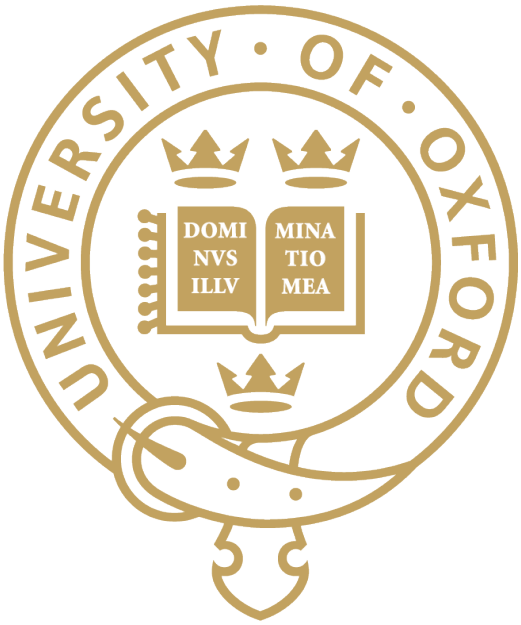
Investigating the transcriptional role of FoxP in decision making in *Drosophila melanogaster*



Laura Jayne Mee Lee Chan Wah Hak

Corpus Christi College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2018

Acknowledgements

I am deeply grateful to everyone who has made this thesis possible and supported me, both scientifically and personally, during the past four years. First and foremost, I would like to thank Gero Miesenböck, who keeps amazing me with his breadth of knowledge and his creativity. My research would not have been possible without his unwavering trust and support and the academic freedom he provided. I would also like to thank Lukas Groschner and Shamik DasGupta, who have mentored me at different stages of my PhD. I am proud of what you have both achieved so far, and I am sure you will be able to 'fly' higher!

I also want to acknowledge past and present members of the Miesenböck lab, for being the best lab siblings I could ask for! I am grateful to Anissa Kempf who has been tremendously helpful during my PhD. Thank you for your endless advice and help. Thanks to Diogo Pimentel, Peter Hasenhuetl and Timothy Wong for always being there for me and for proof-reading this thesis. Katerina Christodoulou, Jan Kropf, Cliff Talbot, Eleftheria Vrontou, Susanne Szydlowski and Laura Garmendia Sanchez- thanks for making the days in lab so fun and enjoyable! I promise I'll get into the ice machine and serve ice lollies one day. Evan Harrell and Michael Song immediately took me under their wings and helped me settle in the lab. Vincent Croset and Christoph Treiber, thank you for teaching the basics of fly pushing and bioinformatics. But above all, thank you for helping me keep my figure by eating most of my food!

This work would not have been possible without the help and support of Fiona Woods,

Ruth Brain, Manrique Moreno Gasulla, Bethan McSparron and Ammerins de Haan. Thank you to Jerome Nicod from the Wellcome Trust for Human Genetics for his help with the DamID-sequencing. This work was supported by a Wellcome Trust OXION studentship.

I am eternally grateful to my parents and my siblings for their unconditional support and love. You have always gone out of your way to make everything possible for me! Even though we live in four different countries, I have always felt like I could count on all of you.

And finally, to Bill, thank you for teaching me how to open ‘Terminal’ and how to code. If it weren’t for you, I would still be looking blankly at my sequencing data. Thank you for my favourite housemate and MarioKart opponent. You have supported me through the good and bad times and have always encouraged me to aim higher. I am excited to see what lies ahead for us.

Abstract

Transcription factors play crucial roles in the expression of genes and thus control the cellular abundance of proteins, which in turn define cell fates, developmental processes, cellular states and ultimately behaviour. One family of transcription factors of special interest in molecular and developmental neurobiology, cognitive science, linguistics and evolutionary anthropology is the evolutionarily conserved forkhead box P (FoxP) proteins, which have been implicated in vocal communication, motor learning, cognitive development, and decision-making. To fully understand the role of FoxP family members in these processes, it is important to analyse how they impact cellular functions through the regulation of gene expression.

At the molecular level, the study of transcription factors involves identifying their target genes and specific binding patterns, analysing the effects of transcription factor-DNA interactions on gene expression and understanding gene regulatory networks. The *Drosophila* FoxP gene gives rise to 2 isoforms, of seemingly differential expression. Elucidating the binding patterns, the putative targets and the gene regulatory networks of each isoform will give an insight into why the isoforms are differentially expressed and how the different isoforms influence different behaviours, such as decision-making, differentially.

In order to analyse the putative targets of FoxP in the *Drosophila* brain, targeted DNA adenine methyltransferase identification (targeted DamID or TaDa), coupled with next generation sequencing was carried out. The binding profiles were consistent with FoxP regulating gene expression at the level of transcription by acting on silencers and promoter

regions of its target genes. Consensus FoxP binding sites were determined by analysing for enrichment of common motifs in FoxP-bound regions. Targeted DamID identified about 600 targets for each isoform, with less than 50% overlapping genes. Most of the putative targets of both isoforms play crucial roles in the nervous system, in agreement with the known or suspected roles of FoxP. These targets were confirmed with a complementary method, chromatin immunoprecipitation (ChIP) *in vitro* in neuronal cells and *in vivo* in *Drosophila* brains. Among the targets of the FoxPC isoform is the gene encoding the voltage-gated potassium channel Shal, confirming biophysical evidence that FoxP sets neuronal integration and behavioural decision times by controlling the abundance of Shal in a small population of FoxP-positive, decision-critical neurons.

One of the major challenges of transcription factor studies is that binding sites are common in the genome and binding per se does not necessarily have a functional effect. The functional consequences of the transcription factor-DNA binding were tested by luciferase reporter assays and translating ribosome affinity purification (TRAP). Like its mammalian orthologues, *Drosophila* FoxP, acts as a transcriptional repressor, decreasing mRNA levels of the most highly-bound genes.

Since most transcription factors work in a cooperative manner to regulate gene expression, potential FoxPC cofactors were identified to allow a more in-depth understanding of gene regulation, especially in the decision-making process. Yeast two-hybrid and co-immunoprecipitation assays identified three transcription factors, glial cells missing (gcm), Sp1-like factor for pairing sensitive-silencing (Spps) and Dichaete (D) as potential interaction partners. ChIP and luciferase assays confirmed that FoxPC cooperates with these transcription factors to regulate its target genes. Together, these results offer not only

a global view of gene expression control by FoxP but also a molecular explanation for the decision phenotype of *FoxP* mutant flies.

Acknowledgements **2**

Abstract **4**

Figures **10**

Tables **11**

Abbreviations **12**

1 Introduction **14**
1.1 Gene expression and regulation *14*
1.2 Transcription factors *15*
1.2.1 Mechanisms of gene expression control by transcription factors 16
1.2.2 Structure of transcription factors 18
1.2.3 Regulation of transcription factor activity 21
1.2.4 Combinatorial Control of gene expression 22
1.2.5 Families of transcription factors 23
1.2.6 Forkhead transcription factors 24
1.2.7 Forkhead box P family, FoxP, in vertebrates 25
1.2.8 FoxP in Drosophila melanogaster 26
1.3 Aims and objectives *30*

2 Methods **31**
2.1 Fly Strains *31*
2.2 Cell Culture *32*
2.3 Targeted DNA adenine methyltransferase identification *32*
2.3.1 Plasmid construction 32
2.3.2 Targeted DamID 34
2.3.3 Sequencing data quality control 35
2.3.4 Sequencing data analysis 36
2.3.5 Overlap with known CRMs, TFBSs and inferred motifs 38
2.3.6 Motif Analysis 38
2.3.7 Gene Ontology Enrichment Analysis 39
2.3.8 Pathway Analysis 39
2.3.9 Motif occurrence by MAST 40
2.4 Chromatin immunoprecipitation (ChIP) *40*
2.4.1 Plasmid Construction 40
2.4.2 ChIP in ML-DmBG3-c2 neuronal cells 41
2.4.3 ChIP in $\alpha\beta_c$ neurons in Drosophila brains 43
2.5 Translating ribosome affinity purification *44*
2.5.1 RNA extraction, reverse transcription and real-time PCR 45
2.6 Dual luciferase reporter assays *46*
2.6.1 Plasmid construction 46
2.6.2 Luciferase assay 47
2.7 Electrophoretic Mobility Shift Assays (EMSA) *48*
2.8 In vitro transcription/ translation and co-immunoprecipitation *49*
2.8.1 Construct preparation 49
2.8.2 Cell-free protein expression 49
2.8.3 Western Blotting 50
2.9 Electrophysiology *51*
2.10 Two- odour discrimination tasks *52*

3 Identifying putative targets of FoxP in the brain **54**
Chromatin immunoprecipitation 54

DNA adenine methyltransferase identification 55
3.1 Aims and objectives *59*
3.2 Results *60*
3.2.1 Validation of the targeted DamID protocol 60
3.2.2 Analysis of DamID-sequencing data 61
3.2.3 Association of FoxPC and FoxPD occupancy to genomic features 68
3.2.4 Motif Identification 73
3.2.5 Characterisation of target genes 77
3.2.6 Identification of pathways, molecular functions and networks of target genes 78
3.2.7 Pathway analysis 86
3.3 Interim Discussion from Chapter 3 *87*

4 Confirming FoxPC targets identified by DNA Adenine Methyltransferase Identification 90
4.1 Aims and objectives *94*
4.2 Results *95*
4.2.1 Analysis of FoxPC binding to promoter regions 97
4.2.2 Analysis of the regulation of targets by FoxPC by chromatin immunoprecipitation 102
4.2.3 Analysis of the regulation of targets by FoxPC in luciferase reporter assays 106
4.2.4 Confirmation of the FoxPC binding motif by electrophoretic mobility shift assays 110
4.2.5 Confirmation of the FoxPC binding motif in luciferase reporter assays 112
4.3 Interim Discussion from Chapter 4 *113*

5 FoxPC regulates its targets as part of a complex 115
5.1 Aims and Objectives *118*
5.2 Results *119*
5.2.1 Identification of putative interaction partners 119
5.2.2 Confirmation of putative interaction partners 120
5.2.3 Analysis of expression levels of interaction partners in FoxPC-positive neurons 122
5.2.4 Analysis of FoxPC transcription complexes binding to target genes 123
5.2.5 Determination of locations of binding motifs in promoter regions of target genes 126
5.2.6 FoxPC transcription complexes and gene regulation 127
5.2.7 FoxPC transcription complexes and biophysical properties of FoxPC-positive cells 130
5.2.8 FoxPC transcription complexes and decision-making 131
5.3 Interim Discussion from Chapter 5 *133*

6 Discussion 137
6.1 FoxP regulates a plethora of genes *137*
6.1.1 Identification of FoxPC and FoxPD targets 141
6.1.2 Comparison to known targets 145
6.1.3 Linking FoxP targets to known behavioural phenotypes 146
6.1.4 FoxPC and signalling pathways 148
6.2 Transcription regulation by FoxP *148*
6.2.1 FoxPC's mechanisms of action 150
6.3 FoxPC regulates its targets as part of a complex *153*
6.4 Conclusions *162*
6.5 Future directions *162*

References 164

Appendix 179
Appendix 1. Number of reads from each sequencing replicate *179*
Appendix 2. FastQC reports of each replicate *180*
Appendix 3. List of candidate CRMs bound by FoxPC and FoxPD *183*
Appendix 4 List of target genes for FoxPC and FoxPD *184*

Appendix 5. List of ontology terms associated with target genes of FoxPC and FoxPD *191*
Appendix 6. Primer List *193*

Figures

Figure 1. Cis-regulatory elements determine gene transcription. 17
Figure 2. Mechanisms by which transcription factors activate or repress transcription. 20
Figure 3. Drosophila FoxP isoforms. 27
Figure 4. FoxP-mutant flies have increased reaction times in an olfactory discrimination task. 28
Figure 5. Proposed mechanism. FoxP represses Shal to regulate dendritic signal propagation. 29
Figure 6. Schematic of ChIP and DamID pipeline. 57
Figure 7. Successful amplification of adenine-methylated fragments. 60
Figure 8. Flowchart of the DamID-seq analysis pipeline. 62
Figure 9. Distribution of mapped reads at a specific region of genome for the three replicates. 63
Figure 10. Correlation between the biological replicates. 65
Figure 11. Binding of FoxPC and FoxPD is enriched at specific sites compared to controls. 67
Figure 12. Assigning binding intervals to genomic features. 70
Figure 13. Metagene profiles of FoxPC and FoxPD binding. 71
Figure 14. i-cisTarget motif analysis of FoxPC and FoxPD datasets. 73
Figure 15. Motif Analysis using MEME. 75
Figure 16. Comparison of FoxPC putative motif with known motifs by TomTom. 76
Figure 17. FoxPC and FoxPD regulate common and unique targets. 78
Figure 18. FoxPC and FoxPD targets are implicated in various functional processes. 79
Figure 19. Breakdown of the signaling and nervous system development GO categories. 80
Figure 20. Heatmaps of the genes contributing to the most overrepresented GO categories. 84
Figure 21. FoxPC binds only 8 ion channel target genes. 85
Figure 22. FoxPC regulates a large number of genes involved in nervous system development. 87
Figure 23. Schematic of the TRAP protocol. 92
Figure 24. Dual luciferase reporter assay schematic. 93
Figure 25. Binding intervals at several FoxPC target loci. 95
Figure 26. MAST output for the locations of the FoxPC putative motif. 96
Figure 27. Tagging FoxPC with an HA tag does not affect its cellular location. 98
Figure 28. Chromatin immunoprecipitation in ML-BG3-C2 neuronal cells. 100
Figure 29. Chromatin Immunoprecipitation in Drosophila brains 102
Figure 30. Expression of FoxPC and its targets in the $\alpha\beta_c$ neurons. 103
Figure 31. Relative transcript levels in the $\alpha\beta_c$ neurons. 105
Figure 32. FoxPC represses the activities of its target promoter regions. 107
Figure 33. FoxPC DNA sequence. 108
Figure 34. Mutating the FoxPC binding domain abrogates repressive action of FoxPC on promoter regions. 109
Figure 35. The wild-type FoxPC DNA binding domain is required for binding to the FoxPC motif. 111
Figure 36. Mutating the FoxPC binding motif affects the regulatory action of FoxPC on its target genes. 112
Figure 37. Principles of the yeast two-hybrid system. 118

Figure 38. Y2H assay uncovers direct interactions between FoxPC and 5 transcription partners. 120
Figure 39. FoxPC binds to gcm, Spps and D. 121
Figure 40. D, gcm and Spps are expressed in FoxPC-positive $\alpha\beta_c$ neurons labelled by NP7175-GAL4. 123
Figure 41. Chromatin immunoprecipitation in Drosophila brains suggests that FoxPC binds to its target promoters as part of a complex. 125
Figure 42. FoxPC and interaction partners' motifs locations. 127
Figure 43. FoxPC forms complexes with different transcription factors to regulate promoter activity. 129
Figure 44. FoxPC, Spps and gcm repress an A-type potassium current in $\alpha\beta_c$ KCs. 131
Figure 45. gcm and Spps in the $\alpha\beta_c$ neurons using NP7175-GAL4 driver line determine reaction times. 133
Figure 46. Approaches for deciphering the FoxP transcriptional regulatory network. 139
Figure 47. Chromosome distribution of FoxPC and FoxPD target genes. 141
Figure 48. Shal is the only voltage-gated potassium channel bound by FoxPC. 147
Figure 49. Presence and locations of FoxPC and FoxPD motifs on target genes. 149
Figure 50. Network motifs of transcriptional regulatory networks. 154
Figure 51. FoxPC binds to numerous transcription factors. 156

Tables

Table 1. Number of peaks with an FDR < 0.01 from the fusion-Dam experiments. 66
Table 2. Summary of the genes tested for their FoxPC occupancy by DamID-seq and the presence of the FoxPC putative motif. 97
Table 3. Known gcm targets. 159

Abbreviations

acj6	Abnormal chemosensory jump 6
ANOVA	Analysis of variance
Bp	Base pairs
ChIP	Chromatin immunoprecipitation
CRM	Cis-regulatory module
Dam	DNA adenine methyltransferase
D	Dichaete
DamID	DNA adenine methyltransferase identification
DamID-seq	DNA adenine methyltransferase identification- sequencing
DNA	Deoxyribonucleic acid
EMSA	Electrophoretic mobility shift assay
FDR	False discovery rate
FoxP	Forkhead box protein P
FoxPC	FoxP protein isoform C
FoxPD	FoxP protein isoform D
gcm	Glial cells missing
GST	glutathione-S-transferase
HA	Human influenza hemagglutinin
Kb	Kilo base pair
MAST	Motif Alignment & Search Tool

MEME	Multiple Em for Motif Elicitation
PCR	Polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
s.e.m	Standard error of the mean
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Shal	Shaker cognate l
spps	Sp1-like factor for pairing sensitive-like factor silencing
TFBS	Transcription factor binding site
TRAP	Translating Ribosome Affinity Purification
TSS	Transcription start site
UAS	Upstream activation sequence
Y2H	Yeast two- hybrid assays
$\alpha\beta_c$	αβ core

Throughout the text, nucleotides (genes and mRNA) are abbreviated according to standard nomenclature in the literature (in italics), and proteins are in roman type.

1

Introduction

1.1 Gene expression and regulation

The nature and development of a living organism are determined by its genes. The genetic content but also the spatial and temporal expression of genes are crucial. Specific expression of genes leads to cellular differentiation, cellular proliferation, developmental patterning and, in adults, regulates the adaptability of living organisms to changes in their external and internal environment. In turn aberrant expression of genes is responsible for a wide range of diseases (Lee and Young, 2013), ranging from cancers (Chen et al., 2002, Makiyama et al., 2005) to neurological disorders (Ding et al., 2008, Kaufmann et al., 2010b).

The fundamental dogma of molecular biology is that deoxyribonucleic acid (DNA) produces ribonucleic acid (RNA), which in turn leads to the production of proteins (Crick, 1970). The process of converting DNA into RNA, known as transcription, is therefore the crucial first step in determining the characteristics or phenotypes of cells. Although gene expression can be controlled post-transcriptionally, for example by RNA splicing and stability (Bashirullah et al., 2001), the most important control point lies at the level of transcription, as it defines the messenger RNA pool available in cells at a given time (Latchman, 1990, Darnell, 1982). Transcription is therefore subject to various regulatory controls, with the three major steps being transcriptional regulation by specific proteins,

regulation of transcript prevalence and induction of chromatin modifications upon protein-DNA binding (Bannister and Kouzarides, 2011).

Transcription is initiated by the assembly of the initiation machinery, which includes RNA polymerase and general initiation factors such as TFIIB and TFIID. RNA polymerases recognise and bind upstream of the transcription start site, at the core promoter region, to begin transcription (Butler and Kadonaga, 2002). After locating the transcription start site, they use one strand of the DNA molecule as a template to generate complementary RNA copies. RNA polymerases are unable to recognise genes and initiate transcription in eukaryotes on their own. To be able to do so, they require specialised proteins called transcription factors. Therefore, transcription factors are crucial in regulating the spatial and temporal specificities of transcription.

1.2 Transcription factors

Transcription factors make up between 0.5% to 8% of the genetic content of eukaryotic genomes (Levine and Tjian, 2003). They can be divided into two broad classes: basal transcription factors which are required for RNA polymerase recruitment for baseline transcription, and specific transcription factors, which are sequence-specific proteins that function to regulate cellular gene expression in response to internal and environmental (external) cues. These proteins bind to specific sequences upstream (5' end) or downstream of the transcription start site, hence modulating chromatin conformation and transcription initiation. Transcription factors recognise and bind to specific DNA sequences directly or via other transcription factors and can therefore recruit RNA polymerases to promoter elements, such as the TATA box, of target genes. They can

induce changes to the structure of chromatin around the transcription start site by interacting with and recruiting chromatin remodelling complexes (Szutorisz et al., 2005). Transcription factors selectively recognise different DNA sequences, or motifs, which are present in sequences found close to the transcription start site. Each transcription factor therefore regulates a specific set of genes (Triezenberg, 1995), allowing for the fine-tuning of gene expression, by positively or negatively influencing the expression of specific genes. Binding of transcription factors to DNA might alter chromatin structure, but this is not sufficient for regulating transcription. Once bound to DNA, transcription factors can act as activators by facilitating the recruitment and stabilisation of the basal transcription machinery. Alternatively, they function as repressors either by preventing the binding or stabilisation of the transcription machinery, or by inhibiting the action of activators.

1.2.1 Mechanisms of gene expression control by transcription factors

The eukaryotic transcriptional machinery consists of two types of regulatory elements: *cis*-acing and *trans*-acting elements. *Cis*-acting elements are DNA sequences in the vicinity of genes, which contain binding sites for *trans*-acting proteins such as transcription factors. The cis-regulatory sequences are complex and differ from gene to gene. They include proximal elements or promoters and more distal elements such as enhancers, silencers and repressors (Figure 1). The RNA polymerase II (Pol II) and its cofactors bind to the core promoter regions, which is part of the proximal region and which contains the transcription start site, the TATA box (TATA), the TFIIB recognition element (BRE), the initiator (INR), the motif ten element (MTE) and the downstream promoter element (DPE) (Smale and Kadonaga, 2003). All of these sequences are required for the formation of the basal transcriptional machinery. CpG islands (CpG) are part of proximal promoter elements upstream of the core promoter and are therefore important in transcription

initiation. Enhancer elements can be found upstream or downstream of the transcription start site. Enhancers usually contain multiple transcription factor binding sites and increase the activities of the promoters, resulting in the upregulation of gene expression by several mechanisms, for example by looping out DNA (Ptashne and Gann, 1997) or changing the DNA conformation (Freeman and Garrard, 1992). Silencers have the opposite effect to enhancers. They can be found in the proximal promoter or distal enhancer. Insulators ensure that genes are only transcribed by their regulatory elements and are not affected by their neighbouring genes' elements (Zhao and Dean, 2004). Locus control regions are a group of enhancers, silencers and insulators, which collectively function to regulate gene expression.

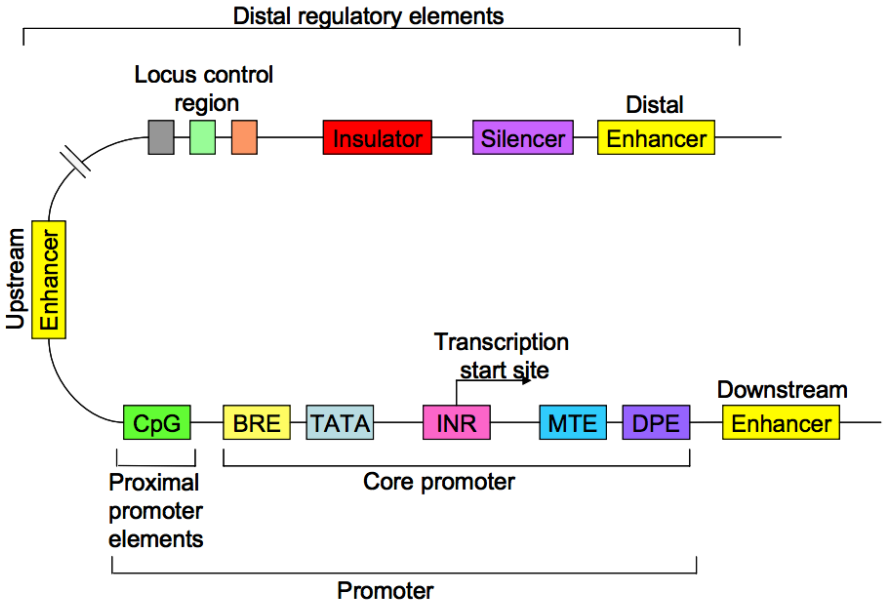


Figure 1. Cis-regulatory elements determine gene transcription.

The core promoter region consists of several sequences- TFIIB recognition sequence (BRE), TATA box (TATA), initiator (INR) encompassing the transcription start site, motif ten element (MTE) and downstream promoter element (DPE), required for the initiation of transcription and the formation of the basal transcription complex. CpG islands together with the core promoter elements form proximal promoter elements. Distal regulatory elements such as enhancers, silencers, locus control regions and insulators can be found upstream or downstream of the transcription start sites and work together to activate or repress proper regulation gene expression.

1.2.2 Structure of transcription factors

Transcription factors have several structural elements that allow them to regulate gene expression. They have a modular structure, comprising a DNA-binding domain and a transcriptional regulatory domain (Collado-Vides et al., 1991). The DNA-binding domain is usually highly conserved, while the regulatory domain is variable (Schwechheimer et al., 1998). The DNA-binding domain recognises specific DNA motifs in the promoter regions of target genes, allowing the selective activation or repression of transcription.

In transcriptional activators, the regulatory domain is an activation domain. The domain directly or indirectly recruits components of the basal transcriptional machinery (RNA polymerase and general transcription factors) or enhances its formation. Activators can also make DNA more accessible for transcription (for example, by promoting acetylation of histone proteins) (Cosma et al., 1999) (Figures 2A and 2B). Transcriptional activators usually recruit other proteins or cofactors to accomplish their functions.

Although most transcription factors act to upregulate transcription, some, known as transcriptional repressors, have an inhibitory effect on transcription. They can either reduce the rate of transcription or completely suppress transcription, in several ways. Firstly, binding of repressors can sterically prevent binding of the transcriptional machinery or activators to their recognition motifs. Secondly, repressors can bind to the general transcription factors and other activators, preventing their recruitment to the transcription start site. Lastly, transcriptional repressors can alter higher-order DNA structure by promoting a tightly packed chromatin structure which does not allow binding of activators (Hanna-Rose and Hansen, 1996) (Figure 2C-2E).

In addition to the DNA-binding domain and the trans-activating domains, some transcription factors also have a protein-protein interaction domain, which allows binding of other transcription factors. This facilitates crosstalk between transcription factors and modulation of their activities. Cooperation among transcription factors is highly common in eukaryotes and provides another level of regulation.

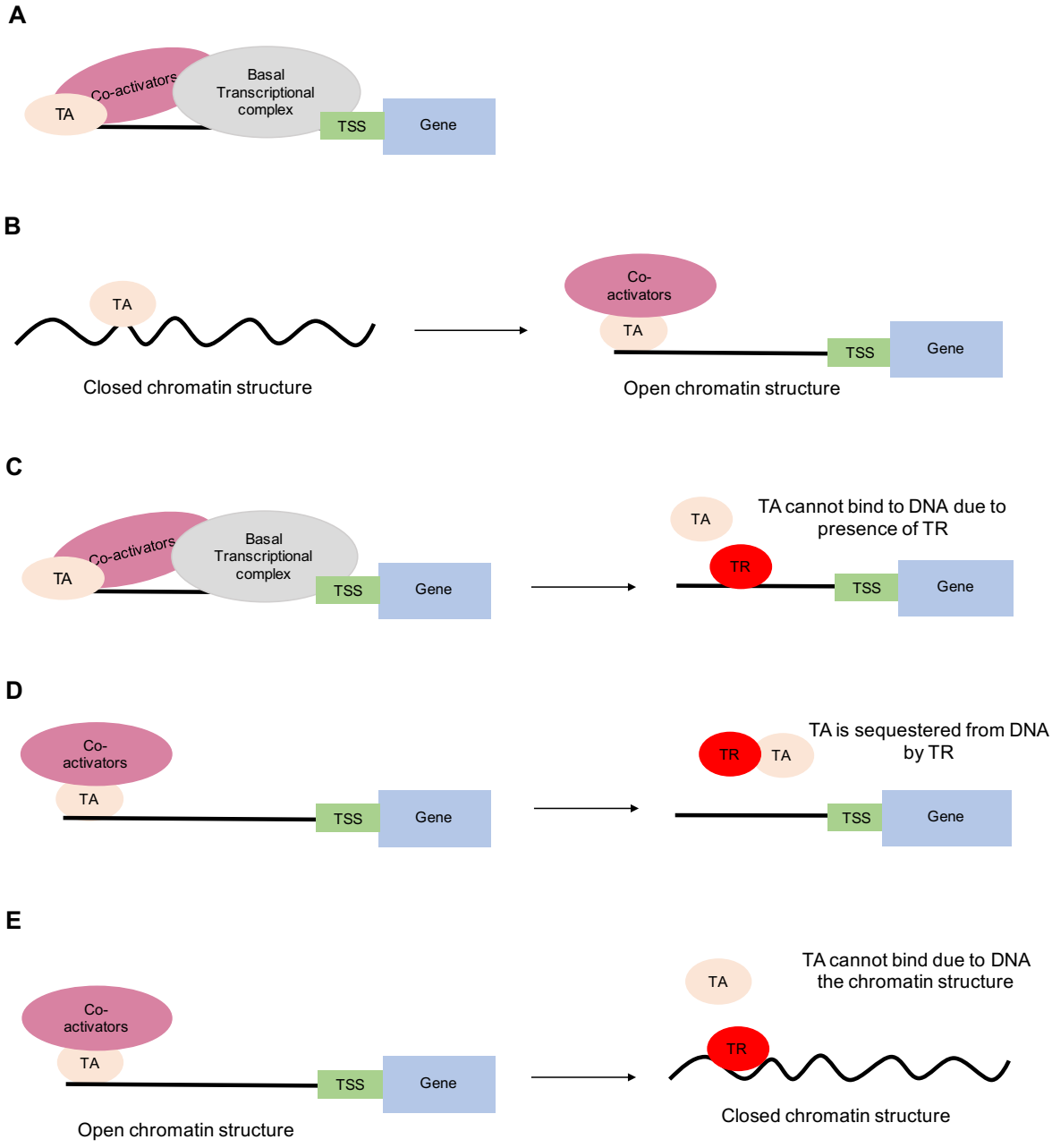


Figure 2. Mechanisms by which transcription factors activate or repress transcription.

*Transcriptional activators (TA) can either act as a mediator between the other activators and the basal transcriptional complex (**A**) or can promote the conversion of compact chromatin structure to a more open structure allowing further recruitment of co-activators (**B**). Inhibitory transcription factors or repressors (TR) repress transcription by binding to the binding site of an activator, hence preventing the latter's binding (**C**). TRs can also sequester TAs, preventing their binding to DNA (**D**). TRs can also prevent binding of activators by modifying the DNA structure (**E**).*

1.2.3 Regulation of transcription factor activity

Rapid responses to internal and external changes are crucial for all organisms. Various signalling pathways are activated to relay external information from the cell surface to the nucleus, with transcription factors often being the end targets (Tootle and Rebay, 2005). Errors in the regulation of transcription factors often lead to diseases as several pathways tend to be affected. To maintain proper regulation of gene expression, transcription factors themselves are also tightly regulated via several mechanisms, including post-translational modifications, multimeric interactions with other partner transcription factors and cooperation with coregulators (Latchman, 1997, Siggers et al., 2011, Salghetti et al., 2001). These mechanisms ensure that transcription factors are spatially and temporally controlled.

Post-translational modifications such as acetylation, methylation and phosphorylation affect the functions of the transcription factors. They might alter their DNA binding affinities (Elenkov et al., 2011) and their cellular locations, and therefore their availability in the nucleus (Parker et al., 2003).

Coregulators do not bind to DNA, but rather to transcription factors, and can regulate their activating or repressing activities, for example by modifying the structure of chromatin, by recruiting the basal transcriptional machinery or by facilitating the unwinding of DNA before transcription (Ugrinova et al., 2009).

1.2.4 Combinatorial Control of gene expression

In eukaryotes, transcription factors recognise sequences that are too short to define unique genomic positions (Morgunova and Taipale, 2017). In addition, the number of transcription factors is much smaller than the number of genes in the genome to allow a one-to-one relationship. Precise gene regulation is achieved by cooperativity between different transcription factors, thus increasing the number of possible combinations. Each eukaryotic gene is under the control of several transcription factor binding sites, or *cis*-regulatory elements. Although one transcription factor may regulate a single process, it is likely to function with different sets of transcription factors at different times at the same *cis*-regulatory module to regulate RNA polymerase activity and to enable different expression patterns (Zinzen et al., 2009, MacArthur et al., 2009). This combinatorial control of gene expression allows complex information processing at individual regulatory elements and provides greater diversity, allowing the same transcription factors to regulate different processes (Karczewski et al., 2011).

Transcription factors can cooperate in different ways. Often, they bind weakly to DNA as monomeric proteins. By binding to each other via protein-protein interaction domains before binding to DNA, the protein-DNA binding is enhanced (Morgunova and Taipale, 2017). In such cases, the DNA recognition sequences of the interacting transcription factors are often located close to each other. In other cases, the transcription factors' interactions are facilitated or mediated by DNA. Binding of one transcription factor to DNA changes the shape or dynamics of DNA, facilitating the binding of other transcription factors (Tahirov et al., 2001).

1.2.5 Families of transcription factors

Transcription factors are grouped together in families based on the structural similarities of their sequence-specific DNA-binding domain (Pabo and Sauer, 1992) and their interaction modes with DNA. The five major classes of transcription factors are helix-turn-helix, helix-loop-helix, zinc finger, basic protein-leucine zipper and β-sheet types.

The helix-turn-helix DNA-binding domains (for example homeodomains) consist of a tri-helical bundle, with the third 'recognition' helix binding to DNA. Variants of the helix-turn-helix domain include the tetra-helical bundle and winged-helix (with two large flexible loops in addition to the helix-turn-helix) domain (Zheng et al., 1999). The latter is present in the forkhead class of transcription factors.

Basic helix-loop-helix proteins have two conserved and functionally distinct domains: the basic domain, which binds to DNA, and the helix-loop-helix domain, which facilitates interactions with other proteins (Jones, 2004).

Members of the leucine zipper transcription factor family include c-Myc and MafB and consist of a highly conserved leucine zipper domain. Leucine zipper domains are dimeric motifs, made up of a basic region which recognises and binds to DNA and the leucine zipper, which is composed of heptad repeats of leucine residues which mediate dimerisation of leucine zipper proteins (Hurst, 1995).

Zinc finger motif-containing proteins, for example TFIIA and Gata4, are the most abundant transcription factors in eukaryotes (Lander et al., 2001). The zinc finger motif has highly adaptable DNA binding specificity, unlike other DNA binding motifs

(Desjarlais and Berg, 1992). The zinc finger domain is made up of a short α-helix and two β-sheets held together by a zinc ion.

β-scaffold motifs use large β-sheet surfaces which recognise and bind to the minor groove of DNA. β-sheet-containing proteins include TATA binding proteins and HMG-box proteins.

1.2.6 Forkhead transcription factors

Forkhead transcription factors have a conserved 100-amino acid, winged helix (or forkhead) DNA-binding domain, with three β-sheets, three α-helices and two loops forming the wings that flank the third β-sheet (Myatt and Lam, 2007). The forkhead domain is evolutionarily conserved from yeast to mammals, with at least 4 proteins in yeast, 16 in *Drosophila melanogaster* and 50 in humans. In addition to the conserved winged helix, forkhead box (Fox) proteins also contain highly divergent transactivation or transrepression domains.

Fox proteins are grouped into 19 subclasses (FOXA-FOXS), depending on homology within and outside the forkhead domain (Kaestner et al., 2000). Even though the forkhead domain is highly conserved, the other protein domains differ quite substantially between members of different subgroups. They can therefore regulate different gene networks and affect gene transcription in different ways: for example, FOXM1 and FOXO3A bind to the same DNA motif and have opposing effects on gene transcription (Gomes et al., 2013). Fox proteins have highly diverse regulatory roles in cellular process, for example signalling, development, organogenesis and language and speech acquisition (Golson and Kaestner, 2016, Le Lay and Kaestner, 2010, Takahashi et al., 2009). As such, Fox proteins

are implicated in various disorders, such as cancer (Kim et al., 2006, Zou et al., 2008) and language and speech disorders (Lai et al., 2001).

1.2.7 Forkhead box P family, FoxP, in vertebrates

The FoxP family differs from the other Fox subfamilies, as it has a C-terminal forkhead domain, instead of an N-terminal one (Chu et al., 2011). Members of the FoxP family also share similar structural motifs, for example a glutamine-rich region involved in transcription activation or repression, and a leucine zipper and a zinc finger which might be involved in homo- and heterodimerisation (Wang et al., 2003). The forkhead domain allows FoxP to further form dimers by domain swapping (Medina et al., 2016). In vertebrates, the FoxP subfamily consists of four members (FoxP1-4), with each member being involved in diverse biological processes. FoxP1 is implicated in macrophage development (Sagardoy et al., 2013) and cognitive development (Hamdan et al., 2010); FoxP2 is crucial in language and speech development (Lai et al., 2001), motor control (Enard et al., 2009) and cognitive development (Lieberman, 2009); FoxP3 is involved in the adaptive immune response (Yagi et al., 2004) and FoxP4 regulates islet cell development and function. FoxP1 and FoxP2 often have overlapping expression patterns in the brain, but are sometimes present in specific locations. For example, FoxP2 is highly expressed in the cerebellum, while the highly homologous FoxP1 is highly expressed in the hippocampus (Ferland et al., 2003, Teramitsu et al., 2004). When the two factors are expressed in the same areas of the brain, they tend to heterodimerise (Li et al., 2004) and hence function in a coordinated manner to regulate transcription. Even though most of the Fox proteins activate transcription, some members of the FoxP family are known to function as transcriptional repressors (Lopes et al., 2006, Shu et al., 2001).

Each *FoxP* member is subject to tissue-specific alternative splicing of functionally important domains, giving rise to different isoforms (for example FoxP1 has 11 isoforms and FoxP2 has 8 isoforms). This suggests that the same protein is differentially expressed in different tissues and can perform different functions in different cells. This alternative splicing mechanism is also present in *Drosophila melanogaster* (Santos et al., 2011). The sequence conservation of the forkhead box and the splicing mechanism in distant organisms such as insects and mammals suggest crucial roles for the different isoforms.

1.2.8 FoxP in Drosophila melanogaster

The majority of the bilateria *Fox* gene families show orthologous relationships with *Fox* genes in other taxa. The *D. melanogaster* sequence of *FoxP* is highly similar to the vertebrate *FoxP2* gene. This single *FoxP* gene consists of six constitutively and two alternatively spliced exons that give rise to a minimum of two transcripts (*RC* and *RD*) (Figure 3) and corresponding protein isoforms, PC and PD (hereafter referred to as FoxPC and FoxPD) (Santos et al., 2011). The two isoforms share an N-terminal zinc finger domain but differ in their C-terminal, DNA-binding forkhead box motifs.

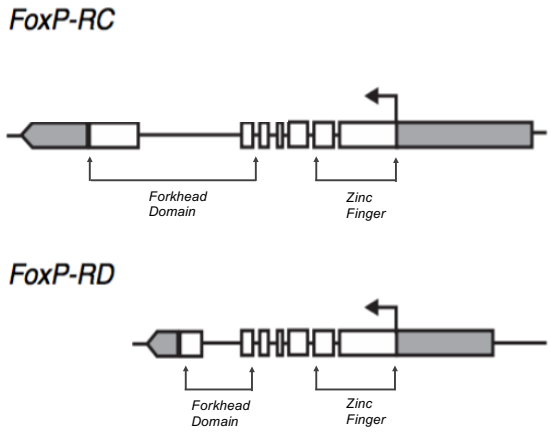


Figure 3. Drosophila FoxP isoforms.

Alternative splicing of the FoxP gene gives rise to at least two distinct transcripts (RC and RD) and corresponding protein isoforms (PC and PD). The zinc finger and the forkhead domains are encoded by exons 1 and 2 and exons 6 and 7, respectively.

FoxP is expressed at all stages of development, with both isoforms thought to be present in the central nervous system (CNS), but only isoform D being present in hemocytes (Santos et al., 2011). This suggests that, like in humans, alternative splicing of the *D. melanogaster* orthologue FoxP could lead to differential expression and function of the two isoforms. In humans, isoforms of human FOXP1, FOXP1 and FOXP1-ES, which differ in their forkhead DNA-binding domains regulate different sets of target genes (Gabut et al., 2011). Like its neuronally expressed vertebrate orthologues, *Drosophila* FoxP has been suggested to play a crucial role in brain processes. Lawton et al. have proposed a role for FoxP in courtship song production (Lawton et al., 2014) and Mendoza et al. in operant conditioning, and learning (habit formation) (Mendoza et al., 2014).

DasGupta et al. found that FoxP affects reaction times in perceptual decision-making, providing a link between FoxP and temporal processing in the brain (DasGupta et al., 2014). FoxP, and especially the FoxPC isoform, is expressed in decision-relevant third-order olfactory neurons of the mushroom body, the so-called Kenyon cells (KCs)

(DasGupta et al., 2014, Groschner et al., 2018). Of the approximately 2000 KCs, a small, genetically distinct population of about 80 neurons per hemisphere in the $\alpha\beta_c$ division (Aso et al., 2014, Tanaka et al., 2008) is critical for perceptual decisions (DasGupta et al., 2014) (Figure 4). The biophysical properties of these neurons—in particular, their ability to integrate synaptic inputs over time—determine the reaction times of flies performing odour discrimination tasks (Groschner et al., 2018). *FoxP* mutants take longer than wild-type flies to discriminate odour intensities with similar accuracy (Figure 4B), suggesting that the process of evidence accumulation over time is impaired.

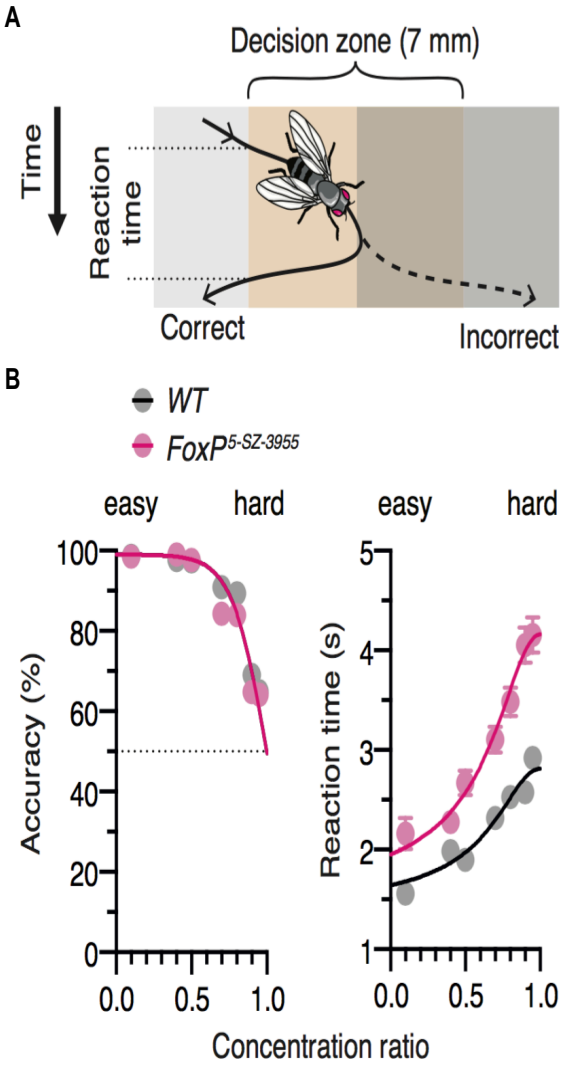


Figure 4. FoxP-mutant flies have increased reaction times in an olfactory discrimination task.

A, Measurement of decision accuracies and reaction times. Two odour streams converge in a 7-mm wide decision zone (orange) at the centre of a narrow chamber. Flies are trained to avoid a specific concentration of MCH using electric shock and must discriminate the reinforced concentration from a lower MCH concentration (2-18ppm); the concentration ratio during testing determines the difficulty of the task. The accuracy of a decision is determined by which side of the chamber the fly exits the decision zone and the time spent inside the decision zone is quantified as the reaction time. ***B****, Flies with a homozygous loss-of-function mutation in the FoxP gene (FoxP5-SZ-3955) take longer than wild-type (WT) flies to form decisions of similar accuracy. Experimental data in panel B are from DasGupta et al., 2014, reprinted with permission from AAAS.*

This function of FoxP in tuning neural circuits for processing time-varying sensory signals creates an interesting avenue for exploration. Since FoxP is a transcription factor, it is expected to modulate the expression of genes required for temporal information processing and hence provides a potential route for dissecting the underlying neural pathways. One plausible mechanism by which FoxP influences the speed of perceptual decision in *Drosophila* is by controlling the abundance of the voltage-gated potassium channel Shal (K_v4) (Groschner et al., 2018) (Figure 5). Evidence from calcium imaging, patch clamp electrophysiology and behavioural experiments suggests that Shal is overexpressed in mutants but direct molecular evidence is still lacking (Groschner et al., 2018). It is also unclear whether FoxP regulates target genes other than Shal, and whether it acts in combination with other transcription factors.

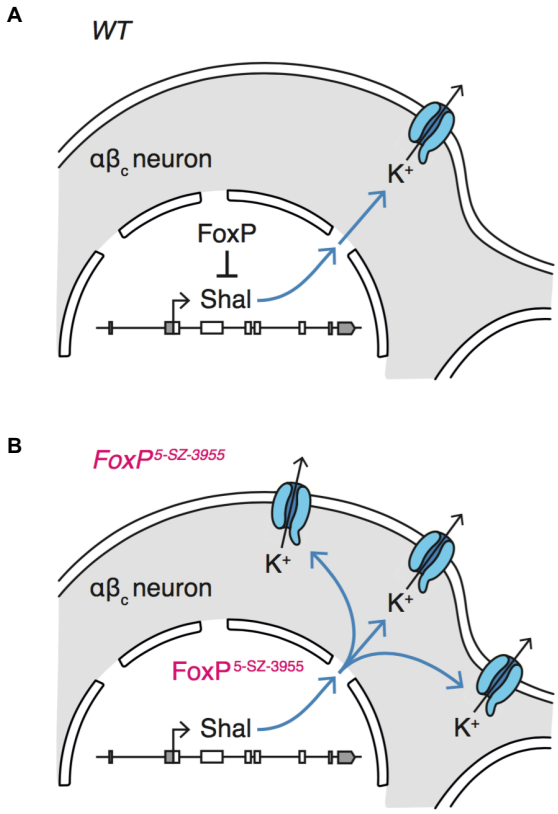


Figure 5. Proposed mechanism. FoxP represses Shal to regulate dendritic signal propagation.
A and B, Simplified schematic depicting the proposed mechanism by which FoxP influences the dentritic integration of sensory evidence. A, In wild-type $\alpha\beta_c$ KCs, FoxP represses the transcription of the Shal gene, maintaining low concentrations of Shal mRNA and protein in the neurons. B, In FoxP5-SZ-3955 mutant $\alpha\beta_c$ KCs, the repressive action of FoxP on Shal transcription is abolished, leading to an increased abundance of functional Shal channels in the plasma membranes of αβc KCs. Figures are from Groschner et al, 2018.

1.3 Aims and objectives

The main aim of this thesis is to use molecular tools to identify the transcriptional regulatory networks of FoxP in *Drosophila melanogaster* by analysing the direct interactions between FoxP and the regulatory elements of its target genes. Deciphering mechanisms of transcriptional control is crucial for understanding how genes specify neural function and, ultimately, behaviour (Ryner et al., 1996, Neville et al., 2014, Blanchard et al., 2010).

One comprehensive and versatile technique used to study the complex interactions between transcription factors and DNA is targeted DNA adenine methyltransferase identification (Targeted DamID). Coupled with deep sequencing, the results from targeted DamID are used to identify the binding sites of transcription factors in specific subsets of cells. Downstream analysis then allows the identification of target genes and the biological processes they are involved in. Furthermore, since transcription factors are defined by their sequence-specific DNA-binding activity, it is important to characterise their binding motif to confirm their DNA interactions.

Specific and dynamic gene expression not only depends on the sequence-specific DNA-binding of one particular transcription factor, but also on the cooperation or competition with other transcription factors. Cooperative binding involves protein-protein interactions between transcription factors bound to adjacent sites of promoter regions. Since the human orthologue of FoxP, FoxP2, has been found to interact with DNA in both dimeric and monomeric form (Stroud et al., 2006), the possibility of combinatorial gene regulation by FoxP in *Drosophila* is tested.

2

Methods

2.1 Fly Strains

All flies were kept on standard cornmeal/agar medium at 25°C and 60% humidity, on a 12-hour light/12-hour dark regime. Experimental flies were heterozygous for all transgenes, including *UAS-FoxP*^{RNAi} (Vienna *Drosophila* Resource Centre [VDRC] ID 15732), *UAS-gcm*^{RNAi} (VDRC ID 110539), *UAS-Spps*^{RNAi} (VDRC ID 45300), *UAS-Dichaete*^{RNAi} (VDRC ID 107194) and *UAS-acj6*^{RNAi} (VDRC ID 105292) (Dietzl et al., 2007) and homozygous for either a wild-type (*WT*) or the mutant *FoxP*$^{5\text{-}SZ\text{-}3955}$ allele (*w*^{1118}*; P{R5}FoxP*$^{5\text{-}SZ\text{-}3955}$, Kyoto Stock Center), as indicated. For TRAP experiments, *NP7175-GAL4* (Murthy et al., 2008) (Tanaka et al., 2008) or *VT030604-GAL4* (VDRC ID 200228) (Kvon et al., 2014) were used to target the expression of UAS-RpL10A::GFP (Ng et al., 2016) to $\alpha\beta_c$ or α'β' KCs, respectively. *NP7175-GAL4* was also used to restrict the expression of UAS-HA and UAS-HA-FoxPC to $\alpha\beta_c$ KCs.

The expression of the DamID constructs (see below) was driven by *elav-GAL4* (Luo et al., 1994). For patch-clamp recordings, *NP7175-GAL4* was used to target the expression of *UAS-CD8::GFP* (Lee and Luo, 1999) to $\alpha\beta_c$ KCs.

2.2 Cell Culture

Neuronal ML-DmBG3-c2 cells and S2 cells were obtained from DGRC and ThermoFisher, respectively. ML-DmBG3-c2 cells were cultured in Schneider's *Drosophila* medium supplemented with 10% fetal bovine serum (FBS) (ThermoFisher Scientific), 10 µg/ml insulin (ThermoFisher Scientific) and 100U/mL penicillin and 100µg/mL streptomycin (ThermoFisher Scientific). S2 cells were cultured in Schneider's *Drosophila* medium (ThermoFisher Scientific) supplemented with 10% FBS and 100U/mL penicillin and 100µg/mL streptomycin. All cells were grown at 25^{o}C.

ML-DmBG3-c2 cells were used for chromatin immunoprecipitation. Approximately 2x10^{6} cells were seeded on 100 mm dishes. Approximately 5x10^{4} of S2 cells were seeded in 96-well glass bottom plates for luciferase reporter assays.

2.3 Targeted DNA adenine methyltransferase identification

2.3.1 Plasmid construction

The pUAST-attB-LT3-NDam vector obtained from Andrea Brand (Southall et al., 2013) was used as template to amplify the mCherry-NDam using a 5' primer that incorporates the NotI restriction enzyme site (GCGGCCGC), TOPO bases (CACC), the Kozak sequence (CAAC) and the first 17 base pairs of the mCheryNDamMyc sequence (ATGGCAACTAGCGGCAT). The 3' primer includes a linker sequence (TCCTCCTGATCCTCCTGA) and the first 13 base pairs of the *FoxP* sequence (ATGGCAACTAGCG).

FoxPC and *FoxPD* CDS were amplified from vectors kindly provided by Korneel Hens using the 3' primers including the last 13 base pairs of the Myc Sequence

(AAGAGGATCTG), a linker sequence (TCCTCCTGATCCTCCTGA) and the first 21 base pairs of the *FoxP* CDS sequence (ATGCATCGGATACATGACGAC). To discriminate between the two possible FoxP isoforms, isoform-specific 3' primers were designed (CTATTTGAGACCCACATACCCAAT and TTATCGATTGTGCTCATTGGCACTA for amplification of *FoxPC* and *FoxPD* respectively).

The resulting PCR products were fused by a second PCR reaction. The fused PCR products were then inserted, by TOPO cloning, into the attB-containing vector, *pMUH*, for site-specific insertion in the *Drosophila* genome, using restriction enzyme digestion (NotI) (NEB) and ligation using T4 DNA ligase (Roche).

Control constructs containing the mCherry-NDamMyc domains but lacking the FoxP sequences were also inserted into the *pMUH* vectors and will be used to estimate baseline DNA methylation levels.

The *pMUH* constructs were inserted into the attP2 integration sites using the phiC31 site-specific integration system (Keravala and Calos, 2008). Embryo injections and screening for transformant flies were performed by BestGene. The transgenic flies were first tested for the presence of the transgenes. Genomic DNA (gDNA) was extracted from 5 flies using the Qiagen DNeasy (Qiagen) Blood and Tissue Kit. Lysis of the brains, binding of DNA to spin columns, removal of any contaminants and elution of DNA were carried out according to manufacturer's protocol.

2.3.2 Targeted DamID

Once the presence of the transgenes was confirmed, the targeted DamID protocol was carried out as described before (Marshall et al., 2016). 200 brains were dissected in DamID buffer (5mM EDTA, 100mM Tris-HCL pH 8.5, 137mM NaCl) for each biological replicate (Dam-FoxPC, Dam-FoxPD and Dam) and flash frozen in liquid nitrogen. Genomic DNA (gDNA) was extracted using the QIAmp DNA Micro Kit (Qiagen). Briefly, digestion of the brain tissues was carried out at 56^{o}C overnight. Precipitation, clean-up and elution of DNA were carried out according to manufacturer's instructions. Due to the inclusion of EDTA in the dissection buffer, RNaseA was only added after DNA elution in 50 µl of water. The quality of the gDNA was checked on a 0.8% agarose gel before proceeding.

For selective amplification of methylated DNA fragments, 43.5 µl of gDNA was digested with 5 µl of NEB CutSmart Buffer and 1.5 µl of DpnI (NEB) overnight at 37^{o}C. DpnI and RNase were then removed from the DpnI-digested DNA using the Qiagen PCR Purification Kit (Qiagen) according to manufacturer's instructions. The DNA was eluted in 32 µl of water.

Double-stranded adaptors (top strand, 5′-CTA ATA CGA CTC ACT ATA GGG CAG CGT GGT CGC GGC CGA GGA-3′; bottom strand, 5′-TCC TGC GCC-3′) were ligated to 15 µl of DpnI digested gDNA for 2 hours at 16^{o}C with 1 µl of T4 DNA ligase (NEB) in a total volume of 20 µl of ligation buffer. The ligase enzyme was inactivated at 65^{o}C for 10 minutes.

To prevent amplification of non-methylated GATCs and to prevent PCR primers from being sequenced, adaptor-ligated DNA was digested with 1 µl of DpnII (NEB) for 2 hours at 37^{o}C in a total volume of 40 µl of DpnII buffer (NEB). Amplification of the methylated fragments was performed with 40 µl of DpnII-digested DNA, 2 µl of Advantage 2 cDNA polymerase enzyme (Clontech), 32 nmoles of dNTPs (Roche) and 100 pmoles of a primer complementary to the double-stranded adaptor (5′-GGT CGC GGC CGA GGA tc-3) in 160 µl total volume of PCR Advantage reaction buffer. The PCR cycling conditions were as follows: activation of the polymerase and nick translation for 10 min at 68°C, followed by one cycle of 1 min at 94°C, 5 min at 65°C, and 15 min at 68°C; three cycles of 1 min at 94°C, 1 min at 65°C, and 10 min at 68°C; and 19 cycles of 1 min at 94°C, 1 min at 65°C, and 2 min at 68°C.

The PCR products were run on a 0.8% agarose gel to visualise the quality of the PCR amplicons. The PCR products were then purified according to manufacturer's instructions using the Qiagen PCR Purification Kit. 2000 ng of amplified DNA from experimental and Dam-only controls were then sent to the Wellcome Centre for Human Genetics (Oxford) for deep sequencing. The libraries were sequenced with 75-bp paired-ends reads on the Illumina HiSeq4000 platform, with a theoretical coverage > 30x.

2.3.3 Sequencing data quality control

Each fastq dataset was first tested using FastQC (version 0.11.5) (http://www.bioinformatics.babraham.ac.uk/ projects/fastqc/) to analyse the read quality (Appendix 2).

2.3.4 Sequencing data analysis

The two fastq files for each sample were concatenated. The ligation adaptor and PCR primer sequences were trimmed from the sequencing reads using cutadapt. The DamID-sequencing analysis was done using the 'damidseq pipeline' publicly available (https://github.com/owenjm/damidseq_pipeline). Briefly, the trimmed reads were aligned to the Flybase dmel_r6.16 assembly of the *Drosophila* genome (ftp://ftp.flybase.net/releases/current/) using Bowtie 2 (version 2.1) (Langmead and Salzberg, 2012) using default settings. Unmapped reads and reads with a mapping quality score <10 were filtered and excluded from downstream analysis. The SAM file was converted to BAM format using SAMtools (version 0.1.9) (Li et al., 2009).

Correlation coefficients between the biological replicates were calculated using the multiBamsummary function in the NGS analysis suite deepTools (Ramirez et al., 2016), with the custom options multiBamSmmary bins –corMethod Pearson. Since the replicates displayed Pearson correlation coefficients > 0.8, the replicates were merged before downstream analysis. The FoxPC and FoxPD binding profiles at the transcription start sites were also generated using the computeMatrix and plotProfile functions in deepTools. ComputeMatrix summarised the read density at ± 1000bp around transcription start sites using the command computeMatrix reference-point –b 1000 –a 1000.

Using the intersectBed function in BEDTools, the number of significant peaks (FDR<0.01) associated with different genomic functions (promoters, 3' UTRs, exons, introns, 3' UTRs and intergenic regions) was determined. The different region coordinates were downloaded from UCSC browser (Karolchik et al., 2003). Promoter regions were defined as regions 1000 bp from the transcription start site.

Since it is expected that all the reads will theoretically begin with GATC, the reads are mapped to a library consisting of pre-binned GATC-flanking DNA sequences (Dmel_BDGP6.GATC.gff). Mapped reads were then assigned to individual GATCs in the genome. All the reads were extended to the next GATC site or to the average fragment length (300bp). The number of extended reads overlapping each GATC fragment was calculated using the BEDTools coverage utility in continuous, non-overlapping 75bp windows (bins) (Quinlan and Hall, 2010). The pipeline from Marshall et al. determined the normalisation factor between Dam-FoxPC or Dam-FoxPD and Dam-only samples based on Gaussian kernel density estimate and calculated algorithmically the optimal number of pseudo-counts added to the samples to reduce background noise. The pipeline then yielded $\log_2$ ratio files by normalising the read counts between Dam-FoxPC or Dam-FoxPD and Dam-only.

The $\log_2$-transformed files were then used for peak calling using the find_peaks script (https://github.com/owenjm/find_peaks/blob/master/find_peaks). All 75-bp windows with binding intensity ($\log_2$ read count ratios) between quantiles 0.95 to 0.99, with a stepping quantile of 0.01, were extracted and merged. To calculate the FDR of each of these candidate peaks, the dataset was randomly shuffled for 100 times, and the frequency of consecutive windows with intensity score above each threshold was calculated. The relationship between the number of consecutive fragments and the frequency of observation of such windows in a random dataset is logarithmic and can therefore be modeled for any window by linear regression. The FDR was calculated as the ratio of observed over expected peaks for a number of consecutive fragments above the intensity threshold. The binding profiles were visualised using Integrative Genomics Viewer (IGV) and Integrated Genome Browser browser (IGB).

Significant binding intervals or peaks (FDR<0.01) overlapping with genes annotated in Flybase database were assigned to the overlapping or closest gene (within 2 kb) using the peaks2genes script (https://github.com/owenjm/find_peaks/blob/master/peaks2genes).

2.3.5 Overlap with known CRMs, TFBSs and inferred motifs

To verify whether FoxPC and FoxPD interact with known *cis*-regulatory modules (CRMs), a list of 16625 known *Drosophila* transcription CRMs was downloaded from the REDfly database (Gallo et al., 2011). Significant peaks (FDR < 0.01 and $\log_2$ ratio > 1.5) were mapped to those features using the intersectBed tool function in the bedTools package.

To identify known putative motifs which are enriched in the peak regions *in silico*, i-cisTarget (http://med.kuleuven.be/lcb/i-cisTarget) was used. Significant peak regions for FoxPC and FoxPD were used as input and are mapped to known CRMs and motifs from several databases (136K regions covering the entire non-coding genome). The default parameters were used and motifs with an enrichment score >3 were considered significant.

2.3.6 Motif Analysis

Sequences underlying the significant peaks were extracted from FoxPC and FoxPD datasets and used as *fasta* input for *de novo* motif analysis using Multiple Expectation Maximization for Motif Elicitation (MEME) (Bailey et al., 2009). The number of potential motifs to find was set to 15, the optimum motif width ranging between 5 and 15 residues and the number of occurrences to zero or one per sequence. Only motifs with an e-value <0.05 were considered significant. The resulting motifs were presented as position weight

matrices. The discovered motifs were then scanned against known motifs using TOMTOM. Default values were used for the TOMTOM search. The query and the target motifs were compared using Pearson Correlation Coefficients and the significance threshold was set to <10.

2.3.7 Gene Ontology Enrichment Analysis

Genes were functionally classified with Gene Ontology (GO) terms using FlyMine (http://www.flymine.org/flymine/bag.do). Over-represented GO terms were determined using the binomial test and p-values were corrected for multiple testing using the Bonferroni method. A corrected p-value <0.05 was considered as significant. The most significant GO terms for each dataset were visualised in a heatmap using the heatmap.2 function in the gplot package of the statistics package R.

2.3.8 Pathway Analysis

Cytoscape 3.5.1 was used to generate gene networks using the GeneMania (D. melanogaster Version 2012-08-02) database application. The genes associated with the top GO term (nervous system development) were used as input to create interaction networks using the input networks: genetic interactions, physical interaction and predicted. Weighting was set at automatic. The network was analysed and the largest connected network was visualized using the following topological parameters: edge colour/width represents edge 'betweenness' (a score which measures the number of shortest path flows that go through an edge), node size represents 'betweenness centrality' (reflects the amount of control that the node exerts over the interactions of other nodes in the network). Edge 'betweenness' and 'betweenness centrality' are defined by the equations below.

$$Edge\ betweenness\ for\ edge\ e = \sum_{s,t \neq v} \frac{\sigma_{st}(e)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(e)$ is the number of those paths that pass through edge e

$$Betweenness\ centrality\ of\ node\ v = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through edge v

2.3.9 Motif occurrence by MAST

To identify the locations of sequences in the promoter regions of target genes, the MAST program (as part of the MEME suite) was used. The sequences of target genes were scanned for the positions of the input sequences or motifs using default settings. Only matches with an E-value smaller than 10 were considered significant.

2.4 Chromatin immunoprecipitation (ChIP)

2.4.1 Plasmid Construction

FoxPC was amplified from *FoxPC* cDNA (gift of Korneel Hens) using the following primers:

AGCCGGAATTCATGTACCCATACGATGTTCCAGATTACGCTCATCGGATACAT GACGACGAG and CGGGGTACCTTATCGATTGTGCTCATTGGCACTA. The forward primer contains the EcoRI restriction site and the HA tag, while the reverse primer contains the KpnI site. The PCR product was purified and ligated into a pUAST-attB vector (Addgene). The HA-only vector was made by annealing the two oligos AGGGAGAATTCATGTACCCATACGATGTTCCAGATTACGCT and

GGTACCTTAAGCGTAATCTGGAACATCGTATGGGTA. The forward oligo bears the EcoRI site and the reverse strand bears the KpnI site. The annealed oligo was then ligated to the pUAST-attB vector.

2.4.2 ChIP in ML-DmBG3-c2 neuronal cells

To obtain a snapshot of the association of FoxPC to DNA, chromatin immunoprecipitation (ChIP) was carried out in cultured neuronal cells using the Tag-ChIP-IT kit (Active Motif). In brief, ML-BG3-C2 cells were grown in 100mm dishes (ThermoFisher Scientific) and transfected with 600 ng of either pUAST-HA-FoxPC or pUAST-HA constructs and pMT-GAL4 using the Effectene transfection reagent (Qiagen). Two days after transfection and induction of GAL4 activity with copper sulfate (ThermoFisher Scientific), cells were fixed by incubation with 1% formaldehyde (SigmaAldrich) for 10 minutes at room temperature with shaking. The fixation reaction was stopped by adding 1X glycine solution to the cells for 5 minutes. Cells were detached from the plates using a rubber policeman and pelleted by centrifugation for 3 minutes at 1,250 g at 4^{o}C. The pellet was washed twice with ice-cold PBS.

The cell pellets were then resuspended in Chromatin Prep Buffer supplemented with proteinase inhibitor cocktail and phenylmethylsulfonyl fluoride (PMSF). After incubation on ice, the cells were lysed using Dounce homogenizers. The nuclear pellets were then resuspended in ChIP buffer and chromatin was sheared by sonication with a microtip sonicator (Cole Parmer) at 25% amplitude for 12 pulses of 30 seconds each, with a 30-second rest on ice between each pulse. The sizes of the resulting sheared chromatin were verified to be between 200-1000 bp.

25 µl of chromatin was stored to be used as the input DNA control. 25 µg of sheared chromatin was incubated with 100 µl of ChIP buffer and with either 8 µg of ChIP grade HA antibody (Abcam) or rabbit IgG antibody as a negative isotype control (Abcam) overnight at 4^{o}C. The ChIP reactions were then incubated with 30 µl protein G agarose beads for 3 hours at 4^{o}C to capture the immunoprecipitated chromatin. The beads-Ig complexes were washed 3 times and then purified from the chromatin-antibody mixture using ChIP Filtration Columns.

Immunoprecipitated DNA was eluted in 100 µl of pre-warmed elution buffer. Crosslinks were reversed at 65^{o}C overnight in both the input DNA and immunoprecipitated DNA samples using 2 µl of 5 M NaCl and protein was removed by adding 2 µl of 10 µg/µl proteinase K. Quantification of the DNA fragments in the input samples and the immunoprecipitated samples (both for HA-tagged FoxPC and HA-only control) was carried out by quantitative PCR (qPCR) using SYBR Green dye and the Roche LightCycler 480® (Roche). Amplification reactions using the SYBR Green I assays contained 1xSYBR Green I PCR Master MIX (Roche), 2.5 µM of each primer and 50 ng of template cDNA. The DNA was purified using the DNA purification column and examined by real-time PCR. Primers specific for *osa, ph-p, Mam, ftz-f1, Mnt, crol, Trpγ, Shal* and *Shaker* promoter regions (1000bp upstream and downstream of the transcription start sites) were used (Appendix 6 for primer sequences). PCR primers were to yield amplicon sizes of roughly 100 bp. The reaction conditions were as follows; denaturation at 95^{o}C for 10 minutes, followed by 40 amplification cycles of 95^{o}C for 15 seconds and 58^{o}C for 20 seconds. An additional step was added whereby the sample was maintained at 72°C for 5 seconds to allow for the fluorescence signal to be read.

The C_t values were recorded after analysis by the Roche LightCycler® 480 software. C_t values of triplicates were checked for minimal variablility and were used to calculate the

fold difference between the experimental sample and normalised input. This was also compared to the fold difference between HA-only control sample and the input. In this case, the dilution factor was 4.

$$\Delta C_t = C_t(ChIP\ DNA) - C_t(input) - log_2(dilution\ factor)$$

$$\Delta\Delta C_t = \Delta C_t(HA - FoxPC) - \Delta C_t(HA - only)$$

$$Fold\ change = 2^{-\Delta\Delta C_t}$$

2.4.3 ChIP in $\alpha\beta_c$ neurons in Drosophila brains

Flies were injected with either pUAST- HA-FoxPC or pUAST-HA vectors by BestGene. The expressions of those transgenes were restricted to the $\alpha\beta_c$ neurons using the NP7175-GAL4 line. Brains from 200 flies were dissected in cold PBS buffer supplemented with 1x proteinase inhibitor cocktail (Roche) and phenylmethylsulfonyl fluoride (Roche). Brains were washed twice and resuspended in 200 µl of the same buffer. Crosslinking was carried out for 15minutes at 37^{o}C with 1% formaldehyde. 0.125M glycine was added for 5 minutes at room temperature to terminate the fixation. The brains were then washed twice with PBS. Cells were lysed by adding 200 µl of lysis buffer (50 mM Tris-HCl, pH 7.6, 1 mM CaCl2, 0.2% Triton X-100 or NP-40, 5 mM butyrate, and 1x proteinase inhibitor cocktail). The brains were homogenised until they were completely dissociated and incubated at room temperature for 10 minutes.

To fragment chromatin, the brain lysates were sheared with a microtip sonicator (Cole Parmer) at power 20 for 5 pulses of 30 seconds each. After sonication, the cell lysate was centrifuged at 4^{o}C to remove cellular debris. The sheared chromatin DNA was diluted by adding 1ml of RIPA buffer (ThermoFisher Scientific) supplemented by protease inhibitor. For quality checks and DNA quantification, 40 µl of diluted chromatin was used as input,

reverse crosslinked using 2 µl of 5 M NaCl overnight at 65°C. The input was then incubated with 2 µl RNAse A (10mg/mL) and 5 µl proteinase K (20mg/mL) at 65°C for 4 hours. The input chromatin DNA was purified by phenol-chloroform extraction (Sigma-Aldrich), quantified and used as control to compare against antibody conjugated samples. DNA fragment sizes were checked after cross-link reversal on an agarose gel.

8 µl of either ChIP grade HA antibody or rabbit IgG antibody was conjugated to 20 µl of magnetic protein A beads (ThermoFisher Scientific) at 4°C for 4 hours with rotation. 800 µL of diluted ChIP sample was immunoprecipitated with the conjugated beads overnight at 4°C with rotation. The immunoprecipitated chromatin was then washed twice with a low-salt wash buffer (1.89 mL 'RIPA buffer' + 315 µL 7x protease inhibitors + 20 µL PMSF), and a high-salt wash buffer (1.89 mL RIPA buffer + 220 µL 3 M NaCl). The beads were then washed with a buffer containing a different salt (0.25 M LiCl, 0.5% NP40, 0.5% NaDOC) to improve the removal of non-specific chromatin interactions. The chromatin was then released from the beads using 1 mL elution buffer (1x TE + 0.2% Triton X-100) and 1 mL of TE buffer. The crosslinked DNA was released as described above. DNA was then purified by phenol-choloroform extraction.

qPCR was then carried out on input DNA, immunoprecpitated DNA from experimental (HA-FoxPC) and control (HA-only) flies as above and the fold enrichment was also calculated as described before. However, in this case, the dilution factor was 20.

2.5 Translating ribosome affinity purification

UAS-RpL10A::GFP flies were crossed with wild-type, $FoxP^{5\text{-}SZ\text{-}3955}$ and $FoxP^{RNAi}$

lines for the labeling of polysomes in $\alpha\beta_c$ KCs using the NP7175-GAL4 line. For the analysis of gene expression levels in whole fly heads, ~400 flies per sample were collected 7-8 days post-elocsion in Eppendorf® tubes, plunge-frozen in liquid nitrogen, and vortexed, to dislocate the heads from the rest of the body. The sample was then applied onto pre-cooled sieves with mesh sizes of 850, 710 and 355µm. Fly heads were collected and homogenised using a mechanical disruptor in 500 µl ice-cold extraction buffer (20 mM HEPES, pH 7.5, 150 mM KCl, 5 mM MgCl2, 1% Triton X-100, 0.5 mM DTT, 100 µg/mL cycloheximide, 100 U/mL RNase OUT (Life Technologies) and 1X Complete Protease Inhibitor (Roche). Lysates were centrifuged at 14,000 *g* for 15 min at 4℃, and the supernatants were applied to 25 µl Protein G Mag Sepharose beads (GE Healthcare) coated with monoclonal mouse anti-GFP antibody (Htz-GFP-19C8, Memorial Sloan-Kettering Monoclonal Antibody Facility). After incubation for 1 h at 4℃, unbound material was collected, and the beads were washed 5 times with 500 µl each of buffer (pH 7.4) containing 20 mM HEPES, 150 mM KCl, 5 mM $MgCl_2$, 0.05% (v/v) Triton X-100, and 40 U/ml RNaseOUT.

2.5.1 RNA extraction, reverse transcription and real-time PCR

RNA was isolated from immunoprecipitated ribosomes and unbound material using the PicoPURE RNAIsolation Kit (Life Technologies); 10-15 ng of RNA were then reverse-transcribed using SuperScript III Reverse Transcriptase and oligo(dT) primers (Invitrogen). cDNA was pre-amplified by multiplex PCR using primers specific for three housekeeping genes (*Gpdh, Tbp* and *Ef1α100E*); *GFP; Fas2* or *trio* as markers for αβ or α'β' KCs, respectively (Grenningloh et al., 1991, Awasaki et al., 2000); the FoxP isoforms and genes of interest (*osa, ph-p, Mam, ftz-f1,Mnt, crol, Trpγ, Shal* and *Shaker*) (Appendix 6). Intron-spanning primers were designed using the Primer3 software (Untergasser et al., 2012). Transcript levels were determined by quantitative real-time PCR on a LightCycler

480 system (Roche) using SYBR Green I PCR Master Mix (Roche) in 10 µl reactions containing 100 nM of each gene-specific primer and 50 ng of pre-amplified cDNA. All samples were run in technical triplicates; non-reverse-transcribed mRNA and water served as negative controls. The reaction conditions were as follows: denaturation at 95^{o}C for 10 minutes, followed by 40 amplification cycles of 95^{o}C for 10 seconds and 62^{o}C for 1 minute. An additional step was added whereby the sample was maintained at 75°C for 5 seconds to allow the fluorescence signal to be read. Melting curves were analysed after amplification, and amplicons were visualised by agarose gel electrophoresis to confirm primer specificity. Relative transcript levels were estimated with the help of the $2\text{-}\Delta\Delta C_t$ method (Livak and Schmittgen, 2001) using the geometric mean of the C_t values of the three housekeeping genes for normalisation.

2.6 Dual luciferase reporter assays

2.6.1 Plasmid construction

The pDONRp4p1R entry vector, the actin promoter and firefly luciferase gene-containing vector (pLucCP4), the renilla luciferase vector (pRenilla) and the pActin vector were gifts of Korneel Hens. The primers used to amplify the *osa, ph-p, Mam, ftz-f1, Mnt, crol, Trpγ, Shal* and *Shaker promoter* regions (± 1000 bp from transcription start sites) from genomic DNA (gDNA) are listed in Appendix 6. The primers include the attB1 and attB4 recombination sites required for BP recombination with the pDONRp4p1r vector. The promoter regions were amplified using the Phusion high fidelity polymerase (NEB). The PCR products were then gel purified using the QIAquick Gel Extraction kit (Qiagen). The purified products were inserted into the pDONRp4p1r vector by a BP reaction using the Gateway BP Clonase II (ThermoFisher Scientific). The promoter regions were then transferred from the resulting entry clone to the destination vector, pLucCP4, using the

Gateway LR Clonase II (ThermoFisher Scientific), such that the promoter regions are found upstream of the firefly luciferase gene (pLuc-Promoter). Osa, Trpγ and Shal constructs where the FoxPC motif is mutated were synthesised by GeneArt (ThermoFisher) and cloned into the pLucCP4 vector.

The coding sequences of the transcription factors FoxPC, Spps, Dichaete, gcm and acj6 were amplified using the primers listed in Appendix 6 and cloned downstream of the constitutive actin promoter in the pActin vector (pActin-TF). To confirm FoxPC binding to its targets, the residues crucial in the FoxPC DNA binding domain (N373, T377, R376, S380 and W397) were mutated to alanines by GeneArt (ThermoFisher).

2.6.2 Luciferase assay

S2 cells were grown in glass-bottom 96-well plates (ThermoFisher Scientific) to a confluency of $5x10^4$. 60 ng of pRenilla, 60ng of pLuc-Promoter and 60ng of pActin-TF were transfected into cells using FuGene (Promega). To measure basal activity level of the promoter regions, the empty pActin vector was used instead of pActin-TF. Reporter assays were carried out using the Dual-Glo Luciferase Assay System (Promega) 48 hours after transfection. Cells were lysed, and after the addition of firefly luciferase substrate, relative light units (RLU) due to firefly luciferase were measured in a microplate reader (LUMIstar Omega), followed by determination of renilla luciferase activity by luminescence. Firefly luciferase activity was normalised to renilla luciferase activity to eliminate any differences in the state of the cells, number of cells, and transfection or lysis efficiencies. The relative luciferase activity was then compared to the basal activity of the cells transfected with the empty vectors. All experiments were performed in triplicates

from independent cell cultures. The data were analyzed using GraphPad Prism software by analysis of variance (ANOVA) followed by unpaired t test (2 experimental conditions).

2.7 Electrophoretic Mobility Shift Assays (EMSA)

S2 cells were transfected with the same HA-tagged FoxPC construct used in ChIP assays and the HA-tagged version of the mutated FoxPC. Proteins in the cytoplasm and nucleus were extracted using the Nuclear Extract Kit (Active Motif) following manufacturer's instructions. The HA-tagged expressing constructs were purified using anti-HA magnetic beads (ThermoFisher Scientific). 0.15 mg of protein extract was incubated with 50 µl of beads overnight at 4^{o}C. The HA-tagged protein was eluted from the beads using 0.1M glycine at pH 2.0.

5'-biotinylated labeled *osa*, *Trpγ* and *Shal* probes were made by amplification of the *osa*, *Trpγ* and *Shal* regulatory regions containing the putative FoxPC motif using the biotin-labeled primers. The PCR products were purified using the Qiagen Purification Kit (Qiagen). 8 µg of biotin-labeled DNA was immobilised onto 40 µl of Dynabeads^{TM} MyOne^{TM} Streptavidin T1 beads (ThermoFisher). The single stranded biotin-labeled probe was released from the beads by incubating at 65^{o}C in 10mM EDTA (Sigma-Aldrich) with 95% formamide (Sigma-Aldrich).

EMSA was conducted using the Lightshift Chemiluminescent EMSA Kit (ThermoFisher). Binding between the FoxPC protein and the probe was carried out in a 20 µl reaction containing 2 µl of 10x binding buffer,1 µl of glycerol, 1 µl of poly(dI-dC), 1 µl of 100mM $MgCl_2$, 1 µl of 1M KCl, 1 µl of 200mM EDTA and 1 µl of 1% NP-40. 100 fmol of biotin-

labeled DNA was incubated with 3 µg of protein extract for 20 minutes at room temperature. The samples were run onto a 6% native polyacrylamide gel for 1 hour at 100V. Complexes were then immobilised onto a positively-charged nylon membrane (GE Healthcare) at 10V for 30 minutes. Transferred DNA was crosslinked to the membrane at 120mJ/cm^2 for 1 minute. The membrane was then incubated with blocking buffer for 20 minutes at room temperature, followed by incubation with the streptavidin-horseradish peroxidase conjugate. After several washes at room temperature, the DNA was visualised using a CCD camera (GeneGnome XRQ, Syngene).

2.8 In vitro transcription/ translation and co-immunoprecipitation

2.8.1 Construct preparation

Plasmids encoding pDONR-gcm, Spps, Dichaete and acj6 were gifts of Korneel Hens. The coding sequences were transferred to a GST-containing pF3A WG (BYDV) Flexi® Vector (Promega) by LR reactions. pDONR-FoxPC (also a gift of Korneel Hens) was cloned into a FLAG-containing vector (Promega).

2.8.2 Cell-free protein expression

To achieve the synthesis of FLAG-tagged FoxPC, GST-tagged gcm, GST-tagged Spps, GST-tagged Dichaete and GST-tagged acj6 *in vitro*, the TnT® SP6 High-Yield Wheat Germ Protein Expression System (Promega) was used. 2 µg of FLAG-tagged FoxPC and 2 µg of one of GST-tagged gcm, Spps, Dichaete were incubated in 60 µl of TnT MasterMix for 2 hours at 25^{o}C. For control reactions, 20 µl of distilled water was used instead of GST-tagged plasmid. 20 µl of the ITT mix was saved to test for expression of the proteins. The rest of the ITT mix was then incubated with 100 µl of GST-coated magnetic beads

(ThermoFischer Scientific) (washed with and resuspended in 500 µl of Equilibration/Wash buffer (50 mM Tris-HCl, 150 mM NaCl) using an end over end rotator at 4^{o}C overnight. 20 µl of unbound fraction (supernatant) was collected for analysis and reduces in LDS and reducing agent. The beads were washed three times with 500 µl of Equilibration/Wash buffer. The proteins were then eluted from the beads and reduced using 10 µl of LDS buffer (ThermoFischer Scientific), 4 µl of NuPAGE Sample Reducing agent (ThermoFischer Scientific) and 26 µl of milliQ water at 70 ^{o}C for 10 minutes.

2.8.3 Western Blotting

Protein in the ITT Mix, and unbound and immunoprecipitated fractions, were then analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) under reducing conditions and blotted onto polyvinylidene fluoride (PVDF) membranes using the iBlot 2 Dry blotting System (ThermoFisher Scientific). The membranes were blocked for 1 hour at room temperature in Tris-buffered saline (pH 7.5) containing 20 mM Tris, 150 mM NaCl, 0.1% (v/v) Tween 20 (TBST) and 5% (w/v) skim milk powder. FoxPC was detected using a rabbit anti-FLAG antibody (1:1000) (F3185, Sigma Aldrich) and peroxidase-conjugated goat anti-rabbit IgG (32260, Invitrogen, 1:5000). After primary incubations at 4^{o}C overnight, the membranes were washed 4 times with TBST, each wash lasting 15 minutes. Secondary incubation was carried out for 2 hours at room temperature. Peroxidase activity was visualised by chemiluminescence using ECL substrate (Thermo Fisher Scientific) and a CCD camera (GeneGnome XRQ, Syngene).

2.9 Electrophysiology

For whole-cell patch-clamp recordings in vivo, male flies aged 6-24 h post eclosion were fixed to a custom mount with soft thermoplastic wax (Agar Scientific). Cuticle, adipose tissue, perineural sheath and trachea were surgically removed in a window large enough to expose the dorsal brain and antennal nerves. The brain was continuously superfused with extracellular solution (pH 7.3) containing 5 mM TES, 103 mM NaCl, 3 mM KCl, 26 mM $NaHCO_3$, 1 mM NaH_2PO_4, 1.5 mM $CaCl_2$, 4 mM $MgCl_2$, 8 mM trehalose, 10 mM glucose, and 7 mM sucrose (275 mOsM, equilibrated with 5% CO_2 and 95% O_2). Patch pipettes (15-17 MΩ) were fabricated from borosilicate glass capillaries with outer and inner diameters of 1.5 and 0.86 mm (Sutter Instruments), using a PC-10 micropipette puller (Narishige), and filled with solution (pH 7.3) containing 10 mM HEPES, 140 mM potassium aspartate, 1 mM KCl, 4 mM MgATP, 0.5 mM Na_3GTP, 1 mM EGTA, and 10 mM biocytin (265 mOsM). Pipettes were visually targeted to green-fluorescent KC somata using a combination of epifluorescence and differential interference contrast on an Axioscop 2 FS mot microscope (Zeiss) equipped with a 60×, 1.0 NA water immersion objective (LUMPLFLN60XW, Olympus) and an X-Cite 120PC Q light source (Excelitas Technologies). $\alpha\beta_c$ KCs labelled by NP6024-GAL4 were considered a homogeneous population and sampled randomly, irrespective of their discernible origin from one of the four neuroblast clones that give rise to the mushroom bodies. Signals were acquired at room temperature (21–23^{o}C) with a MultiClamp 700B amplifier (Molecular Devices), lowpass-filtered at 10 kHz, and sampled at 50 kHz using a Digidata 1440A digitizer controlled through pCLAMP 10 (Molecular Devices). Data were corrected for liquid junction potential and analysed with custom procedures using the NeuroMatic package (http://neuromatic.thinkrandom.com) in Igor Pro (WaveMetrics).

2.10 Two- odour discrimination tasks

7-8 day old male flies were trained to avoid 4-methylcyclohexanol (MCH). After training, the flies had to discriminate the reinforced MCH concentration from a different concentration of the same odour. The flies were analyzed individually in transparent plexiglass chambers (50 mm long, 5 mm wide, 1.3 mm high) (DasGupta et al., 2014). Two independently controlled odour streams entered the chambers through ports at the ends of the long chamber axis, converged at the centre, and left through lateral vents. Printed circuit boards, connected via solid-state relays (Fairchild HSR312L) to a 70-V power supply, served as floors and ceilings. Filtered, flow-controlled (CMOSens, Sensirion), and humidified carrier air was mixed, in a 90 vs 100 or 100 vs 90 ratio, with flow-controlled odour streams drawn through vials containing (MCH), yielding a final flow rate of 0.25 l/min per half-chamber. Twenty chambers were operated simultaneously in an incubator (Sanyo MIR-154) at 25^{o}C. The chambers were backlit by 940-nm LEDs (TSAL6100, Vishay) and imaged using a Stingray F080B CCD camera (Allied Vision Technologies) with an 18-mm lens (Edmund Scientific). A virtual instrument written in LabVIEW (National Instruments) controlled the delivery of odours and electric shock and recorded the positions of the 20 flies as functions of time (Claridge-Chang et al., 2009).

Flies were trained to avoid 20 ppm MCH by pairing twelve consecutive 20-s odour presentations, which were interspersed with 20-s presentations of clean air, with 1 s of electric shock (DasGupta et al., 2014). During a 2-min testing period immediately after training, one half of each chamber was filled with MCH at the reinforced concentration while the other half was perfused with a lower concentration (18 ppm). Experimental data were processed in MATLAB (The MathWorks) (DasGupta et al., 2014). Position measurements were smoothed by computing 0.5-s moving averages and filtered to exclude

animals making <2 end-to-centre runs per testing period. Each chamber contained a central, 7-mm wide decision zone. Correct and incorrect choices during the test period were combined into individual accuracy scores, which were used to calculate population averages. The time between entry into and exit from the decision zone was recorded as the reaction time (DasGupta et al., 2014). A fly's trajectory within the decision zone, viewed along the long chamber axis, typically consisted of an entry phase at high velocity, a pause at the odour interface, and an accelerated exit phase. Reaction times were normalized to each fly's instantaneous walking speed, estimated as the average speed from the point of entry into the decision zone to the first reduction to 10% of the entry velocity (DasGupta et al., 2014). Reaction time distributions were compared by Kolmogorov–Smirnov test, using the Bonferroni correction to adjust the level of statistical significance.

3

Results

Identifying putative targets of FoxP in the brain

Transcription factors are crucial for decoding genomic information, leading to the expression of specific, coordinated and unique combinations of proteins in every cell type. Binding of transcription factors to specific DNA regulatory sequences modulates the rate and level of gene transcription, and therefore protein synthesis. An essential aspect of understanding a transcription factor's biological role is important to map its binding sites and therefore its possible targets. Recent developments have revolutionalised the identification of transcription factor binding *in vivo*. Genome-wide techniques for mapping protein-DNA interactions, such as nuclease sensitivity and electrophoretic mobility assays, chromatin immunoprecipitation (ChIP) and DNA adenine methyltransferase identification (DamID) now allow the identification of transcription factor binding at high resolution, without any prior knowledge about their targets.

Chromatin immunoprecipitation

The most widely used and the *de facto* standard to analyse the binding of proteins to DNA in a biological context is chromatin immunoprecipitation (ChIP) (Bernstein et al., 2004, Buck and Lieb, 2004, Ren and Dynlacht, 2004). ChIP allows the transcription factor of interest to be selectively immunoprecipitated from a chromatin preparation to identify the DNA sequences bound to it. This is possible by first reversibly cross-linking DNA targets

with protein using formaldehyde to create strong covalent links between them. The chromatin is then sonicated into small fragments, and transcription factor-DNA complexes are immunoprecipitated using antibodies specific to the protein of interest. After reversal of the chemical cross-links, the ChIP-enriched DNA is purified and can be identified by labelling and hybridisation to tilling arrays (ChIP-on-chip), direct high-throughput sequencing (ChIP-seq) or quantified using quantitative real-time PCR (ChIP-qPCR) (Figure 6A).

Even though this technique has been used for over a decade, it has several drawbacks. ChIP can only yield insights into transcription factor-DNA interactions at given times when proteins are fixed to DNA. Furthermore, ChIP can only identify DNA molecules which are bound directly with the proteins of interest as formaldehyde cross-links proteins and DNA within ~2 Å of each other. The success of ChIP depends on the availability of very specific antibodies for immunoprecipitation. In the case of FoxP, no ChIP-grade validated antibodies are available. Hence, immunoprecipitation of FoxP-DNA complexes is not possible in flies.

DNA adenine methyltransferase identification

DNA adenine methyltransferase identification (DamID) is an alternative method to ChIP, which overcomes the above mentioned challenges encountered with ChIP (Vogel et al., 2007). DamID relies on marking DNA binding sites of transcription factors by taking advantage of the DNA adenine-N6 methyltransferase, Dam, from *Escherichia coli*. Dam will freely diffuse in the nucleus and will be recruited to all accessible GATC sites, resulting in random methylation. Fusing the protein of interest to Dam increases steric hindrance and decreases the random, non-specific, methylation rate. The fusion protein

will be mostly recruited to native binding sites of the protein of interest, resulting in methylation of adenine residues in neighbouring GATC sequences ($GA^{me}TC$). An increased methylation ratio at binding sites will be observed for Dam-fusion proteins compared to Dam-only proteins.

DamID is possible because methylated adenine is not present in eukaryotes and the methylation obtained with Dam can therefore be exploited. Genomic DNA is extracted after expression of the fusion protein and methylated DNA is enriched by digesting with methylation sensitive restriction enzymes (DpnI and DpnII), ligation of PCR adapters and amplification. *In vivo* protein-DNA interaction maps can then be identified by microarray technologies (DamID-chip) or next-generation sequencing (DamID-seq) (Figure 6B). DamID-seq is now the preferred method for analysing DamID libraries because it has no hybridization bias and allows one to assess methylation at every GATC site in the genome.

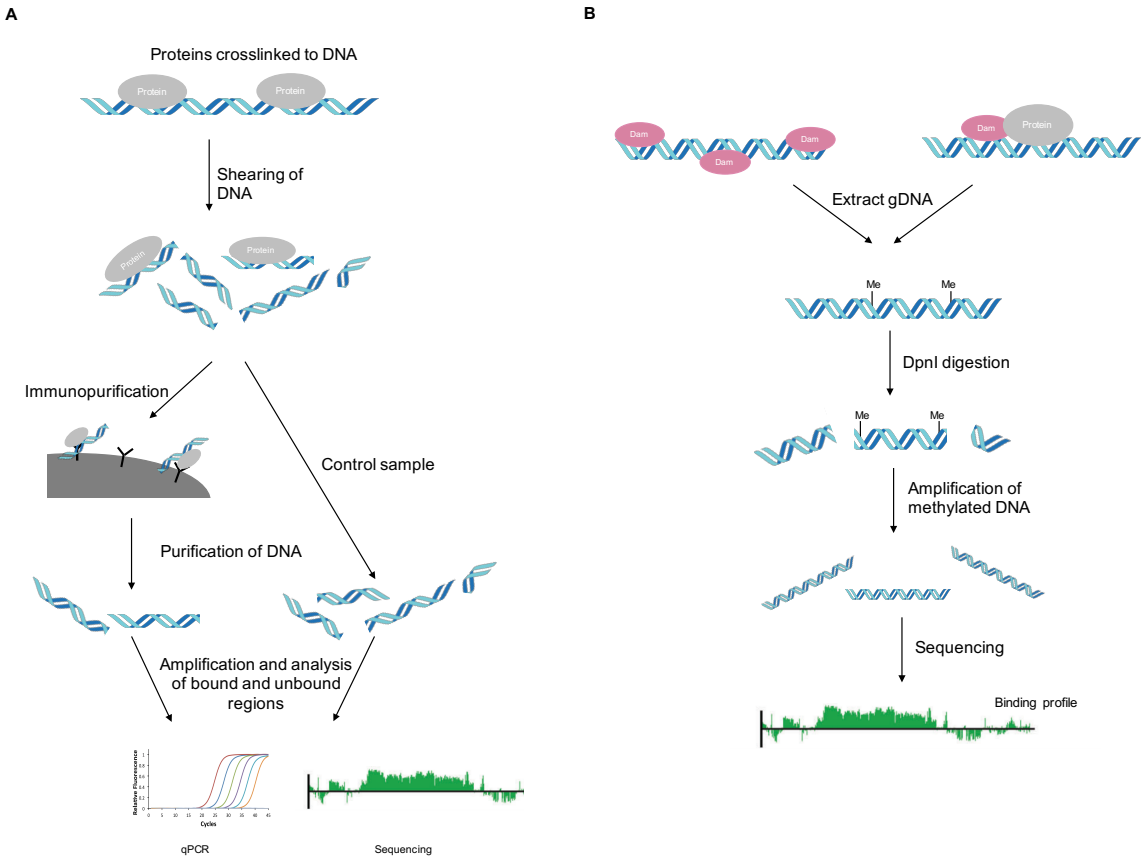


Figure 6. Schematic of ChIP and DamID pipeline.

*ChIP and DamID can both be used for determining protein-DNA binding profiles. In ChIP (**A**), proteins of interest are covalently crosslinked to DNA. Genomic DNA (gDNA) is isolated and sheared by sonication into fragments. A highly specific antibody, which recognises the protein of interest is added and immunoprecipitation is then carried out to isolate protein-DNA complexes. The DNA is then released from the protein and purified. The protein binding sites can then be uncovered by qPCR and sequencing. DamID (**B**) uses the expression of Dam only or a fusion protein of Dam with the protein of interest. gDNA is extracted and digested by DpnI. The digested fragments are amplified by PCR and can be determined by microarrays and sequencing.*

The major disadvantage of DamID is the overexpression of the exogenous Dam protein. Due to its processive methylation activity, very low expressions of Dam are usually sufficient to methylate DNA (Urig et al., 2002). Overexpression of Dam will therefore cause saturating methylation levels and create artefactual DNA binding sites, especially at open chromatin loci. This has been overcome by translation reinitiation-mediated DamID

(Southall et al., 2013). The expression of Dam is kept minimal by the phenomenon of ribosome reinitiation. This arises because the eukaryotic ribosome does not always detach from the mRNA at the stop codon of the first open reading frame (ORF) and therefore can carry on translating at the second downstream ORF on the transcript. The frequency of reinitiation is inversely proportional to the length of the primary ORF (Kozak, 2001). The Dam-fusion proteins are encoded downstream of a UAS enhancer as a secondary ORF. Efficiency of translation of the Dam-fusion protein is therefore greatly reduced. In *Drosophila*, the minimal expression of the Dam-fusion protein can also be restricted to specific cells using a GAL4 driver line.

Even though this modified technique has addressed many issues encountered with DamID, its application is still limited by its resolution. The resolution of DamID is limited by the median GATC interval (about 205 bp in *Drosophila melanogaster*, FlyBase release 6) in the genome and the extent of methylation from the Dam binding site (up to 5 kb from any given target locus) (van Steensel and Henikoff, 2000).

The main advantage of DamID is that it does not require fixation and specific antibodies that can be major sources of artefacts. Unlike ChIP, DamID does not give a snapshot of protein-DNA interactions, but rather gives an average of interactions over the span of a few hours and can therefore detect transient interactions.

3.1 Aims and objectives

To decipher the regulation of gene expression by the two FoxP isoforms, targeted DamID was chosen to identify their targets in all neurons in the fly brain. TaDa is efficient in fruit flies as GATC sequences are frequently present in the *Drosophila* genome (median GATC spacing being about 205 bp) and adenine methylation is transcriptionally and developmentally inert. In addition to providing an insight into the regulatory activity of FoxPC and FoxPD, the results from targeted DamID, coupled with sequencing (referred to as DamID-seq in this thesis) were also used for post-processing and functional analysis of the bound genomic regions. This allows for a better understanding of the regulatory and functional relationships between FoxPC and FoxPD and their targets.

Since the defining feature of TFs is their sequence-specific DNA-binding activity, it is of interest to characterise their intrinsic sequence specificity. Hence, one of the post-processing analyses that were carried out is *de novo* motif discovery using MEME. FoxPC and FoxPD binding sites were identified and compared to other known binding sites using TOMTOM. New putative co-regulators and co-factors were identified by correlating each DamID-seq data set to known cis-regulatory elements (CRMs) and transcription factor binding sites (TFBSs). The final analysis that was performed is the association of each significant peak to candidate target genes and the functional classification of those genes into gene ontology (GO) groups by Flymine. Classification of genes allowed the analysis of their function and their involvement in different networks.

3.2 Results

3.2.1 Validation of the targeted DamID protocol

To identify FoxPC and FoxPD direct targets using targeted DamID, flies were injected with either *mCherry-Dam-FoxPC* or *mCherry-Dam-FoxPD* constructs. Dam-only constructs were also used in control flies. The DamID experiment was carried out according to the well-established protocol by Marshall et al., (Marshall et al., 2016). The *Dam-FoxPC* and *Dam-FoxPD* transgenes were expressed in all neurons by using *elav-GAL4*. 200 brains were dissected per replicate and genomic DNA was extracted. Methylated sites were digested by DpnI, ligated to a PCR adaptor and amplified before being sequenced. Three biological replicates were performed for Dam-FoxPC, Dam-FoxPD and Dam-only control. The quality of the samples sent for sequencing was verified (Figure 7). A smear, due to the amplification of all GATC fragments, ranging from 100 bp to 2500 bp was seen for all samples treated with DpnI and T4 ligase.

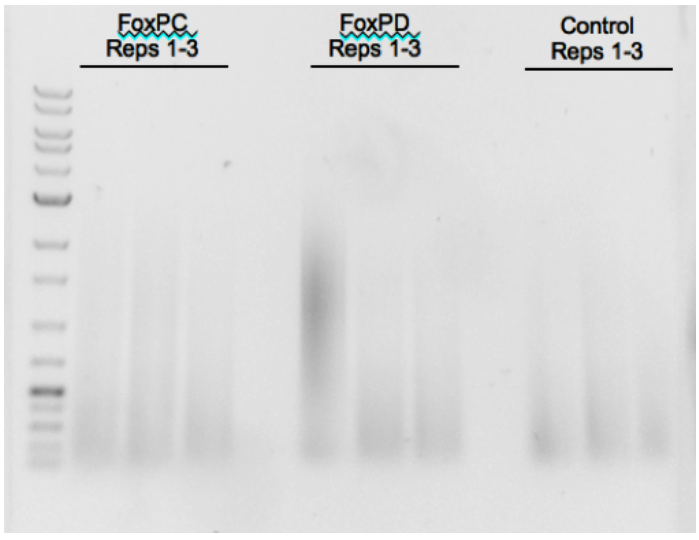


Figure 7. Successful amplification of adenine-methylated fragments.
Three biological replicates for Dam-FoxPC, Dam-FoxPD and Dam-only controls were tested after PCR amplification for the presence of amplified DNA fragments.

3.2.2 Analysis of DamID-sequencing data

Sequencing of the DamID products generated about 10^6 high quality reads per sample, 99% of which mapped to the *Drosophila* genome (see Appendix 1 for the number of reads obtained for each sample). The quality of the sequencing data was quickly analysed using FastQC (see Appendix 2 for individual graphs of read quality). The "per base sequence quality" chart showed reliable base calls across the entirety of the reads and the "sequence quality histogram" showed that the average quality per read was high (Phred scores: 30–40) (Appendix 2 right and left panel, respectively).

The DamID-seq analysis was carried out as shown in the flowchart (Figure 8). In brief, the DamID adapter sequences were first trimmed from each sequencing fragment. The reads were mapped to the *Drosophila* genome, extended, counted and visualised. Differentially enriched GATC fragments for FoxPC-Dam vs Dam-only or FoxPD vs Dam-only were then called and fragments with an FDR < 0.01 and $\log_2$ fold change >1.5 were extracted. Downstream analyses included searches for *de novo* and known enriched motifs in the peaks, annotation of the peaks, analysis of GO term enrichment and network analysis.

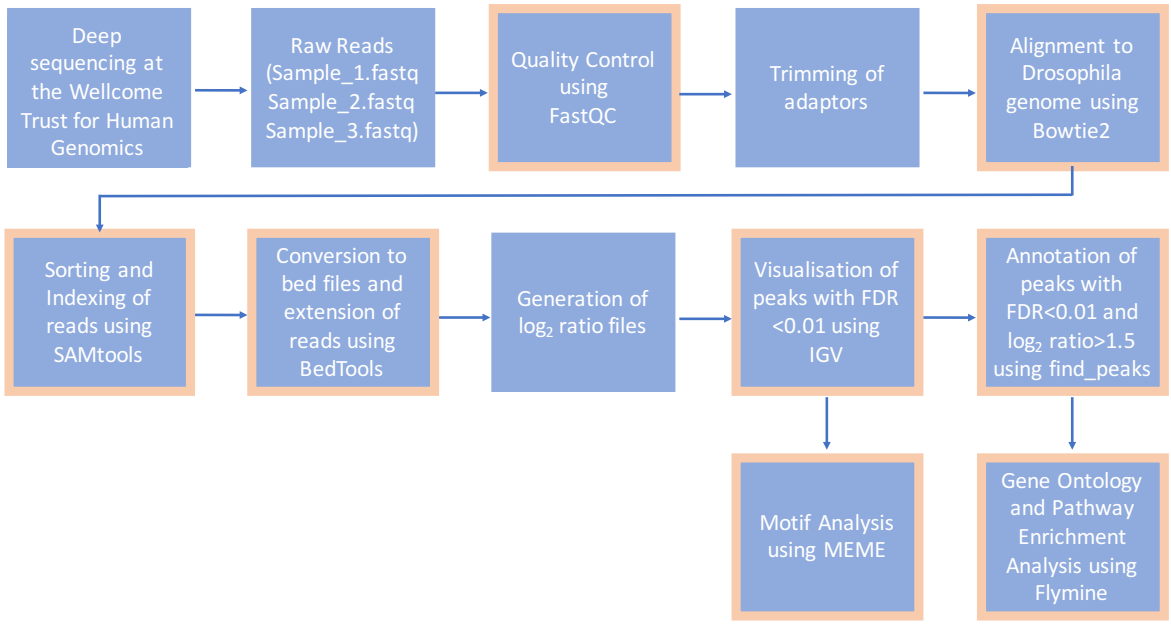


Figure 8. Flowchart of the DamID-seq analysis pipeline.

Quality control was performed on the raw DamID-seq reads using FastQC; the reads were then mapped to the Drosophila genome using Bowtie2. Log_2 ratio peak files were then generated comparing Dam-FoxPC and Dam-FoxPD to Dam-only controls. Those peaks were analysed for common binding sequences using MEME. Genes associated with the peaks were retrieved and concatenated into Gene Ontology terms using Flymine. Boxes with no outline represent the input or output files. Boxes outlined show the task done and the tool used.

The three biological replicates were compared to each other to ensure good correlation between them. After aligning to the *Drosophila* genome, the number of reads at every genomic location was counted and visualised in IGV. The resulting FoxPC-Dam and FoxPD-Dam methylation profiles show consistent binding peaks for the replicates at specific locations, confirming the reproducibility of the technique (top and middle panels in Figure 9). The binding profile of the Dam protein (bottom panel in Figure 9) shows that it binds at basal levels throughout the genome, with a few prominent peaks which occur at accessible chromosomal regions. These peaks, however, are clearly distinct from the additional strong binding peaks associated with FoxPC and FoxPD binding, revealing the success of the experiment.

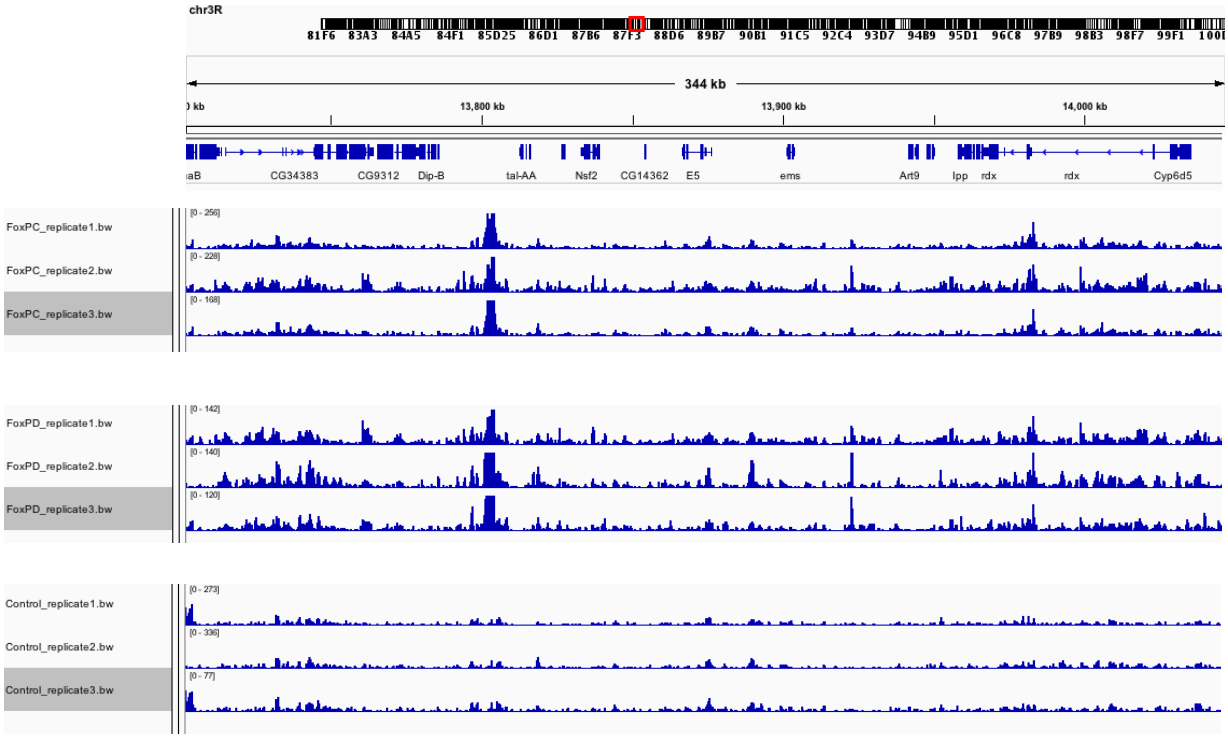


Figure 9. Distribution of mapped reads at a specific region of genome for the three replicates.
The top panel shows the highly correlated sequencing coverage between the three biological FoxPC-Dam replicates. The middle panel shows the correlation between the three FoxPD-Dam biological replicates and the bottom panel shows the correlation between the Dam replicates. The y-axis represents the number of reads at a specific location in the genome and the x-axis shows the reads at different locations on chromosome 3R.

Given that this is a crude way of determining the correlation between the replicates, the total correlation between replicates, calculated by comparing all the reads in the genome, was determined. This was done by comparing read numbers found in 50 bp bins across the genome. The three FoxPC and FoxPD replicates show high pairwise Pearson correlation coefficients of 0.87-0.99 and 0.88-0.92, respectively (Figure 10A and 10B). These reasonably high values confirm the robust experimental procedure and specific binding of FoxPC and FoxPD to their targets. The different correlation coefficients between replicates can be explained by variation in the total sequencing reads and therefore the extent of coverage. Comparing the three Dam-only replicates to each other results in much

lower correlation coefficients of 0.14-0.69 (Figure 10C), consistent with a lack of notable binding preferences and random binding at low levels across the entire genome.

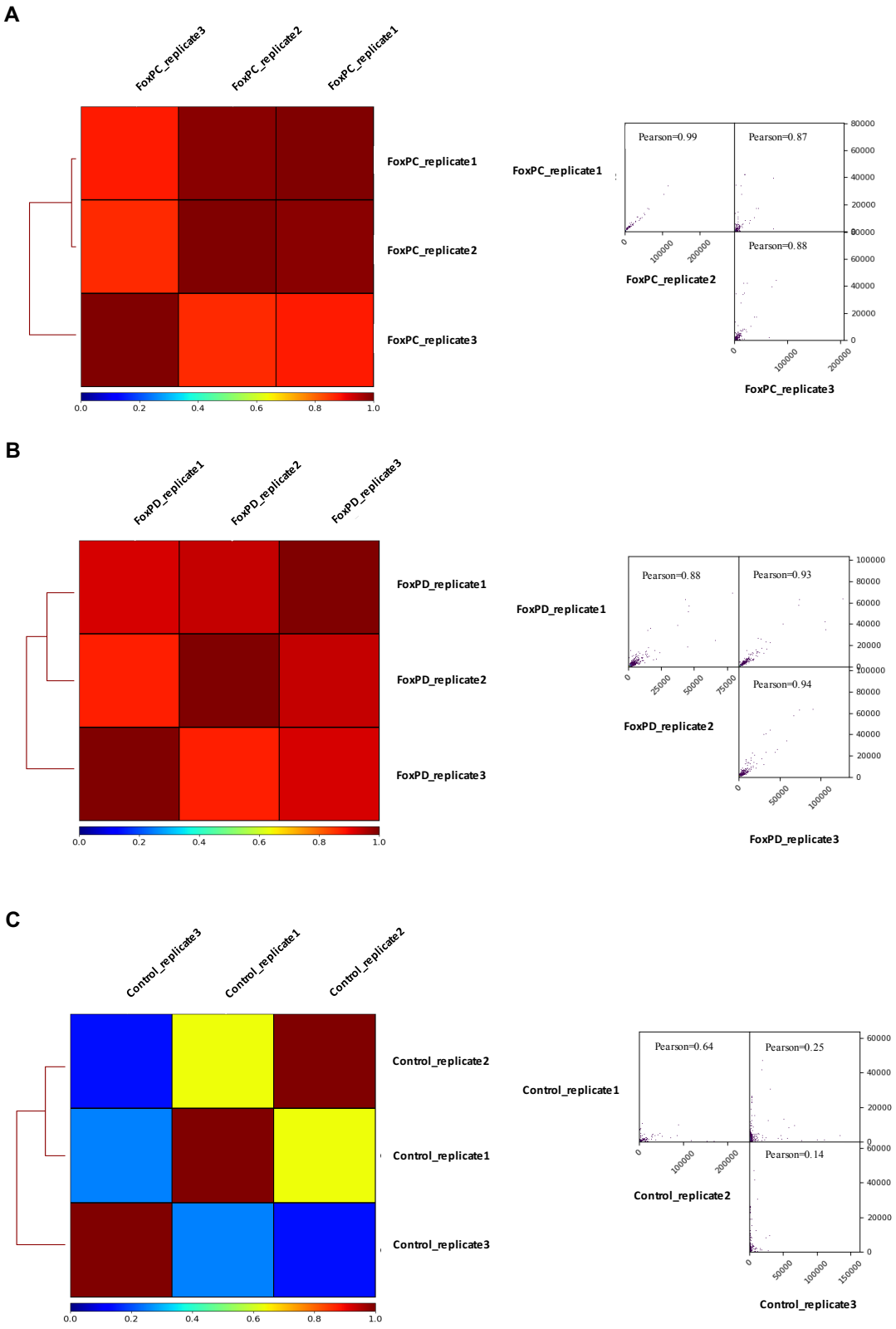


Figure 10. Correlation between the biological replicates.

*The left panels show heatmaps of Pearson correlation of the sequencing coverage of the three biological replicates (**A**- Dam-FoxPC, **B**- Dam-FoxPD, **C**-Dam-control). The colour intensities represent the correlation coefficients. The right panels show scatter plots demonstrating the*

pairwise correlations. The scatter plots on the right show that there is a high correlation between the three Dam-FoxPC (0.87-0.99) and Dam-FoxPD replicates (0.88-0.94). The correlation between the Dam-control pairs ranges from low (0.14) to moderate (0.64).

Since the DamID-seq seems robust, yielding three reproducible datasets, the reads for all replicates were fed to the custom bioinformatics pipeline by Owen Marshall (Marshall et al., 2016). By normalizing the Dam-fusion experiments to the Dam control, a false discovery rate value (FDR) was assigned to each peak. A highly stringent FDR of < 0.01 was chosen to exclude any false positive peaks and to obtain only significant binding intervals. Peaks with an FDR greater than 0.01, and found outside any gene boundary or more than 2 kb upstream or downstream of the transcription start and end sites were discarded. This resulted in more than 800 peaks being called for the Dam-fusion protein replicates (Table 1).

Table 1. Number of peaks with an FDR < 0.01 from the fusion-Dam experiments.

Sample	Number of peaks
FoxPC_replicate1	897
FoxPC_replicate2	871
FoxPC_replicate3	808
FoxPD_replicate1	931
FoxPD_replicate2	831
FoxPD_replicate3	957

To eliminate any random Dam binding to open chromatin, $\log_2$ ratio binding profiles were generated for the Dam-fusion proteins. The numbers of reads at each position were counted in the control samples and subtracted from the number of reads at the same position for the fusion proteins. An example of normalised DamID data for one replicate

of FoxPC-Dam and FoxPD-Dam is shown in Figure 11. The methylation maps show strong signal at certain regions, resulting from binding of FoxPC or FoxPD. Regions are considered as being bound by FoxPC or FoxPD only if they have at least 1.5-fold more normalised reads than the Dam-only controls.

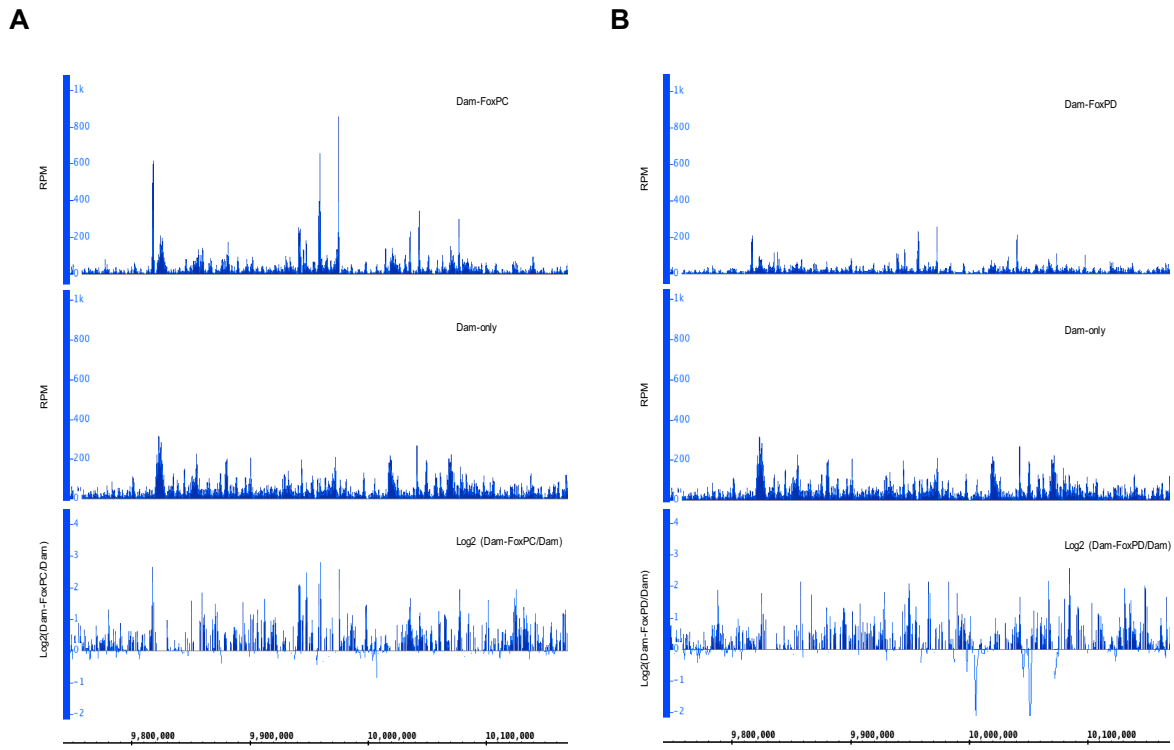


Figure 11. Binding of FoxPC and FoxPD is enriched at specific sites compared to controls.
*Binding profiles for Dam-FoxPC (**A**) or Dam-FoxPD (**B**) (top panels) and Dam-control (middle panels). Only peaks with an FDR < 0.01 are plotted. The bottom panels show the enrichment of methyladenine marked DNA from FoxPC- or FoxPD-Dam over Dam-only controls (log_2 scale). The y-axes of the top two panels show the number of reads.*

3.2.3 Association of FoxPC and FoxPD occupancy to genomic features

Transcription factors function by binding to specific transcriptional regulatory regions, for example promoters and enhancers (or silencers), to control the expression of target genes. It is crucial to elucidate the binding pattern of FoxPC and FoxPD to better understand their regulatory roles in *Drosophila melanogaster*. In order to do so, the distribution of the peaks in relation to genomic features, for example introns, exons, intergenic regions, 5'UTR and 3'UTRs (Figure 12A), was determined. On the basis of the *Drosophila* genome annotation from Flybase r6.16, the significant peaks (FDR < 0.01 and $\log_2$ fold change > 1.5) were mapped onto the genome. Promoter regions were defined as regions 1000bp from the transcription start site (TSS).

Intersecting the DamID peaks with the features gave similar distribution patterns for the different Dam-FoxPC biological replicates. Most of the peaks lie within genic regions rather than intergenic regions (Figure 12B). Within genes, 19.34% of the binding intervals are found in promoter regions and 54.40% in the introns. These two features account for almost 75% of the total features corresponding to peaks. A moderate percentage (10.70%) of the peaks falls within exons, intergenic regions (6.98%) and 5'UTRs (5.78%). The 3'UTRs contain the least number of FoxPC peaks (2.79%). The intergenic regions represent all the regions outside the boundaries of a gene. The high percentage of binding of FoxPC and FoxPD to introns and promoters is reassuring as this is in agreement with their regulatory role in transcription. The lower number of peaks which were seen at the promoter regions upstream of the TSS can be explained by a much lower GATC occurrence at those positions. A similar observation was made by Aughey et al., (Aughey et al., 2018).

Since methylation occurs depending on the random occurrence of GATC sites close to the binding sites of Dam or Dam-fusion proteins, it is difficult to determine exactly where the exact binding site is within a binding interval. Therefore, all the GATCs in the genome were also mapped onto the same genomic features to better understand the distribution of the DamID peaks. Most of the GATCs are found in exons and introns (34.53% and 37.98% respectively) (Figure 12B). The high intronic GATC distribution and the generally high distribution of introns in the genome could explain why the majority of the Dam-FoxPC and Dam-FoxPD peaks are associated with intronic regions.

It should be noted, however, that the distances between binding intervals and genomic features do not always accurately reflect the distances between the actual binding sites and the features. Therefore, even though associating each peak to genomic features gives a good indication of where they might possibly bind, in order to regulate transcription, it should be taken into account that DamID does not always accurately represent the true binding sites of the transcription factors under study as there is a bias towards GATC-rich loci.

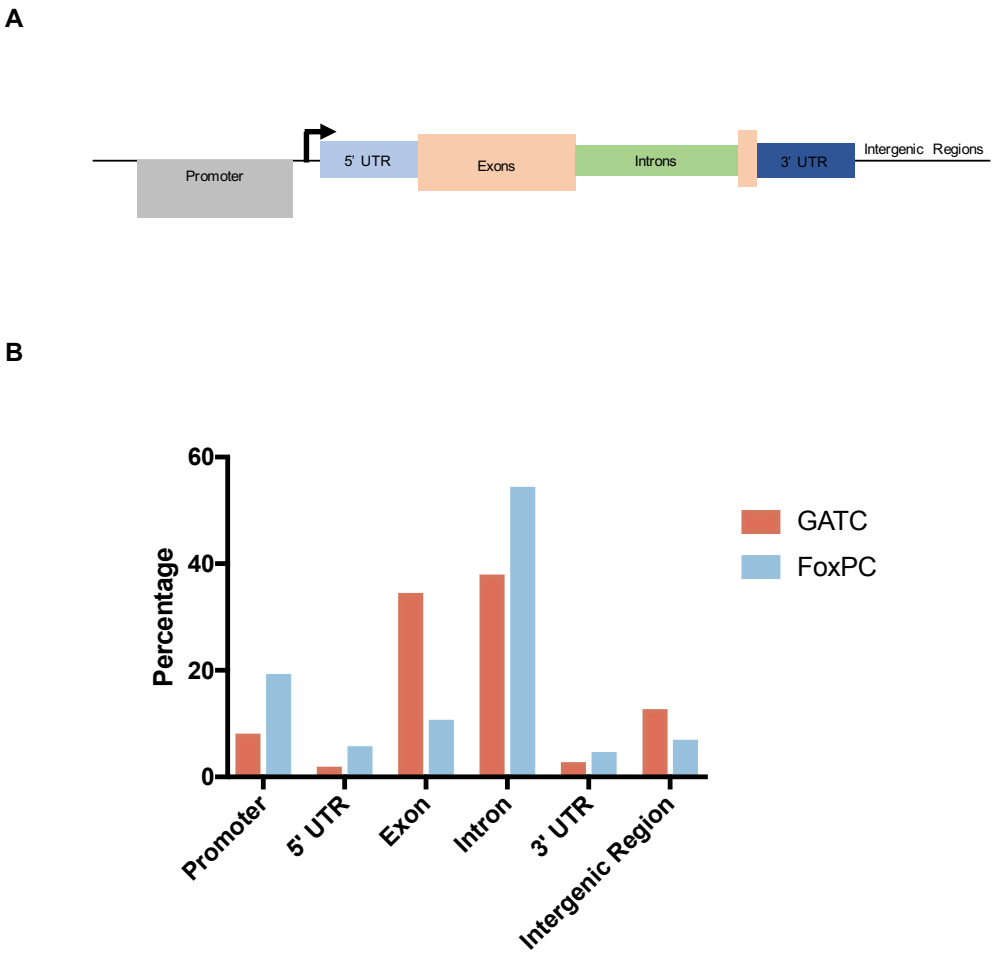


Figure 12. Assigning binding intervals to genomic features.

A, Features of a gene. The arrow represents the transcription start site. ***B****, Percentage of genes mapped to each feature. Red bars represent the GATC coverage in genome, while the blue bars correspond to FoxPC bound regions.*

Although mapping of the peaks to the functional regions revealed the distribution over the genome, the intensity of binding at those regions was not determined. Histograms for peak scores associated with a specific genomic region (in this case, the TSS) for all genes were plotted using the deepTools package (Ramirez et al., 2016). The peaks were sorted into 50 bp bins and an average binding profile was calculated and plotted on a meta-gene (Figure 13). The results show that most of the peaks lie very close to the TSS (within 500 bp). Even though there is a high binding affinity on either side of the TSS, there is a slight skew towards downstream binding sites, which suggests a high amount of binding to

regulatory regions, such as enhancers or silencers, in the introns. The highest binding was seen around the TSS, with the intensity of binding decreasing over gene bodies after the TSS. Regions downstream of the TSS were also highly bound, which potentially means that FoxPC binds to the proximal core promoter regions. Therefore, it can be concluded that FoxPC and FoxPD mainly bind very close to the TSS regions.

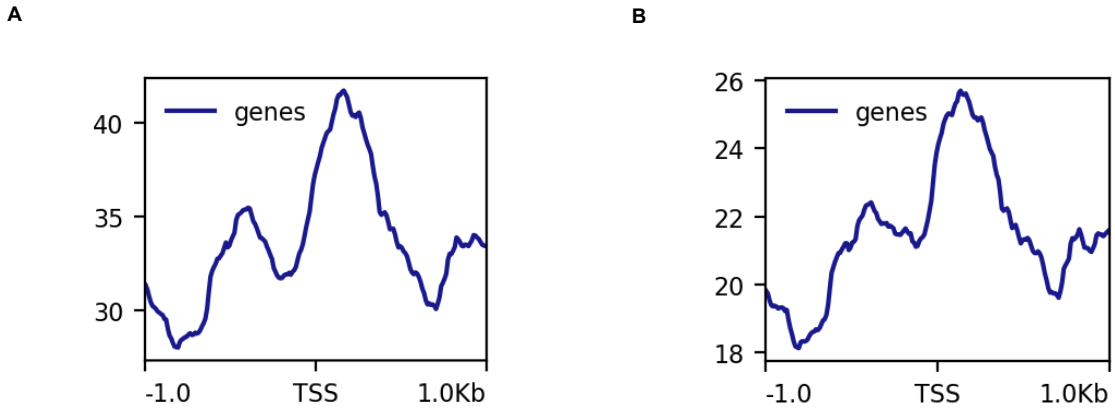


Figure 13. Metagene profiles of FoxPC and FoxPD binding.

Average binding profile of peaks binding to TSS regions (± 1000 bp) over all genes scaled into a meta-gene. The average number of reads per 50 bp bins are shown on the y-axis, relative to the meta-gene coordinates shown on the x-axis.

To further investigate the interaction of FoxPC and FoxPD with regulatory regions in introns, the significant peaks were also mapped onto known *cis*-regulatory modules (CRMs) obtained from the REDfly database (Gallo et al., 2011). These CRMs include enhancers and silencers outside of core promoter regions. The locations of 16,625 known *Drosophila* transcription CRMs were compared with the DamID-seq data from FoxPC and FoxPD. About 28% of the peaks from the datasets overlap with 607 known CRMs, providing an insight into the potential targets of FoxPC and FoxPD. Some of the CRMs bound by FoxPC uncovered by this analysis include Ftz-f1, eve, rpr, Sox15, lola and gcm.

The relatively low value of mapped CRMs is expected as not every CRM in the *Drosophila* genome has been characterised. Also, the peaks spanning the GATC regions might not have the minimum degree of overlap, as set by the 'intersectBed' function in bedTools, with known CRMs to be considered significantly overlapping.

To identify enriched regulatory features (including CRMs and motifs) *in silico*, i-*cis*Target (http://med.kuleuven.be/lcb/i-cisTarget) was used. The significant peaks (FDR < 0.01 and $\log_2$ fold change > 1.5) were treated as co-regulated sequences and mapped against publicly available known and inferred regulatory datasets (including CIS-BP (Weirauch et al., 2014), TRANSFAC (Matys et al., 2006), JASPAR (Mathelier et al., 2014) and FlyFactorSurvey (Zhu et al., 2011a).

i-cisTarget retains regions with a minimal overlap of 40% with any known or inferred CRMs (set of 136K regions covering the entire non-coding genome). The analysis was carried out using bedfiles containing significant binding intervals obtained from the DamID-seq analysis. All parameters were set to default settings and only motifs with an enrichment score > 3 were extracted. The full list of motifs is shown in Appendix 3. The top 12 enriched sequences for FoxPC and FoxPD are shown in Figure 14. Almost all of the top enriched motifs extracted correspond to forkhead binding domain proteins, for example slp1, foxo, FoxK and croc for the FoxPC dataset and croc, slp1, FoxK for the FoxPD dataset. Sox-related binding motifs make up the rest of the top enriched motifs. Both top enriched motifs, slp1 and croc (enrichment ratios 5.71157 and 5.18648) are forkhead domain proteins which have repressive roles on transcription activity in the central nervous system, a role in which FoxPC and FoxPD have been implicated.

Strikingly, the inferred binding motif of FoxP (enrichment score of 4.20 in FoxPC and 4.86 in FoxPD) is listed as one of the most enriched CRMs. This strongly indicates that the DamID-seq peaks obtained correspond to the actual binding sites of FoxPC and FoxPD and confirms the validity of the DamID-seq experiments. The slight differences in the binding motifs obtained for the two isoforms can be explained by their different DNA binding domains.

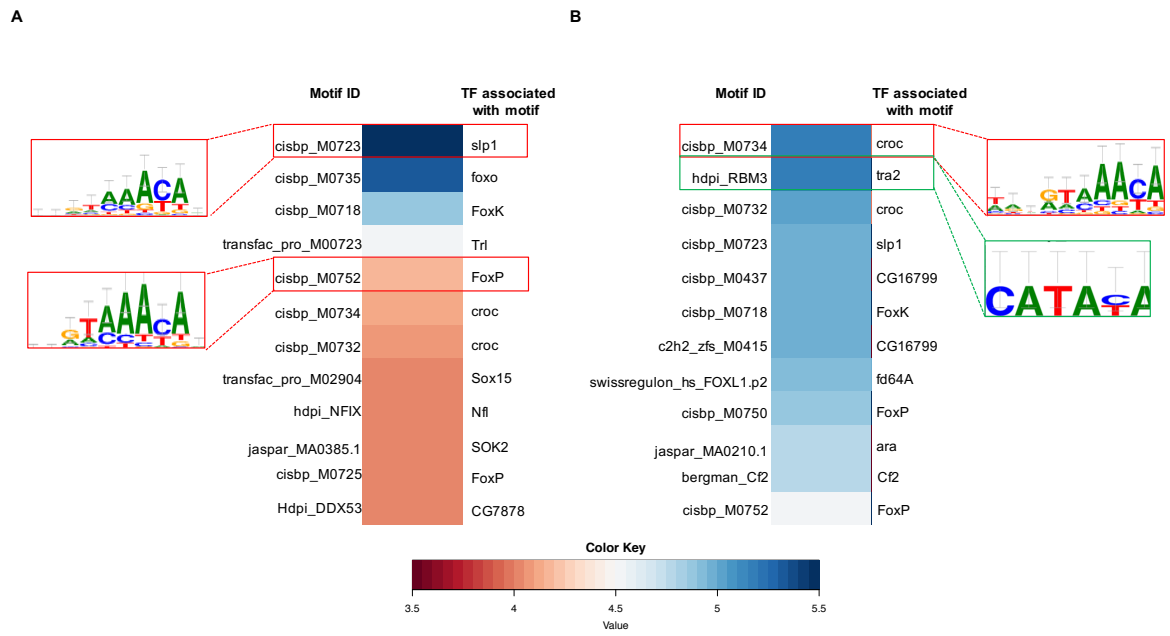


Figure 14. i-cisTarget motif analysis of FoxPC and FoxPD datasets.

*A heatmap of the 12 most enriched motifs is shown for both FoxPC (**A**) and FoxPD (**B**) and are ranked according to their enrichment scores (as dictated by the colour key). Motif IDs are shown on the left and the associated candidate transcription factors on the right. The most enriched motif is boxed in red for both FoxPC and FoxPD and the sequence logos are shown on the left and right for FoxPC and FoxPD, respectively. The inferred FoxP motif and the known tra2 motif are also shown.*

3.2.4 Motif Identification

Specific transcription factors recognise and bind to similar DNA sequences, or motifs, which are represented using models such as position weight matrices (Inukai et al., 2017).

Since there is a strong indication that the DamID-seq experiments are robust and that the peaks obtained are regions known to be bound by forkhead domain proteins, an attempt was made to predict the binding motifs of FoxPC and FoxPD *de novo*.

The Multiple Expectation Maximization for Motif Elicitation (MEME)-suite was used to analyse regions underlying significant DamID peaks and scans the list for recurring patterns. The number of potential motifs was set to 15, with motifs having an optimum width ranging between 5 and 15 residues, with zero or one occurrence per sequence. Only motifs with an e-value <0.05 were considered significant. A number of motifs was obtained, with the 10 top most significant motifs being shown as position weight matrices in Figure 15. The most significant motif for both FoxPC and FoxPD datasets is the CACACA motif. However, this is likely a false positive as it is a low-complexity microsatellite repeat DNA, which is present in a large number of sequences (Ellegren, 2004). The second most significant motif (A(A/G)CAACAA) is very similar to the motif predicted by Nita et al. (AAACAACAACAA) by SELEX (Nitta et al., 2015) and contains the core region of the motif inferred by Cis-BP (Figure 15). This motif appears 630 times in the regions bound by FoxPC, as determined by the MEME-tool, Find Individual Motif Occurrences (FIMO). This relatively high occurrence rate infers that the FoxPC binding sites uncovered are robust. With regards to all the evidence pointing towards the robustness of the DamID-seq data and to its similarity to inferred motifs, A(A/G)CAACAA likely represents the genuine FoxPC-binding motif.

The two FoxP isoforms have slightly different sequence specificities, as would be expected on the basis of their different DNA binding domains. In the case of FoxPD, the most significant motifs include CAG(C/A)G(C/A)(A/T)(G/T)(C/A), ATA(T/C)AT and

(G/C)GAAA(A/T). The putative FoxPC motif recurs as the ninth most enriched motif for FoxPD, albeit with a much lower e-value than in the FoxPC dataset. The third motif corresponds to the tra2 motif (which was found to be the second most enriched motif by i-cisTarget analysis (Figure 14). This suggests that this motif could be the putative binding motif for FoxPD.

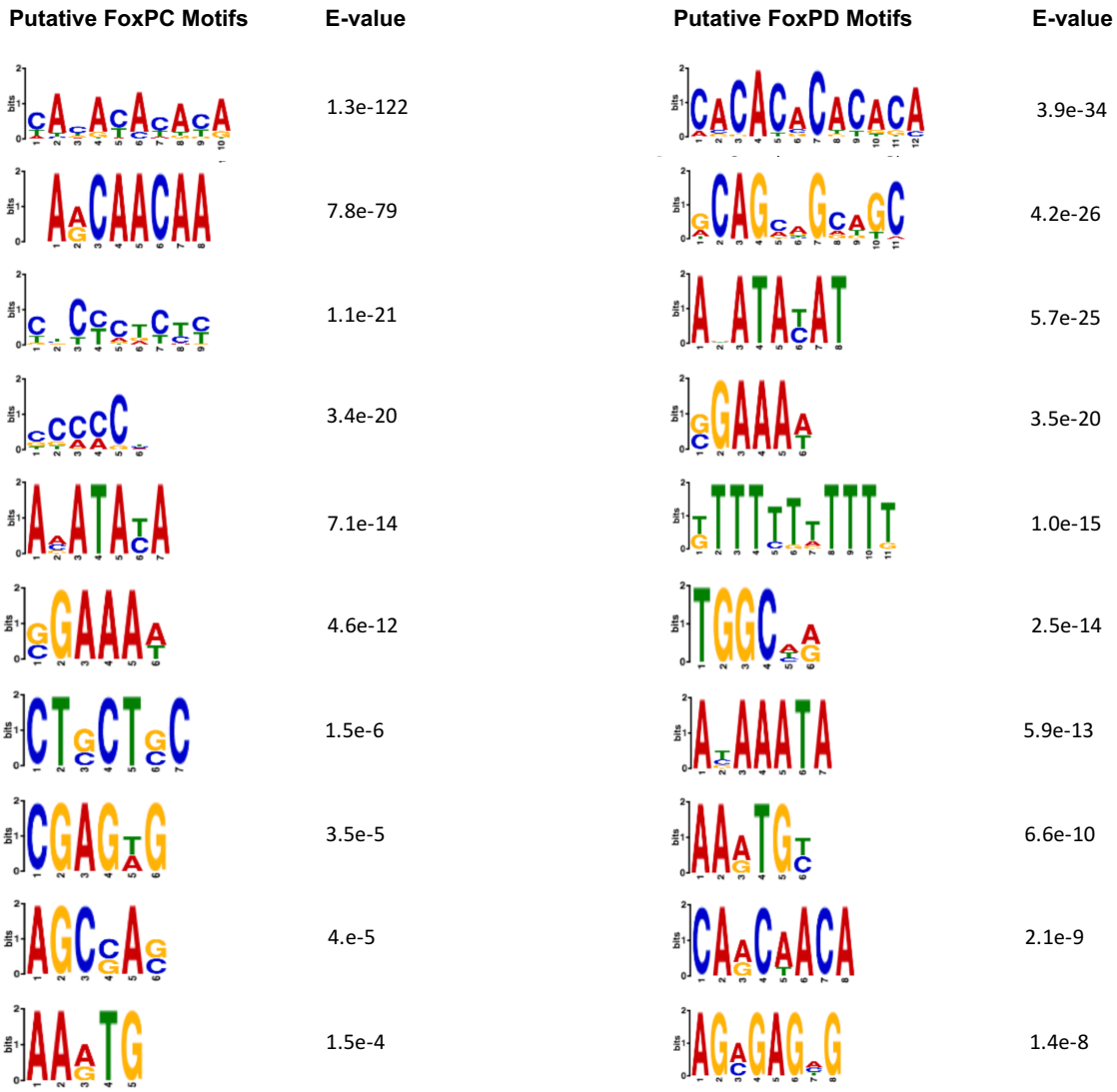


Figure 15. Motif Analysis using MEME.

Sequences underlying the most significant peaks were used to determine enriched regions in the dataset. The ten most significant FoxPC motifs are shown on the left and the FoxPD motifs on the right. The e-value corresponds to the significance of the motif. The overall height of each stack indicates the consensus sequence at that position (measured in bits) and the height of each nucleotide symbol (A, C, G or T) within the stack reflects the frequency of occurrence of the nucleotides at that position.

As Cis-BP only provides inferred motifs, the A(A/G)CAACAA motif was used as input for comparison with any known motifs, using TOMTOM which is part of the MEME-suite package. TOMTOM quantifies the similarity between the query motif and known motifs and ranks the latter in order of similarity and significance. Default values were used for the TOMTOM search.

TOMTOM identified 13 motifs from the Fly Factor Survey database and the OnTheFly database as similar to the putative FoxPC motifs. The most significant matches (p-values between 1.30e-05 and 3.58e-03) include the Adult enhancer factor 1(Aef1), longitudinals lacking (lola) and the hunchback (hb) binding motifs (Figure 16). These proteins are zinc finger binding domain proteins, with known repressive functions in the fat body (Aef1) (Falb and Maniatis, 1992) and nervous system development (Lola and hb) (Gates et al., 2011, Novotny et al., 2002). It is possible that FoxPC, via its zinc finger domain, binds to those proteins to form a complex which is involved in regulation of gene expression.

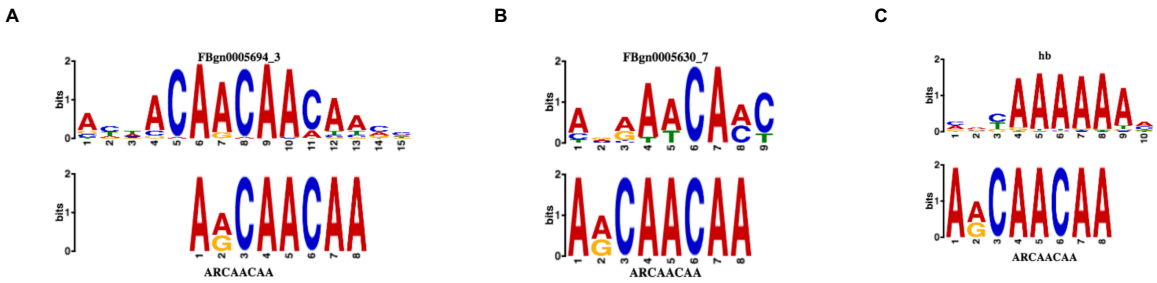


Figure 16. Comparison of FoxPC putative motif with known motifs by TomTom.
*Alignment of putative FoxPC motif (bottom panel) to Aef1 binding motif (**A**), Lola binding motif (**B**) and hb binding motif (**C**) using TOMTOM.*

3.2.5 Characterisation of target genes

To identify the putative targets of FoxPC and FoxPD, the positions of the binding intervals were compared to those of gene bodies. Only peaks with an FDR < 0.01 and $\log_2$ fold change > 1.5) were used to identify putative target genes. The genes were called when the peaks directly overlapped with the gene regions, or in the absence of a direct overlap, the closest gene to the peak was called.

A list of 535 genes was produced for the Dam-FoxPC replicates and 615 genes for the Dam-FoxPD-replicates, respectively. The full list of genes and the $\log_2$ fold change are appended (Appendix 4). Some of the top targets identified from the DamID-seq were previously identified as targets of FoxP2, for example Mnt (Spiteri et al., 2007) and semaphorins (Vernes et al., 2011). The lists also contain target genes whose REDfly-documented CRMs overlap with FoxPC and FoxPD binding intervals, for example Ftz-f1, Sox, rpr and ci.

The lists of genes for each isoform were compared to each other to identify common and unique targets. About 35.7% of the FoxPC targets (130 genes) were also found to be FoxPD targets (Figure 17), suggesting that although the two isoforms regulate common genes, the majority of targets are in fact unique genes, potentially giving rise to differential expression patterns and isoform-specific roles.

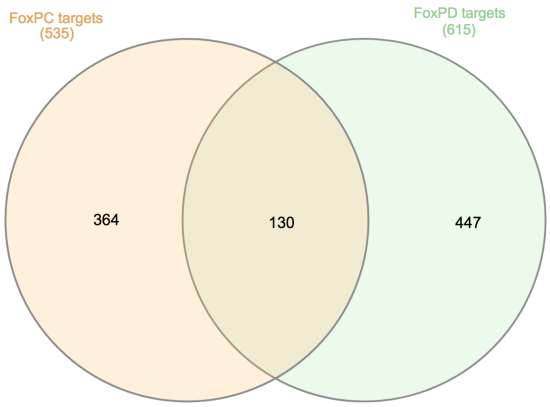


Figure 17. FoxPC and FoxPD regulate common and unique targets.

Venn diagram showing the overlap between the two sets of genes regulated by FoxPC and FoxPD. Significant binding intervals (FDR<0.01) and $\log_2$ ratio>1.5 were assigned to the overlapping or closest gene (within 2 kb). About 35.7% of the datasets are common between the two isoforms.

3.2.6 Identification of pathways, molecular functions and networks of target genes

To analyse the functional characteristics of the target genes, they were then used as input for gene ontology and network analysis using Flymine (http://www.flymine.org/flymine/bag.do) (Lyne et al., 2007) and Cytoscape for both FoxP isoforms. Flymine searches for biased representations of gene function annotations within lists of genes and assigns a Holm-Bonferroni corrected p-value to each biological process GO term. The p-values represent the probability that the number of query genes will be present in that GO term when compared to the size of the query list and the number of genes normally present in this GO category in the whole genome. This allows the identification of terms that are over-represented in the dataset. The full list of gene ontology terms, their p-values and the associated genes are shown in Appendix 5.

The FoxPC datasets resulted in 145 terms with a p-value <0.05 and the FoxPD datasets yielded 92 significant GO terms. The top 20 terms for each data set are shown in the heatmap below, GO term associated with FoxPC target genes being on the left and those

linked to FoxPD targets being on the right (Figure 18). GO terms with a lower p-value are shown in red and those with a higher p-value are shown in blue.

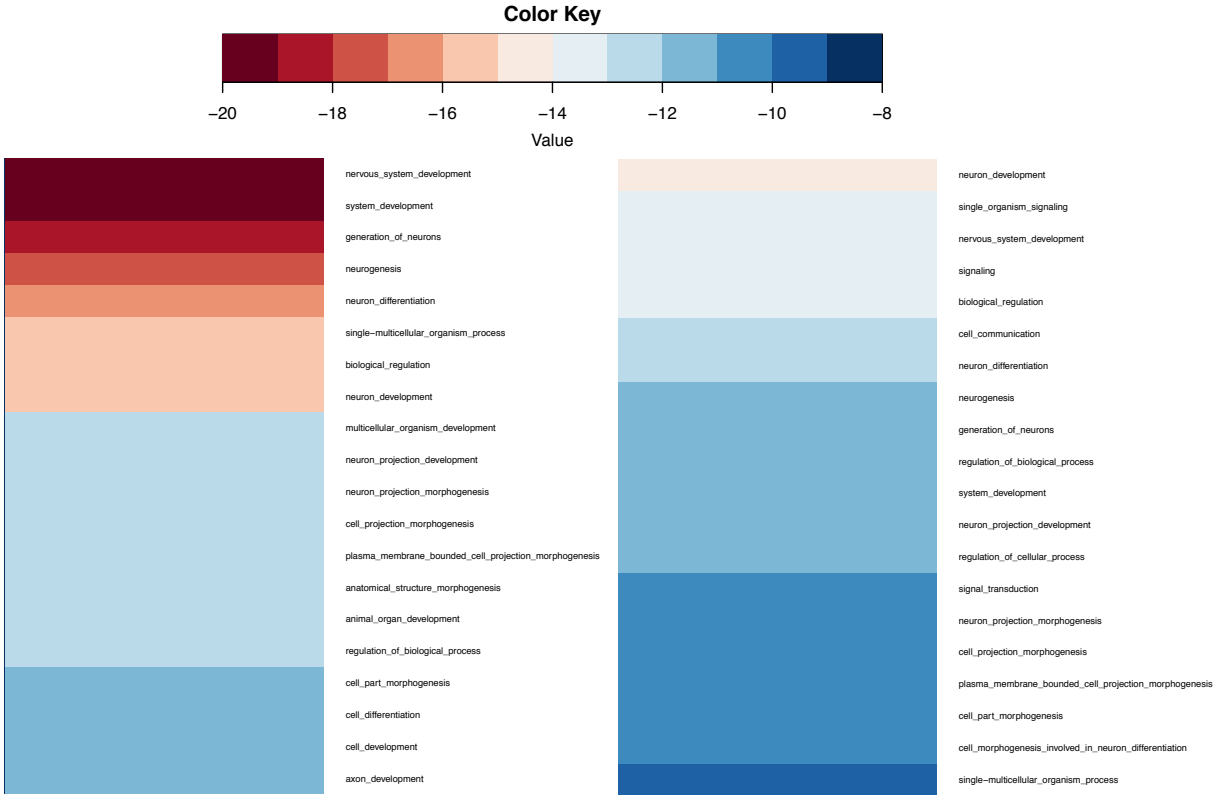


Figure 18. FoxPC and FoxPD targets are implicated in various functional processes.
GO term analysis for targets of FoxPC (right) and FoxPD (left). The scale represents the log p-value, with a low p-value being shown in red and a high p-value in blue.

Since the known roles of FoxP are attributed to its functions in the nervous system (Lawton et al., 2014, Mendoza et al., 2014, DasGupta et al., 2014, Groschner et al., 2018), it is expected that FoxPC and FoxPD targets would be nervous system-related. Indeed, the most significant GO term for both FoxPC and FoxPD targets is “nervous development” (p-values 5.25E-21 and 5.28E-15). Other significant NSD descendent ontology terms in the list include “generation of neurons”, “neuron differentiation”, “neuron development” and “axon development” (Figure 19). The GO analysis suggests that most of the target genes regulated by FoxPC and FoxPD have neural functions. This is in agreement with

GO categories associated with FoxP2 targets in humans (Vernes et al., 2007, Konopka et al., 2009, Spiteri et al., 2007). In addition, overrepresentation of genes involved in signaling processes was also observed (Figure 19), which is not surprising as proper development of the central nervous system involves multiple signaling pathways, for example retinoic acid, Wnt and Notch signaling pathways.

The same GO terms appear in the global GO enrichment analysis of biological functions for FoxPC and FoxPD, albeit with different degrees of significance. However, unique GO terms are also present. In addition to NSD categories, FoxPC also seems to be involved in organ development, including eye and wing morphogenesis, and cognitive processes, such as learning and memory (Appendix 5). FoxPD is uniquely involved in metabolic processes, such as the regulation of primary and RNA metabolic processes. This gives an indication of the biological processes that set the two isoforms apart.

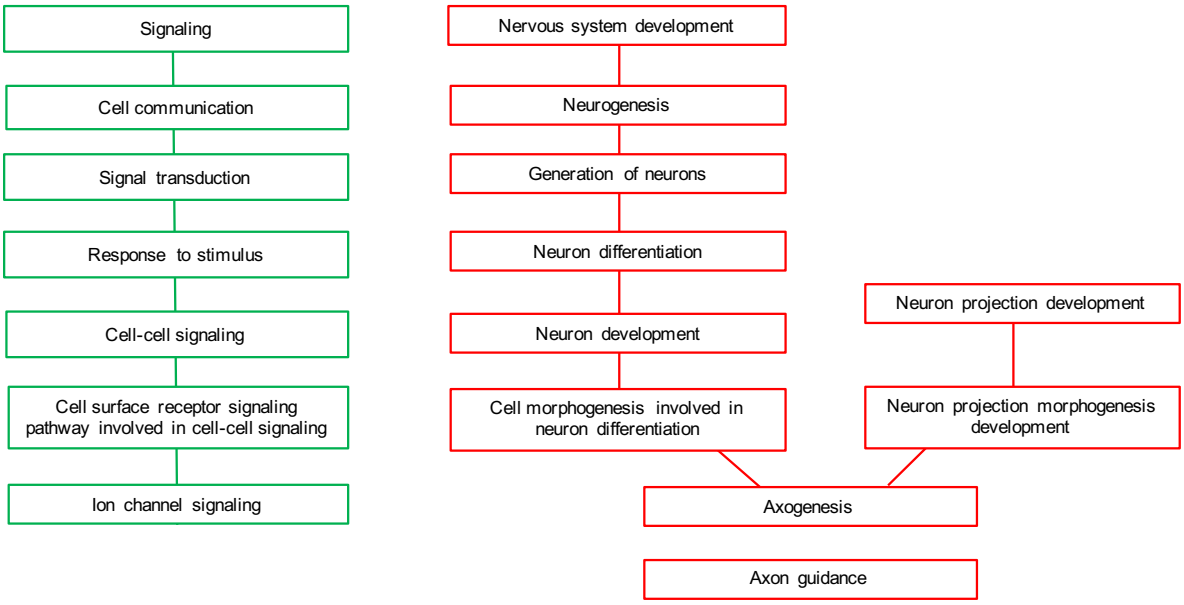
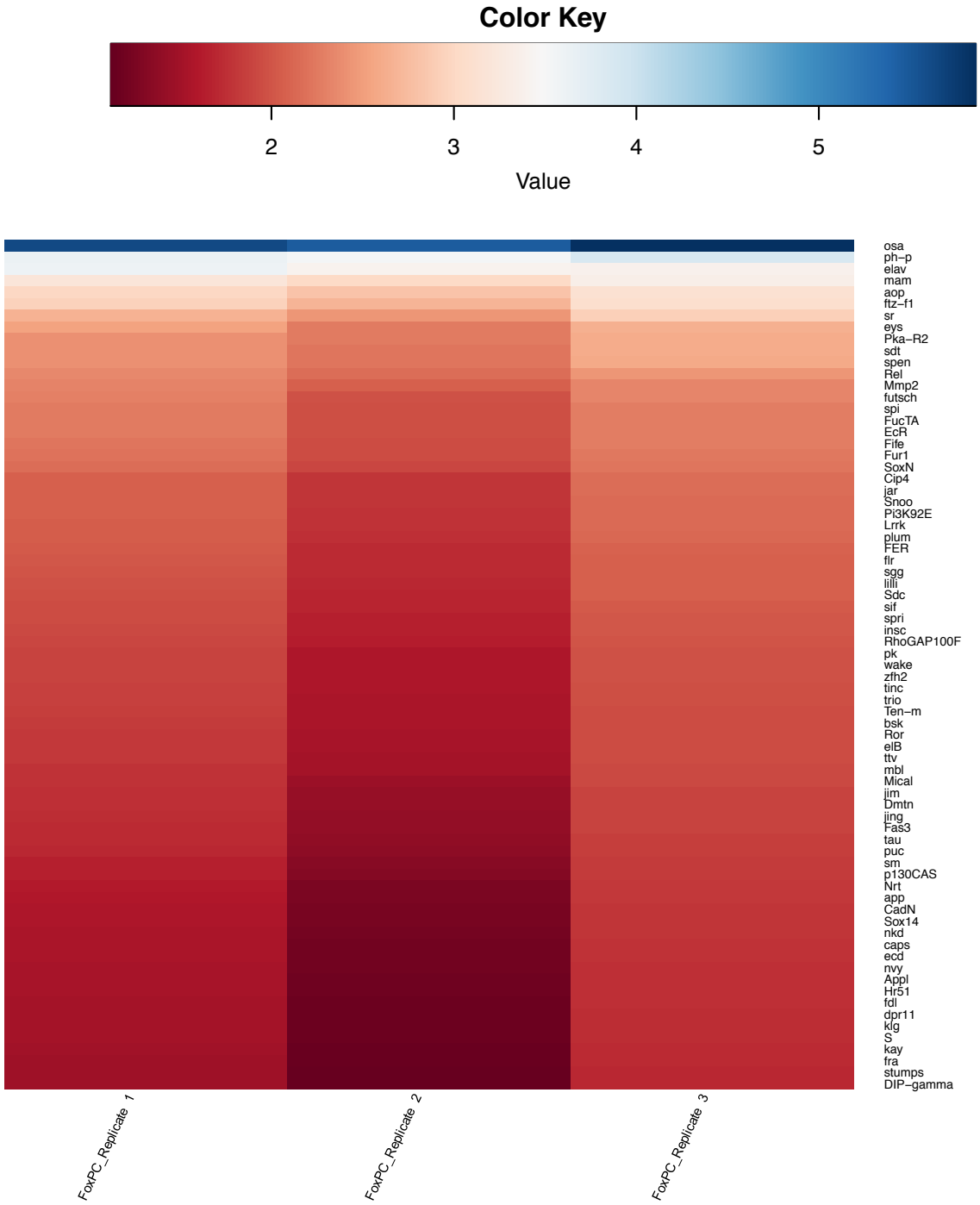
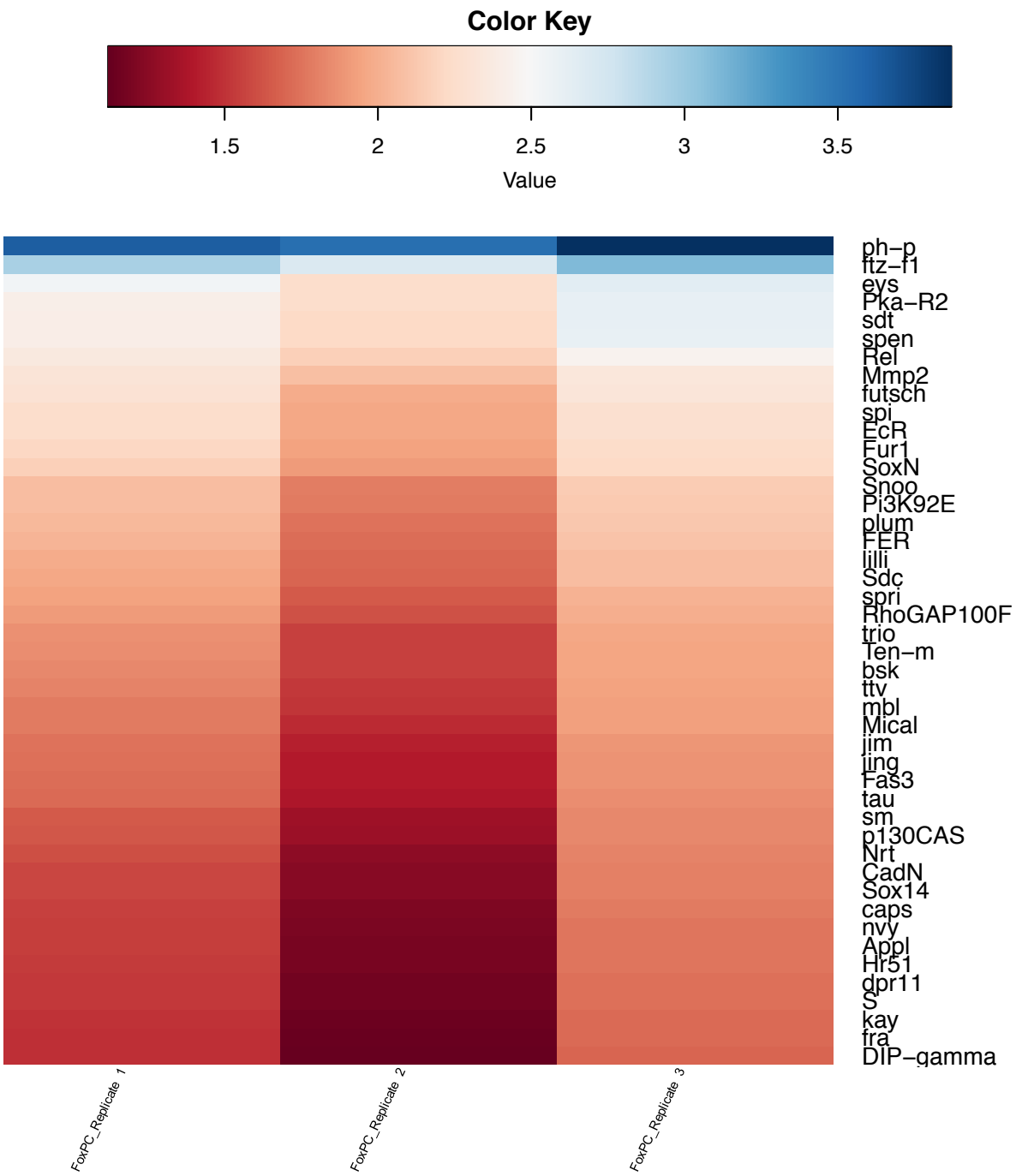


Figure 19. Breakdown of the signaling and nervous system development GO categories.
Most significant gene ontology categories suggest that FoxP plays role in known functional pathways. The descendants of the two most significant GO terms, signaling and nervous system development, are boxed in green and red, respectively.

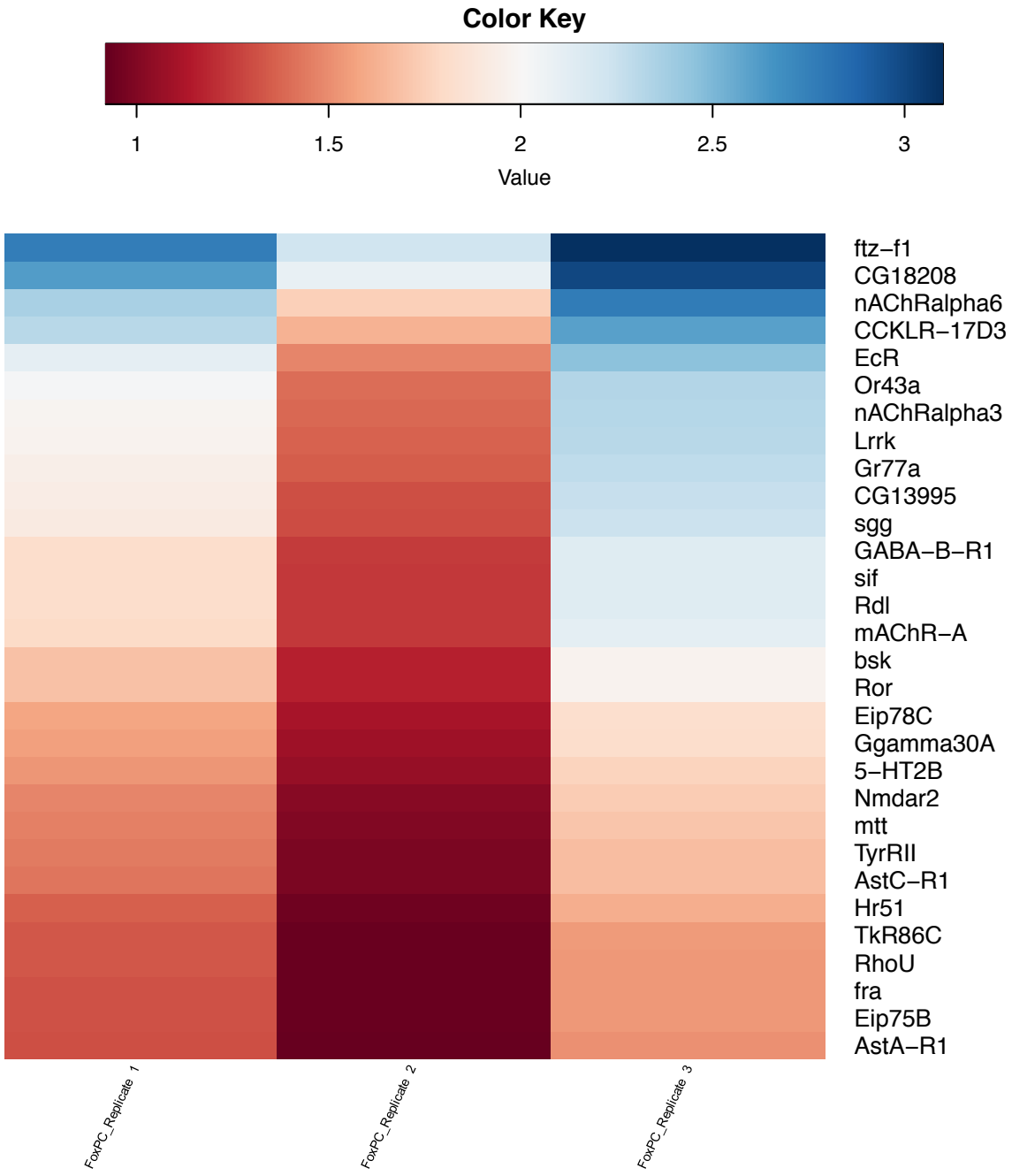
To get an idea of which genes are involved in the different biological processes, the gene lists from different replicates were classified into the most significant GO terms, for example nervous system development, neuron development, signal transduction and biological regulation (Figure 20).



GO term: Nervous System Development



GO term: Nervous Neurogenesis



GO term: Signal Transduction Activity

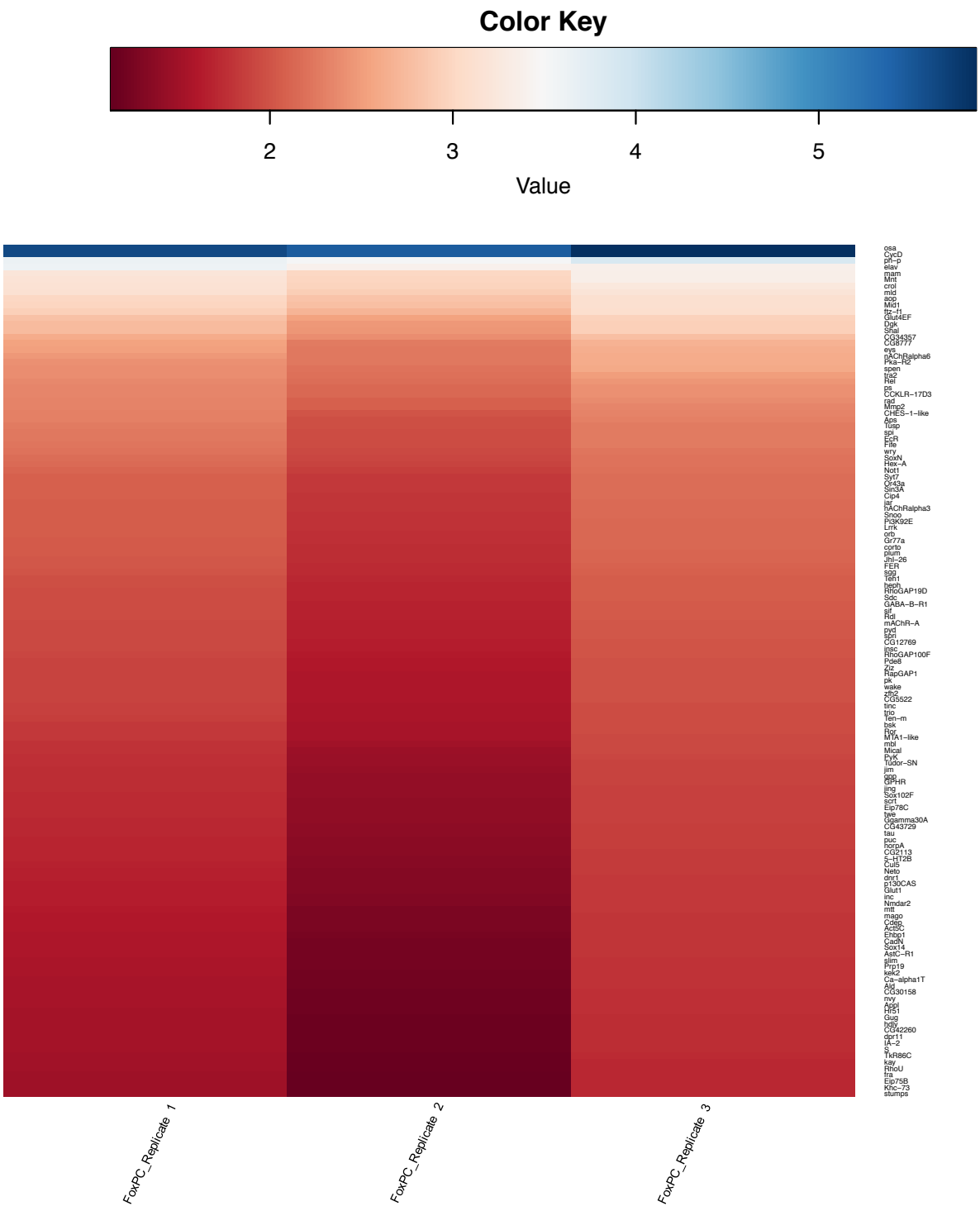


Figure 20. Heatmaps of the genes contributing to the most overrepresented GO categories.

All the genes contributing to nervous system development, neurogenesis, signal transduction activity and biological regulation are shown for the different FoxPC replicates. The color key represents the log2fold change from the DamID-seq analysis.

Previous work in the lab, to which I contributed, found that FoxP, more specifically FoxPC, controls the abundance of the voltage-gated potassium channel, Shal. (Groschner et al., 2018). This in turn controls the excitability and determines the integrative properties of neurons and hence decision times. A large-scale analysis of other ion channels (voltage-gated, two-pore potassium, chloride and calcium channels) did not reveal any other putative target of FoxPC. This was verified by mapping the DamID-seq target list to a list of all the ion channels obtained from FlyBase. Only 8 ion channel genes seem to be potential FoxPC targets (Figure 21). Reassuringly, Shal is the second most significant ion channel target of FoxPC, in accordance with the data from Groschner et al.,.

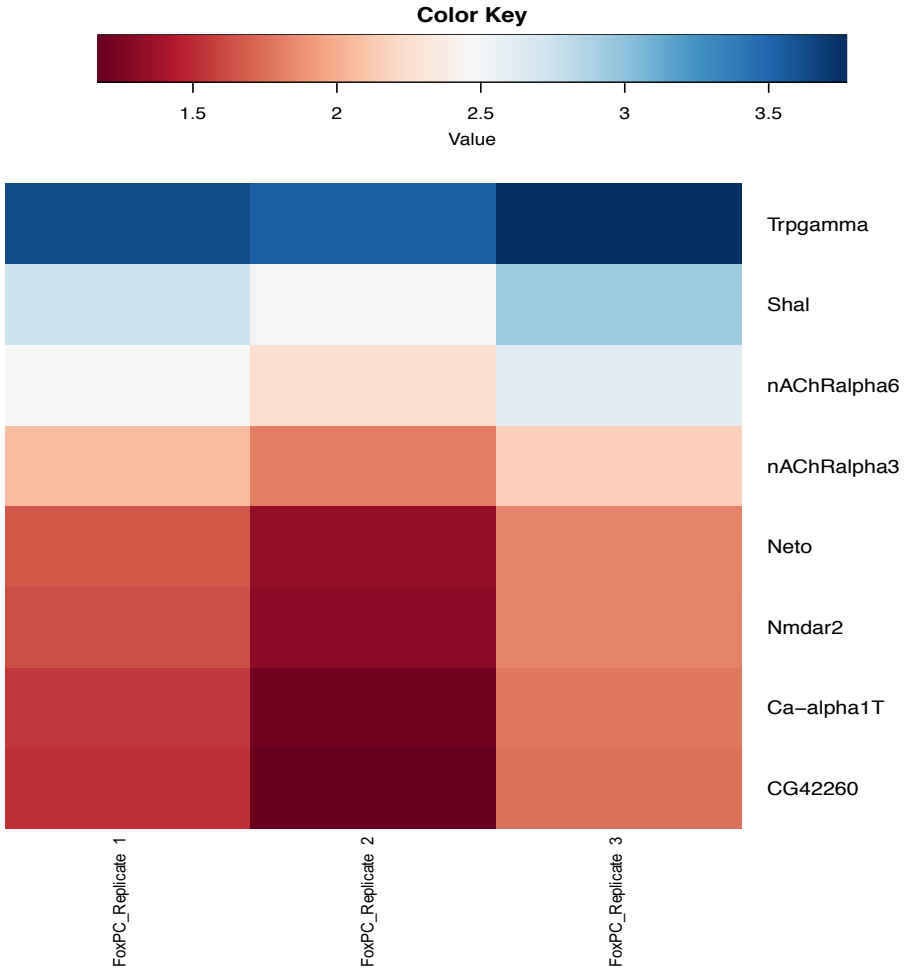


Figure 21. FoxPC binds only 8 ion channel target genes.

The DamID-seq dataset was compared to a list of all the ion channels in Drosophila from FlyBase. Only 8 ion channels are potentially regulated by FoxPC and are ranked by their FoxPC occupancy in the heatmap.

3.2.7 Pathway analysis

A network analysis of the genes associated with the most enriched GO term ("nervous system development") and which contain the putative FoxPC motif was then built using GeneMania, a Cytoscape plugin that stores interaction information from different sources. This allows the identification of active pathways and possibly the connection of molecular information (transcriptome) with the phenotype in study, in this case the decision making phenotype associated with FoxPC (Papatheodorou et al., 2015).

In Figure 22, the network is composed of nodes (representing the genes) and edges (which show physical interactions between the nodes). The edges of a node represent the degree of the node, the edge width and colour represents 'edge betweenness' which is the number of shortest paths through a certain edge. 'Edge betweenness' therefore is the degree to which an edge makes other connections possible. The node size represents 'betweenness centrality' which reflects the control that the node exerts over interactions of other nodes in the network. 'Betweenness centrality' quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.

The network analysis produced a network of 43 connected components. The network with the most connected components was extracted. 32 of these components are putative direct FoxPC targets (as determined by DamID-seq analysis), with the majority of the genes being involved in axogenesis and neuron development. The nodes with the highest degrees (most edges) are Ten-m and aop. Teneurin-m (Ten-m) has well established roles in neural development (Mosca et al., 2012, Hattori et al., 2013). Anterior open (aop) is a transcription factor which acts as a transcriptional repressor and is critical for the development of the nervous system (Zhu et al., 2011b). The network analysis suggests that

FoxPC plays critical roles in the *Drosophila* brain by regulating the expression of genes involved in development of the nervous system.

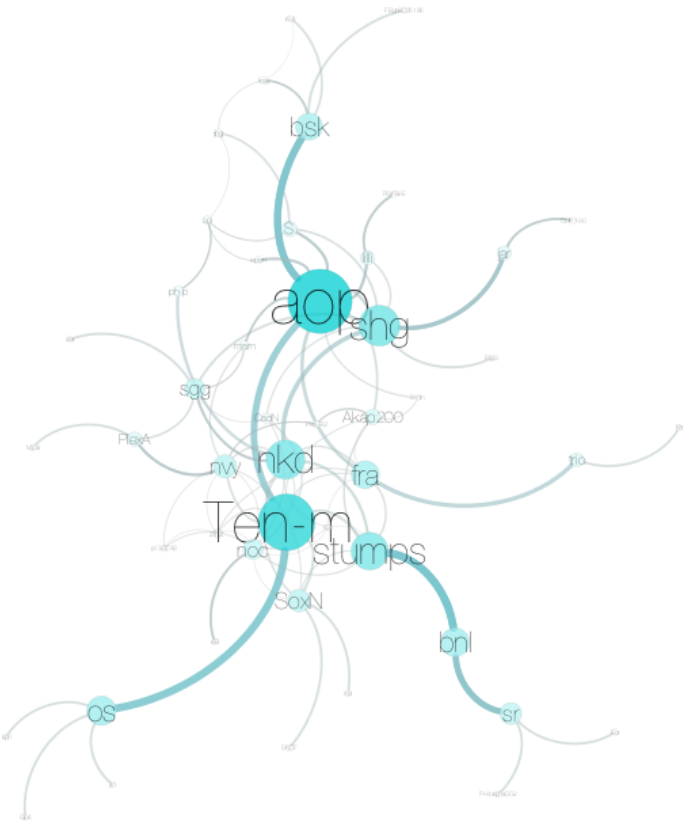


Figure 22. FoxPC regulates a large number of genes involved in nervous system development.
A network analysis of the top target genes associated with the "nervous system development" GO term was created using Cytoscape. The largest connection is shown. The size of the nodes and thickness and colour of the edges represent centrality and edge betweenness, respectively. The nodes labelled in blue are FoxPC target genes.

3.3 Interim Discussion from Chapter 3

In this chapter, putative targets of FoxPC and FoxPD were identified by targeted DamID. This method was chosen over ChIP-seq as no high quality FoxP antibody is commercially available and attempts to produce such an antibody in-house were unsuccessful.

FoxPC and FoxPD were fused to the DNA adenine methyltransferase (Dam) and their binding sites identified by deep sequencing. The experiments are robust, as can be inferred from their reproducibility between biological replicates (pairwise Pearson correlation > 80%), their significantly different binding profiles from the Dam-control binding and the successful identification of known targets. The motifs identified for each transcription factor also match inferred and predicted binding motifs from Cis-BP and Nitta et al. (Nitta et al., 2015).

FoxPC and FoxPD have some common targets, but mostly regulate different genes, which can be explained by their different DNA binding domains (Santos et al., 2011). Most of the putative targets are involved in nervous system development as determined by the GO analysis, which confirms the crucial roles of FoxP in the developing nervous system. The human homologue FoxP2 has been involved in neurodevelopmental disorders, such as autism spectrum disorder (Feuk et al., 2006). The identification of the neural transcriptional network, including its component genes for example Aop, Ten-m and FoxP in *Drosophila* may allow for a better understanding of these disorders in humans.

The second most significant GO term is “signaling”, which incorporates different signaling pathways for example retinoic acid pathway, ion channel signaling pathways and signal transduction. This is consistent with studies that implicate FoxP2 in signaling pathways to affect neuronal differentiation (Devanna et al., 2014).

Some of the targets identified in this study have been identified elsewhere using ChIP-chip screens for FoxP2 binding, suggesting that the targets and pathways regulated by FoxP are conserved during evolution. One of the targets identified in the DamID-seq is the voltage-gated channel, Shal. This channel was reported to be a direct target and is negatively regulated by FoxPC in an independent analysis I performed: translating ribosome affinity

purification (TRAP) (Groschner et al., 2018).

DamID-seq is uniquely capable of mapping the binding profiles of mutant proteins. Mutating the amino acids important for DNA binding (as inferred by FoxP2 mutations (Santos et al., 2011), carrying out DamID-seq and comparing it to the datasets from the wild-type FoxPC and FoxPD will therefore be useful for confirming the targets of FoxPC and FoxPD, for identifying aberrant patterns and for further understanding how mutating these proteins affect neural processes, for example the integration of sensory information (DasGupta et al., 2014, Groschner et al., 2018).

4

Results

Confirming FoxPC targets identified by DNA Adenine Methyltransferase Identification

In the previous chapter, the targets of FoxPC and FoxPD were identified using DamID, coupled with deep sequencing. As with most bioinformatics analysis, it is crucial to choose the appropriate sequencing platform, library preparation, sequencing pipeline, analysis software, parameters and statistical analysis. These vary from analysis to analysis and even the most optimal approach will likely have some limitations. Even though the computational analyses done on DamID-sequencing datasets seem robust, it is necessary to validate the targets experimentally using complementary methods. Since the main goal of this thesis is to gain insight into how FoxPC regulates the decision-making process in *Drosophila* (DasGupta et al., 2014, Groschner et al., 2018) by identifying its targets and its regulatory capabilities, this chapter will focus on the FoxPC target genes. The three methods that were used to confirm putative targets of FoxPC are chromatin immunoprecipitation (ChIP), translating ribosome affinity purification (TRAP), and dual luciferase reporter assays. Electrophoretic mobility shift assays (EMSA) were then carried out to confirm binding of FoxPC to its binding motif found in the target genes.

Chromatin immunoprecipitation (ChIP) was described in the previous chapter. In brief, ChIP allows the identification of physical interaction between a protein of interest and

genomic loci (Kaufmann et al., 2010a, Pose and Yant, 2016). ChIP relies on the crosslinking of proteins to their target loci *in vivo*, the isolation and shearing of chromatin, the immunoprecipitation of DNA-protein complexes by a highly specific antibody and the release of fragments of bound DNA. Coupled with qPCR (ChIP-qPCR), hybrisidation to arrays (ChIP-Chip) or next generation sequencing (ChIP-seq), the fragments can be purified and identified accurately. ChIP and DamID have been shown to produce similar transcription factor association profiles (Greil et al., 2006, Moorman et al., 2006). ChIP-qPCR is therefore an ideal technique to verify the putative targets of FoxPC obtained from DamID-seq.

Translating ribosome affinity purification (TRAP) is a method that allows the transcriptomic analysis of a genetically defined population of cells (Figure 23). It is a method that uses both cell type specific targeting of a transgene, enhanced GFP-tagged ribosomal protein L10A (RpL10A), and indirect targeting of mRNA (Heiman et al., 2014). In *Drosophila*, the GAL4/UAS system can be exploited to restrict the expression of this transgene (Thomas et al., 2012). The fusion protein incorporates efficiently into ribosomes and polysomes. Affinity purification of the GFP-tagged polysomes and release of the translated mRNAs from the ribosomes provide an enriched pool of mRNAs, which can be used to determine the gene expression profile of specific cells by quantitative PCR (qPCR) or RNA sequencing. TRAP is a highly sensitive method, as it does not require dissociation or fixation of tissues. However, the protein-coding mRNA associated with ribosomes constitutes a very small fraction (5%) of total cellular RNA, with ribosomal RNA (rRNA) making up about 80% of the total RNA. Therefore, most of the purified RNA will be rRNA and only a small amount will be mRNA. Coupling TRAP with amplification of the

complementary DNA by qPCR allows the analysis of the small amount of mRNA isolated.

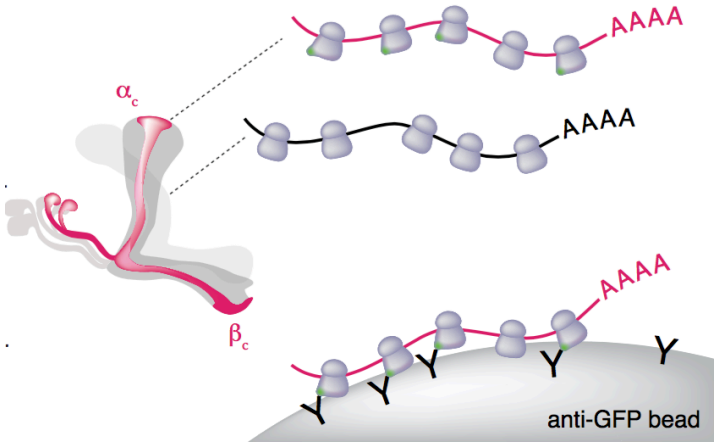


Figure 23. Schematic of the TRAP protocol.
eGFP labelled ribosomes and translating mRNAs are immunoprecipitated using anti-GFP beads from specific neurons in the fly brain.

Luciferase reporter assays are sensitive assays which allow the high-throughput and rapid analysis of transcription factor-DNA binding. In addition to validating the binding of FoxPC to the targets, luciferase assays will provide a more functional and quantitative aspect of the interaction (Solberg and Krauss, 2013). Reporter assays rely on measurable properties of genetic reporters, for example bioluminescence from luciferases. Luciferases are cloned downstream of a DNA sequence of interest, for example a promoter region, in the absence of other eukaryotic promoter elements and are transfected into cells. The ability of the promoter region to drive expression of the luciferase is determined by directly measuring the enzymatic activity of the luciferases. Comparison of the luciferase activity in an experimental vector (containing a promoter) with that without a promoter (empty vector) gives an indication of the activity of the promoter. The activity of the promoter can then be manipulated using transcription factors which bind to and regulate the promoters. In this way, reporter assays enable the accurate quantification of small

changes in transcription resulting from transcription factor regulation rather than simply a binary evaluation of expression.

Dual luciferase reporter assays use two reporters, firefly and Renilla luciferases, and provide more accurate and reproducible results (Figure 24). The firefly luciferase is found downstream of the promoter of interest and is used as the experimental reporter. The Renilla luciferase is coupled to a constitutive promoter whose activity is unaffected by experimental conditions and thus works as an internal control reporter. Normalisation to the Renilla luciferase activity minimises variabilities that undermine experimental accuracy, for example differences in the number of cells and the efficiencies of cell transfection and lysis.

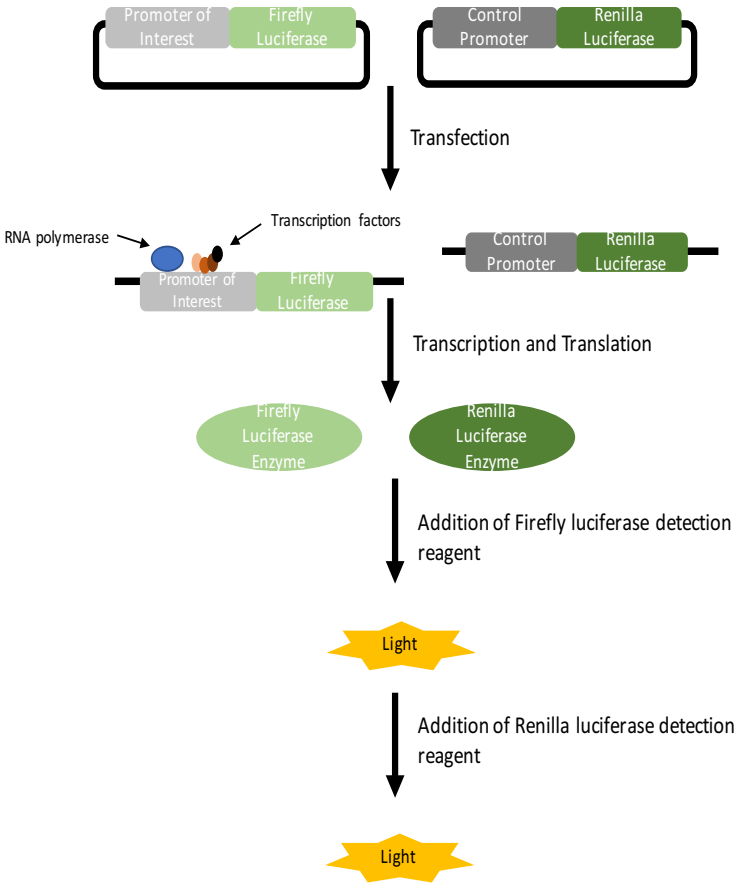


Figure 24. Dual luciferase reporter assay schematic.

The dual luciferase assay uses the firefly and the Renilla luciferases, with the former being the experimental reporter which indicates the activity of the promoter, while the latter is used as a control reporter. They are both transfected into cells, together with the transcription factors of interest. The activities of the firefly and Renilla luciferases are measured sequentially using a luminometer.

Electrophoretic mobility shift assays (EMSA) do not only detect transcription factor-DNA interactions in vitro with high sensitivity, they also allow the confirmation of consensus sequences (Fried, 1989). They exploit the differences in electrophoretic mobility (or the rate of migration) of free nucleic acids and protein-nucleic acid complexes. Labeled target DNA probes, which contain the consensus binding sequences, are incubated with transcription factors to allow any complexes to form. The complexes are resolved from free DNA by electrophoresis. The free and bound labelled DNA can then be visualised using autoradiography, fluorescence dyes or chemiluminescence.

4.1 Aims and objectives

A list of 535 genes was obtained from the DamID-seq analysis of FoxPC targets. Since testing all the candidate targets is impossible, the following strategy was used to prioritise which targets were were taken forward for further analysis:

1) Targets with the highest degrees of FoxPC occupancy (determined by the $\log_2$ fold change obtained from the DamID-seq analysis)
2) Targets with functional roles in the most over-represented GO terms (nervous system development, signaling and biological regulation)
3) Targets containing the putative FoxPC binding motif (determined in the previous chapter by MEME).

Once suitable candidates for testing have been selected, chromatin immunoprecipitation, translating ribosome affinity purification and luciferase assays were carried out to validate the DamID-seq targets and to understand their regulation by FoxPC. These three methods were chosen to confirm the reliability of the DamID technique, the pipeline used to

analyse the sequencing data and therefore to confidently establish FoxPC targets. The functionality of FoxPC binding motif (A(A/G)CAACAA) was confirmed in electrophoretic mobility shift assays.

4.2 Results

The 2-3 most significant genes from each of the most overrepresented GO categories (nervous system development, signaling and biological regulation) are *osa, ph-p, mam, ftz-f1, nAChRα6, cycD, Mnt, crol, Trp γ* and *Shal*. Figure 25 shows that these genes are some of the most highly FoxPC-bound genes: the y-axis represents the $\log_2$ fold change (from -6 to 6), which was obtained from the DamID-seq analysis in the previous chapter and the gene loci are labelled at the bottom of each graph.

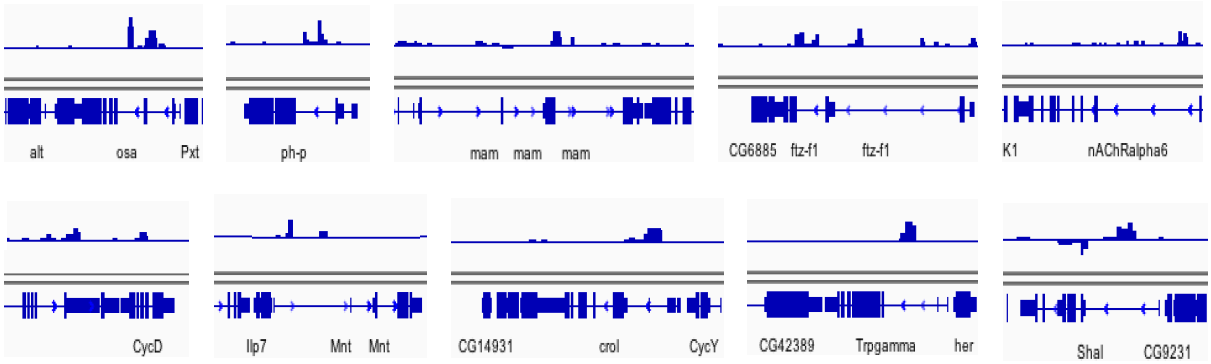


Figure 25. Binding intervals at several FoxPC target loci.

The binding of FoxPC at the most enriched genes in each category was visualised using IGV. The y-axis shows the $\log_2$ fold ratio of the peaks at each locus.

The 10 candidate genes were analysed for the presence of the FoxPC motif determined by the MEME software. This was done using default settings in MAST (Motif Alignment & Search Tool) (http://www.sdsc.edu/MEM) (Bailey et al., 2009), which is part of the MEME Suite. MAST searches a list of sequences (target sequences extracted from the

most significant peaks) for matches to a given motif, in this case the putative FoxPC motif (A(A/G)CAACAA). Since most of the binding sites of FoxPC were found to be either upstream or slightly downstream of the transcription start site (Figure 13), sequences ± 1000bp of the transcription start sites were used as input. Only sequences with high match scores (< 10) were retrieved. The locations of the different motifs with respect to the most significant target genes are shown below (Figure 26 and Table 2).

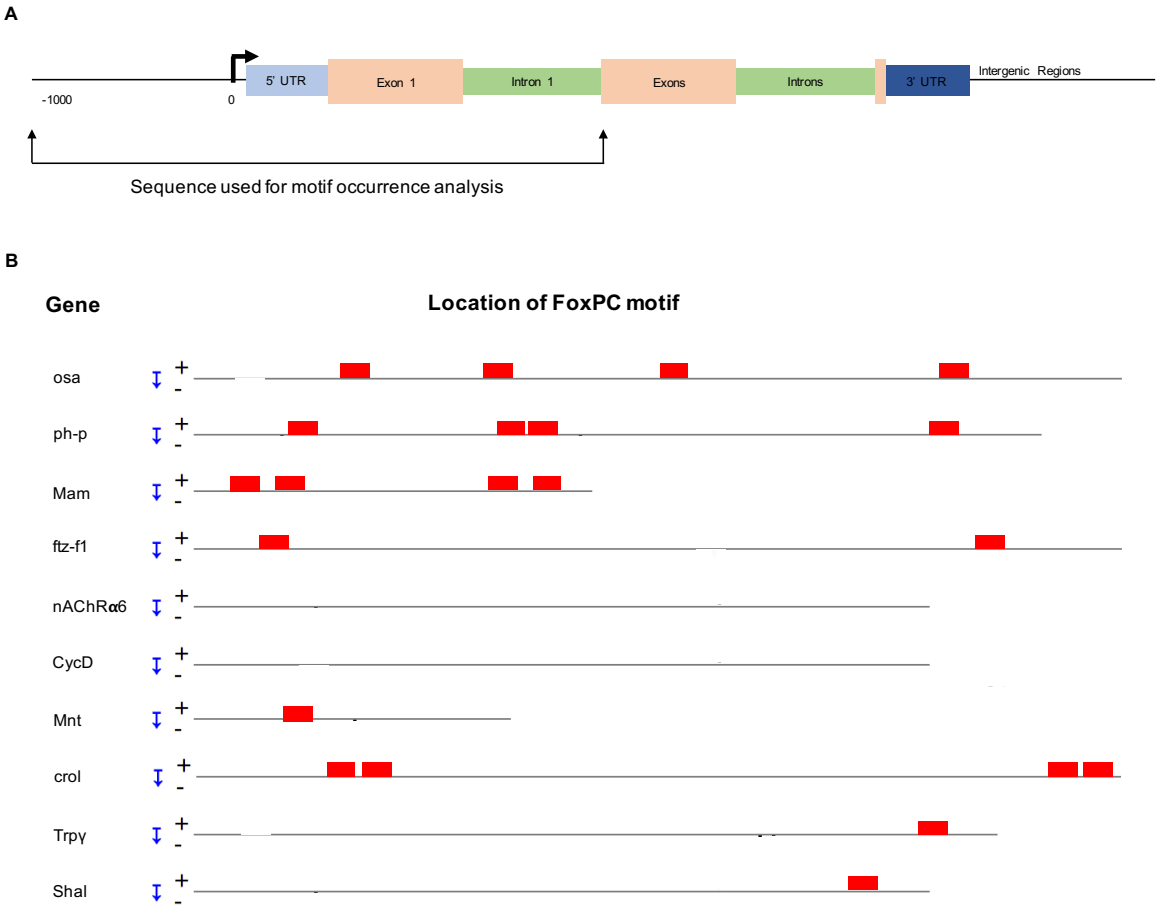


Figure 26. MAST output for the locations of the FoxPC putative motif.

*The occurrence and locations of the FoxPC putative motif was determined using MAST for the most significant FoxPC targets. The sequence analysed for each gene includes the sequence 1000 bp upstream of the start site, the first exon and the first intron (**A**). The putative motif was found in most of the query sequences, with the exception of nAChRα6 and CycD (**B**).*

Table 2. Summary of the genes tested for their FoxPC occupancy by DamID-seq and the presence of the FoxPC putative motif.

Gene Name	GO Term(s)	Log2 fold change	Presence of motif
osa	Nervous System Development and Biological Regulation	5.62	Yes
ph-p	Nervous System Development and Biological Regulation	4.0	Yes
Mam	Nervous System Development and Biological Regulation	2.98	Yes
ftz-fl	Nervous System Development and Signal Transduction	2.4	Yes
nAChRα6	Signal Transduction	2.36	No
cycD	Biological Regulation	3.8	No
Mnt	Biological Regulation	2.88	Yes
crol	Biological Regulation	2.81	Yes
Trpγ	Ion Channel Activity	3.28	Yes
Shal	Ion Channel Activity and Biological Regulation	2.56	Yes

4.2.1 Analysis of FoxPC binding to promoter regions

Chromatin immunoprecipitation was then used as a complementary method to validate the results from DamID-seq. Since high-quality FoxPC antibodies are not available, FoxPC was fused to the HA tag, for which highly sensitive antibodies exist. HA-tagged FoxPC vectors were first transfected into ML-BG3-C2 neuronal cells. An empty HA vector was also included as a control to account for artefacts and background from overexpression of the tg. Immunofluorescence of cells transfected with *HA-tagged FoxPC* revealed that the HA-tagged FoxPC localise to the nucleus (Figure 27), suggesting that the localisation and function of the FoxPC are not affected by introducing the HA tag.

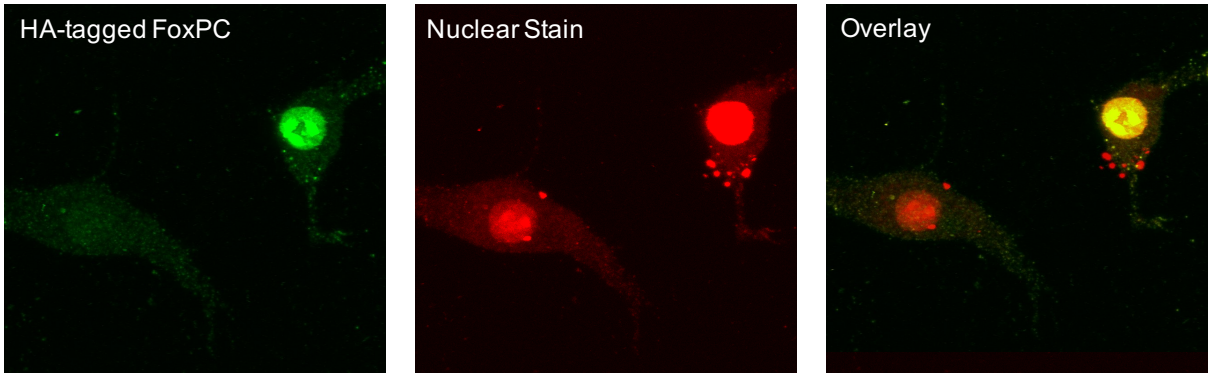


Figure 27. Tagging FoxPC with an HA tag does not affect its cellular location.
Cells transfected with an HA-FoxPC were stained with an anti-HA antibody (left panel) and a nuclear stain (propidium iodide) (middle panel). The overlay is shown in the right panel.

After transfection, HA-tagged FoxPC was crosslinked to its target DNA sequences and purified from cell lysates using a ChIP-grade anti HA antibody. To account for non-specific binding by the antibody, an isotype antibody control was used. Binding of FoxPC at the different loci in HA-FoxPC-transfected cells was always compared to the isotype control and to the HA-only control and normalised to the input DNA (sample before immunoprecipitation). The target DNA sequences were then amplified using primers specific to the promoter regions (regions ± 1000bp from transcription start sites) of candidate genes. Several primers were tested, each spanning 100 bp regions (Figure 28B), which is the optimal amplicon length for qPCR. Only the results from primer sets resulting in an enriched fold change compared to the controls are shown in Figure 28C. The fold change was calculated by normalising the cycle threshold (C_t) values obtained from the immunoprecipitated samples to those obtained from the input samples under experimental and control conditions.

ChIP revealed significant FoxPC binding at most of the promoter regions tested (enriched fold change > 1) when normalised against the inputs and compared to both the isotype and the HA-only controls. Consistent with the DamID-seq data, the *osa* promoter showed the

highest FoxPC binding (enriched fold change ~ 9), validating both approaches. As a control, the effect of FoxPC on the *Shaker* locus was also determined. Shaker was not identified as a FoxPC target from the DamID-seq, does not contain the putative FoxPC binding motif and is not regulated by FoxPC (Groschner et al., 2018). The ChIP results show that FoxPC does not bind to the Shaker promoter regions (enriched fold change < 1), suggesting that the DamID-seq and MAST results are valid.

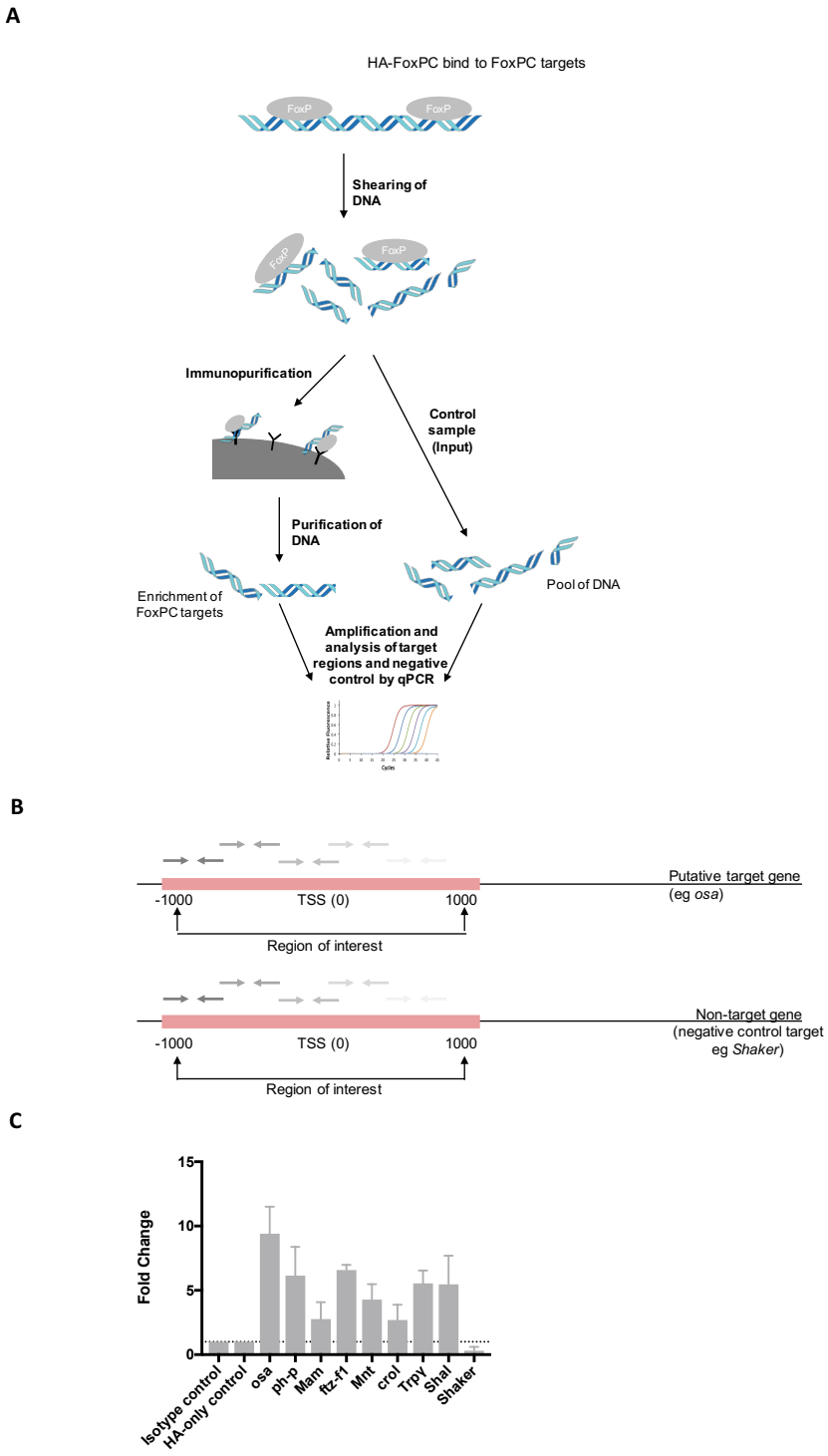


Figure 28. Chromatin immunoprecipitation in ML-BG3-C2 neuronal cells.

*ChIP was carried out in neuronal cells transfected with HA-tagged FoxPC (A). The binding of FoxPC to known target regions (osa, ph-p, Mam, ftz-f1, Mnt, crol, Trpγ and Shal) and known non-target gene (Shaker) was quantified by qPCR using sets of primers spanning ± 1000 bp from the TSS (**B**). The binding of FoxPC to the osa, ph-p, Mam, ftz-f1, Mnt, crol, Trpγ, Shal and Shaker promoters in immunoprecipitated samples is normalised to the inputs (samples before immunoprecipitation) and compared to the isotype and HA-only controls. FoxPC binds to all the promoters tested, except the Shaker promoter (**C**). Bars represent the means, and the error bars the s.e.m. Values represent results from three biological replicates.*

To complement these analyses in transfected cells with studies *in vivo*, where the presence of cofactors might modulate the DNA binding of FoxP, ChIP experiments were also carried out in transgenic flies expressing either HA-tagged FoxPC or HA-only control. Again, isotype controls were included in the experiments, which were performed in triplicates. Optimisation of the sonication protocol, by varying the time course of sonication and the number of cycles, was required for efficient and homogenous shearing of chromatin (Figure 29A). Sonication protocol 2 (30 seconds on, 30 seconds off for 10 cycles) yielded appropriately sized fragments for quantification using qPCR. Comparable results were seen in both ChIP experiments *in vitro* in neuronal cells and *in vivo* in *Drosophila* brains. All of the promoter regions showed binding by FoxPC when normalised to the input sample and compared to the controls. Once more, the *Shaker* promoter did not show any FoxPC binding (Figure 29B).

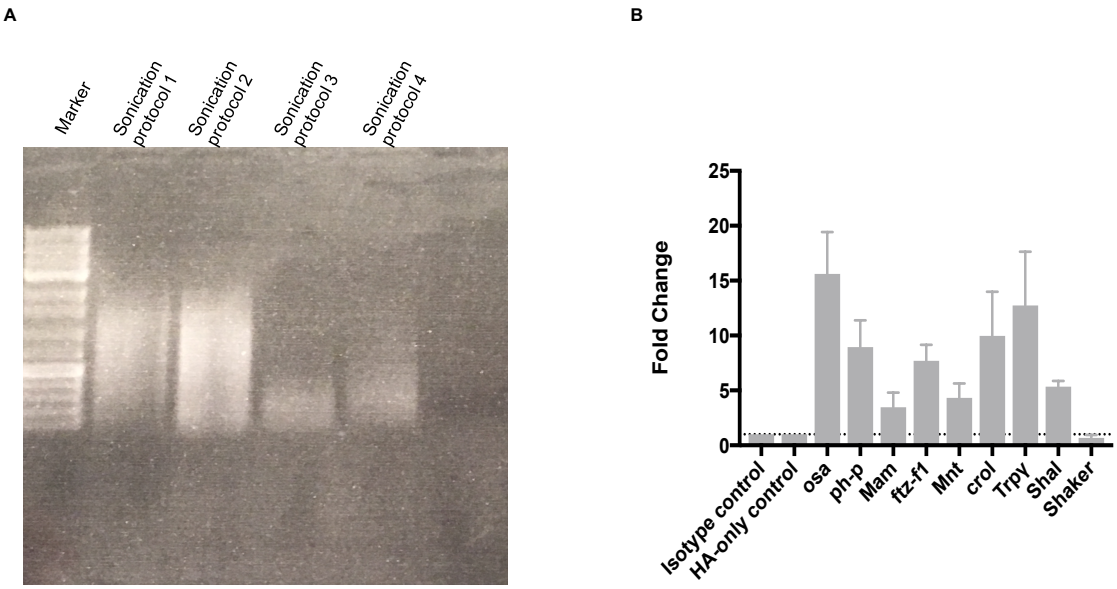


Figure 29. Chromatin Immunoprecipitation in Drosophila brains.

HA-tagged FoxPC or HA-only control constructs were injected in flies and expressed in $\alpha\beta_c$ neurons using NP7175-GAL4. ***A****, Several sonication protocols were tested (Protocol 1- 30seconds on, 30 seconds off for 5 cycles, Protocol 2- 30 seconds on, 30 seconds off for 10 cycles, Protocol 3- 20 seconds on, 30 seconds off for 5 cycles and Protocol 4- 20 seconds on, 30 seconds off for 10 cycles). Successful shearing of chromatin only occurred under specific conditions (Protocols 1 and 2) DNA fragments ranging from 100 bp to 2.5 kb were seen on a 1% agarose gel.* ***B****, Binding of FoxPC to the promoter regions of osa, ph-p, Mam, ftz-f1, Mnt, crol, Trpγ, Shal and Shaker from the immunoprecipitated samples were quantified using qPCR with gene-specific primers. Bars represent the mean value from three biological replicates and errors bars represent s.e.m.*

4.2.2 Analysis of the regulation of targets by FoxPC by chromatin immunoprecipitation

For DNA binding to be physiologically relevant, FoxPC and its target genes need to be expressed in the same neurons. The expression of the 8 genes that constitute validated FoxPC binding targets was therefore tested in FoxPC-positive neurons of the mushroom body (Groschner et al., 2018, Schatton and Scharff, 2017). Translating ribosome affinity purification (TRAP) was used to isolate actively translated mRNAs from two subsets of neurons of the mushroom body, the $\alpha\beta_c$ and α'β' KCs, using the *NP7175-GAL4* and *VT030604-GAL4* lines, respectively (Figure 30A and 30B). The expression levels of

FoxPC were then quantified by qPCR and compared between the two subsets. To account for the variations in RNA input, the efficiency of reverse transcription of mRNA to cDNA and the PCR itself (Kozera and Rapacz, 2013), the expression levels are normalised against three reference genes (*Gpdh, Tbp* and *Efα100E*), which were chosen because they are expressed at stable, non-regulated levels.

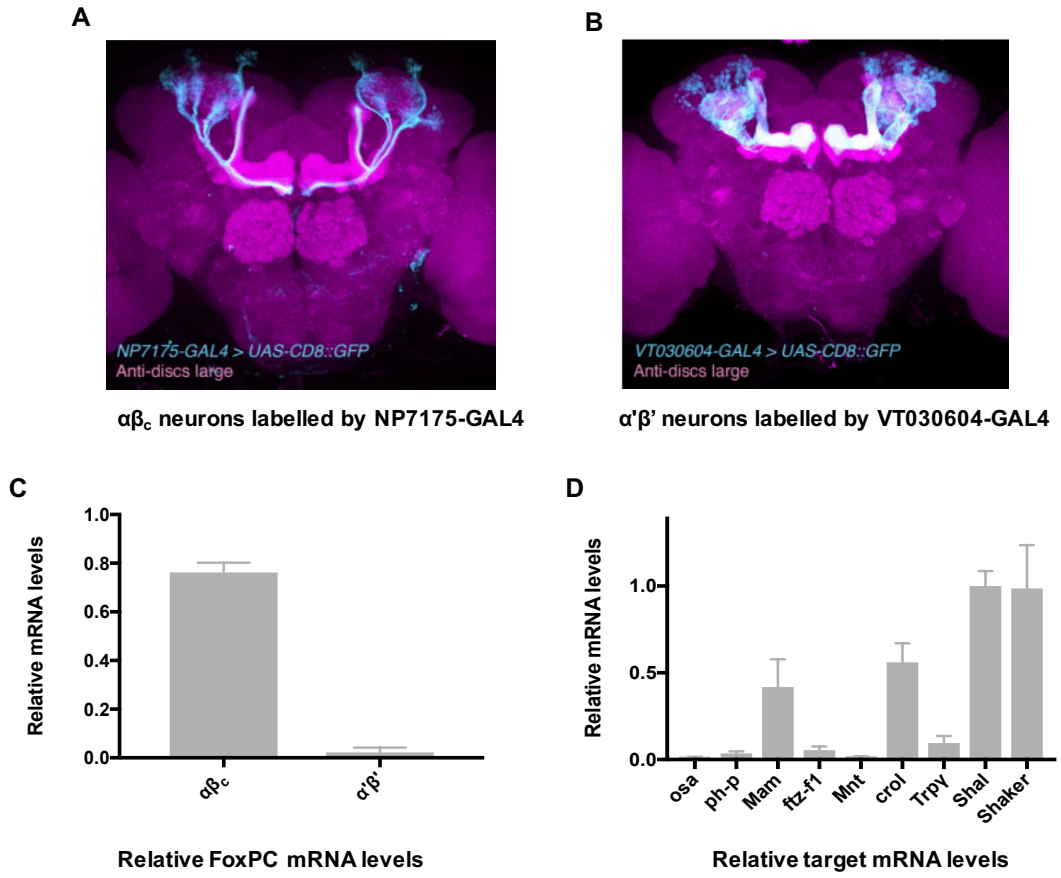


Figure 30. Expression of FoxPC and its targets in the αβc neurons.

*Expression patterns of the two GAL4 lines used for TRAP. Maximum intensity projections of confocal image stacks of the central brains of flies expressing UAS-CD8::GFP (cyan) under the control of NP7175-GAL4 (**A**) and VT030604-GAL4 (**B**). Synaptic structures were counterstained with an antibody against discs large (magenta). **C**, Quantification of FoxPC mRNA in neurons labelled by each GAL4 line using TRAP and qPCR. Values indicate results from 10 replicates. **D**, Levels of polysome-bound transcripts in αβc neurons relative to the geometric mean of three reference gene transcripts. The bars represent the means and the error bars the s.e.m.*

TRAP coupled to qPCR showed an abundance of *FoxPC* mRNA in $\alpha\beta_c$ neurons but little, if any, in α'β' neurons (Figure 30C), meaning that the levels of FoxPC vary between different subsets of KCs. This suggests that $\alpha\beta_c$ KCs are a principal site of FoxPC's action in the mushroom body, confirming an earlier indication by Das Gupta et al., (DasGupta et al., 2014). To be able to examine the effects of FoxPC on its target genes in the $\alpha\beta_c$ neurons, it was necessary to first ensure that they are expressed in those neurons. The levels of the target genes in the $\alpha\beta_c$ neurons were measured using TRAP, coupled to qPCR using gene-specific primers. *Mam, crol, Shal* and the non-target *Shaker* are expressed at relatively high levels (compared to the geometric mean of the three reference genes), while *osa, ph-p, ftz-f1* and *Mnt* mRNAs are detected only in trace amounts (Figure 30D).

The availability of FoxP-mutant ($FoxP^{5\text{-}SZ\text{-}3955}$) flies and flies expressing hairpin RNA directed against the *FoxP* transcript ($FoxP^{RNAi}$) was exploited in *NP7175-GAL4*-restricted TRAP experiments to determine whether FoxPC activates or represses its putative targets. After confirming the loss of *FoxPC* transcript from $\alpha\beta_c$ neurons in homozygous $FoxP^{5\text{-}SZ\text{-}3955}$ mutant and FoxP-knockdown flies (Figure 31A), target gene expression levels were measured. Combined results from at least 4 biological replicates are shown in Figure 31B. The mRNA levels of all the target genes increase in the absence of FoxPC. In some cases, for example, *osa, ph-p, ftz-f1, Trpγ* and *Shal*, the expression levels increase drastically (> 10 fold) when FoxPC levels are depleted. Since the DamID-seq data revealed the highest $\log_2$ ratios for those genes, it is expected that they show the highest FoxPC occupancy and are hence strongly repressed by FoxPC, in some cases (*osa*, *ph-p* and *ftz-f1*) to an extent that the corresponding transcripts are virtually absent from the $\alpha\beta_c$ neurons of wild-type flies.

These results infer that binding of FoxPC to its putative target genes is functionally

relevant and demonstrate that FoxPC negatively regulates, or represses, their transcription. As mentioned previously, FoxPC does not bind to *Shaker* and differences in the level of FoxPC are therefore not expected to have an effect on its expression levels. Indeed, the mRNA levels of *Shaker* remained unchanged when FoxPC levels were decreased or knocked down. This proves the robustness of both the DamID and TRAP experiments

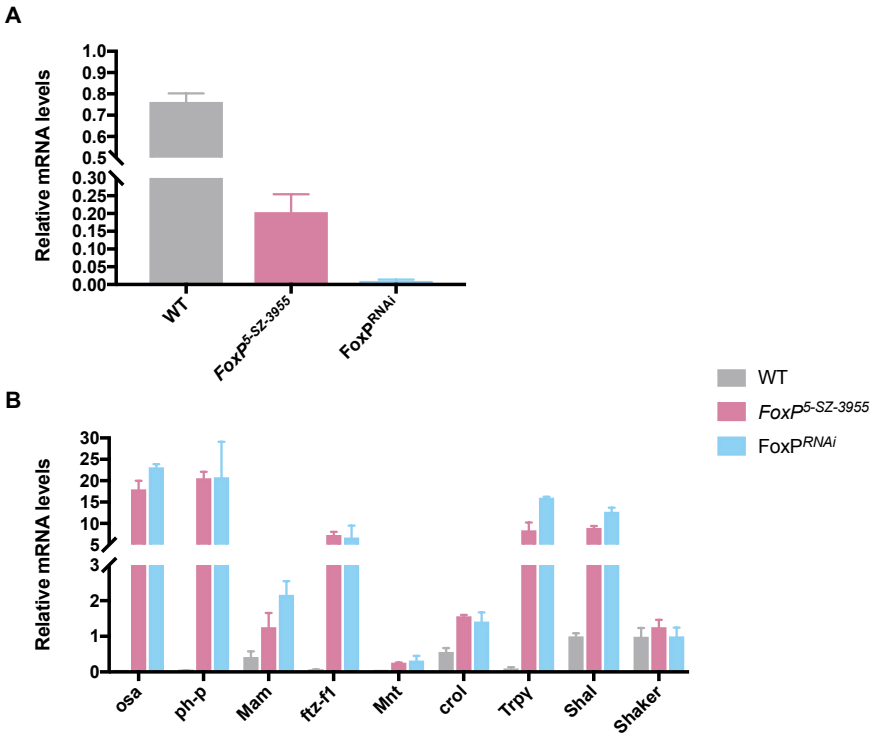


Figure 31. Relative transcript levels in the αβc neurons.

*TRAP was used to isolate mRNAs from the αβc neurons (using the NP7175-GAL4 driver line) in wild-type (WT), FoxP$^{5\text{-}SZ\text{-}3955}$ and FoxP^{RNAi} flies. After reverse transcription, the resulting cDNA was quantified by qPCR using primers specific to the genes of interest. (**A**) Expression levels of FoxPC transcript are highly decreased in the FoxP-mutant (pink bars) and FoxP^{RNAi} flies (blue bars). (**B**) Expression levels of the FoxPC putative targets (osa, ph-p, Mam, ftz-f1, Mnt, crol, Trpγ and Shal) are increased in FoxP$^{5\text{-}SZ\text{-}3955}$ and FoxP^{RNAi} flies compared to wild-type flies.*

4.2.3 Analysis of the regulation of targets by FoxPC in luciferase reporter assays

Another complementary method which was used to analyse the functional consequences of transcription factor-promoter DNA binding was dual luciferase reporter assays. Since the promoter regions for most genes have not been annotated, the 1000 base pairs upstream and downstream of the transcription start sites were used. Promoter regions of the 8 target genes were cloned downstream of a minimal promoter (*Act5C* promoter) and upstream of a sequence encoding firefly luciferase. The promoter drives the expression of the firefly luciferase, such that the luminescence emitted by the luciferase is proportional to the activity of the promoter. The renilla luciferase, which uses a different substrate and causes the emission of light at a different wavelength, is also transfected as a control to eliminate variations in cell numbers, transfection and lysis efficiencies. Transfecting the luciferases in cells overexpressing FoxPC will allow the efficiency of binding and regulation between FoxPC and the different promoter regions to be assessed. Transfections with the promoters and the luciferases, but without FoxPC, reflect background luminescence.

FoxPC expression repressed activity of the firefly luciferase reporters driven by *osa, ph-p. Mam, ftz-f1, Mnt, crol, Trpγ* and *Shal* promoters (Figure 32). The decrease in luciferase signal driven by the different promoters is highly significant ($p<0.005$), suggesting that FoxPC binds strongly to these promoters and regulates their activity (Figure 32A-H). No significant change was seen when the *Shaker* promoter was transfected into cells overexpressing FoxPC. This suggests that FoxPC does not affect the activity of the *Shaker* promoter region (Figure 32I). The repression of the promoter activity is not complete, which suggests that FoxPC might work in concert with other transcription factors or that other distal regulatory elements, which are found further from the transcription start sites,

are required. The different extent of repression could be explained by the different binding frequency of FoxPC to the different promoter regions. Different transcription factor complexes, made up of transcription factors having different repressive activities, might be required.

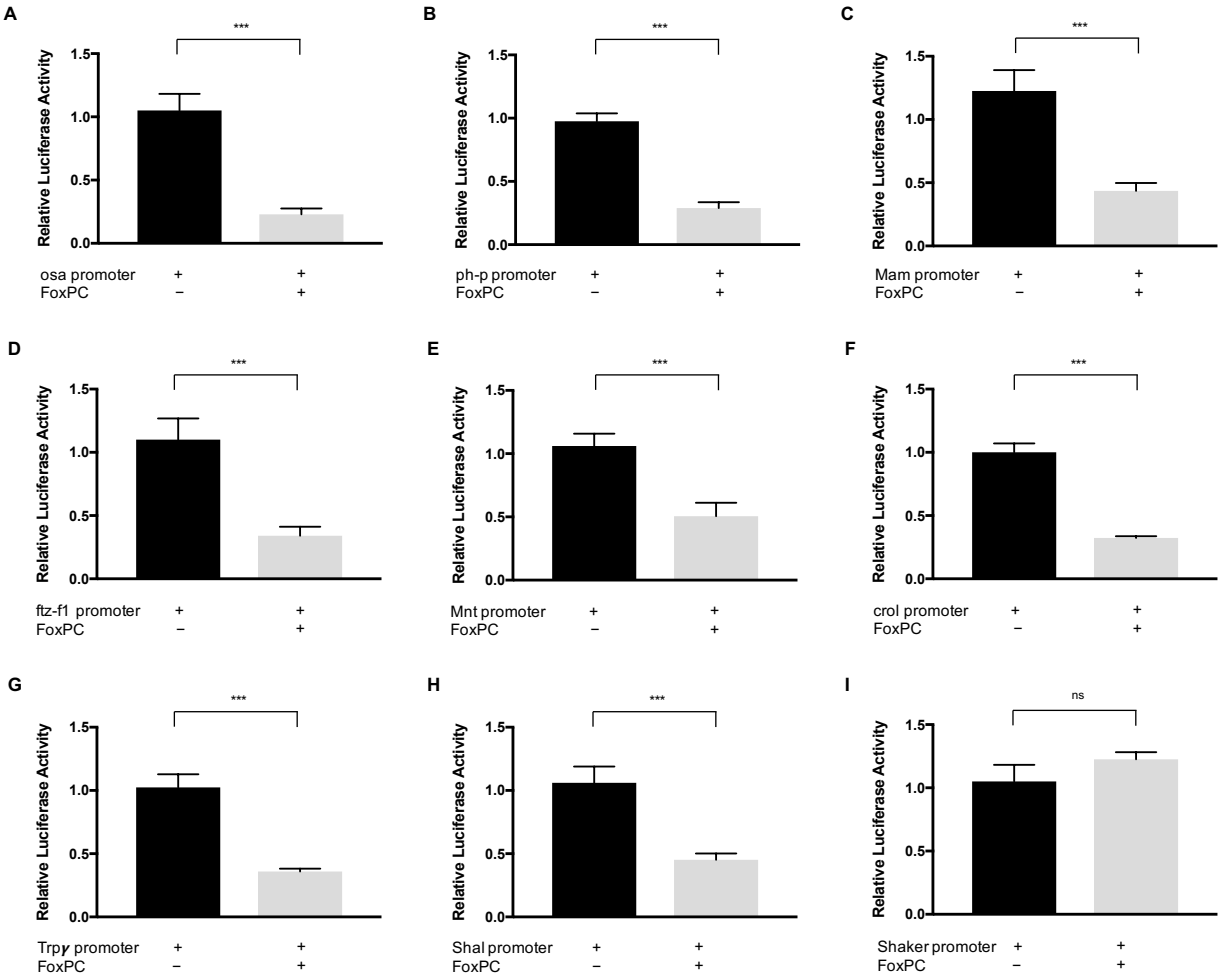


Figure 32. FoxPC represses the activities of its target promoter regions.

*Luciferase assays were carried out using S2 cells transfected with (grey bars) or without FoxPC (black bars), and promoters of the target genes (osa (**A**), ph-p (**B**), Mam (**C**), ftz-f1 (**D**), Mnt (**E**), crol (**F**), Trpγ (**G**) and Shal (**H**)). No change in firefly luciferase activity is noted in the presence of the Shaker promoter (**I**). Results are expressed as mean ± SD of three independent experiments. p values were calculated using the Student's t test. *** $p<0.001$, ns non significant.*

To further validate the binding of FoxPC to the target genes, the FoxPC binding domain was mutated and the difference in binding determined. Santos et al. (Santos et al., 2011) have reported 5 amino acid residues (N550, H554, R553, S557 and W573; FoxP2 numbering), found in the FoxP mammalian orthologue, FOXP2, which form critical contacts with DNA. Since these amino acids, except for H554, are also present in FoxPC (N373, T377, R376, S380 and W397; FoxPC numbering) (Figure 33) they are likely to be important for FoxPC-DNA binding and gene regulation in Drosophila. These amino acids were therefore mutated to alanines, and the effects on the DNA binding and target gene regulation were determined in luciferase assays.

MHRIHDDEYSEDAKESDFKSSIQKEISVKSRHQISIPDICSDAVKNNCFPPTGFLNNSITFASHVVKCS
SPASSIDESSTAAQQHESNPHMHIQGQHMMAPVPDLGFYNVPEFISEQEKLMFSDAERFLRSKDNEVCN
NDFSYMHDEFAMRKYYHPLFAHGICRWPGCEMDLEDITSFVKHLNTEHGLDDRSTAQARVQMQVVSQLE
SHLQKERDRLQAMMHHLYLSKQLLSPTKIDRKDVPGREGKFCRSPLTVNSIGRPIRQTNSPSPLNLPMV
NSTNLCSIKKRNHDKNTFSINGGLPYMLERAGLDVQQEIHRNREFYKNADVRPPFTYASLIRQAIIDSP
DKQLTLNEIYNWFQNTFCYFRRNAATWKNAIRTNLSLHKCFVRYEDDFGSFWMVDDNEFVKRRHLSRGR
PRKYEPSSSPNSCQSGNGVPTDKNPCDNCTQHCTSLPPGADNPLDSNNPNDLGRIGCLPYCGSDGLSKA
SKDYSNMDSGMIESNSHLTIDEYSTNMYESSANEHNR

Figure 33. FoxPC DNA sequence.

The DNA binding domain is shown in blue and the amino acids important for DNA binding domain are shown in red.

When the mutant FoxPC (FoxPC*) was transfected into cells, the luminescence measured from the firefly luciferase was brought back to baseline luminescence measured when the empty control vector was used. The firefly luciferase activity significantly increased when transfecting the mutant FoxPC compared to measurements with the wild-type FoxPC (Figure 34A-H). Mutating the FoxPC DNA-binding domain therefore abrogated the binding activity of FoxPC and its repression of the *osa, ph-p, Mam, ftz-f1, Mnt, crol, Trpγ* and *Shal* promoters. It was found that both wild-type and mutant FoxPC do not affect the *Shaker* promoter activity. This is in agreement with the results from the DamID-seq data which revealed that FoxPC does not bind to *Shaker* (Figure 34I).

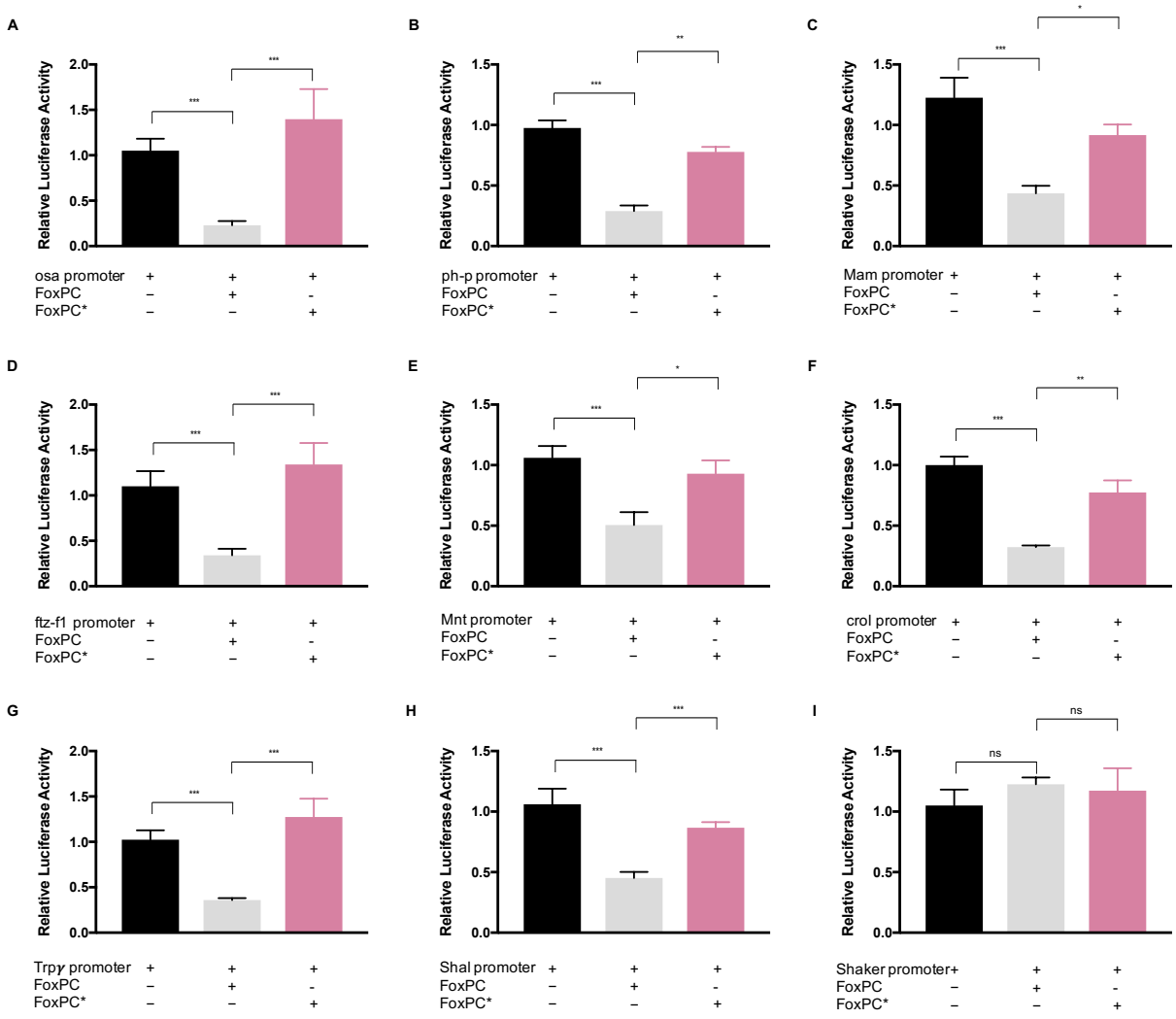


Figure 34. Mutating the FoxPC binding domain abrogates repressive action of FoxPC on promoter regions.

Luciferase assays were carried out in S2 cells using both wild-type FoxPC (grey bars) and mutant FoxPC (FoxPC) (pink bars) and promoter regions of osa (**A**), ph-p (**B**), Mam (**C**), ftz-f1 (**D**), Mnt (**E**), crol (**F**), Trpγ (**G**) and Shal (**H**) and Shaker (**I**). When the DNA binding domain is mutated, FoxPC no longer binds to its target genes and its repressive effect on their activities is relieved (pink bars). No change in firefly luciferase activity is noted in the presence of the Shaker promoter (**I**) when the FoxPC binding domain is mutated. The activity of the promoters return to baseline levels. Bars represent means of at least 6 biological replicates. Error bars represent s.e.m. *** $p<0.001$, ** $p<0.01$, * $p<0.05$ and ns means non significant.*

4.2.4 Confirmation of the FoxPC binding motif by electrophoretic mobility shift assays

There is growing evidence that the DamID-seq experiments and analysis are robust as some of the targets of FoxPC have been validated by several methods, including ChIP, TRAP and luciferase assays. Since they all have the putative FoxPC binding motif (CAACAAAG) in their regulatory regions, the next logical step was to validate the binding of FoxPC to this sequence. The FoxPC putative motif in the *osa, Trpγ* and *Shal* regulatory region was mutated from CAACAAAG to GCCGCCGCC and electrophoretic mobility shift assays (EMSA) were carried out. The probes spanning the CAACAAAG motif or the mutated GCCGCCGCC motif in the regulatory regions were labelled with biotin and incubated with either the FoxPC protein or the mutant FoxPC protein. When biotin-labelled *osa, Trpγ* or *Shal* was incubated with FoxPC (Figure 35 lane 2), a shift was observed compared to biotin-labelled *osa, Trpγ* or *Shal*-only controls (Figure 35 lane 1). This is due to the slower migration of the bound DNA in the FoxPC-DNA complex. However, this shift disappears when either the putative FoxPC motif is mutated or the FoxPC DNA-binding domain is mutated (Figure 35 lanes 3 and 4). This strongly suggests that FoxPC binds to its target genes at the CAACAAAG motif and that residues N373, T377, R376, S380 and W397 are crucial for DNA binding.

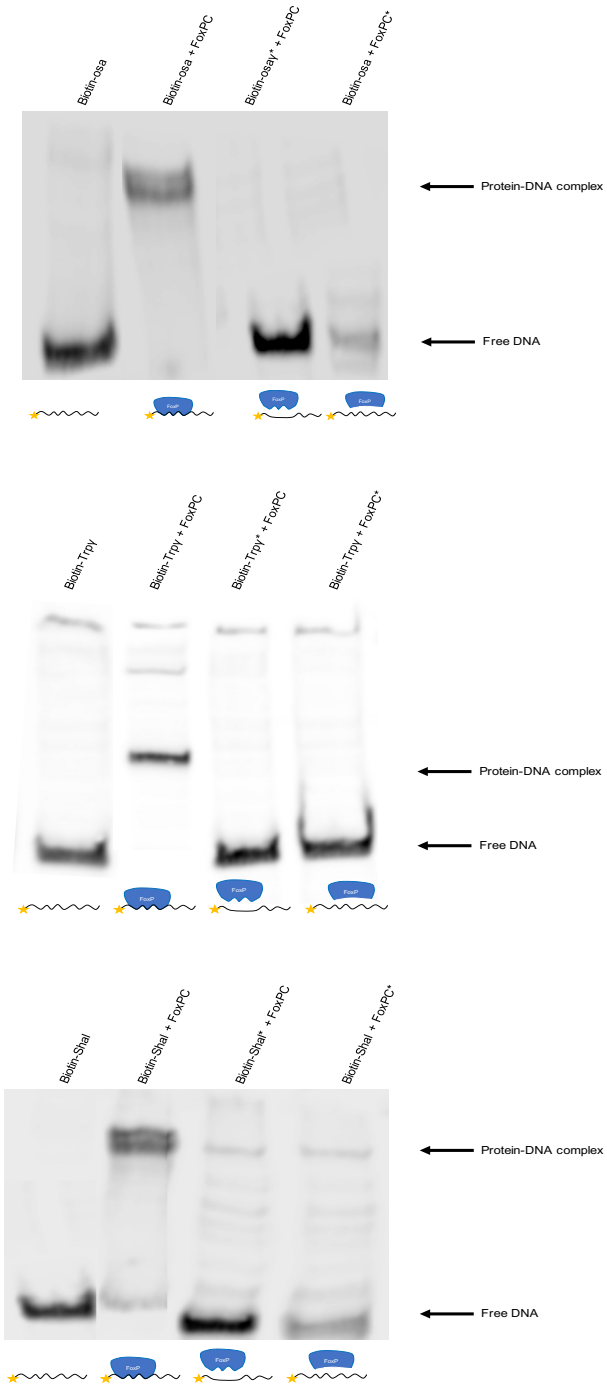


Figure 35. The wild-type FoxPC DNA binding domain is required for binding to the FoxPC motif.

Biotin-labelled probes were detected in reactions containing only the probes (lane 1), both the probe (Biotin-osa, Biotin-Trpγ and Biotin-Shal) and wild-type FoxPC (lane 2), mutated probes (Biotin-osa, Biotin Trpγ* and Biotin-Shal*) and wild-type FoxPC (lane 3) and probes incubated with mutant FoxPC (FoxPC*) (lane 4). The free labelled probes are found at the bottom of the gel, and the bound DNA migrates slower along the gel.*

4.2.5 Confirmation of the FoxPC binding motif in luciferase reporter assays

Binding to the putative motif was then tested functionally in luciferase reporter assays. Theroretically, mutating the FoxPC binding domain and the FoxPC consensus binding sequence in its target genes should have the same effect on the repressive activity of FoxPC. As shown above, wild-type FoxPC does not recognise the mutated consensus sequence and mutant FoxPC no longer binds to the wild-type binding motif. These two scenarios should prevent FoxPC from negatively regulating its target genes. This was tested with the *osa* and *Trpγ* regulatory regions.

Indeed, FoxPC no longer repressed the activity of the firefly luciferase reporter driven by the mutated *osa* and *Trpγ* promoter, with the activity of the luciferase reporter increasing by about two-fold (Figure 36). This demonstrates that FoxPC binds to this consensus motif to regulate the expression of its target genes.

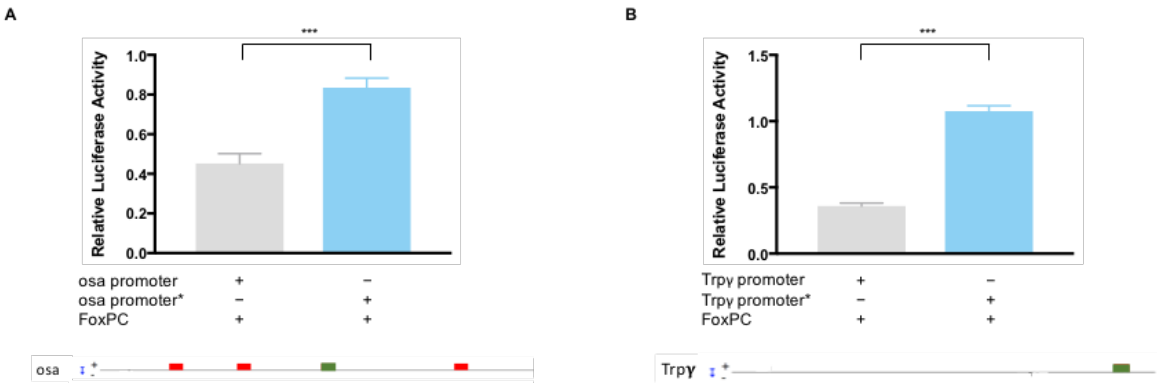


Figure 36. Mutating the FoxPC binding motif affects the regulatory action of FoxPC on its target genes.

*Luciferase assays carried out in S2 cells show that FoxPC binds to the CAACAAAG sequence in the regulatory regions of osa (**A**) and Trpγ (**B**). Mutating the closest motif (green) to and upstream of the TSS brings the firefly luciferase activities back to basal levels (blue bars). *** p<0.001.*

4.3 Interim Discussion from Chapter 4

In this chapter, the putative targets identified by DamID-seq were confirmed in four independent approaches: chromatin immunoprecipitation (ChIP), translating ribosome affinity purification (TRAP), luciferase reporter assays and electrophoretic mobility shift assays (EMSA). In addition to confirming the targets, further insights were gained into the functional consequences of FoxPC-target interactions. Genes found to be bound by FoxPC were also found to be the ones whose transcription was regulated by FoxPC. Their regulatory regions have several FoxPC binding sites (as determined by MAST), which are decorated by FoxPC, as seen in ChIP assays in cell culture and in the intact fly.

TRAP and luciferase reporter assays showed that FoxPC negatively regulates, or represses, the expression of its target genes. This confirms data from Groschner et al., (Groschner et al., 2018) in *Drosophila* and several precedents that FoxP proteins act as repressors in other species (Takahashi et al., 2009, Wang et al., 2003). The FoxP orthologue in humans acts mostly as a transcriptional activator (Spiteri et al., 2007), but can function as a repressor for a small number of genes. The results of this chapter show that FoxPC negatively regulates the expression of all 8 of its significant targets. However, testing its effect on more target genes or carrying out RNA-sequencing might reveal positive regulation also by FoxPC.

Even though the luciferase reporter assays suggest that FoxPC represses the transcription of its target genes, they also showed that FoxPC alone acts as a relatively weak repressor and, like many transcription factors, might therefore work in complex with other transcription factors. This possibility will be explored further in the next chapter.

The amino acid residues important for the repressor activity of FoxPC were determined by mutagenesis coupled with luciferase reporter assays and EMSA. These residues are conserved across species, as they were also found to be present in the human orthologue, FoxP2. To validate the importance of those residues *in vivo*, DamID-seq experiments with the FoxPC mutant constructs could be carried out. Comparing the results with those obtained with wild-type FoxPC would consolidate the findings in this chapter.

The luciferase assays and EMSA also suggest that the putative FoxPC motif identified in the previous chapter is correct. Mutating this motif prevents FoxPC binding to its target promoter regions and hence FoxPC regulation of these promoters. Given that the CAACAAAG sequence has been tested by three techniques, previously by systematic evolution of ligands by exponential enrichment (SELEX) (Nitta et al., 2015), and in this chapter by EMSA and luciferase assays, this sequence can therefore be taken as the genuine FoxPC binding motif.

5

Results

FoxPC regulates its targets as part of a complex

Understanding the mechanisms of eukaryotic transcription regulation remains a challenge. Unlike prokaryotes, eukaryotes usually regulate gene expression in a combinatorial manner, involving the concerted action of several proteins that bind DNA and interact with the transcription machinery (Wolberger, 1999). The resulting complexity affords additional opportunities for the regulation of gene expression that is important in the development of higher organisms. The multicomponent protein assemblies, or enhanceosomes, are usually made up of either several sequence-specific DNA-binding proteins that bind to DNA and each other (Grosschedl, 1995) or several sequence-specific DNA-binding proteins and other proteins recruited by the DNA-bound proteins. These proteins are usually regulatory proteins, such as transcription factors that can dimerise or oligomerise with other transcription factors affecting the rate of transcription initiation. Each transcription factor can also interact with cofactors, such as histone- and DNA-modifying enzymes, to act as positive or negative regulators. Proper spatial and temporal gene expression depends on the presence of several cis-regulatory DNA sequences, for example promoters and enhancers or silencers. These sequences are recognised and bound by several specific transcription factors such that most transcription factors act in complexes.

Combinations of different DNA-binding proteins recognise extended, and more specific, DNA elements to control gene expression more precisely than would be possible with simple operons. The combinatorial network also enables cells to integrate diverse signals (Chen et al., 1998) and increases the sensitivity of the rates of transcription to different transcription factor concentrations (Rosenfeld et al., 2005, Setty et al., 2003). This is possible because transcription factors exhibit spatio-temporal expression patterns and hence form complexes with different transcription factors in different cell types and at different stages of development. The oligomerisation of transcription factors is therefore crucial in the formation of intricate regulatory networks and the differential fine-tuning of gene expression.

Oligmerisation among transcription factors is possible due to their modular structure. In addition to DNA-binding domains, they also possess activation or repression domains that recruit cofactors to form the transcriptional machinery to regulate transcription. Hence, a transcription might either upregulate or downregulate transcription depending on their partner proteins. Many transcription factors have been shown to be capable or heterodimerising or oligomerising, generally through the presence of basic region-leucine zipper (bZIP), basic helix-loop-helix (bHLH), MADS box and Rel homology domains. Dimerisation of transcription factors through these DNA binding domains enables conformational changes to form a functional dimer that allow them to interact with their target genes.

In the previous chapter, the repressive action of FoxPC on its targets genes was elucidated. However, this does not provide a complete picture, as it is highly likely that FoxPC cooperates with other transcription factors as a dimer or through higher order complexes

which recognise different regulatory elements with different degrees of affinity. This inference is supported by luciferase reporter assays, where FoxPC alone does not completely abolish the activity of its target promoters. The structure of the FoxPC protein could give an insight into its combinatorial control of gene expression. The N-terminal zinc finger domain is not directly involved in repression of gene expression, but could be important for binding transcription factors and other proteins (Li et al., 2004). The bi-partite DNA binding domain of the C-terminal forkhead domain of FoxPC could also be partly responsible for complex formation with other proteins. The domain consists of a DNA-binding region and a leucine zipper, which could dimerise with other leucine zipper domain-containing transcription factors.

One of the most effective techniques to detect protein-protein interactions among a large number of potential partners is the yeast two-hybrid (Y2H) system (Fields and Song, 1989). This system exploits the ability of the transcription factor GAL4 to initiate transcription only when its DNA-binding domain (DBD) and its activation domain (AD) are in close proximity. In order to identify interactions between two proteins, for example transcription factors, one of the proteins is fused to the DBD (forming the bait protein) and the other to the AD (giving rise to the prey protein). If the two proteins interact when coexpressed in yeast, the DBD and the AD will be brought close to each other, leading to the transcription of a reporter gene (for example GFP, luciferase, lacZ or HIS3) under the regulation of an upstream activation sequence (UAS) (Figure 37). The reporter gene then allows detection of binding between the two proteins. HIS3 and LacZ reporters are most commonly used; the HIS3 reporter gene allows growth of yeast in medium lacking histidine and the LacZ reporter causes the cells to turn blue in the presence of an X-Gal substrate.

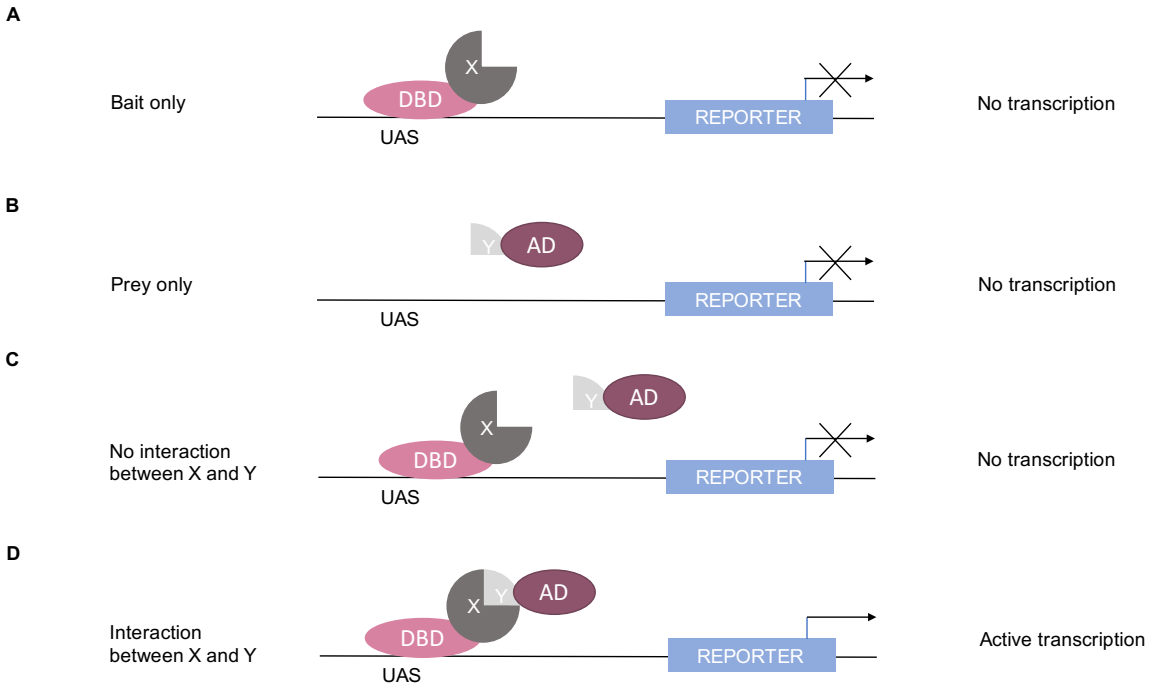


Figure 37. Principles of the yeast two-hybrid system.

*To test for interactions between two proteins (X and Y), fusion proteins are expressed in yeast. The bait is made by fusing one protein to the GAL4 DNA-binding domain (DBD and the prey by fusing the other protein with the GAL4 activation domain (AD). The DBD binds to the upstream activation sequence (UAS) which directs the expression of the reporter gene. Transcription of the reporter gene does not occur if only one of the bait or prey is present (**A** and **B**). If the DBD and AD are not brought close to each other (**C**), no reporter activation is seen, Transcription of the reporter only occurs if the DBD and AD are in close proximity, that is if X and Y interact with each other (**D**).*

5.1 Aims and Objectives

Analysing the protein assemblies regulating transcription is essential for a mechanistic understanding of transcription factor function. To identify possible FoxPC interaction partners, a high-throughput screen of transcription factor interactions was conducted using the yeast two-hybrid (Y2H) system, using FoxPC as the prey and a library of transcription factors as the bait.

Since the Y2H system is prone to false positives, the interaction partners identified were confirmed using co-immunoprecipitation of in vitro expressed proteins and Western Blot. Using the targets identified in the previous chapters, any differential binding and transcription regulation of the target genes by the FoxPC-interaction partner complexes was determined by chromatin immunoprecipitation (ChIP)-qPCR and luciferase reporter assays, respectively. The regulatory impact of the transcription factor complexes on known neuronal and behavioural phenotypes, in this case elevated A-type potassium currents and decision making, were tested in electrophysiological recordings and odour discrimination tasks.

5.2 Results

5.2.1 Identification of putative interaction partners

70 transcription factors were cloned into expression vectors containing the GAL4 AD domain and used in two-hybrid screening to discover interactions with FoxPC fused to the GAL4 DBD domain by Johanna Jung and Dr Korneel Hens (Centre for Neural Circuits and Behaviour). To reduce the positive interaction hit rate, two reporter genes (HIS3 and LacZ) were used in independent experiments. Out of the 70 transcription factors, five transcription factors, glial cells missing (gcm), Sp1-like factor for pairing sensitive-silencing (Spps), Dichate (D), Abdominal-B (AbdB) and Atf6 were found to possibly interact with FoxPC. Examples of Y2H results for FoxPC and Atf6 and Spps are shown in Figure 38. FoxPC was also found to homodimerise, as is common with zinc finger domain proteins.

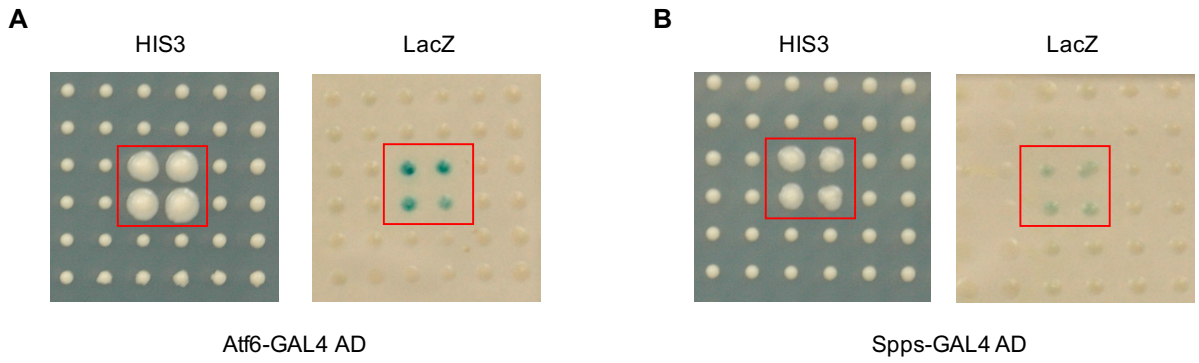


Figure 38. Y2H assay uncovers direct interactions between FoxPC and 5 transcription partners. *Y2H assays identified direct interactions between Atf6-GAL4 AD and FoxPC-GAL4 DBD (**A**) and Spps-GAL4 AD and FoxPC-GAL4 DBD (**B**) using the HIS3 and LacZ reporter genes.*

5.2.2 Confirmation of putative interaction partners

Even though yeast-two hybrid assays are powerful tools in high-throughput studies to detect interacting proteins in living cells, they are often prone to errors and can identify false positives (Williamson and Sutcliffe, 2010). Therefore, it is important to validate the interactions using other assays, such co-immunoprecipitation and Western Blot. FLAG-tagged FoxPC and GST-tagged gcm, Spps, Dichaete, Atf6 and Abd-B were first expressed *in vitro* by a coupled transcription and translation system. The template DNAs under the control of the bacteriophage SP6 promoter were first transcribed into mRNA, followed by the translation of the mRNA into protein. This system can quickly produce large amounts of the modified proteins, which can be used for downstream detection of protein-protein interactions.

Interactions between each GST-tagged putative interaction partner and FoxPC were then validated by co-immunoprecipitation, followed by Western Blot. Since overexpressing proteins can lead to the appearance of non-specific complexes and false interactions, a negative control (GST-tagged acj6) was included. acj6, a transcription factor expressd in the peripheral olfactory system (Lee and Salvaterra, 2002, Komiyama et al., 2003) was not

identified as a potential FoxPC interaction partner in the Y2H. To eliminate the background or non-specific binding to the anti-GST beads, co-immunoprecipitation was also carried out between FLAG-tagged FoxPC and water (lane 7 in all blots in Figure 39).

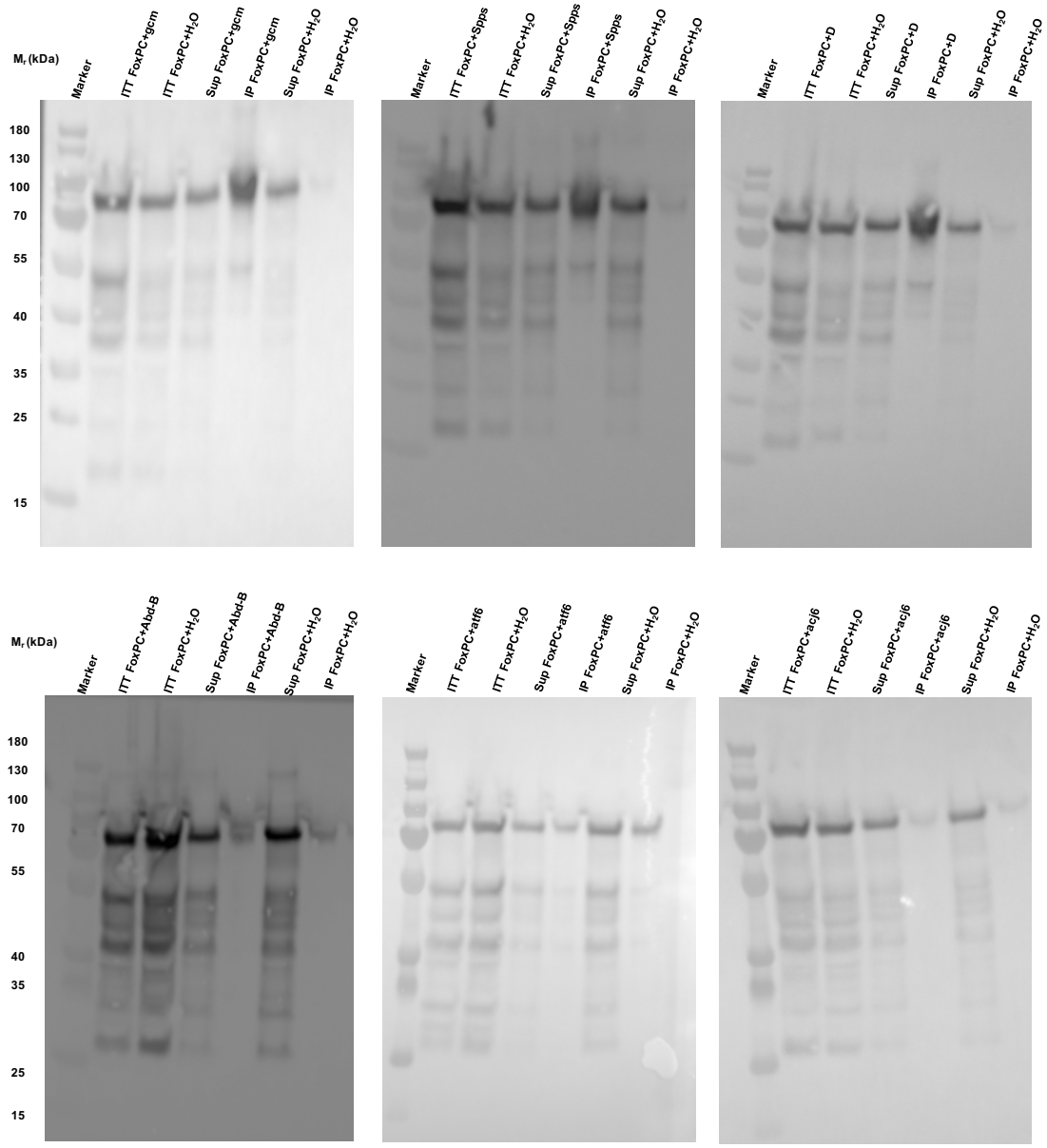


Figure 39. FoxPC binds to gcm, Spps and D.

Co-immunoprecipitation between the FLAG-tagged FoxPC and GST-tagged gcm, Spps, D, Abd-B, atf6 and acj6 was carried out. The proteins were expressed in vitro in ITT mixes (lanes 2 and 3) and then subject to immunoprecipitation with GST-coated beads. The supernatants (Sup) show the unbound proteins (lanes 4 and 6) and the immunoprecipitated fractions (IP) contain the proteins bound to the beads (lanes 5 and 7). Controls using water instead of GST-tagged proteins were included (lanes 3, 6 and 7). The band at ~70 kDa corresponds to the molecular weight of FLAG-tagged FoxPC.

Co-immunoprecipitation using anti-GST beads to isolate the GST-tagged proteins and immunoblotting using an anti-FLAG antibody, which recognises the FLAG-tagged FoxPC, confirmed most of the interactions uncovered by the Y2H assay. GCM, spps and Dichaete bind strongly to FoxPC (Figure 40, top row). Western Blot analysis revealed a band at ~70kDa, corresponding to the molecular weight of the FLAG-tagged FoxPC, in the immunoprecipitated fraction (lane 5 in all blots). On the other hand, no prominent bands above background were seen when Atf6 and Abd-B were incubated with FoxPC, suggesting no direct interaction between these transcription factors and therefore possible false positives in the Y2H (Figure 40, bottom left and centre). Reassuringly, FoxPC does not bind to the negative control, acj6, confirming the Y2H result (Figure 40, bottom right).

5.2.3 Analysis of expression levels of interaction partners in FoxPC-positive neurons

Interactions between FoxPC and its interaction partners only have physiological significance if the two interaction partners are expressed in the same neurons. The expression levels of *gcm, Spps* and *Dichaete* were therefore analysed in $\alpha\beta_c$ neurons, which express high amounts of FoxPC (Figure 30). All three partners identified by both the Y2H and co-immunoprecipitation were found to be expressed in those neurons, relative to the reference genes (Figure 40). No expression of *acj6* was seen, in accordance with previous results (Certel et al., 2000).

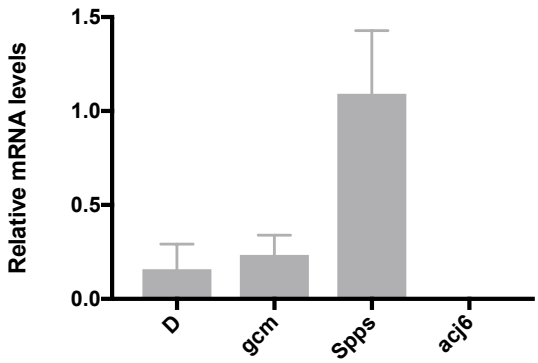


Figure 40. D, gcm and Spps are expressed in FoxPC-positive αβ$_c$ neurons labelled by NP7175-GAL4.

mRNA from the αβ$_c$ neurons was isolated by TRAP and quantified by qPCR using primers specific for D, gcm, Spps and acj6. Levels of polysome-bound transcripts are shown relative to the geometric mean of three reference gene transcripts. The bars represent the means and the error bars the s.e.m.

5.2.4 Analysis of FoxPC transcription complexes binding to target genes

The identification of the different interaction partners suggests that FoxPC could form complexes with different partners to regulate different genes. This was tested using the target genes identified in the previous chapters: *osa, ph-p, Mam, ftz-f1, Mnt, crol, Trp γ* and *Shal*.

Brains from flies expressing both HA-tagged FoxPC and RNAi hairpin loops against each of the interaction partners in the αβ$_c$ neurons (using the NP7175-GAL4 driver line) were used in ChIP experiments. The DNA fragments bound by FoxPC were then amplified and quantified using real-time PCR. The goal was to determine the importance of the interaction partners in FoxPC's binding to its target genes. As in Chapter 4, the input chromatin (material before immunoprecipitation) was used to normalise for background levels. Primers spanning regions upstream and downstream of the transcription start sites were used. These experiments showed that the FoxPC target genes are bound, and possibly regulated, by different FoxPC complexes (dimers or trimers). The *osa* promoter

regions are bound by both FoxPC and Spps (Figure 41A). In the absence of Spps, the binding of FoxPC to the *osa* regulatory regions is abolished. *ph-p, ftz-f1, Trp γ* and *Shal* (Figure 41B, D, G and H) are regulated by FoxPC, Spps and gcm. Binding of FoxPC to *Mam* is affected by the presence of gcm only (Figure 41C). Binding of FoxPC to the *Mnt* promoter regions is also decreased when gcm levels are depleted, although this decrease is not significant (as determined by the student's t-test) (Figure 41E). *crol* expression levels seem to be significantly affected by the presence of FoxPC only. As a control, the binding of FoxPC to the promoter regions of its target genes was analysed in the absence of acj6. FoxPC binding is not affected by the absence of acj6 (grey bars in all charts),implying that FoxPC does not require acj6 to regulate its target genes.

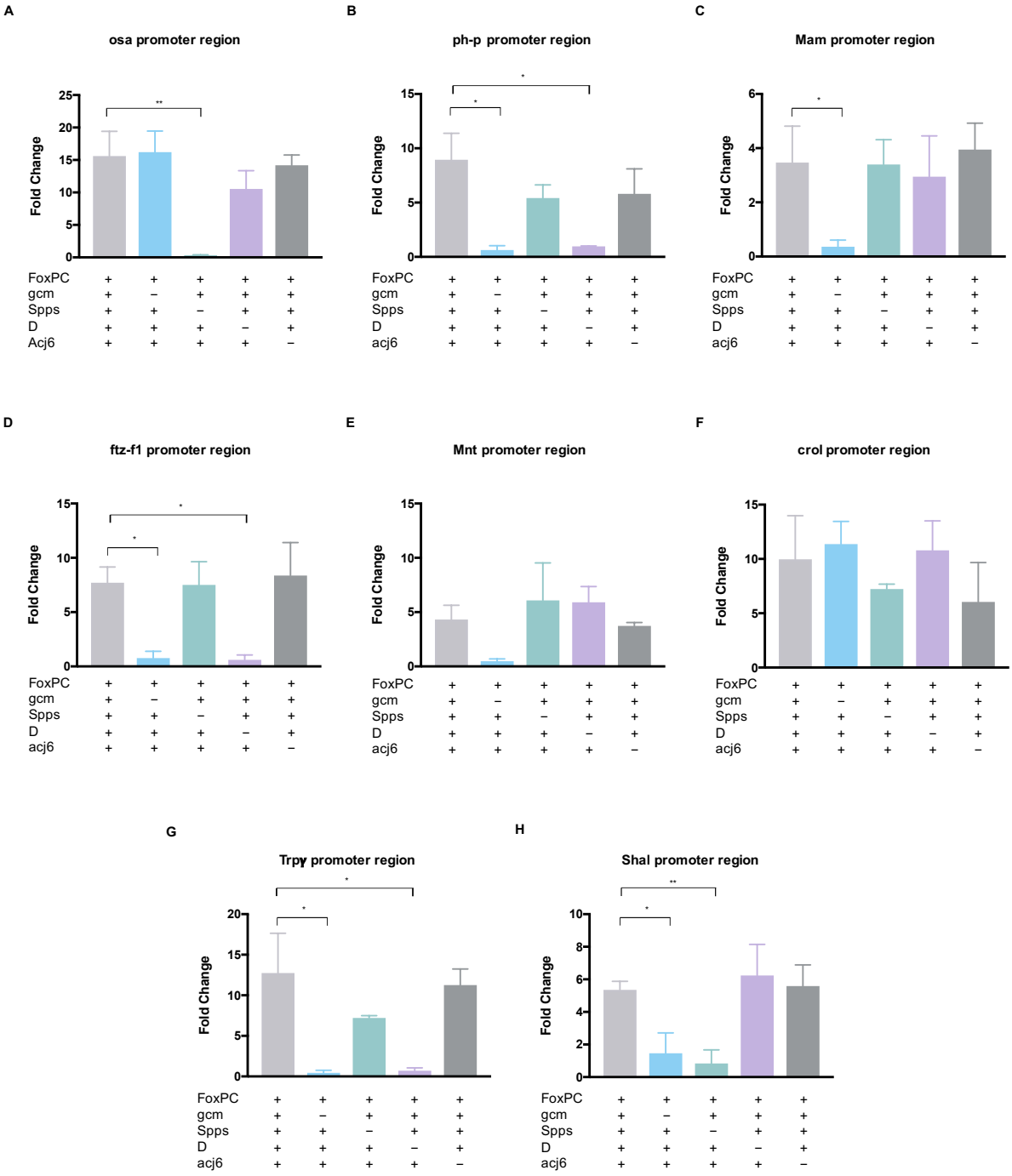


Figure 41. Chromatin immunoprecipitation in Drosophila brains suggests that FoxPC binds to its target promoters as part of a complex.

*Binding of FoxPC to the different regulatory regions (osa (**A**), ph-p (**B**), Mam (**C**), ftz-f1 (**D**), Mnt (**E**), crol (**F**), Trpγ (**G**) and Shal (**H**)) in the absence of its interaction partners (gcm (blue bar), Spps (green bar) and D (purple bar)) and a negative control (acj6 (dark grey bar). The change in binding is compared to binding in wild-type flies (light grey bar). ** p<0.01 and * p< 0.05.*

5.2.5 Determination of locations of binding motifs in promoter regions of target genes

Binding of FoxP to other transcription factors suggest that they act in tandem to regulate the expression of different genes. This could happen if the promoter regions or the cis-regulatory modules (CRMs) of the target genes contain both the FoxPC motif and that of the co-associated transcription factor partners (Spitz and Furlong, 2012), such that they bind to and regulate the target DNA in a directly cooperative manner. Another scenario, indirect cooperative gene regulation, would require binding of FoxPC to its targets and the recruitment of the interaction partners and other proteins to form a functional repressor complex. The former was tested using MAST, which looks at the occurrences and locations of motifs in DNA sequences. The gcm, Dichaete and Spps motifs (ATGCGGG(C/T), AACAAT and GGGGCG, respectively) used in the analysis had previously been identified (Cattenoz et al., 2016, Aleksic et al., 2013, Brown et al., 2005). Since FoxPC was found to bind close to the transcription start site of the target genes and CRMs can span several hundred base pairs to 1 kilobase (kb) in length, the 1 kb region upstream and downstream of the transcription start sites were used for analysis. Only motif matches with an E-value less than 10 are displayed. The MAST output (Figure 42) results match the results seen from the ChIP-qPCR experiments (Figure 41). The regulatory region of *osa* harbours binding sites for Spps and FoxPC while those of *ph-p, ftz-f1* and *Trpγ* contain binding sites for FoxPC, gcm and Dichaete. The *Shal* regulatory region contains the FoxPC, Spps and gcm motifs. The *Mam* and *Mnt* promoters both contain the FoxPC and gcm motifs. This suggests that the non-significant decrease in FoxPC binding to the *Mnt* promoter in the absence of gcm is either due to experimental variability or to the requirement of another transcription factor to form a complex with FoxPC and gcm. The *crol* promoter only contains FoxPC motifs. None of the promoter regions contain the acj6 motif, in accordance with the ChIP-qPCR results. The *Shaker (Sh)*

promoter does not harbour any of the binding sites, consistent with its insensitivity to FoxPC. The fact that some motifs have different orientations or are found on different strands does not affect their regulatory function (Lis and Walther, 2016).

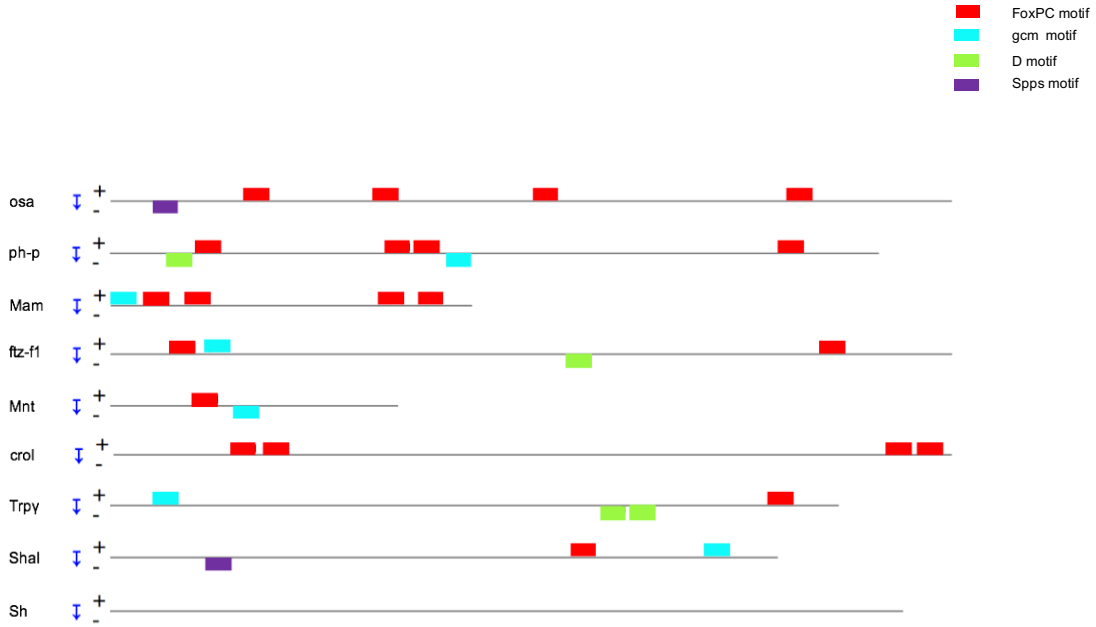


Figure 42. FoxPC and interaction partners' motifs locations.

MAST output showing the locations of the FoxPC (red), gcm (blue), D (green) and Spps (purple) motifs. 1000 bp upstream of the TSS and the region up to the end of the first intron were scanned for the presence of the different motifs. All of the input sequences harbor different combinations of motifs except crol, which only contains binding sites for FoxPC. Shaker (Sh) does not harbor any of the motifs tested.

5.2.6 FoxPC transcription complexes and gene regulation

Binding of transcription factors to their binding sites on target genes does not necessarily have a biological significance. The regulatory effects of the FoxPC-complexes on the activities of the target promoters were therefore tested using luciferase reporter assays (Figure 43). S2 cells were transfected with FoxPC and the different interaction partners, together with the promoter regions of the target genes driving the expression of firefly luciferase. The effects on the promoter activities were determined by comparing

luminescence outputs from cells transfected with FoxPC and the interaction partners (grey bars) to control cells, which do not overexpress any of the transcription factors (black bar). The results showed that FoxPC and its partners cooperatively affect the activities of the promoters they bind to, possibly by increasing their DNA-binding activities. FoxPC, gcm, Spps and Dichaete have significant repressive effects on their target genes even when they are expressed on their own, meaning that they act as active repressors. In the presence of FoxPC-transcription factor complexes (that is when both FoxPC and its interaction partners are present), the repressive effect on the target genes is enhanced in most cases. For example, in the presence of both FoxPC and Spps, the activity of the *osa* promoter is at least half of the activity when only either FoxPC or Spps are present. When gcm and Dichaete are present along with FoxPC, the luciferase activity driven by the *osa* promoter is comparable to baseline luciferase activity, suggesting that most of the FoxPC repression has been relieved. This could be explained by the formation of FoxPC-gcm or FoxPC-Dichaete complexes, hence preventing FoxPC from binding to its target sequence and from repressing the promoter activity. FoxPC forms different complexes with gcm, Spps and Dichaete to repress the activities of its different promoter targets to different extents. The FoxPC-gcm complex has a more pronounced effect on the *ph-p* and *ftz-f1* promoters than the FoxPC-Dichaete complex (Figure 43B and D). This could be due to different strengths of the individual promoters. These results imply that FoxPC, like most transcription factors, acts in a combinatorial manner and works in concert with other transcription factors to enhance each other's regulatory activities. FoxPC could form functional dimers with each of its interaction partners to trigger different regulatory events. The finding that different dimers can have the same effect on the same promoter (for example FoxPC-gcm and FoxPC-spps on the *Shal* promoter (Figure 43H) suggests

that there is regulatory redundancy, or that FoxPC could form higher-order oligomers, by binding to two or more transcription factors.

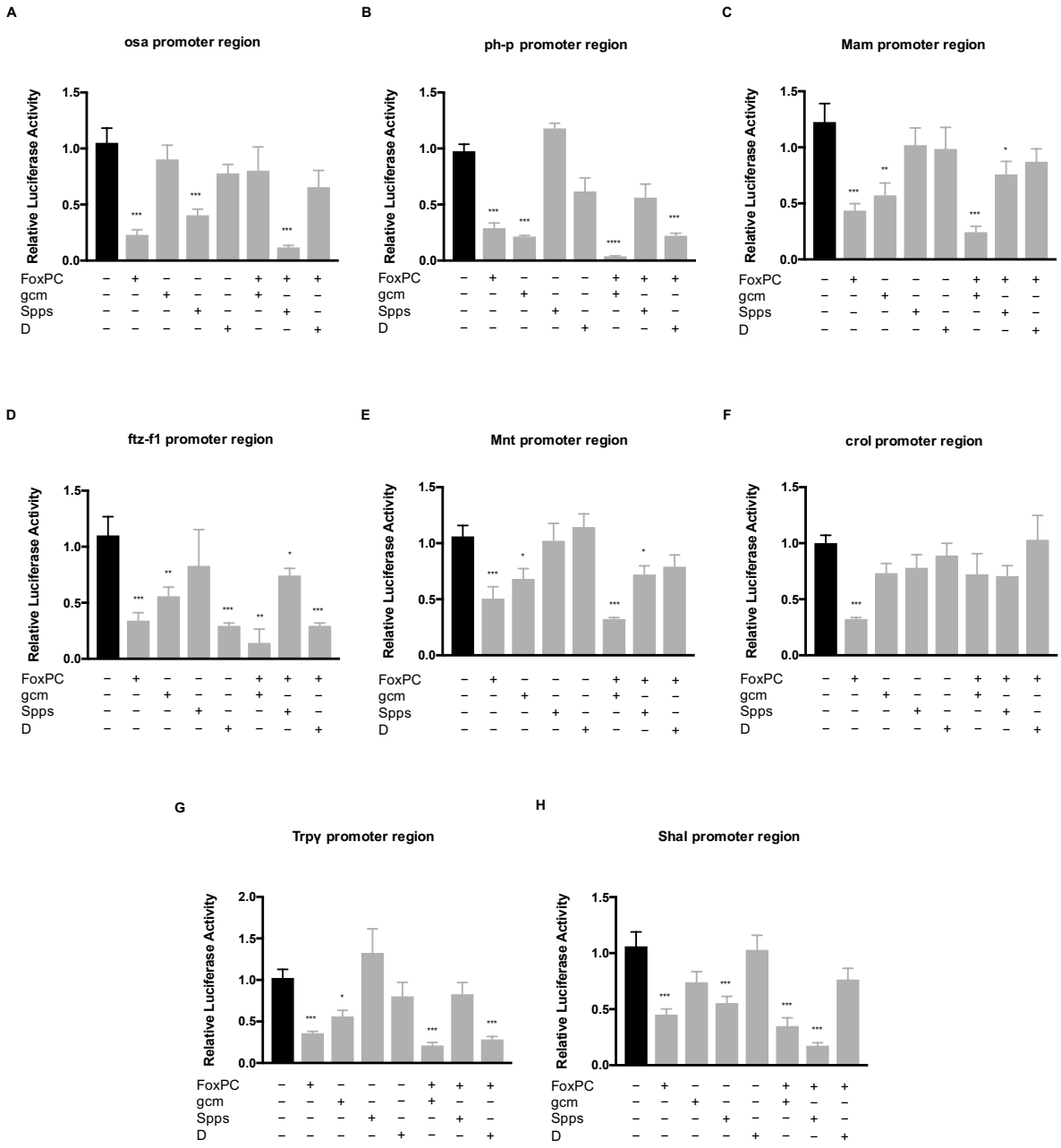


Figure 43. FoxPC forms complexes with different transcription factors to regulate promoter activity.

*FoxPC, gcm, Spps and Dichaete were transfected into S2 cells on their own (bars 2-4), or together with FoxPC (bars 5-7). The effects of each transcription factor and their interactions on target promoter regions (osa (**A**), ph-p (**B**), Mam (**C**), ftz-f1 (**D**), Mnt (**E**), crol (**F**), Trpγ (**G**) and Shal (**H**)) were determined by luciferase reporter assays. The firefly luciferase activities in each experimental scenario are compared to the basal firefly luciferase activity (black bar). *** p< 0.001, ** p< 0.01 and * p<0.05.*

5.2.7 FoxPC transcription complexes and biophysical properties of FoxPC-positive cells

FoxPC was shown to repress the voltage-dependent A-type potassium current (I_A) mediated by Shal in the FoxPC-positive $\alpha\beta_c$ KCs (Groschner et al., 2018). Judging from ChIP and luciferase reporter data, it is likely that some of FoxPC's interaction partners also participate in this process. The consequences for I_A currents of removing these interaction partners were therefore determined in whole-cell voltage-clamp experiments *in vivo* by Dr Lukas Groschner (Centre for Neural Circuits and Behaviour). The loss of FoxP almost doubled the membrane density of the I_A current (Figure 45A) compared to control currents, while the sustained delayed rectifier current (I_K) does not change significantly (Figure 45 B), in agreement with data from Groschner et al. (Groschner et al., 2018). A significant increase in the membrane density of the I_A current was also seen when the expressions of gcm and Spps were knocked down in $\alpha\beta_c$ neurons. The decrease was, however, smaller than that seen with $FoxP^{RNAi}$ flies. When Dichaete and acj6 levels were knocked down, the I_A density was comparable to that in control flies, suggesting that they do not affect Shal expression levels. Similar I_K currents (~17 pA/pF) are seen in gcm-, Spps-, Dichaete- and acj6-deficient flies (Figure 45B), confirming that neither FoxPC nor its interaction partners regulate the channels carrying these currents (Shab and Shaw). The membrane capacitance (C_m) was indistinguishable (~2.5 pF) between all genotypes tested (Figure 45C). Given that membrane capacitance is a function of membrane area, this finding rules out differences in cell sizes as being the cause of changes in I_A currents.

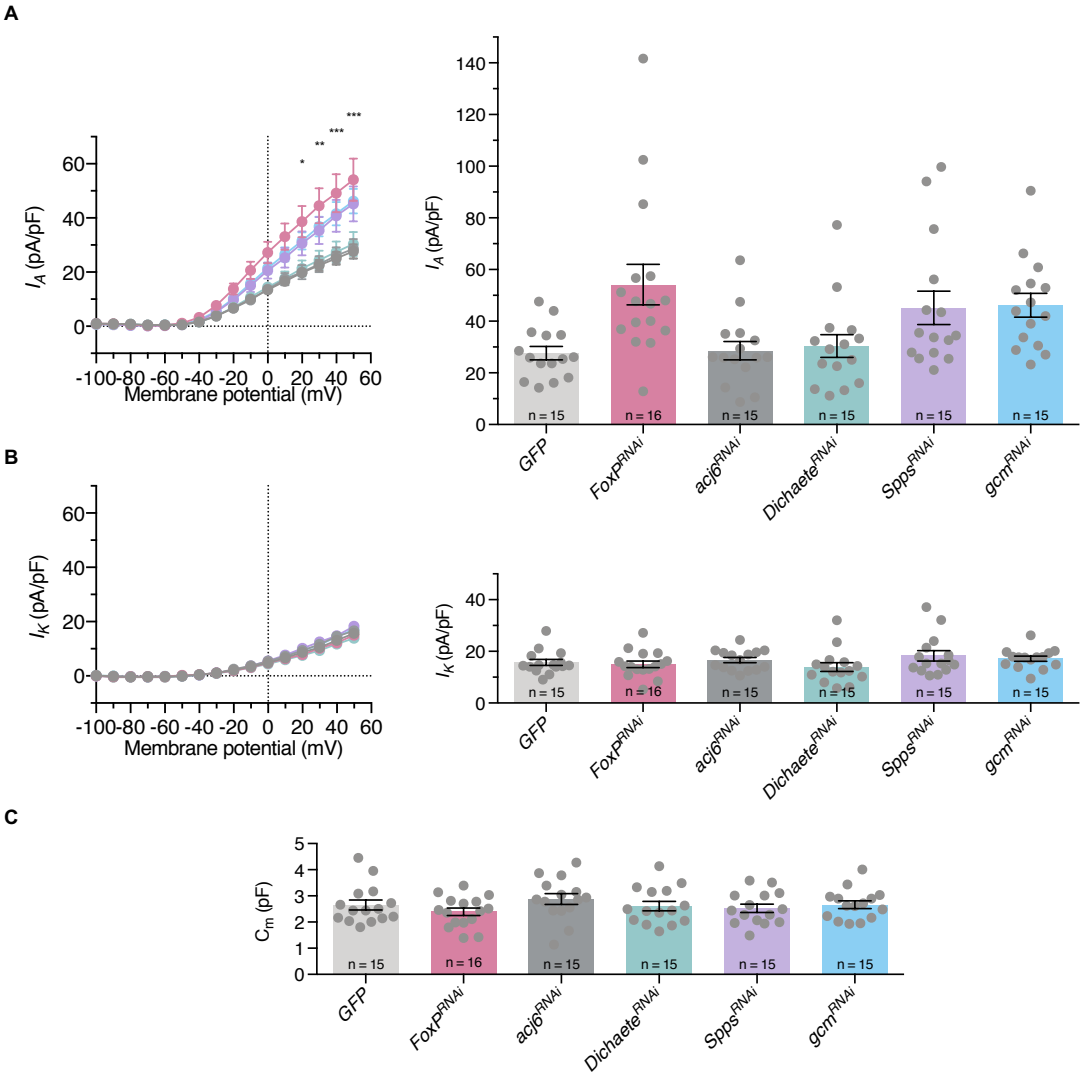


Figure 44. FoxPC, Spps and gcm repress an A-type potassium current in αβ$_c$ KCs.

*I_A, I_K and C_m measurements for wild-type (GFP) (light grey curves and bars), FoxP-deficient (pink curves and bars), acj6-deficient (dark grey curves and bars), Dichaete-deficient (green curves and bars), Spps-deficient (purple curves and bars) and gcm-deficient (blue curves and bars). Circles, means ± s.e.m.; solid lines, Boltzmann fits; asterisks, significant differences from wild-type in post-hoc comparisons following two-way ANOVA **A**, Average A-type potassium current densities. **B**, Average delayed rectifier current densities. **C**, Membrane capacitance (Cm) measurements.*

5.2.8 FoxPC transcription complexes and decision-making

FoxPC has been shown to affect the integrative properties of the αβ$_c$ neurons and decision making in *Drosophila (Groschner et al., 2018)*. The effects of the interaction partners at a behavioural level were therefore examined. Flies were trained to avoid a specific

concentration (20 ppm) of 4-methylyclohexanol (MCH) by pairing odour exposures with electric foot shock (Claridge-Chang et al., 2009, DasGupta et al., 2014). The animals then had to discriminate the reinforced concentration of MCH from a lower intensity (18 ppm) of the same odour (Figure 45). Note that this (small) concentration difference corresponds to the highest difficulty in this task as shown in Figure 4B. This level of difficulty of discrimination was chosen because (*i*) reaction times increase and (*ii*) performance decreases with difficulty. Thus, this difficulty level pushes the decision-making capabilities of the animal to the limit and should enable the detection even of subtle behavioural changes due to genetic interventions.

Flies expressing hairpin *RNA* against gcm and Spps in FoxPC-positive cells ($\alpha\beta_c$ neurons) had similar decision accuracies, but significantly increased reaction times, compared to control flies. Expressions of hairpin *RNAi* against Dichaete and acj6, however, yielded similar reactions times to the wild-type values. This means that only gcm and Spps affect the integrative properties of the $\alpha\beta_c$ neurons, possibly through the formation of FoxPC-gcm and FoxPC-Spps transcription complexes. The disruption of these complexes is expected to extend decision times through the de-repression of Shal (Groschner et al., 2018).

These behavioural data, like the luciferase assay results, suggest that FoxPC, gcm and Spps, but not Dichaete and acj6, expression levels affect Shal promoter activity, and hence Shal expression levels. The data also reaffirm that the Shal currents in the $\alpha\beta_c$ neurons are crucial in FoxPC's role in decision making.

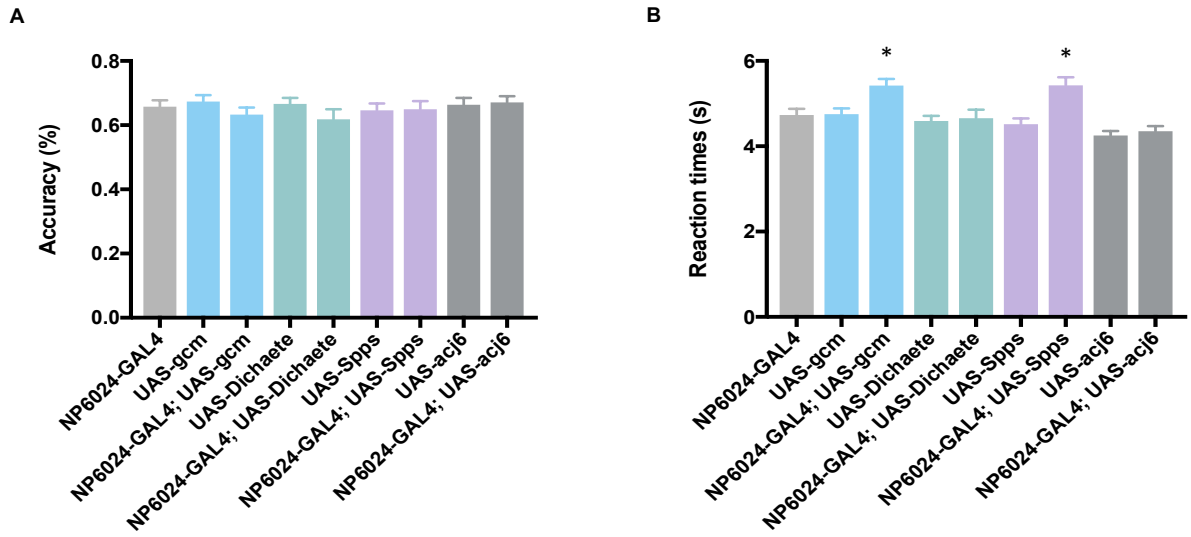


Figure 45. gcm and Spps in the αβ$_c$ neurons using NP7175-GAL4 driver line determine reaction times.

Knocking down the levels of gcm, Dichaete, Spps and acj6 in the αβ$_c$ neurons affects reaction times, but not accuracies in 150 flies tested. ***A****, Accuracies of the parental (GAL4 and UAS) flies are similar to the experimental flies (GAL4-driven UAS).* ***B****, Driving the expression of gcm^{RNAi} and Spps^{RNAi} in the αβ$_c$ neurons significantly increase the reaction times compared to the parental controls. Data are means ± s.e.m.); asterisks, significant differences of experimental flies from both parental controls (accuracy: Dunn's test following one-way Kruskal–Wallis test; reaction time: Kolmogorov-Smirnov test with Bonferroni-corrected P < 0.00625).*

5.3 Interim Discussion from Chapter 5

Many members of eukaryotic transcription factor gene families require interactions with an identical molecule or another transcription factor to form functional dimers to bind DNA. The choice of interaction partner depends on the cellular context and determines the regulatory effects that affect cell processes. Dimerisation is crucial for differential recognition of DNA regulatory elements and regulation of genes, increased proximity of regulatory proteins and enhancement of DNA-binding specificity (Klemm et al., 1998). Transcription factor families which commonly form homotypic dimers include leucine zippers, helix-loop-helix and the MADS box transcription factors.

In this chapter, I showed that FoxPC acts in heteromeric complexes to regulate the expression of its target genes. 70 most promising transcription factors were tested for interaction with FoxPC by yeast two-hybrid (Y2H) assays, yielding 5 potential interaction partners: gcm, Spps, Dichaete, Atf6 and Abd-B. Since Y2H assays often result in false positives, the interactions were further tested by co-immunoprecipitation assays with FoxPC. Three of the positive hits, gcm, Spps and Dichaete, from Y2H were confirmed and taken forward to analyse the functions of different FoxPC transcription complexes. A negative control, acj6, which does not interact with FoxPC as determined by Y2H and co-immunoprecipitation, was also used to confirm the validity of the functional assays. The putative interaction partners are present in FoxPC-positive neurons as determined by TRAP experiments and are therefore able to bind each other *in vivo*. Acj6, on the other hand, is not present in these neurons.

Chromatin immunoprecipitation showed that FoxPC forms complexes with different combinations of transcription factors on different promoter regions. The complexes potentially have different properties and specific functions as they regulate different genes. Strinkingly, some of FoxPC's target genes were already identified as targets of gcm, including *ph-p* and *ftz-f1* (Cattenoz et al., 2016). This confirms that FoxPC and gcm cooperate to regulate certain genes.

Luciferase assays showed that the repressive effect on the target genes' promoter regions is enhanced when both FoxPC and its interaction partners are present. Cooperative binding of transcription factors is a well-known phenomenon (Graves, 1998). Full repression of the target genes was not observed in any case, which could be explained by the absence of other possible interaction partners, unphysiologically high promoter activity in the

synthetic reporter constructs or a role for FoxP in titrating mRNA dosage rather than switching transcription off completely, as suggested by the functional data of Groschner et al. The list of transcription factors tested was not exhaustive. To get a complete picture of FoxPC's regulatory actions, the full list of transcription factors in *Drosophila* would have to be tested.

Transcription factors which cooperate in multi-protein complexes are often found bound to the same stretch of DNA (Aguilar and Oliva, 2008). This means that their binding sites will be clustered within regulatory regions. This was tested using MAST to identify the locations of the binding motifs of the different interaction partners. The MAST output correlates with the ChIP and luciferase reporter assays as they show that only the transcription factors which bind to the promoter regions affect promoter activity. Another interesting information that can be gathered from the MAST output is that the number of FoxPC binding motifs present in the promoter regions possibly affects the $\log_2$ ratio (and hence the intensity of binding) obtained from the DamID-seq analysis and the degree of repressive action of FoxPC, on its own, on its target genes.

The importance of the FoxPC heterodimers on the biophysical properties of FoxPC-positive cells was tested by *in vivo* whole-cell voltage clamp recordings. The expression levels of FoxPC, gcm and Spps affect the membrane density of the I_A current, which was attributed to the abundance of the dendritic voltage-gated potassium channel, Shal (K_V4), at the cell membrane (Groschner et al., 2018). This correlates with the ChIP and luciferase data, which showed that FoxPC, gcm and Spps also bind to the *Shal* promoter regions. Involvement of the potential interaction partners in the integration of sensory evidence, partially attributed to Shal currents, was confirmed in two-odour discrimination tasks. gcm

and Spps affect reaction times, albeit to a smaller degree that FoxPC (Groschner et al., 2018).

6

Discussion

6.1 FoxP regulates a plethora of genes

Gene regulation is vital for the biological fitness of multicellular organisms. The major regulatory step in the maintenance and regulation of protein content occurs at the level of transcription and is carried out by transcription factors. Sets of genes transcribed into mRNA determine the protein content of each cell and hence their states and their response to stimuli. The key concepts of transcriptional control were discovered in bacterial systems (Jacob and Monod, 1961). It was established that DNA-binding transcription factors bind to specific control elements (cis-elements) and recruit the transcription machinery. However, transcriptional-regulatory networks in eukaryotes are significantly more complex. They often involve several transcription factors, which bind in a cooperative manner to regulatory sequences to influence the expression of their target genes.

Despite the considerable amount of work aimed at understanding eukaryotic gene transcription, it is still unclear how individual genes are selectively turned on and off by specific transcription factors. In this thesis, I aimed to take a step closer to understanding this by studying how FoxP influences the expression of its targets in *Drosophila melanogaster*. More specifically, I took advantage of the recently discovered influence of the FoxPC on the integration of sensory information during decision-making to investigate the link between gene regulation, its consequences for the biophysics of neurons, and,

ultimately, its regulation of complex, adaptive behaviours. Using a combination of molecular techniques, including FoxP-tethered DNA adenine methyltransferase identification (DamID), chromatin immunoprecipitation (ChIP), luciferase reporter assays and bioinformatics analyses, FoxPC's direct targets, its binding motif and its regulatory effects on its targets were investigated.

In *Drosophila melanogaster*, two isoforms of FoxP, FoxP-RC and FoxP-RD (referred to as FoxPC and FoxPD in this thesis) arising from an alternative splicing event that joins a unique last exon (exon 8) with the common body of the gene, are present. The resulting proteins share dimerisation motifs but possess distinct C-terminal forkhead domains, affecting their DNA recognition sequence (Santos et al., 2011). Both forkhead domains are conserved throughout evolution, suggesting that both domains are functionally significant. The different putative DNA binding domains of the two FoxP isoforms suggest their possible regulation of different target genes and could also explain their differential expression. For example, other studies in *Drosophila* and in humans have also reported differential expressions of alternatively spliced transcription factors in the brain such as the transcription factor FoxP1 (Wang et al., 2003). Furthermore, as isoforms from a single locus are important in regulating different physiological pathways in different neuronal substrates, it is important to analyse the binding patterns of both isoforms and the functionality of binding.

In this thesis, a thorough and systematic approach (depicted in the flowchart in Figure 46) was devised to delineate the transcriptional regulatory network in which FoxP operates, with particular emphasis on the decision-relevant isoform, FoxPC. DamID-sequencing was used to first identify targets and binding motifs of both FoxPC and FoxPD. The FoxPC targets were tested using ChIP, expression profiling using TRAP and luciferase

reporter assays. Cooperating partners were then identified by yeast two-hybrid assays and validated by ChIP, TRAP, reporter, electrophysiological and behavioural assays.

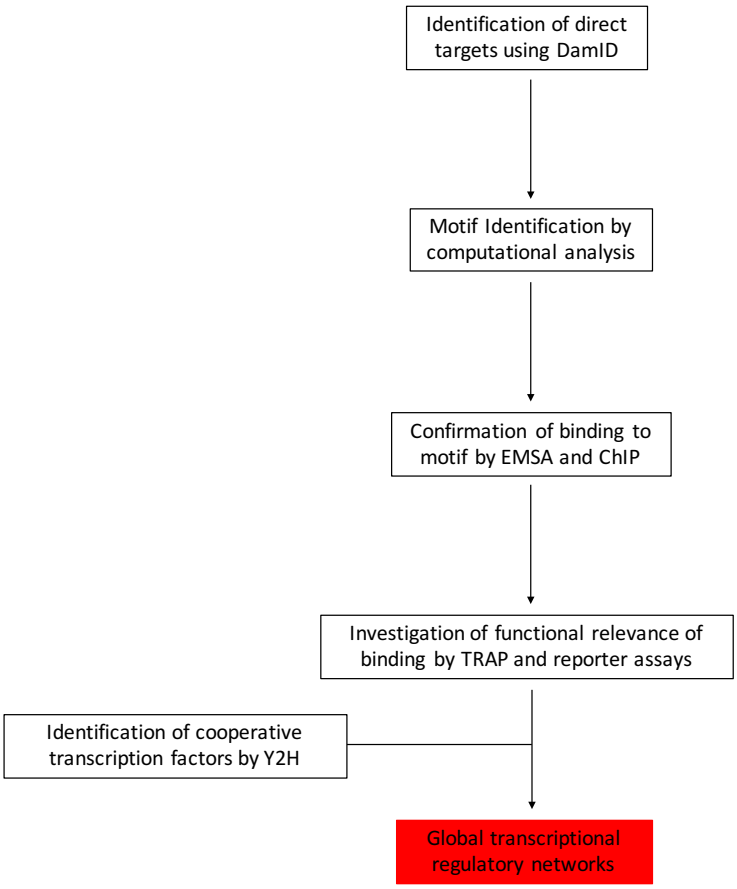


Figure 46. Approaches for deciphering the FoxP transcriptional regulatory network.
A combination of methods can be used to delineate the FoxPC transcriptional regulatory networks.

Targeted DNA adenine methyltransferase identification (targeted DamID or TaDa) was carried out, generating the *in vivo* binding profiles of FoxPC and FoxD. The TaDa experiments resulted in three high quality and high confidence binding datasets for each isoform. Even though DamID-seq is gaining popularity for profiling protein-DNA interactions, there are still some limitations to the technique. Low expressions of Dam are sufficient to randomly methylate easily-accessible DNA. This was, however, eliminated by normalising each experimental dataset to the control binding profiles. The experimental

and control binding profiles had a very low correlation coefficient, with FoxPC and FoxPD binding appearing as distinct peaks. Furthermore, fusion of Dam to FoxPC and FoxPD might affect their normal functions due to steric hindrance. The Dam methylase was therefore fused to the N-terminus of the two isoforms, to minimise the impact on the DNA-binding capacities and affinities. This however, may affect dimerisation of FoxP with other proteins required for their gene regulation activities. Overexpression of the *Dam-FoxPC* and *Dam-FoxPD* transgenes could also affect their locations and binding sites. For this reason, it was necessary to test the putative targets by other complementary methods, such as ChIP and reporter assays.

Furthermore, the binding profiles might not accurately represent the true binding of the transcription factors. Dam methylates GATC sites in the genome, which can spread to approximately 2 kb from the actual binding site (van Steensel and Henikoff, 2000). Therefore, assigning peaks to the nearest transcription unit might not be ideal. Association of a peak with a gene needs to be experimentally verified because it does not necessarily infer direct regulation of the genes.

The ChIP experiments confirmed the binding of FoxPC to regions close to the transcription start sites of some of the putative targets. Dual luciferase reporter assays not only showed that FoxPC binds to the promoter regions of these target genes, they also showed that FoxPC negatively regulates their activities. This is not surprising given that FoxP isoforms have been shown to have repressive functions in humans (Chokas et al., 2010, Roll et al., 2010). Mutating the critical DNA-binding residues in FoxPC abolished the repressive effect on the target genes, suggesting that FoxPC regulates its target genes through direct DNA binding. Although only a subset of targets was confirmed, we are a

step closer to determining the ultimate function of FoxPC and the networks it might regulate.

6.1.1 Identification of FoxPC and FoxPD targets

The gene target lists obtained from the DamID-seq experiments are conservative because stringent parameters were used for the binding analysis. Therefore, the numbers of genes uncovered are likely to be an underestimate. Both FoxPC and FoxPD bind extensively across the genome, on all chromosomes (Figure 47) suggesting that they play crucial roles in various processes. The distribution of chromosome binding is similar in FoxPC and FoxPD for chromosomes 2R, 2L, 3L, 3R and Y. FoxPD seems to regulate more genes on the X chromosome than FoxPC. This could explain why the second most highly similar motif to the FoxPD binding sequence in that of transformer2 (tra2) (Figure 14), tra2 being involved in sex determination in *Drosophila* by regulating doublesex (dsx) and fruitless expressions (Rideout et al., 2007, Lam et al., 2003).

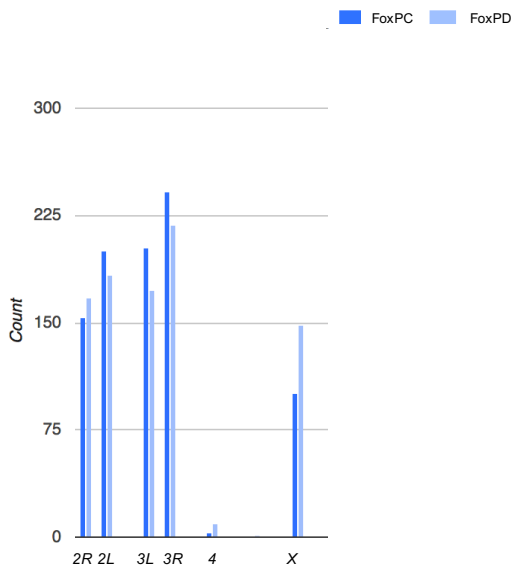


Figure 47. Chromosome distribution of FoxPC and FoxPD target genes.
Both isoforms bind extensively across the genomes. The x-axis shows the chromosome names and the y-axis shows the number of genes binding at each chromosome.

FoxPC and FoxPD had comparable overlaps between their binding sites (average Pearson correlation coefficient of 0.70 between the binding profiles and 37.5% overlap between the gene sets). The residues important in protein-DNA interactions in FoxPC and FoxPD are essentially similar, with only one change in residue from histidine to threonine in FoxPC (Santos et al., 2011). This could explain why they bind to common targets but are still able to regulate different ones. The differential expression of the two isoforms could partly explain their regulation of different genes. In neurons where both FoxPC and FoxPD are expressed, they could act in concert to regulate certain genes. The identification of unique binding targets is also indicative of unique functions of the two isoforms.

Examples of differentially bound genes include *cyclinD (cycD), Tyramine receptor II (TyrRII), SoxNeuro (SoxN*) and protein kinase A isoforms *(Pka-R2 and Pka-C1*) by FoxPC only and *Neuroligin 3 (Nlg3*) and *ether a go-go (eag)* by FoxPD.

cycD is one of the principal regulators of the cell cycle during development in *Drosophila.* It forms a complex with cyclin-dependent kinase 4 to regulate cellular growth and proliferation (Datar et al., 2000, Buttitta et al., 2007). *cycD* is also a known target of the group B *Sox* genes, *Dichaete* and *SoxN* (Aleksic et al., 2013). This is interesting as even though cycD is identified as a putative FoxPC target, it does not contain the putative binding motif (Figure 27). This could be due to FoxPC forming a heterodimer with Dichaete or another transcription factor before binding to the regulatory regions of *cycD*.

TyrRII (TAR3) is a biogenic amine-activated G protein-coupled receptor (GPCR) (Bayliss et al., 2013). Like other tyramine receptors, TyrRII plays a role in learning and memory

(Brembs et al., 2007). TyrRII is highly expressed in the mushroom bodies of *Drosophila* (El-Kholy et al., 2015), some divisions of which also contain high levels of FoxPC (Figure 27). It is conceivable that FoxPC represses TyrRII in the KC divisions where it is present, or that it acts to upregulate the expression of TyrRII. In this case, regulation of TyrRII by FoxP would be different from the repressive effect shown in this thesis. However, since mammalian FoxP orthologues are known to act as both activators and repressors (Takahashi et al., 2009, Chakraborty et al., 2017, Konopka et al., 2009), this too could be a plausible scenario.

SoxN is crucial for the development of embryonic central nervous system during the dorso-ventral patterning of the neuroectoderm and the formation of neuroblasts (Overton et al., 2002) and in axonal patterning in the later stages of embryogenesis (Girard et al., 2006).

cAMP-dependent protein kinase regulatory and catalytic subunits (Pka-R2 and Pka-C1) are unique FoxPC targets. This is a notable observation because PKA is not only involved in olfactory learning and memory (Gervasi et al., 2010), but is a key component of the machinery of slow synaptic transmission (Greengard, 2001). PKA is one of the components of the most connected 'nervous system development' pathway regulated by FoxPC (Figure 22). Due to its known functions, this putative target of FoxPC is an interesting avenue to explore.

The unique FoxPD target, Nlg3, is a cell adhesion molecule which plays a role in neuromuscular junction development and synaptic maturation (Xing et al., 2014).This is an interesting target, which could be followed up, given that mutations in Nlg3 in humans have been implicated in autism (Jamain et al., 2003).

Eag is a voltage-gated potassium channel, which regulates motor neuron excitability (Srinivasan et al., 2012) and is also involved in memory formation (Griffith et al., 1994). FoxPD could therefore also mediate olfactory learning in other neurons in the mushroom body.

The regulation of different target genes by two isoforms has also been shown in zebra finch, whereby the two isoforms of FoxP2 contribute differently to vocal learning by affecting different transcriptional networks (Burkett et al., 2018). The three isoforms of fruitless also display different binding patterns, while also showing overlap between regulated genes (Vernes, 2014, Neville et al., 2014).

Most of the genes identified through the DamID-seq screens are involved in neural processes, for example neuron development and neurogenesis, as determined by the gene ontology (GO) analysis. This is coherent with the most common biological processes associated with FoxP2 in the CNS of mammals (Konopka et al., 2009, Enard et al., 2009) and with FoxP in *Drosophila* (Mendoza et al., 2014, DasGupta et al., 2014). Another highly enriched GO category is signalling. This category has been reported as one of the overrepresented categories in FoxP2 RNA-seq data (Oswald et al., 2017). FoxPC and FoxPD datasets showed enrichment in genes involved in ion channel activity, for example nicotinic acetylcholine receptor subunits (*nAChRα6 and nAChRα3*), NMDA receptor (*Nmdar2*) and *Shal*. This is not surprising considering that ion channels are crucial in nervous system function. nAChRs are the predominant excitatory neurotransmitter receptors in the central nervous system of *Drosophila* (Fayyazuddin et al., 2006, Takemura et al., 2011) and NMDA receptors are involved in connectivity in circuit plasticity (Zhang et al., 2013). Shal channels regulate dendritic signal propagation

(Groschner et al., 2018). The control of ion channels, and signaling, might be a way by which FoxP governs neural and neuronal development.

6.1.2 Comparison to known targets

The DamID-seq datasets were assessed for the presence of any known targets of FoxP. Unfortunately, little is known about the regulation of genes by FoxP in *Drosophila* and the gene lists were therefore scanned for known FoxP1 or FoxP2 target genes. Several targets which have been identified as FoxP1/2 targets by ChIP-chip or RNA sequencing were present in my dataset. From the DamID-seq datasets, FoxPC and FoxPD both appear to bind to Semaphorins (especially *Sema2b*), which are involved in axon growth and guidance (Ayoob et al., 2006). It has been established that the human orthologue, FoxP2, regulates genes involved in neurite outgrowth (Vernes et al., 2011). Their ChIP-chip assays also identified semaphorins as key targets of FoxP2. One the most highly bound FoxPC target gene is *cyclinD*. This is in accord with previous studies that have linked FoxP1 to the regulation of the cell cycle (Leishman et al., 2013).

Ph-p which was also found as a highly bound FoxPC and FoxPD target is a subunit of the Polycomb repressive complex. This complex has been found to bind to and negatively regulate the expression of FoxP2 in humans (Rastegar-Pouyani et al., 2016), creating the potential for feedback regulation between these two genes. Lrrk2 which is involved in wingless/Int (Wnt) and Notch signalling (Berwick and Harvey, 2013, Imai et al., 2015), was also found to be a FoxP2 target by RNA-sequencing. Mnt (or Mad) and Shal (K_v4) have been identified as two of 142 FoxP2 targets in humans (Spiteri et al., 2007). Shal is a highly-bound target of FoxPC and FoxPD. The negative regulatory effect of FoxP on K_v4 channels was also shown in basal ganglia (specifically the external globus pallidus), where neurons that express FoxP2 show reduced K_v4 currents (Hernandez et al., 2015). This

finding is noteworthy, because (*i*) the mushroom body can to a certain extent be considered functionally analogous to the basal ganglia (Strausfeld and Hirth, 2013, Cohn et al., 2015) and (*ii*) the basal ganglia are critical for decision-making in mammals (Barron et al., 2015, Ding and Gold, 2013).

The high degree of overlap between the FoxP targets identified in *Drosophila* and the known targets of FoxP2 in humans suggests that the targets and roles of FoxP proteins, for example in nervous system development and signalling, and perhaps higher order behaviour, are highly conserved throughout evolution.

6.1.3 Linking FoxP targets to known behavioural phenotypes

FoxP has been implicated in several behaviours in *Drosophila,* including courtship song production (Lawton et al., 2014), operant self-learning (Mendoza et al., 2014) and perceptual decision-making (DasGupta et al., 2014, Groschner et al., 2018). The role of FoxP in these processes can be explained by the regulatory effect of FoxP on specific transcriptional networks described here.

The DamID-seq datasets obtained showed that FoxP potentially binds to and regulates both fruitless (*fru*) and doublesex (*dsx*), which are known to control courtship song production (Goodwin et al., 2000, Villella and Hall, 1996). This could suggest a more direct role of FoxP, rather than the previously proposed indirect consequences of motor control in courtship (Lawton et al., 2014).

FoxP was also implicated in operant conditioning and learning. $FoxP^{5\text{-}SZ\text{-}3955}$ mutants showed a decrease in operant self-learning and this effect was attributed to a decrease in

the FoxPC isoform only (Mendoza et al., 2014). The self-learning mechanism thought to be dependent on Protein Kinase A (PKA) and Protein Kinase C (PKC) (Brembs and Plendl, 2008, Colomb and Brembs, 2016). My DamID-seq datasets showed that PKA is a putative target exclusive to FoxPC (as mentioned above). FoxPC could therefore act in the formation of certain forms of memory via the cAMP/PKA pathway.

FoxP has also been implicated in the integration of sensory information required for perceptual decisions (DasGupta et al., 2014, Groschner et al., 2018). The voltage-gated potassium channel Shal (K_V4) shapes the biophysical properties of decision-relevant neurons ($\alpha\beta_c$ KCs) so that synaptic inputs can be accumulated over time. DamID-seq identified the Shal locus as a direct target of FoxP. None of the other K_v channels (Shaker, Shab, Shaw and Shawl) are bound by FoxPC (Figure 48), confirming an inference drawn from previous functional studies (Groschner et al., 2018).

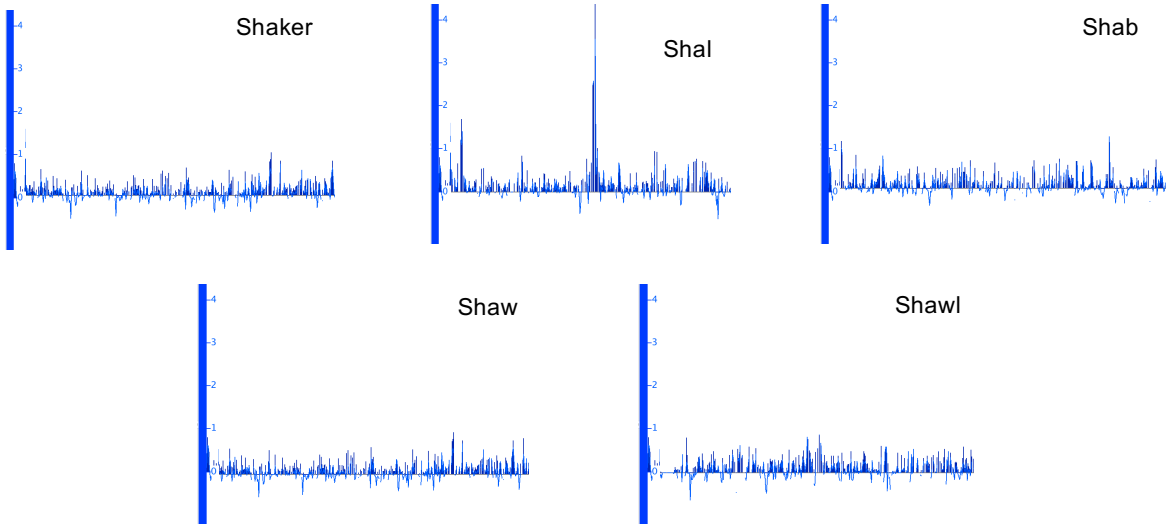


Figure 48. Shal is the only voltage-gated potassium channel bound by FoxPC.
FoxPC does not significantly bind to the Sh, Shab, Shaw and Shawl loci. The y-axis shows the log_2 fold change obtained from the DamID-seq analysis.

6.1.4 FoxPC and signalling pathways

FoxPC's involvement in nervous system development by regulating the Wnt signaling pathway is highly consistent with the role of FoxP2 in the regulation of signaling pathways (Vernes et al., 2011). The Wnt pathway controls numerous cellular processes during brain development, including neuronal connectivity and neuromuscular junction formation (Logan and Nusse, 2004, Packard et al., 2002). The Wnt signaling pathway consists of several components, including the Wg ligand, Frizzled receptor (fz) which recruits the disheveled (Dsh) protein to form a complex together with Armadillo (Arm), Axin, APC and Shaggy (sgg) (Waghmare and Page-McCaw, 2018). This receptor complex ultimately forms a transcription complex in the nucleus, together with the transcription factor, Pangolin (Pan), resulting in the transcription of a plethora of genes involved in developmental and differentiation processes.

By binding to fz, sgg and Pan, FoxPC regulates the major components of this developmentally important pathway. FoxPC also regulates other modulators of this pathway, including SoxNeuro (SoxN), CycD, PKA and Dally (Cadigan and Peifer, 2009, Chao et al., 2007, Kikuchi et al., 2006). This suggests that the regulatory action of FoxPC on both the components and the modulators might act as a safeguard mechanism to keep the Wnt signalling pathway in check. This would mean that, in those cases, FoxPC mostly acts during development, consistent with the role of FoxP2 in the developmental control of the central nervous system in mammals (Takahashi et al., 2009).

6.2 Transcription regulation by FoxP

Consensus sequences, or motifs, were identified by scanning the FoxPC and FoxPD binding intervals for enriched sequences using MEME. The second most enriched and

significant FoxPC motif (A(A/G)CAACAA), found in about 71.54% of target genes, was identified by HT-SELEX (Nitta et al., 2015). The same motif was found in motif enrichment analyses of FoxPD bound regions, but with a much lower E-value. Another putative motif is ((G/A)CAG(C/A)AG(C/A)GC). This motif was not uncovered in the FoxPC datasets and could thus represent a unique FoxPD binding motif.

The FoxPC and FoxPD motifs are sometimes found in the same genes, for example in *osa*, *ph-p* and *Mnt* (Figure 49), while some of the genes contain binding sites for either isoform (*PKA* only contains binding sites for FoxPC and *Nlg3* only harbours FoxPD binding motifs). This suggests that FoxPC and FoxPD can recognise the same targets or different promoter regions. It also suggests that FoxPC and FoxPD might dimerise in order to regulate certain genes. This is a known scenario in humans, where FoxP1, FoxP2 and FoxP4 dimerise to differentially regulate FoxP2 targets (Sin et al., 2015).

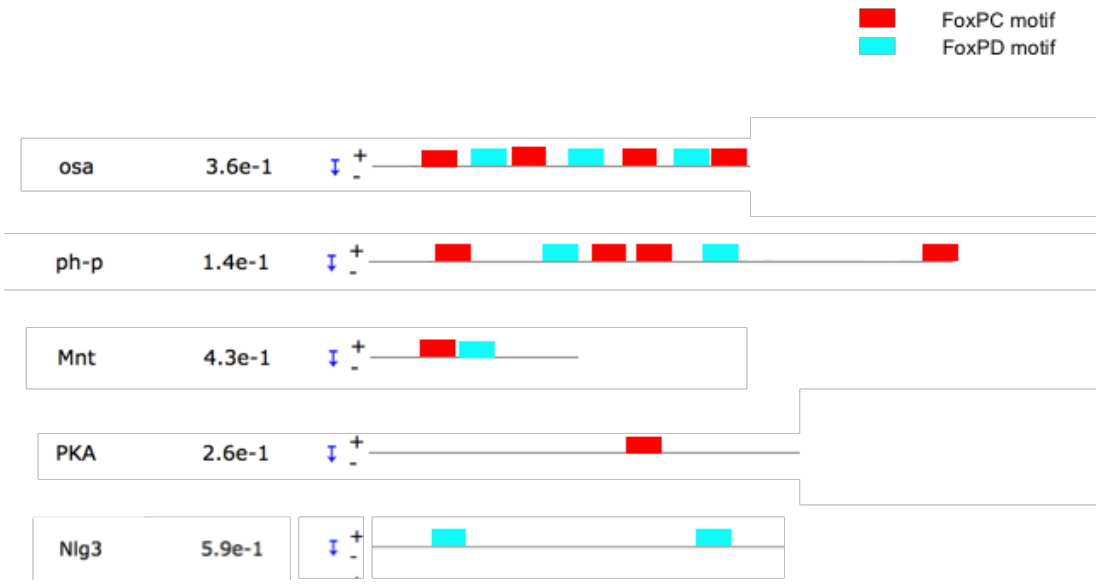


Figure 49. Presence and locations of FoxPC and FoxPD motifs on target genes.
Regions close to the transcription start sites of the genes were scanned for the presence of the FoxPC motif (A(A/G)CAACAA) (red box) and the FoxPD motif ((G/A)CAG(C/A)AG(C/A)GC) (blue box). Osa, ph-p and Mnt harbour both motifs, while PKA and Nlg3 only contain the FoxPC and FoxPD motif, respectively.

The motifs identified are highly similar to other forkhead protein motifs, identified by i-cisTarget. The top 12 cis-regulatory modules uncovered by analysis of the FoxP binding intervals belong to forkhead proteins, for example sloppy paired 1(slp1), foxo, and croc. This is not surprising because the shared forkhead domain is highly conserved RYAAAYA (where R is either adenine (A) or guanine (G) and Y can be cytosine (C) or thymine (T)) (Chen et al., 2016). The slight differences in their binding motifs could indicate how they are differentially expressed in an organism and how they regulate different target genes to control specific biological processes.

Given the strong indication that the DamID-sequencing data had identified a *bona fide* FoxPC motif, this was tested using electrophoretic mobility assays and luciferase reporter assays (Figures 35 and 36). The motif was mutated in the promoter regions of three of the most highly bound FoxPC targets (*osa, Trp γ* and *Shal*). FoxPC's binding and repressive effect on those targets was abolished when the consensus binding site was mutated, strongly confirming the A(A/G)CAACAA sequence as the FoxPC binding motif.

6.2.1 FoxPC's mechanisms of action

Regulation at intronic regions

The binding profiles of FoxPC and FoxPD show a large number of peaks (about 50%) in the intronic regions close to and downstream of the transcription start sites of genes, typically in the first intron (Figures 12 and 13). This could either be due to a larger number of GATC sites in those regions, compared to the promoter regions upstream of the transcription start sites, or to binding to large introns containing cis-regulatory modules or silencers. While it is possible that this could be an artefact due to long intronic sequences potentially being bound more frequently in general, it is possible that binding to intronic

regions could also be a feature of transcriptional control by FoxPC and FoxPD. The latter scenario was tested by comparing the peak locations to known enhancer regions obtained from the REDfly database, noting that this is only an informative test as the REDfly database does not contain the full genome-wide dataset of enhancer regions. The analysis showed that 28% of the peaks overlapped with known enhancers, with FoxPC and FoxPD binding to the enhancer regions of some of their target genes, for example ftz-f1, osa, cyclins rpr. This confirms that FoxPC and FoxPD regulate transcription by binding to classical enhancer regions close to the transcription start sites. A similar pattern exists in humans, whereby the first introns of most genes play critical roles in transcriptional control (Li et al., 2012). By binding to intron-located silencers, FoxPC could physically block the transcriptional elongation process or interfere with the assembly of the transcriptional complex (Ogbourne and Antalis, 1998).

Regulation at core promoter regions

The meta gene profiles (Figure 13) show two major binding locations: close to and upstream and close to and downstream of the transcription start site. This is not surprising as the putative FoxPC and FoxPD motifs are highly enriched close to the transcription start sites of their target genes. The occurrence of binding intervals very close to the transcription start sites (< 80 bp) suggests that FoxPC and FoxPD could interact directly with core promoter elements. They could also bind to other transcription factors at regulatory modules further away from the transcription start sites and be brought close to the transcription start sites. In the latter case, when the transcription factor partner is capable of bending DNA, for example high-mobility-group domain proteins, such as SoxN and Dichaete (Kamachi and Kondoh, 2013), FoxPC or FoxPD will be brought close to the transcriptional machinery by DNA looping. By binding to proximal promoter

elements, FoxP could prevent the basal transcriptional complex formation and hence inhibit the initiation of transcription.

Hence, FoxP could regulate transcription by two distinct mechanisms. In addition to binding to the core promoter regions (~80 bp from the transcription start site), FoxPC and FoxPD also bind to elements (silencers) upstream and downstream of transcription start sites, which contain their putative motif. These elements may act independently or synergistically with the core promoter, closer to the transcription start sites, to prevent transcription. Repressive elements have been found to occur between 500 bp and 1000 bp upstream of the transcription start sites (Cooper et al., 2006), correlating perfectly with the meta gene binding profiles (Figure 13). These two mechanisms could explain the decrease in expressions of the target genes observed by TRAP (Figure 31B). By binding to its target genes, FoxPC can decrease promoter activity or inhibit the transcript elongation.

A small number of peaks fall within the intergenic regions, which have no classical regulatory function. However, accessible DNA can act as a low-affinity reservoir for transcription factors, maintaining their optimum amount available in the nucleus (Lin and Riggs, 1975, MacArthur et al., 2009). This is especially important if the transcription factors act in a combinatorial manner. In this way, those binding sites are not directly involved in gene transcription but nevertheless important for fine-tuning gene expression.

Only very few peaks are present in the 5' and 3' UTR regions (5.78 % and 2.79%, respectively). Regulation at those sequences occurs post-transcription by affecting mRNA stability, localisation and translational efficiency (van der Velden and Thomas, 1999, Thomas et al., 2012, Mignone et al., 2002).

6.3 FoxPC regulates its targets as part of a complex

As discussed above, FoxPC likely plays a crucial role in nervous system development. As such, gene regulation by FoxPC must be tightly controlled and specific, such that its target genes are expressed or repressed at the appropriate time and in the right cellular compartments. Technically, since motifs are relatively short, transcription factors could bind to thousands of sites in the genome. This is limited by concerted action of multiple transcription factors at cis-regulatory modules.

Transcription factors participate in transcriptional regulatory networks or programmes, which include the transcription factors, DNA binding motifs and other cofactors. The networks are defined by combinations of network motifs, which are small networks with distinct properties (Shen-Orr et al., 2002). Some of these are shown in Figure 50: in the feed-forward loop motif (Figure 50A), one transcription factor (TF1) regulates the expression of another (TF2), with both of them ultimately regulating the expression of a target one. TF1 acts as a master regulator and hence any change in its expression levels can lead to massive effects in gene regulation of the targets. The single input motif consists of only one transcription factor regulating different target genes (Figure 50B). This ensures that genes are temporally and spatially expressed together, when required. The multiple input motif (Figure 50C) requires several transcription factors which bind together to the promoter or enhancer regions of the same target genes. The subfamily of the multiple input motifs -dense overlapping regulons (Figure 50D)- consist of different regulators that combinatorially regulate a set of targets, with each target being regulated by different combinations of transcription factor combinations.

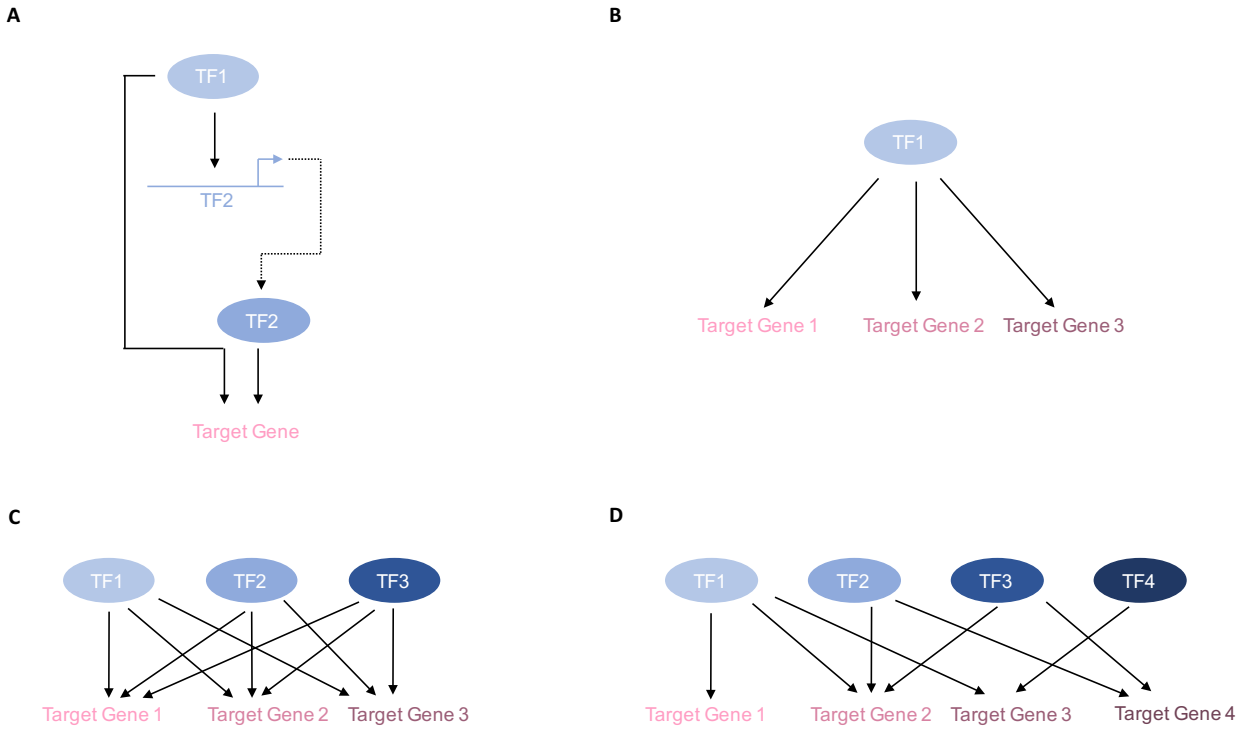


Figure 50. Network motifs of transcriptional regulatory networks.

*Gene expression is regulated by combinations of motifs, some of which are shown. **A**, feedforward loop. **B**, single input motif. **C**, multiple input motif. **D**, dense overlapping region. Solid black arrows transcription factor- DNA interactions or transcription factor regulation (in **A**). Dotted arrows indicate translation of proteins.*

It was hypothesised that, as with most transcription factors, the binding of FoxPC alone might not be sufficient to exert tight regulatory control of the target genes. Hence, other transcription factors might be required to bind in the same cis-regulatory elements and function synergistically. Binding to other transcription factors is possible due to the modular structure of FoxPC. In addition to the forkhead DNA-binding domain and the polyglutamine regulatory domain, two other functional domains—a zinc finger domain and a leucine zipper motif—are present. In humans, both DNA-binding by the forkhead domain and dimerisation via the leucine zipper and zinc finger domain are required for gene repression (Wang et al., 2003). Based on my findings, a similar mechanism is highly likely in *Drosophila* as an additional layer of regulation of gene expression.

Comparing the FoxPC targets with those of known transcription factors resulted in highly significant overlaps between binding patterns of FoxPC and other transcription factors involved in nervous system development, for example Hb, SoxN and Dichaete. This suggests that FoxPC might interact with those transcription factors during nervous system development. By looking at the DamID-seq data, several transcription factors were identified as bound FoxPC targets (Figure 51), including SoxN, ftz-f1, CHES-1-like and Sox14. This could be an example of a feedforward loop (Figure 50A), where FoxPC regulates the expression of another transcription factor, such as SoxN, and then interacts with it to regulate the expression of common target genes. The interaction network of SoxN has been analysed by DamID and ChIP (Ferrero et al., 2014). SoxN directly binds to numerous genes crucial in nervous system development, including *Dichaete, lola, beat-1a, Insc, gcm* and *Sema-1a*. *Beat-a1, Insc* and *Sema-1* were also identified in the FoxPC DamID-seq dataset.

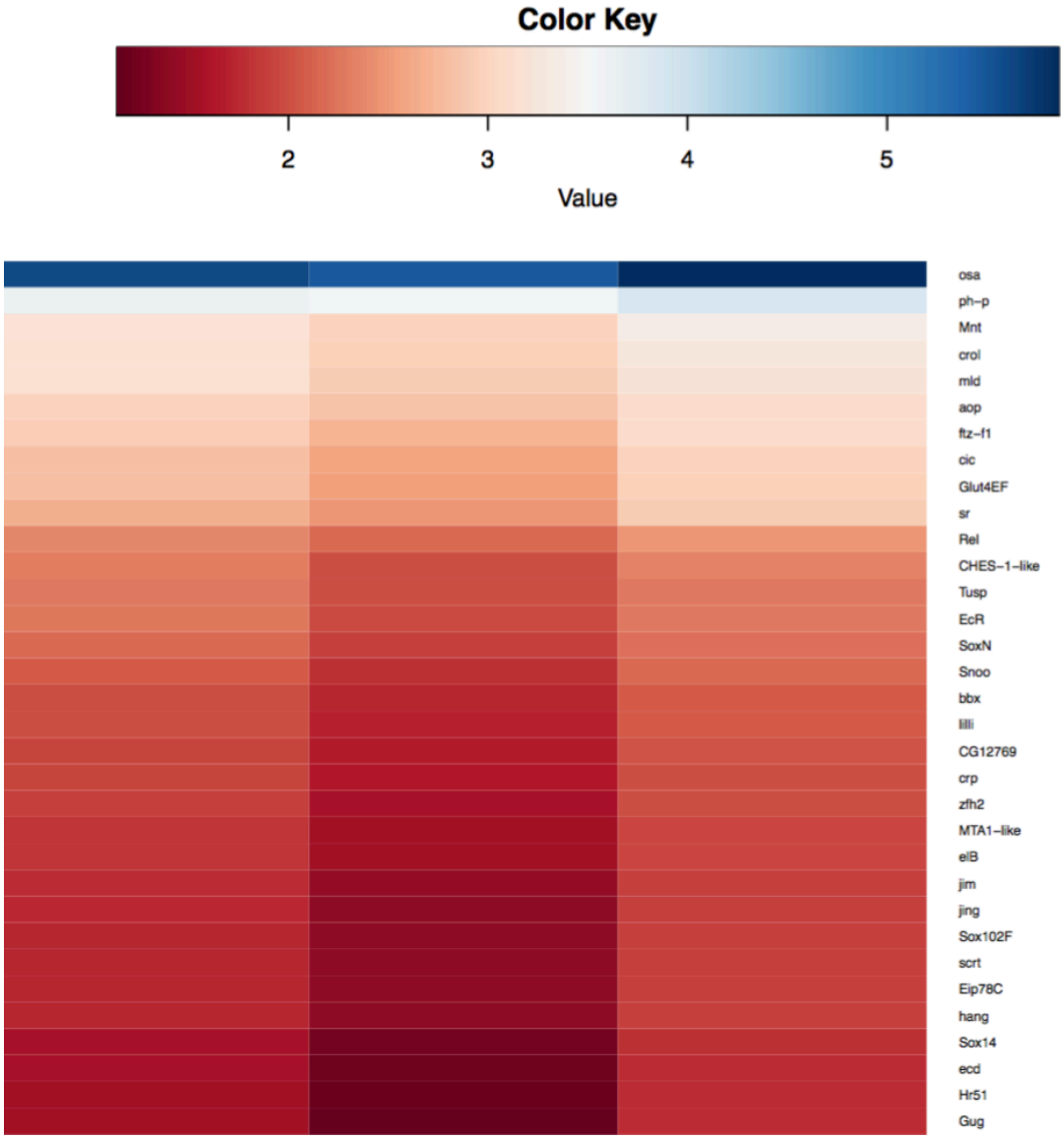


Figure 51. FoxPC binds to numerous transcription factors.

Classification of the FoxPC target genes revealed several transcription factors. The color key represents the log_2 fold change from the DamID-seq analysis.

Comparing the FoxPC motif to known motifs by FIMO showed that it is highly similar to the Hb and lola motifs. This does not prove any direct interactions, but could point towards putative cofactors that should be studied in more detail. The three most similar motifs belong to zinc finger transcription factors, Adult enhancer factor 1 (Aef1), longitudinals lacking (Lola) and hunchback (hb). Aef1, Lola and hb could form dimers with FoxPC, via their zinc finger domains, to affect its role in gene regulation and nervous

system development. This is interesting as both mammalian orthologues of SoxN and hb, have been found to interact with FoxP1, FoxP2 and FoxP3 (Bond et al., 2016, Pan et al., 2009).

Taken together, these results show that FoxPC acts as a transcriptional regulator, possibly together with other transcription factors, for example SoxN, Hb and lola to regulate the expression of genes in various aspects of nervous system development, including controlling the neuronal fate (for example by *ftz-f1*), neurogenesis (by *Insc*) and eye development (by *Eys*).

Another way to test the transcription network motif is by examining the direct interaction between FoxPC and other transcription factors by yeast two-hybrid assays and co-immunoprecipitation. 70 known transcription factors were chosen, based on their dimerisation domains and their known interactions with mammalian FoxP. Three of the transcription factor, glial cells missing (gcm), Sp1-like factor for pairing sensitive-silencing (Spps) and Dichaete (D), bound strongly to FoxPC. Gcm and Spps are zinc finger transcription factors, which likely heterodimerise with FoxPC at its N-terminal zinc finger domain. Unfortunately, the Y2H assays did not yield convincing interactions between FoxPC and SoxN and FoxPC and Hb, probably due to the high rate of false negatives associated with the technique (Bruckner et al., 2009). Another plausible scenario is that FoxPC does not directly interact with SoxN and Hb, but acts as a repressor of SoxN and Hb binding, hence preventing the activation of their target genes.

Gcm is the primary regulator of glial versus neuronal fate determination by activating glial genes and repressing neuronal genes (Cattenoz et al., 2016, Hosoya et al., 1995, Jones et

al., 1995). Gcm plays a role in events post-gliogenesis by maintaining differentiated cells by binding to chromatin remodelling factors and cell cycle regulators (Egger et al., 2002). Gcm levels therefore need to be tightly controlled to regulate its activating and repressive activities (Jones, 2005).

The Y2H and co-immunoprecipitation results are interesting as FoxPC binds to the gcm CRMs (as determined by the i-cisTarget analysis). This suggests that FoxPC could bind to the *gcm* CRM to repress gcm expression and could also bind to the gcm protein to regulate its activator activity. The same scenario is seen between slp1/Foxg1 and GCM (Mondal et al., 2007). Mondal et al., showed that Slp1 inhibits GCM activity by protein-protein interactions and not by competing with its transcription binding sites. If FoxPC is expressed in the same cells as gcm during gliogenesis, it could control the expression levels of gcm to ensure proper neuron and glial differentiation.

During the later stages of development, FoxPC and gcm could cooperate to repress the expression of neuron-specific genes. Previous research found that gcm causes the activation and repression of transcription of its downstream targets, which are involved in various processes (Table 3) (Egger et al., 2002). Strikingly, many of the downregulated genes were found to be FoxPC targets, for example *lola, ey, sgg, Sox14, ph-p, CycE, CadN* and *ftz*. This strongly confirms that FoxPC and gcm cooperate to regulate the expression of genes involved in various processes required to maintain neuronal cells, including axogenesis, cellular proliferation and regulation of gene expression. This was confirmed by ChIP-qPCR and luciferase reporter assays which revealed that the repressive effects of FoxPC on *ph-p* and *ftz-f1* were reinforced when both FoxPC and gcm were present. In some cases, the target genes, for example *Cyclin D*, do not contain the putative FoxPC

binding domain. One possible explanation for their identification as target genes could be due to FoxPC binding to gcm, which then binds to those target genes. This brings the FoxPC-fused Dam in close proximity to the target genes, resulting in methylation identified by the DamID technique.

Table 3. Known gcm targets.

Functional Class	Examples of downregulated genes
Transcription factor	Slp1, lola, ey, Ubx, Mef2
Kinase/Phosphatase	Sgg, Egfr, rdgC, bsk
DNA binding/ chromatin binding	Sox14, ph-p, Dichaete
Cell cycle regulation	CycE
Cell adhesion	Ten-m, CadN

Spps is a polycomb group (PcG) recruiter, which brings PcG proteins to the Polycomb response elements (PREs). PcG proteins are important regulators of gene expression (Kassis et al., 2017), maintaining key developmental genes in their silenced states through methylation. There is previous evidence that Fox transcription factors regulate the recruitment to and dissociation of PcG proteins from their target genes (Bracken and Helin, 2009). Some of the targets of PcG proteins are *SoxNeuro, Lim3* and *wingless*. The interaction between FoxPC and Spps at the target genes brings the PcG proteins close to the genes to repress their transcription. The ChIP and luciferase reporter experiments showed that the osa promoter is bound by and affected by FoxPC and Spps. This is not surprising as osa acts as part of the Brahma-associated proteins (BAP) complex which is

required for the antagonism of PcG proteins (Chalkley et al., 2008). Therefore, the repression of osa expression allows PcG proteins to act on their target genes. FoxPC could therefore help fine-tune the regulatory effect on activation and repression of developmental genes. Voltage-gated potassium (K_v) channel genes are known targets of PcG regulation (Ryland et al., 2015), which explains why the Shal promoter is under negative regulation of FoxPC and Spps. The knockdown of both FoxPC and Spps causes an increase in Shal currents, suggesting that these transcription factors recruit PcG protein to regulate Shal expression.

The interaction between FoxPC and Dichaete reaffirms the role of FoxPC in several pathways in development. Dichaete is a member of the high-mobility group (HMG) transcription factors. It is crucial in embryonic segmentation, nervous system development and hindgut formation (Nambu and Nambu, 1996, Aleksic et al., 2013, Overton et al., 2002, Sanchez-Soriano and Russell, 2000). Dichaete is very similar to SoxN, making the interaction between Dichaete and FoxPC highly plausible. FoxPC and Dichaete bind to common characterised regulatory elements catalogued in RedFly.

Both FoxPC and Dichaete bind to transcriptional regulators important in central nervous system development, for example SoxN, Kr and lola. The high similarity of the FoxPC putative motif to the lola and hunchback (hb) motifs, which are known targets of Dichaete, strongly suggests that there is similarity between their regulatory targets. FoxPC, together with Dichaete, could therefore play a role in differentiating neural cells by controlling the switch between neuroblast self-renewal to cell-cycle withdrawal and differentiation or apoptosis (Maurange et al., 2008). Consistent with the hypothesis that FoxPC might regulate the decision to exit the cycle, two of the targets of FoxPC and Dichaete are rpr

and Cyclin E. The enrichment of the GAGA-binding factor, Trl, motif in both the FoxPC (Figure 14A) and Dichaete binding intervals (Aleksic et al., 2013) suggest a regulatory role early in development.

FoxPC and Dichaete can also regulate signalling pathways during nervous system development, for example by targeting fz and nkd in the Wnt pathway and Su(H) and heph in the Notch signalling pathway. This suggests that FoxP may act as a regulatory hub, together with Dichaete, to affect genes involved in different processes. The binding of FoxPC to both Dichaete, and potentially Trl target genes, and Spps is important in the maintenance of homeotic gene expression. Trl and Spps are members of the trithorax and Polycomb groups of genes, which give rise to the restricted pattern of genes to ensure proper development (Bejarano and Busturia, 2004). Therefore, FoxPC might play a much bigger role in development than currently thought.

It can be concluded that FoxPC regulates some of its targets by the dense overlapping region motif. FoxPC can regulate targets on its own, for example in the case of crol and it can also work with different combinations of transcription factors, including gcm, Spps and Dichaete, to regulate different subsets of genes. From MAST analysis of the transcription factor binding sites (Figure 43), it was often observed that FoxPC, gcm and Spps binding sites were close to each other. However, in some cases where Dichaete binding sites are present, for example in *ftz-f1* and *Trpγ*, the FoxPC, Dichaete and any other transcription factor binding sites, were slightly further away from each other. This could be explained by the DNA-bending capacities of Dichaete (Ma et al., 1998). Upon DNA binding, Dichaete could bend the neighbouring stretch of DNA, bringing other transcription factor complexes close to the transcription start sites.

Since some of its partners have both activating and repressing activities, it is not simply the timing of expression of transcription factors that affects the temporal nature of gene regulatory networks. The combinatorial interplay between several transcription factors, each having different expression windows and their time-dependent occupancy at regulatory regions are crucial parameters in gene expression.

6.4 Conclusions

From the DamID-seq experiments, targets of FoxPC and FoxPD have been identified. The two isoforms of FoxP regulate some common target genes, which are involved in nervous system development. The two isoforms also regulate the expression of unique genes, which suggests that they possibly participate in different pathways, for example in learning and integration of sensory information during decision-making.

FoxPC binds to genes involved in crucial processes during development, for example neurogenesis and gliogenesis. To achieve its functions, FoxPC likely interacts with different interaction partners, including gcm, Spps and Dichaete. These transcription factors have established roles in the nervous system development.

6.5 Future directions

Although the DamID-seq experiments seem to be robust, with many known target genes of FoxP in *Drosophila* and FoxP2 in humans identified in the datasets, the conclusions could be strengthened by mutating the FoxPC DNA binding domain and comparing the binding profiles to those obtained using wild-type FoxPC.

Even though the DamID-seq gives a clear idea of the target genes of FoxPC, the regulatory effects of FoxPC binding to those genes remain unknown. Comparison of RNA-sequencing from wild-type and FoxPC mutant flies might give an indication of whether FoxPC has both activating and repressive effects, like its mammalian counterpart FoxP2.

FoxPC is known to be involved in the temporal integration of sensory information by the regulation of its target Shal. This process likely involves other target genes regulated by FoxPC. The finding that decision-making of *Drosophila* can be quantified within the same conceptual and mathematical framework as human decision-making (Groschner et al., 2018) raises hopes for the genetic dissection of circuits underpinning cognition. This may ultimately lead to an understanding of how mobile animals collect and integrate sensory evidence and transform this information into complex, adaptive behaviours.

References

AGUILAR, D. & OLIVA, B. 2008. Topological comparison of methods for predicting transcriptional cooperativity in yeast. *BMC Genomics,* 9**,** 137.

ALEKSIC, J., FERRERO, E., FISCHER, B., SHEN, S. P. & RUSSELL, S. 2013. The role of Dichaete in transcriptional regulation during Drosophila embryonic development. *BMC Genomics,* 14**,** 861.

ASO, Y., HATTORI, D., YU, Y., JOHNSTON, R. M., IYER, N. A., NGO, T. T., DIONNE, H., ABBOTT, L. F., AXEL, R., TANIMOTO, H. & RUBIN, G. M. 2014. The neuronal architecture of the mushroom body provides a logic for associative learning. *Elife,* 3**,** e04577.

AUGHEY, G. N., ESTACIO GOMEZ, A., THOMSON, J., YIN, H. & SOUTHALL, T. D. 2018. CATaDa reveals global remodelling of chromatin accessibility during stem cell differentiation in vivo. *Elife,* 7.

AWASAKI, T., SAITO, M., SONE, M., SUZUKI, E., SAKAI, R., ITO, K. & HAMA, C. 2000. The Drosophila trio plays an essential role in patterning of axons by regulating their directional extension. *Neuron,* 26**,** 119-31.

AYOOB, J. C., TERMAN, J. R. & KOLODKIN, A. L. 2006. Drosophila Plexin B is a Sema-2a receptor required for axon guidance. *Development,* 133**,** 2125-35.

BAILEY, T. L., BODEN, M., BUSKE, F. A., FRITH, M., GRANT, C. E., CLEMENTI, L., REN, J., LI, W. W. & NOBLE, W. S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res,* 37**,** W202-8.

BANNISTER, A. J. & KOUZARIDES, T. 2011. Regulation of chromatin by histone modifications. *Cell Res,* 21**,** 381-95.

BARRON, A. B., GURNEY, K. N., MEAH, L. F., VASILAKI, E. & MARSHALL, J. A. 2015. Decision-making and action selection in insects: inspiration from vertebrate-based theories. *Front Behav Neurosci,* 9**,** 216.

BASHIRULLAH, A., COOPERSTOCK, R. L. & LIPSHITZ, H. D. 2001. Spatial and temporal control of RNA stability. *Proc Natl Acad Sci U S A,* 98**,** 7025-8.

BAYLISS, A., ROSELLI, G. & EVANS, P. D. 2013. A comparison of the signalling properties of two tyramine receptors from Drosophila. *J Neurochem,* 125**,** 37-48.

BEJARANO, F. & BUSTURIA, A. 2004. Function of the Trithorax-like gene during Drosophila development. *Dev Biol,* 268**,** 327-41.

BERNSTEIN, B. E., HUMPHREY, E. L., LIU, C. L. & SCHREIBER, S. L. 2004. The use of chromatin immunoprecipitation assays in genome-wide analyses of histone modifications. *Methods Enzymol,* 376**,** 349-60.

BERWICK, D. C. & HARVEY, K. 2013. LRRK2: an eminence grise of Wnt-mediated neurogenesis? *Front Cell Neurosci,* 7**,** 82.

BLANCHARD, F. J., COLLINS, B., CYRAN, S. A., HANCOCK, D. H., TAYLOR, M. V. & BLAU, J. 2010. The transcription factor Mef2 is required for normal circadian behavior in Drosophila. *J Neurosci,* 30**,** 5855-65.

BOND, J., DOMASCHENZ, R., ROMAN-TRUFERO, M., SABBATTINI, P., FERREIROS-VIDAL, I., GERRARD, G., ASNAFI, V., MACINTYRE, E., MERKENSCHLAGER, M. & DILLON, N. 2016. Direct interaction of Ikaros and Foxp1 modulates expression of the G protein-coupled receptor G2A in B-lymphocytes and acute lymphoblastic leukemia. *Oncotarget,* 7**,** 65923-65936.

BRACKEN, A. P. & HELIN, K. 2009. Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nat Rev Cancer,* 9**,** 773-84.

BREMBS, B., CHRISTIANSEN, F., PFLUGER, H. J. & DUCH, C. 2007. Flight initiation and maintenance deficits in flies with genetically altered biogenic amine levels. *J Neurosci,* 27**,** 11122-31.
BREMBS, B. & PLENDL, W. 2008. Double dissociation of PKC and AC manipulations on operant and classical learning in Drosophila. *Curr Biol,* 18**,** 1168-71.
BROWN, J. L., GRAU, D. J., DEVIDO, S. K. & KASSIS, J. A. 2005. An Sp1/KLF binding site is important for the activity of a Polycomb group response element from the Drosophila engrailed gene. *Nucleic Acids Res,* 33**,** 5181-9.
BRUCKNER, A., POLGE, C., LENTZE, N., AUERBACH, D. & SCHLATTNER, U. 2009. Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci,* 10**,** 2763-88.
BUCK, M. J. & LIEB, J. D. 2004. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics,* 83**,** 349-60.
BURKETT, Z. D., DAY, N. F., KIMBALL, T. H., AAMODT, C. M., HESTON, J. B., HILLIARD, A. T., XIAO, X. & WHITE, S. A. 2018. FoxP2 isoforms delineate spatiotemporal transcriptional networks for vocal learning in the zebra finch. *Elife,* 7.
BUTLER, J. E. & KADONAGA, J. T. 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev,* 16**,** 2583-92.
BUTTITTA, L. A., KATZAROFF, A. J., PEREZ, C. L., DE LA CRUZ, A. & EDGAR, B. A. 2007. A double-assurance mechanism controls cell cycle exit upon terminal differentiation in Drosophila. *Dev Cell,* 12**,** 631-43.
CADIGAN, K. M. & PEIFER, M. 2009. Wnt signaling from development to disease: insights from model systems. *Cold Spring Harb Perspect Biol,* 1**,** a002881.
CATTENOZ, P. B., POPKOVA, A., SOUTHALL, T. D., AIELLO, G., BRAND, A. H. & GIANGRANDE, A. 2016. Functional Conservation of the Glide/Gcm Regulatory Network Controlling Glia, Hemocyte, and Tendon Cell Differentiation in Drosophila. *Genetics,* 202**,** 191-219.
CERTEL, S. J., CLYNE, P. J., CARLSON, J. R. & JOHNSON, W. A. 2000. Regulation of central neuron synaptic targeting by the Drosophila POU protein, Acj6. *Development,* 127**,** 2395-405.
CHAKRABORTY, S., PANDA, A. K., BOSE, S., ROY, D., KAJAL, K., GUHA, D. & SA, G. 2017. Transcriptional regulation of FOXP3 requires integrated activation of both promoter and CNS regions in tumor-induced CD8(+) Treg cells. *Sci Rep,* 7**,** 1628.
CHALKLEY, G. E., MOSHKIN, Y. M., LANGENBERG, K., BEZSTAROSTI, K., BLASTYAK, A., GYURKOVICS, H., DEMMERS, J. A. & VERRIJZER, C. P. 2008. The transcriptional coactivator SAYP is a trithorax group signature subunit of the PBAP chromatin remodeling complex. *Mol Cell Biol,* 28**,** 2920-9.
CHAO, A. T., JONES, W. M. & BEJSOVEC, A. 2007. The HMG-box transcription factor SoxNeuro acts with Tcf to control Wg/Wnt signaling activity. *Development,* 134**,** 989-97.
CHEN, L., GLOVER, J. N., HOGAN, P. G., RAO, A. & HARRISON, S. C. 1998. Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature,* 392**,** 42-8.
CHEN, X., CHEUNG, S. T., SO, S., FAN, S. T., BARRY, C., HIGGINS, J., LAI, K. M., JI, J., DUDOIT, S., NG, I. O., VAN DE RIJN, M., BOTSTEIN, D. & BROWN, P. O. 2002. Gene expression patterns in human liver cancers. *Mol Biol Cell,* 13**,** 1929-39.

CHEN, X., JI, Z., WEBBER, A. & SHARROCKS, A. D. 2016. Genome-wide binding studies reveal DNA binding specificity mechanisms and functional interplay amongst Forkhead transcription factors. *Nucleic Acids Res,* 44**,** 1566-78.
CHOKAS, A. L., TRIVEDI, C. M., LU, M. M., TUCKER, P. W., LI, S., EPSTEIN, J. A. & MORRISEY, E. E. 2010. Foxp1/2/4-NuRD interactions regulate gene expression and epithelial injury response in the lung via regulation of interleukin-6. *J Biol Chem,* 285**,** 13304-13.
CHU, Y. P., CHANG, C. H., SHIU, J. H., CHANG, Y. T., CHEN, C. Y. & CHUANG, W. J. 2011. Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: insight into its domain swapping and DNA binding. *Protein Sci,* 20**,** 908-24.
CLARIDGE-CHANG, A., ROORDA, R. D., VRONTOU, E., SJULSON, L., LI, H., HIRSH, J. & MIESENBOCK, G. 2009. Writing memories with light-addressable reinforcement circuitry. *Cell,* 139**,** 405-15.
COHN, R., MORANTTE, I. & RUTA, V. 2015. Coordinated and Compartmentalized Neuromodulation Shapes Sensory Processing in Drosophila. *Cell,* 163**,** 1742-55.
COLLADO-VIDES, J., MAGASANIK, B. & GRALLA, J. D. 1991. Control site location and transcriptional regulation in Escherichia coli. *Microbiol Rev,* 55**,** 371-94.
COLOMB, J. & BREMBS, B. 2016. PKC in motorneurons underlies self-learning, a form of motor learning in Drosophila. *PeerJ,* 4**,** e1971.
COOPER, S. J., TRINKLEIN, N. D., ANTON, E. D., NGUYEN, L. & MYERS, R. M. 2006. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res,* 16**,** 1-10.
COSMA, M. P., TANAKA, T. & NASMYTH, K. 1999. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell,* 97**,** 299-311.
CRICK, F. 1970. Central dogma of molecular biology. *Nature,* 227**,** 561-3.
DARNELL, J. E., JR. 1982. Variety in the level of gene control in eukaryotic cells. *Nature,* 297**,** 365-71.
DASGUPTA, S., FERREIRA, C. H. & MIESENBOCK, G. 2014. FoxP influences the speed and accuracy of a perceptual decision in Drosophila. *Science,* 344**,** 901-4.
DATAR, S. A., JACOBS, H. W., DE LA CRUZ, A. F., LEHNER, C. F. & EDGAR, B. A. 2000. The Drosophila cyclin D-Cdk4 complex promotes cellular growth. *EMBO J,* 19**,** 4543-54.
DESJARLAIS & BERG, J. M. 1992. Redesigning the DNA-binding specificity of a zinc finger protein: a data base-guided approach. *Proteins,* 13**,** 272.
DEVANNA, P., MIDDELBEEK, J. & VERNES, S. C. 2014. FOXP2 drives neuronal differentiation by interacting with retinoic acid signaling pathways. *Front Cell Neurosci,* 8**,** 305.
DIETZL, G., CHEN, D., SCHNORRER, F., SU, K. C., BARINOVA, Y., FELLNER, M., GASSER, B., KINSEY, K., OPPEL, S., SCHEIBLAUER, S., COUTO, A., MARRA, V., KELEMAN, K. & DICKSON, B. J. 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature,* 448**,** 151-6.
DING, L. & GOLD, J. I. 2013. The basal ganglia's contributions to perceptual decision making. *Neuron,* 79**,** 640-9.
DING, N., ZHOU, H., ESTEVE, P. O., CHIN, H. G., KIM, S., XU, X., JOSEPH, S. M., FRIEZ, M. J., SCHWARTZ, C. E., PRADHAN, S. & BOYER, T. G. 2008. Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. *Mol Cell,* 31**,** 347-59.

EGGER, B., LEEMANS, R., LOOP, T., KAMMERMEIER, L., FAN, Y., RADIMERSKI, T., STRAHM, M. C., CERTA, U. & REICHERT, H. 2002. Gliogenesis in Drosophila: genome-wide analysis of downstream genes of glial cells missing in the embryonic nervous system. *Development,* 129**,** 3295-309.

EL-KHOLY, S., STEPHANO, F., LI, Y., BHANDARI, A., FINK, C. & ROEDER, T. 2015. Expression analysis of octopamine and tyramine receptors in Drosophila. *Cell Tissue Res,* 361**,** 669-84.

ELENKOV, I., PELOVSKY, P., UGRINOVA, I., TAKAHASHI, M. & PASHEVA, E. 2011. The DNA binding and bending activities of truncated tail-less HMGB1 protein are differentially affected by Lys-2 and Lys-81 residues and their acetylation. *Int J Biol Sci,* 7**,** 691-9.

ELLEGREN, H. 2004. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet,* 5**,** 435-45.

ENARD, W., GEHRE, S., HAMMERSCHMIDT, K., HOLTER, S. M., BLASS, T., SOMEL, M., BRUCKNER, M. K., SCHREIWEIS, C., WINTER, C., SOHR, R., BECKER, L., WIEBE, V., NICKEL, B., GIGER, T., MULLER, U., GROSZER, M., ADLER, T., AGUILAR, A., BOLLE, I., CALZADA-WACK, J., DALKE, C., EHRHARDT, N., FAVOR, J., FUCHS, H., GAILUS-DURNER, V., HANS, W., HOLZLWIMMER, G., JAVAHERI, A., KALAYDJIEV, S., KALLNIK, M., KLING, E., KUNDER, S., MOSSBRUGGER, I., NATON, B., RACZ, I., RATHKOLB, B., ROZMAN, J., SCHREWE, A., BUSCH, D. H., GRAW, J., IVANDIC, B., KLINGENSPOR, M., KLOPSTOCK, T., OLLERT, M., QUINTANILLA-MARTINEZ, L., SCHULZ, H., WOLF, E., WURST, W., ZIMMER, A., FISHER, S. E., MORGENSTERN, R., ARENDT, T., DE ANGELIS, M. H., FISCHER, J., SCHWARZ, J. & PAABO, S. 2009. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell,* 137**,** 961-71.

FALB, D. & MANIATIS, T. 1992. A conserved regulatory unit implicated in tissue-specific gene expression in Drosophila and man. *Genes Dev,* 6**,** 454-65.

FAYYAZUDDIN, A., ZAHEER, M. A., HIESINGER, P. R. & BELLEN, H. J. 2006. The nicotinic acetylcholine receptor Dalpha7 is required for an escape behavior in Drosophila. *PLoS Biol,* 4**,** e63.

FERLAND, R. J., CHERRY, T. J., PREWARE, P. O., MORRISEY, E. E. & WALSH, C. A. 2003. Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. *J Comp Neurol,* 460**,** 266-79.

FERRERO, E., FISCHER, B. & RUSSELL, S. 2014. SoxNeuro orchestrates central nervous system specification and differentiation in Drosophila and is only partially redundant with Dichaete. *Genome Biol,* 15**,** R74.

FEUK, L., KALERVO, A., LIPSANEN-NYMAN, M., SKAUG, J., NAKABAYASHI, K., FINUCANE, B., HARTUNG, D., INNES, M., KEREM, B., NOWACZYK, M. J., RIVLIN, J., ROBERTS, W., SENMAN, L., SUMMERS, A., SZATMARI, P., WONG, V., VINCENT, J. B., ZEESMAN, S., OSBORNE, L. R., CARDY, J. O., KERE, J., SCHERER, S. W. & HANNULA-JOUPPI, K. 2006. Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. *Am J Hum Genet,* 79**,** 965-72.

FIELDS, S. & SONG, O. 1989. A novel genetic system to detect protein-protein interactions. *Nature,* 340**,** 245-6.

FREEMAN, L. A. & GARRARD, W. T. 1992. DNA supercoiling in chromatin structure and gene expression. *Crit Rev Eukaryot Gene Expr,* 2**,** 165-209.

FRIED, M. G. 1989. Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis,* 10**,** 366-76.
GABUT, M., SAMAVARCHI-TEHRANI, P., WANG, X., SLOBODENIUC, V., O'HANLON, D., SUNG, H. K., ALVAREZ, M., TALUKDER, S., PAN, Q., MAZZONI, E. O., NEDELEC, S., WICHTERLE, H., WOLTJEN, K., HUGHES, T. R., ZANDSTRA, P. W., NAGY, A., WRANA, J. L. & BLENCOWE, B. J. 2011. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell,* 147**,** 132-46.
GALLO, S. M., GERRARD, D. T., MINER, D., SIMICH, M., DES SOYE, B., BERGMAN, C. M. & HALFON, M. S. 2011. REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in Drosophila. *Nucleic Acids Res,* 39**,** D118-23.
GATES, M. A., KANNAN, R. & GINIGER, E. 2011. A genome-wide analysis reveals that the Drosophila transcription factor Lola promotes axon growth in part by suppressing expression of the actin nucleation factor Spire. *Neural Dev,* 6**,** 37.
GERVASI, N., TCHENIO, P. & PREAT, T. 2010. PKA dynamics in a Drosophila learning center: coincidence detection by rutabaga adenylyl cyclase and spatial regulation by dunce phosphodiesterase. *Neuron,* 65**,** 516-29.
GIRARD, F., JOLY, W., SAVARE, J., BONNEAUD, N., FERRAZ, C. & MASCHAT, F. 2006. Chromatin immunoprecipitation reveals a novel role for the Drosophila SoxNeuro transcription factor in axonal patterning. *Dev Biol,* 299**,** 530-42.
GOLSON, M. L. & KAESTNER, K. H. 2016. Fox transcription factors: from development to disease. *Development,* 143**,** 4558-4570.
GOMES, A. R., ZHAO, F. & LAM, E. W. 2013. Role and regulation of the forkhead transcription factors FOXO3a and FOXM1 in carcinogenesis and drug resistance. *Chin J Cancer,* 32**,** 365-70.
GOODWIN, S. F., TAYLOR, B. J., VILLELLA, A., FOSS, M., RYNER, L. C., BAKER, B. S. & HALL, J. C. 2000. Aberrant splicing and altered spatial expression patterns in fruitless mutants of Drosophila melanogaster. *Genetics,* 154**,** 725-45.
GRAVES, B. J. 1998. Inner workings of a transcription factor partnership. *Science,* 279**,** 1000-2.
GREENGARD, P. 2001. The neurobiology of slow synaptic transmission. *Science,* 294**,** 1024-30.
GREIL, F., MOORMAN, C. & VAN STEENSEL, B. 2006. DamID: mapping of in vivo protein-genome interactions using tethered DNA adenine methyltransferase. *Methods Enzymol,* 410**,** 342-59.
GRENNINGLOH, G., REHM, E. J. & GOODMAN, C. S. 1991. Genetic analysis of growth cone guidance in Drosophila: fasciclin II functions as a neuronal recognition molecule. *Cell,* 67**,** 45-57.
GRIFFITH, L. C., WANG, J., ZHONG, Y., WU, C. F. & GREENSPAN, R. J. 1994. Calcium/calmodulin-dependent protein kinase II and potassium channel subunit eag similarly affect plasticity in Drosophila. *Proc Natl Acad Sci U S A,* 91**,** 10044-8.
GROSCHNER, L. N., CHAN WAH HAK, L., BOGACZ, R., DASGUPTA, S. & MIESENBOCK, G. 2018. Dendritic Integration of Sensory Evidence in Perceptual Decision-Making. *Cell,* 173**,** 894-905 e13.
GROSSCHEDL, R. 1995. Higher-order nucleoprotein complexes in transcription: analogies with site-specific recombination. *Curr Opin Cell Biol,* 7**,** 362-70.
HAMDAN, F. F., DAOUD, H., ROCHEFORT, D., PITON, A., GAUTHIER, J., LANGLOIS, M., FOOMANI, G., DOBRZENIECKA, S., KREBS, M. O.,

JOOBER, R., LAFRENIERE, R. G., LACAILLE, J. C., MOTTRON, L., DRAPEAU, P., BEAUCHAMP, M. H., PHILLIPS, M. S., FOMBONNE, E., ROULEAU, G. A. & MICHAUD, J. L. 2010. De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. *Am J Hum Genet,* 87**,** 671-8.

HANNA-ROSE, W. & HANSEN, U. 1996. Active repression mechanisms of eukaryotic transcription repressors. *Trends Genet,* 12**,** 229-34.

HATTORI, Y., USUI, T., SATOH, D., MORIYAMA, S., SHIMONO, K., ITOH, T., SHIRAHIGE, K. & UEMURA, T. 2013. Sensory-neuron subtype-specific transcriptional programs controlling dendrite morphogenesis: genome-wide analysis of Abrupt and Knot/Collier. *Dev Cell,* 27**,** 530-44.

HEIMAN, M., KULICKE, R., FENSTER, R. J., GREENGARD, P. & HEINTZ, N. 2014. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nat Protoc,* 9**,** 1282-91.

HERNANDEZ, V. M., HEGEMAN, D. J., CUI, Q., KELVER, D. A., FISKE, M. P., GLAJCH, K. E., PITT, J. E., HUANG, T. Y., JUSTICE, N. J. & CHAN, C. S. 2015. Parvalbumin+ Neurons and Npas1+ Neurons Are Distinct Neuron Classes in the Mouse External Globus Pallidus. *J Neurosci,* 35**,** 11830-47.

HOSOYA, T., TAKIZAWA, K., NITTA, K. & HOTTA, Y. 1995. glial cells missing: a binary switch between neuronal and glial determination in Drosophila. *Cell,* 82**,** 1025-36.

HURST, H. C. 1995. Transcription factors 1: bZIP proteins. *Protein Profile,* 2**,** 101-68.

IMAI, Y., KOBAYASHI, Y., INOSHITA, T., MENG, H., ARANO, T., UEMURA, K., ASANO, T., YOSHIMI, K., ZHANG, C. L., MATSUMOTO, G., OHTSUKA, T., KAGEYAMA, R., KIYONARI, H., SHIOI, G., NUKINA, N., HATTORI, N. & TAKAHASHI, R. 2015. The Parkinson's Disease-Associated Protein Kinase LRRK2 Modulates Notch Signaling through the Endosomal Pathway. *PLoS Genet,* 11**,** e1005503.

INUKAI, S., KOCK, K. H. & BULYK, M. L. 2017. Transcription factor-DNA binding: beyond binding site motifs. *Curr Opin Genet Dev,* 43**,** 110-119.

JACOB, F. & MONOD, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol,* 3**,** 318-56.

JAMAIN, S., QUACH, H., BETANCUR, C., RASTAM, M., COLINEAUX, C., GILLBERG, I. C., SODERSTROM, H., GIROS, B., LEBOYER, M., GILLBERG, C., BOURGERON, T. & PARIS AUTISM RESEARCH INTERNATIONAL SIBPAIR, S. 2003. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet,* 34**,** 27-9.

JONES, B. W. 2005. Transcriptional control of glial cell development in Drosophila. *Dev Biol,* 278**,** 265-73.

JONES, B. W., FETTER, R. D., TEAR, G. & GOODMAN, C. S. 1995. glial cells missing: a genetic switch that controls glial versus neuronal fate. *Cell,* 82**,** 1013-23.

JONES, S. 2004. An overview of the basic helix-loop-helix proteins. *Genome Biol,* 5**,** 226.

KAESTNER, K. H., KNOCHEL, W. & MARTINEZ, D. E. 2000. Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev,* 14**,** 142-6.

KAMACHI, Y. & KONDOH, H. 2013. Sox proteins: regulators of cell fate specification and differentiation. *Development,* 140**,** 4129-44.

KARCZEWSKI, K. J., TATONETTI, N. P., LANDT, S. G., YANG, X., SLIFER, T., ALTMAN, R. B. & SNYDER, M. 2011. Cooperative transcription factor associations discovered using regulatory variation. *Proc Natl Acad Sci U S A,* 108**,** 13353-8.

KAROLCHIK, D., BAERTSCH, R., DIEKHANS, M., FUREY, T. S., HINRICHS, A., LU, Y. T., ROSKIN, K. M., SCHWARTZ, M., SUGNET, C. W., THOMAS, D. J., WEBER, R. J., HAUSSLER, D., KENT, W. J. & UNIVERSITY OF CALIFORNIA SANTA, C. 2003. The UCSC Genome Browser Database. *Nucleic Acids Res,* 31**,** 51-4.

KASSIS, J. A., KENNISON, J. A. & TAMKUN, J. W. 2017. Polycomb and Trithorax Group Genes in Drosophila. *Genetics,* 206**,** 1699-1725.

KAUFMANN, K., MUINO, J. M., OSTERAS, M., FARINELLI, L., KRAJEWSKI, P. & ANGENENT, G. C. 2010a. Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). *Nat Protoc,* 5**,** 457-72.

KAUFMANN, R., STRAUSSBERG, R., MANDEL, H., FATTAL-VALEVSKI, A., BEN-ZEEV, B., NAAMATI, A., SHAAG, A., ZENVIRT, S., KONEN, O., MIMOUNI-BLOCH, A., DOBYNS, W. B., EDVARDSON, S., PINES, O. & ELPELEG, O. 2010b. Infantile cerebral and cerebellar atrophy is associated with a mutation in the MED17 subunit of the transcription preinitiation mediator complex. *Am J Hum Genet,* 87**,** 667-70.

KERAVALA, A. & CALOS, M. P. 2008. Site-specific chromosomal integration mediated by phiC31 integrase. *Methods Mol Biol,* 435**,** 165-73.

KIKUCHI, A., KISHIDA, S. & YAMAMOTO, H. 2006. Regulation of Wnt signaling by protein-protein interaction and post-translational modifications. *Exp Mol Med,* 38**,** 1-10.

KIM, I. M., ACKERSON, T., RAMAKRISHNA, S., TRETIAKOVA, M., WANG, I. C., KALIN, T. V., MAJOR, M. L., GUSAROVA, G. A., YODER, H. M., COSTA, R. H. & KALINICHENKO, V. V. 2006. The Forkhead Box m1 transcription factor stimulates the proliferation of tumor cells during development of lung cancer. *Cancer Res,* 66**,** 2153-61.

KLEMM, J. D., SCHREIBER, S. L. & CRABTREE, G. R. 1998. Dimerization as a regulatory mechanism in signal transduction. *Annu Rev Immunol,* 16**,** 569-92.

KOMIYAMA, T., JOHNSON, W. A., LUO, L. & JEFFERIS, G. S. 2003. From lineage to wiring specificity. POU domain transcription factors control precise connections of Drosophila olfactory projection neurons. *Cell,* 112**,** 157-67.

KONOPKA, G., BOMAR, J. M., WINDEN, K., COPPOLA, G., JONSSON, Z. O., GAO, F., PENG, S., PREUSS, T. M., WOHLSCHLEGEL, J. A. & GESCHWIND, D. H. 2009. Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature,* 462**,** 213-7.

KOZAK, M. 2001. Constraints on reinitiation of translation in mammals. *Nucleic Acids Res,* 29**,** 5226-32.

KOZERA, B. & RAPACZ, M. 2013. Reference genes in real-time PCR. *J Appl Genet,* 54**,** 391-406.

LAI, C. S., FISHER, S. E., HURST, J. A., VARGHA-KHADEM, F. & MONACO, A. P. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature,* 413**,** 519-23.

LAM, B. J., BAKSHI, A., EKINCI, F. Y., WEBB, J., GRAVELEY, B. R. & HERTEL, K. J. 2003. Enhancer-dependent 5'-splice site control of fruitless pre-mRNA splicing. *J Biol Chem,* 278**,** 22740-7.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM,

J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, Y., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.

LATCHMAN, D. S. 1990. Cell-type-specific splicing factors and the regulation of alternative RNA splicing. *New Biol,* 2**,** 297-303.

LATCHMAN, D. S. 1997. Transcription factors: an overview. *Int J Biochem Cell Biol,* 29**,** 1305-12.

LAWTON, K. J., WASSMER, T. L. & DEITCHER, D. L. 2014. Conserved role of Drosophila melanogaster FoxP in motor coordination and courtship song. *Behav Brain Res,* 268**,** 213-21.

LE LAY, J. & KAESTNER, K. H. 2010. The Fox genes in the liver: from organogenesis to functional integration. *Physiol Rev,* 90**,** 1-22.

LEE, M. H. & SALVATERRA, P. M. 2002. Abnormal chemosensory jump 6 is a positive transcriptional regulator of the cholinergic gene locus in Drosophila olfactory neurons. *J Neurosci,* 22**,** 5291-9.

LEE, T. I. & YOUNG, R. A. 2013. Transcriptional regulation and its misregulation in disease. *Cell,* 152**,** 1237-51.

LEISHMAN, E., HOWARD, J. M., GARCIA, G. E., MIAO, Q., KU, A. T., DEKKER, J. D., TUCKER, H. & NGUYEN, H. 2013. Foxp1 maintains hair follicle stem cell quiescence through regulation of Fgf18. *Development,* 140**,** 3809-18.

LEVINE, M. & TJIAN, R. 2003. Transcription regulation and animal diversity. *Nature,* 424**,** 147-51.

LI, H., CHEN, D. & ZHANG, J. 2012. Analysis of intron sequence features associated with transcriptional regulation in human genes. *PLoS One,* 7**,** e46784.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, S., WEIDENFELD, J. & MORRISEY, E. E. 2004. Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. *Mol Cell Biol,* 24**,** 809-22.

LIEBERMAN, P. 2009. FOXP2 and Human Cognition. *Cell,* 137**,** 800-2.

LIN, S. & RIGGS, A. D. 1975. The general affinity of lac repressor for E. coli DNA: implications for gene regulation in procaryotes and eucaryotes. *Cell,* 4**,** 107-11.
LIS, M. & WALTHER, D. 2016. The orientation of transcription factor binding site motifs in gene promoter regions: does it matter? *BMC Genomics,* 17**,** 185.
LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25**,** 402-8.
LOGAN, C. Y. & NUSSE, R. 2004. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol,* 20**,** 781-810.
LOPES, J. E., TORGERSON, T. R., SCHUBERT, L. A., ANOVER, S. D., OCHELTREE, E. L., OCHS, H. D. & ZIEGLER, S. F. 2006. Analysis of FOXP3 reveals multiple domains required for its function as a transcriptional repressor. *J Immunol,* 177**,** 3133-42.
LUO, L., LIAO, Y. J., JAN, L. Y. & JAN, Y. N. 1994. Distinct morphogenetic functions of similar small GTPases: Drosophila Drac1 is involved in axonal outgrowth and myoblast fusion. *Genes Dev,* 8**,** 1787-802.
LYNE, R., SMITH, R., RUTHERFORD, K., WAKELING, M., VARLEY, A., GUILLIER, F., JANSSENS, H., JI, W., MCLAREN, P., NORTH, P., RANA, D., RILEY, T., SULLIVAN, J., WATKINS, X., WOODBRIDGE, M., LILLEY, K., RUSSELL, S., ASHBURNER, M., MIZUGUCHI, K. & MICKLEM, G. 2007. FlyMine: an integrated database for Drosophila and Anopheles genomics. *Genome Biol,* 8**,** R129.
MA, Y., NIEMITZ, E. L., NAMBU, P. A., SHAN, X., SACKERSON, C., FUJIOKA, M., GOTO, T. & NAMBU, J. R. 1998. Gene regulatory functions of Drosophila fish-hook, a high mobility group domain Sox protein. *Mech Dev,* 73**,** 169-82.
MACARTHUR, S., LI, X. Y., LI, J., BROWN, J. B., CHU, H. C., ZENG, L., GRONDONA, B. P., HECHMER, A., SIMIRENKO, L., KERANEN, S. V., KNOWLES, D. W., STAPLETON, M., BICKEL, P., BIGGIN, M. D. & EISEN, M. B. 2009. Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol,* 10**,** R80.
MAKIYAMA, K., HAMADA, J., TAKADA, M., MURAKAWA, K., TAKAHASHI, Y., TADA, M., TAMOTO, E., SHINDO, G., MATSUNAGA, A., TERAMOTO, K., KOMURO, K., KONDO, S., KATOH, H., KOIKE, T. & MORIUCHI, T. 2005. Aberrant expression of HOX genes in human invasive breast carcinoma. *Oncol Rep,* 13**,** 673-9.
MARSHALL, O. J., SOUTHALL, T. D., CHEETHAM, S. W. & BRAND, A. H. 2016. Cell-type-specific profiling of protein-DNA interactions without cell isolation using targeted DamID with next-generation sequencing. *Nat Protoc,* 11**,** 1586-98.
MATHELIER, A., ZHAO, X., ZHANG, A. W., PARCY, F., WORSLEY-HUNT, R., ARENILLAS, D. J., BUCHMAN, S., CHEN, C. Y., CHOU, A., IENASESCU, H., LIM, J., SHYR, C., TAN, G., ZHOU, M., LENHARD, B., SANDELIN, A. & WASSERMAN, W. W. 2014. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res,* 42**,** D142-7.
MATYS, V., KEL-MARGOULIS, O. V., FRICKE, E., LIEBICH, I., LAND, S., BARRE-DIRRIE, A., REUTER, I., CHEKMENEV, D., KRULL, M., HORNISCHER, K., VOSS, N., STEGMAIER, P., LEWICKI-POTAPOV, B., SAXEL, H., KEL, A. E. & WINGENDER, E. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res,* 34**,** D108-10.

MAURANGE, C., CHENG, L. & GOULD, A. P. 2008. Temporal transcription factors and their targets schedule the end of neural proliferation in Drosophila. *Cell,* 133**,** 891-902.
MEDINA, E., CORDOVA, C., VILLALOBOS, P., REYES, J., KOMIVES, E. A., RAMIREZ-SARMIENTO, C. A. & BABUL, J. 2016. Three-Dimensional Domain Swapping Changes the Folding Mechanism of the Forkhead Domain of FoxP1. *Biophys J,* 110**,** 2349-2360.
MENDOZA, E., COLOMB, J., RYBAK, J., PFLUGER, H. J., ZARS, T., SCHARFF, C. & BREMBS, B. 2014. Drosophila FoxP mutants are deficient in operant self-learning. *PLoS One,* 9**,** e100648.
MIGNONE, F., GISSI, C., LIUNI, S. & PESOLE, G. 2002. Untranslated regions of mRNAs. *Genome Biol,* 3**,** REVIEWS0004.
MONDAL, S., IVANCHUK, S. M., RUTKA, J. T. & BOULIANNE, G. L. 2007. Sloppy paired 1/2 regulate glial cell fates by inhibiting Gcm function. *Glia,* 55**,** 282-93.
MOORMAN, C., SUN, L. V., WANG, J., DE WIT, E., TALHOUT, W., WARD, L. D., GREIL, F., LU, X. J., WHITE, K. P., BUSSEMAKER, H. J. & VAN STEENSEL, B. 2006. Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. *Proc Natl Acad Sci U S A,* 103**,** 12027-32.
MORGUNOVA, E. & TAIPALE, J. 2017. Structural perspective of cooperative transcription factor binding. *Curr Opin Struct Biol,* 47**,** 1-8.
MOSCA, T. J., HONG, W., DANI, V. S., FAVALORO, V. & LUO, L. 2012. Trans-synaptic Teneurin signalling in neuromuscular synapse organization and target choice. *Nature,* 484**,** 237-41.
MURTHY, M., FIETE, I. & LAURENT, G. 2008. Testing odor response stereotypy in the Drosophila mushroom body. *Neuron,* 59**,** 1009-23.
MYATT, S. S. & LAM, E. W. 2007. The emerging roles of forkhead box (Fox) proteins in cancer. *Nat Rev Cancer,* 7**,** 847-59.
NAMBU, P. A. & NAMBU, J. R. 1996. The Drosophila fish-hook gene encodes a HMG domain protein essential for segmentation and CNS development. *Development,* 122**,** 3467-75.
NEVILLE, M. C., NOJIMA, T., ASHLEY, E., PARKER, D. J., WALKER, J., SOUTHALL, T., VAN DE SANDE, B., MARQUES, A. C., FISCHER, B., BRAND, A. H., RUSSELL, S., RITCHIE, M. G., AERTS, S. & GOODWIN, S. F. 2014. Male-specific fruitless isoforms target neurodevelopmental genes to specify a sexually dimorphic nervous system. *Curr Biol,* 24**,** 229-41.
NG, F. S., SENGUPTA, S., HUANG, Y., YU, A. M., YOU, S., ROBERTS, M. A., IYER, L. K., YANG, Y. & JACKSON, F. R. 2016. TRAP-seq Profiling and RNAi-Based Genetic Screens Identify Conserved Glial Genes Required for Adult Drosophila Behavior. *Front Mol Neurosci,* 9**,** 146.
NITTA, K. R., JOLMA, A., YIN, Y., MORGUNOVA, E., KIVIOJA, T., AKHTAR, J., HENS, K., TOIVONEN, J., DEPLANCKE, B., FURLONG, E. E. & TAIPALE, J. 2015. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *Elife,* 4.
NOVOTNY, T., EISELT, R. & URBAN, J. 2002. Hunchback is required for the specification of the early sublineage of neuroblast 7-3 in the Drosophila central nervous system. *Development,* 129**,** 1027-36.
OGBOURNE, S. & ANTALIS, T. M. 1998. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J,* 331 (Pt 1)**,** 1-14.
OSWALD, F., KLOBLE, P., RULAND, A., ROSENKRANZ, D., HINZ, B., BUTTER, F., RAMLJAK, S., ZECHNER, U. & HERLYN, H. 2017. The FOXP2-Driven

Network in Developmental Disorders and Neurodegeneration. *Front Cell Neurosci,* 11**,** 212.
OVERTON, P. M., MEADOWS, L. A., URBAN, J. & RUSSELL, S. 2002. Evidence for differential and redundant function of the Sox genes Dichaete and SoxN during CNS development in Drosophila. *Development,* 129**,** 4219-28.
PABO, C. O. & SAUER, R. T. 1992. Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem,* 61**,** 1053-95.
PACKARD, M., KOO, E. S., GORCZYCA, M., SHARPE, J., CUMBERLEDGE, S. & BUDNIK, V. 2002. The Drosophila Wnt, wingless, provides an essential signal for pre- and postsynaptic differentiation. *Cell,* 111**,** 319-30.
PAN, F., YU, H., DANG, E. V., BARBI, J., PAN, X., GROSSO, J. F., JINASENA, D., SHARMA, S. M., MCCADDEN, E. M., GETNET, D., DRAKE, C. G., LIU, J. O., OSTROWSKI, M. C. & PARDOLL, D. M. 2009. Eos mediates Foxp3-dependent gene silencing in CD4+ regulatory T cells. *Science,* 325**,** 1142-6.
PAPATHEODOROU, I., OELLRICH, A. & SMEDLEY, D. 2015. Linking gene expression to phenotypes via pathway information. *J Biomed Semantics,* 6**,** 17.
PARKER, M. H., SEALE, P. & RUDNICKI, M. A. 2003. Looking back to the embryo: defining transcriptional networks in adult myogenesis. *Nat Rev Genet,* 4**,** 497-507.
POSE, D. & YANT, L. 2016. DNA-Binding Factor Target Identification by Chromatin Immunoprecipitation (ChIP) in Plants. *Methods Mol Biol,* 1363**,** 25-35.
PTASHNE, M. & GANN, A. 1997. Transcriptional activation by recruitment. *Nature,* 386**,** 569-77.
QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-2.
RAMIREZ, F., RYAN, D. P., GRUNING, B., BHARDWAJ, V., KILPERT, F., RICHTER, A. S., HEYNE, S., DUNDAR, F. & MANKE, T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res,* 44**,** W160-5.
RASTEGAR-POUYANI, S., KHAZAEI, N., WEE, P., MOHAMMADNIA, A. & YAQUBI, M. 2016. Role of Hepatic-Specific Transcription Factors and Polycomb Repressive Complex 2 during Induction of Fibroblasts to Hepatic Fate. *PLoS One,* 11**,** e0167081.
REN, B. & DYNLACHT, B. D. 2004. Use of chromatin immunoprecipitation assays in genome-wide location analysis of mammalian transcription factors. *Methods Enzymol,* 376**,** 304-15.
RIDEOUT, E. J., BILLETER, J. C. & GOODWIN, S. F. 2007. The sex-determination genes fruitless and doublesex specify a neural substrate required for courtship song. *Curr Biol,* 17**,** 1473-8.
ROLL, P., VERNES, S. C., BRUNEAU, N., CILLARIO, J., PONSOLE-LENFANT, M., MASSACRIER, A., RUDOLF, G., KHALIFE, M., HIRSCH, E., FISHER, S. E. & SZEPETOWSKI, P. 2010. Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex. *Hum Mol Genet,* 19**,** 4848-60.
ROSENFELD, N., YOUNG, J. W., ALON, U., SWAIN, P. S. & ELOWITZ, M. B. 2005. Gene regulation at the single-cell level. *Science,* 307**,** 1962-5.
RYLAND, K. E., SVOBODA, L. K., VESELY, E. D., MCINTYRE, J. C., ZHANG, L., MARTENS, J. R. & LAWLOR, E. R. 2015. Polycomb-dependent repression of the potassium channel-encoding gene KCNA5 promotes cancer cell survival under conditions of stress. *Oncogene,* 34**,** 4591-600.

RYNER, L. C., GOODWIN, S. F., CASTRILLON, D. H., ANAND, A., VILLELLA, A., BAKER, B. S., HALL, J. C., TAYLOR, B. J. & WASSERMAN, S. A. 1996. Control of male sexual behavior and sexual orientation in Drosophila by the fruitless gene. *Cell,* 87**,** 1079-89.

SAGARDOY, A., MARTINEZ-FERRANDIS, J. I., ROA, S., BUNTING, K. L., AZNAR, M. A., ELEMENTO, O., SHAKNOVICH, R., FONTAN, L., FRESQUET, V., PEREZ-ROGER, I., ROBLES, E. F., DE SMEDT, L., SAGAERT, X., MELNICK, A. & MARTINEZ-CLIMENT, J. A. 2013. Downregulation of FOXP1 is required during germinal center B-cell function. *Blood,* 121**,** 4311-20.

SALGHETTI, S. E., CAUDY, A. A., CHENOWETH, J. G. & TANSEY, W. P. 2001. Regulation of transcriptional activation domain function by ubiquitin. *Science,* 293**,** 1651-3.

SANCHEZ-SORIANO, N. & RUSSELL, S. 2000. Regulatory mutations of the Drosophila Sox gene Dichaete reveal new functions in embryonic brain and hindgut development. *Dev Biol,* 220**,** 307-21.

SANTOS, M. E., ATHANASIADIS, A., LEITAO, A. B., DUPASQUIER, L. & SUCENA, E. 2011. Alternative splicing and gene duplication in the evolution of the FoxP gene subfamily. *Mol Biol Evol,* 28**,** 237-47.

SCHATTON, A. & SCHARFF, C. 2017. FoxP expression identifies a Kenyon cell subtype in the honeybee mushroom bodies linking them to fruit fly alphabetac neurons. *Eur J Neurosci,* 46**,** 2534-2541.

SCHWECHHEIMER, C., SMITH, C. & BEVAN, M. W. 1998. The activities of acidic and glutamine-rich transcriptional activation domains in plant cells: design of modular transcription factors for high-level expression. *Plant Mol Biol,* 36**,** 195-204.

SETTY, Y., MAYO, A. E., SURETTE, M. G. & ALON, U. 2003. Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci U S A,* 100**,** 7702-7.

SHEN-ORR, S. S., MILO, R., MANGAN, S. & ALON, U. 2002. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat Genet,* 31**,** 64-8.

SHU, W., YANG, H., ZHANG, L., LU, M. M. & MORRISEY, E. E. 2001. Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. *J Biol Chem,* 276**,** 27488-97.

SIGGERS, T., DUYZEND, M. H., REDDY, J., KHAN, S. & BULYK, M. L. 2011. Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. *Mol Syst Biol,* 7**,** 555.

SIN, C., LI, H. & CRAWFORD, D. A. 2015. Transcriptional regulation by FOXP1, FOXP2, and FOXP4 dimerization. *J Mol Neurosci,* 55**,** 437-48.

SMALE, S. T. & KADONAGA, J. T. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem,* 72**,** 449-79.

SOLBERG, N. & KRAUSS, S. 2013. Luciferase assay to study the activity of a cloned promoter DNA fragment. *Methods Mol Biol,* 977**,** 65-78.

SOUTHALL, T. D., GOLD, K. S., EGGER, B., DAVIDSON, C. M., CAYGILL, E. E., MARSHALL, O. J. & 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**,** 101-12.

SPITERI, E., KONOPKA, G., COPPOLA, G., BOMAR, J., OLDHAM, M., OU, J., VERNES, S. C., FISHER, S. E., REN, B. & GESCHWIND, D. H. 2007. Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. *Am J Hum Genet,* 81**,** 1144-57.

SPITZ, F. & FURLONG, E. E. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet,* 13**,** 613-26.
SRINIVASAN, S., LANCE, K. & LEVINE, R. B. 2012. Contribution of EAG to excitability and potassium currents in Drosophila larval motoneurons. *J Neurophysiol,* 107**,** 2660-71.
STRAUSFELD, N. J. & HIRTH, F. 2013. Deep homology of arthropod central complex and vertebrate basal ganglia. *Science,* 340**,** 157-61.
STROUD, J. C., WU, Y., BATES, D. L., HAN, A., NOWICK, K., PAABO, S., TONG, H. & CHEN, L. 2006. Structure of the forkhead domain of FOXP2 bound to DNA. *Structure,* 14**,** 159-66.
SZUTORISZ, H., DILLON, N. & TORA, L. 2005. The role of enhancers as centres for general transcription factor recruitment. *Trends Biochem Sci,* 30**,** 593-9.
TAHIROV, T. H., INOUE-BUNGO, T., MORII, H., FUJIKAWA, A., SASAKI, M., KIMURA, K., SHIINA, M., SATO, K., KUMASAKA, T., YAMAMOTO, M., ISHII, S. & OGATA, K. 2001. Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBFbeta. *Cell,* 104**,** 755-67.
TAKAHASHI, H., TAKAHASHI, K. & LIU, F. C. 2009. FOXP genes, neural development, speech and language disorders. *Adv Exp Med Biol,* 665**,** 117-29.
TAKEMURA, S. Y., KARUPPUDURAI, T., TING, C. Y., LU, Z., LEE, C. H. & MEINERTZHAGEN, I. A. 2011. Cholinergic circuits integrate neighboring visual signals in a Drosophila motion detection pathway. *Curr Biol,* 21**,** 2077-84.
TANAKA, N. K., TANIMOTO, H. & ITO, K. 2008. Neuronal assemblies of the Drosophila mushroom body. *J Comp Neurol,* 508**,** 711-55.
TERAMITSU, I., KUDO, L. C., LONDON, S. E., GESCHWIND, D. H. & WHITE, S. A. 2004. Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. *J Neurosci,* 24**,** 3152-63.
THOMAS, A., LEE, P. J., DALTON, J. E., NOMIE, K. J., STOICA, L., COSTA-MATTIOLI, M., CHANG, P., NUZHDIN, S., ARBEITMAN, M. N. & DIERICK, H. A. 2012. A versatile method for cell-specific profiling of translated mRNAs in Drosophila. *PLoS One,* 7**,** e40276.
TOOTLE, T. L. & REBAY, I. 2005. Post-translational modifications influence transcription factor activity: a view from the ETS superfamily. *Bioessays,* 27**,** 285-98.
TRIEZENBERG, S. J. 1995. Structure and function of transcriptional activation domains. *Curr Opin Genet Dev,* 5**,** 190-6.
UGRINOVA, I., PASHEV, I. G. & PASHEVA, E. A. 2009. Nucleosome binding properties and Co-remodeling activities of native and in vivo acetylated HMGB-1 and HMGB-2 proteins. *Biochemistry,* 48**,** 6502-7.
UNTERGASSER, A., CUTCUTACHE, I., KORESSAAR, T., YE, J., FAIRCLOTH, B. C., REMM, M. & ROZEN, S. G. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res,* 40**,** e115.
URIG, S., GOWHER, H., HERMANN, A., BECK, C., FATEMI, M., HUMENY, A. & JELTSCH, A. 2002. The Escherichia coli dam DNA methyltransferase modifies DNA in a highly processive reaction. *J Mol Biol,* 319**,** 1085-96.
VAN DER VELDEN, A. W. & THOMAS, A. A. 1999. The role of the 5' untranslated region of an mRNA in translation regulation during development. *Int J Biochem Cell Biol,* 31**,** 87-106.

VAN STEENSEL, B. & HENIKOFF, S. 2000. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol,* 18**,** 424-8.

VERNES, S. C. 2014. Genome wide identification of fruitless targets suggests a role in upregulating genes important for neural circuit formation. *Sci Rep,* 4**,** 4412.

VERNES, S. C., OLIVER, P. L., SPITERI, E., LOCKSTONE, H. E., PULIYADI, R., TAYLOR, J. M., HO, J., MOMBEREAU, C., BREWER, A., LOWY, E., NICOD, J., GROSZER, M., BABAN, D., SAHGAL, N., CAZIER, J. B., RAGOUSSIS, J., DAVIES, K. E., GESCHWIND, D. H. & FISHER, S. E. 2011. Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet,* 7**,** e1002145.

VERNES, S. C., SPITERI, E., NICOD, J., GROSZER, M., TAYLOR, J. M., DAVIES, K. E., GESCHWIND, D. H. & FISHER, S. E. 2007. High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *Am J Hum Genet,* 81**,** 1232-50.

VILLELLA, A. & HALL, J. C. 1996. Courtship anomalies caused by doublesex mutations in Drosophila melanogaster. *Genetics,* 143**,** 331-44.

VOGEL, M. J., PERIC-HUPKES, D. & VAN STEENSEL, B. 2007. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat Protoc,* 2**,** 1467-78.

WAGHMARE, I. & PAGE-MCCAW, A. 2018. Wnt Signaling in Stem Cell Maintenance and Differentiation in the Drosophila Germarium. *Genes (Basel),* 9.

WANG, B., LIN, D., LI, C. & TUCKER, P. 2003. Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. *J Biol Chem,* 278**,** 24259-68.

WEIRAUCH, M. T., YANG, A., ALBU, M., COTE, A. G., MONTENEGRO-MONTERO, A., DREWE, P., NAJAFABADI, H. S., LAMBERT, S. A., MANN, I., COOK, K., ZHENG, H., GOITY, A., VAN BAKEL, H., LOZANO, J. C., GALLI, M., LEWSEY, M. G., HUANG, E., MUKHERJEE, T., CHEN, X., REECE-HOYES, J. S., GOVINDARAJAN, S., SHAULSKY, G., WALHOUT, A. J. M., BOUGET, F. Y., RATSCH, G., LARRONDO, L. F., ECKER, J. R. & HUGHES, T. R. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell,* 158**,** 1431-1443.

WILLIAMSON, M. P. & SUTCLIFFE, M. J. 2010. Protein-protein interactions. *Biochem Soc Trans,* 38**,** 875-8.

WOLBERGER, C. 1999. Multiprotein-DNA complexes in transcriptional regulation. *Annu Rev Biophys Biomol Struct,* 28**,** 29-56.

XING, G., GAN, G., CHEN, D., SUN, M., YI, J., LV, H., HAN, J. & XIE, W. 2014. Drosophila neuroligin3 regulates neuromuscular junction development and synaptic differentiation. *J Biol Chem,* 289**,** 31867-77.

YAGI, H., NOMURA, T., NAKAMURA, K., YAMAZAKI, S., KITAWAKI, T., HORI, S., MAEDA, M., ONODERA, M., UCHIYAMA, T., FUJII, S. & SAKAGUCHI, S. 2004. Crucial role of FOXP3 in the development and function of human CD25+CD4+ regulatory T cells. *Int Immunol,* 16**,** 1643-56.

ZHANG, Z. W., PETERSON, M. & LIU, H. 2013. Essential role of postsynaptic NMDA receptors in developmental refinement of excitatory synapses. *Proc Natl Acad Sci U S A,* 110**,** 1095-100.

ZHAO, H. & DEAN, A. 2004. An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. *Nucleic Acids Res,* 32**,** 4903-19.

ZHENG, N., FRAENKEL, E., PABO, C. O. & PAVLETICH, N. P. 1999. Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev,* 13**,** 666-74.

ZHU, L. J., CHRISTENSEN, R. G., KAZEMIAN, M., HULL, C. J., ENUAMEH, M. S., BASCIOTTA, M. D., BRASEFIELD, J. A., ZHU, C., ASRIYAN, Y., LAPOINTE, D. S., SINHA, S., WOLFE, S. A. & BRODSKY, M. H. 2011a. FlyFactorSurvey: a database of Drosophila transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res,* 39**,** D111-7.

ZHU, S., BARSHOW, S., WILDONGER, J., JAN, L. Y. & JAN, Y. N. 2011b. Ets transcription factor Pointed promotes the generation of intermediate neural progenitors in Drosophila larval brains. *Proc Natl Acad Sci U S A,* 108**,** 20615-20.

ZINZEN, R. P., GIRARDOT, C., GAGNEUR, J., BRAUN, M. & FURLONG, E. E. 2009. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature,* 462**,** 65-70.

ZOU, Y., TSAI, W. B., CHENG, C. J., HSU, C., CHUNG, Y. M., LI, P. C., LIN, S. H. & HU, M. C. 2008. Forkhead box transcription factor FOXO3a suppresses estrogen-dependent breast cancer cell proliferation and tumorigenesis. *Breast Cancer Res,* 10**,** R21.

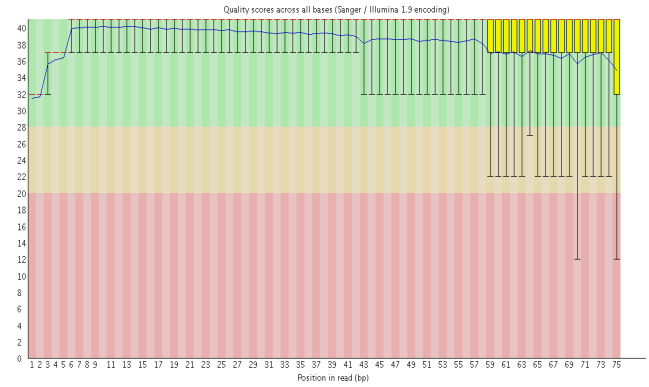
Appendix

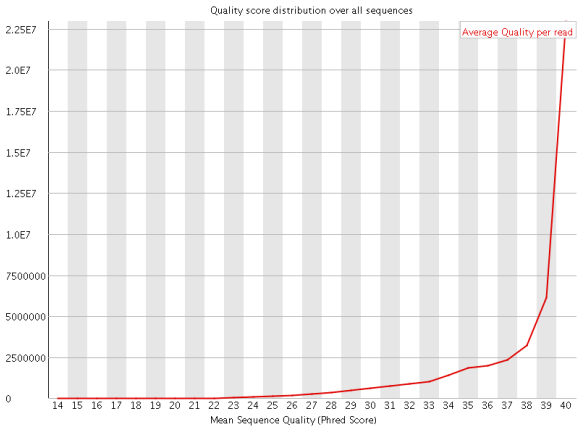
Appendix 1. Number of reads from each sequencing replicate

Sample	Number of reads
FoxPC_replicate1	33711982
FoxPC_replicate2	34460892
FoxPC_replicate3	30803932
FoxPD_replicate1	33690540
FoxPD_replicate2	33761082
FoxPD_replicate3	32659498
Control_replicate1	34574002
Control_replicate2	38759059
Control_replicate3	25094532

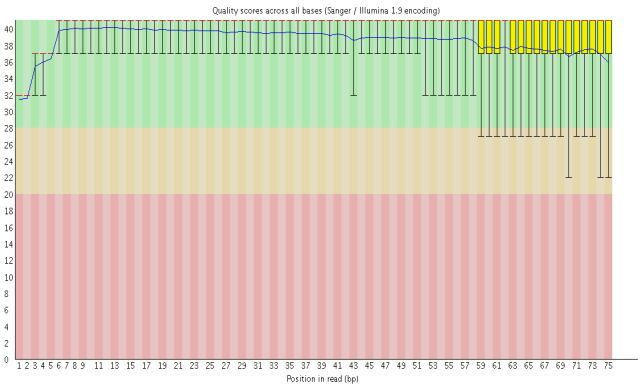
Appendix 2. FastQC reports of each replicate

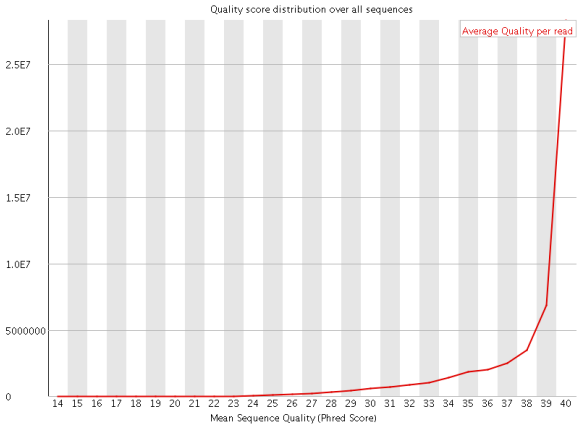
FoxPC_replicate 1



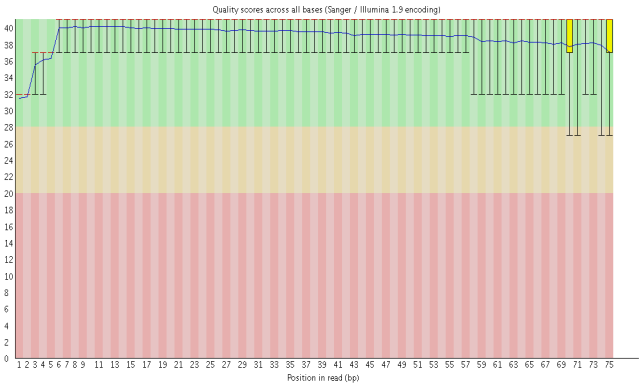


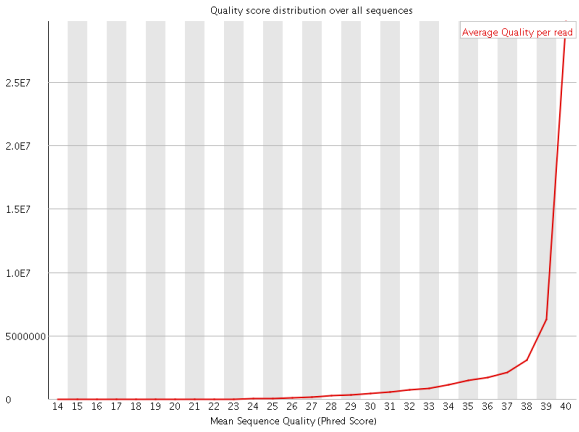
FoxPC_replicate 2



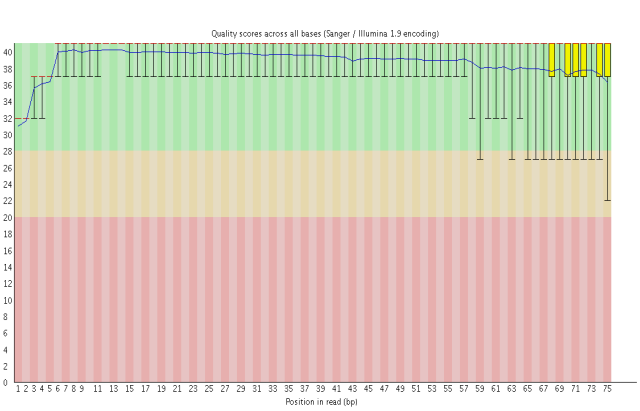


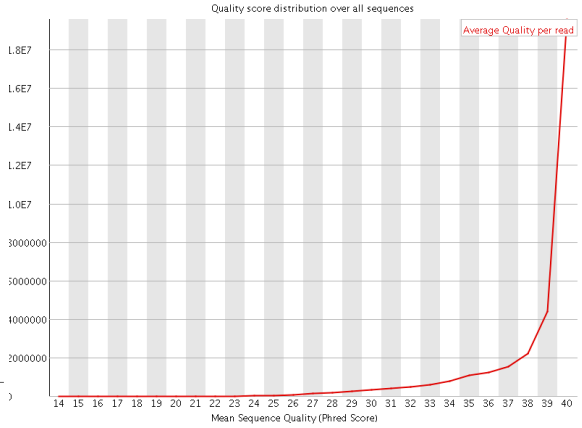
FoxPC_replicate3



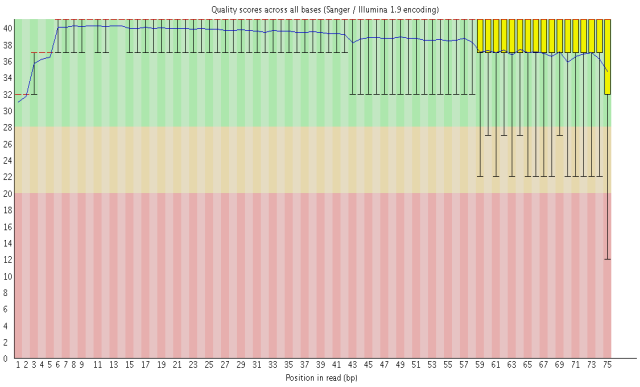


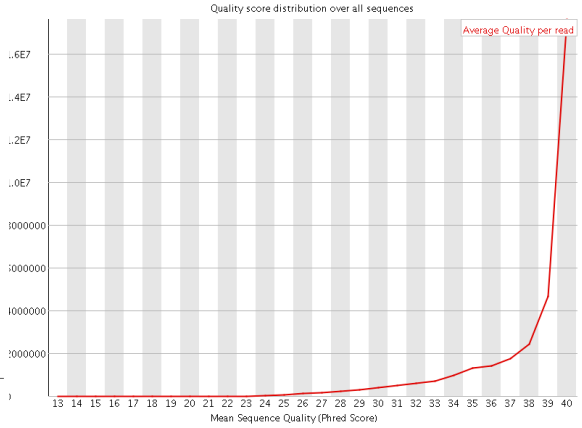
FoxPD_replicate 1



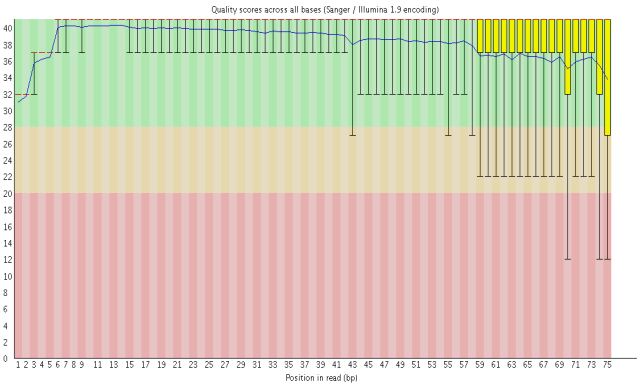


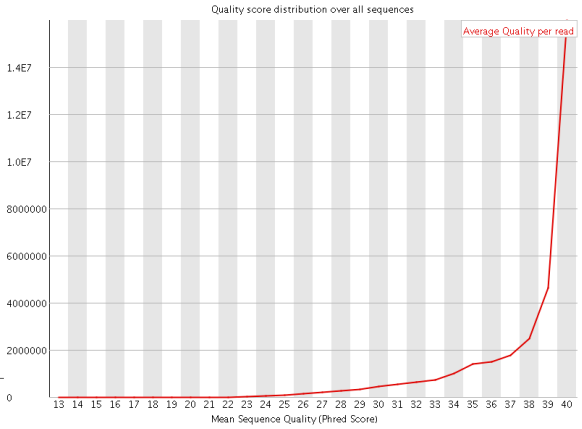
FoxPD_replicate2



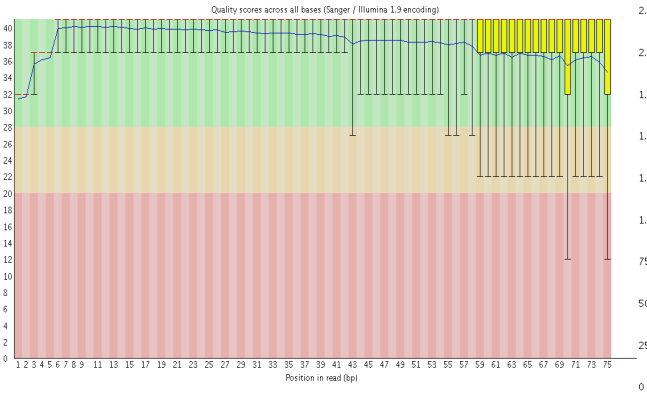


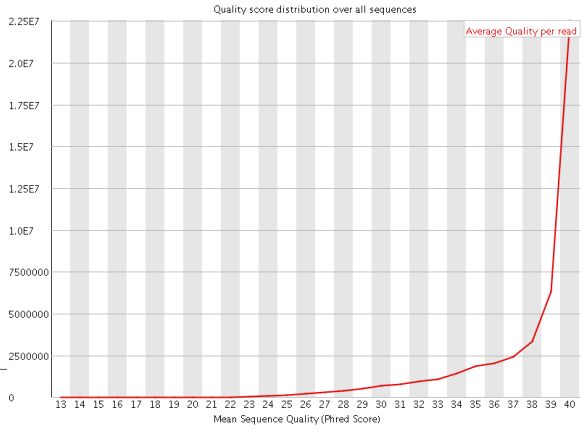
FoxPD_replicate3



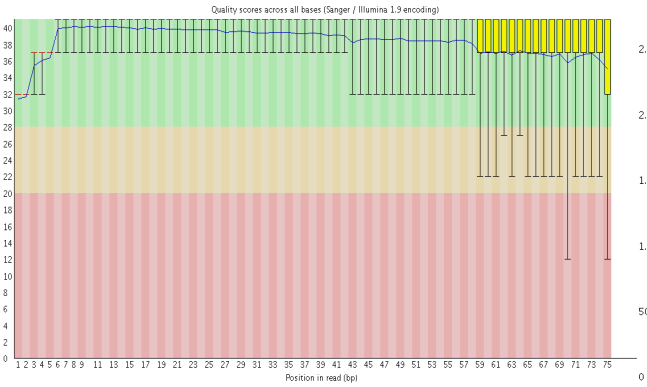


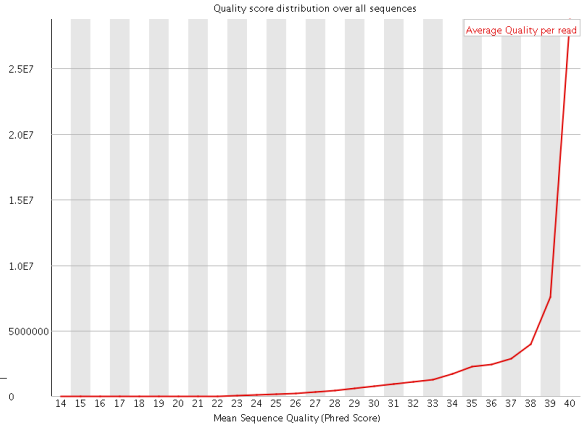
Control_replicate1



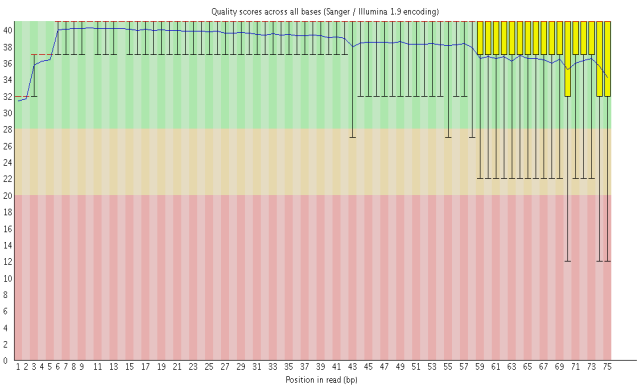


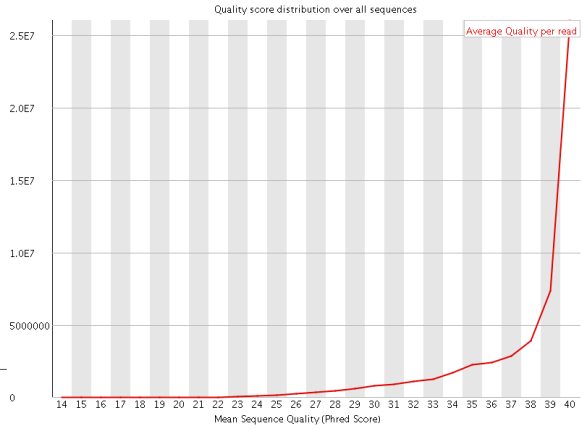
Control_replicate2





Control_replicate3





Appendix 3. List of candidate CRMs bound by FoxPC and FoxPD

Candidate CRMs for FoxPC are shown in the left table, while those for FoxPD are shown in the right one

CRM	normalised enrichment score	Possible TF
cisbp__M0723	5.71157	slp1
cisbp__M0086	5.58654	
cisbp__M1751	5.47277	
homer__CTCTCTCTCY_GAGA-repeat	5.41814	
cisbp__M0735	5.33844	foxo
cisbp__M0757	5.29889	
cisbp__M0756	5.26697	
cisbp__M0730	5.1997	foxo
cisbp__M0727	5.15143	
cisbp__M0099	4.97461	
cisbp__M0748	4.94298	
cisbp__M0721	4.87552	
cisbp__M0718	4.87435	foxK
cisbp__M0434	4.8347	
elemento__CTCTCTC	4.82089	
cisbp__M0087	4.72974	
cisbp__M0408	4.61557	
transfac_pro__M02049	4.57749	
transfac_pro__M00723	4.51492	Trl
cisbp__M2014	4.42328	Trl
flyfactorsurvey__Trl_FlyReg_FBgn0013263	4.42034	Trl
elemento__AGAGAGAG	4.38833	
tiffin__TIFDMEM0000009	4.3706	
elemento__AGAGAGCG	4.28141	
homer__KTGTTTGC_PHA	4.2678	
c2h2_zfs__M0404	4.25615	
cisbp__M0731	4.23167	foxo
transfac_pro__M00317	4.21953	
transfac_public__M00253	4.21865	
cisbp__M0752	4.19623	FoxP
cisbp__M0734	4.14962	croc
neph__UW	4.10439	
cisbp__M0732	4.07668	croc
transfac_pro__M02904	4.04251	Sox15
hdpi__NFIX	4.0241	Nfi
jaspar__MA0385.1	4.0146	SOK2
cisbp__M0473	4.01216	
cisbp__M0725	4.01167	FoxP
hdpi__DDX53	4.00951	CG7878
transfac_pro__M01621	3.99629	
hdpi__NOC2L	3.96957	
hdpi__BOLL	3.94049	
transfac_pro__M07665	3.92081	
transfac_pro__M01780	3.89819	
cisbp__M0378	3.84992	
cisbp__M0750	3.81643	
predrem__nrMotif1363	3.80987	
hdpi__ZNF671	3.7945	
cisbp__M0374	3.78687	
swissregulon__sacCer__SOK2	3.78148	
hdpi__WISP2	3.7805	
cisbp__M1483	3.75916	
transfac_pro__M01787	3.75475	
stark__RSWGAGMRHRR	3.74682	
hdpi__HLCS	3.72508	
hocomoco__SMAD1_HUMAN	3.72362	
yetfasco__YPR104C_2203	3.71852	
tfdimers__MD00590	3.71245	
hocomoco__FOXG1_HUMAN.H10MO.D	3.69268	
cisbp__M6017	3.67956	
cisbp__M0377	3.65488	
scertf__harbison.SOK2	3.65028	
cisbp__M0542	3.63246	
AD4_FOXI1_RCATWCCNNNNNNNNNNNNNR	3.62297	
hdpi__CANX	3.61807	
cisbp__M5455	3.61406	
transfac_pro__M03577	3.60759	
yetfasco__YMR016C_404	3.60094	
cisbp__M1033	3.60064	
tfdimers__MD00511	3.59017	
stark__ACANACA	3.58057	
transfac_pro__M02942	3.57519	
cisbp__M1485	3.57284	
hdpi__CENTG1	3.55443	
cisbp__M0290	3.55032	
transfac_pro__M01307	3.54288	
neph__UW.Motif.0336	3.54219	
hdpi__SMAD4	3.5417	
homer__CTYTCTYTCTCTCTC_GAGA-repeat	3.53387	
hdpi__FEZF2	3.52985	
hdpi__ZNF193	3.52848	
cisbp__M1301	3.47551	
yetfasco__YER109C_67	3.46719	
hdpi__CKMT1B	3.45671	
aipale__FOXJ2_DBD_RTAAACAWMAACA_rep	3.43517	
scertf__spivak.PHD1	3.43459	
yetfasco__YKL043W_2153	3.42401	
cisbp__M0370	3.41843	
transfac_pro__M07266	3.41784	
hdpi__INSM1	3.39464	
hdpi__HSPA1L	3.37633	
taipale__Foxj3_DBD_RTAAACATMAACA	3.37457	
cisbp__M0205	3.36948	
cisbp__M0628	3.36918	
predrem__nrMotif404	3.3685	
cisbp__M0375	3.36164	
hdpi__MYEF2	3.35959	
transfac_pro__M00624	3.35939	

CRM	normalised enrichment score	Possible TF
transfac_pro__M08193	6.21997	
cisbp__M0720	5.85305	
cisbp__M0721	5.26452	
cisbp__M0734	5.18648	croc
predrem__nrMotif2061	5.17741	
hdpi__RBM3	5.16611	tra2
cisbp__M0757	5.05071	
cisbp__M0732	5.04803	croc
cisbp__M0723	4.99451	slp1
cisbp__M0437	4.98285	CG16799
cisbp__M0718	4.97905	FoxK
c2h2_zfs__M0415	4.96765	CG16799
cisbp__M0727	4.96223	
swissregulon__hs__FOXL1.p2	4.92559	fd64A
cisbp__M0750	4.86492	FoxP
cisbp__M1564	4.82565	
jaspar__MA0210.1	4.7881	ara
bergman__Cf2	4.78267	Cf2
cisbp__M1563	4.70367	
jaspar__MA1058.1	4.68821	
cisbp__M1571	4.67442	
cisbp__M2544	4.67361	Cf2
transfac_public__M00013	4.67062	Cf2
tiffin__TIFDMEM0000034	4.63849	
stark__RTATATRTRB	4.61072	Cf2
jaspar__MA0015.1	4.55796	Cf2
cisbp__M1849	4.53916	Cf2
cisbp__M0752	4.52214	FoxP
cisbp__M2019	4.52006	ara
tiffin__TIFDMEM0000062	4.50425	fd64A
cisbp__M1865	4.48261	
cisbp__M0756	4.44318	
predrem__nrMotif2189	4.385	
jaspar__MA0233.1	4.38216	
cisbp__M0626	4.35505	
cisbp__M1570	4.34775	
cisbp__M0748	4.27498	
cisbp__M0725	4.26489	FoxP
cisbp__M4232	4.26236	
yetfasco__YDR081C_1050	4.26109	
hdpi__CKMT1B	4.25557	CG4546, CG3027
transfac_pro__M03850	4.21431	Sox102F
cisbp__M0614	4.19364	
cisbp__M1561	4.17068	
hdpi__MSI1	4.16146	Rbp6
cisbp__M0066	4.16065	
cisbp__M0875	4.16024	
cisbp__M0735	4.15431	foxo
cisbp__M0607	4.14488	Kdm2
cisbp__M2041	4.10657	mirr
transfac_pro__M03854	4.0561	
cisbp__M0749	4.04834	foxo
cisbp__M0440	4.04201	foxo, fd102C
cisbp__M0682	4.03664	
tiffin__TIFDMEM0000068	4.02711	
stark__GTATNWATA	4.01946	
cisbp__M1566	3.99178	
cisbp__M1130	3.98773	
cisbp__M1864	3.98256	
transfac_pro__M01216	3.96538	
swissregulon__sacCer__SPT15	3.95626	
cisbp__M0627	3.94237	
cisbp__M0503	3.93568	
flyfactorsurvey__Cf2-II_FlyReg_FBgn0000286	3.93528	
jaspar__MA1057.1	3.93508	
cisbp__M1572	3.92388	
jaspar__MA0379.1	3.92063	
swissregulon__sacCer__SIG1	3.92063	
transfac_pro__M01631	3.92038	
yetfasco__YER068W_556	3.89818	
flyfactorsurvey__Cf2-PB_SOLEXA_FBgn0000286	3.87451	
cisbp__M0408	3.82424	
cisbp__M1567	3.82252	
cisbp__M0847	3.81816	
cisbp__M0741	3.8096	
cisbp__M0898	3.80579	
stark__STATAWAWR	3.74767	
cisbp__M0278	3.7244	
cisbp__M1569	3.72192	
cisbp__M1167	3.71457	
cisbp__M0731	3.71376	
cisbp__M0231	3.71087	
cisbp__M0615	3.70611	
flyfactorsurvey__Mirr_Cell_FBgn0014343	3.69765	
transfac_pro__M00713	3.69668	
cisbp__M1734	3.65351	
flyfactorsurvey__Ara_Cell_FBgn0015904	3.60719	
predrem__nrMotif414	3.60607	
cisbp__M0821	3.59365	
cisbp__M0611	3.59071	
cisbp__M0635	3.58884	
cisbp__M0605	3.58408	
cisbp__M0411	3.58012	
cisbp__M1559	3.57627	
cisbp__M1162	3.56811	
cisbp__M1565	3.55463	
cisbp__M0628	3.54814	
transfac_pro__M01599	3.54475	
cisbp__M1556	3.541	
flyfactorsurvey__BtbVII_SANGER_FBgn0012049	3.53725	

Appendix 4 List of target genes for FoxPC and FoxPD

FoxPC targets

osa	5.62
CR46035	5.62
CR46035-RA	5.62
CycD	3.77
CG8909	3.77
CG30389	3.63
tRNA:CR30206	3.63
tRNA:G3:57BCa	3.63
tRNA:CR30208	3.63
CG4266	3.63
CG8405	3.61
ph-p	3.59
Trpgamma	3.28
sfl	3.25
CG4467	3.21
CG4449	3.21
elav	3.18
arg	3.18
CG14246	3.16
larp	3.14
osa	3
mam	2.98
Cow	2.94
Mnt	2.88
CG10910	2.85
crol	2.81
mld	2.81
aop	2.8
Mid1	2.78
ftz-f1	2.76
CR45223	2.75
CR45223-RB	2.75
CG14322	2.75
CR45223-RA	2.75
cic	2.69
CG30419	2.66
Glut4EF	2.66
Art4	2.66
CG18208	2.62
CR45633	2.62
CR45633-RA	2.62
CG31324	2.59
Dgk	2.57
Shal	2.56
UbcE2H	2.55
CG2258	2.55
sr	2.53
CG6678	2.53
fas	2.47
amon	2.47
CG34357	2.41
Mur82C	2.41
Uhg4	2.4
Uhg4-RA	2.4
CG8078	2.4
snoRNA:Or-CD2	2.4
CG8777	2.4
CG17230	2.4
Mid1	2.38
CG9967	2.37
eys	2.37
nAChRalpha6	2.36
CG30116	2.36
Glut4EF	2.36
Pka-R2	2.34
ftz-f1	2.34
CR45939	2.34
CR45939-RA	2.34
sdt	2.34
spen	2.33
CG34353	2.33
tra2	2.32
scaRNA:PsiU5-44	2.32
CG12868	2.32
CG33469	2.32
CG33468	2.32
Nmdmc	2.32

FoxPD targets

ph-p	4.83
osa	4.59
CR46035	4.59
CR46035-RA	4.59
elav	4.27
arg	4.27
CG8405	4.16
Trpgamma	4.03
Not1	4.03
CG1814	4.03
CR46257	3.96
CR46257-RA	3.96
crol	3.96
ftz-f1	3.96
CG30389	3.96
tRNA:CR30206	3.91
tRNA:G3:57BCa	3.78
tRNA:CR30208	3.78
CG4266	3.78
alpha-Man-IIa	3.78
ps	3.77
ko	3.67
CR44933	3.50
CR44933-RA	3.36
Sin3A	3.16
CR30055	3.16
CR30055-RC	3.16
CR30055-RB	3.13
CR30055-RD	3.09
crp	3.09
orb	3.08
CG14246	3.05
spi	3.05
jar	3.03
tRNA:CR31130	3.03
UbcE2H	3.03
CG2258	3.03
aop	3.03
Mid1	3.03
Pi3K92E	2.99
Lrrk	2.99
Fs	2.95
CR43730	2.95
CR43730-RB	2.91
Jhl-26	2.89
CG30100	2.84
CG42372	2.83
CG7747	2.82
sr	2.82
Kdm4B	2.81
CG4467	2.81
CG4449	2.81
skd	2.81
CG10585	2.81
Rdl	2.80
Axs	2.80
38231	2.80
let-7-C	2.77
let-7-C-RA	2.77
mir-100	2.77
let-7	2.77
Myo81F	2.75
CG11399	2.75
CG13624	2.73
EcR	2.72
AstC-R1	2.71
CR45246	2.69
CR45246-RA	2.69
CG42337	2.69
CG6678	2.69
tlk	2.69
Mnt	2.69
pHCl	2.69
CG3107	2.68
CG44837	2.68
CG5906	2.68
Sprn	2.65

Rel	2.32
Mst85C	2.32
alpha-Man-IIa	2.31
ps	2.31
CCKLR-17D3	2.31
CG4174	2.3
CG13380	2.3
CG32199	2.3
CR43635	2.27
CR43635-RA	2.27
rad	2.26
Mmp2	2.25
CR44070	2.25
CR44070-RA	2.25
CR44804	2.23
CR44804-RA	2.23
cv-c	2.23
sbb	2.22
ome	2.21
CG42666	2.21
mir-981	2.21
CG32806	2.21
CDase	2.2
PH4alphaEFB	2.2
CHES-1-like	2.19
X11Lbeta	2.18
futsch	2.17
Aps	2.16
Dscam3	2.16
Tusp	2.14
CG15572	2.14
spi	2.12
FucTA	2.12
EcR	2.11
Fife	2.11
rap	2.11
Pp2C1	2.11
CG15172	2.09
CR43700	2.09
CR43700-RA	2.09
Sidpn	2.09
Fur1	2.09
CG42540	2.08
wry	2.07
CG8641	2.07
CG34030	2.07
CR46017	2.07
CR46017-RA	2.07
CR44942	2.07
CR44942-RA	2.07
SoxN	2.06
CR45541	2.06
CR45541-RA	2.06
Hex-A	2.06
CG8746	2.05
CR44779	2.05
CR44779-RA	2.05
CR44779-RB	2.05
CR44779-RC	2.05
CG3655	2.05
CG15630	2.03
Not1	2.03
CG1814	2.03
CR46257	2.03
CR46257-RA	2.03
CG42795	2.03
Syt7	2.03
Or43a	2.02
Sin3A	2.02
CR30055	2.02
CR30055-RC	2.02
CR30055-RB	2.02
CR30055-RD	2.02
DIP-beta	2.02
Cip4	2.01
CR45054	2
CR45054-RB	2
CG30497	1.98
CG45093	1.98
jar	1.98
tRNA:CR31130	1.98
nAChRalpha3	1.98
Snoo	1.97
Pi3K92E	1.97
Lrrk	1.97
CG31324	1.97
Ac76E	2.64
Fife	2.63
mspo	2.63
CG43175	2.63
tRNA:CR31573	2.63
Epac	2.63
CG43060	2.63
PlexA	2.63
CG42666	2.58
CG32806	2.58
DIP-beta	2.58
Yeti	2.56
CG15803	2.56
CG11360	2.56
myo	2.56
CG17018	2.56
Cul5	2.55
CG11873	2.54
CG32850	2.53
PMCA	2.53
Ziz	2.53
Obp28a	2.53
RapGAP1	2.53
PVRAP	2.52
sdt	2.50
CYLD	2.50
CR44968	2.50
CR44968-RA	2.50
CR44968-RB	2.49
Sec71	2.49
CG4050	2.49
mago	2.49
Magi	2.49
CR44520	2.44
CR44520-RB	2.44
blanks	2.44
FucTA	2.43
CG33970	2.42
CR42765	2.39
CR42765-RB	2.39
CR42765-RC	2.39
4EHP	2.37
shot	2.36
DJ-1alpha	2.36
AGO1	2.36
CG16758	2.35
CR45425	2.34
CR45425-RA	2.33
Spn	2.33
jim	2.32
sta	2.32
inc	2.31
Nmdar2	2.31
Irp-1A	2.31
CR45648	2.30
CR45648-RA	2.30
CG4813	2.30
CG1344	2.30
Not3	2.30
CG4174	2.30
CG13380	2.30
not	2.30
tRNA:CR32200	2.29
Aps	2.29
Shal	2.29
pum	2.28
D1	2.26
brat	2.26
Nlg3	2.26
CR44332	2.26
CR44332-RA	2.26
CG6330	2.26
Tsp97E	2.25
mam	2.25
mir-4978	2.25
Glut4EF	2.25
kuz	2.25
Myo81F	2.24
Ca-alpha1D	2.24
lola	2.24
lola	2.23
CR45135	2.23
CR45135-RA	2.23
CG13055	2.23
CG13054	2.23
mub	2.22

CR46060	1.97
CR46060-RA	1.97
Sema-1a	1.96
CR44412	1.96
CR44412-RA	1.96
orb	1.96
Crys	1.95
CG13252	1.95
Gr77a	1.95
corto	1.95
plum	1.95
CR46083	1.95
CR46083-RA	1.95
CR44909	1.94
CR44909-RA	1.94
CR43730	1.94
CR43730-RB	1.94
Jhl-26	1.94
CG30100	1.94
CG42372	1.94
CG7747	1.94
FER	1.94
CR46145	1.94
CR46145-RA	1.94
CG1607	1.94
CG9135	1.93
CG13995	1.93
CG42238	1.93
CR46077	1.93
CR46077-RA	1.93
CG10222	1.92
flr	1.92
sgg	1.91
waw	1.9
bbx	1.9
CR45612	1.9
CR45612-RA	1.9
CR45611	1.9
CR45611-RA	1.9
sIgA	1.9
lilli	1.89
CG34302	1.89
CR42858	1.89
CR42858-RA	1.89
Teh1	1.89
CG1888	1.88
CG31221	1.88
CG42238	1.87
heph	1.87
RhoGAP19D	1.87
Art2	1.86
Sdc	1.86
CG11873	1.84
CG14760	1.83
beat-IIa	1.83
CG11873	1.83
CG42784	1.82
CR43856	1.82
CR43856-RA	1.82
GABA-B-R1	1.82
CG45186	1.82
sif	1.82
Rdl	1.82
CG42492	1.82
CG42342	1.81
CR44997	1.81
CR44997-RA	1.81
mAChR-A	1.8
CR44813	1.8
CR44813-RA	1.8
slik	1.8
pyd	1.79
CG31668	1.78
CG13784	1.78
ico	1.78
CR44989	1.78
CR44989-RA	1.78
beat-IIb	1.78
CG44257	1.78
spri	1.78
CG13624	1.77
CR45707	1.76
CR45707-RA	1.76
CG12769	1.76
insc	1.76
CG1275	1.76

scro	2.22
vri	2.21
CG32264	2.21
RapGAP1	2.21
CG42784	2.21
ome	2.21
CG17177	2.19
CG5890	2.19
Tsp96F	2.18
CG1671	2.18
PCB	2.18
Nprl3	2.18
upSET	2.18
Rim	2.17
14-3-3zeta	2.17
Pfk	2.17
mamo	2.17
CG15431	2.15
CG43055	2.15
nvd	2.15
Lk6	2.15
Cda5	2.15
CR45339	2.15
CR45339-RA	2.15
CR45340	2.14
CR45340-RA	2.14
dbr	2.13
Hr46	2.13
CG12912	2.12
fz	2.12
CR45025	2.11
CR45025-RB	2.11
CG13482	2.11
CG34354	2.11
Sxl	2.11
CG33309	2.10
cac	2.10
tna	2.09
CG6418	2.09
Ptp10D	2.09
RpS13	2.09
CSN8	2.08
CG17684	2.08
CG7881	2.08
CR44127	2.07
CR44127-RA	2.07
CR44127-RB	2.07
tra2	2.05
scaRNA:PsiU5-44	2.05
CG12868	2.05
CG33469	2.05
CG33468	2.05
hid	2.04
tinc	2.04
Lar	2.03
CG46244	2.03
Oda	2.03
CR46194	2.03
CR46194-RA	2.03
Cct5	2.03
Desat1	2.03
CG5538	2.03
CG18549	2.03
Dscam3	2.03
spen	2.03
RapGAP1	2.03
Npc1a	2.02
Prosalpha6	2.02
CG4908	2.01
Ca-alpha1D	2.01
CR45916	1.99
CR45916-RA	1.99
Shal	1.99
tok	1.99
Syt7	1.99
nAChRalpha3	1.99
Src42A	1.98
mbl	1.97
Rdl	1.96
CG6767	1.96
CR45804	1.95
CR45804-RA	1.95
CR45121	1.95
CR45121-RA	1.95
Sfp77F	1.93
nrm	1.93

CG2034	1.76
oxt	1.76
crp	1.75
CG34356	1.75
RhoGAP100F	1.75
let-7-C	1.74
let-7-C-RA	1.74
mir-100	1.74
Pde8	1.74
heph	1.74
CG8034	1.74
Ziz	1.73
Obp28a	1.73
RapGAP1	1.73
kek3	1.73
pk	1.73
wake	1.73
orb	1.73
zfh2	1.73
CR44443	1.72
CR44443-RA	1.72
resilin	1.72
CG5522	1.72
Eip63F-1	1.72
Task6	1.72
CG9588	1.72
tinc	1.72
trio	1.71
sif	1.71
CG31523	1.71
CG34380	1.7
Rgl	1.7
Ten-m	1.7
CG32450	1.7
alpha-Man-Ia	1.7
CG2909	1.7
bsk	1.69
CG31717	1.69
Ror	1.69
PVRAP	1.69
CG34347	1.69
MTA1-like	1.68
elB	1.67
CG44837	1.67
CG5906	1.67
CG15431	1.66
ttv	1.66
CG45263	1.66
LRP1	1.65
Tina-1	1.65
CG15861	1.65
CG15803	1.65
CG4983	1.64
CG43167	1.64
mbl	1.64
CG17896	1.64
CG17778	1.64
Dyrk2	1.63
CG42337	1.63
Mical	1.63
CG3909	1.63
CG5382	1.63
CG5380	1.63
PyK	1.63
dsx	1.63
CG4078	1.63
Tudor-SN	1.62
snoRNA:Psi18S-996	1.62
mRpL17	1.62
jim	1.62
Dmtn	1.62
gpp	1.62
GPHR	1.61
CG17839	1.61
CR44702	1.61
CR44702-RA	1.61
CG34353	1.61
CG14434	1.61
CG43736	1.61
jing	1.6
Fife	1.6
Sox102F	1.6
CG15630	1.59
Fas3	1.59
trio	1.59
scrt	1.58
ci	1.92
CR43957	1.91
CR43957-RB	1.90
CR43957-RA	1.90
Mmp2	1.90
CR44070	1.90
CR44070-RA	1.90
Sema-2b	1.89
CG4398	1.89
CR45400	1.89
CR45400-RA	1.89
Pdk1	1.89
CR45805	1.88
CR45805-RA	1.88
Spn	1.88
CG32295	1.88
CR43087	1.88
CG7408	1.88
CG15186	1.88
Cals	1.88
rdo	1.88
CG17684	1.88
Mapmodulin	1.88
CR44417	1.88
CR44417-RA	1.88
scrib	1.87
Scp1	1.87
mbl	1.87
sbb	1.87
klu	1.87
bru-3	1.87
Myo81F	1.87
CG13954	1.86
CG13739	1.86
CG14330	1.85
Pez	1.85
Cpr	1.85
CG40006	1.85
CG4424	1.85
CG4854	1.85
trem	1.84
CG11155	1.83
CG32017	1.83
Nipped-A	1.83
dpr2	1.83
CG33129	1.83
exp	1.83
Drl-2	1.83
Vps60	1.83
CG7510	1.83
CR44322	1.82
CR44322-RA	1.81
CG42827	1.81
CG42828	1.81
CG4393	1.80
raw	1.80
MESR3	1.79
Hsromega	1.79
Hsromega-RF	1.79
Hsromega-RB	1.78
Hsromega-RG	1.78
Actbeta	1.78
zyd	1.78
Piezo	1.77
l(2)35Cc	1.77
yuri	1.77
CG30497	1.77
CG45093	1.77
CR44510	1.77
CR44510-RA	1.77
CR44511	1.77
CR44511-RA	1.77
CR43150	1.77
CR43150-RB	1.77
CR43988	1.77
CR43988-RA	1.76
CG13405	1.75
FMRFaR	1.75
kst	1.75
beat-IIIc	1.74
CR44396	1.74
CR44396-RA	1.74
CR44395	1.74
CR44395-RA	1.74
CR43431	1.74
CR43431-RA	1.73

Eip78C	1.58
sdt	1.58
hang	1.58
CG14234	1.58
twe	1.57
CG4935	1.57
crp	1.57
CG17486	1.57
Ggamma30A	1.56
CR45262	1.56
CR45262-RA	1.56
CG43729	1.56
tau	1.56
Mlc1	1.56
ttv	1.55
mv	1.55
SCOT	1.55
CR45827	1.55
CR45827-RA	1.55
CR45828	1.55
CR45828-RA	1.55
Cip4	1.55
mRpS6	1.55
CG33514	1.55
puc	1.55
CR44331	1.55
CR44331-RA	1.55
CR44331-RB	1.55
fru	1.55
CG34356	1.54
CR45732	1.54
CR45732-RA	1.54
CG33062	1.54
Glut4EF	1.54
CG2113	1.53
Cpr65Ea	1.53
Cpr65Eb	1.53
Rbp6	1.53
5-HT2B	1.53
CG4467	1.53
ftz-f1	1.52
CR45939	1.52
CR45939-RA	1.52
CG14073	1.52
Tusp	1.52
Cul5	1.52
CG11873	1.52
CG43658	1.52
CR44651	1.51
CR44651-RA	1.51
Mur89F	1.51
Neto	1.51
Mnt	1.5
CR44997	1.5
CR44997-RA	1.5
CR44143	1.49
CR44143-RA	1.49
sm	1.49
px	1.49
CG11362	1.49
dnr1	1.49
Dic61B	1.49
p130CAS	1.49
Glut1	1.49
CR44935	1.49
CR44935-RA	1.49
ps	1.49
CG4810	1.48
CG43450	1.48
CG3809	1.48
inc	1.48
Nmdar2	1.48
CG42265	1.48
CG42339	1.48
CG7848	1.47
CG13295	1.47
CG6610	1.47
ssp6	1.47
CG32432	1.47
Kap3	1.47
CG34348	1.47
mtt	1.46
CG13954	1.46
CG4050	1.46
mago	1.46
Magi	1.46
CG32945	1.73
CG43427	1.73
CG33927	1.73
bru-2	1.73
mtt	1.73
knrl	1.72
CG17514	1.72
PlexB	1.72
CR45124	1.72
CR45124-RA	1.71
CR44440	1.71
CR44440-RA	1.71
CG42674	1.71
CR43685	1.70
CR43685-RA	1.70
CR43685-RB	1.70
ps	1.70
CG18208	1.70
GluClalpha	1.70
heph	1.70
CG34393	1.70
CR41257	1.69
CR41257-RB	1.69
rl	1.69
CG32432	1.68
MstProx	1.68
CG17816	1.68
CR43456	1.68
CR43456-RA	1.68
CR43457	1.67
CR43457-RA	1.67
CR45651	1.67
CR45651-RA	1.67
CG32264	1.67
ppk27	1.67
CR45368	1.67
CR45368-RA	1.67
CG18518	1.67
CG12017	1.66
CG13827	1.66
orb	1.66
CR44949	1.66
CR44949-RA	1.66
nompB	1.66
CG8401	1.65
SP2353	1.65
UQCR-11	1.65
ctrip	1.65
5-HT2B	1.65
CR46124	1.65
CR46124-RA	1.65
Syn	1.65
CG12420	1.65
CG12817	1.64
X11Lbeta	1.64
Or65c	1.64
CG13300	1.63
Rdl	1.63
CG34356	1.63
bru-3	1.63
Rbp6	1.63
dpr21	1.63
CG43064	1.63
CR44702	1.63
CR44702-RA	1.63
CR45398	1.63
CR45398-RA	1.63
lap	1.63
CG14609	1.63
Trpm	1.63
A2bp1	1.63
miple1	1.62
miple2	1.62
CG32845	1.62
rpr	1.62
CG42666	1.62
Adar	1.62
CG1494	1.62
BicC	1.62
CG45084	1.62
beat-Ia	1.62
CG44837	1.62
CR43624	1.60
CR43624-RA	1.60
Fur1	1.60
fred	1.60

Nrt	1.46
Cdep	1.46
Act5C	1.46
CG15431	1.45
tRNA:CR31963	1.45
CG3124	1.45
CG13541	1.45
CR43334	1.45
CR43334-RA	1.45
CR43334-RC	1.45
app	1.45
CG3339	1.45
Ehbp1	1.44
CG8963	1.44
TyrRII	1.44
CadN	1.43
Sox14	1.43
dpr6	1.43
bru-3	1.43
AstC-R1	1.43
CR44126	1.42
CR44126-RA	1.42
slim	1.42
CG5189	1.42
CG30120	1.42
Prp19	1.42
px	1.42
nkd	1.42
CG43175	1.42
tRNA:CR31573	1.42
tinc	1.42
Pdcd4	1.42
kek2	1.41
CG33309	1.41
CG42673	1.41
Glut4EF	1.41
CR43847	1.41
CR43847-RA	1.41
Ca-alpha1T	1.41
CG5059	1.4
CR46061	1.4
CR46061-RB	1.4
Ald	1.4
CG44422	1.4
caps	1.39
Cad87A	1.39
CG14731	1.39
alpha-Man-IIb	1.39
CR45211	1.39
CR45211-RA	1.39
oxt	1.38
CG13807	1.38
ecd	1.38
CG43778	1.37
CG30158	1.37
nvy	1.37
CG8138	1.37
CG8508	1.37
Appl	1.37
CG11319	1.36
CR44371	1.36
CR44371-RA	1.36
Hr51	1.36
Gug	1.36
CG31191	1.36
hdly	1.36
CG43867	1.36
rad	1.36
fdl	1.35
CG42260	1.35
dpr11	1.35
klg	1.35
IA-2	1.34
S	1.34
l(1)G0007	1.34
Ady43A	1.33
CG6149	1.33
CR44548	1.33
CR44548-RA	1.33
CG32206	1.33
TkR86C	1.33
CG14691	1.33
kay	1.33
RhoU	1.33
CR44894	1.33
CR44894-RB	1.33
Spn28Db	1.60
Snoo	1.60
Rh7	1.60
CCAP-R	1.60
CG14540	1.60
CG16848	1.60
CG16956	1.60
CG42784	1.60
CG43850	1.60
CR44932	1.60
CR44932-RA	1.60
Antp	1.59
CCAP-R	1.59
bru-2	1.59
Oatp74D	1.59
pzg	1.59
Cpr78Ca	1.58
Myo81F	1.58
ATP8B	1.58
RluA-1	1.58
CG31814	1.58
CR44696	1.57
CR44696-RA	1.57
CG43362	1.57
CR44693	1.57
CR44693-RA	1.57
Sfp36F	1.57
CR43358	1.57
CR43358-RB	1.57
CR43358-RA	1.57
Cht3	1.57
mim	1.56
CheB42c	1.56
CR32111	1.56
CR32111-RC	1.56
CR32111-RD	1.56
CG45263	1.56
CG31323	1.56
alpha-Man-Ia	1.56
kek2	1.56
dimm	1.56
FucTA	1.56
Msh6	1.56
CG6959	1.55
CG2772	1.55
A2bp1	1.55
JIL-1	1.55
Hnf4	1.55
CG4168	1.55
BBS4	1.54
Atg1	1.54
CG17839	1.54
CR44584	1.54
CR44584-RA	1.54
CG5681	1.54
Prosbeta5R2	1.54
sxc	1.54
toc	1.54
tai	1.54
rk	1.54
CadN2	1.54
CG17018	1.54
DCX-EMAP	1.54
CG31345	1.54
CG31183	1.53
CG42673	1.53
slo	1.53
fs(1)h	1.52
mys	1.52
PAPLA1	1.52
CG43394	1.51
CG31606	1.51
CG42671	1.50
CR46050	1.50
CR46050-RA	1.50
Gbs-76A	1.50
fal	1.50
CR46048	1.50
CR46048-RA	1.50
Irp-1A	1.50
CR45648	1.50
CR45648-RA	1.50
CG4813	1.49
CG45049	1.49
KrT95D	1.49
CG5888	1.49

CR44894-RA	1.33
Sema-1a	1.32
fra	1.32
CR44535	1.32
CR44535-RB	1.32
CR44535-RA	1.32
PVRAP	1.32
Eip75B	1.32
Khc-73	1.31
CG42268	1.31
CR45407	1.31
CR45407-RA	1.31
CG34451	1.31
stumps	1.31
CG14853	1.31
CG16791	1.31
DIP-gamma	1.31
CR45669	1.31
CR45669-RA	1.31
AstA-R1	1.31
fas	1.3
Mid1	1.3
CR44828	1.3
CR44828-RA	1.3
Ncc69	1.3
CG42342	1.3
CG7080	1.3
CG43137	1.3
CR45352	1.29
CR45352-RA	1.29
CR31840	1.29
CR31840-RB	1.29
brat	1.29
shn	1.29
CG11537	1.29
CR45819	1.29
CR45819-RA	1.29
bol	1.29
CG43060	1.29
mAChR-B	1.29
CG15186	1.28
crp	1.27
Mlh1	1.27
CG14757	1.27
LRP1	1.27
AGO1	1.27
Rpn13	1.27
Rh50	1.27
CG14669	1.27
Calx	1.27
CG30419	1.26
CG3967	1.26
Oatp74D	1.26
edin	1.26
CG34347	1.26
CG42541	1.26
inaE	1.26
AMPdeam	1.26
mamo	1.26
fwd	1.25
klg	1.25
Smr	1.25
CR43963	1.25
CR43963-RA	1.25
CR43963-RB	1.25
CG32506	1.25
Su(H)	1.24
CadN	1.24
mbl	1.24
mbl	1.24
bru-3	1.24
tou	1.23
CG43729	1.23
CG30472	1.23
dally	1.23
Ccn	1.23
Oct-TyrR	1.23
Dscam3	1.23
ATbp	1.23
GlcAT-P	1.22
SKIP	1.22
beat-IIa	1.21
lmd	1.21
CG17198	1.21
CG17197	1.21
woc	1.21
brat	1.49
CG43367	1.49
Cpr64Aa	1.49
Deaf1	1.48
trc	1.48
eag	1.48
GalNAc-T1	1.48
Khc-73	1.48
tn	1.48
cpx	1.47
CG9780	1.47
mir-929	1.47
rdx	1.46
Irk1	1.46
twi	1.46
CG9896	1.46
bru-3	1.46
bmm	1.46
Tdrd3	1.46
CG34383	1.45
cnc	1.45
fzo	1.45
CG31693	1.45
CG14879	1.45
Manf	1.45
Gyc32E	1.45
CG6287	1.45
Rdl	1.45
btsz	1.45
kek2	1.45
spag4	1.45
Fas3	1.44
Galphao	1.44
Roc2	1.44
tRNA:M2:48Bb	1.44
CG4250	1.44
CG42867	1.44
CR45269	1.44
CR45269-RA	1.44
CR42868	1.44
CR42868-RB	1.44
CG30269	1.44
CG30273	1.44
pros	1.44
CR46008	1.43
CR46008-RA	1.43
Ptp99A	1.43
CG33978	1.42
CR45198	1.42
CR45198-RA	1.42
CR44596	1.41
CR44596-RA	1.41
CR44595	1.40
CR44595-RA	1.40
CR43357	1.40
CR43357-RB	1.40
CR43357-RA	1.40
fry	1.40
Fbp1	1.40
CG32206	1.40
ms(3)76Ba	1.40
jim	1.40
CG31128	1.40
CG5888	1.40
CG34362	1.40
CR44488	1.40
CR44488-RA	1.40
CG3262	1.39
RpL21	1.39
bol	1.39
CG31760	1.39
CG31861	1.39
l(2)41Ab	1.39
CR44043	1.38
CR44043-RA	1.38
CR44546	1.38
CR44546-RA	1.38
Cad86C	1.38
CR46218	1.38
CR46218-RA	1.38
Cpr92F	1.38
Nipped-A	1.37
hipk	1.37
CG13978	1.37
CG10348	1.37
CG40498	1.36

Appendix 5. List of ontology terms associated with target genes of FoxPC and FoxPD

GO term associated with FoxPC targets	p-value
nervous system development	5.28E-21
system development	1.31E-20
generation of neurons	1.48E-19
neurogenesis	1.24E-18
neuron differentiation	1.27E-17
single-multicellular organism process	2.22E-16
biological regulation	2.41E-16
neuron development	2.61E-16
multicellular organism development	1.66E-13
neuron projection development	3.35E-13
neuron projection morphogenesis	7.20E-13
cell projection morphogenesis	8.04E-13
plasma membrane bounded cell projection morphogenesis	8.04E-13
anatomical structure morphogenesis	8.15E-13
animal organ development	8.99E-13
regulation of biological process	9.30E-13
cell part morphogenesis	2.96E-12
cell differentiation	3.26E-12
cell development	3.53E-12
axon development	7.86E-12
cell morphogenesis	9.04E-12
regulation of cellular process	9.04E-12
cell morphogenesis involved in differentiation	1.75E-11
axonogenesis	1.94E-11
response to chemical	2.38E-11
cell morphogenesis involved in neuron differentiation	3.04E-11
cellular developmental process	4.43E-11
axon guidance	4.46E-11
cellular component morphogenesis	6.88E-11
chemotaxis	8.21E-11
plasma membrane bounded cell projection organization	1.70E-10
neuron projection guidance	1.97E-10
cell projection organization	3.28E-10
developmental process	3.63E-10
multicellular organismal process	6.21E-10
anatomical structure development	6.91E-10
locomotion	7.46E-10
animal organ morphogenesis	7.66E-10
taxis	8.01E-10
response to stimulus	1.15E-09
single-organism developmental process	1.33E-09
cell-cell signaling	1.53E-09
single organism signaling	1.62E-09
signaling	1.88E-09
cell communication	7.76E-09
regulation of cell development	1.37E-08
regulation of multicellular organismal process	1.45E-08
regulation of biological quality	1.84E-08
neuron recognition	2.05E-08
regulation of nervous system development	4.59E-08
behavior	5.26E-08
regulation of neurogenesis	5.46E-08
cell recognition	5.71E-08
appendage development	5.92E-08
appendage morphogenesis	6.47E-08
sensory organ development	9.31E-08
imaginal disc-derived appendage development	1.05E-07
cell adhesion	1.28E-07
imaginal disc-derived appendage morphogenesis	1.47E-07
regulation of multicellular organismal development	1.53E-07
post-embryonic animal organ development	1.58E-07
eye development	1.68E-07
regulation of cell differentiation	1.90E-07
biological adhesion	2.09E-07
cell-cell adhesion	2.18E-07
cell-cell adhesion via plasma-membrane adhesion molecules	4.35E-07
regulation of developmental process	9.92E-07
tube development	1.1716E-06
wing disc development	1.26879E-06
imaginal disc development	1.5621E-06
post-embryonic appendage morphogenesis	1.63946E-06
imaginal disc-derived wing morphogenesis	1.79636E-06
movement of cell or subcellular component	2.87833E-06
tissue development	3.52312E-06
morphogenesis of an epithelium	3.92623E-06
compound eye development	4.09564E-06
epithelium development	4.57402E-06
wing disc morphogenesis	4.80492E-06
tissue morphogenesis	5.97554E-06

GO term associated with FoxPD targets	p-value
neuron development	5.25E-15
single organism signaling	3.05E-14
nervous system development	3.24E-14
signaling	3.55E-14
biological regulation	4.52E-14
cell communication	2.02E-13
neuron differentiation	4.56E-13
neurogenesis	1.09E-12
generation of neurons	1.25E-12
regulation of biological process	2.28E-12
system development	2.56E-12
neuron projection development	4.14E-12
regulation of cellular process	7.28E-12
signal transduction	1.29E-11
neuron projection morphogenesis	1.81E-11
cell projection morphogenesis	1.98E-11
plasma membrane bounded cell projection morphogenesis	1.98E-11
cell part morphogenesis	5.76E-11
cell morphogenesis involved in neuron differentiation	8.42E-11
single-multicellular organism process	1.18E-10
cell morphogenesis involved in differentiation	5.22E-10
multicellular organism development	8.07E-10
cell morphogenesis	9.82E-09
axon development	1.40E-08
axonogenesis	1.41E-08
response to stimulus	3.52E-08
cell differentiation	4.04E-08
plasma membrane bounded cell projection organization	4.23E-08
cell projection organization	7.03E-08
cell development	2.21E-07
cellular developmental process	2.65E-07
behavior	3.09E-07
anatomical structure development	3.19E-07
single-organism developmental process	3.54E-07
developmental process	6.45E-07
cellular response to stimulus	9.69E-07
chemotaxis	1.27201E-06
neuron projection guidance	1.27393E-06
cellular component morphogenesis	1.48294E-06
anatomical structure morphogenesis	1.62097E-06
multicellular organismal process	1.66154E-06
axon guidance	1.77335E-06
response to chemical	1.82847E-06
taxis	3.90954E-06
cell-cell signaling	8.90043E-06
regulation of biological quality	3.18988E-05
locomotion	4.52737E-05
synapse organization	0.000183434
single-organism process	0.000205328
animal organ development	0.000227171
regulation of multicellular organismal process	0.000357908
single-organism behavior	0.000398885
intracellular signal transduction	0.000499381
regulation of multicellular organismal development	0.000652495
cell-cell adhesion via plasma-membrane adhesion molecules	0.000655893
animal organ morphogenesis	0.000980025
growth	0.001193984
cell surface receptor signaling pathway involved in cell-cell signaling	0.001211559
embryonic morphogenesis	0.001232776
single-organism cellular process	0.001391257
developmental growth	0.00199373
cell-cell adhesion	0.003626733
photoreceptor cell development	0.004068081
regulation of cell communication	0.004074424
regulation of signaling	0.004254861
embryo development	0.004786078
negative regulation of nucleobase-containing compound metabolic process	0.006422828
cell adhesion	0.008007106
morphogenesis of an epithelium	0.009405413
biological adhesion	0.010665677
eye photoreceptor cell development	0.011669589
regulation of developmental process	0.016346724
movement of cell or subcellular component	0.018712144
eye development	0.018856038
regulation of nucleobase-containing compound metabolic process	0.019751338
tissue morphogenesis	0.021758115
negative regulation of biosynthetic process	0.021796607
negative regulation of cellular biosynthetic process	0.021796607
regulation of locomotion	0.022432627

GO term associated with FoxPC targets	p-value
signal transduction	6.37405E-06
regulation of cell communication	8.6609E-06
regulation of signaling	9.2167E-06
single-organism behavior	9.48559E-06
imaginal disc morphogenesis	1.28588E-05
post-embryonic animal organ morphogenesis	1.28588E-05
single-organism process	1.31405E-05
cell surface receptor signaling pathway involved in cell-cell signaling	1.75261E-05
post-embryonic development	1.7699E-05
metamorphosis	2.31305E-05
cellular response to stimulus	2.31557E-05
regulation of neuron differentiation	6.30369E-05
synaptic signaling	8.07043E-05
instar larval or pupal morphogenesis	9.77209E-05
tube morphogenesis	0.000103043
regulation of response to stimulus	0.000108741
epithelial tube morphogenesis	0.000119373
post-embryonic animal morphogenesis	0.000120261
single-organism cellular process	0.000125241
negative regulation of cellular process	0.000161159
regulation of signal transduction	0.000169636
regulation of cell morphogenesis involved in differentiation	0.000189741
chemical synaptic transmission	0.000211581
anterograde trans-synaptic signaling	0.000211581
trans-synaptic signaling	0.000211581
eye morphogenesis	0.00025043
sensory organ morphogenesis	0.00025043
positive regulation of cellular process	0.0004746
cellular process	0.000594593
synapse organization	0.000643829
negative regulation of biological process	0.000668965
homophilic cell adhesion via plasma membrane adhesion molecules	0.000693293
instar larval or pupal development	0.000870034
establishment or maintenance of cell polarity	0.00089762
regulation of neuron projection development	0.001214845
Wnt signaling pathway	0.001218025
cell-cell signaling by wnt	0.001418103
compound eye morphogenesis	0.001657964
photoreceptor cell differentiation	0.001711157
learning or memory	0.001727692
cognition	0.001727692
synaptic target recognition	0.002563092
neurological system process	0.002891854
positive regulation of biological process	0.002998092
regulation of cellular component organization	0.003403242
regulation of plasma membrane bounded cell projection organization	0.004538689
cell fate commitment	0.00500332
regulation of cell projection organization	0.005269228
negative regulation of cell communication	0.005931852
negative regulation of signaling	0.005931852
photoreceptor cell development	0.006712068
regulation of axonogenesis	0.007344559
regulation of anatomical structure morphogenesis	0.007360838
regulation of cell morphogenesis	0.008856635
pattern specification process	0.009196699
growth	0.012292311
response to external stimulus	0.013868654
olfactory behavior	0.01683046
cell surface receptor signaling pathway	0.033417114
negative regulation of signal transduction	0.03467176
heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules	0.035165125
synaptic target attraction	0.035625875
motor neuron axon guidance	0.038193431
system process	0.038662026
eye photoreceptor cell development	0.04522991
eye photoreceptor cell differentiation	0.04942084

GO term associated with FoxPD targets	p-value
epithelium development	0.027557814
synaptic signaling	0.028974041
regulation of primary metabolic process	0.029050984
regulation of nervous system development	0.030659151
negative regulation of cellular process	0.031521806
regulation of growth	0.032048499
R7 cell development	0.032684883
axon extension	0.03352949
compound eye photoreceptor development	0.035229766
regulation of signal transduction	0.038304588
regulation of developmental growth	0.040820726
negative regulation of RNA metabolic process	0.041094526
regulation of cellular metabolic process	0.048078721

Appendix 6. Primer List

Name	Application	Forward Primer	Reverse primer
Ph-p	Reporter Assays	GGGGACAACTTTGTATAGAAAAGTTGGCATTTACTTTTACTAGT	GGGGACTGCTTTTTTGTACAAACTTGAGTTGTGACCTCCTCCTTCGCCG
osa	Reporter Assays	GGGGACAACTTTGTATAGAAAAGTTGTGAGTTAAGTTTCCGTTTAA	GGGGACTGCTTTTTTGTACAAACTTGATAATTGCCGCGCCTGCTTGG
ftz-f1	Reporter Assays	GGGGACAACTTTGTATAGAAAAGTTGATTTGAAAACAAATATCGAT	GGGGACTGCTTTTTTGTACAAACTTGATGTGTGATTCATTTCGC
Mam	Reporter Assays	GGGGACAACTTTGTATAGAAAAGTTGGATCGGTTATGTTTGGTAACT	GGGGACTGCTTTTTTGTACAAACTTGAGACAATTTTCGAGCGCTA
Mnt	Reporter Assays	GGGGACAACTTTGTATAGAAAAGTTGAATACCAGCACGAGAAATAC	GGGGACTGCTTTTTTGTACAAACTTGACTCGCTTTTTCCAAGCGATT
Crol	Reporter Assays	GGGGACAACTTTGTATAGAAAAGTTGATGGGCAATTTATATTACTT	GGGGACTGCTTTTTTGTACAAACTTGACAGTGGTCACAACAGGAGTTA
Trpγ	Reporter Assays	GGGGACAACTTTGTATAGAAAAGTTGAACTGTCCGTATTTTAATTAA	GGGGACTGCTTTTTTGTACAAACTTGACTTTGGCCAAATAATTTTGG
Shal	Reporter Assays	GGGGACAACTTTGTATAGAAAAGTTGCAGTTTGTCGAAAACTGTT	GGGGACTGCTTTTTTGTACAAACTTGAGCGATTGTCGATTTAAATAGGC
Shaker	Reporter Assays	GGGGACAACTTTGTATAGAAAAGTTGGTGATTTACTTTGGCCATT	GGGGACTGCTTTTTTGTACAAACTTGACTACTGTTGGTTGCGTGTCGCG
gcm	Reporter Assays	AGGAGAATTCATGGACTACAAAGACGATGACGACAAGGTTTTGAACGG CATGCCTAT	TGCTTAGGTACCGCAATAGATGGGATCCGTGCTG
spps	Reporter Assays	AGGAGAATTCATGGACTACAAAGACGATGACGACAAGCTGCAGTACAA GAGTGCCG	TGCTTAGGTACCGTTGGACGAATTGTCGCCAG
Dichaete	Reporter Assays	AGGAGAATTCATGGACTACAAAGACGATGACGACAAGGCCACCTTATC GACACACCC	TGCTTAGGTACCATAGAGCACCGGAACCGGTCG
FoxPC	Reporter Assays	AGCCGGAATTCATGTACCCATACGATGTTCCAGATTACGCTCATCGGA TACATGACGACGAG	TTATCGATTGTGCTCATTGGCACTAGGTACCCCGGA
acj6	Reporter Assays	CCGGAATTCATGACAATGTCGATGTATTCGA	CGGGGTACCGTATCCAAATCCCGCCGAACC

GAPDH	TRAP	CTTCTTCAGCGACACCCATT	ACCGAACTCGTTGTCGTACC
TBP	TRAP	ACAGGGGCAAAGAGTGAGG	CTTAAAGTCGAGGAACTTTGCAG
EF1	TRAP	GCGTGGGTTTGTGATCAGTT	GATCTTCTCCTTGCCCATCC
Ph-p	TRAP	TGGGCCAAACCAGATCAT	GCGTCACGTTGCATTGAATA
osa	TRAP	ACCGTCACAGCAGTACAGGA	GGCGGATAGGGTGACTGAC
ftz-f1	TRAP	AAGCAGGAAATTCAAATACCTCA	CACCTGTCCCAACGCAAT
Mam	TRAP	GCAACAAAAATTCCTGAAACG	GCCGTTGTTGTTGTTGCTATT
Mnt	TRAP	GGCCCACAAATATGTCAATTC	TCTGCTTGGCCAGTTTCTCT
Crol	TRAP	ACACCCAAATGGATCAACTGT	GTCGACTGCTTTCGCTGAG
Trpγ	TRAP	GGGAACATCTAAAGACAATTCGAC	CGCATTATCTGCTGGTACTTGA
Shal	TRAP	GATGAGGTGGGAAAAGTTGC	TGAACTTATTGCTGTCATTTTGC
Shaker	TRAP	TTAAGTGTTTCCCGACTTGCC	CTTTTTCTCTCTCTGCACGTGC
Ph-p	ChIP	CCCCTTCTCAACCGACAC	TCTTAAATATTTATTTGTA
osa	ChIP	AATGCGTGAATATTCTTCTATGT	TAATTGCCGCGCCTGCTT
ftz-f1	ChIP	GCGTGTCAGTGTGTGCGTA	TCTCTCTTTCTCCGTCT
Mam	ChIP	GTAAAATGTTTATCTATTTAAA	GATAAAAACCGCGTCCGACT
Mnt	ChIP	TGTTGGTATTCGTTTGAGAAT	CGTAAAACTTGGCGCTCGTT
Crol	ChIP	AACTATATCGCTGGTGAGTGCA	CCAGCACTGGCAAGAGGTGT
Trpγ	ChIP	TTTCGGTGGATCACTAAAC	GGGCGAAAAAAACCGTTGGT
Shal	ChIP	GTAAGAGGCGAGAAGTGTAAAA	AATGAGCTAACCGCATGGCAAGG