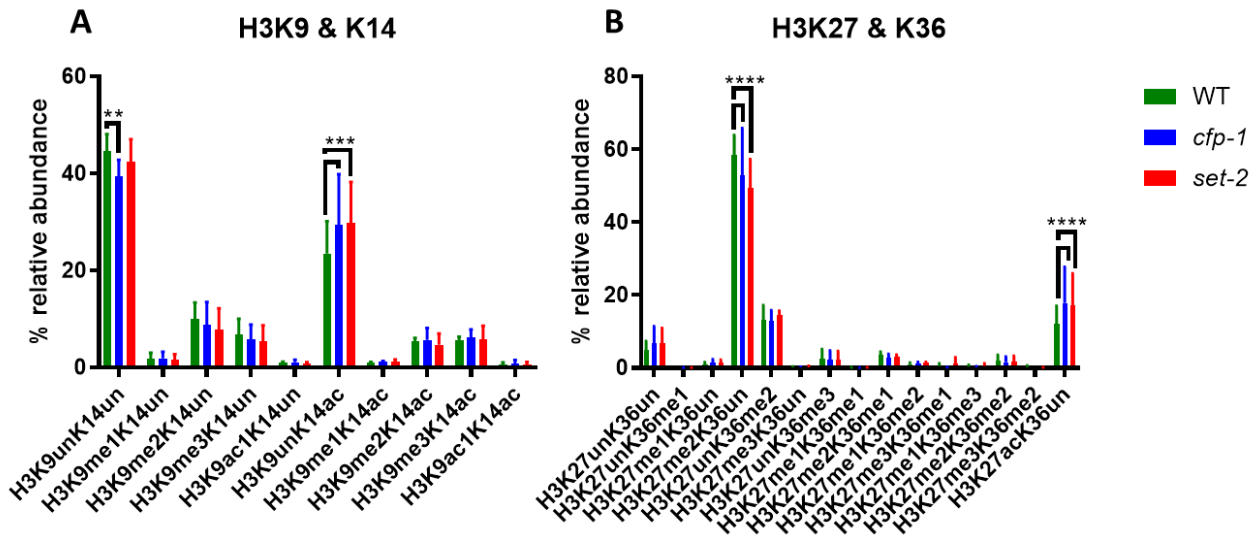


Figure 3.3 – Relative abundances of H3K9-, K14-, K27- and K36-modified peptides. The height of the bars represents the mean of four experiments, the error bars represent the standard deviation. Graph A displays the relative abundances of all proteoforms detected for the KSTGGKAPR peptide (containing K9 and K14), while graph B displays all those detected for the KSAPASGGVKKPHR peptide (containing K27 and K36).

It is notable that for the H3K27/K36-containing peptide, sequence KSAPASGGVKKPHR, the unmodified form of the peptide is not the most abundant, as it is with the other peptides (Figure 3.2, 3.3). H3K27me2K36un is the most abundant proteoform, but its relative abundance is significantly reduced in *cfp-1* (52.76) and *set-2* (49.41) compared to WT (58.36) (Table 3.2, Figure 3.3B). The relative abundance of H3K27acK36un is significantly increased in *cfp-1* (17.85) and *set-2* (17.11) compared to WT (12.06). None of the other proteoforms differ significantly in relative abundance between strains (Figure 3.3B). This indicates a conversion of H3K27me2 to H3K27ac specifically in H3 with unmodified K36. H3K27me2 is an important repressive mark, whereas acetylation tends to be associated with actively expressed chromatin. Such a significant shift in balance of repressive/active marks on one lysine residue suggests an important consequence for the loss of COMPASS function on the regulation of gene expression by H3K27 modification.

3.4.5 H3K23ac levels are significantly elevated in *cfp-1* but not *set-2* mutants

While the relative abundance of H3K14ac and H3K27ac was elevated to a similar extent in *cfp-1* and *set-2* mutants, H3K23ac, another acetylated lysine, is elevated significantly above WT (27.11) in *cfp-1* (31.46) but not *set-2* (29.01) mutants (Table 3.3, Figure 3.4). Correspondingly, the relative abundance of H3K18unK23me1 is reduced in *cfp-1* mutants (15.32) compared to *set-2* (20.52) and WT (18.99). For K23, then, unlike other modified lysines discussed so far, the impact of CFP-1 loss-of-function is not the same as

for that of SET-2. There are no significant differences in the relative abundances of K18-modified proteoforms.

Table 3.3 - Mean relative abundance for H3K23 modifications that significantly differed between WT, *cfp-1* and *set-2*.

Modification	Mean relative abundance			Significance		
	WT	<i>cfp-1</i>	<i>set-2</i>	WT vs <i>cfp-1</i>	WT vs <i>set-2</i>	<i>cfp-1</i> vs <i>set-2</i>
H3K18unK23ac	27.11	31.46	29.01	*p = 0.0253	ns p = 0.4611	ns p = 0.2857
H3K18unK23me1	18.99	15.32	20.52	ns p = 0.0666	ns p = 0.6049	**p = 0.0064

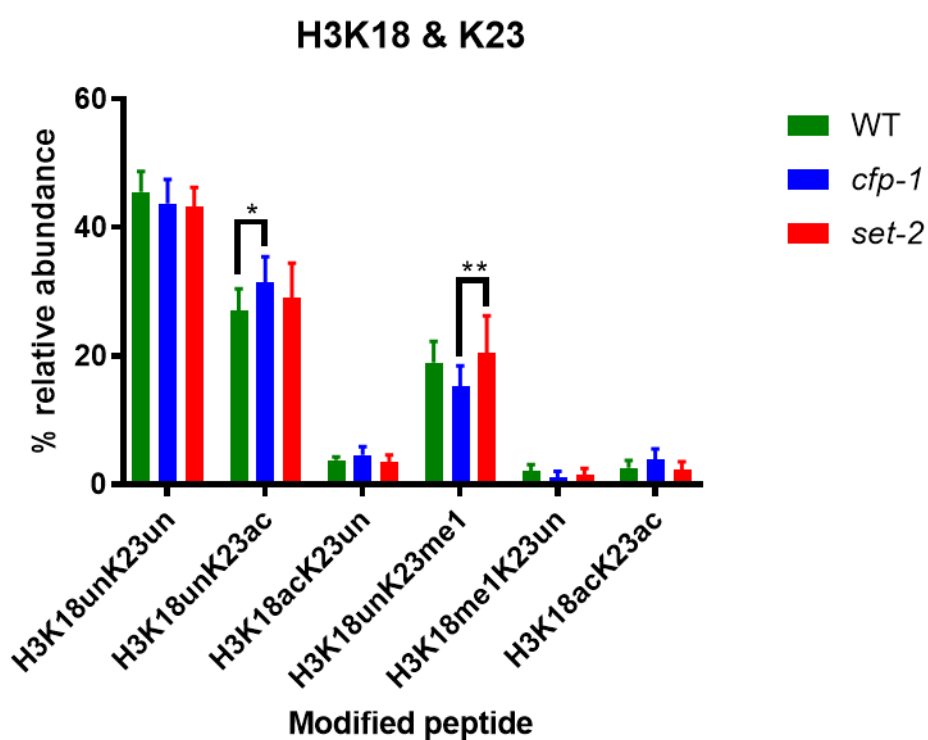


Figure 3.4 – Relative abundances of H3K18- and K23-modified peptides. The height of the bars represents the mean of four experiments, the error bars represent the standard deviation. The graph displays the relative abundances of all proteoforms detected for the KQLATKAAR peptide (containing K18 and K23).

3.5 Discussion

The similarity of the consequences for H3 modification in *cfp-1* and *set-2* mutants are striking, and suggest that SET-2 and CFP-1 proteins have near identical roles, at least in the context of their role in regulating the histone modification landscape.

However, the recent discovery that CFP-1 interacts with the small histone deacetylase (HDAC) complex SIN3S, independently of SET-2/COMPASS (Beurton et al., 2019), challenges the idea that SET-2 and CFP-1 play the same role in chromatin regulation. Moreover, this interaction provides a possible explanation for why *cfp-1* mutants have elevated acetylation at K14, K23 and K27 residues if this is an effect of reduced functionality of SIN3S. Why *set-2* mutants have elevated acetylation at two of the same lysine residues, K14 and K27, is harder to explain. Perhaps COMPASS function is in some way a prerequisite for CFP-1's role in SIN3S. CFP-1 plays a role in the recruitment of SIN3S to H3K4me3-enriched promoters. So, when H3K4me3 is absent, as in both the *set-2* and *cfp-1* mutants, that recruitment cannot happen, and thus the consequence for K14 and K27 acetylation is the same in both mutants. In short, WT H3K4me3 marking is essential for the normal functional interaction for CFP-1 with SIN3S. This explanation is supported by CHIP-seq data showing that the genomic distribution of H3K4me3 is similarly reduced in *cfp-1* and *set-2* mutant embryos (Beurton et al., 2019).

The next chapter discusses the interaction between CFP-1 and the SIN3S complex subunits further, and how experiments that were intended primarily to probe this interaction led to the discovery of a transgenerational worsening of the *cfp-1* mutant phenotype, at the normal growth temperature of 20°C.

Chapter 4. Somatic gene derepression in the *cfp-1* germline is enhanced by RNAi knockdown of *hda-3*, and possibly *hda-1* and *sin-3*

4.1 Abstract

Using the *unc-119::GFP* reporter as a readout for somatic gene derepression, RNAi knockdown of the histone deacetylase (HDAC) subunits *hda-1*, and possibly *hda-3* and *sin-3*, was found to enhance *unc-119::GFP* expression in the germline of *cfp-1(tm6369)* mutants. RNAi of *hda-1* was not found to enhance *unc-119::GFP* expression in *set-2(bn129)* mutants.

4.2 Introduction

Both *set-2* and *cfp-1* mutants were found to have elevated lysine acetylation by LC-MS/MS (chapter 3), raising the possibility that COMPASS interacts with a histone acetyltransferase (HAT) or histone deacetylase (HDAC). Previous RNAi enhancer screens for further reduction of the already low brood size observed in *set-2* and *cfp-1* mutants, found no effect of RNAi knockdown (KD) of major HAT genes *mys-4* and *hat-1* (Pokhrel et al., 2019). However, RNAi KD of *hda-1* and *hda-3*, orthologs of mammalian class I HDACs (Shi and Mello, 1998), and of *sin-3*, a component of the Sin3 HDAC complex, significantly reduced brood size in *cfp-1* mutants, but not in *set-2* mutants (Pokhrel et al., 2019). Low brood size can be symptomatic of germline gene misregulation, as shown previously using the *unc-119::GFP* reporter (Robert et al., 2014). The same reporter was used in these experiments to confirm the genetic interactions indicated by the brood size assay results, and to assess the impact of HDAC RNAi KD on *set-2* and *cfp-1* germline gene regulation.

4.2.1 HDAC complexes in *C. elegans*

HDAC proteins are highly conserved in eukaryotes. All three classes, I, II and IV are found in all sequenced free-living eukaryotes (except fungi) and even eubacteria, thus predating the evolution of histones (Gregoretto et al., 2004). Evidence that HDACs act on non-histone substrates, including tubulin (Hubbert et al., 2002) and tumour suppressor p53 (Ito et al., 2002), raises the intriguing prospect that their primary activity may not be directed against histone substrates, but that this is a secondary function they gained later in their evolution.

HDAC genes have undergone extensive gene duplication and subsequent functional differentiation. Therefore, further subclassifications are necessary to define the functional subgroups. HDAC subclasses are named according to the human proteins contained within them, for example, the HDAC1/2 subclass contains the proteins HsHDAC1 and HsHDAC2. Phylogenetic analysis indicates that the common metazoan ancestor would have contained one HDAC1/2 protein, but that this was duplicated, independently, in both the *C. elegans* and human evolutionary lineages (Gregoretto et al., 2004). Strong sequence constraint between HsHDAC1 and HsHDAC2 suggests little functional divergence, as does their existence in the same complexes, Sin3, NuRD and CoREST.

The Sin3 and CoREST HDAC complexes derive their names from their major subunit. NuRD, an abbreviation of nucleosome remodelling and histone deacetylase complex, contains an Mi2 ATPase as the major subunit. The roles of the NuRD complex in *C. elegans* include embryonic patterning (Solari et al., 1999) and antagonising the Ras pathway in vulval development (Solari and Ahringer, 2000). The Sin3 complex differs from NuRD in that it does not appear to be active in embryos. A deletion mutation in the *sin-3* gene did not produce any embryonic abnormalities but caused defects at larval stages, in male sensory ray patterning and hermaphrodite vulval morphogenesis (Choy et al., 2007). *C. elegans* SIN-3 is the only Sin3-like protein in nematodes, and has a paired amphipathic helix (PAH) and histone deacetylase interaction domain (HID) like its vertebrate counterparts (Laherty et al., 1997). The name CoREST comes from the interaction between this human protein and the REST tumour suppressor (Andres et al., 1999). The *C. elegans* orthologue of human CoREST is *spr-1*, which, like the NuRD and Sin3 complex components, also plays a role in regulating vulval morphogenesis (Bender et al., 2007).

Three genes in *C. elegans*, named *hda-1*, *hda-2* and *hda-3*, were found to share significant homology with mammalian HDAC1 (Shi and Mello, 1998). The encoded proteins each contain the conserved histidine residues, H150, H151 and H188, which have been shown to be important for HDAC activity (Kadosh and Struhl, 1997). Of the three, HDA-1 shares the highest homology with HDAC1, and was found to interact functionally with the HAT CBP-1 in the regulation of endoderm differentiation (Shi and Mello, 1998). *hda-1* was also shown to inhibit vulval development via the SynMuv A and B pathways, which *hda-2* and *hda-3* did not (Solari and Ahringer, 2000). It seems, therefore, that while all three HDAs are capable of HDAC activity, they may have diverged functionally and form different complexes.

4.2.2 The *unc-119::GFP* reporter

The *unc-119::GFP* reporter provides a readout for neuronal gene expression. It was utilised in a study investigating the impact of conserved translational regulators, MEX-3 and GLD-1, on the maintenance of

totipotency in the *C. elegans* germline (Ciosk et al., 2006), confirming that the abnormal cells observed in the central gonad of *mex-3;gld-1* double mutants were indeed differentiated neuronal cells. The authors noted that those cells expressing this reporter had ‘extensive processes’ like those seen in normal *C. elegans* neuronal cells.

A later study (Robert et al., 2014) used the same reporter to assess somatic conversion of germ cells in *set-2* mutants, at the non-permissive temperature of 25°C. They found that sterility and *unc-119::GFP* expression concomitantly increased with each generation after the temperature shift in *set-2(bn129)* mutants, with complete sterility reached by the F4 generation. Notably, the *set-2(ok952)* mutant did not show the same trend, nor the mortal germline (Mrt) phenotype associated with *set-2(bn129)* and other COMPASS subunit loss-of-function mutants. The authors observed ‘axodendritic projections’ in *unc-119::GFP*-expressing cells in *set-2(bn129)* germlines, similar to those noted in other studies (Ciosk et al., 2006, Updike et al., 2014). However, Updike *et al.* found that *unc-119::GFP* positive germ cells lack the nuclear granules that are observed in differentiated neurons, indicating that while the germ cells are reprogrammed to a different fate, they cannot terminally differentiate.

Therefore, due to its prior successful use in characterizing the *set-2(bn129)* mutant, the *unc-119::GFP* reporter was used here to address the question of whether HDAC inactivation might worsen the gene derepression in the *set-2* and *cfp-1* germline. As mentioned previously, *hda-1*, *hda-3* and *sin-3* RNAi reduced brood size in *cfp-1* mutants (Pokhrel et al., 2019), but these *unc-119::GFP* reporter experiments would address whether the reduction in fertility might be due to the impact of *hda-1* and *hda-3* inactivation on gene derepression in the germline.

4.3 Experimental Design

For the *hda-3* and *sin-3* RNAi, *cfp-1* mutant L4 hermaphrodites previously maintained at 20°C were picked to RNAi feeding plates and shifted to the non-permissive temperature of 25°C. At this temperature, *unc-119::GFP* expression was expected to be induced in the germline of the F1 at a frequency of 7.5% as observed in *set-2(bn129)* mutants (Robert et al., 2014). Their progeny, the F1, were divided into those laid in the first 24 hours after RNAi feeding began, and those laid in the next 24 hours. Reporter expression in the ‘0-24h’ and ‘24-48h’ cohort was assessed on the first day of young adulthood. Cohorts were compared to confirm the replication of the trend, and to assess if the exposure of the later cohort to the non-permissive temperature and RNAi feeding from earlier stages of gametogenesis would produce a more pronounced effect.

RNAi knockdown of *hda-1* results in embryonic lethality, and therefore this experiment had to be performed differently. Instead, synchronised *cfp-1*, *set-2* and WT L1s were directly picked onto the RNAi plates, grown to adulthood at 25°C, and assessed for reporter expression (Fig. 4.1) on the first day of young adulthood.

An empty vector (EV) RNAi negative control was performed with each experiment, for both the WT and mutant. The results from two independent replicates are presented here in tables. It was inappropriate to pool the data from each replicate because the manner in which the statistical testing was conducted (2.4.5) required that each experiment be considered individually.

4.4 Results

The aim of this experiment was to determine whether RNAi of the *HDAC1* orthologues *hda-1* and *hda-3*, and the Sin3 HDAC corepressor complex scaffold encoding *sin-3*, could enhance gene derepression in the germline of *cfp-1* mutants. This would shed further light on the mechanism behind a previous finding that *hda-1*, *hda-3* and *sin-3* RNAi further reduced the brood size of *cfp-1*, but not *set-2* mutants (Pokhrel et al., 2019).

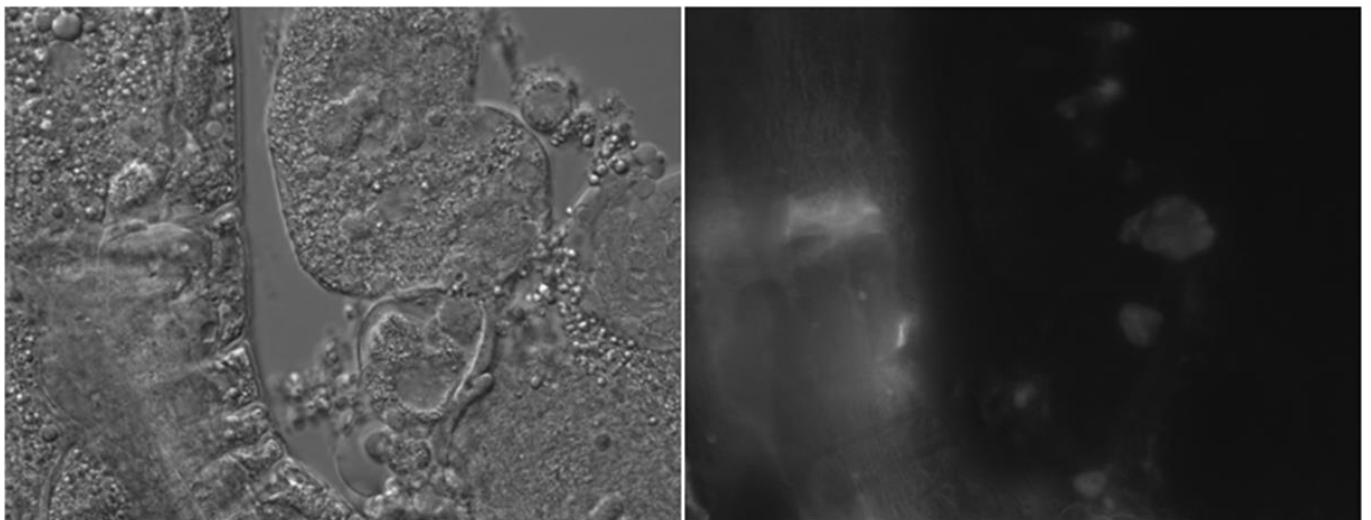


Figure 4.1 - *unc-119::GFP*-expression in a dissected gonad. The panel on the left shows the DIC image and the right panel shows GFP expression. The DIC image shows a section the worm's body on the left and a section of the dissected gonad next to it. Clusters of GFP-expressing cells can be seen in the gonad. This is a *cfp-1* negative RNAi control.

4.4.1 *hda-3* RNAi significantly enhanced *cfp-1* germline gene derepression

In all experiments, *hda-3* RNAi significantly enhanced *unc-119::GFP* expression in *cfp-1* mutant germlines (Table 4.1). There was no *unc-119::GFP* expression in WT worms subjected to *hda-3* RNAi, which gives confidence that the inactivation of HDA-3 cannot induce germline gene derepression independently of CFP-1. In the second replicates, which tended to have a higher n due to improved speed and efficiency with scoring, the enhancing effect is more significant. Significance is reduced in the 24-48 hr dataset compared with the 0-24 hr dataset, probably because the *cfp-1* worms on the EV control plates had *unc-119::GFP* expression in a greater proportion of their germlines (0.51 vs 0.33 average, Table 4.1). When the 0-24 hr and 24-48 hr data are pooled the enhancing effect is most significant.

Table 4.1 – *unc119::GFP* expression scoring data for *hda-3* RNAi in *cfp-1* and WT germlines. The table displays the raw number of germlines scored for each condition and the calculated proportion of *unc119::GFP*-expressing (+ve) germlines. The reported *p* value for each experiment refers to the likelihood of observing this result assuming the null hypothesis of no interaction between the *cfp-1* and *hda-3* genes (see 2.4.5 for details of how this was calculated).

Day 1/0-24hr data		WT		<i>cfp-1</i>		Probability of no genetic interaction
		EV	<i>hda-3</i>	EV	<i>hda-3</i>	
rep 1	GFP+ve	0	0	17	18	**p = 0.0013
	GFP-ve	39	37	23	6	
	Total	39	37	40	24	
	proportion +ve	0	0	0.43	0.75	
rep 2	GFP+ve	0	0	11	20	***p = 0.00021
	GFP-ve	54	50	36	22	
	Total	54	50	47	42	
	proportion +ve	0	0	0.23	0.48	
average		0	0	0.33	0.61	
Day 2/24-48hr data		WT		<i>cfp-1</i>		Probability of no genetic interaction
		EV	<i>hda-3</i>	EV	<i>hda-3</i>	
rep 1	GFP+ve	0	0	17	23	*p = 0.014
	GFP-ve	32	29	13	6	
	Total	32	29	30	29	
	proportion +ve	0	0	0.57	0.79	
rep 2	GFP+ve	0	0	24	32	**p = 0.0085
	GFP-ve	59	60	30	19	
	Total	59	60	54	51	
	proportion +ve	0	0	0.44	0.63	
average		0	0	0.51	0.71	
Day 1&2/0-48hr data		WT		<i>cfp-1</i>		Probability of no genetic interaction
		EV	<i>hda-3</i>	EV	<i>hda-3</i>	
rep 1	GFP+ve	0	0	34	41	***p = 0.000028
	GFP-ve	71	66	36	12	
	Total	71	66	70	53	
	proportion +ve	0	0	0.49	0.77	
rep 2	GFP+ve	0	0	35	52	***p = 0.000016
	GFP-ve	113	110	66	41	
	Total	113	110	101	93	
	proportion +ve	0	0	0.35	0.56	
average		0	0	0.42	0.67	

4.4.2 *sin-3* RNAi may enhance *cfp-1* germline gene derepression.

Compared to *hda-3* RNAi, the trend of *sin-3* RNAi enhancing *unc-119::GFP* expression is less significant and consistent across experiments (Table 4.2). For the first replicate, there was a significant enhancement of *unc-119::GFP* expression both in the 0-24h ($p=0.0025$) and 24-48h ($p=0.049$) datasets. For the second replicate, only the 24-48h data were significant ($p=0.048$). This may be partly due to the small (6.25%) proportion of WT germlines expressing *unc-119::GFP* following *sin-3* RNAi in the second replicate 0-24h data, whereas none did in the first replicate. In the 24-48h data, the proportion of WT germlines expressing *unc-119::GFP* is similar between replicates (6.45 vs 6.78%). As with the *hda-3* RNAi data, the pooled 0-48h data gives the most significant enhancement trend.

Table 4.2 – *unc119::GFP* expression scoring data for *sin-3* RNAi in *cfp-1* and WT germlines. The table displays the raw number of germlines scored for each condition and the calculated proportion of *unc119::GFP*-expressing (+ve) germlines. The reported *p* value for each experiment refers to the likelihood of observing this result assuming the null hypothesis of no interaction between the *cfp-1* and *sin-3* genes (see 2.4.5 for details of how this was calculated).

Day 1/0-24hr data		WT		<i>cfp-1</i>		Probability of no genetic interaction
		EV	<i>sin-3</i>	EV	<i>sin-3</i>	
rep 1	GFP+ve	0	0	12	19	**p = 0.0025
	GFP-ve	38	38	22	12	
	Total	38	38	34	31	
	proportion +ve	0	0	0.35	0.61	
rep 2	GFP+ve	0	3	15	18	ns p = 0.25
	GFP-ve	60	45	35	24	
	Total	60	48	50	42	
	proportion +ve	0	0.06	0.30	0.43	
average		0	0.03	0.33	0.52	
Day 2/24-48hr data		WT		<i>cfp-1</i>		Probability of no genetic interaction
		EV	<i>sin-3</i>	EV	<i>sin-3</i>	
rep 1	GFP+ve	0	2	24	27	*p = 0.049
	GFP-ve	27	29	14	6	
	Total	27	31	38	33	
	proportion +ve	0	0.06	0.63	0.82	
rep 2	GFP+ve	0	4	20	31	*p = 0.048
	GFP-ve	60	55	30	23	
	Total	60	59	50	54	
	proportion +ve	0	0.07	0.40	0.57	
average		0	0.07	0.52	0.70	
Day 1&2/0-48hr data		WT		<i>cfp-1</i>		Probability of no genetic interaction
		EV	<i>sin-3</i>	EV	<i>sin-3</i>	
rep 1	GFP+ve	0	2	36	46	**p = 0.0011
	GFP-ve	65	67	36	18	
	Total	65	69	72	64	
	proportion +ve	0	0.03	0.50	0.72	
rep 2	GFP+ve	0	7	35	49	*p = 0.018
	GFP-ve	120	100	65	47	
	Total	120	107	100	96	
	proportion +ve	0	0.07	0.35	0.51	
average		0	0.05	0.43	0.61	

4.4.3 *hda-1* RNAi may enhance *unc-119::GFP* expression in *cfp-1* but not *set-2* mutants

As discussed in section 4.3 above, *hda-1* RNAi causes embryonic lethality and therefore feeding had to begin at the L1 stage, with the same worms scored on the first day of young adulthood. Thus, it was not possible to obtain a 24-48h dataset in these experiments; the effect measured would be influenced by the age of the worms being scored as well as duration of exposure to RNAi.

In *cfp-1* mutants, enhancement of *unc-119::GFP* expression by *hda-1* RNAi was statistically significant in the first replicate ($p=0.0042$) but not in the second ($p=0.07$) (Table 4.3). It is worth noting, however, that this is very close to the $p=0.05$ cut-off for significance.

In *set-2* mutants, by contrast, *hda-1* RNAi does not induce any more *unc-119::GFP* expression than in the WT controls (Table 4.3). This is consistent with the brood size assay results described in (Pokhrel et al., 2019).

Table 4.3 – *unc119::GFP* expression scoring data for *hda-1* RNAi in *cfp-1*, *set-2* and WT germlines. The table displays the raw number of germlines scored for each condition and the calculated proportion of *unc-119::GFP*-expressing (+ve) germlines. The reported *p* value for each experiment refers to the likelihood of observing this result assuming the null hypothesis of no interaction between the *cfp-1/set-2* and *hda-1* genes (see 2.4.5 for details of how this was calculated).

Day 1/0-24hr data		WT		<i>cfp-1</i>		Probability of no genetic interaction
		EV	<i>hda-1</i>	EV	<i>hda-1</i>	
rep 1	GFP+ve	0	4	12	14	**p = 0.0042
	GFP-ve	41	24	26	5	
	Total	41	28	38	19	
	proportion +ve	0	0.14	0.32	0.74	
rep 2	GFP+ve	1	12	11	24	ns p = 0.070
	GFP-ve	44	43	25	17	
	Total	45	55	36	41	
	proportion +ve	0.02	0.22	0.31	0.59	
average		0.01	0.18	0.31	0.66	

Day 1/0-24hr data		WT		<i>set-2</i>		Probability of no genetic interaction
		EV	<i>hda-1</i>	EV	<i>hda-1</i>	
rep 1	GFP+ve	0	10	4	12	ns p = 0.85
	GFP-ve	49	41	54	38	
	Total	49	51	58	50	
	proportion +ve	0	0.20	0.07	0.24	
rep 2	GFP+ve	0	15	4	7	ns p = 0.15
	GFP-ve	57	35	43	23	
	Total	57	50	47	30	
	proportion +ve	0	0.30	0.09	0.23	
average		0	0.25	0.08	0.24	

4.4.4 The proportion of *unc-119::GFP*-expressing germlines increases over generations in *cfp-1* mutants

Over the course of performing the RNAi experiments, it was noted that the proportion of *unc-119::GFP*-expressing germlines in the *cfp-1* empty vector (EV) controls was increased in older-generation worms (Table 4.4) (Figure 4.2A). In other words, those *cfp-1* mutants that had been maintained for many generations since outcrossing, compared to those more recently outcrossed, had greater *unc-119::GFP* expression. For example, in *cfp-1* empty vector (EV) controls at generation number 22 since outcrossing, 23.4% of germlines expressed *unc-119::GFP* in the 0-24h cohort. By generation 41 this had increased to 42.5%, and by generation 53 it was 73.3%.

Table 4.4 - *unc-119::GFP* expression increases over generations in *cfp-1* mutants on EV negative control RNAi plates. The table displays the raw number of germlines scored for each condition and the calculated proportion of *unc-119::GFP*-expressing (+ve) germlines. No 24-48h data is available for G41 and G53 because these were controls for the *hda-1* RNAi experiments in which this data was not collected.

Day 1/0-24h data	G22	G23	G40	G41	G45	G53
GFP+ve	11	15	12	17	20	22
GFP-ve	36	35	22	23	25	8
Total	47	50	34	40	45	30
% GFP+ve	0.23	0.30	0.35	0.43	0.44	0.73
Day 2/24-48h data	G22	G23	G40	G41	G45	G53
GFP+ve	24	20	24		20	
GFP-ve	30	30	14		9	
Total	54	50	38		29	
% GFP+ve	0.44	0.40	0.63	no data	0.69	no data

To rule out that this trend could be due to changes in experimental technique or plate quality, early (G5) versus later (G60) generation worms were compared using the same experimental design as the RNAi experiments (L4 mothers picked to new plates and shifted to 25°C, F1 progeny scored as YAs), except that the worms were grown on standard fresh NGM plates. The same trend was observed (Figure 4.2B)(Table 4.5). In both the 0-24h and 24-48h datasets the proportion of *unc-119::GFP*-expressing germlines was significantly greater in *cfp-1* mutants 56 generations post-outcrossing (G56) than those 5 generations post-outcrossing (G5) ($p=0.0046$ and $p=0.0024$ respectively). As seen for the *sin-3* and *hda-3* RNAi experiments, the pooled 0-48h data gives the most significant result.

Table 4.5 - *unc-119::GFP* expression is significantly increased in late vs early generation *cfp-1* mutants grown on NGM plates. The table displays the raw number of germlines scored for each condition and the calculated proportion of *unc-119::GFP*-expressing (+ve) germlines. The *p* values reported were calculated using Fisher's exact test.

Day 1/0-24h data	WT	<i>cfp-1</i> G5	<i>cfp-1</i> G56	G5 vs G56 significance
GFP+ve	0	9	19	** <i>p</i> = 0.0046
GFP-ve	46	36	18	
Total	46	45	37	
% GFP+ve	0%	20%	51%	
Day 2/24-48h data	WT	<i>cfp-1</i> G5	<i>cfp-1</i> G56	G5 vs G56 significance
GFP+ve	0	15	12	** <i>p</i> = 0.0024
GFP-ve	38	30	3	
Total	38	45	15	
% GFP+ve	0%	33%	80%	
Day 1&2/0-48h data	WT	<i>cfp-1</i> G5	<i>cfp-1</i> G56	G5 vs G56 significance
GFP+ve	0	24	31	*** <i>p</i> = 0.0002
GFP-ve	84	66	21	
Total	84	90	52	
% GFP+ve	0%	27%	60%	

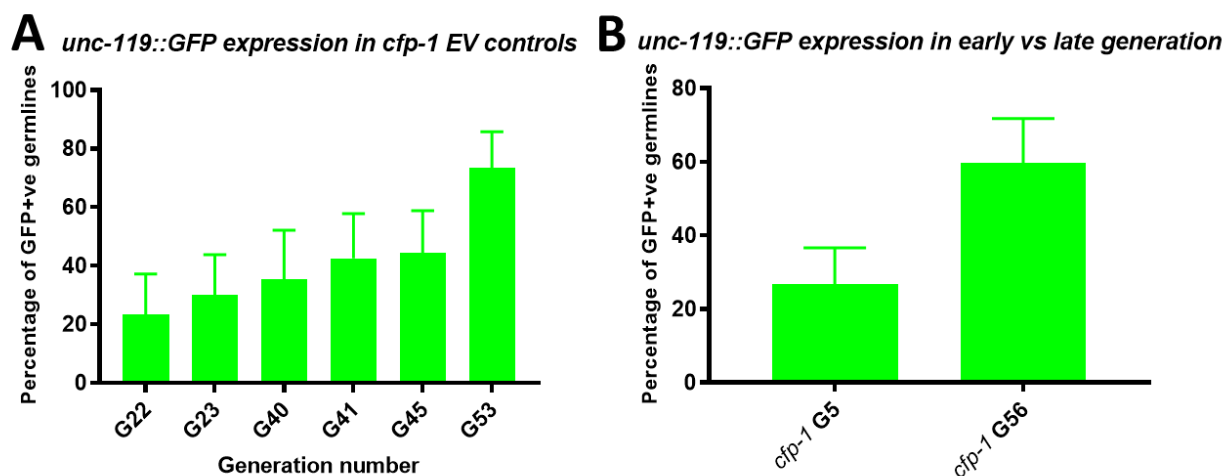


Figure 4.2 - *unc-119::GFP* expression increases over generations in *cfp-1* mutants. The left panel, A, presents the percentage of *unc-119::GFP* expressing germlines observed in *cfp-1(tm6369)* young adults in the 0-24h cohort grown at 25°C on EV control RNAi plates over the course of experiments reported in this chapter. The right panel, B, shows the percentage of *unc-119::GFP* expressing germlines observed in the pooled 0-48h *cfp-1(tm6369)* young adults grown at 25°C on NGM plates. In both graphs, error bars represent 95% confidence intervals.

4.5 Discussion

RNAi KD of *hda-3* significantly enhanced *unc-119::GFP* expression, and thus gene misregulation, in *cfp-1* mutants. The effect was similar but less significant for *hda-1* and *sin-3* RNAi in *cfp-1* mutants. Only *hda-1* RNAi was tested in *set-2* mutants, but it induced *unc-119::GFP* expression only to the same extent as the WT control.

Comparing the 24-48h to the 0-24h cohorts across experiments suggests that the induction of *unc-119::GFP* expression is increased with time in the *cfp-1* EV controls, which is most likely an effect of the longer exposure to the non-permissive temperature. The same effect was reported in *set-2* mutants (Robert et al., 2014), with the authors speculating that higher temperatures may increase the frequency of stochastic events at the single cell level that lead to acquisition of somatic cell fate. The induction of *unc-119::GFP* expression in *set-2* EV controls (7.7% average, Table 4.3), measured in the *hda-1* RNAi experiment, is similar to the 7.5% reported by (Robert et al., 2014) however the impact of temperature was not tested on *set-2* mutants here. It is noted that *unc-119::GFP* expression in *cfp-1* EV controls from the same experiment was even higher (31%, Table 4.3), indicating that *cfp-1* mutants are more sensitive to the impact of high temperature on germline transdifferentiation.

The WT control results reveal that while *hda-3* RNAi does not induce any *unc-119::GFP* expression in the WT background, *sin-3* RNAi did in approximately 7% of germlines scored (except in the 0-24h data of the first replicate). RNAi of *hda-1* did so to an even greater extent, with 22% of WT germlines expressing *unc-119::GFP*. Therefore, it appears that knockdown of *hda-1* and *sin-3* causes a strong loss of gene repression in the germline, independent of the *cfp-1* mutation.

The increase in *unc-119::GFP* expression over generations in *cfp-1* mutants was an interesting discovery which may explain why the second replicate data for the *sin-3* and *hda-1* RNAi experiments was less significant, as the *cfp-1* mutants used for EV controls had been maintained (under standard growth conditions) for an additional few weeks, and thus were at a later generation by the time the experiments were carried out. With the control *unc-119::GFP* expression already so high in the second replicate data, maybe it masked the enhancing effect of HDAC RNAi. This observation highlights the need for controlling for the generation number of *cfp-1* and probably other COMPASS mutants in future experiments.

Chapter 5. Wild-type descendants of COMPASS

mutants inherit the mutant long-lived

phenotype, and this is associated with a loss of

H3K4me2 and gain of H3K9me3 in the long-lived

descendants

5.1 Abstract

Following on from experiments reporting that WT descendants of *set-2(ok952)* mutants could inherit the extended longevity phenotype of their mutant ancestors (Greer et al., 2011), the reproducibility of this result was tested in the *set-2(bn129)* mutant background. Not only finding that WT descendants of *set-2(bn129)* mutants could, too, inherit the extended longevity phenotype, these experiments also showed that the strength and duration of the phenotype is increased when the long-lived WTs are descended from a late generation (maintained by self-fertilisation 60 generations post outcrossing, G60) *set-2(bn129)* mutant as compared to descendants from a G10 *set-2(bn129)* mutant. LC-MS/MS analysis of germline derived histones found that H3K4me2 levels are reduced in F3 and F4 WT descendants of *set-2(bn129)* mutants, and that H3K9me3 levels are increased, compared to WT controls.

5.2 Introduction

The previous chapter discussed the germline gene misregulation in *cfp-1(tm6369)* and *set-2(bn129)* mutants, manifest by *unc-119::GFP* reporter expression, which appears to worsen over generations.

This chapter will focus on another phenotype of COMPASS mutants, increased longevity (Greer et al., 2010), which is inherited by their wild-type descendants (Greer et al., 2011). This might be considered a different form of transgenerational epigenetic inheritance to the transgenerational worsening of gene expression control discussed previously, but potentially involving similar mechanisms.

A previous *C. elegans* study reported that RNAi-mediated knockdown of key components of the COMPASS complex, namely *ash-2*, *wdr-5* and *set-2*, extended lifespan (Greer et al., 2010). Greer et al. also

demonstrated that this lifespan extension required the presence of an intact adult germline and the continuous production of mature oocytes (Greer et al., 2010). The following year, Greer et al. reported that genetically wild-type descendants from long-lived *wdr-5* and *set-2* mutant ‘grandparents’ display a longevity phenotype, that persists for two further generations of the genetically wild-type descendants and then ceases to be inherited (Figure 5.1) (Greer et al., 2011).

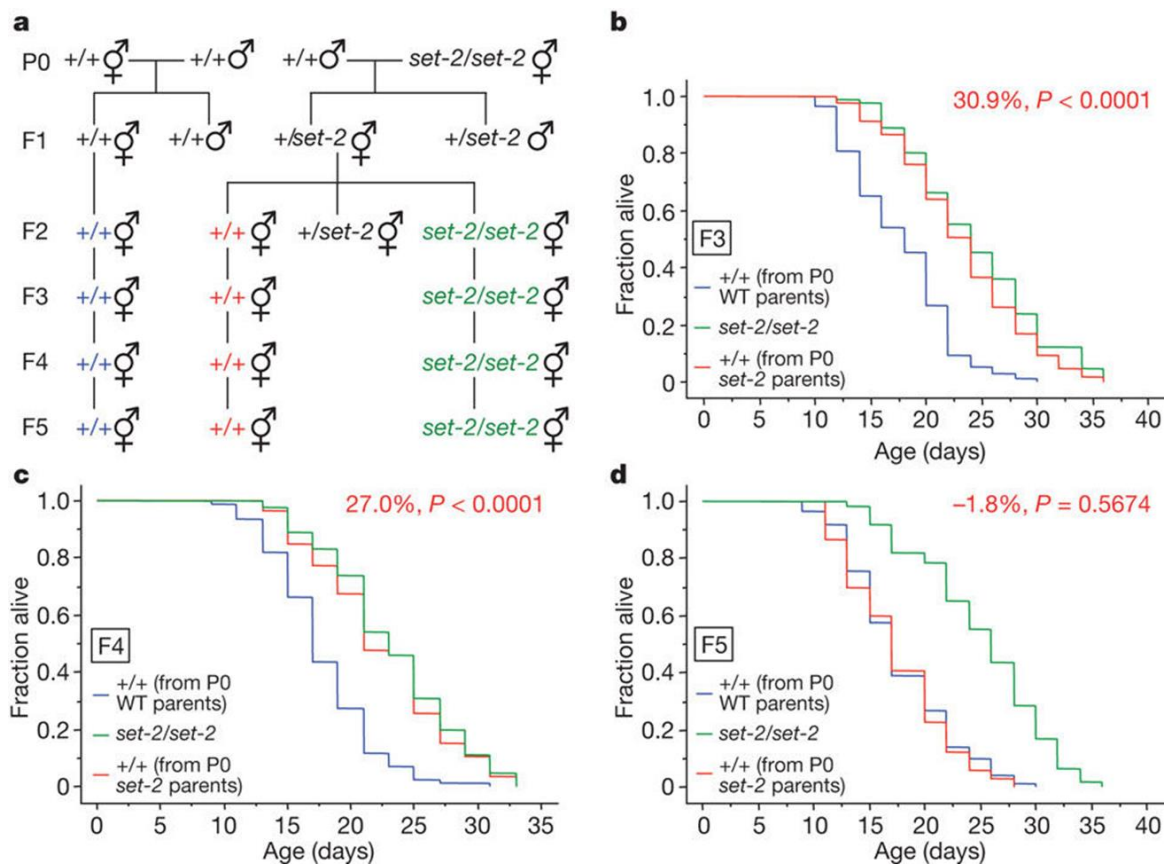


Figure 5.1 - Lifespan extension of *set-2(ok952)* descendants. Panel **a** illustrates the crosses that were set up and how the lineages were isolated to obtain the homozygous WT and homozygous *set-2(ok952)* mutant F3, F4 and F5. Panels **b**, **c** and **d** illustrate how the lifespan of the wild-type descendants (red line) is comparable to their mutant ‘cousins’ (green line) for the F3 and F4, but by the F5 the wild-type descendants’ lifespan is comparable with wild-type controls (blue line). *P* values are derived from log-rank comparisons of wild-type (blue line) controls with wild-type descendants (red line). Figure from (Greer et al., 2011).

More recently, *C. elegans* gene expression from an integrated heterochromatic multicopy array was found to be elevated, transgenerationally, by high temperature heat-shock (25°C) (Klosin et al., 2017). The elevated expression took 14 generations, after the growth temperature was reduced back to 20°C, to return to basal levels. In this example, the elevated expression was associated with heritable loss of the heterochromatin mark H3K9me3 in the vicinity of the transgene.

The different experiments discussed here report different numbers of generations over which transgenerational epigenetic inheritance can persist, but nevertheless there is always a limit, a generation by which the phenotype returns to the wild type. A sensible model to explain these observations might be that it takes multiple generations for wild-type levels of chromatin modifications to be restored after, for example, the return of enzymatic function or the removal of environmental stress.

This model is supported by the heat shock experiments in *C. elegans* (Klosin et al., 2017), which found that inactivating the H3K9 trimethylase *set-25* removed the temperature sensitivity of array expression and prevented transmission of variation in transgene expression. These observations led the authors to propose a model whereby the temperature-sensitive SET-25's activity is rapidly restored after a return to low temperature, but the re-establishment of repression via deposition of H3K9me3 takes many generations due to mechanisms present which maintain epigenetic states across life-cycles.

This model of gradual recovery (of H3K9me3) following restoration of H3K9 histone methyltransferase (HMT) function may not be applicable to COMPASS/H3K4me3. Immunocytochemistry data shows wild-type levels of H3K4me3 in the genetically wild-type long-lived descendants of *wdr-5* mutants (Greer et al., 2011), apparently ruling out the explanation that the phenotype persisted while H3K4me3 levels were recovering. The preferred explanation of Greer et al. was that the altered gene regulation in the long-lived wild-type descendants might be associated with heritable changes in H3K4me3 only at certain loci. Nevertheless, with the limitations of immunocytochemistry in providing robust quantitative data, the recovery model in the *set-2(bn129)* background was tested using LC-MS/MS.

5.3 Experimental design

A lifespan assay was pursued as described in (Greer et al. 2011) but using *set-2(bn129)* instead of *set-2(ok952)* to establish whether the results for the *set-2* mutant could be replicated. As discussed in 1.4.2.3, these deletion alleles have different predicted outcomes on SET-2 function; *set-2(ok952)* homozygotes, unlike *set-2(bn129)*, may not be entirely loss-of-function mutants.

In these experiments, the lifespan of ~90 worms per condition was monitored and recorded over an up to 40 day period, however long the lifespan of the most long-lived individual. To set up the first experiment, a cross was set up between *set-2(bn129)* hermaphrodites maintained for 10 generations post-outcrossing (G10), and WT males (Fig. 5.2). An F1 heterozygote was isolated from the cross progeny and then mutant and WT lineages were established from their F2 progeny, as described in 2.3.1.2.1. A simultaneous cross between WT hermaphrodites and males was set-up, and one F1 and one F2 singled, to prepare the control WT F3 (as illustrated in Figure 5.1A). Lifespan assays were conducted simultaneously for experimental and

control populations. As such, the assay was set up by picking 12 worms to 9 NGM/OP50 plates for each condition: control WT F3, WT descendant F3 (F3+/+) and *set-2(bn129)* descendant F3 (F3-/-). They had hatched on the same day, which was set as day 0, and they were checked on alternate days (2,4,6 etc. post-hatching) when any deaths and/or exclusions (see 2.5.3) were recorded. The F4 and F5 experiments were set up in the same way, four days after the preceding generation. When setting up the F3 lifespan assay, four worms each from the control WT F3, WT descendant F3 (F3+/+) and *set-2(bn129)* descendant F3 (F3-/-) populations were picked to separate NGM/OP50 plates to lay the F4s. The same procedure was carried out when setting up the F4 lifespan assay, four worms each from the control WT F4, WT descendant F4 (F4+/+) and *set-2(bn129)* descendant F4 (F4-/-) populations were picked to separate NGM/OP50 plates to lay the F5s.

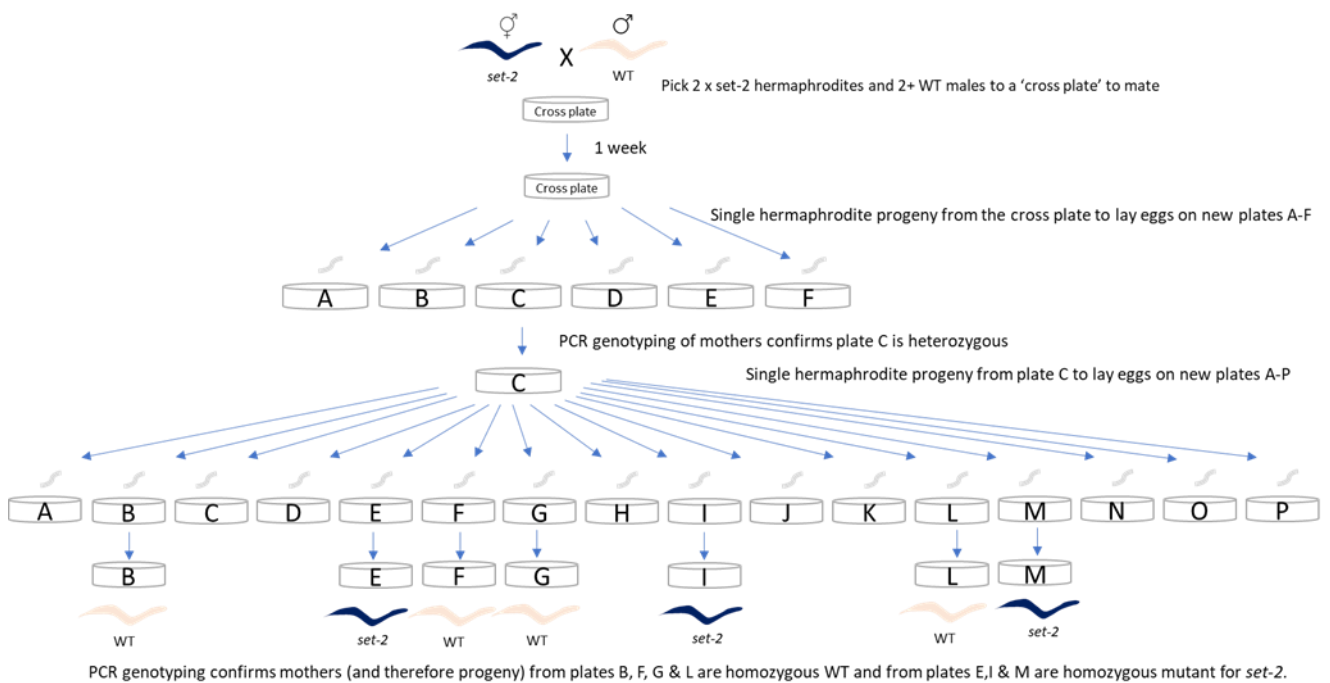


Figure 5.2 - The crossing and genotyping procedure used to isolate WT and mutant descendants of *set-2*.

Briefly, the cross is set up, a sample of the F1 is genotyped to identify and isolate a heterozygote. The heterozygote is left to self-fertilise and the F2 progeny are isolated, in sufficient number to be sure of isolating sufficient homozygotes. Finally, the F2 are genotyped, the genotype of their F3 progeny is inferred from this.

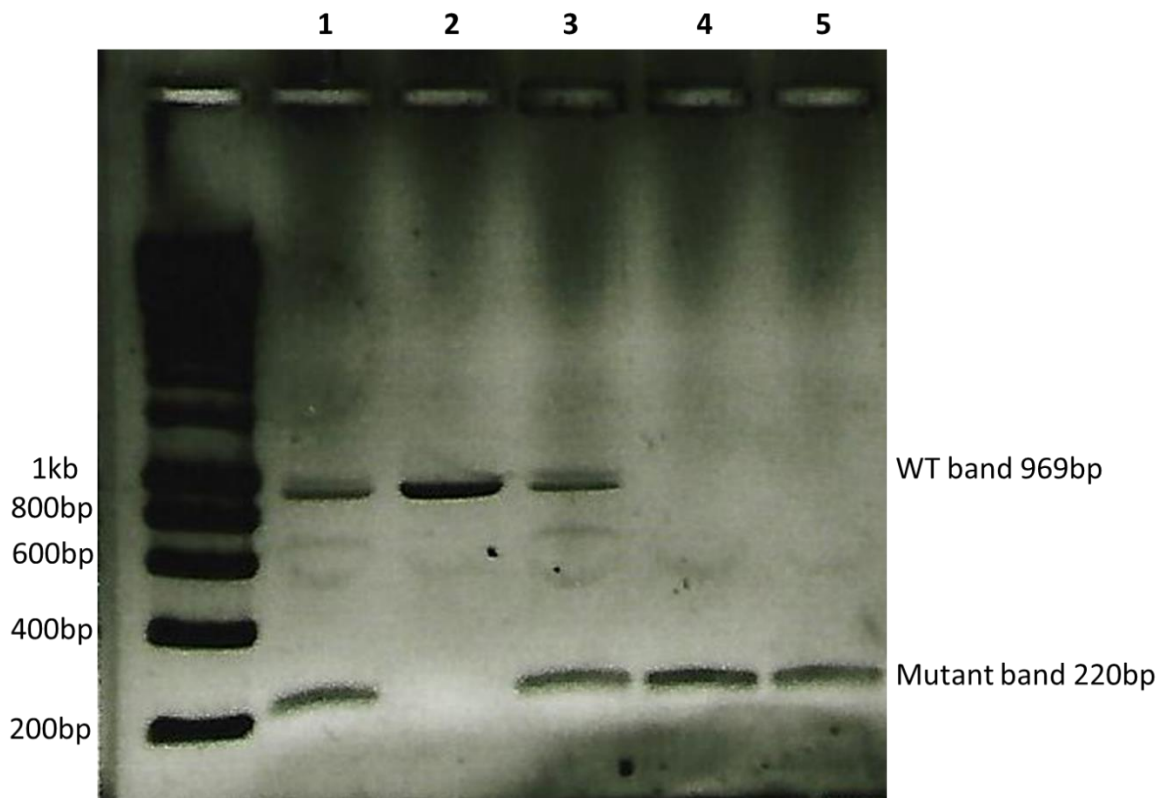


Figure 5.3 – Example genotyping results for the F2 progeny. The sizes of the DNA ladder bands are shown on the left, and those from the genotyping experiment on the right. See Table 2.4 for details of the primers. Lanes 1 and 3 show a heterozygous sample, lane 2 a homozygous WT, and lanes 4 and 5 a homozygous mutant.

The second experiment was set up just the same as the first except that *set-2(bn129)* hermaphrodites maintained for 60 generations post-outcrossing (G60), were used in the initial cross.

5.4 Results

5.4.1 WT descendants of *set-2(bn129)* also have increased longevity in the F3 and F4 generations

WT F3 and F4 descendants of *set-2(bn129)* G10 hermaphrodites had significantly ($p=0.0005$, $p<0.0001$, respectively) increased lifespan compared to WT controls of the corresponding generation (Table 5.1, Figure 5.2A). The median lifespan of the WT F3, WT F4 and WT F5 controls was consistent at 18 days, whereas for the F3 +/+ and F4 +/+ (G10 descendants) this was 20 days. The WT lifespans are similar to those reported in Greer et al., which were 18, 17 and 17 days for WT F3, F4 and F5, respectively. However, the F3 +/+ and F4 +/+ lifespans reported by Greer et al. were longer, at 24 and 23 days, respectively. So, while the trend is replicated, the extent of lifespan extension reported by Greer et al. was greater.

5.4.2 WT descendants of *set-2(bn129)* are not quite back to WT lifespan by the F5 generation

The F5 +/+ lifespan reported by Greer et al. is 17 days, the same as for their WT controls, indicating a complete recovery of the WT lifespan phenotype by the F5 generation. In this experiment, the lifespan of the F5 +/+ (G10 descendants) was 20 days (Table 5.1). Looking at the graph (Figure 5.2A), the F5 +/+ curve does appear shifted to the left, away from the F3 +/+ and F4 +/+ curves and closer to the WT control lines. The difference between WT F5 and F5 +/+ median lifespans is not significant ($p=0.0982$) but large enough to doubt that they could be considered the same. Taken together these observations suggest an incomplete transition back to WT longevity in the F5 +/+.

Table 5.1 – Median lifespan and statistical significance of the WT descendant’s longevity increase as compared to WT controls. P values are from pairwise comparison of each WT control with its generation-matched WT *set-2* descendants e.g. WT F3 vs F3 +/+, WT F4 vs F4 +/+.

Generation and genotype	Median lifespan (days)	Log-rank test significance
WT F3	18	
F3 +/+ G10	20	p = 0.0005
F3 +/+ G60	22	p < 0.0001
WT F4	18	
F4 +/+ G10	20	p < 0.0001
F4 +/+ G60	20	p < 0.0001
WT F5	18	
F5 +/+ G10	20	p = 0.0982
F5 +/+ G60	22	p = 0.0004

5.4.3 The extent of the longevity extension of WT descendants of *set-2(bn129)* mutants depends on the generation of the mutant ancestor

The longevity of genotypically wild-type descendants of *set-2(bn129)* mutants was now examined as before but with consideration to generation number since outcrossing. In other words, the longevity of descendants of generation 10 (G10) *set-2(bn129)* hermaphrodites was compared with descendants of later generation 60 (G60) hermaphrodites. Considering the transgenerational worsening of COMPASS phenotypes discussed in chapter 4, it was expected that the increased longevity of WT descendants from later generation *set-2(bn129)* mutants may be more pronounced compared to descendants from an earlier generation. This expectation was found to be correct. F3 +/+ descendants of a G60 *set-2(bn129)* hermaphrodite (F3 +/+ G60) had a median lifespan of 22 days, compared to 20 days for F3 +/+ G10, that was more significantly different (p<0.0001) from the 18 days of the WT F3 control (Table 5.1). The F4 +/+ G60 and F4 +/+ G10 median lifespans were both 20 days, but this just reflects the poor resolution of medians in this assay and qualitative comparison of the survival curves (Figure 5.4a & b) reveals a greater shift from the WT control for F4 +/+ G60. Moreover, while F5 +/+ G10 is not significantly different to WT F5, F5 +/+ G60 is still highly significantly more long-lived than WT F5. This observation suggests that not only does a later generation genetically wild-type *set-2(bn129)* ancestor have a more pronounced increase in lifespan, but the number of generations over which increased longevity can be inherited in WT descendants is increased.

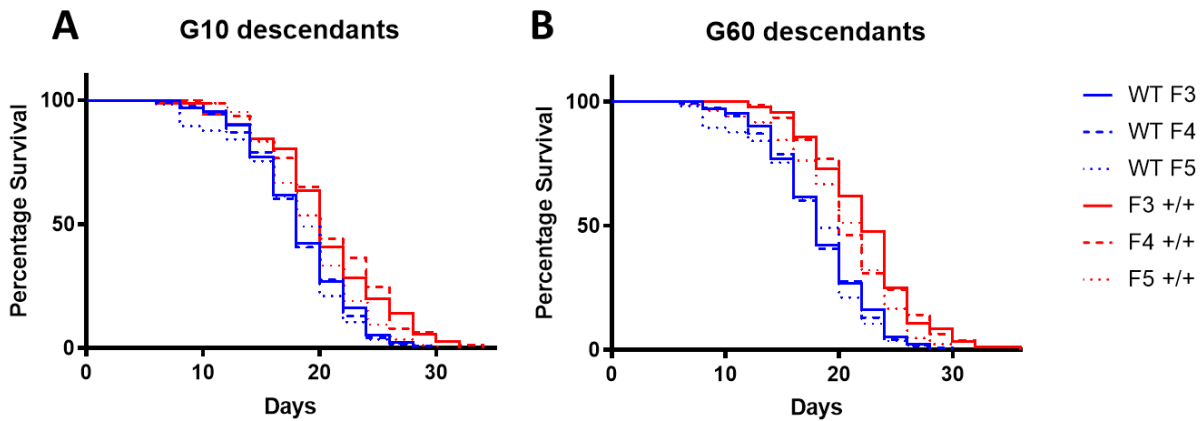


Figure 5.4 - Lifespan extension of WT descendants of *set-2(bn129)* mutants depends on the generation of the mutant ancestor. Survival data are presented as Kaplan-Meier plots, generated in GraphPad/Prism. Colour of the line differentiates between WT control and WT descendant, whereas style of the line differentiates the generation number post-cross.

5.4.4 Later generation *set-2(bn129)* mutants develop slowly compared to an earlier generation

Given that WT descendants of early and late generation *set-2(bn129)* mutants show phenotypic differences, then it seemed likely the *set-2(bn129)* mutants themselves would too. Increasing *unc-119::GFP* expression over generations in *set-2(bn129)* mutants has been reported previously (Robert et al., 2014). However, comparing the developmental progression of early versus late generation *set-2(bn129)* mutants revealed that the latter develop more slowly (Fig. 5.5). Given 60 hours to develop from the L1 stage, 84% of the wild type and 7% of the *set-2* G4 population had reached the young adult stage, but none of the *set-2* G60 population had (Table 5.2). The G60 population had 19% still in L3 or mid-L4 stages compared to 5% for the G10 population.

Table 5.2 – Developmental stages reached by WT, *set-2* generation 4 and generation 60 mutants after 60 hours at 20°C. The table displays the raw counts and calculated percentages. Developmental stage was ascertained by viewing gonad structure (2.6).

Stage	WT		<i>set-2</i> G4		<i>set-2</i> G60	
	Raw	Normalized	Raw	Normalized	Raw	Normalized
L3	0	0%	1	1%	4	4%
Mid L4	0	0%	4	4%	15	15%
Late L4	16	16%	87	88%	78	80%
Adult	81	84%	7	7%	0	0%
Total	97		99		97	

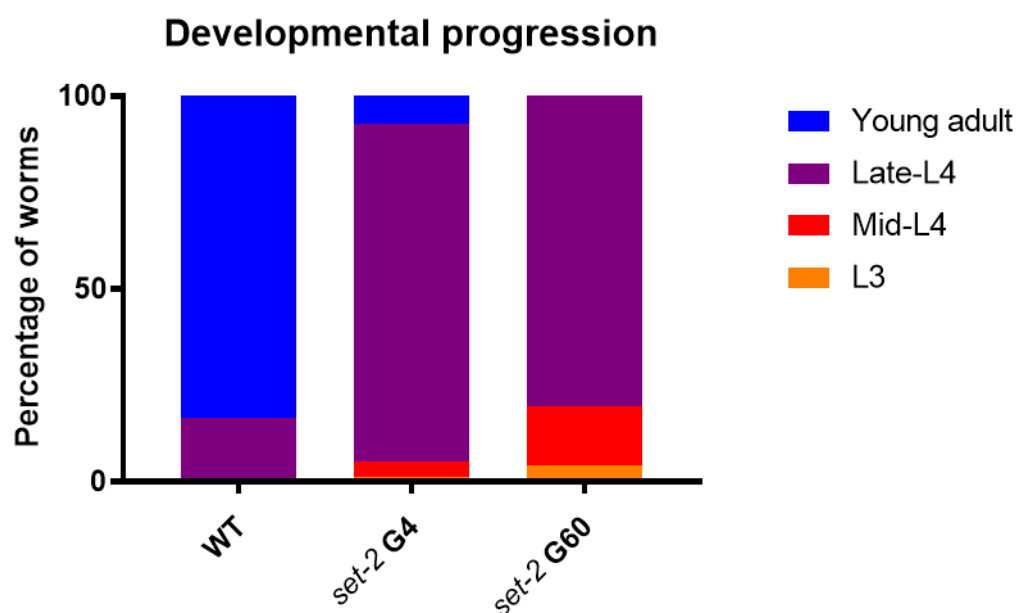


Figure 5.5 – Later generation (G60) *set-2(bn129)* mutants develop more slowly than an earlier generation (G4). The different coloured fractions of the bars represent the percentage of each developmental stage recorded after newly hatched wild-type and *set-2 (bn129)* larval stage 1 (L1) worms were allowed to develop for 60 hours at 20 °C.

5.4.5 Increased longevity in F3 and F4 WT descendants of *set-2(bn129)* mutants is accompanied by a decrease in H3K4me2 and an increase in H3K9me3

To clarify whether the longevity of WT descendants of *set-2* mutants is linked to delayed recovery of H3K4me3, germline-derived histones from F3, F4, F5 and F10 descendants were analysed by LC-MS/MS to compare histone PTM levels and identify H3K4 and H3K9 modifications that might be altered in the long-

lived F3 and F4 compared with WT. Unexpectedly, it was not H3K4me3 levels that were reduced but H3K4me2 levels (Figure 5.6A). H3K4me2 levels were reduced by approximately 30% in F3 +/+ and F4 +/+ *set-2(bn129)* descendants but recovered to the WT control level in the F5 +/+ and remained at that level in the F10 +/+ (Figure 5.6A). H3K4me3 levels appear to increase slightly in the F3 and F4 descendants, which could be due to increased conversion of H3K4me2 to H3K4me3 after restoration of COMPASS function (Figure 5.6A).

H3K9me3 levels, by contrast, are elevated in F3 and F4 +/+ *set-2(bn129)* descendants and fall to WT levels in the F5 and F10 (Figure 5.6B). H3K9me3 is a heterochromatic mark implicated in several *C. elegans* transgenerational epigenetic inheritance studies (Klosin et al., 2017, Woodhouse et al., 2018, Kalinava et al., 2018).

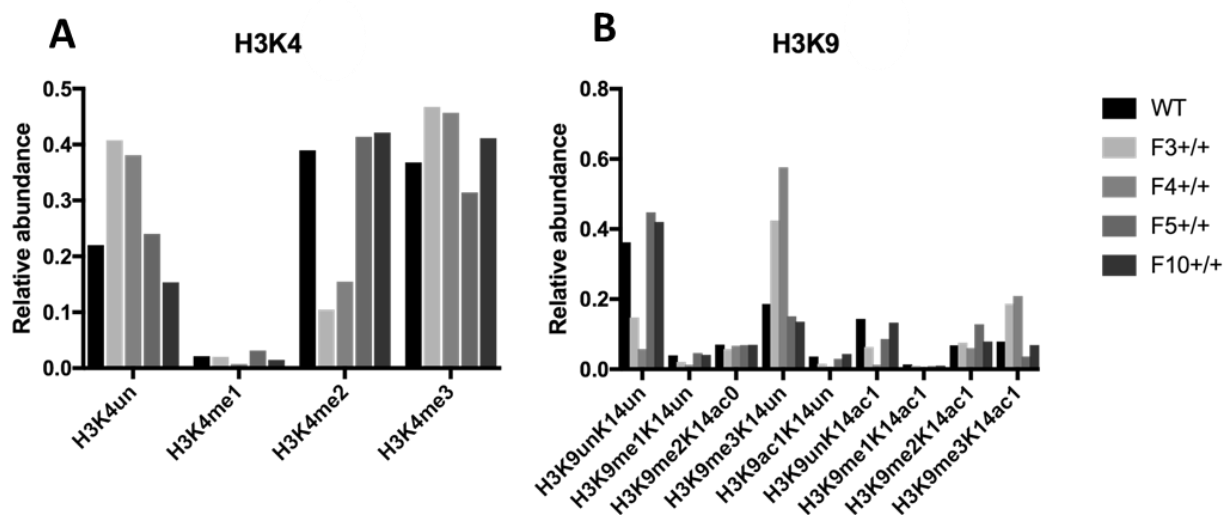


Figure 5.6 - LC-MS/MS profiling of H3 PTMs in germlines of WT descendants of *set-2(bn129)*. The data presented corresponds to the best dataset (see discussion) of three replicates. The height of the bars, or relative abundance, represents the percentage intensity of each modified peptide (e.g. H3K4me3, H3K9me3K14un) detected, out of the total intensity detected for all proteoforms of the peptide (displayed as labels on the x-axis).

5.5 Discussion

Taken together, these results confirm that WT descendants of *set-2(bn129)* mutants inherit an extended longevity phenotype, as reported for the descendants of *set-2(ok952)* mutants (Greer et al., 2011). H3K4me2 and H3K9me3 levels in wild-type descendants of the *set-2(bn129)* mutants show a recovery pattern consistent with the 5-generation timeline proposed in that same study. These experiments further

demonstrate that inheritance of increased longevity in descendants of *set-2(bn129)* mutants is more pronounced when these mutants have been without COMPASS function for a prolonged series of generations. Therefore, not only does loss of COMPASS function lead to worsened phenotypes in homozygous mutants over generations of self-fertilisation, but the inheritance of worsened phenotypes even in genetically WT descendants.

Chapter 6. Assessing the recovery of histone modifications in the germlines of wild-type descendants of COMPASS mutants using immunofluorescence microscopy

6.1 Abstract

Here, an immunofluorescence (IF) microscopy approach was employed to attempt to reproduce the LC-MS/MS finding that H3K4me2 levels are reduced in F3 and F4 WT descendants of *set-2(bn129)* mutants, and that H3K9me3 levels are increased, compared to WT controls. Overall, the IF results did not consistently support the mass spectrometry data for H3K4me2 but reaffirmed the H3K9me3 trend.

6.2 Introduction

With the mass spectrometry results showing such a striking change in H3K4me2 and H3K9me3 levels in F3 and F4 WT descendants of *set-2(bn129)* hermaphrodites (Chapter 5), it was important to attempt to reproduce these findings using an independent technique. Immunofluorescence microscopy (IF) uses fluorophore-conjugated antibodies to visualise the location and/or approximate amount of a substance present in a biological tissue. The technique is widely used in *C. elegans* and has been successfully used to compare levels of histone modifications in the germline (Greer et al., 2011, Woodhouse et al., 2018). Therefore, IF was used to visualise and compare levels of specific histone modifications: H3K4me2, H3K4me3, H3K9me2 and H3K9me3, in *C. elegans* adult germlines from WT (+/+) and mutant (-/-) descendants of *set-2(bn129)* mutants, and WT controls.

6.3 Experimental design

Like for the lifespan assays discussed in chapter 5, crosses were set up with a late (G70) *set-2(bn129)* mutant and an early (G10) *set-2(bn129)* mutant, in order to see whether phenotype severity was increased in the later generation mutant's descendants. A control cross was set up between a WT hermaphrodite and a WT male to provide the WT F3, F4 and F5 controls. Germlines from the F3, F4 and F5 WT (+/+) and mutant (-/-)

descendants of *set-2(bn129)* mutants, as well as from the WT F3, F4 and F5 controls, were dissected, immunostained and imaged.

Immunostaining was performed using the freeze-crack method as described previously (Strome and Wood, 1983). Briefly, YAs were dissected with a needle in 7 μ L M9 to release intact gonads on a poly-lysine coated slide (Figure 6.1). Each slide has three equal sized square wells, but only the left hand side (LHS) and right hand side (RHS) wells were used for dissection/staining/imaging, leaving the middle well empty, to use for labelling. The LHS/RHS distinction is important because worms were consistently dissected on the LHS first, then the RHS, to minimise the risk of smudging. The extra few seconds that the LHS sample spent at room temperature before snap freezing compared to the RHS, may have affected the results, as discussed later in the chapter.

The poly-lysine coating encourages sticking of the gonad to the slide, which is achieved by pressing lightly with a coverslip. Then the slide was transferred to a pre-chilled metal plate on dry ice to snap freeze the tissue to allow reagent penetration, before fixing in methanol and applying antibodies.

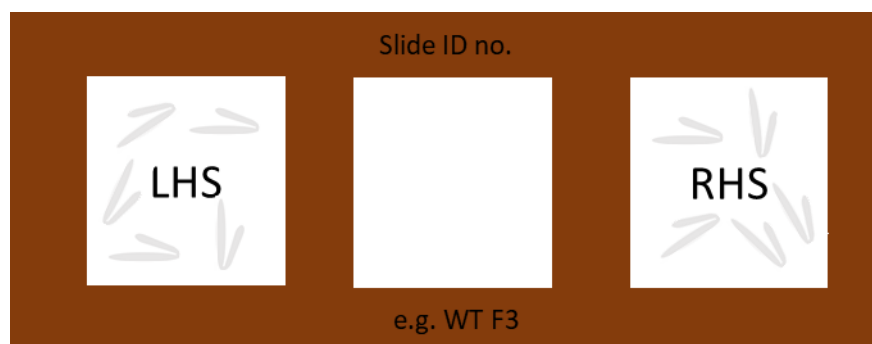


Figure 6.1 – Slides as prepared for imaging. Prior to dissection they were labelled by etching into the brown outer coating in the places illustrated. After coverslip sealing the glass in the middle well was labelled with details of the antibodies used. Only the outer wells were used for the experiment, which are labelled as left hand side (LHS) and right hand side (RHS) because this is how they appear in relation to the orientation of the labelling on the slide. LHS and RHS is how they will be referred to in the main text. The dissected gonads are represented in grey.

Primary antibodies for each histone modification were tested in WT YA *C. elegans* germlines at different concentrations to determine an optimal working concentration for the experiments. The antibody specificity was tested by dot blot (2.7.4). However, this only worked for the H3K4me3 and H3K9me3 antibodies, which were shown to be specific for H3K4me3- and H3K9me3-modified peptides, with no cross-reactivity with the me2-modified peptide, or any of the other trimethylated lysines assayed. Validation information for the

specificity of the H3K4me2 (C15200151 Diagenode) and H3K9me2 (Wako MABI0317) antibodies from the manufacturer had to be relied upon. These antibodies did still work in the concentration optimisation tests so were active.

By using secondary antibodies labelled with Alexa Fluor dyes whose excitation and emission wavelengths overlap minimally (Fig. 6.2), it was possible to image an anti-H3 control (anti-sheep Alexa Fluor 405) and two histone modifications (anti-mouse Alexa Fluor 488 and anti-rabbit Alexa Fluor 594) simultaneously.

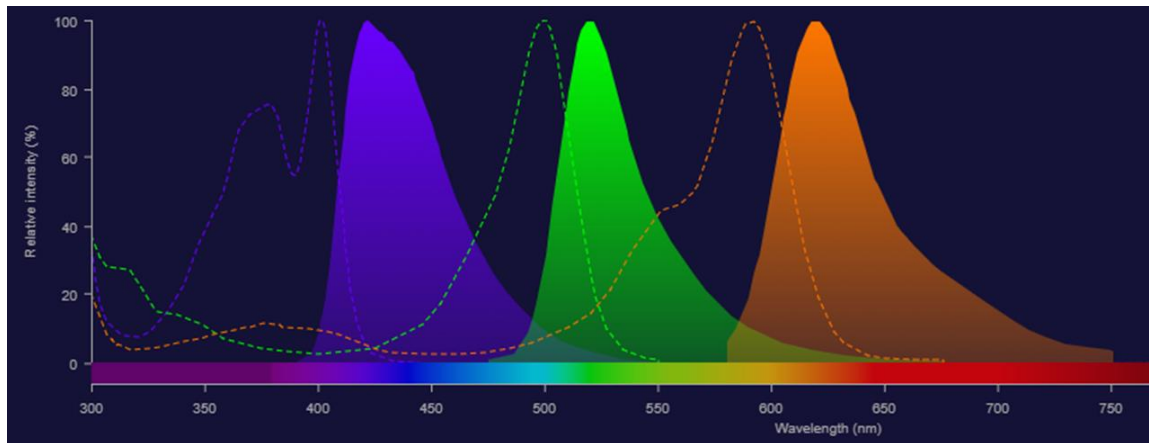


Figure 6.2 – A comparison of the excitation and emission wavelengths (dotted line and filled curves, respectively) of the Alexa Fluor 405, 488 and 594 dyes. Plotted using the Thermofisher ‘SpectraViewer’ tool.

The following three primary antibody combinations were used in each experiment:

1. Anti-H3K4me2 (mouse); anti-H3K4me3 (rabbit); anti-H3 (sheep)
2. Anti-H3K4me2 (mouse); anti-H3K9me3 (rabbit); anti-H3 (sheep)
3. Anti-H3K9me2 (mouse); anti-H3K9me3 (rabbit); anti-H3 (sheep)

In discussion of the results hereafter, these antibody combinations will be referred to as ‘antibody combination 1’, ‘antibody combination 2’, ‘antibody combination 3’. The LHS and RHS well of each slide was stained with the same antibody combination, and thus provides a technical replicate for each assay. 6 gonads were imaged from the LHS and 6 from the RHS, giving 12 gonads in total for each experimental condition, that is, generation (F3, F4, F5), genotype (WT, +/- or *set-2(bn129)* mutant -/-) and antibody combination. Imaging was performed on a Zeiss LSM880 + Airyscan Upright Microscope, using confocal microscopy with laser excitation of the three fluorophores. Fluorescent signal intensity was quantified from a subset of manually selected nuclei (Fig. 6.4) in the pachytene zone of the gonad (Fig. 6.3) using FIJI software (2.6.7). All fluorescence intensities from the histone modifications were divided by the fluorescence intensity from the anti-H3 control antibody (Alexa Fluor 405), to normalise the data.

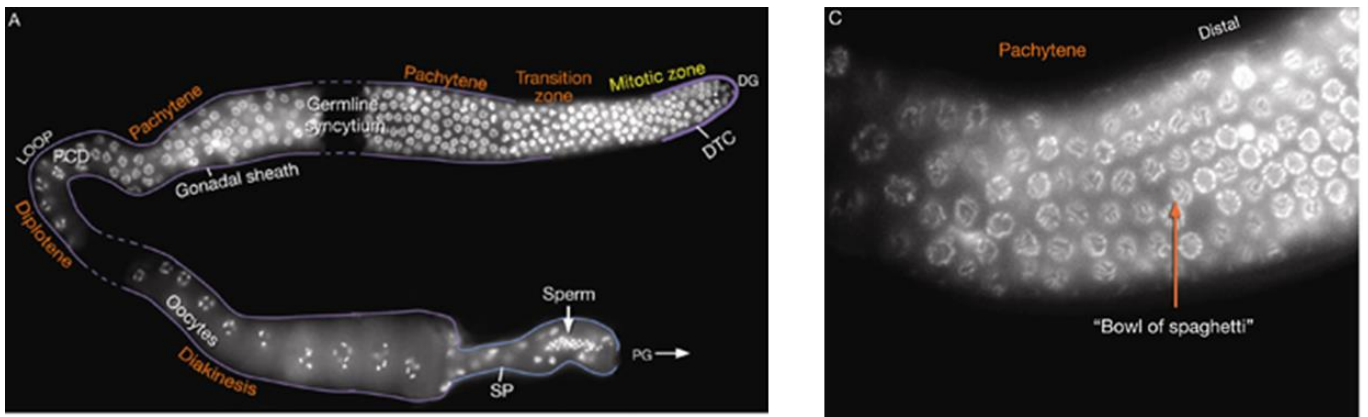


Figure 6.3 – The organisation of the hermaphrodite *C. elegans* gonad. The mitotic zone, at the distal tip, contains a stem cell population. Nuclei (stained here with DAPI) progress through the transition zone, where they begin to undergo meiosis I, in to the pachytene zone, where they continue to progress through the stages meiosis as described in the orange text. The ‘bowl of spaghetti’ morphology of nuclei approaching the loop is due to homologous chromosomes aligning side by side. Adapted from wormatlas.org.

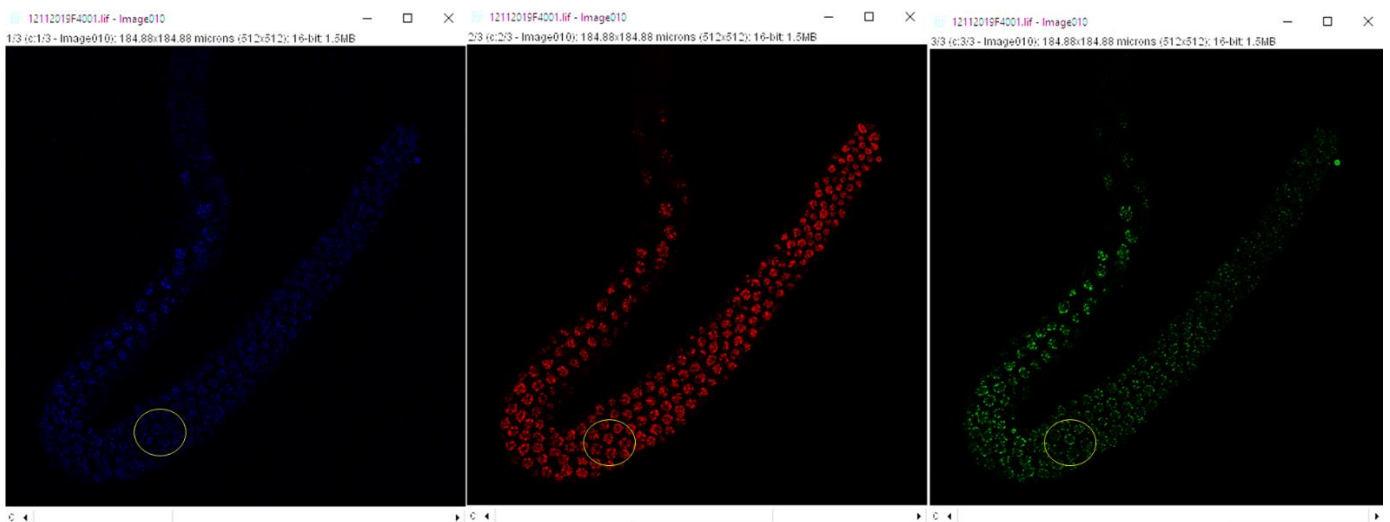


Figure 6.4 Selection of nuclei for fluorescence quantification – The yellow ring illustrates positioning (in the pachytene zone, just before the loop when the nuclei acquire the ‘bowl of spaghetti morphology’) and the approximate size of area selected in FIJI. The green and red fluorescence signal is divided by the blue to normalise for size and number of nuclei in the area selected.

6.4 Results

In evaluating these data there were three main questions to address:

1. Do the data agree with the mass spectrometry results showing H3K4me2 levels reduced and H3K9me3 levels increased in the F3 +/+ and F4 +/+ descendants of *set-2(bn129)* mutants? It should be noted that the mass spectrometry data was only derived from histones extracted from the germlines of early-generation descendants, so it is only appropriate to compare the early generation IF results with the mass spectrometry findings.
2. Is there agreement in the data between different experimental conditions? E.g., do H3K4me2 levels appear consistent in the IF whether co-imaged with H3K4me3 or H3K9me3?
3. Is there agreement between results for the early generation and late generation experiments?

The following considers each of these questions for each histone modification tested.

6.4.1 H3K4me2

6.4.1.1 Agreement with mass spectrometry results

There are four independent sets of results for H3K4me2 for each of the three generations, F3, F4, F5 (Table 6.1). There are separate results for antibody combinations 1 and 2, for both the descendants of early and late-generation *set-2(bn129)* mutants.

In all these sets of results, whether for early or late generations of *set-2(bn129)* mutants, there is no statistically significant difference in mean fluorescence signal for H3K4me2 between WT F3 and F3 +/+, or WT F4 and F4 +/+ (Table 6.1) (Figure 6.5 A, B, D, E, G, H, J, K). Indeed, the F3 +/+ mean appears lower than the WT F3 mean in only one of the four sets of results (early descendants, antibody combination 1) (Figure 6.5A). These data, therefore, do not support the mass spectrometry results which suggested that the H3K4me2 levels are reduced in the germlines of F3 +/+ and F4 +/+ descendants of *set-2(bn129)* mutants as compared to WT controls (chapter 5, supplementary figure 8.1). According to the IF, the wild type descendants of *set-2(bn129)* mutants appear to exhibit the wild type levels of H3K4me2.

6.4.1.2 Agreement between antibody combinations

In the early generation set, there is good agreement between antibody combination 1 and 2 data with regard to the significantly lower mean fluorescence of the mutant (-/-) descendants as compared to the corresponding WT controls (Figure 6.5 A,B,C,D,E,F). This is consistent with the reduction in H3K4me2 seen in

set-2(bn129) mixed embryos (chapter 1) and germlines (Supplementary Figure 8.1A and 8.2A) and lends confidence to the reliability of these results for the *set-2(bn129)* early generation.

With regard to the difference between the mean anti-H3K4me2 fluorescence of WT descendants compared to the WT control, however, antibody combination 1 gave a small reduction in signal for F3 +/+ and F4 +/+ gonads (Figure 6.5 A&B) compared to WT controls, whereas with antibody combination 2 the mean fluorescence signals are very similar (Figure 6.5 D&E). The F5 +/+ mean fluorescent signal is above WT F5 for both antibody combinations, significantly so for antibody combination 1 (Figure 6.5 C). This increase, in F5 +/+ levels of H3K4me2 compared to WT, was not seen in the mass spectrometry data (chapter 5, supplementary figure 8.1A & 8.2A).

6.4.1.3 Agreement between early and late generation experiments

In the late generation data the reduction in mean fluorescence in mutants compared to WT is much less consistent than in the early generation data, only appearing for combination 1 F4 (Figure 6.5 H) and combination 2 F3 (Figure 6.5 J). The significant increase in F4 -/- compared to WT F4 (Figure 6.5 K) looks anomalous given the wide distribution of the data points and the disagreement with the general trend.

Overall, the trends in the late generation data appear less consistent than in the early generation data for H3K4me2. There is more agreement between combination 1 and 2 in the early data than between early and late in the combination 1, or 2 data.

6.4.1.4 Agreement between LHS and RHS

In two of the graphs (Figure 6.5B & K) the data points for F4 -/- appear to form two separate clusters, suggesting potential variation between the LHS and RHS. Indeed, calculation of the individual means for each side reveals that in the combination 1 early F4 -/- data (Figure 6.3B) the LHS mean is 0.57 and the RHS mean 0.26. In the combination 2 late F4 -/- data (Figure 6.5K) the LHS mean is 0.816 and the RHS mean is 1.14. The difference could be due to one side of the slide not having properly set mounting medium, or not being completely sealed and drying out prior to imaging, consequently exaggerating the signal. As discussed in the experimental design (6.3), there was a consistent difference in the way the LHS and RHS were prepared because worms were always dissected on the LHS first, then the RHS, to minimise the risk of smudging. Consequently, the dissected gonads on the LHS spent an additional few seconds at room temperature, compared to the RHS, before snap freezing. Here, though, whether the RHS or LHS is affected appears to be random, because in the first of these examples the LHS mean is higher, while in the second example the RHS mean is higher.

It is impossible to be certain which side to accept as the truer mean. Given the overall trend of H3K4me2 fluorescence being lower in *set-2(bn129)* mutants than WT controls across experiments, it seems likely that the lower mean is the more acceptable in these examples.

Table 6.1 – Summary of results for all IF experiments for H3K4me2. The ‘combination’ column refers to whether H3K4me2 was co-stained and co-imaged with H3K4me3 (combination 1) or with H3K9me3 (combination 2). The p-values presented in the ‘significance’ column refer to the results of a one-way ANOVA with Tukey’s multiple comparison, comparing the WT control mean with either the WT descendant (+/+) or mutant descendant (-/-) mean. The abbreviation ‘ns’ means the p-value is non-significant, * means $p < 0.05$, ** means $p < 0.01$, *** means $p < 0.001$ and **** means $p < 0.0001$ (at this level of significance the exact p-value is not reported by GraphPad/Prism, in which the statistical testing was performed).

Generation and genotype	Histone modification	Combination	Early (G10) descendants		Late (G70) descendants	
			Mean fluorescence	Significance	Mean fluorescence	Significance
WT F3	H3K4me2	1	0.7236		0.3003	
F3 +/+			0.6145	ns p = 0.2064	0.3287	ns p = 0.4962
F3 -/-			0.3279	****p < 0.0001	0.3073	ns p = 0.9568
WT F4			0.5666		0.4266	
F4 +/+			0.4689	ns p = 0.1203	0.4	ns p = 0.6956
F4 -/-			0.4182	*p = 0.0111	0.2635	****p < 0.0001
WT F5			0.251		0.6291	
F5 +/+			0.3053	***p = 0.0004	0.7873	ns p = 0.1762
F5 -/-			0.1808	****p < 0.0001	0.6035	ns p = 0.9483
WT F3		2	0.7878		0.2408	
F3 +/+			0.7923	ns p = 0.9968	0.2648	ns p = 0.1227
F3 -/-			0.3388	****p < 0.0001	0.1964	**p = 0.002
WT F4			0.444		0.3777	
F4 +/+			0.466	ns p = 0.7358	0.3599	ns p = 0.9848
F4 -/-			0.2211	****p < 0.0001	0.9783	****p < 0.0001
WT F5			0.3031		0.3513	
F5 +/+			0.3248	ns p = 0.3338	0.3088	ns p = 0.1754
F5 -/-			0.2197	****p < 0.0001	0.3188	ns p = 0.3369

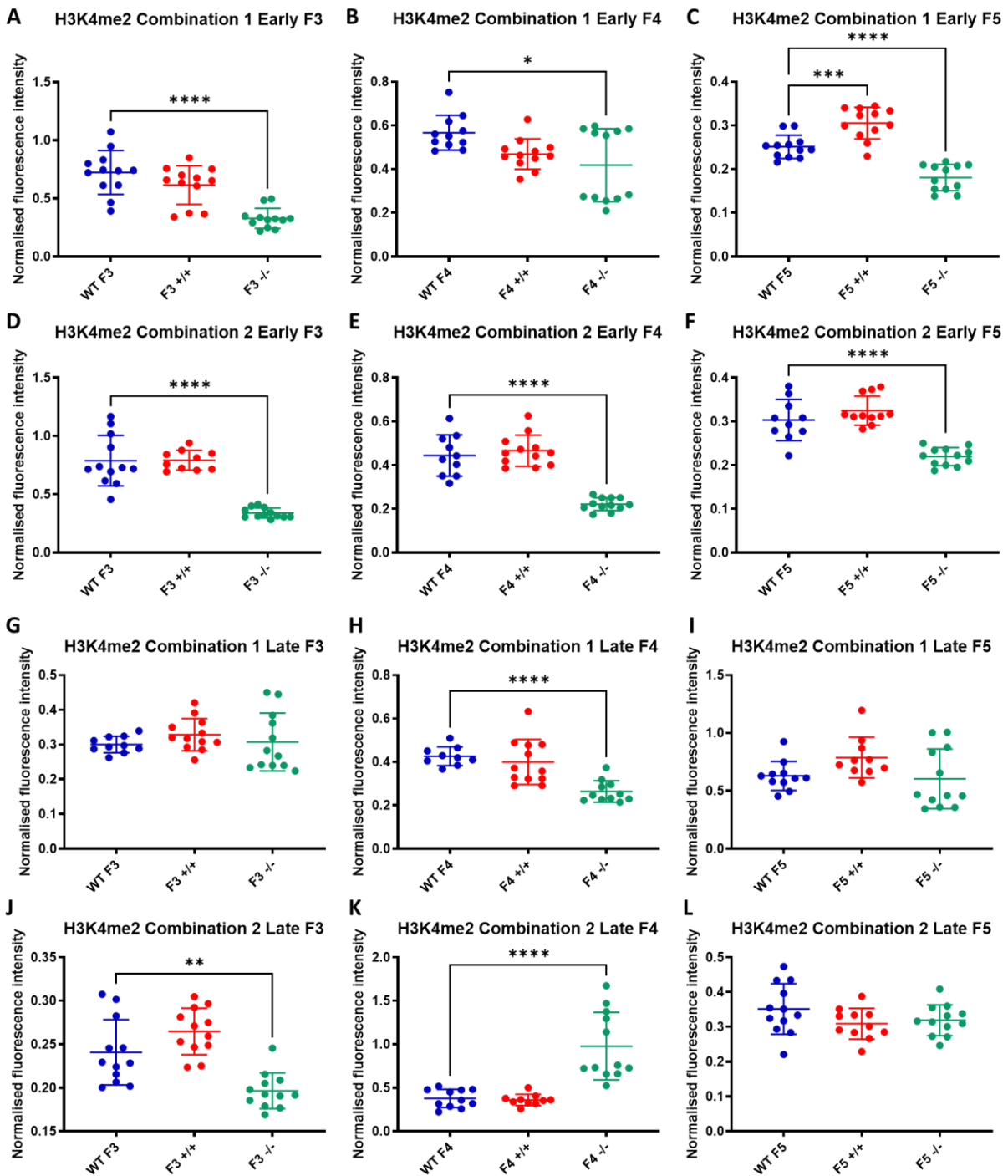


Figure 6.5 – Anti-H3K4me2 immunofluorescence intensity in the germline of descendants of early and late generation *set-2(bn129)* mutants. Each circle represents a single fluorescence intensity value for each individual gonad imaged. Here, fluorescence intensities for the LHS and RHS of each slide have been combined, giving an *n* of 12 for each condition. The middle horizontal line represents the mean of the 12 data points, and the shorter horizontal lines show the standard deviation. All fluorescence intensities have been normalized to the H3 control. There is variation in fluorescence intensities between graphs, but each graph represents a separate experiment (i.e. slides prepared and imaged on different days) and thus comparisons are only appropriate within graphs rather than between them.

6.4.2 H3K4me3

H3K4me3 was stained and imaged with only antibody combination 1. Therefore, there are just single datasets to compare, for early and late generations of *set-2(bn129)* mutants (Table 6.2).

6.4.2.1 Agreement with mass spectrometry results

The H3K4me3 early F3 results show a reduction in mean fluorescence intensity for F3 +/+ and F3 -/- compared to WT F3 (Figure 6.6A), which differs to the LC-MS/MS results in which F3 +/+ and F3 -/- H3K4me3 relative abundance is slightly elevated above WT control levels (supplementary figure 8.1B). Therefore, the agreement between the different techniques is poor for the F3 data.

Agreement is better for the F4 and F5 generations. F4 -/- and F5 -/- H3K4me3 levels are lower compared to WT controls in both IF and LC-MS/MS data, albeit only slightly in the IF results. However, the increase in F5 +/+ H3K4me3 levels compared to WT controls (Figure 6.6C) in the IF data is not seen in the MS data (supplementary figure 8.1B).

6.4.2.2 Agreement between early and late generation experiments

Largely there is consistency between the two datasets suggesting little change in H3K4me3 levels arising from the length of maintenance of the *set-2(bn129)* genotype through successive generations. Indeed, there is little difference in the levels of H3K4me3 between WT controls, and the wild type and mutant descendants of *set-2(bn129)* mutants apparent in any of these IF data.

The mean anti H3K4me3 fluorescence signal for F3 +/+ is slightly reduced compared to that for WT F3 in both early and late *set-2(bn129)* generations (Figure 6.6 A&D). The elevation of the signal for F5 +/+ above WT F5 is also seen in both datasets, but the early generation mean fluorescence signal is similar between WT F5 and F5 -/- (Figure 6.6 C) while the late generation signal for F5 -/- is significantly increased above that for WT F5 (Figure 6.6 F). Given the wide range and distribution in data points in the late generation F5 -/- data, this distinction is not convincing. Indeed, comparison of the LHS (1.40) and RHS (0.72) means for the late generation dataset suggests that one side is skewing the mean. It's not possible to be certain that the LHS or RHS is the erroneous one, but because the -/- signal is equal to or lower than the WT control signal in all other experiments, it seems more likely that the LHS is artificially skewing the mean fluorescence signal higher than it should be, rather than the RHS skewing the mean lower.

Table 6.2 - Summary of results for all IF experiments for H3K4me3. The p-values presented in the 'significance' column refer to the results of a one-way ANOVA with Tukey's multiple comparison, comparing the WT control mean with either the WT descendant (+/+) or mutant descendant (-/-) mean. Combination 1 refers to the use of anti-H3K4me2 antibodies in combination with anti-H3K4me3 antibodies. The abbreviation 'ns' means the p-value is non-significant, * means $p < 0.05$, ** means $p < 0.01$, *** means $p < 0.001$ and **** means $p < 0.0001$ (at this level of significance the exact p-value is not reported by GraphPad/Prism, in which the statistical testing was performed).

Generation and genotype	Histone modification	Combination	Early (G10) descendants		Late (G70) descendants	
			Mean fluorescence	Significance	Mean fluorescence	Significance
WT F3	H3K4me3	1	0.5153		0.408	
F3 +/+			0.4092	*p = 0.0293	0.3328	*p = 0.0428
F3 -/-			0.338	***p = 0.0002	0.387	ns p = 0.7528
WT F4			0.4903		0.5043	
F4 +/+			0.4874	ns p = 0.9976	0.4456	ns p = 0.3882
F4 -/-			0.4684	ns p = 0.8668	0.3603	**p = 0.0064
WT F5			0.3165		0.6633	
F5 +/+			0.4183	**p = 0.0011	0.8117	ns p = 0.4638
F5 -/-			0.2742	ns p = 0.2574	1.003	*p = 0.0265

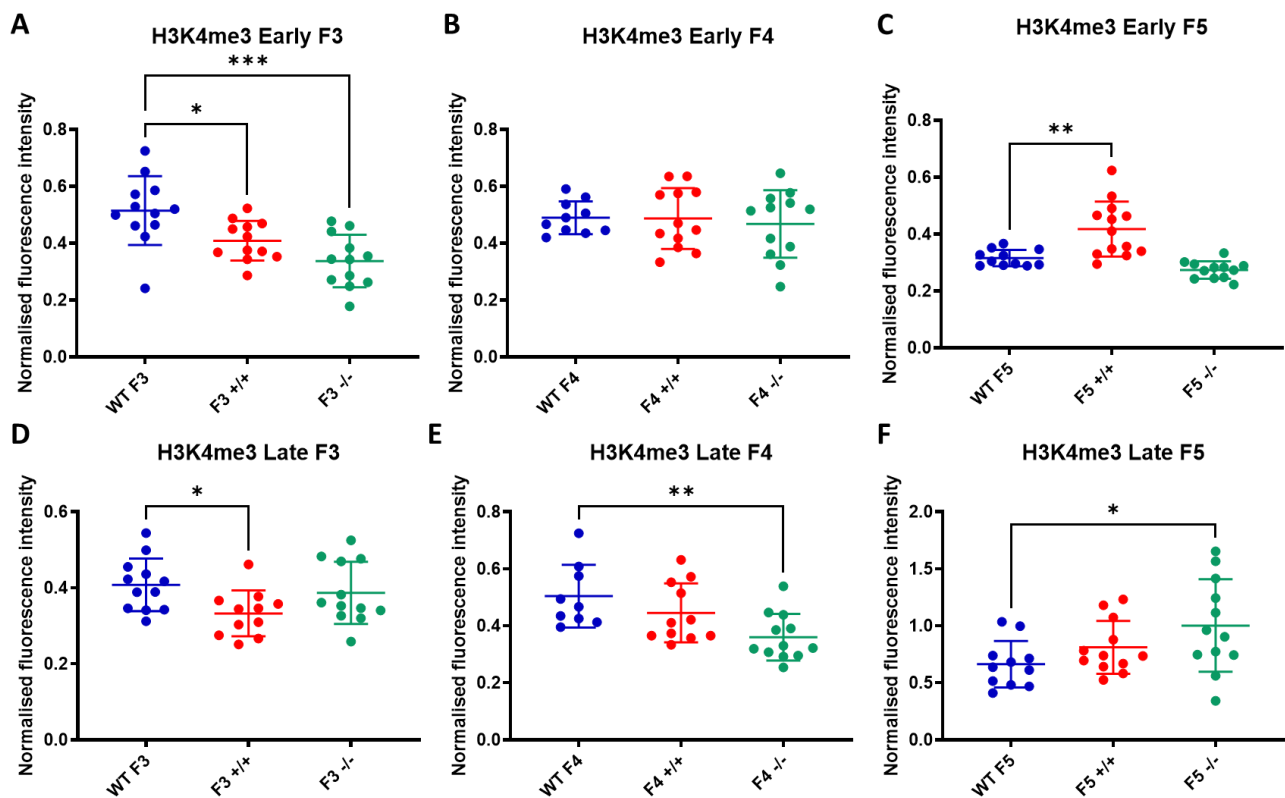


Figure 6.6 – Anti-H3K4me3 immunofluorescence intensity in the germline of descendants of early and late generation *set-2(bn129)* mutants. Each circle represents a single fluorescence intensity value for each individual gonad imaged. Here, fluorescence intensities for the LHS and RHS of each slide have been combined, giving an n of 12 for each condition. The middle horizontal line represents the mean of the 12 data points, and the shorter horizontal lines show the standard deviation. All fluorescence intensities have been normalized to the H3 control. There is variation in fluorescence intensities between graphs, but each graph represents a separate experiment (i.e. slides prepared and imaged on different days) and thus comparisons are only appropriate within graphs rather than between them.

6.4.3 H3K9me2

As with the H3K4me3 data, H3K9me2 was stained and imaged only with antibody combination 3 and so there are just single early and late generation datasets to compare (Table 6.3).

6.4.3.1 Agreement with mass spectrometry results

Overall, the agreement between the LC-MS/MS and early generation IF data for H3K9me2 is good. For F3 descendants, there is little difference between WT, +/+ and -/- H3K9me2 fluorescence intensity (Figure 6.7A) or relative abundance (Supplementary Figure 8.1C). The same similarity is seen for WT F4 and F4 +/+ (Figure 6.7B) in the IF and LC-MS/MS data. However, while H3K9me2 relative abundance appears elevated compared to WT in the LC-MS/MS data, in the IF results it is reduced (Figure 6.7B), the opposite trend.

F5 +/+ fluorescence intensity and relative abundance is elevated slightly above WT (Figure 6.7C, Supplementary Figure 8.1C). In the IF data, unlike the similar relative abundance seen in the LC-MS/MS results (Supplementary Figure 8.1C), F5 -/- fluorescence intensity is significantly higher than WT F5. However, the F5 -/- data is very broadly distributed, and inspection of the LHS (0.51) and RHS (0.77) F5 -/- means reveals the latter may be skewing the mean higher than it should be, just as with the H3K4me3 late F5 data discussed previously. Therefore, this may not be a biologically valid result.

6.4.3.2 Agreement between early and late generation experiments

The concurrence between the early and late datasets is poor, as in the latter there is a significant increase in F3 +/+, F3 -/-, F4 +/+ and F4 -/- H3K9me2 levels above their respective WT controls (Figure 6.7 D & E), which is not seen in the former (Figure 6.7 A & B). However, the consistency of the trends in the late F3 and F4 data suggest that this might be a biologically meaningful result and a true consequence of the length of maintenance of the *set-2(bn129)* genotype through successive generations.

Table 6.3 - Summary of results for all IF experiments for H3K9me2. The p-values presented in the 'significance' column refer to the results of a one-way ANOVA with Tukey's multiple comparison, comparing the WT control mean with either the WT descendant (+/+) or mutant descendant (-/-) mean. Combination 3 refers to the use of anti-H3K9me2 antibodies in combination with anti-H3K9me3 antibodies. The abbreviation 'ns' means the p-value is non-significant, * means $p < 0.05$, ** means $p < 0.01$, *** means $p < 0.001$ and **** means $p < 0.0001$ (at this level of significance the exact p-value is not reported by GraphPad/Prism, in which the statistical testing was performed).

Generation and genotype	Histone modification	Combination	Early descendants		Late descendants	
			Mean fluorescence	Significance	Mean fluorescence	Significance
WT F3	H3K9me2	3	0.5961		0.1705	
F3 +/+			0.5942	ns p = 0.9992	0.2877	****p < 0.0001
F3 -/-			0.67	ns p = 0.3514	0.2713	***p = 0.0001
WT F4			0.425		0.3558	
F4 +/+			0.4497	ns p = 0.5776	0.4684	*p = 0.023
F4 -/-			0.3358	**p = 0.0021	0.4666	*p = 0.0256
WT F5			0.3006		1.601	
F5 +/+			0.3721	ns p = 0.2895	1.239	ns p = 0.0841
F5 -/-			0.6398	****p < 0.0001	1.531	ns p = 0.904

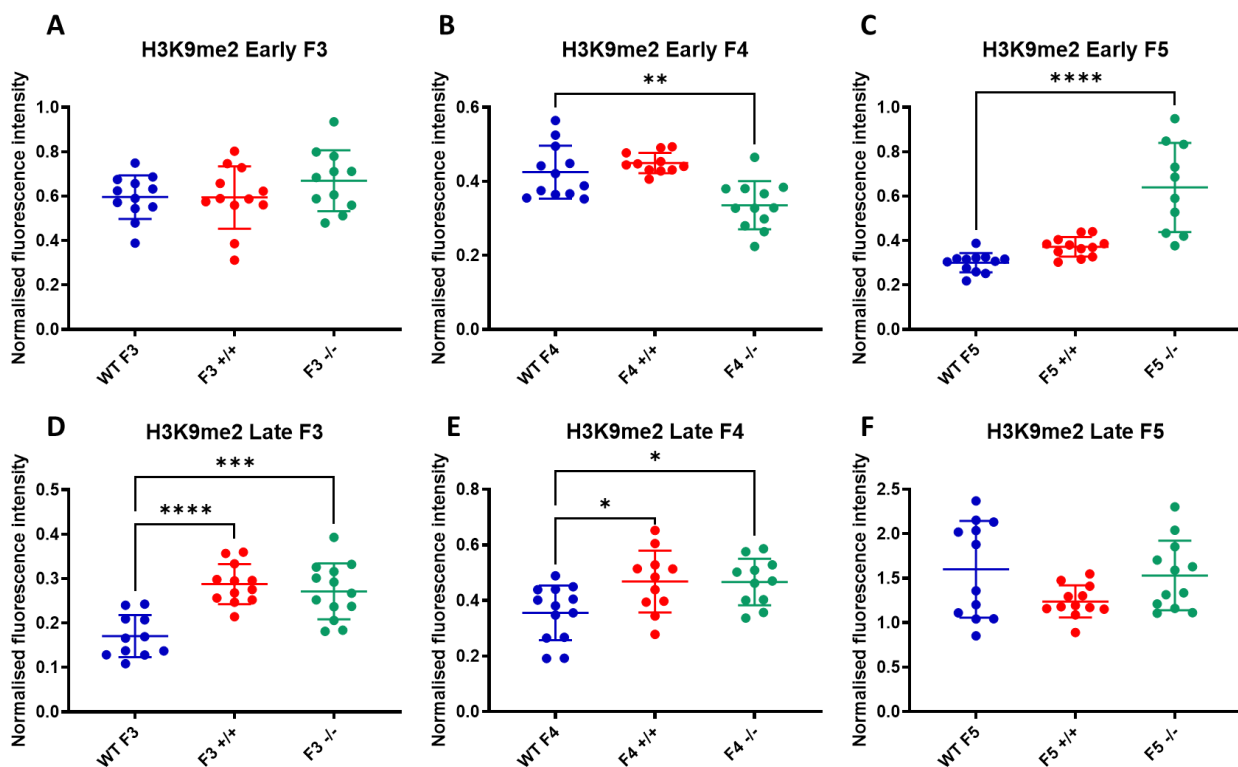


Figure 6.7 – Anti H3K9me2 immunofluorescence intensity in the germline of descendants of early and late generation *set-2(bn129)* mutants. Each circle represents a single fluorescence intensity value for each individual gonad imaged. Here, fluorescence intensities for the LHS and RHS of each slide have been combined, giving an n of 12 for each condition. The middle horizontal line represents the mean of the 12 data points, and the shorter horizontal lines show the standard deviation. All fluorescence intensities have been normalized to the H3 control. There is variation in fluorescence intensities between graphs, but each graph represents a separate experiment (i.e. slides prepared and imaged on different days) and thus comparisons are only appropriate within graphs rather than between them.

6.4.4 H3K9me3

As for H3K4me2, H3K9me3 was stained and imaged with two antibody combinations (2 and 3), once with early generation descendants and once with late generation descendants for each, so there are four independent datasets to compare.

6.4.4.1 Agreement with mass spectrometry results

Unlike the H3K4me2 results, there is better agreement between the IF and LC-MS/MS results for H3K9me3. The LC-MS/MS results revealed increased relative abundance of H3K9me3 in F3 +/+, F3 -/- and F4 +/- compared to WT (Supplementary Figure 8.1). In the early (Figure 6.8A) antibody combination 2 data, F3 +/+ has a significantly higher mean anti-H3K9me3 immunofluorescence signal than WT F3. However, in the early generation antibody combination 3 data the F4 +/- mean is significantly higher than the WT F4 (Figure 6.8E), despite this not being the case for the early generation antibody combination 3 F3 data (Figure 6.8D).

6.4.4.2 Agreement between antibody combinations

In the early generation data, the agreement between antibody combinations 2 and 3 is poor for the F3 results. With antibody combination 3, the F3 fluorescence intensity means are very similar (Figure 6.8D) whereas with antibody combination 2, the F3 +/+ and -/- fluorescence intensity means are significantly higher than for the WT control (Figure 6.8A). In the F4 data, the slight elevation of the F4 -/- fluorescence intensity mean above that for the WT F4 control is consistent between the two antibody combinations (Figure 6.8B & E), but F4 +/- has a significantly higher fluorescence intensity mean than the WT F4 control only in the antibody combination 3 data. The nonsignificant difference between the WT F5 control and F5 +/- fluorescence intensity means is consistent across both antibody combinations, but the F5 -/- fluorescence intensity mean is elevated significantly above that for the WT F5 control only in the antibody combination 3 data (Figure 6.8F). Comparison of the LHS (0.13) and RHS (0.2) fluorescence intensity means for these early generation antibody combination 3 data suggests that, as with the H3K9me2 data, which came from the same slide, the RHS data is skewing the overall mean upwards. Therefore, the LHS fluorescence intensity mean (0.13) may better reflect the biology. When this LHS fluorescence intensity mean is compared against the WT F5 control (0.119, Table 6.4), the difference is non-significant ($p=0.2422$).

In the late generation data, agreement between antibody combinations 2 and 3 is good for the WT control versus the +/- data, with an increase in the fluorescence signal intensity for +/- compared to WT for F3 and F4 (Figure 6.8 G, H, J, K), and a reduction in the F5 (Figure 6.8 I & L) for both antibody combinations. The relationship between WT control and -/- mutants is consistent but more variable in strength, with -/-

fluorescence intensity means elevated above WT in the F3 and F4 generations to different extents: slightly for antibody combination 2 F3 (Figure 6.8 G), significantly for antibody combination 3 F3 and F4 (Figure 6.8 J & K) and highly significantly for antibody combination 2 F4 (Figure 6.8 H). In the F5 late generation data, antibody combinations 2 and 3 agree with regard to the significantly lower mean fluorescence intensity in F5 +/- compared to WT F5 (Figure 6.8 I & L). However, the F5 -/- vs WT trend is opposite between the two antibody combinations, being lower in combination 2 (Figure 6.8 I) but similar in combination 3 (Figure 6.8 L).

Table 6.4 - Summary of results for all IF experiments for H3K9me3. The p-values presented in the 'significance' column refer to the results of a one-way ANOVA with Tukey's multiple comparison, comparing the WT control mean with either the WT descendant (+/+) or mutant descendant (-/-) mean. Combination 3 refers to the use of anti-H3K9me2 antibodies in combination with anti-H3K9me3 antibodies. The abbreviation 'ns' means the p-value is non-significant, * means $p < 0.05$, ** means $p < 0.01$, *** means $p < 0.001$ and **** means $p < 0.0001$ (at this level of significance the exact p-value is not reported by GraphPad/Prism, in which the statistical testing was performed).

Generation and genotype	Histone modification	Combination	Early descendants		Late descendants	
			Mean fluorescence	Significance	Mean fluorescence	Significance
WT F3	H3K9me3	2	0.1262		0.1905	
F3 +/+			0.1748	**p = 0.0019	0.3033	****p < 0.0001
F3 -/-			0.1744	**p = 0.0021	0.192	ns p = 0.9939
WT F4			0.4514		0.1237	
F4 +/+			0.5198	ns p = 0.7011	0.1475	ns p = 0.4544
F4 -/-			0.539	ns p = 0.5612	0.6001	****p < 0.0001
WT F5			0.1558		1.441	
F5 +/+			0.1434	ns p = 0.1982	0.9805	*p = 0.0311
F5 -/-			0.1461	ns p = 0.3666	0.8813	**p = 0.0063
WT F3		3	0.1503		0.1234	
F3 +/+			0.1457	ns p = 0.8945	0.1544	*p = 0.0255
F3 -/-			0.1443	ns p = 0.8316	0.1542	*p = 0.0236
WT F4			0.3001		0.09683	
F4 +/+			0.4491	**p = 0.0044	0.1253	***p = 0.0007
F4 -/-			0.317	ns p = 0.9108	0.1175	**p = 0.0098
WT F5			0.1119		0.2509	
F5 +/+			0.1188	ns p = 0.8701	0.187	***p = 0.0001
F5 -/-			0.1697	***p = 0.0009	0.2654	ns p = 0.521

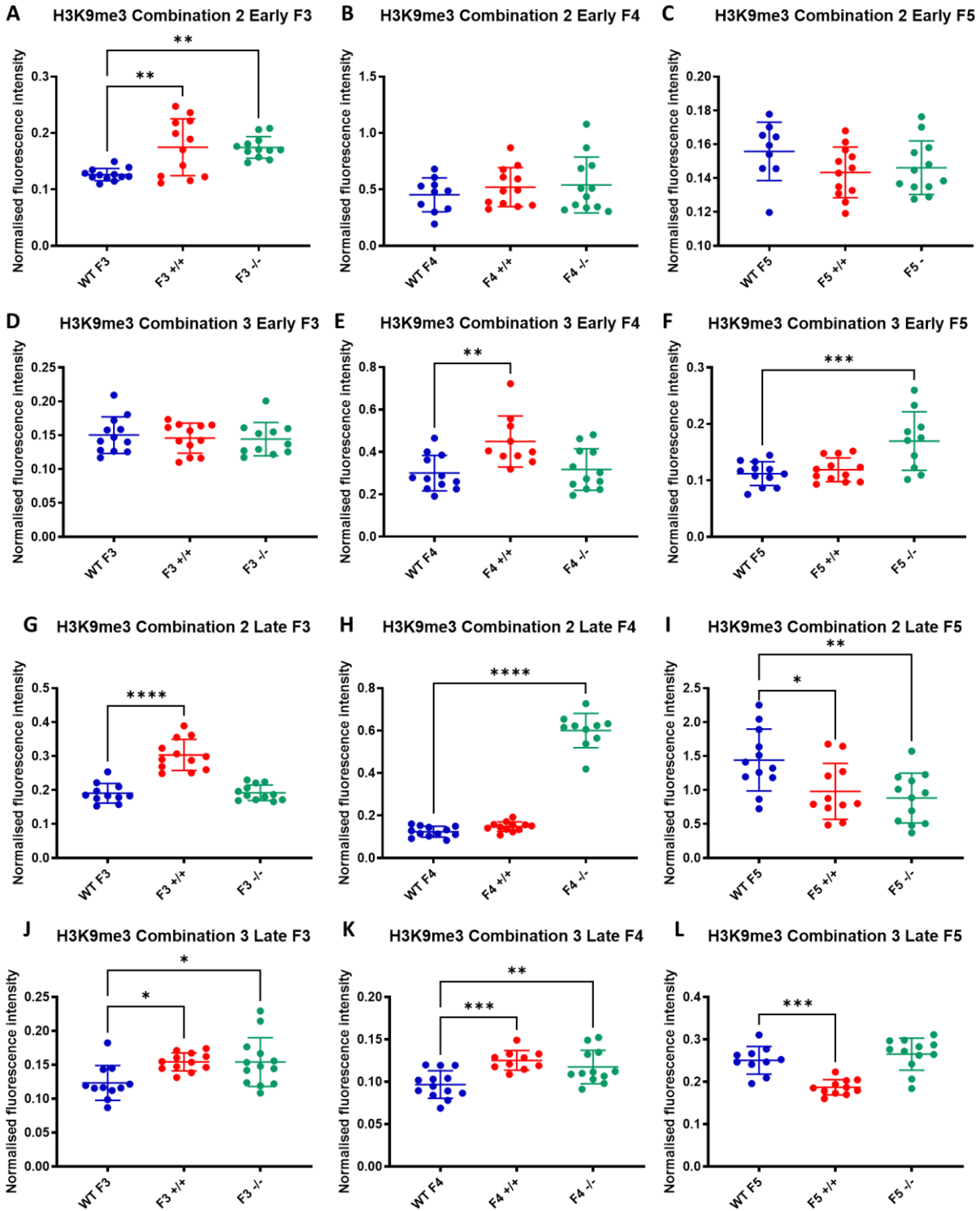


Figure 6.8 – Anti-H3K9me3 immunofluorescence intensity in germline of descendants of early and late generation set-2(bn129) mutants. Each circle represents a single fluorescence intensity value for each individual gonad imaged. Here, fluorescence intensities for the LHS and RHS of each slide have been combined, giving an n of 12 for each condition. The middle horizontal line represents the mean of the 12 data points, and the shorter horizontal lines show the standard deviation. All fluorescence intensities have been normalized to the H3 control. There is variation in fluorescence intensities between graphs, but each graph represents a separate experiment (i.e. slides prepared and imaged on different days) and thus comparisons are only appropriate within graphs rather than between them.

6.5 Discussion

There is a great deal of variability in these data. To be confident in asserting any biological conclusions requires repeating of experiments and more extensive validation of the antibodies. Several different primary antibodies against the same histone modification could be assessed.

It is striking how similar the trends for some modifications in the same antibody combination appear. For example, antibody combination 3 for late generation H3K9me3 and H3K9me2 (Figure 6.5 D, E, F, Figure 6.6 J, K, L), or antibody combination 1 for early generation H3K4me2 and H3K4me3 in the F3 data (Figure 6.4 A, Figure 6.3 A). This might lead to concern about the antibodies and potential cross-reaction between them. However, the antibody combination 3 early generation H3K9me3 and early generation H3K9me2 data look less similar, especially for the F4s. Likewise for antibody combination 1, the early F4 and F5 and late data across all generations show different patterns for H3K4me2 and H3K4me3. This would suggest that the antibodies are not interacting and any similarity between results for different modifications is coincidental and a biologically valid result.

Moreover, the 2D imaging and analysis performed here may be too simplistic to accurately compare the quantity of histone modifications in a nucleus. Perhaps a 3D approach, taking Z-stacks through nuclei as reported in (Mutlu et al., 2018), would provide a more accurate and more reproducible assessment of the biology.

Nevertheless, the reproducibility of the H3K9me3 increase in F3 +/- germlines is an encouraging result, especially since it agrees with the LC-MS/MS results reported for this modification in chapter 5. Changes in H3K9me3 levels have been reported in multiple studies of transgenerational epigenetic inheritance (Woodhouse et al., 2018, Kalinava et al., 2018, Ni et al., 2016), so to link this to COMPASS/H4K4me3 by two different experimental methods helps build a fuller picture of histone modification cross-talk in this active area of research.

Chapter 7. Discussion

7.1 Summary of results

The LC-MS/MS study of mixed embryo-derived histones described in Chapter 3 revealed a reduction in H3K4me3 to negligible amounts in *set-2(bn129)* and *cfp-1(tm6369)* mutants, compared to WT mixed embryos. This result suggests that *set-2(bn129)* and *cfp-1(tm6369)* are indeed loss-of-function alleles, whose effect is to render the COMPASS complex incapable of depositing H3K4me3, globally. This agrees with ChIP-seq data showing that the genomic distribution of H3K4me3 is similarly reduced in *cfp-1* and *set-2* mutants, implying that CFP-1 is needed for SET-2 activity at promoters (Beurton et al., 2019).

There was also a significant reduction in H3K4me2, albeit not to the same extent as H3K4me3, suggesting that there are other H3K4 KMTs active in *set-2* and *cfp-1* mutant embryos. These are not so well characterised as the COMPASS complex owing to the challenges of experimenting with the *set-16* mutant (Fisher et al., 2010). However, recent immunoprecipitation experiments have found that WDR-5.1 is a member of the SET-16/MLL complex, as well as COMPASS (Beurton et al., 2019).

An incidental finding of the LC-MS/MS study was that H3K14ac and H3K27ac levels are significantly elevated in *cfp-1* and *set-2* mutants and to very similar extents. The similarity of this result for both mutants, combined with the drastic reduction of H3K4me3 observed, indicates that the increase in these acetylation modifications, K14ac and K27ac, is a direct consequence of the lack of COMPASS function and loss of H3K4me3 (Figure 7.1).

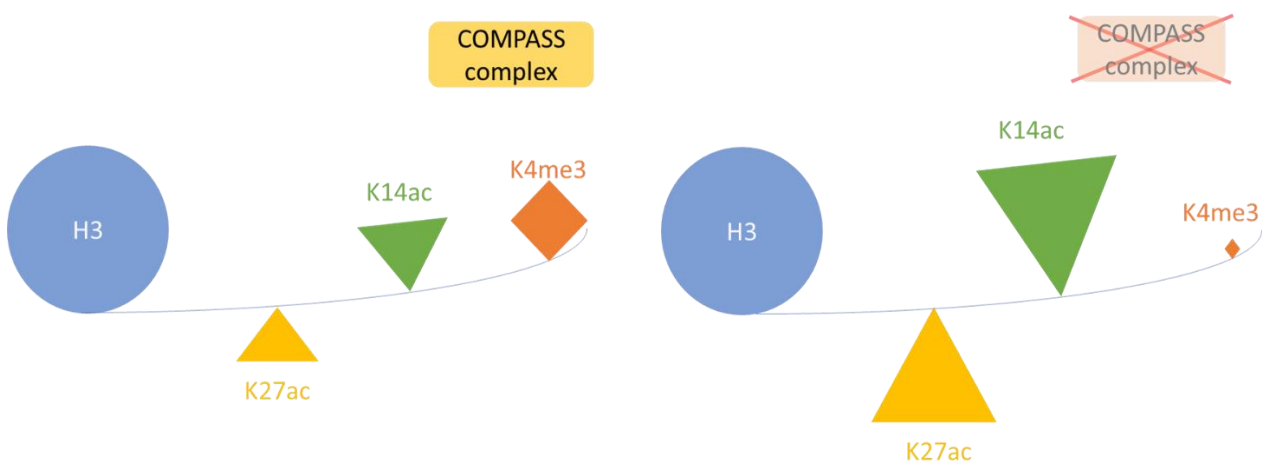


Figure 7.1 – In the absence of COMPASS function, H3K4me3 levels fall and H3K14ac and H3K27ac levels rise. The illustration presents two conditions, an H3 N-terminal tail with an active COMPASS complex, and

one without. Levels of three key histone modifications from chapter 3, H3K4me3, H3K14ac and H3K27ac, are represented by the relative size (between conditions) of the coloured shapes attached to the H3 N-terminal tail.

How this increase in acetylation happens, mechanistically, is unknown. The full extent of ‘cross-talk’ between different histone modifiers is still being elucidated. Recently, CFP-1 was shown to physically interact with the SIN-3 HDAC complex comprising SIN-3 and HDA-1 (Beurton et al., 2019). Moreover, a potential genetic interaction between *cfp-1* and *sin-3* was reported in Chapter 4, which describes experiments in which *cfp-1* mutants were subjected to RNAi knockdown of HDAC subunits *hda-3*, *hda-1* and *sin-3*. Expression of the *unc-119::GFP* reporter in the germlines of *cfp-1* mutants and WT controls was recorded, as a readout for gene derepression. The proportion of *unc-119::GFP* expressing germlines was significantly increased in germlines of *cfp-1* mutants subjected to *hda-3* RNAi, as compared to negative control RNAi, consistently across all experiments. For *sin-3* and *hda-1* RNAi, the effect on *cfp-1* mutants was the same but not consistently significant, in the second replicate. Overall, the experiments support a genetic interaction between *cfp-1* and the tested HDAC subunits, RNAi knockdown of which is able to enhance the germline gene derepression observed in *cfp-1* mutants.

Notably, *hda-1* RNAi did not enhance *unc-119::GFP* expression in *set-2(bn129)* mutants here, suggesting that CFP-1 is interacting with HDA-1, and possibly SIN-3 and HDA-3, independently of COMPASS. This is supported by the LC-MS/MS data showing that *cfp-1* mixed embryo-derived histones have H3K23ac levels elevated, albeit less significantly than K14 and K27ac, whereas *set-2* mutants do not. Testing of *set-2* mutants with *sin-3* and *hda-3* RNAi, and of both *cfp-1* and *set-2* mutants with RNAi of other HDAC genes, is needed to confirm these interactions and/or lack of them.

The larger P-values obtained for the probability of a genetic interaction between *cfp-1* and *sin-3/hda-1* in the second replicate, and consequent non-significant results, may be partly explained by the increased *unc-119::GFP* expression observed in the germlines of *cfp-1* negative control RNAi worms as the experiment progressed. When the generation number since outcrossing of the *cfp-1* controls was taken into account, a clear trend of increased *unc-119::GFP* expression with increased generation number was shown. Such a trend may be thought of as a ‘transgenerational worsening’, and fits with a model proposed previously, in which the acquisition of somatic cell fate, as indicated by germline *unc-119::GFP* expression, may require a threshold effect of chromatin changes accumulated over time, whose frequency may increase over generations (Robert et al., 2014).

Not just for *unc-119::GFP* expression was this transgenerational phenotypic worsening observed, but also with the slower developmental progression of late vs early generation *set-2(bn129)* mutants, as discussed in chapter 5. Abnormal development has also been observed as an effect of the loss of H3K4me3 in mammalian systems (Siklenka et al., 2015).

Moreover, this chapter demonstrated a heritable component to this transgenerational worsening. WT descendants of late-generation *set-2(bn129)* mutants inherited a more pronounced and longer lasting (over more generations) longevity phenotype compared to the WT descendants of early generation *set-2(bn129)* mutants. The same effect has been observed in *wdr-5* mutants of the COMPASS complex (Dr Teresa Lee, Katz lab, personal communication).

7.1.1 Model

Taken together, these results can be summarized in the following model (Figure 7.2). Acquired chromatin changes accumulate in *C. elegans set-2* mutants over generations of self-fertilisation. This accumulation manifests itself in worsening mutant phenotypes, similar to the Mrt phenotype noted among many chromatin modifier mutants, the more generations that have passed since outcrossing. These phenotypes, for example germline gene derepression (Robert et al., 2014), slow development, increased longevity (Greer et al., 2010), are observed in early generation *set-2* mutants, but are more penetrant at later generations. In the case of germline derepression, this effect was also observed for *cfp-1* mutants, but speed of development and longevity were not tested in *cfp-1* mutants. Therefore, the model is primarily based on *set-2* mutants but is likely to apply to other COMPASS mutants, given the essential roles of other subunits of the complex (Li and Kelly, 2011, Xiao et al., 2011) and should be tested in them.

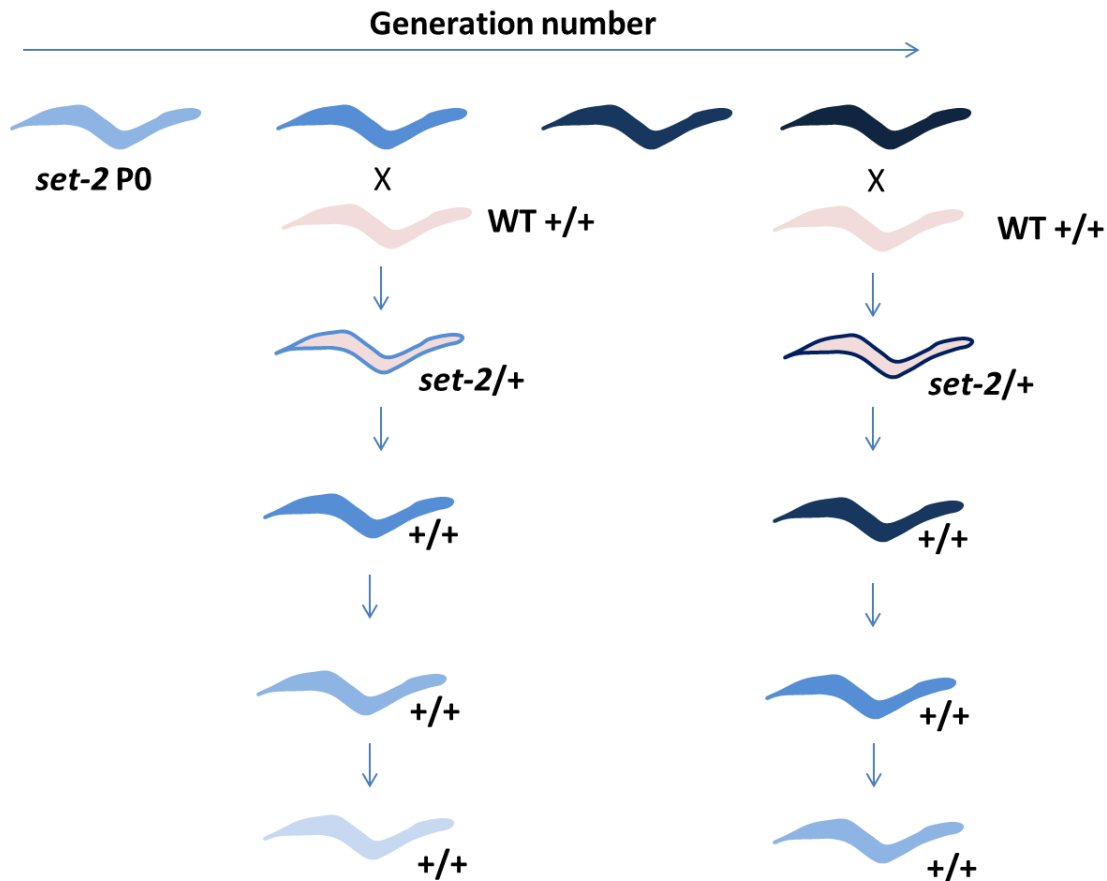


Figure 7.2 – A model for the heritability of transgenerationally accumulated chromatin changes. WT worms are represented in pink, *set-2* mutants in blue. Both colours are combined in the heterozygote, which has a pink interior and a blue outline. Accumulating chromatin changes are represented by darkening blue, which happens over generations, as illustrated by the arrow. The shade of blue is also employed to illustrate how after crossing a *set-2* hermaphrodite with a WT male, the extent of accumulated chromatin changes in the *set-2* parent determines how long these changes persist in the WT descendants, contributing to the transgenerational epigenetic inheritance of the parental mutant phenotype.

The second part of the model is based solely on experiments assessing the longevity of WT descendants of *set-2(bn129)* mutants, and none of the other phenotypes mentioned above. In this part of the model, the accumulated chromatin changes are heritable, which in WT descendants manifests as a persistence of the parental mutant phenotype, in this case increased longevity. The more generations the mutant parent has been maintained by self-fertilisation, the more pronounced the mutant phenotype in WT descendants, and the more generations of WT descendants the mutant phenotype persists for.

7.1.2 The role of histone modifications in the model

A major histone modification change noted in *cfp-1* and *set-2* mutants, besides the loss of H3K4me3, was a gain in acetylation of lysines 14 and 27. Further LC-MS/MS and/or antibody-based approaches could be employed to test whether acetylation, on these lysines and possibly others, accumulates over generations (Figure 7.3). More acetylation modifications contribute to a more open chromatin structure, as discussed in 1.1.4.2, which could explain, at least in part, why gene derepression is worsened in the germlines of later generation *cfp-1* and *set-2* mutants (Robert et al., 2014).

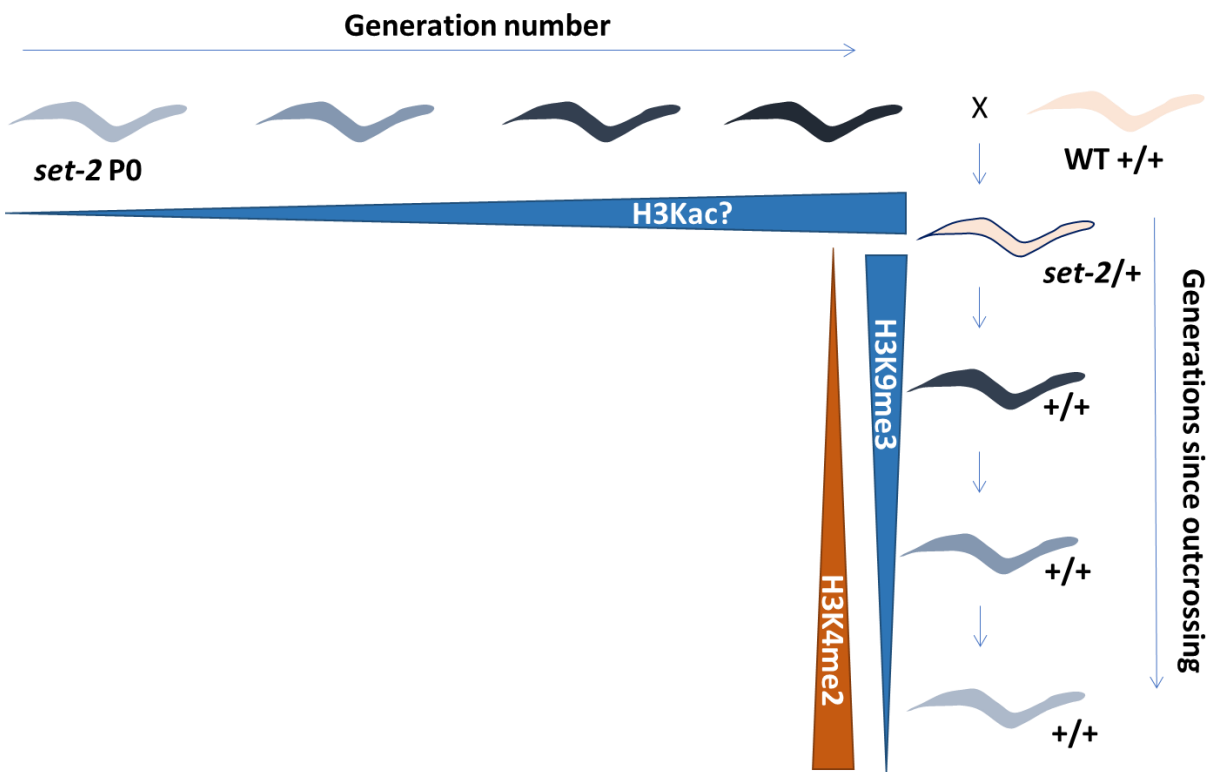


Figure 7.3 – The role of histone modifications in the model for heritability of transgenerationally accumulated chromatin changes. The accumulating chromatin changes and the *set-2* mutant/WT distinction are represented identically to Figure 7.2, but the putative increase and decrease of different histone modifications over generations is represented by expanding or shrinking bars.

However, when germline-derived histones from WT descendants of *set-2(bn129)* mutants were compared with WT controls, lysine acetylation was not notably different, refuting the suggestion that lysine acetylation could be the accumulating chromatin change responsible for the inheritance of extended longevity in F3 and F4 WT descendants. Histone modifications that were altered in these descendants compared with WT controls were H3K4me2, which was reduced in F3 and F4 descendants, and H3K9me3, which was increased (Figure 7.3). The H3K9me3 result was exciting because this modification appears to be necessary for the

inheritance of gene expression memory (Klosin et al., 2017) and plays a role in TEI via small RNAs (Guerin et al., 2014).

Owing to the difficulty of obtaining sufficient germlines by manual dissection to provide enough histones for LC-MS/MS, reproducing these results was challenging. An alternative method, immunofluorescence (IF) microscopy, was used to quantify H3K4me2 and H3K9me3 levels in WT and mutant descendants of *set-2(bn129)* mutants. The data obtained was very variable and the experiment would need repeating to be confident in making any conclusions, but overall the IF data supported the LC-MS/MS results for H3K9me3, in that it did appear elevated in F3 and F4 WT descendants. No reduction in H3K4me2 in F3 and F4 WT descendants was observed by IF microscopy, however. To be convinced of this trend, both LC-MS/MS and IF microscopy experiments would need to be repeated.

7.2 Further work

This thesis has highlighted the need for future experiments with mutants of the COMPASS complex, and indeed other chromatin modifiers, to control for generation number of the mutants post-outcrossing with WT. As shown in chapters 4 and 5, generation number impacts severity of *cfp-1* and *set-2* mutant phenotypes, which may confound results when not controlled for.

Practically, controlling for generation number could be achieved by freezing *C. elegans* stocks at set intervals post-outcrossing, enabling the building of a 'generational library' of different mutant strains, to be thawed for use in experiments when required.

Chapter 3 demonstrated the utility and success of LC-MS/MS with *C. elegans* mixed embryo-derived histones. The method could be used again with other histone-modifier mutants, to confirm whether particular deletion alleles do indeed completely inactivate the complex.

LC-MS/MS experiments with germline-derived histones, however, highlighted the need for an alternative method for obtaining germline nuclei. Only in one of three replicates did manual dissection provide sufficient histones for complete detection of all major proteoforms of the TKQTAR (H3K4-containing) and KSTGGKAPR (H3K9 & K14-containing) peptides, which is needed for confidence in assigning relative abundances. Absence of one proteoform will skew relative abundances of those detected.

One potential way of obtaining more germline nuclei would be to break up YA worms with two strokes in a metal wheaton tissue grinder, which would be enough force to rupture the gonad and release primarily pachytene nuclei, without much somatic nuclei contamination (Prof. Enrique Martinez-Perez, personal

communication). Nuclei could then be collected from the supernatant after centrifuging the ground sample, as described in 2.3.1.1.5. The drawback of this method compared to manual dissection and collection would be the greater probability of contamination, but the enhanced retrieval of histones might be an acceptable trade-off. It would be informative to repeat the germline LC-MS/MS experiments using this method, to see if the H3K4me2 reduction and H3K9me3 increase in F3 and F4 WT descendants could be reproduced.

Chapter 8. Supplementary data

8.1 Germline mass spectrometry data for *set-2(bn129)* descendants

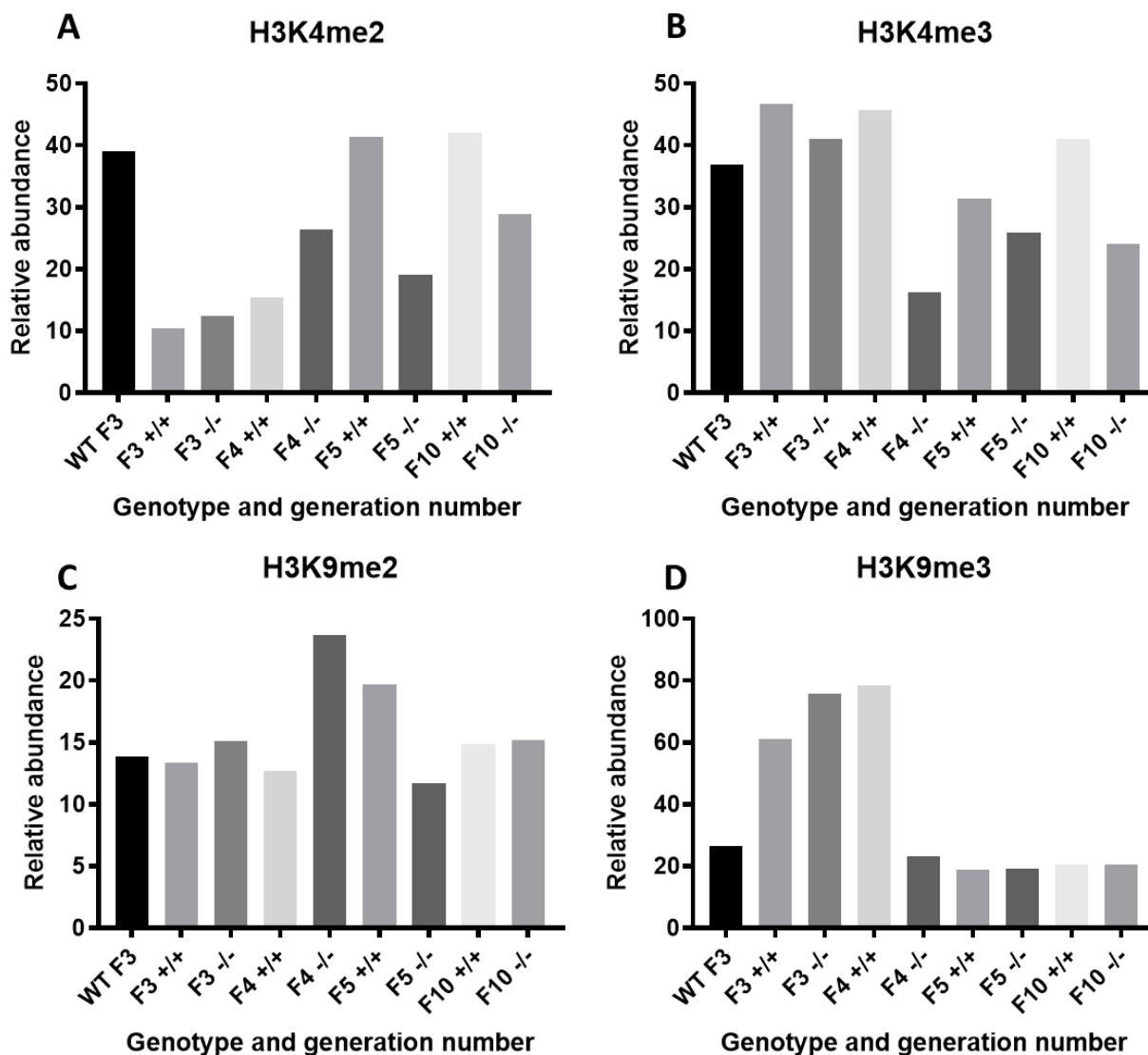


Figure 8.1 – Relative abundance of di- and trimethylated H3K4 and H3K9 in WT and mutant descendants of *set-2(bn129)*, and WT controls. The data presented corresponds to the fullest dataset out of three replicates performed, in which all modified proteoforms were detected. The height of the bars, or relative abundance, represents the percentage intensity of the modified peptide in question out of the total intensity detected for all proteoforms of the peptide.

Chapter 9. Bibliography

- AICHLER, M. & WALCH, A. 2015. MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice. *Laboratory Investigation*, 95, 422-431.
- ANDRES, M. E., BURGER, C., PERAL-RUBIO, M. J., BATTAGLIOLI, E., ANDERSON, M. E., GRIMES, J., DALLMAN, J., BALLAS, N. & MANDEL, G. 1999. CoREST: A functional corepressor required for regulation of neural-specific gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 9873-9878.
- ARDEHALI, M. B., MEI, A., ZOBECK, K. L., CARON, M., LIS, J. T. & KUSCH, T. 2011. Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. *Embo Journal*, 30, 2817-2828.
- BANNISTER, A. J., SCHNEIDER, R. & KOUZARIDES, T. 2002. Histone methylation: Dynamic or static? *Cell*, 109, 801-806.
- BARSKI, A., CUDDAPAH, S., CUI, K. R., ROH, T. Y., SCHONES, D. E., WANG, Z. B., WEI, G., CHEPELEV, I. & ZHAO, K. J. 2007. High-resolution profiling of histone methylations in the human genome. *Cell*, 129, 823-837.
- BARSTEAD, R., MOULDER, G., COBB, B., FRAZEE, S., HENTHORN, D., HOLMES, J., JEREBIE, D., LANDSDALE, M., OSBORN, J., PRITCHETT, C., ROBERTSON, J., RUMMAGE, J., STOKES, E., VISHWANATHAN, M., MITANI, S., GENGYO-ANDO, K., FUNATSU, O., HORI, S., IMAE, R., KAGE-NAKADAI, E., KOBUNA, H., MACHIYAMA, E., MOTOHASHI, T., OTORI, M., SUEHIRO, Y., YOSHINA, S., MOERMAN, D., EDGLEY, M., ADAIR, R., ALLAN, B. J., AU, V., CHAUDHRY, I., CHEUNG, R., DADIVAS, O., ENG, S., FERNANDO, L., FISHER, A., FLIBOTTE, S., GILCHRIST, E., HAY, A., HUANG, P., HUNT, R. W., KWITKOWSKI, C., LAU, J., LEE, N., LIU, L., LORCH, A., LUCK, C., MAYDAN, J., MCKAY, S., MILLER, A., MULLEN, G., NAVAROLI, C., NEIL, S., HUNT-NEWBURY, R., PARTRIDGE, M., PERKINS, J., RANKIN, A., RAYMANT, G., REZANIA, N., ROGULA, A., SHEN, B., STEGEMAN, G., TARDIF, A., TAYLOR, J., VEIGA, M., WANG, T., ZAPF, R. & CONSORTI, C. E. D. M. 2012. Large-Scale Screening for Targeted Knockouts in the Caenorhabditis elegans Genome. *G3-Genes Genomes Genetics*, 2, 1415-1425.
- BENDER, A. M., KIRIENKO, N. V., OLSON, S. K., ESKO, J. D. & FAY, D. S. 2007. lin-35/Rb and the CoREST ortholog spr-1 coordinately regulate vulval morphogenesis and gonad development in C-elegans. *Developmental Biology*, 302, 448-462.
- BEURTON, F., STEMPOR, P., CARON, M., APPERT, A., DONG, Y., CHEN, R. A. J., CLUET, D., COUTE, Y., HERBETTE, M., HUANG, N., POLVECHE, H., SPICHTY, M., BEDET, C., AHRINGER, J. & PALLADINO, F. 2019. Physical and functional interaction between SET1/COMPASS complex component CFP-1 and a Sin3S HDAC complex in C. elegans. *Nucleic Acids Research*, 47, 11164-11180.
- BIRD, A. P. 1980. DNA METHYLATION AND THE FREQUENCY OF CPG IN ANIMAL DNA. *Nucleic Acids Research*, 8, 1499-1504.
- BIRD, A. P. 1986. CPG-RICH ISLANDS AND THE FUNCTION OF DNA METHYLATION. *Nature*, 321, 209-213.
- BLACK, J. C., VAN RECHEM, C. & WHETSTINE, J. R. 2012. Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Molecular Cell*, 48, 491-507.
- BOESL, U. 2017. TIME-OF-FLIGHT MASS SPECTROMETRY: INTRODUCTION TO THE BASICS. *Mass Spectrometry Reviews*, 36, 86-109.
- BRENNER, S. 1974. GENETICS OF CAENORHABDITIS-ELEGANS. *Genetics*, 77, 71-94.
- CANTERBURY, J. D., MERRIHEW, G. E., MACCOSS, M. J., GOODLETT, D. R. & SHAFFER, S. A. 2014. Comparison of Data Acquisition Strategies on Quadrupole Ion Trap Instrumentation for Shotgun Proteomics. *Journal of the American Society for Mass Spectrometry*, 25, 2048-2059.
- CARLONE, D. L., LEE, J. H., YOUNG, S. R. L., DOBROTA, E., BUTLER, J. S., RUIZ, J. & SKALNIK, D. G. 2005. Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein. *Molecular and Cellular Biology*, 25, 4881-4891.
- CARLONE, D. L. & SKALNIK, D. G. 2001. CpG binding protein is crucial for early embryonic development. *Molecular and Cellular Biology*, 21, 7601-7606.

- CHEN, R. A. J., STEMPOR, P., DOWN, T. A., ZEISER, E., FEUER, S. K. & AHRINGER, J. 2014. Extreme HOT regions are CpG-dense promoters in *C. elegans* and humans. *Genome Research*, 24, 1138-1146.
- CHOY, S. W., WONG, Y. M., HO, S. H. & CHOW, K. L. 2007. *C. elegans* SIN-3 and its associated HDAC corepressor complex act as mediators of male sensory ray development. *Biochemical and Biophysical Research Communications*, 358, 802-807.
- CIOSK, R., DEPALMA, M. & PRIESS, J. R. 2006. Translational regulators maintain totipotency in the *Caenorhabditis elegans* germline. *Science*, 311, 851-853.
- CLAYTON, A. L., ROSE, S., BARRATT, M. J. & MAHADEVAN, L. C. 2000. Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation. *Embo Journal*, 19, 3714-3726.
- CLOUAIRE, T., WEBB, S., SKENE, P., ILLINGWORTH, R., KERR, A., ANDREWS, R., LEE, J. H., SKALNIK, D. & BIRD, A. 2012. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes & Development*, 26, 1714-1728.
- COOPER, D. N., TAGGART, M. H. & BIRD, A. P. 1983. UNMETHYLATED DOMAINS IN VERTEBRATE DNA. *Nucleic Acids Research*, 11, 647-658.
- COUSTHAM, V., BEDET, C., MONIER, K., SCHOTT, S., KARALI, M. & PALLADINO, F. 2006. The *C. elegans* HP1 homologue HPL-2 and the LIN-13 zinc finger protein form a complex implicated in vulval development. *Developmental Biology*, 297, 308-322.
- COUTEAU, F., GUERRY, F., MULLER, F. & PALLADINO, F. 2002. A heterochromatin protein 1 homologue in *Caenorhabditis elegans* acts in germline and vulval development. *Embo Reports*, 3, 235-241.
- DILLON, S. C., ZHANG, X., TRIEVEL, R. C. & CHENG, X. D. 2005. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biology*, 6.
- DOERR, A. 2015. DIA mass spectrometry. *Nature Methods*, 12, 35-35.
- EISSENBERG, J. C. & SHILATIFARD, A. 2010. Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Developmental Biology*, 339, 240-249.
- FILION, G. J., VAN BEMMEL, J. G., BRAUNSCHWEIG, U., TALHOUT, W., KIND, J., WARD, L. D., BRUGMAN, W., DE CASTRO, I. J., KERKHOVEN, R. M., BUSSEMAKER, H. J. & VAN STEENSEL, B. 2010. Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in *Drosophila* Cells. *Cell*, 143, 212-224.
- FISHER, K., SOUTHALL, S. M., WILSON, J. R. & POULIN, G. B. 2010. Methylation and demethylation activities of a *C. elegans* MLL-like complex attenuate RAS signalling. *Developmental Biology*, 341, 142-153.
- GARCIA, B. A., MOLLAH, S., UEBERHEIDE, B. M., BUSBY, S. A., MURATORE, T. L., SHABANOWITZ, J. & HUNT, D. F. 2007. Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature Protocols*, 2, 933-938.
- GAYDOS, L. J., WANG, W. C. & STROME, S. 2014. H3K27me and PRC2 transmit a memory of repression across generations and during development. *Science*, 345, 1515-1518.
- GELATO, K. A. & FISCHLE, W. 2008. Role of histone modifications in defining chromatin structure and function. *Biological Chemistry*, 389, 353-363.
- GILL, G. 2004. SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? *Genes & Development*, 18, 2046-2059.
- GREER, E. L., MAURES, T. J., HAUSWIRTH, A. G., GREEN, E. M., LEEMAN, D. S., MARO, G. S., HAN, S., BANKO, M. R., GOZANI, O. & BRUNET, A. 2010. Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in *C. elegans*. *Nature*, 466, 383-U137.
- GREER, E. L., MAURES, T. J., UCAR, D., HAUSWIRTH, A. G., MANCINI, E., LIM, J. P., BENAYOUN, B. A., SHI, Y. & BRUNET, A. 2011. Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans*. *Nature*, 479, 365-U204.
- GREGORETTI, I. V., LEE, Y. M. & GOODSON, H. V. 2004. Molecular evolution of the histone deacetylase family: Functional implications of phylogenetic analysis. *Journal of Molecular Biology*, 338, 17-31.
- GUERIN, T. M., PALLADINO, F. & ROBERT, V. J. 2014. Transgenerational functions of small RNA pathways in controlling gene expression in *C. elegans*. *Epigenetics*, 9, 37-44.

- HEARD, E. & MARTIENSSEN, R. A. 2014. Transgenerational Epigenetic Inheritance: Myths and Mechanisms. *Cell*, 157, 95-109.
- HENIKOFF, S. & GREALLY, J. M. 2016. Epigenetics, cellular memory and gene regulation. *Current Biology*, 26, R644-R648.
- HON, G. C., HAWKINS, R. D. & REN, B. 2009. Predictive chromatin signatures in the mammalian genome. *Human Molecular Genetics*, 18, R195-R201.
- HU, Q. Z., NOLL, R. J., LI, H. Y., MAKAROV, A., HARDMAN, M. & COOKS, R. G. 2005. The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry*, 40, 430-443.
- HUANG, S. 2002. Histone methyltransferases, diet nutrients and tumour suppressors. *Nature Reviews Cancer*, 2, 469-476.
- HUBBERT, C., GUARDIOLA, A., SHAO, R., KAWAGUCHI, Y., ITO, A., NIXON, A., YOSHIDA, M., WANG, X. F. & YAO, T. P. 2002. HDAC6 is a microtubule-associated deacetylase. *Nature*, 417, 455-458.
- ITO, A., KAWAGUCHI, Y., LAI, C. H., KOVACS, J. J., HIGASHIMOTO, Y., APPELLA, E. & YAO, T. P. 2002. MDM2-HDAC1-mediated deacetylation of p53 is required for its degradation. *Embo Journal*, 21, 6236-6245.
- JANSSEN, K. A., SIDOLI, S. & GARCIA, B. A. 2017. Recent Achievements in Characterizing the Histone Code and Approaches to Integrating Epigenomics and Systems Biology. *Proteomics in Biology, Pt B*, 586, 359-378.
- JENUWEIN, T. & ALLIS, C. D. 2001. Translating the histone code. *Science*, 293, 1074-1080.
- KABASENCHE, W. P. & SKINNER, M. K. 2014. DDT, epigenetic harm, and transgenerational environmental justice. *Environmental Health*, 13.
- KADOSH, D. & STRUHL, K. 1997. Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. *Cell*, 89, 365-371.
- KALINAVA, N., NI, J. Z., GAJIC, Z., KIM, M., USHAKOV, H. & GU, S. G. 2018. C-elegans Heterochromatin Factor SET-32 Plays an Essential Role in Transgenerational Establishment of Nuclear RNAi-Mediated Epigenetic Silencing. *Cell Reports*, 25, 2273-+.
- KAMATH, R. S., FRASER, A. G., DONG, Y., POULIN, G., DURBIN, R., GOTTA, M., KANAPIN, A., LE BOT, N., MORENO, S., SOHRMANN, M., WELCHMAN, D. P., ZIPPERLEN, P. & AHRINGER, J. 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*, 421, 231-237.
- KELLY, W. G. & FIRE, A. 1998. Chromatin silencing and the maintenance of a functional germline in *Caenorhabditis elegans*. *Development*, 125, 2451-2456.
- KLOSIN, A., CASAS, E., HIDALGO-CARCEDO, C., VAVOURI, T. & LEHNER, B. 2017. Transgenerational transmission of environmental information in C-elegans. *Science*, 356, 316-319.
- KORNBERG, R. D. 1977. STRUCTURE OF CHROMATIN. *Annual Review of Biochemistry*, 46, 931-954.
- KORNBERG, R. D. & LORCH, Y. 1992. CHROMATIN STRUCTURE AND TRANSCRIPTION. *Annual Review of Cell Biology*, 8, 563-687.
- LAHERTY, C. D., YANG, W. M., SUN, J. M., DAVIE, J. R., SETO, E. & EISENMAN, R. N. 1997. Histone deacetylases associated with the mSin3 corepressor mediate Mad transcriptional repression. *Cell*, 89, 349-356.
- LEE, J. H. & SKALNIK, D. G. 2005. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian set1 histone H3-Lys(4) methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *Journal of Biological Chemistry*, 280, 41725-41731.
- LEE, J. H., VOO, K. S. & SKALNIK, D. G. 2001. Identification and characterization of the DNA binding domain of CpG-binding protein. *Journal of Biological Chemistry*, 276, 44669-44676.
- LEI, H., OH, S. P., OKANO, M., JUTTERMANN, R., GOSS, K. A., JAENISCH, R. & LI, E. 1996. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development*, 122, 3195-3205.
- LI, T. G. & KELLY, W. G. 2011. A Role for Set1/MLL-Related Components in Epigenetic Regulation of the *Caenorhabditis elegans* Germ Line. *Plos Genetics*, 7.
- LIN, S., WEIN, S., GONZALES-COPE, M., OTTE, G. L., YUAN, Z. F., AFJEHI-SADAT, L., MAILE, T., BERGER, S. L., RUSH, J., LILL, J. R., ARNOTT, D. & GARCIA, B. A. 2014. Stable-isotope-labeled Histone Peptide Library

- for Histone Post-translational Modification and Variant Quantification by Mass Spectrometry. *Molecular & Cellular Proteomics*, 13, 2450-2466.
- MACLEAN, B., TOMAZELA, D. M., SHULMAN, N., CHAMBERS, M., FINNEY, G. L., FREWEN, B., KERN, R., TABB, D. L., LIEBLER, D. C. & MACCOSS, M. J. 2010. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, 26, 966-968.
- MAILE, T. M., IZRAEL-TOMASEVIC, A., CHEUNG, T., GULER, G. D., TINDELL, C., MASSELOT, A., LIANG, J., ZHAO, F., TROJER, P., CLASSON, M. & ARNOTT, D. 2015. Mass Spectrometric Quantification of Histone Post-translational Modifications by a Hybrid Chemical Labeling Method. *Molecular & Cellular Proteomics*, 14, 1148-1158.
- METZLER-GUILLEMAIN, C., DEPETRIS, D., LUCIANI, J. J., MIGNON-RAVIX, C., MITCHELL, M. J. & MATTEI, M. G. 2008. In human pachytene spermatocytes, SUMO protein is restricted to the constitutive heterochromatin. *Chromosome Research*, 16, 761-782.
- MILLER, P. E. & DENTON, M. B. 1986. THE QUADRUPOLE MASS FILTER - BASIC OPERATING CONCEPTS. *Journal of Chemical Education*, 63, 617-622.
- MILLER, T., KROGAN, N. J., DOVER, J., ERDJUMENT-BROMAGE, H., TEMPST, P., JOHNSTON, M., GREENBLATT, J. F. & SHILATIFARD, A. 2001. COMPASS: A complex of proteins associated with a trithorax-related SET domain protein. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 12902-12907.
- MINSHULL, T. C., COLE, J., DOCKRELL, D. H., READ, R. C. & DICKMAN, M. J. 2016. Analysis of histone post translational modifications in primary monocyte derived macrophages using reverse phase x reverse phase chromatography in conjunction with porous graphitic carbon stationary phase. *Journal of Chromatography A*, 1453, 43-53.
- MOHAN, M., HERZ, H. M., SMITH, E. R., ZHANG, Y., JACKSON, J., WASHBURN, M. P., FLORENS, L., EISENBERG, J. C. & SHILATIFARD, A. 2011. The COMPASS Family of H3K4 Methylases in Drosophila. *Molecular and Cellular Biology*, 31, 4310-4318.
- MUTLU, B., CHEN, H. M., MORESCO, J. J., ORELO, B. D., YANG, B., GASPAR, J. M., KEPPLER-ROSS, S., YATES, J. R., HALL, D. H., MAINE, E. M. & MANGO, S. E. 2018. Regulated nuclear accumulation of a histone methyltransferase times the onset of heterochromatin formation in C. elegans embryos. *Science Advances*, 4.
- NAKAMURA, T., MORI, T., TADA, S., KRAJEWSKI, W., ROZOVSKAIA, T., WASELL, R., DUBOIS, G., MAZO, A., CROCE, C. M. & CANAANI, E. 2002. ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Molecular Cell*, 10, 1119-1128.
- NELSON, D. A., PERRY, M., SEALY, L. & CHALKLEY, R. 1978. DNASE-I PREFERENTIALLY DIGESTS CHROMATIN CONTAINING HYPERACETYLATED HISTONES. *Biochemical and Biophysical Research Communications*, 82, 1346-1353.
- NG, H. H., ROBERT, F., YOUNG, R. A. & STRUHL, K. 2003. Targeted recruitment of set1 histone methylase by elongating pol II provides a localized mark and memory of recent transcriptional activity. *Molecular Cell*, 11, 709-719.
- NI, J. Z., KALINAVA, N., CHEN, E., HUANG, A., TRINH, T. & GU, S. G. 2016. A transgenerational role of the germline nuclear RNAi pathway in repressing heat stress-induced transcriptional activation in C. elegans. *Epigenetics & Chromatin*, 9.
- OOI, S. L., HENIKOFF, J. G. & HENIKOFF, S. 2010. A native chromatin purification system for epigenomic profiling in Caenorhabditis elegans. *Nucleic Acids Research*, 38, 14.
- PASSARGE, E. 1979. HEITZ, EMIL AND THE CONCEPT OF HETEROCHROMATIN - LONGITUDINAL CHROMOSOME DIFFERENTIATION WAS RECOGNIZED 50 YEARS AGO. *American Journal of Human Genetics*, 31, 106-115.
- PERRY, R. H., COOKS, R. G. & NOLL, R. J. 2008. ORBITRAP MASS SPECTROMETRY: INSTRUMENTATION, ION MOTION AND APPLICATIONS. *Mass Spectrometry Reviews*, 27, 661-699.
- POKHREL, B., CHEN, Y. & BIRO, J. J. 2019. CFP-1 interacts with HDAC1/2 complexes in C. elegans development. *Febs Journal*, 286, 2490-2504.

- REA, S., EISENHABER, F., O'CARROLL, N., STRAHL, B. D., SUN, Z. W., SCHMID, M., OPRAVIL, S., MECHTLER, K., PONTING, C. P., ALLIS, C. D. & JENUWEIN, T. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406, 593-599.
- RICHMOND, T. J. & DAVEY, C. A. 2003. The structure of DNA in the nucleosome core. *Nature*, 423, 145-150.
- RIDSDALE, J. A., HENDZEL, M. J., DELCUVE, G. P. & DAVIE, J. R. 1990. HISTONE ACETYLATION ALTERS THE CAPACITY OF THE H1 HISTONES TO CONDENSE TRANSCRIPTIONALLY ACTIVE COMPETENT CHROMATIN. *Journal of Biological Chemistry*, 265, 5150-5156.
- ROBERT, V. J., MERCIER, M. G., BEDET, C., JANCZARSKI, S., MERLET, J., GARVIS, S., CIOSK, R. & PALLADINO, F. 2014. The SET-2/SET1 Histone H3K4 Methyltransferase Maintains Pluripotency in the *Caenorhabditis elegans* Germline. *Cell Reports*, 9, 443-450.
- ROGUEV, A., SCHAFT, D., SHEVCHENKO, A., AASLAND, R., SHEVEHENKO, A. & STEWART, A. F. 2003. High conservation of the Set1/Rad6 axis of histone 3 lysine 4 methylation in budding and fission yeasts. *Journal of Biological Chemistry*, 278, 8487-8493.
- ROGUEV, A., SCHAFT, D., SHEVCHENKO, A., PIJNAPPEL, W., WILM, M., AASLAND, R. & STEWART, A. F. 2001. The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *Embo Journal*, 20, 7137-7148.
- ROTHSTEIN, M. A., HARRELL, H. L. & MARCHANT, G. E. 2017. Transgenerational epigenetics and environmental justice. *Environmental Epigenetics*, 3.
- SANCHEZ, R. & ZHOU, M. M. 2011. The PHD finger: a versatile epigenome reader. *Trends in Biochemical Sciences*, 36, 364-372.
- SARG, B., FASERL, K., KREMSER, L., HALFINGER, B., SEBASTIANO, R. & LINDNER, H. H. 2013. Comparing and Combining Capillary Electrophoresis Electrospray Ionization Mass Spectrometry and Nano-Liquid Chromatography Electrospray Ionization Mass Spectrometry for the Characterization of Post-translationally Modified Histones. *Molecular & Cellular Proteomics*, 12, 2640-2656.
- SCHIAVONE, D., GUILBAUD, G., MURAT, P., PAPADOPOULOU, C., SARKIES, P., PRIOLEAU, M. N., BALASUBRAMANIAN, S. & SALE, J. E. 2014. Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *Embo Journal*, 33, 2507-2520.
- SCHNEIDER, J., WOOD, A., LEE, J. S., SCHUSTER, R., DUEKER, J., MAGUIRE, C., SWANSON, S. K., FLORENS, L., WASHBURN, M. P. & SHILATIFARD, A. 2005. Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. *Molecular Cell*, 19, 849-856.
- SCHOTT, S., COUSTHAM, V., SIMONET, T., BEDET, C. & PALLADINO, F. 2006. Unique and redundant functions of *C.elegans* HP1 proteins in post-embryonic development. *Developmental Biology*, 298, 176-187.
- SCHURENBERG, M., DREISEWERD, K. & HILLENKAMP, F. 1999. Laser desorption/ionization mass spectrometry of peptides and proteins with particle suspension matrixes. *Analytical Chemistry*, 71, 221-229.
- SHI, Y. & MELLO, C. 1998. A CBP/p300 homolog specifies multiple differentiation pathways in *Caenorhabditis elegans*. *Genes & Development*, 12, 943-955.
- SHI, Y. & WHETSTINE, J. R. 2007. Dynamic regulation of histone lysine methylation by demethylases. *Molecular Cell*, 25, 1-14.
- SHI, Y. J., LAN, F., MATSON, C., MULLIGAN, P., WHETSTINE, J. R., COLE, P. A., CASERO, R. A. & SHI, Y. 2004. Histone demethylation mediated by the nuclear arnine oxidase homolog LSD1. *Cell*, 119, 941-953.
- SHIBUE, M., MANT, C. T. & HODGES, R. S. 2005. Effect of anionic ion-pairing reagent hydrophobicity on selectivity of peptide separations by reversed-phase liquid chromatography. *Journal of Chromatography A*, 1080, 68-75.
- SIDOLI, S., LIN, S., XIONG, L., BHANU, N. V., KARCH, K. R., JOHANSEN, E., HUNTER, C., MOLLAH, S. & GARCIA, B. A. 2015. Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH) Analysis for Characterization and Quantification of Histone Post-translational Modifications. *Molecular & Cellular Proteomics*, 14, 2420-2428.

- SIDOLI, S., SCHWAMMLE, V., RUMINOWICZ, C., HANSEN, T. A., WU, X. D., HELIN, K. & JENSEN, O. N. 2014. Middle-down hybrid chromatography/tandem mass spectrometry workflow for characterization of combinatorial post-translational modifications in histones. *Proteomics*, 14, 2200-2211.
- SIDOLI, S., VANDAMME, J., SALCINI, A. E. & JENSEN, O. N. 2016. Dynamic changes of histone H3 marks during *Caenorhabditis elegans* lifecycle revealed by middle-down proteomics. *Proteomics*, 16, 459-464.
- SIKLENKA, K., ERKEK, S., GODMANN, M., LAMBROT, R., MCGRAW, S., LAFLEUR, C., COHEN, T., XIA, J. G., SUDERMAN, M., HALLETT, M., TRASLER, J., PETERS, A. & KIMMINS, S. 2015. Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science*, 350, 13.
- SIMONET, T., DULERMO, R., SCHOTT, S. & PALLADINO, F. 2007. Antagonistic functions of SET-2/SET1 and HPL/HP1 proteins in *C. elegans* development. *Developmental Biology*, 312, 367-383.
- SOLARI, F. & AHRINGER, J. 2000. NURD-complex genes antagonise Ras-induced vulval development in *Caenorhabditis elegans*. *Current Biology*, 10, 223-226.
- SOLARI, F., BATEMAN, A. & AHRINGER, J. 1999. The *Caenorhabditis elegans* genes *egl-27* and *egr-1* are similar to MTA1, a member of a chromatin regulatory complex, and are redundantly required for embryonic patterning. *Development*, 126, 2483-2494.
- STRAHL, B. D. & ALLIS, C. D. 2000. The language of covalent histone modifications. *Nature*, 403, 41-45.
- STROME, S. & WOOD, W. B. 1983. GENERATION OF ASYMMETRY AND SEGREGATION OF GERM-LINE GRANULES IN EARLY C-ELEGANS EMBRYOS. *Cell*, 35, 15-25.
- SUGANUMA, T. & WORKMAN, J. L. 2011. Signals and Combinatorial Functions of Histone Modifications. In: KORNBERG, R. D., RAETZ, C. R. H., ROTHMAN, J. E. & THORNER, J. W. (eds.) *Annual Review of Biochemistry*, Vol 80. Palo Alto: Annual Reviews.
- TYKOCINSKI, M. L. & MAX, E. E. 1984. CG DINUCLEOTIDE CLUSTERS IN MHC GENES AND IN 5' DEMETHYLATED GENES. *Nucleic Acids Research*, 12, 4385-4396.
- UPDIKE, D. L., KNUTSON, A. K., EGELHOFER, T. A., CAMPBELL, A. C. & STROME, S. 2014. Germ-Granule Components Prevent Somatic Development in the *C. elegans* Germline. *Current Biology*, 24, 970-975.
- URBAN, P. L. 2016. Quantitative mass spectrometry: an overview. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, 374.
- VANDAMME, J., SIDOLI, S., MARIANI, L., FRIIS, C., CHRISTENSEN, J., HELIN, K., JENSEN, O. N. & SALCINI, A. E. 2015. H3K23me2 is a new heterochromatic mark in *Caenorhabditis elegans*. *Nucleic Acids Research*, 43, 9694-9710.
- VOO, K. S., CARLONE, D. L., JACOBSEN, B. M., FLODIN, A. & SKALNIK, D. G. 2000. Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Molecular and Cellular Biology*, 20, 2108-2121.
- WINTER, S. & FISCHLE, W. 2010. Epigenetic markers and their cross-talk. In: LIPPS, H. J., POSTBERG, J. & JACKSON, D. A. (eds.) *Essays in Biochemistry: Epigenetics, Disease and Behaviour*. London: Portland Press Ltd.
- WOODHOUSE, R. M., BUCHMANN, G., HOE, M., HARNEY, D. J., LOW, J. K. K., LARANCE, M., BOAG, P. R. & ASHE, A. 2018. Chromatin Modifiers SET-25 and SET-32 Are Required for Establishment but Not Long-Term Maintenance of Transgenerational Epigenetic Inheritance. *Cell Reports*, 25, 2259-+.
- XIAO, Y., BEDET, C., ROBERT, V. J. P., SIMONET, T., DUNKELBARGER, S., RAKOTOMALALA, C., SOETE, G., KORSWAGEN, H. C., STROME, S. & PALLADINO, F. 2011. *Caenorhabditis elegans* chromatin-associated proteins SET-2 and ASH-2 are differentially required for histone H3 Lys 4 methylation in embryos and adult germ cells. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 8305-8310.
- XU, L. & STROME, S. 2001. Depletion of a novel SET-domain protein enhances the sterility of *mes-3* and *mes-4* mutants of *Caenorhabditis elegans*. *Genetics*, 159, 1019-1029.
- YERGEY, J., HELLER, D., HANSEN, G., COTTER, R. J. & FENSELAU, C. 1983. ISOTOPIC DISTRIBUTIONS IN MASS-SPECTRA OF LARGE MOLECULES. *Analytical Chemistry*, 55, 353-356.

- YUAN, Z. F., LIN, S., MOLDEN, R. C., CAO, X. J., BHANU, N. V., WANG, X. S., SIDOLI, S., LIU, S. C. & GARCIA, B. A. 2015. EpiProfile Quantifies Histone Peptides With Modifications by Extracting Retention Time and Intensity in High-resolution Mass Spectra. *Molecular & Cellular Proteomics*, 14, 1696-1707.
- ZIPPO, A., SERAFINI, R., ROCCHIGIANI, M., PENNACCHINI, S., KREPELOVA, A. & OLIVIERO, S. 2009. Histone Crosstalk between H3S10ph and H4K16ac Generates a Histone Code that Mediates Transcription Elongation. *Cell*, 138, 1122-1136.