Deposition of methyl marks on H3K4 (histone 3 lysine 4)
and their significance during Caenorhabditis elegans
embryogenesis

A thesis submitted to The University of Manchester for the degree of
DOCTOR OF PHILOSOPHY in the Faculty of Life Sciences

2014

SIYAO WANG
List of Contents

List of Contents........................................................................................................2
List of Figures ..............................................................................................................8
List of Tables ..............................................................................................................10
Abstract ...................................................................................................................11
Declaration ..............................................................................................................12
Copyright statement ...............................................................................................12
List of Abbreviation .................................................................................................13
Acknowledgements .................................................................................................16

Chapter 1: General Introduction ..........................................................................17

1.1 Introduction to chromatin .............................................................................18

1.1.1 Chromatin structure......................................................................................18
1.1.2 Regulation of transcriptional initiation .........................................................20
1.1.3 Regulation of transcriptional elongation.......................................................21
1.1.4 Regulating transcription in the context of chromatin ..................................23

1.2 Histone 3 lysine 4 methylation .....................................................................23

1.2.1 The “writer” of H3K4 methylation: the MLL/SET/COMPASS complex ......24

1.2.1.1 The H3K4 methyltransferases ..................................................................24
1.2.1.2 The core complex components ..................................................................26
1.2.1.3 The MLL/SET/COMPASS complexes from yeast to human ...............28
1.2.1.4 The recruitment of the MLL/SET/COMPASS complex .......................35
1.2.2 Readers: Downstream effectors of the H3K4 methylation .........................37
1.2.3 “Erasers” of H3K4 methylation: Demethylases ...........................................41

1.3 C. elegans as a model ...............................................................................42

1.3.1 Embryogenesis of C. elegans ...................................................................45
Chapter 3: Lineage specific trimethylation of H3 on lysine 4 during *C. elegans* early embryogenesis

ABSTRACT

3.1 Introduction

3.2 Materials and Methods

3.2.1 Strains and general maintenance

3.2.2 Co-immunofluorescence

3.2.3 Image processing

3.2.4 Peptide competition assay

3.2.5 Western blot analysis

3.2.6 Embryonic and fertility analysis

3.3 Results

3.3.1 H3K4me3 in early embryo and on chromosome X

3.3.2 Cell lineage specific deposition of H3K4me3 at the eight-cell stage

3.3.3 The MLL complex is responsible for acquisition of H3K4me3 during early embryogenesis

3.3.4 H3K4me3 deposition remains low in germline blastomeres from rbr-2 mutants

3.3.5 Absence of PIE1 activates deposition of H3K4me3 in P lineage

3.3.6 H3K4me3 deposition is robust in the germline

3.4 Discussion

3.5 References

Chapter 4: The *C. elegans* MLL/SET/COMPASS complex core component RBBP-5 can antagonise accumulation of H3K4me2/me3 in primodial germ cells
Chapter 4: The core components of the MLL/SET/COMPASS complexes aid RNA Pol II processivity and prevent somatic cells from adopting germline characteristics in Caenorhabditis elegans embryos..............119

ABSTRACT ..................................................................................................................120

5.1 Introduction ........................................................................................................121
5.2 Materials and Methods

5.2.1 Strains and general maintenance

5.2.2 Generation of worm strains

5.2.3 Co-immunofluorescence for embryos and larvae

5.2.4 Co-immunofluorescence for gonad

5.2.5 Western Blot

5.2.6 Cell fate challenge assay

5.2.7 RNAi by feeding

5.2.8 Antibodies

5.3 Results

5.3.1 The core components of SET/MLL/COMPASS complex are essential for the viability of C. elegans.

5.3.2 H3K4 methylation level is further reduced in the double mutants compare to the single mutants.

5.3.3 Transcriptional elongation (pSer2), but not initiation (pSer5), is reduced in the core component mutants.

5.3.4 The determinant of germline blastomeres, P granules, are ectopically expressed in the somatic cells of the core component mutants.

5.3.5 Removal of the core components causes a decrease of developmental plasticity in somatic blastomeres.

5.3.6 The role of the core components on regulation of PGL-1 ectopic expression is independent of transcription.

5.4 Discussion

5.5 References

Chapter 6: General Discussion

6.1 H3K4me3 is a lineage specific mark in early embryos of C. elegans
6.2 The differential activities of WDR-5, RBBP-5 and ASH-2 may explain their differential roles on longevity ................................................... 160

6.3 How does the H3K4methylation affect transcription state? .............. 161

6.4 H3K4me3 on the bivalent promoters might be crucial for maintaining the lineage specification ........................................................................ 161

6.5 Future perspective .............................................................................. 162

6.6 Conclusion ........................................................................................... 163

Chapter 7: Appendix .................................................................................. 164

Chapter 8: References ............................................................................... 175

Final words count: 39610
List of Figures

Figure 1. Genome compaction and histone modifications .................................................. 19
Figure 2. Regulation of transcriptional initiation and elongation by transcription factors ...... 22
Figure 3. 3D reconstruction and modeling of the Set1/COMPASS complex ......................... 28
Figure 4. Components of the MLL/SET/COMPASS complexes from yeast to human .......... 29
Figure 5. Downstream effectors of the H3K4 methylation ............................................. 41
Figure 6. Anatomy of C. elegans ..................................................................................... 44
Figure 7. Transcription during early embryogenesis of C. elegans .................................... 47
Figure 8. A model of interaction between PIE-1, synMuv B proteins and MES proteins ...... 52
Figure 9. RNAi by feeding ............................................................................................... 64
Figure 10. The H3K4me3 mark during early embryonic development ................................ 72
Figure 11. H3K4me3 deposition is enriched in the AB lineage ........................................ 74
Figure 12. H3K4me3 deposition is resumed in the somatic lineage .................................. 75
Figure 13. WDR-5 is required to generate the H3K4me3 mark .......................................... 76
Figure 14. Mutants of the MLL complex show reduced levels of H3K4me3 ..................... 78
Figure 15. H3K4me3 deposition remains low in germline blastomeres in rbr-2 mutants ...... 80
Figure 16. H3K4me3 deposition is activated in germline blastomeres in pie-1 mutants ...... 82
Figure 17. Germline deposition of H3K4me3 in N2 and wdr-5 mutants ............................ 84
Figure 18. Depiction of the embryogenesis process in C. elegans .................................... 99
Figure 19. H3K4me1 deposition is mainly regulated by the RBBP-5 in embryos ............... 103
Figure 20. H3K4me2 is differentially affected by the loss of WDR-5, RBBP-5, and ASH-2 .................................................................................................................... 104
Figure 21. H3K4me3 deposition is largely affected in wdr-5, rbbp-5 and ash-2 mutants during embryogenesis .............................................................................................................. 105
Figure 22. Line charts of H3K4 methylation level in WT (N2), wdr-5(-), rbbp-5(-) and ash-2(-) ........................................................................................................................................ 107
Figure 23. Western blot analysis of the H3K4 methylation in WT, wdr-5, rbbp-5 and ash-2 ..................................................................................................................................... 108
Figure 24. RBBP-5 and ASH-2 act positively on the regulation of H3K4me2 in PGCs ...... 110
Figure 25. Introduction to the MLL/SET/COMPASS complexes in C. elegans and generation of double mutants ........................................................................................................ 124
Figure 26. The rbbp-5 ash-2 double mutant is lethal ......................................................... 130
Figure 27. H3K4 methylation is further reduced in embryos and germline of double mutants, compare to single mutants ................................................................. 133

Figure 28. The pSer2, but not the pSer5, is reduced in the two double mutants ........... 136

Figure 29. One of the P granules, PGL-1, is ectopically expressed in somatic cells of core components mutants ........................................................................................................ 140

Figure 30. The other two P granules, PGL-3 and GLH-1, are ectopically expressed in somatic cells of mutants ........................................................................................................ 142

Figure 31. The somatic cells of core components’ mutants lose their developmental plasticity ........................................................................................................ 145

Figure 32. PGL-1 expression is not affected by cdk-9 (RNAi) ........................................ 147

Figure 33. The regulatory role of the MLL/SET/COMPASS complexes in C. elegans’ embryos ........................................................................................................ 150

Figure 34. The model of H3K4me-involved lineage specification of C. elegans .......... 159
List of Tables

Table 1. Post-translational histone modifications and their associated functions .................. 19
Table 2. Strains used and generated in this study ............................................................... 57
Table 3. Primer used in this project .................................................................................. 58
Table 4. Embryonic viability and fertility analysis of mutants of the MLL core complex ...... 85
Table 5. Characterisation of the embryonic lethality and brood size of mutants .............. 129
Table 6. Characterisation of the pSer2 defect .................................................................... 137
Abstract

Institution: The University of Manchester

Name: Siyao Wang

Degree Title: PhD Molecular Cancer Studies

Thesis Title: Deposition of methyl marks on H3K4 (histone 3 lysine 4) and their significance during Caenorhabditis elegans embryogenesis

Date: 2014

Covalent modifications of the histone proteins can alter chromatin structure and have crucial effects on DNA-based processes. For example, methylation at lysine 4 on histone 3 (H3K4) is correlated with activation of transcription. H3K4 methylation is deposited by a variety of complexes called MLL/SET/COMPASS complexes, which have in common a core complex including WDR-5, RBBP-5 and ASH-2. It remains unclear whether the deposition of H3K4 methylation by the core complex is important for transcriptional regulation or just a consequence of transcription. In order to start investigating this question, we developed a system in which we could track the changes of H3K4 methylation at a cellular level by using the C. elegans embryonic and larval development system. In this system, we found that H3K4 methylation is lineage and stage specific. Importantly, H3K4me2/me3 are deposited according to transcription states (Chapter 3). In order to link this observation to the MLL/SET/COMPASS complexes, we investigated the contribution of each core component toward deposition of H3K4 methyl marks. We found that the contribution of each core component toward H3K4 methylation is different according to the methyl mark and developmental stage involved. Unexpectedly, we also revealed that RBBP-5 and ASH-2 can antagonise deposition of H3K4 methylation in the primordial germ cells (Chapter 4). We next characterised the phenotypic consequences of altering H3K4 methylation in embryos. Our results indicate that removal of the core components causes defects in C. elegans viability and lineage specification (Chapter 5). Importantly, we also uncovered the role of the MLL/SET/COMPASS complex on RNA Pol II processivity. In summary, this study shed light on the individual role of MLL/SET/COMPASS core components in the deposition of H3K4 methyl marks, and it revealed that H3K4 methylation has an impact on transcriptional regulation.
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DoculInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.
List of Abbreviation

"Win" motif: WDR5 interaction motif
AdoHcy: S-adenosyl-L-homocysteine
AdoMet: S-adenosyl-L-methionine
AMA-1: AMAnitin resistant -1
ASH-2: A trithorax-related protein-2
ATAC complex: ADA2A-containing complex
BPTF: Bromodomain and PHD finger transcription factor
C. elegans: Caenorhabditis elegans
CDK: Cyclin-dependent kinase
CFCA: Cell fate challenge assay
CFP-1: CXXC finger protein 1
CHD1: Chromo-ATPase/helicase-DNA binding domain 1
CHROMO-domain: Chromatin organization modifier domain
COMPASS: Complex Proteins Associated with Set1
Cps: COMPASS
CTD: C-terminal domain
DPY-30: DumPY-30
DRB: 5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole
ERα: Estrogen receptor α
GLH-1: Germ Line Helicase-1
H3K27me3/2: Tri-/di-methylation on histone 3 lysine 27
H3K4me3/me2/me1: Tri-/di-/mono-methylation on histone 3 lysine 4
HAT: Histone acetyltransferase
HDAC1: Histone DeAcetylase-1
HLH-1: Helix Loop Helix-1
HMTs: Histone methyltransferases
Hox gene: Homeotic genes
ING proteins: Inhibitor of growth proteins
JmjC domain: Jumonji C domain
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>KMTs</td>
<td>Lysine methyltransferases</td>
</tr>
<tr>
<td>Let-418</td>
<td>LETHal-418</td>
</tr>
<tr>
<td>lincRNAs</td>
<td>Long intergenic non-coding RNAs</td>
</tr>
<tr>
<td>LSD</td>
<td>Lysine demethylases</td>
</tr>
<tr>
<td>MEG</td>
<td>Maternal-effect germ-cell defective</td>
</tr>
<tr>
<td>MEP-1</td>
<td>Mog interacting, Ectopic P granules -1</td>
</tr>
<tr>
<td>MES</td>
<td>Maternal effect sterile</td>
</tr>
<tr>
<td>MEX</td>
<td>Muscle EXcess</td>
</tr>
<tr>
<td>Mvp</td>
<td>Multiple ventral protrusions</td>
</tr>
<tr>
<td>NOS-1/2</td>
<td>NanOS related protein-1/2</td>
</tr>
<tr>
<td>NSL</td>
<td>Non-specific lethal</td>
</tr>
<tr>
<td>NURF</td>
<td>Nucleosome remodelling factor</td>
</tr>
<tr>
<td>OMA</td>
<td>Oocyte MAtration defective</td>
</tr>
<tr>
<td>Paf1</td>
<td>Polymerase-associated factor</td>
</tr>
<tr>
<td>PAR-3</td>
<td>Abnormal embryonic PARtitioning of cytoplasm-3</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PELP1</td>
<td>Proline glutamic acid and leucine-rich protein 1</td>
</tr>
<tr>
<td>PGCs</td>
<td>Primordial Germline cells</td>
</tr>
<tr>
<td>PHD domain</td>
<td>Plant Homeo Domain</td>
</tr>
<tr>
<td>PIE-1</td>
<td>Pharynx and Intestine in Excess-1</td>
</tr>
<tr>
<td>PRC2</td>
<td>Polycomb Repressive Complex 2</td>
</tr>
<tr>
<td>PRMT-1</td>
<td>Protein arginine methyltransferases-1</td>
</tr>
<tr>
<td>P-TEFb</td>
<td>Positive transcription elongation factor b</td>
</tr>
<tr>
<td>PTMs</td>
<td>Post-translational modifications</td>
</tr>
<tr>
<td>RAR</td>
<td>Retinoic acid receptor</td>
</tr>
<tr>
<td>RBBP-5</td>
<td>Retinoblastoma-binding protein-5</td>
</tr>
<tr>
<td>RBR-2</td>
<td>RB (Retinoblastoma Binding protein) Related Protein-2</td>
</tr>
<tr>
<td>RNA Pol II</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td><em>Saccharomyces cerevisiae</em></td>
</tr>
<tr>
<td>SAGA</td>
<td>Spt-Ada-Gcn5 acetyltransferase</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>SAM</td>
<td>S-adenosyl-L- [methyl-3H] methionine</td>
</tr>
<tr>
<td>SEC</td>
<td>Super elongation complex</td>
</tr>
<tr>
<td>SET domain</td>
<td>Su(var)3-9, Enhancer-of-zeste, trithorax domain</td>
</tr>
<tr>
<td>SILAC</td>
<td>Stable isotope labelling by amino acids in cell culture</td>
</tr>
<tr>
<td>SLIK</td>
<td>SAGA-like</td>
</tr>
<tr>
<td>SynMuv B</td>
<td>Synthetic Multivulval class B</td>
</tr>
<tr>
<td>TAF3</td>
<td>Transcription associate factor 3</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>The MLL complex</td>
<td>The myeloid/lymphoid or mixed-lineage leukemia complex</td>
</tr>
<tr>
<td>The NURD complex</td>
<td>The Mi-2/nucleosome remodeling and deacetylase complex</td>
</tr>
<tr>
<td>TOR</td>
<td>The target of rapamycin</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start site</td>
</tr>
<tr>
<td>UTX</td>
<td>Ubiquitously transcribed tetratricopeptide repeat, X chromosome</td>
</tr>
<tr>
<td>WDR-5</td>
<td>WD repeat-containing protein-5</td>
</tr>
<tr>
<td>XLMR</td>
<td>X-linked mental retardation</td>
</tr>
</tbody>
</table>
Acknowledgements

Firstly I would like to thank my main supervisor Dr. Gino Poulin for his patience and guidance throughout my project. The time and effort he contributed to this project made it more productive and interesting. I would also like to thank my advisor Dr. Catherine Millar and my co-supervisor Dr. Alan Whitmarsh. The help and guidance they both provided is invaluable. In this project, all nematode strains were provided by the Caenorhabditis Genetics Center, which is funded by the NIH National Center Research Resources (NCRR).

Secondly, I would like to thank all past and present members of Poulin’s lab. Special thanks go to Dr. Kate Fisher, for her generous help with methods and inspiration throughout my PhD. Additionally, I would also like to give my special thanks to Dr. Fiona Gee, Dr. Feng Xue, Elizabeth Wilkins, and Ulrike Klemstein, whose help and advice has been greatly appreciated. Importantly, their friendship made my time here a wonderful one.

I would like to thank all of my friends for their companies and encouragement, which made the last four years much easier. Finally, I would like to thank my parents for their endless love and support.
Chapter 1

General Introduction
1.1 Introduction to chromatin

Chromatin state is defined by the interaction of the genomic DNA and histones leading to the compaction of the genome. There are two basic functions of chromatin: The first is to compact DNA into a smaller volume in order to fit the genome into the nucleus of the cell; the second is to regulate all DNA-based processes, such as transcriptional regulation, DNA replication and DNA repair (Gerasimova and Corces, 2001). In order to maintain these two functions, the chromatin state must be highly organised and dynamic. The genome undergoes several levels of folding to increase condensation, and mechanisms need to be in place to allow the transcription factors and RNA polymerases to access DNA according to the requirement of the cell (Saumweber, 1987). Increasing evidence suggest that the chromatin state has a fundamental and direct role in the process of regulating gene transcription (Berger, 2007; Kouzarides, 2007; Shilatifard, 2006; Smith and Shilatifard, 2010).

1.1.1 Chromatin structure

In chromatin, long strain of DNA wraps around histones to form a “beads on a string” structure (Kornberg, 1974; Kornberg and Lorch, 1999). The “bead” in this structure is the basic unit of chromatin called nucleosome, which functions in compacting DNA strands and regulating DNA–based processes (Kornberg, 1974; Luger et al., 1997). The nucleosome consists of DNA and four core histones. In the eukaryotic cells, histones can be grouped into five major classes: H1, H2A, H2B, H3 and H4 (Kornberg, 1974; Kornberg and Lorch, 1999). H1 is the linker histone, which functions in locking the DNA into the nucleosome and allowing the formation of higher order structure; H2A, H2B, H3 and H4 are the core histones on which 147 base pairs of DNA is wrapped around to form the nucleosome (Kornberg, 1974; Kornberg and Lorch, 1999). These four core histones are highly conserved throughout evolution, and they all share a feature of a long tail protruding from the nucleosome on the N-terminal side of the amino acid structure (Alva et al., 2007). Many post-translational modifications can be covalently deposited on residues of these histone tails, especially on the tails of the histone H3 and H4 (Berger, 2002; Rea et al., 2000; Strahl and Allis, 2000). These histone modifications include methylation, acetylation, phosphorylation, ADP ribosylation, ubiquitination and SUMOylation (Figure 1) (Table 1) (Kouzarides, 2007). These modifications can impact on the chromatin state. Indeed, it is known that these modifications establish an inheritable “histone code” that
facilitates transcription activation or repression in eukaryotic cells (Abmayr and Workman, 2012; Berger, 2002; Jenuwein and Allis, 2001).

Figure 1. Genome compaction and histone modifications (Adapted from Shilatifard, 2012)

The long strand of DNA is compacted within the nucleus of the cell through its interactions with histone proteins, formed by the basic unit of chromatin, the nucleosome. Structural studies demonstrated that the N-terminal tails of histones protrude outward beyond the nucleosome. Many of the post-translational modifications can be deposited on these tails of histones, providing a landing pad for diverse transcription factors, and chromatin regulators. Three modifications are indicated: Histone H3 lysine 4 (H3K4) methylation (me), Histone H3 lysine 79 (H3K79) methylation (me) and histone H2B monoubiquitination (ub).

<table>
<thead>
<tr>
<th>Chromatin modifications</th>
<th>Residues Modified</th>
<th>Functions regulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetylation</td>
<td>K-ac</td>
<td>Transcription, Repair, Replication, Condensation</td>
</tr>
<tr>
<td>Methylation (lysines)</td>
<td>K-me1 K-me2 K-me3</td>
<td>Transcription, Repair</td>
</tr>
<tr>
<td>Methylation (arginines)</td>
<td>R-me1 R-me2a R-me2s</td>
<td>Transcription</td>
</tr>
<tr>
<td>Phosphorylation</td>
<td>S-ph T-ph</td>
<td>Transcription, Repair, Condensation</td>
</tr>
<tr>
<td>Ubiquitylation</td>
<td>K-ub</td>
<td>Transcription, Repair</td>
</tr>
<tr>
<td>Sumoylation</td>
<td>K-su</td>
<td>Transcription</td>
</tr>
<tr>
<td>ADP ribosylation</td>
<td>E-ar</td>
<td>Transcription</td>
</tr>
<tr>
<td>Deimination</td>
<td>R&gt;Cit</td>
<td>Transcription</td>
</tr>
<tr>
<td>Proline isomerization</td>
<td>P-cis&gt; P-trans</td>
<td>Transcription</td>
</tr>
</tbody>
</table>

Table 1. Post-translational histone modifications and their associated functions (Adapted from Kouzarides, 2007)

K= Lysine; R= Arginine; S= Serine; T= Threonine; E= Glutamic acid; P= Proline
1.1.2 Regulation of transcriptional initiation

Gene transcription is regulated at two interconnected levels: the first level involves the basic transcription apparatus and transcription factors; the second level involves chromatin and its regulator (Lee and Young, 2013; Li et al., 2007). Most relevant to my project are the transcription factors that play a role in either transcriptional initiation or transcriptional elongation.

The first step of transcription is the formation of the preinitiation complex (Sikorski and Buratowski, 2009; Thomas and Chiang, 2006). The preinitiation complex comprises the RNA polymerase II and a group of general transcription factors (such as TFIIID) that recognize specific promoter elements, such as TATA sequence and the downstream promoter element (DPE) (Sikorski and Buratowski, 2009). These transcription factors normally contribute to transcription initiation by recruiting coactivators including the Mediator and P300 (Juven-Gershon and Kadonaga, 2010; Malik and Roeder, 2010; Taatjes, 2010). Many studies have highlighted the importance of the Mediator, as it is essential for activator-dependent transcription by acting as a bridge between the RNA Pol II and general transcriptional activators, initiation factors, and other components of the transcription machinery. (Borggrefe and Yue, 2011; Casamassimi and Napoli, 2007; Conaway and Conaway, 2011; Malik and Roeder, 2010; Taatjes, 2010). The formation of the perinitiation complex is initially unstable because the DNA is in a closed conformation (Kim et al., 2000). To open the double-stranded DNA, the melting of the promoter requires the recruitment of the transcription factor TFIIH. TFIIH contains an ATP-dependent helicase that binds to the RNA Pol II and open the DNA double-strands to initiate mRNA synthesis at the transcription start site (TSS). However, this RNA synthesis frequently results in the production of short and abortive RNAs that are less than 10 nucleotides. Only if the RNA length is longer than 10 nucleotides that transcription is considered successful. Then, the RNA Pol II can enter the main body of a gene and produce a full-length transcript. This step is called promoter escape, which depends on the phosphorylation on the serine 5 (Ser5) residue of the carboxyl-terminal region (CTD) of RNA Pol II by TFIIH (Figure 2A). After promoter escape, transcriptional initiation is completed and elongation can proceed (Levine, 2011; Sikorski and Buratowski, 2009).
1.1.3 Regulation of transcriptional elongation

Good experimental evidence has demonstrated that the RNA Pol II arrests after the production of a nascent RNA (30-50nt in length) (Gilmour, 2009; Saunders et al., 2006; Shilatifard, 1998). Further studies reveal that this promoter-proximal pausing of RNA Pol II is a common feature of developmentally regulated genes (Guenther et al., 2007; Muse et al., 2007; Shilatifard, 2012). Thus, the release of the paused RNA Pol II and its entry into the elongation phase of transcription is a crucial mechanism of gene control during development.

After transcriptional initiation, negative elongation factors bind to the RNA Pol II and arrest it at the promoter proximal position. Two factors, DRB sensitivity-inducing factor (DSIF) and NELF, are acting as negative elongation factors required for generating paused RNA Pol II (Missra and Gilmour, 2010; Peterlin and Price, 2006; Wu et al., 2005). Release of the paused RNA Pol II and entry to the productive elongation requires the recruitment of the positive transcription elongation factor b (P-TEFb) (Figure 2 B). One subunit of the P-TEFb, the Cdk9 kinase, is responsible for catalysing phosphorylation at serine 2 (Ser2) of CTD of RNA Pol II. This phosphorylation is required for the transition from paused Pol II to elongating Pol II (Boehm et al., 2003) (Figure 2 B). In addition, P-TEFb also phosphorylates the “E” subunit of the NELF and the SPT5 subunit of DSIF. The phosphorylation on SPT5 ejects the DSIF from the NELF, and transforms the DSIF to a positive elongation factor travelling with RNA Pol II during elongation (Peterlin and Price, 2006; Price, 2008). Then, other elongation factors come into this complex and establish a steady rate of transcriptional elongation. These elongation factors include TFII S that aids in restarting arrested RNA Pol II and stimulates the elongation rate of Pol II, the PAF complex that generates transcription-associated histone modifications, and the super elongation complex (SEC), which is required for the rapid induction of transcription in response to environmental stimuli during development and in cancer pathogenesis (Fish and Kane, 2002; Luo et al., 2012a; Luo et al., 2012b; Smith et al., 2011; Zhou et al., 2012). The number of factors involved both in positively or negatively regulating RNA Pol II suggest that the mechanism to ensure proper elongation is under strict control during development.
Figure 2. Regulation of transcriptional initiation and elongation by transcription factors (Levine, 2011)

(A) Regulation of transcriptional initiation. Transcription factor THIID binds to the specific DNA sequence, such as TATA box, and leads to recruitment of other general transcriptional machinery, like TFIID, TFIIB and RNA Pol II. This perinitiation complex is unstable, as the complex is in a close conformation. Recruitment of TFIIH opens conformation of the complex and onset of transcription. TFIIH also deposits phosphorylation on CTD of RNA Pol II to maintain the stable transcription. (B) Regulation of transcriptional elongation. Top panel presents the paused RNA Pol II model, as the RNA Pol II is paused downstream of the DPE element. At this stage, the RNA Pol II contains phosphorylation on Ser5 of CTD and binds with two negative elongation factors, NELF and DSIF. The recruitment of P-TEFb helps the release of NELF and deposition of P-Ser2 on CTD, thereby induces the transcriptional elongation.
1.1.4 Regulating transcription in the context of chromatin

Transcription is also regulated by proteins that can alter the chromatin architecture either by ATP-dependent chromatin remodelling or by deposition of histone modifications. Transcription activation coupled with recruitment of ATP-dependent chromatin remodelling complexes can alter the structure of chromatin to facilitate access of the basal transcription apparatus and transcription factors to the genomic DNA (Li et al., 2007). In addition, post-translational modification on histone tails, such as histone methylation, acetylation, ubiquitination, play an important role on regulation of transcription; some histone modifications change the overall histone charge leading to alterations in chromatin architecture. An example of this is histone acetylation, as it neutralizes the basic charge of the lysine residue. Any alteration of the histone charge can potentially disturb the stabilization of higher–order chromatin structure. Indeed, in vitro experiment confirm that the acetylation of H4K16 has a negative effect on the generation of higher-order chromatin structure (Kouzarides, 2007; Shogren-Knaak et al., 2006). Many chromatin modifications are required for the recruitment of proteins that contribute to transcriptional control. For example, methylation on histone 3 lysine 4 residue (H3K4me) can recruit PHD domain containing proteins, such as BPTF and ING2 proteins. BPTF is a component of the NURF chromatin remodelling complex. The recruitment of the NURF complex tethers the SNF2L ATPase to active Hox gene expression (Wysocka et al., 2006); ING2 proteins bring the Sin3a-HDAC1 histone deacetylase complex to active proliferation-related genes during DNA damage (Shi et al., 2006). Thus, as the methylation cannot alter the overall positive charge of the lysine residue, the finding of these downstream proteins highlights the role of the H3K4 methylation that is to tether enzymatic activities onto chromatin, and facilitate transcriptional regulation. My project investigated how H3K4 methylation is implemented and its role in cell specification during early embryogenesis. Thus, I will next describe where H3K4 methylation is found within the genome and explain how methyl groups can specifically be transferred onto H3K4.

1.2 Histone 3 lysine 4 methylation

Genomic scale analysis reveals that H3K4 methylation is a mark of active transcription at both promoters and gene bodies, and enhancers whether they are active or not. H3K4 can be mono-, di- or tri-methylated (H3K4me1/me2/me3), and each methyl state has specific distribution pattern and correlates with distinct functions. A high level of H3K4me3 is
associated with the 5’ end of virtually all active genes, while H3K4me2 is spread throughout these genes (Bernstein et al., 2005; Ng et al., 2003a; Ruthenburg et al., 2007; Santos-Rosa et al., 2002). Although the H3K4me2/me3 is mainly correlated with active transcription, in some cases, these marks also serve as repressive marks when they are deposited on the 3’ end of certain genes through mediating 3’ end anti-sense transcription in yeast (Margaritis et al., 2012). A genome-wide study of embryonic stem cells reveals that H3K4me3 can be co-localised with tri-methylation on lysine 27 of histone H3 (H3K27me3), a repressive mark of transcription, found at the promoters of certain lineage-specific genes, such as Hox genes. This unique pattern of histone modifications is called the “bivalent” mark, which correlates with development and cell fate decision (Bernstein et al., 2006; Mikkelsen et al., 2007; Vastenhouw and Schier, 2012). H3K4me1 is deposited on the gene body of active genes and enhancer regions regardless of their activity (Chen et al., 2013; Kaikkonen et al., 2013). However, the enrichment of H3K4me1 and H3K27 acetylation serves as marks for active enhancers (Chen et al., 2013; Kaikkonen et al., 2013). This state-specific distribution pattern of H3K4me implies that each state of H3K4me is associated with specific transcriptional activities.

1.2.1 The “writer” of H3K4 methylation: the MLL/SET/COMPASS complex

Methylation on H3K4 is primarily deposited by conserved SET-domain containing Histone MethylTransferases (HMTs) of the SET/MLL family (reviewed in (Shilatifard, 2012). These HMTs associate with the core complex to establish the active form of these enzymes. Structural studies have identified the minimal core complex required for full histone H3K4 methylation in yeast and humans. These highly conserved subunits are: Cps30/Wdr5, Cps 50/RbBP5, Cps60/Ash2L and Cps25/Dpy30, respectively in yeast and in humans (Shilatifard, 2012).

1.2.1.1 The H3K4 methyltransferases

A conserved SET domain is the key feature of the histone lysine methyltransferases (HKMTs), including the H3K4 methyltransferases (Jenuwein et al., 1998; Rea et al., 2000). All of these members of KMT2 share a C-terminal conserved SET domain, which is essential for the H3K4 methylation. SET takes its name from the Drosophila genes Su (var)3-9, Enhancer of zeste [E(z)] and Trx (Jones and Gelbart, 1993; Stassen et al., 1995; Tschiersch et al., 1994b). These SET-domain containing HKMTs have been divided into eight classes of enzyme,
KMT1-KMT8 (Allis et al., 2007). The enzymes responsible for catalysing methylation on H3K4 belong to the KMT2 class. The first identified KMT2 class member was the H3K4 methyltransferase, Set1, in yeast. The yeast Set1-containing macromolecular complex was isolated from nuclear extracts by chromatographic method (Miller et al., 2001). This complex was named as the Set1/COMPASS complex, which is the only H3K4 methyltransferase in yeast. However, there are more than one Set1/COMPASS-like complexes in other species. For example, mammal contains six Set1-like H3K4 MTs belonging to the SET and MLL family (Set1A, Set1B, MLL1-4), and each of them forms a COMPASS-like complex (Shilatifard, 2008).

The function of the SET domain is to transfer the methyl group from the S-adenosyl-L-methionine (AdoMet) to the amino group of the lysine residue on histones, producing a methylated lysine residue and a cofactor byproduct S-adenosyl-L-homocysteine (AdoHcy) (Trievel et al., 2002; Zhang et al., 2003b). The structural analysis reveals that the binding site for the AdoMet and the histone substrate are located at the opposite sides of the SET domain. Therefore, a narrow hydrophobic channel that runs through the core of the SET domain is formed to link the substrate lysine and the AdoMet, and transfers the methyl group from AdoMet to the ε-amino group of lysine (Trievel et al., 2002; Zhang et al., 2003b). This structure of the SET domain permits multiple rounds of lysine methylation. A side chain of an aromatic residue from the post-SET region forms an interaction between the aromatic π rings and the adenine moiety of the AdoMet. The cumulative effect of this interaction transfers the labile methyl group to the methyltransfer pore and allows processive lysine methylation (Dillon et al., 2005; Trievel et al., 2002). In addition, the specificity of the methylation states (mono-, di- and tri-methylation) is determined by a critical tyrosine (Y) at position 3942 of MLL1 (Zhang et al., 2003b). Mutation of the Y3942 residue to a phenylalanine (F) converts the SET domain of MLL1 from a mono-methyltransferase to a tri-methyltransferase (Patel et al., 2009). These studies raise a question of how the MLL1/COMPASS regulates state-specificity in vivo. One hypothesis is that the MLL1 core components (WDR-5, RBBP-5, ASH-2 and DPY-30) may alter the conformation of the key Y3942 in the SET domain, thus allowing the di-and tri-methylation of H3K4.
1.2.1.2 The core complex components

H3K4 methyltransferases exhibit very low intrinsic methyltransferases activity. Their full activity can only be achieved in the presence of the core complex formed by four essential components: Cps30/Wdr5, Cps50/RbBP5, Cps60/Ash2L and Cps25/Dpy30, respectively in yeast and humans (Dou et al., 2006; Southall et al., 2009). A crystal structure of the MLL1 SET domain reveals that, in order to establish a full methyltransferase activity, the inserted SET motif (iSET motif) of the SET domain need to be re-orientated to form the optimal conformation (Southall et al., 2009). This re-orientation requires the addition of the core components (WDR5, RbBP5, Ash2L and Dpy30), which greatly enhance the catalytic activity of the SET domain (Southall et al., 2009). This study suggests that core components of the MLL/SET/COMPASS complex contribute to enhance the catalytic activity of the SET domain proteins through structural regulation.

\textit{SET^{\text{Win}}-WDR5-H3 interaction}

One of these core components is WDR5, a WD-40 repeat beta-propeller containing protein. This component is essential for the assembling of the MLL/SET/COMPASS complexes. Previous studies indicate that WDR5 binds to the amino terminal region of H3, especially H3K4me2 (Wysocka et al., 2005). This led to the proposal that WDR5 is crucial for the processivity of the SET domain to go from di-methylation to tri-methylation. Subsequently, structural studies reveal that the axial cavity of WDR5 specifically recognises the H3 arginine 2 side chain. This binding leaves the target residue, H3K4, exposed to the methyltransferase. Another binding segment to the WDR5 axial cavity was identified within MLL1. This motif is located in an unstructured region outside the N-flanking region of the SET domain. This motif responsible for the interaction between MLL1 and WDR5 is known as the “Win” motif (WDR5 interaction motif) (Patel et al., 2008; Song and Kingston, 2008). These findings suggest a model by which WDR5 serves as a scaffold presenting lysine 4 residue of histone tail to the methyltransferase for further methylation.

\textit{WDR5-RbBP5 interaction}

In addition to the interaction between WDR5 and histone H3 or with the “Win” motif of MLL1, it was demonstrated that WDR5 can also form a stable heterodimer with another core component, RbBP5. RbBP5 is a WD-40 beta-propeller containing protein. Structural study
reveals that the tail region of the RbBP5 interacts with a binding site of WDR5. This RbBP5 binding site of WDR5 is located on the opposite face of the WDR5 beta-propeller in which WDR5 interacts with histone H3 (Avdic et al., 2011; Odho et al., 2010). These interactions form a SET\textsuperscript{Wen}-WDR5-RbBP5 tripartite complex, which plays an important role in catalysing H3K4 methylation (Avdic et al., 2011). In vitro studies demonstrated that WDR5 alone does not stimulate the catalytic activity of the SET domain protein in the HMT activity assay, but addition of the WDR5- RbBP5 subcomplex can increase the HMT activity (Odho et al., 2010). Therefore, these studies propose that in an active MLL/SET/COMPASS complex, WDR5 secures the methyltransferase or the lysine residue on one side, and its effector RbBP5 on the other side, to facilitate the productive interaction between SET domain and RbBP5 beta-propeller.

**RbBP5-Ash2L-DPY30 interaction**

Another study on the MLL1 complex identified an interaction between RbBP5 and Ash2L. Ash2L is a trithorax protein and an important component of the MLL/SET/COMPASS complexes (Steward et al., 2006). The role of the Ash2L on H3K4 methylation is dependent on its two interacting partners, RbBP5 and a dimer of DPY30. A conserved C-terminal SPRY domain of Ash2L was recognized as a binding site for the RbBP5 (Cao et al., 2010), and a small motif, DPY30 binding motif (DBM), is responsible for the interaction between Ash2L and DPY30 (Chen et al., 2012). Importantly, this RbBP5-Ash2L heterodimer can also directly interact with AdoMet and H3 and stimulates the methyltransferases activity of the MLL1 (Cao et al., 2010). The Ash2L-DPY30 interaction has recently been reported to be important for the H3K4 tri-methylation (Jiang et al., 2011). Further, work using 3D crystal EM reconstruction of the yeast and human MLL/SET/COMPASS complexes revealed a Y-shape architecture with Ash2L/(DPY30)\(_2\) forming the base, and WDR5 and RbBP5 localizing on the top two lobes (Figure 3). The SET domain of the H3K4 methyltransferase is located at the juncture of these core components, and acts as a platform for catalysis (Takahashi et al., 2011). Taken together, these structural studies have established the configuration of the MLL\textsuperscript{Wen}-WDR5-RbBP5-Ash2L-DPY30 complex.
1.2.1.3 The MLL/SET/COMPASS complexes from yeast to human

The first member of the MLL/SET/COMPASS complexes is the yeast H3K4 Set1/COMPASS complex (Miller et al., 2001; Roguev et al., 2001). This complex is highly conserved from yeast to human (Figure 4). In C. elegans, two Set1-related proteins, SET-2 and SET-16, were identified in COMPASS-like complexes (Fisher et al., 2010; Simonet et al., 2007). In Drosophila, there are three Set1/COMPASS-related complexes, dSet1/COMPASS, Trx/COMPASS and Trr/COMPASS (Mohan et al., 2011). It has been found that for each of these three Drosophila complexes there are two homologous complexes in mammalian (Figure 4). They are the dSet1-like Set1A and Set1B/COMPASS, the Trx-like MLL1 and MLL2/COMPASS, and the Trr-like MLL3 and MLL4/COMPASS complexes (Eissenberg and Shilatifard, 2010; Shilatifard, 2008). I herein discuss the function of the MLL/SET/COMPASS complexes in three species: S. cerevisiae, C. elegans, and in mammals.
Figure 4. Components of the MLL/SET/COMPASS complexes from yeast to human

The names of MLL/SET/COMPASS complexes from yeast to human are listed in the left lane. The middle lane shows the name and the structure of the SET-domain containing methyltransferases for each species. The right lane shows the components existing in each species. The conserved core components are shown in red.
The yeast Set1/COMPASS complex

Methyltransferase activity at H3K4 was firstly discovered in yeast. In 2001, a SET domain-containing complex was isolated from *S. cerevisiae* culture using a chromatographic method (Miller et al., 2001). This study identified a SET domain containing protein, Set1, existing within a macromolecular complex named COMPASS, for COMplex of Proteins ASSociated with Set1. Following analysis revealed that this Set1/COMPASS complex consist of Set1 and six other polypeptides of about 25, 30, 35, 40, 50, and 60 kDa. These additional components are known as Cps (COMPASS) 25, Cps30, Cps35, Cps40, Cps50 and Cps60, respectively (Miller et al., 2001). In yeast, the Set1/COMPASS complex is able to catalyse mono-, di-, and tri-methylation on H3K4.

The SET domain of Set1 contains methyltransferase activity, as any alteration to the SET domain of Set1 results in a loss of H3K4me1/me2/me3 in yeast. However, Set1 alone is not the active form of this enzyme, other subunits are required to regulate and stabilize the enzymatic activity of Set1. Indeed, the minimal subunit composition required for full H3K4 methylation includes Set1, Cps25, Cps30, Cps50 and Cps60, as removal of any of these causes a significant loss of H3K4me1/me2/me3 in yeast cells (Takahashi et al., 2011). Cps35 is required for the Rad6/Bre1-dependent Set1/COMPASS recruitment to chromatin (Discussed later in the recruitment section). Cps35 is also the only subunit of COMPASS essential for the viability of yeast. However, the importance of Cps35 for viability is due to its role within other complexes (cleavage and polyadenylation factor complex) but not in COMPASS (Nedea et al., 2008). Cps40 is required for proper H3K4me3, as the yeast cell lacking Cps40 is defective in H3K4me3, yet shows no detectable effect on H3K4me1 and H3K4me2 (Schneider et al., 2005). These studies have first demonstrated that the SET domain-containing methyltransferase functions within a macromolecular complex, the COMPASS complex, to catalyse H3K4 methylation.

The *C. elegans* SET-2/SET-16/COMPASS complex

Two SET domain-containing methyltransferases have been identified in *C. elegans*, SET-2 and SET-16. SET-2 is the ortholog of yeast Set1, and SET-16 belongs to the mammalian MLL family. The difference between these two HMTs is the N-terminal structure: SET-2 contains a
RRM region, whereas SET-16 contains PHD domains and an FY-rich region. The differences in their structure suggest a distinct role of each HMT. It was proposed that although both SET-2 and SET-16 are required for H3K4 methylation in *C. elegans*, SET-2 plays a predominant role. Unpublished data (Elizabeth Wilkins and Gino Poulin) have shown that SET-16 acts early in embryonic development and specifically for H3K4me3. Whereas SET-2 acts later and specifically for H3K4me3/me2. Surprisingly, none of these impacts on H3K4me1. Of note, *set-16* deletion is inviable, probably due to its early activity. RNAi against *set-16* can cause embryonic lethality and Multivulva developmental defect, indicating of hyper-activation of the RAS signalling pathway (Fisher et al., 2010). Deletion of *set-2* can suppress sterility observed in *hpl-1/2*. These proteins are homologous to HP1, heterochromatin protein 1, which is associated with repression of transcription. This led to the proposal that SET-2 acts in establishing and/or maintaining euchromatin (Simonet et al., 2007). Interestingly, recent work has identified an H3K4me1 MT, SET-30. SET-30 is involved in suppressing the sterility produced by the deletion of *spr-5*, a demethylase targeting H3K4me1/me2 (Greer et al., 2014). These data together, indicate that both SET-2 and SET-16 are important writer of H3K4me2/me3 and H3K4me3 and that SET-30 is probably producing the H3K4me1 state.

Similar to the yeast *Set1/COMPASS* complex, SET-2 and SET-16 also exist within a macromolecular complex, the COMPASS-like complex, to enhance their enzymatic activity. The *SET-2/COMPASS* and *SET-16/COMPASS* complexes share four core components, WDR-5 (Cps30), RBBP-5 (Cps50), ASH-2 (Cps60) and DPY-30 (Cps25), although there are also some subunits specific to each complex. Increasing evidences have shown that these four core components are essential for embryonic H3K4me2/me3 in *C. elegans*, as previous papers show that knockdown of WDR-5, RBBP-5, ASH-2 and DPY-30 results in a reduction of H3K4me2/me3 in embryos (Li and Kelly, 2011; Wang et al., 2011b; Xiao et al., 2011), but the effect of the core components on H3K4 methylation in adult cells (including somatic cells and germ cells) is very limited. These core components are also involved in the regulation of longevity in *C. elegans*, *wdr-5* mutant exhibits a long-live phenotype, as well as depletion of RBBP-5 and ASH-2 by RNAi shorten the worms’ life span (Greer et al., 2011). However, the mechanism by which the components regulate longevity is still not understood. In addition, a recent paper reveals that the WDR-5 paralog, WDR-5.2, acts redundantly during the spermatogenesis-oogenesis transformation (Li and Kelly, 2014). Interestingly, this role of the
WDR-5 and WDR-5.2 is independent from their roles in COMPASS complex, suggesting that the core components might also function within other complexes. Another example is DPY-30, it has been demonstrated that DPY-30 not only associates with the COMPASS complex, but it also has functions within the dosage compensation complex (DCC) to repress the X-linked gene expression (Pferdehirt et al., 2011). These studies indicate that each core component may have additional functions not necessarily related to H3K4 methylation.

WDR82 (Cps35) and CFP-1 (Cps40) are subunits specific to the SET-2/COMPASS complex. WDR82 is not required for H3K4 methylation, but it is involved in the specification of left-right asymmetry in the ASEL and ASER neuron pair (ASE left and right) (Poole et al., 2011). CFP-1 (CXXC finger protein 1) is a CXXC finger containing protein, which is able to recruit SET-2/COMPASS complex to unmethylated CpGs-enriched promoters and deposit H3K4me3 (Chen et al., 2014). UTX-1 is a subunit specific to SET-16/COMPASS complex. UTX-1 is a JmjC domain-containing protein belonging to the KDM6 family, which is responsible for demethylating H3K27me3/H3K27me2, a negative mark of transcription. UTX-1 is crucial for the embryonic viability and post-embryonic development (Fisher et al., 2010; Vandamme et al., 2012; Vandamme and Salcini, 2013). However, a recent study suggests that the developmental role of UTX-1 is independent of its catalytic activity, but is likely mediated by its role within the SET-16/COMPASS complex (Vandamme et al., 2012; Vandamme and Salcini, 2013). Taken together, UTX-1 is essential for the development of C. elegans due to its role within the SET-16/COMPASS complex, and it also establishes the crosstalk between the negative mark, H3K27me3, and the positive mark, H3K4me3. Interestingly, these two marks often co-localise on “bivalent” promoters, which is a unique chromatin landscape with H3K4me3 and H3K27me3 marks at promoter of key developmental genes (such as Hox genes) (Vastenhouw and Schier, 2012).

1.2.1.3.3 The mammalian MLL/SET/COMPASS complex

Six COMPASS-like complexes were found in mammalian cells: two Set1/COMPASS homologs, Set1A/Set1B/COMPASS, and four MLL/COMPASS complexes, MLL1-4/COMPASS. All of these complexes share the same set of core components (WRAD), which are essential for the stability and enzymatic activity of HMTs, but there are also some subunits
unique to each complex. These specific subunits for each complex lead to distinct biological outcome.

Two components, CXXC1 and Wdr82 (the yeast Cps40 and Cps35, respectively), are specifically found within the Set1A and Set1B/COMPASS complexes. Previous studies indicate that reduction of Wdr82 results in a decrease of the Set1A/B protein level, followed by a global loss of the H3K4 tri-methylation level, even though the functional MLL1-4/COMPASS complexes are still present in those cells. It demonstrates that, similar to yeast Set1/COMPASS and worm SET-2/COMPASS, the Set1A/ Set1B/COMPASS complexes play a predominant role on the deposition of H3K4 tri-methylation in mammals. As the Set1A/B/COMPASS complexes in mammalian cells function as the major H3K4 methyltransferases, the MLL1-4/COMPASS complexes are thought to have gene-specific functions.

Previous studies have demonstrated that the MLL1/MLL2/COMPASS complexes positively regulate Hox gene expression. Removal of the MLL1 in mouse embryonic stem (ES) cells causes embryonic lethality. Moreover, the expression of Hoxa7 and Hoxc9 is impaired in MLL1 heterozygote (+/-)(Terranova et al., 2006; Yu et al., 1995). A subsequent study confirms the role of MLL1 as a regulator of the Hox gene expression. By analysing the global pattern of H3K4 methylation in the MLL1 mutant, it reveals that MLL1 is required for the tri-methylation on H3K4 of less than 5% of promoter carrying the H3K4me3 modification. These MLL1 targeted genes include Hox genes, albeit not all Hox genes (Guenther et al., 2005). Further study finds that the removal of the MLL1 results in the decrease of the RNA Pol II recruitment, reduced expression and loss of H3K4 tri-methylation at these genes (Wang et al., 2009). Importantly, this role of MLL1 in the Hox gene regulation is also associated with leukemic pathogenesis (Tenney and Shilatifard, 2005; Yokoyama et al., 2004).

To understand the importance of the MLL1/MLL2/COMPASS complexes in Hox genes regulation, the deletion of menin, a component specific to MLL1/MLL2/COMPASS complexes, was used to examine the effect on H3K4 methylation at Hox gene clusters (Wang et al., 2009). The result shows that the H3K4 methylation on all Hox genes is abolished in the menin mutant (Wang et al., 2009). This study demonstrates that the MLL1/MLL2/COMPASS complexes are
the major methyltransferases for the *Hox* genes, as no effect on H3K4 methylation and expression of the *Hox* genes is detected by depleting other MLL/SET/COMPASS complexes.

Similar to the MLL1/MLL2/COMPASS complex, deletion of the MLL3/MLL4 results in embryonic lethality (Eissenberg and Shilatifard, 2010). However, no effect on the homeotic genes was detected in the MLL3/MLL4 mutants. It indicates that the MLL3/MLL4/COMPASS complexes are not major regulators of the homeotic genes, in contrast to the MLL1/MLL2/COMPASS complexes.

Recently, two studies in *Drosophila* and mammalian cells revealed that the Trr-related MLL3/MLL4/COMPASS complexes are major H3K4 monomethyltransferases functioning at enhancer regions and required for the enhancer-promoter communication for tissue-specific genes (Herz et al., 2012; Hu et al., 2013a). Depletion of the MLL3/MLL4 results in a great loss of H3K4me1 and H3K27ac on enhancer regions. Moreover, as the mutation of the MLL3/MLL4 is frequently found in multiple types of cancer, this study also suggests that the MLL3/MLL4 mutation associated with cancer may be caused by enhancer malfunction (Morin et al., 2011; Parsons et al., 2011).

One of the components specific to the MLL3/MLL4 complexes is UTX, a demethylase of H3K27 methylation (Herz et al., 2010). The role of the MLL3/MLL4 on the regulation of enhancer activity indicates that the H3K27 demethylase UTX can be recruited to the enhancer region within the MLL3/MLL4/COMPASS complexes (Herz et al., 2010; Herz et al., 2012). In mammalian ES cells, the inactive/poised enhancer is coupled with H3K27 methylation, which prevents H3K27 from being acetylated. Recruitment of the UTX within the MLL3/MLL4 complexes to those inactive enhancer regions would allow the H3K27 to be demethylated and then acetylated (Herz et al., 2012). This leads to a model whereby the H3K4 monomethyltransferases MLL3/4 and the H3K27 demethylase UTX may function together to regulate the transition from inactive/poised state to active state of enhancers. However, this model still needs to be formally demonstrated.

In summary, the Set1/COMPASS homologs, *C. elegans* SET-2/COMPASS and mammalian Set1A/Set1B/COMPASS, are major methyltransferases responsible for depositing global
H3K4 methylation. The MLL/COMPASS complexes, *C. elegans* SET-16/COMPASS and mammalian MLL1-4/COMPASS, contribute to H3K4 methylation on specific gene loci, such as Hox genes and enhancer regions.

1.2.1.4 The recruitment of the MLL/SET/COMPASS complex

Above, I have explained the role that the different components of the MLL/SET/COMPASS complex play in the process of methylation at H3K4. I will next describe how the MLL/SET/COMPASS complex is recruited to target genes. There are mainly five recruitment mechanisms to target genes. The complex can be recruited by 1) the basal transcriptional machinery, 2) specific DNA sequence, 3) long non-coding RNA, or 4) nuclear hormones and 5) DNA damage. Importantly, the recruitment of the MLL/SET/COMPASS can involves the combination of different mechanisms.

H3K4 methylation is enriched at promoters and bodies of actively transcribed genes. A common character of these locations have is the presence of the basal transcriptional machinery, thus it is a possibility that the transcriptional machinery could be involved in the recruitment of the MLL/SET/COMPASS complex. Indeed, a global proteomic analysis in *S. cerevisiae* reveals that the polymerase-associated factor (Paf1) complex, which associates with elongating RNA Pol II, is required for the recruitment of the Set1/COMPASS complex to gene promoters and proper H3K4 methylation (Wood et al., 2003a; Wood et al., 2003b). Further studies in mammals have found that Paf1 has a dual function by serving as a landing pad for the SET/MLL1/COMPASS complex on the RNA Pol II and promoting the H2B monoubiquitination through Rad6/Bre1 (Wood et al., 2003a). H2B monoubiquitination stimulates the SET/MLL1/COMPASS complex to generate the H3K4me2/me3 (Wood et al., 2003a). Depletion of the Rad6/Bre1 results in a loss of H3K4me2/me3, but does not affect the recruitment of the Set1/COMPASS complex in yeast (Wood et al., 2003b). In this Rad6/Bre1-regulated H3K4 methylation, one component of the Set1/COMPASS complex, Cps35 (or Wdr82 in mammals), is essential for H3K4 tri-methylation (Wu et al., 2008). In summary, the importance of Paf1 in H3K4 methylation and its link to RNA Pol II support the concept that the transcriptional machinery is crucial for the recruitment of the MLL/SET/COMPASS complex.
Other experiments have revealed that CpG islands can act as specific gene sequences for recruitment of the MLL/SET/COMPASS complex. A study in mammals reveals that a SET1A/B/COMPASS specific subunit, CXXC1, can be recruited to unmethylated CpG islands to deposit H3K4 methylation (Miller et al., 2001). The genome-wide profiling indicates that CXXC1 is located at 80% of CpG islands, most of which enriched for H3K4 methylation (Bird et al., 1985; Lee et al., 2001; Miller et al., 2001). Moreover, immunoprecipitation experiments suggest that unmethylated CpG islands are the binding sites for the SET1/COMPASS complexes. Consistent with the mammalian studies, the C. elegans CXXC1 homolog, CFP-1, is also involved in the recruitment of SET-2/COMPASS to the unmethylated CpGs-enriched promoters. The interaction between the CXXC1 subunit and the unmethylated CpG islands is due to the presence of the CXXC domain within the CXXC1 protein (Miller et al., 2001). In mammals, the trithorax-like MLL1 and MLL2 methyltransferases also contain CXXC domain within their polypeptide sequence, suggesting that MLL1/2 potentially can be recruited to CpG islands as well (Tate et al., 2009). Interestingly, the yeast homolog of the CXXC1, Cps40, lacks a CXXC domain, but has a PHD domain, and the drosophila counterpart of the MLL familay, trithorax, also lacks a CXXC domain, suggesting that although the MLL/SET/COMPASS complexes are highly conserved in function and structure, their recruitment mode can be differ from each other.

Long intergenic non-coding RNAs (lincRNAs) also play an important role in targeting the Trithorax-like MLL1/2 complexes. For example, the HOTTP RNA, a lincRNA transcribed from the 5' tip end of the Hoxa locus, binds to the WDR5-MLL1 complex and targets it to the 5' Hoxa locus, leading to H3K4me3 deposition and transcriptional activation (Wang et al., 2011a). Another study in mammals has identified another lincRNA, Mistral, which activates transcription of Hoxa6 and Hoxa7 through recruiting MLL1/COMPASS complex to the Mistral gene (Bertani et al., 2011). This recruitment triggers changes in chromosome conformation, followed by activation of Hoxa6 and Hoxa7 transcription. These results connect lincRNAs with the recruitment of MLL1/COMPASS complex to Hox gene clusters to implement H3K4 methylation on this locus and activate transcription.

Unlike the MLL1/COMPASS complex, the recruitment of Trr-like MLL3/MLL4/COMPASS to the promoter of target genes can be regulated by a nuclear receptor signalling pathway. In the
presence of hormone ecdysone, the hormone receptor can form a heterodimeric complex with its substrate, and translocates together with the MLL3/MLL4/COMPASS complex into the nucleus to activate ecdysone-induced genes transcription (Sedkov et al., 2003). The MLL3/MLL4 can also be recruited to its target genes in response to the P53-mediated DNA damage signalling. Study shows that the MLL3/MLL4/COMPASS specific subunit NCOA6 binds to P53 and directs the MLL3/MLL4/COMPASS complexes to the promoter of the P53 targeted genes under the DNA damage signalling (Lee et al., 2009). Thus H3K4 methyltransferases can be recruited to target genes by specific signalling pathway.

Apart from these four mechanisms of recruitment of MLL/SET/COMPASS complex, it is generally appreciated that the MLL/SET/COMPASS complexes contain a wealth of potential histone modification binding domains, suggesting that the recruitment of the complex can be regulated by other histone modifications (Lee et al., 2010). In yeast, asymmetric H3R2 di-methylation prevents recruitment of the Set1/COMPASS complex and globally affects the H3K4me3 (Kirmizis et al., 2007). This role of the H3R2 di-methylation is also found in mammalian cells, as it interferes with the recruitment of various COMPASS complexes in mammals. Therefore the recruitment of the H3K4 methyltransferases is very intricate and still only partially understood.

1.2.2 Readers: Downstream effectors of the H3K4 methylation

In the previous sections, I described where H3K4 methylation is found in the genome, how it is deposited by the conserved MLL/SET/COMPASS complex, and how it can be recruited to its targets. I will next explain the function pertaining to recognition of the methyl marks at H3K4. There are two mechanisms by which histone modifications can mediate their effect. The first mechanism is by directly altering the charge of histones. The second is by the recruitment of non-histone proteins that recognise specific histone modifications. In the first case, histone modifications can affect the higher-order of chromatin structure by altering the contact between nucleosomes or the interaction between histone and DNA. Biophysical analysis indicates that any modifications altering the positive histone charge will cause alteration in chromatin architecture (Kouzarides, 2007; shogren-Knaak et al., 2006). However, methylation on H3K4 does not significantly change the overall basic charge of the nucleosome. Thus, theoretically, the H3K4 methylation cannot affect the chromatin architecture directly.
implying that H3K4 methylation must regulate chromatin events mainly through the second mechanism. (Tony Kouzarides 2007)

This mechanism involves proteins that are recruited by histone modifications via specific recognition domains. H3K4 methylation can be recognized by chromo domain, tudor domain, MBT domain and PHD domain (Smith and Shilatifard, 2010). Structural analysis reveals that an “aromatic cage” within these domains is responsible for the methyl-lysine recognition (Jacobs and Khorasanizadeh, 2002; Nielsen et al., 2002). Many proteins containing these domains have been found recruited to the H3K4 methylation marks; their bindings bring about other proteins that can either further modified histone tails (such as HATs) or interact physically with the basal transcriptional machinery. Below I provide examples of this recognition process.

BPTF (bromodomain and PHD finger transcription factor), a component of the nucleosome remodelling factor (NURF), recognizes the H3K4me3/2 via its PHD domain (Figure 5). NURF is an ATP-dependent remodelling complex required for maintaining the Hox gene expression pattern during development. Loss of H3K4me3 impairs the recruitment of NURF and its associated ATPase activity to the Hoxc8 promoter. Disruption of this recruitment causes aberrant regulation of Hox gene expression (Wysocka et al., 2006). This study concludes that NURF can be recruited to promoters of Hox genes by binding to methylation on H3K4, and modulates transcription initiation through ATP-dependent chromatin remodelling.

Similar to BPTF, CHD1 (chromo-ATPase/helicase-DNA binding domain 1) is also involved in the ATP-dependent chromatin remodelling process (Pray-Grant et al., 2005). CHD1 is a component of the histone acetyltransferases SAGA (Spt-Ada-Gcn5 acetyltransferase) and SLIK (SAGA-like), and preferentially acetylates H3 and H2B. Histone acetylation directly establishes more accessible chromatin architecture, and recruits bromodomain-containing transcription regulator to active transcription sites (Pray-Grant et al., 2005). Studies have demonstrated that the two chromodomains of CHD1 specifically interact with the methyl mark on H3K4 and then subsequently activates transcription. This interaction has been confirmed in yeast and humans (Sims et al., 2005).
Members of the ING proteins family are correlated with both acetylation and deacetylation (Shi and Gozani, 2005). The PHD finger containing ING3-5 proteins associate with H3K4 methylation and catalyse histone acetylation (Figure 5), which is consistent with the role of H3K4 methylation in activation of transcription. Thus the ING3-5 proteins establish a link between H3K4 methylation and histone acetylation (Shi et al., 2006; Taverna et al., 2006). However, another effector of the H3K4 methylation, the ING2 protein, recruits the repressive histone deacetylases complex Sin3/HDAC1 to silence the highly active, proliferation-specific genes upon DNA damage (Figure 5) (Shi et al., 2006). Thus, in this case, the interaction between the H3K4 methylation and the ING proteins can be linked to both transcriptional activation and repression.

The double tudor domain, JMJD2A, is a demethylase protein containing a Jmjc-domain. The double tudor domain forms an aromatic cage to interact with the tri-methyl marks on H3K4 and H4K20 (Figure 5) (Huang et al., 2006). The JMJD2A is known as a trimethyl-demethylase responsible for converting H3K9me3 to H3K9me2. An increase with H3K9me2 within the gene body is correlated with gene repression, suggesting that the JMJD2A has a role in gene repression (Shi and Whetstine, 2007). Therefore, H3K4 methylation can be coupled with demethylation as an additional mechanism to down regulate gene expression.

A study using stable isotope labelling by amino acids in cell culture (SILAC) has revealed a direct interaction between the H3K4 methylation and TAF3, a subunit of the transcription factor TFIID (Lauberth et al., 2013). TAF3 binds to H3K4 methylation via its PHD domain (Figure 5). This interaction is highly selective for H3K4 methylation. As TAF3 plays a crucial role in transcription initiation, this finding suggests a direct connection between the H3K4 methylation and the basal transcription machinery.

PHF8 is a coactivator for retinoic acid receptor (RAR) and a Jmjc domain-containing H3K9me2 demethylase activity. Studies have revealed that the PHD domain of PHF8 binds to the H3K4me2/me3 and co-localises with H3K4me3 at transcription initiation site (Kleine-Kohlbrecher et al., 2010). This study suggests that H3K4me3 is coupled with H3K9me2 demethylation to maintain transcription activation, as the H3K9me1 is associated with active transcription. Interestingly, a point-mutation within the Jmjc domain of PHF8 has been
correlated with X-linked mental retardation (XLMR) (Kleme-Kohlbrecher et al., 2010). Therefore, this finding provides further evidence that H3K4 methylation can impact on a specific disease.

Among the methyl-binding proteins, few of them act in a methylation-state specific manner. Most of them recognize both the H3K4me2 and H3K4me3 through either the PHD domain or the chromo-like domain. However, two effectors have been identified that only recognizes di-methylation on H3K4. PELP1 (Proline glutamic acid and leucine-rich protein 1), a human oestrogen receptor (ERα) co-regulator, preferentially binds to H3K4me2 and H3K9me2 through a glutamic acid rich region within its C-terminus (Nair et al., 2010). PELP1 interacts with a histone lysine demethylase of the KDM1 family targeting H3K9me2. Recruitment of PELP1 to H3K4me2-enriched regions leads to demethylation of H3K9me2, thereby activating transcription of ERα target genes. Nardilysin (NDc) is the other H3K4me2-specific binding protein. It is linked to gene repression (Li et al., 2012). However, no domain has been identified responsible for this H3K4me2-specific interaction (Li et al., 2012). These studies together reveal that the recognition of H3K4 methylation can impact on other histone modifications, which in turn will regulate gene expression. However, as different states of the H3K4 have distinct distribution pattern and biological function, how the effector recognizes H3K4 methylation on specific gene locus or in different cellular context remains unclear.
1.2.3 “Erasers” of H3K4 methylation: Demethylases

Considering that H3K4 methylation impacts on gene expression, it is important to regulate, in addition to its deposition, its removal. There are two classes of demethylases identified so far: LSD1 (also known as BHC110/KDM1) responsible for removing the monomethyl- and dimethyl-marks of H3K4, and the jumonji (JmjC) domain classes, which remove monomethyl-, dimethyl- and trimethyl- marks from H3K4. LSD1 removes the methyl mark through an amine oxidase reaction (Shi et al., 2004). It was shown that down-regulation of LSD1 results in an increasing level of H3K4 methylation in mammal (Shi et al., 2004). Moreover, this H3K4me2 demethylase, LSD1, is conserved from yeast to human, In C. elegans, removal of SPR-5, the ortholog of LSD1, produce a failure to erase H3K4me2 in primodial germ cells, resulting in defects in oogenesis and spermatogenesis (Katz et al., 2009). This demethylation activity suggests that LSD1’s role could be to regulate transcriptional repression. However, LSD1 can also act as a transcriptional coactivator by demethylation of H3K9 (Metzger et al., 2005; Shi et al., 2004). These observations imply that LSD1 has distinct transcriptional functions that remain difficult to delineate (Shi and Whetstine, 2007).
The JmjC domain was first identified within the Jumonji proteins, which are named by their unique cross-like neural tube phenotype. The JmjC mediated demethylation is performed not only at lysine (especially tri-methylation on lysine residue), but also at arginine residues. 27 different JmjC domain containing proteins have been identified in mammals, and 15 of them are thought to be lysine or arginine demethylases, including, JARID1 (demethylase of H3K4me2/me3) (Christensen and Agger et al., 2007), and the UTX/JMJD3 group (demethylase of H3K27me1/me2/me3) (Cloos and Christensen et al., 2008). *C. elegans* encodes a homolog of the JARID1 demethylase, RBR-2. Depletion of RBR-2 displays a multivulva or vulvaless phenotype and a significant increase in H3K4 methylation in larval and adult stages. It suggests that the RBR-2 acts as a key demethylase of H3K4 methylation and is essential for reliable development of the vulva. Interestingly, UTX is also a subunit of the MLL3/MLL4/COMPASS complex, suggesting that the H3K27 demethylase and H3K4 methyltransferases may function together. Indeed, removal of its *C. elegans* homolog UTX-1 produced a Mutiple ventral protrusions (Mvp) phenotype, which was also observed in the depletion of the SET-16 (*C. elegans* homolog of MLL3/MLL4/COMPASS) and the core components (Fisher et al., 2010). Biochemical analysis confirmed that UTX-1 is a component of the SET-16/COMPASS complex (Fisher et al., 2010; Vandamme et al., 2012; Vandamme and Salcini, 2013) (mentioned in 1.2.1.3.2). In mammal, the H3K27 demethylase UTX within MLL3/4/COMPASS complex is thought to regulate the transition from inactive to active state of enhancers (Herz et al., 2010; Herz et al., 2012) (mentioned in 1.2.1.3.3). Moreover, an inverse correlation between H3K27me3 and H3K4me3 has been identified on the promoters of *Hox* genes (Bernstein et al., 2006). Thus, the co-regulation of the demethylase UTX and the methylase MLL/SET/COMPASS complex is crucial for development.

### 1.3 *C. elegans* as a model

*Caenorhabditis elegans* (*C. elegans*) has been used as a model organism for molecular and development biological research since 1974. *C. elegans* is a transparent and living-free nematode (Brenner, 1974b). There are many reasons to choose it as a model organism in my project: firstly, it is a transparent nematode, thus its development and differentiation processes can been directly observed under microscope (Figure 6) (Stiernagle, 2006). Secondly, *C. elegans* is easy to maintain in the lab, they feed on bacterial, and strains can be frozen, which
allows long-term maintenance (Stiernagle, 2006). *C. elegans* exists primarily as self-fertilized hermaphrodites (Sulston and Horvitz, 1977), and males only comprise 0.05% of the total population (Kimble and Hirsh, 1979; Sulston and Horvitz, 1977). Therefore, mutants and specific phenotypes can be preserved stably during the *C. elegans* reproduction. Thirdly, *C. elegans* is amenable to RNA interference (RNAi) (Fire et al., 1998) to knock down gene expression. RNAi by feeding in *C. elegans* allows us to examine the function of genes in a simple, specific and efficient way. Fourthly, the genome of *C. elegans* has been completely sequenced (Brenner, 1974b). All published strains and information related to *C. elegans* have been collected in the online database. Importantly, H3K4 methylation is known as a mark of active transcription in *C. elegans*. *C. elegans* contains two MLL/SET/COMPASS complexes, the SET-2/COMPASS and SET-16/COMPASS, which are homologues to mammalian Set1A/B/COMPASS and MLL3/4/COMPASS, respectively (Fisher et al., 2010; Simonet et al., 2007). Although the core components of MLL/SET/COMPASS complex are highly conserved through species, the mechanism by which each core component contributes to the regulation of H3K4 methylation-associated processes is still unclear. We herein use the embryogenesis of *C. elegans* as a model to investigate the contribution of each core component toward H3K4me deposition and its effect on transcription, lineage specification and embryonic development.

Embryogenesis of *C. elegans* has gained increasing attention for studying the epigenetic regulation of transcription. This process is particularly important as transcription is highly regulated during early embryogenesis. Transcription is activated in somatic cells from 8-cell stage of embryogenesis, but remains inactive in germline blastomeres. Thus, early embryogenesis is normally used as a model for studying the role of chromatin modifications on regulation of gene expression and lineage specification.

(A) DIC image of an adult *C. elegans*. Scale bar 0.1mm (B) schematic drawing of the structure of *C. elegans*. 
1.3.1 Embryogenesis of C. elegans

Early embryogenesis is completed by the generation of six founder cells, which develop into specialised cell lineages. These six founder cells, AB, MS, E, C, D, and P4, are generated through four asymmetric divisions, and each asymmetric division produces a larger somatic blastomere (AB, EMS, C and D) and a smaller germline blastomere (P1, P2, P3 and P4) (Mello et al., 1992b). After the generation of P4, the germline lineage is completely separated from the somatic lineages. The AB, MS, E, C and D cells are destined to develop into hypodermis, neurons, muscle and pharynx, and the germline blastomere P4 symmetrically divides to yield two primordial germ cells (PGCs), Z2 and Z3, at the 88-cell stage (Strome, 2005). These two PGCs do not divide anymore during embryogenesis and remain close to each other in the embryo (Figure 7 A). Z2 and Z3 resume divisions only in the first larval stage (L1 stage) after feeding, and gradually generate ~8000 germ cells. Germ cell differ widely from the somatic cells (Strome, 2005). Germ cells must maintain the pluripotency state to generate reproductive cells and pass the genomic information to the next generation, whereas somatic cells go through differentiation, development and are destined to die.

1.3.1.1 Transcriptional regulation during early embryogenesis of C. elegans

In many species, transcriptional activity is shut down during oogenesis and reactivated during the early development of the embryo (Batchelder et al., 1999; Evans et al., 1994; Ghosh and Seydoux, 2008). In C. elegans, this transcriptional silencing occurs as the oocytes enter the diakinesis stage of meiosis and undergo maturation (Davidson, 1986). This silencing is maintained until the 4-cell stage of embryo, when low level of transcription can be observed in certain somatic cells (Kelly et al., 2002; Schisa et al., 2001). Transcription gradually increases in all somatic cells, and after Z2-Z3 stage, transcription becomes robust in all somatic cells, but remains at a low level in PGCs. Chromatin modifications marking active transcription state, like deposition of H3K4me2/me3 and H4K8ac, are detected in somatic cells throughout embryogenesis, but are depleted from the germline lineage. The earliest mark of this segregation is H3K4me3 (Wang et al., 2011b). Importantly, this transcriptional pattern corresponds to the distribution of two phosphorylation states on the CTD of RNA Pol II (Seydoux and Dunn, 1997). Phospho-Ser5 (pSer5), a mark of transcriptional initiation, is present in all somatic blastomeres but low in germline blastomeres, and phosphor-Ser2 (pSer2), another phosphorylation state on RNA Pol II relating to transcriptional elongation, is
present in all somatic cells during embryogenesis but absent from the germ blastomeres (Figure 7B) (Kelly et al., 2002; Schisa et al., 2001; Seydoux and Dunn, 1997). pSer2 and pSer5 appear in the Z2 and Z3 cells at the birth of the PGCs and coincide with the degradation of PIE-1, a determinant of germline blastomeres, and general transcriptional repressor (see below). However, transcription in Z2 and Z3 is still at a basal level (Mello et al., 1992a; Mello et al., 1996; Tenenhaus et al., 1998).

Although transcription is dynamic in somatic lineage, restrictive transcription is particularly important in the germline lineage during the whole embryogenesis process. Inactive transcription in germline lineages is essential for preventing the initiation of somatic program and maintaining the pluripotency in germ cells (Mello et al., 1992a). Many transcription-associated chromatin modifications are involved in maintaining the distinction between somatic and germline lineages (Bender et al., 2004; Mello et al., 1992a; Yuzyuk et al., 2009). Therefore, the embryogenesis of *C. elegans* provides an ideal model to study the role of H3K4 methylation on lineage specification and pluripotency maintenance during embryogenesis.
Figure 7. Transcription during the early embryogenesis of *C. elegans*

(A) Depiction of the early embryogenesis of *C. elegans*. Germline lineage is indicated as orange circles; Somatic cells with inactive transcription are shown in grey circles, and somatic cells with active transcription are indicated as blue circles. The On the right side, six stages of embryo are depicted, the number of cells is indicated, and the P cells and somatic cells are in are in dark and light grey, respectively. (B) One stage (P4 stage) is chosen to exemplify the pSer2/5 level in embryos. In staining, pSer2 and pSer5 are poor in the germline blastomeres and enriched in somatic cells. DAPI indicates nuclei within the embryos. Arrows indicate the germline blastomere P4 cell.

1.3.2 Lineage specification of *C. elegans*

Lineage specification during embryogenesis is a fascinating process as it generates two fundamentally different programs: the somatic program and the germline program. The somatic program produces cells that will eventually die; while the germline program maintains the immortality of a species. In *C. elegans*, the specification of germline lineage is due to the asymmetric distribution of specific cytoplasmic complexes, referred to as germ plasm, into the germline blastomeres (Lesch and Page, 2012). The germ plasm contains maternally encoded proteins and RNAs, which are inherited from the mother. These proteins and RNAs are segregated to the germline blastomere side before the asymmetric division. Thus, the germ plasm is hypothesised to specify the germ cell fate. Further studies revealed that the germ plasm maintains the germ cell fate through 1) inhibits somatic gene expression, 2) establishes repressive chromatin modifications and 3) regulates the germline-specific gene expression. Many components within the germ plasm have been studied, including the P granules and PIE-1. Moreover, apart from the germ plasm complexes, many proteins have been identified that are implicated in the soma/germline cell fate decision (Lesch and Page, 2012). Herein, we
focus on two important components of the germ plasm, P granules and PIE-1, and a number of chromatin regulators implicated in germline specification.

1.3.2.1 P granules

P granules are initially inherited from the mother and segregated into the germline blastomeres after the first asymmetric division, and exist in all germ cells throughout the germline development (Strome and Wood, 1982). The partition of P granules into germline blastomeres is a progressive process: firstly, moving to the posterior cell with bulk cytoplasm; secondly, tethering on the nucleus to the posterior side of the cell; thirdly, degrading the granules left in the somatic sister cell. Eventually, the majority of P granules is segregated into the P4 cell in which they localise around the outside of the nuclear envelope, and they will stay at this location throughout the rest of germline development (Updike and Strome, 2010; Updike and Strome, 2009).

P granules contain maternal mRNAs and proteins. Most of the proteins are RNA-binding proteins, including the P granule assembly proteins (PGL-1, PGL-2 and PGL-3), the members of VASA-related germline helicase family (GLH-1, GLH-2 and GLH-3), and the meiotic regulator (OMA-1 and OMA-2) (Updike and Strome, 2010). They are required for the normal germline development, as depletion of the P granules results in sterility and germline developmental defects. In addition, many proteins within the germ plasm are enriched on P granules, such as PIE-1, MEX-1, MEX-3, MEG-1, MEG-2 and Sm proteins, raising the possibility that the P granules organise and deliver these maternal proteins to the germ cells (Guedes and Priess, 1997; Lesch and Page, 2012; Updike and Strome, 2010). Therefore, it is hypothesised that the P granules are essential for the specification of germline.

However, recent studies indicate that mutations disrupting P granules partition are still able to segregate other germ plasm complexes into the germline lineage, and in adulthood these mutants are fertile (Gallo et al., 2010). Thus, although the P granules are crucial for the germline development, they are not unique determinant of the germline cells.
1.3.2.2 PIE-1

PIE-1 is a CCCH zinc finger-containing RNA-binding protein. It acquires its name due to the “pharynx and intestine excess” phenotype produced in pie-1 mutant worm (Mello et al., 1992b; Mello et al., 1996). This important finding has shown that PIE-1 prevents germ line cells from transforming into somatic cells.

The location of PIE-1 is dynamic during embryogenesis. pie-1 mRNA is maternally expressed in gonad, oocyte and all blastomeres until 4-cell stage. After 4-cell stage, pie-1 mRNA is degraded in somatic cell, and gradually enriches in germline blastomeres (Mello et al., 1992b; Reese et al., 2000). After translation of the pie-1 mRNA, PIE-1 protein localises initially uniformly in the cytoplasm and nucleus and then gradually become more concentrated in the nucleus of germline blastomeres. During each asymmetric division, PIE-1 is primarily segregated into germline blastomeres, but a low or un-detectable level of PIE-1 can be left in its somatic sisters (Mello et al., 1992b; Reese et al., 2000). PIE-1 is enriched in germline blastomeres until 100-cell stage. Around the time of gastrulation, the level of PIE-1 starts to decline in Z2 and Z3. (Mello et al., 1992b; Mello et al., 1996; Tenenhaus et al., 1998).

The main function of PIE-1 is to act as a general transcriptional repressor in germline blastomeres (Batchelder et al., 1999; Ghosh and Seydoux, 2008). Removal of PIE-1 dramatically increases the level of transcription in germline blastomeres and causes a germline-to-soma transformation (Batchelder et al., 1999; Mello et al., 1996; Seydoux and Dunn, 1997; Seydoux et al., 1996). My work has also shown that in absence of PIE-1 H3K4me3 reappears in germline blastomeres and PGCs (Wang et al., 2011b). Moreover, ectopic expression of PIE-1 in somatic cells results in transcriptional repression (Mello et al., 1992b). PIE-1 represses transcription by blocking RNA Pol II mediated transcriptional elongation (Batchelder et al., 1999; Seydoux and Dunn, 1997). As mentioned earlier, transcriptional elongation is associated with pSer2 on RNA Pol II CTD, which is catalyzed by the kinase of P-TEFb, Cdk9 (Ng et al., 2003b). In germline blastomeres, PIE-1 contains a motif YAPMAPT, which is homologous to the unphosphorylated CTD region, and competes for the binding of P-TEFb (Ghosh and Seydoux, 2008; Zhang et al., 2003a). Therefore, in the presence of PIE-1, P-TEFb is recruited away from RNA Pol II, thereby preventing the phosphorylation of Ser2 on RNA Pol II. This mechanism is thought to inhibit transcriptional
elongation in germline blastomeres. Moreover, PIE-1 also inhibits initiation (Ghosh and Seydoux, 2008). It contains a region outside the YAPMAPT responsible for blocking Ser5 phosphorylation (Ghosh and Seydoux, 2008). In wild type, the germline blastomeres pSer2 cannot be detected and levels of pSer5 are low, although pSer5 signal can be detected at two prominent foci of germ cell nuclei (Figure 7 B). In the pie-1 deleted mutant, both the pSer2 and pSer5 signals are detected in germline blastomeres; their patterns are the same as observed in somatic cells, i.e. pSer5 signal is evenly distributed and no longer forms foci (Seydoux and Dunn, 1997). In sum, PIE-1 inhibits transcription activation through preventing the phosphorylation on Ser2 and Ser5 residue on RNA Pol II in germline blastomeres.

1.3.2.3 The soma-to-germline transformation

PIE-1 and P granules are required for maintaining germline characteristics in germline blastomeres and PGCs, there are also many proteins responsible for preventing the somatic from acquiring germline characteristics (soma-to-germline transformation), such as the synthetic Multiple Vulva class B (synMuv B) proteins, and the Maternal Effect Sterile (MES) proteins.

The synMuv B proteins contain C. elegans LIN-35, HPL-2, E2F components, EFL-1 and DPL-1, and members of NuRD complex, MEP-1 and LET-418 (Fay and Yochem, 2007). They are a group of conserved transcriptional repressors that have been implicated in soma/germline distinction. The synMuv B mutants show a number of defects in lineage specification, several of which indicate that the somatic cells acquired germline characteristics, such as ectopic expression of P granules and meiosis proteins in somatic cells (Unhavaithaya et al., 2002; Wang et al., 2005). A study has shown that the synMuv B mutants arrest as young larvae by increasing temperature, perhaps because the expression of germline genes compromises the intestinal cell fate (Petrella et al., 2011). Surprisingly, these defects can be rescued by loss of the chromatin modifiers, MES-2/3/6 and MES-4, suggesting that the SynMuv B proteins antagonizes the MES proteins in somatic cells, thereby preventing the establishment of germline chromatin state (Figure 8) (Petrella et al., 2011). Moreover, one of the synMuv B proteins, MEP-1, is able to interact with the germline blastomere determinant, PIE-1 (Unhavaithaya et al., 2002). Therefore, a mechanism was proposed to explain how the two different cell fates segregate during embryogenesis. The authors suggest that the expression
of synMuv B in germline blastomeres is inhibited by PIE-1, leading to expression of MES proteins that promote expression of germline characteristics (Figure 8). However, in somatic cells, there is no PIE-1 and MES proteins are therefore inhibited and germline characteristics repressed. In this model (Figure 8), the chromatin modifier MES proteins are prominent and I thus describe their function in the following section.

1.3.2.4 MES proteins and H3K27 methylation

Transcriptional regulation in germline lineage also depends on chromatin-based mechanisms for repression. Z2 and Z3 have high levels of the transcriptional repressive mark, H3K27me3 (Bender et al., 2004; Schaner et al., 2003). The enzyme for depositing this modification is the Polycomb group orthologous, MES-2/MES-3/MES-6, in C. elegans. The MES-2/3/6 was first identified in a genetic screen designed to find maternal factors required for fertility. Mutation of the MES proteins results in a maternal effect sterile (MES): homozygous mothers (F1) are viable and fertile, but their progenies are sterile (Bender et al., 2004). This phenotype suggests that the MES proteins are crucial for the germline development.

The C. elegans MES-2/3/6 form a complex similar to the Drosophila PRC2 Polycomb complex, which is responsible for depositing H3K27 methylation and transcriptional repression (Strome, 2005). In Drosophila, the PRC2 complex is essential for repressing the expression of developmental regulators and maintaining specific cell types (Schuettengruber et al., 2007). In mouse ES cells, the PRC2 complex prevents the expression of developmental regulators and maintains pluripotency (Niwa, 2007). The C. elegans PRC2, MES proteins, is necessary for silencing X chromosome and methylating H3K27 in both somatic and germline cells. The role of the MES proteins in somatic cells is still only partially defined. A study indicates that one component of the MES proteins, MES-2, influences global chromatin organization and promote the loss of developmental plasticity in somatic cells of C. elegans embryos (Yuzyuk et al., 2009). In sum, the MES proteins form a complex orthologous to the conserved Polycomb Group Complex, and methylate H3K27. Their main function is to maintain germline development through inhibiting developmental regulators and developmental plasticity (Figure 8).
Figure 8. A model of interaction between PIE-1, synMuv B proteins and MES proteins

The SynMuv B proteins and MES proteins are responsible for preventing the somatic cells from adopting germline characteristics (Strome, 2005). In somatic cells (shown in the green circle), the MEP-1/synMuv B proteins can inhibit the expression of MES proteins, and therefore maintain the somatic cell fate (Petrella, 2011). However, in the germline cells (blue circle), PIE-1 blocks the function of the MEP-1/synMuv B proteins. Then, the MES proteins express in the germline cells, and promote germline characteristics (Yuzuk, et al, 2009).

1.4 Aims of the project

H3K4 methylation is one of the most studied chromatin modifications. There is a variety of complexes involved in its deposition but they all have in common a core complex comprising WDR-5, RBBP-5 and ASH-2. The configuration of the core complex is known, but the individual contribution of each component toward H3K4 methylation in metazoans remains undefined. A priori model based on the configuration of the core complex suggested that the deletion of each core component would contribute similarly toward the methylation states at H3K4.

In order to test this model, we needed to develop a system in which we could rely on to track how the methylation states of H3K4 changes through time. C. elegans embryogenesis is ideal for this: transcription is differentially regulated according to two fundamental lineages (somatic and germline); zygotic transcription is activated at a specific stage of embryogenesis; and each cell of embryo can be tracked and labelled by immunofluorescence or GFP. Moreover, null mutations for WDR-5, RBBP-5 and ASH-2 are available.
As the deposition of all three methyl marks at H3K4 can be tracked in *C. elegans* embryos, my first objective was to characterise all the methylation states at H3K4 during defined stages of *C. elegans* embryogenesis (Chapter 3). As the dynamic deposition pattern of H3K4 methylation was established, I sought to determine the contribution of each core component towards the methylation at H3K4 during embryogenesis using the core complex mutants (Chapter 4). Then I endeavoured to characterise the phenotypic consequences of altering H3K4 methylation on the developing embryos (Chapter 5).

### 1.5 Alternative format and contribution of co-authors

The thesis is presented in the alternative format in accordance with the rules and regulations of the University of Manchester. There are three result chapters which are presented in manuscript form in the style suitable for their intended journals of submission. The contribution of each author for each manuscript is listed below.

**Chapter 3:** Lineage specific trimethylation of H3 on lysine 4 during *C. elegans* early embryogenesis

**Contribution of authors:** This manuscript is about the investigation of H3K4me3 deposition pattern during embryogenesis of *C. elegans*. The second author, Dr. Kate Fisher, conducted all of the western blot analysis. The Confocal images processing, GLP-1 staining and the immunofluorescence in Figure 10 was performed by my supervisor, Dr. Gino Poulin. As the first author of this manuscript, I conducted the majority of the immunofluorescence in this study, and I also was fully responsible for writing the text of this manuscript. Dr. Gino Poulin, and Dr. Kate Fisher reviewed and provided comments.

**Chapter 4:** Cellular resolution analysis of H3K4 methylation levels during *C. elegans* development reveals an antagonistic role for the chromatin-associated protein RBBP-5 during L1 diapause

**Contribution of authors:** This manuscript is about the contribution of each core components on H3K4 me1/me2/me3 deposition. Dr. Kate Fisher has conducted all of the western blot analysis.
My supervisor, Dr. Gino Poulin, gave advice of establishing the embryonic system and guidance on all experimental work. As the first author, I performed the vast majority of the experiments, and I also was fully responsible for writing the text of this manuscript. Dr. Gino Poulin, and Dr. Kate Fisher reviewed and provided comments.

Chapter 5: The core components of the MLL/SET/COMPASS complexes aid RNA Pol II processivity and prevent somatic cells from adopting germline characteristics in Caenorhabditis elegans embryos

Contribution of authors: This manuscript is about the characterisation of the phenotypic consequences caused by removing the core components. Dr. Kate Fisher performed the H3K4me1/me2/me3 western blot analysis. C. elegans cDNA which are used for generating cdk-9 RNAi feeding strain was prepared by Dr. Kate Fisher. The majority of the RNAi strains were obtained from Ahringer library. My supervisor, Dr. Gino Poulin, provides advice and guidance on all experiment. As the first author, I performed the vast majority of the experiments and are fully responsible for the writing the text. Dr. Gino Poulin, and Dr. Kate Fisher reviewed and provided comments.
Chapter 2

Materials and Methods
2.1 Strains and general maintenance

Strains were maintained as previously described (Brenner, 1974b). The strains used in this study are: **RB1304** wdr-5(ok1417), **RB1025** set-2 (ok952), **PFR327** set-2 (bn129), **OL0058** rbbp-5(tm3463), **PFR403** ash-2(tm1905), **ZR1** rbr-2(tm1231), **JJ532** pie-1(zu154) unc-25(e156)/qC1 dpy-19(e1259) glp-1(q339), **MT2495** lin-15B (n744), **KM267** hs::hlh-1, **PFR210** swd-3::gfp (wdr-5::gfp) and Bristol N2 as wild type (Table 2).

2.2 Generation of double mutants

Worm strains generated in this study are: **OL0076** wdr-5 (ok1417); rbbp-5 (tm3463), **OL0073** wdr-5 (ok1417); ash-2 (tm1905), rbbp-5 (tm3463) ash-2 (tm1905) (inviable), **OL0084** wdr-5 (ok1417); hs::hlh-1, **OL0085** rbbp-5 (tm3463); hs::hlh-1, **OL0086** ash-2 (tm1905); hs::hlh-1, **OL0107** wdr-5 (ok1417); rbbp-5 (tm3463); hs::hlh-1, **OL0108** wdr-5 (ok1417); ash-2 (tm1905); hs::hlh-1, **OL0089** lin-15B (n744); hs::hlh-1 and **OL0090** wdr-5 (ok1417); hs::hlh-1; wdr-5::GFP. Crosses were performed by placing 3 hermaphrodites with 6 males on a 3cm plate, which is seeded with OP50 strain of *E. coli*. Male was generated by placing L4 hermaphrodites at 30 °C for 3-4 hours. Genotypes of all strains were confirmed by single worm PCR (Table 2).

2.3 Single worm PCR

2.3.1 Worm lysis

One worm was placed in 3 μl lysis buffer (1x Bioline NH₄ buffer, 1mg/ml proteinase K), frozen in -80°C for 1h, thawed, and frozen for another 30min. Worm was lysed at 65 °C for 90min, followed by 95°C for 30min.

2.3.2 PCR amplification

PCR reaction was carried out in a 30 μl volume, containing 1x Bioline NH₄ buffer, 1.5mM MgCl₂, 0.2 mM dNTPs (Bioline), 1.5 μl (~150 pmol) of each primer (Primers used in this project was listed in table 2) and 0.5 μl BioTaq (Bioline). Reaction was performed at 96°C for 2min, 30 cycles were performed of 96°C for 30s, 55°C for 30s, followed by 72°C for 1 min per kbp of the product length. Samples were separated on agarose gel (1% (w/v)) in 0.5x TBE (0.04 M Tris, 0.04 M H3BO3, 0.001 M EDTA, pH 8.0) with SybrSafe dye. Gel was running at 100V for 1h, and visualized under UV light.
<table>
<thead>
<tr>
<th>Strain name</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains used in this study</strong></td>
<td></td>
</tr>
<tr>
<td>RB1304</td>
<td>wdr-5.1(ok1417) III</td>
</tr>
<tr>
<td>RB1025</td>
<td>set-2(ok952) III</td>
</tr>
<tr>
<td>PFR327</td>
<td>set-2(bn129) III</td>
</tr>
<tr>
<td>OL0058</td>
<td>rbbp-5(tm3463) II</td>
</tr>
<tr>
<td>PFR403</td>
<td>ash-2(tm1905) II</td>
</tr>
<tr>
<td>ZR1</td>
<td>rbr-2(tm1231) IV</td>
</tr>
<tr>
<td>JJ532</td>
<td>pie-1(zu154) unc-25(e156)/qC1 [dpy-19(e1259) glp-1(q339)] III</td>
</tr>
<tr>
<td>MT2495</td>
<td>lin-15B(n744) X</td>
</tr>
<tr>
<td>KM267</td>
<td>heat-shock hlih-1</td>
</tr>
<tr>
<td>PFR210</td>
<td>N2 qals4(Ex16[swd-3(wdr-5)::gfp+rol-6(su1006)])</td>
</tr>
<tr>
<td><strong>Strains generated in this study</strong></td>
<td></td>
</tr>
<tr>
<td>OL0076</td>
<td>wdr-5.1(ok1417) III;rbbp-5(tm3463) II</td>
</tr>
<tr>
<td>OL0073</td>
<td>wdr-5.1(ok1417) III;ash-2(tm1905) II</td>
</tr>
<tr>
<td>OL0084</td>
<td>wdr-5.1(ok1417) III; hs::hlh-1</td>
</tr>
<tr>
<td>OL0085</td>
<td>rbbp-5(tm3463) II; hs::hlh-1</td>
</tr>
<tr>
<td>OL0086</td>
<td>ash-2(tm1905) II; hs::hlh-1</td>
</tr>
<tr>
<td>OL0107</td>
<td>wdr-5.1(ok1417) III;rbbp-5(tm3463) II; hs::hlh-1</td>
</tr>
<tr>
<td>OL0108</td>
<td>wdr-5.1(ok1417) III;ash-2(tm1905) II; hs::hlh-1</td>
</tr>
<tr>
<td>OL0089</td>
<td>lin-15B(n744) X; hs::hlh-1</td>
</tr>
<tr>
<td>OL0090</td>
<td>wdr-5.1(ok1417) III;qals4(Ex16[swd-3(wdr-5)::gfp+rol-6(su1006)]); hs::hlh-1</td>
</tr>
</tbody>
</table>

**Table 2.** Strains used and generated in this study.
<table>
<thead>
<tr>
<th>Gene name</th>
<th>Deletion/insertion allele</th>
<th>Primers: F and R</th>
<th>Extension time/min</th>
<th>Product size in N2 / bp</th>
<th>Product size in mutant/bp</th>
</tr>
</thead>
</table>
| **wdr-5.1** | ok1417 | F:ATTGTGTGTTCGCTGTGCAT  
R:CGTATTTGCTCTCGTGTCAT | 2.5 | 2239 | 1544 |
| **rbbp-5** | tm3463 | F:ACTTCAGAATGCCTCTCGGA  
R:CCCATGCTGGATTCAGAAGA | 1.5 | 1505 | 500 |
| **ash-2** | tm1905 | F: GTCGCCGTCGATAAGCACTC  
R: CTGCCCTCGGTAGTGCTCCT | 2.5 | 2185 | 1582 |
| **hs::hlh-1** | HSP16.41::hlh-1 | F: CACGCTTTGTTCTAGTGCATC  
R: GGGGTTGTTCGTCGTTCCT | 1 | 0 | 990 |

<table>
<thead>
<tr>
<th>Vector name</th>
<th>Insert</th>
<th>PCR primers</th>
<th>PCR template</th>
</tr>
</thead>
</table>
| L4440 cdk-9 | cdk-9 gene base 171-946 | F: GCTTGAAGTGAAGCAGCAACTAT  
R: ACAGCTCATTTCTGGACTCTGTG | N2 cDNA |

Table 3. Primer used in this project
2.4 Antibodies

Primary Antibodies used in co-immunofluorescence: mouse anti-H3K4me3 (Abcam, ab10121:100), rabbit anti-H3K4me3 (Abcam, ab8580 1:100), rabbit anti-H3K4me2 (Abcam, ab32356, 1:100), mouse anti-H3K4me2 (Millipore CMA303, 1:5,000), rabbit anti-H3K4me1 (Abcam ab8895, 1:500,000), rabbit anti-H3K27me3 (Millipore/Upstate, 07-449, 1:50), rabbit anti-H3K27me2 (Millipore/Upstate, 07-452, 1:100), and rabbit anti-H3 (Abcam, ab1791, 1:500), mouse anti-P-Ser2 (Covance MMS-129R, 1:50), mouse anti-P-Ser5 (Covance MMS-134R, 1:50), mouse anti-Myosin (developmental studies hybridoma bank 5-6, 1:100), mouse anti-GLP-1 (a gift from Judith Kimble, 1:10,000), rabbit anti-PGL-1, rat anti-PGL-3 and rabbit anti-GLH-1 (Gifts from Susan Strom, 1:10,000), and mouse anti-MyoD (Gift from Susan Strom, 1:100).

The secondary antibodies used in co-immunofluorescence: Alexa Fluor® 594 Goat Anti-Rat IgG (H+L) (Life technologies A-11007, 1:1000), DyLight 594 AffiniPure Goat Anti-Mouse IgG (H+L) (Jackson ImmunoResearch 115-515-146, 1:200), DyLight 488 AffiniPure Goat Anti-Rabbit IgG (H + L) (Jackson ImmunoResearch 111-485-144, 1:200), DyLight 488 Goat Anti-Mouse IgM mu chain (Abcam ab97007, 1:200) and DAPI (Sigma 28718-90-3, 2 µg/µl). Our controls performed without primary antibody show that these secondary antibodies do not produce significant background (data not shown).

Primary antibodies used in western-blot analysis: rabbit anti-H3K4me3 (Abcam, ab8580 1:1,000), rabbit anti-H3K4me2 (Abcam, ab32356, 1:1,000), rabbit anti-H3K4me1 (Abcam ab8895, 1:100,000), and rabbit anti-histone H3 (Abcam ab1791, 1:1,000).

Secondary antibodies in western-blot analysis: Peroxidase-conjugated AffiniPure Goat Anti-Mouse IgG (H+L) (Jackson ImmunoResearch 115-035-003, 1:10,000), Peroxidase-conjugated AffiniPure Goat Anti-Rabbit IgG (H+L) (Jackson ImmunoResearch 111-035-003, 1:10,000), Horse Radish Peroxidase (HRP) conjugated Goat Anti-Mouse IgM mu chain (Abcam ab97230, 1:500).
2.5 Co-immunofluorescence for embryos

Immunofluorescence by freeze crack was performed on 0.3% polylysine treated slides. Slides were treated by 75μl of 0.3% polylysine, dried 10 min at 70 °C, quickly rinsed in distilled water and excess liquid was wiped off. The slides were 3 × 14 mm printed wells slides from Fisher Scientific LTD UK. About 30 mothers were placed in a well with a drop of M9 buffer to wash off bacteria. These mothers were transferred into a 7 μl drop of M9 on the polylysine treated well by using an eyelash picker. Embryos were released from these mothers by dissecting with a syringe needle. A cover slip (22 mm × 50 mm) was applied at a right angle and placed at −80 °C for at least 20 min. The cover slip was then promptly removed and the slide with embryos was methanol fixed at −20 °C for 10 min, washed 5 min in PBS, and then followed by two washes in PBS-tween 0.2%. Primary antibody was diluted in PBS-tween 0.2% and incubated overnight at 4 °C in a humid chamber, following with three washes in PBS-tween 0.2%. Secondary antibody was incubated 2 h at 37 °C. Three washes were performed as described above. 5 μl Mowiol was applied to preserve fluorescence.

The fixation for anti-pSer2 and anti-pSer5 is different from others. After freeze crack, the slide was placed in -20 °C MeOH for 30sec, transferred to 1×PBS, 0.08mM HEPES (pH6.9), 1.6mM MgSO4, 0.8mM EGTA, 3.7% Formaldehyde for 30 min. Washes were performed in PBT (1×PBS+0.1% Triton+0.1%BSA).

2.6 Co-immunofluorescence for gonad

A small drop of 2mM tetramisole was placed on a poly-lysine treated 3-well slides. 30 adult worms were transferred to the drop, and paralyzed in the tetramisole. Slice the adults near the heads and tails with two syringe needles. After dissection, the gonads will extrude automatically and stick to the slides. Then a cover slide was placed and followed by the same treatment as mentioned in 2.5.

2.7 Co-immunofluorescence for larvae

For larvae staining, the L1 stage Larvae were synchronized and washed off from plates. These larvae were transferred to a 1.5ml eppendorf and diluted to appropriate concentration.
Put a drop of larvae (around 100 larvae) on a polylysine treated slide. Apply the same fixation and immunofluorescence as mentioned above.

### 2.8 Image processing

Confocal images were analysed using imageJ software. Individual slices were selected for every cell to discriminate between overlapping cells. We used eight-cell stage embryos in which AB descendants are distinctive from P1 descendants, *i.e.* the AB cells enter mitosis faster than the P1 lineage. We expressed the results as a ratio of the H3K4me3 signal over the H3K27me2 signal for each cells.

### 2.9 Peptide competition assay

Peptides were preincubated in three fold excess with the anti-H3K4me3 (ab1012) at room temperature for 30 min prior to immunostainings. Immunostainings was performed as described above. The peptides used were produced by Abcam (ab1432, H3K4me3) and (ab1782, H3K27me3).

### 2.10 Western Blot

Whole worm protein extracts were prepared by harvesting synchronized worms. A minimum volume of 20 μl pelleted worms was collected for each protein sample, washed in PBS buffer, boiled in 1× Sample buffer (60mM Tris-HCl, pH 6.8, 2% (w/v) sodium dodecyl sulfate, 0.01% (v/v) bromophenol blue, 10% (w/v) glycerol) containing 100 mM DTT and sonicated using a BioRuptor UCD-200 (Diagenode). Western blots were performed following the standard protocol with the BioRad Mini Trans-Blot cell system. Briefly, proteins were segregated on a polyacrylamide gel at 120 V for 15min, followed by 200V for 60min. Then the proteins were transferred to Amersham Hybond-P membrane (GE Healthcare) at 100 V for 30-90min (depends on the size of the proteins) and the membrane was blocked in 10% milk in PBS-T overnight at 4°C. On the next day, the membrane was incubated with primary antibody in PBS-T at room temperature for 1h, washed 3x 15min in PBS-T, and incubated with secondary antibody with shaking for 1h. The detection was performed using the Amersham ECL Plus Western blotting Detection system (GE Healthcare) and Kodak BioMax MR Film. Antibodies against actin and histone H3 were used for normalisation.
2.11 Embryonic and fertility analysis

To calculate brood size and embryonic lethality, ten clones per strain were transferred onto fresh plates every day until the worms had finished laying. The number of embryos and progeny were counted each day. Embryos were considered dead if they did not hatch after more than 24 hours and appeared grossly abnormal.

2.12 Cell fate challenge assay

The *hs:: hlh-1* transgene was crossed into *wdr-5, rbbp-5, ash-2, wdr-5; rbbp-5, ash-2* and *lin15B* strains. Adult hermaphrodites were bleached to harvest embryos. P1-P3 stages embryos were selected by using a mouth pipette and transferred to 100 μl of M9 buffer. These embryos were incubated at 20°C for 75 min, and the temperature was increased at 0.1°C/s to 33°C. Embryos were incubated at 33°C for 30 min, and then cool down to 20°C for overnight. Follow by a myosin staining as described above.

2.13 Cloning

Generation of the *cdk-9* (RNAi) feeding vector was performed by the traditional cloning technology. Initial PCR fragment was amplified from the cDNA, which was generated from mRNA of N2 worms.

2.14 Preparation of the *C. elegans* cDNA

Total RNA was extracted from N2 worms using the Trizol Plus RNA Purification System (Invitrogen) following the manufacturer’s instructions. cDNA was synthesised from this RNA using the Cloned AMV First Strand cDNA Synthesis Kit (Invitrogen) according to the instructions.

2.15 PCR amplification of insert *cdk-9* fragment

PCR was performed to amplify the *cdk-9* fragment using specific primer pairs (Table 2). Reaction was carried out in a 50 μl volumes, containing 1x Bioline NH₄ reaction buffer, 2.5 mM MgCL₂, dNTP, BioTaq (Bioline) and 2.5 μl of each primer. Following 2 min at 96 °C, 32 cycles
were performed of 96 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min, and incubated at 72 °C for 1 h to improve the A-tailing. Product was kept at -20 °C.

2.16 Generation of RNAi feeding vector

The cdk-9 (RNAi) feeding vector were generated by ligation at 4 °C overnight, using 0.6 U of T4 DNA ligase in 1x T4 DNA ligase buffer (Roche) and 1:5 molar ration of T-tailed L4440 vector to PCR product.

L4440 vector carrying cdk-9 fragment was transferred into DH5α competent bacteria cells using a standard heat shock transformation protocol. Briefly, the DH5α competent cells were incubated with L4440 vector on ice for 15 min, heat shocked at 42 °C for 45 s. Left to recover on ice for up to 5 min, incubated with 300 μl of LB medium for 1 h at 37 °C with shaking, plated onto 9 cm plates containing LB agar (0.2 M NaCl, 1% (w/v) bacto-peptone, 0.5% (w/v) agar) and 100 μg/ml ampicillin, and incubated at 37 °C overnight.

Bacterial colonies were screened for the insert by clony PCR using the T7 primers, and a single colony was grown overnight at 37 °C with shaking in 5 ml LB with 100 μg/ml ampicillin. From this culture, stock was made and stored at -80 °C in LB glycerol (20%(v/v) glycerol in LB medium) and plasmid DNA was purified from the cell pellet using the QIAprep Spin Miniprep Kit, following the manufacturer’s instruction. The L4440 vector carrying cdk-9 fragment was verified by sequencing by GATC Biotech using the specific primers.

2.17 cdk-9 (RNAi) by feeding (ama-1 and cdk-9)

RNAi by feeding was performed by using HT115 (DE3) bacteria, which was transformed with the L4440 vectors containing a fragment corresponding to a target gene. Bacterial feeding strain of ama-1 was obtained from the Ahringer library. The cdk-9 (RNAi) feeding clone was generated by cloning technology from cDNA of N2 worms (Kamath and Ahringer, 2003).

Bacterial feeding strains were re-streaked on Omnitray flat plates (Nunc) containing LB agar, 15 μg/ml tetracycline and 25μg/ml carbenicillin, and incubated at 37 °C overnight. A single clone of bacteria was grown in LB medium with 100 μg/ml ampicillin for 6 hour at 37 °C.
shaking incubator. Then bacteria were spotted into six-well plates (Greiner) containing NGM agar, 1mM IPTG, 50units/ml nystatin and 25µg/ml carbenicillin. Bacteria carried empty vector was used as a control and followed exactly same protocol.

RNAi six-well plates were labelled before starting the experiment (Figure 9). Around 20 synchronized N2 L1-L2 stage worms were placed in the first well and maintained at 20°C. After these worms reached adulthood, they were transferred to the second well and maintained at 20°C. 24 hours later, they were transferred to the well 3. All experiments were duplicated in the lower wells.

Once all the worms had reached adulthood (around 48 h), a sterile phenotype was observed in the ama-1 and cdk-9 (RNAi) experiment. In the last well, significant different in brood size should be observed between cdk-9 (RNAi) group and empty vector (EV) group.

An embryonic immunofluorescence against pSer2 and pSer5 was performed to confirm the knock-down efficiency following the RNAi experiment. In the immunofluorescence, level of pSer2/5 was significantly reduced comparing to the EV (RNAi) group, indicating that the RNAi was working properly.

![Figure 9. RNAi by feeding](image-url)
Chapter 3

Lineage specific trimethylation of H3 on lysine 4
during C. elegans early embryogenesis

Siyao Wang, Kate Fisher, Gino B. Poulin*

Faculty of Life Sciences, Michael Smith Building, The University of
Manchester, Oxford Road, Manchester, M13 9PL, UK

Accepted Journal:

Developmental Biology 355 (2011) 227–238
ABSTRACT

In many organisms early embryogenesis is characterised by a period refractory to transcription. In *Caenorhabditis elegans*, the one-cell embryo is transcriptionally inactive, but at around eight-cell stage transcription is activated in the somatic lineage. This model suggests that histone tail modifications associated with activation of transcription, such as di- or trimethylation of histone 3 on lysine 4 (H3K4me2/me3) should be enriched in the somatic lineage. Here, we have investigated the deposition of H3K4me3 during embryogenesis and found that it is more dynamic than anticipated. In the eight-cell stage embryo, H3K4me3 deposition is poor in the germline blastomere, as expected, but surprisingly three somatic blastomeres also remain poor in H3K4me3. All the other somatic blastomeres show robust deposition of H3K4me3. Interestingly, the three somatic blastomeres poor in H3K4me3 are descendants of the first germline blastomere, implying an activity that impedes on H3K4me3 deposition in these cells. In contrast, the deposition of H3K4me2 and H3K27me2/me3 is not lineage restricted. Taken together, our data reveal that H3K4me3 deposition is highly regulated according to the cell lineage involved.
3.1 Introduction

One of the most fascinating processes of early embryogenesis is the production of cells with two fundamentally different programs: the somatic program, which produces cells that are destined to die, and the germline program responsible for immortality of the species. Hence, the germline blastomeres face a dilemma at each cell division: to retain totipotency or to engage the somatic program.

Transcription is highly regulated during early embryogenesis and is important to establish the distinction between somatic and germline fates (Mello et al., 1992, 1996; Seydoux and Strome, 1999). Transcription transits rapidly from a generally inactive state in all lineages to an active state specifically in the somatic lineage (Edgar et al., 1994; Seydoux and Dunn, 1997; Seydoux and Fire, 1994). In contrast, the germline lineage remains poor in transcriptional activity.

An important determinant of somatic versus germline lineage is PIE-1 (Mello et al., 1992). PIE-1 has been shown to segregate in the P lineage and to disappear when P4 undergoes the final symmetric division (Mello et al., 1996). PIE-1 is an RNA binding protein that can repress transcription in human cell culture (Zhang et al., 2003). Deletion of pie-1 causes a dramatic effect on the embryonic lineage; the P2 lineage loses its germline program and acquires a somatic program, causing the production of extra intestinal cells (Mello et al., 1996; Seydoux and Strome, 1999). Therefore, PIE-1 is essential to suppress the somatic program, possibly by general repression of transcription.

There is strong evidence that chromatin organisation is important for germline development. Work on MES-2, MES-3, and MES-6, the Polycomb Group-like complex in *Caenorhabditis elegans*, has shown that these prevent degeneration of the germline and sterility through deposition of H3K27me2/me3 marks (Bender et al., 2004; Capowski et al., 1991; Shin and Mello, 2003). It was also shown that defects in erasure of the H3K4me2 marks in the Z2 and Z3 cells compromises germline immortality (Katz et al., 2009; Schaner et al., 2003; Schaner and Kelly, 2006). Furthermore, inactivation of members of the NuRD (nucleosome remodelling deacetylase) complex, a chromatin remodelling complex involved in repression of transcription, causes the ectopic expression of germline markers in the soma (Unhavaithaya et al., 2002).
Therefore, chromatin organisation is critical to retain the germline program and also to prevent its expression in the somatic lineage.

Di- or trimethylation marks on H3K4 and H3K27 correlate with activation of transcription and with repression of transcription, respectively. Consequently, H3K4me2/me3 and H3K27me2/me3 are mainly found at different loci (Cao et al., 2002; Ringrose and Paro, 2004; Ruthenburg et al., 2007). Despite this anti-correlation, a limited number of loci have been found co-occupied by H3K4/K27me3 marks. These loci tend to be transcriptionally inactive and to encode for developmentally regulated transcription factors ready to be activated at the appropriate time (Azuara et al., 2006; Bernstein et al., 2006).

Methylation marks are deposited and removed by enzymes. A number of these enzymes are part of the MLL3/SET1 complex in humans, and we will refer herein to its C. elegans counterpart as the MLL complex. The MLL complex has a methyltransferase activity targeting H3K4 and a demethylase activity targeting H3K27 (Agger et al., 2007; Fisher et al., 2010; Issaeva et al., 2007; Lan et al., 2007; Lee et al., 2007). The complex can therefore deposit methyl marks on H3K4 (activation of transcription) and remove methyl marks from H3K27 (relieving repression of transcription) (Kouzarides, 2007). The MLL complex requires core components to be effective: WDR-5, ASH-2, and RBBP-5 (Ruthenburg et al., 2007). The MLL complex can also contain at least two methyltransferases (SET-16 or SET-2) (Fisher et al., 2010; Simonet et al., 2007). All these components are conserved in humans.

Here, we investigated the levels of di- and trimethylation at both H3K4 and H3K27 sites during early embryogenesis. Since these marks correlate with either activation or repression of transcription, we anticipated that the germline blastomeres would be poor in H3K4me2/me3 and rich in H3K27me2/me3 relatively to the somatic blastomeres. Interestingly, only the H3K4me3 mark shows levels of deposition that differ between the different lineages. Indeed, the germline blastomeres and importantly some of the somatic blastomeres remain poor in H3K4me3. We also show using mutants that the MLL complex is active at an early stage in both the somatic and germline blastomeres. Interestingly, both PIE-1 and RBR-2 (a demethylase targeting H3K4) can affect H3K4me3 levels during early embryogenesis, but only PIE-1 can prevent H3K4me3 deposition in the germline blastomeres. Taken together, our data
show that the deposition of the H3K4me3 mark is unique and dynamic compared to other epigenetic marks investigated here or reported by others.

### 3.2 Materials and Methods

#### 3.2.1 Strains and general maintenance

Strains were maintained as previously described (Brenner, 1974). The strains used in this study were: RB1304 wdr-5(ok1417), RB1025 set-2 (ok952), OL0058 rbbp-5(tm3463), PFR403 ash-2(tm1905), ZR1 rbr-2(tm1231), JJ532 pie-1(zu154) unc-25 (e156)/qC1 dpy-19(e1259) glp-1(q339) and Bristol N2 as wild type (Table 2).

#### 3.2.2 Co-immunofluorescence

Immunofluorescence by freeze crack was performed on polylysine treated slides at 0.3%. Slides treated by application of 75 μl of polylysine 0.3%, dried 10 min at 70 °C, quickly rinsed in distilled water and excess liquid wiped off. The slides were 3 × 14 mm printed wells from Fisher Scientific LTD UK. About 30 mothers were placed in a well with a drop of M9 buffer, to wash off bacteria. These mothers were transferred into a 5-6 μl drop of M9 onto the polylysine treated well using an eyelash. A cover slip (22 mm × 50 mm) was applied at a right angle and the slide placed at −80 °C for at least 20 min. The cover slip was then promptly removed and embryos methanol fixed at −20 °C for 10 min, washed 5 min in PBS, and then followed by two washes in PBS-tween 0.2%. Primary antibody was incubated overnight at 4 °C in a humid chamber. Washes performed as described above. Secondary antibody was incubated 2 h at 37 °C. Washes performed as described above. Mowiol was applied to preserve fluorescence. Primary antibodies used: anti-H3K4me3 (Abcam, ab1012 and ab8580), anti-H3K4me2 (Abcam, ab32356), anti-H3K27me3 (Millipore/Upstate, 07-449), anti-H3K27me2 (Millipore/Upstate, 07–452), and anti-H3 (Abcam, 1791). The secondary antibodies used are from Jackson ImmunoResearch: DyLight 594 AffiniPure Goat Anti-Mouse IgG (H+L) (cat.115-515-146) and DyLight 488 AffiniPure Goat Anti-Rabbit IgG (H + L) (cat.111-485-144). Our controls performed without primary antibody show that these secondary antibodies do not produce significant background (data not shown).
3.2.3 Image processing
Confocal images were analysed using imageJ software. Individual slices were selected for
every cell to discriminate between over-lapping cells. We used eight-cell stage embryos in
which AB descendants are distinctive from P1 descendants, i.e. the AB cells enter mitosis
faster than the P1 lineage. We expressed the results as a ratio of the H3K4me3 signal over the
H3K27me2 signal for each cells.

3.2.4 Peptide competition assay
Peptides were pre-incubated in three folds excess with the anti-H3K4me3 (ab1012) at room
temperature for 30min prior to immunostainings. Immunostainings were performed as
described above. The peptides used were produced by Abcam (ab1432, H3K4me3) and
(ab1782, H3K27me3).

3.2.5 Western blot analysis
Western blot analysis was performed as described in (Fisher et al., 2010) with a few
modifications. Briefly, embryo protein extracts were prepared by bleaching young mothers
followed by washing the embryos four times with M9 buffer. A minimum volume of 20 μl
pelleted embryos was collected for each protein sample, boiled in Laemmli buffer containing
100 mM DTT and sonicated.

3.2.6 Embryonic and fertility analysis
To calculate brood size and embryonic lethality ten clones per strain were transferred onto
fresh plates every day until the worms had finished laying. The number of embryos and
progenies were counted each day. Embryos are considered dead if they did not hatch after
more than 24 h and appeared grossly abnormal.
3.3 Results

3.3.1 H3K4me3 in early embryo and on chromosome X

The *C. elegans* early embryo is a good model to study how the somatic lineage differs from the germline lineage, because these lineages can be tracked *in vivo*. During early embryogenesis, the lineage that will generate the gametes is called the P lineage (P0, P1, P2, P3, and P4). P0-P3 originate from four successive asymmetric cell divisions that also give rise to four somatic blastomeres (or founder cells) (Sulston et al., 1983). P4 will later divide symmetrically to produce the primordial germ cells (PGCs), Z2 and Z3 (Figure 10 A and B). At that stage, the germline programme is distinct from the somatic programme (Schaner and Kelly, 2006; Seydoux and Strome, 1999).

It was previously shown that transcription in two- and four-cell stage embryos is poor (Edgar et al., 1994; Seydoux and Fire, 1994). Therefore, we investigated whether H3K27me2/me3, the repression-associated marks, and H3K4me2/me3, the activation-associated marks, are detectable by immunofluorescence at these stages. We found that H3K4me2/me3, and H3K27me2/me3 are detectable marks and appear evenly distributed between the different cells of the embryos at both two- and four-cell stages (Figure 10 C, data not shown, and Schaner et al., 2003). Therefore, the methylation marks analysed herein are all present during early embryogenesis, and each mark is distributed at comparable levels between cells.

The H3K4me2, H3K27me2, and H3K27me3 marks have been studied during post-embryonic germline development. During germline development the X chromosome is transcriptionally inactive (Reinke et al., 2000). Accordingly, the X chromosome is devoid of the H3K4me2 mark (Kelly et al., 2002; Reuben and Lin, 2002), and enriched for both the H3K27me2/me3 marks (Bender et al., 2004). We predicted that H3K4me3 should adopt a similar localization as H3K4me2, and indeed we found a chromosomal area devoid of H3K4me3 (Figure 10D). This pattern of deposition suggests that the antibody against H3K4me3 recognised the correct epitope. Nevertheless, we additionally performed a peptide competition assay, and show that a peptide trimethylated at K4 can abrogate the H3K4me3 signal in embryos and adult germline, but a peptide trimethylated at K27 has no effect (Figure 10E). Taken together, the pattern of deposition, the competition assays, and other data provided below indicate that this antibody is specific to the H3K4me3 mark.
Figure 10. The H3K4me3 mark during early embryonic development

(A) Depiction of the early embryo lineage. The one cell embryo first divides asymmetrically to generate the AB and P lineages. The AB lineage is entirely formed of somatic cells. The P lineage produces the germline blastomeres (P1, P2, P3, and P4 (green and blue)), the somatic blastomeres (EMS, C and D (green)), and the primordial germ cells (Z2 and Z3 (blue)). The latter will generate all the gamete of the adult hermaphrodite. (B) Depiction of the two-, four-, and eight-cell stages of the embryo. In green are the somatic blastomeres, and in green and blue are the germline blastomeres. (C) The H3K4me3 and H3K27me2 marks are detected in all the cells of the embryos at two- and four-cell stages. Co-immunostaining for H3K4me3, H3K27me3 and DNA by DAPI are shown as well as the corresponding DIC picture. (D) A chromosomic area at pachytene stage is depleted of the H3K4me3 mark. E) Competition assays using peptides as indicated followed by immunostaining provide further evidence that the H3K4me3 antibody is specific. Two-cell stage embryos are presented, but similar data were also obtained at later stages of development and also in the germlines.
3.3.2 Cell lineage specific deposition of H3K4me3 at the eight-cell stage

Transcription has been shown to be active at the eight-cell stage (Edgar et al., 1994). Interestingly, it is at this stage that we observed a striking difference in H3K4me3 deposition. In contrast, the other methylation marks (H3K4me2, H3K27me2, and H3K27me3) remained unchanged. The cells poor in H3K4me3 are the germline blastomere (P3) and three somatic blastomeres: MS, E, and C (Figure 11A). The four somatic blastomeres enriched for H3K4me3 are all AB descendants (Figure 11A). We ensured that these four cells are descendants of the AB lineage, by performing co-immunofluorescence with an antibody against GLP-1. GLP-1 is a Notch homolog found at the membrane and within the cytoplasm. Importantly, it is expressed only in the AB lineage at the early stages of embryogenesis (Evans et al., 1994). At four-cell stage only ABa and ABp express GLP-1, confirming that GLP-1 is specific to the AB lineage (Figure 11B). We next analysed the eight-cell stage and found that only four cells are positive for GLP-1 expression, and that these cells show high levels of H3K4me3 deposition (Figure 11C). This indicates that H3K4me3 deposition is high in the AB lineage. We also performed a co-staining using an antibody against PGL-1, which confirms that levels of H3K4me3 are low in the P3 cell (Supplementary Figure 1). We next quantified the abundance of H3K4me3 relative to H3K27me2 using confocal microscopy (Figure 11D). We analysed 80 cells from ten embryos, and found a 3.3 fold enrichment in H3K4me3 deposition in the AB descendants compared with the P1 descendants (Figure 11E). We have also produced 3D movies to help visualize these data (Supplementary Figure 2). Accordingly, similar results were obtained from another commercially available antibody against H3K4me3 (Supplementary Figure 3). Taken together, these data lead us to conclude that the H3K4me3 mark is regulated according to the cell lineage. Indeed, the somatic blastomeres derived from P1, as well as P3, are resistant to accumulation of H3K4me3.

Since at eight-cell stage we have observed four H3K4me3-depleted cells, we expected that following an additional cell division, eight cells (seven somatic cells and P4) would remain depleted of H3K4me3, but this is not the case. Indeed, only three somatic cells and P4 remain poor in H3K4me3. This implies that all of the descendants of the four H3K4me3-depleted cells are capable of acquiring the H3K4me3 mark (Figure 12A). Accordingly, high levels of H3K4me3 are detected in all somatic cells at the post-gastrulation stage (Figure 12B).
Figure 11. H3K4me3 deposition is enriched in the AB lineage

(A) Eight-cell stage embryo showing that H3K4me3 levels are high in the AB lineage, but H3K27me2 levels are similar regardless of the lineage. The AB lineage is indicated by a dotted line. (B) Co-immunostainings against PGL-1 and H3K4me3 at four-cell stage show that PGL-1 is specific to the AB lineage. (C) Co-immunostainings against PGL-1 and H3K4me3 at eight-cell stage show that H3K4me3 and PGL-1 are detected in the same cells, the AB descendants. (D) A confocal photograph (stacking of multiple slices) shows that H3K4me3 (red) and H3K27me2 (green) only co-localise in half the cells. (E) Quantification using confocal images of co-immunostainings against H3K4me3, H3K27me2 and DAPI. Data are from 80 cells (ten embryos) and four separate experiments. Error bars represent standard error of the mean (SEM).
3.3.3 The MLL complex is responsible for acquisition of H3K4me3 during early embryogenesis

We next addressed whether the deposition of H3K4me3 during early embryogenesis requires the MLL complex. We used mutants of the core components (wdr-5, rbbp-5 and ash-2) as well as a set-2 mutant. SET-2 is one of the two methyltransferases targeting H3K4 that have been shown to be part of the complex (Fisher et al., 2010; Simonet et al., 2007). Initially, we performed co-immunofluorescence at two- and four-cell stages in the wdr-5 mutant. This mutant has already been shown to be impaired in deposition of H3K4me3 (Fisher et al., 2010; Simonet et al., 2007), but not at early stages of embryonic development. We started our analysis at the two-cell stage, and could still detect H3K4me3 deposition, though the signal appeared weaker than in wild type worms (Figure 13A). However, at four-cell stage, we were unable to detect the H3K4me3 mark (Figure 13B). Noteworthy, we found no defect in levels of deposition of H3K27me2 (Figures 13A and B) or of H3K27me3 (data not shown).
Figure 13. WDR-5 is required to generate the H3K4me3 mark

(A) Two-cell stage N2 embryo in which the H3K4me3 and H3K27me2 marks can be detected. (B) Two-cell stage wdr-5 embryo in which the H3K4me3 mark appears diminished, and in which the H3K27me2 marks is not affected detected. (C) Four-cell stage N2 embryo in which the H3K4me3 and H3K27me2 marks can be detected. (D) Four-cell stage wdr-5 embryo showing that the H3K4me3 mark is now undetectable, but H3K27me2 deposition is unaffected.
Additionally, we analysed the four-cell stage and found that the H3K4me3 mark is undetectable in wdr-5, rbbp-5, ash-2 and set-2 mutants compared to N2 (Figure 14A). We performed the same systematic analysis at post-gastrulation stage. We obtained the same result with wdr-5, rbbp-5 and ash-2 mutants, but surprisingly H3K4me3 is detectable in the set-2 mutant (Figure 14B). To ensure that these observations based on immunofluorescence are not due to masking of specific epitopes, we performed western blot analysis using extracts from wdr-5, rbbp-5, and ash-2 mutants and compared levels of H3K4me3 to wild type embryo extracts. We found that in all the mutant cases, H3K4me3 is significantly decreased, which is consistent with our immunofluorescence data (Figure 14C). We conclude that the MLL complex is required for appropriate deposition of H3K4me3 during embryogenesis.
Figure 14. Mutants of the MLL complex show reduced levels of H3K4me3

(A) The core MLL complex, rbpp-5, and wdr-5, as well as the methyltransferase set-2, are required for deposition of H3K4me3 at the eight-cell stage of embryogenesis. (B) The core MLL complex, rbpp-5, and wdr-5, as well as the methyltransferase set-2, are required for deposition of H3K4me3 at post gastrulation. (C) Western blot analysis against H3K4me3 using different antibodies shows that embryonic extracts from wdr-5, rbpp-5, and ash-2 have reduced levels of H3K4me3.
3.3.4 H3K4me3 deposition remains low in germline blastomeres from rbr-2 mutants

Our observations are consistent with a dynamic and lineage specific regulation of H3K4me3 deposition during embryogenesis, in which the MLL complex must be playing an important role in cells rich in H3K4me3. However, poor deposition of H3K4me3 could be explained by both reduced MLL activity and increased demethylase activity. Two types of demethylases targeting H3K4 have been described in C. elegans: LSD1-like and RBR-2 (Christensen et al., 2007; Cloos et al., 2008; Greer et al., 2010; Katz et al., 2009). The LSD1 ortholog, SPR-5, can demethylate H3K4me2/me1 and has been shown to be active in the PGCs (Katz et al., 2009). On the other hand, RBR-2 can act on H3K4me3/me2, which makes RBR-2 a good candidate, since it can target H3K4me3 (Christensen et al., 2007). To test whether demethylation by RBR-2 contributes to the H3K4me3 pattern of deposition, we used an rbr-2 mutant, and analysed its effect at multiple stages, including the eight-cell stage. At this stage, we found that the rbr-2 embryos deposit H3K4me3 in a pattern similar to wild type animals (Figure 15A, eight-cell stage). Importantly, we could not detect H3K4me3 deposition in the P4 or later in Z2 and Z3, the PGCs (Figure 15A, P4 and Z2-Z3). However, when we compared the H3K4me3 levels between the rbr-2 mutant and wild type, we observed an enrichment of H3K4me3 in the somatic cells (Figure 15B). We quantified this observation using western blot analysis, and detected about four times the level of H3K4me3 in rbr-2 embryos compared with wild type, with no effect on H3K27me3 deposition (Figure 15C). Therefore, RBR-2 is not involved in preventing H3K4me3 deposition in the P lineage, but could play a role in regulating somatic deposition of H3K4me3.
Figure 15. H3K4me3 deposition remains low in germline blastomeres in rbr-2 mutants

(A) H3K4me3 deposition is not affected in germline blastomeres of rbr-2 mutant embryos. (B) Comparison between rbr-2 and N2 embryos suggests an elevated deposition of H3K4me3 in rbr-2 embryos. (C) Comparison between rbr-2 and N2 embryos using western blot analysis shows an elevated deposition of H3K4me3 in rbr-2 embryos. K4_r indicates blotting using the rabbit anti-H3K4me3 (ab8580), H3K4_m, the mouse anti-H3K4me3(ab1012), K27 a rabbit anti-H3K27me3, and H3 a rabbit anti-H3.
3.3.5 PIE1 inhibits deposition of H3K4me3 in P lineage

PIE-1 has been shown to be critical to maintain the germline fate through repression of transcription in the P lineage (Mello et al., 1992; Mello et al., 1996; Seydoux and Strome, 1999; Zhang et al., 2003). Herein, we have shown that deposition of H3K4me3, an activation mark, is low in the P lineage. Hence, we tested whether the absence of PIE-1, a repressor of transcription, could promote the deposition of H3K4me3 in the P lineage, and the somatic descendants of P1. Using a pie-1 mutant and performing similar analysis as presented above, we did not observe an effect on the deposition of H3K4me3 at two- and four-cell stages (Figures. 16A). However, we found that at the eight-cell stage, the embryos depleted of PIE-1 acquire the H3K4me3 mark in germline blastomeres and in the somatic cells in which H3K4me3 is normally poor (Figure 16B). Strikingly, even at later stages of development, we observed deposition of H3K4me3 in PGCs, Z2 and Z3 (Figure 16C). Therefore PIE-1 is critical to prevent accumulation of H3K4me3 in germline blastomeres and in the somatic descendants of P1.
Figure 16. H3K4me3 deposition is activated in germline blastomeres in pie-1 mutants

(A and B) Immunostainings at two- and four-cell stages with pie-1 embryos showing that H3K4me3 deposition appears normal. (C) An eight-cell stage pie-1 embryo showing that the H3K4me3 mark can be detected in all cells, including P3. (D) A later stage of a pie-1 embryo showing that H3K4me3 is present in P4.
3.3.6 H3K4me3 deposition is robust in the germline

The effect that the *wdr-5* mutant has on levels of H3K4me3 in embryo is striking (Figure 14). But surprisingly, previous studies have shown that L4 or adults express levels of H3K4me3 that can easily be detected by Westerns blot analysis (Fisher et al., 2010; Simonet et al., 2007). One explanation (other than the sensitivity of the methods) for these results is that the germ cells of the *wdr-5* animals maintain robust deposition of H3K4me3. Hence, we tested this possibility by performing immunostaining in the gonads of *wdr-5* mutants. We found that in the wild type gonads, the H3K4me3 pattern of deposition is dynamic. It is at a lower level in the mitotic zone, but at a higher level in the other zones (Figure 17A). In contrast, the differential distribution of the H3K27me2 modification is similar between the different zones, in particular if the DAPI signal is taken into consideration (Figure 17A). Importantly, in the *wdr-5* gonads, the H3K4me3 mark is detectable, but its deposition pattern has changed (Figure 17B). It appears that the zone normally low in H3K4me3 levels is extended in *wdr-5* mutants (Figure 17B). This may explain the source of H3K4me3 detected by western blot analysis in *wdr-5* animals (Fisher et al., 2010; Simonet et al., 2007), and also suggests that H3K4me3 deposition in the germline is more robust than in the embryo.
Figure 17. Germline deposition of H3K4me3 in N2 and wdr-5 mutants

(A) Wild type germlines showing that the deposition of the H3K4me3 mark is low level in the mitotic zone. The H3K27me2 mark is evenly distributed relative to DAPI staining. (B) The wdr-5 mutant can still deposit H3K4me3 in the germline, but the zone in which the levels are low is extended.

Importantly, our data from embryos and the germline imply that embryonic viability and fertility could be affected in mutants of the core MLL complex. Hence, we assessed these functions in the wdr-5, rbbp-5 and ash-2 mutants. We found that embryonic viability and fertility for all mutants is compromised at both 20°C and 25°C. But surprisingly, these functions are affected differentially. For example, the wdr-5 mutant produces 45% embryonic lethality at 25°C compared with 16% for rbbp-5 (p<0.005) and 0% for ash-2 (p<0.00005) (Table 3). On the other hand, brood size is more affected in ash-2 (6 progeny, p<0.0005) or rbbp-5 (19 progeny,
p<0.001) than in wdr-5 (54 progeny) (Table 3). Perhaps this is an indication that the MLL core complex can associate with different partners according to the site of expression. Despite this extra layer of complexity, our data support the view that proper embryonic development and fertility requires deposition of methyl groups on H3K4.

*p<0.01, **p<0.005, ***p<0.0001, ****p<0.000001

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>% emb. leth.</th>
<th>aver. brood size</th>
</tr>
</thead>
<tbody>
<tr>
<td>N2, 20 °C</td>
<td>0</td>
<td>204</td>
</tr>
<tr>
<td>N2, 25 °C</td>
<td>0.5</td>
<td>149</td>
</tr>
<tr>
<td>wdr-5, 20 °C</td>
<td>4**</td>
<td>111**</td>
</tr>
<tr>
<td>wdr-5, 25 °C</td>
<td>45***</td>
<td>54***</td>
</tr>
<tr>
<td>rbbp-5, 20 °C</td>
<td>2</td>
<td>44***</td>
</tr>
<tr>
<td>rbbp-5, 25 °C</td>
<td>16</td>
<td>19****</td>
</tr>
<tr>
<td>ash-2, 20 °C</td>
<td>5*</td>
<td>38****</td>
</tr>
<tr>
<td>ash-2, 25 °C</td>
<td>0</td>
<td>6****</td>
</tr>
</tbody>
</table>

Table 4. Embryonic viability and fertility analysis of mutants of the MLL core complex
3.4 Discussion

Transcriptional activity during early embryogenesis is poor in the germline blastomeres, but processive in the somatic blastomeres (Edgar et al., 1994; Schaner et al., 2003; Schaner and Kelly, 2006; Seydoux and Dunn, 1997; Seydoux and Fire, 1994; Seydoux and Strome, 1999). Previous studies have investigated the deposition of methylation and acetylation marks during early embryogenesis and found that H3K4me2 is actively removed from the PGCs later during embryogenesis (>26-cell stage). But prior to that stage, the levels of H3K4me2 (and other marks) remain equivalent in all cells of the early embryo (Katz et al., 2009; Schaner et al., 2003; Schaner and Kelly, 2006). Similarly, we found that H3K4me2 and H3K27me2/me3 levels remain equivalent prior to the generation of PGCs. However, we uncovered that the H3K4me3 mark has a unique distribution during embryogenesis. Indeed, some of the somatic blastomeres and the germline blastomeres fail to accumulate the H3K4me3 marks. Interestingly, this characteristic is regulated by PIE-1. Therefore, it is likely that deposition of the H3K4me3 mark is important for the development of the germline blastomeres.

Prior to this study there was no evidence that the chromatin architecture of germline blastomeres could be different from somatic blastomeres. Indeed, most data indicated that, at least at a superficial level, both types of blastomeres were equivalent in deposition of numerous methylation and acetylation marks thereby implying that their chromatin architecture was similar. Here, we provide new insights into a possible role that a specific mark (H3K4me3) may play to ensure that the chromatin architecture adopted in the germline blastomeres is in accordance with its transcriptional status. Consistent with this role, we found that in absence of PIE-1, deposition of H3K4me3 is restored in germline blastomeres, and that the differential deposition between somatic blastomeres at the eight-cell stage is abolished. Therefore, these data imply interplay between the general transcriptional repressor PIE-1 and the chromatin architecture of the germline blastomeres.

The observation that specific somatic blastomeres fail to accumulate the H3K4me3 mark was unexpected. Interestingly, the somatic blastomeres ‘resistant’ to the deposition of the third methylation mark on H3K4 are derived from the germline blastomere P1. Hence, we would like to raise the possibility that the P1 germline blastomere produces an activity that prevents the MLL complex to deposit the third methylation mark on H3K4 or/and that it stimulates its
removal by a still undiscovered demethylase. However, this putative activity from P1 may not affect directly the MLL complex or the demethylase, but could affect the deposition or removal of other marks on the chromatin leading to the effect characterised herein. Our data appear to rule out RBR-2 as a potential demethylase, since it is not required to maintain the H3K4me3 deposition pattern at the eight-cell stage (Figure 15). Finally, considering that the germline blastomere is a type of stem cell, our findings may provide new insights into how totipotency (and perhaps also pluripotency) can be maintained by controlling the deposition of the third methyl mark on H3K4.
Supplementary Fig 1. Co-immunostaining at eight-cell stage against H3K4me3 and PGL-1 to ensure that the PGL-1 positive cell is negative for H3K4me3.

Supplementary Fig 2. 
(Movie is available online http://www.sciencedirect.com/science/article/pii/S0012160611002429) 3D Movie that sequentially animates the immunostaining results against H3K4me3 (red), H3K27me2 (green), and DAPI (blue) in a representative eight-cell stage embryo.
Supplementary Fig 3. Immunostaining using a different commercially available antibody against H3K4me3 (ab8580). Representative embryos at two-, four-, eight- and >16-cell stage are shown. These results are similar to the results obtained with the anti-H3K4me3 (ab1012).
Supplementary Fig 3. Immunostaining using a different commercially available antibody against H3K4me3 (ab8580). Representative embryos at two-, four-, eight- and >16-cell stage are shown. These results are similar to the results obtained with the anti-H3K4me3 (ab1012).
3.5 References


Chapter 4

The *C. elegans* MLL/SET/COMPASS complex core component RBBP-5 can antagonise accumulation of H3K4me2/me3 in primordial germ cells

Siyao Wang, Kate Fisher, Gino B. Poulin*

Faculty of Life Sciences, Michael Smith Building, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK

Intended Journal: Developmental Biology
ABSTRACT

Post-translational modifications on histone tails are crucial to determine the transcriptional state of cells. Methylation on histone 3 at lysine 4 (H3K4) is a conserved modification required for establishing and maintaining an active transcriptional state. WDR-5, RBBP-5, and ASH-2 are core components of the MLL/SET/COMPASS complex necessary for methylation at H3K4. These three core components physically interact to form a Y-shaped structure, which anchors the catalytic part of methyltransferases at its centre. This structural architecture suggests that the removal of any components should dramatically affect the deposition of methyl marks at H3K4. We systematically examined this prediction during embryonic and larval development. We found that the removal of each component affects the stability of the complex differentially according to the methyl mark and the developmental stage involved. But surprisingly, we found that RBBP-5 can antagonise H3K4me2/me3 accumulation in primordial germ cells of starved young larvae (L1). This suggests that RBBP-5 plays a role in sensing nutrient availability in the environment of newly hatched embryos. Taken together, we present a unique developmental study of a metazoan system that systematically examined the methylation levels at H3K4 and the contribution of each core component towards this process. Moreover, this work revealed a novel function for RBBP-5 that may be linked to the nutrient sensing response that controls whether young larvae will continue or arrest their development.
4.1 Introduction

Chromatin is central to preserving and regulating access to genetic information. Chromatin is composed of 147 base pairs of DNA wrapped around an octamer of core histones (H2A, H2B, H3, and H4) to form a “beads on a string” structure (Kornberg, 1974; Kornberg and Lorch, 1999). The “bead” in this structure is the nucleosome, the basic repetitive unit of chromatin (Kornberg, 1974; Luger et al., 1997). Structural studies of the nucleosome revealed that histones have a globular part interacting with the DNA and an unstructured N-terminal tail protruding outwards (Alva et al., 2007). The N-terminal tail is subjected to post-translational modifications (PTMs) such as acetylation, phosphorylation, ubiquitylation, and methylation (Berger, 2002; Rea et al., 2000). It is thought that these PTMs can form a code or a signalling platform that can control the accessibility to DNA (Berger, 2002; Shilatifard, 2012).

Methylation at lysine 4 on Histone 3 (H3K4) is conserved from yeast to humans. H3K4 methylation comes in three different forms: mono-, di-, and trimethyl (H3K4me1, me2, and me3) (Ruthenburg et al., 2007; Strahl and Allis, 2000; Strahl et al., 1999). Each methylation form has its own pattern of deposition and functional significance: H3K4me1 marks enhancers (Herz et al., 2012); H3K4me2 is found at active enhancers and promoters (Bernstein et al., 2005); and H3K4me3 peaks at transcription start sites (Bernstein et al., 2005; Ng et al., 2003a; Santos-Rosa et al., 2002). Moreover, H3K4me3 generates, when coupled with H3K27me3, bivalent domains producing a poised state in which developmentally regulated genes await a signal to resolve their transcriptional state (Bernstein et al., 2006; Hu et al., 2013b; Mikkelsen et al., 2007). Interestingly, readers of methylation states at H3K4 tend to associate with components of the basal transcriptional machinery, indicating that H3K4 methylation could be a powerful mean to interpret the underlying genomic information (Lauberth et al., 2013).

Methylation at H3K4 is deposited by subgroups of the MLL/SET/COMPASS complex each comprised of unique histone methyltransferases (HMTs) associated with the core complex components WDR-5, ASH-2, RBBP-5 and DPY-30 (Avdic et al., 2011; Dou et al., 2006; Patel et al., 2009; Takahashi et al., 2011). The structural configuration of the core complex is known and indicates that each core complex component should contribute similarly to the deposition of methyl marks at H3K4. However, a study has shown that C. elegans ASH-2 has an inhibitory activity towards H3K4me2 in embryonic and adult germ cells (Xiao et al., 2011).
Raising the possibility that other core complex components could also have additional and specific activities that still remain to be revealed.

We tested this possibility using the *C. elegans* embryo system as well as the Primordial Germ Cells (PGCs) of L1 (larval stage 1; see Figure 18). Transcription is highly regulated during *C. elegans* embryogenesis. In the one-cell stage embryo, transcription is inactive but as the embryo divides transcription becomes active in the somatic lineage, whilst remaining mostly quiescent in the germline lineage (Batchelder et al., 1999; Seydoux et al., 1996; Seydoux and Strome, 1999). We, and others, have shown that H3K4me2/me3 levels become elevated in the somatic lineage when zygotic transcription transits from generally inactive to fully active (Li and Kelly, 2011; Wang et al., 2011). We have also shown that the lack of H3K4me3 in the germline lineage depends on PIE-1 (Wang et al., 2011), a general repressor of transcription (Batchelder et al., 1999; Ghosh and Seydoux, 2008; Mello et al., 1992; Mello et al., 1996; Tenenhaus, 2001). Regulation of H3K4me2/me3 in the PGCs of L1 is a very dynamic process; initially embryonic PGCs are low in H3K4me2/me3 level, these levels remain low in starved L1, but in response to nutrients, the levels of H3K4me2/me3 increase rapidly (Schaner et al., 2003). Thus, the embryo and L1’s PGCs coupled to immunofluorescence are systems offering an excellent opportunity to study, at cellular resolution, how the core complex components (WDR-5, ASH-2 and RBBP-5) contribute towards the dynamic deposition of all three H3K4 methyl marks during development.

Using these systems and genetic deletions predicted to be null alleles for *wdr-5*, *rbbp-5* and *ash-2*, we examined the deposition of H3K4me1/me2/me3. Interestingly, we found that RBBP-5 has an antagonistic activity towards H3K4me2/me3 deposition solely in starved L1s, hence contrasting with ASH-2’s antagonistic activity previously described (Xiao et al., 2011). Taken together, these results suggest a novel mechanism by which nutrients’ absence or presence would regulate RBBP-5’s activities and hence connect nutrient-sensing and chromatin organisation.
Figure 18. Depiction of the embryogenesis process in *C. elegans*

In our study, we divided the embryogenesis into six stages, according to the generation of P cells (germline blastomeres (P0-P4) and PGCs (Z2 and Z3)). During embryogenesis, the zygote, P0, goes through four successive asymmetric divisions to generate four germline blastomeres (P1, P2, P3 and P4). At the 100-cell stage, the germline blastomere, P4, symmetrically divided into two PGCs, Z2 and Z3. P cells are in orange, and somatic cells are in grey. On the right side, the six studied stages of embryogenesis are depicted, the number of cells is indicated, and the P cells and somatic cells are in dark and light grey, respectively.
4.2 Materials and Methods

4.2.1 Strains and general maintenance
Strains used in this study were maintained as previously described (Brenner, 1974a). The strains are: **RB1304 wdr-5 (ok1417)**, **OL0058 rbbp-5 (tm3463)**, **PFR403 ash-2 (tm1905)**, and Bristol N2 as wild type (Table 2).

4.2.2 Co-immunofluorescence of embryos
Immunofluorescence by freeze crack was performed on polylysine treated slides at 0.3%. Slides treated by application of 75μl of polylysine 0.3%, dried 10 min at 70 °C, quickly rinsed in distilled water and excess liquid wiped off. The slides were 3 × 14 mm printed wells from Fisher Scientific LTD UK. About 30 mothers were placed in a well with a drop of M9 buffer, to wash off bacteria. These mothers were transferred into a 5-6μl drop of M9 onto the polylysine treated well using an eyelash. Embryos were dissected from these mothers by a syringe needle. A cover slip (22 mm × 50 mm) was applied at a right angle and the slide placed at −80 °C for at least 20 min. The cover slip was then promptly removed and embryos methanol fixed at −20 °C for 10 min, washed 5 min in PBS, and then followed by two washes in PBS-tween 0.2%. Primary antibody was incubated overnight at 4 °C in a humid chamber. Washes performed as described above. Secondary antibody was incubated 2 h at 37 °C. Washes performed as described above. Mowiol was applied to preserve fluorescence.

4.2.3 Co-immunofluorescence for L1
For larvae staining, adult hermaphrodites of each mutant were washed off from plates with M9 buffer and dissolved in bleaching buffer (1 M NaOH in 10% sodium hypochlorite solution). After two additional M9 washes, the embryos were collected in 500 μl M9 buffer. After 48h, put a drop of starved L1 larvae (around 100 larvae) on a poly-lysine treated well. Apply the same fixation and staining as mentioned above. The remaining L1 larvae were fed *E. coli* (OP50 strain) for 30min before fixation and staining.

4.2.4 Antibodies
Primary antibodies used: anti-H3K4me3 (Abcam ab1012, 1:100), anti-H3K4me2 (Millipore CMA303, 1:5000), anti-H3K4me1 (Abcam ab8895, 1:500000), and anti-PGL-1 (a gift from Judith Kimble, 1:100). The secondary antibodies used are from Jackson ImmunoResearch: DyLight 594 AffiniPure Goat Anti-Mouse IgG (H+L) (cat.115-515-146, 1:200) and DyLight 488 AffiniPure Goat Anti- Rabbit IgG (H + L) (cat.111-485-144, 1:200). Our controls performed without primary antibody show that these secondary antibodies do not produce significant background (data not shown).

4.2.5 Western blot analysis

Western blot analysis was performed as described in (Fisher et al., 2010) with a few modifications. Briefly, embryo protein extracts were prepared by bleaching young mothers followed by washing the embryos four times with M9 buffer. A minimum volume of 20 μl pelleted embryos was collected for each protein sample, boiled in Laemmli buffer containing 100 mM DTT and sonicated.
4.3 Results

To perform our systematic analysis, we divided embryogenesis into six stages according to the generation of germline blastomeres, also called P cells. Fertilisation generates the first P cell, P0. P0 will undergo four successive asymmetric cell divisions producing P1, P2, P3 and P4, and corresponding to 2-, 4-, 8- and 16-cell stages, respectively. As the embryo reaches about 100 cells, P4 divides symmetrically to generate two Primordial Germ Cells (PGCs), called, Z2 and Z3 (Figure 18). The levels of H3K4me2/me3 are low in these PGCs. PGCs remain mitotically and transcriptionally quiescent until the L1 larval stage is reached. However, when the L1s resume feeding, the levels of H3K4me2/me3 increase and mitosis is activated (Schaner et al Dev Cell 2003). We used both the embryogenesis and the switch between starved and fed L1s to delineate the role of each core components play in these developmental systems.

4.3.1 H3K4me1 deposition is most sensitive to the loss of RBBP-5

We first analysed the effect that each core component has on the deposition of methyl marks at H3K4 during embryogenesis (Figure 18). We performed immunofluorescence on embryos and found that levels of H3K4me1 are depleted from the P1 stage in absence of RBBP-5 when compared with wild type embryos (Figure 19). A transient effect altering H3K4me1 levels is also observed in absence of either WDR-5 or ASH-2 from P2 to P4 stages (Figure 19). The levels of H3K4me1 are similar to wild type at the Z2-Z3 stage suggesting that a redundant mechanism is active later during development. Overall, RBBP-5 is the predominant player ensuring that levels of H3K4me1 are maintained throughout embryogenesis.
Figure 19. H3K4me1 deposition is mainly regulated by the RBBP-5 in embryos

Embryos are stained with a rabbit anti-H3K4me1 antibody and DAPI. A strong reduction of H3K4me1 was observed in rbbp-5 mutant. In the wdr-5 and ash-2 mutants, the slight reduction of H3K4me1 was only detected at the P2 and P3 stages (not shown). Exposure time is the same for each condition and the experiments were repeated at least three times.
4.3.2 H3K4me2 deposition is differentially impaired by the loss of WDR-5, RBBP-5 or ASH-2

We next examined the contribution of each core component towards H3K4me2 deposition. In the wdr-5 mutants, a reduction in the H3K4me2 levels could be observed from the P1 till the P4 stage. However, at Z2-Z3 stage, H3K4me2 levels increased at levels comparable to wild type, indicating that a redundancy mechanism is active during late embryogenesis and compensate for the loss of WDR-5. In contrast, absence of RBBP-5 impairs H3K4me2 deposition from P1 till Z2-Z3 stage. Absence of ASH-2 has also a temporal effect on H3K4me2 levels. In the early stages of embryogenesis, P1 and P2 stages, no or little effect on H3K4me2 levels are observed. However, from the P3 stage, H3K4me2 levels are gradually depleted to become almost not detectable later at P4 and Z2-Z3 stages (Figure 20). Taken together, these results indicate that the role of each component in the deposition of H3K4me2 differs from each other. These differences suggest that the deposition of H3K4me2 may rely on different core component activities depending on the stages of embryogenesis.

![Figure 20. H3K4me2 is differentially affected by the loss of of WDR-5, RBBP-5, and ASH-2](image)

Embryos were stained with DAPI and mouse anti-H3K4me2 antibody. The deposition of H3K4me2 is dramatically reduced in the rbbp-5 mutant at P1, P2, P4 and Z2-Z3 stages. However, the wdr-5 mutant impairs H3K4me2 deposition from P2 till P4 stage. The ash-2 mutant impairs H3K4me2 deposition from P4 till the Z2-Z3 stage. Exposure time is the same for each condition and the experiments were repeated at least three times.
4.3.3 H3K4me3 deposition is sensitive to the loss of WDR-5, RBBP-5 or ASH-2

Using the same approach as above, we investigated the core components’ contribution towards the deposition of H3K4me3. We found that the deletion of wdr-5, rbbp-5 or ash-2 greatly affects the levels of H3K4me3, especially following the P2 stage. The P1 and P2 stages still show a detectable reduction in each mutant compared to wild type (Figure 21). Thus, H3K4me3 levels are very sensitive to the loss of any of the core components tested herein.

Figure 21. H3K4me3 deposition is largely affected in wdr-5, rbbp-5 and ash-2 mutants during embryogenesis

Embryos were dissected from the WT and mutants, and probed with mouse anti-H3K4me3 antibody. DNA is co-stained with DAPI. H3K4me3 deposition is depleted from the nucleus of wdr-5, rbbp-5 and ash-2. Exposure time is the same for each condition and the experiments were repeated at least three times.
4.3.4 H3K4 methylation levels in the one-cell embryo are not affected by the loss of WDR-5, RBBP-5, or ASH-2

Using immunofluorescence on wild type embryos, we could detect all three H3K4 methylation forms at all stages. However, we observed that the levels in the one-cells stage embryo are lower than at later stages of embryonic development. We analyzed these relative temporal levels between stages in the wild type embryo using DAPI signal to normalise the H3K4me1, me2, or me3 signals. We found differences in the accumulation of these methyl marks. All three marks levels decrease from P0 to P1 stages, but H3K4me1 and me2 levels increase from P1 to P2 stage to reach stable levels subsequently (Figure 22; N2). On the other hand, H3K4me3 accumulation is delayed and remains lower at P1, P2, and P3 stages than at P0 (Figure 22; N2). H3K4me3 levels then increase at P4 stage to reach stable levels later (Figure 22; N2).

We next analysed this relative pattern of deposition in absence of WDR-5, RBBP-5, or ASH-2. In accordance with the data presented above, RBBP-5 is the most important contributor of all methylation forms at H3K4 (Figure 22; rbbp-5(-)). ASH-2 also plays a predominant role for accumulation of H3K4me3 starting from P1 stage. It also affects H3K4me1/me2 levels, but somehow H3K4me1 levels are compensated and increase at Z2-Z3 stage (Figure 22; ash-2(-)). WDR-5 acts from P1 or P2 stage during embryogenesis on all forms of H3K4 methylation. But H3K4me1 and H3K4me2 effects can be compensated for from Z2-Z3 stage. Taken together, these experiments show that despite forming the core complex each core component has a degree of specificity or at least a preferential activity especially towards H3K4me1 and me2, since H3K4me3 levels are dramatically affected by the loss of any of the core components and never compensated for.
Figure 22. Line charts of H3K4 methylation level in WT (N2), wdr-5(-), rbbp-5(-) and ash-2(-)

The line charts of the H3K4me1 (blue), H3K4me2 (red) and H3K4me3 (green) level at different stages of wild type and mutants. We quantified the abundance of H3K4me1/me2/me3 relative to DAPI using immunofluorescences. For each stage of embryos, 6-8 cells were analysed. The results were expressed in a percentage of the H3K4me signal level over the DAPI signal level. Compare to the wild type, all mutants are impaired in H3K4me1/me2/me3 deposition to different degrees.

4.3.5 Western blots confirmed that H3K4me1/me2/me3 levels are reduced in absence of WDR-5, RBBP-5, or ASH-2

Our immunofluorescence analysis has shown that deposition of methyl marks at H3K4 is altered to varying degrees in each mutant of the core complex. We used Western blot analysis to ensure that our immunofluorescence analyses were not biased by changes in chromatin architecture. We prepared mixed stage embryo extracts from wdr-5, rbbp-5 and ash-2 mutants, and compared the H3K4me1/me2/me3 levels to wild type (Figure 23A and B). We found a striking depletion of all forms of H3K4 methylation in absence of RBBP-5 or ASH-2 (Figure 23A and B). Depletion WDR-5 also reduces the levels of H3K4me1/me2/me3 levels but to a lesser degree for H3K4me1 and H3K4me2 levels. Even though temporal information were lost in these Western blot experiments, the results is still broadly in accordance with the immunofluorescence experiments (Figure 22). Of note, when exposure time was increased in these Western and immunofluorescence experiments, we observed the presence of residual methylation of all three forms at H3K4 in all three mutants (data not shown).
Figure 23. Western blot analysis of the H3K4 methylation in WT, wdr-5, rbbp-5 and ash-2

(A) Western blot was performed with lysates from mixed stage embryos of WT, wdr-5, rbbp-5 and ash-2. The blots are probed with a rabbit anti-H3K4me3, a mouse anti-H3K4me2 and a rabbit anti-H3K4me1 antibody. (B) Quantification of the H3K4 methylation signal from western blot data. Error bars represent standard error of the mean (SEM). Statistics was performed with one-way ANOVA compared to N2, * p<0.01, ** p<0.005, *** p<0.0001.
4.3.6 Absence of RBBP-5 causes reappearance of H3K4me2/me3 in PGCs of unfed L1s

It has been shown that levels of H3K4me1/me2/me3 in adult germ cells are not dramatically reduced by the loss of WDR-5 or ASH-2 (Li and Kelly, 2011; Wang et al., 2011). The ash-2 mutant has even been reported to cause an increase in H3K4me2 levels ASH-2 (Xiao et al., 2011). This finding raises the question as when are the levels of H3K4me1/me2/me3 restored in germlines in these core complex mutants? In absence of RBBP-5, H3K4me1 is undetectable in all cells including the embryonic PGCs, but H3K4me1 reappears in starved L1s (Figure 24). This shows that RBBP-5 is no longer important for deposition of H3K4me1 in L1 and that its deposition is independent of feeding. However, a very different picture emerges when H3K4me2/me3 levels are analysed. We found that both these marks reappear in starved L1s (Figure 24), indicating that RBBP-5 has an antagonising activity towards the accumulation of H3K4me2/me3. We have never observed this effect with wild type L1s, which is consistent with a previous report (Schaner et al., 2003). This study also shows that the PGCs of fed L1s resume accumulation of H3K4me2/me3 (Schaner et al., 2003). We thus tested whether any of these core components could be required in restoring the levels of H3K4me2/me3 in fed L1s. We found that none of these components are required for the deposition of H3K4me2/me3 (Figure 24). Surprisingly, we found the opposite; the absence of RBBP-5 causes an increase in the levels of H3K4me2/me3 of starved L1s (Figure 24). This is different from a previous study with ash-2(−) mutants showing an increase in H3K4me2 levels from embryogenesis (Xiao et al., 2011) but not linked to nutrient exposure. We also observed this effect in embryos, which persists in starved L1s. In conclusion, RBBP-5 has a unique activity antagonizing H3K4me2/me3 accumulation specifically in starved L1s, indicating that RBBP-5 is important for sensing absence of nutrients.
Embryos and L1 larvae from N2, wdr-5, rbbp-5 and ash-2 mutants were stained with anti-H3K4me3/me2/me1, as indicated, and anti-PGL-1 (marker of PGCs) was used to aid in identifying the PGCs. (A) H3K4me3 is not detectable in PGCs of N2 and mutants. (B) H3K4me2 is undetectable in PGCs of N2, wdr-5(-) and rbbp-5(-), however it is detected in ash-2(-). (C) H3K4me1 levels are reduced in the PGCs of rbbp-5(-). (D) Unfed L1 larvae enter diapause stage. H3K4me3 is absent in PGCs of N2, but reappeared in PGCs of rbbp-5(-). (E) H3K4me2 appears in PGCs of rbbp-5(-) and ash-2(-). (F) H3K4me1 is detected in PGCs of all stains. (G) L1 larvae were put on OP50 stain for 30min, H3K4me3 appears in PGCs of all strains. (H) H3K4me2 can be observed in PGCs of all strains after feeding. PGCs are indicated by white dash circles.
4.4 Discussion

We have used genetic deletions to examine the contributions of the core components towards H3K4 methylation. This approach has allowed us to follow the dynamic changes in the H3K4 methylation levels at a cellular resolution both during embryogenesis and in the larval PGCs. This study shows that each core component is required for a fully active MLL/SET/COMPASS complex, yet each core component plays a differential role depending on the developmental stage and the methyl mark involved. Interestingly, we have also revealed that RBBP-5 has an antagonistic role in the accumulation of H3K4me2/me3 during L1 diapause. It remains unclear how RBBP-5 can perform this task, but we can envisage at least two models: the recruitment of a demethylase or the displacement of a methyltransferase, such as SET-2 or SET-16. These are the two H3K4 methyltransferases so far identified in C. elegans. In the first model, demethylation of H3K4me3/me2/me1 can be performed by the lysine demethylase 5 family of jumonji containing protein complexes and demethylation of H3K4me2/me1 can be performed by the conserved Flavin Adenine Dinucleotide (FAD)-dependent demethylase, LSD1 (Christensen et al., 2007; Katz et al., 2009; Shi et al., 2004). There are reports physically linking H3K4 demethylase activities to Trithorax (a Drosophila H3K4 methyltransferase) (Eissenberg et al., 2007; Lee et al., 2007), but the interplay with any core components remain to be investigated. The second model would imply a drastic change in protein structure that could be produced by post-translational modifications (PTMs) of RBBP-5. This PTM could for example produce a change in the conformation of RBBP-5 that would prevent the H3K4MT to associate with the MLL/SET/COMPASS complex.

Our study strongly suggests that RBBP-5 is the least ‘replaceable’ of the three core components followed by ASH-2 (Figure 22). A potential explanation to the relation between RBBP-5, ASH-2 and the HMTs activity is based on the structural analysis of the conserved MLL1 SET domain. These studies proposed a model by which the interaction among core components and MLL1 induces re-orientation of the SET-I sub-domain of MLL1 to an optimal conformation of its catalytic site (Southall et al., 2009). The change in conformation would involve the insertion of aromatic residues between H3 and methyl group to lock-in the optimal orientation as shown by in vitro HMT assays. These experiments suggest a predominant role for RBBP-5 in H3K4 methylation (Odho et al., 2010). However, another in vitro study performed with the yeast homologues of RBBP-5 (Cps50), ASH-2 (Cps60), and WDR-5
(Cps30), indicate that all components contribute equally to methylation of H3K4 (Takahashi et al., 2011). Taking into account the in vivo context of our results, RBBP-5 is the core components that affect the most the accumulation of methyl marks at H3K4.

During the course of this work, we have observed the presence of residual methyl marks at H3K4 in all mutants analysed. This is consistent with other studies (Li and Kelly, 2011; Xiao et al., 2011). These results indicate that a redundant mechanism is present in embryos to maintain basal levels of methylation at H3K4. This mechanism appears particularly efficient to compensate for the loss of WDR-5 but less efficient to compensate for the loss of RBBP-5 and ASH-2. There are at least three possible types of redundancy that could explain these results: structural robustness of the core complex, paralogous redundancy amongst each core component, and a putative HMTs activity targeting H3K4 independent of the core complex. In the first case, there is good evidence suggesting that the core complex is robust even in vitro. In mammals, WDR5 and Ash2L can directly interact with the SET domain region of the HMTs (Cao et al., 2010; Odho et al., 2010). RbBP5 can be directed to the SET-I domain by either interacting with WDR5 via its tail region or with Ash2 via its WD40 beta-propeller motif (Cao et al., 2010; Odho et al., 2010). These interactions suggest that removing a single core component may not inhibit the recruitment of the other two components. This is also supported by in vitro HMT assays showing that the removal of a single core component abolishes H3K4MT activity, but a small amount of methylation, in particular H3K4me1, is nevertheless still detectable (Takahashi et al., 2011). These results are in accordance with our findings. In the second case, redundancy amongst each of the core component could be due to paralogues. WDR-5 has two other paralogues: WDR-5.2 and WDR-5.3. These putative paralogues may take over the function of WDR-5 (also called WDR-5.1) in its absence. A recent study describes redundancy between WDR-5.1 and WDR-5.2 for TRA-1 dependent repression (Li and Kelly, 2014). RBBP-5 has no clear paralogues, but present weak homology with all three WDR-5s, most probably due to the presence of WD40 domains, and are most likely not true paralogues. ASH-2 shows no sequence homology with other C. elegans proteins. Thus, RBBP-5 and ASH-2 are unlikely to be replaced by paralogues. The last possibility is the existence of HMTs acting independently of the core complex. It has been reported that the SET domain of MLL1 and the SET domain of SET-16 can display H3K4 monomethyltransferase activity and weak dimethyltransferase activity in vitro (Dou et al., 2006;
Fisher et al., 2010; Southall et al., 2009; Steward et al., 2006). Thus, it is possible that H3K4MT activity independent of the core complex could additionally participate in the production of the residual methylation described in this study.

An important limitation of our study relates to the genomic location of these characterised effects on H3K4 methylation. It would be of interest in future to determine whether the alterations are due to modest effect at the global levels at many genomic locations or strong depletion at limited genomic locations. ChIP-seq would be adequate to address this aspect. Even though H3K4 methylation is not essential for viability in yeast (Krogan et al., 2002; Roguev et al., 2001), this conclusion is difficult to draw for our study, since residual methylation at H3K4 is still present in each the core complex mutant tested. C. elegans viability might be due to this residual methylation. Thus, it is possible that H3K4 methylation is actually essential to viability in metazoan, but further work directed at completely eliminating H3K4 methylation will be required.

In conclusion, we show that core components can display differential and antagonistic activities towards H3K4 methylation depending on the development stage, the feeding state, and the methyl mark involved. Crucially, we have revealed that the removal of RBBP-5 causes erroneous accumulation H3K4me2/me3 in PGCs of starved L1s. This study supports other work showing that the core components ASH-2 can antagonize accumulation of methyl marks at H3K4 in the embryonic PGCs (Xiao et al., 2011). In contrast to RBBP-5’s antagonistic effect acting solely in PGCs of starved L1. In aggregate, these results are further supporting biochemical evidence that, in metazoan, many subgroups of the multi-proteins MLL/SET/COMPASS complex exist some with overlapping and, crucially, many of them likely to convey specialised functions.
4.5 References


Chapter 5

The core components of the MLL/SET/COMPASS complexes aid RNA Pol II processivity and prevent somatic cells from adopting germline characteristics in Caenorhabditis elegans embryos

Siyao Wang, Kate Fisher, Gino B. Poulin*

Faculty of Life Sciences, Michael Smith Building, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK

Intended Journal: Development
ABSTRACT

The MLL/SET/COMPASS complexes are the histone methyltransferases (HMTs) for catalysing Histone 3 Lysine 4 methylation (H3K4me), which is a mark of active transcription. The *C. elegans* contains two MLL/SET/COMPASS complexes, and they share four core components, WDR-5, RBBP-5, ASH-2 and DPY-30. Previous studies have shown that removal of any core components affect deposition of H3K4me during embryogenesis of *C. elegans*, but cannot abolish the enzymatic activity of the HMTs completely. Therefore, in order to study the significance of this residual H3K4me during embryogenesis, we generated three double mutants, *wdr*-*5; rbbp*-5, *wdr*-*5; ash*-2, *rbbp*-5 *ash*-2. Unexpectedly, one of the double mutants, *rbbp*-5 *ash*-2, is inviable; embryonic lethality and significant reduction of H3K4 methylation were observed in the two viable double mutants. Further investigation revealed a defect of RNA Pol II processivity in the single and double mutants. This observation 1) provides evidence that the MLL/SET/COMPASS complexes are involved in the regulation of transcription during embryogenesis of *C. elegans*. Moreover, we also found that 2) these HMTs also prevent the somatic cells from adopting germline characteristics in embryos. Taken together, our study reveals two fundamental roles of the MLL/SET/COMPASS core components during embryogenesis of *C. elegans*. Importantly, we demonstrated that these two roles are independent from each other, implying that different deposition pattern and level of H3K4 methylation may link to diverse biological outcome.
5.1 Introduction

‘Epigenetic regulation’ is an inherited state of gene regulation that is independent of the genetic information encoded within the DNA itself. Many classes of proteins have been demonstrated to be involved in this epigenetic regulation through altering chromatin architecture. One class of these proteins is the chromatin associated SET-domain containing proteins. Many of the SET domain containing proteins are known as the histone (H) or lysine (K) methyltransferases (HMTs or KMTs), which catalyse methylation on histone tails (Shilatifard, 2012; Tschiersch et al., 1994a). Many post-translational modifications on histone tails are involved in chromatin remodelling, DNA replication and gene expression processes (Berger, 2002). For example, the KMT2 class proteins, Set1/MLL proteins, are enzymes for catalysing histone 3 lysine 4 methylation (H3K4me), which is associated with active transcription (Krogan et al., 2002; Miller et al., 2001; Roguev et al., 2001; Schneider et al., 2005; Shilatifard, 2012). Mutation or translocation of the Set1-related gene MLL1 is correlated with haematological malignancies in human (Rowley, 1998; Schneider et al., 2005; Tenney and Shilatifard, 2005). However, the significance of the Set1/MLL HMTs during development and oncogenesis is still unclear.

The founding member of the SET domain containing H3K4 methyltransferase is the yeast Set1 protein, which is isolated within a macromolecular complex named Set1/COMPASS (COMplex of Proteins ASSociated with Set1) complex. Set-1 is the only H3K4 methyltransferase in S. cerevisiae (Krogan et al., 2002; Miller et al., 2001; Shilatifard, 2006). Following this identification, more than six mammalian homologs (Set1A, Set-1B, MLL1-4-like proteins) were found within the COMPASS-like complexes functioning as H3K4 methyltransferases, suggesting this MLL/SET/COMPASS complex is highly conserved through evolution (Cho et al., 2007; Eissenberg and Shilatifard, 2010). It is thought that there are several MLL/SET/COMPASS complexes, defined as including a Set1/MLL family member and a group of core components (WDR5, RBBP5, ASH2 and DPY30) (Smith and Shilatifard, 2010). Importantly, the Set1/MLL proteins alone are not enzymatically active, but Set1/MLL within the COMPASS complexes is the active form of these enzymes (Schneider et al., 2005). *In vitro* studies find that the isolated Set1/MLL proteins exhibit a weak methyltransferase activity, but the addition of the four core components (WDR5, RBBP5, ASH2 and DPY30) dramatically stimulates this enzymatic activity and forms the minimal subunits composition
required for H3K4 methylation (Dou et al., 2006). A further structural study revealed a Y-shaped architecture of the yeast and mammalian MLL/SET/COMPASS complexes in vitro. The WDR5 and Rbp5 locate on the top two lobes and Ash2L/Dpy30 forms the base. The SET domain of the Set1/MLL is located in the centre of the WDR5, Rbp5 and Ash2L/Dpy30 module (Figure 25A) (Takahashi et al., 2011). Our and others studies have demonstrated that WDR5, Rbp5 and Ash2 are required for the proper methyltransferase activity of the Set1/MLL proteins, and different core components affect methyl states to different degrees (See chapter 4).

H3K4 methylation has three different forms, the mono, di- and tri-methylation (H3K4me1/me2/me3). H3K4me1 is recently known as a mark of active enhancers, and H3K4me2/me3 are marks of active transcription. In *S. cereviseae*, the sole methyltransferases of H3K4, Set1/COMPASS, is associated with transcription initiation (Ng et al., 2003a). Subsequent global proteomic analysis in *S. cereviseae* (GPS) revealed that the Set1/COMPASS can be recruited to the promoters by Paf1 elongation complex (Mueller et al., 2006; Wood et al., 2003b). This recruitment requires monoubiquitination on H2B, which associates with other transcriptional elongation factor to regulate transcriptional elongation (Dover et al., 2002; Lee et al., 2007; Wood et al., 2003a). These results together indicate that the Set1/COMPASS may be involved in the transition between transcriptional initiation and elongation. In addition, this machinery is conserved from yeast to humans, as the mammalian Set1A and Set1B/COMPASS complexes are also recruited by Paf1 elongation complex (Shilatifard, 2012). Thus, it is shown that the MLL/SET/COMPASS complex is associated with the basal transcriptional machinery, especially the transcriptional elongation process.

These previous studies lead to a question of whether the MLL/SET/COMPASS complexes deposited H3K4 methylation affects gene transcription. Therefore, my project was designed to investigate the contribution of the MLL/SET/COMPASS core components on transcriptional regulation. Embryogenesis of *C. elegans* provides an excellent model to investigate this question. *C. elegans* has two methyltransferases, the Set1-like SET-2 and the MLL3/4-like SET-16 (Fisher et al., 2010; Simonet et al., 2007). They share the same group of core components as in mammalian cell, WDR-5, RBP-5, ASH-2 and DPY-30 (Shilatifard, 2012) (Figure 25 A). Transcription level of *C. elegans* is highly dynamic during embryogenesis.
Transcription is active in the somatic lineage after 4-cell stage but remains low in the germline lineages (Shin and Mello, 2003). According to our previous study, the H3K4me2/me3 is lineage specifically distributed during embryogenesis according to the transcriptional status. Moreover, we used the loss-of-function alleles of *C. elegans*, wdr-5, rbbp-5 and ash-2 (Figure 25 B) to investigate how these core components contribute to H3K4 methylation during embryogenesis. Our previous results show that each core component of the MLL/SET/COMPASS complexes plays a preferential role on depositing H3K4 methylation during embryogenesis, and residual methyl marks were detectable in all single mutants. This study suggests that these three core components (WDR-5, RBBP-5 and ASH-2) have function redundantly to regulate methyltransferases activity. Indeed, by generating double mutants (Figure 25 C), we found that the *rbbp-5 ash-2* mutant is inviable, and the other two viable double mutants exhibit a further reduction of H3K4 methylation and a defect in the RNA Pol II processivity. Moreover, we also observed that the somatic cells of the mutants acquire germline characteristics, indicating that these core components are required for maintaining the distinction between somatic and germline lineage during *C. elegans* embryogenesis. In addition, our results have demonstrated that ectopic expression of germline characteristics is not caused by defect in RNA Pol II processivity. In summary, this study reveals two independent roles of the MLL/SET/COMPASS core components during embryogenesis of *C. elegans*, and provides insight into the mechanism of histone modifications on regulation of transcription and lineage distinction in metazoans.
Figure 25. Introduction to the MLL/SET/COMPASS complexes in C. elegans and generation of double mutants

(A) The two homolog of the MLL/SET/COMPASS complexes in C. elegans. (B) The loss-of-function alleles of WDR-5, RBBP-5 and ASH-2 we used in this study. (C) By performing the genetic cross, we try to generate three double mutants combination.
5.2 Materials and Methods

5.2.1 Strains and general maintenance

Strains used in this study were maintained as previously described (Brenner, 1974a). Bristol N2 was used as wild type, other strains are RB1304 wdr-5 (ok1417), OL0058 rbbp-5 (tm3463), PFR403 ash-2 (tm1905), MT2495 lin-15B (n744), KM267 hs::hlh-1 and PFR210 swd-3::gfp (wdr-5::gfp) (Table 2).

5.2.2 Generation of worm strains

Worm strains generated in this study are: OL0076 wdr-5 (ok1417); rbbp-5 (tm3463), OL0073 wdr-5 (ok1417); ash-2 (tm1905), rbbp-5 (tm3463) ash-2 (tm1905) (inviable), OL0084 wdr-5 (ok1417); hs::hlh-1, OL0085 rbbp-5 (tm3463); hs::hlh-1, OL0086 ash-2 (tm1905); hs::hlh-1, OL0107 wdr-5 (ok1417); rbbp-5 (tm3463); hs::hlh-1, OL0108 wdr-5 (ok1417); ash-2 (tm1905); hs::hlh-1, OL0089 lin-15B (n744); hs::hlh-1 and OL0090 wdr-5 (ok1417); hs::hlh-1; wdr-5::GFP (Table 2). Crosses were performed by placing 3 hermaphrodites with 6 males on a 3cm plate, which is seeded with OP50 strain of E. coli. Male was generated by placing L4 hermaphrodites at 30 °C for 3-4 hours. Genotypes of all strains were confirmed by a single worm PCR.

5.2.3 Co-immunofluorescence for embryos and larvae

Immunofluorescence by freeze crack was performed on polylysine treated slides at 0.3%. Slides treated by application of 75μl of polylysine 0.3%, dried 10 min at 70 °C, quickly rinsed in distilled water and excess liquid wiped off. The slides were 3 × 14 mm printed wells from Fisher Scientific LTD UK. About 30 mothers were placed in a well with a drop of M9 buffer, to wash off bacteria. These mothers were transferred into a 5-6 μl drop of M9 onto the polylysine treated well using an eyelash. Embryos were dissected from these mothers by a syringe needle. A cover slip (22 mm × 50 mm) was applied at a right angle and the slide placed at −80 °C for at least 20 min. The cover slip was then promptly removed and embryos methanol fixed at −20 °C for 10 min, washed 5 min in PBS, and then followed by two washes in PBS-tween 0.2%. Primary antibody was incubated overnight at 4 °C in a humid chamber. Washes
performed as described above. Secondary antibody was incubated 2 h at 37 °C. Washes performed as described above. Mowiol was applied to preserve fluorescence.

For quantification, 6-8 cells were chosen and analysed by ImageJ from each stage of embryo. The abundance of H3K4me1/me2/me3 is relative to DAPI, and a line chart was drawn for wild type and all mutants.

For larvae staining, adult hermaphrodites of each mutant were washed off from plates with M9 buffer and dissolved in bleaching buffer (1 M NaOH in 10% sodium hypochlorite solution). After two additional M9 washes, the embryos were collected in 500 μl M9 buffer. After 48h, put a drop of starved L1 larvae (around 100 larvae) on a poly-lysine treated well. Apply the same fixation and staining as mentioned above. The remaining L1 larvae were fed E. coli (OP50 strain) for 30min before fixation and staining.

5.2.4 Co-immunofluorescence for gonad
A small drop of 2mM tetramisole was placed onto a poly-lysine treated 3-well slides. 30 adult worms were transferred to the drop, and paralyzed in the tetramisole. Slice the adults near the heads and tails with two syringe needles. After dissection, the gonads will extrude automatically and stick to the slides. Then a cover slide was placed and followed by the same treatment as mentioned in the last paragraph.

5.2.5 Western Blot
Western blot analysis was performed as described previously with a few modifications (Fisher et al., 2010). Briefly, embryo protein extracts were prepared by bleaching young mothers followed by washing the embryos four times with M9 buffer. A minimum volume of 20 μl pelleted embryos was collected for each protein sample, boiled in Laemmli buffer containing 100 mM DTT and sonicated. Western blots were performed following the standard protocol with the BioRad Mini Trans-Blot cell system.
5.2.6 Cell fate challenge assay

The hs:: hlh-1 transgene was crossed into wdr-5, rbbp-5, ash-2, wdr-5; rbbp-5, wdr-5; ash-2, and lin15B strains. Adult hermaphrodites were bleached to harvest embryos. P1-P3 stages embryos were selected by using a mouth pipette and transferred to 100 μl of M9 buffer. These embryos were incubated at 20°C for 75min, and the temperature was increased at 0.1°C/s to 33°C. Embryos were incubated at 33°C for 30min, and then cool down to 20°C for overnight. Follow by a myosin staining as described above.

5.2.7 RNAi by feeding

RNAi experiments were performed using HT115 (DE3) bacteria carrying the L4440 vector. The L4440 empty vector (EV) and cdk-9 (RNAi) vector were transformed to the HT115 (DE3) bacteria grown for 6 hours and spotted on six-well plates. The N2 adult hermaphrodites were bleached, and embryos were placed on the first well of the RNAi plate. After these worms had hatched and reached L4 stage, they were transferred to the second well and allowed to lay eggs. After the F1 worms were hatched, they were transferred to the third well. As the F1 reached the adulthood, immunofluorescences against myosin were followed as described above.

5.2.8 Antibodies

The primary Antibodies used in this study were: anti-H3K4me3 (Abcam ab1012, 1:100), anti-H3K4me2 (Millipore CMA303, 1:5000), anti-H3K4me1 (Abcam ab8895, 1:500000), anti-PGL-1 (a gift from Judith Kimble, 1:100), anti-PGL-3 and anti-GLH-1 (Gift from Susan Strom), anti-pSer2 (Covance MMS-129R, 1:50), anti-pSer5 (Covance MMS-134R, 1:50), anti-Myosin (developmental studies hybridoma bank 5-6, 1:100) and anti-MyoD (Gift from Susan Strom). The secondary antibodies used were from Jackson ImmunoResearch: DyLight 594 AffiniPure Goat Anti-Mouse IgG (H+L) (cat.115-515-146, 1:200) and DyLight 488 AffiniPure Goat Anti-Rabbit IgG (H + L) (cat.111-485-144, 1:200). Anti- mouse IgM (Abcam ab97007, 1:200), Anti-mouse IgM (HRP) (Abcam ab97230, 1:500)
5.3 Results

5.3.1 The core components of SET/MLL/COMPASS complex are essential for the viability of C. elegans.

According to our previous study, we hypothesised that the core components of MLL/SET/COMPASS complexes function redundantly, although residual methylation of H3K4 is still detectable in each single mutants. To investigate this hypothesis, we decided to generate three double mutants, wdr-5; rbbp-5, wdr-5; ash-2 and rbbp-5 ash-2, by performing genetic crosses.

We firstly generated the rbbp-5 ash-2 mutant. As the deletion alleles of rbbp-5 and ash-2 are located at the same chromosome (Chromosome II), the chance of recombination is around 5%. In order to get the recombination strain, we single-cloned 136 progeny from the rbbp-5 (+/-) ash-2 (-/-) heterozygote (Figure 26). By performing PCR to confirm the genotype of these progeny, I was able to identify the rbbp-5 (-/-) ash-2 (-/-) and rbbp-5 (+/-) ash-2 (-/-) combinations. However, we were unable to generate the rbbp-5(-/-) ash-2 (-/-) homozygote according to the low chance of recombination. We then single-cloned the progeny derived from the rbbp-5 (-/-) ash-2 (-/-) and rbbp-5 (+/-) ash-2 (-/-) heterozygotes, and performed PCR to identify their genotypes. According to the Mendel's law of segregation, one out of four progeny should be rbbp-5 (-) ash-2 (-) homozygote. Unexpectedly, the PCR analysis revealed that none of the progeny were rbbp-5 (-) ash-2 (-) homozygotes. To confirm this observation, we re-examined genotypes of all the progeny derived from the rbbp-5 (-/-) ash-2 (+/-) and rbbp-5 (-/-) ash-2 (-/-) heterozygotes. Our PCR result indicates that all of these viable progenies are either rbbp-5 (-) or ash-2 (-) single mutants, suggesting that the rbbp-5 (-) ash-2 (-) combination is inviable. Moreover, the rbbp-5 (-/-) ash-2 (+/-) and rbbp-5 (+/-) ash-2 (-/-) heterozygotes could not be found in these progenies, suggesting that the viability of these heterozygotes may be attributable to maternal effect, which can be diluted over two generations. As the defect of viability was not found in the rbbp-5 (-) or ash-2 (-) single mutants, it is clearly evident that the functions of RBBP-5 and ASH-2 are redundant and essential for the viability of C. elegans.

The viability defect observed in the rbbp-5 ash-2 homozygote promotes us to generate the wdr-5; rbbp-5 and wdr-5; ash-2 double mutants by performing genetic crosses. Unlike the
rbbp-5 ash-2, both of the wdr-5; rbbp-5 and wdr-5; ash-2 are viable. In addition, some of the wdr-5; rbbp-5 mutants displays a Multiple Vulva (Muv) phenotype (less than 5%). After we obtained the wdr-5; rbbp-5 and wdr-5; ash-2 mutants, the embryonic lethality and brood size were characterised (Table 4). Compare to the wild type, the embryonic lethality is increased to 41% in wdr-5; rbbp-5 and 38% in wdr-5; ash-2 mutant. However, according to our previous paper, the embryonic lethality of the single mutants is only increased to 4%, 2% and 5% in wdr-5, rbbp-5 and ash-2 mutants, respectively (Wang et al., 2011). Moreover, the average brood size of the two viable double mutants is reduced to 21 and 39 in wdr-5; rbbp-5 and wdr-5; ash-2 mutants, which is much lower than the brood size of single mutants. According to these characterisation data, we conclude that WDR-5, RBBP-5 and ASH-2 work redundantly with each other for maintaining the viability of C. elegans.

\*p<0.01, \**p<0.005, ***p<0.0001

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>% emb. leth.</th>
<th>aver. brood size</th>
</tr>
</thead>
<tbody>
<tr>
<td>N2</td>
<td>0</td>
<td>204</td>
</tr>
<tr>
<td>wdr-5</td>
<td>4**</td>
<td>111**</td>
</tr>
<tr>
<td>rbbp-5</td>
<td>2</td>
<td>44***</td>
</tr>
<tr>
<td>ash-2</td>
<td>5*</td>
<td>38***</td>
</tr>
<tr>
<td>wdr-5; rbbp-5</td>
<td>41***</td>
<td>21***</td>
</tr>
<tr>
<td>wdr-5; ash-2</td>
<td>38***</td>
<td>39***</td>
</tr>
<tr>
<td>ash-2 rbbp-5</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 5. Characterisation of the embryonic lethality and brood size of mutants
Figure 26. The \textit{rbbp-5 ash-2} double mutant is lethal

We segregated the progeny from the \textit{rbbp-5} (-/-) \textit{ash-2} (+/-) heterozygotes. The genotypes of 136 progeny were characterised and the number of each genotype was labelled. In this characterisation, we could not find the \textit{rbbp-5 ash-2} homozygote. We next segregated the progeny from the \textit{rbbp-5} (-/-) \textit{ash-2} (-/-) and \textit{rbbp-5} (+/-) \textit{ash-2} (+/-) heterozygote and the \textit{rbbp-5 ash-2} homozygote is still inviable. The Red crosses indicate that we could not find this combination.
5.3.2 H3K4 methylation level is further reduced in the double mutants compared to the single mutants

As the core components are responsible for the deposition of H3K4 methylation, we next addressed whether this H3K4 methylation is further reduced in the double mutants compared to the single mutants and wild type. As we were unable to maintain the rbbp-5 ash-2 mutant, we used the two viable double mutants to perform the immunofluorescence with antibodies against the H3K4me1/me2/me3. Six different stages of embryogenesis (P0, P1, P2, P3, P4 and Z2-Z3 stages) were chosen to measure the deposition of H3K4 methylation. In order to visualize the changes of methylation level between different stages of embryogenesis, we analysed the data as described in our previous paper (See chapter 4). 6-8 cells were chosen for each stage and mutant to analyse. We expressed the results as a percentage of the H3K4me signal over the DAPI signal for each cell, and plotted a line chart for wild type and all mutants (Figure 27A). According to our previous study, each component has a preferential role on depositing the methyl marks, but residual methylation on H3K4 was still detectable in all single mutants (Figure 27A). In the wdr-5; rbbp-5 mutant, H3K4me1/me2/me3 is dramatically reduced and stays at a low or undetectable level after P0 stage. This reduction of H3K4me1/me2/me3 is more obvious at late stages of embryogenesis (P4 and Z2-Z3 stages). Similar to the wdr-5; rbbp-5 mutant, H3K4me2/me3 level is further reduced in the wdr-5; ash-2 mutant compared to single mutants, albeit it is still detectable at early stages. More significantly, H3K4me1 is only slightly affected in the wdr-5 and ash-2 mutants, but dramatically reduced in the wdr-5; ash-2 double mutant (Figure 27A).

To confirm this observation, western blot analysis was carried out to measure the methylation level in all mutants (Figure 27 B), and the signal level was quantified and analysed (Figure 27 C). Compare to our previous western blot analysis for single mutants (see chapter 4), we found that the methylation level is further reduced in the wdr-5; rbbp-5 double mutant than in the single mutants (Figure 27 C). Due to the limitation of time, this western blot has only been conducted once. In order to confirm this observation, more analyses are required to carry out. We next measured H3K4me1/me2/me3 level in the post-embryonic stages; the immunofluorescence against H3K4me1/me2/me3 in adult germ cells was performed (Figure 27 D). According to previous work, H3K4me3 is mainly reduced in the mitotic and transition zone of the germline in wdr-5 and rbbp-5 mutants (Li and Kelly, 2011). The distribution pattern
of H3K4me3 in the double mutants is similar to wdr-5 and rbbp-5 mutants, as it is lower in the mitotic and transition zone than in other zones. Interestingly, H3K4me2 deposition is affected mainly in the mitotic zone of the germline in the wdr-5; rbbp-5 mutant. This reduction of H3K4me2 was not observed in the single mutants and wild type. No reduction of H3K4me1 was observed in both double and single mutants. Therefore, according to the immunofluorescence and western blot results, the enzymatic activity of the MLL/SET/COMPASS complex for catalysing H3K4me1/me2/me3 is further reduced in the two viable double mutants than in the single mutants, suggesting the core components work redundantly for depositing H3K4 methylation in embryos and adult germ cells.
Figure 27. H3K4 methylation is reduced in the embryos and germline of double mutants, compare to single mutants

(A) The line charts of the H3K4me1 (blue), H3K4me2 (red) and H3K4me3 (green) level at different stages of wild type and mutants. We quantified the abundance of H3K4me1/me2/me3 relative to DAPI using immunofluorescences. For each stage of embryos, 6-8 cells were analysed. Compare to the wild type, all mutants are impaired in H3K4me1/me2/me3 deposition to different degrees. Moreover the H3K4me1/me2/me3 is further decreased in the two double mutants compare to single mutants. (B) The adult germline of double mutants are impaired in the H3K4me2/me3 deposition. The dash lines indicate the border of the distal area in which the H3K4 methylation is poor. No obvious decrease of the H3K4me1 level was observed in the adult gonad. Scale bar is 50μm.
5.3.3 Transcriptional elongation (pSer2), but not initiation (pSer5), is reduced in the core component mutants.

It is well known that di-/tri-methylation on H3K4 is positively correlated with transcriptional activation. As we found that H3K4me2/me3 is further reduced in the embryos of the two viable double mutants, we predicted that transcription is affected in the two double mutants. A good indicator of active transcription is the phosphorylation states at CTD of RNA Polymerase II: Phosphorylation on the Ser5 residue (pSer5) of the CTD on RNA Pol II is linked to transcriptional initiation, while phosphorylation on the Ser2 residue (pSer2) indicates transcriptional elongation. In order to investigate whether the transcriptional states are affected in mutants, we carried out immunofluorescence with antibodies specific to pSer5 and pSer2 in wild type and mutants (Figure 28 A and B). In the wild type, the C. elegans pSer5 and pSer2 signal is deposited in the nucleus of somatic cells from 4-cell stage, but poor in the germline blastomeres until the Z2-Z3 stage (Seydoux and Strome, 1999). We firstly examined the pSer5 level in all the mutants, and no decrease was detected in all mutants compare to wild type (Figure 28 A). We next performed the same analysis with pSer2 antibody (Figure 28 B). Surprisingly, a reduction of pSer2 was observed in some embryos of wdr-5; rbbp-5 and wdr-5; ash-2 mutants, but no obvious reduction was found in other mutants (Figure 28 B). Interestingly, this pSer2 reduction was only observed in embryos at later embryogenesis (Z2-Z3 stage embryos), but not at early stages, suggesting that later stages of embryogenesis may be more sensitive to the removal of the core components than the early stages. We next characterised the pSer2 defect in wild type and mutants, and we noticed that the percentage of embryos with the pSer2 defect is much lower in single mutants compared to double mutants (Table 5). In single mutants, the percentage of Z2-Z3 stage embryos with P-Ser-2 defect is 1.75%, 4% and 3% in wdr-5, rbbp-5 and ash-2 mutants, respectively. However, this percentage is increased to 28% and 35% in wdr-5; rbbp-5 and wdr-5; ash-2. Western bolt was also performed to confirm the reduction of the pSer2 in the single and double mutants (Figure 28 C). We prepared mixed stage embryo extracts from all mutants, and compared the H3K4me1/me2/me3 levels between wild type and mutants. Indeed, consistent with our immunofluorescence data, the level of pSer2 is greatly reduced in wdr-5; rbbp-5 and wdr-5; ash-2 compared to N2, whereas no significant difference in pSer5 was detected among all mutants. Surprisingly, a reduction of pSer2 was also observed in rbbp-5 and ash-2 single mutants by western blot analysis, suggesting immunofluorescence may not be sensitive to
detect minor reductions. It is also indicated that RBBP-5 and ASH-2 are involved in the regulation of transcriptional elongation. In addition, we found that the level of pSer2/pSer5 is slightly increased in *wdr-5* mutants, although this is a non-statistics difference. Therefore, we concluded from these data that the double mutants are impaired in the transcriptional elongation, but not initiation, during later embryogenesis. This indicates that the MLL/SET/COMPASS complexes may be important for the transition between transcriptional initiation and elongation in *C. elegans* embryo.
Figure 28. The pSer2, but not the pSer5, is reduced in the two double mutants

(A) Embryonic immunofluorescence against DAPI and the transcriptional initiation mark, pSer5. (B) A decrease of the transcriptional elongation mark, pSer2, was observed in Z2-Z3 stage embryos of double mutants, but was not found in all single mutants. (C) Western blot against the pSer5 and pSer2. The general level of pSer2 is decreased in two double mutants. Error bars represent standard error of the mean (SEM). Statistics was performed with one-way ANOVA compared to N2, * p<0.01, ** p<0.005, *** p<0.0001.
<table>
<thead>
<tr>
<th>Genotypes</th>
<th>% pSer2 defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>N2</td>
<td>0</td>
</tr>
<tr>
<td>wdr-5</td>
<td>1.75</td>
</tr>
<tr>
<td>rbbp-5</td>
<td>4</td>
</tr>
<tr>
<td>ash-2</td>
<td>3</td>
</tr>
<tr>
<td>wdr-5; rbbp-5</td>
<td>28***</td>
</tr>
<tr>
<td>wdr-5; ash-2</td>
<td>35***</td>
</tr>
<tr>
<td>ash-2; rbbp-5</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 6. Characterisation of the pSer2 defect
5.3.4 The markers of germline blastomeres, P granules, are ectopically expressed in the somatic cells of the core component mutants.

According to our immunofluorescence and western blot analysis, removal of the core components affects transcriptional elongation during embryogenesis. The defect in gene transcriptional elongation may lead to further consequences, such as defect in embryonic development, differentiation and cell fate decision. We then asked what the consequences of removing the core components of the embryogenesis process are. Previous studies revealed that the H3K4me2/me3 deposition is lineage specific during embryogenesis (Wang et al., 2011b). In addition, H3K4me2/me3 is ectopically expressed in the germline blastomeres by removing a germline determinant, PIE-1, which absence causes a germline-to-soma transformation (Mello et al., 1996; Tenenhaus et al., 1998). Thus, we asked if the MLL/SET/COMPASS complexes are important for maintaining the distinction between somatic and germline lineages. To test these possibilities, we used the core components mutants to investigate whether the somatic cells can acquire germline characteristics. One important characteristic of the germline blastomeres is the expression of the germline marks, P granules (Hanazawa et al., 2011). We then performed immunofluorescence with an antibody against one of the P granules, PGL-1, in wild type and all core component mutants (Figure 29 A). The SynMuv B group of proteins are known as the regulator for preventing the somatic cell from adopting the germline cell fate, and removal of the SynMuv B protein results in an ectopic expression of P granules in embryos and larvae (Petrella et al., 2011). Thus we used one of the SynMuv B mutants, lin-15B, as a positive control in this study. In the wild type, P granules are normally expressed in the cytoplasm of the germline blastomeres and two PGCs. In the core component mutants, PGL-1 signal was detected in the cytoplasm of some somatic cells at P4 and Z2-Z3 stage of embryos, but not at early stages. Most of the somatic cells with ectopic PGL-1 expression are adjacent to the germline blastomeres P4 or Z2/Z3. This observation suggests that those somatic cells may be the descendant of the germline lineage. This PGL-1 phenotype was characterised at P4 and Z2-Z3 stage separately (Figure 29 B). Indeed, our characterisation suggests that removal of the core components significantly increases the PGL-1 ectopic expression phenotype. According to this characterisation, a significant difference was observed between double and single mutants, suggesting that the core components are partially redundant for regulating the PGL-1 distribution. This redundancy was also confirmed by the PGL-1 staining performed in the L1 stage larvae.
(Figure 29 C). In the L1 stage, PGL-1 is normally detected only in the two PGCs, Z2 and Z3. However, PGL-1 was ectopically expressed in somatic cells of double mutants, but not in all single mutants. Moreover, this PGL-1 ectopic expression was also detected in the embryos and larvae of lin-15B mutant, as previous reported (Petrella et al., 2011). This observation together with the embryonic immunofluorescence suggests that the PGL-1 distribution is regulated redundantly by the MLL/SET/COMPASS core components. These core components are essential for preventing somatic cells from acquiring the germline markers during late embryogenesis and larvae stage.
Figure 29. One of the P granules, PGL-1, is ectopically expressed in somatic cells of core components mutants

(A) The P4 stage embryos were stained with DAPI and PGL-1 antibody and characterised for the PGL-1 ectopic expression. The mutation of SynMuv B group proteins, lin-15B, was used as a positive control. In the wild type, PGL-1 is only located at the cytoplasm of P4 cell. In the mutants, the PGL-1 was detected in some adjacent somatic cells of P4 cell. The percentage of the PGL-1 phenotype is significantly increased in mutants compared to wild type. (B) The similar analysis was performed in the Z2-Z3 stage embryos. The PGL-1 signal, which is normally expressed in Z2 and Z3 cells, was detected in many somatic cells of the mutant embryos. The percentage of this defect is increased in all mutants. The red circles indicate the cells having PGL-1 expression. The number of embryos was characterised was labelled in the Parentheses. Statistics was performed with the chi-square for difference. * p<0.01, ** p<0.005, *** p<0.0001. (C) L1 stage larvae were stained with DAPI and PGL-1 antibody, and ectopic PGL-1 signal was detected in double mutants and our positive control.
To confirm this P granules defect, we examined the other two components of P granules, PGL-3 and GLH-1 (Figure 30 A and B). These two components of P granules have the same distribution pattern as the PGL-1 in the wild type embryos, and ectopic expression of these two proteins was observed in the SynMuv B mutants. In our study, similar to the PGL-1 defect, the ectopic expression of PGL-3 and GLH-1 was observed in all mutants at P4 and Z2-Z3 stages of embryogenesis. However, the distribution pattern of the ectopic PGL-3 and GLH-1 marks differ from the PGL-1, as these marks are widespread over the whole embryos instead of locating at certain descendants of germline lineage. By characterising the PGL-3 and GLH-1 phenotype, we confirmed that the ectopic PGL-3 and GLH-1 expression is significantly increased in the core component mutants, and no significant difference of this defect was observed between the single mutants and the double mutants (Figure 30 A and B). This result suggests that the regulatory roles of the core components on regulation of the PGL-3 and GLH-1 distribution are not redundant. In larvae staining, unlike the PGL-1 defect, the ectopic PGL-3 and GLH-1 expression was detected in the L1 stage larvae of single mutants (Figure 30 C). It confirmed that these core components have a non-redundant role on regulation of PGL-3 and GLH-1 distribution during late embryogenesis and early larvae stages. Taken together, we concluded that the core components of the SET/ MLL/COMPASS complex prevent the somatic cells from acquiring the expression of germline marks, P granules.
Figure 30. P granules, PGL-3 and GLH-1, are ectopically expressed in the somatic cells of the core component mutants

(A) Immunofluorescence against DAPI and PGL-3 was performed in wild type, wdr-5, rbbp-5, wdr-5; rbbp-5 and lin-15B. The PGL-3 expression was detected in somatic cells of embryos from wdr-5, rbbp-5, wdr-5; rbbp-5 and lin-15B. This phenotype was characterised, and significant increase of the PGL-3 defect percentage was observed in mutants. (B) The same analysis was performed for GLH-1. The GLH-1 signal was detected in somatic cells of all mutants. The characterisation indicates that this phenotype is significantly increased in all mutants. The number of embryos was characterised was indicated in the Parentheses. Statistics was performed with the chi-square for difference. * p<0.01, ** p<0.005, *** p<0.0001. (C) PGL-3 and GLH-1 staining in L1 stage larvae. The signal of PGL-3 and GLH-1 was detected in rbbp-5 and lin-15B mutants.
5.3.5 Removal of the core components causes a decrease of developmental plasticity in somatic blastomeres.

To further test if the MLL/SET/COMPASS complexes are essential for the distinction between somatic and germline cells during embryogenesis, we tested another important characteristic of the germline blastomeres; the low developmental plasticity. Developmental plasticity is much higher in somatic blastomeres than in germline blastomeres, as it allows the somatic blastomeres to potentially differentiate into different cell types (Fukushige and Krause, 2005). However, the germline blastomeres have low developmental plasticity, preventing them from adopting other cell fates. In order to measure the level of the developmental plasticity in embryos, we performed a cell fate challenge assay (Fukushige and Krause, 2005; Yuzyuk et al., 2009) (Figure 31 A). In the cell fate challenge assay, a myogenic transcription factor, hlh-1, is over-expressed in C. elegans. hlh-1 initiates a transcriptional cascade and eventually express the terminal skeletal muscle gene products. The integrated heat-shock hlh-1 (hs::hlh-1) strain can transform the majority of somatic blastomeres to muscle-like cells under a short heat-shock at 2 or 4-cell stage when the somatic blastomeres have high developmental plasticity. Because the germline blastomeres lack developmental plasticity; those cells will not be able to be transformed to the muscle cells in the cell fate challenge assay. An immunofluorescence against myosin was conducted to detect the myosin expression level in the embryos (Figure 31 B). Myosin is normally expressed in the body wall muscle cells, which is located at the outer layer the embryos. However, under the cell fate challenge treatment, the majority of the somatic cells become muscle cells and myosin can be detected in most cells under microscope. By analysing the percentage of embryos with non-ectopic myosin expression (normal myosin expression pattern), we are able to measure the somatic blastomeres’ developmental plasticity. For example, an increase of this percentage indicates that the somatic cells are resistant to the cell fate challenge and refers to a loss of developmental plasticity in the somatic blastomeres.

We hypothesised that the removal of the MLL/SET/COMPASS core components results in a loss of developmental plasticity in somatic cells. To test this hypothesis, we crossed the hs::hlh-1 transgene into the core components mutants and lin-15B. Then the cell fate challenge assay was used to analyse the developmental plasticity of all strains. The embryos with non-ectopic myosin expression were characterised after performing myosin staining.
(Figure 31 C). In the wild type, only 2.5% of embryos were resistant to the cell fate transformation and displayed a normal myosin expression pattern. However, in the core components mutants, a great increase of embryos with non-ectopic myosin expression was detected in both single and double mutants (Figure 31 C). In addition, no significant difference in this percentage between single and double mutants was found, suggesting that this role of core components is non-redundant (Figure 31 C). However, no effect was observed in the lin-15B mutant, implying that although the SynMuv B proteins are responsible for preventing the somatic cells from adopting germline cell fate, they do not affect the developmental plasticity of somatic blastomeres as the core components do (Figure 31 C). We also transferred the wdr-5 gene into the wdr-5; hs::hlh-1 background to rescue this developmental plasticity defect. Indeed, the percentage of resistant embryos was greatly reduced in this strain, albeit is still higher than hs::hlh-1. Taken together, these results suggest that the removal of the core components causes the somatic cells resistance to adopt the muscle-like cell fate under the cell fate challenge assay. However, according to pervious studies that transcription of certain genes can be down-regulated in mutants; a possibility is that the transcription of hlh-1 is impeded in the core components mutants. The repression of hlh-1 can also result in a failure of the muscle cell transformation, therefore produces a false positive result. To exclude this possibility, we used an antibody specific to the vertebrate homolog of HLH-1, MyoD, to confirm that the HLH-1 is over expressed in all mutants (Figure 31 D). Ectopic MyoD signal can be detected at high level within both mutants and hs::hlh-1. This result confirms that the cell fate resistance of somatic blastomeres is due to the loss of developmental plasticity in the core components mutants. In general, our results indicate that the core components of the SET/MLL/COMPASS complex are required for maintaining the developmental plasticity of somatic blastomeres.
Figure 31. The somatic cells of core components’ mutants lose their developmental plasticity

(A) A model of the cell fate challenge assay. In wild embryos, the somatic blastomere (red cell) can develop into different cell types (indicated by different colors), but germline blastomere (green cell) can only adopt germ cell fate. In the cell fate challenge assay, the hs::hlh-1 was transferred into embryos, which can transform the majority of somatic cells to muscle cells (red rounds), but the germ cell is resistant to this transformation. (B) Under the cell fate challenge assay, embryos with high developmental plasticity were stained with myosin antibody and DAPI, and the myosin was detected all over the embryos. The embryos had low developmental plasticity cannot be transformed to muscle cells, thus the myosin signal was only detected in the outer edge of the embryos as in normal embryos. (C) The number of the embryos with non-ectopic myosin staining was counted and characterised for wild type and mutants. Significant increase in the percentage was detected in all core components mutants, but not detected in lin-15B mutant. The number of embryos was characterised was indicated in the Parentheses. Statistics was performed with the chi-square for difference. * p<0.01  ** p<0.005  *** p<0.0001. (D) Embryos with either ectopic myosin expression or non-ectopic expression were stained with DAPI and MyoD. MyoD is highly expressed in all embryos, suggesting that the hs::hlh-1 transgene is working properly.
5.3.6 The role of the core components on regulation of PGL-1 ectopic expression is independent of transcription

According to our results, we have already identified two roles of these three core components during *C. elegans*’ embryogenesis. One is that the core components is involved in maintaining the processivity of RNA Pol II; while the other one is that these core components can prevent somatic cells from adopting germline characteristics. We then addressed the next question, which was whether these two roles of the core components are linked to each other. To answer this question, we performed RNAi experiment to knock down the enzyme responsible for depositing pSer2, CDK-9, and examined the expression pattern of PGL-1 (Lee and Young, 2013; Smith and Shilatifard, 2013) (Figure 32). Knock down of CDK-9 caused defective gastrulation and embryonic arrest at Z2-Z3 stage, consistent with previous study (Bowman et al., 2013). Our pSer2 staining confirmed the efficiency of this *cdk*-9 (RNAi), as pSer2 signal was greatly reduced in the *cdk*-9 (RNAi) compared to the EV (RNAi) treatment. Then, a co-immunofluorescence against PGL-1 was carried out and shown that the PGL-1 was detectable only in the cytoplasm of germline blastomeres and two PGCs. In contrast, a reduction of PGL-1 signal level was observed in some embryos, suggesting the transcription of PGL-1 is affected by the *cdk*-9 (RNAi) at some degrees. This result indicates that the P granules ectopic expression we observed in the core components mutants may not be due to the defect in transcriptional elongation. It also suggests that the core components of the MLL/SET/COMPASS complexes maintain the distinction between somatic and germline lineage independently from their roles on regulation of processivity of RNA Pol II.
Figure 32. PGL-1 expression is not affected by the cdk-9 (RNAi)

(A) In P4 stage embryos, PGL-1 was detected only in the P4 cells from both EV or cdk-9 (RNAi) animals. (B) the pattern of PGL-1 is not affected by cdk-9 (RNAi) at Z2-Z3 stage of embryos.
5.4 Discussion

By generating double mutants of the MLL/SET/COMPASS core components, we expected to completely abolish the methyltransferase activity of the MLL/SET/COMPASS complexes in C. elegans. Indeed, our data indicates that the level of H3K4 methylation is further reduced in the double mutants compared to the single mutants and wild type. The great reduction of H3K4 methylation is associated with severe defects in viability and lineage specification, implying the importance of the MLL/SET/COMPASS core components for the proper development of C. elegans. Importantly, our data suggest that the MLL/SET/COMPASS complex has an impact on RNA Pol II processivity. Moreover, it also shows that the core components function redundantly for depositing H3K4 methylation. By the limitation of time, we were unable to understand the mechanism of this redundancy. However, based on our previous work that each core component has preferential role on H3K4 deposition, we suggest that each core component deposits H3K4 methyl marks on distinct gene loci. Therefore, by depleting more components, H3K4 methylation would be removed from more genes. To confirm this possibility, RNA-seq and Chip-sequence needs to be conducted to identify the effects on H3Kme1/me2/me3 deposition pattern in each mutant.

In this study, we revealed that the removal of the core components affects the RNA Pol II processivity and transcriptional elongation, as pSer2 is significantly decreased in the single and double mutants of core components. Although the MLL1/2/COMPASS complexes have been reported that they are required for transcription activation of hox genes in mammals (Wang et al., 2009), there is no evidence to support that the MLL/SET/COMPASS complex can regulate transcriptional elongation in C. elegans. Immunofluorescence and western blot analysis have shown a striking reduction of pSer2 in the two viable double mutants, suggesting that the core components are required for transcriptional elongation throughout the genome. This indicates that H3K4 methylation may recruit downstream effectors that promote pSer2 deposition. Many downstream effectors of H3K4me1/me2/me3 have been identified, and some of them are associated with the basal machinery of transcription, such as TAF3 (Lauberth et al., 2013). TAF3, a subunit of the TFIID, can recognise H3K4 methylation via its PHD domain and play a crucial role in transcriptional initiation. Thus, TAF3 can be a potential candidate that connects the H3K4 methylation to transcription activation. We conclude that
H3K4 methylation may be an essential chromatin landmark for directing the basal transcriptional machinery to activate gene expression throughout the genome of *C. elegans*.

Our western blot analysis also uncovers an unexpected result that the level of pSer2 is slightly increased in *wdr*-5 mutants, indicating that the MLL/SET/COMPASS complexes may have a repressive role on transcription regulation. The increase of pSer2 was only observed in *wdr*-5 mutants, but not in the other single and double mutants. There are two models that could explain this observation. The first model is that WDR-5 has a repressive role apart of its role within the MLL/SET/COMPASS complex. Apart from forming the COMPASS complex, WDR-5 has been identified as a subunit within two acetyltransferase complexes, the non-specific lethal (NSL) HAT complex (Cai et al., 2010; Zhao et al., 2013) and the ADA2A-containing (ATAC) complex (Suganuma et al., 2008). Both of these two complexes exhibit an acetyltransferase activity on histone H4, especially at H4K16. Although acetylation at H4K16 is associated with transcriptional activation, it can also be linked to transcriptional repression (Suganuma et al., 2008). Thus, WDR-5 may have two separate roles on regulation of transcription: a positive role within the COMPASS complex, and a repressive role within other complexes. The second mechanism is that WDR-5 deposits H3K4 methylation on specific gene loci and recruit repressors to inhibit transcriptional elongation. Many H3K4 methylation downstream effectors can be involved in this model. For example, a tudor domain-containing demethylase JMJD2A can recognize H3K4me2/me3 and convert H3K9me3 to H3K9me2 (Huang et al., 2006). Increasing level of H3K9me2 is correlated with transcriptional repression (Shi and Whetstine, 2007). However, we have not identified yet which genes are up-regulated in *wdr*-5 mutant. Therefore, RNA-seq analysis needs to be carried out to investigate the repressive role of the COMPASS core components in the future.

Another key finding is the acquisition of germline characteristics in the somatic cells of the mutant embryos. Prior to this work, no evidence has been shown that the MLL/SET/COMPASS complexes are involved in maintaining the distinction between the somatic and germline lineages. Interestingly, H3K27 methylation, a repressive mark of transcription, is known to restrain somatic characteristics in embryos of *C. elegans* and to maintain the lineage specification (Yuzyuk et al., 2009). In addition, H3K27me3 and H3K4me3 co-localise at the promoters of certain developmental genes to form the “bivalent” mark
(Bernstein et al., 2006). Taken together, we hypothesise that depletion of H3K4me3 on developmental genes may switch on the H3K27me3 deposition on these genes, resulting in the acquisition of germline characteristics in somatic cells. Although germline characteristics, P granules expression and low developmental plasticity, were observed in the somatic cells of mutants’ embryos, we still cannot conclude that those somatic cells are actually transformed to germline cells, as more in vitro and in vivo evidences are required. Germline blastomeres share many characteristics with cancer cells (Shin and Mello, 2003). Thus our results may provide new insights into the pathology of tumorgenesis.

Importantly, the regulatory role of these core components on the lineage distinction is independent from the transcriptional elongation machinery. Our results show that inhibition of the transcriptional elongation process does not affect the P granules expression. Taken together, our study leads to a model that the core components of the SET/ MLL/COMPASS complex have two regulatory roles in the embryos (Figure 33). The first role is to regulate processivity of RNA Pol II; and the second role is to protect the somatic cells from adopting germline characteristics during the embryogenesis of C. elegans.

Figure 33. The regulatory role of the MLL/SET/COMPASS complexes in C. elegans’ embryos

The MLL/SET/COMPASS complexes are involved in two separate roles of regulation in embryos of C. elegans: to aid the RNA Pol II processivity, and to restrain the germline characteristics in somatic cells.
5.5 References


Chapter 6

General Discussion
In the work presented in this thesis, we endeavoured to investigate the individual contribution of WDR-5, RBBP-5 and ASH-2 towards H3K4 methylation, and the phenotypic consequences caused by removing these core components. In order to study this question, we used *C. elegans*’ embryogenesis and larval development to establish a system in which we could follow the relation between H3K4me1/me2/me3 and transcriptional states at a cellular resolution. In this system, we revealed that WDR-5, RBBP-5 and ASH-2 have different activities toward H3K4 methylation. Unexpectedly, RBBP-5 exhibits an antagonistic activity that prevents the deposition of H3K4me2/me3 during diapause L1 stage. In the last chapter, we uncovered that deposition of H3K4 methylation is crucial to RNA Pol II processivity and lineage specification, suggesting the importance of H3K4 methylation during embryogenesis. Herein, we will discuss our observations from the following aspects.

### 6.1 H3K4me3 is a lineage specific mark in early embryos of *C. elegans*

H3K4me3 is commonly known as an epigenetic mark of active genes. Prior to this work, no evidence has shown that the deposition of the H3K4me3 could be different between somatic and germline lineage during early embryogenesis of *C. elegans*. Our immunofluorescence data revealed that, in the *C. elegans* embryos, the deposition of H3K4me3 is robust in the somatic cells, but is low in the germline blastomeres and their somatic descendants. This observation raises the possibility that inhibitors may exist in germline blastomeres to prevent the germline blastomeres from H3K4me3 deposition. One candidate is the H3K4 demethylase. Although we have ruled out the RBR-2 as a H3K4 demethylase in the germline blastomeres, other unknown H3K4 demethylases may be involved in the regulation of H3K4me3 in the germline blastomeres. The other potential candidate is PIE-1, the determinant of the germline blastomeres (Mello et al., 1992a; Mello et al., 1996; Tenenhaus et al., 1998). During each asymmetric division, PIE-1 is primarily segregated into the germline blastomere, but a low or undetectable level of PIE-1 is left in its somatic sister (Reese et al., 2000). The distribution of H3K4me3 and PIE-1 seems to be mutually exclusive, which suggests that PIE-1 may be responsible for inhibiting H3K4me3 deposition. Indeed, our data show that removal of PIE-1 results in the accumulation of H3K4me3 in the germline blastomeres and their somatic sisters, suggesting that PIE-1 is a key inhibitor of H3K4me3 in germline blastomeres. The next
question is how does PIE-1 inhibit H3K4me3 deposition in the germline blastomerses? It is well
known that PIE-1 suppresses transcriptional elongation through blocking the interaction
between P-TEFb and CTD of RNA Pol II, thereby preventing phosphorylation on Ser2 residue
of RNA Pol II (Batchelder et al., 1999). Based on this model, we hypothesise that PIE-1 may
inhibit H3K4me3 deposition in germline lineage through directly interacting with the H3K4
methyltransferases MLL/SET/COMPASS complexes; or blocking the recruitment of
MLL/SET/COMPASS complexes to P-TEFb, or Pol II. However, according to our work,
removal of the catalytic component of P-TEFb, Cdk-9, causes a great loss of pSer2 deposition
on RNA Pol II, without affecting the H3K4me3 deposition in somatic cells, suggesting that the
MLL/SET/COMPASS complexes are not recruited by Cdk-9 or pSer2 at CTD to the
nucleosome (data not shown). Therefore, more evidences will be required to uncover the
mechanism of H3K4me3 regulation by PIE-1 in C. elegans embryos.

The role of H3K4me deposition in somatic lineage has been further investigated in Chapter 5.
Our data indicates that depletion of the MLL/SET/COMPASS complexes causes defects in
RNA Pol II processivity and lineage specification. Based on these observations, a model was
proposed to explain the mechanism of H3K4me-involved lineage specification (Figure 34).
This model suggests that the MLL/SET/COMPASS complexes are essential for the
maintenance of somatic cell program through depositing H3K4me1/me2/me3 and in turn
activating transcription. However, the catalytic activity of MLL/SET/COMPASS is repressed by
PIE-1 in germline blastomeres. The repression of the MLL/SET/COMPASS complexes allows
the expression of germline characteristics, like P granules expression and limited
developmental plasticity, in germline lineage.
Figure 34. The model of H3K4me-involved lineage specification of C. elegans

The somatic lineage (red background) is the lineage with active transcription and it is destined to die, while the germline lineage (green background) with inactive transcription is responsible for immortality of the species. H3K4 methylation is deposited by MLL/SET/COMPASS complexes (the Y-shape model) in somatic lineage, and therefore activates transcription. In addition, H3K4 methylation is also required for preventing somatic cells from adopting germline fate. However, in the germline lineage, the deposition of H3K4me is inhibited by PIE-1, which might directly interacts with SET/MLL/COMPASS complex. The presence of PIE-1 and absence of H3K4me leads to the onset of germline program.
6.2 The differential activities of WDR-5, RBBP-5 and ASH-2 may explain their differential roles on longevity

According to our observations in chapter 4 and 5, WDR-5 exhibits different, even opposite, activities from RBBP-5 and ASH-2, although all of them functions together within the same complexes. For example, RBBP-5 and ASH-2 is required for the pSer2 accumulation during embryogenesis, whereas WDR-5 has a repressive role on deposition of pSer2 on RNA Pol II; RBBP-5 and ASH-2 prevent the feeding-dependent accumulation of H3K4me2/me3 in PGCs during late embryogenesis and diapause L1, while WDR-5 has no effect on this process. These observations imply that deletion of WDR-5, RBBP-5 and ASH-2 may cause different phenotypic consequences in adulthood.

Previous work has shown that members of the MLL/SET/COMPASS complex regulate lifespan of *C. elegans*. Significant lifespan extension was observed in *ash-2 (RNAi)*, *wdr-5 (ok1417)* and *set-2 (ok952)* mutants. As ASH-2, WDR-5 and SET-2 are required for H3K4me3 deposition, the author suggests that deposition of H3K4me3 is detrimental for worm longevity (Greer et al., 2010). However, our preliminary data suggests that null deletion of RBBP-5 and ASH-2 shorten worm lifespan, and only deletion of WDR-5 extends lifespan. Our observations do not support the previous suggestion that H3K4me3 prevents longevity. According to our preliminary data, longevity is not simply correlated with H3K4me3 deposition.

We herein proposed two models to explain the longevity phenotype we observed in the core component mutants. Firstly, the premature accumulation of H3K4me2/me3 in PGCs of *rbbp-5 (tm3463)* and *ash-2 (tm1905)* accelerates the ageing process of *C. elegans*. In this model, RBBP-5 and ASH-2 may be involved in a nutrient-dependent signalling pathway, and food consumption is normally correlated with ageing (Piper and Bartke, 2008). The second model is that the up-regulation of pSer2 detected in *wdr-5 (ok1417)* activates a group of ageing-related genes to prolong lifespan. In the opposite, the decrease of pSer2 in *rbbp-5 (tm3463)* and *ash-2 (tm1905)* prevents the activation of those genes and leads to a short lifespan. Importantly, the regulation of longevity may be a combination of these two models.
6.3 How does the H3K4methylation affect transcription state?

Although H3K4 methylation is known as a mark of active genes, whether this mark affects gene transcription is unclear. Our study shows that the loss of H3K4 methylation is associated with embryonic lethality and a significant reduction of transcriptional elongation mark, pSer2 at CTD, suggesting that the H3K4 methylation is involved in the regulation of transcription. The embryonic lethality observed in double mutants implies that H3K4 methylation may be essential for transcription of developmental genes, such as Hox genes. The embryonic lethality was also previously observed by removing MLL1/MLL2/COMPASS complexes in mammal (Wang et al., 2009). In ESCs, MLL1/MLL2/COMPASS depletion causes a loss of H3K4me3 on the promoters of hox genes and a reduction of pSer2 on these genes, suggesting that H3K4me3 is required for regulating transcription of hox genes (Wang et al., 2009). However, C. elegans does not contain homolog of MLL1/MLL2/COMPASS complexes. Therefore, it is particular interesting to understand how the C. elegans MLL/SET/COMPASS regulate hox genes expression required for viability.

Many downstream effectors could be involved in the H3K4 methylation-dependent transcriptional regulation. One candidate is TAF3, a subunit of TFIID, which is crucial for transcriptional initiation. Decrease of H3K4 methylation impairs the recruitment of TAF3, and thereby causes a defect in transcriptional initiation to elongation conversion (Lauberth et al., 2013). Thus, it is possible that H3K4 methylation interacts with basal transcriptional machinery to influence gene expression.

6.4 H3K4me3 on the bivalent promoters might be crucial for maintaining the lineage specification

Germline characteristics, P granules, expression and low developmental plasticity were detected in the somatic cells of core components depleted embryos, indicating that the core components also play a role on lineage specification during C. elegans embryogenesis. This lineage specification defect was observed in all mutants, which share a common feature that the H3K4me3 is low during embryogenesis. It suggests that H3K4me3 may be an important mark for maintaining the somatic program. Moreover, the role of H3K4me3 on lineage specification seems to be independent from transcription, as we cannot detect this lineage
specification defect in the \textit{cdk-9 (RNAi)} and \textit{ama-1 (RNAi)} (Data not shown). One explanation is the “bivalent promoter” (Vastenhouw and Schier, 2012). H3K4me3 (positive mark) together with H3K27me3 (negative mark) can mark the bivalent promoters of lineage specific genes (Bernstein et al., 2006; Vastenhouw and Schier, 2012). Thus, it is possible that loss of H3K4me3 prompts the deposition of H3K27me3 on these lineage specific genes, and thereby causes an over-expression of germline-specific genes. To test this possibility, we can knock-down the H3K27 methyltransferases, the Polycomb proteins, to rescue this lineage specification defect. It is also possible that there are other downstream effectors of H3K4me3 involved in maintaining lineage specification in somatic cells. However, due to the time limitation of my project, I was unable to identify the downstream effectors of H3K4me3 responsible for lineage specification.

\subsection*{6.5 Future perspective}

As my project was designed to study the core components-dependent H3K4 methylation and its significance during embryogenesis and early larval stage of \textit{C. elegans}, the next objective is to further investigate the significance of the core components in the adulthood of \textit{C. elegans}. Our preliminary data has shown a longevity defect in mutants of the core components, thus more characterisations need to be carried out to confirm the role of the core components on ageing process of \textit{C. elegans}. Many signalling pathways are involved in regulation of ageing. It is important to understand which pathway the core components are involved in and how they differently regulate longevity. To answer this question, an RNAi screen is required to identify the candidates who potentially function with the core components to regulate longevity. Alternatively, the deposition patterns of H3K4me1/me2/me3 in \textit{wdr-5 (ok1417)}, \textit{rbbp-5 (tm3463)} and \textit{ash-2 (tm1905)} may be different. Understanding the deposition pattern of H3K4me1/me2/me3 in each mutant may reveal the roles of the core components on longevity. Thus, Chip-seq with antibodies against H3K4me1/me2/me3 needs to be performed in future.

Another novel finding of my project is the regulatory role of the core components on RNA Pol II processivity. This finding leads to a possibility that H3K4 methylation is involved in transcriptional regulation of genes. Moreover, the lethality observed in \textit{rbbp-5; ash-2} suggests
that a great loss of H3K4 methylation causes a defect in developmental genes’ expression, such as *hox* genes. Therefore, RNA-seq and RT-PCR techniques will be required to measure the transcription level of genes in mutants of the core components.

### 6.6 Conclusion

Epigenetic regulation has become a hot topic in the last decade, as mis-regulation of epigenetic marks is associated with many human diseases, such as cancer. Thus, it is crucial to understand how these epigenetic marks control DNA-based processes. The work presented in this thesis focus on the deposition of an epigenetic mark, H3K4 methylation, and its significance in embryogenesis of *C. elegans*.

In summary, our date demonstrates that:

1. H3K4me3 is a lineage specific mark during early embryogenesis of *C. elegans*, and can be inhibited by the germline blastomere determinant, PIE-1.

2. Each core component of the MLL/SET/COMPASS complexes contributes differently toward deposition of methylation at H3K4. RBBP-5 has a repressive role on H3K4me2/me3 deposition in PGCs during diapause L1 stage.

3. Deposition of H3K4 methylation is essential for viability, lineage specification and RNA Pol II processivity of *C. elegans* embryo, suggesting that it has an impact on transcriptional regulation.

As H3K4 methylation is highly conserved from yeast to human, this study is important for understanding the significance of epigenetic regulation during development and carcinogenesis in human.
Chapter 7

Appendix
I have contributed partially to the paper presented in this chapter. This paper uncovers that the double bromodomain protein, BET-1, and its sumoylation are both required for the maintenance of muscle myosin level of *C. elegans*. As the third author of this publication, I conducted all of the immunofluorescence against myosin in adult and embryo of *C. elegans*.

**Maintenance of muscle myosin levels in adult *C. elegans* requires both the double bromodomain protein BET-1 and sumoylation**

Kate Fisher1,*, Fiona Gee1,*, Siyao Wang1, Feng Xue1, Stefan Knapp3, Martin Philpott1, Christopher Wells2, Mirian Rodriguez1, L. Buster Snoke1, Jan Kammenga1, and Gino B. Poullet1

1Faculty of Life Sciences, Michael Smith Building, University of Manchester, Oxford Road, Manchester M13 9PT, UK
2Nutfield Department of Clinical Medicine, Old Road Campus Research Building, University of Oxford, Oxford OX3 7BN, UK
3Laboratory of Neurobiology, Wageningen Universiteit, Druivenlandsesteeg 1, 6708 PB, Wageningen, The Netherlands

*These authors contributed equally to this work.

*Author for correspondence (kfisher@manchester.ac.uk)

Biology Open 1: 1354–1363
doi: 10.1242/bio.20130897
Reserved 2nd July 2013
Accepted 7th October 2013

Summary

Attenuation of RAS-mediated signalling is a conserved process essential to control cell proliferation, differentiation, and apoptosis. Cooperative interactions between histone modifications such as acetylation, methylation and sumoylation are crucial for proper attenuation in *C. elegans*, implying that the proteins recognising these histone modifications could also play an important role in attenuation of RAS-mediated signalling. We sought to systematically identify these proteins and found BET-1. BET-1 is a conserved double bromodomain protein that recognises acetyl-lysines on histone tails and maintains the stable fate of various lineages. Unexpectedly, adults lacking both BET-1 and SUMO-1 are depleted of muscle myosin, an essential component of myofilaments. We also show that this muscle myosin depletion does not occur in all animals at a specific time, but rather that the penetrance of the phenotype increases with age. To gain mechanistic insights into this process, we sought to delay the occurrence of the muscle myosin depletion phenotype and found that it requires caspase activity and MEK-dependent signalling. We also performed transcription profiling on these mutants and found an up-regulation of the FGF receptor, egl-15, a tyrosine kinase receptor acting upstream of MEK. Consistent with a MEK requirement, we could delay the muscle phenotype by systemic or hypodermal knock down of egl-15. Thus, this work uncovered a caspase- and MEK-dependent mechanism that acts specifically on ageing adults to maintain the appropriate net level of muscle myosin.

© 2013. Published by The Company of Biologists Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Key words: Bromodomain, Sumo, Body wall muscle, *C. elegans*
SUMO and BET-1 maintain muscle myosin levels in adults

We set out to identify readers that recognize chromatin modifications and genetically interact with the sumoylation pathway to prevent hyperactivation of the LET-60 signalling cascade. We used RNAi to deplete all predicted readers and identified CHD-3, HPL-2, and BET-1. CHD-3 and HPL-2 are chromodomain proteins recognizing methylated histone tails and were previously shown to play a role in LET-60 attenuation (Connolly et al., 2006; Solari and Alarcón, 2008). BET-1 is a conserved double bromodomain protein of the BET family required for establishment and maintenance of stable fate in various lineages (Shibata et al., 2010). BET-1 shares homology with both human BRD2 and BRD4, and is a likely homolog of BRD4 because of a putative P-TEFb interaction motif not present in BRD2 (Bisgrove et al., 2007). BET-1, like other BET proteins, physically associates with acetyl-lysines on histone tails (Shibata et al., 2010).

Low molecular weight inhibitors such as JQ1 and I-BET151 can efficiently target acetyl-lysine binding sites of BET proteins (Dawson et al., 2011; Delmore et al., 2011; Filipakopoulos et al., 2010; Nicoulenec et al., 2003; Zuber et al., 2011). In caudal myotrons, the inhibition of BRD4 leads to downregulation of the euceneogens aUTC and other growth promoting and apoptotic genes (Delmore et al., 2011). This specific transcriptional regulation has recently been attributed to the effect of BRD3 on super-enhancers (Loven et al., 2013).

Herein we performed a targeted RNAi screen and identified BET-1 as a novel SUMO interactor. Unexpectedly, we found that SMO-1 and BET-1 act together to maintain net muscle myosin levels in ageing adults. We show that muscle myosin depletion requires caspase activities and the FQR receptor/MEK signalling pathway to manifest. Interestingly, I-BET151 and JQ1 are activated under muscle catabolic conditions induced by insulin resistance (Du et al., 2004).

Materials and Methods

Strains and general maintenance

Strains were maintained at 20°C as described (Bangay, 1974), unless stated. For full list of strains see supplementary material Table S2. Of note, the muscle phenotype has been analysed using either bet(hsp46) or bet(hsp46.625) but all processed data are with bet(hsp46).

Identification of putative readers of chromatin marks

To identify genes for use in the targeted RNAi screen, the Predictomics database (release WS1996), using the WormBase data mining tool.

RNAi experiments

RNAi screens of the ~250 gene set were performed similarly to those described previously (Karkanis et al., 2003; Poulin et al., 2005). Briefly, individual cultures were used to inoculate three wells on a six-well plate (Poulin et al., 2005), around 10 concentrated smo-1/2 l3.4 d7 stage worms were placed in the upper well for each bacterial strain and the plates maintained at 20°C. After 48 h, 5 worms from the upper well were transferred to the lower well. The F3 progeny were scored for the Msp (multiple vertical projection) phenotype. RNAi clones giving Msp in one or more smo-1/sm-2 animals and none or two or more smo-1/sm-2 animals were selected for further analysis. These criteria took into account the 1:2 background Msp in smo-1/2 smo-3/4 animals. RNAi clones were obtained from the Abelagen RNAi library (Kanchan et al., 2002) and the Vidal RNAi library (Ruhl et al., 2004). All positive RNAi clones were verified by sequencing. All further RNAi experiments were performed similarly as described above.

Immunofluorescence staining of embryos and muscle myosin

Immunofluorescence of muscle myosins on embryos or adults was performed by freeze-crack method as previously described (Wang et al., 2013) with the following adaptation for staining of embryos: each mother was cut open in the middle using a sharp needle. Four-day-old adults (wild type or mutants) were grown from L1 on OP50, picked onto slides and freeze-cracked. Slides were incubated with anti-myosin heavy chain A (5-6 (1:100) or 5-15 (1:100), Developmental Studies Hybridoma Bank, University of Iowa) overnight at 4°C in a humid chamber.

Following washes slides were then incubated with Dylight 594 AffiniPure Goat Anti-Mouse IgG (H+L) (1:200) (cat (115-515-146), Jackson Immunoresearch). Primary were then washed in PBS and myosin was visualized with DAPI. The image data and Immunofluorescence recovery were exported from the microscope control software (ZEISS 2005) into GraphPad Prism to determine the half-time for full recovery for individual cells and averages were calculated from 15-20 cells per treatment point.

Western blot analysis

Whole worm protein extracts at the indicated stages were prepared by harvesting synchronized worms, washing the pellet in 1× PBS and boiling in 1× sample buffer containing 100 mM DTT and 0.1% tri-n-butylphosphine. Quantification performed using ImageJ software.

Microarrays

Sample preparation: Whole worm extracts for microarray analysis were prepared by placing ~50 mothers on 10 cm NGM plates for each strain. Gravid progeny were bleached and the eggs put on NGM plates with no food. After 24 hours the synchronized L1 were washed off and placed on plates with food. Once the F1 progeny had reached L4 the worms were harvested, washed once in MP buffer and frozen at −80°C.

RNA preparation: Two replicates of bet(hsp46), smo-1/2 mutants, and smo-1/2 double mutants and the corresponding balanced strains as a control were processed for microarray analysis. Nematode pellets were incubated with 3% beta-mercaptoethanol and 800 μg/mL proteinase K at 55°C, 500 rpm shaking for 60 minutes. Total RNA was extracted from these pellets using RNeasy Micro kit (Qiagen) according to manufacturer instructions.

Microarray analysis: The extracted RNA was processed for microarray performance; the platform used for that purpose was C. elegans C22 Gene Expression Microarray 4×4H (Agilent technologies), following manufacturer instructions. Raw data (supplementary material Table S3) was extracted from the scanned images by the Agilent feature extract software. Data was normalized in the statistical programming environment “R” using the limma package (Smyth, 2004). For within array normalization we used the Lowess method and for the between array normalization we used the quantile method. No background correction was needed. A linear model was used to determine the differently expressed genes.

Quantitative RT-PCR

Worm pellets were prepared by harvesting synchronized L4 worms, washing the pellet in 1× PBS and freezing at −80°C. Total RNA was extracted from these pellets using TRIzol (Invitrogen) and first strand cDNA synthesis was performed using the SuperScript VILO (Invitrogen) kit (Invitrogen), according to the manufacturer’s instructions. Quantitative RT-PCR was performed using the FastStart SYBR Green Master (ROX) mix (Roche) on a StepOnePlus Real Time PCR System (Applied Biosystems). Two biological samples for each strain were prepared, and for every biological replicate, a triplicate of two serial dilutions was analysed. axtv was used as the internal reference for data normalization. qRT-PCR levels were determined by comparing the unknown samples to a standard curve of known relative amounts. Primers used are listed in supplementary material Table S4.

JQ1 and U0126 treatments of nematodes

Synchronized L1 larvae were transferred onto NGM plates prepared with the indicated concentration of JQ1 (Filipakopoulos et al., 2010) or U0126 (Almog et al., 2010).

Results

BET-1 genetically interacts with SUMO

Deposition and removal of post-translational modifications (PTMs) on histone tails play an important role in
SUMO and BET-1 maintain muscle myosin levels in adults

transcriptional regulation, which in turn impacts on the process of attenuation of LET-60-mediated signalling in C. elegans (Cui et al., 2006; Lipscck, 2004; Lu and Horvitz, 1998; Poulin et al., 2005). These FMRs can act by either altering the electrostatic interactions between histones and DNA or by creating recognition sites for specialised proteins often referred to as readers of the epigenetic code (Kozanides, 2007). To test the latter mode of action, we performed an RNAi screen targeting all known predicted readers (~200 genes; supplementary material Table S1). Since the sumoylation pathway has been shown to genetically interact with many chromatin complexes involving attenuation of LET-60 (Leight et al., 2005; Poulin et al., 2005), we performed the screen in a SUMO-compromised strain (sno-1/lp). We selected our candidates according to the observation of SUMO-associated phenotypes. The main expected phenotype being the multivulva (MUV) phenotype (Breday et al., 2004; Leight et al., 2005; Poulin et al., 2005) and its superficial manifestation the multiventral protrusion phenotype (MVP) (Fisher et al., 2010); these phenotypes indicate a hyperactivation of the LET-60/LIN-45/MEK-2/MPK-1 signalling cascade. We identified three candidates: two conserved chromodomain proteins: CHD-3 (Solari and Ahringer, 2000) and HPL-2 (Cousshart et al., 2006); and the double bromodomain BET-1. We focused this study on BET-1, which was previously shown to recognize acetyl-lysines on histone tails and to maintain cell fate in various lineages (Shibuta et al., 2010).

BET-1 and SUMO prevent muscle myosin depletion in adults

Following the identification of BET-1 by RNAi screening, we generated a null mutant sno-1/lp bet-1/lp and assessed whether we could detect a genetic interaction during vulva development. Surprisingly, we could not find an interaction in the vulva. Instead, we found that the single bet-1/lp mutant or RNAi against bet-1 can produce the Muv (multiple vulva) phenotype, but the additional loss of sno-1/lp does not aggravate the Muv phenotype (data not shown). However, during these investigations, we noticed that an important proportion of these double sno-1/lp bet-1/lp mutants lost their ability to crawl early in adulthood. We quantified this observation by assessing loss of locomotion. In this established assay (Herndon et al., 2002), locomotion can fall into three exclusive categories: fully mobile (A), mobile following prodding (B), immobile (C). This latter category is defined by animals incapable of producing a full body movement following prodding whilst remaining alive, which is determined by head movements. We measured loss of locomotion from larval stage 1 (L1) for each single mutant, the double mutant, and wild type. We found that a proportion of the single bet-1/lp or sno-1/lp mutant lost locomotion earlier than wild type (45.5%, n=44; and 23%, n=48, respectively). However, the decline in locomotion is accelerated and more penetrant in the double sno-1/lp bet-1/lp mutants (73%, n=45). The double sno-1/lp bet-1/lp mutants are statistically different (Fisher's exact test) from N2 and the singles (P=4x10^{-9} compared with N2), 0.1x10^{-7} (compared with bet-1/lp), 3x10^{-7} (compared with sno-1/lp). Fig 1A; supplementary material Fig S1). Further, to provide an overview of the life history of each population, we present the entire dataset as box plots from day 1 till day 21. The number of days spent in each category for each strain at a specific time point can be selected (supplementary material Movie 1) or alternatively the entire 21-day assay can be viewed as a movie (supplementary material Movie 1) showing how the population for each category changes according to time. In conclusion, the locomotion assay shows that BET-1 and SUMO can act individually and in cooperation to prevent loss of locomotion in adults.

Time-associated loss of locomotion is a natural phenomenon attributed to a failure of maintaining a functional ageing muscle mass (Herndon et al., 2002). We therefore explored the possibility that bet-1/lp and sno-1/lp mutants are experiencing a decline in muscle functionality. To test this, we directly measured muscle myosin (MYO-3), a major component of myofibrils, using immunofluorescence stainings and Western blot analysis. We first surveyed the levels of adult muscle myosin by immunostainings on wild type animals versus singles and double mutants. We used a modified freeze-crack method, which involves cutting open the mothers roughly in the middle, allowing the antibody to penetrate thoroughly the muscles. All wild type animals are successfully stained using this adaptation. At day one of adulthood, we found that the MYO-3 signal was depleted in 4.7% of double sno-1/lp bet-1/lp mutants while all wild type animals and the respective single mutants all remained positive for MYO-3 (Fig 1B). Interestingly, at day four of adulthood, the phenotype reached a penetrance of 37.5% for sno-1/lp bet-1/lp animals. A slight increase was observed with single mutants (bet-1/lp mutants: 2.7%; sno-1/lp mutants: 4.8%) but no effect was found on wild type (Fig 1B; Table 1). One day later at day five, the double sno-1/lp bet-1/lp reached 42.5% and the respective single mutants about 10% (Fig 1B). Depletion was never observed in late embryos (Fig 1C). Taken together, these experiments indicate a time-associated defect in net levels of muscle myosin for all mutants.

We next verified whether the observations obtained by immunostainings could also be detected by Western blot analysis, hence eliminating possible artefacts due to the staining procedure. We prepared whole worm extracts from double sno-1/lp bet-1/lp mutants and compared with extracts from wild type animals at three time points: L4, day-one adults and day-four adults. We found that muscle myosin levels are similar at L4 and day-one adults, but depleted at day-four adults (Fig 1D). We quantified this effect relative to wild type sample and found that levels of myosin are down to 68.8% in average for sno-1/lp bet-1/lp samples (SEM 6.5%; n=4). This is consistent with our immunostaining experiments (Fig 1B) and thus validates the occurrence of the muscle myosin depletion phenotype. We also observed degradation products from day-one adults, but only in wild type animals, we do not know the significance of this reproducible effect. Of note, we also confirmed the muscle myosin depletion phenotype using a different antibody against MYO-3 (supplementary material Fig S2A). Moreover, we used an antibody against pararxysmin (supplementary material Fig S2B/C) and observed the same effect, i.e. immunofluorescence shows that ~30% of double sno-1/lp bet-1/lp mutants are depleted in pararxysmin and Western blot analysis shows that pararxysmin is down to 66% in average (SEM 2.8%; n=4).

Finally, our data from immunostainings and Western blot analysis show that the muscle myosin depletion phenotype has an onset in adults, but it does not rule out that the causal defect occurs prior to adulthood, since the genetic delets are always present. To circumvent this problem, we performed acute sno-1/lp(RNAi) treatments on young bet-1/lp adults followed by immunostainings against MYO-3 on day four adults. We observed that 28.1% (n=34) of bet-1/lp sno-1/lp(RNAi) are MYO-3 depleted compared with 0% (n=32) for the control RNAi
SUMO and BET-1 maintain muscle myosin levels in adults

**Fig. 1.** Loss of locomotion and depletion of muscle myosin in smo-1/ bet-1<sub>Y</sub> animals. (A) Histograms of cloned individual animals followed using a locomotion assay from larval stage one (L1) until day four (D4) adult showing that locomotion of the double mutants is aggravated by the loss of both BET-1 and SMO-1. Blue depicts crawling animals, red, animals requiring prodding; green, immobile animals; and purple, dead animals. (B) Immunostaining against MYO-3 showing double smo-1/ bet-1<sub>Y</sub> mutants displaying a high frequency of the depletion of muscle myosin phenotype at day four and five adult. (C) Western blot analysis using an anti-MYO-3 antibody showing that at day four adult, MYO-3 is depleted. Ten age-matched animals were used for each lane. We repeated the experiments four times and a representative blot is shown. Scale bars: 100 μm (B), 10 μm (C).

**treatment (P=0.001; Chi square test of association).** This experiment using acute RNAi treatments against smo-1 on young bet-1<sub>Y</sub> adults shows that the depletion of muscle myosin phenotype can occur after the establishment of muscle development and therefore provide further evidence that the depletion of muscle myosin is consistent with a defect during adulthood.

**Caspase-dependent depletion of muscle myosin**

Muscle myosin levels are regulated by both synthesis and proteolysis. Since the phenotype described herein has an onset in adulthood and that the bulk of muscle myosin is synthesised prior to adulthood, we hypothesised that the depletion of muscle myosin is more consistent with excessive degradation of muscle myosin. There are four proteolysis systems described for mammalian muscles: the proteosome (Mitch and Goldberg, 1996), the lysosome (Sandri, 2013), calpains (Sortimachi and Ono, 2012) and caspases (Du et al., 2004). The first three have been shown to function in C. elegans (Etheridge et al., 2012). We reasoned that the proteosome and lysosome systems are unlikely to be the primary system acting on muscle myosin because of their inefficiency at directly targeting myofilaments components, such as muscle myosin (Du et al., 2004). Calpains are activated by disruption of the integrin attachment complex (Etheridge et al., 2012), which produces a very different muscle phenotype than the muscle myosin depletion phenotype. We therefore investigated whether a caspase-mediated system could be erroneously activated in smo-1<sub>Y</sub> bet-1<sub>Y</sub> double mutants. To test this, we blocked the caspase cascade in smo-1<sub>Y</sub> bet-1<sub>Y</sub> mutants using ced-3 or ced-4 loss of function mutants. CED-3 (the
Table 1. Immunostaining analysis of muscle myosin levels at day 10 adulthood to assess the muscle myosin depletion phenotype. Not applicable (n/a) for egl-15 mutant or RNAI, since the worms are egg laying defective and eventually burst before any analysis can be conducted. Sterility avoids this problem. The * indicates a P value<0.01 by Chi square test of association. Denm: Depletion of muscle myosin.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>N</th>
<th>% Deem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>smo-1(f)</td>
<td>63</td>
<td>3</td>
</tr>
<tr>
<td>bet-1(f)</td>
<td>63</td>
<td>3</td>
</tr>
<tr>
<td>smo-1(f) bet-1(f)</td>
<td>104</td>
<td>38</td>
</tr>
<tr>
<td>cod-3(f)</td>
<td>69</td>
<td>3</td>
</tr>
<tr>
<td>cod-4(f)</td>
<td>61</td>
<td>0</td>
</tr>
<tr>
<td>smo-1(f) bet-1(f) cod-3(f)</td>
<td>55</td>
<td>3</td>
</tr>
<tr>
<td>cod-6(f)</td>
<td>55</td>
<td>3</td>
</tr>
<tr>
<td>EV (RNAi)</td>
<td>223</td>
<td>0</td>
</tr>
<tr>
<td>EV (RNAi); smo-1(f) bet-1(f)</td>
<td>162</td>
<td>34</td>
</tr>
<tr>
<td>spar-7 (RNAi); smo-1(f) bet-1(f)</td>
<td>129</td>
<td>22</td>
</tr>
<tr>
<td>sur-7 (RNAi); smo-1(f) bet-1(f)</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>evp-13 (RNAi); smo-1(f) bet-1(f)</td>
<td>98</td>
<td>10</td>
</tr>
<tr>
<td>egl-15 (RNAi); smo-1(f) bet-1(f)</td>
<td>118</td>
<td>0</td>
</tr>
<tr>
<td>egl-15 (RNAi)</td>
<td>190</td>
<td>15</td>
</tr>
<tr>
<td>egl-15(f)</td>
<td>9/7</td>
<td>0</td>
</tr>
<tr>
<td>cl-1(f)</td>
<td>64</td>
<td>0</td>
</tr>
</tbody>
</table>

downstream caspase) and CED-4 (the apoptotic protease-activating factor) are required for most apoptotic events occurring in C. elegans (Ellis and Horvitz, 1986; Mora et al., 1993). The triple smo-1(f) bet-1(f) cod-3(f) or smo-1(f) bet-1(f) cod-4(f) mutants were analysed by immunostaining against muscle myosin. We observed that most of these caspase-defective triple mutants maintain muscle myosin levels at day four adult (CED-3: 98% and CED-4: 95%; Fig 2A; Table 1). We verified this result using Western blot service against MYO-3 and confirmed that muscle myosin levels are in average higher in absence of SMO-1, BET-1 and CED-3 than in absence of SMO-1 and BET-1 (127%, SEM 3%, n=4; Fig 2B), albeit at levels remaining below the wild type levels (Fig 2B). Similar results were obtained using the anti-paramyosin antibody (131%, SEM 8.6%, n=3; supplementary material Fig S2D). Furthermore, we performed the locomotion assay on the triple smo-1(f) bet-1(f) cod-3(f) mutants and observe a significant increase in the proportion of animals retaining locomotion for the triple smo-1(f) bet-1(f) cod-3(f) mutants, when compared with the double smo-1(f) bet-1(f) mutants (from 67%, n=45 to 27%, n=45 at P=0.0003; Fisher’s exact Test) (Fig 2C), indicating that inactivation of the caspase system causes a delay rather than a rescue of the loss of locomotion phenotype. We also present the delay effect as animated box plots for each locomotion category and strain (supplementary material Movie 2). Taken together, these results provide strong evidence that a caspase system is functional to maintain muscle homeostasis in C. elegans.

The muscle myosin depletion phenotype is MEK-dependent. Both somatostatin and BET-1 are important regulators of transcription. We therefore postulated that changes in their transcription profiles could provide insights into the muscle myosin depletion phenotype. However, we encountered a technical problem; we could not extract sufficient materials from single or double smo-1(f) bet-1(f) homozygous escapers. To palliate to this problem...
issue, we instead used heterozygotes (see materials and methods). As anticipated, transcription is not strikingly affected in these heterozygous animals. However, focusing our analysis on known components of the LET-60 signalling pathway, we found two possibly up-regulated positive regulators of the pathway: egl-15, (the FGFR receptor; DeVore et al., 1995) and sur-7 (a cation transporter (Yoder et al., 2004)). To verify these results, we performed quantitative RT-PCR on independent biological samples (in the double smo-1 cargo heterozygous background) and tested the levels of expression of six genes, including the egl-15 and sur-7 candidates (Fig. 3B). We found that both egl-15 and sur-7 are up-regulated significantly by about 2- and 1.5-fold, respectively (Fig. 3B), egl-15 showing a very strong p-value (Fig. 3B). We also detected a slight up-regulation (~1.5-fold) for let-60 (Fig. 3B). In contrast, ctx-1, ptp-2 and egl-17 are not significantly affected (Fig. 3B). The up-regulation of egl-15 is of interest since the FGF receptor can activate the LET-60/MEK signalling pathway and when hyperactivated increase muscle cells proteolysis as measured using a reporter assay (Kecel et al., 1998; Sundararam, 2006; Szewczyk and Jacobsen, 2003). On the other hand, SUR-7 has not been linked to proteolysis in muscle cells, yet it acts as a positive regulator of LET-60/MEK signalling by regulating levels of cytoplasmic zinc ions through sequestration in the endoplasmic reticulum (Yoder et al., 2004). It is also known that elevated levels of zinc ions increase phosphorylation of the scaffold protein KSR, preventing its association with RAF and MEK and instead favouring an inhibitory association with I3-3 (Mueller et al., 2001). Thus, up-regulation of both EGL-15 and SUR-7 are consistent with an increase in LET-60/MEK-mediated signalling.

These results from the expression profile data suggested the possibility that the absence of BET-1 and SMO-1 could lead to

**Fig. 3.** Up-regulation of egl-15 and sur-7 in smo-1 cargo heterozygous mutants. (A) Transcription profiles obtained from microarrays of smo-1 cargo heterozygous and wild type larvae. The balance strain (k237) is used as a reference. Displayed as a color-coded matrix are the results from 56 genes known to act within the LET-60 signalling pathway. Black indicates no change; blue and yellow indicate down- and up-regulation in log2 scale, respectively. The raw data and calculated p-values are available in supplementary material Table S3. Microarray analysis for differential expression was performed using the LIMMA (Linear Models for Microarrays (Smyth, 2004)) package with Bioconductor R. (B) Quantitative RT-PCR confirming up-regulation of egl-15 and sur-7 in smo-1 cargo heterozygous mutants. At least three independent samples were used. P values calculated using the Student T-test and error bars are ± SEM.
hypereactivation of the EGL-15/LET-60/LIN-45/MEK-2 signalling pathway that in turn could initiate muscle myosin depletion. We dampened the LET-60 signalling pathway using the MEK inhibitor U0126 (Morgan et al., 2010) and measured the effect on muscle myosin levels at day four adult. Remarkably, we found that U0126-treated smo-1p bet-1p double mutants can maintain muscle myosin levels. 52% of DMSO-treated smo-1p bet-1p animals displayed depletion of muscle myosin compared with 2% of U0126-treated animals (Fig. 4A). Importantly, this experiment indicates that the conserved FGF receptor/RAS/RAF/MEK signalling pathway is required for the muscle myosin depletion phenotype to manifest. It is also consistent with other studies linking hypereactivation of the LET-60 signalling pathway with protein degradation in muscles (Szewczyk and Jacobson, 2003; Szewczyk et al., 2007; Szewczyk et al., 2002).

Since we found that egl-15 and sur-7 are overexpressed in double mutants (Fig. 3A), we next addressed specifically whether these could play an important role in the muscle myosin depletion phenotype. To this end, we depleted smo-1p bet-1p double mutants of egl-15 by either performing RNAi or using a reduced function allele; we found that the penetrance of the myosin depletion phenotype is decreased by 56%, from 34% to 15%, (P=0.001; Table 1) and by 35%, from 34% to 22%, (P=0.042; Table 1), respectively. We also tested sur-7 (RNAi) and observed a decrease in the penetrance by 76%, from 34% to 11% (P<0.001; Table 1). Thus, the overexpression of EGL-15 and SUR-7, the rescue experiments by the MEK inhibitor, and depletion of either EGL-15 or SUR-7 together provide strong evidence that the EGL-15/LET-60/MEK signalling pathway is required for the muscle myosin depletion phenotype to fully manifest.

Hydopidermal depletion of egl-15 or sur-7 rescues the muscle myosin phenotype

It has been previously shown that the EGL-15/LET-60/LIN-45/MEK-2 signalling cascade, in addition to producing the muscle cell phenotype, can also cause a Ctr phenotype (Huang and Stern, 2004; Koele et al., 1998). Of note, cdc-1 mutants do not display the muscle myosin depletion phenotype (Table 1) and are morphologically different from the double smo-1p bet-1p mutants in the phenotypic analysis (Fig. S2). The exact relationship between the muscle phenotype and the Ctr phenotype remains unclear. However, the anatomical locus of activity for the EGL-15 signalling cascade, to produce the Ctr phenotype, is the hypoderm (Huang and Stern, 2004) rather than the muscle itself. With this in mind, we sought to identify the tissue in which EGL-15 and SUR-7 are required to produce muscle myosin depletion. To this end, we performed hydopidermal and body wall muscle specific RNAi against egl-15 and sur-7. This established tissue specific RNAi system takes advantage of an RNAi insensitive strain lacking RDE-1, in which tissue-specific re-expression of RDE-1 reactivates RNAi sensitivity in the targeted tissue (Quodota et al., 2007). We assessed whether knocking down egl-15 or sur-7 in either the hypoderm or the body wall muscles could rescue the muscle myosin depletion phenotype. We found that only hydopidermal RNAi, of either egl-15 or sur-7, can do so. Depleting EGL-15 or SUR-7 reduces the penetrance by 47% and 50%, from 32% down to 17% (P=0.016) and 16% (P=0.011), respectively (Table 2). No rescuing effect could be detected by targeting either egl-15 or sur-7 in body wall muscles. Since we cannot rule out discrepancies in RNAi efficiency, we cannot rule out the possibility of a muscle activity (Table 2). Despite this caveat, these data show that hydopidermal EGL-15 (and SUR-7) signalling is implicated in the depletion of muscle myosin phenotype.

JQ1-treated SUMO mutants display the muscle myosin depletion phenotype

So far we have shown that BET-1 acts together with the SUMOylation pathway to prevent muscle myosin depletion in adults. We next wanted to address whether the recognition of acetyl-lysines is important in the depletion of muscle phenotype. We blocked reading of acetyl-lysines using a small molecule compound inhibitor of BET proteins, JQ1 (Dawson et al., 2011;
Table 2. EGL-15 and SUR-7 are active in the hypoderm. The depletion of muscle myosin D genu phenotype was assessed using immunostaining against muscle myosin. Significant p values of 0.011 and 0.0162 by chi square test of association.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>N</th>
<th>% Denim</th>
</tr>
</thead>
<tbody>
<tr>
<td>EV(RNAi) ; smo-1 bet-1 rol-6; kia9 (hypoderm)</td>
<td>104</td>
<td></td>
</tr>
<tr>
<td>egl-13(km8); smo-1 bet-1 rol-6; kia9 (hypoderm)</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>sur-17(km8); smo-1 bet-1 rol-6; kia9 (hypoderm)</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>EV(RNAi) ; smo-1 bet-1 rol-6; kia9;20 (muscle)</td>
<td>152</td>
<td></td>
</tr>
<tr>
<td>egl-13(km8); smo-1 bet-1 rol-6; kia9;20 (muscle)</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>sur-17(km8); smo-1 bet-1 rol-6; kia9;20 (muscle)</td>
<td>82</td>
<td></td>
</tr>
</tbody>
</table>

Delmore et al., 2011; Filippopoulos et al., 2010; Nicodeme et al., 2010; Zuber et al., 2011). If recognition of acetyl-lysines is involved, we should detect the muscle myosin depletion phenotype when sma-1 forward mutations are treated with increasing amount of JO1 (2.5, 10 and 25 μM). Using immunostaining, we found that JO1-treated sma-1 forward animals indeed display the muscle myosin depletion phenotype (Fig. 3 B). JO1 treatment of wild type animals did not cause the depletion of muscle myosin phenotype (data not shown). From this, we concluded that recognition of acetyl-lysines is important to prevent depletion of muscle myosin in adults. Since it is likely that most of these recognised acetyl-lysines are on histone tails, the data suggest that the muscle myosin depletion phenotype implicates a defect at the epigenetic level. Of note, we confirmed that JO1 can block BET-1’s association with acetyl-lysines on histones using FRAP (supplementary material Fig. S4), in accordance with another study showing that BET-1 can associate with acetylated histones (Shibata et al., 2010).

Discussion
This study provides novel mechanistic insights into the pathways that ensure maintenance of muscle myosin levels in ageing adults and likely to influence the complex behaviour of locomotion. We present a novel muscle phenotype characterised by the depletion of adult muscle myosin. Our investigation shows a number of specific characteristics associated with this phenotype: it is caspase- and MEK-dependent, it requires hypodermal EGL-15 activity, and the muscle myosin depletion is observed only in adults and getting progressively more severe as the animals are ageing.

Transcriptional regulation of EGL-15 and SUR-7
We have found that in absence of BET-1 and SMO-1 the FGF receptor, egl-15, and the cation diffusion facilitator sur-7 are up-regulated. Since BET-1 associates with acetyl-lysines on histone tails (supplementary material Fig. S4) (Shibata et al., 2010), it is a possibility that their expression are regulated by this histone modification and therefore implicating an epigenetic mechanism. This possibility is consistent with our experiments showing that the ability to recognize acetyl-lysines is crucial to prevent muscle myosin depletion in adults (Fig. 4B, C). However, the up-regulation of both sur-7 and egl-15 (Fig. 3) suggests that acetyl-lysines could be interpreted as a signal for repression by BET-1, even though acetyl-lysines on histone tails are generally associated with activation of transcription. An alternative explanation for this repression effect is that BET-1 could be required to maintain the expression of a repressor that in turn acts on sur-7 and egl-15. Further work will be needed to distinguish between these mechanisms.

Non-cell autonomous EGL-15 activity
It is intriguing that the depletion of muscle myosin phenotype is apparent only in adults (in non-dividing cells), suggesting an important role for BET-1 and SMO-1 in muscle myosin homeostasis. Furthermore, we show that this phenotype is likely to involve a non-cell-autonomous mechanism. Interestingly, previous mosaic analysis on the Clear phenotype caused by hyperactivation of LET-60 signalling revealed that the anatomical locus of activity for the EGL-15 signalling cascade is hypodermal (Huang and Stern, 2004). Similarly, muscle myosin depletion is influenced by hypodermal EGL-15 activity (Table 2). Even though it is not obvious how hyperactivation of the EGL-15 signalling in the hypoderm can lead to muscle myosin depletion, there is a physical association between the muscles and the hypoderm. A recent report has shown that calpains mediate integrin attachment complex maintenance of adult muscles in C. elegans (Eatheridge et al., 2012). Integrin attachment complexes fulfil multiple functions in muscles (Moerman and Williams, 2006), one of which is to anchor body wall muscles to the basement membrane. Since hypodermal cells are also linked to the basement membrane (Moerman and Williams, 2006), this physical association could mediate signalling events between muscles and hypodermis. Hence, hyperactivation of the EGL-15/LET-60/MEK signalling pathway in the hypoderm could produce a defect in signalling events between muscle and hypodermis, leading to the muscle myosin depletion phenotype.

Premature loss of locomotion and depletion of muscle myosin
We have found that loss of locomotion occurs prematurely in single and especially in double sma-1, bet-1 mutants. Loss of locomotion can be observed before the depletion of muscle myosin. This sequence of events strongly suggests that another function impacting on locomotion is impaired in these mutants. Locomotion is a complex behaviour involving muscles, neurons and muscle attachments to the cuticle via the hypoderm. We have shown that a defect in signalling occurs in the hypoderm and that muscle cells are depleted in muscle myosin. It remains to be investigated whether neurons are affected, since BET-1 has been shown to act in stabilising neuronal cell fate (Shibata et al., 2010). It is also a possibility that the loss of locomotion is the primary defect, leading to the depletion of muscle myosin. There is however a discrepancy between the percentage of animals losing locomotion and the percentage of animals depleted in muscle myosin; there are at least twice as many animals losing locomotion that there are animals depleted in muscle myosin. Further, inactivation of CED-3 in sma-1, bet-1 mutants allowed maintenance of muscle myosin levels. However, we observe the same discrepancy aforementioned between loss of locomotion at day four and depletion of muscle myosin. Thus, the depletion of...
SUMO and BET-1 maintain muscle myosin levels in adults

muscle myosin, at day four adult, appears unlikely to be induced by loss of toeconomom. Unless a number of animals have been in an immobilised state longer than others prior to analysis and that those particular animals are depleted in muscle myosin. Finally, the muscle myosin depletion phenotype that we described produces an effect resembling muscle atrophy. However, it is unclear whether this phenotype is actually a premature manifestation of sarcopenia, the loss of muscle mass due to aging, which occurs naturally in C. elegans (Herndon et al., 2002) or whether it is a muscular pathology. Taken together, our study has identified bet-1 and smo-1 as important players in the maintenance of adult muscle myosin levels through a caspase- and MEK-dependent mechanism, which could be relevant to muscle ageing and/or a muscular pathology.

Acknowledgements
We would like to thank Prof. Andrew Sharrock, Prof. Stephen Taylor, and Dr Verena Wolffram for critical reading of the manuscript, Azorny Le for RNAi clones, Prof. Hiroshi Soga for bet-1 strains, and Prof. Nathaniel Szezweyk for helpful discussion. Some strains were provided by the CGC (Carnegie Institution of Washington). The 5-6, 5-14 and 5-23 monoclonal antibodies used by Hill, F. Epstein were obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained by The University of Iowa, Department of Biology, Iowa City, IA 52242, USA.

Funding
G.B.P. is funded by the MRC [Career Development Award G0600127]. F.G.S. is supported by a BBSRC studentship, M.R.K.F., F.X. and J.K. are supported by EU FP7 project PASTA [STREP 202608]. G.B.P. and K.F.P. are also supported by the Wellcome Trust [097920/Z/11/Z]. The CGC is funded by NIH Office of Research Infrastructure Programs [P40 OD010440].

Competing Interests
The authors have no competing interests to declare.

References


Chapter 8

References


Nair, S.S., Nair, B.C., Cortez, V., Chakravarty, D., Metzger, E., Schule, R., Brann, D.W., Tekmal, R.R., Vadlamudi, R.K., 2010. PELP1 is a reader of histone H3 methylation that facilitates oestrogen receptor-alpha target gene activation by regulating lysine demethylase 1 specificity. EMBO reports 11, 438-444.


trimethylation by COMPASS and the regulation of gene expression. Molecular cell 19, 849-856.


