



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par

l'Université Toulouse III - Paul Sabatier

Discipline ou spécialité

Biologie moléculaire, cellulaire et du Développement

Présentée et soutenue par
Hadi Boukhatmi

Le 26 Octobre 2012

Mise en place de l'identité des muscles au cours de la spécification des myoblastes chez la *Drosophile*.

Ecole doctorale

Biologie, Santé, Biotechnologies (BSB)

Unité de recherche

Centre de Biologie du Développement

Directeurs de thèse

Laurence Dubois et Alain Vincent

Membres du jury

Pr Cathy Soula, président

Dr Sigolene Meilhac, rapporteur

Dr Laurent Perrin, rapporteur

Dr Stéphane Zaffran, examinateur

Dr Laurence Dubois, co-directeur de thèse

Dr Alain Vincent, directeur de thèse



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par

l'Université Toulouse III - Paul Sabatier

Discipline ou spécialité

Biologie moléculaire, cellulaire et du Développement

Présentée et soutenue par
Hadi Boukhatmi

Le 26 Octobre 2012

Mise en place de l'identité des muscles au cours de la spécification des myoblastes chez la *Drosophile*.

Ecole doctorale

Biologie, Santé, Biotechnologies (BSB)

Unité de recherche

Centre de Biologie du Développement

Directeurs de thèse

Laurence Dubois et Alain Vincent

Rapporteurs

Dr Sigolene Meilhac, rapporteur

Dr Laurent Perrin, rapporteur

Membres du jury

Pr Cathy Soula, Président

Dr Sigolene Meilhac, rapporteur

Dr Laurent Perrin, rapporteur

Dr Stéphane Zaffran, rapporteur

Dr Laurence Dubois, co-directeur de thèse

Dr Alain Vincent, directeur de thèse

Remerciements

Merci à toi Alain, pour ton aide, ta disponibilité, et surtout d'avoir cru en moi. Ce fut un réel plaisir et un privilège de travailler avec toi. Promis, dans l'avenir je penserai à m'entraîner moins et à faire beaucoup plus de contrôles.

Merci à toi Laurence pour ton coaching, ta patience, ton amitié, et pour tout ce que tu m'as appris. Grâce à toi je sais maintenant que manger des madeleines prévient le mal de dos. To be suivre ;).

Mes sincères remerciements à Cathy Soula, Sigolene Meilhac, Laurent Perrin, Stéphane Zaffran pour avoir pris le temps de juger mon travail.

Merci à toi Michèle, pour ta présence, ton aide si précieuse et pour les nombreuses discussions qu'on a pu avoir et qui m'ont beaucoup aidé à avancer dans mes travaux.

Merci à toi Jean-Louis pour tous les bons moments passés ensemble. Ne t'inquiètes pas un jour on retrouvera la boîte bleue de Renée.

Merci à Laetitia et Jonathan d'avoir partagé avec moi votre passion et votre enthousiasme pour la recherche.

Merci à Joanna, pour ton amitié, ta gentillesse et pour ton soutien lors de mon M2R et de ma thèse. Vive la Pologne !

Merci à tous les membres de notre équipe Delphine, Rami, Virginie, Nathalie, Caroline M, Caroline D, Yanick, Mathilde, Ismaël, Isabelle, pour votre bonne humeur au quotidien. Vous allez tous me manquer.

Merci à ma tutrice Muriel pour tes idées, tes conseils, ton investissement dans mon travail et surtout d'avoir répondu présente à chaque fois que j'ai eu besoin de toi.

Merci aux copains pour tous les moments que l'on a pu partager : Elodie, je te souhaite bonne chance pour ton aventure suisse et rendez vous sur les bassins. Yacine et Billel zoufri forever ☺.

Merci à tous les membres du CBD.

A mes chères sœurs Yasmine et Leila.

A ma bien aimée Myriam.

A mes très chers parents.

Résumé

La formation des muscles squelettiques au cours de l'embryogenèse de la drosophile est un modèle d'étude du contrôle génétique de la différenciation cellulaire. La formation de chaque muscle comprend quatre étapes successives: spécification d'un groupe promusculaire, sélection d'un progéniteur (PC) à partir de ce groupe, division asymétrique de ce progéniteur pour donner des cellules fondatrices de muscles (FC) ; fusion de chaque FC avec des myoblastes compétents (FCM), suivie de la différenciation musculaire. Chaque muscle squelettique est composé d'une fibre. Chaque muscle présente des propriétés spécifiques de taille, forme, position, attachement, et patron d'innervation. Ces propriétés sont groupées sous le terme d'identité musculaire. Cette identité est conférée par l'expression dans chaque PC/FC d'une combinaison de Facteurs de Transcription identitaires (FTi). Notre laboratoire étudie ce processus, en utilisant comme point d'entrée l'expression et les rôles du FTi Collier (Col) au cours du développement d'un muscle dorso-latéral, le muscle DA3 (Dorsal Acute 3).

Au cours de la première partie de ma thèse, j'ai étudié la régulation transcriptionnelle de *col* durant les phases de spécification des groupes promusculaires et de sélection du PC à l'origine du muscle DA3. Partant de prédictions bioinformatiques j'ai caractérisé le module cis régulateur (CRM) de *col* actif durant ces phases (CRM précoce). Un CRM « tardif », actif du stade progéniteur à la complétion de la formation du muscle DA3, avait été préalablement caractérisé dans l'équipe. Afin de déterminer plus précisément les fenêtres temporelles d'activité des deux CRM mésodermiques de *col*, j'ai mis au point un nouveau gène rapporteur comportant un intron permettant de détecter les transcrits primaires. Ceci m'a permis de montrer que les CRM précoce et tardif reproduisent ensemble l'expression endogène de *col*. La caractérisation du CRM précoce de *col* m'a aussi permis de suivre le destin des FCM du groupe promusculaire Col dans les embryons tardifs et de montrer que ces FCM contribuent uniquement à des muscles dorsaux-latéraux.

Au cours de la deuxième partie de ma thèse, j'ai caractérisé le rôle, inconnu jusqu'alors, du FT à domaine LIM-Homeodomaine Tailup (*Tup*)/*Islet1* dans la myogenèse. J'ai d'abord montré que *Tup* est spécifiquement exprimé dans les 4 muscles les plus dorsaux. L'analyse de mutants m'a permis de montrer qu'en absence de *Tup*, le muscle dorsal DA2 exprime Col et est transformé en muscle dorso-latéral de type DA3. J'ai ensuite montré que le PC du DA2 est à l'origine de la FC DA2 et d'un précurseur musculaire adulte (AMP). Ce PC est sélectionné à partir du groupe promusculaire Col quand les cellules de ce groupe expriment encore le FT à homéodomaine Tinman/*NKx2.5*. Tin active *tup* dans le PC DA2. *Tup*, en retour, réprime *col* et cette répression permet de distinguer les identités musculaires DA2 et DA3.

En conclusion, mes travaux de thèse m'ont permis de proposer un nouveau modèle permettant de relier le processus de spécification des progéniteurs au contrôle temporel et spatial de l'expression des FTi. Une vision dynamique de ce processus de spécification permet de mieux comprendre le programme identitaire propre à chaque muscle. L'analyse des interactions entre Tin, *Tup*, et Col au cours de la formation des muscles dorsaux révèle de nouveaux parallèles avec les interactions entre *Nkx2.5*, *Islet*, EBF au cours de la formation des muscles pharyngaux chez les chordés.

Listes des abréviations

ADN Acide désoxyribonucléique
AMP Adult Muscle Precursor
Ap Apterous
Ara Araucan
ASM Atrial siphon muscle
bHLH basic Helix-Loop-Helix
BLRP Biotin ligase recognition peptide
BMP Bone Morphogenetic Protein
BRM Brahma containing
Caps Capricious
Cau Caupolican
COE Collier/Olfactory-1/Early B cell Factor
Col Collier
Con Connectin
CRM Cis-regulatory module
DA Dorsal acute
DME Dorsal Muscle Enhancer
DO Dorsal oblique
Dpp Decapentaplegic
DT Dorsal transverse
dTCF Drosophila T Cell Factor
Duf Dumbfounded
EBF Early B-cell Factor
E^{col}-CRM Early *collier* cis regulatory module
EGF Epidermal Growth Factor
EPC Eve pericardial cell
Eve Even-skipped
FC Founder Cell
FCM Fusion Competent Myoblast
FT Facteur de transcription
FTi Facteur de transcription identitaire
GFP Green fluorescent protein
GMC Ganglion mother cell
GP Groupe Promusculaire
H2b Histone 2b
Hbs Hibris
Htl Heartless
IgSF Immunoglobulin superfamily
Insc Inscuteable
Iro-C Iroquois gene complex
Isl1 Islet-1
Kr Kruppel
L'sc Lethal of Scute
Lb Ladybird
Lbx1 Ladybird-like
L^{col}-CRM Late *collier* cis regulatory module
LL Lateral Longitudinal
Lms Lateral muscles scarcer
LRR Leucine-Rich Repeat
LT Latero-transverse
MA Muscle Alaire
Mad Mother Against dpp

MADS MCM1, Agamous, Deficiens, Serum reponse factor
MAPK Mitogen Activated Protein Kinase
MAT Muscles alaires transversaux
Mbc Myoblast city
MEC Matrice extracellulaire
Mef2 Myogenic enhancer factor 2
MET Microscopie électronique en transmission
MHC Myosin Heavy Chain
MHE Muscle and Heart Enhancer
Moe Moesin
MRF Myogenic Regulatory Factor Family
Msx1 Muscle segment homeobox 1
Myog Myogenin
Nau Nautilus
Org-1 Optomotor-blind-related-gene-1
PC Progenitor Cell
Poxm Pox meso
PTB Phosphotyrosine-binding motif
RFP Red fluorescent protein
Rst Roughest
RTK Récepteur à tyrosine kinase
SBM Segment border muscle
Shh Sonic hedgehog
Slp Sloppy paired
Sna Snail
Sns Sticks-and-Stones
Sr Stripe
Su(H) Suppressor of Hairless
SvpC Seven Up Cardioblasts
TCS Territoire cardiaque secondaire
Tey Teyrha-Meyhra
Tin Tinman
Tup Tailup
TVC Trunk Ventral Cells
Twi Twist
Wg Wingless
Zfh1 Zinc finger homeodomain 1

Introduction

A. La myogenèse squelettique chez les vertébrés. (Page 17)

- 1. Le patron musculaire (Page 17)**
- 2. Origine des muscles squelettiques chez les vertébrés ; les somites (Page 17)**
- 3. Aspects moléculaires et tissulaires de la myogenèse (Page 19)**
- 4. MRF (Myogenic Regulatory Factor Family) : principaux acteurs de la différenciation des myoblastes (Page 21)**

B. La drosophile : modèle d'étude de la myogenèse. (Page 23)

1. Le mésoderme somatique chez la drosophile (Page 25)

- a. Le mésoderme somatique à l'origine des muscles
- b. Régionalisation antero-postérieure du mésoderme
- c. Régionalisation dorso-ventrale du mésoderme
- d. Régulation transcriptionnelle de la myogenèse chez la drosophile

2. Etapes de formation des muscles somatiques chez la drosophile (Page 35)

- a. Spécification des groupes promusculaires et sélection des progéniteurs
- b. Les cellules fondatrices
- c. Division asymétrique des progéniteurs : Les divers lignages.
- d. Fusion des myoblastes: Description cellulaire et moléculaire
- e. Attachement des fibres musculaires à l'épiderme
- f. Innervation

3. Mise en place du programme d'identité musculaire au cours de la spécification des myoblastes (Page 49)

- a. Slouch et Apterous: les premiers gènes d'identité musculaire identifiés
- b. Lb et S59 : Exemple de répression croisée entre FTi
- c. Expression d'Even-skipped (Eve) dans des muscles dorsaux
- d. Interactions entre Lms, Ap et Msh et formation des muscles LT
- e. Morphologie des muscles dorso-latéraux : La combinatoire Collier/Nautilus.
- f. Araucan et Caupolican ; régulation post-traductionnelle de l'activité des FTi.
- g. Les muscles ventraux : le rôle de Poxm

4. Les muscles alaires (Page 61)

5. Les cellules précurseurs des muscles adultes (Page 63)

6. Les réseaux de régulation du mésoderme cardio-pharyngal chez les chordes ; le Facteur Islet 1 (Page 65)

7. Conclusion et nouvelles questions (Page 67)

Résultats

A. Régulation transcriptionnelle de Collier et identité musculaire. (Page 71)

Article 1. Multi-step control of muscle diversity by Hox proteins in the drosophila embryo.

1. Caractérisation du E^{col}-CRM de *col*.
2. Fenêtres temporelles d'activité des CRM.

B. Analyse du destin des FCM issus du groupe promusculaire Col. (Page 75)

C. *tup/islet* , un nouveau gène d'identité musculaire. (Page 109)

Article 2. Tup/Islet1 integrates time and position to specify muscle identity in Drosophila.

Résultats supplémentaires (Page 143)

1. Tup réprime la transcription de *col* via le L^{Col}CRM
2. Autorégulation directe de Tup via le DME (CRM3.3)

D. Le rôle d'Org-1 et Tup dans la spécification et/ou la réalisation de l'identité des muscles alaires. (Page 147)

1. Tup est exprimé et requis pour la formation des MA
2. Tup est exprimé au cours des étapes précoces de la formation des MA
3. Org-1 agit en amont de Tup dans le lignage MA

Discussion (Page 153)

Matériel et méthodes (Page 169)

Bibliographie (Page 175)

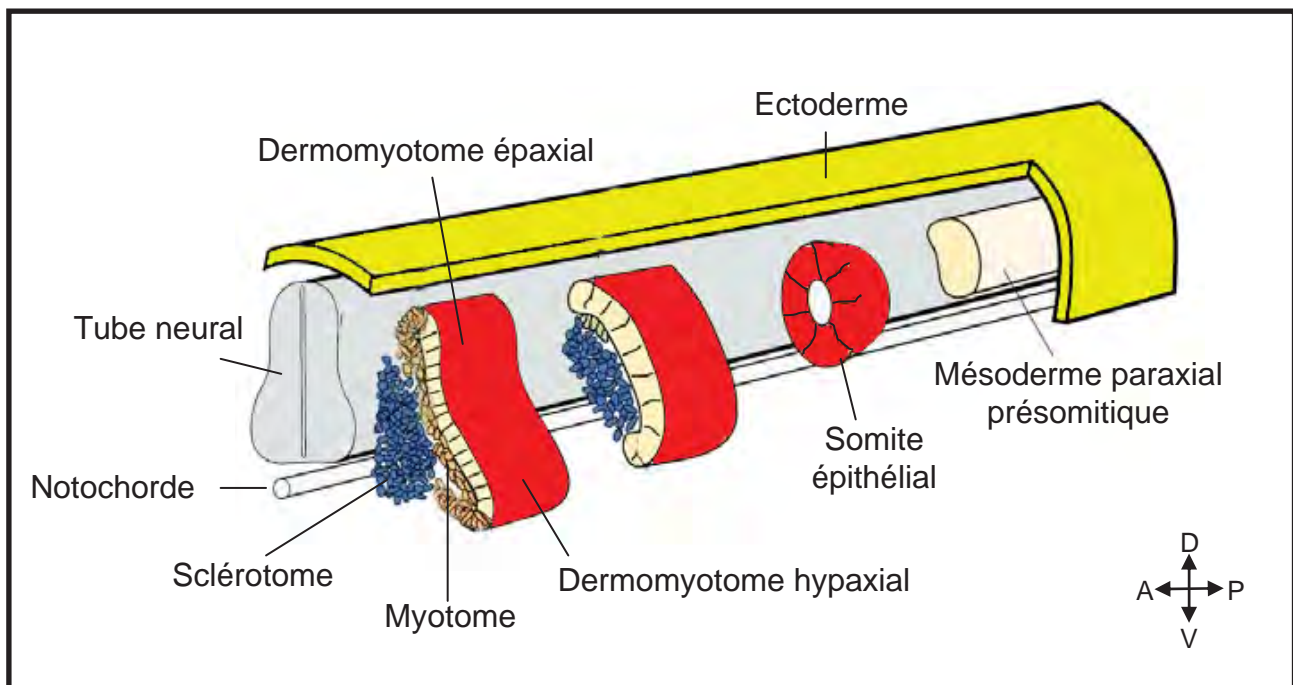


Figure 1. Représentation schématique de la somitogenèse chez les vertébrés. Le mésoderme paraxial se segmente en structures épithéliales, les somites, à l'origine du myotome, du sclérotome et du dermomyotome épaxial et hypaxial. *D'après Buckingham M. et al., 2003*

A. La myogenèse squelettique chez les vertébrés.

1. Le patron musculaire

Un atlas d'anatomie humaine illustre bien la diversité et la complexité de la musculature qui permet l'ensemble des mouvements coordonnés et involontaires propres à chaque espèce animale. On peut distinguer trois types musculaires : i) Les muscles squelettiques de formes et tailles variées qui assurent les mouvements, le maintien du corps, et la stabilité des articulations. Les muscles squelettiques sont formés de fibres musculaires ou myofibrilles dont l'organisation en sarcomères leur confère un aspect strié. Chaque fibre est une cellule syncytiale multi-nucléée. Leur activité est contrôlée par le système nerveux central. ii) les muscles lisses dont la contraction est involontaire, sous le contrôle du système nerveux autonome. Cette musculature assure les fonctions dites végétatives telles que le transit intestinal ou le transport du sang. Les muscles lisses sont composés de cellules contractiles fusiformes mononucléées. iii) le muscle cardiaque qui a l'aspect d'un muscle strié, mais dont les contractions sont autonomes, bien que le rythme et la puissance des contractions puissent être modulés par des stimuli issus du système autonome ou des stimuli hormonaux. La musculature se met en place au cours du développement embryonnaire grâce à un processus appelé myogenèse. Bien que l'étude cellulaire et moléculaire de la myogenèse ait été principalement réalisée sur des organismes vertébrés, le contrôle de la diversité morphologique des muscles squelettiques a été peu abordée. Les chapitres suivants de l'introduction sont essentiellement consacrés à cette question.

2. Origine des muscles squelettiques chez les vertébrés ; les somites

Il existe environ 600 muscles squelettiques chez l'homme, répartis en trois grandes classes : muscles du tronc, des membres, et de la tête. Les muscles du tronc et des membres dérivent de progéniteurs issus de structures épithéliales, appelées somites. Les somites sont des structures transitoires dérivant de la segmentation du mésoderme paraxial localisé de part et d'autre du tube neural et de la notochorde. Les somites sont régionalisés avec une structure ventrale appelée sclérotome et une structure dorsale appelée dermomyotome. Cette régionalisation dorso-ventrale est dépendante de signaux émis par les tissus voisins, le tube neural, la notochorde, l'ectoderme dorsal et le mésoderme latéral (Figure 1). Outre les muscles, les somites sont à l'origine de plusieurs types cellulaires. Le sclérotome donne le squelette du tronc et le dermomyotome donne, comme son nom l'indique, le derme. Les muscles des membres proviennent de la partie latéro-ventrale du dermomyotome appelée dermomyotome hypaxial (Christ and Ordahl, 1995; Buckingham, 2001). Le syndetome, à

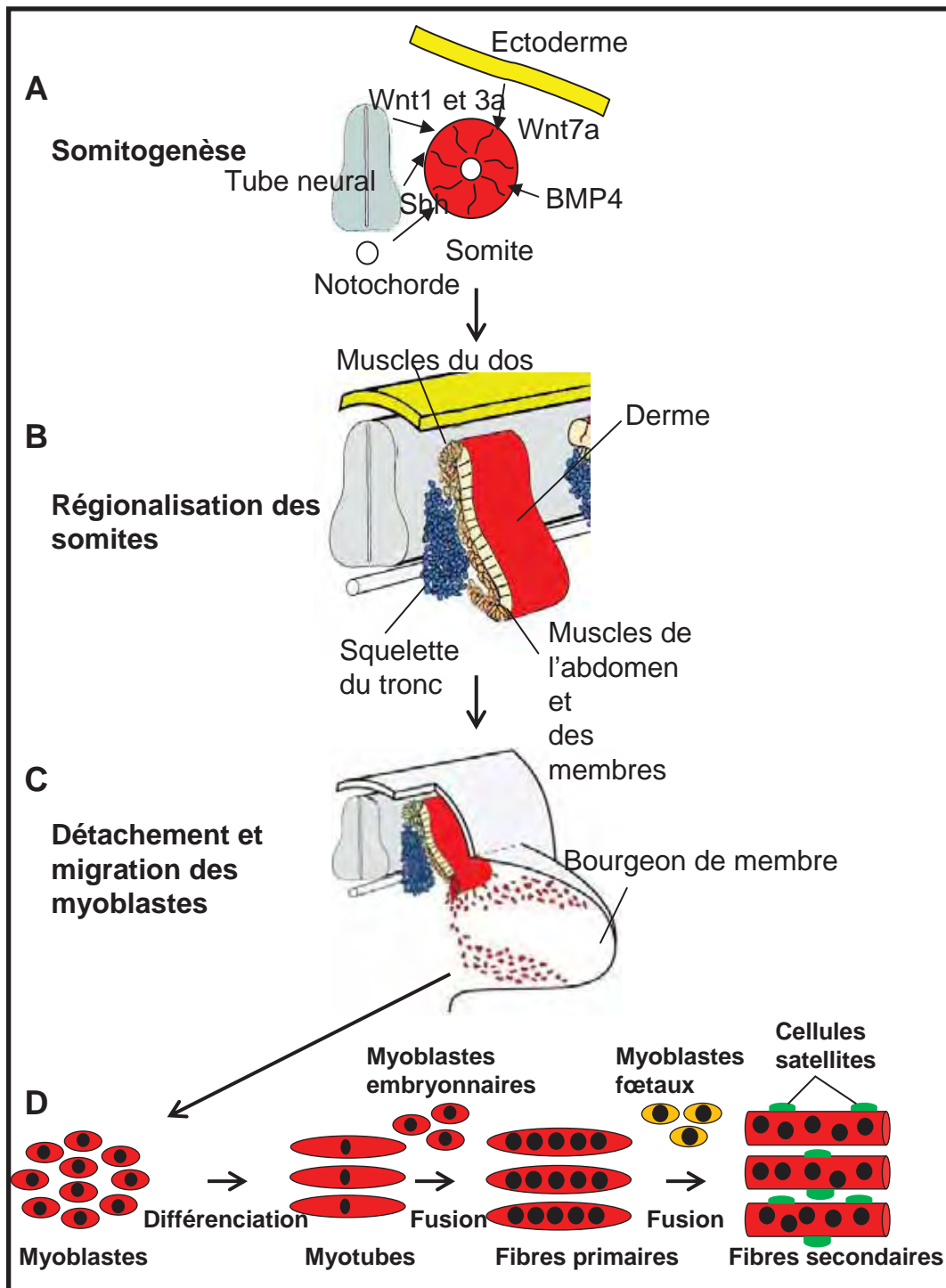


Figure 2. Représentation schématique de la myogenèse squelettique des muscles des membres chez les vertébrés. Plusieurs étapes sont requise la formation des muscles à partir des somites.

l'origine des tendons des muscles du tronc, dérive du sclérotome (Brent et al., 2003). Mis à part les muscles de la langue et du larynx qui dérivent aussi du dermomyotome hypaxial, les muscles de la tête dérivent du mésoderme non segmenté situé en position rostrale du territoire somitique (Noden, 1983; Trainor and Tam, 1995). De même, le squelette de la tête est issu du mésoderme non segmenté ainsi que des crêtes neurales. Nous reviendrons en plus de détail sur l'origine des muscles de la tête dans la partie discussion.

Des études génétiques principalement effectuées sur les modèles souris et poulet ont montré que les signaux Wnt, BMP (Bone Morphogenetic Protein) et Shh (Sonic hedgehog) sécrétés par les tissus adjacents aux somites induisent la régionalisation des somites et la formation du dermomyotome et du sclérotome (Figure 2). La différenciation du dermomyotome hypaxial est contrôlée par de signaux BMP et Wnt7a qui proviennent, respectivement, du mésoderme paraxial et de l'ectoderme (Figure 2 A-B). L'action combinée de Wnt1 ou Wnt3, et de Shh en provenance du tube neural et de la notochorde permet la différenciation du dermomyotome épaxial (partie dorsale du dermomyotome) qui est à l'origine des muscles du dos (Figure 2 A-B). Ainsi, les signaux Wnt1 et 3, Shh, BMP sont au cœur du processus de spécification des myoblastes (Bryson-Richardson and Currie, 2008). Finalement, en émettant le signal Shh, la notochorde induit la différenciation du sclérotome (Fan and Tessier-Lavigne, 1994).

3. Aspects moléculaires et tissulaires de la myogenèse

On distingue deux grandes étapes de la myogenèse, la myogenèse embryonnaire, dite primaire et la myogenèse fœtale, secondaire (Zhang and McLennan, 1998; Wigmore and Evans, 2002). Après formation des somites, les myoblastes à l'origine des fibres primaires se détachent du dermomyotome et s'étendent pour former des fibres mononuclées ou myotubes, s'étendant sur toute la longueur du somite (Figure 2 C). D'autres myoblastes embryonnaires fusionnent ensuite avec les myotubes, pour former des fibres multinuclées appelées fibres primaires (Figure 2 D) (Kelly and Zacks, 1969). Les fibres primaires forment les muscles lents, ou fibres de l'endurance plus efficaces en métabolisme aérobie. La seconde vague de myogenèse est caractérisée par la fusion de myoblastes fœtaux entre eux où avec les fibres primaires pour former les fibres secondaires à l'origine des muscles rapides (Figure 2 D) (Duxson et al., 1989). Les muscles rapides sont les muscles de l'effort intense en conditions d'anaérobie. Les myoblastes qui forment les muscles des membres, migrent vers les territoires précurseurs des membres (bourgeons). Cette migration requiert l'activité des FT à homéo-domaine Lbx1 (Lady bird-like) et Msx1 (Muscle segment homeobox) (Jagla et al., 1995; Mennerich et al., 1998; Houzelstein et al., 1999) et du

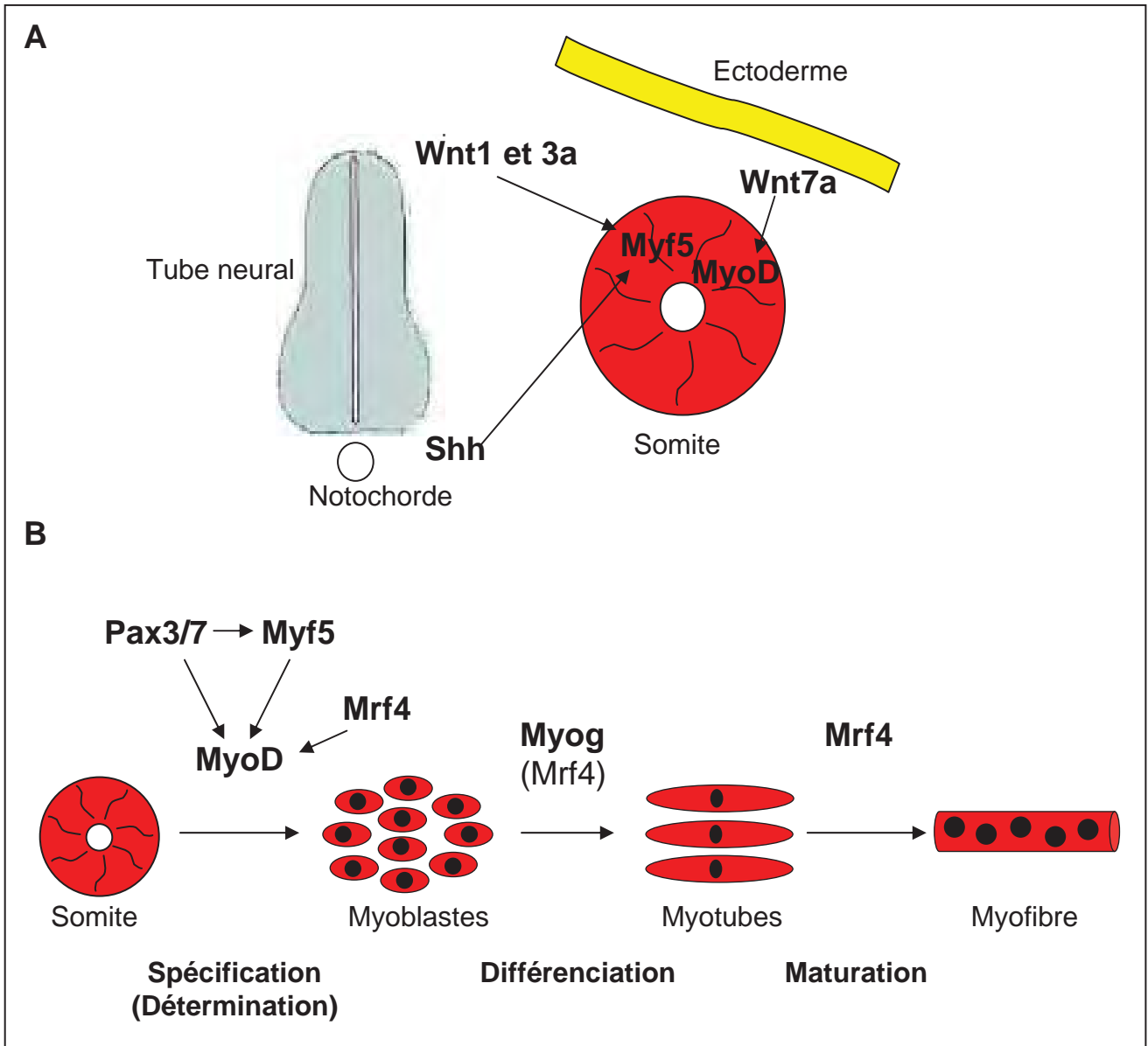


Figure 3. Induction et relation hiérarchique des MRF au cours de la myogenèse. **A.** Les signaux nécessaires à l'induction des facteurs myogéniques sont Shh, issu de la notocorde et le tube neural ventral, Wnt1 et 3a issus du tube neural dorsal, et Wnt7a issu de l'ectoderme. **B.** Le cœur transcriptionnel des différentes étapes de la myogenèse chez la souris.

récepteur tyrosine kinase C-met (Dietrich et al., 1999). Il est important de noter qu'à la fin de l'embryogénèse, des cellules appelées cellules satellites ou cellules souches du muscle squelettique apparaissent en périphérie des fibres musculaires (Figure 2 D) (Mauro, 1961). Ces cellules sont dans un état quiescent et peuvent être activées afin de contribuer à la croissance post-natale où bien à la régénération des muscles en cas de dommage (Collins et al., 2005; Montarras et al., 2005). La myogenèse qui recouvre les différentes étapes de différenciation des myoblastes en fibres musculaires fonctionnelles est orchestrée par un réseau de facteurs de transcription (FT) myogéniques appelés MRF.

4. MRF (Myogenic Regulatory Factor Family) : principaux acteurs de la différenciation des myoblastes.

L'identification du facteur de transcription (FT) MyoD fut une découverte capitale dans la compréhension des mécanismes moléculaires qui gouvernent la myogenèse (Davis et al., 1987; Weintraub et al., 1991). MyoD appartient à la famille des facteurs de transcription bHLH (basic Helix-Loop-Helix) et a la capacité de convertir différents types cellulaires (ex Fibroblastes, Neurones) en myoblastes. Peu de temps après cette découverte, trois autres FT appartenant à la même famille bHLH et ayant les mêmes propriétés myogéniques que MyoD ont été caractérisés. Ces FT sont Myf5, Myogenin (Myog) et MRF4 (Braun et al., 1989; Edmondson and Olson, 1989; Miner and Wold, 1990). Ces 4 protéines bHLH sont exprimées dans les muscles squelettiques et regroupées sous le nom des MRF. Des expériences de cultures de mésoderme pré-somitique ont montré que l'expression précoce des MRF est induite par des signaux émis par les tissus voisins (Munsterberg and Lassar, 1995). Ces signaux sont Wnt, émis par l'ectoderme dorsal et la partie dorsale du tube neurale (Tajbakhsh et al., 1998), et Shh qui provient de la plaque du plancher (partie ventrale du tube neural) et la notochorde (Figure 3 A) (Borycki et al., 1998; Borycki et al., 1999).

De nombreuses études ont porté sur les réseaux de régulation impliquant les MRF et leur dynamique d'expression. Ces études, principalement basées sur des analyses de perte de fonction ont permis de décrypter le rôle de chaque MRF au cours de la myogenèse. Il a été montré que les facteurs Myf5, Mrf4 coopèrent avec Pax3 pour activer MyoD au cours des étapes précoces lors de la spécification des myoblastes (Tajbakhsh et al., 1997) (Figure 3 B). Durant les étapes tardives, Myog est requis pour la différenciation des myoblastes en myotubes et Mrf4 est impliqué dans la maturation des myofibres (Venuti et al., 1995; Kassam-Duchossoy et al., 2004) (Figure 3 B). Via leur domaine HLH, les MRF forment des hétéro-dimères avec d'autres FT à domaine bHLH qui sont exprimés de manière ubiquitaire, les E-protéines (Lassar et al., 1991; Conway et al., 2004). Ces hétéro-dimères régulent des gènes cibles en se fixant sur des séquences d'ADN appelées E-Box (CANNTG). D'autres FT

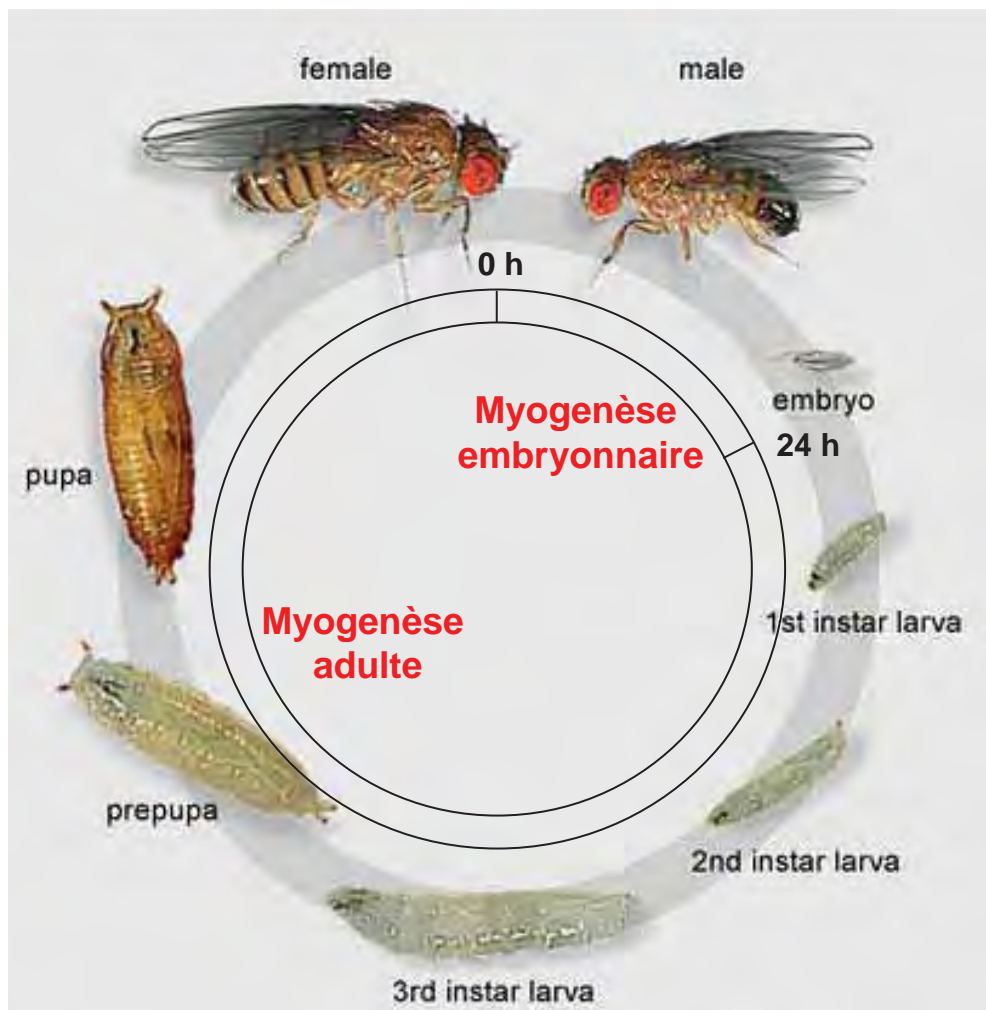


Figure 4. Deux vagues de myogenèse, embryonnaire, puis adulte, au cours du développement de la drosophile.

interviennent dans les phases précoces de la myogenèse des vertébrés, notamment les FT à domaine Paired (Pax3 et Pax7), qui sont requis pour la fusion des myotubes, et des FT de la famille Six (Six1 et 4) dont le rôle est essentiel pour la formation des fibres rapides (Richard et al., 2011 ; Bryson-Richardson and Currie, 2008). Chez la souris, plusieurs MRF ont des activités au moins partiellement redondantes. Par exemple, l'inactivation de Myf5 ou de MyoD n'entraîne pas de défauts majeurs de la musculature. Par contre, la double mutation entraîne la perte totale des muscles squelettiques (Berkes and Tapscott, 2005). Mef2 (myogenic enhancer factor 2) est une autre classe de FT qui est au coeur de la myogenèse et qui potentialise l'action des MRFs. Mef2 appartient à la famille des MADS (MCM1, Agamous, Deficiens, Serum response factor) et regroupe quatre protéines (Mef2 a-d) qui se lient à l'ADN en reconnaissant une séquence du type C/TTA(A/T)⁴TAG/A (Black and Olson, 1998). Contrairement aux MRF, Mef2 n'est pas capable d'induire, seul, la myogenèse. Un des rôles de Mef2 est de former des hétéro-dimères avec les MRF, via une interaction directe entre les domaines MADS et HLH, afin d'augmenter la spécificité de régulation des gènes cibles des MRF (Molkentin et al., 1995). En résumé, les cellules somitiques prolifèrent et deviennent des myoblastes sous le contrôle de Myf5, MyoD, MRF4. Les myoblastes se différencient ensuite en myocytes sous le contrôle de Myog, MyoD MRF4 et Mef2.

B. La drosophile : modèle d'étude de la myogenèse

La relative simplicité et la rapidité des modèles de cultures cellulaires ont permis d'identifier des acteurs majeurs impliqués dans la myogenèse, tels les MRF, et de caractériser leurs fonctions ex vivo. En parallèle, des études d'embryologie classique sur des embryons vertébrés, en particulier le poulet ont apporté une description cellulaire et tissulaire de la formation du mésoderme, la somitogenèse et l'origine des divers types de muscles. Les études initiées avec la découverte de MyoD in cellulo ont été replacées dans le contexte du développement embryonnaire par des études génétiques réalisées chez la souris. Ces études génétiques ont montré que les diverses étapes de la myogenèse sont contrôlées par un réseau de régulations transcriptionnelles, et que la chronologie d'expression des MRF est décisive. Malgré les connaissances accumulées au cours des années sur le rôle des divers MRF, de Pax3/Pax7, et de Mef2 au cours de la de la myogenèse, le contrôle génétique et moléculaire de **la diversité morphologique des muscles squelettiques des vertébrés reste largement inconnu**. La drosophile est, depuis 20 ans, le modèle principal d'étude de ce contrôle. C'est le modèle que j'ai utilisé au cours de ma thèse. La drosophile se prête en effet très bien aux analyses génétiques et moléculaires, et désormais génomiques, du développement. La grande diversité d'outils de génétique classique et de génétique reverse

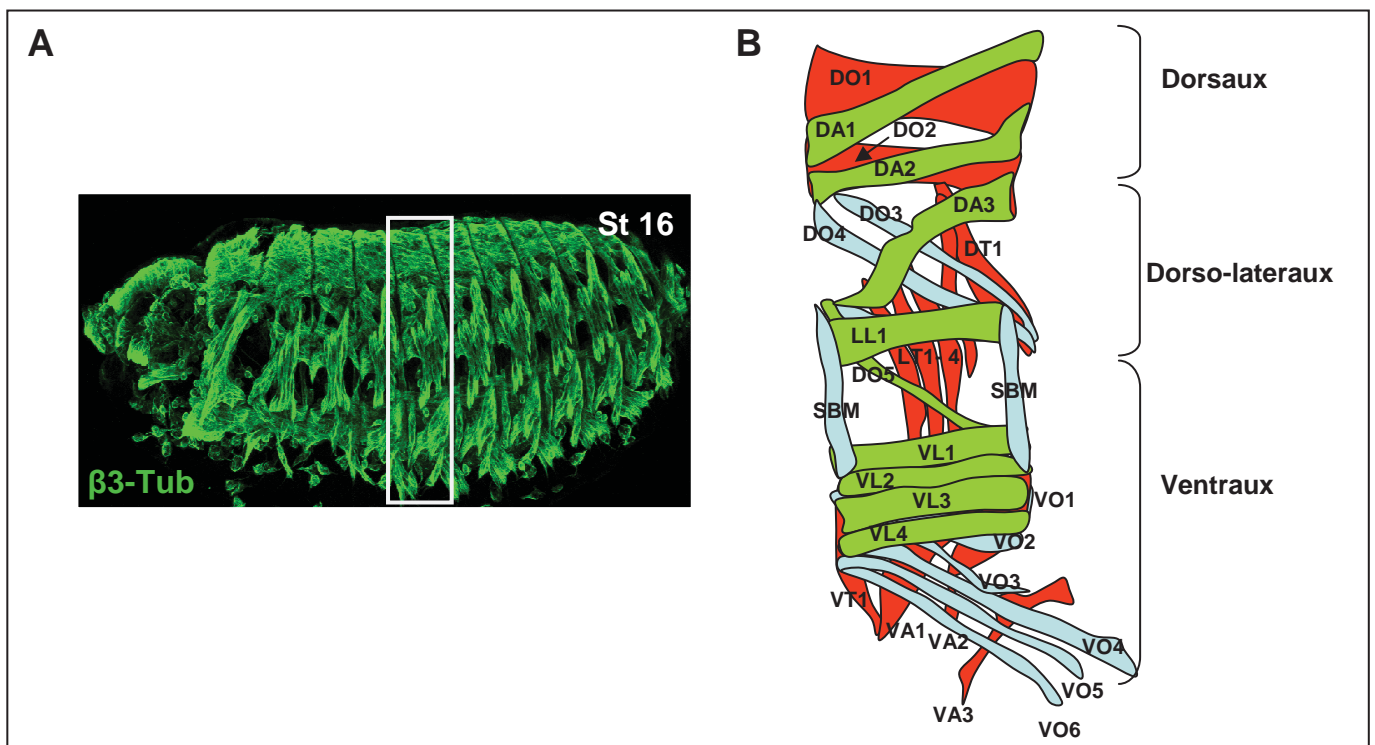


Figure 5. Musculature somatique de l'embryon de drosophile. **A.** Immunocoloration réalisée avec l'anticorps Anti β3-tubulin montrant le patron musculaire d'un embryon au stade 16. **B.** Représentation schématique du patron musculaire d'un segment abdominal. D : dorsal, L : latéral, V : ventral. O:oblique, A:aigu, L: Longitudinal, T transversal. SBM muscle à la frontière segmentaire

disponibles chez la drosophile ne cesse de s'accroître, avec comme exemples récents, les collections de petites délétions ciblées du génome, d'insertions du transposon MiMic ou de lignées rapporteur (<http://flybase.org/reports/FBtp0051711.html>, Venken et al., 2011). Par ailleurs, la conservation de structure et fonction des protéines entre la drosophile et les vertébrés permet de comparer directement les résultats obtenus dans ces deux organismes. Enfin, la moindre complexité du génome de la mouche limite le problème de redondance fonctionnelle entre gènes dupliqués et facilite les études d'éléments cis-régulateurs.

La drosophile est un insecte diptère holométabole. A 25°C, son cycle de développement dure environ 10 jours. Après un développement embryonnaire de 24h, les larves éclosent et sont capables de se mouvoir et se développer dans un milieu de culture. Après trois stades de développement larvaire, les larves forment des pupes, site de la métamorphose en adultes. Pendant la pupaison, la majorité des tissus larvaires sont histolysés et remplacés par des tissus adultes. C'est le cas de la majorité des muscles squelettiques. La myogenèse chez la drosophile comprend donc deux vagues. Une première vague embryonnaire qui met en place la musculature somatique de la larve. Cette vague aboutit également à la spécification des cellules précurseurs des muscles adultes (AMP : Adult Muscle Precursor), des muscles cardiaques, et des muscles alaires qui connectent les cellules cardiaque à l'épiderme. La deuxième vague se produit lors de la métamorphose. La majorité des muscles adultes, dont les muscles des pattes et les muscles thoraciques ventro-latéraux « du vol » sont issus d'AMPs associés aux disques imaginaux. Les muscles dorso-longitudinaux du vol sont eux, issus du remodelage de muscles larvaires (Fernandes et al., 1991) (Figure 4). Puisque mes travaux de thèse ont essentiellement porté sur la myogenèse embryonnaire, le reste de l'introduction sera principalement consacré à nos connaissances de ce processus.

1. Le mésoderme somatique chez la drosophile

La musculature squelettique (somatique) de la larve de drosophile est composée d'environ trente muscles différents par hémisegment (Figure 5). La formation de la musculature intègre un programme « générique » de différenciation musculaire et un programme « identitaire » qui se traduit par une forme, taille, orientation, site d'attachement propre à chaque muscle. L'orientation relative et la position de chaque muscle est indiquée dans une nomenclature couramment utilisée et que nous avons adoptée dans notre équipe (ex DO pour dorsal oblique, VL pour ventro-latéral etc. Figure 5 B) Le patron musculaire et ses variations entre segments le long du corps, sont très stéréotypées, indiquant un contrôle génétique très strict.

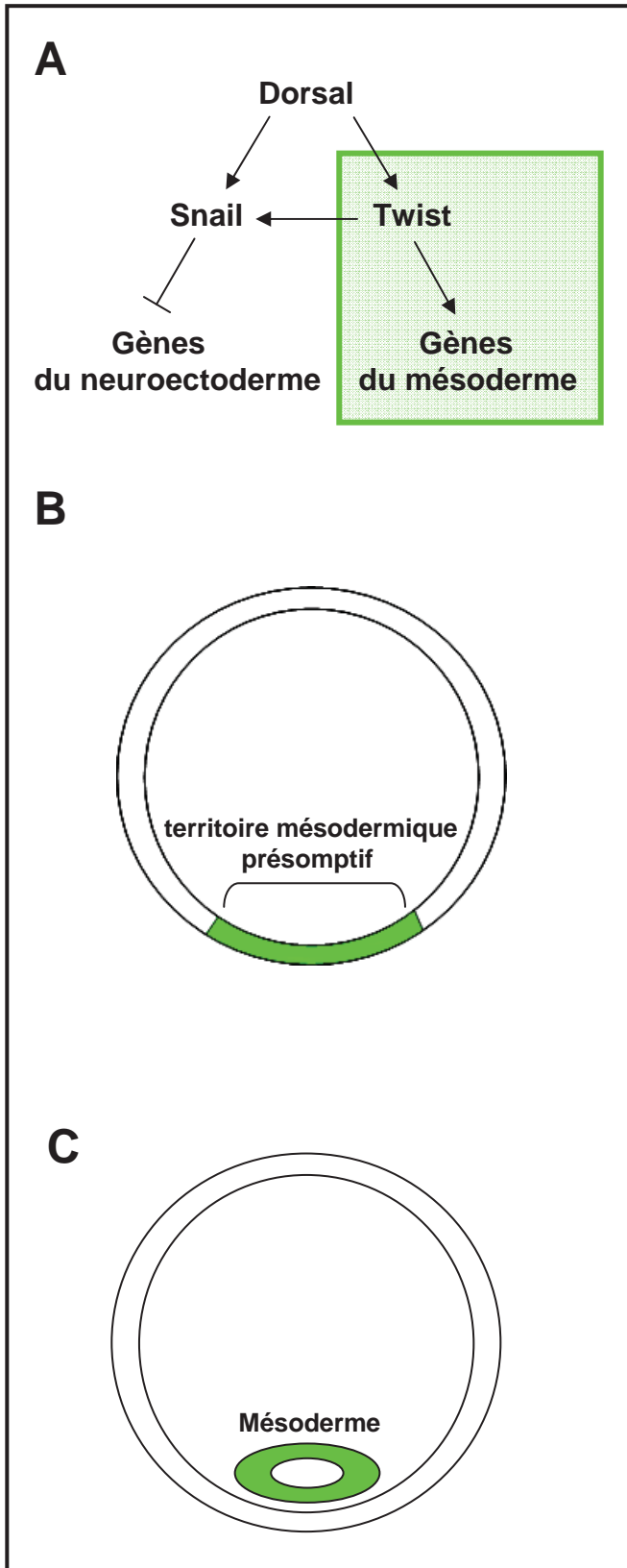


Figure 6. Formation du mésoderme chez la drosophile. **A.** Schéma transcriptionnel. **B.** spécification. **C.** invagination.

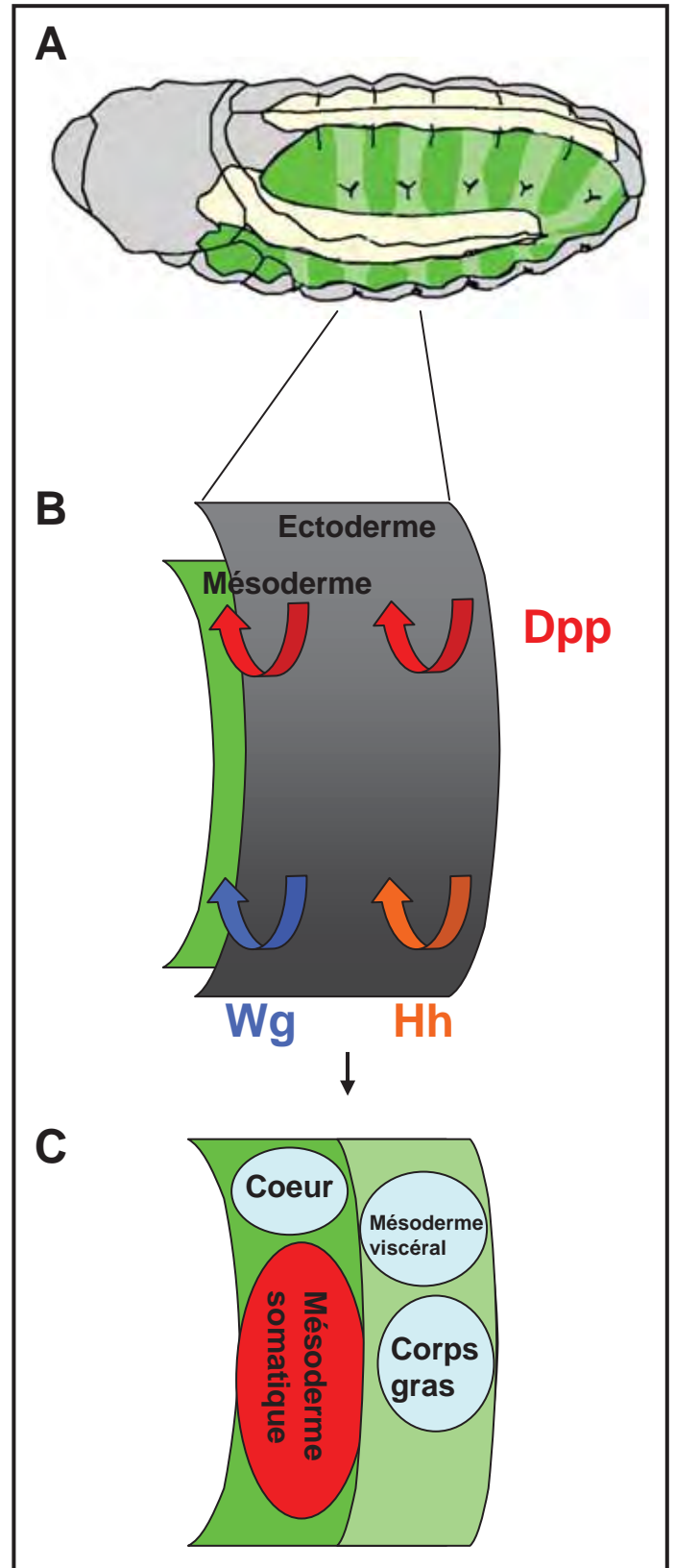


Figure 7. signaux contrôlant la subdivision du mésoderme. Vert foncé : haut niveau de Twi , vert pâle: bas niveau de Twi. **B, C** Schéma d'un segment.

a. Le mésoderme somatique à l'origine des muscles

Les muscles squelettiques de la larve sont issus du mésoderme somatique. Le mésoderme est spécifié dans la région ventrale de l'embryon de drosophile au stade blastoderme. La spécification du mésoderme est initiée par l'activité de deux FT exprimés ventralement, sous le contrôle du morphogène maternel Dorsal (NF- κ B chez les vertébrés) : le FT à domaine b-HLH Twist (Twi) et le FT à doigt de zinc Snail (Sna) (Riechmann et al., 1998) (Figure 6). Outre ses fonctions précoces dans le contrôle de l'invagination du mésoderme pour former un feuillet interne (Sweeton et al., 1991), Twi est requis pour l'activation de gènes impliqués dans la spécification des différents territoires mésodermiques (Thisse et al., 1987; Leptin, 1991). Sna réprime l'expression des gènes neurogéniques dans le mésoderme et le neuroectoderme ventral (Thisse et al., 1988; St Johnston and Nusslein-Volhard, 1992) (Alberga et al., 1991; Kosman et al., 1991) (Figure 6). Le mésoderme est à l'origine des muscles somatiques (squelettiques), viscéraux, cardiaques et du corps gras de la drosophile (Bate, 1993; Reuter and Leptin, 1994). La spécification de ces différents territoires reflète une régionalisation intrasegmentaire, selon les axes antero-postérieur (A/P) et dorso-ventral (D/V), similaire à la régionalisation des somites. Cette régionalisation est sous le contrôle de divers signaux en provenance de l'ectoderme, les mêmes que ceux impliqués dans la régionalisation des somites des vertébrés : Dpp (famille des TGF-B : apparentée aux BMP) exprimé dans l'ectoderme dorsal ; Wingless (Wg, homologue des Wnt) et Hedgehog (Hh, homologue de Shh) (Lawrence et al., 1995; Wu et al., 1995; Azpiazu et al., 1996) exprimés, respectivement, dans les régions antérieure et postérieure de chaque segment (Figure 7).

b. Régionalisation antero-postérieure du mésoderme.

Au sein de chaque segment (ou parasegment), le mésoderme est régionalisé selon l'axe A/P sous le contrôle des voies de signalisation Wg et Hh et des gènes de segmentation *even-skipped (eve)* et *sloppy paired (slp)* (Azpiazu et al., 1996; Riechmann et al., 1997) (Figure 8). Ceci se traduit par une évolution de l'expression de Twi, d'abord uniforme dans tout le mésoderme, en un domaine de haut niveau d'expression et un domaine de faible niveau d'expression (Borkowski et al., 1995; Baylies and Bate, 1996). L'activation de Slp, un régulateur positif de Twi, et Wg dans la partie antérieure du segment, crée un domaine à forte expression de Twi. A l'inverse, Eve régule Hh dans la partie postérieure du segment résultant en un domaine de faible expression de Twi (Bate and Rushton, 1993; Baylies et al., 1995). Le mésoderme somatique et le mésoderme cardiaque sont issus du domaine à forte expression de Twi, et le mésoderme viscéral et le corps gras sont issus du domaine où Twi est exprimé à faible dose. La voie de signalisation Notch contribue aussi à la régulation de

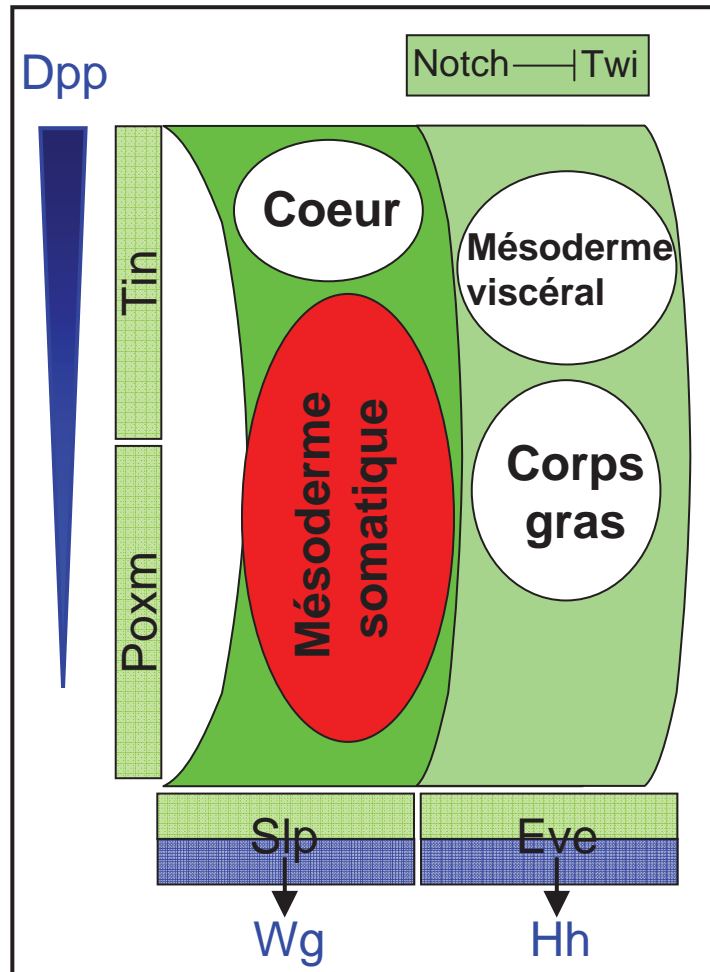


Figure 8. Facteurs de transcription mésodermiques agissant en amont (Slp, Eve) et en aval (Poxm, Tin) des signaux ectodermiques. Eve et Slp agissent aussi directement dans le mésoderme.

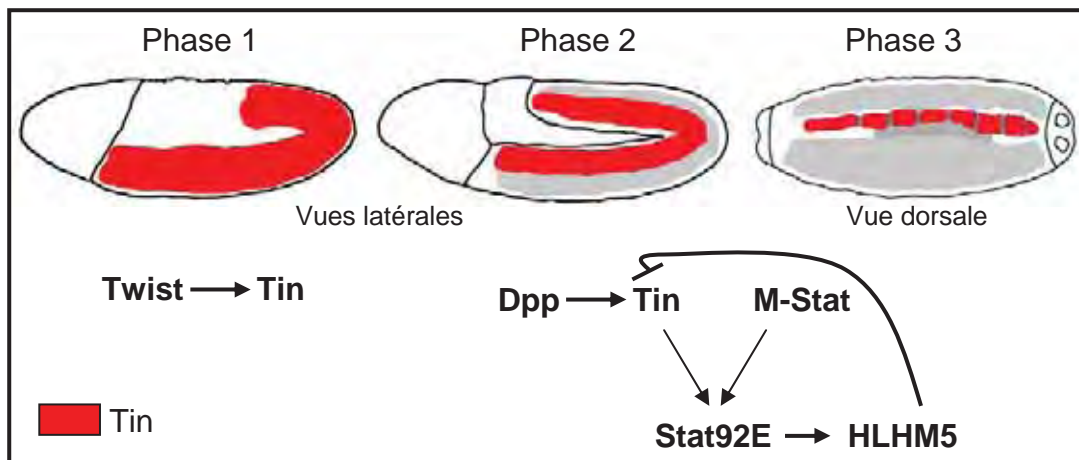


Figure 9. Étapes d'expression de Tinman (Tin) au cours de l'embryogenèse.
Adapté de Zaffran et al., 2006

niveau d'expression de *Twi* dans le mésoderme en établissant des domaines d'expression à faible ou à forte dose de *Twi*. Cette régulation se fait via l'effecteur transcriptionnel de la voie Notch, Suppressor of Hairless *Su(H)* (Tapanes-Castillo and Baylies, 2004) (Figure 8).

c. Régionalisation dorso-ventrale du mésoderme.

En 1994 M. Bate et ses collaborateurs ont montré que la voie *Dpp* est requise pour l'activation des gènes du mésoderme (Staehling-Hampton et al., 1994). Un an après cette découverte, M. Frasch a identifié le premier gène cible de *Dpp* dans le mésoderme, le FT à homéodomaine *Tinman* (*Tin*, homologue de *NKx2.5* chez les vertébrés) (Frasch, 1995). L'expression de *Tin* est d'abord détectée dans tout le mésoderme avant d'être restreinte au mésoderme dorsal où son activité est nécessaire pour spécifier les cellules cardiaques, le mésoderme viscéral ainsi que des muscles dorsaux (Azpiazu and Frasch, 1993). L'étude de la régulation transcriptionnelle de *Tin* a montré que plusieurs activateurs et répresseurs modulent son expression spatio-temporelle (Xu et al., 1998). L'expression de *Tin* peut être schématisée en quatre étapes successives. *Tin* est initialement activé par *Twi* dans tout le mésoderme (phase 1). La phase 2 correspond au maintien de cette expression dans le mésoderme dorsal, sous l'action de *Dpp*. La phase 3 est caractérisée par le maintien de *Tin* uniquement dans les progéniteurs cardiaques et viscéraux. La phase 4 correspond à la restriction de l'expression de *Tin* aux cellules cardiaques (Yin et al., 1997; Xu et al., 1998; Venkatesh et al., 2000; Zaffran et al., 2006) (Figure 9). Très récemment, il a été établi que la voie de signalisation *JAK/Stat* assure la transition entre les phases 2 et 3, en réprimant indirectement *Tin* via l'activation du FT *HLHm5* (un des 7 membres du complexe « *enhancer of split* » : *E (spl)-C*). Cette régulation est initiée par l'activation de *Stat92E* (effecteur transcriptionnel de la voie *JAK/Stat*) sous l'effet conjugué de *Tin* et *M-Stat* (contribution maternelle de *Stat92E* : préalablement phosphorylé par l'activité des ligands *Upd* et *Upd2*) (Johnson et al., 2011) (Figure 9). *Tin* étant un acteur clef de la spécification du mésoderme dorsal, un enjeu majeur était d'identifier ses gènes cibles. Dans cette optique, l'équipe d'Eileen Furlong a mis en place une stratégie de *Chip-on-chip* permettant de dresser une carte des gènes régulés par *Tin in vivo*. Cette étude a permis de montrer que *Tin* régulait l'activité de *Stat* (Liu et al., 2009) et de proposer qu'il régulait les FT *islet1/tup* (*Tup*) et *Collier* (*Col*). Nous y reviendrons.

La spécification du mésoderme ventral est aussi, indirectement, contrôlée par l'activité de *Dpp*. En effet, *Dpp* réprime dorsalement *Pox meso* (*Poxm* : un gène de la famille *Pax*) et restreint son expression au mésoderme ventral (Staehling-Hampton et al., 1994) (Figure 8). Le rôle de *Poxm* a été étudié en détail au cours de la myogenèse. Il a été montré que *Poxm*

est requis pour la spécification des muscles ventraux et quelques muscles latéraux (Duan et al., 2007). En conclusion, ces données permettent de schématiser la régionalisation D/V du mésoderme en partie dorsale contrôlée par Tin et une partie ventrale sous le contrôle de Pox meso. Afin de compléter ce schéma, il serait nécessaire de replacer le rôle de Collier (Col) dans la spécification du mésoderme dorso-latéral (Enriquez et al., 2012). Ce rôle sera abordé dans le chapitre discussion des résultats.

En résumé, le territoire à l'origine des muscles squelettiques de la drosophile correspond au territoire soumis à l'action de Wg et l'action transitoire de Dpp. La distinction entre mésoderme cardiaque et le mésoderme somatique dorsal dépend du niveau et de la durée d'expression de Tin.

d. Régulation transcriptionnelle de la myogenèse chez la drosophile.

Comme nous l'avons vu précédemment, les MRF sont des acteurs clef de la myogenèse des vertébrés. Ils ont été largement étudiés chez les vertébrés. Cependant, l'étude des réseaux transcriptionnels impliquant les MRF a été confrontée à deux obstacles. Le premier concerne les expériences de perte de fonction qui ont permis d'accéder aux fonctions précoces des MRF mais pas à leurs fonctions tardives à cause de la létalité embryonnaire. Le deuxième obstacle concerne la redondance fonctionnelle entre différents MRF. La perte de fonction d'un des MRF peut être partiellement compensée par la présence d'un autre facteur. Quelle est la situation chez la drosophile ?

Twi est le gène de drosophile dont la fonction est la plus proche des MRF vertébrés. Comme les MRF (considérés ici collectivement), Twi a un rôle précoce essentiel au cours de la formation des myoblastes. Des analyses de perte de fonction utilisant des mutants thermosensibles ont montré que Twi a également un rôle plus tardif au cours de la spécification des muscles et des précurseurs musculaires adultes. De même, comme les MRF, Twi a la capacité d'induire le programme myogénique ; un gain de fonction de Twi dans l'épiderme induit la formation de myoblastes ectopiques exprimant des marqueurs musculaires (Baylies and Bate, 1996). Twi agit soit sous forme d'homodimère et active des gènes de la myogenèse, soit sous forme d'hétérodimère avec le FT Daughterless (Da), l'équivalent des E-protéines vertébrés et réprime ces mêmes gènes. L'activité répressive est assurée par le domaine REP de la protéine Da et est contexte-dépendante. Elle intervient dans les domaines à faible expression de Twi et participe à la spécification de destins alternatifs, le corps gras et le mésoderme viscéral (Castanon et al., 2001; Wong et al., 2008). Plus récemment, il a été montré que la spécificité d'action de Twi faisait aussi appel à des mécanismes de remodelage de la chromatine. La liaison entre Twi et le complexe de remodelage de la chromatine BRM (Brahma containing) est assurée par l'Akirin une protéine

nucléaire interagissant directement avec Twi. Cette interaction permet de moduler de manière sélective l'expression des gènes cibles de Twi (Nowak et al., 2012).

Paradoxalement, le gène *nautilus* (Nau : FT à domaine bHLH) qui est le gène de drosophile le plus apparenté aux MRF (Approximativement 90 % de similitude de séquence avec MyoD (Michelson et al., 1990; Paterson et al., 1991), ne semble pas avoir de fonction équivalente aux MRF. Nous reviendrons sur Nau plus tard dans l'introduction.

Mef2 joue également un rôle clé dans la différenciation des muscles squelettiques larvaires chez la drosophile (Taylor, 1995; Black and Olson, 1998). Mef2 est initialement exprimé dans l'ensemble du mésoderme, puis son expression est restreinte au mésoderme somatique et aux précurseurs des cellules cardiaques (Lilly et al., 1994; Nguyen et al., 1994). Comme pour Twi, et contrairement à son homologue chez les vertébrés, l'expression ectopique de Mef2 dans l'épiderme enclenche le programme myogénique (Lin et al., 1997). Inversement, dans un contexte mutant pour Mef2 les muscles somatiques larvaires, viscéraux, cardiaques, alaires et pharyngaux ne se forment pas (Bour et al., 1995). Les muscles adultes issus du remodelage des muscles larvaires sont aussi malformés (Soler et al., 2012). En absence de Mef2, les myoblastes et précurseurs musculaires se forment mais ne sont pas capables de se différencier en fibres fonctionnelles (Ranganayakulu et al., 1995). Ces données indiquent que le rôle de Mef2 durant la différenciation des muscles somatiques chez la drosophile se rapproche de celui de Myog chez les organismes vertébrés.

Twi et Mef2 sont donc les deux FT myogéniques de la drosophile dont la fonction semble être la plus proche des MRF et de Mef2 vertébrés. Une analyse du transcriptome d'embryons exprimant différentes doses de Mef2 a montré que les gènes cibles de Mef2 sont régulés de manière dose dépendante (Elgar et al., 2008). Des expériences de Chip-chip recherchant des cibles directes de Mef2 ont par ailleurs montré que Mef2 régule essentiellement des gènes de différenciation musculaire (Sandmann et al., 2006). La recherche des gènes cibles directes de Twi a également été menée par des expériences de Chip-chip à différentes périodes du développement embryonnaire (période allant de la spécification des myoblastes à la période de fusion des myoblastes). Ces analyses ont montré que Twi se fixe sur de nombreux gènes impliqués dans la spécification du mésoderme comme Tinman (Tin) et Mef2 mais aussi des FT impliqués dans des programmes de différenciation muscle-spécifiques (Sandmann et al., 2007) voir section résultats. L'activation de Mef2 au cours de la différenciation des myoblastes chez les vertébrés est sous le contrôle des MRF. Cette régulation semble être conservée au cours de l'évolution puisque chez la drosophile, Mef2 est aussi activé par Twi (Sandmann et al.,

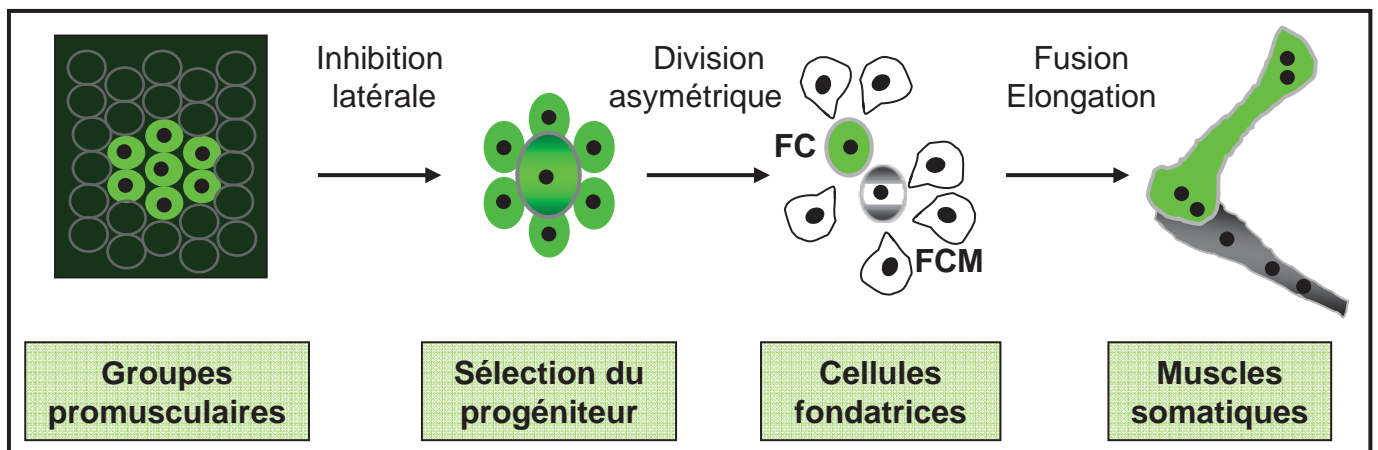


Figure 10. Principales étapes de formation des muscles squelettiques chez la drosophile
Adapté de Baylies et al., 1998

2007). Pour conclure, il est important de retenir que des FT clef et réseaux de régulation de la myogenèse semblent bien conservés entre la drosophile et les vertébrés. Ceci suggère que l'étude du contrôle de la diversité morphologique des muscles chez la drosophile apporte des connaissances fondamentales à notre compréhension générale de la formation de la musculature.

2. Etapes de formation des muscles chez la drosophile

Suite à la spécification du mésoderme somatique, on peut schématiquement distinguer quatre étapes successives dans la formation des muscles squelettiques (Figure 10). La première étape est la définition, au sein du mésoderme, de groupes (clusters) de cellules équivalentes, appelés groupes promusculaires. La deuxième étape est la sélection d'une cellule progéniteur (PC) à partir de chaque groupe promusculaire via le processus d'inhibition latérale. Chaque cellule progéniteur subit ensuite une division asymétrique pour générer soit deux cellules dites fondatrices (FC), soit une FC et une cellule d'un lignage différent (AMP où progéniteur de cellules cardiaques). Les cellules non sélectionnées deviennent des myoblastes « naïfs » (FCM : Fusion Competent Myoblast), qui ont la compétence de fusionner avec les FC. La quatrième étape est la fusion de chaque FC avec un nombre défini de FCM pour former une fibre multi-nucléée de taille, forme et orientation spécifiques (Bate and Rushton, 1993; Carmena et al., 1995; Ruiz Gomez and Bate, 1997). Contrairement aux vertébrés où chaque muscle strié est formé de multiples fibres, chaque muscle de la larve de drosophile est composé d'une seule fibre multi-nucléée. La simplicité de cette musculature permet de suivre toutes les étapes de la myogenèse, de la spécification des myoblastes à l'attachement du muscle à l'épiderme.

a. Spécification des groupes promusculaires et sélection des progéniteurs

Les cellules progéniteurs (PC) sont sélectionnées à partir de groupes de cellules équivalentes, les « groupes promusculaires ». 18 groupes promusculaires (GP) ont été cartographiés sur la base de l'expression du gène bHLH proneural Lethal of Scute (L'sc), entre 5h et 7h de développement (stades 10 et 11). Ils apparaissent, à position fixe et de façon séquentielle au sein du mésoderme somatique. Le positionnement de ces GP est soumis à une information de position issue de l'ectoderme (Carmena et al., 1995) (Figure 11). Plusieurs voies de signalisation sont impliquées, comme l'a montré l'étude particulièrement détaillée des GP 2 et 15 (Figure 12) qui sont, respectivement, à l'origine des muscles dorsaux DO2 et DA1 (Carmena et al., 1998a) et qui est brièvement rappelée ci-dessous. La première étape mise en évidence est la définition d'un pré-domaine de

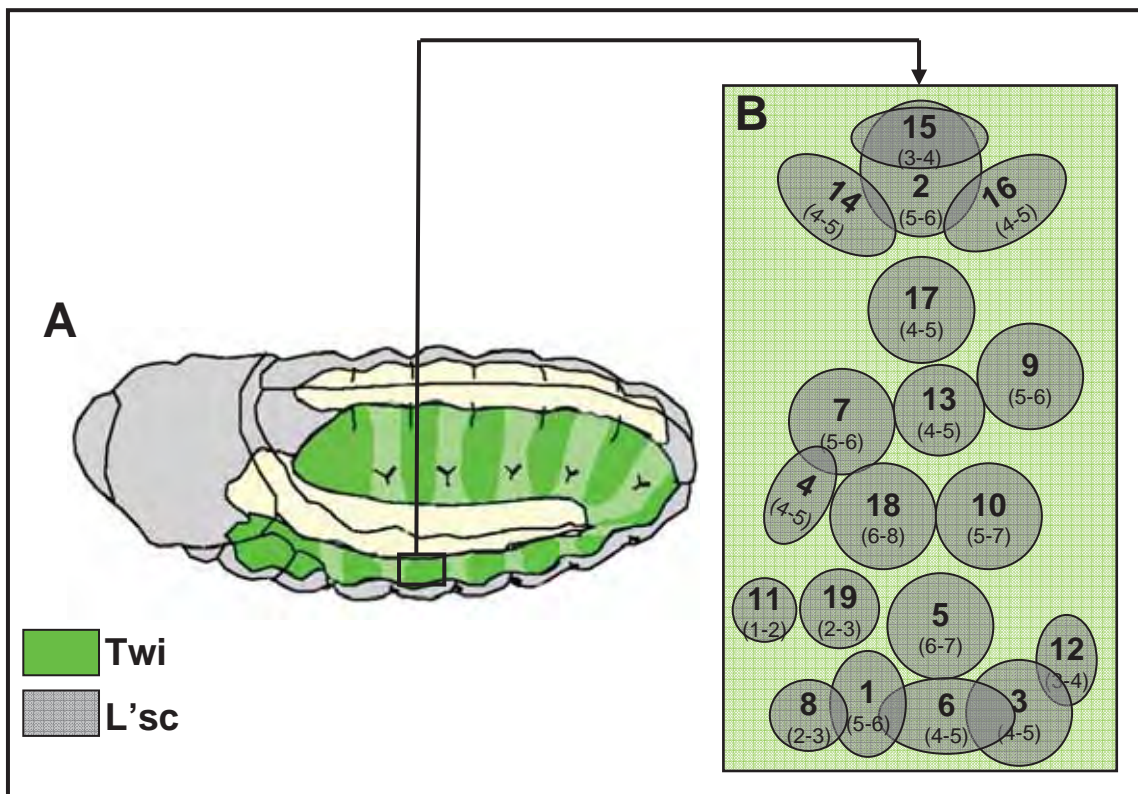


Figure 11. Localisation des groupes promusculaire (GP) exprimant L'sc au sein d'un segment abdominal. Le nombre de cellule formant chaque GP est indiqué. (*Adapté de Carmena et al., 1995*)

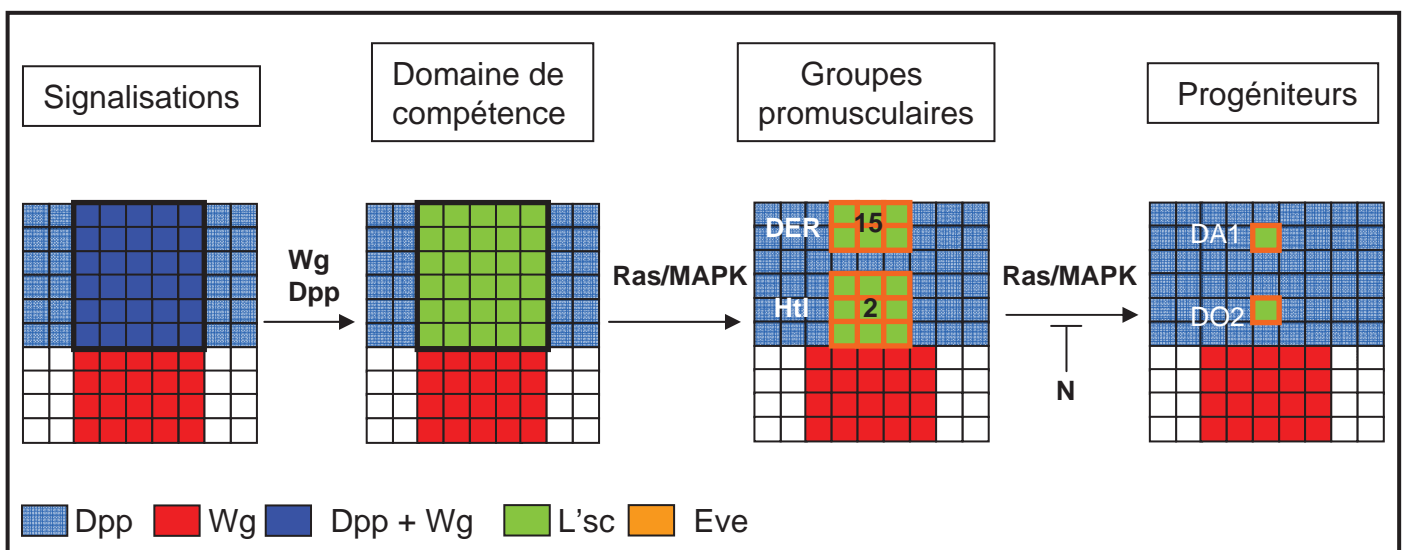


Figure 12. Schéma d'intégration des signaux extrinsèques au cours des étapes de spécification des groupes promusculaires et des progéniteurs (*d'après Carmena et al., 1998*)

compétence (pré-cluster) correspondant à l'intersection des voies de signalisations antéro-postérieure Wg et dorso-ventrale Dpp. Ce pré-cluster exprime L'sc et est capable de répondre à la signalisation Ras/MAPK (Mitogen Activated Protein Kinase). L'activation de la voie Ras/MAPK, via les récepteurs à tyrosine kinase (RTK) de type EGF (DER) et FGF (Heartless : Htl) exprimés dans les myoblastes restreint l'expression de L'sc aux GP 2 et 15 issus du pré-cluster. A ce stade, toutes les cellules des deux GP commencent à exprimer le FT Eve et les GP 2 et 15 ne se distinguent l'un de l'autre que par leur niveau relatif de signalisation Ras/MAPK. Un PC est sélectionné à partir de chacun de ces GP et maintient un haut niveau d'expression de L'sc. La sélection des PC, via le processus d'inhibition latérale Notch/Delta, entraîne la répression de la signalisation Ras/MAPK et l'activation de la voie N dans les cellules non sélectionnées (FCM) (Halfon et al., 2000) (Figure 12). Dans des mutants affectant la voie de signalisation Notch (N), toutes les cellules maintiennent un haut niveau de L'sc et gardent un destin « PC-like » (Carmena et al., 1995).

Le modèle de sélection des PC à partir de GP, similaire à la sélection des neuroblastes à partir de groupes d'équivalence dans le neuro-ectoderme, établi en 1995 à partir de l'étude de deux GP/PC exprimant Eve a été ensuite confirmé par l'étude de l'expression d'autres FTi (S59, Col) dans d'autres clusters spécifiques latéraux dorso-latéraux (Crozatier and Vincent, 1999; Knirr et al., 1999). Cependant, à ce jour, aucun autre marqueur général des GP que L'sc n'a été identifié. Paradoxalement, dans un mutant L'sc, très peu de muscles sont affectés, indiquant un rôle mineur de ce gène. En 2007, M.Noll et ses collaborateurs ont montré que le gène *poxm* avait une fonction redondante avec *l'sc* durant les étapes de spécification du domaine de compétence ventral (pré-cluster ventral). Cette découverte a permis d'étendre au domaine ventral le modèle de spécification des GP décrit préalablement pour les muscles dorsaux et latéraux et d'expliquer le phénotype mineur observé dans le mutant *l'sc* (Duan et al., 2007). Le modèle de positionnement des clusters et de sélection des PC proposé à partir de l'étude d'Eve laissait donc de nombreuses zones d'ombre, parmi lesquelles les mécanismes d'interprétation de position par les myoblastes et le processus de sélection/spécification de différents PC à partir d'un même pré-cluster ou cluster. Nous reviendrons sur ces questions.

b. Les cellules fondatrices

Au cours du processus de sélection, les PC délaminent à partir des groupes promusculaires et se positionnent en position latérale, sous l'ectoderme avant de se diviser de manière asymétrique. Chaque PC donne naissance à deux cellules fondatrices (FC) qui sont des cellules post-mitotiques. La notion de FC a été évoquée pour la première fois en 1983 par RK Ho lorsqu'il a identifié chez le criquet une classe distincte de myoblastes qui préfiguraient la localisation des futurs muscles et étaient capables de fusionner avec d'autres myoblastes afin de former une fibre. Il a appelé ces myoblastes « cellules pionnières » et a montré que dans des conditions d'ablation des cellules pionnières, les muscles ne se forment pas (Ho et al., 1983). M. Bate a ensuite montré chez la drosophile l'existence de 30 cellules pionnières réparties dans chaque segment le long de l'axe dorso-ventral et qu'il a appelés myoblastes fondateurs (FC). Ces myoblastes ont la particularité d'être plus volumineux que les autres myoblastes avec lesquels ils ont la capacité de fusionner pour former une fibre musculaire. Il a ainsi émis l'hypothèse qu'il y avait deux types de myoblastes, les FC équivalents aux cellules pionnières caractérisées chez le criquet, et les myoblastes compétents pour la fusion (FCM), et que les FC contiennent toute l'information nécessaire pour former une fibre musculaire (Bate, 1990). Des données de M. Frasch ont confirmé cette hypothèse en montrant que le FT à homéodomaine S59 (Slouch) est exprimé dans un sous ensemble de FC et requis pour la formation des muscles issus de ces FC. Le même groupe a montré qu'après fusion, les FCM expriment *de novo* S59 et adoptent donc la même identité que la FC avec laquelle ils ont fusionné (Dohrmann et al., 1990). En 1992, Bougouin C a renforcé l'idée que les FC contiennent l'information propre à chaque muscle avec l'analyse d'Apterous (Ap). Comme S59, Ap est exprimé dans un sous ensemble de FC et requis pour la formation des muscles issues de ces FC (Bougouin et al., 1992). Finalement, la caractérisation du mutant *myoblaste city* (*mbc* : mutation qui affecte le processus de fusion) a montré qu'en absence de fusion avec des FCM, les FC gardent la capacité à former une fibre, dans ce cas mononucléée, qui exprime des marqueurs de différenciation musculaire comme la MHC (Myosin Heavy Chain) et a les mêmes propriétés morphologiques, d'attachement et d'innervation que les fibres sauvages (Rushton et al., 1995).

Aujourd'hui, il est communément admis que les caractéristiques de taille, forme, orientation, innervation propres à chaque muscle reflètent l'expression, par chaque FC, d'une combinatoire spécifique de FT appelés FT identitaires (FTi). Plusieurs FTi ont été étudiés en détail : *Kruppel* (*Kr*), *even-skipped* (*eve*), *collier* (*col*), *ladybird* (*lb*) (Ruiz-Gomez et al., 1997; Carmena et al., 1998a; Jagla et al., 1998, Crozatier et Vincent, 1999). Une question demeure en suspens, qui est de savoir si les FCM issus d'un GP donné sont naïfs ou au contraire déjà

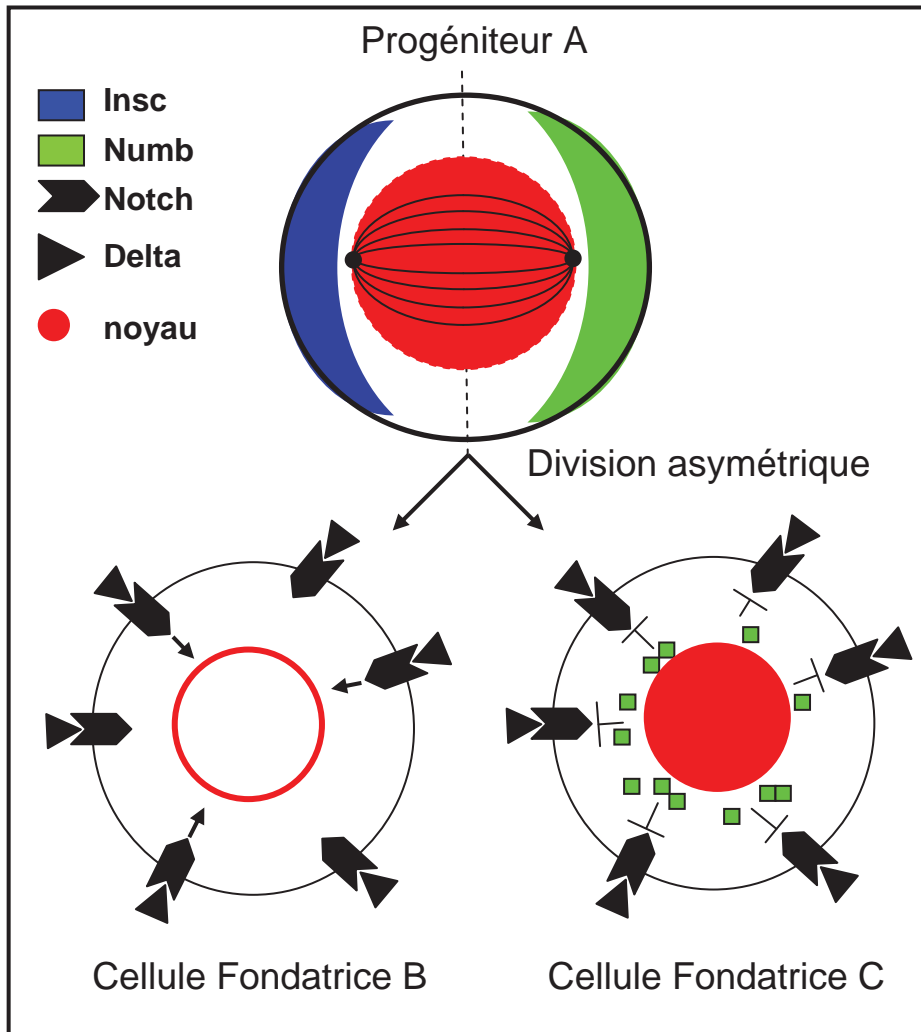


Figure 13. Représentation schématique du processus de division asymétrique d'un progéniteur au cours de la myogenèse chez la *Drosophile*.

restreints dans leur capacité à contribuer à la formation d'un sous ensemble de muscles. Nous reviendrons sur cette question dans la section résultats.

c. Division asymétrique des progéniteurs : Les divers lignages

La division des PC est une division asymétrique qui génère deux cellules qui diffèrent entre elles et de la cellule mère. Les mécanismes de cette division sont similaires à ceux mis en jeu dans les divisions asymétriques des neuroblastes qui génèrent un neuroblaste et une GMC (ganglion mother cell). La division asymétrique des neuroblastes repose sur la fonction de la protéine Inscuteable (Insc) (Carmena et al., 1998b; Ward and Skeath, 2000). Insc est une protéine adaptatrice localisée au cortex apical des neuroblastes en interphase et qui coordonne l'orientation du fuseau mitotique et la localisation de la protéine Numb au cortex basal, de telle sorte que Numb est sélectivement héritée par la GMC. L'asymétrie initiale est le contact du neuroblaste avec l'ectoderme. *insc* a été identifié indépendamment pour son rôle dans la myogenèse (mutant *not enough muscles (nem)*) (Knirr et al 1997). Dans ce cas, l'asymétrie initiale reste à identifier. Numb est une protéine à domaine PTB (phosphotyrosine-binding motif), un domaine d'interaction avec la protéine Notch (N) et régule négativement la signalisation N. La ségrégation asymétrique de Insc et Numb dans le PC au cours de la mitose crée une asymétrie d'activation de la signalisation N qui est active dans la FC ayant hérité Insc et inactive dans la FC ayant hérité Numb (Guo et al., 1996) (Figure 13). La voie N réprime la transcription de FTi préalablement activés dans le PC. Le résultat net est deux cellules jumelles qui expriment des sous-ensembles différents de FTi et possèdent donc des identités distinctes, à l'origine de muscles différents. Il faut noter que ces identités sont une variation de l'identité du PC. Il faut aussi noter que la division d'un PC peut donner soit deux FC d'identités différentes, soit une FC et un précurseur de cellules cardiaques ou péri-cardiaques, ou encore une FC et une cellule précurseur d'un muscle adulte (AMP : Adult Muscle Precursor) (Speicher et al., 2008; Figeac et al., 2010). Quatre groupes de FC peuvent être distingués en fonction de leur position selon l'axe D/V. Les FC dorsales, dorso-latérales, latérales, et ventrales. Ces FC apparaissent suivant un timing bien précis (Nose et al., 1998; Beckett and Baylies, 2007).

d. Fusion des myoblastes: Description cellulaire et moléculaire

Chacune des 30 fibres musculaires de l'embryon de drosophile est générée à partir de la fusion entre une FC et plusieurs FCM. Le nombre de FCM fusionnées est contrôlé par les FTi et les gènes homéotiques et varie en fonction de la taille des muscles (Bataille et al., 2010; Enriquez et al., 2010; de Jossineau et al., 2012). Ce nombre varie de 5 à 30. Le

processus de fusion lui-même est l'objet de très nombreuses études qui ne peuvent être que très brièvement résumées ici.

Avant de fusionner, les FCM doivent migrer et adhérer à la FC avec lesquelles ils vont fusionner. Les FCM émettent donc des filopodes qui leur permettent d'explorer leur environnement et d'identifier les FC. D'un point de vue moléculaire, la reconnaissance et l'adhésion entre FC et FCM est contrôlée par des protéines d'adhésion cellulaire appartenant à la superfamille des immunoglobulines (IgSF) et exprimées de manière dissymétrique entre FC et FCM. Les FC expriment les protéines d'adhésion Dumbfounded (Duf) et Roughest (Rst) et les FCM expriment Sticks-and-Stones (Sns), Hibris (Hbs) et Rst (Bour et al., 2000; Ruiz-Gomez et al., 2000; Dworak et al., 2001; Strunkelnberg et al., 2001).

L'adhésion entre FC et FCM via des interactions hétérotypiques entre les molécules d'adhésion cellulaire, Duf ou Rst et Sns, induit le remodelage de leur cytosquelette d'actine, initiée par la polymérisation d'actine-G (G pour globulaire) au niveau du contact entre les deux cellules, formant ainsi des foci membranaires d'actine-F (F pour filamenteuse). Ce remodelage du cytosquelette qui dépend de l'activité de la voie Rac et met en jeu de nombreuses protéines adaptatrices est indispensable au processus de fusion (Sens et al., 2010; Haralalka et al., 2011). L'enrichissement focal des réseaux d'actine-F est observé dans les deux cellules et précède la formation de pores de fusion (Haralalka et al., 2011). L'accumulation de foci d'actine-F dans les FCM requiert l'activité de la protéine FCM-spécifique Myoblast city (Mbc) et induit la formation de structures « protrusives » dont le rôle est de générer des tensions physiques qui permettent la pénétration des FCM dans les FC. A ce jour aucun lien moléculaire définitif n'a été établi entre le remodelage du cytosquelette et la formation des vésicules et des pores de fusion.

Toutes les protéines cytoplasmiques impliquées dans le remodelage du réseau d'actine qui précède la fusion sont conservées au cours de l'évolution. Une question est donc de savoir si le concept de fusion asymétrique FC/FCM utilisé chez la drosophile est généralisable à d'autres organismes, notamment les vertébrés. Des études menées dans ce but ont montré que Dock1 et Dock5 (les homologues de *mbc*) sont impliqués dans la fusion des myoblastes chez le poisson zèbre (Moore et al., 2007; Srinivas et al., 2007). D'autres études ont montré que Kirre, un récepteur membranaire spécifique des FC et impliqué dans la fusion FC/FCM joue aussi un rôle dans la fusion des myoblastes chez les vertébrés (Srinivas et al., 2007). Pour autant, un processus de reconnaissance hétérotypique entre deux types de myoblastes et un rôle de ce mode de fusion dans le contrôle de l'identité musculaire n'ont pas, pour le

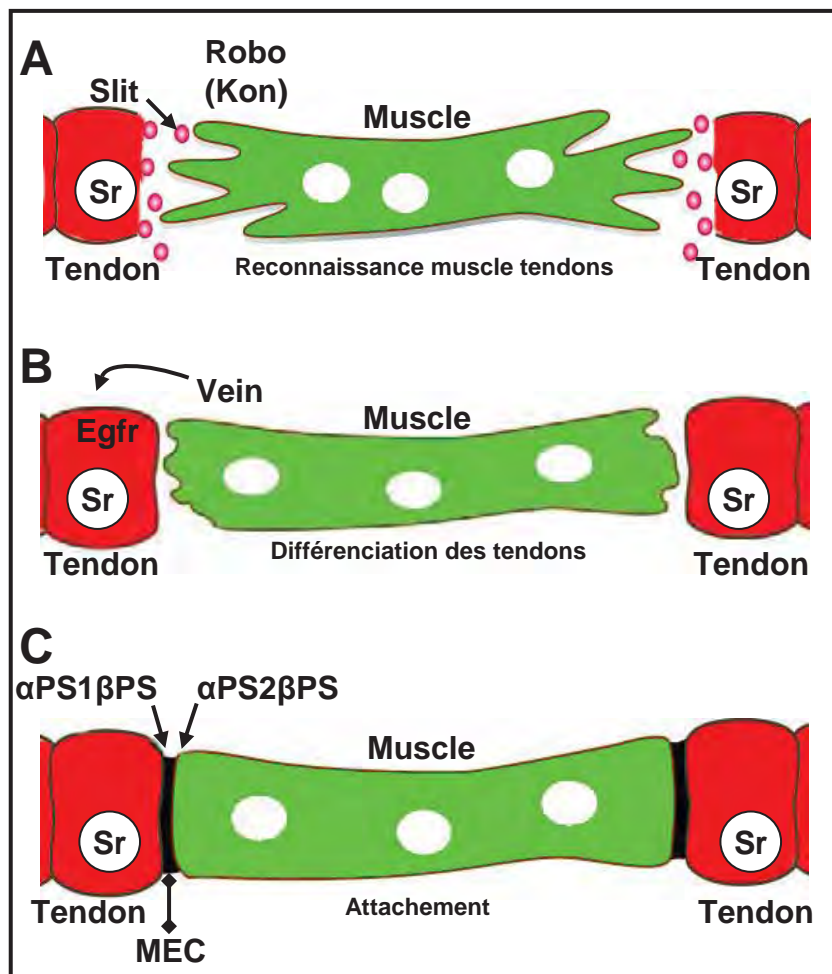


Figure 14. Schéma des étapes d'attachement des muscles aux tendons. *Adapté de Schweitzer R et al., 2010*

moment été mis en évidence chez les vertébrés et restent un sujet de débat (Powell and Wright, 2012).

e. Attachement des fibres musculaires à l'épiderme

Au cours du processus de fusion, la fibre en croissance s'allonge pour atteindre ses sites d'attachement à l'épiderme, où elle interagit avec des cellules de tendons (Schnorrer and Dickson, 2004; Schweitzer et al., 2010). Le facteur de transcription à doigt de zinc Stripe (Sr) est nécessaire et suffisant pour spécifier les précurseurs des cellules de tendons qui occupent des positions fixes dans l'épiderme (exosquelette) (Frommer et al., 1996; Becker et al., 1997). Dans des mutants *Sr*, les tendons ne sont pas spécifiés et la musculature est complètement désorganisée. Inversement, un gain de fonction de *Sr* dans le mésoderme induit la formation de tendons ectopiques et provoque une ré-orientation de muscles (Vorbruggen and Jackle, 1997). Ces données montrent d'une part que *Sr* est le principal FT requis pour la formation des tendons, et d'autre part que la position des cellules exprimant *Sr* dicte les sites d'attachement des muscles à l'épiderme. Si la spécification des sites d'attachement est indépendante de celle des muscles, la différenciation terminale des muscles et des tendons est initiée par les contacts établis entre la myofibre et les cellules de tendon (Becker et al., 1997). La différenciation terminale des tendons est induite par la sécrétion, par les muscles, de la molécule Vein, un ligand de la voie EGF (Epidermal Growth Factor). Vein sécrété par les muscles s'accumule dans l'espace intercellulaire entre muscles et tendons (Strumpf and Volk, 1998). En se fixant sur le récepteur EGFR exprimé par les tendons, il active la voie EGF et la maturation finale des tendons (Figure 14 B).

La liaison des muscles aux sites d'attachements doit résister aux forces exercées par les contractions musculaires. Elle se fait par l'intermédiaire d'une matrice extracellulaire (MEC) située entre le muscle et le tendon. Cette liaison implique des récepteurs appelés intégrines. Ces récepteurs présents sous forme d'hétérodimères, α PS2/ β PS dans les muscles et α PS1/ β PS dans les tendons se lient à des composants de la MEC comme les protéines Tigrin et Laminin, sécrétées respectivement par les tendons et les muscles, assurant ainsi la jonction muscle/tendon (Figure 14 C), (Brown, 2000; Brown et al., 2002). Dans des mutants affectant certaines chaînes d'intégrine, les muscles et les tendons forment bien des pseudo-contacts, mais les muscles se détachent dès que les embryons commencent à se contracter (MacKrell et al., 1988; Brown, 1993).

Le patron extrêmement stéréotypé de la musculature larvaire ainsi que les phénotypes mutants de FTi tels que *Col* ou *Tup* (voir résultats) indiquent que chaque muscle a la capacité de sélectionner ses sites d'attachement. Les signaux par lesquels les tendons pourraient contribuer à la guidance des muscles restent encore largement méconnus. Le

récepteur Robo exprimé par un sous-ensemble de muscles permet la guidance de ces muscles vers les cellules de tendon exprimant le ligand Slit (Figure 14 A) (Kramer et al., 2001). La protéine KonTiki est exprimée à la surface des lamellipodes de certains des muscles au cours de leur extension et interagit avec le récepteur à glutamate pour la guidance de ces muscles mais le ligand potentiel exprimé par les cellules de tendon reste inconnu (Schnorrer et al., 2007). Finalement, un autre récepteur RTK le récepteur Derailed (Drl) est exprimé dans un sous-ensemble de muscles, en particulier les muscles LT1, LT2 et LT3. Dans un mutant *drl*, les muscles LT dépassent fréquemment leur site normal d'attachement et utilisent des sites ectopiques (Callahan et al., 1996). Très récemment, Wnt5 a été identifié comme le ligand de Drl actif pour la guidance musculaire (Lahaye et al., 2012). Wnt5 est exprimé par les cellules de tendon et les muscles, et la question des mécanismes de choix des sites d'attachement n'est pas résolue.

L'analyse de mutants hypomorphes pour *col* a montré que le processus d'attachement des muscles est plus élaboré qu'initialement pensé. En effet cette analyse a révélé l'existence d'une étape intermédiaire (stade 14) durant laquelle le muscle DA3 est attaché transitoirement à un site différent de ses sites finaux d'attachement. Au stade 16 cet attachement transitoire est perdu dans les embryons sauvages et maintenu dans les embryons mutants pour *col*. L'interprétation donnée par (Enriquez et al., 2012) est un processus d'attachement en 2 phases : attachement transitoire de la FC (à sa position de naissance) et exploration d'autres sites puis une phase de résolution sélectionnant les « meilleurs » sites finaux. Cette interprétation reste une hypothèse de travail sur laquelle je reviendrai dans la discussion du phénotype mutant *tup*.

f. Innervation

La contraction des muscles larvaires est sous le contrôle de l'innervation par des motoneurones (neurones moteurs) qui deviennent actifs à la fin de l'embryogenèse. Cette activité permet à la larve de s'extraire de la membrane vitelline. Chaque segment abdominal comprend un patron stéréotypé de 36 motoneurones innervant les 30 muscles (Halpern et al., 1991; Landgraf et al., 1997; Hoang and Chiba, 2001). Le choix des connexions entre les motoneurones et les muscles est imposé par des molécules secrétées ou localisées à la surface des cellules. Parmi ces molécules on retrouve les protéines LRR (Leucine-Rich Repeat) qui jouent aussi un rôle majeur dans la guidance des motoneurones (Kurusu et al., 2008). Elles peuvent avoir un effet attractif comme Connectin (Con) et Capricious (Caps) (Nose et al., 1992; Shishido et al., 1998) ou un effet répulsif comme la protéine membranaire Toll (Halfon et al., 1995). Les muscles contribuent aussi à la stéréotypie des connexions

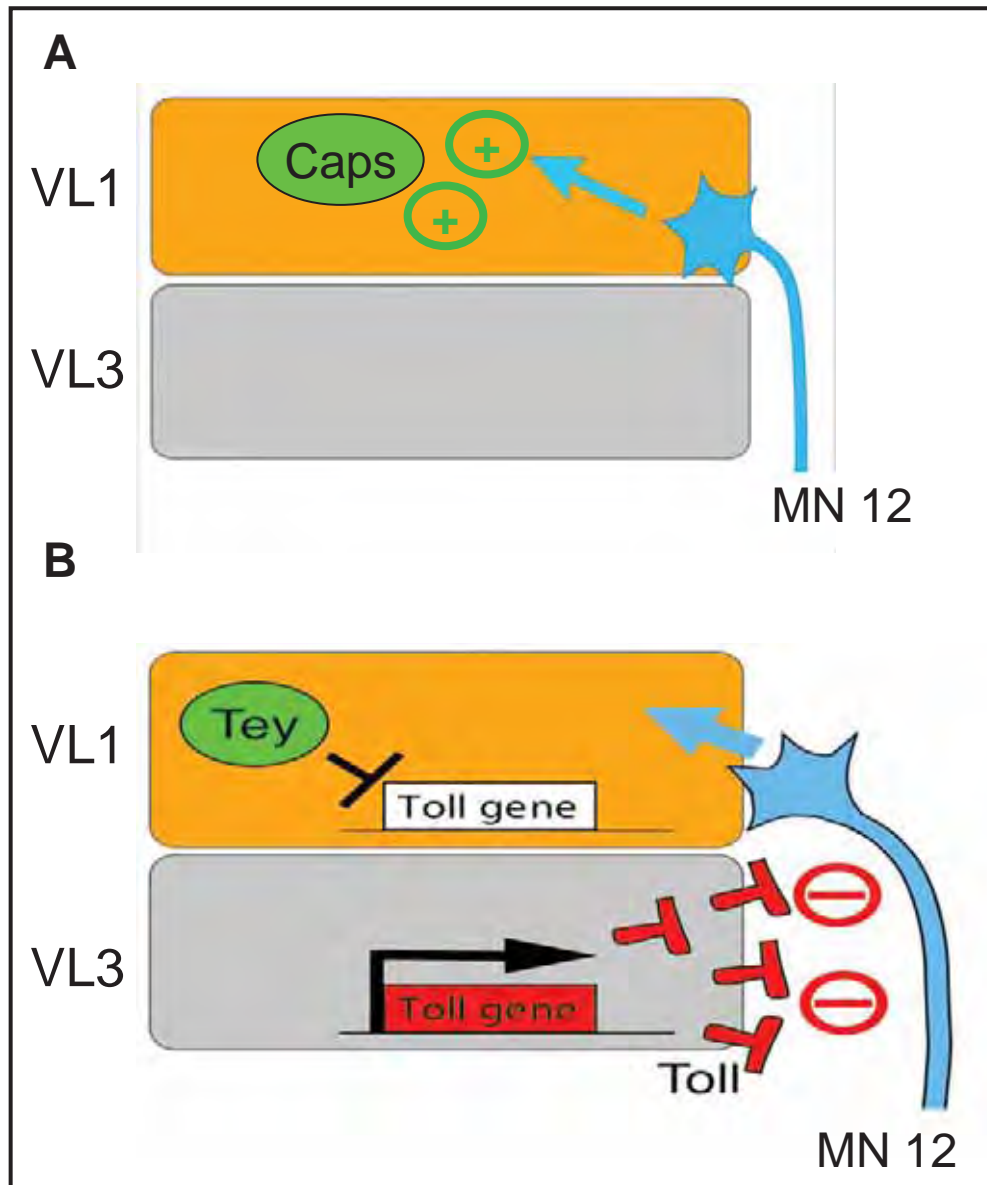


Figure 15. Innervation sélective des muscles par les motoneurones. **A.** L'innervation du muscle VL1 par le motoneurone 12 est guidée par Capricious. **B.** Toll est exprimé dans le muscle VL3 et réprimé dans le muscle VL1 par le FT Tey. Toll agit comme une molécule répulsive du MN 12. *Adapté de Nose, 2012*

muscle/motoneurone en émettant des extensions filopodiales, appelées « Myopodes » qui permettent d'instaurer des signaux bidirectionnels essentiels pour la reconnaissance neuromusculaire et l'établissement d'un patron d'innervation invariant. L'exemple de l'innervation du muscle VL1 par le motoneurone M12 illustre bien ce processus. Les myopodes du muscle VL1 arborent à leurs extrémités la protéine Caps. Des expériences d'imagerie sur embryon vivant ont montré que Caps joue un rôle clé dans la reconnaissance et la formation des synapses entre M12 et le muscle VL1 (Kohsaka and Nose, 2009; Nose, 2012). A l'inverse, l'expression de Toll dans le muscle VL3 inhibe la connexion entre le M12 et le muscle VL3. Récemment, il a été montré que Toll est spécifiquement réprimé dans le VL1 par le FTi Tey (Teyrha-Meyhra) (Inaki et al., 2010). Cet exemple souligne le rôle du code des FTi dans la spécificité du patron des connexions neuromusculaires (Figure 15).

3. Mise en place du programme d'identité musculaire au cours de la spécification des myoblastes.

La morphologie spécifique de chaque muscle larvaire de la drosophile reflète l'existence d'un programme « identitaire » intégré au programme myogénique général. L'hypothèse du contrôle de l'identité musculaire par l'expression d'une combinatoire spécifique de FT dit « identitaires » (FTi) exprimés dans le PC/FC à l'origine de chaque muscle a été émise très tôt (Bate, 1990; Bourgouin et al., 1992). Bien que de nombreux FTi aient été identifiés au cours des années, notre compréhension de l'activation des FTi et des réseaux de régulation auxquelles ils contribuent contrôlé reste très fragmentaire. L'expression des FTi est dynamique au cours de la myogenèse et diffère selon le FTi et le lignage musculaire. Afin de résumer l'état des données actuelles, j'ai choisi de présenter d'abord la découverte des premiers FTi puis de compléter en prenant comme fil conducteur l'axe dorso-ventral de l'embryon.

a. Slouch et Apterous: les premiers gènes d'identité musculaire identifiés

La caractérisation du FT à homéodomaine Slouch (S59) par M. Frasch et ses collaborateurs, en 1990, a permis de confirmer l'existence des cellules fondatrices exprimant des programmes spécifiques. S59 est d'abord exprimé dans un groupes promusculaire à l'origine des muscles VA1, VA2, VA3 et d'une AMP ventrale. Un peu plus tard, il est aussi exprimé dans les PC à l'origine des muscles DO4 et DT1, et VT1 et LO1 (Dohmann et al., 1990). Au cours de la différenciation musculaire, S59 est maintenu exclusivement dans un sous ensemble de muscles, les muscles DT1, VT1 et VA2, qui sont respectivement issus de la division asymétrique des PC DO4/DT1, VT1/LO1 et VA2/VA3. L'absence d'expression dans

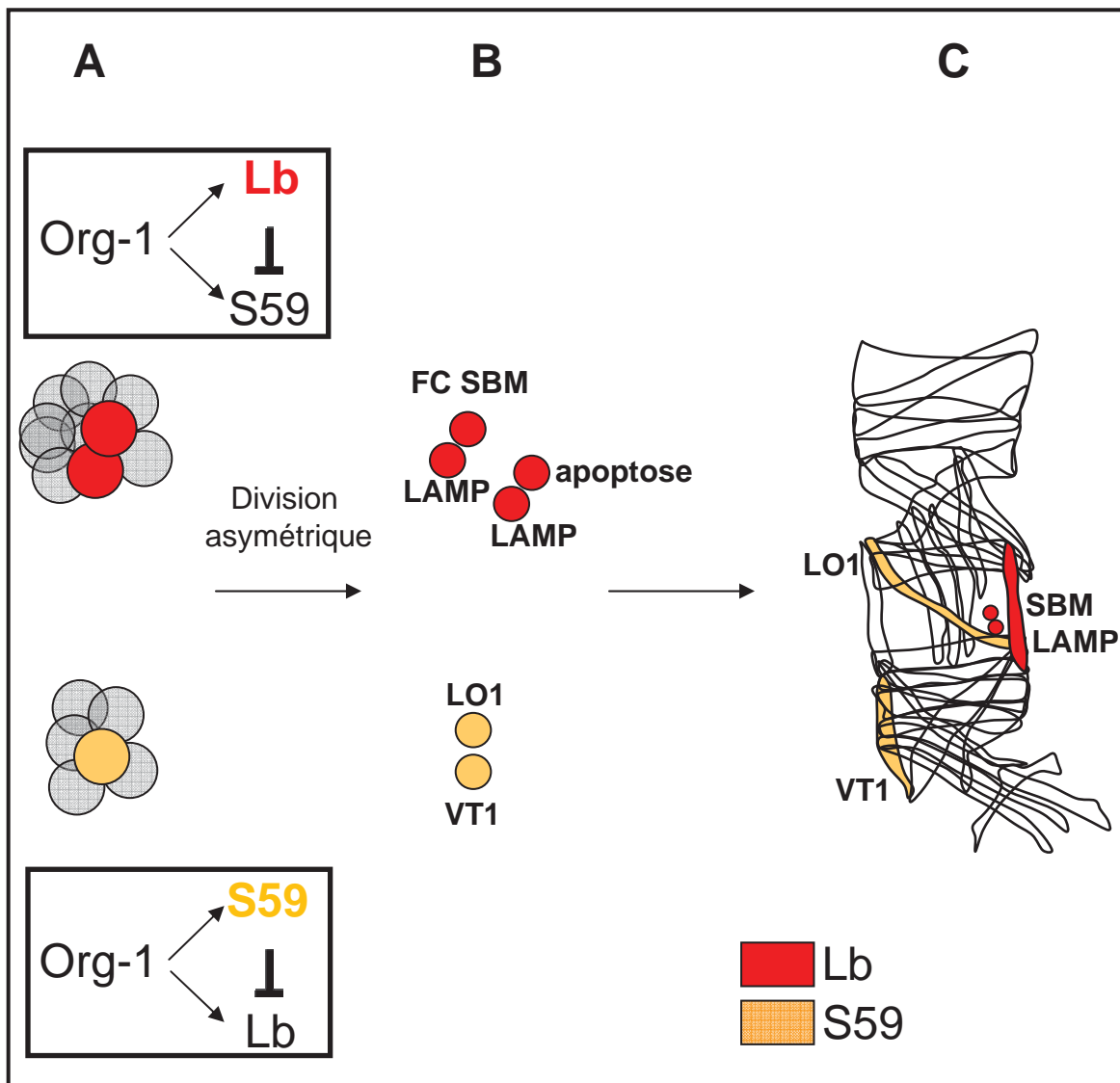


Figure 16. Modèle de répression croisée entre deux FTi. Org-1 active Lb and S59 dans 3 PC. La répression croisée entre Lb et S59 distingue les destins SBM et LO1/VT1. Le biais initial dans cette répression demeure inconnu.

les muscles jumeaux illustre bien le rôle de la division asymétrique des PC dans la diversification des programmes identitaires. L'analyse d'embryons mutants S59 a montré que le muscle DO4 est transformé en muscle DT1, c.a.d., le lignage jumeau, le muscle LO1 est transformé en muscle SBM (nous y reviendrons) et que les muscles VA3 et VT1 sont absents. L'ensemble des données d'expression et des phénotypes mutants indique que S59 est un FTi qui est exprimé et requis à différentes étapes dans la formation de sous-ensembles de muscles, y compris de façon transitoire. Au cours du processus de différenciation des fibres DT1, VT1, VA2, S59 est détecté dans les noyaux des FCM nouvellement fusionnés, suggérant que tous les noyaux d'une même fibre musculaire acquièrent l'identité de la cellule fondatrice (Dohrmann et al., 1990). Des résultats similaires obtenus par hybridation *in situ* avec des sondes introniques permettant de détecter les transcrits primaires de *col* ont montré que la propagation du programme identitaire de la FC aux noyaux de FCM n'était pas due à l'accumulation passive du même FTi dans tous les noyaux d'un syncytium mais correspondait à une reprogrammation transcriptionnelle active (Crozatier and Vincent, 1999; Dubois et al., 2007). Ce mode de reprogrammation, qui assure que tous les noyaux d'une fibre musculaire en cours de différenciation acquièrent la même identité restait cependant à généraliser.

Apterous (Ap), un FT appartenant à la famille des LIM-homéodomaine, est impliqué dans divers processus au cours du développement. Il est exprimé dans plusieurs groupes promusculaires et maintenu dans les muscles latero-transverses (LT1-4) et ventraux (VA2-3). Ap est requis pour la formation des muscles LT1-3. Inversement, on observe une duplication de muscles LT dans des conditions d'expression ectopique d'Ap (Bourgouin et al., 1992). Apterous agit donc comme un FTi des muscles LT. En conclusion, les études pionnières sur Slou et Ap ont conforté l'hypothèse de cellules fondatrices des muscles dont l'identité était le reflet de leur expression de combinaisons de FTi spécifiques.

b. Lb et S59 : Exemple de répression croisée entre FTi

Lady bird (Lb) est un FTi à homéodomaine exprimé dans le muscle SBM (muscle localisé le long de la frontière segmentaire) (Jagla et al., 1998). L'expression de Lb est détectée dans un groupe promusculaire à partir duquel deux PC exprimant Lb sont sélectionnés (Figure 16A). La division asymétrique d'un de ces deux PC est à l'origine du SBM et d'une AMP latérale (LAMP) ; la division de l'autre PC génère une autre LAMP et une cellule qui rentre en apoptose (Figure 16 B). À la fin de l'embryogenèse, Lb est exprimé dans le SBM et dans les deux LAMP (Figure 16 C). Pour rappel, S59 est exprimé dans le PC LO1/VT1, un PC situé à une position adjacente au PC SBM/LAMP exprimant Lb (Figure 16 A B). Dans des

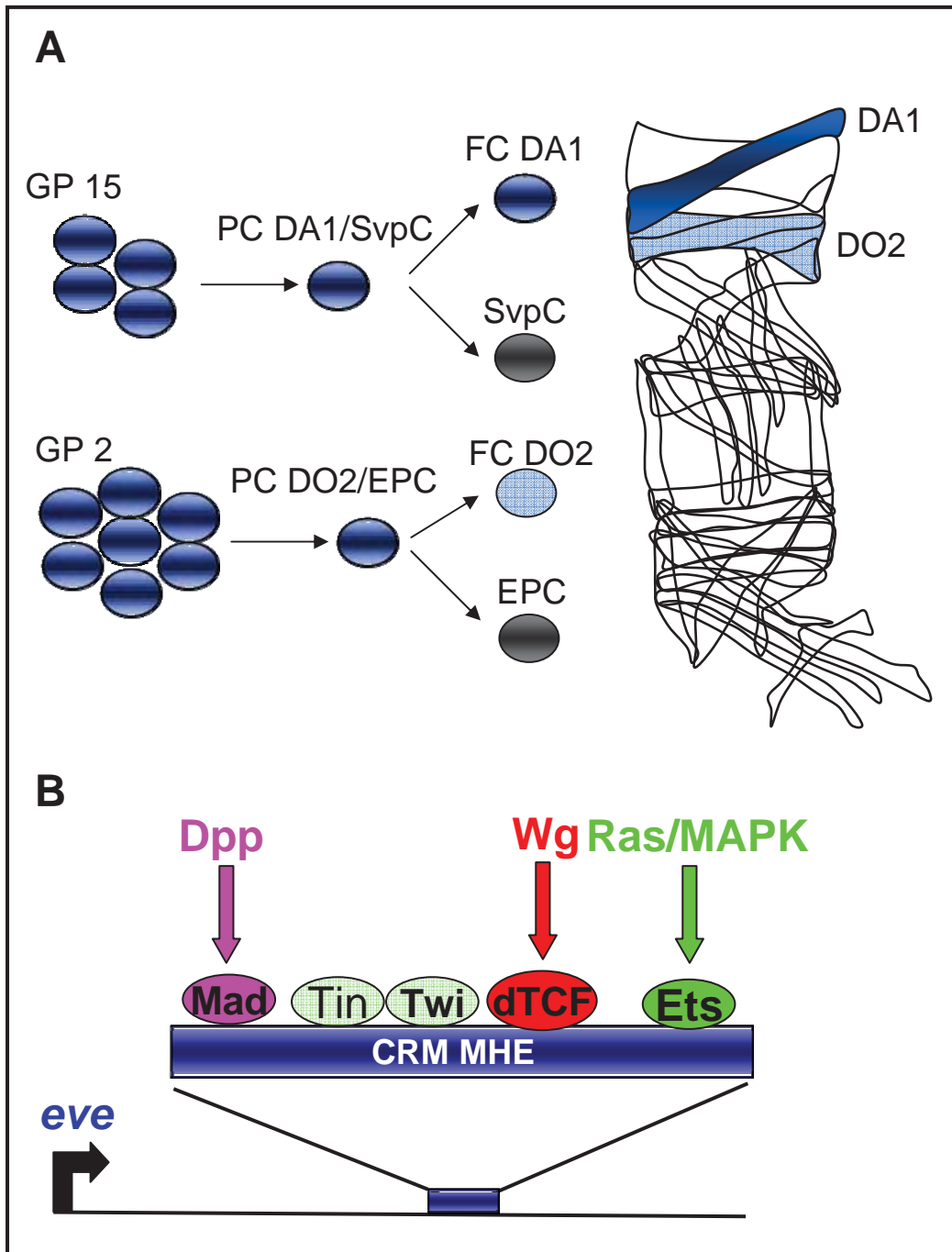


Figure 17. Intégration de l'information de position par le CRM MHE d'*eve*. **A.** Expression d'*Eve* (bleu) dans les GP15 et 2 à l'origine des FC DA1 et DO2. **B.** Schéma du CRM MHE, des FT qui s'y lient et des signalisations mises en jeu.

embryons mutants S59, le PC LO1/VT1 exprime de manière ectopique Lb, avec comme conséquence une duplication du lignage SBM et la perte des muscles LO1 et VT1. Ceci montre que S59 réprime Lb dans le progéniteur LO1/VT1. Inversement, S59 est réprimé par Lb dans le PC SBM/LAMP. Le rapprochement de ces données a révélé un mécanisme de répression mutuelle établi entre deux FTi, qui permet de conférer des identités distinctes à des PC dont les coordonnées spatiales sont très proches (Jagla et al., 1998; Junion et al., 2007). Plus récemment, il a été établi que l'homologue du FT Tbx1 de souris, le FT Org-1 (Optomotor-blind-related-gene-1 ; famille des FT à T-Box) est un régulateur direct de Lb et de S59 dans les PC LO1/VT1 et SBM/LAMP. La nature du biais menant à une répression croisée entre ces 2PC reste cependant inconnue (Schaub et al., 2012) (Figure 16).

c. Expression d'Even-skipped (Eve) dans des muscles dorsaux

Eve est un FT à homéodomaine qui a d'abord été caractérisé comme un gène de segmentation (Nusslein-Volhard and Wieschaus, 1980; Frasch et al., 1987), puis impliqué dans la spécification de neurones du système nerveux central. Aux premières étapes de la myogenèse, dans le mésoderme, Eve est d'abord exprimé dans les groupes promusculaires GP2 et GP15 d'où sont sélectionnés les PC DO2/EPC (Eve pericardial Cell) et DA1/SvpCs (Seven Up Cardioblasts) (Doe et al., 1988; Carmena et al., 1998a). L'expression d'Eve est ensuite seulement maintenue dans le muscle DA1 et un sous-ensemble de cellules pericardiales (Figure 17 A) (Speicher et al., 2008). Pourtant, dans un mutant *eve*, les muscles DA1 et DO2 ne se forment pas, montrant l'importance de la fonction d'Eve au stade progéniteur (Carmena et al., 1998a). M Halfon et ses collaborateurs se sont intéressés à la régulation transcriptionnelle d'*eve* dans le mésoderme dorsal et ont identifié un module cis régulateur CRM (CRM), le MHE (Muscle and Heart Enhancer) (Figure 17 B). Ce CRM intègre une information tissulaire (mésoderme) conférée par Twi et Tin (Halfon et al., 2000) et une information positionnelle médiée par les voies de signalisation Wingless (Wg), Dpp (TGF β) et à récepteur tyrosine kinase (Ras/MAPK). Ces voies activent respectivement les FT dTCF (T Cell Factor), Mad (Mother Against dpp) et Ets. En condensé, l'étude des fonctions mésodermiques d'Eve a montré que Eve est un FTi dorsal et que son activation dans deux groupes promusculaires intègre une information tissulaire et une information de position. L'ensemble de ces résultats pionniers a permis pour la première fois de relier l'identité de muscles à la position de leur PC/FC.

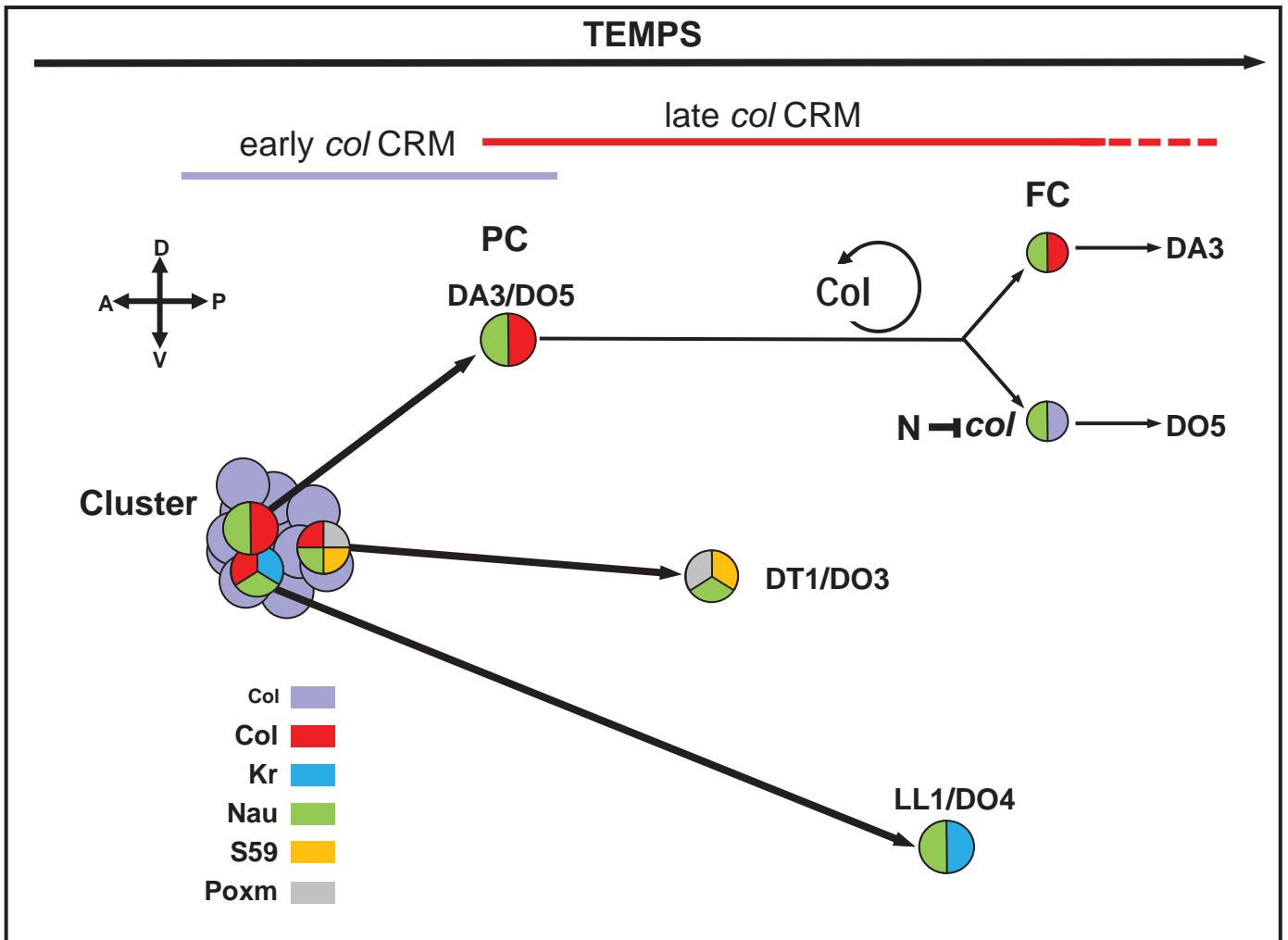


Figure 18. Deux CRM contrôlent l'expression de *col* au cours de la myogenèse. 3 PC sont séquentiellement issus du groupe promusculaire exprimant Col. Seul, le premier né continue d'exprimer Collier, il est à l'origine du muscle DA3. D'après. *Enriquez et al., 2012*

d. Interactions entre Lms, Msh et Ap et formation des muscles LT

Lms (*lateral muscles scarcer*) et Msh (muscle segment homeobox), des FT à homéodomaine, et Krüppel un FT à doigts de zinc sont trois autres FTi des muscle latéro-transverses LT (Nose et al., 1998; Muller et al., 2010). Lms est exprimé à toutes les étapes de la formation des 4 muscles LT1-4. Dans un contexte perte de fonction *lms*, les muscles LT sont partiellement affectés. Cependant, dans un double mutant *lms* et *ap*, le phénotype musculaire est aggravé, suggérant une collaboration entre ces deux gènes dans l'établissement de l'identité des LT. Msh est exprimé dans les PC/FC des muscles LT, mais pas dans les phases plus tardives. Dans un mutant *msh*, l'expression de Lms est sévèrement diminuée, suggérant que Msh est un régulateur positif de Lms dans les muscles LT. Comme nous l'avons vu précédemment Lb est un FTi du muscle SBM. Un gain de fonction de Lb obtenu par expression ectopique, provoque la répression de Lms (Muller et al., 2010). Ces données ont renforcé la conclusion que le contrôle de l'identité musculaire implique une combinatoire de FTi, agissant en réseaux, et l'importance des régulations croisées entre les FTi.

e. Morphologie des muscles dorso-latéraux : La combinatoire Collier/Nautilus.

Collier (Col) appartient à la famille des FT COE (Col/Olf/EBF), des facteurs de structure très conservée au cours de l'évolution des métazoaires. Les FT COE comportent un domaine de liaison à l'ADN qui leur est spécifique, associé à un domaine de dimérisation de type HLH atypique, et se lient à l'ADN sous forme d'homodimères avec un site consensus de liaison palindromique (Dubois and Vincent, 2001; Daburon et al., 2008). Au cours du développement embryonnaire, Col est d'abord exprimé dans le parasegment PS0 séparant la tête du tronc, d'où ce nom, puis dans le muscle DA3, divers neurones du système nerveux central et périphérique, et la glande lymphatique (organe hématopoïétique larvaire) (Crozatier et al., 1996, 1999, 2004 ; Crozatier and Vincent, 1999, 2008). Au cours de la myogenèse, l'expression de Col est d'abord détectée dans un large groupe promusculaire localisé dorso-latéralement, initialement montré comme à l'origine de deux (segments thoraciques) ou trois (segments abdominaux) PC dorso-latéraux. Ces PC apparaissent de manière séquentielle dans l'ordre chronologique suivant : DA3/DO5, DT1/DO3 (uniquement dans les abdominaux), LL1/DO4. L'expression de Col est ensuite maintenue dans le muscle DA3 jusqu'à la fin de la myogenèse et ce maintien nécessite un processus d'autorégulation directe et l'activité de protéines Hox (Figure 18) (Crozatier and Vincent, 1999) ; (Dubois et al., 2007; Enriquez et al., 2012). J'ai contribué à montrer que l'activation et le maintien de la transcription de *col* dépendaient de 2 CRM séparés (voir résultats et Figure 18). Le

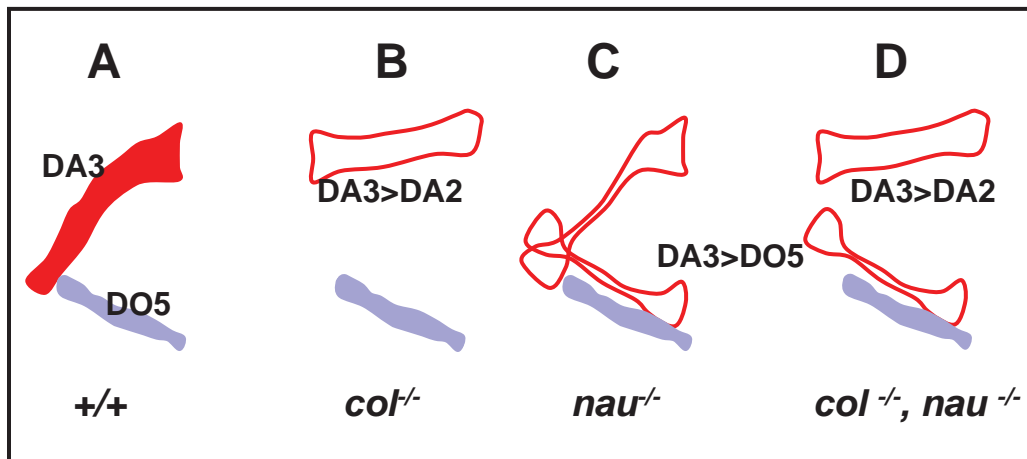


Figure 19 La Combinatoire Collier/Nautilus contrôle la morphologie du muscle DA3. Le muscle DA3 est cerclé de rouge, son jumeau, le DO5 est mauve. L'expression de Col est indiquée par le remplissage rouge. *D'après Enriquez et al., 2012*

mécanisme par lequel Col n'est maintenu que dans le muscle DA3 reste cependant à identifier. Dans un mutant *col*, il y a transformation morphologique du muscle dorso-latéral DA3 en un muscle de type dorsal DA2 dans 80% des segments. La perte de fonction de *col* induit aussi des défauts dans les autres muscles dorso-latéraux. L'ensemble de ces données montre que Col est un FTi dorso-latéral (Enriquez et al., 2012).

Nautilus (Nau), l'unique orthologue drosophile des MRF vertébrés est exprimé dans les PC/FC dorsaux aussi bien que dorso-latéraux. Son rôle exact dans la spécification de l'identité musculaire a été l'objet de nombreuses controverses (Michelson et al., 1990; Paterson et al., 1991; Balagopalan et al., 2001; Wei et al., 2007). Jonathan Enriquez, doctorant dans notre l'équipe de 2006 à 2010 et Laurence Dubois ont donc réexaminé le rôle de Nau au cours de la myogenèse et plus particulièrement au cours de la spécification du muscle DA3. L'analyse de la musculature des embryons mutants *nau* par microscopie électronique à balayage a d'abord montré que la majorité des fibres étaient beaucoup plus fines que dans les embryons sauvages. Cette première observation a révélé une fonction générale de Nau dans le contrôle de la taille des muscles. Une analyse plus détaillée a montré que la morphologie du muscle DA3 était sévèrement affectée : le muscle DA3 adopte soit la morphologie du muscle DO5, soit une morphologie bifide DA3+DO5, une transformation différente de la transformation DA3>DA2 observée dans les mutants *col* (Figure 19 A-C). En absence de *nau* et d'une dose de *col* (*col*^{1/+}; *nau*^{ck187}/*nau*^{ck187}), on observe une fréquence élevée de transformations DA3>DA2 (Figure 19 D). L'ensemble de ces résultats et la régulation transitoire de *col* par Nau ont montrés que *nau* agit à la fois comme facteur général régulateur de la taille des fibres et comme facteur identitaire. En outre, l'analyse des doubles mutants *col/nau* a étayé pour la première fois expérimentalement l'hypothèse d'un contrôle combinatoire de l'identité musculaire par les FTi (Enriquez et al., 2012), une hypothèse évoquée 20 ans auparavant, suite à la découverte de l'expression d'Apterous et de Slouch dans des sous-ensembles de muscles partiellement recouvrants (Bourgouin 1992, voir ci-dessus) .

f. Araucan et Caupolican ; régulation post-traductionnelle de l'activité des FTi.

Araucan (Ara) et Caupolican (Cau) sont deux gènes du complexe Iro-C (Iroquois gene Complex). Ces gènes codent pour des FT à homéodomaine, de fonction conservée au cours de l'évolution. Le groupe de Mar Ruiz-Gomez s'est intéressé au rôle de ces gènes au cours de la myogenèse et a montré que Ara et Cau sont impliqués dans la spécification des muscles LT où ils assurent des fonctions redondantes (Figure 20). Dans un contexte de perte de fonction de *ara* et *cau*, les muscles LT3-4 sont transformés en muscles ventraux

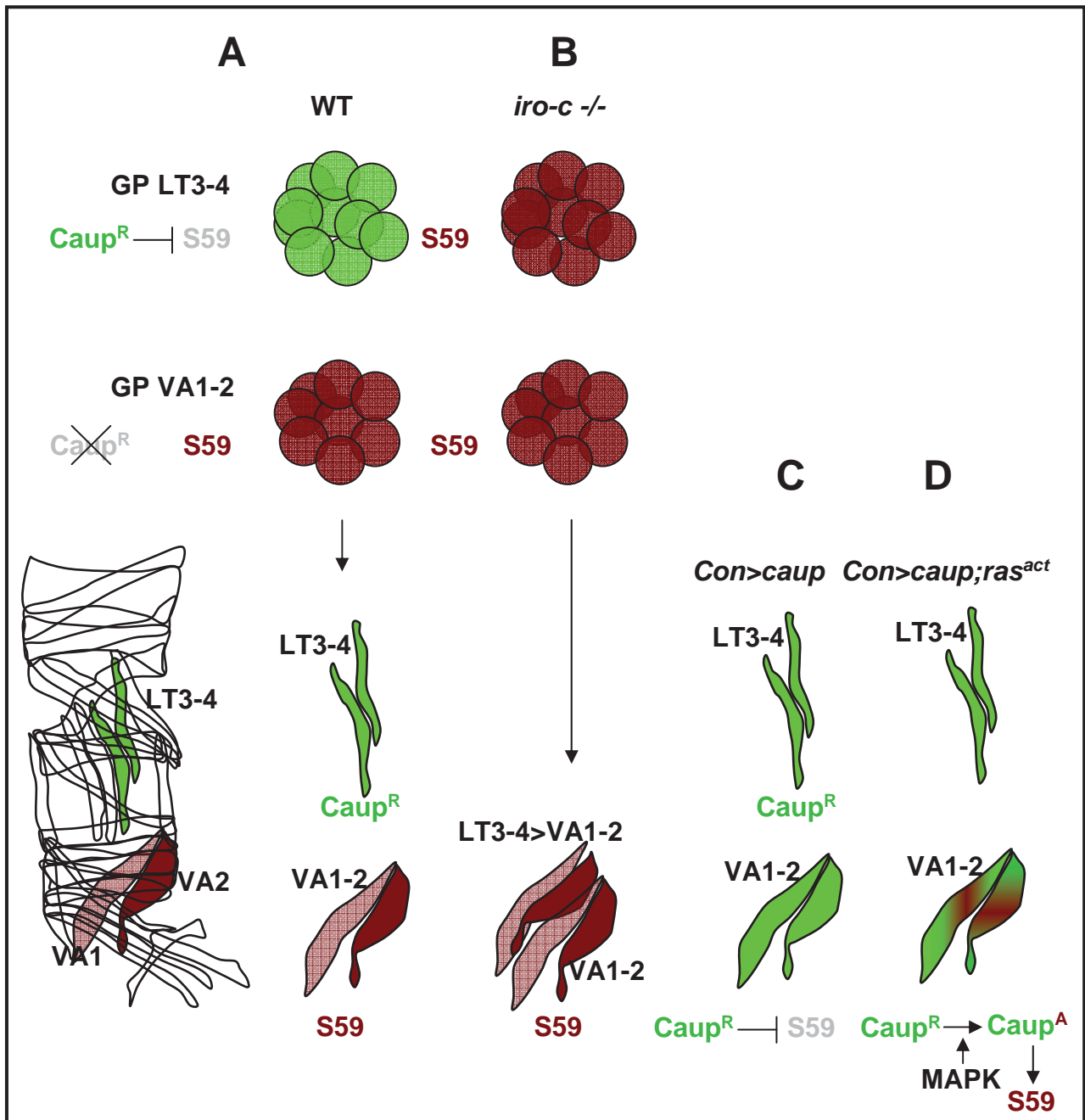


Figure 20. Régulation post-traductionnelle de Ara et Cau au cours de la formation des muscles latéraux et ventraux. Cette figure montre que la régulation post-traductionnelle du FTi Cau par phosphorylation le convertit de répresseur en activateur. Cette régulation contrôle l'identité des muscles latéraux et ventraux. *Adapté de Carrasco-Rando et al., 2011*

VA1-2. Cette transformation coïncide avec une expression ectopique de S59 et montre que dans une situation sauvage S59 est réprimé par Cau au cours de la spécification des LT (Figure 20 A-B). Les auteurs ont ensuite montré que la répression de S59 par Cau impliquait une interaction directe de Cau sur le promoteur de S59 et que cette interaction était modulée par l'activité de la voie Ras/MAPK. En effet, l'activité répresseur de Cau corrèle avec l'absence de signalisation Ras/MAPK dans le groupe promusculaire LT3-4. Inversement, la surexpression d'une forme active de Ras entraîne la phosphorylation de Cau et sa conversion de répresseur en activateur (Figure 20 C-D). L'étude des fonctions d'Ara et Cau dans la spécification de l'identité des muscles LT a révélé un autre exemple de répression croisée entre différents FTi et de plus montré que la régulation post-traductionnelle des FTi peut être décisive dans l'établissement de l'identité musculaire (Carrasco-Rando et al., 2011).

g. Les muscles ventraux : le rôle de Poxm

Nous avons vu précédemment qu'un des acteurs principaux de la spécification du mésoderme ventral est Poxm (Staehling-Hampton et al., 1994). Au cours de la régionalisation dorso-ventrale du mésoderme, Dpp réprime dorsalement l'expression de Poxm et par conséquent l'expression de Poxm est restreinte au mésoderme ventral. Poxm est exprimé d'abord dans un large domaine de compétence (appelé pré-cluster ventral), puis maintenu dans les muscles DT1 et VA1-3. La perte de fonction de Poxm conduit à des défauts peu pénétrants des muscles ventraux (VL1-3, VO1-2, VO4-6), de quelques muscles latéraux (LT3-4, SBM) et du muscle DT1. Ces phénotypes sont aggravés dans un double mutant *poxm//sc*, indiquant une redondance fonctionnelle et ou des régulations croisées entre ces deux gènes. Au stade progéniteur, Poxm active le gène identitaire S59 dans le PC DT1/DO3. Poxm est aussi nécessaire au maintien de l'expression de S59 dans le muscle VA2. L'ensemble de ces données renforce la conclusion qu'un même FTi peut assurer diverses fonctions à diverses étapes de la myogenèse. Dans le cas de Poxm, une fonction précoce avec la spécification du domaine de compétence à l'origine de plusieurs muscles ventraux, et une fonction plus tardive dans la régulation de S59 (Duan et al., 2007). De manière intéressante, et en contradiction avec cette épistasie, le phénotype mutant Poxm du muscle DT1 est moins pénétrant que le phénotype mutant S59, suggérant que des régulations compensatoires existent dans le PC exprimant les deux facteurs.

4. Les muscles alaires.

Chez la drosophile, le cœur a l'aspect d'un tube (d'où le nom tube cardiaque), et est localisé dorsalement. Le cœur s'étend le long de l'axe antéropostérieur, du segment T2 au segment A7. On peut distinguer morphologiquement l'aorte, segments T2->A3/A4 du cœur proprement dit, segments A4-A7 (Figure 21 A). L'activité pulsatile du cœur est contrôlée par un pace maker autonome et assure la circulation de l'hémolymphe (l'équivalent du sang chez la drosophile) (Medioni et al., 2008). L'organogenèse du cœur a lieu au cours de l'embryogenèse, à partir de progéniteurs issus du mésoderme dorsal (Bodmer and Venkatesh, 1998). Le tube cardiaque est constitué de deux types cellulaires : a) Les cardiomyocytes contractiles qui, alignés par paires, forment le tube cardiaque propre. B) Les cellules péricardiaques qui entourent les cardiomyocytes (Figure 21 A). Une des fonctions potentiellement assurées par ces cellules est de filtrer l'hémolymphe (nephrocyte). Sept muscles, dit muscles alaires, sont disposés bilatéralement et transversalement au cœur dans sa région abdominale, un par segment. Ils sont connectés aux cellules péricardiaques (LaBeau et al., 2009) (Figure 21 A-C). (Lehmacher et al., 2012). A ce jour, les fonctions de ces muscles restent inconnues. Puisque mes travaux de thèse m'ont amené à m'intéresser au développement embryonnaire de ces muscles, je résume ci-dessous nos connaissances actuelles aux plans cellulaire et moléculaire.

Les muscles alaires (MA) sont attachés dorsalement aux cellules péricardiaques et latéralement à l'épiderme. Une analyse détaillée par microscopie électronique en transmission (MET) a révélé que les MA sont attachés au réseau de matrice extracellulaire (MEC) entourant les cellules péricardiaques positives pour Svp. Il a été proposé que ce réseau de MEC apporte la flexibilité permettant d'amortir les tensions entre les deux types cellulaires générées par les contractions musculaires. L'autre extrémité de chaque MA est reliée à l'épiderme à des sites d'attachements de type tendon sans que ces tendons aient été caractérisés en plus de détail. Comme les muscles squelettiques, chaque MA est formé d'une seule fibre multinucléée qui comporte environ cinq noyaux. Du stade embryonnaire au stade larvaire, les MA développent des myofilaments contractiles organisés en sarcomères. L'ensemble de ces données indique que les MA sont des muscles striés (Lehmacher et al., 2012). On connaît à ce jour très peu le contrôle moléculaire de la formation des MA au cours du développement embryonnaire. En étudiant les fonctions du FT à T-box *Org-1*, M. Frasch et ses collaborateurs ont récemment identifié une des cellules exprimant *Org-1* dans chaque segment comme étant le progéniteur du MA. Ce PC dérive du mésoderme dorso-latéral. *Org-1* reste exprimé dans les MA jusqu'à la fin du développement embryonnaire (Figure 21 B). Dans des embryons mutants pour *org-1*, les MA ne sont pas spécifiés (Schaub et al., 2012). Les trois paires de MA antérieurs (segments A1-A3) expriment *Ubx* et les quatre

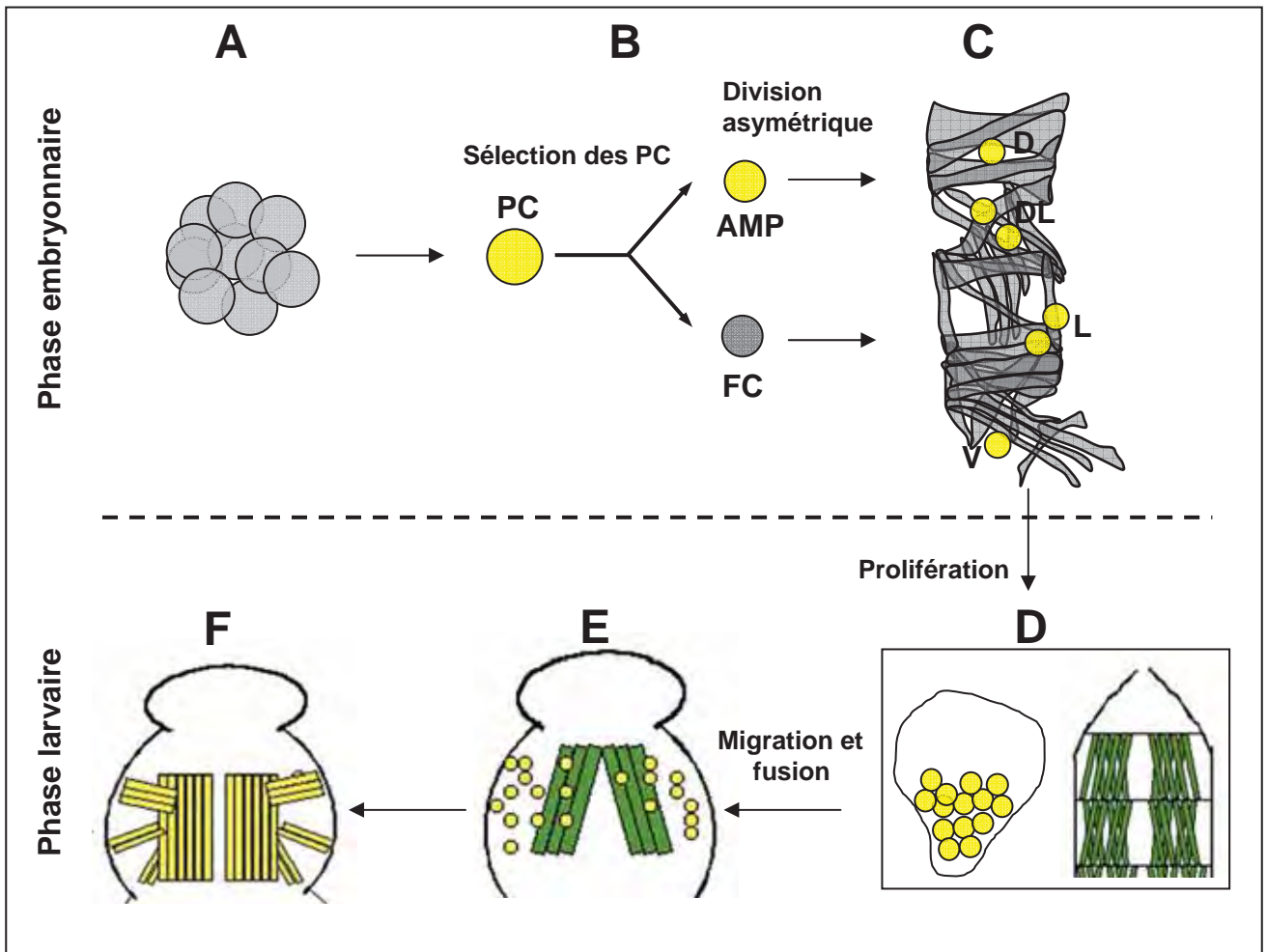


Figure 22 Schéma simplifié des étapes de spécification et de formation des muscles adultes chez la drosophile. D'après Dutta et al., 2004.

paires postérieures expriment Abd-A (segments A4-A7). Des études de perte et de gain de fonction de ces gènes *hox* ont montré que leur activité est requise pour le développement des MA (LaBeau et al., 2009). Un autre FT exprimé dans les MA est le FT Tailup (*Tup/islet1*, FT appartenant à la famille des LIM-homéodomaine, homologue de *Islet-1* chez les vertébrés) sans qu'aucune donnée fonctionnelle n'ait été associée à cette expression (Tao et al., 2007) (Figure 21 C). Nous reviendrons en détail sur les fonctions possibles de *Tup*, dans la partie résultats.

5. Les cellules précurseurs des muscles adultes.

Contrairement aux muscles larvaires, les muscles adultes sont formés de plusieurs fibres, une structure proche de celle des muscles squelettiques vertébrés. Les muscles dorso-longitudinaux du vol sont issus du remodelage de muscles larvaires (Fernandes et al., 1991). Cependant, la majorité des muscles adultes, dont les muscles des pattes et les muscles thoraciques ventro-latéraux « du vol » sont issus de la prolifération et la différenciation de précurseurs musculaires adultes (AMP) associés aux disques imaginaires (Broadie and Bate, 1991). Les AMP sont issues de la division asymétrique de PC dont une des cellules filles devient une FC, l'autre un AMP (Figure 22 A-B) (Ruiz Gomez and Bate, 1997). A la fin de l'embryogenèse, chaque segment abdominal contient six AMP positionnées selon un patron bien défini : 1 AMP dorsale (DAMP), 2 dorso-latérales (DLAMP), 2 latérales (LAMP), et 1 ventrale (VAMP) (Figure 22 C) (Bate et al., 1991). A ce stade, les AMP expriment spécifiquement une cible de la voie de signalisation Notch, *Hm* (Holes in muscle) et le FT *Zfh1*. Ces deux FT permettent de supprimer la capacité de *Mef-2* à activer le programme de différenciation musculaire, maintenant ainsi les AMP dans un état indifférencié (Figeac et al., 2010). La voie de signalisation EGF est aussi impliquée dans le maintien de l'état quiescent des AMPs. Des FTi spécifiques sont exprimés dans chaque AMP comme *Lb*, *S59*, *Poxm*, reflétant leur position. Ces FTi semblent requis pour la spécification des AMP, mais par un mécanisme qui reste inconnu. Après spécification, les AMP forment un réseau de cellules interconnectées par des extensions cytoplasmiques. Ces connections sont aussi observées entre AMP localisés dans des segments différents. La destruction de ces connexions par des expériences d'ablation laser a montré qu'elles étaient nécessaires au maintien de la position des AMP dans l'embryon et de leur état quiescent (Figeac et al., 2010).

Les AMP prolifèrent au cours des stades larvaires sans se différencier. Ainsi, chaque segment de l'abdomen d'une larve au 3^{ème} stade larvaire contient six groupes d'AMP, chaque groupe est composé de 8 à 15 cellules. 24 heures après la pupaison et hormis les muscles obliques qui servent de matrice à la formation des muscles dorso-longitudinaux,

tous les muscles larvaires sont histolysés. Les AMP fusionnent entre elles ou avec les muscles obliques afin de former les muscles adultes (Figure 22 D-F) (Fernandes et al., 1991). Des expériences d'ablation des AMP ont montré que contrairement aux FC, une AMP peut contribuer à plusieurs muscles adultes.

6. Les réseaux de régulation du mésoderme cardio-pharyngal, chez les chordés ; le Facteur Islet 1.

Le facteur de transcription à LIM homédomaine Islet (Tup) a d'abord été identifié pour son rôle dans la régulation de l'expression du gène de l'insuline chez le rat (Karlsson et al., 1990). La similitude de séquence avec les protéines Lin-11 et Mec-3 du nématode *C. elegans* a montré la présence d'un nouveau domaine à doigt de zinc conservé au cours de l'évolution. Ce domaine, appelé LIM pour Lin/Islet/Mec, est un domaine d'interaction protéine-protéine. L'expression d'Islet 1 au cours du développement a ensuite été détectée dans des neurones du PNS et des motoneurones dans la moelle épinière des rongeurs (Thor et al., 1991) et les neurones primaires chez le poisson-zèbre; Islet1 est nécessaire à la différenciation des motoneurones (Korzh et al., 1993). Thor et Thomas (1997) ont identifié Isl1 chez la drosophile et montré qu'il était requis pour la navigation des axones des motoneurones et la sécrétion de neurotransmetteurs, dopamine et sérotonine. Islet1 et *tail-up (tup)* correspondent au même gène. Bien que *tup* ait été identifié dans « le » crible de Nusslein-Volhard, Wieschaus et Kluding (1984), en raison d'un phénotype « de rétraction incomplète de la bandelette germinale et de tête élargie », ce n'est que récemment que son rôle dans la formation du cœur et de la glande lymphatique, le site de l'hématopoïèse définitive de la drosophile a été décrit en détail (Tao et al., 2007 ; Mann et al., 2009). Entre temps, l'expression de Islet1 avait identifié une population de progéniteurs prolifératifs et contribuant à la formation du cœur chez la souris (Cai et al., 2003). En fait, Islet1, comme Tbx1, est un marqueur précoce du territoire cardiaque secondaire (TCS) chez les vertébrés. Pour rappel Nkx2.5 et Tbx1, respectivement les orthologues de Tin et Org-1 chez la drosophile, sont (aussi) exprimés dans le croissant cardiaque chez les vertébrés (territoire cardiaque primaire). Le TCS contribue au ventricule droit, l'ostium aortique et la base de l'aorte et l'ostium du tronc pulmonaire (Vincent and Buckingham, 2009). L'analyse de l'expression d'Islet1 dans le TCS et au cours de la formation du cœur chez le poulet et la souris a révélé l'existence de progéniteurs communs au cœur et à des muscles de la tête, plus exactement les muscles issus des arcs branchiaux 1 et 2. Le TCS est donc du mésoderme cardio-pharyngal (Tirosh-Finkel et al., 2006; Nathan et al., 2008). Des expériences de lignage rétrospectif réalisées chez la souris ont ensuite montré la ségrégation précoce de deux lignages cardio-pharyngaux : un donne les muscles dérivant de

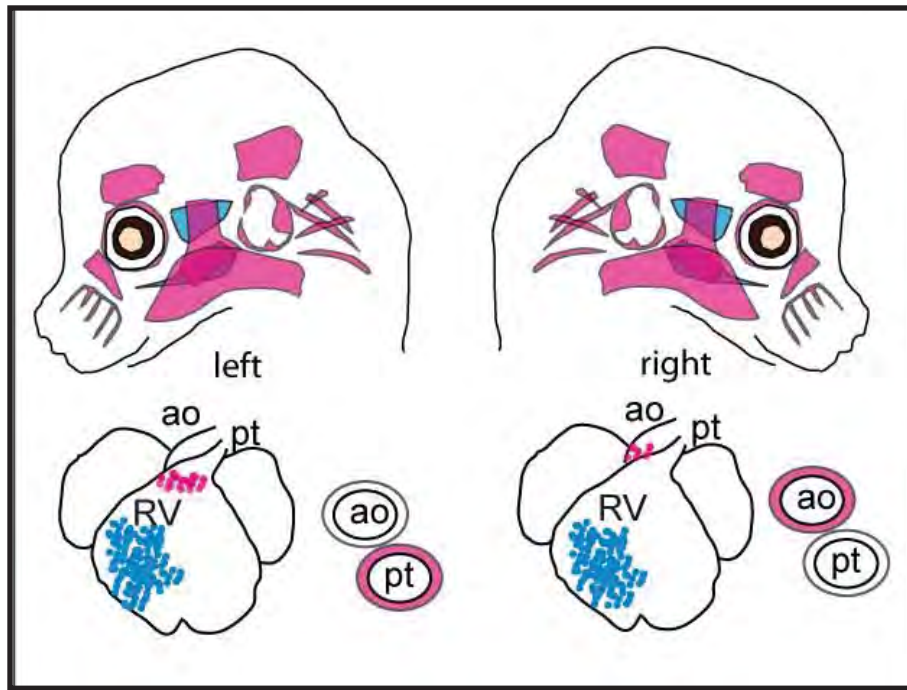


Figure 23. Les progéniteurs présents dans le territoire cardiaque secondaire (TCS) sont à l'origine de muscles de la face (haut) et de structures cardiaques (bas). 2 lignages sont présents parmi ces progéniteurs : **1.** Muscles de la mastication appelé muscle masséter (dérivés bleus) et ventricule droit (RV), **2.** Muscles faciaux (dérivés rouges) et bases de l'aorte (ao) et du tronc pulmonaire (pt). *Adapté de F. Lescroart et al., Development, 2010*

l'arc branchial 1 (Masseter et Temporalis) et le ventricule cardiaque droit ; l'autre donne les muscles dérivant de l'arc branchial 2 (muscles faciaux) et la base de l'aorte et du tronc pulmonaire (Figure 23) (Lescroart et al., 2010; Tzahor and Evans, 2011). L'étude de l'expression des orthologues de Nkx2.5 (NK4) et Islet chez le tunicier *Ciona intestinalis*, une espèce appartenant au phylum le plus proche des vertébrés, a montré qu'Islet est exprimé dans les précurseurs des muscles pharyngaux (siphon atrial et muscles longitudinaux) notés ici ASM (Stolfi et al., 2010; Tolkin and Christiaen, 2012). Ces précurseurs sont eux-mêmes issus de la division de précurseurs ventraux exprimant les facteurs NK4 (Nkx2.5), Hand, GATA4 à l'origine du cœur rudimentaire de *Ciona* et des ASM, suggérant que l'existence du mésoderme cardio-pharyngal est commun à l'ensemble des chordés (Tolkin and Christiaen, 2012). Ces travaux ont aussi révélé que Ci-COE, l'orthologue de Col est co-exprimé avec Islet1 et Tbx-1 (Org-1 chez la drosophile) dans les ASM (Isl1 et Tbx-1 sont aussi exprimés dans le cœur). De l'ensemble de ces données, on peut conclure que le nœud des réseaux de régulation transcriptionnelle spécifiant le cœur secondaire et les muscles pharyngaux des chordés et les vertébrés comprend les FT Nkx2.5, Tbx1 et Islet1. Les études sur *Ciona* et l'expression de Xebf2 et Xebf3 (un orthologue de collier chez le xénope) dans le mésoderme des arcs branchiaux du xénope (Dubois et al., 1998; Stolfi et al., 2010) suggèrent en outre que les FT COE puissent être des acteurs majeurs de la spécification des muscles pharyngaux, sans avoir de rôle dans la formation du cœur.

7. Conclusion et nouvelles questions

20 années de recherche sur les mécanismes de contrôle de l'identité musculaire et du rôle des FTi au cours de la myogenèse de la drosophile ont mis en évidence la complexité des réseaux de régulations assurés par un ensemble de FT très divers, exprimés de manière spécifique, et à différentes étapes de la myogenèse dans différents lignages musculaires. Les recherches sur l'expression d'*eve* dans le mésoderme dorsal ont montré que la spécification de l'identité musculaire passe par l'activation transcriptionnelle de FTi spécifiques, à une position donnée dans le mésoderme, connectant ainsi information de position et identité musculaire. Les études menées sur le rôle des facteurs Ap, Lb et Slou dans la spécification des muscles LT et SBM ont révélé l'importance des régulations croisées entre différents FTi. L'étude du rôle d'Ara et Cau a permis de souligner l'existence de modifications post-traductionnelles dans ces régulations croisées. Enfin, la notion de contrôle combinatoire de l'identité musculaire par un ensemble (code) de FTi est maintenant bien étayée. L'étude comparée de l'expression et des fonctions de Col et Nau dans le lignage DA3 a permis de montrer que le contrôle combinatoire de l'identité musculaire

mettait en jeu à la fois des régulations transcriptionnelles croisées et la régulation, par chaque FTi d'une combinatoire, de propriétés spécifiques des muscles.

Malgré ces avancées, beaucoup de questions restaient posées au début de ma thèse, en 2008, parmi lesquelles on peut distinguer celles ayant trait au processus de spécification de l'identité musculaire et celles ayant trait au processus de réalisation de cette identité. Au cours de ma thèse, je me suis plus particulièrement intéressé au premier aspect. J'ai d'abord contribué à l'étude de l'intégration de l'information tissulaire et positionnelle dans l'activation du gène *col* en identifiant le CRM responsable de cette activation. Ceci m'a conduit à étudier plus en détail le mécanisme de relais entre l'activation promusculaire et le maintien de l'expression de *col* dans le muscle DA3. J'ai pu mettre en évidence un mécanisme type « passage de témoin » qui a lieu au stade PC. J'ai ensuite étudié la fonction du FT *Tup/Islet1*, dans la spécification des muscles dorsaux. Cette étude a révélé de nouveaux parallèles entre le développement du cœur et des muscles dorsaux chez la drosophile et le développement du cœur et des muscles pharyngaux chez les chordés. D'autres observations, non publiées et jointes à la section résultats ont trait à la question du lignage des FCM et à la formation des muscles alaires.

Résultats

A. Régulation transcriptionnelle de Collier et identité musculaire.

Article 1. Multi-step control of muscle diversity by Hox proteins in the *drosophila* embryo. Jonathan Enriquez, Hadi Boukhatmi, Laurence Dubois, Anthony A. Philippakis, Martha L. Bulyk, Alan M. Michelson, Michèle Crozatier and Alain Vincent.

Dans cet article, notre laboratoire a montré que le stade PC est l'étape clé qui intègre les informations de position intra- et inter-segmentaire dans le processus de spécification de l'identité musculaire. Alors que l'activation transcriptionnelle de *col* dans le groupe promusculaire à l'origine du PC DA3/DO5 est observée dans tous les segments, le muscle DA3 ne se forme que dans les segments T2-A7. Cette étude montre que le maintien de l'expression de *col* dans le PC DA3/DO5, dépend des gènes homéotiques *Antp*, *Ubx* et *abdA* et que cette régulation est directe, et agit sur un CRM mésodermique tardif de *col* (L^{col} -CRM) préalablement identifié par L. Dubois et al, (2007). Les protéines homéotiques contrôlent aussi le nombre de noyaux incorporés dans une fibre musculaire donnée, ou plus précisément la variation de ce nombre entre les différents segments.

Ma contribution à cet article a été d'identifier le CRM précoce responsable de l'activation mésodermique de *col* dans un groupe promusculaire (E^{col} -CRM, voir ci-dessous). J'ai ensuite déterminé précisément la fenêtre temporelle d'activité des CRM précoce et tardif grâce à la mise au point d'un nouveau gène rapporteur lacZ comportant un intron exogène qui permet de détecter les transcrits primaires du transgène. L'ensemble a mis en évidence un mécanisme de relais entre les CRM précoce et tardif de *col* qui se produit au stade PC et est requis pour le maintien de l'expression de Col et l'identité du muscle DA3. Ce « mécanisme de passage de témoin » est dépendant de l'accumulation de Col et de l'activité des protéines homéotiques.

1. Caractérisation du E^{col} -CRM de *col*

L'étude de la régulation transcriptionnelle du FTi Eve a permis d'identifier le CRM responsable de la transcription d'*eve* dans des myoblastes à l'origine de deux muscles dorsaux et de cellules péricardiaques (voir introduction). Ce CRM intègre une information tissulaire (FTs Twi et Tin) et une information de position médiée par les voies de signalisation Wg, Dpp et RAS/MAPK (FTs dTCF, Mad et Pnt) (Halfon et al., 2000). La

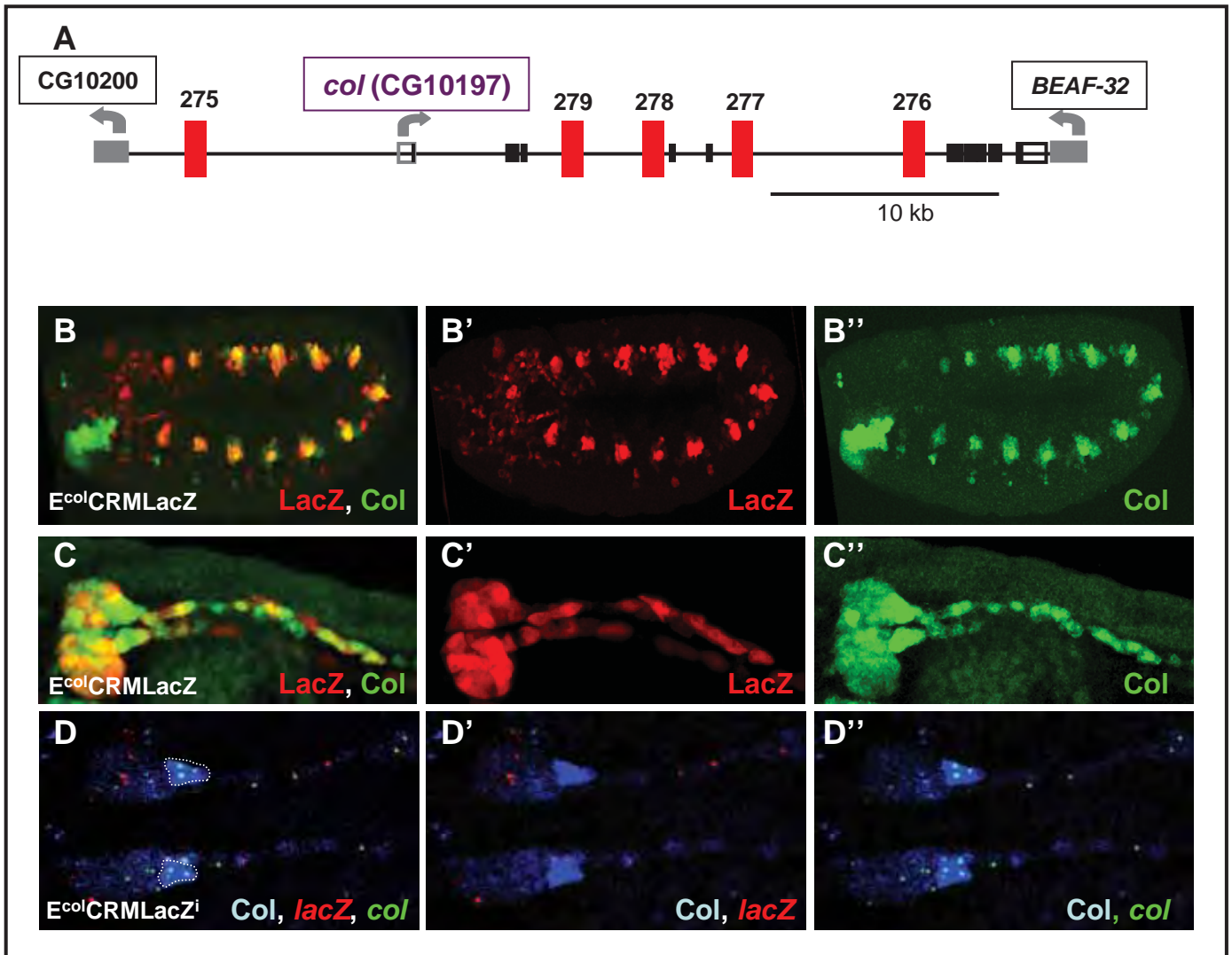


Figure 24. Identification du CRM précoce de collier ($E^{col}CRM$). **A.** Représentation schématique de la séquence génomique de *collier*. Les régions codantes et les régions 3' et 5' UTR sont représentés respectivement par des boîtes noires et blanches. Les boîtes rouges sont les CRM prédits par *Codefinder* (Philippakis et al., 2006) et sont numérotés de 275 à 279. **B-C''.** Immunomarquage anti-Collier (vert) et anti-LacZ (rouge) sur des embryons $E^{col}CRM$ -LacZ aux stades 10 (B-B'') et 16 (C-C''). **B-B''.** Le transgène $E^{col}CRMLacZ$ reproduit le patron d'expression de Col dans les groupes promusculaires Col. **C-C''.** $E^{col}CRM$ -LacZ est détecté dans la glande lymphatique. **D-D''.** Double hybridation in situ *lacZ*, *col* et immunomarquage Col sur les lignées $E^{col}-CRM/lacZ^1$. La transcription du *lacZ* n'est pas détectée dans le PSC (entouré dans D).

comparaison des transcriptomes de myoblastes isolés à partir d'embryons mutants ou sur-exprimant chacun de ces 5 FT a suggéré que cette combinaison de FT régulait l'expression d'autres FTi que Eve, dont *col* (Estrada et al., 2006). Afin d'identifier les CRM correspondants, l'équipe de Martha Bulyk a développé une approche bio-informatique appelée « Code Finder » qui permet d'identifier des séquences ADN enrichies en sites de fixation pour au moins 4 des 5 FT de cette combinaison (Philippakis et al., 2006). 5 séquences, prédictions 275-279 (Figure 24 A), situées soit en amont soit dans les introns de *col* répondent à ce critère d'enrichissement et sont donc des CRM possibles. J'ai testé l'activité de ces séquences *in vivo*, par la technique de transgénèse site-spécifique AttB/AttP (Bischof et al., 2007). Un seul des 5 CRM prédits, le CRM 276, a la capacité d'activer le gène rapporteur *lacZ* dans le groupe promusculaire Col (Figure 24 B-B"). Les autres « CRM » que j'ai testés (277, 278, 279), comme le 275 testé indépendamment par L. Dubois (2007), ne sont pas actifs dans le mésoderme. L'analyse détaillée de l'activité du CRM 276 maintenant appelé E^{col}-CRM a aussi révélé une expression dans la glande lymphatique embryonnaire, un autre site d'expression de *col* endogène (Figure 24 C-C"). J'ai ensuite recherché dans la séquence nucléotidique du E^{col}-CRM la présence de sites de liaisons pour les FT utilisés pour les prédictions bio-informatiques (Article 1 Fig S2). J'ai identifié six sites de liaison pour le FT Twi et deux pour Tin. Ces sites sont situés dans des fragments validés *in vivo* par immunoprécipitation de la chromatine (Sandmann et al., 2007). Par ailleurs un site Ets, et un site Mad, sont aussi présents dans le E^{col}-CRM. Le fragment 276 testé initialement a une longueur de 2 kb, Mathilde De Taffin a continué la caractérisation de ce fragment et a montré qu'un fragment de 1kb reproduit l'expression de Col dans le groupe promusculaire, mais pas dans la glande lymphatique (rapport M2R, M.De Taffin 2009).

2. Fenêtres temporelles d'activité des CRM

L'approche généralement utilisée pour tester *in vivo* l'activité d'un CRM est de le placer en amont d'un gène rapporteur codant pour une protéine GFP ou LacZ. Cependant, la stabilité de ces protéines constitue un handicap si on veut déterminer les fenêtres temporelles d'activité. Afin de remédier à ce problème, j'ai choisi d'insérer un intron exogène dans le gène rapporteur *lacZ*, permettant de détecter les transcrits primaires et donc les fenêtres d'activation et de répression de ces CRM avec précision. Ce nouveau gène rapporteur est appelé *lacZ*. Le choix de l'intron devait répondre à plusieurs critères: 1. Ne pas présenter de similitude de séquence significative avec le génome de *D. melanogaster*. 2. Posséder une taille compatible avec une hybridation *in situ* efficace (Nous avons opté pour environ 2kb). 3. être épissé efficacement, sans biais tissulaire. Nous avons donc choisi l'intron 1 du gène *βtubulin56D* de *D. virilis* qui est épissé efficacement dans tous les tissus de *D. virilis*

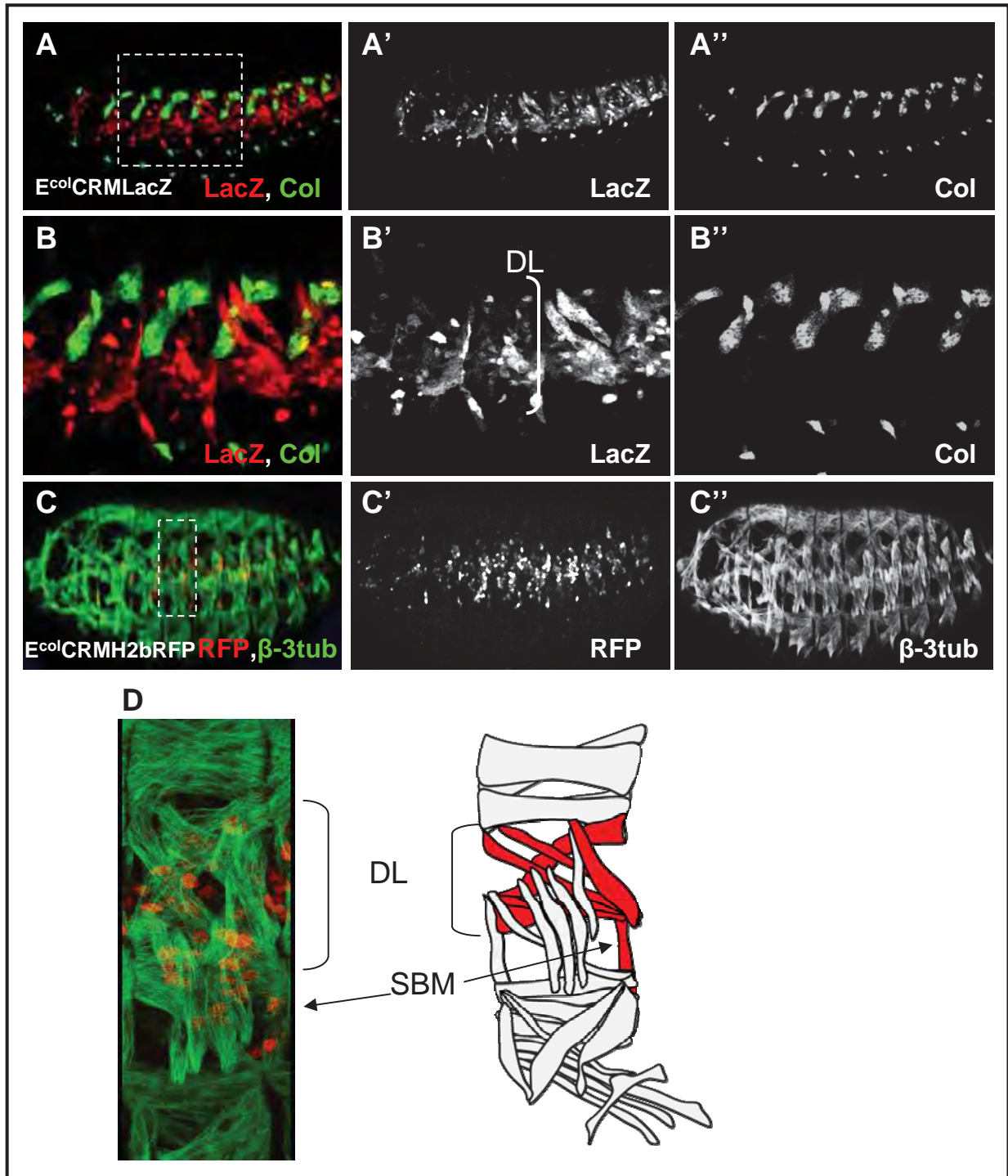


Figure 25. Devenir des FCM issus du groupe promusculaire Col. A-B''. Double immunomarquage Col+LacZ sur des embryons au stade 16 issus de lignées $E^{col}CRMLacZ$. Le LacZ perdure dans les muscle dorso-latéraux (DL). **C-C''.** Double immunomarquage RFP et β 3tubulin sur des embryons $E^{col}CRMH2bRFP$ au stade 16. **D.** Droite, grossissement du segment encadré dans C. Gauche, schéma récapitulatif des muscles ayant incorporé des FCM issus du groupe promusculaire Col (rouge).

(<http://flybase.org/reports/FBgn0207606.html>). J'ai donc généré des constructions plaçant *LacZ* sous le contrôle des E^{col} -CRM et L^{col} -CRM. J'ai vérifié que l'intron *β tubulin56D* est correctement épissé chez *D. melanogaster* en m'assurant que les patrons d'accumulation de la protéine LacZ sont identiques en présence ou absence de cet intron. Des doubles hybridations in situ, avec des sondes *col* et *lacZ*, sur les embryons E^{col} -CRM/*lacZ* ou L^{col} -CRM/*lacZ* m'ont ensuite permis de déterminer que le E^{col} -CRM est actif dans le groupe promusculaire à l'origine du PC DA3/DO5 et que le L^{col} -CRM est activé dans le PC DA3/DO5 et reste actif pendant la différenciation du muscle DA3. De manière intéressante, la période de recouvrement de l'activité de ces 2 CRM est limitée au PC. Ceci a permis de révéler un mécanisme dit de passage de témoin, l'activité du CRM précoce étant nécessaire à l'activation du CRM tardif. Indépendamment, l'analyse de l'expression du E^{col} -CRM/*lacZ* dans la glande lymphatique montre que son expression est exclue du PSC (Post Signaling Center : la niche hématopoïétique) à la fin de l'embryogenèse (Figure 24 D-D"). Ce résultat suggère que 2 CRM différents sont responsable de l'expression précoce de *col* dans les cellules précurseur de la glande lymphatique et le PSC.

En conclusion, mes résultats ont montré que le E^{col} -CRM intègre une information de position et une information tissulaire permettant son activation dans un groupe promusculaire dans tous les segments. La protéine Col accumulée active le L^{col} -CRM dans un des PC issus de ce groupe permettant le maintien de l'expression de Col dans le lignage DA3. Ce mécanisme de passage de témoin met en jeu une autorégulation directe de Col et une coopération des protéines Hox et du FT Nau. La restriction de ce processus à un seul des PC exprimant Col indique que d'autres acteurs, responsable de cette spécificité, restent à identifier.

B. Analyse du destin des FCM issus du groupe promusculaire Col

La caractérisation du E^{col} -CRM ma permis d'aborder une question restée en suspens depuis longtemps : les FCM issus d'un même groupe promusculaire sont ils réellement naïfs ou, au contraire programmés pour contribuer à la formation d'un sous ensemble spécifique de muscles ?

Afin de répondre à cette question, j'ai commencé par utiliser les lignées E^{col} -CRM-LacZ et exploiter la stabilité de la protéine LacZ pour suivre le destin des FCM issus du groupe promusculaire Col dans les embryons tardifs (Figure 25 A-B"). Mes premières observations ont montré que ces FCM contribuent principalement aux muscles dorsaux-latéraux (Figure 25 B-B"). Cependant, d'une manière surprenante, un double immuno-marquage Col/LacZ montre que le muscle DA3, le plus dorsal des muscles dorso-latéraux, est peu positif pour LacZ (Figure 25 B-B"), suggérant que les FCM incorporés dans la fibre DA3 pourraient être

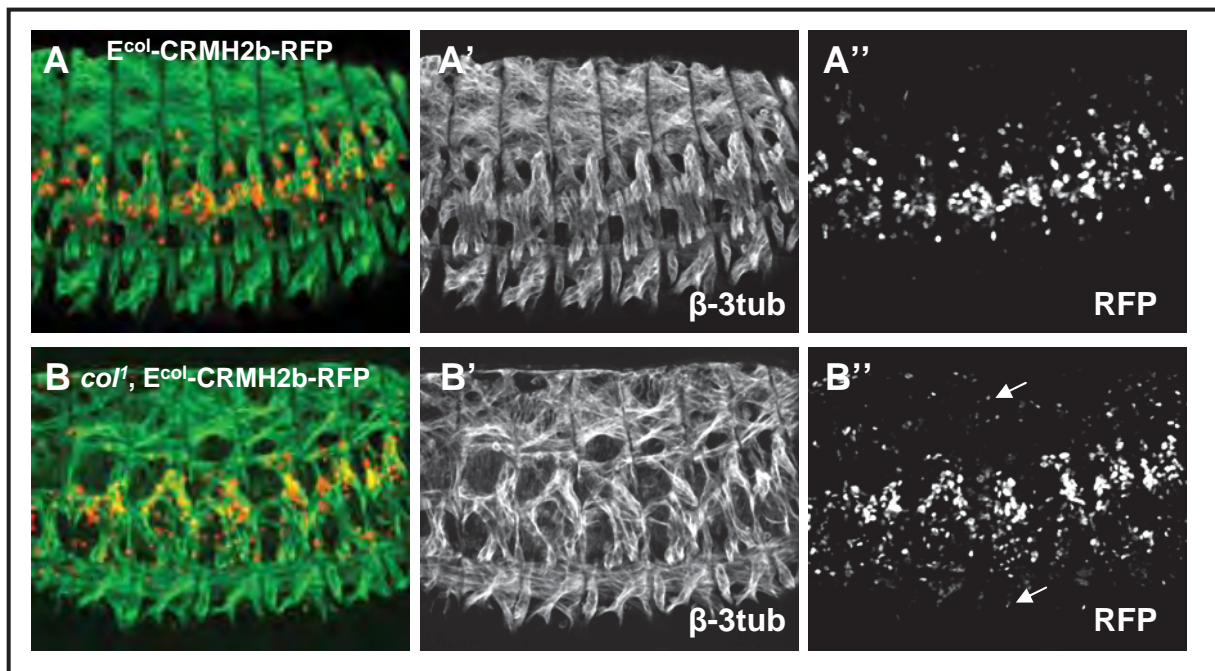


Figure 26. Impact de l'expression de *col* sur le devenir des FCM. Double immunomarquage RFP et β 3-tubulin sur des embryons $E^{col}CRMH2bRFP$ (A-A'') et col' ; $E^{col}CRMH2bRFP$ (B-B') au stade 16. Contrairement au contrôle (A-A''), dans un mutant col' (B-B'') les FCM (rouge) sont plus dispersés (flèches).

d'origine dorsale au groupe promusculaire Col. Afin de compléter cette analyse, j'ai exprimé une protéine de fusion Histone2b-RFP sous le contrôle du E^{col} -CRM (Figure 25 C-C"). L'expression transitoire de l'Histone2b-RFP permet de marquer spécifiquement et sélectivement les noyaux des cellules du groupe promusculaire Col et de les suivre dans des embryons tardifs. Une double immuno-coloration RFP et β 3-Tubulin (marqueur musculaire) a confirmé que les FCM du groupe promusculaire Col contribuent principalement à la formation des muscles dorso-latéraux, mais peu au muscle DA3 (Figure 25 C-C"). Ces FCM participent également à la formation du muscle SBM qui est un muscle latéral dérivant d'un progéniteur qui n'est pas issu du groupe promusculaire Col (Figure 25 D). L'ensemble de ces résultats suggèrent que plutôt qu'une information identitaire conférée par l'expression transitoire de Col, la répartition des FCM reflète leur position initiale dans le mésoderme. Afin d'étayer ou réfuter cette hypothèse, nous avons répété l'expérience dans des embryons mutants pour *col*. Le transgène rapporteur E^{col} -CRMH2B-RFP a été introduit dans le mutant nul *col*¹, (*col*¹; E^{col} -CRMH2b-RFP), dans lequel les muscles dorso-latéraux sont généralement présents mais présentent des défauts morphologiques (Enriquez et al., 2012) (Figure 26 A'-B'). Un double marquage RFP/ β 3-tubulin montre une plus grande dispersion des FCM du groupe promusculaire Col qui sont maintenant détectés dans des muscles plus ventraux et dorsaux (Figure 26 A''-B''). Une interprétation possible est que l'expression transitoire de Col restreint la dispersion des FCM, ce qui serait la première indication que les FCM ne sont pas strictement naïfs. Cependant, ces expériences méritent d'être répétées avec l'analyse d'autres marqueurs et dans d'autres contextes mutants (e.g., *tup*).

L'observation d'un faible marquage du muscle DA3 dans les expériences de lignage E^{col} -CRM-LacZ (Figure 25 B-B') nous a incités à examiner en plus de détail la répartition des noyaux lacZ⁺ et LacZ⁻ dans ce muscle. Pour cela, j'ai généré un transgène permettant de visualiser les contours cellulaires du muscle DA3. Ce transgène, L^{col} CRMMoeGFP, place une GFP membranaire (Moesin-GFP) sous le contrôle du L^{col} -CRM (Figure 27 A-A'). Un triple marquage GFP, RFP et Mef2 (pour visualiser les noyaux de tous les myoblastes) d'embryons L^{col} CRMMoeGFP ; E^{col} -CRMH2B-RFP permet de voir quels noyaux (Mef2) au sein du muscle DA3 (GFP-positif) sont issus du groupe promusculaire Col (RFP-positif) ou d'un autre groupe promusculaire (RFP-négatif) (Figure 27 B-B'). Les résultats indiquent qu'environ la moitié des noyaux du muscle DA3 est RFP-positif (Figure 27 B-B' et D). La même expérience a été réalisée pour le muscle dorso-latéral DT1 dont le PC est aussi issu du groupe promusculaire Col. Dans ce cas, nous avons utilisé un transgène rapporteur S59-GFP (Figure 27 C-C'). A nouveau, une moitié environ des noyaux du muscle DT1 sont marqué par la RFP (Figure 4 C-C' et D). Ces données, générées en collaboration avec Laëtitia Bataillé, nous conduisent à conclure que les FCM incorporés dans les muscles ne

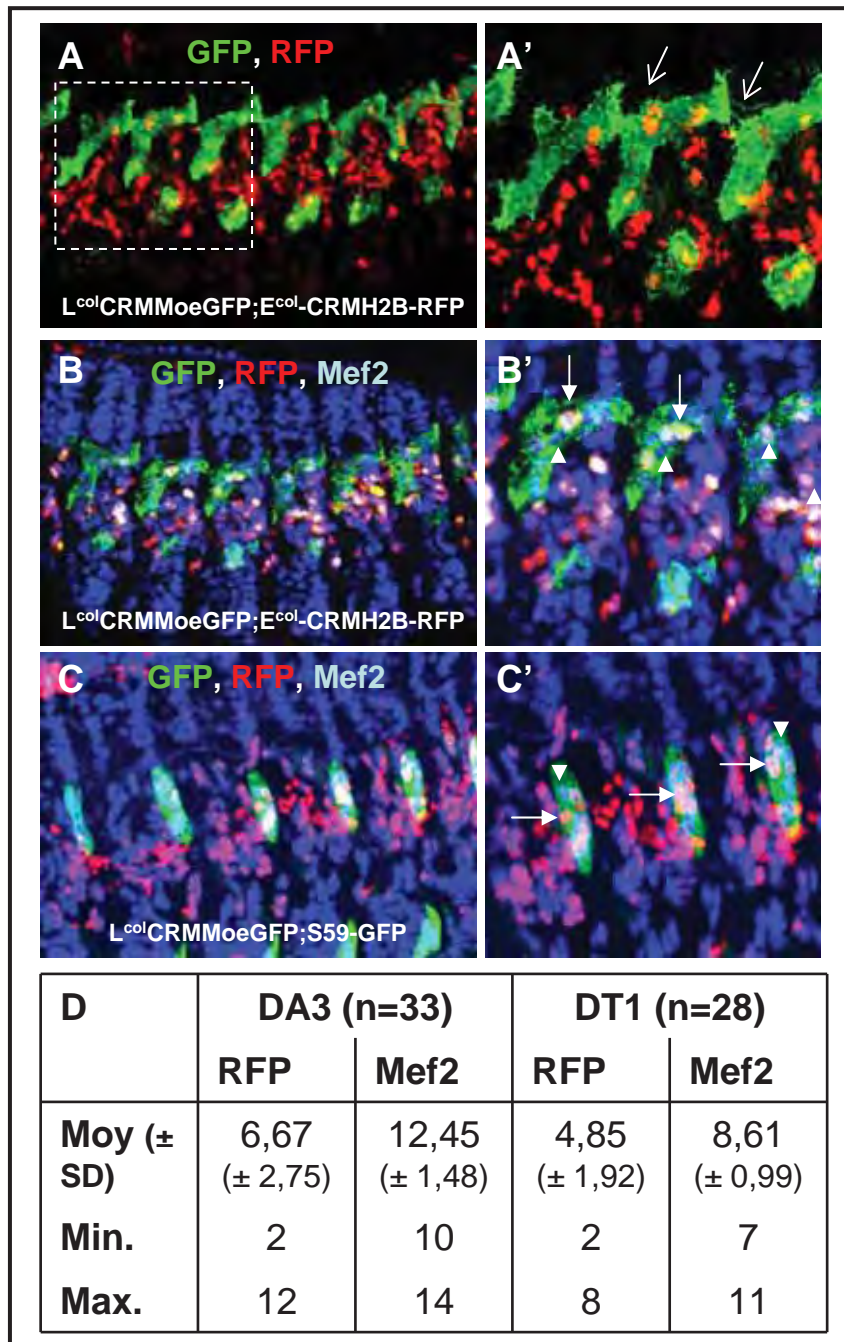


Figure 27. Contribution des FCM du groupe promusculaire Col aux muscles DA3 et DT1. A-A'. Immunomarquage GFP et RFP sur embryons $E^{col}CRMH2bRFP$; $L^{col}CRMMoeGFP$ au stade 15. **A'**. Agrandissement des 3 segments indiqué dans A. **A'**. Le transgène $L^{col}CRMMoeGFP$ marque le contour du muscle DA3 et les myopodes (voir flèches). **B-C'**. Triple immunomarquage GFP, RFP, Mef2 sur les embryons $E^{col}CRMH2BRFP$; $L^{col}CRMMoeGFP$ (B-B') et $E^{col}CRMH2BRFP$; S59GFP (C-C') au stade 15. Les muscles DA3 et DT1 contiennent soit des noyaux doublement marqués RFP et Mef2 (Flèche B' et C') soit des noyaux marqués uniquement avec Mef2 (pointe de flèche B' et C'). **D.** Comptage de noyaux RFP/Mef2 contribuant aux muscles DA3 et DT1.

sont pas tous issus du même groupe promusculaire et ne semblent donc pas avoir une même identité.

Multi-step control of muscle diversity by Hox proteins in the *Drosophila* embryo

Jonathan Enriquez¹, Hadi Boukhatmi¹, Laurence Dubois¹, Anthony A. Philippakis², Martha L. Bulyk², Alan M. Michelson³, Michèle Crozatier¹ and Alain Vincent^{1,*}

SUMMARY

Hox transcription factors control many aspects of animal morphogenetic diversity. The segmental pattern of *Drosophila* larval muscles shows stereotyped variations along the anteroposterior body axis. Each muscle is seeded by a founder cell and the properties specific to each muscle reflect the expression by each founder cell of a specific combination of 'identity' transcription factors. Founder cells originate from asymmetric division of progenitor cells specified at fixed positions. Using the dorsal DA3 muscle lineage as a paradigm, we show here that Hox proteins play a decisive role in establishing the pattern of *Drosophila* muscles by controlling the expression of identity transcription factors, such as Nautilus and Collier (Col), at the progenitor stage. High-resolution analysis, using newly designed intron-containing reporter genes to detect primary transcripts, shows that the progenitor stage is the key step at which segment-specific information carried by Hox proteins is superimposed on intrasegmental positional information. Differential control of *col* transcription by the Antennapedia and Ultrabithorax/Abdominal-A paralogs is mediated by separate cis-regulatory modules (CRMs). Hox proteins also control the segment-specific number of myoblasts allocated to the DA3 muscle. We conclude that Hox proteins both regulate and contribute to the combinatorial code of transcription factors that specify muscle identity and act at several steps during the muscle-specification process to generate muscle diversity.

KEY WORDS: Hox proteins, Cis-regulatory modules, Collier (Knot)/EBF, Nautilus/MyoD, Myogenesis, *Drosophila*

INTRODUCTION

Anatomy drawings illustrate the stereotyped patterns of skeletal muscles that are essential for coordinated movements. Each muscle has its own name/identity, reflecting its specific properties and function. The genetic and molecular bases of muscle identity remain, however, largely unknown. The rather simple pattern of *Drosophila* embryonic/larval skeletal muscles makes it an ideal model with which to study this process (Bate 1990; Bate 1993). Every (hemi)segment of the *Drosophila* larva contains ~30 different somatic muscles, each composed of a single multinucleate fiber. Whereas all muscles express the same myogenic program (e.g. of sarcomeric proteins), each has its own identity characterized by its specific position and orientation with respect to the dorsoventral (D/V) and anteroposterior (A/P) axes, size, sites of attachment on the epidermis and innervation (Bate 1993; Baylies et al., 1998; Knirr et al., 1999). Each syncytial muscle fiber is seeded by a 'founder' myoblast (founder cell, FC). FCs possess the unique property of being able to undergo multiple rounds of fusion with fusion-competent myoblasts (FCMs). The current view is that muscle identity reflects the expression by each FC of a specific combination of 'identity' transcription factors (iTFs), with *Apterous*, *Collier* (Col; *Knot*), *Even-skipped* (*Eve*), *Krüppel*, *Ladybird*, *Nautilus* (Nau) and *Slouch* (*Slou*; *S59*) being among the best characterized (Bourgouin

et al., 1992; Ruiz-Gomez et al., 1997; Jagla et al., 1998; Crozatier and Vincent, 1999; Knirr et al., 1999; Balagopalan et al., 2001; Fujioka et al., 2005; Dubois et al., 2007). FCs originate from the asymmetric division of progenitor cells, which are themselves singled out from promuscular clusters (equivalence groups) by Notch (N)-mediated lateral inhibition (Carmena et al., 1995; Ruiz Gomez and Bate, 1997). Each muscle progenitor/FC is specified at a specific A/P and D/V position within the somatic mesoderm. This position determines the final location of the muscle(s) issued from this progenitor.

The abdominal A2-A7 segments present the same final muscle pattern. The patterns of the thoracic T2-T3 and abdominal A1 segments are variations on this pattern, whereas the first thoracic segment (T1) and eighth abdominal segment (A8) present fewer and more diversified muscles (Bate and Rushton, 1993). Although it is well established that the Hox transcription factors are major regulators for patterning of the animal body, our understanding of how each Hox protein specifies distinct morphological features and cellular identities within each body part is still fragmentary (Hueber and Lohmann, 2008; Mann et al., 2009). The segment-specific aspects of the *Drosophila* musculature represent an interesting paradigm with which to address this question (Greig and Akam, 1993; Michelson, 1994). Based on the expression pattern of Nau, which is the *Drosophila* ortholog of the mammalian bHLH myogenic factor MyoD, in different Hox conditions, Michelson (Michelson, 1994) suggested that segmental differences in the somatic muscle pattern reflect the regulation of muscle iTFs by Hox proteins. Subsequent studies of *apterous* expression in thoracic lateral muscles supported this notion and suggested that Hox activity could control the segment-specific variation in the number of myoblasts allocated to a specific group of muscles (Capovilla et al., 2001). As a whole, however, when and how homeotic genes act during the muscle-specification program remain to be established.

¹Centre de Biologie du Développement, UMR 5547 CNRS/UPS, IFR 109 Institut d'Exploration Fonctionnelle des Génomes, 118 route de Narbonne, 31062 Toulouse cedex 9, France. ²Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA. ³Laboratory of Developmental Systems Biology, National Heart, Lung and Blood Institute National Institutes of Health, 31 Center Drive, Bethesda, MD 20892, USA.

*Author for correspondence (vincent@cict.fr)

Here, we addressed this question, focusing on the dorsal acute DA3 muscle lineage, which depends upon the combinatorial activity of Nau and Col, which is the *Drosophila* ortholog of mammalian early B-cell factors (EBFs) (Michelson et al., 1990; Keller et al., 1997; Balagopalan et al., 2001). Our data show that Hox activity is superimposed on mesoderm-specific and positional information at the progenitor stage to establish the pattern of somatic muscles and acts at several independent steps during the muscle-specification process. Very few studies have addressed the role of Hox genes in controlling the fate of skeletal muscle precursors in vertebrates (Alvares et al., 2003). Our data provide a new framework for studying the integration of Hox information into the generic process of myogenesis and the generation of muscle diversity.

MATERIALS AND METHODS

Drosophila genetics

The strains used were *w¹¹⁸* as wild-type (wt) reference (Bloomington Stock Center, IN, USA), *Antp¹* (Abbott and Kaufman, 1986), *Ubx¹* (Bloomington Stock Center), *Antp²⁵*; *Ubx¹* (Ernesto Sanchez-Herrero, Madrid, Spain), *rP298-lacZ* (Nose et al., 1998), *UAS-Antp* (Heuer et al., 1995), *UAS-Ubx*, *UAS-abdA* (Michelson, 1994), *UAS-col* (Vervoort et al., 1999), *UAS-nau* (Keller et al., 1997), *24B-Gal4* (Brand and Perrimon, 1993), *rP298-Gal4* (Menon and Chia, 2001) and *sns-Gal4* (Kocherlakota et al., 2008). The mutant strains were balanced over marked (TM3 *twist-lacZ*) chromosomes. All Gal4-UAS crosses were performed at 25°C.

Plasmid constructions and transgenic lines

A *Bam*HI-*Nae*I DNA fragment containing the *attB* site from the *pUASTattB* vector (Bischof et al., 2007) was substituted for the P-element-containing *Nsi*I-*Avr*II fragment in the H-Pelican *lacZ* transformation vector (Barolo et al., 2000) (GenBank AF242361.2) to generate *attB-inslacZ*, in which *lacZ* is under the control of the minimal *hsp70* promoter and flanked by *gypsy* insulator sequences. The 4_0.9, 2.6_0.9 and CRM276 *col* genomic fragments were inserted into the *attB-inslacZ* and/or *attB-inslacZ²* vectors. Each *attB* construct was inserted at position 49D on the second chromosome by injection into *nosC31^{NLS}*; *Zh8* embryos (Bischof et al., 2007).

Construction of an intron-containing *lacZ²* reporter gene

The unique intron of the *D. virilis* β -*tubulin56D* (*βtub56D*) gene (Dvir|GJ20466; FlyBase ID FBgn0207606) was inserted into the *lacZ* coding region between Asp119 and Val120 by standard PCR-based cloning. The 5' to 3' sequences of the created splice junctions are as follows: 5'-CGTGAGGT-AGGTCTCG-3'; the intron sequence is underlined. This resulted in a single C-to-G nucleotide change (bold) in the *lacZ* coding sequence, resulting in an Asp235-to-Glu substitution. The resulting intron-containing *lacZ* gene, denoted *lacZ²*, was used to generate *attB-inslacZ²* from *attB-inslacZ*. Mutagenesis of the consensus AbdA/Ubx binding site TAATTA (Ekker et al., 1994) to TGGGGA was by PCR.

Immunohistochemical staining and in situ hybridization

Embryos were fixed and processed for antibody staining and/or in situ hybridization as described (Crozatier et al., 1996). Primary antibodies were: mouse anti-Col (1/100) (Dubois et al., 2007); rabbit anti-Mef2 and anti-Nau (1/100) (provided by Eileen Furlong, Heidelberg, Germany and Bruce Paterson, Bethesda, MD, USA, respectively); mouse anti- β -galactosidase (Promega, 1/1000); and mouse anti-Ubx and anti-Antp (1/100, Developmental Studies Hybridoma Bank, IA, USA). Secondary antibodies were: Alexa Fluor 488-conjugated goat anti-rabbit and goat anti-mouse; Alexa Fluor 555-conjugated goat anti-rabbit and goat anti-mouse; Alexa Fluor 647-conjugated goat anti-mouse (all Molecular Probes, 1/300); and biotinylated goat anti-mouse (Vector Laboratories, 1/1000). Double fluorescence in situ hybridization and immunostaining using intronic probes and Col antibodies were as described (Dubois, 2007).

Sequence alignments and transcription factor binding sites

Pairwise sequence alignments of *col* upstream and CRM276 sequences from various *Drosophila* species (<http://flybase.org/cgi-bin/gbrowse/>) were performed using NCBI-BLAST (bl2seq), Genome Browser (University of

California, Santa Cruz, CA, USA) and Evoprimer (NINDS, NIH, Bethesda, MD, USA) and manually edited by eye. The search for individual binding sites for transcription factors made use of Cis Analyst (<http://rana.lbl.gov/cis-analyst/>) and FlyEnhancer (<http://opengenomics.org/>) and manual inspection based on the matrices for binding sites of Twi, Tin, dTCF, Mad and Pnt (Philippakis et al., 2006). Access to the Tin and Twi in vivo binding sites (Sandmann et al., 2006; Sandmann et al., 2007; Zinzen et al., 2009) was via the E. Furlong lab site (<http://furlonglab.embl.de/data/>).

RESULTS

Segment-specific properties of the DA3 muscle

Col expression in stage 15 embryos showed that the DA3 muscle forms in the thoracic (T) T2-T3 segments and abdominal (A) A1-A7 segments, but not in the T1 segment and was thinner in T2-T3 than in A1-A7 (Fig. 1A). Since the size of *Drosophila* larval muscles correlates with their number of nuclei (Demontis and Perrimon, 2009), we counted the number of nuclei in the DA3 muscle. Fewer nuclei were present in T2 and T3 (six on average) than in A1-A7 (eight on average) (Fig. 1B), showing that this number is segment specific. Col expression in the somatic mesoderm is first detected at embryonic stage 10, in a cluster of cells at the same dorsal position in all trunk segments, including T1 (Fig. 1C). This cluster gives rise to the DA3/DO5 progenitor in T2 and T3 and to the DA3/DO5 and DO4/DT1 progenitors in A1-A7 (Fig. 1D). Following asymmetric division of the DA3/DO5 progenitor, *col* transcription is maintained in the DA3 FC but is repressed in the sibling DO5 FC, a repression mediated by N; it is also not maintained in the DO4 and DT1 FCs (Crozatier and Vincent, 1999). Nau is expressed in the same progenitors as Col (Fig. 1C,D). However, Nau and Col co-expression is only transient. *col* transcription is maintained in the DA3 FC and is activated in the nucleus of each FCM incorporated into the growing DA3 myofiber, whereas *nau* is transiently transcribed in the DA3 lineage and is activated in FCM nuclei incorporated into the DO5 myofiber (Dubois et al., 2007). This leads to a specific accumulation of Col and Nau in the DA3 and DO5 muscles, respectively (Fig. 1E). Thus, the progenitor/FC stage is the specific step in the DA3/DO5 lineage at which Nau and Col are expressed together. One Col- and Nau-expressing progenitor is found in T2 and T3, two progenitors in A1-A7 and none in T1, correlating with the final muscle pattern (Bate and Rushton, 1993; Crozatier and Vincent, 1999). Transient co-transcription of *col* and *nau* was nevertheless observed in a cell issued from the Col-expressing cluster in T1, although at a very low level compared with other segments (Fig. 1F). This indicates that a positive input required to upregulate the expression of these two iTFs in progenitor cells is missing in T1. In summary, transient expression of Nau and Col at the progenitor stage is segment specific and foreshadows the segment-specific formation of DA3/DO5 and DO4/DT1 muscles (Fig. 1G).

Two independent cis-regulatory modules (CRMs) control the different phases of *col* transcription in the DA3 muscle

A *lacZ* reporter gene containing 4 kb of the *col* upstream cis-regulatory region (*4_0.9-lacZ*; Fig. 2A) reproduces *col* expression in the DA3 FC and muscle, but not in promuscular clusters (Dubois et al., 2007). An enhancer that drives Eve expression in dorsal promuscular clusters at the origin of the DA1 and DO2 muscles (see Fig. S1 in the supplementary material) has been described (Halfon et al., 2000; Knirr and Frasch, 2001; Speicher et al., 2008). This enhancer integrates positional information issued from the ectoderm with mesoderm-intrinsic information via the binding of five different

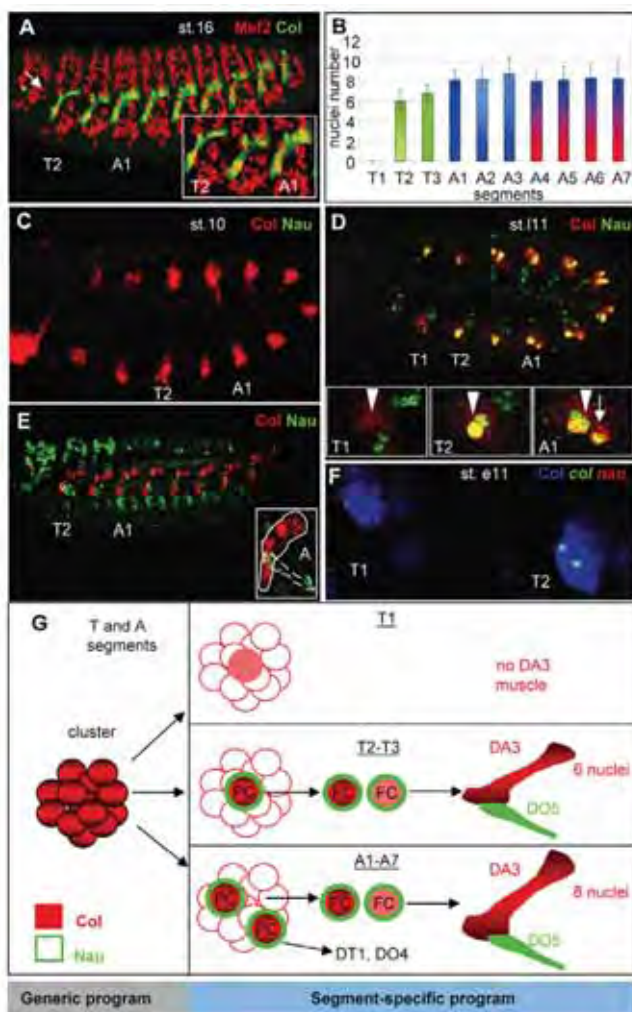


Fig. 1. Segment-specific formation and properties of the DA3 muscle. (A) *Mef2* (red) and *Col* (green) expression in a stage (st.) 16 *Drosophila* embryo, showing the entire muscle pattern and DA3 muscle, respectively. The arrow indicates the absence of DA3 muscle in the T1 segment. The inset shows an enlarged view of the T2, T3 and A1 segments. Note that the DA3 muscle is smaller in T than in A segments. (B) The segment-specific number of nuclei in the DA3 muscle (number of segments counted=30). The mesodermal domain of expression of *Antp* is in green, *Ubx* in blue and *AbdA* in red (Bate, 1993) (see Fig. S3 in the supplementary material). (C-E) Comparison of expression of *Col* (red) and *Nau* (green) in stage 10 (C), late stage 11 (st.11) (D) and stage 15 (E) embryos. Insets in D show the *Col*-expressing cluster in T1 (arrowhead), the DA3/DO5 progenitor in T2, and DA3/DO5 (arrowhead) and the DT1/DO4 progenitors (arrow) in A1. The inset in E is an enlarged view of the abdominal DA3 (circled) and DO5 (dotted circle) muscles that express *Col* and *Nau*, respectively. (F) Double in situ hybridization for *col* (green) and *nau* (red) primary transcripts and immunostaining for *Col* (blue). The T1 and T2 segments of an early stage 11 embryo are shown. Two dots per nucleus reflect the two alleles. All embryos are oriented anterior to the left and dorsal to the top. (G) *Col* (red) and *Nau* (green) expression during the process of DA3 muscle formation in different trunk segments, highlighting the segment-specific aspects. PC, progenitor; FC, founder cell.

transcription factors: Mothers against *dpp* (*Mad*), dTCF (*Pangolin*), Pointed (*Pnt*), Twist (*Twi*) and Tinman (*Tin*) (Carmena et al., 1998; Halfon et al., 2000; Knirr and Frasch, 2001). *Mad* and dTCF are the

downstream effectors of Decapentaplegic (*Dpp*) and Wingless (*Wg*) signaling, respectively, with these ectodermal signals defining a positional grid within each segment. *Pnt* is an effector of the Ras/MAPK pathway that contributes to the specification of equivalence groups within the mesoderm. *Twi* and *Tin* are mesoderm-specific transcription factors (Halfon et al., 2000).

By comparing gene expression profiles for myoblasts of different genotypes representing perturbations of the *Wg*, *Dpp*, *Ras* and *N* pathways, Estrada et al. identified ~160 genes that were possibly regulated similarly to *eve* in the mesoderm, including *col* (Estrada et al., 2006). Using an in silico search, with the ModuleFinder protocol (Philippakis et al., 2006), we identified five non-coding sequences that were selectively enriched for combinations of *Tin*, *Twi*, *Pnt*, *Mad* and dTCF binding sites within the *col* gene. Each was individually tested by transgenic reporter analysis (data not shown). The intron-located sequence with the highest prediction score, despite the absence of a conserved *Mad* binding site (see Fig. S2 in the supplementary material), which is referred to below as CRM276 (Fig. 2A), drove *lacZ* expression in promuscular clusters. Double staining of stage 10-11 embryos showed a precise overlap between CRM276-*lacZ* and *Col* expression (data not shown).

The stability of *lacZ* mRNA and β -gal protein prevents, however, a precise determination of temporal aspects of the activity of CRMs, a central aspect of our study. To circumvent this problem, we introduced the *Drosophila virilis* *β tub56D* intron into the *lacZ* coding region (*lacZⁱ*, see Materials and methods) and generated reporters in which *lacZⁱ* is placed under control of the CRM276 and 4_0.9 *col* fragments (4_0.9-*lacZⁱ* and CRM276-*lacZⁱ*, respectively; Fig. 2B,C). The patterns of *lacZ* expression were identical when driven by intron-devoid (not shown) and intron-containing transgenes (Fig. 2B,C), demonstrating that the intron is efficiently spliced. In situ hybridization with a mixture of *col* and *lacZⁱ* intronic probes confirmed at the primary transcript level that CRM276 precisely reproduces *col* activation and integrates the same A/P and D/V positional information (Fig. 2D). CRM276-*lacZⁱ* transcription was subsequently restricted to the DA3/DO5 progenitor, identical to endogenous *col*, indicating that it is subject to repression by *N* (Fig. 2E). Unlike *col*, however, CRM276-*lacZⁱ* transcription was very weak in progenitors and was not detected beyond that stage (Fig. 2F,G). Thus, CRM276 is an early mesodermal enhancer that imparts positional and mesodermal information to *col* activation in a specific promuscular cluster. Conversely, the 4_0.9 CRM was only active from the progenitor stage and was activated in the nuclei of FCMs that have fused with the DA3 FC (Fig. 2H-K). CRM276 and 4_0.9 CRM together account for all aspects of *col* transcription, including its repression by *N* signaling, during the successive steps of selection and asymmetric division of the DA3/DO5 progenitor (Fig. 2I).

The precise determination of temporal windows of activity established that the activity of the position-specific CRM276 is transient and is relayed by another 4_0.9 CRM in progenitor cells. As it does not operate in T1, this relay is segment specific, providing evidence that it is at the progenitor stage that segment-specific information superimposes on positional information provided by segmentation and D/V patterning genes (Fig. 2L).

Hox proteins are required for DA3 muscle specification and allocate its number of nuclei

The register of mesodermal expression of the thoracic Hox proteins *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*) and *Abdominal-A* (*AbdA*) is schematized in Fig. 3A (Bate, 1993) (see Fig. S3 in the supplementary material; data not shown) (Bate and Rushton, 1993). Comparison between

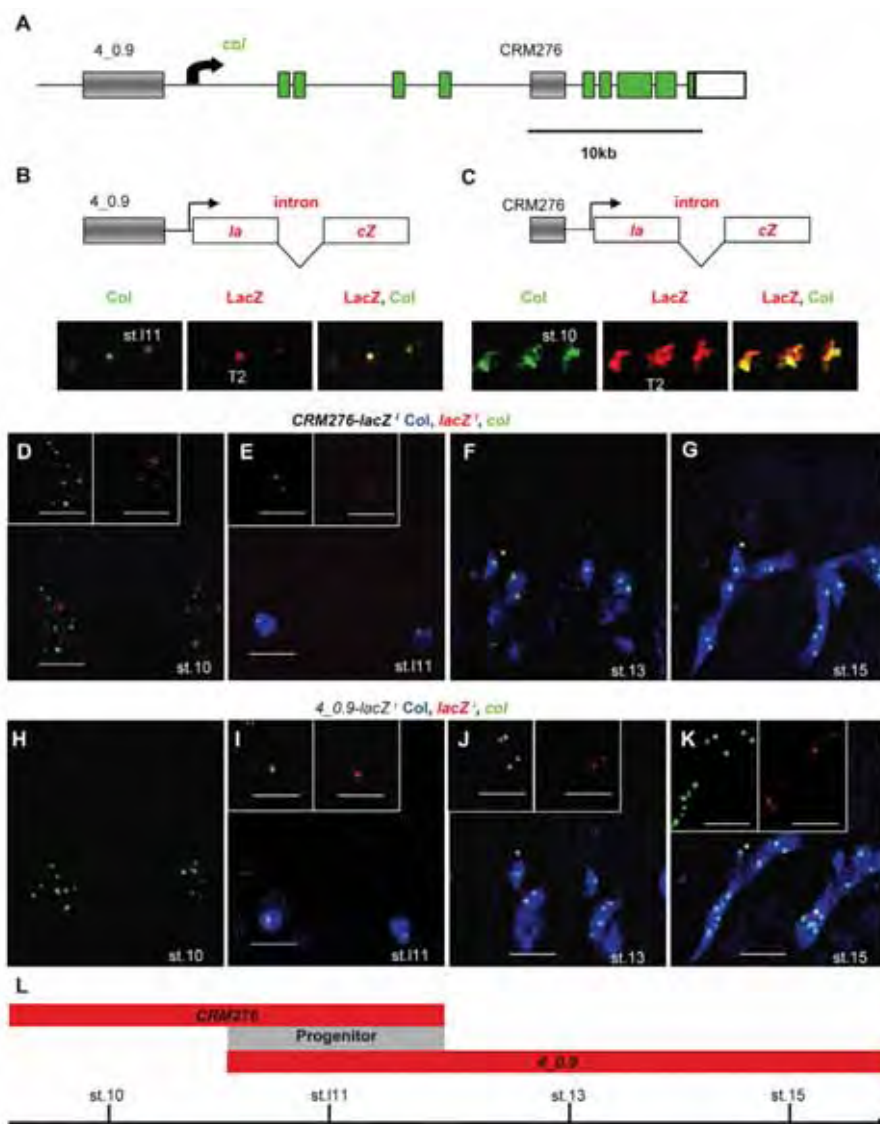


Fig. 2. Two separate CRMs control *col* transcription at different steps during DA3 muscle specification. (A) The *Drosophila col* genomic region. Coding exons are in green. The transcription start site is indicated by an arrow. Gray boxes represent the muscle CRMs 4_0.9 and 276. (B,C) The 4_0.9-*lacZ*¹ and CRM276-*lacZ*¹ transgenes, in which the *lacZ* coding sequence (red) is split by the *Drosophila virilis* *βtub56D* intron. The transcription start site is indicated by an arrow. Beneath are shown (B) stage 11 4_0.9-*lacZ*¹ and (C) stage 10 CRM276-*lacZ*¹ embryos immunostained for Col (green) and β-galactosidase (red, LacZ), or both, as indicated. (D-K) Double in situ hybridization for *col* (green) and *lacZ* (red) primary transcripts and Col immunostaining (blue) on (D-G) CRM276-*lacZ*¹ and (H-K) 4_0.9-*lacZ*¹ embryos. Developmental stages are indicated in each panel. The T2 and T3 segments are shown. Insets show only *lacZ*¹ or *col* transcripts. Note that the red and green dots corresponding to *lacZ*¹ and *col*, respectively, do not always overlap because the two genes are located at separate chromosomal positions, with two dots per gene per nucleus reflecting the two alleles. The asterisk (F,G,J,K) indicates a Col-expressing multidendritic (md) neuron. CRM276-*lacZ*¹ transcription is only active in promuscular clusters and in the corresponding progenitor (D,E). 4_0.9-*lacZ*¹ transcription starts in the DA3/DO5 progenitor (I) and is active in DA3 nuclei until stage 15 (J,K). (L) The temporal windows of activity of CRM276 and 4_0.9, which specifically overlap at the progenitor stage (gray).

4_0.9-*lacZ* and either Scr, Antp or Ubx expression in stage 11 embryos (see Fig. S3A,B in the supplementary material; data not shown) confirmed that the Col-expressing progenitor in T2 and T3 expresses Antp and that the two *col*-positive progenitors in A1-A7 express Ubx. Overlapping expression with the FC-specific marker, *rP298-lacZ* (Nose et al., 1998), in late stage 11 embryos showed that Antp and Ubx expression is maintained in T2-T3 and A1-A7 FCs, respectively (see Fig. S3C,D in the supplementary material). We looked at the effect of overexpressing either Antp or Ubx (Fig. 3B,C) or AbdA (not shown) throughout the mesoderm on the specification of the DA3 muscle. Pan-mesodermal expression of either Hox protein (Brand and Perrimon, 1993; Michelson, 1994) resulted in the formation of an ectopic Col-expressing muscle in T1, at the same position as in other segments (Fig. 3B,C). Thus, providing Antp, Ubx or AbdA activity is sufficient to convert the Col-expressing promuscular cluster in T1 into a DA3 muscle. We observed, however, that the size of this ectopic DA3 muscle varied depending on which Hox protein was expressed. When Antp was overexpressed, the number of nuclei in this muscle was identical to that in wild-type (wt) T2 and T3 (Fig. 3D). However, when either Ubx or AbdA was overexpressed, this number was

converted to that in wt A segments (Fig. 3E and data not shown), a clear example of posterior prevalence (Duboule and Morata, 1994). These data show that Antp, Ubx and AbdA activities in the mesoderm control the formation of the DA3 muscle and regulate its number of nuclei.

Hox proteins control the number of muscle progenitors expressing Col and Nau

We then looked at when during the muscle-specification process expression of Col and Nau was modified by pan-mesodermal expression of Antp or Ubx. In stage 11 embryos, residual Col expression was detected in all segments in cells that do not become progenitors; high levels of Col and Nau co-localized in the DA3/DO5 progenitor in T2-T3 and in the DA3/DO5 and DO4/DT1 progenitors in A1-A7 (Fig. 1D,G; Fig. 4A). Pan-mesodermal expression of Antp resulted in high-level expression of Col and Nau in one dorsal progenitor in T1 (Fig. 4B), a pattern normally restricted to T2 and T3 (Fig. 4A). Pan-mesodermal expression of Ubx resulted in high-level expression of Col and Nau in two progenitors in all three T segments (Fig. 4C), a pattern typical of wt A1-A7 segments (see Fig. 1D). We conclude that

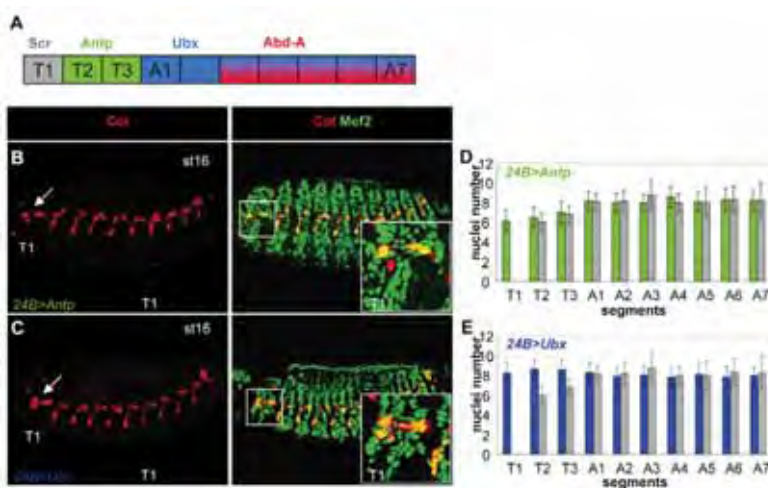


Fig. 3. Hox proteins are required for the formation of the DA3 muscle and allocate its number of nuclei. (A) A color-coded representation of Scr, Antp, Ubx and AbdA expression in the *Drosophila* trunk segments. (B,C) Col (red) and Mef2 (green) expression in stage 16 embryos expressing Antp (B) or Ubx (C) in all mesodermal cells (*24B-Gal4* driver). Col staining alone is shown on the left. The arrow points to the DA3-like muscle that forms in T1. (D,E) The number of nuclei per DA3 muscle in each T1 to A7 segment. The genotypes are shown at the top, with the Hox color code as in Fig. 1. The gray bars show wild-type (wt) nuclei numbers. Number of segments counted=30.

Hox activity in the mesoderm controls the segment-specific number of progenitors expressing Col and Nau that emerge from a pre-defined promuscular cluster.

In order to complement the Hox gain-of-function data, we followed Col expression and the DA3 muscle lineage in Hox loss-of-function mutant embryos, using null alleles (Fig. 4D-I). No DA3 muscle formed in T2 and T3 in *Antp* mutant embryos (Fig. 4E). In *Ubx* mutant embryos, the DA3 muscle pattern was similar to that of the wt (not shown), presumably owing to the derepression of Antp expression in the A1 and A2 segments (Hooper, 1986). Accordingly, we observed that *Antp*, *Ubx* double-mutant embryos lacked a DA3 muscle in all T and A1-A2 segments (Fig. 4F). Close examination of stage 11 *Antp* mutant embryos showed that one cell in each thoracic segment transiently accumulated more Col protein than the other cells of the cluster, similar to T1 in wt embryos. This transient, low-level Col accumulation confirms that the process of progenitor selection is initiated and prematurely aborts in the absence of Antp input (Fig. 4H). In *Antp*, *Ubx* double mutants, Col upregulation in progenitors was only observed in A3-A7 (Fig. 4I), further confirming that Hox-dependent upregulation and the final muscle pattern are linked (Fig. 4J).

Antp and Ubx/AbdA control *col* expression in muscle progenitors via distinct cis-regulatory elements

The loss of *col* expression in Hox mutant embryos and the pattern of *4_0.9-lacZ* expression together indicated that the *4_0.9* CRM mediates *col* regulation by Hox proteins (Figs 2 and 4; Fig. 5A). The observation that a reporter construct containing only 2.6 kb of *col* upstream DNA (*2.6_0.9-lacZ*) was only active in A progenitors (Fig. 5B,C) further raised the possibility that differential control by the Antp and Ubx/AbdA paralogs could involve different cis elements. To investigate this, we compared *2.6_0.9-lacZ* and *4_0.9-lacZ* expression in stage 11 embryos when either Antp or Ubx was expressed throughout the entire mesoderm. Unlike *col*, *2.6_0.9-lacZ* was not activated by Antp in T1, consistent with the fact that it is not expressed in T2 and T3 in wt embryos (Fig. 5E; see also Fig. 4). By contrast, *2.6_0.9-lacZ* was activated by Ubx (or AbdA, not shown) in two progenitors in each segment, including all three T segments, thereby reproducing the pattern of *col* expression under these conditions (Fig. 5F). These results showed that upregulation of *col* expression by Ubx/AbdA, and not Antp, involves cis elements present within the *2.6_0.9 col* upstream region. Conversely, *4_0.9-*

lacZ expression was upregulated by Antp in one progenitor in T1, therefore reproducing Col expression under the same conditions (Fig. 5D; Fig. 4B). These data show that Antp and Ubx/AbdA control *col* expression in muscle progenitors via distinct cis-regulatory elements (Fig. 5C).

Ubx/AbdA regulation of *col* transcription in muscle progenitors is direct

Since *2.6_0.9-lacZ* is only expressed in A segments, we asked whether its regulation by Ubx/AbdA is direct. An in silico search for high-affinity Hox binding sites [TAATTA (Ekkert et al., 1994; Affolter et al., 2008)] within the *2.6_0.9* CRM identified two sites that are conserved at the same relative position in several *Drosophila* species (Fig. 5C and see Fig. S4 in the supplementary material). Each of these sites, named Hox1 and Hox2, was individually mutated (TAAT to GGG) within the *2.6_0.9-lacZ* construct, giving rise to the *2.6_0.9^{Hox1}-lacZ* and *2.6_0.9^{Hox2}-lacZ* transgenes. *2.6_0.9^{Hox2}-lacZ* expression was no longer detected in progenitors (Fig. 5G), whereas the Hox1 mutation had little if any effect. Mutation of both Hox1 and Hox2 sites had no additional effect over the Hox2 mutation alone (data not shown), indicating that Hox2 is required for *col* regulation by Ubx/AbdA in abdominal muscle progenitors. Expression of *2.6_0.9^{Hox2}-lacZ* recovered later during DA3 muscle development (Fig. 5H), consistent with our previous finding that cis elements responsible for *col* activation in the DA3-recruited FCM are all contained within the *2.3_1.6* interval, which does not cover the Hox2 site (Dubois et al., 2007). We conclude that a single Hox binding site mediates the specific upregulation of *col* expression by Ubx/AbdA in A progenitors (Fig. 5C).

Forced expression of Nau plus Col bypasses the need for Hox activity in forming a DA3 muscle, but not in allocating its nuclei number

Our finding that Hox activity is required in progenitors for prolonged expression of Nau and Col and for implementation of the muscle differentiation process suggested that forced expression of these iTFs in progenitor cells might bypass the need for Hox activity. We therefore repeated the *24B-Gal4*×*UAS*-overexpression experiments, but using both Col and Nau in place of a Hox protein. To visualize the DA3 muscle, we used the *2.6_0.9-lacZ* reporter (Fig. 6A). As previously shown (Dubois et al., 2007), *24B-Gal4*-driven expression of Nau alone does not affect *2.6_0.9-lacZ* expression, whereas expression of Col results in strong activation in the DA2 and VL1-2

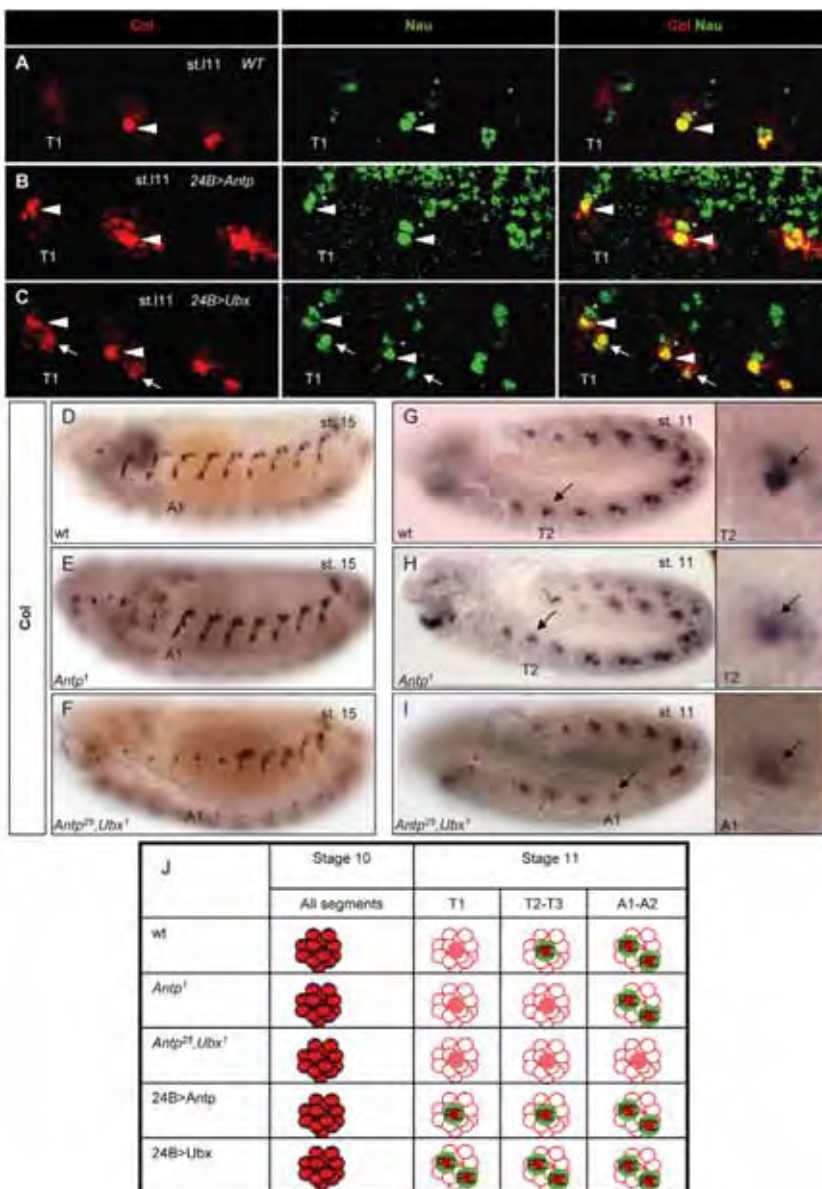


Fig. 4. Hox proteins control the number of muscle progenitors expressing Col and Nau. (A-C) Col (red) and Nau (green) expression in T segments of late stage 11 *Drosophila* embryos. Overlap (yellow) is shown on the right. (A) One progenitor expressing high levels of Col and Nau is found in wt T2 (arrowhead) and T3 and in T1 upon expression of Antp (B) or Ubx (C). (C) Ubx induces a second progenitor (arrow) to express high levels of Col and Nau in all three T segments, similar to wt A segments (see Fig. 1D). The asterisk indicates another muscle progenitor that expresses Nau, but not Col, in wt T segments and sometimes low levels of Col upon Hox overexpression. (D-I) Col expression in stage 15 (D-F) and 11 (G-I) embryos. The DA3 muscle is lacking in T2-T3 and T2-A2 in *Antp* (E) and *Antp*; *Ubx* (F) mutants, respectively, correlating with the absence of high-level Col-expressing progenitors (H,I). The remaining Col-positive nucleus corresponds to that of an md neuron, as shown by its expression of *pickpocket* (Crozier and Vincent, 2008). Enlarged views of the Col-expressing cluster/progenitor in T2 (G,H) and A1 (I) show that Col upregulation in progenitor cells does not occur in the absence of Hox protein. Residual Col protein indicates the position of the Col-expressing cluster. (J) Col (red) and Nau (green) expression in muscle progenitors in different trunk segments in Hox mutants and Hox-overexpression conditions.

muscles in T2 to A7, in DT1 in A segments, plus, more sporadically, in a few other muscles (Fig. 6B,C) (Dubois et al., 2007). The same pattern was observed upon expression of Col plus Nau, with the notable exception of additional activation in a muscle in T1. This ectopic muscle is at the same position and shows the same orientation and attachment sites as the ectopic DA3 muscle induced by Hox proteins (see Fig. 3). We infer from this observation that high levels of Nau plus Col can bypass the need for a Hox protein in implementing the DA3 muscle formation process. It strengthens our conclusion that the upregulation of expression of a combination of iTFs by Hox proteins plays a decisive role in converting positional information within the mesoderm into a specific muscle pattern.

However, the ectopic DA3 muscle that formed in T1 upon forced expression of Col and Nau displayed very few nuclei (two to three), indicating that Hox activity is essential for allocating this muscle a normal number of nuclei. To better define in which myoblasts Hox was acting, we repeated the Ubx-overexpression experiments using two different drivers, *rp298-Gal4* and *sns-Gal4*, which are specific for FCs and muscle precursors and for FCs, respectively (Menon

and Chia, 2001; Kocherlakota et al., 2008). We found that the number of nuclei in the T2 and T3 segments was converted to that in A segments upon Ubx expression in FCs (Fig. 6E,F) but not in FCMs. These results show that Hox factors act to regulate the size of each muscle. Ubx expression in FCs did not, however, promote the formation of an ectopic DA3 muscle in T1 (Fig. 6E), strengthening our conclusion that Hox activity is required at the progenitor stage to implement the muscle differentiation process.

A model for the transcriptional history of a *Drosophila* muscle

Segment-specific upregulation of *col* and *nau* expression by Hox proteins suggests the following model for the transcriptional history of the DA3 muscle (Fig. 7). The first step is the activation of *col* transcription in a cluster of mesodermal cells in all T and A segments. This is controlled by the early-acting CRM276 enhancer, which integrates positional information issued from the ectoderm, mesoderm-intrinsic cues and repression by N signaling in non-progenitor cells. In a second step, maintenance of Col and Nau

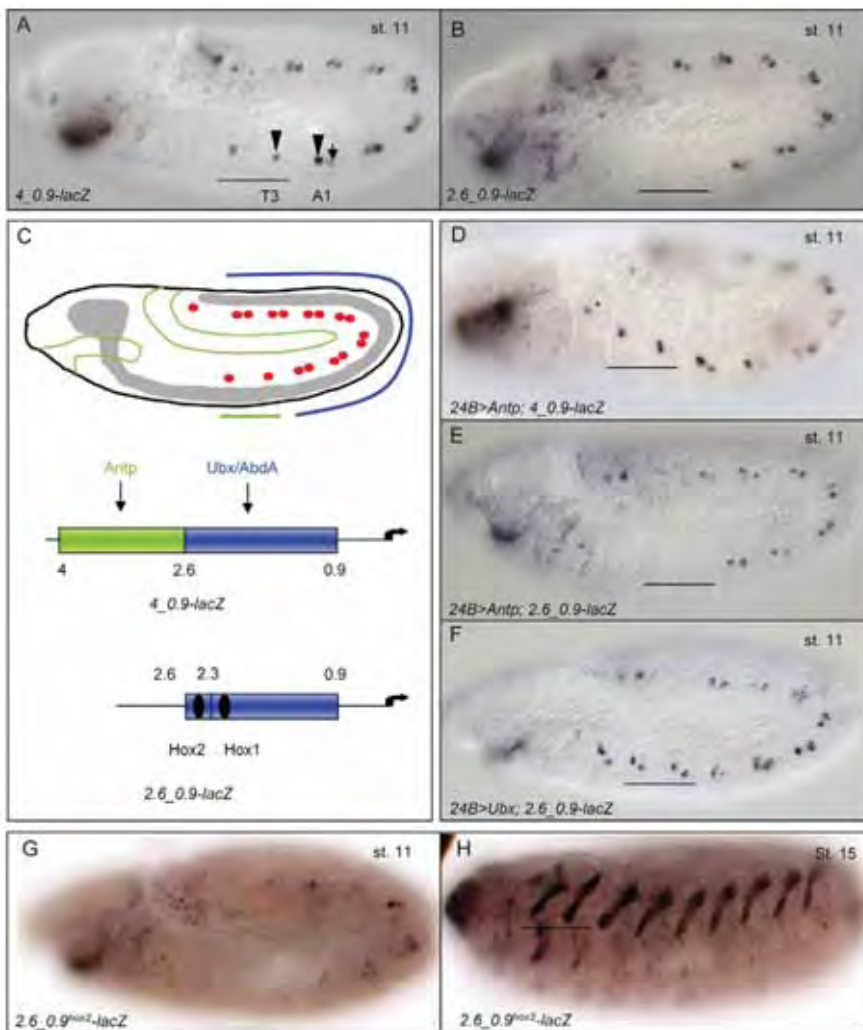


Fig. 5. Hox regulation of *col* transcription at the progenitor stage is modular.

(A,B,D-H) Lateral views of stage 11 *Drosophila* embryos expressing *lacZ* in the DA3/DO5 (A, arrowhead) and DT1/DO4 (A, arrow) progenitors under control of either the *4_0.9* or *2.6_0.9* CRM. (A,B) *lacZ* expression in T2 and T3 (horizontal bar) in wt embryos requires the *4_2.6* CRM. (D-F) *2.6_0.9-lacZ* expression is induced in T2 and T3 by Ubx but not by Antp; *4_0.9-lacZ* expression is induced by Antp in all segments, including in T1 (asterisk in D). (G) Mutation of the Hox2 site abolishes *2.6_0.9-lacZ* expression in progenitors. (H) *lacZ* expression in the DA3 muscle at stage 15 is independent of the *4_2.6* CRM and Hox2. (C) Schematic of a stage 11 embryo with the Col-expressing progenitors shown as red dots and domains of mesodermal Antp and Ubx/AbdA expression as green and blue lines, respectively. Beneath are the *4_0.9-lacZ* and *2.6_0.9-lacZ* constructs showing the Antp- and Ubx-responsive elements in green and blue, respectively. Position of the Hox1 and Hox2 sites is indicated (see Fig. S4 in the supplementary material).

expression in the DA3/DO5 progenitor relies upon Hox factors and is mediated by another CRM, the *4_0.9* CRM, which contains separate elements for regulation by Antp and Ubx/AbdA. The *4_0.9* CRM is also subject to repression by N, leading to restriction of *col* transcription to the DA3 FC. In a third step, implementation of the DA3 muscle differentiation process requires positive and direct Col autoregulation, which converts all of the DA3 muscle nuclei to the same transcriptional program (Dubois et al., 2007). In essence, Col accumulation in the DA3/DO5 progenitor is required to maintain its own expression in the DA3 FC and muscle precursor, thus representing a case of forward autoregulation. Finally, Hox proteins collaborate with iTFs to control the number of myoblasts assigned to the DA3 muscle. Central to our model is the switch in regulation from the early to late muscle CRM that occurs at the progenitor stage and requires Hox activity, thereby linking positional information along the A/P axis to muscle diversity.

DISCUSSION

Cis-reading of positional information: intersecting computational predictions and ChIP-on-chip data

Eve expression in the DA1 muscle lineage provided the first paradigm for studying the early steps of muscle specification. Detailed characterization of an *eve* muscle CRM showed that positional and tissue-specific information were directly integrated at the level of

CRMs via the binding of multiple transcription factors, including dTCF, Mad, Pnt, Tin and Twi (Carmena et al., 1998; Halfon et al., 2000; Knirr and Frasch, 2001). Based on this transcription factor code and using the ModuleFinder computational approach (Philippakis et al., 2006), we have identified a CRM, CRM276, that precisely reproduces the early phase of *col* transcription. This CRM also drove expression in cells of the lymph gland, another organ that is issued from the dorsal mesoderm where *col* is expressed (Crozatier et al., 2004). Parallel to our study, two *col* genomic fragments were selectively retrieved in chromatin immunoprecipitation (ChIP-on-chip) experiments designed to identify *in vivo* binding sites for Twi, Tin or Mef2 in early embryos (Sandmann et al., 2007; Liu et al., 2009). One fragment overlaps with CRM276. Based on this overlap and interspecies sequence conservation, we tested a 1.4 kb subfragment of CRM276 that retained most of the transcription factor binding sites identified by ModuleFinder and found that it specifically reproduced promuscular *col* expression (M. de Taffin, personal communication). This *in vivo* validation shows that intersecting computational predictions and ChIP-on-chip data should provide a very efficient approach to identify functional CRMs on a genome-wide scale (Zinzen et al., 2009).

The *eve* and *col* early mesodermal CRMs are activated at distinct A/P and D/V positions. We are now in a position to undertake a comparison of these two CRMs, in terms of the number and relative

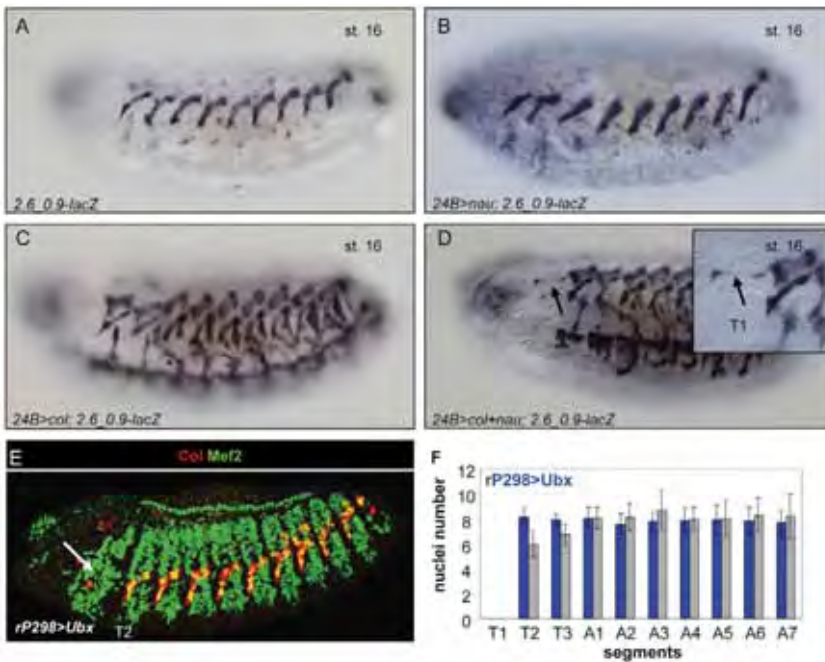


Fig. 6. Expression of Nau and Col can bypass the Hox requirement for DA3 muscle specification. (A-D) *lacZ* expression in stage 16 *2.6_0.9-lacZ* *Drosophila* embryos upon pan-mesodermal expression of Nau (B), Col (C), Col plus Nau (D) or neither protein (A). Only co-expression of Col plus Nau results in the formation of a DA3-like muscle in T1 (arrow). This ectopic muscle has only two nuclei (inset in D). (E) Double immunostaining of stage 15 embryos for Mef2 (green) and Col (red) showing Col expression upon expression of *Ubx* in all FCs. (F) The number of nuclei per DA3 muscle in each segment of *rP298>Ubx* embryos is in blue. Wt values are in gray. Number of segments counted=30.

spacing of common activator and repressor sites and their expanded combinatorial code, in order to understand how different mesodermal cis elements perform a specific interpretation of positional information.

The Hox code relays positional information in a segment-specific manner

A progenitor is selected from the Col promuscular cluster in T2 and T3 but not T1. One cell issued from the Col-expressing promuscular cluster in T1 nevertheless shows transiently enhanced Col

expression, suggesting that the generic process of progenitor selection is correctly initiated in T1. This process aborts, however, in the absence of a Hox input, as shown by the loss of progenitor Col expression and DA3 muscle in specific segments in Hox mutants. The similar changes in Nau and Col expression observed under Hox gain-of-function conditions allow us to conclude that the expression of iTFs is regulated by Hox factors at the progenitor stage. The superimposition of Hox information onto the intrasegmental information thereby implements the iTF code in a segment-specific manner and establishes the final muscle pattern. Unlike DA3, a

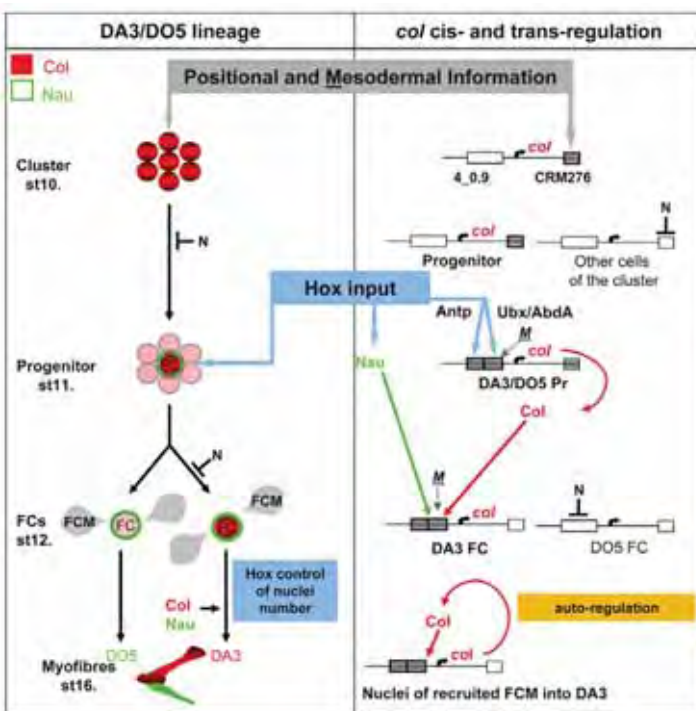


Fig. 7. Transcriptional control of *Drosophila* muscle identity: the central role of Hox proteins. Expression of Col (red) and Nau (green) in the DA3 muscle lineage is shown on the left, with the corresponding cis control of *col* expression represented on the right. *col* transcription is activated in one promuscular cluster per trunk segment under control of the early-acting CRM276, which integrates A/P and D/V positional information, mesoderm-intrinsic cues and repression by N signaling. This results in Col expression in one (T segments) or two (A1-A7 segments) dorsal progenitors. Upregulation of *col* (red) and *nau* (green) transcription in the DA3/DO5 progenitor at stage 11 is controlled by Antp and Ubx/AbdA and is mediated by two separate elements of the 4_0.9 CRM. This is the key step, when segment-specific variations are imposed on the segmental muscle pattern. Following progenitor division, restriction of *col* transcription to the DA3 FC requires positive regulation by Col and Nau and repression by N. From this stage on, Col is required for activating its own transcription in all FCM nuclei incorporated into the DA3 myofiber. Hox activity in FCs controls the number of myoblasts allocated to the DA3 muscle, independently of the control of *nau* and *col*. The key regulatory steps are shaded and centered: early positional information is relayed by segment-specific Hox information at the progenitor stage. The active and inactive CRMs are indicated by gray and white boxes, respectively. Binding of mesodermal transcription factors (M), such as Twi and Mef2, to CRM 276 and 4_0.9 is suggested by phylogenomic footprinting (see Figs S3 and S4 in the supplementary material) and ChIP-on-chip data (Sandmann et al., 2006; Sandmann et al., 2007).

number of specific muscles are found in both T1 and T2-A7 (Bate and Rushton, 1993), such as the Eve-expressing DA1 muscle; other muscles form in either abdominal or thoracic segments, as illustrated by the pattern of Nau expression in stage 16 embryos (Michelson, 1994) (Fig. 1E). This diversity in segment-specific patterns indicates that Hox regulation of iTF expression is iTF and/or progenitor specific.

Modular cis regulation of *col* transcription by Hox proteins

As early as 1994, Hox proteins were proposed to regulate the segment-specific expression of iTFs. Seven years later, Capovilla et al. characterized an *apterous* mesodermal enhancer (apME680) active in the LT1-4 muscles and proposed that regulation by Antp was direct (Capovilla et al., 2001). However, mutation of the predicted Antp binding sites present in apME680 abolished its activity also in A segments, suggesting that some of the same sites were bound by Ubx/AbdA. We now have evidence that the regulation of *col* expression by Ubx/AbdA in muscle progenitors is direct and involves a single Hox binding site. However, regulation by Antp does require other cis elements. It remains to be seen whether regulation by Antp is also direct. Since Antp, Ubx and AbdA display indistinguishable DNA-binding preferences in vitro (Ekker, 1994), the modular regulation of *col* expression by different Hox paralogs suggests that other cis elements and/or Hox collaborators contribute to Hox specificity (Mann et al., 2009). Direct regulation of *col* by Ubx has previously been documented in another cellular context, that of the larval imaginal haltere disc, via a wing-specific enhancer (Hersh and Carroll, 2005). In this case, Ubx directly represses *col* expression by binding to several sites, contrasting with *col*-positive regulation via a single site in muscle progenitors. This is the second example, in addition to CG13222 regulation in the haltere disc, of direct positive regulation by Ubx via a single binding site (Hersh et al., 2007). Hox 'selector' proteins collaborate on some cis elements with 'effector' transcription factors that are downstream of cell-cell signaling pathways (Grienerberger et al., 2003; Walsh and Carroll, 2007; Mann et al., 2009). In the DA3 lineage, it seems that Dpp, Wg and Ras signaling act on one *col* cis element and the Hox proteins on others. The regulation of *col* expression by Hox proteins in different tissues via different CRMs provides a new paradigm to decipher how different Hox paralogs cooperate and/or collaborate with tissue- and lineage-specific factors to specify cellular identity (Brodu et al., 2004; Gebelein et al., 2004; Walsh and Carroll, 2007; Stobe et al., 2009).

Hox proteins control the number of myoblasts allocated to each muscle

The DA3 muscle displays fewer nuclei in T2 and T3 than in A1-A7, an opposite situation to that described for an aggregate of the four LT1-4 muscles. Capovilla et al. proposed that the variation in the number of LT1-4 nuclei was controlled by Hox proteins (Capovilla et al., 2001). Our studies of the DA3 muscle extend this conclusion by showing that the variations due to Hox control are specific to each muscle and are exerted at the level of FCs. Since the number of nuclei is both muscle- and segment-specific, Hox proteins must cooperate and/or collaborate with various iTFs to differentially regulate the nucleus-counting process. As such, Hox proteins contribute to the combinatorial code of muscle identity. Identifying the nature of the cellular events and genes that act downstream of the iTF/Hox combinatorial code and that are involved in the nucleus-counting process represents a new challenge.

Acknowledgements

We thank the Bloomington Stock Center, S. Abmayr, S. Menon, E. Sanchez-Herrero and F. Karch for fly stocks; B. Patterson and E. Furlong for antibodies; F. Karch and members of our laboratory for critical reading of the manuscript; and E. Furlong for communicating unpublished data. We acknowledge the help of the Toulouse RIO Imaging Platform and B. Ronsin and A. Leru for confocal microscopy. This work was supported by CNRS and Ministère de la Recherche et de la Technologie (MRT), Université Paul Sabatier, Association Française contre les Myopathies (AFM) and the US National Institutes of Health. J.E. was supported by fellowships from MRT and AFM and H.B. by a fellowship from MRT. Deposited in PMC for release after 12 months.

Competing interests statement

The authors declare no competing financial interests.

Supplementary material

Supplementary material for this article is available at <http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.045286/-DC1>

References

- Abbott, M. K. and Kaufman, T. C. (1986). The relationship between the functional complexity and the molecular organization of the Antennapedia locus of *Drosophila melanogaster*. *Genetics* **114**, 919-942.
- Affolter, M., Slattery, M. and Mann, R. S. (2008). A lexicon for homeodomain-DNA recognition. *Cell* **133**, 1133-1135.
- Alvares, L. E., Schubert, F. R., Thorpe, C., Mootoosamy, R. C., Cheng, L., Parkyn, G., Lumsden, A. and Dietrich, S. (2003). Intrinsic, Hox-dependent cues determine the fate of skeletal muscle precursors. *Dev. Cell* **5**, 379-390.
- Balagopalan, L., Keller, C. A. and Abmayr, S. M. (2001). Loss-of-function mutations reveal that the *Drosophila nautilus* gene is not essential for embryonic myogenesis or viability. *Dev. Biol.* **231**, 374-382.
- Barolo, S., Carver, L. A. and Posakony, J. W. (2000). GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in *Drosophila*. *Biotechniques* **29**, 726, 728, 730, 732.
- Bate, M. (1990). The embryonic development of larval muscles in *Drosophila*. *Development* **110**, 791-804.
- Bate, M. (1993). The mesoderm and its derivatives. In *The Development of Drosophila melanogaster*, vol. 2 (ed. M. Bate and A. Martinez-Arias), pp. 1013-1090. Cold Spring Harbor, NY: 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.
- Baylies, M. K., Bate, M. and Ruiz Gomez, M. (1998). Myogenesis: a view from *Drosophila*. *Cell* **93**, 921-927.
- Bischof, J., Maeda, R. K., Hediger, M., Karch, F. and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc. Natl. Acad. Sci. USA* **104**, 3312-3317.
- Bourgouin, C., Lundgren, S. E. and Thomas, J. B. (1992). *Apterous* is a *Drosophila* LIM domain gene required for the development of a subset of embryonic muscles. *Neuron* **9**, 549-561.
- Brand, A. H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401-415.
- Brodu, V., Elstob, P. R. and Gould, A. P. (2004). EGF receptor signaling regulates pulses of cell delamination from the *Drosophila* ectoderm. *Dev. Cell* **7**, 885-895.
- Capovilla, M., Kambris, Z. and Botas, J. (2001). Direct regulation of the muscle-identity gene *apterous* by a Hox protein in the somatic mesoderm. *Development* **128**, 1221-1230.
- 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., Gisselbrecht, S., Harrison, J., Jimenez, F. and Michelson, A. M. (1998). Combinatorial signaling codes for the progressive determination of cell fates in the *Drosophila* embryonic mesoderm. *Genes Dev.* **12**, 3910-3922.
- Crozatier, M. and Vincent, A. (1999). Requirement for the *Drosophila* COE transcription factor Collier in formation of an embryonic muscle: transcriptional response to notch signalling. *Development* **126**, 1495-1504.
- Crozatier, M. and Vincent, A. (2008). Control of multidendritic neuron differentiation in *Drosophila*: the role of Collier. *Dev. Biol.* **315**, 232-242.
- Crozatier, M., Valle, D., Dubois, L., Ibsouda, S. and Vincent, A. (1996). *collier*, a novel regulator of *Drosophila* head-development, is expressed in a single mitotic domain. *Curr. Biol.* **6**, 707-718.
- Crozatier, M., Ubeda, J. M., Vincent, A. and Meister, M. (2004). Cellular immune response to parasitization in *Drosophila* requires the EBF orthologue *collier*. *PLoS Biol.* **2**, E196.
- Demontis, F. and Perrimon, N. (2009). Integration of Insulin receptor/Foxo signaling and dMyc activity during muscle growth regulates body size in *Drosophila*. *Development* **136**, 983-993.
- Dubois, L., Enriquez, J., Daburon, V., Crozet, F., Lebreton, G., Crozatier, M. and Vincent, A. (2007). Collier transcription in a single *Drosophila* muscle

- lineage: the combinatorial control of muscle identity. *Development* **134**, 4347-4355.
- Duboule, D. and Morata, G.** (1994). Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet.* **10**, 358-364.
- Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E. and Beachy, P. A.** (1994). The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. *EMBO J.* **13**, 3551-3560.
- Estrada, B., Choe, S. E., Gisselbrecht, S. S., Michaud, S., Raj, L., Busser, B. W., Halfon, M. S., Church, G. M. and Michelson, A. M.** (2006). An integrated strategy for analyzing the unique developmental programs of different myoblast subtypes. *PLoS Genet.* **2**, e16.
- Fujioka, M., Wessells, R. J., Han, Z., Liu, J., Fitzgerald, K., Yusibova, G. L., Zamora, M., Ruiz-Lozano, P., Bodmer, R. and Jaynes, J. B.** (2005). Embryonic even-skipped-dependent muscle and heart cell fates are required for normal adult activity, heart function, and lifespan. *Circ. Res.* **97**, 1108-1114.
- Gebelein, B., McKay, D. J. and Mann, R. S.** (2004). Direct integration of Hox and segmentation gene inputs during *Drosophila* development. *Nature* **431**, 653-659.
- Greig, S. and Akam, M.** (1993). Homeotic genes autonomously specify one aspect of pattern in the *Drosophila* mesoderm. *Nature* **362**, 630-632.
- Grienenberger, A., Merabet, S., Manak, J., Iltis, I., Fabre, A., Berenger, H., Scott, M. P., Pradel, J. and Graba, Y.** (2003). Tgfbeta signaling acts on a Hox response element to confer specificity and diversity to Hox protein function. *Development* **130**, 5445-5455.
- Halfon, M. S., Carmena, A., Gisselbrecht, S., Sackerson, C. M., Jimenez, 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.
- Hersh, B. M. and Carroll, S. B.** (2005). Direct regulation of knot gene expression by Ultrabithorax and the evolution of cis-regulatory elements in *Drosophila*. *Development* **132**, 1567-1577.
- Hersh, B. M., Nelson, C. E., Stoll, S. J., Norton, J. E., Albert, T. J. and Carroll, S. B.** (2007). The UBX-regulated network in the haltere imaginal disc of *D. melanogaster*. *Dev. Biol.* **302**, 717-727.
- Heuer, J. G., Li, K. and Kaufman, T. C.** (1995). The *Drosophila* homeotic target gene centrosomin (*cnn*) encodes a novel centrosomal protein with leucine zippers and maps to a genomic region required for midgut morphogenesis. *Development* **121**, 3861-3876.
- Hooper, J. E.** (1986). Homeotic gene function in the muscles of *Drosophila* larvae. *EMBO J.* **5**, 2321-2329.
- Hueber, S. D. and Lohmann, I.** (2008). Shaping segments: Hox gene function in the genomic age. *BioEssays* **30**, 965-979.
- Jagla, T., Bellard, F., Lutz, Y., Dretzen, G., Bellard, M. and Jagla, K.** (1998). ladybird determines cell fate decisions during diversification of *Drosophila* somatic muscles. *Development* **125**, 3699-3708.
- Keller, C. A., Erickson, M. S. and Abmayr, S. M.** (1997). Misexpression of nautilus induces myogenesis in cardioblasts and alters the pattern of somatic muscle fibers. *Dev. Biol.* **181**, 197-212.
- Knirr, S. and Frasch, M.** (2001). Molecular integration of inductive and mesoderm-intrinsic inputs governs even-skipped enhancer activity in a subset of pericardial and dorsal muscle progenitors. *Dev. Biol.* **238**, 13-26.
- Knirr, S., Azpiazu, N. and Frasch, M.** (1999). The role of the NK-homeobox gene slouch (*S59*) in somatic muscle patterning. *Development* **126**, 4525-4535.
- Kocherlakota, K. S., Wu, J. M., McDermott, J. and Abmayr, S. M.** (2008). Analysis of the cell adhesion molecule sticks-and-stones reveals multiple redundant functional domains, protein-interaction motifs and phosphorylated tyrosines that direct myoblast fusion in *Drosophila melanogaster*. *Genetics* **178**, 1371-1383.
- Liu, Y. H., Jakobsen, J. S., Valentin, G., Amarantos, I., Gilmour, D. T. and Furlong, E. E.** (2009). A systematic analysis of Tinman function reveals Eya and JAK-STAT signaling as essential regulators of muscle development. *Dev. Cell* **16**, 280-291.
- Mann, R. S., Lelli, K. M. and Joshi, R.** (2009). Hox specificity unique roles for cofactors and collaborators. *Curr. Top. Dev. Biol.* **88**, 63-101.
- Menon, S. D. and Chia, W.** (2001). *Drosophila* rolling pebbles: a multidomain protein required for myoblast fusion that recruits D-Titin in response to the myoblast attractant Dumbfounded. *Dev. Cell* **1**, 691-703.
- Michelson, A. M.** (1994). Muscle pattern diversification in *Drosophila* is determined by the autonomous function of homeotic genes in the embryonic mesoderm. *Development* **120**, 755-768.
- Michelson, A. M., Abmayr, S. M., Bate, M., Arias, A. M. and Maniatis, T.** (1990). Expression of a MyoD family member prefigures muscle pattern in *Drosophila* embryos. *Genes Dev.* **4**, 2086-2097.
- Nose, A., Ishiki, T. and Takeichi, M.** (1998). Regional specification of muscle progenitors in *Drosophila*: the role of the *rsh* homeobox gene. *Development* **125**, 215-223.
- Philippakis, A. A., Busser, B. W., Gisselbrecht, S. S., He, F. S., Estrada, B., Michelson, A. M. and Bulyk, M. L.** (2006). Expression-guided in silico evaluation of candidate cis regulatory codes for *Drosophila* muscle founder cells. *PLoS Comput. Biol.* **2**, e53.
- Ruiz Gomez, M. and Bate, M.** (1997). Segregation of myogenic lineages in *Drosophila* requires numb. *Development* **124**, 4857-4866.
- Ruiz-Gomez, M., Romani, S., Hartmann, C., Jackle, H. and Bate, M.** (1997). Specific muscle identities are regulated by Kruppel during *Drosophila* embryogenesis. *Development* **124**, 3407-3414.
- Sandmann, T., Jensen, L. J., Jakobsen, J. S., Karzynski, M. M., Eichenlaub, M. P., Bork, P. and Furlong, E. E.** (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., Stolz, V. and Furlong, E. E.** (2007). A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev.* **21**, 436-449.
- Speicher, S., Fischer, A., Knoblich, J. and Carmena, A.** (2008). The PDZ protein Canoe regulates the asymmetric division of *Drosophila* neuroblasts and muscle progenitors. *Curr. Biol.* **18**, 831-837.
- Stobe, P., Stein, M. A., Habring-Muller, A., Bezdan, D., Fuchs, A. L., Hueber, S. D., Wu, H. and Lohmann, I.** (2009). Multifactorial regulation of a hox target gene. *PLoS Genet.* **5**, e1000412.
- Vervoort, M., Crozatier, M., Valle, D. and Vincent, A.** (1999). The COE transcription factor Collier is a mediator of short-range Hedgehog-induced patterning of the *Drosophila* wing. *Curr. Biol.* **9**, 632-639.
- Walsh, C. M. and Carroll, S. B.** (2007). Collaboration between Smads and a Hox protein in target gene repression. *Development* **134**, 3585-3592.
- Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M. and Furlong, E. E.** (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**, 65-70.

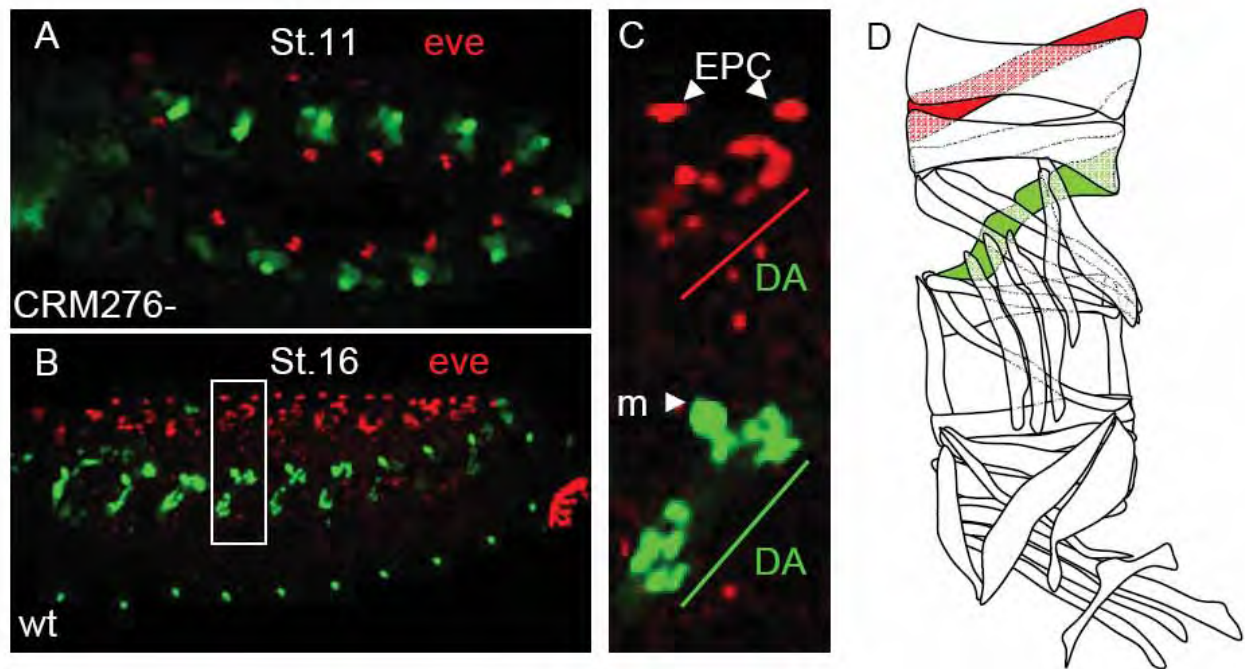


Fig. S1. The distinct A/P and D/V positions of the DA1 and DA3 clusters and muscles. (A) Lateral view of a stage 10 *CRM276-lacZ* embryo stained for Eve (red) and β -gal (green), showing that the Eve and Col-expressing clusters have distinct positions within the mesoderm. (B) Lateral view of a stage 16 embryo stained for Eve and Col. (C) An enlarged view of the A2 segment from B. The same relative positions of Eve and Col expression are conserved. (D) Diagram of the somatic musculature of an abdominal hemisegment (Bate and Rushton, 1993), showing the DA1 muscle in red and the DA3 muscle in green. md, multidendritic neurons expressing Col; EPCs, Eve pericardial cells.

TGGCCGGGATAAAGTCGATGGCTTAAGCTTAGAGTTAGAATTTAGTAGGAGCACTAAATA
TTAGGTTATAAATAGCTTGAACCTTATTAACAATATTATACTATCAGCTTAAAAGTTAATA **Tin**
TTTATATTTAAGTACCATCCTTAAATTCCACTTTGTGCCAGATAGAGCT**TGAAGTGCTAT**
CTTCTGTTCAATTGGCACTTTTTAGCCAATCGGGCAGAAACATTTAAGGACATTCGGAGC
AGCTAAAAAACAATAATCTCCGTCCAATCGAAGAGAAAC**GCCGCCAA**GTCCGCATCCC **Mad**
CTCGC**CCAAATATGAATGACCAATTATGCACCTTCGTTGGC**CTACGTGACCGAGCTGCCA
ATTTTGGGCCCATCGAAAATCCCAAACGGAGGTGTCCCGATGTCATGTGCG**AACATTTGA** **Twi**
TGATGCGTAGCTTTTCATTAAATGAAAATTTTCGCGTTGCTGGCAAACAAAAGGACAAAG
AAGCGGCCAAAGGCGCTAACTTTGCTGGCAAATGGAAGGCC**AAGATGGGAGA**CAC**TTTTGA** **dTCF**
TTATTGACAGGACGCGAGTGGAAATTTTAAATTGCCCCAAGTGGGTGGTGGTGTCTAT**** **Tin**
GGTTGGTTGGTTTTACCGAGGGGAAAAATCCTTGAATTGGTCCAGGACTTGTGTTTTCGGTT
AGATTTCCCCAAAATGCAATATCCTCGTCTTGAAGA**CATATGCAAATGG**GAGGGGGCACC **Twi**
CGGTG**AAATTTATGTGTCCGTGCAGGATATTTTGAAAACATTGTAAT**CAAGCAAGTCGAG****
GGGCAGGATGCT**TTCCGAGAAAGTCAATTAGCTGCCATA**AAAATGCAGGCTGGCCTCGAA
GTGATGAGTGGCTTGGTGTCTCGGGGATGAAGTCATCGCCGGGCAAGTGAATCATTATC
CTGGAGGATCCCCCGTCCCTTAGA**CACTCGATCGGTATTCAAATTTCAA**CTGCAGTTGCA**** **Ets, Twi**
TCGTGAGTTTTCCACGTTCTGTTGCCGCACCGTAATTCGATATTGTATCTGACAGATATAC****
ACGGCTCTCAAATCTCGAGGGATGGCGCACTGGATTGCTTGTGTTAATGATGATGAGCCC
GCTCCGATGCTCTTTTTTTTTCCGTAATGGATAACATCCATCCACATCT**CACATCTGTTTT** **Twi**
TTCTTCGCAAAGTCCAACAAGTTCACAGGCATTCTCGATAAGCACCACTTGGCTTTTA
ATGCTGAATATAACAAGTAATACAATAACAATTTATTTTAGTAAGCCGAAGTTTCTATGCC
CTAGCAATTTTAGAATTTTAGCAATTTTATTTACATATTTGTAATAACAAGTTTATCA
TTATGTTTATGAAGTATGGTATACTATAGTACTATATTTAAAAAGGTATCTAGTTCAATT
TGAGAT**TATTGTATATAGAAAACACTTGGTGATTTGTTACTATAACTTTAAAACCCGT**
TTTCAGGGTATAAGTTTTACGCATAAAAATAAATAAATAACTTGTTTACACTTTGTGATTT
GGACTTATTGGATACTACTCAATGGTTTTAGTTTTATCAGTGTTAATAGAAAATATTGCTT
TTAGTGGTTATCTCGAGTGCCAGTTTCTTATCGCCCGCTCAGAGATTTATATGCTTTAGA
AATTCAAAGAGCAAAAAGAGCTATTTAGGCTGAAGTTTTTATAGGGAGACCGAATT**ATTAA**
AAAAAAAAAAAAACAAGTGTCCGGAAGTATAAACAATTTGCGACACTCCCATCGGACAAG
GACATATGCAGATGCCACCCGCTGGCCAGAGATCAACCCTCCAGCTGCTGCTAATTTGT**** **Twi**
TCAATTTGTGCGAGTGGCAAACAACGGAGAAAAATATGATGAAAATTTGCCGGCACAACGC
CGCCGCTTTTCCATCTCTCCCCCCCCCCCCCATCGCTGATCGGAAAGG**AAAATATTATT**
TCGGCAAATAAAAAGTTGGCAGGGGATTTAGATAGGGCGTATCATGTGCGCGGCGGATG
CATAAAAGTATTT**CAG**CATTTAAATGC**AACGCACACGGGTGTCAGTTTCATTGGATTGG** **Mef-2**
TTTCGAGACGGATTGGGTTGGGTTGGCTGGGCTAACAGGCTCAC**ATCCTG**GCCGCAGGA****
CATTGCGCGGCTGTCA**CCCACTTATCGCAGTGGCACTCACTACC**CGCGGGCC**AAACA** **Twi**
TCTGCCCAGAAAAGCCAAGCTAACC**GGCA**CAAGTGGCAGCAAAAAGGCG**CAACC****CGTCAT
TAGCGGCTCGTAAAGTCGCCTTCGAGGAGGCGAACATCGTAAAGACTCTTAATTGAAGCT
GGCATAAATCCACA**ACTGATTGATAATTGCCAGCACAAATGGGAAAAACGAGTCCAATGCC**
ATTTATGGTCCATCCTCAGTCTGCCCTCACAAAATAGAATAGCTGCAGGAATGGTT**

Fig. S2. Nucleotide sequence of *D. melanogaster* CRM276. FlyBase positions 2R: 10664067-10666462, identified by ModuleFinder (Philippakis et al., 2006). The >20-bp sequence motifs that are more than 90% conserved at the same relative positions in *D. melanogaster*, *D. pseudobscura* and *D. virilis* are shaded in blue; additional shorter conserved motifs are shaded in gray. The putative binding sites for dTCF (orange), Ets (purple), Mad (pale blue), Tin (red) and Twi (green) used by ModuleFinder are in bold characters and their position indicated in the margin. A putative site for Mef2 (brown) is also indicated. The sequence of the DNA fragment identified by genome-wide mapping of Twi in vivo binding sites (Sandmann et al., 2007) is underlined. The 3' end of the CRM276 subfragment, which reproduces *col* expression only in promuscular clusters, is shaded in red.

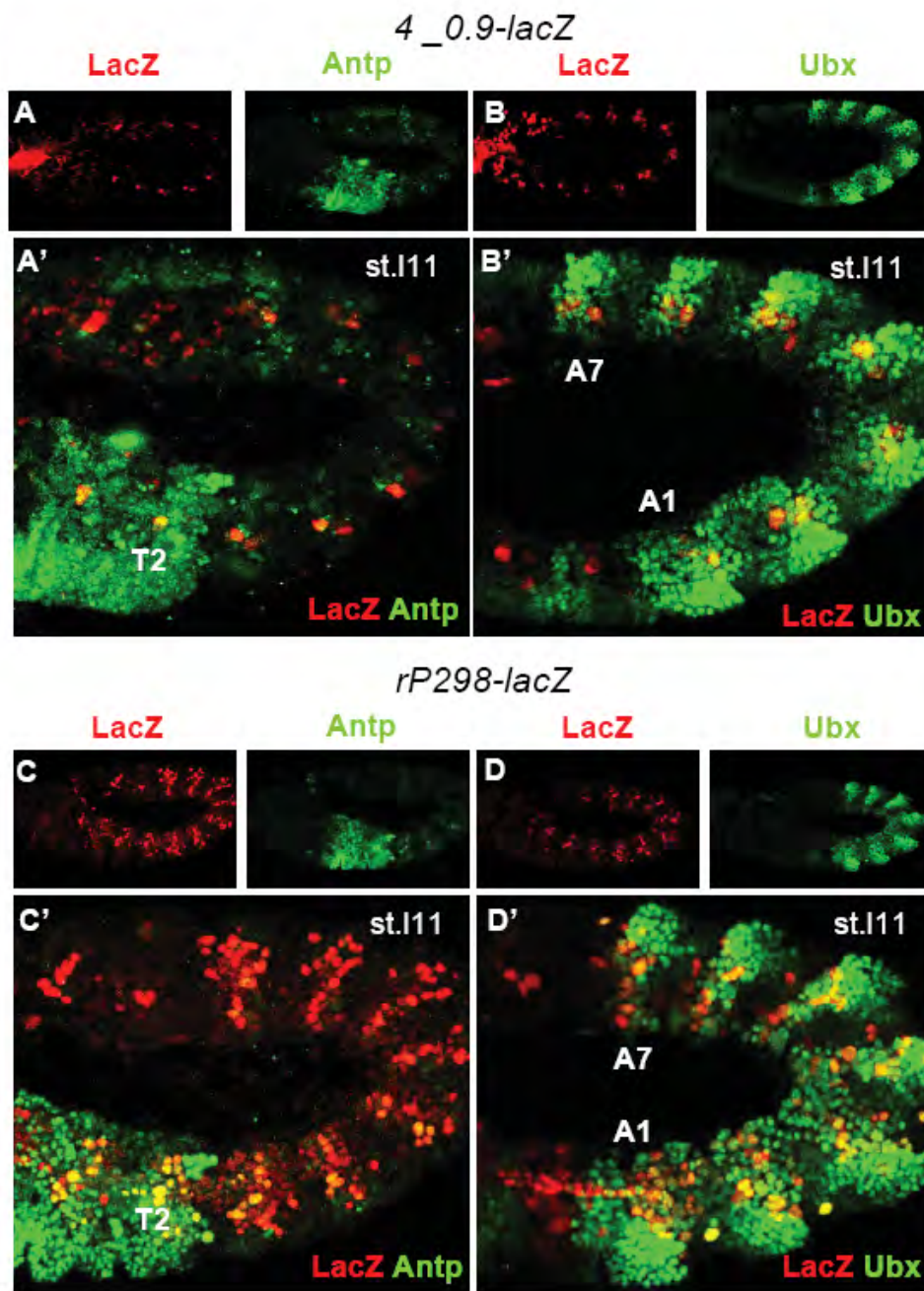


Fig. S3. Antp and Ubx expression in the DA3/D05 muscle progenitor and founder cells. Lateral views of stage 11 *4_0.9-lacZ* embryos (**A,B**) and late stage 11 *rP298-lacZ* embryos (**C,D**) double stained for β -gal (red) and (green) Antp (**A,C**) or Ubx (**B,D**). (**A'-D'**) Enlarged views of **A-D**. Antp and Ubx are expressed in the DA3/D05 muscle progenitor and founder cells in the T2-T3 and A1-A7 segments, respectively.

CTCGAGAGCGAGCTGGAAAC**TAATTA**AAATGTTTCGCGCCTCAAATTTCC **Hox²**
TTATTTATGACCCCTGTGAATGTCTTGCATATGGACGGAGTTGAGATCCTTCTCGTTTTCT **Twi**
 TCCTTTTTTTGTTTACGAGTGACAAATGGGACATTTACATGCGAGTCAGTTTGCATATG
 GAATTATTCATGGTTTCTTTAAAGATGGCGCAGGATGTGATGTGCGGGAAAAGGATGAG
 GGTTTCCTCAAGGAAGTCGGAAATAGAAGTGGTTTCTTTTCATGCTACCAATATGGGGGC
 ACATAAACTTGCAATTTGATGAGCTAAATGTAACATACACTATTTAAATACATTTTTTA
 TGATAAGTGAACCTAAAGTCACACTAACTTTTGAAAATTTGATTGACTTCTACAATGTGT
 GTTGTTTAATTTCTTATTTTATATTTTTTTAAATATCGAAAATCTACAAATCCGCTTATG
 TTTAAAAGTCAAGCCGCTGGCTAATTGACAAAATGTGTAATTGTGGCGATGAGAGTCCTT
 CCGATTGTCACTCTCCCAACCCTCCGTAATCCCCTAAGTCACCATAGGGGTTGGGTAAAA
 ATCAAATGCGGAGAATGTACCAACAAATTATTTAGCAATTTGGCTAGTGCCTAGTCCGGC
 GCCATGTAAATCCAATTGTAACAATTGAAATCAAATTCGGGGGCATCGCCACCGAAAG
 GGGGTGGCATGGGTTAAAGGGTCATGGTGTCCATGACAGCTGTCCGGCCGGGAAACTC **Nau**
 CACTGAGGATTGGCGGTGATAAAAGGGGTTAATCGAACACGCGGACCACAGCTTTTGGAT
 GGAATTTTCTTACCTTTGCCGAAAAAGAAAATTTATTTGTCAAAGGGCGTGGAATTTCC
 CGGTTCTGGTTATT**TTGGTTAATTAGCGATGCTCTGGGGACATCTT** **Hox¹, Col**
 CAGTCTGCTCTTTTCGCCTTTGATTTACACATAAAATTTGTGCGACCTAGCAGCTCTGGGGAAATGTTCCAGA
 ACTTCTCAAAAATAGACCAACAAGTGTGGGTGTGGGGAAAATTCGGGAAAAAGAGTTGCC **Mef-2**
 GTCTTGTGGTTAACGGAGTTTTCGGTCTTGTGAGAAGTTAATGAGATTAACCAGCCCCC
 CCAAATTAATTCACGGAAAAGTCAAACAAGCTCTCTCTTTTCCCTCGTCCATTCAATCG
 GTCACAGGCAGAAACCGCAAATGCGGAAAATGCAAATTTGTTTTCATGCGAACCCAGAAC
 CTCAAAAACCAATGCCATCCCCTGACCTTTTTTACCTTTTGTTCCTAACGAGGCGTG
 AAAATTTTGTGGGCGGTTCCGGTAACTAATCCCTTGTGTTGGAAGAGAAAGAGGAAGAG
 GCATTTGGATGCGAAGGGTATGCCTATAGGCACTCAACTCTGTTTGACAACCTTCTTCTGC
 CTAATTTATGTGCGAAAGTCAGAAGGGTTAAACCCGAAAATGATGACTAATCAAGTCTGT
 GCTTACAACAAATTTATAAAACAAAAGGCAAGATATAAAGGGAAGAAAGTGAACCTATGAA
 ACAACAACAGGATTACAACCTGCCTTTACGTACATATAACTATAACTTCATATCGCAAAA
 CATTATGAATGTAATTATGTACGTTTTCGTTATACATGTAATAGAAAAGCGATAACATCATTA
 AATGTATTAATGTGTTATTTTTTTATCTAAAAAGTCACAGAAGAATTC

Fig. S4. Nucleotide sequence of the *D. melanogaster col* upstream region from position -2626 to position -916 that corresponds to the site of fusion of 2.6_0.9c or 4_0.9c with *lacZ* in *attB* reporter genes. The restriction sites used for cloning are in italics. The >20-bp sequence motifs that are more than 90% conserved at the same relative positions in *D. melanogaster*, *D. pseudobscura* and *D. virilis* are shaded in blue; additional shorter conserved motifs are shaded in gray. Conserved Hox binding sites TAATTA (Egger et al., 1994) are in bold red and numbered in the margin. Binding sites for Col (purple), Mef2 (brown), Nau (blue) and Twi (green) are also indicated.

C. *tup/islet* , un nouveau gène d'identité musculaire.

Article 2 *Tup/Islet1* integrates time and position to specify muscle identity in *Drosophila*. Hadi Boukhatmi, Jean Louis Frenedo, Jonathan Enriquez, Michèle Crozatier, Laurence Dubois and Alain Vincent

Dans cet article, nous décrivons un rôle, jusqu'ici inconnu, du FT à domaine LIM-Homéodomaine Tailup (*Tup*)/*Islet1*, dans le mésoderme dorsal. *Tup* a été précédemment caractérisé pour ses rôles dans la spécification des neurones moteurs et la navigation axonale, et la formation du cœur et de la glande lymphatique. Notre étude révèle une fonction essentielle de *Tup* dans la spécification des muscles dorsaux. Il met aussi en évidence l'importance de la régulation temporelle de l'expression des FTi dans le processus de spécification de l'identité des progéniteurs musculaires.

Le point de départ de cette étude est un crible génétique systématique, entrepris par Laurence Dubois, afin d'identifier l'ensemble des gènes requis pour la formation du muscle DA3. Ce crible est basé sur l'analyse de déficiences chromosomiques partiellement chevauchantes, couvrant toute la longueur des chromosomes 2 et 3. Il a permis d'identifier une région chromosomique dont la délétion provoque l'activation ectopique de *Col* dans le muscle dorsal DA2 et la transformation morphologique de ce muscle en muscle DA3. L'analyse détaillée de cette région a montré que l'absence du gène *tup* est à l'origine de ce phénotype (Article 2 Fig 1). J'ai d'abord corrélé ce phénotype avec l'expression de *Tup* dans les quatre muscles dorsaux DA1, DO1, DA2, DO2, puis entrepris une analyse détaillée du patron de transcription de *tup* au cours de la myogenèse par hybridation in situ avec des sondes introniques pour *tup* et *col* (Article 2 Fig 4). Les résultats ont révélé que le PC à l'origine du muscle DA2 est sélectionné très tôt à partir du groupe promusculaire exprimant *Col*, avant le PC DA3/DO5, et qu'il co-transcrit transitoirement *col* et *tup*. Ce PC se divise de manière asymétrique pour générer la FC DA2 et un précurseur musculaire adulte (AMP) dorso-latéral (Article 1 Fig 2). Des expériences de gain et de perte de fonction de *tup* m'ont permis de montrer que *tup* réprime la transcription de *col* dans le PC DA2/AMP. Cette répression impose une identité DA2 alors qu'en absence de *tup*, l'expression maintenue de *col* impose une identité DA3 (Article 2 Fig 3). Le résultat le plus frappant est que l'activation de *tup* dans un seul des PC exprimant *col* reflète la dynamique d'expression de *Tin* qui agit en amont de *tup* (Article 2 Fig 5). Une régulation spatiale et temporelle est donc impliquée dans la spécification de l'identité DA2.

Je me suis ensuite intéressé à la régulation transcriptionnelle de *tup* dans le mésoderme somatique dorsal et montré que, comme *col*, l'expression de *tup* est régulée par deux CRM indépendants. Un CRM précoce qui dépend de l'activité de Tin et est responsable de l'accumulation de Tup dans le PC DA2. Cette accumulation est nécessaire à la répression de *col* et à l'initiation d'une boucle d'autorégulation positive de *tup* dans le lignage DA2, via un second CRM, tardif. Ce mécanisme de « passage de témoin » entre deux CRM d'abord mis en évidence pour *col* (Article 1) nous permet de proposer qu'un mécanisme de relais permettant une mémoire de l'identité du PC pourrait être général.

Les résultats décrits dans cet article montrent une fonction essentielle de *tup* dans la spécification de l'identité des muscles dorsaux ; l'expression mutuellement exclusive de *tup* et *col* permet la diversification morphologique des muscles dorsaux et dorsaux latéraux. Chez le tunicier *Ciona intestinalis*, l'expression de Ci-COE (l'orthologue de Col) impose un destin « muscle pharyngal » aux progéniteurs issus du territoire cardiaque secondaire qui expriment Islet. Un chevauchement d'expression de EBF-Islet est aussi observé dans les arcs branchiaux 1 et 2 des vertébrés qui contribuent à la formation des muscles faciaux. L'ensemble suggère que des régulations croisées entre Tin/Nkx2.5, Col/EBF et Islet/Tup et leurs rôles dans la diversification des muscles ont été conservés au cours de l'évolution.

Tup/Islet1 integrates time and position to specify muscle identity in *Drosophila*

Hadi Boukhatmi, Jean Louis Frenedo, Jonathan Enriquez*, Michèle Crozatier, Laurence Dubois[‡] and Alain Vincent[‡]

SUMMARY

The LIM-homeodomain transcription factor Tailup/Islet1 (Tup) is a key component of cardiogenesis in *Drosophila* and vertebrates. We report here an additional major role for *Drosophila* Tup in specifying dorsal muscles. Tup is expressed in the four dorsal muscle progenitors (PCs) and *tup*-null embryos display a severely disorganized dorsal musculature, including a transformation of the dorsal DA2 into dorsolateral DA3 muscle. This transformation is reciprocal to the DA3 to DA2 transformation observed in *collier* (*col*) mutants. The DA2 PC, which gives rise to the DA2 muscle and to an adult muscle precursor, is selected from a cluster of myoblasts transiently expressing both Tinman (Tin) and Col. The activation of *tup* by Tin in the DA2 PC is required to repress *col* transcription and establish DA2 identity. The transient, partial overlap between Tin and Col expression provides a window of opportunity to distinguish between DA2 and DA3 muscle identities. The function of Tup in the DA2 PC illustrates how single cell precision can be reached in cell specification when temporal dynamics are combined with positional information. The contributions of Tin, Tup and Col to patterning *Drosophila* dorsal muscles bring novel parallels with chordate pharyngeal muscle development.

KEY WORDS: Tailup/Islet1, Collier/EBF, Tinman/Nkx2.5, Myogenesis, Temporal cascades, *Drosophila*

INTRODUCTION

Understanding how a cell acquires its identity is a fundamental problem in developmental biology. The somatic musculature of the *Drosophila* embryo is a classical model with which to study the regulatory processes that generate cellular diversity in response to positional information provided by embryonic body axes. Each hemi-segment shows a stereotypical arrangement of 30 somatic muscles connected to the exoskeleton, one alary muscle that attaches to the dorsal vessel, which is the equivalent of the heart, and six muscle stem-like cells [known as adult muscle precursors (AMPs)] (Bate and Rushton, 1993; Bate et al., 1991; Figeac et al., 2010; LaBeau et al., 2009). Each skeletal muscle is a syncytium seeded by a founder cell (FC), which undergoes multiple rounds of fusion with fusion-competent myoblasts. FCs are born from the asymmetric division of progenitor cells (PCs), themselves selected by Notch (N)-mediated lateral inhibition from equivalence groups of myoblasts, termed promuscular clusters at distinct positions within the somatic mesoderm (Carmena et al., 1995; Ruiz-Gómez and Bate, 1997). Characterization of selected muscle lineages has shown that muscle shape, size and orientation reflect the expression of specific ‘identity’ transcription factors (iTFs) in each PC and FC (Bate and Rushton, 1993; Baylies et al., 1998; Frasch, 1999; Tixier et al., 2010). Maintaining the expression of a subset of iTFs in one FC, but not its sibling (Ruiz-Gómez et al., 1997; Rushton et al., 1995), and propagation of this iTF code into all nuclei to the growing myofibre (Crozatier and Vincent, 1999; Dubois et al., 2007; Knirr et al., 1999) determine the final identity of each muscle. There are four dorsal muscles (DA1, DA2, DO1, DO2)

and six dorsolateral (DL) muscles (DA3, DO3, DO4, DO5, DT1 and LL1) (Bate and Rushton, 1993; Nose et al., 1998). Detailed studies of the iTF Even-skipped (Eve) expression in the DA1 and DO2 lineages showed that *eve* cis-regulation integrates positional information provided by ectodermal Decapentaplegic (Dpp) and Wingless (Wg) signalling, tissue-specific information provided by the mesodermal transcription factors Twist (Twi) and Tinman (Tin), and Receptor Tyrosine Kinase (RTK) signals (Carmena et al., 1998; Duan et al., 2007; Knirr et al., 1999). Although these pioneering studies have provided a useful framework, there remain many gaps in our understanding of how each PC acquires a specific identity according to its position.

From a genetic screen designed to identifying new muscle identity genes (L.D., unpublished), we found that removal of the LIM homeodomain TF Tailup/Islet1 (Tup) causes a DA3 duplication. Islet1 was first characterized in vertebrates, as a regulator of insulin expression in the pancreas (Karlsson et al., 1990) and in motoneuron specification (Pfaff et al., 1996). Recently, it has been shown to be necessary for second heart field (SHF)-derived cells to populate the heart (Laugwitz et al., 2005; Moretti et al., 2006). *Drosophila tup* was initially characterized for its role in neuronal identity and axon pathfinding (Thor and Thomas, 1997), and was later shown to be required for development of the lymph gland and the specification of cardiac cells (Mann et al., 2009; Tao et al., 2007).

We report here that Tup is expressed in the four dorsal muscles and the alary muscles, and that the dorsal musculature is severely disorganized in *tup*-null embryos. We then focus our studies on the DA2 to DA3 muscle transformation. The DA2 PC is selected from myoblasts transiently expressing both Col and Tin. Tup activation in this PC during this time window is required to repress *col* transcription and impose the DA2 identity. Tup is then required for maintaining its own expression in dorsal muscles beyond the PC stage, suggesting a handover relay mechanism between two, early and late, cis-regulatory modules. Our findings provide a new dynamic view of transcriptional regulation that patterns the

Université de Toulouse 3, Centre de Biologie du Développement, UMR 5547 CNRS and FRBT, 118 route de Narbonne, F-31062 Toulouse cedex 09, France.

*Present address: Department of Biochemistry and Molecular Biophysics, Columbia University, 701 West 168th Street, HHSC 1104 New York, NY 10032, USA

[‡]Authors for correspondence (dubois@cict.fr; vincent@cict.fr)

Drosophila dorsal musculature and novel parallels with specification of pharyngeal muscles in chordates.

MATERIALS AND METHODS

Genetics

Drosophila mutant and transgenic strains were: *tup^{ex4}* (de Navascués and Modolell, 2010), *col¹* (Crozatier et al., 1999), *rp298-lacZ* (Nose et al., 1998), *Mhc-GFP* (Eric Olson and Elisabeth Chen, Dallas, USA), *twist-Gal4* (Baylies and Bate, 1996), *IsletH-GFP*, *UAS-tup* (Thor and Thomas, 1997), *N^{55e11}* (Salzberg et al., 1994), *tin^{EC40}* (Bodmer, 1993) and *Vg-lacZ* (Richard Cripps, Albuquerque, USA). Mutant strains were balanced over marked chromosomes: *Cyo wg-lacZ*; *Cyo dfd-EYFP*; *TM3 twist-lacZ*. *Df(2L)Exel7072* and *Df(2L)Exel7073* were obtained from the Bloomington Stock Center.

Constructions of transgenic reporter lines

The *4_0.9col-moeGFP* construct fuses a *col* genomic DNA fragment that contains 4 kb of upstream sequence and 560 bp of transcribed sequence extending into the Col protein region, in frame with a D-moesin GFP fusion (Polesello et al., 2002). It was inserted into the *AttP-pS3AG* transgenesis vector (Williams et al., 2008). *tup DME-lacZ* was constructed by placing a PCR-amplified *tup* genomic fragment, FlyBase positions 2L:18897397..18900688 (release: r5.42), in the *attB-inslacZ* vector (Enriquez et al., 2010). *4_0.9col-moeGFP* and *DME-lacZ* were inserted at position 68A4 on the third chromosome by injection into nosC31NLS;attP2 embryos (Bischof et al., 2007; Markstein et al., 2008).

Immunohistochemistry and in situ hybridization

Antibody staining and in situ hybridization with intronic probes were as described previously (Dubois et al., 2007). Primary antibodies were: guinea pig anti-Col (1:5000; Adrian Moore, Saitama-Ken, Japan), anti-L'sc (Stephen Crews, Chapel Hill, USA), anti-SrB (Talila Volk, Rehovot, Israel), mouse anti-Col (1:100) (Dubois et al., 2007), anti-Tup (1:100; Developmental Studies Hybridoma Bank), anti-GFP (1:100; Torrey Pines Biolabs), anti-β-galactosidase (1:1000; Promega), rabbit anti-Tin (1:750; Manfred Frasch, Erlangen, Germany), anti-Nau (1:100; Bruce Paterson, Bethesda, USA), anti-Kr (Ralf Pflanz, Goettingen, Germany), anti-Eve (1:1000) (Kosman et al., 1998), anti-b3-tubulin (1:5000; Renate Renkawitz-Pohl, Marburg, Germany), anti-Zfh1 (1:5000; Alain Garces, Montpellier, France), anti-Vg (Andrew Simmonds, Edmonton, Canada) and rat anti-Runt (1:300) (Kosman et al., 1998). Secondary antibodies were: Alexa Fluor 488- and 555-conjugated antibodies (1:300; Molecular Probes), and biotinylated goat anti-mouse (1:1000; Vector Laboratories). The *tup* intronic probe spans 2 kb of the first intron. Confocal sections were acquired on a Leica SP5 microscope at 40× magnification and 1024/1024 pixel resolution. Images were assembled using Photoshop software.

tup CRM predictions

tup CRM predictions were based on the positions of Tin, Twi and Mef2 in vivo binding sites (Sandmann et al., 2007; Finzen et al., 2009). These data were accessed via the Eileen Furlong's lab webpage (http://furlonglab.embl.de/data/browse_chip_mod/).

RESULTS

tup: a new muscle identity transcription factor

Our knowledge of the regulatory program that controls the exquisite pattern of *Drosophila* larval muscles is still fragmentary. To identify new players, we undertook a genetic screen, starting from a collection of small overlapping deficiencies covering the second chromosome and using Col expression as a read-out. Two overlapping deficiencies, *Df(2L)Exel7072* and *Df(2L)Exel7073*, revealed a duplication of the Col-expressing DA3 muscle in most segments (81%, *n*=249; 82%, *n*=230, respectively; Fig. 1A-D). The genomic region deleted in both deficiencies includes two genes, *short spindle 3* (*ssp3*) and *tup* (Fig. 1A). *tup^{ex4}* null embryos (de Navascués and Modolell, 2010) show a DA3 duplication in 78% of segments (*n*=102) (Fig. 1E-E') and a

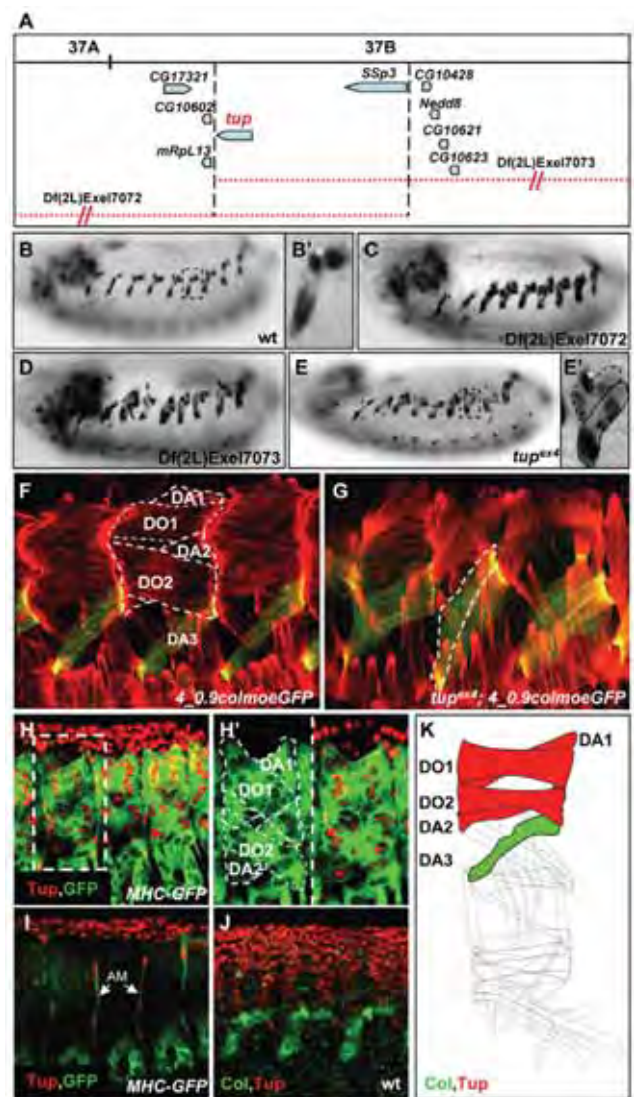


Fig. 1. *tup*: a muscle identity gene. (A) The 37A-37B chromosomal interval and overlapping *Df(2L)Exel7072* and *Df(2L)Exel7073* deficiencies (dotted red lines). The genes surrounding and within the deleted genomic region are indicated by blue arrows. (B-E') Lateral views of stage 15 embryos stained for Col. (B,B') Wild type (wt), (C) *Df(2L)Exel7072*, (D) *Df(2L)Exel7073* and (E,E') *tup^{ex4}* homozygous embryo. (B',E') Enlarged views of one segment. A duplication of the DA3 muscle is observed in absence of *tup*. Asterisks indicate dorsal md neurons that express Col. (F,G) Dorsolateral views of stage 16 *4_0.9colmoeGFP* embryos stained with phalloidin to visualize muscle fibres. Three segments are shown. (F) Wild-type embryo; the contours of the DA1, DA2, DO1 and DO2 dorsal muscles can be unambiguously identified (dotted white lines); GFP marks the DA3 muscle. (G) *tup^{ex4}* embryo; the dorsal muscles are ill-formed and GFP is expressed in a second, DA3-like muscle (framed). (H-I) Dorsolateral views of two segments of stage 15 Myosin Heavy Chain (MHC)-GFP embryos stained for Tup (red) and GFP (green), showing Tup expression in the four dorsal muscles (H) and the alary muscles (white arrows in I). (H') High-magnification view of the region framed in H; left, GFP staining outlining the dorsal muscles (dotted white circled). (J,K) Tup (red) and Col (green) expression at stage 15; schematic in K.

similar phenotype is observed in hemizygous *tup^{ex4}/Df(2L)Exel7072* embryos (not shown). Lack of *tup* is therefore responsible for the DA3 duplication phenotype

observed in Df(2L)-Exel7072 and Df(2L)Exel7073 embryos. To better visualize the DA3 muscle, we expressed a *moe*-GFP fusion protein under control of a DA3-specific cis-regulatory module (CRM) (4_0.9*colmoe*GFP). Double phalloidin and GFP staining confirmed the transformation of one dorsal muscle into a DA3-like muscle in *tup^{ex4}* embryos (Fig. 1F,G). Phalloidin staining also revealed that all dorsal muscles are affected to some extent in *tup^{ex4}* embryos, with individual muscles often not recognisable from segment to segment, showing that *tup* activity is crucial for patterning the dorsal musculature.

Tup contributes to the iTF code of dorsal muscle PCs

Each muscle can be uniquely identified by its position and shape at stage 15. Staining for *Tup* showed its expression in the four dorsal muscles (Fig. 1H,H'), complementary to DA3 *Col* expression (Fig. 1H,J,K), in addition to alary muscles (Fig. 1I) (Tao et al., 2007). As muscle identity is conferred by activation of specific combinations of iTFs in each PC, we looked at *Tup* expression at the PC stage. As general marker of PCs/FCs, we used *rp298lacZ* (Beckett and Baylies, 2007; Enriquez et al., 2012; Nose et al., 1998). Staining of late stage 11 *rp298lacZ* embryos for *lacZ* and either *Tup* or *Col*, showed that *Tup* expression is expressed in the four dorsal-most PCs (Fig. 2A). The dorsal FCs have been positioned relative to each other and identified, based on the expression of specific iTFs, except for DA2 (Beckett and Baylies, 2007). We tentatively identified the DA2 PC as a cell expressing *Nautilus* (*Nau*), but not *Col*, and located immediately dorsal to the DA3/DO5 PC (Enriquez et al., 2010; Enriquez et al., 2012). Double staining for *Tup* and either *Eve* or *Kr* confirmed that this cell expresses *Tup*, and is *Eve* and *Kr* negative (Fig. 2D,E). Thus, *Tup* is expressed in the four dorsal PCs and contributes to their specific iTF code. *Kr* expression is reduced in the DO1 muscle in *tup* hypomorphic mutants (Mann et al., 2009). Accordingly, *Kr* expression is completely lost in *tup^{ex4}* embryos, confirming that it crucially depends upon *Tup* (data not shown). On the contrary, *Eve* and *Runt* remain expressed in the DA1 and DO2 muscles, respectively, although an abnormal clustering of *Runt*-expressing nuclei correlates well with the abnormal morphology of the DO2 muscle in *tup* mutant embryos (supplementary material Fig. S1). Together, these data indicate that *Tup* acts both upstream of some iTFs and in combination with others in specifying dorsal muscle identities.

The DA2 muscle originates from a mixed muscle/AMP lineage

Only two dorsal muscle lineages, DA1 and DO2, were previously established, based on *Eve* expression. Both are mixed muscle/*Eve* pericardial cell (EPC) precursor lineages (Carmena et al., 1998; Speicher et al., 2008). To establish the DA2 lineage, we used a 7 kb *tup* cis-regulatory region, *IsletH*, which was initially characterized for its ability to drive GFP expression in a subset of post-mitotic motoneurons and interneurons (Thor and Thomas, 1997). We found that *IsletH* is active at stage 11 in the DA2 PC, as identified by its position, immediately dorsal to the DA3/DO5 PC (Fig. 2G). Following *IsletH*-GFP expression showed that the DA2 PC divides at stage 12. Double staining for GFP and *Zfh1*, a general AMP marker (Figeac et al., 2010; Sellin et al., 2009), established that this division is asymmetric, with one daughter cell becoming an AMP (Fig. 2H). The DA2 PC thus gives rise to a mixed somatic muscle/AMP lineage. Staining of stage 16 embryos for *Zfh1* and GFP showed that the DA2 sibling AMP is the dorsal-

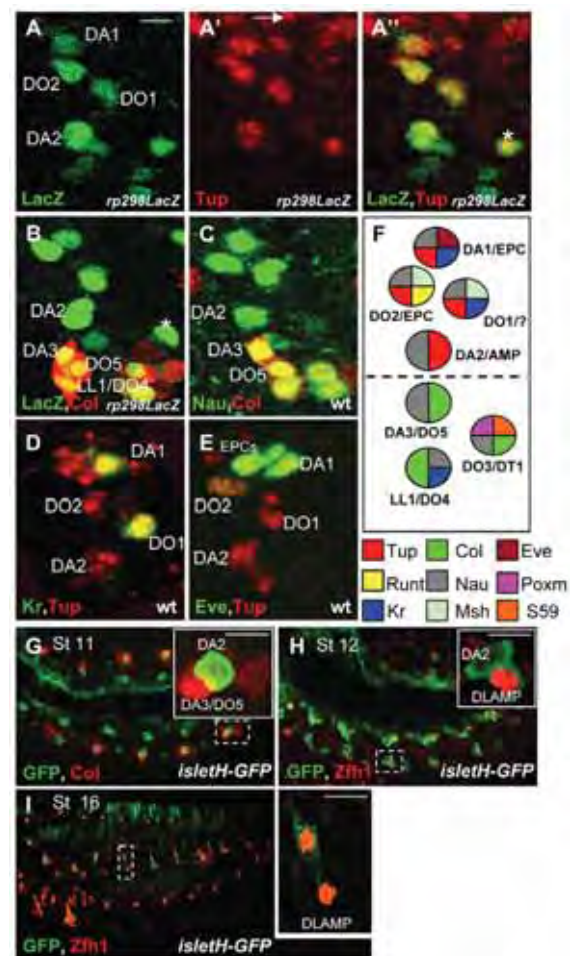


Fig. 2. *Tup*: the iTF code of dorsal PCs and the DA2/AMP lineage.

(A-G) Staining of late stage 11 embryos for different iTFs. (A-B) *rp298lacZ* embryos stained for *lacZ* (green) and (red), either *Tup* (A-A') or *Col* (B). *Tup* is expressed in cardioblasts (arrow in A'), the DA1, DO2, DO1 and DA2 PCs and a cell of unknown identity (asterisk in A',B). *Col* is expressed in the DA3/DO5, DO3/DT1 and LL1/DO4 PCs. (C-E) Wild-type embryos stained for *Nau* (green) and *Col* (red) (C), *Tup* (red) and (green) *Kr* (D) or *Eve* (E), as indicated in each panel. The dorsal region of one abdominal segment is shown. The DA2 PC expresses *Tup*, but neither *Eve* nor *Kr*. (F) Schematic map of the dorsal and DL PCs, at late stage 11, with the iTF code colour-coded. (G-I) Staining of stages 11, 12 and 16 *isletH*-GFP embryos for GFP (green) and either *Col* (G) or *Zfh1* (H,I) (red). (G) The DA2 PC expresses *IsletH*-GFP and divides into the DA2 FC and the dorsal-most DLAMP (H,I). Insets in G-I are enlarged views of framed areas. Scale bars: 10 μ m.

most DLAMP (Fig. 2I). This AMP still expresses *Zfh1* in *tup* mutant embryos, suggesting that *tup* is not required for its specification (not shown).

Tup is required to specify the DA2 muscle identity, via *Col* repression

To determine which muscle ectopically activates *Col* expression and adopts a DA3 morphology upon loss of *tup*, we double stained wild-type and *tup^{ex4}* embryos for *Col* and *Vestigial* (*Vg*). *Vg* marks the DA1, DA2, DA3 and LL1 muscles (Bate et al., 1993). In wild-type embryos, only the DA3 muscle expresses *Col* (Fig. 3A,A'). In

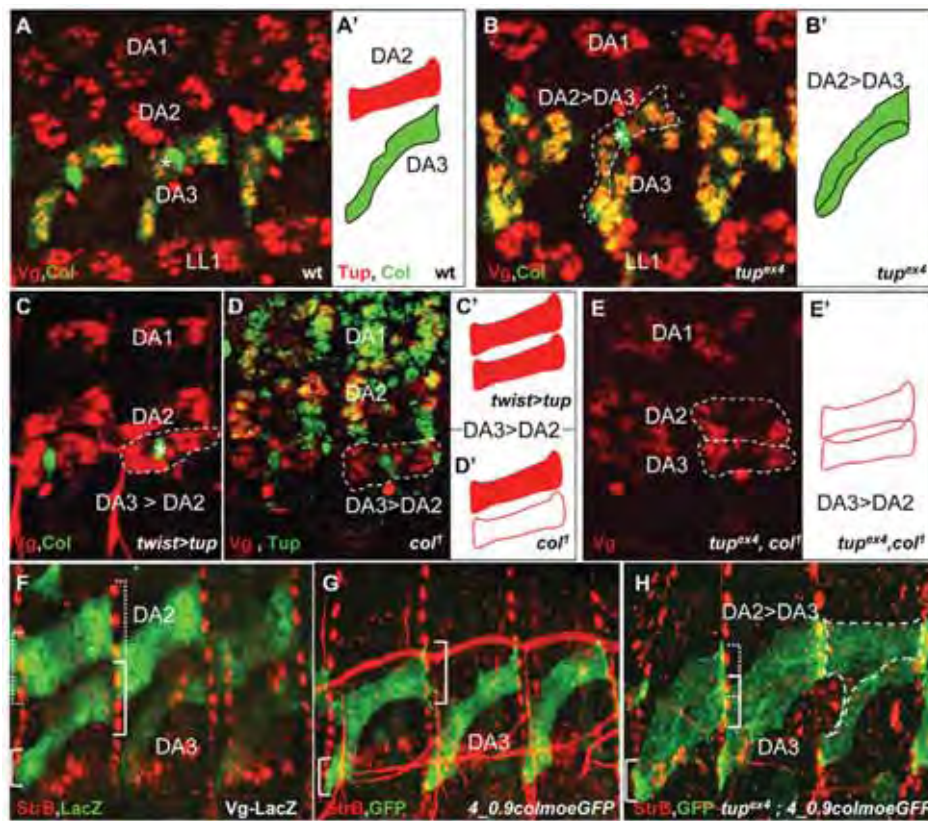


Fig. 3. Tup activity distinguishes between the DA2 and DA3 fates by repressing Col. Dorsolateral views of stage 15 embryos. (A) Wild-type, (B) *tup^{ex4}* (C) *twist>tup*, (D) *col¹* and (E) *tup^{ex4}, col¹* embryos stained for Vg (red) and either Col (A-C) or Tup (D) (green). (A) Vg labels the DA1, DA2, DA3 and LL1 muscles, and Col the DA3 muscle. (B) The DA2 muscle expresses Col and adopts DA3 (DA2>DA3) morphology. (C) The DA3 muscle expresses Tup and adopts DA2 (DA3>DA2) morphology. (D) The DA3 muscle (dashed circled) adopts DA2 morphology without expressing Col. (E) The DA2 and DA3 muscles display DA2 morphology. (A', B', C', E') Schematic drawings of the DA2 and DA3 phenotypes with Tup and Col expression in red and green, respectively. (F) Vg-*lacZ* embryo stained for *lacZ* (green) and SrB (red), showing the anterior and posterior attachment sites of the DA2 (dotted white brackets) and DA3 (white brackets) muscles. (G, H) *4_0.9colmoeGFP* (G) and *tup^{ex4}; 4_0.9colmoeGFP* (H) embryos stained for GFP (green) and SrB (red). In *tup^{ex4}* embryos, the DA3 and DA2>DA3 attachment sites overlap; triangular-shaped muscles (white dotted circled) are observed. Three abdominal segments are shown in A, B, F-H; two abdominal segments are shown in C-E. The white asterisks in A-C indicate Col expression in multidendritic neurons.

tup^{ex4} embryos, the DA2 muscle also expresses Col, correlating with its abnormal morphology, showing that it is transformed into DA3 in absence of *tup* activity (Fig. 3B, B').

The DA2>DA3 transformation observed in *tup* mutants is also observed upon Col ectopic expression in the entire mesoderm (Twi>Col) (Enriquez et al., 2012). This phenotype similarity suggested that Tup could antagonize Col activity. To test this possibility, we overexpressed Tup in the entire mesoderm (Twi>Tup embryos). Staining for Col and Vg showed a loss of Col expression in the DA3 muscle (Fig. 3C) and high frequency of DA3>DA2 transformations (88% of segments, $n=77$), similar to a *col* loss-of-function phenotype (Fig. 3D). We conclude that Tup is able to repress Col expression in the DA3 muscle, correlating with the DA3>DA2 identity shift (Fig. 3C', D'). The opposite DA2>DA3 and DA3>DA2 transformations observed in *tup* mutants or *col* gain-of-function and *col* mutants or *tup* gain-of-function, respectively, show that Col and Tup distinguish between the DA2 and DA3 muscle identities. Of note, however, Col is activated in the DA2>DA3 muscle in *tup^{ex4}* embryos, whereas the DA3>DA2 transformation in *col* embryos occurs without Tup activation (Fig. 3B', D'). This difference suggested that Tup is not

instructive for the DA2 fate. To test this idea further, we looked at the muscle pattern of *col¹; tup^{ex4}* double mutant embryos. β 3-Tubulin staining showed a cumulative phenotype; that is, dorsal muscle defects and a lack of DA3 muscle (supplementary material Fig. S2). Staining *col¹; tup^{ex4}* embryos for Vg confirmed that, in absence of *tup* and *col* activity, the DA2 and DA3 muscles display a DA2 morphology (Fig. 3E, E'). It indicates that DA2 is default morphology and suggests that DA2 iTFs other than Tup remain to be identified.

To describe more precisely the DA2>DA3 morphological transformation in *tup* mutants, we compared the epidermal attachment sites of the DA2 and DA3 muscles in wild-type and in *tup^{ex4}* embryos. Staining Vg-*lacZ* embryos for *lacZ* and Stripe B (StrB), a marker of tendon cells (Volk and VijayRaghavan, 1994; Volohonsky et al., 2007), indicated that the wild-type DA2 and DA3 attachment sites are not overlapping (Fig. 3F). StrB staining of wild-type and *tup^{ex4}*; *4_0.9colmoeGFP* embryos showed an overlap between the wild-type DA3 and *tup^{ex4}* DA2>DA3 insertion sites (Fig. 3G, H). Thus, in absence of *tup*, the DA2 muscle targets the DA3 tendon cells. Some partial DA2>DA3 transformations were also observed in *tup* mutant

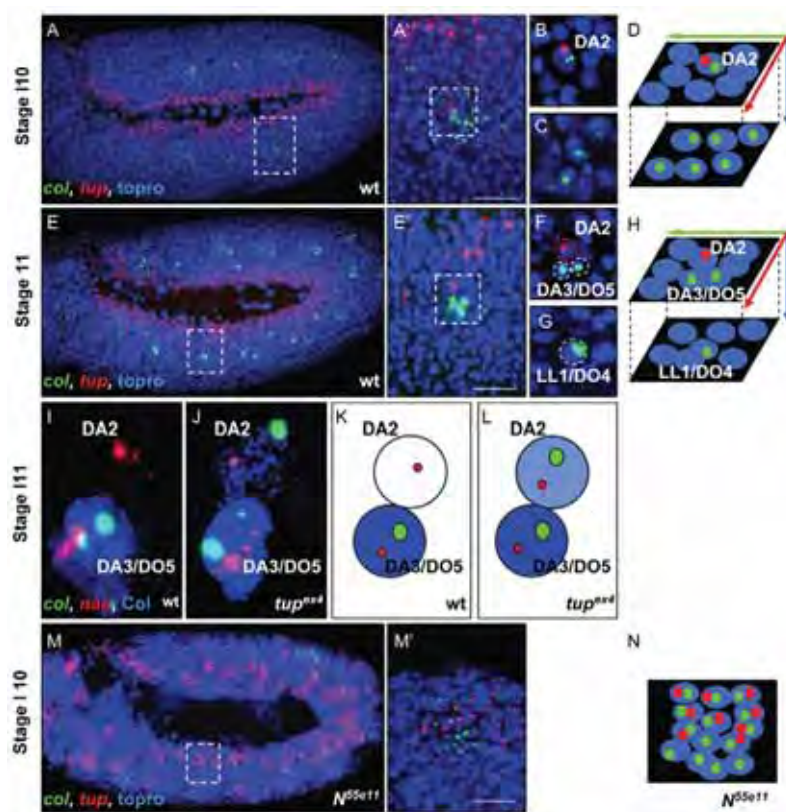


Fig. 4. Transcriptional repression of *col* by *Tup* in the DA2 PC. (A-H) *col* (green) and *tup* (red) primary transcripts. Nuclei are stained by Topro (blue). (A,A') Late stage 10 embryo; *tup* is detected in a single cell among a *col*-positive cluster; outlined region is enlarged in A'. (B-D) High-magnification views of two z-sections of the squared region in A'. The DA2 PC singles out from the *col* cluster, shown schematically in D. (E,E') Stage 11 embryo. The outlined region in E is enlarged in E'. (F-H) High-magnification views of different z-sections of the squared region in E', shown schematically in H. The DA2 PC maintains *tup* transcription (red circled in F). *col* is transcribed in the DA3/DO5 and LL1/DO4 PCs (white circles in F and G, respectively). (D,H) The green, red and blue arrows indicate the anteroposterior, dorsoventral and mediolateral axes, respectively. (I-L) Late stage 11 embryos. *col* (green) and *nau* (red) primary transcripts in wild-type (I) and *tup^{ex4}* (J) embryos. In *tup^{ex4}* embryos, *col* transcription and Col protein (blue) are detected in the DA2PC, shown schematically in K,L. (M-N) *col* (green) and *tup* (red) primary transcripts in late stage 10 *N^{55e11}* embryos. Nuclei are stained by Topro (blue). M' is an enlargement of the outlined region in M. Several myoblasts co-transcribe *col* and *tup*, shown schematically in N. Scale bar: 10 μ m.

embryos, with triangular shaped muscles (Fig. 3H). This phenotype underlines the existence of two successive steps in the attachment of elongating muscles to specific tendon cells, a transient attachment, followed by a selection of the final attachment sites, with the second step being tightly regulated by the iTF code (Enriquez et al., 2012). Whether iTFs other than Col and Tup contribute to the robustness of the selection of DA2 versus DA3 attachment sites remains an unresolved issue.

Tup represses *col* transcription in the DA2 progenitor

Temporal aspects of Col expression in the DL PCs showed that they are specified sequentially. The DA3/DO5 PC is specified first, followed by the DO3/DT1 PC (only in the A1-A7 segments; not further considered here) and, finally, the LL1/DO4 PC (Enriquez et al., 2012). We previously noticed that the Nau-positive cell now identified as the DA2 PC (Fig. 2), singles out from the Col-positive cluster earlier than the DA3/DO5 PC and transiently expresses low amounts of Col (Enriquez et al., 2012). This observation suggested that the repression of Col by Tup was exerted in the DA2 PC, raising the question of the relative kinetics of *tup* and *col* transcription in this cell. To address this question, we used in situ hybridization with intronic probes to detect *tup* and *col* nascent transcripts. Double in situ hybridization revealed that *tup* is strongly transcribed in only one cell among the cluster of myoblasts that transcribe *col* at stage 10 (Fig. 4A,A'). Examination of single confocal sections showed that this unique *tup*- and *col*-positive cell has singled out and adopted a more external position than the rest of the cluster (Fig. 4B-D). By stage 11, however, whereas *tup* remains transcribed, *col* transcription is not detected anymore in this cell, unlike in the DA3/DO5 FCs and LL1/DO4 PC, which is

singling out at this stage (Fig. 4E-H). We conclude that *col* and *tup* are co-transcribed in the DA2 PC, prior to *col* extinction. To determine whether *tup* maintenance and *col* extinction are linked, we compared *col* transcription in the DA2 PC in wild-type and *tup^{ex4}* embryos, using *nau* transcription as an internal reference. In wild-type embryos, *col* and *nau* transcripts are detected in the DA3/DO5 PC, identified by the accumulation of Col protein, while only *nau* transcripts are detected in the DA2 PC (Fig. 4I,L). In *tup^{ex4}* mutant embryos, *nau* and *col* transcripts, and Col protein are detected in the DA2 PC (71% of segments; $n=60$) in addition to the DA3/DO5 PC (Fig. 4J,L), showing that Tup represses *col* transcription in the DA2 PC. Promuscular *col* transcription at stage 10 is not affected in *twi>tup* embryos (supplementary material Fig. S3), confirming that Tup repression of *col* is exerted at the PC stage. The lower level of Col protein in the DA2, compared with DA3 PC, in *tup* mutant embryos (Fig. 4J), suggests, however, another level of *col* regulation.

Eve and S59 expression in selected lineages showed that PC-specific expression of some iTFs is established in two steps: first, activation in an equivalence group of competent cells, followed by restriction to PC(s), which involves N-mediated lateral inhibition (Buff et al., 1998; Carmena et al., 2002; Knirr and Frasch, 2001). *col* expression in the DL PCs follows this scheme (Fig. 4A-F) (Crozatier and Vincent, 1999; Enriquez et al., 2010). To determine whether *tup* expression in the DA2 lineage is first activated in an equivalence group, we looked at *tup* transcription in N mutant (*N^{55e11}*) embryos where lateral inhibition is defective. In absence of N, several myoblasts co-transcribe *col* and *tup* (Fig. 4M,M'), revealing the existence of a group of cells from which the DA2 PC is selected. Only the dorsal-most cells of the *col* cluster transcribe *tup* (Fig. 4M,N), providing an explanation of why *col* repression

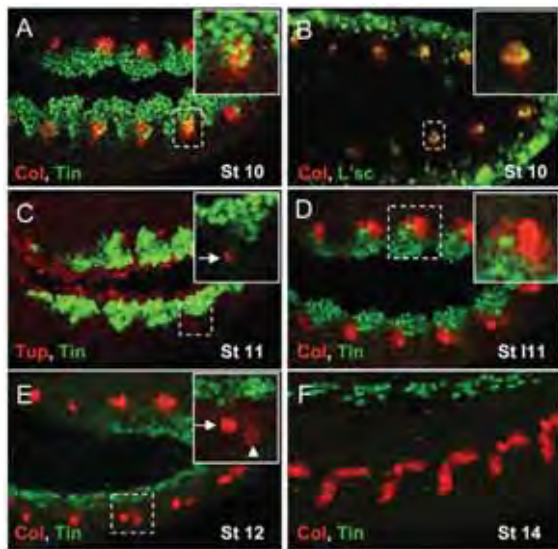


Fig. 5. The DA2 PC singles out from a promuscular cluster transiently co-expressing Col and Tin. (A,B) Col (red) and Tin (A) or L'sc (B) (green) expression in stage 10 embryos. The small region of Tin/Col expresses L'sc. (C) Tup (red) and Tin (green) expression in stage 11 embryos at the onset of dorsal restriction of Tin expression. Tup labels the DA2 PC (white arrow). Other Tup- and Tin-positive cells include dorsal muscle PCs and cardiac progenitors (Mann et al., 2009). (D-F) Tin (green) and Col (red) expression at stage 11 (D), 12 (E) and 14 (F). (D) Dorsal restriction of Tin expression and ventral expansion of Col expression is observed. (E) The DA3/DO5 (arrow) and DO3/DT1 (arrowhead) PCs have singled out; insets in A-D correspond to single confocal sections. (F) Tin expression in the heart.

by Tup only occurs in the DA2 PC. In turn, it raised the question of which mechanisms control the different registers of *tup* and *col* activation.

The DA2 progenitor is issued from a promuscular cluster transiently expressing Tin

Tin is a major regulator of cell fate in the dorsal mesoderm (Bodmer, 1993; Liu et al., 2009), which shows a dynamic expression pattern (Azpiazu and Frasch, 1993; Bodmer, 1993; Johnson et al., 2011; Xu et al., 1998; Yin et al., 1997). Tin is first expressed in the entire trunk mesoderm, followed by upregulation in the dorsal mesoderm, in response to Dpp signalling from the dorsal ectoderm (stage 10, Fig. 5A). It is during this second phase that Tin has been proposed to specify different dorsal mesoderm fates. Later, Tin expression is restricted to cardiac progenitors (late stage 11, Fig. 5D), before directing heart cell diversification (Fig. 5E,F). Of specific interest here, is the relationship between Tin, Col and Tup expression during the second phase of Tin expression. At stage 10, a small group of myoblasts express both Col, Tin and the proneural protein Lethal of scute (L'sc) (Fig. 5A,B), suggesting that it corresponds to the equivalence group of *tup/col* transcribing myoblasts revealed in *N* mutant embryos (Fig. 4N). Accordingly, residual Tin is detected in the DA2 PC (Fig. 5C). The overlap between Tin and Col expression is transient, however, because at late stage 11, while the Col-expressing cluster has expanded ventrally, Tin expression has retracted dorsally (Johnson et al., 2011; Reim and Frasch, 2005), such that Col and Tin are now

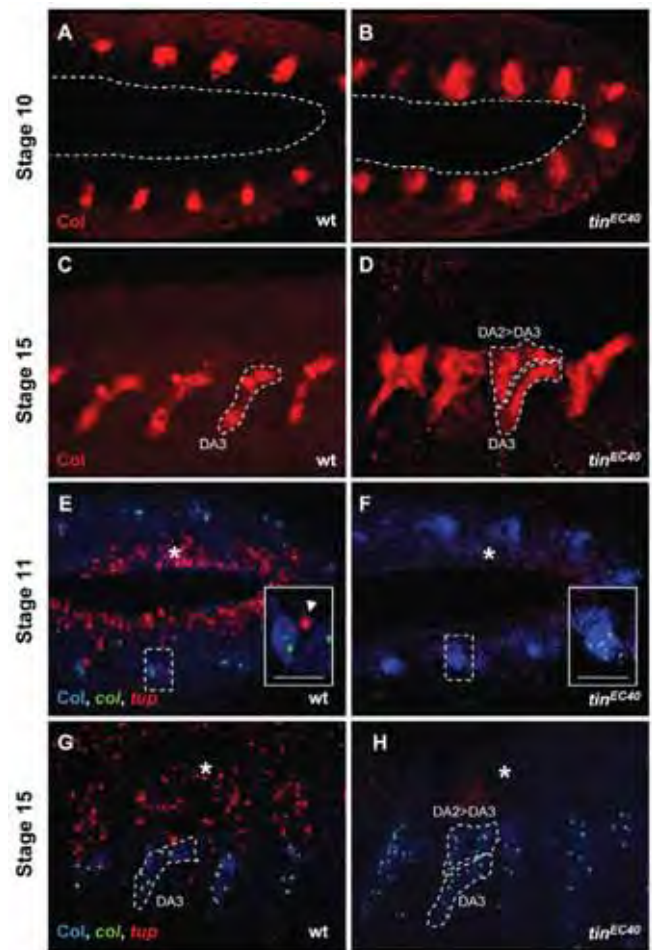


Fig. 6. Tin is required for *tup* transcription in the DA2 PC and DA2 identity. (A-D) Col expression in wild-type (A,C) and *tin*^{EC40} (B,D) embryos. In A,B, the dotted-white line delineates the amnioserosa. In *tin* mutants, Col expression extends dorsally (B); the DA3 muscle (dotted-white circled in one segment) is duplicated (D). (E-H) *col* (green) and *tup* (red) primary transcripts in wild-type (E,G) and *tin*^{EC40} (F,H) embryos. Col protein is in blue. Stage 11 *tup* transcription in the dorsal mesoderm (white asterisk) and DA2 PC (white arrow), is lost in *tin*^{EC40} embryos. (H) *tup* transcription in muscles (asterisk) is lost, whereas *col* is ectopically transcribed in the DA2>DA3 muscle. Scale bars: 10 μ m.

expressed in complementary domains (Fig. 5D). As a consequence, the DA3/DO5 (and DT1/DO3) progenitors, which are born more ventrally, and later than the DA2PC, are selected from cells that do not anymore express Tin (Fig. 5E).

Tup acts downstream of Tin in the DA2 PC

Tin role in patterning the dorsal muscles has not been studied in detail, except for its requirement for Eve expression in the DO2 and DA1 progenitors (Knirr and Frasch, 2001). In silico search and chip-on-chip experiments suggested that Tin could directly regulate the expression of several iTFs, including *col* (Enriquez et al., 2010; Liu et al., 2009; Philippakis et al., 2006). Promuscular Col expression is not lost in *tin*^{EC40}-null embryos, however, but extends dorsally, to reach the dorsal edge of the embryo proper and abut the amnioserosa, which is expanded in these mutants (Bodmer, 1993) (Fig. 6A,B). Thus, during its first

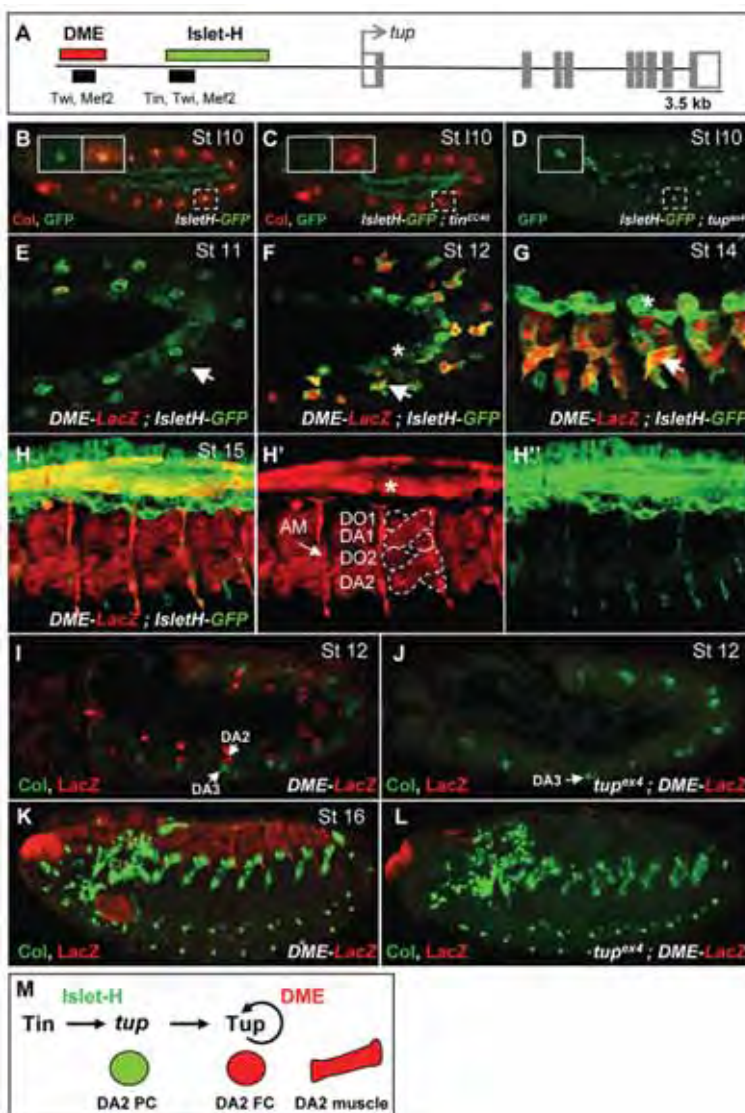


Fig. 7. Two separate cis-regulatory regions control *tup* transcription in the DA2 lineage. (A) Schematic view of the *tup* genomic region. The transcription start is indicated by an arrow; coding exons are in grey. The IsletH (Thor and Thomas, 1997) and DME cis-regulatory regions are represented by green and red boxes, respectively. Positions of meso-ChIP CRMs are indicated by black boxes. (B-D) IsletH-GFP expression in the DA2 PC at late stage 10 (B), is lost in *tin^{EC40}* (C) but not in *tup^{ex4}* (D) embryos. (E-H') *lacZ* (red) and GFP (green) expression in DME-*lacZ*/IsletH-GFP embryos. (E) DME activity in the DA2 PC (white arrow) is first detected at stage 12. (G) IsletH and DME activity in elongating dorsal muscles, including the DA2 muscle (white arrowhead); IsletH is also expressed in heart cells (asterisk). (H-H') Only DME activity is detected in dorsal muscles in stage 15, in addition to alary muscles (white arrow) and cardiomyocytes (white asterisk). (I, L) DME-*lacZ* (red) and Col (green) expression at stage 12 (I, J) and 15 (K, L). DME is inactive in the DA2 PC (J), and dorsal and alary muscles (L) in *tup^{ex4}* embryos. (M) Scheme of the *tup* CRM handover, allowing maintenance of *tup* expression in dorsal muscles, independently of Tin.

phase of expression, Tin sets the dorsal limit of Col expression, via a mechanism that remains unknown. At stage 15, Col is ectopically expressed in the DA2 muscle in 56% of *tin^{EC40}* segments ($n=146$), correlating with a DA3-like morphology (Fig. 6C,D). This mutant phenotype, reminiscent of *tup* mutants, prompted us to compare *tup* and *col* transcription between wild-type and *tin^{EC40}* embryos. At stage 11, while *col* is transcribed in several PCs in *tin* mutant embryos, *tup* transcription is strongly reduced in all dorsal cells and lost in the DA2 progenitor (Fig. 6F), indicating that Tin acts upstream of *tup* in specifying DA2 identity. Consistent with this loss of *tup* transcription, *col* remains transcribed in the DA2 muscle throughout differentiation (Fig. 6H). However, it is not expressed in more dorsal muscles where *tup* expression is also lost.

In summary, the staggered patterns of Tin and Col activation in stage 10 embryos define three distinct types of PCs along the DV axis: the DA1, DO1 and DO2 PCs, which express Tin and Tup; the DA2 PC, which expresses Tin, Tup and Col; the three DL PCs, which express Col but not Tin. Both spatial and temporal aspects of Tin, Tup and Col regulation are involved in distinguishing the DA2/AMP and DA3/DO5 identities (Fig. 8).

Two separate cis-regulatory regions control Tup activation and maintenance in the DA2 lineage

The absence of *tup* transcription and IsletH expression in the DA2 PC in *tin^{EC40}* embryos (Fig. 7B,C) suggested that activation of *tup* transcription by Tin is mediated by IsletH, consistent with in vivo binding of Tin (Sandmann et al., 2007) (Fig. 7A). IsletH activity in the DA2 muscle is only transient, however (Fig. 7H), indicating the need for (an)other cis-regulatory region(s) responsible for maintaining *tup* transcription throughout muscle development. To identify other CRMs active in the mesoderm, we profited from data concerning the position of ChIP-validated binding sites of Tin, Twist and Mef2, which define so-called meso-ChIP CRMs (Liu et al., 2009; Sandmann et al., 2007; Zinzen et al., 2009). One such CRM maps to IsletH and overlaps *tup*-F4, a 1.5 kb IsletH subfragment previously shown to drive GFP in the heart, lymph gland and alary muscles (Tao et al., 2007); *tup*-F4 and IsletH give identical GFP patterns in dorsal muscles (data not shown). A second Meso-ChIP CRM located upstream of IsletH binds Twi and Mef-2, but not Tin (Fig. 7A). An overlapping DNA region was retrieved from an in silico search for non-coding sequences selectively enriched for combinations of Twi-, Tin- and Ets-binding

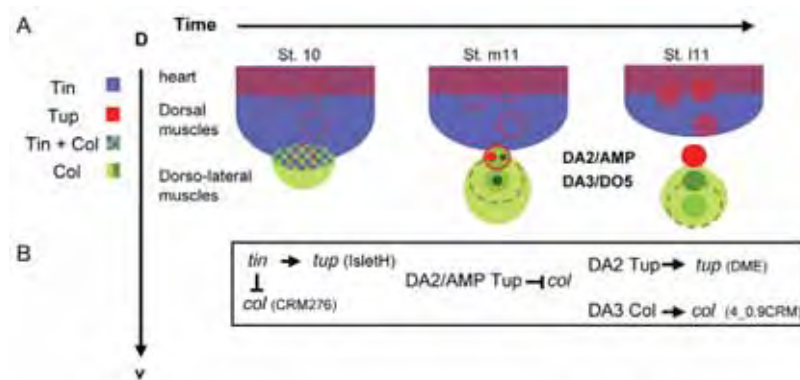


Fig. 8. Model for the sequential specification of the DA2/AMP and DA3/DO5 PCs. (A) Tin (blue) and Col promuscular expression (pale green) partly overlap at stage 10. The early born DA2 PC is specified from the Col- and Tin-expressing myoblasts. At mid stage 11, Tin and Col expression about each other, owing to dorsal regression of Tin expression. Tin-mediated activation of *tup* transcription (red dot) in the DA2 PC, leads to repression of *col* transcription (dark green dot). The second-born DA3/DO5 PC transcribes *col* but not *tup*. At late stage 11, Tin expression becomes limited to cardiac progenitors. Tup and Col are expressed in other dorsal and DL PCs, respectively. (B) *col* promuscular/PC expression is driven by an early CRM (CRM276); *tup* activation by Tin in the DA2 PC is via IsletH. Tup and Col subsequently autoregulate their own expression in the DA2 and DA3 lineages, via the 4_0.9CRM (Enriquez et al., 2010) and DME, respectively.

sites (Philippakis et al., 2006; Warner et al., 2008) (A. Aboukhabil and M. Bulyk, personal communication). Taking into account ChIP-ChIP data, and sequence conservation during evolution, we identified a 3 kb fragment, designated below as DME (dorsal muscle enhancer), that drives *lacZ* expression in heart cells, the dorsal and the alary muscles (Fig. 7A,H; supplementary material Fig. S4). DME starts to be active in the DA2 FC at stage 12 (Fig. 7E,F) and remains active in the dorsal and alary muscles up to stage 16, at a time when IsletH-driven expression is no longer detected (Fig. 7G,H). The sequential activity of IsletH and DME in the DA2 lineage was reminiscent of the sequential activity of an early CRM (CRM276) and a late CRM (4_0.9*col* CRM) in the control of mesodermal *col* transcription (Enriquez et al., 2010). It raised the possibility of a similar relay process, whereby Tup protein expressed in the DA2 PC under control of IsletH, would subsequently auto-regulate its own expression via DME. To test this possibility, we examined IsletH and DME-*lacZ* expression in *tup^{ex4}* embryos and found that, while IsletH remains active (Fig. 7D), DME-*lacZ* expression is completely lost from the DA2 PC (Fig. 7I,J). Staining stage 16 embryos confirmed that DME activity in the dorsal muscles is dependent upon Tup (Fig. 7K,L). In situ hybridization confirmed that *tup* transcription in dorsal muscles is lost in the absence of *tup* (not shown). We conclude that Tup-positive autoregulation, mediated directly or indirectly by DME, is required for *tup* expression in dorsal muscles throughout differentiation (Fig. 7M).

DISCUSSION

Drosophila Tup/Islet plays key roles in specification of neurons (Thor and Thomas, 1997) and cardiac and lymph gland progenitor cells (Mann et al., 2009; Tao et al., 2007). We show here that it is required for somatic muscle development.

Intertwined cascades of iTF regulation in embryonic and adult muscle progenitors

The pattern of *rp298lacZ* expression and the three-dimensional arrangement of FCs distinguished four groups: dorsal, dorsolateral, lateral and ventral (Beckett and Baylies, 2007; Nose et al., 1998).

Whether this topology reflects specific genetic programs has remained unclear. Tup and Col are expressed in the four dorsal and three DL PCs, respectively, supporting the notion of DV regionalization of the somatic mesoderm. This notion was evoked by regional Pox meso (Poxm) expression in most ventral and lateral FCs (Duan et al., 2007). As other known iTFs are only expressed in subsets of dorsal PCs/FCs (Tixier et al., 2010), it raised the possibility that Tup could reside at the top of dorsal iTF cascades. Our data show that this is not the case, as Tup, although required for Kr expression in the DO1 PC (Mann et al., 2009) (data not shown) and for Col repression in the DA2 PC, is not required for expression of Eve, Runt and Vg in the DA1, DO2 and DA2 and DA1 lineages, respectively (Fig. 2; supplementary material Fig. S1). Likewise, Col is required for expression of some iTFs but not others in DL PCs (Enriquez et al., 2012). Together, the patterns of Col, Eve, Kr, Poxm, Runt, Tup and Vg expression in wild-type and *tup* or *col* mutant conditions underline the intertwined, combinatorial nature of transcriptional regulatory networks specifying muscle identity. The DA2 PC gives rise to the DA2 muscle/DL AMP mixed lineage. Each abdominal hemisegment features six AMPs at stereotypical positions (Figeac et al., 2010). Other AMPs originate from mixed lineages, e.g. the ventral VA3/AMP and lateral SBM/AMP lineages. The VA3/AMP and SBM/AMP PCs express Poxm and S59, and Lb, respectively (Jagla et al., 1998; Ruiz-Gómez and Bate, 1997). Tup expression in the DA2/AMP lineage confirms that different AMPs express different iTFs at the time of specification. Whether, as for somatic muscles, the iTF code confers specific properties to each AMP remains an unresolved issue.

Temporal progression of Tin expression translates into unique muscle identities

How PCs born at similar positions in the somatic mesoderm come to express different combinations of iTFs and acquire distinct identities has remained elusive. For example, what distinguishes the fate of the two Eve-expressing PCs, which are sequentially born from the same dorsal cluster, is unknown (Carmena et al., 2002; Carmena et al., 1998). One other example is the expression of S59

and Lb, each in one of two abutting ventrolateral PCs: the LO1/VT1 and SBM PCs. Activation of both Lb and Slo expression in the two PCs is controlled by the same upstream regulator, Org-1 (Schaub et al., 2012). Subsequent reciprocal secondary cross-repression results in exclusive S59 or Lb expression (Junion et al., 2007; Keller et al., 1997), but the nature of the presumed positional bias responsible for the oriented resolution of this cross-repression has not been explored. In the case of the adjacent DA2 and DA3 PCs, we show here that Tup activation by Tin in the DA2 PC is instrumental in distinguishing between DA2 and DA3 identities. The DA2 PC is selected from a small group of cells at the intersection between Tin and Col expression domains. Thus, the relative registers of *tin* and *col* expression along the DV axis provide precise positional information. Another key is timing. The overlap between Tin and Col expression is only transient, such that only the earlier-born Col-expressing PC expresses Tin. This provides a unique temporal window for Tup activation and Col repression (Fig. 8). The transient overlap is due to the dorsal restriction of Tin expression to cardial cells during stage 11 (Yin et al., 1997). This dynamic process is controlled by JAK-STAT signalling activity in the mesoderm, which is itself modulated by Tin activity (Johnson et al., 2011; Liu et al., 2009). The key function of Tup in the DA2 PC, which is to distinguish between two muscle identities, illustrates how cell identity can be specified with single-cell precision when temporal dynamics are combined with positional information.

The CRM handover process: a general mode of muscle identity propagation?

Some iTFs are expressed during all steps of myogenesis, from promuscular stage to muscle attachment (Baylies et al., 1998; Frasch, 1999). Schematically, two major phases of expression can be distinguished: (1) PC specification when multiple iTFs are expressed in different PCs and extensive cross-regulation occurs, leading to FC-specific iTF patterns (Enriquez et al., 2012; Junion et al., 2007; Knirr et al., 1999; Ruiz-Gómez and Bate, 1997) (Fig. 2; supplementary material Fig. S1); and (2) muscle differentiation when the FC pattern is maintained and propagated into the syncytial fibre (Rushton et al., 1995) via transcriptional activation of the iTF code in newly fused FCs (Crozier and Vincent, 1999; Dubois et al., 2007; Knirr et al., 1999). Analysis of *col* regulation in the DA3 lineage showed that these two phases rely on two separate, early (CRM276) and late (4_0.9) CRMs, the activity of the late CRM requiring Col provided under the control of the early CRM. We termed this auto-regulatory loop a CRM handover mechanism (Dubois et al., 2007; Enriquez et al., 2010). Here, we have provided evidence that *tup* transcriptional regulation in the DA2 muscle follows the same rule. On the one hand, *tup* activation by Tin is mediated by an early CRM, IsletH; on the other, *tup* expression is maintained in differentiating muscles via a late CRM, DME, the activity of which depends upon Tup (Fig. 7). We propose that this handover relay mechanism could be a widespread mode of iTF regulation, as it efficiently links early steps of muscle specification in response to positional information with final muscle identity.

New parallels between the transcriptional regulatory networks that pattern *Drosophila* dorsal muscles and vertebrate jaw muscles

Tup and Tin are key components of the transcriptional regulatory cascade that controls early cardiogenesis, with Tin acting to activate Tup, the expression of which then persists after Tin has ceased to be

expressed. We have now shown that a similar cascade operates in the somatic muscle mesoderm. Tup and Tin expression in both the heart and dorsal somatic muscles recalls Nkx2.5 (Tin ortholog) and Islet1 expression in the pharyngeal mesoderm, which contributes to some head muscles and part of the vertebrate heart (Kelly et al., 2001; Lescroart et al., 2010). Nkx2.5 is required for deployment of the SHF (Moretti et al., 2006; Prall et al., 2007) and Islet1 marks SHF progenitors that contribute both to the right ventricle and the arterial pole of the forming heart and a subset of skeletal pharyngeal muscles (Cai et al., 2003; Nathan et al., 2008; Tzahor and Evans, 2011). Similarly, in the simple chordate *Ciona intestinalis*, Nk4 (Tin/NKx2.5) marks the cardio-pharyngeal mesoderm at the origin of the heart and atrian siphon muscles (ASMs) that are evocative of vertebrate pharyngeal muscles. Islet-expressing cells also contribute to ASMs (Stolfi et al., 2010; Tolkin and Christiaen, 2012), suggesting an evolutionarily conserved link between cardiac and pharyngeal muscle development. Interestingly, the ascidian Col/EBF ortholog Ci-COE, is expressed in ASM precursors and is a crucial determinant of the ASM fate (Stolfi et al., 2010), reminiscent of *Xenopus* XCoe2 expression and requirement in pharyngeal arches for aspects of jaw muscle development (Dubois et al., 1998; Green and Vetter, 2011). It is now well established that distinct genetic networks govern skeletal myogenesis in the vertebrate head and trunk (Bothe and Dietrich, 2006; Sambasivan et al., 2011). The repertoire of TFs differentially deployed in the head mesoderm includes Tbx1, the *Drosophila* ortholog of which, Org-1, has recently been shown to act as a muscle iTF (Schaub et al., 2012). Tin/Nkx2.5, Tup/Islet1, Org-1/Tbx1 and Col/EBF may thus be part of a repertoire of transcription factors co-opted and diversified to regulate muscle patterning in *Drosophila* trunk and head muscle patterning in chordates.

Acknowledgements

We thank the Bloomington Stock Center and colleagues for *Drosophila* strains and antibodies; Martha Bulyk for sharing unpublished PhylCRM predictions; and Laetitia Bataillé, Caroline Monod, David Cribbs and Stéphane Zaffran for critical reading of the manuscript. We acknowledge the help of Brice Ronsin, Toulouse RIO Imaging platform, and Frédéric Luce and Julien Favier for maintenance of fly stocks.

Funding

This work was supported by Centre National de la Recherche Scientifique (CNRS) and Ministère de la Recherche et de la Technologie (MRT), Université Paul Sabatier, Association Française contre les Myopathies (AFM) [14859-SR MYOLOGIE to L.D.] and Fondation pour la Recherche Médicale (FRM) [DEQ20090515429 to A.V.]. H.B. and J.E. were supported by fellowships from MRT (H.B.), and from FRM and AFM (J.E.).

Competing interests statement

The authors declare no competing financial interests

Supplementary material

Supplementary material available online at <http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.083410/-DC1>

References

- 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.
- 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. *Dev. Suppl.* **1993**, 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 Ruiz Gomez, M. (1998). Myogenesis: a view from *Drosophila*. *Cell* **93**, 921-927.

- Beckett, K. and Baylies, M. K. (2007). 3D analysis of founder cell and fusion competent myoblast arrangements outlines a new model of myoblast fusion. *Dev. Biol.* **309**, 113-125.
- Bischof, J., Maeda, R. K., Hediger, M., Karch, F. and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc. Natl. Acad. Sci. USA* **104**, 3312-3317.
- Bodmer, R. (1993). The gene tinman is required for specification of the heart and visceral muscles in *Drosophila*. *Development* **118**, 719-729.
- Bothe, I. and Dietrich, S. (2006). The molecular setup of the avian head mesoderm and its implication for craniofacial myogenesis. *Dev. Dyn.* **235**, 2845-2860.
- Buff, E., Carmena, A., Gisselbrecht, S., Jiménez, F. and Michelson, A. M. (1998). Signalling by the *Drosophila* epidermal growth factor receptor is required for the specification and diversification of embryonic muscle progenitors. *Development* **125**, 2075-2086.
- Cai, C. L., Liang, X., Shi, Y., Chu, P. H., Pfaff, S. L., Chen, J. and Evans, S. (2003). Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev. Cell* **5**, 877-889.
- Carmena, A., Bate, M. and Jiménez, 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., Gisselbrecht, S., Harrison, J., Jiménez, F. and Michelson, A. M. (1998). Combinatorial signaling codes for the progressive determination of cell fates in the *Drosophila* embryonic mesoderm. *Genes Dev.* **12**, 3910-3922.
- Carmena, A., Buff, E., Halfon, M. S., Gisselbrecht, S., Jiménez, F., Baylies, M. K. and Michelson, A. M. (2002). Reciprocal regulatory interactions between the Notch and Ras signaling pathways in the *Drosophila* embryonic mesoderm. *Dev. Biol.* **244**, 226-242.
- Crozatier, M. and Vincent, A. (1999). Requirement for the *Drosophila* COE transcription factor Collier in formation of an embryonic muscle: transcriptional response to notch signalling. *Development* **126**, 1495-1504.
- Crozatier, M., Valle, D., Dubois, L., Ibsouda, S. and Vincent, A. (1999). Head versus trunk patterning in the *Drosophila* embryo; collier requirement for formation of the intercalary segment. *Development* **126**, 4385-4394.
- de Navascués, J. and Modolell, J. (2010). The pronotum LIM-HD gene tailup is both a positive and a negative regulator of the proneural genes achaete and scute of *Drosophila*. *Mech. Dev.* **127**, 393-406.
- Duan, H., Zhang, C., Chen, J., Sink, H., Frei, E. and Noll, M. (2007). A key role of Pox meso in somatic myogenesis of *Drosophila*. *Development* **134**, 3985-3997.
- Dubois, L., Bally-Cuif, L., Crozatier, M., Moreau, J., Paquereau, L. and Vincent, A. (1998). XCo2, a transcription factor of the Col/Olf-1/EBF family involved in the specification of primary neurons in *Xenopus*. *Curr. Biol.* **8**, 199-209.
- Dubois, L., Enriquez, J., Daburon, V., Crozet, F., Lebreton, G., Crozatier, M. and Vincent, A. (2007). Collier transcription in a single *Drosophila* muscle lineage: the combinatorial control of muscle identity. *Development* **134**, 4347-4355.
- Enriquez, J., Boukhatmi, H., Dubois, L., Philippakis, A. A., Bulyk, M. L., Michelson, A. M., Crozatier, M. and Vincent, A. (2010). Multi-step control of muscle diversity by Hox proteins in the *Drosophila* embryo. *Development* **137**, 457-466.
- Enriquez, J., de Taffin, M., Crozatier, M., Vincent, A. and Dubois, L. (2012). Combinatorial coding of *Drosophila* muscle shape by Collier and Nautilus. *Dev. Biol.* **363**, 27-39.
- Figeac, N., Jagla, T., Aradhya, R., Da Ponte, J. P. and Jagla, K. (2010). *Drosophila* adult muscle precursors form a network of interconnected cells and are specified by the rhomboid-triggered EGF pathway. *Development* **137**, 1965-1973.
- Frasch, M. (1999). Controls in patterning and diversification of somatic muscles during *Drosophila* embryogenesis. *Curr. Opin. Genet. Dev.* **9**, 522-529.
- Green, Y. S. and Vetter, M. L. (2011). EBF proteins participate in transcriptional regulation of *Xenopus* muscle development. *Dev. Biol.* **358**, 240-250.
- Jagla, T., Bellard, F., Lutz, Y., Dretzen, G., Bellard, M. and Jagla, K. (1998). ladybird determines cell fate decisions during diversification of *Drosophila* somatic muscles. *Development* **125**, 3699-3708.
- Johnson, A. N., Mokalled, M. H., Haden, T. N. and Olson, E. N. (2011). JAK/Stat signaling regulates heart precursor diversification in *Drosophila*. *Development* **138**, 4627-4638.
- Junion, G., Bataillé, L., Jagla, T., Da Ponte, J. P., Tapin, R. and Jagla, K. (2007). Genome-wide view of cell fate specification: ladybird acts at multiple levels during diversification of muscle and heart precursors. *Genes Dev.* **21**, 3163-3180.
- Karlsson, O., Thor, S., Norberg, T., Ohlsson, H. and Edlund, T. (1990). Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. *Nature* **344**, 879-882.
- Keller, C. A., Erickson, M. S. and Abmayr, S. M. (1997). Misexpression of nautilus induces myogenesis in cardioblasts and alters the pattern of somatic muscle fibers. *Dev. Biol.* **181**, 197-212.
- Kelly, R. G., Brown, N. A. and Buckingham, M. E. (2001). The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. *Dev. Cell* **1**, 435-440.
- Knirr, S. and Frasch, M. (2001). Molecular integration of inductive and mesoderm-intrinsic inputs governs even-skipped enhancer activity in a subset of pericardial and dorsal muscle progenitors. *Dev. Biol.* **238**, 13-26.
- Knirr, S., Azpiazu, N. and Frasch, M. (1999). The role of the NK-homeobox gene slouch (S59) in somatic muscle patterning. *Development* **126**, 4525-4535.
- LaBeau, E. M., Trujillo, D. L. and Cripps, R. M. (2009). Bithorax complex genes control alary muscle patterning along the cardiac tube of *Drosophila*. *Mech. Dev.* **126**, 478-486.
- Laugwitz, K. L., Moretti, A., Lam, J., Gruber, P., Chen, Y., Woodard, S., Lin, L. Z., Cai, C. L., Lu, M. M., Reth, M. et al. (2005). Postnatal Isl1+ cardioblasts enter fully differentiated cardiomyocyte lineages. *Nature* **433**, 647-653.
- Lescroart, F., Kelly, R. G., Le Garrec, J. F., Nicolas, J. F., Meilhac, S. M. and Buckingham, M. (2010). Clonal analysis reveals common lineage relationships between head muscles and second heart field derivatives in the mouse embryo. *Development* **137**, 3269-3279.
- Liu, Y. H., Jakobsen, J. S., Valentin, G., Amarantos, I., Gilmour, D. T. and Furlong, E. E. (2009). A systematic analysis of Tinman function reveals Eya and JAK-STAT signaling as essential regulators of muscle development. *Dev. Cell* **16**, 280-291.
- Mann, T., Bodmer, R. and Pandur, P. (2009). The *Drosophila* homolog of vertebrate Islet1 is a key component in early cardiogenesis. *Development* **136**, 317-326.
- Markstein, M., Pitsouli, C., Villalta, C., Celniker, S. E. and Perrimon, N. (2008). Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat. Genet.* **40**, 476-483.
- Moretti, A., Caron, L., Nakano, A., Lam, J. T., Bernshausen, A., Chen, Y., Qyang, Y., Bu, L., Sasaki, M., Martin-Puig, S. et al. (2006). Multipotent embryonic Isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. *Cell* **127**, 1151-1165.
- Nathan, E., Monovich, A., Tirosh-Finkel, L., Harrelson, Z., Rousso, T., Rinon, A., Harel, I., Evans, S. M. and Tzahor, E. (2008). The contribution of Islet1-expressing splanchnic mesoderm cells to distinct branchiomeric muscles reveals significant heterogeneity in head muscle development. *Development* **135**, 647-657.
- Nose, A., Ishiki, T. and Takeichi, M. (1998). Regional specification of muscle progenitors in *Drosophila*: the role of the msh homeobox gene. *Development* **125**, 215-223.
- Pfaff, S. L., Mendelsohn, M., Stewart, C. L., Edlund, T. and Jessell, T. M. (1996). Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell* **84**, 309-320.
- Philippakis, A. A., Busser, B. W., Gisselbrecht, S. S., He, F. S., Estrada, B., Michelson, A. M. and Bulyk, M. L. (2006). Expression-guided in silico evaluation of candidate cis regulatory codes for *Drosophila* muscle founder cells. *PLoS Comput. Biol.* **2**, e53.
- Polesello, C., Delon, I., Valenti, P., Ferrer, P. and Payre, F. (2002). Dmoesin controls actin-based cell shape and polarity during *Drosophila* melanogaster oogenesis. *Nat. Cell Biol.* **4**, 782-789.
- Prall, O. W., Menon, M. K., Solloway, M. J., Watanabe, Y., Zaffran, S., Bajolle, F., Biben, C., McBride, J. J., Robertson, B. R., Chaulet, H. et al. (2007). An Nkx2-5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation. *Cell* **128**, 947-959.
- Reim, I. and Frasch, M. (2005). The Dorsocross T-box genes are key components of the regulatory network controlling early cardiogenesis in *Drosophila*. *Development* **132**, 4911-4925.
- Ruiz-Gómez, M. and Bate, M. (1997). Segregation of myogenic lineages in *Drosophila* requires numb. *Development* **124**, 4857-4866.
- Ruiz-Gómez, M., Romani, S., Hartmann, C., Jäckle, H. and Bate, M. (1997). Specific muscle identities are regulated by Krüppel during *Drosophila* embryogenesis. *Development* **124**, 3407-3414.
- 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.
- Salzberg, A., D'Evelyn, D., Schulze, K. L., Lee, J. K., Strumpf, D., Tsai, L. and Bellen, H. J. (1994). Mutations affecting the pattern of the PNS in *Drosophila* reveal novel aspects of neuronal development. *Neuron* **13**, 269-287.
- Sambasivan, R., Kuratani, S. and Tajbakhsh, S. (2011). An eye on the head: the development and evolution of craniofacial muscles. *Development* **138**, 2401-2415.
- Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolz, V. and Furlong, E. E. (2007). A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev.* **21**, 436-449.
- Schaub, C., Nagaso, H., Jin, H. and Frasch, M. (2012). Org-1, the *Drosophila* ortholog of Tbx1, is a direct activator of known identity genes during muscle specification. *Development* **139**, 1001-1012.

- Sellin, J., Drechsler, M., Nguyen, H. T. and Paululat, A.** (2009). Antagonistic function of Lmd and Zfh1 fine tunes cell fate decisions in the Twi and Tin positive mesoderm of *Drosophila melanogaster*. *Dev. Biol.* **326**, 444-455.
- Speicher, S., Fischer, A., Knoblich, J. and Carmena, A.** (2008). The PDZ protein Canoe regulates the asymmetric division of *Drosophila* neuroblasts and muscle progenitors. *Curr. Biol.* **18**, 831-837.
- Stolfi, A., Gainous, T. B., Young, J. J., Mori, A., Levine, M. and Christiaen, L.** (2010). Early chordate origins of the vertebrate second heart field. *Science* **329**, 565-568.
- Tao, Y., Wang, J., Tokusumi, T., Gajewski, K. and Schulz, R. A.** (2007). Requirement of the LIM homeodomain transcription factor tailup for normal heart and hematopoietic organ formation in *Drosophila melanogaster*. *Mol. Cell Biol.* **27**, 3962-3969.
- Thor, S. and Thomas, J. B.** (1997). The *Drosophila* islet gene governs axon pathfinding and neurotransmitter identity. *Neuron* **18**, 397-409.
- Tixier, V., Bataillé, L. and Jagla, K.** (2010). Diversification of muscle types: recent insights from *Drosophila*. *Exp. Cell Res.* **316**, 3019-3027.
- Tolkin, T. and Christiaen, L.** (2012). Development and evolution of the ascidian cardiogenic mesoderm. *Curr. Top. Dev. Biol.* **100**, 107-141.
- Tzahor, E. and Evans, S. M.** (2011). Pharyngeal mesoderm development during embryogenesis: implications for both heart and head myogenesis. *Cardiovasc. Res.* **91**, 196-202.
- Volk, T. and VijayRaghavan, K.** (1994). A central role for epidermal segment border cells in the induction of muscle patterning in the *Drosophila* embryo. *Development* **120**, 59-70.
- Volohonsky, G., Edenfeld, G., Klämbt, C. and Volk, T.** (2007). Muscle-dependent maturation of tendon cells is induced by post-transcriptional regulation of stripeA. *Development* **134**, 347-356.
- Warner, J. B., Philippakis, A. A., Jaeger, S. A., He, F. S., Lin, J. and Bulyk, M. L.** (2008). Systematic identification of mammalian regulatory motifs' target genes and functions. *Nat. Methods* **5**, 347-353.
- Williams, T. M., Selegue, J. E., Werner, T., Gompel, N., Kopp, A. and Carroll, S. B.** (2008). The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* **134**, 610-623.
- Xu, X., Yin, Z., Hudson, J. B., Ferguson, E. L. and Frasch, M.** (1998). Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the *Drosophila* mesoderm. *Genes Dev.* **12**, 2354-2370.
- 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.
- Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M. and Furlong, E. E.** (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**, 65-70.

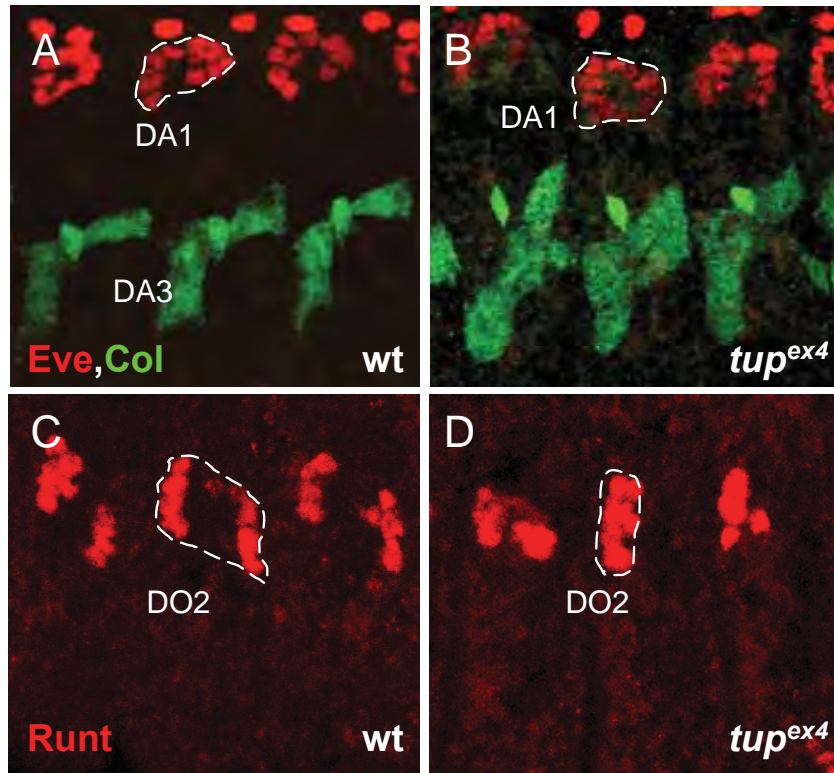


Fig. S1. Eve and Runt expression in dorsal muscles in *tup* mutant embryos. Dorsolateral views of three abdominal segments of stage 15 embryos are shown. (A) Wild-type embryo. Eve (red) and Col (green) are expressed in the DA1 (dotted circled in one segment) and DA3 muscles, respectively. (B) DA1 Eve expression is preserved in *tup^{ex4}* embryos. (C) Wild-type Runt expression in two rows of nuclei in the DO2 muscle (white-dotted circled in one segment). (D) Runt expression reveals an abnormal clustering of the DO2 nuclei.

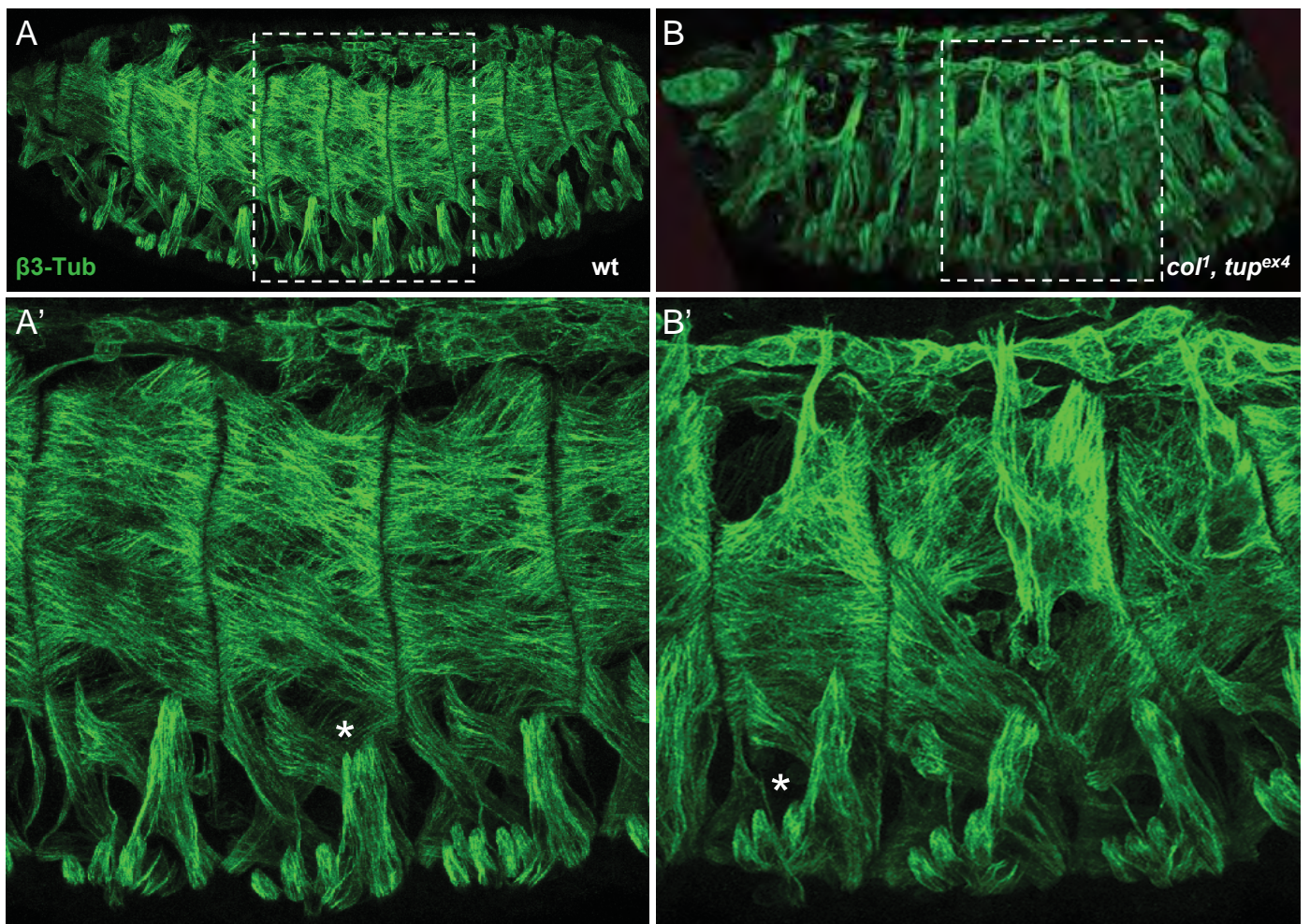


Fig. S2. Cumulative muscle defects in double *col¹, tup^{ex4}* mutant embryos. (A,B) β 3-Tubulin staining of stage 16 wild-type (A) and *col¹, tup^{ex4}* (B) embryos. (A',B') High-magnification views of the dorsal and dorsolateral muscles in three segments (framed region in A,B). (B,B') Major disorganization of the dorsal musculature and lack of muscle at the DA3 position (white asterisk) are observed.

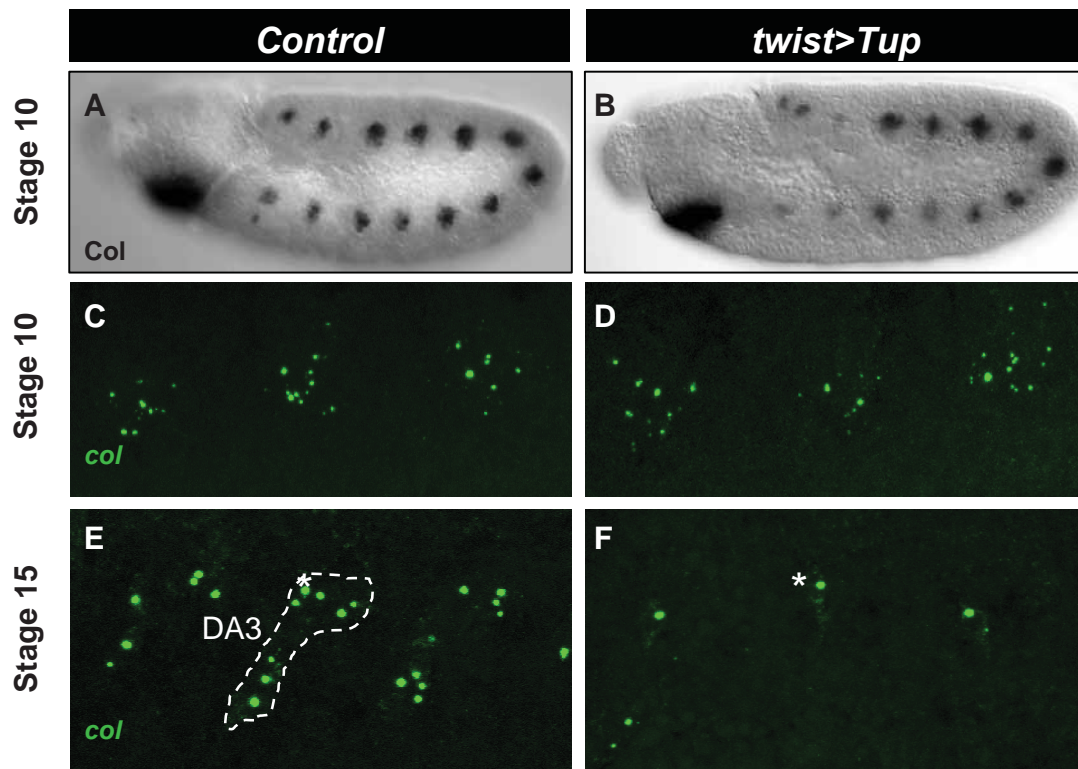


Fig. 3. Tup does not repress *col* transcription prior to the PC stage. (A,B) Col expression in wild-type (A) and *twi>tup* (B) embryos at stage 10. (C-F) *col* transcription in wild-type (C,E) and *twi>tup* (D,F) embryos. (A-D) Promuscular *col* expression is unaffected in *twi>tup* conditions; (E,F) DA3 *col* transcription is fully repressed at stage 15. The DA3 muscle is dotted circled in one wild-type segment. The white asterisk indicates *col* transcription in multidendritic neurons.


```

GGAAATGCAGcgcggcTCGTCAgTGTCAaaagt cggacatcggacgactgaggCATTACGAAATATCT 75
TGATAACTCAATCTccgAATCTTAAAAGggggaggatactgcggAAAAcGCCAGCGTCATggaagtcagccTGT 150
CAGTCagctcaGATCccagcgGATCATGACTtcgcagttcaatggagcaggagcgggtggaatattgccacaaag 225
ccgtggccacAAATTGGCTGATCGGTGCCGGCTTGAAGTCAATTTGCgaCGCTTAGGcccacggatcgttcga 300
tcgatcgatcgattggcatcagctcgtactctcatctgactgaaaatcggaaaactgacttgccgactgacggact 375
aaatctgccgagcaagtACTAAGTAcctggggaTGAGgGACAAACGAGCGCGAAAGTGCccacaAAACATATC 450 Twist
GCaCACAATTATccaacTGACAGtgACGATGTGTTAaGAAATGCCaAAAtacaagcaggttttcgaggttgagaa 525
ctaggtgatactgtggaaatgttgttttagtactttaagtaaagtaaagtaaagtttcgattaacctttataaac 600
taataactaactagtttcagaacaattaaatatgtagaaaccaatgcccgtcacattgaagcctccgaaaaccg 675
cgagatgcaatacaccgcttgataccctttctcctacaaaaacaaacaccgaaatggaatatgaaaatgatg 750
aaggaagttaatggaaaatctgtggcaccacaaaggccagggccacggactggaagcggcagccgtggatgcagca 825
gcagcggatcccagggaaaccagagggcGCACATGGcGCCccctgtgcGGCATTATCAggggectccgGaGCAC 900 Mef2
TATATTTAGACgacgGGCTAATGGATGcAgCagggaaaaatagaggcagataggtacaatgggtgagggagcggg 975
tggcagggcagatccgatccgaagcgaaccgaaccGGcaCaGGCCACACTTTCAAATTCATATTTATTGTCTAGGC 1050
TGcGGaCTGAGTGCATTTTTATATATGGCAGTCCATTTGCgcgcgactctcccaccagctcccaaaaACCAAAAT 1125
AAACAATAATTgaacatagcacgcaagaaatgaatgagtttcattagccacctactatttgaaaaacttaatagg 1200
atgtatataaatgtcataagaattcaggaactgaatcaattcctcagaatatagaactcattgaaaatattgtcct 1275
cgagagtgcacttaaatcttcttgcgtgtatgagcggAAAGGAAATCGCattgtagtgcacaatcagggggatac 1350
aaggagccgTGTAggccgtgAAATGCgtGGCATTTTAAAGGTTCCACATTTTGGGGGTCAaaaagt cggggGCT 1425
AATTGCATTGATTAGCTGACGGaGCgttttgccgcgcatcaaagtgccttcgcggccaggagaaatgatggggc 1500
aacggtgggttaaggcggggtactGAGGCAAGTCAGCTGTCAcTCAAGTGGCGACATGTTTACTTgggtggcagt 1575
caaaaacggaaAAAAaAAAaCcttgttgcctcttgaaagccacaaaaccaaagccaaaaaaaaaaaaaaaaaa 1650
aaaaaaaaacgagaaacgagaaaaaatccatagaaaagAAAaAAGAAACCAGCCGaaaAAACGGAAACgTTTGAGc 1725
CaTGCCGCAAAATgAAAGAAAAcatttTgttGTTTgCATTCAATTTCAATATTTtccctctcatggttattgggt 1800
ggtgtgagatggtattagaagatgcttgaattccaatctaccgagcgcgatcgatcgggtggagtaaaaagagcgggG 1875
ATCAAAGGTgAGCactccggagggtcgacggtcgatgggcagtcgcctaagcgcgtggcgatgccaaattgac 1950
ataatgtggcggttaacaaattcccaacaatgcacatccatacaagatccaacggtcaggatcagggctgcaca 2025
agtgaccaggggactataaagatattgaagattgaatcaaaagatatttctcaagggaggaagtgtaaaatga 2100
aatattattgggatactttggtatatacgaattagaatgaagagatgagattttgtatatttgcataccaaaacca 2175
atcaatcttaatagatgattcgttgcgtgacatgcgctttaatataattcccatcacgtaatgctctacatacatt 2250
ccccagccctaactctctgaaattcccctccaattgagtcgatctgaagaggcggccaaaggcgtatctttt 2325
tggaggggagcgtgcattggcgatcggcgttcgatcgcgcgcttgataaaaccgagccaaagccgaaaAAGGTG 2400
TAGAAAAAGTAAATGACAaatgcaGCATCAcagggatggatggatgggtggatggatggatggatgagattgag 2475
ggctgatctttggatcggcagacagggcattaggctaatcggatggcgggtgggtcaccagacagatattgtaa 2550
caccagagagataagaccgaggaacaaaagtgccgcctttgtgcgcatcgcaacgcttttcacgcctccgcc 2625
agcctccagatgtgttccctccccctaaaaccgaatctgaatcggaaatctctatctgtatctgcatgtggacga 2700
gcatctttgtatctccgcactcggcgttccactgtgtgtgtttgcccagagatgagttaaacgggggataagcc 2775
gctttttagtagtaagcatgagcgccttaaagagacctgagctcaggtaaacctgtagagtttagtgtacgaaggc 2850
caCAAAATGGCCgAGGCGtcgctgTAAATTCGATTTGCATTtGGccaaggggaaatccGCcagtGCCCgtggcC 2925
TAATgCgTTAATCACTTTcGAAAGCGTcAATTGAATGtCgCGTgCAGGAAcctcGCAATTTtgattgagggga 3000
cgagatgagaTCGaAATCAactacgaaAATCAACTGCTCTAATAAATTGTcgcgccgaatgacttgtgtggcatgg 3075
gagagtgggagttccttgggggtgACGCCTTAcacttgtccctcaattcgatccaagcgcctccgagaaaccac 3150
ccatctgcatccatccgctggcgaatggccGCCACGTGtGtGACAATTGTTGTTTATACTAACAATTTGTGAG 3225
GCAGGAAGATGAACATTTAgctGGATAACCCAATCctcttGCCAcGcctTGATTGCCCGCcaatgtgg 3300

```

Fig. S4. Nucleotide sequence of *D. melanogaster* Tup-DME. FlyBase positions (release: r5.42) 2L: 18897397..18900688. Conserved sequences in *D. melanogaster*, *D. sechellia*, *D. simulans*, *D. yakuba*, *D. erecta*, *D. pseudoobscura*, *D. persimilis*, *D. ananassae* and *D. willistoni* are in black bold characters. Putative binding sites for Mef2 (purple) and Twi (yellow) were positioned using Jaspar, Consite and Genomatix (MatInspector) matrix models. Their position is indicated in the margin. Only matrix similarities greater than 0.9 were considered. In grey is region corresponding to the ChIP-ChIP data from E. E. Furlong laboratories (http://furlonglab.embl.de/data/browse_chip_mod/). This region includes late binding sites for twist (6-8 hours) and for Mef2 (6-8 hours, 8-10 hours and 10-12 hours).

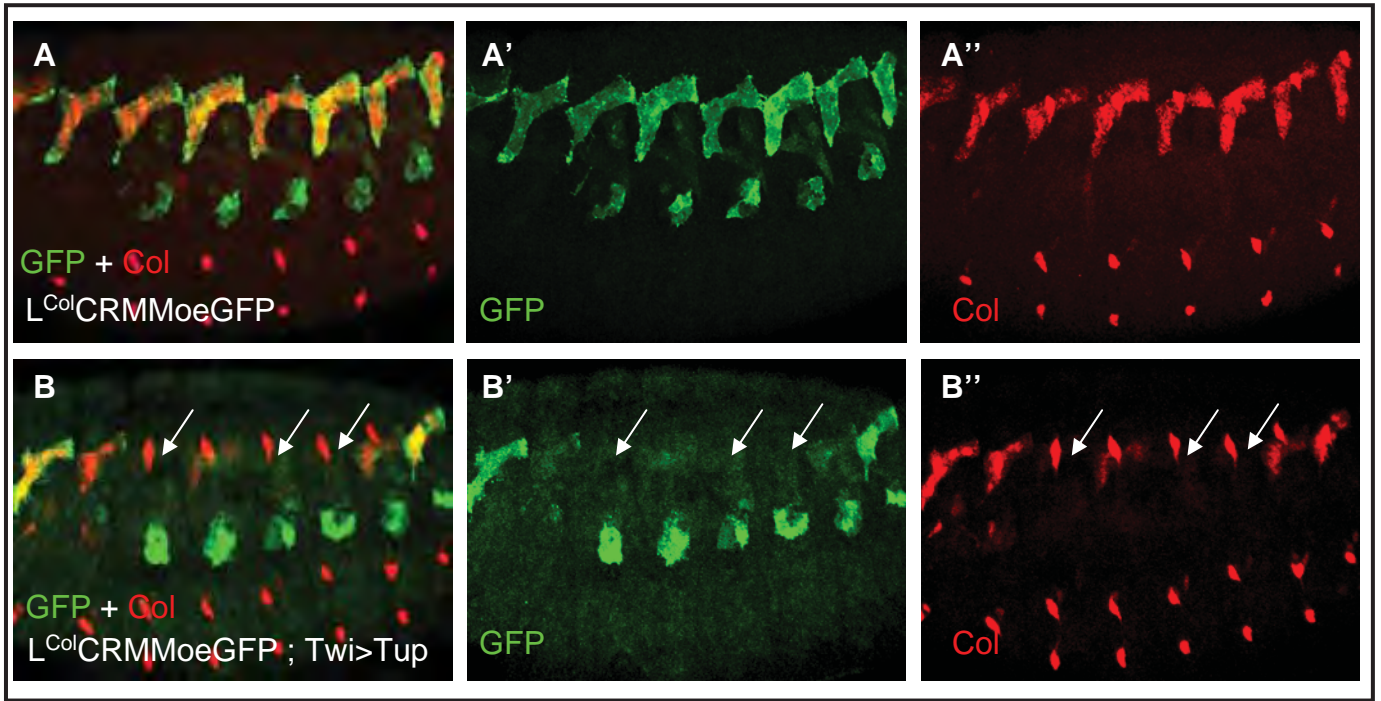


Figure 28. Tup réprime la transcription de *col* directement ou indirectement via le $L^{col}CRM$. A-B Double immunocoloration GFP/Col d'embryons $L^{col}CRM MoeGFP$ (A-A'') et $L^{col}CRM MoeGFP ; Twi>Tup$ (B-B'') au stade 15. Contrairement aux contrôles (A-A''), la surexpression de Tup dans tout le mésoderme inhibe l'expression de Col et l'activité du $L^{col}CRM$ (B-B'') (voir flèches). Le $L^{col}CRM$ reste actif dans les segments où Col n'est pas complètement réprimé (B-B'').

Résultats supplémentaires.

1. Tup réprime la transcription de *col* via le L^{Col}CRM.

Les hybridations in situ avec des sondes introniques montrent que *col* et *tup* sont co-transcrits dans le progéniteur du muscle DA2 (PC DA2). Cependant, seul *tup* est maintenu dans la FC DA2 puis le muscle en cours de différenciation (Article 2. Figure 4 B et F). par ailleurs, l'analyse du mutant *tup* montre que la transformation du muscle DA2 en muscle DA3 est associée à l'expression ectopique de Col dans le DA2 > DA3. A l'opposé, l'expression de Col est réprimée dans des conditions de gain de fonction de Tup (Twi>Tup) (Article 2 ; Figure 3 A et C, et 3S). L'ensemble montre que Tup a la capacité de réprimer *col*, une activité permettant de distinguer l'identité des muscles DA2 et DA3 (Article 2). Cependant, l'expression pan-mésodermique de Tup n'affecte pas l'expression promusculaire de *col*. Pour rappel, l'expression de *col* à ce stade est sous le contrôle d'un CRM, précoce E^{Col}CRM (Article 1). Ces résultats suggèrent que Tup agit donc sur le CRM *col* tardif, le L^{Col}CRM (Article 2. Figure S3). Afin de confirmer cette hypothèse, j'ai reproduit les expériences de gain de fonction de Tup (Twi>Tup) avec une lignée exprimant la MoeGFP sous le contrôle du L^{Col}CRM (Figure 28 A et B). Une double immuno-coloration GFP et Col montre que, contrairement aux embryons sauvages (Figure 28 A-A''), l'expression de la GFP est absente dans la majorité des segments, dans les embryons L^{Col}CRMMoeGFP ; Twi>Tup (Figure 28 B-B''). Ces données permettent de conclure que Tup réprime *col* via le L^{Col}CRM. De manière intéressante, une expression résiduelle de Col est détectée dans les segments où la GFP est exprimée (Figure 28 B-B''), suggérant fortement que l'activité du L^{Col}CRMMoeGFP reste dépendante du processus d'autorégulation par Col (voir introduction, Figure 18). L'analyse de la séquence nucléotidique du L^{Col}CRM n'a révélé aucun site potentiel de fixation de Tup, indiquant que la répression de *col* par Tup est indirecte et met en jeu un tiers facteur qui reste à identifier.

2. Autorégulation directe de Tup via le DME (CRM3.3).

Les résultats décrits dans l'article 2 montrent que la régulation transcriptionnelle de *tup* dans les muscles somatiques met en jeu deux CRM séparés : Un CRM précoce qui dépend de Tin et active *tup* dans le PC DA2 ; un deuxième CRM, tardif (DME) qui prend le relais et permet le maintien de la transcription de *tup* au cours de la différenciation des muscles dorsaux. Ce CRM est soumis à un processus d'autorégulation. La question restait de savoir si cette autorégulation est directe, via la liaison de Tup sur le DME, ou indirecte. Le DME que j'ai initialement caractérisé a une taille de 3 kb et comporte trois blocs de séquence fortement

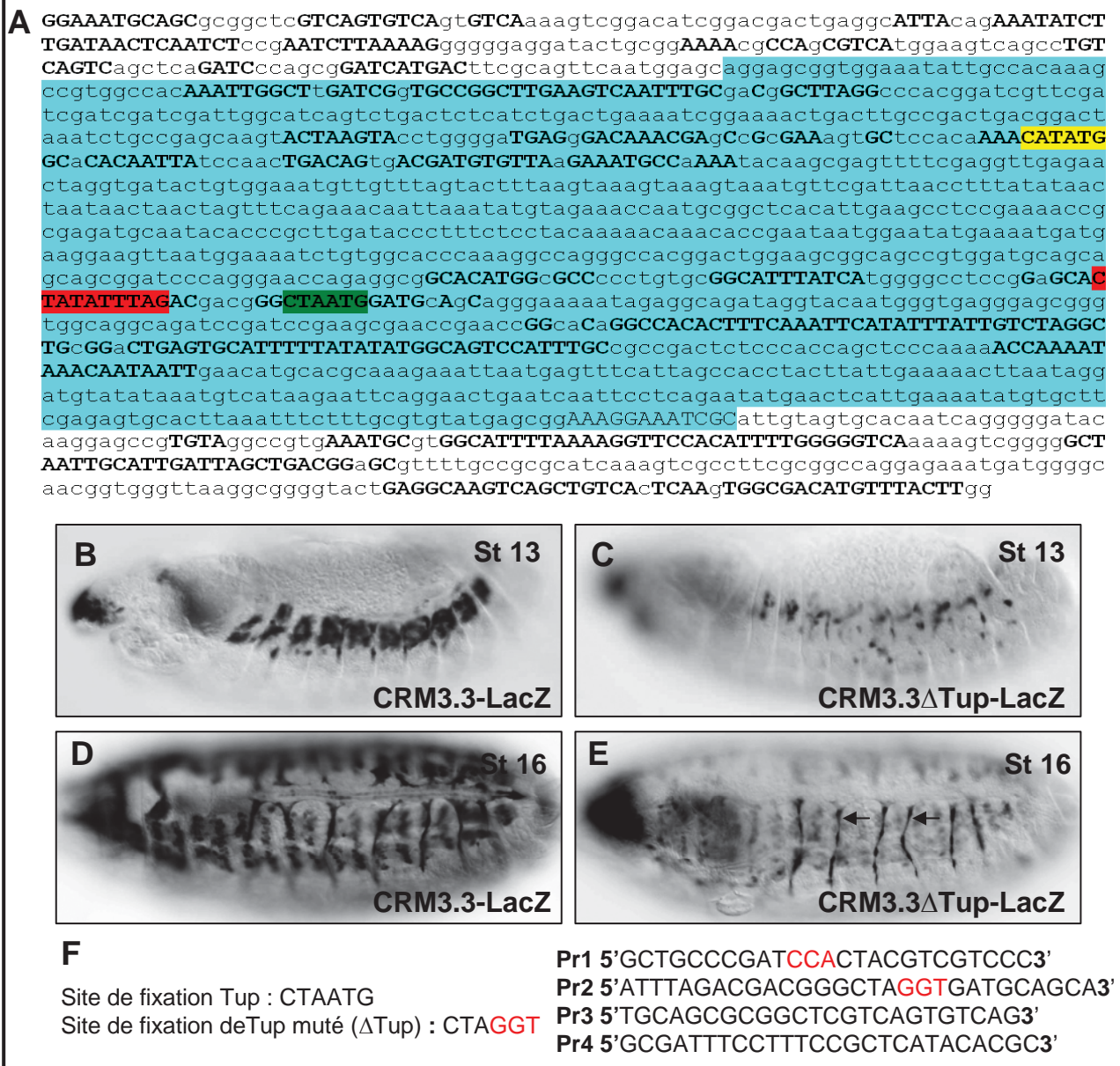


Figure 29. L'autorégulation de Tup est directe via le CRM 3.3. **A.** Séquence nucléotidique du CRM 3.3. Les lettres en majuscule désignent les positions conservées entre les espèces de drosophiles suivantes : *D. melanogaster*, *D. sechellia*, *D. simulans*, *D. yakuba*, *D. erecta*, *D. pseudoobscura*, *D. persimilis*, *D. ananassae* and *D. willistoni*. Les sites de fixation prédits pour Twi (Jaune), Mef2 (Rouge), Tup (Vert) sont indiqués. Le fragment d'ADN immuno-précipité dans des expériences de Chip-chip avec les anticorps anti-Twi et anti-Mef2 est surligné en bleu turquoise (http://furlonglab.embl.de/data/browse_chip_mod/). **B-C.** Immunocoloration LacZ d'embryons CRM3.3-LacZ (B,D) et CRM3.3ΔTup-LacZ (C,E) aux stades 13 et 16. La mutation du site de fixation de Tup réduit fortement l'activité du CRM3.3 dans les muscle dorsaux (comparer B, D et C, E). L'activité du CRM3.3 dans les muscles altères n'est pas affectée (Flèches dans E). **F.** Séquences des oligonucléotides (Pr1-4) choisis pour muter le site de fixation de Tup (voir Matériels et méthodes).

conservée entre les 12 espèces de drosophiles dont le génome est séquencé (Fig S4. Article 2). Avec la coopération de Jean-Louis Frendo (CR1 CNRS dans l'équipe) nous avons décidé de tester indépendamment l'activité de chacun de ces blocs par transgénèse avec un gène rapporteur LacZ. Cette analyse a permis d'identifier un fragment de 1.5 kb qui reproduit l'expression du DME (Figure 29, A, B, D). Ce fragment, appelé CRM 3.3, contient un site prédit *in silico* de liaison de Tup, un site Twi et un site Mef2, tous conservés, suggérant qu'ils sont fonctionnels (Figure 29, A). A l'appui de cette hypothèse, le CRM 3.3 recouvre un fragment immuno-précipité dans des expériences de Chip-ChiP avec des anticorps anti-Mef2 et anti-Twi, montrant que ces FT s'y lient *in vivo* (Figure 29 A et Sandmann et al., 2007). La présence dans le CRM 3.3 d'un site consensus Tup (CATTAG) conservé à la même position dans les 12 espèces de drosophile dont le génome est séquencé, suggérait que l'autorégulation par Tup est directe, via ce site (Figure 29 A). J'ai testé cette hypothèse en mutant ce site prédit dans le CRM 3.3 placé en amont du LacZ (Figure 29 B-F). Les résultats obtenus montrent que la mutation du site de fixation de Tup diminue drastiquement l'expression du transgène rapporteur dans les muscles dorsaux (Figure 29 B-E) alors qu'elle n'affecte pas l'expression du gène rapporteur dans les muscles alaires (Figure 29 E). On peut donc déduire de ces résultats que le site de liaison de Tup présent dans le CRM 3.3 est nécessaire au maintien de l'expression de *tup* dans les muscles dorsaux en cours de différenciation (Figure 29) et que l'autorégulation de Tup est probablement directe, via ce CRM. Cette hypothèse devra être confirmée par l'analyse des séquences cis régulatrices de *tup* en contexte génomique.

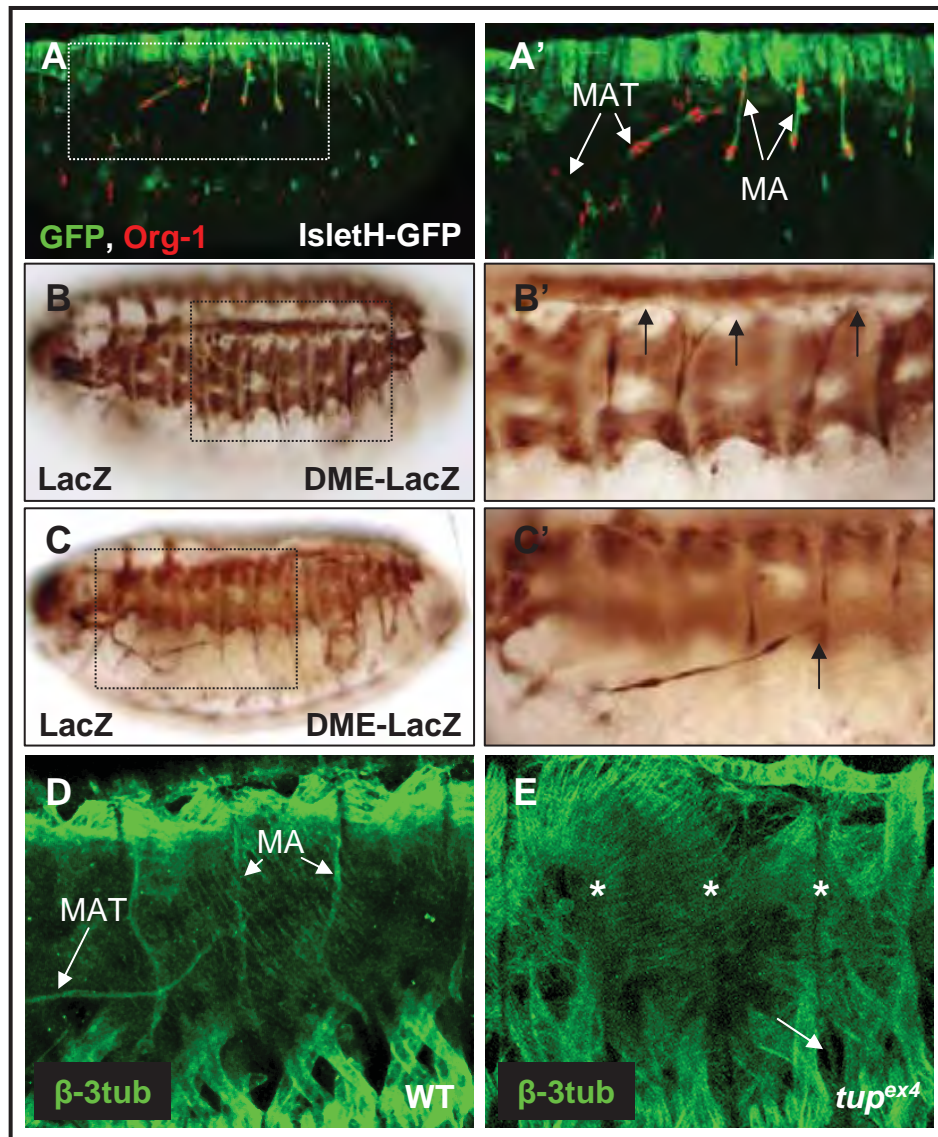


Figure 30. *Tup* est exprimé et requis dans les muscles alaires (MA) A-A'. Double immunomarquage GFP, Org1 d'embryons IsletH-GFP montrant que Org-1 et *Tup* sont co-exprimés dans les MA et les muscles alaires transversaux (MAT). **B-C'.** Marquage LacZ sur des embryons DME-LacZ. **B'.** Grossissement des quatre segments abdominaux encadrés dans B. Les flèches montrent que les MA de deux segments adjacents sont connectés entre eux dorsalement. **C'.** Agrandissement des segments indiqués dans C. Le plan choisi montre que les MA et les MAT sont liés (flèche dans C'). **D-E.** Embryons sauvages (D) ou *tup^{ex4}* (E) marqués avec la β -3tub. Les MA ne sont pas formés dans un mutant *tup^{ex4}* (Étoiles et flèches dans D et E). Tout les embryons sont au stade 15.

D. Le rôle d'Org-1 et Tup dans la spécification et/ou la réalisation de l'identité des muscles alaires.

tup est exprimé dans les muscles alaires (MA) (Figure 21 C ; Tao et al., 2007 ; article 2). Le FT à T-box Org-1 est également exprimé dans les MA, et requis pour leur spécification (Schaub et al., 2012). Je me suis donc intéressé à la morphologie des MA embryonnaires et aux relations transcriptionnelles entre Tup et Org-1 au cours de la spécification de ces muscles.

1. Tup est exprimé et requis pour la formation des MA.

Afin de compléter l'analyse de l'expression de Tup dans les MA, j'ai comparé l'expression de Org-1 et IsletH-GFP, une lignée rapporteur de l'expression de *tup* dans le mésoderme (article 2). Un double marquage pour Org-1 et GFP confirme une co-expression de Org-1 et Tup dans les MA (Figure 30 A-A'). On note également une co-expression de la GFP et Org-1 dans des fibres transversales localisées dans la partie antérieure de l'embryon, et qui ont cette morphologie étirée caractéristique des muscles alaires. Nous proposons d'appeler ces fibres, jamais décrites jusqu'à présent, MAT pour muscles alaires transversaux (Figures 30 A et A'). Afin de visualiser avec plus de détail la structure des MA et des MAT, nous avons utilisé une autre lignée rapporteur de *tup*, DME-LacZ qui reproduit l'expression tardive de Tup (Article 2). L'expression de LacZ dans des embryons aux stades 15 et 16 montre que les MA sont interconnectés à leurs extrémités dorsales (Figure 30 B et B'). L'extrémité postérieure des MAT semble également connectée aux MA (Figure 30 C et C'). En résumé, nos données montrent que Tup et Org-1 sont co-exprimés dans les MA, et que les MA et les MAT forment un réseau de muscles interconnectés.

La co-expression de Org-1 et Tup dans les MA pose au moins deux questions : Le rôle de Tup et les relations d'épistasie possibles entre les deux gènes. J'ai donc comparé les muscles alaires dans des embryons sauvages et des embryons mutants *tup^{ex4}* (un mutant nul, de Navascues and Modolell, 2010) avec un marquage anti- β 3-tubulin. J'ai observé que dans un mutant *tup^{ex4}* les MA ne se différencient pas (Figure 30 D et E). Néanmoins, dans une fraction des segments, on observe des fibres isolées, localisées dorso-latéralement, qui ressemblent à des MA mais ne s'attachent pas au tube cardiaque contrairement aux embryons sauvages (Figure 30 E).

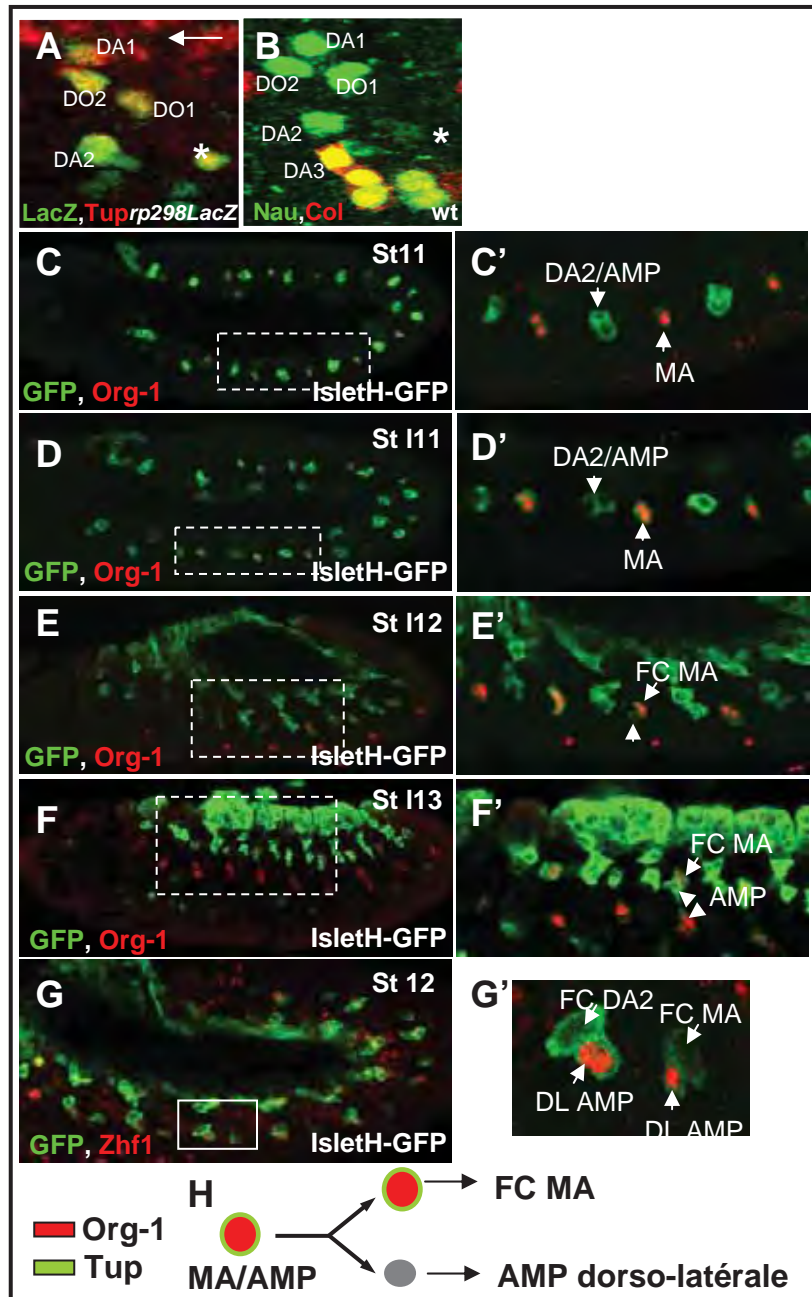


Figure 31. Lignage du PC MA. A. Marquage *Tup/LacZ* sur des embryons *rp298LacZ* (Duf). En plus des quatre PC des muscles dorsaux, et les cellules cardiaques (flèche), *Tup* est exprimé dans une cellule exprimant *rp298lacZ* (asterisk). B. Marquage *Col+Nau*. La cellule exprimant *Tup* et *rp298lacZ* (asterisk) exprime ni *Col* ni *Nau*. D-G. Double marquage *GFP* et *Org-1* d'embryons *Islet-H-GFP* à différents stades de développement. C-C' La cellule *Duf+* et *Nau-* (A,B) exprime aussi *Org-1* indiquant que c'est le PC du MA. D-G. Ce PC se divise; une des deux cellules maintient *Org-1* et est la FC MA; l'autre cellule fille perd l'expression de *Org-1* et active *Zfh1*: c'est une AMP (DL). F-F'. L'AMP DL est connectée avec une AMP L exprimant *Org-1*. H. Schéma récapitulatif du lignage du PC MA.

2. *Tup* est exprimé au cours des étapes précoces de la formation des MA.

Le phénotype mutant *tup^{ex4}* suggère que *tup* est requis au cours des étapes précoces de la formation des MA. Le PC MA a été identifié par l'expression d'Org-1 (Schaub et al., 2012). Au début du stade 11, une cellule exprimant *Tup* est située à une position proche du PC DA2/LAMP. Cette cellule exprime *Dumbfounded* (*Duf*) mais pas *Nau*, contrairement aux PC/FC des muscles dorsaux (Figure 31 A-B ; article 2). L'expression de *Duf* dans cette cellule suggère qu'elle a la capacité de fusionner avec des FCM pour former une fibre multinucléée, et en fait donc un bon candidat pour être le PC du MA. Une double immunocoloration Org-1 et GFP d'embryons *IsletH-GFP* montre que *Tup* et Org-1 sont co-exprimés dans cette cellule, l'expression de *IsletH-GFP* permettant de suivre le devenir de cette cellule (Figure 31 C-C'). Elle se divise de manière asymétrique, pour donner deux cellules filles (Figure 31 D-D') dont une seule maintenant l'expression d'Org-1, l'identifiant comme la FC MA (Figure 31 E-E'). À partir du stade 13, cette FC fusionne avec des FCM. La cellule jumelle établit des connexions avec une AMP latérale qui exprime Org-1 (Figeac et al., 2010 ; Schaub et al., 2012), l'identifiant comme une AMP dorso-latérale (Figure 31 F-F'). Afin de conforter cette hypothèse j'ai réalisé un double marquage GFP/*Zfh1* (marqueur des AMP) sur des embryons *IsletH-GFP* (Figeac et al., 2010, article 2). Ce double marquage confirme que la cellule sœur de la FC MA exprime *Zfh1* et confirme que c'est une AMP dorso-latérale (Figure 31 G, G', H ; article 2). L'ensemble de ces données montre que *Tup* et Org-1 sont co-exprimés dans les MA à partir du stade PC. L'absence d'expression de *Nau* suggère que le programme de différenciation myogénique des muscles alaires est distinct de celui des autres fibres. Cette hypothèse devra être testée.

3. Org-1 agit en amont de *Tup* dans le lignage MA.

La co-expression de *Tup* et Org-1 dans les PC/FC des MA soulève la question d'une relation d'épistasie possible entre ces deux FT. L'expression normale d'Org-1 dans le PC MA, dans un mutant *tup*, indique que *Tup* n'agit pas en amont d'Org-1 dans cette cellule. Afin de tester la réciproque, j'ai analysé l'expression de *Tup* dans un mutant *org-1* (*org-1^{OJ487}* Schaub et al., 2012) (Figure 32 A-B). Des immunocolorations anti-*Tup* montrent la perte d'expression de *Tup* dans les PC/FC des MA d'embryons *org-1^{OJ487}* (Figure 32 A'-B'). Ceci démontre qu'Org-1 agit en amont de *Tup* dans ce lignage. L'absence de muscles alaires dans les embryons mutants *Org-1^{OJ487}* et *tup^{ex4}* suggère que l'activation de *tup* par Org-1 au stade PC est essentielle à la spécification de l'identité « muscle alaire » (Figure 32 C-D). Plusieurs questions restent posées : Existe-t-il d'autres cibles essentielles d'Org-1 dans le PC MA. La régulation de Org-1 sur *Tup* est-elle directe, et dans ce cas, quels autres FT contrôlent la

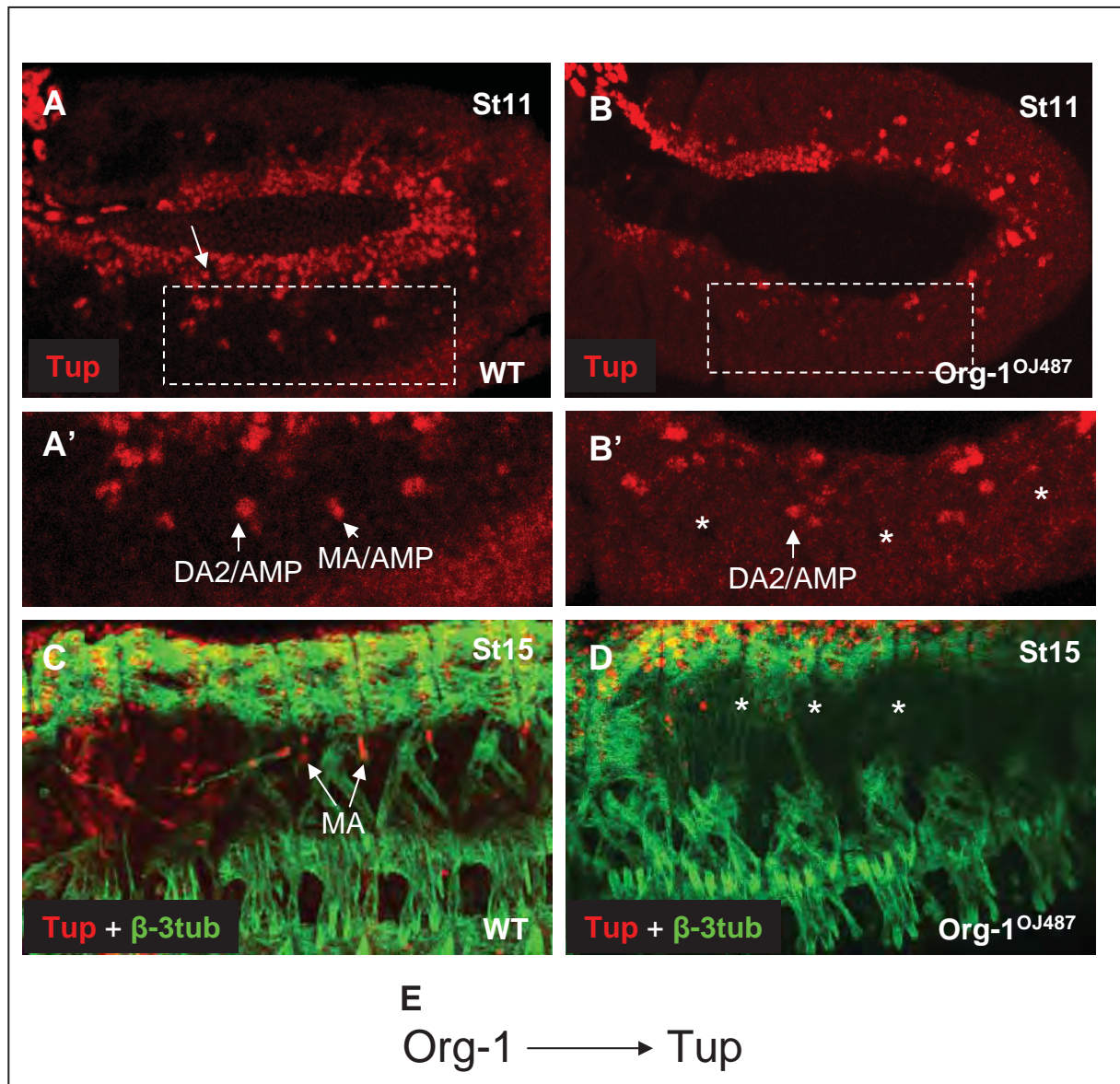


Figure 32. Org-1 active Tup au cours de la spécification des muscles alaires. A-B. Embryons sauvages (A) ou *org-1^{OJ487}* (B) marqués avec l'anticorps Anti-Tup. A ce stade Tup est exprimé dans : les cellules cardiaques (flèche dans A), les progéniteurs des muscles dorsaux dont le DA2, le PC MA/AMP (A'-B'). L'expression de Tup dans le PC MA/AMP est perdue dans un mutant *org-1^{OJ487}* (A'-B'). **C-D.** Double marquage Tup/β-3-Tub sur des embryons sauvage (C) et *Org-1^{OJ487}* montre que les MA ne sont pas spécifiés dans le mutant *Org-1^{OJ487}*. **E.** Régulation Org-1 et Tup dans les MA.

spécificité de cette régulation ? Les cribles en cours dans le laboratoire devraient permettre d'identifier d'autres acteurs de la spécification et de la différenciation des muscles alaires. Nous reviendrons sur la question de la fonction de ces muscles et de leurs équivalents fonctionnels possibles dans d'autres phyla dans la discussion.

Discussion

Mes travaux de thèse ont visé à mieux comprendre les mécanismes de mise en place du programme identitaire qui est responsable de la diversité morphologique des muscles squelettiques chez la drosophile. Les données actuelles sur le contrôle transcriptionnel de l'identité musculaire sont fragmentaires. Afin d'avoir une vision plus intégrée, j'ai choisi d'étudier deux niveaux de ce contrôle. Le premier est l'étude de l'activation transcriptionnelle de deux FTi ainsi que les régulations croisées entre eux. Le second est d'analyser le rôle de ces interactions dans la spécification de l'identité de deux muscles issus de deux progéniteurs différents. Les résultats obtenus ont permis d'établir un modèle liant une séquence dynamique de sélection des progéniteurs et l'implémentation d'un programme identitaire. Ils ont aussi soulevé de nouvelles questions et ouvert de nouvelles perspectives. Dans cette partie discussion, je reviens sur des résultats qui n'ont pas ou peu été discutés dans les articles publiés, les interprétations et les suites possibles à leur donner.

1. Activation transcriptionnelle des FTi ; Intégration des informations de position.

La première étape de spécification de l'identité musculaire est la « lecture », par les myoblastes, de leur position au sein du mésoderme. Cette lecture se traduit, en partie, par l'activation de l'expression de FTi spécifiques. Les premières études ayant lié position et activation de FTi spécifiques ont concerné l'expression de Eve dans le mésoderme dorsal. Les équipes de A. Michelson et M. Frasch ont caractérisé un CRM contrôlant l'expression mésodermique de Eve, le MHE. Ce CRM intègre des informations de position, les signaux Wg, Dpp, et Ras/MAPK et une information tissulaire apportée par les FT mésoderme-spécifiques Tin et Twi. Il intègre aussi la signalisation N impliquée dans la sélection des progéniteurs (Halfon et al., 2000). Les équipes d'A. Michelson et M. Bulyk ont ensuite utilisé une recherche bio-informatique (Codefinder) visant à identifier systématiquement les séquences génomiques enrichies en sites de liaison pour les FT dTCF, Mad, Ets, Tin et Twi mésodermiques, une composition prédictive de CRM mésodermiques dorsaux. Un sous ensemble d'environ 150 gènes sélectionnés comme gènes potentiellement exprimés dans les FC par des expériences de transcriptome, dont le gène *col*, ont été soumis à cette analyse. Ces prédictions, fournis par le laboratoire de M. Bulyk, m'ont permis d'identifier le CRM précoce de Col (E^{col} CRM), par une approche de transgène rapporteur. L'utilisation d'un gène rapporteur comportant un intron (*LacZ*ⁱ) a montré que ce CRM reproduit précisément la transcription du gène endogène. L'existence d'un autre CRM redondant (shadow) n'est bien

sûr pas exclue, mais aucun des autres CRM prédits par Codefinder n'est actif dans mes mains. De manière intéressante, les E^{col}CRM et MHE sont activés dans deux groupes promusculaires spatialement non recouvrant, montrant qu'ils intègrent différemment l'information de position issue de l'ectoderme, bien que probablement via la liaison des mêmes FT. Quels sont les mécanismes sous-tendant cette différence? Une caractérisation plus détaillée du E^{col}CRM devrait permettre de répondre à cette question. Elle pourra impliquer des mutations de sites de liaison des FT ayant servi de base à la prédiction bio-informatique, suivi de la mesure de l'impact de ces mutations sur l'expression d'un gène rapporteur. Dans ce cadre, une attention particulière devra être apportée aux sites conservés au cours de l'évolution. Une stratégie plus ambitieuse sera de caractériser le E^{col}CRM dans son contexte génomique normal. Pour cela, on pourra utiliser un transgène P[acman] ou FlyFos, contenant respectivement environ 80kbp ou 40kbp de séquence génomique *col*, c.à.d, l'intégralité de la région transcrite plus des régions amont et aval. Mathilde de Taffin qui a continué de caractériser le E^{col}CRM au cours de son M2R est actuellement en train d'étiqueter le gène *col* dans un FlyFos afin de pouvoir distinguer son expression de l'expression de la copie endogène. Le E^{col}CRM du gène *col* étiqueté peut être modifié par recombineering chez *E. Coli* avant intégration dans le génome de la drosophile grâce au système de transgénèse site spécifique Atp/AttB.

Outre les prédictions bio-informatiques, les données de Chip-chip du laboratoire d'E. Furlong suggèrent que Tin se lie *in vivo* sur le E^{col}CRM au cours de la période de spécification des groupes promusculaires. Pourtant, dans un mutant Tin, Col est exprimé dans le mésoderme ; son domaine d'expression est même élargi dorsalement, montrant que Tin n'est pas requis pour activer *col* (Article 2). L'équipe d'E. Furlong a montré que des FT mésodermiques, dont Tin, sont recrutés d'une façon systématique sur les CRM « actifs » sans forcément avoir de rôle essentiel, et que la combinaison de divers FT confère la spécificité d'activité de ces CRM (Junion et al., 2012). Une possibilité est que le E^{col}CRM soit un exemple de ce recrutement « collectif », sans rôle spécifique pour la liaison de Tin. L'expansion dorsale du domaine d'expression de Col dans des embryons mutant *tin* suggère néanmoins l'existence d'un répresseur dorsal agissant en amont de *col* et en aval de Tin. Le FTi Eve est un très bon candidat. En effet des données (très) préliminaires montrent que la surexpression de Eve (*Tw>Eve*) diminue l'expression promusculaire de Col. En outre, il a été précédemment montré qu'Eve réprime *col* dans le parasegment 1, au moment de la segmentation. Il faut cependant noter que dans ce cas, Eve agit via un CRM différent du E^{col}CRM (Crozatier et al., 1999).

Aucun site consensus dTCF (effecteur de la voie de signalisation Wg) n'a pu être identifié sur la séquence du E^{col}CRM. Les travaux de Crozatier et Vincent, 1999 ont montré que la

signalisation *wg* définit la limite antérieure du domaine d'expression promusculaire de *col*. Dans un mutant *wg*, le groupe promusculaire Col est élargi latéralement à l'ensemble du segment. La possibilité que Slp, en aval de *wg* réprime antérieurement *col* reste à tester. Les PC exprimant Col sont cependant sélectionnés normalement dans un mutant *wg*, indiquant que la restriction de l'expression promusculaire aux progéniteurs, qui implique la voie Notch (Crozatier et Vincent, 1999) et est médiée par le E^{col}CRM (Enriquez et al., 2010) se fait normalement malgré le défaut initial d'expression.

2. Propagation sélective de l'identité. Un passage de témoin

L'expression de *col* au cours de la myogenèse est contrôlée par deux CRM séparés, le E^{col}CRM et le L^{col}CRM. L'utilisation de transgènes rapporteurs comportant un intron (LacZⁱ) a permis de mettre en évidence un processus de relais entre ces deux CRM, qui se produit au stade progéniteur et assure le maintien de *col* dans le PC puis la cellule fondatrice DA3 (Enriquez et al., 2010). L'accumulation de Col dans le PC DA3 est un prérequis puisqu'il permet d'enclencher un processus d'autorégulation directe impliquant la liaison de Col sur le L^{col}CRM (Dubois et al., 2007 ; Enriquez et al., 2010). On a appelé ce mécanisme passage de témoin. Il faut noter que, dans le cas de Col, ce mécanisme est spécifique du PC DA3/DO5 alors que la protéine Col est aussi présente dans les PC DT1/DO3 et LL1/DO4. Les facteurs requis pour cette sélectivité restent à identifier. Plusieurs candidats possibles ont été révélés par le crible génétique effectué par Laurence Dubois dans l'équipe. L'étude de la régulation transcriptionnelle de *tup* et l'identification d'un CRM précoce et d'un CRM tardif séparés, nous ont permis de mettre en évidence que le même mécanisme de passage de témoin est requis pour le maintien de l'expression de Tup (Article 2). L'activité du CRM précoce Islet-H est aussi relayée par un CRM tardif, CRM 3.3 et ce relais dépend, comme pour *col*, d'un mécanisme d'autorégulation direct impliquant la liaison de Tup sur le CRM 3.3. En conclusion, le mécanisme de relais entre deux CRM pourrait être un mécanisme général de la régulation transcriptionnelle des FTi, qui permette de faire le lien entre intégration de l'information de position et réalisation de l'identité musculaire. L'étude d'autres FTi et de ce processus en contexte génomique devra maintenant conforter cette hypothèse.

La taille du CRM IsletH appelé maintenant E^{tup}CRM initialement identifié par Thor et Thomas, (1997) est de 7kb. Un fragment plus petit semble reproduire l'activité de IsletH (TupF4 ; Tao et al., 2007). Une analyse plus systématique est cependant nécessaire afin de disséquer la régulation de *tup* en détail. Cette analyse, entreprise par Jean-Louis Frendo, est facilitée par la disponibilité des lignées transgéniques Gal4 générées par le centre de

recherche Janelia (<http://flystocks.bio.indiana.edu/Browse/misc-browse/Janelia.php>). Les lignées Janelia contiennent des fragments d'ADN génomique d'environ 3kb, situés soit en amont de la transcription, soit dans des introns des gènes et légèrement chevauchants, qui sont fusionnés au gène Gal4 de levure via un promoteur synthétique. Pour le gène *tup*, il existe 23 lignées dont plusieurs contiennent des fragments partiellement chevauchants avec E^{tup}CRM et L^{tup}CRM. Leur croisement avec une lignée UAS-GFP (ou LacZ) permettra d'identifier les différents CRM de *tup*, dont celui actif dans les muscles alaires, et de réduire la taille du E^{tup}CRM. Une deuxième étape est de définir les fragments minimaux portant l'activité du E^{tup}CRM et du L^{tup}CRM, puis d'identifier les FT agissant directement.

3. Dynamique de sélection des PC et programme identitaire.

Notre équipe a récemment montré que plusieurs PC sont sélectionnés à partir du groupe promusculaire Col (Enriquez et al., 2012). Ces PC sont sélectionnés dans un ordre bien précis. Cet aspect dynamique, séquentiel de naissance des PC, bien que noté pour les PC exprimant Eve, n'avait pas été, jusqu'à présent, connecté directement à un processus de spécification d'identité. La comparaison détaillée de l'expression de *tup* et *col* avec des sondes introniques m'a permis de montrer que le premier PC à être sélectionné à partir du groupe promusculaire Col est le PC à l'origine du muscle DA2 qui est un muscle qui n'exprime pas Col. En raison de sa sélection précoce, le PC DA2 est issu de cellules exprimant encore Tin au moment de cette sélection. Durant cette fenêtre temporelle, Tin active *tup* qui à son tour réprime *col*. Au contraire, le PC du muscle DA3 est sélectionné quand l'expression de Tin est déjà restreinte au mésoderme plus dorsal (transition phase 2 phase 3). L'absence de Tin ne permet pas l'activation de *tup* et la répression de *col*. Cette séquence temporelle permet d'imposer une identité DA2 puis une identité du DA3 à deux PC issus du même groupe promusculaire, à des positions équivalentes dans le mésoderme. En connectant information de position (lue par les E-CRM) et séquence temporelle de sélection des PC, nos travaux ont permis de comprendre comment deux progéniteurs issus du même groupe promusculaire acquièrent des identités distinctes. Il reste maintenant à identifier le mécanisme de répression de *col* par Tup. On a vu que la répression se produit probablement sur le L^{col}CRM, bien que ce CRM ne semble pas comporter de site de fixation prédit de Tup. Une possibilité est que Tup active ou coopère avec un répresseur empêchant le processus d'autorégulation nécessaire au maintien de *col*.

Par ailleurs, le PC LL1/DO4 est sélectionné à la même position, mais plus tardivement que le PC DA3/DO5. Les mécanismes (gènes) responsables de la distinction de ces deux identités restent aussi à déterminer.

4. Rôle des FTi dans la reconnaissance des sites d'attachements.

L'analyse du processus d'attachement à l'épiderme du muscle DA3 dans des embryons sauvages et mutants a révélé deux étapes distinctes dans l'attachement des muscles à des sites précis de l'épiderme. Au stade 14, le muscle DA3 est attaché à trois sites et montre une forme triangulaire. Au stade 16, seuls les sites d'attachement distaux sont maintenus, le site central est perdu. Ce site central est donc un site d'attachement transitoire. Ces observations ont permis de proposer l'existence d'une phase d'attachement exploratoire, suivie d'une sélection des sites définitifs (Enriquez et al., 2012). Dans des mutants nuls pour *col*, le site transitoire remplace le site définitif antérieur. Dans des mutants hypomorphes *col*, le site transitoire est maintenu, donnant ainsi au muscle DA3 une forme triangulaire. De ces observations, on peut conclure que les FTi contrôlent la transition exploration/attachement définitif. L'analyse du mutant *tup^{ex4}* (Article 2 Figure 3 H) a révélé que dans une fraction des segments, le muscle DA2 adopte aussi une forme triangulaire avec trois sites d'attachement. Ceci suggère un mécanisme d'exploration/sélection général à tous les muscles et hautement dépendant du code des FTi. Afin de vérifier cette hypothèse, il faudrait rééditer les caractérisations morphologiques précises faites sur les muscles DA2 et DA3 avec d'autres muscles et passer à une analyse en imagerie du vivant (voir ci-dessous). Si nos données montrent que la résolution des sites d'attachement des muscles DA2 et DA3 est régulée par les FTi Col et Tup, les molécules impliquées en aval de ces FTi restent à identifier. L'identification des gènes cibles directs de ces deux FTi, respectivement dans les muscles DA2 et DA3, est un projet en soi. Une stratégie que j'avais envisagée au début de ma thèse était d'exprimer une version de la protéine Col étiquetée avec un motif de biotinylation (BLRP) par la protéine bactérienne BirA, sous le contrôle du L^{col}CRM. L'idée était ensuite de réaliser des expériences de Chip-Seq après sélection des complexes Col^{BLRP}/ADN avec des colonnes de streptavidine. Cette stratégie a été abandonnée puis reprise par Mathilde de Taffin en utilisant comme base un FlyFos (voir ci-dessus). Maintenant que le L^{tup}CRM est identifié, on pourrait utiliser des versions BLRP des protéines Tup et Col pour comparer leurs cibles respectives. Cela paraît idéal sur le papier mais nécessite encore beaucoup de mises au point.

5. Régionalisation du mésoderme : Collier et le territoire dorso-latéral

Actuellement, deux FT ont été décrits comme essentiels au processus de régionalisation dorso-ventrale du mésoderme : Tin pour la partie dorsale (phase 2 d'expression) et Poxm pour la partie ventrale et latérale (voir introduction). On pourrait y rajouter un rôle de Col dans la spécification du territoire dorso-latéral puisque notre équipe a montré que la perte de fonction de *col* induit un défaut de tous muscles dorso-latéraux (Enriquez et al., 2012). Nos expériences de lignage ont aussi montré que les myoblastes naïfs issus du groupe promusculaire Col contribuent majoritairement à la formation des muscles dorso-latéraux (Figure 25) même si ce n'est pas le cas pour le muscle DA3. L'ensemble suggère que Col pourrait avoir une fonction précoce de « facteur dorso-latéral » du mésoderme au même titre que proposé pour Poxm dorso-latéralement et Tup dorsalement (article 2). Cette régionalisation initiale est affinée par des régulations croisées entre ces FT. Il n'est pas évident que Tup, Poxmeso et Col agissent en amont de réseaux « régionaux » de régulation. Au vu des résultats acquis (article 2 et Enriquez et al., 2012), la situation semble plus compliquée. Il est probable que les régulations croisées entre différents facteurs, à différentes étapes de la spécification des PC et FC aboutissent à la diversité des codes FTi et permettent la transformation d'une information graduelle positionnelle et temporelle, en multiples identités musculaires (Figure 33).

6. Filmer la myogenèse sur des embryons vivants.

Jusqu'à présent, la visualisation de l'ensemble des étapes de formation des muscles a été essentiellement faite sur des embryons fixés. Les étapes tardives ont été bien décrites en imagerie du vivant, sur des embryons sauvages mais la résolution utilisée n'a pas permis, par exemple de détecter l'attachement transitoire des muscles (Schnorrer et al., 2004). En tirant profit de notre caractérisation des deux CRM de *col* il devrait être désormais possible d'avoir une vision dynamique et en temps réel de l'ensemble du processus de formation d'un muscle. L'expression dans le même embryon de deux transgènes rapporteurs, E^{col}CRM-H2BRFP qui permet d'exprimer une RFP membranaire dans le groupe promusculaire Col et L^{col}CRM-MoeGFP qui permet de visualiser la formation du muscle DA3 devrait permettre de suivre, *in vivo*, le processus de sélection du progéniteur DA3/DO5, la migration de la FC DA3 puis les étapes de fusion élongation et attachement du muscle DA3.

Le transgène L^{col}CRM-MoeGFP permet aussi de marquer les myopodes du muscle DA3 (Figure 27). Ses structures, encore peu étudiées, sont nécessaires aux mécanismes de migration des FC et de guidance des muscles. Il sera intéressant d'utiliser le

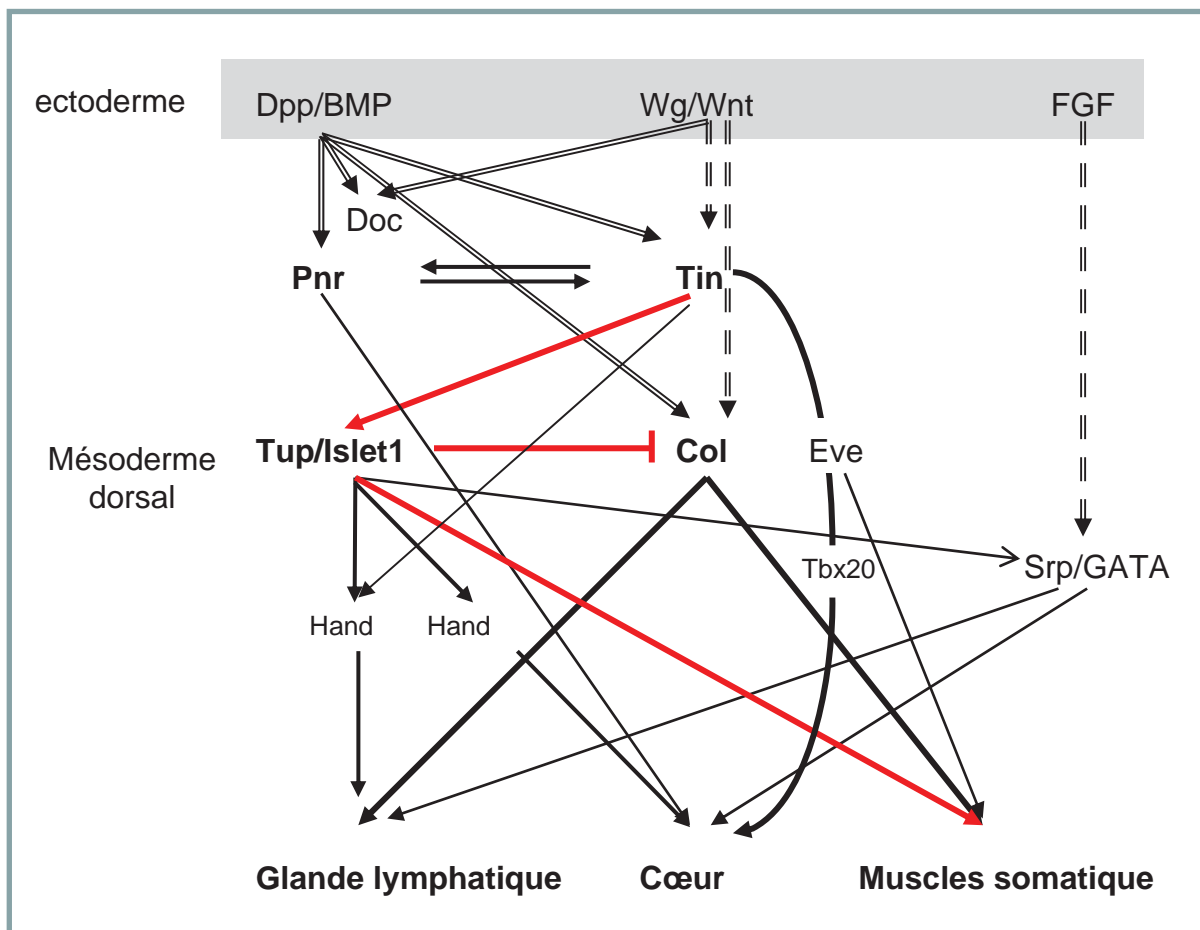


Figure 33. Réseaux de régulation génique dans le mésoderme dorsal de la drosophile. en rouge les interactions que nous avons décrit dans l'Article 2. *Adapté de Han and Olson, 2005; Tao et al., 2007; Mann et al., 2009; Ryu et al., 2011*

L^{col}CRMMoeGFP comme marqueur des myopodes dans des cribles génétiques visant à identifier de nouvelles molécules impliquées dans les mécanismes de migration et d'attachement musculaire.

7. La, ou les fonctions des muscles alaires ?

Les MA sont positionnés sous les muscles squelettiques et cette conformation les rend difficilement accessibles à l'observation directe d'embryons entiers. Un autre obstacle à l'étude des MA était, jusqu'à récemment, l'absence de marqueurs spécifiques. Nos travaux récents et ceux de Schaub et al 2012 ont montré que Org-1 et Tup sont exprimés précocement dans les MA et jouent un rôle déterminant dans leur développement. Une analyse plus détaillée des phénotypes mutants peut maintenant être entreprise. Tup ou Org-1 peuvent en outre être utilisés comme marqueurs des MA dans des cribles génétiques visant à identifier de nouveaux acteurs impliqués dans la différenciation de ces muscles et leur attachement au cœur et à l'épiderme. Quelle est la fonction des MA ? C'est une question à laquelle il est difficile de répondre en absence de mutants affectant spécifiquement ces muscles. Si des études morphologiques ont permis de postuler qu'ils sont nécessaires au soutien du tube cardiaque en position dorsale lors des mouvements de la larve et que leur contraction permet de réguler le flux de l'hémolymphe dans le cœur, ces hypothèses restent à tester. L'identification de CRM spécifiques de ces muscles devrait maintenant permettre d'affecter spécifiquement leur développement par l'utilisation d'ARN interférence et contourner la létalité embryonnaire associée à des mutations des gènes comme *org-1* et *tup*.

8. conservation évolutive des réseaux transcriptionnels au cours de la spécification musculaire et cardiaque.

Chez les vertébrés, le territoire cardiaque secondaire (TCS) est à l'origine d'un sous-ensemble de muscles faciaux et du cœur. Ce territoire est issu du mésoderme cardio-pharyngal. Un territoire cardio-pharyngal est également retrouvé chez le chordé *Ciona intestinalis* et est à l'origine des muscles du siphon atrial (ASM) et du cœur (Tolkin and Christiaen, 2012). Chez la drosophile, certains progéniteurs du mésoderme dorsal sont des progéniteurs mixtes à l'origine de cellules cardiaques et de muscles squelettiques ; par contre aucun lignage cardio-pharyngal n'a été décrit. Néanmoins, l'hypothèse qu'un tel territoire puisse exister chez la drosophile est soutenue, entre autres, par l'expression de

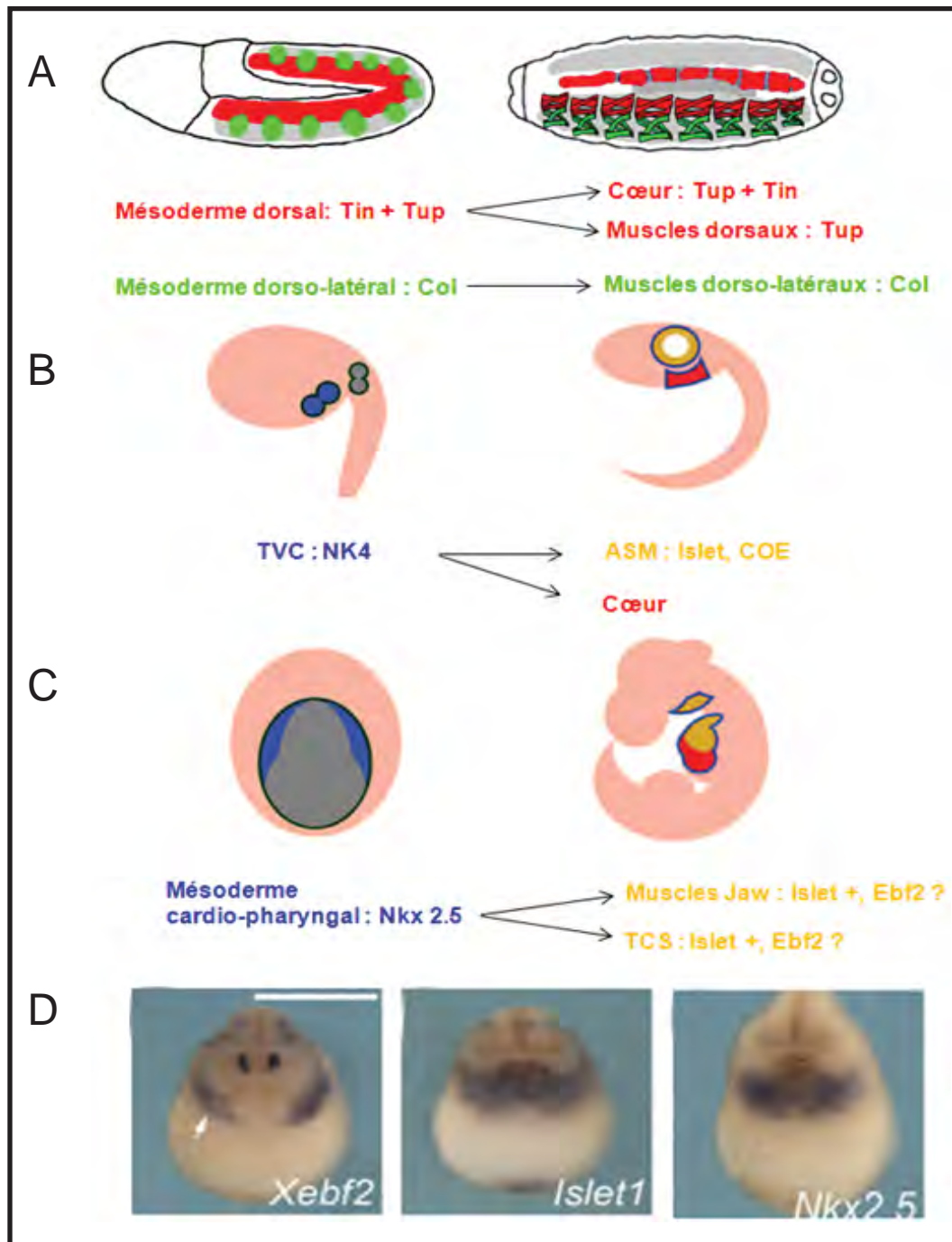


Figure 34. Facteurs régulant la spécification du mésoderme dorsal et du mésoderme cardio-pharyngal chez la drosophile, *Ciona intestinalis*, et les vertébrés. **A.** Alors que les muscles dorsaux ainsi que le cœur sont issus du mésoderme dorsal exprimant *Tup* et *Tin*, les muscles dorsolatéraux proviennent du mésoderme dorsolatéral exprimant *Col*. **(B-C).** Lignée cardio-pharyngal chez *Ciona* (B) et chez les vertébrés (C) **D.** Expression de *Xebf2*, *Islet1*, *NKx2.5* dans le mésoderme latéral chez le Xénope. Adapté de Stolfi et al., 2010.

Tup dans les muscles pharyngaux (donnés préliminaires). Cette hypothèse pourra être testée prochainement grâce à des expériences de lignage.

La spécification du mésoderme dorsal chez la drosophile met en jeu un réseau très élaboré de régulations transcriptionnelles, incluant les FT Tin, Tup, Col, mais aussi d'autres FT conservés au cours de l'évolution (GATA, Tbx, Pnr, Doc, Hand) (Figure 33). Parmi ces interactions, on retrouve l'activation de Tup par Tin qui a d'abord été décrite pour son rôle dans la spécification des cellules cardiaques (Mann et al., 2009). Nous avons maintenant montré que cette régulation s'étend à la spécification des muscles dorsaux. Cette activation de Tup dans les muscles permet ensuite la répression de Col et distingue l'identité dorsale et dorso-latérale. Les réseaux d'interactions entre ces différents acteurs sont aussi retrouvés chez les chordés et les vertébrés dans la formation du cœur et des muscles de la face (Figure 34). Chez *Ciona intestinalis*, NK4 (Tin/NKx2.5) est exprimé dans le territoire cardiopharyngal appelé TVC (Trunk Ventral Cells), ce territoire donne les précurseurs des ASM et les cellules cardiaques. Une fois séparés du TVC, les précurseurs des ASM expriment COE (Col/EBF) et Islet (Tup/Islet1). L'activité transcriptionnelle de COE est nécessaire pour la spécification des ASM et la répression du destin cardiaque (Stolfi et al., 2010). La question qui se pose maintenant est de savoir si NK 4 agit en amont de COE et d'islet dans le TVC. Chez les vertébrés, NKx2.5 et Islet1 sont tout les deux exprimés et requis avec d'autre FT (ex Hand, GATA, FoxF-1) au cours de la spécification des cellules du mésoderme cardiopharyngal (Vincent and Buckingham; Cai et al., 2003). Islet1 est ensuite maintenu dans le TCS et dans les muscles faciaux. On ne sait pas encore si EBF (Orthologue de Collier) est exprimé dans les muscles faciaux (Figure 34). Il a été montré par Laurence Dubois et plus récemment par Green et Vetter que Xebf2 et Xebf3 sont exprimés dans le mésoderme des arcs branchiaux chez le xénope, une expression recouvrant celle d'Islet1 et Nkx2.5 (Figure 34) (Dubois et al., 1998) (Green and Vetter, 2011). Xebf2 et Xebf3 collaborent avec Myo-D (Nau chez la drosophile) (Green and Vetter, 2011). Pour rappel, Col et Nau collaborent également au cours de la spécification du muscle DA3 chez la drosophile (voir introduction et Enriquez et al., 2012). L'ensemble de ces données permet de considérer que même si le sens des régulations transcriptionnelles varie, l'ossature du réseau impliqué dans la formation du mésoderme cardio-musculaire chez la drosophile et le mésoderme cardiopharyngal chez *Ciona* et les vertébrés est conservée.

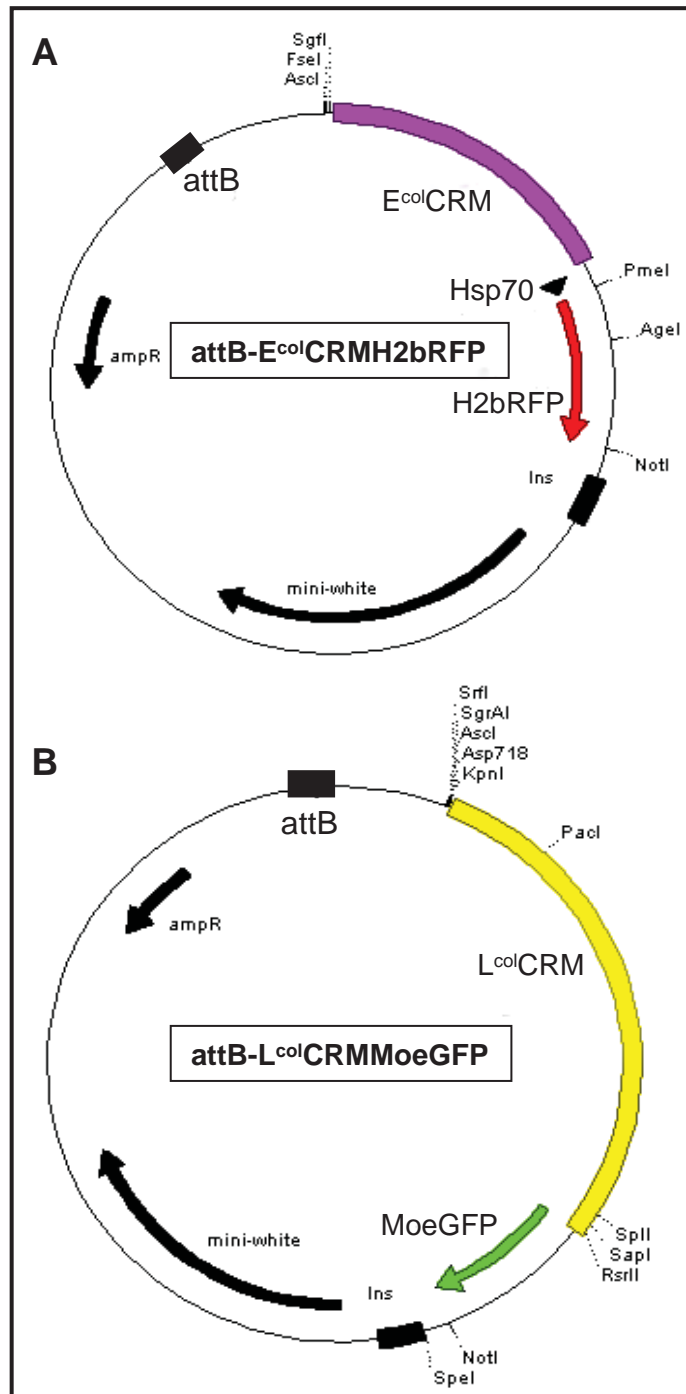


Figure 35. Représentation schématique des plasmides attB-E^{col}CRMH2bRFP (A) et attB-L^{col}CRMMoeGFP (B). A. Le E^{col}CRM, situé en amont du promoteur minimal Hsp70 est représenté en violet, la séquence codante H2bRFP en rouge. B. La séquence du L^{col}CRM est représentée en jaune, la séquence codante MoeGFP en vert. Les deux plasmides comportent un site attB requis pour l'insertion site-spécifique du transgène, une séquence « insulateur » (Ins) localisée en 3' du gène rapporteur, un gène de résistance à l'ampicilline (ampR), et un gène *mini-white* qui permet de sélectionner les drosophiles transgéniques qui ont les yeux rouges.

Matériels et Méthodes.

1. Construction des transgènes pour générer les lignées reportrices $E^{col}CRMH2bRFP$ et $L^{col}CRMMoeGFP$.

Dans ce paragraphe, je détaille les stratégies de clonage utilisées pour construire les transgènes rapporteurs ayant servi aux expériences de lignage (pages 72, 74, 76), et dans l'Article 2.

a. AttB- $E^{col}CRMH2bRFP$ (Figure 35 A). Le fragment comportant la séquence codante H2bRFP provient du plasmide P13-pc5-H2B-mRFP (Wallingford Laboratory). Il a été introduit dans le plasmide KS en utilisant les enzymes de restriction XbaI/NotI, puis ré-isolé à partir de ce vecteur par coupure avec NotI et XmaI, afin d'être substitué à la séquence codante GFP dans le vecteur AttB-PS3AG (Williams et al., 2008) coupé par AgeI et NotI. Ce clonage intermédiaire nous a permis d'obtenir le transgène AttB-H2bRFP. Le fragment d'ADN correspondant au $E^{col}CRM$ cloné dans le plasmide KS, préparé par digestion enzymatique Sall/SpeI, a été introduit dans le transgène AttB-H2bRFP coupé par XhoI et NheI. Le transgène AttB- $E^{col}CRMH2bRFP$ a été inséré dans le génome de la drosophile en utilisant la plateforme Zh8, à la position 49D du chromosome 2 (Bischof et al., 2007).

b. AttB- $L^{col}CRMMoeGFP$ (Figure 35 B). Le $L^{col}CRM$ a été préparé à partir du transgène P[col5-LacZ] (Crozatier et Vincent, 1999) après digestion par NgOMVI, suivie d'un traitement « Mung Bean Nucléase » pour générer des extrémités franches, et d'une seconde digestion XbaI. Le fragment obtenu comprend 4 kb d'ADN en amont du début de la transcription du gène *col* et 560 pb de région transcrite, incluant l'ATG initiateur. Il a été inséré, en phase, en amont de la séquence codante de la D-moesin GFP, dans le vecteur PSD7 (Polesello et al., 2002) préalablement coupé par NaeI et SpeI. Le gène chimère généré, $L^{col}CRMMoeGFP$ a été isolé à partir de PSD7 par coupure avec NotI et KpnI, et inséré dans le vecteur de transgénèse AttB-PS3AG préalablement coupé avec NotI et KpnI. Le transgène AttB- $L^{col}CRMMoeGFP$ a été inséré au niveau de la plateforme AttPII, localisé à la position 68A4 du chromosome 3. Ce transgène est aussi appelé *4_0.9col-moeGFP* dans l'article 2.

2. Mutation du site putatif de fixation de Tup dans le CRM3.3.

J'ai utilisé la méthode des quatre oligonucléotides (oligos). Les oligos amorces Pr1 et Pr2, sont en partie complémentaires et portent la mutation du site Tup (CTAATG en CTAGGT) (Figure 29 F). Pr3 et Pr4 sont complémentaires aux extrémités 5' et 3' du CRM 3.3, respectivement. Avec ces amorces j'ai réalisé plusieurs réactions d'amplification PCR

polymérase choisie pour ces amplifications est la *Phusion* (Finnzysms) qui a pour particularité de générer des extrémités franches aux fragments amplifiés. J'ai ainsi obtenu deux sous-fragments du CRM3.3, partiellement recouvrants, au niveau du site de fixation de Tup muté. Ces deux sous-fragment PCR ont été purifiés et mélangés afin de servir de matrice à la dernière réaction PCR, effectuée avec le couple d'amorces Pr3 et Pr4 et la polymérase Dynazyme (Finazyme). Cette polymérase rajoute un résidu Adénosine aux extrémités 3' et permet ainsi de faire du « TA cloning ». Le fragment amplifié, correspondant au CRM 3.3 complet comportant la mutation Tup, a d'abord été cloné dans un plasmide KS (par TA cloning), puis ré-isolé par coupure avec NotI avant d'être inséré dans le vecteur de transgénèse AttB-LacZ (Enriquez et al., 2010) linéarisé avec NotI.

Bibliographie

- Alberga, A., Boulay, J. L., Kempe, E., Dennefeld, C. and Haenlin, M. (1991) 'The snail gene required for mesoderm formation in *Drosophila* is expressed dynamically in derivatives of all three germ layers', *Development* 111(4): 983-92.
- 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(7B): 1325-40.
- Azpiazu, N., Lawrence, P. A., Vincent, J. P. and Frasch, M. (1996) 'Segmentation and specification of the *Drosophila* mesoderm', *Genes Dev* 10(24): 3183-94.
- Balagopalan, L., Keller, C. A. and Abmayr, S. M. (2001) 'Loss-of-function mutations reveal that the *Drosophila* nautilus gene is not essential for embryonic myogenesis or viability', *Dev Biol* 231(2): 374-82.
- Bataille, L., Delon, I., Da Ponte, J. P., Brown, N. H. and Jagla, K. (2010) 'Downstream of identity genes: muscle-type-specific regulation of the fusion process', *Dev Cell* 19(2): 317-28.
- Bate, M. (1990) 'The embryonic development of larval muscles in *Drosophila*', *Development* 110(3): 791-804.
- Bate, M. (1993) The Mesoderm and Its Derivatives. in M. Bate and A. Martinez Arias (eds.) *The Development of *Drosophila melanogaster**, vol. 2: Cold Spring Harbor Laboratory press.
- Bate, M. and Rushton, E. (1993) 'Myogenesis and muscle patterning in *Drosophila*', *C R Acad Sci III* 316(9): 1047-61.
- 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(1): 79-89.
- Baylies, M. K. and Bate, M. (1996) 'twist: a myogenic switch in *Drosophila*', *Science* 272(5267): 1481-4.
- Baylies, M. K., Martinez Arias, A. and Bate, M. (1995) 'wingless is required for the formation of a subset of muscle founder cells during *Drosophila* embryogenesis', *Development* 121(11): 3829-37.
- Becker, S., Pasca, G., Strumpf, D., Min, L. and Volk, T. (1997) 'Reciprocal signaling between *Drosophila* epidermal muscle attachment cells and their corresponding muscles', *Development* 124(13): 2615-22.
- Beckett, K. and Baylies, M. K. (2007) '3D analysis of founder cell and fusion competent myoblast arrangements outlines a new model of myoblast fusion', *Dev Biol* 309(1): 113-25.
- Berkes, C. A. and Tapscott, S. J. (2005) 'MyoD and the transcriptional control of myogenesis', *Semin Cell Dev Biol* 16(4-5): 585-95.
- 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: 167-96.
- Bodmer, R. and Venkatesh, T. V. (1998) 'Heart development in *Drosophila* and vertebrates: conservation of molecular mechanisms', *Dev Genet* 22(3): 181-6.
- Borkowski, O. M., Brown, N. H. and Bate, M. (1995) 'Anterior-posterior subdivision and the diversification of the mesoderm in *Drosophila*', *Development* 121(12): 4183-93.
- Borycki, A. G., Brunk, B., Tajbakhsh, S., Buckingham, M., Chiang, C. and Emerson, C. P., Jr. (1999) 'Sonic hedgehog controls epaxial muscle determination through Myf5 activation', *Development* 126(18): 4053-63.
- Borycki, A. G., Mendham, L. and Emerson, C. P., Jr. (1998) 'Control of somite patterning by Sonic hedgehog and its downstream signal response genes', *Development* 125(4): 777-90.
- Bour, B. A., Chakravarti, M., West, J. M. and Abmayr, S. M. (2000) '*Drosophila* SNS, a member of the immunoglobulin superfamily that is essential for myoblast fusion', *Genes Dev* 14(12): 1498-511.
- 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(6): 730-41.
- Bourgouin, C., Lundgren, S. E. and Thomas, J. B. (1992) 'Apterous is a *Drosophila* LIM domain gene required for the development of a subset of embryonic muscles', *Neuron* 9(3): 549-61.

Braun, T., Bober, E., Buschhausen-Denker, G., Kohtz, S., Grzeschik, K. H. and Arnold, H. H. (1989) 'Differential expression of myogenic determination genes in muscle cells: possible autoactivation by the Myf gene products', *EMBO J* 8(12): 3617-25.

Brent, A. E., Schweitzer, R. and Tabin, C. J. (2003) 'A somitic compartment of tendon progenitors', *Cell* 113(2): 235-48.

Broadie, K. S. and Bate, M. (1991) 'The development of adult muscles in *Drosophila*: ablation of identified muscle precursor cells', *Development* 113(1): 103-18.

Brown, N. H. (1993) 'Integrins hold *Drosophila* together', *Bioessays* 15(6): 383-90.

Brown, N. H. (2000) 'Cell-cell adhesion via the ECM: integrin genetics in fly and worm', *Matrix Biol* 19(3): 191-201.

Brown, N. H., Gregory, S. L., Rickoll, W. L., Fessler, L. I., Prout, M., White, R. A. and Fristrom, J. W. (2002) 'Talin is essential for integrin function in *Drosophila*', *Dev Cell* 3(4): 569-79.

Bryson-Richardson, R. J. and Currie, P. D. (2008) 'The genetics of vertebrate myogenesis', *Nat Rev Genet* 9(8): 632-46.

Buckingham, M. (2001) 'Skeletal muscle formation in vertebrates', *Curr Opin Genet Dev* 11(4): 440-8.

Cai, C. L., Liang, X., Shi, Y., Chu, P. H., Pfaff, S. L., Chen, J. and Evans, S. (2003) 'Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart', *Dev Cell* 5(6): 877-89.

Callahan, C. A., Bonkovsky, J. L., Scully, A. L. and Thomas, J. B. (1996) 'derailed is required for muscle attachment site selection in *Drosophila*', *Development* 122(9): 2761-7.

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(19): 2373-83.

Carmena, A., Gisselbrecht, S., Harrison, J., Jimenez, F. and Michelson, A. M. (1998a) 'Combinatorial signaling codes for the progressive determination of cell fates in the *Drosophila* embryonic mesoderm', *Genes Dev* 12(24): 3910-22.

Carmena, A., Murugasu-Oei, B., Menon, D., Jimenez, F. and Chia, W. (1998b) 'Inscuteable and numb mediate asymmetric muscle progenitor cell divisions during *Drosophila* myogenesis', *Genes Dev* 12(3): 304-15.

Carrasco-Rando, M., Tutor, A. S., Prieto-Sanchez, S., Gonzalez-Perez, E., Barrios, N., Letizia, A., Martin, P., Campuzano, S. and Ruiz-Gomez, M. (2011) '*Drosophila* araucan and caupolican integrate intrinsic and signalling inputs for the acquisition by muscle progenitors of the lateral transverse fate', *PLoS Genet* 7(7): e1002186.

Castanon, I., Von Stetina, S., Kass, J. and Baylies, M. K. (2001) 'Dimerization partners determine the activity of the Twist bHLH protein during *Drosophila* mesoderm development', *Development* 128(16): 3145-59.

Christ, B. and Ordahl, C. P. (1995) 'Early stages of chick somite development', *Anat Embryol (Berl)* 191(5): 381-96.

Collins, C. A., Olsen, I., Zammit, P. S., Heslop, L., Petrie, A., Partridge, T. A. and Morgan, J. E. (2005) 'Stem cell function, self-renewal, and behavioral heterogeneity of cells from the adult muscle satellite cell niche', *Cell* 122(2): 289-301.

Conway, K., Pin, C., Kiernan, J. A. and Merrifield, P. (2004) 'The E protein HEB is preferentially expressed in developing muscle', *Differentiation* 72(7): 327-40.

Crozatier, M., Valle, D., Dubois, L., Ibensouda, S. and Vincent, A. (1996) 'Collier, a novel regulator of *Drosophila* head development, is expressed in a single mitotic domain', *Curr Biol* 6(6): 707-18.

Crozatier, M., Valle, D., Dubois, L., Ibensouda, S. and Vincent, A. (1999) 'Head versus trunk patterning in the *Drosophila* embryo; collier requirement for formation of the intercalary segment', *Development* 126(19): 4385-94.

Crozatier, M. and Vincent, A. (1999) 'Requirement for the *Drosophila* COE transcription factor Collier in formation of an embryonic muscle: transcriptional response to notch signalling', *Development* 126(7): 1495-504.

Crozatier, M. and Vincent, A. (2008) 'Control of multidendritic neuron differentiation in *Drosophila*: the role of Collier', *Dev Biol* 315(1): 232-42.

Daburon, V., Mella, S., Plouhinec, J. L., Mazan, S., Crozatier, M. and Vincent, A. (2008) 'The metazoan history of the COE transcription factors. Selection of a variant HLH motif by mandatory inclusion of a duplicated exon in vertebrates', *BMC Evol Biol* 8: 131.

Davis, R. L., Weintraub, H. and Lassar, A. B. (1987) 'Expression of a single transfected cDNA converts fibroblasts to myoblasts', *Cell* 51(6): 987-1000.

de Jossineau C, Bataillé L, Jagla T, Jagla K. (2012). 'Diversification of muscle types in *Drosophila*: upstream and downstream of identity genes', *Curr Top Dev Biol* 98: 277-301

Dietrich, S., Abou-Rebyeh, F., Brohmann, H., Bladt, F., Sonnenberg-Riethmacher, E., Yamaai, T., Lumsden, A., Brand-Saberi, B. and Birchmeier, C. (1999) 'The role of SF/HGF and c-Met in the development of skeletal muscle', *Development* 126(8): 1621-9.

Doe, C. Q., Smouse, D. and Goodman, C. S. (1988) 'Control of neuronal fate by the *Drosophila* segmentation gene even-skipped', *Nature* 333(6171): 376-8.

Dohrmann, C., Azpiazu, N. and Frasch, M. (1990) 'A new *Drosophila* homeo box gene is expressed in mesodermal precursor cells of distinct muscles during embryogenesis', *Genes Dev* 4(12A): 2098-111.

Duan, H., Zhang, C., Chen, J., Sink, H., Frei, E. and Noll, M. (2007) 'A key role of Pox meso in somatic myogenesis of *Drosophila*', *Development* 134(22): 3985-97.

Dubois, L., Bally-Cuif, L., Crozatier, M., Moreau, J., Paquereau, L. and Vincent, A. (1998) 'XCo2, a transcription factor of the Col/Olf-1/EBF family involved in the specification of primary neurons in *Xenopus*', *Curr Biol* 8(4): 199-209.

Dubois, L., Enriquez, J., Daburon, V., Crozet, F., Lebreton, G., Crozatier, M. and Vincent, A. (2007) 'Collier transcription in a single *Drosophila* muscle lineage: the combinatorial control of muscle identity', *Development* 134(24): 4347-55.

Dubois, L. and Vincent, A. (2001) 'The COE--Collier/Olf1/EBF--transcription factors: structural conservation and diversity of developmental functions', *Mech Dev* 108(1-2): 3-12.

Duxson, M. J., Usson, Y. and Harris, A. J. (1989) 'The origin of secondary myotubes in mammalian skeletal muscles: ultrastructural studies', *Development* 107(4): 743-50.

Dworak, H. A., Charles, M. A., Pellerano, L. B. and Sink, H. (2001) 'Characterization of *Drosophila* hibris, a gene related to human nephrin', *Development* 128(21): 4265-76.

Edmondson, D. G. and Olson, E. N. (1989) 'A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program', *Genes Dev* 3(5): 628-40.

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 U S A* 105(3): 918-23.

Enriquez, J., Boukhatmi, H., Dubois, L., Philippakis, A. A., Bulyk, M. L., Michelson, A. M., Crozatier, M. and Vincent, A. (2010) 'Multi-step control of muscle diversity by Hox proteins in the *Drosophila* embryo', *Development* 137(3): 457-66.

Enriquez, J., de Taffin, M., Crozatier, M., Vincent, A. and Dubois, L. (2012) 'Combinatorial coding of *Drosophila* muscle shape by Collier and Nautilus', *Dev Biol* 363(1): 27-39.

Fan, C. M. and Tessier-Lavigne, M. (1994) 'Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog', *Cell* 79(7): 1175-86.

Fernandes, J., Bate, M. and Vijayraghavan, K. (1991) 'Development of the indirect flight muscles of *Drosophila*', *Development* 113(1): 67-77.

Figeac, N., Jagla, T., Aradhya, R., Da Ponte, J. P. and Jagla, K. (2010) '*Drosophila* adult muscle precursors form a network of interconnected cells and are specified by the rhomboid-triggered EGF pathway', *Development* 137(12): 1965-73.

Frasch, M. (1995) 'Induction of visceral and cardiac mesoderm by ectodermal Dpp in the early *Drosophila* embryo', *Nature* 374(6521): 464-7.

Frasch, M., Hoey, T., Rushlow, C., Doyle, H. and Levine, M. (1987) 'Characterization and localization of the even-skipped protein of *Drosophila*', *Embo J* 6(3): 749-59.

Frommer, G., Vorbruggen, G., Pasca, G., Jackle, H. and Volk, T. (1996) 'Epidermal egr-like zinc finger protein of *Drosophila* participates in myotube guidance', *Embo J* 15(7): 1642-9.

Green, Y. S. and Vetter, M. L. (2011) 'EBF proteins participate in transcriptional regulation of *Xenopus* muscle development', *Dev Biol* 358(1): 240-50.

Guo, M., Jan, L. Y. and Jan, Y. N. (1996) 'Control of daughter cell fates during asymmetric division: interaction of Numb and Notch', *Neuron* 17(1): 27-41.

Halfon, M. S., Carmena, A., Gisselbrecht, S., Sackerson, C. M., Jimenez, 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(1): 63-74.

Halfon, M. S., Hashimoto, C. and Keshishian, H. (1995) 'The *Drosophila* toll gene functions zygotically and is necessary for proper motoneuron and muscle development', *Dev Biol* 169(1): 151-67.

Halpern, M. E., Chiba, A., Johansen, J. and Keshishian, H. (1991) 'Growth cone behavior underlying the development of stereotypic synaptic connections in *Drosophila* embryos', *J Neurosci* 11(10): 3227-38.

Haralalka, S., Shelton, C., Cartwright, H. N., Katzfey, E., Janzen, E. and Abmayr, S. M. (2011) 'Asymmetric Mbc, active Rac1 and F-actin foci in the fusion-competent myoblasts during myoblast fusion in *Drosophila*', *Development* 138(8): 1551-62.

Ho, R. K., Ball, E. E. and Goodman, C. S. (1983) 'Muscle pioneers: large mesodermal cells that erect a scaffold for developing muscles and motoneurons in grasshopper embryos', *Nature* 301(5895): 66-9.

Hoang, B. and Chiba, A. (2001) 'Single-cell analysis of *Drosophila* larval neuromuscular synapses', *Dev Biol* 229(1): 55-70.

Houzelstein, D., Auda-Boucher, G., Cheraud, Y., Rouaud, T., Blanc, I., Tajbakhsh, S., Buckingham, M. E., Fontaine-Perus, J. and Robert, B. (1999) 'The homeobox gene *Msx1* is expressed in a subset of somites, and in muscle progenitor cells migrating into the forelimb', *Development* 126(12): 2689-701.

Inaki, M., Shinza-Kameda, M., Ismat, A., Frasch, M. and Nose, A. (2010) '*Drosophila* Tey represses transcription of the repulsive cue Toll and generates neuromuscular target specificity', *Development* 137(13): 2139-46.

Jagla, K., Dolle, P., Mattei, M. G., Jagla, T., Schuhbauer, B., Dretzen, G., Bellard, F. and Bellard, M. (1995) 'Mouse *Lbx1* and human *LBX1* define a novel mammalian homeobox gene family related to the *Drosophila* ladybird genes', *Mech Dev* 53(3): 345-56.

Jagla, T., Bellard, F., Lutz, Y., Dretzen, G., Bellard, M. and Jagla, K. (1998) 'ladybird determines cell fate decisions during diversification of *Drosophila* somatic muscles', *Development* 125(18): 3699-708.

Johnson, A. N., Mokalled, M. H., Haden, T. N. and Olson, E. N. (2011). 'JAK/Stat signaling regulates heart precursor diversification in *Drosophila*', *Development* 138(21): 4627-38.

Junion, G., Bataille, L., Jagla, T., Da Ponte, J. P., Tapin, R. and Jagla, K. (2007) 'Genome-wide view of cell fate specification: ladybird acts at multiple levels during diversification of muscle and heart precursors', *Genes Dev* 21(23): 3163-80.

Karlsson, O., Thor, S., Norberg, T., Ohlsson, H. and Edlund, T. (1990) 'Insulin gene enhancer binding protein *Isl-1* is a member of a novel class of proteins containing both a homeo- and a Cys-His domain', *Nature* 344(6269): 879-82.

Kassar-Duchossoy, L., Gayraud-Morel, B., Gomes, D., Rocancourt, D., Buckingham, M., Shinin, V. and Tajbakhsh, S. (2004) '*Mrf4* determines skeletal muscle identity in *Myf5:Myod* double-mutant mice', *Nature* 431(7007): 466-71.

Kelly, A. M. and Zacks, S. I. (1969) 'The histogenesis of rat intercostal muscle', *J Cell Biol* 42(1): 135-53.

Knirr S, Breuer S, Paululat A, Renkawitz-Pohl R. (1997). 'Somatic mesoderm differentiation and the development of a subset of pericardial cells depend on the not enough muscles (*nem*) locus, which contains the inscuteable gene and the intron located gene, *skittles*', *Mech Dev* 67(1): 69-81.

Knirr, S., Azpiazu, N. and Frasch, M. (1999) 'The role of the NK-homeobox gene *slouch* (*S59*) in somatic muscle patterning', *Development* 126(20): 4525-35.

Kohsaka, H. and Nose, A. (2009) 'Target recognition at the tips of postsynaptic filopodia: accumulation and function of Capricious', *Development* 136(7): 1127-35.

Korzh, V., Edlund, T. and Thor, S. (1993) 'Zebrafish primary neurons initiate expression of the LIM homeodomain protein Isl-1 at the end of gastrulation', *Development* 118(2): 417-25.

Kosman, D., Ip, Y. T., Levine, M. and Arora, K. (1991) 'Establishment of the mesoderm-neuroectoderm boundary in the Drosophila embryo', *Science* 254(5028): 118-22.

Kurusu, M., Cording, A., Taniguchi, M., Menon, K., Suzuki, E. and Zinn, K. (2008) 'A screen of cell-surface molecules identifies leucine-rich repeat proteins as key mediators of synaptic target selection', *Neuron* 59(6): 972-85.

LaBeau, E. M., Trujillo, D. L. and Cripps, R. M. (2009) 'Bithorax complex genes control alary muscle patterning along the cardiac tube of Drosophila', *Mech Dev* 126(5-6): 478-86.

Lahaye, L. L., Wouda, R. R., de Jong, A. W., Fradkin, L. G. and Noordermeer, J. N. (2012) 'WNT5 interacts with the Ryk receptors doughnut and derailed to mediate muscle attachment site selection in Drosophila melanogaster', *PLoS One* 7(3): e32297.

Landgraf, M., Bossing, T., Technau, G. M. and Bate, M. (1997) 'The origin, location, and projections of the embryonic abdominal motorneurons of Drosophila', *J Neurosci* 17(24): 9642-55.

Lassar, A. B., Davis, R. L., Wright, W. E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D. and Weintraub, H. (1991) 'Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo', *Cell* 66(2): 305-15.

Lawrence, P. A., Bodmer, R. and Vincent, J. P. (1995) 'Segmental patterning of heart precursors in Drosophila', *Development* 121(12): 4303-8.

Lehmacher, C., Abeln, B. and Paululat, A. (2012) 'The ultrastructure of Drosophila heart cells', *Arthropod Struct Dev*.

Leptin, M. (1991) 'twist and snail as positive and negative regulators during Drosophila mesoderm development', *Genes Dev* 5(9): 1568-76.

Lescroart, F., Kelly, R. G., Le Garrec, J. F., Nicolas, J. F., Meilhac, S. M. and Buckingham, M. (2010) 'Clonal analysis reveals common lineage relationships between head muscles and second heart field derivatives in the mouse embryo', *Development* 137(19): 3269-79.

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 U S A* 91(12): 5662-6.

Lin, M. H., Bour, B. A., Abmayr, S. M. and Storti, R. V. (1997) 'Ectopic expression of MEF2 in the epidermis induces epidermal expression of muscle genes and abnormal muscle development in Drosophila', *Dev Biol* 182(2): 240-55.

Liu, Y. H., Jakobsen, J. S., Valentin, G., Amarantos, I., Gilmour, D. T. and Furlong, E. E. (2009) 'A systematic analysis of Tinman function reveals Eya and JAK-STAT signaling as essential regulators of muscle development', *Dev Cell* 16(2): 280-91.

MacKrell, A. J., Blumberg, B., Haynes, S. R. and Fessler, J. H. (1988) 'The lethal myospheroid gene of Drosophila encodes a membrane protein homologous to vertebrate integrin beta subunits', *Proc Natl Acad Sci U S A* 85(8): 2633-7.

Mauro, A. (1961) 'Satellite cell of skeletal muscle fibers', *J Biophys Biochem Cytol* 9: 493-5.

Medioni, C., Astier, M., Zmojdzian, M., Jagla, K. and Semeriva, M. (2008) 'Genetic control of cell morphogenesis during Drosophila melanogaster cardiac tube formation', *J Cell Biol* 182(2): 249-61.

Mennerich, D., Schafer, K. and Braun, T. (1998) 'Pax-3 is necessary but not sufficient for Ibx1 expression in myogenic precursor cells of the limb', *Mech Dev* 73(2): 147-58.

Michelson, A. M., Abmayr, S. M., Bate, M., Arias, A. M. and Maniatis, T. (1990) 'Expression of a MyoD family member prefigures muscle pattern in Drosophila embryos', *Genes Dev* 4(12A): 2086-97.

Miner, J. H. and Wold, B. (1990) 'Herculin, a fourth member of the MyoD family of myogenic regulatory genes', *Proc Natl Acad Sci U S A* 87(3): 1089-93.

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(7): 1125-36.

Montarras, D., Morgan, J., Collins, C., Relaix, F., Zaffran, S., Cumano, A., Partridge, T. and Buckingham, M. (2005) 'Direct isolation of satellite cells for skeletal muscle regeneration', *Science* 309(5743): 2064-7.

Moore, C. A., Parkin, C. A., Bidet, Y. and Ingham, P. W. (2007) 'A role for the Myoblast city homologues Dock1 and Dock5 and the adaptor proteins Crk and Crk-like in zebrafish myoblast fusion', *Development* 134(17): 3145-53.

Muller, D., Jagla, T., Bodart, L. M., Jahrling, N., Dodt, H. U., Jagla, K. and Frasch, M. (2010) 'Regulation and functions of the *lms* homeobox gene during development of embryonic lateral transverse muscles and direct flight muscles in *Drosophila*', *PLoS One* 5(12): e14323.

Munsterberg, A. E. and Lassar, A. B. (1995) 'Combinatorial signals from the neural tube, floor plate and notochord induce myogenic bHLH gene expression in the somite', *Development* 121(3): 651-60.

Nathan, E., Monovich, A., Tirosh-Finkel, L., Harrelson, Z., Rousso, T., Rinon, A., Harel, I., Evans, S. M. and Tzahor, E. (2008) 'The contribution of *Islet1*-expressing splanchnic mesoderm cells to distinct branchiomic muscles reveals significant heterogeneity in head muscle development', *Development* 135(4): 647-57.

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 U S A* 91(16): 7520-4.

Noden, D. M. (1983) 'The role of the neural crest in patterning of avian cranial skeletal, connective, and muscle tissues', *Dev Biol* 96(1): 144-65.

Nose, A. (2012) 'Generation of neuromuscular specificity in *Drosophila*: novel mechanisms revealed by new technologies', *Front Mol Neurosci* 5: 62.

Nose, A., Isshiki, T. and Takeichi, M. (1998) 'Regional specification of muscle progenitors in *Drosophila*: the role of the *msh* homeobox gene', *Development* 125(2): 215-23.

Nose, A., Mahajan, V. B. and Goodman, C. S. (1992) 'Connectin: a homophilic cell adhesion molecule expressed on a subset of muscles and the motoneurons that innervate them in *Drosophila*', *Cell* 70(4): 553-67.

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(3): e1002547.

Nusslein-Volhard, C. and Wieschaus, E. (1980) 'Mutations affecting segment number and polarity in *Drosophila*', *Nature* 287(5785): 795-801.

Paterson, B. M., Walldorf, U., Eldridge, J., Dubendorfer, A., Frasch, M. and Gehring, W. J. (1991) 'The *Drosophila* homologue of vertebrate myogenic-determination genes encodes a transiently expressed nuclear protein marking primary myogenic cells', *Proc Natl Acad Sci U S A* 88(9): 3782-6.

Powell, G. T. and Wright, G. J. (2012) 'Do muscle founder cells exist in vertebrates?', *Trends Cell Biol.*

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(1): 169-81.

Reuter, R. and Leptin, M. (1994) 'Interacting functions of snail, twist and huckebein during the early development of germ layers in *Drosophila*', *Development* 120(5): 1137-50.

Richard, A. F., Demignon, J., Sakakibara, I., Pujol, J., Favier, M., Strohlic, L., Le Grand, F., Sgarioto, N., Guernec, A., Schmitt, A. (2011). 'Genesis of muscle fiber-type diversity during mouse embryogenesis relies on *Six1* and *Six4* gene expression', *Dev Biol* 359(2): 303-20.

Riechmann, V., Irion, U., Wilson, R., Grosskortenhaus, R. and Leptin, M. (1997) 'Control of cell fates and segmentation in the *Drosophila* mesoderm', *Development* 124(15): 2915-22.

Riechmann, V., Rehorn, K. P., Reuter, R. and Leptin, M. (1998) 'The genetic control of the distinction between fat body and gonadal mesoderm in *Drosophila*', *Development* 125(4): 713-23.

Ruiz-Gomez, M., Coutts, N., Price, A., Taylor, M. V. and Bate, M. (2000) '*Drosophila* dumbfounded: a myoblast attractant essential for fusion', *Cell* 102(2): 189-98.

Ruiz-Gomez, M., Romani, S., Hartmann, C., Jackle, H. and Bate, M. (1997) 'Specific muscle identities are regulated by Kruppel during Drosophila embryogenesis', *Development* 124(17): 3407-14.

Ruiz Gomez, M. and Bate, M. (1997) 'Segregation of myogenic lineages in Drosophila requires numb', *Development* 124(23): 4857-66.

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(7): 1979-88.

Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V. and Furlong, E. E. (2007) 'A core transcriptional network for early mesoderm development in Drosophila melanogaster', *Genes Dev* 21(4): 436-49.

Sandmann, T., Jensen, L. J., Jakobsen, J. S., Karzynski, M. M., Eichenlaub, M. P., Bork, P. and Furlong, E. E. (2006) 'A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development', *Dev Cell* 10(6): 797-807.

Schaub, C., Nagaso, H., Jin, H. and Frasch, M. (2012) 'Org-1, the Drosophila ortholog of Tbx1, is a direct activator of known identity genes during muscle specification', *Development* 139(5): 1001-12.

Schnorrer, F. and Dickson, B. J. (2004) 'Muscle building; mechanisms of myotube guidance and attachment site selection', *Dev Cell* 7(1): 9-20.

Schnorrer, F., Kalchauer, I. and Dickson, B. J. (2007) 'The transmembrane protein Kon-tiki couples to Dgrip to mediate myotube targeting in Drosophila', *Dev Cell* 12(5): 751-66.

Schweitzer, R., Zelzer, E. and Volk, T. (2010) 'Connecting muscles to tendons: tendons and musculoskeletal development in flies and vertebrates', *Development* 137(17): 2807-17.

Sens, K. L., Zhang, S., Jin, P., Duan, R., Zhang, G., Luo, F., Parachini, L. and Chen, E. H. (2010) 'An invasive podosome-like structure promotes fusion pore formation during myoblast fusion', *J Cell Biol* 191(5): 1013-27.

Shishido, E., Takeichi, M. and Nose, A. (1998) 'Drosophila synapse formation: regulation by transmembrane protein with Leu-rich repeats, CAPRICIOUS', *Science* 280(5372): 2118-21.

Soler, C., Han, J. and Taylor, M. V. (2012) 'The conserved transcription factor Mef2 has multiple roles in adult Drosophila musculature formation', *Development* 139(7): 1270-5.

Speicher, S., Fischer, A., Knoblich, J. and Carmena, A. (2008) 'The PDZ protein Canoe regulates the asymmetric division of Drosophila neuroblasts and muscle progenitors', *Curr Biol* 18(11): 831-7.

Srinivas, B. P., Woo, J., Leong, W. Y. and Roy, S. (2007) 'A conserved molecular pathway mediates myoblast fusion in insects and vertebrates', *Nat Genet* 39(6): 781-6.

St Johnston, D. and Nusslein-Volhard, C. (1992) 'The origin of pattern and polarity in the Drosophila embryo', *Cell* 68(2): 201-19.

Staehling-Hampton, K., Hoffmann, F. M., Baylies, M. K., Rushton, E. and Bate, M. (1994) 'dpp induces mesodermal gene expression in Drosophila', *Nature* 372(6508): 783-6.

Stolfi, A., Gainous, T. B., Young, J. J., Mori, A., Levine, M. and Christiaen, L. (2010) 'Early chordate origins of the vertebrate second heart field', *Science* 329(5991): 565-8.

Strumpf, D. and Volk, T. (1998) 'Kakapo, a novel cytoskeletal-associated protein is essential for the restricted localization of the neuregulin-like factor, vein, at the muscle-tendon junction site', *J Cell Biol* 143(5): 1259-70.

Strunkelberg, M., Bonengel, B., Moda, L. M., Hertenstein, A., de Couet, H. G., Ramos, R. G. and Fischbach, K. F. (2001) 'rst and its paralogue kirre act redundantly during embryonic muscle development in Drosophila', *Development* 128(21): 4229-39.

Sweeton, D., Parks, S., Costa, M. and Wieschaus, E. (1991) 'Gastrulation in Drosophila: the formation of the ventral furrow and posterior midgut invaginations', *Development* 112(3): 775-89.

Tajbakhsh, S., Borello, U., Vivarelli, E., Kelly, R., Papkoff, J., Duprez, D., Buckingham, M. and Cossu, G. (1998) 'Differential activation of Myf5 and MyoD by different Wnts in explants of mouse paraxial mesoderm and the later activation of myogenesis in the absence of Myf5', *Development* 125(21): 4155-62.

Tajbakhsh, S., Rocancourt, D., Cossu, G. and Buckingham, M. (1997) 'Redefining the genetic hierarchies controlling skeletal myogenesis: Pax-3 and Myf-5 act upstream of MyoD', *Cell* 89(1): 127-38.

Tao, Y., Wang, J., Tokusumi, T., Gajewski, K. and Schulz, R. A. (2007) 'Requirement of the LIM homeodomain transcription factor tailup for normal heart and hematopoietic organ formation in *Drosophila melanogaster*', *Mol Cell Biol* 27(11): 3962-9.

Tapanes-Castillo, A. and Baylies, M. K. (2004) 'Notch signaling patterns *Drosophila* mesodermal segments by regulating the bHLH transcription factor twist', *Development* 131(10): 2359-72.

Taylor, M. V. (1995) 'Muscle development. Making *Drosophila* muscle', *Curr Biol* 5(7): 740-2.

Thisse, B., el Messal, M. and Perrin-Schmitt, F. (1987) 'The twist gene: isolation of a *Drosophila* zygotic gene necessary for the establishment of dorsoventral pattern', *Nucleic Acids Res* 15(8): 3439-53.

Thisse, B., Stoetzel, C., Gorostiza-Thisse, C. and Perrin-Schmitt, F. (1988) 'Sequence of the twist gene and nuclear localization of its protein in endomesodermal cells of early *Drosophila* embryos', *Embo J* 7(7): 2175-83.

Tirosh-Finkel, L., Elhanany, H., Rinon, A. and Tzahor, E. (2006) 'Mesoderm progenitor cells of common origin contribute to the head musculature and the cardiac outflow tract', *Development* 133(10): 1943-53.

Tolkin, T. and Christiaen, L. (2012) 'Development and evolution of the ascidian cardiogenic mesoderm', *Curr Top Dev Biol* 100: 107-42.

Trainor, P. A. and Tam, P. P. (1995) 'Cranial paraxial mesoderm and neural crest cells of the mouse embryo: co-distribution in the craniofacial mesenchyme but distinct segregation in branchial arches', *Development* 121(8): 2569-82.

Tzahor, E. and Evans, S. M. (2011) 'Pharyngeal mesoderm development during embryogenesis: implications for both heart and head myogenesis', *Cardiovasc Res* 91(2): 196-202.

Venkatesh, T. V., Park, M., Ocorr, K., Nemaceck, J., Golden, K., Wemple, M. and Bodmer, R. (2000) 'Cardiac enhancer activity of the homeobox gene tinman depends on CREB consensus binding sites in *Drosophila*', *Genesis* 26(1): 55-66.

Venken, K. J., Schulze, K. L., Haelterman, N. A., Pan, H., He, Y., Evans-Holm, M., Carlson, J. W., Levis, R. W., Spradling, A. C., Hoskins, R. A. et al. (2011) 'MiMIC: a highly versatile transposon insertion resource for engineering *Drosophila melanogaster* genes', *Nat Methods* 8(9): 737-43.

Venuti, J. M., Morris, J. H., Vivian, J. L., Olson, E. N. and Klein, W. H. (1995) 'Myogenin is required for late but not early aspects of myogenesis during mouse development', *J Cell Biol* 128(4): 563-76.

Vincent, S. D. and Buckingham, M. E. (2009). 'How to make a heart: the origin and regulation of cardiac progenitor cells', *Curr Top Dev Biol* 90: 1-41.

Vorbruggen, G. and Jackle, H. (1997) 'Epidermal muscle attachment site-specific target gene expression and interference with myotube guidance in response to ectopic stripe expression in the developing *Drosophila* epidermis', *Proc Natl Acad Sci U S A* 94(16): 8606-11.

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(22): 4959-69.

Wei, Q., Rong, Y. and Paterson, B. M. (2007) 'Stereotypic founder cell patterning and embryonic muscle formation in *Drosophila* require nautilus (MyoD) gene function', *Proc Natl Acad Sci U S A* 104(13): 5461-6.

Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T. K., Turner, D., Rupp, R., Hollenberg, S. et al. (1991) 'The myoD gene family: nodal point during specification of the muscle cell lineage', *Science* 251(4995): 761-6.

Wigmore, P. M. and Evans, D. J. (2002) 'Molecular and cellular mechanisms involved in the generation of fiber diversity during myogenesis', *Int Rev Cytol* 216: 175-232.

Wong, M. C., Castanon, I. and Baylies, M. K. (2008) 'Daughterless dictates Twist activity in a context-dependent manner during somatic myogenesis', *Dev Biol* 317(2): 417-29.

- Wu, X., Golden, K. and Bodmer, R. (1995) 'Heart development in *Drosophila* requires the segment polarity gene *wingless*', *Dev Biol* 169(2): 619-28.
- Xu, X., Yin, Z., Hudson, J. B., Ferguson, E. L. and Frasch, M. (1998) 'Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the *Drosophila* mesoderm', *Genes Dev* 12(15): 2354-70.
- 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(24): 4971-82.
- Zaffran, S., Reim, I., Qian, L., Lo, P. C., Bodmer, R. and Frasch, M. (2006) 'Cardioblast-intrinsic *Tinman* activity controls proper diversification and differentiation of myocardial cells in *Drosophila*', *Development* 133(20): 4073-83.
- Zhang, M. and McLennan, I. S. (1998) 'Primary myotubes preferentially mature into either the fastest or slowest muscle fibers', *Dev Dyn* 213(1): 147-57.

Summary

The somatic musculature of the *Drosophila* embryo is a classical model to study the regulatory processes that generate cellular diversity. Muscle formation is a multistep process: the first step is the specification, within the mesoderm, of a group of competent cells, called promuscular cluster. The second step is the selection of a progenitor cell (PC) from this cluster. Asymmetric division of each PC then generates muscle founder cells (FC). Finally, each FC undergoes a fusion process with fusion competent myoblasts (FCM) to generate a muscle fiber. Each muscle is formed of a single multinucleate fiber. Each *Drosophila* muscle has a specific identity, as it can be distinguished by its position, shape, orientation, attachment, and innervation pattern. Muscle identity reflects the expression by each PC/FC of a specific combination of identity Transcription Factors (iTF). In the laboratory, we study the control of muscle identity, using as entry point, the expression and requirement of the iTF Collier (Col) during development of a dorso-lateral (DA3) muscle.

I started my PhD by characterizing *col* transcriptional regulation during early steps of DA3 muscle formation. Starting from computational predictions, I identified an early *col* cis regulatory module (Early CRM) responsible for *col* activation in a promuscular cluster. A late *col* CRM, active from the PC stage, had previously been characterized in the laboratory. To determine with more precision the temporal windows of activity of each of these CRM, I designed a novel intron-containing reporter gene in order to detect primary transcripts. This allowed me to show that the late and the early CRMs together reproduce precisely the endogenous *col* expression pattern. Characterization of the early mesodermal *col* CRM also allowed to do lineage experiments and determine the fate of FCMs that transiently express Col at the promuscular stage. I found that these myoblasts contribute mostly to dorso-lateral muscles.

During the second part of my thesis, I described a new role of the LIM-homeodomain TF Tailup/Islet1 (Tup) in specifying dorsal muscles. I first showed that Tup is specifically expressed in the four dorsal muscles. In *tup* null mutants, on one hand, the dorsal musculature is severely disorganized and, on the other hand, the dorsal DA2 muscle ectopically expresses Col and is transformed into a dorso-lateral DA3-like muscle. I showed that the DA2 PC is singled out from the Col promuscular cluster when cells of this cluster still express (transitorily) the homeodomain TF Tinman/Nkx2.5 (Tin). The DA2 PC gives rise to the DA2 FC and a (dorso-lateral) adult muscle precursor (AMP). *tup* activation by Tin in the DA2 PC is required to repress *col* and establish a DA2 instead of DA3 identity.

In conclusion, my work allowed to propose a model which connects a temporal sequence of transcriptional regulation of iTFs to the specification of muscle PC identity and final muscle pattern. It provides a novel, dynamic view of how muscle identity is specified. These findings also provide novel parallels with the specification of pharyngeal muscles in vertebrates.

Key words: *Drosophila*, Myogenesis, Muscle identity program, Cis-regulatory modules, Transcriptional regulatory networks.

Hadi BOUKHATMI

Mise en place de l'identité des muscles au cours de la spécification des myoblastes chez la Drosophile

Sous la direction de Laurence Dubois et Alain Vincent

Soutenue à Toulouse, le 26 Octobre 2012

Résumé

La formation des muscles squelettiques au cours de l'embryogenèse de la drosophile est un modèle d'étude du contrôle génétique de la différenciation cellulaire. La formation de chaque muscle comprend quatre étapes successives: spécification d'un groupe promusculaire, sélection d'un progéniteur (PC) à partir de ce groupe, division asymétrique de ce progéniteur pour donner des cellules fondatrices de muscles (FC) ; fusion de chaque FC avec des myoblastes compétents (FCM), suivie de la différenciation musculaire. Chaque muscle squelettique est composé d'une fibre. Chaque muscle présente des propriétés spécifiques de taille, forme, position, attachement, et patron d'innervation. Ces propriétés sont groupées sous le terme d'identité musculaire. Cette identité est conférée par l'expression dans chaque PC/FC d'une combinatoire de Facteurs de Transcription identitaires (FTi). Notre laboratoire étudie ce processus, en utilisant comme point d'entrée l'expression et les rôles du FTi Collier (Col) au cours du développement d'un muscle dorso-latéral, le muscle DA3 (Dorsal Acute 3).

Au cours de la première partie de ma thèse, j'ai étudié la régulation transcriptionnelle de *col* durant les phases de spécification des groupes promusculaires et de sélection du PC à l'origine du muscle DA3. Partant de prédictions bioinformatiques j'ai caractérisé le module cis régulateur (CRM) de *col* actif durant ces phases (CRM précoce). Un CRM « tardif », actif du stade progéniteur à la complétion de la formation du muscle DA3, avait été préalablement caractérisé dans l'équipe. Afin de déterminer plus précisément les fenêtres temporelles d'activité des deux CRM mésodermiques de *col*, j'ai mis au point un nouveau gène rapporteur comportant un intron permettant de détecter les transcrits primaires. Ceci m'a permis de montrer que les CRM précoce et tardif reproduisent ensemble l'expression endogène de *col*. La caractérisation du CRM précoce de *col* m'a aussi permis de suivre le destin des FCM du groupe promusculaire Col dans les embryons tardifs et de montrer que ces FCM contribuent uniquement à des muscles dorsaux-latéraux.

Au cours de la deuxième partie de ma thèse, j'ai caractérisé le rôle, inconnu jusqu'alors, du FT à domaine LIM-Homeodomaine Tailup (Tup)/Islet1 dans la myogenèse. J'ai d'abord montré que Tup est spécifiquement exprimé dans les 4 muscles les plus dorsaux. L'analyse de mutants m'a permis de montrer qu'en absence de Tup, le muscle dorsal DA2 exprime Col et est transformé en muscle dorso-latéral de type DA3. J'ai ensuite montré que le PC du DA2 est à l'origine de la FC DA2 et d'un précurseur musculaire adulte (AMP). Ce PC est sélectionné à partir du groupe promusculaire Col quand les cellules de ce groupe expriment encore le FT à homéodomaine Tinman/NKx2.5. Tin active *tup* dans le PC DA2. Tup, en retour, réprime *col* et cette répression permet de distinguer les identités musculaires DA2 et DA3.

En conclusion, mes travaux de thèse m'ont permis de proposer un nouveau modèle permettant de relier le processus de spécification des progéniteurs au contrôle temporel et spatial de l'expression des FTi. Une vision dynamique de ce processus de spécification permet de mieux comprendre le programme identitaire propre à chaque muscle. L'analyse des interactions entre Tin, Tup, et Col au cours de la formation des muscles dorsaux révèle de nouveaux parallèles avec les interactions entre Nkx2.5, Islet, EBF au cours de la formation des muscles pharyngaux chez les chordés.

Mots Clefs :

Drosophile, Myogenèse, Programme d'identité musculaire, Eléments cis-régulateurs, Réseaux de régulation transcriptionnelle.

Discipline :

Biologie Moléculaire, Cellulaire et du Développement.