Imperial College London

A novel role for condensin in *Drosophila* neurons

Amira Hassan

Imperial College London

Department of Life Sciences

A thesis submitted for the degree of Doctor of Philosophy

April 2020

Declaration of originality

I declare that this thesis and the work presented in it are wholly my own. Any work of others used in this thesis has been appropriately referenced and acknowledged.

Amira Hassan

Copyright declaration

‘”The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.’”

Abstract

Condensins are multi-protein complexes that have key functions such as chromosome compaction, regulation of gene expression and 3D chromatin conformations. Despite the wealth of knowledge on condensins in dividing cells, their role in post-mitotic, specialised cells remains largely unknown. In this thesis I explore a novel role for condensin I subunit Cap-G in *Drosophila* neurons. *In vivo* evidence suggests that condensin I subunits Cap-G and Barren are expressed and localised in neurons. The generation of an exclusively post-mitotic, neuron-specific knockdown of Cap-G revealed an exciting role for condensin in maintaining neuronal identity and health. Removal of Cap-G from post-mitotic neurons resulted in severe developmental arrest, behavioural defects and an overall disruption of gene expression. Furthermore, knockdown animals exhibited a downregulation of neuron-specific genes and an upregulation of ectopic, non-neuron specific genes, suggesting loss of neuronal identity. Cap-G DNA binding profiles in neuronal stem cells and neurons reveal dynamic, cell-specific binding of Cap-G across cell types. Moreover, genes bound by Cap-G significantly overlap with the disrupted genes detected in knockdown animals. Overall, the evidence suggests that Cap-G is necessary for healthy neuronal development and maturation and has a neuron-specific, gene regulatory role in *Drosophila*. The work in this thesis provides the first evidence that condensin is necessary in post-mitotic neurons for transcriptome stability and maintenance of neuronal identity.

Acknowledgments

I want to express my deepest gratitude to Dr Tony D Southall for being a fantastic supervisor and mentor. I consider myself a very lucky student for having a supervisor who is present, supportive and always encouraging. Throughout my PhD your advice and guidance has been essential, but most importantly you always had faith in me and my abilities, even when I didn't have it in myself. Your passion and dedication to research is contagious and has motivated me to approach science with novelty and curiosity.

A very special thanks goes to Dr Gabriel N Aughey, because I don't think I could have made it this far without him. Gabriel dedicated so much time to me and my project, mentoring me and teaching me not only the technicalities of being a scientist, but also how to think like one. Gabriel has this innovative thinking and hunger for knowledge which is inspiring. He helped me gain confidence and always pushed me to my limits, setting high expectations because he knew I could meet them. Thank you for being there and for all the constructive and thought-provoking discussions we had.

A special thanks goes to Emma, for being the best work buddy I could have hoped for. You were always there by my side, whether it was to help with experiments or offering an ear to listen, thank you my dear friend. A number of students came to the lab, but none were quite as great as Pablo. His curiosity and dedication revived the passion in my project even in the toughest times. Thank you for all your help. Finally, I thank Alicia and Colin for sharing their knowledge and supporting me throughout my project. During my time in the Southall lab I have grown as a scientist and I have been challenged and praised by my colleagues, who have become dear friends. They helped me through the highs and lows and made my PhD experience the best it could possibly be.

The PhD came with its fair share of frustrations and struggles, and I wouldn't have made it without the unwavering support of my family and friends. A special thanks goes to my wonderful parents, who always believed in me, I hope I have made you proud. To my dearest Nehal, thank you for your patience and for always being there by my side through thick and thin. You have made these past years of PhD some of the best of my life.

Finally, I would like to thank the whole *Drosophila* community, especially Prof Heidmann for sharing reagents and fly-lines. These were instrumental in carrying out my PhD experiments. Ultimately, I thank the BSSRC for providing the funding to pursue this project and the Imperial College London Department of Life Sciences for the opportunity to learn and gain excellent research experience.

Table of Contents

Abstract 4
Acknowledgments 5
List of Figures 11
List of Tables 13
Abbreviations 14
Chapter 1 Introduction 16
1.1 Neurodevelopment 16
1.2 *Drosophila* neurogenesis 17
1.3 Neuronal differentiation 21
1.3.1 Asymmetric segregation of neuron-specific factors to the GMCs 21
1.3.2 Factors that maintain neuronal differentiation 22
1.4 Lola-N interacts with condensin complex subunit Cap-G 24
1.5 *Drosophila* as a model to study neurodevelopment 25
1.6 SMC protein complexes 28
1.7 The condensin complex: Structure 28
1.8 Canonical functions of the condensin complex 29
1.8.1 Condensin complex and the DNA loop extrusion mechanism 29
1.8.2 Chromosome assembly and segregation 33
1.9 Condensin and regulation of gene expression 34
1.9.1 Condensin and RNA Polymerases 34
1.9.2 Condensin at Promoter / Enhancer regions 36
1.9.3 Condensin DNA binding sites and TF interactions 37
1.9.4 Condensin and cell-specific gene expression 38
1.9.5 Condensin and gene silencing 38
1.9.6 Contradictory evidence for a condensin role in gene regulation 40
1.10 Condensin and 3D genome organisation 42
1.10.1 Chromatin Topology 42
1.10.2 Condensin and 3D chromatin organisation 44
1.11 Condensin and the nervous system 46

1.11.1 Chromatin remodelling and maintenance of neuronal fate 46
1.11.2 Condensin mutations impact neurodevelopment 47
1.11.3 The gap in our knowledge: a post-mitotic role of condensin 48
1.12 Summary of project aims 48
Chapter 2 Materials and Methods 50
2.1 *Drosophila* melanogaster: Husbandry and maintenance 50
2.2 Cell-specific degradation of GFP-tagged proteins using the deGradFP system 52
2.3 Generation of stochastic clones 52
2.4 Embryo microinjection 53
2.5 Behavioural and phenotypic assays 53
2.5.1 Developmental Survival Assay 53
2.5.2 Survival assay 54
2.5.3 Locomotion Assay 54
2.5.4 Negative Geotaxis (climbing) assay 54
2.6 Immunostaining and microscopy 55
2.6.1 Larval and adult brain tissue 55
2.6.2 Embryonic tissue 55
2.7 TUNEL Staining 57
2.8 *In-situ* hybridisation 58
2.9 Image acquisition and analysis 59
2.9.1 Average nuclear area quantification 59
2.9.2 Automatic quantification of cell numbers 59
2.9.3 Quantification of live and *in vivo* GFP signal intensity 60
2.9.4 Statistical significance and plots 61
2.10 DNA cloning 61
2.10.1 Genomic DNA extraction 61
2.10.2 Polymerase Chain Reaction (PCR) 61
2.10.3 Competent cells transformation 62
2.10.4 Gibson Assembly 62
2.10.5 Restriction digest 63
2.11 Targeted DamID 63
2.11.1 Cloning of *UAS-LT3-Cap-G-Dam* vector for transgenic fly generation 63

2.11.2 Targeted DamID library preparation 64
2.12 Analysis of Targeted DamID sequencing data 65
2.12.1 Data accessibility and file formats for statistical analysis 65
2.12.2 Peak calling 65
2.12.3 Peak annotation and De Novo Motif discovery 66
2.12.4 Accessible chromatin and CRM analysis 66
2.12.5 Correlations and PCA 67
2.12.6 Plotting binding profiles 67
2.12.7 Peak intersection analysis 68
2.12.8 Gene Ontology Analysis 68
2.13 RNA extraction 68
2.14 Analysis of RNA-seq experiments 69
2.14.1 Gene mapping and summarisation 69
2.14.2 Quality control 70
2.14.3 Differential expression analysis 70
2.14.4 Tissue-specificity profiling from FlyAtlas 2 71
Chapter 3 Cap-G is essential in post-mitotic neurons for healthy nervous system development 72
3.1 Introduction 72
3.1.1 Neuronal development 72
3.1.2 Condensin I cell localisation and expression 73
3.1.3 The deGradFP system to knockdown GFP-tagged proteins 75
3.1.4 Cell-specific Cap-G knockdown with the deGradFP system 78
3.1.5 Alternative methods to generate cell-specific Cap-G KD/mutant 79
3.1.6 Chapter Aims 82
3.2 Results 83
3.2.1 *Cap-G* is expressed in neurons 83
3.2.2 Subunits of the condensin I complex Cap-G and Barren localise in neurons 84
3.2.3 Cap-G co-localises with Lola-N in neuronal nuclei 91
3.2.4 Cell-specific knockdown of Cap-G with deGradFP 95
3.2.5 Cap-G KD in neurons leads to reduced survival and behavioural defects across development 97
3.2.6 Cap-G KD in neurons results in impaired mobility in flies 99

3.2.7 Increased levels of DNA damage in Cap-G KD neurons 100
3.2.8 Cap-G KD leads to increased neuronal death 104
3.2.9 Characterising the effects of *elav-GAL4* driven Cap-G KD 106
3.2.10 Cap-G KD driven by *elav-GAL4* has no impact on cell division 110
3.3 Discussion 112
3.3.1 Condensin I localisation and distribution in *Drosophila* neurons 112
3.3.2 A post-mitotic role for Cap-G in neurons 114
Chapter 4 Cap-G binds dynamically to chromatin in neuronal stem cells and neurons 119
4.1 Introduction 119
4.1.1 Condensin binding at multiple genomic regions in the DNA 119
4.1.2 3D genome organisation and chromatin states 122
4.1.3 The condensin complex and the chromatin landscape 124
4.1.4 Targeted DamID to profile cell-specific binding 127
4.1.5 Profiling cell-specific binding of Cap-G in NSCs and neurons 130
4.1.6 Chapter aims 132
4.2 Results 133
4.2.1 Profiling cell-specific Cap-G binding in the *Drosophila* CNS 133
4.2.2 Cap-G binds to unique genes across cell types 137
4.2.3 Gene Ontology of Cap-G bound genes reveal multiple tissue-specific terms. 140
4.2.4 Cap-G binding is depleted at accessible chromatin and enriched at repressive chromatin 144
4.2.5 Cap-G binding overlaps with Lola-N binding sites in mature neurons and is differentiation-dependent 150
4.3 Discussion 157
4.3.1 Cap-G binding profiles in NSCs and neurons 157
4.3.2 Cap-G binding in NSCs and neurons suggests a role in chromatin organisation 161
4.3.3 Cap-G and Lola-N interact to maintain neuronal fate 163
Chapter 5 Cap-G knockdown leads to severe disruption of neuronal transcriptomes 166
5.1 Introduction 166
5.1.1 Gene expression regulation by the condensin complex 166

5.1.2 Contradictory evidence for gene expression by condensin and gaps in our knowledge 169
5.1.3 Chapter aims 170
5.2 Results 172
5.2.1 Knockdown of Cap-G in neurons leads to disruption of gene expression in the CNS 172
5.2.2 *Elav-KD* and *nSyb-KD* display similar gene expression differences but at different severities 177
5.2.3 Depletion of Cap-G in neurons leads to downregulation of CNS genes 182
5.2.4 Ectopic upregulation of non-CNS genes in *Cap-G KD* animals 185
5.2.5 Cap-G binding significantly overlaps with differentially expressed genes 188
5.3 Discussion 191
5.3.1 Cap-G regulates gene expression 191
5.3.2 Cap-G mediates neuron-specific gene regulation 192
Chapter 6 Discussion 196
6.1 Overview and summary of key findings 196
6.2 Cap-G in newly born and mature neurons 197
6.3 Cap-G and Lola-N interaction and its impact on neuronal maturation 200
6.4 Cap-G as a regulator of gene expression in neurons 201
6.5 A role for the condensin I complex or just for Cap-G? 204
6.6 Implications for Health 207
6.7 Concluding remarks 208
Chapter 7 Bibliography 209
Chapter 8 Appendix 221

List of Figures

Chapter 1

Figure 1.1. *Drosophila* CNS and neuron differentiation. 19
Figure 1.2. *Drosophila* life cycle. 26
Figure 1.3. Condensin structure and loop extrusion mechanism 32
Figure 1.4. Summary of known condensin functions. 41

Chapter 3

Figure 3.1. *Drosophila* embryo and larval brain anatomy. 74
Figure 3.2 Schematic of the DegradFP system. 77
Figure 3.3 Generating neuron specific Cap-G KD using the deGradFP system. 81
Figure 3.4 *Cap-G* expression in neurons. 85
Figure 3.5 Cap-G^{EGFP} localises in neuronal nuclei in embryos. 87
Figure 3.6 Cap-G^{EGFP} localises in neuronal nuclei in larval brains. 88
Figure 3.7 Barren^{EGFP} localises in neurons 90
Figure 3.8 Cap-G^{EGFP} colocalises with Lola-N in neurons. 92
Figure 3.9 Cap-G^{EGFP} partially colocalises with HP1 in neurons. 94
Figure 3.10. Cap-G^{EGFP} knockdown in neurons. 96
Figure 3.11. Cap-G knockdown in neurons results in premature developmental arrest and severe mobility defects. 98
Figure 3.12 Cap-G KD leads to increased DNA damage in larval neurons. 101
Figure 3.13 Cap-G knockdown has no effect on neuronal nuclei area. 103
Figure 3.14 Increased neuronal apoptosis in *elav-KD*. 105
Figure 3.15 Cap-G^{EGFP} levels remain constant in NSCs between *elav-KD* and *control*. 107
Figure 3.16 Live Cap-G^{EGFP} levels remain unchanged between *elav-KD* and controls. 109
Figure 3.17. *elav-KD* has no effect on mitotic cell numbers. 111

Chapter 4

Figure 4.1 DamID experimental pipeline. 126
Figure 4.2 Targeted DamID for cell-specific profiling of Cap-G binding. 129
Figure 4.3 Cell specific profiling of Cap-G using TaDa. 131
Figure 4.4 Cap-G binding is dynamic across cell types. 134
Figure 4.5. Cap-G binds to NSC and cell-cycle genes in neurons. 136
Figure 4.6. Cap-G binds to unique genes and is enriched in gene bodies. 138
Figure 4.7. Gene ontology enrichment analysis for Cap-G bound genes. 141
Figure 4.8 Enriched Motif analysis for Cap-G binding. 143
Figure 4.9. Cap-G binding is depleted at enhancers and open chromatin. 146
Figure 4.10 Cap-G binding analysis against accessible chromatin. 147

Figure 4.11. Cap-G binding is enriched at repressive chromatin states in NSCs. . 149
Figure 4.12. Cap-G and Lola binding positively correlates in mature neurons...... 152
Figure 4.13. Cap-G and Lola binding significantly overlaps in neurons...... 154
Figure 4.14. Cap-G binding at Lola peaks is differentiation-dependent. 156

Chapter 5

Figure 5.1 Positive correlation between Cap-G KDs replicates and experimental conditions...... 173
Figure 5.2. RNAseq example profiles of chr 3L show similarity between replicates. 174
Figure 5.3. PCA between control and experimental samples. 176
Figure 5.4. Differential gene expression analysis shows misregulation of genes. 178
Figure 5.5 MA plot of RNA-seq datasets...... 180
Figure 5.6. Overlapping DEGs between *elav-KD* and *nSyb-KD*...... 181
Figure 5.7. Example of DEG tracks for both *elav-KD* and *nSyb-KD*...... 184
Figure 5.8 GO analysis of downregulated genes for *elav-KD* and *nSyb-KD.* 186
Figure 5.9 GO analysis for upregulated genes for *elav-KD* and *nSyb-KD*...... 187
Figure 5.10. Upregulated genes are ectopically expressed in *Cap-G KDs* CNS...... 189
Figure 5.11. Differentialy expressed genes are bound by Cap-G in neurons...... 190

Chapter 6

Figure 6.1 Summary of key findings...... 198
Figure 6.2. Hypotheses for Cap-G functions in neurons...... 205

Appendix

Figure 8.1. UAS-LT3-Cap-G-Dam construct generated..... 222
Figure 8.2 UAS-Cap-G-CRISPR/Cas9 mutant plasmid designed..... 224
Figure 8.3 Lola-N binding correlation and QC between elav and nSyb replicates226

List of Tables

Table 1.1. Condensin complex I and II subunit names 31
Table 2.1 List of *Drosophila* strains used in this study 50
Table 2.2 List of antibodies used in this study. 56
Table 2.3 Detector scales used per experiment. 60
Table 2.4 Enzymes used in this study. 63
Table 3.1 Cell-specific Cap-G knockdown / knockout methods attempted. 80
Table 4.1 Summary of condensin genomic binding 120
Table 5.1 Condensin roles in gene expression regulation 168

Abbreviations

AC	Accessible Chromatin
CATaDa	Chromatin Accessibility Targeted DamID
CB	Central Brain
ChIP-seq	Chromatin Immunoprecipitation Sequencing
CRM	Cis-regulatory modules
CT	Chromosome Territory
DAM	DNA adenine methyltransferase
DamID	DNA adenine methyltransferase Identification
DCC	Dosage compensation complex
DEG	Differentially expressed genes
EMSA	Electrophoretic mobility shift assay
ER-α	Estrogen receptor α
ESC	Embryonic stem cell
FC	Fold change
FDR	False Discovery Rate
GMC	Ganglion Mother cell
GO	Gene Ontology
Hi-C	Chromosome conformation capture
HiChIP	Protein-centric chromatin conformation
Hrs	Hours
INP	Intermediate neural progenitor
KD	Knockdown
Mins	Minutes
NSC	Neuronal Stem cell
OL	Optic Lobe
OPC	Optic proliferation centre
PCA	Principal Component Analysis
PEV	Positional effect variegation
POI	Protein of interest
RED	Regulatory Element Database
RNAP	RNA Polymerases
ROI	Region of interest
SMC	Structural maintenance of chromosomes
TAD	Topologically Associated Domains
TaDa	Targeted DamID
TES	Transcription end sites
TF	Transcription Factor
TFIIIC	RNA Polymerase III
TSS	Transcriptional start sites
TTS	Transcriptional terminal sites
VNC	Ventral Nerve Cord

Preface

Part of the data presented in Chapters 3, 4 and 5 has been published in the journal eLife under the title "Condensin I subunit Cap-G is essential for proper gene expression during the maturation of post-mitotic neurons" (1). Authors listed on the publication are Amira Hassan, Pablo Araguas Rodriguez, Stefan K. Heidmann, Emma L. Walmsley, Gabriel N. Aughey and Tony D. Southall.

AH conceived, designed and performed the experiments and data analysis. PAR helped generating replicates for behavioural assays and optimising the TUNEL staining protocol. ELW helped with tissue dissections and RNA extraction. SKH provided fly lines and reagents. TDS and GNA helped conceive and design experiments. AH, GNA and TDS wrote the manuscript.

Data availability

All raw sequence files and processed files presented in this thesis have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (accession number GSE142112). The peak calling results and gene lists used for analysis are available in the repository for this thesis https://doi.org/10.14469/hpc/7267.

Chapter 1 Introduction

1.1 Neurodevelopment

The nervous system is one of the most complex organs, where billions of neurons are shaped and organised into networks vital to our body functions and behaviours. Understanding how a fully functioning nervous system is assembled from just an initial handful of cells has fascinated scientists for generations. During neurodevelopment the proliferation and specialisation of progenitor cells will give rise to neurons, ultimately populating the whole nervous system. In neurogenesis, the progenitor cells begin to differentiate into neurons, where changes occur at the genomic and molecular level. These changes as a cell differentiates are known as epigenetic changes, which will define their neuronal fate. The epigenetic landscape of a cell at the early stages of differentiation is highly plastic and it can be manipulated and altered, although this becomes increasingly difficult as a cell matures (2,3). Therefore, factors and genomic changes that dynamically form as a cell matures are necessary to maintain its terminally differentiated state (4).

Before being fully functioning, the newly-born neurons need to migrate to their precise location and begin a period of maturation. As neurons mature, morphological changes occur where they generate axons and dendrites to then form precise connections with other neurons via their synapses (5). However, dissecting the morphological changes and network assembly offers only part of the picture. To fully understand the complexity of the nervous system we need to identify the molecular mechanisms behind the formation of each individual neuron. Currently there is a considerable gap in our knowledge of the epigenetics of neuronal maturation, only addressed by a handful of studies (1,6-8).

The genomic processes underlying neuronal differentiation have been the focus of recent neurodevelopment research. Regulation of gene expression, changes in chromatin accessibility and the 3D genome organisation all occur as neurons differentiate from progenitor cells (9-14). However, chromatin remodelling and changes to the genomic organisation in neurons are only just starting to be investigated. Nuclear architecture and chromatin-associated proteins are key drivers of chromatin landscape changes, so investigating these factors and mechanisms in neurons could further our understanding of neuronal maturation.

1.2 *Drosophila* neurogenesis

The nervous system is one of the most complex organs, and research has focused on understanding how stem cells generate the vast and varied populations of neurons. *D. melanogaster* has a relatively simple nervous system, with many similarities to the vertebrate nervous system, making it an excellent model to study the mechanisms underlying neuronal development. Neurogenesis describes the process by which neuronal cells are formed from progenitor stem cells. At the heart of neurogenesis lie the neuronal stem cells and in *D. melanogaster* they are known as Neuroblasts. For clarity, I will refer to them in this thesis as Neuronal Stem Cells (NSCs). The *Drosophila* nervous system anatomy changes from embryonic to larval and ultimately adult brain. In embryos, the neuroectoderm is found at the ventrolateral region and it forms the brain and the ventral nerve cord (VNC), which is broadly segmented in thorax and abdominal regions (15). The larval brain is broadly segmented in the central brain (CB), optic lobes (OL) and the VNC (Figure 1.1A).

In *Drosophila* there are two distinct waves of neurogenesis occurring at the embryonic and larval stages respectively. At the embryonic stage, all the neurons that will populate

the larval central nervous system are formed, however, these amount to only 10% of the adult neurons (16,17) (Figure 1.1A). At the end of embryonic development, majority of abdominal NSCs in the VNC undergo controlled cell death (apoptosis), except for NSCs located in the brain and thoracic regions, which enter a period of quiescence until 8-10 hours after larval hatching (15). The second wave of neurogenesis is then initiated in the larval stage, where quiescent NSCs are re-activated and begin proliferating, giving rise to 90% of the adult neurons (18).

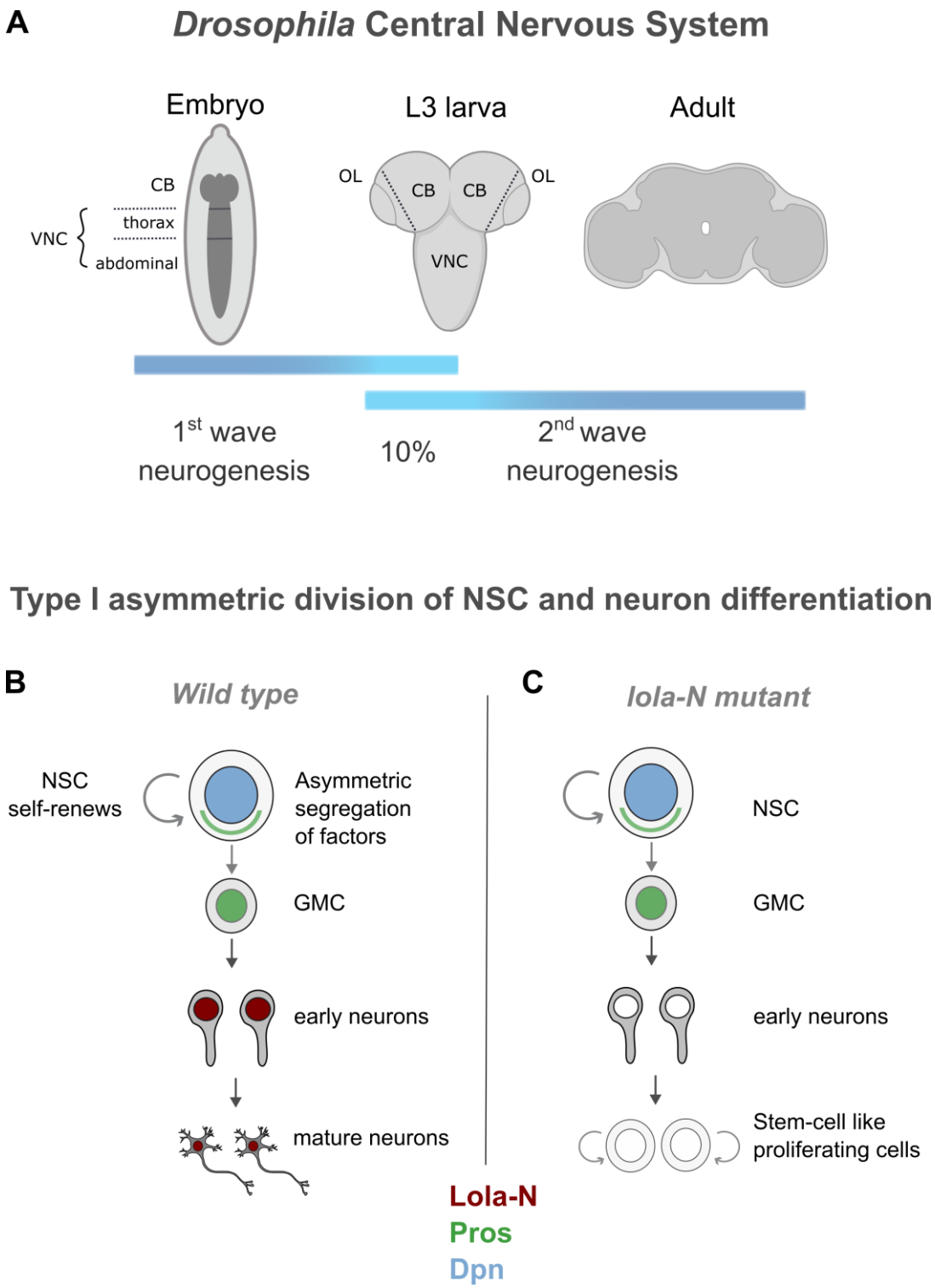


Figure 1.1. *Drosophila* CNS and neuron differentiation.

A) CNS at three main developmental stages. 1st wave of neurogenesis generates all embryonic and larval neurons, and 10% of adult neurons. 2nd wave generates 90% of adult neurons. (Created with BioRender.com) **B)** Asymmetric division of Type I NSCs in wild type *Drosophila*. Prospero factor asymmetrically segregates to the GMC, inducing differentiation. GMC divides once to produce newly differentiated neurons, expressing *lola-N*. Neurons further develop into fully mature neurons. **C)** In *lola-N* mutants, early neurons dedifferentiate into NSC-like, proliferating cells (19).

The first wave of neurogenesis begins with delamination of NSCs from the neuroepithelium. Acquisition of NSC identity and its delamination is orchestrated by lateral inhibition of Notch signalling, which leads to the expression of stem-cell specific genes (20). The specification of NSCs is also dictated by the expression of Sox and Snail family genes (e.g *SoxN, worniu)* (21-26). Finally, the embryonic NSCs can be identified by positional identity and temporal expression of NSC-specific sequential markers, which have been thoroughly studied and reviewed (15,18).

The spatial and temporal distribution of NSCs form distinct stem cell populations, which exhibit different proliferation patterns and behaviours. Three main types of stem cell divisions have been described in the *Drosophila* CNS, namely the Type 0, I and II (27). The Type I NSCs are located in embryonic thorax and abdomen sections of the VNC as well as the larval central brain. These NSCs undergo asymmetric cell division, giving rise to another NSC and a Ganglion Mother Cell (GMC). The GMC then divides once more to produce fully differentiated neurons (Figure 1.1B) (28). In the larval central brain the majority of NSCs adopt this type of cell division however, a small number of Type II dividing NSCs have been identified (18). During the Type II division the NSC will asymmetrically divide to produce an Intermediate Neural Progenitor (INP) which matures and divides asymmetrically a set number of times generating more INPs and ultimately a GMC (29,30). In addition, in the larval optic lobes there is the optic proliferation centre (OPC), which contains symmetrically dividing NSCs (15). Finally, in the embryonic VNC the Type 0 NSCs have been identified, which will asymmetrically divide to give rise to another NSC and a terminally differentiated neuron, without the intermediate GMC (31-33). The diverse divisions of NSCs are a tightly regulated process where molecular factors precisely manipulate the self-renewal of NSCs and the

differentiation of daughter cells (34). In the following paragraphs I will review our current understanding of factors driving neuronal differentiation in *Drosophila.*

1.3 Neuronal differentiation

Neuronal differentiation is primarily driven by the timely action of gene regulatory factors that turn off stem cell/progenitor genes and promote the expression of neuronal genes. In *Drosophila*, differentiation is initiated and maintained by a variety of mechanisms, including segregation of molecular factors, transcriptional regulation and chromatin compaction. Once the neuron is fully differentiated, molecular and transcription factors are in place to maintain its fate. The reversion of terminally differentiated cells to a stem-cell like state is a process known as dedifferentiation.

1.3.1 Asymmetric segregation of neuron-specific factors to the GMCs

Asymmetric segregation of cell-fate determinants to the GMCs is a key mechanism that promotes differentiation. For example, during asymmetric division of Type I NSCs the factors Prospero, Brain tumour (Brat) and Numb are localised to the basal cortex of a NSC and then segregated to the GMC (35-38) (Figure 1.1B). These three factors have different roles in determining cell fate. For instance, Numb inhibits the proliferative state of a cell by repressing Notch signalling receptors (39). Brat is a translation regulator that inhibits the growth and self-renewal properties in GMCs, inducing differentiation (36,37). Prospero is a transcription factor (TF) that after segregation enters the GMC nucleus to regulate target genes. Prospero silences cell-division genes such as *cyclin A* and *E* and directly represses NSC-specific genes such as *deadpan* (*dpn*) (40,41). Moreover, Prospero is able to bind to differentiation-inducing genes, but it is not sufficient to activate expression of said genes, and likely requires interacting partners to induce differentiation

(42). In *prospero* mutants GMCs fail to differentiate and continue to abnormally proliferate leading to the formation of tumours (40). Whilst in *brat* mutants, the INPs of Type II NSCs are affected, as they fail to differentiate and proliferate uncontrollably, also forming tumours (37,43). These events clearly identify a need for factors that timely terminate the mitotic ability of cells and induce differentiation. However, more evidence is arising that regulation is also essential in post-mitotic, differentiated cells, where specific factors are in place to maintain cell fate.

1.3.2 Factors that maintain neuronal differentiation

A nascent research field in *Drosophila* aims to identify factors involved in the maintenance of neuronal differentiation. These are factors in place that maintain the neuronal state of a cell and prevent its reversion back to a proliferative cell, namely a dedifferentiation phenotype. A handful of transcription factors have been described that prevent dedifferentiation, however how they interplay together to regulate this phenomenon remains poorly understood. The differentiated state of embryonic and optic lobe neurons is maintained by an isoform of the zinc-finger transcription factor longitudinals-lacking (Lola), Lola-N (19). The newly-born, optic lobe neurons in *lola-N* mutants dedifferentiate to NSC-like cells and express NSC-specific genes such as *deadpan*. These NSC-like cells then begin to divide in an uncontrolled fashion, eventually forming tumours (Figure 1.1C) (19). This phenotype is attributed to the derepression of silenced stem-cell and cell-cycle genes in neurons. Lola-N is a transcriptional repressor, as it binds NSC-specific (e.g *dpn*) and cell cycle genes (eg. *cyclin E*) which are then re-activated upon depletion of Lola-N (19). Interestingly, no dedifferentiation phenotype is observed in *lola-N* knockdown in adult neurons, which highlights the importance of Lola-N at early stages

of neuronal differentiation, but also that other factors could be in place to maintain the differentiation of neurons at later developmental stages.

Another TF, nervous fingers 1 (Nerfin-1) is involved in maintenance of neuronal differentiation in the larval central brain, medulla neurons of the optic lobes and ventral nerve cord (44,45). *nerfin-1* loss-of-function mutants generate tumours leading to dedifferentiation of neurons into over-proliferating stem-cell like cells (44). Nerfin-1 inhibits the Notch-signalling pathway, hence repressing NSCs self- renewal ability (45,46). Furthermore, proteins that are not TFs have also been implicated in maintenance of neuronal fate. An RNA-splicing protein, midlife crisis (Mdlc), has been identified to maintain neuronal fate in the larval CNS (47). *mdlc* mutants repress Prospero and ectopically express NSC marker *deadpan* but are unable to proliferate (47).

These recent findings have uncovered regulators essential for the maintenance of neuronal differentiation. However, little is known about the interplay of these factors and how they link together in a regulatory network. A recent study highlights the protein Slimb as a post-translational regulator of Nerfin-1 by targeting it for degradation (48). This degradation is key in regulating the abundance of Nerfin-1 in the tissue, as overexpression of *nerfin-1* results in premature differentiation of NSCs (44). Moreover the TF Scalloped (Sd) has been described as an interactor of Nerfin-1, as both proteins form a complex and are involved in maintenance of the differentiated state of medulla neurons (49). Upon depletion of Sd the neurons dedifferentiate and ectopically express stem-cell and mitotic markers (49). Finally both TFs have been found to bind to and regulate the expression of genes necessary to maintain neuronal fate, such as NSC-specific genes and Notch-signalling pathway genes (49). Overall, this active field of research highlights the importance of identifying molecular factors and their interactors,

to fully understand the networks and mechanisms involved in maintenance of neuronal fate.

1.4 Lola-N interacts with condensin complex subunit Cap-G

The transcription factor Lola has well-described roles in neurodevelopment. It modulates axon guidance and maintains neuronal differentiation in embryonic and larval optic lobe neurons (19,50). The *lola* gene locus encodes for 20 protein isoforms that differ at the C-terminal end, which bears one of 17 different Zn-finger domains or none at all (50,51). A recent loss-of-function screen, aimed at creating mutants for each of these isoforms, discovered that isoform Lola-F regulates axonal guidance, whilst Lola-O regulates neurotransmitter expression (52). These recent findings highlight the varied role of Lola in neurodevelopment and emphasise the importance of continued research around the isoforms of this factor. Lola-N is a tumour suppressor and has the ability to silence genes, however, the mechanism by which gene silencing is achieved remains largely unknown (19). Therefore, identifying interacting partners could offer potential insights on the mechanisms underlying neuronal differentiation and gene silencing.

A yeast-two-hybrid (Y2H) screen was performed in the Southall lab to identify proteins that interact with Lola-N (unpublished, Southall, T.D). The screen resulted in a variety of hits and amongst them the protein Cap-G was identified. This finding was somewhat surprising, as Cap-G is a subunit of the chromatin-associated condensin I complex which has an important role in chromosome condensation during cell-division (53). Therefore, finding Cap-G to interact with an exclusively post-mitotic, neuron-specific factor like Lola-N was of great interest and resulted in the foundations for my project. Extensive research efforts have been made to aid our understanding of how chromatin remodelling and chromatin-associated proteins could play a role in neuronal differentiation and

maintenance of neuronal fate (1,6-8). However, I explore these later in this review (Section 1.11.1), as prior to that I wish will elaborate on our current knowledge of condensin; chromatin topology and ultimately tie together the fields of chromatin remodelling and the epigenetics of neuronal maturation.

1.5 *Drosophila* as a model to study neurodevelopment

In this study I employ *Drosophila melanogaster* as the animal model as its simple nervous system, short generation time and genetic tractability make it an excellent model for studying neurodevelopment. Having explored our knowledge of neuronal differentiation in *Drosophila,* in the following section I wish to further justify the reasoning behind selecting this model to study neuronal differentiation and maturation.

The *Drosophila* life cycle takes approximately ten days at 25 ˚C, which includes four main developmental stages: embryo, larvae, pupae and adult fly (Figure 1.2) (54). After egg fertilisation, the embryo will develop into a larva in just 24 hours. Over the next four days the larvae undergo three moulting phases known as L1, L2 and L3. The final L3 larvae are known as the "wandering" larvae, and during pre-pupation they display distinct behaviours such as crawling along the edges of the vials. The pupal stage lasts 5 days, during which the pupae undergo metamorphosis to generate the adult flies, which have an average lifespan of 80 days. *Drosophila* is an extremely useful model in developmental biology as flies have a relatively short generation time and lifespan. Moreover, cultures are easy to maintain making experiment timelines more manageable than vertebrate models. The embryonic, L3 larval and adult stages are the most used in this study to address key neuronal development questions using a variety of techniques such as immunostaining, behavioural and survival experiments.

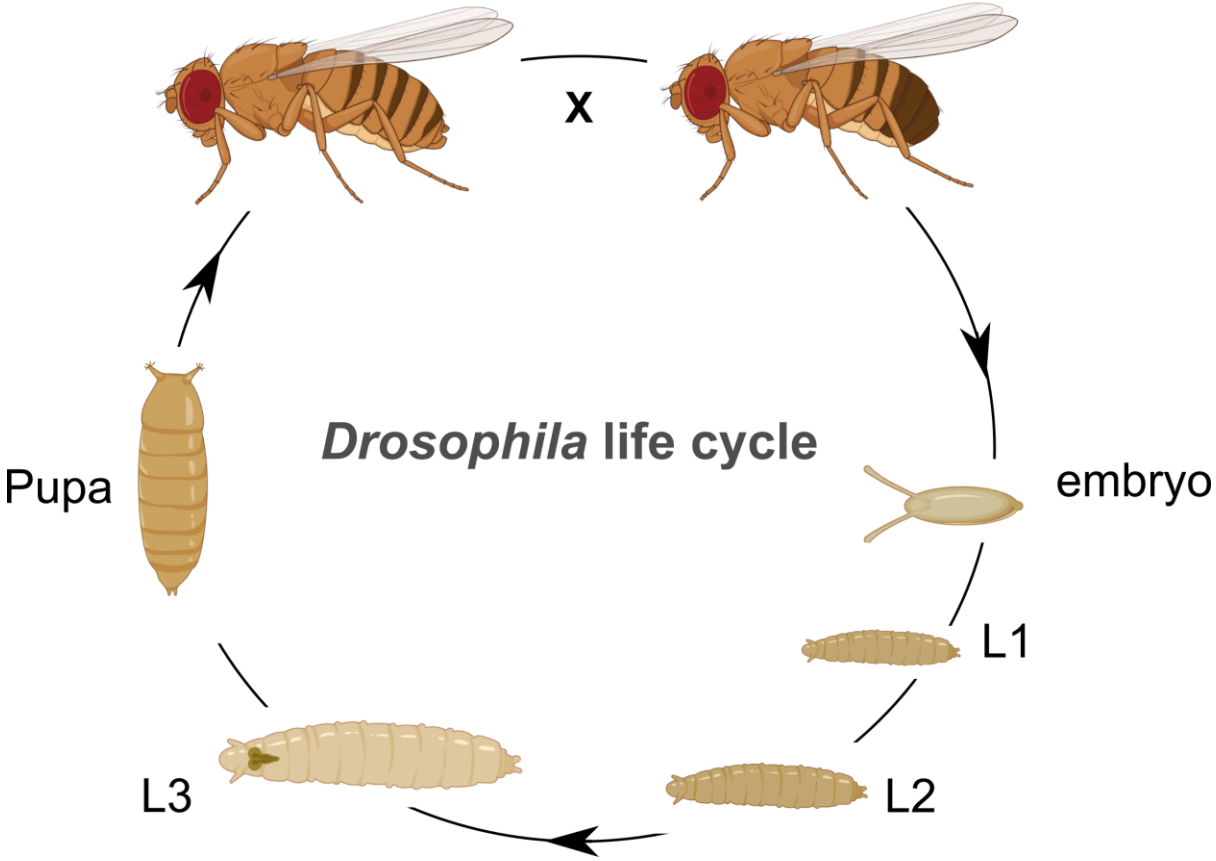


Figure 1.2. *Drosophila* life cycle.

There are five main developmental stages, embryo, three larval moulting stages and pupae. The full life cycle takes approximately 10 days total at 25^{0}C. (Created with BioRender.com)

Another aspect of *Drosophila* that makes it an excellent animal model is the wealth of genetic tools available to manipulate specific tissues and cells. The most widely used tool to induce cell-specific manipulation in *Drosophila* is the *GAL4/UAS* system (55). The scientific community has built a vast array of lines able to genetically modify almost any tissue at the cellular and molecular resolution. Another advantage of *Drosophila* is that its entire genome has been sequenced and annotated in great detail (56,57). These resources have been instrumental in accelerating our progress and understanding of core biological processes in development, as well as improving our understanding of human development and disease. Around 75% of human disease-related genes are conserved and detected in fruit-flies emphasising its importance as a model for medical research (58).

1.6 SMC protein complexes

Cells undergo strictly regulated cell divisions, where high fidelity DNA replication give rise to genetically identical daughter cells. Duplicated DNA is organised into structured, compact chromosomes, ready to be transferred to the daughter cells. This packaging of DNA is driven by highly conserved, elaborate machinery in the form of multi-protein complexes such as cohesin and condensin. Both of these complexes promote DNA looping, but each has distinct roles during mitosis. Cohesins facilitate sister chromatid tethering (cohesion) during interphase and mitosis, and is involved in genome-wide 3D organisation of chromatin (59). On the other hand, the condensin complex is necessary to drive chromatin compaction, which packages the DNA into chromosomes and facilitates segregation during mitosis (53,60,61). Both complexes consists of two structural maintenance of chromosomes (SMC) ATPase subunits, a kleisin subunit, and HEAT-subunit (a structural motif consisting of an array of helices) (60) (Figure 1.3). Each subunit plays an essential role in the ability of the complex to modify and shape DNA, ensuring successful chromosomal cohesion and condensation. In this review I will focus mainly on condensin complexes structure and function, further information on cohesins has been covered in multiple reviews (61-63).

1.7 The condensin complex: Structure

In all eukaryotes there are two types of condensin complexes, named condensin I and II, except for yeast, as it only has condensin I (53) (Table 1.1). Moreover, in *C. elegans* the dosage compensation complex (DCC) is classified as a third condensin complex, closely related to condensin I (64). The condensin I and II complexes share a core SMC-dimer but vary in their non-SMC subunits (60). The subunits are conserved across species and fall under different nomenclature (Table 1.1). Here, I will utilise the *Drosophila* names,

unless otherwise specified. Condensin core structure consists of SMC2 and SMC4 subunits, two coiled 'arms', each with an ATP-binding domain 'head' on one end. A heterodimer 'hinge' forms at the opposite end, adapting a 'V' shape. The non-SMC subunits of condensin I and II consists of a kleisin subunit (Barren / CAP-H2) and two HEAT-subunits (CAP-D2, CAP-G or CAP-D3, CAP-G2) (60) (Figure 1.3A) (Table 1.1). In *Drosophila*, no orthologue for Cap-G2 (condensin II) has been identified, which led to the speculation that Cap-G is present in both complexes (65). However, its exclusive association with condensin I has been described (66). Immunoprecipitation experiments with condensin II proteins failed to show interactions between Cap-G and subunits Cap-H2 and Cap-D3, arguing that Cap-G is not associated with, and does not perform, condensin II functions in *Drosophila* (66). Both complexes and their respective subunits have essential canonical and non-canonical roles in living cells. The mechanism by which mitotic chromosomes are condensed and organised is the canonical role of condensin, thoroughly described in the literature (53,60,62).

1.8 Canonical functions of the condensin complex

1.8.1 Condensin complex and the DNA loop extrusion mechanism

There have been several hypotheses attempting to describe how the condensin complex manipulates DNA from long strands into the packaged, condensed form necessary for cell division. Initiation of chromosome condensation requires the direct contact of condensin with DNA strands and the subsequent formation of DNA loops. The 'hinge' region of the SMC-dimer can bind to single stranded DNA (ssDNA) but not double stranded (dsDNA) (67). Interestingly, in *S. cerevisiae* the HEAT- subunits mediate binding to the dsDNA, as complexes lacking the Cap-G (Ycg1) subunit fail to associate to DNA (68). Therefore, the association of condensin to DNA via its HEAT-subunits is essential to initiate

chromosome condensation. Many describe a DNA loop extrusion (69-71) or chiral loop (53,60) to initiate the formation of chromosomal loop domains. Recent findings in *S. cerevisiae* describe a novel working model for the DNA loop extrusion by condensin (Figure 1.3B) (72-74).

Condensin associates to DNA via the Barren-Cap-G (Brn1-Ycg1) sub-complex, which contains a DNA interaction site (73). Subsequently, the kleisin subunit Barren forms a peptide loop around the DNA helix to anchor the condensin complex and avoid dissociation, known as 'safety-belt mechanism' (73). The binding of the non-SMC-subcomplex stimulates activation of the ATPase domains of the SMC-dimer, which initiates ATP hydrolysis (68,73). ATP hydrolysis is essential in activating the DNA translocase activity of the condensin complex, acting as a mechanochemical motor able to reel DNA (74). Real time visualisation of condensin loop extrusion activity describes the DNA reeling as asymmetric (72). In summary, these findings describe the DNA loop extrusion by means of condensin binding to DNA, anchoring and initiating ATP-hydrolysis-dependent, asymmetrical DNA translocation (72-74) (Figure 1.3B). This model is in contrast with the original theoretical models that described the DNA loop extrusion as symmetrical (69,70). Moreover, research in *Xenopous eggs* suggests that DNA looping by condensin is cell-cycle dependent, suggesting that cells in a dividing state adopt DNA looping as a form of chromatin organisation (75).

Table 1.1. Condensin complex I and II subunit names

Condensin I					
Classification	**ATPase**		**Kleisin**	**HEAT**	
Subunits	SMC2	SMC4	Cap-H	Cap-G	Cap-D2
D. melanogaster	SMC2	SMC4	Barren	Cap-G	Cap-D2
C. elegans	MIX-1	SMC4	DPY-26	Cap-G1	DPY-28
S. cerevisiae	SMC2	SMC4	Brn1	Ycg1	Ycs4
S. pombe	Cut14	Cut3	Cnd2	Cnd3	Cnd1
Vertebrates	CAP-E	CAP-C	Cap-H	Cap-G	Cap-D2
Condensin II					
Classification	**ATPase**		**Kleisin**	**HEAT**	
Subunits	SMC2	SMC4	Cap-H2	Cap-G2	Cap-D3
D. melanogaster	SMC2	SMC4	Cap-H2	-	Cap-D3
C. elegans	MIX-1	SMC4	KLE2	Cap-G2	HCP-6
S. cerevisiae	-	-	-	-	-
S. pombe	-	-	-	-	-
Vertebrates	CAP-E	CAP-C	Cap-H2	Cap-G2	Cap-D3

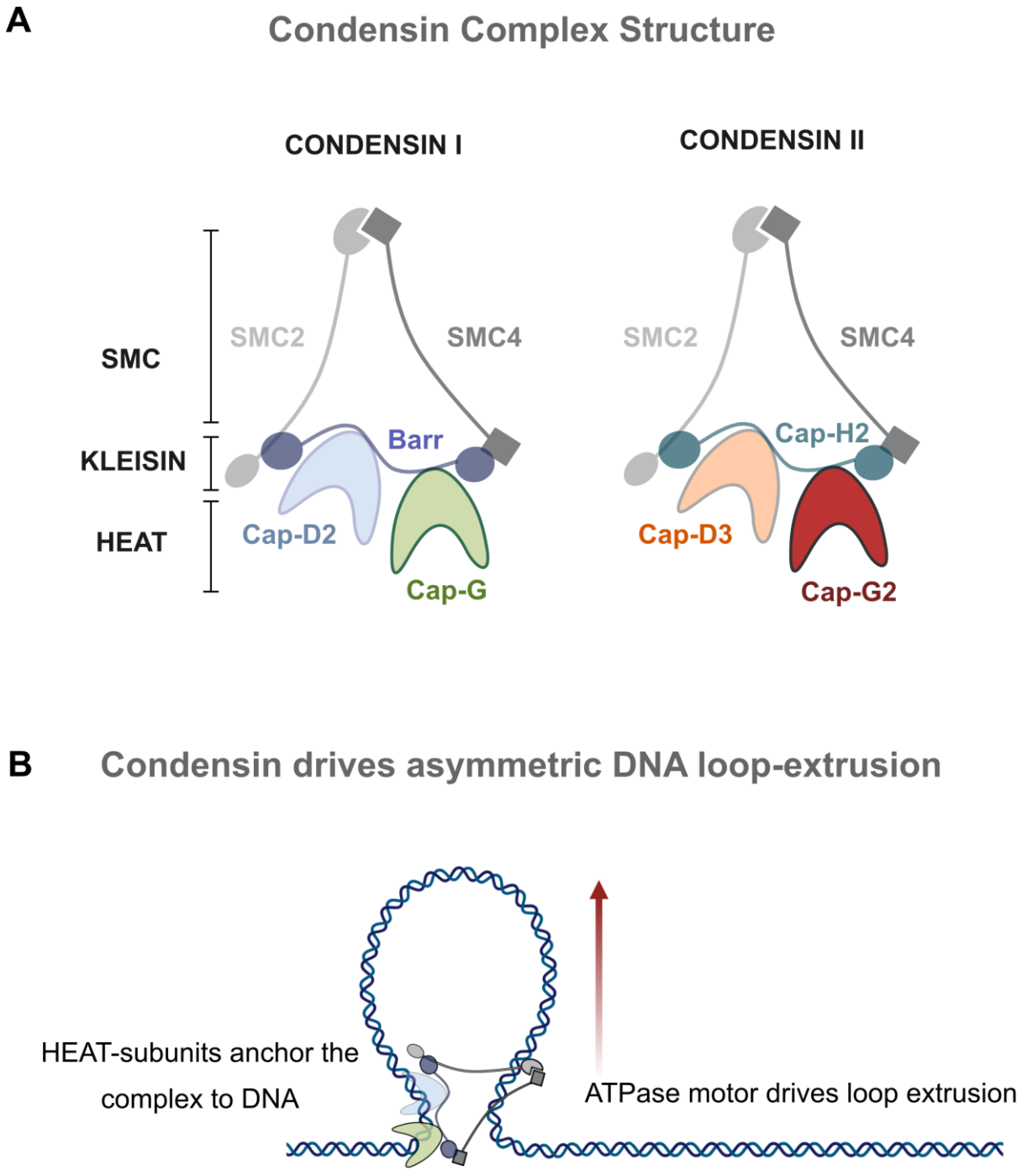


Figure 1.3. Condensin structure and loop extrusion mechanism

A) Structure of condensin I and II highlighting the SMC2/4-dimer and the non-SMC subunits; Barren/Cap-H2, Cap-D2/D3 and Cap-G/G2. **B)** DNA loop extrusion mechanism. Cap-D2 and Cap-G anchor to the DNA via their HEAT-motif. Barren forms a safety belt loop around the DNA and a mechanochemical motor asymmetrically pushes DNA through the SMC-dimer arms, forming a chromosomal loop. (Created with BioRender.com)

1.8.2 Chromosome assembly and segregation

Over the last two decades studies have focused on describing the role of condensin complexes in chromosome condensation, segregation and ultimately cell division (60). In most eukaryotes condensin I and II carry out overlapping, as well as distinct roles during mitosis. The complexes are both involved in mitosis but at different stages, as each complex is sequentially recruited to accurately shape chromosomes during mitosis. Condensin II is present in the nucleus during interphase and contributes to early phase chromosome assembly, whilst condensin I can only access chromosomes after breakdown of nuclear envelope (60). A simplified current model suggests that condensin II binds to DNA and extrudes DNA loops at the S-phase, driving the shortening of chromosomes. Meanwhile condensin I initiates lateral compaction of the chromosomes and further organise the DNA loops to form mitotic chromatids and then chromosomes (60).

Every subunit of the condensin complex plays an essential role in this intricate chromosome condensation process. Disruption of the complex leads to failed mitotic events and often lethality of the organism. For example in *Drosophila,* mutations to the subunit Barren lead to abnormal chromosome segregation during anaphase; resulting in embryonic lethality (76). Moreover, *Drosophila* metaphase cells with depleted Barren result in over-compaction of chromosome regions, leading to failed segregation (77). The subunit Cap-D2 is essential for centromere resolution and complex I stability (78). Meanwhile in *Xenopous* eggs, the subunits Cap-G and Cap-D2 are necessary for correct assembly of chromosome axes (79). In mouse NSCs, the depletion of condensin complex I and II leads to aberrant segregation and organisation of chromosomes, ultimately

resulting in cell death (80). Overall, the condensin subunits play an essential role in executing the canonical condensin function of mitotic chromosome compaction.

The subunit Cap-G has also been extensively studied across species. In *Drosophila, Cap-G* mutants are embryonic lethal (81). The *Cap-G* mutant flies display failure in chromosome condensation at early mitotic sages (prometaphase) followed by impaired sister chromatid separation at anaphase (81). Interestingly, in autosomes heterochromatin-rich regions were lost in *Cap-G* mutants, observed as a loss of bright, punctuate DAPI foci (81). Moreover, hemizygous flies for a semi-lethal *Cap-G* allele were lethal at the larval stage, and displayed severe mitotic phenotypes in larval brains (81). These mitotic cells failed to resolve sister chromatid arms, leading to chromosome segregation errors at anaphase and aneuploidy (81). Therefore, disruption of Cap-G in *Drosophila* severely impacts cell division and survival, emphasising the importance of this subunit during mitosis. Overall, these examples highlight that each condensin subunit is necessary for correct chromosome segregation and ultimately cell division.

1.9 Condensin and regulation of gene expression

The importance of condensin extends beyond its canonical function of physically compacting chromosomes. Growing areas of research aim to describe condensin roles linked to regulation of gene expression, transcription and chromatin organisation (Figure 1.4) (82,83).

1.9.1 Condensin and RNA Polymerases

The condensin complex has been implicated in promoting gene expression by interacting with transcription factors and localising at promoter and enhancer regions. Several studies describe the interaction between condensin and RNA Polymerases across

multiple species. For example, in *S. pombe* SMC2 binding overlaps with RNA Pol II transcription sites (84). Moreover, in mouse embryonic stem cells (ESCs), binding enrichment correlation was observed between condensin II and RNA Pol II (85). Finally, in chicken DT40 cells, condensin I binding overlaps with RNA Pol I-transcribed genes (86). Condensin binding is also documented at RNA Pol III (TFIIIC) regions such as tRNAs in multiple species (86-89). In yeast, condensin peaks coincide with tRNA genes and small cytoplasmic and nuclear RNA genes, but there is no evidence of condensin directly promoting transcription of these sites (87). A more recent study in mouse ESCs showed that Cap-H2 physically interacts and partially co-localises with TFIIIC (89). This has also been reported in *Drosophila* Kc167 cells, where condensin I (Barren) and II (Cap-H2) also localise at tRNA genes (90). The recruitment of condensin to tRNA gene regions is facilitated by TFIIIC (89,91). TFIIIC recognises a specific B-box sequence in the DNA, recruiting and facilitating condensin binding to that specific site (87). Taken together these studies suggest that condensin subunits can interact with RNA Polymerases and be recruited to specific DNA sites to potentially regulate transcription.

In bacteria *B. subtilis*, RNA Polymerases (RNAP) have been described as "moving-barriers" to condensin translocation on the DNA, slowing down loop extrusion (92). The speed of condensin translocation is dependent on the level of transcription, where highly transcribed regions, such as rRNA genes, significantly slow down condensin (92). The authors suggest that condensin translocation along DNA is affected by transcription, where condensin slows down and bypasses RNAP-dense regions to continue loop extrusion. Therefore, detection of condensin peaks at RNAP genomic sites may be an artefact of condensin stalling at those sites, rather than an indicator of direct binding of condensin and hence a correlation between condensin and transcription.

1.9.2 Condensin at Promoter / Enhancer regions

Condensin complex binding has been detected at enhancer/promoter regions however, there is conflicting knowledge on the localisation of condensin I vs II at these loci. For example, in *Drosophila* Kc167 cells, condensin II subunit Cap-H2 localises to and antagonises enhancer-promoter regions interaction (93). This contrasts previous reports from the same cells where Cap-H2 and condensin I subunit Barren overlapped with very few known enhancer regions (90). In mouse ESCs, recruitment of Cap-H2 to enhancer regions during interphase is mediator-dependant, and mutations to Cap-H2 resulted in reduced gene expression (85). Enrichment of Cap-H2 has been observed at promoter regions and the bodies of transcribed genes in chicken DT40 mitotic cells (86). Clearly, Cap-H2 is the most implicated condensin II subunit at enhancer regions across species, whilst condensin I localisation still remains poorly understood. Evidence from *C. elegans* suggests condensin I and II binding is enriched at transcription start site (TSS) of highly expressed genes (88). Moreover, a study on condensin I in human cancerous cells shows localisation of Cap-G to active enhancers, promoting estrogen-dependent gene expression (94). Cap-G appears to be required for effective enhancer promoter looping, though its localisation at enhancers is estrogen and interphase-dependant, with localisation lost in mitotic cells, as well as in the absence of estrogen stimuli (94). Taken together, these studies suggest that condensin loading onto enhancer regions may be dependent on interactions with co-factors, as well as the stage and type of cells analysed. Moreover, enhancer-localisation may be an evolved transcriptional regulation function, predominantly of condensin II as not enough evidence is yet present to define a clear role of condensin I in enhancer-promoter interactions.

1.9.3 Condensin DNA binding sites and TF interactions

Further research focused on dissecting the genome-binding profiles of condensin across multiple species, to identify a potential role for condensin in transcription regulation. In *S. cerevisiae* ChIP-seq experiments targeting condensin I subunits SMC4 and Barren (Brn1) have been carried out in G1-phase and metaphase cells (87). The authors observed that DNA-binding sites were unchanged between dividing and interphase cells, with an enrichment at centromere regions (87). A study in *C. elegans* carried out ChIP-seq analysis in interphase cells by tagging each condensin subunit of condensin I and II (88). Condensin II binding sites overlap significantly with binding sites of LIN-13 (88), a zinc-finger TF involved in vulval development (95). This interaction led the authors to suggest a model in metazoans where TFs recognise specific motifs and recruit condensin to the DNA binding site (88). Evidence for this model are seen for condensin II in *Drosophila*, where Cap-H2 interacts with TF Mrg15 (96). The loading of Cap-H2 onto chromatin is regulated by Mrg15, and this interaction maintains chromatin compaction in interphase cells (96).

Some condensin I subunits also rely on co-factor proteins for loading onto the DNA. For example in yeast quiescent cells, Cap-H binding is dependent on TF Msn2, as upon depletion of Msn2 there is a global reduction in condensin binding (97). In human breast cancer cells condensin I interacts with TF Estrogen Receptor α (ER- α) (94). ChIP-seq for the Cap-G subunit showed a large overlap with ER- α peaks, also seen with subunit Cap-H2 (94). Interestingly, Cap-G peaks were enriched at intergenic and intronic regions with less than 5% of peaks detected at gene promoters (94). The authors report that Cap-G physically interacts with ER- α via its DNA- binding domain and this interaction is essential for successful loading of condensin onto active enhancers in response to

estrogen stimuli (94). These studies present evidence that TF/condensin interactions can modulate different gene regulatory functions of condensin, which can be dependent on stimuli and cell-type.

1.9.4 Condensin and cell-specific gene expression

Studies in interphase and mitotic cells show that condensin I and II regulate cell-specific gene expression in multiple species. For example in mouse ESCs, condensin II is associated with activating the expression of cell-specific genes (85,89). In human embryonic kidney cells, Cap-H2 is associated with activating expression of highly transcribed, housekeeping genes (89). Moreover, in human breast cancer cells, condensin I (Cap-G) knockdown leads to reduced expression of estrogen-target genes (94). There is evidence for condensin-mediated gene expression in post-mitotic cells. For example in *Drosophila* post-mitotic fat body cells, Cap-D3 regulates transcriptional activation of anti-microbial gene clusters, suggesting a role for condensin II in immunity (98). Finally, another study suggests that T-cells in mice fail to differentiate upon depletion of Cap-H2, as condensin II is required for gene silencing (99). These studies suggest that condensin complexes functions can be cell-specific and vary between different cell types, raising questions on the role of condensin in unexplored cells, especially in a post-mitotic state.

1.9.5 Condensin and gene silencing

Gene silencing is essential in avoiding premature, or inappropriate, transcription of genes in any given cell. The condensin complexes have been associated with gene repression in multiple species and across several cell types. One suggested mechanism of condensin-mediated gene silencing is by direct binding of subunits to specific regions. For example in *C. elegans,* the DCC binds to specific DNA sequences known as recruitment on X (*rex*) sites (100). This binding leads to transcriptional repression of X-linked genes by half, to

balance expression in hermaphrodites (64). Moreover, condensin II binding has a suggested direct repressive effect, as mutants display an upregulation of gene expression due to disrupted gene silencing (88).

Important insights from *Drosophila* have revealed other mechanisms of gene silencing by condensin. For example, condensin II negatively regulates transvection to repress gene expression, which is a transcription regulation mechanism promoted by the direct interaction of homologous allelic regions (101,102). Furthermore, heterozygous flies for a *Cap-G* mutant allele display wing notches and a rough eye phenotype, attributing it to a regulatory role of Cap-G in heterochromatin regulation (81). Gene silencing by condensin has also been observed via its interaction with other complexes. For example, Barren interacts with the chromatin-repressing Polycomb complex to silence homeotic genes (103).

Chromatin compaction by condensin can impact gene expression. For example in yeast, chromatin compaction driven by condensin represses transcription in quiescent cells (97). This is supported by previous observations in mouse T-cells, in which condensin II depletion causes chromatin decompaction and an increase in gene expression, disrupting the quiescent state of cells (99). In *S. cerevisiae*, condensin is responsible for clustering of ribosomal DNA loci to the nucleolus as well as clustering of tRNA genes (91,104). This leads to silencing of adjacent RNA pol II transcriptional promoters, preventing transcription (105-107). Finally, some evidence suggests that silencing of gene expression by condensin has an impact on cell-differentiation. In murine cells, Cap-G2 has a potential role in erythroid cell differentiation, by repressing transcription of proliferation genes via chromosome condensation (82).

1.9.6 Contradictory evidence for a condensin role in gene regulation

Recent studies in yeast argue that condensin does not directly regulate gene expression. For example, in budding and fission yeast transcriptome alterations observed in condensin-mutant cells are due to aberrant segregation of chromosomes (108). Accumulation of RNA levels in the cells is an indirect effect of chromosome missegregation, which results in reduced RNA degradation in post-mitotic cells (108). Therefore, the authors suggest that all previously observed gene regulation functions by condensin are an artefact of failed chromosome segregation, challenging the current understanding of condensin's role in gene expression. Recent studies also argue that condensin has no direct effects on gene expression. For example in yeast, the depletion of Cap-D2 and Cap-G has no effect on overall gene expression (109). Moreover, in post-mitotic mouse hepatocytes, depletion of condensin II subunit Cap-H2 resulted in modest changes to gene expression, hardly affecting the transcriptome of the cells (110). These findings reveal that further research is needed to understand how condensin function in gene expression may have evolved across species and whether it is cell-type dependant, especially in a post-mitotic context.

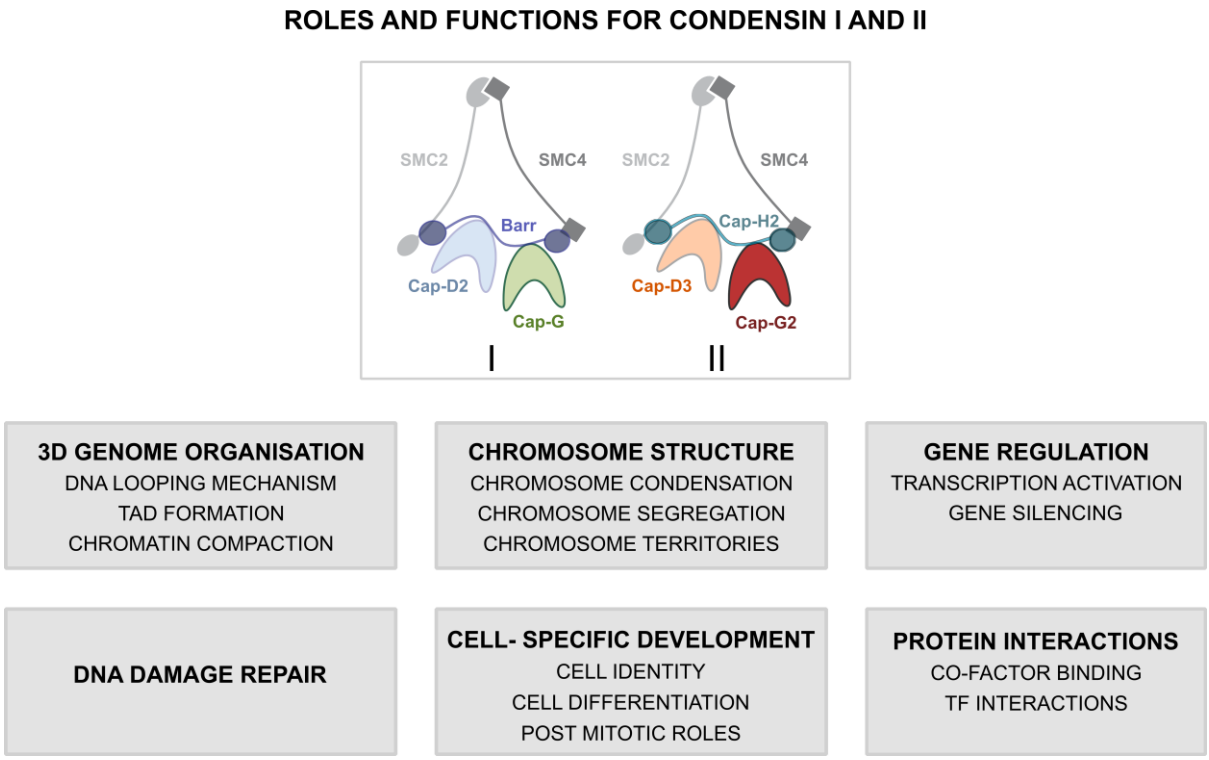


Figure 1.4. Summary of known condensin functions.

1.10 Condensin and 3D genome organisation

1.10.1 Chromatin Topology

The metazoan genome is organised in 3D architectural structures, which are dynamically formed depending on cell-cycle and developmental stages of a cell (111). These structures are organised in hierarchical order and have multiple size scales. The first and smallest feature is the "beads-on-a-string" form where double-stranded DNA wraps around nucleosomes (112). In the next secondary structure, the nucleosomes fold and are stacked into a solenoid form, producing a helix of adjacent nucleosomes. Subsequently, ranging between 100 kb to 1 Mbp in size are the topologically associated domains (TADs), which enclose genomic regions that interact with each other at higher frequency than with the rest of the genome (113-115). TADs are distinct domains separated by boundaries enriched in CTCF binding sites and highly transcribed genes (115-117). The next structure in the topology hierarchy is the chromatin compartments, these are large, multi-mega base genomic regions that can interact with each other over long-range distances. The genome is split into two main chromatin compartments, A and B, where A is active chromatin which is largely transcribed and B is mostly inactive chromatin, with some transcribed regions (117,118). Finally, the largest chromatin topology observed is the segregation of chromatin into distinct chromosome territories (CT) (111).

The mapping of the 3D organisation of the genome has started to emerge in recent years due to the development of molecular methods and cheaper high-throughput sequencing (117). Chromatin conformation capture methods such as 3C and Hi-C are being used to detect long-range interactions between genomic loci, generating contact maps which display a genome-wide view of interactions (117). Moreover, the development of

techniques such as HiChIP have allowed the profiling of long-range contact interactions of a protein of interest (119). Overall, these methods are advancing our understanding of how the 3D genome is organised, what proteins are involved in orchestrating this organisation as well as explain the relationship between transcription and chromatin topology.

Chromatin states are defined genomic regions, generally based on the binding of key proteins or histone modifications associated with active transcription or repression. The main chromatin states in *Drosophila* encompass two permissive states termed "yellow" and "red", where histone marker H3K27ac is a hallmark of the "red" state (120). In addition, there are two heterochromatin sates marked by the HP1 and PcG (Polycomb Group) proteins and finally a third repressive state, termed "Black" chromatin (120,121). This chromatin partitioning in *Drosophila* is known as the five-colour state model (121). To aid the discovery of these 5 chromatin states the accuracy of DamID was exploited to profile the binding of 53 different chromatin-associated proteins, which varied from enzymes to transcriptional machinery components and known complexes, resulting in high resolution, genome-wide location maps of said factors (121). Using computational analysis and classification strategies, the binding of these proteins was broadly partitioned into the 5 colour states known and widely used today in *Drosophila* (120,121). Since establishing this model, the *Drosophila* chromatin landscape has been partitioned further, where profiling of 18 key histone modifications and subsequent computational modelling resulted in a 9-states model (122). Overall, this research highlights how dynamic chromatin is and further insights suggest that states are variable and dependent on the cell type and stage analysed (120).

Recent studies focused on understanding the relationship between chromatin topology and chromatin states, to further understand whether it is the 3D topology that determines the state of chromatin, or vice versa (117). Chromatin compartments are large structures that encompass active (A) and inactive (B) regions of the genome. However, evidence suggest that smaller "compartmental domains" exist and segregate into A/B domains, but at a smaller scale similar to known accessible and inaccessible regions (~10 kb) (123). These compartmental domains have been identified in *Drosophila,* and their formation is dependent on the chromatin state of those regions (123). The authors profiled interactions between inactive (H3K27me3) and active (H3K27ac) histone features using HiChIP and compared them to known Hi-C data, revealing that the histone interactions correlated with A/B compartments even at a smaller scale (~ 10 kb) (123). Altogether this work suggests a relationship between chromatin topology and chromatin states, where chromatin states display interactions which lead to the formation of active/ inactive compartments. However, further research is necessary to understand how chromatin compartments are established and what other proteins and mechanisms are necessary for this 3D genome organisation.

1.10.2 Condensin and 3D chromatin organisation

The tertiary organisation of the genome is carefully orchestrated by architectural proteins that bind to chromatin to reorder and shape its 3D structure. Cohesin and condensin are chromatin associated complexes known to have a role in genome organisation. At the higher organisational scale, there is evidence that condensin has a role in maintaining nuclear structure, chromosome compaction and territories. In *Drosophila,* condensin II subunit Cap-H2 distributes heterochromatin and centromeric regions at specific loci to the nuclear periphery, forming distinct CTs (124). Depletion of

Cap-H2 and Cap-D3 disrupts nuclear architecture in *Drosophila* Kc167 and S2 cells, leading to deformed, larger nuclei (125). Moreover, Cap-H2 is necessary for chromosome compaction and homolog pairing in *Drosophila* (96). At the smaller scale, the cohesin complex has been shown to be essential in genome organisation, and when depleted of cohesin TAD boundaries are lost (63,117,126,127). However, a role for the condensin complex in TAD organisation remains largely unknown. Some research from *Drosophila* Kc167 cells suggests that condensin II is necessary for formation of TADs, as Cap-H2 localises at strong TAD-borders and if depleted leads to disruption of TAD structures (90,93). In *C. elegans*, the DCC complex (related to condensin I) binding at endogenous *rex* sites in the genome promotes TAD formation on the X chromosome (128). Moreover, in yeast cells, mutations to the condensin I subunit Cap-H2 resulted in global decompaction of the genome, where a shift from short to long- range interactions and loss of locus-specific interactions is observed (109). Overall, further research is necessary to dissect the role of condensin in the organisation of chromatin topology.

The functionality of TADs in the genome has been recently disputed, as disruption of TADs in multiple species resulted in minimal changes to gene expression (129-132). For example in *Drosophila,* extensive genetic alterations that fuse, delete and shuffle TADs resulted in modest changes to gene expression (130). Moreover in yeast, genome decompaction and loss of short-range interactions did not result in significant changes in gene expression (109). In addition, there is conflicting evidence on the role of condensin in TAD formation and maintenance. A study in post-mitotic murine hepatocytes shows that disruption of Cap-H2 has no effect on TAD formation (110). These studies question how important chromatin topology really is for maintenance of the transcriptome. Taken altogether, the condensin complex has a better established role in higher chromatin

organisation, whilst its role in forming or maintaining smaller, active/inactive compartments and/or TADs remains unclear.

1.11 Condensin and the nervous system

1.11.1 Chromatin remodelling and maintenance of neuronal fate

Accurate regulation of gene expression is key in ensuring appropriate differentiation, maturation and maintenance of neuronal fate. The molecular mechanisms underlying gene expression changes involve a complex interplay of transcription factors and chromatin remodelling. The activation or repression of genes depends on the accessibility of the chromatin at a given stage. Cell differentiation drives the compaction of chromatin into a more inaccessible state, whilst in dividing and early differentiating cells the chromatin is more open and accessible (133).

Chromatin dynamically changes as a cell transitions from a stem cell towards a differentiated neuronal state (7,14). For example in neuronal progenitor cells, certain enhancer/promoter regions were accessible to promote neurogenesis, which then were compacted in differentiated neurons (8,38). In *Drosophila* a five colour chromatin state model has been suggested, in which states are defined by the occupancy of key proteins (120) (Section 1.10.1). In adult neurons, most NSC-specific genes were encompassed within a repressive HP1 state, whilst in NSCs, neuronal genes were in a highly silenced "Black" chromatin state (7). Therefore, the chromatin sate and thus the genomic binding of the key proteins also change upon neuronal differentiation (7,120). During neuronal maturation, there are also gradual changes that occur to the chromatin accessibility of the cell. A newly differentiated neuron has more plastic and accessible chromatin, compared to more compact and inaccessible chromatin in mature neurons (14). Moreover, the transition from an immature to a mature neuron is also coupled with

changes to the chromatin states (7). Overall, these studies suggest that chromatin remodelling is in place to restrict cell fate plasticity, although the mechanisms and factors driving this process remain largely unexplored.

Dissecting the relationship between chromatin associated complexes and neurodevelopment could provide insights on the epigenetic changes that drive neuronal maturation. There is limited knowledge on the factors involved in progressing neuronal maturation, although recent research has started to progress this field (1,6-8). In the Southall lab the neuron-specific transcription factor Lola-N has been found to interact with condensin I subunit Cap-G in a Y2H screen (unpublished, Southall, T.D) (Section 1.4). Given that the condensin complex is involved in chromatin compaction and remodelling in mitotic cells (53), this interaction suggested that a similar role could be adapted in neurons, and it could be potentially relevant in neuronal differentiation and maturation. Therefore, I will explore our current knowledge of condensin in neurodevelopment, highlight the gaps in the field which pose the key questions addressed in this work.

1.11.2 Condensin mutations impact neurodevelopment

All the literature presented above highlight the dynamic roles condensins have in regulating gene expression and ultimately impacting cell identity and fate. Condensin mutations in mitotic and interphase cells eventually disrupt post-mitotic daughter cells, leading to severe phenotypes. In mouse NSCs, condensin II mutations disrupted nuclear architecture and resulted in apoptosis of NSCs and post-mitotic neurons (80). In *Drosophila*, condensin I is implicated in regulation of position effect variegation (PEV) (78,81,134), where heterozygous *Cap-G* mutants showed wing notches and rough eyes phenotypes in flies (81). Furthermore, *Barren* mutants displayed loss of peripheral nervous system neurons and an increase in the size of remaining neurons (76). In

zebrafish, *Cap-G* mutants display severe apoptosis of proliferative retinal cells, but not of differentiated retinal cells (135). These studies are examples of how impacting condensin in mitotic cells has repercussions on development of neuronal daughter cells. However, none of these studies explore condensin depletion in a post-mitotic context in neurons, leaving a considerable gap in our knowledge of condensin functions beyond cell division.

1.11.3 The gap in our knowledge: a post-mitotic role of condensin

Virtually all the studies to date have observed the roles of condensin in a mitotic or interphase cell-state, but there is limited knowledge of cell-specific, post-mitotic roles for condensin. However, evidence in *Drosophila* suggests a post-mitotic role for condensin II in immunity, where Cap-D3 regulates transcriptional activation of anti-microbial gene clusters in post-mitotic fat body cells (98). There is also evidence that cohesin, closely related to condensin, has a post mitotic role in *Drosophila* (136). The cohesin complex was found to be expressed in neurons, and post-mitotic knockdown of cohesin in neurons resulted in lethality at the pupal stage (136). Post-mitotic knockdown of cohesin in mushroom body neurons and cholinergic neurons resulted in defects in axonal pruning and impaired locomotion in larvae respectively, suggesting a non-mitotic role for cohesin in the *Drosophila* CNS (136). Condensins and cohesins are closely related multi-protein complexes, therefore, it is possible that condensin could also play a role in post-mitotic neurons.

1.12 Summary of project aims

Despite the wealth of knowledge on condensins, their role in post-mitotic cells remains largely unexplored. This project aims to describe a novel role for condensin I subunit Cap-G in *Drosophila* neurons. Condensin may be involved in altering the epigenetic state of

neurons, promoting differentiation, maturation and ultimately maintaining neuronal fate. Therefore to address this hypothesis in this thesis I planned to achieve the following:

- Characterise *in vivo* condensin I complex localisation and expression in neurons
- Generate neuron-specific Cap-G knockdown (KD)
- Identify effects of neuron-specific Cap-G KD on animal behaviour and survival
- Identify effects of neuron-specific Cap-G KD on neuronal cell health
- Identify effects of neuron-specific Cap-G KD on neuronal transcriptome
- Profile Cap-G binding to DNA in neuronal stem cells and neurons

Chapter 2 Materials and Methods

2.1 *Drosophila* melanogaster: Husbandry and maintenance

The Bloomington Stock Centre provided the fly strains in this study, unless stated otherwise. Standard maintenance procedure on medium and at 25 °C was followed for all the *Drosophila* stocks. Table 1 shows a list of the genotypes used and generated in this study.

Table 2.1 List of *Drosophila* strains used in this study

Genotype	Description	Source
w^{1118}	Wild type	
P(GawB)elav^{C155} –GAL4 (elav-GAL4)	Pan-neuronal driver for immature and mature neurons.	#458
w1118; P(y+t7.7 w+mC=GMR57C10-GAL4)attP2 (nSyb-GAL4)	Neuronal driver for mature neurons that are forming synapses.	#39171
w; ; P(w^{+mC}=UAS-Nslmb-vhhGFP4) (w*;;UAS-deGradFP)*	deGradFP reporter line to allow cell-specific degradation of GFP-tagged protein. (137)	# 38421
w; Cap-G^{EGFP}*	GFP-tagged Cap-G from its endogenous locus (138)	Courtesy of the Heidmann Lab
Ay-GAL4,UAS-myr-td-Tomato/Cyo	Ubiquitous driver with FRT cassette for generating FLP-out clones.	Courtesy of the Brand Lab
Cap-G^{EGFP}; UAS-Nslmb-vhhGFP4 (Cap-G^{EGFP};UAS-deGradFP)	DegradFP fly line for cell-specific *Cap-G^{EGFP}* degradation.	I generated it in the Southall Lab.
Ay-GAL4,UAS-myr-td-Tomato/Cyo; UAS-deGradFP	Ubiquitous driver with FRT cassette to generate FLP-out clones expressing degradFP.	I generated it in the Southall Lab.
elav-GAL4; Cap-G^{EGFP}	Neuronal driver line in a *Cap-G^{EGFP}* genetic background.	I generated it in the Southall Lab.

w; ; Daughterless-GAL4/TM6*	Ubiquitous driver.	# 55851
w; Cap-G^{EGFP}; Daughterless-GAL4/TM6*	Ubiquitous driver line in a *Cap-G^{EGFP}* genetic background, for positive controls.	I generated it in the Southall Lab.
w; Cap-G^{EGFP}; nSyb-GAL4*	Mature neurons driver line in a *Cap-G^{EGFP}* genetic background.	I generated it in the Southall Lab.
hs-FLP; Cap-G-EGFP / CyO	*Cap-G^{EGFP}* with flippase for generation of stochastic clones.	I generated it in the Southall Lab.
elav-GAL4;;tub-GAL80^{ts}	To drive Dam-expression in neurons for TaDa experiments.	Southall Lab stocks.
w; tub-GAL80^{ts};nSyb-GAL4*	To drive Dam-expression in mature neurons for TaDa experiments.	Southall Lab stocks.
wor-GAL4;tub-GAL80^{ts}	To drive Dam-expression in NSCs for TaDa experiments.	(139).
yw, hs-FLP; If/CyO	Heat shock-activated flippase line.	Southall Lab stocks.
UAS-Cap-G-Dam	Dam-fused Cap-G to detect specific DNA binding using Targeted DamID (140)	I generated it in the Southall Lab.
UAS-LT3-Dam	Dam-only control for background binding.	(141)
UAS-Lola-N-Dam	Dam-fused Lola-N to detect specific DNA binding using Targeted DamID. (19,140)(19,140)(19,141)(Marshall et al., 2016, Southall et al., 2014)(Marshall et al., 2016, Southall et al., 2014)(Marshall et al., 2016, Southall et al., 2014)(20,141)(19,140)(19,140)(19,140)(19,140)	
UAS-mCD8-GFP (III)	Fluorescent reporter line.	#30003

2.2 Cell-specific degradation of GFP-tagged proteins using the deGradFP system

The deGradFP system (137) (described in Chapter 3, Section 3.1.3) is used to drive cell-specific degradation of the Cap-G^{EGFP} protein. All driver and reporter lines were put in a homozygous Cap-G^{EGFP} background to ensure no wild type Cap-G was present in the animals (Table 2.1). As a positive control for the system, ubiquitous degradation of Cap-G^{EGFP} was driven using *w*; Cap-G^{EGFP}; Daughterless-GAL4/TM6* crossed to *Cap-G^{EGFP}; UAS-deGradFP*. For cell-specific knockdown experiments, the drivers *elav-GAL4; Cap-G^{EGFP}* and *w*; Cap-G^{EGFP}; nSyb-GAL4* were crossed to *Cap-G^{EGFP}; UAS-deGradFP* to drive degradation of Cap-G^{EGFP} in all neurons and mature/synapse forming neurons respectively. The negative controls used for knockdown experiments involved crossing *elav-GAL4* and *nSyb-GAL4* to *UAS-deGradFP*. Finally, the lines *Cap-G^{EGFP}*, *elav-GAL4; Cap-G^{EGFP}* and *w*; Cap-G^{EGFP}; nSyb-GAL4* and *Cap-G^{EGFP}; UAS-deGradFP* were also used as negative controls for behavioural assays when appropriate (Section 2.5).

2.3 Generation of stochastic clones

To induce Cap-G^{EGFP} degradation in stochastic cell clones as a positive control, I combined the deGradFP system with the heat shock 'flip-out' (hs-FLP) system (142). To obtain clones in the optic lobes of the CNS, the lines *hs-FLP; Cap-G^{EGFP}* and *Ay-GAL4,UAS-myr-td-Tomato/Cyo; UAS-deGradFP* were crossed and the progeny larvae were heat shocked at 48 or 72 hrs post laying by incubating at 37 °C for 35 mins. Once they developed into 3rd instar larvae they were dissected and stained accordingly (Section 2.6).

2.4 Embryo microinjection

Fly crosses were setup in collection cages with apple juice agar plates to promote egg laying [3% Agar, 30% apple juice and 2% methyl paraben dissolved in 100% ethanol]. Embryos were collected every 30 minutes and de-chorionated manually on a sticky slide covered with double sided-tape. The embryos were gently rolled over, allowing the chorion to stick to the tape and the de-chorionated embryos were gently placed on clear apple juice agar. Embryos were then aligned on a slide covered in a layer of glue (Scotch, double-sided tape dissolved in heptane), to ensure correct orientation and no movement during injections. Embryos were left to air-dry on the slide for 8 minutes and then covered in Halocarbon oil (Sigma). The DNA sample was added to a glass micro-capillary needle and injected into the posterior end of 150-200 embryos. Post-injection, the embryo slides were left at RT on apple juice plates and larvae were collected and placed in normal food vials to develop to adult animals. The adults that successfully eclosed were crossed to w^{1118} and the resulting progeny were allowed to develop into adults and screened for orange eyes. The orange eyes were a marker for successful transgenic insertions, so those adults were collected and crossed to a balancer line of interest, generating the fly-line desired.

2.5 Behavioural and phenotypic assays

All animals were kept at 25 °C unless otherwise stated.

2.5.1 Developmental Survival Assay

For the developmental survival assay, three biological replicates per genotype were used. Flies were allowed to lay for 5 hrs and then 100 embryos per replicate were collected. After 24 hrs, L1 larvae were transferred to a food vial and allowed to develop. At 5 days post-laying the number of pupae was scored. Finally, at 10 days post-laying the total

number of eclosed adults were recorded. Statistical significance was measured using Logrank and the Gehan-Breslow-Wilcoxon tests in Graphpad Prism version 8.

2.5.2 Survival assay

Survival assay in adults was performed by having three technical replicates, with 20 flies per replicate, separating male and female animals upon eclosion. Death was scored daily until total number of flies deceased. The flies were transferred to fresh food vials every 3 days. Survival curve and statistical significance was measured using Logrank and the Gehan-Breslow-Wilcoxon tests in Graphpad Prism version 8.

2.5.3 Locomotion Assay

Locomotion assay was performed on 3rd instar larvae using 10 biological replicates and 3 technical replicates. Individual larvae were setup in clear agar plates against a 0.5 cm^2 square grid. Larvae were left to acclimatise for 1 minute and locomotion was video-recorded. Videos were analysed using Icy-Bioimage Analysis (143) where larval crawling was visually tracked and locomotion was scored as distance (number of squares) covered per minute (144).

2.5.4 Negative Geotaxis (climbing) assay

Negative geotaxis assays were performed on 10 male flies and 10 technical replicates per genotype (144). Flies were set up in clear climbing vials against a grid and a camera. Video recording was started before sharply tapping flies to the bottom of vials. Videos were analysed using Icy-Bioimage Analysis (143), the frame in which the first fly reached the top of the vial was extracted. The coordinates of each fly (bottom of vial to fly's topmost distance) were used to calculate the average distance climbed per replicate. Statistical

significance between samples was measured by Kruskal-Wallis one-way ANOVA test and plots were generated using the software GraphPad Prism version 8.

2.6 Immunostaining and microscopy

2.6.1 Larval and adult brain tissue

The brain tissue was collected at 3rd instar larvae and adult (5 days old) stages by dissecting individuals in 1x PBS. Tissue was fixed for 25 minutes in 4% formaldehyde (Sigma) diluted in PBST (0.3% Triton X-100 in PBS, Sigma). Steps were carried out at room temperature unless stated otherwise. All washes were done in PBST for 10 mins intervals. Upon fixative removal, the brains were washed three times for 30 mins. After washing, the tissue was incubated in blocking serum (2% Normal Goat Serum (NGS) in PBST) for 15-30 mins. The tissues were incubated with primary antibodies at 4 °C overnight. Subsequently, brains were washed three times (5 mins washes) and then incubated with secondary antibodies for 1.5 hrs, rocking at RT. Finally, tissues were washed six times (10 mins washes) and mounted in Vectashield mounting medium (Vector laboratory). Glass microscope slides and 22 mm squared coverslips were used and sealed with clear nail varnish.

2.6.2 Embryonic tissue

Embryos were kept at 25 °C and collected every 12-15 hrs. Embryos were placed in a small sieve and washed with distilled water. The embryos were de-chorionated by submerging them in 50% bleach (diluted in water) for 2 mins. All bleach was carefully washed off the embryos using distilled water. The embryos were placed in a 2 mL tube containing fixative solution in a 1:1 ratio of 4% formaldehyde: Heptane (Sigma), vortexed for 30 seconds and then placed on a rotating wheel (speed 40 rpm) for 20 mins. The lower fixative phase was replaced with 100% methanol and vortexed to remove the vitelline

membrane, allowing embryos to sink and be collected. Embryos were rinsed twice in 100% methanol, at this stage they were either stored at -20 °C or re-hydrated for 5 mins in 50% methanol (in 1 x PBS). After re-hydration, embryos were washed three times in PBST at 5 minute intervals. Blocking and antibody staining procedure as stated in section 2.6.1 was performed. Stained embryos were serially incubated in 50%, 70% and 90% glycerol (in distilled water) before mounting on microscope slides.

Table 2.2 List of antibodies used in this study.

Primary Antibody	**Source**	**Final Concentration**
Anti-GFP (chicken)	Abcam	1:2000
Anti-HP1 (mouse)	DSHB	1:500
Anti-Lola-N (rabbit)	Southall Lab (19)	1:10
Anti-elav (c) (rat)	DSHB	1:500
Anti-deadpan (guinea pig)	Brand Lab	1:1000
Anti-pH3 (rabbit)	Millipore	1:500
Anti-ɣ-H2AV (mouse)	DSHB	1:200
Anti-RFP (rat)	Takara Bio	1:500
Secondary Antibody		
Alexa Fluor® 488	Life Technologies	
Alexa Fluor® 546	Life Technologies	1:200

Alexa Fluor® 633	Life Technologies	
Dyes		
DAPI (DNA stain)	Thermo Fisher	1:20000

2.7 TUNEL Staining

For *in vivo* detection of apoptosis in embryos, I utilised the TUNEL (terminal deoxynucleotidyl transferase-dUTP nick end labelling) assay using an adapted commercial kit, the Click-iT® TUNEL Alexa Fluor® 549 (ThermoFisher). The TUNEL assay marks for apoptotic cells by introducing modified dUTPs at the 3'-OH ends of fragmented DNA (hallmark of apoptosis) using the TdT enzyme (terminal deoxynucleotidyl transferase). The Click-iT® TUNEL assay was chosen as it allows for fluorescent detection of apoptosis and can be used in conjunction with antibody staining, to identify cells of interest. Click-iT® TUNEL assay involves a two-step process: firstly the TdT reaction which attaches the EdUTPs at the site of DNA damage, secondly the click-chemistry technology attaches a fluorescent dye onto the EdUTPs. In this experiment the provided protocol was followed with minor changes, highlighted below.

The embryos were collected and fixed as previously described (Section 2.6.2). The fixed embryos were incubated in 100 µl 1x Proteinase K for 30 mins at RT and subsequently post-fixed in 4% formaldehyde for 15 mins at RT. The embryos were washed 3 times in PBST (5 mins) and then rinsed in 1x PBS to remove all detergent. The tissue was incubated in TdT Reaction Buffer for 10 mins at 37°C. After incubation, the TdT reaction mixture containing TdT buffer, EdUTPs and TdT enzyme (as per protocol) were added and incubated at 37°C for 1 hr. The embryos were rinsed with ddH_2O and washed with

PBST (5 mins) and then rinsed with PBS. The Click-iT® TUNEL reaction mixture containing the dye and catalyst (per protocol) is added to the embryos and incubated at 37°C for 30 mins, protected from light. Finally, the embryos are washed in PBST (2x 5 mins) and then rinsed with PBS, ready for blocking and primary antibody incubation as previously described (Section 2.6.2).

2.8 *In-situ* hybridisation

Single Molecule Fluorescent *in-situ* hybridisation (smFISH) was performed on wild type (w^{1118}) embryos and *elav-GAL4;;UAS-mcd8-GFP* larval brains to detect *Cap-G-mRNA* in neurons. The tissues were treated following a previously published protocol, which can be combined with immunofluorescence to co-visualise transcripts and proteins (145). The RNA probes were designed using LGC Biosearch Technologies' Stellaris® RNA FISH Probe Designer. There are two known isoforms for *Cap-G* in flies, *Cap-G-RG* and *Cap-G-RF* and a total of 6 exons, with exon 4 being the largest and conserved in both isoforms. To minimise background the probes were designed against *Cap-G* exon 4, generating 48 probe sequences. The *Cap-G* probe was generated by LGC Biosearch Technologies with fused Quasar® 570 dye. Multiple molar concentrations of the probe were tried, and 3 µM seemed to give the best imaging results.

All buffers and reagents are used as per published protocol unless otherwise stated (145). The tissue was dissected and fixed. After fixation, the tissue was rinsed in 0.3% PBS-Tween-20 (1 x PBS with 0.3% Tween-20 (ThermoFisher)). The tissue was blocked in 2% NGS for 15-30 mins at RT. The tissue was incubated in Wash Buffer for 5 mins at 37°C. The tissue was then incubated at 37°C for 8-15 hrs (maximum) in 100 µl of hybridisation buffer including 3 µM *Cap-G* probe and any primary antibody of interest. Sample was then rinsed 3x in Wash Buffer and followed by three 15 mins washes in Wash Buffer at RT.

After the washes, the tissue was incubated with appropriate secondary antibodies and then mounted, ready for fluorescent imaging (Section 2.6).

2.9 Image acquisition and analysis

Samples were imaged using Zeiss LSM 510 and Leica CF8 (courtesy of the Imperial College FILM facility) confocal microscopes. Analysis of acquired images was performed using Fiji (146), Icy-Bioimage Analysis (143), and Ilastik (147) softwares.

2.9.1 Average nuclear area quantification

DAPI staining distinctly marks the DNA in the nuclei of cells, so it was used as a reference to measure the nuclear area of neurons in embryos. To automate the nuclear area detection, the Ilastik software was used for bulk image segmentation (147). The software operates on machine learning principles, so it was given 3 sets of "training" images with the cell (DAPI +) vs background manually highlighted, to help the software identify the cells. The images are then segmented to provide distinct enclosed shapes (nuclei) in a black and white image. After training, the software was able to apply the algorithm to a whole set of new images and create accurate segmentations. The segmented images are then analysed in Fiji using the Analyze Particles Plugin, where each nuclei is identified as a single region of interest (ROI) and the nuclear area in pixels is calculated. A total of 7 biological replicates (embryos) were used and in each replicate >1000 cells were segmented and the average area calculated.

2.9.2 Automatic quantification of cell numbers

The Icy plugin Spot Detector (148) was used to automatically quantify the total number of cells for NSCs (Dpn+), dividing cells (pH3+), apoptotic cells (TUNEL+) and DNA damaged cells (γ-H2Av+). For all analyses there was a Minimum Size Filtering of 10-25

pixels, the detector scales were adjusted per experiment and kept constant across control and experimental images (Table 2.3).

Table 2.3 Detector scales used per experiment.

Detector	Ph3+ cells	Dpn+ cells	TUNEL+ cells	ɣ-H2AV+ cells
Scale 1	disabled	disabled	disabled	disabled
Scale 2	50	disabled	disabled	50
Scale 3	100	100	disabled	100
Scale 4	disabled	150	40	100
Scale 5	disabled	disabled	100	disabled

2.9.3 Quantification of live and *in vivo* GFP signal intensity

To detected whether degradation of Cap-G^{EGFP} was prematurely occurring in dividing cells I quantified the live GFP signal intensity in dividing cells in Fiji. During imaging a z-stack of the embryonic VNC was taken to cover most of the dividing cells layer (5-6 slices/ embryo). Actively dividing cells displayed highest GFP levels with distinct foci, so these were segmented taking the high intensity pixels using Otsu's thresholding (0, 2056). The segmented image is analysed using Analyze particles (setting size = 1.8 – Infinity) to detect dividing cells as individual, outlined ROIs. Mean pixel intensity per ROI (dividing cell) was calculated using Analyze Particles plugin in Fiji (> 30 cells/ embryo). Average pixel intensity of ROI was calculated across z-stack slices, for a total of 10 biological replicates.

To quantify the *in vivo* GFP levels in NSCs, images were analysed from three biological replicates per condition (Chapter 3- Section 3.2.9). The Dpn positive cells were selected as individual ROIs as described above. The average pixel intensity detected for GFP and for Dpn were extracted using the Analyze Particles plugin in Fiji (> 70 cells / embryo). The GFP / Dpn ratio of pixel intensity was calculated per cell. Statistical significance was analysed using a Nested t-test.

2.9.4 Statistical significance and plots

Upon extracting the quantitative data using the methods described above, the statistical significance between the experimental and control data was analysed using GraphPad Prism version 8 for Windows. For two-population experiments, non-parametric t-test's was used, to compare multiple groups Kruskal-Wallis test one-way ANOVA and significant difference displayed a p-value < 0.05. Plots were generated using GraphPad Prism version 8 for Windows.

2.10 DNA cloning

2.10.1 Genomic DNA extraction

A fly with the genotype of interest was ground in 'Squishing buffer' [10 mM Tris-HCL pH 8.2, 1 mM EDTA, 25 mM NaCl] with 200 $\mu g.mL^{-1}$ of Proteinase K at 37 °C for 30 mins. An incubation of 2 mins at 95 °C deactivated the enzyme and the sample was stored at 4 °C.

2.10.2 Polymerase Chain Reaction (PCR)

Genomic DNA or another template (e.g. *UAS-LT3-Dam*) were used for amplification of the strand of interest using a standard PCR cycle in the SureCycler 8800 (Agilent Technologies). The extension time and annealing temperature were adjusted according to the amplicon's size and primers' melting temperature. PCR mix is as follows: 0.5 U

Phusion® DNA Polymerase, 5x Phusion® buffer, 0.4 µM forward and reverse primers, 0.3 µM dNTPs and template DNA. Primers used are available in the Appendix. Diagnostics were done using agarose gel electrophoresis; 1.5% agarose in 1x TBE [90 mM Tris, 90 mM boric acid (pH 8.3), 2 mM EDTA] and 0.1 µg/ml GelRed™ (Biotium).

2.10.3 Competent cells transformation

The plasmid of interest was transformed in competent cells DH5α. All following steps were carried out at 4°C. 1-5 µg of vector DNA was added to the cells and incubated for 20 minutes. Cells were heat-shocked at 42°C for 30 seconds and then immediately placed on ice for 5 mins. 250 µl of SOC medium was added to the cells and incubated at 37°C for 1 hour to optimise growth. Cells were streaked on an L-Agar plate (Agar (Sigma) and 1x L-Broth (Miller)) with the relevant antibiotic to collect colonies the next day. Individual colonies were picked and grown overnight in L-Broth media with the relevant antibiotic. The plasmid was extracted from the overnight cultures according to instructions of the Sigma GenElute™ Plasmid purification Miniprep kit.

2.10.4 Gibson Assembly

PCR products of interest were cloned into a vector using Gibson assembly procedure (149). Equimolar concentration of PCR products to vector was added up to a 10 µl volume. Subsequently, 10 µl NEBuilder® HiFi DNA Assembly Master Mix was added to the sample for a total of 20 µl reaction. To control for successful ligation, a control restriction digestion was performed on the plasmids extracted. Successfully cloned samples were sent to Source Bioscience for DNA sequencing.

2.10.5 Restriction digest

Restriction enzymes are used to diagnose ligation success. Table 2.4 shows a summary of the enzymes used, all obtained from NEB. All reactions were carried out at 37°C for 2 hours.

Table 2.4 Enzymes used in this study.

Restriction enzymes	
EcoRI-HF®	BglII
XbaI	NotI-HF
XhoI	HindII
BbsI	

2.11 Targeted DamID

2.11.1 Cloning of *UAS-LT3-Cap-G-Dam* vector for transgenic fly generation

To profile cell-specific DNA binding for Cap-G I generated a *UAS-LT3-Cap-G-Dam* line to use in Targeted DamID experiments. The isoform *Cap-G-PF* was amplified by PCR from *Drosophila* cDNA library. The *Dam* and *mCherry* constructs used for the primary ORF in the TaDa cassette were amplified by PCR from *pUAST-LT3-Dam* (141). The *Dam* construct was fused on the C-terminus of Cap-G using fusion PCR at equimolar concentrations (1:1) of both *Cap-G-PF* and *Dam* using annealing temperatures of 55°C, 62°C and 68°C, which all provided positive construct results. The *pUAST-attB* (150) plasmid was linearised in a double-digest reaction using the restriction enzymes Not-I-HF and XhoI. Finally, a three-

way Gibson Assembly with the *mCherry* construct, *Cap-G-Dam* construct and the linearised *pUAST-attB* vector resulted in successful cloning of the *pUASTattB-LT3-Cap-G-Dam* plasmid required to generate the *UAS-LT3-Cap-G-Dam* transgenic line (Appendix). The vector was then injected in syncytial embryos (Section 2.4).

To infer cell-specificity in our Targeted DamID (141) experiments I used the following driver lines: *wor-Gal4; tub-Gal80ts* for neuronal stem cells, *elav-Gal4; tub-GAL80ts* for immature and mature neurons and *tub-GAL80ts; nSyb-Gal4* for mature neurons. The *UAS-LT3-Cap-G-Dam* line was used to profile Cap-G binding, the *UAS-LT3-Dam-Lola-N* (19) line for Lola-N binding and the Dam-only line *tub-GAL80ts; UAS-LT3-Dam* (141) was used as a control for background methylation and Chromatin Accessibility TaDa experiments (14).

2.11.2 Targeted DamID library preparation

UAS-LT3-Cap-G-Dam or *UAS-LT3-Dam-Lola-N* animals were crossed to the desired driver line and embryos collected at 25°C for 4 hrs. To obtain third instar larval tissue, embryos were raised at 18 °C for seven days post collection and subsequently placed at 29 °C for 24 hrs to induce Dam- expression. Two technical replicates were collected, with 30 larval brains per replicate dissected in PBS and kept in PBS-EDTA at –80 °C until processed. Extraction of Dam-methylated DNA from the tissue and generation of genomic libraries were performed as previously described (140). In summary, the genomic DNA is extracted using the QIAamp DNA Micro Kit (Qiagen) and methylated DNA is digested using DpnI enzyme. The digested DNA is purified to extract only the methylated sequences using PCR purification kit (Qiagen). Adaptors are ligated onto digested, methylated DNA fragments and a further DpnII digest ensures removal of any unmethylated DNA sequences. The adaptor-ligated DNA is purified and then amplified by

PCR. Finally, the amplified methylated DNA is sheared in a BioRuptor Plus sonicator (HIGH, 8 cycles, 30 sec ON/30 sec OFF). The sequencing DNA library is prepared as per protocol (140) and adaptors are chosen based on Illumina HiSeq recommendations. A maximum of 6 samples were pooled in a single library and ready for Illumina HiSeq 2500 single-end 50 bp sequencing.

2.12 Analysis of Targeted DamID sequencing data

Sequencing data obtained for both replicates was normalised against *Dam-only* control binding and mapped against the *Drosophila* genome (r6.03) using the published *damid_seq_pipeline* (140). This results in GFF files to be used for further statistical analysis. The normalised data was visually inspected in Signalmap (Roche).

2.12.1 Data accessibility and file formats for statistical analysis

All raw sequence files and processed files have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (accession number GSE142112). The files used for statistical analysis vary depending on the tools used. The output GFF files were converted into plain text bedGraph files using command line (Appendix). The bedGraph files were then converted to an indexed, binary bigwig file using the *bedGraphToBigWig* command which was compatible with most tools described below (151). Finally, if only coordinate files of regions were required, BED files would be used, which include the Chromosome, Start and Stop features only.

2.12.2 Peak calling

Significant peaks in the DNA binding profiles were called and mapped to genes using a custom Perl program that allows for the identification of broadly bound regions (available at https://github.com/tonysouthall/Peak_calling_DamID) (140). The peak

calling uses an algorithm where a false discovery rate (FDR) is calculated and assigned to all potential peaks. Then, only peaks with an FDR less than 0.01% were classified as significant. This outputs a file with the coordinates and name of each gene identified to be within 5 kb of a peak (with no intervening genes). Only significant peaks present in both replicates were used to form the final peak file, stored in a coordinates (BED) file for further statistical analyses. The peak files generated and the unique genes identified are available in this repository https://doi.org/10.14469/hpc/7267.

2.12.3 Peak annotation and De Novo Motif discovery

The significant peaks were processed using the HOMER Statistical software to annotate the peaks and for *De Novo* motif discovery (152). The *annotatePeaks* pipeline was used to annotated peaks in relation to important genomic regions found at the centre of the analysed peak (152). For *de novo* motif discovery the pipeline *findMotifsGenome.pl* was used to find enriched motifs in genomic positions of interest (e.g. Cap-G binding peaks) (152). The "mask" parameter was used to mask sequence repeats and the "size" parameter was used as the default 200 bp, which means the region analysed is of a fixed size rather than being dependant on the peak size.

2.12.4 Accessible chromatin and CRM analysis

For CATaDa data analysis Dam only reads were mapped to release 6.22 of the *Drosophila* genome and binned to GATC regions using a previously described pipeline (14). Subsequently, the data was normalised by the total number of reads per million (each read was divided by the total number of reads and multiplied by 1,000,000). Regions of open chromatin (Chapter 4, Section 4.2.4) were defined by calling peaks on CATaDa data using a previously described pipeline, similar to the TaDa peak-calling (Section 2.12.2). The data for known *Drosophila cis*- regulatory modules (CRM) were extracted from the

REDfly database (153). The coordinates of these genomic regions were converted to BED format for comparison to TaDa data.

2.12.5 Correlations and PCA

Deeptools was used for analysis of sequencing data described below (154). The genomic regions analysed were BED coordinate files of either the known *Drosophila* genes (r6.22) or defined genomic regions, e.g. Lola peaks (Chapter 4, Section 4.2.5). Using the *multiBigwigSummary* function multiple files were summarised into an array of average scores, based on 10 kb bins that span the entire genome. The output array was used for correlation analysis between samples, using *plotCorrelation* to run Spearman's Correlation test and visualise it in either a scatterplot or a clustered heatmap. The array was also used to plot Principal Correlation Analysis using the *plotPCA* function.

2.12.6 Plotting binding profiles

To analyse the binding profiles, the detected scores per genomic regions were collected in a matrix using *computeMatrix.* The inputs required were the binding BigWig files of interest and the genomic coordinate (BED) files. For whole genome profiles I used the "scale-region" parameter which stretched or shrunk all BED regions into a chosen fixed length (e.g. 2 kb) (Chapter 4, Section 4.2.2). For analysis of enhancer regions and open / closed chromatin regions I used the "reference-point" parameter, which refers to a fixed position in the BED coordinates (e.g. centre) (Chapter 4, Section 4.2.4). Once the appropriate matrix was produced, it was used as the input file for the *plotProfile* or the *plotHeatmap* tools necessary to plot the scores with the associated genomic regions, generating binding profiles and heatmaps respectively.

2.12.7 Peak intersection analysis

Another genomic analysis toolset suite used was Bedtools (155). BED files of interested were sorted using the *Bedtools sort.* Overlap analysis of peak-files from different datasets was done using *Bedtools Intersect* and statistical significance determined by Fisher's test using *Bedtools Fisher*.

2.12.8 Gene Ontology Analysis

Peak calling algorithms also generated a list of unique genes bound by the protein of interest. For functional profiling of these genes I performed over-representation analysis for Gene Ontology (GO) terms using the R package *clusterProfiler* (156). The enriched GO terms subontology analysed was "Biological Processes" (BP), and terms with a p-value < 0.01 and a q-value < 0.05 were defined as significant. GO Terms of Level > 5 were selected using the *dropGO* function. Barplots and dotplots of the most significant GO terms were generated using the *clusterProfiler* functions integrated with the ggplot2 package (156).

2.13 RNA extraction

To investigate transcriptional changes in the *Cap-G KDs* the tissue was processed to extract sufficient RNA needed to prepare sequencing libraries. RNA extraction was performed on three biological replicates, for *Cap-G KD* and *controls* respectively (Section 2.2). Total RNA was extracted from dissected 3rd instar larvae CNS, with 35 brains per replicate, using a standard TRI Reagent® (Sigma) extraction protocol. The larval brains were dissected in PBS in batches of 10, placed in 175 µl ice-cold TRI reagent® and then flash-frozen in dry-ice. Tissue was homogenised in TRI reagent® and samples for a single replicate were pooled together. Samples were covered in TRI reagent® to a total volume of 800 µl/ tube and incubated at RT for 5 mins. To each replicate tube 160 µl of Phenol-Chloroform (Sigma) was added, shaking vigorously, and then incubated at RT for 3 mins.

To allow phase separation to occur the tubes were centrifuged at 4°C at 11,000 g for 15 mins. The upper, aqueous phase containing the RNA was transferred to a fresh tube. The RNA was precipitated by adding 1 µl of co-precipitant GlycoBlue™ and 400 µl of isopropanol (Sigma) and briefly mixing. The samples were incubated for 1 hr at – 20°C then subsequently centrifuged at 4°C at 11,000 g for 10 mins. After centrifugation the supernatant was removed and the pellet was washed with 1 ml 75% ethanol (diluted in RNAse-free ddH_2O) for a final centrifugation step at 4°C at 6,000 g for 5 mins. After carefully removing the supernatant the pellet was left to air-dry (not completely) for 5 mins before re-suspending the RNA in RNAse-free ddH_2O. Library preparation and sequencing was performed by Beijing Genomics Institute (BGI).

2.14 Analysis of RNA-seq experiments

All raw sequence files and processed files have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (accession number GSE142112). The gene lists resulting from the differential expression analysis (Section 2.14.3) and the GO analysis results are available in the repository https://doi.org/10.14469/hpc/7267.

2.14.1 Gene mapping and summarisation

The RNA sequencing data was mapped to the *Drosophila* genome (release 6.22) from the annotation database "Drosophila_melanogaster.BDGP6.dna.toplevel.fa" using STAR (57,157), resulting in individual, indexed BAM files per replicate. The RNAseq data was visually inspected in Integrative Genome Browser (IGV). Using Deeptools, the BAM files were summarised into a unique array of read coverages using *multiBAMsummary* (154). This array was then used to visualise between and within sample group differences using the *plotPCA* or the *plotCorrelation* functions. The mapped data was then processed in R (version 3.5.2) using the Rsubread package, to count the number of reads detected per

gene (158). Using the function *featureCounts* the number of reads/genes were counted for all individual replicates and collected in a single matrix (158). The generated count matrix was further processed for quality control and differential expression analysis.

2.14.2 Quality control

The count matrices and the sample information (control or experimental) were imported into R (version 3.5.2) and assembled into *DESeqDataSetFromMatrix* objects for processing using the DESeq2 package (159). The imported data was filtered to exclude any genes that had low counts across all samples (less than 10), as these genes were not expressed in the tissue. The gene counts for each sample were normalised using the *estimateSizeFactors* function to allow for comparison of expression levels between samples. The normalised counts were visualised into MA plots using the *plotMA* function (159) .

2.14.3 Differential expression analysis

Differential expression analysis was performed using standard procedure implemented in DESeq2 (159). Briefly, the *DESeq2* function pipeline runs: estimate size factors, estimate dispersions, fitting generalised linear models based on negative binomial distribution and finally compute Wald statistics. Genes that had an adjusted p-value (FDR) < 0.05 and a $\log_2$ fold change greater than 1 (for upregulated) or less than -1 (for downregulated) were classified as significant. This generated a list of differentially expressed genes (DEG), used for further functional analysis using the *clusterProfiler* package (156) (Section 2.12.8). To visualise the most significant DEG, volcano plots were generated using EnhancedVolcano R package (160) and heatmaps were generated using pheatmap package (161).

2.14.4 Tissue-specificity profiling from FlyAtlas 2

Tissue expression data from FlyAtlas 2 was used to determine tissue of origin of upregulated genes (162). The following analysis was carried out using Pandas in Python (163). The code generated for scraping the FlyAtlas 2 database and analysing the files is available at https://github.com/amh111/FlyAtlas2Scraper. Tissue expression data from the FlyAtlas 2 online database was extracted for each individual gene of interest and collected in a single array. The "Larval FPKM" values were used as a filtering threshold for expression, where any genes displaying a value above 2 were considered as expressed in that specific tissue. All genes that were expressed (Larval FPKM > 2) in the "Brain/CNS" were removed from the list, as I was interested in only identifying the genes that were ectopically expressed in the brain. The remaining genes were filtered to exclude any non –expressed genes (Larval FPKM < 2) and the rest was summarised to display the total number of genes expressed in each tissue.

Chapter 3 Cap-G is essential in post-mitotic neurons for healthy nervous system development

3.1 Introduction

3.1.1 Neuronal development

Drosophila melanogaster has a relatively simple nervous system, making it an attractive model to study neurodevelopment. The two main developmental stages that are the focus of this study are the embryonic ventral nerve cord (VNC) and the 3rd instar larval brain (Figure 3.1). In both stages, the cells that populate the nervous system include the neuronal stem cells (NSCs) which will asymmetrically divide to self-renew and form an intermediate ganglion mother cell (GMC), which will divide once to give rise to neurons (18) (Figure 3.1). The neurons deriving from the GMCs are differentiating, post-mitotic cells as they specialise and do not have any ability to divide. In the embryonic VNC these cells separate into clear layers, making them ideal for visualisation and quantification (Figure 3.1A). The layers help identify the younger, newly born neurons by their proximity to the dividing cells.

At the larval stage, the brain anatomy includes optic lobes, which host the Optic Proliferation Centres, a hub of highly mitotic cells, the central brain, where the majority of NSCs and neurons lie, and finally the ventral nerve cord where axonal bundles gather to innervate the larval body with neurons (15,18) (Figure 3.1B). This study will focus on the NSCs and neurons, from immature newly born to fully synaptic, mature neurons. At each cell stage specific markers are expressed that help identify the cells, for example NSCs express *deadpan (dpn)* whilst newly born neurons express factors such as *elav* and

lola-N. Elav and Lola-N localise in neuronal nuclei and are normally used as neuronal markers (19).

3.1.2 Condensin I cell localisation and expression

Condensin is a multi-protein complex necessary for appropriate chromosome condensation and segregation during mitosis (53,60). It is evolutionary conserved across species, largely retaining its function and conformation (60). This chapter focuses on the condensin I complex of *Drosophila*, formed of the ATPases SMC2/4, the kleisin subunit Barren and the HEAT-repeat subunits Cap-G and Cap-D2 (Chapter 1, Figure 1.3). The condensin complex subunits are actively expressed in mitotic and interphase cells, though the localisation of each subunit differs depending on the cell stage (53,60). In *Drosophila* interphase cells the condensin I subunits exhibit different distribution to mitotic cells. During mitosis the complex subunits assemble in the nucleus ready to undertake the DNA-looping mechanism to drive chromosome condensation (72-74). In interphase, Barren localises to the cytoplasm whilst Cap-G appears to be nuclear (66,164).

Published RNA-seq data indicates significant levels of condensin subunit transcripts in neurons, including *Cap-G*, mostly at the embryonic and larval stages (10,162). Moreover, there is evidence of condensin I in murine NSCs, but not in neurons (80). The subunit Cap-G was detected in a Y2H screen to identify interactors of Lola-N, a neuron-specific transcription factor (19) (Southall, T.D., unpublished). This Cap-G/ Lola-N interaction exposed the potential for a condensin subunit to have a post mitotic role in neurons. These data gave us confidence in further exploring the role of Cap-G in *Drosophila* neurons.

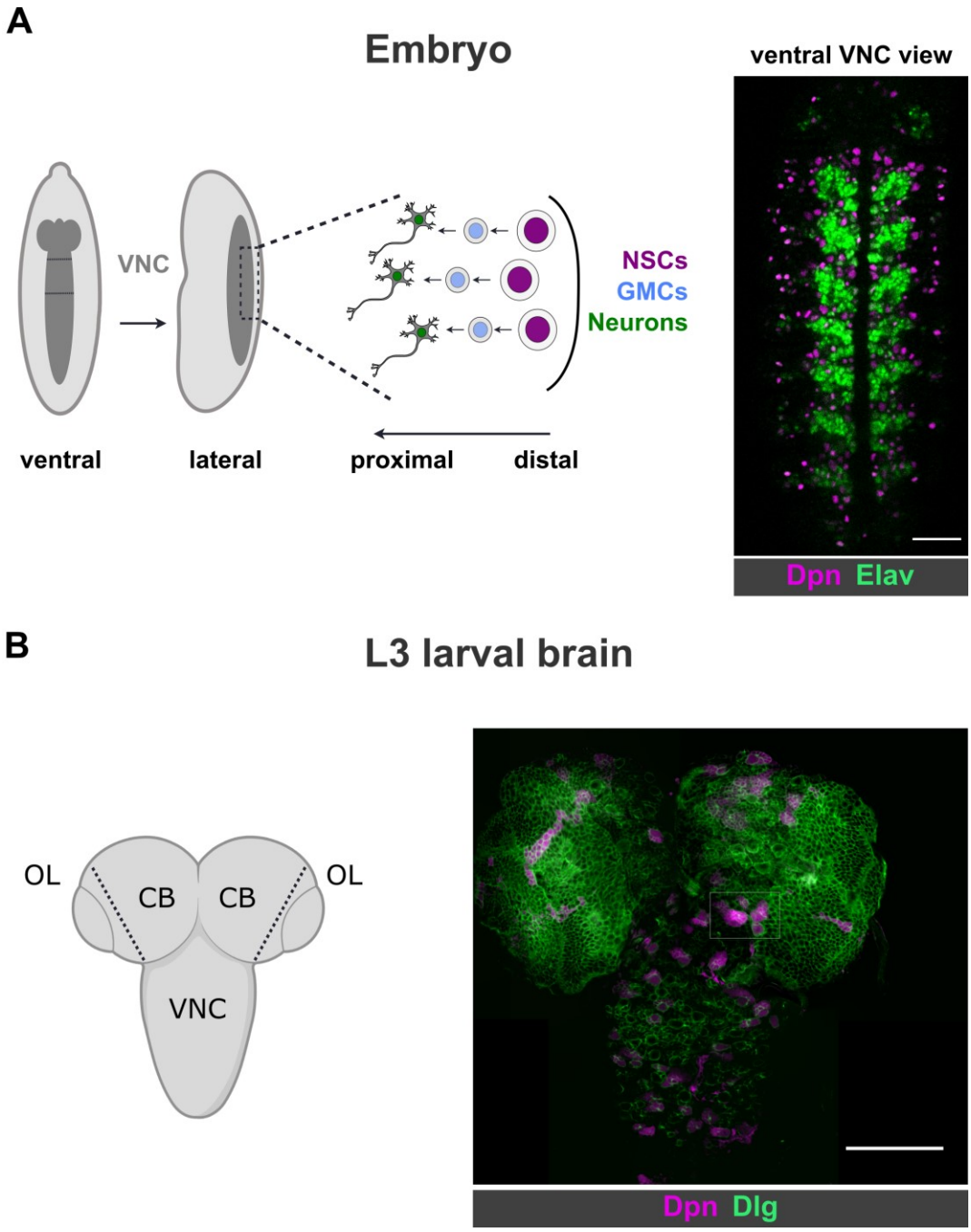


Figure 3.1. *Drosophila* embryo and larval brain anatomy.

A) Embryo diagram showing Ventral nerve cord and layers distal to proximal of NSCs, GMCs and neurons. Fluorescence staining of *wild type* embryonic VNC marking neurons (Elav) and NSCs (Dpn). Scale bar 30 µm. **B)** Third instar (L3) larval brain diagram showing Optic Lobes (OL), Central brain (CB) and VNC. Fluorescence staining of L3 brain showing cell membranes (Dlg) and NSCs (Dpn). Scale bar 100 µm.

Condensin is characterised in mitotic and interphase cells, but there is limited knowledge of its presence or potential role in post-mitotic cells. To address this question, we exploit the genetic tractability of *Drosophila* to observe Cap-G localisation *in vivo,* as well as create neuron-specific Cap-G knockdown (KD) animals. Instrumental to this study is the CRISPR knock-in line *Cap-G*^{EGFP}, donated by collaborator Stefan Heidmann. This line expresses C-terminally EGFP-tagged Cap-G from its endogenous locus, is homozygous viable and fertile and display no ill effects. Moreover, the line is able to recapitulate native Cap-G expression and localisation (138).

3.1.3 The deGradFP system to knockdown GFP-tagged proteins

To investigate a putative neuronal role for Cap-G, I used the deGradFP system to create neuron-specific Cap-G KD animals (137) (Figure 3.2). The deGradFP system is a technique that allows degradation of a GFP-tagged protein of interest. This technique exploits the Ubiquitin Proteasome Pathway of protein degradation. A series of enzymes (E1, E2 and E3) attach ubiquitin molecules to proteins being targeted for degradation (165). The ubiquitin-tagged proteins are then degraded by proteasomes and lysosomes. The E3 family of enzymes determines the specificity of the degradation pathway. An E3 enzyme is the multi-protein complex SKP1-CUL1-F-box protein ligase (SCF). This protein ligase is formed of common subunits SKP1 and CUL1, whilst the F-Box Protein (FBP) subunit changes depending on the substrate, hence determining the enzyme specificity. In *Drosophila*, the Slmb is an F-box protein that recruits the SKP1 and CUL1 factors to mediate ubiquitin transfer onto target proteins (166,167). To hijack the pathway specificity, Caussinus et al. modified Slmb to engineer the new FBP NSlmb-vhhGFP4, where they fused a single-domain GFP antibody fragment to the N-terminal of Slmb (137). The engineered FBP mediates the ubiquitin transfer on GFP-tagged proteins

resulting in their degradation by the proteasome. The authors coined the technical name for this method deGradFP (degrade Green Fluorescent Protein).

The strength of the deGradFP system is that it can be coupled with the *GAL4/UAS* system to allow for cell-specific and timed expression of the NSlmb-vhhGFP4 protein (Figure 3.3). The *GAL4/UAS* system is a powerful tool developed in *Drosophila* to investigate gene expression (55). This method exploits the yeast protein GAL4, a transcriptional activator which recognises and binds specifically to UAS (upstream activator sequence). The system involves two key parts, the driver lines (or GAL4-lines) which are genetically modified GAL4 proteins coupled with tissue or cell -specific genes, and the reporter lines (or UAS-lines) which have the *UAS* sequence upstream a target gene of interest. When the driver and reporter fly lines are crossed, the GAL4 will bind to the *UAS* and activate the transcription of the target gene, only in the tissue of interest. There is a wealth of *GAL4* and reporter lines available in the *Drosophila* community, making it a versatile model for investigating genes of interest in specific tissues.

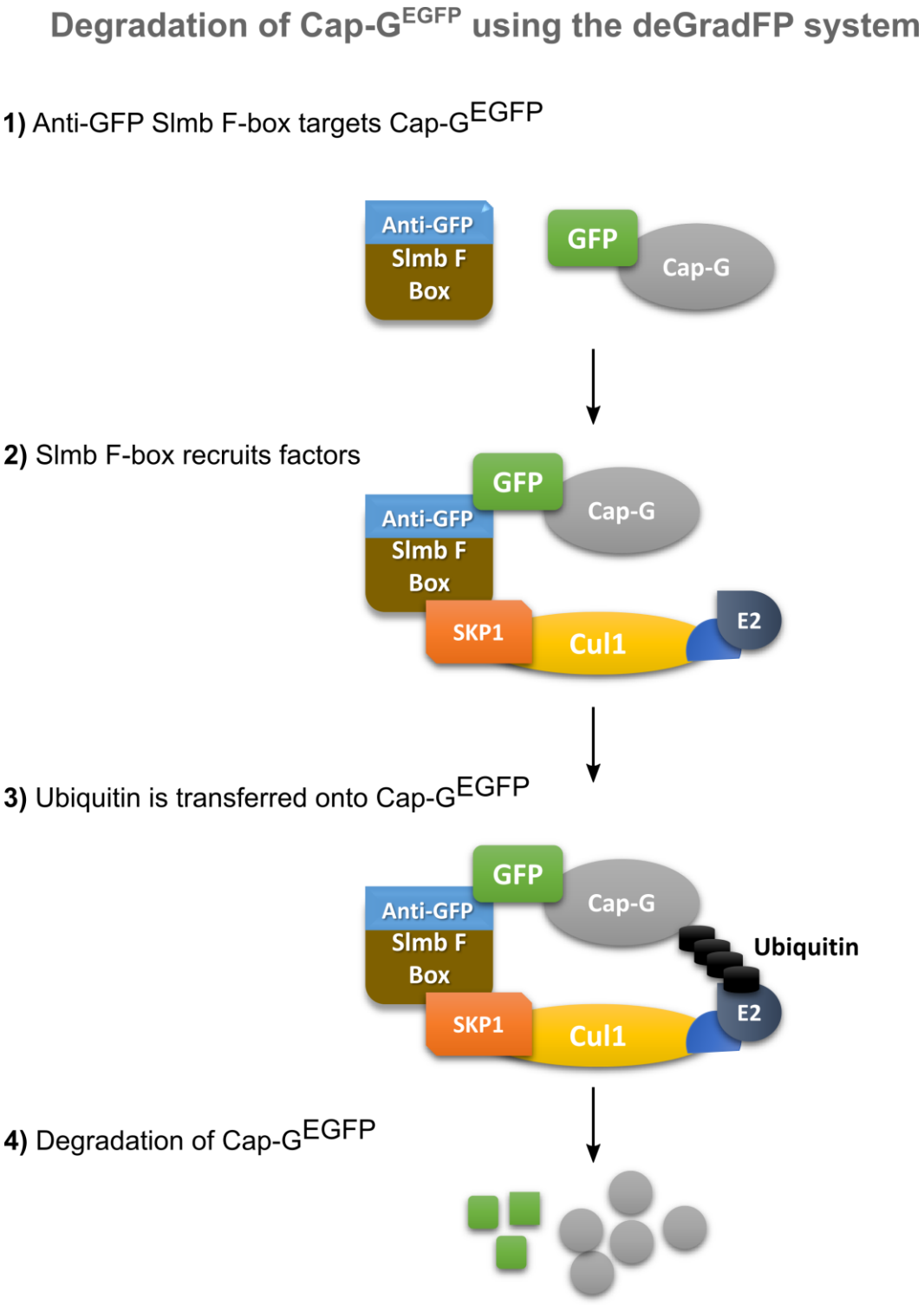


Figure 3.2 Schematic of the DegradFP system.

1) Slmb F box engineered with anti-GFP antibody will target Cap-G^{EGFP} in the cells. **2)** SKP1, Cul1 and E2 enzyme are recruited by the Slmb protein **3)** Cap-G^{EGFP} is tagged with ubiquitin and targeted for degradation **(4)**.

3.1.4 Cell-specific Cap-G knockdown with the deGradFP system

The deGradFP system has been combined with the *GAL4/UAS* system, generating a reporter line *UAS- NSlmb-vhhGFP4*, referred to as *UAS-deGradFP* hereafter (137). The *UAS-deGradFP* when crossed to a *GAL4* driver of interest allows for specific knockdown of target proteins in a cell-type of interest. In this chapter I aimed to drive cell-specific knockdown of Cap-G^{EGFP} in neurons using the deGradFP system. Therefore, the *GAL4* lines and the *UAS-deGradFP* reporter were modified to be in a homozygous *Cap-G^{EGFP}* background, so that all Cap-G protein present in the organism is GFP-tagged and can be targeted by the deGradFP system (Figure 3.3A). Two neuronal, post-mitotic drivers were used: *elav-GAL4* which is expressed in all neurons, from newly born to mature, and *nSyb-GAL4* which drives expression solely in more mature neurons which have begun synapse formation (Figure 3.3B). Cap-G knockdown experiments with these two drivers are referred to as *elav-KD* and *nSyb-KD* respectively hereafter.

Limitations of the deGradFP system include the fact that it infers a knockdown of the protein, which may not result in a null phenotype, as some non-degraded protein may still remain in the tissue. Moreover, the deGradFP effectiveness and specificity is limited by the driver lines available and used for the experiments. If the *GAL4* lines used have low or non-specific tissue expression it could affect the resulting knockdown. Despite these potential limitations, the deGradFP system has been used effectively to degrade protein of interest in specific cell types, such as spermatocytes and neurons (168,169). The deGradFP system was used to successfully degrade GFP-tagged proteins in the *Drosophila* CNS, and in some instances the deGradFP KDs recapitulated known null mutant phenotypes (169). Moreover, the deGradFP system was able to target maternally inherited GFP-tagged protein in both neurons and spermatocytes (168,169). These

reports provide a solid foundation for using the deGradFP system as a tool to generate tissue-specific knockdowns, especially in the CNS, which is the focus of this study.

3.1.5 Alternative methods to generate cell-specific Cap-G KD/mutant

A mutant *Cap-G* line has been previously described, which is an EP-element insertion and has an embryonic lethal phenotype (81). This line was not suitable for this study, as it cannot be manipulated to be cell-specific. Therefore, alternative methods were tested to try and generate cell-specific *Cap-G* knockout or Cap-G knockdown animals. Extensive measures were taken to try and generate cell-specific *Cap-G* knockout or knockdowns using multiple methods, however, none of the methods described below resulted in a faithful *Cap-G* mutant phenotype of embryonic lethality, making them unsuitable for this study (Table 3.1).

In *Drosophila*, generation of cell-specific CRISPR/Cas9 mutants has been described (170). The simultaneous expression of multiple gRNAs was coupled with the *GAL4/UAS* system to achieve multiplexed, cell-specific genome editing (170). Initially, I generated transgenic flies expressing gRNAs for cell-specific targeting of *Cap-G* following the protocol described (170) (Appendix). However in positive controls, ubiquitous expression of the *Cap-G CRISPR/Cas9* mutants was unable to recapitulate *Cap-G* null mutant phenotypes, therefore I discarded this approach.

Alternative knockdown methods were attempted, such as RNA interference (RNAi), which target the nascent mRNA transcripts of interest and result in gene silencing (171). However, positive controls of the *Cap-G RNAi* line available resulted in viable phenotypes when ubiquitously expressed in the fly tissues, making it unusable for this study (Table 3.1). Finally, another method is an adaptation of RNAi, the *in vivo* GFP interference

(iGFPi), which reduces the expression of GFP-tagged genes (172). This iGFPi could be used to knockdown the expression of Cap-G^{EGFP} in a cell specific manner. This would still require the *Cap-G^{EGFP}* line and the *UAS/GAL4* system, similar to the deGradFP system. The phenotypes obtained with the deGradFP system have been shown to be comparable, and almost identical to iGFPi when the drivers are continuously expressed throughout development (169). Given the redundancy of these systems I decided to generate the cell-specific Cap-G knockdown only using the deGradFP system.

Table 3.1 Cell-specific Cap-G knockdown / knockout methods attempted.

Method	Phenotype	Source
UAS-Cap-G-RNAi (KK line)	Viable	VDRC #108841
UAS-Cap-G-gRNAs (Cell-specific CRISPR/Cas9 mutants)	Viable	I generated following published protocol (170)
deGradFP system	Lethal	(137)

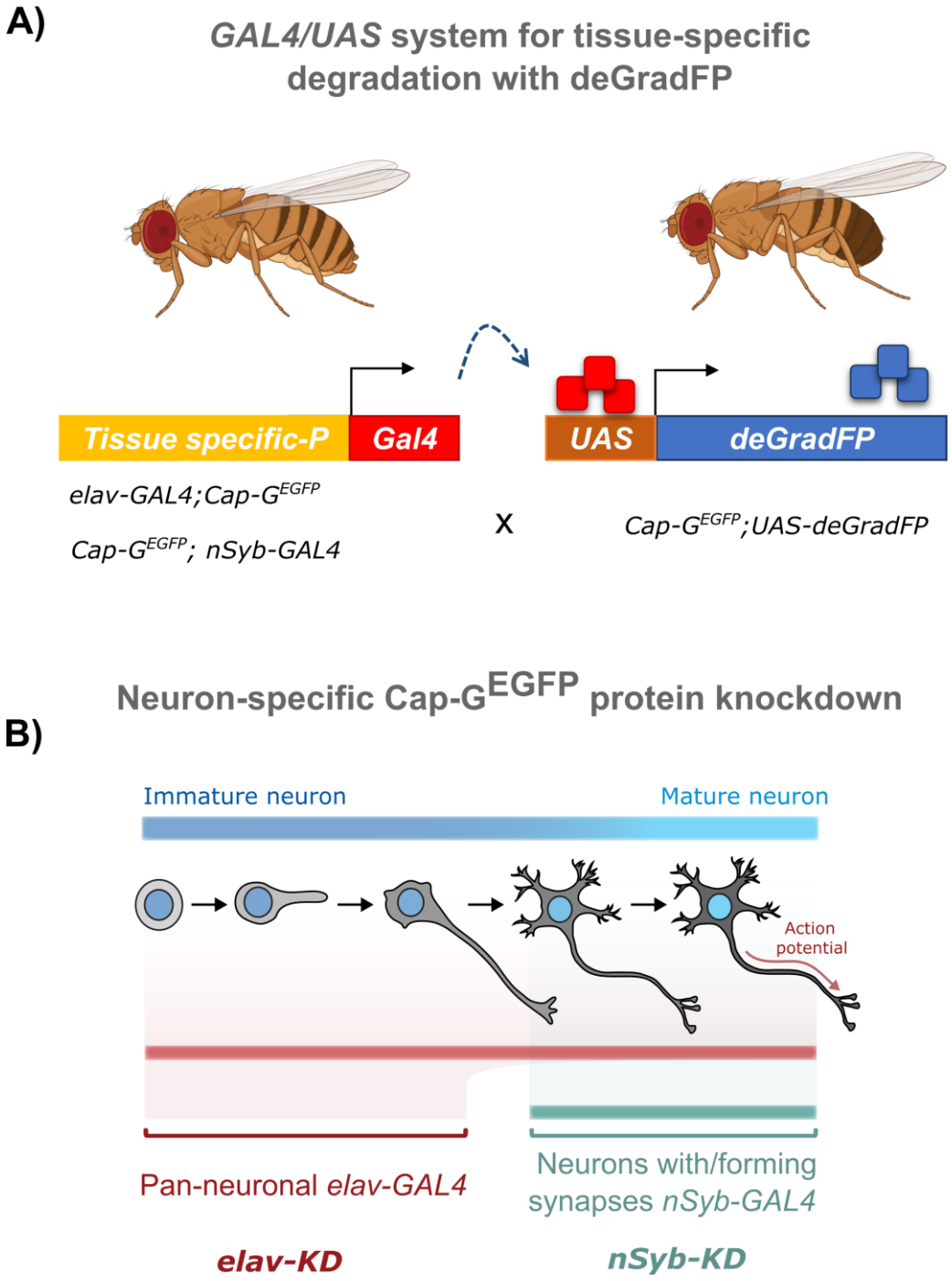


Figure 3.3 Generating neuron specific Cap-G KD using the deGradFP system.

A) The *GAL4/UAS* system where Tissue specific driver induces GAL4 expression, then the GAL4 binds to Upstream Activating Sequence (UAS) which induces expression of the deGradFP. All the lines used are in a *Cap-G*^{EGFP} homozygous genetic background. **B)** Schematic of experimental plan for Cap-G knockdown driven by *elav-Gal4* (*elav-KD*) expressed in all neurons and *nSyb-Gal4* (*nSyb-KD*) expressed in mature neurons.

3.1.6 Chapter Aims

In this chapter I aim to characterise the expression pattern and protein localisation of the condensin I subunit Cap-G in the *Drosophila* nervous system. Currently there is little understanding of condensin in post-mitotic cells and there is no evidence for the effects of condensin knockdown in post-mitotic neurons. In this chapter I aim to generate a robust and reliable post-mitotic, cell-specific knockdown to characterise Cap-G function in neurons. I present evidence for the localisation of Cap-G in neurons across embryonic and larval tissues. Moreover, I successfully generated and characterised neuron-specific Cap-G KD animals. These animals have impaired motility, reduced survival and neurons display increased levels of DNA damage. Overall in this chapter I demonstrate that Cap-G plays a crucial role in the functioning and healthy development of the fly nervous system.

3.2 Results

3.2.1 *Cap-G* is expressed in neurons

A Y2H screen in the Southall lab revealed an interaction between the TF Lola-N and the condensin I subunit Cap-G (unpublished, Southall T.D.). In the *Drosophila* brain, Lola-N is expressed in neurons, so the Cap-G/Lola-N interaction suggests that Cap-G may be also present in neurons. Whilst condensin activity has been characterised in neural stem cells (80), the role of condensin complexes in post-mitotic neurons has not yet been studied in any species. Published RNA-seq data indicates significant levels of condensin subunit transcripts in neurons, including *Cap-G* (10,162). To determine whether *Cap-G* is expressed in neurons I carried out fluorescent *in-situ* hybridisation on embryonic and larval brain tissues. The method used to detect *Cap-G* mRNA is a single-molecule fluorescent *in situ* hybridisation (smFISH), which has been shown to reliably work on whole-mount *Drosophila* larval and adult brains (145). This method also enables co-staining of tissue with antibodies, to mark different cell types, without affecting the probes.

In w^{1118} stage 15 embryos there was ubiquitous expression of *Cap-G* mRNA in all cells of the VNC (Figure 3.4A). *Cap-G* mRNA was also detected in the cytoplasm of neurons marked by the nuclear protein Elav (Figure 3.4A). This suggests active expression of *Cap-G* in embryonic neurons. Further to embryonic tissue I characterised the expression pattern of *Cap-G* in 3rd instar larval brains. *Cap-G* mRNA is ubiquitously expressed in cells of the Central Brain and Optic Lobes of larval brains (Figure 3.4B). In these larval brains, the neurons are marked by mCD8-GFP, driven by *elav-Gal4*, to fluorescently label the neuronal membrane. Sections of the central brain that highlight large cells are likely NSCs, as there are expected strong levels of *Cap-G* mRNA but lack the mCD8-GFP marker (Figure

3.4B). The neighbouring mCD8-GFP-positive neurons also display strong levels of *Cap-G* mRNA, confirming that *Cap-G* is expressed in larval brain neurons. Together these results provide *in vivo* evidence of condensin I subunit *Cap-G* being expressed in post-mitotic neurons of the *Drosophila* nervous system.

3.2.2 Subunits of the condensin I complex Cap-G and Barren localise in neurons

Upon confirming the active expression of *Cap-G* mRNA *in vivo* I wanted to characterise the distribution of Cap-G protein in *Drosophila* neurons. To visualise Cap-G in the CNS I used a *Cap-G*^{EGFP} CRISPR knock-in line, which expresses EGFP-tagged Cap-G and recapitulates native function and distribution (138). Originally, immunostaining experiments were performed with an antibody against Cap-G (66) but the antibody used did not reliably recognise Cap-G in wild-type tissues (data not shown). Therefore the *Cap-G*^{EGFP} line was the most suitable tool available for visualisation of Cap-G *in vivo.*

Firstly, to validate whether the *Cap-G*^{EGFP} line recapitulated native expression, I investigated GFP-signal in dividing cells, where condensin I should be most prevalent. In the VNC of stage 14 embryos, Cap-G^{EGFP} was strongly detected in NSCs, marked by Dpn (Figure 3.5). Larval brains also displayed strong Cap-G^{EGFP} signal in the optic proliferation centre and the NSCs (Figure 3.6). In *Drosophila* , NSCs and mitotic cells require Cap-G and the condensin complex subunits for appropriate cell division (81). Therefore, strong GFP signal detected in dividing cells increased confidence in the accuracy and suitability of the *Cap-G*^{EGFP} fly line for experiments.

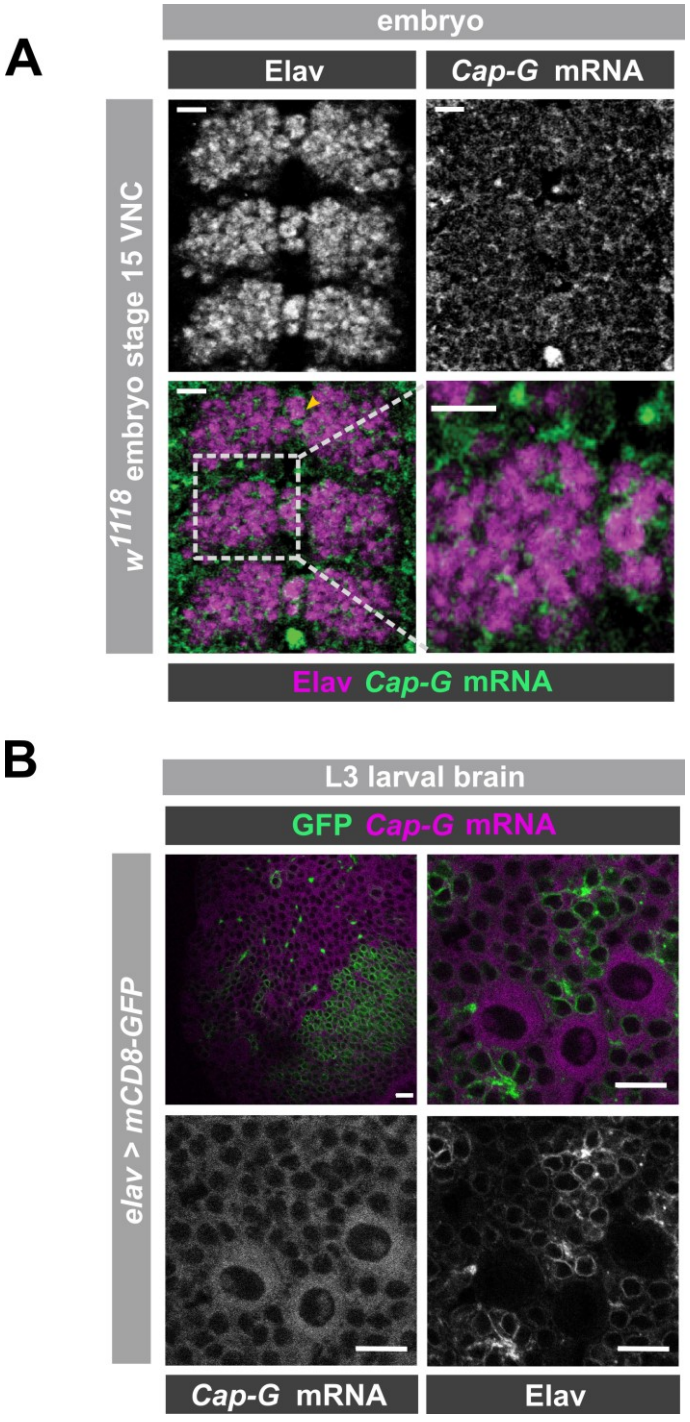


Figure 3.4 *Cap-G* expression in neurons.

A) *w1118* embryo (stage 15, anterior top). *Cap-G* mRNA is ubiquitously present in neuronal cytoplasm, neurons marked by Elav. **B)** Optic lobe of 3rd instar larvae. *Cap-G* mRNA is ubiquitously expressed in the lobe and central brain. mCD8-GFP marks neuronal membrane. Central brain section shows expression of *Cap-G* mRNA in neurons marked by elav-driven GFP expression. Scale bars 10 μm.

Since establishing *Cap-G*^{EGFP} as a fly line suitable for visualisation of native Cap-G, I aimed to investigate whether there were visually detectable levels of Cap-G in post-mitotic neurons. There is currently no reported *in vivo* data for any condensin subunit in neurons. A study on mouse cortex cells identified condensin I and II subunits in NSCs, but absent from neurons (80). Secondly, I aimed to characterise whether Cap-G had a discrete distribution in neurons, and whether it was cytoplasmic or nuclear. In interphase *Drosophila* cells Cap-G appears to localise to the nucleus, whilst the rest of the condensin subunits are localised in the cytoplasm, prior to initiation of cell division (66,76). In the embryonic VNC Cap-G^{EGFP} was localised in neurons, as it co-localised with the neuronal marker Elav (Figure 3.5). Cap-G^{EGFP} localises to neuronal nuclei, as Elav marks neuronal nuclei (Figure 3.5). Cap-G^{EGFP} also localises to neuronal nuclei in the larval central brain (Figure 3.6). These experiments provide *in vivo* evidence of condensin I subunit Cap-G localises to post-mitotic, neuronal nuclei in embryonic and larval tissues.

In embryos, neuronal differentiation occurs in the distal to ventral direction, with newly born neurons localising at the VNC edges closest to the progenitor cells (Figure 3.1A). In larval central brain, the cells closest to the NSCs are often the progenitors and the newly born neurons resulting from a mitotic and differentiation event (15). Interestingly, in both larval and embryonic tissue, the Cap-G^{EGFP} signal appears stronger in the neurons closest to NSCs. In embryos, Cap-G^{EGFP} signal is detected in newly born neurons as well as the rest of the VNC (Figure 3.5). In larval brains, most obvious Cap-G^{EGFP} signal is detected in close proximity of NSCs (Figure 3.6B)

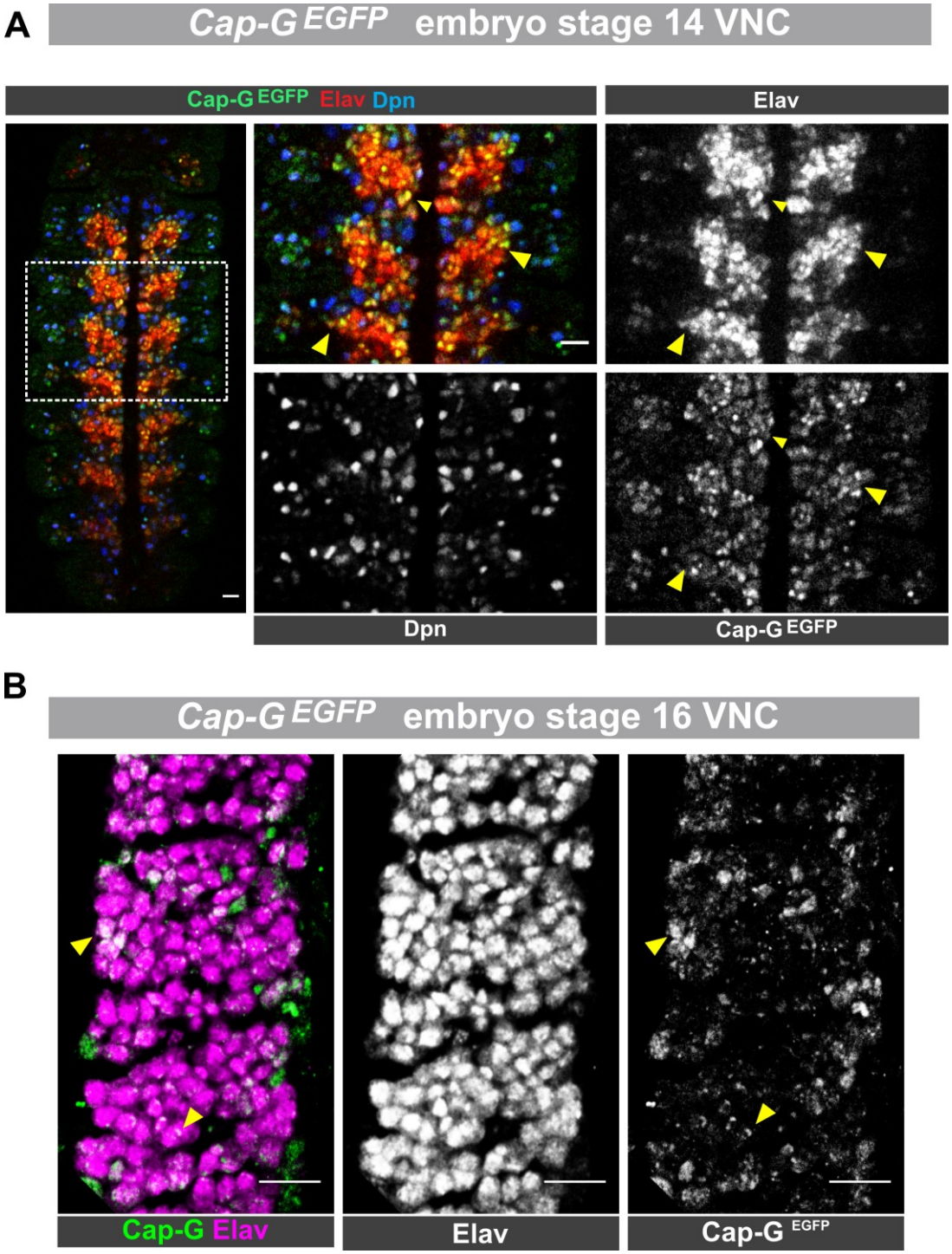


Figure 3.5 Cap-G^{EGFP} localises in neuronal nuclei in embryos.

A) *Cap-G^{EGFP}* embryonic VNC (stage 14, anterior top). Neurons marked by Elav and NSCs by Dpn. Cap-G^{EGFP} strongly co-localises with Dpn. Highlighted VNC segments (dotted box) show Cap-G^{EGFP} co-localise with Elav (neurons) marked by yellow arrowheads, in proximity of NSCs. **B)** Section of *Cap-G^{EGFP}* embryonic VNC (stage 16, anterior top). Cap-G^{EGFP} signal is widespread in neuronal nuclei at the distal edges of the VNC (yellow arrowheads). Neurons in the centre of the VNC show distinct foci distribution of Cap-G^{EGFP} (yellow arrowheads). Scale bars 10 μm.

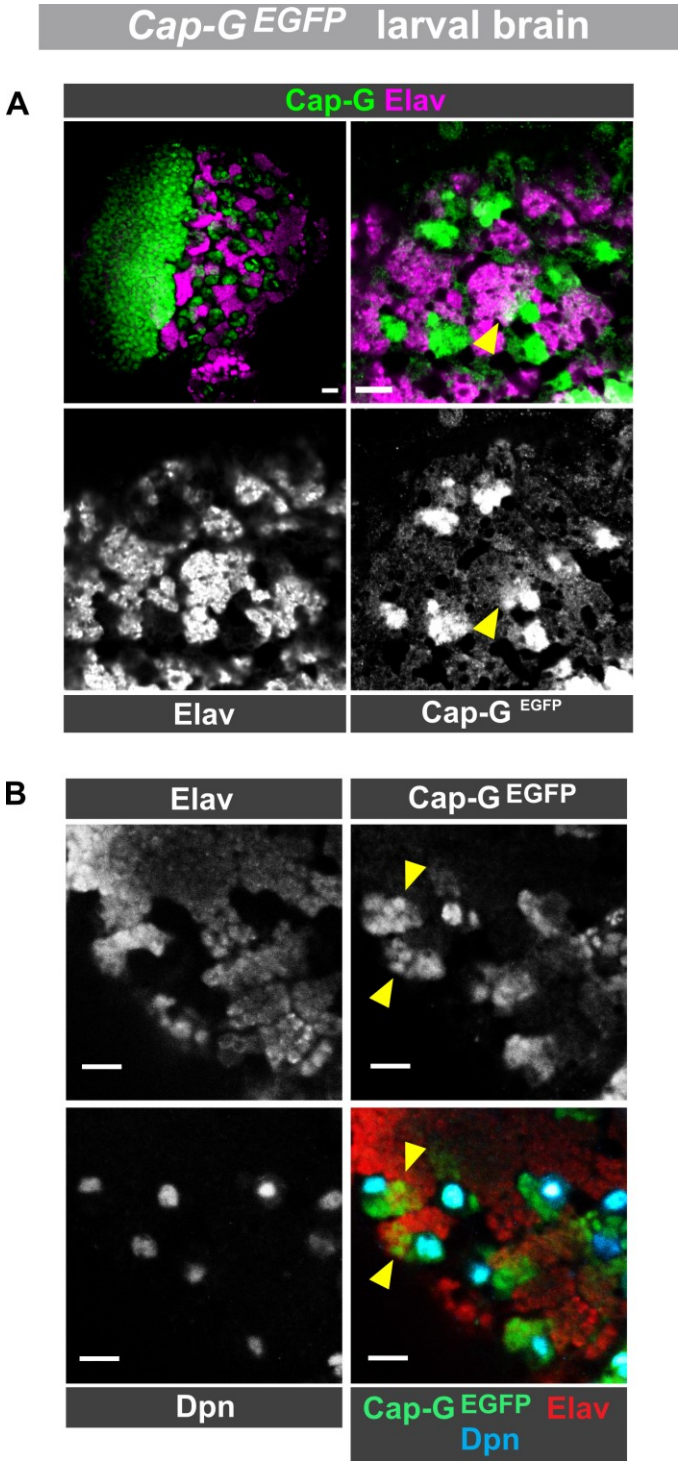


Figure 3.6 Cap-G^{EGFP} localises in neuronal nuclei in larval brains.

Optic lobe of 3rd instar *Cap-G^{EGFP}* larvae. **A)** Cap-G^{EGFP} is strongly present in the optic proliferation centre and the central brain (CB). CB sections show Cap-G^{EGFP} is present in neuronal nuclei as it overlaps with marker Elav (yellow arrowheads). **B)** Cap-G^{EGFP} localises in neurons closest to dividing cells (yellow arrowheads). Scale bars 10 µm.

In addition to Cap-G, a second condensin I subunit, Barren, was investigated in the context of neuronal localisation. The fly line *barren*^{EGFP} is another *CRISPR knock-in* line which expresses EGFP-tagged Barren and recapitulates native function and distribution (138). I aimed to investigate the localisation of a second condensin subunit to understand whether localisation to neurons was exclusive to Cap-G, or possibly involved the whole condensin I complex.

The Barren^{EGFP} protein was detected in cells all across the embryonic VNC (Figure 3.7A). Surprisingly, Barren localisation appears to be predominantly in the cytoplasm of neurons (Figure 3.7A), unlike Cap-G which is nuclear. In larval brains Barren^{EGFP} strongly localised to cells in the optic proliferation centre as well as the Dpn positive cells in the central brain (Figure 3.7B). Moreover, Barren^{EGFP} co-localises with neuronal marker Elav (Figure 3.7B). Barren^{EGFP} localisation also appears to be cytoplasmic in both neurons and NSC in larval neurons. However, in the larval brain the localisation of Barren^{EGFP} is widespread and the high intensity signal makes it difficult to distinctly characterise the localisation of Barren, which could also be in the nucleus (Figure 3.7B). Overall these results indicate that the condensin subunits Cap-G and Barren localise to neurons across *Drosophila* developmental stage. Since Cap-G showed the most distinct nuclear localisation, we decided to focus on the neuronal role of Cap-G for the rest of this study.

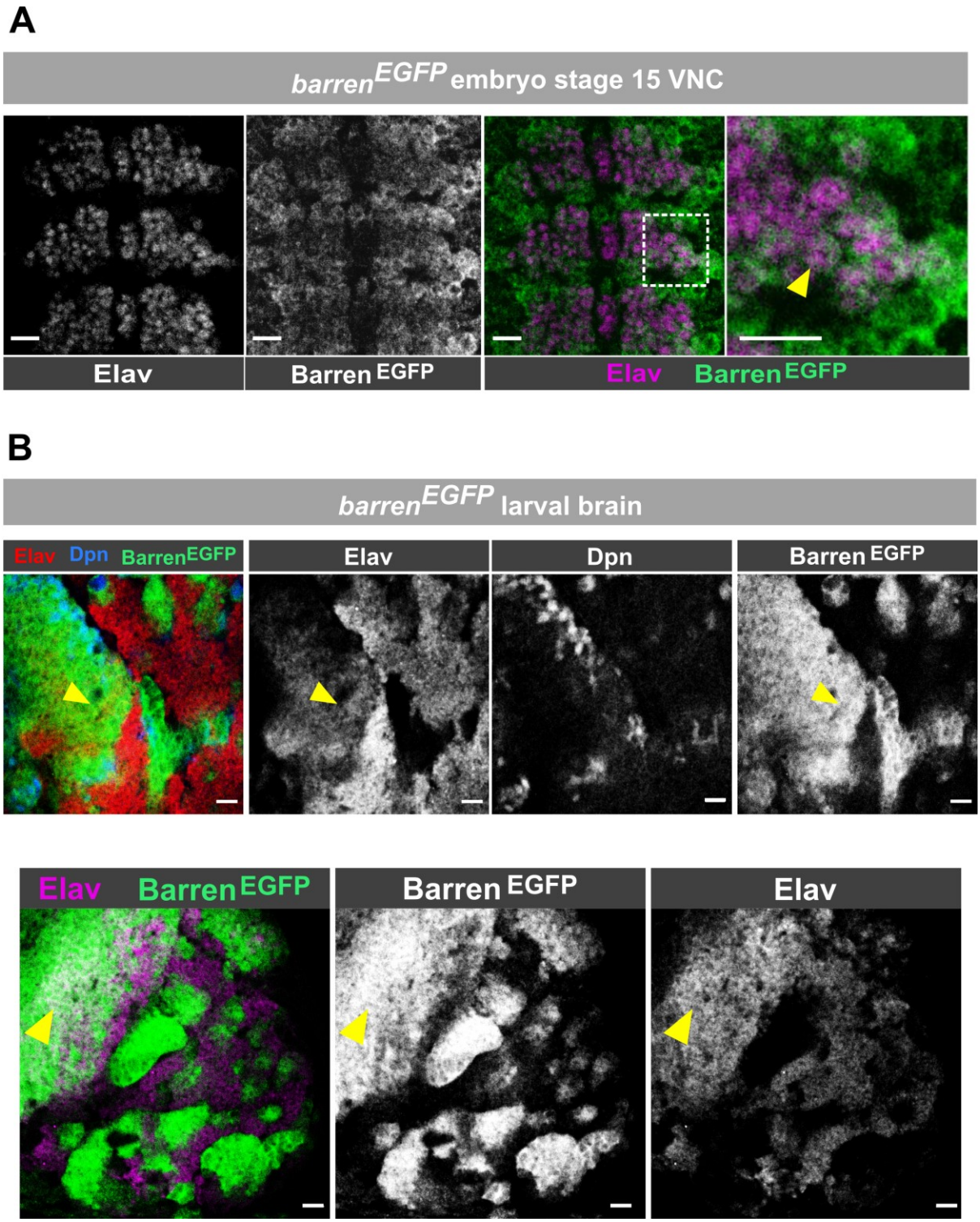


Figure 3.7 Barren^{EGFP} localises in neurons

A) *barren^{EGFP}* embryo (stage 15). Barren^{EGFP} is present in all cells of the VNC, including Elav positive neurons. All images show ventral view of embryonic VNC (anterior -top). **B)** *barren^{EGFP}* larval brain. Barren^{EGFP} detected in the optic lobes and colocalsies with neurons marked by Elav (yellow arrowhead). CB section show Barren^{EGFP} colocalises with NSCs (Dpn) and neurons close to the dividing cells. Scale bars 10 μm.

3.2.3 Cap-G co-localises with Lola-N in neuronal nuclei

Upon validating Cap-G localisation in neurons, I aimed to see whether Cap-G and Lola-N colocalised *in vivo* in *Drosophila* embryos. Lola-N is a neuron-specific TF that is present in neuronal nuclei (19) and an interaction was detected between Cap-G and Lola-N in a Y2H screen (unpublished, Southall T.D). To further understand the interaction between Lola-N and Cap-G I carried out experiments to observe potential co-localisation *in vivo.* Cap-G and Lola-N co-localise in neuronal nuclei across embryonic stages (Figure 3.8). Stage 14 embryos show widespread distribution of Cap-G in neuronal nuclei co-localising with Lola-N (Figure 3.8A). Previously I observed prominent distribution of Cap-G^{EGFP} in neurons nearest to the NSCs and progenitors, suggesting that Cap-G is present in newly differentiated neurons (Figure 3.5). Given Lola-N is a TF present in neurons from their birth onwards (19), co-localisation of Cap-G with Lola-N further confirms that Cap-G is present in newly born neurons.

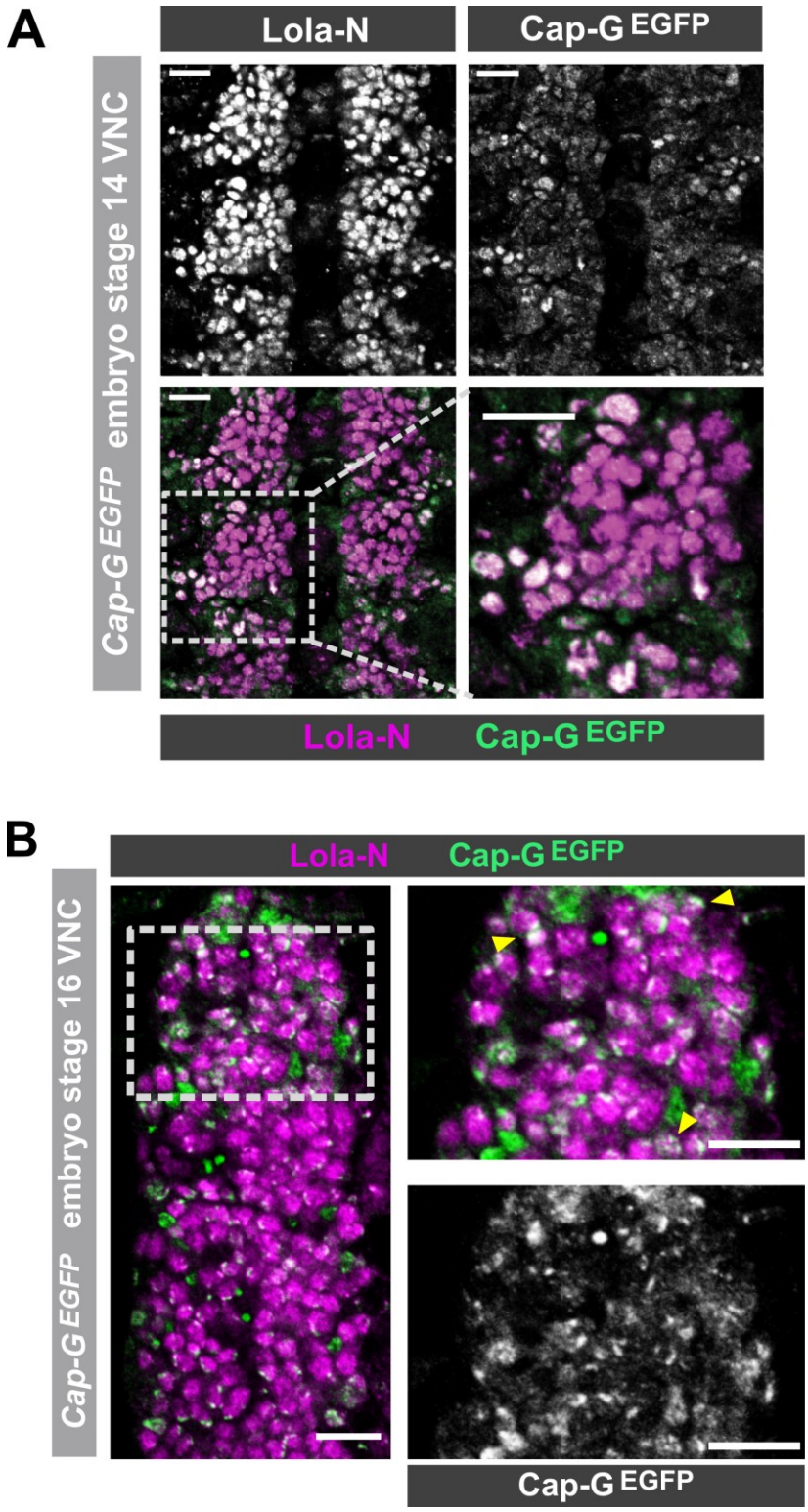


Figure 3.8 Cap-G^{EGFP} colocalises with Lola-N in neurons.

A) *Cap-G^{EGFP}* embryo (stage 14, anterior top). Cap-G^{EGFP} co-localises with Lola-N in neuronal nuclei of the VNC. **B)** *Cap-G^{EGFP}* embryo (stage 16, anterior top) Cap-G^{EGFP} co-localises with Lola-N in distinct foci in neuronal nuclei marked by yellow arrowheads. Scale bars 10 µm.

Interestingly, the distribution of Cap-G within the nucleus seems to differ across embryonic stages. At stage 16, Cap-G distribution in neurons marked with Lola-N is more restricted in the nucleus to a few discrete foci (Figure 3.8B). This is distinct to the broader distribution of Cap-G in the nucleus at the earlier stage (Figure 3.8A). Given the role of the condensin complex in chromatin compaction, these imaging results could suggest that later in development Cap-G localises at heterochromatin sites in nuclei, which could explain the foci distribution observed. Previous reports demonstrate that *Drosophila* Cap-G co-localises with heterochromatin protein 1 (HP1) in interphase cells (66). From this evidence I hypothesised that Cap-G could co-localise with HP1 in neurons. Therefore, to investigate the nature of the Cap-G foci distribution in neurons I co-stained *Cap-G*^{EGFP} embryos with an HP1 marker. In embryonic VNC, Cap-G^{EGFP} partially co-localises with HP1 in neurons (Figure 3.9). Cap-G localises in distinct foci in neurons, however, when compared to HP1 staining those foci only partially overlap (Figure 3.9B). Moreover, in larval brains Cap-G broadly co-localises with HP1 in neurons, though Cap-G signal in neurons did not display distinct foci as in embryos, but is more widespread across the nucleus (Figure 3.9C). These results suggest that Cap-G distribution partially resembles heterochromatin in neuronal nuclei across development.

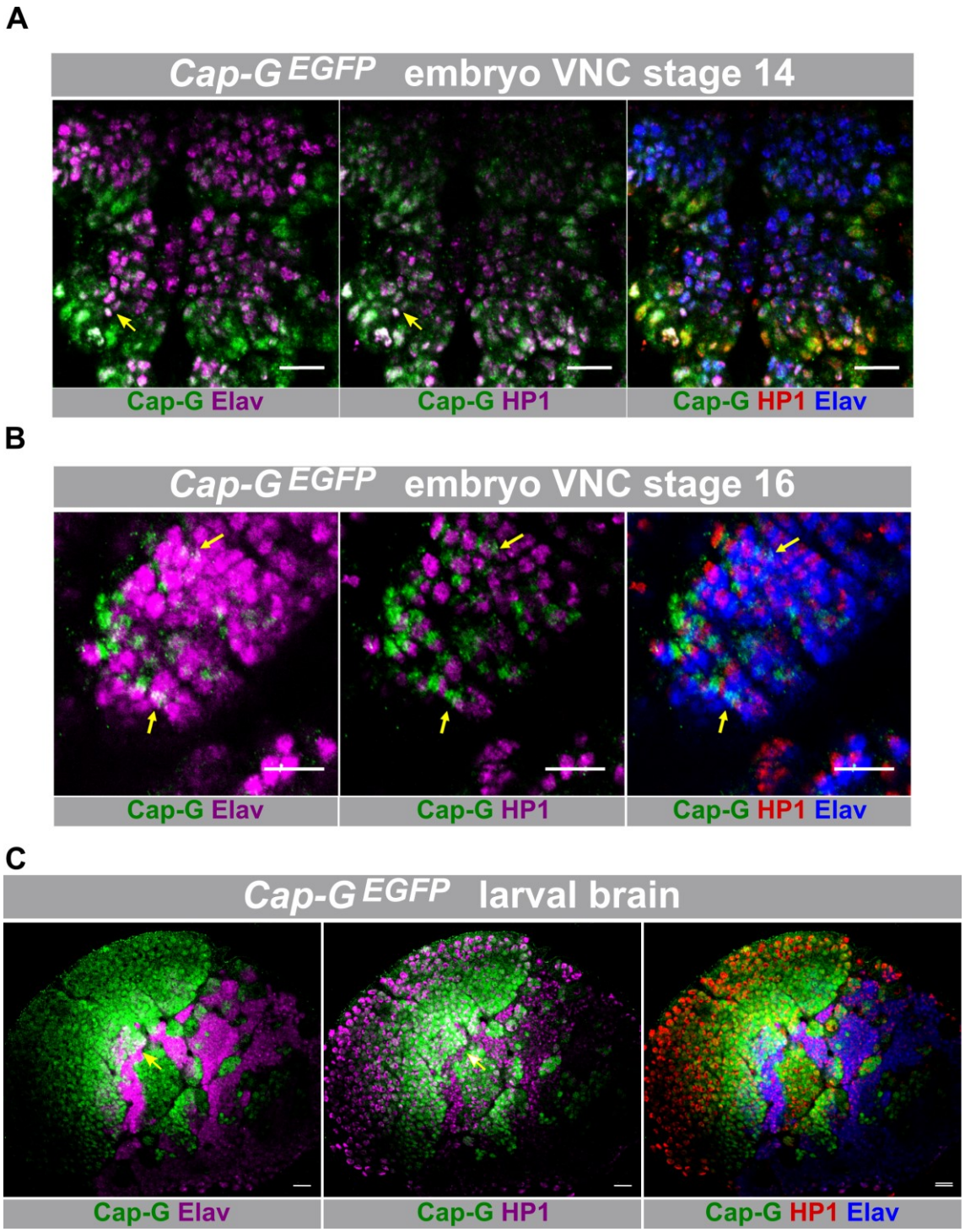


Figure 3.9 Cap-G^{EGFP} partially colocalises with HP1 in neurons.

A) *Cap-G^{EGFP}* embryo (stage 14) VNC segment. Cap-G^{EGFP} distribution is widespread in the neuronal nuclei (Elav) and partially colocalises with HP1 (yellow arrows). **B)** *Cap-G^{EGFP}* embryo (stage 16) VNC segment. Partial co-localisation between Cap-G^{EGFP} and HP1 in neurons (Elav) and foci are highlighted by yellow arrows. **C)** *Cap-G^{EGFP}* larval optic lobe. Cap-G^{EGFP} signal is widespread in nuclei of larval brain cells and has no distinct foci. Partial co-localisation of Cap-G^{EGFP} and HP1 in neurons (yellow arrow). Scale bars 10 µm.

3.2.4 Cell-specific knockdown of Cap-G with deGradFP

The presence of Cap-G in neurons suggests a post-mitotic function for condensin in the CNS. To investigate this putative neuronal role, I used the deGradFP system to create neuron-specific Cap-G KD animals (137). Firstly, I aimed to assess whether the deGradFP system could effectively induce degradation of the Cap-G^{EGFP} protein. Ubiquitous degradation of Cap-G^{EGFP} with *daughterless-GAL4* resulted in 100% embryonic lethality. This is consistent with previous reports of *Cap-G* mutant embryonic lethality in *Drosophila* (81,173). To further confirm and visualise the effectiveness of deGradFP as a knockdown tool I designed a positive control experiment to induce Cap-G^{EGFP} degradation in stochastic cell clones. I combined the deGradFP system with the heat-shock 'flip-out' (hs-FLP) system (142), which exploits a ubiquitous driver line intrinsic of FRT cassette to inhibit *GAL4* expression. Upon heat- shock, the *hs-FLP* line activates expression of the *flipase (Flp)*, which will remove the FRT cassette and allow *GAL4* expression. In turn, this expresses the *UAS-deGradFP* line solely in stochastic clones, marked by a fluorescent reporter protein, and induce Cap-G^{EGFP} degradation. In larval optic lobes, drastic loss of Cap-G^{EGFP} signal is detected in stochastic clones compared to the neighbouring cells (Figure 3.10A). This result further confirms that the deGradFP system is suitable for inducing neuron-specific degradation of Cap-G^{EGFP}.

Immunostaining of *elav-KD* embryos showed a reduction in GFP signal in neurons, indicating that a robust neuronal knockdown was achieved using the deGradFP method (Figure 3.10B). I performed knockdown experiments using both *GAL4* lines to better characterise the role of Cap-G during neuronal maturation, assessing its effect on survival, behavioural and genomic traits.

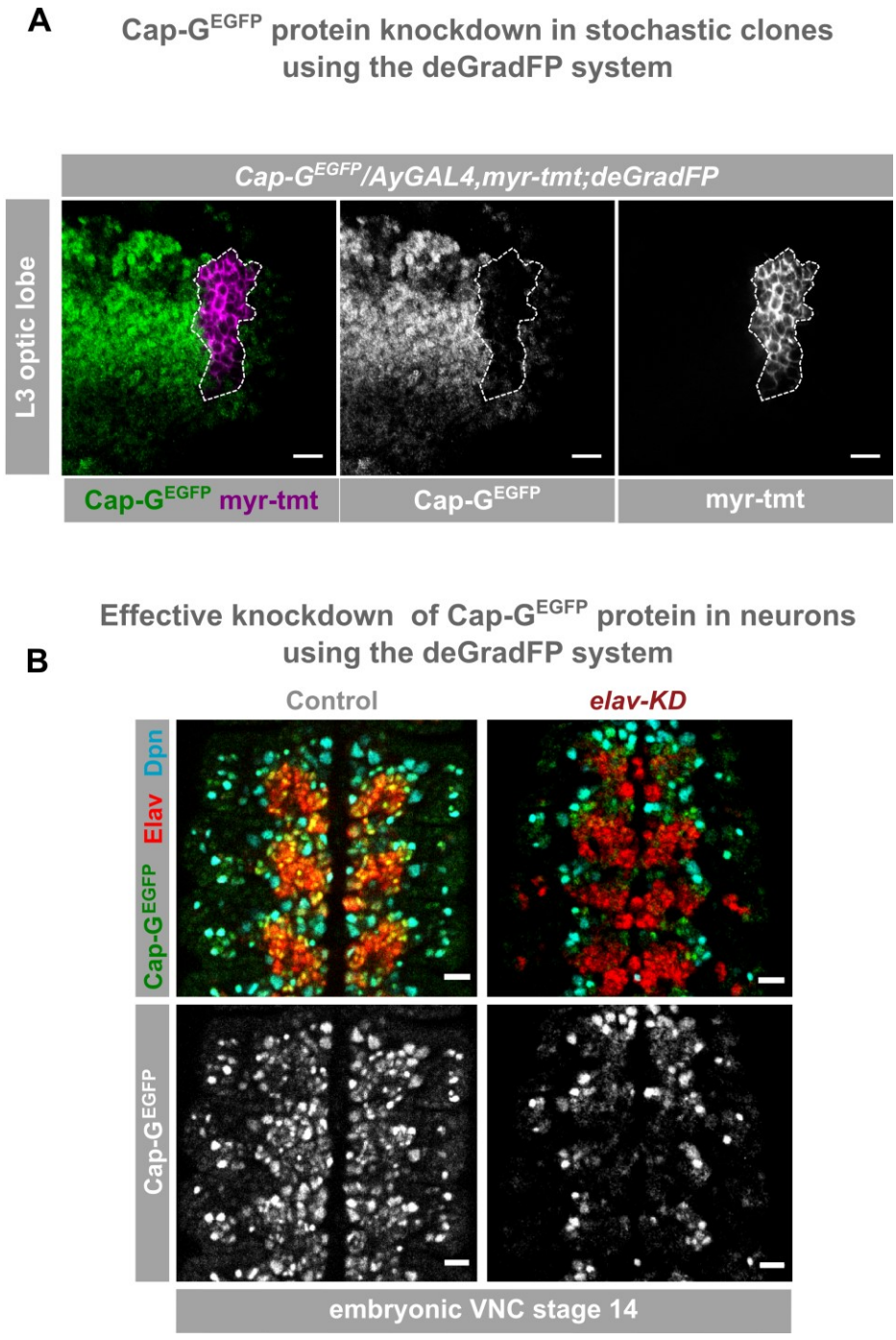


Figure 3.10. Cap-G^{EGFP} knockdown in neurons.

A) Optic lobe 3rd instar larvae. Cap-G^{EGFP} degradation driven in clones marked by myr-tmt. Effective knockdown of Cap-G^{EGFP} in clones in the optic proliferation centre. Scale bar 10 µm. **B)** Effective knockdown of Cap-G^{EGFP} in neurons. Single slice of embryonic VNC (stage 14, anterior top) showing reduced Cap-G^{EGFP} levels in neurons marked by Elav in *elav-KD* compared to *Cap-G^{EGFP}* control. Cap-G^{EGFP} is still detected in NSCs marked by Dpn in both Cap-G^{EGFP} control and elav-KD. Scale bars 10 µm.

3.2.5 Cap-G KD in neurons leads to reduced survival and behavioural defects across development

To study condensin I in a neuronal context I induced the degradation of Cap-G specifically in neurons, resulting in a cell-specific knockdown of Cap-G. Removal of Cap-G from *Drosophila* neurons led to a severe developmental mortality phenotype (Figure 3.11A). Both *elav-KD* and *nSyb-KD* animals displayed increased mortality at pupal and adult stages when compared to *Cap-G*^{EGFP} controls. In *elav-KD* animals, only 17% of embryos developed into pupae, and only 4% successfully eclosed to produce adult animals, compared to 71% and 66% reaching pupal or adult stages respectively in controls (Figure 3.11A). Embryos from *nSyb-KD* flies also displayed a severe survival defect, with only 25% pupating and 21% adults flies eclosing. Survival analysis using both the Logrank and the Gehan-Breslow-Wilcoxon tests showed a significant difference between both *Cap-G KDs* and *control* flies ($p < 0.0001$, n = 300).

Whilst the number of animals surviving to pupation was not significantly different between *elav-KD* and *nSyb-KD,* significantly more *nSyb*-KD flies survived to adult stages. The surviving 4% *elav-KD* adults all had severe mobility impairment, where most adults were unable to move and died shortly after eclosing. Despite appearing morphologically normal, the eclosed *nSyb-KD* adults displayed a significantly higher mortality rate when compared to *Cap-G*^{EGFP} controls (Figure 3.11B). On average the male *nSyb-KD* flies lived 20 days less compared to controls. Female *nSyb-KD* animals fared marginally better, surviving on average 15 days less than controls. Survival analysis using both the Logrank and the Gehan-Breslow-Wilcoxon tests showed a significant difference between the *nSyb-KD* and *control* flies ($p < 0.0001$, n = 60). Overall, these results suggest that Cap-G in post-mitotic neurons is necessary for normal development of the CNS and survival.

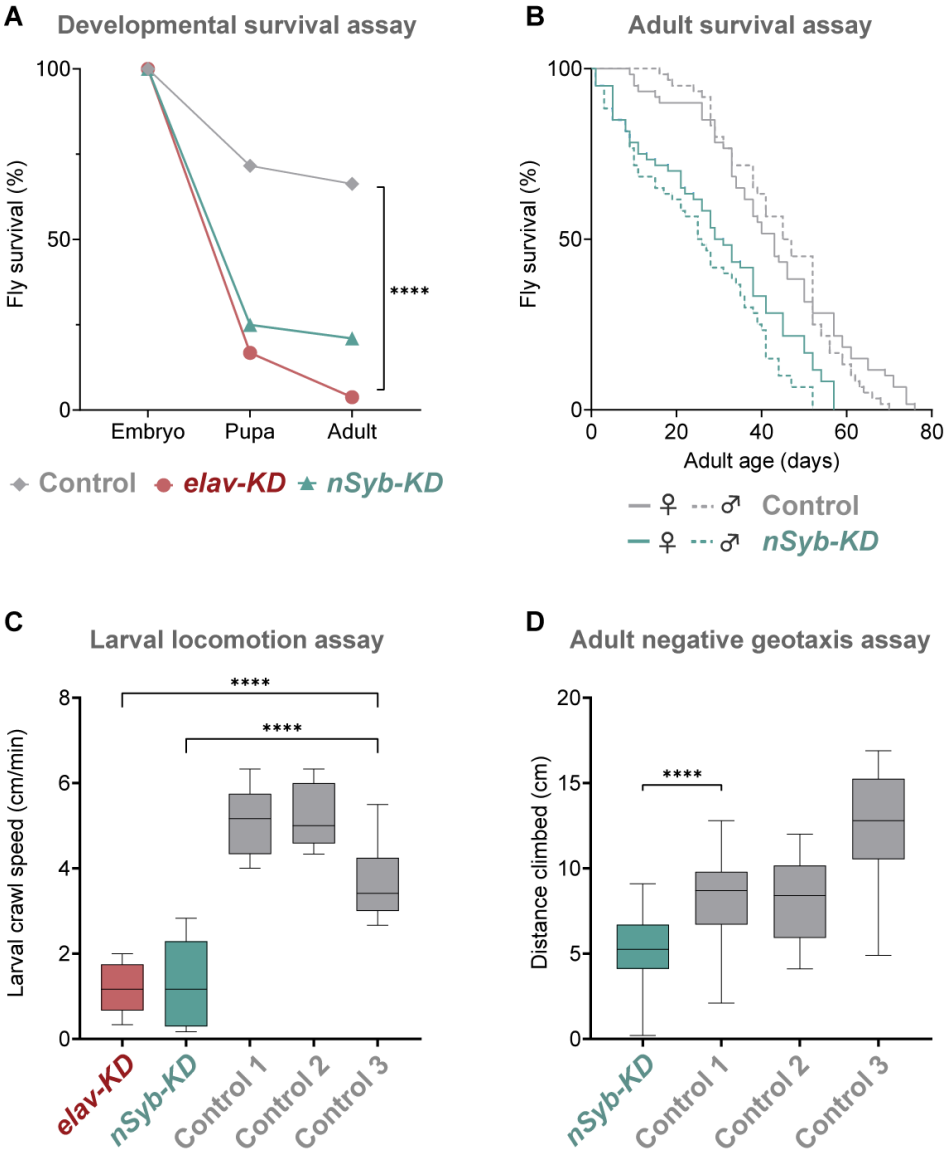


Figure 3.11. Cap-G knockdown in neurons results in premature developmental arrest and severe mobility defects.

A) Survival of developing flies recorded at pupal and adult stages. *Cap-G^{EGFP}* flies were used as the control genotype. Survival expressed as a percentage of the initial total number of embryos. 300 biological replicates/genotype. Logrank test and weighted Gehan-Breslow-Wilcoxon model (****, $p < 0.0001$). **B)** Survival of adult nSyb-*KD* and *Cap-G^{EGFP}* control recorded daily until all flies deceased. Survival expressed as a percentage of starting flies, 60 biological replicates/genotype. Logrank test and weighted Gehan-Breslow-Wilcoxon model (****, $p < 0.0001$). **C)** Locomotion assay of L3 larvae from 3 independent experiments. Control 1 = *Cap-G^{EGFP}; UAS-deGradFP*. Control 2 = *elav-Gal4; Cap-G^{EGFP}*. Control 3 = *Cap-G^{EGFP}; nSyb-Gal4*. Mean crawling speed from 10 technical replicates for each of the 30 biological replicates/ genotype. Kruskal-Wallis test one-way ANOVA (****, $p < 0.0001$). **D)** Negative geotaxis assay from 3 independent experiments. Control 1 = *Cap-G^{EGFP}; nSyb-Gal4*. Control 2 = *Cap-G^{EGFP}; UAS-deGradFP*. Control 3 = *nSyb-Gal4; UAS-deGradFP*. Mean distance climbed from 10 technical replicates for each of the 30 biological replicates/genotype. Kruskal-Wallis test one-way ANOVA (****, $p < 0.0001$).

3.2.6 Cap-G KD in neurons results in impaired mobility in flies

Larval locomotion is a distinct behavioural trait in *Drosophila* larvae and commonly used as a proxy to assess healthy neuronal development (174). Knockdown of Cap-G in neurons severely reduced larval locomotion ability. Both *elav-KD* and *nSyb-KD* animals displayed similar severity in their locomotion defects when compared to controls (Figure 3.11C). On average, the Cap-G knockdown larvae moved 4 cm/min less than the control animals ($p < 0.0001$, n =30). Mobility impairment carried over onto adult flies in both genotypes (Figure 3.11D). *elav-KD* adults displayed a very severe mobility phenotype, as they were unable to move after eclosion and died shortly after. As a few *nSyb-KD* animals successfully eclosed as adults, I was able to carry out another mobility assay known as negative geotaxis (climbing) assay (175). This exploits the innate fly response to climb against gravity, as flies are firmly tapped to the bottom of the tube and their upward distance climbed is recorded and scored.

The *nSyb-KD* adults showed a compromised ability to climb, with the mean climbing distance of ~5 cm, as compared to ~9 cm or ~13 cm for control genotypes (Kruskal-Wallis test one-way ANOVA , $p < 0.0001$, n = 30) (Figure 3.11D). Behavioural phenotypes can be affected by the genetic background of the animals. Therefore, all possible control genotypes were used to account for any differences caused by the genetic background. The controls included the *GAL4* lines used to drive the knockdown and the *UAS-deGradFP* line, all in a *Cap-G*^{EGFP} homozygous background. Overall, these data show that knockdown of Cap-G in neurons results in motor defects in both larvae and adult flies.

3.2.7 Increased levels of DNA damage in Cap-G KD neurons

The behavioural phenotypes observed suggest that Cap-G depletion affects the healthy development of the nervous system. Analysis of the brain tissue *in vivo* did not show any gross morphological defects in larval or adult brains, suggesting that Cap-G KD may lead to molecular and genomic instability, given the established role of condensin in chromosome organisation. Studies in yeast suggest the condensin complex regulates DNA damage response in interphase cells (176,177). To determine whether Cap-G KD had an effect on DNA damage response in neurons, I quantified DNA damage levels *in vivo* using ɣ-H2Av antibody as a marker for double-stranded DNA breaks (DSB) (178). The total number of ɣ-H2Av positive neurons were quantified as a percentage of the total number of ɣ-H2Av positive cells observed in the tissue.

In the larval CNS I observed an increase in neuronal DNA damage in both *elav-KD* and *nSyb-KD*, when compared to *Cap-G*^{EGFP} controls (Figure 3.12). In *elav-KD* and *nSyb-KD* we observed a two/threefold increase of ɣ-H2Av positive neurons compared to controls, (Kruskal-Wallis test one-way ANOVA, $p < 0.001$, n=10) (Figure 3.12B). There was no significant difference in the DNA damage between *elav-KD* and *nSyb-KD* tissues, an interesting result given the more severe survival phenotype seen previously (Figure 3.11).

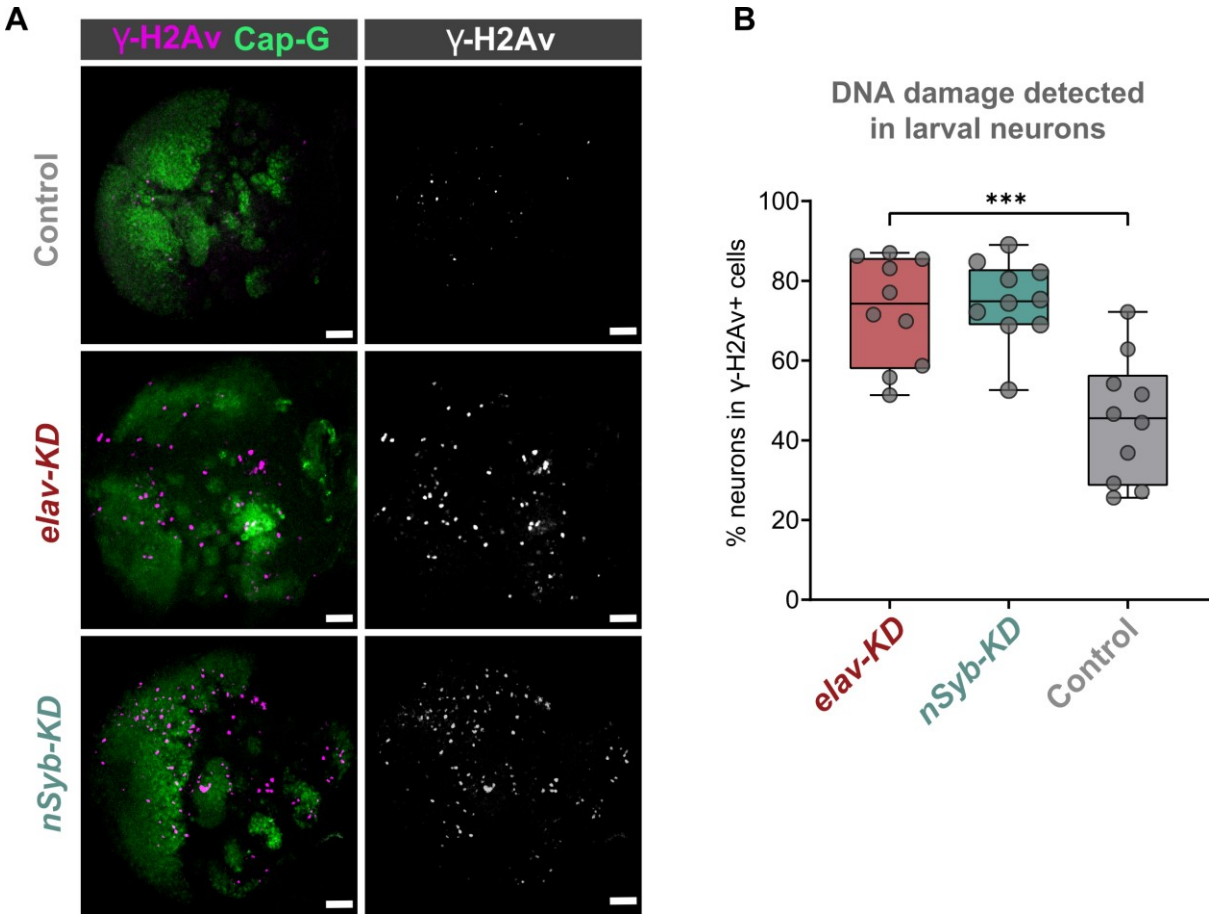


Figure 3.12 Cap-G KD leads to increased DNA damage in larval neurons.

A) Larval optic lobes for *Cap-G*^{EGFP} controls and *Cap-G KD* genotypes. ɣ-H2Av indicates DNA damage, Cap-G signal is strongest in dividing cells. Scale bars 10 µm. **B)** Quantification of ɣ-H2Av+ neurons as percentage of total ɣ-H2Av+ cells. 10 biological replicates per genotype. Kruskal-Wallis test one-way ANOVA (***, $p < 0.001$).

Condensin complex has an established role in chromatin compaction in mitotic cells, which impacts the shape, area and size of nuclei (82). In *Drosophila,* mitotic cells in the larval brains of heterozygous *Cap-G* mutant flies display diffused DAPI staining, accounting for aneuploidy and aberrant chromosome segregation in the nuclei (81). Moreover, in *Drosophila* S2 and Kc167 cell lines the depletion of condensin II subunits Cap-H2 and Cap-D3 resulted in an increased nuclear size, disrupting the area and shape of the nucleus (125). These studies report effects on mitotic and interphase cells, but there is no current knowledge on whether condensin affects chromatin compaction in post-mitotic cells. To investigate whether Cap-G has any role in regulating chromatin compaction in neurons I analysed the nuclear area of embryonic neurons in both *elav-KD* and *nSyb-KD* animals. DAPI staining was used to mark the DNA content in nuclei and Elav to mark the nuclei of neurons. There was no significant difference in neuronal nuclei area between both *Cap-G KDs* and the *Cap-G*^{EGFP} control (Figure 3.13). The average area ranged around 900-1000 pixel (~1000 cells per replicate analysed, 7 biological replicates). These results suggest that Cap-G in neurons may carry a role independent of the chromatin compaction function of the condensin complex.

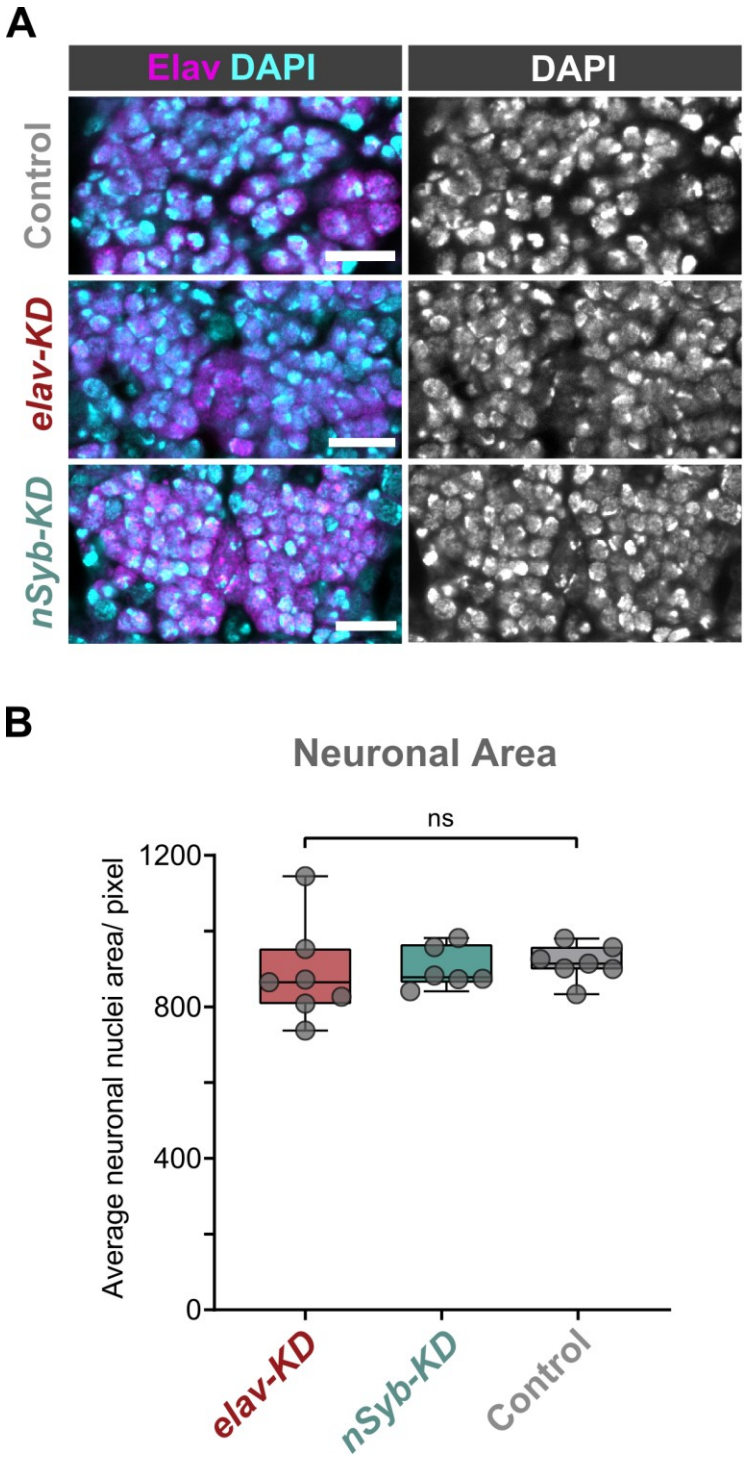


Figure 3.13 Cap-G knockdown has no effect on neuronal nuclei area.

A) Section of embryonic VNC (ventral view) stage 15 for all genotypes. Elav to mark neuronal nuclei and DAPI as nuclei stain. Control *Cap-G*^{EGFP}, *elav-KD* and *nSyb KD* genotypes. **B)** Quantification of neuronal nuclei area. Pixel area average of at least 1000 cells per replicate was analysed. 7 biological replicates per genotype. No statistical difference determined by one-way ANOVA (p> 0.5).

3.2.8 Cap-G KD leads to increased neuronal death

Further to the DNA damage observed in *Cap-G KD* animals (Figure 3.12) I wished to investigate whether there was an increase in neuronal apoptosis as a result of Cap-G KD. A previous report of condensin mutations in the mouse developing cortex led to increased apoptosis of NSCs and neurons (80). In zebrafish, *Cap-G* mutants displayed a significant increase in apoptosis of proliferating retina cells (135). Given these previous findings, I assumed that any degradation of Cap-G in newly-born neurons could result in a detectable increase in cell death.

TUNEL staining was done on *elav-KD* and *Cap-G*^{EGFP} embryos staged 14-16 and the total number of apoptotic neurons (TUNEL +ve, Elav +ve) was quantified (Figure 3.14). There was a significantly higher number of apoptotic neurons in *elav-KD* embryos compared to *Cap-G*^{EGFP} controls (Figure 3.14B). On average ~20 apoptotic neurons were detected in control embryos, compared to ~35 detected in *elav-KD* across 9 biological replicates (Figure 3.14B). Despite being a statistically significant difference, the effect on neuronal cell death observed is still mild.

As visual observations showed a number of TUNEL-positive cells that did not overlap with neurons, I wished to quantify whether there were any differences in apoptotic non-neuronal cells between *elav-KD* and control embryos. The area analysed was the embryonic VNC, therefore these could represent NSCs and other mitotic cells such as GMCs. There was no significant difference in the amount of apoptotic non-neuronal cells between *elav-KD* and *controls* (Figure 3.14C). Taken together these results suggest that knockdown of Cap-G in neurons results in a mild increase in neuronal death, with no effect on apoptosis of NSCs and dividing cells at the embryonic stage.

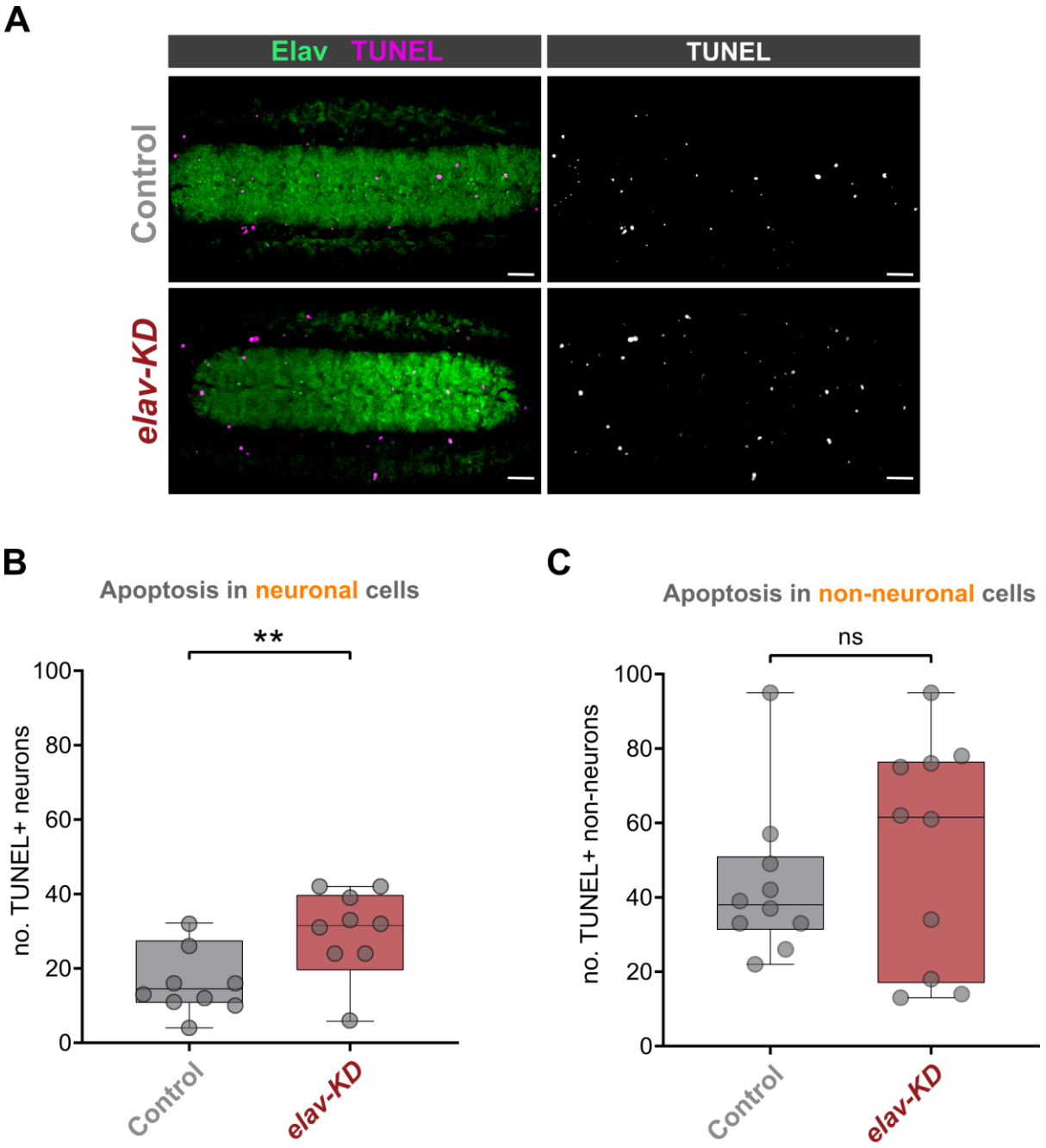


Figure 3.14 Increased neuronal apoptosis in *elav-KD*.

A) Staining for apoptotic cells in embryonic VNC stage 16 for *Cap-G*^{EGFP} (control) and elav-*KD*. Neurons marked in green by Elav and apoptotic cells in magenta by TUNEL staining. Anterior left, posterior right. Maximum Z-projection. Scale bar 10 µm. **B-C)** Quantification of apoptosis in embryonic VNC stage 15/16 for *Cap-G*^{EGFP} and elav-*KD*. **B)** Significant difference in number of apoptotic neurons (TUNEL+, Elav +), t-test p = 0.006. Outliers have been identified and removed using the Grubbs' test. 9 replicates per genotype. C) No significant difference in number of apoptotic non-neuronal cells (TUNEL +, Elav -) between *Cap-G*^{EGFP} control and elav-*KD*. 10 replicates per genotype.

3.2.9 Characterising the effects of *elav-GAL4* driven Cap-G KD

Given the vital role of Cap-G in dividing cells I had to ensure the tools used for protein knockdown were exclusively post-mitotic. It has been reported that *elav-GAL4* can drive low level expression in embryonic NSCs (179), which could interfere with the intended post-mitotic, *elav*-driven Cap-G^{EGFP} knockdown. Therefore, I undertook a careful characterisation of the *elav-KD* effects on NSCs and dividing cells. First, I analysed whether the *elav-GAL4* would drive degradation of Cap-G^{EGFP} in NSCs as well as neurons. Imaging of *elav-KD* embryos shows no significant difference in GFP levels in NSCs, marked by Dpn, compared to *Cap-G^{EGFP}* controls (Figure 3.15). However, the GFP-degradation levels may be subtle and difficult to detect without quantitative measures. Therefore, to determine whether low level degradation of Cap-G^{EGFP} was occurring, I quantified *in vivo* GFP levels in NSCs. There is no significant difference in the GFP levels in Dpn-positive cells between *elav-KD* and *Cap-G^{EGFP}* embryos (Figure 3.15B). Quantification was carried out by normalising the average pixel intensity of Cap-G^{EGFP} protein levels against Dpn signal. The levels are then represented as a ratio of EGFP/Dpn average signal intensity per cell. Across replicates for both *elav-KD* and *Cap-G^{EGFP}* controls the cells displayed a similar distribution in the signal ratio detected (Figure 3.15B).

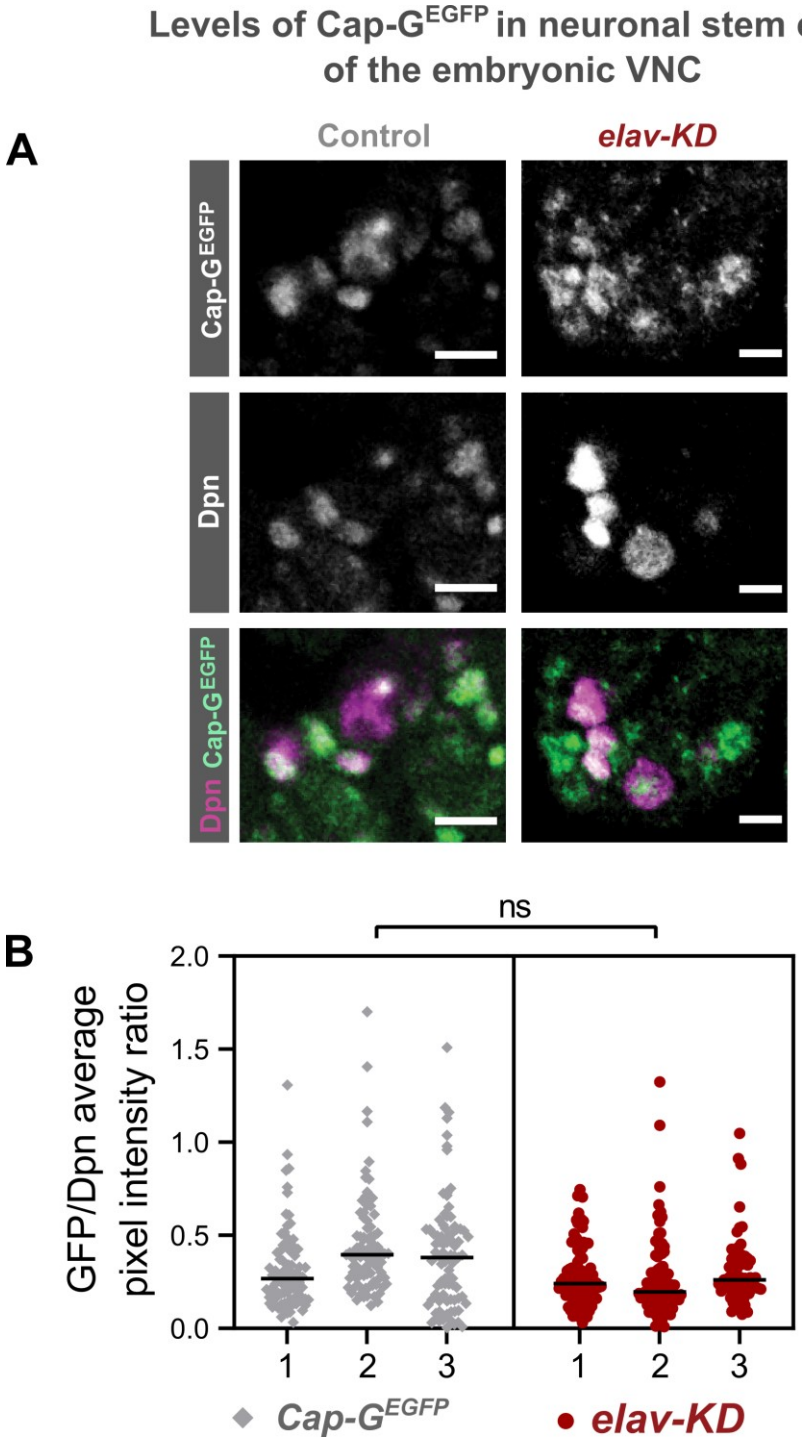


Figure 3.15 Cap-G^{EGFP} levels remain constant in NSCs between *elav-KD* and *control*.

A) Representative images showing Cap-G^{EGFP} in NSCs, marked by Dpn Note that NSCs show variable Cap-G expression depending on cell-cycle stage. Scale bars 5 µm. **B)** Quantification of Cap-G^{EGFP} signal in NSCs normalised against Dpn signal. Y-axis displays ratio of EGFP/Dpn average signal intensity per cell. The average pixel intensity is calculated for >70 cells/ embryo, with 3 biological replicates per genotype (*Cap-G^{EGFP}* control and *elav-KD*). No significant difference as per Nested t-test ($p > 0.05$).

The *in vivo* data presented above does not take into account other mitotic cells present in the VNC, for example the GMCs, which could still be affected by low-level Cap-G^{EGFP} degradation. Therefore, I quantified the GFP-fluorescence levels in live mitotic cells of the VNC in *elav-KD* and *Cap-G^{EGFP}* control embryos. There was no significant difference in the levels of Cap-G^{EGFP} in mitotic cells between the *elav-KD* and control embryos (Figure 3.16). Mitotic cells displayed high intensity GFP signals, which was used to threshold the live images and allowed for analysis of mitotic cells only, isolated as individual Regions of Interest (ROI) (Figure 3.16A). The average pixel intensity of mitotic cells ranged around 1700 pixels for both *elav-KD* and control embryos (Figure 3.16B). A potential reduction of GFP intensity in mitotic cells may affect the total number of cells detected by the thresholding algorithm. Therefore, I quantified the number of thresholded mitotic cells between both genotypes and noted no significant difference in both *elav-KD* and *Cap-G^{EGFP}* animals (Figure 3.16B). These results indicate that a robust neuronal knockdown was achieved using the deGradFP method and conclude that any low-level expression of deGradFP in NSCs is not sufficient to significantly reduce Cap-G^{EGFP} levels.

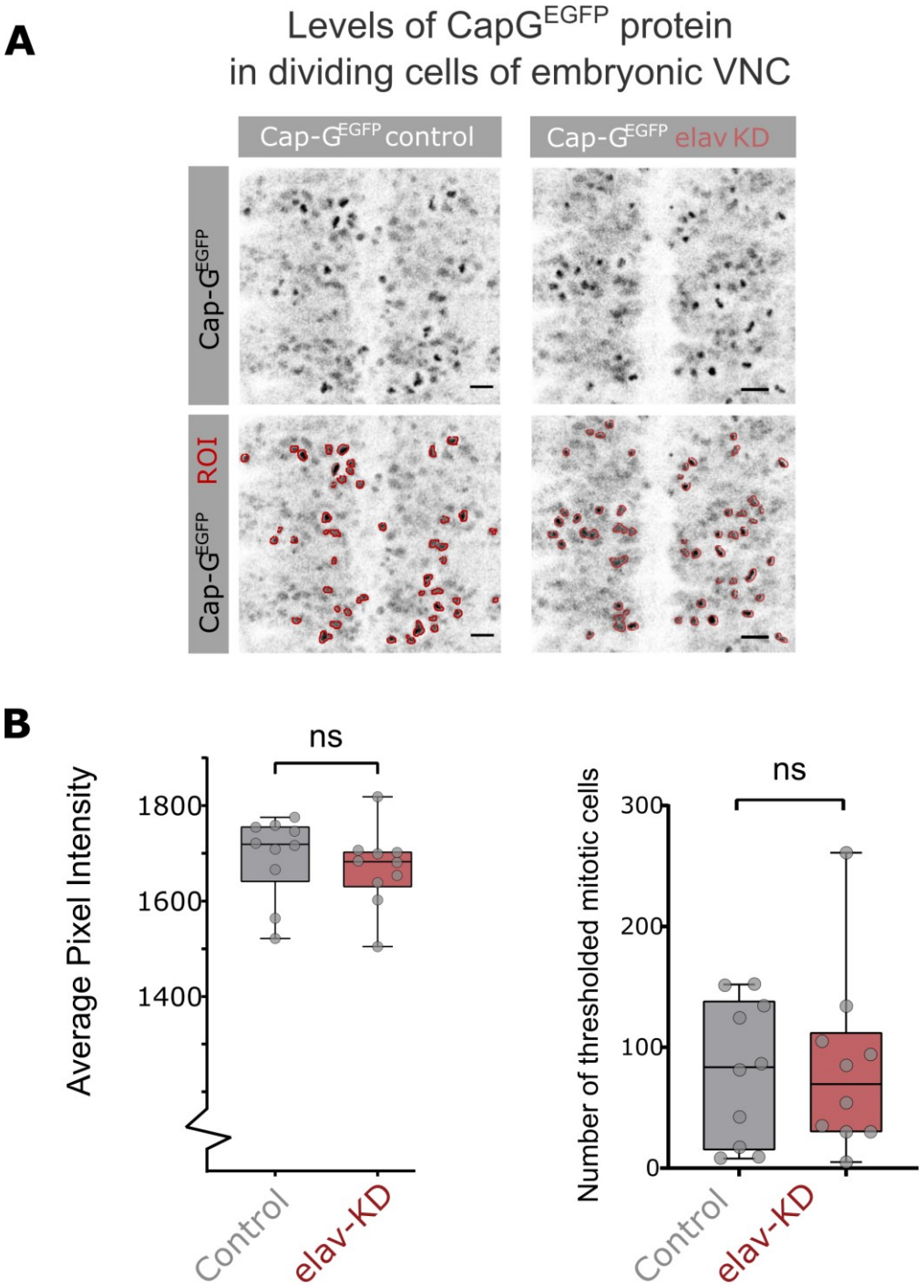


Figure 3.16 Live Cap-G^{EGFP} levels remain unchanged between *elav-KD* and controls.

A) Quantification of Cap-G^{EGFP} protein levels in dividing cells in *Cap-G^{EGFP}* control and *elav-KD* embryonic VNC (stage 14 /15– anterior top). Pixel intensity thresholding was performed to produce images in which dividing cells were defined as regions of interest (ROI) whilst neurons (lower pixel intensity) were excluded from the analysis. **B)** The average pixel intensity of dividing cells is calculated for >30 cells/ embryo, with 10 biological replicates per genotype. No significant difference in Cap-G^{EGFP} levels, t-test ($p > 0.5$). Number of mitotic cells (ROIs) remains constant. Scale bar 10 μm.

3.2.10 Cap-G KD driven by *elav-GAL4* has no impact on cell division

Since condensin has a well characterised function in mitotic cells, any depletion of condensin subunits from these cells would affect their ability to divide. To further confirm the post-mitotic specificity of *elav*-driven Cap-G KD I quantified the number of dividing cells, NSCs and mitotic NSCs in both *elav-KD* animals *and Cap-G*^{EGFP} controls. The embryonic stages analysed ranged from stage 14 to 16, when the highest number of NSCs and dividing cells is detectable (15).

There was no detectable difference in the number of NSCs, dividing cells nor dividing NSCs in both *elav-KD and Cap-G*^{EGFP} controls (Figure 3.17). A total of ~350 NSCs (Dpn-positive) were detected in both *elav-KD* and *Cap-G*^{EGFP} control embryos across 10 biological replicates (Figure 3.17A). I used the M-phase marker pH3, to account for all dividing cells in the VNC, including the GMCs (180). On average, 200 pH3-positive cells were detected in both genotypes, confirming there is not a significant difference in the number of mitotic cells between *elav-KD* and controls (Figure 3.17B). Finally, there was no difference in the overall number of dividing NSCs (Dpn, pH3-positive cells) between the Cap-G KD and controls (Figure 3.17B). These results suggest that there is no effect on NSCs in *elav-KD* embryos and that the method is specific to post-mitotic cells.

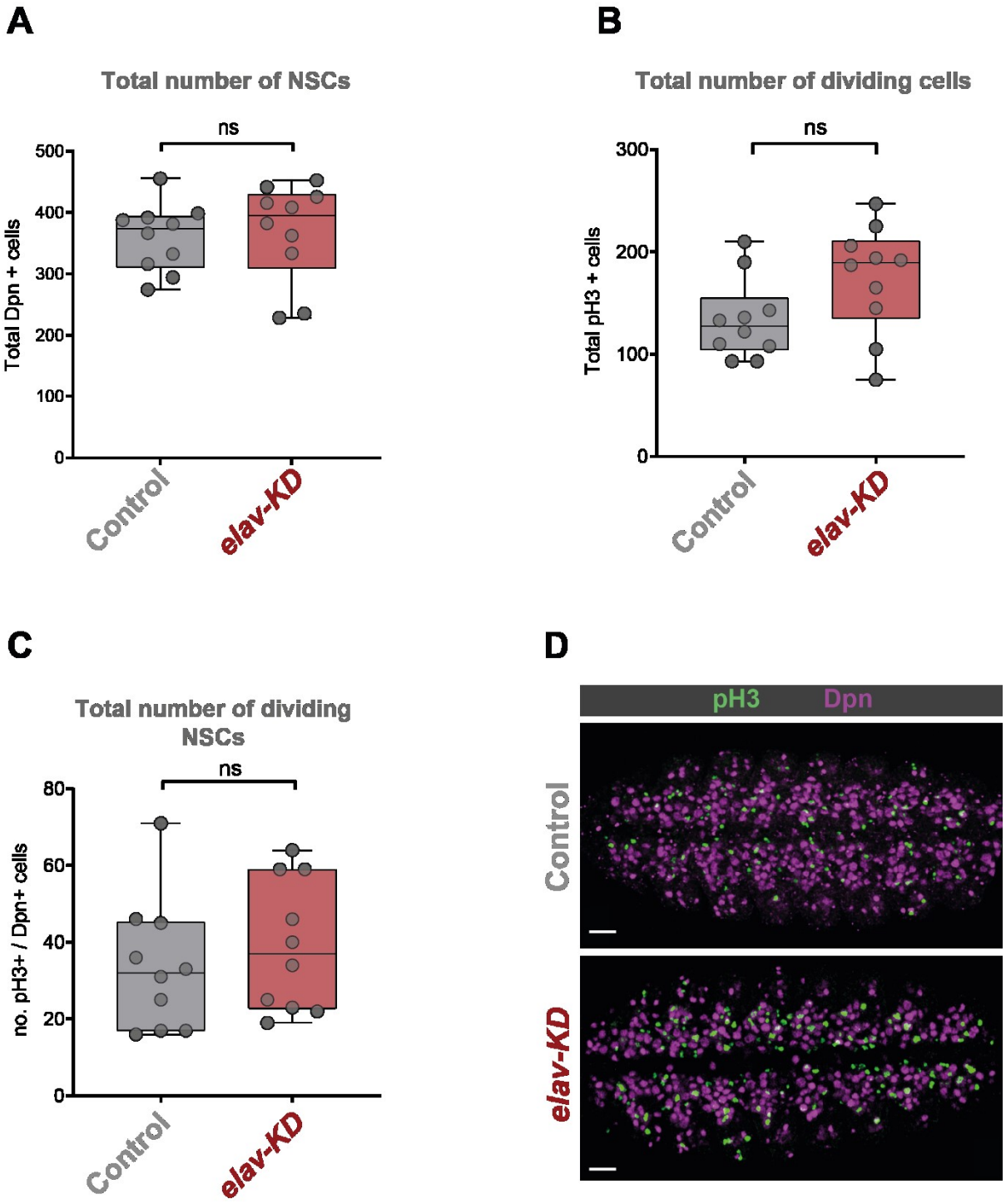


Figure 3.17. ***elav-KD*** **has no effect on mitotic cell numbers.**

Quantification of NSCs numbers and cell division in *Cap-G*^{EGFP} control and *elav-KD* individuals. Z-stacks of embryonic VNC stage 15 for *Cap-G*^{EGFP} and *elav-KD* were used. **A)** Quantification of total number of NSCs marked by Dpn. **B)** Quantification of total number of mitotic cells marked by pH3. **C)** Quantification of total number of dividing NSCs as number of pH3+/Dpn+ cells. 10 replicates per genotype. No significant difference as per t-test ($p > 0.05$). **D)** Maximum Z-projection of stage 15 embryonic VNC for both *Cap-G*^{EGFP} and *elav-KD* showing quantified NSCs (Dpn) and dividing cells (pH3). Scale bar 10 µm.

3.3 Discussion

3.3.1 Condensin I localisation and distribution in *Drosophila* neurons

Cell differentiation and maturation during development requires a tightly regulated interplay of transcription factors, co-factors and chromatin remodelling. A family of chromatin remodelling proteins are the condensin complexes, necessary to condense DNA into chromosomes and ensure effective cell division. The role for condensin in dividing cells is well established (53), but little is known about potential roles of condensin in post-mitotic cells. In this chapter I presented evidence for a novel role of condensin I subunit Cap-G in *Drosophila* neurons. Firstly, I characterised *Cap-G* expression patter *in vivo* and in-situ hybridisation results show ubiquitous expression of *Cap-G* mRNA in embryonic and larval neurons. These observations are consistent with previously published RNA-seq data (10,162). Moreover, in agreement with the RNA-seq data, I was unable to detect *Cap-G* expression in adult brains.

In order to characterise localisation of condensin I subunits Cap-G and Barren in neurons I utilised fluorescently tagged lines for both subunits: *Cap-G*^{EGFP} and *Barren*^{EGFP}. Both of these lines have been described to express the genes from their endogenous locus and to recapitulate the wild type expression of the subunits (138). Imaging of Cap-G^{EGFP} displayed its nuclear distribution in neurons. In contrast, Barren^{EGFP} localised to the cytoplasm of neurons. These observations are in agreement with the described localisation of condensin I subunits in *Drosophila* interphase cells, where Cap-G localised to the nucleus whilst subunits SMC2, Cap-D2 and Barren localised in the cytoplasm (66,76). This difference in localisation suggest that in post-mitotic cells, condensin I subunits may localise individually and perform roles distinct from the whole condensin complex. It is important to note that due to the lack of reliable antibodies for Cap-G and

Barren the localisation observed could not be validated by alternative methods. For example, I attempted to stain *wild-type* tissue with a previously described anti-Cap-G (66) antibody, however, the antibody failed to reliably recognise Cap-G in immunofluorescence experiments. Despite this, the EGFP tagged lines were viable and with no apparent ill effects, reinforcing that the observations made mirror the wild type condensin subunits.

Prior to the immunofluorescence experiments, a detected Y2H screen initially suggested presence of Cap-G in neurons, due to a detected interaction between Cap-G and Lola-N (unpublished, Southall T.D.). Stainings of embryonic VNC show that Cap-G and Lola-N co-localise in neuronal nuclei, further suggesting a potential Lola-N/ Cap-G interaction in neurons. Lola-N is a neuron-specific transcription factor essential in newly born neurons, as it maintains the differentiated state of the cell (19). Removal of Lola-N from newly-born neurons causes a de-differentiation phenotype, where neurons express stem-cell genes and proliferate uncontrollably causing tumours (19). However, when removing Lola-N from mature neurons these effects do not occur (19). Cap-G distribution was predominantly observed in early neurons, closest to the differentiating cell lineage. Given this observation and the fact that Cap-G co-localises with Lola-N, it is plausible to speculate that Cap-G may also be necessary at the early stages of neuronal development. However, it is possible that the Cap-G observed in such proximity of the progenitors may also be maternally inherited protein, and not relate to a function for Cap-G in early neurons.

In late embryonic stages, Cap-G^{EGFP} distribution resemble distinct foci in neurons further from the progenitor cells. A previous study in interphase cells observed similar foci localisation of Cap-G, describing it as a “non-homogenous distribution” in the

nucleoplasm that co-localises with HP1 (66). However, I observed only partial, restricted co-localisation between HP1 and Cap-G^{EGFP} in neurons. As this distribution was only observed in embryonic tissue and not larval it is plausible that it is temporally regulated and sparse in neuronal cells. In *Drosophila* nurse cells, Cap-H2 is involved in the formation of chromosome territories, by distributing heterochromatin and centromeric regions at the nuclear periphery (124). This foci localisation of Cap-G at some heterochromatin could relate to a function in spatially organising the genome in neuronal nuclei in *Drosophila*.

3.3.2 A post-mitotic role for Cap-G in neurons

This study presented first evidence for a robust post-mitotic knockdown of Cap-G in neurons, revealing a novel post-mitotic role. Cap-G knockdown in neurons resulted in phenotypes such as reduced survival, impaired mobility and genomic instability. There were no gross morphological changes to the CNS or the overall animal at any stage of development, despite the observed phenotypes. *Cap-G* mutants in *Drosophila* are homozygous lethal, however heterozygous animals for some alleles showed rough eyes phenotypes (81). I did not observe any eye phenotype in my neuron-specific knockdowns (which include eye tissue), confirming the robustness of the knockdown and that the phenotypes observed may be a consequence of a separate post-mitotic role for Cap-G.

Cap-G knockdown animals were generated using the deGradFP system coupled with *GAL4/UAS*, resulting in highly specific degradation of GFP-tagged proteins in a cell-type of interest (137). The deGradFP has been shown to successfully degrade proteins in the *Drosophila* CNS and in some instances recapitulate null mutant phenotypes (169). In my experiments, ubiquitous degradation of Cap-G^{EGFP} resulted in embryonic lethality, the same phenotype observed in *Drosophila Cap-G* null mutants (81). This result gave

confidence in the effectiveness of the KD method used. Further validation of Cap-G KD is seen in clones, which showed that GFP was successfully degraded, but not completely depleted. In neuron-specific Cap-G KD some GFP signal is still detectable in neuronal cells, suggesting that the knockdown of Cap-G from neurons is not a complete depletion. This could be attributed to the two/ three-hours onset time of the deGradFP system (137). Despite this, the phenotypes of the knockdown are very severe, further emphasising the importance of Cap-G in neurons.

To ensure a robust, post-mitotic neuronal phenotype I induced Cap-G knockdown in two overlapping populations of cells, immature to fully differentiated neurons (expressing *elav*) and mature, fully differentiated neurons (expressing *nSyb*). Interestingly, survival phenotypes exhibited by the *elav-KD* were more severe than the *nSyb-KD*. This could be due to the maturation stage of the cells targeted by each driver. The *nSyb-GAL4* targets fewer neurons which have already matured and lost plasticity, whilst *elav-GAL4* also targets newly born neurons which are in a more transient differentiating sate. Given I observed Cap-G distribution mostly in newly born neurons, I can assume that Cap-G may play a more prominent role in the early life of neurons. For example, Cap-G may be required for terminal differentiation of newly born neurons and aid their maturation. The deGradFP system is able to effectively degrade maternally inherited protein (169), therefore, if the Cap-G^{EGFP} observed close to the progenitors were to be inherited by the neurons, it would still be depleted. Results with *nSyb-KD* show impaired mobility and survival in adult flies, suggesting that, even if at a lesser degree, Cap-G may be required to maintain proper neuronal identity and function. Overall, condensin I seems to have implications on effective neuronal development and the overall functioning of the *Drosophila* nervous system.

One of the drivers used in this study, *elav-GAL4*, has previously reported low-level expression in NSCs (179). This expression could interfere with the intent of generating an exclusive post-mitotic knockdown of Cap-G. Therefore, I took extensive qualitative and quantitative measures to assess whether any Cap-G^{EGFP} degradation occurred in NSCs. Image analysis of *in vivo* and live images did not detect any decrease in Cap-G^{EGFP} levels in mitotic cells and NSCs and the number of mitotic cells and NSCs also remained constant. Furthermore, functional analysis on NSCs show that their ability to divide remained unaffected as well as apoptotic ratio. In the mouse cortex, depletion of condensin I resulted in increased death of NSCs (80). Therefore, the absence of an increase in apoptosis in progenitors suggests that Cap-G in NSCs is not degraded in my experiments.

The knockdown and control populations observed often displayed a widespread dataset, which may be attributed to the variations in staining quality as well as the intensity of the KD driven by Cap-G, which may differ across generations. Multiple biological replicates were performed in the analysis to account for this spread in the population. Taken together, characterisation of the *elav-KD* animals displays no ill effects in mitotic cells, suggesting that the Cap-G KD is indeed in post-mitotic neurons. However, there is some possibility that Cap-G knockdown may occur at levels too low to be detected by the measures used, or in an undetected subset of progenitor cells, giving rise to phenotypes related to aberrant chromosome segregation. That phenotypes were observed with two independent neuronal GAL4 lines (*elav-Gal4* and *nSyb-Gal4*) provides further confidence in a *bona fide* role for Cap-G in neurons.

Further to the behavioural phenotypes, I observed an increase in DNA damage in both Cap-G knockdowns. Previous studies in yeast have implicated the condensin complex in

the regulation of the DNA damage response during interphase (176,177). Moreover in *Drosophila,* condensin II subunit Cap-D3 has been implicated in inhibiting double stranded break (DSB) formation (181). In contrast, condensin I in human cells was not involved in DSB repair, but only in single stranded break (182). Moreover, similar levels of DNA damage were detected in both *elav-KD* and *nSyb-KD,* suggesting that potential involvement of Cap-G in DSB repair in neurons may be independent of the neuronal maturation stage. Overall, results suggest that condensin I may be involved in DNA damage repair pathways in *Drosophila.* Consistent with the increase in DNA damage in neurons, I also observed a mild increase in neuronal apoptosis in *elav-KD* embryos. Similarly, significant apoptosis was seen in the proliferating cells of zebrafish retinas in *Cap-G* mutants (135). These findings suggest that despite not causing morphological changes to neurons, Cap-G depletion leads to genomic instability and is necessary to maintain healthy neuronal cells.

Cap-G depletion from neurons has no effects on neuronal nuclei area. Imaging of DAPI signal across neuronal nuclei did not show any changes in nuclear area, nor diffusion of signal. These results were somewhat surprising as knockdown of condensin has been shown to affect nuclear area (97,125). In quiescent yeast cells, knockdown of condensin I subunit SMC4 results in an increase in nuclear volume, used as a measure of chromatin compaction (97). In *Drosophila,* condensin II Is required to maintain nuclear architecture of interphase cells, as knockdown of Cap-H2 and Cap-D3 results in an increase in the nuclear area (125). Moreover, in *Drosophila Cap-G* heterozygous mutants there was a diffusion of DAPI staining in mitotic cells of the larval brain (81). As the results in this study are from post-mitotic knockdown of Cap-G in neurons, it suggests that a neuronal role for Cap-G may not be related to maintenance of nuclear architecture. Overall Cap-G

plays a regulatory role in neuronal development as it is necessary for fly survival and effective functioning of the nervous system.

Chapter 4 Cap-G binds dynamically to chromatin in neuronal stem cells and neurons

4.1 Introduction

4.1.1 Condensin binding at multiple genomic regions in the DNA

Condensin complexes have the ability to bind to DNA and drive compaction of chromatin, impacting on gene regulation and chromatin organisation (82,83). The condensin I complex is able to anchor itself onto DNA via its HEAT- repeat subunits, Cap-G and Cap-D3 (68). An electrophoretic mobility shift assay (EMSA) showed that Cap-G and Cap-D3 have strong affinity to double-stranded DNA *in vitro* (68). Moreover, in yeast and human cells the HEAT-repeat subunits are responsible for loading the whole condensin complex ring onto the chromosomes (68). The ability of condensin to directly associate to DNA led to the hypothesis that condensin could target specific genomic regions. To date the DNA binding profiles for condensin I and II have been studied in mitotic and interphase cells, mostly using ChIP-seq methods (Table 4.1). However, condensin I binding was largely the same in both mitotic and interphase cells in yeast (87), so these cell types are not enough to have a comprehensive view on condensin function. Currently, there is a gap in our knowledge of condensin, as binding patterns in post-mitotic, terminally differentiated cells are unknown. Given condensin are highly conserved, chromatin-associating complexes, they may carry important roles in cell development and differentiation. Therefore studies on condensin in more specialised cell types are necessary to increase our understanding of these complexes.

Table 4.1 Summary of condensin genomic binding

Condensin I	***Species/tissue***	***Binding regions***	***Method***	***References***
SMC4, BrnI (Barren)	*S. pombe*	tRNA, scRNA, snRNA genes	ChIP-seq	(87)
Cap-D2, Cap-G, Cap-H (Barren)	*C. elegans*	tRNAs, TSS (promoters), H3k27ac open chromatin markers	ChIP-seq	(88)
SMC2, Cap-H	Chicken DT40 cells	promoters, TSS of active genes, non-coding RNAs, tRNAs and rRNAs, gene bodies	ChAP-seq	(86)
Barren	*D. melanogaster* Kc167 cells	tRNA genes	ChIP-seq	(90)
Cap-G	Human breast cancer cells	Estrogen-dependent localisation at enhancers, ER-α peaks, intergenic and intronic regions	ChIP-seq	(94)
Cap-G	*D. melanogaster* NSCs, neurons (elav) and mature neurons (nSyb)	Gene bodies, repressed genes, Lola-N binding sites, repressive chromatin e.g. Black, PcG, HP1	Targeted DamID	This study (1)
BrnI (Barren)	*S. cerevisiae*	**In mitotic cells:** tRNAs, centromere **Quiescent cells:** increased number of peaks, localises at divergent gene promoters, overlaps Msn2 binding	ChIP-seq	(97)
Condensin II				
Cap-H2	*D. melanogaster* Kc167 cells	enhancers	ChIP-seq, Hi-C	(93)
Cap-H2, Cap-D3, Cap-G2	*C. elegans*	TSS, tRNAs, open chromatin H3k27ac, enhancers, LIN-13 sites	ChIP-seq	(88)
Cap-H2	mouse ESCs	TFIIIC sites, enhancers	ChIP-seq	(85,89)
Cap-H2, Cap-D3	mouse ESCs	CTCF sites, TFIIIC	Hi-C, ChIP-seq	(90)

Condensin binding has been characterised at multiple genomic features across different species with some disagreements in the data (Table 4.1). Condensin has been implicated with active transcription sites, as in *C. elegans* binding concentrated around TSS of active genes, whilst in yeast and mouse cells condensin was enriched at RNA Pol II sites (84,85,87). Similarly, in chicken DT40 cells condensin I binding overlapped with RNA Pol I sites (86). Multiple studies report condensin I binding at RNA Pol III (TFIIIC) sites, such as tRNAs and rRNAs (86-89) (Table 4.1). Taken together these studies suggest that condensin binding coincides with binding sites of RNA Polymerases (RNAP) and potentially affect transcription. However in chicken DT40 cells, condensin binding is not restricted to RNAP sites, as SMC2 and Cap-H binding is also observed in the bodies of transcribed genes (86). Overall, condensin is able to localise to multiple genomic regions, and the loci may be dependent on cell stage and species.

Condensin binding has also been associated with co-factor interactions, whereby condensin subunits are recruited to specific loci by other proteins. For example, condensin binding in *C. elegans* overlapped with binding sites for TF LIN-13 (88). In human cancer cells, ChIP-seq data revealed large overlaps between Cap-G and TF Estrogen Receptor α (ER- α) (94). Moreover, Cap-G physically interacts with ER- α via its DNA- binding domain which facilitates the recruitment of condensin onto active enhancers in response to estrogen stimuli (94). In *Drosophila,* the condensin II subunit Cap-H2 interacts with DNA-binding protein Mrg15, which partially regulates levels of Cap-H2 at chromatin sites (96). These studies highlight the importance of identifying condensin interactor proteins, to unveil more specific, co-factor dependent functions of condensins.

4.1.2 3D genome organisation and chromatin states

Chromatin topology refers to a hierarchy of structures that organise the chromatin in a 3D conformation in the nucleus. The genome is separated into large multi mega-base structures known as chromatin compartments. These compartments are split into broadly active chromatin (compartment A) and broadly inactive chromatin (compartment B) (117). The genomic sequences present in the active/ inactive chromatin compartments can interact with each other, and these interactions can occur over long range distances across the length of a chromosome (117). Within the chromatin compartments, there are smaller structures, ranging from 100 kb- 1 Mbp in size, known as topologically associated domains (TADs) (113-115). These TADs encompass genomic loci which interact at higher frequency with each other than with other regions of the genome, for example enhancers and their target genes (183). Most TAD regions are separated by "boundaries", which are genomic loci enriched for CTCF binding sites (a transcriptional repressor) and highly transcribed genes (111,117).

Generally, TADs are the smaller sub-structures found within the larger chromatin compartments, however, recent research identified smaller structures (10-50 kb) referred to as "compartmental domains" (as reviewed in (117)). Compartmental domains resemble the compartments A and B and are formed by the segregation of active and inactive chromatin, just at a smaller scale (123,126). These domains have been identified in other species, including *D. melanogaster* and *C. elegans* (123). Chromatin compartments and compartmental domains are the most unstudied topology, as there is little understanding of the proteins and chromatin changes involved in creating active/inactive regions, though recent work has focused on addressing these gaps (117).

In *Drosophila,* genome organisation is highly dynamic and chromatin has been broadly divided into different "states", identified by key associating proteins (120). Chromatin states can include two euchromatin (accessible/open) "red" and "yellow" states, two heterochromatin (inaccessible/closed) states marked by HP1 and Polycomb Group (PcG) and finally a repressive "Black" chromatin state making up 50% of the genome (121). The genomic regions present in these chromatin states are variable depending on the cell type and developmental stage. Therefore, chromatin landscape of a cell can influence gene expression and hence its identity (120). Cell differentiation drives progressive compaction and inaccessible chromatin to form, whilst in dividing and early differentiating cells the chromatin is generally more accessible (7,14,133). For example, in *Drosophila* NSCs, genomic regions such as enhancers required for neurogenesis are accessible, but subsequently compacted in differentiated neurons (14). Similarly, in neurons, most stem cell and mitotic genes are repressed in an HP1 heterochromatin state (7). Chromatin landscape changes clearly play an important role in cell development, however, the mechanisms underlying these processes remain largely unknown.

There is a relationship between chromatin topology and the transcriptional active/ inactive state of chromatin (marked by histone features). In *Drosophila*, high resolution Hi-C data (250 bp) allowed the identification of smaller (~10 kb) compartmental domains (123). These are smaller than TADs and correspond to the active/inactive chromatin states. The authors collated HiChIP data for histone modification H3K27me3 (inactive) and H3K27ac (active) which captures long range interactions between these two features (protein-centric chromatin conformation method) (119,123). Comparing the interactions between these histone features revealed the presence of smaller compartmental domains that matched the A/B compartments known from Hi-C data

(123). These results suggest that in *Drosophila* the chromatin segregates into compartmental domains depending on the chromatin state of those regions, revealing a relationship between topology and active/inactive states. Therefore, increasing our understanding of what proteins orchestrate chromatin landscape changes would offer a window into understanding how active/inactive chromatin compartments could be formed.

4.1.3 The condensin complex and the chromatin landscape

Condensin binding has been observed at enhancer and promoter regions, which are accessible chromatin regions contained in TADs and generally known as a hub of active transcription (111). For example, the Cap-H2 subunit of condensin II localises to enhancers in chicken and mammalian cells (85,86). Condensin localisation at enhancer regions in *Drosophila* remains unclear. A study suggests that Cap-H2 localises to enhancer regions (93), however this is not observed in a separate study on the same cell type, which reports scarce overlap of condensin I and II with enhancers (90). Most evidence of enhancer binding is for condensin II, whereas there is limited knowledge for condensin I subunits (Table 4.1). During interphase of human cancer cells, Cap-G binding is hardly detected at promoter regions however, upon estrogen stimuli strong binding to enhancers is detected (94). Our knowledge of condensin I localisation at enhancers is still poorly understood, especially in the context of specific cell types. Condensin localisation to enhancers may be dependent on other factors such as co-factor interactions, cell stage and species studied.

There is some evidence suggesting that condensin complex interactions with chromatin leads to global landscape changes, affecting gene expression. For example, in mouse T-cells and yeast, condensin II maintains the quiescent state of cells by driving chromatin

compaction (97,99). In T-cell development chromatin condensation increases as the cell differentiates from immature thymocyte onto a mature one, and this process is dependent on Cap-H2 function (99). Loss of Cap-H2 from mature thymocytes results in an opening of the chromatin (99). Chromatin compaction could relate to chromatin topology changes, however it remains unclear whether chromatin compaction is a cause or consequence of mass 3D genome reorganisation.

Condensin and cohesin are chromatin associated complexes with the ability to interact with chromatin and influence 3D genome organisation. In mammals, cohesin has been characterised as the SMC complex required for TAD formation (63,117,126,127). On the other hand, a role for condensin in TAD formation still remains unclear. Most evidence is for subunit Cap-H2, where it strongly localises at TAD borders and is necessary for maintenance of TAD structure (90,93). Similarly, in *C. elegans* disruption of the condensin-like DCC complex results in loss of TAD formation (128). These studies support a role for condensin II in TAD formation however, there is scarce evidence for condensin I in this process. Contradictory evidence for a role of condensin in TAD formation has emerged, where depletion of Cap-H2 in mouse post-mitotic hepatocytes does not alter TAD conformation (110). Further research is necessary to understand how condensin affects chromatin topology, and whether it changes the transcriptional landscape to impact cell development, especially in post-mitotic cells.

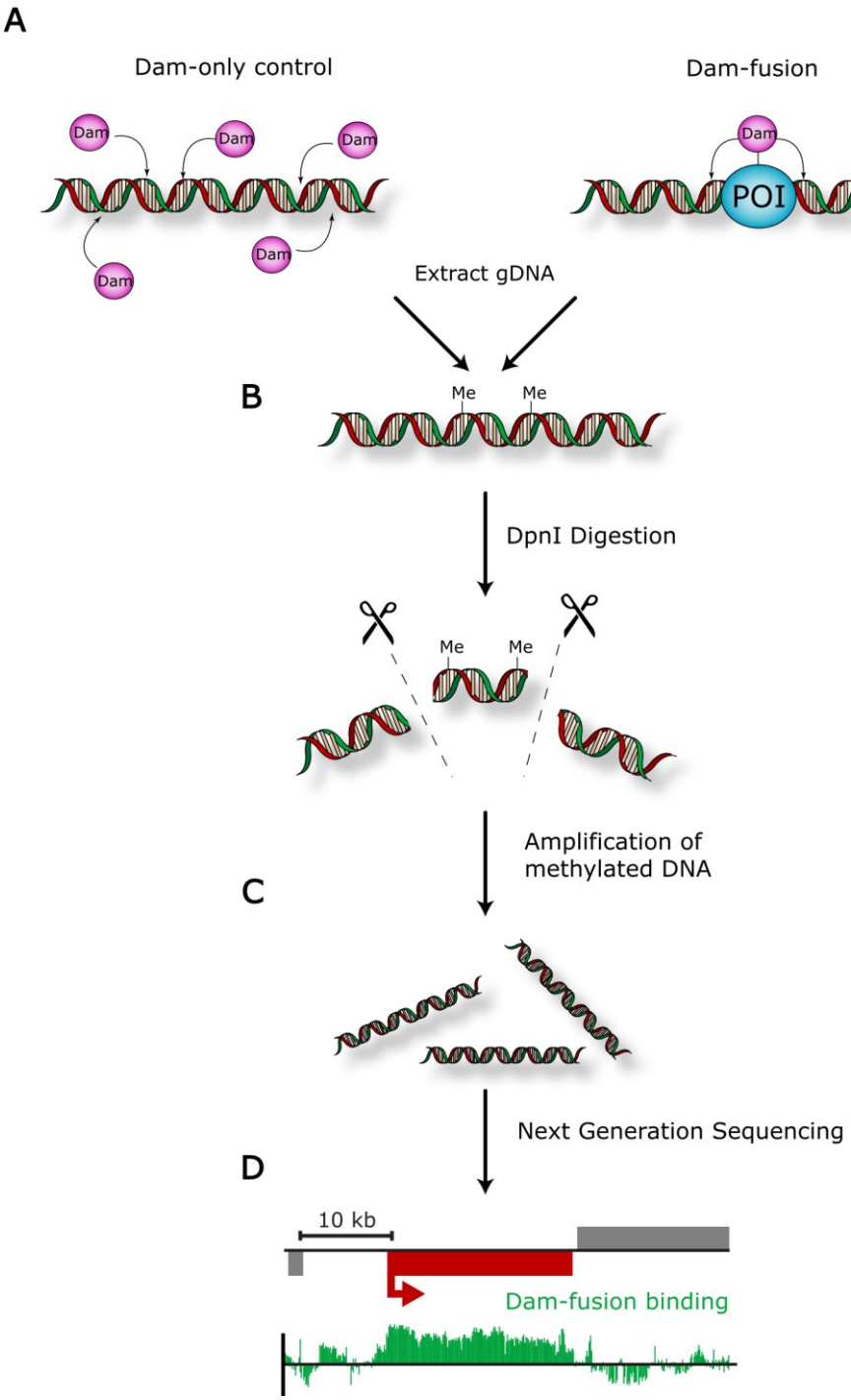


Figure 4.1 DamID experimental pipeline.

A) Dam is fused to a protein of interest (POI). Dam methylates adenine on GATC motifs on DNA sequences nearby POI binding sites. Dam-only binding is used to control for background methylation. **B)** Methylated DNA is extracted and cleaved using DpnI endonuclease. **C)** Methylated DNA is amplified using ligation-mediated PCR. **D)** Dam-fusion binding is normalised against Dam-only binding, providing a genome-wide binding pattern of the POI. Figure adapted from review (184).

4.1.4 Targeted DamID to profile cell-specific binding

In this chapter I use a cell-specific, DNA profiling technique known as Targeted DamID (TaDa) (140,141). This technique is based on the DNA adenine methyltransferase identification (DamID) method (Figure 1). DamID exploits the *E. coli* DNA methyltransferase enzyme (Dam) which methylates adenine residues in GATC motifs (184). Dam is fused to any protein of interest and once expressed at low levels, its DNA binding pattern is marked by methylation at nearby GATC sites (Figure 4.1A). Methylated DNA regions can be cleaved by the endonuclease DpnI and enriched using ligation-mediated PCR (Figure 4.1B, C) (184). The DNA is sequenced using Next Generation Sequencing technology to reveal genome-wide binding pattern of the protein of interest (Figure 4.1D). The control samples express only the Dam, to account for background methylation. The final sequencing data for the protein of interest is normalised against the Dam-only control before conducting any data analysis.

In *Drosophila,* TaDa has been developed to allow for cell-specific profiling of DNA binding without the need to isolate cells (140). TaDa uses the *GAL4/UAS* system to drive the expression of Dam-fusion protein in specific cell types (Figure 4.2A). Excessive Dam expression can be toxic to the animals, so to overcome this, the Dam-fusion construct is encoded as a secondary open reading frame (ORF) downstream of the UAS enhancer (184). The translation of a secondary ORF is reduced in this system due to the ribosome reinitiation phenomen (185). The primary ORF expresses a non-intrusive, fluorescent protein (mCherry) and is upstream of the Dam-fusion construct, so upon GAL4 activation, the primary ORF is translated at high levels, whilst the Dam-fusion remains at low, non-toxic levels (Figure 4.2A). Moreover, the primary ORF also reduces Dam-fusion

translation levels sufficiently to remove any non-specific Dam expression, resulting in cell-specific methylation (184,186).

In a recent study, analysis of TaDa control samples has been adapted to reveal accessible chromatin sites in the cell types of interest, a method called Chromatin Accessibility TaDa (CATaDa) (14)(Figure 4.2B). This method takes advantage of the fact that the untethered Dam preferentially methylates accessible chromatin sites in the specific-cell types used. The Dam-only controls are usually needed to normalise the TaDa data, but with CATaDa these control samples become a powerful dataset that allows profiling of accessible chromatin sites in a cell type of interest.

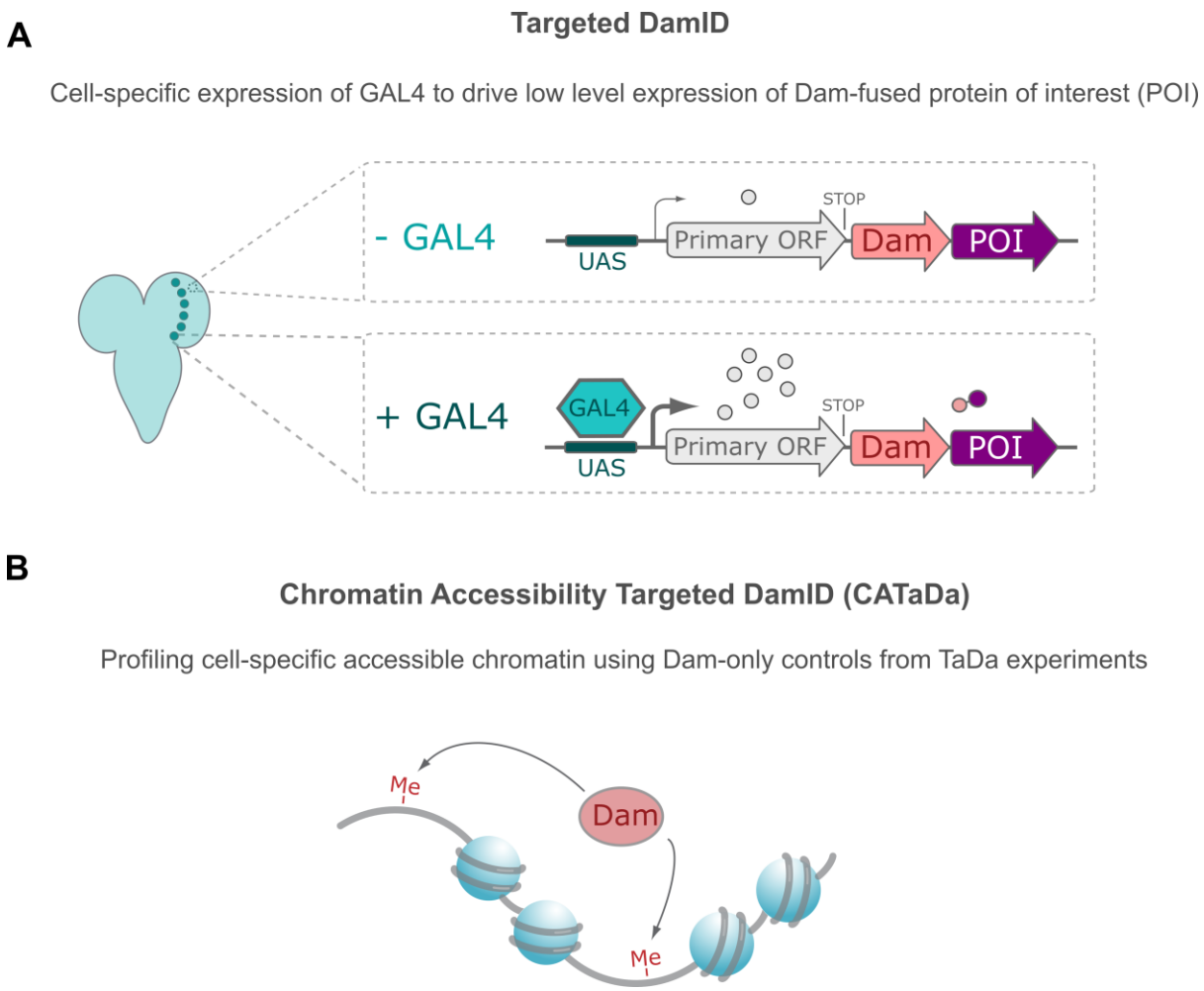


Figure 4.2 Targeted DamID for cell-specific profiling of Cap-G binding.

A) Targeted DamID (TaDa) exploits the *GAL4/UAS* system to express Dam in a cell specific manner, e.g in larval neurons. GAL4 binds to UAS, driving non-toxic, low-level expression of *Dam-fusion* by using primary ORF as an intermediate to attenuate translation. **B)** CATaDa uses the control, Dam-only background methylation as a read out for accessible chromatin sites in cells of interest. Figure adapted from review (187).

4.1.5 Profiling cell-specific binding of Cap-G in NSCs and neurons

The current literature portrays diverse binding patterns for condensin, seemingly dependent on transcription, protein interactions and also chromatin architecture. However, these data are limited to mitotic and interphase cells, revealing a gap in our knowledge of condensin binding in post-mitotic cells. In this chapter I aimed to characterise cell-specific Cap-G binding profiles in *Drosophila* NSCs and neurons using TaDa. Therefore, I generated the *UAS-LT3-Cap-G-Dam* transgenic *Drosophila* line (Chapter 2, Section 2.11) and utilised it to investigate Cap-G binding to DNA in different cell types of the nervous system (Figure 4.3). The cell types of interest used in this study were targeted using cell-specific driver lines. For NSCs, *wor-GAL4* driver was used to profile cell-specific Cap-G binding. For neuronal binding I utilised the *elav-GAL4* and *nSyb-GAL4* lines previously used and for the reasons delineated in Chapter 3. The cell types and stages analysed will be referred to in this chapter as NSCs (*wor-GAL4*), elav (*elav-GAL4*) and nSyb (*nSyb-GAL4*) hereafter. Moreover, in this chapter I will compare Cap-G binding profiles to the open chromatin regions for these cell types, using the CATaDa technique. Profiling these three cell types offers an insight in how condensin binding changes from a mitotic cells to a post-mitotic cell, how it relates to open chromatin and whether any cell-specific binding can be detected.

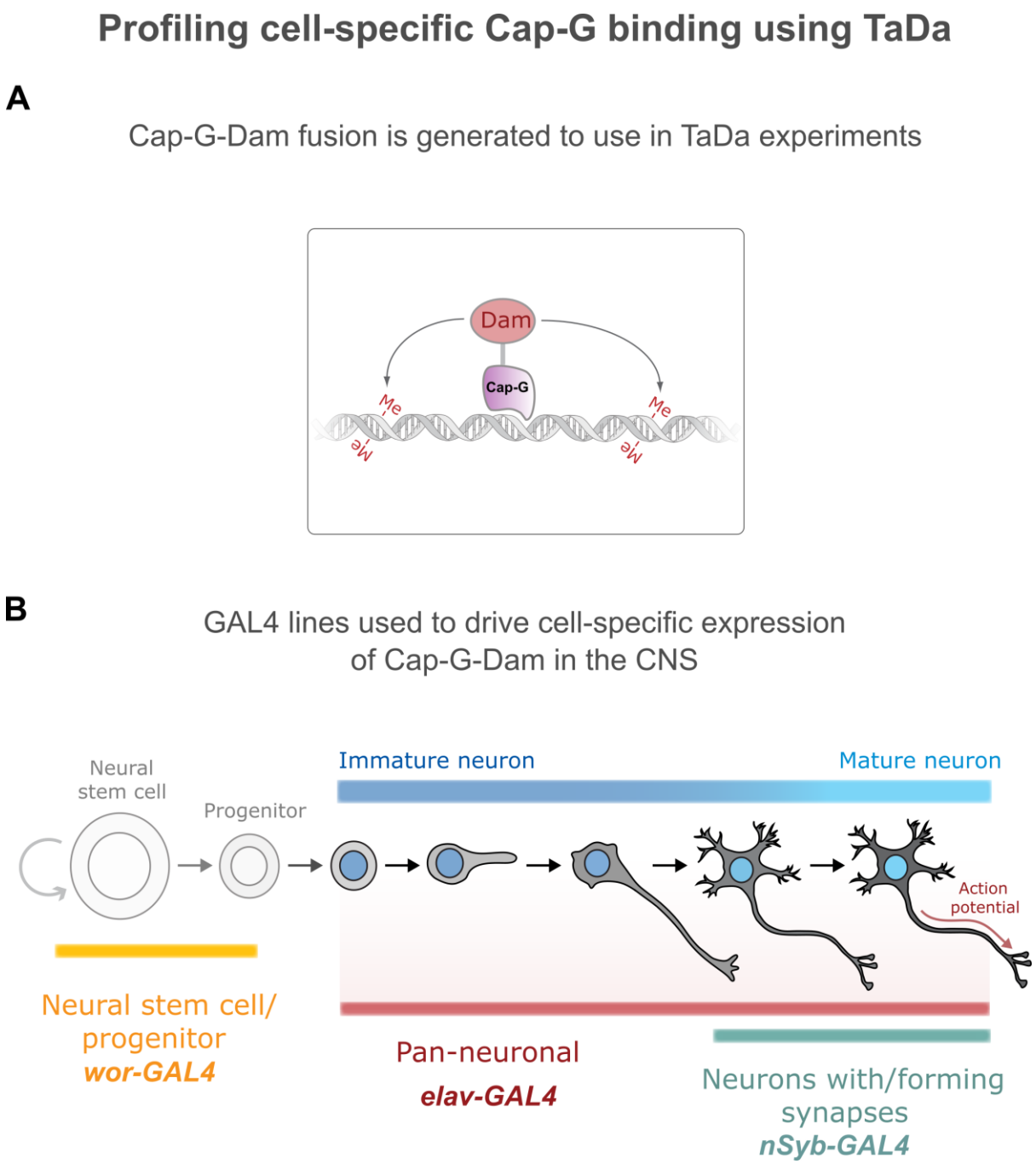


Figure 4.3 Cell specific profiling of Cap-G using TaDa.

A) Cap-G-Dam is generated to use in TaDa experiments. Dam will methylate GATC motifs adjacent to Cap-G binding sites therefore enabling profiling of Cap-G binding in a cell-specific manner. **B)** Cap-G binding is profiled in the *Drosophila* larval CNS in three selected cell types: NSCs (*wor-GAL4*) for dividing, progenitor cells, newly born and mature neurons (*elav-GAL4*) and synapse-forming, mature (*nSyb-GAL4*) neurons.

4.1.6 Chapter aims

In the previous chapter I identified Cap-G localisation in neuronal nuclei, its partial localisation with HP1 and severe developmental defects upon knockdown. It is plausible to speculate that condensin may play a role in regulating chromatin structure depending on cell type, to aid cell differentiation and effective neuronal development. To address this hypothesis, my aim was to characterise the chromatin binding profile of the condensin I subunit Cap-G. Using Targeted DamID I was able to produce genome-wide DNA binding profiles for Cap-G, as well as chromatin accessibility data for NSCs and neurons. I analysed Cap-G binding and correlated it with known chromatin states, offering an insight on a potential post-mitotic role for Cap-G in neurons. Moreover, I explore the interaction between Cap-G and Lola-N by analysing and comparing their binding profiles in neurons, to understand whether Lola-N is a potential co-factor for Cap-G necessary to drive neuronal maturation. Analysis of the binding profiles reveal that Cap-G binds dynamically across cell types. Cap-G binding is depleted in open and accessible chromatin across all cell types and conversely, is enriched at inaccessible chromatin sites in NSCs. Lola-N and Cap-G show a significant overlap in their binding sites, specifically in mature neurons. Overall this chapter presents data for the binding of the condensin I subunit Cap-G in post-mitotic neurons.

4.2 Results

4.2.1 Profiling cell-specific Cap-G binding in the *Drosophila* CNS

In my observations from the previous chapter, Cap-G localises to NSCs and neuronal nuclei in the larval CNS. In neurons, Cap-G distribution also resembled foci and partially co-localised with heterochromatin (Chapter 3). Building on this knowledge I hypothesised that Cap-G may associate with chromatin in neurons. Characterisation of Cap-G DNA binding profiles in neurons could further illuminate on a role in chromatin remodelling and gene regulation in the CNS. Moreover, it would be interesting to characterise whether genomic localisation of Cap-G differs between differentiated, post-mitotic neurons compared to mitotic NSCs.

Two biological replicates were performed for each condition and the DamID methyl-PCR required for the TaDa experiments yielded strong amplification for all cell types, indicating that Cap-G-Dam associates with chromatin in post-mitotic cells. Quality control analysis of the biological replicates was carried out to ensure the data would yield reliable results. There is a strong positive correlation between replicates for each of NSCs, elav and nSyb conditions with spearman's correlation coefficient values ~0.9 (Figure 4.4A). This gives confidence in the replicability of the sequencing data obtained for Cap-G binding. Moreover, I analysed how the different conditions studied correlated with each other. There appears to be a mild positive correlation between the different cell types (Figure 4.4B). The poorest correlation (rho = 0.38) is observed between nSyb cells and NSCs, which is to be expected given the drastically different cell stages. The elav cells display a similar correlation for both NSCs and elav (rho = 0.45) which faithfully represents the more intermediate state of elav cells (Figure 4.4B).

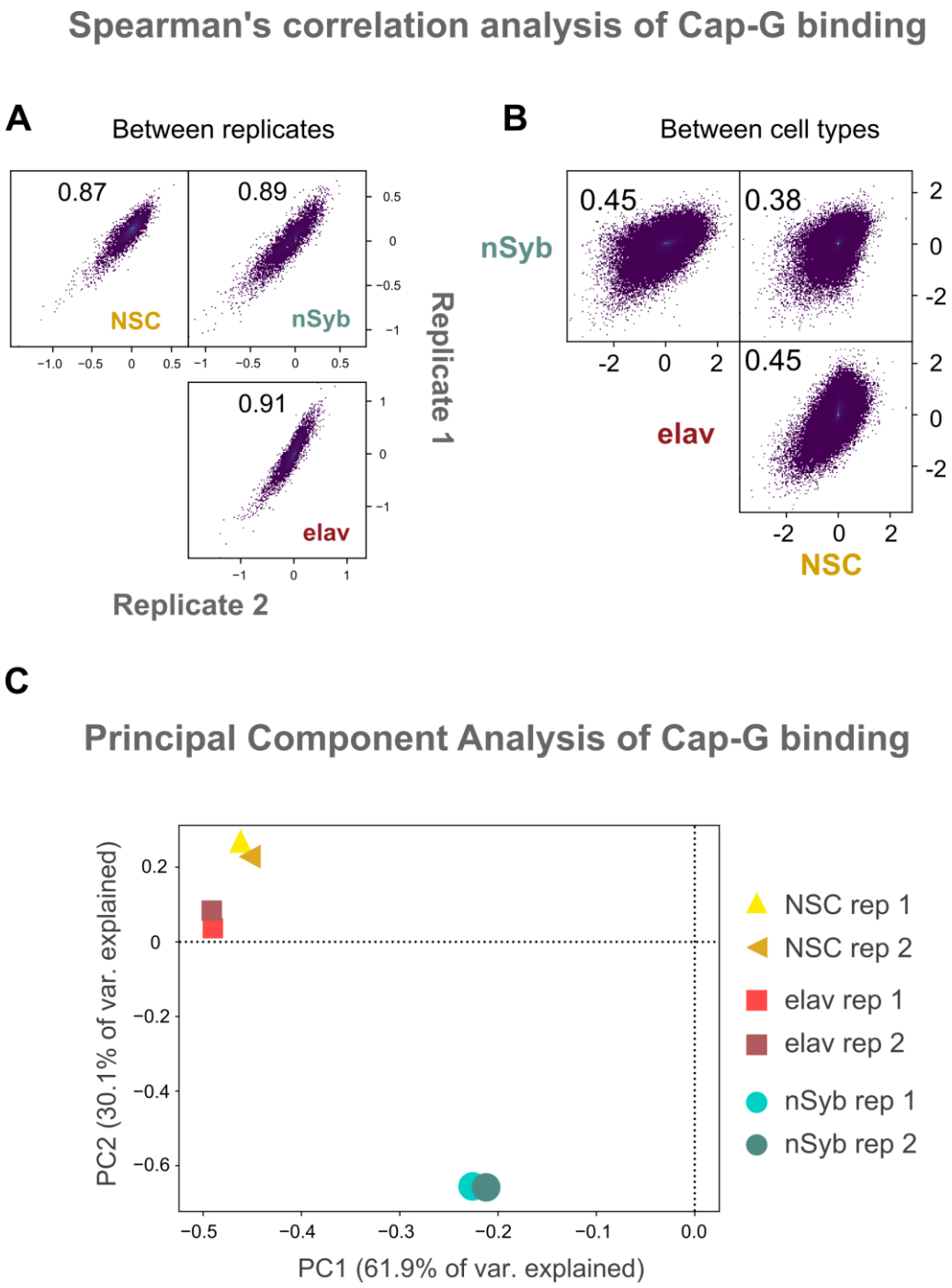


Figure 4.4 Cap-G binding is dynamic across cell types.

Statistical analysis of Cap-G binding in NSCs (NSC), newly born neurons (elav) and mature neurons (nSyb). **A)** Strong positive correlation between replicates for Cap-G mapped binding reads across all cell types (2 per cell type, rho ~ 0.9). **B)** Mild positive correlation between Cap-G binding across all cell types (spearman's rho < 0.5). Lowest correlation observed between NSCs and nSyb neurons (rho= 0.38) whilst elav correlates similarly between NSCs and nSyb neurons (rho= 0.45). **C)** Principal Component Analysis (PCA) shows separate clustering of all cell types, whilst replicates of each cell type cluster together. There is less variance detected between NSC and elav as they cluster in proximity of each other, separate from nSyb.

In order to assess whether the Cap-G binding displays greater variability between experimental conditions than between replicates I carried out Principal Component Analysis (PCA). Differences are represented by separate principal components and are sorted depending on the amount of variability between the populations analysed. PCA of Cap-G binding profiles for NSCs, elav and nSyb displays that biological replicates cluster together (Figure 4.4C), further emphasising similarity between replicates. More importantly, each of the NSCs, elav and nSyb cell types cluster separately and independently from each other (Figure 4.4C). Altogether, this analysis suggests that Cap-G binding varies and dynamically changes depending on the cell type. Generally, Cap-G genomic localisation reflects the differences between each of the NSCs and neuronal cells from different maturation stages.

A peak-calling algorithm identified unique regions significantly bound by Cap-G, present in both replicates for each condition studied (Chapter 2, Section 2.12). Unique Cap-G binding sites dynamically change from NSC to differentiating and mature neurons, suggesting a cell-specific role for Cap-G in the CNS (Figure 4.5). In nSyb neurons, Cap-G peaks are detected at known NSC-specific genes such as *dacapo, deadpan, HLHmγ* and *asense* (Figure 4.5A). No peaks at those genes are detected in NSCs nor elav datasets. Moreover, cell-cycle genes such as *cyclin-E* and *string (stg)* display a similar trend (Figure 4.5B). These genes are known to be actively repressed in neurons (188) as they are necessary for NSC identity and proliferation (15,18,46). Whilst binding of Cap-G is frequently seen in the gene body, in some cases peaks are observed in the surrounding locus as in the case of *stg* in which several peaks are present in the upstream regulatory region (Figure 4.5B).

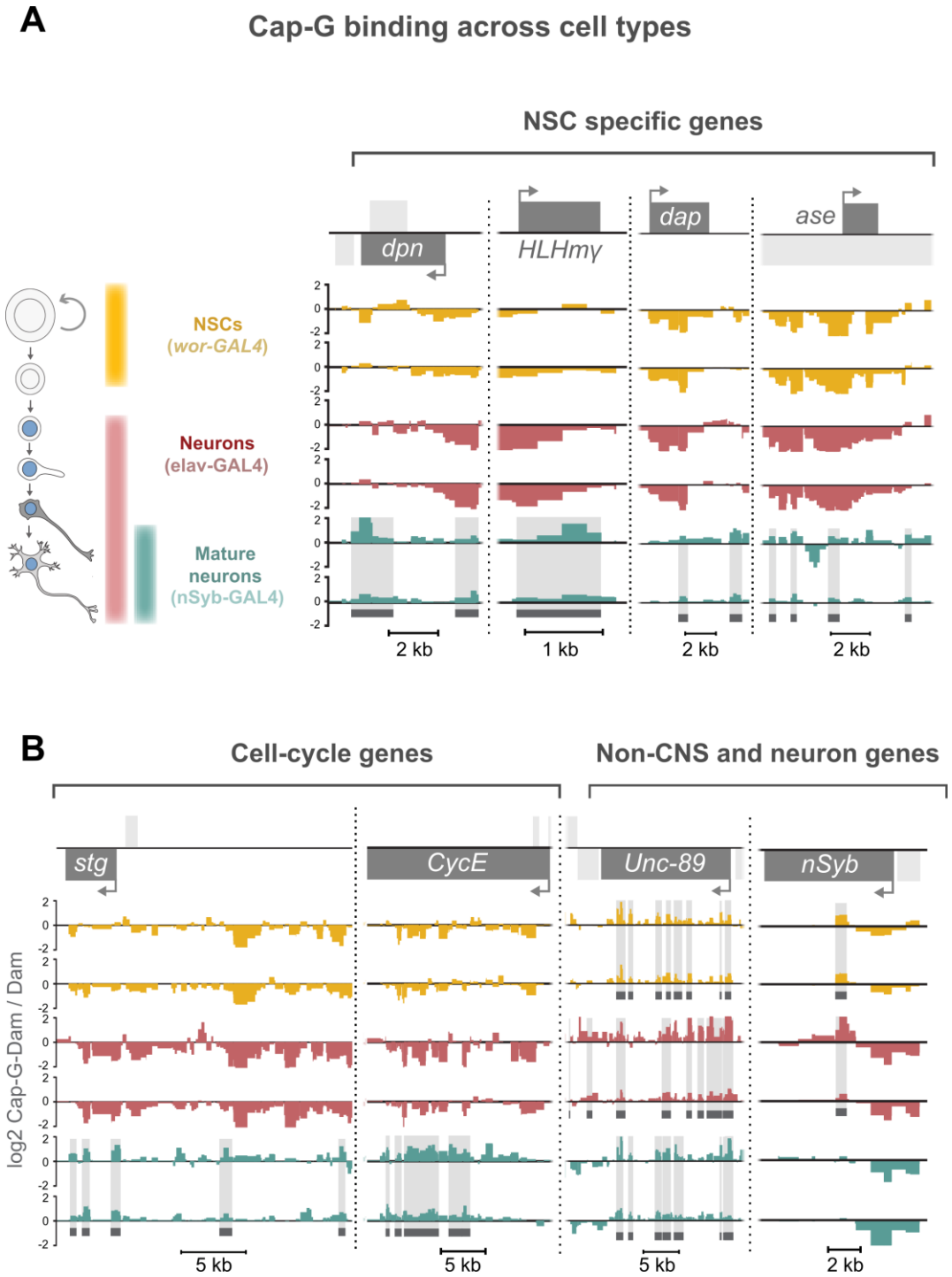


Figure 4.5. Cap-G binds to NSC and cell-cycle genes in neurons.

Cap-G binding at example genes in NSCs, elav and nSyb neurons. The top row shows the gene annotation tracks and light grey boxes represent other genes within the same region. Horizontal dark grey bars that match vertical grey shading over binding indicate statistically significant peaks. Y-axis represents ratio of $\log_2$ Cap-G-Dam/Dam reads. **A)** Cap-G binds at NSC specific genes with peaks detected in nSyb neurons. **B)** Cap-G binds to cell-cycle genes in neurons and to neuronal gene *nSyb* in NSCs. The gene *Unc-89* (not expressed in CNS) shows strong Cap-G binding in all cell types.

In contrast, when looking at neuron-specific genes, such as *nSyb,* peaks are detected in NSC and elav but not in nSyb neurons (Figure 4.5B). Peaks at the *nSyb* gene are also detected in elav neurons as the majority of cells are likely to be in the early, immature stages and therefore have not yet formed synapses. Finally, Cap-G also binds to non-neuronal genes in all cell-types, such as *Unc-89*, which is expressed in muscle tissues, but not the CNS (162) (Figure 4.5B). Interestingly, marked differences are apparent in Cap-G binding between the two post-mitotic cell types, elav and nSyb, suggesting that Cap-G binding in the early life of the neuron differs from more mature neurons. Taken together these data suggest that Cap-G binding is cell-specific and dynamic, likely to localise at repressed genes across cell types.

4.2.2 Cap-G binds to unique genes across cell types

As Cap-G binding appeared in gene bodies and at some upstream sites, I carried out an annotation analysis to associate Cap-G peaks to known genomic features in *Drosophila.* Majority of Cap-G peaks across cell types corresponded to gene bodies (Figure 4.6). About 30-40% peaks were detected at introns and exons but depleted in intergenic and non-coding regions. The second largest distribution of peaks, ~17%, is seen at promoters /transcriptional start sites (TSS) and transcriptional terminal sites (TTS) (Figure 4.6A). Finally, I did not observe any significant peaks at tRNA or rRNA genes, in contrast to previously reported data for condensin binding in *S. cerevisiae* (in which there is only one condensin complex) or condensin II binding in mouse ESCs (87,89).

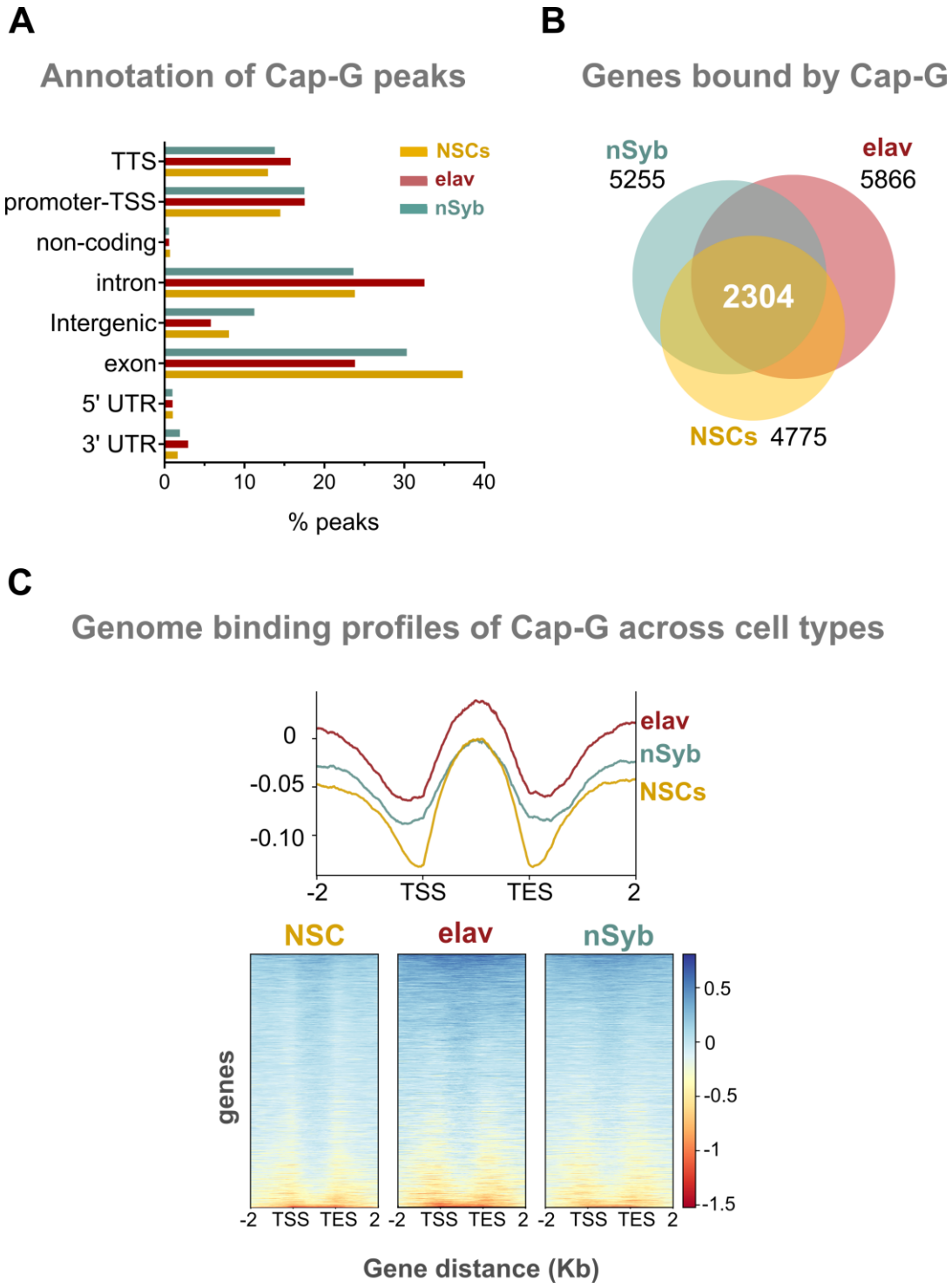


Figure 4.6. Cap-G binds to unique genes and is enriched in gene bodies.

A) Bar chart displaying genomic annotation of Cap-G peaks. Percentage of peaks per feature are displayed and show an enrichment for Cap-G binding in gene bodies (introns and exons) whilst binding at non-coding regions is depleted. **B)** Venn diagram showing unique genes bound by Cap-G and the total overlap between all cell types. **C)** Signal profile and heatmaps of Cap-G binding across the *Drosophila* genes for all three cell-types. Regions show a pseudo-length of 2 kb mid-gene region and up/downstream an anchored TSS and TES. Binding is enriched mid-gene and depleted at TSS and TES.

Upon annotation of peaks I was curious to understand the dynamic binding of Cap-G seen across cell types. Therefore, from the peak-calling analysis I extracted the genes uniquely bound by Cap-G in each of the NSCs, elav and neurons. Overall, the number of genes bound by Cap-G is similar, with 4775, 5866 and 5255 genes identified in NSC, elav and nSyb datasets respectively (Figure 4.6B). There is an overlap of 2304 genes across all cell types and pairwise analysis with Fisher's exact test suggests significant overlap between each cell type (p-value< 1×10^{20}). Despite significant overlaps, Cap-G binds to unique genes in each cell type, with 827, 1138 and 776 genes unique to NSC, elav and nSyb respectively.

In order to characterise the general binding pattern, binding scores for Cap-G were calculated per genomic regions (all known *Drosophila* genes, see Chapter 2, Section 2.12.6) and summarised in a heatmap and signal profile plot. The signal profiles of Cap-G at transcribed regions are similar across NSCs, elav and nSyb cells (Figure 4.6C). Cap-G profiles show a depletion of binding at TSS and TES of *Drosophila* genes, with a mild enrichment mid-gene (Figure 4.6C), in agreement with the previous peak annotation results (Figure 4.6A). The positive binding detected for Cap-G in all cell types has low logarithmic values, which could explain the negative values seen in the profiles. Overall, this data shows that Cap-G binds to unique genes across cell types, and that the binding distribution is enriched in gene bodies.

4.2.3 Gene Ontology of Cap-G bound genes reveal multiple tissue-specific terms

To understand the biological relevance of genes bound by Cap-G, Gene Ontology (GO) terms analysis for Biological Processes was carried out for all three cell types (Figure 4.7). In NSCs, elav and nSyb cells the genes bound by Cap-G relate to a variety of enriched terms, some neuron-specific and other non-tissue specific terms (Figure 4.7). The terms change depending on the cell type, for example, in NSCs some CNS-specific terms such as "Neuromuscular Synaptic Transmission" appear whilst most belong to non-CNS specific tissues, such as wing discs and tracheal system (Figure 4.7). These non-tissue specific terms suggest that Cap-G in NSCs binds to silenced genomic regions, possibly as part of condensin I activity of chromosome compaction in dividing cells. Interestingly, elav-genes are enriched for terms relating to differentiating neurons such as "Neuron projection development" and "Cell morphogenesis in neuron differentiation", indicating that genes bound by Cap-G in elav reflect the early differentiation stage of the cells. Finally, in nSyb neurons, mostly non-CNS specific terms such as "midgut development", "muscle cell differentiation" and "leg disc development" appear. These terms possibly refer to genes that are not tissue-specific and that are probably silenced in the CNS. The variety of terms obtained suggests that Cap-G binding to genes changes as the cell develops, potentially silencing genes not necessary in neurons to aid differentiation and maturation.

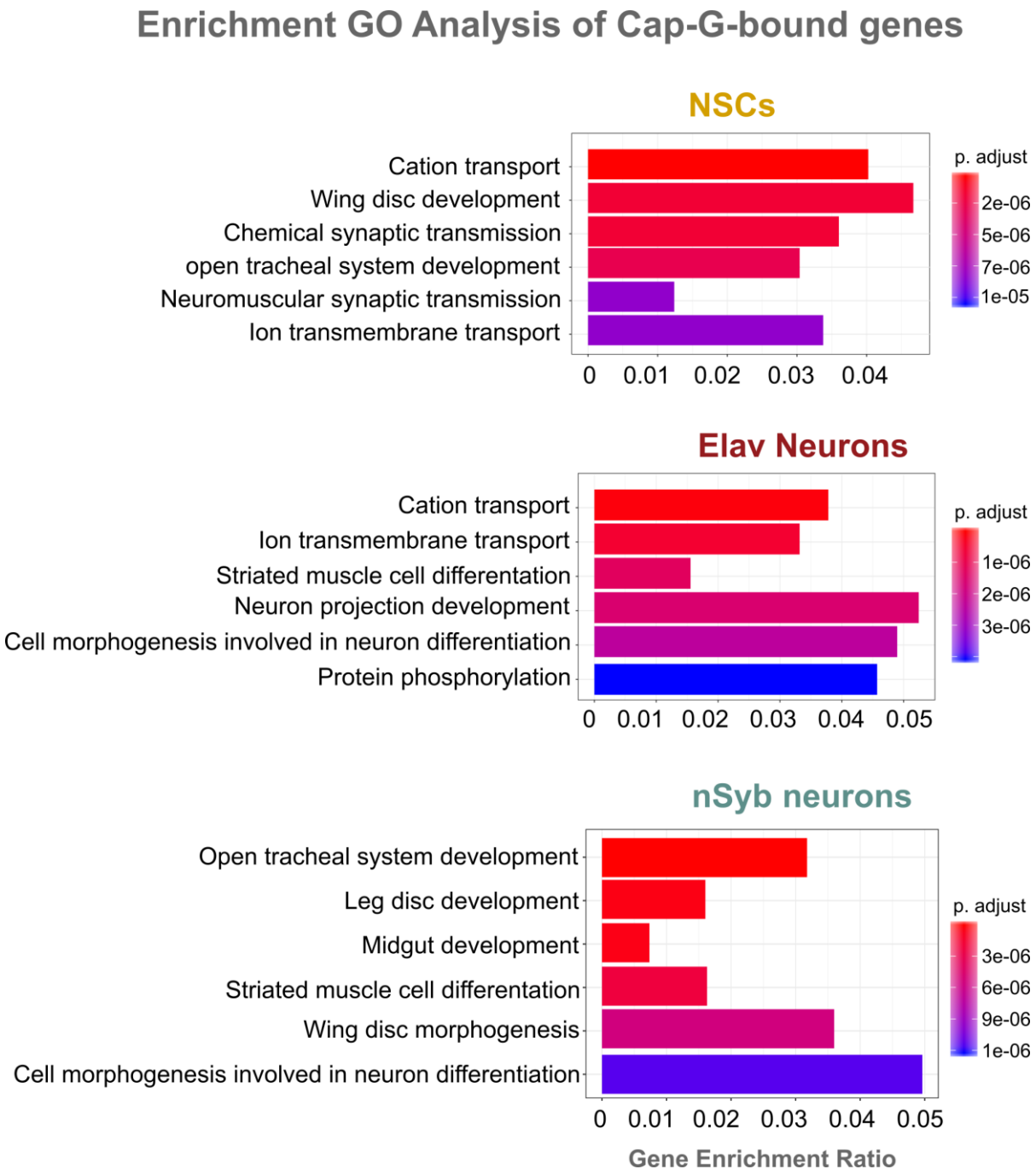


Figure 4.7. Gene ontology enrichment analysis for Cap-G bound genes.

GO terms enriched in gene lists for each of NSCs, elav and nSyb neurons. X-axis represents gene enrichment ratio and colours show the significance by adjusted p-value. Common terms in NSCs and neurons relate to non-CNS specific terms, such as "open tracheal system development" and "Wing disc development" possibly highlighting repressed genes across development. In elav, neuron differentiation terms such as "neuron projection development" reflect genes for the intermediate stage of newly born, differentiating neurons. In nSyb neurons, most enriched terms are non-CNS specific, suggesting Cap-G bound genes belong to repressed loci.

Using DNA motif analysis I identified sequence patterns that were enriched in the datasets and likely have a biological significance. The *de novo* motifs were identified as enriched compared to background sequences, and then were "best matched" to already known transcription factors (TFs) with a p-value cut off for significance at p <1e-12. Unique enriched motifs were identified for Cap-G binding in all cell types (Figure 4.8). In all cell types enriched motifs matched known homeobox genes motifs, such as *abd-A* (189), detected in both NSCs and nSyb neurons, and *Optix* (190) identified in NSCs and elav neurons. In NSCs and nSyb neurons a common motif that matched transcriptional repressor pleiohomeotic (pho), which recruits Polycomb group proteins to chromatin (191) (Figure 4.8). Possibly suggesting involvement of Cap-G in gene repression in NSCs and neurons. In nSyb cells, Cap-G binding motifs matched to TFs such as Giant (gt) and Cut (Ct) which are involved in pattern specification and neuronal identity of peripheral nervous system respectively (192,193) (Figure 4.8). Importantly, motif enrichment analysis can suffer from the low resolution of Targeted DamID, which is limited to GATC motif availability (187). From these motifs it is plausible to speculate that Cap-G has unique co-factor associations, which may mediate gene regulation and vary dependent on the cell type and stage of neuronal maturation.

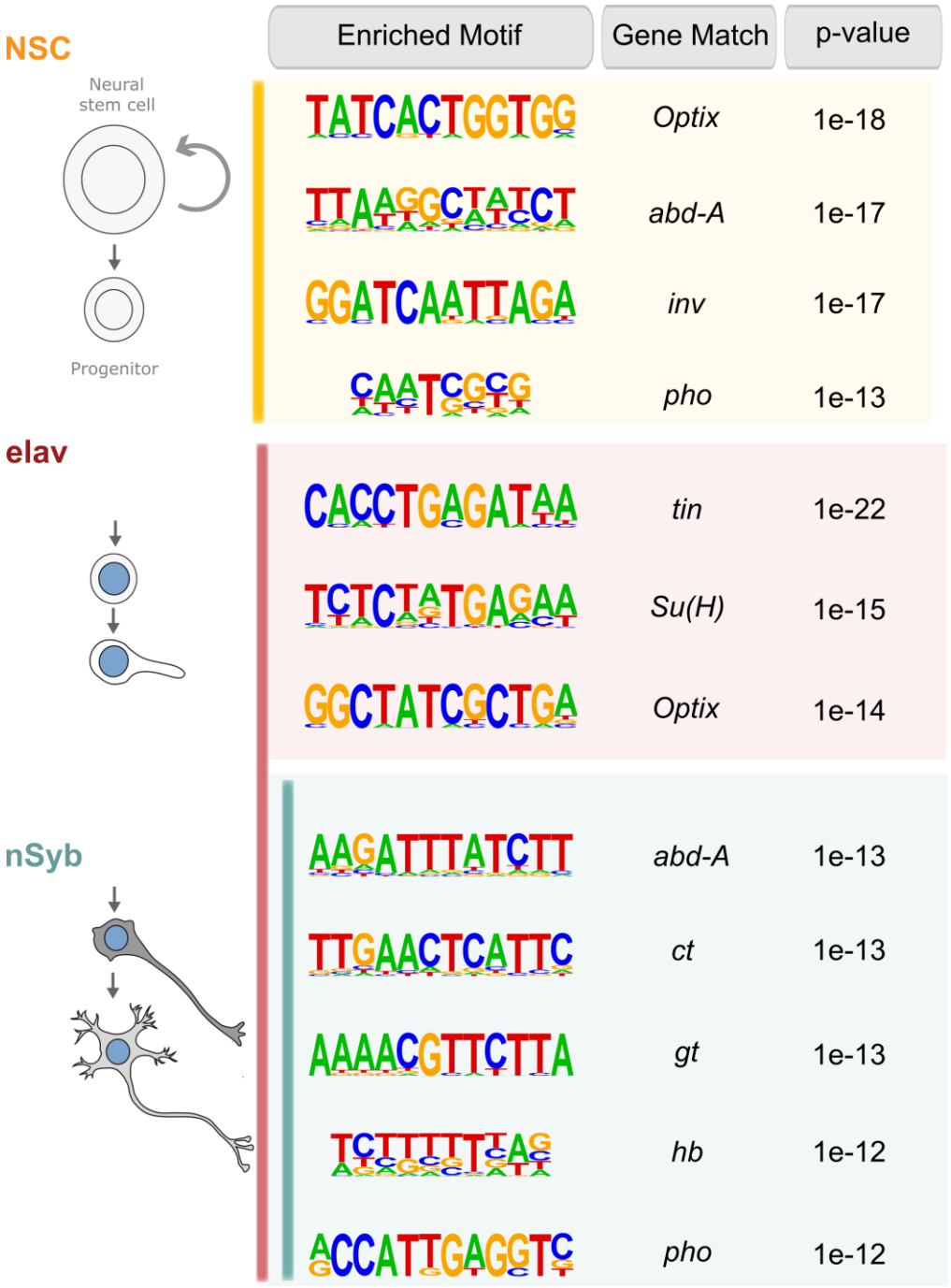


Figure 4.8 Enriched Motif analysis for Cap-G binding.

De novo motif analysis of Cap-G peaks to identify enriched DNA motifs for each of NSCs, elav and nSyb cell types. Enriched motifs were matched to already known motifs relating to *Drosophila* genes, significance detected by p-value cut-off of p < 1e-12. The genes matched to the motifs relate to transcription factors (TFs). In both NSCs and nSyb the transcriptional repressor pho is identified. In nSyb, genes identified are TFs involved in neuronal identity and pattern specification. Motifs extracted from TaDa data require careful consideration due to the low resolution of the data.

4.2.4 Cap-G binding is depleted at accessible chromatin and enriched at repressive chromatin

As a neuron undergoes differentiation, its chromatin state shifts from being accessible and "open" to a more "closed" and inaccessible state (14). The accessible chromatin states correspond to enhancer and active transcription regions. Previous evidence suggests Condensin I and II subunits localise at enhancer regions (85,86,93,94). Therefore, I wished to analyse Cap-G binding in relation to chromatin accessibility in the relevant cell type of interest. The CATaDa method was used, where the reanalysis of control Dam-only data (originally used for Targeted DamID experiments) represents the open chromatin regions in each cell type used (14). The calculated scores for Cap-G binding are plotted over the accessible chromatin (AC) regions. This generates signal profiles to characterise distribution of Cap-G binding at open chromatin regions across all three cell types. The "centre" of each accessible chromatin region is taken as a reference point and the Cap-G binding is plotted for 2 kb either side. Signal profiles show a general depletion of Cap-G binding at accessible chromatin sites (Figure 4.9A). The depletion at AC appears to increase gradually as the cells differentiates, with the strongest depletion observed in nSyb neurons. In the distribution heatmap, a subset of open chromatin regions show positive, enriched binding of Cap-G in all three cell types (Figure 4.9A).

Previous evidence suggests that condensin associates with open chromatin and enhancer regions in *C. elegans* (88). Similarly, the condensin II subunit, Cap-H2 localises at enhancer regions in *Drosophila* Kc167 cells (93). As I observed mild enrichment of Cap-G at some AC sites, I hypothesised that Cap-G could bind to known regulatory sequences in *Drosophila.* To address this I analysed the Cap-G binding distribution at regions from the REDfly database, a collection of *cis*-regulatory elements (CRMs) and transcription factor binding sites (153). Cap-G binding appears to be depleted from known *Drosophila*

regulatory regions across all cell three types (Figure 4.9B). These results suggest that Cap-G binding does not broadly associate with CRMs and is only mildly enriched in some open chromatin regions, suggesting Cap-G may have a gene regulatory role independent of accessible chromatin.

The accessible chromatin identified for each of NSCs, elav and nSyb is further analysed in relation to the Cap-G binding detected for these cell types. There is a negative correlation between the AC and Cap-G binding in all three cell types (Figure 4.10A). Moreover, PCA shows that Cap-G binding clusters separately from AC for each cell type. Interestingly, the accessible chromatin regions from NSCs and elav cluster more closely together, compared to mature nSyb neurons (Figure 4.10B). As PCA displays the variance between populations, a possible explanation may relate to the different chromatin states displayed by immature and mature neurons (14).

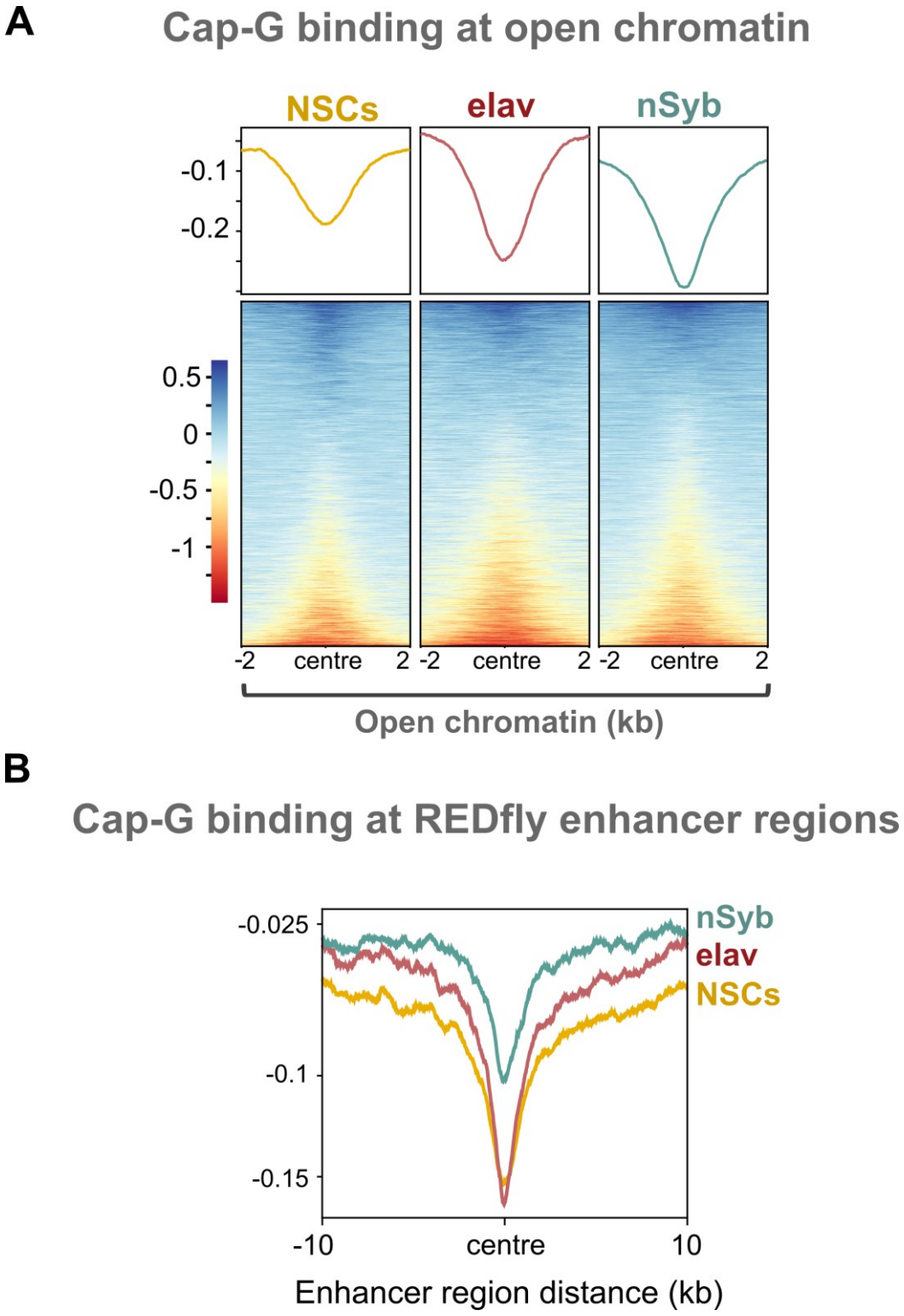


Figure 4.9. Cap-G binding is depleted at enhancers and open chromatin.

A) Average Cap-G binding is depleted at accessible chromatin for all cell types. Plot shows 2 kb up/downstream centre of an open chromatin region. **B)** Cap-G binding is depleted at known *Drosophila* Regulatory Elements (REDfly). Profiles plotted against centre of enhancer region and 10 kb up/downstream.

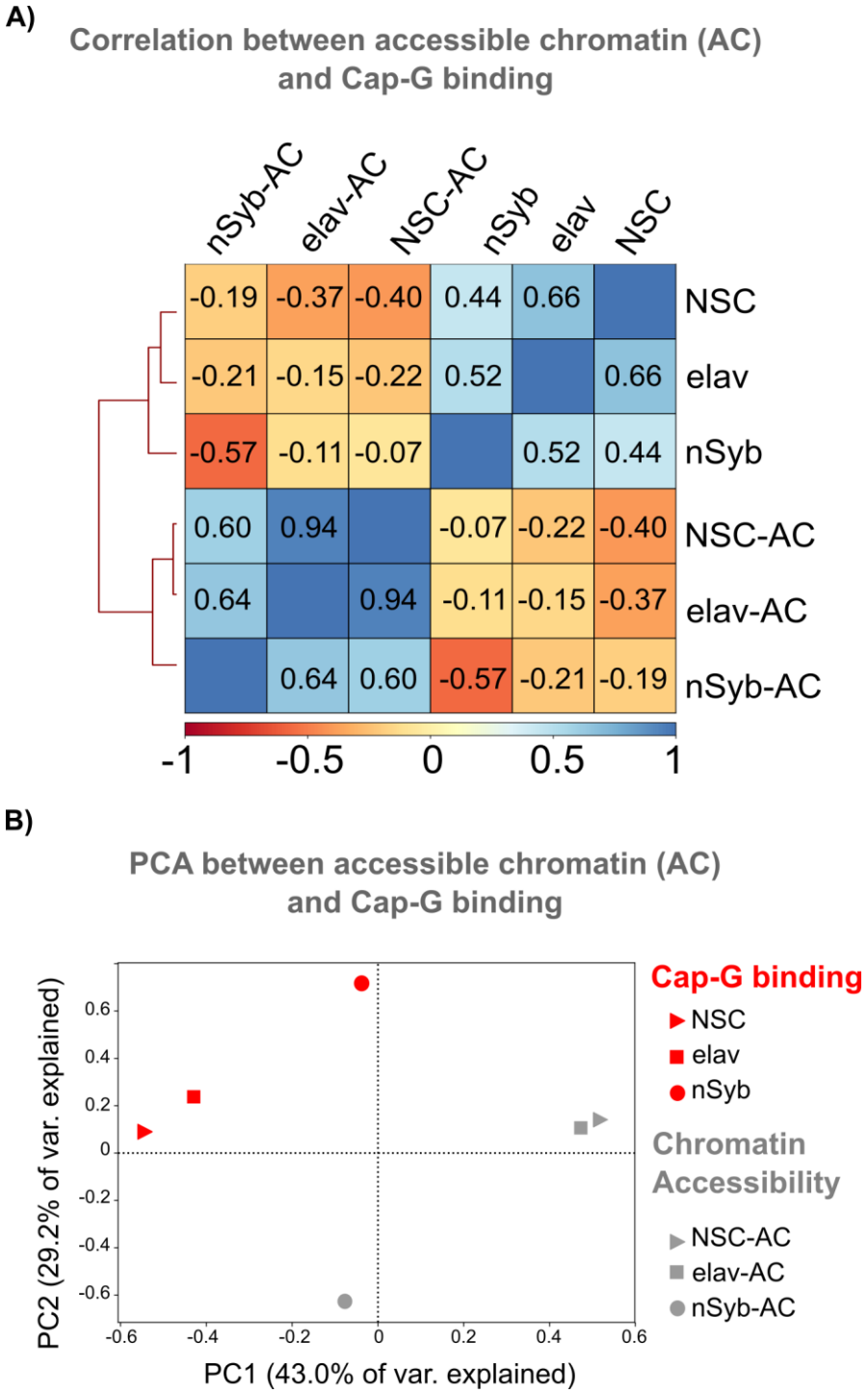


Figure 4.10 Cap-G binding analysis against accessible chromatin.

A) Correlation heatmap between Cap-G binding and Accessible Chromatin (AC) per cell type. Cap-G binding negatively correlates with AC in each cell type. Numbers represent Spearman's rank correlation coefficient. Cap-G binding and AC samples cluster separately by k-means. **B)** PCA of Cap-G binding and AC per cell type. Cap-G binding (red) and AC (grey) cluster separately. Elav and NSC cluster in proximity of each other reflecting their greater similarity, whilst nSyb is more distant, observed in both AC and Cap-G binding samples.

It has been demonstrated that chromatin can be divided into several discrete states (121). These states include repressive chromatin environments such as Polycomb (PcG-repressive), HP1 and "Black" chromatin and permissive states, defined as TrxG, Yellow (non-TrxG) and PcG chromatin (120). As I observed Cap-G mostly depleted at accessible chromatin, I asked whether it could localise to some of these states, offering an insight on Cap-G binding in relation to chromatin. Published Targeted DamID data in *Drosophila* profiled proxy protein binding and was used to define chromatin states in NSCs and adult neurons (7). The published data used *wor-GAL4* for targeting NSCs in larval brains, which matched the cell type used in my experiments. Therefore, I was able to compare Cap-G binding to chromatin states in NSCs.

Cap-G binding profiles showed an enrichment at repressive chromatin states and depletion at permissive states in NSCs (Figure 4.11). Cap-G was particularly strongly associated with 'black' chromatin – a prevalent repressive state that does not incorporate traditional heterochromatin markers. Similarly, Cap-G binding was enriched at PcG repressive state, followed by milder enrichment at TrxG repressive and then HP1 states. In contrast, Cap-G binding was relatively depleted at permissive chromatin states, particularly the 'red' TrxG state (Figure 4.11). Taken together, this analysis suggests that Cap-G in NSCs is mostly enriched at repressive chromatin. Given there was no matching data available for the neuronal cells analysed in these studies, Cap-G binding at chromatin states in neurons could not be profiled. Overall, across NSCs and neurons Cap-G binding is depleted at accessible chromatin sites and enriched at repressive chromatin states in NSCs, suggesting that Cap-G has a potential role in chromatin compaction to induce differentiation.

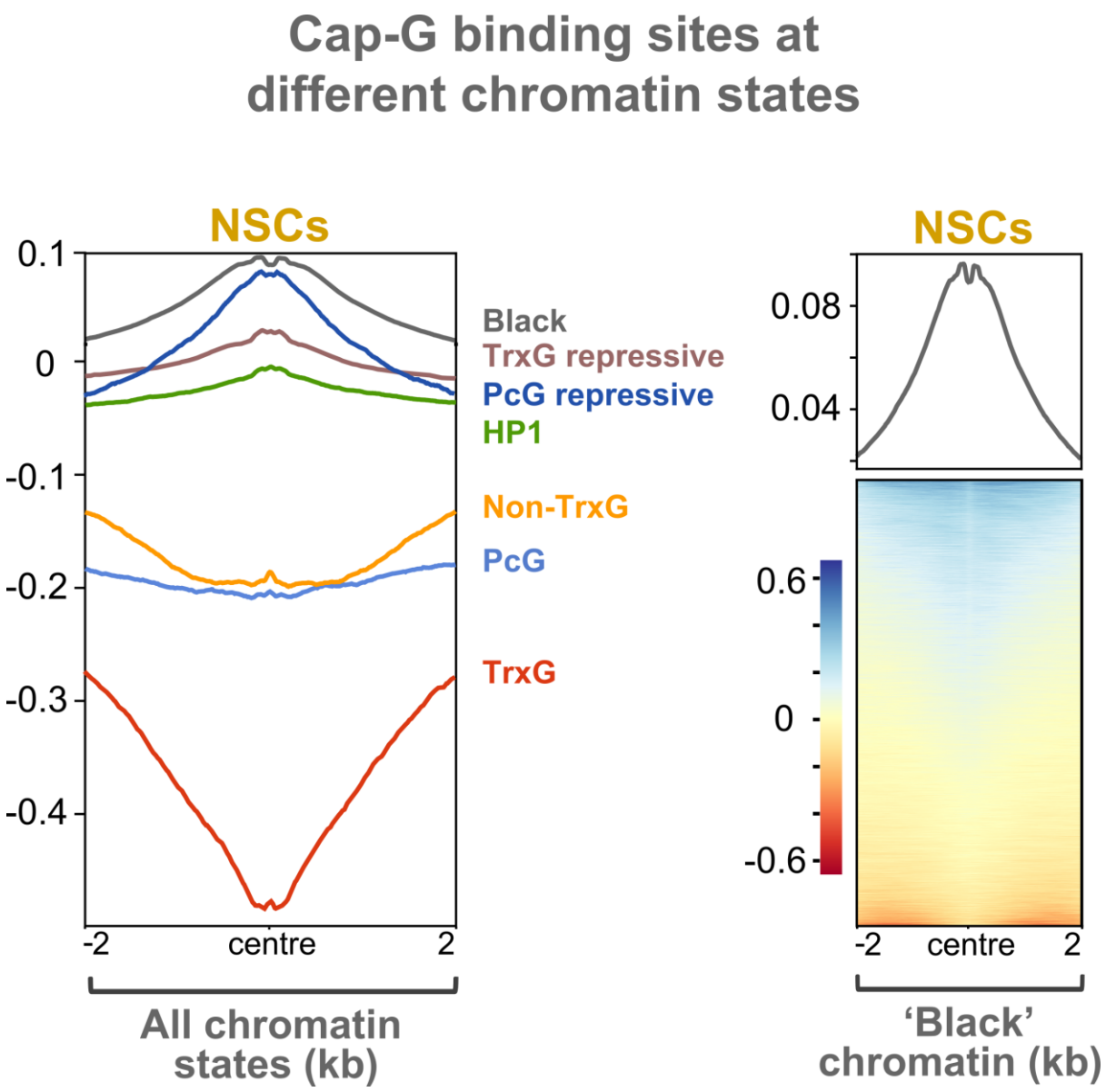


Figure 4.11. Cap-G binding is enriched at repressive chromatin states in NSCs.

Signal profile for Cap-G binding in different chromatin states in NSCs. Cap-G binding is enriched in repressive states (Black, HP1, TrxG-repressive, and PcG repressive), and depleted in permissive chromatin states (non-TrxG, TrxG, and PcG). Heatmap is shown for most highly enriched (black) state.

4.2.5 Cap-G binding overlaps with Lola-N binding sites in mature neurons and is differentiation-dependent

This project was inspired by an interaction detected between the TF Lola-N and Cap-G in a yeast-two-hybrid screen (unpublished, Southall T.D). Lola-N is a transcription factor expressed as soon as the neuron is born and it is crucial to maintain neuronal fate (19). The interaction between a post-mitotic transcription factor and a condensin subunit was unexpected, as there was no prior evidence of a role for condensin in post-mitotic neurons. As I confirmed localisation of Cap-G in neurons, and its co-localisation with Lola-N (Chapter 3, Section 3.2.3), I aimed to characterise whether there were shared binding loci between Lola-N and Cap-G at the genomic level in neurons. Therefore, I generated Lola-N binding profiles in elav and nSyb cells using Targeted DamID. The *UAS-Lt3-Lola-N-Dam* line was used (19) and the same drivers *elav-GAL4* and *nSyb-GAL4* were used to profile Lola-N binding and allow faithful comparisons with the generated Cap-G binding data.

Statistical analysis for quality control of the replicate datasets generated for Lola-N binding in elav and nSyb neurons was carried out (Appendix, Figure 8.3). There was a strong positive correlation between each replicate for elav and nSyb neurons with a Spearman's correlation value of 0.9 and 0.75 respectively (Appendix, Figure 8.3). Moreover, Principal Component Analysis displays a clustering of each replicate, whilst the two different neuronal subtypes cluster separately, suggesting a difference in Lola-N binding between elav and nSyb neurons. This is also observed in correlation analysis between Lola-N binding in elav and nSyb neurons, which displays a negative Spearman's coefficient value of -0.02, suggesting that the two binding datasets do not correlate (Figure 4.12A). Such a drastic change in genomic binding between elav and nSyb neurons

for Lola-N was unexpected, but I continued using these datasets to compare to Cap-G binding.

When comparing Cap-G and Lola-N binding in elav neurons I observed a mild negative correlation, in contrast to a strong positive correlation in nSyb neurons (Figure 4.12A). The poor correlation seen between Cap-G and Lola-N suggests a drastic difference in their genomic binding in immature, elav neurons. In mature neurons, there is a strong linear trend, with spearman's coefficient value of 0.79, suggesting that Cap-G and Lola-N binding is very similar. In agreement, PCA between Cap-G and Lola-N binding shows clustering of nSyb datasets, whilst elav binding for both Lola-N and Cap-G clusters less clearly (Figure 4.12B).

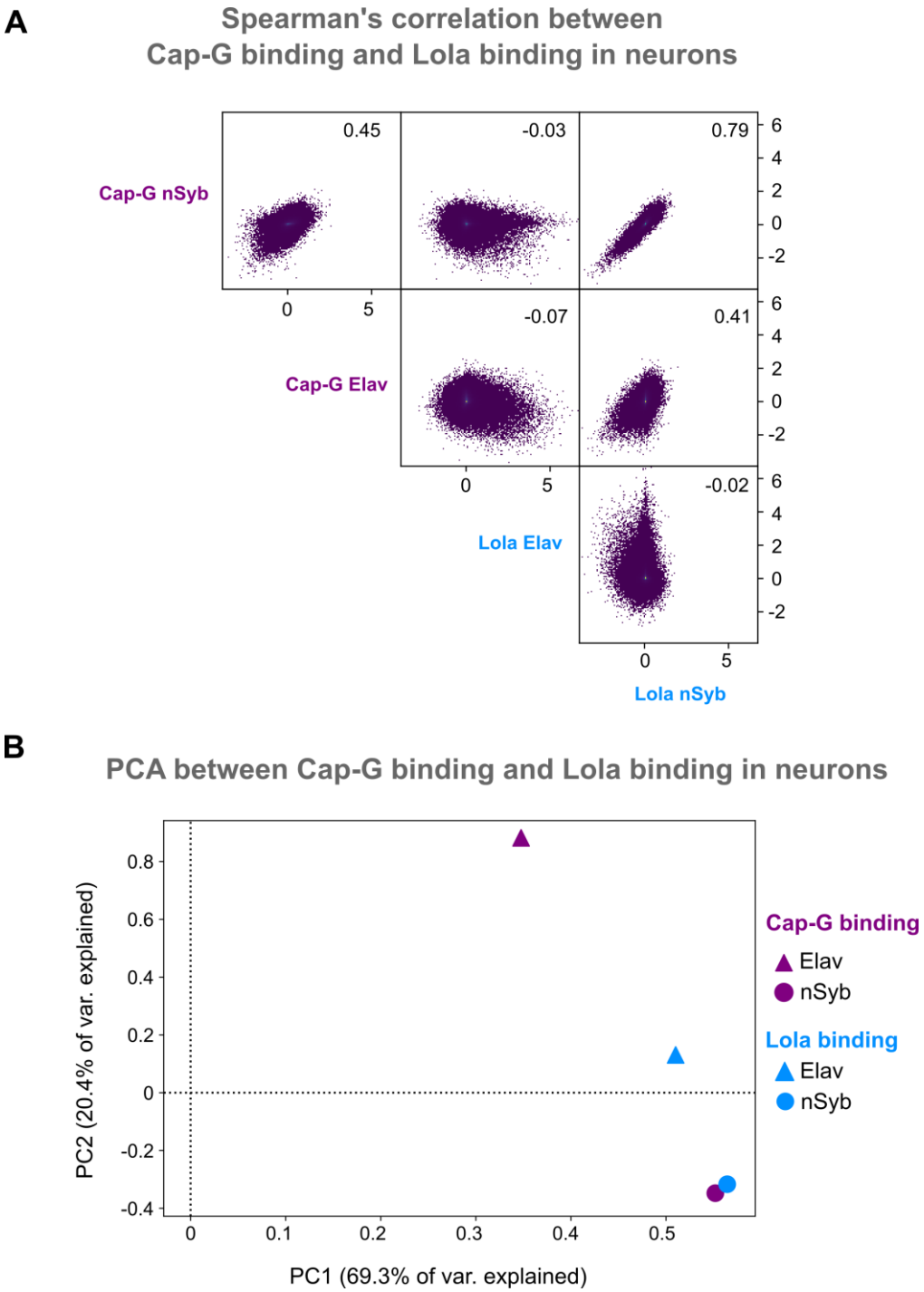
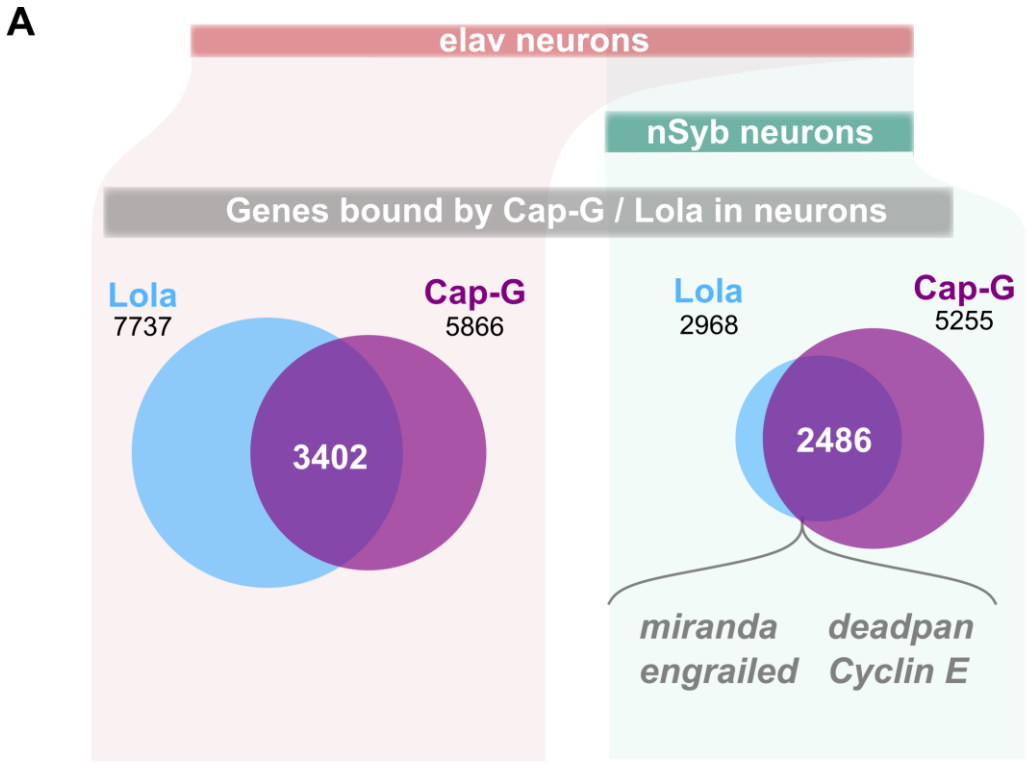


Figure 4.12. Cap-G and Lola binding positively correlates in mature neurons.

A) Correlation scatterplot between Cap-G and Lola binding in elav and nSyb neurons. Cap-G binding negatively correlates with Lola in elav. Strong positive correlation between Lola and Cap-G binding in nSyb data. Numbers represent Spearman's rank correlation coefficient. **B)** PCA of Cap-G / Lola binding per cell type. Cap-G binding (purple) and Lola (blue) cluster separately in elav. In nSyb Cap-G/Lola binding cluster in proximity of each other reflecting their greater similarity.

Overall, there were detectable differences in the number of genes bound by Lola-N and Cap-G in elav and nSyb neurons (Figure 4.13). In elav, Lola-N binds to 7737 genes, compared to 5866 genes bound by Cap-G. Interestingly, Lola-N peaks in nSyb neurons corresponded to far fewer genes, binding to 2968 total genes compared to 7737 in elav neurons. This suggests that there is a reduction in Lola-N targets in more mature neurons, compared to early differentiating, elav neurons. There were 3402 genes commonly bound by Cap-G and Lola-N in elav, compared to 2486 genes in nSyb neurons (Figure 4.13A). There is 1041 genes that remain shared by Cap-G and Lola-N between elav and nSyb neurons. This amounts to 31 % of elav co-bound genes and 42% of nSyb co-bound genes. Therefore, even though there is a reduction in Lola-N targets in more mature neurons, a large proportion of the shared genomic loci between Lola-N and Cap-G are still detected in both cell types. The genes shared between Cap-G and Lola-N in both cell-types were highly significant, fisher's exact test ($p < 0.1 \times 10^{20}$). In nSyb cells, some NSC-specific genes bound by Lola-N and Cap-G were identified, such as *miranda* and *engrailed* (Figure 4.13A). Moreover, cell division genes such as *Cyclin E* were also identified in nSyb cells. These are known Lola-N targets and are actively repressed in neurons (19), suggesting that Cap-G and Lola-N binding may act in concert to supress gene expression.



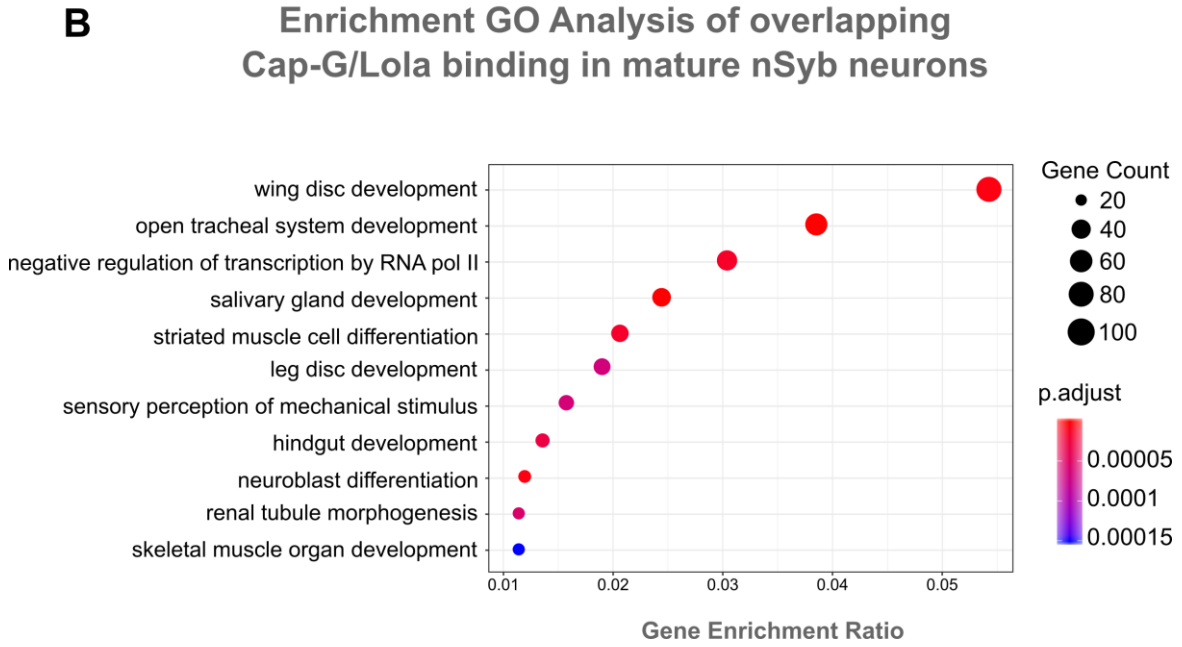


Figure 4.13. Cap-G and Lola binding significantly overlaps in neurons.

A) Venn diagrams representing the number of genes bound by Cap-G and Lola in elav and nSyb neurons, with total number of overlap. Significance of overlap analysed by Fisher's test p-value < 0.1×10^{20}. Example genes bound by both Lola and Cap-G are shown. **B)** Enrichment GO term analysis of genes bound by both Lola and Cap-G in nSyb neurons. GO terms reflect repressed gene regions.

To characterise the biological relevance of the shared genes bound by Cap-G and Lola-N I carried out GO term analysis (Figure 4.13B). Some NSC-specific GO terms were enriched, such as "neuroblast differentiation", which are not expressed in neurons. These terms suggest that Cap-G/Lola-N binding suppresses gene expression. Similar to the GO terms observed for Cap-G binding in nSyb (Figure 4.7), the majority of Cap-G/Lola-N shared genes appear non-CNS specific, belonging to tissues such as the hindgut, muscle and salivary glands (Figure 4.13B). In summary, there is a significant overlap in Lola and Cap-G binding in neurons, with greater similarity observed in the mature, nSyb neurons. These data suggest that Cap-G and Lola repress gene expression to ensure correct differentiation and maturation of neurons.

The similarity in peaks observed and genes called suggest that the binding profiles of Cap-G and Lola-N at all known *Drosophila* genes could also be similar. In elav neurons the Cap-G and Lola-N binding are very different, as Lola binding is enriched at TSS and TES, as expected of a transcription factor (Figure 4.14). Interestingly, as the neuron matures, the Lola-N binding resembles Cap-G binding, where it is depleted at TSS and TES sites, with a mild increase in binding seen mid-gene (Figure 4.14A). This shift in Lola-N binding pattern suggests that it could be a result of neuronal maturation. Thus, it was interesting to investigate whether Cap-G binding at Lola-N specific sites was also dependent on differentiation of the cell. Cap-G binding distribution at Lola peaks is enriched in neurons when compared to NSCs (Figure 4.14B). This suggests that binding of Cap-G at Lola-N-bound genomic sites increases as a result of cell differentiation from NSCs to neurons. Cap-G and Lola-N interaction at genomic regions is more prominent in mature neurons and appears differentiation-dependent, suggesting that Cap-G/Lola-N may interact to regulate neuronal maturation.

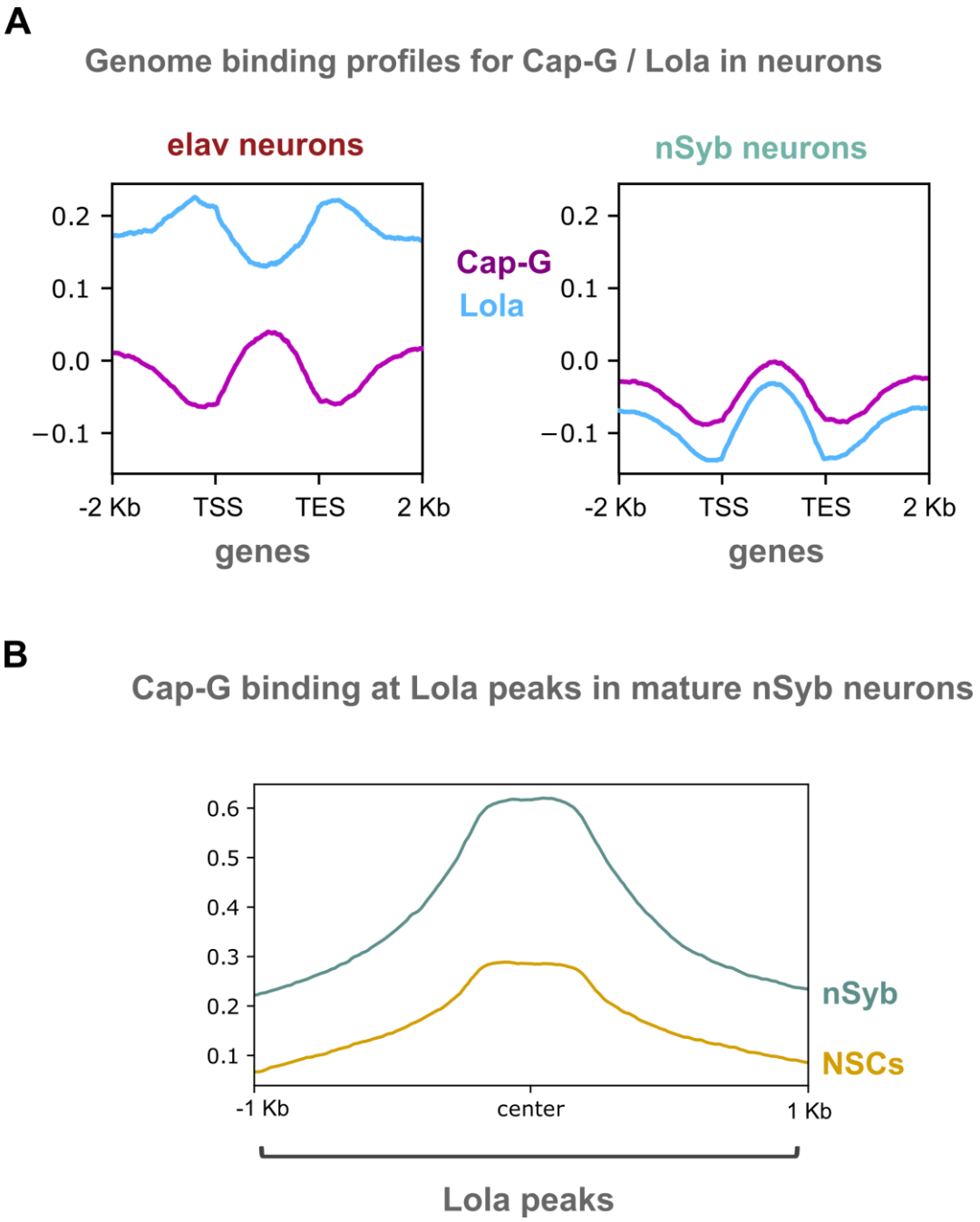


Figure 4.14. Cap-G binding at Lola peaks is differentiation-dependent.

A) Signal binding profiles at known *Drosophila* genes for Cap-G and Lola in elav and nSyb neurons. Cap-G and Lola binding profiles display reversed pattern in elav cells, with Lola enriched at TSS and TES in contrast to Cap-G which is depleted. In nSyb Cap-G/Lola binding profile patterns overlap, with enrichment in gene bodies. **B)** Analysis of Cap-G binding profiles at known Lola peaks. Low enrichment detected in NSCs, followed by high enrichment in nSyb neurons, suggesting localisation of Cap-G at Lola peaks is dependent on differentiation.

4.3 Discussion

In this chapter I characterised the binding profiles for Cap-G in NSCs, newly-born and mature neurons. Using the TaDa technique I generated cell-specific binding profiles for Cap-G and compared them with accessible chromatin data in all three cell types. This is the first evidence for condensin binding profiling in post-mitotic cells, providing exciting new insights into a role for Cap-G in neurons.

4.3.1 Cap-G binding profiles in NSCs and neurons

The Cap-G binding profiles obtained show dynamic changes in binding in NSCs, immature and mature neurons. The data from these three cell types clustered separately, though it still showed positive correlations. These suggest that there are similarities in the binding data between cell types. These results are somewhat surprising as the cell types analysed are very different, from mitotic NSCs to post-mitotic newly differentiated neurons and mature neurons. However, a previous study in yeast suggests that condensin binding profiles remain similar between dividing and interphase cells (87). A similar phenomenon may occur *in vivo* in *Drosophila,* as I also observed similar Cap-G profile binding patterns between all cell types. The annotation of peaks also shows an enrichment in Cap-G binding at intron and exons in all cell types analysed. This is in agreement with previous ChIP-seq data for Cap-G in human cancer cells, where binding is enriched at intergenic and intronic regions (94), although I detected limited binding at intergenic regions in my datasets. In chicken DT40 cells condensin I subunits Cap-H (Barren) and SMC2 displayed 24% and 31% of peaks in bodies of transcribed genes respectively (86). These percentages are similar to the ones observed for intron/exon peak distribution of Cap-G across NSCs and neurons. A subunit of the condensin II complex, Cap-H2, also localised to gene bodies of elongating genes (194). This evidence

suggests that condensin binding can therefore be enriched at gene bodies and may be dependent on the cell type analysed.

To date, condensin binding has been profiled using mostly ChIP-seq methodology (Table 4.1). The use of ChIP-seq as a technique for analysing binding profiles of chromatin-associating SMC-complexes, like condensins, has been disputed in the literature (61). In this study I used TaDa as the technique of choice to profile Cap-G binding. There are some advantages and disadvantages in choosing TaDa over ChIP for this particular study. Firstly, using TaDa allowed me to profile cell-specific binding *in vivo* without the need to isolate cells (184). The cell-specificity of TaDa also allowed me to profile binding in NSCs as well as two distinct neuronal stages, observing how the patterns change across neuron development and maturation. In ChIP-seq, individual cell populations are required and the sample needs to be fixed (195), offering a "frozen in time/snapshot" of the chromatin at the time of collection. In contrast, TaDa doesn't involve sample fixation but rather collects binding profiles *in vivo* from a 16-hrs expression period, offering a more representative picture of the chromatin-association pattern of the protein of interest (184). Moreover, ChIP relies on the specificity of antibodies against the proteins of interest, whilst TaDa uses the chromatin-associating protein itself to detect the binding profiles, bypassing specificity-related issues. However, a limitation of TaDa is that it has lower resolution than ChIP-seq, as it depends on the distribution and availability of GATC sites along the genome (184). Overall, the use of TaDa allowed to generate novel, cell-specific and robust genomic binding profiles for Cap-G in post-mitotic neurons.

Both techniques need to be carefully considered in the context of chromatin-remodelling ability of condensin. The condensin complex creates bridges and DNA loops, so what may appear as individual binding loci may in fact be genomic regions that are bridged closely

together by condensin. Therefore, the condensin-dependent chromatin folding and 3D conformation may be reflected in the protein-binding profiles. This hypothesis has been suggested for ChIP-seq maps (61) and it is also possibly true for TaDa-generated maps. It would be interesting to compare binding profiles generated with Hi-C (chromatin conformation capture experiments) to dissect the individual binding of condensin from 3D genomic interactions.

Annotation of Cap-G peaks revealed mild enrichment at promoter-TSS sites in NSCs, elav and nSyb neurons. The promoter-TSS sites are usually associated with RNA Polymerases (RNAP), and multiple studies observe condensin at these sites (86-89). In chicken DT40 cells condensin I localises at RNA Pol I regions (86). In yeast and mammalian cells, condensin co-localises with RNA Pol II binding (84,85,87). However, recent evidence from *Drosophila* Kc167 cells suggests that RNA Pol II distribution is independent of Cap-H2 (194). The lack of enrichment of Cap-G peaks at TSS in my dataset proposes that there is no overlap between RNAP and Cap-G binding sites. A recent study suggests that as condensin translocates along the DNA to form loops, it encounters a "barrier" of RNAP machinery and is stalled at those regions (92). It is then plausible that condensin accumulates at RNAP barriers, and that the binding detected is an artefact resulting from the sample fixation required in ChIP-seq. However, comparisons between Cap-G and RNAP binding in NSCs and neurons would be necessary to definitively rule out any overlaps.

There is a wealth of evidence supporting co-localisation of condensin subunits with TFIIIC binding sites, such as tRNAs, in multiple species (86-89). In *Drosophila* Kc167 cells, condensin I subunit Barren and condensin II subunit Cap-H2 localise at tRNAs loci (90). This evidence has been obtained mostly from ChIP-seq methods (Table 4.1) and in yeast

there have been reports of false-positive signals from ChIP-seq maps that suggest protein-binding with tRNAs regions and highly-transcribed genes (196). In my data analysis I was unable to detect any overlap between Cap-G peaks and known TFIIIC genomic regions. Taking all this into account, the observations of condensin binding at RNAP sites need to be carefully considered. As in my study I use a different methodology, which does not suffer from some ChIP-mapping pitfalls, it may more accurately represent condensin binding. However, this study also focuses on specialised, post-mitotic cell types, not directly comparable to previous studies. Therefore, it is plausible that Cap-G binding in post-mitotic neurons exhibits different patterns than those previously observed in mitotic and interphase cells.

I observed depletion of Cap-G binding at open chromatin regions across all three cell types. Some signal was detected at some accessible chromatin, but these did not match any known *Drosophila* enhancers. These findings disagree with previously published data. There is evidence that condensin II subunit Cap-H2 binds to enhancer/promoter regions in murine and chicken cells (85,86,93). However, there is limited evidence for condensin I binding at enhancer regions. In *C. elegans,* condensin I is reported to localise to TSS of actively expressed genes (88). Conversely, Cap-G does not localise to enhancer regions in mitotic human cancer cells (94). This agrees with my findings of depleted Cap-G at enhancers in NSCs. As there is scarce evidence for condensin I at enhancer regions this may hint at different roles between the condensin I and II complex in active transcription. Furthermore, the enhancer localisation may be dictated by the presence of co-factors or dependent on the cell type, therefore it may not be necessary in the post-mitotic cells.

Interestingly, across cell-types Cap-G appeared to bind to actively repressed genes as well as non-tissue specific genes. Moreover, the binding dynamically changed between NSCs and neurons, and the binding was specific to the cell type observed. In yeast cells, as they transition from a log phase (highly mitotic) to a quiescent state, there is an increase in Cap-H binding to quiescent specific loci (97). Therefore, condensin binding can dynamically change as the cell progresses to a different stage. Cap-G appears to directly bind to NSC and cell-cycle specific genes in neurons, which are actively silenced. There is evidence that direct condensin binding leads to gene repression, for example in *C. elegans* the DCC complex (related to condensin I) binds to specific loci in the genome to induce repression of X-linked genes in hermaphrodites (64). Moreover, condensin binding is associated with direct repressive effects in *C. elegans* and yeast (88,97). Overall, condensin binding is observed across species, however the evidence in this study is the first relating to post-mitotic cells, suggesting that this feature of condensin may be maintained in different cell stage and cell types.

4.3.2 Cap-G binding in NSCs and neurons suggests a role in chromatin organisation

Chromatin topology describes the 3D genome organisation in nuclei. The topologically associated domains (TADs) are a structural feature which encompass genomic loci that interact with each other at high frequency, such as enhancers (183). There is a poor understanding of the role for condensin I in TAD formation. Most of the evidence is for condensin II, where Cap-H2 is involved in TAD formation and maintenance in *Drosophila* (90,93). As I did not observe Cap-G localising at specific enhancer regions it may not associate with TAD regions nor regulate interactions within the domains in *Drosophila* neurons. However, the data obtained in this study is not sufficient to draw clear

conclusions on the role of condensin in TAD formation, and further research is necessary to explore this relationship in neurons.

Chromatin states is a way of describing the proteins and histone modifications associated with genes that are expressed or silenced. Therefore, these dynamically change between cell stages and different cell types. To understand whether Cap-G is involved or associates with repressive/ permissive chromatin states I compared my NSCs data to previously published work (7). Cap-G binding was strongly enriched at repressive chromatin states and depleted at permissive chromatin sates. Unfortunately, due to the lack of corresponding chromatin state data for larval neurons I was unable to compare the chromatin states and Cap-G binding for those cell types. It would be interesting to generate chromatin states data for the neuronal cells used in this study, to compare it to the Cap-G biding. Nonetheless, given the depletion at open chromatin and the overall similarities between the cell types, it can be assumed that Cap-G binding is also enriched at repressive chromatin in neurons.

The TADs are sub-structures found within larger structures, known as chromatin compartments which segregate the genome into active/inactive regions. The chromatin compartments encompass active (A) chromatin and inactive (B) genomic regions, which are able to interact with each other (117). Recently, there is emerging evidence of smaller chromatin compartments, described as compartmental domains. In *Drosophila,* these compartmental domains A have been described to be similar in size to regions corresponding to active chromatin states (123). Therefore there is a suggested relationship between the active/inactive chromatin compartments and the chromatin states (117). In my data Cap-G binding was depleted at accessible chromatin and enriched at repressive chromatin states, so it is plausible to assume that condensin, as an

architectural protein complex, may have a role in regulating large-scale chromatin topology changes. There is little known about the proteins involved in formation of active/ inactive chromatin compartments. From my data I speculate that Cap-G may have a role in regulating chromatin compartmental domains, to drive differentiation by segregating genomic regions of heterochromatin. However, further research on Cap-G's role in chromatin topology are needed to elucidate on this hypothesis.

The binding at repressive chromatin offers interesting insights on a potential Cap-G role in neurons. In NSCs, Cap-G is most likely associated with the rest of the condensin complex to drive chromosome compaction during cell division (197). In this instance chromatin compaction may work in tandem with gene silencing, hence the Cap-G localisation at repressive chromatin is logical. Condensin-mediated chromatin compaction drives cell silencing in murine and yeast cells, maintaining the quiescent cell state (97,99). Therefore, in my data, Cap-G localisation at repressive chromatin states may suggest that Cap-G facilitates chromatin compaction, to induce differentiation by silencing gene expression.

4.3.3 Cap-G and Lola-N interact to maintain neuronal fate

Investigating Cap-G in neurons stemmed from preliminary Y2H data that identified an interaction between the neuron-specific TF Lola-N and Cap-G (unpublished, Southall T.D.). To further investigate this interaction, I analysed and compared the Cap-G and Lola-N binding profiles in neurons. There was limited similarity between Cap-G and Lola-N binding in elav-cells, with negative correlation of samples and reciprocal patterns of their binding profiles across *Drosophila* genes. Cap-G depletion from newly born neurons had more severe effects on fly development than depletion in mature neurons (Chapter 3). In newly born neurons Cap-G may have a role independent of Lola-N, still relevant to

neuronal differentiation. However, for the mature nSyb neurons, the Lola-N average binding pattern resembled that of Cap-G, with an enrichment detected in gene bodies. Moreover, the genes bound by Lola-N and Cap-G in mature neurons significantly overlapped, including known NSC and cell-cycle specific genes. Lola-N actively represses said genes, as newly born neurons in *lola-N* mutants dedifferentiate to create stem-cell like, proliferative cells that result in tumours (19). Therefore, since Cap-G and Lola-N share binding at these repressed genes, it suggests a role for Cap-G in maintenance of neuronal fate by gene repression.

Localisation of Cap-G and Lola-N at shared genomic loci suggests that they interact to regulate gene expression. There were apparent differences in binding between elav and nSyb neurons, with larger similarity between Lola-N and Cap-G detected in nSyb neurons. Lola-N is crucial in newly born neurons, a small window of development before neuronal maturation, as depletion of Lola-N from mature neurons has no dedifferentiation effects (19). A potential hypothesis is that, as the neuron matures, Lola-N recruits Cap-G to its binding sites to drive chromatin compaction and hence more permanently silence genes. This is also suggested by my observations that Cap-G binding at Lola-N sites increases as the cell shifts from NSCs to mature nSyb neurons. In yeast, as the cells enter quiescence, Cap-H2 is directed to specific loci by TF Msn2 to induce gene silencing (97). Moreover in *Drosophila* S2 cells, the DNA-binding protein Mrg15 regulates the level of Cap-H2 at chromatin sites and their interaction is crucial in maintaining chromosome compaction and function (96). Finally, Lola-N may direct Cap-G to its binding loci to drive gene silencing and induce effective neuronal maturation.

There is evidence in the literature for the recruitment of condensin to specific loci by cofactor proteins, leading to gene expression and development. For example, in human

cancer cells, Cap-G interacts with the DNA binding domain of TF ER- α, which results in loading of Cap-G to specific genomic regions upon estrogen stimuli (94). Also in *C. elegans,* condensin II binding sites largely overlap with those of TF LIN-13, needed for vulval development (88,95). My motif analysis for Cap-G binding identified different sets of motifs in the profiled cell types, some of which are known to be involved in transcriptional repression and specification of neuronal identity, such as *pho* (191). Overall, other factors could be recruiting Cap-G to specific loci to drive gene repression in neurons and to infer a cell-specific role for Cap-G in neurons.

Chapter 5 Cap-G knockdown leads to severe disruption of neuronal transcriptomes

5.1 Introduction

5.1.1 Gene expression regulation by the condensin complex

The condensin complex has been well studied in mitotic cells, however, there is evidence arising that condensin does more than just package DNA during mitosis (60). Importantly, the condensin complex has been associated with regulation of gene expression across a number of species (Table 5.1).

Condensin can manipulate chromatin into a compact state, resulting in silencing of gene expression (Table 5.1). For example, condensin-dependent chromatin compaction is a conserved feature of *S. cerevisiae* and human fibroblast quiescent cells (97). In yeast quiescent cells, mutations to the SMC4 subunit resulted in chromatin decompaction and in gene de-repression, re-activating transcription (97). Condensin II has also been associated with maintenance of a quiescent cell state in mice (99). In naive T-cells, mutations to Cap-H2 resulted in failed silencing of gene expression thereby losing the quiescent cell state (99).

In some instances the condensin complex can regulate and manipulate genomic region interactions to affect gene expression. For example, in yeast, condensin can repress transcription by clustering rDNA and tRNA regions, which leads to silencing of adjacent RNA pol II promoters (91,104-107). In *Drosophila,* a process known as transvection is when homologous allelic regions interact to regulate transcription (102). Condensin II negatively regulates transvection to ultimately repress gene expression (101). Moreover, Cap-G seems to have a repressive effect on gene expression by regulating

heterochromatin structures in *Drosophila* (81). Examples from *C. elegans* also support a silencing role for condensin. For example, the dosage compensation complex (DCC), closely related to condensin I, binds to specific DNA regions resulting in gene repression (64,100). Furthermore, condensin II binding is suggested to directly repress genes in *C. elegans* (88). Disruption of condensin does not exclusively lead to a loss of repressive state, but can also activate gene expression. In *Drosophila* SG4 cells, knockdown of Cap-D3 (condensin II) results in changes to the chromatin structure adjacent to retrotransposon regions, which in turn affect the expression of neighbouring genes (181). Moreover, loss of Cap-D3 resulted in increased methylation of the retrotransposon loci leading to an increase in transcription of nearby genes (181).

The interaction of condensin with other proteins and complexes can also result in gene expression changes. In *Drosophila,* the chromatin-repressing Polycomb complex silences homeotic genes via its interaction with condensin I subunit Barren (103). In addition, Cap-G appears to interact with TF Estrogen Receptor α (ER- α) in human cancer cells to promote expression of estrogen-target genes (94). Condensin interactions with other complexes and factors still remain largely unexplored and could provide further evidence to cell-specific, gene regulatory roles.

There are also cell-specific roles for condensin in regulating gene expression, where it impacts cell identity and transcription specificity. For example, in mouse erythroid cells, condensin II subunit Cap-G2 silences proliferation-specific genes, inducing cell differentiation (82). Furthermore, cell-specific genes and housekeeping genes are activated by condensin II in mouse ESCs and human embryonic kidney cells (85,89).

Table 5.1 Condensin roles in gene expression regulation

Condensin I			
Subunit	***Species/ cell lines***	***Gene Regulation roles***	***References***
Cap-G	human cancer cells	promote transcription activation of estrogen-target genes	(94)
Condensin I	*S. cerevisiae*	Clustering rDNA and tRNA to silence RNA Pol II transcription	(91,104)
Cap-G	*D. melanogaster*	heterochromatin gene expression	(81)
Barren	*D. melanogaster*	Interaction with chromatin-repressing Polycomb complex to silence homeotic genes	(103)
SMC4	*S. cerevisiae*	Condensin-dependent chromatin compaction represses transcription	(97)
DCC complex	*C. elegans*	Transcriptional repression of X-linked genes	(64)
Condensin II			
Cap-H2	mouse T-cells	Silence IL-2 target genes by chromatin compaction to maintain quiescence	(99)
Cap-H2, Cap-D3, Cap-G2	*C elegans*	repressive function	(88)
Cap-H2	Mouse ESCs	activate gene expression of ESC genes	(85)
Cap-H2	Human embryonic kidney cells	Activate gene expression of housekeeping genes	(89)
Cap-H2, Cap-D3, Cap-G2	*C. elegans*	Direct binding leads to repressive effects	(88)
Cap-H2	*D. melanogaster*	Negative regulation of transvection leads to gene silencing	(101)
Cap-D3	*D. melanogaster SG4 cells*	Regulate chromatin structure around retrotransposon loci to maintain neighbouring gene expression	(181)
Cap-H2	Mouse post-mitotic hepatocytes	Mild effects on gene expression	(110)
Cap-G and Cap-D2	*S. cerevisiae*	No changes to gene expression upon loss of condensin	(109)
Cap-D3	*D. melanogaster* post-mitotic fat body cells	Activate transcription of AMP gene clusters	(98)

5.1.2 Contradictory evidence for gene expression by condensin and gaps in our knowledge

The literature described above supports a role for condensin in regulation of gene expression, however, recent studies contradict this view. For example, studies in yeast, and mouse hepatocytes argue that condensin depletion has no impact on overall gene expression (109,110). In *S. cerevisiae*, mutations to condensin I subunits Cap-G and Cap-D2 resulted in global chromatin decompaction but with insignificant effects on gene expression (109). Moreover, the authors report that they did not observe direct correlation of condensin binding and gene silencing (109). A recent study describes the effects of condensin II depletion in post-mitotic mouse hepatocytes (110). Hi-C data on adult liver cells reveal that mutations to the subunit Cap-H2 resulted in no changes to the chromatin conformation (110). Moreover, quantification of gene expression showed a modest change of only 64 genes, which was not enough to deem the condensin II as necessary for the maintenance of the transcriptome in mouse liver cells (110).

A common way to assess the effects of condensin depletion on gene regulation is by analysing RNA levels from mutant organisms or cell lines. A recent study in yeast has shown that condensin mutant cells have disrupted RNA levels as a result of mis-segregation of chromosomes (108). These accumulated RNA levels lead to changes in gene expression, but as an indirect effect of aberrant chromosome segregation, rather than a direct role for condensin in gene regulation. This finding raises the possibility that gene expression changes found in previous condensin mutant studies may be artefacts. Investigating condensin in a post-mitotic environment would exclude any artefacts from cell division. Taken together, these findings challenge our understanding of a role for condensin in gene regulation, and highlight a gap in our knowledge of specific, post-mitotic cell types.

Our knowledge of condensin functions are mostly limited to mitotic cells and seldom to specific cell types, especially post-mitotic cells. One study in *Drosophila* suggests condensin I subunit Cap-D3 regulates gene expression in post-mitotic, fat body cells (98). The study showed that *Cap-D3* heterozygous mutants have compromised immunity due to disrupted expression of anti-microbial gene clusters (98). These observations in post-mitotic, fat-body cells revealed that condensin subunits can have further, unidentified cell-specific and post-mitotic roles. The focus of this study is to understand condensin in a post-mitotic context in the nervous system, and to date there is no current evidence for a role for condensin regulating neuronal gene expression.

5.1.3 Chapter aims

In this chapter I aim to investigate the role of Cap-G in regulation of gene expression in post-mitotic neurons. Data from previous chapters of this thesis suggests that Cap-G is present in neuronal nuclei and that it binds dynamically to DNA in neurons (Chapter 3, 4). Knockdown of Cap-G from neurons resulted in severe behavioural, survival phenotypes and increased DNA damage and apoptosis in neurons (Chapter 3). These findings suggest that changes may be occurring at the transcriptome level in neurons. Therefore, I aim to characterise the effects of Cap-G knockdown on the transcriptome of post-mitotic neurons and how they relate to the observed phenotypes. To address this I implemented RNA-seq experiments to detect differentially expressed genes in Cap-G KD animals. Finally, I aimed to investigate how these mis-expressed genes related to the observed phenotypes as well as the Cap-G binding profiles in neurons. In this chapter I present evidence for a post-mitotic, neuron specific gene regulatory role for condensin I subunit Cap-G. I demonstrate that Cap-G knockdown exclusively in post-mitotic neurons (both immature and mature) results in disruption of gene expression. Gene expression

changes include a downregulation in CNS-specific genes and an ectopic upregulation of non-CNS genes in the knockdown tissues. Moreover, the differentially expressed genes partially overlap with genes bound by Cap-G in neurons. Overall this chapter identifies a role for Cap-G in regulation of gene expression in *Drosophila* neurons.

5.2 Results

5.2.1 Knockdown of Cap-G in neurons leads to disruption of gene expression in the CNS

The data presented in this thesis so far suggests a role for Cap-G in post-mitotic neurons. Cap-G localises to neuronal nuclei and severe developmental phenotypes result from its depletion (Chapter 3). Moreover, I profiled Cap-G binding in neurons and identified enriched binding at neuron-specific and repressed genes in neurons, as well as inaccessible chromatin states (Chapter 4). All these observations suggest that Cap-G may be involved in regulation of gene expression, necessary for correct neuronal development and overall functioning of the nervous system. To definitively assess whether Cap-G impacts gene expression in neurons, I carried out RNA-seq experiments on *Cap-G KD* animals. I used the neuron-specific *Cap-G KD* animals previously described (Chapter 3) to detect any potential changes in the neuronal transcriptome upon Cap-G depletion. I induced neuron-specific KD using the drivers *elav-GAL4* and *nSyb-GAL4* as previously described (Chapter 3, Section 3.1.4) and refer to them in this chapter as *elav-KD* and *nSyb-KD* respectively hereafter. Knockdown of Cap-G presented similar behavioural phenotypes with both drivers, but with different severities (Chapter 3, Section 3.2.5), so I continued using these two drivers to observe if such differences were also evident at the transcriptome level.

I performed RNA-seq experiments on both *elav-KD* and *nSyb-KD* third instar larval CNS tissues. The controls used in this experiment included the driver and deGradFP in the absence of GFP-tagged Cap-G (the *elav-Gal4 > UAS-deGradFP* and *nSybGal4 > UAS-deGradFP*), to account for any off-target effects of the deGradFP system. Three biological replicates were obtained for each of the *Cap-G KD* animals and *controls*, each with 30 larval brains per sample.

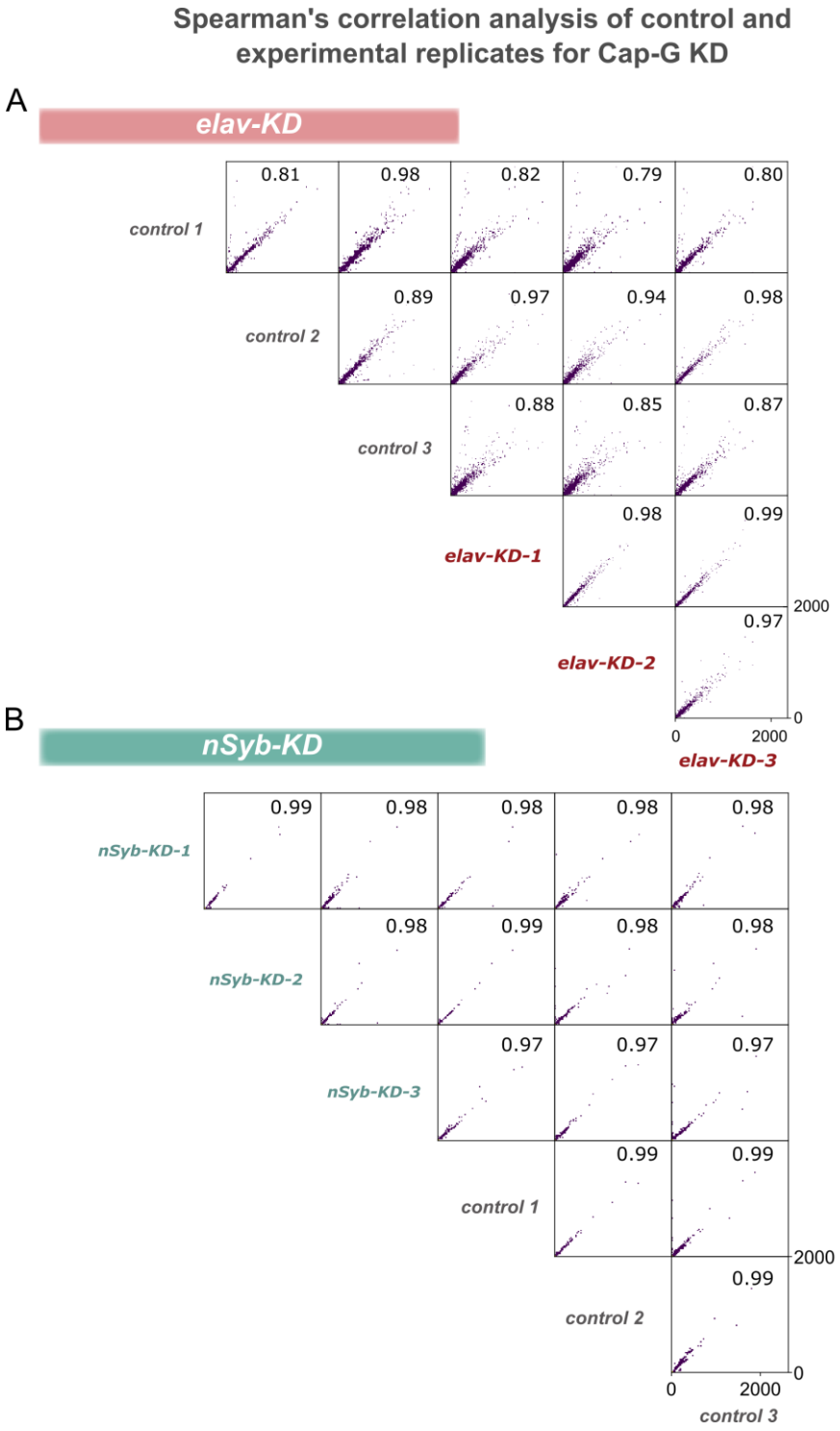


Figure 5.1 Positive correlation between Cap-G KDs replicates and experimental conditions.

Spearman's correlation analysis shows strong positive correlation between replicates for both *controls* and *elav-KD, nSyb-KD* animals. Numbers in panes represent Spearman's correlation co-efficient. The scores obtained for each replicate are computed against consecutive genome bins of 10 kb for the whole *Drosophila* genome. Control genotypes are *elav-Gal4; UAS-deGradFP* and *nSyb-Gal4; UAS-deGradFP* for *elav-KD* and *nSyb-KD* respectively.

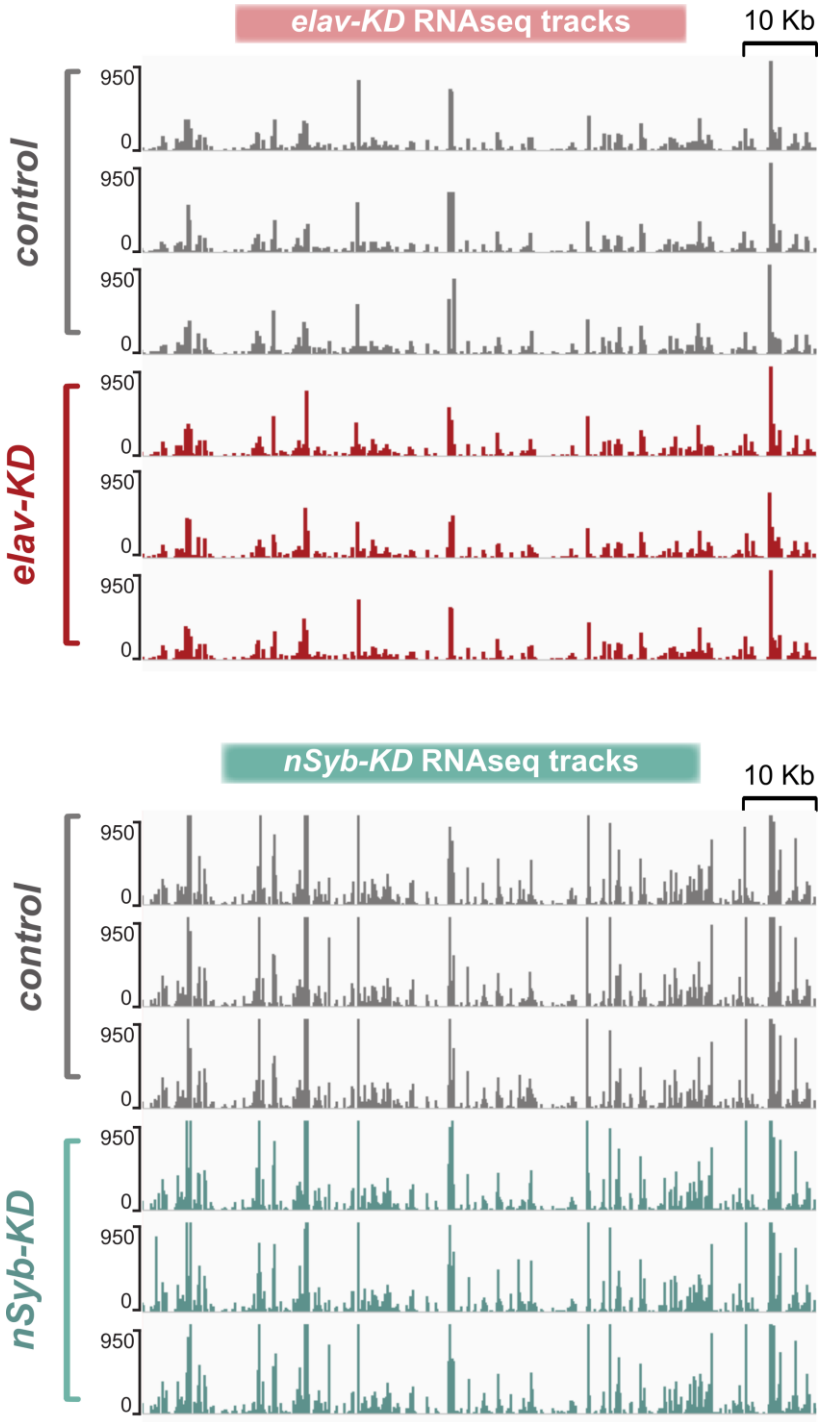


Figure 5.2. RNAseq example profiles of chr 3L show similarity between replicates.

Example of the mapped sequencing data tracks for chr 3L for *elav-KD* and *nSyb-KD*. Three replicates for each of the control and experimental samples is shown. Y-axis shows raw Fragments Per Kilobase of transcript per Million reads (FPKM).

Quality control analysis of the sequencing data obtained reveals overall similar signal detected between replicates, for both *control* and *Cap-G KD* samples. There is strong positive correlation between the genomic scores of the replicates for each of controls and experimental samples in both *Cap-G KDs* (Figure 5.1). There is also a strong correlation between the control and experimental data for both *Cap-G-KDs* (Spearman's rho ~0.9). Finally, the similarity is also visible upon inspection of signal tracks for both *Cap-G KD* samples and their respective controls (Figure 5.2).

In order to analyse whether the *controls* and *Cap-G-KD* samples were different from one another, I carried out a Principal Component Analysis to sort the data in order of variance (importance) (Figure 5.3). As expected, in *elav-KD* two separate, distinct clusters can be seen, each corresponding to either control replicates or experimental replicates (Figure 5.3A). The individual replicates for either controls or experimental samples cluster in close proximity to one another, suggesting similarity in the replicates and giving confidence in the quality of the data obtained. In *nSyb-KD* the clustering is not as clear, with one main cluster of replicates from both controls and experimental samples and an individual replicate for *nSyb-KD* sample clustering separately (Figure 5.3B). This unusual clustering of all replicates could be due to very small changes between the control and experimental replicates, not detectable by PCA. Moreover, PCA is not able to detect the origin of the variance, i.e. whether it is "between groups" or "within groups". Overall, the RNA-seq data quality is sufficient to carry out further statistical analysis.

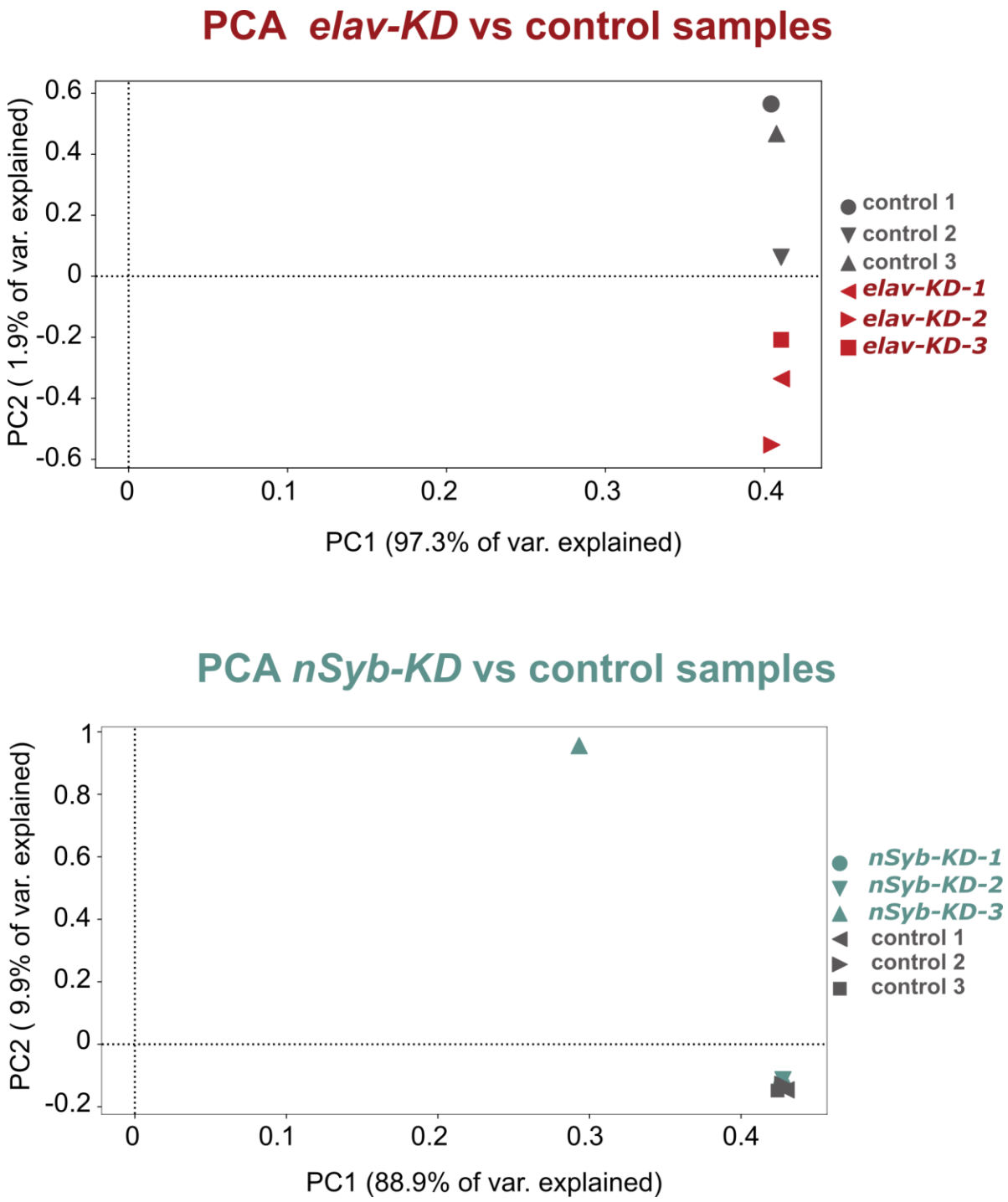


Figure 5.3. PCA between control and experimental samples.

Individual PCA for *elav-KD* and *nSyb-KD*. In *elav-KD* control replicates (grey) cluster separately from experimental replicates (red). Individual replicates cluster together for both controls and experimental samples. In *nSyb-KD* control (grey) and experimental (blue) replicates cluster except for one experimental replicate, suggesting less variance between control and experimental samples.

5.2.2 *Elav-KD* and *nSyb-KD* display similar gene expression differences but at different severities

Once the quality of the sequencing data was confirmed, I carried out differential gene expression analysis to identify significant differentially expressed genes (DEGs) in both *elav-KD* and *nSyb-KD* samples. Differential expression analysis exploits a logarithmic fold value, known as fold change (FC), to describe how much a gene is up or downregulated in a tissue. To identify significant changes in gene expression I used a $\log_2$ FC greater than 1 for upregulation (marking a doubling in expression) or less than -1 for downregulation, and a False Discovery Rate (FDR) < 0.05.

Cap-G knockdown resulted in disruption of the neuronal transcriptome, as a significant number of genes are differentially expressed in both *Cap-G KD* samples (Figure 5.4). Knockdown of Cap-G in *elav-KD* animals resulted in a strong disruption of gene expression (Figure 5.4A), whilst more modest changes are observed in *nSyb-KD* animals (Figure 5.4B). In *elav-KD* a total of 2668 genes were differentially expressed, of which 1356 were upregulated and 1308 downregulated in the larval CNS. In comparison, *nSyb-KD* significantly disrupted 278 genes, with 152 upregulated and 126 downregulated genes. The differential expression is also evident in an MA plot, which shows the intensity ratio between the mean of normalised read counts and the $\log_2$ FC in each of the *Cap-G KDs* (Figure 5.5). The MA plot for *elav-KD* is more widespread, suggesting larger differences between the samples when compared to the *nSyb-KD* plot, where the data clusters mostly around zero on the y-axis, suggesting greater similarity. These data suggest that Cap-G KD in neurons results in disruption of gene expression.

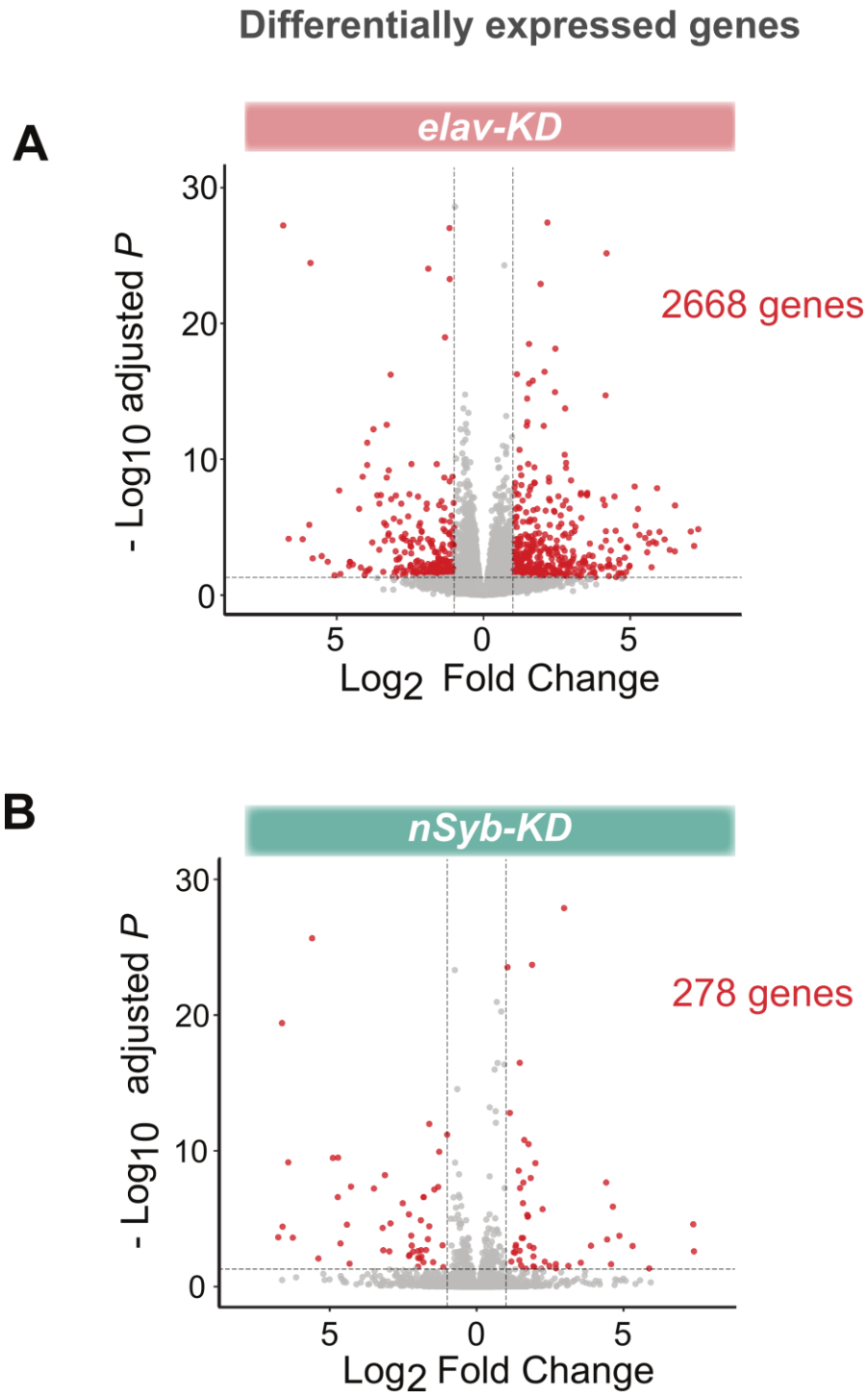


Figure 5.4. Differential gene expression analysis shows misregulation of genes.

Volcano plots showing differentially expressed genes (DEG) in *elav-KD* **(A)** and *nSyb-KD* **(B)** respectively. Significant differential expression marked in red by a Log2 fold change >1 or <-1 and FDR <0.05. Total number of DEG in *elav-KD* is 2668 compared to 278 in *nSyb-KD*.

Despite the overlap between the cells targeted by *elav-GAL4* and *nSyb-GAL4*, only 83 DEGs overlapped in both cell types. That amounts to only 3% of DEGs in *elav-KD* compared to 30% of DEGs in *nSyb-KD*. Overall, the pattern by which these overlapping genes are disrupted remains largely the same (Figure 5.6) with the exception of a few genes. For example, the gene *stall* is involved in negative regulation of neuron migration (198) and is upregulated in *nSyb-KD* but downregulated in *elav-KD*. The RNA-seq data obtained from both *Cap-G-KDs* also resulted in unique DEGs for each of *nSyb-KD* and *elav-KD.* From the numbers obtained it appears that 195 DEGs in *nSyb-KD* are not identified in *elav-KD* analysis, despite the overlap in the targeted cell types.

Signal tracks of overlapping DEGs also show the similarity in gene expression changes between *elav-KD and nSyb-KD* samples (Figure 5.7). For example, neuron-specific gene *hig* (199) is heavily downregulated in both *elav-KD* and *nSyb-KD* larval brains (Figure 5.7A). On the other hand, the non-CNS specific gene *Gadd45*, involved in oxidative stress response (200) is upregulated in both *Cap-G KD* samples (Figure 5.7B). Furthermore, an oocyte-specific gene *tej* (201) is upregulated in *nSyb-KD* but not detected in the analysis for *elav-KD.* Similarly, gene *CG12129* is downregulated in *elav-KD*, but not detected in the *nSyb-KD* analysis. Overall these data suggest that Cap-G is required in neurons for appropriate gene expression throughout development. Moreover, given the difference in severity of gene expression changes, Cap-G appears to have a more essential role at the early stages of neuronal maturation.

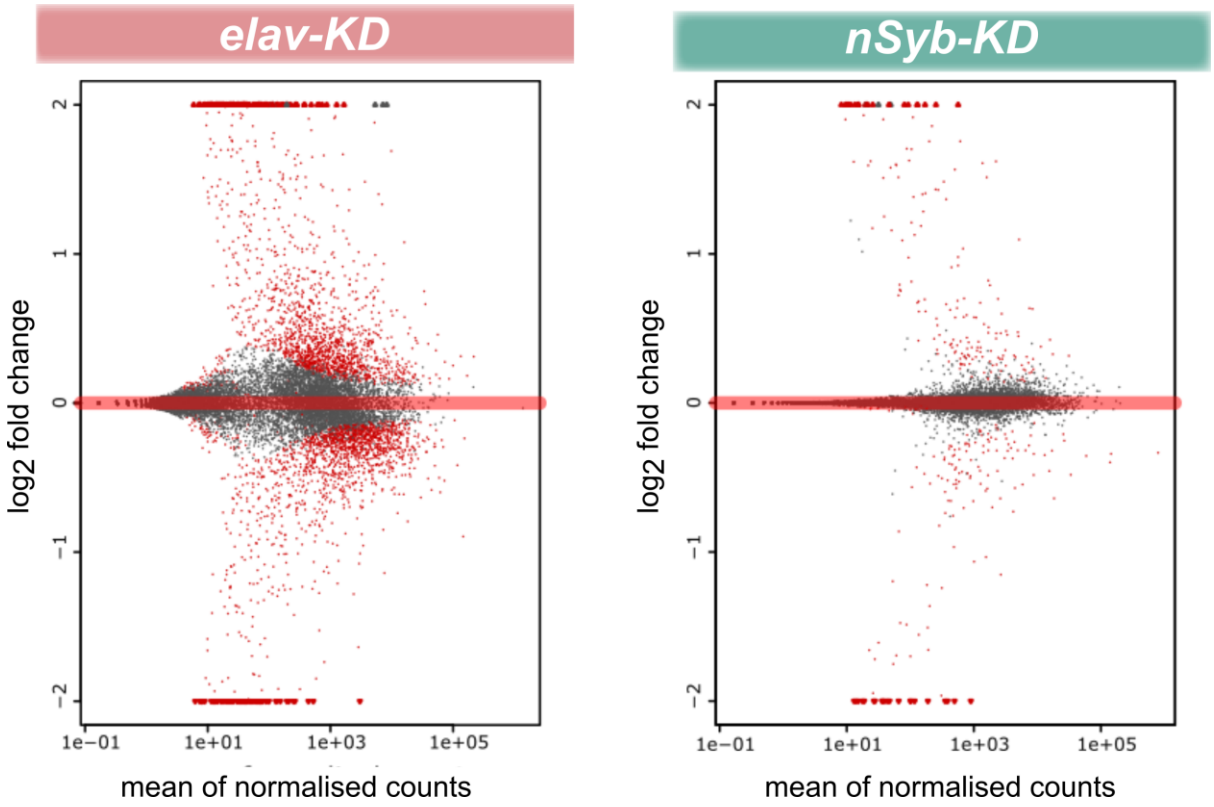


Figure 5.5 MA plot of RNA-seq datasets.

Differentially expressed genes analysis showing ratio between mean of normalised counts and $\log_2$ fold change values, significant genes marked in red. Red line on zero axis shows no difference in expression, for genes marked in grey.

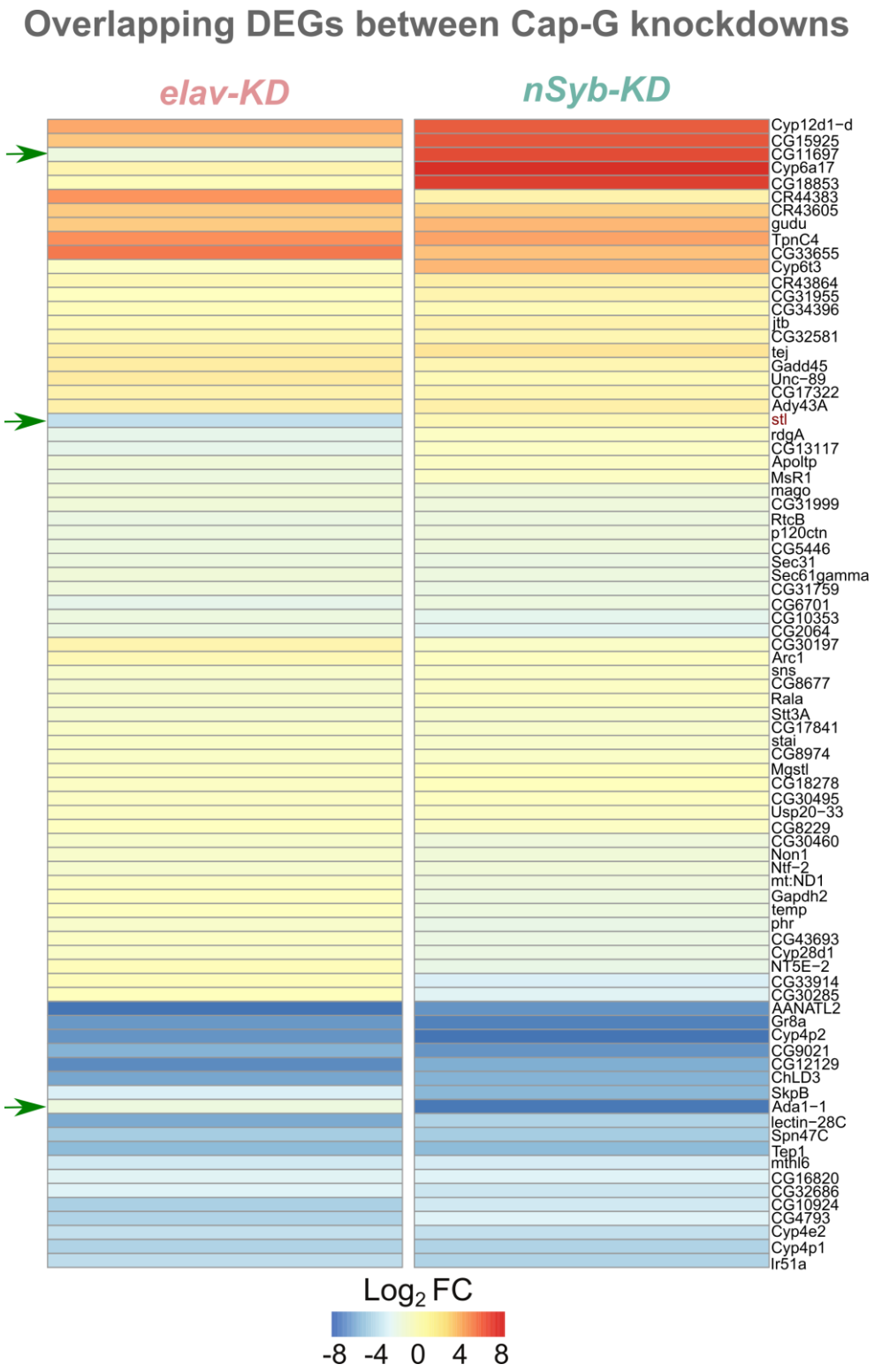


Figure 5.6. Overlapping DEGs between *elav-KD* and *nSyb-KD*.

Differentially expressed genes overlapping in both *Cap-G KDs* show similar trend of misexpression, except for a handful of genes (green arrows). Heatmap colours refer to $\log_2$ fold change values with red being most upregulated (positive values) and blue being most downregulated (negative values). Total of 83 genes identified and analysed.

5.2.3 Depletion of Cap-G in neurons leads to downregulation of CNS genes

To understand the nature of the differentially expressed genes in the *Cap-G KD* animals, I analysed enriched Gene Ontology (GO) terms associated with both downregulated and upregulated genes (Figure 5.8, 5.9). Each DEG analysed is associated with a GO term and the subontology "Biological Processes" was used to identify biologically relevant, tissue-specific or function-specific terms. Enrichment ratio analysis of each identified term was carried out, where "Gene Ratio = number of observed divided by the number of expected genes from each GO term".

In Cap-G KD animals I observed differences in the GO terms yielded for downregulated genes of both *elav-KD* and *nSyb-KD* animals. In *elav-KD* the genes downregulated belonged mostly to CNS-specific GO terms (Figure 5.8). For example, in *elav-KD* terms such as "neuron projection development", "axon guidance" and "cell morphogenesis involved in neuron differentiation" were enriched in downregulated genes (Figure 5.8). In contrast, *nSyb-KD* downregulated genes were enriched for more vague, non-CNS terms, such as "nucleotide metabolic process", "cellular respiration" and "ATP synthesis electron transport" (Figure 5.8).

GO analysis of upregulated genes for both Cap-G KD animals also displayed differences in the GO terms between the two neuronal populations (elav and nSyb) (Figure 5.9). Interestingly, in *nSyb-KD* animals displayed an upregulation in neuronal-specific terms such as "neuron projection development" (Figure 5.9B). Moreover, some non-CNS terms were also enriched *nsyb-KD*, such as "midgut development", "open tracheal system development" and "leg and muscle development" (Figure 5.9B). In *elav-KD* the terms were more vaguely associated with metabolic processes, but surprisingly, also with terms

such as “mitotic cell cycle process” and “mitotic nuclear division” appearing in the analysis (Figure 5.9A). In summary, the GO term analysis revealed that the DEGs belong to a variety of terms, suggesting disruption of neuron-specific gene expression as well as some ectopic expression of non-neuronal genes in the CNS of *Cap-G KD* animals.

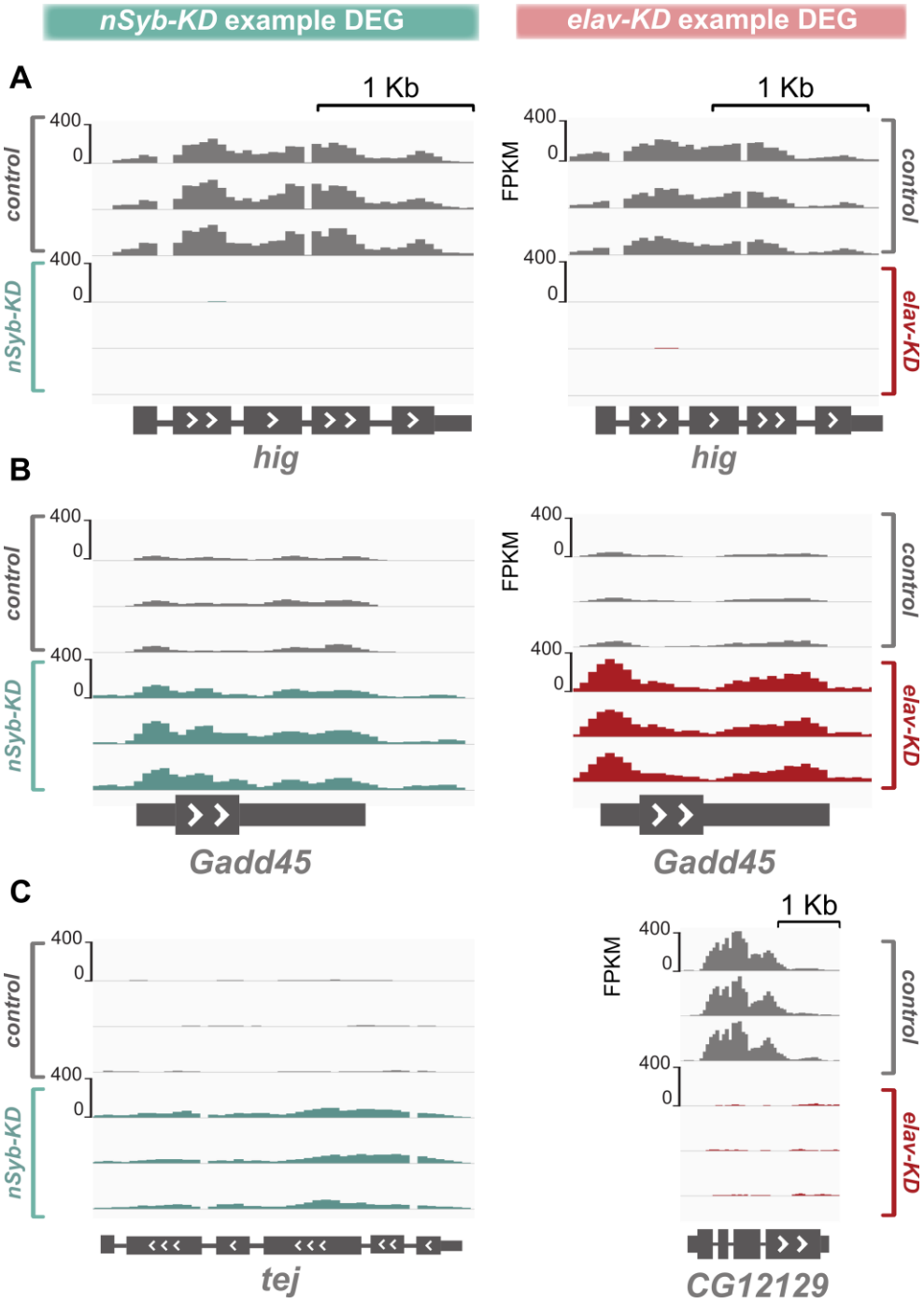


Figure 5.7. Example of DEG tracks for both *elav-KD* and *nSyb-KD*.

Control and experimental replicates shown. **A)** Significant downregulation of *hig* a neuron-specific gene seen in both *Cap-G KDs.* **B)** Significant upregulation of *Gadd45* a non-CNS specific gene seen in both *Cap-G KDs*. **C)** Individual gene examples of misexpression, in *nSyb-KD tej,* an oocyte-specific gene, is upregulated and in *elav-KD CG12129* is downregulated.

5.2.4 Ectopic upregulation of non-CNS genes in *Cap-G KD* animals

In *Cap-G KD* animals I observed an upregulation of non-CNS specific genes, which led me to investigate the tissues in which these genes are normally expressed. Moreover, I investigated whether these genes were in fact ectopically expressed in the brain. I extracted the baseline expression of these genes in the wild type CNS using previously published RNA-seq data from FlyAtlas 2 (162). I extracted gene expression data from larval tissues for the list of upregulated genes in *Cap-G KD* animals. To identify the tissues in which the genes are normally expressed I used the FPKM values > 2 as a threshold for actual expression (FPKM < 2 was considered background). With this method I was able to identify in which tissues (including the CNS), the upregulated genes were normally expressed. I excluded all the genes with detected expression in the CNS from the analysis and tissue expression data of the remaining non-CNS genes was summarised.

Tissue expression analysis revealed that the majority of upregulated genes in both *elav-KD and nSyb-KD* are not normally expressed in the larval CNS (Figure 5.10). Surprisingly, 73% of upregulated genes in *elav-KD*, and 58% in *nSyb-KD* were non-CNS specific, belonging to tissues such as the midgut, trachea and fat body. Moreover, the tissue expression distribution of ectopic genes was very similar in both *Cap-G KDs*. Each of the 6 tissues analysed amounted to ~16% of the upregulated genes. This is interesting given the large discrepancy in the original total number of differentially expressed genes detected for each of the *Cap-G KDs* (Figure 5.4). Taken together, these data suggest that the upregulated genes in *Cap-G KD* tissues are ectopically expressed in the CNS as a result of Cap-G loss.

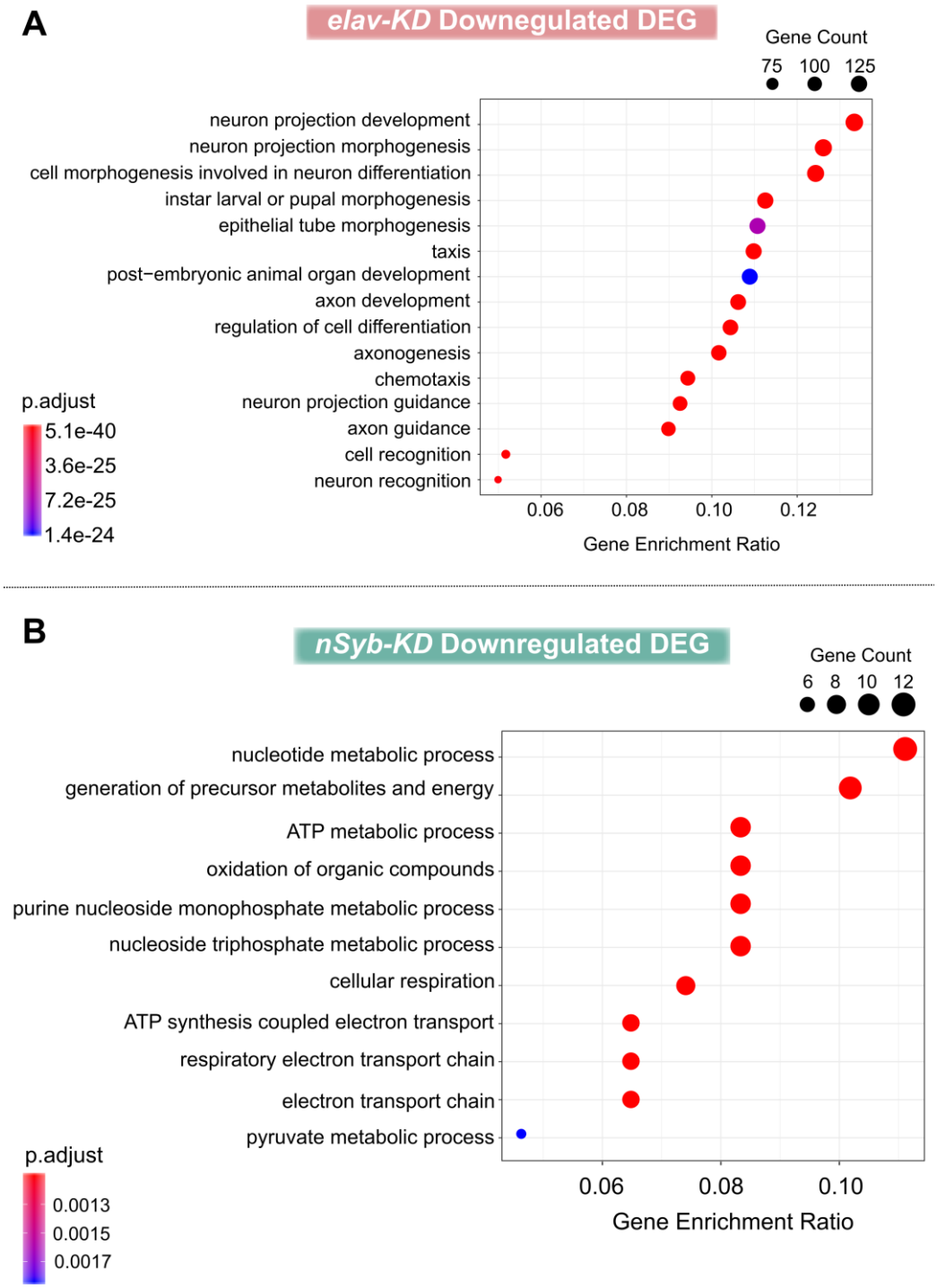


Figure 5.8 GO analysis of downregulated genes for *elav-KD* and *nSyb-KD.*

A) Downregulated DEGs in *elav-KD* show mostly neuron-specific terms. **B)** Downregulated genes in *nSyb-KD* showing metabolic-related GO terms. Gene enrichment ratio shown on x-axis, circle size correlates with number of genes detected and colour with p-adjust value.

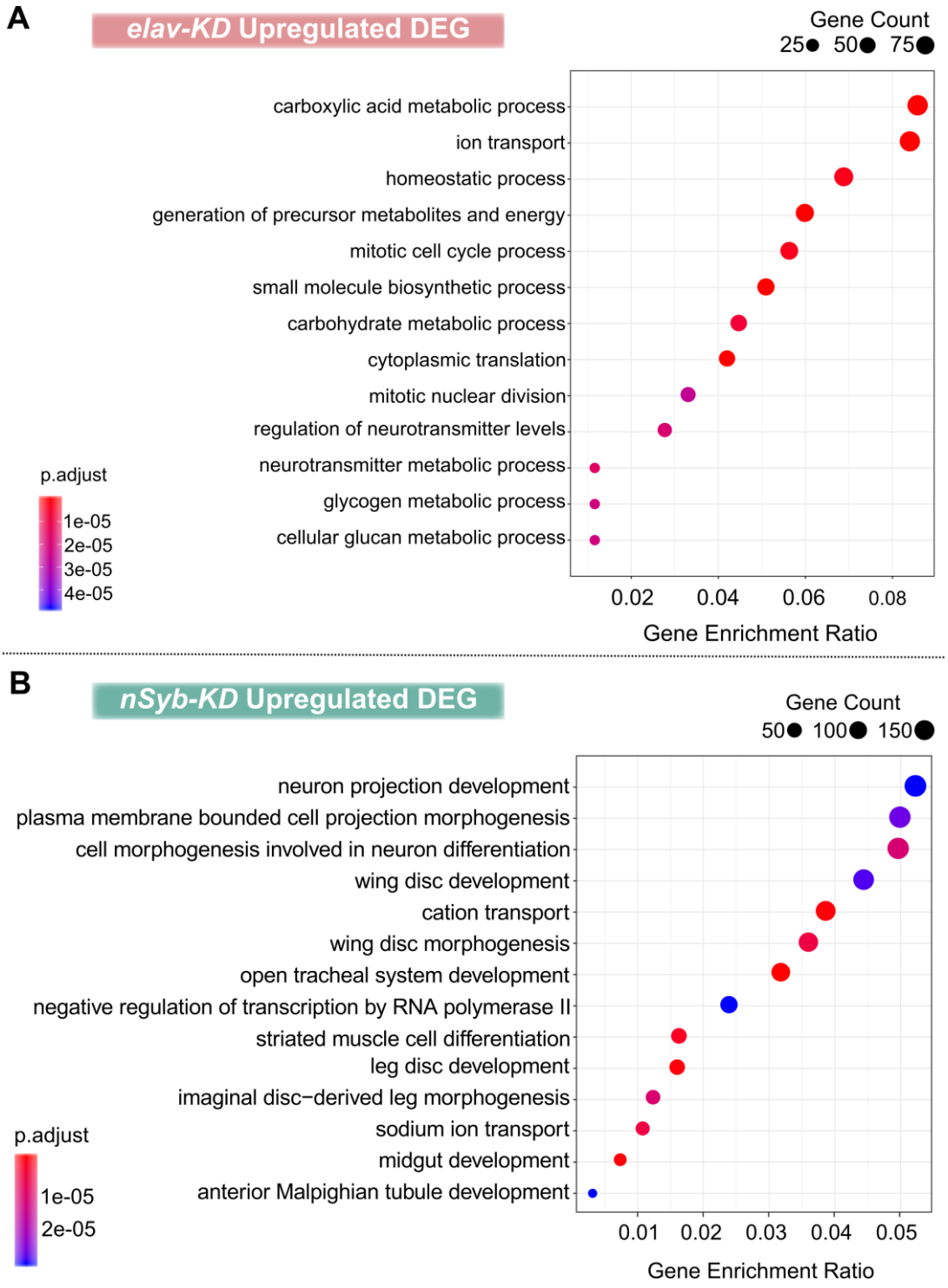


Figure 5.9 GO analysis for upregulated genes for *elav-KD* and *nSyb-KD*.

A) Upregulated DEG in *elav-KD* showing metabolism and mitotic-related enrichment in terms. **B)** Upregulated genes in *nSyb-KD* showing muscle, midgut and some neuron-specific enrichment of terms. Gene enrichment ratio shown, circle size correlates with number of genes detected and colour with p-adjust value.

5.2.5 Cap-G binding significantly overlaps with differentially expressed genes

In the previous chapter I characterised the DNA binding profiles for Cap-G in elav (newly born to mature) and nSyb (mature) neurons (Chapter 4). Cap-G binding changed dynamically across these cell stages, with specific genes identified as bound by Cap-G in each of elav and nSyb neurons (Chapter 4, Section 4.2.1). Knockdown of Cap-G in neurons would result in a loss of Cap-G binding to those unique genes and potentially affect gene expression. Therefore, to determine whether disruption of gene expression was a direct consequence of loss of Cap-G binding in the genome, I compared the RNA-seq data to the previously observed Cap-G binding profiles.

I extracted the genes uniquely bound by Cap-G in elav and nSyb neurons and analysed the amount of overlap with the differentially expressed genes detected in *elav-KD* and *nSyb-KD* samples. There was a significant overlap between Cap-G bound genes from the TaDa data and the DEGs for both *elav* and *nSyb-KD* (Figure 5.11). In elav neurons, 80% of downregulated genes and 72% of upregulated genes were directly bound by Cap-G. These overlaps are highly significant (Fisher's exact test, $p < 10^{-20}$). In mature neurons, 63% of upregulated genes and 65% of downregulated genes were bound by Cap-G (Fisher's exact test, $p < 10^{-20}$). Interestingly, the overlapping DEGs only amount to a small fraction of the genes bound by Cap-G (Figure 5.11). In *nSyb-KD* the up and down regulated genes represent ~2% of Cap-G bound genes. Moreover, in *elav-KD*, up and down regulated genes amount to ~17% of Cap-G bound genes. These data suggest that part of the regions directly bound by Cap-G are disrupted upon loss of Cap-G in neurons, proposing a role for Cap-G binding in direct regulation of gene expression.

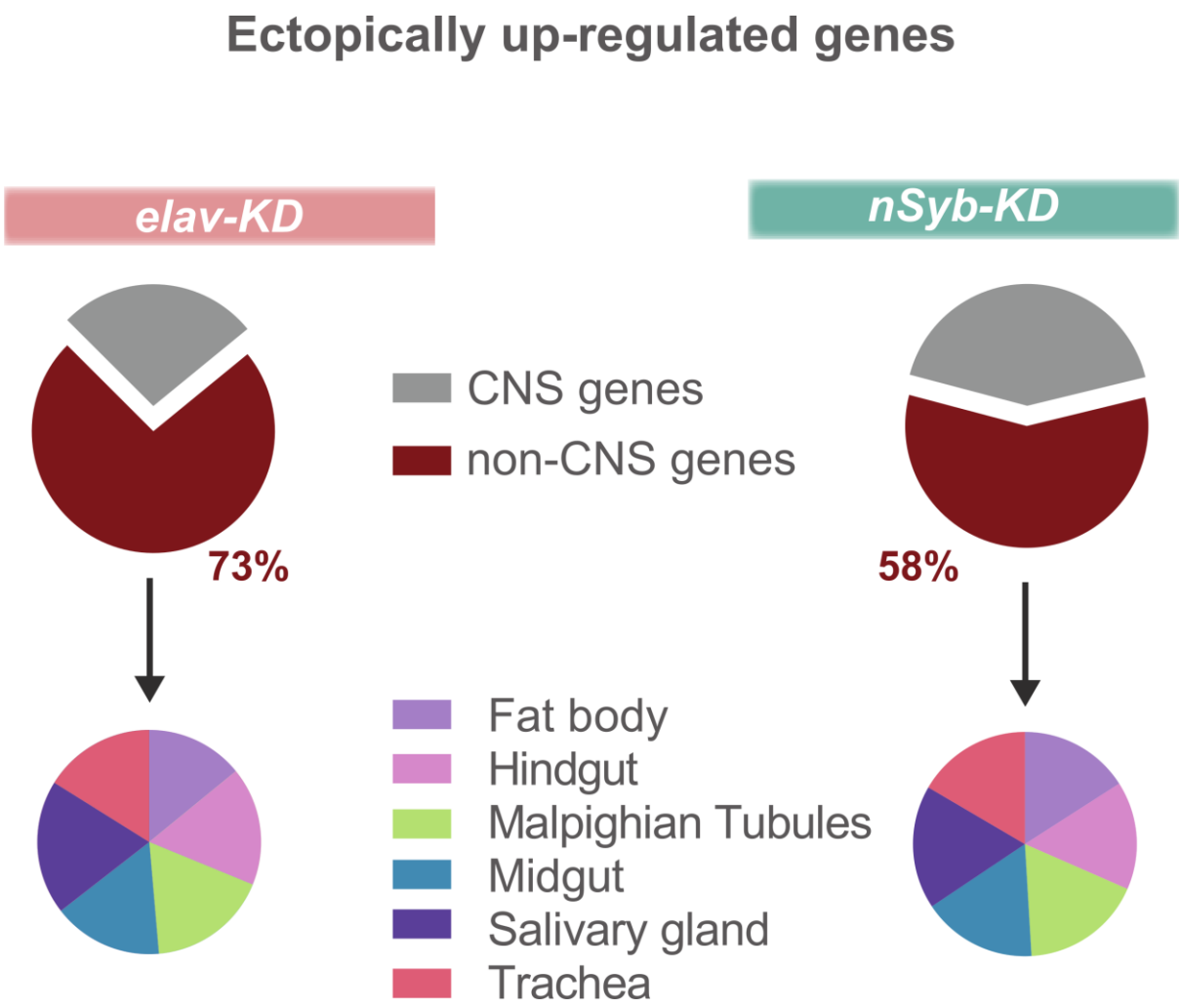


Figure 5.10. Upregulated genes are ectopically expressed in *Cap-G KDs* CNS.

The majority of upregulated differentially expressed genes, 73% in *elav-KD* and 58% in *nSyb-KD* are not normally expressed in the CNS. Upregulated genes are ectopically expressed and belong to tissues such as Fat Body and Midgut. Distribution of misexpression is similar between both *Cap-G-KDs*.

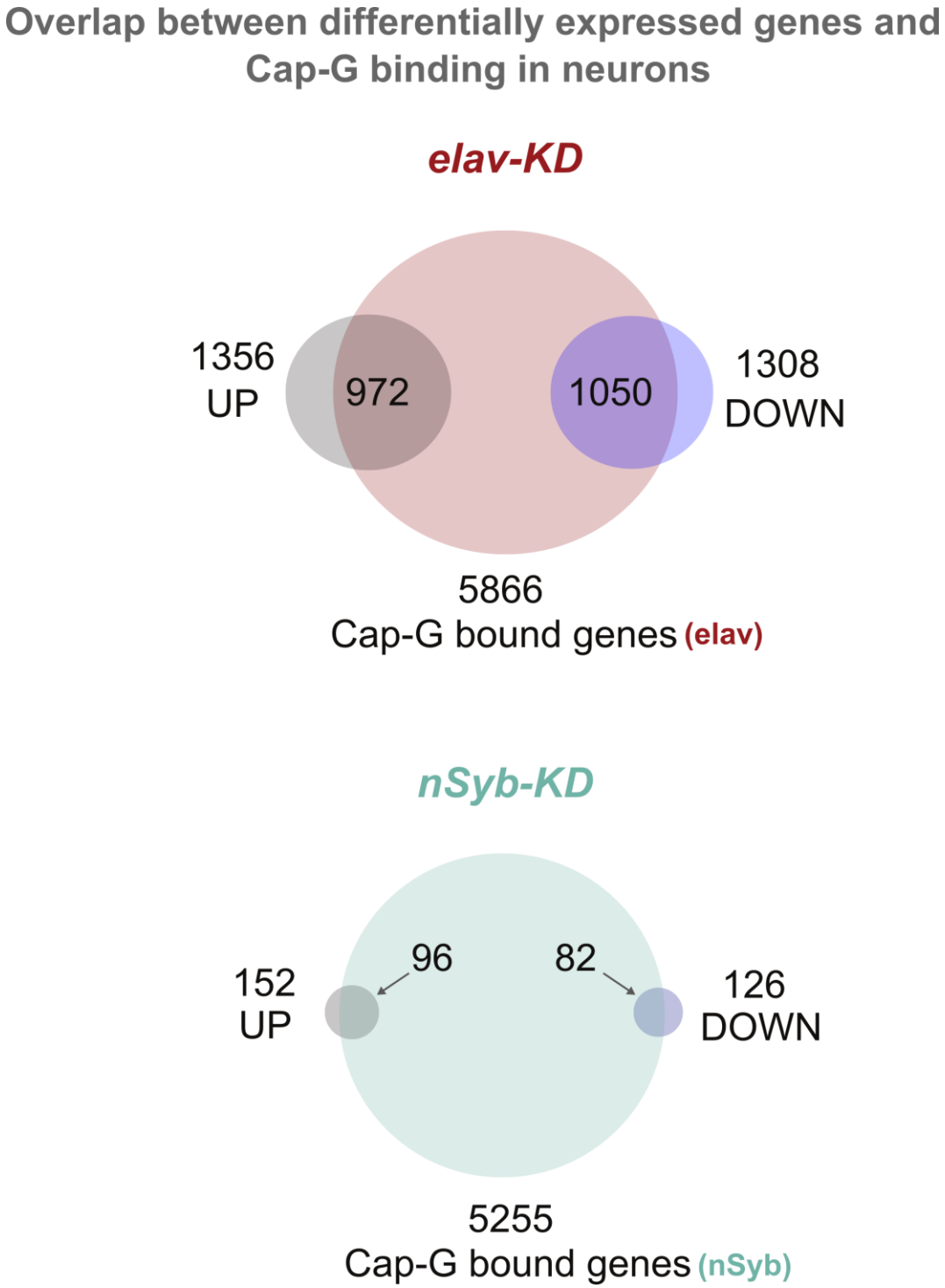


Figure 5.11. Differentialy expressed genes are bound by Cap-G in neurons.

Venn diagram showing number of DEGs, either up or down-regulated that correspond to Cap-G binding loci in *elav-KD* and *nSyb-KD*. In *elav-KD*, 80% of downregulated and 72% of upregulated DEGs are bound by Cap-G in elav neurons. In *nSyb-KD*, 65% of downregulated and 63% of upregulated DEGs are bound by Cap-G in nSyb neurons. Fisher's exact test ($p < 1x\ 10^{-20}$).

5.3 Discussion

5.3.1 Cap-G regulates gene expression

In this thesis I have carried out extensive characterisation of the effects of a neuron-specific Cap-G knockdown on *Drosophila* (Chapter 3). Having observed phenotypes such as reduced survival, impaired mobility and an increase in DNA damage, I hypothesised that Cap-G KD may lead to disruption of gene expression in neurons. Moreover, I observed unique binding of Cap-G to genomic regions as well as heterochromatin in neurons (Chapter 4). All these observations led to the next logical step of investigating gene expression changes in Cap-G KD animals by means of RNA-seq experiments. To date, there is no described role for gene regulation for any condensin subunit in post-mitotic neurons. In this chapter I presented data that suggests a role for Cap-G in regulation of gene expression in *Drosophila* neurons.

A recent study in yeast argues that changes to RNA levels seen in condensin mutants and knockdown experiments are an artefact, resulting from aberrant chromosome segregation during mitosis (108). In my experimental design I took careful consideration of this and generated exclusively post-mitotic, neuron-specific Cap-G KD animals (Chapter 3). Therefore, my RNA-seq results were most likely free from any known artefacts, as neuronal cells do not undergo mitosis. Furthermore, I provided evidence that NSCs health and mitotic ability was not affected by Cap-G KD (Chapter 3). This work provides exclusively post-mitotic evidence for a role for condensin I in transcriptional regulation in neurons.

My RNA-seq data obtained from *Cap-G KD* animals provides strong evidence for a role for Cap-G in gene regulation. I observed strong disruption to gene expression in both *elav-KD* and *nSyb-KD* animals. It is important to note that the RNA-seq data is derived from

whole larval brains, which also include NSCs and mitotic progenitors. Given the KD is only driven in post-mitotic neurons, there should not be any changes detected from non-neuronal cells. However, it is plausible that Cap-G KD in neurons may lead to non-autonomous cell effects, affecting the transcriptome of neighbouring cells which could have been detected in my dataset. Potential future experiments with single-cell RNA-seq could help provide more specificity to the results. Despite this, the quality controls of the replicates infer robustness to the data and downstream analyses suggest neuronal-specificity of the results.

5.3.2 Cap-G mediates neuron-specific gene regulation

Analyses of differentially expressed genes suggest that Cap-G is involved in neuron-specific regulation of gene expression. In Cap-G KD animals, neuron-specific genes were downregulated in *elav-KD*, whilst in *nSyb-KD* mostly non-CNS genes were downregulated. Moreover, in both *Cap-G KDs* a large majority of non-CNS genes were upregulated in neurons. The ectopic expression of non-CNS terms can be associated with loss of neuronal identity, which could result in defective functioning of the nervous system. These observations are consistent with, and could explain, the survival and mobility phenotypes observed upon Cap-G KD in neurons (Chapter 3). There is some evidence from post-mitotic cells for cell-specific, condensin-mediated regulation of gene expression (98,99). For example, loss of Cap-H2 in mouse T-cells led to failed differentiation of immature cells due to disrupted gene silencing (99). Moreover in *Drosophila* adult fat-body cells, condensin II activates transcription of innate-immunity genes (98). Once depleted of Cap-D3, the adult flies cannot effectively clear bacteria due to impaired immune response (98). From my data and previously published evidence, it is plausible to assume that

condensin I and II subunits may have varied cell-specific, gene regulatory roles, especially in a post-mitotic context.

In the *nSyb-KD* dataset, some upregulated genes also matched neuron-specific terms, suggesting that Cap-G may be also involved in maintaining correct levels of gene expression in neurons. Supporting evidence for this hypothesis is seen in SG4 somatic *Drosophila* cells, where Cap-D3 appears to regulate and maintain global levels of retrotransposon expression (181). In addition, my RNA-seq data shows comparable levels of up- and downregulation of genes in both *Cap-G KDs*. This suggests that Cap-G does not have an exclusively repressive nor activator role in neurons, but a more generic regulatory role. Condensin I and II have been implicated in both activation and repression of transcription across a variety of species and cell types (Table 5.1). In particular, condensin I is mostly associated with a repressive role, though some evidence suggests that Cap-G activates cell-specific genes in human cancer cells (94). In contrast, studies in yeast and mouse hepatocytes showed no significant changes in gene expression upon condensin depletion (109,110). In conclusion, the role for condensin complexes in gene regulation may not yet be so clear, as it appears to vary depending on the species, cell type and developmental stage studied.

Cap-G KD was driven in two cell populations, mature neurons (*nSyb-KD*) and newly born, immature neurons (*elav-KD*). In *Cap-G KD* animals there were more severe survival phenotypes in *elav-KD* compared to *nSyb-KD* (Chapter 3), which correspond to the more drastic changes in gene expression observed in *elav-KD* animals. In addition, there was a relatively small overlap of differentially expressed genes between the *elav-KD* and *nSyb-KD* conditions. This discrepancy could be attributed to the different neuronal populations targeted, although there is some overlap in the targets, the cells are at distinctly different

developmental stages. The newly born neurons have a more transient, plastic chromatin conformation in the early stages of differentiation, compared to a more closed, mature state in synapse-forming neurons (14). The difference in the cell stages is also obvious in the Cap-G binding profiles for immature neurons and mature neurons (Chapter 4). There was a distinct difference in Cap-G binding between the elav and nSyb neurons, comparable to the difference detected between NSCs and each of the neuronal stages (Chapter 4). This suggests that Cap-G function changes and adapts alongside neuronal maturation. Hence, the loss of Cap-G will have a different impact on the transcriptome depending on the neuronal stages. Moreover, the RNA-seq data obtained is from larval CNS, which is still developing, and therefore could have relatively few fully mature, nSyb neurons, compared to the number of immature, elav neurons. The discrepancy in the cell numbers targeted by the two drivers could affect the data obtained, where more modest changes may be masked by unaffected neighbouring cells, and therefore appear as a less severe global change in the transcriptome.

In order to understand the mechanism by which Cap-G regulates gene expression I compared the Cap-G binding data in neurons (Chapter 4) to the differentially expressed genes obtained. I observed that a proportion of the mis-expressed genes was directly bound by Cap-G in neurons. The binding overlap was comparable for up- and downregulated genes, further suggesting a generic regulatory role for Cap-G in neurons. The data proposes that Cap-G regulates gene expression via direct binding, however, a subset of disrupted genes was not directly bound by Cap-G, yet still mis-expressed. Therefore, it is more likely that Cap-G also has indirect effects on gene expression, potentially by means of chromatin remodelling. An example is observed for condensin I in yeast, where the SMC4 subunit can regulate gene expression by directly binding to

genomic regions, as well as regions remote from its binding sites by compacting the chromatin (97).

Cap-G binding appeared enriched at repressive chromatin in neurons, although some binding was detected at open chromatin sites (Chapter 4). Therefore Cap-G may act alongside the condensin I complex to regulate gene expression by means of remodelling 3D chromatin conformation in neurons. In *Drosophila,* heterozygous *Cap-G* mutants failed to repress positional effect variegation (PEV) effects, resulting in an increase in red-eye pigment (81). PEV is a phenomenon where heterochromatin structure mediates effects on gene expression. Therefore, disruption of PEV upon Cap-G depletion suggests that Cap-G regulates heterochromatin structure to silence gene expression (81). In *Drosophila* SG4 cells, depletion of Cap-D3 results in loss of repressive chromatin boundaries and spreading of histone repressive marks, affecting transcription of nearby genes (181). Cap-G may have a similar function in post-mitotic neurons, where it maintains boundaries between chromatin environments and hence regulate gene expression. This broad structural regulation may explain why I observed both a repression and de-repression of genes upon Cap-G KD, further emphasising a gene regulatory role for Cap-G in neurons.

Chapter 6 Discussion

6.1 Overview and summary of key findings

Our understanding of condensin complexes is beginning to extend beyond its well-defined, canonical role of compacting DNA into chromosomes. These multi-protein complexes have now been implicated in gene regulation, 3D genome architecture and co-factor interactions, revealing that its roles are highly variable and possibly cell-specific (60,82). However, an aspect of the condensin complex that remains unappreciated is its role in a post-mitotic context. By definition, post-mitotic cells do not divide, therefore, there is no need for chromosome compaction, so what role do the condensin subunits have in terminally differentiated cells? As most studies focus on understanding condensin in mitotic and interphase cells there is a considerable gap in our knowledge in this area.

In this thesis I presented evidence for a post-mitotic role of the condensin I subunit Cap-G in *Drosophila* neurons (Figure 6.1). I demonstrated that Cap-G is expressed in neurons, and is localised in neuronal nuclei across embryonic and larval stages of development. Neuron-specific knockdown of Cap-G results in severe gene expression changes, which in turn impact animal survival, mobility and increased DNA damage in neurons. Chromatin profiling of Cap-G in neurons reveals an enrichment of binding at silenced genes and compacted chromatin. Previous evidence suggests that Cap-G interacts with the neuron-specific TF Lola-N (unpublished, Southall T.D.), which is supported by their shared binding sites in neurons. Overall this study offers a novel appreciation for a role of Cap-G in neurons, having wider implications on our understanding of condensin in differentiated, post-mitotic cells. Furthermore, it offers insights into gene regulatory mechanisms that act during neuronal maturation. In this chapter, I will explore how this

research has contributed to the field, and what questions and future directions can be explored and taken forward from these data.

6.2 Cap-G in newly born and mature neurons

I observed stronger behavioural phenotypes upon Cap-G KD in early differentiating neurons, compared to more mature neurons (Chapter 3). This matched the results in chapter 5, where KD of Cap-G in younger neurons exhibits a more severe disruption of gene expression compared to a KD in more mature ones. Taken together, these findings suggest that Cap-G may have a temporal role in the development of neurons, dynamically changing as the cells differentiate and mature. These data further suggest that the role of Cap-G is more important, and has wider implications, earlier in the life of neurons. This is the case for TF Lola-N, which is expressed in newly-born neurons and is essential in preventing dedifferentiation of neurons (19). However once the neuron has matured, Lola-N is still being expressed but no longer required to maintain differentiation, as depletion of Lola-N in mature neurons does not lead to dedifferentiation phenotypes (19).

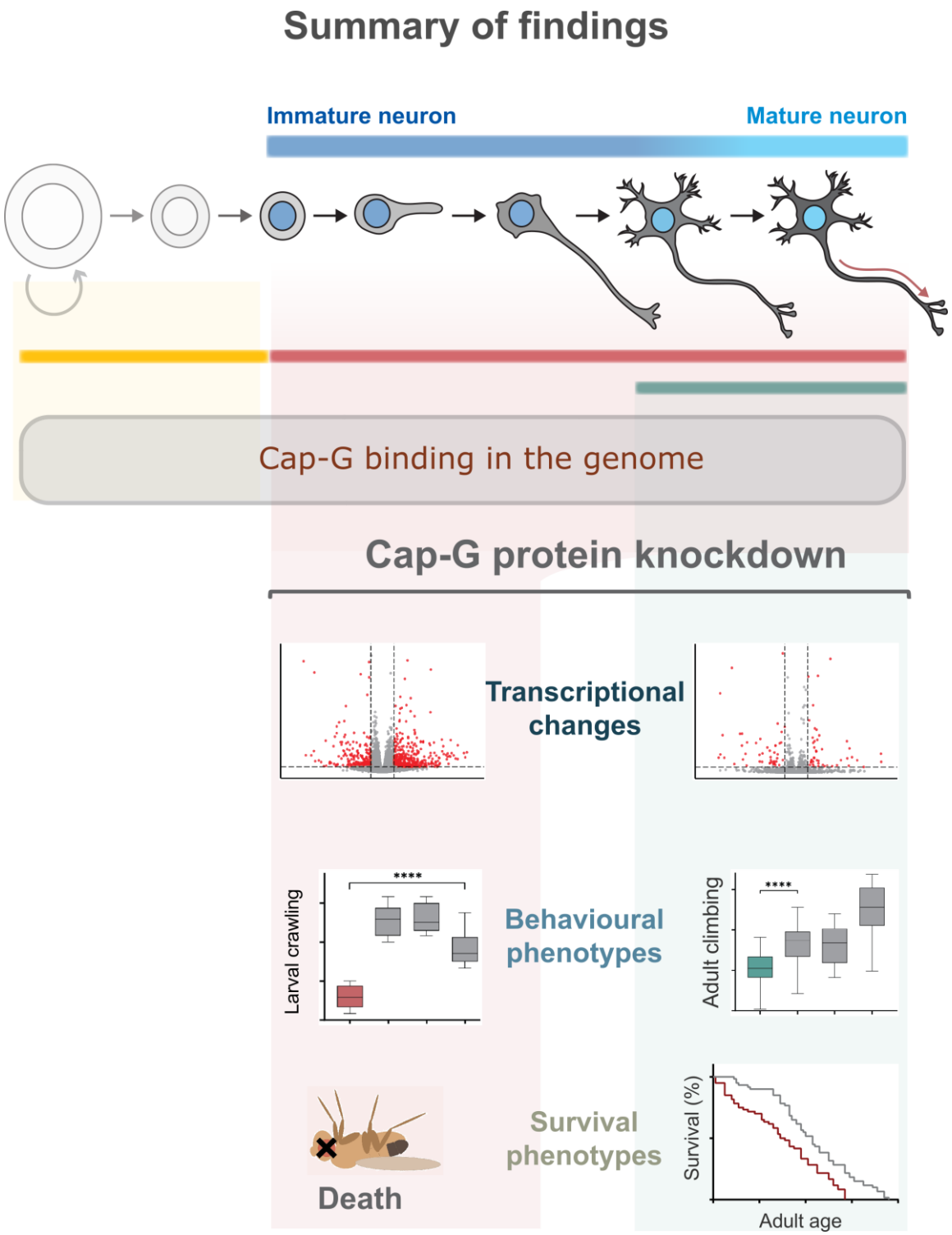


Figure 6.1 Summary of key findings.

Neurogenesis diagram from NSC to immature and mature neurons. Cap-G binding profiles were revealed for NSCs, immature and mature neurons using Targeted DamID (Chapter 4). Neuron-specific knockdown of Cap-G resulted in transcriptional changes at differential severity between immature and mature neurons (Chapter 5). Behavioural and survival phenotypes were also observed in Cap-G KD animals (Chapter 3). (Adapted from (1))

To investigate whether Cap-G has a specific temporal role, experiments could be designed to study the narrow developmental windows of "newly-born" and "mature neurons" separately. There are technical restrictions that increase the difficulty of temporally defining newly-born neurons. One of them is the fact that neurons are being born asynchronously during nervous system development. Therefore, tools are required in order to target specific stages of neuronal maturation. In *Drosophila*, GAL4 lines are the most widely used tools to infer cell-specificity during experiments using the *GAL4/UAS* system (55). Therefore, generating GAL4 lines that could target newly born neurons, as well as other stages of maturation would be highly desirable.

The current lines available include *elav-GAL4*, which is expressed in all neurons, from their birth until fully mature, and *nSyb-GAL4* which is exclusively expressed in synapse-forming, more mature neurons. These GAL4 drivers are missing the crucial early developmental window in neurons. Even though *elav-GAL4* does include the "newly-born" neuronal stage, some reported expression of *elav-GAL4* in NSCs in embryos has questioned the specificity of this tool and made it less appealing to the community (179). In my work I took extensive care to control for this reported expression, as I ensured there was no impact on mitotic cells upon knockdown of Cap-G. I did not detect a change in mitotic cell numbers, Cap-G^{EGFP} levels of expression nor on their ability to divide (Chapter 3, Section 3.2.9). However, it is still plausible that undetected, low-level knockdown of Cap-G may have occurred in progenitor cells due to the *elav-GAL4* line. Therefore, the generation of a driver line exclusively expressed in newly born neurons would be extremely useful for future experiments. Lola-N is a TF expressed as soon as the neuron is born, throughout its life (19). It could be used to potentially generate a *lola-N-GAL4* line which would offer better cell-specificity than the current *elav-GAL4* by avoiding any expression in NSCs. As *lola-N* is still expressed throughout neuronal maturation,

further specificity to the "newly-born" neuronal window could be inferred by exploiting the GAL80 protein, which actively represses GAL4 expression (202). When combined with tissue-specific proteins, GAL80 can prevent the expression of GAL4 in those tissues. A potential *Lola-GAL4* line combined with an *nSyb-GAL80* would exclude expression of GAL4 in mature neurons, inducing specific expression of GAL4 only in the newly-born neurons. It is evident that further research is needed to develop temporally-specific tools to improve our understanding of neuronal differentiation and maturation

6.3 Cap-G and Lola-N interaction and its impact on neuronal maturation

The observed interaction of Cap-G with Lola-N in a Y2H screen, their co-localisation in neurons, and binding at similar DNA regions, all suggest an interaction between these two proteins. There is emerging evidence for condensin subunits interacting with TFs, and that their recruitment to specific loci is dependent on these TFs (94,96,97). The data presented in this thesis does not allow us to fully understand the relationship between Cap-G and Lola-N yet. For example, further investigation is needed to assess whether the role for Cap-G in neurons is dependent on a potential Lola-N interaction. Moreover, Lola-N could direct and orchestrate Cap-G binding to specific DNA regions, resulting in gene regulation. Investigating this relationship could offer further insights on the mechanism and proteins involved in inducing and maintaining the differentiation of *Drosophila* neurons.

Potential insights could be gained by observing Cap-G in a *lola-N* mutant background. Immunostaining experiments could reveal whether the localisation of Cap-G to the nucleus alters in *lola-N* mutant neurons, potentially suggesting that Cap-G is recruited to neuronal nuclei by Lola-N. Moreover, to understand if Cap-G recruitment to specific DNA

regions is Lola-N dependent, an experiment could be designed to profile Cap-G binding in *lola-N* mutant neurons, revealing whether the binding changes, or remains unaffected. The reverse may be true, where Lola-N requires Cap-G to perform its function, however it is unlikely as in my *Cap-G KD* animals I did not observe dedifferentiation phenotypes similar to those seen in *lola-N* mutants (19). It is understandable that *Cap-G KD* animals do not exhibit dedifferentiation, given the plasticity and redundancy of the protein networks maintaining neuronal fate (188).

6.4 Cap-G as a regulator of gene expression in neurons

The role of condensin in gene regulation has been suggested in multiple species across multiple studies (64,85,89,94,97-99). However, these studies have been recently questioned by a report suggesting that gene expression changes and phenotypes are an indirect effect caused by aberrant chromosome segregation, rather than a direct role of condensin (108). As these studies all observed cells in a mitotic/interphase state, or used mutant animals/cell lines which disrupt condensin during chromosome compaction and segregation, it is difficult to extract the real condensin role in gene regulation. In this study, I only induced degradation of condensin I (Cap-G) in post-mitotic neurons, and took careful characterisation methods to ensure no mitotic cells were affected which would impact my results (Chapter 3). This thesis provides evidence for a post-mitotic role of Cap-G in gene expression regulation. The changes in gene expression coupled with the dysfunctional animal behaviours suggest that upon loss of Cap-G the neuronal identity of cells is compromised, resulting in failure to properly mature and operate as units of a fully functioning nervous system (Chapter 5, Chapter 3).

Evidence of gene expression changes observed in *Cap-G KD* animals suggest a role for Cap-G in gene regulation (Chapter 5). When looking at genomic binding data, Cap-G is

enriched at closed chromatin, compared to open chromatin/enhancer regions, suggesting a silencing role (Chapter 4). This is further reinforced when observing the genes disrupted in *Cap-G KD* animals, where downregulated genes were neuron-specific whilst upregulated genes were non-tissue specific. Normally, non-tissue specific genes are silenced in neurons by a variety of factors necessary to maintain cell identity (188). Chromatin states also change as the neuron differentiates and matures, with chromatin increasingly becoming less accessible, less plastic and acquiring more permanent repressive markers (7,14). The mechanism by which Cap-G potentially regulates gene expression remains elusive. Does direct binding of Cap-G at genomic regions induce gene silencing? Or does Cap-G alter chromatin states during neuronal maturation, resulting in gene silencing? Is Cap-G necessary to remodel neuronal chromatin? Further experiments can be designed to address the questions presented above. To understand whether Cap-G affects the chromatin architecture in neurons, 3D chromatin capture experiments (Hi-C) could be performed on isolated wild type and *Cap-G KD* neurons. These would offer insights on whether Cap-G is necessary to maintain certain chromatin topology in neurons.

Cap-G binding is less abundant at accessible chromatin and enhancer regions, suggesting that Cap-G is not promoting active transcription in this context (Chapter 4). However, Cap-G binding is present at a small proportion of open chromatin sites, suggesting that Cap-G may regulate some enhancer region interactions. Moreover, *Cap-G KD* animals show a decrease in neuron-specific gene expression, potentially because certain enhancer/promoter interactions could have been disrupted in the absence of Cap-G, affecting transcription (Chapter 5). Hi-C experimental data from *Cap-G KD* brains would provide the necessary information to understand whether Cap-G is indeed required for establishing, or maintaining, proper enhancer/promoter interactions in both immature

and mature neurons. However, there is conflicting evidence in the literature where changes in chromatin topology and compaction do not affect gene expression (109). For example, the disruption of highly conserved topologically associated domains (TADs) does not seem to impact on gene expression (129-132). Maintenance of gene expression could be orchestrated not only via stable chromatin topology, but also the interaction of a multitude of factors and proteins, to maintain redundancy and robustness in the system. Further research will provide exciting insights on the interaction between chromatin and proteins in regulating cell identity and differentiation.

There is a detailed understanding of what markers characterise each chromatin state (120). As chromatin is highly transient in differentiating cells, further research is needed to identify what factors and proteins orchestrate these changes. To investigate whether Cap-G has a role in manipulating chromatin states, ATAC-seq experiments, profiling the accessible chromatin state of isolated cells could be performed (203). Profiling the accessible chromatin in Cap-G-depleted neurons would detect potential changes resulting from Cap-G degradation, providing important insights for understanding the role of Cap-G in chromatin remodelling. Another key question is whether Cap-G actively compacts chromatin in a closed state, or does it "maintain" an already compacted state? Data presented in this thesis is not sufficient to answer this question, however, it does point towards a role for Cap-G in actively compacting chromatin. For example, the more severe phenotypes, both at the behavioural and molecular level, are only observed when Cap-G is knocked down during early in the life of the neuron, where the chromatin is in a more transient and plastic state, with regions yet to reach their fully compacted state (14).

6.5 A role for the condensin I complex or just for Cap-G?

A key question arising from this study is to what extent is the entire multi-protein, condensin I complex involved in neurons? The findings from my work mostly focused on characterising the role of the subunit Cap-G, with some evidence presented for subunit Barren localising in neurons (Chapter 3, Section 3.2.2). Apart from the HEAT-domain subunit Cap-G, the complex is comprised of four other proteins: two SMC ATPases, the kleisin subunit Barren, and the other HEAT-domain subunit Cap-D2 (53). There is some evidence of expression of these subunits in neurons (10,162). Therefore, it would be interesting to investigate whether the roles I described for Cap-G in neurons are exclusive to this subunit, or whether they require the fully assembled complex. Experiments that investigate this would include neuron-specific profiling of DNA binding for each of the other subunits and compare it to the Cap-G binding presented here. For more functional profiling, knockdown of the other subunits in neurons, to observe whether similar phenotypes occur, could also shed light on this matter. If it is a unique role of Cap-G in neurons, and no other condensin subunits are involved, then a yeast-two-hybrid screen to find interacting proteins of Cap-G, would help to begin to understand how Cap-G might be performing such a non-canonical function.

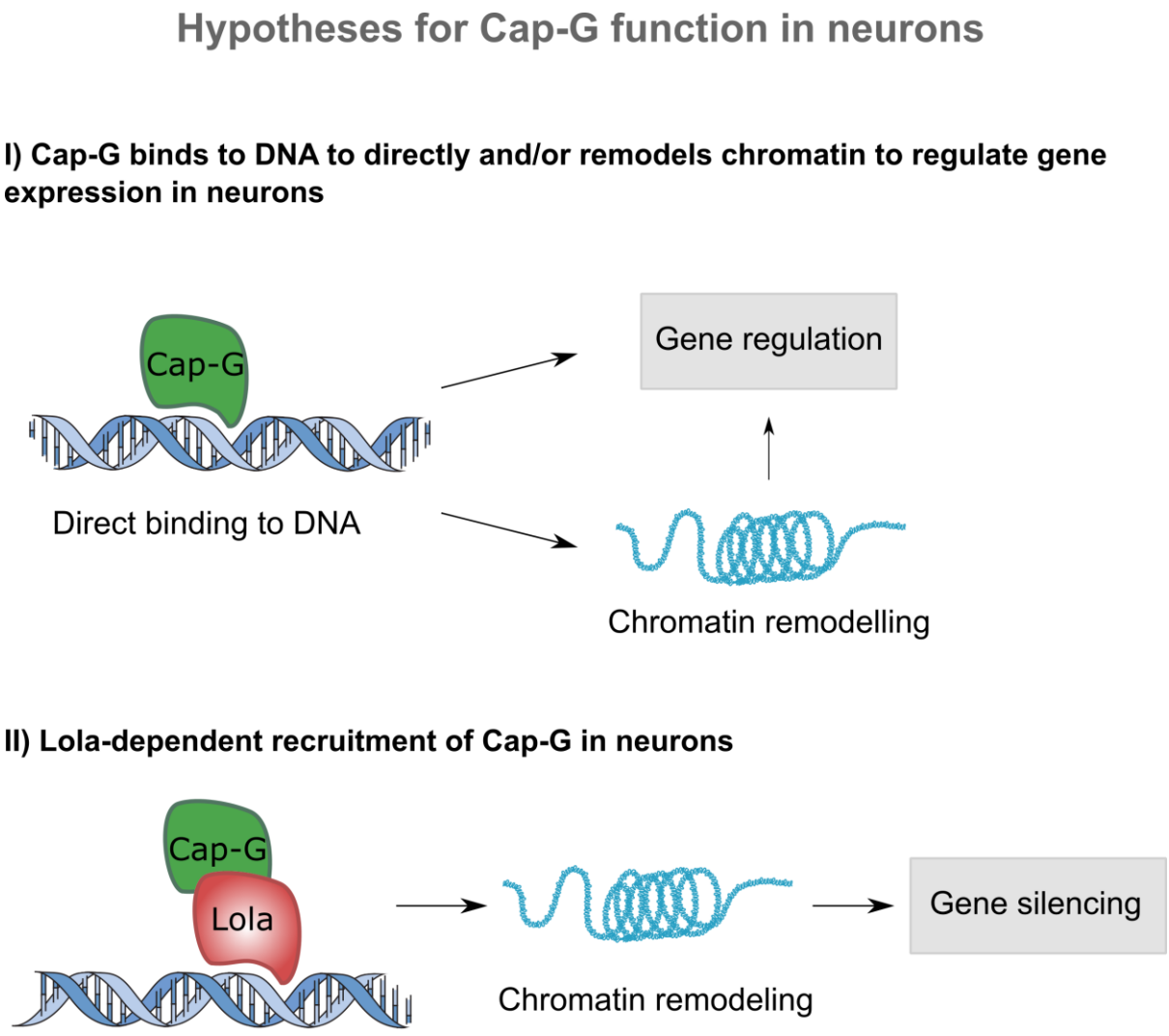


Figure 6.2. Hypotheses for Cap-G functions in neurons.

I) Current working hypothesis suggests that Cap-G binds to DNA, possibly directly regulating gene expression. Moreover, Cap-G binding could induce chromatin remodelling of 3D structure, impacting enhancer/promoter regions and closed chromatin conformations. Subsequently this drives the regulation of gene expression in neurons and promote neuron maturation. **II)** Hypothesis of Cap-G mechanisms via Lola-N dependent recruitment. Cap-G is recruited to Lola-N binding sites to compact chromatin and silence specific genes.

These experiments would not definitively address whether the canonical role of condensin is necessary in neurons, which is its ability to reel DNA loops thanks to the ATPase machinery provided by the SMC subunits (72,73). In bacteria, a point mutation to the SMC subunit results in a dominant negative phenotype, where the ATP hydrolysis activity is impaired, but the SMC maintains its dimeric structure (204). The point mutation data provided in the aforementioned study could be used to generate a *UAS-SMC-mutant* fly line. The *GAL4/UAS* system could drive the expression of dysfunctional condensin in neurons, to observe whether impairing the canonical condensin complex has similar effects to those seen in *Cap-G KD* animals, which could suggest involvement of the whole complex and not just Cap-G. Furthermore, if the whole complex is necessary in neurons and dependent on ATP-hydrolysis, it would suggest that the DNA looping mechanism is involved in neuronal DNA compaction, and that it is not exclusive to mitotic cells (60).

The evidence presented in this thesis, together with current knowledge in the field, offers the potential to speculate hypotheses for the role of Cap-G in neurons. The most plausible hypothesis from the data is that Cap-G binds to specific regions of DNA, remodels the chromatin conformation to regulate closed chromatin as well as open/enhancer regions, regulating gene expression (Figure 6.2). This regulation of the transcriptome would maintain neuronal-identity and induce neuronal maturation, which are lost upon Cap-G depletion (Chapter 3). Moreover, another interesting hypothesis arises from the possible interaction between Cap-G and Lola-N (Chapter 4). Cap-G could be recruited to specific DNA sites by Lola-N and then induce chromatin compaction of Lola-N-bound regions, resulting in more permanent silencing of these genes (Figure 6.2). Further research is necessary to definitively dissect Cap-G function in neurons, and this study provides promising evidence for the pursuit of research on condensin in post-mitotic neurons.

6.6 Implications for Health

Research of condensin complex functions has relevance not only in increasing our understanding of cell development, but also of our knowledge of human health and disease. Mutations observed in the condensin subunits have been implicated in cancer and microcephaly (205-207). In microcephaly, the development of a small brain is attributed to failed neuronal stem cell divisions, due to chromosome segregation defects upon loss of condensin (207). These diseases are mostly linked to the loss of the canonical condensin role of chromosome compaction and segregation. However, in this study, I provide evidence that condensin is also required in post-mitotic cells to ensure proper gene expression and a healthy nervous system. This suggests that condensin plays a role throughout development of the nervous system, including differentiation and maturation of neurons, after cell division occurred. This work emphasises the importance of exploring proteins and complexes outside of their canonical role, as it can lead to interesting, novel paths to further our understanding of neurodevelopment.

Neurodevelopmental disorders have been usually linked to observations in neuronal synaptic dysfunction, and emerging evidence suggests that chromatin remodelling and regulation plays an essential role in healthy neuron formation (208,209). Mutations in subunits of some chromatin remodelling complexes lead to defects in neuronal identity and maturation, affecting axon differentiation and the establishment of learning and memory, resulting in intellectual disability and neurodevelopmental disorders (208). Therefore, we are beginning to appreciate that a functioning nervous system relies on establishing and maintaining an epigenetic state in maturing neurons (209). My research identified that the chromatin-associated complex condensin is required for effective neuronal maturation. Therefore, continuous research of proteins involved in maintaining

and shaping the neuronal chromatin will help dissect the epigenetic changes that occur in neurons as they mature. This will further our understanding of how defects in chromatin remodelling can lead to neurodevelopmental disorders and potentially identify targets to advance medical research in this field.

6.7 Concluding remarks

In recent years, our understanding of the condensin complex has evolved; it is not just a machine to loop DNA and compact chromosomes, but has more important and varied roles, such as regulating gene expression and manipulating chromatin conformation. However, most research has focused on the role of condensin in mitotic and interphase cells, with limited evidence of a post-mitotic role. The importance of condensin in specific cell states and tissues has been implied, but not addressed. In this thesis, I present data for a novel role of Cap-G in neurons, challenging our understanding of the condensin complex. I demonstrated that Cap-G is essential in maintaining neuronal identity and maturation, potentially by altering the epigenetic state of differentiating and maturing neurons. My research opens a window into the emerging field of chromatin-associated complexes as mediators of epigenetic changes in neurons. This work provides an important stepping stone towards furthering our understanding of condensin in neurons and begin to appreciate the varied functions of this complex.

Chapter 7 Bibliography

1. Hassan, A., Rodriguez, P. A., Heidmann, S. K., Walmsley, E. L., Aughey, G. N., and Southall, T. D. (2020) Condensin I subunit Cap-G is essential for proper gene expression during the maturation of post-mitotic neurons. *Elife* **9**, e55159
2. Takahashi, K., and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676
3. Vierbuchen, T., Ostermeier, A., Pang, Z. P., Kokubu, Y., Sudhof, T. C., and Wernig, M. (2010) Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035-1041
4. Pasque, V., Jullien, J., Miyamoto, K., Halley-Stott, R. P., and Gurdon, J. B. (2011) Epigenetic factors influencing resistance to nuclear reprogramming. *Trends Genet* **27**, 516-525
5. Cajal, S. (1890) Sobre la aparición de las expansiones celulares en la médula embrionaria *Gaceta Sanitaria de Barcelona* **12**, 413-419
6. Gallegos, D. A., Chan, U., Chen, L. F., and West, A. E. (2018) Chromatin Regulation of Neuronal Maturation and Plasticity. *Trends Neurosci* **41**, 311-324
7. Marshall, O. J., and Brand, A. H. (2017) Chromatin state changes during neural development revealed by in vivo cell-type specific profiling. *Nat Commun* **8**, 2271
8. Venkatesh, I., Simpson, M. T., Coley, D. M., and Blackmore, M. G. (2016) Epigenetic profiling reveals a developmental decrease in promoter accessibility during cortical maturation in vivo. *Neuroepigenetics* **8**, 19-26
9. Zeisel, A., Hochgerner, H., Lonnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Haring, M., Braun, E., Borm, L. E., La Manno, G., Codeluppi, S., Furlan, A., Lee, K., Skene, N., Harris, K. D., Hjerling-Leffler, J., Arenas, E., Ernfors, P., Marklund, U., and Linnarsson, S. (2018) Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014 e1022
10. Berger, C., Harzer, H., Burkard, T. R., Steinmann, J., van der Horst, S., Laurenson, A. S., Novatchkova, M., Reichert, H., and Knoblich, J. A. (2012) FACS Purification and Transcriptome Analysis of Drosophila Neural Stem Cells Reveals a Role for Klumpfuss in Self-Renewal. *Cell Reports* **2**, 407-418
11. Zhang, S., Moy, W., Zhang, H., Leites, C., McGowan, H., Shi, J., Sanders, A. R., Pang, Z. P., Gejman, P. V., and Duan, J. (2018) Open chromatin dynamics reveals stage-specific transcriptional networks in hiPSC-based neurodevelopmental model. *Stem Cell Res* **29**, 88-98
12. Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., Xu, X., Lv, X., Hugnot, J. P., Tanay, A., and Cavalli, G. (2017) Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572 e524
13. Chathoth, K. T., and Zabet, N. R. (2019) Chromatin architecture reorganization during neuronal cell differentiation in Drosophila genome. *Genome Res* **29**, 613-625
14. Aughey, G. N., Estacio Gomez, A., Thomson, J., Yin, H., and Southall, T. D. (2018) CATaDa reveals global remodelling of chromatin accessibility during stem cell differentiation in vivo. *Elife* **7**
15. Harding, K., and White, K. (2018) Drosophila as a Model for Developmental Biology: Stem Cell-Fate Decisions in the Developing Nervous System. *J Dev Biol* **6**
16. Prokop, A., and Technau, G. M. (1991) The origin of postembryonic neuroblasts in the ventral nerve cord of Drosophila melanogaster. *Development* **111**, 79-88
17. Green, P., Hartenstein, A. Y., and Hartenstein, V. (1993) The embryonic development of the Drosophila visual system. *Cell and tissue research* **273**, 583-598
18. Homem, C. C. F., and Knoblich, J. A. (2012) Drosophila neuroblasts: a model for stem cell biology. *Development* **139**, 4297-4310
19. Southall, T. D., Davidson, C. M., Miller, C., Carr, A., and Brand, A. H. (2014) Dedifferentiation of neurons precedes tumor formation in Lola mutants. *Dev Cell* **28**, 685-696

20. Campos-Ortega, J. A., and Jan, Y. N. (1991) Genetic and molecular bases of neurogenesis in Drosophila melanogaster. *Annual review of neuroscience* **14**, 399-420
21. Nambu, P. A., and Nambu, J. R. (1996) The Drosophila fish-hook gene encodes a HMG domain protein essential for segmentation and CNS development. *Development* **122**, 3467-3475
22. Overton, P. M., Meadows, L. A., Urban, J., and Russell, S. (2002) Evidence for differential and redundant function of the Sox genes Dichaete and SoxN during CNS development in Drosophila. *Development* **129**, 4219-4228
23. Zhao, G., and Skeath, J. B. (2002) The Sox-domain containing gene Dichaete/fish-hook acts in concert with vnd and ind to regulate cell fate in the Drosophila neuroectoderm. *Development* **129**, 1165-1174
24. Buescher, M., Hing, F. S., and Chia, W. (2002) Formation of neuroblasts in the embryonic central nervous system of Drosophila melanogaster is controlled by SoxNeuro. *Development* **129**, 4193-4203
25. Aleksic, J., Ferrero, E., Fischer, B., Shen, S. P., and Russell, S. (2013) The role of Dichaete in transcriptional regulation during Drosophila embryonic development. *BMC genomics* **14**, 861
26. Arefin, B., Parvin, F., Bahrampour, S., Stadler, C. B., and Thor, S. (2019) Drosophila Neuroblast Selection Is Gated by Notch, Snail, SoxB, and EMT Gene Interplay. *Cell reports* **29**, 3636-3651. e3633
27. Knoblich, J. A., Jan, L. Y., and Jan, Y. N. (1995) Asymmetric segregation of Numb and Prospero during cell division. *Nature* **377**, 624-627
28. Knoblich, J. A. (2008) Mechanisms of asymmetric stem cell division. *Cell* **132**, 583-597
29. Bello, B. C., Izergina, N., Caussinus, E., and Reichert, H. (2008) Amplification of neural stem cell proliferation by intermediate progenitor cells in Drosophila brain development. *Neural Development* **3**
30. Boone, J. Q., and Doe, C. Q. (2008) Identification of Drosophila type II neuroblast lineages containing transit amplifying ganglion mother cells. *Developmental Neurobiology* **68**, 1185-1195
31. Baumgardt, M., Karlsson, D., Salmani, B. Y., Bivik, C., MacDonald, R. B., Gunnar, E., and Thor, S. (2014) Global programmed switch in neural daughter cell proliferation mode triggered by a temporal gene cascade. *Developmental cell* **30**, 192-208
32. Baumgardt, M., Karlsson, D., Terriente, J., Díaz-Benjumea, F. J., and Thor, S. (2009) Neuronal subtype specification within a lineage by opposing temporal feed-forward loops. *Cell* **139**, 969-982
33. Karcavich, R., and Doe, C. Q. (2005) Drosophila neuroblast 7-3 cell lineage: a model system for studying programmed cell death, Notch/Numb signaling, and sequential specification of ganglion mother cell identity. *Journal of comparative neurology* **481**, 240-251
34. Homem, C. C., Repic, M., and Knoblich, J. A. (2015) Proliferation control in neural stem and progenitor cells. *Nature Reviews Neuroscience* **16**, 647-659
35. Doe, C. Q., Chu-LaGraff, Q., Wright, D. M., and Scott, M. P. (1991) The prospero gene specifies cell fates in the Drosophila central nervous system. *Cell* **65**, 451-464
36. Betschinger, J., Mechtler, K., and Knoblich, J. A. (2006) Asymmetric segregation of the tumor suppressor brat regulates self-renewal in Drosophila neural stem cells. *Cell* **124**, 1241-1253
37. Bello, B., Reichert, H., and Hirth, F. (2006) The brain tumor gene negatively regulates neural progenitor cell proliferation in the larval central brain of Drosophila. *Development* **133**, 2639-2648
38. Wang, J. T., Kunzevitzky, N. J., Dugas, J. C., Cameron, M., Barres, B. A., and Goldberg, J. L. (2007) Disease gene candidates revealed by expression profiling of retinal ganglion cell development. *J Neurosci* **27**, 8593-8603
39. Couturier, L., Vodovar, N., and Schweisguth, F. (2012) Endocytosis by Numb breaks Notch symmetry at cytokinesis. *Nature Cell Biology* **14**, 131-139

40. Choksi, S. P., Southall, T. D., Bossing, T., Edoff, K., de Wit, E., Fischer, B. E., van Steensel, B., Micklem, G., and Brand, A. H. (2006) Prospero acts as a binary switch between self-renewal and differentiation in Drosophila neural stem cells. *Dev Cell* **11**, 775-789
41. Li, L., and Vaessin, H. (2000) Pan-neural Prospero terminates cell proliferation during Drosophila neurogenesis. *Genes & Development* **14**, 147-151
42. Southall, T. D., and Brand, A. H. (2009) Neural stem cell transcriptional networks highlight genes essential for nervous system development. *EMBO J* **28**, 3799-3807
43. Bowman, S. K., Rolland, V., Betschinger, J., Kinsey, K. A., Emery, G., and Knoblich, J. A. (2008) The tumor suppressors Brat and Numb regulate transit-amplifying neuroblast lineages in Drosophila. *Dev Cell* **14**, 535-546
44. Froldi, F., Szuperak, M., Weng, C. F., Shi, W., Papenfuss, A. T., and Cheng, L. Y. (2015) The transcription factor Nerfin-1 prevents reversion of neurons into neural stem cells. *Genes Dev* **29**, 129-143
45. Xu, J., Hao, X., Yin, M. X., Lu, Y., Jin, Y., Xu, J., Ge, L., Wu, W., Ho, M., Yang, Y., Zhao, Y., and Zhang, L. (2017) Prevention of medulla neuron dedifferentiation by Nerfin-1 requires inhibition of Notch activity. *Development* **144**, 1510-1517
46. Chia, W., Somers, W. G., and Wang, H. (2008) Drosophila neuroblast asymmetric divisions: cell cycle regulators, asymmetric protein localization, and tumorigenesis. *J Cell Biol* **180**, 267-272
47. Carney, T. D., Struck, A. J., and Doe, C. Q. (2013) midlife crisis encodes a conserved zinc-finger protein required to maintain neuronal differentiation in Drosophila. *Development* **140**, 4155-4164
48. Lin, X., Wang, F., Li, Y., Zhai, C., Wang, G., Zhang, X., Gao, Y., Yi, T., Sun, D., and Wu, S. (2018) The SCF ubiquitin ligase Slimb controls Nerfin-1 turnover in Drosophila. *Biochem Biophys Res Commun* **495**, 629-633
49. Vissers, J. H., Froldi, F., Schröder, J., Papenfuss, A. T., Cheng, L. Y., and Harvey, K. F. (2018) The Scalloped and Nerfin-1 Transcription Factors Cooperate to Maintain Neuronal Cell Fate. *Cell reports* **25**, 1561-1576. e1567
50. Goeke, S., Greene, E. A., Grant, P. K., Gates, M. A., Crowner, D., Aigaki, T., and Giniger, E. (2003) Alternative splicing of lola generates 19 transcription factors controlling axon guidance in Drosophila. *Nat Neurosci* **6**, 917-924
51. Ohsako, T., Horiuchi, T., Matsuo, T., Komaya, S., and Aigaki, T. (2003) Drosophila lola encodes a family of BTB-transcription regulators with highly variable C-terminal domains containing zinc finger motifs. *Gene* **311**, 59-69
52. Dinges, N., Morin, V., Kreim, N., Southall, T., and Roignant, J.-Y. (2017) Comprehensive Characterization Of The Complex Lola Locus Reveals A Novel Role In The Octopaminergic Pathway Via Tyramine Beta-Hydroxylase Activation. *bioRxiv*, 132027
53. Hirano, T. (2012) Condensins: universal organizers of chromosomes with diverse functions. *Genes Dev* **26**, 1659-1678
54. Jennings, B. H. (2011) Drosophila – a versatile model in biology & medicine. *Materials Today* **14**, 190-195
55. Brand, A. H., and Perrimon, N. (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401-415
56. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., and Galle, R. F. (2000) The genome sequence of Drosophila melanogaster. *Science* **287**, 2185-2195
57. Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., Millburn, G., Osumi-Sutherland, D., Schroeder, A., and Seal, R. (2009) FlyBase: enhancing Drosophila gene ontology annotations. *Nucleic acids research* **37**, D555-D559
58. Reiter, L. T., Potocki, L., Chien, S., Gribskov, M., and Bier, E. (2001) A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster. *Genome research* **11**, 1114-1125

59. Skibbens, R. V. (2019) Condensins and cohesins - one of these things is not like the other! *J Cell Sci* **132**
60. Hirano, T. (2016) Condensin-Based Chromosome Organization from Bacteria to Vertebrates. *Cell* **164**, 847-857
61. Jeppsson, K., Kanno, T., Shirahige, K., and Sjogren, C. (2014) The maintenance of chromosome structure: positioning and functioning of SMC complexes. *Nat Rev Mol Cell Biol* **15**, 601-614
62. Hagstrom, K. A., and Meyer, B. J. (2003) Condensin and cohesin: more than chromosome compactor and glue. *Nat Rev Genet* **4**, 520-534
63. Yuen, K. C., and Gerton, J. L. (2018) Taking cohesin and condensin in context. *PLoS Genet* **14**, e1007118
64. Meyer, B. J. (2010) Targeting X chromosomes for repression. *Current opinion in genetics & development* **20**, 179-189
65. Ono, T., Losada, A., Hirano, M., Myers, M. P., Neuwald, A. F., and Hirano, T. (2003) Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. *Cell* **115**, 109-121
66. Herzog, S., Nagarkar Jaiswal, S., Urban, E., Riemer, A., Fischer, S., and Heidmann, S. K. (2013) Functional dissection of the Drosophila melanogaster condensin subunit Cap-G reveals its exclusive association with condensin I. *PLoS Genet* **9**, e1003463
67. Griese, J. J., Witte, G., and Hopfner, K.-P. (2010) Structure and DNA binding activity of the mouse condensin hinge domain highlight common and diverse features of SMC proteins. *Nucleic acids research* **38**, 3454-3465
68. Piazza, I., Rutkowska, A., Ori, A., Walczak, M., Metz, J., Pelechano, V., Beck, M., and Haering, C. H. (2014) Association of condensin with chromosomes depends on DNA binding by its HEAT-repeat subunits. *Nat Struct Mol Biol* **21**, 560-568
69. Goloborodko, A., Marko, J. F., and Mirny, L. A. (2016) Chromosome compaction by active loop extrusion. *Biophysical journal* **110**, 2162-2168
70. Goloborodko, A., Imakaev, M. V., Marko, J. F., and Mirny, L. (2016) Compaction and segregation of sister chromatids via active loop extrusion. *Elife* **5**
71. Alipour, E., and Marko, J. F. (2012) Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic acids research* **40**, 11202-11212
72. Ganji, M., Shaltiel, I. A., Bisht, S., Kim, E., Kalichava, A., Haering, C. H., and Dekker, C. (2018) Real-time imaging of DNA loop extrusion by condensin. *Science*
73. Kschonsak, M., Merkel, F., Bisht, S., Metz, J., Rybin, V., Hassler, M., and Haering, C. H. (2017) Structural Basis for a Safety-Belt Mechanism That Anchors Condensin to Chromosomes. *Cell* **171**, 588-600 e524
74. Terakawa, T., Bisht, S., Eeftens, J. M., Dekker, C., Haering, C. H., and Greene, E. C. (2017) The condensin complex is a mechanochemical motor that translocates along DNA. *Science* **358**, 672-676
75. Golfier, S., Quail, T., Kimura, H., and Brugués, J. (2019) Cohesin and condensin extrude loops in a cell-cycle dependent manner. *bioRxiv*, 821306
76. Bhat, M. A., Philp, A. V., Glover, D. M., and Bellen, H. J. (1996) Chromatid segregation at anaphase requires the barren product, a novel chromosome-associated protein that interacts with topoisomerase II. *Cell* **87**, 1103-1114
77. Piskadlo, E., Tavares, A., and Oliveira, R. A. (2017) Metaphase chromosome structure is dynamically maintained by condensin I-directed DNA (de)catenation. *Elife* **6**
78. Savvidou, E., Cobbe, N., Steffensen, S., Cotterill, S., and Heck, M. M. (2005) Drosophila CAP-D2 is required for condensin complex stability and resolution of sister chromatids. *J Cell Sci* **118**, 2529-2543
79. Kinoshita, K., Kobayashi, T. J., and Hirano, T. (2015) Balancing acts of two HEAT subunits of condensin I support dynamic assembly of chromosome axes. *Developmental cell* **33**, 94-106

80. Nishide, K., and Hirano, T. (2014) Overlapping and non-overlapping functions of condensins I and II in neural stem cell divisions. *PLoS Genet* **10**, e1004847
81. Dej, K. J., Ahn, C., and Orr-Weaver, T. L. (2004) Mutations in the Drosophila condensin subunit dCAP-G: defining the role of condensin for chromosome condensation in mitosis and gene expression in interphase. *Genetics* **168**, 895-906
82. Wallace, H. A., and Bosco, G. (2013) Condensins and 3D Organization of the Interphase Nucleus. *Curr Genet Med Rep* **1**, 219-229
83. Dowen, J. M., and Young, R. A. (2014) SMC complexes link gene expression and genome architecture. *Curr Opin Genet Dev* **25**, 131-137
84. Nakazawa, N., Sajiki, K., Xu, X., Villar-Briones, A., Arakawa, O., and Yanagida, M. (2015) RNA pol II transcript abundance controls condensin accumulation at mitotically up-regulated and heat-shock-inducible genes in fission yeast. *Genes Cells* **20**, 481-499
85. Dowen, J. M., Bilodeau, S., Orlando, D. A., Hubner, M. R., Abraham, B. J., Spector, D. L., and Young, R. A. (2013) Multiple structural maintenance of chromosome complexes at transcriptional regulatory elements. *Stem Cell Reports* **1**, 371-378
86. Kim, J. H., Zhang, T., Wong, N. C., Davidson, N., Maksimovic, J., Oshlack, A., Earnshaw, W. C., Kalitsis, P., and Hudson, D. F. (2013) Condensin I associates with structural and gene regulatory regions in vertebrate chromosomes. *Nat Commun* **4**, 2537
87. D'Ambrosio, C., Schmidt, C. K., Katou, Y., Kelly, G., Itoh, T., Shirahige, K., and Uhlmann, F. (2008) Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. *Genes Dev* **22**, 2215-2227
88. Kranz, A. L., Jiao, C. Y., Winterkorn, L. H., Albritton, S. E., Kramer, M., and Ercan, S. (2013) Genome-wide analysis of condensin binding in Caenorhabditis elegans. *Genome Biol* **14**, R112
89. Yuen, K. C., Slaughter, B. D., and Gerton, J. L. (2017) Condensin II is anchored by TFIIIC and H3K4me3 in the mammalian genome and supports the expression of active dense gene clusters. *Science advances* **3**, e1700191
90. Van Bortle, K., Nichols, M. H., Li, L., Ong, C.-T., Takenaka, N., Qin, Z. S., and Corces, V. G. (2014) Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome biology* **15**, R82
91. Haeusler, R. A., Pratt-Hyatt, M., Good, P. D., Gipson, T. A., and Engelke, D. R. (2008) Clustering of yeast tRNA genes is mediated by specific association of condensin with tRNA gene transcription complexes. *Genes Dev* **22**, 2204-2214
92. Brandao, H. B., Paul, P., van den Berg, A. A., Rudner, D. Z., Wang, X., and Mirny, L. A. (2019) RNA polymerases as moving barriers to condensin loop extrusion. *Proc Natl Acad Sci U S A* **116**, 20489-20499
93. Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H. Q., Ong, C. T., Cubenas-Potts, C., Hu, M., Lei, E. P., Bosco, G., Qin, Z. S., and Corces, V. G. (2015) Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Mol Cell* **58**, 216-231
94. Li, W., Hu, Y., Oh, S., Ma, Q., Merkurjev, D., Song, X., Zhou, X., Liu, Z., Tanasa, B., and He, X. (2015) Condensin I and II complexes license full estrogen receptor α-dependent enhancer activation. *Molecular cell* **59**, 188-202
95. Meléndez, A., and Greenwald, I. (2000) Caenorhabditis elegans lin-13, a member of the LIN-35 Rb class of genes involved in vulval development, encodes a protein with zinc fingers and an LXCXE motif. *Genetics* **155**, 1127-1137
96. Smith, H. F., Roberts, M. A., Nguyen, H. Q., Peterson, M., Hartl, T. A., Wang, X. J., Klebba, J. E., Rogers, G. C., and Bosco, G. (2013) Maintenance of interphase chromosome compaction and homolog pairing in Drosophila is regulated by the condensin cap-h2 and its partner Mrg15. *Genetics* **195**, 127-146
97. Swygert, S. G., Kim, S., Wu, X., Fu, T., Hsieh, T. H., Rando, O. J., Eisenman, R. N., Shendure, J., McKnight, J. N., and Tsukiyama, T. (2019) Condensin-Dependent Chromatin Compaction Represses Transcription Globally during Quiescence. *Mol Cell* **73**, 533-546 e534

98. Longworth, M. S., Walker, J. A., Anderssen, E., Moon, N. S., Gladden, A., Heck, M. M., Ramaswamy, S., and Dyson, N. J. (2012) A shared role for RBF1 and dCAP-D3 in the regulation of transcription with consequences for innate immunity. *PLoS Genet* **8**, e1002618
99. Rawlings, J. S., Gatzka, M., Thomas, P. G., and Ihle, J. N. (2011) Chromatin condensation via the condensin II complex is required for peripheral T-cell quiescence. *The EMBO journal* **30**, 263-276
100. McDonel, P., Jans, J., Peterson, B. K., and Meyer, B. J. (2006) Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex. *Nature* **444**, 614
101. Hartl, T. A., Smith, H. F., and Bosco, G. (2008) Chromosome alignment and transvection are antagonized by condensin II. *Science* **322**, 1384-1387
102. Duncan, I. W. (2002) Transvection effects in Drosophila. *Annual review of genetics* **36**, 521-556
103. Lupo, R., Breiling, A., Bianchi, M. E., and Orlando, V. (2001) Drosophila chromosome condensation proteins Topoisomerase II and Barren colocalize with Polycomb and maintain Fab-7 PRE silencing. *Molecular cell* **7**, 127-136
104. Bhalla, N., Biggins, S., and Murray, A. W. (2002) Mutation of YCS4, a budding yeast condensin subunit, affects mitotic and nonmitotic chromosome behavior. *Molecular biology of the cell* **13**, 632-645
105. Kinsey, P. T., and Sandmeyer, S. B. (1991) Adjacent pol II and pol III promoters: transcription of the yeast retrotransposon Ty3 and a target tRNA gene. *Nucleic acids research* **19**, 1317-1324
106. Hull, M. W., Erickson, J., Johnston, M., and Engelke, D. R. (1994) tRNA genes as transcriptional repressor elements. *Molecular and cellular biology* **14**, 1266-1277
107. Kendall, A., Hull, M. W., Bertrand, E., Good, P. D., Singer, R. H., and Engelke, D. R. (2000) A CBF5 mutation that disrupts nucleolar localization of early tRNA biosynthesis in yeast also suppresses tRNA gene-mediated transcriptional silencing. *Proceedings of the National Academy of Sciences* **97**, 13108-13113
108. Hocquet, C., Robellet, X., Modolo, L., Sun, X. M., Burny, C., Cuylen-Haering, S., Toselli, E., Clauder-Munster, S., Steinmetz, L., Haering, C. H., Marguerat, S., and Bernard, P. (2018) Condensin controls cellular RNA levels through the accurate segregation of chromosomes instead of directly regulating transcription. *Elife* **7**
109. Paul, M. R., Markowitz, T. E., Hochwagen, A., and Ercan, S. (2018) Condensin Depletion Causes Genome Decompaction Without Altering the Level of Global Gene Expression in Saccharomyces cerevisiae. *Genetics* **210**, 331-344
110. Abdennur, N., Schwarzer, W., Pekowska, A., Shaltiel, I. A., Huber, W., Haering, C. H., Mirny, L., and Spitz, F. (2018) Condensin II inactivation in interphase does not affect chromatin folding or gene expression. *BioRxiv*, 437459
111. Mir, M., Bickmore, W., Furlong, E. E. M., and Narlikar, G. (2019) Chromatin topology, condensates and gene regulation: shifting paradigms or just a phase? *Development* **146**
112. Kornberg, R. D. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-871
113. Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., and Sedat, J. (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385
114. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458-472
115. Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380

116. Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., and Lander, E. S. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680
117. Rowley, M. J., and Corces, V. G. (2018) Organizational principles of 3D genome architecture. *Nat Rev Genet* **19**, 789-800
118. Nora, E. P., Goloborodko, A., Valton, A.-L., Gibcus, J. H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. A., and Bruneau, B. G. (2017) Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930-944. e922
119. Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J., and Chang, H. Y. (2016) HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* **13**, 919-922
120. Delandre, C., and Marshall, O. J. (2019) United colours of chromatin? Developmental genome organisation in flies. *Biochem Soc Trans* **47**, 691-700
121. Filion, G. J., van Bemmel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., Brugman, W., de Castro, I. J., Kerkhoven, R. M., Bussemaker, H. J., and van Steensel, B. (2010) Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143**, 212-224
122. Kharchenko, P. V., Alekseyenko, A. A., Schwartz, Y. B., Minoda, A., Riddle, N. C., Ernst, J., Sabo, P. J., Larschan, E., Gorchakov, A. A., and Gu, T. (2011) Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**, 480-485
123. Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., Wang, P., Ruan, Y., and Corces, V. G. (2017) Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell* **67**, 837-852 e837
124. Bauer, C. R., Hartl, T. A., and Bosco, G. (2012) Condensin II promotes the formation of chromosome territories by inducing axial compaction of polyploid interphase chromosomes. *PLoS Genet* **8**, e1002873
125. George, C. M., Bozler, J., Nguyen, H. Q., and Bosco, G. (2014) Condensins are Required for Maintenance of Nuclear Architecture. *Cells* **3**, 865-882
126. Rao, S. S., Huang, S.-C., St Hilaire, B. G., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K.-R., Sanborn, A. L., Johnstone, S. E., Bascom, G. D., and Bochkov, I. D. (2017) Cohesin loss eliminates all loop domains. *Cell* **171**, 305-320. e324
127. Seitan, V. C., Faure, A. J., Zhan, Y., McCord, R. P., Lajoie, B. R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., and Fisher, A. G. (2013) Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome research* **23**, 2066-2077
128. Crane, E., Bian, Q., McCord, R. P., Lajoie, B. R., Wheeler, B. S., Ralston, E. J., Uzawa, S., Dekker, J., and Meyer, B. J. (2015) Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240-244
129. Paliou, C., Guckelberger, P., Schöpflin, R., Heinrich, V., Esposito, A., Chiariello, A. M., Bianco, S., Annunziatella, C., Helmuth, J., and Haas, S. (2019) Preformed chromatin topology assists transcriptional robustness of Shh during limb development. *Proceedings of the National Academy of Sciences* **116**, 12390-12399
130. Ghavi-Helm, Y., Jankowski, A., Meiers, S., Viales, R. R., Korbel, J. O., and Furlong, E. E. (2019) Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nature genetics* **51**, 1272-1282
131. Williamson, I., Kane, L., Devenney, P. S., Flyamer, I. M., Anderson, E., Kilanowski, F., Hill, R. E., Bickmore, W. A., and Lettice, L. A. (2019) Developmentally regulated Shh expression is robust to TAD perturbations. *Development* **146**, dev179523

132. Despang, A., Schöpflin, R., Franke, M., Ali, S., Jerković, I., Paliou, C., Chan, W.-L., Timmermann, B., Wittler, L., and Vingron, M. (2019) Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nature genetics* **51**, 1263-1271
133. Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011) Open chromatin in pluripotency and reprogramming. *Nature reviews Molecular cell biology* **12**, 36-47
134. Cobbe, N., Savvidou, E., and Heck, M. M. (2006) Diverse mitotic and interphase functions of condensins in Drosophila. *Genetics* **172**, 991-1008
135. Seipold, S., Priller, F. C., Goldsmith, P., Harris, W. A., Baier, H., and Abdelilah-Seyfried, S. (2009) Non-SMC condensin I complex proteins control chromosome segregation and survival of proliferating cells in the zebrafish neural retina. *BMC Dev Biol* **9**, 40
136. Pauli, A., Althoff, F., Oliveira, R. A., Heidmann, S., Schuldiner, O., Lehner, C. F., Dickson, B. J., and Nasmyth, K. (2008) Cell-type-specific TEV protease cleavage reveals cohesin functions in Drosophila neurons. *Dev Cell* **14**, 239-251
137. Caussinus, E., Kanca, O., and Affolter, M. (2011) Fluorescent fusion protein knockout mediated by anti-GFP nanobody. *Nat Struct Mol Biol* **19**, 117-121
138. Kleinschnitz, K., Vießmann, N., Jordan, M., and Heidmann, S. K. (2020) Condensin I is required for faithful meiosis in Drosophila males. *Chromosoma*, 1-20
139. Albertson, R., Chabu, C., Sheehan, A., and Doe, C. Q. (2004) Scribble protein domain mapping reveals a multistep localization mechanism and domains necessary for establishing cortical polarity. *J Cell Sci* **117**, 6061-6070
140. Marshall, O. J., Southall, T. D., Cheetham, S. W., and Brand, A. H. (2016) Cell-type-specific profiling of protein–DNA interactions without cell isolation using targeted DamID with next-generation sequencing. *Nature protocols* **11**, 1586
141. Southall, T. D., Gold, K. S., Egger, B., Davidson, C. M., Caygill, E. E., Marshall, O. J., and Brand, A. H. (2013) Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: assaying RNA Pol II occupancy in neural stem cells. *Dev Cell* **26**, 101-112
142. Harrison, D. A., and Perrimon, N. (1993) Simple and efficient generation of marked clones in Drosophila. *Current Biology* **3**, 424-433
143. De Chaumont, F., Dallongeville, S., Chenouard, N., Hervé, N., Pop, S., Provoost, T., Meas-Yedid, V., Pankajakshan, P., Lecomte, T., and Le Montagner, Y. (2012) Icy: an open bioimage informatics platform for extended reproducible research. *Nature methods* **9**, 690
144. Nichols, C. D., Becnel, J., and Pandey, U. B. (2012) Methods to assay Drosophila behavior. *JoVE (Journal of Visualized Experiments)*, e3795
145. Yang, L., Titlow, J., Ennis, D., Smith, C., Mitchell, J., Young, F. L., Waddell, S., Ish-Horowicz, D., and Davis, I. (2017) Single molecule fluorescence in situ hybridisation for quantitating post-transcriptional regulation in Drosophila brains. *Methods* **126**, 166-176
146. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., and Schmid, B. (2012) Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676
147. Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., Schiegg, M., Ales, J., Beier, T., and Rudy, M. (2019) ilastik: interactive machine learning for (bio) image analysis. *Nature Methods*, 1-7
148. Olivo-Marin, J.-C. (2002) Extraction of spots in biological images using multiscale products. *Pattern recognition* **35**, 1989-1996
149. Gibson, D. G., Smith, H. O., Hutchison, C. A., 3rd, Venter, J. C., and Merryman, C. (2010) Chemical synthesis of the mouse mitochondrial genome. *Nat Methods* **7**, 901-903
150. Bischof, J., Maeda, R. K., Hediger, M., Karch, F., and Basler, K. (2007) An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A* **104**, 3312-3317
151. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S., and Karolchik, D. (2010) BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204-2207

152. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., and Glass, C. K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589
153. Rivera, J., Keränen, S. V. E., Gallo, S. M., and Halfon, M. S. (2018) REDfly: the transcriptional regulatory element database for Drosophila. *Nucleic acids research* **47**, D828-D834
154. Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., and Manke, T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* **44**, W160-W165
155. Quinlan, A. R., and Hall, I. M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842
156. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology* **16**, 284-287
157. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21
158. Liao, Y., Smyth, G. K., and Shi, W. (2019) The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic acids research* **47**, e47-e47
159. Love, M. I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550
160. Blighe, K. (2019) EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.2. 0.
161. Kolde, R., and Kolde, M. R. (2015) Package 'pheatmap'. *R Package* **1**
162. Leader, D. P., Krause, S. A., Pandit, A., Davies, S. A., and Dow, J. A. T. (2017) FlyAtlas 2: a new version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucleic acids research* **46**, D809-D815
163. McKinney, W. (2010) Data structures for statistical computing in python. in *Proceedings of the 9th Python in Science Conference*, Austin, TX
164. Oliveira, R. A., Heidmann, S., and Sunkel, C. E. (2007) Condensin I binds chromatin early in prophase and displays a highly dynamic association with Drosophila mitotic chromosomes. *Chromosoma* **116**, 259-274
165. Ciechanover, A. (1998) The ubiquitin–proteasome pathway: on protein death and cell life. *The EMBO journal* **17**, 7151-7160
166. Jiang, J., and Struhl, G. (1998) Regulation of the Hedgehog and Wingless signalling pathways by the F-box/WD40-repeat protein Slimb. *Nature* **391**, 493-496
167. Ciechanover, A., and Ben-Saadon, R. (2004) N-terminal ubiquitination: more protein substrates join in. *Trends in cell biology* **14**, 103-106
168. Raychaudhuri, N., Dubruille, R., Orsi, G. A., Bagheri, H. C., Loppin, B., and Lehner, C. F. (2012) Transgenerational propagation and quantitative maintenance of paternal centromeres depends on Cid/Cenp-A presence in Drosophila sperm. *PLoS biology* **10**
169. Nagarkar-Jaiswal, S., Lee, P.-T., Campbell, M. E., Chen, K., Anguiano-Zarate, S., Gutierrez, M. C., Busby, T., Lin, W.-W., He, Y., and Schulze, K. L. (2015) A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila. *Elife* **4**, e05338
170. Port, F., and Bullock, S. L. (2016) Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nature methods* **13**, 852
171. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *nature* **391**, 806-811

172. Neumüller, R. A., Wirtz-Peitz, F., Lee, S., Kwon, Y., Buckner, M., Hoskins, R. A., Venken, K. J., Bellen, H. J., Mohr, S. E., and Perrimon, N. (2012) Stringent analysis of gene function and protein–protein interactions using fluorescently tagged genes. *Genetics* **190**, 931-940
173. Jager, H., Rauch, M., and Heidmann, S. (2005) The Drosophila melanogaster condensin subunit Cap-G interacts with the centromere-specific histone H3 variant CID. *Chromosoma* **113**, 350-361
174. Lanson Jr, N. A., Maltare, A., King, H., Smith, R., Kim, J. H., Taylor, J. P., Lloyd, T. E., and Pandey, U. B. (2011) A Drosophila model of FUS-related neurodegeneration reveals genetic interaction between FUS and TDP-43. *Human molecular genetics* **20**, 2510-2523
175. Gargano, J. W., Martin, I., Bhandari, P., and Grotewiel, M. S. (2005) Rapid iterative negative geotaxis (RING): a new method for assessing age-related locomotor decline in Drosophila. *Exp Gerontol* **40**, 386-395
176. Aono, N., Sutani, T., Tomonaga, T., Mochida, S., and Yanagida, M. (2002) Cnd2 has dual roles in mitotic condensation and interphase. *Nature* **417**, 197-202
177. Chen, E. S., Sutani, T., and Yanagida, M. (2004) Cti1/C1D interacts with condensin SMC hinge and supports the DNA repair function of condensin. *Proc Natl Acad Sci U S A* **101**, 8078-8083
178. Lake, C. M., Holsclaw, J. K., Bellendir, S. P., Sekelsky, J., and Hawley, R. S. (2013) The development of a monoclonal antibody recognizing the Drosophila melanogaster phosphorylated histone H2A variant (gamma-H2AV). *G3 (Bethesda)* **3**, 1539-1543
179. Berger, C., Renner, S., Lüer, K., and Technau, G. M. (2007) The commonly used marker ELAV is transiently expressed in neuroblasts and glial cells in the Drosophila embryonic CNS. *Developmental dynamics: an official publication of the American Association of Anatomists* **236**, 3562-3568
180. Giet, R., and Glover, D. M. (2001) Drosophila aurora B kinase is required for histone H3 phosphorylation and condensin recruitment during chromosome condensation and to organize the central spindle during cytokinesis. *J Cell Biol* **152**, 669-682
181. Schuster, A. T., Sarvepalli, K., Murphy, E. A., and Longworth, M. S. (2013) Condensin II subunit dCAP-D3 restricts retrotransposon mobilization in Drosophila somatic cells. *PLoS Genet* **9**, e1003879
182. Heale, J. T., Ball, A. R., Jr., Schmiesing, J. A., Kim, J. S., Kong, X., Zhou, S., Hudson, D. F., Earnshaw, W. C., and Yokomori, K. (2006) Condensin I interacts with the PARP-1-XRCC1 complex and functions in DNA single-strand break repair. *Mol Cell* **21**, 837-848
183. Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014) Functional and topological characteristics of mammalian regulatory domains. *Genome research* **24**, 390-400
184. Aughey, G. N., and Southall, T. D. (2015) Dam it's good! DamID profiling of protein-DNA interactions. *Wiley Interdiscip Rev Dev Biol*
185. Kozak, M. (2001) Constraints on reinitiation of translation in mammals. *Nucleic Acids Res* **29**, 5226-5232
186. van den Ameele, J., Krautz, R., and Brand, A. H. (2019) TaDa! Analysing cell type-specific chromatin in vivo with Targeted DamID. *Current opinion in neurobiology* **56**, 160-166
187. Aughey, G. N., Cheetham, S. W., and Southall, T. D. (2019) DamID as a versatile tool for understanding gene regulation. *Development* **146**
188. Froldi, F., and Cheng, L. Y. (2016) Understanding how differentiation is maintained: lessons from the Drosophila brain. *Cell Mol Life Sci* **73**, 1641-1644
189. Bello, B. C., Hirth, F., and Gould, A. P. (2003) A pulse of the Drosophila Hox protein Abdominal-A schedules the end of neural proliferation via neuroblast apoptosis. *Neuron* **37**, 209-219
190. Gold, K. S., and Brand, A. H. (2014) Optix defines a neuroepithelial compartment in the optic lobe of the Drosophila brain. *Neural development* **9**, 18

191. Fritsch, C., Brown, J. L., Kassis, J. A., and Muller, J. (1999) The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. *Development* **126**, 3905-3913
192. Capovilla, M., Eldon, E. D., and Pirrotta, V. (1992) The giant gene of Drosophila encodes a b-ZIP DNA-binding protein that regulates the expression of other segmentation gap genes. *Development* **114**, 99-112
193. Grueber, W. B., Jan, L. Y., and Jan, Y. N. (2003) Different levels of the homeodomain protein cut regulate distinct dendrite branching patterns of Drosophila multidendritic neurons. *Cell* **112**, 805-818
194. Rowley, M. J., Lyu, X., Rana, V., Ando-Kuri, M., Karns, R., Bosco, G., and Corces, V. G. (2019) Condensin II Counteracts Cohesin and RNA Polymerase II in the Establishment of 3D Chromatin Organization. *Cell Rep* **26**, 2890-2903 e2893
195. Poorey, K., Viswanathan, R., Carver, M. N., Karpova, T. S., Cirimotich, S. M., McNally, J. G., Bekiranov, S., and Auble, D. T. (2013) Measuring chromatin interaction dynamics on the second time scale at single-copy genes. *Science* **342**, 369-372
196. Teytelman, L., Thurtle, D. M., Rine, J., and van Oudenaarden, A. (2013) Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proceedings of the National Academy of Sciences* **110**, 18602-18607
197. Paul, M. R., Hochwagen, A., and Ercan, S. (2019) Condensin action and compaction. *Curr Genet* **65**, 407-415
198. Lhamo, T., and Ismat, A. (2015) The extracellular protease stl functions to inhibit migration of v'ch1 sensory neuron during Drosophila embryogenesis. *Mechanisms of development* **137**, 1-10
199. Hoshino, M., Matsuzaki, F., Nabeshima, Y.-i., and Hama, C. (1993) hikaru genki, a CNS-specific gene identified by abnormal locomotion in Drosophila, encodes a novel type of protein. *Neuron* **10**, 395-407
200. Moskalev, A., Plyusnina, E., Shaposhnikov, M., Shilova, L., Kazachenok, A., and Zhavoronkov, A. (2012) The role of D-GADD45 in oxidative, thermal and genotoxic stress resistance. *Cell Cycle* **11**, 4222-4241
201. Patil, V. S., and Kai, T. (2010) Repression of retroelements in Drosophila germline via piRNA pathway by the Tudor domain protein Tejas. *Current Biology* **20**, 724-730
202. Ma, J., and Ptashne, M. (1987) The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80. *Cell* **50**, 137-142
203. Yan, F., Powell, D. R., Curtis, D. J., and Wong, N. C. (2020) From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome biology* **21**, 22
204. Schwartz, M. A., and Shapiro, L. (2011) An SMC ATPase mutant disrupts chromosome segregation in Caulobacter. *Mol Microbiol* **82**, 1359-1374
205. Woodward, J., Taylor, G. C., Soares, D. C., Boyle, S., Sie, D., Chathoth, K., Vukovic, M., Tarrats, N., Jamieson, D., and Campbell, K. J. (2016) Condensin II mutation causes T-cell lymphoma through tissue-specific genome instability. *Genes & development* **30**, 2173-2186
206. Leiserson, M. D., Vandin, F., Wu, H.-T., Dobson, J. R., Eldridge, J. V., Thomas, J. L., Papoutsaki, A., Kim, Y., Niu, B., and McLellan, M. (2015) Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nature genetics* **47**, 106
207. Martin, C.-A., Murray, J. E., Carroll, P., Leitch, A., Mackenzie, K. J., Halachev, M., Fetit, A. E., Keith, C., Bicknell, L. S., and Fluteau, A. (2016) Mutations in genes encoding condensin complex proteins cause microcephaly through decatenation failure at mitosis. *Genes & development* **30**, 2158-2172
208. Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018) The chromatin basis of neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. *Prog Neuropsychopharmacol Biol Psychiatry* **84**, 306-327

209. Iwase, S., Bérubé, N. G., Zhou, Z., Kasri, N. N., Battaglioli, E., Scandaglia, M., and Barco, A. (2017) Epigenetic etiology of intellectual disability. *Journal of Neuroscience* **37**, 10773-10782

Chapter 8 Appendix

Generation of *UAS-LT3-Cap-G-Dam* construct for TaDa experiments

Primers used in this study.

DNA template	Primers 5' to 3'
mCherry_Dam_FW	AGGGAATTGGGAATTCGTTAACAGATCTGCATGGCAACTAGCGGCATGGTTAGTAAAGGA
mCherry_Dam_RV	CTTTGGTTTGGCCATGTTATTATGCGGTACCA
Cap-G_FW	TGGTACCGCATAATAACATGGCCAAACCAAAG
Cap-G_RV	GCGATTTTTCTTCATCAGATCCTCTTCAGAGATGAGTTTCTGTTCTTTCCTCCTGCTGCG
Dam_FW	GCGCAGCAGGAGGAAAGAACAGAAACTCATCTCTGAAGAGGATCTGATGAAGAAAAATCGC
Dam_RV	TTCACAAAGATCCTCTAGAGGTACCCTCGATTAACCGGCTTTTTTCGCGGGTGAAACGACTCC
PCFD6_seq_FW	CGAGCGGAGACTCTAGCG
PCFD6_seq_RV	ATGCAGTACACGTTTAGGCA
PCFD6seqfwd	TTAGAGCTTTAAATCTCTGTAGGTAG
PCFD6seqrev	ATGCAGTACACGTTTAGGCA
Cap-G_CRISPR_FW1	GCGGCCCGGGTTCGATTCCCGGCCGATGCAGAAAGCCGTCCTAGGGGCTGAGGGTTTCAGAGCTATGCTGGAAACA
Cap-G_CRISPR_RV1	CCCGCACATACGTTATCGCATCTTGCACCAGCCGGGAATCGAACCC
Cap-G_CRISPR_FW2	AGATGCGATAACGTATGTGCGGGGTTTTAGAGCTAGAAATAGCAAG
Cap-G_CRISPR_RV2	ATTTTAACTTGCTATTTCTAGCTCTAAAAC CCCAACGTGCGTCAGTGCATCATTGCACCAGCCGGGAATCGAACCC

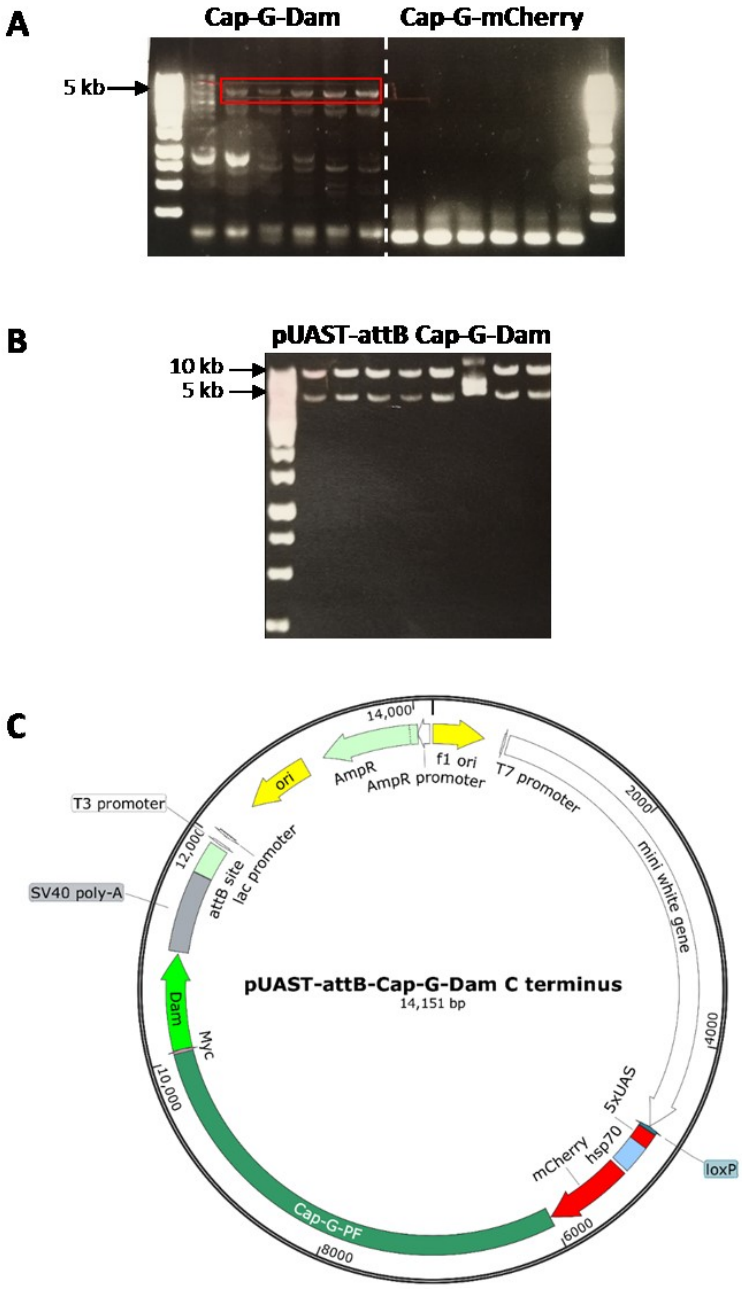


Figure 8.1. UAS-LT3-Cap-G-Dam construct generated

(A) Agarose gel image of successful fusion PCR of *Cap-G-Dam* construct by obtaining the correct band size of 4293 bp. *mCherry-Cap-G* fusion was not successful. **(B)** Agarose gel image of enzyme digest showing successful cloning of the *pUAST-attB-LT3-Cap-G-Dam* vector, further confirmed by sequencing data. **(C)** Diagram of the *pUAST-attB-LT3-Cap-G-Dam* plasmid injected in the embryos.

Appendix

Design of CRISPR/Cas9 Cap-G mutants:

Cap-G gRNA oligos designed:

>2R 2R:9021176..9021198 (+ strand) class=CRISPR length=23
GAAAGCCGTCCTAGGGGCTGAGG
(middle of 2nd exon)

>2R 2R:9021402..9021424 (- strand) class=CRISPR length=23
AGATGCGATAACGTATGTGCGGG
(beginning of 3rd exon)

>2R 2R:9021646..9021668 (- strand) class=CRISPR length=23
ATGATGCACTGACGCACGTTGGG
(middle of 3rd exon)

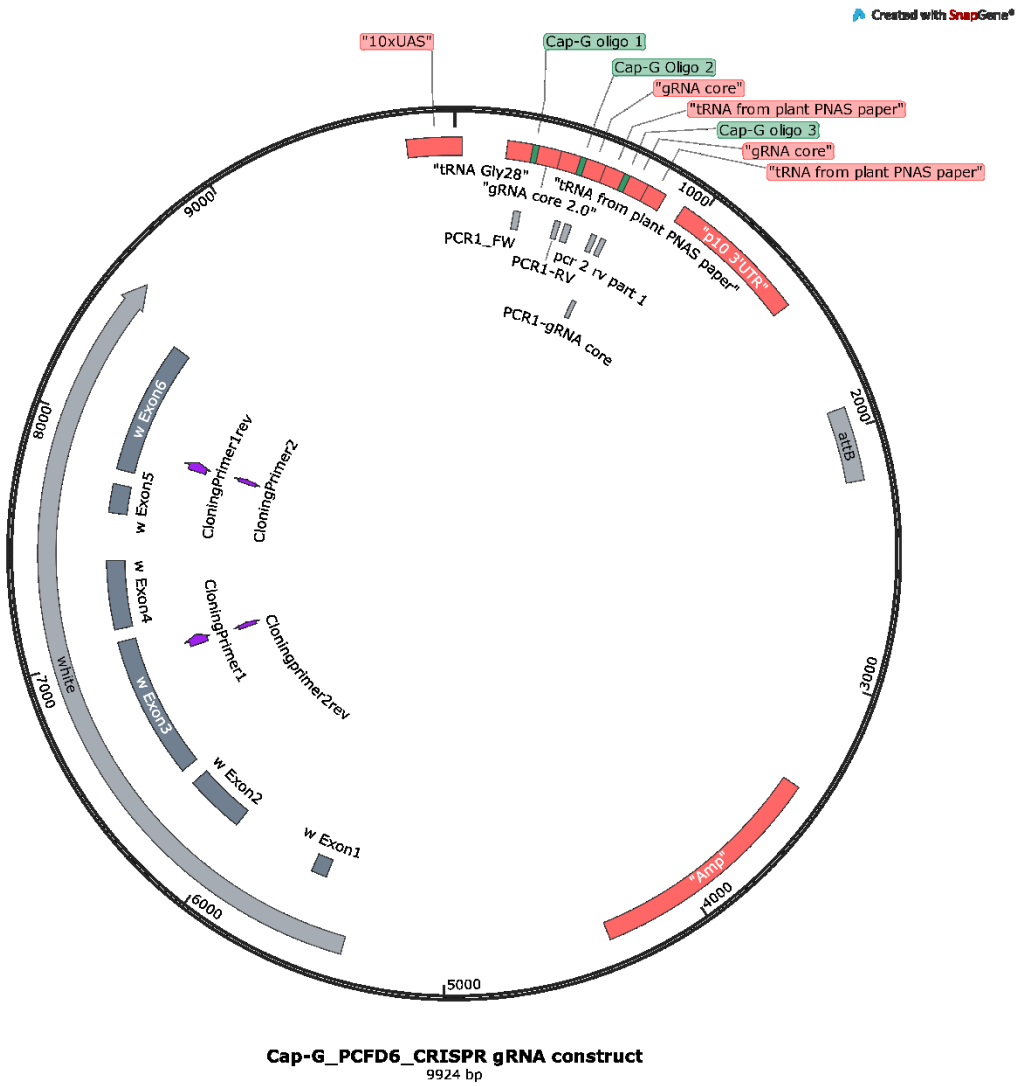


Figure 8.2 UAS-Cap-G-CRISPR/Cas9 mutant plasmid designed

Map of plasmid used to generate *UAS-Cap-G-CRISPR/Cas9* mutants, with three gRNAs designed and inserted between Cap-G exons. Plasmid generated following published protocol (170).

Appendix

Supporting code for sequencing data analysis

LINUX Commands to convert GFF to Bedgraph

Get gff file (output from perl script)

Delete all mitochondrial sequences:

$ grep –v “mito” filename(*write the name of your file)*.gff > output.gff

Remove all columns except for chromasome (1), startcoord (4), end coord(5), and score(6)

$ cut –f1,4,5,6 input(previous output name).gff > output.bedgraph

(or all in one: $ grep –v “mito” filename.gff | cut –f1,4,5,6 > output.bedgraph)

Navigate to folder containing bedgraph file, chromosome sizes file (dm6.chromsizes), and bedgreaphtobigwig executable.

Folder containing: bedgraph exe, chr sizes, output bedgraph to be converted

Type:

./bedGraphToBigWig inputfile.bedgraph dm6.chromSizes outputfile.bw

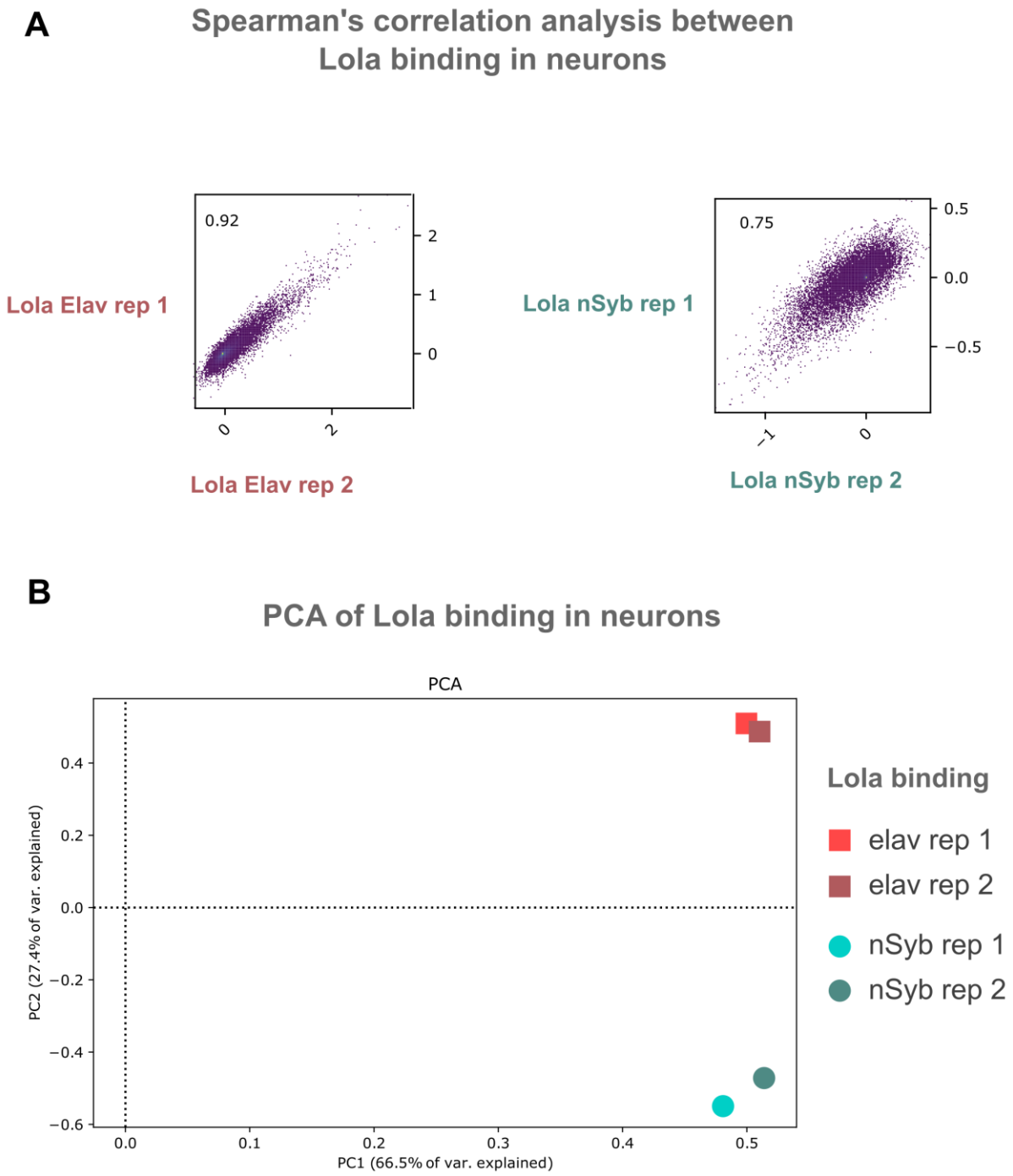


Figure 8.3 **Lola-N binding correlation and QC between elav and nSyb replicates**

Statistical analysis of Lola-N binding in newly born neurons (elav) and mature neurons (nSyb). **A)** Strong positive correlation between replicates for Lola-N mapped binding reads across all cell types (2 per cell type, rho ~ 0.9, 0.75). **B)** Principal Component Analysis (PCA) shows separate clustering of elav and nSyb neuronal stages, whilst replicates of each stage cluster together.