

**An investigation into Mef2-interacting
proteins during embryonic muscle
development of *Drosophila***

Alina Anamaria Cepraga

Cardiff University

PhD, 2020



DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed (candidate) Date

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD

Signed (candidate) Date

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third Party Editors by Research Degree Students. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed (candidate) Date

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed (candidate) Date

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loans **after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.**

Signed (candidate) Date

Acknowledgements

First I would like to thank my supervisor Michael Taylor for sharing his invaluable guidance and unwavering dedication for research that has guided me throughout my PhD and has shaped me into the scientist I am today. I would also like to thank Eileen Furlong at EMBL Heidelberg for hosting me in her lab to carry out a significant part of my work and for her invaluable advice in solving challenges. Thank you to Anne Claude Gavin at EMBL Heidelberg for sharing reagents and her labs knowledge on protein purification. Thank you to Nick Kent for always providing sensible feedback and encouragement. I am grateful to Helen White-Cooper for her advice and support throughout the years and especially in her help finalising my PhD.

Thanks of course to fly labmates past and present – Clare, David (x2), Gaia, Gordon, Jesús, Karen, Kevin, Leanne, Louise, Ryan, Sami, Sarah, Simona and to everyone else in W3.25, Cardiff. Thank you for your support, friendship and being a strong support system. To the people who have made EMBL a wonderful experience - Aleksander, Anna, Bingqing, Daniel, Denisa, Eva, Francis, Ignacio, James, Juan, Lucia, Matilda, Olga, Pierre, Raquel, Sourabh, Srishti, Tibor – thank you for the insightful discussions and all the coffee breaks we shared.

I am grateful also to my review committee, Nick Brown and Karl Swann, who have taken the time to assess my work and offer excellent feedback that helped improve the quality of this thesis.

My work would not be completed without the support of Peter Watson, Kerrie Thomas and the entire school team who has offered reliable advice and assistance in complex matters. Special thanks go to Martina and Maria who have helped me find practical solutions to impossible situations.

Credit also goes to all the people who have listened and offered advice and support, the fly kitchen ladies, mass spectrometry core facility and other support staff who have enabled my work.

I am grateful to my family for always believing in me and supporting every goal I have had. And thank you Bene, you will no longer be the only Dr in the Haus!

It took a team effort to get here, but WE DID IT!

Summary

The present thesis was an investigation into developing a full picture of the Mef2-interacting proteins in muscle differentiation of *Drosophila melanogaster* by using a mixture of wet-lab and computational approaches. Understanding how Mef2 is regulated is of general importance due to its pivotal role in cell differentiation programs. An overview of the transcription factor's interacting proteins contributes a significant part to understanding how Mef2 performs its functions. The proteins identified to interact with Mef2 at its endogenous level and in its normal pattern of expression and activity in the context of development were identified through a complex purification approach that can detect systematically protein complexes and networks. The method of choice was Tandem Affinity Purification (TAP), which offers a good recovery of the Mef2 used as bait as well as a low level of contaminants. The identified candidates were subjected to an analytical pipeline based on *in silico* database mining for expression and protein interactions information, as well as literature querying for existing functional information. Based on the derived information, it was possible to profile the type of proteins Mef2 interacts with during development. New candidates confirmed to be associated with muscle differentiation were shown to be quite distant towards Mef2 when known functionality based on literature was used as a relatedness coefficient. One of the identified candidates in the screen, HDAC4, was used for further biological validation *in vivo*. HDAC4 was extracted during a screen of 11-13 h old embryos, the developmental time point where terminal muscle differentiation occurs. The biological assessment confirmed that the two proteins Mef2 and HDAC4 interact in the context of myogenesis as HDAC4 was found to repress the expression of a Mef2 target protein in somatic muscle. The knowledge derived from this screen suggests that Mef2 does not only interact with muscle specific genes during myogenesis.

Contents

Acknowledgements	iii
Summary	iv
Contents	v
Abbreviations	x
List of Figures	xii
List of Tables	xiii
Chapter 1: Introduction	1
1.1 Overview	2
1.2 <i>Drosophila</i> life cycle	2
1.3 Embryonic development.....	6
1.4 Formation of the mesodermal germlayer and its subdivision	6
1.5 The larval musculature	10
1.5.1 The body wall musculature	10
1.5.2 The visceral musculature.....	11
1.5.3 The heart	13
1.6 Muscle as a model system for cell differentiation programs.....	14
1.7 Mef2 in <i>Drosophila</i>	19
1.8 Mef2 in vertebrates.....	28
1.9 Aims and objectives	32
Chapter 2: Materials and methods	36
2.1 Fly rearing, fly strains and antibodies	37
2.2 TAP purification experiments specific methods	40
2.2.1 Making Mef2 TAP purification compatible.....	40
2.2.2 <i>Drosophila</i> large population maintenance	40
2.2.3 Embryo collection from large collection plates for larval boxes set up....	43
2.2.4 Embryo staging and collection.....	44

2.2.5 Embryos preparation for TAP	45
2.2.6 Lysis of <i>Drosophila</i> embryos for TAP purifications	45
2.2.7 GS-TAP purification	46
2.2.8 Western blots.....	49
2.2.9 Sample preparation and mass spectrometry analysis	50
2.3 Bioinformatic analysis specific methods.....	51
2.3.1 Normalised Spectral Abundance Factor and estimation of contaminants	51
2.3.2 Data collection and identification of protein interaction networks.....	52
2.3.3 Identification of shared and unique seeds in the analysed samples	53
2.3.4 Protein Identifiers.....	53
2.3.5 Shortest distance derivation	53
2.3.6 Expression data	53
2.3.7 Gene expression specificity scale.....	54
2.3.8 An expression filter to find genes active in particular contexts	54
2.3.9 Gene ontology enrichment	55
2.3.10 Modular analysis	55
2.3.11 Functional similarity	56
2.3.12 Pathway enrichment and ontology analysis	56
2.3.13 Data availability statement.....	56
2.4 HDAC4 as a potential Mef2 interactor specific methods.....	57
2.4.1 Embryo collection and fixation.....	57
2.4.2 Immunohistochemistry – single antibody staining	57
2.4.3 Immunohistochemistry - double antibody staining.....	58
2.4.4 Immunohistochemistry using fluorescent antibodies.....	58
2.4.5 Larval muscle scoring and analysis.....	59
2.4.6 Hatching and survival assay.....	59
2.4.7 Preparation of Digoxigenin-labelled RNA probe	59

2.4.8 RNA <i>In Situ</i> Hybridisation.....	61
2.4.9 The GAL4-UAS Expression System	62
Chapter 3: Tandem affinity purification of Mef2 interacting proteins	64
3.1 Introduction	65
3.1.1 Experimental approach	67
3.2 Results	70
3.2.1 Embryo collections for TAP purification.....	70
3.2.2 Detection of the tagged Mef2 bait.....	72
3.2.3 Tracking Mef2 during TAP purification as influenced by the amount of solubilised bait	74
3.2.4 TAP purification of Mef2-GSTAP and associated complexes from 2 g of O/N embryos	80
3.2.5 TAP purification of Mef2 complexes from 2 g of 11-13 h embryos	84
3.2.6 Improving bottleneck steps of TAP purifications	87
3.2.7 TAP purification of Mef2 complexes from 7.7 g of O/N embryos.....	90
3.3 Discussion	94
3.3.1 Input material required.....	94
3.3.2 Tracking Mef2 during TAP purification.....	94
3.3.3 Optimisation of protein extraction from embryos prior to purification	97
3.3.4 TAP purification of Mef2-GSTAP and associated complexes from 2 g of O/N embryos	98
3.3.5 TAP purification of Mef2 complexes from 2 g of 11-13 h embryos	100
3.3.6 Improving bottleneck steps of TAP purifications	101
3.3.7 TAP purification of Mef2 complexes from 7.7 g of O/N embryos.....	102
3.3.8 Purification yield.....	103
Chapter 4: Bioinformatic analysis of Mef2-associated proteins as determined by proteomics studies.....	104
4.1 Introduction	105

4.1.1 Experimental approach	106
4.2 Results	108
4.2.1 Datasets of candidate Mef2-interacting proteins to be investigated	108
4.2.2 Experimental characteristics of candidate datasets	108
4.2.3 Classification of candidate proteins based on spectral counts identified by MS	112
4.2.4 Identification of best candidates for Mef2 regulation during myogenesis	116
4.2.5 Overlaps between TAP and published datasets of Mef2 interactors.....	120
4.2.6 Connectivity between nodes depends on annotation level of the input proteins.....	123
4.2.7 All candidate proteins are closely associated with Mef2	130
4.2.8 Connectivity between candidate proteins.....	133
4.2.9 Topological evaluation of protein-protein interaction networks.....	139
4.2.10 Functional analysis based on gene expression	146
4.2.11 Functional analysis of individual datasets using gene ontology	149
4.2.12 Functional analysis of all candidates using functional similarity	154
4.3 Discussion	158
4.3.1 Datasets analysed	158
4.3.2 Variety of computational techniques applied.....	158
4.3.3 Analysis of protein abundance derived from mass spectrometry	159
4.3.4 Identification of HDAC4 as a promising candidate for further study.....	161
4.3.5 Analysis of candidate proteins' network connectivity.....	163
4.3.6 Analysis of candidate proteins' functional relationships.....	165
4.3.7 Identification of a genetic network working around Mef2 in late muscle development	167
4.3.8 Further validation of the Mef2 interactome	169
Chapter 5: HDAC4 as a potential Mef2 interactor identified by the screen	171

5.1 Introduction	172
5.1.1 The Mef2 and HDAC axis in vertebrates.....	172
5.1.2 Histone deacetylases in <i>Drosophila</i>	175
5.1.3 HDAC4, the only class IIa HDAC in <i>Drosophila</i>	177
5.1.4 Mef2 in <i>Drosophila</i> muscle and its regulators	178
5.1.5 Experimental approach	179
5.2 Results	180
5.2.1 HDAC4 expression in <i>Drosophila</i> embryos	180
5.2.2 Colocalization of HDAC4 with Mef2	186
5.2.3 Functional comparison to human HDAC5.....	190
5.2.4 Generation of targeted deletion HDAC4 null mutants.....	199
5.2.5 Hatching and survival assay of potential HDAC4 null mutants	203
5.2.6 Rescue of lethality of HDAC4 null mutants	203
5.2.7 Muscle phenotype analysis for the potential HDAC4 null mutants.....	204
5.3 Discussion	209
5.3.1 HDAC4 is expressed in an unspecific pattern in <i>Drosophila</i> embryos and does not localise to the nucleus of muscle precursors.....	209
5.3.2 <i>Drosophila</i> HDAC4 is able to repress myogenesis in embryos.....	210
5.3.3 HDAC4 null mutants are lethal but have a normal muscle phenotype...	213
Chapter 6: General discussion.....	214
6.1 The main findings of the work	215
6.2 The role of large scale purification methods in studying protein interactions	218
6.3 Highlighting important biological processes from complex data using bioinformatic analyses.....	220
6.4 Biological validation of candidates identified from purification data.....	223
6.5 Conclusions and Perspective	224
7. References	226

Abbreviations

Act57B	Actin 57B
Actn	α actinin
Brm	Brahma
BSA	Bovine serum albumin
CF	Cytosolic fraction
CTRL	control
DroID	The Drosophila Interactions Database
FRT	Flippase recognition target
FT	flow through
GFP	Green Fluorescent Protein
GO	gene ontology
GS-TAP	a TAP tag with two Protein G modules, a streptavidin binding peptide (SBP), separated by two TEV protease cleavage sites
HDAC	Histone Deacetylase
Jar	jaguar
Mef2	Myocyte enhancer factor 2
modENCODE	Model Organism ENCYclopedia Of DNA Elements
NES	Nuclear export signal
NLS	Nuclear localization signal
NSAF	normalised spectral abundance factor
PBS	Phosphate-buffered saline
PBS-Tx	0.1% Triton X-100 in PBS
PBT	0.1% Tween-20 in PBS
Pmax	the percent maximum or <i>pmax</i> filter for a RNA level expressed as a percent of its maximal level in the tissue or stage where it is expressed in
PPI	protein-protein interaction(s)
PyK	Pyruvate kinase
RNAi	RNA interference
S2R+	Schneider-2 receptor plus

SBP	Streptavidin Binding Petptide
Sls	sallimus
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SWI/SNF complex	SWItch/Sucrose Non-Fermentable complex
TAP	tandem affinity purification
TEV protease	Tobacco Etch Virus protease
TF	transcription factor
Tm 1/2	Tropomyosin 1/2
UAS	upstream activating sequence
Up	upheld
WT	wildtype
wupA	wings up A
YFP	Yellow Fluorescent Protein

List of Figures

Figure 1.1: Drosophila life cycle and muscle development	3
Figure 1.2: Development of the mesoderm and somatic musculature	8
Figure 1.3: Modes of regulation of Mef2 activity	22
Figure 2.1: DNA Sequence of Mef2 exon X with inserted GSTAP-tag	41
Figure 2.2: Mef2GSTAP fusion protein structure	42
Figure 2.3: A Drosophila population cage	42
Figure 2.4: Tandem affinity purification (TAP) procedure	48
Figure 3.1: Collected embryos of Mef2-GSTAP line for TAP purifications that require large amounts of starting biological material	71
Figure 3.2: Identification of tagged Mef2 in the lysate of Mef2-GSTAP embryos	73
Figure 3.3: Options for observing the Mef2-GSTAP bait during TAP purification	76
Figure 3.4: Testing different lysis methods to optimise the recovery of tagged Mef2 protein for use in TAP purifications	78
Figure 3.5: TAP pilot experiment with 2g of O/N embryos expressing Mef2-GSTAP	82
Figure 3.6: TAP pilot experiment with 2g of 11-13 embryos expressing Mef2-GSTAP	85
Figure 3.7: Optimisation of the first steps of TAP purification	88
Figure 3.8: Final TAP experiment with 7.7g of O/N embryos expressing Mef2-GSTAP	92
Figure 4.1: Determining the best candidate for biological testing	118
Figure 4.2: Overlap between different datasets of candidates of Mef2-associated proteins obtained by affinity purification studies	121
Figure 4.3: Analysis of PPI networks of the datasets generated based on DroID derived data	124
Figure 4.4: Numbers of detected interaction partners in DroID for each candidate protein in the starting datasets	128
Figure 4.5: Number of intermediate proteins between Mef2 and associated proteins present in each PPI network	131
Figure 4.6: Classification of Mef2 associated proteins on levels depending on their distance to Mef2 in seed-only subnetworks	135
Figure 4.7: Topological characterisation of protein network generated from a unified seed list and interactions between seeds stored in DROID	144
Figure 4.8: Expression distribution of the protein seeds based on the expression filter pmax	147
Figure 4.9: GO terms associated with Mef2	150
Figure 4.10: Mef2 GO terms enriched in one or more dataset	152
Figure 4.11: 11-13MEF specific GO term clusters and their constituent proteins	153
Figure 4.12: Functional analysis of seeds based on the network generated from the most functionally related seeds	156
Figure 5.1: mRNA and protein expression of HDAC4 in Drosophila embryos	184

Figure 5.2: Relative mRNA expression of Mef2 and two co-expressed candidates across embryogenesis	187
Figure 5.3: Colocalisation of YFP-HDAC4 with Mef2 during Drosophila embryogenesis.	189
Figure 5.4: Structure of HDACs and transgenically expressed mutants	191
Figure 5.5: HDAC5 Δ C overexpression downregulates Mef2 activity	196
Figure 5.6: HDAC4 Δ C overexpression downregulates Mef2 activity	197
Figure 5.7: Overexpression of Drosophila HDAC4 and human HDAC5 affects muscle differentiation in late stage embryos	198
Figure 5.8: HDAC4 deletion screening	200
Figure 5.9: HDAC4 successful deleted region screening	202
Figure 5.10: Hatching and survival assay of potential HDAC4 null mutants	205
Figure 5.11: Genetic scheme of attempted rescue experiment of HDAC4 del6 using a UAS-HDAC4 construct and a DaGal4 driver	206
Figure 5.12: Rescue of HDAC null mutants by genomic duplications of the X chromosome	207
Figure 5.13: Muscle phenotype in HDAC deletion mutants	208
Figure 6.1: Graphical abstract	217

List of Tables

Table 2.1: Drosophila lines	38
Table 2.2: Antibodies	39
Table 2.3: Stages of embryos collected	44
Table 2.4: UAS-Constructs used to overexpress HDAC	62
Table 4.1: Protein datasets	109
Table 4.2: Short list of candidates for Mef2 regulation during myogenesis	119
Table 4.3: Characteristics of networks based on seeds from all datasets, filtered or not by CTRL edges	142
Table S1: Raw spectral counts and hit ranking of all proteins identified by TAP/MS	Figshare
Table S2: Mef2 GO biological process terms	Figshare
Table S3: High-interest candidates identified by each analysis	Figshare
Table S4: Overlap of TAP datasets with muscleRNAi screen	Figshare
Tables S5ff: Data corresponding to individual figures	Figshare

Chapter 1: Introduction

1.1 Overview

The main question this project aims to answer is how the transcription factor Mef2 is regulated during muscle development. Mef2 is the central transcription factor that governs the development of muscle. Pinpointing how Mef2 is regulated by other proteins is crucial to our understanding of muscle differentiation. The project therefore focussed primarily on identifying the Mef2 interactome in its physiological context during development. The model organism selected for this study was *Drosophila melanogaster* because this organism expresses one Mef2 gene and can be cultured to obtain large sample quantities for protein purification. Additionally, any candidate proteins identified can be tested *in vivo* with relative ease due to the range of existing genetic tools. The main approach taken to identify proteins that interact with Mef2 during myogenesis has been to perform a large scale study of protein purification that maintains physical protein-protein interactions under physiological conditions, thereby isolating not only proteins directly interacting with Mef2 but also further indirectly bound components of Mef2 protein complexes.

1.2 *Drosophila* life cycle

The *Drosophila* life cycle consists of four distinct stages (Figure 1.1A): embryo, larva, pupa and adult (Bate and Martinez-Arias, 1993; Hartenstein, 1993). The *Drosophila* females deposit the fertilised egg on the surface of the food and the egg develops externally. This aspect makes *Drosophila* embryos easily accessible for collection in relatively high numbers. The embryonic development takes 24 hours and comprises 17 stages of development (Campos-Ortega and Hartenstein, 1997). The first and second instar stage each last for 24 hours and it takes another 48 hours for the third instar larvae to grow enough to pupate. Most of the larval tissues are histolysed during pupation and the adult fly is formed during the metamorphosis process. The pupal stages are distinguishable by changes in colour, which range from the white colour of the larva that becomes stationary for the prepupal stage to the brownish shade of the pupal case. At the end of metamorphosis the eyes develop pigmentation and become visible and the wings turn black. The pupal stage lasts for approximately 4 days and it ends with the adult emerging from the pupal case. It takes approximately 11 days from egg to adult fly at 25°C.

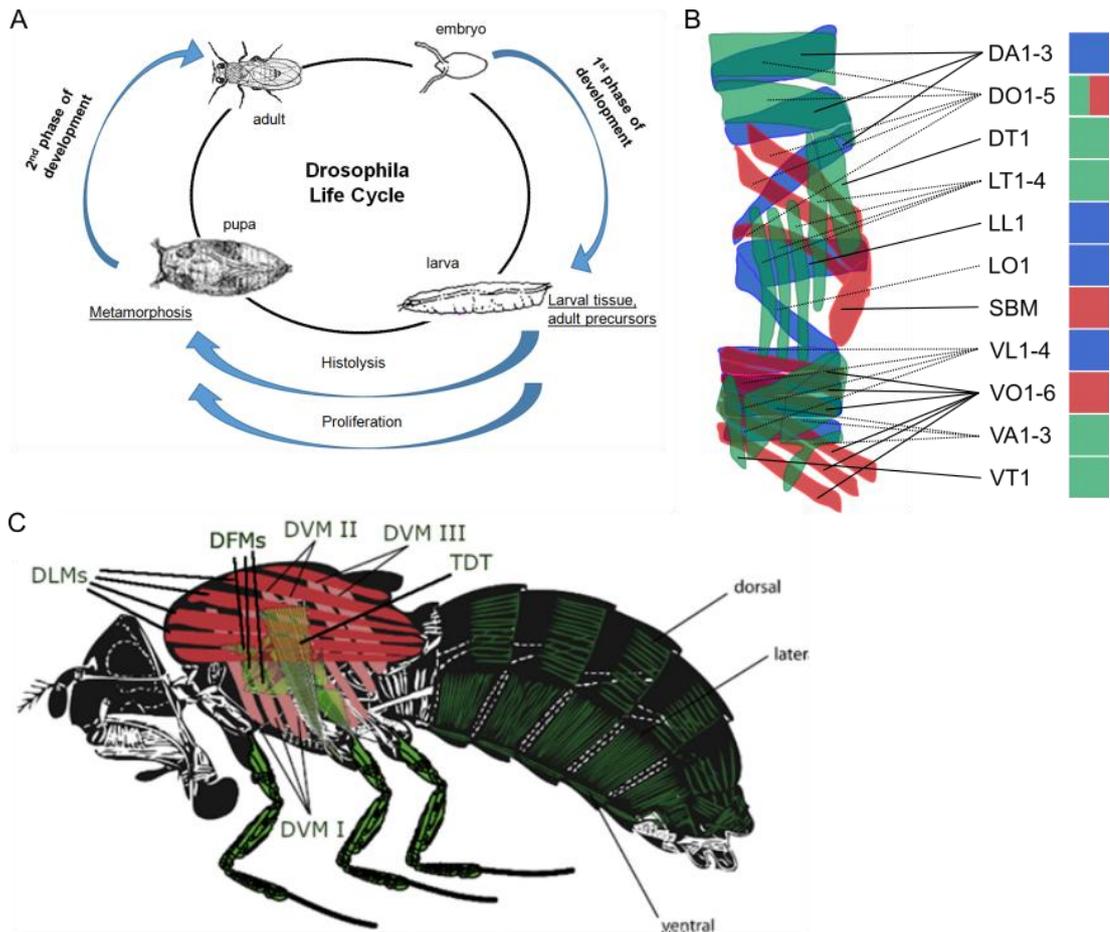


Figure 1.1: *Drosophila* life cycle and muscle development

(A) *Drosophila* has four distinct stages: embryo, larva, pupa and adult. Larva and adult stages are motile and have distinct types of muscles. The larval muscles are formed in the first phase of myogenesis occurring during embryogenesis. The adult has a variety of muscles that are formed de novo from adult muscle precursors (AMPs) or through differentiation of larval muscles that do not get histolysed during metamorphosis. The adult muscles are formed in the second phase of development.

(B) Anatomical muscle pattern in a larval body hemisegment. Muscles are named according to their location along the dorso-ventral axis (D = dorsal, L = lateral, V = ventral) and its orientation (A = acute, O = oblique, T = transverse, L = longitudinal). SBM = segment border muscle. Colour indicates muscle depth: green = external, red = intermediate, blue = internal. Adapted from (Weitkunat and Schnorrer, 2014).

(C) Scheme of muscles in adult *Drosophila*. Thorax muscles: Indirect flight muscles, fibrillar type: Dorso-ventral muscles (DVMs, pink) located along the dorso-ventral axis, Dorso-longitudinal muscles (DLMs, red) located along the antero-posterior

axis. The direct flight muscles (DFMs, green) are located at the base of the wing insertion and are tubular. The Tergal depressor of the trochanter (TDT) or jump muscles are located from the dorsal thorax to the second leg of the thorax. Abdominal muscles located in the posterior part of the fly have a more uniform fate compared to the thoracic muscles, they present a stereotyped repetitive pattern in each segment. Abdominal muscles are smaller and more numerous than thoracic muscles. Representation taken from (Laurichesse and Soler, 2020).

Two of the *Drosophila* developmental stages, larva and adult, are mobile and present different types of muscles that are adapted to the respective lifestyle (Bothe and Baylies, 2016). The larval musculature has a repeated muscle pattern that consists of 30 distinct muscles that are formed from individual muscle cells developed during embryonic myogenesis. Each muscle name is indicative of the position and the orientation of each muscle fibre which presents a distinctive size, shape, orientation, number of nuclei, tendon attachment and innervation (Figure 1.1B). Adult musculature is formed either *de novo* from precursor muscle cells or through transdifferentiation of larval muscles. The adult muscles are multifibrillar and present distinct muscle structures (Figure 1.1C): fibrillar muscles (structure specific for the Indirect Flight Muscles (IFM)) and tubular muscles (structure specific for abdominal, leg, jump, and direct flight muscles (DFMs)) (Dobi et al., 2015). The present study only focused on muscles formed during embryogenesis.

There are three main types of muscles in the embryo: somatic, cardiac and visceral. All these muscles are derived from the mesoderm that is formed in the ventral part of the embryo. The somatic musculature presents in each abdominal hemisegment of the embryo a pattern of thirty distinct muscle fibres with specific characteristics. Each muscle fibre is derived from a single founder cell (FC) that will fuse to fusion competent myoblasts (FCMs) to differentiate into a syncytial myotube. During the pupal stage, most of the larval muscles are destroyed and adult muscles are formed *de novo* from adult precursor muscle cells that are related to the embryonic precursor cells. The adult muscle progenitors (AMP) are formed in the embryo and are kept in an undifferentiated state until metamorphosis when they will give rise to abdominal and thoracic myofibers *de novo* (Dobi et al., 2015). There are two types of larval muscles that escape histolysis and together to AMP will be remodelled into adult fibres: abdominal intersegmental muscles are transformed into temporary dorsal oblique muscles and the three larval oblique muscles (LOMs) will form the dorso-longitudinal muscles (DLMs) (Soler and Taylor, 2009). Both *de novo* and remodelled muscles present a common myoblast proliferation program to form segment-specific muscle fibres.

1.3 Embryonic development

There are 17 stages of development during *Drosophila* embryogenesis (Campos-Ortega and Hartenstein, 1997). Once fertilised, the egg's nucleus undergoes several rounds of divisions before the nuclei migrate to the periphery of the egg. This is followed by cellularisation, which consists of approximately 6000 cells and completes two to three hours after egg laying (AEL). After this, the embryo undergoes gastrulation, whereby the three germ layers are formed (Figure 1.2A): endoderm, ectoderm and mesoderm. The endoderm gives rise to the midgut derived from two independent regions that develop separately in the anterior and posterior part of the embryo and use cell migration to connect in late embryogenesis. The ectoderm is split into regions with different fates: the foregut and hindgut are formed from cells located anteriorly and posteriorly, adjacent to the respective endodermal primordia. The nervous system is derived from neuroblasts originating from the neuroectoderm located ventrally, while the lateral and dorsal ectoderm will form the epidermis and the tracheal system. The dorsal side of the embryo will form the amnioserosa, while the ventral side of the embryo is occupied by the mesoderm. This germ layer differentiates into fat body and all muscles, including heart, pharyngeal muscles, visceral muscles, and somatic muscles (Leptin, 2004). Since this work focuses on muscle development in the *Drosophila* embryo, the subdivision of the mesoderm is explored further in more detail.

1.4 Formation of the mesodermal germlayer and its subdivision

The development of muscles from the mesodermal germ layer can be subdivided into three steps. First, the fate specification of mesodermal cells (see Figure 1.2A), second, the specification of a subset of mesodermal cells as muscle progenitors that express muscle-specific genes (see Figure 1.2B), and third, the muscle differentiation program by which muscle progenitors develop into individual muscles (see Figure 1.2C) (reviewed in Dobi et al., 2015). The mesodermal cell fate is specified by the maternally deposited transcription factor Dorsal, which is located on the ventral side of the embryo as a result of embryonic dorso-ventral axis formation (Roth et al., 1989; Rushlow et al., 1989). Two target genes of Dorsal, *snail* and *twist*, are then both required for the ventral cells to adopt the mesodermal identity (Boulay et al.,

1987; Ip et al., 1992; Jiang et al., 1991; Kosman et al., 1991; Pan et al., 1991; Thisse et al., 1991). To achieve this, Snail represses genes expressed by the other germ layers, and Twist activates the late mesodermal genes. Near embryonic stage 8, all cells that express both of these invaginate as part of gastrulation and undergo two cell divisions while spreading along the ectodermal cells towards the dorsal side (Borkowski et al., 1995; Dobi et al., 2015; Leptin, 2004; Leptin and Grunewald, 1990; Riechmann et al., 1997). Twist is a helix-loop-helix protein and activates transcription of both the homeobox gene *tinman* (*tin*) and the MADS domain transcription factor *mef2* (Taylor, 1995; Yin et al., 1997). All three proteins, Twist, Tinman and Mef2, are uniformly expressed in the early mesoderm and together they regulate its developmental transcriptional network (Ip et al., 1992; Lilly et al., 1995; Nguyen et al., 1994; Sandmann et al., 2006; Taylor et al., 1995; Yin et al., 1997).

As the mesodermal cells spread across the ectoderm, they enter signalling domains established by the ectodermal cells (Figure 1.2B). Along the dorso-ventral axis, Decapentaplegic (Dpp) influences mesoderm subdivision, while along the anterior-posterior axis it is influenced by Wingless (Wg) and Hedgehog (Hh) (Baylies et al., 1998; Borkowski et al., 1995; Frasch, 1995; Halfon et al., 2000; Staehling-Hampton et al., 1994; Tixier et al., 2010). Dpp is located dorsally in the ectoderm, and promotes the dorsal mesodermal fate in those cells that migrate the furthest (Azpiazu and Frasch, 1993; Dobi et al., 2015; Frasch, 1995). By helping to maintain *tinman* expression, Dpp not only promotes the formation of the dorsal mesoderm but also inhibits ventrally expressed genes (Dobi et al., 2015; Staehling-Hampton et al., 1994).

Along the anterior-posterior axis, pair-rule genes act to subdivide the mesoderm once the germ band is extended. This includes *even-skipped* (*eve*) and *sloppy-paired* (*slp*). Areas subject to *eve* develop into the visceral musculature, fat body, gonadal mesoderm, mesodermal glial cells and some dorsal somatic muscles due to the influence of Dpp and Hh (Baylies et al., 1998; Tixier et al., 2010), while mesoderm in the *slp* domains receive only Wg and form the somatic musculature and the heart (Borkowski et al., 1995; Lee and Frasch, 2000; Riechmann et al., 1997; Tixier et al., 2010). These parts of the mesoderm form the outer layer that maintains contact with the ectoderm, while the *eve* layer detaches from the ectoderm (Ruiz-Gómez, 1998).

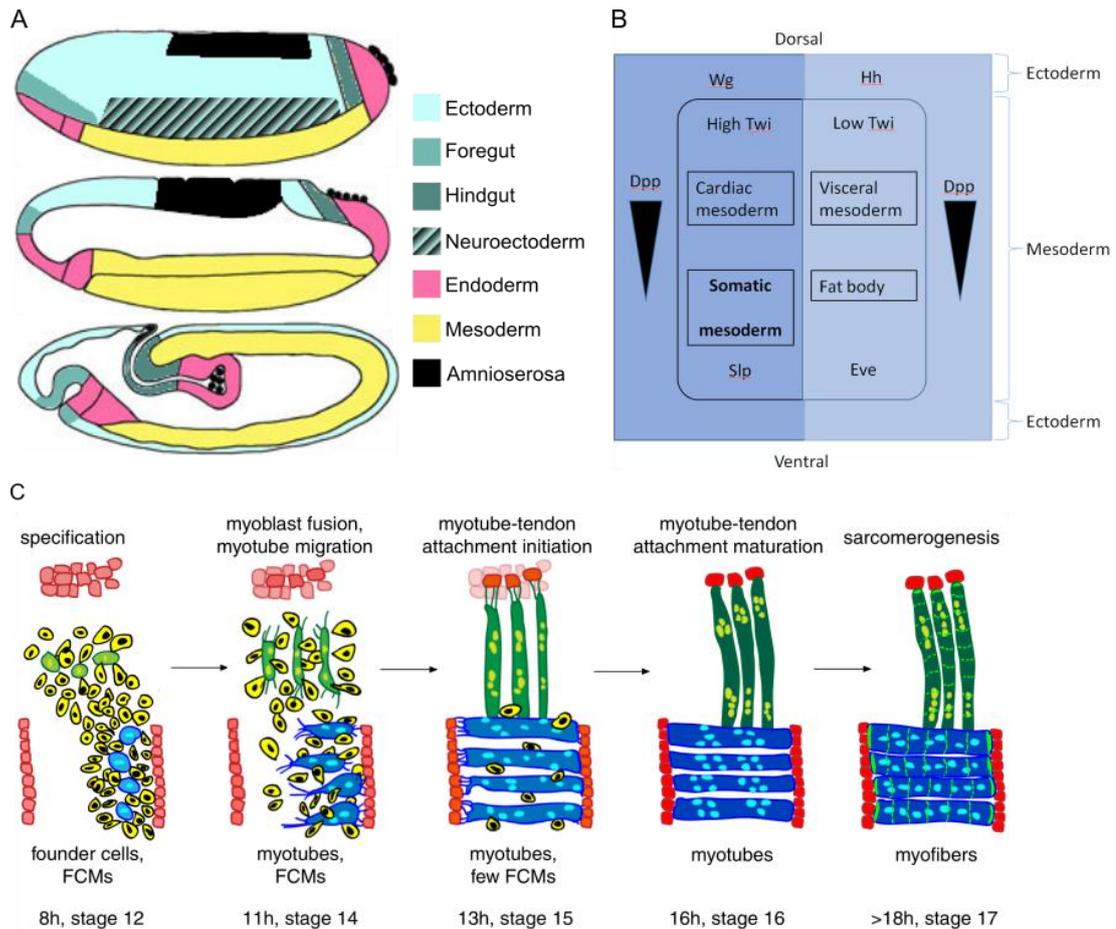


Figure 1.2: Development of the mesoderm and somatic musculature

(A) *Drosophila* gastrulation, the process by which the undifferentiated blastoderm separates into the three germlayers, ectoderm, mesoderm and endoderm. The amnioserosa, also called extraembryonic tissue, undergoes apoptosis and is fully reabsorbed by other cells during the late stages of embryogenesis. Adapted from (Leptin, 1999).

(B) Genetic patterning of the mesoderm shortly after gastrulation. Ectodermal signals pattern the mesoderm. Dpp patterns the dorso-ventral axis, while Wg and Slp pattern the antero-posterior axis. Myogenic competence domains are determined by Twi, which has a high expression in Slp sections and low expression in Wg sections. Somatic mesoderm forms in the antero-ventral part of each segment. Cardiac precursors form from the most dorsal portions of high-Twi domains of each segment. In the posterior part of each segment, in low-Twi domains, visceral muscles and fat body are specified. See main text for detail. Adapted from (Tixier et al., 2010).

(C) Embryonic development of somatic musculature by the example of four ventral-longitudinal muscles (VL1-4, blue) and three lateral transverse muscles (LT1-3, green). Approximate timing and embryonic developmental stage is given below each diagram. The process can be roughly subdivided into five stages, which are explained in section 1.5.1. Adopted from (Weitkunat and Schnorrer, 2014).

By this point, the subdivision of the mesoderm also becomes apparent in the expression of Twist. The external mesodermal layer in contact with the ectoderm expresses high amounts of Twist while those cells that detached express low levels (Bate et al., 1993; Borkowski et al., 1995; Tixier et al., 2010). This subdivision of Twist expression differs between the segments depending on the specific mesodermal tissues that develop in each of them (Baylies and Bate, 1996; Castanon and Baylies, 2002; Tixier et al., 2010).

1.5 The larval musculature

1.5.1 The body wall musculature

At the end of embryogenesis, the somatic muscles have developed a complex pattern of syncytial striated muscles located under the surface of the ectoderm (Bate, 1990; Bate et al., 1993). There are in total 30 multinucleated fibres per abdominal segment (Bate, 1990). Each fibre is formed from a single founder cell (FC), to which several fusion-competent myoblasts (FCMs) fuse during myogenesis (Abmayr and Pavlath, 2012).

Muscle development begins in parallel to the onset of germ band retraction (Schnorrer and Dickson, 2004). The first process is the fusion of doublets or triplets of cells in the ventral mesoderm. These small group are the future syncytia of the ventral muscles, but such syncytia also form dorsally and laterally by the end of germ band retraction. The exact positions are defined by the location of the precursors, which act as a sort of blueprint for the muscle pattern (Bate and Rushton, 1993; Campos-Ortega and Hartenstein, 1997). Each precursor founder cell fuses with fusion competent myoblasts while growing in size starting from stage 12. Simultaneously, the syncytium expresses structural proteins and begins forming attachment sites with the ectoderm. By the end of embryonic stage 15, the characteristic points where muscles insert on the ectoderm have formed, marking the finalisation of the muscle pattern (Figure 1.2C). Each fibre is formed by a founder cell (FC) and fusion-competent myoblasts (FCM), whereby each FC seeds a specific type of muscle (Baylies et al., 1998; Chen and Olson, 2004; Weitkunat and Schnorrer, 2014).

FCs are selected out of clusters of cells within domains of high Twist expression that express lethal of scute (*l'sc*). *l'sc* expression is gradually restricted by Notch-mediated lateral inhibition, specifying a single remaining cell expressing *l'sc* as an FC (Carmena et al., 1995; Dobi et al., 2015), the remaining cells becoming FCMs (Baylies et al., 1998; Carmena et al., 1998; Chen and Olson, 2004). FCs then undergo asymmetric cell division, giving rise to either two daughter FCs, or one FC and one adult muscle progenitor (AMP). Before developing into FCs, muscle progenitor cells express different combinations of identity genes, which are not necessarily maintained during division into two daughter cells. As a result, when FCs are specified at the onset of germ band shortening, each of them expresses a specific combination of identity genes (Dobi et al., 2015; Tixier et al., 2010). These genes are characteristic of individual muscles and determine muscle differentiation of specific myofibers with different sizes, shapes, attachment sites and number of nuclei (Abmayr and Pavlath, 2012; Baylies et al., 1998; Chen and Olson, 2004). During the fusion process, FCMs are recruited to the characteristic pattern of identity genes of the founder cells they fuse with. Therefore, the syncytial muscle precursors and, later, the muscles themselves can be identified by their pattern of identity gene expression. Although each founder has a characteristic identity gene profile, the process of fusion with FCMs is similar (Figure 1.2C). The process can be described as 1) cell attraction of the FCM by the FC or in later fusion events by a syncytial myotube, 2) adhesion of FC or syncytial myotube with the FCM mediated by cell-surface molecules, 3) alignment and fusion, which involves formation of a plaque through actin polymerisation and vesicle recruitment from which FCM invasive structures will protrude into the FC/myotube, 4) absorption of the FCM into the myotube (Abmayr and Pavlath, 2012; Doberstein et al., 1997; Dobi et al., 2015).

1.5.2 The visceral musculature

Visceral muscles surround the larval midgut and are involved in the peristaltic movement of digestion (Campos-Ortega and Hartenstein, 1997). Visceral muscles comprise two types: binucleated circular muscle fibres that derive from the trunk visceral mesoderm and multinucleated longitudinal muscles coming from the caudal visceral mesoderm. The two types form a web around the midgut and their correct

development is required for proper separation of the gut into four chambers (Klapper et al., 2001; Rudolf et al., 2014; San Martin and Bate, 2001; Schröter et al., 2006).

The visceral mesoderm is derived from cells expressing low levels of *twist* that also express the transcription factor Bagpipe (Bap) that is activated by Tin. Mutations in the two genes result in disruption of the visceral mesoderm. Visceral mesoderm starts by segregation of clusters of cells expressing *bap* and *binou* (bin) in the dorsal mesoderm (Azpiazu and Frasch, 1993; Zaffran et al., 2001). These two factors cooperate to form rows of visceral cells on each side of the egg yolk. This process is followed by visceral myoblasts fusion, a similar process of fusion of FCs with FCMs as for somatic musculature. The circular muscles form binucleated syncytia interconnected with multiple cytoplasmic bridges and with the ability to stretch in order to enclose the gut along the dorso-ventral axis. The longitudinal FC myoblasts originate from the caudal visceral mesoderm and migrate anteriorly along the trunk visceral mesoderm and most probably fuse with FCMs after the circular muscles fusion occurs (Aghajanian et al., 2016; Klapper et al., 2001; Martin et al., 2001; Rudolf et al., 2014; Schröter et al., 2006). This myoblast fusion process is independent from contact with adjacent tissue, but elongation processes from late visceral myogenesis are dependent on contact with the endoderm (Aghajanian et al., 2016).

During its development, the midgut forms three constrictions which require the correct differentiation of the visceral mesoderm (Azpiazu and Frasch, 1993). Bap is required for visceral mesoderm development but is not maintained in visceral muscle differentiation. Tin activates Mef2 and β 3-tubulin (Cripps et al., 1998) in the developing visceral mesoderm where Mef2 is the central component of the differentiation process (Lilly et al., 1995). In mutants of Mef2 the midgut forms and *tin* and *bap* are expressed at normal levels (Lilly et al., 1995), but the gut fails to contract and the non-contracting midgut musculature expresses very little myosin (Bour et al., 1995). However, β 3-tubulin is not a target of Mef2 in the visceral musculature (Damm et al., 1998), its expression being activated by the cooperation and partial redundancy of Bin and Bap (Zaffran and Frasch, 2002).

1.5.3 The heart

The *Drosophila* heart, also called the dorsal vessel, is a dorsomedial muscular tube that pumps haemolymph throughout the body in a back to front direction. It consists of two types of cells: cardioblasts and pericardial cells (Cripps and Olson, 2002; Zaffran and Frasch, 2002). The cardioblasts express muscle proteins and form the contractile heart tube. They are arranged in two rows at the dorsal midline and are surrounded by four rows of pericardial cells. No muscle proteins are expressed in the pericardial cells (Ward and Skeath, 2000). The different heart cell types derive from the dorsal mesodermal cells expressing high amounts of Twist. As embryogenesis progresses, heart precursors migrate from each side of the embryo to the dorsal midline.

Around stage 10, Tin is activated early by Dpp and restricted to the dorsal mesoderm. During stage 11, the mesoderm displays alternating domains of high and low Twist expression. The low Twist cells will form the precursors of the circular visceral muscles, while the high Twist domains will generate the precursors for the heart and some dorsal muscles (Figure 1.2B). The heart progenitors further divide into two rows of bilaterally symmetrical cells. One row will form the cardioblasts and one row will form the pericardial cells (Bodmer and Frasch, 2010). Tin is involved in specification of both types of heart cells (Bodmer and Frasch, 2010; Reim and Frasch, 2010). Tin activates Mef2 which is involved in the differentiation of cardioblasts. Mef2 activates expression of pan-muscular structural genes in cardioblasts such as *Act57B*, *Mhc*, *TnI*, *Prm*. β 3-tubulin, which seems to be a cardiac specific gene (Bryantsev and Cripps, 2009), is expressed in differentiated cardioblasts, however it is not a direct Mef2 target in these cells, but rather of Tin (Kremser et al. 1999; Damm et al, 1998). Complex networks of genes are involved in the specification of the various heart cell populations. At the dorsal midline differential gene expressions allow to distinguish the cardioblasts from pericardial cells. The genes characteristically expressed in cardioblasts during development are *mef2* (Lilly et al. 1995; Taylor et al. 1995), *myosin*, *β 3-tubulin* (Damm et al., 1998) and *tin* (Azpiazu and Frasch, 1993). Genes involved in the functioning of the mature cardiac tube have distinct expression patterns: some are expressed throughout the tube and most often their expression is Mef2-dependent, others are restricted to Tin-

expressing cells and they are direct targets of Tin. The ostium cells, which form the inflow valves at the end of the heart, present a distinct characteristic in that Tin initially induces expression of both Svp and Mef2 that will activate specific ostial genes, but the expression of the two TFs becomes Tin-independent (Bryantsev and Cripps, 2009).

1.6 Muscle as a model system for cell differentiation programs

Cell development involves patterns of gene expression which are regulated spatio-temporally via the activity of transcription factors (TFs) and cell signalling cascades. Cis-regulatory modules (CRMs) or enhancer elements are transcriptional regulatory regions where multiple transcription factors bind DNA in order to orchestrate activation of various batteries of target genes (Bonn and Furlong, 2008). CRMs can be simultaneously bound by multiple TFs in a dynamic manner, and different concentrations and combinatorial interactions between co-occupying TFs can radically change the impact of an enhancer on the transcription of its downstream gene(s). Furthermore, even for two TFs that elicit similar phenotypes in loss-of-function mutant embryos and thus appear to perform similar functions, the molecular relationship between these two proteins can result in a diversity of transcriptional responses within the same cell depending on the cell's external environment. Many developmental processes in *Drosophila* embryos are governed by hierarchical networks of TFs that integrate at transcriptional level signals from environmental stimuli or intercellular interactions (Cunha et al., 2010). The embryonic developmental processes can be classified as either differentiation processes that form dozens of tissue types or dramatic temporal transitions as found in gastrulation or segmentation (Zhou et al., 2019).

The segmentation network of the *Drosophila* embryo that defines its antero-posterior axis is characterised by the interaction of transcriptional activators with broad expression profile with the transcriptional repressor that have restricted activity domains. For example, the second stripe CRM of *eve* is activated under the influence of Bicoid and Hunchback, two maternal factors expressed in the anterior part of the embryos and the expression of the same enhancer elements is repressed by the TFs

Giant and Krüppel that are expressed only in specific regions (Bonn and Furlong, 2008; Small et al., 1991; Stanojevic et al., 1991).

The dorso-ventral axis is defined by a concentration gradient of the TF Dorsal which is highly expressed in the ventral region and at lower levels in the lateral and dorsal domains of the embryo (Bonn and Furlong, 2008). The dorso-ventral gradient of Dorsal expression occurs under the transmembrane receptor Toll signalling which is active ventrally and is able to transport the Dorsal TF from the cytoplasm in the nucleus. In the dorsal region Dorsal remains excluded from the nucleus since Toll is inactive in this region (Roth et al., 1989; Rushlow et al., 1989). Dorsal target genes have different affinity in their binding site for their transcriptional activation. Snail and twist, two of its ventral targets, require high levels of Dorsal since they have low affinity binding sites and they specify the cells of the future mesoderm. Genes that specify the ectoderm and extend dorsally have higher affinity binding sites for Dorsal. While a gene like *rho* expressed in the neuroectoderm still requires low levels of Dorsal to activate its expression (Ip et al., 1992), a completely dorsally expressed gene like *zen* is only expressed in areas of the embryo where Dorsal is not able to repress its transcriptional activation (Cai et al., 1996). Dorsal can repress transcription by interacting with other TFs (Dubnicoff et al., 1997).

The mesoderm network which regulates the subdivision of the mesoderm into primordia of different muscle types is a multi-layered process, compared to the D-V process which occurs in a single epidermal layer. This network could be described as a TF cascade which is started by Twist which regulates the expression of many TFs such as Tinman and Mef2. Tinman further regulates a subnetwork that will define the dorsal mesoderm, while Mef2 activates the differentiation network for somatic muscle. The CRMs regulated by the TFs in this network integrate inputs from TFs and signalling cascades (Bonn and Furlong, 2008). Lameduck is a transcription factor found downstream of Twist, Tin and Mef2 and is expressed in precursors of the somatic muscle and the visceral muscle. Enhancers bound by Lmd are cobound by the three others in a combinatorial manner to activate muscle precursor genes. Genes expressed both in FCMs and FCs, are co-bound by Lmd, Tin and Twi, while target genes exclusive to FCMs are co-bound by Lmd, Tin and Mef2. All four TFs' target genes are expressed predominantly in both myoblast types. These genes have a

diversity of expression pattern resulting from the diversity of transcriptional activators and their binding behaviour and no one interaction is sufficient to account for all features of FCM gene expression. Enhancers of FCM-exclusive genes and genes expressed in both FCMs and FCs that are bound by Lmd and the other TFs show chromatin modification associated with active enhancers, specifically H3K4me1, H3K27ac, and H3K79me3 (Busser et al., 2012). Most of Lmd-bound CRMs are co-bound by Mef2 and the responses generated by these two transcription factors can be varied within the same cell type (Cunha et al., 2010).

Myogenesis is a process of cell differentiation that consists of a stage of progenitor cells proliferation, followed by activation of muscle-specific genes that lead to differentiation of the myoblasts. These will later fuse into myotubes. In embryonic myogenesis the muscle fibres are derived from mesodermal structures that form the template for the subsequent muscle tissue. Myogenesis has represented a paradigm for cellular differentiation processes as it has offered insights into basic cellular mechanisms such as signalling, transcriptional and post-transcriptional control of cell fate, cell fusion and cell differentiation (Gunage et al., 2017).

Many properties of the muscle differentiation process are shared between *Drosophila* and higher organisms. *Drosophila* is the model organism that has offered important advances in understanding the way myoblasts fuse, identification of muscle-specific promoters, the role of myonuclear positioning in myofibrillar function and mechanisms that lead to muscle formation. *Drosophila* larval muscles are formed by a single syncytial fiber, while adult muscles have multiple fibers. Vertebrates only contain multiple syncytial muscles (Bentzinger et al., 2012; Chaturvedi et al., 2017).

Muscle tissue is derived from the mesoderm, a germ layer that together with the endoderm and the ectoderm give rise to all the tissues that form an organism. Myogenesis can be divided into several distinct phases: cells that will generate the muscle precursors start from mesodermal structures that under positive and negative external signalling activate a battery of transcription factors and chromatin remodelling factors. These factors translate this signalling into gene expression and microRNA programs that give myogenic identity to the muscle progenitor cells (Bentzinger et al., 2012). The mesoderm starts to form segments along the anterior-posterior (A/P) and dorsal-ventral (D/V) axes and populations of muscle progenitor

cells are formed at specific locations. Within the muscle progenitor cells population individual cells become singled out compared to their neighbours and will form founder cells. The neighbours will become fusion-competent myoblasts that will fuse with a single founder cell to form a muscle fibre. Founder cells have a very important role in establishing the particular characteristics of different muscles fibres by activating specific muscle identity gene products. By fusing with competent myoblasts, mature syncytial myotubes are formed that will eventually attach to the epidermis. In more complex myotubes, primary fibres are formed that will fuse to secondary fibres while the muscle grows (Lemke and Schnorrer, 2017).

Muscle formation is influenced by several signalling cascades, including Wnt, TGF β and Hedgehog family members (Baylies et al, 1998). Transcription factors like Twist (Baylies and Bate, 1996; Castanon and Baylies, 2002) and Mef2 (Bour et al., 1995; Lilly et al., 1995; Ranganayakulu et al., 1995) are the key activators of muscle differentiation and of gene-specific programs. Their roles have been conserved from *Drosophila* to vertebrates. Mef2 promotes muscle differentiation and is present throughout the muscle differentiation process from early mesoderm to later stages where it remains detectable in all muscle types (somatic, visceral, pharyngeal and the heart musculature). Mef2 is activated early in development by Twist (Taylor et al. 1995), which binds directly to its regulatory sequence (Cripps et al., 1998).

The *mef2* gene is first detected in the *Drosophila* embryo in the ventral furrow cells during gastrulation. Its expression is restricted to the mesoderm during germ-band extension and its expression matches the pattern of the *twi* gene. Around stage 10, the mesoderm separates into two different cell layers in order to form the primordia of the visceral muscle and heart precursor cells in the dorsal mesoderm and the somatic musculature in the ventral mesoderm. The *mef2* expression is present in all these precursors and its expression is maintained in later stages of embryonic development (Lilly et al., 1994; Nguyen et al., 1994).

The Mef2 protein is first detected at stage 11 during germband expansion throughout the mesoderm. During the germband expansion process and in later stages of embryonic development, the Mef2 protein is observable in the visceral muscles, cardiac cells and in segmentally repeating clusters of the forming somatic

musculature. The localisation of the Mef2 protein is restricted to the nucleus of muscle progenitor cells and their differentiated derivatives (Bour et al., 1995).

Mef2 protein in turn binds to the consensus regulatory sequence YTAWWWWTAR (Andres et al., 1995) and activates a large series of target genes with broad expression profiles (Elgar et al., 2008; Junion et al., 2005; Sandmann et al., 2006). Some of the first genes that were identified as direct target genes were Tropomyosin (Lin et al, 1996), β 3-tubulin (Damm et al, 1998) and Paramyosin (Arredondo et al, 2001), which are proteins involved in muscle structure, while Actin57B (Kelly et al, 2001) is the major myofibrillar actin expressed in muscle during embryogenesis. Large-scale studies pinpointing Mef2 binding sites across the genome have shown that Mef2 can activate genes important in muscle differentiation both directly and indirectly (Elgar et al., 2008; Junion et al., 2005; Sandmann et al., 2006; Taylor and Hughes, 2017).

The precursors of the adult muscles (AMPs) begin to form during the early stages of *Drosophila* embryonic development and commit to the muscle fate. Unlike other myoblasts around them which differentiate into larval muscles, they retain an undifferentiated state (Bate et al, 1991). It has been suggested that these cells activate a specific lineage program to maintain their identity, and simultaneously prevent their peer myoblasts from acquiring the same fate (Ciglar et al., 2014).

Progenitor muscle cells begin expressing Mef2 long before they begin expressing early target genes and thereby enter the process of differentiating into muscles. One proposed explanation for this is that regulatory factors repress Mef2 and inhibit its activity. In vertebrates, two repressors of Mef2 are known among other candidates: Twist and class IIa Histone Deacetylases (HDACs). However, Twist in *Drosophila* is required for Mef2 expression in the mesoderm, which would contradict a simultaneous role as a repressor (Cripps et al., 1998; Taylor et al., 1995).

Mef2 is expressed throughout the myogenesis process that generates muscles from the respective mesodermal cells both in muscle precursor cells and adult myocytes. Muscle development depends on Mef2 expression and a lack of this transcription factor results in failure of myogenesis (Bour et al., 1995). Both early and late muscle-specific target genes are activated by Mef2 (Elgar et al., 2008; Junion et al., 2005;

Sandmann et al., 2006; Taylor and Hughes, 2017). In *mef2* mutants, genes characteristic for founder cell are expressed, therefore FCs are formed and specified (Bour et al., 1995; Lilly et al., 1995; Ranganayakulu et al., 1995), but later stages of muscle development are impaired. Mef2 is required both for fusion and for activation of the genes involved in the differentiation process.

1.7 Mef2 in *Drosophila*

Early in Mef2 research, the role of the gene in muscle development at *Drosophila* embryonic level has been used to place this gene at the centre of muscle differentiation. Later work in *Drosophila* has established that Mef2 is involved in activation of transcription in other tissues (Blanchard et al., 2010; Clark et al., 2013). Mef2 is an important hub of a regulatory network that consists of intricate feed forward and feedback loops that integrate all types of regulation of a gene and its products. Regulation of Mef2 in *Drosophila* involves mechanisms that regulate its activity at transcriptional, translational, and post-transcriptional levels.

To understand how Mef2 performs its function most of the studies looked at the activation of its target genes and how their transcription is activated. A large scale genomic study correlated the enhancer regions bound by Mef2 during embryonic muscle development with the expression profiles of the genes activated by these enhancers. This study identified at least 211 direct targets of Mef2 with three different temporal profiles: (1) enhancers bound continuously bound by Mef2 (representing approximately half of the analysed enhancers), (2) enhancers bound early but not late by Mef2 (representing 21% of the enhancers), (3) enhancers bound late but not early (representing 32% of the enhancers). The expression of the target genes coincided with the incidence of a Mef2 binding event of their respective enhancer (Sandmann et al., 2006). Moreover, a study using an allelic series of Mef2 mutants to profile gene expression of its targets has shown that the correct timing of their expression is correlated with the level of Mef2 protein. Target genes expressed earlier in development were shown to require lower levels of Mef2 proteins for their activation compared to genes expressed later in development. Overexpression of Mef2 at earlier time points in the muscle differentiation program coincided with an earlier onset of expression of late-expressing genes (Elgar et al., 2008).

The three groups of genes with different temporal profiles show differences in terms of the number of Mef2 regulatory sites present in their enhancers. Groups (1) and (3) showed enrichment of one or several Mef2 sites per fragment, while early-bound genes contain as many Mef2 sites as the rest of the genome (Sandmann et al., 2006). Certain genes have been shown to require the co-binding of tissue-specific transcription factors together with Mef2 at their CRM in order to fine-tune the timing of expression and levels of their genes. For example, the *Act57B* gene locus is part of the group of enhancers that are bound by Mef2. However, *Act57B* is expressed in the embryo as early as Stage 11 and its expression has been shown to require the presence of Mef2 for its activation at this stage of development (Elgar et al., 2008). Artificially increasing Mef2 levels at this stage was not sufficient to induce premature activation of this locus and only through the additive input of Mef2 and Lmd the *Act57B* expression was shown to be induced in stage 11 (Cunha et al., 2010). Therefore, the way Mef2 is able to find specific DNA sites in order to elicit transcription of certain genes earlier in development must be through cooperation with other cofactors and TFs.

The activation of Mef2 target genes seems to require a very fine balancing of Mef2 levels, since mis-expression of Mef2 leads to defective phenotypes in the tissue it is active in. Loss-of-function mutants of *mef2* do not form differentiated somatic muscle, despite the fact that muscle progenitors are able to develop, the fusion process and any further differentiation process fails (Bour et al., 1995; Lilly et al., 1995; Taylor, 1995). When Mef2 is overexpressed in the mesoderm, the differentiation process of the heart, the somatic and visceral muscles is disrupted (Gunthorpe et al., 1999). In the pacemaker neurons, reduced Mef2 expression causes loss of circadian rhythms due to dampening of the molecular rhythms, while overexpression of Mef2 desynchronises the pacemaker neurons resulting in longer and complex locomotor behaviours in adult flies (Blanchard et al., 2010). Based on the phenotypic responses observed in various tissues where Mef2 was mis-expressed, a particular aspect becomes apparent: there is a certain range of Mef2 levels that are compatible with normal progression of the molecular processes Mef2 is involved in. The regulation of Mef2 levels in *Drosophila* can manifest at transcriptional, translational and post-transcriptional level.

The regulation of Mef2 at gene level displays three types of activators: 1) Activators that turn on *mef2* transcription and co-bind enhancers together with Mef2 to activate target genes (for example Twi, Lmd) 2) Activators that are a direct target of Mef2 and in a feedback loop bind *mef2* enhancers to increase Mef2 production (for instance CF2, Vg) 3) Mef2 acts as its own transcriptional activator. However, the regulation of Mef2 occurs at mRNA and protein levels as well.

In *Drosophila*, Mef2 is part of a functional network with various proteins that modulate its activity at different molecular levels, often forming feed-forward and feedback loops. These network partners can be divided in 4 categories:

1) Autoregulation of Mef2 of its own enhancer in late muscle differentiation.

Mef2 protein is able to bind to a specific enhancer of the *mef2* gene that is active in differentiated muscle cells (Figure 1.3A). The enhancer is active in all embryonic muscle tissues, with detection of expression at late stage 12 in longitudinal visceral muscle precursors, starting with stage 14 in myoblasts of somatic muscle. By stage 16 this Mef2 enhancer showed strong activity in the visceral and somatic muscle, but not in the dorsal vessel. In third instar larvae the expression is maintained in the other muscles and also detectable in the dorsal vessel. The onset of activity of this enhancer resembles the mechanism of late differentiation of all the three muscle types: first the visceral muscles start terminal differentiation, followed by the somatic muscle cells and the dorsal vessel is the last to start differentiation. The activity of this enhancer is only detectable in muscle cells, though it can be ectopically activated by transgenically expressed Mef2. In absence of the Mef2 site, overexpressed Mef2 was unable to drive activation of the autoregulatory enhancer (Cripps et al., 2004).

2) Mef2 levels can be modulated by the products of its own target genes.

The *Drosophila* Chorion factor 2 (CF2) is a protein that collaborates with Mef2 to activate a number of structural muscle genes (e.g. *Actin57B*, *TnI*, *mhc*) (Gajewski and Schulz, 2010; Tanaka et al., 2008), while its own expression is dependent on Mef2 activation (Figure 1.3A). CF2 is detected in all types of muscle lineages in the embryo and its mRNA is not detectable in a null mutant of *mef2* (Bagni et al., 2002), thus the CF2 gene expression is directly or indirectly regulated by Mef2. The model proposed for the interaction between the two proteins is the following: at stage 11

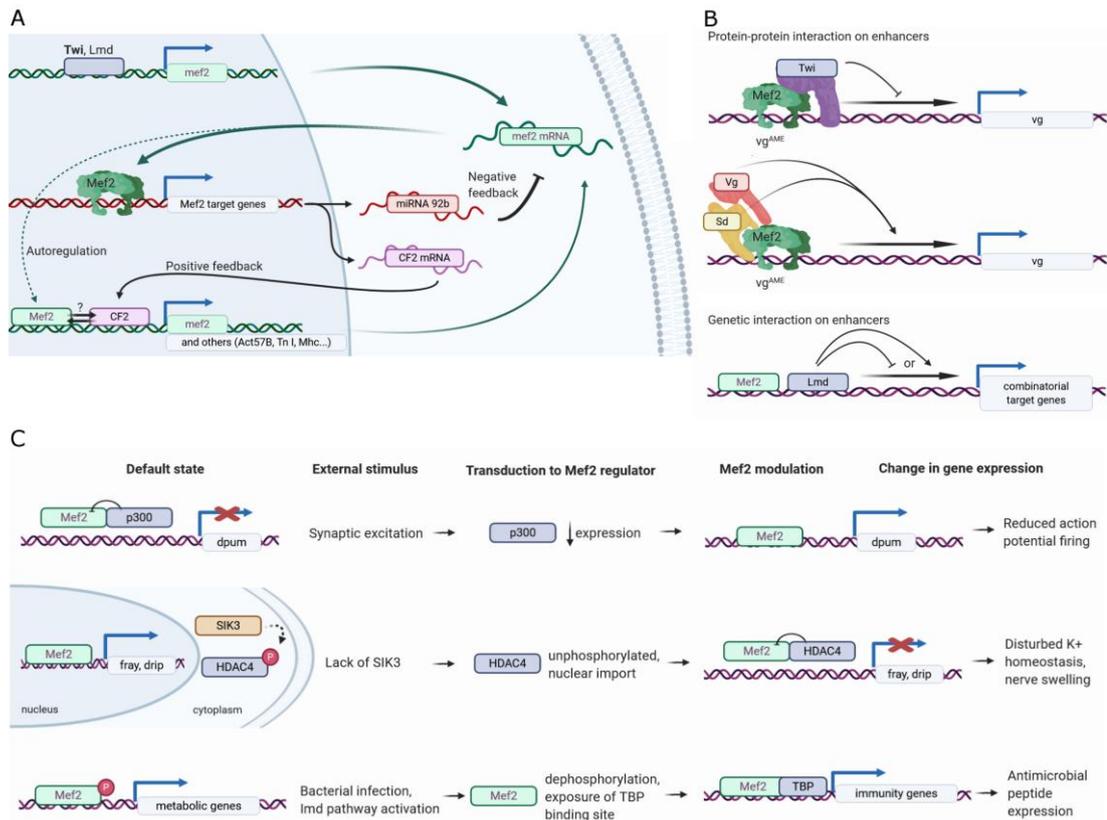


Figure 1.3: Modes of regulation of Mef2 activity

(A) Regulation of Mef2 expression in muscle. Twist (Twi) initiates expression of Mef2. Later, both Twi and Lameduck (Lmd) are able to bind upstream of Mef2 and drive its expression. Mef2 target genes include Chorion factor 2 (CF2), which feeds back positively by binding enhancers of the *mef2* gene to stabilise Mef2 levels. Together, Mef2 and CF2 activate target genes including structural muscle proteins. Mef2 also forms an autoregulatory feedback loop by binding its own enhancer. Another target, microRNA 92b, feeds back negatively onto Mef2 levels by repressing translation of *mef2* mRNA.

(B) Examples of interactions of Mef2 with other transcription factors during *Drosophila* development. Upper diagrams: Mef2 physically interacts with Twi, Vestigial (Vg) and Scalloped (Sd) on the vestigial adult muscle enhancer (vg^{AME}). The interaction with Twi represses *vg* expression, while interaction with Vg and Sd synergistically enhances *vg* expression. Bottom diagram: Mef2 and Lmd co-bind a multitude of enhancers throughout the genome, leading to diverse target gene expression profiles within the same cell context.

(C) Responses of Mef2 activity to external stimuli. Upper diagram: In neurons, Mef2 physically interacts with p300 and modulates specific target genes with effects on action potential firing depending on synaptic input activity. Middle diagram: Mef2 is the nuclear effector of salt-inducible kinase 3 (SIK3) for maintaining K^+ and water homeostasis in glial cells. Histone deacetylase 4 (HDAC4) can translocate to the nucleus when it is unphosphorylated, and repress Mef2 protein from activating transcription. Bottom diagram: In fat body cells, infection signalling triggers dephosphorylation of Mef2, which can then bind TATA binding protein (TBP). This interaction leads to a switch in target genes.

mef2 transcription is activated by *Twf* and *Mef2* is able to maintain its transcription in a *Twf* independent manner by binding to its own enhancer. At stage 12 the *Mef2* protein activates *CF2* transcription, the protein in combination with *Mef2* being able to activate expression of structural proteins and to maintain high levels of *Mef2* in a *CF2* concentration dependent manner. When *CF2* was overexpressed in the muscle lineages, higher levels of *Mef2* were detected, while when *CF2* was knocked-down the *Mef2* levels were reduced compared to the wild-type control, indicating a positive feedback loop (Arredondo et al., 2017).

Another product of a *Mef2* target gene that is able to regulate *Mef2* activity is miR-92b, a negative regulator of *Mef2*. *Mef2* is able to activate expression of miR-92b by directly binding to its enhancers in the heart, somatic and visceral muscles, which present three *Mef2*-binding sites. *Mef2* mRNA contains two targeting sites for miR-92b in its 3' UTR region which allow repression of the *Mef2* translation process. When miR-92b is overexpressed, *Mef2* mRNA and protein levels are decreased and defects in muscle attachment are observed. When miR-92b is deleted, *Mef2* levels are elevated, leading to abnormalities in muscle development. In this case the interaction between miR-92b and *Mef2* forms a negative feedback loop that is able to stabilise *Mef2* levels in order to ensure the muscle development process can continue under normal conditions (Chen et al., 2012).

- 3) *Mef2* transcriptional activators that require collaboration with *Mef2* protein to activate target genes.

One protein in this category is *Twist*, the bHLH transcription factor that induces mesodermal cells down the path of myogenesis. *Mef2* gene expression is only initiated via activation of *Twist* binding its enhancer region in the early stages of embryonic development. A significant part of CRMs bound by the *Mef2* protein in early embryonic development (totalling 42%) is co-bound by *Twist*. *Twist* and *Mef2* have partial overlap in the timing of their expression during embryonic development (Adryan and Teichmann, 2010; Lilly et al., 1994) and the co-bound CRMs regulate genes involved in early mesoderm development. In contrast, *Mef2* alone selectively drives expression of genes involved in aspects of later differentiation such as muscle attachment and sarcomere structure (Sandmann et al., 2006).

Twist protein is able to physically bind Mef2 protein via its C-terminal part and repress its activity (Figure 1.3B). The effects of this interaction can result in repression of the vg^{AME} enhancer, an element that is repressed in swarming myoblasts surrounding developing fibres in *Drosophila* pupae and is a direct target of Mef2 during IFM differentiation. The vg^{AME} enhancer was not found among the embryonic CRMs bound by Mef2 during myogenesis. Twist alone is not able to repress the expression of this enhancer in absence of Mef2 in S2 cells (Bernard et al., 2009).

Lmd (Figure 1.3B, lower diagram) is a tissue-specific modulator of Mef2 activity in *Drosophila* embryogenesis at stages 10-13. Lmd binds a locus upstream of the Mef2 gene and regulates its expression. The Mef2 protein co-binds 68.8% of the Lmd enhancers with a similar temporal profile and these direct target genes are also regulated by Mef2. The transcriptional behaviour of these genes is divergent to the regulatory effects of Lmd and Mef2, showing either an additive, cooperative or antagonistic response. The genes were separated into three clusters based on their changes in expression in *lmd* and *mef2* loss-of-function mutants compared to wildtype controls. Cluster I contains genes that are downregulated in both *lmd* and *mef2* mutants and consists mainly of structural muscle proteins. Cluster II genes have differential responses in the two mutants: upregulated in *lmd* mutants, but unchanged or slightly altered expression in *mef2* mutants. Cluster III genes are decreased in expression in *lmd* mutants and have increased expression in *mef2* mutants. The two TFs were overexpressed ectopically in the ectoderm in order to test by colorimetric fluorescent *in situ* hybridisation the transcriptional response of the shared target genes. When expressed ectopically in the ectodermal stripes, candidate genes from the three clusters had varied responses: cluster I genes were able to be activated ectopically by Mef2 alone and/or by co-expressing the two TF; cluster II genes were activated ectopically by overexpressing Lmd alone and/or by Lmd and Mef2 together; cluster III genes were expressed in the ectoderm under Mef2 overexpression, but their levels were attenuated when Lmd was co-expressed with Mef2, showing that the two TFs have opposing inputs on these target genes. Integrating this enhancer binding information with the transcriptional responses of target genes it becomes clear that there is flexibility in the way the two TFs activate transcription, inducing synergistic, additive or repressive effects (Cunha et al., 2010).

Interestingly, the *lmd* gene is directly regulated by Twi and its expression is dependent on Twi binding its enhancer (Bonn et al., 2012; Furlong et al., 2001; Zinzen et al., 2009). Enhancers found in somatic FCMs were bound by combinations of Twi, Lmd and Mef2, suggesting that Twi is able to start a series of feedforward loops in FCMs that are able to induce gene regulation by Mef2 and Lmd in these cells. These feedforward loops are able to offer spatiotemporal specificity to gene expression patterns (Busser et al., 2012).

Scalloped (Sd) and vestigial (Vg) are two myogenic factors that physically interact with Mef2 both during embryonic muscle development and indirect flight muscles development (Bernard et al., 2009; Deng et al., 2009). The two factors have differential expression in the embryo, Sd being expressed in a subset of developing somatic muscles and in the dorsal vessel, while Vg is found only in the embryonic somatic muscle. The overexpression of the two myogenic markers in developing embryos of stages 12 - 15 can repress Mef2 target genes expression during embryonic muscle differentiation without affecting Mef2 protein levels. The Sd and Vg proteins can help modulate the activity of Mef2 in late-stage *Drosophila* myogenesis, with effects in specific cell types. It is important to note that *vg* and *mef2* are co-expressed in progenitors of somatic muscles as early as stage 11, while Sd and Mef2 are present in cardiac muscle cells, where Vg is not expressed. The three proteins were shown to form a tripartite complex that is expressed in certain stage 16 somatic muscles (Deng et al., 2009). During IFM development the three proteins are found throughout the differentiation process, from muscle progenitors to fully formed muscles. Vg is involved in both establishing IFM identity in myoblasts (Bernard et al., 2003; Sudarsan et al., 2001) and throughout the differentiation process (Bernard et al., 2006). Vg^{AME} is the enhancer that activates *vg* expression in adult muscles and is specifically active in unfused myoblast that undergo differentiation. Mef2 is able to bind this enhancer and activate its expression and the presence of Sd next to Mef2 can synergistically increase the activation of this enhancer *in vivo* (Figure 1.3B, middle diagram). When Vg is cotransfected with Sd and Mef2, the highest levels of activation of the Vg^{AME} are detected in S2 cells. The presence of Vg as part of the activation complex of its own transcription could show the presence of a positive feedback loop where Vg reinforces its own expression (Bernard et al., 2009).

4) Mef2 can influence responsiveness to signalling.

In neurons, synaptic activity regulates the levels of expression of p300 (Figure 1.3C). At normal synaptic activity levels, p300 can repress Mef2 protein by binding it, independently of the p300 histone acetylase activity. Lower levels of this protein encountered under enhanced synaptic activity do not interfere with activation of expression by Mef2, allowing Mef2 to bind the enhancer of *dpum* and induce its transcription. The dPum protein is able to reduce action potential firing by repressing paralytic mRNA translation. This repression is required under enhanced synaptic activity to regulate the neuron's downstream signalling (Lin and Baines, 2019).

The Salt-inducible kinase 3 (SIK3) is the central node that transduces the signalling pathway for K^+ and water homeostasis in glial cells (Figure 1.3C, middle diagram). Knockdown of Mef2 or SIK3 in glial cells by RNAi causes a nerve swelling phenotype, which can be further exacerbated by nuclear HDAC4. Mef2 acts downstream of SIK3, since overexpressing Mef2 in SIK3 mutants rescues the swelling phenotype. When looking at the HDAC4 localisation in glial cells under mis-expression of SIK3, the subcellular localisation of HDAC4 is regulated by SIK3 levels. When SIK3 is overexpressed, HDAC4 accumulates in the cytoplasm and the nerve swelling is suppressed, while when SIK3 is downregulated, HDAC4 accumulates in the nucleus and nerve swelling is exacerbated. The overexpression of a phosphorylation-defective form of HDAC4 in a SIK3 knockdown glial background exacerbates the swelling phenotype compared to when a wild-type form of HDAC4 is overexpressed. Thus, SIK3 is able to suppress the inhibitory role of HDAC4 in glial cells in order to maintain K^+ and water homeostasis. *Fray* and *drip*, two Mef2 target genes with roles K^+ and water homeostasis, were shown to be downregulated under Mef2 and SIK3 knockdown in glial cells. Overexpression of these genes in a SIK3 mutant background is able to rescue the nerve swelling phenotype. Likewise, when larvae of SIK3 mutants are fed a pan-HDAC inhibitor diet, the nerve defects in these larvae are rescued. These genetic interaction experiments point to a model where HDAC4 is trapped in the cytoplasm of glial cells when phosphorylated by SIK3, and Mef2 in the nucleus is able to activate transcriptional programs that control water and K^+ homeostasis. When HDAC4 is dephosphorylated, it shuttles to the nucleus, represses the activity of Mef2 is repressed and transcription of target genes does not

occur (Li et al., 2019). A potential way how HDAC4 is able to repress Mef2 activity is by direct protein-protein interaction via its conserved Mef2 binding site (Yang and Seto, 2008) and physically blocking its access to the DNA. HDAC4 has been shown to be able to change the distribution of Mef2 in the nucleus of neurons when overexpressed (Fitzsimons et al., 2013).

In the *Drosophila* adult fat body, Mef2 acts as a transcriptional switch activating either a battery of genes involved in metabolic functions or a group of immunity proteins necessary to fight bacterial infections (Figure 1.3C, bottom diagram). In healthy flies, Mef2 is phosphorylated at T20 and activates the expression of anabolic enzymes. As flies get infected by gram-negative bacteria, Mef2 loses its T20 phosphorylation which allows direct binding to TATA Binding Protein (TBP). Mef2 and TBP bind enhancers of antimicrobial peptides necessary to fight the infection. When the switch is made from the phosphorylated to dephosphorylated state of Mef2, a reduction in anabolic enzymes is observed. Mef2 activity is critical in immunity activation, since when Mef2 is knocked down in the fat body and an infection occurs, those flies show reduced resistance. The loss of metabolic genes expression is mediated by the *imd* pathway signalling that occurs under Gram-negative infection (Clark et al., 2013).

Based on the multiple levels of regulation of Mef2 in *Drosophila*, it becomes apparent that protein interactions can account for specific outcomes of target genes expression and for different functions that Mef2 can fulfil in various tissues. There are distinctions in responses to Mef2 levels, activity and transcription not only between different cell types, but also within the same cell. Mef2 is able to activate batteries of genes under the impulse of diverse environmental signals and in order to ensure specificity for each context it cooperates with other interaction partners present in the nucleus to modulate gene expression. The interaction can be cooperative, additive, synergistic or even antagonistic.

1.8 Mef2 in vertebrates

Mef2 was first identified in mammalian cell culture (Gossett et al., 1989) and its target DNA element, an A/T rich sequence, was identified in nearly all known muscle genes in skeletal muscle (Black and Olson, 1998). Mef2 is part of the MADS

box family of transcription factors that are essential for muscle genes expression (Black and Olson, 1998; Taylor, 1995).

The *myocyte enhancer factor 2 (mef2)* gene family in vertebrates has four members, denoted Mef2a, Mef2b, Mef2c and Mef2d which share over 65% identity in the Mef2 domain and 90% similarity in the MADS box (Breitbart et al., 1993; Chambers et al., 1992; Leifer et al., 1993; Martin et al., 1993; McDermott et al., 1993; Morisaki et al., 1997; Pollock and Treisman, 1991; Ticho et al., 1996; Wu et al., 2011; Yu et al., 1992), while only one *mef2* gene is conserved in *Drosophila* (Bour et al., 1995; Lilly et al., 1995; Nguyen et al., 1994). The genes encoding these proteins are located on different chromosomes in the vertebrate genomes (Hobson et al., 1995; Martin et al., 1993; Wu et al., 2011). All of these proteins are expressed in a tissue-specific manner and can be alternatively spliced to form multiple isoforms with functional differences (Martin et al., 1994; Zhu and Gulick, 2004; Zhu et al., 2005). The four Mef2 proteins bind to a consensus DNA sequence as homo- or heterodimers by their MADS-box domain (Andres et al., 1995; Pollock and Treisman, 1991; Wu et al., 2011). The Mef2-specific domain is adjacent to the MADS-box and allows high-affinity DNA binding and interaction with other factors (McKinsey et al., 2002a). The expression patterns of the four vertebrate Mef2 proteins have partial overlap and regulate temporally and spatially gene expression in body development and maintenance (Black and Cripps, 2010; Edmondson et al., 1994; Potthoff and Olson, 2007).

The four Mef2 genes are involved in regulation of myogenesis where they interact with other transcription factors to achieve this role. The most well-known transcription factor interactors are the myogenic BHLH regulatory factor family (MRF) that consists of MyoD, Myf5, Myogenin and MRF4 (Black and Olson, 1998; Molkentin et al., 1995; Ornatsky et al., 1997; Weintraub et al., 1991). Mef2 interacts with members of the MRF family in cell culture where synergistically muscle specific genes are activated (Kaushal et al., 1994; Molkentin et al., 1995). The four genes that encode the proteins Mef2a,-b, -c, -d in vertebrates have a functional relationship with the MRF transcription factor family in myogenesis. The MRFs are able to convert fibroblast into myoblasts and the co-transfection of Mef2 can enhance this ability (Molkentin et al., 1995). In mammalian skeletal myogenesis, cellular

signalling activates the expression of MyoD and Myf5 that induce expression of Myogenin, the TF that will activate Mef2 expression (Wang et al., 2001). Myogenin and Mef2c can self-activate their own promoter and can bind their promoters reciprocally in order to maintain their expression levels (Edmondson et al., 1992; Ridgeway et al., 2000; Wang et al., 2001). Muscle differentiation occurs in vertebrates under the cooperation of Mef2 proteins with myogenic bHLH factors.

The four Mef2 proteins have partially redundant functions during myogenesis, but not all of them are essential for myoblast differentiation. Mef2a is required for the differentiation process while Mef2b, -c and -d are dispensable. These proteins have redundant functions in myoblast development, but there are also particular gene programs regulated by only one individual of the Mef2 family. When knocking down a specific Mef2 isoform in C2C12 cells, certain groups of target genes are misregulated and their expression cannot be rescued by the overexpression of another isoform. In this study it was found that Mef2a regulates specifically the most abundant group of proteins, followed by Mef2b, -c and -d in a descending trend. Many genes are sensitive to the activity of two isoforms but very few genes were found to be dependent on all four isoforms (Estrella et al., 2015).

Mef2 genes are able to modulate their own expression via a negative-feedback loop which involves activation of expression of HDAC9, a class IIa HDAC. During *in vitro* muscle differentiation, HDAC9 levels increase followed by a decline to a basal rate as the differentiation process continues and Mef2 protein levels reach a certain threshold. The decline in HDAC9 levels coincides with the increase in Mef2 levels (Haberland et al., 2007). HDAC9 is found to be expressed in skeletal muscle in mice during embryogenesis, but is downregulated postnatally (Zhang et al., 2001). When present in the nucleus, HDAC9 can repress the transcriptional activation via a direct protein-protein interaction with Mef2 (McKinsey et al., 2001). Signalling can induce HDAC class IIa phosphorylation by calcium dependent kinases that produce HDAC9 nuclear export and removal of Mef2 activity repression (Chang et al., 2005).

Mef2 modulates the expression of muscle specific microRNA, miR-1 and miR-133 by inducing the activation of an enhancer located in a region separating the coding regions of these two genes. The two miRNAs have opposing roles: miR-1 inhibits myoblast growth, promoting differentiation, while miR-133 activates myoblast

growth and suppresses differentiation (Chen et al., 2006). The miRNAs post-transcriptionally regulate gene expression by disrupting the translation and stability of the mRNA (He and Hannon, 2004). HDAC4 is a target of miR-1 both during myogenesis and in chondrocytes, whose mRNA is targeted as part of a regulation mechanism to induce differentiation of muscle cells (Chen et al., 2006; Li et al., 2014). Taken together, Mef2 is able to activate miR-1 that will repress HDAC4, a repressor of Mef2 transcriptional activity, thus creating a positive regulatory feedback loop to promote muscle differentiation (Potthoff and Olson, 2007).

Mef2 can also interact with polyoma virus enhancer activator 3 (PEA3) in satellite cells to induce differentiation (Taylor et al., 1997). In *Drosophila* only recently a population of satellite cells have been described and in these cells Mef2 levels are repressed in order to repress muscle differentiation (Boukhatmi and Bray, 2018; Chaturvedi et al., 2017). The role of Mef2 in muscle development further involves cooperation with Homeobox proteins such as Tinman (Cripps et al., 1999), and GATA factors (GATA4) (Morin et al., 2000). The different Mef2 isoforms play roles in a variety of tissues including the nervous system (Shalizi and Bonni, 2005), the immune system (Rao et al., 1998), adipocytes (Sharma and Goalstone, 2005), the endothelium (Lin et al., 1998), chondrocytes (Arnold et al., 2007) and bones (Verzi et al., 2007).

Mef2 activity can be also negatively regulated by interacting transcription factors. Mouse Twist binds to the Mef2 transcriptional activation domain and inhibits its activity (Spicer et al., 1996). In *Drosophila*, Twist is required for *mef2* expression in the mesoderm, thus Twist is a positive regulator of *mef2* in *Drosophila* (Cripps et al., 1998; Taylor et al., 1995).

Genomic studies in vertebrates to study Mef2 function and regulation have been performed in cell culture (Estrella et al., 2015), but not *in vivo* during embryonic development. Undertaking such studies *in vivo* are complicated by the fact that the 4 genes have overlapping expression in multiple tissues. The direct interactors of Mef2c have been reviewed recently and several of these protein were found to be involved in the role of Mef2c in different cell types: development of muscle cells, endothelial cells, immune cell, neurocyte, chondrocyte and other interactors. These

totalled 30 distinct interaction partners and some of these also interact with the other isoforms (Dong et al., 2017).

1.9 Aims and objectives

Drosophila Mef2 is expressed throughout embryonic development by every type of muscle cell, both at early and late stages. Its gene regulatory activity contributes to many different processes in muscle development, ranging from cell differentiation to syncytial fusion and establishment of the sarcomere. Mef2 likely regulates the expression of at least 230 genes in *Drosophila* (Junion et al., 2005; Sandmann et al., 2006). This variety of functions in various cell types must require regulatory systems that govern the activity of Mef2 itself in turn, ensuring that the appropriate subset of Mef2 target genes is activated while the rest remains inactive depending on the needs of each Mef2-expressing cell at different stages. Some aspects of how Mef2 is regulated are already known, such as its accumulation throughout myogenesis and the switching of target genes based on the level of Mef2 present in a cell (Sandmann et al., 2006). However, there are also several known examples of gene regulatory activity where Mef2 requires the physical interaction with one or more cofactors to induce expression of target genes, or where Mef2 is repressed by cofactors. Given the vast number of target genes, study approaches investigating individual candidates seeking to identify the regulatory programme that governs their response to Mef2 activity are unlikely to be efficient to yield a comprehensive picture of Mef2 regulation. It would thus be valuable to complement the existing genomic data on Mef2 DNA binding with proteomic data to identify the cofactors that confer spatial and temporal specificity to its function.

Two large scale affinity purification studies have attempted to use *Drosophila* S2 cells for genome-wide mapping of protein interaction networks (Guruharsha et al., 2011; Rhee et al., 2014). While Guruharsha and colleagues focused on protein complexes derived from whole lysate extractions, Rhee and colleagues focused on transcription-related protein interactions, therefore only using nuclear extracts for purification experiments. Mef2 was used in both studies as a bait protein, among others. In both experiments a variant of Mef2 tagged with FLAG-HA was transiently expressed in S2R+ cells to identify interacting partners. The Mef2 list of interaction

partners derived from whole lysate S2 cells totalled 101 proteins (Guruharsha et al., 2011). From the nuclear extracts a total of 681 unique bait-prey protein interactions were identified for Mef2 (Rhee et al., 2014). Only 67 proteins were found to interact with Mef2 in both of the studies.

The experimental approach used in these studies has an important limitation in the fact that Mef2 was overexpressed for the purification experiments. In *Drosophila*, Mef2 expression and activity is very tightly regulated in time and space and the expression of its target genes is sensitive to particular levels of the protein. Therefore an overexpression of the protein could enrich for particular interactions that are susceptible to binding Mef2 when it is expressed at a particular threshold, especially at higher levels of expression. Moreover, the S2 cells might not express the most representative interactome for Mef2 since it is believed that these cells are likely derived from haemocytes. Mef2 is known to be expressed in different *Drosophila* cell types, primarily the muscle, neuronal and fat body cells.

The current study was intended to identify interaction partners for Mef2 in *Drosophila* embryonic myogenesis. Cell culture is an attractive experimental set up for performing affinity purification studies that require a considerable amount of material. However the low expression levels of Mef2 in S2 cell (The modENCODE Consortium et al., 2010) would imply the necessity to overexpress the protein and one of the first aims of our study was to investigate Mef2 protein interactions at physiological levels. Moreover, studying the Mef2 interactome in embryos at different stages of development would allow to differentiate the dynamics of the Mef2 protein interactions at different time points. Mef2 is expressed at lower levels early in muscle development and at higher levels during late myogenesis (Sandmann et al., 2006). Different enhancers are susceptible to different levels of Mef2 proteins rather than different isoforms (Gunthorpe et al., 1999; Sandmann et al., 2006). Moreover, certain genes have been shown to require certain Mef2 levels to activate their own expression and to physically interact with Mef2 to turn on other target genes (Arredondo et al., 2017). Therefore, in order to get a clear picture of the Mef2 interacting network in the muscle, the best experimental set up is to extract the lysate from the same particular tissue via a method that allows expression of the prey at physiological levels.

The present study therefore aimed to identify as comprehensively as possible the collective of proteins that interact with Mef2. Since Mef2 is a transcriptional activator, its full interactome might contain primarily proteins that interact with it to regulate its induction of gene expression, as well as potentially proteins that are directly regulated by Mef2 on the basis of a protein-protein interaction. This aim was tackled in three objectives, which reflect the three main chapters of this thesis.

The first objective towards this aim was to perform protein purification experiments to extract Mef2 and proteins physically bound to it. The resulting extract could then be subjected to mass spectrometry to identify the proteins contained and create a list of candidate Mef2-interacting proteins. To implement this, a transgenic *Drosophila* line was used where Mef2 is endogenously fused to a GS-TAP tag. This tag contains specific sites that can be targeted using antibody- and streptavidin coated beads to bind and extract Mef2 from a protein extract. Using appropriate buffer conditions, the physical interactions between Mef2 and its binding partners can be maintained. This procedure is known as tandem affinity purification (TAP) and usually produces extracts of higher purity than single-step purification procedures such as co-immunoprecipitation (Li, 2011).

The second objective was to use bioinformatic methods to enrich the protein datasets obtained using TAP/MS with functional and contextual information such that meaningful conclusions can be drawn based on the presence of proteins in the extracts. This is necessary firstly as a quality control, because no purification is without contaminants that would lead to false conclusions if TAP/MS results are interpreted naively. Secondly, this objective takes advantage of the rich computational resources available for the *Drosophila* model organism, covering among others genome-wide screens of interactions between genes and proteins, functional annotations for many genes, and large-screen expression studies of mRNA and other gene products across development and across different tissue types. Using this wide array of data, rich information can be extracted when a meaningful list of protein candidates is analysed.

The third objective was to study further at least one candidate Mef2 regulator identified by the proteomic study. The chosen candidate was HDAC4, which was a potential Mef2 regulator candidate not only based on the proteomics work but also

based on the role of its vertebrate homologues in vertebrate Mef2 repression. Having been identified by TAP, this candidate should physically interact with Mef2, either directly or indirectly as part of a larger protein complex. The aim of these molecular biological and genetic experiments was thus to determine whether HDAC4 plays a role in muscle development and whether it is associated with regulating expression of Mef2 target genes. If so, this would suggest that HDAC4 could act as a repressor of Mef2 in muscle development.

Chapter 2: Materials and methods

2.1 Fly rearing, fly strains and antibodies

Experiments on *Drosophila* were performed at two locations: Cardiff University and EMBL Heidelberg. The flies used in the experiments were cultivated on fly food prepared from recipes specific for the location. All the fly stocks were kept at 18°C and flipped every four weeks. When making crosses the flies were kept at 25°C. All egg collections were made at 25°C for defined time points on a suitable food substrate consisting of apple juice agar plates for small scale collections and in population cages. All the plates were supplemented with a generous amount of 100% yeast paste placed at the centre of the plate.

The fly food at Cardiff University was a mixture containing cornmeal, dextrose, yeast, nipagin and agar. The fly food at EMBL contained malt extract, corn powder, molasses, dry yeast, agar, soy powder, propionic acid and nipagin.

When collecting flies for crosses, the stock tubes were amplified and kept at 25°C during the day and 18°C during the night to optimise female virgins collection. The laying cages were kept at 25°C for all experiments regardless of the size of the collection. Large scale collections were carried out for TAP purifications in population cages, while any other experiments were performed in small collection cages. In all experiments a wild-type control was used.

Table 2.1: *Drosophila* lines

Name	Genotype	Source/Ref	Experiment
Oregon-R (OR)	Wildtype	Lab resource EMBL Fly Facility Resource	TAP purification dHDAC4 experiments
Mef2GSTAP	W*;;Mef2GSTAP	E. Furlong	TAP purification
UAS-dHDAC4	W*;;UASdHDAC4 WT	E. Olsen	Overexpression experiments in muscle for dHDAC4
UAS- dHDAC4ΔC	W*;;UASdHDAC4ΔC	E. Olsen	Overexpression experiments in muscle for dHDAC4
UAS- dHSAC4SA	W*;;UASdHDAC4	E. Olsen	Overexpression experiments in muscle for dHDAC4
UAS-hHDAC5	W*;;UASdHDAC5 WT	E. Olsen	Overexpression experiments in muscle for hHDAC5
UAS- hHDAC5ΔC	W*;;UAShHDAC5ΔC	E. Olsen	Overexpression experiments in muscle for hHDAC5
UAS- hHDAC5SA	W*;;UAShHDAC5SA	E. Olsen	Overexpression experiments in muscle for hHDAC5
CPTI77	w1118,PBac{602.P.SVS- 1}HDAC4CPTI000077	(Knowles- Barley et al., 2010)	Colocalization of Mef2 and dHDAC4
HDAC4 del6	PBac{RB}HDAC4 ^{e04575/e02449} /FM 7 ftzlacZ	J. Han, M. Taylor	dHDAC4 LOF mutants
HDAC4 del48	PBac{RB}HDAC4 ^{e03932/e02449} /FM 7 ftzlacZ	J. Han, M. Taylor	dHDAC4 LOF mutants
handGFP	If/CyO;C3.1 hand-GFP	(Sellin et al., 2006)	Positive control for dHDAC4/GFP expression
DaGal4	w;;DaGal4	(Wodarz et al., 1995)	Rescue experiments for dHDAC4 LOF mutants
Twiptwip-Gal4	w*; P{GAL4-twi.B}2	(Baylies and Bate, 1996)	Gal4 for overexpression experiments in muscle
Dp(1;3)DC266	w1118; Dp(1;3)DC266, PBac{DC266}VK00033	Bloomington (# 30383)	Rescue experiments for dHDAC4 LOF mutants
Dp(1;3)DC267	w1118; Dp(1;3)DC267, PBac{DC267}VK00033	Bloomington (# 30384)	Rescue experiments for dHDAC4 LOF mutants

Table 2.2: Antibodies

Antibody	Host	Source	Working dilution	Experimental purpose
Anti-TAP	Rabbit	Sigma	1/10000 or 1/5000	Detection of GSTAP-tagged Mef2
Anti-SBP	Mouse	Santa Cruz Biotechnology	1/500	Detection of GSTAP-tagged Mef2
Anti-Mef2 (4 different strains)	Rabbit	EMBL in house production	1/1000	Detection of GSTAP-tagged Mef2
Anti-GFP	Mouse	Abcam	1/100	Detection of a GFP-tagged protein used as control for Mef2 levels estimations
Anti-Rabbit-HRP	Goat	Santa Cruz Biotechnology	1/1000	Secondary antibody
Anti-Mouse-HRP	Goat	Santa Cruz Biotechnology	1/1000	Secondary antibody
Anti- B3-tubulin	Rabbit	R. Renkawitz-Pohl	1/1000	Analysis of somatic muscle development during embryogenesis
Anti-DIG-AP Fab fragments	Sheep	Roche	1/2000	RNA <i>in situ</i> hybridisation
Anti-GFP	Mouse	Sigma	1/2000	Study of HDAC4 expression
Anti-lacZ	Rabbit	Molecular probes, Invitrogen	1/5000	Staining for FM7-ftzlacZ balancer in order to distinguish heterozygous from homozygous embryos
Anti-Mef2	Rabbit	Bruce Pearce	1/1000	Colocalization studies with HDAC4
Anti-rabbit 488 IgG	Goat	Molecular probes, Invitrogen	1/200	Secondary antibody
Anti-mouse 546 Igg	Goat	Molecular probes, Invitrogen	1/200	Secondary antibody
Biotinylated anti-rabbit	Goat	Vector Laboratories	1/200	Secondary antibody
Biotinylated anti-mousee	Goat	Vector Laboratories	1/200	Secondary antibody

2.2 TAP purification experiments specific methods

2.2.1 Making Mef2 TAP purification compatible

This study was intended to systematically identify the proteins that associate with Mef2 during *Drosophila* embryogenesis. For these purposes a Tandem Affinity Purification (TAP) experimental strategy was developed where a *Drosophila* line expressing a GSTAP tagged Mef2 was used as bait. The *Drosophila* line was prepared in the Furlong lab. The line was generated via homologous recombination. The transgenic line contained a knock-in of a GS-TAP at the C-terminus of Mef2 (Figure 2.2), in frame with its 10th exon (Figure 2.1). The tag contained two Protein G domains, two TEV-protease cleavage sites, and one streptavidin-binding peptide (SBP) moiety, placing the Protein G domains at the extreme of the C-terminus. The Mef2 protein itself weighs roughly 55 kDa, while the entire tag is estimated to weigh approximately 45.8 kDa. The SBP peptide has a molecular mass of 9 kDa, one TEV cleavage site has a molecular weight of 1.7 kDa and one Protein G domain weighs 14.4 kDa. The molecular weights were estimated with ExPASy based on sequence analysis (Artimo et al., 2012). When performing experiments like Western blots the observed molecular weight may vary from the estimated one due to post translational modifications, relative changes or other experimental factors. The tag used is a variation of the original yTAP tag that has been shown to have better efficiency in purification experiments (Bürckstümmer et al., 2006).

2.2.2 *Drosophila* large population maintenance

Cages for *Drosophila* populations were plastic cylindrical tubes covered at every end with a mesh that did not allow the flies to escape. The mesh was secured on the tubes with rubber bands. For the rear part a shorter mesh, long enough to allow complete coverage of the circular area, was used. For the front part a longer, sleeve-like mesh, was used (Figure 2.3). The frontal mesh allowed handling of the food/egg collection plates in and out of the cage with minimization of the flies escaping. This mesh was knotted to avoid flies escaping. All population cages were kept in rooms that maintained a constant 25°C temperature and provided a day-night cycle.

Prior to populations set-up, *Drosophila* were expanded in bottles and the adult flies

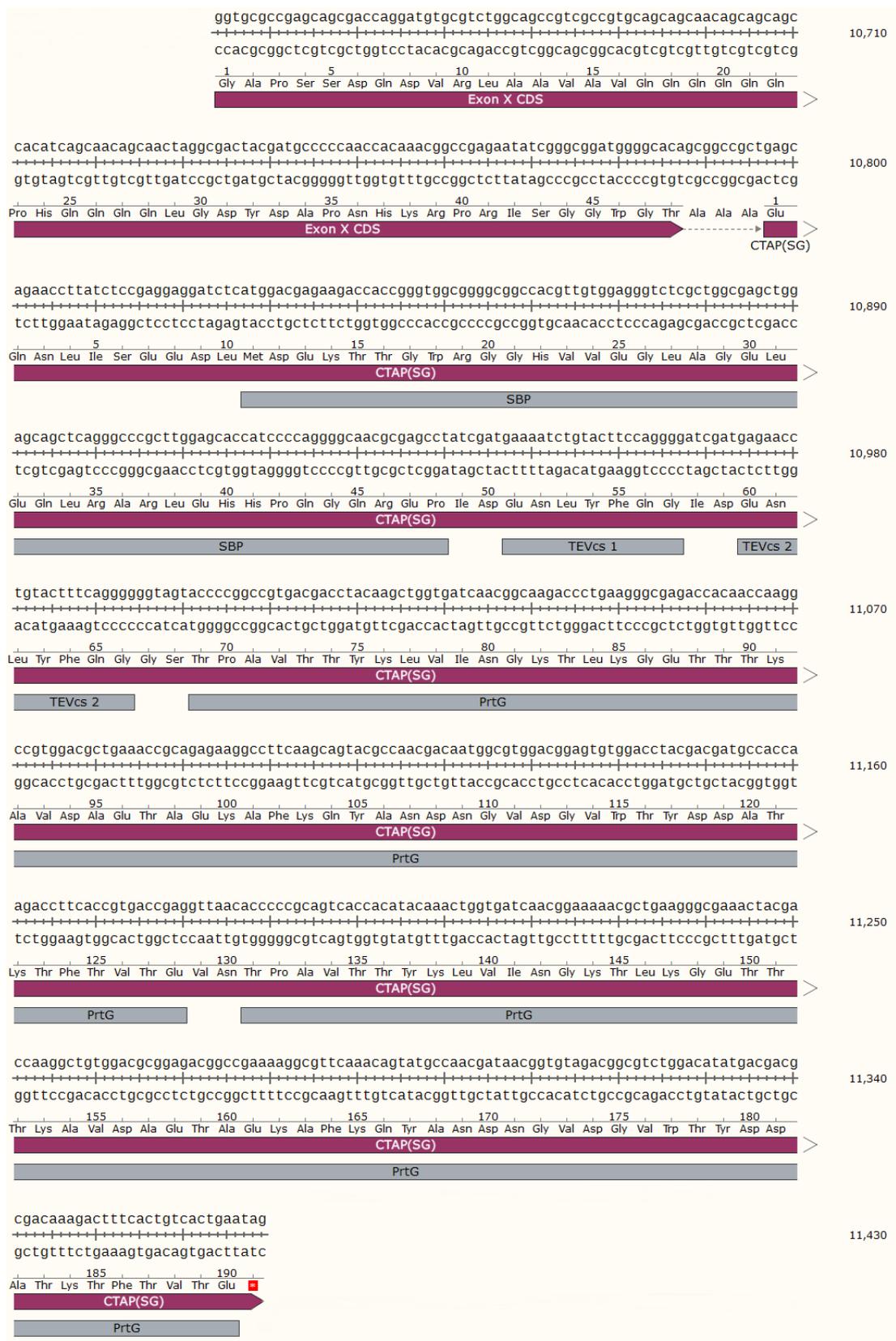


Figure 2.1: DNA Sequence of Mef2 exon X with inserted GSTAP-tag

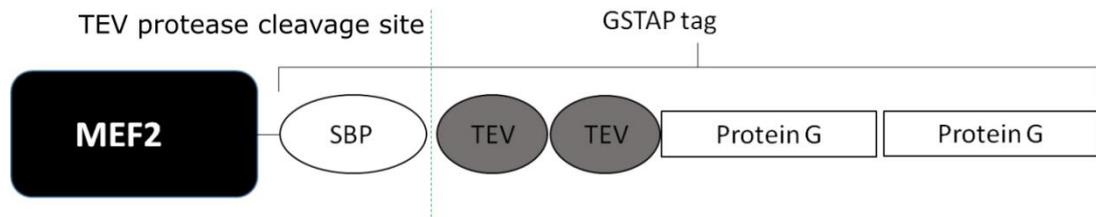


Figure 2.2: Mef2GSTAP fusion protein structure

Domain structure of GSTAP-tagged Mef2. Cleavage site for the tobacco etch virus (TEV) protease is labelled. SBP: streptavidin-binding peptide.



Figure 2.3: A *Drosophila* population cage

A population cage consisting of a cylindrical plastic tube which has the ends covered by mesh that impedes the flies from the population to escape. The larger mesh at the front that is knotted allows access to the inside of the cage for feeding and eggs collections.

were transferred to the cage. For the next cycles of *Drosophila* flies maintained in population cages the adult flies were collected from “larval boxes”. Before addition of the adult flies to the population cages, the flies were anesthetised with CO₂ and poured into a conical shaped flask. The flies were weighted and added in sequential steps to the population cages until a maximum weight of 30 g of total flies/cage was added. The flies were allowed to recover and two apple-juice plates supplemented with yeast were added to the bottom of the tube. The two apple-juice plates had a double role: as the only food source for the flies inside the cages and as the collection plates for embryos.

Once the flies were transferred into the cages, the population was maintained and expanded through alternating cycles of embryos collection for larval boxes set-up, growth of new adult flies in larval boxes that will be used to prepare new population cages. Each cycle of egg collection for larval boxes preparation did not last longer than 14 days and the cages were supplied with fresh food twice a day to avoid larval growth inside the population cages.

2.2.3 Embryo collection from large collection plates for larval boxes set up

In order to collect the embryos for larval boxes three large sieves having different mesh sizes were used. The top sieve had 710 µm mesh and did not allow any dead adult flies to go through, the intermediate one had a 356 µm mesh which did not allow fly heads, and potential larvae to go through, while the bottom one (112 µm) collected all the embryos. The use of the three sieves one on top of the other allowed to collect only the embryos and to wash off any debris such as yeast. Distilled water (dH₂O) and brushes were used to allow transfer of the embryos from the plates to the sieves. The plates were apple juice agar plates supplemented with yeast where the females had deposited the eggs during collections. The cleaned embryos with dH₂O were transferred from the bottom sieve to a cylinder and dechorionated for 2 min in 50% bleach. A magnetic stir bar was used to mix the embryos during the dechorionation step. The dechorionated embryos were transferred back to the 112 µm sieve and washed with dH₂O to remove any traces of bleach. The embryos were then transferred to a cylinder in 1x PBS + 0.1% TritonX 100 (PBS-Tx). Using a filter pump to remove the PBS-Tx, the embryos were placed onto a filter paper. On the filter paper the embryos were washed once with 70% ethanol and once in 1x PBS-

Tx. The filter containing the embryos was placed in the larval boxes on top of the food and spread evenly on the food surface with a small amount of PBS-Tx. The layer of embryos transferred in one larval box did not exceed a height of 0.5 cm to avoid overcrowding the box.

The larval boxes were plastic Tupperware boxes with the lid presenting a hole in the middle covered with mesh to allow oxygen flow without allowing the larvae to escape. Tupperware boxes were tightly sealed such that the larvae did not escape. The food was placed at the bottom of the boxes on a few layers of tissue paper. The larval boxes were shifted between 18°C and 25°C incubator to delay or speed up development such that the adult flies were synchronised for population cages set-up.

2.2.4 Embryo staging and collection

The correlation of desired developmental stages was transformed in time points of development at 25°C. Developmental 2 h windows were used to capture the most representative stages of muscle development in *Drosophila* embryos. The staged embryos were collected in two-hour windows of egg laying on apple juice-agar plates and then aged at 25°C until the desired age. The collected stages are described in Table 2.1. Each egg laying period for embryo collection was preceded by three consecutive 1 h prelays (1 h long egg-laying time points). The prelays allowed to collect any eggs that the flies did not manage to lay due to the overcrowding of the plates with eggs, therefore these eggs would be of unknown age and could have biased the experimental collection. After the two hour egg-laying period the plates were removed from the population cage and the embryos were allowed to develop until the desired age. O/N embryos collections were timed as 18h of continuous egg laying time, no separate aging step required.

Table 2.3: Stages of embryos collected

Stage	Egg laying time 25°C	Aging time 25°C
4-6 h	2 h	4 h
6-8h	2 h	6 h
10-12 h	2 h	10 h
11-13 h	2 h	11 h
O/N (0-18 h)	18 h	0 h

2.2.5 Embryos preparation for TAP

The stage collected embryos as described in section 2.2.4 were transferred from the plates into the three sieve system described in section 2.2.3 and thoroughly washed with dH₂O to remove any debris. For chorion removal the embryos were transferred into a cylinder containing 50% bleach where they were incubated for 2 min under stirring. For washing off the bleach the embryos were transferred through a sieve which contained a detachable mesh of 112 µm pore size. This sieve allowed thorough washing of the embryos with dH₂O in order to remove traces of bleach. The removable mesh containing the embryos was placed on a tissue paper to absorb excess moisture. The embryos were left to air dry for approximately 5 min. When suitably dry the embryos were transferred with a spatula into a 15 ml or 50 ml tube, weighed and snap frozen in liquid Nitrogen. The snap frozen embryos were stored at -80°C.

2.2.6 Lysis of *Drosophila* embryos for TAP purifications

All steps were performed at 4°C unless otherwise stated. The embryos were removed from the -80°C freezer and placed on ice to partially thaw. The ratio of lysis buffer volume to amount of embryos used was 1:3, meaning for 1 g of embryos to be lysed, 3 ml of lysis buffer were used. The lysis buffer composition was the following: 50 mM Tris pH 7.5, 125 mM NaCl, 1.5 mM MgCl₂, 0.2% NP40, 5% Glycerol, 1 mM EDTA, 1 mM DTT, 10 µg/ml DNase, 25 mM NaF, 1 mM Na₃VO₄ (phosphatase inhibitors), protease inhibitors mix (2 mM AEBSF, 40 µM E-64, 0.5 µM Aprotinin, 1 µM Leupeptin, 60 µM Pepstatin A, 10 µM Bestatin (Sigma)). The EDTA, DTT, DNase, protease and phosphatase inhibitors were added fresh before lysis. The buffer conditions were derived from Bürckstümmer et al and the protease inhibitors mix was based on a protocol routinely used in the Gavin lab (Bürckstümmer et al., 2006; F. O'Reilley, personal communication).

The partially thawed embryos were transferred to a 15 ml Wheaton douncer provided with two pestles, loose and tight. The loose pestle was used to break the cell membranes, while the tight pestle is used to mechanically disrupt/lyse the cells. The homogenisation was performed with 20x strokes with the loose pestle followed by 20 strokes with the tight pestle. If the required volume of lysis buffer surpassed 8ml, the

homogenisation was done in subsequent steps. Another volume of lysis buffer was used to rinse all the tubes and the Wheaton douncer and 5 complete strokes with the loose pestle and 5 complete strokes with the tight pestle are used to homogenise any leftover material from the wall that was not homogenised.

The homogenate was transferred to thick wall Beckham Polycarbonate Thick Wall centrifuge tubes and centrifuged at 100000 g/49000 rpm for 1 h in order to clarify it. A lipid layer was formed at the surface of the sample after the ultracentrifugation. The cytosolic fraction was carefully removed with a 1 ml syringe with an orange tip (25 G). The removal of lipids was done carefully to avoid mixing the clarified supernatant or to collect any debris from the pellet. To recover as much as possible from the cytosolic fraction that got mixed with the lipid layer, further centrifugation steps were applied. Three subsequent steps of centrifugation at 14000 rpm for 5 min of the mixed lysate, followed by needle extraction were performed. A new needle was used for each new extraction of the cytosolic fraction as the lipids tended to stick to the needle. The obtained cytosolic fraction was immediately used as input material for the TAP purification.

2.2.7 GS-TAP purification

This TAP purification protocol was adapted from a protocol from the Gavin lab that was described in several studies that were using proteins tagged with the γ TAP as bait (Kühner et al., 2009; Puig et al., 2001; Rigaut et al., 1999). The modifications made to the protocol were taken from the Bürckstümmer et al study because we were working with a different tag, specifically the GSTAP (Figure 2.4) (Bürckstümmer et al., 2006).

All steps were performed at 4°C unless otherwise stated. IgG-Sepharose (IgG Sepharose 6 Fast Flow, GE Healthcare) beads were washed twice in lysis buffer by mixing 400 μ l of slurry (200 μ l of actual IgG beads) with 1.5 ml 1x lysis buffer in a 15 ml Falcon tube (Figure 2.4). The beads were spun down at 1800 rpm for 1 min and the buffer was removed. The cytosolic fraction of the *Drosophila* embryos was added to the washed IgG beads and incubated for 2h on a spinning wheel. The IgG beads bound by the bait protein were spun down at 1800 rpm for 1 min and the supernatant (biological material free of bait and associated proteins) was removed.

The bound IgG beads were resuspended in 400 μ l 1x lysis and transferred to a small Mobitec (0.8 ml, M1002) column. An additional 1 ml of lysis buffer was used to rinse the falcon tube for any remaining beads left on the walls of the falcon tube. The beads in the Mobitec column were washed with 10 ml 1x lysis buffer followed by 4 ml of TEV cleavage buffer (TEVcb, 10 mM Tris pH 7.5, 100 mM NaCl, 0.1% NP40, 2 mM DTT) and the buffers were allowed to pass through the column by gravity flow. A mixture of 400 μ l TEVcb and 21 μ l TEV protease (final concentration 0.05 μ g/ μ l) was added to cleave the bound bait protein off the beads. The cleavage was done by incubating the beads with the TEV protease for 1h at 16°C, while shaking at 500 rpm in a thermoshaker.

200 μ l of Streptavidin Sepharose beads (Ultralink Immobilized Streptavidin Plus, Thermo Scientific) slurry (100 μ l beads) were pipetted into a new Mobitec column and washed with a total of 3 ml TEVcb. The TEV cleavage product was directly eluted onto the new Mobitec column containing the Streptavidin beads. When everything was eluted, the column was sealed and incubated for 1h on a spinning wheel to bind the cleaved bait protein to the streptavidin beads. The buffer, now free of bait protein was allowed to flow through and the bound streptavidin beads were washed with 6 ml TEVcb. The bait protein was elute off the beads by incubating with 150 μ l of 2 mM biotin (Sigma) for 5 min. The proteins were eluted off the column by gravitational flow into a 1 ml microcentrifuge tube, snap frozen in liquid nitrogen and stored at -80°C until required.

Both extraction procedures that were used as a basis to develop the protocols for extraction of Mef2 interacting proteins are protocols optimised for purification of solubilised proteins from whole cell extracts (Bürckstümmer et al., 2006; Kühner et al., 2009). The Mef2 proteins that was used as bait is a nuclear protein and could be successfully solubilised using this procedure (see western blot results in Chapter 3, e.g. Figure 3.4). Therefore the procedure should extract both nuclear and cytosolic proteins that interact with Mef2 in *Drosophila* embryos.

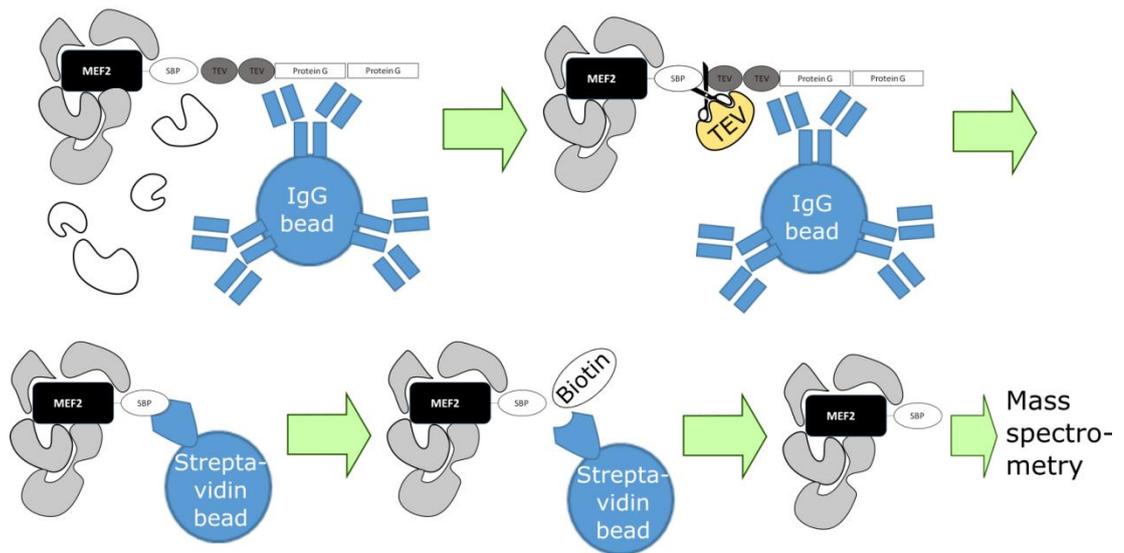


Figure 2.4: Tandem affinity purification (TAP) procedure

The cell extract is incubated with IgG beads that recognise protein G to isolate the bait and its interacting protein complexes. The beads are purified from the suspension and incubated with TEV to release the bait. The resulting solution is incubated with streptavidin beads to clean up the bait further and remove TEV protease. The beads are purified and incubated with biotin, which competes with streptavidin on the beads for binding to SBP.

2.2.8 Western blots

2.2.8.1 SDS-polyacrylamide gel electrophoresis

SDS-polyacrylamide gels were prepared fresh for routine western blot analysis of lysis samples or TAP purifications. For 1.0 cm thick gels a 10% resolving gel was prepared (375 mM Tris pH 8.8, 10% Acrylamide (30% acrylamide/bisacrylamide solution, Biorad), 0.1% sodium dodecyl sulphate (SDS), 0.2% ammonium persulfate (APS, Thermo Fisher), 0.5% Tetramethylethylenediamine (TEMED, Bio-Rad)) and a 5% stacking gel (125 mM Tris pH 6.8, 5% Acrylamide, 0.1% SDS, 0.2% APS, 0.1% TEMED). For small gels a volume of 5 ml of 10% resolving gels and 3 ml of stacking gel were sufficient to obtain a 72 x 86 mm gel. For large SDS-polyacrylamide gels 50 ml of 10% resolving gel and 30 ml of 5% stacking gel were prepared to make a 594mm x 841mm gel. For small gels the maximum volume to be loaded per well was of 30 µl and for large gels of 100 µl. The PageRuler Prestained Protein Ladder (Thermo Fisher) was used as a reference for protein sizes both for gel analysis and western blots. The loading for Input samples and cytosolic fraction was done depending on total protein content as determined by Bradford assay not exceeding 20 µg of total protein loaded per lane. For other types of samples a 10% volume of the sample was loaded unless otherwise stated. For loading the protein samples, an in-house 4x SDS loading dye (200 mM Tris pH 6.8, 400 mM DTT, 8% SDS, 0.4% Bromophenol blue, 40% glycerol) was added to the samples that were denatured by boiling for 10 min at 100°C. The gel was run in Laemmli electrophoresis running buffer (25 mM Tris base, 192 mM Glycine, 0.1% SDS, pH 8.8). For small gels the Mini Portean Casting and Electrophoresis (Biorad) system was used, while for casting and running the large gels an EMBL in house system was available. Precast gradient gels (4-20% Mini-PROTEAN TGX Gels, Biorad) were used for analysis of the TAP purification experiments and for preparation of eluted proteins for gel extraction before MS analysis.

2.2.8.2 Protein transfer

Following SDS-PAGE, the gels and membranes were equilibrated in transfer buffer (25 mM Tris, 192 mM Glycine, 20% (v/v) Methanol pH 8.3). The transfer of proteins was done on Nitrocellulose membranes (Thermo Fisher) via a semi-dry transfer method. The membrane, gel and Whatman papers (3 layers each side) were

thoroughly soaked in transfer buffer. The transfer was done overnight at very low voltage (25 mA, depending on the size of the membrane) due to improved resolution of the bands.

2.2.8.3 Antibody incubation

All the steps were performed at room temperature unless otherwise stated. After transfer, the membranes were washed 3 times 20 min in TBST (19 mM Tris base, 137 mM NaCl, 0.1% Tween-20 (Biorad)), and blocked for 1 h in 5% milk (low fat powder diluted in TBST). The primary antibody was diluted appropriately in 1xTBST+ 5% milk and incubated overnight at 4°C with shaking. The next day the membrane was rinsed 3 times in fresh TBST and washed 3 times 20 min in TBST while shaking. The HRP-conjugated secondary antibody was diluted in 1xTBST+ 5% milk and incubated for 1 h at room temperature with the membrane. After incubation with the secondary antibody the membrane was rinsed 3 times followed by three 20 min washes in TBST.

2.2.8.4 Signal detection

The membranes were coated in Amersham ECL Detection Kit (GE Healthcare) and incubated in the dark for 5 min with the developing solution. The excess developing solution was removed, the membranes were covered with clear film and transferred to a film developing cassette that allowed transport of the membrane in darkness to the developing room. The signal detection was performed in the dark on film (Amersham Hyperfilm ECL, GE Healthcare) by exposing the film to the signal for various amounts of times: as short as 1 sec and as long as 30 min. The films were developed on a Kodak RP X-OMAT Processor Machine, Model M6B.

2.2.9 Sample preparation and mass spectrometry analysis

The sample preparation and mass spectrometry identification was performed by the Proteomics Core Facility at EMBL. The TAP purified proteins were loaded onto gradient SDS-polyacrylamide gels (4-20% Mini-PROTEAN TGX Gels, Biorad), run a few centimetres into the gel and stained by a Methanol free Coomassie Stain. The samples were prepared by in-gel digestion with trypsin, the peptides were extracted from the gel by incubation with acetonitrile and cleaned and eluted by ZipTip. The extracted peptides were loaded onto a Thermo Orbitrap Velos Pro and Q-Exactive

where they were analysed by a AmCit/ β -Casein/MSA 1h method. The identified data was compared against the *Drosophila* sequence available in UNIPROT and the spectra were parsed using Scaffold. The reported outputs were the list of identified protein in each sample together with the number of unique peptides and the unweighted spectrum count (that show how often peptides belonging to a protein was selected for MS). The filters applied to compile the final list of data was the following: minimum 2 peptides per protein needed to be identified for that candidate to make it to the final list and a peptide needed to have a minimum Mascot score of 20 such that the peptide is taken into account for a protein.

2.3 Bioinformatic analysis specific methods

2.3.1 Normalised Spectral Abundance Factor and estimation of contaminants

The Normalised Spectral Abundance Factor (NSAF) was determined for each protein in each sample based on the total spectral counts identified by mass spectrometry. To calculate NSAF, first all proteins' total spectral counts were normalised to each protein's respective sequence length as reported on Uniprot, because larger proteins are more likely to generate more peptides that can be detected by MS (Zybailov et al., 2006):

$$\text{Size - adjusted counts} = \frac{\text{Total spectral counts}}{\text{Protein sequence length}}$$

NSAF was then calculated by dividing each protein's size-adjusted counts by the sum of all size-adjusted counts in the sample:

$$NSAF_P = \frac{\text{Size - adjusted counts}_P}{\sum_{\text{Sample}} \text{Size - adjusted counts}}$$

To classify candidate proteins as potential contaminants or specific bait-binding proteins, a ratio of vector magnitudes (α) was calculated to compare NSAFs between purifications containing bait (TAP experiments on Mef2-GSTAP expressing embryos) and purifications containing no bait (TAP experiments on WT embryos). For each protein present in at least one of the included datasets, a vector was constructed, containing the respective NSAF values of that protein across the four

TAP datasets. These vectors were combined into a matrix such that each row represented one protein and each column represented one TAP experiment. The α coefficient was then calculated according to the formula (Sardiu et al., 2008):

$$\alpha = \sqrt{\frac{y_{i1}^2 + y_{i2}^2}{x_{i1}^2 + x_{i2}^2}}$$

where $x_{i1/2}$ = NSAF in Mef2-GSTAP embryos (ON and 11-13), and $y_{i1/2}$ = NSAF in WT embryos (ON and 11-13). Proteins for which $\alpha > 1$ were classified as potential contaminants. Proteins only present in control samples have $x_{i1} = 0$ and $x_{i2} = 0$, leading to a division by 0 and making the α coefficient invalid. These proteins were classified as contaminants by default.

2.3.2 Data collection and identification of protein interaction networks

To construct the Protein-protein interaction (PPI) network for the Mef2-interacting candidates the *Drosophila* Interaction Database (DroID) (Murali et al., 2011; Yu et al., 2008) version 2015_12 was used. The database included 561,842 interactions, experimentally detected in 3 yeast two hybrid studies, curated literature-derived PPI from other databases, two large scale coAP/MS studies, *Drosophila* interlogs predicted from experimental data in yeast, worm and human, genetic interactions, miRNA-gene interactions and transcription factor-gene interactions.

TAP identified candidate proteins, Mef2 interacting proteins identified in a co-affinity purification (CoAP) study in S2R+ cells (Rhee et al., 2014) and overlapping candidates have been used as starting proteins to generate interaction networks. The interaction network representation is based on graph theory, mathematical structures that are used to model pairwise relationships between objects. A graph consists of nodes and edges that connect these nodes. In a PPI, one protein represents one node, while an edge between the nodes signifies the existence of some sort of interaction between them. These networks represent the framework to assess if the identified candidates during the purification studies are functionally relevant. The proteins used to generate the networks will be referred to as seeding nodes (seeds). Two proteins are considered to interact in the generated network if there is any known association that can be derived from the DroID database.

The network was generated computationally via the DroID plugin compatible with the Cytoscape 2.8.3 software. This plugin allows to extract all the interaction data available in the DroID database connected to all the proteins provided as input nodes (seeds). The network building function ensures that there are no redundant proteins in the connected network.

2.3.3 Identification of shared and unique seeds in the analysed samples

The Mef2-interacting candidates shared among the different purification experiments were identified using the Draw Venn Diagrams web applications which compared the list of proteins and represented them as Venn Diagrams.

2.3.4 Protein Identifiers

Uniprot database (The UniProt Consortium, 2019) was used to map between UniPROTKB accession numbers, FlyBase gene identifiers and Gene Symbols. Flybase IDs that map the genes encoding the analysed proteins were used as the identifiers for PPI network generation and analysis. UniPROTKB accession numbers were used to compute functional similarity between proteins. Gene Symbols were used to map the proteins with RNAseq information for their transcript.

2.3.5 Shortest distance derivation

The shortest path between identified candidate proteins and Mef2 was calculated using the NetworkX 1.10 module in Python 2.7. The distance could take three values: 0 if there is a direct interaction with Mef2, 1 if another common interaction partner is present between Mef2 and the other seed and 2 if two intermediates stand between Mef2 and the analysed candidate. Based on the algorithm used to generate the PPI network, if two intermediate partners are required to connect one seed to Mef2, at least one of the interaction partners needs to be a seed.

2.3.6 Expression data

To identify potential Mef2 interacting candidate proteins acting in similar biological contexts at particular time points the expression of their analysis was performed. The genes encoding the proteins were classified based on the specificity of expression across developmental stages. RNA-Seq data from 30 developmental stages spanning

from *Drosophila* embryos to adult have been downloaded from the modENCODE project (Graveley et al., 2011). Based on their expression pattern across all developmental stages the proteins were classified as ubiquitous if they were expressed across all developmental stages, stage specific if they were expressed predominantly in one and no more than four consecutive stages and non-ubiquitous non-specific (NUNS) if a particular pattern could be identified. A specificity scale was devised to quantitatively classify the genes based on their expression pattern (Murali et al., 2014).

2.3.7 Gene expression specificity scale

The expression specificity (E_{s_i}) of a gene across developmental stages is a fraction of the total abundance of that gene across all the 30 developmental stages and was calculated as follows:

$$E_{s_i} = \frac{RNA\ abundance\ stage\ i}{\sum_{i=1}^{30} RNA\ abundance\ stage\ i}$$

Where *RNA abundance stage i* is the raw expression of a particular gene at stage i. The values obtained will be between 0 and 1. The sum of all expression specificities for each gene across all developmental stages equals 1.

Genes labelled as specific have E_{s_i} values of ≥ 0.19 ; genes expressed across all stages and are labelled as ubiquitous have stage E_i values > 0.005 across all time points; NUNS were genes with nonzero E_{s_i} values ≤ 0.005 . Genes with 0 E_{s_i} values across all stages were labelled as not expressed.

2.3.8 An expression filter to find genes active in particular contexts

The expression filter was created on the hypothesis that a protein is more likely to be active at expression levels that approach its maximal level across developmental stages. Therefore a gene's expression level was calculated at each stage as a percentage of the maximal expression value achieved across developmental stages. Each gene had one percent maximum (pmax) value for each stage.

The pmax of a gene in stage i is calculated as follows:

$$pmax_i = \frac{RNA\ abundance\ stage\ i}{\max\ RNA\ abundance}$$

Where *RNA abundance stage i* is the raw expression of a particular gene at stage *i* and *max RNA abundance* is the raw expression of the genes maximal expression level. The values obtained will be between 0 and 1. The *pmax* of the developmental stage where the gene is maximally expressed equals 1. The RNA expression data was derived from RNAseq data from modENCODE (Graveley et al., 2011).

A *pmax* value of above 0.45 is considered to give the highest confidence that the curated list of gene are active in a similar biological context.

The connection between developmental stages and the *pmax* values of all the analysed proteins was analysed by creating a distance matrix and then clustering them hierarchically. The Morpheus platform was used to plot the values (Gould, 2012).

2.3.9 Gene ontology enrichment

The candidate proteins were annotated functionally by gene ontology (GO) enrichment. The GO enrichment analysis was performed using BINGO (Maere et al., 2005), a tool that determines overrepresented GO categories in a set of genes. BINGO version 3.0.3 compatible with Cytoscape version 3.3.0 was used. The Gene Ontology version 1.2, release 2016-04-20 associated with *Drosophila melanogaster* model organism was retrieved from the Gene Ontology Consortium website (Ashburner et al., 2000). The overrepresented GO categories were determined by a hypergeometric statistical test with a Benjamini and Hochberg False Discovery Rate correction. The *Drosophila* genes association with specific GO terms was downloaded from FlyBase (McQuilton et al., 2012) release 2016_02.

2.3.10 Modular analysis

Scale free biological networks are modular (Hartwell et al., 1999), many functionally related nodes tending to cluster together to perform a particular function (Davis et al., 2015). The densely connected seed nodes of the network were analysed using the MCODE version 1.4.2 (Bader and Hogue, 2003) from Cytoscape. For MCODE

clustering the default parameters: degree cut-off value of 2, node score cut-off of 0.2 and k-core clustering value of 2 were used.

2.3.11 Functional similarity

Functional similarity between proteins was computed based on a semantic measure of the GO terms associated with them. FunSimMat release 6.0 (Schlicker and Albrecht, 2008, 2010) was used to measure the functional similarity between each pair of proteins. For this analysis the proteins identified by their FlyBase IDs were mapped to UniProtKB accession numbers. If multiple UniProtKB entries were associated to a single initial gene, all the mapped entries were included in the semantic analysis where the maximum functional similarity obtained by any of them was considered for further exploration. The functional similarity is based on GO terms obtained from the Gene Ontology Annotation database release January 2012 (Barrell et al., 2009). The functional similarity score obtained from summing up the score for each GO category (biological processes, molecular function, cellular component) was used as the criterion to construct a similarity network. The nodes in this network corresponded to proteins identified in the Mef2 interacting candidates and an edge was connecting two nodes if the functional similarity score was ≥ 0.7 (Zanon et al., 2013).

2.3.12 Pathway enrichment and ontology analysis

The ClueGO plugin version 2.2.5 (Bindea et al., 2009) was used to analyse the clusters obtained for modular analysis of the networks derived for the study of Mef2-interacting candidates.

2.3.13 Data availability statement

All supplemental data files can be found in the figshare repository at <https://doi.org/10.6084/m9.figshare.12445634>. In addition to tables S1 to S4, this provides: tables with overlap groups in Venn diagrams, pmax filter dataset (section 4.2.10), Mef2 biological process GO terms and GO term clustering dataset (section 4.2.11), functional similarity matrix (section 4.2.12).

2.4 HDAC4 as a potential Mef2 interactor specific methods

2.4.1 Embryo collection and fixation

The *Drosophila* strains used in Chapter 5 for embryo collections can be found in Table 2.1. Embryos were collected in small collection cages on fresh apple juice-agar plates supplemented with fresh yeast for 18h at 25°C to allow collection of an array of embryonic developmental stages. Embryos were transferred from the apple juice-agar plate using a brush and water to a basket with a wire mesh base. The eggs were washed with water to clean the yeast away. A solution of 50% bleach was used to remove the chorion membrane (approximately 2 minutes). Dechoriation was monitored under a dissecting microscope and was allowed to proceed until the embryos became shiny and the dorsal appendages were removed. Dechorionated embryos were rinsed thoroughly with water, dried on tissue paper and transferred into a petri dish containing heptane and afterwards transferred to a 2ml tube containing a final volume of 1 ml heptane. Another 1 ml of 7.4% paraformaldehyde/PBS was added to the tube and embryos were then fixed for 20 minutes under constant shaking. Afterwards, all embryos were resuspended into the heptane phase and the paraformaldehyde (lower phase) was removed using a glass Pasteur Pipette. A mixture of 1:1 of heptane and methanol was used to remove the vitelline membrane by vortexing for 30 s. The embryos were washed 3 times in methanol to remove any traces of heptane and were stored at -20°C in methanol until required.

2.4.2 Immunohistochemistry – single antibody staining

The immunohistochemistry protocol is derived from (Rushton et al., 1995). Fixed embryos were rehydrated using a 1:1 mixture of methanol and 1x PBS containing 0.1% Triton X-100 (PBS-Tx), washed 3 times in 1x PBS-Tx and then blocked in PBS-Tx + 0.5% bovine serum albumin (BSA) (Sigma) for 30 minutes at room temperature. Embryos were then incubated under constant shaking overnight at 4°C with primary antibodies diluted in PBS-Tx (for working dilutions of specific antibodies used in Chapter 5 please see Table 2.2). All the primary antibodies used in Chapter 5 were preabsorbed against very young (0-2h) fixed OR embryos and these antibody stocks were diluted prior to incubation with fixed embryos. Embryos were

washed 3 times with PBS-Tx for 10 min and incubated with a 1:200 dilution of biotinylated secondary antibodies (anti-mouse, anti-rabbit, all from Vector Laboratories) for one hour at room temperature with shaking. The signal was amplified using the reagents A and B from the Vectastain Elite ABC kit (Vector Laboratories) diluted 1:100 in PBS-Tx. This solution required 30 min pre-incubation at room temperature before adding it to the embryos to allow formation of the amplification system. Incubation of embryos with the A+B complex solution was for 30 min at room temperature without shaking. The stain was then developed with 0.5mg/ml 3,3'-diaminobenzadine-tetrahydrochloride (DAB) (Sigma) and 0.02% hydrogen peroxide and observed under a dissecting microscope. To stop the colour developing reaction, the DAB and hydrogen peroxide were washed away with PBS-Tx. Embryos were mounted in 80% glycerol, staged and selected for microscopic analysis. OR embryos were stained in each case and used as wildtype controls. Embryos were viewed on a *Zeiss Axioskop* microscope in bright field.

2.4.3 Immunohistochemistry - double antibody staining

Double antibody stainings were performed as described in the previous section with the following modifications: When primary antibodies were raised in the same animal, embryos were incubated overnight at 4°C with the first (weaker) antibody and developed using nickel salts which give a dark coloration. Embryos were then washed, blocked with PBS-Tx + 0.5% BSA and incubated overnight at 4°C with the second primary antibody. This was developed the next day without nickel salts and resulted in a brown stain.

2.4.4 Immunohistochemistry using fluorescent antibodies

For fluorescent double stainings, secondary antibodies conjugated with fluorescent dyes were used (Vector laboratories). In all cases, primary antibodies were raised in different animals (see Table 2.2) and incubation of embryos with primary antibodies was done as described in section 2.4.2. Embryos were incubated in the dark with the secondary antibodies at room temperature for 2 hours and, after washing, they were mounted in 80% Glycerol. Embryos were viewed on a Leica TCS SP2 AOBS spectral confocal microscope.

2.4.5 Larval muscle scoring and analysis

Qualitative analysis of the effect on somatic musculature in the different experiments was performed by examination of the muscle pattern. Each of the 30 muscle was analysed systematically following specific criteria: presence/absence; shape. Based on the severity of the phenotype the effects were classified as: 1) wild-type - if no defects were observed; 2) weak – if the muscle pattern is wildtype and only a reduced number of muscles are misshapen; 3) moderate - most muscles are present and correctly shaped; 4) Severe- the muscle pattern is significantly affected, many muscles are misshapen or missing but a slight outline of the expected muscle pattern is still observable; 5) Extreme - no muscle is formed correctly. The wild-type muscle pattern was used as a reference for comparison.

2.4.6 Hatching and survival assay

Fly lines carrying a FM7 balancer with an ActGFP marker were generated for the potential HDAC4 null mutants (HDAC4 del6 and HDAC4 del48). Eggs were collected on apple juice-agar plates for 2h at 25°C and the obtained embryos were allowed to develop further for another 7h at 25°C. Embryos were observed under a fluorescence microscope and the embryos without GFP signal were selected. These embryos were homozygous for the generated deletion. 100 GFP negative embryos were aligned on a new agar plate and incubated for another 24h at 25°C. At this stage any first instar larvae were collected and transferred to a new apple agar juice-plate. The counting for second instar and third instar larvae was performed after 48h and 72 h at 25°C. Third instar larvae were transferred into tubes and allowed to reach pupal stage. As a control, 100 OR embryos were aligned on an apple juice agar plate and the hatched number of larvae was counted as previously described.

2.4.7 Preparation of Digoxigenin-labelled RNA probe

The DNA template for the *HDAC4* RNA probe was a 5528bp cDNA derived from clone RE18386 ordered from the BDGP collection. Prior to transcription, DNA templates were linearised using a NotI enzyme, the reaction digest being incubated for 2h at 37°C. The protocol for synthesising DIG-labelled RNA probes from linearised DNA was adapted in the Taylor lab from protocols described by others

(Casey and Davidson, 1977; Urrutia et al., 1993). Transcription reactions were carried out in 50 μ l tubes containing more than 2 μ g of linearised template DNA, 1 μ l of 10X transcription buffer (Roche), 1 μ l of DIGNTP mix (Roche), 0.5 μ l of RNase inhibitor (Roche) and 1 μ l of T3 RNA polymerase (Roche) for the antisense probe, T7 RNA polymerase (Roche) for the sense probe. The final volume of each probe synthesis reaction was then adjusted to 10 μ l with H₂O. Reactions were incubated at 37°C for 2 hours. After transcription, the DNA template was removed by the addition of 1 μ l of DNaseI buffer, 6 μ l of H₂O and 3 μ l of DNaseI RNase free (10U/ μ l; Roche) and the mix was incubated at 37°C for 15 minutes. The RNA probes were fragmented by adding 80 μ l of 125 mM sodium carbonate (pH 10.2), and incubating the probe at 60°C for 15 minutes in the hybridisation oven. The alkaline hydrolysis of the probes was stopped by adding a volume of 50 μ l 7.5M ammonium acetate and storing on ice. For RNA probe precipitation, the reaction was further mixed on ice, with 375 μ l 100% ethanol for 10 min and centrifuged at maximum speed for 15 min. The remaining pellet was air dried, and resuspended in 30 μ l TE:formamide (1:1). A spot test was performed in order to estimate the yield of the DIG-labelled RNA. A DIG-labelled control (Boehringer) of known concentration was used to determine the yield of the probe. Dilutions of both the control and the probes (all in RNase-free H₂O) were spotted and UV cross-linked in a UV Stratalinker to a positively-charged nylon membrane (Roche). The membrane was washed twice for 5 min in Blocking Solution (10 % blocking reagent from Roche dissolved in maleic acid buffer (100 mM maleic acid, 150 mM NaCl, pH 7.5)), followed by a 30 min blocking step. The membrane was incubated with an anti-DIG-AP Fab fragment antibody diluted 1/2000 in blocking solution for 30 min at room temperature. There were two washing steps of 5 min in blocking solution, followed by two washes of 10 min in Washing Solution (100 mM Tris pH 9.5, 100 mM NaCl, 5 mM MgCl₂, 0.1% Tween). The spots were developed in a solution containing 4.5 μ l NBT and 3.5 μ l BCIP per 1 ml of developing/washing solution, incubating in the dark for an appropriate time and the reaction was stopped by washing thoroughly with PBT. The yield was determined by comparing the intensities of the synthesized RNA probes with the control. Probes were diluted with TE:formamide (1:1) to a final volume of 25 ng/ μ l after the quantification procedure.

2.4.8 RNA *In Situ* Hybridisation

In situ hybridisations were carried out according to standard protocol (Taylor, 2000) with the use of Digoxigenin-labelled RNA probes. Fixed embryos were stored at -20°C in methanol and for RNA *in situ* hybridisation they need to be rehydrated. The rehydration process involves sequential replacement of methanol with 4% paraformaldehyde (PFA): 1) incubation of embryos for 2 min at room temperature in a mixture of 3:2 methanol to 4% PFA, 2) incubation of embryos for 5 min at room temperature in a mixture of 1:3 methanol to 4% PFA, 3) incubation of embryos in 1 ml of 4% PFA for 10 min with gentle shaking. The rehydrated embryos were rinsed 3 times and washed 3 times for 5 min in 1X PBS containing 0.1% tween-20 (PBT). Prehybridisation of the embryos at 55°C in a humidified chamber/hybridisation oven for 1 hour in hybridisation buffer (50% formamide (Fluka), 4X SSC (Sigma), 1X Denhardt's solution (Sigma), 250µg/ml yeast tRNA (Invitrogen), 250µg/ml salmon testis DNA (Sigma), 50µg/ml heparin (Sigma), 0.1% Tween-20) was necessary. The actual hybridisation of the embryos was done overnight at 55°C with 5 µl of 25 ng/µl DIG-labelled RNA probe in 0.5 ml hybridisation buffer, both the probe diluted in buffer and embryos being incubated at 55°C before mixing. The following day, embryos were washed 4 times in a washing solution at 55°C (50% formamide, 2X SSC, 0.1% tween-20). The first 3 washes lasted several hours during the day, the last wash was overnight. Prior to antibody detection, the embryos were rinsed once in PBT and washed in 1 ml PBT for 30 min under gentle shaking. For detection an anti-DIG-AP FAB fragments antibody (Roche) diluted in PBT + 5% normal goat serum was used. Incubation of the embryos and antibody was done for 90 min at room temperature with gentle shaking. The embryos were then washed 4 times in PBT, 20 min per wash. In order to prepare the embryos for staining development the embryos were rinsed twice and washed once for 5 min in AP buffer (100 mM Tris pH 9.5, 100 mM NaCl, 50 mM MgCl₂, 0.1% Tween). The colouring was developed in the dark in presence of 2.7 µl NBT (stock concentration of 9 µl/ml of 4-NitroBlue Tetrazolium Chloride Solution from Roche) and 2.1 µl BCIP (stock concentration 7 µl/ml of 5-Bromo-4-chloro-3-indolyl-phosphate, 4-toluidine salt from Roche) diluted in 0.3 ml AP buffer. Embryos were washed thoroughly in PBT to stop the colouring reaction does not develop further than necessary and mounted in 80% glycerol. A Zeiss Axioskop microscope was used for observation of the stained embryos.

2.4.9 The GAL4-UAS Expression System

The GAL4/UAS system was used to mis-express *Drosophila* HDAC4 and human HDAC5 in specific cells or tissues within the embryo. The system has two components: the GAL4 driver stock that expresses the yeast transcriptional activator protein under a tissue/cell specific promoter/enhancer, and the UAS stock that carries a transgene whose expression will be modulated by the GAL4 upstream activating sequence (UAS). In GAL4 and UAS crosses, the transgene is expressed in the same pattern and tissue as the GAL4 protein. The promoter that regulates GAL4 expression can induce the ectopic expression of a transgene (Sonnenfeld, 2009).

In the experiment used to observe the effects of overexpressing HDAC4 in muscle, females homozygous for the *twi-Gal4*; *twi-Gal4* drivers were crossed with males homozygous for each UAS construct (see tables 2.1 and 2.4 for UAS constructs used). Crosses were set up in collection cages where flies were allowed to lay eggs on apple juice-agar plates at 25°C for 18h. These 0-18h collections contained embryos of stages 1-17. Stages 12 and 16 were identified by morphology and used to observe the effects of overexpressing different Class IIa HDAC constructs in muscle.

Table 2.4: UAS-Constructs used to overexpress HDAC

UAS construct	Role
UAS-dHDAC4	Construct allows overexpression of full length <i>Drosophila</i> HDAC4
UAS-dHDAC4ΔC	Construct allows overexpression of constitutively active <i>Drosophila</i> HDAC4, due to deletion of C-terminus including the HDAC domain
UAS-dHSAC4SA	Construct contains a <i>Drosophila</i> HDAC4 variant in which all the serine residues have been mutated to alanine. These Ser residues are important in the process involving shuttling of class IIa HDACs between cytoplasm and nucleus
UAS-hHDAC5	Construct allows overexpression of full length <i>human</i> HDAC5
UAS- hHDAC5ΔC	Construct allows overexpression of constitutively active <i>human</i> HDAC5, due to the deletion of the C-terminus which includes the HDAC domain and NES
UAS- hHDAC5SA	Construct contains a <i>human</i> HDAC5 variant in which all the serine residues have been mutated to alanine. These Ser residues are important in the process involving shuttling of class IIa HDACs between cytoplasm and nucleus

Chapter 3: Tandem affinity purification of Mef2 interacting proteins

3.1 Introduction

Cell differentiation represents the core biological process to produce specialised tissues during development, and are implemented by molecular switches like transcription factors (TFs) that are able to activate batteries of genes. Uncovering the molecular mechanisms that lead certain central elements to coordinate the activation of specific genes via cis-promoters is considered essential for understanding how organisms develop (Britten and Davidson, 1969; Davidson and Britten, 1974; García-Bellido, 1975; Newman, 2020). Myogenesis represents a paradigm for the cell differentiation program and its gene expression program is regulated by key transcription factors like MRFs and the family of Mef2 genes in vertebrates and Twist and Mef2 in *Drosophila* (see Introduction). The Mef2 transcription factor family plays a pivotal role in muscle development both in *Drosophila* and vertebrate musculature where it partakes in TF cascades responsible to activate muscle specific genes (see Introduction sections 1.7 and 1.8 for specific examples) (Black and Olson, 1998; Potthoff and Olson, 2007; Taylor, 1995). *Drosophila* has had a central role in uncovering the function of Mef2 in muscle development together with studies in mammalian cell culture (Taylor and Hughes, 2017). In mammals there are four closely related Mef2 genes (Mef2a, -b, -c, -d), while *Drosophila* has a single gene. In *Drosophila*, Mef2 null mutants are not able to form differentiated muscle - although specification occurs the progenitors fail to fuse (Bour et al., 1995; Lilly et al., 1995; Ranganayakulu et al., 1995) - while a loss of function study in vertebrates showed that MRFs require Mef2 to convert fibroblasts to myoblasts and C2C12 myoblasts fail to form myotubes in absence of Mef2 (Ornatsky et al., 1997; Snyder et al., 2013). Therefore, to understand muscle formation in *Drosophila* it is crucial to understand the function of Mef2 and of muscle specific proteins that interact with it during myogenesis.

Genomic chromatin immunoprecipitation (ChIP) studies in *Drosophila* embryos have shown the broad role Mef2 plays in the orchestration of muscle differentiation (Junion et al., 2005; Sandmann et al., 2006). Hundreds of genes expressed at different time points during the differentiation program were regulated by Mef2 and showed differential transcriptional responses to different levels of Mef2 and to co-activation by Mef2 and Lmd or Mef2 and Twi (Bernard et al., 2009; Cunha et al.,

2010; Sandmann et al., 2006, 2007). In mammalian cell culture it was shown that the four Mef2 isoforms can activate separate gene groups and that very few genes are responsive to the activation of all the four Mef2 proteins (Estrella et al., 2015).

Both in *Drosophila* and in mammals the activity of Mef2 is at the centre of a variety of biological contexts, where it coordinates specific gene programs based on signalling cues transduced to the nucleus by different kinases, or other activators and repressors. A selection of these partners have been found in muscle cells, while others interact with Mef2 in the fat body and the neurons. In *Drosophila*, Mef2 is present both in embryonic and adult muscle progenitor cells, hours before many of its target genes are expressed (Ranganayakulu et al. 1995; Taylor, 1995) and the activation of its target genes have been shown to require the presence of other TFs like Twist, Lmd, CF2, Sd and Vg for their activation. Some of these activation partners have been shown only to bind the same CRM to activate target genes, while other like Twist, Sd and Vg physically bind to Mef2 as proven by studies performed in S2 cells (Arredondo et al., 2017; Bernard et al., 2009; Cunha et al., 2010; Sandmann et al., 2007). Even in the same cell type, the Mef2 target genes have a range of expression profiles during the differentiation program (Elgar et al., 2008; Junion et al., 2005; Sandmann et al., 2006; Cunha et al., 2010) and Mef2 can interact differentially with its interaction partners in different muscle types (e.g. cardiac, visceral, somatic) which leads to varied transcriptional patterns (Busser et al., 2012; Deng et al., 2009). Mef2 is involved in adult neurons to regulate circadian rhythmic behaviour where its expression is regulated by the Clock and Cycle circadian transcription complex (Blanchard et al., 2010; Sivachenko et al., 2013). In the fat body Mef2 acts as a molecular switch for anabolic function and immune response depending on its phosphorylation state and the ability to bind TBP (Clark et al., 2013). Moreover, the Mef2 family displays diverse functions in many mammalian cell types: muscle, nerve, vasculature, T-cells, chondrocytes (Potthoff and Olson, 2007), with different Mef2 isoforms showing overlapping but also converging functions. Many of the proteins that directly interact with Mef2c in various biological contexts have been recently reviewed and they total 30 direct binding partners (Dong et al., 2017). Most of these proteins were identified through small scale studies where the interaction of each protein with Mef2 was assessed. Two large scale studies

performed in S2 cells provide lists of *Drosophila* Mef2 candidate direct interaction partners.

Considering how complex the activity of Mef2 is in *Drosophila* and the diverse transcriptional responses it can generate, it is clear there are protein complexes that modulate the function of Mef2 in time and space in order to prevent early activation of genes in early muscle development, that regulate its activity during the differentiation program, as well as allow its activity in different cellular contexts required to regulate gene expression. Some of these direct protein interactions have been documented, however these are probably part of larger protein complexes that interact with Mef2 and are able to regulate its activity during muscle development (Guruharsha et al., 2011; Rhee et al., 2014). Considering the particular characteristics of the Mef2 behaviour in muscle cells, it becomes apparent that most of its interactions happen sporadically and in a context specific manner. Sometimes the interaction with the same protein can lead to opposing functions, either activating or repressing, most probably brought into effect by interaction with other proteins that can induce post translational modifications either to the two proteins themselves or to the chromatin surrounding the enhancers (Bernard et al., 2009; Busser et al., 2012).

It was the aim of this study to identify Mef2-interacting proteins in muscle differentiation in embryos, preferably as part as distinct protein complexes that are able to effect a particular function. Identifying and understanding the role of these proteins is of general importance since they represent an important part of the picture which describes the role of Mef2 in cell differentiation of muscle cells. Some protein interactions with Mef2 have been described, with little information about the role in development and the S2 large scale extractions cannot offer the resolution for a comprehensive screen of Mef2-interacting protein during development.

3.1.1 Experimental approach

Advances in complex protein purification approaches and subsequent identification of candidates by mass spectrometry have allowed systematic identification of protein complexes and networks (Köcher and Superti-Furga, 2007). Tandem affinity purification (TAP) represents the method of choice for proteome analysis when the subject of interest are long-range interactions, i.e. protein complexes consisting of

more than a bait (protein studied) and its direct interactor proteins, particularly when little prior information on expected interactors is available. The TAP method implies the fusion of the TAP tag with the targeted protein and its insertion into the host cell or organism. Cell extracts are prepared and the bait protein (TAP-tagged target protein) and its interaction partners are extracted via two subsequent steps of purification/elution. Compared to single-purification protocols, this usually reduces the amount of contaminating proteins. The TAP tag consists of two IgG binding domains from *Staphylococcus aureus* Protein A, a TEV cleavage site and a calmodulin binding protein (CBP; Figure 2.2). This tag represents the canonical TAP tag and was initially developed in yeast (Rigaut et al., 1999). Both N-terminal and C-terminal TAP tags are available, where the domains are inverted such that Protein A is always located furthest from the target protein (Puig et al., 2001). This tag will be referred to as yTAP where “y” stands for yeast. The yTAP tag has been successfully used in yeast, bacteria, *Drosophila* and mammalian cell culture to purify protein complexes (Bouwmeester et al., 2004; Butland et al., 2005; Gavin et al., 2002; Veraksa et al., 2005).

In the purification process Protein A binds strongly to an IgG matrix and requires the use of TEV protease to elute the bait protein and its interaction partners under physiological conditions. The eluted material is further bound to Calmodulin-coated beads in the presence of Ca^{2+} , the following washing steps eliminating contaminants and remaining TEV protease. The bound material is eluted off the beads with EGTA under mild buffer conditions. This method allowed for purification of high-yield protein complexes under native conditions (Puig et al., 2001).

Different versions of the TAP tag have been developed to reduce the amount of necessary input material, reduce the percentage of contaminants and increase the overall yield. The more bait protein is recovered, the more associated partners are obtained. The new tags developed involved substitution of Protein A with Protein G and CBP with streptavidin binding peptide (SBP). Permutations of the mentioned components resulted in 4 different tags: a Protein A-CBP tag (the original yTAP), Protein G-CBP tag, Protein A-SBP tag and Protein G-SBP tag (GS-TAP). Of the four tags tested, the GS-TAP was shown to render the best results in mammalian cell culture. A 10-fold yield increase was observed for the GS-TAP tag compared to

yTAP (Bürckstümmer et al., 2006). The performance of the GS-TAP tag compared to yTAP was tested in the *Drosophila* system both in cell culture and embryos. The GS-TAP tagged protein expressed in *Drosophila* whole embryos gave a 3-fold higher elution yield compared to its yTAP tagged version. A lower load of contaminants was observed in the GS-TAP purification than in the yTAP purification. The use of additional TEV cleavage sites in the GS-TAP tag increased the cleavage time after the first purification step (Kyriakakis et al., 2008).

The advantages offered by the GS-TAP tag for the purification procedure are the following: less input material is required, a reduction in contaminants left in the final elute is observable and the overall yield is higher than for TAP purification with the yTAP tag. The Furlong lab developed a line of C-terminally tagged Mef2. The GS-TAP tag was knocked in frame with the Mef2 gene after its tenth exon. After the tenth exon the tag continues at the C-terminus with the SBP, two TEV cleavage sites and two Protein G domains. The knock-in was obtained by homologous recombination and the obtained stock was homozygous viable. The localisation of the protein in the cell was not affected by the tag.

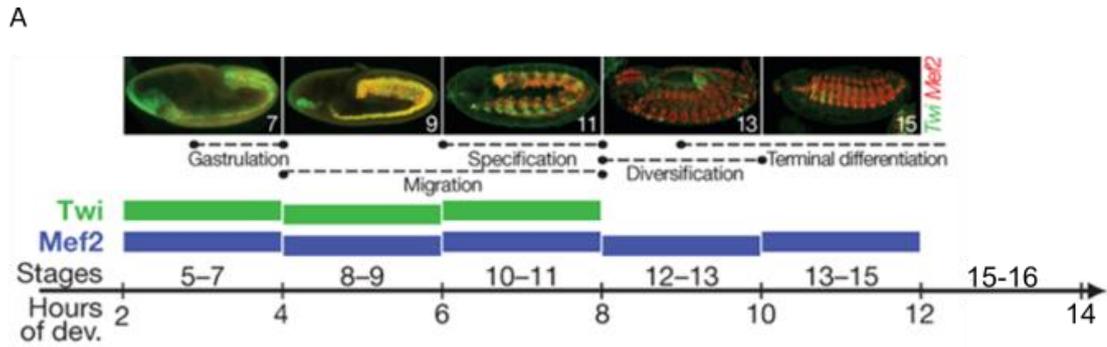
The modifications to the TAP purification procedure generated by the use of the GS-TAP are the following: the first part of the purification step is the same (IgG beads are used to sequester the bait protein and the associated proteins from the lysate mixture, followed by the release of the bait by TEV protease cleavage), while the second step requires the use of streptavidin sepharose beads (instead of calmodulin beads) to bind the TEV-protease cleaved proteins and eluted with biotin (instead of EGTA).

3.2 Results

3.2.1 Embryo collections for TAP purification

Tandem affinity purification is a method to purify protein complexes in different model systems. Despite the wide range of applicability of this method in various organisms, the yield of the purified proteins is low. Therefore, a large quantity of biological input material is necessary to obtain a sufficient amount of eluted proteins to be detected by mass spectrometry (MS). In mammalian cells, the typical quantity of harvested cells for TAP purification amounted to 5×10^8 - 1×10^9 cells (Bürckstümmer et al., 2006). The range of quantities of input material used in the past to perform TAP purifications from *Drosophila melanogaster* embryos is very broad. The TAP purification of a nuclear protein used a nuclear extract from 30 to 40 grams of embryos (Klymenko et al., 2006). Another study performing TAP purifications in *Drosophila* embryos both for cytosolic and nuclear proteins reported a range of 3-6 grams of embryos for an experiment (Veraksa et al., 2005). A study comparing the efficiency of the use of the GS-TAP tag to the classical yTAP reported that 5 grams of embryos were necessary for one TAP purification (Kyriakakis et al., 2008).

Mef2, the target protein of the TAP purification, is known to be expressed very early in development and is present in the embryo until the end of embryogenesis. A combination of genetic and genomic studies have revealed that muscle genes regulated by Mef2 have a lower requirement of Mef2 activity if they are expressed early compared to the genes expressed at a later time point (Sandmann et al., 2006). A model in which Mef2 activity levels increase during muscle differentiation was proposed (Elgar et al., 2008), although without specifying whether "Mef2 activity" refers to levels of expression of the *mef2* gene, or the fraction of Mef2 protein that is activating transcription of target genes. If the activity levels in this sense correlate with the amount of protein expressed, it is expected that Mef2 protein expression is lower in early development compared to late development (Figure 3.1A). Applying this hypothesis to TAP purification, one would expect that a lower amount of starting material is required for extraction from late stage embryos compared to earlier stage embryos. Taking these into consideration, embryos of 4 staging windows were collected of the Mef2-GSTAP line such that enough input material for 2-3 large scale



B

Staging	Mef2TAP embryos collected (g)	WT embryos collected (g)
4-6h	19.4	12.8
6-8h	19.0	15.7
10-12h	10.8	8.8
11-13h	6.8	8.2
O/N (0-18h)	71.4	54.3

Figure 3.1: Collected embryos of Mef2-GSTAP line for TAP purifications that require large amounts of starting biological material

A) Developmental timeline of *Drosophila* embryos at 25°C (Campos-Ortega and Hartenstein, 1997), highlighting expression of Mef2 and the mesodermal determinant Twist, which induces Mef2 expression. Collection of staged embryos was performed in 2 h windows that allowed to identify the most representative stages for Mef2 activity during embryogenesis. The picture was adapted from (Zinzen et al., 2009)

B) Total amounts of staged Mef2-GSTAP and wildtype (WT) collected embryos to cover large scale experiments of TAP purifications of the same stage. Wildtype embryos were collected as controls to identify non-specifically binding proteins.

TAP purifications was available (Figure 3.1B), although due to time constraints only the 11-13h staging window was analysed as a representative of muscle development.

3.2.2 Detection of the tagged Mef2 bait

To identify systematically the proteins that associate with Mef2 during *Drosophila* embryogenesis, the Mef2-GSTAP *Drosophila* line developed in the Furlong laboratory at EMBL was used. The transgenic line contained a knock-in of a GSTAP at the C-terminus of Mef2 (Figure 2.2), in frame with its 10th exon (Figure 2.1). The tag contained two Protein G domains, two TEV-protease cleavage sites, and one streptavidin-binding peptide (SBP) moiety, placing the Protein G domains at the extreme of the C-terminus. Immunohistochemistry against the tagged Mef2 protein revealed that its expression recapitulated the expression pattern of the wildtype Mef2 protein during embryogenesis (data not shown). The first step in adapting the TAP purification protocol for the tagged Mef2 protein (Figure 2.4) was to identify a method that allows tracking the extraction process at each step. For proteins, Western blot was the method of choice to assess the success of the optimisation steps, both in preparing embryonic protein extracts and during the TAP purification experiments. Both small and large scale experiments were performed.

Western blot required an antibody that can specifically recognise Mef2 and distinguish the bands for tagged and untagged Mef2. To obtain this, four different bleeds of anti-Mef2 antibodies were tested (Figure 3.2). The antibodies were prepared in-house by the Furlong lab and previously used in ChIP and immunostaining. All antibodies were tested on the following biological materials: embryos from the Mef2-GSTAP transgenic line, wildtype *Drosophila* embryos (Oregon R), HeLa cells and S2 cells. Each sample was homogenised on ice with a Wheaton douncer in the trial lysis buffer 1 (TL1): 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM KCl, 5% glycerol, 0.2% NP-40, 1.5 mM MgCl₂, 0.2% IGEPAL, 1 mM DTT, 0.001% DNase and protease inhibitors (Pefabloc/AEBSF 2 mM, E-64 40 uM, Aprotinin 0.5 mM, Leupeptin 1 uM, Pepstatin A 60 uM, Bestatin 10 uM). All four antibodies produced a band close to 100 kDa in lysates of embryos collected overnight (O/N, 0-18h) from the Mef2-GSTAP line. A band close to 70 kDa was dominant in all other samples. This suggests the 70kDa band corresponds to untagged Mef2 and the 100kDa band to tagged Mef2. The DRC4 antibody (Figure

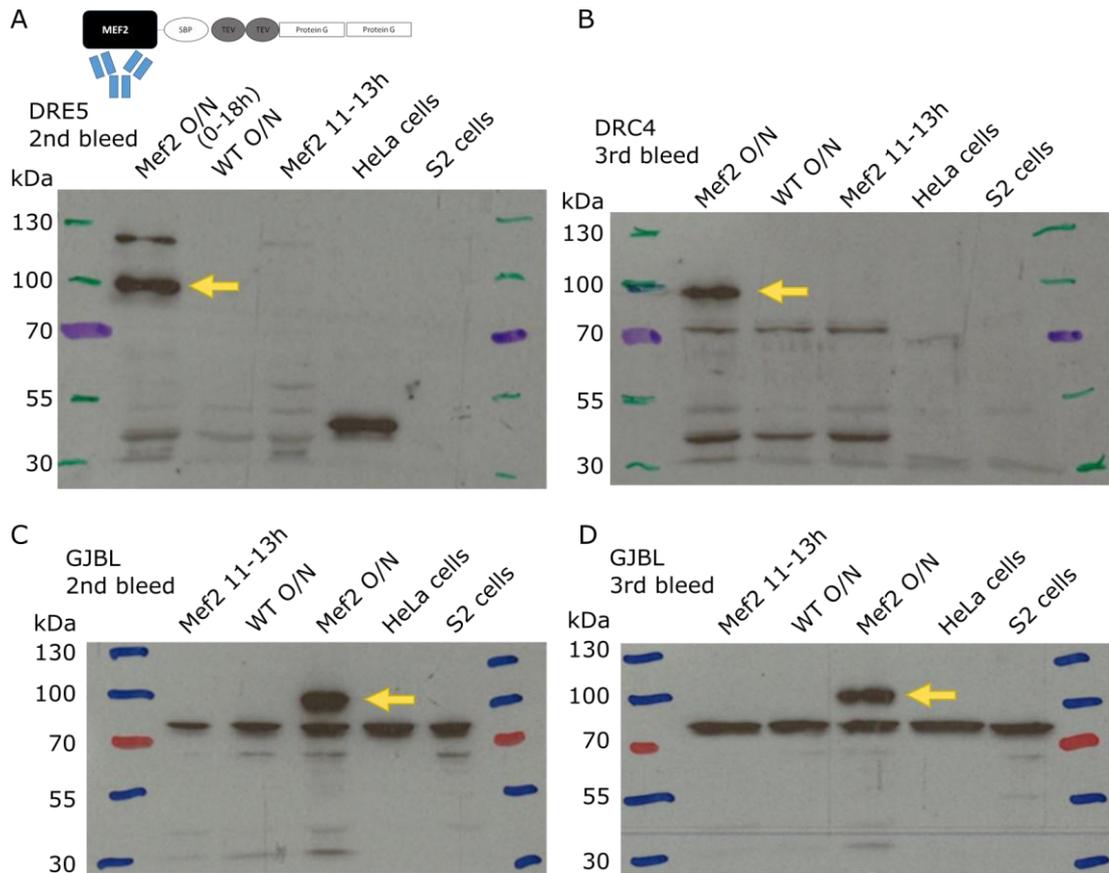


Figure 3.2: Identification of tagged Mef2 in the lysate of Mef2-GSTAP embryos

Lysates from different biological samples were loaded onto a blot and incubated with four different anti-Mef2 antibodies raised against *Drosophila* Mef2. (A-D) show a blot of one antibody, each tested against the same set of lysates. The antibodies were grown in house at EMBL and the names were coded internally. Wildtype embryos were used as a control to distinguish the tagged from the untagged Mef2. HeLa cells extracts were included as a control for cross reactivity with human Mef2, while S2 cells extracts were included as a potential negative control as no endogenous Mef2 protein was reported in S2 cells so far. The yellow arrow indicates the tagged Mef2 band.

3.2B) was able to recognise both variants of Mef2 with relatively low background signal, and was chosen for future use. "Anti-Mef2" in below sections refers to this antibody.

3.2.3 Tracking Mef2 during TAP purification as influenced by the amount of solubilised bait

It is important to be able to follow the bait protein during the purification process as problems during intermediate steps could affect the outcome of the procedure and affect the overall yield. The less protein is obtained from the TAP purification process, the less associated proteins the MS can identify. To track the Mef2 protein during the TAP purification process there were three possibilities: an anti-Protein G (anti-TAP) antibody, an anti-SBP antibody or an anti-Mef2 antibody. The anti-TAP antibody would only be useful to follow the bait before the TEV protease cleavage step, while an anti-Mef2 and anti-SBP antibody would be appropriate to observe the entire purification process. To test this hypothesis a lysate obtained from 500 mg of Mef2-GSTAP O/N embryos was split in half and one of the halves was incubated with TEV protease for 1h at 16°C. The lysates derived from Wildtype *Drosophila* embryos (WT) and HeLa cells were used as positive controls for the untagged Mef2 and as negative controls for antibodies targeting the GSTAP tag. The anti-SBP and anti-TAP antibodies detected only a 96 kDa band in the uncleaved Mef2-GSTAP O/N embryos (Figure 3.3A & B, second lane). When the lysate from Mef2-GSTAP O/N embryos was incubated with the TEV-protease, the 96 kDa band disappeared and a shorter 19 kDa band appeared in the anti-TAP blot, consistent with the now cleaved tag, which is much smaller in size (Figure 3.3B, third lane). Similarly, in the anti-Mef2 blot, the upper band disappeared and the 77 kDa band became stronger, suggesting that Mef2 with the leftover SBP tag has a similar size as untagged Mef2 (Figure 3.3C, compare second and third lane). The anti-SBP antibody was not able to detect the cleaved Mef2 (Mef2-SBP; Figure 3.3A, third lane), which made it inappropriate as an antibody for observation of the purification process. The anti-TAP and anti-Mef2 antibodies were also tested on lysates from staged Mef2-GSTAP 11-13h embryos, with essentially the same results (Figure 3.3D & E).

Having identified a suitable antibody, the lysis method was the first step to optimise. Several lysis methods were tested for their ability to extract and solubilise the Mef2-

GSTAP protein (Figure 3.4). Since just homogenising the embryos and clarifying the lysate proved to be quite inefficient, two new lysis methods were compared: longer incubation of the homogenate with the lysis buffer and a hypotonic extraction. For the standard lysis the embryos were resuspended in TL1, the tissues mechanically sheared with the loose pestle and the grinding process was finished off with the tight pestle. The loose pestle should break apart tissues while leaving most of the cells and nuclei intact, while the tight pestle should help the shearing of all the cellular components. The lysis buffer was designed to create an isolating and stable environment for the proteins of interest. This is of vital importance to the TAP purification procedure, because long range interactions between proteins are only maintained under close to physiological conditions. The modified procedure with extended incubation took advantage of additives contained in TL1, such as detergents, which should allow membranes to be disrupted more easily. After dissociating tissues using the loose pestle, the lysate was therefore incubated on ice for 30 min to partially lyse cells in this modified procedure. Subsequently, the lysate was ground with the tight pestle and clarified to solubilise the proteins. For the hypotonic extraction, a TL1 lysis buffer without NaCl was used during the two steps of mechanical shearing in pestles. After this, the salt was added and the lysate was incubated on ice for 30 min to allow the osmotic pressure created by the addition of salt to disrupt any unbroken cellular components.

To quantify the amounts of Mef2 extracted by each lysis procedure, the total protein concentration of each lysate was determined using a Bradford assay and equal amounts were loaded onto a blot (see Methods). Additionally, a standard of known amount of GFP was loaded in serial dilutions (Figure 3.4A) and quantified by densitometry (Figure 3.4B). It was shown that regardless of the lysis method, the amounts of Mef2 that can be extracted from embryos varies by stage (Figure 3.4A & C). The comparison of the supernatant to the pellet has shown that the standard lysis procedure had the highest recovery rate of solubilised bait (Figure 3.4D & E). The hypotonic extraction was slightly more efficient than the lysis with extended incubation, though overall the initial lysis method was the most efficient and the least time consuming.

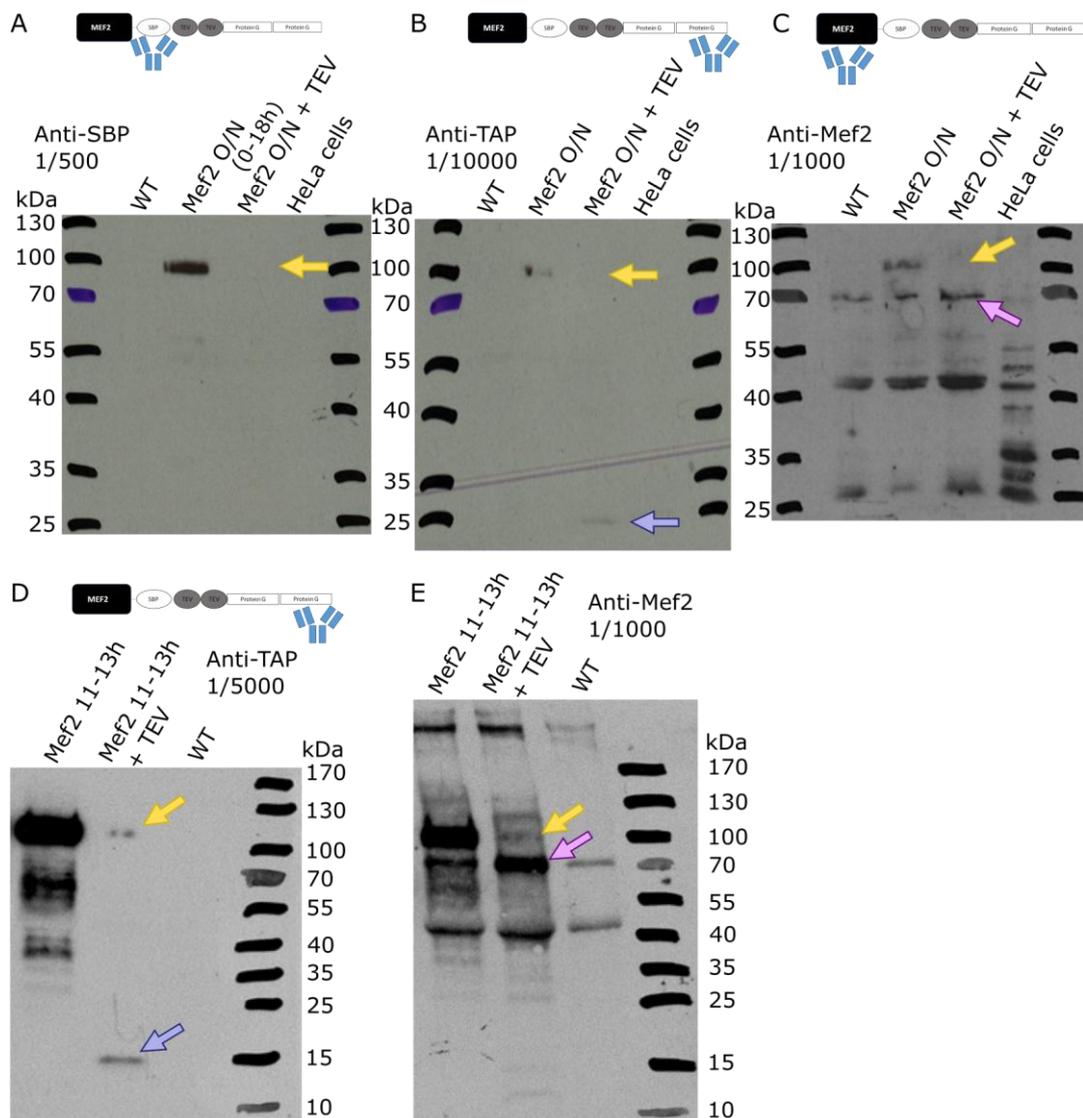


Figure 3.3: Options for observing the Mef2-GSTAP bait during TAP purification

Western blots testing three methods to detect Mef2-GSTAP during the purification. The antibodies were tested on overnight (O/N) collected Mef2-GSTAP embryo lysates that were incubated with or without protein TEV protease. The lysates of HeLa cells and WT embryos were used as negative control. Yellow arrow indicates the tagged uncleaved bait, pink arrow indicates the Mef2 protein after tag cleavage, blue indicates the cleaved Protein G domains.

A) Anti-SBP antibody that should detect the SBP in the tag before and after cleavage by TEV protease

B) Anti-TAP/ Protein G antibody that can detect the tagged bait when uncleaved or the cleaved Protein G domains of the bait

C) Anti-Mef2 antibody that can detect both the cleaved and uncleaved bait.

D&E) Extracts of 11-13 h staged Mef2-GSTAP embryos were used to confirm results of the anti-TAP (D) and anti-Mef2 (E) antibodies.

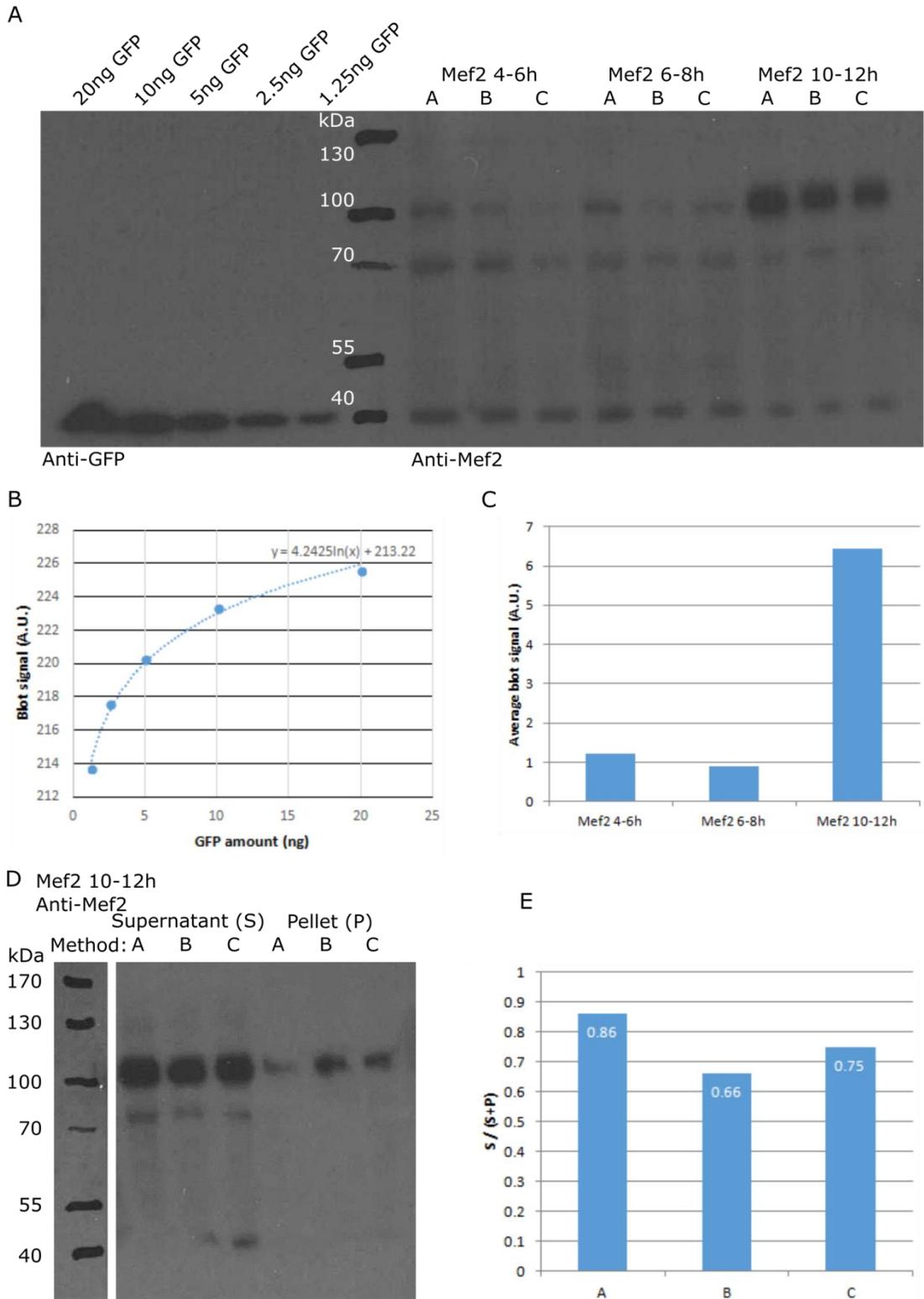


Figure 3.4: Testing different lysis methods to optimise the recovery of tagged Mef2 protein for use in TAP purifications

The three different lysis methods were liquid homogenisation (labelled A on the blots), liquid homogenisation with extended incubation (labelled B on the blots), and

hypotonic shock extraction (labelled C on the blots). Each lysis methods was applied to embryos staged 4-6h, 6-8h and 10-12h. Bradford assays were used to determine total protein concentrations of each sample such that equal amounts could be loaded.

A) To determine the amounts of clarified Mef2-tagged protein contained in each lysate, a western blot detected by the anti-Mef2 antibody was prepared in correlation with a serial dilution of a purified GFP of known concentration.

B) Densitometry-based analysis to determine amounts of Mef2 extracted in the different homogenisation methods. A correlation between amount loaded and signal detected was determined based on GFP quantification and a similar correlation was used to determine the Mef2 bait extracted at different stages.

C) Efficiency of the methods was determined by comparing the signal in the supernatant containing the solubilised protein that can be used in TAP, and the pellet that contains unextracted bait.

D) Densitometric quantification of the relative Mef2 content of supernatant and pellet in each lysate.

3.2.4 TAP purification of Mef2-GSTAP and associated complexes from 2 g of O/N embryos

The first large scale TAP experiment was realised starting from 2.366 g of O/N Mef2-GSTAP embryos using a procedure adapted from a TAP purification protocol used in the Gavin lab to purify yTAP-tagged proteins from Mycoplasma cells. Since the second part of the purification using the GSTAP tag required the use of Streptavidin beads rather than Calmodulin beads, this part of the purification protocol was adapted from the method described by Kyriakakis et al. (2008). In this study, the streptavidin-coated beads were boiled in SDS-loading buffer to release the bound purified complexes from the beads. The Bürckstümmer et al. (2006) study, which reported on the different TAP tag variants and their efficiency on purifying complexes from mammalian cells, used a 1 mM biotin solution to elute proteins from streptavidin beads instead. According to personal communication, others performing the biotin elution method with other tags have used a variety of biotin concentrations ranging from 2 mM to 10 mM. To perform the elution of Mef2-GSTAP and associated proteins, the streptavidin beads were therefore incubated in 150 µl of 2 mM biotin solution (1.5 times beads volume) for 5 min at 4°C and eluate was recovered by gravitational flow. A control experiment of the same TAP protocol was performed on wildtype embryos.

To observe the evolution of the TAP purification process, analytical fractions were collected throughout the procedure (Figure 3.5A). The fractions were analysed by Western blot and densitometry to determine the efficiency of each intermediate step and identify bottlenecks. The aliquots were taken at the following step: input extracts before clarifying the sample (Input); supernatant containing solubilised protein (cytosolic fraction, CF); supernatant after incubating the CF with the IgG beads (IgG flow-through, FT); uncleaved bait protein left on the IgG beads after incubation with TEV protease (IgG beads); bait protein cleaved by TEV protease (TEV eluate); flow-through after incubating the streptavidin (SA) beads with the TEV eluate (Streptavidin FT); bait protein not eluted off the streptavidin beads (Streptavidin beads); final bait eluted with biotin that was used for MS analysis (Final eluate). Note that any protein detected in the IgG FT, IgG beads, SA FT and SA beads fractions represents protein lost in the purification procedure.

Densitometry was performed on the relevant bands obtained on the Western blot (Figure 3.5C). This refers to the bands for uncleaved Mef2 (100kDa) up to the TEV cleavage step, and the lower band for cleaved Mef2 (70kDa) after the cleavage. Both bands were quantified in the IgG beads fraction to assess the performance of TEV cleavage itself. The obtained densitometry values were corrected by the specific portion of each purification step loaded on the gel. 93% of the bait was solubilised (Figure 3.5B, CF vs. Input, normalised by amounts loaded). This solubilised bait is considered 100% of bait to be purified (Figure 3.5C), of which 27% bound to IgG beads as suggested by the 73% found in the IgG FT. After washing the beads and adding TEV protease, 3.3% of the initial bait remained in total. This estimate is the sum of 1.2% uncleaved bait (IgG beads, 100kDa band), 1.3% cleaved bait remaining bound to beads (IgG beads, 70kDa band) and 0.8% recovered bait (TEV eluate). This suggests that up to 23.7% of bait was lost in the washing steps before TEV cleavage. After the incubation of the TEV eluate with the streptavidin beads, washed with 6 ml cleavage buffer and eluted with biotin, only 0.27% was eluted in the final fraction, which amounts to 1/3 recovery of bait from the TEV eluate.

For the MS analysis both purified samples (100 µl eluate each from O/N Mef2-GSTAP or WT embryos) were concentrated with Amicon centrifugal filters with a 10k cut-off to a volume of 30 µl. The samples were run a few centimetres into a 4-15% Biorad gradient gel and the lane was cut out entirely and processed for analysis by the Proteomics Core Facility at EMBL. Running the samples so shortly into the gel allowed reducing the amount of gel needed to extract the peptides from the gel. Even though no bands were visible by Coomassie staining (Figure 3.5D), three proteins were identified by MS in the Mef2-GSTAP sample: Mef2, Hsc70-3 (a heat shock protein) and Actin 5C. The most abundant was Mef2 (the bait). In the WT control, one protein was identified, which was also among the proteins in the Mef2-GSTAP sample (Actin 5C). The high abundance of the bait in the sample shows that the purification procedure was successful. However, the amount of bait was not sufficient to permit detecting many associated proteins. The two proteins pulled down are often regarded as contaminants in other pull down studies (Veraksa et al., 2005), but have been shown to interact with Mef2 in other organisms (Sala et al., 2014; Sonnemann et al., 2006; Yang et al., 2009).

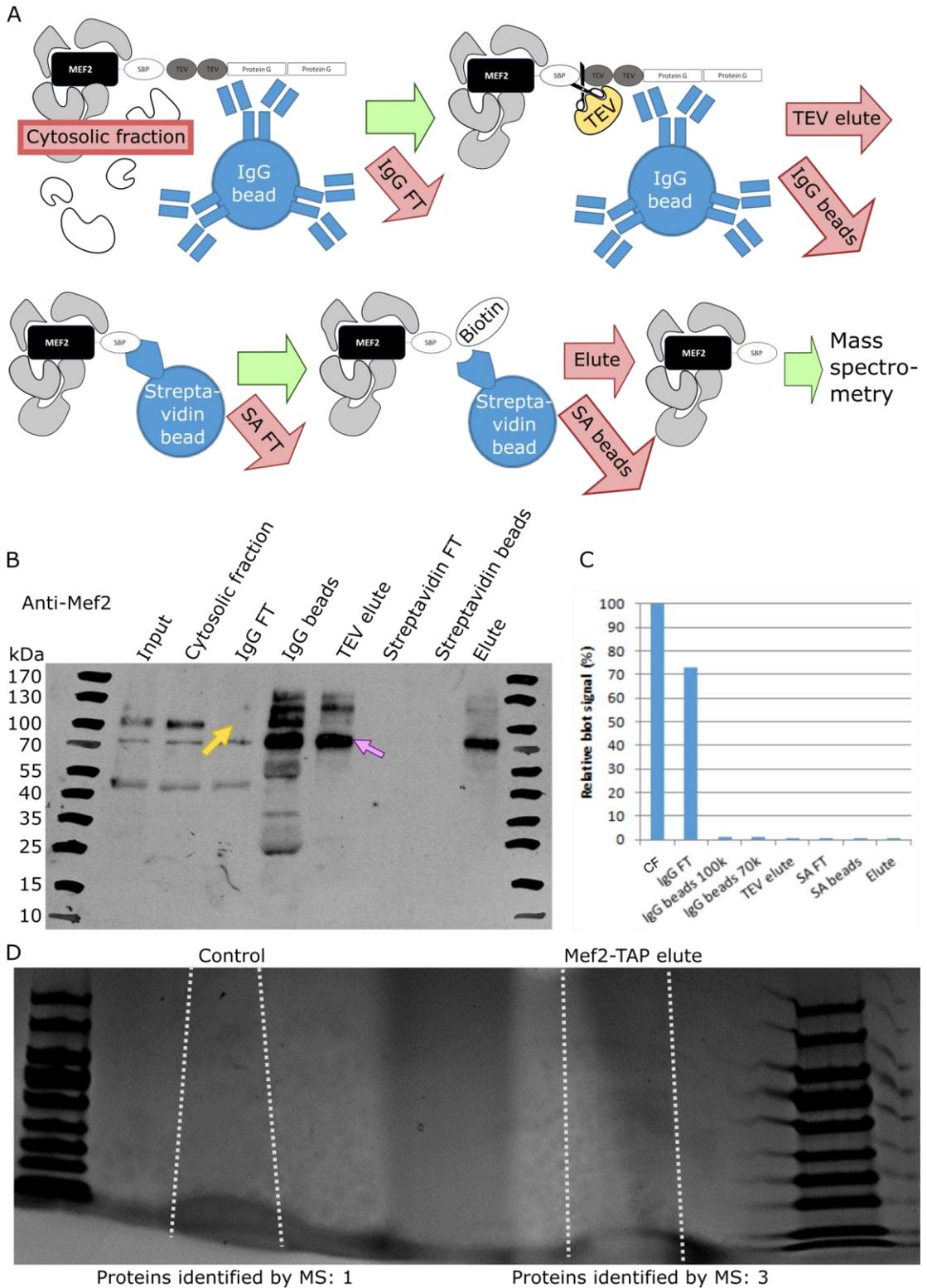


Figure 3.5: TAP pilot experiment with 2g of O/N embryos expressing Mef2-GSTAP

A) Illustration of the aliquots collected during the purification procedure to observe the loss of bait at each crucial step of the purification: Input, input extracts before clarifying the sample; clear cytosolic fraction (CF), solubilised proteins fraction; IgG

FT (flow through), flow through after incubating the CF with the IgG beads, any detected protein is lost protein; IgG beads, uncleaved bait protein left on the IgG beads after incubation with TEV protease; TEV elute, cleaved bait protein by TEV protease; Streptavidin F, flow through after incubating the streptavidin beads with the TEV elute; Streptavidin beads, bait protein not eluted off the streptavidin beads; Eluate, final bait eluted with biotin that was used for MS analysis.

B) Western blot of the aliquots collected during purification, stained against Mef2. Yellow arrow: the band corresponding to uncleaved Mef2 (100k) is not present in the supernatant after bead purification. Pink arrow: Band corresponding to cleaved Mef2 (70k).

C) Quantification of the Mef2 bait identified in each analytical fraction in B) based on densitometry after correction by the relative amount loaded onto the gel. CF, IgG FT and IgG beads 100k quantify the upper band of uncleaved Mef2 near 100kDa, all others quantify the lower band of cleaved Mef2 near 70kDa.

D) Coomassie stained gel of the proteins purified by TAP. The control sample was extracted from WT embryos. The lanes marked were cut out from this gel for analysis by mass spectrometry (MS). Even though no bands were visible, a small number of proteins were identified by MS. The results from this experiment are in Table S1 ("ONMEF 2g pilot" column), see data availability statement in section 2.3.13).

3.2.5 TAP purification of Mef2 complexes from 2 g of 11-13 h embryos

The second TAP experiment was performed on 11-13 h staged embryos, collected from the Mef2-GSTAP or WT lines. Based on the first experiment, a higher amount of bait was desired. However, a much larger amount of bait protein per gram of input embryos was expected compared to the O/N embryos based on the previous experiments for optimising the lysis procedure. A similar quantity of embryos was therefore used as input material to prepare the lysate (~2 g of embryos). In the first TAP experiment, a large amount of bait remained on the IgG beads after TEV cleavage, with only 0.8% of bait being recovered in the TEV eluate. The concentration of TEV protease was therefore increased from 0.05 $\mu\text{g}/\mu\text{l}$ as in the Gavin Protocol to 0.1 $\mu\text{g}/\mu\text{l}$ as in the Kyriakakis study in order to increase the removal of the bound bait from the IgG beads.

The analytical fractions were collected during the TAP purification procedure as in the first experiment and were analysed by western blot (Figure 3.6A) and densitometry. The recovery of bait after solubilisation was 79%, lower than 93% solubilisation in the first experiment. The fraction of bait bound to IgG beads was 20% and in the TEV eluate 0.9%, while 1% was cut but remained stuck on the IgG beads and 0.8% was uncut by the TEV protease (Figure 3.6B). The ratio of total cleaved to uncleaved bait was therefore changed from 1.7 in the previous purification from O/N embryos to 2.5 in the staged embryos. The final percentage of bait purified was 0.35%, a slightly better yield than the first experiment (0.27%). Considering that the absolute amount of protein was higher in the 11-13h embryos as input, a much larger amount of bait should be present in this final eluate.

The MS results suggested that indeed more bait was present, since more co-purified proteins were identified. While previously only 3 proteins were detected in the O/N embryos (bait and two candidate proteins), 56 proteins were identified in the 11-13h embryos (Mef2 bait and 55 candidate proteins). Only 5 proteins were detected in the 11-13 h WT control.

To compare in more detail the improvements brought by the use of more TEV protease to cleave the Mef2-GSTAP off the IgG beads, the analytical fractions of the O/N and 11-13 h experiments obtained after the incubation with the protease were

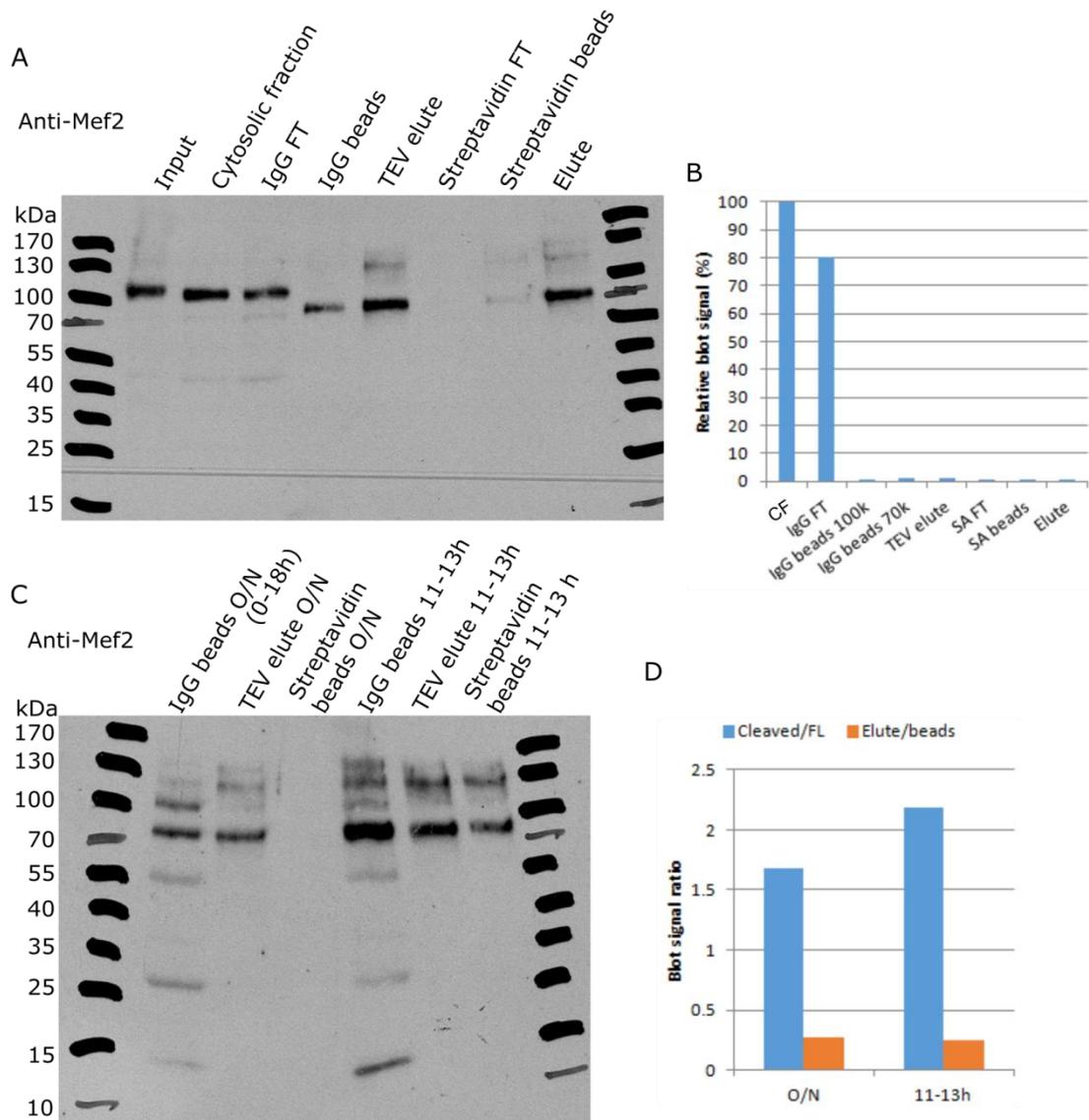


Figure 3.6: TAP pilot experiment with 2g of 11-13 embryos expressing Mef2-GSTAP

A) Aliquots were collected during the purification procedure and analysed by Western blot to observe the loss of bait at each crucial step of the purification: Input, input extracts before clarifying the sample; clear cytosolic fraction (CF), solubilised proteins fraction; IgG FT (flow through), flow through after incubating the CF with the IgG beads, any detected protein is lost protein; IgG beads, uncleaved bait protein left on the IgG beads after incubation with TEV protease; TEV elute, cleaved bait protein by TEV protease; Streptavidin F, flow through after incubating the streptavidin beads with the TEV elute; Streptavidin beads, bait protein not eluted off the streptavidin beads; Eluate, final bait eluted with biotin that was used for MS analysis.

B) Quantification of the Mef2 bait identified in each analytical fraction in (A) based on densitometry after correction by the relative amount loaded onto the gel. CF, IgG FT and IgG beads 100k quantify the upper band of uncleaved Mef2 near 100kDa, all others quantify the lower band of cleaved Mef2 near 70kDa.

C) Comparison of the efficiency of Mef2 cleavage off the IgG beads in the 2g O/N TAP experiment compared to 2 g of 11-13 h embryos.

D) Ratios between the full length tagged bait (100 kDa band) and the cleaved bait (70 kDa band) or between the eluted cleaved protein and the bait remaining on the IgG beads. The total amount of cleaved bait is the sum of the lower band in the IgG beads and TEV elute lanes, adjusted by amount loaded. The total amount on beads is the sum of the two bands in the respective IgG beads lane of (C).

loaded on a Western blot (Figure 3.6C) and analysed by densitometry. The comparison of the cleaved to the uncleaved Mef2 showed that the increased concentration of TEV protease was able to cut a larger proportion of the tagged bait protein, however the yield obtained in the eluate was not significantly larger (Figure 3.6D). Despite this, the amount of bait and associated proteins in the 11-13 h eluate was sufficient to cross the detection threshold for MS and return a more appropriate number of identified proteins. Given that the change in TEV protease concentration did not impact the fraction of recovered bait substantially, this improvement is likely owing to a higher absolute amount of bait in the 11-13 h embryos.

3.2.6 Improving bottleneck steps of TAP purifications

Based on these two experiments, the main bottlenecks impacting the amount of bait (and bound proteins) available for MS were identified as the following: 1) the amount of input bait; 2) the efficiency of the bait binding to the IgG beads (70-80% of bait lost); 3) the recovery of cut bait after TEV cleavage (75% of remaining bait lost). The binding efficiency of the bait to IgG beads could be increased by using a larger amount of beads to achieve saturation during the incubation with the lysate. To improve recovery of the cleaved bait, a different elution method was chosen. The elution of the cleaved protein from the IgG beads so far was done by gravitational flow, which did not prove the most efficient.

To test the impact of the amount of IgG beads on the recovery of bait in subsequent steps, a small scale experiment focussing on the first part of the TAP purification procedure was designed (Figure 3.7A). 2 g of O/N Mef2-GSTAP embryos were used as input material for the experiment and lysed using the standard procedure. 5 ml of cytosolic fraction were obtained and divided into four fractions of increasing volume: 500 μ l (1xCF), 1 ml (2xCF), 1.5 ml (3xCF), and 2 ml (4xCF). Each fraction was incubated with 20 μ l of IgG beads to assess the binding capacity of the IgG beads from an increasing amount of *Drosophila* embryo lysate. The beads were recovered, washed twice and incubated with TEV protease to cleave and release the bait. The cleaved bait was eluted twice, first by centrifugation at 400 g for 2 min and again by washing the beads in 100 μ l of cleavage buffer (TEVcb), followed by centrifugation. Analytical fractions of each step were analysed by Western blot as previously, though additional samples from washes were included (Figure 3.7B).

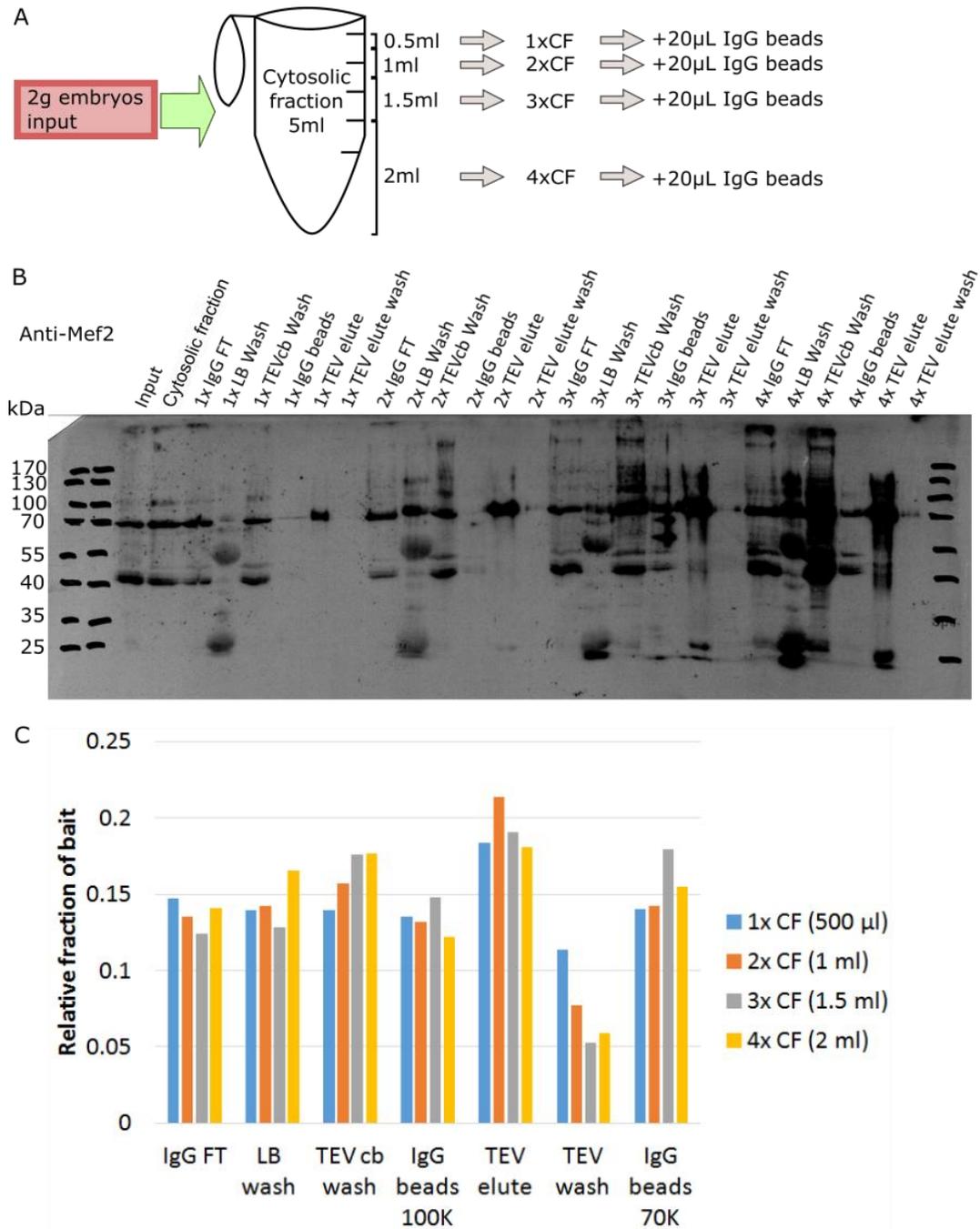


Figure 3.7: Optimisation of the first steps of TAP purification

A) 2 g of O/N embryos were lysed and split into four fractions of increasing volume, each to be incubated with the same amount of IgG beads. The aliquots therefore receive decreasing concentrations of beads.

B) Detection of bait in aliquots taken during incubation with the IgG beads and release of the bound bait by TEV protease cleavage. IgG FT, bait not adsorbed by IgG beads; LB wash, wash of bait-coated IgG beads with lysis buffer; TEVcb wash,

wash of bait-coated IgG beads with TEV protease cleavage buffer; IgG beads, left-over bait on the IgG beads after washes and TEV protease cleavage (obtained by boiling beads after elution; typically some bait stays uncleaved and some cleaved bait also remains); TEV elute, cut bait protein eluted off the beads by centrifugation; TEV elute wash, 100 μ l wash of post-cleavage IgG beads in TEV cleavage buffer prior to boiling, collected with centrifugation. An equal volume of corresponding analytical fractions was loaded, e.g. for TEV elute, 4 μ l was loaded from the 1xCF TEV elute, 4 μ l from 2xCF TEV elute etc. The cleaved Mef2 band is near 70kDa, the uncleaved Mef2 near 100kDa.

C) Quantification of (B), highlighting distribution of bait across the different fractions. For each sample (1xCF, 2xCF etc.), the total bait detected across the different analytical fractions was measured. Each bar represents the respective fraction of this total.

The distribution of bait across the different fractions was similar regardless of the ratio of lysate to beads (Figure 3.7C), indicating consistency of the extraction procedure during these steps. To assess the efficiency of bait binding to IgG beads, the densitometry values of bait detected in the IgG FT were compared to the total amount of bait detected in the input. If the binding capacity of the beads exceeded the amount of bait, the efficiency should be unaffected by the amount of lysate. For 1xCF and 2xCF, 91% and respectively 81% of the bait bound to the beads, while only 69% was bound in the 3xCF and 51% in the 4xCF samples, indicating that the binding capacity of the beads was saturated, potentially already in the 1xCF sample. The increasing amount of contaminants (blot signal outside of the 70kDa and 100kDa bands, Figure 3.7B) suggests that additionally to lower efficiency of binding the bait, the beads were increasingly binding other proteins unspecifically. After TEV cleavage, approximately 35% of the bait was successfully eluted off the beads in the first two more concentrated samples, and approximately 25% in 3xCF and 4xCF. Increasing the amount of IgG beads used for bait capture should therefore improve bait binding and yield.

3.2.7 TAP purification of Mef2 complexes from 7.7 g of O/N embryos

Taking into account the conclusions derived from the previous experiments, the GS-TAP purification protocol was modified to improve the steps that have an important impact of the overall yield of the purification process. This third TAP purification experiment used 7.73 g of O/N Mef2-GSTAP embryos (and in parallel 7.5g of WT embryos as a control sample) as input as the previous optimisation test suggested that a higher input amount is a simple way of obtaining more bait for elution and interaction capture. The following additional modifications to the previous protocol were made: 1) The quantity of IgG beads was increased from 200 µl to 400 µl in total to increase binding capacity, and the lysate was split into smaller fractions to improve the exposure of the lysate to the bead surface area during the incubation. All the beads were then collected into the same Mobitec column; 2) For washing the IgG beads after binding the bait, the ratio of lysis buffer to TEV cleavage buffer was increased from 2.5:1 to 4:1; 3) To increase recovery of bait after the TEV cleavage, the elution from the IgG beads was done by a short centrifugation at 400 g, followed by a wash (Maeda et al., 2014). The concentration of TEV protease was maintained

at 1 µg/ µl as used for the 11-13 h purification. All other steps were performed and analytical fractions were collected as described previously.

The analytical fractions in this experiment produced an untidy blot with strong signal for all the samples even at very low film exposure (1 s; Figure 3.8A), indicating high total protein concentrations. The blot was analysed by densitometry where possible (Figure 3.8B). The quantification showed the following: 16% of the bait was successfully bound to the IgG beads, 0.9% was eluted after TEV protease cleavage and 1.7% was cut by the TEV protease, but was not recovered from the IgG beads. The final eluted bait amounted to 0.45%. This yield of the TAP purification was sufficient to identify a substantial number of proteins by MS, 310 in total. In the control TAP purification performed on 7.5 g of WT embryos, 121 proteins were identified. The proteins were also visible by staining with Coomassie blue on the gel before extraction for MS analysis (Figure 3.8C). A full list of all proteins identified by MS can be found in Table S1 in the Figshare repository (see data availability statement, section 2.3.13).

The overlaps between proteins identified in the different MS runs are illustrated in Figure 3.8D and E. The majority (56%) of the 310 proteins identified from O/N Mef2-GSTAP embryos (O/N MEF) were not found in any of the other samples, while only a quarter (27%) of the 56 candidates identified in 11-13 h embryos (11-13 MEF) were unique to that sample. Only 5 proteins were found in all four samples; the same 5 that were present in the sample purified from 11-13 h staged wildtype embryos. The O/N wildtype embryos (O/N CTRL) revealed 3 unique proteins (2% of 121 total), indicating that the proteins purified when no GSTAP tag is present in the sample are not simply a subset of unspecifically eluted, i.e. contaminant, proteins from a GSTAP sample. A more in-depth bioinformatic analysis was therefore carried out.

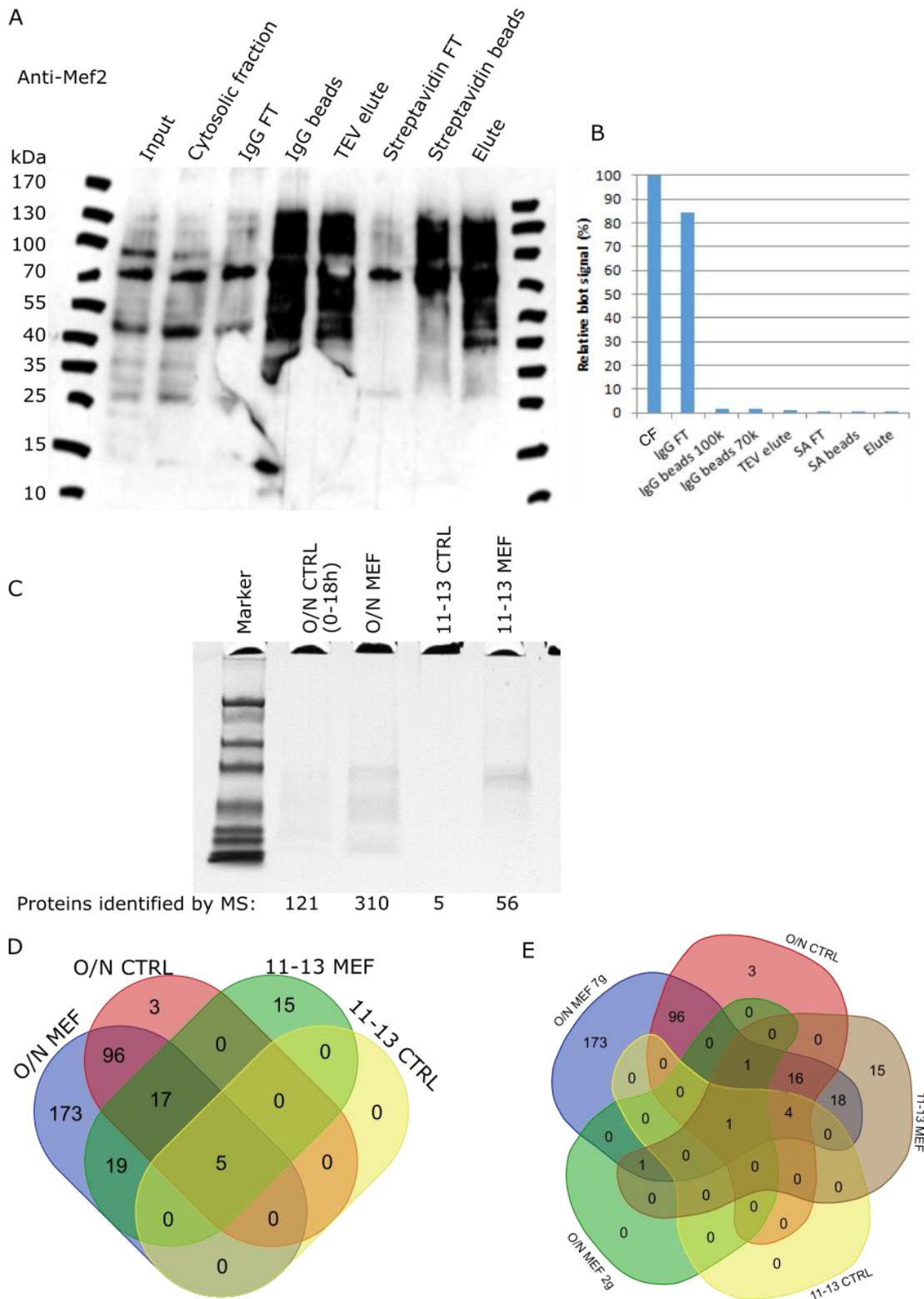


Figure 3.8: Final TAP experiment with 7.7g of O/N embryos expressing Mef2-GSTAP

A) Anti-Mef2 treated Western blot of the aliquots collected during the purification procedure to observe the loss of bait at each crucial step of the purification: Input, input extracts before clarifying the sample; cytosolic fraction (CF), solubilised

proteins fraction; IgG FT (flow through), flow through after incubating the CF with the IgG beads, any detected protein is lost protein; IgG beads, uncleaved bait protein left on the IgG beads after incubation with TEV protease; TEV elute, cleaved bait protein by TEV protease; Streptavidin F, flow through after incubating the streptavidin beads with the TEV elute; Streptavidin beads, bait protein not eluted off the streptavidin beads; Elute, final bait eluted with biotin that was used for MS analysis.

B) Quantification of the Mef2 bait identified in each analytical fraction of A) based on densitometry. The values were adjusted by the respective amount of each fraction loaded onto the gel and refer to the band of uncleaved Mef2 (near 100kDa) before TEV cleavage, and respectively the band of cleaved Mef2 (near 70kDa) after cleavage. The IgG beads fraction contained both cleaved and uncleaved Mef2, which was separately measured and represented on the plot (100k and 70k).

C) Coomassie stained gel of the proteins eluted from the TAP purification. O/N (0-18h) samples are from the final purification experiment using roughly 7 g of Mef2-GSTAP (O/N MEF) or WT (O/N CTRL) embryos, while the 11-13 h samples are from the previous purification experiment using roughly 2 g of Mef2-GSTAP (11-13 MEF) or WT (11-13 CTRL) embryos (section 3.2.5). The lists of proteins pulled down in each sample and total spectral counts can be found in Table S1 in the Figshare repository (see data availability statement, section 2.3.13).

D) Overlap of the Mef2 interaction candidates identified from the samples in (C). Lists of genes in each overlap category can be found in the Figshare repository (see data availability statement, section 2.3.13).

E) As (D), but also showing overlap with the sample purified from 2 g O/N Mef2-GSTAP embryos in the first TAP experiment (O/N MEF 2g).

3.3 Discussion

3.3.1 Input material required

The creation of the yTAP tag and the development of the TAP method has contributed to significant advances in the identification of protein complexes in a systematic manner. By taking advantage of the GSTAP tag introduced by Bürckstümmer et al. (2006) it was possible to identify candidate proteins that interact with Mef2 during *Drosophila* embryogenesis. The tag is supposed to require relatively low input biological material for performing tandem affinity purifications (Bürckstümmer et al., 2006). However, a small amount of input material still requires large scale collections of *Drosophila* embryos to provide enough bait for the TAP purification, such that the bait and co-purified proteins could be identified by MS. The collection of embryos required cycles of population expansion and eggs collections and several batches of collected embryos needed to be collated for one experiment. Over a 2 hours time window, one collection provided on average 250 mg, therefore even for a relatively small scale TAP experiment (e.g. 2 grams of input embryos) several collection rounds were required. Four different specific developmental time points were collected, as well as 18h unstaged collections that were labelled as O/N embryos. The collections required coordination of two different large *Drosophila* populations: the Mef2-GSTAP line, as well as the control wildtype embryos from the Oregon R (OR) line. Only collection of the input material required a solid 3 months-worth of egg collections and a well-defined collections schedule since only one collection of each stage was possible to achieve per day during a 14 days collection time window. During each cycle at which the fly population was in the correct laying stage, embryos had to be collected for different purposes as follows: 4 different specific developmental time points, O/N embryos and collection for population maintenance and expansion. In itself the collection of the required input biological material is a significant time investment.

3.3.2 Tracking Mef2 during TAP purification

By taking advantage of the properties of TAP and the GSTAP tag, the study was focused on purifying a tagged Mef2 protein from *Drosophila* embryos. Since no specific antibody was known for the GSTAP tag, a series of rabbit anti-Mef2

antibodies were tested on *Drosophila* embryos extracts for identification of Mef2 on western blots. The different bleeds were generated in house by the Furlong lab and were tested, which had previously been used in other types of assays (ChIP, immunohistochemistry). All the tested bleeds showed on the Western blots a specific band of 96 kDa for the GSTAP-tagged Mef2 and a 76 kDa band for the untagged Mef2 protein as determined by densitometric analysis. All reported Mef2 protein isoforms have molecular weights around 55kDa (see section 2.2.1), however there are no reports of the apparent molecular weight of Mef2 on Western blots when extracted from *Drosophila* embryos. This discrepancy could be due to post-translational modifications. Despite this, the 76kDa band was consistent between wildtype and TEV-protease treated Mef2-GSTAP extracts. Additionally, the 96kDa band only appeared in Mef2-GSTAP extracts that had not been treated with protease, and was the only band appearing with the anti-TAP antibody, indicating that the bands correspond respectively to GSTAP-tagged Mef2 and wildtype or cleaved Mef2. The anti-Mef2 antibody chosen for further experiments was DRC4 3rd since it gave the least amount of background. A 76 kDa band was also identified in S2 cells. The expression of Mef2 in HeLa cells was shown in other studies (Ornatsky and McDermott, 1996; Perry et al., 2009), but in S2 cells only studies with Mef2 that was transiently expressed through expression of exogenous constructs were done (Guruharsha et al., 2011). The background observed during incubation with the anti-Mef2 antibodies could be due to the use of β -mercaptoethanol in the gel sample buffer, an artefact which was observed in other cases (Hanukoglu, 1990). Other causes could be that an excessive amount of lysate loaded onto the gel may cause extra bands or a result of protein aggregation that could not be resolved by SDS and boiling.

To be able to observe the Mef2-GSTAP dynamics during the TAP purification two antibody options were available: antibodies against the tag or against Mef2. There is an anti-Protein G antibody that could be used to observe the cleavage of the tag in the initial steps of the purification. However, it could not be used to observe the TEV-cleaved Mef2 because the Protein G domains of the tag are distal to the TEV cleavage site. Another option to target the tag was an anti-SBP antibody that should be able to identify both the full length tag as well as the cleaved tag, since the SBP domain is proximal to the TEV cleavage site. The anti-Mef2 antibody would be a

feasible option, however previous testing showed that it gave some unspecific background. The antibodies targeting the tag recapitulated the 96 kDa band representing the tagged Mef2 as previously identified by the Mef2 antibodies. The anti-Protein G antibody was able to identify both the Mef2-GSTAP protein, as well as the part of the tag cleaved off by TEV protease. Despite offering a very clean signal, this antibody could not be used further due to the inability to observe the cleaved bait during the second part of the TAP purification. The anti-SBP antibody was not able to detect the bait after cleavage, even though the SBP moiety is not cleaved off by TEV protease, and the successful purification of bait by streptavidin beads in the downstream steps confirms that the SBP moiety remained on the bait. For these reasons all further investigations performed by Western blots were done by incubating with the anti-Mef2 DRC4 3rd antibody. The inability of the anti-SBP antibody to recognise the cleaved peptide could be due the cutting of the tag changing the structure of the recognised epitope, or the binding of the SBP peptide is too close to the end of the peptide chain which makes the interaction with the antibody inefficient.

It is common practice for Western blots to include staining for a loading control, particularly when protein quantities are compared between bands on the blot. A common choice is to stain actin or tubulin, under the assumption that their expression would be more or less uniform between samples in different lanes and their bands should therefore be equal sizes if the blotted gel was loaded properly. This has been omitted in this study because to serve its purpose of standardising a blot, the protein chosen as a loading control needs to be consistent across lanes. In this study, the different lanes in most blots contain samples from different stages of the purification procedure, and therefore by their nature contain varying amounts of various types of proteins. There are no proteins which could reasonably be expected to be present in all samples, let alone in consistent amounts. The purpose of most blots in this study was to assess the relative amount of bait present in each volume of liquid produced during the purification procedure. Similar studies have likewise omitted loading controls for TAP analytical fractions (Kyriakakis et al., 2008). The more important step to provide relevant quantifications was therefore to normalise the band intensities as determined by densitometry to the fraction of the total experimental volume that the band represented. Where this reasoning did not apply, a GFP

standard was instead included on the blot to ascertain loading consistency. No Western blots were performed for the WT control purifications because staining these for wildtype Mef2 (or any other) protein would not provide relevant information on purification quality.

3.3.3 Optimisation of protein extraction from embryos prior to purification

To homogenise the embryos, a mixture of mechanical shearing and chemical lysis was used. However, the amount of solubilised protein can vary depending on the lysis method. The lysis buffer was designed such that the isolated proteins were maintained under stable conditions and the compatibility with the first part of the TAP purification procedure was maintained. Tris base was used as the desired buffering system due to its compatibility with many proteins that were purified by TAP. Other additives were used to increase the stability of the isolated proteins: DNase was added to the lysis buffer to decrease viscosity of the lysate due to the release of the nucleic acids, NP40 was added to enable membrane disruption, DTT to deter oxidative damage, EDTA to chelate metal ions from proteases, Mg²⁺ and glycerol for stabilisation and protease inhibitors to protect the extracted proteins. No protease inhibitors were used in the TEV cleavage step to avoid inhibition of the bait cutting process.

Three different methods of lysis were tested using this buffer: 1) Liquid homogenisation, very quick homogenisation of the embryos resuspended in lysis buffer using a Wheaton douncer followed by immediate clarifying of the extract by ultracentrifugation; 2) Liquid homogenisation with incubation, an additional 30 min incubation on ice after homogenisation, before clarifying the extract by ultracentrifugation; this modification is often seen in published studies (Kyriakakis et al., 2008); 3) Hypotonic shock extraction. The aim of this step was to optimise the amount of solubilised bait extracted from the embryos. The lysis buffer already contained additives meant to improve the lysis process. For the third lysis option the incubation with a hypotonic version of the lysis buffer was intended to swell the cells and make them burst more easily under osmotic pressure, while the addition of high salt was meant to extract the proteins, in particular the nuclear ones. The hypotonic method was more efficient than liquid homogenisation with extended incubation as shown by the ratio of bait contained in the final lytic pellet to solubilised bait in the

supernatant. However, liquid homogenisation without extended incubation was favoured because it was equally efficient as hypotonic shock extraction, but also less time consuming. Using this method reduced the time required for the procedure by half compared to the other extraction methods, which could be a significant factor in maintaining physiological binding of Mef2 with its interacting proteins.

All three extraction methods showed that the amount of Mef2-GSTAP extracted from 10-12h embryos was many folds higher than in equal amounts of earlier staged embryos (4-6h and 6-8h). The Mef2 extracted from 4-6 h embryos was slightly more than from 6-8h samples. This aspect showed that for earlier staged embryos more input material was required for a TAP purification compared to a later staged one. It was assumed that O/N embryos would represent a rough average of the Mef2-GSTAP protein expression across time. Considering the difference of Mef2-GSTAP expression across the stages, the input required for a TAP purification from staged embryos should therefore be different from the amount of O/N embryos required. Earlier stages should require more input than O/N because of the lower Mef2 expression, while later stages should require less input than O/N.

3.3.4 TAP purification of Mef2-GSTAP and associated complexes from 2 g of O/N embryos

The first TAP purification was performed on 2 grams of O/N Mef2-GSTAP embryos, an amount of starting material on the lower scale of what was reported in previous TAP studies in *Drosophila* (Kyriakakis et al., 2008; Veraksa et al., 2005). The goal of this experiment was to identify a minimum amount of starting material required, to enable economical use of the collected material and keep enough available for repeat experiments.

The lysis was performed as described, however the binding of Mef2-GSTAP to the IgG beads was extremely low as estimated by densitometry. The low binding of the Mef2-GSTAP to the IgG beads could be due to saturation of the IgG beads, either because more bait than the maximum binding capacity of the beads was available or because other proteins present in the lysate were able to non-specifically interact with the IgG beads such that the majority of them were masked for binding by the tagged Mef2. An important conclusion from this initial test was that despite having

an efficient cleavage of the bound bait, the elution of the cleaved bait, i.e. separating the cleaved bait from the IgG beads, was also a limiting step.

The ability of mass spectrometry (MS) to identify Mef2 as the most abundant protein in the elute sample showed that in principle the purification protocol worked for extracting Mef2 and associated proteins. Additionally, this result also suggested that the amount/number of Mef2-associated proteins that can be identified by MS is proportional to the bait eluted after the purification. A sufficient amount of bait needs to be pulled down to have proteins above the detection threshold of MS. The main two strategies on how to obtain more bait were to increase the amount of input embryos from which to extract bait, as well as to improve the intermediate steps of the purification procedure such that more bait and associated protein is extracted at each step to obtain the best possible overall yield.

The extraction from wildtype (WT) embryos as a control sample was intended to identify the most abundant proteins in the *Drosophila* embryos that are able to bind non-specifically to the beads and can withstand the different binding/elution steps as well as the washes. The proteins most commonly identified as probable contaminants of TAP purification are abundant cytosolic proteins such as myosins, heat shock proteins, actins and ribosomal proteins (Veraksa et al., 2005). The MS results of the WT embryos was very clean, only Actin5C was identified. However, in the Mef2 sample both an actin and a heat shock protein were detected. This is consistent with the tissue of interest being muscle, a cell in which myosin, actins and other generally abundant proteins are present and functionally enriched (Uhlen et al., 2015). The result therefore highlights a difficulty in the interpretation of the Mef2 sample compared to the WT sample. Due to its role in muscle development, Mef2-associated proteins are likely to be muscle-enriched. However, once developed, muscles are a large protein reservoir in the embryo and are also a significant source of contaminant proteins in purification studies. Most proteins commonly considered contaminants may as a consequence be purified both specifically as Mef2-associated proteins, and unspecifically from WT embryos. Proteins found in both samples can therefore not necessarily be excluded, but rather may be valid interactor candidates.

3.3.5 TAP purification of Mef2 complexes from 2 g of 11-13 h embryos

The first experiment using 2 grams of O/N embryos as starting material for TAP purification showed that this amount of embryos was not sufficient to detect Mef2-associated proteins by MS. To assess if using more input bait was able to improve the detection rate while maintaining a low total sample volume, 2 grams of 11-13h staged embryos were used for the second experiment. These embryos were expected to yield more bait than a similar amount of O/N embryos because of the higher expression of Mef2 during these late embryonic stages. The amount of Mef2 extracted from 10-12h embryos in the previous lysis optimisation also suggested that the 11-13h sample should provide increased amounts of Mef2 bait. Also, an increased concentration of TEV protease was used to enhance the cleavage of the bait off the beads and thus increase the amount of eluted cleaved Mef2-GSTAP. A direct comparison of the ratios of cleaved to uncleaved Mef2 of the two experiments showed that using more TEV protease did substantially improve cleavage efficiency. Despite this, the relative percentage of eluted cleaved Mef2 was only slightly larger than in the previous TAP experiment from 2 grams of O/N embryos. Due to the larger absolute amount of bait in the 11-13h staged embryos, the final yield of obtained bait was sufficient to allow detection of 55 associated proteins.

It is important to note that despite having a similar amount of embryos lysed for the input material, the biological background is very different between the O/N embryos and 11-13h staged samples. While the 11-13h time point enriched for proteins involved in terminal muscle differentiation, the O/N embryos contain an average overview of Mef2 associated proteins, most probably complexes that interact with it throughout development. Depending on the proportion between proteins that specifically interact with Mef2 during late stages of myogenesis and the fraction of the O/N samples constituted by embryos from those stages, late-stage complexes might still be identified from O/N embryos. It is more likely that complexes that overlap between the two samples are involved in interactions with Mef2 for a longer time during development, rather than only towards the end of embryogenesis. The control sample was quite clean, only 5 protein were identified: Actin5C (which was also detected in the first TAP purification experiment), heat shock protein, two

vitellogenin proteins and elongation factor 1-alpha. This last detected protein is not one of the standard proteins classified as TAP contaminants.

3.3.6 Improving bottleneck steps of TAP purifications

In both TAP experiments the binding of Mef2-GSTAP to the IgG beads was quite inefficient although the amount of total protein should have been within the recommended range. A recommended starting point was to incubate 100 mg of total protein lysate with 150 μ l IgG beads (Kaiser et al., 2008). The total protein detected in the clarified lysates for both the O/N embryos and 11-13h embryos was below the recommended starting incubation amounts, therefore there should have been enough beads available to allow efficient binding of the entire Mef2 bait to IgG beads. This suggests that the more likely problem was non-specific binding of other proteins to the beads, hindering binding of the bait. To test this hypothesis, another 2 grams of O/N embryos were lysed and the cytosolic fraction containing 6.8mg of total protein was split into four fractions of increasing size (500 μ l, 1 ml, 1.5 ml, 2 ml), each to be incubated with the same amount of IgG beads (20 μ l). Considering that an amount of 10 mg of total protein was recommended as suitable to be incubated with 15 μ l IgG beads (Kaiser et al., 2008), 20 μ l of beads should have been sufficient to bind most if not all of the bait present in the cytosolic fraction in total, and even more so when split into smaller fractions where the largest fraction contained only 3.4 mg of protein.

The experiment showed almost full binding of bait in the small volumes of cytosolic fraction (500 μ l, 1 ml), while less than 70% were successfully bound to the beads in the larger volume fractions. IgG bead binding efficiency therefore did not depend on the total amount of protein, as the smaller fractions contained less total protein. Rather, considering that the amount of beads was kept constant, these fractions had a higher ratio of beads to protein, both bait and non-specific proteins. Consequently, this experiment suggests that with a lower bead-to-protein ratio, non-specifically binding proteins are able to attach to the IgG beads during incubation, thus interfering with the affinity binding of the tagged bait. This aspect was particularly observable in the wash aliquots where the Mef2 antibody was able to detect nonspecific proteins removed during the washes. The blots obtained for the 1xCF and 2xCF samples were much cleaner than the ones of the 3xCF and the 4xCF. This

could be a direct consequence of an increased amount of non-specifically bound proteins extracted with the beads and then eluted off them in the subsequent steps. Their detection on the blot additionally suggests that the proteins that interact non-specifically with the IgG beads also cross-react with the Mef2 antibody.

Considering the above, a more efficient and clean option for binding Mef2-GSTAP to the IgG beads was to increase the amount of beads as well as to incubate the beads with less volume of cytosolic fraction such that the ratio of non-specific proteins to IgG beads was kept at a minimum.

3.3.7 TAP purification of Mef2 complexes from 7.7 g of O/N embryos

The use of almost 8 grams of O/N embryos was successful in providing a large set of Mef2-associated proteins that were identified by MS and had a significant amount of spectral counts associated with each protein, thus the amounts of co-purified proteins were in quantities above the detection threshold. A total of 121 proteins were identified in the control experiment and 310 in the Mef2-GSTAP samples. The binding efficiency to IgG beads was lower in this experiment than the previous runs. This was despite the modifications to the protocol, such as incubating the lysate in smaller fractions and with more beads, that should have improved this according to the optimisation tests. A possible explanation for this could be that even though the amount of beads was doubled, this larger amount of input embryos may have required even more IgG beads to provide efficient binding. However, the practical restrictions of the downstream steps, in particular the size of sample containers, limited the amount of beads that could be used. However, despite the relatively low yield of individual steps of the purification, the overall yield of the purification was sufficient to provide information on candidates interacting with Mef2.

Taking into account all the steps taken to improve the yield obtained by purifying Mef2 by tandem affinity purification and the reduced effects registered for each modified strategy, it appears that it is quite difficult to standardise the TAP purification protocols. The primary requirement of any TAP purification experiment is to provide sufficient amount of bait protein such that the final elute target protein and associated peptides are above the detection threshold of the MS. The protein composition given by the biological context of the analysed *Drosophila* sample and

the amount of eggs required to provide a sufficient amount of bait are the two main characteristics that influence the outcome of the TAP experiment. Therefore, selecting a generous amount of starting material and of IgG beads can ensure that the TAP experiment will have the highest chances of success. Using the detection by Coomassie of the eluted proteins represents a crude, but quite reliable measure in assessing the ability to detect the purified proteins by MS, as previously found by others (Kyriakakis et al., 2008).

3.3.8 Purification yield

The GSTAP tag was initially used in mammalian cell culture and was reported as giving ten-fold improvement in yield compared with the canonical yTAP tag (Bürckstümmer et al., 2006). For *Drosophila* embryos, a three-fold increase in yield compared to the yTAP tag was reported (Kyriakakis et al., 2008). In this study Kyriakakis et al. tested both tags in S2 cells and 0-6h *Drosophila* embryos. The yTAP and the GSTAP tags were used to extract a cytoplasmic protein as bait for co-purification. In their analysis, they reported a 2.7% yield for the yTAP tag, and respectively 8.4% for the GSTAP tag when extracting from *Drosophila* embryos. The test performed in S2 cells showed a similar relationship between the two tags but with lower overall efficiency (8.1% GSTAP vs. 6.8% yTAP). In the present study, the best yield achieved despite many adjustments of the protocol and several optimisation trials was 0.45%, a 1.7-fold improvement from the 0.27% yield of the first TAP experiment. However, this yield was still substantially lower than in reported studies. A possible explanation for this is that the bait used here was a predominantly nuclear protein, which can interfere with purification unless very large amounts of input material are used, or specific steps are taken to first isolate nuclei. *Drosophila* embryo collections at such a scale were not feasible for this study. Yet, the relative fraction of bait extracted does not reflect the absolute amount of bait present, which was nevertheless sufficient to identify proteins interacting with Mef2 both from the 2g 11-13h extraction and the 7g O/N extraction. The yield also has no impact on the quality of the data obtained using MS or the confidence in the identified proteins as specific interactors of Mef2. The primary drawback of low yield efficiency is that very large amounts of input material are needed to extract a high enough absolute amount of protein for detection by MS.

Chapter 4: Bioinformatic analysis of Mef2-associated proteins as determined by proteomics studies

4.1 Introduction

Transcription factors sit at the core of differentiation processes and usually they are involved in complex protein-protein interactions (PPI) and protein-DNA interactions in order to govern the different gene expression programs. The identity of the specific regulatory networks active in a particular cell or tissue defines much of that system's phenotype. Large-scale purification studies to extract systematically the molecular interactions that take place in such networks have created a wealth of available experimental data. Due to the increasing popularity of specialised databases on model organisms and the standardisation of the different available datasets it has become more easily accessible to extract experimental data performed in other biological systems and correlate it to customised tasks.

The advances of complex protein purification methods coupled with Mass Spectrometry (MS) allow to systematically identify protein complexes and protein networks. While experimental approaches like the yeast two-hybrid system, affinity purifications, chemical cross linking, chemical foot printing, protein arrays and fluorescence resonance energy transfer are helpful in determining PPIs, computational approaches are necessary to estimate which connections are the most reliable. Through the use of databases such as STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) (Franceschini et al., 2013; von Mering et al., 2005), DIP (Database of Interacting Proteins) (Salwinski et al., 2004) and Predictome (Mellor et al., 2002) one can easily derive associations between proteins. Usually the interactions between proteins coming from such databases means a functional relationship between the two connected proteins, but not necessarily an actual physical interaction.

The Mef2 TAP purification performed in Chapter 3 revealed tens to hundreds of candidate proteins that interact with Mef2 in *Drosophila* embryogenesis. Mef2 is active in a lot of biological contexts and is responsible in activating genes with diverse expression profiles. Understanding how Mef2 is regulated is of particular interest as the protein was identified to be expressed throughout embryogenesis, therefore being active both in undifferentiated cells and fully differentiated tissues. The protein has been associated with a wide diversity of functions ranging from cellular functions such as the regulation of gene expression and nucleic acid

metabolism or biosynthesis to more complex, tissue relevant processes like mesodermal and muscle development, locomotion, and immune response.

4.1.1 Experimental approach

In this study the candidates identified by TAP/MS as Mef2 binding proteins are bioinformatically screened as interactors that form a connected network. The proteins identified by MS are expected to contain: 1) false positives, i.e. contaminants that do not interact with Mef2; 2) direct interactors of Mef2; 3) as the aspect unique to TAP, indirectly interacting proteins pulled down as part of Mef2 protein complexes. To increase our understanding of Mef2 regulation through PPIs, two bioinformatics approaches were chosen. Such analyses have been shown to be useful in understanding the quality as well as the functionality of the experimental data (Guruharsha et al., 2011; Murali et al., 2014; Rhee et al., 2014).

The first approach is based on the spectral counting method, in which the data is treated semi-quantitatively in that the total number of peptides identified for a protein by MS is interpreted as a reflection of the abundance of that protein in the sample. This analysis then assumes that proteins detected in wildtype samples are contaminants and compares the abundance of each protein in the Mef2-GSTAP datasets with their abundance in wildtype datasets to classify all candidates as binders or contaminants (section 4.2.3). The next step then aims to narrow down the list of candidates or to rank candidates and identify the most promising Mef2 interactor candidate for further study (section 4.2.4). This analysis thus minimises false-positives while having to accept false-negatives (ignoring potentially valuable candidates). Integrated with literature background, this analysis yielded HDAC4 as a promising candidate Mef2 interactor during muscle development, which was further studied in Chapter 5.

The second approach was based on network models and investigated the functional relatedness between candidates. This was intended firstly to validate the experimental success of the extraction, but also aimed to identify potential functional modules, i.e. clusters of proteins that might form complexes with Mef2. This approach takes into account that proteins that can be unspecifically extracted from wildtype embryos through TAP could nevertheless be Mef2 interactors. Thus, the

analysis considers all candidates to have equal probability of being a real interactor from the start. It aims to minimise the loss of information by not discarding potential false-negatives and instead integrates the datasets with published data of PPIs (sections 4.2.6 to 4.2.9), gene expression (section 4.2.10) and functional annotations (sections 4.2.11 and 4.2.12). Using the paradigm of "guilt by association", the better characterised a relationship between proteins is, the stronger their association in the network becomes. Similarly, correlations in gene expression and functional annotation imply a higher probability that proteins are part of shared molecular pathways or physically interact with each other. By integrating these data, we can validate the quality of the extraction since it should be reflected in shared annotations between the pulled down proteins, and gain a deeper understanding of the biology reflected in each dataset.

Mef2 was endogenously GSTAP-tagged in *Drosophila* embryos, purified by tandem affinity purification (TAP), and the eluate was analysed by mass spectrometry (MS). Parallel negative control purifications were performed from wildtype (WT) *Drosophila* embryos of equivalent age using the same TAP protocol and analysis. These controls are aimed to help distinguish contaminants from specific binders among the proteins identified by MS in the purified samples. The bioinformatic analyses in this chapter also considered as positive controls Mef2 interactors identified in published studies that used other affinity purification methods analysed by MS (Rhee et al., 2014; Guruharsha et al., 2011). The published controls are meant to test the validity of the two approaches selected to interpret the Mef2-TAP purifications results.

4.2 Results

4.2.1 Datasets of candidate Mef2-interacting proteins to be investigated

The TAP process is a two-step extraction method where protein complexes containing the GSTAP-tagged Mef2 are pulled down from whole cell lysates of *Drosophila* embryos and analysed by MS. The datasets of candidate proteins obtained and their biological sources are listed in Table 4.1 and are described in more detail in section 4.2.2. A total of 328 unique candidate proteins were identified from *Drosophila* embryos across all TAP samples, of which 3 were unique to WT embryos and another 118 were found both in WT and tagged-Mef2 purified samples.

Two other datasets were used in the analyses, both resulting from single-step affinity purifications of FLAG-HA tagged Mef2 expressed in S2R+ cells analysed via MS. The Guruharsha et al. (2011) purifications were performed on S2R+ whole cellular lysates, while Rhee et al. (2014) used the same cell lines but extracted protein from nuclear fractions only. Both affinity purifications were performed as part of large scale studies and the elimination of false positive results was assessed via a statistical approach which accounted for protein abundance and consecutive observations in sequentially performed experiments. The Guruharsha study reported a total of 101 unique Mef2-interactors (DPiM), while 681 were described in the other study (S2-CoAP).

The candidate proteins reported in the two published studies were identified via their Flybase gene identifiers. The lists of candidate proteins from TAP purifications were mapped to Flybase gene identifiers via their UNIPROT identifiers. To allow comparison between all the datasets, the protein identifiers of all candidate proteins used in these analyses were converted to Flybase gene identifiers according to database release 2016_02. These identifiers are compatible with the *Drosophila* Interaction Database (DROID) version 2015_12 which represents the basis of building PPIs for the network analysis approach.

4.2.2 Experimental characteristics of candidate datasets

Mef2 like many transcription factors is able to activate or repress transcription through binding DNA at specific locations. In order to regulate cellular processes it

Table 4.1: Protein datasets

Label	Description	Size
Mef2GSTAP Datasets		
ONCTRL	0-18h embryos, whole cells, negative (WT) purification	121
ONMEF	0-18h embryos, whole cells, Mef2GSTAP purification	310
11-13CTRL	11-13 h embryos, whole cells, negative (WT) purification	5
11-13MEF	11-13 h embryos, whole cells, Mef2GSTAP purification	56
External datasets		
S2-CoAP	S2R+ cells, nuclear extracts, Mef2-FLAG-HA purification	681
DPiM	S2R+ cells, whole cells, Mef2-FLAG-HA purification	101
Derived datasets		
ON/S2-CoAP	Overlap between ONMEF and S2-CoAP candidates	102
11-13/S2-CoAP	Overlap between 11-13MEF and S2-CoAP candidates	29
ONMEF*	ONMEF candidates classified as specific by spectral counting	227
11-13MEF*	11-13MEF candidates classified as specific by spectral counting	50
Combined datasets		
Mef2 candidates	Union of all Mef2GSTAP datasets and S2-CoAP candidates	888

interacts with other proteins such as other transcription factors, chromatin modifiers and other cofactors. These interactions are crucial for gene expression modulation and they define the outcome of cell differentiation.

Three TAP experiments (detailed in chapter 3) were performed to identify candidate proteins that interact with Mef2 in *Drosophila* embryos to modulate muscle tissue development. The proteins extracted by TAP were identified using mass spectrometry (MS). There were two bulk collections in which embryos of all development stages were collected overnight (O/N) and subjected to the TAP

protocol. Overnight collections contain embryos of ages in the range 0-18h, and thus span all stages of embryonic development. Besides various adaptations to the purification protocol, the main difference between the two experiments consisted in the amount of embryos used as input: 2 g of O/N embryos in the first and respectively 7 g in the second. Since only 2 proteins were identified in the TAP purification from 2g of O/N embryos, this sample was not studied further in this chapter. In the third experiment, candidate proteins that are potential interactors of Mef2 during late embryonic myogenesis were identified by purification of proteins from 11-13h staged embryos. By this stage, the embryos have completed development of the larval musculature. Additionally to the collections from Mef2-GSTAP expressing embryos, each of the three experiments was accompanied by a parallel extraction from equally collected (O/N or 11-13h staged) wildtype (WT) embryos, where no bait is present and thus any identified proteins passed through the purification by non-specific binding. The data derived from WT embryos will be treated as an internal control for the experimental purification procedure.

The list of proteins for each sample returned by the proteomics facility at EMBL contained the following information: Description of the protein identified, the UNIPROT accession number, the protein's predicted molecular weight as derived from the sequence available in the database, and number of spectral counts. Two separate protein lists were rendered for each sample, containing total spectral counts and unique spectral counts respectively. The database used to identify the peptides rendered from the sample was UNIPROT, restricted to *Drosophila melanogaster* sequences.

Prior to mass spectrometric analysis, trypsin is used to digest the proteins into smaller peptides of 7 to 11 amino acids length. The resulting peptides are broken down by the mass spectrometer and the fragmentation spectra are recorded. These spectra are afterwards matched to a database which contains sequence specific data of known model organisms. Mass spectrometry detection of peptides does not provide chemical sequencing of these proteins, but uses the recorded spectra for each peptide and an algorithmic comparison with a databases of proteins to match the sample peptides to known proteins. Some fragments can match the sequence of more than one protein. Spectra that can be matched to only one specific protein are

classified as unique spectral counts. The total spectral count for a protein thus contains both its unique spectral counts and those spectra that were matched to it but could also have been derived from a different protein.

To test the validity of the obtained data, we tested the range of spectra obtained in other studies where Mef2 was bait. The *Drosophila* Protein Interaction Map (DPiM; Guruharsha et al., 2011) was built from clones transiently expressing FLAG-HA tagged *Drosophila* proteins in S2R+ cells. Using co-affinity purification, the proteins that interact with the tagged proteins were extracted and identified via tandem MS. Mef2 was one of the tagged protein that was successfully expressed and its interaction partners were identified. The spectral counts of these proteins were reported and were used as a quality control for the present Mef2 TAP purification results. A single purification experiment for the Mef2 dataset was reported in this study and we are not aware if the raw data reported is the consequence of several biological or technical replicates. A total of 101 unique proteins were identified in the affinity purification of FLAG-HA-Mef2 from whole lysates of S2R+ cells expressing transiently the tagged-Mef2. This dataset will be referred to as "DPiM".

Rhee et al. (2014) used a similar FLAG-HA tagged Mef2 clone taken from the Universal Proteomics Resource (Yu et al., 2011), a part of the Berkeley *Drosophila* Genome Project (BDGP) to build transcription factor networks in *Drosophila*. The FLAG-HA tagged Mef2 was transiently expressed in S2R+, nuclear extracts of these cells were obtained and single-step affinity purification with Mef2 as a bait was used to extract the nuclear interactome of Mef2 in S2R+ cells. To analyse the purified proteins, the sample was trypsin-digested and the peptides obtained were analysed by tandem MS and mapped to a database. Due to the similar way of analysing the data, the spectral values reported in this study were used as a control for the spectral counting methodology that was applied to semi-quantitatively analyse the candidate proteins obtained in our TAP experiments (see section 4.2.3). 643 unique bait-prey interactions and two replicates were reported for the Mef2 clones. The published data include only total spectral counts for each candidate protein. The second replicate identified only 215 proteins, all of which were also present in the first replicate and therefore represent a 32% consistency between the two replicates. This dataset will be referred to as "S2-CoAP".

4.2.3 Classification of candidate proteins based on spectral counts identified by MS

The higher the number of matched spectra for a protein, the higher the likelihood of the identification being valid, however a minimum of 2 spectra per protein is expected before a protein is considered to be identified with confidence. The number of matched spectra for a protein can be considered an indication of its concentration in the sample, however in order to make estimations of relative concentrations of different proteins, each protein should have at least above 4 spectral counts for a minimum of reasonable linearity. It is important to note that if a protein has 0 spectral counts in a sample, it is not possible to conclude that this protein is absent from the sample. Besides an actual concentration below the detection threshold of MS, it is for example also possible that all peptides detected for a protein fall below the confidence cutoffs that are applied when mapping detected spectra to the protein database. Spectral counting is therefore a semi-quantitative way to estimate amounts of proteins in the analysed samples and it is mostly reliable if we compare the different proteins in the same sample.

The overnight collections of *Drosophila* embryos were subjected to TAP purification and the eluate was analysed via tandem MS. WT embryos were used as control for the purification procedure and were treated in parallel to the Mef2-GSTAP expressing embryos. The resulting MS datasets were labelled ONCTRL (for the WT embryos) and ONMEF (for the Mef2-GSTAP embryos) and a total of 121 proteins, respectively 310 proteins were matched during MS analysis. Only one MS run was performed per sample. It is important to note that the purification is based on antibodies against Protein G, and a streptavidin-binding moiety, neither of which are present in WT *Drosophila* embryos. Proteins identified in the WT sample are therefore expected to be contaminants of the purification process that are able to bind unspecifically to the antibody-coated and the streptavidin-coated beads, and could not be removed during the various washing steps of the protocol.

The staged Mef2-GSTAP embryo collections (11-13MEF) were treated in a similar manner as the overnight collections and a WT staged collection (11-13CTRL) was purified in parallel as a control. The purification process of 11-13CTRL and 11-13MEF resulted in 5 proteins identified in the control sample and 56 proteins

identified in the Mef2-GSTAP expressing embryos. The Mef2-interacting candidate proteins identified in these samples should show a snapshot of the complexes that interact with the transcription factor towards the end of the larval muscle developmental process.

The range of spectral counts of all the TAP purifications and the two published studies was compared. The samples expressing Mef2-GSTAP (ONMEF, 11-13MEF) were the only ones having proteins with total spectral counts for any protein higher than 100. To estimate the abundance of proteins semi-quantitatively based on total spectral counts, a minimum threshold of 4 total spectral counts was considered. While proteins can be confidently identified with fewer spectral counts than this, such low numbers of spectral counts are very unreliable estimators of the concentration of the protein in the analysed sample. Only the overnight samples had more than half of the proteins with spectral counts above this threshold (66% of ONMEF and 60% of ONCTRL). In all other samples, the majority of proteins had a spectral count of 4 or lower. Only 26% of the 11-13MEF proteins had 4 or more counts, while only one protein in the 11-13CTRL reached the 4 total spectral counts threshold. Of the proteins in the published studies, 38% (S2-CoAP) and respectively 28% (DPiM) of the proteins met the threshold.

The length of a protein can influence the number of spectra obtained for it in the MS analysis, generally the longer the protein the more spectra are available. Therefore a protein that is larger can seem more abundant than a smaller one due to the possibility of generating more spectra corresponding to the larger number of peptides per molecule after tryptic digestion. To compensate for the effects protein length has on spectral counts, a normalised spectral abundance factor (NSAF) was created (Zybailov et al., 2006). NSAF is calculated for each protein in a sample by dividing the total spectral counts (TSC) by the protein's length (L), this ratio being further divided by the sum of all TSC/L of all the proteins in that particular sample. The numbers obtained for NSAF are between 0 and 1 and the closer the NSAF of a protein is to 1, the more abundant the protein is. NSAF allows to compare the abundance of the different proteins in the same mixture and to compare the abundance of various proteins in different samples (Paoletti et al., 2006).

To be able to compare the spectral counts of the identified candidates, the NSAF of each protein was calculated. All the studies included in this analysis involved purification of proteins where Mef2 acted as a bait except for the wildtype controls where no bait was expressed. Therefore, Mef2 is expected to be enriched in all datasets where it acted as bait, and should be among the most abundant proteins if the purification procedure was successful and specific. When ranked by NSAF, Mef2 was the most abundant protein in 11-13MEF and the second most abundant in ONMEF. In S2-CoAP, Mef2 ranked 40th (top 7%), while in DPiM it ranked 29th (top 29%). Therefore, for the TAP purifications, Mef2 was among the first proteins while in the single-step purifications from the published studies it was not among the most abundant proteins. There were only 2 proteins that were identified in all samples (including controls) and only 4 proteins that were present in all samples except controls ("Mef2 samples"; 4 including the bait). The 2 proteins that were identified in all the samples had ranks relatively close to the bait, while the 3 non-bait proteins that are only found in Mef2 samples have very low ranks compared to the bait. One of the 3, RpS3A, is a ribosomal protein which was more abundant than the bait in DPiM, while it is less abundant in S2-CoAP. In the TAP purifications this protein is significantly less abundant compared to the bait. Since the values of NSAF for the same protein differ significantly between samples, it is difficult to estimate the stoichiometry of these protein in relation to Mef2. The 2 proteins that are found in controls and samples are Act5C and Hsc70-4, two housekeeping genes, and their abundance is very close to the bait during TAP experiments and more abundant than the bait in FLAG affinity purifications. Since very few proteins are found in all samples, and due to the inconsistency of relative abundances between samples, only comparing the NSAF values or ranks of proteins with samples and between samples does not present a reliable method to prioritise the list of candidate proteins.

The separation between specific and nonspecific binders is essential to distinguish likely Mef2-interacting candidates from contaminants. The NSAFs of proteins purified by TAP were therefore used in a second approach, whereby statistical analysis is used to compare specific (bait-containing) purifications with control (bait-less) purifications. The detected proteins were represented as vectors containing the respective NSAF values of each protein across the four TAP datasets. These vectors were used to build a matrix where the rows represent proteins and columns represent

experiments. The experiments include both control experiments where WT embryos were subjected to TAP purifications and the specific samples extracted from Mef2-GSTAP expressing embryos. To determine if a protein was a contaminant, a ratio of vector magnitudes (α) was calculated as follows:

$$\alpha = \sqrt{\frac{y_{i1}^2 + y_{i2}^2}{x_{i1}^2 + x_{i2}^2}}$$

where $x_{i1/2}$ = NSAF in Mef2-GSTAP embryos (ON and 11-13), and $y_{i1/2}$ = NSAF in WT embryos (ON and 11-13). Proteins for which $\alpha > 1$ are considered contaminants (Sardiu et al., 2008). The published studies were not included in this analysis since there were no negative controls available to be contrasted against as the described method requires. Note that proteins only present in control datasets have no α coefficient because $x_{i1} = 0$ and $x_{i2} = 0$. Such proteins are instead classified as contaminants by default.

The ratio of vector magnitudes was calculated for all candidate proteins in the TAP experiments. Based on this α coefficient determination, the candidates were then classified as below or above 1. Of the total 328 potential candidates, 83 candidates had an α coefficient larger than 1, and were thus classified as unspecific binders. Since the α coefficient is a cross-dataset measure, we tested how many proteins in each dataset had been classified as a contaminant. All 83 contaminants are present in the ONMEF dataset. In the 11-13MEF sample, 6 proteins were identified as contaminants, 4 of which were proteins also present in the 11-13CTRL dataset, while the remaining 2 were also present in ONCTRL. The 3 proteins present uniquely in ONCTRL were regarded as contaminants by default and removed from further analysis.

Based on this algebraic analysis of NSAF abundance, 6 proteins were classified as contaminants in 11-13MEF, while 50 were regarded as specific Mef2-binding proteins. ONMEF has 227 proteins considered specific binder, while 83 were contaminants. These specific binders were grouped into two new datasets and were used further in the following network analysis model to contrast how the network behaviour changes when potential contaminants are removed. The two new datasets

are denoted as ONMEF* which includes the 227 specific interactors, respectively 11-13MEF* has 50 proteins.

4.2.4 Identification of best candidates for Mef2 regulation during myogenesis

The aim of the proteomic study is to identify candidates that regulate Mef2 activity to activate target genes. In this first approach, we aim to identify the most promising individual Mef2 interactor candidates for further study (Figure 4.1A). The second approach in the following chapters instead aims to identify potential protein complexes or functionally cohesive groups of proteins from a systems-level point of view.

Many studies show that Mef2 is able to cooperate with other proteins in a context dependent manner that will create a variety of transcriptional responses. The most promising candidates for further study should thus be found in 11-13MEF*, since proteins bound to Mef2 during this stage would most likely be interacting with it in the context of myogenesis. Since the two large scale studies performed in S2 cells and ONMEF* contain proteins that interact with Mef2 respectively in a non-muscle cell type and in non-specific developmental stage contexts, we used them here to narrow down the shortlist of candidates to those that are specific for the 11-13h staging. Thus the datasets 11-13MEF* and ONMEF* were compared and the proteins only found in the specifically timed sample were extracted. This reduced the list of 50 proteins down to 15 proteins found only in 11-13MEF*. These proteins were further shortlisted by comparing the 15 proteins with the Mef2 specific interactors extracted in the DPiM and S2-CoAP studies. None of the 15 proteins left in the 11-13MEF sample were found in the DPiM study, while 7 proteins were pulled down together with Mef2 in the S2-CoAP study. Therefore the final list of candidates comprises 8 proteins specifically found in the 11-13h embryos. The other proteins were eliminated from further analysis. Since these proteins were found interacting with Mef2 in more than one experimental setup, they are likely true interactors of Mef2, though potentially not specific to myogenesis. The list of 8 candidates was analysed for known associations using the STRING database (Szklarczyk et al., 2019) and the network showed that 6 of the proteins are likely to be part of a muscle specific complex (Figure 4.1B). The 2 remaining proteins were dHDAC4, a class IIa HDAC that modulates gene expression together with transcription factors and

Brahma, an ATP-dependent helicase, a subunit of the Brahma complex that acts as a transcriptional activator.

To assess whether any of the final 8 candidates is already known to participate in muscle development, we cross-referenced this list with a *Drosophila* RNAi screen where 10461 genes were specifically silenced in muscle using a Mef2-Gal4 driver (Schnorrer et al., 2010). The primary screen in this study assessed viability, posture, locomotion and ability to fly of flies expressing each RNAi construct in their muscles. The secondary screen assessed the morphology of the larval body wall muscle under RNAi targeting 436 genes that showed embryonic or larval lethality. The observed phenotypes were categorised either as muscle morphology defects (missing muscles, rounded muscles, split myofibril) or as sarcomeric organisation defects (fading Z, spotty Z, clumpy Z). This study will be further referred to as "muscleRNAi".

Most of the genes corresponding to the short-listed proteins in dataset specific for the 11-13h embryos (7/8) were also examined in the muscleRNAi primary screen (the gene *up* was not screened). Only 3 of the genes were part of the secondary screen for larval wall musculature defects. Five of 7 genes were lethal, of which one also showed a larval muscle morphology defect (*wupA*), and two showed a sarcomere defect (*sls* and *tm2*). The other two were not screened for muscle phenotypes (*zip* and *jar*; see Table 4.2). These five, together with *up*, form the group identified by the STRING database as a muscle-specific complex and the proteins they encode are also well-known and central building blocks of the sarcomere in muscles (Vigoreaux, 2001). Since sarcomeric proteins should be localised in the cytoplasm, and Mef2 as a transcription factor should be nuclear, we obtained the gene ontology annotation for these genes to assess their cellular component (localisation) and molecular function (Table 4.2). Two of the sarcomeric proteins were already annotated with both cytoplasmic and nuclear localisation (*wupA* and *sls*), suggesting that presence in the nucleus might be a general feature of sarcomeric proteins.

The only genes that were not associated with lethality in the muscleRNAi screen were dHDAC4 and Brahma, both localised to the nucleus and involved in transcriptional regulation according to their gene ontology annotation. Out of all the 8 candidates, these two proteins are likely to interact with Mef2 to regulate gene

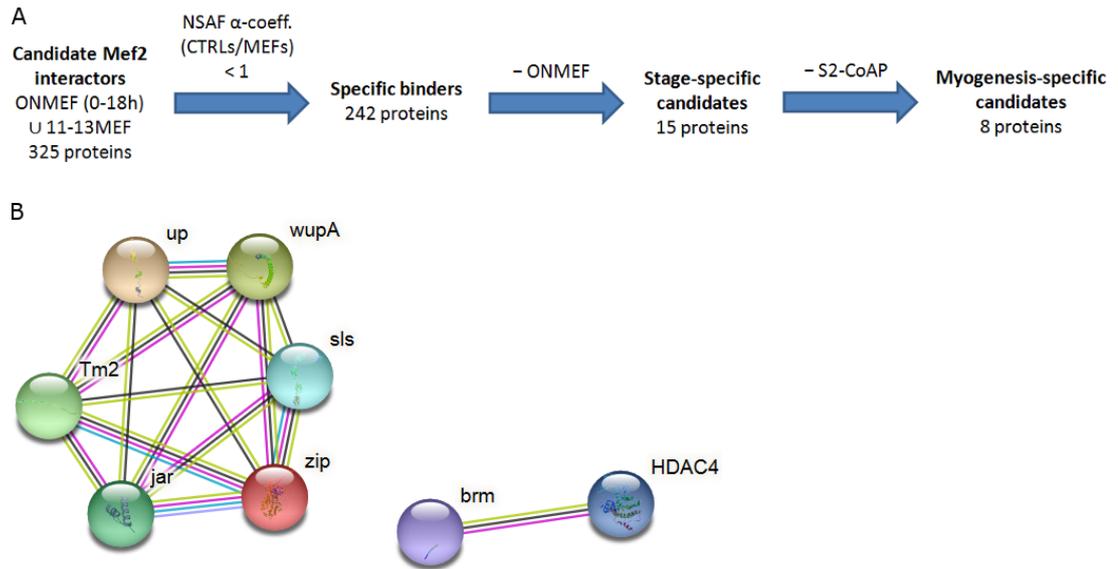


Figure 4.1: Determining the best candidate for biological testing

(A) Workflow to narrow down candidate list to identify proteins that interact with Mef2 specifically during myogenesis.

(B) STRING database network for 8 myogenesis-specific candidates. Connections between proteins include experimentally validated physical and genetic interactions, but also predicted relationships inferred from orthologues.

Table 4.2: Shortlist of candidates for Mef2 regulation during myogenesis

Gene/Protein	MuscleRNAi phenotype	Gene Ontology Cellular component	Gene Ontology Molecular function
wupA/ Troponin I	Lethal (rounded muscle)	Cytoplasm (striated muscle thin filament)/nucleus	Actin binding ATP binding
sls/Titin	Lethal (spotty Z)	Cytoplasm (striated muscle myosin thick filament)/nucleus	Actin binding Protein binding Structural constituent of muscle
Tm2/ Tropomyosin2	Lethal (spotty Z)	Cytoplasm (actin binding)	Actin filament binding
zip/Myosin II	Lethal	Cytoplasm	Myosin light chain binding Protein binding
jar/Myosin VI	Lethal	Cytoplasm	Calmodulin binding Actin binding Myosin light chain binding Protein binding
up/ Troponin T	n.d.	Cytoplasm (striated muscle thin filament)	Calcium ion binding Tropomyosin binding
brm/Brahma	Wildtype	Nucleus	Transcription factor binding Transcription coactivator binding Protein binding
dHDAC4	Wildtype	Cytoplasm/nucleus	Histone deacetylase activity Protein deacetylase activity

target activation in the nucleus. Interestingly, the STRING database predicts a functional link between these two proteins based on co-expression and biochemical data from orthologues. Brahma is a very promising candidate and two other subunits of the Brahma complex were pulled down from the extracts of 11-13h embryos (moira and Bap55). These proteins had been removed from the shortlist because they were also identified as Mef2 interactors in the S2-CoAP study. dHDAC4 was the only candidate out of the 8 proteins that seemed to act independently and not part of a bigger complex. Additionally, in vertebrates, class IIa HDACs have been shown to interact with Mef2 in muscle cells to regulate transcription. HDAC4 was thus selected as a candidate for further study to test whether the function of class II HDACs in myogenesis might be conserved in *Drosophila*.

4.2.5 Overlaps between TAP and published datasets of Mef2 interactors

After the analysis of the TAP results and the other Mef2-interacting candidate datasets by spectral counting, we looked into characterising the likelihood of the proteins interacting with the transcription factor based on network models. While the prioritising approach of the previous analysis is useful to guide future experiments, this second approach should give a more general view on the type of protein complexes Mef2 comes in contact with in *Drosophila*. Most of the datasets of proteins pulled down by Mef2 in different experimental set ups have identified some candidates also found in the other datasets and some candidates that are specific to that particular sample. To identify the overlap between the different datasets and the degree of divergence between lists, a comparison of the candidates based on Venn diagrams was performed. It is important to note that the comparison of dataset was based on the presence or absence of a candidate in each sample and not on a protein's abundance.

The analysed datasets were the staged TAP purifications performed on 11-13h embryos (11-13MEF – data from Mef2-GSTAP embryos, 11-13CTRL – data from wildtype [WT] embryos) and the O/N embryos extracted from over 7 grams of input material (ONMEF - data from Mef2-GSTAP embryos, ONCTRL – data from WT embryos). The protein sets from the samples, respectively controls, have been analysed for any overlap in the proteins they contained (Figure 4.2A-C). All datasets overlapped at least partially with every other dataset. It is noticeable that the proteins recovered in the ONCTRL and 11-13CTRL samples overlapped completely with the data from Mef2-GSTAP embryos, except for 3 proteins that were unique to ONCTRL (Figure 4.2B, top diagram). 49 of the 56 proteins in 11-13MEF were also found in ONMEF.

Almost 1/3 of the proteins in the ONMEF sample were also identified in the Rhee et al. (2014) CoAP/MS study where Mef2 associated proteins were co-purified from S2 cells (S2-CoAP), and 50% of the 11-13MEF candidates overlapped with the S2-CoAP dataset (Figure 4.2D-E). It was assumed that the presence in both datasets would suggest a higher confidence of a protein specifically interacting with Mef2. Therefore, the intersection of the TAP Mef2 samples with the Rhee CoAP/MS derived data was calculated and two new datasets were generated for later use to

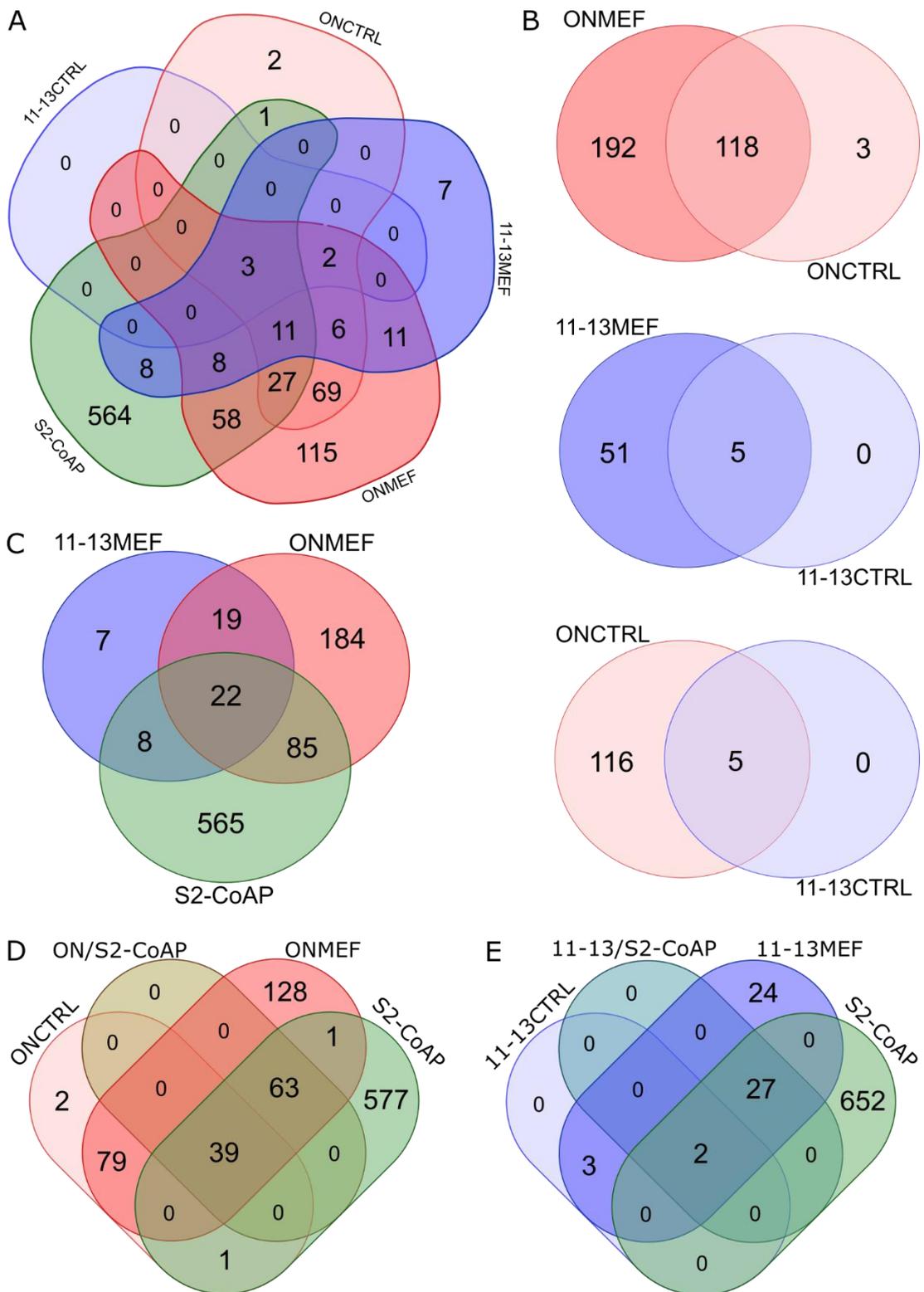


Figure 4.2: Overlap between different datasets of candidates of Mef2-associated proteins obtained by affinity purification studies

Supporting files listing protein IDs in each Venn diagram can be found in the figshare repository (see data availability statement, section 2.3.13).

- A) Overlap between all datasets derived from purification experiments: 11-13CTRL (11-13h staged wildtype embryos), 11-13MEF (11-13h staged Mef2-GSTAP embryos), ONCTRL (overnight wildtype embryos), ONMEF (overnight Mef2-GSTAP embryos), S2-CoAP (FLAG-HA tagged Mef2 overexpressed in S2 cells, (Rhee et al., 2014)).
- B) Overlap of candidate proteins between ONCTRL and ONMEF, respectively 11-13CTRL and 11-13MEF, or ONCTRL and 11-13CTRL
- C) Overlap between the three samples purified from tagged-Mef2 expressing biological systems: ONMEF, 11-13MEF and S2-CoAP
- D) Overlap between the samples purified from overnight collected embryos: ONMEF, ONCTRL, S2-CoAP and ON/S2-Co-AP (intersection between ONMEF and S2-CoAP datasets)
- E) Overlap between the samples purified from staged collected embryos: 11-13MEF, 11-13CTRL, S2-CoAP and 11-13/S2-CoAP (intersection between 11-13MEF and S2-CoAP datasets)

generate network models. These overlap datasets are denoted ON/S2-CoAP (overlap between ONMEF and S2-CoAP) and 11-13/S2-CoAP (overlap between 11-13MEF and S2-CoAP) respectively and are illustrated in Figure 4.2D & E. The overlap between the 11-13MEF candidates and the DPiM dataset (Guruharsha et al., 2011) is very low since only 8 proteins overlap. The ONMEF has only 15 proteins overlap with the DPiM dataset, it was unlikely that integrating these data would allow more insights into the TAP datasets, and relevant aspects of Mef2 function during development might be obscured rather than reinforced. Hence, this control was not studied any further via the network model approach.

The datasets ONMEF* and 11-13MEF* which contain all the specific binders as identified by the spectral counting method represent 73% of the candidates for ONMEF, respectively 89% of the 11-13MEF dataset. The ONMEF* and 11-13MEF* were compared with the 11-13/S2CoAP and ON/S2CoAP datasets and a number of 18 proteins were found in all the sets. Only 50% of 11-13MEF* were found in 11-13/S2CoAP, while only 80 proteins were common between ONMEF* and ON/S2CoAP.

4.2.6 Connectivity between nodes depends on annotation level of the input proteins

Protein Interaction Networks are a straightforward way to visualise lists of proteins and the connections between them. For the lists of Mef2 interactors each protein-protein interaction network will be prepared starting from the list of candidate proteins and the connections between these proteins will be added based on information derived from a database which compiles information based on experimental, biochemical and genetic data available. In this way of visualising lists of candidates, each protein becomes a dot and if two proteins have a functional connection their dots are link together by a line. If two candidates and Mef2 are known to interact among each other, then our network would have three dots representing the proteins and three lines representing the functional relationships between each pair of proteins. This triangle would be found in the network representation of two separate lists of Mef2 candidate interactors, if these candidates were found in both of the lists. If there are candidates that are found only in a specific list, the dots of these unique candidates are linked with Mef2 only in the

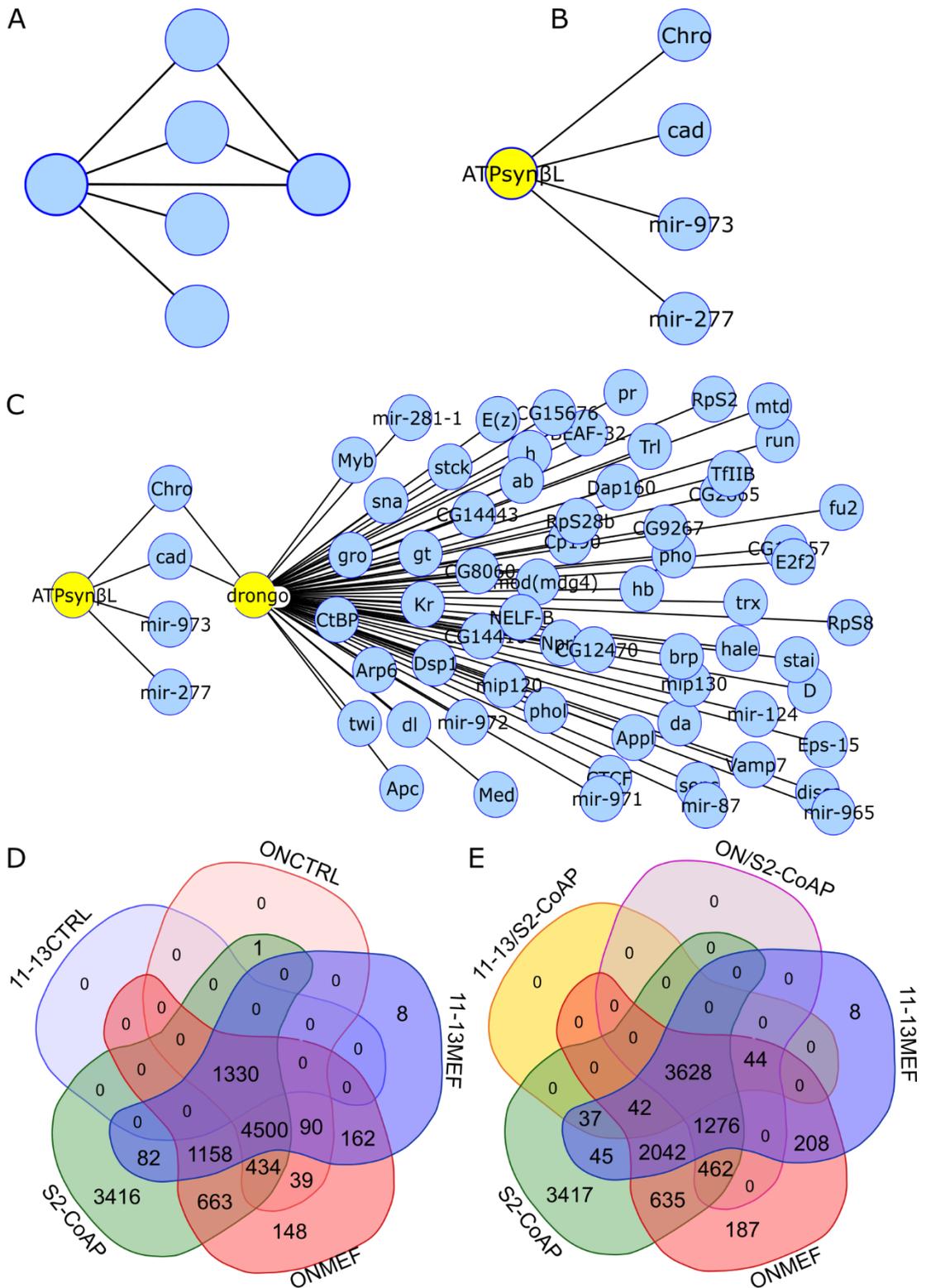


Figure 4.3: Analysis of PPI networks of the datasets generated based on DroID derived data

A) Schematic representation of a protein-protein interaction (PPI) network based on graph theory, where the nodes are proteins and the edges are known associations stored in the DroID database.

B) Representation of how the DroID Cytoscape plugin retrieves the information concerning a given protein used as input. The plugin receives a list of seeds (candidate proteins). For each seed, it retrieves a list of known interaction partners and the binary interactions stored for each seed-partner pair from the DroID database.

C) Schematic how the DroID Cytoscape reading algorithms merges information retrieved for every protein provided as input node (seed). If two seeds (yellow) have a common interaction partner, the seeds will be connected through that intermediate partner (blue). If the DroID database contained a known interaction between two seeds, the two nodes will be directly connected in the PPI network (not illustrated).

D) Overlap of the PPI networks obtained using each labelled dataset as seeds for the DroID plugin: 11-13CTRL (11-13h staged wildtype embryos), 11-13MEF (11-13h staged GSTAP embryos), ONCTRL (O/N wildtype embryos), ONMEF (O/N Mef2-GSTAP embryos), S2-CoAP (Mef2 overexpressed in S2 cells, (Rhee et al., 2014)). Note that the number of identified interaction partners is several folds larger than the number of seeds (compare with Figure 4.2). Supporting files listing protein IDs in each Venn diagram can be found in the figshare repository (see data availability statement, section 2.3.13).

E) Overlap of the PPI networks obtained for ONMEF, 11-13MEF, S2-CoAP, ON/S2Co-AP and 11-13/S2Co-AP. The 11-13MEF* and ONMEF* networks are omitted due to practical constraints of displaying Venn diagrams but are explained in the main text.

network representation of the list it was originally found in. Therefore, overlapping candidates between two datasets have the ability to give a similar topology of the network of the two list, while the unique candidates should give divergence to these networks since new functional relationships are added.

The model used to construct the protein-protein interaction (PPI) network (Figure 4.3A) for each considered dataset was based on the DroID Cytoscape Plugin which is able to retrieve information regarding a set of input proteins (referred further as seeds) from the DroID database. The information retrieved includes binary physical PPIs derived from various yeast two hybrid studies, physical PPIs derived from co-immunoprecipitation studies, literature-curated genetic interactions, transcription factor-gene relationships, miRNA-gene interactions and predicted protein interactions derived from data in human, yeast and worm. Based on the obtained information from the database, the algorithm adds to the list of seed proteins the identified interaction partners and registers all seeds and interaction partners as nodes in a network. Any known association between two proteins is registered in the network as an undirected edge between the two proteins' nodes, regardless of the nature of that association (Figure 4.3B). Thus, if two seeds have a known association, they will be connected by an edge. If two seeds interact with the same non-seed protein, a network path is created between them through this common protein/node (Figure 4.3C).

Each obtained PPI network contains two categories of nodes: nodes derived from seed proteins ("seeds") and nodes derived from interaction partners of seed proteins as identified by DroID ("interaction partners"). The connectivity (number of edges) of a node derived from a seed protein is influenced by the number of different known associations of that protein. The connectivity of a node representing an interaction partner depends on the number of seeds that interact directly with that protein. Some nodes derived from seeds can have low connectivity if they have a low number of known associations; this is typically the case for proteins with little literature background. A node from an interaction partner can have a high connectivity if it is linked with several seed proteins. Proteins from the datasets for which DroID returned no known association in the database are not included in the network. Note that the interaction partners for seeds as identified by DroID were only used to

determine how well-characterised the proteins in the different datasets were and how well-connected the respective PPI networks were (this section and the one following). After this step, all further analysis considered only the seeds, i.e. candidate Mef2-interactors, themselves, and the interactions (connections) between them that were found in the DroID database.

To test if the seed proteins in our datasets were well characterised, the connectivity of nodes derived from them (Figure 4.4A) was compared to the connectivity obtained using random lists of proteins (Figure 4.4B), whereby the random lists had the same number of proteins as the datasets tested. From Figure 4.4, it is observed that the proteins in all the datasets have a high number of neighbours (more than 10 for almost all seeds, mean connectivity above 200 interaction partners per seed). The WT control samples each had a slightly higher mean connectivity than the corresponding sample derived from embryos expressing tagged Mef2. The proteins extracted in the staged samples had a greater connectivity than overnight samples.

The fact that the proteins used as input are well connected signifies that the proteins of interest are characterised and there is enough biological background available to assess the validity of the data obtained using functional relatedness and network modelling.

Using the DroID plugin, one PPI network was generated for each dataset, where two proteins were connected if there was any known biological interaction between them (genetic interaction, PPI, TF-gene relationship, binary interaction derived from yeast two hybrid studies, or predicted from other organism; Figure 4.3B-C). For each dataset, all candidates contained in the dataset were used as seeds to query known interaction partners and generate the respective network. This resulted in a total of 9 networks. The networks are denoted by the same name as the dataset they are derived from (S2-CoAP, ONMEF, 11-13MEF, ONMEF*, 11-13MEF*, ON/S2-CoAP, 11-13/S2-CoAP, ONCTRL, 11-13CTRL). The number of nodes in the experimental datasets' PPI network are illustrated in Figure 4.3D.

Since the derived datasets (the two datasets of non-contaminant Mef2-interactors as determined by the NSAF α coefficient, and the two overlap datasets) are subsets of

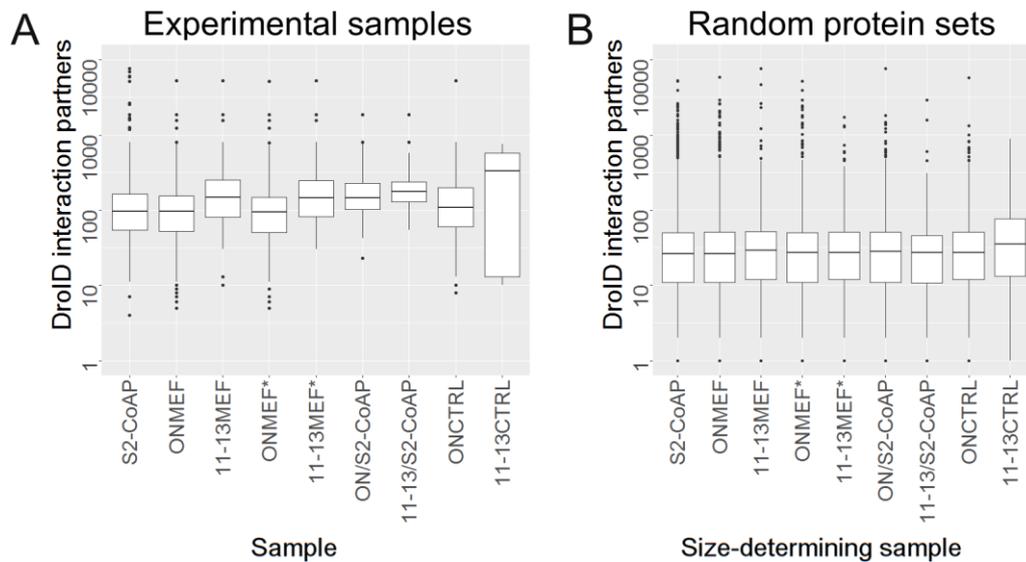


Figure 4.4: Numbers of detected interaction partners in DroID for each candidate protein in the starting datasets

A) Number of interaction partners identified per seed by DroID for each of the 9 datasets used for network analysis.

B) Distribution of the interaction partners identified by DroID using random lists of proteins as seeds. For each dataset, ten lists of proteins chosen randomly from the *Drosophila* genome were generated, where the number of proteins on each list corresponded to the number of proteins in the respective dataset.

ONMEF and 11-13MEF respectively, also the PPI networks derived from them are subnetworks of the ONMEF and 11-13MEF networks. Consequently, neither of the overlap samples have any unique nodes (Figure 4.3E). It is striking to note that even though the overlap between the datasets, i.e. the seed proteins, was minimal (compare Figure 4.2A), the vast majority of nodes in the PPI networks were shared between several networks (Figure 4.3D & E). Correspondingly, most datasets contain a substantial number of unique proteins (Figure 4.2A), but only very few nodes in the PPI networks are unique to any network (Figure 4.3D & E). Those seeds which were shared between datasets necessarily also lead to all of those seeds' interaction partners identified by DroID being shared. However, those seeds which were not shared might have been expected to result in divergent additional interaction partners. This striking overlap suggested that the proteins pulled down in the analysed samples were able to identify, partially overlapping parts of a closely connected larger genetic network.

There were overall 1042 nodes/proteins that were shared between all 9 PPI networks. Another 678 proteins appeared in all networks derived from the 7 Mef2 datasets but not in ONCTRL and 11-13CTRL. This includes both physically interacting proteins, but also proteins otherwise included in the network as implied by the different types of connections that DroID queries (see above). 93% of the proteins in these groups had direct connections with Mef2. Only the S2-CoAP network had a substantial number of nodes that could not be found in the networks generated by any other dataset (Figure 4.3D).

The comparison between networks around different samples of Mef2 interactors allows to assess how distinct the candidates extracted in each case are and if they capture very distinct parts of the Mef2 interactome. The high numbers of known interactions show that the proteins that were co-purified with Mef2 are well characterised proteins. Yet, while there is only partial overlap between different lists of candidates, the overlaps between the networks were very large, indicating that the unique candidates seem to connect to Mef2 via a common set of links. Since these common connections were present in datasets representing different biological contexts, these proteins could be part of a core biological network in which Mef2 is embedded across time and cell types. The large subnetwork that was unique to S2

cell data and not present in the *Drosophila* networks may reflect the different biological context of Mef2 in cultured cells. Similar subnetworks unique to the *Drosophila* sets, though very small, were also present and could reflect different Mef2 contexts during development.

4.2.7 All candidate proteins are closely associated with Mef2

To determine if the seed proteins in each dataset are indeed part of the same network component with Mef2, the network distance from each seed to Mef2 was analysed. For each dataset containing Mef2 as a seed (all sets, except ONCTRL and 11-13CTRL), the shortest path network distance (ND) within the PPI networks generated by DroID of each seed to Mef2 was computed. The highest ND to Mef2 found in any network was 2. ND can therefore take up three distinct values: 0 corresponds to direct interactions between Mef2 and a candidate protein (Figure 4.5A), 1 indicates the presence of one intermediate in the shortest path (Figure 4.5B), 2 indicates two intermediates in the shortest path (Figure 4.5C). Since the ND is determined within the DroID networks, intermediate nodes in this case can be either other seed proteins or interaction partners identified by DroID. Note that for ND=2, at least one of the intermediates must be another seed (Figure 4.5B). In the control datasets (11-13CTRL, ONCTRL) the seed list did not contain Mef2. However, the PPI networks generated by DroID contained Mef2 as an interaction partner of some of the seeds. Therefore the analysis of shortest distance for each candidate to Mef2 was calculated for the controls as well. Figure 4.5G shows the ND for all seeds and datasets.

The first important observation was that all proteins had a path to Mef2 within their respective PPI network, indicating that none of the networks contain any fully separate subnetworks without connections to the rest. This confirms that all proteins identified by mass spectrometry as Mef2 interactors are already known to be directly or indirectly connected to Mef2. Most of the proteins are connected to Mef2 through another intermediate protein (ND=1, Figure 4.5G). In all samples except 11-13MEF and 11-13CTRL, over 70% of the seeds require a different protein as an intermediate to connect to Mef2. Only very few seeds require more than one intermediate to link to Mef2 (ND=2): 2 seeds in S2-CoAP, 3 seeds in 11-13CTRL and 10 seeds in ONCTRL. The ratio between indirect (ND=1) to direct (ND=0) interactors of Mef2

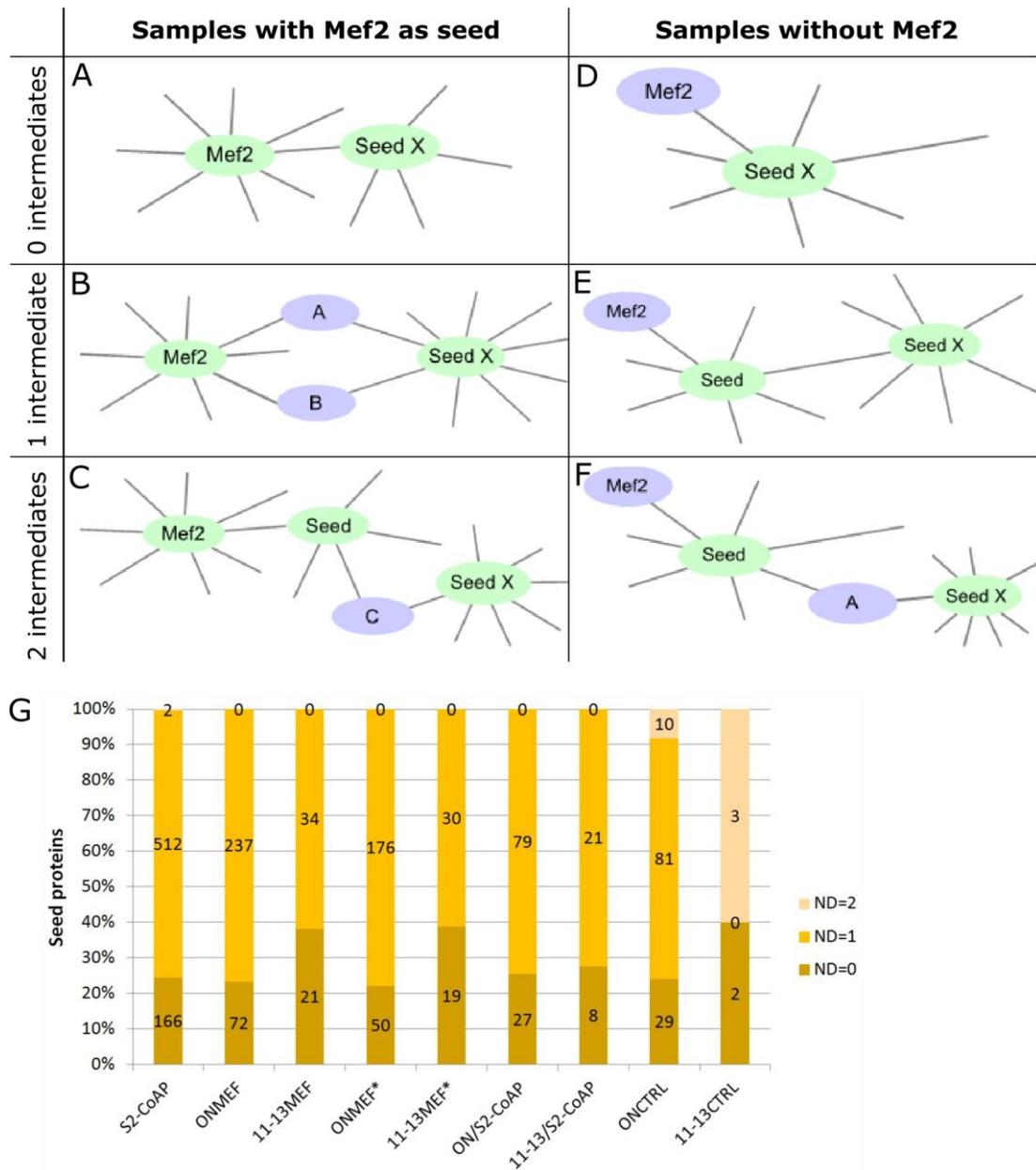


Figure 4.5: Number of intermediate proteins between Mef2 and associated proteins present in each PPI network

Diagrams of network distance between Mef2 and other seed proteins when Mef2 is either a seed or an interaction partner in the PPI. With Mef2 as a seed, each other seed (seed X) can have A) a direct interaction/no intermediate proteins, B) 1 intermediate protein, which can be another seed or an interaction partner identified by DroID, C) 2 intermediate proteins, of which necessarily at least one must be another seed.

When Mef2 is not a seed but was itself identified by DroID as an interaction partner, each seed (seed X) can have D) a direct interaction / no intermediate proteins E) 1

intermediate protein, which must be another seed, F) 2 intermediate proteins, of which at least one must be another seed.

G) Distance of Mef2 to each seed in the PPI networks generated from the 9 datasets. ND, network distance. Network distances larger than 2 are possible in theory but were not found in any of the datasets analysed.

was roughly 1.5 in networks generated from datasets derived from staged *Drosophila* embryos (11-13MEF, 11-13MEF*, 11-13CTRL), while in the overnight samples the ratio is above 3. This suggests that in overnight samples Mef2 is involved in more indirect connections than in the 11-13h staged context.

The networks created to analyse the relationship between candidates in the datasets and Mef2 consist of both Mef2 interactor candidates identified via protein purification and other proteins that are predicted to be part of the PPI network since they have direct functional connections with candidates. These common “intermediate” proteins allow to create a path within the network between any two candidates considered. Many proteins have a direct link to Mef2, but the majority require an intermediate protein to connect to it. However, only the CTRL and the S2-CoAP networks have third-order candidates (2 proteins distant from Mef2), and none are further apart than this, indicating that all networks are densely centred around Mef2.

The identification of both direct and indirect links to Mef2 suggests that all the datasets have identified proteins that are either direct interactors, or part of a complex that works together with Mef2. This is most striking for the CTRL datasets, since these would normally be assumed to be contaminants that should not interact with Mef2. Since all networks form a single connected component it can be assumed that these networks are densely packed with connections and that focusing only on the connectivity among seed proteins will highlight which complexes they are part of in their interaction with Mef2.

4.2.8 Connectivity between candidate proteins

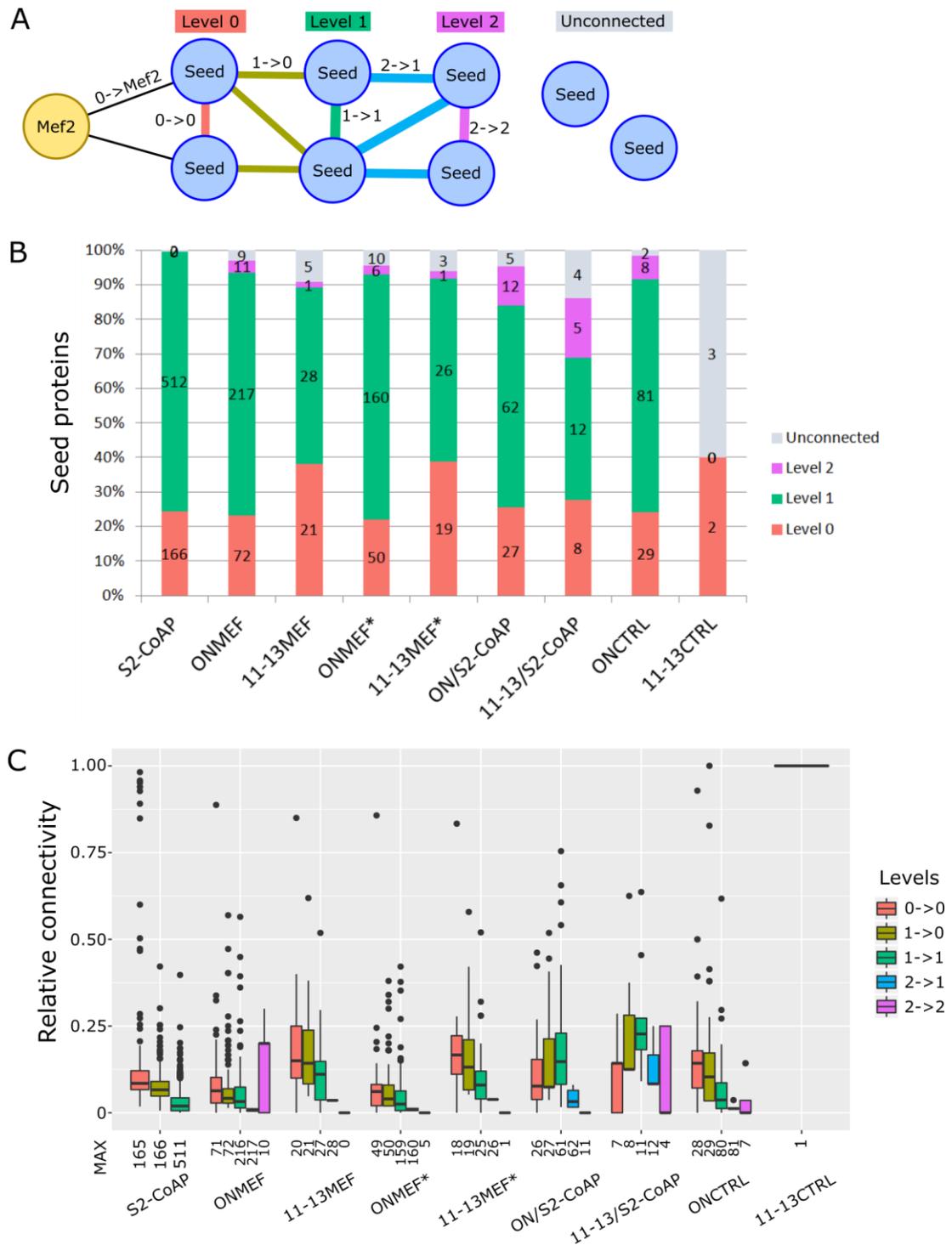
The seeds were found to form either direct or indirect connections to Mef2 in the analysed networks. The connectivity of the seeds with regards to other seeds and Mef2 was investigated in order to determine if they form a connected subnetwork. This analysis is similar to the previous step analysing network distances in the concept of connectivity, but it selectively analyses seeds and disregards all interaction partners added by DroID to test whether seeds are well-connected between themselves. To achieve this purpose, all non-seed interaction partners were removed from the 9 PPI networks, leaving behind only the seed nodes and the

connections between them. The seeds were then classified into different levels: 0 level seeds had direct connections with Mef2, 1st level proteins interact with Mef2 through 0 level seeds, 2nd level proteins need two other seeds to interact with Mef2 (seeds from levels 0 and 1; Figure 4.6A). Next, the connectivity (number of edges) between all seeds on each respective level was computed (within-level connectivity). Subsequently, the connectivity between each two levels was computed, i.e. the number of edges from seeds on the upper level to seeds on the level below (between-level connectivity).

For all 9 datasets except 11-13CTRL, most of the seeds had a direct or indirect connection to Mef2 through other seeds (Figure 4.6B). Indeed, the overall distances of seeds from Mef2 when only connections between seeds are taken into account are quite similar to the distances in the full networks that include all interaction partners identified by DroID (compare Figure 4.6B with Figure 4.5G). However, in the seeds-only subnetworks, a small number of seeds are not connected to Mef2 at all (Figure 4.6B, grey bars), indicating that these seeds depended on a non-seed interaction partner to be connected to the full PPI network. These proteins do not form a separate connected group among themselves, but are isolated both from the seeds subnetwork and the other seeds in this category.

The great majority of the seeds are either direct interactors of Mef2 or on level 1, followed by level 0 proteins. The number of level 2 proteins is very low in all the samples. The low number of unconnected and level 2 seeds strengthens the previous finding that the candidates contained in the datasets are part of a highly connected network with many known interactions.

Additionally to their distance to Mef2, the connectivity between the seeds classified as level 0, 1 and 2 was studied in all the datasets (Figure 4.6C). This was done to compare the connectivity of seeds with Mef2 to their connectivity with each other, as a way of assessing how central Mef2 is within the network. The connectivity in this case is the number of connections within or between the respective levels, normalised by the number of potential connections (Figure 4.6C, labelled as "MAX"), which depends on the number of seeds on the respective level (compare to Figure 4.6B). This normalisation is necessary because large networks naturally are likely to contain a larger absolute number of connections, and the sizes of the



if two other seeds are required to connect to Mef2. The seeds on the same levels or adjacent levels can be interconnected.

B) Distribution of seeds on levels in each dataset.

C) Relative connectivity between seeds within each level and between levels. Connectivity is the number of connections of each seed divided by the maximum possible connections, which in turn depends on the number of seeds on the same level (respectively the level below). MAX, maximum potential connections per seed

different subnetworks diverge substantially. Comparing absolute numbers of connections would thus not provide much information except that one subnetwork is larger than the other.

In all the samples that have level 2 seeds, these proteins have reduced connectivity among themselves and with level 1 proteins, making them the least connected seeds. Level 1 and level 0 proteins tend to have the highest connectivity between the seeds. In all datasets except the overlap-sets (S2-CoAP, ONMEF, 11-13MEF, ONMEF*, 11-13MEF* and ONCTRL) the trend of the connectivity between seeds showed the following pattern: direct interactors of Mef2 are the most interconnected seeds among themselves. The connectivity of level 1 proteins to direct interactors was the second-highest, and level 1 proteins were the third-most connected group of seeds. This pattern indicates networks which are centred on Mef2 and gradually become less connected the further away from Mef2 nodes are. The S2-CoAP subnetwork derived from the published study contained only level 0 and level 1 proteins, but their connectivity followed the same pattern.

However, in the overlap datasets ON/S2-CoAP and 11-13/S2-CoAP, the pattern was reversed, the level 1 proteins being the most interconnected while Mef2 direct interactors are the third-most connected group. This indicates the opposite, i.e. networks where nodes further away from Mef2 become gradually more interconnected. This suggests that the candidates that are shared between the TAP purification datasets (ONMEF and 11-13MEF) and the published study (S2-CoAP) are less likely to contain complexes that directly include Mef2 and instead may be enriched for long-range interactors that can be pulled down with Mef2 e.g. through adapter proteins. Furthermore, the consistency of this trend also in ONCTRL indicates that this control sample is also Mef2-centred, even though Mef2 was only added to the network as an interaction partner identified by DroID. This suggests that the control samples primarily contain Mef2 interaction partners of interest, even though the proteins in these samples were purified due to unspecific binding to the purification columns and not due to specific binding to Mef2. Consistent with this, 11-13MEF* and ONMEF* ONMEF*, which were filtered as likely specific binders based on the CTRL datasets, have a very similar behaviour as the networks that still contained seeds categorised as potential contaminants (as determined in section

4.2.2). This means that the potential contaminants are not distinguishable from specific binders based on their connectivity characteristics both towards Mef2 and between each other, which argues against the assumption that they are contaminants.

The previously identified trend of more connectivity with Mef2 in staged samples (11-13MEF and 11-13MEF*; Figure 4.5G) is also maintained when intermediate interactor proteins are removed from the networks and only the connectivity of seeds with Mef2 and between seeds is taken into account. These samples had a larger fraction of seeds on level 0 (direct connection to Mef2; Figure 4.6B), and their level 0 seeds also had nearly twice the relative connectivity between each other compared to level 0 seeds in other datasets (Figure 4.6C). This further increases the confidence that these datasets contain complexes that directly interact with Mef2.

Overall the samples display a strong connectivity which seems to indicate the extracted candidates are part of well characterised networks that are able to indicate the biological processes that Mef2 is involved in during *Drosophila* development. The characteristic that stands out from the connectivity among seeds in all these networks is that there is an intense crosstalk between different potential candidates and it is connectivity information alone does not make it possible to extract information about certain proteins that could form distinct protein complexes. Together with the previous observation that the extended networks for each dataset form roughly the same connected component suggests that Mef2 is involved in an intricate regulatory network that is present around it in most contexts. Some of these candidates were also identified in control experiments and these proteins are often regarded as contaminants in purification experiments and are discarded from further analysis. Interestingly, when looking at their connectivity, Mef2 is one potential protein that links these together as an intermediate protein. Also, most unique candidates were introduced in the network of another list as intermediate proteins by the DROID algorithm. This suggests that these unique candidates are likely present in the other biological contexts where they haven't been identified by MS, but they were not pulled down due to dynamics in concentrations or differences in experimental conditions which are able to preserve better certain types of interactions compared to others.

The fact that the proteins are very well interconnected makes further bioinformatic analyses possible, and suggests that the candidates can form modules that cluster together to perform a function. It is impossible to estimate from a network model if these networks represent protein complexes that interact with Mef2 at the same time or physically, but these analyses can provide important directions for further experiments.

4.2.9 Topological evaluation of protein-protein interaction networks

The final aim of the network analysis was to use a network cluster detection algorithm to identify cohesive subgroups within a network of all available candidates. If such groups are identified, these could be Mef2-interacting protein complexes. However, the clustering algorithms could be impacted by the presence of contaminants, since these could interconnect otherwise discrete subgroups of the full network. The previous analysis of interconnectivity in DroID-generated PPI networks suggested that the control samples purified from WT embryos (ONCTRL and 11-13CTRL) contain a substantial portion of potential Mef2-interacting candidate proteins and it would thus be preferable to retain them for further analysis to identify potential Mef2-interacting complexes. We thus wanted to assess how the seeds in the CTRL datasets and the connections between them influence the networks of non-CTRL datasets. Seeds from all datasets were pooled for this purpose, resulting in the "Mef2 candidates" dataset containing 888 candidate proteins (not counting Mef2 itself).

As previously, known interactions / connections between seeds were extracted from the DroID database, resulting in a large network containing one node for each seed and edges between nodes according to the PPIs stored in DroID. Two networks were derived from this. First, control connections were identified, i.e. connections between two seeds, where both seeds were present in either of the control datasets (ONCTRL or 11-13CTRL). The first derived network ("Without CTRL edges") was created by removing all control connections from the large network. The second derived network ("CTRL only") was created by removing all connections except those where at least one of the two seeds was present in a control dataset. This was intended to test whether the seeds from control samples connect into a network of different

character than seeds not present in controls. The resulting networks are characterised in Table 4.3.

The control connections that were removed to create the first derived network constituted 26.5% of all edges. Along with these edges, 13.6% of seeds were eliminated (becoming entirely unconnected to the main network). Even though such a large number of edges was removed, the network metrics were only marginally different, meaning that the overall layout of the network was conserved. In contrast, the network composed only of CTRL-adjacent edges contained only 26.7% of the original edges. Despite this drastic reduction, the network retained 88.6% of all seeds, a similar amount of seeds as the "without CTRL" network. The characteristics of this network were markedly different, reflecting a substantially less connected network. Besides the reduced density and increased heterogeneity, nodes in this network had substantially fewer neighbours (also reflected by the reduced number of nodes with higher degrees). The network was also less centralised, reflected in reduced clustering coefficients, a higher diameter and radius, and higher shortest path lengths. These differences suggest that the control-adjacent network does not conserve the centralisation on Mef2 that is typical of the other PPI networks. In contrast, removing control-only edges did not disturb this centralisation, suggesting that control-only edges were peripheral in the network relative to Mef2. Overall, these results suggest that the presence of seeds that are also contained in control datasets, or the PPIs associated with them, do not substantially impact the network and thus it should not be necessary to remove either of them for the downstream analysis. In the interest of avoiding false negatives (excluding proteins that are genuine Mef2-interacting proteins), the complete network was therefore used in the following step (Figure 4.7A).

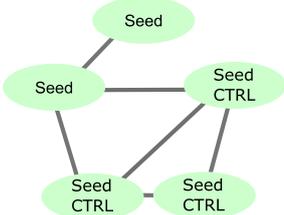
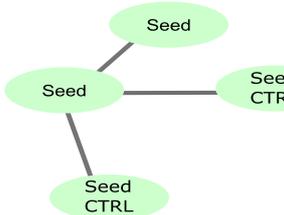
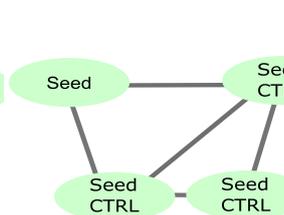
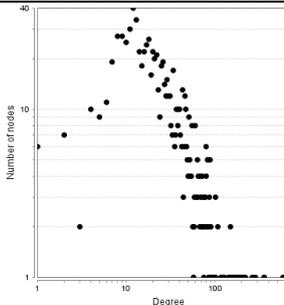
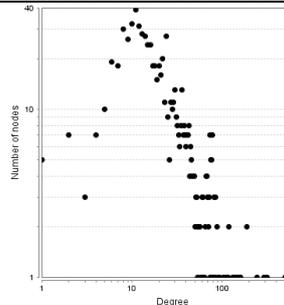
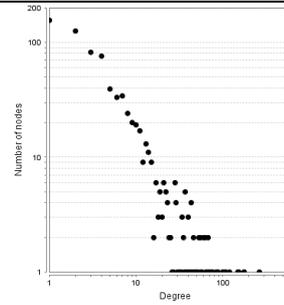
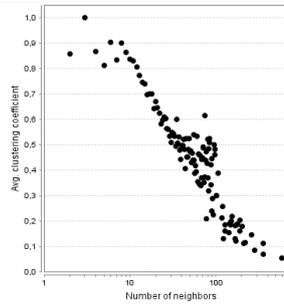
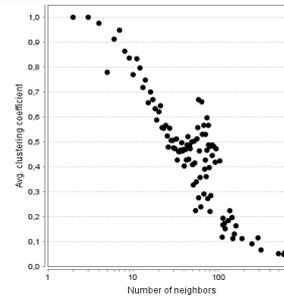
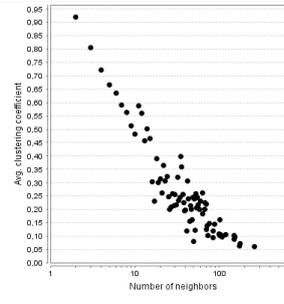
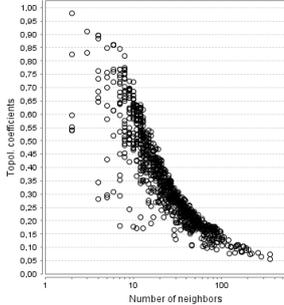
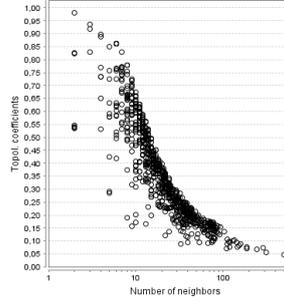
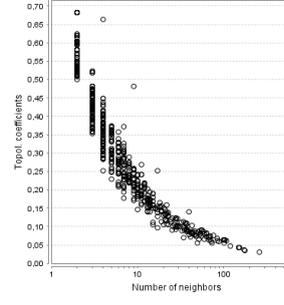
Based on the "guilt by association" paradigm, an identification of potential cohesive subunits of the main network was attempted. In principle, highly-connected subnetworks could reflect biologically relevant groups of genes, such as regulatory networks that act together at particular times or in particular biological contexts. To identify potential subnetworks, the MCODE plugin was used in Cytoscape to extract clustered groups of nodes. Running MCODE on the complete network of Mef2

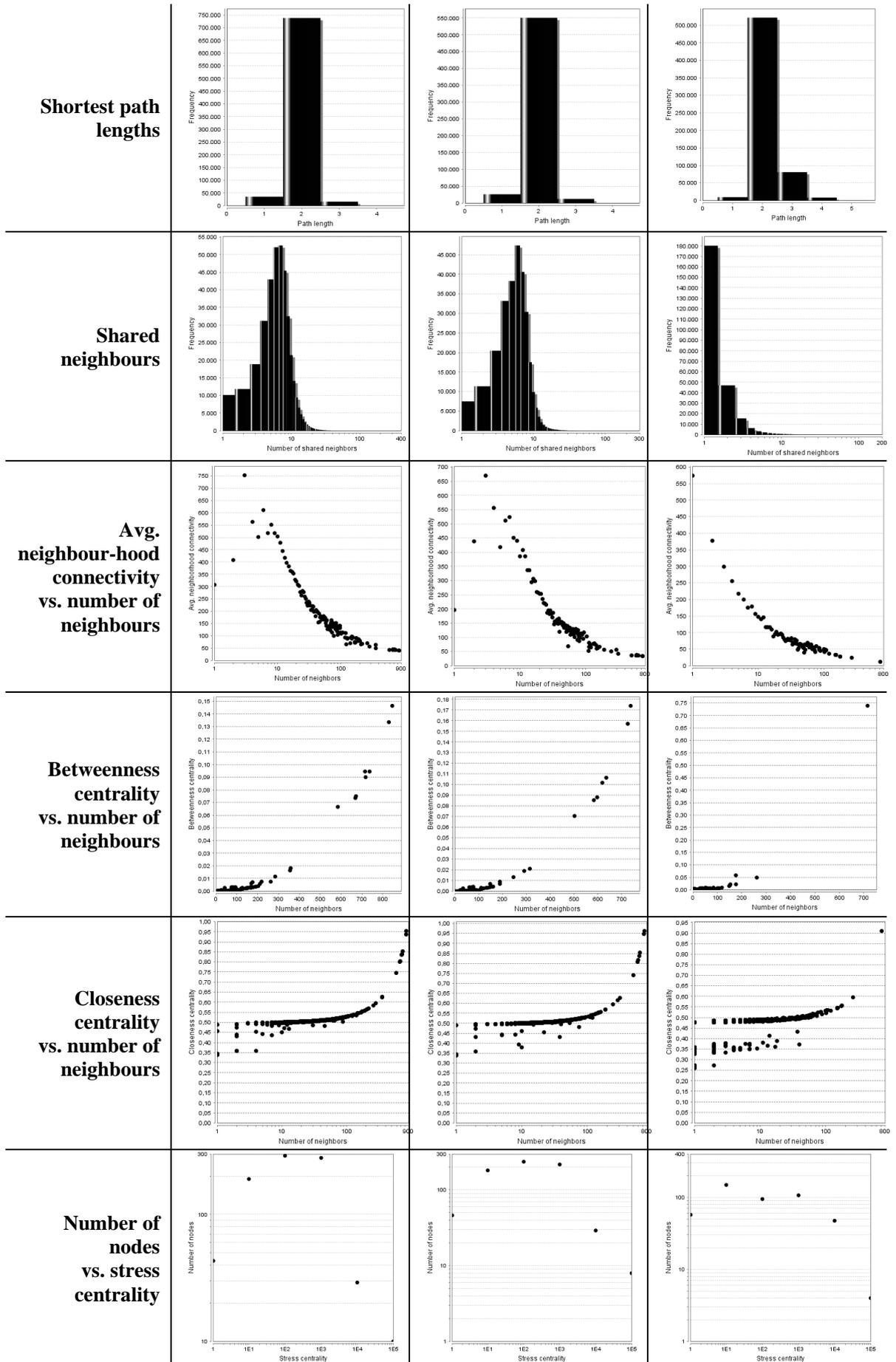
candidates identified 10 clusters (Figure 4.7B). However, without further analysis, it was not possible to interpret whether any of these clusters were biologically relevant.

To test whether proteins contained in each cluster were associated with coherent biological contexts or functions, a gene ontology (GO) enrichment analysis was carried out for each cluster using the ClueGO plugin for Cytoscape. This showed that the clusters were diverse in functions (representative GO network of the first MCODE cluster in Figure 4.7C), and they were therefore not analysed further. To test the previous hypothesis that removing control-only edges had no significant effect on the functional characteristics of the network, the same analysis using MCODE and ClueGO was carried out for the two networks derived earlier ("without CTRL" and "CTRL only"). The same cluster as shown in Figure 4.7C was found in all cases. However, no other clusters or their functionality were maintained. Therefore, even though the network characteristics were maintained when removing CTRL edges, the functional assessment of the network was impacted. This corroborated the previous decision to retain seeds from control samples and their edges, as removing either of them changed the functional characteristics of the network. Furthermore, the cluster that was retained was characterised as related to ribosome and translation. Proteins of this category of function are often considered contaminants. Therefore, the method of adjusting for control samples by removing control-only edges as tested here also did not remove clusters in the network that could be considered potential contaminants.

In conclusion, the PPI network approach was not able to identify functionally coherent subgroups of candidate proteins, and other approaches to achieve this were therefore explored in the next sections. However, this analysis highlighted that the assumption that proteins in the control datasets are contaminants simplifies the issue. Thus, while the control datasets could be used to narrow down the candidate list for the identification of a single best candidate (as done in section 4.2.4), on the systems level the dynamics of biological conditions seem to be reflected in the linking of candidates and not in the presence or absence of candidates. Therefore, when assessing if candidates are able to interact with Mef2 it is important to look at the types of associations that occur and not in removing the proteins that fit all criteria of a "perfect" interactor.

Table 4.3 Characteristics of networks based on seeds from all datasets, filtered or not by CTRL edges

	All seeds & edges	Without CTRL edges	CTRL edges
Schematic			
Seeds / Nodes	889	768	788
Edges	17756	13056	4700
Isolated seeds	0	0	0
Diameter	4	4	5
Radius	2	2	3
Density	0.045	0.044	0.015
Heterogeneity	1.861	1.9	2.794
Centralisation	0.909	0.919	0.896
Avg. neighbors	39.946	34.0	11.929
Number of nodes (y) vs. degree (x)			
Avg. clustering coefficient vs number of neighbours			
Topol coefficients vs. number of neighbours			



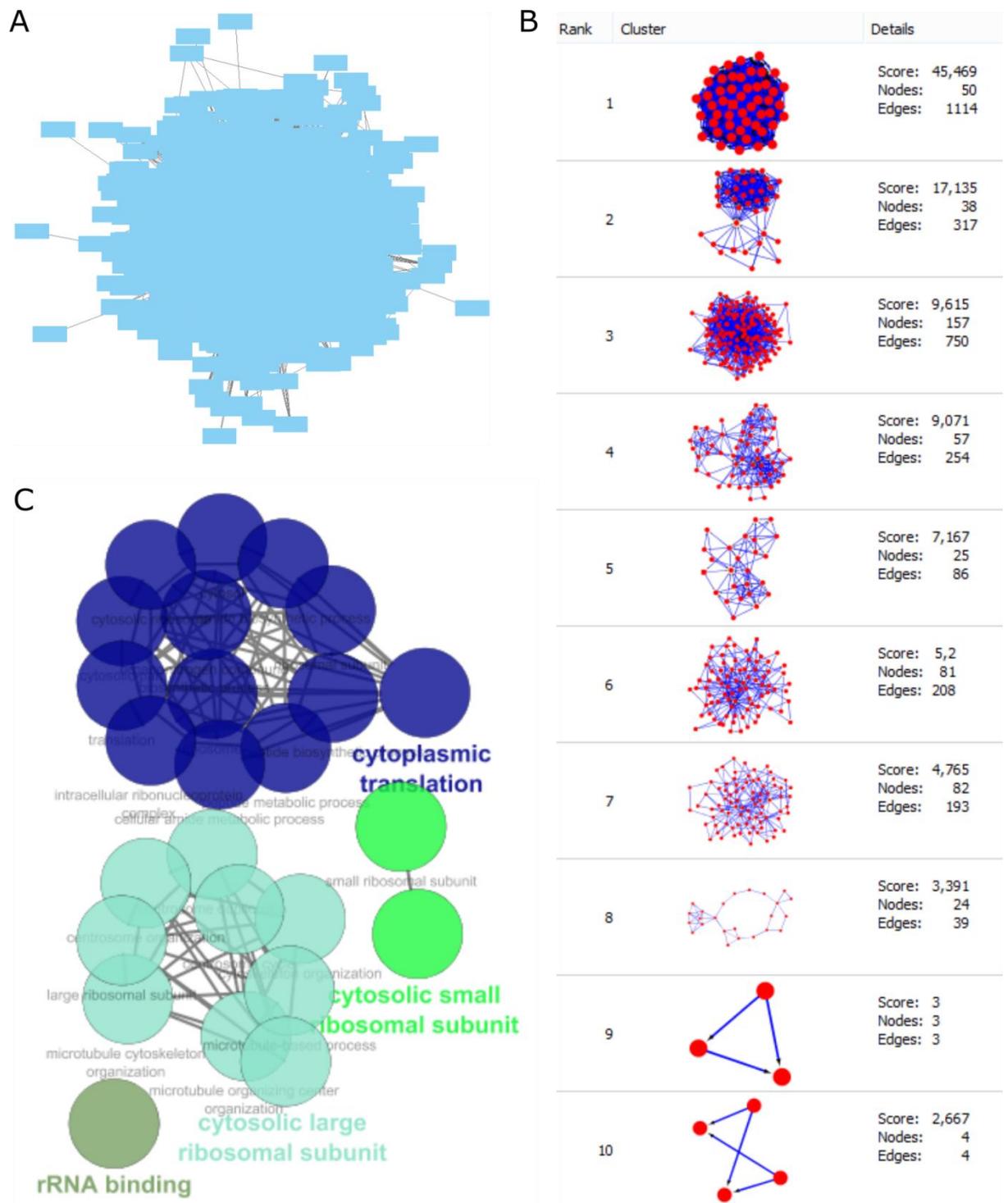


Figure 4.7: Topological characterisation of protein network generated from a unified seed list and interactions between seeds stored in DROID

A) Illustration of the complete network including all seeds from all samples as nodes and edges representing interactions derived from the DroID database

B) Clusters identified by the MCODE clustering algorithm in the network displayed in (A)

C) Gene ontology terms enriched in the most significant and interconnected cluster identified by MCODE in the network depicted in (A), as determined by ClueGO.

4.2.10 Functional analysis based on gene expression

The network analysis can show if there is experimental data available to suggest that the candidates can act in similar biological contexts. To test whether the candidates had functional and contextual connections with each other we looked at the expression data of each candidate in *Drosophila* embryos from modENCODE and used this information to assess the likelihood of coexpression and functional similarity (Graveley et al., 2011). To study these two parameters we took advantage of the percent max (pmax) filter, which allows to estimate the likelihood of genes being expressed and active in a particular context (Murali et al., 2014). The list of candidates included all proteins identified in all datasets even if they were also identified in control samples.

This analysis assumes that a gene is more likely to be functional at a time point when its expression is maximal. The pmax index is a relative measure of RNA expression of a particular gene at a particular stage during embryonic development, whereby the stage with the highest expression has $pmax = 1$ and all other stages have a pmax according to the fraction of RNA expression at that stage. The closer a gene's pmax is to 1, the more likely it is the gene product is functionally active at that stage. Many proteins that interact with each other are co-expressed in time and space (Murali et al., 2014). Therefore their pmax values should correlate across development and should be closer to 1 at stages when they are active. To test whether any of the candidate proteins identified in this study were co-expressed during *Drosophila* embryonic development, the pmax value was calculated for all candidate genes based on RNA expression values obtained from the modENCODE database (Graveley et al., 2011) and a hierarchical clustering algorithm was used to estimate the coexpression of genes (Gould, 2012). This showed that the majority of candidate genes had their highest expression during early embryogenesis (Figure 4.8A). However, a relatively small cluster of genes was co-expressed during the later stages of embryogenesis (Figure 4.8B), showing a similar pattern as Mef2, whose expression levels in muscles increase towards the late stages of embryogenesis. This gene group was associated with muscle-related GO terms, specifically sarcomere organisation (Figure 4.8C), and may therefore constitute a potential gene network that interacts with Mef2 during muscle development.

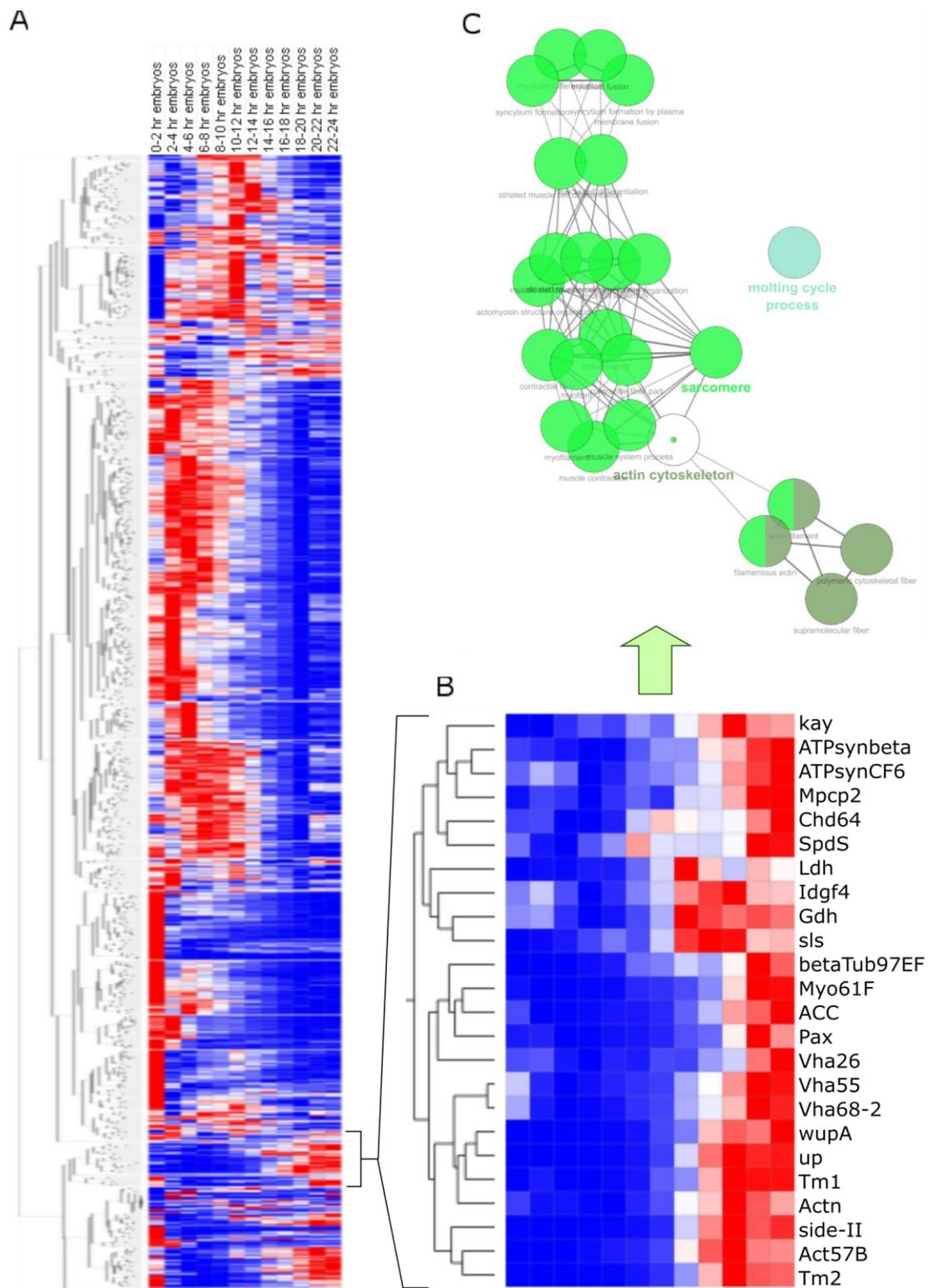


Figure 4.8: Expression distribution of the protein seeds based on the expression filter pmax

A) Heatmap of the expression distribution of the seeds based on relative expression at different time points during embryogenesis. Colour indicates the pmax value, i.e.

RNA expression level relative to the gene's expression peak, at each embryonic stage. Colour scale: blue: $p_{max} = 0$, red: $p_{max} = 1$. Each line in the heatmap corresponds to one gene. The genes have been arranged using a hierarchical clustering algorithm. Supporting file with the complete p_{max} matrix can be found in the figshare repository (see data availability statement, section 2.3.13).

B) Identification of a cluster of genes expressed with a similar relative expression in late embryos that are involved sarcomere maintenance. The relevant portion of the dendrogram produced by hierarchical clustering is shown on the left, illustrating the similarity of these genes' p_{max} distribution.

C) GO terms enriched in the cluster identified in (B), determined by ClueGO in Cytoscape.

Some of these genes were also among the candidates identified as myogenesis-specific Mef2 interactors in the semiquantitative analysis of section 4.2.4 (Tm2, wupA, up, sls) and have been tested for roles in muscle in a specific RNAi screen (Schnorrer et al, 2010). The group identified here is more extensive, however, and includes further cytoskeletal proteins such as Actin57B, alpha-Actinin, beta-Tubulin and Myosin 61F, as well as proteins involved in myoblast fusion such as Lactate dehydrogenase and Paxillin. Therefore, using gene expression data and hierarchical clustering it is possible to identify groups of proteins that are co-expressed and are highly likely to act together with Mef2 in specific biological processes.

4.2.11 Functional analysis of individual datasets using gene ontology

While pmax as analysed above provides information on gene expression, gene ontology (GO) terms reflect known aspects of a gene's function. Mef2 is annotated with an unusually large number of GO terms, reflecting a diversity of biological functions (Figure 4.9). The biological roles of Mef2 are assumed to reflect also the functions of proteins it interacts with to perform these functions. To try to classify the candidates based on the functions they might have in connection with Mef2, they were checked against Mef2 specific GO terms.

GO enrichment was calculated for each dataset and the enrichment p-values for the various GO terms enriched in each sample were collated into a single table for comparison. Another dimension of resolution was added to the analysis by calculating GO enrichment in each case for the whole dataset (ALL), the subset of seeds with direct interaction to Mef2 (Level 0), indirect interaction through one other seed (Level 1) or two (Level 2). This refers to levels as in section 4.2.8 where the interconnectivity between seeds relative to Mef2 was analysed. A hierarchical clustering algorithm was used to cluster the GO terms according to their similarity of enrichment across datasets. The resulting list of terms which are enriched in one or more samples contains a total of 1729 terms and can therefore not be displayed in a practical manner. The list was therefore narrowed down to only those terms which are also annotated for Mef2 itself (Figure 4.10).

From the list, clusters which were enriched specifically in the 11-13MEF sample were isolated and the proteins which contributed to those GO term clusters identified.

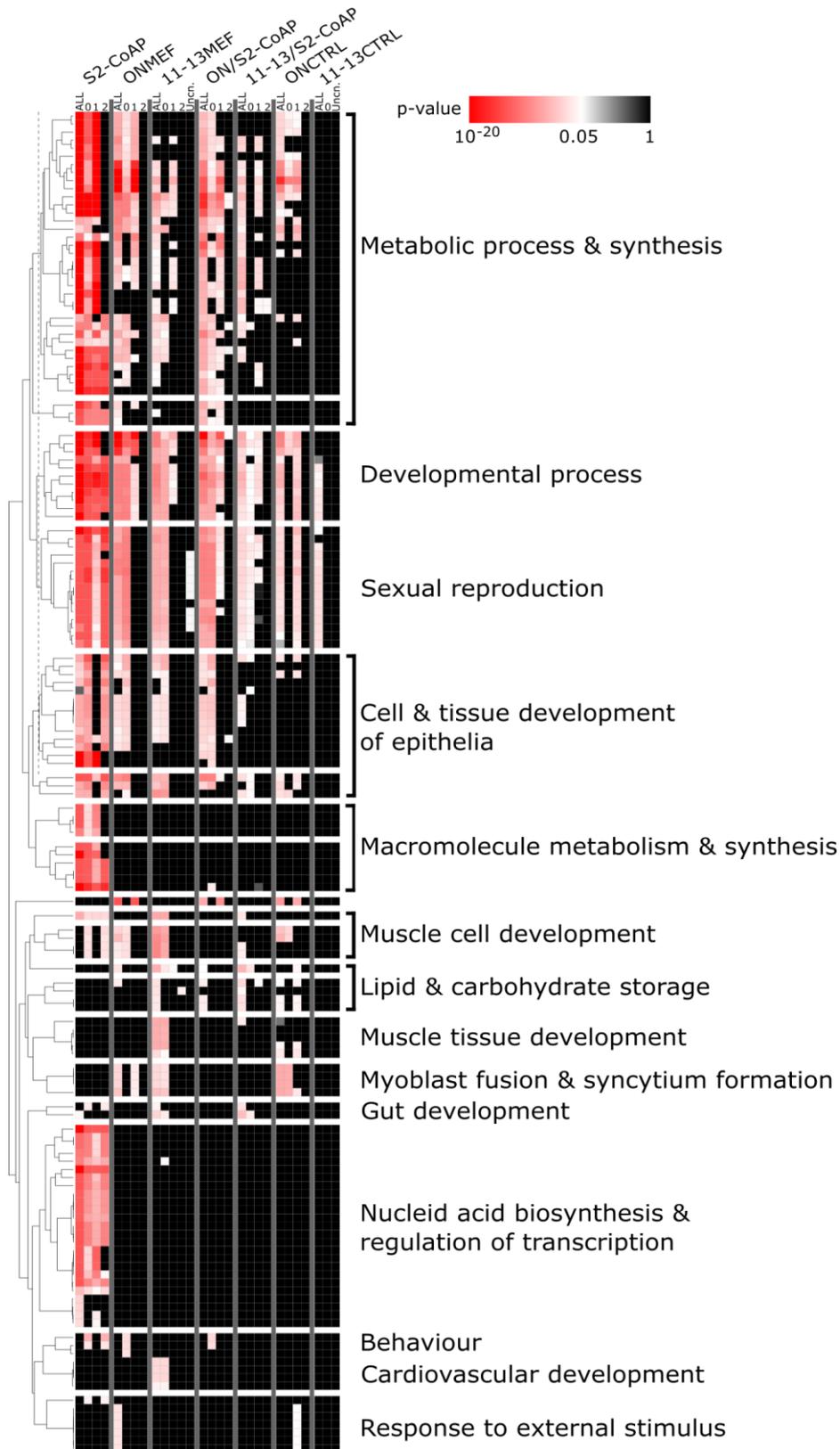


Figure 4.10: Mef2 GO terms enriched in one or more dataset

The terms have been arranged according to similarity in their enrichment across different datasets (clustering dendrogram on the left). Each line in the data matrix

represents one GO term and each column represents the enrichment of each term in one set of seeds. The colour of each cell represents the enrichment p-value of the term in that dataset. GO enrichment was calculated for each sample for the whole dataset (ALL), the subset of seeds with direct interaction to Mef2 (0), indirect interaction through one other seed (1) or two (2) according to interactions found in the DroID database (see section 4.2.8). Individual terms have been manually summarised into broader categories for better displayability. Supporting files with the complete list of GO terms and GO enrichment matrix can be found in the figshare repository (see data availability statement, section 2.3.13).

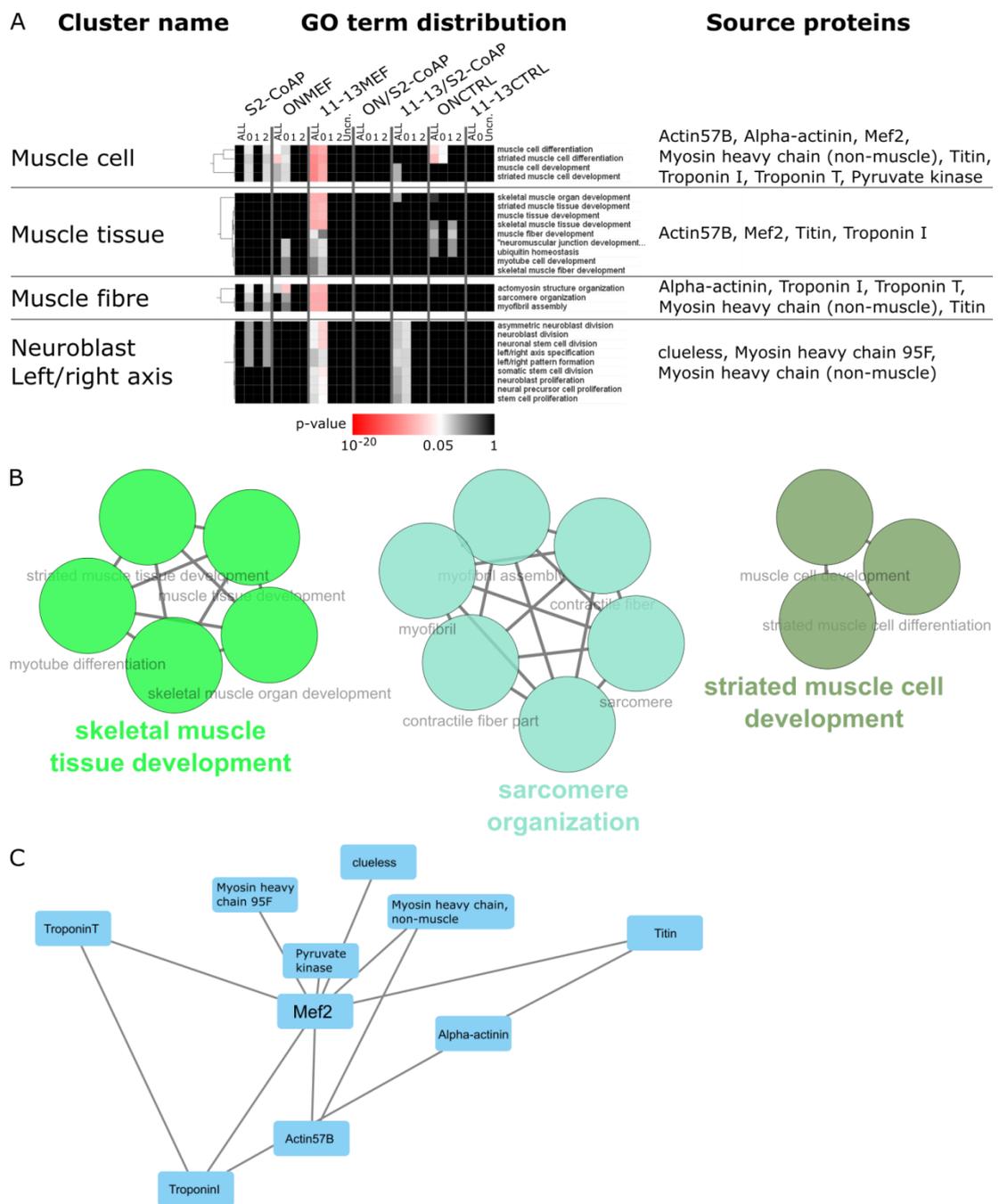


Figure 4.11: 11-13MEF specific GO term clusters and their constituent proteins

A) Higher-zoom representation of the GO term clusters of interest (specifically enriched in 11-13MEF and the subset of seeds with direct interaction to Mef2) identified from Figure 4.9. Proteins identified to be causing the enrichment of these terms. Cluster names were assigned manually.

B) Graph of all GO terms enriched in the proteins from (A).

C) Protein-protein interaction network of the proteins from (A) derived from connections stored in the DroID database.

All clusters came from a mostly overlapping list of 10 proteins (Figure 4.11A). To assess whether these genes generally have shared functions outside of the context of the present datasets, GO enrichment was calculated for the 10 identified proteins. As seen from the GO terms enriched in these proteins (Figure 4.11B), these proteins are mostly muscle related. Additionally, the DroID-based interactions associated with this group of genes was retrieved (Figure 4.11C). This showed that most proteins in this group, additionally to their interaction with Mef2, already have known associations with others from the group. The co-purification of these proteins with Mef2 from 11-13h staged Mef2-GSTAP expressing *Drosophila* embryos suggests that additionally to the known genetic relationships, these proteins may physically interact with each other. Several proteins of this group are also present in the co-expressed cluster identified using the pmax expression filter (section 4.2.9), specifically Troponin I (*wupA*), Troponin T (*up/wupB*), Titin (*sls*) and α -Actinin (*actn*) (compare Figure 4.11C and Figure 4.8B).

Taken together, sections 4.2.9 and 4.2.10 show that using network topology information, gene expression data and gene ontology terminology can be integrated to break down the highly interconnected candidate network of Mef2 to extract protein networks with cohesive expression patterns and functional annotations, suggesting that they interact with Mef2 in a specific biological context.

4.2.12 Functional analysis of all candidates using functional similarity

The previous approach to identify candidates of interest based on GO annotation was successful, but could not be applied to the whole dataset in this form. Besides being very time-consuming, the complexity of this analysis is likely to result in human error and bias. To evaluate GO annotations on a large scale for the complete network of Mef2 candidates, the functional similarity score (FunSim) was therefore used. This is a pairwise score where for any two proteins, a semantic measure of similarity between the GO terms associated with them can be calculated. This eliminates the earlier problem encountered with GO enrichment analysis that the large number of GO terms makes any further analysis impracticable. The FunSim score integrates all GO annotations and returns a measure of similarity, which can then be used to identify groups of proteins with similar existing functional annotations.

A FunSim matrix was computed for all pairwise relationships between all 889 Mef2 candidates using FunSimMat (Schlicker and Albrecht, 2008, 2010). Based on these scores, a FunSim network was created where an edge was added between two seed proteins if $\text{FunSim} > 0.7$. If a seed did not have any edges by this criterion, the seed was eliminated from the network (Figure 4.12A). The primary finding of this analysis was that most candidates formed a tight cluster with Mef2, indicating high FunSim scores and therefore much overlap in their GO annotations.

In the previous section, 10 muscle-related proteins were identified based on their GO-term specificity for 11-13MEF and their known associations stored in DroID. Despite these correlations, in a network representation based on functional similarity as this one, these proteins were quite distant from Mef2. This reflects that these proteins are specific and similar to each other in their functional annotation, whereas Mef2 shares part of its function with them but simultaneously has more diverse unrelated functions, leading to a relatively low FunSim score between Mef2 and the identified muscle proteins.

Three quite distant clusters were observable in the functional similarity network. The three clusters were related functionally to translation initiation (Figure 4.12B), cytosolic translation with the help of ribosomes (Figure 4.12C) and proteasomal degradation (Figure 4.12D). Proteins that contribute to these functions are typically ubiquitous and highly expressed, making them commonly found contaminants in protein purification studies (Zanon et al., 2013).

While the previous approach using only the GO terms for biological process (section 4.2.10) was able to identify some proteins with associations to muscle, it seems that when comparing all GO terms including those for cellular component and molecular function for each pair of candidates, the most clearly distinguishable protein groups are related to housekeeping functions. According to the "guilt by association" interpretation, these protein clusters should be unlikely to be found in similar biological contexts as Mef2 and thus could also constitute contaminants in the datasets where they were originally found. Overall however, the tight clustering of the vast majority of candidates around Mef2 reaffirms that the datasets contain a large number of proteins that interact with Mef2 in all of its functions.

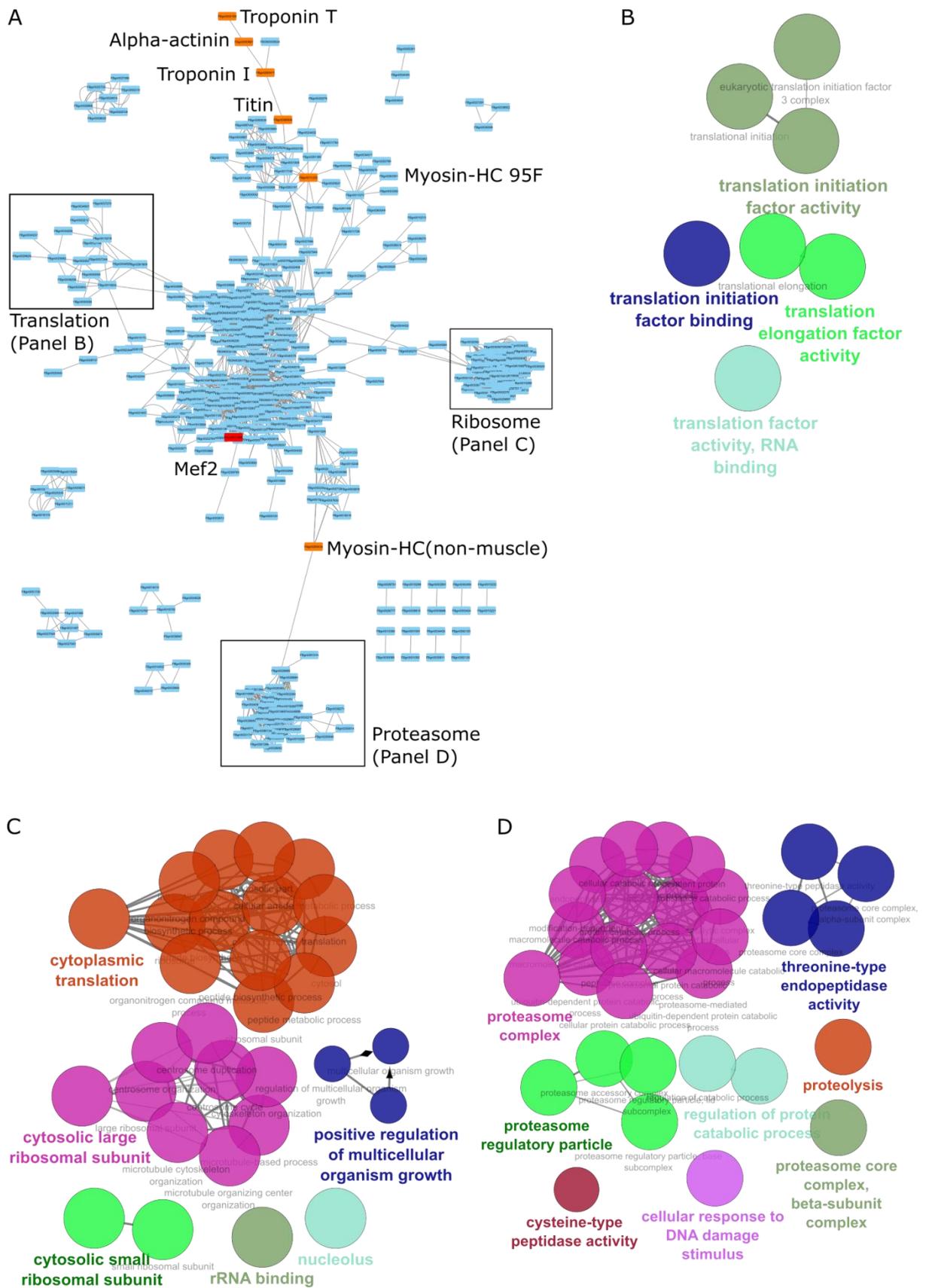


Figure 4.12: Functional analysis of seeds based on the network generated from the most functionally related seeds

A) Representation of the Functional Similarity network obtained by linking seed nodes that have a FunSim score derived from GO terms annotations above 0.7. Red: Mef2. Orange: 11-13MEF GO-specific proteins identified in Figure 4.9 and 4.10. Boxes: Large clusters of interest due to their isolation from the main network. The complete FunSim network can be found in the figshare repository (see data availability statement, section 2.3.13)

B-D) Functional characterisation (GO enrichment analysis) of the proteins forming the most distant clusters that are connected to the main component of the network by very few proteins

4.3 Discussion

4.3.1 Datasets analysed

The datasets of Mef2 interacting proteins were derived from quite different biological backgrounds: one sample included biological material from *Drosophila* embryos of all developmental stages that should give an overview of the most consistent functions that Mef2 and its interactors are involved in (ONMEF), while 11-13h embryos concentrate more on complexes that are involved in muscle differentiation (11-13MEF). Equivalent control samples from all stages wildtype embryos (ONCTRL) and 11-13h wildtype embryos (11-13CTRL) were also tested in order to detect unspecific binders that could be purified from *Drosophila* embryos. The S2-CoAP study (Rhee et al., 2014) was performed in S2 cells, therefore should offer more insight into the cellular functions and specifically the transcriptional regulation partners of Mef2. Despite the fact that the contexts were so different and the exact conditions during the purifications had variations, many of the proteins associated with Mef2 overlapped in the three studies. The creation of the overlap datasets ON/S2-CoAP and 11-13/S2-CoAP was aimed to test if the genes that were shared between datasets identify complexes that are partners of Mef2 in diverse biological contexts to achieve particular functions.

4.3.2 Variety of computational techniques applied

Two approaches were applied to analyse the Mef2-interacting candidate lists derived from TAP purification experiments: 1) a very conservative approach in which Mef2 interaction candidates were prioritised according to spectral counts, specificity to the biological conditions and functional roles according to published literature and 2) a computational approach using network properties, gene expression and gene ontology data derived from published large scale datasets as ways to investigate the different complexes that the candidates might form.

For the first approach spectral counts and NSAF were used to determine protein abundance within a sample and the α -coefficient to assess the enrichment of each protein in Mef2-specific datasets compared to control datasets. After classifying the candidates as specific and non-specific binders, the ONMEF and S2-CoAP datasets

were used as additional controls, and the proteins in these two sets were eliminated to narrow down the candidates to a shortlist of Mef2 interaction partners present only in the 11-13 hour-old embryos. This approach was considered as an efficient way of short-listing candidates that are involved in regulation of Mef2 target genes in muscle development. Literature and expert judgment was then used to identify the most interesting candidate from this shortlist for further study in Chapter 5.

The second analysis made use of a variety of different computational methods to better understand the sets of proteins that were co-purified with Mef2 in the TAP purification experiments. As control datasets the S2-CoAP candidate list was used as a positive control, while ONCTRL, 11-13CTRL were used as negative controls.

By the most naive viewpoint, all proteins contained in the datasets would be considered proteins that physically interact with Mef2, because only the GSTAP-tagged Mef2 bait construct contains the specific protein domains for which the purification procedure selects. The parallel purifications performed from wildtype (WT) embryos that express no bait demonstrate that this assumption is false, and that certain proteins are able to bind non-specifically and withstand the washes during the purification procedure. The fact that most proteins identified in WT extracts were also present in Mef2 extracts implies that at least some proteins in the datasets are likely to be contaminants that were purified not due to a specific physical interaction with Mef2 but due to non-specific binding. Additionally, the TAP purification method is able to not only extract proteins directly bound to the bait, but also long-range interactions, i.e. proteins indirectly bound to the bait as part of protein complexes. The aim of this computational analysis was to estimate which proteins might be contaminants, and which are likely to be genuine Mef2-interacting proteins. These interactors then would be likely to be involved in regulating Mef2 activity. During the analysis, a specific emphasis was placed on identifying candidates that might be involved in a genetic network related to the role of Mef2 in muscle development and cell differentiation.

4.3.3 Analysis of protein abundance derived from mass spectrometry

The first analytical approach used the abundance of proteins identified by MS to estimate which proteins might be contaminants, by comparing the abundances in

samples purified from Mef2-GSTAP bait expressing *Drosophila* embryos to the same protein's abundances in samples purified from WT embryos. Protein abundance is reflected in the total number of spectral counts that the MS identified as originating from each protein. The total spectral counts of proteins obtained from TAP experiments in this study were compared to values from the published Mef2 pulldown studies as a step of quality control. This showed that of the proteins purified by TAP, a substantially larger fraction had abundances considered sufficient for semi-quantitative analysis.

To compare between samples, a more reliable measure of protein abundance was required, as total spectral counts depend on many sample-specific factors and also on the specific protein. Larger proteins have an inherently higher likelihood of generating more peptides identifiable by MS. To account for this, the normalised spectral abundance factor (NSAF; (Zybailov et al., 2006)) was analysed rather than the raw counts. Using NSAF, a statistical analysis was permissible whereby effectively each protein's abundance in samples from Mef2-bait expressing embryos was compared to its abundance in samples from WT embryos. If a protein's abundance was higher in controls, it was classified as a contaminant. This identified a substantial number of potential contaminants.

This analysis is based primarily on the spectral counts identified for each protein by MS, and therefore by the concentration of each protein in the purified sample. However, the amount of each protein that was purified does not necessarily reflect that protein's importance as a Mef2-interacting protein or its relevance to muscle development. For example, a highly abundant protein could be enriched due to a strong physical interaction with Mef2, in which case it would be a very interesting candidate. On the other hand, the high abundance could be due to a generalised high expression of the protein, combined with the ability to bind non-specifically during the purification.

As another indication of sample quality, proteins with a higher abundance in the sample than the bait are generally considered to have a high likelihood of being contaminants, because a specific binder would require an almost permanent physical interaction and an n-to-1 stoichiometry with respect to the bait to reach an abundance higher than the bait. In this respect, the TAP purified samples had the Mef2 bait

either as the most abundant or second-most abundant protein, while both published datasets had a substantial number of proteins with abundances higher than the bait. This suggests that overall the TAP experiments provided samples of higher purity than the published datasets.

Conversely, a protein with low abundance could still be an interesting candidate that plays an important role, but whose abundance is low because its protein expression is highly specific to a small subset of cells, or because its activity is not correlated to a high amount of protein expression. Consequently, the classification into specific binder and contaminant by NSAF and the α coefficient is an estimate and needs to be viewed with caution. To assess whether this classification is corroborated by further analysis, two additional datasets were created, containing only those proteins that were classified as specific Mef2-binding proteins in the overnight and respectively staged samples.

4.3.4 Identification of HDAC4 as a promising candidate for further study

Besides creating additional subsets of candidates to be analysed by computational techniques, the main aim of the first branch of analysis was to prioritise candidates that are likely to regulate myogenesis together with Mef2. As a first step spectral counts were used to calculate NSAFs for each protein in the list. Working with NSAFs is preferable to working with spectral counts since this factor accounts for differences in size and length of different proteins, a property which can impact the number of spectral counts obtained for each protein. While the NSAF allows to compare the different proteins extracted in the same sample, the α -coefficient allows to compare the results of the same protein between samples. Since some proteins were purified from controls and in Mef2-GSTAP embryos, the α -coefficient of these candidates allows to assess if they are real contaminants or this is a case of both an unspecific and specific pull down. One such examples is the Act57B protein which is the most abundant extracted protein in all samples, including controls. This gene is a known target of Mef2 and is one of the few genes that encode myofibrillar actin in *Drosophila* (Kelly et al., 2001; Tobin et al., 1990), and its α -coefficient has shown that the interaction with Mef2 is specific, not just a contaminant. Moreover, Act5C and Act88F were co-purified with Mef2 from S2 cells nuclear extracts, therefore an interaction between Mef2 and the muscle specific Act57B (Rhee et al., 2014) would

not be unexpected. Therefore, in order to separate contaminants from specific interactions the α -coefficient was used as a separation criterion. The proteins classified as specific binders this way were retained in the 11-13MEF* and ONMEF* datasets.

Next, the 11-13MEF* and ONMEF* sets were compared for shared hits. The candidates found only in 11-13MEF* were analysed further, since the biological material the proteins were extracted from should enrich for differentiated muscle specific partners of Mef2, while ONMEF* should contain specific interactors of Mef2 that it binds at any point during embryonic *Drosophila* development. There were 15 such candidates found, of which 7 were also found in S2 cells (Rhee et al., 2014) and were thus discarded as not specific to myogenesis. The remaining 8 candidates were analysed by STRING database, GO terms and for functional relevance in a muscle specific RNAi study (Schnorrer et al., 2010; Szklarczyk et al., 2019).

Among the last 8 shortlisted candidates there were 6 well-known sarcomeric proteins, which implies a cytoplasmic localisation. The remaining two proteins were Brm and HDAC4, two nuclear proteins that are known to be involved in transcriptional regulation. When looking how many of the 11-13h shortlisted candidates had a role in muscle, as determined by an RNAi approach that provided data on 7 of the 8 candidates, only 5 candidates were lethal when knocked down in muscles. Of these 5, 3 were also tested for muscle phenotypes, which were abnormal in all three cases. Considering the analysis was interested in identifying a candidate protein that is able to regulate transcription, the decision in choosing a candidate for biological testing was done between the two proteins with characterised functions of gene regulation in other cell types and model organisms.

The Brm protein is a component of the Brahma complex, the SWI/SNF remodelling complex homologue identified in yeast. In *Drosophila*, this protein has been shown to repress dedifferentiation of intermediate neural progenitors into neuroblasts in the larval body (Koe et al., 2014). The entire complex cooperates with histone deacetylase 3 and the Earmuff transcription factor to stop neuroblast overgrowth in larval brains. Interestingly, other Brahma complex subunits (Moira, Bap55) have been pulled down together with Mef2 from S2 cells and in our study from 11-13h

embryos, though their presence in the S2 dataset had ranked them lower in the priority list. The Moira subunit has been shown to bind the Mef2 enhancer together with Twist and Akirin in *Drosophila* 2-4 and 4-6 hour-old embryos (Nowak et al., 2012). In human cancer cell lines the silencing of *brm* has been associated to the interaction between Mef2d and HDAC9, a class IIa HDAC (Dong et al., 2017; Di Giorgio et al., 2018).

HDAC4 is the only class IIa HDAC in *Drosophila*, while in vertebrates there are 4 different proteins belonging to this category. The interaction between Mef2 and class IIa HDACs has been widely studied in mammalian cell culture, including muscle cells (see Introduction and Chapter 5). The cooperation between Mef2 and HDAC4 in *Drosophila* is not as frequently addressed, only a link between HDAC4 and Mef2 being described in adult neurons and larval glial cells (Fitzsimons et al., 2013; Li et al., 2019). The mammalian HDAC9, the class IIa HDAC that together with Mef2d represses *brm*, is also a target gene of Mef2 and the protein is able to bind different Mef2 variants and repress their transcriptional activity in muscle differentiation (Haberland et al., 2007).

Taking into account the literature available for each of the two candidates and the materials available in the lab at the time of performing experiments, the decision was made that HDAC4 was a suitable candidate for further biological testing. The results of the biological testing of HDAC4 in *Drosophila* embryos are detailed in Chapter 5.

4.3.5 Analysis of candidate proteins' network connectivity

To account for the potential of candidates being specific Mef2-binders regardless of their abundance in MS, a protein-protein interaction (PPI) network analysis was carried out which assessed all proteins in the different datasets regardless of their total spectral counts. The information on known PPIs was queried from the *Drosophila* Interaction Database (DroID), which contains interactions of various nature: physical PPIs, for example derived from yeast two-hybrid studies, genetic interactions sourced from literature, transcription factor-gene relationships, microRNA-gene interactions and predicted protein interactions based on homology to human, yeast and worm. There is no accepted standard method for weighing different kinds of PPIs from DroID relative to each other. Therefore, all interactions

were included in this analysis and the nature of the resulting connections between proteins was not interpreted in detail. Rather, the connections were seen as unspecific associations requiring further substantiation.

The information derived from DroID included both interactions between the candidates submitted in the query to the database, and connections with other potential interaction partners not contained in the list of candidates. The second part, predicted potential interaction partners, showed a surprising relationship between the published studies and the TAP samples. While the lists of candidates had some overlap but many proteins unique to each sample, the predicted interaction partners for all samples were largely shared between the samples. This suggested that potentially each sample is a subset of a large genetic network around Mef2.

The availability of large numbers of interactions showed that the proteins pulled down in the various purification datasets were well characterised and formed a well interconnected network based only on interactions between the candidates identified in the datasets rather than on predicted new interactors. The fact that the proteins were so well connected not only with the seeds, but were also well studied due to the large availability of interactors derived from the database made it possible to study the behaviour of the proteins in the network as a whole rather than only small subgroups of proteins. A high interconnectivity between the seeds was also a strong indicator that the derived information from the screen was reliable. Based on DroID interactions, it was also possible to infer which proteins are more likely to be direct binding partners of Mef2 (these proteins were referred to as "level 0") and which might form longer range interaction (referred to as "level 1" and "level 2"). A comparison of level 0 proteins with data derived from yeast two hybrid studies should be able to biologically confirm these interactions.

An interesting finding during the assessment of interconnectivity between candidate proteins was that the samples based on 11-13 h staged Mef2-GSTAP embryos had more proteins with close connections to Mef2. In the subset of this sample that was shared with the S2-CoAP study, this characteristic disappeared, and the connectivity towards Mef2 was similar to the non-staged samples and the S2-CoAP study. This suggested that those proteins of the 11-13 h dataset which were not shared with the S2-CoAP study were closely connected to Mef2.

4.3.6 Analysis of candidate proteins' functional relationships

Since all the seeds formed a very well connected component it was of interest to determine which proteins are most likely to functionally be related to Mef2. Two different aspects of functional relatedness were analysed: co-expression and correlation in functional annotations.

For co-expression analysis, the percent maximum (pmax) filter for RNA expression was used. In this method, all absolute RNA expression values determined for a particular gene across *Drosophila* embryonic development are obtained from modENCODE, a large-scale study of *Drosophila* gene expression in developmental stages and different tissues (Graveley et al., 2011). The highest expression value (max) is then equalised to 1, and all other values are normalised into fractions of this maximum (percent maximum or pmax). The resulting pmax distribution reflects the relative expression of a gene across development and allows comparisons between different genes. This analysis was applied to all Mef2 interacting protein candidates, and hierarchical clustering was used to identify groups of genes whose expression patterns across time correlate. Specifically, a cluster was identified which is expressed in the late stages of embryogenesis, coinciding with the final stages of muscle development. This cluster was revealed to consist of proteins related to sarcomere organisation and the actin cytoskeleton, suggesting that this might be a genetic network that interacts with Mef2 during this developmental process. Indeed, at this late stage, the larval musculature is fully formed and muscle associated proteins are mainly focused on sarcomere organisation rather than cell differentiation.

To analyse functional annotations, the gene ontology (GO) was used. GO terms are in many cases manually curated annotations available in the Gene Ontology Database, describing the gene's molecular function, the biological processes it is involved in, and the cellular component where gene products are active. The common method of analysing GO terms is to query the GO terms for a given list of proteins, and then to compute for each term an enrichment p-value compared to the random presence of the term across the genome ("GO enrichment analysis"). Most of the GO terms with which Mef2 itself is annotated were also enriched in the candidate lists of the purification studies, confirming that the purifications were successful on a

basic level. By analysing the GO annotations of the proteins in the datasets it was possible to confirm the biological context from which the proteins were extracted, further supporting the quality of the datasets. S2-CoAP samples were mainly enriched in cell biological functions of Mef2 such as gene expression, metabolism and biosynthesis of macromolecules. The 11-13h proteins were found to be associated mainly with GO terms associated with muscle related genes. Overnight embryos were enriched for a mixture of the functions identified in the S2-CoAP study and the 11-13h sample. A strong component of all the GO terms enriched in all the samples including controls was metabolic functions. These metabolic function terms were enriched in a network consisting of all Mef2 candidates regardless of whether connections derived from WT samples (negative controls) were excluded. Assuming that the 11-13h staging should enrich for muscle-related Mef2 interacting proteins, the enrichments were screened for GO terms which were enriched specifically in this dataset. A small group of terms fit this criterion, which were terms related to muscle development and cell differentiation. When in reverse the proteins were identified that had contributed to the enrichment of these terms, it was found that all of these proteins were related to skeletal muscle development and sarcomere organisation.

Finally, the GO terms of all candidates were used to calculate a matrix of functional similarity scores between each pair of candidates, which were then used to create a network where proteins are associated depending on their functional similarity as reflected by GO annotations. This network allowed the categorisation of three separated clusters as possible contaminants, since they were remote from the rest of the network and only indirectly functionally related to Mef2. All other proteins formed a single large cluster around Mef2, meaning that they could not be separated into discreet subgroups based on their functional similarities, or dissimilarities, with each other. If Mef2 interacted with specific discrete protein complexes for different functions, the functional similarity comparison should have separated these complexes from each other. This suggests that even if Mef2 forms persistent protein complexes, the same complex can then be involved in different biological functions.

4.3.7 Identification of a genetic network working around Mef2 in late muscle development

In addition to the shortlist of semiquantitatively identified candidates, the biocomputational analysis identified three small sets of proteins as interesting candidates for further study related to muscle development: 1) The list of candidates from 11-13MEF that did not overlap with the S2-CoAP study. These proteins were interesting because their interconnectivity with Mef2 was higher than the remaining samples. 2) The list of candidates that is co-expressed during late embryogenesis. These proteins were interesting because their expression peaks correlate with Mef2 and occur at the stage relevant to late myogenesis. 3) The list of candidates that caused an enrichment of muscle-related GO terms specifically in the 11-13MEF sample. All three sets share most of the sarcomeric proteins that were also in the semiquantitative shortlist (up, wupA, Tm2, sls, jar), though each of these three lists included some more cytoskeletal proteins that had not been identified semiquantitatively, such as Paxillin, Myosin 61F, Tubulins and α -Actinin. The full lists of candidates can be found in the figshare repository (see data availability statement, section 2.3.13).

A striking feature of these candidates is that they are classified as mainly cytoplasmic due to their nature as sarcomere components. Since they were co-purified with Mef2, this implies that they directly or indirectly interact with Mef2 protein physically, but Mef2 is a transcription factor and should be present primarily in the nucleus. For these proteins to encounter each other, either Mef2 would thus need to be present in the cytoplasm or the candidates would have to be present in the nucleus. Both are in principle possible. There is already evidence that proteins like actin and tropomyosin can display some nuclear functions. The troponin-tropomyosin complex (composed of WupA, Tm1, Tm2 proteins) that is known to regulate muscle contraction through controlling actin-myosin interaction in a Ca^{2+} -dependent manner. The same complex has been shown to regulate nuclear functions, like stable chromosomal integrity and cell polarity in early *Drosophila* embryos (Sahota et al., 2009). It cannot be excluded that such a function could be performed in muscle cells and it depends on interaction with Mef2. Actin57B is the actin type specific for muscle fibres, with well characterised cytoplasmic functions as part of the cytoskeleton and is a well-known

target gene of Mef2 (Kelly et al., 2001). Recently, the idea that actins can also act as transcriptional regulators has come back in focus. Actin is associated with all three eukaryotic RNA polymerases (Fomproix and Percipalle, 2004; Grummt, 2006; de Lanerolle and Serebryanny, 2011; Philimonenko et al., 2004; Visa and Percipalle, 2010) and experimental data suggests that it might be involved in transcription initiation and elongation. Actin also interacts with a myosin 1C isoform called NM1 which in turn interacts with Pol I in rDNA transcription. Actin binds Pol I and NM1 binds translation initiation factor IA to facilitate the assembly of the preinitiation complex (Philimonenko et al., 2004; Venit et al., 2018; Visa, 2005). Considering that Mef2 is involved in transcription activation and Actins are involved in formation of the transcription preinitiation complex, a situation when these proteins like Act57B have a nuclear moonlighting function in muscle is plausible. While for the remaining sarcomeric proteins, no moonlighting function in regulating transcription has been described, it is conceivable that they could similarly be shuttled to the nucleus at low concentrations and interact with Mef2 in order to feed back on their own expression. The presence of a variety of sarcomeric and cytoskeletal proteins in the pull down datasets suggests that this could be a generalised feature of sarcomeric components to advance sarcomere maturation.

In addition to cytoskeletal proteins, all three sets identified further muscle-related proteins involved in transcriptional control like Brahma and Histone Deacetylase 4 (HDAC4), but also a variety of metabolic enzymes such as Pyruvate kinase, Lactate dehydrogenase, Glutamate dehydrogenase, and several mitochondrial proteins like ATP synthase subunits. An intricate link between metabolism, specifically glycolysis and autophagy, and myoblast fusion has been demonstrated for *in vitro* systems and zebrafish embryonic development (Fortini et al., 2016; Tixier et al., 2013). Next to their metabolic functions, these enzymes have been known to execute "moonlighting" functions, acting for example as protein kinases and regulating a vast variety of cellular processes (Kejiou et al., 2019; Lu and Hunter, 2018). The prevalence of metabolic components next to muscle components in these high-interest candidate lists is consistent with the idea that late myogenesis involves far-reaching modulation of cellular metabolism, and this way supports the energy-demanding process of muscle growth. Furthermore, transcriptional activation by Mef2 includes metabolic target genes at least in fat body cells (Clark et al., 2013). It

could thus be a possible transcriptional effector that could be regulated to achieve such adaptation of metabolism during myogenesis.

4.3.8 Further validation of the Mef2 interactome

The main drawback of this analysis was that the connections drawn between proteins are unspecific, and physical interactions cannot be estimated using purely computational means. Further experiments would therefore be necessary to validate the list of candidates. One possible route for this would be a knockdown screen expressing RNA interference (RNAi) constructs targeting each candidate under a muscle-specific Gal4 driver. One such study was previously performed by Schnorrer et al. (2010) and 82% of all the candidates pulled down together with Mef2 in our TAP experiment were tested. Out of 266 candidates tested, 101 showed a wildtype phenotype, while 164 proteins had a defect relating to locomotion, flight or posture. Only 31 of these candidates showed any observable defect in larval muscle morphology or sarcomeric organisation. However, some of these candidates that tested negative in the above study have been shown to regulate Mef2 activity in muscle in other organisms.

Through the use of the pmax parameter, which was calculated based on RNA-seq information derived from the modENCODE database, it was possible to correlate the function of genes in time. A recent method has taken advantage of the Drop-seq protocol to map at single-cell resolution the *Drosophila* embryo transcriptome (Karaiskos et al., 2017). Dynamics of transcription derived from mRNA-seq data have also been used to study the main events of indirect flight muscle development and using “indicator proteins” whose expression correlates with important developmental transitions to validate the transcriptionally significant events (Spletter et al., 2018). Taking advantage of such tools and approaches, in particular trying to link RNA-seq data with proteomic data extracted from carefully staged embryos, should be able to offer a better understanding of what events take place in connection with Mef2 at significant developmental events and what partners partake in them. Therefore, performing Drop-seq extractions in parallel to TAP-purification from early (4-6h, 6-8h) and late embryos (10-12h, 11-13h) should offer a comprehensive outlook on Mef2 interactome and its dynamics in muscle development.

To verify physical protein-protein interactions, one could perform affinity purifications in reverse, using the interactor candidates as bait, or use a classic system such as yeast two hybrid, or one of the newer fluorescence based systems with split fluorophores (Cabantous et al., 2013).

Chapter 5: HDAC4 as a potential Mef2 interactor identified by the screen

5.1 Introduction

5.1.1 The Mef2 and HDAC axis in vertebrates

Specialised tissues are formed through cell differentiation programs during development, starting from progenitor cells with the ability to turn into specialised cell types under cues of cell signalling cascades that use molecules that shuttle into the nucleus to interact with transcription factors. Transcription factors are able to activate specific gene programs to induce cell differentiation (Perrimon et al., 2012). In order to be able to understand the regulation of a developmental process, it is important to uncover the link between interactions occurring at molecular level to the effects generated on a phenotypic level via mis-expression studies. Myogenesis represents a classic paradigm in the study of regulation of differentiation programs controlled by key transcription factors and other transcriptional regulators (Bentzinger et al., 2012).

Myocyte enhancer factor 2 (Mef2) is a transcription factor that is the major regulator of gene expression and differentiation in muscle, conserved from flies to humans (Black and Olson, 1998). Mef2 was first identified in mammalian cell culture and many of the regulatory molecules that interact with it during myogenesis have been characterised in this system, including class IIa HDACs (Gossett et al., 1989; Yu et al., 1992). In vertebrates there are four closely related Mef2 genes, while *Drosophila* has a single Mef2 gene. A similar situation is encountered for class IIa HDACs. The only *Drosophila* class IIa HDAC is dHDAC4, while HDAC4, -5, -7 and -9 are the four vertebrate homologues.

The Mef2 proteins are part of the family of MADS-box proteins and were initially characterised in vertebrates as important regulators downstream of the MyoD family of transcription factors, while in *Drosophila* the only Mef2 gene is activated by the mesodermal determining factor Twist (Potthoff and Olson, 2007). The central role of Mef2 in muscle differentiation was first identified in *Drosophila* embryos and the wide range of muscle specific genes it activates have emphasised the crucial importance of studying how Mef2 works in order to understand how muscle is made. The role of class IIa HDACs in mammalian muscle in connection to Mef2 has been

widely characterised in cell culture, while to date nothing is known about the role of dHDAC4 in *Drosophila* muscle.

In vertebrates, the four Mef2 transcription factors (Mef2a, -b, -c, -d) have divergent as well as overlapping functions, which complicates the study of functional roles specific for each isoform. Studies have shown that they act as central regulators in a range of cell types (skeletal, cardiac and smooth muscle, brain, neural crest, lymphocytes and bone), governing a diversity of developmental programs (cell proliferation, survival, apoptosis and differentiation) (Potthoff and Olson, 2007) and a trusted interaction partner in all of these contexts is a class IIa HDAC. The interaction between Mef2 and a class IIa HDAC in the nucleus leads to transcriptional repression, irrespective of the context and a direct physical interaction is expected between the two since all the class IIa HDACs present a Mef2 binding domain at their N-terminus (Jayathilaka et al., 2012).

Mammals have 17 histone deacetylases that are grouped in 3 subtypes based on their homology with the yeast HDACs: Rpd3 (Class I), HDA1 (Class II), and Sir (Class III). The class I and II HDACs are more closely related in terms of sequence and base the hydrolysis of the acetyl-lysine amide bond on Zn-catalysis. The Class III HDACs do not show sequence similarity with the other two Classes and use NAD as the acetyl group acceptor (Blander and Guarente, 2004). The Class I HDACs have 4 subtypes: HDAC1, 2, 3 and -8, and Class II HDACs are further split into the subgroups Class IIa (proteins HDAC4, 5, 7, 8) and Class IIb (only two proteins, HDAC6 and HDAC10). HDAC11 is evolutionarily not particularly homologous to either Rpd3 or HDA1, therefore it cannot be assigned to either Class I or II HDACs, and some studies assign it to a distinct category of HDACs entitled Class IV.

The mammalian class IIa HDAC proteins have sequence similarity in their catalytic domain, the extended long N-terminal domains and the C-terminal tails. The N-terminus contains one Mef2 binding domain, and two phosphorylation sites specific for interaction with 14-3-3 proteins and a nuclear localisation signal (NLS). These proteins also contain a nuclear export sequence (NES) located in the C-terminus region (Yang and Grégoire, 2005). These proteins perform their regulatory roles mainly independent of their deacetylase ability and they associate with class I

HDACs in macromolecular complexes to induce chromatin modifications (Fischle et al., 2002; Lahm et al., 2007).

Their subcellular localization carries an important role in regulating their repressive activities, as these proteins shuttle in and out of the nucleus. Various kinases activated by extracellular stimuli can alter the balance between nuclear import and export of these repressors (Clocchiatti et al., 2013). Mechanisms of nuclear retention of class IIa HDACs involve interaction with Mef2 and other factors such as Parathyroid hormone related peptide (PTHrP) and Forskolin, which activate protein kinase A (PKA), which in turn can phosphorylate a Serine residue in the NLS of HDAC5 and block nuclear export (Borghi et al., 2001; Ha et al., 2010).

In order to reside on chromatin, class IIa HDACs associate with selected transcription factors such as the Mef2 family. The binding of class IIa HDACs to Mef2s has been extensively characterised in mammals. The Mef2 binding site is conserved among the four mammalian HDACs. For HDAC4 and HDAC5 this motif overlaps with a Ca^{2+} /calmodulin-binding site. Ca^{2+} /calmodulin activates specific kinases that phosphorylate HDACs for 14-3-3 protein binding, thus promoting their nuclear export (McKinsey et al., 2001). In a signal-responsive manner, Mef2 recruits HDACs to repress transcription (Yang and Grégoire, 2005). Repression of Mef2 activity by class IIa HDACs leads to repression of myoblast differentiation (McKinsey et al., 2000) and chondrocyte hypertrophy (Vega et al., 2004). The binding between HDACs and Mef2 is dynamic and it can be stabilised in myoblasts if the PI3K pathway is blocked (Serra et al., 2007). Class IIa HDACs are able to increase Mef2 sumoylation which will decrease its transcriptional ability (Grégoire and Yang, 2005). Other transcription factors that class IIa HDACs regulate include: SRF, Runx2, GATA, Forkhead (Clocchiatti et al., 2011).

Mammalian class IIa HDACs are expressed in a tissue specific manner. HDAC4, -5, -9 show their highest expression in the heart, skeletal muscle and brain, while HDAC7 is mostly expressed in the thymus. Knockout of HDAC4 and HDAC7 in mice led to embryonic lethal abnormalities, while HDAC5 and -9 knockout mice presented cardiac hypertrophy in advanced age. Defects in mobility and breathing were the causes leading to embryonic lethality (Clocchiatti et al., 2013).

The role of class IIa HDACs in muscle development has been studied in cell culture and all the four isoforms are able to repress muscle differentiation by repressing Mef2-dependent transcription activation. Their interaction with Mef2 is histone deacetylase activity independent and they shuttle between the cytoplasm and the nucleus under regulation of signalling cues. In order to stop transcriptional repression and allow muscle differentiation to take place, class IIa HDACs are excluded from the nucleus via phosphorylation mechanisms. The repressive roles of HDAC4, -5, -7 and -9 in muscle are redundant since differentiation can continue in absence of a specific isoform (Clocchiatti et al., 2011). However, when a mutation that rendered the NES of HDAC5 inactive and HDAC5 was trapped in the nucleus, muscle differentiation was not impacted (McKinsey et al., 2001), which implies that class IIa HDACs undergo regulation in the nucleus as well. Degradation of class IIa HDACs and Mef2 activation of genes can cause a switch in muscle fibre properties from fast and glycolytic to slow and oxidative (Potthoff et al., 2007). In muscle, all class IIa HDACs are regulated at translational levels by miRNAs and only one of the four isoform (HDAC9) is a direct transcriptional target of Mef2 (Chen et al., 2006; Haberland et al., 2007). MITR is an isoform of HDAC9 that is able to bind Mef2, does not have a HDAC deacetylase domain and is exclusively nuclear (Sparrow et al., 1999; Zhou et al., 2000).

5.1.2 Histone deacetylases in *Drosophila*

There are orthologues in *Drosophila* for Class I and II HDACs. In class I, HDAC1 and HDAC2 are orthologous to dHDAC1 and HDAC3 is similar to dHDAC3. The four subtypes of Class IIa HDACs have only one orthologue named dHDAC4 in *Drosophila*. DHDAC2 is homologous to HDAC6, while dHDACX is the orthologue of HDAC11 (Foglietti et al., 2006).

When knocking down the Class I and II HDACs in *Drosophila* S2 cells via RNAi, only dHDAC1 showed an increase in histone acetylation and only dHDAC1 and HDAC3 created a gene expression signature. All the other HDACs when knocked down did not show any transcriptional effects leading to the hypothesis that they are involved in interacting with nonhistone substrates. Interestingly, knocking down dHDAC2 increased tubulin acetylation, an effect that has also been observed in

mammals for HDAC6 which acts as a tubulin deacetylase (Foglietti et al., 2006; Haggarty et al., 2003).

A study looking at expression of HDACs at different stages of *Drosophila* development showed that Class I HDACs dHDAC1 and dHDAC3 were highly expressed in the embryo, with dHDAC3 being also highly expressed in the adult (Cho et al., 2005). Class II HDACs, meaning dHDAC4 (IIa) and dHDAC2 (IIb) reached highest expression in the adult. All the Class I and II HDACs had their lowest expression in the embryo. The class III HDAC Sir2 has a distinct pattern, with the highest expression in the embryo, lowest in the larvae and a recovery of expression to half of the embryo levels during pupal and adult stages. In terms of subcellular localisation, different HDACs have distinct patterns of expression in S2 cells when overexpressed: dHDAC1 and Sir2 are mainly nuclear, dHDAC2 is predominantly cytoplasmic and dHDAC4 and dHDAC3 shuttle between the nucleus and cytoplasm.

When overexpressed in S2 cells each individual HDAC was able to elevate only its own expression levels, no upregulation or repression of another HDAC was observed, which suggests that each type of HDAC has a distinct role compared to the others. The ratio of upregulated to downregulated genes when one HDAC is overexpressed is distinct for each case. HDAC1 overexpression resulted in more genes upregulated than downregulated, while dHDAC4 and dHDAC2 showed a bias towards more downregulated genes. DHDAC3 overexpression produced a similar number of up and downregulated genes. These results show that HDACs can also activate transcription, not only repress it. When comparing the identity of genes that were mis-expressed when a particular HDAC was overexpressed it was concluded that each HDAC regulates a different subset of genes. Therefore HDACs in *Drosophila* have distinct roles and can regulate a variety of target genes. When assessing the type of genes mis-expressed by the different HDACs, some patterns emerge. Class I HDACs are involved in *Drosophila* development, in particular during embryonic development. The class IIa HDAC regulates organization of the cytoskeleton and dHDAC2 regulates genes involved in olfactory receptor activity, cell cycle and axon guidance. Sir2 was involved in misregulation of genes involved in glycolysis, immune response and aging (Cho et al., 2005).

5.1.3 HDAC4, the only class IIa HDAC in *Drosophila*

Unlike mammals, lower organisms have only one class IIa HDAC which contains the conserved Mef2 domain. HDAC4, the only class IIa HDAC in *Drosophila*, has two putative 14-3-3 binding site and a high sequence similarity to the identified NLS in mammals. To date, the interaction of HDAC4 with Mef2 in *Drosophila* myogenesis has not been analysed. However, its functions in other *Drosophila* tissues have been investigated.

According to *in situ* hybridisation studies dHDAC4 has a dynamic expression in the *Drosophila* embryo (Zeremski et al., 2003). At pre-blastoderm stages, the gene was expressed ubiquitously, with expression concentrated around the nuclei in the syncytium. About 2 h after egg laying the expression became localised to a broad anterior domain within the embryo, followed by the formation of seven additional stripes during cellularization. As development continues with gastrulation, the expression of dHDAC4 seems to follow a segment-polarity pattern containing 14 stripes. When analysing the expression of dHDAC4 in segmentation gene mutant backgrounds, it was found that hunchback, knirps, and giant were able to activate expression of the gene since the stripes expressed in the area of those particular segmentation genes were missing. When expressed in an even-skipped mutant background, the pattern of expression of dHDAC4 had broader, unresolved stripes compared to the wildtype, therefore even-skipped is able to repress the dHDAC4 expression in areas where it is expressed. In human 293 cell lines dHDAC4 was shown to possess catalytic capabilities and the histone deacetylase activity can be suppressed by a HDAC specific inhibitor.

The shuttling of dHDAC4 between the nucleus and the cytoplasm is used in the fat body of *Drosophila* to regulate insulin response. When phosphorylated by SIK3, dHDAC4 is trapped in the cytoplasm where it is unable to interact with FOXO. Lipolysis and gluconeogenesis are dependent on FOXO activation of specific target genes. To bind DNA, FOXO requires to be deacetylated. When fasting occurs, SIK3 is deactivated and dHDAC4 translocates to the nucleus and deacetylates FOXO. DHDAC4 is able to activate FOXO-dependent transcription via dephosphorylating the TF directly, or by associating with other HDAC complexes (Wang et al., 2011).

A recent study identified that dHDAC4 plays a role in modulation of long-term courtship memory. The repressive effects of HDAC4 on long-term memory in *Drosophila* seems to occur through interaction with Mef2 (Fitzsimons et al., 2013), which has been implicated in long-term memory regulation in *Drosophila* (Blanchard et al., 2010). In glial cells, HDAC4 regulates Mef2 target gene activation under the influence of SIK3. When phosphorylated by SIK3, HDAC4 accumulates in the cytoplasm and Mef2 is able to activate target genes with roles in water and K⁺ homeostasis. If dHDAC4 accumulates in the nucleus, Mef2 gene activation is repressed and nerves swell (Li et al., 2019).

5.1.4 Mef2 in *Drosophila* muscle and its regulators

Drosophila has two phases of the myogenic program: the first occurs in the embryo to form the larval musculature (somatic, cardial, visceral), while the second one occurs during metamorphosis to form the adult muscles (abdominal muscles, indirect flight muscles, leg muscles). In the embryo Mef2 is expressed in the mesoderm and its muscle derivatives that will form muscle fibres through fusion of founder cells to fusion competent myoblasts. Mef2 is expressed in muscle progenitors and continues to be expressed throughout the muscle differentiation process. Mef2 initiates the program that regulates myoblast fusion around stage 12 of development and drives the differentiation of the resulting myotube into a contractile fibre (Bour et al., 1995). The somatic muscle is formed by a pattern of 30 distinct multinucleated muscle fibres located in each abdominal hemisegment, with defined size, shape and attachments (Bate, 1990) given by the identity genes of the originating founder cell. During muscle differentiation, Mef2 activates target genes with a range of expression profiles (Elgar et al., 2008; Junion et al., 2005; Sandmann et al., 2006), due to target genes responding differentially to Mef2 levels. *In vivo* occupancy experiments revealed that Mef2 binds dynamically to enhancers, although it is expressed continuously. One group of enhancers is bound only early in development, while another group is bound at late developmental stages (Sandmann et al., 2006).

In adult myogenesis, Mef2 is first detected in the myoblasts of late third instar larvae (Soler and Taylor, 2009) and it is required for the formation of adult fibrillar and tubular muscles. These muscles have structurally and physiologically distinct

properties and they form through remodelling of larval muscles or *de novo* assembly. Down-regulating Mef2 during adult myogenesis leads to failure of AMPs and FCMs fusion or muscle splitting depending on the timing of altering Mef2 activity. AMPs are precursors of muscles that were set aside and remained undifferentiated during myogenesis (Soler et al., 2012).

The interaction between dHDAC4 and Mef2 has been documented in *Drosophila* neurons and many of the characteristics found in mammals are also found in *Drosophila*.

Considering the diverse interactions between HDAC4 and Mef2 transcription factors to regulate various developmental processes in mammals, we believe HDAC4 represents one of the co-regulators of Mef2 that are able to modulate its activity. There is substantial evidence in mammals that supports the importance of class IIa HDACs-Mef2 interaction during muscle development. Considering these, we are interested to characterise the role of HDAC4 in *Drosophila* myogenesis. Based on the information from vertebrates, we hypothesise that HDAC4 is an inhibitor of muscle differentiation, the protein interacting with Mef2 to repress muscle specific genes. Therefore, HDAC4 is a potential muscle differentiation inhibitor keeping progenitor cells in an undifferentiated state. The TAP experiments identified dHDAC4 as a Mef2-interacting protein in 11-13h old embryos, which corresponds to stage 14-15 when the muscle pattern is fully differentiated. However, in mammals class IIa HDACs have been shown to affect myogenesis also at myoblast levels, therefore the interaction between Mef2 and dHDAC4 will be assessed both in early and late embryogenesis.

5.1.5 Experimental approach

To study the relationship between HDAC4 and Mef2, classic molecular biology techniques and genetics were used. This included RNA *in situ* hybridisation and immunohistochemistry to test expression of HDAC4, as well as Gal4/UAS-driven overexpression of fly HDAC4 and human HDAC5 transgenes. Using Flippase-mediated recombination, two new null alleles of the *HDAC4* gene were generated for phenotypic loss-of-function studies.

5.2 Results

5.2.1 HDAC4 expression in *Drosophila* embryos

A gene's expression pattern can provide information on its possible roles during a developmental process. By analyzing both the expression at RNA and protein level of HDAC4 it is possible to assess in which developmental processes the protein is potentially involved in.

Gene model

The HDAC4 gene is located in the *Drosophila* genome on the X chromosome, cytogenetic map 11E8-11E9, Sequence location X:13,262,687..13,285,632 [-] (Flybase FB2018_05, released Oct 16, 2018). There are 8 transcripts assigned to this gene and a number of unique peptides. The protein isoforms do not share the same N-terminus, but they overlap in the C-terminal region, an aspect which can impact the creation of loss-of-function mutants.

RNA expression

To determine if HDAC4 is expressed in *Drosophila* embryos, the gene expression was initially analysed via published RNA-seq data. The modENCODE temporal expression data (Graveley et al., 2011) comprises RNA-seq data of genes' expression throughout *Drosophila* development from early embryonic stages to adulthood. For the embryonic part of development, the analysis of the genes expression was performed in 2 h embryonic developmental windows. HDAC4 has a relatively low expression in *Drosophila* and the peak is achieved in the 8-10h old embryos. The peak of expression of HDAC4 is observed in a 10 h interval, starting at 6 h AEL and ending at 16 h of development. Based on this information the gene should be expressed in embryos and in order to analyse where the gene is expressed *in situ* hybridisation was prepared.

The expression of *HDAC4* RNA in embryos was analysed using *in situ* hybridisation. An RNA *in situ* probe was prepared based on a cDNA clone (FBcI0212840) gifted by the *Drosophila* Genomics Research Centre. The *in situ* hybridization assay was performed on wildtype embryos using an antisense probe. The control was a sense probe generated from the same cDNA clone. This control is particularly important to

distinguish the background from the signal of the probe labelling HDAC4 mRNA. HDAC4 has 8 annotated transcripts in *Drosophila*. The cDNA used to generate the RNA *in situ* probe represents the mRNA sequence of the HDAC4-B variant. However, due to shared exons between the 8 isoforms of HDAC4 it was possible to target all the mRNA variants. The RNA *in situ* protocol involves using small RNA from the probe used that are able to interact in the experiment with HDAC4 transcripts.

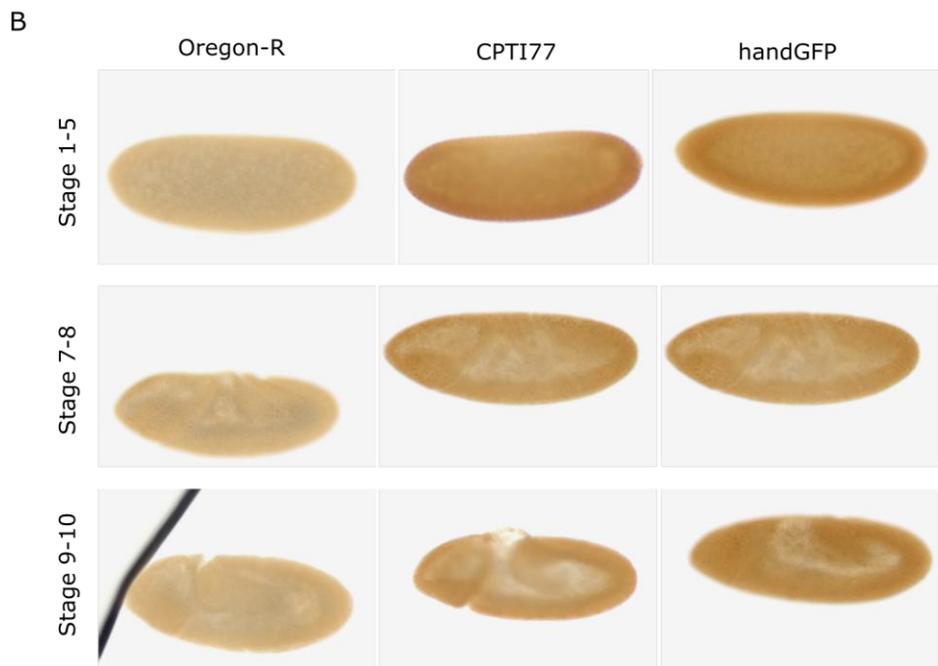
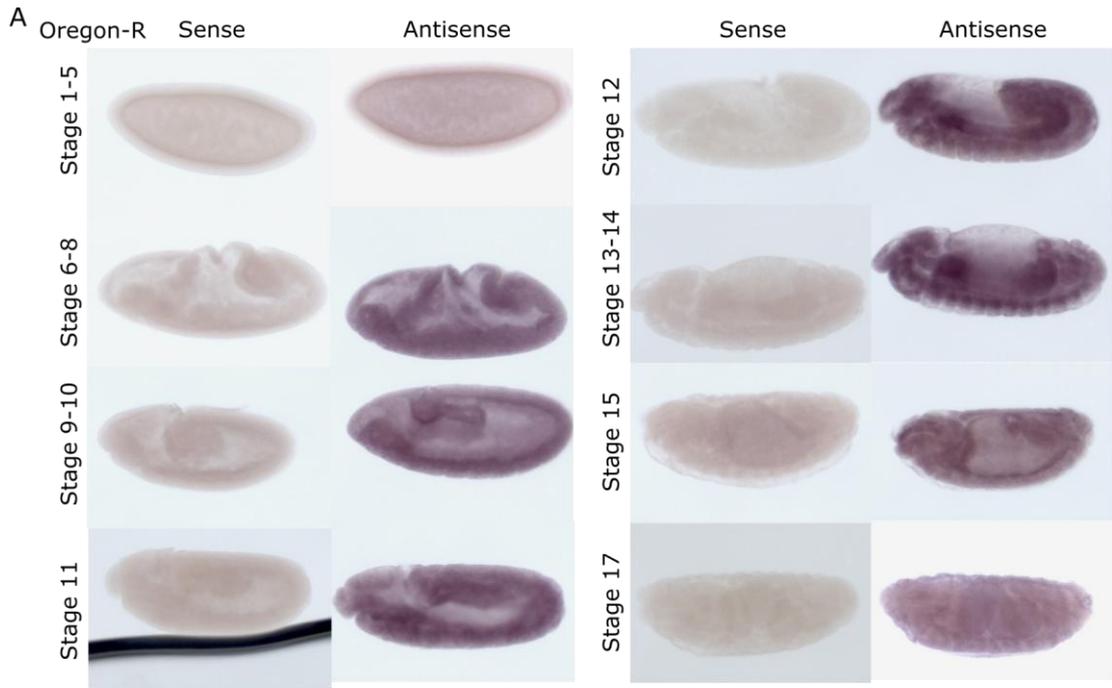
Based on our data the mRNA of HDAC4 is expressed throughout the embryo at all embryonic stages (Figure 5.1A). Expression of HDAC4 specific mRNA is detected in very early embryos, thus indicating a potential role of HDAC4 in muscle progenitor cells together with Mef2. There are particular regions of the embryos where the *in situ* results show a higher expression, however it is difficult to assess the specificity of such tissues considering that the embryos have more opaque areas in the wild-type embryos.

Protein expression

To complement the gene expression study, the protein expression in *Drosophila* embryos was analysed as well. HDAC4 protein expression was investigated using immunohistochemistry on embryos containing a YFP insertion in the *HDAC4* region (line CPTI77). By using an anti-GFP antibody it was possible to visualise the expression of the fusion protein (Figure 5.1B).

The CPTI77 line used to study HDAC4 protein expression was generated using a piggyBAC protein tag construct containing a splice acceptor and donor site, purification tags for StrepII and FLAG and a functional YFP exon. The YFP insertion in the CPTI77 line was mapped in the N-terminal region of HDAC4 (site 13174889), in frame with the HDAC4 sequence.

The expression of YFP-fused HDAC4 was studied using an anti-GFP antibody generated in Rabbit. The expression of HDAC4 in *Drosophila* embryo can be observed in Figure 4.1. OregonR/wildtype embryos were used as a negative control, while embryos expressing a handGFP construct were used as a positive control. Based on our data the protein is expressed throughout the whole embryos at all embryonic stages. No particular areas seem to be enriched for HDAC4. By



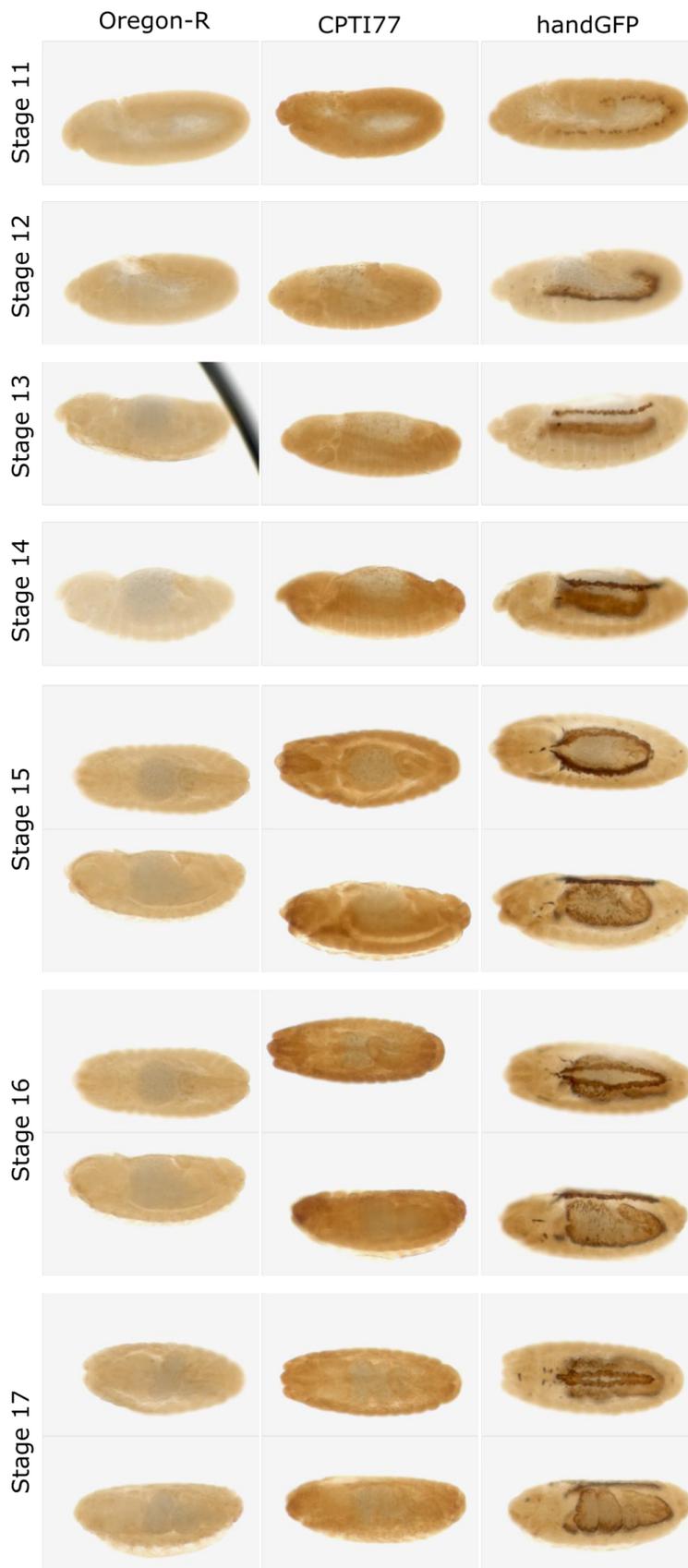


Figure 5.1: mRNA and protein expression of HDAC4 in *Drosophila* embryos

A) Wildtype embryos were incubated with either an antisense RNA probe to detect *hdac4* mRNA, or with a sense RNA probe as a negative background control. Lateral views are shown.

B) Immunohistochemistry using an antibody against GFP to assess HDAC4 protein expression using a YFP-HDAC4 protein trap line. Oregon-R: Wildtype control not expressing GFP. CPTI77: transgenic line containing a YFP insertion resulting in a YFP-HDAC4 fusion protein. handGFP: Positive control with high expression of GFP. Lateral views are shown. Note that most anti-GFP antibodies crossreact with YFP.

comparison the hand protein is ubiquitously expressed in early stages, while starting from mesodermal differentiation the proteins is present in the visceral mesoderm and the dorsal tube and the visceral musculature. The ubiquitous expression of HDAC4 seems to suggest that the protein could be expressed in muscle.

HDAC4 and Mef2 co-expression across development

HDAC4 was suggested as a Mef2 interacting protein as part of a TAP purification study from 11-13h embryos where GSTAP-tagged Mef2 acted as bait. The validity of the interaction was determined both via spectral counting (where HDAC4 was determined as a specific binder as determined via statistical analysis) and through network models. In a PPI network constructed on information derived from DroID database HDAC4 was determined as a direct interactor of Mef2. A direct interactor in this network model does not necessarily imply a physical binding between Mef2 and HDAC4, but a very strong connection in functional relatedness. In this PPI network, HDAC4 was found to interact with 17 other proteins, 5 of which were found to also be direct interactors of Mef2. The other 12 proteins are contextually connected to HDAC4 as well as the other 5 direct interactors.

A gene is considered more likely to be active in contexts where it is expressed close to its maximal level. To determine a subnetwork of genes active in specific contexts an expression filter was defined in order to determine if two genes are likely to be expressed and active in the same tissue or stage. A gene's expression level in each stage or tissue (p_{max}) can be calculated as a percentage of its level in the tissue or stage where it is maximally expressed (more detailed explanation in the previous chapter). Based on different filters tested it was determined that proteins that have a p_{max} higher than 45% in the same tissue or stage are involved in the same biological context (Murali et al., 2014). To determine if Mef2 and HDAC4 are active in the same developmental stage during *Drosophila* embryonic development, the p_{max} for these proteins was determined throughout the different embryonic stages of development. The expression information was determined from RNAseq data from modENCODE which was measured throughout the lifetime of a fly. The p_{max} of Mef2 and HDAC4 was compared at different embryonic developmental stages (recorded in 2 h windows). A third gene that shows a similar behaviour in the

HDAC4 subnetwork in terms of connectivity with Mef2, other level 0 seeds and level 1 proteins was included for comparison.

In the embryos, HDAC4 has the highest expression in 0-2h embryos, afterwards its level drop under the considered threshold of 0.45. At 10-12h of development HDAC4 levels start increasing slowly close to a significant pmax expression, a second expression peak being obtained at 20-22h. Mef2 expression levels slowly increase in the early developmental stages, the maximal peak being reached at 8-10h. From then on the levels start decreasing, but maintain a value above the threshold until 16-18h embryonic development. In late embryonic development Mef2 levels go under the threshold but stay at a higher level than in early embryonic development. The Ubi-p63E protein has relatively low levels in early and late embryonic development. A tripling of the expression level is observed in 6-8 h embryos and stays similarly high until 12-14h of development when the level suddenly drop to very low levels in late embryogenesis.

It is important to note that in the TAP purification proteomic study, HDAC4 was co-purified with Mef2 from 11-13h embryos while Ubi-p63E was found both in Mef2 bait-expressing overnight embryos and wildtype control purification samples. Based on spectral counting both HDAC4 and Ubi-p63E were classified as specific binders. According to the analysis of the pmax expression filters it is visible that at 10-12h of development all the three analysed genes are above the activity threshold and should be involved in the same biological context. In the case of HDAC4 and Mef2 the developmental stage of 14-18h seems to show another context when the two genes could be acting in the same context.

5.2.2 Colocalization of HDAC4 with Mef2

In order to better understand the role of HDAC4 in muscle, a colocalisation study between HDAC4 and Mef2 was carried out. Mef2 is expressed in both muscle progenitor cell and differentiated muscle, and we tested whether HDAC4 is expressed in these cells, in which subcellular localisation, and at what time during development.

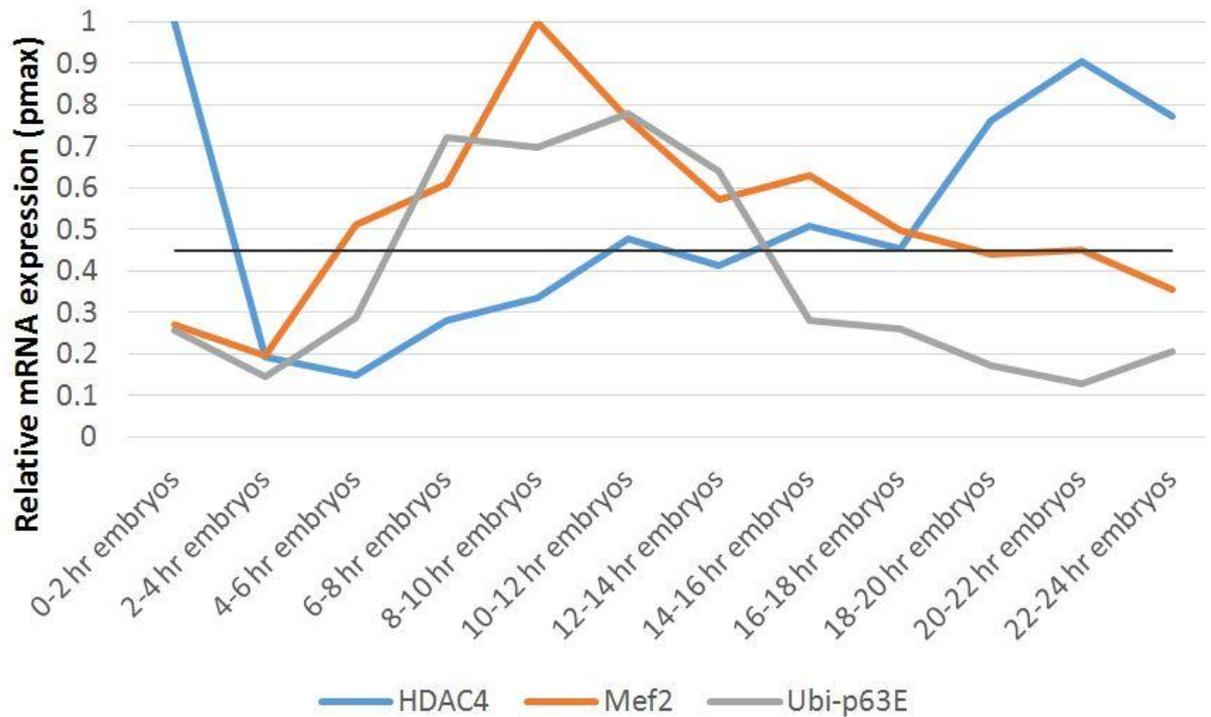


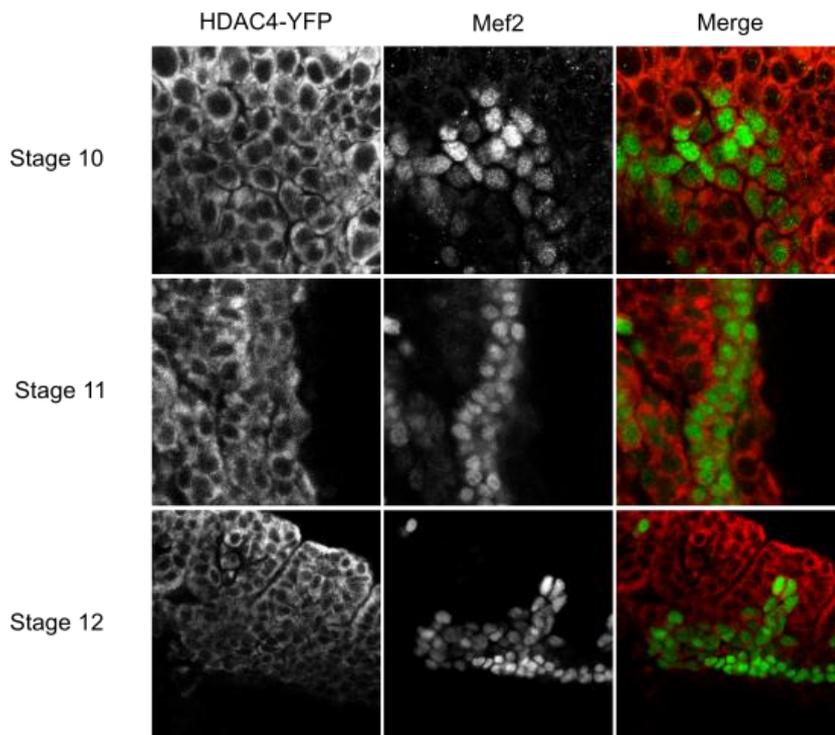
Figure 5.2: Relative mRNA expression of Mef2 and two co-expressed candidates across embryogenesis

HDAC4 and Ubi-p63E were identified as potentially interesting candidates that are part of a muscle-specific group of Mef2-interacting proteins in the previous TAP experiments. Graphs show percent max (pmax), the relative RNA expression level at each stage. Pmax at each stage is the fraction of the expression at the stage with the highest level. The expression peak for Ubi-p63E did not occur during embryogenesis, hence no value on the graph is 1. Black line: 0.45 cutoff, a gene is considered to be actively expressed when above this cutoff.

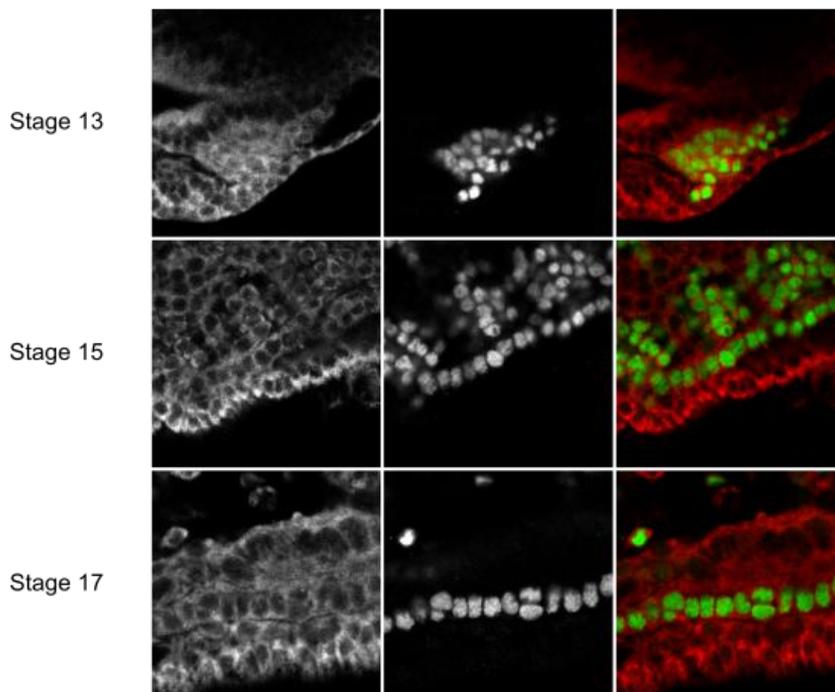
The relationship between Mef2 and class IIa HDACs has been widely documented in mammalian studies. The interactions between the two types of proteins was studied in many tissues and cell types such as muscle, brain, neurons, thymocytes. In myogenesis, class IIa HDACs act as repressors by binding directly to Mef2 at some critical promoters for the muscle differentiation programme (McKinsey et al., 2001; Lu et al., 2000). It is believed that the HDAC maintains the muscle progenitor cells in a repressive state until the appropriate myogenic differentiation signal is delivered. When muscle differentiation is underway, class IIa HDACs shuttle to the cytoplasm. A similar interaction was reported in *Drosophila* neurons (in particular Kenyon cells) between HDAC4 and Mef2. The HDAC4 protein and Mef2 were reported to colocalise at specific loci within the nucleus of neurons when HDAC4 was overexpressed. This overexpression had an impact on long term memory because plasticity genes were being repressed (Fitzsimons et al., 2013; Sando et al., 2012). The nuclear bodies observed in this case have also been reported in mammalian cell culture. The nuclear localization domain of the MEF2 family of transcription factors shows member-specific features and mediates the nuclear import of histone deacetylase 4 (Borghi et al., 2001; Miska et al., 1999). The localisation of HDAC4 in the *Drosophila* brain has been reported as a cytoplasmic halo around the nucleus and a punctate nuclear localisation (Fitzsimons et al., 2013).

We used the CPTI77 *Drosophila* line to study the colocalisation of HDAC4 to Mef2 in muscle cells throughout embryogenesis. A primary antibody against GFP was used to observe the localization of HDAC4 since this antibody detects the YFP insert of this stock. It was possible to identify embryonic regions that will result in the formation of muscle due to the specific localisation of Mef2 in the mesoderm and to the nuclei of the muscle fibre at different development stages of *Drosophila*. The specific expression of Mef2 in the mesoderm is observed starting with stage 9 and continues until the end of embryogenesis in the developing muscle fibers.

We observed that HDAC4 expression throughout the embryos had a honeycomb-like pattern. This expression pattern seems to be due to a cytoplasmic, rather than nuclear expression of the protein. Although it is known that class IIa HDACs shuttle between the cytoplasm and the nucleus. Interestingly, at no embryonic stage the YFP-HDAC4 and Mef2 signals overlapped. Considering that Mef2 expression is mainly nuclear,



Assumed repressed state of Mef2 by HDAC4 in progenitor cells



Assumed activated state of Mef2 by HDAC4 during myotube formation

Figure 5.3: Colocalisation of YFP-HDAC4 with Mef2 during *Drosophila* embryogenesis.

Confocal micrographs of muscle precursors during *Drosophila* embryogenesis. Red: YFP-HDAC4 stained with anti-GFP in the CPTI77 transgenic line. Green: Mef2 antibody.

this stain seems to suggest that the HDAC4 protein is mainly expressed in the cytoplasm. However in order to confirm this hypothesis, the colocalisation of Mef2 expression with a nuclear specific marker like DAPI is required.

5.2.3 Functional comparison to human HDAC5

The role of class IIa HDACs has been widely studied in mammalian systems, however not much is known about their role in *Drosophila melanogaster*, in particular at embryonic level. There are 4 mammalian class IIa HDACs compared to only one in *Drosophila* which is referred to as HDAC4. To understand the role of HDAC4 in myogenesis in *Drosophila*, the effects of overexpressing this protein in the embryo was compared to the effects of overexpressing the human HDAC5 protein which has been shown to repress myogenesis by interacting with Mef2.

Mammalian class IIa HDACs have sequence similarity in their catalytic domain, the extended long N-terminal domains and the C-terminal tails (Figure 5.4A). The N-terminus contains one Mef2 binding domain, and two phosphorylation sites specific for interaction with 14-3-3 proteins and an NLS signal. The C-terminus contains the catalytic deacetylase domain and the nuclear export sequence (Di Giorgio et al., 2015). *Drosophila* HDAC4 (dHDAC4) has a deacetylase catalytic domain at its C-terminus, as well as a Mef2 binding site at its N-terminus. *Drosophila* HDAC4 is relatively highly conserved, with 57% amino acid identity and 84% similarity to human HDAC4 across the deacetylase domain C terminus, and 35% identity and 59% similarity across the whole protein (Fitzsimons et al., 2013). The Mef2 binding domain is conserved in dHDAC4 and the Serine residues that are involved in the shuttling in and out of the nucleus are conserved. The NLS is located in between the two serine residues and is hidden when the HDAC activity is inhibited. The C-terminal histone deacetylase domain is maintained in the *Drosophila* HDAC4 as well.

The activity of HDAC4 in *Drosophila* embryos and its involvement in muscle development was studied by modulating its activity. The effects on muscle formation and on Mef2 protein levels was studied at two important events in muscle development: at stage 12 when mesodermal specification occurs and at stage 16 when the muscle pattern should be fully developed. The structural characteristics of

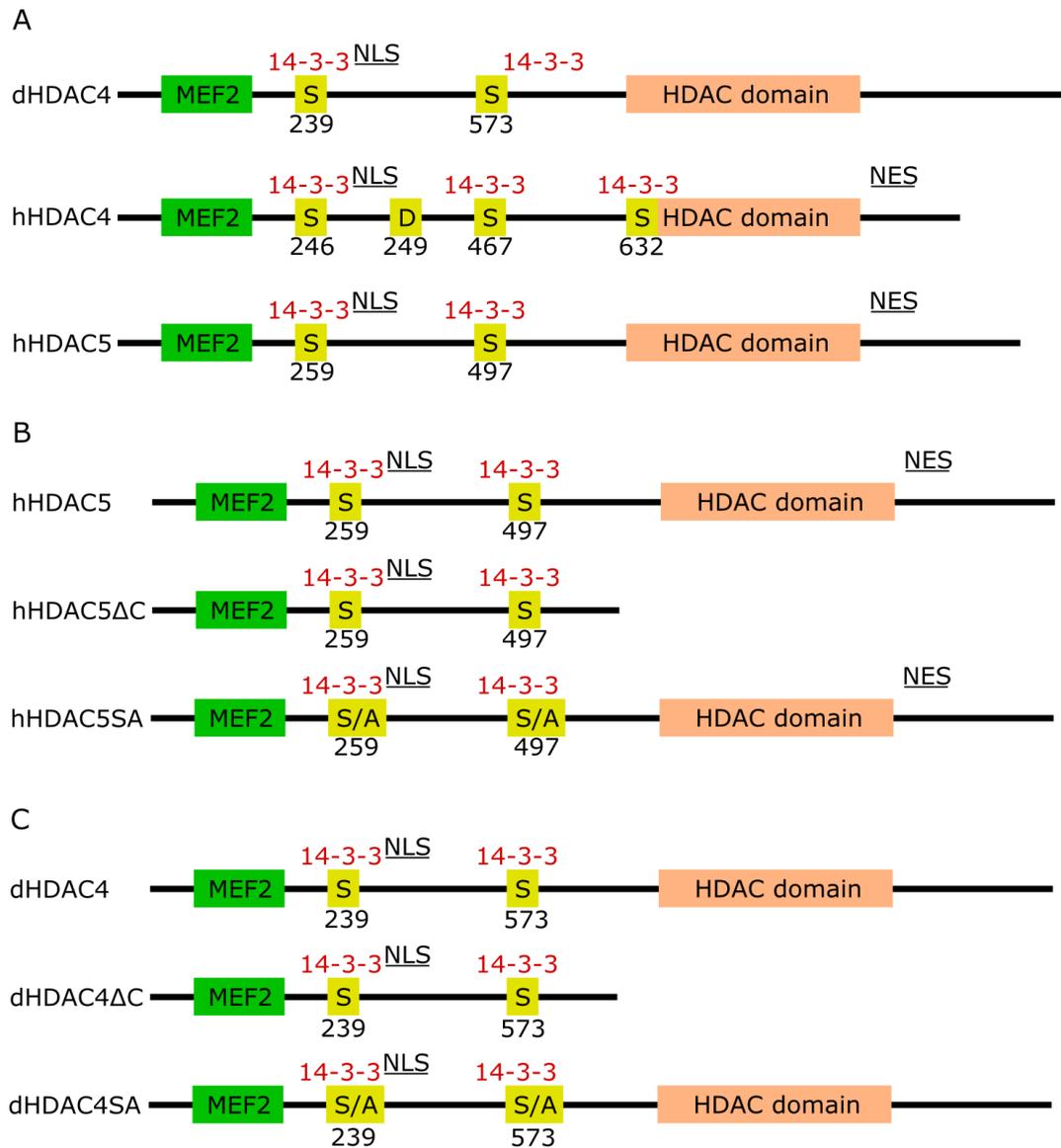


Figure 5.4: Structure of HDACs and transgenically expressed mutants

A) Structure of *Drosophila* HDAC4 (dHDAC4) in comparison to human class IIa HDACs 4 and 5.

B) Wildtype and mutant versions of human HDAC5 that were transgenically expressed in this study.

C) Wildtype and mutant versions of *Drosophila* HDAC4 that were transgenically expressed in this study.

MEF2, Mef2-binding domain; S, serine phosphorylation site for interaction with 14-3-3 proteins; NLS, nuclear localisation signal; NES, nuclear export signal; numbers below S denote residue number.

the class IIa HDACs were taken into consideration and transgenic variants were also tested in the same experimental set-up. The overexpressed protein should be able to shuttle between the nucleus and the cytoplasm as no modification was made to its nuclear export/import system. The deletion of a portion of the C-terminus results in the elimination of the HDAC domain which should create a version similar to Mef2-interacting transcription repressor (MITR) that contains only the non-catalytic N-terminus part of HDAC9 (Sparrow et al., 1999).

It is important to note that MITR is able to repress Mef2 mediated transcription, however not the one specific for muscle development (Zhang et al., 2001). MITR is a protein that is mainly nuclear and its inhibitory function is blocked when the two conserved Ser residues are phosphorylated causing the interaction with the 14-3-3 chaperone that alters MITR's nuclear distribution. Human HDAC5 was found to be able to repress Mef2 mediated transcription in absence of its HDAC domain as long as the two repressive regions were present (Lemercier et al., 2000). However the presence of HDAC5 in the cytoplasm is necessary in order to allow execution of the myogenic program (McKinsey et al., 2000). In the case of human HDAC4 both the Mef2-binding domain and the HDAC domains were required to repress muscle differentiation. The protein is maintained in the cytoplasm of myotubes by interaction with the 14-3-3 chaperone and when the myogenic signal retracts it is able to enter the nucleus, which most probably enables Mef2 to repress transcription of particular subsets of target genes (Miska et al., 2001).

In order to test the above mentioned characteristics in dHDAC4 and hHDAC5, both proteins were expressed in *Drosophila* embryos under the *twi-Gal4;twi-Gal4* driver. Taking into account the information available from literature, both proteins are expected to be able to shuttle between the cytoplasm and the nucleus based on their phosphorylation state of the two conserved serine residues. The *Drosophila* HDAC4 is assumed not to have a nuclear export signal (NES) in its C-terminus while all the human versions of Class IIa HDACs do (Yang and Grégoire, 2005). The NES is eliminated together with the HDAC domain in the hHDAC5 Δ C construct. In this situation, the variant of hHDAC5 is expected to be sequestered in the nucleus and be able to repress myogenesis since it can achieve this even in the absence of its catalytic domain. The exchange of Ser259 and Ser498 to alanine interferes with the

shuttling of hHDAC5 from the nucleus to the cytoplasm and it has been shown that the protein cannot be properly transported out of the nucleus in such mutants, therefore muscle development is expected to be repressed in these experiments (McKinsey et al., 2000). Since the hHDAC5 protein is expressed in a new system this mechanistic characteristics could be impacted and the effects could have less strong effects compared to cell culture.

In adult flies, dHDAC4 has been reported to bind Mef2 when overexpressed in so called nuclear bodies, which was not observed when dHDAC4 was not present in the nucleus or when it was not overexpressed (Fitzsimons et al., 2013).

To study the interaction between HDAC4 and Mef2 the two HDACs, dHDAC4 and hHDAC5 were overexpressed in *Drosophila* embryos. The effects on early myogenesis was studied on embryos undergoing mesodermal specification equivalent to stage 12. The effects on late myogenesis was studied in embryos of stages 16 in order to have the developmental age where the muscle pattern should be fully developed. Both the full length and the mutants were overexpressed under similar conditions. The readout of the overexpression results were the following: stage 12 embryos were immunostained with a Mef2 antibody to analyse effects on the protein level, the effects on a target gene of Mef2 in the somatic mesoderm was assessed via *in situ* RNA hybridisation experiments of *β3-tubulin*. In late stage myogenesis the effects on the β3-Tubulin protein levels was assessed via immunostaining for that protein.

β3-Tubulin has a cell-type-specific relationship with Mef2 in *Drosophila* embryos. The genes expression has a Mef2 binding site in its enhancer and full expression in the somatic muscle is dependent of the TF starting from stage 12 (Damm et al., 1998). The visceral enhancer of β3-Tubulin does not present a Mef2 binding site and its activation is independent of Mef2. Him is a negative regulator of Mef2 activity and is able to repress transcriptional activation of Mef2 target genes in somatic musculature (Liotta et al., 2007). β3-Tubulin is a Mef2 target that is responsive to Him mis-expression during muscle development. Since we hypothesise that dHDAC4 is also a repressor of Mef2 activity during myogenesis, β3-Tubulin is the target gene of choice to study effects of overexpressing Class IIa HDACs in *Drosophila* muscle.

The effects of hHDAC5 and its variants overexpression was studied in early myogenesis by looking at the effects on the levels of Mef2 protein and one of its targets *β3-tubulin*. Mef2 is expressed both in the visceral and the somatic mesoderm, however only in the somatic mesoderm Mef2 has *β3-tubulin* as a target gene. It is visible that both Mef2 protein levels and *β3-tubulin* expression are unaffected by overexpression of hHDAC5 and hHDAC5SA (Figure 5.5A). In the case of hHDAC5ΔC it is noticeable a reduction in *β3-tubulin* levels in the somatic mesoderm (Figure 5.5B).

In stage 12 embryos, when mesodermal specification occurs, it is observable that dHDAC4 and all its variants dHDAC4ΔC and dHDAC4SA are able to inhibit Mef2-regulated gene transcription by affecting expression of its target genes, but not of the Mef2 protein itself whose levels seem wildtype (Figure 5.6A). The full length dHDAC4 was able only partially to inhibit transcription of *β3-tubulin*, a Mef2 target gene in the somatic musculature (Figure 5.6B). The effects of overexpressing dHDAC4 in the *Drosophila* embryonic musculature are more severe compared to the effects of gain of function experiment of hHDAC5.

In late stage myogenesis the effects on muscle development are more severe (Figure 5.7). To quantify the effects of the overexpression experiments the muscle pattern phenotype was split into 5 categories: 1) wild-type: the muscle pattern presents no defects; 2) weak: a very small number of muscles are affected, mostly displaying defects in shape but the pattern appears mostly wild-type; 3) moderate: Most muscles are present and correctly shaped; 4) Severe: The muscle pattern is significantly affected, many muscles are misshapen or missing but a slight outline of the expected muscle pattern is still observable; 5) Extreme: No muscle is formed correctly.

The effects in late myogenesis are stronger than in early development, however the pattern of having stronger effects while overexpressing dHDAC4 compared to overexpressing hHDAC5 is maintained. The dHDAC4 embryos had very severe to extreme phenotypes of muscle disruption in late embryos. The least severe of the three versions was observed when overexpressing dHDAC4ΔC since there were still some muscle fibres at the stereotypical location that could give an impression of a reminiscent muscle pattern. In dHDAC4 and dHDAC4SA the embryos were missing

most muscles and it was not possible to conclude which type of fibre the fragments originated from. These muscles had very little plasticity, did not maintain the usual rounded shape and the cuticle broke easily when handled.

In contrast, the muscle pattern when overexpressing hHDAC5 and its mutants was still recognisable. The hHDAC5 gain-of-function embryos had a very weak phenotype, only the ventral muscles displaying a shaping problem due to extension of these muscles into the ventral side of the embryo more than in wildtypes. The consequence of this defect resulted in the embryo appearing less rounded and more elongated on the ventral side. Both the hHDAC5 Δ C and the hHDAC5SA muscle phenotype could be classified as severe since most of the muscles were missing or misshapen, but the gross pattern was still observable. The hHDAC5SA was comparable to the severity of dHDAC4 Δ C phenotype, while the hHDAC5 Δ C seemed to be milder. When looking at the Mef2 expression in late embryogenesis it is clear that most of the nuclei are not properly positioned in these mutants.

Based on the phenotypes obtained in early development it is possible to conclude that most probably class IIa HDACs do not play a very strong role in muscle progenitor cells to repress Mef2 activity. It is likely that the default localisation for HDAC4 in myoblasts is in the cytoplasm as it was observed in the colocalisation study (section 5.2.2). The same effect is most probably applicable for hHDAC5 in this system. The lack of effects on muscle specific genes expression early in myogenesis when the full-length HDACs are overexpressed could support an active sequestration of the class IIa HDAC in early myogenesis in order to allow muscle differentiation, a pattern which seems to support the mechanistic model proposed of hHDAC5 in mammalian cell culture. The fact that there is a reduction in *β 3-tubulin* expression for the mutants lacking a HDAC domain suggests that dHDAC4 is able to effect muscle repression despite lacking its catalytic domain. The lower effects in myogenesis observed for hHDAC5 could be due to the fact that the mechanisms regulating class IIa HDACs in *Drosophila* are able to compensate for the gain of function. The lack of effects in constitutively nuclear-localised hHDAC5 could be due to other proteins that are able to block the activity of hHDAC5SA in the nucleus.

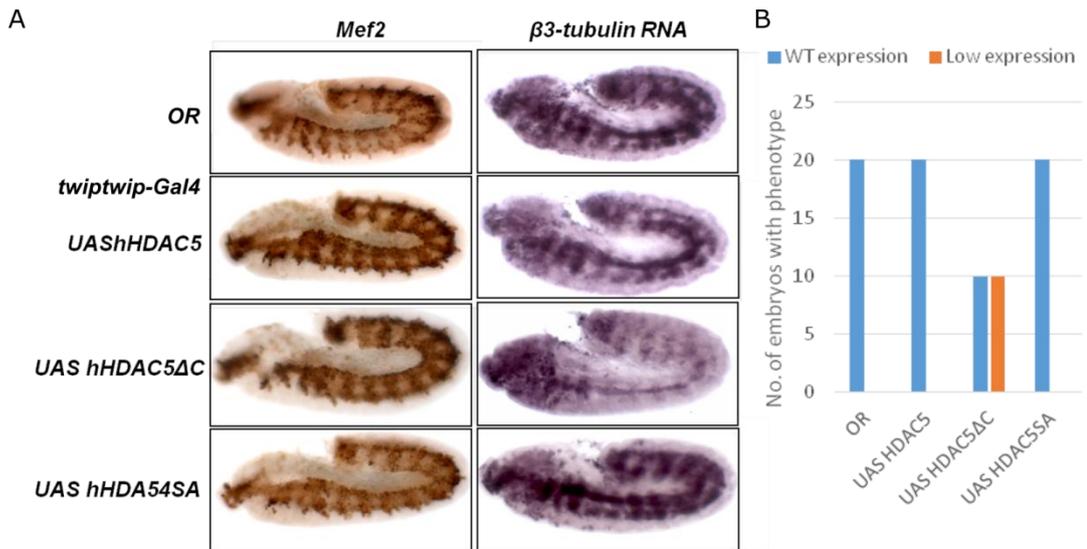


Figure 5.5: HDAC5ΔC overexpression downregulates *Mef2* activity

Expression of *UAS HDAC5*, *UAS hHDAC5ΔC*, *UAS hHDA54SA* driven by *twi-Gal4*; *twi-Gal4* at 25°C. The overexpression of human HDAC5 and its modified versions HDAC5ΔC (a HDAC4 version lacking its histone deacetylase domain) and HDAC5 SA (lacks the phosphorylation site ensuring shuttling of the HDAC between the cytoplasm and the nucleus) does not affect *Mef2* protein expression, but HDAC5ΔC downregulates its $\beta3$ -*tubulin* target gene expression in the somatic mesoderm. *Mef2* protein is visualised by an immunostain of stage 12 embryos. Expression of a *Mef2* target in the developing somatic muscle, $\beta3$ -*tubulin*, is visualised by *in situ* hybridisation of stage 12 embryos. OR, Oregon-R (wildtype).

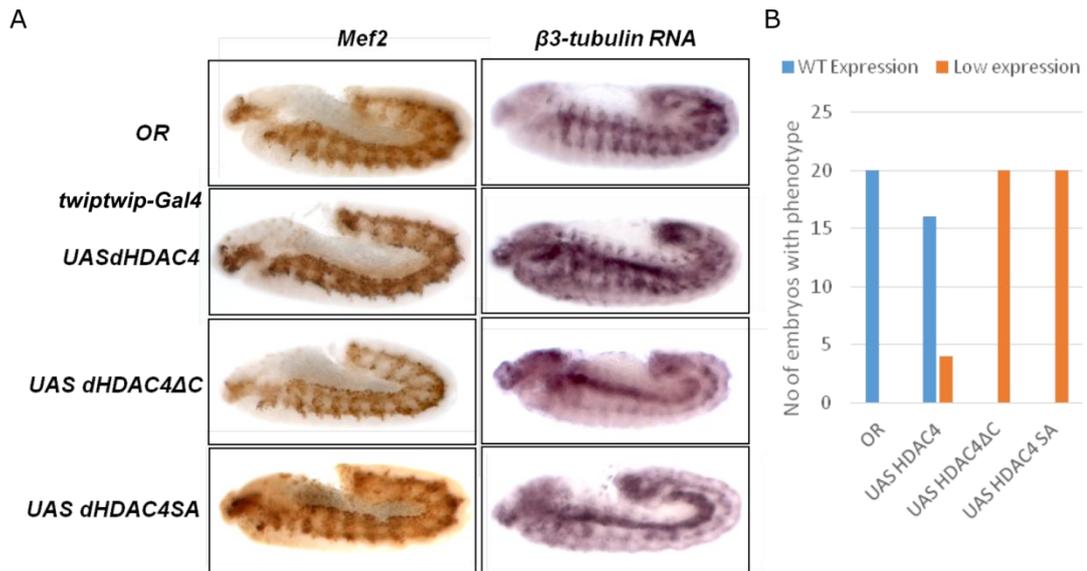


Figure 5.6: HDAC4ΔC overexpression downregulates Mef2 activity

Expression of UAS-HDAC4 driven by *twi-Gal4*; *twi-Gal4* at 25°C. The overexpression of HDAC4 and its modified versions HDAC4ΔC (a HDAC4 version lacking its histone deacetylase domain) and HDAC4 SA (lacks the phosphorylation site ensuring shuttling of the HDAC between the cytoplasm and the nucleus) does not affect Mef2 protein expression, but downregulates its *β3-tubulin* target gene expression in the somatic mesoderm. Mef2 protein is visualised by an immunostain of stage 12 embryos. Expression of a Mef2 target in the developing somatic muscle, *β3-tubulin*, is visualised by *in situ* hybridization of stage 12 embryos. OR, Oregon-R (wildtype).

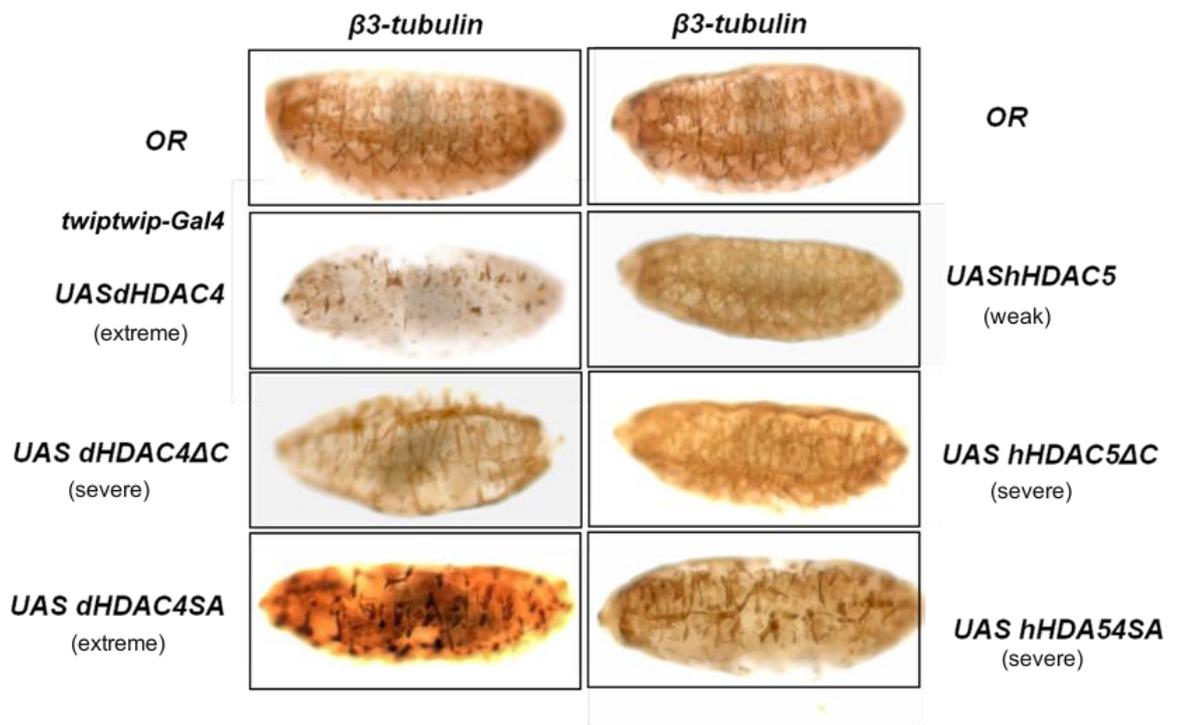


Figure 5.7: Overexpression of *Drosophila* HDAC4 and human HDAC5 affects muscle differentiation in late stage embryos

Stage 16 embryos were immunostained for β 3-Tubulin protein. Descriptions in brackets denote the phenotype category caused by each transgene. At least 10 embryos were examined for each condition. Quantification was omitted since phenotypes were consistent. OR, Oregon-R (wildtype).

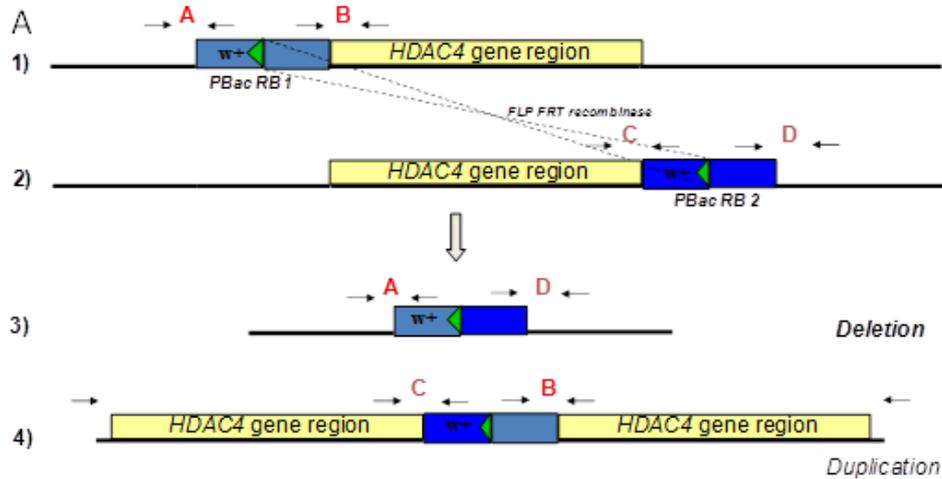
5.2.4 Generation of targeted deletion HDAC4 null mutants

A classic genetic approach to functionally characterise the role of a specific molecule is to do gene perturbation studies, in which the molecule of interest is either overexpressed or its activity is impaired. The loss-of-function approach can be achieved through two methods: 1) by generating a null mutant of the protein of interest; 2) the use of a knockdown approach such as RNA interference, in which the expression of the protein of interest is inhibited at RNA level. The second approach is dependent on the existence of a specific Gal4 driver for the tissue of interest, as well as the strength of that driver. A targeted deletion of HDAC4 locus was attempted in order to obtain a null mutant that would not depend on the strength of the driver.

A targeted deletion generated by P-element transrecombination was used to generate *HDAC4* null mutants. There were three potential P-elements within the *HDAC4* region that allowed partial or complete deletion of the gene (Figure 5.8C). Recombination between compatible pBAC{RB} elements can result in either a duplication or a deletion of the DNA region between the two inserted elements. The pBAC{RB} e04575 element inserted within the *HDAC4* gene region was paired with pBAC{RB} e02449 located upstream of the *HDAC4* region to generate a partial deletion of HDAC4 targeting the N-terminal part of the protein.

The recombination of the elements pBAC{RB} e02449 and the pBAC{RB} e03932 (downstream of the *comt* gene) allowed complete removal of *HDAC4* together with the *comt* gene. The compatibility of the combined sites was given by the directionality of their FRT that allow transrecombination by the Flippase enzyme. The recombination experiments were performed by Ms Jun Han.

Because all elements contained *w+* insertions, eye colour could not be used as a marker to screen for a recombination event. A red *w+* eye colour would be observable in the event of a duplication, a deletion or the occurrence of an unrecombined element. Recombination events were screened solely by PCR using a combination of a P-element specific primer and a genomic primer targeting the region around the element. A successful deletion event took place only when a PCR



B

Event	Primer pair give product?			
	A	B	C	D
1) RB 1 no rec	yes	yes	no	no
2) RB 2 no rec	no	no	yes	yes
3) DELETION	yes	no	no	yes
4) Duplication	no	yes	yes	no

C

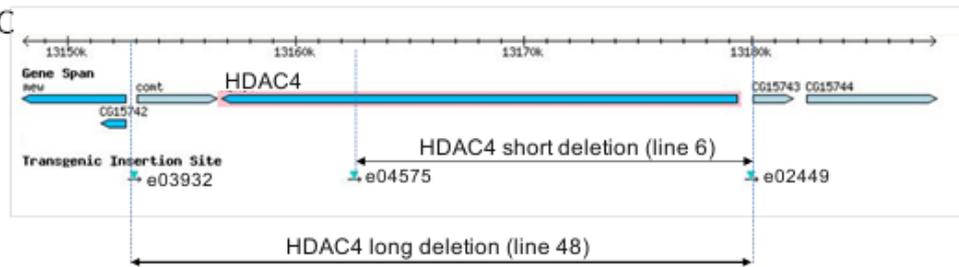


Figure 5.8: HDAC4 deletion screening

A) Recombination strategy with recombination event taking place at the *HDAC4* locus during the FLP/FRT recombination.

B) Possible combinations of primer pairs and their interpretation during screening.

C) Deleted genomic regions in the two lines that scored positive during the screening process (they had a positive PCR A and D and a negative PCR B and C).

product for the region targeting the downstream region of the first P-element, respectively the upstream region of the second element was obtained.

For each combination of P-elements 50 different crosses were tested for recombination events. Most of the lines were non-recombination events, less than 5% resulting in duplication or deletions. Only line 6 for the e04575/ e02449 recombination and line 48 for the e03932/ e02449 combination were confirmed to contain the targeted deletions. Because the two lines presented no red eyed males it was concluded that the two deletions were lethal, therefore stocks were maintained over an FM7 balancer.

In order to check if the recombination events successfully deleted the HDAC4 regions predicted, we performed PCR studies in which two primer pairs targeting a part of the deleted region were tested. PCR was performed on previously screened homozygous null embryos since the balanced embryos would give a positive fragment for the deleted region.

Based on the PCR screen performed, it was confirmed that the HDAC4 deletion 6 line is indeed a null mutant for the HDAC4 gene, while further analysis of the HDAC4 deletion 48 line is required, since the deletion 48 genomic DNA line gives a positive fragment in the region of bPBAC (Figure 5.9B), although if the deletion event was successful that part of the genome should not exist in the mutants. It is important to note that the positive fragment is observed in the PCR where the genomic DNA came from embryos, while in the pupa the fragment for the bPBAC primer region is negative as expected if the mutant HDAC4 deletion 48 is a true deletion. A potential contamination with a heterozygous embryo of the genomic DNA could be the cause of this observed fragment.

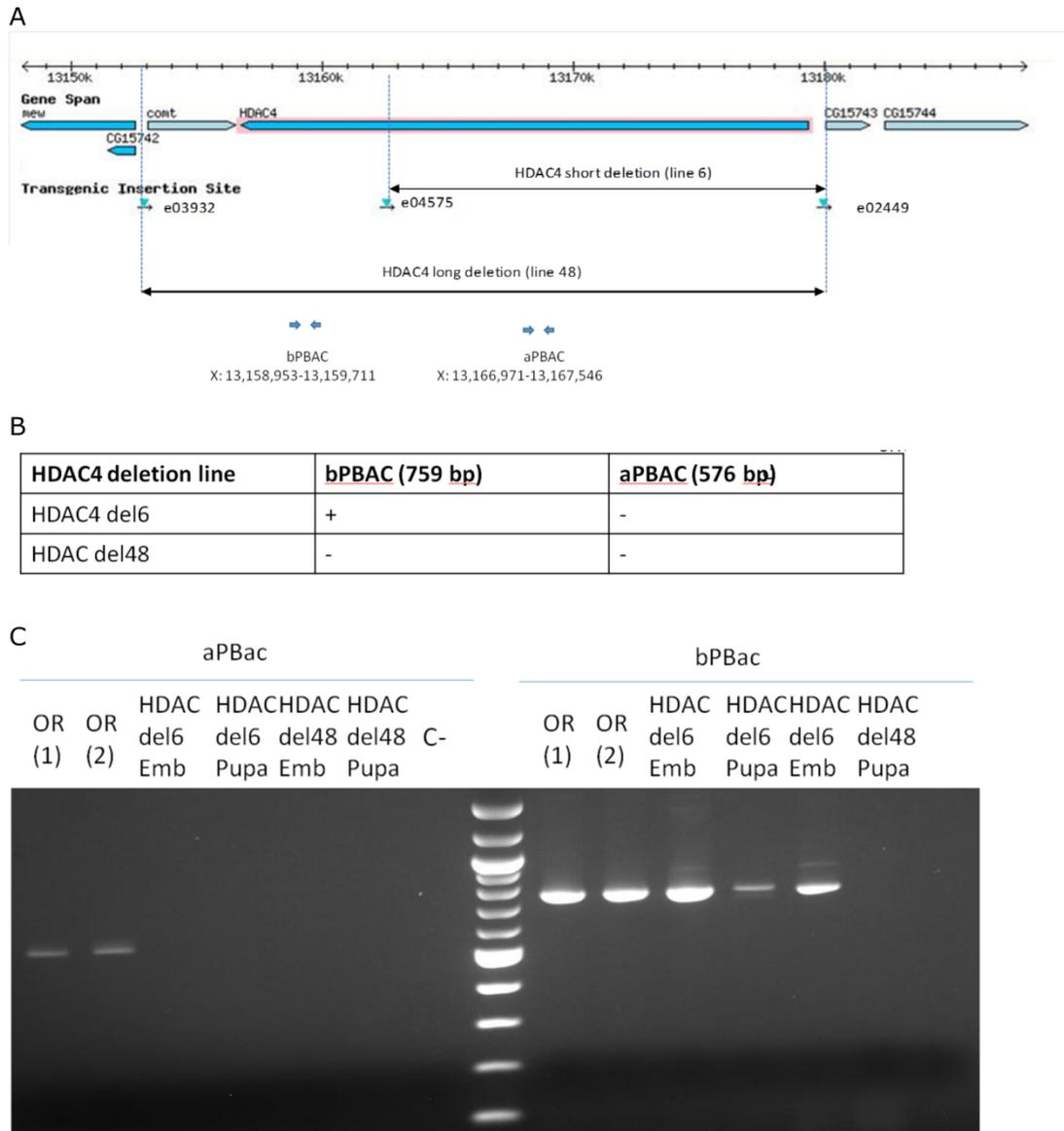


Figure 5.9: HDAC4 successful deleted region screening

A) Map of the location of the two deleted regions in the genome.

B) The table shows possible combinations of primer pair and the expected for a successful generation of a null mutants.

C) Agarose gel showing PCR results of fragments located inside the deleted regions. Samples: Positive control: OR1 (genomic DNA from Oregon-R embryos), OR2 (genomic DNA from OR pupa), HDAC4 del6 Emb/Pupa (genomic DNA extracted from HDAC4 deletion 6 homozygous embryos/pupa), HDAC4 del48 Emb/Pupa (genomic DNA extracted from HDAC4 deletion 48 homozygous embryos/pupa), Negative control: C- (H₂O instead of genomic DNA template).

5.2.5 Hatching and survival assay of potential HDAC4 null mutants

All the investigated lines were homozygous lethal at adult stage and in order to test at what stage the lethality occurred a hatching and survival assay was performed for each line confirmed positive as a potential HDAC4 deletion. The lethality of the two HDAC4 targeted deletions was initially determined based on the absence of any red eyed males in the initial cross, only white eyed males being viable.

In order to distinguish the homozygous embryos from the heterozygous ones, the heterozygous maintained stocks were crossed over an FM7,ActGFP balancer. Based on the fluorescence of the ActGFP marker it was possible to select only the homozygous embryos for further inspection. Since the homozygous embryos should not carry the GFP marker, embryos that did not show fluorescence were selected for the hatching and survival assay. The autofluorescence of the chorion membrane made it very difficult to separate the homozygous embryos from the rest at early stages, therefore the embryos were allowed to reach stage 14 before selecting the homozygous embryos.

The hatching and survival assay performed on homozygous mutants of the two potential null mutants showed that both lines die at pupal level. Only 35% for the HDAC4 del 48 and 42% for the HDAC4 del6 reach this developmental stage. None of the homozygous flies develop to adulthood.

Lethality is one of the most striking phenotypes observed in these HDAC4 null mutants and in order to show that this phenotype is due to HDAC4 loss of function, a rescue experiment of HDAC4 was attempted. The rescue of HDAC4 deletion 6 lethality was attempted.

5.2.6 Rescue of lethality of HDAC4 null mutants

In order to prove that lethality is due to HDAC4 deletion rescue experiments using a UAS-HDAC4 construct were designed. A rescue experiment for deletion HDAC4 deletion 6 using a Da-Gal4 universal driver was attempted. The deletion stocks are maintained over FM7 balancers due to the lethality of the mutated X chromosome, which means that the deletion is transmitted from generation to generation in females, but not in males (all males in the stable line have FM7). If overexpressing

HDAC4 rescues the deficiency, then it would be possible for males with the mutated X chromosome to survive. The characteristic of males containing the FM7 balancer is the presence of Bar eyes. The presence of males having wild-type eyes was therefore used to screen for males with a successfully rescued deletion (Figure 5.11).

The rescue experiment performed was not successful since no wild-type eyed males were observed. Only bar eyed progeny, both male and females resulted from the final cross. The strength of the DaGal4 driver can influence the outcome of the rescue experiment, therefore the use of another ubiquitous driver is advisable before it can be concluded whether this X chromosome is lethal due to HDAC4 deficiency or for other reasons.

However, it was possible to obtain viable males with the mutated HDAC4 deletion 6 chromosome by crossing the heterozygous null mutants females with a line containing an insertion of the deleted region of the X-chromosome on another chromosome (Figure 5.12). However, because the insertion does not exclusively contain the deleted region, but also neighbouring regions, it is not possible to finally conclude that the lethality phenotype is due only to the HDAC4 gene deletion.

5.2.7 Muscle phenotype analysis for the potential HDAC4 null mutants

To test whether the deletion of HDAC4 disrupts muscle development, homozygous HDAC4 deletion embryos were screened for muscle-related phenotypes. The types of muscles analysed included: somatic, visceral, heart and pharyngeal. Any difference in muscle pattern formation was determined by qualitative phenotypic comparison with wild-type embryos. All embryos analysed are homozygous for the deletion as indicated by absence of the *ftz-lacZ* element indicative of the FM7 balancer. Both deletion mutants showed no disruptions in the somatic muscle pattern (Figure 5.13). All the other types of muscle tissue did not show any defects.

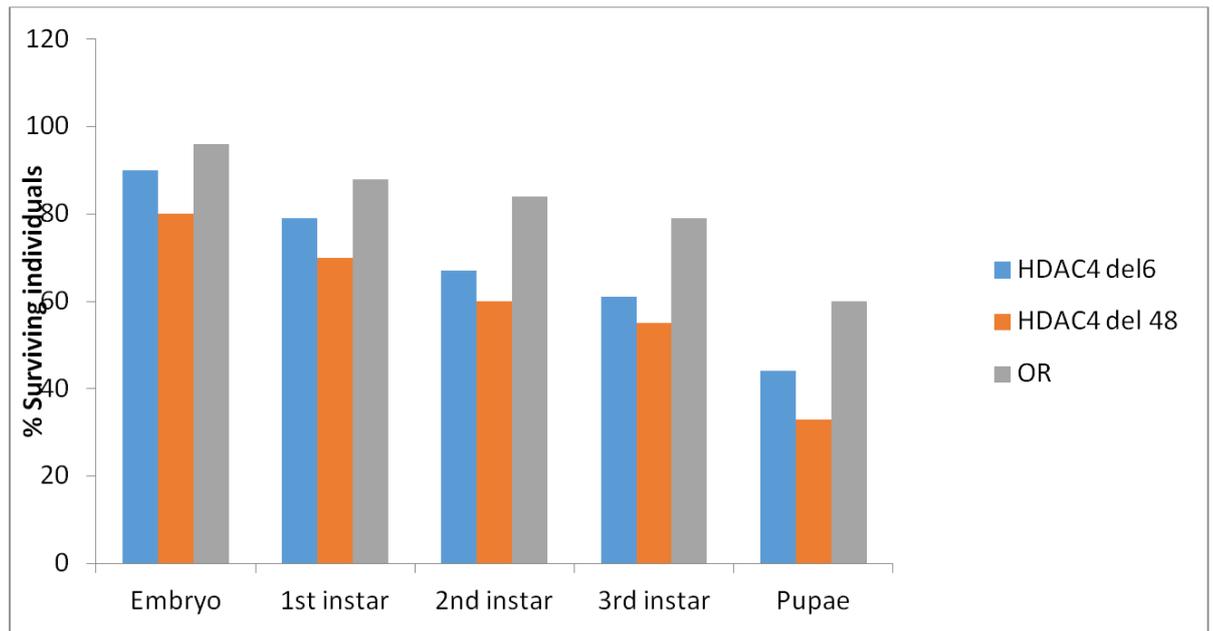


Figure 5.10: Hatching and survival assay of potential HDAC4 null mutants

The HDAC4 deletion lines were balanced over a fluorescent balancer chromosome to allow screening for homozygous null embryos. Embryos collected from each line were counted and allowed to develop. Once the cultures reached each of the given stages, the remaining surviving individuals were counted. The hatching and survival assay performed on homozygous mutants of the two potential null mutants showed that both lines die at pupal level. Only 35% for the HDAC4 del 48 and 42% for the HDAC4 del6 reach this developmental stage. None of the homozygous flies develop to adulthood.

1a) V♀ HDAC del6/FM7 ftz lacZ x ♂ FM6/Y;; TM3 Ser/Sb

Select v ♀ HDAC del6/FM6;;TM3 Ser/+

1b) V♀ w1/FM6;;TM3Ser/Sb x ♂ w/Y;;DaGal4/DaGal4

Select ♂ FM6/Y;;Dagal4/Sb

2a) v ♀ HDAC del6/FM6;;TM3 Ser/+ x ♂ FM6/Y;;Dagal4/Sb

Select v ♀ HDACdel6/FM6;; DaGal4/TM3Ser

2b) V♀ w1/FM7grhlacZ; sco/CyO x ♂ w/Y; UASHDAC4/ UASHDAC4;
TM3/TM6

Select ♂ FM7 grhlacZ; UASHDAC4/sco; TM3/+ or TM6/+

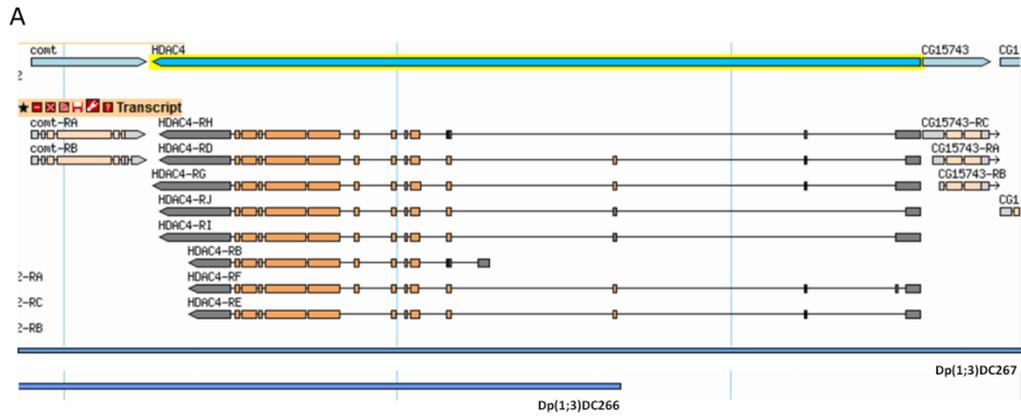
3) v ♀ HDACdel6/FM6;; DaGal4/TM3Ser x ♂ FM7 grhlacZ; UASHDAC4/sco;
TM3/+ or TM6/+

If rescue worked: ♂ HDACdel6/Y;UASHDAC4/+;DaGal4/+ or TM3 or TM6

(i.e. if there are males with wildtype eyes the rescue worked)

Figure 5.11: Genetic scheme of attempted rescue experiment of HDAC4 del6 using a UAS-HDAC4 construct and a DaGal4 driver

All the crosses, including the overexpression experiment were performed at 25°C.



B

Mutant	Dp(1;3)DC266		Dp(1;3)DC267	
	FM7/Y	HDACdel/Y	FM7/Y	HDACdel/Y
HDAC4 del6	13	4	5	1
HDAC4 del48	13	18	-	-

Figure 5.12: Rescue of HDAC null mutants by genomic duplications of the X chromosome

A) HDAC4 locus and regions replicated by the duplications Dp(1;3)DC266 and Dp(1;3)DC267, lines available commercially.

B) Number of surviving HDAC4 null hemizygous males compared to males with hemizygous balancer. Duplication of each region of X on the third chromosome.

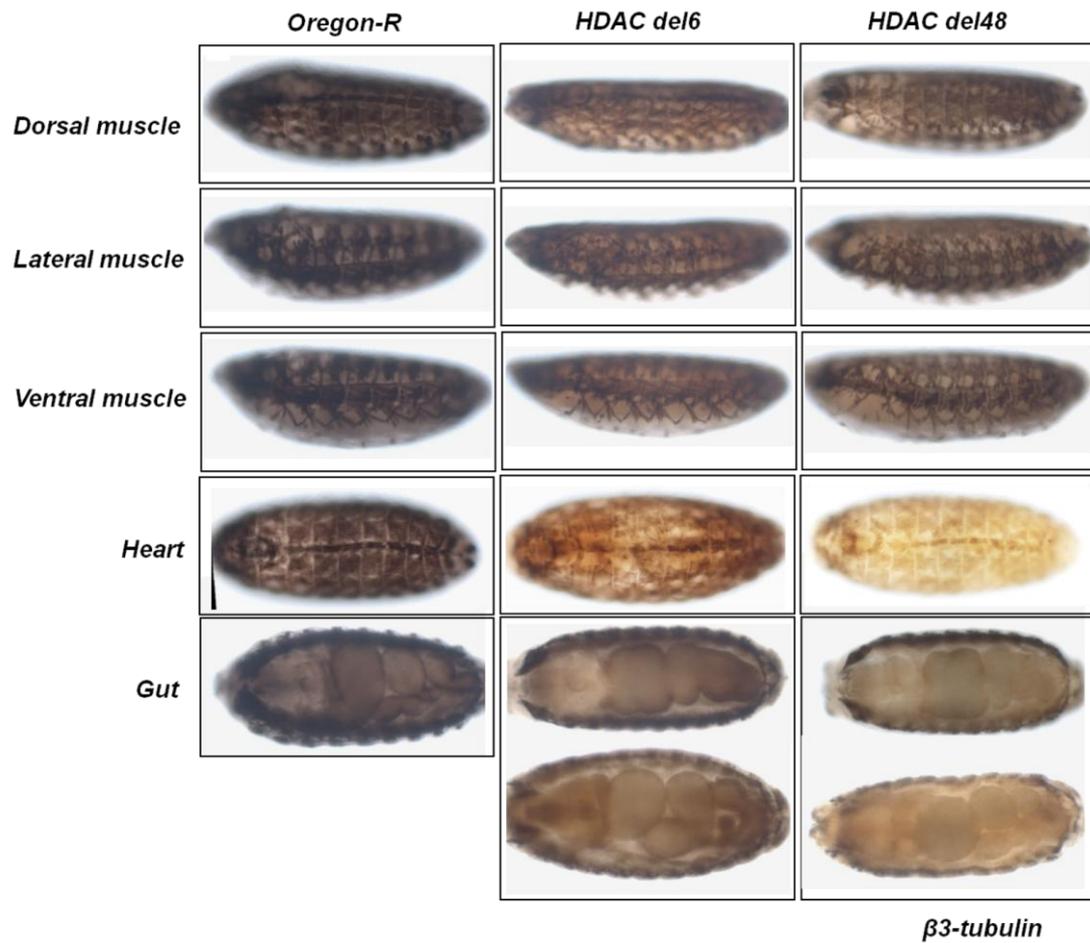


Figure 5.13: Muscle phenotype in HDAC deletion mutants

Immunostain against β 3-Tubulin in late-stage embryos. Genotype of the embryos as indicated in each column. The embryos with HDAC4 deletions were screened against *ftz-lacZ* for homozygosity.

5.3 Discussion

5.3.1 HDAC4 is expressed in an unspecific pattern in *Drosophila* embryos and does not localise to the nucleus of muscle precursors

The HDAC4 protein was discovered in the 11-13MEF sample as a possible interactor of Mef2. The interaction of class IIa HDACs with Mef2 in vertebrates has been widely studied, including in myogenesis, however little is known by the interaction of Mef2 with HDAC4 in *Drosophila* embryos. Class IIa HDACs are a class of HDAC isoform that shuttle between the nucleus and the cytoplasm. In vertebrates, they are specifically expressed in certain tissues like muscle, heart, brain or neurons, unlike other classes of HDACs. In mammalian cells there are four class IIa HDACs: HDAC4, 5, 7 and 9. In their nuclear roles these proteins have a repressive role and are involved in some developmental and differentiation processes. HDAC4 regulates chondrocyte hypertrophy, skeletal development and neural death (Vega et al., 2004; Bolger & Yao, 2005). HDAC7 is involved in regulation of T-cell development and vascular integrity (Kasler & Verdin, 2007; Bolden, Peart & Johnstone, 2006). The last two class IIa HDACs are responsible in cardiac hypertrophy (Zhang et al., 2002).

It was therefore surprising to find that HDAC4 in *Drosophila* embryos did not show specificity to any particular tissue. Furthermore, we found no colocalisation between Mef2 and HDAC4. Although HDAC4 seems to be expressed throughout the embryo, it was not present in the nucleus in any muscle-related cells at any stage of embryogenesis. Unless the activity of HDAC4 in muscle cells can be confirmed through other assays, the interaction between Mef2 and the class IIa HDAC4 in the *Drosophila* muscle remains elusive. The presence of *Drosophila* class IIa HDAC in both nucleus and cytoplasm has only been confirmed in S2 cells (Foglietti et al., 2006), and *Drosophila* adult neurons (Fitzsimons et al., 2013). In its subcellular localisation, dHDAC4 therefore seems more similar to hHDAC4, which was also distributed in the cytoplasm in cultured cells but could be shuttled into the nucleus by Mef2 (Borghi et al., 2001).

5.3.2 *Drosophila* HDAC4 is able to repress myogenesis in embryos

Nucleocytoplasmic shuttling is an important mechanism to regulate the activity of class IIa HDACs in mammals. When present in the nucleus these enzymes are able to repress transcription by interacting with nonhistone targets and not by performing a deacetylation function. The intracellular localisation of the class IIa HDACs seems to be dependent on signalling cues and upon being phosphorylated they are excluded to the cytoplasm (Bertos et al., 2004; Khochbin et al., 2001; McKinsey et al., 2001, 2002b). The default intracellular localisation of HDAC5 and HDAC7 is nuclear in proliferating myoblasts and upon differentiation they are transferred to the cytoplasm (Dressel et al., 2001; McKinsey et al., 2000). HDAC4 on the other hand is cytoplasmic in undifferentiated myotubes and when fusion has occurred it is translocated to the nucleus of C2C12 cells. In myotubes HDAC4 is found both in the cytoplasm and the nucleus. When present in the nucleus HDAC4 can change the pan-nuclear pattern of Mef2 into distinct clusters of nuclear bodies, a characteristic that is found in other class IIa HDACs like HDAC5 and -7 (Downes et al., 2000). HDAC4 can be actively and rapidly transported outside of the nucleus and it seems that the number of Mef2 molecules present in the nucleus could act as the limiting factor that stops HDAC4 translocation into the cytoplasm (Chan et al., 2003).

dHDAC4 is mainly cytoplasmic throughout the *Drosophila* developing embryonic stage. A similar pattern was observed in neurons of *Drosophila* adult brains when dHDAC4 was overexpressed and its nuclear localisation was noted in only a subset of neuronal nuclei. A punctate redistribution of Mef2 localisation pattern in the nuclei of certain subsets of neurons when dHDAC4 was overexpressed was also detected (Fitzsimons et al., 2013). dHDAC4 is also expressed in the *Drosophila* larval fat body and its intracellular localisation is regulated by dietary status. Upon fasting dHDAC4 is shuttled into the nucleus, while under feeding conditions dHDAC4 is phosphorylated by SIK3 and translocates to the cytoplasm (Wang et al., 2011). All these data from *Drosophila* seem to express that the dHDAC4 is responsive to the same regulatory mechanisms identified for class IIa HDACs in mammalian cell culture. The line we used to study the colocalisation of Mef2 with dHDAC4 in the embryo expressed dHDAC4 at physiological levels, aspect which could impact the ability to identify nuclear expression of dHDAC4 if the levels are

under the detection limit of the confocal microscope. Considering that dHDAC4 seems to be mainly cytoplasmic and its nucleus import occurs under very specific conditions, and due to sequence similarity, we postulate that dHDAC4 is regulated by similar mechanisms that control the activity of the mammalian HDAC4.

When overexpressed in the embryo, dHDAC4 was able to impact the development of somatic musculature and the expression of B3-tubulin, a gene that is both a Mef2 target and that expresses a structural protein that can be used as a marker for staining the muscle pattern in *Drosophila* embryos (Damm et al., 1998). The direct interaction between Mef2 and dHDAC4 was also identified in our pull down study from *Drosophila* embryos where Mef2 acted as the bait. Taken together, we can hypothesise that the two proteins are able to interact in the nuclei of muscle cells to repress the expression of certain Mef2 target genes. Mammalian HDAC4 was shown to specifically modulate the activation of structural and contractile genes that are expressed in a Mef2-dependent manner in muscle (Cohen et al., 2009). The expression of B3-tubulin in muscle was downregulated when dHDAC4 was overexpressed in *Drosophila* muscle. This experiment seems to highlight that the group of structural proteins activated by Mef2 can have their expression repressed by dHDAC4. The larval musculature fails to form properly when dHDAC4 is overexpressed, therefore it seems dHDAC4 is able to affect muscle integrity.

Nucleoplasmic shuttling is an important mechanism to regulate dHDAC4 activity in *Drosophila*. The predominant cytoplasmic localisation of dHDAC4 could be related to a rapid translocation of the protein from the nucleus. This could be the reason why the effects of the full length dHDAC4 overexpression is less penetrant compared to the forms of dHDAC4 that cannot be exported from the nucleus. Additionally, the effects of dHDAC4 overexpression worsen as the development process gets further along. This aspect emphasises that there needs to be a balance of input from dHDAC4 throughout the muscle development process, most probably highly dependent on signalling cues.

The two modified versions of dHDAC4 cannot be shuttled out of the nucleus and have a stronger ability to repress Mef2 activity and muscle development. This effect is visible early in muscle development since expression of Mef2 genes seems to be impacted starting from the myoblast stage. The absence of the HDAC catalytic

domain does not seem to impact the ability of dHDAC4 to repress transcription during myogenesis. The four Class IIa HDACs from vertebrates also present an N-terminal domain that is able to repress Mef2 mediated transcription. Mammalian HDAC4 was also shown to repress expression of structural proteins even when overexpressing a form that lacks the phosphorylation prone serine residues required for nuclear translocation or a deacetylase defective form of the protein (Cohen et al., 2009).

The overexpression of the human HDAC5 protein in the *Drosophila* embryo seems to have some limited effects in *Drosophila* myogenesis. The consequences are visible in later stages of development, not from the onset of myogenesis. The limited penetrance could be given by the fact that the hHDAC5 and dHDAC4 are not as similar and the affinity of *Drosophila* Mef2 is reduced for this mammalian class IIa HDAC. When looking at the overexpression experiments of the three variants of hHDAC5 in *Drosophila* embryos, only the hHDAC5 Δ C overexpression had any effect on transcriptional regulation of B3-tubulin. Considering that both hHDAC5 Δ C and hHDAC5SA should be mainly nuclear, a repression phenotype would be more likely for the two variants compared to the case of overexpressing the full length protein. The lack of a repression phenotype on transcription observable in stage 12 embryos for the phosphorylation defective hHDAC5 could be due to the fact that hHDACSA has a reduced binding affinity, closely comparable to the one of full length hHDAC5, but lower when compared to hHDAC5 Δ C. The effects of overexpressing hHDAC5 and all its variants are mainly observable in late stage myogenesis. The mainly nuclear variants have a stronger capacity in inducing muscle damage compared to the full length protein. The defects observed so late in development could be due to the fact that the dHDAC4 levels present in the nucleus need to be tightly regulated throughout the development process and the constant increased levels surpass the capacity of compensatory mechanisms available. Overall the effects of overexpressing the human HDAC5 in *Drosophila* embryos is reduced compared to the overexpression of the *Drosophila* version itself. The degree of effects observed could be a consequence of affinity binding differences between hHDAC5 and dHDAC5 and/or due to the fact that human HDAC4 is a more appropriate orthologue for the *Drosophila* HDAC4.

Similarly to other class II a HDACs, dHDAC4 was found capable of repressing myogenesis in *Drosophila* embryos. The overexpression of constitutively active versions of dHDAC4 showed that it can repress Mef2 target genes even in absence of its HDAC domain. This behaviour is thus conserved with vertebrate class IIa HDACs, in which a tyrosine residue in the catalytic site is exchanged to histidine, reducing the deacetylase activity to close to none. Instead, class IIa HDACs function as corepressors of transcription factors in mammals (Fischle et al., 2002). The overexpressions have also shown that shuttling the protein out of the nucleus is necessary to allow muscle differentiation. Considering this ability to repress muscle development, dHDAC4 could function like hHDAC4 in cell culture in that Mef2 can stop its shuttling outside of the nucleus (Fischle et al., 2002). Another alternative could be this suggests that muscle precursors have a regulatory mechanism excluding dHDAC4 from the nucleus to prevent it from disrupting myogenesis. If this was the case, a cytoplasmic localisation as found here would be expected.

5.3.3 HDAC4 null mutants are lethal but have a normal muscle phenotype

After establishing that overexpression of HDAC4 has negative effects on embryonic myogenesis, we postulated that decreasing the amount of HDAC4 would release the constraint on the muscle tissue and potentially result in excessive muscle growth. However, mutants which do not express a functional HDAC4 showed wild-type muscle patterns. This suggests that wild-type levels of HDAC4 are required for normal formation of muscles and that the loss of this protein either is compensated by other regulatory mechanism or by the presence of maternally deposited HDAC4. The maternal contribution could be addressed by generating germline clones and studying gene expression of Mef2 target genes in such mutants.

Despite the absence of a muscle phenotype, HDAC4 null mutants were not viable. It is not possible to assess why the null mutants do not reach adulthood. It is known that embryos that do not form a muscle pattern are unable to hatch. However, the muscle pattern was normal in these embryos and can therefore be excluded as a cause for lethality. These data also contradict a previous suggestion that HDAC4 is involved in patterning the segments of the embryo (Zeremski et al., 2002). Taking into consideration the broad expression in the embryo and the lack of myogenesis defects it seems that HDAC4 might impact another tissue.

Chapter 6: General discussion

6.1 The main findings of the work

The work presented here explored types of Mef2 interacting proteins in the development of muscles in *Drosophila melanogaster* embryos. The central aim of my work was to conduct a large scale purification screen across embryonic development to identify Mef2 interacting proteins in a high throughput manner. This led to a three-part structure of my work. The first part was an exploratory Mef2 protein purification study under different conditions, which established the ground work for systematic extraction of Mef2 interacting proteins from *Drosophila* embryos and thereby enabled the appropriate investigation of the results of the screen. The second step was the analysis of the candidate lists based on computational approaches and literature data available and the validation of its results. The third part was the investigation of the function of HDAC4 in myogenesis, a candidate identified in the screen as potentially unique among Mef2 interacting proteins.

This work presents the first comprehensive study of Mef2 interacting proteins in *Drosophila* embryos, taking into account the expression of the bait protein under physiological conditions, and attempting to compile a complete list of candidates both from tightly staged embryos and a sample covering all developmental stages. The initial analysis (described in chapter 3) mainly answered the question: 1) Can we extract Mef2 interacting proteins from *Drosophila* embryos, if the bait is expressed at "normal" levels? The answer was affirmative, extraction of candidates being possible both from late-staged embryos and an "average" sample that spanned all developmental stages. These lists of proteins pulled down with Mef2 had both overlapping and unique hits.

The second question considered: 2) How to analyse lists of candidate proteins generated via a purification study to better understand the Mef2 interactome? Two approaches were adopted to answer this question: a) a narrowing down of a candidate list extracted from staged embryos based on spectral counts, the use of genome wide datasets to explore links between them and to extract an interesting candidate for further biological testing; b) a general approach where the obtained Mef2 interactome particularities are explored. Both approaches proved valuable in understanding which candidates have interesting biological functions as it relates to

Mef2. Approach (a) highlighted two interesting candidates (HDAC4 and Brahma) found in the nucleus that potentially modulate with Mef2 target gene activation. Additionally it was found that Mef2 is able to interact with certain proteins that are reported as cytoplasmic, while literature categorises Mef2 as only nuclear. Mef2 is able to create protein interactions not only with muscle structural proteins, but also ribosomal proteins and proteasomal complexes (as shown by approach b).

Lastly, I characterised the function of HDAC4 in muscles, since this was an interesting candidate found to directly interact with Mef2 and that seemed not to be part of a bigger protein complex. By conducting overexpression experiments, I was able to show that HDAC4 functions in muscles and its mis-expression can affect muscle differentiation and gene activation. Of particular interest could be that despite being reported to shuttle between nucleus and cytoplasm, HDAC4 was mainly cytoplasmic in *Drosophila* embryos throughout development.

6.2 The role of large scale purification methods in studying protein interactions

The potential functions that a protein can carry out are defined by its ability to interact physically with other molecules in the cell, including ions, metabolites, lipids, RNA, DNA and other proteins. Proteomics has become a widely-used technique to identify and analyse proteins at large scale. Traditional yeast two-hybrid studies have addressed similar questions by testing thousands of binary protein-protein interactions and constructing "interactome" maps by connecting proteins based on positive interactions in the yeast two-hybrid setup (Giot et al., 2003; Stanyon et al., 2004). In the last two decades, the use of co-affinity purification has been increasing, whereby protein extracts are generated that contain a mixture of proteins "pulled down" by transgenically expressing a tagged version of a protein of interest ("bait"). The tag is then used for purification, and proteins interacting with the protein in the desired biological context ("prey") are co-purified with the bait. Proteins contained in the mixture can be identified with high sensitivity but low throughput using for example Western blot, or with low sensitivity but high throughput using mass spectrometry (MS). In contrast to the yeast two-hybrid method, this allows screening 1-to-n interactions rather than testing one binary interaction at a time. Consequently, datasets from such studies tend to be very large, providing a wealth of information on protein-protein interactions that was unobtainable previously. Two hallmark studies that were also integrated in the present analysis produced large interaction networks for, respectively, most proteins of the *Drosophila melanogaster* proteome (Guruharsha et al., 2011), or specifically for 459 *Drosophila* transcription factors (Rhee et al., 2014). Inevitably, some quality must be sacrificed in exchange for quantity. In the case of both of these studies, the data were obtained from S2R+ cell cultures, meaning that the data on protein-protein interactions obtained are not contextualised with development and different tissue types. However, as more datasets become available, our confidence in individual associations will improve, and it is through such large-scale studies that predicting the protein interaction networks of new proteins of interest will become more reliable. Candidates of interest extracted from this data must nevertheless be validated individually before conclusions about their functions can be drawn.

The successful application of TAP experiments is very specific to the type of biological materials used and many tag variants have been optimised for a particular system. Such type of purification experiments could be applied at organism level in single cell organisms like yeast (where the method was initially developed) and bacteria and lower level multicellular model systems like *Drosophila* and *C. elegans* (Kyriakakis et al., 2008; Pepper et al., 2012; Puig et al., 2001; Rigaut et al., 1999; Veraksa et al., 2005; Viala and Bouveret, 2017; Zanin et al., 2011). In mammals, only cell culture TAP purifications were performed and different types of tags were developed to try to optimise the recovery of native complexes (Bürckstümmer et al., 2006; Ma et al., 2012). The original TAP tag consisted of protein A and calmodulin-binding peptide (CBP) moieties (Puig et al., 2001; Rigaut et al., 1999). A later version used FLAG and HA peptide tags for sequential immuno-affinity purification (Nakatani and Ogryzko, 2003). The GSTAP system used here is a combination of two protein G domains, together with a streptavidin binding peptide (SBP) (Kyriakakis et al., 2008), while the BIOTAP system has a Protein A moiety (Rigaut et al., 1999) and a biotinylation targeting signal consisting of a 75-amino acid sequence derived from a *P. shermanii* transcarboxylase (Alekseyenko et al., 2015; Guerrero et al., 2005). Another variant called the LAP tag contains GFP (or mCherry) and S peptide and was used successfully in *C. elegans* (Zanin et al., 2011). Despite the different tags available, the principles of the purification protocol remain similar: biological material is lysed to generate whole extracts or cytoplasmic or nuclear extracts, protein native complexes are extracted via two subsequent affinity purification steps and the eluted proteins are analysed most commonly by MS. Generally these experiments can generate quite a long list of candidate proteins that need to be screened via computation means to extract the best hits.

In this study we were able to extract Mef2 interacting proteins from *Drosophila* embryos using a TAP purification protocol. The idea of tagging Mef2 with TAP for purification has been successfully applied in mammalian cell culture as well, where a TAP-tagged Mef2a was used to co-purify with Mef2c and to concentrate the low expressing Mef2a in COS7 and HeLa cell in order to allow determining its phosphorylation states via Mass Spectrometry (Aridgides et al., 2002; Cox et al., 2003).

6.3 Highlighting important biological processes from complex data using bioinformatic analyses

Such large amounts of data require computational methods to extract meaning from them. A major problem in the field of proteomics studies as of yet is standardisation, both in technical aspects, and in the analysis of obtained data. The experimental methods for purifying proteins vary substantially between studies because they inevitably have to be adjusted to each specific scientific question. As illustrated in chapter 3, a variety of affinity tags have been developed just for tandem affinity purification (TAP) and there are many more variants of purification tags for different protocols depending on the biological context. Recent efforts to make proteomics more robust have included quantitative proteomics (Schubert et al., 2017) whereby amino acids labelled with stable isotopes are incorporated into the proteins analysed by MS, as well as some statistical approaches (Aggarwal and Yadav, 2016). However, there is no standard set of methods used to analyse proteomic datasets and consequently few studies analyse their data using similar techniques.

The use of proteomics in this study represents a comparably small-scale application, focussing much effort on optimising the purification of a single bait. It is commonplace at this level for studies to not perform an in-depth analysis but instead to discard all but a small number of the most abundant proteins identified by MS. The remaining proteins are then manually curated and candidates are selected for further experiments based on literature context. The data obtained from TAP extracts for Mef2 lent itself to a more in-depth analysis because the TAP procedure maintains physiological conditions and thus preserves protein complexes beyond direct physical bait-prey interactions. The subsequent data analysis used several methods that were previously described in different published studies. The largest part of the analysis followed a similar pipeline as the ParkinTAP project (Zanon et al., 2013). In this study, a similar procedure of network analysis was used to correlate proteins purified using Parkin as bait with databases of known proteins related to parkinsonism. To analyse the Mef2 datasets, the procedures had to be adapted for the *Drosophila* model system, and suitable databases had to be identified. The results showed that both the datasets derived from unstaged (overnight-collected) and staged (11-13 h) *Drosophila* embryos likely contained few genuine contaminants, since the

vast majority of proteins had known associations with Mef2 and a functional comparison showed that the vast majority of proteins in the datasets cluster around Mef2 when their annotation with gene ontology (GO) terms is taken into account. It should therefore be interesting for future research to determine the biological relationship between Mef2 and each of the identified interaction candidates.

Similar methods of analysis have been used in two TAP studies targeting other TFs, with the specific aim of distinguishing proteins that TFs interact with while bound to chromatin, and while not bound to chromatin. The first, a study of 56 human TFs in HEK293T cells (mostly Forkhead box TFs but also including Mef2a), showed that TFs interact with diverse complexes on and off chromatin (Li et al., 2015). The "off chromatin" fraction in this study should be methodologically equivalent to the cytosolic fraction used here, while the "on chromatin" fraction, the pellet, was not used in our case. The complexes found in the solubilised fractions of TFs were significantly different than the chromatin bound complexes obtained from protein complexes pulled down from the pelleted fractions. The soluble fractions did not only contain proteins released from the chromatin, rather also proteins that interact with TFs independently of DNA binding. While the "on chromatin" section enriched for transcription related proteins, the soluble fractions enriched for kinases, peptidase and transmembrane proteins. Many of the analysed TFs were classified mainly nuclear via immunohistochemistry, like Mef2, but they were detected both in the soluble fraction and the chromatin fractions. A similar behaviour was observed in the case of Mef2a and many of the proteins purified together with Mef2a from soluble fractions were involved in post-translational regulation and trafficking. This could imply that the Mef2 purification experiments performed here potentially enriched for complexes that have functions in post-translational regulation, rather than transcriptional modulation, which would explain why we did not pull down any of the TFs with which Mef2 is known to physically interact in *Drosophila*, like Vestigial, Scalloped or Twist. Another explanation for this could also be that TAP is a relatively conservative pulldown method that tends to identify only a fraction of the full interactome (Liu et al., 2004). However, there were transcription associated complexes found both in on chromatin and off chromatin (soluble) purified fractions, explaining why it was possible to identify transcriptional modulators among the candidates of Mef2 interacting proteins extracted from solubilised fractions.

The second study focussed on MyoD as the bait, a basic helix-loop-helix (bHLH) TF known to induce muscle differentiation together with Mef2 in vertebrates (Boyarchuk et al., 2016). The interactions of this protein were studied in HeLa cells via TAP purification of a Flag-HA tagged MyoD, in this case specifically in nuclear soluble and chromatin bound fractions. The MyoD partners found included TFs and co-repressors, the SWI/SNF complex and other chromatin remodellers, RNA processing factors, kinases, histones, histone chaperones, protein trafficking factors and DNA replication factors. Any ribosomal proteins, translation-initiation factors, DNA repair factors and the tubulin isoforms were directly regarded as contaminants and excluded. Several of these protein groups were present both on and off chromatin, though every group had unique members only found in one fraction. In follow-up experiments, MyoD was found to interact in muscle cells with proteins that functionally act as TFs, transcription co-repressors and SWI/SNF complexes when considering the chromatin bound partners. The interacting partners retrieved from the soluble fraction included TFs, chromatin remodelers, transcription co-repressors, SWI/SNF complexes and RNA processing factors. Our study identifying Mef2 partners in *Drosophila* muscle recovered many proteins of the same groups: transcription factors, SWI/SNF subunits, transcriptional regulators, transcription co-repressors, but also ribosomal proteins, translation-initiation factors, actin and tubulin isoforms and several heat shock proteins. Many of the protein belonging to the last 5 categories would be considered contaminants. However, Mef2 has been shown to interact with proteins like heat shock proteins in different contexts (Wang et al., 2019; Yang et al., 2009; Zhang et al., 2014). Additionally, Mef2c directly binds to the MyoD-E12 heterodimer and activates transcription necessary for myogenesis to occur (Molkentin and Olson, 1996). However, no Mef2 isoform was pulled down with MyoD in the TAP purification, despite the fact that the Brg1 subunit known to mediate the interaction between MRFs and Mef2s was recovered (Ohkawa et al., 2006).

The candidates extracted via TAP purification of Mef2 in *Drosophila* embryos were derived from a non-chromatin bound fraction. Given published results, this could have biased the candidates towards post-translational regulation, rather than transcriptional regulation. However, this does not mean that proteins involved in modulation of transcription cannot be found in the soluble fraction. Most of the

MyoD interaction partners belonged to similar types of complexes both when looking at the chromatin bound sample or the soluble fraction. Given the technical challenges, involving additional optimisations, and potential risk for other contaminants in the pellet, the soluble fraction was thus the best way to get an overview of the types of complexes a certain transcription factors comes in contact with.

6.4 Biological validation of candidates identified from purification data

We selected one candidate of the dataset that appeared interesting based on its functional association with muscle-related proteins and its known association with muscle in other organisms. HDAC4 was indeed able to repress muscle development when a mutant variant was overexpressed that was lacking the nuclear export sequence and would thus localise to the nucleus constitutively. However, in the reverse experiment where we deleted the *hdac4* gene, we found no muscle-associated phenotype in HDAC4 null embryos. This could be simply due to redundancy and therefore compensation by another regulatory mechanism, or due to rescue by maternally deposited HDAC4. If neither of these were the case, this could suggest that HDAC4 represses other cell types, potentially non-muscle cells, from activating muscle-specific transcription. The broad expression pattern of *hdac4* mRNA and HDAC4 protein would allow for such a function. The fact that HDAC4 is found to bind Mef2 in late myogenesis, suggests that its role goes beyond the known function in vertebrates of impeding muscle differentiation by interfering with the fusion process (Miska et al., 2001). The fly HDAC4 seems to closely mimic the activity of the mammalian HDAC4, therefore a role for HDAC4 in the maintenance and/or modulation of fibre type-specific gene transcription could also be present in *Drosophila* (Cohen et al., 2015).

These experiments can serve as a proof-of-concept that larger-scale studies and careful computational analysis can indicate promising avenues of further research. We attempted to use *Drosophila* embryos as an *in vivo* model for acquiring candidates specific to a developmental process, which was in principle successful. However, this work also required significant investments.

The TAP purification screen has proven to be able to create a full picture of the kind of proteins Mef2 interacts with. The array of functions discovered has shown that Mef2 is involved in many different functions but that are very much interconnected. The reliability of these functions is supported both by the facts that most of the identified candidates were well characterised and that many of them were purified together with Mef2 under very different conditions.

The coherence of the data proven through *in silico* assessment leads us to believe that using TAP purification to understand the way a protein performs its activity at a proteomics level is a productive approach. TAP screens in themselves are quite time intensive and require a significant input of biological material to obtain valuable results. The ability to validate such results through computational approaches is a quicker and more systematic way than to perform genetic screens. The methodology described in this study could serve as a pipeline to confirm the most reliable candidates that can be further used for *in vivo* validation. One important aspect of the analysed samples was to highlight that just disregarding the unspecific binders pulled down in the control can bias the results. Mef2 shows to play an important role in metabolism and proteolysis, some of the proteins being classified as contaminants in other studies.

6.5 Conclusions and Perspective

In this work, we focussed on the characterisation of Mef2 partners in *Drosophila* embryos, partially because all known regulation mechanisms related to Mef2 activity primarily relate to activation of target genes and only as a secondary consequence to how the Mef2 protein dynamics are kept in check during development. Prior to the advent of proteomics, performing large scale protein purification studies was challenging, because generating the required biological input material was difficult and the analytical methods available did not allow for high throughput results. In many prior studies, interactions between two proteins were assessed in small scale studies or in yeast two hybrid test. While in many tissues the key players were discovered by genetic studies, a systematic interactome approach is required to understand which proteins are associated with different molecular functions in order to achieve cell differentiation. My work thus provides a valuable reference of Mef2

interacting proteins that should enrich our understanding of how Mef2 acts in muscle, and allows future work on muscle development to proceed with clearer directions.

The main part of my thesis comprises a large scale screen of Mef2 interacting proteins, executed using a tandem affinity purification approach that has proven in many studies to provide substantial benefits over other types of extraction methods. By knocking-in the GS-TAP tag, Mef2 was made compatible for extraction experiments at physiological conditions. This approach, combined with computational analytical methods, revealed a number of protein complexes that modulate Mef2 activity in muscle. It further implicated Mef2 in regulatory mechanisms that involve direct interactions with many of its target gene products that are involved in muscle structure. Of particular interest could be Actin57B and WupA, neither of which have been characterised previously as shuttling to the nucleus in muscle cells. Of interest would be to study in more detail if these genes translocate to the nucleus to perform “moonlight functions”.

Lastly, I characterised the function of HDAC4 in *Drosophila* embryos, since this was one of two candidates whose presence in fully differentiated myotubes could imply a role in transcriptional regulation through direct binding to Mef2 and its role in late embryogenesis has not been studied. By conducting mis-expression experiments, I was able to show that HDAC4 functions in muscle cells and can affect muscle specific transcription. Mef2 protein levels are not impacted when HDAC4 is overexpressed, but activation of its target genes and muscle differentiation are impacted. Therefore, direct binding of HDAC4 to repress Mef2 activity in muscle is plausible. It will be interesting in the future to identify if HDAC4 colocalises with Mef2 in nuclear bodies of muscle cells as observed in neurons and if there is a direct functional link between HDAC4, the Brahma complex and Mef2 in regulation of muscle maintenance in differentiated myotubes.

7. References

- Abmayr, S.M., and Pavlath, G.K. (2012). Myoblast fusion: lessons from flies and mice. *Development* *139*, 641–656.
- Adryan, B., and Teichmann, S.A. (2010). The developmental expression dynamics of *Drosophila melanogaster* transcription factors. *Genome Biol.* *11*, R40.
- Aggarwal, S., and Yadav, A.K. (2016). False Discovery Rate Estimation in Proteomics. In *Statistical Analysis in Proteomics*, K. Jung, ed. (New York: Humana Press), pp. 119–128.
- Aghajanian, P., Takashima, S., Paul, M., Younossi-Hartenstein, A., and Hartenstein, V. (2016). Metamorphosis of the *Drosophila* visceral musculature and its role in intestinal morphogenesis and stem cell formation. *Dev. Biol.* *420*, 43–59.
- Alekseyenko, A.A., McElroy, K.A., Kang, H., Zee, B.M., Kharchenko, P. V, and Kuroda, M.I. (2015). BioTAP-XL: Cross-linking/Tandem Affinity Purification to Study DNA Targets, RNA, and Protein Components of Chromatin-Associated Complexes. *Curr. Protoc. Mol. Biol.* *109*, 21.30.1-21.30.32.
- Andres, V., Cervera, M., and Mahdavi, V. (1995). Determination of the consensus binding site for MEF2 expressed in muscle and brain reveals tissue-specific sequence constraints. *J. Biol. Chem.* *270*, 23246–23249.
- Aridgides, L.J., Stacey, M., Brihn, L., Scott, D., and Osgood, C. (2002). Fluorescence In Situ Hybridization on Sperm Using Alkaline Denaturation. *Biotechniques* *33*, 266–267.
- Arnold, M.A., Kim, Y., Czubryt, M.P., Phan, D., McAnally, J., Qi, X., Shelton, J.M., Richardson, J.A., Bassel-Duby, R., and Olson, E.N. (2007). MEF2C transcription factor controls chondrocyte hypertrophy and bone development. *Dev. Cell* *12*, 377–389.
- Arredondo, J.J., Marco Ferreres, R., Maroto, M., Cripps, R.M., Marco, R., Bernstein, S.I., and Cervera, M. (2001). Control of *Drosophila* paramyosin/miniparamyosin gene expression. Differential regulatory mechanisms for muscle-specific transcription. *J. Biol. Chem.* *276*, 8278–8287.

Arredondo, J.J., Vivar, J., Laine-Menéndez, S., Martínez-Morentin, L., and Cervera, M. (2017). CF2 transcription factor is involved in the regulation of Mef2 RNA levels, nuclei number and muscle fiber size. *PLoS One* *12*, e0179194.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* *40*, W597-603.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* *25*, 25–29.

Azpiazu, N., and Frasch, M. (1993). Tinman and bagpipe: Two homeo box genes that determine cell fates in the dorsal mesoderm of *Drosophila*. *Genes Dev.* *7*, 1325–1340.

Bader, G.D., and Hogue, C.W. V (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* *4*, 2.

Bagni, C., Bray, S., Gogos, J.A., Kafatos, F.C., and Hsu, T. (2002). The *Drosophila* zinc finger transcription factor CF2 is a myogenic marker downstream of MEF2 during muscle development. *Mech. Dev.* *117*, 265–268.

Barrell, D., Dimmer, E., Huntley, R.P., Binns, D., O'Donovan, C., and Apweiler, R. (2009). The GOA database in 2009--an integrated Gene Ontology Annotation resource. *Nucleic Acids Res.* *37*, D396-403.

Bate, M. (1990). The embryonic development of larval muscles in *Drosophila*. *Development* *110*, 791–804.

Bate, M., and Martinez-Arias, A. (1993). *The Development of Drosophila melanogaster* (New York: Cold Spring Harbor Laboratory Press).

Bate, M., and Rushton, E. (1993). Myogenesis and muscle patterning in *Drosophila*. *C. R. Acad. Sci. III.* *316*, 1047–1061.

Bate, M., Rushton, E., and Currie, D. a (1991). Cells with persistent twist expression

are the embryonic precursors of adult muscles in *Drosophila*. *Development* 113, 79–89.

Bate, M., Rushton, E., and Frasch, M. (1993). A dual requirement for neurogenic genes in *Drosophila* myogenesis. *Development* 119, 149–161.

Baylies, M.K., and Bate, M. (1996). twist: a myogenic switch in *Drosophila*. *Science* 272, 1481–1484.

Baylies, M.K., Bate, M., and Gomez, M.R. (1998). Myogenesis: A view from *Drosophila*. *Cell* 93, 921–927.

Bentzinger, C.F., Wang, Y.X., and Rudnicki, M.A. (2012). Building muscle: molecular regulation of myogenesis. *Cold Spring Harb. Perspect. Biol.* 4.

Bernard, F., Lalouette, A., Gullaud, M., Jeantet, A., Cossard, R., Zider, A., Ferveur, J., and Silber, J. (2003). Control of apterous by vestigial drives indirect flight muscle development in *Drosophila*. *Dev. Biol.* 260, 391–403.

Bernard, F., Dutriaux, A., Silber, J., and Lalouette, A. (2006). Notch pathway repression by vestigial is required to promote indirect flight muscle differentiation in *Drosophila melanogaster*. *Dev. Biol.* 295, 164–177.

Bernard, F., Kasherov, P., Grenetier, S., Dutriaux, A., Zider, A., Silber, J., and Lalouette, A. (2009). Integration of differentiation signals during indirect flight muscle formation by a novel enhancer of *Drosophila* vestigial gene. *Dev. Biol.* 332, 258–272.

Bertos, N.R., Gilquin, B., Chan, G.K.T., Yen, T.J., Khochbin, S., and Yang, X.-J. (2004). Role of the tetradecapeptide repeat domain of human histone deacetylase 6 in cytoplasmic retention. *J. Biol. Chem.* 279, 48246–48254.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pagès, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25, 1091–1093.

Black, B.L., and Cripps, R.M. (2010). Myocyte Enhancer Factor 2 Transcription

Factors in Heart Development and Disease. In *Heart Development and Regeneration*, N. Rosenthal, and R.P. Harvey, eds. (Oxford: Academic Press), pp. 673–699.

Black, B.L., and Olson, E.N. (1998). Transcriptional Control of Muscle Development By Myocyte Enhancer Factor-2 (Mef2) Proteins. *Annu. Rev. Cell Dev. Biol.* *14*, 167–196.

Blanchard, F.J., Collins, B., Cyran, S.A., Hancock, D.H., Taylor, M. V, and Blau, J. (2010). The transcription factor Mef2 is required for normal circadian behavior in *Drosophila*. *J. Neurosci.* *30*, 5855–5865.

Blander, G., and Guarente, L. (2004). The Sir2 Family of Protein Deacetylases. *Annu. Rev. Biochem.* *73*, 417–435.

Bodmer, R., and Frasch, M. (2010). Development and Aging of the *Drosophila* Heart. In *Heart Development and Regeneration*, N. Rosenthal, and R.P. Harvey, eds. (Oxford: Academic Press), pp. 47–86.

Bonn, S., and Furlong, E.E.M. (2008). cis-Regulatory networks during development: a view of *Drosophila*. *Curr. Opin. Genet. Dev.* *18*, 513–520.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E.M. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* *44*, 148–156.

Borghgi, S., Molinari, S., Razzini, G., Parise, F., Battini, R., and Ferrari, S. (2001). The nuclear localization domain of the MEF2 family of transcription factors shows member-specific features and mediates the nuclear import of histone deacetylase 4. *J. Cell Sci.* *114*, 4477–4483.

Borkowski, O.M., Brown, N.H., and Bate, M. (1995). Anterior-posterior subdivision and the diversification of the mesoderm in *Drosophila*. *Development* *121*, 4183–4193.

Bothe, I., and Baylies, M.K. (2016). *Drosophila* myogenesis. *Curr. Biol.* *26*, R786–91.

- Boukhatmi, H., and Bray, S. (2018). A population of adult satellite-like cells in *Drosophila* is maintained through a switch in RNA-isoforms. *Elife* 7, e35954.
- Boulay, J.L., Dennefeld, C., and Alberga, a. (1987). The *Drosophila* developmental gene *snail* encodes a protein with nucleic acid binding fingers. *Nature* 330, 395–398.
- Bour, B.A., O'Brien, M.A., Lockwood, W.L., Goldstein, E.S., Bodmer, R., Taghert, P.H., Abmayr, S.M., and Nguyen, H.T. (1995). *Drosophila* MEF2, a transcription factor that is essential for myogenesis. *Genes Dev.* 9, 730–741.
- Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P.-O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S., et al. (2004). A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. *Nat. Cell Biol.* 6, 97–105.
- Boyarchuk, E., Robin, P., Fritsch, L., Joliot, V., and Ait-Si-Ali, S. (2016). Identification of MyoD Interactome Using Tandem Affinity Purification Coupled to Mass Spectrometry. *J. Vis. Exp.* e53924.
- Breitbart, R.E., Liang, C.S., Smoot, L.B., Laheru, D.A., Mahdavi, V., and Nadal-Ginard, B. (1993). A fourth human MEF2 transcription factor, hMEF2D, is an early marker of the myogenic lineage. *Development* 118, 1095–1106.
- Britten, R.J., and Davidson, E.H. (1969). Gene regulation for higher cells: a theory. *Science* 165, 349–357.
- Bryantsev, A.L., and Cripps, R.M. (2009). Cardiac gene regulatory networks in *Drosophila*. *Biochim. Biophys. Acta* 1789, 343–353.
- Bürckstümmer, T., Bennett, K.L., Preradovic, A., Schütze, G., Hantschel, O., Superti-Furga, G., and Bauch, A. (2006). An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat. Methods* 3, 1013–1019.
- Busser, B.W., Huang, D., Rogacki, K.R., Lane, E.A., Shokri, L., Ni, T., Gamble, C.E., Gisselbrecht, S.S., Zhu, J., Bulyk, M.L., et al. (2012). Integrative analysis of the zinc finger transcription factor *Lame duck* in the *Drosophila* myogenic gene

regulatory network. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20768–20773.

Butland, G., Peregrín-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., et al. (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* 433, 531–537.

Cabantous, S., Nguyen, H.B., Pedelacq, J.-D., Koraïchi, F., Chaudhary, A., Ganguly, K., Lockard, M.A., Favre, G., Terwilliger, T.C., and Waldo, G.S. (2013). A New Protein-Protein Interaction Sensor Based on Tripartite Split-GFP Association. *Sci. Rep.* 3, 2854.

Cai, H.N., Arnosti, D.N., and Levine, M. (1996). Long-range repression in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. U. S. A.* 93, 9309–9314.

Campos-Ortega, J.A., and Hartenstein, V. (1997). *The Embryonic Development of Drosophila melanogaster* (Berlin: Springer).

Carmena, A., Bate, M., and Jimenez, F. (1995). lethal of scute, a proneural gene, participates in the specification of muscle progenitors during *Drosophila* embryogenesis. *Genes Dev.* 9, 2373–2383.

Carmena, A., Murugasu-Oei, B., Menon, D., Jiménez, F., and Chia, W. (1998). inscuteable and numb mediate asymmetric muscle progenitor cell divisions during *Drosophila* myogenesis. *Genes Dev.* 12, 304–315.

Casey, J., and Davidson, N. (1977). Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. *Nucleic Acids Res.* 4, 1539–1552.

Castanon, I., and Baylies, M.K. (2002). A Twist in fate: Evolutionary comparison of Twist structure and function. *Gene* 287, 11–22.

Chambers, A.E., Kotecha, S., Towers, N., and Mohun, T.J. (1992). Muscle-specific expression of SRF-related genes in the early embryo of *Xenopus laevis*. *EMBO J.* 11, 4981–4991.

Chan, J.K.L., Sun, L., Yang, X.-J., Zhu, G., and Wu, Z. (2003). Functional

characterization of an amino-terminal region of HDAC4 that possesses MEF2 binding and transcriptional repressive activity. *J. Biol. Chem.* 278, 23515–23521.

Chang, S., Bezprozvannaya, S., Li, S., and Olson, E.N. (2005). An expression screen reveals modulators of class II histone deacetylase phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8120–8125.

Chaturvedi, D., Reichert, H., Gunage, R.D., and VijayRaghavan, K. (2017). Identification and functional characterization of muscle satellite cells in *Drosophila*. *Elife* 6, e30107.

Chen, E.H., and Olson, E.N. (2004). Towards a molecular pathway for myoblast fusion in *Drosophila*. *Trends Cell Biol.* 14, 452–460.

Chen, J.-F., Mandel, E.M., Thomson, J.M., Wu, Q., Callis, T.E., Hammond, S.M., Conlon, F.L., and Wang, D.-Z. (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat. Genet.* 38, 228–233.

Chen, Z., Liang, S., Zhao, Y., and Han, Z. (2012). miR-92b regulates Mef2 levels through a negative-feedback circuit during *Drosophila* muscle development. *Development* 139, 3543–3552.

Cho, Y., Griswold, A., Campbell, C., and Min, K.T. (2005). Individual histone deacetylases in *Drosophila* modulate transcription of distinct genes. *Genomics* 86, 606–617.

Ciglar, L., Girardot, C., Wilczyński, B., Braun, M., and Furlong, E.E.M. (2014). Coordinated repression and activation of two transcriptional programs stabilizes cell fate during myogenesis. *Development* 141, 2633–2643.

Clark, R.I., Tan, S.W.S., Péan, C.B., Roostalu, U., Vivancos, V., Bronda, K., Pilátová, M., Fu, J., Walker, D.W., Berdeaux, R., et al. (2013). MEF2 is an in vivo immune-metabolic switch. *Cell* 155, 435–447.

Clocchiatti, A., Florean, C., and Brancolini, C. (2011). Class IIa HDACs: from important roles in differentiation to possible implications in tumorigenesis. *J. Cell. Mol. Med.* 15, 1833–1846.

- Clocchiatti, A., Di Giorgio, E., Ingrao, S., Meyer-Almes, F.-J., Tripodo, C., and Brancolini, C. (2013). Class IIa HDACs repressive activities on MEF2-dependent transcription are associated with poor prognosis of ER⁺ breast tumors. *Int. J. Radiat. Oncol.* 27, 942–954.
- Cohen, T.J., Barrientos, T., Hartman, Z.C., Garvey, S.M., Cox, G.A., and Yao, T. (2009). The deacetylase HDAC4 controls myocyte enhancing factor-2-dependent structural gene expression in response to neural activity. *FASEB J.* 23, 99–106.
- Cohen, T.J., Choi, M.-C., Kapur, M., Lira, V.A., Yan, Z., and Yao, T.-P. (2015). HDAC4 regulates muscle fiber type-specific gene expression programs. *Mol. Cells* 38, 343–348.
- Cox, D.M., Du, M., Marback, M., Yang, E.C.C., Chan, J., Siu, K.W.M., and McDermott, J.C. (2003). Phosphorylation Motifs Regulating the Stability and Function of Myocyte Enhancer Factor 2A. *J. Biol. Chem.* 278, 15297–15303.
- Cripps, R.M., and Olson, E.N. (2002). Control of Cardiac Development by an Evolutionarily Conserved Transcriptional Network. *Dev. Biol.* 246, 14–28.
- Cripps, R.M., Black, B.L., Zhao, B., Lien, C.L., Schulz, R.A., and Olson, E.N. (1998). The myogenic regulatory gene Mef2 is a direct target for transcriptional activation by Twist during *Drosophila* myogenesis. *Genes Dev.* 12, 422–434.
- Cripps, R.M., Zhao, B., and Olson, E.N. (1999). Transcription of the myogenic regulatory gene Mef2 in cardiac, somatic, and visceral muscle cell lineages is regulated by a Tinman-dependent core enhancer. *Dev. Biol.* 215, 420–430.
- Cripps, R.M., Lovato, T.L., and Olson, E.N. (2004). Positive autoregulation of the Myocyte enhancer factor-2 myogenic control gene during somatic muscle development in *Drosophila*. *Dev. Biol.* 267, 536–547.
- Cunha, P.M.F., Sandmann, T., Gustafson, E.H., Ciglar, L., Eichenlaub, M.P., and Furlong, E.E.M. (2010). Combinatorial binding leads to diverse regulatory responses: Lmd is a tissue-specific modulator of Mef2 activity. *PLoS Genet.* 6, e1001014.

Damm, C., Wolk, A., Buttgerit, D., Löher, K., Wagner, E., Lilly, B., Olson, E.N., Hasenpusch-Theil, K., and Renkawitz-Pohl, R. (1998). Independent regulatory elements in the upstream region of the *Drosophila* beta 3 tubulin gene (beta Tub60D) guide expression in the dorsal vessel and the somatic muscles. *Dev. Biol.* *199*, 138–149.

Davidson, E.H., and Britten, R.J. (1974). Proceedings: Molecular aspects of gene regulation in animal cells. *Cancer Res.* *34*, 2034–2043.

Davis, D., Yaveroğlu, Ö.N., Malod-Dognin, N., Stojmirovic, A., and Pržulj, N. (2015). Topology-function conservation in protein-protein interaction networks. *Bioinformatics* *31*, 1632–1639.

Deng, H., Hughes, S.C., Bell, J.B., and Simmonds, A.J. (2009). Alternative requirements for Vestigial, Scalloped, and Dmef2 during muscle differentiation in *Drosophila melanogaster*. *Mol. Biol. Cell* *20*, 256–269.

Doberstein, S.K., Fetter, R.D., Mehta, A.Y., and Goodman, C.S. (1997). Genetic analysis of myoblast fusion: Blown fuse is required for progression beyond the prefusion complex. *J. Cell Biol.* *136*, 1249–1261.

Dobi, K.C., Schulman, V.K., and Baylies, M.K. (2015). Specification of the somatic musculature in *Drosophila*. *Dev. Biol.* *4*, 357–375.

Dong, C., Yang, X.-Z., Zhang, C.-Y., Liu, Y.-Y., Zhou, R.-B., Cheng, Q.-D., Yan, E.-K., and Yin, D.-C. (2017). Myocyte enhancer factor 2C and its directly-interacting proteins: A review. *Prog. Biophys. Mol. Biol.* *126*, 22–30.

Downes, M., Ordentlich, P., Kao, H.Y., Alvarez, J.G., and Evans, R.M. (2000). Identification of a nuclear domain with deacetylase activity. *Proc. Natl. Acad. Sci. U. S. A.* *97*, 10330–10335.

Dressel, U., Bailey, P.J., Wang, S.C.M., Downes, M., Evans, R.M., and Muscat, G.E.O. (2001). A Dynamic Role for HDAC7 in MEF2-mediated Muscle Differentiation. *J. Biol. Chem.* *276*, 17007–17013.

Dubnicoff, T., Valentine, S.A., Chen, G., Shi, T., Lengyel, J.A., Paroush, Z., and

Courey, A.J. (1997). Conversion of dorsal from an activator to a repressor by the global corepressor Groucho. *Genes Dev.* *11*, 2952–2957.

Edmondson, D.G., Cheng, T.C., Cserjesi, P., Chakraborty, T., and Olson, E.N. (1992). Analysis of the myogenin promoter reveals an indirect pathway for positive autoregulation mediated by the muscle-specific enhancer factor MEF-2. *Mol. Cell. Biol.* *12*, 3665–3677.

Edmondson, D.G., Lyons, G.E., Martin, J.F., and Olson, E.N. (1994). Mef2 Gene Expression Marks the Cardiac and Skeletal Muscle Lineages During Mouse Embryogenesis. *Development* *120*.

Elgar, S.J., Han, J., and Taylor, M. V (2008). mef2 activity levels differentially affect gene expression during *Drosophila* muscle development. *Proc. Natl. Acad. Sci.* *105*, 918–923.

Estrella, N.L., Desjardins, C.A., Nocco, S.E., Clark, A.L., Maksimenko, Y., and Naya, F.J. (2015). MEF2 transcription factors regulate distinct gene programs in mammalian skeletal muscle differentiation. *J. Biol. Chem.* *290*, 1256–1268.

Fischle, W., Dequiedt, F., Hendzel, M.J., Guenther, M.G., Lazar, M.A., Voelter, W., and Verdin, E. (2002). Enzymatic activity associated with class II HDACs is dependent on a multiprotein complex containing HDAC3 and SMRT/N-CoR. *Mol. Cell* *9*, 45–57.

Fitzsimons, H.L., Schwartz, S., Given, F.M., and Scott, M.J. (2013). The histone deacetylase HDAC4 regulates long-term memory in *Drosophila*. *PLoS One* *8*, e83903.

Foglietti, C., Filocamo, G., Cundari, E., De Rinaldis, E., Lahm, A., Cortese, R., and Steinkühler, C. (2006). Dissecting the biological functions of *Drosophila* histone deacetylases by RNA interference and transcriptional profiling. *J. Biol. Chem.* *281*, 17968–17976.

Fomproix, N., and Percipalle, P. (2004). An actin–myosin complex on actively transcribing genes. *Exp. Cell Res.* *294*, 140–148.

- Fortini, P., Ferretti, C., Iorio, E., Cagnin, M., Garribba, L., Pietraforte, D., Falchi, M., Pascucci, B., Baccarini, S., Morani, F., et al. (2016). The fine tuning of metabolism, autophagy and differentiation during in vitro myogenesis. *Cell Death Dis.* 7, e2168–e2168.
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., Von Mering, C., et al. (2013). STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41.
- Frasch, M. (1995). Induction of visceral and cardiac mesoderm. *Nature* 374, 464–467.
- Furlong, E.E., Andersen, E.C., Null, B., White, K.P., and Scott, M.P. (2001). Patterns of gene expression during *Drosophila* mesoderm development. *Science* 293, 1629–1633.
- Gajewski, K.M., and Schulz, R.A. (2010). CF2 represses Actin 88F gene expression and maintains filament balance during indirect flight muscle development in *Drosophila*. *PLoS One* 5, e10713.
- García-Bellido, A. (1975). Genetic control of wing disc development in *Drosophila*. In *Cell Patterning*, R. Porter, and J. Rivers, eds. (Basel: Ciba Foundation), pp. 161–182.
- Gavin, A.-C., Bösch, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147.
- Di Giorgio, E., Gagliostro, E., and Brancolini, C. (2015). Selective class IIa HDAC inhibitors: myth or reality. *Cell. Mol. Life Sci.* 72, 73–86.
- Di Giorgio, E., Hancock, W.W., and Brancolini, C. (2018). MEF2 and the tumorigenic process, hic sunt leones. *Biochim. Biophys. Acta - Rev. Cancer* 1870, 261–273.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E., et al. (2003). A protein interaction map of *Drosophila melanogaster*. *Science* 302, 1727–1736.

Gossett, L.A., Kelvin, D.J., Sternberg, E.A., and Olson, E.N. (1989). A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. *Mol. Cell. Biol.* 9, 5022–5033.

Gould, J. (2012). Morpheus - Versatile matrix visualization and analysis software.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., et al. (2011). The developmental transcriptome of *Drosophila melanogaster*. *Nature* 471, 473–479.

Grummt, I. (2006). Actin and myosin as transcription factors. *Curr. Opin. Genet. Dev.* 16, 191–196.

Guerrero, C., Tagwerker, C., Kaiser, P., and Huang, L. (2005). An integrated mass spectrometry-based proteomic approach: quantitative analysis of tandem affinity-purified in vivo cross-linked protein complexes (QTAX) to decipher the 26 S proteasome-interacting network. *Mol. Cell. Proteomics* 5, 366–378.

Gunage, R.D., Dhanyasi, N., Reichert, H., and VijayRaghavan, K. (2017). *Drosophila* adult muscle development and regeneration. *Semin. Cell Dev. Biol.* 72, 56–66.

Gunthorpe, D., Beatty, K.E., and Taylor, M. V. (1999). Different Levels, but Not Different Isoforms, of the *Drosophila* Transcription Factor DMEF2 Affect Distinct Aspects of Muscle Differentiation. *Dev. Biol.* 215, 130–145.

Guruharsha, K.G., Rual, J.-F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., Beekman, C., Wong, C., Rhee, D.Y., Cenaj, O., et al. (2011). A protein complex network of *Drosophila melanogaster*. *Cell* 147, 690–703.

Ha, C.H., Kim, J.Y., Zhao, J., Wang, W., Jhun, B.S., Wong, C., and Jin, Z.G. (2010). PKA phosphorylates histone deacetylase 5 and prevents its nuclear export, leading to the inhibition of gene transcription and cardiomyocyte hypertrophy. *Proc. Natl.*

Acad. Sci. U. S. A. *107*, 15467–15472.

Haberland, M., Arnold, M.A., McAnally, J., Phan, D., Kim, Y., and Olson, E.N. (2007). Regulation of HDAC9 gene expression by MEF2 establishes a negative-feedback loop in the transcriptional circuitry of muscle differentiation. *Mol. Cell Biol.* *27*, 518–525.

Haggarty, S.J., Koeller, K.M., Wong, J.C., Grozinger, C.M., and Schreiber, S.L. (2003). Domain-selective small-molecule inhibitor of histone deacetylase 6 (HDAC6)-mediated tubulin deacetylation. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 4389–4394.

Halfon, M.S., Carmena, A., Gisselbrecht, S., Sackerson, C.M., Jiménez, F., Baylies, M.K., and Michelson, A.M. (2000). Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell* *103*, 63–74.

Hanukoglu, I. (1990). Elimination of non-specific binding in Western blots from non-reducing gels. *J. Biochem. Biophys. Methods* *21*, 65–68.

Hartenstein, V. (1993). *Atlas of Drosophila Development* (New York: Cold Spring Harbor Laboratory Press).

Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999). From molecular to modular cell biology. *Nature* *402*, C47-52.

He, L., and Hannon, G.J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* *5*, 522–531.

Hobson, G.M., Krahe, R., Garcia, E., Siciliano, M.J., and Funanage, V.L. (1995). Regional chromosomal assignments for four members of the MADS domain transcription enhancer factor 2 (MEF2) gene family to human chromosomes 15q26, 19p12, 5q14, and 1q12-q23. *Genomics* *29*, 704–711.

Ip, Y.T., Park, R.E., Kosman, D., Yazdanbakhsh, K., and Levine, M. (1992). Dorsal-twist interactions establish snail expression in the presumptive mesoderm of the *Drosophila* embryo. *Genes Dev.* *6*, 1518–1530.

- Jayathilaka, N., Han, A., Gaffney, K.J., Dey, R., Jarusiewicz, J.A., Noridomi, K., Philips, M.A., Lei, X., He, J., Ye, J., et al. (2012). Inhibition of the function of class IIa HDACs by blocking their interaction with MEF2. *Nucleic Acids Res.* *40*, 5378–5388.
- Jiang, J., Kosman, D., Ip, Y.T., and Levine, M. (1991). The dorsal morphogen gradient regulates the mesoderm determinant twist in early *Drosophila* embryos. *Genes Dev.* *5*, 1881–1891.
- Junion, G., Jagla, T., Duplant, S., Tapin, R., Da Ponte, J.-P., and Jagla, K. (2005). Mapping Dmef2-binding regulatory modules by using a ChIP-enriched in silico targets approach. *Proc. Natl. Acad. Sci.* *102*, 18479–18484.
- Kaiser, P., Meierhofer, D., Wang, X., and Huang, L. (2008). Tandem affinity purification combined with mass spectrometry to identify components of protein complexes. *Methods Mol. Biol.* *439*, 309–326.
- Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., and Zinzen, R.P. (2017). The *Drosophila* embryo at single-cell transcriptome resolution. *Science* *358*, 194–199.
- Kaushal, S., Schneider, J.W., Nadal-Ginard, B., and Mahdavi, V. (1994). Activation of the myogenic lineage by MEF2A, a factor that induces and cooperates with MyoD. *Science* *266*, 1236–1240.
- Kejiou, N.S., Ilan, L., Aigner, S., Luo, E., Rabano, I., Rajakulendran, N., Najafabadi, H.S., Angers, S., Yeo, G.W., and Palazzo, A.F. (2019). Pyruvate Kinase M Links Glucose Availability to Protein Synthesis. *BioRxiv* 715086.
- Kelly, K.K., Meadows, S.M., and Cripps, R.M. (2001). *Drosophila* MEF2 is a direct regulator of Actin57B transcription in cardiac, skeletal, and visceral muscle lineages. *Mech. Dev.* *110*, 39–50.
- Khochbin, S., Verdel, A., Lemercier, C., and Seigneurin-Berny, D. (2001). Functional significance of histone deacetylase diversity. *Curr. Opin. Genet. Dev.* *11*, 162–166.

Klapper, R., Heuser, S., Strasser, T., and Janning, W. (2001). A new approach reveals syncytia within the visceral musculature of *Drosophila melanogaster*. *Development* 128, 2517–2524.

Klymenko, T., Papp, B., Fischle, W., Köcher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Müller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev.* 20, 1110–1122.

Knowles-Barley, S., Longair, M., and Armstrong, J.D. (2010). BrainTrap: a database of 3D protein expression patterns in the *Drosophila* brain. *Database* 2010.

Köcher, T., and Superti-Furga, G. (2007). Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat. Methods* 4, 807–815.

Koe, C.T., Li, S., Rossi, F., Wong, J.J.L., Wang, Y., Zhang, Z., Chen, K., Aw, S.S., Richardson, H.E., Robson, P., et al. (2014). The Brm-HDAC3-Erm repressor complex suppresses dedifferentiation in *Drosophila* type II neuroblast lineages. *Elife* 3, e01906.

Kosman, D., Ip, Y.T., Levine, M., and Arora, K. (1991). Establishment of the mesoderm-neuroectoderm boundary in the *Drosophila* embryo. *Science* 254, 118–122.

Kremser, T., Gajewski, K., Schulz, R. a, and Renkawitz-Pohl, R. (1999). Tinman regulates the transcription of the beta3 tubulin gene (betaTub60D) in the dorsal vessel of *Drosophila*. *Dev. Biol.* 216, 327–339.

Kühner, S., van Noort, V., Betts, M.J., Leo-Macias, A., Batische, C., Rode, M., Yamada, T., Maier, T., Bader, S., Beltran-Alvarez, P., et al. (2009). Proteome organization in a genome-reduced bacterium. *Science* 326, 1235–1240.

Kyriakakis, P., Tipping, M., Abed, L., and Veraksa, A. (2008). Tandem affinity purification in *Drosophila*: The advantages of the GS-TAP system. *Fly (Austin)*. 2, 229–235.

Lahm, A., Paolini, C., Pallaoro, M., Nardi, M.C., Jones, P., Neddermann, P.,

- Sambucini, S., Bottomley, M.J., Lo Surdo, P., Carfi, A., et al. (2007). Unraveling the hidden catalytic activity of vertebrate class IIa histone deacetylases. *Proc. Natl. Acad. Sci.* *104*, 17335–17340.
- de Lanerolle, P., and Serebryanny, L. (2011). Nuclear actin and myosins: Life without filaments. *Nat. Cell Biol.* *13*, 1282–1288.
- Laurichesse, Q., and Soler, C. (2020). Muscle development: a view from adult myogenesis in *Drosophila*. *Semin. Cell Dev. Biol.*
- Lee, H.H., and Frasch, M. (2000). Wingless effects mesoderm patterning and ectoderm segmentation events via induction of its downstream target sloppy paired. *Development* *127*, 5497–5508.
- Leifer, D., Krainc, D., Yu, Y.T., McDermott, J., Breitbart, R.E., Heng, J., Neve, R.L., Kosofsky, B., Nadal-Ginard, B., and Lipton, S.A. (1993). MEF2C, a MADS/MEF2-family transcription factor expressed in a laminar distribution in cerebral cortex. *Proc. Natl. Acad. Sci. U. S. A.* *90*, 1546–1550.
- Lemercier, C., Verdel, A., Galloo, B., Curtet, S., Brocard, M.P., and Khochbin, S. (2000). mHDA1/HDAC5 histone deacetylase interacts with and represses MEF2A transcriptional activity. *J. Biol. Chem.* *275*, 15594–15599.
- Lemke, S.B., and Schnorrer, F. (2017). Mechanical forces during muscle development. *Mech. Dev.* *144*, 92–101.
- Leptin, M. (1999). Gastrulation in *Drosophila*: the logic and the cellular mechanisms. *EMBO J.* *18*, 3187–3192.
- Leptin, M. (2004). Gastrulation in *Drosophila*. In *Gastrulation: From Cells to Embryos*, C.D. Stern, ed. (New York: Cold Spring Harbor Laboratory Press), pp. 91–104.
- Leptin, M., and Grunewald, B. (1990). Cell shape changes during gastrulation in *Drosophila*. *Development* *110*, 73–84.
- Li, Y. (2011). The tandem affinity purification technology: an overview. *Biotechnol. Lett.* *33*, 1487–1499.

- Li, H., Russo, A., and DiAntonio, A. (2019). SIK3 suppresses neuronal hyperexcitability by regulating the glial capacity to buffer K⁺ and water. *J. Cell Biol.* *218*, 4017–4029.
- Li, P., Wei, X., Guan, Y., Chen, Q., Zhao, T., Sun, C., and Wei, L. (2014). MicroRNA-1 regulates chondrocyte phenotype by repressing histone deacetylase 4 during growth plate development. *FASEB J.* *28*, 3930–3941.
- Li, X., Wang, W., Wang, J., Malovannaya, A., Xi, Y., Li, W., Guerra, R., Hawke, D.H., Qin, J., and Chen, J. (2015). Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. *Mol. Syst. Biol.* *11*, 775.
- Lilly, B., Galewsky, S., Firulli, A.B., Schulz, R.A., and Olson, E.N. (1994). D-MEF2: a MADS box transcription factor expressed in differentiating mesoderm and muscle cell lineages during *Drosophila* embryogenesis. *Proc. Natl. Acad. Sci.* *91*, 5662–5666.
- Lilly, B., Zhao, B., Ranganayakulu, G., Paterson, B.M., Schulz, R. a, and Olson, E.N. (1995). Requirement of MADS domain transcription factor D-MEF2 for muscle formation in *Drosophila*. *Science* *267*, 688–693.
- Lin, W., and Baines, R.A. (2019). Myocyte enhancer factor-2 and p300 interact to regulate the expression of homeostatic regulator Pumilio in *Drosophila*. *Eur. J. Neurosci.* *50*, ejn.14357.
- Lin, Q., Lu, J., Yanagisawa, H., Webb, R., Lyons, G.E., Richardson, J.A., and Olson, E.N. (1998). Requirement of the MADS-box transcription factor MEF2C for vascular development. *Development* *125*, 4565–4574.
- Lin, T.Y., Viswanathan, S., Wood, C., Wilson, P.G., Wolf, N., Fuller, M.T., Alphey, L., Jimenez, J., White-Cooper, H., Dawson, I., et al. (1996). Coordinate developmental control of the meiotic cell cycle and spermatid differentiation in *Drosophila* males. *Development* *122*, 1331–1341.
- Liotta, D., Han, J., Elgar, S., Garvey, C., Han, Z., and Taylor, M. V. (2007). The Him Gene Reveals a Balance of Inputs Controlling Muscle Differentiation in *Drosophila*. *Curr. Biol.* *17*, 1409–1413.

- Liu, H., Sadygov, R.G., and Yates, J.R. (2004). A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Descr. Anal. Chem.* *76*, 4193–4201.
- Lu, Z., and Hunter, T. (2018). Metabolic Kinases Moonlighting as Protein Kinases. *Trends Biochem. Sci.* *43*, 301–310.
- Ma, H., McLean, J.R., Chao, L.F.-I., Mana-Capelli, S., Paramasivam, M., Hagstrom, K.A., Gould, K.L., and McCollum, D. (2012). A highly efficient multifunctional tandem affinity purification approach applicable to diverse organisms. *Mol. Cell. Proteomics* *11*, 501–511.
- Maeda, K., Poletto, M., Chiapparino, A., and Gavin, A.-C. (2014). A generic protocol for the purification and characterization of water-soluble complexes of affinity-tagged proteins and lipids. *Nat. Protoc.* *9*, 2256–2266.
- Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* *21*, 3448–3449.
- Martin, B.S., Ruiz-Gómez, M., Landgraf, M., and Bate, M. (2001). A distinct set of founders and fusion-competent myoblasts make visceral muscles in the *Drosophila* embryo. *Development* *128*, 3331–3338.
- Martin, J.F., Schwarz, J.J., and Olson, E.N. (1993). Myocyte enhancer factor (MEF) 2C: a tissue-restricted member of the MEF-2 family of transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* *90*, 5282–5286.
- Martin, J.F., Miano, J.M., Hustad, C.M., Copeland, N.G., Jenkins, N.A., and Olson, E.N. (1994). A Mef2 gene that generates a muscle-specific isoform via alternative mRNA splicing. *Mol. Cell. Biol.* *14*, 1647–1656.
- McDermott, J.C., Cardoso, M.C., Yu, Y.T., Andres, V., Leifer, D., Krainc, D., Lipton, S.A., and Nadal-Ginard, B. (1993). hMEF2C gene encodes skeletal muscle- and brain-specific transcription factors. *Mol. Cell. Biol.* *13*, 2564–2577.
- McKinsey, T.A., Zhang, C.L., Lu, J., and Olson, E.N. (2000). Signal-dependent

nuclear export of a histone deacetylase regulates muscle differentiation. *Nature* 408, 106–111.

McKinsey, T.A., Zhang, C.L., and Olson, E.N. (2001). Identification of a signal-responsive nuclear export sequence in class II histone deacetylases. *Mol. Cell. Biol.* 21, 6312–6321.

McKinsey, T.A., Zhang, C.L., and Olson, E.N. (2002a). MEF2: a calcium-dependent regulator of cell division, differentiation and death. *Trends Biochem. Sci.* 27, 40–47.

McKinsey, T.A., Zhang, C.L., and Olson, E.N. (2002b). Signaling chromatin to make muscle. *Curr. Opin. Cell Biol.* 14, 763–772.

McQuilton, P., St Pierre, S.E., Thurmond, J., and FlyBase Consortium (2012). FlyBase 101--the basics of navigating FlyBase. *Nucleic Acids Res.* 40, D706-14.

Mellor, J.C., Yanai, I., Clodfelter, K.H., Mintseris, J., and DeLisi, C. (2002). Predictome: a database of putative functional links between proteins. *Nucleic Acids Res.* 30, 306–309.

von Mering, C., Jensen, L.J., Snel, B., Hooper, S.D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M.A., and Bork, P. (2005). STRING: Known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 33.

Miska, E.A., Karlsson, C., Langley, E., Nielsen, S.J., Pines, J., and Kouzarides, T. (1999). HDAC4 deacetylase associates with and represses the MEF2 transcription factor. *EMBO J.* 18, 5099–5107.

Molkentin, J.D., Black, B.L., Martin, J.F., and Olson, E.N. (1995). Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell* 83, 1125–1136.

Morin, S., Charron, F., Robitaille, L., and Nemer, M. (2000). GATA-dependent recruitment of MEF2 proteins to target promoters. *EMBO J.* 19, 2046–2055.

Morisaki, T., Sermsuvitayawong, K., Byun, S.H., Matsuda, Y., Hidaka, K., Morisaki, H., and Mukai, T. (1997). Mouse Mef2b gene: unique member of MEF2 gene family.

J. Biochem. *122*, 939–946.

Murali, T., Pacifico, S., Yu, J., Guest, S., Roberts, G.G., and Finley, R.L. (2011). DroID 2011: A comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for *Drosophila*. *Nucleic Acids Res.* *39*.

Murali, T., Pacifico, S., and Finley, R.L. (2014). Integrating the interactome and the transcriptome of *Drosophila*. *BMC Bioinformatics* *15*, 177.

Nakatani, Y., and Ogryzko, V. (2003). Immunoaffinity Purification of Mammalian Protein Complexes. *Methods Enzymol.* *370*, 430–444.

Newman, S.A. (2020). Cell differentiation: What have we learned in 50 years? *J. Theor. Biol.* *485*, 110031.

Nguyen, H.T., Bodmer, R., Abmayr, S.M., McDermott, J.C., and Spoerel, N.A. (1994). D-mef2: a *Drosophila* mesoderm-specific MADS box-containing gene with a biphasic expression profile during embryogenesis. *Proc. Natl. Acad. Sci.* *91*, 7520–7524.

Nowak, S.J., Aihara, H., Gonzalez, K., Nibu, Y., and Baylies, M.K. (2012). Akirin Links Twist-Regulated Transcription with the Brahma Chromatin Remodeling Complex during Embryogenesis. *PLoS Genet.* *8*, e1002547.

Ohkawa, Y., Marfella, C.G.A., and Imbalzano, A.N. (2006). Skeletal muscle specification by myogenin and Mef2D via the SWI/SNF ATPase Brg1. *EMBO J.* *25*, 490–501.

Ornatsky, O.I., and McDermott, J.C. (1996). MEF2 protein expression, DNA binding specificity and complex composition, and transcriptional activity in muscle and non-muscle cells. *J. Biol. Chem.* *271*, 24927–24933.

Ornatsky, O.I., Andreucci, J.J., and McDermott, J.C. (1997). A dominant-negative form of transcription factor MEF2 inhibits myogenesis. *J. Biol. Chem.* *272*, 33271–33278.

Pan, D., Huang, J.D., and Courey, A.J. (1991). Functional analysis of the *Drosophila* twist promoter reveals a dorsal-binding ventral activator region. *Genes Dev.* *5*, 1892–

1901.

Paoletti, A.C., Parmely, T.J., Tomomori-Sato, C., Sato, S., Zhu, D., Conaway, R.C., Conaway, J.W., Florens, L., and Washburn, M.P. (2006). Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 18928–18933.

Pepper, A., Bhogal, B., and Jongens, T. (2012). Tandem Affinity Purification in *Drosophila* Heads and Ovaries. *Bio-Protocol* *2*.

Perrimon, N., Pitsouli, C., and Shilo, B.-Z. (2012). Signaling mechanisms controlling cell fate and embryonic patterning. *Cold Spring Harb. Perspect. Biol.* *4*, a005975.

Perry, R.L.S., Yang, C., Soora, N., Salma, J., Marback, M., Naghibi, L., Ilyas, H., Chan, J., Gordon, J.W., and McDermott, J.C. (2009). Direct interaction between myocyte enhancer factor 2 (MEF2) and protein phosphatase 1alpha represses MEF2-dependent gene expression. *Mol. Cell. Biol.* *29*, 3355–3366.

Philimonenko, V. V., Zhao, J., Iben, S., Dingová, H., Kyselá, K., Kahle, M., Zentgraf, H., Hofmann, W.A., de Lanerolle, P., Hozák, P., et al. (2004). Nuclear actin and myosin I are required for RNA polymerase I transcription. *Nat. Cell Biol.* *6*, 1165–1172.

Pollock, R., and Treisman, R. (1991). Human SRF-related proteins: DNA-binding properties and potential regulatory targets. *Genes Dev.* *5*, 2327–2341.

Potthoff, M.J., and Olson, E.N. (2007). MEF2: a central regulator of diverse developmental programs. *Development* *134*, 4131–4140.

Potthoff, M.J., Wu, H., Arnold, M.A., Shelton, J.M., Backs, J., McAnally, J., Richardson, J.A., Bassel-Duby, R., and Olson, E.N. (2007). Histone deacetylase degradation and MEF2 activation promote the formation of slow-twitch myofibers. *J. Clin. Invest.* *117*, 2459–2467.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* *24*, 218–229.

- Ranganayakulu, G., Zhao, B., Dokidis, A., Molkenin, J.D., Olson, E.N., and Schulz, R. a (1995). A series of mutations in the D-MEF2 transcription factor reveal multiple functions in larval and adult myogenesis in *Drosophila*. *Dev. Biol.* *171*, 169–181.
- Rao, S., Karray, S., Gackstetter, E.R., and Koshland, M.E. (1998). Myocyte enhancer factor-related B-MEF2 is developmentally expressed in B cells and regulates the immunoglobulin J chain promoter. *J. Biol. Chem.* *273*, 26123–26129.
- Reim, I., and Frasch, M. (2010). Genetic and genomic dissection of cardiogenesis in the *Drosophila* model. *Pediatr. Cardiol.* *31*, 325–334.
- Rhee, D.Y., Cho, D.Y., Zhai, B., Slattery, M., Ma, L., Mintseris, J., Wong, C.Y., White, K.P., Celniker, S.E., Przytycka, T.M., et al. (2014). Transcription factor networks in *Drosophila melanogaster*. *Cell Rep.* *8*, 2031–2043.
- Ridgeway, A.G., Wilton, S., and Skerjanc, I.S. (2000). Myocyte enhancer factor 2C and myogenin up-regulate each other's expression and induce the development of skeletal muscle in P19 cells. *J. Biol. Chem.* *275*, 41–46.
- Riechmann, V., Irion, U., Wilson, R., Grosskortenhaus, R., and Leptin, M. (1997). Control of cell fates and segmentation in the *Drosophila* mesoderm. *Development* *124*, 2915–2922.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Séraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* *17*, 1030–1032.
- Roth, S., Stein, D., and Nüsslein-Volhard, C. (1989). A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. *Cell* *59*, 1189–1202.
- Rudolf, A., Buttgereit, D., Jacobs, M., Wolfstetter, G., Kesper, D., Pütz, M., Berger, S., Renkawitz-Pohl, R., Holz, A., and Önel, S.F. (2014). Distinct genetic programs guide *Drosophila* circular and longitudinal visceral myoblast fusion. *BMC Cell Biol.* *15*, 27.
- Ruiz-Gómez, M. (1998). Muscle patterning and specification in *Drosophila*. *Int. J.*

Dev. Biol. 42, 283–290.

Rushlow, C.A., Han, K., Manley, J.L., and Levine, M. (1989). The graded distribution of the dorsal morphogen is initiated by selective nuclear transport in *Drosophila*. *Cell* 59, 1165–1177.

Rushton, E., Drysdale, R., Abmayr, S.M., Michelson, A.M., and Bate, M. (1995). Mutations in a novel gene, myoblast city, provide evidence in support of the founder cell hypothesis for *Drosophila* muscle development. *Development* 121, 1979–1988.

Sahota, V.K., Grau, B.F., Mansilla, A., and Ferrús, A. (2009). Troponin I and Tropomyosin regulate chromosomal stability and cell polarity. *J. Cell Sci.* 122, 2623–2631.

Sala, G., Stefanoni, G., Arosio, A., Riva, C., Melchionda, L., Saracchi, E., Fermi, S., Brighina, L., and Ferrarese, C. (2014). Reduced expression of the chaperone-mediated autophagy carrier hsc70 protein in lymphomonocytes of patients with Parkinson's disease. *Brain Res.* 1546, 46–52.

Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., and Eisenberg, D. (2004). The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* 32, D449-51.

San Martin, B., and Bate, M. (2001). Hindgut visceral mesoderm requires an ectodermal template for normal development in *Drosophila*. *Development* 128, 233–242.

Sandmann, T., Jensen, L.J., Jakobsen, J.S., Karzynski, M.M., Eichenlaub, M.P., Bork, P., and Furlong, E.E.M. (2006). A Temporal Map of Transcription Factor Activity: Mef2 Directly Regulates Target Genes at All Stages of Muscle Development. *Dev. Cell* 10, 797–807.

Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V., and Furlong, E.E.M. (2007). A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev.* 21, 436–449.

Sando, R., Gounko, N., Pieraut, S., Liao, L., Yates, J., and Maximov, A. (2012).

HDAC4 governs a transcriptional program essential for synaptic plasticity and memory. *Cell* 151, 821–834.

Sardiu, M.E., Cai, Y., Jin, J., Swanson, S.K., Conaway, R.C., Conaway, J.W., Florens, L., and Washburn, M.P. (2008). Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. *Proc. Natl. Acad. Sci. U. S. A.* 105, 1454–1459.

Schlicker, A., and Albrecht, M. (2008). FunSimMat: a comprehensive functional similarity database. *Nucleic Acids Res.* 36, D434-9.

Schlicker, A., and Albrecht, M. (2010). FunSimMat update: new features for exploring functional similarity. *Nucleic Acids Res.* 38, D244-8.

Schnorrer, F., and Dickson, B.J. (2004). Muscle building; mechanisms of myotube guidance and attachment site selection. *Dev. Cell* 7, 9–20.

Schnorrer, F., Schönbauer, C., Langer, C.C.H., Dietzl, G., Novatchkova, M., Schernhuber, K., Fellner, M., Azaryan, A., Radolf, M., Stark, A., et al. (2010). Systematic genetic analysis of muscle morphogenesis and function in *Drosophila*. *Nature* 464, 287–291.

Schröter, R.H., Buttgereit, D., Beck, L., Holz, A., and Renkawitz-Pohl, R. (2006). Blown fuse regulates stretching and outgrowth but not myoblast fusion of the circular visceral muscles in *Drosophila*. *Differentiation*. 74, 608–621.

Schubert, O.T., Röst, H.L., Collins, B.C., Rosenberger, G., and Aebersold, R. (2017). Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat. Protoc.* 12, 1289–1294.

Sellin, J., Albrecht, S., Kölsch, V., and Paululat, A. (2006). Dynamics of heart differentiation, visualized utilizing heart enhancer elements of the *Drosophila melanogaster* bHLH transcription factor Hand. *Gene Expr. Patterns* 6, 360–375.

Serra, C., Palacios, D., Mozzetta, C., Forcales, S. V, Morantte, I., Ripani, M., Jones, D.R., Du, K., Jhala, U.S., Simone, C., et al. (2007). Functional interdependence at the chromatin level between the MKK6/p38 and IGF1/PI3K/AKT pathways during

muscle differentiation. *Mol. Cell* 28, 200–213.

Shalizi, A.K., and Bonni, A. (2005). brawn for brains: the role of MEF2 proteins in the developing nervous system. *Curr. Top. Dev. Biol.* 69, 239–266.

Sharma, G., and Goalstone, M.L. (2005). Dominant negative FTase (DNFTalpha) inhibits ERK5, MEF2C and CREB activation in adipogenesis. *Mol. Cell. Endocrinol.* 245, 93–104.

Sivachenko, A., Li, Y., Abruzzi, K.C., and Rosbash, M. (2013). The transcription factor Mef2 links the drosophila core clock to Fas2, neuronal morphology, and circadian behavior. *Neuron* 79, 281–292.

Small, S., Kraut, R., Hoey, T., Warrior, R., and Levine, M. (1991). Transcriptional regulation of a pair-rule stripe in *Drosophila*. *Genes Dev.* 5, 827–839.

Snyder, C.M., Rice, A.L., Estrella, N.L., Held, A., Kandarian, S.C., and Naya, F.J. (2013). MEF2A regulates the Gtl2-Dio3 microRNA mega-cluster to modulate WNT signaling in skeletal muscle regeneration. *Development* 140, 31–42.

Soler, C., and Taylor, M. V (2009). The Him gene inhibits the development of *Drosophila* flight muscles during metamorphosis. *Mech. Dev.* 126, 595–603.

Soler, C., Han, J., and Taylor, M. V (2012). The conserved transcription factor Mef2 has multiple roles in adult *Drosophila* musculature formation. *Development* 139, 1270–1275.

Sonnemann, K.J., Fitzsimons, D.P., Patel, J.R., Liu, Y., Schneider, M.F., Moss, R.L., and Ervasti, J.M. (2006). Cytoplasmic gamma-Actin Is Not Required for Skeletal Muscle Development but Its Absence Leads to a Progressive Myopathy. *Dev. Cell* 11, 387–397.

Sonnenfeld, M.J. (2009). GAL4/UAS Expression System. In *Encyclopedia of Neuroscience*, M.D. Binder, N. Hirokawa, and U. Windhorst, eds. (Berlin, Heidelberg: Springer), pp. 1662–1666.

Sparrow, D.B., Miska, E.A., Langley, E., Reynaud-Deonauth, S., Kotecha, S., Towers, N., Spohr, G., Kouzarides, T., and Mohun, T.J. (1999). MEF-2 function is

modified by a novel co-repressor, MITR. *EMBO J.* 18, 5085–5098.

Spletter, M.L., Barz, C., Yeroslaviz, A., Zhang, X., Lemke, S.B., Bonnard, A., Brunner, E., Cardone, G., Basler, K., Habermann, B.H., et al. (2018). A transcriptomics resource reveals a transcriptional transition during ordered sarcomere morphogenesis in flight muscle. *Elife* 7, e34058.

Staebling-Hampton, K., Jackson, P.D., Clark, M.J., Brand, A.H., and Hoffmann, F.M. (1994). Specificity of bone morphogenetic protein-related factors: cell fate and gene expression changes in *Drosophila* embryos induced by decapentaplegic but not 60A. *Cell Growth Differ.* 5, 585–593.

Stanojevic, D., Small, S., and Levine, M. (1991). Regulation of a segmentation stripe by overlapping activators and repressors in the *Drosophila* embryo. *Science* 254, 1385–1387.

Stanyon, C.A., Liu, G., Mangiola, B.A., Patel, N., Giot, L., Kuang, B., Zhang, H., Zhong, J., and Finley, R.L. (2004). A *Drosophila* protein-interaction map centered on cell-cycle regulators. *Genome Biol.* 5, R96.

Sudarsan, V., Anant, S., Guptan, P., VijayRaghavan, K., and Skaer, H. (2001). Myoblast diversification and ectodermal signaling in *Drosophila*. *Dev. Cell* 1, 829–839.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613.

Tanaka, K.K.K., Bryantsev, A.L., and Cripps, R.M. (2008). Myocyte enhancer factor 2 and chorion factor 2 collaborate in activation of the myogenic program in *Drosophila*. *Mol. Cell. Biol.* 28, 1616–1629.

Taylor, M. V. (1995). Muscle Development: Making *Drosophila* muscle. *Curr. Biol.* 5, 740–742.

Taylor, M. V (2000). A novel *Drosophila*, *mef2*-regulated muscle gene isolated in a subtractive hybridization-based molecular screen using small amounts of zygotic mutant RNA. *Dev. Biol.* 220, 37–52.

Taylor, M. V, and Hughes, S.M. (2017). *Mef2* and the skeletal muscle differentiation program. *Semin. Cell Dev. Biol.* 72, 33–44.

Taylor, J.M., Dupont-Versteegden, E.E., Davies, J.D., Hassell, J.A., Houlé, J.D., Gurley, C.M., and Peterson, C.A. (1997). A role for the ETS domain transcription factor PEA3 in myogenic differentiation. *Mol. Cell. Biol.* 17, 5550–5558.

Taylor, M. V., Beatty, K.E., Hunter, H.K., and Baylies, M.K. (1995). *Drosophila* MEF2 is regulated by *twist* and is expressed in both the primordia and differentiated cells of the embryonic somatic, visceral and heart musculature. *Mech. Dev.* 50, 139–141.

The modENCODE Consortium, Roy, S., Ernst, J., Kharchenko, P. V, Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., et al. (2010). Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* 330, 1787–1797.

The UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515.

Thisse, C., Perrin-Schmitt, F., Stoetzel, C., and Thisse, B. (1991). Sequence-specific transactivation of the *Drosophila* *twist* gene by the dorsal gene product. *Cell* 65, 1191–1201.

Ticho, B.S., Stainier, D.Y., Fishman, M.C., and Breitbart, R.E. (1996). Three zebrafish MEF2 genes delineate somitic and cardiac muscle development in wild-type and mutant embryos. *Mech. Dev.* 59, 205–218.

Tixier, V., Bataillé, L., and Jagla, K. (2010). Diversification of muscle types: Recent insights from *Drosophila*. *Exp. Cell Res.* 316, 3019–3027.

Tixier, V., Bataille, L., Etard, C., Jagla, T., Weger, M., DaPonte, J.P., Strahle, U., Dickmeis, T., and Jagla, K. (2013). Glycolysis supports embryonic muscle growth by

promoting myoblast fusion. *Proc. Natl. Acad. Sci.* *110*, 18982–18987.

Tobin, S.L., Cook, P.J., and Burn, T.C. (1990). Transcripts of individual *Drosophila* actin genes are differentially distributed during embryogenesis. *Dev. Genet.* *11*, 15–26.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., et al. (2015). Tissue-based map of the human proteome. *Science* *347*, 1260419–1260419.

Urrutia, R., McNiven, M.A., and Kachar, B. (1993). Synthesis of RNA probes by the direct in vitro transcription of PCR-generated DNA templates. *J. Biochem. Biophys. Methods* *26*, 113–120.

Vega, R.B., Matsuda, K., Oh, J., Barbosa, A.C., Yang, X., Meadows, E., McAnally, J., Pomajzl, C., Shelton, J.M., Richardson, J.A., et al. (2004). Histone deacetylase 4 controls chondrocyte hypertrophy during skeletogenesis. *Cell* *119*, 555–566.

Venit, T., Xie, X., and Percipalle, P. (2018). Actin in the Cell Nucleus. In *Nuclear Architecture and Dynamics*, C. Lavelle, and J.-M. Victor, eds. (Cambridge, MA: Academic Press), pp. 345–367.

Veraksa, A., Bauer, A., and Artavanis-Tsakonas, S. (2005). Analyzing protein complexes in *Drosophila* with tandem affinity purification-mass spectrometry. *Dev. Dyn.* *232*, 827–834.

Verzi, M.P., Agarwal, P., Brown, C., McCulley, D.J., Schwarz, J.J., and Black, B.L. (2007). The transcription factor MEF2C is required for craniofacial development. *Dev. Cell* *12*, 645–652.

Viala, J.P.M., and Bouveret, E. (2017). Protein–Protein Interaction: Tandem Affinity Purification in Bacteria. In *Bacterial Protein Secretion Systems*, L. Journet, and E. Cascales, eds. (New York: Humana Press), pp. 221–232.

Vigoreaux, J.O. (2001). Genetics of the *Drosophila* flight muscle myofibril: a window into the biology of complex systems. *Bioessays* *23*, 1047–1063.

Visa, N. (2005). Actin in transcription. *EMBO Rep.* *6*, 218–219.

Visa, N., and Percipalle, P. (2010). Nuclear Functions of Actin. *Cold Spring Harb. Perspect. Biol.* 2, a000620.

Wang, B., Moya, N., Niessen, S., Hoover, H., Mihaylova, M.M., Shaw, R.J., Yates, J.R., Fischer, W.H., Thomas, J.B., and Montminy, M. (2011). A hormone-dependent module regulating energy balance. *Cell* 145, 596–606.

Wang, C., Arrington, J., Ratliff, A.C., Chen, J., Horton, H.E., Nie, Y., Yue, F., Hrycyna, C.A., Tao, W.A., and Kuang, S. (2019). Methyltransferase-like 21c methylates and stabilizes the heat shock protein Hspa8 in type I myofibers in mice. *J. Biol. Chem.* 294, 13718–13728.

Wang, D.Z., Valdez, M.R., McAnally, J., Richardson, J., and Olson, E.N. (2001). The *Mef2c* gene is a direct transcriptional target of myogenic bHLH and MEF2 proteins during skeletal muscle development. *Development* 128, 4623–4633.

Ward, E.J., and Skeath, J.B. (2000). Characterization of a novel subset of cardiac cells and their progenitors in the *Drosophila* embryo. *Development* 127, 4959–4969.

Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T.K., Turner, D., Rupp, R., and Hollenberg, S. (1991). The *myoD* gene family: nodal point during specification of the muscle cell lineage. *Science* 251, 761–766.

Weitkunat, M., and Schnorrer, F. (2014). A guide to study *Drosophila* muscle biology. *Methods* 68, 2–14.

Wodarz, A., Hinz, U., Engelbert, M., and Knust, E. (1995). Expression of crumbs confers apical character on plasma membrane domains of ectodermal epithelia of *Drosophila*. *Cell* 82, 67–76.

Wu, W., Folter, S. de, Shen, X., Zhang, W., and Tao, S. (2011). Vertebrate Paralogous MEF2 Genes: Origin, Conservation, and Evolution. *PLoS One* 6, e17334.

Yang, X.-J., and Grégoire, S. (2005). Class II histone deacetylases: from sequence to function, regulation, and clinical implication. *Mol. Cell. Biol.* 25, 2873–2884.

Yang, X.-J., and Seto, E. (2008). The Rpd3/Hda1 family of lysine deacetylases: from

bacteria and yeast to mice and men. *Nat. Rev. Mol. Cell Biol.* 9, 206–218.

Yang, Q., She, H., Gearing, M., Colla, E., Lee, M., Shacka, J.J., and Mao, Z. (2009). Regulation of neuronal survival factor MEF2D by chaperone-mediated autophagy. *Science* 323, 124–127.

Yin, Z., Xu, X.L., and Frasch, M. (1997). Regulation of the twist target gene tinman by modular cis-regulatory elements during early mesoderm development. *Development* 124, 4971–4982.

Yu, C., Wan, K.H., Hammonds, A.S., Stapleton, M., Carlson, J.W., and Celniker, S.E. (2011). Development of Expression-Ready Constructs for Generation of Proteomic Libraries. In *Methods in Molecular Biology (Methods and Protocols)*, Vol 723, C. Wu, ed. (New York: Humana Press), pp. 257–272.

Yu, J., Pacifico, S., Liu, G., and Finley, R.L. (2008). DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. *BMC Genomics* 9, 461.

Yu, Y.T., Breitbart, R.E., Smoot, L.B., Lee, Y., Mahdavi, V., and Nadal-Ginard, B. (1992). Human myocyte-specific enhancer factor 2 comprises a group of tissue-restricted MADS box transcription factors. *Genes Dev.* 6, 1783–1798.

Zaffran, S., and Frasch, M. (2002). Early signals in cardiac development. *Circ. Res.* 91, 457–469.

Zaffran, S., Küchler, A., Lee, H.H., and Frasch, M. (2001). biniou (FoxF), a central component in a regulatory network controlling visceral mesoderm development and midgut morphogenesis in Drosophila. *Genes Dev.* 15, 2900–2915.

Zanin, E., Dumont, J., Gassmann, R., Cheeseman, I., Maddox, P., Bahmanyar, S., Carvalho, A., Niessen, S., Yates, J.R., Oegema, K., et al. (2011). Affinity Purification of Protein Complexes in *C. elegans*. *Methods Cell Biol.* 106, 289–322.

Zanon, A., Rakovic, A., Blankenburg, H., Doncheva, N.T., Schwienbacher, C., Serafin, A., Alexa, A., Weichenberger, C.X., Albrecht, M., Klein, C., et al. (2013). Profiling of Parkin-binding partners using tandem affinity purification. *PLoS One* 8,

e78648.

Zeremski, M., Stricker, J.R., Fischer, D., Zusman, S.B., and Cohen, D. (2003). Histone deacetylase dHDAC4 is involved in segmentation of the *Drosophila* embryo and is regulated by gap and pair-rule genes. *Genesis* 35, 31–38.

Zhang, C.L., McKinsey, T.A., and Olson, E.N. (2001). The transcriptional corepressor MITR is a signal-responsive inhibitor of myogenesis. *Proc. Natl. Acad. Sci.* 98, 7354–7359.

Zhang, L., Sun, Y., Fei, M., Tan, C., Wu, J., Zheng, J., Tang, J., Sun, W., Lv, Z., Bao, J., et al. (2014). Disruption of chaperone-mediated autophagy-dependent degradation of MEF2A by oxidative stress-induced lysosome destabilization. *Autophagy* 10, 1015–1035.

Zhou, N., Jiang, Y., Bergquist, T.R., Lee, A.J., Kacsoh, B.Z., Crocker, A.W., Lewis, K.A., Georghiou, G., Nguyen, H.N., Hamid, M.N., et al. (2019). The CAFA challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens. *Genome Biol.* 20, 244.

Zhou, X., Richon, V.M., Rifkind, R., and Marks, P. (2000). Identification of a transcriptional repressor related to the noncatalytic domain of histone deacetylases 4 and 5. *Proc. Natl. Acad. Sci.* 97, 1056–1061.

Zhu, B., and Gulick, T. (2004). Phosphorylation and alternative pre-mRNA splicing converge to regulate myocyte enhancer factor 2C activity. *Mol. Cell. Biol.* 24, 8264–8275.

Zhu, B., Ramachandran, B., and Gulick, T. (2005). Alternative pre-mRNA splicing governs expression of a conserved acidic transactivation domain in myocyte enhancer factor 2 factors of striated muscle and brain. *J. Biol. Chem.* 280, 28749–28760.

Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M., and Furlong, E.E.M. (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* 462, 65–70.

Zybailov, B., Mosley, A.L., Sardi, M.E., Coleman, M.K., Florens, L., and Washburn, M.P. (2006). Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* 5, 2339–2347.