# Regulatory roots of cranial muscle diversity : developmental origins of muscular resilience

Alexandre Grimaldi

Sorbonne Université

ED515-Complexité du vivant

*Département de Biologie du Développement et des Cellules Souches*

*Unité Cellules Souches et Développement, CNRS-UMR3738*

*Institut Pasteur Paris France*

# Regulatory roots of cranial muscle diversity

***Developmental origins of muscular resilience***

# Fondements régulatoires de la diversité des muscles faciaux

***Origines développementales de la résilience musculaire***

Alexandre Grimaldi

Thèse de doctorat de Sciences

Dirigée par Shahragim Tajbakhsh

Présentée et soutenue publiquement le 08/10/2020

Devant un jury composé de :

Mme. la Dr. Delphine Duprez ---------- Présidente du jury

M. le Dr. Jaime Carvajal ---------- Rapporteur

M. le Dr. Ramkumar Sambasivan---------- Rapporteur

M. le Dr. Pascal Maire----------Examinateur

Mme. la Pr. Ana Cumano ----------Examinatrice

M. le Dr. Robert Kelly----------Examinateur

M. le Pr. Shahragim Tajbakhsh ---------- Directeur de thèse

*To my grand-father, who hereditarily or not, passed his love of science on to me.*

*A mon grand-père, qui de manière héréditaire ou non, m'a transmis son amour de la science.*

# Acknowledgements

First and foremost, I would like to deeply thank all the members of the jury for accepting to evaluate my work. In these particular times where we all have to constantly re-establish our priorities; I sincerely appreciate that they take the time to examine my work of those last 4 years.

I would like to particularly thank Dr. Ramkumar Sambasivan and Dr. Jaime Carvajal for accepting to review this manuscript. Their contribution and expertise in this field has profoundly shaped this project and I am looking forward to their comments and suggestions.

I would also like to thank Dr. Pascal Maire for having been part of my thesis committee throughout these years, and provided me with great comments and encouragements.

I would like to thank Shahragim, for his mentoring, his trust and his availability. Despite his tight schedules, I have never felt that he was too busy for me scientifically or non scientifically. I have learnt, grown and become even more passionate evolving in his group, and I believe he is, in big part, responsible for that. Thank you for your dedication, mentorship and trust.

I would like to thank all the major contributors to the exploitation of scRNAseq data, whose work greatly benefitted this thesis, particularly the Satija Lab, the Theis Lab, the Aerts Lab, the Trapnell Lab, the Karchenko Lab, the Linnarsson and the Galon Lab.

I want to thank Caro, who has been by my side during all these years. You have grown even more patient, kind, and supportive of me over the years and even though these values are inherent to your nature, I do not want to just take them for granted. You have kept me grounded, sane and passionate, and no matter what I will ever try to do, I know that you will respect me and believe in me. Thank you.

I would like to thank my friends and colleagues, especially Maria and Dani, with whom I have shared great moments inside and outside the lab. We might have become more like-minded after 4 years, but I think we were already well matched when we met. There is something very enjoyable in being quickly understood, scientifically or non-scientifically, and this is how I

always felt with you. I am lucky I had you both during these years, and I cannot imagine them without you. Thank you for providing me with such comfort, support, patience, and friendship.

Je voudrais remercier tous mes collègues du laboratoire de Shahragim passés et présents, en particulier Glenda et Eglantine pour leur aide et soutien dès mon arrivée et ce jusqu'à maintenant, Sébastien pour toute son aide et nos longues discussions d'analyse de scRNAseq et d'autres, Marc, pour sa bonne humeur, son énergie paranormale et tout le travail qu'il fournit avec maintenant Priscilla et Chloé (et Gilles Sylvain par le passé) pour nous permettre de travailler dans les meilleures conditions.

J'aimerais remercier tous mes autres amis et collègues du département, Eric, Laure, Pier, Juan avec qui j'ai toujours passé de très bons moments.

J'aimerais remercier mes amis musiciens, particulièrement Selene. On a tellement partagé en peu de temps que c'est comme si on s'était toujours connus, et on ne se perdra jamais vraiment. Merci à Pedro de m'avoir fait découvrir un nouveau monde musical et de nouvelles inspirations dont j'ignorais l'existence. Merci à Khalil, Pier et Germano et tous les Ceceritos pour tous ces bons moments partagés à jouer ensemble, au Music Lab ou sur scène.

Merci à tous les membres du département Cellules Souches et Développement de l'Institut Pasteur pour avoir contribué de près ou de loin à mon épanouissement professionnel et personnel en organisant et en participant aux évènements sociaux et scientifiques, au cours desquels j'ai pu découvrir de nouveaux collègues et amis. Je garde de très bons souvenirs de toutes les rencontres, discussions, soirées et concerts organisés par les étudiants, PostDocs, techniciens, ingénieurs, PIs et membres de l'admnistration du département, qui ont réellement renforcé mon attachement à l'institut et enrichi mon cadre de travail.

Merci à mes amis de longue date, Ulysse, Arthur et Kevin, qui ont toujours été là et qui le seront toujours. Et bien sûr j'aimerais remercier ma famille, mes parents, mon frère et mes grands-parents, qui m'ont toujours soutenu et ont toujours veillé à mon bien-être où que je sois.

# Index


**INDEX** ..... **4**

**INTRODUCTION** ..... **11**

**CHAPTER I** ..... **15**

**THE RULES OF MYOGENESIS: A COMMON GROUND** ..... **15**

1. MESODERMAL ORIGIN ..... 15
2. THE MYOGENIC REGULATORY FACTORS ..... 17
3. FINE-TUNING THE FUNCTION OF THE MYOGENIC REGULATORY FACTORS ..... 19
   - *A. Cofactors and repressors* ..... *19*
   - *B. microRNAs* ..... *20*
4. MYOGENESIS: A MULTISTEP PROCESS ..... 22
   - *A. Primary and secondary myogenesis* ..... *22*
   - *B. Fetal and perinatal stages* ..... *23*
     - a) Muscle stem cells self-regulate ..... 23
     - b) Support from stromal cells ..... 24
5. ADULT MUSCLE STEM CELLS AND THEIR NICHE ..... 24
   - *A. Satellite cells: the muscle stem cells* ..... *25*
   - *B. The muscle stem cell niche* ..... *25*
     - a) Intrinsic and autocrine control of MuSC quiescence ..... 25
     - b) Extrinsic and paracrine control of MuSC quiescence ..... 27
     - c) Extrinsic plasticity: myogenic interstitial cells ..... 27

**CHAPTER II** ..... **30**

**AN INTRIGUING HETEROGENEITY** ..... **30**

1. MYOPATHIES AND REGENERATIVE CAPACITIES ..... 30
   - *A. Limb* ..... *30*
   - *B. Extraocular muscles* ..... *31*
   - *C. Masseter* ..... *31*
   - *D. Pharyngeal muscles* ..... *31*
   - *E. Diaphragm* ..... *32*
   - *F. Laryngeal muscles* ..... *32*
2. DISTINCT GENETIC REQUIREMENTS OF SKELETAL MUSCLE PROGENITORS ..... 32

**CHAPTER III** ..... **38**

**INTRINSIC CUES TO DIVERSITY** ..... **38**

1. TRUNK AND LIMB MYOGENESIS ..... 38
   - *A. Somites are the source of trunk and limb skeletal muscles* ..... *38*
   - *B. Initiation of limb muscles* ..... *41*
   - *C. Gene regulatory networks* ..... *42*
1. HEAD MYOGENESIS ..... 45
   - *A. The pharyngeal arches: the source of branchiomeric muscles* ..... *46*
   - *B. The caudal arch muscles* ..... *48*
   - *C. Somitic head muscles: neck and tongue* ..... *51*
   - *D. Gene regulatory network* ..... *52*
     - a) Tbx1 ..... 53
     - b) Tcf21/Msc ..... 54
     - c) Six family ..... 55
     - d) Pitx2 ..... 56
     - e) Lhx2 ..... 56
     - f) Isl1 ..... 56
     - g) cMet ..... 57
     - h) Cxcr4/Sdf1 ..... 57

**CHAPTER IV** ..... **59**

**EXTRINSIC CUES TO DIVERSITY** ..... **59**

3. TRUNK AND LIMB CONNECTIVE TISSUE GUIDES MUSCLE FORMATION ..... 59
*A. Primaxial muscle connective tissue* ..... *60*
*B. Abaxial muscle connective tissue* ..... *61*
a) Connective tissue of ventral muscles ..... 61
b) Limb muscle connective tissue ..... 62
3. HEAD CONNECTIVE TISSUE ..... 65
*A. The neural crest guides head muscle formation* ..... *65*
a) Origin and contribution ..... 66
b) Role of NCC in muscle patterning ..... 66

**CHAPTER V** ..... **71**

**WHEN INTRINSIC MEETS EXTRINSIC** ..... **71**

1. IN THE EMBRYO ..... 71
2. IN THE ADULT ..... 72
3. A POTENTIAL BIPOTENT PARADIGM: MEDIAL CPM ..... 73

**CHAPTER VI** ..... **77**

**MUSCLE MEMORY** ..... **77**

1. MUSCLE IDENTITY: INTRINSIC OR EXTRINSIC? ..... 78
2. ROLE FOR DEVELOPMENTAL REGULATORS IN ADULT ..... 79

**RESULTS** ..... **81**

PART 1 ..... 83
BIPOTENT PROGENITORS GIVE RISE TO MYOGENIC AND CONNECTIVE TISSUE IN THE DEVELOPING EMBRYO ..... 83
PART 2 ..... 145
GENETIC BIRTHMARKS REGULATE CRANIAL MYOGENIC DIVERSITY ..... 145

**DISCUSSION** ..... **177**

1. MYF5+ BIPOTENT PROGENITORS GIVE RISE TO CRANIAL MUSCLES AND ASSOCIATED CONNECTIVE TISSUE ..... 178
2. GENETIC BIRTHMARKS REGULATE CRANIAL MYOGENIC DIVERSITY ..... 180
3. A FIBROGENIC STATE IN MYOGENIC PROGENITORS ..... 181
4. SCRNASEQ ..... 182
*Advantages* ..... *182*
*Limitations* ..... *184*
*Extracting valuable information* ..... *186*
5. EVOLUTION OF CAUDAL ARCHES ..... 186
*Axial muscles as an ancestral feature* ..... *187*
*Vertebrates* ..... *187*
*Remodeling of caudal arches in vertebrates* ..... *188*

**ANNEXES** ..... **191**

**REFERENCES** ..... ERREUR ! SIGNET NON DEFINI.

# Abbreviations:

bHLH: basic helix-loop-helix

CPM: Cardiopharyngeal mesoderm

ECM: Extracellular matrix

EOM: Extraocular muscles

ESM: Esophageal striated muscle

GBM: Genetic birthmark

LPM: Lateral plate mesoderm

MyHC: Myosing Heavy Chain

MRF: Myogenic regulatory factor

MuSC: Muscle stem cell

NCCs: Neural crest cells

PA: Pharyngeal arch

EMT: Epithelial to mesenchymal transition

# Résumé

Les muscles squelettiques sont présents dans tout le corps et présentent un niveau surprenant d'hétérogénéité, dans leur susceptibilité aux maladies, potentiel de régénération ou capacités métaboliques. Cette diversité est également retrouvée au cours du développement embryonnaire où les cellules myogéniques et non myogéniques établissent le système musculo-squelettique. La tête et le cou sont constitués d'une grande variété de muscles qui remplissent des fonctions essentielles, mais nous en savons peu sur la biologie des muscles craniofaciaux. Ces structures sont associées à l'émergence de cellules de la crête neurale (CCN) qui donnent naissance à la plupart des tissus non myogéniques crâniens et qui sont cruciales à la formation des muscles. Cependant, certains muscles crâniens sont privés de CCN, et nous ignorons comment les cellules myogéniques et non myogéniques contribuent à ces domaines. Cette thèse fournit des preuves démontrant que les progéniteurs en amont du muscle se détournent du programme myogénique pour donner naissance au tissu conjonctif. Nous avons utilisé une approche de single-cell RNAseq non biaisée et restreinte avec différentes lignées transgéniques de souris à des stades embryonnaires distincts, des marquages in situ et de nouvelles méthodes analytiques, et avons montré que les progéniteurs bipotents issus du mésoderme exprimant le gène de détermination musculaire Myf5 donnent naissance au muscle squelettique et au tissu conjonctif anatomiquement associé dans les muscles partiellement privés de CCN. Cette transition est caractérisée par une complémentarité de signalisation de récepteurs tyrosine kinase entre les cellules musculaires et non musculaires, ainsi que par des modules régulateurs distincts. Les muscles crâniens proviennent également de différentes lignées qui impliquent l'activité de cascades de régulation génique spécifiques. Ici, nous avons utilisé une approche non biaisée et large pour découvrir des modules de régulation spécifiques qui sous-tendent différentes populations de cellules myogéniques dans la tête et à travers plusieurs stades de développement. Certaines de ces « tâches de naissance génétiques » uniques sont des facteurs de transcription spécifiques et sont conservées dans les cellules souches musculaires adultes, ce qui indique que leur importance potentielle est de fournir les propriétés uniques qui ont été signalées pour différentes populations de cellules souches musculaires. Enfin, ces études utilisent des méthodes analytiques inédites qui bénéficient des dernières avancées algorithmiques et offrent de nouvelles perspectives pour la découverte de processus biologiques à partir de données à haut débit.

# Abstract

Skeletal muscles are found throughout the body and they display a surprising level of heterogeneity in properties and function. For example, some muscles are specifically susceptible to diseases, and some have better regenerative potential or different metabolic capacities. Diversity is also found during embryonic development where myogenic and non-myogenic cells establish the musculoskeletal system. The head and neck are comprised of a wide variety of muscles that perform essential functions such as feeding, breathing and vocalising, yet little is known about craniofacial muscle biology. Novel structures are associated with the emergence of neural crest cells (NCC) which give rise to most craniofacial connective tissue, cartilage and bone and are crucial for muscle morphogenesis. However, some cranial muscles are deprived of NCC, and it is unclear how myogenic and non-myogenic cells contribute to those domains. This thesis provides evidence demonstrating that upstream progenitors redirect from the myogenic program to give rise to the muscle-associated connective tissue that supports the formation of muscular structures. We employed unbiased and lineage-restricted single-cell RNAseq using different mouse transgenic lines at distinct embryonic stages, in situ labelling, and new analytical methods, and show that bipotent progenitors expressing the muscle determination gene *Myf5* give rise to skeletal muscle and anatomically associated connective tissue in distinct muscle groups spatiotemporally. Notably, this property was restricted to muscles with only partial contribution from NCCs suggesting that in their absence, the balance of myogenic and connective tissue cells is undertaken by somite-derived or cranial-derived mesoderm. This transition is characterised by a complementarity of tyrosine kinase receptor signalling between muscle and non-muscle cells, as well as distinct regulatory modules. Cranial muscles also originate from different lineages that involve the activity of specific gene regulatory cascades. Here, we used an all-inclusive unbiased approach to uncover specific regulatory modules that underlie different myogenic cell populations in the head and across multiple developmental stages. Some of these unique "genetic birthmarks" are specific transcription factors, and are retained in adult muscle stem cells pointing to their potential importance is delivering the unique properties that have been reported for different muscle stem cell populations. Finally, these studies employ novel computational methods that benefit from the latest algorithmic advancements and they provide prospects for the discovery of new biological processes from high throughput data.

# INTRODUCTION

# Objectives

Skeletal muscles accounts for about 40% of the body mass, and they are required for voluntary movement. Building and preserving muscle tissue is widely regarded as a determining factor for maintaining a regulated metabolism, reducing risks of obesity, diabetes, cardiovascular defects, increasing longevity, and delaying the onset of ageing. Loss of muscle function during injury, disease or ageing can significantly hamper quality of life and can be a direct cause of death. In recent years, numerous studies have demonstrated remarkable improvements in therapeutic approaches for various myopathies, as illustrated by the use of viral vectors for targeted gene therapy. Although these biotechnological advances are undeniably promising, currently most treatments are disease-specific. To date, numerous disorders cannot be treated, and they worsen with age. Additionally, skeletal muscles are inherently heterogeneous, such that distinct diseases will affect different muscles. Skeletal muscles possess different proliferative and regenerative capacities and arise from diverse developmental origins. Throughout their lifetime, they will receive signaling cues from their microenvironment that will affect their commitment and self-renewal capacities. This raises a number of questions: 1) What specific factors underly muscle diversity? 2) Can certain factors safeguard specific muscles against lasting damage, or confer plasticity? 3) What are the relative contributions of intrinsic and extrinsic cues on muscle function? This thesis aims at initiating a comprehensive 4D map of the developing head and neck musculature, with a specific focus on the unique molecular cues characterising various muscles groups and how they may control cell fate decisions. Building on this knowledge can shed light on the specific pathological outcomes of various muscle diseases and inform on new therapeutic approaches.

# Chapter I

# The rules of myogenesis: a common ground

## 1. Mesodermal origin

All skeletal muscles derive from mesodermal embryonic progenitors (**Figure 1**). Mesoderm is formed during early embryogenesis (3rd week in humans, embryonic day (E) 6 in mice), during a process known as gastrulation (Solnica-Krezel and Sepich, 2012). Cells located in the inner cell mass of the embryo, called epiblast cells, undergo an epithelial-to-mesenchymal transition (EMT) and delaminate along a primitive streak towards the ventral side to first form a second embryonic layer called endoderm. Subsequently, epiblasts will continue to delaminate ventrally, proliferate, leading to the appearance of a third middle layer: the mesoderm. By the end of gastrulation, three embryonic layers are formed: the ectoderm, mesoderm and endoderm. The ectoderm gives rise to multiple tissues including skin, central and peripheral nervous system, while endoderm gives rise to the digestive tract and glands. Mesoderm, on the other hand, generates heart, kidney, blood, reproductive system, bones, skeletal and smooth muscle, connective tissues, tendons, ligaments, vasculature, dermis and cartilage (Carlson, 2014). Another population in vertebrates, called the "neural crest cells" (NCCs), derive from the ectoderm and gives rise to most structural elements of the face and contributes to innervation, vasculature, melanocytes, bone, cartilage and connective tissue (Douarin and Kalcheim, 1999).

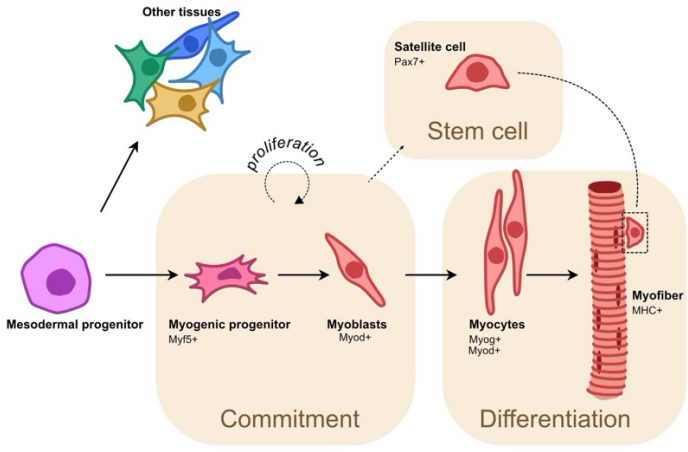



**Figure 1. Common myogenic history.**

Mesodermal cells give rise to myogenic progenitors that will express the myogenic regulatory factors (MRFs), proliferate and commit to myogenesis. A pool of Pax7+ stem cells will be generated and maintained for repair or growth.

## 2. The myogenic regulatory factors

Skeletal muscle formation has long been considered to be a prime example to study cell fate determination and lineage progression. More than 3 decades ago, the stem cell field was marked by the seminal discovery of the transdifferentiation potential of *Myod*. The induction of this single transcription factor could activate the myogenic program in fibroblasts and generate differentiated myofibres (Davis et al., 1987). This work inspired a search for similar conversion potential in other tissues, which was for some time unsuccessful. However, this breakthough laid the conceptual foundation for experiments that eventually led to the Nobel prize discovery of the Yamanaka factors 20 years later, and which allowed the production of induced-pluripotent-stem ("iPS") cells (Takahashi and Yamanaka, 2006).

In vivo, myogenesis is achieved through the timely expression of "myogenic regulatory factors" (MRFs). These genes named *Myod*, *Myf5*, *Mrf4* (*Myf6*), and *Myog* (*myogenin*) are bHLH (basic helix-loop-helix) transcription factors that are thought to be the products of multiple duplication events of *Myf5* during evolution (Atchley et al., 1994; Megeney and Rudnicki, 1995) (**Figure 2**). During mouse embryonic myogenesis, progenitors first activate *Myf5*, followed by *Mrf4* (in most muscles) then *Myod* (**Figure 1**). Individually or combined, these myogenic fate determinants give rise to committed myoblasts. *Myogenin* expression in *Myod*+ myoblasts promotes differentiation and the expression of *Mymk* (*Myomaker*) and *Mymx* (*Myomixer*), 2 molecules located in the cell membrane that allow fusion of myoblasts into elongated multi-nucleated myofibres (Bi et al., 2017; Millay et al., 2013; Zhang et al., 2017). Myofibres are enriched in acto-myosin protein complexes, which upon maturation and further additive fusion of myoblasts constitute the primary units of contraction (Rayment et al., 1993). In the adult, Pax7+ (a gene encoding a paired box transcription factor) cells are muscle stem cells (MuSCs, known as "satellite cells"). During homeostasis in the mouse, these cells reside under the basal lamina of the myofibres, do not express *Myod* and are quiescent (**Figure 1**). Upon injury or damage, disruption of the basal lamina will activate these cells that will enter the cell cycle, proliferate symmetric or asymmetrically to generate new Myod+ myoblasts and regenerate the injured muscle (Dumont et al., 2015; Evano and Tajbakhsh, 2018; Relaix and Zammit, 2012; Zammit et al., 2006).

Despite their common ancestry, the individual MRFs exhibit differential capacities to promote myogenesis due to their different motifs outside the highly conserved bHLH domain

(**Figure 2**) (Fong and Tapscott, 2013; Tapscott, 2005). In vitro studies using domain swapping (Fong and Tapscott, 2013; Tapscott, 2005), and in vivo genetic studies using single and compound mutants provided critical information on the requirement and sufficiency of each MRF in initiating embryonic and fetal myogenesis (Braun et al., 1990; Comai and Tajbakhsh, 2014; Rudnicki et al., 1992; Rudnicki et al., 1993; Sambasivan et al., 2009; Tajbakhsh et al., 1996; Tajbakhsh et al., 1997; Winter et al., 1992; Kassar-Duchossoy et al., 2004). These and other loss-of-function experiments, as well as experiments in avians, underscored the intrinsic heterogeneity of skeletal muscles which deploy combinatorial programs to safeguard myogenesis (**Table 1**, **Figure 7**). Yet, genetic studies in the mouse showed that in the absence of the 3 determination genes *Myf5*, *Mrf4*, and *Myod*, no muscle is formed in the embryo (Kablar et al., 1997; Kassar-Duchossoy et al., 2004; Rudnicki et al., 1993). In this context, *Myog* is not sufficient to initiate myogenesis, suggesting that the C terminal domain of *Myf5* and *Myod* is necessary for activation of downstream targets. In addition, deletion of *Myog* leads to lethality at birth, but it is dispensable for postnatal life (Hasty et al., 1993; Meadows et al., 2008; Venuti et al., 1995). Although embryonic myogenesis seems occur normally in these mutants, fetal myogenesis is impaired and muscle formation is halted before birth (Hasty et al., 1993; Nabeshima et al., 1993; Venuti et al., 1995).

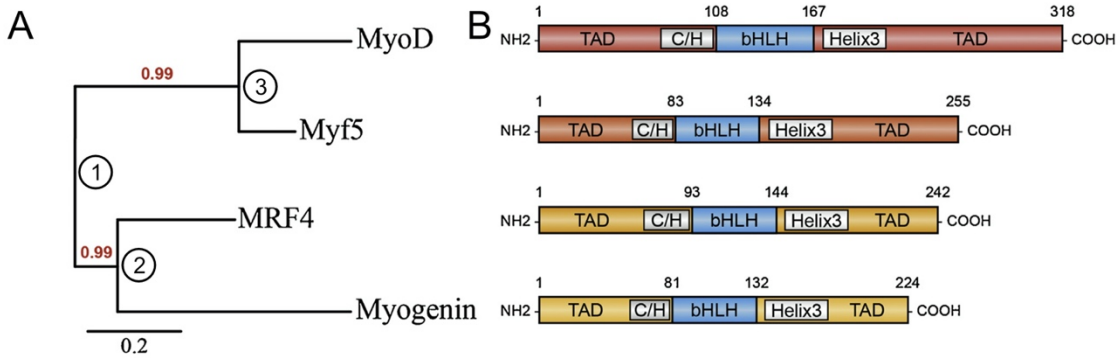



**Figure 2. Ancestry and structure of the MRFs, bHLH transcription factors.**

(A) Phylogeny tree of the MRFs. 3 duplication events were reported to take place during evolution, leading to 4 MRFs (Atchley et al., 1994)
(B) Different structures of the MRFs, suggesting different activation potential. Scheme adapted from (Hernández-Hernández et al., 2017).

# 3. Fine-tuning the function of the myogenic regulatory factors

## A. Cofactors and repressors

To activate transcription, the MRF bHLH proteins heterodimerize with the ubiquitously expressed E proteins (Lassar et al., 1991). This bHLH-E complex will bind E-box (enhancer box) DNA motifs at the level of a promoter or enhancer (**Figure 3A-B**). E-box DNA sequences (CANNTG, where N can be any nucleotide) are located throughout the genome, raising the issue of specificity for muscle genes. In this context, flanking regions of an E-box will influence its specificity and affinity for various MRFs (Yutzey and Konieczny, 1992).

The best characterised cofactors associated with MRFs are members of the *Mef2* family (Black et al., 1998; Potthoff and Olson, 2007). MEF2 proteins (myocyte enhancer factors 2; 4 members: a, b, c and d) bind a MADS domain and they play a central in many differentiation processes by potentiating transcription (Potthoff and Olson, 2007). Other cofactors include RUNX, PPARG and MYB (MacQuarrie et al., 2012). The interaction of *Myod* (and other MRFs) with E proteins can be altered by the expression of the *Id* gene family members that encode bHLH proteins that lack a transactivation domain and that will sequester E proteins and prevent *Myod* from activating E-box containing genes (Jen et al., 1992). Apart from bHLH containing proteins, E-boxes can also bind zinc-finger proteins such as the transcription repressor SNAIL, which recruits histone modifiers HDAC1/2 to inhibit transcription (Soleimani et al., 2012). Interestingly though, SNAIL binds to E-boxes that have a G/C-rich central dinucleotide, which are often found in genes expressed in myotubes, but not in myoblasts genes. *Snail* overexpression blocks differentiation while knocking it down induces precocious differentiation (Soleimani et al., 2012). ZEB1 was shown to have a similar effect by binding G/C-rich domains and associating with CtBP (C-terminal binding proteins, which are transcriptional regulators). Similarly, disrupting *Zeb1* function induces precocious differentiation (Siles et al., 2013). The PBX/MEIS complex (a homeodomain transcription factor complex) was shown to promote MYOD activation of *Myog* by providing a non-canonical second E-box element to *Myog* (Tapscott, 2005). MYOD can also directly bind the BAF60c subunit of the SWI/SNF remodeling complex, consistent with the higher chromatin remodeling capacity of MYOD over MYF5 (Conerly et al., 2016). Interestingly, this complex is found at the promoter region of untranscribed MYOD target genes in myoblasts (Forcales et al., 2011). Upon differentiation cues, BAF60c is phosphorylated by p38 (MAPK14) and the

MYOD/BAF60c complex becomes integrated in a Brg1-based SWI/SNF ATP-dependent chromatin remodeling complex to robustly activate transcription (Forcales et al., 2011).

## B. microRNAs

Regulation can also occur indirectly, through the action of miRNAs. miRNAs have emerged as important post-transcriptional regulators of myogenesis by acting on RNA stability and translation through binding of the 3'UTR region (Xu et al., 2019) (**Figure 3C-D**). For example, miR1, miR206, miR27 and miR486 repress the expression of upstream genes *Pax3* and *Pax7* (paired box transcription factors) in Myod+/Myf5+ cells, to promote their commitment and differentiation. Our group recently showed that also in the adult, miR708 is induced by Notch signaling, thereby maintaining adult muscle stem cells in their niche and preventing their entry into the cell cycle (Baghdadi et al., 2018). It was reported that *Myf5* transcripts can be targeted by miR31 (Crist et al., 2012), preventing precocious activation of myogenesis in myogenic progenitors and in the central nervous system where *Myf5* is transcribed during embryogenesis but no protein is produced (Daubas et al., 2000; Crist et al., 2012; Gayraud-Morel et al., 2012). As indicated above, *Snail* can act as a repressor, and its activity is regulated by miR30a and miR206 which prevent accumulation of SNAIL protein during the onset of differentiation. *Mef2* genes act at the onset of differentiation by promoting (like *Myod*) a miRNA gene cluster: miR1/206/133 (Liu et al., 2007). A feedback circuit operates where miR1 regulates HDAC4 (histone deacetylase, a transcriptional repressor) that inhibits MEF2C, while miR133 targets the mRNA of *Maml1*, a transcriptional coactivator of *Mef2* (Cesana et al., 2011). Additionally, miR133 preRNA also encodes a competing endogenous (ce) long non-coding RNA, named linc-MD1, that binds to and sequesters miR133 but also miR135, which targets the mRNA for *Mef2c* (Cesana et al., 2011). In another context, adult stem cells transcribe *Prdm16* that encodes a key regulator of myogenic to brown fat cell fate, however miR133 targets *Prdm16* mRNA and prevents its function (Yin et al., 2013). It was reported that when miR133 is disrupted, MuSCs give rise to brown adipocytes. The authors proposed a role for that mechanism where cold exposure allows conversion of MuSCs to thermogenic brown fat cells (Yin et al., 2013).

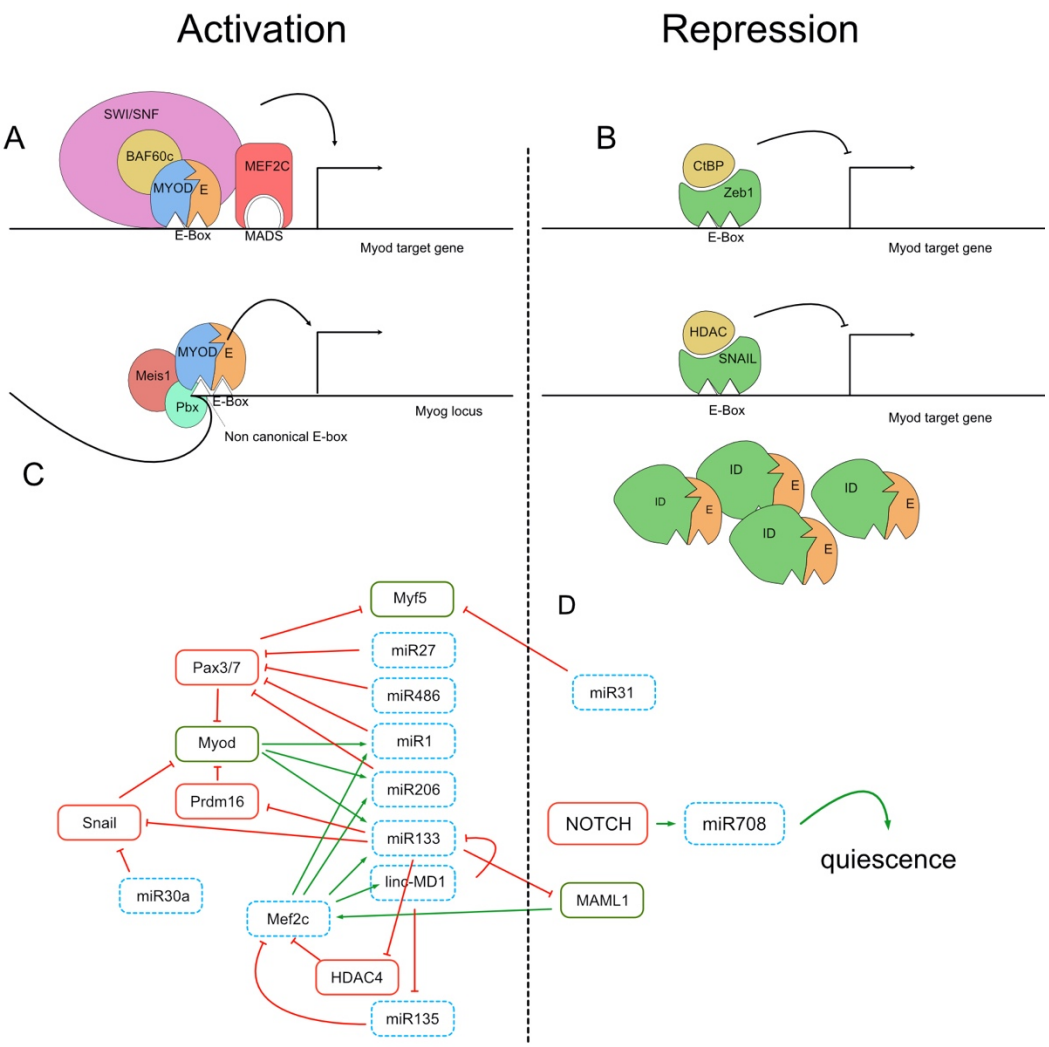



**Figure 3. Activation and repression of MRF activity.**

(A-B) Cofactors like MEF2C and MEIS1/PBX promote MYOD function in activating pro-myogenic genes, while ZEB1 and SNAIL recruit transcription repressors. ID segregates E proteins in order to reduce MYOD binding
(C-D) Numerous micro RNAs act as intermediate molecules to fine tune myogenesis

# 4. Myogenesis: a multistep process

## A. Primary and secondary myogenesis

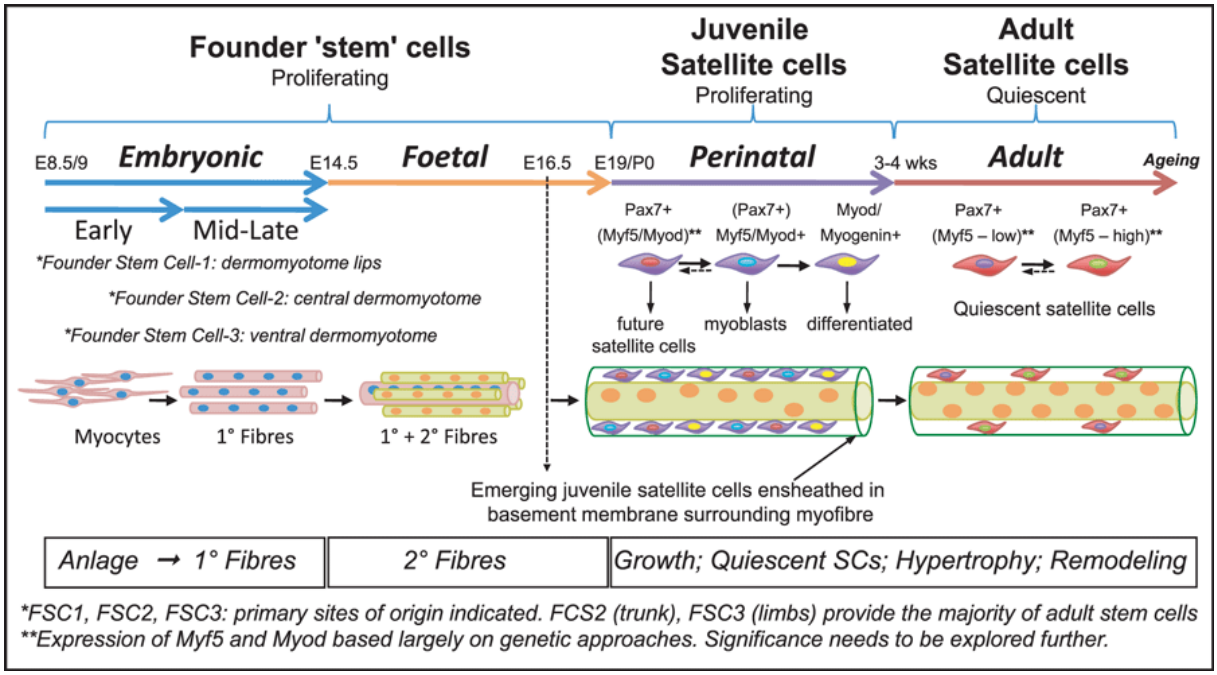



**Figure 4: Myogenesis is a multistep process.**
Myogenic progenitors first form primary fibres first and this is followed by a second wave of secondary fibres which results in fusion of secondary myoblast between themselves, or with the primary fibres. At perinatal stages, growth is supported by proliferating Pax7+ cells which will adopt a quiescent state in the adult (Tajbakhsh, 2009).

Muscle formation has been defined as a multistep process, most studied in limb and trunk muscles, but the principles appear to apply also to most cranial-derived muscles. In the trunk, myogenic progenitors start appearing in the trunk at around E8.5, and primary myogenesis takes place from about E10.5 to E12.5 (Comai and Tajbakhsh, 2014; Duxson et al., 1989). This process involves the fusion of early myoblasts into primary fibres that contain several myonuclei (Kalcheim et al., 1999). This results in the formation of "myotomes" in early myogenesis, and other muscles from mid-embryogenesis in an anteroposterior developmental gradient. From E14.5 to birth, a second wave of progenitors will fuse using primary myofibres as a scaffold to form secondary multinucleated fibres (**Figure 4**) (Comai and Tajbakhsh, 2014; Deries et al., 2010; Duxson et al., 1989). The latter are gradually surrounded by a basal lamina (Duxson et al., 1989). Embryonic and fetal myoblasts also display heterogeneity in the Myosin Heavy Chain (MyHC) isoform that they express, categorized as "fast" and "slow", with

different ATPase activities, contraction speeds and associated metabolism. While embryonic fibres can possess both types (but more generally slow), fetal fibres usually consists of "fast" type (Biressi et al., 2007; Kelly and Rubinstein, 1980; Wigmore and Evans, 2002). Notably, different muscles have different compositions of fast and slow fibre types, which will be developped in **Chapter II**.

After birth, skeletal muscle enters a hypertophic phase where perinatal stem cells continuously divide and contribute to the growing muscles. The majority of this expansion happens within the first 2 weeks after birth (White et al., 2010). Following postnatal growth, a subpopulation is set aside and these cells act as a reservoir of quiescent stem cells lodged in a highly regulated niche, under the basal lamina of the muscle fibres.

The different progenitor states that were identified during development were shown to differ in their expansion capacities in vitro and their morphological aspects. Embryonic myoblasts are elongated cells, and are more prone to differentiation in vitro (and thus proliferate less), forming mononucleated or oligonucleated myofibres. Fetal myoblasts appear as triangular in shape, form larger colonies and fuse into longer multinucleated fibres. Their differentiation is inhibited by growth factors Tgfβ and Bmp4, contrary to their embryonic counterparts (Cossu et al., 2000; Biressi et al., 2007a; Biressi et al., 2007b; Yablonka-Reuveni and Seifert, 1993).

## B. Fetal and perinatal stages

Another common feature of all skeletal muscles is the dramatic growth taking place at fetal and perinatal stages, marked by the high proliferation of Pax7+/Myf5+ cells. Importantly, these stages coinside with the emergence of a reserve pool of muscle stem cells (Messina and Cossu, 2009).

### a) Muscle stem cells self-regulate

Muscle stem cells will shape their niche to a certain extent by depositing extracellular matrix (ECM) components including high levels of collagen VI and fibronectin (Tierney and Sacco, 2016). Tenascin-C, which is known to modulate proliferation and differentiation in other niche contexts, was shown to be expressed only by fetal progenitors and required for fetal MuSC proliferation and differentiation (Tierney et al., 2016).

A major intercellular signaling pathway that is required to maintain MuSCs is mediated by Notch. Notch is expressed by prenatal MuSCs at multiple developmental stages, and it interacts with Delta-like1/2 (Dll1/2) and Jagged1/2 to regulate their maintenance. Disruption of Dll1 or Notch effector RBPJ at fetal stages leads to hypotrophy and the exhaustion of the upstream cells (Schuster-Gossler et al., 2006; Vasyutina et al., 2007). Constitutive Notch signaling on the other hand blocks lineage progression and differentiation, but maintains the founder stem cell population in the absence of differentiated cells (Mourikis et al., 2012). In addition, fetal fibre growth requires the expression of *Nfix*, which is a potential target of Pax7 (Messina et al., 2010).

b) Support from stromal cells

As development proceeds to perinatal growth, stromal mesenchymal cells, which have been most extensively studied in the adult, play important roles. One such cell population identified as PW1+/Sca1+ (a zinc-finger transcription factor and a cell surface protein respectively) interstitial cells (PICs), secrete IGF1 which promotes proliferation and differentiation of MuSCs (Formicola et al., 2014; Pannerec et al., 2013). In *Pax7*-null mice at perinatal stages, the MuSC pool is depleted and PICs increase in number (Mitchell et al., 2010). Other populations of Tcf4+ (Mathew et al., 2011) mesenchymal stromal cells are present in the perinatal niche and they promote differentiation and maturation of muscle fibres (Biferali et al., 2019). Other populations include endothelial cells, which secretome induces proliferation of MuSC (Kostallari et al., 2015). Much of these interactions are carried over to the adult stem cell niche.

# 5. Adult muscle stem cells and their niche

Skeletal muscle is a particularly resilient tissue. Owing to its stem cells, this tissue possesses an outstanding capability to recover from repeated injuries, resulting in multiple rounds of degeneration/regeneration without major long-term consequences (Bentzinger et al., 2013). However, a number of disorders can impair muscle function, such as muscular dystrophies, cachexia (cancer-induced muscle wasting), diabetes, and ageing (Ali and Garcia, 2014). In many pathologies, a chronic cycle of impaired stem cell function, muscle weakening, disuse, inflammation and fibrosis undermines muscle integrity. Therefore, understanding the mechanisms safeguarding satellite cell function has been at the core of the discipline for many years.

## A. Satellite cells: the muscle stem cells

Postnatal and adult satellite cells are marked by the expression of *Pax7*, a paired domain transcription factor. It is thought that *Myod* but not *Myf5* requires Pax7 since *Myf5* can initiate myogenesis in the absence of *Pax7* in adult satellite cells in vitro (Relaix et al., 2006). The function of *Pax7* in adult stem cells has been controversial. A first study showed that even when both *Pax3* (another member of the family with overlapping function in the embryo) and *Pax7* where conditionally deleted in adult MuSCs, these cells were still present and muscle regeneration could take place (Lepper et al., 2009). Conversely, conditional deletion of *Pax7* in satellite cells before P21 leads to defective regeneration, but not after (Lepper et al., 2009). The authors proposed that myogenic cells differentiate prematurely in this context. These results suggested that Pax7 was required within a specific period in the context of combined demands for growth and repair following injury (Seale et al., 2000). Follow-up studies then showed that long-term depletion of *Pax7* in satellite cells resulted in failed regeneration following injury (Günther et al., 2013; Maltzahn et al., 2013; Relaix and Zammit, 2012).

## B. The muscle stem cell niche

The concept of the stem cell niche was first proposed in the hematopoietic system, based on the observation that stem cells require communication with surrounding cells to maintain their stem cell properties (Schofield, 1978). The MuSC niche is home to many molecular and mechanical interactions between the satellite cell and its environment that control its maintenance and activation. During puberty, the expansion of satellite cells and muscle growth will progressively stop, as MuSCs begin entering quiescence from late perinatal stages and homeostasis is established (Tajbakhsh, 2009). In the adult niche, MuSCs are in contact with the basal lamina rich in ECM components on one side and the myofibre on the other (**Figure 5**).

### a) Intrinsic and autocrine control of MuSC quiescence

During homeostasis, satellite cells are mostly quiescent in the majority of muscles. As indicated above, Notch signaling was shown to be a main factor in sustaining this state. Notch activity represses *Myod*, and the downregulation of its canonical target genes (HeyL and Hes1) allows cell cycle entry during activation (Lemos et al., 2015; Ontell and Kozeka, 1984). When

its effector RBPJ is disrupted, MuSCs exit the quiescent state (Bi et al., 2016; Mourikis and Tajbakhsh, 2014).

In addition to Pax7, quiescent satellite cells express Calcr and Odz4 as markers (2 transmembrane proteins). Odz4 deletion induces a reduction in muscle mass/satellite cell pool, and MuSCs have a prolonged proliferation and increased differentiation in vitro (Ishii et al., 2015). Calcr was shown to act on quiescence, and its disruption results in an increase in cell-cycle related genes (Yamaguchi et al., 2015). Our group recently showed that ColV, expressed by satellite cells and under the control of Notch/RBPJ, can bind Calcr and regulate quiescence cell-autonomously (Baghdadi et al., 2018b). Similarly, it was shown that autocrine ANG1/TIE2 signaling in a subset of MuSCs, as well as paracrine signaling, promotes cell cycle exit (Abou-Khalil et al., 2009). MuSC also produce their own ECM components, as well as receptors like Syndecan3/4 (SDC3/4) which bind a number of growth factors like HGF, FGF, TGFB1 and VEGF (Xian et al., 2009). A number of other factors have been shown to regulate MuSC quiescence and activation suggesting that this is a tightly regulated process (Evano and Tajbakhsh, 2018).

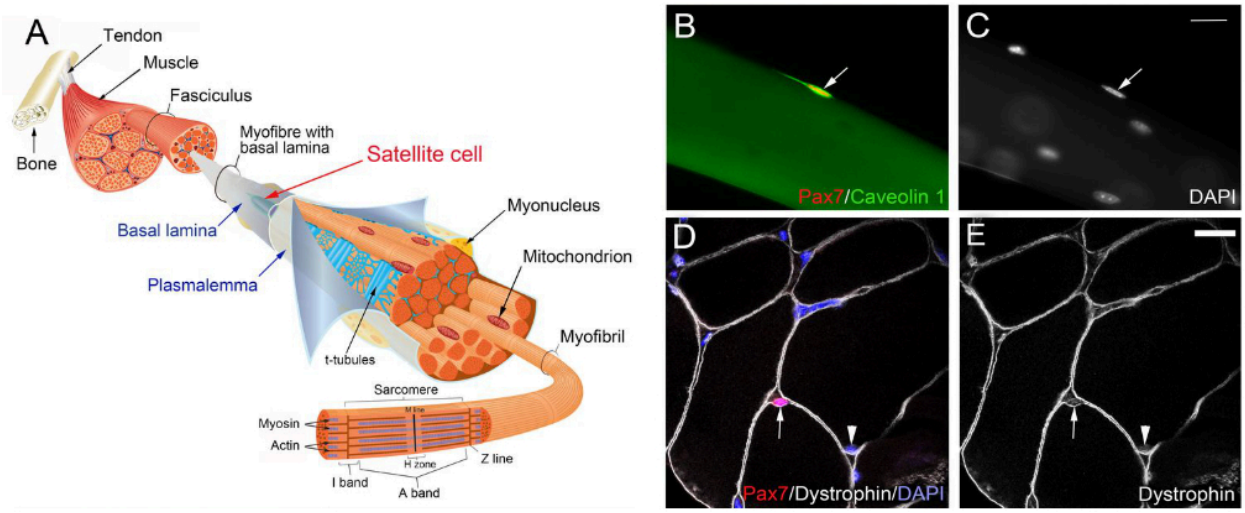



**Figure 5. The adult muscle stem cell: the satellite cell.**

Satellite cells are found at the periphery of adult myofibre, under the basal lamina, and are marked by *Pax7* (Relaix and Zammit, 2012; Tajbakhsh, 2009).
(A) Schematic view a myofibre in the context of the entire muscle.
(B-C) Pax7 and Caveolin 1 immunostaining highlighting a satellite cell.
(D-E) Pax7/Dystrophin immunostaining of a cross-sectioned muscle highlighting the peripheral location of the satellite cells, under the basal lamina.

b) Extrinsic and paracrine control of MuSC quiescence

The ECM is comprised of many structural proteins that support maintenance of the MuSC quiescent state by providing a physical barrier and sequestering growth factors (Gohring et al., 1998). Smooth muscles and pericytes, produce IGF1 and angiopoietin 1 (ANG1 signaling through TIE2) which promotes cell cycle exit and transition to quiescence of MuSCs (Abou-Khalil et al., 2009; Kostallari et al., 2015).

Extrinsic signals coming from fibroblast populations, immune cells, endothelial cells, pericytes and the myofibre can all impact on MuSC fate. For example, Oncostatin M produced by the myofibre can promote quiescence of the associated MuSCs (Sampath et al., 2018). During injury, immune and inflammatory response can act directly on MuSC. Neutrophils and macrophages secrete TNFα which results in a lower expression of *Pax7* and *Notch* thereby promoting their activation (Acharyya et al., 2010; Palacios et al., 2010). A non-myogenic interstitial population, identified based on the expression of PDGFRα and their fibro-adipogenic potential in vitro, have been shown to contribute to fibrosis if not sufficiently cleared (Joe et al., 2010; Uezumi et al., 2010). These cells, called FAPs (fibroadipogenic progenitors), secrete IL-6 among other factors, and promote myogenic differentiation (Biferali et al., 2019; Joe et al., 2010). In addition, it was recently shown that WISP1 expression by FAPs regulates myogenic expansion and asymmetric cell division, and this activity is impaired during ageing (Lukjanenko et al., 2019). Finally, Follistatin and IL-10 secreted by FAPs were also suggested to promote myogenic differentiation (Lemos et al., 2012; Mozzetta et al., 2013).

c) Extrinsic plasticity: myogenic interstitial cells

Interstitial cells play a critical role in the maintenance and activation of MuSCs, and skeletal muscle regeneration. In addition, some of these cells present an intriguing myogenic potential. PW1+ cells were shown to contribute to myofibres during regeneration (Mitchell et al., 2010). Similarly, Pax7-/Twist2+ interstitial cells were shown to give rise to fast myofibres in normal and regenerative conditions (Liu et al., 2017; Joe et al., 2010; Uezumi et al., 2010). Notably, recent scRNAseq experiments have provided a clearer view on the local environment of MuSCs in the adult muscles, including cells with myogenic potential, expressing smooth muscle and mesenchymal markers (Giordani et al., 2019; Rubenstein et al., 2020). Of interest, these stromal cells with myogenic potential cannot regenerate the muscle following injury, at

least over a short-term period, when Pax7+ cells are genetically ablated with diphtheria toxin (Lepper et al., 2011; Murphy et al., 2011; Sambasivan et al., 2011).

Taken together, from perinatal stages to adulthood, the MuSC niche is a highly regulated environment, where MuSC receive signaling and mechanical cues from various sources, while also directing their own fate (**Figure 6**).

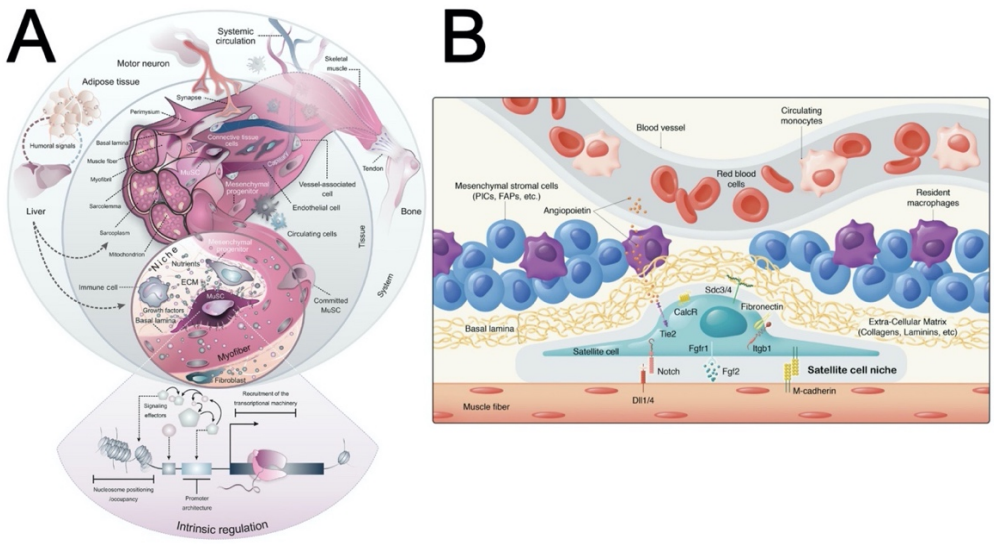



**Figure 6. The muscle stem cell niche.**

(A) An integrated schematic view of the MuSC in the context of extrinsic (systemic and local), and intrinsic cues. (ECM: Extracellular Matrix, MuSC: Muscle Stem Cell) (Mashinchian et al., 2018).
(B) Scheme the signaling networks between the MuSCs and their local environment.(Evano and Tajbakhsh, 2018).

# Chapter II

# An intriguing heterogeneity

## 1. Myopathies and regenerative capacities

Given that contraction is the principal function of all skeletal muscles, it would be reasonable to assume that identical structural proteins would be found in all muscle groups. Yet, skeletal muscles display a wide range of differences in properties. First, skeletal muscles in different anatomical locations are more or less prone to degenerative diseases like dystrophies (**Figure 7).** These disorders are characterised by a progressive muscle weakness that affects the limb, face, body, heart, respiratory and swallowing muscles to variable degrees (Mercuri and Muntoni, 2013). Different classes of myopathies have been reported depending on the muscles affected and the genetic defect underlying the individual forms of the disease. In the most prevalent cases, the disorder is associated with a genetic mutation that affects an extracellular, cytoskeletal, membranous or enzymatic protein. Most of these structural elements are ubiquitous to skeletal muscles, which makes the selective aspect of these diseases puzzling. The origins of these discrepancies are largely unknown, but it is possible that distinct muscles are able to compensate for the loss of certain proteins by mobilising other mechanisms to maintain function.

### A. Limb

Therapeutic studies involving muscle stem cells have for the most part used limb as a source of donor cells. This is due to the easily accessible nature of the limb and its size, as collection of tissue samples from individuals would be relatively innocuous. Notably, limb myogenic cells expand to a limited extent and for relatively short periods in vitro, presumably due to the lack of defined culture conditions promoting their expansion, and a limited stem cell capacity in vitro (Konieczny et al., 2013; Miller et al., 1997; Tremblay et al., 1993; Wilschut et al., 2012).

## B. Extraocular muscles

The extraocular muscles (EOM) are a set of specialized muscles that control the movements of the eye. These muscles are uniquely spared in Duchenne muscular dystrophy and dystrophic animal models, and have mostly positive outcomes in numerous diseases including ageing (Kaminski et al., 2002; Man et al., 2005; Schoser and Pongratz, 2006; Valdez et al., 2012). Moreover, they possess a super-fast myosin-heavy-chain isoform, *Myh13*, as well as unique distinct metabolic and cytoskeletal transcriptomic signatures (Cheng et al., 2004; Spencer and Porter, 2006). In culture conditions, EOM stem cells proliferate at a higher rate than their diaphragm and limb counterparts (Stuelsatz et al., 2015). This high proliferative capacity in vitro is also maintained in aged mice. The same study also measured proliferation capacities of dystrophic satellite cells, again showing an in vitro potential that is maintained compared to limb and diaphragm. Also, when grafted into the hindlimb of an immune-compromised mouse, EOM stem cells performed better than those from the limb (Stuelsatz et al., 2015). EOM MuSCs continuously contribute myonuclei to myofibres in the adult in the absence of injury (Keefe et al., 2015). In addition, EOM cells possess a higher portion of PW1+ interstitial cells compared to limb, and this has been suggested to provide resistance cues within the satellite cell niche (Formicola et al., 2014; Stuelsatz et al., 2014)

## C. Masseter

Masseter muscles control movements of the jaw and possess a lower regenerative potential than *Tibialis anterior* (TA; lower limb) muscles, with a lower number of satellite cells during regeneration (Pavlath et al., 1998). The number of satellite cells per masseter myofibre increases with age, but their proliferative capacity declines (Ono et al., 2009).

## D. Pharyngeal muscles

Pharyngeal muscles are critical for proper feeding, speech and breathing. These muscles also receive sustained contribution of myonuclei from their stem cell pool during homeostasis (Randolph et al., 2015). Ablation of satellite cells showed that they are required for maintaining myofibre size and nuclear count of pharyngeal muscles (Kim et al., 2020; Randolph et al., 2015). In both an ageing and an oculopharyngeal muscular dystrophy mouse model, muscles of the pharynx were affected to a greater extent than other muscles, leading to dysphagia (Randolph et al., 2014).

### E. Diaphragm

Diaphragm MuSCs have been shown to possess particular properties compared with other somitic muscles. The expression of the upstream gene *Pax3* is maintained in satellite cells postnatally (Day et al., 2007; Relaix et al., 2006; Stuelsatz et al., 2012). They also possess a higher proliferative potential and reduced differentiation capacity compared to limb and craniofacial muscles in vitro (Ippolito et al., 2012). Additionally, diaphragm muscle stem cells continuously give rise to muscle in the adult, while myofibre size remains unchanged (Keefe et al., 2015).

### F. Laryngeal muscles

A similar observation was made in rabbit laryngeal muscles (Goding et al., 2005). Also, elevated Erk phosphorylation was detected in laryngeal muscle cultures compared with limb (Walz et al., 2008). Another study found myogenic cells expressing *Myod* and *Myog*, still present in thyroarytenoid muscles 2 years post-denervation in humans. This points to a unique prolonged regenerative potential of laryngeal muscles (Donghui et al., 2009).

## 2. Distinct genetic requirements of skeletal muscle progenitors

Another level of muscle heterogeneity was noted in the developing embryo regarding the requirements for different MRFs. Over the last 3 decades, multiple combinatorial loss-of-function studies were done to better characterise the overlapping and unique functions of the MRFs for embryonic and fetal myogenesis. Most studies compared trunk muscles (referred to as epaxial and hypaxial muscles, see **Chapter IV**), limb, Pharyngeal arch 1/2 muscles (corresponding to masticatory and facial muscles respectively, see **Chapter III**) and extraocular muscles. Notably, most of these studies examined at few selected muscles. For example, caudal head muscles, derived from the caudal arches during development, were not reported in any of these studies, presumably because they are not visible in whole-mount stainings.

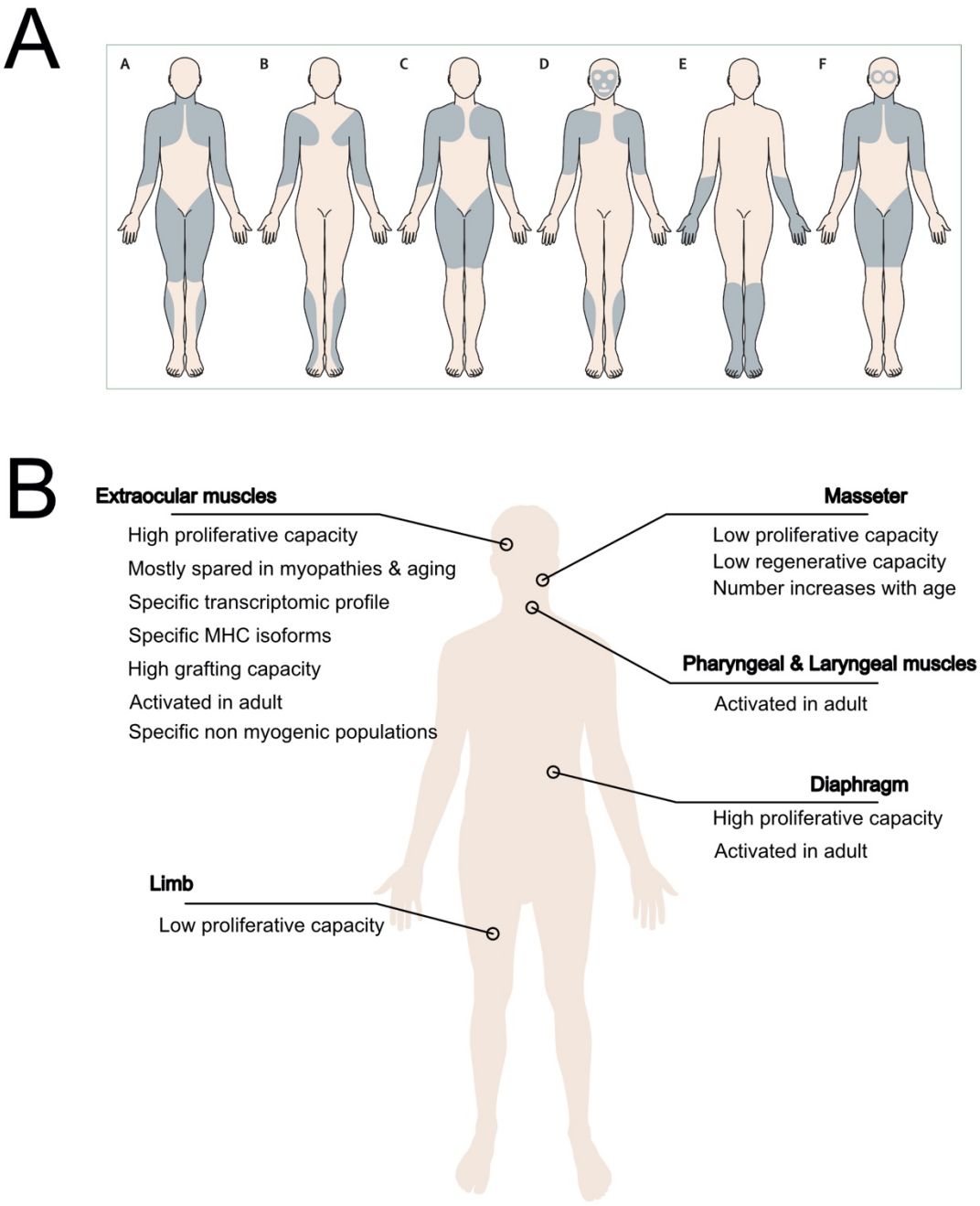



**Figure 7. Skeletal muscle heterogeneity**

(A) Defective muscles in various myopathies. A: Duchenne and Becker muscular dystrophy. B: Emery-Dreifuss muscular dystrophy. C: Limb girdle muscular dystrophy. D: Facioscapulohumeral muscular dystrophy. E: Distal muscular dystrophy. F: Oculopharyngeal muscular dystrophy.
(B) Muscle-specific features of satellite cells.

Importantly, initial studies using single mutants of *Myf5* or *Mrf4* used mouse models where both *Myf5* and *Mrf4* expression were affected (Kassar-Duchossoy et al., 2004; Yoon et al., 1997). These genes are linked on chromosome 10 in mouse, and possess juxtaposed regulatory sequences (Carvajal et al., 2001). Specific mutant alleles have been generated since then (Kassar-Duchossoy et al., 2004). Multiple common and muscle-specific requirements have emerged from these studies, which are summarized in **Table 1** and **Figure 7**.

Briefly, myogenic progenitors require the expression of either *Myf5*, *Mrf4* or *Myod* for embryonic myogenesis (Kablar et al., 1997; Rudnicki et al., 1993), and *Myog* for foetal myogenesis (Hasty et al., 1993; Nabeshima et al., 1993; Venuti et al., 1995). Notably, the EOM and some deep epaxial muscles require the expression of both *Myf5* and *Mrf4*; in their absence, these muscles do not form (Kassar-Duchossoy et al., 2004; Sambasivan et al., 2009). Interestingly, other cranial muscles such as the masticatory muscles remained unaffected in those studies, pointing to a distinct regulatory hierarchy among head muscles (Sambasivan et al., 2009) (see **Chapter III**). Interestingly, *Myod* deletion leads to delayed myogenesis in anterior arch and limb muscles, but it is compensated in trunk and EOM by *Myf5* and *Mrf4* (Kablar et al., 1997; Kassar-Duchossoy et al., 2004; Sambasivan et al., 2009). As pharyngeal and laryngeal muscles were not specifically explored in these studies, we examined their phenotype in *Myf5*;*Mrf4* double mutant embryos, and found that muscle integrity was conserved (**Annex 1A-B**).

It is therefore apparent that skeletal muscles are heterogeneous with regard to initiation of myogenesis, activation of MRFs, response to myopathies, and regenerative/proliferative potential. The molecular mechanisms underlying these differences remain largely undefined. Both extrinsic and intrinsic factors are susceptible to affect cellular behaviours including proliferation, differentiation, fusion, migration, and cell death. These differences in properties and function may be deeply rooted in prenatal development, as the emergence of skeletal muscle masses also refects diversity.

| | Trunk | Limb | Pharyngeal arch 1/2 | Caudal arches | Extraocular |
|---|---|---|---|---|---|
| **MyoG null** (1) (2) (3) (Lethal at birth) | Normal embryonic myogenesis<br>No fetal myogenesis<br>(MyoG expressed) | Normal embryonic myogenesis<br>No fetal myogenesis<br>(MyoG expressed) | Normal embryonic myogenesis<br>No fetal myogenesis<br>(MyoG expressed) | ? | ? |
| **MyoD null** (4) (5) (7) (Lethality Background dependent) | Normal | Delayed | Delayed | ? | Normal |
| **Myf5 null** (6) (7) (Viable) | Partially normal<br>(Some epaxial muscles absent) | Normal | Normal | ? | Normal |
| **Mrf4 null** (8) (Viable) | Normal | Normal | Normal | ? | ? |
| **Myf5: Mrf4 null** (6) (7) (Lethal at birth) | Delayed<br>(Some epaxial muscles absent) | Normal | Normal | Normal (13) | Absent<br>(Apoptotic progenitors) |
| **MyoD : Mrf4 null** (9) (Lethal at birth) | Normal embryonic myogenesis<br>No fetal myogenesis<br>(MyoG expressed) | Delayed embryonic myogenesis<br>No fetal myogenesis<br>(MyoG expressed) | ? | ? | ? |
| **MyoD : Myf5 null** (6) (10) (12) (Lethal at birth) | Normal embryonic myogenesis<br>No fetal myogenesis<br>(MyoG expressed) | Absent<br>(Residual fibers) | Absent<br>(Residual fibers) | ? | Absent |
| **MyoD : Myf5 : Mrf4 null** (5) (11) (Lethal at birth) | Absent<br>(No MyoG expression) | Absent<br>(No MyoG expression) | Absent<br>(No MyoG expression) | ? | ? |

1: Hasty et al., Nature, 1993
2: Nabeshima et al., Nature, 1993
3: Venuti et al., The Journal of Cell Biology, 1995
4: Rudnicki et al., Cell, 1992
5: Kablar et al., Development, 1997
6: Kassar-Duchossoy et al., Nature, 2004
7: Sambasivan et al., Developmental Cell, 2009
8: Zhang et al., Genes and development, 1995
9: Rawls et al., Development, 1998
10: Kassar-Duchossoy et al., Genes and Development, 2005
11: Rudnicki et al., Cell, 1993
12: Sambasivan, unpublished data
13: Grimaldi, unpublished data, 2017

Severity

**Table 1. Requirement for MRFs in embryonic and fetal myogenesis.**

Different compound mutants of the MRFs have been made to identify their distinct properties. In doing so, a first level of diversity was revealed.

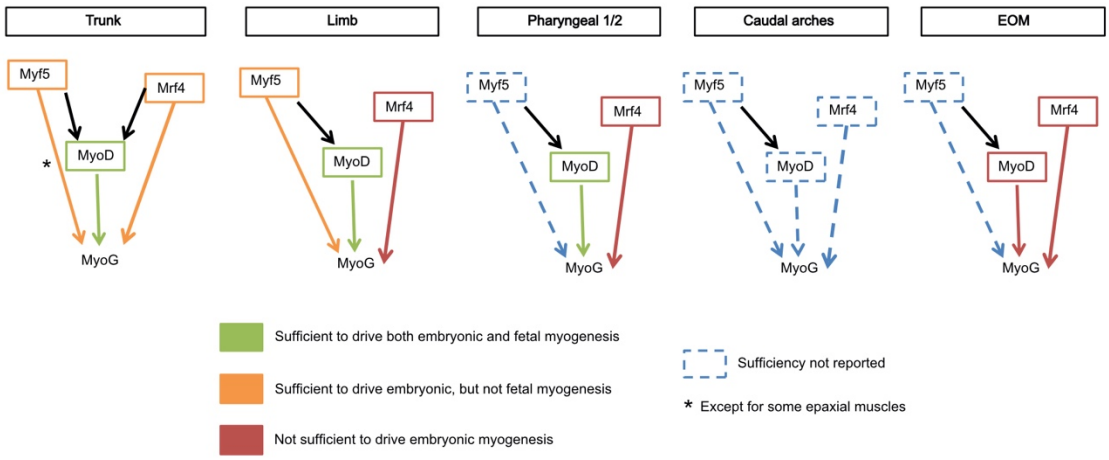



**Figure 8. Sufficiency of MRFs.**

Specific requirement and sufficiency of MRFs in different muscles of the body point at a muscle-specific intrinsic regulatory hierarchy (Buckingham and Rigby, 2014).

# Chapter III

# Intrinsic cues to diversity

For more than 30 years, a number of studies have focused on the morphogenic, genetic and molecular components at work in the establishment of various skeletal muscles, and revealed an unsuspected molecular diversity. It appears clear now that intricate and sometimes non-overlapping gene regulatory networks are required for the activation of myogenesis in distinct parts of the body. Multiple layers of heterogeneity are present within large muscle groups and it is tempting to consider these as multi-dimensional modules, a concept often suggested in an evolutionary context (Espinosa-Soto, 2018; Martik et al., 2019).

## 1. Trunk and limb myogenesis

### A. Somites are the source of trunk and limb skeletal muscles

Trunk, limb, and some neck and head muscles derive from segmented paraxial mesodermal structures known as somites. They are transient epithelial segments on the dorsal side of the embryo, flanking the developing neural tube. Sequentially, each newly formed somite transitions from a ball-like shape and progressively undergoes and epithelial-to-mesenchymal transition (EMT) as cells commit to various lineages (**Figure 8**). The most dorsal part is called the dermomyotome and remains epithelial for an extended period until about E12 in mouse (Comai and Tajbakhsh, 2014; Kalcheim et al., 1999). This region will give rise to skeletal muscle, dorsal dermis, smooth muscle, endothelium (Eichmann et al., 1997; Huber et al., 2004; Yvernogeau et al., 2011), and brown fat (Shapira et al., 2017). Ventrally to the dermomyotome lies the syndetome, giving rise to tendons (Brent et al., 2003) and the sclerotome, giving rise to cartilage and bone of the vertebrae and ribs (Brent et al., 2003). Under the control of signals from surrounding tissues, myogenic progenitors emerge from the dermomyotome and undergo a progressive EMT to populate the muscle forming regions of the embryo (trunk, limbs, tongue and some neck muscles) (Tosney et al., 1994). Trunk muscles are divided into epaxial and hypaxial muscles (Shih et al., 2008; Wotton et al., 2014).

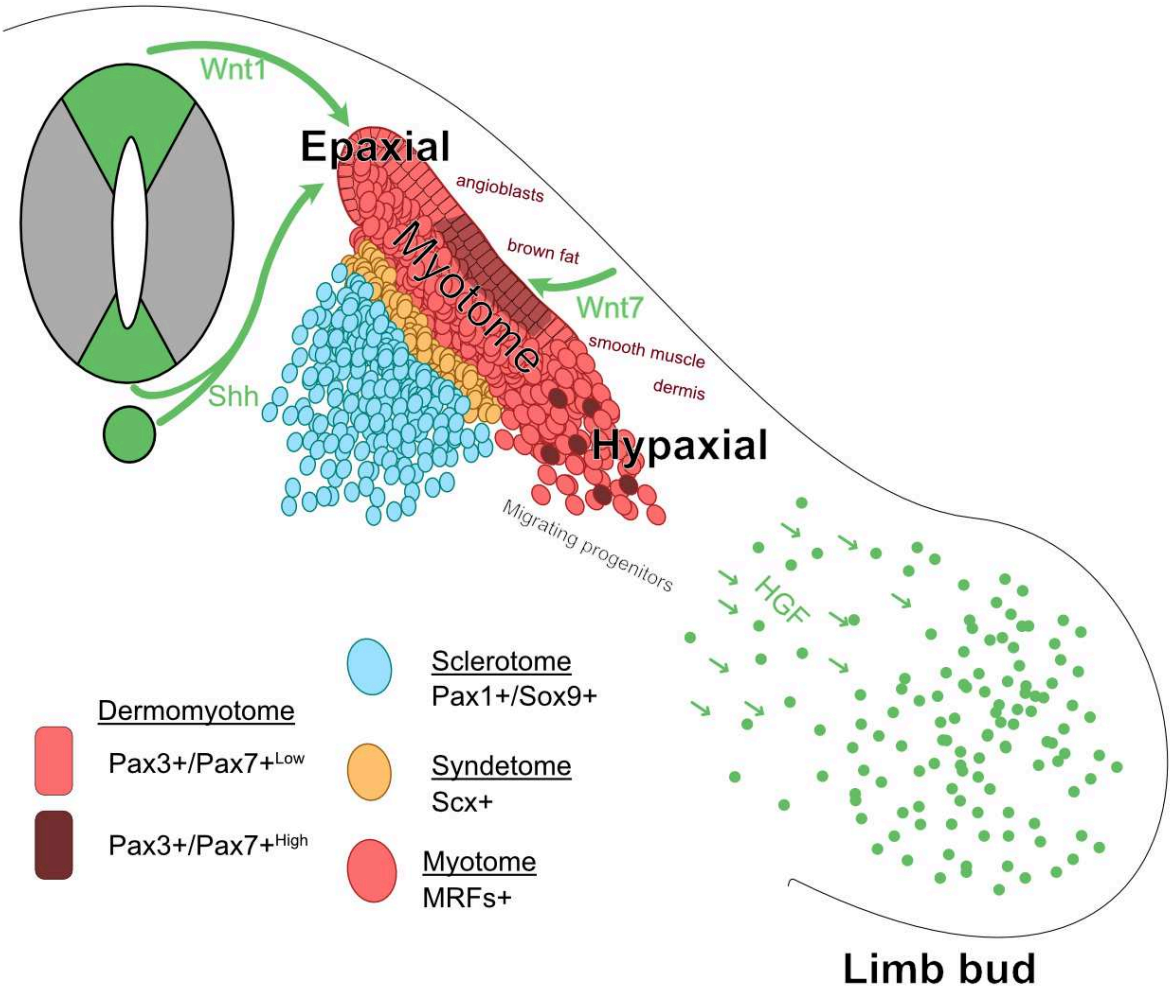



**Figure 9. Somitic myogenesis at limb level.**

Pax3+/Pax7+ progenitors delaminate from the dermomyotome to form the myotome. Epaxial muscles arise from the dorsomedial lip of the dermomyotome while hypaxial muscle arise from the ventrolateral lip. Limb progenitors migrate from the hypaxial region to the limb bud following an HGF gradient coming from lateral plate mesoderm. Epaxial progenitors receive signals from the notochord/floor plate and the dorsal neural tube to activate the expression of the Myf5. Cartilage and bone arise from the sclerotome as Pax1+/Sox9+ progenitors and Scleraxis+ tendon progenitors arise from the syndetome.

Epaxial muscles refer to the muscles of the back and they are innervated by the dorsal branch of the spinal nerves (Fetcho, 1987). Hypaxial muscles are located ventrolaterally and comprise the major muscle mass in amniotes, including muscles of the tongue, diaphragm, abdomen, and limb (Wotton et al., 2014). The trunk hypaxial muscles are innervated by the ventral branch of the spinal nerves (Dietrich et al., 1998; Fetcho, 1987).

Subdivisions within the somite are observable from the earliest stages where epaxial muscle progenitors that emerge require Shh from the notochord and Wnt1/Wnt3a from the neural tube to activate *Myf5* in the dorsal dermomyotome (Summerbell et al., 2000; Teboul et al., 2002). In contrast, hypaxial precursors rely on Wnt7a signals emanating from the dorsal ectoderm that will activate *Myf5* and *Mrf4* which in turn induce *Myod* expression and myogenic commitment (Cossu et al., 2000; Kassar-Duchossoy et al., 2004). This appears consistent with *Myf5 (also* affecting *Mrf4)* and *Myod* single mutant phenotypes, where epaxial and hypaxial muscles are affected selectively, although other MRFs are able to compensate this effect and promote an almost normal skeletal muscle development in most regions (Kablar et al., 1997) (**Table 1**, **Figure 7**). To that series of molecular events is added a temporal control, since *Myf5* expression is first detected in the epaxial domain, then it appears on the opposite hypaxial side of somites (most prominently at the interlimb level), almost concomitant with the expression of *Mrf4*. *Mrf4* expression is subsequently restricted to differentiated cells from mid-embryogenesis. *Myod* expression follows that of *Myf5* and *Mrf4*, but first in the hypaxial domain, and subsequentially in the epaxial region (Kassar-Duchossoy et al., 2004). Regulation of the *Myf5/Mrf4* locus is complex, where the promoters of these genes are separated by about 9 kb. Multiple enhancer sequences upstream of the *Myf5/Mrf4* locus have been identified (Chang et al., 2004). Studies with transgenic mice coupling these enhancers with heterologous promoters showed a variety of expression patterns in the dermomyotome (Hadchouel et al., 2003). Notably, it was shown that *Myf5* possesses a large regulatory sequence 100 kb upstream of its promoter, allowing multiple levels of spatiotemporal control (Moncaut et al., 2013). This level of regulation is adjusted intricately by transcription balancing sequences (called TRABS) that act as cryptic promoters and finely control the expression of Myf5 and Mrf4 specifically (Carvajal et al., 2008). The epaxial activation of *Myf5* by Shh and Wnts is performed through binding of GLI1 and TCF transcriptional effectors on the "early epaxial element" (EEE) (Summerbell et al., 2000; Teboul et al., 2002). ZIC1 and ZIC2 (zinc finger transcription factors) function as important GLI cofactors is this activation (Pan et al., 2011).

Myogenic progenitors delaminate first from the dorsal then ventral dermomyotome to form the primary myotomes in the trunk. Hypaxial progenitors delaminate from the ventro-lateral lip of the dermomyotome and migrate to more distal locations (e.g. body wall, limbs, diaphragm, tongue) where they will give rise to differentiated muscles. As the dermomyotome continues to release progenitors from the dorsal and ventral edges while remaining epithelial, the central dermomyotome undergoes a full EMT from about E11, to generate a reserve pool

of progenitors that assure continued muscle growth in the trunk (Gros et al., 2005; Kassar-Duchossoy et al., 2005; Relaix et al., 2005). It has been suggested that the balance between differentiating and proliferating precursors is undertaken by NOTCH/DELTA lateral inhibition, as shown in *Drosophila*, consistent with similar mechanisms in the adult stem cell niche where Notch represses myogenic differentiation (Baghdadi et al., 2018a; Baghdadi et al., 2018b; Baylies et al., 1998; Mourikis et al., 2012; Vasyutina et al., 2007). En1 (Engrailed1) and Sim1 (an homeobox and a bHLH transcription factors) have been identified as specific markers of epaxial and hypaxial regions of the dermomyotome respectively (Cheng et al., 2004b).

It is interesting to comment here that epaxial muscles initiate and terminate their formation within the dorsal side of the embryo, while hypaxial muscles displace to the ventral region to reach their final anatomical location. This transition involves either active migration through stroma or passive displacement accompanying morphogenic remodeling (Tajbakhsh, 2009).

## B. Initiation of limb muscles

Limb muscle progenitors appear as migratory Pax3+ precursors that delaminate from the ventrolateral lip of dermomyotomes opposing the forming limb bud and migrate through the adjacent lateral plate mesoderm (LPM) (Deries and Thorsteinsdóttir, 2016).This process requires a set of genes and signals to establish migratory competence and to maintain cells undifferentiated until migration is complete. One of the main genes in that context, is the pair *cMet/Hgf*. *cMet* encodes a tyrosine kinase receptor that directs migration towards a physical gradient of released HGH, its ligand. In the context of the limb, myogenic progenitors require the expression of *cMet* cell-autonomously and the expression of Hgf from the LPM to delaminate and migrate (Bladt et al., 1995; Dietrich et al., 1999). *Lbx1* is also expressed in premigratory hypaxial cells and it is necessary for migration of lateral but not ventral progenitors (Gross et al., 2000). Lbx1 activates the expression of *Cxcr4* (encoding SDF1 receptor) (Vasyutina et al., 2005). Loss of function of *Lbx1* leads to the absence of lateral forelimb and hindlimb muscles. In addition, myogenic progenitors at the level of the limb are misplaced into the adjacent somites (Schäfer and Braun, 1999). Pax3 promotes the expression of *Lbx1* and *cMet*, providing migratory competence to somitic progenitors. In *Pax3* mutant embryos, *cMet* expression is abrogated and limb muscles do not delaminate to migrate away from the dermomyotome (Dietrich et al., 1999; Relaix et al., 2004). Although necessary for

migration, *Lbx1* is not required for delamination from the dermomyotome (Schäfer and Braun, 1999). *Lbx1* is a potential direct target of Pax3, but its expression is limited to migratory cells, and *Hox* genes (homeobox genes involved in axial segment identity) contribute to that axial specificity (Alvares et al., 2003).

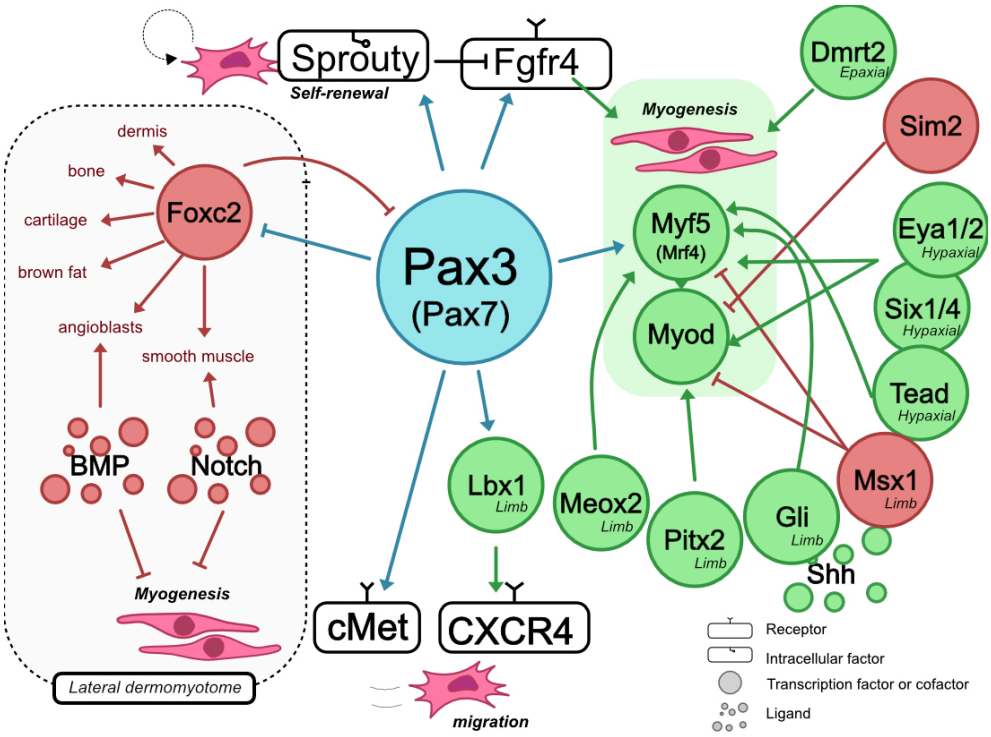



**Figure 10. Regulatory network of trunk and limb myogenesis.**
*Pax3/7* regulate myogenic commitment, self-renewal and migration within the dermomyotome. *Foxc2* and *Pax3* reciprocally repress each other to produce myogenic and non-myogenic cells.

## C. Gene regulatory networks

Muscle development originating from somites is regulated by 2 paired domain transcription factors, *Pax3* and *Pax7* (Relaix et al., 2004; Relaix et al., 2006). Unlike the MRFs, *Pax3* and *Pax7* are not muscle specific, but expressed in other tissues including neurectoderm, brain, neural tube and neural crest (Mansouri and Gruss, 1998). *Pax3* expression is found throughout the dermomyotomes, in particular in myogenic progenitors that have not yet activated the myogenic program, then it is rapidly downregulated following the expression of *Myf5* (Groves et al., 2005). *Pax3* deletion leads to the absence of limb muscles, and later apoptosis of hypaxial progenitors (Auerbach, 1954; Bober et al., 1994; Franz, 1989). *Pax3* also

acts on the balance of progenitor/differentiating cells by activating *Fgfr4* to promote differentiation and *Sprouty* to maintain an undifferentiated myogenic pool (Lagha et al., 2008). Therefore, *Pax3* has a dual role in regulating trunk myogenesis.

Pax3 also activates *Myf5* in epaxial muscles by promoting *Dmrt2*, that binds the EEE (Sato et al., 2010). Pax3 also directly binds the regulatory elements of the limb and the hypaxial somites of *Myf5* to initiate myogenesis in these regions. These 2 domains are also subject to extrinsic cues: Tead activity (Hippo pathway) in the hypaxial and Gli (Shh) in the limb (Gustafsson et al., 2002; Ribas et al., 2011).

*Myod* regulation is carried out in part by an enhancer at -20 kb (CE) which does not appear to be a target of Pax3, but of Pax7 (Hu et al., 2008; Tapscott, 2005). *Pax3* expression overlaps with that of *Pax7* in the central dermomyotome and in *Pax3;Pax7* double mutants, muscle development is arrested from mid-embryogenesis in the trunk after initial formation of the myotome (Relaix et al., 2005). As development proceeds, Pax7 replaces Pax3 in the majority of myogenic progenitors, but some limb and trunk muscles and the diaphragm continue to express *Pax3* (Montarras et al., 2013). However, this expression cannot compensate for the lack of Pax7 in *Pax7* knock out mice (Soleimani et al., 2012b).

Another transcription factor, Foxc2 (forkhead-box), is also expressed at this time in the somite and appears to have an antagonistic effect with Pax3 (Lagha et al., 2009). Foxc2 represses *Pax3* and favours non-myogenic lineages while Pax3 reciprocally represses Foxc2 to promote myogenesis. Foxc2 was found to increase the vascular/smooth muscle potential of dermomyotomal cells (Lagha et al., 2009). In chick, NOTCH and BMP signaling also promote endothelial/smooth muscle cell fate over myogenic in the dermomyotome (Ben-Yair and Kalcheim, 2008). In mouse embryogenesis, clonal relationship was found between *Pax3*-derived skeletal muscle and smooth muscle of the aorta, indicative of the multipotent nature of Pax3+ cells in the dermomyotome (Esner et al., 2006).

*Six* gene family members are homeodomain transcription factors that act in conjunction with co-factors Eya1 and Dach throughout myogenesis (Grifone et al., 2007; Maire et al., 2020; Wu et al., 2014). Specifically, Eya1 inhibits Dach, a corepressor of the *Six* complex. In addition, Eya1 helps to recruit CBP (CREB-binding protein, a transcription activator) to promote transcription by Six (Maire et al., 2020). Ectopic activation of *Eya* and *Six* in the chick promotes the expression of *Pax3* and the MRFs. During embryonic myogenesis, *Six1*, *Six4*,

*Eya1* and *Eya2* are expressed in the dermomyotome and Pax3+ progenitors (Grifone et al., 2007; Wurmser et al., 2020). Their expression is then maintained in differentiated progenitors. *Eya1* and 2 are most specifically expressed in the epaxial and hypaxial regions of the anterior and middle region, but not the posterior somites. In *Six1/4* and *Eya1/2* double mutants, all hypaxial muscles are absent and epaxial muscle are less affected (Grifone et al., 2005; Grifone et al., 2007). Here, *Pax3* expression is lost in the hypaxial region, and cells undergo apoptosis or are found mislocated, and change fates. *Pax3* is thought to be normally upregulated by *Six* through a hypaxial enhancer (Brown et al., 2005). Furthermore, Six acts on regulatory sequences of the MRFs, notably *Mrf4* and *Myf5*, on a hypaxial specific regulatory element (Giordani et al., 2007). Six also binds the CE enhancer of *Myod* at -20 kb as well as a distal enhancer at -6 kb (DRR, distal regulatory region) (Relaix et al., 2013). In addition, *Six* controls the expression of *Myog* through proximal regulatory elements (Spitz et al., 1998). *Six* has not been shown to activate the EEE region of the *Myf5* locus, suggesting that its role in epaxial development might be less prominent, as observed *Six1/4* mutants. Six1/4 also activate the expression of fast fibre genes, and Sox6, which represses the expression of slow fibre type genes (Niro et al., 2010). In this context, it was also shown that Six1 and Eya1 can convert slow and fast fibre types when their expression is artificially increased. Notably, Six1 also plays a role in adult stem cells, were it regulates their self-renewing potential (Grand et al., 2012).

Meox1 and Meox2 are transcription factors expressed in somites. Meox2 is required for limb muscle formation as an activator of myogenesis, particularly in the forelimb (Mankoo et al., 1999). In *Meox2* mutants, *Pax3* is downregulated and *Myf5* expression is delayed. As for repressors of myogenesis, *Msx1* and *Sim2* are expressed in the migratory limb progenitors, and they repress the transcription of *Myod* by recruiting the repressive Polycomb complex at its -20kb enhancer region (Havis et al., 2012; Wang et al., 2011). In *Msx1* mutants, *Myf5* is upregulated (Wang et al., 2011). *Msx1* is regulated by Tcf4, a key transcription factor for connective tissue development (Miller et al., 2007). Forced expression of *Msx1* in fibroblasts prevents transdifferentiation by *Myod*, in accordance with its role as a repressor of myogenesis. *Msx1* function is necessary to delay the onset of differentiation and maintain the progenitor pool (Bendall et al., 1999). Interestingly, Msx1 is a key transcription factor for craniofacial development and it promotes cranial neural crest survival and patterning, a tissue that does not give rise to skeletal muscles (Ishii et al., 2005; Montarras et al., 2013; Soleimani et al., 2012b).

*Pitx* genes also play a prominent role in myogenesis (Hernandez-Torres et al., 2017). These transcription factors were first described for their involvement in pituary development, but were found to act on trunk and craniofacial development, left/right asymmetry as well as multiple other organs (Gage et al., 1999). *Pitx2* is the most extensively studied gene in myogenesis out of the 3 members of the paired-related homeodomain transcription factors. In the embryo, *Pitx2* is expressed in myogenic progenitors while *Pitx3* is expressed in differentiated cells. In *Pitx2* loss-of-function mice, limb muscles are severely affected (Hernandez-Torres et al., 2017). Pitx2 can directly activate *Myod* by binding to the -20kb enhancer, and in the absence of *Pitx2*, *Myod* is downregulated. However, Pitx2 was also reported to promote proliferation in the somite, as proliferation was reduced in its absence (Kioussi et al., 2002). In the trunk, *Pitx2;Myf5;Mrf4* triple mutants show a complete lack of muscle, a similar phenotype to the *Pax3* mutant, suggesting that *Pitx2* is a downstream effector of *Pax3* (L'Honoré et al., 2010). Indeed, when *Pax3* is overexpressed, *Pitx2* was found to be upregulated, and thus to be a potential direct target (Lagha et al., 2010).

# 1. Head myogenesis

Muscles of the head and neck form a highly heterogeneous group. The intricate bone arrangement of the mammalian skull is accompanied by a complex network of more than 60 muscles. They allow a wide range of movements, with tremendous differences in force. For instance, humans can crush food with strong mandibular muscles while maintaining breathing, facial expressions, speech and eye movements, all with astonishing coordination. This complexity is joined by a great diversity of origin and distinct regulatory mechanisms (Arnold, 2020; Chai and Maxson, 2006; Heude et al., 2018; Kelly, 2013; Noden and Francis-West, 2006; Sambasivan et al., 2009; Tzahor, 2015).

Cranial and neck musculature can be divided into two groups depending on their origin: those derived from unsegmented cranial paraxial mesoderm, (cardiopharyngeal mesoderm (CPM) and prechordal mesoderm (PCM)), and those derived from somites. Those derived from unsegmented cranial mesoderm include EOMs, mandibular, facial, pharyngeal, soft palate and laryngeal muscles, as well as few muscles of the neck (comprising the cucullaris-derived muscles, discussed below). Somite-derived muscles include the tongue, and some neck muscles (Michailovici et al., 2015; Schubert et al., 2018) (**Figure 11**).

## A. The pharyngeal arches: the source of branchiomeric muscles

CPM is contained in "pharyngeal arches" (PAs), which are transient structures in development, visible as multiple pairs of swellings forming sequentially on the ventrolateral sides of the developing foregut (**Figure 11**). Together with the frontonasal prominence, which will form the forehead and the nose, the pharyngeal arches give rise to most of the face and neck. Each pharyngeal arch contains neural crest mesenchyme, a mesodermal core, innervation and vascularization (Frisdal and Trainor, 2014). Neural crest cells (NCCs), are a vertebrate-specific population of multipotent progenitors that give rise to most of the craniofacial structures, including cartilage, bone, connective tissue and innervation (Douarin and Kalcheim, 1999; Gans and Northcutt, 1983; Green et al., 2015; Kuratani, 2007; Noden, 1983; Santagati and Rijli, 2003; Ziermann et al., 2018).

Between each arch, the endoderm and ectoderm are pinched together, forming pharyngeal pouches and clefts (Dennis, 2019). Pharyngeal pouches give rise to the eustachian tube, tonsils, and glands such as the thymus, parathyroid glands and ultimobranchial bodies (Carlson, 2014). Pharyngeal clefts give rise to the external auditory tube, and cervical sinus (Carlson, 2014; Dennis, 2019). The organization of the PAs is also highly conserved from invertebrate chordates such as amphioxus (even in absence of NCCs), to vertebrate agnathans and gnasthostomes including avians, squamates, and mammals. The number of arches varies between species but all developing vertebrate possesses this pharyngeal conformation (Graham, 2001; Miyashita, 2016; Poopalasundaram et al., 2019). Since this organization is thought to originate from gills ("*branchiae*" in latin), the muscles derived from these arches are sometimes referred to as "branchiomeric muscles". In mammals, there are 5 pairs of PAs. Classically, the numbering of these arches has been 1, 2, 3, 4 and 6, arguing that derivatives of the 5$^{th}$ arch degenerate as they appear in development (Frisdal and Trainor, 2014). Recent reports have since questioned this nomenclature, arguing that there is no convincing justification for this odd numbering (Graham et al., 2019). However, this issue being rather recent, herein, we will use the traditional nomenclature of arch 4 and 6 while referring to them as "caudal arches".

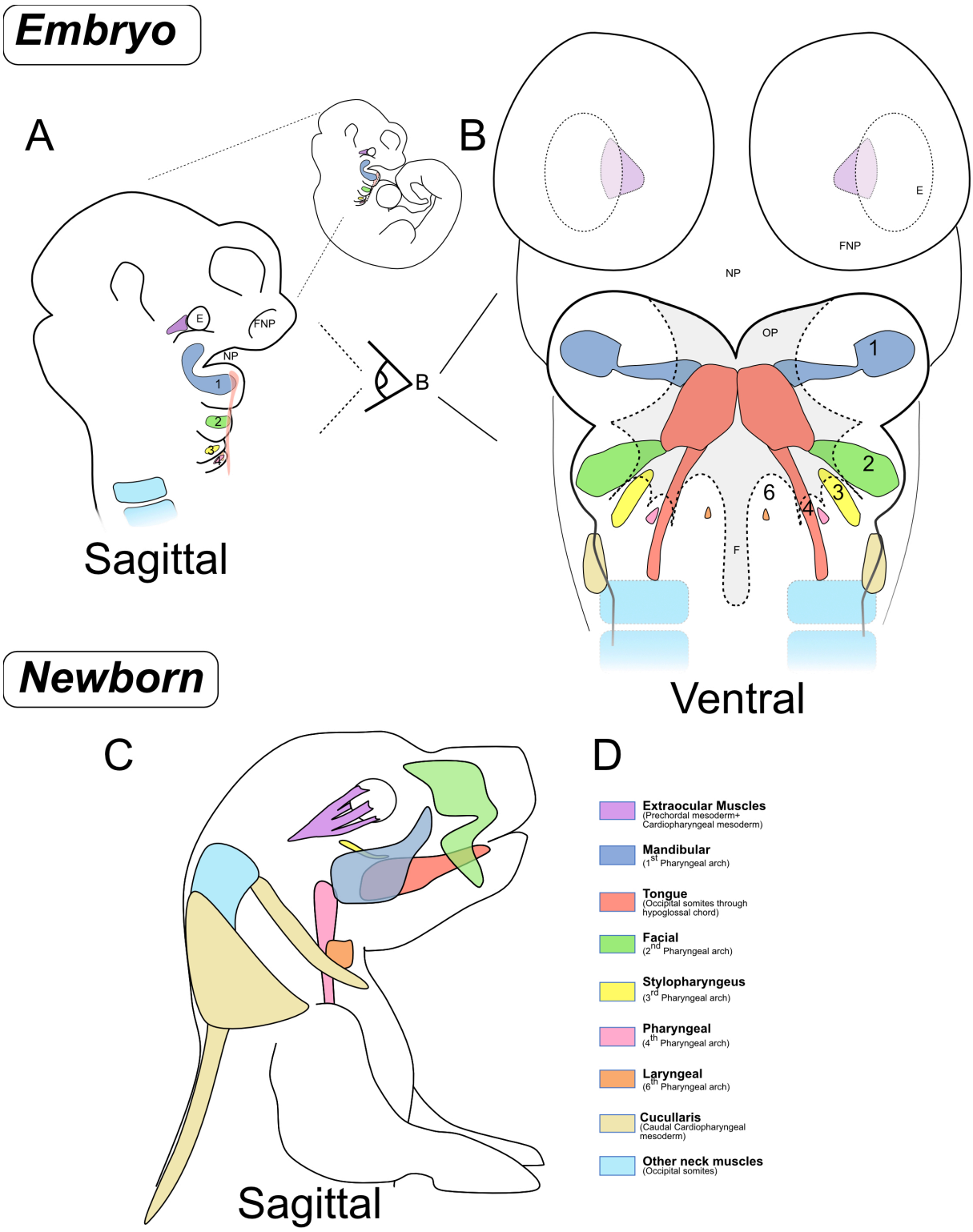



**Figure 11. Head myogenic progenitors and their derivatives.**

(A-B) Sagittal and ventral schematic views of the developing mouse embryo, showing the anatomy of the early pharyngeal apparatus and its myogenic compartments. E: Eye, F: Foregut, FNP: Frontonasal process, NP: Nasopharynx, OP: Oropharynx
(C-D) Sagittal schematic view of a newborn mouse highlighting the derivatives of the embryonic structures.

Head muscles arise from CPM, with the exception of the EOM, that derive from CPM and prechordal mesoderm (Diogo et al., 2015; Evans and Noden, 2006; Noden and Francis-West, 2006). PCM is defined mainly by its location, anterior to the notochord. It is known to contribute to the EOMs, but its boundaries with regards to the CPM (and their relevance for embryogenesis) are not well established (Ziermann et al., 2018). CPM progenitors were shown to give rise to both cranial and cardiac muscles, with specific clonal relationships (Diogo et al., 2015). Interestingly, lineage studies in mouse demonstrated that specific head muscles have a clonal ancestry with specific parts of the heart (Lescroart et al., 2010; Lescroart et al., 2015) These results suggest that 3 main "CPM fields" of multipotent progenitors can be distinguished: 1)EOM-PA1-Right Ventricle, 2) PA2-Pulmonary trunk-Aorta, 3) Caudal arches-Atria-Caval veins-Pulmonary vein (**Figure 12**). Interestingly, these 3 clonal CPM fields might correspond to the 3 main streams of NCCs from the dorsal side of the embryo to the ventral side. Indeed, in the pharyngeal arches, neural crest is completely surrounding the CPM mesodermal core (Douarin and Kalcheim, 1999; Frisdal and Trainor, 2014). During this phase, CPM myogenic progenitors and NCCs are juxtaposed, and NCCs have been shown to provide cues to pattern cranial muscles once myogenesis has been initiated (Rinon et al., 2007). This intricate relationship will be developed further in **Chapter IV**.

## B. The caudal arch muscles

Muscles derived from the caudal pharyngeal arches (4-6) comprise the soft palate muscles (except the tensor veli palatini and the tensor tympani which derive from the first arch), pharyngeal constrictor, laryngeal, esophageal, trapezius and sternocleidomastoid muscles. In comparison with muscles of the first and second arches, these muscles have been less well investigated, and no markers are currently available (Lescroart et al., 2015; Naumann et al., 2017; Ziermann et al., 2018). To clarify the genetic regulatory networks operating in the caudal arches, we investigated the expression patterns of *Hox* genes and found that *Hoxb4* expression specifies the anterior boundary of the caudal arches (**Annex 2**). More specifically, *Hoxb4* delineates the mesenchymal frontier between arch 3 and 4. The origin of caudal arch-derived muscles has long been a subject of controversy, and most of them are absent in avian models and reptiles (Smith, 1992).

## B. The cardiopharyngeal mesoderm

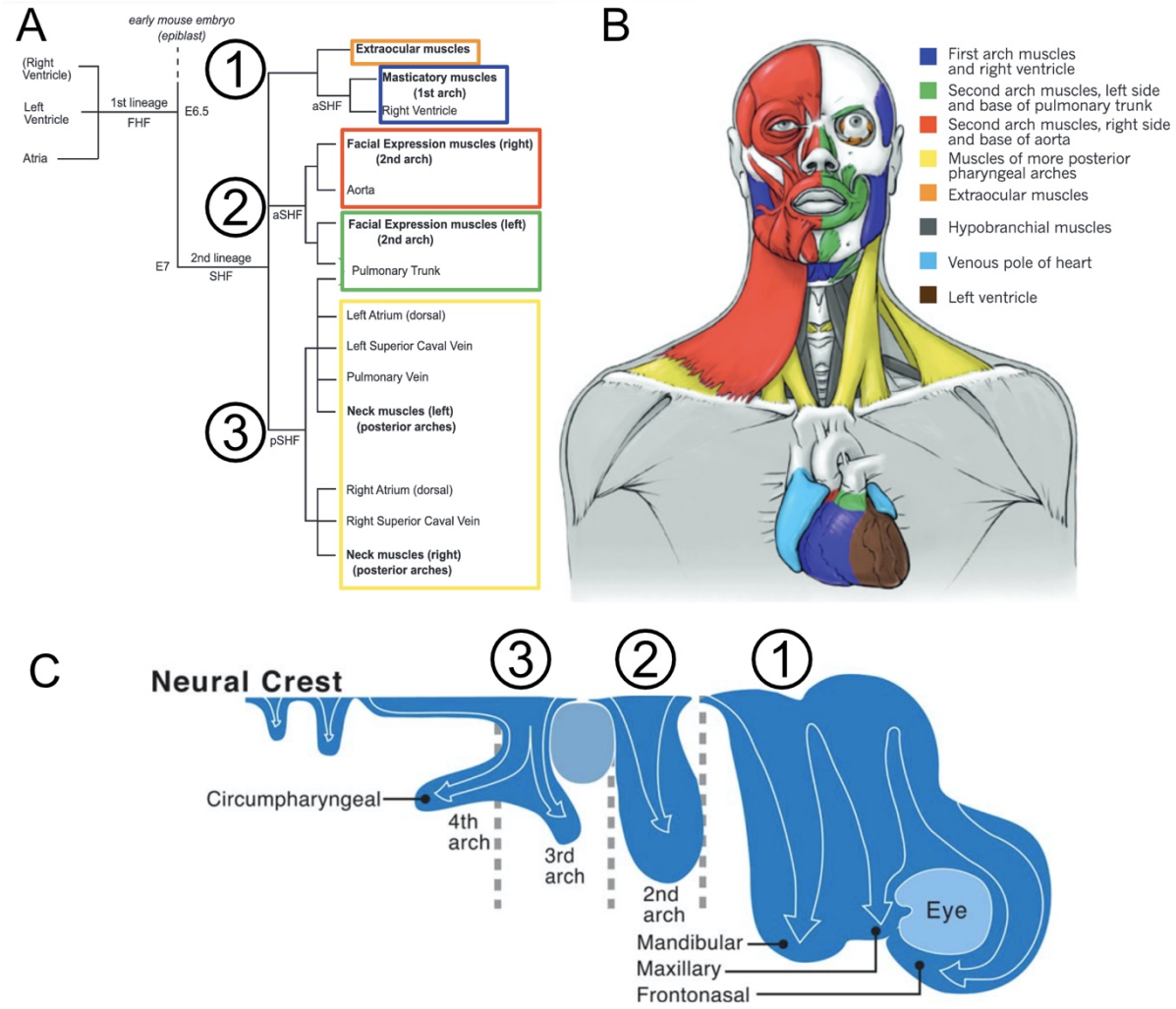



**Figure 12. Branchiomeric and cardiac muscles have common ancestors.**

(A) Hierarchical representation of cardiopharyngeal progenitors. FHF: First heart field (an initial population of cardiac progenitor giving rise to the left ventricle), SHF: Second wave of cardiac progenitors, associated with head muscles), aSHF, pSHF: anterior and posterior heart field. Adapted from (Buckingham, 2017).

(B) Ilustration of head and cardiac muscle associations. From (Diogo et al., 2015)
(C) Schematic representation of neural crest migration stream as they populate the cranial region. From (Ziermann et al., 2018).

This lack of information is possibly related to the facts that these muscles are not all directly visible in a whole mount preparation of mouse embryo, their development timeline is shifted compared with anterior muscles and their anatomy is particularly complex (**Figure 13**).

The first evidence of muscle formation in pharyngeal, soft palate and laryngeal muscles starts at E12.5 and E11.5 in mouse, when most other cranial muscles have undergone terminal embryonic differentiation (Grimaldi et al., 2015) (present study). The esophagus striated muscles (ESM), were shown to be the most posterior CPM-derived muscles that start to be muscularised at around E13.5, and skeletal muscle formation continues into postnatal stages in an anteroposterior gradient with the concomittant displacement of smooth muscle cells (Baguma-Nibasheka et al., 2019; Gopalakrishnan et al., 2015). Previously thought to originate from smooth muscle to striated transdifferentiation (Patapoutian et al., 1995; Rishniw et al., 2003), ESM progenitors actively migrate along a smooth muscle scaffold from the pharyngeal end towards the stomach, through the action of MET/HGF signaling (Comai et al., 2019).

Laryngeal muscles are one of the most recently acquired in vertebrates and are linked to a substantial variety of social and behavioural characteristics of mammals by allowing refined vocalizations (Smith, 1992). Despite their importance in communication, their development is poorly understood and their origin remains controversial, as retroviral labeling experiments have identified them to be derived from the 1$^{st}$ occipital somites in the chick (Evans and Noden, 2006). However, as the authors of that study indicate, the boundary between the first occipital somite and the unsegmented CPM is particularly challenging to delineate, hinting that these muscles could arise from CPM in chick. In mouse, these muscles were shown to be CPM-derived, marked by the expression of *Isl1* (Heude et al., 2018; Tabler et al., 2017). Accordingly, using a *Mef2c-AHF-Cre* (a specific *Mef2c* anterior heart field enhancer) line (Dodou et al., 2004), a recent cell lineage tracing approach showed that these muscles, some of their cartilages, and their surrounding mesenchyme are CPM-derived (Adachi et al., 2020).

Owing to their location adjacent to the developing foregut, laryngeal myogenic progenitors are in range of *Shh* signaling emanating from the endoderm (Sagai et al., 2017). This proximity was found to be relevant in the context NCC expansion into the laryngeal region (Tabler et al., 2017). Indeed, contrary to previously reports, the NCC posterior frontier is located in the larynx, where it only gives rise to most of the thyroid cartilage and associated connective tissues. In *Fuz* mutant mice, where cilia-mediated *Shh* signaling is disrupted, this frontier is lost and NCC surrounds the foregut, leading to massive morphological defects. More subtle disturbance of the pathway using *Gli3* loss-of-function leads to milder effects, impacting on vocalization frequencies (Tabler et al., 2017). Importantly, the laryngeal muscles in avians where most of their vocalization capacity originates from the syrinx are greatly reduced

compared with mammals (Kingsley et al., 2018). This specialized muscle structure is located at the base of the bronchi and allows tight control of air flow (developed in discussion).

Although cranial mesodermal progenitors appear relatively early in development, as morphogenesis progresses anteroposteriorly, head muscles differentiate later than those in the somites. It is interesting to note that Pax7+ stem cells also appear later, from mid-embryogenesis, after the onset of myogenesis (Nogueira et al., 2015).

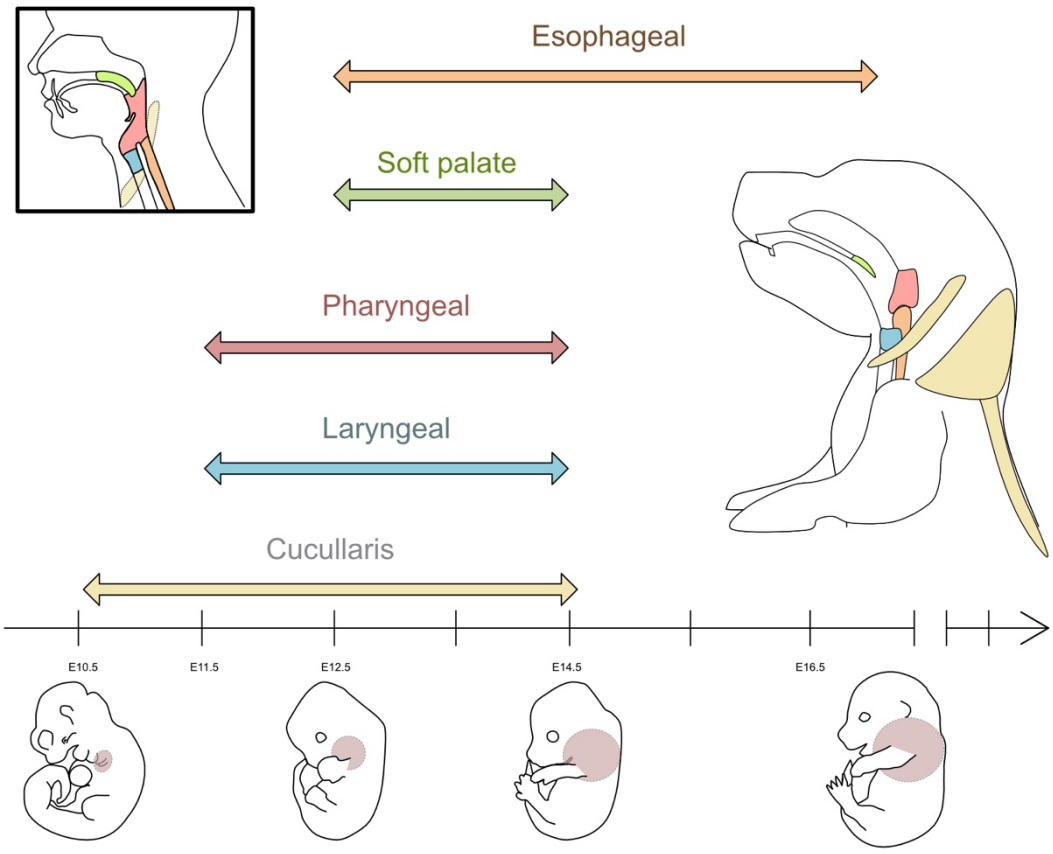



**Figure 13. Developmental timeline of caudal arch-derived muscles.**
Arrows represent the extend of full differentiation across the muscle anlage.
Darker area in embryos represent the expansion of the caudal arch-derived muscles over time.

## C. Somitic head muscles: neck and tongue

The tongue is the most anteriorly-located migratory hypaxial muscle originating from somites. Myogenic progenitors migrate from the occipital somites into the tongue primordium

along the hypoglossal chord (Huang et al., 1999) and invade the neural crest derived mesenchyme starting at E11.5 (Han et al., 2012; Parada and Chai, 2015). This migration depends on the expression of *Pax3/cMet/Hgf*, but not *Lbx1* (Amano et al., 2002; Bladt et al., 1995; Dietrich et al., 1999; Gross et al., 2000). Throughout development, tongue myogenic progenitors will receive promoting factors from interstitial neural crest cells and the lining epithelium, which will be developped further in **Chapter IV** (Cobourne et al., 2018). Tongue muscles originate from a Pax3+/Mesp1+ lineage, specific to occipital somite derivatives. Whether this specificity provides distinct phenotypic characteristic is currently unknown.

Somitic neck muscles are also derived from the occipital somites. In addition to *Pax3*, and similar to tongue muscles, these muscles are marked by *Mesp1*, a marker of anterior mesoderm and heart, unlike the rest of somitic-derived muscles in the trunk and limbs (Heude et al., 2018). Similar to the tongue, neck muscles only exhibit slight morphological defects in *Pax3* mutants, while more posterior somitic muscles are absent (Heude et al., 2018). However, in contrast to tongue muscles, somitic neck muscles are not migratory and are unaffected in *cMet* mutants (Vasyutina et al., 2005). This genetic program that overlaps with CPM-derived muscles (Mesp1+/Pax3+/Lbx1-), might impart specific properties to these muscles, although this remains unexplored.

## D. Gene regulatory network

Some heterogeneity was described regarding the specific requirement of MRFs for different cranial muscles. For example, *Myf5*;*Mrf4* double mutant embryos lack EOMs, yet other craniofacial muscles are not overtly perturbed. Using $Myf5^{nlacZ/nlacZ}$ homozygous mice, (introduction of *nlacZ* disrupts both *Myf5* and linked *Mrf4 in cis*), leads to apoptotic β-gal+ progenitors in the extraocular region (Sambasivan et al., 2009). *Myod* and *Myog* expression were greatly reduced in EOM compared to PA1 muscles. In addition, analysis of $Myf5^{loxP/loxP}$ and $Mrf4^{-/-}$ embryos (affecting *Myf5* only and Mrf4 only, respectively) showed a mild phenotype in the EOM for each single mutant, and no phenotype in the PA1 (Kassar-Duchossoy et al., 2004; Sambasivan et al., 2009). Interestingly, when *Myod* was introduced into the *Myf5* locus, development of the EOM was rescued. (Kassar-Duchossoy et al., 2004; Sambasivan et al., 2009). These observations suggest a compensatory mechanism for the loss of *Myf5* and *Mrf4* that would allow activation of *Myod*, taking place in PA1, but not in EOMs (Sambasivan et al., 2009).

### a) Tbx1

CPM development depends on the function of *Tbx1*, a T-box transcription factor implicated in the most common microdeletion in human: DiGeorge syndrome (Kelly et al., 2004). This disease is marked by cardiovascular and craniofacial defects. Studies in mice have shown *Tbx1* to be the major contributor to the disease, and knock out experiments recapitulate a number of human cardiovascular and craniofacial phenotypic traits (Jerome and Papaioannou, 2001; Kelly et al., 2004). Importantly, the human disease condition reflects a heterozygous context while heterozygous mice do not seem to exhibit any severe intrinsic muscle phenotype (Jerome and Papaioannou, 2001). Interestingly, phenotypic variability is also observed in muscles derived from the 1$^{st}$ arch (mandibular). Indeed, while all of the more caudal muscles are absent in *Tbx1* mutants, and myogenesis in compromised from cells in 2$^{nd}$ arch, mandibular muscles are only partially affected (Grifone et al., 2008). *Tbx1* is expressed in mesoderm, endoderm and ectoderm, but absent in neural crest (Vitelli et al., 2002). Endodermal expression was shown to play a role in pharyngeal arch segmentation, leading to suggestion that the myogenic defects seen in arch 2 to 6 could be indirect (Arnold et al., 2006). However, mesodermal-specific expression of *Tbx1* in a *Tbx1* null background rescued most of the defects found in the mutant, suggesting that myogenic cells do require *Tbx1* cell-autonomously (Zhang et al., 2006). These observations were supported by recent studies using chimeric mice from our group (Comai et al., 2019). However, it is to be noted that *Tbx1* promotes *Fgf10*, so the myogenic defects seen in *Tbx1* mutants might be in part due to proliferation defects, which depends on FGF signaling (Kong et al., 2014). Analysis *Tbx1;Myf5* double mutant embryos showed that 1$^{st}$ arch muscles were compromised, placing Tbx1 as a potential regulator of *Myod*, complementary to *Myf5* in this region (Sambasivan et al., 2009). This lead us to question the compensatory potential of Tbx1 in maintaining the integrity of craniofacial musles in *Tbx1*$^{+/-}$*;Myf5*$^{nlacZ/nlacZ}$ (**Annex 1C**). In these mutants, *Myf5* and *Mrf4* are disrupted, and a single copy of *Tbx1* is functional. We did not find obvious defect in muscles of the face, mandible and larynx. However, a more detailed analysis needs to be performed to assess the complete phenotype of these muscles. Several reports have highlighted the role of transcription factors other than *Tbx1* in cranial myogenesis, thereby adding further layers on complexity within this genetic regulatory network (**Figure 14**).

b) Tcf21/Msc

*Tcf21* and *Msc* are transcription factors expressed in cranial myogenic progenitors before MRF expression. In *Tcf21;Msc* double mutant embryos, masticatory muscles are absent and *Myf5* expression is reduced. In *Msc* mutants, *Myf5* and *Myod* expression are reduced but *Tbx1* is upregulated, indicating that cells are less prone to myogenic commitment. Regulatory sequences of *Myf5* and *Myod* bind TCF21 and MSC to initiate myogenenesis in pharyngeal arches (Moncaut et al., 2012). MSC was found to bind similar targets as MYOD (MacQuarrie et al., 2012).

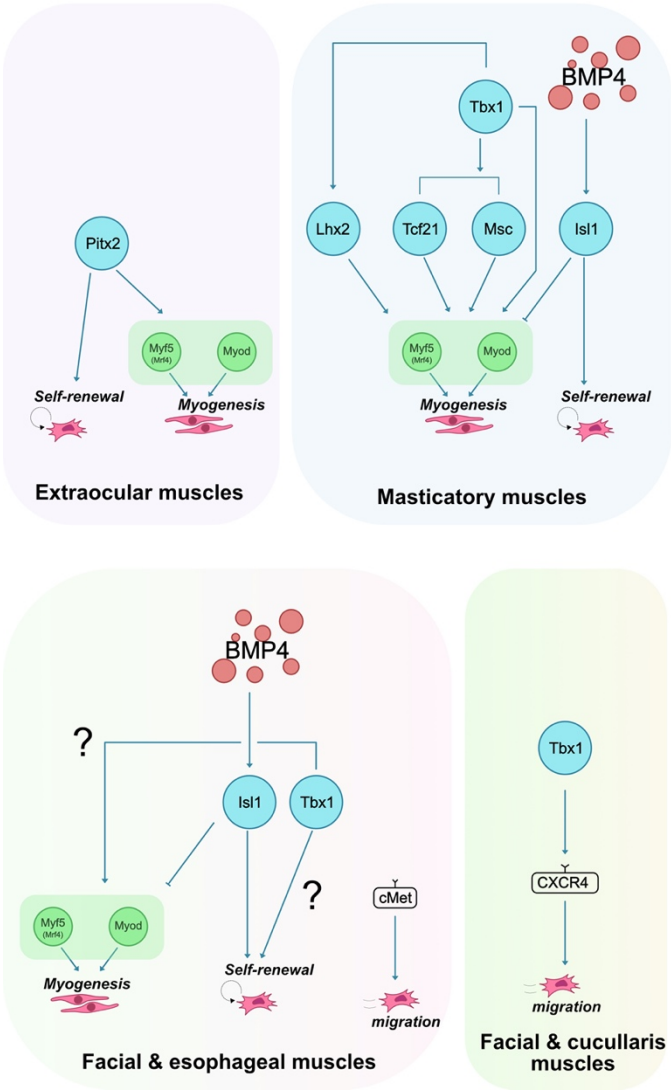



**Figure 14. Gene regulatory networks of branchiomeric muscles.**

Little is known about the genetic hierarchy governing head myogenesis. It is likely that *Tbx1* regulated both expansion of progenitors at early stages and activation of myogenesis at the appropriate timing. More mechanistic studies need to be carried out to uncover the gene requirement of other pharyngeal arch muscles.

c) Six family

*Six* genes appear to have a more restricted role during craniofacial development. *Six1/4* double mutants do not show any phenotype in the head, most likely because of compensatory effects from other *Six* family members (Maire et al., 2020). However, analysis of *Six1;Eya1* double mutant mice revealed hypoplasic muscles of the extraocular region in addition to muscles originating from PA1 and PA2. Furthermore, comparing *Six1* and *Eya1* expression in *Tbx1* mutant embryos, and *Tbx1* expression in *Six1;Eya1* double knock-outs suggested that *Tbx1* lies upstream of *Six1/Eya1* (Guo et al., 2011; Maire et al., 2020). Interestingly, *Six1a* disruption in zebrafish leads to the absence of EOM, suggesting a rearrangement of *Six* genes required in the head myogenesis hierarchy during evolution, and the relationship with *Tbx* in cranial mesoderm in zebrafish remains unknown (Lin et al., 2009).

d) Pitx2

In the case of the extraocular muscles, deletion of *Pitx2* causes apoptosis in the muscle primordia, reminiscent of the *Myf5*;*Mrf4* double mutant (Sambasivan et al., 2009; Zacharias et al., 2010). In the 1st pharyngeal arch, the absence of upstream myogenic cells results in a specification defect and reduction of the overall size of the primitive jaw, suggesting that *Pitx2* acts also on other tissues (Shih et al., 2007). In this case, early markers of PA1 muscle progenitors like *Tcf21* and *Msc* were also absent. *Pitx2* is more highly expressed in the EOM than any other muscle examined, and it promotes the transcription of EOM specific genes (Zhou et al., 2012). Taken together, *Pitx2* is considered to be an essential upstream regulator of the extraocular muscles, controlling both proliferation and lineage commitment. It has been proposed that *Pitx2* might contribute to the unique features that EOM exhibit, such as sparing in some myopathies (Kaminski et al., 2002; Man et al., 2005; Schoser and Pongratz, 2006; Stuelsatz et al., 2015). Considering its involvement in other muscles (particularly the 1st arch), it is also thought to be a potent compensatory effector to initiate myogenesis, as shown in the partial mandibular defect seen in the PA1 derivatives in *Tbx1* mutant (Grifone et al., 2008).

e) Lhx2

Lhx2 is a transcription factor that plays a role in pharyngeal arch myogenesis. *Lhx2* acts downstream of *Tbx1*, *Pitx2*, and *Tcf21*, which directly regulate it, and upstream of *Myf5* (Harel et al., 2012). *Lhx2* mutants has defect in muscle specification and patterning in the head.

In *Tbx1;Lhx2* double mutant embryos, head muscles are absent (Harel et al., 2012). In addition, more Isl1+ cells and less Pax7+ cells are found in the PA1, indicating that *Lhx2* might play a role in promoting myogenic commitment in the CPM.

### f) Isl1

*Isl1* a marks cardiopharyngeal mesoderm, both in the cardiac and muscle lineages. The early lethality of *Isl1* mutant embryos (before myogenesis is established) has made functional studies of this regulator challenging (Cai et al., 2003). Nevertheless, *Isl1* was shown to repress myogenic differentiation in chick (Tzahor and Evans, 2011), and can promote proliferation within the cardiac lineage (Cai et al., 2003). *Isl1* expression was found to be upregulated following exogeneous administration of BMP4, consistent with the anti-myogenic effect of BMP4 in the somites, while promoting cardiac markers (Harel et al., 2009). More recently, our group developed a chimeric model to study the cell autonomous requirement of *Isl1* (Comai et al., 2019). The focus of this study was on the esophageal muscle, a muscle that was described to be derived from cardiopharyngeal mesoderm (Gopalakrishnan et al., 2015). The analysis of *Isl1* null chimaeric embryos showed that lineage marked *Isl1* null cells did not contribute to striated esophageal muscles, suggesting a cell autonomous requirement of *Isl1* for that muscle (Comai et al., 2019). The contribution of *Isl1* null cells to other cranial-derived muscles remains to be explored.

### g) cMet

cMET is a receptor for HGF ligand and this signaling pathway promotes migration of progenitors, well studied for their involvement in limb myogenesis (Birchmeier and Brohmann, 2000; Bladt et al., 1995; Buckingham and Rigby, 2014). In the head, *cMet* is necessary for tongue, facial muscles (PA2), and esophagus striated muscle (caudal arch) development (Comai et al., 2019; Prunotto et al., 2004). Interestingly, in the *cMet* mutant, esophageal muscle progenitors are found at the base of the pharynx and they do not migrate posteriorly into the esophagus, yet the laryngeal and pharyngeal muscles form normally in these mutants (Comai et al., 2019).

### h) Cxcr4/Sdf1

CXCR4 and its ligand SDF1 is a known signaling pathway necessary for cell migration, and together with Met signaling, it is required for limb and tongue development

(Vasyutina et al., 2005). In keeping with the requirement for Met for migration of 2$^{nd}$ arch progenitors, these cells require the expression of *Cxcr4*. In *Cxcr4* mutants, myogenic progenitors were shown to be reduced in PA2-derivatives but normal in PA1-derivatives (Yahya et al., 2020). When beads containing CXCR4 inhibitors were applied to the proximal part of PA2, myogenic expression was reduced. Conversely, when SDF was applied, more myogenic expression was observed in PA2. Surprisingly, the authors also reported reduction or absence of non somitic neck muscles, such as the trapezius muscles and sternocleidomastoid (Yahya et al., 2020). Given that these muscles have not been reported to be actively migratory, it would be interesting to reevaluate their disposition in other mutant models, including *Met* and *Lbx1* knock-outs.

# Chapter IV

# Extrinsic cues to diversity

Throughout the body, skeletal muscles are associated with a variety of tissues including nerves, vasculature, immune cells, tendons, bones and connective tissue. Interestingly, muscle-associated connective tissue (MCT) plays a predominant role in embryonic development, adult contractile function and force transmission, and injury repair (Biferali et al., 2019; Nassari et al., 2017; Passerieux et al., 2007; Sefton and Kardon, 2019). Anatomically in the adult, MCT has been defined as 3 layers in muscle: the epimysium, perimysium and endomysium (Passerieux et al., 2007). Each myofibre is ensheathed in endomysium, while perimysium surrounds the fascicle, and epimysium surrounds the whole muscle and connects the myotendinous junction (Passerieux et al., 2007).

MCT cells are challenging to categorize since some specific features are often shared among different populations. Production of ECM structural molecules like collagens and proteoglycans, fibro/osteo/adipogenic potential in vitro, and transcription factors are often common features (Muhl et al., 2020). As a consequence, terms like “mesenchymal stromal/stem cells”, “connective tissue”, “fibroblastic/fibrogenic”, and “pericytes” are found used interchangeably. However, MCTs taken from different parts of the body were found to have specific properties in human, possibly related to their developmental origin (Sacchetti et al., 2016). Thus, it is important to understand how MCT arise in development and how they interact with myogenic progenitors to establish skeletal muscles at various locations in the body.

## 3. Trunk and limb connective tissue guides muscle formation

Trunk and limb connective tissues derive from mesoderm. Some MCT can arise from the somites, and associate with epaxial or hypaxial muscles. These muscles are named “primaxial” (or “closer to the axis”, i.e. the neural tube) (Burke and Nowicki, 2003). Lateral plate mesoderm (LPM), can also give rise to connective tissue, such as in the limb and some hypaxial muscles (Prummel et al., 2020). Hypaxial muscles with LPM-derived MCT are known as “abaxial” (or “not on the axis”). The separation between somite-derived and LPM-derived MCT is called the “lateral somitic frontier” (Burke and Nowicki, 2003; Nowicki et al., 2003).

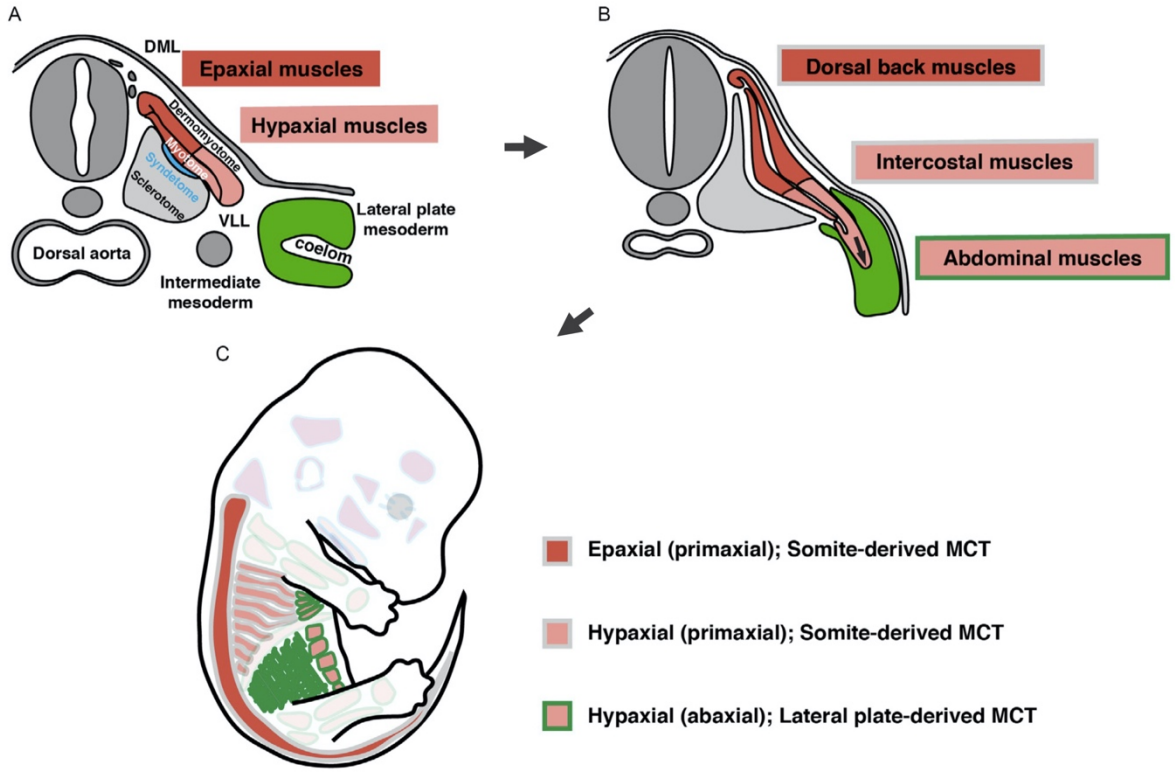



**Figure 15. Epaxial and hypaxial muscle connective tissue development.**
MCT can arise from the somite and associate with primaxial muscles (epaxial or hypaxial), or from the LPM and associated with abaxial muscles (hypaxial only) (Sefton and Kardon, 2019).

## A. Primaxial muscle connective tissue

In addition to epaxial and hypaxial muscles, the somites give rise to many other lineages, including dermis, adipocytes, smooth muscle, angioblasts, cartilage, bone, tendon, and connective tissue. As myogenic cells originate from the dermomyotome, mesenchymal progenitors arise from the sclerotome, located ventromedially. Another compartment, called the syndetome, forms from the cranial and caudal edges of the sclerotome, upon signals from the myotome (Brent and Tabin, 2004; Brent et al., 2003; Brent et al., 2005). Although no study has yet formally showed it, it is presumably in this context of dense mesenchymal heterogeneity that somitic connective tissue forms (Burke and Nowicki, 2003). Avian transplantation experiments have shown that some somitic muscles possess somite-derived connective tissue (Saberi et al., 2017) (**Figure 15**). Transgenic labeling of Scleraxis (Scx), a transcription factor found in the syndetome, show Scx+ cells in epaxial muscles (Deries et al., 2010). Taken together, it is likely that primaxial MCT arises from the syndetome. However, more detailed

analyses need to be carried out to understand the specific mechanisms regulating MCT as opposed to tendon or chondrogenic fates.

The morphogenic events establishing muscle patterns remain unclear. Sequential whole-mount immunofluorescences showed that epaxial muscles form through a multistep process of translocation, re-orientation, elongation, and cleavage of the myotome (Deries et al., 2010). Consistenly with that, the same authors showed in another study that epaxial muscle formation in the somites is associated with changes in ECM components (Deries et al., 2011). Both the dermomyotome and the myotome appear to be surrounded by laminin-rich ECM that is actively remodeled at each stage of myogenic development. When myocytes start appearing, the laminin layer disassembles and fibronectin is deposited around myogenic cells. This suggests that fibronectin deposition may be associated with myogenic maturation in the epaxial muscles (Deries et al., 2011). Given that fibroblasts were found within the myotome, it is possible that these fibroblasts participate to muscle formation by regulating fibronectin expression.

## B. Abaxial muscle connective tissue

Abaxial MCT is derived from lateral plate mesoderm (LPM). However, the anterior-most boundary of LPM is not well defined and it is possible that the nomenclature of anterior LPM, CPM, and PCM overlap in this region (Prummel et al., 2020). More generally, LPM gives rise to blood, heart, the cardiovascular system, kidneys, smooth muscle, skeleton and connective tissue and it is located lateral to the somites (Durland et al., 2008). Although some markers have been used to track LPM during development in different animal models, these markers do not always coincide resulting in conflicting results (Prummel et al., 2020). Abaxial muscles comprise those in the limbs and some hypaxial muscles including the ventral region (Burke and Nowicki, 2003; Nowicki et al., 2003) (**Figure 15**).

### a) Connective tissue of ventral muscles

*Hox* genes define regional identity of tissues along the anteroposterior (rostrocaudal) axis (Krumlauf, 1994) (discussed in **Chapter VI**). Muscles along the body axis express *Hox* genes corresponding to their position. Interestingly, when somitic hypaxial progenitors were transplanted into a different region of the body, they expressed *Hox* genes of the local LPM (Nowicki and Burke, 2000). This data suggests that LPM-derived MCT may instruct local

identity to the developing muscle. A study examining the effect of the loss of *Pitx2* in abaxial abdominal muscles suggested that *Pitx2*-driven repressive regulation of *Hox* genes guides muscle patterning (Eng et al., 2012). In *Pitx2* mutant embryos, abdominal muscles are mispatterned as they elongate and sustain normal MRF expression. The authors concluded that perturbations of these muscles arose from a cell non-autonomous role of the LPM-derived MCT determining muscle progenitor orientation (Eng et al., 2012). In an LPM-specific context using a *Prx1-Cre* line, loss of *Fat1*, an atypical Fat-like cadherin planar cell polarity molecule, resulted in a hypoplasic ventral cutaneous maximus muscle, with deformed myofibres (Helmbacher, 2018). This study provided a clear example of the implication of LPM-derived MCT in muscle formation.

The diaphragm is an essential mammalian muscle and its development was recently examined in detail (Sefton et al., 2018). From sequential temporal views of diaphragm myogenesis, it appears that the pleuroperitoneal folds, which precede muscle expansion and innervation, first establish a scaffold. This LPM-derived MCT expresses *Gata4* as a marker (which is affected in human congenital defects). *Gata4* promotes the secretion of myogenesis-inducing factors and thus promotes myogenic expansion.

b) Limb muscle connective tissue

Limb muscle development has been extensively studied due to its accessibility in chick/quail transplantation studies, and the availability of genetic tools for known markers of the different regions regulating muscle morphogenesis and myogenesis. Myogenic and connective tissue lineages can be traced in mouse using: $Pax3^{Cre}/Pax7^{Cre}/Pax7^{CreER}$ (for the muscle compartment) and *Tg:Prx1-Cre/Tg:Prx1-CreER* (for the LPM-derived MCT) (Durland et al., 2008; Engleka et al., 2005; Hasson et al., 2010; Lepper et al., 2011; Logan et al., 2002; Murphy et al., 2011). Additionally, markers of MCT have been identified as *Tcf4*, and *Osr1* (for a subset of limb MCT) (Kardon et al., 2003; Vallecillo-García et al., 2017). Limb MCT, as well as bone, tendon and cartilage derive from LPM (Prummel et al., 2020). Interestingly, a lineage study performed in chick showed that limb tendon and MCT derive from one lineage while perichondrium and cartilage derive from a separate lineage (Pearse et al., 2007). It is unclear if such distinction is operating in the somites. LPM mesenchymal progenitors first form the limb primordium (called "limb bud") starting from E9.5. This process initiates with the expression of *Tbx5* and *Tbx4* in the forelimb and hindlimb respectively (induced by *Hox* genes, discussed in **Chapter III**). These T-box transcription factors establish the identity of the limb

and initiate an *Fgf10-Fgf8* feedback loop between mesenchyme and overlying ectoderm promoting limb expansion in the proximo-distal axis. Specification in the anteroposterior axis is dictated by *Shh* signaling emanating from the zone of polarizing activity, promoting FGF production in the apical ectodermal ridge. These 2 signaling centers promote the expansion of underlying LPM. During this process, HGF/SF and SDF1 are secreted by LPM-derived mesenchyme to attract Met+/Cxcr4+ myogenic precursors to the site of terminal differentiation, as described above (Dietrich et al., 1999; Heymann et al., 1996; Vasyutina et al., 2005) (**Figure 16**). In addition to these chemotractant mechanisms, the LPM also produces repulsive signals, in the form of EPHRINA5 (ligand) that acts on EPHRIN4 (receptor)-expressing myogenic progenitors. This interaction was suggested to result in repulsion of myogenic cells away from the peripheral limb bud towards the central region, where it is not active (Swartz et al., 2001).

Once myogenic progenitors arrive into the limb bud, muscle patterning occurs resulting in alignment of muscle fibres along their future trajectories leading to proper muscle attachments to the bones. During this process also, LPM-derived MCT have been proposed to play a predominant role in patterning (**Figure 16**). Strinkingly, even in absence of muscle, non-muscle cells are organized in a "muscle-like" arrangement (Grim and Wachtler, 1991).

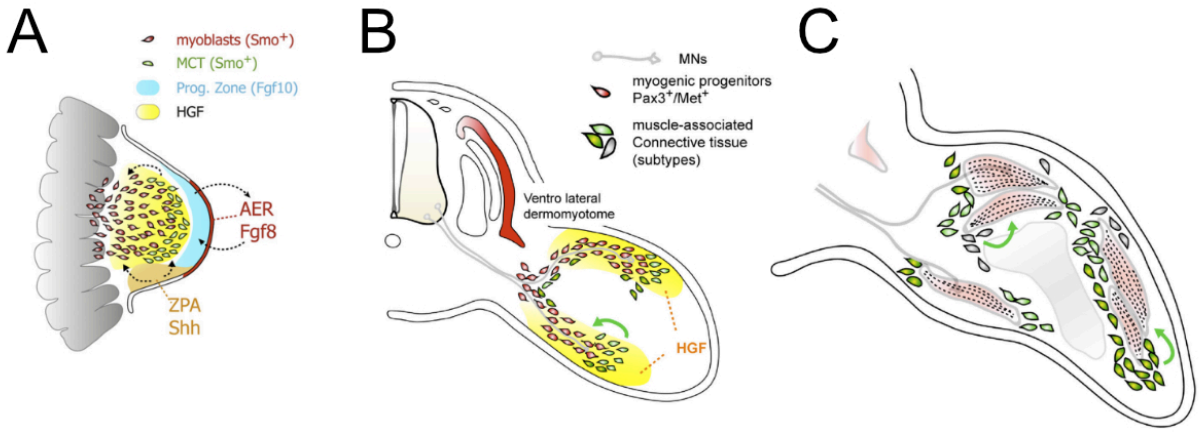



**Figure 16. Limb muscle and connective tissue development.**

(A) Feedback loops between mesenchyme and signaling centers initiate limb bud formation
(B) Chemotractant signals (HGF) and repulsive signals (Ephrins, not shown) guide muscle progenitors into the limb bud.
(C) Connective tissue subtypes carrying patterning identity support proper myofibre orientation. Adapted from (Helmbacher and Stricker, 2020).

*Tcf4* was shown to play a central role in MCT prepatterning. Tcf4+ cells were found in the limb bud arranged in a muscle-specific conformation in a muscle-less limb. Furthermore, loss of function and gain of function experiments of *Tcf4* resulted in mispatterned or ectopic limb muscle (Kardon et al., 2003). These results suggest that limb MCT guides naïve myogenic progenitors through attractive and repulsive signals, and promotes myogenic growth in a prepatterned limb bud to establish proper orientation and function.

Importantly, a few upstream transcription factors of MCT fibroblasts have been identified such as *Tbx3-5, Hoxa11/d11* and *Osr1*. *Tbx3* and *Tcf4* were shown to affect muscle patterning in a muscle-independent manner, as $Pax3^{Cre}$-driven deletion does not lead to muscle phenotypes (Colasanto et al., 2016; Mathew et al., 2011). Notably, *Osr1* and *Tbx5* regulate the number and shape of Myod+ progenitors, while *Tbx5* appears to control myogenic differentiation (Colasanto et al., 2016; Vallecillo-García et al., 2017). Importantly, when *Osr1* is depleted, limb LPM-derived cells adopt a more chondrogenic/tenogenic behaviour, which suggests a cell fate regulatory role of *Osr1* on LPM fibroblasts (Vallecillo-García et al., 2017). Sefton and colleagues have described 3 potential mechanisms of action of these transcription factors in muscle patterning (Sefton and Kardon, 2019). First, by directly regulating fibroblastic cell fate, thus providing competent MCT cells for muscle guidance. Second, by patterning the MCT cells, such as in *Hoxa11/d11* mutants which display patterning defects of MCT and associated muscles (Swinehart et al., 2013). Lastly, they may act on the secretion of molecules by the MCT, which are thought to be responsible for muscle guidance. In *Tbx5* mutants, MCT cells express less N-CADHERIN and β-CATENIN, and in a more disorganized manner (Hasson et al., 2010). *Osr1* on the other hand, promotes secretion of SDF, BMP4 and COLVI (Vallecillo-García et al., 2017). In addition, members of the FGF family emanating from the LPM promote proliferation of myogenic progenitors (Robson and Hughes, 1996).

Taken together, extensive work in limb development has shown that the MCT plays an instrumental role in appendicular muscle formation through secreted molecules (ex. positive and negative chemotaxis, growth factors, and ECM components), and prepatterning capacity.

# 3. Head connective tissue

## A. The neural crest guides head muscle formation

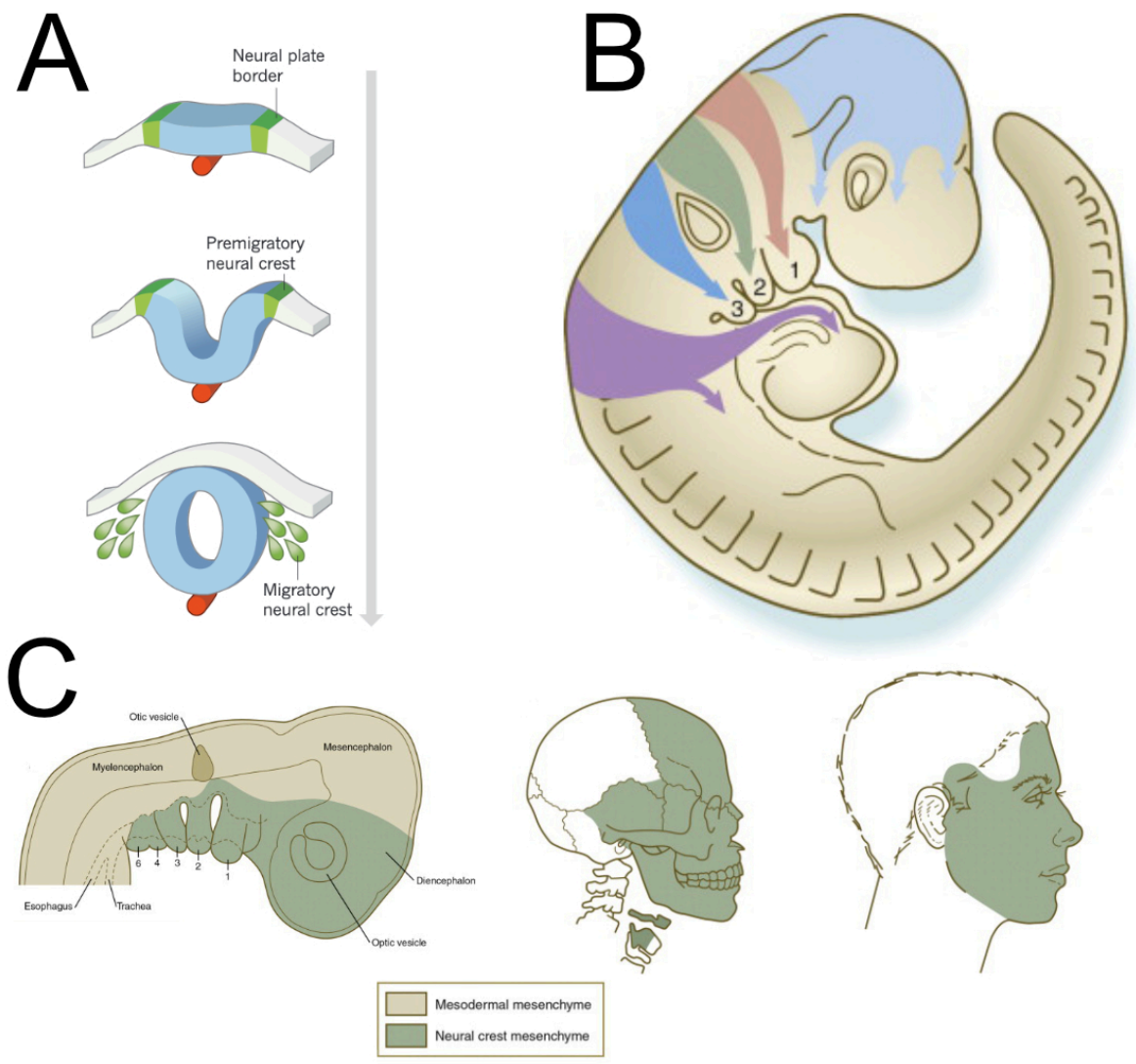



**Figure 17. Contribution of neural crest cells to the head.**

(A) Neural crest cells delaminate and migrate from the dorsal side of the neural tube during fusion (Green et al., 2015).
(B) NCCs further migrate towards the ventral side of the embryo and invade the pharyngeal and cardiac regions (Carlson, 2014).
(C) These cells give rise to most structures of the face (Carlson, 2014).

a) Origin and contribution

Neural crest cells (NCCs) are a multipotent population of cells that delaminate from the neural folds during neurulation. Following EMT, these cells migrate dorsoventrally on each side of the embryo and populate the pharyngeal arches, skull and frontonasal process. Their contribution to the head is extensive, as most bones and connective tissue of the face, jaw, bones of the middle ear (malleus, incus, and stapes), as well as the cartilages of the neck are NCC-derived (Chai and Maxson, 2006; Evans and Noden, 2006; Jiang et al., 2002; Noden, 1983; Noden and Trainor, 2005; Ziermann et al., 2018) (**Figure 17**). NCCs also give rise to the odontoblasts and pulp of teeth, sensory neurons and glia that integrate with the ectoderm-derived sensory placode neurons, pericytes, and smooth muscle surrounding the vasculature. In the trunk, NCCs give rise to melanocytes, adrenal gland cells, neurons of the dorsal root ganglia, sympathetic chain, and enteric nervous system (Douarin and Kalcheim, 1999; Vega-Lopez et al., 2017). This active migration starts around week 4 in human and E8.5 in mouse. NCCs are specific to vertebrates and play a critical role in cranial and cardiovascular development as they give rise to most of the craniofacial structures and the outflow tract of the heart. Cells derived from the neural crest express *Wnt1* during delamination and *Sox10* during subsequent migration, which conveniently allows tracing of these populations by transgenic markers (Debbache et al., 2018).

b) Role of NCC in muscle patterning

Similar to limb myogenic cells and their LPM-derived MCT, the interactions between cranial muscles and NCCs has been widely studied. Muscle patterning by NCCs was demonstrated when presumptive NCCs from PA1 were grafted into PA2 or PA3. The resulting ectopic anomalies resembled formation of a beak and eustachian tube-like depression, characteristic of 1$^{st}$ pharyngeal pouch derivatives (Noden, 1983). This striking result suggested that NCCs are pre-instructed with identity information as they propagate through the head. Additionally, patterns of muscles similar to those found in the arch of origin were observed in the donor, suggesting that NCCs can also pattern associated muscles. Similarly, transplantation experiments done between quail and duck, which have significant differences in beak shape, showed that transplanting quail NCC in duck donor embryos led to quail-like beak features and quail-like attachment sites from duck-derived muscles (Tokita and Schneider, 2009).

Studies in axolotl showed that removal of NCCs severely impacted muscle patterning, but not initial positioning (Ericsson et al., 2004). Similarly, a seminal study by Rinon et al. focused on the initiation of myogenesis and patterning defects of myogenic cells in the absence of neural crest. Consistently with the previous study, results indicated that while NCCs are dispensable for muscle initiation, they are necessary for patterning the developing muscle (Rinon et al., 2007). It was shown that NCCs act to repress WNTs and BMPs and promote proliferation, differentiation and proper orientation of muscle fibres (Rinon et al., 2007; Tzahor et al., 2003). In the extraocular context, *Pitx2* is required cell autonomously for both myogenic and neural crest cells (Evans and Gage, 2005; Zacharias et al., 2010). NCC-specific deletion of *Pitx2* leads to misalignment of EOM fibres (Evans and Gage, 2005).

Notably, a substantial part of the head is not under the regulation of *Hox* genes (Couly et al., 1998). These homeobox transcription factors are well known for determining identity of segments along the anteroposterior axis (Burke, 1999; Ehehalt et al., 2004; Jacob et al., 1975; Kieny et al., 1972). However, head morphology is regulated in part by another set of homeobox transcription factors called the *Dlx* genes (distal-less) (Depew et al., 2005; Frisdal and Trainor, 2014; Heude et al., 2010). Similar to the combinatorial "Hox code" found in the body, a "Dlx code" in the head was found to govern the identity of structural components in the proximo-distal (dorso-ventral) axis (Depew et al., 2005; Sumiyama and Tanave, 2020) (**Figure 18**). *Dlx* genes are expressed in neural crest-derived mesenchyme, and their misexpression has led to striking homeotic transformations. For example, in a *Dlx5;Dlx6* double mutant background, jaw identity is lost leading to a duplicated maxilla in the mandibular domain, with vibrissae and palatal rugae (Depew et al., 2002). Heude et al. demonstrated that NCC activity in non-skeletal elements of *Dlx5/6* was required for jaw muscle patterning (Heude et al., 2010). In addition, analysis of human patients with hemifacial macrosomia displaying normal bone development but affected jaw muscles suggests that NCC-derived MCT defects in those patients are responsible for masticatory muscle mispatterning (Heude et al., 2011).

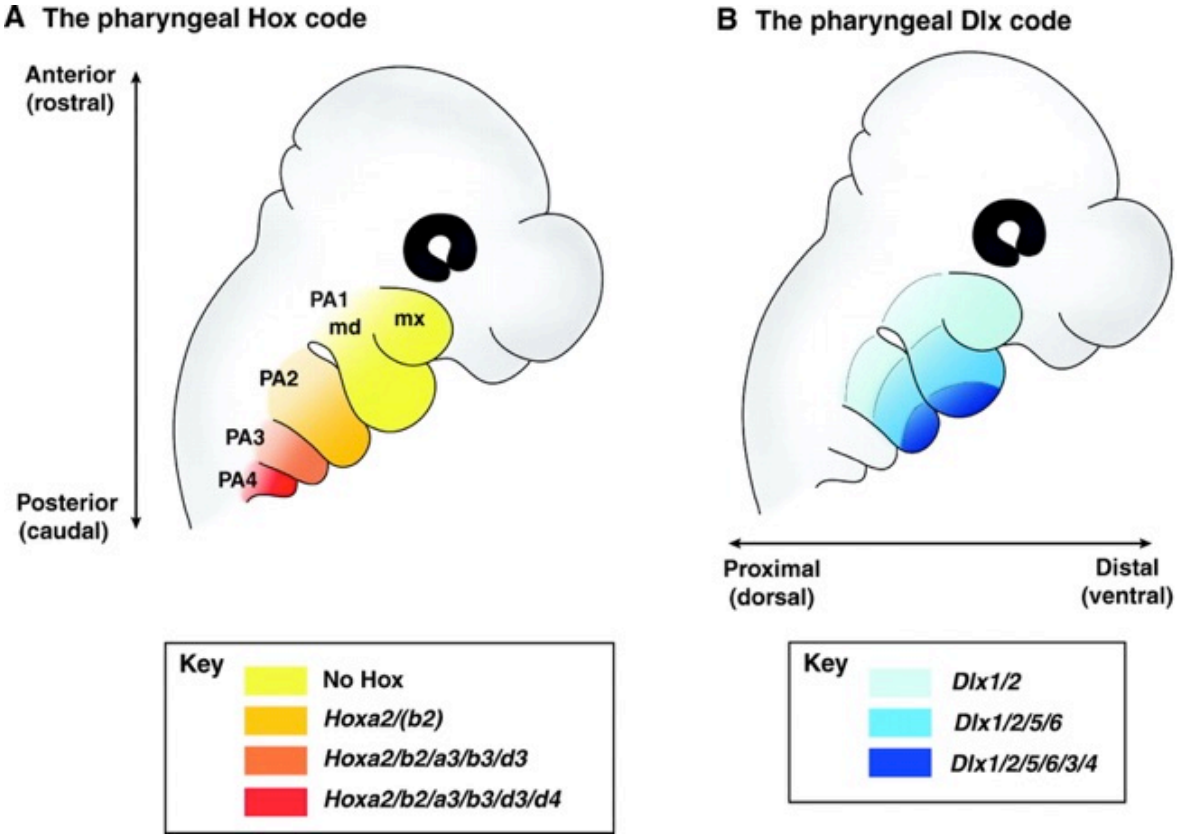



**Figure 18. Hox code and Dlx code in the developing head**
The mandibular and maxillar process of the first arch are not subject to *Hox* expression. However, the *Dlx* code determine the identity of neural crest-derived structures.
From (Minoux and Rijli, 2010)

Tongue muscles have proven to be another valuable model for the study of tissue-tissue interactions (Cobourne et al., 2018; Han et al., 2012; Han et al., 2014; Iwata et al., 2013; Millington et al., 2017). These muscles derive from the occipital somites, and populate the neural-crest-filled tongue primordium following an anterior migration along the hypoglossal chord (Han et al., 2012). In addition, the tongue length spans from PA1 to PA4, thus potentially receiving positional cues at different anteroposterior positions (Parada et al., 2012). As indicated above, migration of tongue progenitors relies on MET/HGF signaling, although it is not clear if HGF is only expressed along the hypoglossal chord or if NCCs are also a source of chemotractant molecules, similarly to the LPM for limb muscles (Dietrich et al., 1999). Additionally, NCC-specific cilia-dependent GLI activity is necessary for migration and survival of tongue progenitors (Millington et al., 2017). In the tongue bud, cell autonomous and non-

autonomous TGFβ signaling promotes tongue muscle proliferation and differentiation (Han et al., 2012; Han et al., 2014; Hosokawa et al., 2010; Iwata et al., 2013).

Finally, analysis of parrots, which have evolved a second jaw, suggests that the emergence of new muscles is associated with appearance of novel NCC-derived structures (Tokita et al., 2013).

Taken together, these studies indicate that NCC is the major constituent of cranial muscle connective tissue, and it is instrumental in guiding head muscle development.

In summary, skeletal muscle formation critically depends on surrounding stromal cells for proliferative and positional cues

# Chapter V

# When intrinsic meets extrinsic

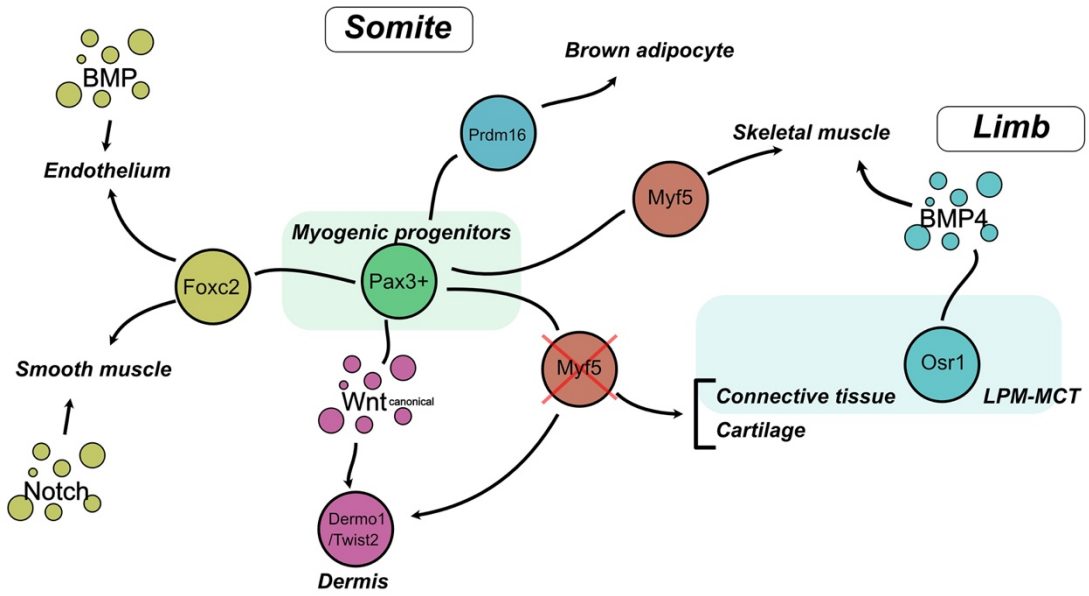



**Figure 19. Promiscuity of myogenic and non-myogenic fates in somitic development.**

Upstream myogenic progenitors give rise to a variety of cell types in embryos. Several studies have described a fine balancing act between myogenic and non-myogenic fates. Additionally, limb LPM connective tissue was reported to merge with myogenic cells at the myotendinous junction.

## 1. In the embryo

Trunk, limb and some cranial muscles derive from multipotent progenitors in the somite, which give rise to dermis, endothelium, smooth muscle, brown adipocytes, cartilage, bone, tendon and connective tissues (Christ et al., 2007) (**Figure 19**). Retrospective clonal analysis showed that myogenic progenitors and aortic smooth muscle share a common ancestry, and *Pax3* lineage tracing showed that cells of the blood vessel wall originate from somitic Pax3+ cells (Esner et al., 2006). The regulators ensuring a proper balance between the production of myogenic and non-myogenic cells have been partially elucidated (Buckingham and Relaix, 2007). For example, vascular contribution from Pax3+ cells, which operates first, was reported to be controlled by reciprocal inhibition between *Pax3* and *Foxc2* (Lagha et al., 2009). BMP and NOTCH signaling promote endothelium and smooth muscle production respectively (Ben-Yair and Kalcheim, 2008). *Prdm16* is a molecular switch between myogenic and brown adipocyte lineages (Seale et al., 2008). Surprisingly, lineage-traced *Myf5*-expressing cells from

the central dermomyotome give rise to brown adipocytes and dermis in the embryo (Jimenez et al., 2006; Shan et al., 2013). βcatenin (effector of canonical Wnt signaling), was found to be necessary and sufficient for the formation of dermis cells from En1+ dorsal dermomyotomal cells (Atit et al., 2006).

Interestingly, the *Myf5*$^{nlacZ/nlacZ}$ mouse line allows to interrogate the fate of progenitor cells in which *Myf5* expression is depleted (Tajbakhsh et al., 1996a; Tajbakhsh et al., 1996b). In homozygous mutants, the cells that would normally express *Myf5* in the intercostal muscles change fate and integrate into the ribs as chondrocytes (Tajbakhsh et al., 1996a). Dorsal βgal+ cells express dermis markers and ventral cells express Scx+, a connective tissue marker (Tajbakhsh et al., 1996a). Additonally, the *Myf5*-lineage has been found to give rise to chondrocytes in the ribs, by Cre-mediated recombination (Haldar et al., 2008).

Human limb muscle development was recently investigated by scRNAseq (Korrapati et al., 2019). Intriguingly, a substantial portion of skeletal muscle progenitors had a lower expression of myogenic markers and higher expression of connective tissue markers (identified as "skeletal muscle mesenchyme" by the authors). Similarly, 2 simultaneous studies currently under review suggest that fibroblastic cells directly fuse with the developing myofibre to promote myotendinous attachment (Lima et al., 2020; Yaseen-Badarneh et al., 2020). Specifically, in one study, *Osr1*-expressing LPM-derived connective tissue cells were reported to contribute to chick and mouse limb striated muscle by fusing with the myofibre at the myotendinous junction, following a BMP-induced fibroblast-to-myoblast conversion (Lima et al., 2020).

## 2. In the adult

Following muscle injury and repair, both PICs and mesangioblasts were shown to contribute to myofibres (Dellavalle et al., 2007; Mitchell et al., 2010). These cells are marked by the expression of PW1, which was shown to confer myogenic potential and the ability to cross the blood vessel wall (Bonfanti et al., 2015). Other interstitial cells, negative for *Pax7* and expressing *Twist2* (a marker for dermal lineage), was shown to participate to type IIb (fast glycolytic fibres) myofibres both in homeostatic and regenerating conditions (Liu et al., 2017). Recent work using scRNAseq identified a population of cells expressing smooth muscle and mesenchymal markers (Itga7+/Vcam1-, named "SMMCs") that have a myogenic potential in vitro (Giordani et al., 2019).

Another study where stem cells were cultured in vitro with aged mouse serum showed that about 10% of MuSCs acquire a fibrogenic phenotype, similarly to aged mouse stem cells, due to the activation of Wnt signaling (Brack et al., 2007). Other teams have reported fibro-adipogenic potential from satellite cell-derived myoblast cultures or single myofibre cultures (Asakura et al., 2001; Shefer et al., 2004; Wada et al., 2002). However, some of these reports have been challenged due to concerns regarding contamination or in vitro bias (Uezumi et al., 2014). Other studies have characterised MuSCs as unipotent and proposed that fibrosis arises uniquely from mesenchymal progenitors during impaired regeneration (Uezumi et al., 2010; Uezumi et al., 2011; Uezumi et al., 2014). Stem cells have been shown to convert to brown adipocyte under cold condition, and that this conversion was repressed in normal state by miR133, under the control of *Myod* (Yin et al., 2013). This switch is regulated by *Prdm16*, a transcription factor promoting *Pparγ* (Peroxisome Proliferator Activated Receptor Gamma) expression, leading to brown adipocyte fate (Seale et al., 2008; Yin et al., 2013).

## 3. A potential bipotent paradigm: medial CPM

Several studies have suggested that the extent of contribution of neural crest to cranial mesenchyme needs reconsideration (Adachi et al., 2020; Comai et al., 2020; Grimaldi et al., 2015; Heude et al., 2018; Tabler et al., 2017) (**Figure 20**). Although a great majority of the head mesenchyme derives from neural crest, several studies have demonstrated that some subdomains (often located dorsomedially) of cranial MCT is of mesodermal origin. For example, laryngeal cartilages that comprise the thyroid, cricoid and arytenoid cartilages have been considered for some time to be all neural crest derived, like most of the upper hard tissues (Matsuoka et al., 2005). However, recent studies have shown that the medial part of the thyroid cartilage is derived from mesoderm and that the cricoid and arytenoid cartilage surrounding the larynx are also mesodermal-derived in both mouse and chick (Evans and Noden, 2006; Heude et al., 2018; Tabler et al., 2017). Notably, the extraocular muscles, acromiotrapezius, sternocleidomastoid, pharyngeal constrictor, and laryngeal muscles are all partly embedded in mesoderm (Adachi et al., 2020; Comai et al., 2020; Grimaldi et al., 2015; Heude et al., 2018; Tabler et al., 2017) (**Figure 20**). Therefore, CPM gives rise to both mesenchymal and muscular tissues, as is the case for somitic mesoderm in primaxial muscles.

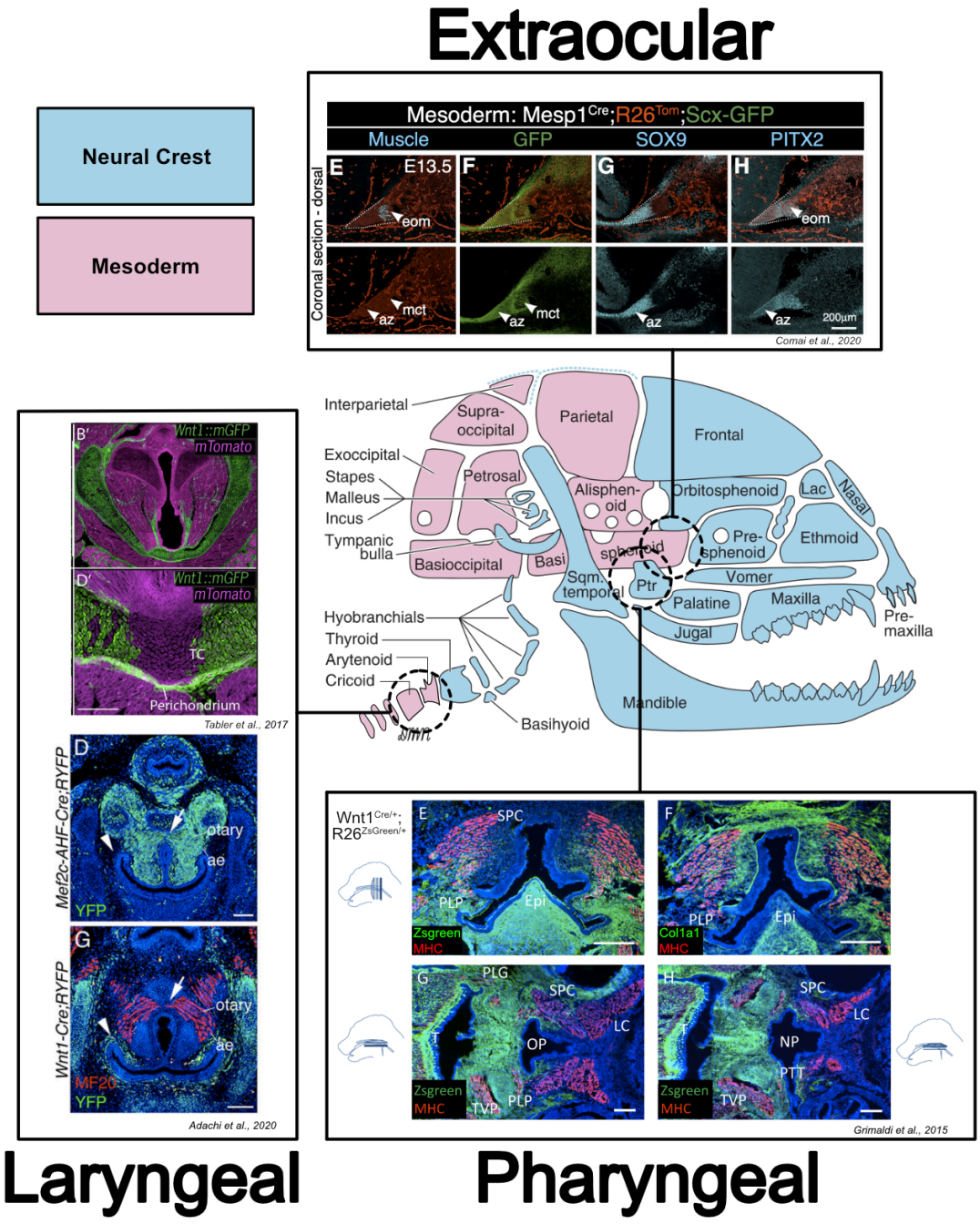



**Figure 20. A new frontier for neural crest: CPM connective tissue.**
Cumulative evidence on the prominent role of cardiopharyngeal mesoderm connective tissue. Extraocular, laryngeal and pharyngeal muscles are associated with mesodermal (and potentially all CPM) connective tissue. Adapted from (Adachi et al., 2020; Comai et al., 2020; Grimaldi et al., 2015; Tabler et al., 2017). Illustration adapted from (Noden and Trainor, 2005).

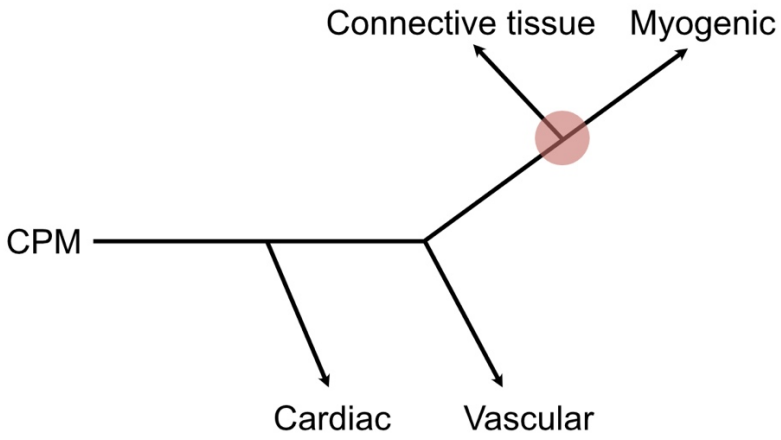



**Figure 21. Presumed branchpoint for CPM bipotent progenitors**

CPM gives rise to multiple tissues in the head, including muscle associated connective tissues. The cell fate determinants impinging on this decision are currently unknown.

Interestingly, detailed analyses of *Tbx1* heterozygous mutants revealed patterning defects of somitic and non-somitic muscles associated with CPM-derived connective tissue, indicating that CPM-derived MCT also possess a muscle-patterning role (Adachi et al., 2020). In addition to cardiac muscles, CPM also gives rise to the pharyngeal arch arteries (smooth muscle and endothelium), a transient vascular structure that supply each pharyngeal arch and that will give rise to all anterior vasculature following coordinated expansion and selective degeneration (Hiruma et al., 2002; Whitesell et al., 2019). However, in contrast to somitic mesoderm, CPM does not appear to be overtly segmented into myogenic and non-myogenic compartments. Although it has been shown to be regionalized in terms of expression patterns in the chick into an anterior- and posterior pole, no clear morphological boundary was established for its cardiac, cranial and vascular components (Bothe and Dietrich, 2006; Sambasivan et al., 2011b). Further, these expression patterns were not found in zebrafish (Wang et al., 2019). In this context, it remains obscure how in the absence of neural crest, multipotent progenitors decide which fate to adopt, spatially and temporally (**Figure 21**).

# Chapter VI

# Muscle memory

It is now clear that distinct gene regulatory networks are required for cell-autonomous myogenic initiation during skeletal muscle development. On the other hand, MCT appears to have a critical role in muscle patterning, with limited impact on muscle initiation during development (Nassari et al., 2017; Rinon et al., 2007; Sefton and Kardon, 2019). In adult muscle, a permanent cross-talk between myogenic and non-myogenic cells operates, and following muscle injury and repair to maintain and regenerate skeletal muscle (Biferali et al., 2019). In that regard, muscles are heterogeneous and have varying degrees of repair efficiency, proliferative capacities, and possess various contractile features (Kelly and Buckingham, 2000; Ono et al., 2009; Sambasivan et al., 2009; Schoser and Pongratz, 2006; Stuelsatz et al., 2015).

An obvious question to ask is whether these developmental programs are still active in adult myofibres or in MuSCs. For instance, in the context of fly neurogenesis, multiple neuron subtypes are generated by the expression of unique transcription factor networks. The identity of these subtypes needs to be assured; otherwise critical communication pathways could be impaired. Eade and colleagues reported that unique TF networks were actively maintained in each subtype throughout life, and suggested that this safeguards cell identities (Eade et al., 2012). In keeping with this notion, "cellular memory modules" were reported to couple with Polycomb (repressive) and Trithorax (activating), potent set of genes known to maintain homeotic genes (master genes of body plan) in the embryo (Francis and Kingston, 2001). These modules act as epigenetic inheritance mechanisms during *Drosophila* imaginal wing morphogenesis (Maurange and Paro, 2002). Could similar memory mechanisms apply to skeletal muscle? And could they account for some of the muscle heterogeneity reported?

To answer these questions, we must first assess the presence and the potency of such intrinsic marks in defining muscle identity during development. Second, we need to assess the extent to which these developmental regulatory nodes are conserved in adult muscle. Finally, we would need to investigate the mechanisms regulated by these factors that may impinge of cellular behaviours and phenotypic outcomes.

# 1. Muscle identity: intrinsic or extrinsic?

A long-standing question in stem cell biology is how much do intrinsic parameters dictate the unique features of progenitors (i.e. their identity), compared to environmental cues. It is clear that both sources influence cell behaviours, but their relative contributions might vary between cell types (Davies et al., 2018). This problem has been broadly discussed in the context of muscle development with discordant conclusions (Alvares et al., 2003). Somites appear to be morphologically identical during early embryogenesis, but they will form a wide variety of anatomically distinct epaxial, hypaxial and limb muscles along the body axis. This involves a tight control of positional information that will establish the identity of progenitors, while repressing that of adjacent somites.

The axial skeleton is patterned by *Hox* genes, where combinatorial expression patterns establish the "Hox code" (Burke, 1999). Transplantation experiments demonstrated that non-myogenic somitic precursors (dermal and sclerotomal) carry their identity through the expression of Hox genes (Ehehalt et al., 2004; Jacob et al., 1975; Kieny et al., 1972). In contrast, when interlimb somites were grafted at the limb level, appendicular muscles were still observed, suggesting that somitic myogenic progenitors are mostly naïve and they respond to environmental cues (Christ et al., 1977). However, several other studies have challenged this view and suggested that positional identity is present in somitic muscle progenitors (Alvares et al., 2003; Donoghue et al., 1992; Grieshammer et al., 1992; Murakami and Nakamura, 1991). One possible explanation for that dissonance is that FGF signals coming from the appendicular region are able to promote the expression of *Lbx1* in non-migratory precursors (Alvares et al., 2003). This would suggest that observations made in the context of the limb might not apply to other levels of the body.

During rib development, studies have shown that *Myf5* and *Mrf4* expression in myogenic cells, under the control of *Hox* genes, induce the expression of *Pdgf* and *Fgf* ligands to promote the development of the ribs (Vinagre et al., 2010). This work indicates that in the context of rib formation, intrinsic *Hox* positional information is integrated within muscle progenitors, and translated into supportive signaling for surrounding skeletal structures.

## 2. Role for developmental regulators in adult

Recent work under review showed that *Hoxa10* that is expressed in embryonic limb muscles, is maintained through adulthood and is necessary for sustained proliferative potential of MuSCs (Yoshioka et al., 2020). This muscle-specific activity is associated with specific methylation patterns on the *Hoxa* cluster. Disruption of *Hoxa10* in MuSCs leads to genomic instability and collapse of mitotic capacity cell-autonomously.

In other studies, expression data by RT-qPCR showed that some transcription factors of EOM and PA1 developmental programs are found in the adult, while others are lost (Sambasivan et al., 2009). Specifically, *Pitx1/2* and *Alx4* of the EOM were retained when compared with the limb-derived satellite cells, whereas *Tbx1* from PA1 was lost (Sambasivan et al., 2009). Interestingly, *Tbx1* is expressed in the adult limb after the onset of myogenic commitment, while it is rapidly downregulated in pharyngeal arches and its mutation results in only partial alterations of PA1 muscle development (Dastjerdi et al., 2007; Grifone et al., 2008; Nathan et al., 2008). Similarly, *Lbx1* is significantly upregulated in the adult limb compared with PA1, suggesting active conservation of some genes in migratory abaxial muscles. In contrast, when MuSCs from EOM were cultured in vitro and differentiated, they were unable to recapitulate the adult muscle fibre phenotype, which can be distinguished by the unique expression of *Myh13* and *Myh15* (Sambasivan et al., 2009). In keeping with these observations, when grafted into an injured limb, EOM could contribute to endogenous regenerating fibres efficiently, but no expression of *Myh13* or *Myh15* was detected (Sambasivan et al., 2009).

A recent study from our group characterised the transcriptomic profile of EOM satellite cells post-graft in the limb TA muscle. After self-renewal, EOM satellite cells were shown to express limb-specific *Hox* genes, while maintaining 10% of their initial genetic signature. This suggests that a significant epigenetic resistance of MuSC-specific genes is at play, even in following transplantation and exposure to a heterotopic niche (Evano et al., 2020).

Upstream myogenic regulators such as *Pax3*/*Pax7*, *Tbx1* and *Pitx2* actively contribute to the formation of other tissues, sometimes in close proximity to muscle (Gage et al., 1999b; Huynh et al., 2007; Mansouri and Gruss, 1998; Meilhac and Buckingham, 2018). *Pax3*/*Pax7* are expressed in regions of the nervous system and in neural crest cells. This suggests that in the case of the tongue, myogenic progenitors and their associated NCC-derived MCT both

derive from *Pax3*-expressing progenitors, but they represent 2-independent lineages (Blake and Ziman, 2014; Monsoro-Burq, 2015). *Tbx1* on the other hand is expressed in CPM-derived endothelial cells, which suggests that *Tbx1* alone cannot instruct myogenic fate (Dastjerdi et al., 2007). Further, as the somite-derived tongue muscles mature, they express *Tbx1*, which is required for myogenesis (Okano et al., 2008). Therefore, upstream factors can be coopted and redeployed by various lineages, thus indicating that in the absence of robust identity determinants such as *Hox* genes, it is plausible that cell identity is maintained by a combination of transcription factors in the head, rather than by a single determinant.

Taken together, one can envisage a model where developmental extrinsic and intrinsic cues would impinge on myogenic progenitors based on their anatomical location. Different gene regulatory networks operating to initiate myogenesis would have to be maintained through lineage progression and growth to ensure robust cell identities. In this framework, certain key regulators could be maintained as genetic modules through adulthood in muscle stem cells, and thus contribute to their apparent phenotypic variability (**Figure 22**).

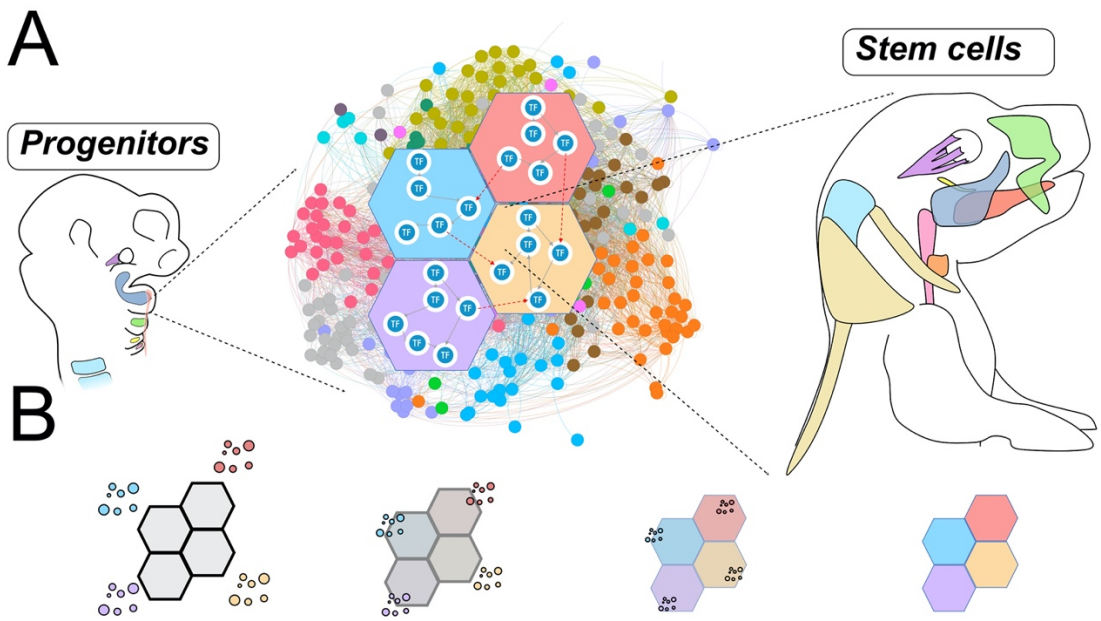



**Figure 22. Hypothetical model of muscle memory.**

A) Schematic representation of the model. Distinct myogenic progenitors initiate the expression of various transcription factor regulatory networks during embryonic myogenesis. These genetic modules are maintained through adulthood and coopted by MuSCs, contributing to their diversity.
B) Scheme of extrinsic and intrinsic cues establishing these various modules.

# RESULTS

# Part 1

# Bipotent progenitors give rise to myogenic and connective tissue in the developing embryo

# Bipotent progenitors give rise to myogenic and connective tissue in the developing embryo

Contributors: Alexandre Grimaldi[1,2], Glenda Comai[1,2], Sébastien Mella[1,2], Shahragim Tajbakhsh[1,2]

[1]Stem Cells & Development Unit, 25 rue du Dr. Roux, Institut Pasteur, 75015 Paris, France
[2]UMR CNRS 3738, Institut Pasteur, Paris, France

## ABSTRACT

The interplay between skeletal muscle and associated stromal cells assures the coordinated development of muscle functional units. How these distinct cell populations arise from embryonic precursor cells and cooperate is poorly understood. Stromal cell populations can be of mesodermal or neural crest origin, whereas skeletal muscles are derived exclusively from mesoderm, thereby raising the question of whether these cells can have coherent embryological origins. Here, we used multiple lineage-traced single cell RNAseq datasets combined with anatomical cell positioning and gene expression analysis to investigate the emergence and patterning of skeletal muscle and stromal cell populations within distinct cranial muscle groups in mouse. We identify Myf5$^{+}$ bipotent progenitors giving rise to muscle and associated connective tissue with complementary receptor-ligand communication streams. We present a novel computational pipeline that unveiled a common gene regulatory network controlling connective tissue fate at multiple stages of development and in diverse muscle groups. This method can be applied to other biological systems and might be of interest to identify key regulators of transitory states. Our study shows that progenitors seemingly destined to myogenic fate redirect to connective tissue to give rise to muscle-associated structural components. Following this bifurcation, muscle and connective tissue cells acquire complementary signalling features and maintain spatial proximity. Collectively, our results reveal unexpected adaptability of myogenic ancestral cells to redirect their fate during morphogenesis of complex structures.

## INTRODUCTION

Throughout the lifetime of an organism, stromal cells that are associated with skeletal muscles play critical roles in providing structural support and molecular cues (Ahmad et al., 2020; Biferali et al., 2019; Kardon et al., 2003; Nassari et al., 2017; Sefton and Kardon, 2019). The majority of muscle-associated connective tissues in the head is derived from cranial neural crest cells (NCCs) (Douarin and Kalcheim, 1999; Grenier et al., 2009; Heude et al., 2018; Noden, 1983; Noden and Trainor, 2005). The extend of this contribution was recently redefined in extraocular (EOM), laryngeal and pharyngeal muscles (Adachi et al., 2012; Comai et al., 2020; Grimaldi et al., 2015; Heude et al., 2018; Noden and Epstein, 2010; Tabler et al., 2017). A common feature of these muscles is a continuous mesoderm-derived dorsal component, whereas the rest of the muscle mass is embedded in neural crest-derived mesenchyme. It remains unclear how the coordinated emergence of myogenic and connective tissue cell types takes place during development, and what is the significance of connective tissue origin. Although some information is emerging regarding these populations in trunk and limb muscles, these relationships are less well understood for craniofacial muscles where most studies have focused on the neural crest (Grenier et al., 2009; Helmbacher and Stricker, 2020; Nassari et al., 2017; Noden and Trainor, 2005; Sefton and Kardon, 2019).

Cranial mesodermal progenitors give rise to at least 4 distinct lineages: bone/cartilage, connective tissue/tendon, vasculature, and skeletal muscle Noden and Trainor, 2005; Bildsoe et al., 2013; Adachi et al., 2020; Heude et al. 2018; Chai and Maxson, 2006). Similarly, along the trunk axis, paraxial mesoderm gives rise to skeletal muscles and associated connective tissues (Burke and Nowicki, 2003; Deries et al., 2010; Saberi et al., 2017). Upon signals emanating from the neural tube, notochord, ectoderm and lateral plate, the dermomyotome (the dorsal portion of the somite) undergoes an epithelial-to-mesenchymal transition and gives rise to several cell types (Cossu et al., 1996; Dietrich et al., 1998; Ikeya and Takada, 1998; Munsterberg et al., 1995; Pourquié et al., 1996; Christ et al., 2007; Ben-Yair et al., 2008). However, unlike the somites, head mesoderm is unsegmented, raising the question of how spatiotemporal control of these cellular identities is established.

Cardiopharyngeal mesoderm constitutes the major portion of cranial mesoderm and it has cardiovascular potential, which manifests in the embryo as regions of clonally related cardiac and craniofacial muscles (Diogo et al., 2015; Lescroart et al., 2010; Lescroart et al., 2015; Swedlund and Lescroart, 2019). This skeletal muscle/cardiac branchpoint has been the subject of intense investigation in several model organisms including ascidians, which possess an ancestral cardiopharyngeal mesoderm population arising from bipotent cells, in avians and mouse (Wang et

al., 2019b). However, the issue of connective tissue divergence from this lineage has not been addressed.

Here, we employed unbiased and lineage-restricted single-cell transcriptomics using multiple transgenic mouse lines, in situ labelling, and new analytical methods, and show that bipotent progenitors expressing the muscle determination gene *Myf5* give rise to skeletal muscle and anatomically associated connective tissue in distinct muscle groups spatiotemporally. Notably, this property was restricted to muscles with only partial contribution from NCCs suggesting that in the absence of NCCs, somite-derived or cranial-derived mesoderm emerges as a source of connective tissue. This transition is characterised by a complementarity of tyrosine kinase receptor signalling between muscle and non-muscle cells, as well as distinct regulatory modules. Therefore, our work exploits new analytical methods to unveil key regulators of transitory states and further defines the intimate intercellular communication network between these two cell types.

## RESULTS

### Myogenic and non-myogenic mesodermal populations coexist within distinct cranial lineages

Somitic (*Pax3*-dependent) and cranial (*Isl1/Tbx1/Pitx2-dependent*) mesoderm give rise to diverse cell types including those of the musculoskeletal system (Figure 1A). However, little is known about the spatiotemporal control of skeletal muscle formation within these mesodermal lineages. To clarify the regulatory cascades underlying these distinct mesodermal programs, we performed scRNAseq analysis using a broad cranial mesoderm lineage-tracing strategy in *Mesp1*$^{Cre/+}$;*R26*$^{mTmG/+}$ mouse embryos at E10.5. Expectedly, a large portion of the cells obtained corresponded to adipogenic, chondrogenic, sclerotomal, endothelial, and cardiovascular cells (Figure 1B, Figure S1A-B). Using *Pax3*, *Pitx2*, *Isl1*, *Myf5* and *Myod* we identified myogenic progenitors that contribute to craniofacial muscles (Figure1C, Figure S1-B).

We thus filtered the dataset for these cells which allowed them to clearly separate based on their origin and anatomical location (Figure1D-E, Figure S1C). Surprisingly, about half of the cells exhibited a connective tissue signature, including a strong bias towards *Prrx1*, a marker for lateral plate mesoderm, and *Col1a1* (Figure 1E-F). Strikingly, the expression of Pdgfra, (a well-defined marker of stromal cells (Farahani and Xaymardan, 2015)), was robustly anticorrelated with the expression of its ligand Pdgfa, and associated with non-myogenic genes. Conversely, Pdgfa expression correlated with a myogenic cell state (Figure1F, Figure S1D). Therefore, we identified anatomically distinct muscle and closely-associated connective tissue progenitors.

### Transcriptional trajectories reveal a myogenic to non-myogenic transition

To better understand the short-term lineage relationship between myogenic and non-myogenic cells, we exploited the unspliced and spliced variants of our scRNAseq data, and computed the RNA velocity in each cell, using a recently published tool (Bergen et al., 2020) (Figure 2). RNA velocity interrogates the relative abundance of unspliced and spliced gene variants, which depends on the rates of transcription, degradation, and splicing to infer directional trajectories (Bergen et al., 2020; Manno et al., 2018). Surprisingly, RNA velocity-inferred trajectories indicated that Myf5+ cells from the myogenic compartment contributed to non-myogenic cells (Figure 2A-C). These calculations were based on gene- and cluster-specific dynamics, which yields higher accuracy than the initially described RNA velocity method, while providing quantitative metrics for quality control (Figure S2B and Methods).

Latent time refers to a local advancement of a cell through a lineage. Applied to our data, it suggested that some Myf5+ cells exhibited characteristics of a more upstream cell state compared to the non-myogenic cells, particularly in myogenic cells originating from somites (Figure 2C). A powerful feature of this method, in addition to being quantitative, is the ability to infer “driver genes” that are responsible for most of the calculated RNA velocity, hence actively transcribed, or repressed (Bergen et al., 2020). Therefore, these genes can identify transitory states underlying cell fate decisions. We used this approach to uncover the driver genes that were responsible for the velocity found in anterior somites, and that contribute to some head and neck muscles (Figure 2D, Table1). Although the RNA velocity confidence index was found to be higher in other regions examined, this domain displayed the most consistent directionality (Figure S2B).

Cell cycle status constitutes a potential bias in scRNAseq data, especially with heterogeneous populations undergoing cellular expansion, commitment and differentiation (McDavid et al., 2016). To eliminate this potential bias, cell cycle genes were consistently regressed out during preprocessing and directional trajectories were overlaid with cell cycle phase visualization for comparisons (Figure S2A). This analysis showed that trajectories observed in the anterior somites could not be explained by cell cycle alone (Figure S2A-B). Top transcribed driver genes included *Cdh6, Ednra, Ebf2, Meis1, Ptn, Nfia, Sim1, Pdgfra, Prrx1* and *Fli1,* that are associated with adipogenesis, fibroblast development, and vasculature formation (Figure 2E) (Jimenez et al., 2006; Gulyaeva et al., 2019; Wang et al., 2018).

RNA velocity refers to the relative abundance of unspliced and spliced variants, which depend on the rates of transcriptional, degradation and splicing (Manno et al., 2018). When unspliced variants are more abundant than spliced variants for a given gene, this gene is being actively transcribed (if splicing and degradation rates are considered constant). Conversely, when spliced variants are more abundant than unspliced variants, the gene is being repressed. The rates of degradation, splicing and transcription constitute parameters that can be inferred by specific algorithms (Bergen

et al., 2020; Manno et al., 2018). Hence, variant abundance of genes can be visualised as "phase portraits", which display unspliced (Y-axis) and spliced (X-axis) transcript counts for a given gene in each cell, overlaid with the inferred model of transcriptomic dynamics for that gene. We noted that *Pdgfra* appeared as a driver gene in the anterior somite, and found it to be activated along the inferred trajectory, whereas Pdgfa expression decreased rapidly (Figure 2F).

Taken together, RNA velocity analysis indicated that Myf5+/Pdgfa+ cells shifted towards a non-myogenic fate, by downregulating these 2 markers and activating Pdgfra expression. Although *Myf5*-expressing cells were reported to contribute to adipocytes and dermal cells arising from the somites (Jimenez et al., 2006; Shan et al., 2013), here we report this phenomenon taking place in pharyngeal arch 2, but not pharyngeal arch 1, with high confidence, albeit the cell number was relatively low (96 and 182 cells respectively, and 432 for the anterior somites) (Figure 2A-B, Figure S2A). This observation would be coherent with the full contribution of neural crest to the 1$^{st}$ arch connective tissue, suggesting that mesoderm-derived connective tissue is not required there.

**_Myf5_-derived lineage contributes to connective tissue cells in the absence of neural crest**

Given the low number of cells in from the EOM and pharyngeal arch clusters from the E10.5 dataset, it was unclear if the emergence of *Myf5*-derived non-myogenic cells could be found also in these regions. To assess the extent of contribution of *Myf5*-derived Pdgfra+ non-myogenic cells to different muscle groups, we examined the EOM, caudal arch muscles, and anterior somites at E14.5. We combined a *Myf5*-lineage reporter mouse with a reporter for non-myogenic cells, *Myf5$^{Cre/+}$; R26$^{TdTomato/+}$; Pdgfra$^{H2BGFP/+}$* and immunostained for the commitment and differentiation-specific myogenic markers Myod and Myog, respectively (Figure 3). Notably, we observed double-positive (muscle and connective tissue) cells in the EOM, laryngeal and anterior somite muscles, but not in masseter and tongue muscles (Figure 3A-E). In agreement with our scRNAseq velocity analysis, this observation suggests that the *Myf5* lineage contributes to muscle-associated connective tissue in territories where it arises from a mesodermal, but not neural crest, origin (summarized in Figure 3F).

To determine the spatiotemporal contribution of these connective tissue cells to muscles, we assessed by histology the extent and nature of *Myf5*-derived connective tissue in the head at birth (Figure S3A-B), and in the EOM at P30 using *Myf5$^{Cre/+}$; R26$^{mTmG/+}$* mice. We immunostained with Sox9 (expressed in cartilage (Lefebvre et al., 2019), and at lower levels in various connective tissues), and Pax7 (muscle stem cell marker (Zammit et al., 2006b)) and found that Myf5-derived cells contributed to various connective tissues. This contribution included adipogenic regions from the anterior somites and the EOM, tenocytes at the medial tendon attachment in the EOM, chondrocytes and perichondrium in the skull base and sphenoid primordium, close to the EOM medial attachment (Figure S3C-D). Further, coronal sections of the EOM at P30 showed a

contribution of *Myf5*-derived cells to intraconal adipose tissue, as well as interstitial cells intermingled with muscle fibres (Figure S3C-D). These observations suggest that *Myf5*-derived connective tissue cells are found in multiple regions of the head and neck, and could play a long-lasting role in establishing the EOM niche.

**Myf5 contribution to connective tissue is sustained through muscle initiation**

Although we identified *Myf5*-derived non-myogenic cells in the anterior somites at E10.5 by scRNAseq, it was not clear if this population was self-sustaining or continuously generated throughout development. In addition, given that cranial myogenesis is asynchronous (Nogueira et al., 2015), we could not confidently assess other regions at this timepoint owing to the lower number of cells obtained. To address this issue, we performed 3 more scRNAseq experiments ranging from E11.5 to E14.5, using historical *Myf5* cell lineage tracing ($Myf5^{Cre/+}; R26^{mTmG/+}$) and contemporary *Myf5* labelling ($Myf5^{GFP\text{-}P/+}$; Figure 4, Figures S2-3). In accordance with the mesodermal lineage tracing, we observed a similar phenomenon in all 3 datasets, where cells that appeared to belong to muscle anlagen of EOM, somitic and caudal arches progressed towards a non-myogenic state (Figure 4A'-C'). Once again, this progression appeared to be mostly independent of the cell cycle, indicating that these trajectories most likely reflect variations in cell identity (Figure S2A). However, it was not clear if these non-myogenic cells could represent various cell types depending on their anatomical location and stage, including cartilage, adipocytes, fibroblasts, dermis and pericytes. Therefore, we performed a gene set enrichment analysis network analysis combining the differentially expressed genes of non-myogenic clusters of all stages. We found that all stages contributed equally to "GO Molecular Function" and "Reactome pathways" terms in spite of relatively diverse gene expression signatures (Figure S4) (Bindea et al., 2009). This finding suggests that these non-myogenic cells are relatively homogeneous in gene signatures throughout cranial muscles when they emerge from common progenitors. Highly significant terms hinted at a myogenic-supporting role, providing muscle progenitors with extracellular matrix components, and contributing to neuronal guidance (Figure 4E). Among these terms, Pdgf signalling and receptor kinase activity implied that these cells might interact with the juxtaposed myogenic cells.

**Myf5-derived myogenic and non-myogenic cells can maintain a molecular crosstalk after they split from common progenitors**

To investigate potential cell-cell communication between myogenic and non-myogenic cells, we examined the anatomical proximity and the specific signalling complementarity of the *Myf5*-derived cell population. We took advantage of our $Myf5^{Cre/+};R26^{mTmG/+}$ E11.5 dataset, and focused on the EOM region which was clearly identifiable as an anatomically distinct cluster, based on the co-expression of *Pitx2* and *Alx4* (Bothe and Dietrich, 2006) (Figure 5A). Here, RNA velocity revealed a strong bi-directional cell-fate, myogenic or non-myogenic (Figure 5A).

Using neural crest and mesodermal lineage tracing coupled with a contemporary *Myf5* reporter (neural crest: *Wnt1*$^{Cre/+}$;*R26*$^{mTmG/+}$;*Myf5*$^{nLacZ/+}$; mesodermal: *Mesp1*$^{Cre/+}$;*R26*$^{mTmG/+}$;*Myf5*$^{nLacZ/+}$), we could not identify neural-crest derived *Myf5* expressing cells (nlacZ+) in the EOM at E13.5 (Figure S6). In agreement with the E10.5 data, the EOM at E11.5 presented a strong dichotomy in Pdgfa and Pdfgra between myogenic and non-myogenic cells. Immunostaining using Myod and Myog antibodies on *Myf5*$^{Cre/+}$;*R26*$^{TdTomato/+}$;*Pdgfra*$^{H2BGFP/+}$ embryos at E12.5 confirmed that the EOM contained *Myf5*-derived myogenic and non-myogenic cells in close proximity, preferentially at the muscle origin at the base of the skull (Figure S7). Thus, we interrogated the relative proximity of these cells once the muscle masses are well individuated, by performing high-resolution in situ RNAscope with *Pdgfa* and *Pdgfra* on *Myf5*$^{Cre/+}$;*R26*$^{mTmG/+}$ fetuses at E14.5 (Figure 5C-D). We observed a complementary pattern of *Pdgfa* and *Pdgfra* transcripts within *Myf5*-derived (membrane GFP+) cells, indicating that these cells could potentially maintain signalling crosstalk at least up to fetal stages in the EOM.

Gene set enrichment analysis of EOM myogenic and non-myogenic driver genes revealed that transmembrane receptor protein kinase and SMAD activity are shared terms between the 2 clusters, indicating that specific complementary signalling networks could be actively maintained between them (Figure S8D). To explore this observation in more detail, we examined the dynamic induction of tyrosine kinase ligands and receptors in the EOM. Notably, *Bmpr1b* and *Ephb1* were found to be among the top 100 driver genes of the myogenic EOM compartment, indicating that myogenic commitment is associated with upregulation of these receptors in the EOM. Strikingly, 2 of their respective ligands *Bmp4* and *Efnb1*, were found specifically in non-myogenic cells. These results suggested that complementary paracrine signalling network operates between myogenic and non-myogenic *Myf5*-derived cells, while cellular juxtaposition is maintained through fetal stages in the EOM.

**Myogenic and non-myogenic states are associated with specific gene regulatory networks**

The directional trajectories inferred by RNA velocity in the EOM at E11.5 showed a strong bipolarity in fate with a higher velocity confidence index at each end of the myogenic and non-myogenic domains (Figure S8B). This suggests that the anticipated cell fate is ambiguous at the interface between myogenic and non-myogenic cells. Conversely, cells that are located on either side of this central region can be identified with greater confidence as committed to myogenic or non-myogenic fates. In addition, our previous gene set enrichment analysis revealed transcription factors as major effectors that are implicated at every stage in the non-myogenic cluster (Figure 4D-D'). This suggested that this cell fate transition might be dictated by the activation of a combinatorial of transcription factors that together reshape transcriptional output.

To identify these factors, we used a regulatory network inference algorithm called SCENIC (Aibar et al., 2017). This tool allows regrouping of sets of correlated genes into regulons (i.e. a transcription factor and its targets) based on binding motifs and co-expression. Use of this pipeline significantly reduced the number of variables from thousands of genes to a few hundred regulatory modules, while preserving the general aspect of the data, particularly the bipartite distribution of myogenic and non-myogenic cells in UMAP (Figure 6A). The top regulons of this analysis revealed the active transcription factors underlying myogenic and non-myogenic cell fates in the EOM at E11.5. Notably, *Myf5*, *Pitx1*, *Mef2a* and *Six1*, transcription factors known to be implicated in myogenic development (Buckingham and Rigby, 2014; Maire et al., 2020), appeared among the top regulons in myogenic cells whereas *Fli1*, *Ebf1*, Ets1, *Foxc1*, *Meis1* and *Six2*, genes known for their involvement in adipogenic, vascular, mesenchymal and tendon development (Jimenez et al., 2006; López-Delgado et al., 2020; Noizet et al., 2016; Truong and Ben-David, 2000; Whitesell et al., 2019; Yamamoto-Shiraishi and Kuroiwa, 2013), constituted some of the highly active non-myogenic transcription factors (Figure 6B).

To facilitate visualization of transcription factor modules, we built a network comprised solely of transcription factors that are coregulated, by removing all targets that were not regulons themselves. This resulted in a graphical representation where connections between nodes indicate inferred regulation. To better highlight interconnected nodes, we used a force-directed algorithm, which assigns spring-like attraction and repulsion qualities to node edges. By doing so, some nodes dispersed while others organized into tightly interconnected modules, thus revealing potential transcription factor associations (Figure 6C). This analysis indicated that many of the top regulons that we identified in both myogenic and non-myogenic cells likely coregulate each other, for example, *Prrx2*, *Fli1, Foxc2, Ebf1, Ets1, Emx2 and Twist1*.

**Key transcription factors underlie non-myogenic fate transitions at various stages and anatomical locations**

Myf5+ bipotent progenitors were observed at multiple stages and anatomical locations, and they yielded a relatively homogeneous population expressing common markers associated with extracellular matrix components, cell adhesion molecules, and tyrosine kinase signalling. However, it is possible that the regulatory mechanisms guiding this transition could be distinct in different locations in the head. Therefore, we set out to explore the common molecular switches underlying this cell fate decision. To do so, we developed a pipeline where we combined the list of driver genes (transcriptionally dynamic genes) at the start of the non-myogenic trajectory with the most active regulons in the non-myogenic region. In other words, we connected each driver gene with its potential regulon, if it were present. This resulted in a network consisting of the most active transcription factors and the most transcriptionally dynamic genes found at the non-myogenic branchpoint. We performed this operation for each dataset independently and displayed them as

individual networks (Figure 7A-D). Finally, we overlapped the list of these “driver regulators” to identify the common transcription factors guiding the non-myogenic cell fate decision (Figure 7E-F). Notably, *Foxp2*, *Hmga2*, *Meis1*, *Meox2* and *Tcf7l2* were found in all 4 datasets as key driver regulators, and thus are likely to bear significant weight in the non-myogenic transition (Figure 7F).

## DISCUSSION

Distinct fates can potentially emerge through the specification of cells positioned within a signalling domain, or through direct lineage ancestry from bipotent or multipotent cells. Here we addressed this issue that defines the emergence of embryonic cell lineages by employing single-cell transcriptomics and complementary histological analysis. Specifically, we examined how myogenic and associated connective tissue cells arise to form craniofacial muscles. We show that bipotent progenitors with a bias in the myogenic signature can give rise to skeletal muscle and anatomically associated connective tissue cells spatiotemporally (Figure 8). Also, mesodermal- and neural crest cell-derived connective tissue cells can both contribute to the formation of muscle functional unit. Our work provides new analytical methods that combine state-of-the-art algorithms to tie in dynamic intercellular communication networks with their regulators and can be adapted to other tissues and organs.

### Leveraging scRNAseq for identifying lineage relationships in muscle and connective tissue

Our study relies extensively on scRNAseq data and computational methods for extracting valuable information from cell-lineage traced datasets (Chen et al., 2019). Given an appropriate experimental design, scRNAseq can provide powerful tools for discovering new cellular relationships during lineage progression and cell fate decisions (Cao et al., 2019; He et al., 2020; Lescroart et al., 2018; Nguyen et al., 2018). We exploited the latest algorithms to uncover directional trajectories between mesodermal transitory states (Bergen et al., 2020). Unsuspected cell fate decisions were validated with genetic tools and contemporary labelling in situ. We inferred gene regulatory networks by binding motifs and correlated expressions (Aibar et al., 2017; Sande et al., 2020) and identified transcription factor interactions that were visualized in a force-directed graph to highlight co-regulating modules. These methods were then combined to isolate the most meaningful common regulators across multiple datasets. We consider that this approach provides a valuable pipeline for refining large scRNAseq datasets and exploring primary sources of transcriptomic variation.

**Non-myogenic contribution of Myf5 lineage**

The *Myf5*-lineage has been shown to contribute to brown adipocytes, neurons, pericytes and rib cartilage (Daubas et al., 2000; Haldar et al., 2008; Sebo et al., 2018; Stuelsatz et al., 2014). Interestingly, when Myf5 expression is disrupted, cells acquire non-myogenic fates and contribute to cartilage, dermis and connective tissue (Tajbakhsh et al., 1996a). These studies suggest that *Myf5*-expression alone may not be sufficient to promote robust myogenic fate in multiple regions of the developing embryos. The Myod-lineage was reported not to contribute to rib cartilage and has not been shown to give rise to connective tissue cells (Wood et al., 2020). This is consistent with the higher chromatin-remodelling capacity of Myod over Myf5 and its classification as defining the committed myogenic cell state (Conerly et al., 2016; Tapscott, 2005). Our observations are consistent with this view as we noted that Myod-expressing cells rarely showed a directional trajectory towards a connective tissue state, but instead towards myogenic differentiation. In the EOM, it was proposed that *Myf5*-derived fibroadipogenic cells arise from a neural-crest population with an ancestral expression of *Myf5* (Stuelsatz et al., 2014). However, that study did not perform single-cell analyses of the neural-crest lineage and *Myf5* expression. In our study, we found no neural-crest derived cells expressing *Myf5* during EOM tissue genesis at E13.5 (using *Wnt1*$^{cre/+}$;*R26*$^{mTmG/+}$;*Myf5*$^{nLacZ/+}$). Nevertheless, we cannot exclude the possibility that *Myf5* was expressed prior to E13.5 in the neural crest and that the βGAL protein was degraded by E13.5. However, it appears clear that *Myf5*-expressing cells continue to contribute to EOM non-myogenic cells at the mid-embryonic stages E11.5-E12.5, as indicated by our results with *Myf5*$^{Cre/+}$;*R26*$^{mTmG/+}$ and *Myf5*$^{GFP\text{-}P/+}$ scRNAseq datasets.

**Role of tyrosine kinase signalling during myogenesis**

Tyrosine kinase receptors have been implicated in a number of developmental aspects of both muscle and associated connective tissue (Arnold et al., 2020; Brent and Tabin, 2004; Han et al., 2012; Han et al., 2014; Knight and Kothary, 2011; Michailovici et al., 2015; Miwa and Era, 2015; Olson and Soriano, 2009; Tallquist et al., 2000; Tzahor et al., 2003; Vinagre et al., 2010). For example, the differentiation of fetal myoblasts is inhibited by growth factors Tgfb and Bmp4 (Cossu et al., 2000). Epha7 signalling is expressed in embryonic and adult myocytes and promotes differentiation (Arnold et al., 2020). Particularly, we noticed a striking and lasting complementarity in the expression of Pdgfa and Pdgfra through embryonic stages, in the myogenic and non-myogenic progenitors respectively. Pdgf ligands emanating from hypaxial myogenic cells under the control of *Myf5* were shown to be necessary from rib cartilage development (Tallquist et al., 2000; Vinagre et al., 2010). Additionally, Pdgfra promotes fibroblastic expansion during fibrosis (Olson and Soriano, 2009). Interestingly, we found that *Pdgfa* expression was reduced in cells expressing high levels of *Myog* at the fetal stage (data not shown). Therefore, *Myf5*-derived myogenic progenitor cells might guide non-myogenic *Myf5*-derived expansion, which in turn provides ligands and extracellular matrix components to favour myogenic development and patterning.

**Cell identity of non-myogenic Myf5-derived cells**

Recent efforts have been made to further characterise anatomically distinct fibroblastic populations using single-cell transcriptomics (Muhl et al., 2020). However, unique markers could not be characterised (Muhl et al., 2020; Ramirez et al., 2020), making identification of subtypes challenging. Tcf4/Tcf7l2 was identified as a master regulator of fibroblastic fate during muscle-associated connective tissue development although being also expressed in myogenic progenitors at lower levels (Kardon et al., 2003; Mathew et al., 2011; Miller et al., 2007; Sefton and Kardon, 2019). We also report this gene in our integrated analysis to be one of the main regulators of connective tissue fate.

Other transcription factors have been linked to skin fibroblast fates including Tcf4, Six2, Meox2, Egr2 and Foxs1, and their repression favours a myofibroblastic potential (Noizet et al., 2016). Six2 and Meox2 were also found in our data, which raises the question of the shared genetic programs between myofibroblastic cells and fibroblastic cells derived from progenitors primed for myogenesis during development.

Gene set enrichment analyses suggested that Myf5-derived connective tissue cells perform a similar function in contributing to the developing muscles, despite a relative heterogeneity in the genes underlying these terms. It is possible however that the distinct molecular functions of such genes have not been yet fully characterised, and thus could not be discriminated in our analysis.

Interestingly, *Prrx1*, a marker for lateral plate mesoderm, was differentially expressed the connective tissue population at various stages (Durland et al., 2008). Although lateral plate mesoderm is clearly identifiable in the trunk, its anterior boundaries in the head are uncertain (Prummel et al., 2020). It is possible that this population largely overlaps with cardiopharyngeal mesoderm and more specifically lateral splanchnic mesoderm as suggested previously (Nathan et al., 2008; Prummel et al., 2020). More detailed analyses of *Prrx1*, *Isl1* and *Myf5* lineages need to be carried out to delineate the specific boundaries of each progenitor contribution to cranial connective tissue.

Finally, it is possible that this connective tissue-like state constitutes a transitory phase of myogenic expansion. Thus, this specific state may in part be reversible and some cells might reintegrate the myogenic compartment at later stages of development, while others can fully commit to connective tissue lineages. In adult muscle stem cell-derived myoblasts, expression of *Pdgfra* induces pericyte features and increase self-renewal capacities (Gerli et al., 2019). Further studies would be required to address the presence of such cell states in vivo during embryonic development and adulthood.

## CONCLUSIONS

Taken together, we report an unsuspected cell fate redirection of immature myogenic cells to connective tissue. This phenomenon occurs in multiple regions of the developing head, which are in part deprived of neural-crest cells. This transition is accompanied with complementary tyrosine kinase signalling between the myogenic and non-myogenic compartments, and this relationship is maintained as muscles are established in the fetus. Using new scRNAseq analytical methods, we show that transcription factors such as Prrx1/2, Twist1, Fli1, Foxp2, Tcf7l2, Meis1, Meox1/2 are the main regulators of this transcriptomic remodelling (Figure 8). This cellular resilience could lead to the emergence of various muscle-associated connective tissue in the adult head and may bear important consequences for muscle stem and niche cell communication in the adult.

## MATERIALS & METHODS

### scRNAseq data generation

For E10.5 to E12.5 embryos, the cranial region above the forelimb was dissected in ice-cold 3% FBS and mechanically dissociated. The same procedure was applied at E14.5 but the dissection was refined to the pharyngeal and laryngeal regions. Tissues were then digested in TrypLE (ThermoFisher Cat #: 12604013) during 3 rounds of 5-minute incubation (37°C, 1400 RPM), interspersed with gentle pipetting to further dissociate the tissue. Cells were resuspended in FBS 3%, filtered, and incubated with Calcein blue (eBioscience, Cat #: 65-0855-39) and Propidium Iodide (ThermoFisher Cat #: P1304MP) for viability check. Viable cells were sorted on BD FACSAria™ III and manually counted using a hemocytometer. RNA integrity was assessed with Agilent Bioanalyzer 2100 to validate the isolation protocol prior to scRNAseq (RIN>8 was considered acceptable). 4000 to 13000 cells were loaded onto 10X Genomics Chromium microfluidic chip and cDNA libraries were generated following manufacturer's protocol. Concentrations and fragment sizes were measured using Agilent Bioanalyzer and Invitrogen Qubit. cDNA libraries were sequenced using NextSeq 500 and High Output v2.5 (75 cycles) kits. Genome mapping and count matrix generation were done following 10X Genomics Cell Ranger pipeline.

### RNA velocity and driver genes

RNA velocity analyses were performed using scvelo (Bergen et al., 2020) in python. This tool allows inferring velocity flow and driver genes using scRNAseq data, with major improvements from previous methods (Manno et al., 2018). First, unspliced and spliced transcript matrices were generated using velocyto (Manno et al., 2018) command line function, which outputs unspliced, spliced, and ambiguous matrices as a single loom file. These files were combined with filtered Seurat objects to yield objects with unspliced and spliced matrices, as well as Seurat-generated annotations and cell-embeddings (UMAP, tSNE, PCA). These datasets were then processed

following scvelo online guide and documentation. Velocity was calculated based on the dynamical model (using *scv.tl.recover_dynamics(adata)*, and *scv.tl.velocity(adata, mode='dynamical')*) and when outliers were detected, differential kinetics based on top driver genes were calculated and added to the model (using *scv.tl.velocity(adata, diff_kinetics=True)*). Specific driver genes were identified by determining the top likelihood genes in the selected cluster. The lists of top 100 drivers for each stage are given in Table1.

**Seurat preprocessing**

scRNAseq datasets were preprocessed using Seurat in R (https://satijalab.org/seurat/) (Butler et al., 2018). Cells with more than 20% of mitochondrial gene fraction were discarded. The number of genes expressed averaged to 4000 in all 4 datasets. Dimension reduction and UMAP generation were performed following Seurat workflow. Doublets were inferred using DoubletFinder v3 (McGinnis et al., 2019). Cell cycle genes, mitochondrial fraction, number of genes, number of UMI were regressed in all datasets following Seurat dedicated vignette. We noticed that cell cycle regression, although clarifying anatomical diversity, seemed to induce low and high UMI clustering (Suppl. Fig. 3I-J). For the E10.5 and E11.5 datasets, 2 replicates were generated from littermates and merged after confirming their similitude. For subsequent datasets (E12.5 and E14.5), no replicates were used. Annotation and subsetting were also performed in Seurat. "Myogenic" and "Non-myogenic" annotations were based on *Pdgfa* and *Pdgfra* expression and myogenic genes. Cells not expressing *Pdgfa* were annotated as "non-myogenic" unless they express *Myf5*. Cells expressing *Pdgfa* were annotated as "myogenic". We noticed that at later stages, *Pdgfa* expression decreases in Myog+ cells. Thus, driver genes of connective tissue at E12.5 and E14.5 were determined using cluster annotations obtained from Leiden-based clustering.

**Gene regulatory network inference and transcription factor modules**

Gene regulatory networks were inferred using SCENIC (R implementation) and pySCENIC (Aibar et al., 2017; Sande et al., 2020) (python implementation). This algorithm allows regrouping of sets of correlated genes into regulons (i.e. a transcription factor and its targets) based on motif binding and co-expression. UMAP and heatmap were generated using regulon AUC matrix (Area Under Curve) which refers to the activity level of each regulon in each cell. The outputted list of each regulon and their targets was subsequently used to create a transcription factor network based on interactions involving genes that are regulons themselves. This operation greatly reduced the number of genes involved, while highlighting co-regulating transcriptional modules. 2 tables were generated: a node table and an edge table. The node table comprised all transcription factors involved, along with a "top50 regulon" annotation ("Top 50 myogenic", "Top 50 non-myogenic", or "Other"). The edge table was comprised of source and target IDs along with a column indicating the weight of the connection, which refers to the correlation of these 2 genes in the data, obtained

from SCENIC correlation matrix. These tables were imported in Gephi (https://gephi.org/), networks were generated and subjected to the "Force-Atlas2" algorithm.

**Driver regulons**

Results from SCENIC and scvelo were combined to identify regulons that could be responsible for the transcriptomic induction of driver genes. Similarly to the steps mentioned above, SCENIC lists of regulons were used to infer connections between transcription factors and driver gene. Networks were generated as explained above, and annotated with "Active regulon" or "driver gene". The lists of individual driver regulons of each dataset were then combined and the most recurring driver regulons were identified.

**Gene set enrichment analysis**

Gene set enrichment analyses were performed on either top markers (obtained from Seurat function FindAllMarkers) or from driver genes (obtained from scvelo), using Cluego (Bindea et al., 2009). "GO Molecular Pathway", "GO Biological Process" and "Reactome pathways" were used independently to identify common and unique pathways involved in each dataset. In all analyses, an enrichment/depletion two-sided hypergeometric test was performed and p-values were corrected using the Bonferroni step down method.

**Mouse strains**

Animals were handled according to European Community guidelines and the ethics committee of the Institut Pasteur (CETEA) approved protocols. The following strains were previously described: *Myf5*$^{Cre}$ (Haldar et al., 2008), *Mesp1*$^{Cre}$ (Saga et al., 1999), *Tg:Wnt1Cre* (Danielian et al., 1998), *R26*$^{TdTom}$ (Ai9;(Madisen et al., 2009)), *R26*$^{mTmG}$(Muzumdar et al., 2007), *Myf5*$^{nlacZ}$ (Tajbakhsh et al., 1996a), *Pdgfra*$^{H2BGFP}$ (Hamilton et al., 2003) and *Myf5*$^{GFP-P}$ (Kassar-Duchossoy et al., 2004).To generate *Myf5*$^{Cre/+}$*;R26*$^{TdTomato/+}$*;Pdgfra*$^{H2BGFP/+}$embryos, *Myf5*$^{Cre/+}$ females were crossed with *Pdgfra*$^{H2BGFP/+}$*;R26*$^{TdTomato/TdTomato}$ males. Mice were kept on a mixed genetic background C57BL/6JRj and DBA/2JRj (B6D2F1, Janvier Labs). Mouse embryos and fetuses were collected between embryonic day (E) E10.5 and E14.5, with noon on the day of the vaginal plug considered as E0.5.

**Immunofluorescence**

Collected embryonic and adult tissues were fixed 2.5h in 4% paraformaldehyde (Electron Microscopy Sciences, Cat #:15710) in PBS with 0,2-0,5% Triton X-100 (according to their stage) at 4°C and washed overnight at 4°C in PBS. In preparation for cryosectioning, embryos were equilibrated in 30% sucrose in PBS overnight at 4°C and embedded in OCT. Cryosections (16-20µm) were left to dry at RT for 30 min and washed in PBS. The primary antibodies used in this study are chicken polyclonal anti-b-gal (Abcam, Cat #: ab9361, dilution 1:1000), mouse monoclonal

IgG1, mouse monoclonal IgG1 anti-Myod (BD Biosciences, Cat# 554130, dilution 1:100), mouse monoclonal IgG1 anti-Pax7 (DSHB, Cat. #: AB_528428 IF (1:20), rabbit anti-mouse Sox9 (Millipore, Cat. #: AB5535, dilution 1/2000), rabbit polyclonal anti-Tomato (Clontech Cat. #: 632496, dilution (1:500)) and chicken polyclonal anti-GFP (Abcam Cat. #: 13970, dilution (1:1000)). Images were acquired using Zeiss LSM780 or LSM700 confocal microscopes and processed using ZEN software (Carl Zeiss).

**RNAscope in situ hybridization**

Embryos for in situ hybridization were fixed overnight in 4% PFA. Embryos were equilibrated in sucrose and sectioned as described for immunofluorescence. RNAscope probes Mm-Pdgfa(411361) and Mm-Pdgfra(480661-C2) were purchased from Advanced Cell Diagnostics, Inc. In situ hybridization was performed using the RNAscope Multiplex Fluorescent Reagent Kit V2 and RNAscope 2.5 HD ReagentKit-RED according to manufacturer's instructions (Wang et al., 2012) and modifications detailed in previous work (Comai et al., 2019).

**Acknowledgements**

We acknowledge funding support from the Institut Pasteur, Association Française contre le Myopathies, Agence Nationale de la Recherche (Laboratoire d'Excellence Revive, Investissement d'Avenir; ANR-10-LABX-73), Association Française contre les Myopathies (Grant #20510), Fondation pour la Recherche Médicale (Grant # FDT201904008277), and the Centre National de la Recherche Scientifique. We gratefully acknowledge the UtechS Photonic BioImaging (Imagopole), C2RT, Institut Pasteur, supported by the French National Research Agency (France BioImaging; ANR-10–INSB–04; Investments for the Future).

**Competing interests**

The authors declare no competing interests.

**REFERENCES:**

**Adachi, N., Takechi, M., Hirai, T. and Kuratani, S.** (2012). Development of the head and trunk mesoderm in the dogfish,Scyliorhinus torazame: II. Comparison of gene expression between the head mesoderm and somites with reference to the origin of the vertebrate head: Gene expression in head cavities. *Evol Dev* **14**, 257–276.

**Adachi, N., Bilio, M., Baldini, A. and Kelly, R. G.** (2020). Cardiopharyngeal mesoderm origins of musculoskeletal and connective tissues in the mammalian pharynx. *Development* **147**, dev185256.

**Ahmad, K., Shaikh, S., Ahmad, S. S., Lee, E. J. and Choi, I.** (2020). Cross-Talk Between Extracellular Matrix and Skeletal Muscle: Implications for Myopathies. *Front Pharmacol* **11**, 142.

**Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C. C., Geurts, P., Aerts, J., et al.** (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature methods* **14**, 1083–1086.

**Arnold, L. L., Cecchini, A., Stark, D. A., Ihnat, J., Craigg, R. N., Carter, A., Zino, S. and Cornelison, D.** (2020). EphA7 promotes myogenic differentiation via cell-cell contact. *Elife* **9**, e53689.

**Bergen, V., Lange, M., Peidli, S., Wolf, F. A. and Theis, F. J.** (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* 1–7.

**Biferali, B., Proietti, D., Mozzetta, C. and Madaro, L.** (2019). Fibro–Adipogenic Progenitors Cross-Talk in Skeletal Muscle: The Social Network. *Front Physiol* **10**, 1074.

**Bildsoe, H., Loebel, D. A. F., Jones, V. J., Hor, A. C. C., Braithwaite, A. W., Chen, Y.-T., Behringer, R. R. and Tam, P. P. L.** (2013). The mesenchymal architecture of the cranial mesoderm of mouse embryos is disrupted by the loss of Twist1 function. *Dev Biol* **374**, 295–307.

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

**Bothe, I. and Dietrich, S.** (2006). The molecular setup of the avian head mesoderm and its implication for craniofacial myogenesis. *Developmental Dynamics* **235**, 2845–2860.

585 **Brent, A. E. and Tabin, C. J.** (2004). FGF acts directly on the somitic tendon progenitors through
586 the Ets transcription factors Pea3 and Erm to regulate scleraxis expression. *Development* **131**,
587 3885–3896.

588 **Buckingham, M. and Rigby, P. W. J.** (2014). Gene Regulatory Networks and Transcriptional
589 Mechanisms that Control Myogenesis. *Dev Cell* **28**, 225–238.

590 **Burke, A. C. and Nowicki, J. L.** (2003). A New View of Patterning Domains in the Vertebrate
591 Mesoderm. *Dev Cell* **4**, 159–165.

592 **Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R.** (2018). Integrating single-cell
593 transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**,
594 411–420.

595 **Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S.,**
596 **Christiansen, L., Steemers, F. J., et al.** (2019). The single-cell transcriptional landscape of
597 mammalian organogenesis. *Nature* **566**, 496–502.

598 **Chai, Y. and Maxson, R. E.** (2006). Recent advances in craniofacial morphogenesis.
599 *Developmental Dynamics* **235**, 2353–2375.

600 **Chen, G., Ning, B. and Shi, T.** (2019). Single-Cell RNA-Seq Technologies and Related
601 Computational Data Analysis. *Frontiers in genetics* **10**, 317.

602 **Comai, G., Heude, E., Mella, S., Paisant, S., Pala, F., Gallardo, M., Langa, F., Kardon, G.,**
603 **Gopalakrishnan, S. and Tajbakhsh, S.** (2019). A distinct cardiopharyngeal mesoderm genetic
604 hierarchy establishes antero-posterior patterning of esophagus striated muscle. *Elife* **8**, e47460.

605 **Comai, G., Tesarova, M., Dupé, V., Rhinn, M., Garcia, P. V., Silva, F. da, Feret, B., Exelby, K.,**
606 **Dollé, P., Carlsson, L., et al.** (2020). Local retinoic acid directs emergence of the extraocular
607 muscle functional unit. *Biorxiv* 2020.01.07.897694.

608 **Conerly, M. L., Yao, Z., Zhong, J. W., Groudine, M. and Tapscott, S. J.** (2016). Distinct Activities
609 of *Myf5* and MyoD Indicate Separate Roles in Skeletal Muscle Lineage Specification and
610 Differentiation. *Dev Cell* **36**, 375–85.

611 **Cossu, G., Kelly, R., Tajbakhsh, S., Donna, S. D., Vivarelli, E. and Buckingham, M.** (1996).
612 Activation of different myogenic pathways: myf-5 is induced by the neural tube and MyoD by
613 the dorsal ectoderm in mouse paraxial mesoderm. *Dev Camb Engl* **122**, 429–37.

**Danielian, P. S., Muccino, D., Rowitch, D. H., Michael, S. K. and McMahon, A. P.** (1998). Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. *Curr Biol* **8**, 1323-S2.

**Daubas, P., Tajbakhsh, S., Hadchouel, J., Primig, M. and Buckingham, M.** (2000). *Myf5* is a novel early axonal marker in the mouse brain and is subjected to post-transcriptional regulation in neurons. *Dev Camb Engl* **127**, 319–31.

**Deries, M., Schweitzer, R. and Duxson, M. J.** (2010). Developmental fate of the mammalian myotome. *Dev Dynam* **239**, 2898–2910.

**Dietrich, S., Schubert, F. R., Healy, C., Sharpe, P. T. and Lumsden, A.** (1998). Specification of the hypaxial musculature. *Dev Camb Engl* **125**, 2235–49.

**Diogo, R., Kelly, R. G., Christiaen, L., Levine, M., Ziermann, J. M., Molnar, J. L., Noden, D. M. and Tzahor, E.** (2015). A new heart for a new head in vertebrate cardiopharyngeal evolution. *Nature* **520**, 466–73.

**Douarin, N. L. and Kalcheim, C.** (1999). The Neural Crest.

**Durland, J. L., Sferlazzo, M., Logan, M. and Burke, A. C.** (2008). Visualizing the lateral somitic frontier in the Prx1Cre transgenic mouse. *J Anat* **212**, 590–602.

**Farahani, R. M. and Xaymardan, M.** (2015). Platelet-Derived Growth Factor Receptor Alpha as a Marker of Mesenchymal Stem Cells in Development and Stem Cell Biology. *Stem Cells Int* **2015**, 362753.

**Gerli, M. F. M. F., Moyle, L. A., Benedetti, S., Ferrari, G., Ucuncu, E., Ragazzi, M., Constantinou, C., Louca, I., Sakai, H., Ala, P., et al.** (2019). Combined Notch and PDGF Signaling Enhances Migration and Expression of Stem Cell Markers while Inducing Perivascular Cell Features in Muscle Satellite Cells. *Stem cell reports* **12**, 461–473.

**Grenier, J., Teillet, M.-A., Grifone, R., Kelly, R. G. and Duprez, D.** (2009). Relationship between neural crest cells and cranial mesoderm during head muscle development. *Plos One* **4**, e4381.

**Grimaldi, A., Parada, C. and Chai, Y.** (2015). A Comprehensive Study of Soft Palate Development in Mice. *PloS one* **10**, e0145018.

**Gulyaeva, O., Dempersmier, J. and Sul, H. S.** (2019). Genetic and epigenetic control of adipose development. *Biochimica Et Biophysica Acta Bba - Mol Cell Biology Lipids* **1864**, 3–12.

**Haldar, M., Karan, G., Tvrdik, P. and Capecchi, M. R.** (2008). Two cell lineages, *myf5* and *myf5*-independent, participate in mouse skeletal myogenesis. *Dev Cell* **14**, 437–45.

**Hamilton, T. G., Klinghoffer, R. A., Corrin, P. D. and Soriano, P.** (2003). Evolutionary Divergence of Platelet-Derived Growth Factor Alpha Receptor Signaling Mechanisms. *Mol Cell Biol* **23**, 4013–4025.

**Han, D., Zhao, H., Parada, C., Hacia, J. G., Bringas, P. and Chai, Y.** (2012). A TGFβ-Smad4-Fgf6 signaling cascade controls myogenic differentiation and myoblast fusion during tongue development. *Dev Camb Engl* **139**, 1640–50.

**Han, A., Zhao, H., Li, J., Pelikan, R. and Chai, Y.** (2014). ALK5-mediated transforming growth factor β signaling in neural crest cells controls craniofacial muscle development via tissue-tissue interactions. *Mol Cell Biol* **34**, 3120–31.

**He, P., Williams, B. A., Trout, D., Marinov, G. K., Amrhein, H., Berghella, L., Goh, S.-T., Plajzer-Frick, I., Afzal, V., Pennacchio, L. A., et al.** (2020). The changing mouse embryo transcriptome at whole tissue and single-cell resolution. *Biorxiv* 2020.06.14.150599.

**Helmbacher, F. and Stricker, S.** (2020). Tissue cross talks governing limb muscle development and regeneration. *Semin Cell Dev Biol* **104**, 14–30.

**Heude, E., Tesarova, M., Sefton, E. M., Jullian, E., Adachi, N., Grimaldi, A., Zikmund, T., Kaiser, J., Kardon, G., Kelly, R. G., et al.** (2018). Unique morphogenetic signatures define mammalian neck muscles and associated connective tissues. *eLife* **7**,.

**Ikeya, M. and Takada, S.** (1998). Wnt signaling from the dorsal neural tube is required for the formation of the medial dermomyotome. *Dev Camb Engl* **125**, 4969–76.

**Jimenez, M. A., Åkerblad, P., Sigvardsson, M. and Rosen, E. D.** (2006). Critical Role for Ebf1 and Ebf2 in the Adipogenic Transcriptional Cascade ▿ †. *Mol Cell Biol* **27**, 743–757.

**Kardon, G., Harfe, B. D. and Tabin, C. J.** (2003). A Tcf4-Positive Mesodermal Population Provides a Prepattern for Vertebrate Limb Muscle Patterning. *Dev Cell* **5**, 937–944.

**Kassar-Duchossoy, L., Gayraud-Morel, B., Gomès, D., Rocancourt, D., Buckingham, M., Shinin, V. and Tajbakhsh, S.** (2004). Mrf4 determines skeletal muscle identity in *Myf5*:Myod double-mutant mice. *Nature* **431**, 466–471.

**Knight, J. D. and Kothary, R.** (2011). The myogenic kinome: protein kinases critical to mammalian skeletal myogenesis. *Skelet Muscle* **1**, 29.

673 **Lefebvre, V., Angelozzi, M. and Haseeb, A.** (2019). SOX9 in cartilage development and disease.
674 *Curr Opin Cell Biol* **61**, 39–47.

675 **Lescroart, F., Kelly, R. G., Garrec, J.-F. F. L., Nicolas, J.-F. F., Meilhac, S. M. M. and**
676 **Buckingham, M.** (2010). Clonal analysis reveals common lineage relationships between head
677 muscles and second heart field derivatives in the mouse embryo. *Development (Cambridge,*
678 *England)* **137**, 3269–79.

679 **Lescroart, F., Hamou, W., Francou, A., Théveniau-Ruissy, M., Kelly, R. G. and Buckingham,**
680 **M.** (2015). Clonal analysis reveals a common origin between nonsomite-derived neck muscles
681 and heart myocardium. *Proceedings of the National Academy of Sciences of the United States*
682 *of America* **112**, 1446–51.

683 **Lescroart, F., Wang, X., Lin, X., Swedlund, B., Gargouri, S., Sànchez-Dànes, A., Moignard,**
684 **V., Dubois, C., Paulissen, C., Kinston, S., et al.** (2018). Defining the earliest step of
685 cardiovascular lineage segregation by single-cell RNA-seq. *Science (New York, N.Y.)* **359**,
686 1177–1181.

687 **López-Delgado, A. C., Delgado, I., Cadenas, V., Sánchez-Cabo, F. and Torres, M.** (2020). Axial
688 skeleton anterior-posterior patterning is regulated through feedback regulation between Meis
689 transcription factors and retinoic acid. *Biorxiv* 2020.03.09.983106.

690 **Madisen, L., Zwingman, T. A., Sunkin, S. M., Oh, S. W., Zariwala, H. A., Gu, H., Ng, L. L.,**
691 **Palmiter, R. D., Hawrylycz, M. J., Jones, A. R., et al.** (2009). A robust and high-throughput
692 Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* **13**, 133–
693 40.

694 **Maire, P., Santos, M. D., Madani, R., Sakakibara, I., Viaut, C. and Wurmser, M.** (2020).
695 Myogenesis control by SIX transcriptional complexes. *Semin Cell Dev Biol* **104**, 51–64.

696 **Manno, G. L., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber,**
697 **K., Kastriti, M. E., Lönnerberg, P., Furlan, A., et al.** (2018). RNA velocity of single cells. *Nature*
698 **560**, 494–498.

699 **Mathew, S. J., Hansen, J. M., Merrell, A. J., Murphy, M. M., Lawson, J. A., Hutcheson, D. A.,**
700 **Hansen, M. S., Angus-Hill, M. and Kardon, G.** (2011). Connective tissue fibroblasts and Tcf4
701 regulate myogenesis. *Dev Camb Engl* **138**, 371–84.

702 **McDavid, A., Finak, G. and Gottardo, R.** (2016). The contribution of cell cycle to heterogeneity in
703 single-cell RNA-seq data. *Nat Biotechnol* **34**, 591–593.

**McGinnis, C. S., Murrow, L. M. and Gartner, Z. J.** (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337.e4.

**Michailovici, I., Eigler, T. and Tzahor, E.** (2015). Current Topics in Developmental Biology. *Part 1: Craniofacial Morphogenesis and Regeneration: From Cells to Tissues to Organs* **115**, 3–30.

**Miller, K. A., Barrow, J., Collinson, J. M., Davidson, S., Lear, M., Hill, R. E. and MacKenzie, A.** (2007). A highly conserved Wnt-dependent TCF4 binding site within the proximal enhancer of the anti-myogenic Msx1 gene supports expression within Pax3-expressing limb bud muscle precursor cells. *Dev Biol* **311**, 665–678.

**Miwa, H. and Era, T.** (2015). Generation and characterization of PDGFRα-GFPCreERt2 knock-In mouse line. *Genesis* **53**, 329–336.

**Muhl, L., Genové, G., Leptidis, S., Liu, J., He, L., Mocci, G., Sun, Y., Gustafsson, S., Buyandelger, B., Chivukula, I. V., et al.** (2020). Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat Commun* **11**, 3953.

**Munsterberg, A. E., Kitajewski, J., Bumcrot, D. A., McMahon, A. P. and Lassar, A. B.** (1995). Combinatorial signaling by Sonic hedgehog and Wnt family members induces myogenic bHLH gene expression in the somite. *Gene Dev* **9**, 2911–2922.

**Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. and Luo, L.** (2007). A global double-fluorescent Cre reporter mouse. *Genesis (New York, N.Y. : 2000)* **45**, 593–605.

**Nassari, S., Duprez, D. and Fournier-Thibault, C.** (2017). Non-myogenic Contribution to Muscle Development and Homeostasis: The Role of Connective Tissues. *Frontiers in Cell and Developmental Biology* **5**, 22.

**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.

**Nguyen, Q. H., Pervolarakis, N., Nee, K. and Kessenbrock, K.** (2018). Experimental Considerations for Single-Cell RNA Sequencing Approaches. *Frontiers Cell Dev Biology* **6**, 108.

**Noden, D. M.** (1983). The role of the neural crest in patterning of avian cranial skeletal, connective, and muscle tissues. *Dev Biol* **96**, 144–165.

**Noden, D. M. and Epstein, M. L.** (2010). Embryonic origins of avian and mammalian laryngeal musculoskeletal structures. *Faseb J* **24**, 172.1-172.1.

**Noden, D. M. and Trainor, P. A.** (2005). Relations and interactions between cranial mesoderm and neural crest populations. *Journal of anatomy* **207**, 575–601.

**Nogueira, J. M., Hawrot, K., Sharpe, C., Noble, A., Wood, W. M., Jorge, E. C., Goldhamer, D. J., Kardon, G. and Dietrich, S.** (2015). The emergence of Pax7-expressing muscle stem cells during vertebrate head muscle development. *Front Aging Neurosci* **7**, 62.

**Noizet, M., Lagoutte, E., Gratigny, M., Bouschbacher, M., Lazareth, I., Crollius, H. R., Darzacq, X. and Dugast-Darzacq, C.** (2016). Master regulators in primary skin fibroblast fate reprogramming in a human ex vivo model of chronic wounds. *Wound Repair Regen* **24**, 247–262.

**Olson, L. E. and Soriano, P.** (2009). Increased PDGFRalpha activation disrupts connective tissue development and drives systemic fibrosis. *Dev Cell* **16**, 303–13.

**Pourquié, O., Fan, C.-M., Coltey, M., Hirsinger, E., Watanabe, Y., Bréant, C., Francis-West, P., Brickell, P., Tessier-Lavigne, M. and Douarin, N. M. L.** (1996). Lateral and Axial Signals Involved in Avian Somite Patterning: A Role for BMP4. *Cell* **84**, 461–471.

**Prummel, K. D., Nieuwenhuize, S. and Mosimann, C.** (2020). The lateral plate mesoderm. *Dev Camb Engl* **147**, dev175059.

**Ramirez, A. K., Dankel, S. N., Rastegarpanah, B., Cai, W., Xue, R., Crovella, M., Tseng, Y.-H., Kahn, C. R. and Kasif, S.** (2020). Single-cell transcriptional networks in differentiating preadipocytes suggest drivers associated with tissue heterogeneity. *Nat Commun* **11**, 2117.

**Saberi, M., Pu, Q., Valasek, P., Norizadeh-Abbariki, T., Patel, K. and Huang, R.** (2017). The hypaxial origin of the epaxially located rhomboid muscles. *Ann Anat - Anatomischer Anzeiger* **214**, 15–20.

**Saga, Y., Miyagawa-Tomita, S., Takagi, A., Kitajima, S., Miyazaki, J. i and Inoue, T.** (1999). MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Dev Camb Engl* **126**, 3437–47.

**Sande, B. V. de, Flerin, C., Davie, K., Waegeneer, M. D., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., et al.** (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* 1–30.

**Sebo, Z. L., Jeffery, E., Holtrup, B. and Rodeheffer, M. S.** (2018). A mesodermal fate map for adipose tissue. *Development* **145**, dev166801.

**Sefton, E. M. and Kardon, G.** (2019). Connecting muscle development, birth defects, and evolution: An essential role for muscle connective tissue. *Curr Top Dev Biol* **132**, 137–176.

**Shan, T., Liang, X., Bi, P., Zhang, P., Liu, W. and Kuang, S.** (2013). Distinct populations of adipogenic and myogenic *Myf5*-lineage progenitors in white adipose tissues. *J Lipid Res* **54**, 2214–24.

**Stuelsatz, P., Shearer, A. and Yablonka-Reuveni, Z.** (2014). Ancestral *Myf5* gene activity in periocular connective tissue identifies a subset of fibro/adipogenic progenitors but does not connote a myogenic origin. *Developmental biology* **385**, 366–79.

**Swedlund, B. and Lescroart, F.** (2019). Cardiopharyngeal Progenitor Specification: Multiple Roads to the Heart and Head Muscles. *Csh Perspect Biol* a036731.

**Tabler, J. M., Rigney, M. M., Berman, G. J., Gopalakrishnan, S., Heude, E., Al-Lami, H. A., Yannakoudakis, B. Z., Fitch, R. D., Carter, C. M., Vokes, S. A., et al.** (2017). Cilia-mediated Hedgehog signaling controls form and function in the mammalian larynx. *Elife* **6**, e19153.

**Tajbakhsh, S., Rocancourt, D. and Buckingham, M.** (1996a). Muscle progenitor cells failing to respond to positional cues adopt non-myogenic fates in myf-5 null mice. *Nature* **384**, 266–270.

**Tajbakhsh, S., Bober, E., Babinet, C., Pournin, S., Arnold, H. and Buckingham, M.** (1996b). Gene targeting the myf-5 locus with nlacZ reveals expression of this myogenic factor in mature skeletal muscle fibres as well as early embryonic muscle. *Developmental dynamics* **206**, 291–300.

**Tallquist, M., Weismann, K. and Development, H.-M.** (2000). Early myotome specification regulates PDGFA expression and axial skeleton development.

**Tapscott, S. J.** (2005). The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* **132**, 2685–2695.

**Truong, A. H. and Ben-David, Y.** (2000). The role of Fli-1 in normal cell function and malignant transformation. *Oncogene* **19**, 6482–6489.

**Tzahor, E., Kempf, H., Mootoosamy, R. C., Poon, A. C., Abzhanov, A., Tabin, C. J., Dietrich, S. and Lassar, A. B.** (2003). Antagonists of Wnt and BMP signaling promote the formation of vertebrate head muscle. *Gene Dev* **17**, 3087–3099.

**Vinagre, T., Moncaut, N., Carapuço, M., Nóvoa, A., Bom, J. and Mallo, M.** (2010). Evidence for a Myotomal Hox/Myf Cascade Governing Nonautonomous Control of Rib Specification within Global Vertebral Domains. *Dev Cell* **18**, 655–661.

**Wang, F., Flanagan, J., Su, N., Wang, L.-C., Bui, S., Nielson, A., Wu, X., Vo, H.-T., Ma, X.-J. and Luo, Y.** (2012). RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagnostics Jmd* **14**, 22–9.

**Wang, H., Liu, C., Liu, X., Wang, M., Wu, D., Gao, J., Su, P., Nakahata, T., Zhou, W., Xu, Y., et al.** (2018). MEIS1 Regulates Hemogenic Endothelial Generation, Megakaryopoiesis, and Thrombopoiesis in Human Pluripotent Stem Cells by Targeting TAL1 and FLI1. *Stem Cell Rep* **10**, 447–460.

**Wang, W., Niu, X., Stuart, T., Jullian, E., Mauck, W. M., Kelly, R. G., Satija, R. and Christiaen, L.** (2019). A single-cell transcriptional roadmap for cardiopharyngeal fate diversification. *Nature Cell Biology* **21**, 674–686.

**Whitesell, T. R., Chrystal, P. W., Ryu, J.-R. R., Munsie, N., Grosse, A., French, C. R., Workentine, M. L., Li, R., Zhu, L. J., Waskiewicz, A., et al.** (2019). foxc1 is required for embryonic head vascular smooth muscle differentiation in zebrafish. *Developmental biology* **453**, 34–47.

**Wood, W. M., Otis, C., Etemad, S. and Goldhamer, D. J.** (2020). Development and patterning of rib primordia are dependent on associated musculature. *Dev Biol*.

**Yamamoto-Shiraishi, Y. and Kuroiwa, A.** (2013). Wnt and BMP signaling cooperate with Hox in the control of Six2 expression in limb tendon precursor. *Dev Biol* **377**, 363–374.

**Zammit, P. S., Relaix, F., Nagata, Y., Ruiz, A. P., Collins, C. A., Partridge, T. A. and Beauchamp, J. R.** (2006). Pax7 and myogenic progression in skeletal muscle satellite cells. *J Cell Sci* **119**, 1824–1832.

# Figure 1

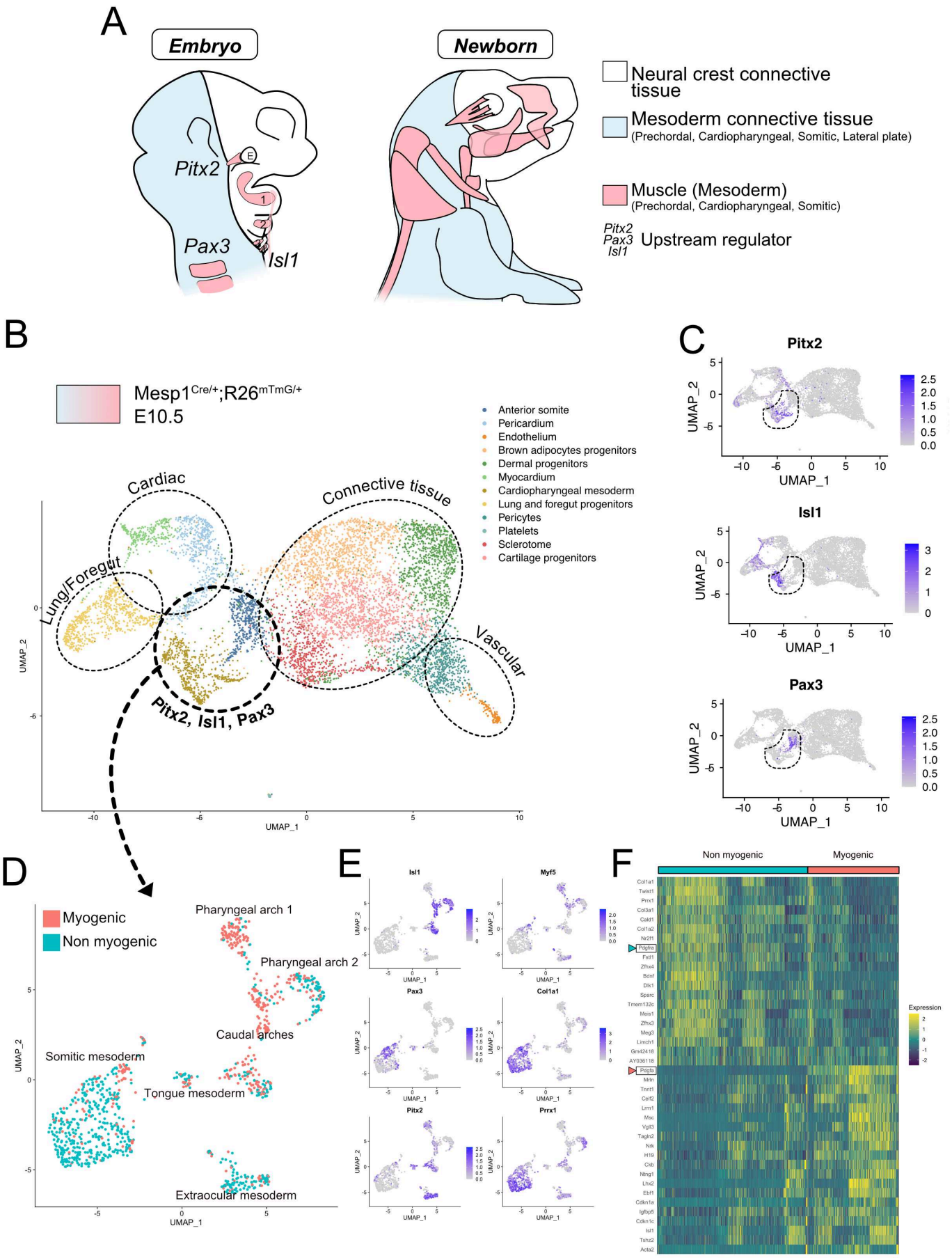

A
Embryo
Newborn
Pitx2
Pax3
Isl1
Neural crest connective tissue
Mesoderm connective tissue
(Prechordal, Cardiopharyngeal, Somitic, Lateral plate)
Muscle (Mesoderm)
(Prechordal, Cardiopharyngeal, Somitic)
Pitx2
Pax3
Isl1
Upstream regulator
B
Mesp1Cre/+;R26mTmG/+
E10.5
Anterior somite
Pericardium
Endothelium
Brown adipocytes progenitors
Dermal progenitors
Myocardium
Cardiopharyngeal mesoderm
Lung and foregut progenitors
Pericytes
Platelets
Sclerotome
Cartilage progenitors
Cardiac
Connective tissue
Lung/Foregut
Vascular
Pitx2, Isl1, Pax3
UMAP_1
UMAP_2
C
Pitx2
Isl1
Pax3
D
Myogenic
Non myogenic
Pharyngeal arch 1
Pharyngeal arch 2
Caudal arches
Somitic mesoderm
Tongue mesoderm
Extraocular mesoderm
E
Isl1
Myf5
Pax3
Col1a1
Pitx2
Prrx1
F
Non myogenic
Myogenic
Expression

**Figure 1. scRNAseq reveals non-myogenic populations of cranial mesoderm lineages.**
(A) Scheme of connective tissue origin in the head and known mesodermal upstream regulators. E: Eye, 1-4: Pharyngeal arches 1-4. (B) UMAP of *Mesp1*$^{Cre/+}$*; R26*$^{mTmG/+}$ E10.5 scRNAseq with main cell types highlighted. (C) UMAP expression plots of Pitx2, Isl1 and Pax3, indicating the clusters containing the cranial mesoderm upstream progenitors. (D) UMAP of the upstream progenitor subset annotated as "myogenic" and "non-myogenic" based on expression of Pdgfa, Myf5 (myogenic) and Pdgfra (non myogenic). (E) UMAP expression plots of lineages (Cardiopharyngeal mesoderm: Isl1, Extraocular mesoderm: Pitx2, and Somitic mesoderm: Pax3) and myogenic (Myf5) non-myogenic (Col1a1 and Prrx1) genes. (F) Heatmap of top 20 markers of myogenic versus non-myogenic clusters. Highlighted are Pdgfra/Pdgfa genes.

Figure 2

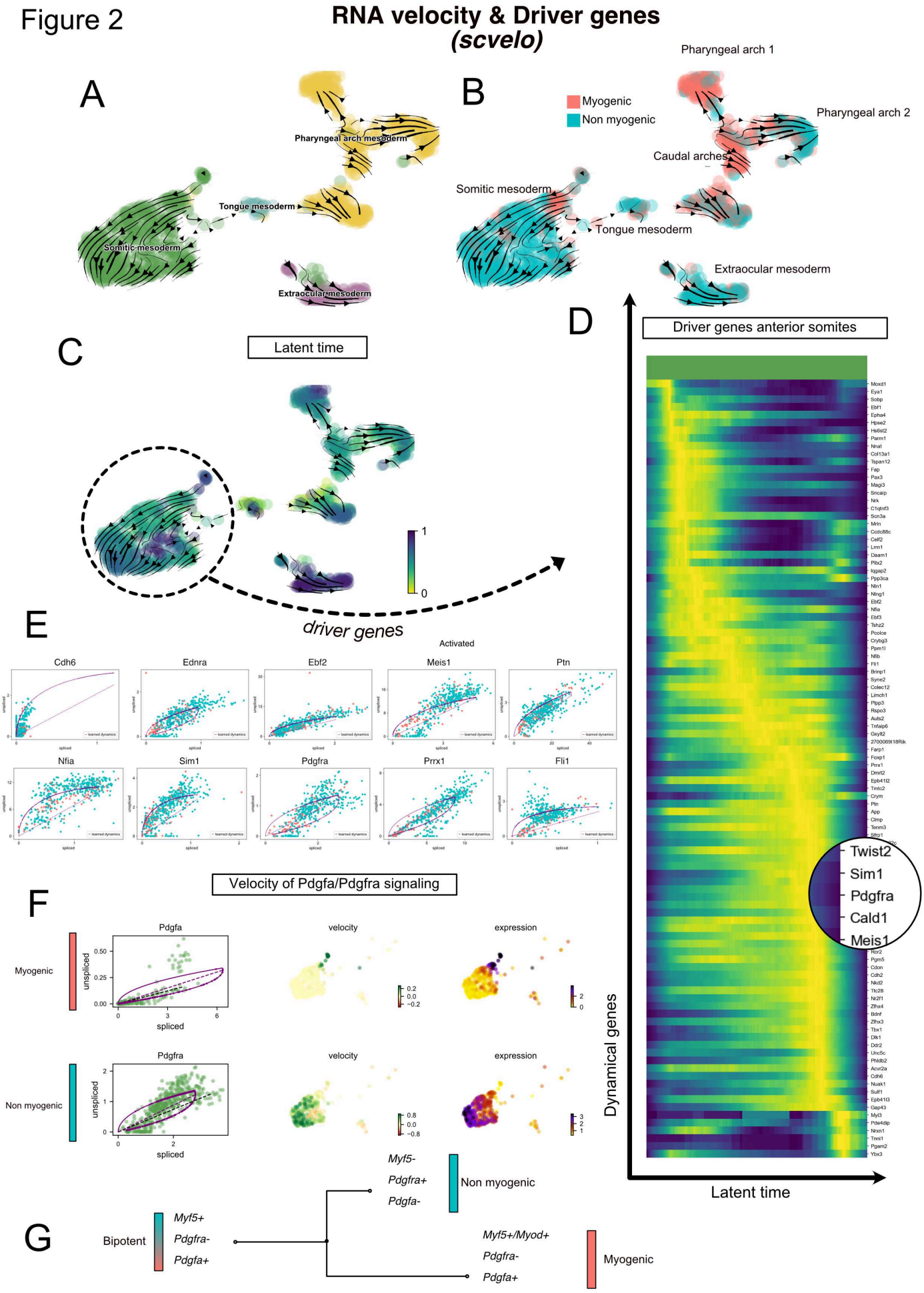
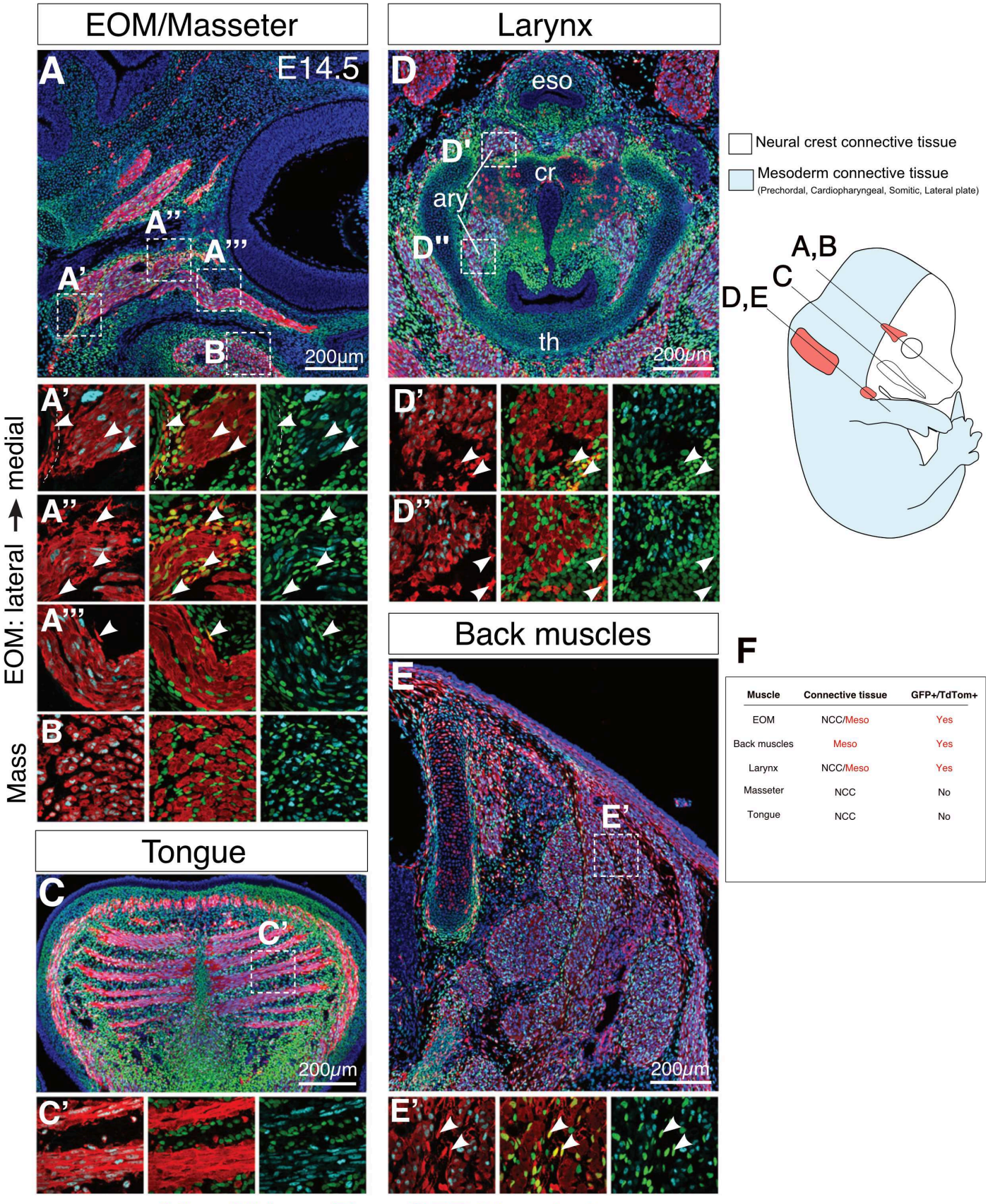

RNA velocity & Driver genes
(scvelo)
A
Pharyngeal arch mesoderm
Tongue mesoderm
Somitic mesoderm
Extraocular mesoderm
B
Myogenic
Non myogenic
Pharyngeal arch 1
Pharyngeal arch 2
Caudal arches
Somitic mesoderm
Tongue mesoderm
Extraocular mesoderm
C
Latent time
1
0
driver genes
D
Driver genes anterior somites
Twist2
Sim1
Pdgfra
Cald1
Meis1
Dynamical genes
Latent time
E
Activated
Cdh6
Ednra
Ebf2
Meis1
Ptn
Nfia
Sim1
Pdgfra
Prrx1
Fli1
F
Velocity of Pdgfa/Pdgfra signaling
Pdgfa
Myogenic
unspliced
spliced
velocity
expression
Pdgfra
Non myogenic
G
Bipotent
Myf5+
Pdgfra-
Pdgfa+
Myf5-
Pdgfra+
Pdgfa-
Non myogenic
Myf5+/Myod+
Pdgfra-
Pdgfa+
Myogenic

**Figure 2. Transcriptomic dynamics reveal a myogenic to non-myogenic transition.**

(A-C) Velocity UMAP plots displaying the main cranial mesoderm lineages (A), anatomical sites (B) and latent time (C). Arrows illustrate the local trajectories based on RNA velocity (relative abundance of unspliced and spliced transcripts). (D) Expression heatmap of driver genes accounting for anterior somite velocity, along latent time, highlighting Pdgfra. (E) Phase portraits of few selected driver genes including activated genes *Cdh6*, *Ednra*, *Ebf2*, *Meis1*, *Ptn*, *Nfia*, *Sim1*, *Pdgfra*, *Prrx1*, *Fli1*. The Y-axis represents the amount of unspliced transcript per cell, and the X-axis represents the number of spliced transcripts per cell. The dynamics of transcription were inferred at a gene- and cluster-specific level (see Methods). (F) Phase portraits, velocity and expression patterns of *Pdgfa* and *Pdgfra*. (G) Inferred model of myogenic and non-myogenic fate decision from a common bipotent progenitor.

Figure 3

$Myf5^{Cre/+};R26^{TdTom/+};Pdgfra^{H2BGFP/+}$
E14.5




| Muscle | Connective tissue | GFP+/TdTom+ |
|---|---|---|
| EOM | NCC/Meso | Yes |
| Back muscles | Meso | Yes |
| Larynx | NCC/Meso | Yes |
| Masseter | NCC | No |
| Tongue | NCC | No |

**Figure 3. Myf5-derived mesodermal connective tissue partially compensates for the lack of neural crest.**

(A-E) Transverse sections of an E14.5 *Myf5*$^{Cre/+}$*; R26*$^{TdTomato/+}$*; Pdgfra*$^{H2BGFP/+}$ embryo immunostained for Myod/Myog. Double positive cells were found in the EOM (preferentially in the medial region), back muscles (anterior somite-derived), and laryngeal muscles, but not in masseter and tongue muscles. White arrowheads indicated cells double-positive GFP/Tomato and negative for Myod/Myog. ary: arytenoid muscles, eso: esophagus, EOM: extraocular muscles, Meso: mesoderm, NCC: neural crest cells. (F) Table summarizing the results, correlated with local connective tissue origin. Presence of mesodermal contribution to muscle connective tissue correlated with the presence of Myf5-derived Pdgfra+ cells.

# Figure 4

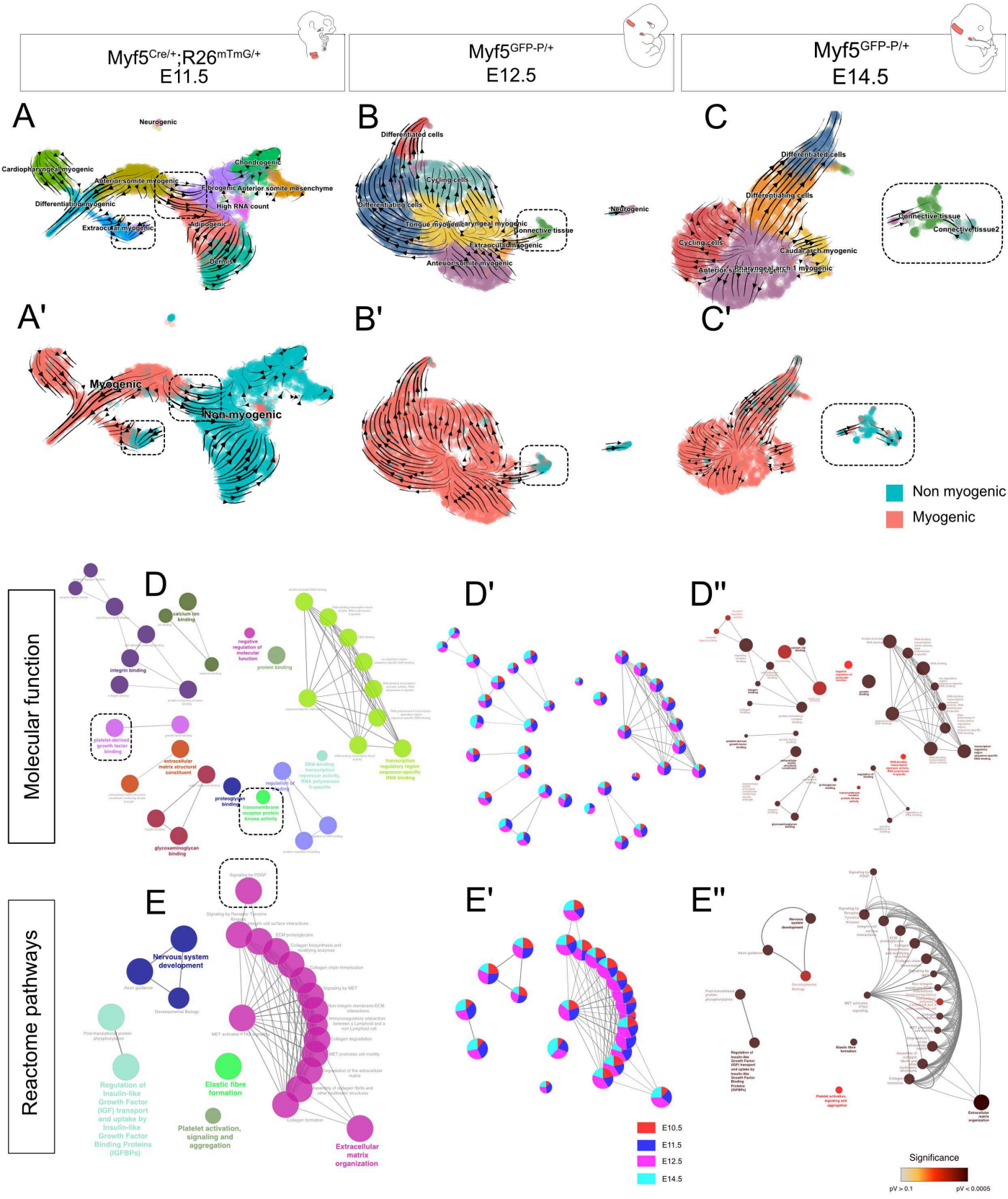

Myf5Cre/+;R26mTmG/+
E11.5
Myf5GFP-P/+
E12.5
Myf5GFP-P/+
E14.5
A
Neurogenic
Cardiopharyngeal myogenic
Anterior somite myogenic
Chondrogenic
Progenic Anterior somite mesenchyme
Differentiating myogenic
High RNA count
Extraocular myogenic
Adipogenic
Dermis
B
Differentiated cells
Cycling cells
Differentiating cells
Tongue myogenic
Connective tissue
Extraocular myogenic
Anterior somite myogenic
Neurogenic
C
Differentiated cells
Differentiating cells
Cycling cells
Caudal arch myogenic
Anterior pharyngeal arch 1 myogenic
Connective tissue
Connective tissue2
A'
Myogenic
Non myogenic
B'
C'
Non myogenic
Myogenic
Molecular function
D
calcium ion binding
integrin binding
negative regulation of molecular function
protein binding
platelet-derived growth factor binding
extracellular matrix structural constituent
proteoglycan binding
transmembrane receptor protein kinase activity
glycosaminoglycan binding
DNA-binding transcription repressor activity, RNA polymerase II-specific
transcription regulatory region sequence-specific DNA binding
D'
D''
Reactome pathways
E
Signaling by PDGF
Nervous system development
Axon guidance
Developmental Biology
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)
Elastic fibre formation
Platelet activation, signaling and aggregation
Extracellular matrix organization
E'
E10.5
E11.5
E12.5
E14.5
E''
Significance
pV > 0.1
pV < 0.0005

**Figure 4. Myf5 contribution to connective tissue is sustained through muscle initiation.**
(A-C) RNA velocity plots of *Myf5*$^{Cre/+}$*; R26*$^{mTmG/+}$ E11.5, *Myf5*$^{GFP/+}$ E12.5 and *Myf5*$^{GFP/+}$ E14.5 datasets displaying cell-type annotation and myogenic and non-myogenic clustering (A'-C'). Myf5-derived non-myogenic cells are found at multiple stages in different anatomical locations. (D-E) Gene set enrichment analysis network of GO Molecular Function and Reactome pathways performed on combined top 100 markers. Non-myogenic genes point towards numerous aspects of fibrogenic cell signature including Pdgf signaling and transmembrane receptor protein kinase activity (highlighted) (D'-E') Relative contribution of each stage to each term node. Most terms harbour a uniform contribution from all 4 stages (D''-E'') Gene set enrichment analysis term significance.

Figure 5

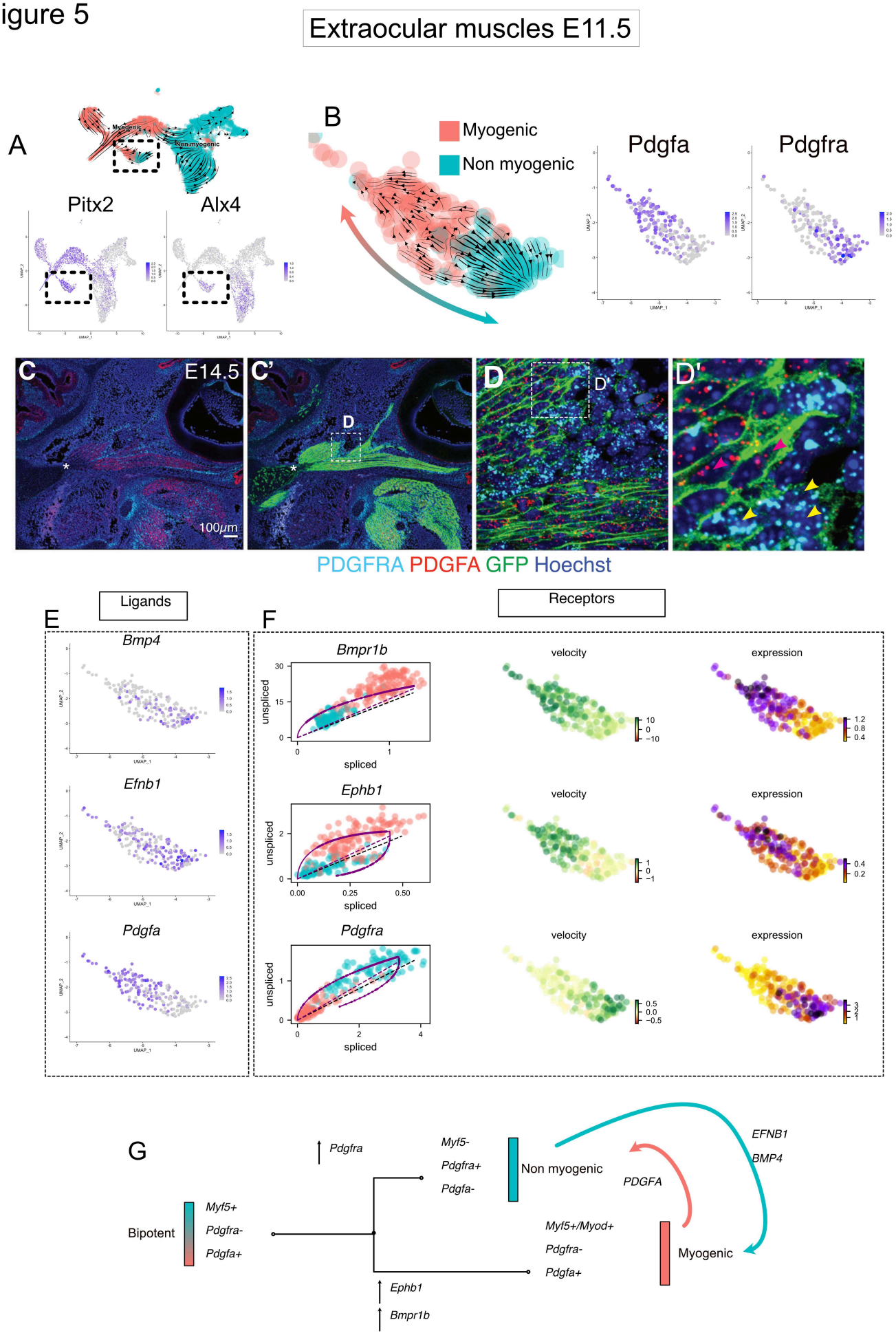

Extraocular muscles E11.5
A
Myogenic
Non-myogenic
Pitx2
Alx4
B
Myogenic
Non myogenic
Pdgfa
Pdgfra
C
E14.5
100µm
C'
D
D
D'
PDGFRA PDGFA GFP Hoechst
E
Ligands
Bmp4
Efnb1
Pdgfa
F
Receptors
Bmpr1b
Ephb1
Pdgfra
unspliced
spliced
velocity
expression
G
Bipotent
Myf5+
Pdgfra-
Pdgfa+
Pdgfra
Myf5-
Pdgfra+
Pdgfa-
Non myogenic
Myf5+/Myod+
Pdgfra-
Pdgfa+
Myogenic
Ephb1
Bmpr1b
EFNB1
BMP4
PDGFA

**Figure 5. Maintenance of signalling cues between Myf5-derived myogenic and non-myogenic cells in the EOM.**

(A) UMAP plots of *Myf5*$^{Cre/+}$*; R26*$^{mTmG/+}$ E11.5, displaying myogenic and non myogenic clusters, EOM markers *Pitx2* and *Alx4* overlaid with RNA velocity trajectories. The EOM cluster is highlighted. (B) UMAP plots of the EOM subset, displaying myogenic and non-myogenic clusters overlaid with RNA velocity trajectories, and *Pdgfa* (myogenic) and *Pdgfra* (non-myogenic) expression plots. (C-D') RNAscope on *Myf5*$^{Cre/+}$*; R26*$^{mTmG/+}$ E14.5 section with *Pdgfra* (cyan) and *Pdgfa* (red) probes. *Myf5*-derived cells are labelled by GFP membrane. The asterisk indicates the medial attachment site, also *Myf5*-derived. (D') High magnification view. Yellow arrowheads indicate *Myf5*-derived *Pdgfra*-expressing cells (non-myogenic). Red arrowheads indicate *Myf5*-derived *Pdgfa*-expressing cells (myogenic). (E-F) Expression patterns of ligands (E), phase portraits, velocity plots and expression patterns of receptors *Bmp*, *Ephrin* and *Pdgf* pathways. Differential expression of ligands in one cluster correlates with the active induction of the complementary receptor in the other cluster. (G) Current model of lineage bifurcation, associated with the expression of complementary signaling molecules.

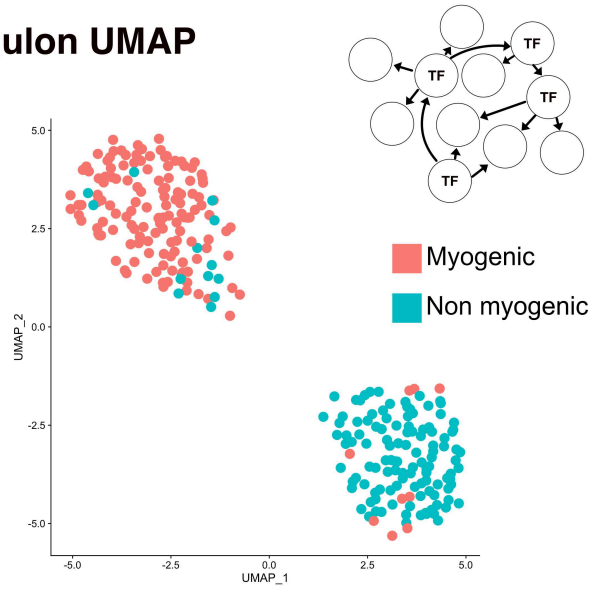
Figure 6

# Regulatory network inference *(SCENIC)*

Extraocular muscles E11.5

A

## Regulon UMAP




B

## Regulon Heatmap


## Transcription factor network

C

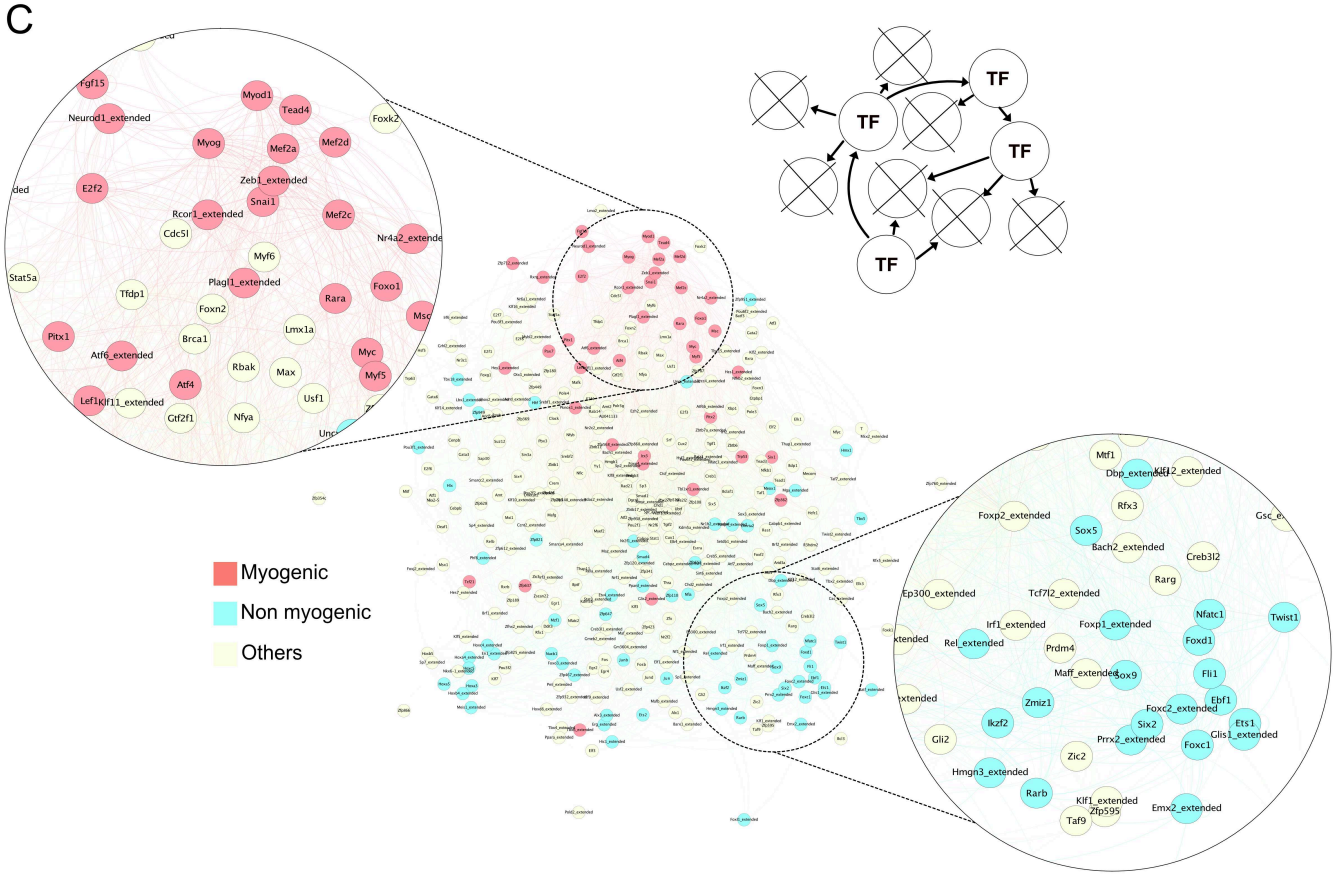


**Figure 6. Myogenic and non-myogenic cell fates are associated with particular gene regulatory networks.**

(A) UMAP of *Myf5*$^{Cre/+}$*; R26*$^{mTmG/+}$ E11.5 EOM based on SCENIC Regulon activity (Area Under Curve score). (B) Heatmap of top regulons (Transcription factor and associated targets). The suffix "_extended" indicates that the regulon includes motifs that have been linked to the TF by lower confidence annotations, for instance, inferred by motif similarity. The number in brackets indicates the number of genes comprising the regulon. (C) Transcription factor network based on regulons, top 50 regulons of each are colored. Opposite modules of non-myogenic (blue) and myogenic (red) are highlighted. The close proximity of these nodes suggests coregulation.

Figure 7

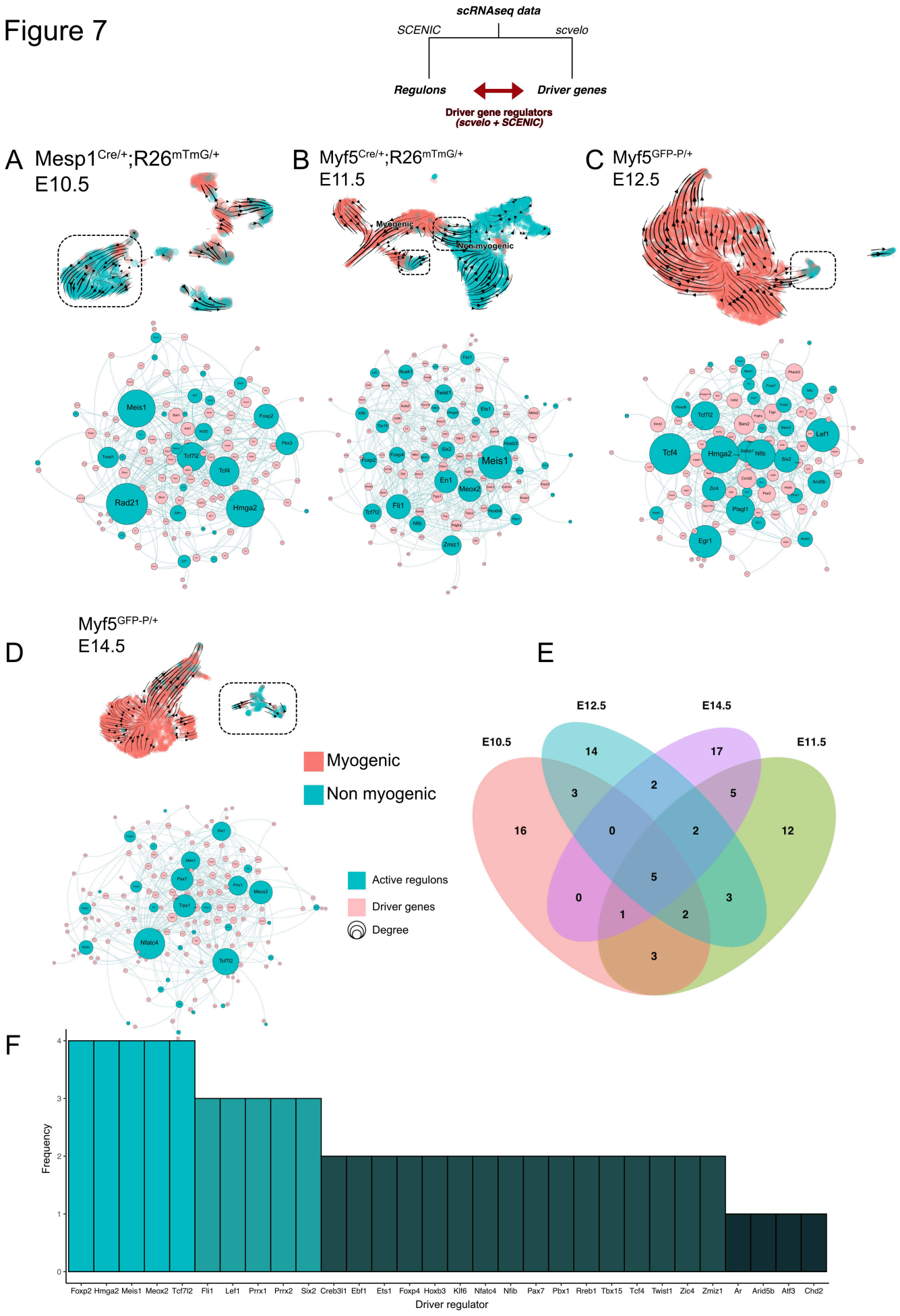

scRNAseq data
SCENIC
scvelo
Regulons
Driver genes
Driver gene regulators
(scvelo + SCENIC)
A Mesp1Cre/+;R26mTmG/+
E10.5
B Myf5Cre/+;R26mTmG/+
E11.5
Myogenic
Non-myogenic
C Myf5GFP-P/+
E12.5
Meis1
Foxp2
Pbx3
Tcf7l2
Tcf4
Rad21
Hmga2
Twist1
Ets1
Six2
Foxp4
En1
Meox2
Fli1
Zmiz1
Lef1
Nfib
Sox2
Arid5b
Plagl1
Egr1
D Myf5GFP-P/+
E14.5
Myogenic
Non myogenic
Active regulons
Driver genes
Degree
Pax7
Prrx1
Trps1
Nfatc4
E E12.5
E14.5
E10.5
E11.5
14
17
3
2
5
16
0
2
12
5
0
3
1
2
3
F
Frequency
0
1
2
3
4
Foxp2 Hmga2 Meis1 Meox2 Tcf7l2 Fli1 Lef1 Prrx1 Prrx2 Six2 Creb3l1 Ebf1 Ets1 Foxp4 Hoxb3 Klf6 Nfatc4 Nfib Pax7 Pbx1 Rreb1 Tbx15 Tcf4 Twist1 Zic4 Zmiz1 Ar Arid5b Atf3 Chd2
Driver regulator

**Figure 7. Key transcription factors are underlying non-myogenic fate transition at various stages and anatomical locations.**

(A-D) Velocity UMAP highlighting the non-myogenic transition at each stage, from which the underlying network was inferred. Driver genes and regulatory networks (regulons) were produced for each stage independently, and a stage-specific network of active transcription factor and associated driver gene targets was built. The size of nodes corresponds to the number of edges (connections) it has, (i.e. the number of driver genes it regulates). (E) Venn diagram illustrating the overlap on of “driver regulons” between each stage. (F) Histogram displaying the frequency of appearance of the most predominant transcription factors as driver regulators (4= present in all 4 datasets as driver regulon, 1= present in only 1 dataset).

Figure 8

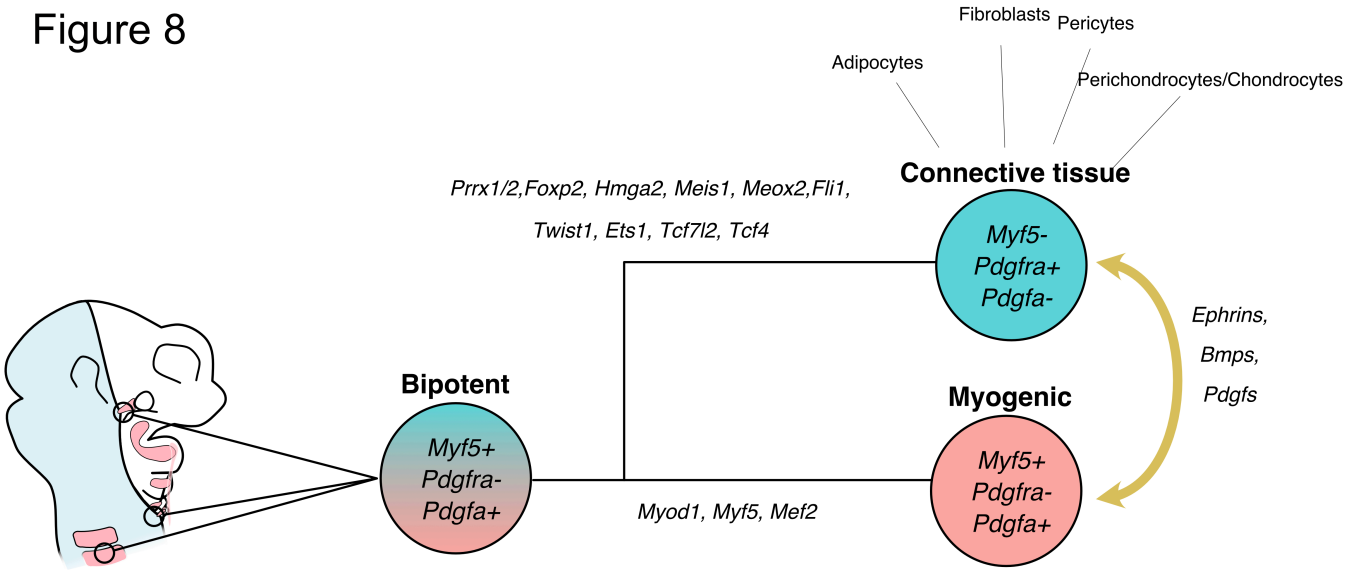
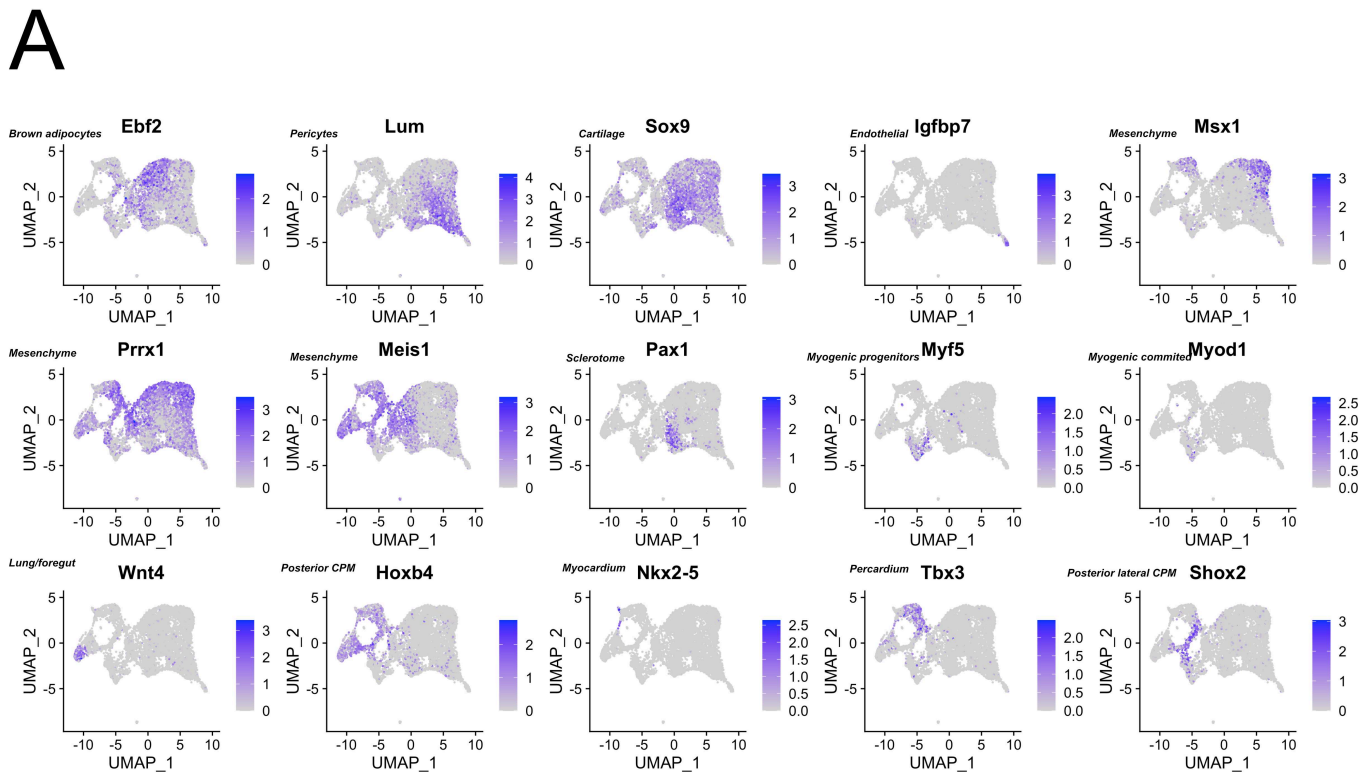
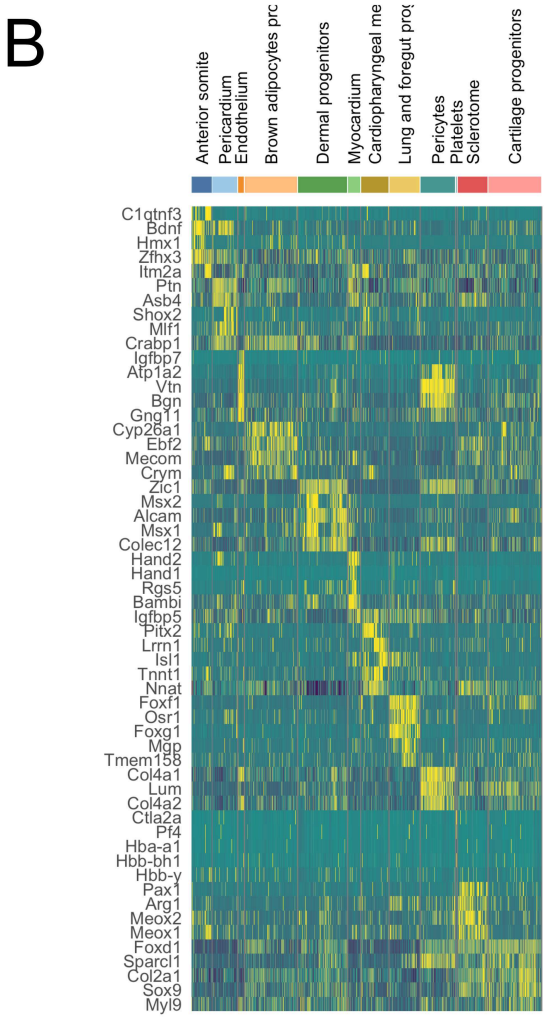

Fibroblasts
Pericytes
Adipocytes
Perichondrocytes/Chondrocytes
Connective tissue
Prrx1/2,Foxp2, Hmga2, Meis1, Meox2,Fli1,
Twist1, Ets1, Tcf7l2, Tcf4
Myf5-
Pdgfra+
Pdgfa-
Ephrins,
Bmps,
Pdgfs
Bipotent
Myf5+
Pdgfra-
Pdgfa+
Myogenic
Myf5+
Pdgfra-
Pdgfa+
Myod1, Myf5, Mef2

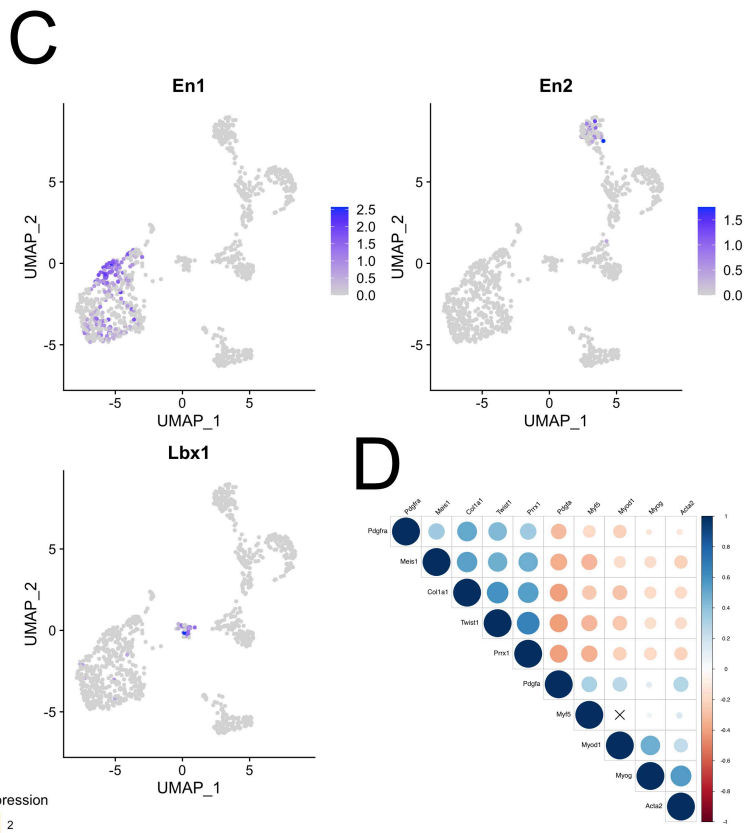
**Figure 8. Model of Myf5+ bipotent progenitors giving rise to muscle and associated connective tissues.**

Scheme illustrating the model of bipotent Myf5+/Pdgfa+ progenitors giving rise to myogenic and non-myogenic cells and discreet parts of the head, deprived of neural crest. Upon activation of a set of transcription factors including *Prrx1/2, Foxp2, Hmga2, Meis1, Meox2, Fli1, Twist1, Ets1, Tcf7l2* and *Tcf4*, a fibrogenic fate is acquired. A molecular dialogue is initiated at the branchpoint including extracellular matrix components and tyrosine kinase signalling such as *Pdgf*, *Ephrins* and *Bmps*.

# Figure S1


A
Brown adipocytes
Ebf2
Pericytes
Lum
Cartilage
Sox9
Endothelial
Igfbp7
Mesenchyme
Msx1
Mesenchyme
Prrx1
Mesenchyme
Meis1
Sclerotome
Pax1
Myogenic progenitors
Myf5
Myogenic commited
Myod1
Lung/foregut
Wnt4
Posterior CPM
Hoxb4
Myocardium
Nkx2-5
Percardium
Tbx3
Posterior lateral CPM
Shox2
UMAP_1
UMAP_2

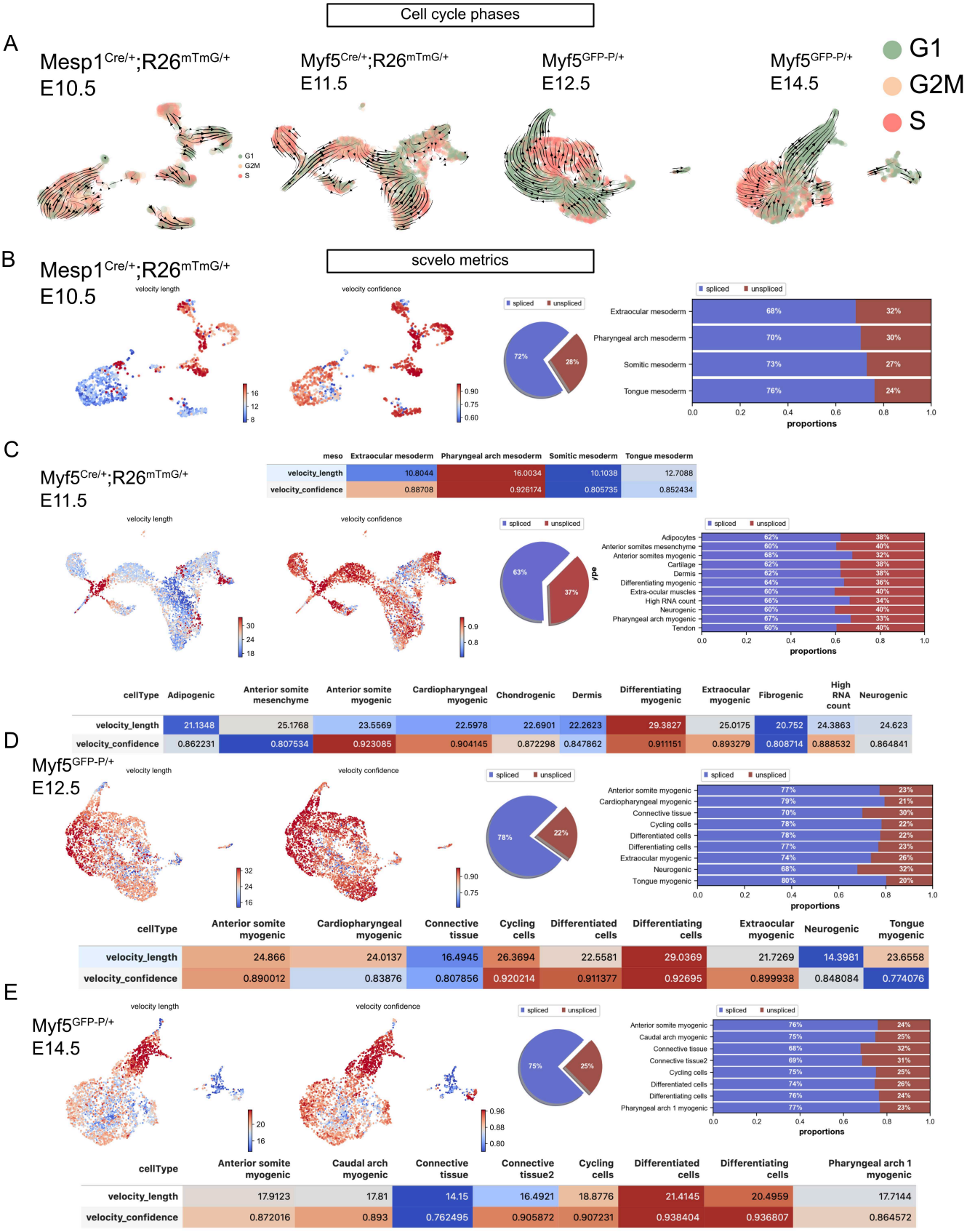

B
Anterior somite
Pericardium
Endothelium
Brown adipocytes prc
Dermal progenitors
Myocardium
Cardiopharyngeal me
Lung and foregut proc
Pericytes
Platelets
Sclerotome
Cartilage progenitors
C1qtnf3
Bdnf
Hmx1
Zfhx3
Itm2a
Ptn
Asb4
Shox2
Mlf1
Crabp1
Igfbp7
Atp1a2
Vtn
Bgn
Gng11
Cyp26a1
Ebf2
Mecom
Crym
Zic1
Msx2
Alcam
Msx1
Colec12
Hand2
Hand1
Rgs5
Bambi
Igfbp5
Pitx2
Lrrn1
Isl1
Tnnt1
Nnat
Foxf1
Osr1
Foxg1
Mgp
Tmem158
Col4a1
Lum
Col4a2
Ctla2a
Pf4
Hba-a1
Hbb-bh1
Hbb-y
Pax1
Arg1
Meox2
Meox1
Foxd1
Sparcl1
Col2a1
Sox9
Myl9
Expression


C
En1
En2
Lbx1
UMAP_1
UMAP_2
D

**Supplemental Figure S1. Myogenic and non-myogenic markers define anterior mesodermal tissues.**

(A) *Mesp1*$^{Cre/+}$;*R26*$^{mTmG/+}$ E10.5 UMAP expression plots of markers of various mesodermal lineages, assisting the identification of clusters. (B) Heatmap of top 5 markers of each cluster of *Mesp1*$^{Cre/+}$;*R26*$^{mTmG/+}$ E10.5. (C) UMAP expression plot of *Mesp1*$^{Cre/+}$;*R26*$^{mTmG/+}$ E10.5 subset of *En2*: a marker of pharyngeal arch 1, *En1*: a marker of epaxial somitic progenitors, *Lbx1*: a marker for tongue progenitors. (D) Correlation plot of myogenic (*Pdgfa, Myf5, Myod1, Myog, Acta2*) and non-myogenic (*Pdgfra, Prrx1, Meis1, Twist1, Osr1, Col1a1*) genes.

# Figure S2


Cell cycle phases
A
Mesp1Cre/+;R26mTmG/+
E10.5
Myf5Cre/+;R26mTmG/+
E11.5
Myf5GFP-P/+
E12.5
Myf5GFP-P/+
E14.5
G1
G2M
S
B
Mesp1Cre/+;R26mTmG/+
E10.5
scvelo metrics
velocity length
velocity confidence
spliced
unspliced
72%
28%
Extraocular mesoderm 68% 32%
Pharyngeal arch mesoderm 70% 30%
Somitic mesoderm 73% 27%
Tongue mesoderm 76% 24%
proportions
C
Myf5Cre/+;R26mTmG/+
E11.5
meso | Extraocular mesoderm | Pharyngeal arch mesoderm | Somitic mesoderm | Tongue mesoderm
velocity_length | 10.8044 | 16.0034 | 10.1038 | 12.7088
velocity_confidence | 0.88708 | 0.926174 | 0.805735 | 0.852434
63%
37%
type
Adipocytes 62% 38%
Anterior somites mesenchyme 60% 40%
Anterior somites myogenic 68% 32%
Cartilage 62% 38%
Dermis 62% 38%
Differentiating myogenic 64% 36%
Extra-ocular muscles 60% 40%
High RNA count 66% 34%
Neurogenic 60% 40%
Pharyngeal arch myogenic 67% 33%
Tendon 60% 40%
proportions
D
Myf5GFP-P/+
E12.5
cellType | Adipogenic | Anterior somite mesenchyme | Anterior somite myogenic | Cardiopharyngeal myogenic | Chondrogenic | Dermis | Differentiating myogenic | Extraocular myogenic | Fibrogenic | High RNA count | Neurogenic
velocity_length | 21.1348 | 25.1768 | 23.5569 | 22.5978 | 22.6901 | 22.2623 | 29.3827 | 25.0175 | 20.752 | 24.3863 | 24.623
velocity_confidence | 0.862231 | 0.807534 | 0.923085 | 0.904145 | 0.872298 | 0.847862 | 0.911151 | 0.893279 | 0.808714 | 0.888532 | 0.864841
78%
22%
Anterior somite myogenic 77% 23%
Cardiopharyngeal myogenic 79% 21%
Connective tissue 70% 30%
Cycling cells 78% 22%
Differentiated cells 78% 22%
Differentiating cells 77% 23%
Extraocular myogenic 74% 26%
Neurogenic 68% 32%
Tongue myogenic 80% 20%
proportions
cellType | Anterior somite myogenic | Cardiopharyngeal myogenic | Connective tissue | Cycling cells | Differentiated cells | Differentiating cells | Extraocular myogenic | Neurogenic | Tongue myogenic
velocity_length | 24.866 | 24.0137 | 16.4945 | 26.3694 | 22.5581 | 29.0369 | 21.7269 | 14.3981 | 23.6558
velocity_confidence | 0.890012 | 0.83876 | 0.807856 | 0.920214 | 0.911377 | 0.92695 | 0.899938 | 0.848084 | 0.774076
E
Myf5GFP-P/+
E14.5
75%
25%
Anterior somite myogenic 76% 24%
Caudal arch myogenic 75% 25%
Connective tissue 68% 32%
Connective tissue2 69% 31%
Cycling cells 75% 25%
Differentiated cells 74% 26%
Differentiating cells 76% 24%
Pharyngeal arch 1 myogenic 77% 23%
proportions
cellType | Anterior somite myogenic | Caudal arch myogenic | Connective tissue | Connective tissue2 | Cycling cells | Differentiated cells | Differentiating cells | Pharyngeal arch 1 myogenic
velocity_length | 17.9123 | 17.81 | 14.15 | 16.4921 | 18.8776 | 21.4145 | 20.4959 | 17.7144
velocity_confidence | 0.872016 | 0.893 | 0.762495 | 0.905872 | 0.907231 | 0.938404 | 0.936807 | 0.864572

**Supplemental Figure S2. Cell cycle phases and scvelo metrics.**

(A) UMAP of each dataset with overlaid velocity and cell cycle phase. Although cell cycle genes were regressed for all datasets, parts of the infered trajectories at later stages correlate with cell cycle phases, but not the myogenic to non-myogenic transition. (B-E) QC metrics of scvelo, including velocity length, velocity confidence and spliced/unspliced abundance per dataset and cell type. The velocity length refers to the speed at which the transition is happening and the velocity confidence measure the coherence of the directionality in this region.

# Figure S3

Myf5$^{Cre/+}$;R26$^{mTmG/+}$

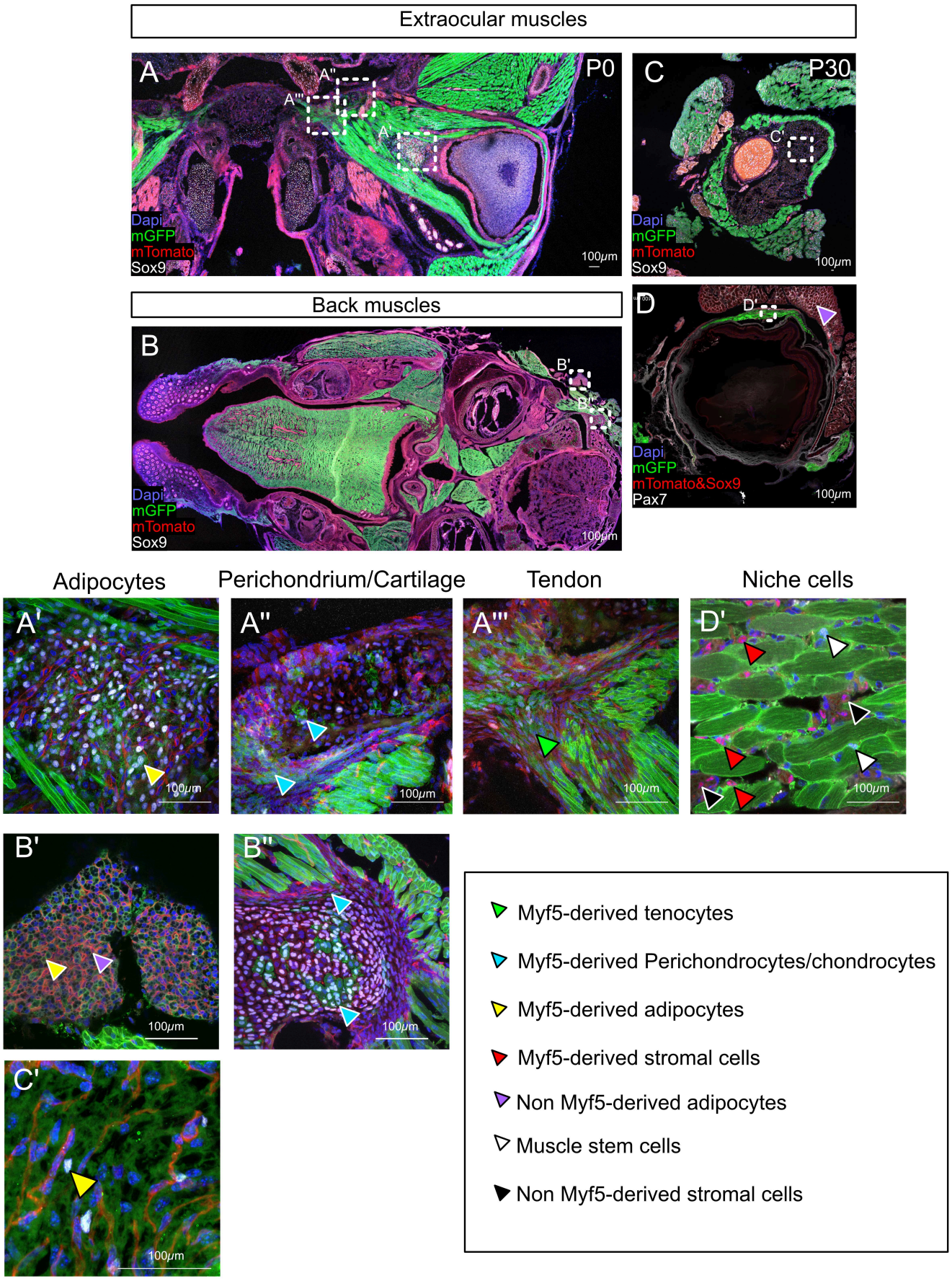
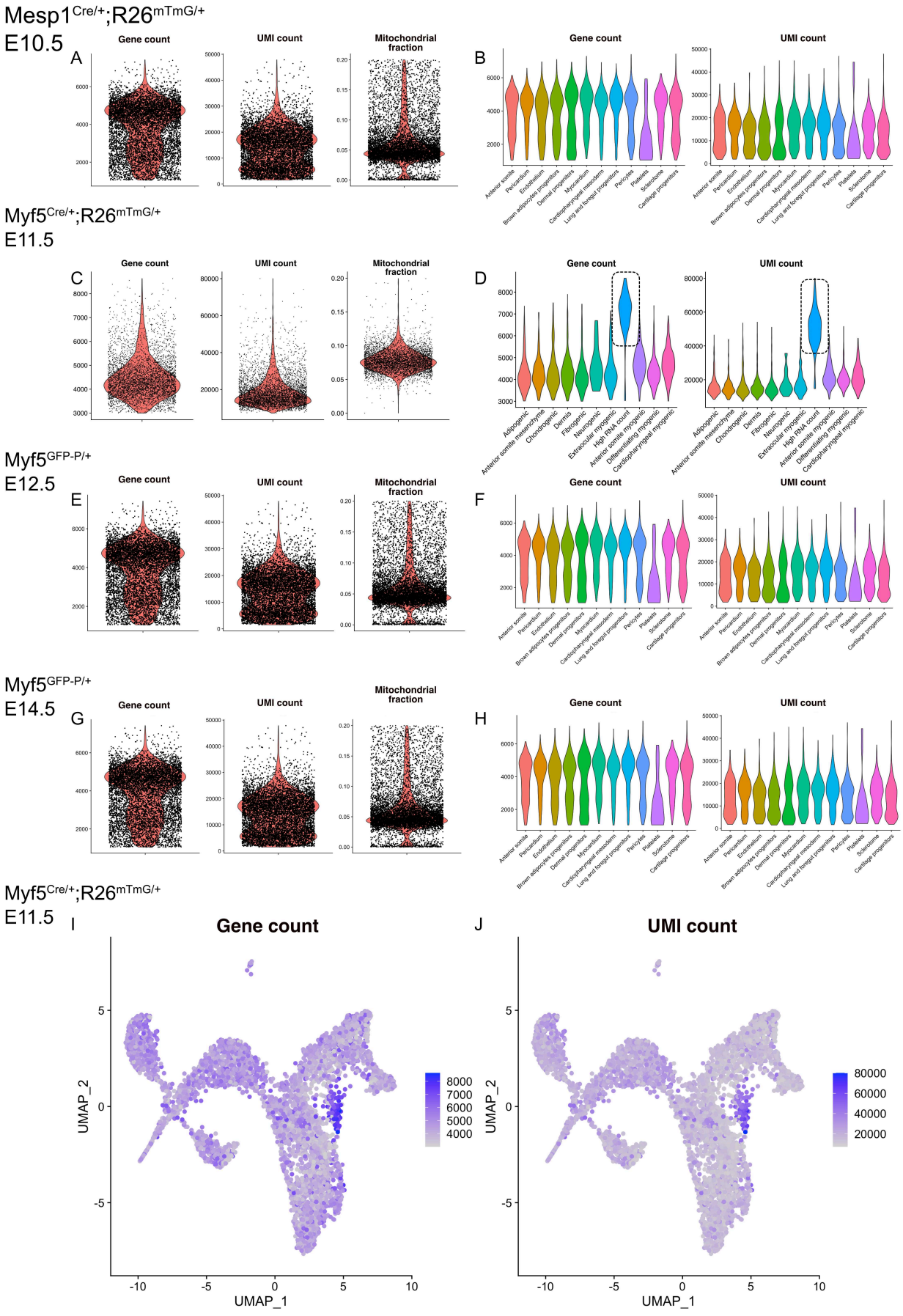


**Figure S3. The fate of *Myf5*-derived connective tissue perinatally.**

(A-B) Transverse sections of *Myf5*$^{Cre/+}$*; R26*$^{mTmG/+}$immunostained for *Sox9* (a marker of cartilage and some connective tissue) at P0 indicating regions of higher magnification in the EOM and back muscles. (C-D) Coronal sections of *Myf5*$^{Cre/+}$*; R26*$^{mTmG/+}$ EOM at P30 immunostained for *Pax7* and *Sox9* indicating the areas of higher magnifications. (A'-D', A''-B'', A''') High magnification of *Myf5* contribution to adipocytes, cartilage, perichondrium, tendon and niche cells.

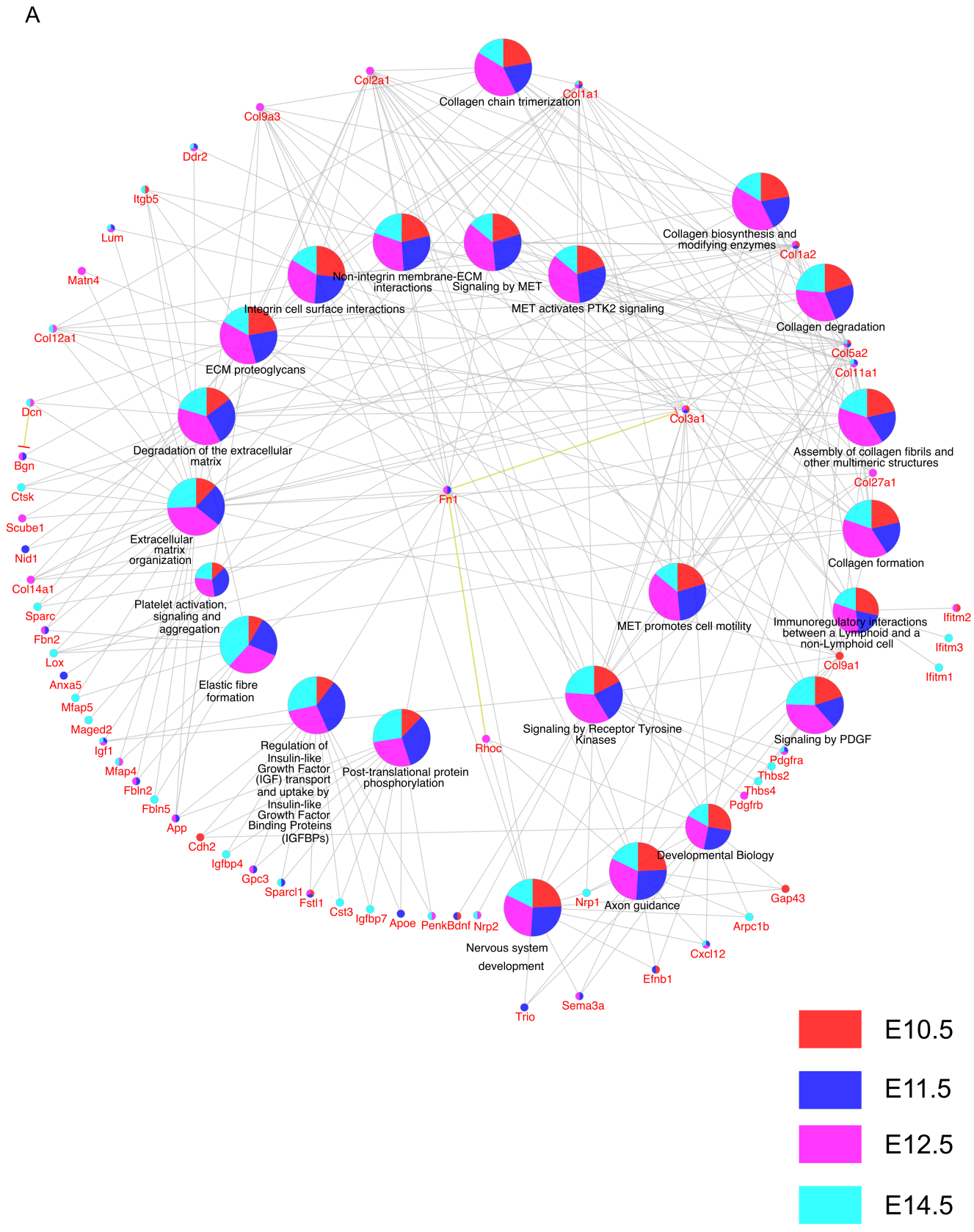
# Figure S4


Mesp1Cre/+;R26mTmG/+
E10.5
A
Gene count
UMI count
Mitochondrial fraction
B
Gene count
UMI count
Myf5Cre/+;R26mTmG/+
E11.5
C
Gene count
UMI count
Mitochondrial fraction
D
Gene count
UMI count
Myf5GFP-P/+
E12.5
E
Gene count
UMI count
Mitochondrial fraction
F
Gene count
UMI count
Myf5GFP-P/+
E14.5
G
Gene count
UMI count
Mitochondrial fraction
H
Gene count
UMI count
Myf5Cre/+;R26mTmG/+
E11.5
I
Gene count
J
UMI count
UMAP_1
UMAP_2

**Supplemental Figure S4. Seurat library pre-processing metrics.**

(A, C, E, G) Violin plots of gene count, UMI count and mitochondrial fraction for each dataset. (B, D, F, H) Gene count and UMI count per cell type for each dataset. Note the “High count” cluster of the E11.5 dataset. (I-J) UMAP of gene count and UMI count of the E11.5 dataset, highlighting the clustering of high UMI cells.

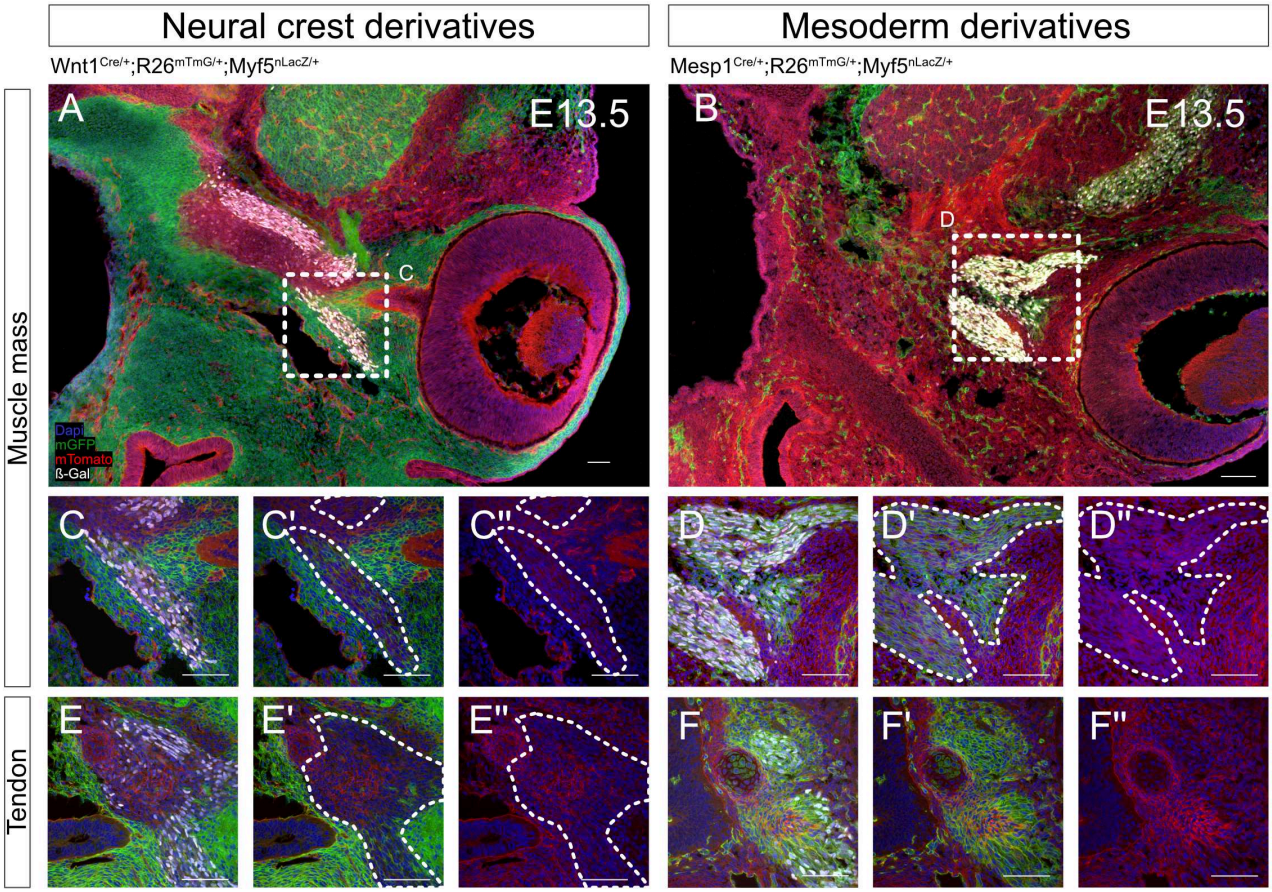
# Figure S5


A
Collagen chain trimerization
Col2a1
Col1a1
Col9a3
Ddr2
Itgb5
Lum
Matn4
Col12a1
Dcn
Bgn
Ctsk
Scube1
Nid1
Col14a1
Sparc
Fbn2
Lox
Anxa5
Mfap5
Maged2
Igf1
Mfap4
Fbln2
Fbln5
App
Cdh2
Igfbp4
Gpc3
Sparcl1
Fstl1
Cst3
Igfbp7
Apoe
PenkBdnf
Nrp2
Trio
Sema3a
Efnb1
Cxcl12
Arpc1b
Gap43
Nrp1
Axon guidance
Nervous system development
Developmental Biology
Pdgfrb
Thbs4
Thbs2
Pdgfra
Signaling by PDGF
Signaling by Receptor Tyrosine Kinases
Rhoc
Post-translational protein phosphorylation
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)
Elastic fibre formation
Platelet activation, signaling and aggregation
Extracellular matrix organization
Degradation of the extracellular matrix
ECM proteoglycans
Integrin cell surface interactions
Non-integrin membrane-ECM interactions
Signaling by MET
MET activates PTK2 signaling
Collagen biosynthesis and modifying enzymes
Col1a2
Collagen degradation
Col5a2
Col11a1
Col3a1
Assembly of collagen fibrils and other multimeric structures
Col27a1
Collagen formation
Fn1
MET promotes cell motility
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell
Col9a1
Ifitm2
Ifitm3
Ifitm1
E10.5
E11.5
E12.5
E14.5

**Supplemental Figure S5. Non-myogenic *Myf5*-derived cells display a similar gene set enrichment analysis.**

(A) Gene set enrichment analysis for Reactome pathways, including genes underlying each term, and their representation in each dataset. Although a slight variability in the specific genes of each stage, their related terms are similar.

Figure S6

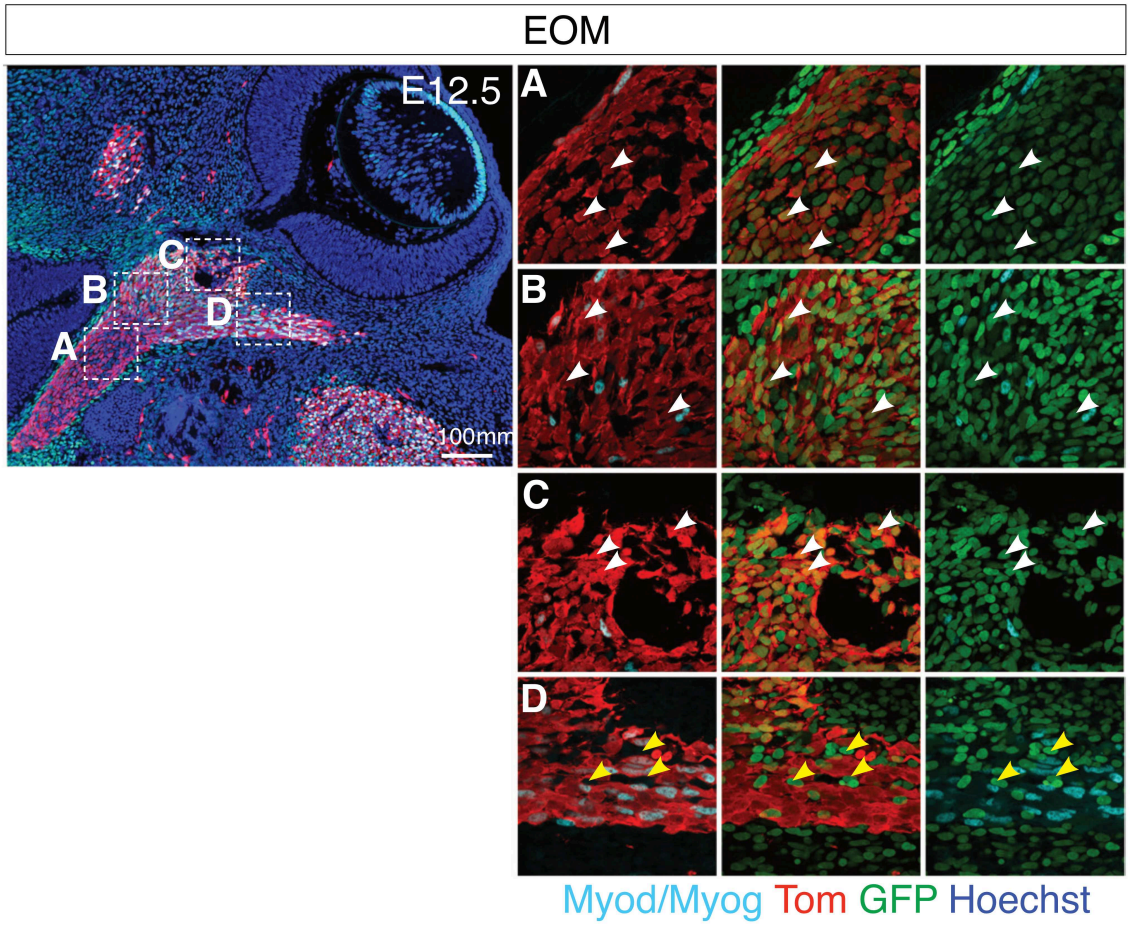

Neural crest derivatives
Mesoderm derivatives
Wnt1Cre/+;R26mTmG/+;Myf5nLacZ/+
Mesp1Cre/+;R26mTmG/+;Myf5nLacZ/+
Muscle mass
Tendon
A
E13.5
B
E13.5
C
D
Dapi
mGFP
mTomato
ß-Gal
C
C'
C''
D
D'
D''
E
E'
E''
F
F'
F''

**Supplemental Figure S6. Myf5-expressing cells are not neural crest derivatives in the EOM at E13.5.**

(A) Transverse cryosections of *Wnt1*$^{Cre/+}$*; R26*$^{mTmG/+}$*; Myf5*$^{nLacZ/+}$ and *Mesp1*$^{Cre/+}$*; R26*$^{mTmG/+}$*; Myf5*$^{nLacZ/+}$ immunostained for Bgal, at the level of the muscle mass of the EOM at E13.5. No Myf5-expressing neural crest-derived cell could be found. (B) Transverse cryosections of *Wnt1*$^{Cre/+}$*; R26*$^{mTmG/+}$*; Myf5*$^{nLacZ/+}$ and *Mesp1*$^{Cre/+}$*; R26*$^{mTmG/+}$*; Myf5*$^{nLacZ/+}$ immunostained for Bgal, at the level of the medial tendon attachment of the EOM at E13.5. No Myf5-expressing neural crest-derived cell could be found.

Figure S7

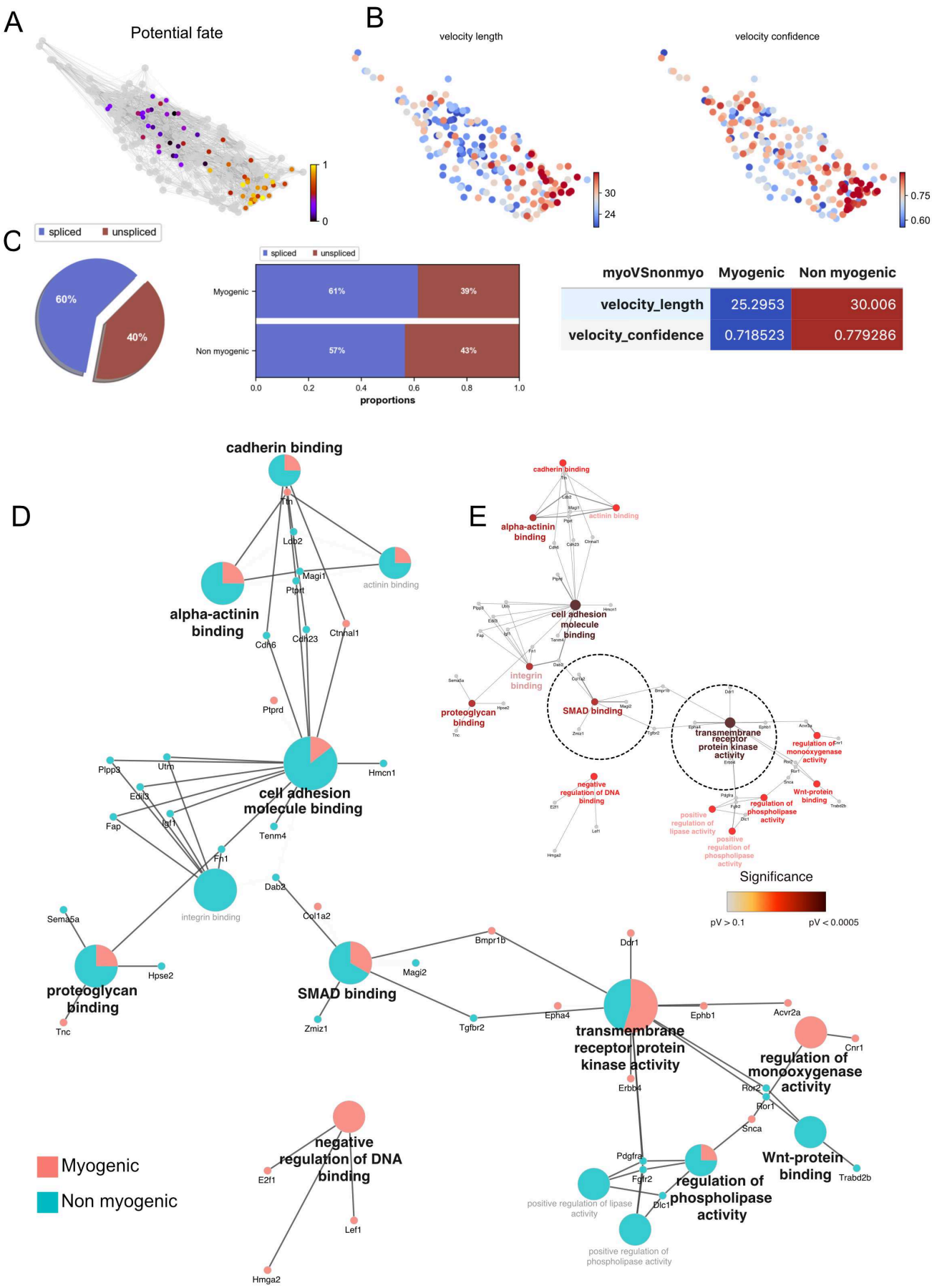
$Myf5^{Cre/+};R26^{TdTom/+};Pdgfra^{H2BGFP/+}$



**Supplemental Figure S7. *Myf5*-derived non-myogenic cells are found preferentially in the central mass in the EOM at E12.5.**

(A-D) Coronal sections of *Myf5*$^{Cre/+}$*; R26*$^{TdTomato/+}$*; Pdgfra*$^{H2BGFP/+}$ embryo at E12.5 highlighting *Myf5*-derived non-myogenic cells from medial attachment (A) to lateral fibres (D). Note the higher portion of double-positive cells in the central region.

# Figure S8


A
Potential fate
1
0
B
velocity length
30
24
velocity confidence
0.75
0.60
C
spliced
unspliced
60%
40%
spliced
unspliced
Myogenic
Non myogenic
61%
39%
57%
43%
0.0
0.2
0.4
0.6
0.8
1.0
proportions
myoVSnonmyo | Myogenic | Non myogenic
velocity_length | 25.2953 | 30.006
velocity_confidence | 0.718523 | 0.779286
D
cadherin binding
Ttn
Ldb2
Magi1
Ptprt
actinin binding
alpha-actinin binding
Cdh6
Cdh23
Ctnnal1
Ptprd
Plpp3
Utrn
Hmcn1
Edil3
cell adhesion molecule binding
Fap
Igf1
Tenm4
Fn1
integrin binding
Dab2
Sema5a
Col1a2
Bmpr1b
Ddr1
Hpse2
proteoglycan binding
Magi2
SMAD binding
Tnc
Zmiz1
Tgfbr2
Epha4
Ephb1
Acvr2a
transmembrane receptor protein kinase activity
Cnr1
regulation of monooxygenase activity
Erbb4
Ror2
Ror1
Snca
Wnt-protein binding
Pdgfra
Trabd2b
Fgfr2
regulation of phospholipase activity
Dlc1
positive regulation of lipase activity
positive regulation of phospholipase activity
negative regulation of DNA binding
E2f1
Lef1
Hmga2
Myogenic
Non myogenic
E
cadherin binding
alpha-actinin binding
actinin binding
cell adhesion molecule binding
integrin binding
proteoglycan binding
SMAD binding
transmembrane receptor protein kinase activity
regulation of monooxygenase activity
negative regulation of DNA binding
Wnt-protein binding
positive regulation of lipase activity
regulation of phospholipase activity
positive regulation of phospholipase activity
Significance
pV > 0.1
pV < 0.0005

**Supplemental Figure S8. EOM non-myogenic cells arise from a myogenic compartment and initiate crosstalk with myogenic cells.**

(A) Example of a potential fate of a unique cell. Scale gradient represents pseudotime.

(B) Velocity length and confidence. Higher confidence is found on both ends of the EOM cluster, and the highest length is found in the non-myogenic portion, indicating that the central region provides cells to both clusters, and that the tip of the non-myogenic cluster likely corresponds to a high determination state. (C) Unspliced and spliced proportions within the EOM subset and each subcluster indicating a relatively high number of unspliced variants in the non-myogenic cluster, which could suggest a high transcriptional activity. (D) Gene set enrichment analysis network, including relative contribution of each cluster to the term and significance levels.

# Table 1

## Driver genes

| E10.5 Anterior somites | E11.5 EOM Myo | E11.5 EOM Non myo | E12.5 Connective tissue | E14.5 Connective tissue1 |
|---|---|---|---|---|
| Tshz2 | Ccdc141 | Zfpm2 | Mgat4c | Dnm1 |
| Eya1 | Mcm6 | Ptxna4 | Cenpv | Pld1 |
| C1qtnf3 | Dync1i1 | Col23a1 | C130073E24Rik | Nrp2 |
| Meis2 | Tpm2 | Edil3 | Tbx3os1 | Ntrk3 |
| Limch1 | Celf2 | Map2 | E330013P04Rik | Tmem132c |
| Moxd1 | Sox6 | Rora | Stk26 | Egflam |
| Epha4 | Tnc | Sema5a | Edil3 | Gpr153 |
| Pitx2 | Magi3 | Colec12 | Fdft1 | Efemp1 |
| Parm1 | Sh3glb1 | Smoc1 | Lima1 | Adamts2 |
| Hpse2 | Parm1 | Ptprt | Trim59 | Brinp1 |
| Lrrn1 | Ephb1 | Ror1 | Meg3 | Vegfc |
| Dmrt2 | Bmpr1b | Dock5 | Gins3 | Twist2 |
| Myl3 | Hells | Map1b | Tpm2 | Itgb5 |
| Fap | Pdgfc | Fn1 | Cdh6 | Gria1 |
| Hs6st2 | Ptprd | Limch1 | Csmd3 | Sned1 |
| Ddr2 | Cnr1 | Tenm4 | Tceal5 | Sorcs3 |
| Cald1 | Sema3d | Rbms3 | Pclaf | Ebf2 |
| Prrx1 | Clcn5 | Srgap3 | Tspan9 | Fam19a1 |
| Magi3 | Chd7 | Tmem132c | Eps8 | Trabd2b |
| Ntn1 | Col25a1 | Sdc2 | Lmna | Ptxdc2 |
| Zfhx3 | Reep1 | Add3 | Dmrt2 | Sh3gl3 |
| Meis1 | Ctnnal1 | Pdgfra | Cpeb4 | Luzp2 |
| Tnni1 | Tpm1 | Gmds | Hpgd | Pdzd2 |
| Crym | Zim1 | St6galnac3 | Rcsd1 | Sema3e |
| Ebf1 | Lmx1a | Epb41l3 | Pdgfra | Rims1 |
| Nr2f1 | Neb | Pde3a | Ptac1 | Epha3 |
| Ntng1 | Atad2 | Tox | Palmd | Cyp7b1 |
| Pgm5 | Dapk2 | Smarca2 | Gucy1a1 | Gem |
| Cdh6 | Prox1 | Ctdspl | Wif1 | Ldb2 |
| Foxp1 | Lsamp | Magi2 | Naalad2 | Scube1 |
| Celf2 | Ttn | Dpysl3 | Smoc2 | Pdgfra |
| Tbx1 | Pls3 | Fgfr2 | Rassf4 | Pde1a |
| Bdnf | Slf2 | Ldb2 | Pttg1 | Nde1 |
| Colec12 | Vat1l | Igf1 | Josd2 | Enpp2 |
| Eya4 | E2f1 | Elk3 | Ptxna4 | Fam107b |
| Sobp | Epb41l2 | Zmiz1 | Eya2 | Stxbp6 |
| Peg3 | Gm28653 | Dlc1 | Nrsn1 | Rerg |
| Pdgfra | Lrrn1 | Nhs | Fign | Prex2 |
| Nrk | Mef2c | Cdkn1c | Inppl1 | Man1a |
| Ptn | St8sia2 | Plpp3 | Rnf152 | Tmem45a |
| Daam1 | Tshz1 | Ebf1 | Lasp1 | Sh3bp4 |
| Dlk1 | Wee1 | Sorbs2 | Mrln | Mcc |
| Unc5c | Slc24a3 | Baz1a | Cdt1 | Ncald |
| Lpar1 | Ncoa1 | Fat4 | Notch3 | Kdelr2 |
| Syne2 | Dek | Golgb1 | Pax3 | Pcdh19 |
| Nkd2 | Kdm5b | Hpse2 | Egfr | Gas7 |
| Brinp1 | Unc13c | Samd4 | Dbf4 | Cpt1c |
| Zfhx4 | Ddr1 | Itga9 | Bcr | Adam22 |
| Nnat | Pip4k2a | Magi1 | Mllt3 | Itgb8 |
| Gxylt2 | Fndc3c1 | Pcdh9 | Nectin1 | Dchs2 |
| Clmp | Rbm24 | Tgfbr2 | Grin3a | Cep350 |
| Ror2 | Rreb1 | Ntf3 | Cbfa2t3 | Oat |
| Nfia | Rragd | Col11a1 | Cdh2 | Rab30 |
| Ebf2 | Acsl3 | Runx1t1 | Anln | Aff2 |
| Ednra | Acvr2a | Tnrc18 | Ccdc6 | Gna14 |
| Fli1 | Zeb1 | Crym | Mcu | Slc29a1 |
| Tspan12 | Rgma | Fap | Fnip2 | Pls3 |
| Ttc28 | Arpp21 | Ppp1r1a | Kcnk13 | Traf3ip1 |
| Nfib | Lef1 | Tes | Sned1 | Rcsd1 |
| Ccdc88c | Nr2f2 | Bicc1 | Nde1 | Lgr4 |
| Col13a1 | Foxo1 | Il1rapl1 | Hipk3 | Zfp9 |
| 2700069I18Rik | Pdzm4 | Alcam | Arhgap11a | Hs3st5 |
| Pcolce | Hmga2 | 2700069I18Rik | Fam8a1 | Aspn |
| Scn3a | Lurap1l | Dab2 | Kif21a | Nrxn1 |
| Acvr2a | Pkig | Cntln | Mtss1 | Rrm1 |
| Auts2 | Ncl | Clmn | Abcd2 | Igfbp7 |
| Col3a1 | CT025619.1 | Rbms1 | Irx5 | Slc35f3 |
| Gap43 | Erbb4 | Tmem2 | Pacs2 | Kif15 |
| Mrln | Cdk14 | Cdh6 | Nab1 | Slc1a3 |
| Pax3 | Kif21a | Lypd6 | Ccnd2 | Bmp6 |
| Sim1 | Zfp704 | Mmp2 | Bok | Dkk2 |
| Epb41l2 | Nasp | Kif5c | Dok5 | Tspan9 |
| Ppp3ca | Plekha5 | Cadm2 | Ncapg | Ets1 |
| Tnfaip6 | Cap2 | Prkg2 | Rfx8 | Gria3 |
| Tmem132c | Snca | Cped1 | Fhod3 | Sox8 |
| Tmem2 | Epha4 | Dtl | Tk1 | Melk |
| Epb41l3 | Atad5 | Ror2 | Asf1b | Ntm |
| Crybg3 | Cntn3 | Utrn | Tek | Synpo2l |
| Nrxn1 | Cacna2d1 | Foxp1 | Arfgef3 | Hlf |
| Farp1 | Pak3 | L3mbtl3 | Rnf182 | Adamts5 |
| Sulf1 | Megf10 | Cdh23 | Kif14 | Ptcb4 |
| Tmtc2 | Tnnt1 | Negr1 | 1810041L15Rik | Cdc25b |
| Pde4dip | Acta2 | Hmcn1 | Rrm2 | Mgat4a |
| Phldb2 | Barx2 | Col26a1 | Fgf5 | Mdfic |
| Plpp3 | Mrln | Fbn2 | Barx2 | Trpc5 |
| Ybx3 | Pgm5 | Ankrd12 | Fli1 | Kif4 |
| Ppm1l | Fmr1 | Lhfp | Jph2 | Plce1 |
| Twist2 | Smc4 | Hs3st3b1 | Dtx4 | Il17rd |
| Nuak1 | Clmp | Adgrl3 | Ncald | Mmp16 |
| Tgfb2 | Alpk2 | Svil | Zic4 | Hhip |
| Sfrp1 | Kctd1 | Mob3b | Dlc1 | Tpx2 |
| Sncaip | Meg3 | Trabd2b | Cdc45 | Ndc80 |
| Tenm3 | Samd5 | Rmst | Gatm | Bub1b |
| Cdh2 | Nrk | Prrx1 | Ssc5d | Hmmr |
| Iqgap2 | Piezo2 | 5330434G04Rik | Phactr2 | Kank4 |
| App | Robo1 | Zfhx3 | Ppp1r14c | Tmeff2 |
| Pgam2 | Col1a2 | Foxp2 | Agl | Nr4a1 |
| Rspo3 | Cntrl | Mpp6 | Tox3 | Aurkb |
| Cdon | Mllt3 | Crispld1 | Aurka | Lrrtm3 |
| Ebf3 | Peg3 | Eya1 | Cdh15 | Cenpq |

**Table 1: Driver genes of each dataset.**

# Part 2

# Genetic birthmarks regulate cranial myogenic diversity

# Genetic birthmarks regulate cranial myogenic diversity

Alexandre Grimaldi[1,2] and Shahragim Tajbakhsh[1,2]

[1]Stem Cells & Development Unit, 25 rue du Dr. Roux, Institut Pasteur, 75015 Paris, France
[2]UMR CNRS 3738, Institut Pasteur, Paris, France

## INTRODUCTION

Skeletal muscles play critical and diverse roles including voluntary movement, feeding, speech and thermoregulation. An unexpected feature that has emerged in the last decades is the extent of heterogeneity that is characteristic of this tissue beyond fast and slow fibre types. Notably, adult skeletal muscle stem cells differ in their proliferative capacities (Stuelsatz et al., 2015), individual muscle groups are subject to specific evolutionary changes (Schubert et al., 2018; Smith, 1992), and muscle masses have distinct susceptibilities to myopathies (Sambasivan et al., 2009; Schoser et al., 2006; Formicola et al., 2014; Mercuri et al., 2013; Kaminski et al., 2002). This diversity can be traced back to development, where different muscle groups arising from distinct myogenic lineages require specific gene regulatory networks to initiate their myogenic programs (Buckingham, 2017; Comai and Tajbakhsh, 2014; Sambasivan et al., 2009). The genetic hierarchies governing cranial muscle development has been defined to a limited extent (Comai et al., 2019; Sambasivan et al., 2011b). Cranial muscle formation begins after most muscles in the trunk and limbs have been initiated, and in the case of the esophagus, this extends to postnatal stages. (Comai et al., 2019; Gopalakrishnan et al., 2015; Grimaldi et al., 2015; Heude et al., 2018; Nogueira et al., 2015; Tabler et al., 2017).

Some developmental programs required for myogenesis in diverse locations continue to be expressed in those same locations in adult muscle stem cells (Evano et al., 2020; Sambasivan et al., 2009; Yoshioka et al., 2020). In spite of extensive analysis of the muscle lineage, limited specific markers of the developing somitic and non-somitic muscle sub-groups have been identified to date (Comai et al., 2019; Heude et al., 2018).

Here, we use lineage analysis of multiple mouse lines, targeted microdissection, and advanced computational methods to build a 4D map of developing cranial muscles and focus on unique genetic programs that guide distinct cranial muscle groups. Using newly developed analytical approaches, we reveal the molecular switches that underpin the active maintenance of progenitor cell states. This process is presumably differentially regulated in each anlage, by a specific set of genes that are continuously expressed that we called “genetic birthmarks” (GBMs). Interestingly, some of these unique regulatory modules are maintained in adult muscle stem cells, while others are lost.

We are surveying the functional relevance of these traits for their ability to act as chromatin remodelers and regulate cell plasticity, extracellular remodeling and myogenic commitment. In doing so, we will provide the genetic foundation underlying skeletal muscle phenotypic diversity, and uncover key transcription factors that can be exploited for therapeutic strategies.

## RESULTS

### Craniofacial muscle initiation is asynchronous

Although recent studies have laid the foundation for identifying the major lineages and genetic programs for craniofacial muscle development (Heude et al., 2018), the morphogenic events and anatomical positioning of myogenic progenitors in the head that assure the reproducible disposition of the various muscle masses are largely unknown. To investigate the early compartmentalization and sequential appearance of myogenic progenitor populations, we performed a series of whole mount immunostainings using *Myf5*$^{Cre/+}$;*R26*$^{mTmG/+}$ embryos (the earliest known marker of myogenic commitment (Tajbakhsh et al., 1996b)) from E9.5 to E13.5 (Tajbakhsh et al., 1996) (Figure 1). We found that extraocular muscles (EOM) and the first and second pharyngeal arch (PA1, PA2) anlagen appear first at E9.5 (Figure 1A), whereas pharyngeal arch 6 muscles (laryngeal muscles) were only visible at E11 (Figure 1D). Notably, by E12.5-E13.5, considerable morphogenic remodeling has taken place, making the boundaries of each anlage difficult to discern (Figure S1). Combining these datapoints, we built a schematic model of the cranial muscle anlagen in relation to their adult derivatives. This model serves as a framework to overlay transcriptomic data.

To identify the genetic signatures of cranial muscles, we adopted an unbiased scRNAseq approach. However, given that craniofacial development is largely an asynchronous process that extends to late stages (Comai et al., 2019; Gopalakrishnan et al., 2015; Grimaldi et al., 2015; Heude et al., 2018; Nogueira et al., 2015), it is unlikely that we could capture all myogenic progenitors at a single timepoint. Therefore, we designed a series of experiments to cover crucial stages in craniofacial muscle development using different transgenic mouse lines and dissection protocols (Figure 1F). We first used *Mesp1*$^{Cre/+}$;*R26*$^{mTmG/+}$ reporter mice which represent a broad cell lineage

tracing strategy for anterior mesoderm at E10.5, to capture cells prior to commitment as well as their derivatives. Then, we used *Myf5*$^{Cre/+}$*; R26*$^{mTmG/+}$ reporter mice at E11.5 where cell-lineage tracing captures myogenic progenitors and their derivatives. Finally, we used *Myf5*$^{GFP\text{-}P/+}$ reporter mice at E12.5 and E14.5, which allow contemporary labelling of myogenic progenitors (Kassar-Duchossoy et al., 2004). All embryos from E10.5 to E12.5 included tissue from the entire region anterior to the forelimb, whereas E14.5 samples included tissues from the tongue, pharyngeal and laryngeal regions to focus on caudal arch-derived muscles (derivatives of PA3 to PA6) (Noden and Epstein, 2010).

**Somite derived muscles expand rapidly during cranial myogenesis**

As indicated, our scRNAseq pipeline generated 4 datasets that included *Mesp1*-traced, *Myf5*-traced, and Myf5-contemporary cells. The Mesp1 lineage includes cardiovasculature, adipocytes, dermis, endothelium and connective tissue (Noden and Trainor, 2005; Bildsoe et al., 2013; Adachi et al., 2020; Heude et al. 2018; Chai and Maxson, 2006). The Myf5 lineage contributes to other tissues, including neuronal cells, adipocytes, dermis, cartilage and connective tissue (Daubas et al., 2000; Haldar et al., 2008; Sebo et al., 2018; Stuelsatz et al., 2014).

To isolate myogenic cells, we resorted to known myogenic markers such as *Myf5* (Tajbakhsh et al., 1996b), and *Pdgf* signalling which segregates early myogenic versus non-myogenic cells (Grimaldi et al., in preparation). Strikingly, following myogenic filtering, multiple populations naturally separated in all 4 datasets (Figure 2). We used previously described markers to annotate each dataset identically: *Barx2* for PA1 (Jones et al., 1997), *Lbx1* for the tongue (Gross et al., 2000), *Pax3* for somitic neck muscles (Heude et al., 2018), *Isl1* for PA2 (Comai et al., 2019; Gopalakrishnan et al., 2015; Harel et al., 2009), *Alx4* for the EOM (Bothe and Dietrich, 2006), *Shox2/Meis2/Tshz1/Tshz2/Hoxb4/Hoxc4* (Coré et al., 2007; Sun et al., 2013; Wang et al., 2020) for caudal arches and *Myog* for differentiating cells (Hasty et al., 1993) (Figure 2A-D).

The relative proportion of each cluster was assessed at every stage (Figure 2A-D, piecharts). We noticed an increasingly predominant proportion of differentiated cells and somitic progenitors from E10.5 to E12.5, both from neck and tongue clusters, presumably due to the considerable increase in the size of these muscles at these stages (Figure 2E, S1) (Heude et al., 2018). In addition, half of the cells collected at E14.5 corresponded to tongue progenitors. This highlighted the importance of using a refined isolation protocol to yield an appropriate representation of each population for our study (Figure 2E). Cell cycle can bear significant weight in scRNAseq data (McDavid et al., 2016), and can be inferred from gene expression by taking advantage of the induction of specific genes at each phase of the cell cycle and scoring them (Tirosh et al., 2016). We noticed a marked increase of cells in G1 at E12.5, correlating with the large expansion of differentiating cells found close to somitic progenitors (Figure 2F).

**Myogenic development of cranial muscles is associated with the expression of unique and co-opted markers**

A handful of markers helped us annotate each anlage from E10.5 to E14.5 (from E10.5 to E12.5 for the EOM and somitic neck muscle, removed at E14.5). We chose to use these markers due to their persistent expression throughout development, yet some of them including *Pax3*, *Barx2* and *Isl1*, were also expressed in other myogenic compartments (Figure 2A-D). To investigate the persistence and specificity of other markers, we generated lists of top markers for each anlage, at each stage. We then combined these lists into a Sankey diagram, to visualize how many of these specific genes were kept or co-opted by other clusters between stages (Figure 3A).

Expectedly, markers defining differentiating cells were well conserved throughout development (Figure 3A, yellow edges). In contrast, between most progenitor compartments, it appeared that a significant portion of markers was shared with other clusters through various stages. This non-specificity could be indicative of 2 possible phenomena: 1) the sequential cell states that embryonic progenitors go through are similar among different anlagen, but asynchronous; 2) the substantial remodelling of the head displaces myogenic progenitors into new anatomical locations where divergent environmental cues will impinge on their transcriptomic state. While these hypotheses are not mutually exclusive, the cases of the EOM and the tongue might in part be explained by anatomical specificity.

Strikingly, EOM markers appeared to be well conserved from E10.5 to E12.5, consistent with the UMAP analyses where the EOM cluster always appeared as noticeably distinct (Figure 1A-D). Accordingly, the anatomical position of the EOM does not change overtly throughout development (Comai et al., 2020). It is thus tempting to speculate that its local molecular environment might also be coherent through development, hence promoting a transcriptomic output that is also coherent. In contrast, tongue myogenic progenitors migrate from the occipital somites to the tongue primordium through the hypoglossal chord (Han et al., 2012; Parada and Chai, 2015; Parada et al., 2012). These cells are displaced from the dorsal side of the embryo to the middle part of the mandibular prominence between E10.5 and E12.5 (Figure1 A) (Parada et al., 2012). This process might underlie the transition from the expression of anterior somitic markers to pharyngeal arch markers that appear from E11.5 to E12.5 (Figure 3A).

In spite of this, this analysis also revealed that a number of markers were consistently maintained in most clusters. In addition, some specific genes could be downregulated at certain key stages and expressed again at a later timepoint, and thus not being captured in this visualization method. Therefore, we compared the specific markers that were kept in at least one stage after E10.5 (Figure 3B-E) and represent them as “word clouds”, where the word size is representative of the

cumulative fold change across all stages (Figure 3B-E). We validated some of these markers by precise microdissection of the regions of interest, and qRT-PCR on E11.5 *Myf5*$^{Cre/+}$*; R26*$^{mTmG/+}$ embryos (Figure S2). Along with known markers, we uncovered a number of unknown markers which are conserved, highly expressed, and specific from the initiation of myogenesis until foetal stages. These included novel markers for extraocular muscles, pharyngeal arches and caudal arches. For example, the homeobox protein SHOX2 that is implicated in craniofacial and cardiovascular pathologies (Sun et al., 2013; Vickerman et al., 2011), appears as a developmental marker of the posterior muscles that are anatomically and clonally associated to the heart (Figure 3B). *Shox2*, *Meis2* and *Tshz2* were reported to be required for palate osteogenesis and soft palate development, muscles mostly derived from caudal arches (Vickerman et al., 2011; Wang et al., 2020). Importantly, published and unpublished work from our group have identified a number of these genes still expressed in adult muscle stem cells and after heterotypic transplantation (data not shown) (Evano et al., 2020; Sambasivan et al., 2009). We thus named them “genetic birthmarks” (GBMs).

**Core regulatory modules underlie cranial muscle diversity**

Most of the GBMs that we identified corresponded to transcription factors. To visualize the activity of these genes dynamically throughout different developmental stages, we merged the first three anatomically matching datasets (E10.5-E11.5-E12.5). As suspected, cells clustered preferentially by stage, cell cycle phase, and anatomical location (Figure 4A-B), thus validating the cluster annotations. Importantly, to assess the dynamic regulation of progenitors, cell cycle genes were not regressed in the combined dataset. We investigated the activity of GBMs using SCENIC, a toolkit for regulatory network inference (Aibar et al., 2017; Sande et al., 2020). This method leverages expression correlation and binding motif databases to infer regulatory relationships.

By regrouping transcription factors and their targets into “regulons”, we were able to measure their activity throughout development (Figure 4C). Several transcription factors were found among the most variable regulons (top 25 out of 99) including *Ebf1*, *Barx2*, *Dmrta2*, *Barhl1*, *Foxp2*, *Zic1*, *Pax3*, *En1*, *Prrx2*, *Shox2*, *Myog* and *Mef2c*. We found that most of the GBMs were active in their respective anatomical locations (Figure 4C). Intriguingly, those of the tongue and facial muscle progenitors (both migratory populations (Dietrich et al., 1999; Prunotto et al., 2004)) were not found as top active regulons. It is possible that the identity of these cells is not fixed as they receive different signals from their local environment.

Next, we interrogated the capacity of these transcription factors to coregulate within specific anatomical locations, thus providing insights into the underlying transcriptional mechanisms governing anatomical identity. We followed an original approach (detailed in Grimaldi et al., in preparation) to build a network of co-regulating transcription factors from SCENIC regulons of the

E11.5 dataset. Strikingly, various transcription factors including previously identified GBMS clustered into modules that were visually identifiable as groups of tightly interconnected nodes (Figure 4D). This result suggests that unique core regulatory systems are found in various cranial muscles during embryogenesis, and could thus contribute to establishing various lineages.

**Myogenic cells homogeneously maintain a progenitor state characterised by unique signalling pathways**

To investigate how these modules associated with GBMs can impart unique identity during myogenesis of distinct muscle groups, we first assessed the developmental trajectories of each myogenic cluster from E10.5 to E12.5 and at E14.5 (Figure 5A-B). To do so, we used scvelo, a recently described RNA velocity tool providing major improvements from previous methods (Bergen et al., 2020; Manno et al., 2018). Expectedly, we noticed 2 main directions in almost all clusters: differentiation, and proliferation/progenitor state (Figure 5A-E). Strikingly, when projected onto PC space (PC1/PC2), cells could be uniformly separated into “progenitor” versus “differentiating” categories (Figure 5E’-E’’). Strikingly, E10.5 cells were localized at the tip of the progenitor trajectory, and do not seem to integrate the “differentiating” half of the population. We annotated these 2 directional trajectories as “progenitor” and “committed” and used scvelo integrated functions to identify the driver genes (i.e. the genes responsible for most of the velocity calculated) underlying these 2 directional trajectories (Figure 5E’’-F). We build a gene set enrichment analysis network based on the top 100 driver genes of each population and uncovered the most relevant molecular functions associated with progenitor and committed states (Bindea et al., 2009). These 2 partitions yielded mostly distinct, non-overlapping terms, apart from terms associated with tyrosine kinase activity and Rho activity. Notably, progenitor driver genes were specifically associated with BMP/SMAD signalling, proteoglycan binding and extracellular matrix constituents. Different studies have reported a role for BMPs as inhibitors of myogenesis in favour of connective tissue. (El-Magd et al., 2013; Harel et al., 2009; Lima et al., 2020; Tzahor et al., 2003; Yamamoto-Shiraishi and Kuroiwa, 2013) suggesting that a similar mechanism is operating for progenitor maintenance during development.

## CONCLUSIONS AND PERSPECTIVES

By using genetically modified mouse lines to mark specific cell lineages, single-cell transcriptomics, anatomical studies, and original computational methods, we have identified distinct core regulatory modules for each craniofacial muscle in the mouse embryo. These genetic birthmarks are mostly retained throughout development and partially redeployed in adult skeletal muscle (data not shown). Characterisation of the molecular features underlying progenitor maintenance in the embryo revealed strikingly analogous features to a connective tissue/stromal cell states. In ongoing studies we will address how these genetic birthmark modules could directly regulate different aspects of myogenesis by transcriptionally controlling specific sets of genes. We are extending the analysis of GBM maintenance in the adult to various representative craniofacial muscles to evaluate the extent of the redeployment of developmental genes in adult stem cells, and their possible involvement in regulating susceptibility to diseases.

## MATERIALS & METHODS

### scRNAseq data generation

For E10.5 to E12.5 embryos, the cranial region above the forelimb was dissected in ice-cold 3% FBS and mechanically dissociated. The same procedure was applied at E14.5 but the dissection was refined to the pharyngeal and laryngeal regions. Tissues were then digested in TrypLE (ThermoFisher Cat #: 12604013) during 3 rounds of 5-minute incubation (37°C, 1400 RPM), interspersed with gentle pipetting to further dissociate the tissue. Cells were resuspended in FBS 3%, filtered, and incubated with Calcein blue (eBioscience, Cat #: 65-0855-39) and Propidium Iodide (ThermoFisher Cat #: P1304MP) for viability check. Viable cells were sorted on BD FACSAria™ III and manually counted using a hemocytometer. RNA integrity was assessed with Agilent Bioanalyzer 2100 to validate the isolation protocol prior to scRNAseq (RIN>8 was considered acceptable). 4000 to 13000 cells were loaded onto 10X Genomics Chromium microfluidic chip and cDNA libraries were generated following manufacturer's protocol. Concentrations and fragment sizes were measured using Agilent Bioanalyzer and Invitrogen Qubit. cDNA libraries were sequenced using NextSeq 500 and High Output v2.5 (75 cycles) kits. Genome mapping and count matrix generation were done following 10X Genomics Cell Ranger pipeline.

### RNA velocity and driver genes

RNA velocity analyses were performed using scvelo (Bergen et al., 2020) in python. This tool allows to infer velocity flow and driver genes using scRNAseq data, with major improvements from previous methods (Manno et al., 2018). First, unspliced and spliced transcript matrices were generated using velocyto (Manno et al., 2018) command line function, which outputs unspliced, spliced, and ambiguous matrices as a single loom file. These files were combined with filtered

Seurat objects to yield objects with unspliced and spliced matrices, as well as Seurat-generated annotations and cell-embeddings (UMAP, tSNE, PCA). These datasets were then processed following scvelo online guide and documentation. Velocity was calculated based on the dynamical model (using *scv.tl.recover_dynamics(adata)*, and *scv.tl.velocity(adata, mode='dynamical')*) and when outliers were detected, differential kinetics based on top driver genes were calculated and added to the model (using *scv.tl.velocity(adata, diff_kinetics=True)*). Specific driver genes were identified by determining the top likelihood genes in the selected cluster.

**Seurat preprocessing**

scRNAseq datasets were preprocessed using Seurat in R (https://satijalab.org/seurat/) (Butler et al., 2018). Cells with more than 20% of mitochondrial gene fraction were discarded. The number of genes expressed averaged to 4000 in all 4 datasets. Dimension reduction and UMAP generation were performed following Seurat workflow. Doublets were inferred using DoubletFinder v3 (McGinnis et al., 2019). Cell cycle genes were regressed in all datasets following Seurat dedicated vignette. We noticed that cell cycle regression, although clarifying anatomical diversity, seemed to induce low and high UMI clustering (Suppl. Fig. 3I-J). For the E10.5 and E11.5 datasets, 2 replicates were generated from littermates and merged after confirming their similitude. For subsequent datasets (E12.5 and E14.5), no replicates were used. Annotation and subsetting were also performed in Seurat. Myogenic cells were isolated based on their expression of Myf5 and Pdgfa and Pdgfra as described in (Grimaldi et al., in preparation).

Merging was performed using Seurat anchors (dims=1:30, anchor.features=40). Mitochondrial fraction, number of genes and number of UMI were regressed during scaling, but no cell cycle regression was applied.

**Gene regulatory network inference and transcription factor modules**

Gene regulatory networks were inferred using SCENIC (R implementation) and pySCENIC (Aibar et al., 2017; Sande et al., 2020) (python implementation). This algorithm allows regrouping of sets of correlated genes into regulons (i.e. a transcription factor and its targets) based on motif binding and co-expression. UMAP and heatmap were generated using regulon AUC matrix (Area Under Curve) which refers to the activity level of each regulon in each cell. The outputted list of each regulon and their targets was subsequently used to create a transcription factor network based on interactions involving genes that are regulons themselves. This operation greatly reduced the number of genes involved, while highlighting co-regulating transcriptional modules. 2 tables were generated: a node table and an edge table. The node table comprised all transcription factors involved, along with a "top50 regulon" annotation ("Top 50 myogenic", "Top 50 non-myogenic", or "Other"). The edge table was comprised of source and target IDs along with a column indicating the weight of the connection, which refers to the correlation of these 2 genes in the data, obtained

from SCENIC correlation matrix. These tables were imported in Gephi (https://gephi.org/), networks were generated and subjected to the “Force-Atlas2” algorithm.

**Gene set enrichment analysis analysis**

Gene set enrichment analysis analyses were performed on driver genes (obtained from scvelo), using Cluego (Bindea et al., 2009). “GO Molecular Pathway was used to identify common and unique pathways involved in each dataset. For statistical analyses, an enrichment/depletion two-sided hypergeometric test was performed and p-values were corrected using the Bonferroni step down method.

**Mouse strains**

Animals were handled according to European Community guidelines and the ethics committee of the Institut Pasteur (CETEA) approved protocols. The following strains were previously described: *Myf5*$^{Cre}$ (Haldar et al., 2008), *Mesp1*$^{Cre}$ (Saga et al., 1999), *Tg:Wnt1Cre* (Danielian et al., 1998), *R26*$^{TdTom}$ (Ai9;(Madisen et al., 2009)), *R26*$^{mTmG}$(Muzumdar et al., 2007), *Myf5*$^{nlacZ}$ (Tajbakhsh et al., 1996a), *Pdgfra*$^{H2BGFP}$ (Hamilton et al., 2003) and *Myf5*$^{GFP-P}$ (Kassar-Duchossoy et al., 2004).
To generate *Myf5*$^{Cre/+}$*; R26*$^{TdTomato/+}$*; Pdgfra*$^{H2BGFP/+}$embryos, *Myf5*$^{Cre/+}$ females were crossed with *Pdgfra*$^{H2BGFP/+}$*:: R26*$^{TdTomato/TdTomato}$ males. Mice were kept on a mixed genetic background C57BL/6JRj and DBA/2JRj (B6D2F1, Janvier Labs). Mouse embryos and fetuses were collected between embryonic day (E) E10.5 and E14.5, with noon on the day of the vaginal plug considered as E0.5.

**Whole-mount immunostainings**

For whole-mount immunostaining, were fixed 2.5h in 4% paraformaldehyde (Electron Microscopy Sciences, Cat #:15710) in PBS with 0,2% Triton X-100 at 4°C and washed overnight at 4°C in PBS. Then, embryos were dehydrated in 50% Methanol in PBS and twice in 100% Methanol, 30min each at RT and kept at -20˚C till needed. Heads were rehydrated, blocked overnight with 3%BSA, 10% Goat serum and 0.5% Triton in PBS. Primary and secondary immunostaining lasted 2 days each for E9.5 to E11.0 embryos, and 7 days each for E12.5 and E13.5 embryos. Antibodies were prepared in blocking solution, with enough volume to cover the entire tissue and place for several days at 4°C with agitation. After immunolabelling, samples washed in 0.1%Tween/PBS, dehydrated in 50% Methanol in PBS, 100% Methanol 30min each at room temperature, and cleared with a mix benzyl alcohol and benzyl benzoate (BABB). The clarified tissues were then mounted for imaging as described in (Yokomizo et al., 2012). Primary antibodies used were: chicken polyclonal anti-GFP (Abcam Cat. #: 13970, dilution (1:1000), mouse monoclonal IgG1 anti-Islet1 (DSHB Cat. #: 40.2D6, dilution 1:500), mouse IgG2A anti-Tuj1 (β3 tubulin) (Ozyme Cat. #. BLE801202, dilution: 1:1000).

**Image acquisition and processing**

Wholemount immunostainings ranging from E9.5 to E11.5 were acquired using Zeiss LSM700 and LSM780 confocal microscopes. E12.5 and E13.5 samples were acquired using a Lavision Biotech UltraMicroscope II light-sheet microscope. Images were processed using ZEN software (Carl Zeiss)

**Acknowledgements**

We acknowledge funding support from the Institut Pasteur, Association Française contre le Myopathies, Agence Nationale de la Recherche (Laboratoire d'Excellence Revive, Investissement d'Avenir; ANR-10-LABX-73), Association Française contre les Myopathies (Grant #20510), Fondation pour la Recherche Médicale (Grant # FDT201904008277), and the Centre National de la Recherche Scientifique. We gratefully acknowledge the UtechS Photonic BioImaging (Imagopole), C2RT, Institut Pasteur, supported by the French National Research Agency (France BioImaging; ANR-10–INSB–04; Investments for the Future).

**Competing interests**

The authors declare no competing interests.

**REFERENCES**

**Adachi, N., Bilio, M., Baldini, A. and Kelly, R. G.** (2020). Cardiopharyngeal mesoderm origins of musculoskeletal and connective tissues in the mammalian pharynx. *Development* **147**, dev185256.

**Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C. C., Geurts, P., Aerts, J., et al.** (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature methods* **14**, 1083–1086.

**Bergen, V., Lange, M., Peidli, S., Wolf, F. A. and Theis, F. J.** (2020). Generalizing RNA velocity to transient cell states through dynamical modelling. *Nat Biotechnol* 1–7.

**Bildsoe, H., Loebel, D. A. F., Jones, V. J., Hor, A. C. C., Braithwaite, A. W., Chen, Y.-T., Behringer, R. R. and Tam, P. P. L.** (2013). The mesenchymal architecture of the cranial mesoderm of mouse embryos is disrupted by the loss of Twist1 function. *Dev Biol* **374**, 295–307.

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

**Bothe, I. and Dietrich, S.** (2006). The molecular setup of the avian head mesoderm and its implication for craniofacial myogenesis. *Developmental Dynamics* **235**, 2845–2860.

**Buckingham, M.** (2017). Gene regulatory networks and cell lineages that underlie the formation of skeletal muscle. *Proc National Acad Sci* **114**, 5830–5837.

**Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R.** (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420.

**Chai, Y. and Maxson, R. E.** (2006). Recent advances in craniofacial morphogenesis. *Developmental Dynamics* **235**, 2353–2375.

**Comai, G. and Tajbakhsh, S.** (2014). Current Topics in Developmental Biology. *Curr Top Dev Biol* **110**, 1–73.

**Comai, G., Heude, E., Mella, S., Paisant, S., Pala, F., Gallardo, M., Langa, F., Kardon, G., Gopalakrishnan, S. and Tajbakhsh, S.** (2019). A distinct cardiopharyngeal mesoderm genetic hierarchy establishes antero-posterior patterning of esophagus striated muscle. *Elife* **8**, e47460.

**Comai, G., Tesarova, M., Dupé, V., Rhinn, M., Garcia, P. V., Silva, F. da, Feret, B., Exelby, K., Dollé, P., Carlsson, L., et al.** (2020). Local retinoic acid directs emergence of the extraocular muscle functional unit. *Biorxiv* 2020.01.07.897694.

**Coré, N., Caubit, X., Metchat, A., Boned, A., Djabali, M. and Fasano, L.** (2007). Tshz1 is required for axial skeleton, soft palate and middle ear development in mice. *Dev Biol* **308**, 407–420.

**Danielian, P. S., Muccino, D., Rowitch, D. H., Michael, S. K. and McMahon, A. P.** (1998). Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. *Curr Biol* **8**, 1323-S2.

**Daubas, P., Tajbakhsh, S., Hadchouel, J., Primig, M. and Buckingham, M.** (2000). Myf5 is a novel early axonal marker in the mouse brain and is subjected to post-transcriptional regulation in neurons. *Dev Camb Engl* **127**, 319–31.

**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 (Cambridge, England)* **126**, 1621–9.

**El-Magd, M. A., Allen, S., McGonnell, I., Otto, A. and Patel, K.** (2013). Bmp4 regulates chick Ebf2 and Ebf3 gene expression in somite development. *Dev Growth Differ* **55**, 710–722.

**Evano, B., Gill, D., Hernando-Herraez, I., Comai, G., Stubbs, T. M., Commere, P.-H., Reik, W. and Tajbakhsh, S.** (2020). Transcriptome and epigenome diversity and plasticity of muscle stem cells following transplantation. *Biorxiv* 2020.05.20.107219.

**Gopalakrishnan, S., Comai, G., Sambasivan, R., Francou, A., Kelly, R. G. and Tajbakhsh, S.** (2015). A Cranial Mesoderm Origin for Esophagus Striated Muscles. *Developmental cell* **34**, 694–704.

**Grimaldi, A., Parada, C. and Chai, Y.** (2015). A Comprehensive Study of Soft Palate Development in Mice. *PloS one* **10**, e0145018.

**Gross, M. K., Moran-Rivard, L., Velasquez, T., Nakatsu, M. N., Jagla, K. and Goulding, M.** (2000). Lbx1 is required for muscle precursor migration along a lateral pathway into the limb. *Dev Camb Engl* **127**, 413–24.

**Haldar, M., Karan, G., Tvrdik, P. and Capecchi, M. R.** (2008). Two cell lineages, myf5 and myf5-independent, participate in mouse skeletal myogenesis. *Dev Cell* **14**, 437–45.

**Hamilton, T. G., Klinghoffer, R. A., Corrin, P. D. and Soriano, P.** (2003). Evolutionary Divergence of Platelet-Derived Growth Factor Alpha Receptor Signaling Mechanisms. *Mol Cell Biol* **23**, 4013–4025.

**Han, D., Zhao, H., Parada, C., Hacia, J. G., Bringas, P. and Chai, Y.** (2012). A TGFβ-Smad4-Fgf6 signaling cascade controls myogenic differentiation and myoblast fusion during tongue development. *Dev Camb Engl* **139**, 1640–50.

**Harel, I., Nathan, E., Tirosh-Finkel, L., Zigdon, H., Guimarães-Camboa, N., Evans, S. M. and Tzahor, E.** (2009). Distinct Origins and Genetic Programs of Head Muscle Satellite Cells. *Dev Cell* **16**, 822–832.

**Hasty, P., Bradley, A., Morris, J. H., Edmondson, D. G., Venuti, J. M., Olson, E. N. and Klein, W. H.** (1993). Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature* **364**, 501–506.

**Heude, E., Tesarova, M., Sefton, E. M., Jullian, E., Adachi, N., Grimaldi, A., Zikmund, T., Kaiser, J., Kardon, G., Kelly, R. G., et al.** (2018). Unique morphogenetic signatures define mammalian neck muscles and associated connective tissues. *eLife* **7**,.

**Jones, F. S., Kioussi, C., Copertino, D. W., Kallunki, P., Holst, B. D. and Edelman, G. M.** (1997). Barx2, a new homeobox gene of the Bar class, is expressed in neural and craniofacial structures during development. *Proc National Acad Sci* **94**, 2632–2637.

**Kassar-Duchossoy, L., Gayraud-Morel, B., Gomès, D., Rocancourt, D., Buckingham, M., Shinin, V. and Tajbakhsh, S.** (2004). Mrf4 determines skeletal muscle identity in Myf5:Myod double-mutant mice. *Nature* **431**, 466–471.

**Lima, J. E. de, Blavet, C., Bonnin, M.-A., Hirsinger, E., Comai, G., Yvernogeau, L., Bellenger, L., Mella, S., Nassari, S., Robin, C., et al.** (2020). BMP signalling directs a fibroblast-to-myoblast conversion at the connective tissue/muscle interface to pattern limb muscles. *Biorxiv* 2020.07.20.211342.

**Madisen, L., Zwingman, T. A., Sunkin, S. M., Oh, S. W., Zariwala, H. A., Gu, H., Ng, L. L., Palmiter, R. D., Hawrylycz, M. J., Jones, A. R., et al.** (2009). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* **13**, 133–40.

**Manno, G. L., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M. E., Lönnerberg, P., Furlan, A., et al.** (2018). RNA velocity of single cells. *Nature* **560**, 494–498.

**McDavid, A., Finak, G. and Gottardo, R.** (2016). The contribution of cell cycle to heterogeneity in single-cell RNA-seq data. *Nat Biotechnol* **34**, 591–593.

**McGinnis, C. S., Murrow, L. M. and Gartner, Z. J.** (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337.e4.

**Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. and Luo, L.** (2007). A global double-fluorescent Cre reporter mouse. *Genesis (New York, N.Y. : 2000)* **45**, 593–605.

**Noden, D. M. and Epstein, M. L.** (2010). Embryonic origins of avian and mammalian laryngeal musculoskeletal structures. *Faseb J* **24**, 172.1-172.1.

**Noden, D. M. and Trainor, P. A.** (2005). Relations and interactions between cranial mesoderm and neural crest populations. *Journal of anatomy* **207**, 575–601.

**Nogueira, J. M., Hawrot, K., Sharpe, C., Noble, A., Wood, W. M., Jorge, E. C., Goldhamer, D. J., Kardon, G. and Dietrich, S.** (2015). The emergence of Pax7-expressing muscle stem cells during vertebrate head muscle development. *Front Aging Neurosci* **7**, 62.

**Parada, C. and Chai, Y.** (2015). Mandible and Tongue Development. *Current topics in developmental biology* **115**, 31–58.

**Parada, C., Han, D. and Chai, Y.** (2012). Molecular and cellular regulatory mechanisms of tongue myogenesis. *Journal of dental research* **91**, 528–35.

**Prunotto, C., Crepaldi, T., Forni, P. E., Ieraci, A., Kelly, R. G., Tajbakhsh, S., Buckingham, M. and Ponzetto, C.** (2004). Analysis of Mlc-lacZ Met mutants highlights the essential function of Met for migratory precursors of hypaxial muscles and reveals a role for Met in the development of hyoid arch-derived facial muscles. *Developmental dynamics* **231**, 582–91.

**Saga, Y., Miyagawa-Tomita, S., Takagi, A., Kitajima, S., Miyazaki, J. i and Inoue, T.** (1999). MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Dev Camb Engl* **126**, 3437–47.

**Sambasivan, R., Gayraud-Morel, B., Dumas, G., Cimper, C., Paisant, S., Kelly, R. G., Kelly, R. and Tajbakhsh, S.** (2009). Distinct regulatory cascades govern extraocular and pharyngeal arch muscle progenitor cell fates. *Developmental cell* **16**, 810–21.

**Sambasivan, R., Kuratani, S. and Tajbakhsh, S.** (2011). An eye on the head: the development and evolution of craniofacial muscles. *Development* **138**, 2401–2415.

**Sande, B. V. de, Flerin, C., Davie, K., Waegeneer, M. D., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., et al.** (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* 1–30.

**Schubert, F. R., Singh, A. J., Afoyalan, O., Kioussi, C. and Dietrich, S.** (2018). To roll the eyes and snap a bite - function, development and evolution of craniofacial muscles. *Seminars in cell & developmental biology*.

**Sebo, Z. L., Jeffery, E., Holtrup, B. and Rodeheffer, M. S.** (2018). A mesodermal fate map for adipose tissue. *Development* **145**, dev166801.

**Smith, K. K.** (1992). The evolution of the mammalian pharynx. *Zool J Linn Soc-lond* **104**, 313–349.

**Stuelsatz, P., Shearer, A. and Yablonka-Reuveni, Z.** (2014). Ancestral Myf5 gene activity in periocular connective tissue identifies a subset of fibro/adipogenic progenitors but does not connote a myogenic origin. *Developmental biology* **385**, 366–79.

**Stuelsatz, P., Shearer, A., Li, Y., Muir, L. A., Ieronimakis, N., Shen, Q. W., Kirillova, I. and Yablonka-Reuveni, Z.** (2015). Extraocular muscle satellite cells are high performance myo-engines retaining efficient regenerative capacity in dystrophin deficiency. *Dev Biol* **397**, 31–44.

**Sun, C., Zhang, T., Liu, C., Gu, S. and Chen, Y.** (2013). Generation of Shox2-Cre allele for tissue specific manipulation of genes in the developing heart, palate, and limb. *Genesis* **51**, 515–522.

**Tabler, J. M., Rigney, M. M., Berman, G. J., Gopalakrishnan, S., Heude, E., Al-Lami, H. A., Yannakoudakis, B. Z., Fitch, R. D., Carter, C. M., Vokes, S. A., et al.** (2017). Cilia-mediated Hedgehog signaling controls form and function in the mammalian larynx. *Elife* **6**, e19153.

**Tajbakhsh, S., Bober, E., Babinet, C., Pournin, S., Arnold, H. and Buckingham, M.** (1996). Gene targeting the myf-5 locus with nlacZ reveals expression of this myogenic factor in mature skeletal muscle fibres as well as early embryonic muscle. *Developmental dynamics* **206**, 291–300.

**Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al.** (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Sci New York N Y* **352**, 189–96.

**Tzahor, E., Kempf, H., Mootoosamy, R. C., Poon, A. C., Abzhanov, A., Tabin, C. J., Dietrich, S. and Lassar, A. B.** (2003). Antagonists of Wnt and BMP signaling promote the formation of vertebrate head muscle. *Gene Dev* **17**, 3087–3099.

**Vickerman, L., Neufeld, S. and Cobb, J.** (2011). Shox2 function couples neural, muscular and skeletal development in the proximal forelimb. *Dev Biol* **350**, 323–336.

**Wang, L., Tang, Q., Xu, J., Li, H., Yang, T., Li, L., Machon, O., Hu, T. and Chen, Y.** (2020). The transcriptional regulator MEIS2 sets up the ground state for palatal osteogenesis in mice. *J Biol Chem* **295**, 5449–5460.

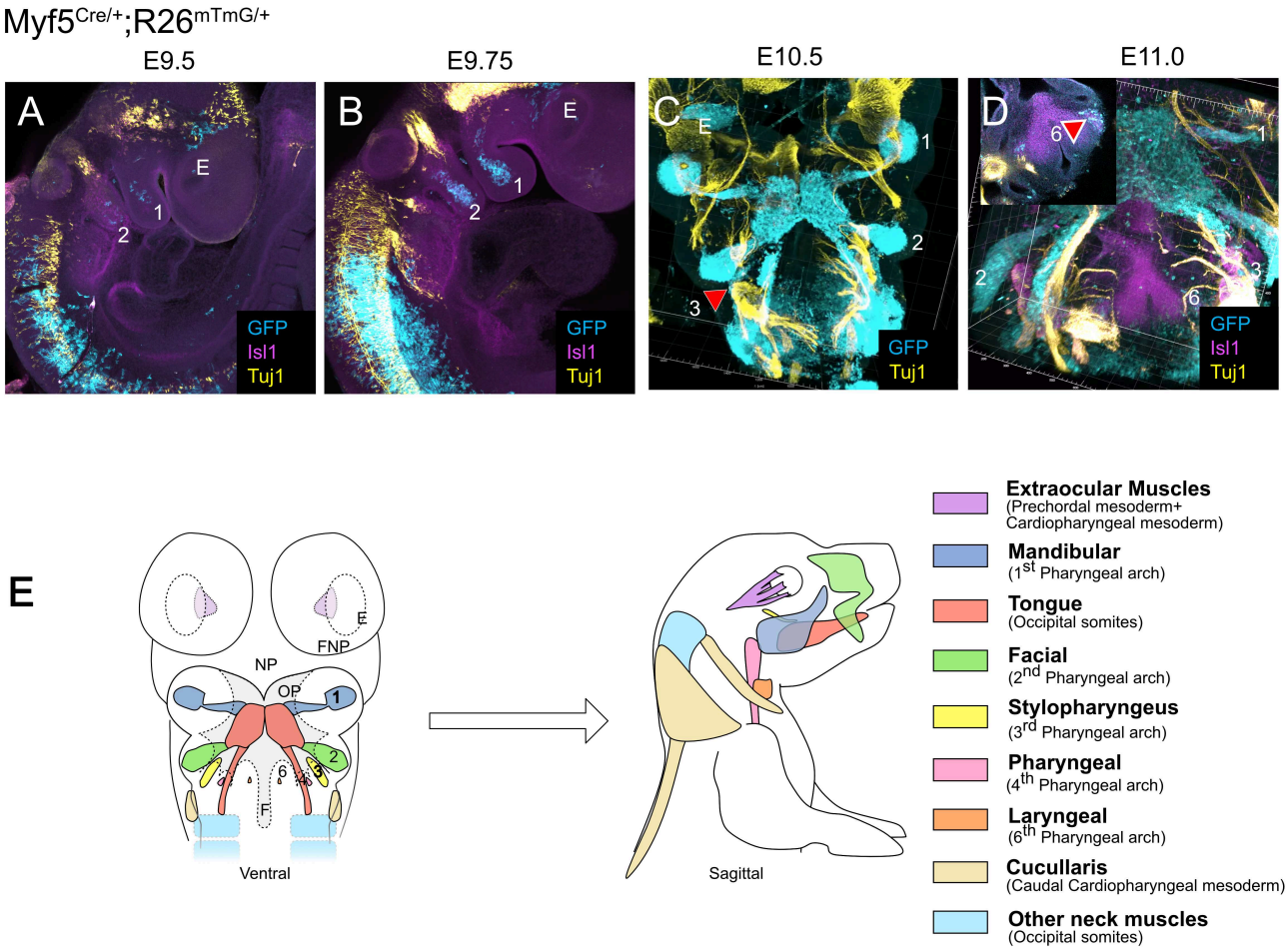
**Yamamoto-Shiraishi, Y. and Kuroiwa, A.** (2013). Wnt and BMP signaling cooperate with Hox in the control of Six2 expression in limb tendon precursor. *Dev Biol* **377**, 363–374.

**Yokomizo, T., Yamada-Inagawa, T., Yzaguirre, A. D., Chen, M. J., Speck, N. A. and Dzierzak, E.** (2012). Whole-mount three-dimensional imaging of internally localized immunostained cells within mouse embryos. *Nat Protoc* **7**, 421–431.

**Yoshioka, K., Nagahisa, H., Miura, F., Araki, H., Kamei, Y., Kitajima, Y., Seko, D., Nogami, J., Tsuchiya, Y., Okazaki, N., et al.** (2020). Hoxa10 mediates positional memory to govern stem cell function in adult skeletal muscle. *Biorxiv* 2020.07.16.207654.

512

# Figure 1


Myf5Cre/+;R26mTmG/+
E9.5
E9.75
E10.5
E11.0
A
B
C
D
E
1
2
3
6
GFP
Isl1
Tuj1
GFP
Tuj1
E
NP
FNP
OP
Ventral
Sagittal
Extraocular Muscles
(Prechordal mesoderm+ Cardiopharyngeal mesoderm)
Mandibular
(1st Pharyngeal arch)
Tongue
(Occipital somites)
Facial
(2nd Pharyngeal arch)
Stylopharyngeus
(3rd Pharyngeal arch)
Pharyngeal
(4th Pharyngeal arch)
Laryngeal
(6th Pharyngeal arch)
Cucullaris
(Caudal Cardiopharyngeal mesoderm)
Other neck muscles
(Occipital somites)

E

F

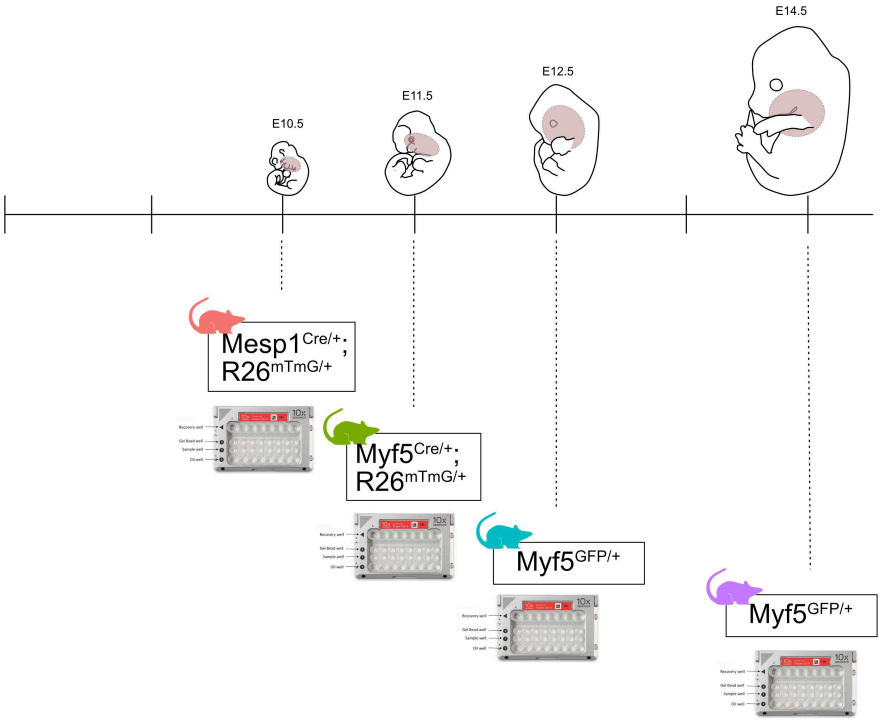

E10.5
E11.5
E12.5
E14.5
Mesp1Cre/+; R26mTmG/+
Myf5Cre/+; R26mTmG/+
Myf5GFP/+
Myf5GFP/+

**Figure 1. Rationale and experimental design.**

(A-D) Whole-mount immunostaining of *Myf5*$^{Cre/+}$;*R26*$^{mTmG/+}$ embryos from E9.5 to E11.0, labelled for GFP (myogenic anlagen), Isl1 (CPM, endoderm), and Tuj1 (neurons), and highlighting the progressive anteroposterior appearance of myogenic progenitors. 1/2/3/6: Pharyngeal arch 1/2/3/4, E: Eye. Arrowhead shows myogenic primordium. (E) Scheme of cranial myogenic compartments in the embryo and some of their derivatives in the adult. (F) Scheme of scRNAseq experimental design highlighting the stages, dissected area, and mouse line used for each dataset. *Mesp1*$^{Cre/+}$;*R26*$^{mTmG/+}$: broad cell lineage tracing of anterior mesoderm capturing pre-committed cells, *Myf5*$^{Cre/+}$;*R26*$^{mTmG/+}$: cell lineage tracing capturing derivatives of myogenic progenitors, *Myf5*$^{GFP\text{-}P/+}$: Contemporary labelling of myogenic progenitors. E10.5 to E12.5 include tissue from all anterior region above the forelimb. E14.5 includes tissue from the tongue, pharyngeal and laryngeal regions.

Figure 2

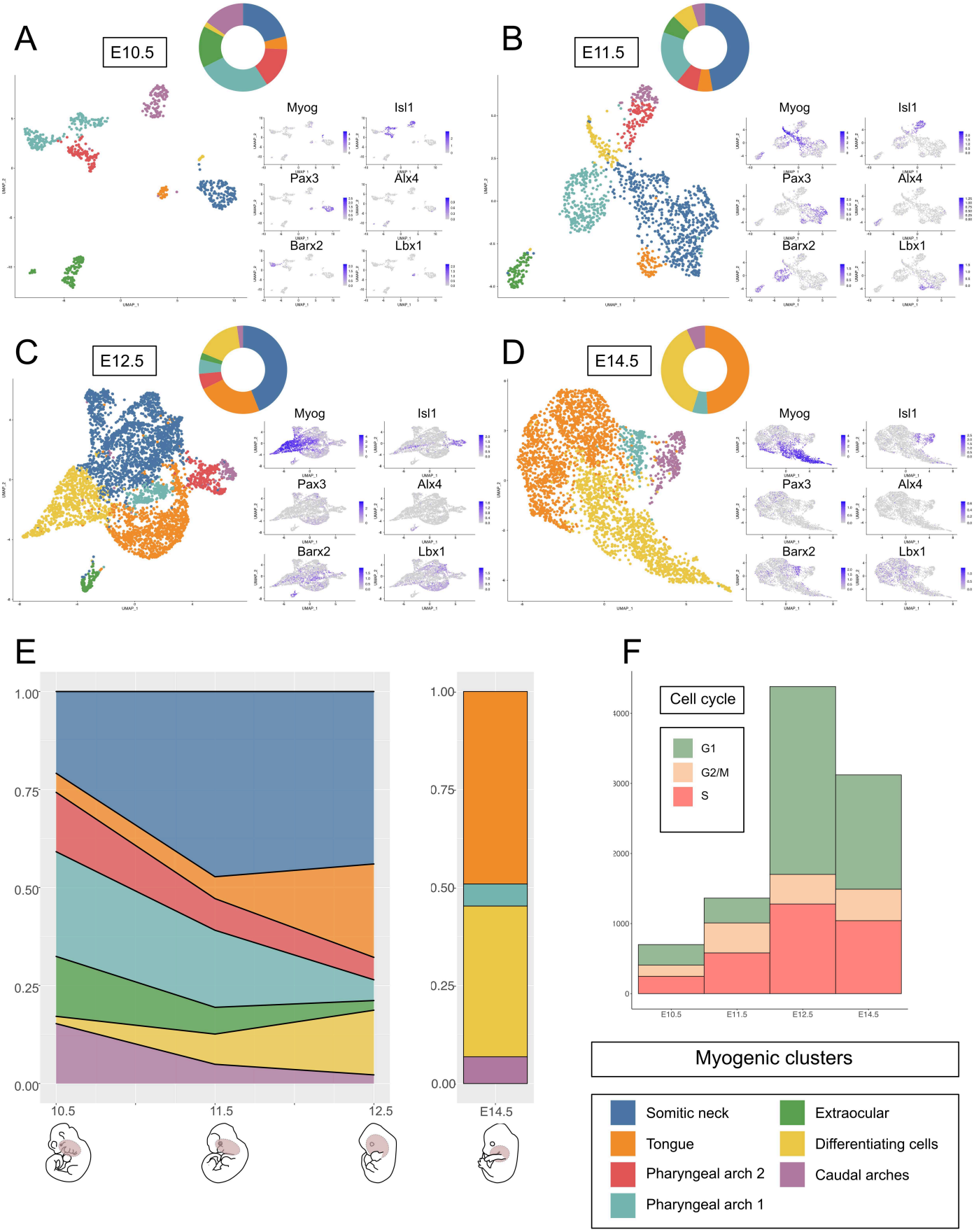
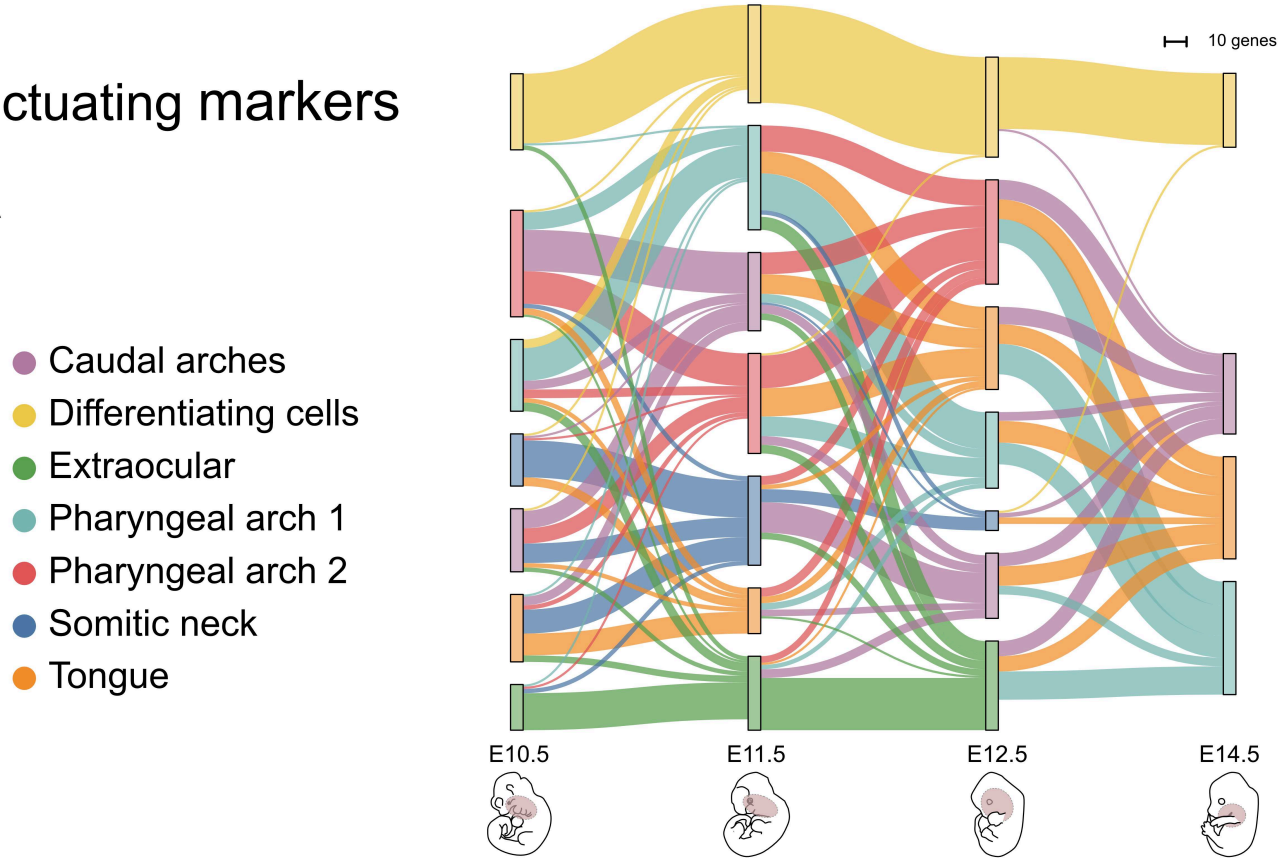

A
E10.5
B
E11.5
C
E12.5
D
E14.5
Myog
Isl1
Pax3
Alx4
Barx2
Lbx1
E
1.00
0.75
0.50
0.25
0.00
10.5
11.5
12.5
E14.5
F
Cell cycle
G1
G2/M
S
4000
3000
2000
1000
0
E10.5
E11.5
E12.5
E14.5
Myogenic clusters
Somitic neck
Tongue
Pharyngeal arch 2
Pharyngeal arch 1
Extraocular
Differentiating cells
Caudal arches

**Figure 2. Cranial myogenic compartments expand asynchronously during development.**
(A-D) UMAP dimension reduction of all 4 datasets, piecharts of the relative proportion of myogenic clusters, and expression pattern of selected markers. *Isl1*: Cardiopharyngeal mesoderm, *Lbx1*: Tongue, *Myog*: Differentiating cells, *Pax3*: Somitic muscle progenitors, *En2*: Epaxial progenitors, *Pitx2*$^{high}$: Extraocular muscle progenitors. (E) Area plot of the relative proportions of each myogenic cluster at each stage. Note the increasing proportion of somitic progenitors, tongue progenitors and differentiated cells. (F) Histogram of the number of cells in each cell cycle phase at each stage. Note the high proportion of G1 cells from E12.5, indicative of differentiating cells. Myogenic cells were filtered from each dataset using myogenic markers, Leiden clustering and Pdgfa/Pdgfra signalling dichotomy (Grimaldi et al., in preparation).

Figure 3

# Fluctuating markers

A

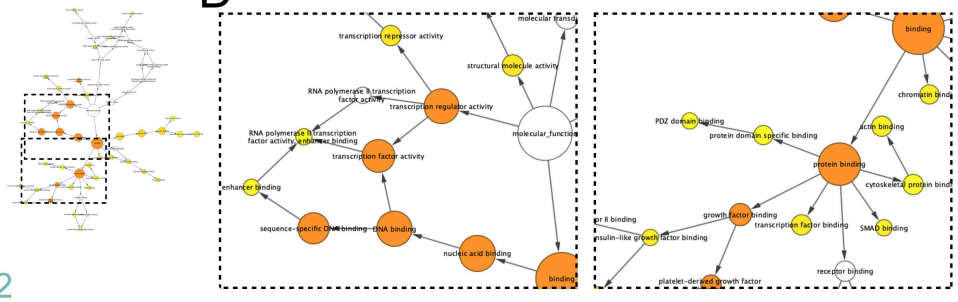



B

# Persisting markers


Progenitors

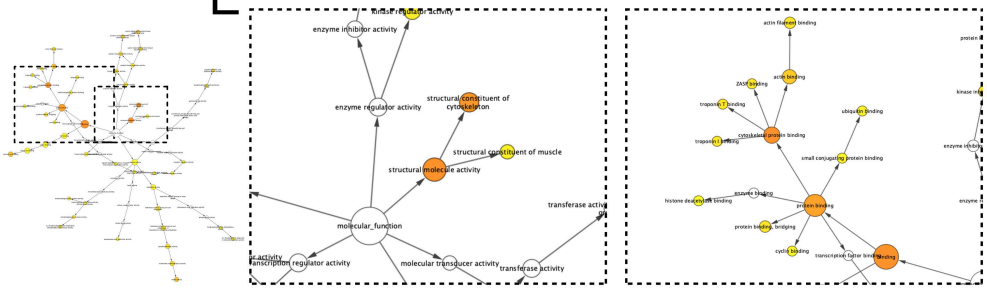
D




Differentiating

C


E




Signficance

**Figure 3. Each myogenic anlage expresses a unique set of persisting transcription factors through development.**

(A) Sankey diagram showing the flow of persisting/non-persisting, common/unique markers. Each node illustrates the top markers of a cluster at a specific stage. The width of the connections between each node is representative of the number of marker genes shared between these 2 conditions. Top marker gene test: Wilcoxon rank-sum test, log fold change: 0.25, maximum number of genes displayed: 50. Markers for differentiating cells and EOM progenitors are highly conserved. (B-C) Wordcloud representation of genes persisting as top markers of each cluster. The size of the word represents the cumulative fold change across all stages. (D-E) Gene set enrichment analysis network from persisting genes of all progenitors (D) and differentiating cells (E). Persisting markers for progenitors are mostly transcription factors while conserved markers for differentiating cells are mostly cytoskeleton molecules.

Figure 4

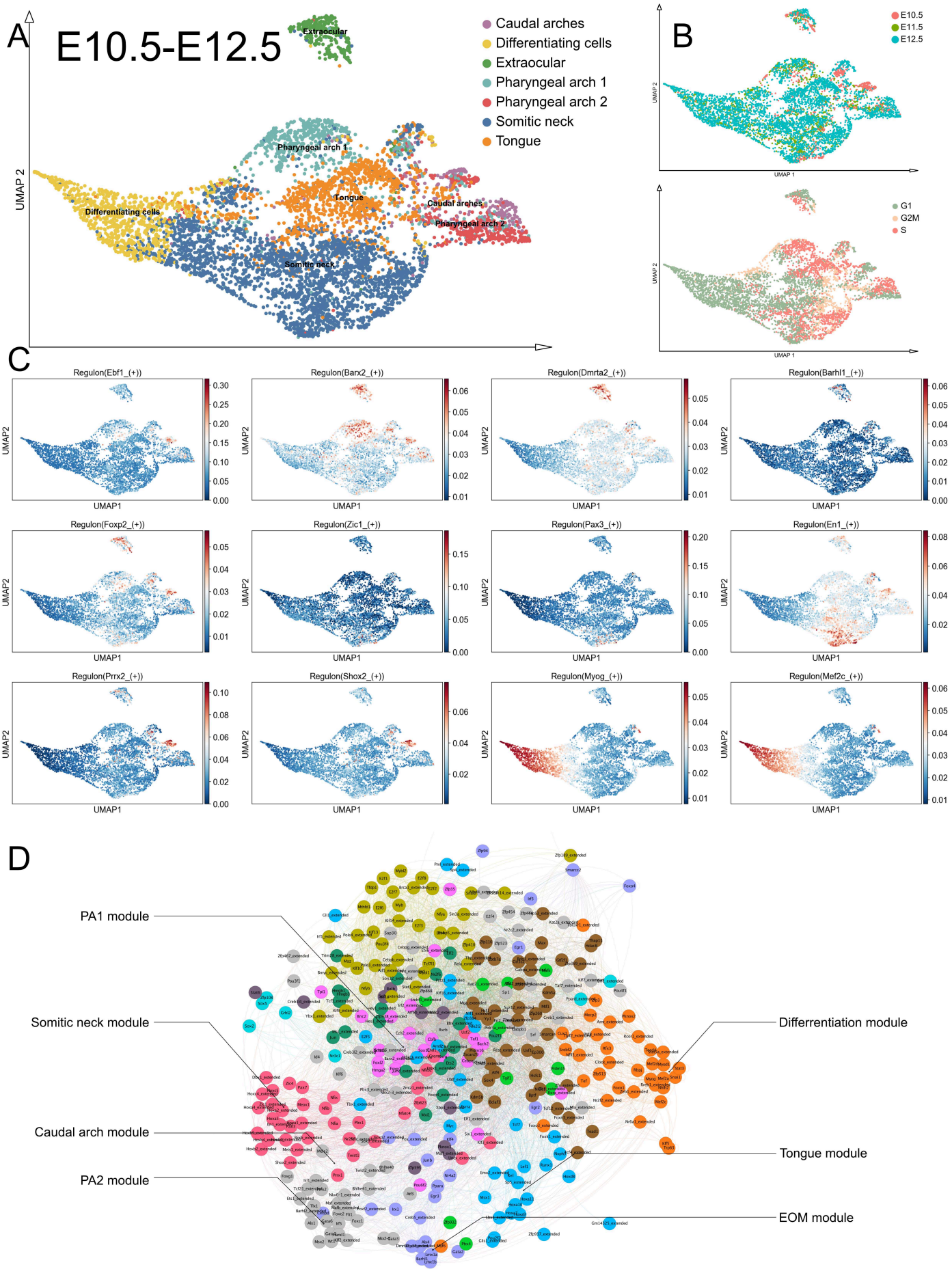

A
E10.5-E12.5
Extraocular
Caudal arches
Differentiating cells
Extraocular
Pharyngeal arch 1
Pharyngeal arch 2
Somitic neck
Tongue
UMAP 2
Pharyngeal arch 1
Tongue
Differentiating cells
Caudal arches
Pharyngeal arch 2
Somitic neck
B
E10.5
E11.5
E12.5
UMAP 2
UMAP 1
G1
G2M
S
C
Regulon(Ebf1_(+))
Regulon(Barx2_(+))
Regulon(Dmrta2_(+))
Regulon(Barhl1_(+))
Regulon(Foxp2_(+))
Regulon(Zic1_(+))
Regulon(Pax3_(+))
Regulon(En1_(+))
Regulon(Prrx2_(+))
Regulon(Shox2_(+))
Regulon(Myog_(+))
Regulon(Mef2c_(+))
UMAP2
UMAP1
D
PA1 module
Somitic neck module
Caudal arch module
PA2 module
Differrentiation module
Tongue module
EOM module

**Figure 4. Distinct transcription factors are active in cranial myogenic progenitors.**
(A) UMAP of merged E10.5 to E12.5 datasets annotated for each myogenic cluster. Cells appear to cluster mostly based on anatomical location. (B) UMAP showing stage, and cell cycle annotations. Stage and cell cycle phase also weigh on cell clustering. (C) Activity of selected regulons, generated with SCENIC. (D) Transcription factor network inferred from regulons. Transcription factor co-regulation is visualized by nodes and edges. Persisting markers of each cluster stand out as coregulating modules.

Figure 5

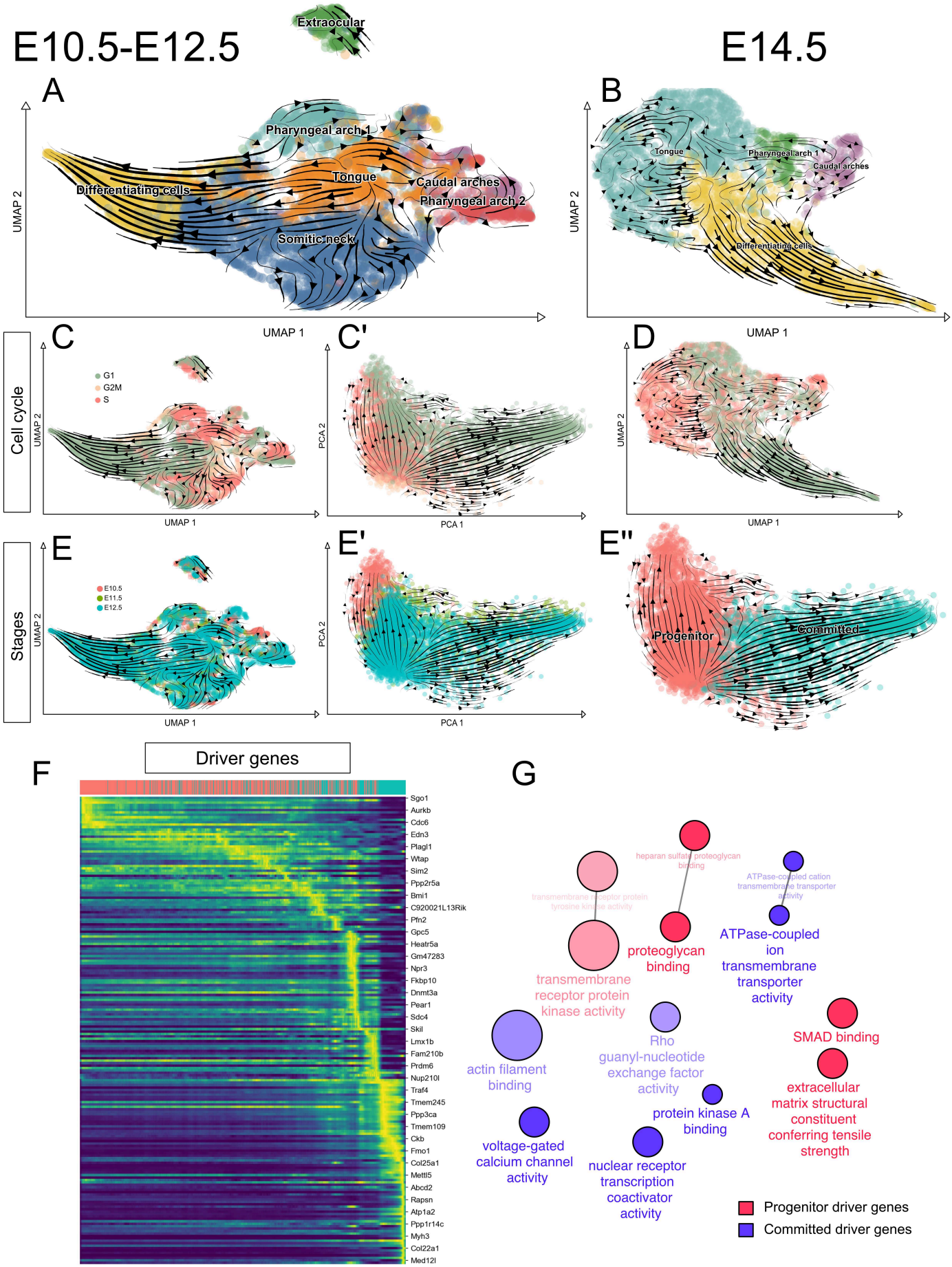
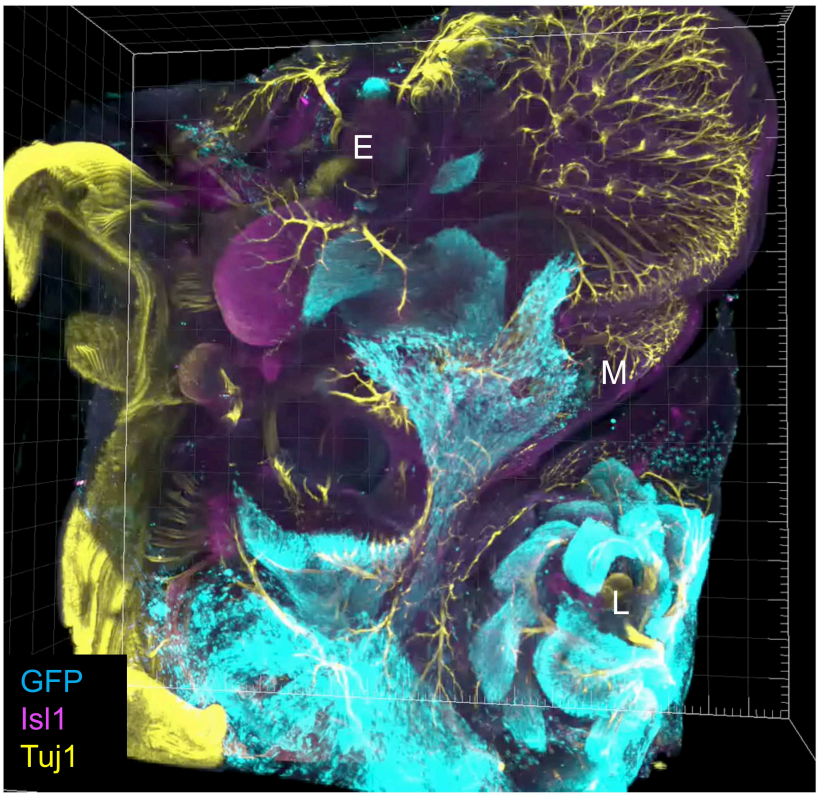

E10.5-E12.5
Extraocular
E14.5
A
Pharyngeal arch 1
Tongue
Caudal arches
Pharyngeal arch 2
Differentiating cells
Somitic neck
UMAP 1
UMAP 2
B
Tongue
Pharyngeal arch 1
Caudal arches
Differentiating cells
Cell cycle
C
G1
G2M
S
C'
PCA 1
PCA 2
D
Stages
E
E10.5
E11.5
E12.5
E'
E''
Progenitor
Committed
F
Driver genes
Sgo1
Aurkb
Cdc6
Edn3
Plagl1
Wtap
Sim2
Ppp2r5a
Bmi1
C920021L13Rik
Pfn2
Gpc5
Heatr5a
Gm47283
Npr3
Fkbp10
Dnmt3a
Pear1
Sdc4
Skil
Lmx1b
Fam210b
Prdm6
Nup210l
Traf4
Tmem245
Ppp3ca
Tmem109
Ckb
Fmo1
Col25a1
Mettl5
Abcd2
Rapsn
Atp1a2
Ppp1r14c
Myh3
Col22a1
Med12l
G
heparan sulfate proteoglycan binding
transmembrane receptor protein tyrosine kinase activity
ATPase-coupled cation transmembrane transporter activity
proteoglycan binding
transmembrane receptor protein kinase activity
ATPase-coupled ion transmembrane transporter activity
Rho guanyl-nucleotide exchange factor activity
SMAD binding
actin filament binding
protein kinase A binding
extracellular matrix structural constituent conferring tensile strength
voltage-gated calcium channel activity
nuclear receptor transcription coactivator activity
Progenitor driver genes
Committed driver genes

**Figure 5. Myogenic cells actively maintain their progenitor state characterised by unique signalling pathways.**

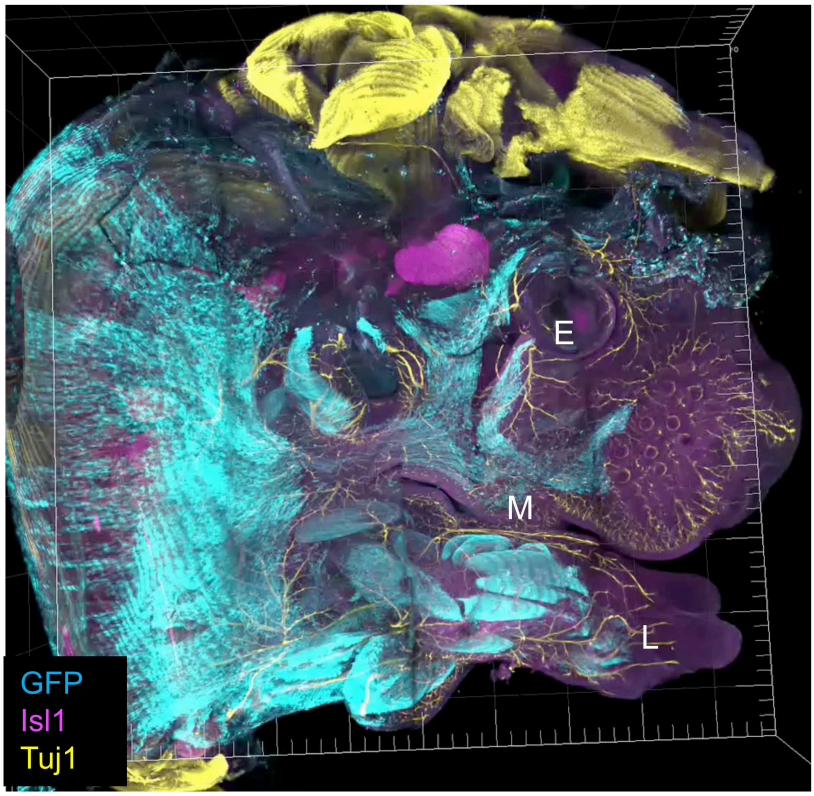
(A-B) RNA velocity streams on merged datasets (A) and E14.5 dataset (B). (C-D) RNA velocity streams with cell cycle phase annotation overlaid onto E10.5-E12.5 UMAP (C), PC1/PC2 (C') and E14.5 UMAP (D). (E-E'') RNA velocity streams with stage annotation overlaid onto E10.5-E12.5 UMAP (E) and E10.5-E12.5 (PC1/PC2). (E'') RNA velocity streams with "progenitor/committed" annotation overlaid onto E10.5-E12.5 PC1/PC2. Myogenic progenitors transition towards commitment, or progenitor maintenance (associated with earlier timepoint states and proliferation). (F) Driver genes characterising progenitor and committed developmental trajectories, along latent time. (G) "GO Molecular function" gene set enrichment analysis network of top 100 driver genes of progenitor and committed populations.

Figure S1

Video 1: E12.5 $Myf5^{Cre/+};R26^{mTmG/+}$




Video 2: E13.5 $Myf5^{Cre/+};R26^{mTmG/+}$



**Supplemental Figure S1: Substantial myogenic expansion at E12.5-E13.5.**

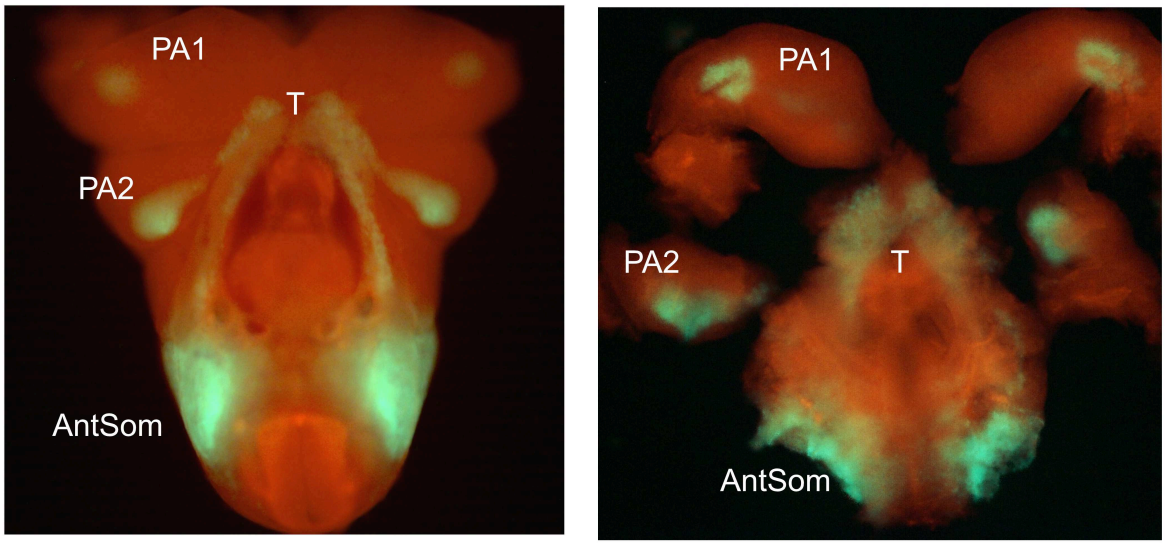
Video 1: 3D reconstruction of a wholemount immunostained and clarified *Myf5$^{Cre/+}$; R26$^{mTmG/+}$* embryo at E12.5, labelled with anti-GFP, anti-Tuj1 and anti-Isl1, imaged by light-sheet microscopy. Note the expansion of the dorsal musculature at E13.5, and the intricate patterns of progenitors in the cranial region. E: Eye, L:Forelimb, M: Mandible.

Video 2: 3D reconstruction of a wholemount immunostained and clarified *Myf5$^{Cre/+}$; R26$^{mTmG/+}$* embryo at E13.5, labelled with anti-GFP, anti-Tuj1 and anti-Isl1, imaged by light-sheet microscopy.

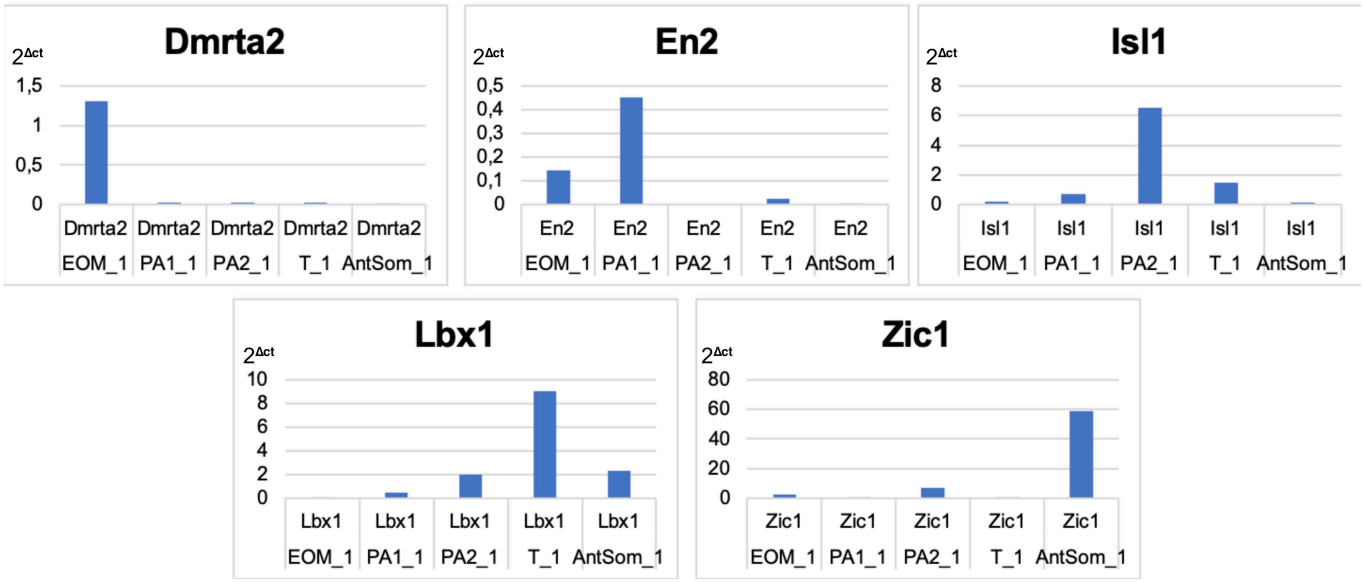
# Figure S2

E11.5 Myf5$^{Cre/+}$;R26$^{mTmG/+}$

A




B



**Supplemental Figure S2: Validation of identified markers.**

(A) Dissection protocol of the pharyngeal arches PA1/2: Pharyngeal arch 1/2, T: Tongue, AntSom: Anterior somites. (B) qPCR results validating the identification of the clusters. N=1

# DISCUSSION

# 1. Myf5+ bipotent progenitors give rise to cranial muscles and associated connective tissue

We describe a Myf5$^{+}$mesodermal population, that gives rise to both myogenic and non-myogenic cells in the embryo. This phenomenon was noted in cranial muscles where neural crest only partially contributes to muscle connective tissue, such as in EOM, laryngeal muscles and somite-derived neck muscles. Bifurcation of myogenic and non-myogenic progenitors is accompanied by complementary tyrosine kinase signaling involving Pdgfs, Ephrins, and Bmps. This analysis also allowed us to identify multiple markers that are associated with vessel development and neuronal cell guidance. Thus, these *Myf5*-derived cells could potentially also participate in establishing neuro-vascular networks within the growing muscle. The transition of Myf5$^{+}$ bipotent cells to non-myogenic Myf5$^{-}$ cells is regulated by a set of transcription factors forming a core regulatory network coopted in different anatomical locations including *Foxp2*, *Hmga2*, *Meis1*, *Meox2,* and *Tcf7l2.*

Further experiments could be carried out to determine the potency of these transcription factors and their targets in inducing connective-tissue-like signatures and inhibiting myogenesis in Myf5$^{+}$ cells. For example, we could perform cell-sorting on *Myf5*-derived/Pdgfra$^{-}$ fetal cells (using *Myf5$^{Cre/+}$; R26$^{TdTomato/+}$; Pdgfra$^{H2GFP/+}$*) and assess their capacity activate Pdgfra (GFP) under specific conditions in vitro. Other reports have suggested that *Myod*-derived cells do not give rise to connective tissue (Wood et al., 2020). Accordingly, we did not observe many Myod$^{+}$ cells that would adopt the same directional trajectories as Myf5$^{+}$ cells. However, more detailed analyses of *Myod$^{Cre/+}$;R26$^{TdTomato/+}$;Pdgfra$^{H2BGFP/+}$* would help assess if this potential is restricted only to Myf5$^{+}$ cells.

A previous study described a *Myf5*-derived population in the EOM that the authors identify as neural crest-derived pericytes (Stuelsatz et al., 2014). However, the authors did not include single-cell co-labeling of neural crest and Myf5 to verify this observation. Apart for differences in the *Myf5$^{Cre}$* allele used (Myf5$^{tm3(cre)Sor/J}$ (Tallquist et al., 2000) previous versus Myf5$^{tm1.1(cre)Mrc}$ (Haldar et al., 2008) present study), we could not detect neural crest-derived cells expressing *Myf5* at E13.5. Also, all Myf5$^{+}$ cells appeared to be mesoderm-derived as verified by examination of *Wnt1$^{Cre/+}$; R26$^{mTmG/+}$* and *Mesp1$^{Cre/+}$; R26$^{mTmG/+}$* embryos. Further, Myf5-derived Pdgfra$^{+}$ cells at E14.5 and Sox9$^{+}$ cells at P0 were found in regions harboring mesoderm-

derived connective tissue (EOM medial attachment and intraconal fat, laryngeal arytenoid muscles, and anterior somites). We also detected *Myf5*-derived interstitial cells subjacent to EOM muscle fibres at P30. Future work will determine the extent of *Myf5*-derived interstitial stromal cells in the EOM and other muscles, as this would suggest an intimate relationship between adult muscle stem cells and clonally related connective tissue. Specifically, we will use $Wnt1^{Cre/+};R26^{TdTomato/+};Pdgfra^{H2GFP/+}$, $Mesp1^{Cre/+};R26^{TdTomato/+};Pdgfra^{H2GFP/+}$ and $Myf5^{Cre/+};R26^{TdTomato/+};Pdgfra^{H2GFP/+}$ mice to assess the contribution of non-neural crest cells to developing muscles and to the adult stem cell niches.

In this study, we present a common gene regulatory network underpinning connective tissue development that arises from $Myf5^{+}$ mesodermal cells at multiple embryonic stages. This transition does not seem to occur in neural-crest embedded muscles, suggesting that this might be a highly regulated process. It is unknown if this potential arises from intrinsic predispositions of anterior mesoderm or if non-neural crest environmental cues contribute to this plasticity.

## 2. Genetic birthmarks regulate cranial myogenic diversity

We present a model for cranial myogenesis, where embryologically distinct myogenic anlagen give rise to discrete adult muscle groups under the control and maintenance of specific regulatory networks, that we named “genetic birthmarks” (GBMs). We identified specific markers of cranial myogenic progenitors including EOM, PA1, PA2, somitic neck muscles, tongue, and notably, caudal arch muscles. A significant proportion of these genes are transcription factors that we found to be specifically active during development. Network-based visualization revealed that these genes form modules that might indicate co-regulation at the level of each muscle anlage. Transcriptomic trajectories exposed the maintenance of a specific progenitor state associated with extracellular matrix production and BMP signaling, reminiscent of connective tissue lineages. This study requires additional analyses and complementary validations by in situ methods. Notably, regulon data from all stages can be combined into a single network to visualize the unique regulatory dynamics of every cluster at each stage. Also, other validations are ongoing in the embryo and adult in the form of qRT-PCRs and RNAscope in situ hybrisations.

As indicated previously, EOM are specifically spared in most diseases (Formicola et al., 2014; Kaminski et al., 2002; Man et al., 2005; Sambasivan et al., 2009; Stuelsatz et al., 2015). We identified the specific regulatory networks of the developing EOM, which for a large part are maintained in adult stem cells (data not shown), and after heterotopic transplantation (Evano et al., 2020). It is thus possible that these GBMs in EOM are responsible in part for the intrinsic and unique properties of EOM stem cells (Stuelsatz et al., 2015). Further analyses will determine the binding potential of EOM-specific GBMs to genes associated with their high proliferative capacities.

## 3. A fibrogenic state in myogenic progenitors

In our work, we present transcriptional changes at the level of embryonic myogenic progenitors that are associated with connective tissue-specific signatures: production of extracellular matrix constituents, proteoglycans, adhesion molecules and tyrosine kinase signaling molecules, including BMPs. Surprisingly, these pathways were found to be dynamically regulated in $Myf5^{+}$ myogenic cells transitioning to a $Myf5^{-}$ non-myogenic state, as well as in myogenic progenitors actively maintaining an undifferentiated pool. It is thus tempting to speculate that this mechanism is an intrinsic property of myogenic cells that allows the integration of mesenchymal features such as extracellular matrix remodeling, migration potential, proliferation and paracrine activities to promote differentiation within the myogenic anlage (Murphy et al., 2011; Sefton and Kardon, 2019; Uezumi et al., 2014). Eventually, in regions where neural crest is absent, this process could be exploited even further and lead to the formation *bona fide* connective tissue by Myf5+ cells. This intriguing observation can be also considered in the context of the EOM, that contain a relatively high amount of *Myf5*-derived connective tissue, including cartilage, adipocytes, tendon and interstitial cells.

Considering the higher proliferative capacity of EOM stem cells (Stuelsatz et al., 2015), one can imagine that the cellular plasticity of EOM during development could be reflected later to the adult phenotype. That is, by regulating a pseudo-connective tissue state possibly through the maintenance of deeply rooted ancestral regulatory networks or GBMs. In this context, adult muscle stem cells have been shown to display fibro-adipogenic behaviors under certain conditions (Uezumi et al., 2014). When cultured with aged mouse serum, some muscle stem cells acquired a fibrogenic phenotype (Brack et al., 2007) and several studies have reported a

fibro-adipogenic potential from satellite cell-derived myoblast cultures or single myofibre cultures (Asakura et al., 2001; Shefer et al., 2004; Wada et al., 2002), although concerns about mesenchymal contamination were raised (Uezumi et al., 2014). Muscle stem cells were shown to convert to adipocytes under cold condition, and this transition is repressed in the normal state by *Myod* (Yin et al., 2013). More work needs to be done to assess the fibrogenic potential of muscle stem cells, and evaluate whether this property is muscle-dependent. Other members of our team are currently exploring this specific state in the EOM adult muscle stem cells, compared to that of the limb.

## 4. scRNAseq

Single-cell transcriptomics has revolutionized developmental biology by providing a method to interrogate a large number of genes at single-cell resolution spatiotemporally and infer lineage trajectories across multiple cell populations simultaneously (Cao et al., 2019; Ibarra-Soria et al., 2018; Lescroart et al., 2018; Pijuan-Sala et al., 2018; Wagner et al., 2018). In doing so, it has redefined the notions of cell type and cell markers and deepened our understanding of cell lineage progression and single-cell fate decisions(Wagner et al., 2016). The work presented in this thesis relies primarily on this technology to investigate craniofacial development in mouse. Thus, it is important to appreciate the advantages and limitations associated with this approach.

***Advantages***

Given adequate sample preparation and sequencing parameters, scRNAseq allows a relatively accurate depiction of the transcriptomic state of a single cell. As such, new markers for various cell types can be identified and differential expression analysis can be performed. In our first study, this approach revealed a strong myogenic versus non-myogenic bias for Pdgf signaling (Pdgfa/Pdgfra) which we validated in vivo. From an initial list of differentially expressed genes, gene set enrichment analysis can assist researchers by summarizing the results into generic terms relative to biological processes, molecular function, protein classes, and more. This approach helped us to identify dominant protein classes within GBMs (markers of cell states at the onset of myogenesis), as well as specifying the biological processes underlying genes characterising Myf5+ connective tissue cell trajectories. Further, differential gene

expression analysis can distinguish conditions that were generated separately and that appear distinct. This requires a preemptive merging of datasets, a challenging process where batch effect bears significant weight, and for which an array of computational methods have been (are being) developed (Tran et al., 2020). In our context, datasets from both *Myf5*$^{Cre/+}$*;R26*$^{mTmG/+}$ at E11.5 and *Mesp1*$^{Cre/+}$*;R26*$^{mTmG/+}$ at E10.5 originated from 2 littermate duplicates that we merged without correction after evaluating that their overlap was extensive. The second study presented here involved merging 3 datasets of 3 stages. This operation was performed using Seurat "anchors", and allowed us to group cells based on their anatomical location (and thus underlying muscle specificity) instead of batch-based (Butler et al., 2018).

The studies described in this thesis include the latest advances in computational methods for RNA velocity (Bergen et al., 2019) and regulatory network inference (Aibar et al., 2017; Sande et al., 2020). RNA velocity (Bergen et al., 2019; Manno et al., 2018) exploits spliced and unspliced transcriptomes to infer lineage progression. That approach is a major improvement from methods solely based on reversed graph embedding to infer cell trajectories (Cao et al., 2019; Qiu et al., 2017; Trapnell et al., 2014). Notably, it reduces the user input (and thus user bias) with regards to the direction of the course taken by individual cells. For instance, even in an ideal context where a single lineage is fully and unequivocally displayed graphically (through UMAP or tSNE coordinates), with progenitors on one end and committed cells on the other end, the fate of an intermediate cell remains ambiguous. A cell in this position could move towards differentiation, or replenish the progenitor pool. It is thus critical for the investigator to reevaluate how committed a given cell is relative to its transcriptomic signature in an in vivo context.

In our work, Myog+ cells were not found to be diverted from their trajectory towards differentiation. Myod+ cells were more inclined to deviate from myogenesis towards an upstream progenitor or connective tissue cell state (although Myf5+/Myod-/Myog- cells had greater potential to do so). The latest iteration of RNA velocity methods allows more accurate trajectory inference and provides a list of "driver genes", ie. genes that are responsible for most of the velocity. We exploited this feature to assess the genes that were actively transcribed during Myf5+ bipotent-to-fibrogenic progression.

We combined these results with a regulatory network inference method (Aibar et al., 2017; Sande et al., 2020), to identify the active transcription factor networks underlying this

transition. This powerful tool harnesses available binding motif databases and correlative expression within the data to build "regulons" (transcription factor and targets). Here, the initial count matrix is reduced dramatically to the most significant regulatory modules. We took advantage of this tool to identify transcription factors regulating driver genes that we presume to lay at the base of connective tissue cell fate decisions.

We also exploited this tool to create networks of TFs, based on putative coregulation patterns. When we subjected the data to force-directed algorithms, nodes of the network organized graphically as modules, thus uncovering potential cooperative systems visually. We propose that this downstream visualization method is used as a more user-friendly and straightforward way to reduce dense networks into a handful of modules comprised of tightly co-regulating TFs. This approach allowed us to reveal core memory modules of cranial myogenic anlagen (named "genetic birthmarks").

### ***Limitations***

Like any method, there are some specific considerations to integrate when using scRNAseq. First, the sequencing depth and the number of cells need to be tailored to the level of heterogeneity of the cell population. If a standard sequencing depth (around 50 000 reads/cell) is used, then the number of cells analysed must be sufficient to reveal the cell subtypes that are of interest. In standard experiments that involve examination of a few thousand cells, the isolation and selection method must be appropriately designed to capture all targeted populations. In our case, this was illustrated by refining the selection method to capture caudal arch myogenic progenitors at later stages, when proliferating and differentiating cells from other regions became too dominant. Further, we adjusted the sample preparation by using contemporary Myf5-labeling instead of Myf5 lineage tracing, thereby restricting the analysis to that cell state. We also adapted our dissection protocols to select more specifically caudal arch progenitors in the laryngeal and pharyngeal regions.

Another consideration is cluster annotation which is a critical step and arguably one of the most prominent sources of inaccurate conclusions in downstream analyses. Considerable efforts have been invested to build comprehensive catalogues of transcriptomic signatures for all cell types to assist in their identification (Ma and Pellegrini, 2019; Wang et al., 2020b; Xu

and Su, 2015; Zand and Ruan, 2020). While well-known markers exist for a variety of cell types, many subtypes and transient states do not have definite transcriptomic signatures, or they are extrapolations from the adult phenotypes which are not necessarily representative of the embryonic ancestral cell. Cell types that have not been clearly defined, such as mesenchymal cells, have transcriptomic profiles that often overlap.

Efforts are being made currently to better discriminate stromal cells into specific cell populations, but no unique transcription factors were found (Mononen et al., 2020; Muhl et al., 2020). In our studies, we used well-known markers to annotate differentiating muscle (*Myog*), epaxial somitic progenitors (*Pax3, En1, Zic1*), hypaxial tongue progenitors (*Lbx1, Sim2*), extraocular muscles (*Pitx1/2, Alx4*) and cardiopharyngeal mesoderm progenitors (*Isl1, Tcf21, Tbx1*). A first annotation of this type led to the discovery of even more specific markers that were conserved across embryonic stages. Also, cartilage (*Sox9, Col2a1*), dermis (*Twist2*), and adipocytes (*Dlk1, Prdm16, Ebf2*) were relatively straightforward to identify. However, pericytes, interstitial fibroblasts, and tenogenic cells were more challenging to discern clearly, as others have reported (Muhl et al., 2020). In addition, it is not clear whether these cells have distinct progenitors in the embryo, or if a common ancestral pool expresses markers equally for multiple lineages.

This technology is also prone to dropouts (i.e. undetected but expressed gene). Dropouts can be reduced with appropriate sequencing depth and transcript representation and by avoiding overwhelming transcripts that would mask the lower expression of others. Regulatory network inference algorithms reduce the influence of dropouts by compressing the data into a higher level of transcription factor activity, in which the lack of one target will be compensated for by the presence of others (Aibar et al., 2017; Sande et al., 2020).

Importantly, scRNAseq is based solely on transcriptomic data. Therefore, post-transcriptional modifications are completely inaccessible. Indeed, most mature miRNA do not possess a polyA tail and thus will not bind polydT primers, the most widely used method to capture mRNA. New methods are emerging to combine mRNA and miRNA single-cell transcriptomics (Wang et al., 2019c). Various approaches combine scRNAseq with proteomics analyses to couple transcriptomic profiles with protein content. However, current techniques only allow the coupling of scRNAseq with cell surface markers (Ha et al., 2020; Keren-Shaul et al., 2019; Mair et al., 2020; Stoeckius et al., 2017). This limitation is particularly relevant in

the context of myogenesis, where several post-transcriptional regulatory mechanisms have been reported(Xu et al., 2019).

***Extracting valuable information***

As single-cell technologies continue to expand and offer more resolution into biological processes, the challenge in handling large datasets becomes increasingly prevalent (Lähnemann et al., 2020). Data science has provided biologists with dimensionality reduction methods that allow visualisation of complex multivariable variations in 2D space, such as principal component analysis, tSNE and UMAP (Becht et al., 2018; Lehrmann et al., 2013; Maaten and Hinton, 2011; McInnes et al., 2018). However, finding biological relevance in a set of differentially expressed genes is a common issue with transcriptomic data. In both studies presented in this thesis, we resorted to a specific strategy to infer regulatory mechanisms that are currently active within certain clusters, and combined these results with the most transcriptionally dynamic genes between clusters, using SCENIC and scvelo (Aibar et al., 2017; Bergen et al., 2020; Sande et al., 2020). We argue that these 2 algorithms provide powerful, biologically relevant dimensionality reduction methods, and they are complementary. Additionally, we propose some downstream methods to condense and extract valuable information from this pipeline. These methods are exploiting network-based graphs and force-directed layouts to offer a visually appealing way to explore interconnected genes (Koutrouli et al., 2020). Three different types of networks were generated, representing: 1) transcription factors-only; 2) driver regulons; and 3) gene set enrichment analysis terms. Respectively, these methods identified: 1) co-regulating modules; 2) the main regulators of transitional states; 3) molecular pathways characterising transitional states. This pipeline will be useful to unravel the molecular mechanisms underpinning lineage progression, and to explore large datasets in a more comprehensive way.

# 5. Evolution of caudal arches

In 2012, type II myosin heavy chain (MyHC) orthologs were found in unicellular organisms, indicating that cellular contraction is likely to be an ancient feature, predating multicellular life(Steinmetz et al., 2012). Nonetheless, the mechanisms employed by metazoans (i.e. “animals”) for locomotion are extremely varied and their evolution is often accompanied by remodeling of entire structures (Deline et al., 2018; Raff, 2008; Tokita and Schneider, 2009).

***Axial muscles as an ancestral feature***

The axial musculature (i.e. trunk) is comprised of repeated segments along the anteroposterior axis. This feature is shared among all chordates (metazoans with notochords) and allows stability and flexibility of the notochord (or vertebrae in vertebrates, where the notochord appears transiently) to support locomotion (Schubert et al., 2018).

The ancestral axial musculature likely resembled that of the cephalochordate Amphioxus. This close relative of vertebrates possesses V-shaped axial segmented muscles from head to tail, separated by myosepta that allowing distinct contractions and facilitate swimming (Mansfield et al., 2015).

The evolution of jawless vertebrates (lampreys and hagfish) is associated with the emergence of epaxial and hypaxial muscles, increasing the range of motions to lateral and dorsoventral flexions (Fetcho, 1987). The development of the epaxial and hypaxial division appears to be controlled by the transcription factor *En1* (*Engrailed1*), which upon misexpression, suppresses markers of the ventral lip domain (Ahmed et al., 2017; Cheng et al., 2004b). In jawed vertebrates, this separation became more prominent as a horizontal septum of connective tissue formed in between (Wotton et al., 2014).

***Vertebrates***

The evolution of vertebrates is marked by the appearance of paired fins and limbs for greater locomotion, and a "new head". Fossils from the Devonian era such as the "Tiktaalik" help to elucidate the genetic changes underlying the morphogenic remodeling of the fin-like skeleton into the tetrapod digited limb (Shubin et al., 2006). During aquatic to terrestrial transition, changes in regulatory sequences of *Hox* genes allowed the appearance of a wrist and digits (Nakamura et al., 2016; Stewart et al., 2019).

Vertebrate evolution is marked by the progressive remodeling of the head and neck region, leading from a transition from passive filter-feeding to predatory life style (Diogo et al., 2015; Naumann et al., 2017; Sambasivan et al., 2011b; Schubert et al., 2018; Tzahor, 2015). In mammals, for instance, this set of acquisitions include extraocular muscles to move the eye, flexible neck musculature, jaw musculature, soft palate, pharyngeal and laryngeal muscles for

coordinated feeding, breathing and vocalizing. Many of these advantageous features are attributed to the neural crest, a specialized tissue emanating from the dorsal neural tube that populates the majority of the cranial region (Green et al., 2015; Martik et al., 2019; Ziermann et al., 2018). In addition, vertebrate cranial features (including muscles) have been substantially remodeled during evolution (Frisdal and Trainor, 2014; Poopalasundaram et al., 2019; Schubert et al., 2018).

***Remodeling of caudal arches in vertebrates***

Muscles of the pharynx are prominently developed in mammals (Smith, 1992). The mammalian pharynx is primarily made of soft tissues that are rarely preserved in fossil records (Zatoń and Broda, 2015). This might explain why their evolution is underappreciated (Smith, 1992). Soft palate, pharyngeal and esophageal muscles are mostly absent in sauropsids (birds and reptiles), and the laryngeal musculature in these animals is poorly developed (Smith, 1992). Amphibians do possess some pharyngeal muscles, but their homology with mammals is debated (Ericsson et al., 2013; Smith, 1992). In addition, a study has found esophageal striated muscles in the bull frogs, which are missing in birds and reptiles (Yoshida, 2001). This indicates that if the pharyngeal muscles of mammals and amphibians are indeed homologous, sauropsids may have lost a significant portion of pharyngeal and esophageal musculature during evolution (Smith, 1992).

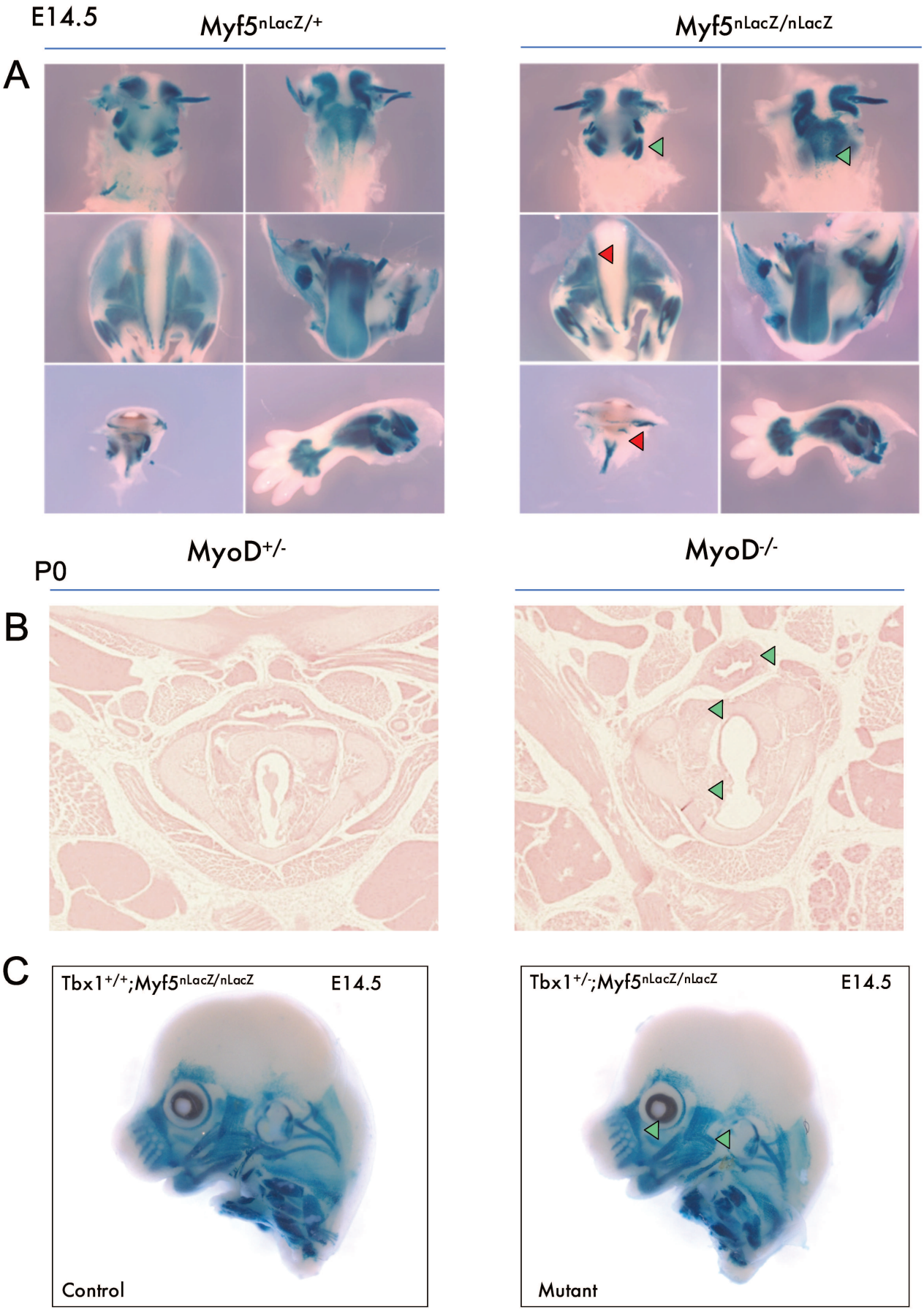
Importantly, birds perform vocalizations primarily through the syrinx, a specialized muscular structure found at the base of the trachea (**Annex 3A**) (Faunes et al., 2017; Kingsley et al., 2018; Schubert et al., 2018). Interestingly, a recent study proposed that syringeal muscles belong to the craniofacial muscle group based on their expression of *Myh13*, which is considered to be a "head-specific" MyHC isoform (Mead et al., 2017). This raises the intriguing possibility of a posteriorization of some caudal arch mesoderm in birds. In this context, it is thus interesting to consider the anatomical positioning of caudal arch derivatives with respect to the neck and cervical segments. Remarkably, the parathyroid gland and ultimobranchial bodies that are derived from posterior pharyngeal pouches, are located in the thoracic region in birds and reptiles (Breit et al., 1998; Clark, 1971). In contrast, they are present in proximity to the larynx in mammals, in the cervical region (between C3-C6), even in giraffes (Harrison, 2009). Consistent with this finding, avians can have up to 25 cervical vertebrae while most mammals only have 7 (Varela-Lasheras et al., 2011). An appealing hypothesis would then be that duplication of cervical segments in birds might have displaced some caudal arch derivatives

posteriorly, giving rise to thoracic thyroid and syringeal elements. Such a phenomenon could involve differential control of *Hox* gene expression and/or a higher rate in the somite segmentation clock, the 2 main known processes tied to establishing the number of skeletal segments (Gomez et al., 2008; Thewissen et al., 2012). Additionally, it is interesting to note that a study in chick where laryngeal muscles were reported to be of somitic origin, and the thyroid cartilage, the only neural crest derived cartilage in this region, is absent (Noden and Epstein, 2010).

We performed some experiments to investigate the extent of posterior arch remodeling in chick, and found that the caudal pharyngeal arches are clearly defined, and positioned latero-ventrally, compared with mouse, where their morphological appearance is less discernable (**Annex 3B-C, J, H**). We also found a significant contribution of $Isl1^{+}$ cells (a marker for cardiopharyngeal mesoderm) to the chick larynx at D7, although it is not clear if these cells are restricted to cartilage or if cartilage and muscles are all $Isl1^{+}$(**Annex 3D**). Further single-cell resolution analyses will need to be done to assess the potential cranial mesoderm origin of bird laryngeal muscles. Notably, we did observe an $Isl1^{+}$ contribution to the mesodermal core of posterior arches in the developing chick (**Annex 3F**), however, a pilot qRT-PCR experiment suggested that although *Isl1* is expressed in chick posterior arches, myogenic factors are not (**Annex 3J**). These results suggest that the laryngeal structures in chick are derived from cardiopharyngeal mesoderm, but laryngeal muscles are not. Further, there may lie an additional cardiopharyngeal mesoderm population posterior to larynx, that could give rise to syringeal muscles. Therefore, more experiments are needed to determine the muscularisation of caudal arches in avians, and the potential cranial mesoderm origin of syringeal muscles.

# ANNEXES

# Annex 1

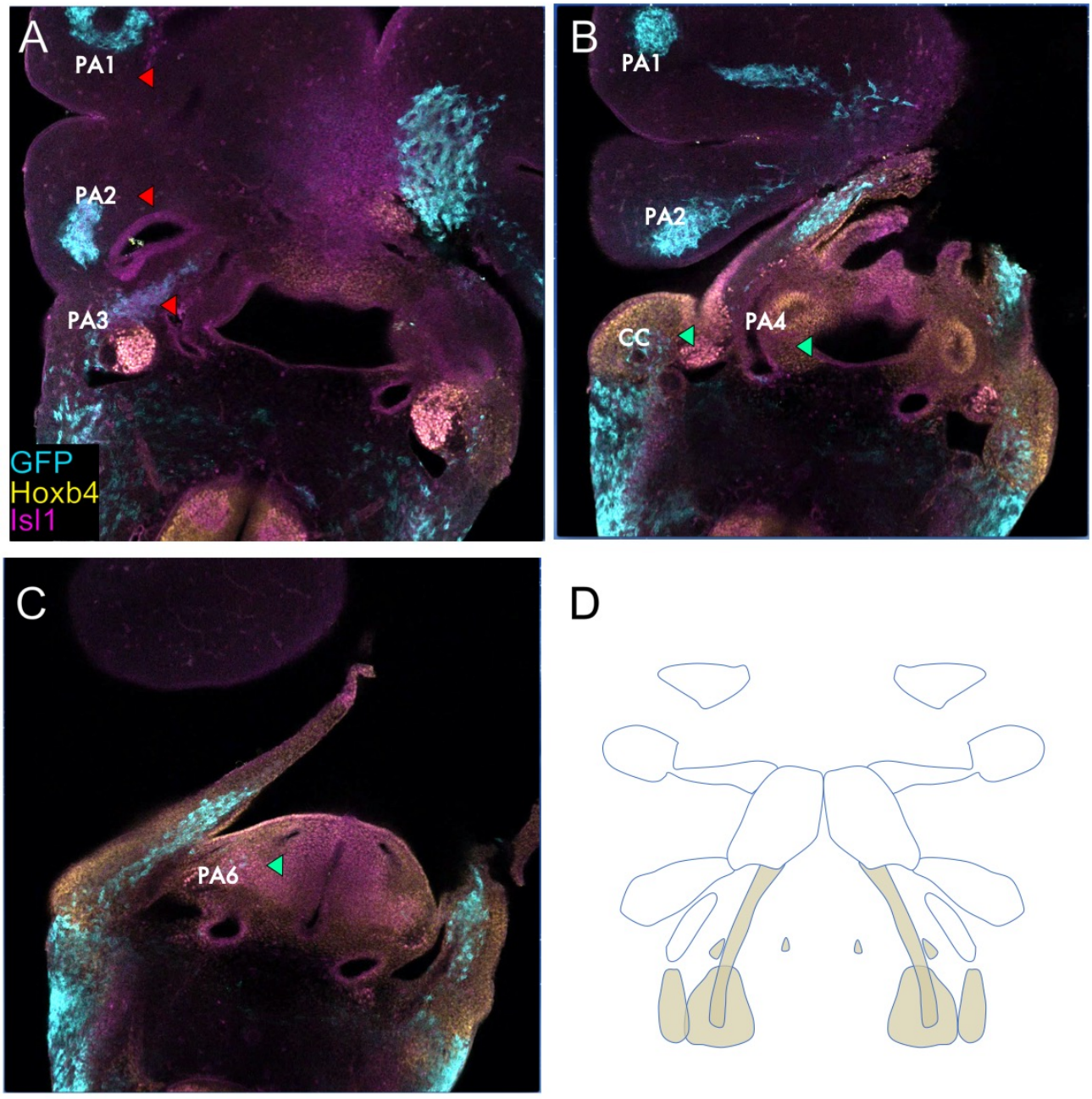

E14.5
Myf5nLacZ/+
Myf5nLacZ/nLacZ
A
P0
MyoD+/-
MyoD-/-
B
C
Tbx1+/+;Myf5nLacZ/nLacZ
E14.5
Control
Tbx1+/-;Myf5nLacZ/nLacZ
E14.5
Mutant

**Annex 1: Assessment of the genetic requirements for caudal arch muscles.**

(A) *Myf5*$^{nlacZ/+}$ (heterozygous control) and *Myf5*$^{nlacZ/nlacZ}$ (mutant) at E14.5. Different regions of the fetus were dissected and X-gal stained to assess muscle integrity. Top to bottom, left to right: Laryngeal muscles, pharyngeal and esophageal muscles, tail epaxial muscles, tongue and mandible, EOMs, forelimb. Green arrowheads indicate intact caudal arch-derived muscles. Red arrowheads indicate affected tail epaxial muscles and EOMs. (B) *Myod*$^{+/-}$ (heterozygous control) and *Myod*$^{-/-}$ (mutant) at P0. Loss of *Myod* did not noticeably affect the esophageal, posterior arytenoid and crico-arytenoid muscles (Green arrowheads). (C) *Tbx1*$^{+/+}$*; Myf5*$^{nlacZ/nlacZ}$ (control) and *Tbx1*$^{+/-}$*; Myf5*$^{nlacZ/nlacZ}$ (*Tbx1* heterozygous mutant) at E14.5, X-gal stained. Loss of 1 allele of *Tbx1* in a *Myf5(Mrf4)* double mutant context did not lead to significant alteration of PA1, PA2 (green arrows) and laryngeal muscles (not shown).

Annex 2

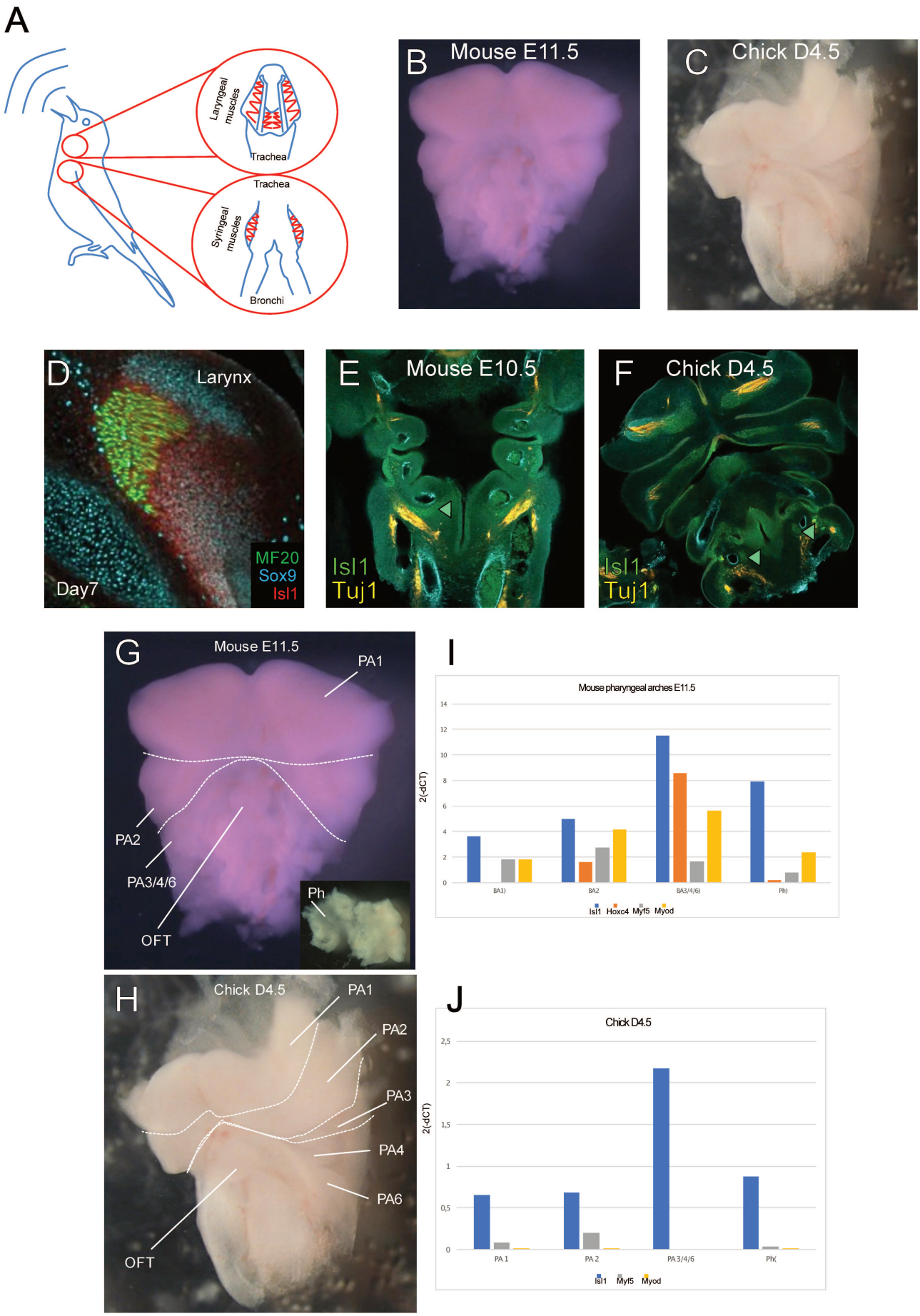
E11.5 Myf5$^{Cre/+}$;R26$^{mTmG/+}$



**Annex 2: Hoxb4 marks caudal arch mesenchyme**

(A-C) Whole mount immunostaining on *Myf5*$^{Cre/+;}$ *R26*$^{mTmG/+}$ E11.5 embryo with Isl1 (CPM/endoderm), GFP (Myogenic cells) and Tuj1 (neuronal marker for innervation), from the ventral to dorsal side of the embryo. Green arrowhead indicates positive staining for Hoxb4 in the 4$^{th}$ and 6$^{th}$ arches. Red arrowhead indicates negative staining in PA1, PA2, and PA3. CC=Cucullaris muscle anlage. (D) Scheme highlighting the frontier of Hoxb4 expression.

# Annex 3


A
Laryngeal muscles
Trachea
Trachea
Syringeal muscles
Bronchi
B
Mouse E11.5
C
Chick D4.5
D
Larynx
MF20
Sox9
Isl1
Day7
E
Mouse E10.5
Isl1
Tuj1
F
Chick D4.5
Isl1
Tuj1
G
Mouse E11.5
PA1
PA2
PA3/4/6
OFT
Ph
I
Mouse pharyngeal arches E11.5
2(-dCT)
14
12
10
8
6
4
2
0
BA1)
BA2
BA3/4/6)
Ph)
Isl1 Hoxc4 Myf5 Myod
H
Chick D4.5
PA1
PA2
PA3
PA4
PA6
OFT
J
Chick D4.5
2(-dCT)
2,5
2
1,5
1
0,5
0
PA 1
PA 2
PA 3/4/6
Ph(
Isl1 Myf5 Myod

**Annex 3: Chick caudal arch remodeling**

(A) Scheme of bird laryngeal and syringeal muscles.

(B-C) Comparative ventral views of dissected pharyngeal arches from mouse E11.5 embryo (B) and equivalent stage chick D4.5 embryo (C).

(D) Immunostaining at D7 of the laryngeal region in chick. Note that Isl1 is expressed throughout the cricoid cartilages.

(E-F) Whole-mount immunostaining of mouse and chick pharyngeal arches at E10.5 and D4.5 respectively, labeling Isl1 (CPM/endoderm) and Tuj1 (innervation). Green arrowheads indicate $Isl1^{+}$ CPM in the caudal arches.

(G-J) Pilot qRT-PCR experiment on dissected pharyngeal arches from mouse E10.5 and chick D4.5 embryos. OFT: Cardiac outflow tract.

**Annex 4: Published article**

Heude, E., Tesarova, M., Sefton, E. M., Jullian, E., Adachi, N., Grimaldi, A., Zikmund, T., Kaiser, J., Kardon, G., Kelly, R. G., et al. (2018). Unique morphogenetic signatures define mammalian neck muscles and associated connective tissues. eLife 7,.

eLIFE elifesciences.org

RESEARCH ARTICLE

# Unique morphogenetic signatures define mammalian neck muscles and associated connective tissues

**Eglantine Heude[1,2], Marketa Tesarova[3], Elizabeth M Sefton[4], Estelle Jullian[5], Noritaka Adachi[5], Alexandre Grimaldi[1,2], Tomas Zikmund[3], Jozef Kaiser[3], Gabrielle Kardon[4], Robert G Kelly[5], Shahragim Tajbakhsh[1,2]***

[1]Department of Developmental and Stem Cell Biology, Institut Pasteur, Paris, France; [2]CNRS UMR 3738, Paris, France; [3]Central European Institute of Technology, Brno University of Technology, Brno, Czech Republic; [4]Department of Human Genetics, University of Utah, Salt Lake City, United States; [5]Aix-Marseille Université, CNRS UMR 7288, IBDM, Marseille, France

**Abstract** In vertebrates, head and trunk muscles develop from different mesodermal populations and are regulated by distinct genetic networks. Neck muscles at the head-trunk interface remain poorly defined due to their complex morphogenesis and dual mesodermal origins. Here, we use genetically modified mice to establish a 3D model that integrates regulatory genes, cell populations and morphogenetic events that define this transition zone. We show that the evolutionary conserved cucullaris-derived muscles originate from posterior cardiopharyngeal mesoderm, not lateral plate mesoderm, and we define new boundaries for neural crest and mesodermal contributions to neck connective tissue. Furthermore, lineage studies and functional analysis of *Tbx1*- and *Pax3*-null mice reveal a unique developmental program for somitic neck muscles that is distinct from that of somitic trunk muscles. Our findings unveil the embryological and developmental requirements underlying tetrapod neck myogenesis and provide a blueprint to investigate how muscle subsets are selectively affected in some human myopathies.
DOI: https://doi.org/10.7554/eLife.40179.001

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 20

**Received:** 17 July 2018
**Accepted:** 17 November 2018
**Published:** 19 November 2018

**Reviewing editor:** Clare Blackburn, MRC Centre for Regenerative Medicine, University of Edinburgh, United Kingdom



## Introduction

The neck is composed of approximately 80 skeletal muscles in humans that allow head mobility, respiration, swallowing and vocalization processes, containing essential elements such as the trachea, esophagus, larynx, and cervical vertebrae. These processes are ensured by a robust network of muscles at the head-trunk interface, a transition zone subjected to a spectrum of human muscle diseases such as dropped head syndrome, oculopharyngeal myopathy, myotonic dystrophy, Duchenne-type dystrophy and congenital muscular disorders (***Emery, 2002***; ***Martin et al., 2011***; ***Randolph and Pavlath, 2015***). Defining the embryology of these distinct muscle groups is critical to understand the mechanisms underlying the susceptibility of specific muscles to muscular dystrophies. While myogenesis at the cranial and trunk levels has been studied extensively, the developmental mechanisms at the basis of neck muscle formation are poorly documented and often debated (***Ericsson et al., 2013***).

In vertebrates, head and trunk muscles arise from different mesodermal origins and their development depends on distinct myogenic programs. At the cranial level, the cardiopharyngeal mesoderm (CPM) resides in pharyngeal arches and gives rise to branchiomeric muscles and the second heart field. CPM specification is initiated by the activation of genes such as *Mesp1*, *Islet1* and *Tbx1*, while *Pax7* subsequently marks muscle stem cells (***Diogo et al., 2015***; ***Kelly et al., 2004***; ***Nathan et al.,***

2008; Saga et al., 1996; Sambasivan et al., 2009). In contrast, *Pax3* and *Pax7* are expressed in somitic mesoderm giving rise to trunk and limb muscles, with *Pax3* then being downregulated in most muscles during fetal stages, while *Pax7* maintains the stem cell pool (Kassar-Duchossoy et al., 2005; Relaix et al., 2005; Tajbakhsh et al., 1997). After the differential specification of cranial and trunk progenitors, the bHLH myogenic regulatory factors (MRFs) Myf5, Mrf4, Myod and Myog regulate myogenic cell fate and differentiation (reviewed in [Comai and Tajbakhsh, 2014; Noden and Francis-West, 2006]).

In early embryos, *Tbx1* is required for robust activation of MRF genes and proper branchiomeric muscle formation (Grifone et al., 2008; Kelly et al., 2004; Kong et al., 2014; Sambasivan et al., 2009). In *Tbx1* mutant embryos, the first pharyngeal arch is hypoplastic and posterior pharyngeal arches do not form resulting in variably penetrant defects of masticatory muscles and absence of muscles derived from more posterior arches (Kelly et al., 2004; Lescroart et al., 2015; Theis et al., 2010). In humans, *TBX1* is a major gene involved in 22q11.2 deletion syndrome (DiGeorge/velo-cardio-facial syndrome), a congenital disease characterized by cardiovascular defects and craniofacial malformations (Papangeli and Scambler, 2013). In contrast, *Pax3* acts upstream of MRF genes in somites and *Pax3* mutants have defects of epaxial and hypaxial muscle formation while double *Pax3/Pax7*-null embryos lack trunk/limb muscles (Brown et al., 2005; Relaix et al., 2005; Tajbakhsh et al., 1997; Tremblay et al., 1998).

The neck constitutes a transition zone characterizing land vertebrates (tetrapods). The major muscle groups in the neck consist of: epaxial back muscles; ventral hypaxial musculature; pharyngeal, laryngeal and esophagus striated muscles located medioventrally; and cucullaris-derived muscles. The cucullaris is a generic term defining putative homologous muscles that are evolutionarily conserved and connect the head and trunk in jawed vertebrates (gnathostomes). In amniotes, the cucullaris represents the embryonic anlage that gives rise to trapezius and sternocleidomastoid muscles which are innervated by the accessory nerve XI (Diogo, 2010; Edgeworth, 1935; Ericsson et al., 2013; Kuratani, 2008; Kuratani et al., 2018; Lubosch, 1938; Tada and Kuratani, 2015).

While the somitic origin of epaxial/hypaxial neck muscles and CPM origin of pharyngeal, laryngeal and esophagus striated muscles are well defined (Gopalakrishnan et al., 2015; Noden, 1983; Tabler et al., 2017), the embryological origin of cucullaris-derived muscles has remained a subject of controversy (Couly et al., 1993; Edgeworth, 1935; Greil, 1913; Huang et al., 1997; Huang et al., 2000; Matsuoka et al., 2005; Noden, 1983; Piatt, 1938; Piekarski and Olsson, 2007). This muscle group was reported to originate either from lateral plate mesoderm (LPM) or CPM populations adjacent to the first three somites in chick and axolotl (Nagashima et al., 2016; Sefton et al., 2016; Theis et al., 2010). However, retrospective lineage analysis indicated that the murine trapezius and sternocleidomastoid muscles are clonally related to second heart-field-derived myocardium and laryngeal muscles, consistent with a CPM origin (Lescroart et al., 2015). Moreover, cucullaris development follows a branchiomeric program and cucullaris-derived muscles were shown to be absent in *Tbx1*-null mice (Kelly et al., 2004; Lescroart et al., 2015; Sefton et al., 2016; Theis et al., 2010). Nevertheless, the source of the cucullaris is still equivocal due to the location of its embryonic anlagen at the interface of cranial, somitic and lateral plate mesodermal populations.

Skeletal elements and muscle-associated connective tissue (MCT) also have distinct embryological origins along the rostro-caudal axis. The connective tissue of branchiomeric and tongue muscles originate from neural crest cells (NCCs) of cranial origin (Evans and Noden, 2006; Köntges and Lumsden, 1996; Noden, 1983; Noden, 1988; Ziermann et al., 2018b). Cranial NCCs also give rise to skeletal components and tendons in the head. In contrast, the skeleton and connective tissue originate from somitic mesoderm in the trunk and from LPM in limbs (Nassari et al., 2017). The neck and shoulder girdle contain skeletal elements and connective tissues of distinct NCC, LPM or somitic origins (Durland et al., 2008; Matsuoka et al., 2005; McGonnell et al., 2001; Nagashima et al., 2016; Tabler et al., 2017; Valasek et al., 2010). It has been suggested that NCCs form both connective tissue and endochondral cells at the attachment sites of neck muscles to shoulders in mouse (Matsuoka et al., 2005). However, studies in non-mammalian animals have contested a NCC contribution to the pectoral girdle (Epperlein et al., 2012; Kague et al., 2012; Ponomartsev et al., 2017).

Therefore, the neck region consists of muscle, skeletal and connective tissue elements of mixed cellular origins, underscoring the difficulty in assigning embryonic identities for these structures. In addition, the genetic requirements for the formation of non-somitic and somitic neck muscles remain

to be defined. To resolve these issues, we used genetic lineage and loss-of-function mice combined with histology, µCT and 3D reconstructions to map the embryological origins of all neck muscles and associated connective tissues. In doing so, we show that cucullaris-derived muscles originate from a posterior CPM population and are differentially affected in *Tbx1*-null mice. Moreover, we identify a unique genetic network involving both *Mesp1* and *Pax3* genes for somite-derived neck muscles and we define a new limit of neural crest contribution to neck connective tissue and shoulder components.

## Results

### Distinct myogenic programs define neck muscle morphogenesis

To investigate the embryological origin of neck muscles in the mouse, we mapped CPM- and somite-derived myogenic cells using lineage-specific *Cre* drivers including *Mef2c-AHF*, *Islet1*, *Mesp1* and *Pax3* (**Figure 1**). The *Mef2c-AHF* (anterior heart field) enhancer is activated in the second heart field and myogenic progenitors of CPM origin (**Lescroart et al., 2010**; **Verzi et al., 2005**). *Islet1* and *Mesp1* genes are both expressed in early CPM and are essential for cardiac development. The *Mesp1* lineage also marks some anterior somitic derivatives (**Cai et al., 2003**; **Harel et al., 2009**; **Saga et al., 2000**; **Saga et al., 1999**). In contrast, *Pax3* is activated in all somitic progenitors and is a key actor during trunk and limb muscle formation (**Relaix et al., 2005**; **Tajbakhsh et al., 1997**; **Tremblay et al., 1998**). Given that the majority of *Mef2c-AHF* derivatives are myogenic cells (**Lescroart et al., 2015**; **Lescroart et al., 2010**; **Verzi et al., 2005**), we analyzed this lineage using $Rosa26^{R\text{-}lacZ/+}$ (*R26R*) reporter mice. *Islet1*, *Mesp1* and *Pax3* genes are also expressed in cells contributing to skeletal components, connective tissues or neurons. To focus on the myogenic lineage, we used $Pax7^{nGFP\text{-}stop/nlacZ}$ ($Pax7^{GPL}$) reporter mice, which mark cells with nuclear β-galactosidase (β-gal) activity following *Cre* recombination (**Sambasivan et al., 2013**).

We first examined embryos after myogenic specification (E10.5 and E11.75), and fetuses when muscles are patterned (E18.5). In $Mef2c\text{-}AHF^{Cre}$;*R26R* embryos, β-gal-positive cells were observed in the mesodermal core of pharyngeal arches at the origin of branchiomeric muscles, in second heart field derivatives, and in the cucullaris anlage (**Figure 1A,E**). A spatiotemporal analysis of the cucullaris using $Myf5^{Cre}$;$Pax7^{GPL}$ and $Myf5^{Cre}$;$R26^{mTmG}$ embryos (**Figure 1—figure supplement 1**) showed that *Myf5*-derived muscle progenitors located at the level of the posterior pharyngeal arches, and adjacent to somites S1-S3 (**Figure 1—figure supplement 1A'**), were innervated by the accessory nerve XI (**Figure 1—figure supplement 1G–G''**). These cells gave rise to the trapezius and sternocleidomastoid muscles (**Figure 1—figure supplement 1A–F'**) thus confirming the identity of the cucullaris anlage in mouse (**Tada and Kuratani, 2015**).

In $Islet1^{Cre}$;$Pax7^{GPL}$ and $Mesp1^{Cre}$;$Pax7^{GPL}$ embryos, labeling was also observed in pharyngeal arch derivatives and the cucullaris (**Figure 1B–C,F–G**), the latter showing less contribution from the *Islet1* lineage. On sections, a subset of the Myod-positive cells in the cucullaris originated from *Islet1*-derived cells (**Figure 1—figure supplement 2A**). Surprisingly, $Pax3^{Cre}$;$Pax7^{GPL}$ embryos also showed *lacZ* expression in the cucullaris at E11.75, although no expression was detected at E10.5 (**Figure 1D,H**). Given that *Pax3* and *Pax7* are also expressed in neural crest cells (**Relaix et al., 2004**), and that these *Pax3*/*Pax7*-derived cells were excluded from the Myod-positive myogenic population at E12.5 after muscle specification (**Figure 1—figure supplement 2B**), they are likely to be of NCC origin. As expected, *Pax3* lineage tracing also labeled the somite-derived myotomes, hypaxial migrating progenitors that form the hypoglossal cord (origin of tongue and infrahyoid muscles), and limb muscle progenitors. Furthermore, the hypaxial anlage, which is located at the proximal limb bud and gives rise to the cutaneous maximus and latissimus dorsi muscles, was *Pax3*-derived (**Figure 1D,H**; **Figure 1—figure supplement 1D'**) (**Prunotto et al., 2004**; **Tremblay et al., 1998**). Unexpectedly, this anlage and the latissimus dorsi muscle were also labeled in $Islet1^{Cre}$;$Pax7^{GPL}$ but not in $Mesp1^{Cre}$;$Pax7^{GPL}$ embryos (**Figure 1F–G,J–K**). On sections at E12.5, Islet1 expression was observed in *Pax3*-derived cells after the emergence of myogenic cells from the proximal limb bud (**Figure 1—figure supplement 2C**). In addition, the *Mesp1* lineage contributed to anterior somitic derivatives during early embryonic development as previously reported (**Loebel et al., 2012**; **Saga et al., 1999**); strong *lacZ* expression was observed in the hypoglossal

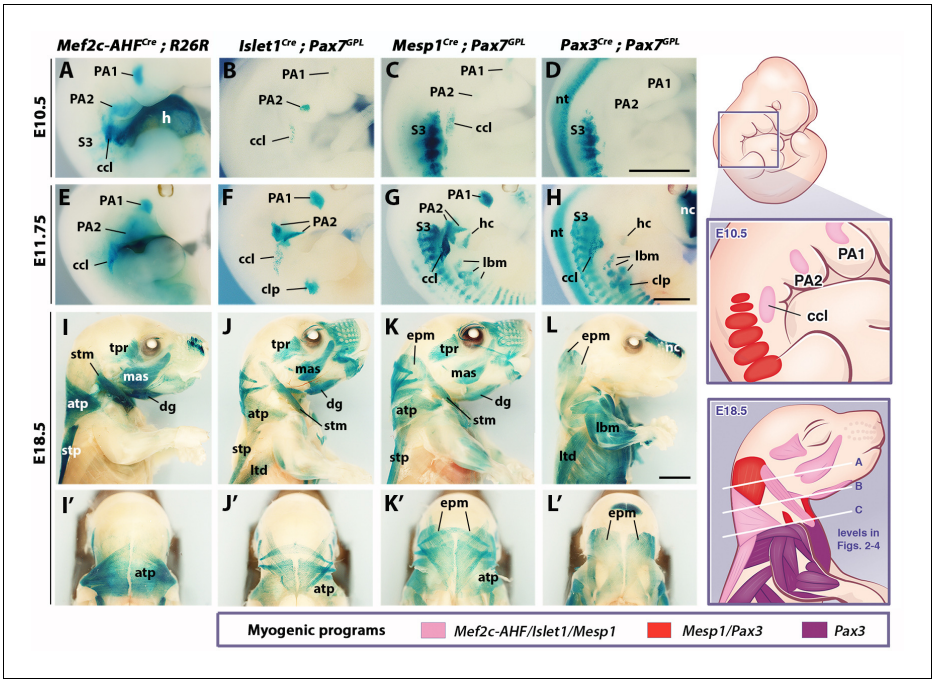



**Figure 1.** Genetic lineage tracing of neck muscle progenitors. Whole-mount X-gal stainings of *Mef2c-AHF*$^{Cre}$; *R26R*, *Islet1*$^{Cre}$;*Pax7*$^{GPL}$, *Mesp1*$^{Cre}$;*Pax7*$^{GPL}$ and *Pax3*$^{Cre}$;*Pax7*$^{GPL}$ mice at E10.5 (**A–D**), E11.75 (**E–H**) and E18.5 (**I–L'**) (n = 3 for each condition). See associated ***Figure 1—supplements 1–3***. (**A–H**) Note labeling of mesodermal core of pharyngeal arches (PAs) and cucullaris anlage (ccl) by *Mef2c-AHF, Islet1* and *Mesp1* lineage reporters; β-gal$^{+}$ cells in anterior somites of *Mesp1*$^{Cre}$ embryos and in the *clp* anlagen of *Islet1*$^{Cre}$ embryos. *Pax3* lineage marked somitic mesoderm. (**I–L'**) *Mef2c-AHF, Islet1* and *Mesp1* lineages marked branchiomeric (mas, tpr, dg) and cucullaris muscles (stm, atp and stp). *Pax3*$^{Cre}$ and *Mesp1*$^{Cre}$ labeled somitic epaxial neck muscles (epm). atp, acromiotrapezius; ccl, cucullaris anlage; clp, cutaneous maximus/latissimus dorsi precursor; dg, digastric; epm, epaxial musculature; h, heart; hc, hypoglossal cord; lbm, limb muscle anlagen and limb muscles; ltd, latissimus dorsi; mas, masseter; nc, nasal capsule; nt, neural tube; PA1-2, pharyngeal arches 1–2; S3, somite 3; stm, sternocleidomastoid; stp, spinotrapezius; tpr; temporal. Scale bars: in D for A-D and in H for E-H, 1000 μm; in L for I-L', 2000 μm.

DOI: https://doi.org/10.7554/eLife.40179.002

The following figure supplements are available for figure 1:

**Figure supplement 1.** Ontogenetic analysis of *Myf5* muscle progenitors at the head-trunk interface.

DOI: https://doi.org/10.7554/eLife.40179.003

**Figure supplement 2.** *Mef2c-AHF, Islet1, Mesp1* and *Pax3* lineage tracings using *lacZ* reporters.

DOI: https://doi.org/10.7554/eLife.40179.004

**Figure supplement 3.** *Mesp1* and *Pax3* lineage tracings in somitic neck muscles using the *Pax7*$^{GPL}$ reporter.

DOI: https://doi.org/10.7554/eLife.40179.005

cord and somites S1-S6. Labeling decreased in more posterior myotomes and in forelimb muscle progenitors compared to *Pax3*$^{Cre}$;*Pax7*$^{GPL}$ embryos (***Figure 1C–D,G–H***).

Lineage tracings with *Mef2c-AHF*$^{Cre}$, *Islet1*$^{Cre}$ and *Mesp1*$^{Cre}$ marked branchiomeric (temporal, masseter, digastric, mylohyoid and pharyngeal) and cucullaris-derived neck muscles (acromiotrapezius, spinotrapezius and sternocleidomastoid), all of which were excluded from the *Pax3* lineage (***Figure 1I–L***, ***Figure 1—figure supplement 2D–G'***). These findings support previous studies showing that cucullaris muscle development is controlled by a branchiomeric myogenic program (***Kelly et al., 2004***; ***Lescroart et al., 2015***; ***Sefton et al., 2016***; ***Theis et al., 2010***). In addition, both

*Mesp1* and *Pax3* lineages labeled somitic neck muscles (***Figure 1K–L'***, ***Figure 1—figure supplement 2F–G'***).

Analysis of different somite-derived neck muscles on sections showed that *Mesp1* and *Pax3* lineages gave rise to the great majority of the Pax7-positive myogenic population (***Figure 1—figure supplement 3***), demonstrating the high recombination efficiency obtained with the *Cre* lines. The results indicate that neck somitic muscles originate from myogenic cells that have expressed both *Mesp1* and *Pax3* genes.

To further investigate the contributions of *Mesp1* and *Pax3* lineages to neck muscles, we examined sections using the $R26^{tdTomato}$ reporter co-immunostained with the myofibre marker Tnnt3 at three representative levels (A, B and C levels in ***Figure 1***; see also ***Figure 2—figure supplement 1***). At anterior levels, while *Pax3* lineage contribution was limited to somite-derived neck muscles, the *Mesp1* lineage marked branchiomeric muscles (mylohyoid, pharyngeal, laryngeal, esophagus), cucullaris-derived muscles (acromiotrapezius and sternocleidomastoid) and somite-derived neck muscles (***Figure 2A–H***, ***Figure 1—figure supplement 2F–G'***, ***Figure 2—figure supplement 2A–H'***). The epaxial and hypaxial neck muscles showed equivalent Tomato expression in both $Mesp1^{Cre};R26^{tdTomato}$ and $Pax3^{Cre};R26^{tdTomato}$ mice. These observations further indicate that *Mesp1* and *Pax3* lineages contribute equivalently to neck muscles derived from anterior somites.

At the shoulder level, we observed less *Mesp1* contribution to more posterior somitic muscles (***Figure 2I–J***). In contrast to that observed at anterior levels, little or no Tomato expression was detected in myofibres of scapular muscles in $Mesp1^{Cre};R26^{tdTomato}$ mice (***Figure 2—figure supplement 2I–J'***). Therefore, *Mesp1* lineage contribution was restricted to epaxial and hypaxial neck muscles, in contrast to pectoral and trunk muscles that originate from the *Pax3* lineage (***Figures 1–2***) (***Table 1***). These observations lead us to propose that three distinct myogenic programs are involved in the formation of neck and pectoral musculature at the head-trunk interface. The branchiomeric and cucullaris-derived muscles depend on a common myogenic program involving *Mef2c-AHF*, *Islet1* and *Mesp1* lineages; the somitic neck muscles that originate from anterior somites derive from both *Mesp1* and *Pax3* lineages; the pectoral muscles derived from more posterior somites depend on the activation of *Pax3* only (***Table 1***).

## Dual neural crest and mesodermal origins of neck connective tissues

To define the cellular origin of neck muscle-associated connective tissue (MCT), we traced the contribution of different embryonic populations using $Mesp1^{Cre};R26^{tdTomato}$ and $Pax3^{Cre};R26^{tdTomato}$ mice as well as $Wnt1^{Cre}$ and $Prx1^{Cre}$ reporters that label NCC and postcranial LPM derivatives, respectively (***Burke and Nowicki, 2003***; ***Danielian et al., 1998***; ***Durland et al., 2008***). Both NCC and LPM populations were reported to contribute to trapezius MCT (***Durland et al., 2008***; ***Matsuoka et al., 2005***). Moreover, it was suggested that the postcranial LPM is a source for cucullaris-derived muscles (***Theis et al., 2010***). A direct comparison of NCC and LPM derivatives allowed us to clarify the contribution of these two populations to cucullaris formation (***Figures 3–4***).

We first investigated the distribution of neck muscles and NCCs using $Myf5^{nlacZ/+}$, $Mef2c\text{-}AHF^{Cre};R26R$, $Pax3^{Cre};R26R$ and $Wnt1^{Cre};R26R$ embryos (***Figure 3—figure supplement 1***). At E10.5, the cucullaris anlage was positioned at the level of posterior pharyngeal arches where *Wnt1*-derived-positive cells were detectable (***Figure 1A–C***, ***Figure 1—figure supplement 1A'***, ***Figure 3—figure supplement 1A–B***). Subsequently, the cucullaris progenitors expanded caudally from E11.5 to E13.5. The posterior limit of the cranial NCC domain also extended posteriorly; however, the *Wnt1*-labeled cells did not cover the posterior portion of cucullaris-derived muscles (***Figure 3—figure supplement 1C–H***). At E14.5, the acromiotrapezius and spinotrapezius attained their definitive position in $Myf5^{nlacZ/+}$ and $Mef2c\text{-}AHF^{Cre};R26R$ embryos (***Figure 3—figure supplement 1I–J***). *Wnt1*-derived cells were observed in the anterior acromiotrapezius muscle, but not in the spinotrapezius that was situated in a *Pax3*-derived domain (***Figure 3—figure supplement 1K–L***). Analysis of whole-mount embryos indicated that the cranial NCCs did not contribute to connective tissue of posterior cucullaris derivatives, in contrast to what was reported previously (***Matsuoka et al., 2005***).

To further analyze NCC contribution to the cervical region at the cellular level, we performed immunostainings on sections for Tomato and Tnnt3 in E18.5 $Wnt1^{Cre};R26^{tdTomato}$ fetuses (***Figure 3***, ***Figure 3—figure supplement 2***). Given that the *Wnt1* lineage is a source of both neuronal and connective tissue derivatives, we associated Tomato immunostaining with Tuj1 that marks neuronal cells and with Tcf4 that labels MCT fibroblasts (***Figure 3***, ***Figure 3—figure supplements 2–3***). At the

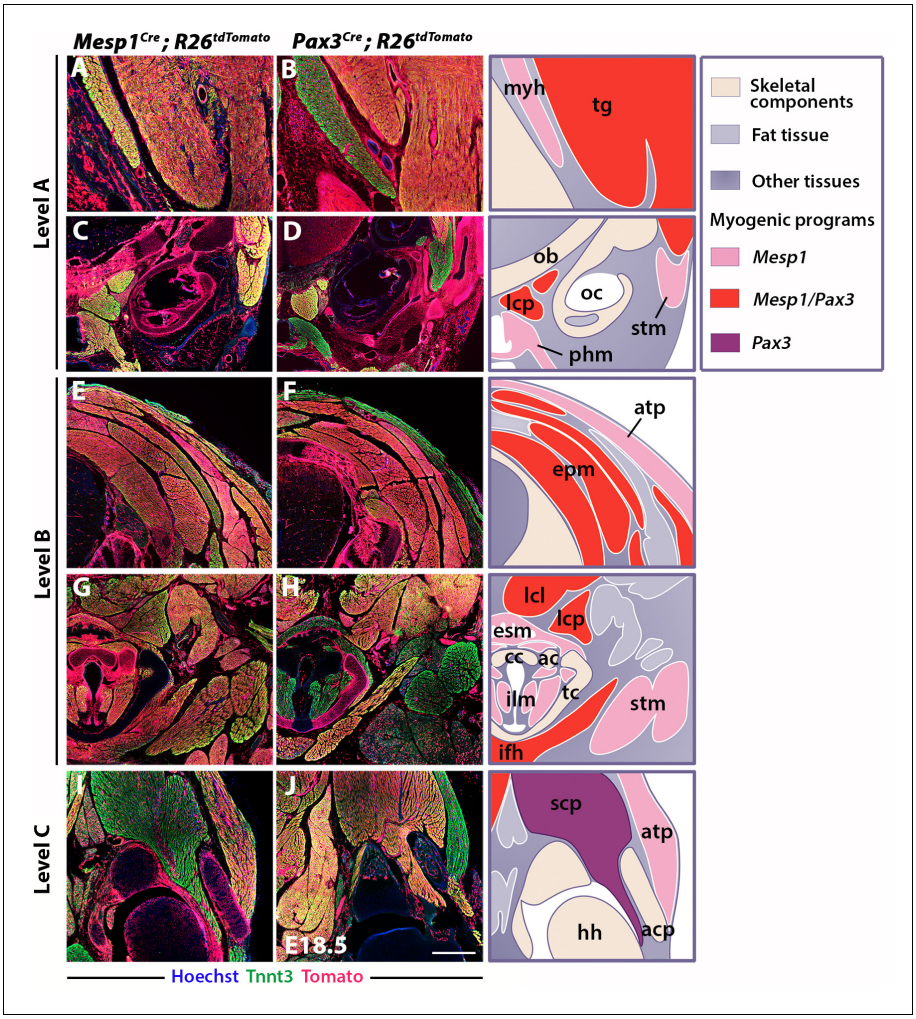



**Figure 2.** Differential contributions of *Mesp1* and *Pax3* lineages to neck and shoulders. Immunostainings on coronal cryosections of E18.5 $Mesp1^{Cre}$;$R26^{tdTomato}$ and $Pax3^{Cre}$;$R26^{tdTomato}$ mice for the myofibre Tnnt3 and Tomato markers at levels indicated in ***Figure 1***. Higher magnifications of selected areas in (**A–J**) are shown in ***Figure 2—figure supplement 2***; (n = 2 for each condition). See also the atlas of neck musculature in ***Figure 2—figure supplement 1***. (**A–H**) $Mesp1^{Cre}$ labeled all neck muscles including branchiomeric (myh, esm, phm and ilm), cucullaris (stm, atp), somitic epaxial (epm) and hypaxial (tg, lcp, lcl, ifh) muscles. $Pax3^{Cre}$ marked somitic muscles. (**I–J**) At shoulder level, *Mesp1*-derived cells did not contribute to posterior somitic myofibres including scapular muscles (scp) compared to that observed in $Pax3^{Cre}$ embryos. ac, arytenoid cartilage; acp, scapular acromion process; atp, acromiotrapezius; cc, cricoid cartilage; epm, epaxial musculature; esm, esophagus striated muscle; hh, humeral head; ifh, infrahyoid muscles; ilm, intrinsic laryngeal muscles; lcl, longus colli; lcp, longus capitis; myh, mylohyoid; ob, occipital bone; oc, otic capsule; phm, pharyngeal muscles; stm, sternocleidomastoid; scp, scapular musculature; tc, thyroid cartilage; tg, tongue. Scale bars: in J for A-B 200 µm, for C-J 400 µm.

DOI: https://doi.org/10.7554/eLife.40179.006

The following figure supplements are available for figure 2:

**Figure supplement 1.** Atlas of neck musculature in mouse.

DOI: https://doi.org/10.7554/eLife.40179.007

**Figure supplement 2.** *Mesp1* and *Pax3* lineage contributions to neck and shoulder muscles.

DOI: https://doi.org/10.7554/eLife.40179.008

**Table 1.** Contribution of *Mef2c-AHF*, *Islet1*, *Mesp1* and *Pax3* lineages to neck and pectoral musculature.

| ***Mef2c/Islet1/Mesp1*-derived muscles** | ***Mesp1/Pax3*-derived muscles** | ***Pax3*- derived muscles** |
|---|---|---|
| Mylohyoid<br>Digastric muscles<br><br>Pharyngeal muscles<br>Intrinsic laryngeal muscles<br>Esophagus striated muscle<br><br>Sternocleidomastoid<br>Acromiotrapezius<br>Spinotrapezius | Epaxial neck muscles (splenius, semispinalis, levator scapula, rhomboid occipitalis, suboccipital and postvertebral muscles)<br><br>Hypaxial neck muscles (tongue muscles*, infrahyoid muscles, longus capitis, longus colli) | Scapular muscles (supraspinatus, Infraspinatus, subscapularis)<br><br>Pectoralis<br><br>Latissimus dorsi†<br>Cutaneous maximus† |
| Branchiomeric myogenic program | Anterior-most somite myogenic program | More posterior somite myogenic program |

*Including intrinsic and extrinsic tongue muscles of somitic origin

†Also derived from an *Islet1* lineage

DOI: https://doi.org/10.7554/eLife.40179.009

cranial level, the MCT of branchiomeric (masseter, mylohyoid), tongue and acromiotrapezius muscles was derived from *Wnt1*- and *Pax3*-lineages but not from the mesodermal *Mesp1* lineage (***Figure 3—figure supplement 2A–B'***, ***Figure 3—figure supplement 3A,F***, ***Figure 3—figure supplement 4A–D,G***). The acromiotrapezius showed a high contribution from *Wnt1*-derived cells while the underlying epaxial muscles had considerably less labeled cells that were limited to the neuronal Tuj1-positive population (***Figure 3A–A'***). The *Wnt1* lineage gave rise to Tcf4-positive fibroblasts in the acromiotrapezius, but not in epaxial neck muscles, where fibroblasts were derived from the *Mesp1* lineage (***Figure 3—figure supplements 3B–C*** and ***4E***). These observations are in accordance with a NCC origin of branchiomeric, anterior trapezius and tongue connective tissue as reported previously (***Matsuoka et al., 2005***).

However, the NCC contribution to connective tissue in the sternocleidomastoid subset of cucullaris-derived muscles appeared more heterogeneous than that observed in the acromiotrapezius. In rodents, the sternocleidomastoid is composed of three individual muscles (cleidomastoid, sternomastoid and cleido-occipitalis portions); a differential NCC contribution to MCT was observed in these muscles. While *Wnt1*-derived NCCs were widely present in the sternomastoid and cleido-occipitalis, the NCC contribution was limited in the cleidomastoid (***Figure 3B–B'***). Indeed, Tcf4-positive fibroblasts in the cleido-occipitalis originated from the *Wnt1* lineage, whereas the majority of MCT fibroblasts in the cleidomastoid were derived from the *Mesp1* lineage (***Figure 3—figure supplements 3D–E*** and ***4F***).

A differential contribution of NCCs to connective tissue was also seen within the laryngeal and infrahyoid musculature. Extensive *Wnt1* lineage contributions to MCT was observed in laryngeal muscles (thyroarytenoid and cricothyroid) that connect to the thyroid cartilage, which is of NCC origin (***Figure 3C–C'***) (***Tabler et al., 2017***). In contrast, the laryngeal muscles (cricoarytenoid and vocal muscles) that link mesoderm-derived laryngeal cartilages (cricoid, arytenoid and medio-caudal portion of the thyroid) did not contain NCC-derived connective tissue (***Figures 2G–H*** and ***3C–C'***) (***Tabler et al., 2017***). In these muscles, the *Wnt1*-derived cells were neuronal, as observed in the esophagus, whereas the MCT fibroblasts were derived from the *Mesp1* lineage (***Figure 3C–C'***, ***Figure 3—figure supplements 2D–D'*** and ***4H***). As another example, *Wnt1*-derived cells contributed to a greater extent to MCT in infrahyoid muscles (thyrohyoid muscles) that connect the hyoid and thyroid cartilage that are of NCC origin, compared to infrahyoid muscles (omohyoid and sternohyoid muscles) that link posteriorly pectoral structures of mesodermal origin (***Figure 3—figure supplement 2C,C', H***; ***Figure 3—figure supplement 3G–H***). These observations suggest that MCT composition within laryngeal and infrahyoid muscles correlates in part with the embryonic origin of the skeletal components to which they attach (***Figure 2G–H***, ***Figure 3C–C'***, ***Figure 3—figure supplement 2C–C', H***).

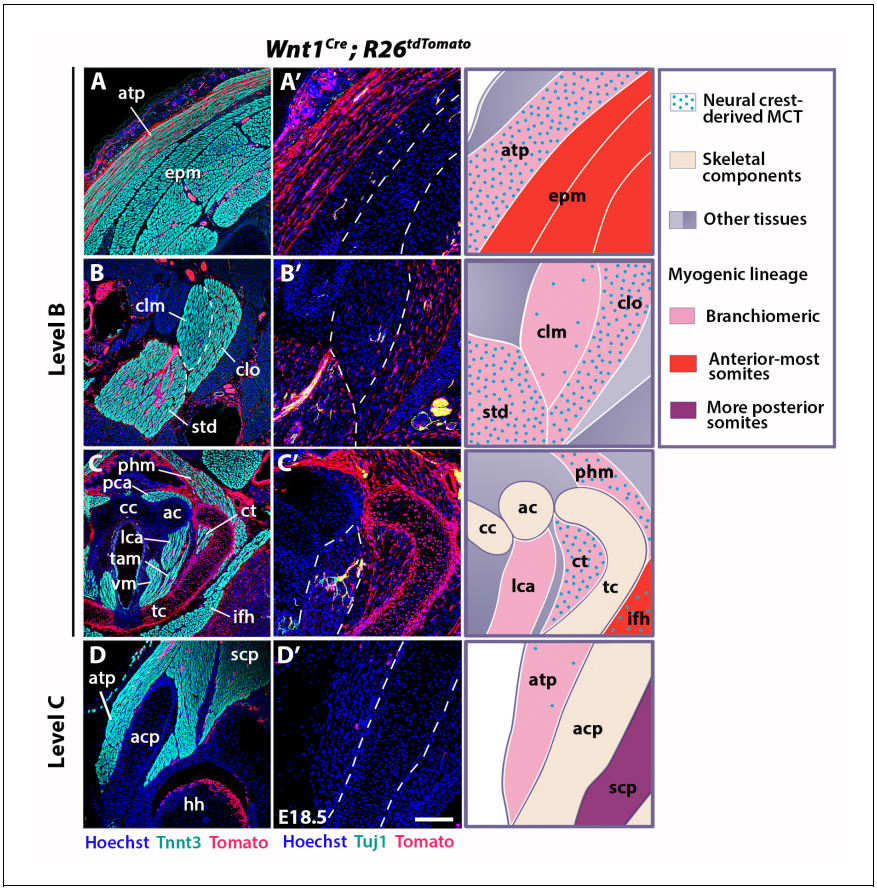



**Figure 3.** Neural crest contribution to neck muscle-associated tissue. Immunostainings on coronal cryosections of E18.5 *Wnt1*$^{Cre}$;*R26*$^{tdTomato}$ mice at levels indicated in ***Figure 1***. Tnnt3/Tomato immunostainings are shown in (**A–D**) and immunostainings for Tuj1/Tomato on selected areas of (**A–D**) are shown with higher magnifications in (**A'–D'**). See associated ***Figure 3—figure supplement 1–4***; (n = 2). (**A–A'**) Note high *Wnt1* contribution in the acromiotrapezius but not in epaxial muscles where *Wnt1*-derived cells marked neuronal cells. (**B–C'**) *Wnt1*-derived cells marked differentially the distinct muscles composing the sternocleidomastoid and laryngeal musculatures. (**D–D'**) At shoulder level, the *Wnt1* cells did not contribute to attachment of acromiotrapezius to scapula. ac, arytenoid cartilage; acp, scapular acromion process; atp, acromiotrapezius; cc, cricoid cartilage; clm, cleidomastoid; clo, cleido-occipitalis; ct, cricothyroid; epm, epaxial musculature; hh, humeral head; ifh, infrahyoid muscles; lca, lateral cricoarytenoid; MCT, muscle-associated connective tissue; pca, posterior cricoarytenoid; phm, pharyngeal muscles; scp, scapular musculature; std, sternomastoid; tam, thyroarytenoid muscle; tc, thyroid cartilage; vm, vocal muscle. Scale bars: in D' for A-D 400 μm for A'-D' 200 μm.

DOI: https://doi.org/10.7554/eLife.40179.010

The following figure supplements are available for figure 3:

**Figure supplement 1.** Distribution of developing neck muscles and neural crest cells.

DOI: https://doi.org/10.7554/eLife.40179.011

**Figure supplement 2.** Neural crest contribution to neck and pectoral structures.

DOI: https://doi.org/10.7554/eLife.40179.012

**Figure supplement 3.** *Wnt1* lineage contribution to connective tissue fibroblasts.

DOI: https://doi.org/10.7554/eLife.40179.013

**Figure supplement 4.** Contribution of *Pax3* and *Mesp1* lineages to connective tissue fibroblasts.

DOI: https://doi.org/10.7554/eLife.40179.014

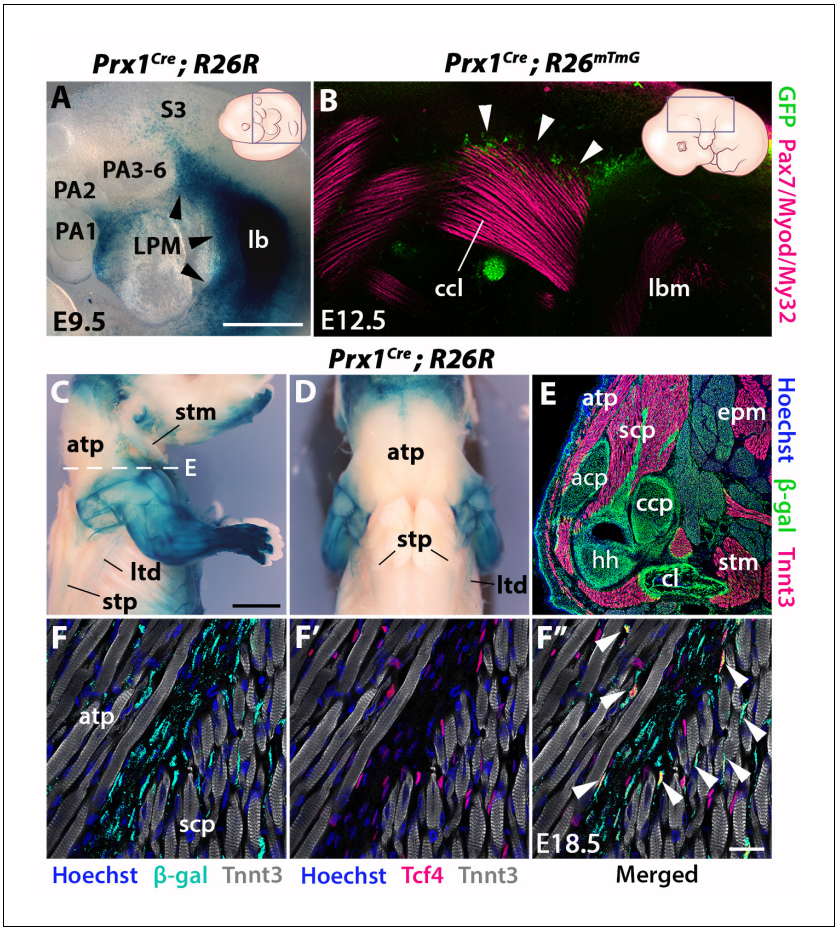



**Figure 4.** *Prx1*-LPM lineage contribution to neck and pectoral girdle. See also ***Figure 4—figure supplement 1***, ***2***. (**A–D**) X-gal stainings of *Prx1*$^{Cre}$;*R26R* reporter mice at E9.5 (n = 3) (**A**) and E18.5 (n = 3) (**C–D**), and immunostaining for GFP and the Pax7/Myod/My32 myogenic markers in *Prx1*$^{Cre}$;*R26*$^{mTmG}$ E12.5 embryo (n = 2) (**B**). Note *Prx1*-derived cells in postcranial LPM (A, black arrowheads) and *Prx1*-derived cells among, but not in, cucullaris myofibres (**B–D**). (**E–F''**) Immunostaining for β-gal, Tnnt3 and Tcf4 on coronal cryosections of E18.5 *Prx1*$^{Cre}$;*R26R* mice (n = 2) showed β-gal$^{+}$ cells constituting the pectoral girdle (E, level C in ***Figure 1***) and in MCT fibroblasts (F-F'', white arrowheads), but not in trapezius myofibres. acp, scapular acromion process; atp, acromiotrapezius; ccl, cucullaris anlage; ccp, scapular coracoid process; cl, clavicle; epm, epaxial musculature; hh, humeral head; lb, forelimb bud; lbm, limb muscle anlagen; LPM, lateral plate mesoderm; ltd, latissimus dorsi; PA1-6, pharyngeal arches 1–6; S3, somite 3; scp, scapular muscles; stm, sternocleidomastoid; stp, spinotrapezius. Scale bars: in A for A, B 500 μm; in C for C-D 2000 μm, for E 500 μm; in F'' for F-F'' 20 μm.

DOI: https://doi.org/10.7554/eLife.40179.015

The following figure supplements are available for figure 4:

**Figure supplement 1.** Comparison of the *Myf5* and *Prx1* lineage tracings.

DOI: https://doi.org/10.7554/eLife.40179.016

**Figure supplement 2.** *Prx1* lineage contribution to neck and limbs.

DOI: https://doi.org/10.7554/eLife.40179.017

Given our findings that connective tissues of neck muscles have differential contributions of NCC and mesodermal populations, we analyzed the caudal connections of the cucullaris-derived muscles to the pectoral girdle (***Figure 3D***, ***Figure 3—figure supplement 2E–H***). The acromiotrapezius

attaches dorsally to the nuchal ligament and ventrally to the scapular acromion process in continuity with the scapular spine. While *Wnt1*-derived cells were present dorsally (***Figure 3A***, ***Figure 3—figure supplement 2E***), this contribution diminished gradually and was undetectable at the insertion on the scapula (***Figure 3D–D'***, ***Figure 3—figure supplement 2F***). Similarly, the sternocleidomastoid muscle showed limited NCC contribution to the attachment sites of the clavicle and sternum (***Figure 3—figure supplement 2G–H***). In contrast to what was previously described (***Matsuoka et al., 2005***), we did not observe NCC contribution to the shoulder endochondral tissue nor to the nuchal ligament (***Figure 3—figure supplement 2E–H***). Taken together, these observations define a novel boundary for neural crest contribution to neck/pectoral components. The posterior contribution limit of neural crest to branchiomeric MCT occurs at the level of laryngeal muscles that connect to NCC skeletal derivatives. Moreover, NCCs do not participate in connecting posterior cucullaris and infrahyoid muscles to their skeletal elements.

To assess the cellular origin of cucullaris connective tissue at posterior attachment sites, we next traced the contribution of lateral plate mesoderm (LPM) to the neck/shoulder region using $Prx1^{Cre}$ reporter mice (***Durland et al., 2008***; ***Logan et al., 2002***) (***Figure 4***, ***Figure 4—figure supplements 1–2***). Analysis of E9.5 embryos showed that *Prx1*-derived cells contribute to the forelimb bud and cells adjacent to the anterior-most somites, but not to pharyngeal arches (***Figure 4A***). At E12.5, the postcranial *Prx1*-derived domain clearly defined the lateral somitic frontier along the rostrocaudal axis (***Durland et al., 2008***) and did not include the cucullaris anlage (***Figure 4—figure supplement 1***, white arrowheads). Whole-mount immunostainings for the myogenic markers Pax7/Myod/My32 and for GFP in $Prx1^{Cre}$;$R26^{mTmG}$ embryos showed that *Prx1*-derived cells were present in the dorsal part of the cucullaris but did not contribute to myofibres (***Figure 4B***, white arrowheads). At E18.5, the *Prx1* lineage marked the limb, scapular and abdominal regions, whereas only a few *Prx1*-derived cells were detected in the cucullaris-derived sternocleidomastoid, acromiotrapezius and spinotrapezius muscles (***Figure 4C–D***). On sections, immunostaining for β-gal and Tnnt3 showed that *Prx1*-derived LPM contributed to limb/shoulder MCT and to skeletal components of the pectoral girdle (***Figure 4E***, ***Figure 4—figure supplement 2A–B***). In contrast, fewer *Prx1*-derived cells were detected in the acromiotrapezius and little or no contribution was observed in the epaxial muscles (***Figure 4E***, ***Figure 4—figure supplement 2B–C***). In addition, only a limited number of *Prx1*-derived cells gave rise to Tcf4-positive fibroblasts in the trapezius muscles, but they contributed more extensively to the fibroblast population in scapular muscles (***Figure 4F–F''***, white arrowheads, ***Figure 4—figure supplement 2D–D''***). Notably, β-gal expression for this lineage was not detected in trapezius myofibres thereby confirming the results obtained at E12.5 (***Figure 4B–F''***, ***Figure 4—figure supplements 1–2***).

Therefore, these observations reveal a dual NCC/LPM origin of trapezius connective tissue, with a decrease of NCC contribution at posterior attachment sites. Moreover, our analysis shows that the postcranial LPM does not give rise to cucullaris myofibres in contrast to what was suggested previously (***Theis et al., 2010***), thus providing further evidence for a branchiomeric origin of the cucullaris.

## Divergent functions of *Tbx1* and *Pax3* in neck development

Given the key role for *Tbx1* and *Pax3* genes in the specification of the CPM and somites respectively, we analyzed the effect of inactivation of these genes on neck muscle formation, compared to the muscle phenotypes observed at cranial and trunk levels.

Analysis has been performed by immunostainings on sections and 3D reconstructions of the neck and pectoral girdle using high-resolution micro-computed tomographic (μCT) scans of control, $Tbx1^{-/-}$ and $Pax3^{-/-}$ fetuses (***Figures 5–6***).

In the early embryo, *Tbx1* is expressed in pharyngeal mesoderm and is required for proper branchiomeric muscle formation (***Grifone et al., 2008***; ***Kelly et al., 2004***). While *Tbx1* is expressed in other cranial populations including the pharyngeal ectoderm and endoderm (***Arnold et al., 2006***; ***Huynh et al., 2007***), the gene is known to be required cell autonomously during CPM myogenesis (***Kong et al., 2014***; ***Zhang et al., 2006***). Analysis of *Tbx1* mutants revealed unexpected features in cucullaris and hypaxial neck muscle formation. As previously described (***Gopalakrishnan et al., 2015***; ***Kelly et al., 2004***), anterior branchiomeric muscles (digastric and mylohyoid) showed phenotypic variations, whereas posterior branchiomeric muscles (esophagus and intrinsic laryngeal muscles) and the acromiotrapezius were severely affected or undetectable (***Figure 5B,E,H***;

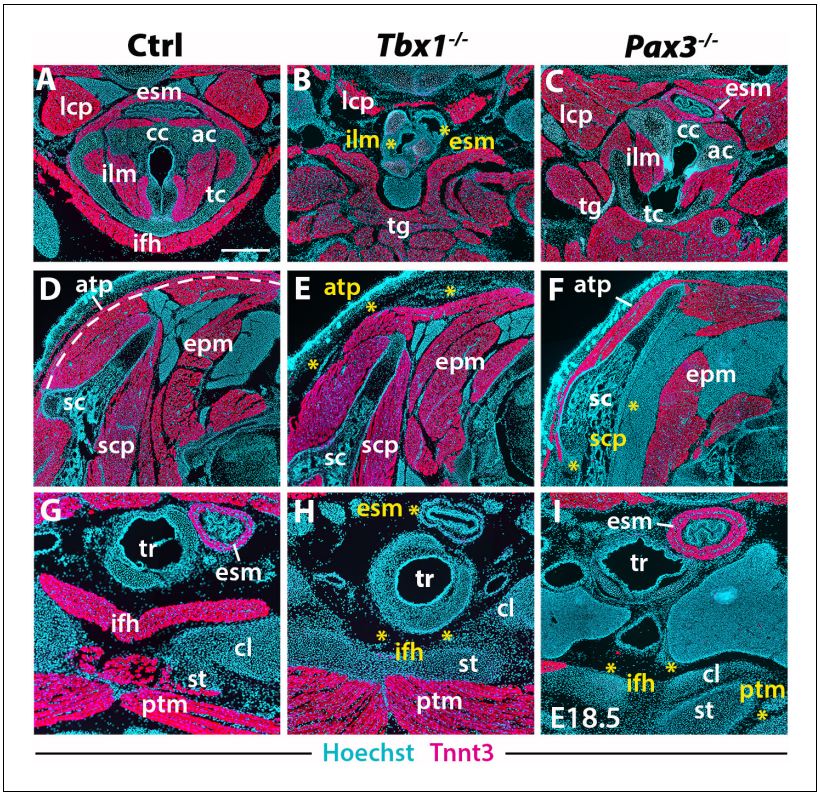



**Figure 5.** Neck muscle phenotypes in *Tbx1* and *Pax3* mutants. (**A–I**) Immunostainings for Tnnt3 on coronal cryosections of control, *Tbx1-null* and *Pax3-null* fetuses at E18.5 (n = 3 each condition). Yellow asterisks indicate missing muscles. Note absence of branchiomeric laryngeal (ilm), esophagus (esm) and trapezius (atp) muscles and severe alteration of somitic infrahyoid muscles (ifh) in *Tbx1* mutants. Scapular (scp) and pectoral (ptm) muscles are missing in *Pax3* mutants. ac, arytenoid cartilage; atp, acromiotrapezius; cc, cricoid cartilage; cl, clavicle; epm, epaxial musculature; esm, esophagus striated muscle; ifh, infrahyoid muscles; ilm, intrinsic laryngeal muscles; lcp, longus capitis; ptm, pectoralis muscles; sc, scapula; scp, scapular muscles; st, sternum; tc, thyroid cartilage; tg, tongue. Scale bars: in A for A-I 500 µm.

DOI: https://doi.org/10.7554/eLife.40179.018

***Figure 6B***) (***Table 2***). However, detailed examination of the cucullaris-derived muscles revealed a heterogeneous dependence on *Tbx1* function that was not reported previously (***Lescroart et al., 2015***; ***Theis et al., 2010***). Unexpectedly, the sternocleidomastoid muscle was present bilaterally but smaller (***Figure 6B***); the different portions (cleido-occipitalis, cleidomastoid and sternomastoid) were unilaterally or bilaterally affected in a stochastic manner. Moreover, while the epaxial neck and scapular muscles were unaffected (***Figure 5E***, ***Figure 6E–H***), the hypaxial neck muscles derived from anterior somites were altered. Indeed, the tongue and longus capitis were reduced and the infrahyoid and longus colli muscles were severely affected or undetectable (***Figure 5B,H***, ***Figure 6E,H***; see interactive 3D PDFs in ***Supplementary file 1–2***).

Analysis of *Pax3* mutants showed that the neck and pectoral muscles were differentially affected. As expected, branchiomeric and epaxial muscles developed normally but displayed morphological differences adapted to malformations noted in some skeletal components (***Figure 5C,F***; ***Figure 6C, I***). However, whereas hypaxial trunk/limb muscles were severely affected or undetectable in *Pax3* mutants (***Figure 5F,I***; ***Figure 6F,I***) (***Tajbakhsh et al., 1997***; ***Tremblay et al., 1998***), surprisingly the majority of hypaxial neck muscles derived from both *Mesp1* and *Pax3* lineages were present. Tongue muscles were reduced in size but patterned, the infrahyoid were hypoplastic, whereas the longus

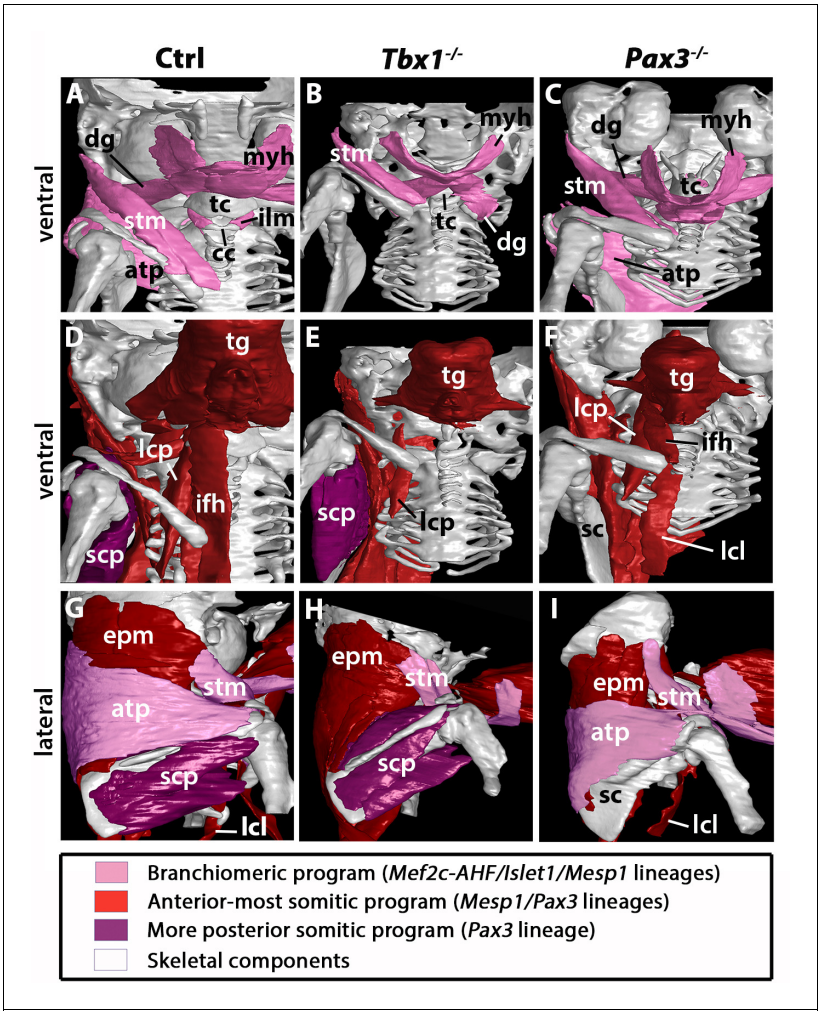



**Figure 6.** 3D reconstructions of neck musculoskeletal system in *Tbx1* and *Pax3* mutants. See interactive 3D PDFs in ***Supplementary file 1–3***; control n = 1; mutants n = 2. (**A–C**) Branchiomeric and cucullaris-derived muscles marked by *Mef2c-AHF/Islet1/Mesp1* lineages are indicated in pink. (**D–F**) Anterior somitic muscles (*Mesp1, Pax3* lineages), in red. (**G–I**) Scapular muscles from more posterior somites (*Pax3* lineage), in violet. atp, acromiotrapezius; cc, cricoid cartilage; dg, digastric muscles; epm, epaxial musculature; ifh, infrahyoid muscles; ilm, intrinsic laryngeal muscles; lcl, longus colli; lcp, longus capitis; myh, mylohyoid; sc, scapula; scp, scapular muscles; stm, sternocleidomastoid; tc, thyroid cartilage; tg, tongue.

DOI: https://doi.org/10.7554/eLife.40179.019

capitis and longus colli were unaffected (***Figure 5C***; ***Figure 6F,I***; see interactive 3D PDF in ***Supplementary file 3***). The phenotypes of the different muscle groups observed in *Tbx1* and *Pax3* mutants are summarized in ***Table 2*** (see also ***Figure 7—figure supplement 1***).

Taken together, these observations revealed that hypaxial muscles in the neck were less affected in *Pax3* mutants than more posterior hypaxial muscles, pointing to distinct requirements for *Pax3* function during neck and trunk muscle formation. In addition, *Tbx1* mutants exhibited more severe phenotypes in hypaxial neck muscles, thus highlighting distinct roles for this gene in branchiomeric and hypaxial neck myogenesis.

**Table 2.** Summary of the neck muscle phenotype observed in *Tbx1*- and *Pax3*-null fetuses.

| | *Tbx1*-null | *Pax3*-null |
|---|---|---|
| Branchiomeric muscles (*Mef2c-AHF/Islet1/Mesp1* lineage) | | |
| Mylohyoid | +/- | ++ |
| Digastric muscles | +/- | ++ |
| Intrinsic laryngeal muscles | – | + |
| Esophagus striated muscle | – | ++ |
| Sternocleidomastoid | +/- | + |
| Acromiotrapezius | – | + |
| Anterior-most somite muscles (*Mesp1/Pax3* lineage) | | |
| Epaxial musculature | ++ | + |
| Longus capitis | +/- | ++ |
| Longus colli | – | ++ |
| Infrahyoid muscles | – | +/- |
| Tongue muscles* | + | + |
| More posterior somite muscles (*Pax3* lineage) | | |
| Scapular muscles | ++ | – |
| Pectoralis | ++ | – |

++,normal; +, altered morphology; +/-, affected; -, severely affected or undetectable

*Including intrinsic and extrinsic tongue muscles of somitic origin

DOI: https://doi.org/10.7554/eLife.40179.020

## Discussion

The embryological origins of neck muscles and connective tissues at the head-trunk interface have been poorly defined largely due to their localization at a transition zone that involves multiple embryonic populations. Using a combination of complementary genetically modified mice and 3D analysis that identifies muscles in the context of their bone attachments, we provide a detailed map of neck tissue morphogenesis and reveal some unexpected features regarding the muscle and connective tissue network.

### Branchiomeric origin of cucullaris-derived muscles

The mammalian neck consists of somitic epaxial/hypaxial muscles, branchiomeric muscles and cucullaris-derived muscles (***Table 1***). The latter constitute a major innovation in vertebrate history, connecting the head to the pectoral girdle in gnathostomes and allowing head mobility in tetrapods (***Ericsson et al., 2013***). Recent studies in different organisms including shark, lungfish and amphibians suggest that the cucullaris develops in series with posterior branchial muscles and that its developmental origin and innervation is conserved among gnathostomes (***Diogo, 2010***; ***Ericsson et al., 2013***; ***Naumann et al., 2017***; ***Noda et al., 2017***; ***Sefton et al., 2016***; ***Tada and Kuratani, 2015***; ***Ziermann et al., 2018a***; ***Ziermann et al., 2017***). However, multiple embryological origins including CPM, LPM and somites have been reported for the cucullaris, underscoring the difficulty in deciphering the morphogenesis of this and other muscles in the head-trunk transition zone (***Huang et al., 2000***; ***Nagashima et al., 2016***; ***Sefton et al., 2016***; ***Theis et al., 2010***).

Our study shows that the cucullaris anlage is innervated by the accessory nerve XI and develops contiguously with the mesodermal core of posterior arches and anterior-most somites 1–3. Our lineage analysis reveals that cucullaris development depends on a branchiomeric myogenic program involving *Mef2c-AHF*, *Islet1* and *Mesp1* lineages in keeping with previous results (***Table 1***) (***Lescroart et al., 2015***; ***Sefton et al., 2016***; ***Theis et al., 2010***). However, our detailed functional analysis and 3D reconstructions lead us to modify the view of the genetic requirements of cucullaris-derived muscles (***Lescroart et al., 2015***; ***Theis et al., 2010***). Notably, these muscles are differentially affected in *Tbx1*-null fetuses; the acromiotrapezius does not form while the sternocleidomastoid is present but reduced. Therefore, *Tbx1* is differentially required for sternocleidomastoid and trapezius formation, suggesting that distinct subprograms regulate cucullaris development.

We also demonstrate that the cucullaris anlage is excluded from the postcranial *Prx1*-derived expression domain, which delineates the trunk LPM field (***Figure 4***). The *Prx1* lineage instead gives rise to connective tissue, thereby excluding a contribution from LPM to cucullaris-derived myofibres. Thus, our results, combined with innervation studies, retrospective clonal analyses and grafting experiments in chick and axolotl (***Lescroart et al., 2015***; ***Nagashima et al., 2016***; ***Sefton et al.,***

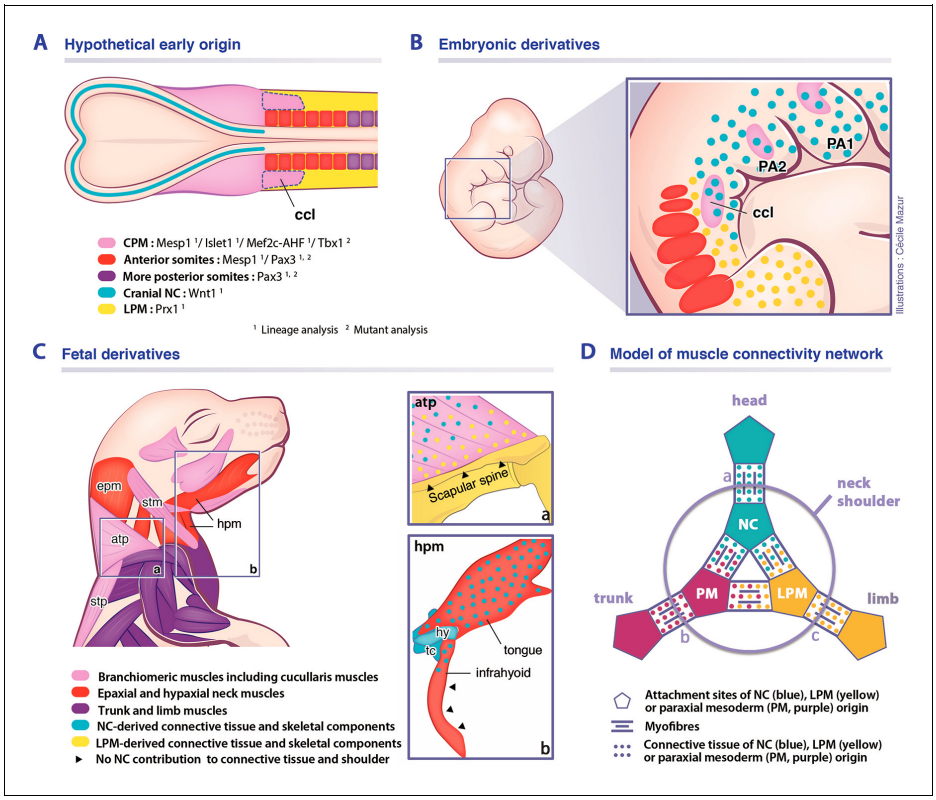



**Figure 7.** Model for musculoskeletal and connective tissue relationships during murine neck development. See also ***Figure 7—figure supplement 1***. (**A, C**) CPM (pink), anterior somites (red) and more posterior somites (violet) muscles are defined by three distinct myogenic programs. (**B**) Note that the cucullaris develops in a NC domain (blue dots), but is excluded from the postcranial LPM (yellow dots). (**C**) Dual NC/LPM origin of trapezius connective tissue is indicated in (**a**). NC contribution to connective tissue extends to tongue and anterior infrahyoid musculature (**b**). (**D**) Mixed origins of muscle connective tissues at the head-trunk-limb interface. Example of representative muscles: (**a**) masseter, (**b**) spinalis dorsi, (**c**) deltoid. atp, acromiotrapezius; ccl, cucullaris; CPM, cardiopharyngeal mesoderm; epm, epaxial neck musculature; hpm, hypaxial neck musculature; hy, hyoid bone; LPM, postcranial lateral plate mesoderm; NC, neural crest; PA1-2, pharyngeal arches 1–2; PM, paraxial mesoderm; stm, sternocleidomastoid; stp, spinotrapezius; tc, thyroid cartilage.

DOI: https://doi.org/10.7554/eLife.40179.021

The following figure supplement is available for figure 7:

**Figure supplement 1.** Muscles affected in *Tbx1* and *Pax3* mutants.

DOI: https://doi.org/10.7554/eLife.40179.022

***2016***; ***Tada and Kuratani, 2015***), suggest that the postcranial extension of the CPM lateral to the first three somites in tetrapod embryos is a source of cucullaris myogenic cells (***Figure 7A***). The discordance with previous studies regarding the origin of the cucullaris is likely due to its proximity to both anterior somites and LPM (***Figure 7A–B***), and consequently, to potential contamination of embryonic sources in grafting experiments (***Couly et al., 1993***; ***Huang et al., 1997***; ***Huang et al., 2000***; ***Noden, 1983***; ***Piekarski and Olsson, 2007***; ***Theis et al., 2010***).

## A unique genetic program for somite-derived neck muscles

Our study also points to a unique *Mesp1/Pax3* genetic program in anterior somites for epaxial/hypaxial neck muscle formation (***Table 1***). While it was shown that the *Mesp1* lineage gives rise to tongue muscles (***Harel et al., 2009***), we demonstrate that it also contributes to all neck muscles. In

chordates, *Mesp* genes appear to play a conserved role in cardiogenesis and axis segmentation. In mouse, *Mesp1* inactivation causes early embryonic death from abnormal heart development, and *Mesp1/Mesp2* double-knockout embryos lack non-axial mesoderm (**Moreno et al., 2008**; **Saga, 1998**; **Saga et al., 2000**; **Satou et al., 2004**; **Sawada et al., 2000**). During early murine development, *Mesp1* shows two waves of activation; initially in the nascent mesoderm destined for extra-embryonic, cranial and cardiac mesoderm at the onset of gastrulation; later during somitogenesis, transient *Mesp1* expression is limited to anterior presomitic mesoderm (**Saga, 1998**; **Saga et al., 1996**; **Saga et al., 2000**; **Saga et al., 1999**). Our lineage analysis shows that *Mesp1* extensively labels the anterior mesoderm, including the CPM and anterior somites 1–6, while contribution decreases in more posterior somites (**Figure 1**) (**Loebel et al., 2012**; **Saga et al., 2000**; **Saga et al., 1999**). Previous fate mapping experiments have shown that the mesoderm of late-streak stage embryos contributes to both CPM and anterior somites (**Parameswaran and Tam, 1995**). It appears that the first wave of *Mesp1* expression defines not only the CPM field but also includes the mesoderm destined for anterior somites. In contrast, the *Mesp1*-labeled cells observed in more posterior somites using the sensitive $Pax7^{GPL}$ reporter may result from the transient wave of *Mesp1* expression in the presomitic mesoderm during axis segmentation. Furthermore, we show that *Mesp1*-derived anterior somites give rise to all epaxial/hypaxial neck muscles in contrast to trunk/limb muscles originating from more posterior somites marked by *Pax3*. The boundary of *Mesp1* lineage contribution to muscles corresponds to the neck/pectoral interface. Our findings indicate that the anterior somitic mesoderm employs a specific transition program for neck muscle formation involving both *Mesp1* and *Pax3* genes implicated in CPM and somitic myogenesis, respectively (**Figure 7A–C**).

Whereas little is known about the function of *Mesp* genes in chordates, there is evidence that *Mesp1* might be differentially required during anterior *versus* posterior somitic formation. In mouse, different *Mesp1* enhancer activities have been observed between CPM/anterior somites and posterior somites indicating that the regulation of *Mesp1* expression might differ in the two embryonic compartments (**Haraguchi et al., 2001**). In zebrafish, quadruple mutants of *Mesp* genes (*Mesp-aa/-ab/-ba/-bb*) lack anterior somite segmentation while the positioning of posterior somite boundaries is unaffected, suggesting distinct requirements for *Mesp* genes in anterior and posterior somites (**Yabe et al., 2016**). Interestingly, during early ascidian development, *Mesp* is expressed in B7.5 founder cells that give rise to both CPM and anterior tail muscles (ATM) (**Satou et al., 2004**). In *Ciona*, the CPM precursors at the origin of heart and atrial siphon (pharyngeal) muscles depend on the ascidian homologs of *Mesp1*, *Islet1* and *Tbx1* (reviewed in [**Diogo et al., 2015**]), indicating that a conserved genetic network promotes chordate myogenesis in the anterior embryonic domain.

Our lineage analysis also reveals an unexpected contribution of *Islet1*-derived cells to the formation of cutaneous maximus and latissimus dorsi muscle progenitors (**Table 1**) (**Prunotto et al., 2004**; **Tremblay et al., 1998**). *Islet1* is activated in a subset of CPM progenitors giving rise to branchiomeric muscles and second heart field myocardium (**Cai et al., 2003**; **Harel et al., 2009**; **Nathan et al., 2008**). At the trunk level, while *Islet1* is widely expressed in the nervous system and in the LPM forming the hindlimb bud (**Cai et al., 2003**; **Yang et al., 2006**), to our knowledge its expression in somitic myogenic cells has not been reported. The cutaneous maximus and latissimus dorsi muscles are missing in both *Pax3* and *Met* mutants (**Prunotto et al., 2004**; **Tajbakhsh et al., 1997**; **Tremblay et al., 1998**). Therefore, the formation of the latissimus dorsi and cutaneous maximus muscles depends on a specific developmental program implicating *Pax3*, *Islet1* and *Met* genes. Given that the latissimus dorsi and cutaneous maximus participated in the gain in mobility of the forelimbs towards the shoulder girdle in tetrapods, our findings provide insights into their genetic and evolutionary origins.

Our detailed analysis of *Tbx1*- and *Pax3*-null mice on sections and in 3D reconstructions now provides a clarified view of neck muscle morphogenesis (**Table 2**). In both *Tbx1* and *Pax3* mutants, whereas the epaxial neck musculature is unaffected, the hypaxial muscles originating from anterior somites are perturbed with a more severe phenotype observed in *Tbx1* mutants (**Table 2**). Whereas no *Tbx1* expression has been reported in early myotomes in somites, *Tbx1* transcripts appear in hypaxial limb and tongue precursors after myogenic specification (**Grifone et al., 2008**; **Kelly et al., 2004**; **Zoupa et al., 2006**). *Tbx1*-null embryos show normal myotomal and limb muscle morphology while the hypoglossal cord is hypoplastic, resulting in reduced tongue musculature (**Table 2**) (**Grifone et al., 2008**; **Kelly et al., 2004**). Therefore, we cannot exclude the possibility that *Tbx1* is activated and plays a role after specification of neck hypaxial muscles (**Okano et al., 2008**;

*Zoupa et al., 2006*). The hypaxial muscle defects might also be secondary to a failure of caudal pharyngeal outgrowth (*Kelly et al., 2004*). While *Tbx1* acts cell autonomously in mesodermal progenitors (*Kong et al., 2014*; *Zhang et al., 2006*), its expression in pharyngeal endoderm might imply an indirect role in CPM myogenesis (*Arnold et al., 2006*). Defects in signaling from pharyngeal endoderm may explain the hypoglossal cord deficiency and the potential non-autonomous role for *Tbx1* in neck hypaxial myogenesis. Detailed analysis of muscle formation in conditional *Tbx1* mutants is needed to resolve the relative roles of *Tbx1* in neck myogenesis.

It has been shown that hypaxial muscles are perturbed to a greater extent than epaxial muscles in *Pax3* mutants (*Tajbakhsh et al., 1997*; *Tremblay et al., 1998*), suggesting a different requirement for *Pax3* in these muscle groups, possibly through differential gene regulation (*Brown et al., 2005*). An unexpected outcome of our analysis was that hypaxial neck muscles (derived from *Mesp1* and *Pax3* lineages) are less perturbed in *Pax3*-null mutants than hypaxial trunk/limb muscles (*Pax3* lineage only) that are severely altered or undetectable (*Table 2*). Our results indicate that *Pax3* is not essential for the formation of neck muscles derived from anterior somites in contrast to hypaxial muscles originating from more posterior somites. These observations support our model that a distinct genetic program governs somitic neck muscles compared to more posterior trunk muscles.

## Connectivity network of the neck and shoulders

Assessing the non-muscle contribution to the neck region is a major challenge due to the extensive participation of diverse cell types from different embryological origins. Previous studies in amphibians, chick and mouse reported that branchiomeric and hypobranchial connective tissue originates from NCCs (*Hanken and Gross, 2005*; *Köntges and Lumsden, 1996*; *Matsuoka et al., 2005*; *Noden, 1983*; *Olsson et al., 2001*; *Ziermann et al., 2018b*). It has been shown that the neural crest provides connective tissue for muscles that link the head and shoulders, whereas mesodermal cells give rise to connective tissue for muscles connecting the trunk and limbs (*Matsuoka et al., 2005*).

Our findings demonstrate that not all branchiomeric muscles are composed of neural crest-derived connective tissue, thereby redefining a new limit for NCC contribution to the neck and shoulders. Unexpectedly, we noted that the contribution of the neural crest lineage is limited in infrahyoid and posterior branchiomeric muscles that connect skeletal components of mesodermal origin. Indeed, it appears that the connective tissue of muscles that link exclusively mesodermal skeletal derivatives is of mesodermal origin. In contrast, the connective tissue of cucullaris-derived muscles is of a mixed origin, first developing in a cranial NCC domain at early stages, then expanding to incorporate connective tissue from both neural crest and LPM populations (*Figure 7B*). While NCCs are present in the anterior acromiotrapezius, sternocleidomastoid and infrahyoid muscles, contribution gradually decreases at posterior attachment sites and is undetectable at scapular level. In parallel, the LPM gives rise to shoulder skeletal components and to connective tissue at the attachment sites of associated musculature including trapezius muscles (*Figure 7C*). Therefore, the dual NCC/LPM origin of the trapezius connective tissue correlates with the embryonic origin of skeletal components to which it is connected.

*Wnt1*$^{Cre}$ and *Sox10*$^{Cre}$ NCC reporter mice were used to show that endochondral cells connecting the cucullaris-derived muscles on the scapula, clavicle and sternum share a common NCC origin with the connective tissue (*Matsuoka et al., 2005*). However, NCCs are not found in pectoral components of fish, axolotl and chick, while contribution to neurocranium is conserved, suggesting that NCC involvement in shoulder formation would be specific to mammals (*Epperlein et al., 2012*; *Kague et al., 2012*; *Piekarski et al., 2014*; *Ponomartsev et al., 2017*). In contrast to this view, our lineage analysis reveals that the neural crest lineage shows limited contribution to cucullaris connective tissue and does not form endochondral cells at the posterior attachment sites (*Figure 7C*). Differences in genetic lineage tracers and reagents might explain these discordant results (*Matsuoka et al., 2005*).

Taken together, our findings indicate that the gradient of neural crest and mesodermal contributions to neck connective tissue depends on the embryonic source of attachment sites. Therefore, it reveals that connective tissue composition in the neck region correlates with the cellular origin of associated skeletal components, independently of the myogenic source or ossification mode, forming a strong link between muscles and bones of the head, trunk and limb fields (*Figure 7D*).

### Evolutionary and clinical perspectives

Our findings demonstrate that the hybrid origin of the skeletal, connective tissue and muscle components of the neck is defined during early embryogenesis. The close proximity of neural crest, CPM, LPM and somitic populations is unique along the body plan and underscores the difficulty in defining their relative contributions to structures in the neck (***Figure 7A–B***). Our results refine the relative contributions of the neural crest and mesodermal derivatives in mouse, thereby providing a coherent view of embryonic components at the head-trunk interface in gnathostomes. Our study highlights the limited NCC contribution to posterior branchiomeric and infrahyoid muscle connective tissue, that is instead of mesodermal origin. This reinforces recent notions suggesting that the cranial NCCs and the postcranial rearrangement of mesodermal populations at the head-trunk interface had been central for the establishment of the neck during gnathostome evolution (***Adachi et al., 2018***; ***Kuratani et al., 2018***; ***Lours-Calet et al., 2014***; ***Nagashima et al., 2016***; ***Sefton et al., 2016***). The contribution of anterior mesoderm in the origin of the neck needs to be elucidated in future studies of gnathostomes.

Our study reveals that neck muscles develop in a complex domain that is distinct from the head and trunk (***Figure 7A–D***), and that might be a contributing factor to pathologies that affect subsets of neck muscles in specific myopathies (***Emery, 2002***; ***Randolph and Pavlath, 2015***). In human, *TBX1* has been identified as a major candidate gene for 22q11.2 deletion syndrome (***Papangeli and Scambler, 2013***). Laryngeal malformations, esophageal dysmotility and shortened neck are frequent in patients. Moreover, the neck deficiencies might not be exclusively due to cervical spine abnormalities but also to neck muscle defects (***Hamidi et al., 2014***; ***Leopold et al., 2012***; ***Marom et al., 2012***). Therefore, our analysis of *Tbx1*-null mutants provides a better understanding of the etiology of the 22q11.2 deletion syndrome and has direct implications in establishing clinical diagnosis in cases where patients present failure in neck-associated functions.

## Materials and methods

**Key resources table**

| Reagent type (species) or resource | Designation | Source or reference | Identifiers | Additional information |
|---|---|---|---|---|
| Strain, strain background (*Mus musculus*) | B6D2F1/JRj | Janvier | | |
| Genetic reagent (*M. musculus*) | *Mef2c-AHF$^{Cre}$* | PMID:16188249 | MGI:3639735 | Dr. Brian L Black (Cardiovascular Research Institute, University of California, USA) |
| Genetic reagent (*M. musculus*) | *Islet1$^{Cre}$* | PMID:11299042 | MGI:2447758 | Dr. Thomas M Jessell (Howard Hughes Medical Institute, Columbia University, USA) |
| Genetic reagent (*M. musculus*) | *Mesp1$^{Cre}$* | PMID:10393122 | MGI:2176467 | Pr. Yumiko Saga (National Institute of Genetics, Japan) |
| Genetic reagent (*M. musculus*) | *Pax3$^{Cre}$* | PMID:22394517 | MGI:3573783 | Dr. Jonathan A. Epstein (Perelman Shool of Medicine, University of Pennsylvania, USA) |
| Genetic reagent (*M. musculus*) | *Myf5$^{Cre}$* | PMID:17418413 | MGI:3710099 | Dr. Mario R Capecchi (Institute of Human Genetics, University of Utah, USA) |
| Genetic reagent (*M. musculus*) | *Wnt1$^{Cre}$* | PMID:9843687 | MGI:J:69326 | Pr. Andrew P. McMahon (Keck School of Medicine of the University of Southern California, USA) |

*Continued on next page*

*Continued*

| Reagent type (species) or resource | Designation | Source or reference | Identifiers | Additional information |
|---|---|---|---|---|
| Genetic reagent (*M. musculus*) | *Prx1$^{Cre}$* | PMID:12112875 | MGI: J:77872 | Dr. Clifford J Tabin (Department of genetics, Harvard Medical School, USA) |
| Genetic reagent (*M. musculus*) | *Pax7$^{GPL}$* | PMID:19531352 | MGI:3850147 | Dr. Shahragim Tajbakhsh (Department of Developmental and Stem Cell Biology, Institut Pasteur, France) |
| Genetic reagent (*M. musculus*) | *Rosa26$^{R\text{-}lacZ}$* | PMID:9916792 | MGI:1861932 | Pr. Philippe Soriano (Icahn School of Medicine at Mt. Sinai, USA) |
| Genetic reagent (*M. musculus*) | *R26$^{mTmG}$* | PMID:17868096 | MGI:3716464 | Pr. Philippe Soriano (Icahn School of Medicine at Mt. Sinai, USA) |
| Genetic reagent (*M. musculus*) | *R26$^{tdTomato}$* | PMID:20023653 | MGI:3809524 | Dr. Hongkui Zeng (Allen Institute for Brain Science, USA) |
| Genetic reagent (*M. musculus*) | *Myf5$^{nlacZ/+}$* | PMID:8918877 | MGI:1857973 | Dr. Shahragim Tajbakhsh (Department of Developmental and Stem Cell Biology, Institut Pasteur, France) |
| Genetic reagent (*M. musculus*) | *Tbx1-null* | PMID:11242110 | MGI:2179190 | Dr. Virginia Papaioannou (Department of Genetics and Development, Columbia University Medical Center, USA) |
| Antibody | Chicken polyclonal anti-β-gal | Abcam | Cat. #: ab9361 | IF (1:1000) |
| Antibody | Rabbit polyclonal anti-β-gal | MP Biomedicals | Cat. #: MP 559761 | IF (1:750) |
| Antibody | Chicken polyclonal anti-GFP | Aves Labs | Cat. #: 1020 | IF (1:500) |
| Antibody | Chicken polyclonal anti-GFP | Abcam | Cat. #: 13970 | IF (1:1000) |
| Antibody | Mouse monoclonal IgG1 anti-Islet1 | DSHB | Cat. #: 40.2D6 | IF (1:1000) |
| Antibody | Mouse monoclonal IgG1 anti-My32 | Sigma | Cat. #: M4276 | IF (1:400) |
| Antibody | Mouse monoclonal IgG1 anti-Myod | Dako | Cat. #: M3512 | IF (1:100) |
| Antibody | Mouse monoclonal IgG1 anti-Pax7 | DSHB | Cat. #: AB_528428 | IF (1:20) |
| Antibody | Rabbit polyclonal anti-Tcf4 | Cell Signalling | Cat. #: C48H11 | IF (1:150) |
| Antibody | Mouse monoclonal IgG1 anti-Tnnt3 | Sigma | Cat. #: T6277 | IF (1:200) |
| Antibody | Rabbit polyclonal anti-Tomato | Clontech | Cat. #: 632496 | IF (1:500) |
| Antibody | Mouse monoclonal IgG2a anti-Pax7 | Ozyme | Cat. #: BLE801202 | IF (1:1000) |
| Software, algorithm | GE phoenix datos\|x 2.0 | GE Sensing and Inspection Technologies GmbH | | |
| Software, algorithm | 3D PDF maker | SolidWorks Corporation | | |
| Software, algorithm | Zen | Zeiss | | |

*Continued on next page*

*Continued*

| Reagent type (species) or resource | Designation | Source or reference | Identifiers | Additional information |
|---|---|---|---|---|
| Chemical compound, drug | X-gal | Fisher | Cat. #: 10554973 | |
| Chemical compound, drug | paraformaldehyde | Electron Microscopy Sciences | Cat. #: 15710 | |
| Chemical compound, drug | Triton X-100 | Sigma | Cat. #: T8787 | |
| Chemical compound, drug | Tween 20 | Sigma | Cat. #: P1379 | |
| Chemical compound, drug | Histoclear II | National Diagnostics | Cat. #: HS-202 | |

## Animals

Animals were handled as per European Community guidelines and the ethics committee of the Institut Pasteur (CTEA) approved protocols (APAFIS#6354–20160809 I2028839). Males carrying the *Cre* driver gene, *Mef2c-AHF*$^{Cre}$ (**Verzi et al., 2005**), *Islet1*$^{Cre}$ (**Srinivas et al., 2001**), *Mesp1*$^{Cre}$ (**Saga et al., 1999**), *Pax3*$^{Cre}$ (**Engleka et al., 2005**), *Myf5*$^{Cre}$ (**Haldar et al., 2007**), *Wnt1*$^{Cre}$ (**Danielian et al., 1998**), *Prx1*$^{Cre}$ (**Logan et al., 2002**), were crossed to reporter females from previously described lines including *Pax7*$^{GPL}$ (**Sambasivan et al., 2013**), *Rosa26*$^{R\text{-}lacZ}$ (*R26R*) (**Soriano, 1999**), *R26*$^{mTmG}$ (**Muzumdar et al., 2007**) and *R26*$^{tdTomato}$ (**Madisen et al., 2010**). *Myf5*$^{nlacZ/+}$ KI mice and mice carrying the *Tbx1*$^{tm1pa}$ allele (referred to as *Tbx1*-null) were previously described (**Jerome and Papaioannou, 2001**; **Kelly et al., 2004**; **Tajbakhsh et al., 1996**). To generate experimental *Pax3*-null fetuses, *Pax3*$^{WT/Cre}$ males and females were intercrossed (**Engleka et al., 2005**) (n = 5 *Tbx1* and *Pax3* mutants analysed including n = 2 by µCT scanning). Mice were crossed and maintained on a B6D2F1/JRj background and genotyped by PCR. Mouse embryos and fetuses were collected between E9.5 and E18.5, with noon on the day of the vaginal plug considered as E0.5.

## X-gal and immunofluorescence stainings

Whole-mount samples were analysed for beta-galactosidase activity with X-gal (0.6 mg/ml) in 1X PBS buffer (D1408, Sigma, St. Louis, MO) containing 4 mM potassium ferricyanide, 4 mM potassium ferrocyanide, 0.02% NP-40 and 2 mM $MgCl_2$ as previously described (**Comai et al., 2014**). For immunostaining on cryosections, foetuses were fixed 3 hr in 4% paraformaldehyde (PFA) (15710, Electron Microscopy Sciences, Hatfield, PA) 0.5% Triton X-100 (T8787, Sigma) at 4˚C, washed overnight at 4˚C in PBS 0.1% Tween 20 (P1379, Sigma), cryopreserved in 30% sucrose in PBS and embedded in OCT for 12–16 µm sectioning with a Leica cryostat (CM3050 S, Leica, Wetzlar, Germany). Cryosections were dried for 30 min and washed in PBS. For immunostaining on paraffin sections, samples were fixed overnight in 4% PFA, dehydrated in graded ethanol series and penetrated with Histoclear II (HS-202, National Diagnostics, Atlanta, GA), embedded in paraffin and oriented in blocks. Paraffin blocks were sectioned at 10–12 µm using a Leica microtome (Reichert-Jung 2035). Sections were then deparaffinized and rehydrated by successive immersions in Histoclear, ethanol and PBS. Samples were then subjected to antigen retrieval with 10 mM Citrate buffer (pH 6.0) using a 2100 Retriever (Aptum Biologics, Rownhams, UK).

Rehydrated sections were blocked for 1 hr in 10% normal goat serum, 3% BSA, 0.5% Triton X-100 in PBS. Primary antibodies were diluted in blocking solution and incubated overnight at 4˚C. Primary antibodies included the following: β-gal (1/1000, chicken polyclonal, ab9361, Abcam, Cambridge, UK; 1/750, rabbit polyclonal, MP 559761, MP Biomedicals, Illkirch, France), GFP (1/500, chick polyclonal, 1020, Aves Labs, Tigard, OR; 1/1000, chick polyclonal, 13970, Abcam), Islet1 (1/1000, mouse monoclonal IgG1, 40.2D6, DSHB), My32 (1/400, mouse monoclonal IgG1, M4276, Sigma), Myod (1/100, mouse monoclonal IgG1, M3512, Dako, Santa Clara, CA), Pax7 (1/20, mouse monoclonal IgG1, AB_528428), Tcf4 (1/150, rabbit polyclonal, C48H11, Cell Signalling, Leiden, Netherlands), Tnnt3 (1/200, monoclonal mouse IgG1, T6277, Sigma), Tomato (1/500, rabbit polyclonal, 632496, Clontech, Shiga, Japan; 1/250, chick polyclonal, 600-901-379, Rockland, Pottstown, PA) and Tuj1 (1/1000, monoclonal mouse IgG2a, BLE801202, Ozyme, Montigny-le-Bretonneux, France). After 3 rounds of

15 min washes in PBS 0.1% Tween 20, secondary antibodies were incubated in blocking solution 2 hr at RT together with 1 µg/ml Hoechst 33342 to visualize nuclei. Secondary antibodies consisted of Alexa 488, 555 or 633 goat anti-rabbit, anti-chicken or anti-mouse isotype specific (1/500, Jackson Immunoresearch, Cambridgeshire, UK). After 3 rounds of 15 min washes in PBS 0.1% Tween 20, slides were mounted in 70% glycerol for analysis.

For whole-mount immunofluorescence staining, embryos were dissected in PBS, fixed in 4% PFA, washed in PBS and stored at −20°C in 100% methanol. After rehydration in PBS, whole mount immunostainings were performed incubating the primary and secondary antibodies for 3 days each. Samples were cleared using benzyl alcohol/benzyl benzoate (BABB) clarification method (***Yokomizo et al., 2012***).

### µCT scan analysis

For µCT scan analysis, the fetuses were treated with the phosphotungstic acid (PTA) contrast agent to well reveal skeletal and muscle structures. After dissection of the cervical region (including the mandible and scapular components, see ***Figure 2—figure supplement 1***), the fetuses were fixed in 4% PFA for 24 hr at 4°C. Samples were then additionally fixed and dehydrated by exchanging the fixative and washing solutions to incrementally increasing ethanol concentrations (30%, 50%, 70%) with 2 days in each concentration to minimize the shrinkage of tissues. To start the contrasting procedure, the embryos were firstly incubated in ethanol-methanol-water mixture (4:4:3) for 1 hr and then transferred for 1 hr into 80% and 90% methanol solution. The staining procedure was then performed for 10 days in 90% methanol 1.5% PTA solution (changed every day with fresh solution) to ensure optimal penetration of the contrast agent. Staining was followed by rehydration of the samples in methanol-grade series (90%, 80%, 70%, 50% and 30%) and stored in sterile distilled water. The samples were placed in polypropylene tubes and embedded in 1% agarose gel to avoid movement artefacts during measurements. µCT scanning was performed using laboratory system GE Phoenix v|tome|x L 240 (GE Sensing and Inspection Technologies GmbH, Hamburg, Germany), equipped with a nanofocus X-ray tube with maximum power of 180 kV/15 W and a flat panel detector DXR250 with 2048 × 2048 pixel2, 200 × 200 $\mu m^2$ pixel size. The µCT scan was carried out at 60 kV acceleration voltage and 200 µA tube current with voxel size of 5.7 µm for all samples. The beam was filtered by a 0.2 mm aluminium filter. The 2200 projections were taken over 360° with exposure time of 900 ms. The tomographic reconstructions were done using the software GE phoenix datos|x 2.0 (GE Sensing and Inspection Technologies GmbH) and data segmentations and visualizations were performed by combination of software VG Studio MAX 2.2 (Volume GraphicsGmbH, Heidelberg, Germany) and Avizo 7.1 (Thermo Fisher Scientific, Waltham, MA) according to (***Tesařová et al., 2016***). The interactive 3D PDFs were set up using 3D PDF maker software.

### Imaging

Images were acquired using the following systems: a Zeiss Axio-plan equipped with an Apotome, a Zeiss stereo zoom microscope V16 or a Zeiss LSM 700 laser-scanning confocal microscope with ZEN software (Carl Zeiss, Oberkochen, Germany). For whole-mount rendering, acquired Z-stacks were 3D reconstructed using Imaris software. All images were assembled in Adobe Photoshop (Adobe Systems, San Jose, CA).

## Acknowledgements

We thank Drs. Claudio Cortes and Françoise Helmbacher for providing transgenic mice and Mirialys Gallardo for technical assistance. We also thank Cécile Mazur for illustrations.

## Additional information

### Funding


| Funder | Grant reference number | Author |
|---|---|---|
| Institut Pasteur | | Eglantine Heude<br>Alexandre Grimaldi<br>Shahragim Tajbakhsh |

| | | |
|---|---|---|
| Agence Nationale de la Recherche | | Eglantine Heude<br>Alexandre Grimaldi<br>Shahragim Tajbakhsh |
| Centre National de la Recherche Scientifique | | Eglantine Heude<br>Estelle Jullian<br>Noritaka Adachi<br>Alexandre Grimaldi<br>Robert G Kelly<br>Shahragim Tajbakhsh |
| French Muscular Dystrophy Association | | Eglantine Heude<br>Alexandre Grimaldi<br>Robert G Kelly<br>Shahragim Tajbakhsh |
| Central European Institute of Technology | | Marketa Tesarova<br>Tomas Zikmund<br>Jozef Kaiser |
| March of Dimes Foundation | | Elizabeth M Sefton<br>Gabrielle Kardon |
| National Institutes of Health | | Elizabeth M Sefton<br>Gabrielle Kardon |
| Fondation pour la Recherche Médicale | | Estelle Jullian<br>Noritaka Adachi<br>Robert G Kelly |
| Fondation Leducq | | Estelle Jullian<br>Noritaka Adachi<br>Robert G Kelly |
| Yamada Science Foundation | | Noritaka Adachi<br>Robert G Kelly |
| Bourses du Gouvernement Français | | Noritaka Adachi |
| project CEITEC 2020 | | Marketa Tesarova<br>Tomas Zikmund<br>Jozef Kaiser |
| The Ministry of Education, Youth and Sports of the Czech Republic under the National Sustainability Programme II | | Marketa Tesarova<br>Tomas Zikmund<br>Jozef Kaiser |
| CEITEC Nano Research Infrastructure | MEYS CR, 2016–2019 | Marketa Tesarova<br>Tomas Zikmund<br>Jozef Kaiser |

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.


### Author contributions

Eglantine Heude, Conceptualization, Validation, Investigation, Visualization, Methodology, Writing—original draft, Writing—review and editing; Marketa Tesarova, Resources, Data curation, Formal analysis, Investigation, Visualization, Methodology; Elizabeth M Sefton, Validation, Investigation, Writing—review and editing; Estelle Jullian, Validation, Investigation; Noritaka Adachi, Alexandre Grimaldi, Investigation, Writing—review and editing; Tomas Zikmund, Resources, Supervision; Jozef Kaiser, Resources, Supervision, Funding acquisition, Project administration; Gabrielle Kardon, Robert G Kelly, Resources, Supervision, Funding acquisition, Project administration, Writing—review and editing; Shahragim Tajbakhsh, Conceptualization, Resources, Supervision, Funding acquisition, Validation, Visualization, Methodology, Project administration, Writing—review and editing


### Author ORCIDs

Elizabeth M Sefton http://orcid.org/0000-0001-6481-612X
Noritaka Adachi http://orcid.org/0000-0002-9482-8436
Shahragim Tajbakhsh http://orcid.org/0000-0003-1809-7202

### Ethics

Animal experimentation: Animals were handled as per European Community guidelines and the ethics committee of the Institut Pasteur (CTEA) approved protocols. (APAFIS#6354-2016080912028839)

### Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.40179.028
Author response https://doi.org/10.7554/eLife.40179.029

## Additional files

### Supplementary files

- Supplementary file 1. Interactive 3D neck reconstruction of a E18.5 control fetus. Download PDF for full details.

DOI: https://doi.org/10.7554/eLife.40179.023

- Supplementary file 2. Interactive 3D neck reconstruction of a E18.5 *Tbx1*-null fetus. Download PDF for full details.

DOI: https://doi.org/10.7554/eLife.40179.024

- Supplementary file 3. Interactive 3D neck reconstruction of a E18.5 *Pax3*-null fetus. Download PDF for full details.

DOI: https://doi.org/10.7554/eLife.40179.025

- Transparent reporting form

DOI: https://doi.org/10.7554/eLife.40179.026

### Data availability

All data generated or analysed during this study are included in the manuscript and supporting files.

## References

**Adachi N**, Pascual-Anaya J, Hirai T, Higuchi S, Kuratani S. 2018. Development of hypobranchial muscles with special reference to the evolution of the vertebrate neck. *Zoological Letters* **4**:5. DOI: https://doi.org/10.1186/s40851-018-0087-x, PMID: 29468087

**Arnold JS**, Werling U, Braunstein EM, Liao J, Nowotschin S, Edelmann W, Hebert JM, Morrow BE. 2006. Inactivation of Tbx1 in the pharyngeal endoderm results in 22q11DS malformations. *Development* **133**:977–987. DOI: https://doi.org/10.1242/dev.02264, PMID: 16452092

**Brown CB**, Engleka KA, Wenning J, Min Lu M, Epstein JA. 2005. Identification of a hypaxial somite enhancer element regulating Pax3 expression in migrating myoblasts and characterization of hypaxial muscle Cre transgenic mice. *Genesis* **41**:202–209. DOI: https://doi.org/10.1002/gene.20116, PMID: 15789408

**Burke AC**, Nowicki JL. 2003. A new view of patterning domains in the vertebrate mesoderm. *Developmental Cell* **4**:159–165. DOI: https://doi.org/10.1016/S1534-5807(03)00033-9, PMID: 12586060

**Cai CL**, Liang X, Shi Y, Chu PH, Pfaff SL, Chen J, Evans S. 2003. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Developmental Cell* **5**:877–889. DOI: https://doi.org/10.1016/S1534-5807(03)00363-0, PMID: 14667410

**Comai G**, Sambasivan R, Gopalakrishnan S, Tajbakhsh S. 2014. Variations in the efficiency of lineage marking and ablation confound distinctions between myogenic cell populations. *Developmental Cell* **31**:654–667. DOI: https://doi.org/10.1016/j.devcel.2014.11.005, PMID: 25490270

**Comai G**, Tajbakhsh S. 2014. Molecular and cellular regulation of skeletal myogenesis. *Current Topics in Developmental Biology* **110**:1–73. DOI: https://doi.org/10.1016/B978-0-12-405943-6.00001-4, PMID: 25248473

**Couly GF**, Coltey PM, Le Douarin NM. 1993. The triple origin of skull in higher vertebrates: a study in quail-chick chimeras. *Development* **117**:409–429. PMID: 8330517

**Danielian PS**, Muccino D, Rowitch DH, Michael SK, McMahon AP. 1998. Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. *Current Biology* **8**:1323–S2. DOI: https://doi.org/10.1016/S0960-9822(07)00562-3

**Diogo RA**. 2010. *Muscles of Vertebrates - Comparative Anatomy, Evolution, Homologies and Development*. Enfield, New Hampshire: Science Publishers. DOI: https://doi.org/10.1201/9781439845622

**Diogo R**, Kelly RG, Christiaen L, Levine M, Ziermann JM, Molnar JL, Noden DM, Tzahor E. 2015. A new heart for a new head in vertebrate cardiopharyngeal evolution. *Nature* **520**:466–473. DOI: https://doi.org/10.1038/nature14435, PMID: 25903628

**Durland JL**, Sferlazzo M, Logan M, Burke AC. 2008. Visualizing the lateral somitic frontier in the Prx1Cre transgenic mouse. *Journal of Anatomy* **212**:590–602. DOI: https://doi.org/10.1111/j.1469-7580.2008.00879.x, PMID: 18430087

**Edgeworth FH**. 1935. *The Cranial Muscles of Vertebrates*. Cambridge University Press.

**Emery AEH**. 2002. The muscular dystrophies. *The Lancet* **359**:687–695. DOI: https://doi.org/10.1016/S0140-6736(02)07815-7

**Engleka KA**, Gitler AD, Zhang M, Zhou DD, High FA, Epstein JA. 2005. Insertion of Cre into the Pax3 locus creates a new allele of Splotch and identifies unexpected Pax3 derivatives. *Developmental Biology* **280**:396–406. DOI: https://doi.org/10.1016/j.ydbio.2005.02.002, PMID: 15882581

**Epperlein HH**, Khattak S, Knapp D, Tanaka EM, Malashichev YB. 2012. Neural crest does not contribute to the neck and shoulder in the axolotl (Ambystoma mexicanum). *PLOS ONE* **7**:e52244. DOI: https://doi.org/10.1371/journal.pone.0052244, PMID: 23300623

**Ericsson R**, Knight R, Johanson Z. 2013. Evolution and development of the vertebrate neck. *Journal of Anatomy* **222**:67–78. DOI: https://doi.org/10.1111/j.1469-7580.2012.01530.x, PMID: 22697305

**Evans DJ**, Noden DM. 2006. Spatial relations between avian craniofacial neural crest and paraxial mesoderm cells. *Developmental Dynamics* **235**:1310–1325. DOI: https://doi.org/10.1002/dvdy.20663, PMID: 16395689

**Gopalakrishnan S**, Comai G, Sambasivan R, Francou A, Kelly RG, Tajbakhsh S. 2015. A Cranial Mesoderm Origin for Esophagus Striated Muscles. *Developmental Cell* **34**:694–704. DOI: https://doi.org/10.1016/j.devcel.2015.07.003, PMID: 26387456

**Greil A**. 1913. Entwicklungsgeschichte des Kopfes und des Blutgefässsystemes von Ceratodus forsteri. II. Die epigenetischen Erwerbungen während der Stadien 39–48. Denkschr Med-Naturwiss Ges Jena. *Denkschriften Der Medicinisch-NaturwissenschaftlichenGesellschaft Zu Jena* **9**:935–1492.

**Grifone R**, Jarry T, Dandonneau M, Grenier J, Duprez D, Kelly RG. 2008. Properties of branchiomeric and somite-derived muscle development in Tbx1 mutant embryos. *Developmental Dynamics* **237**:3071–3078. DOI: https://doi.org/10.1002/dvdy.21718, PMID: 18816853

**Haldar M**, Hancock JD, Coffin CM, Lessnick SL, Capecchi MR. 2007. A conditional mouse model of synovial sarcoma: insights into a myogenic origin. *Cancer Cell* **11**:375–388. DOI: https://doi.org/10.1016/j.ccr.2007.01.016, PMID: 17418413

**Hamidi M**, Nabi S, Husein M, Mohamed ME, Tay KY, McKillop S. 2014. Cervical spine abnormalities in 22q11.2 deletion syndrome. *The Cleft Palate-Craniofacial Journal* **51**:230–233. DOI: https://doi.org/10.1597/12-318, PMID: 24003836

**Hanken J**, Gross JB. 2005. Evolution of cranial development and the role of neural crest: insights from amphibians. *Journal of Anatomy* **207**:437–446. DOI: https://doi.org/10.1111/j.1469-7580.2005.00481.x, PMID: 16313386

**Haraguchi S**, Kitajima S, Takagi A, Takeda H, Inoue T, Saga Y. 2001. Transcriptional regulation of Mesp1 and Mesp2 genes: differential usage of enhancers during development. *Mechanisms of Development* **108**:59–69. DOI: https://doi.org/10.1016/S0925-4773(01)00478-6, PMID: 11578861

**Harel I**, Nathan E, Tirosh-Finkel L, Zigdon H, Guimarães-Camboa N, Evans SM, Tzahor E. 2009. Distinct origins and genetic programs of head muscle satellite cells. *Developmental Cell* **16**:822–832. DOI: https://doi.org/10.1016/j.devcel.2009.05.007, PMID: 19531353

**Huang R**, Zhi Q, Ordahl CP, Christ B. 1997. The fate of the first avian somite. *Anatomy and Embryology* **195**:435–449. DOI: https://doi.org/10.1007/s004290050063, PMID: 9176666

**Huang R**, Zhi Q, Patel K, Wilting J, Christ B. 2000. Contribution of single somites to the skeleton and muscles of the occipital and cervical regions in avian embryos. *Anatomy and Embryology* **202**:375–383. DOI: https://doi.org/10.1007/s004290000131, PMID: 11089929

**Huynh T**, Chen L, Terrell P, Baldini A. 2007. A fate map of Tbx1 expressing cells reveals heterogeneity in the second cardiac field. *Genesis* **45**:470–475. DOI: https://doi.org/10.1002/dvg.20317, PMID: 17610275

**Jerome LA**, Papaioannou VE. 2001. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nature Genetics* **27**:286–291. DOI: https://doi.org/10.1038/85845, PMID: 11242110

**Kague E**, Gallagher M, Burke S, Parsons M, Franz-Odendaal T, Fisher S. 2012. Skeletogenic fate of zebrafish cranial and trunk neural crest. *PLOS ONE* **7**:e47394. DOI: https://doi.org/10.1371/journal.pone.0047394, PMID: 23155370

**Kassar-Duchossoy L**, Giacone E, Gayraud-Morel B, Jory A, Gomès D, Tajbakhsh S. 2005. Pax3/Pax7 mark a novel population of primitive myogenic cells during development. *Genes & Development* **19**:1426–1431. DOI: https://doi.org/10.1101/gad.345505, PMID: 15964993

**Kelly RG**, Jerome-Majewska LA, Papaioannou VE. 2004. The del22q11.2 candidate gene Tbx1 regulates branchiomeric myogenesis. *Human Molecular Genetics* **13**:2829–2840. DOI: https://doi.org/10.1093/hmg/ddh304, PMID: 15385444

**Kong P**, Racedo SE, Macchiarulo S, Hu Z, Carpenter C, Guo T, Wang T, Zheng D, Morrow BE. 2014. Tbx1 is required autonomously for cell survival and fate in the pharyngeal core mesoderm to form the muscles of mastication. *Human Molecular Genetics* **23**:4215–4231. DOI: https://doi.org/10.1093/hmg/ddu140, PMID: 24705356

**Köntges G**, Lumsden A. 1996. Rhombencephalic neural crest segmentation is preserved throughout craniofacial ontogeny. *Development* **122**:3229–3242. PMID: 8898235

**Kuratani S**. 2008. Evolutionary developmental studies of cyclostomes and the origin of the vertebrate neck. *Development, Growth & Differentiation* **50**:S189–S194. DOI: https://doi.org/10.1111/j.1440-169X.2008.00985.x, PMID: 18430164

**Kuratani S**, Kusakabe R, Hirasawa T. 2018. The neural crest and evolution of the head/trunk interface in vertebrates. *Developmental Biology*. DOI: https://doi.org/10.1016/j.ydbio.2018.01.017, PMID: 29408469

**Leopold C**, De Barros A, Cellier C, Drouin-Garraud V, Dehesdin D, Marie JP. 2012. Laryngeal abnormalities are frequent in the 22q11 deletion syndrome. *International Journal of Pediatric Otorhinolaryngology* **76**:36–40. DOI: https://doi.org/10.1016/j.ijporl.2011.09.025, PMID: 22019154

**Lescroart F**, Kelly RG, Le Garrec JF, Nicolas JF, Meilhac SM, 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. DOI: https://doi.org/10.1242/dev.050674, PMID: 20823066

**Lescroart F**, Hamou W, Francou A, Théveniau-Ruissy M, Kelly RG, Buckingham M. 2015. Clonal analysis reveals a common origin between nonsomite-derived neck muscles and heart myocardium. *PNAS* **112**:1446–1451. DOI: https://doi.org/10.1073/pnas.1424538112, PMID: 25605943

**Loebel DA**, Hor AC, Bildsoe H, Jones V, Chen YT, Behringer RR, Tam PP. 2012. Regionalized Twist1 activity in the forelimb bud drives the morphogenesis of the proximal and preaxial skeleton. *Developmental Biology* **362**: 132–140. DOI: https://doi.org/10.1016/j.ydbio.2011.11.020, PMID: 22178153

**Logan M**, Martin JF, Nagy A, Lobe C, Olson EN, Tabin CJ. 2002. Expression of Cre Recombinase in the developing mouse limb bud driven by a Prxl enhancer. *Genesis* **33**:77–80. DOI: https://doi.org/10.1002/gene.10092, PMID: 12112875

**Lours-Calet C**, Alvares LE, El-Hanfy AS, Gandesha S, Walters EH, Sobreira DR, Wotton KR, Jorge EC, Lawson JA, Kelsey Lewis A, Tada M, Sharpe C, Kardon G, Dietrich S. 2014. Evolutionarily conserved morphogenetic movements at the vertebrate head-trunk interface coordinate the transport and assembly of hypopharyngeal structures. *Developmental Biology* **390**:231–246. DOI: https://doi.org/10.1016/j.ydbio.2014.03.003, PMID: 24662046

**Lubosch W**. 1938. Muskeln des Kopfes: Viscerale Muskulatur. In: Bolk L, Göppert E, Kallius E, Lubosch W (Eds). *Handbuch Der Vergleichenden Anatomie Der Wirbeltiere*. Berlin: Urban & Schwarzenberg.

**Madisen L**, Zwingman TA, Sunkin SM, Oh SW, Zariwala HA, Gu H, Ng LL, Palmiter RD, Hawrylycz MJ, Jones AR, Lein ES, Zeng H. 2010. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neuroscience* **13**:133–140. DOI: https://doi.org/10.1038/nn.2467, PMID: 20023653

**Marom T**, Roth Y, Goldfarb A, Cinamon U. 2012. Head and neck manifestations of 22q11.2 deletion syndromes. *European Archives of Oto-Rhino-Laryngology* **269**:381–387. DOI: https://doi.org/10.1007/s00405-011-1745-1, PMID: 21861138

**Martin AR**, Reddy R, Fehlings MG. 2011. Dropped head syndrome: diagnosis and management. *Evidence-Based Spine-Care Journal* **2**:41–47. DOI: https://doi.org/10.1055/s-0030-1267104, PMID: 23637681

**Matsuoka T**, Ahlberg PE, Kessaris N, Iannarelli P, Dennehy U, Richardson WD, McMahon AP, Koentges G. 2005. Neural crest origins of the neck and shoulder. *Nature* **436**:347–355. DOI: https://doi.org/10.1038/nature03837, PMID: 16034409

**McGonnell IM**, McKay IJ, Graham A. 2001. A population of caudally migrating cranial neural crest cells: functional and evolutionary implications. *Developmental Biology* **236**:354–363. DOI: https://doi.org/10.1006/dbio.2001.0330, PMID: 11476577

**Moreno TA**, Jappelli R, Izpisúa Belmonte JC, Kintner C. 2008. Retinoic acid regulation of the Mesp-Ripply feedback loop during vertebrate segmental patterning. *Developmental Biology* **315**:317–330. DOI: https://doi.org/10.1016/j.ydbio.2007.12.038, PMID: 18261720

**Muzumdar MD**, Tasic B, Miyamichi K, Li L, Luo L. 2007. A global double-fluorescent Cre reporter mouse. *Genesis* **45**:593–605. DOI: https://doi.org/10.1002/dvg.20335, PMID: 17868096

**Nagashima H**, Sugahara F, Watanabe K, Shibata M, Chiba A, Sato N. 2016. Developmental origin of the clavicle, and its implications for the evolution of the neck and the paired appendages in vertebrates. *Journal of Anatomy* **229**:536–548. DOI: https://doi.org/10.1111/joa.12502, PMID: 27279028

**Nassari S**, Duprez D, Fournier-Thibault C. 2017. Non-myogenic Contribution to Muscle Development and Homeostasis: The Role of Connective Tissues. *Frontiers in Cell and Developmental Biology* **5**:22. DOI: https://doi.org/10.3389/fcell.2017.00022, PMID: 28386539

**Nathan E**, Monovich A, Tirosh-Finkel L, Harrelson Z, Rousso T, Rinon A, Harel I, Evans SM, 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. DOI: https://doi.org/10.1242/dev.007989, PMID: 18184728

**Naumann B**, Warth P, Olsson L, Konstantinidis P. 2017. The development of the cucullaris muscle and the branchial musculature in the Longnose Gar, (*Lepisosteus osseus*, Lepisosteiformes, Actinopterygii) and its implications for the evolution and development of the head/trunk interface in vertebrates. *Evolution & Development* **19**:263–276. DOI: https://doi.org/10.1111/ede.12239, PMID: 29027738

**Noda M**, Miyake T, Okabe M. 2017. Development of cranial muscles in the actinopterygian fish Senegal bichir, *Polypterus senegalus* Cuvier, 1829. *Journal of Morphology* **278**:450–463. DOI: https://doi.org/10.1002/jmor.20636, PMID: 28182295

**Noden DM**. 1983. The embryonic origins of avian cephalic and cervical muscles and associated connective tissues. *American Journal of Anatomy* **168**:257–276. DOI: https://doi.org/10.1002/aja.1001680302, PMID: 6650439

**Noden DM**. 1988. Interactions and fates of avian craniofacial mesenchyme. *Development* **103**:121–140. PMID: 3074905

**Noden DM**, Francis-West P. 2006. The differentiation and morphogenesis of craniofacial muscles. *Developmental Dynamics* **235**:1194–1218. DOI: https://doi.org/10.1002/dvdy.20697, PMID: 16502415

**Okano J**, Sakai Y, Shiota K. 2008. Retinoic acid down-regulates Tbx1 expression and induces abnormal differentiation of tongue muscles in fetal mice. *Developmental Dynamics* **237**:3059–3070. DOI: https://doi.org/10.1002/dvdy.21715, PMID: 18816858

**Olsson L**, Falck P, Lopez K, Cobb J, Hanken J. 2001. Cranial neural crest cells contribute to connective tissue in cranial muscles in the anuran amphibian, Bombina orientalis. *Developmental Biology* **237**:354–367. DOI: https://doi.org/10.1006/dbio.2001.0377, PMID: 11543620

**Papangeli I**, Scambler P. 2013. The 22q11 deletion: DiGeorge and velocardiofacial syndromes and the role of *TBX1*. *Wiley Interdisciplinary Reviews: Developmental Biology* **2**:393–403. DOI: https://doi.org/10.1002/wdev.75, PMID: 23799583

**Parameswaran M**, Tam PP. 1995. Regionalisation of cell fate and morphogenetic movement of the mesoderm during mouse gastrulation. *Developmental Genetics* **17**:16–28. DOI: https://doi.org/10.1002/dvg.1020170104, PMID: 7554492

**Piatt J**. 1938. Morphogenesis of the cranial muscles of *Amblystoma punctatum*. *Journal of Morphology* **63**:531–587. DOI: https://doi.org/10.1002/jmor.1050630306

**Piekarski N**, Gross JB, Hanken J. 2014. Evolutionary innovation and conservation in the embryonic derivation of the vertebrate skull. *Nature Communications* **5**:5661. DOI: https://doi.org/10.1038/ncomms6661, PMID: 25434971

**Piekarski N**, Olsson L. 2007. Muscular derivatives of the cranialmost somites revealed by long-term fate mapping in the Mexican axolotl (Ambystoma mexicanum). *Evolution & Development* **9**:566–578. DOI: https://doi.org/10.1111/j.1525-142X.2007.00197.x, PMID: 17976053

**Ponomartsev S**, Valasek P, Patel K, Malashichev Y. 2017. Neural crest contribution to the avian shoulder girdle and implications to girdle evolution in vertebrates. *Biological Communications* **62**:26–37. DOI: https://doi.org/10.21638/11701/spbu03.2017.104

**Prunotto C**, Crepaldi T, Forni PE, Ieraci A, Kelly RG, Tajbakhsh S, Buckingham M, Ponzetto C. 2004. Analysis of Mlc-lacZ Met mutants highlights the essential function of Met for migratory precursors of hypaxial muscles and reveals a role for Met in the development of hyoid arch-derived facial muscles. *Developmental Dynamics* **231**:582–591. DOI: https://doi.org/10.1002/dvdy.20177, PMID: 15376315

**Randolph ME**, Pavlath GK. 2015. A muscle stem cell for every muscle: variability of satellite cell biology among different muscle groups. *Frontiers in Aging Neuroscience* **7**:190. DOI: https://doi.org/10.3389/fnagi.2015.00190, PMID: 26500547

**Relaix F**, Rocancourt D, Mansouri A, Buckingham M. 2004. Divergent functions of murine Pax3 and Pax7 in limb muscle development. *Genes & Development* **18**:1088–1105. DOI: https://doi.org/10.1101/gad.301004, PMID: 15132998

**Relaix F**, Rocancourt D, Mansouri A, Buckingham M. 2005. A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. *Nature* **435**:948–953. DOI: https://doi.org/10.1038/nature03594, PMID: 15843801

**Saga Y**, Hata N, Kobayashi S, Magnuson T, Seldin MF, Taketo MM. 1996. MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. *Development* **122**:2769–2778. PMID: 8787751

**Saga Y**. 1998. Genetic rescue of segmentation defect in MesP2-deficient mice by MesP1 gene replacement. *Mechanisms of Development* **75**:53–66. DOI: https://doi.org/10.1016/S0925-4773(98)00077-X, PMID: 9739106

**Saga Y**, Miyagawa-Tomita S, Takagi A, Kitajima S, Miyazaki J, Inoue T. 1999. MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Development* **126**:3437–3447. PMID: 10393122

**Saga Y**, Kitajima S, Miyagawa-Tomita S. 2000. Mesp1 expression is the earliest sign of cardiovascular development. *Trends in Cardiovascular Medicine* **10**:345–352. DOI: https://doi.org/10.1016/S1050-1738(01)00069-X, PMID: 11369261

**Sambasivan R**, Gayraud-Morel B, Dumas G, Cimper C, Paisant S, Kelly RG, Kelly R, Tajbakhsh S. 2009. Distinct regulatory cascades govern extraocular and pharyngeal arch muscle progenitor cell fates. *Developmental Cell* **16**:810–821. DOI: https://doi.org/10.1016/j.devcel.2009.05.008, PMID: 19531352

**Sambasivan R**, Comai G, Le Roux I, Gomès D, Konge J, Dumas G, Cimper C, Tajbakhsh S. 2013. Embryonic founders of adult muscle stem cells are primed by the determination gene Mrf4. *Developmental Biology* **381**:241–255. DOI: https://doi.org/10.1016/j.ydbio.2013.04.018, PMID: 23623977

**Satou Y**, Imai KS, Satoh N. 2004. The ascidian Mesp gene specifies heart precursor cells. *Development* **131**:2533–2541. DOI: https://doi.org/10.1242/dev.01145, PMID: 15115756

**Sawada A**, Fritz A, Jiang YJ, Yamamoto A, Yamasu K, Kuroiwa A, Saga Y, Takeda H. 2000. Zebrafish Mesp family genes, mesp-a and mesp-b are segmentally expressed in the presomitic mesoderm, and Mesp-b confers the anterior identity to the developing somites. *Development* **127**:1691–1702. PMID: 10725245

**Sefton EM**, Bhullar BA, Mohaddes Z, Hanken J. 2016. Evolution of the head-trunk interface in tetrapod vertebrates. *eLife* **5**:e09972. DOI: https://doi.org/10.7554/eLife.09972, PMID: 27090084

**Soriano P**. 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genetics* **21**:70–71. DOI: https://doi.org/10.1038/5007, PMID: 9916792

**Srinivas S**, Watanabe T, Lin CS, William CM, Tanabe Y, Jessell TM, Costantini F. 2001. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Developmental Biology* **1**:4. PMID: 11299042

**Tabler JM**, Rigney MM, Berman GJ, Gopalakrishnan S, Heude E, Al-Lami HA, Yannakoudakis BZ, Fitch RD, Carter C, Vokes S, Liu KJ, Tajbakhsh S, Egnor SR, Wallingford JB. 2017. Cilia-mediated Hedgehog signaling controls

form and function in the mammalian larynx. *eLife* **6**:e19153. DOI: https://doi.org/10.7554/eLife.19153, PMID: 28177282

**Tada MN**, Kuratani S. 2015. Evolutionary and developmental understanding of the spinal accessory nerve. *Zoological Letters* **1**:4. DOI: https://doi.org/10.1186/s40851-014-0006-8, PMID: 26605049

**Tajbakhsh S**, Rocancourt D, Buckingham M. 1996. Muscle progenitor cells failing to respond to positional cues adopt non-myogenic fates in myf-5 null mice. *Nature* **384**:266–270. DOI: https://doi.org/10.1038/384266a0, PMID: 8918877

**Tajbakhsh S**, Rocancourt D, Cossu G, Buckingham M. 1997. Redefining the genetic hierarchies controlling skeletal myogenesis: Pax-3 and Myf-5 act upstream of MyoD. *Cell* **89**:127–138. DOI: https://doi.org/10.1016/S0092-8674(00)80189-0, PMID: 9094721

**Tesařová M**, Zikmund T, Kaucká M, Adameyko I, Jaroš J, Paloušek D, Škaroupka D, Kaiser J. 2016. Use of micro computed-tomography and 3D printing for reverse engineering of mouse embryo nasal capsule. *Journal of Instrumentation* **11**:C03006. DOI: https://doi.org/10.1088/1748-0221/11/03/C03006

**Theis S**, Patel K, Valasek P, Otto A, Pu Q, Harel I, Tzahor E, Tajbakhsh S, Christ B, Huang R. 2010. The occipital lateral plate mesoderm is a novel source for vertebrate neck musculature. *Development* **137**:2961–2971. DOI: https://doi.org/10.1242/dev.049726, PMID: 20699298

**Tremblay P**, Dietrich S, Mericskay M, Schubert FR, Li Z, Paulin D. 1998. A crucial role for Pax3 in the development of the hypaxial musculature and the long-range migration of muscle precursors. *Developmental Biology* **203**:49–61. DOI: https://doi.org/10.1006/dbio.1998.9041, PMID: 9806772

**Valasek P**, Theis S, Krejci E, Grim M, Maina F, Shwartz Y, Otto A, Huang R, Patel K. 2010. Somitic origin of the medial border of the mammalian scapula and its homology to the avian scapula blade. *Journal of Anatomy* **216**:482–488. DOI: https://doi.org/10.1111/j.1469-7580.2009.01200.x, PMID: 20136669

**Verzi MP**, McCulley DJ, De Val S, Dodou E, Black BL. 2005. The right ventricle, outflow tract, and ventricular septum comprise a restricted expression domain within the secondary/anterior heart field. *Developmental Biology* **287**:134–145. DOI: https://doi.org/10.1016/j.ydbio.2005.08.041, PMID: 16188249

**Yabe T**, Hoshijima K, Yamamoto T, Takada S. 2016. Quadruple zebrafish mutant reveals different roles of Mesp genes in somite segmentation between mouse and zebrafish. *Development* **143**:2842–2852. DOI: https://doi.org/10.1242/dev.133173, PMID: 27385009

**Yang L**, Cai CL, Lin L, Qyang Y, Chung C, Monteiro RM, Mummery CL, Fishman GI, Cogen A, Evans S. 2006. Isl1Cre reveals a common Bmp pathway in heart and limb development. *Development* **133**:1575–1585. DOI: https://doi.org/10.1242/dev.02322, PMID: 16556916

**Yokomizo T**, Yamada-Inagawa T, Yzaguirre AD, Chen MJ, Speck NA, Dzierzak E. 2012. Whole-mount three-dimensional imaging of internally localized immunostained cells within mouse embryos. *Nature Protocols* **7**:421–431. DOI: https://doi.org/10.1038/nprot.2011.441, PMID: 22322215

**Zhang Z**, Huynh T, Baldini A. 2006. Mesodermal expression of Tbx1 is necessary and sufficient for pharyngeal arch and cardiac outflow tract development. *Development* **133**:3587–3595. DOI: https://doi.org/10.1242/dev.02539, PMID: 16914493

**Ziermann JM**, Freitas R, Diogo R. 2017. Muscle development in the shark *Scyliorhinus canicula*: implications for the evolution of the gnathostome head and paired appendage musculature. *Frontiers in Zoology* **14**:31. DOI: https://doi.org/10.1186/s12983-017-0216-y, PMID: 28649268

**Ziermann JM**, Clement AM, Ericsson R, Olsson L. 2018a. Cephalic muscle development in the Australian lungfish, *Neoceratodus forsteri*. *Journal of Morphology* **279**:494–516. DOI: https://doi.org/10.1002/jmor.20784, PMID: 29214665

**Ziermann JM**, Diogo R, Noden DM. 2018b. Neural crest and the patterning of vertebrate craniofacial muscles. *Genesis* **56**:e23097. DOI: https://doi.org/10.1002/dvg.23097, PMID: 29659153

**Zoupa M**, Seppala M, Mitsiadis T, Cobourne MT. 2006. Tbx1 is expressed at multiple sites of epithelial-mesenchymal interaction during early development of the facial complex. *The International Journal of Developmental Biology* **50**:504–510. DOI: https://doi.org/10.1387/ijdb.052116mz, PMID: 16586352

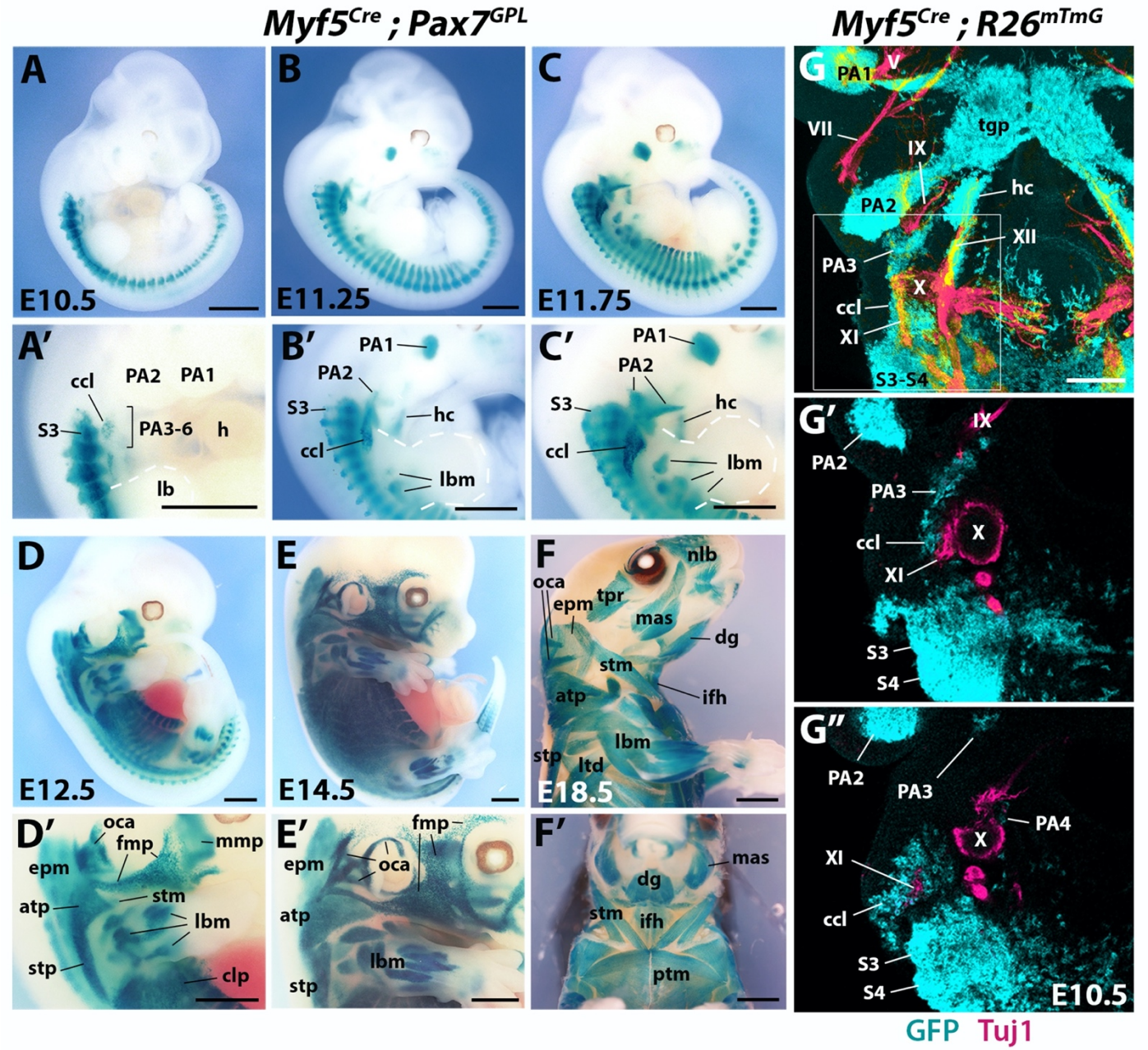



**Figure 1-figure supplement 1**

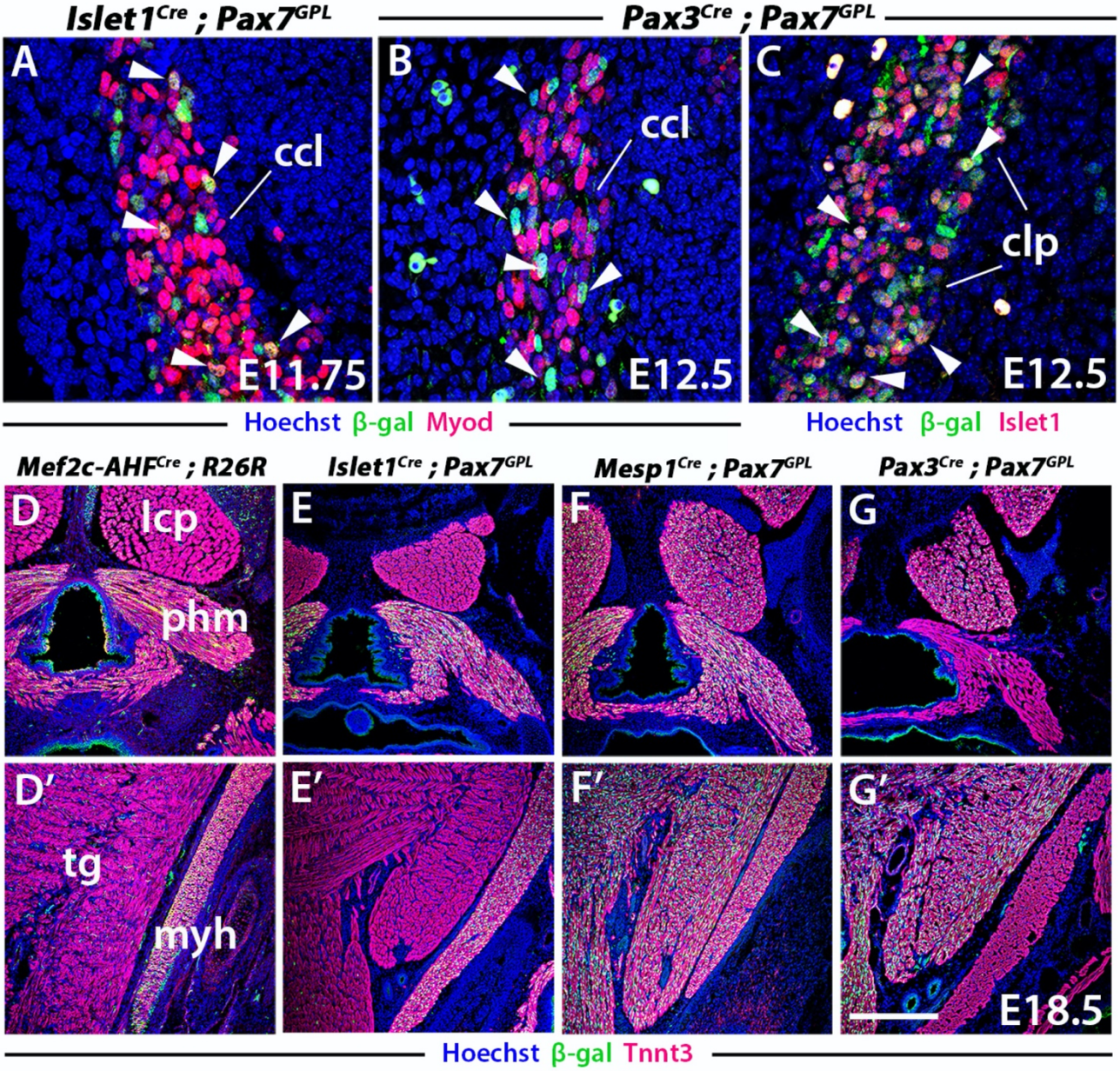



**Figure 1-figure supplement 2**

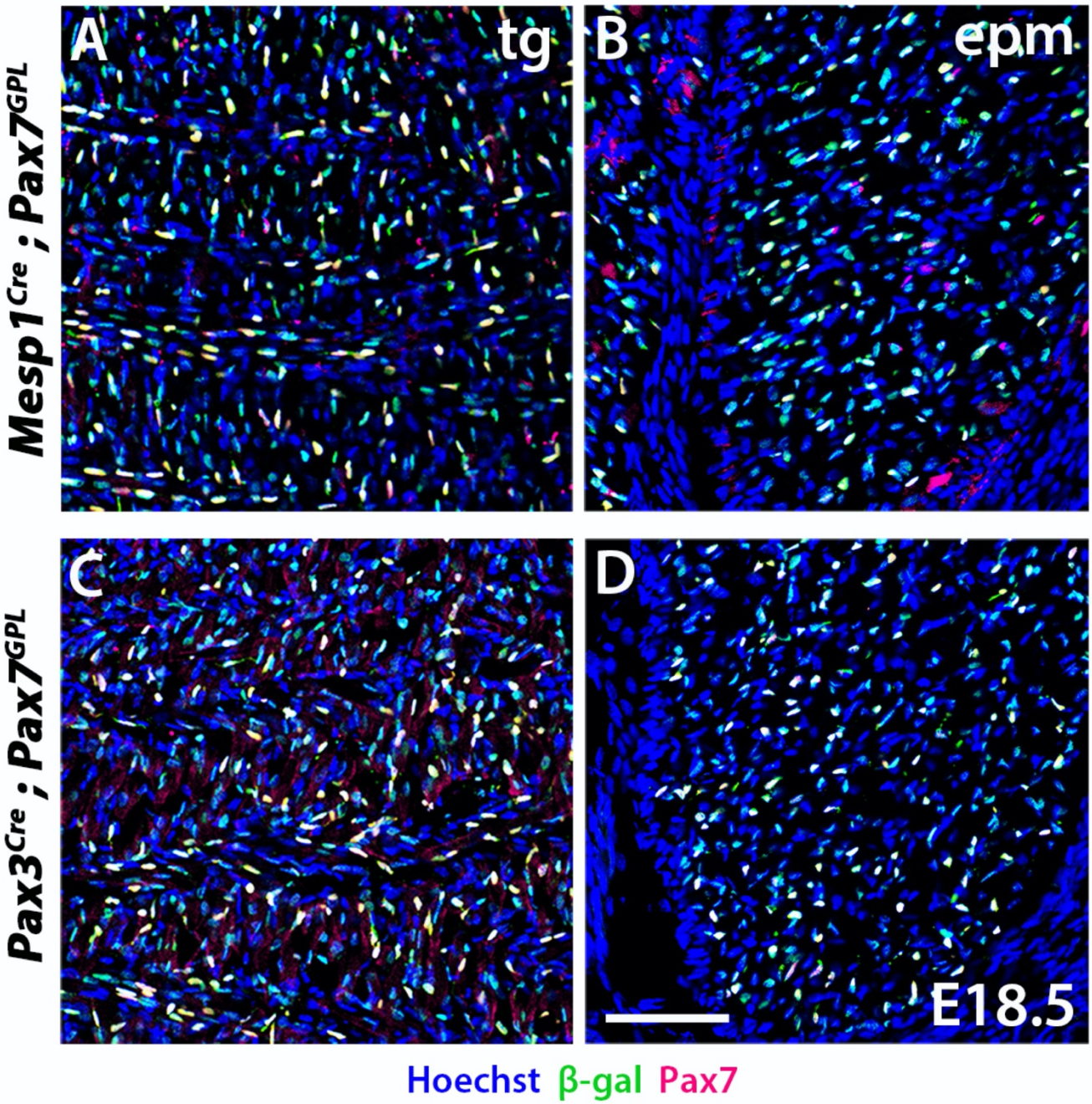



**Figure 1-figure supplement 3**

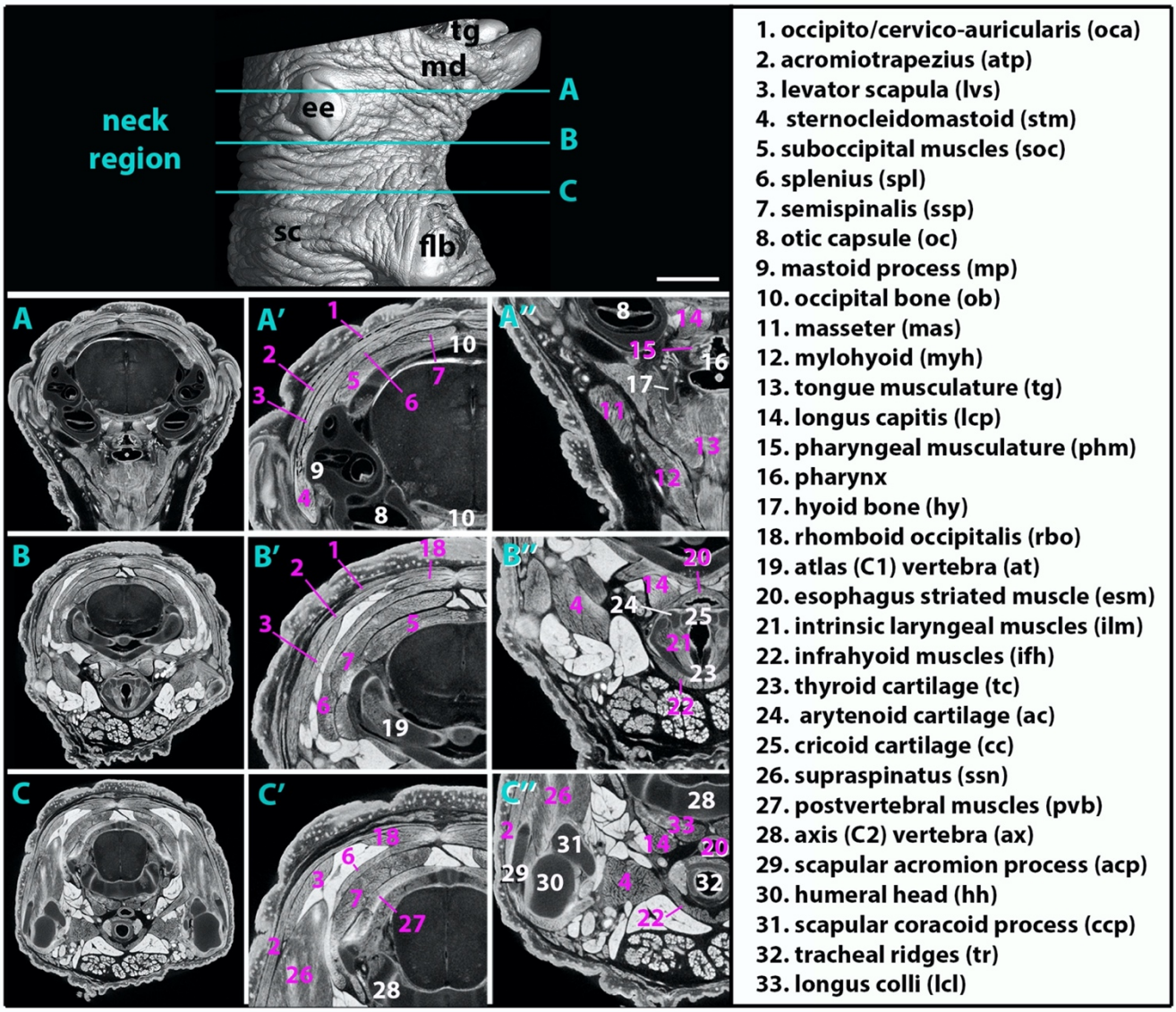



**Figure 2-figure supplement 1**

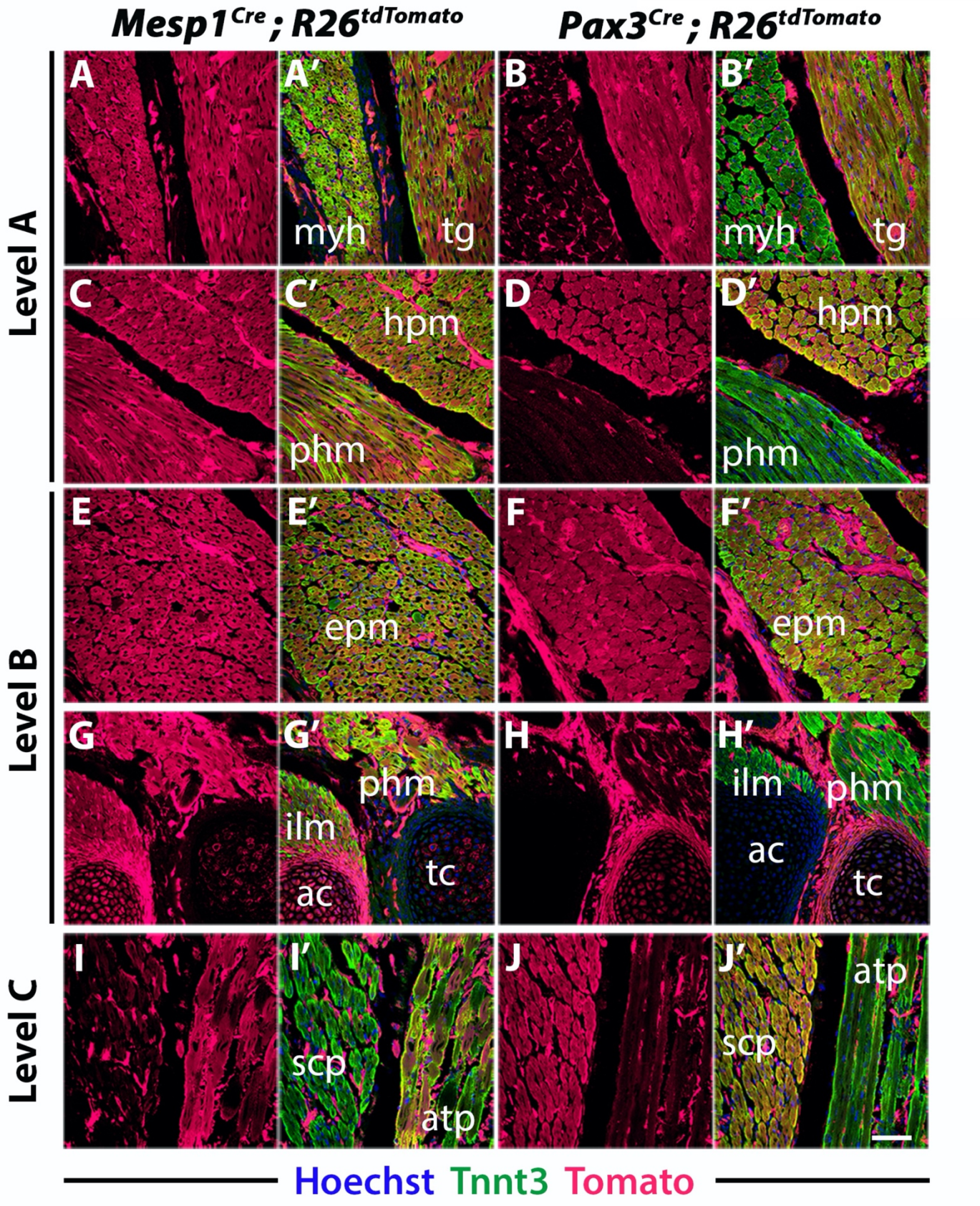



**Figure 2-figure supplement 2**

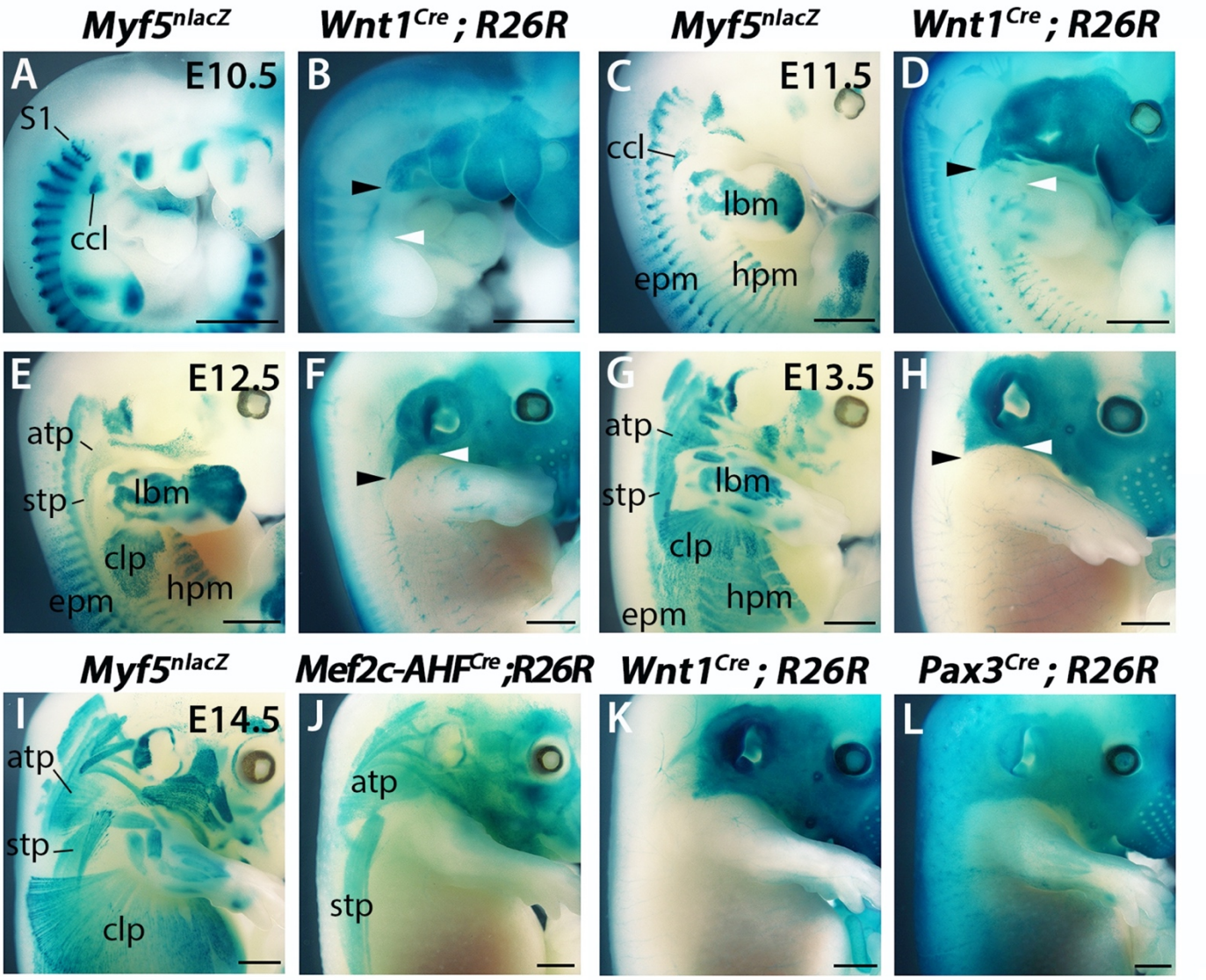



Figure 3-figure supplement 1

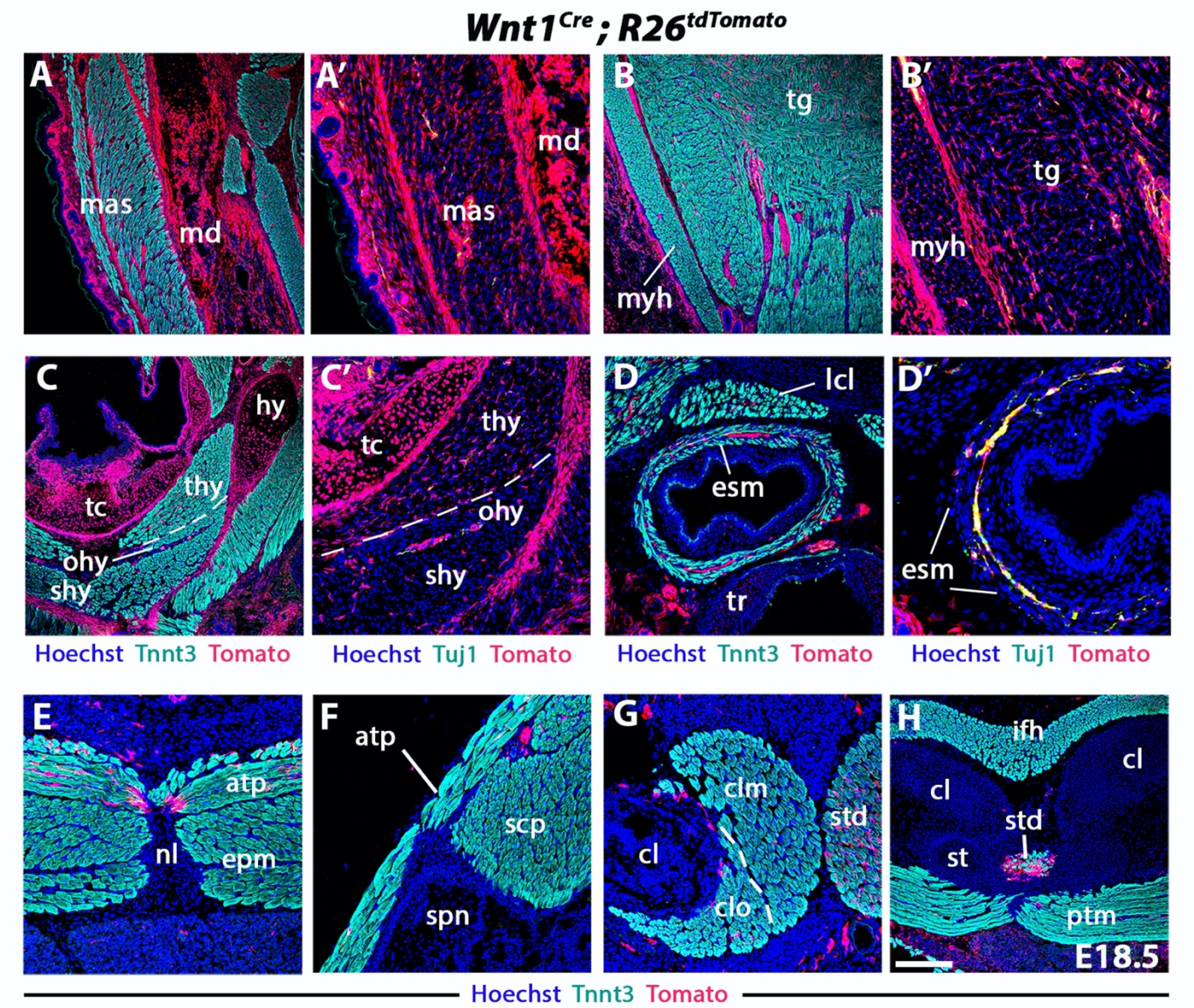



**Figure 3-figure supplement 2**

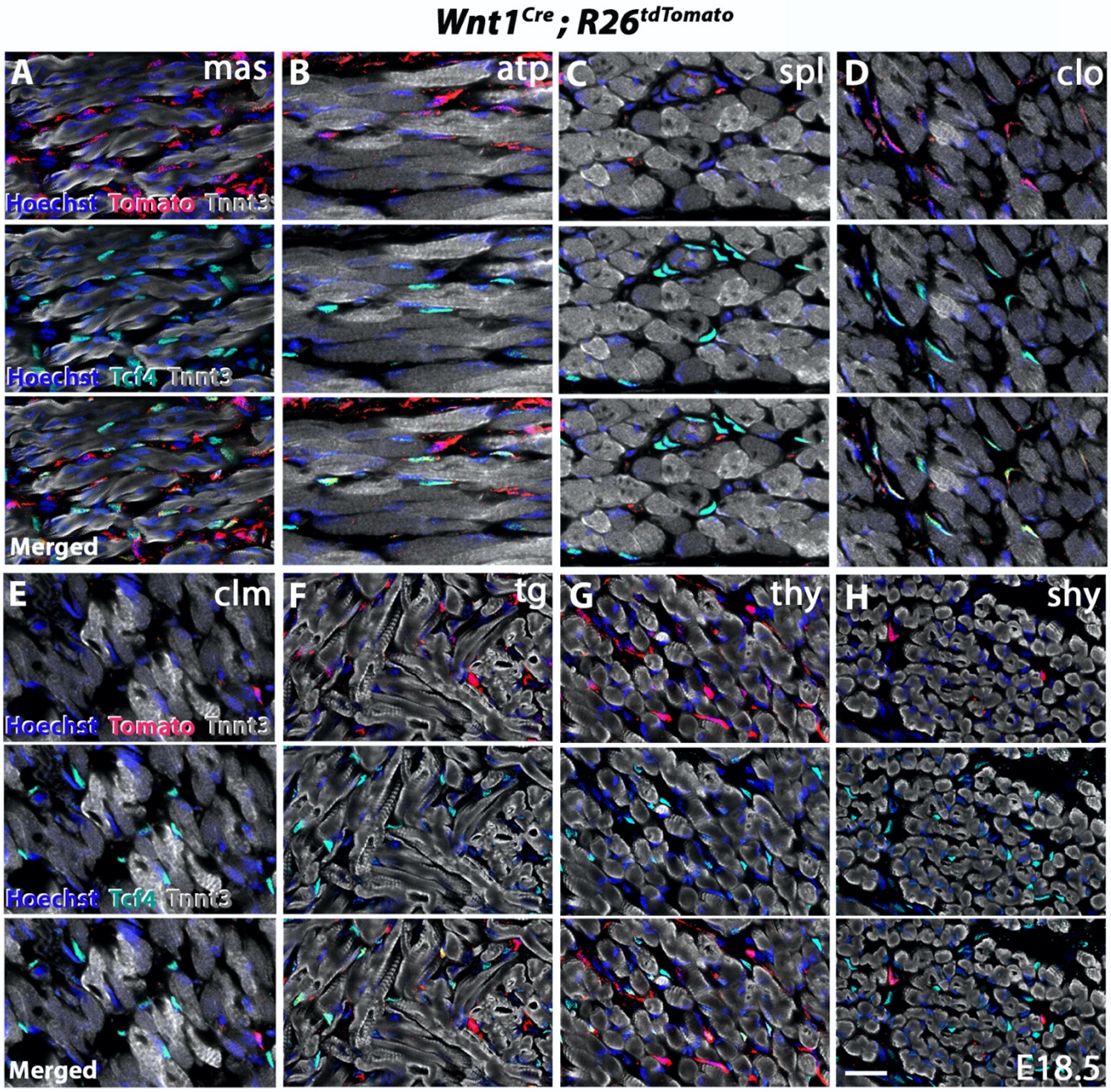



**Figure 3-figure supplement 3**

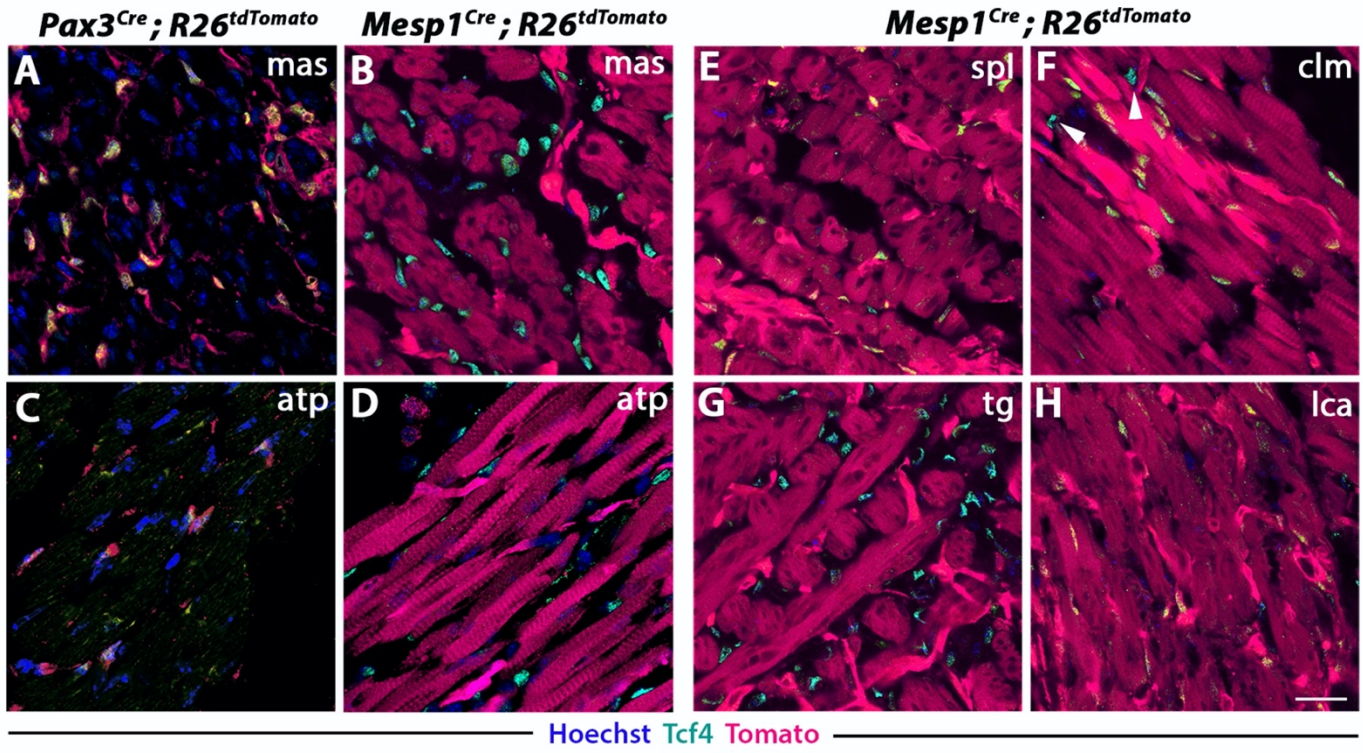



**Figure 3-figure supplement 4**

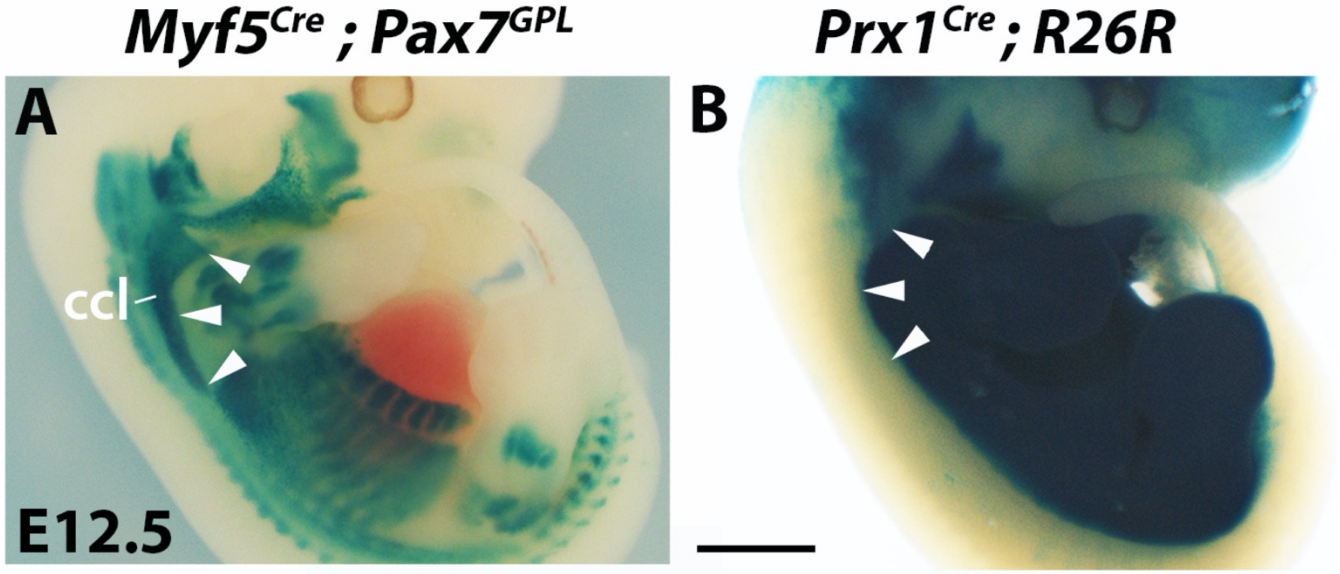



**Figure 4-figure supplement 1**

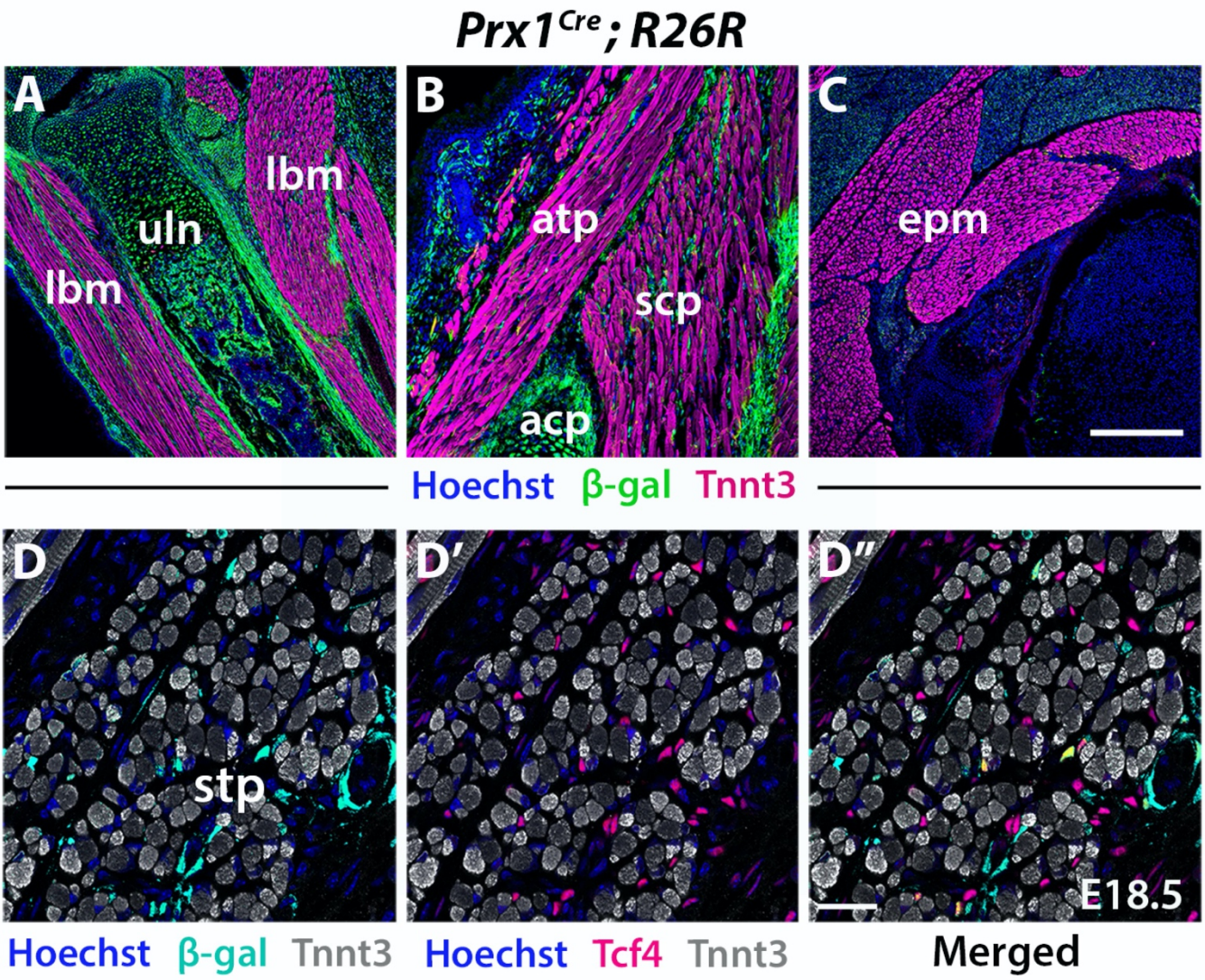



Figure 4-figure supplement 2

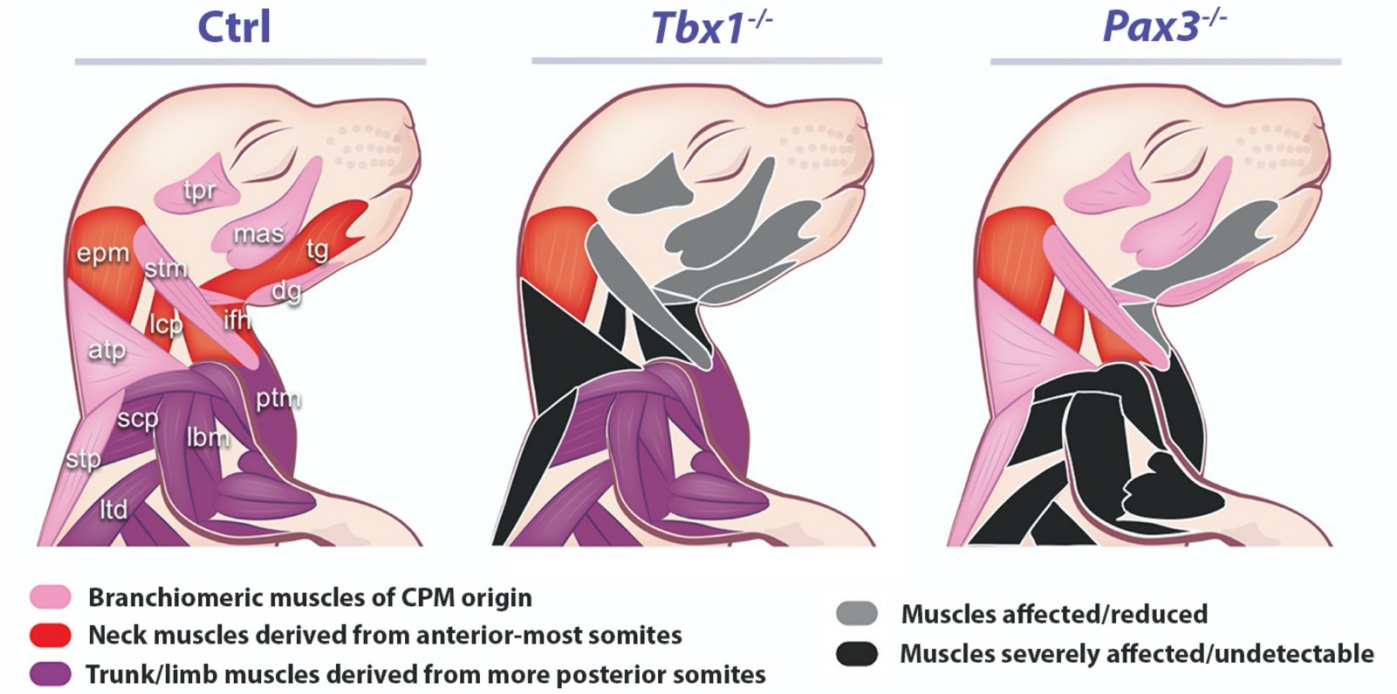



**Figure 7-figure supplement 1**

1

# REFERENCES

**Abou-Khalil, R., Grand, F. L., Pallafacchina, G., Valable, S., Authier, F.-J., Rudnicki, M. A., Gherardi, R. K., Germain, S., Chretien, F., Sotiropoulos, A., et al.** (2009). Autocrine and Paracrine Angiopoietin 1/Tie-2 Signaling Promotes Muscle Satellite Cell Self-Renewal. *Cell Stem Cell* **5**, 298–309.

**Acharyya, S., Sharma, S. M., Cheng, A. S., Ladner, K. J., He, W., Kline, W., Wang, H., Ostrowski, M. C., Huang, T. H. and Guttridge, D. C.** (2010). TNF Inhibits Notch-1 in Skeletal Muscle Cells by Ezh2 and DNA Methylation Mediated Repression: Implications in Duchenne Muscular Dystrophy. *Plos One* **5**, e12479.

**Adachi, N., Bilio, M., Baldini, A. and Kelly, R. G.** (2020). Cardiopharyngeal mesoderm origins of musculoskeletal and connective tissues in the mammalian pharynx. *Development* **147**, dev185256.

**Ali, S. and Garcia, J. M.** (2014). Sarcopenia, Cachexia and Aging: Diagnosis, Mechanisms and Therapeutic Options - A Mini-Review. *Gerontology* **60**, 294–305.

**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.

**Amano, O., Yamane, A., Shimada, M., Koshimizu, U., Nakamura, T. and Iseki, S.** (2002). Hepatocyte growth factor is essential for migration of myogenic cells and promotes their proliferation during the early periods of tongue morphogenesis in mouse embryos. *Developmental Dynamics* **223**, 169–179.

**Arnold, P.** (2020). Evolution of the Mammalian Neck from Developmental, Morpho-Functional, and Paleontological Perspectives. *J Mamm Evol* 1–11.

**Arnold, J. S., Werling, U., Braunstein, E. M., Liao, J., Nowotschin, S., Edelmann, W., Hebert, J. M. and Morrow, B. E.** (2006). Inactivation of Tbx1 in the pharyngeal endoderm results in 22q11DS malformations. *Development* **133**, 977–987.

**Asakura, A., Rudnicki, M. A. and Komaki, M.** (2001). Muscle satellite cells are multipotential stem cells that exhibit myogenic, osteogenic, and adipogenic differentiation. *Differentiation* **68**, 245–253.

**Atchley, W. R., Fitch, W. M. and Bronner-Fraser, M.** (1994). Molecular evolution of the MyoD family of transcription factors. *Proc National Acad Sci* **91**, 11522–11526.

**Atit, R., Sgaier, S. K., Mohamed, O. A., Taketo, M. M., Dufort, D., Joyner, A. L., Niswander, L. and Conlon, R. A.** (2006). β-catenin activation is necessary and sufficient to specify the dorsal dermal fate in the mouse. *Dev Biol* **296**, 164–176.

**Auerbach, R.** (1954). Analysis of the developmental effects of a lethal mutation in the house mouse. *J Exp Zool* **127**, 305–329.

**Babiuk, R. P., Zhang, W., Clugston, R., Allan, D. W. and Greer, J. J.** (2002). Embryological origins and development of the rat diaphragm. *J Comp Neurology* **455**, 477–487.

**Baghdadi, M. B., Firmino, J., Soni, K., Evano, B., Girolamo, D. D., Mourikis, P., Castel, D. and Tajbakhsh, S.** (2018a). Notch-Induced miR-708 Antagonizes Satellite Cell Migration and Maintains Quiescence. *Cell Stem Cell* **23**, 859-868.e5.

**Baghdadi, M. B., Castel, D., Machado, L., Fukada, S.-I. I., Birk, D. E., Relaix, F., Tajbakhsh, S. and Mourikis, P.** (2018b). Reciprocal signalling by Notch-Collagen V-CALCR retains muscle stem cells in their niche. *Nature* **557**, 714–718.

**Baguma-Nibasheka, M., Fracassi, A., Costain, W. J., Moreno, S. and Kablar, B.** (2019). Striated-for-smooth muscle replacement in the developing mouse esophagus. *Histol Histopathol* 18087.

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

**Bendall, A. J., Ding, J., Hu, G., Shen, M. M. and Abate-Shen, C.** (1999). Msx1 antagonizes the myogenic activity of Pax3 in migrating limb muscle precursors. *Dev Camb Engl* **126**, 4965–76.

**Bentzinger, C. F., Wang, Y. X., Dumont, N. A. and Rudnicki, M. A.** (2013). Cellular dynamics in the muscle satellite cell niche. *Embo Rep* **14**, 1062–1072.

**Ben-Yair, R. and Kalcheim, C.** (2008). Notch and bone morphogenetic protein differentially act on dermomyotome cells to generate endothelium, smooth, and striated muscle. *J Cell Biol* **180**, 607–618.

**Bi, P., Yue, F., Sato, Y., Wirbisky, S., Liu, W., Shan, T., Wen, Y., Zhou, D., Freeman, J. and Kuang, S.** (2016). Stage-specific effects of Notch activation during skeletal myogenesis. *Elife* **5**, e17355.

**Bi, P., Ramirez-Martinez, A., Li, H., Cannavino, J., McAnally, J. R., Shelton, J. M., Sánchez-Ortiz, E., Bassel-Duby, R. and Olson, E. N.** (2017). Control of muscle formation by the fusogenic micropeptide myomixer. *Sci New York N Y* **356**, 323–327.

**Biferali, B., Proietti, D., Mozzetta, C. and Madaro, L.** (2019). Fibro–Adipogenic Progenitors Cross-Talk in Skeletal Muscle: The Social Network. *Front Physiol* **10**, 1074.

**Birchmeier, C. and Brohmann, H.** (2000). Genes that control the development of migrating muscle precursor cells. *Curr Opin Cell Biol* **12**, 725–730.

**Biressi, S., Tagliafico, E., Lamorte, G., Monteverde, S., Tenedini, E., Roncaglia, E., Ferrari, S., Ferrari, S., Angelis, M. G. C.-D., Tajbakhsh, S., et al.** (2007). Intrinsic phenotypic diversity of embryonic and fetal myoblasts is revealed by genome-wide gene expression analysis on purified cells. *Dev Biol* **304**, 633–651.

**Black, B. L., Molkentin, J. D. and Olson, E. N.** (1998). Multiple Roles for the MyoD Basic Region in Transmission of Transcriptional Activation Signals and Interaction with MEF2. *Mol Cell Biol* **18**, 69–77.

**Bladt, F., Riethmacher, D., Isenmann, S., Aguzzi, A. and Birchmeier, C.** (1995). Essential role for the c-met receptor in the migration of myogenic precursor cells into the limb bud. *Nature* **376**, 768–771.

**Blake, J. A. and Ziman, M. R.** (2014). Pax genes: regulators of lineage specification and progenitor cell maintenance. *Development* **141**, 737–751.

**Bober, E., Franz, T., Arnold, H. H., Gruss, P. and Tremblay, P.** (1994). Pax-3 is required for the development of limb muscles: a possible role for the migration of dermomyotomal muscle progenitor cells. *Dev Camb Engl* **120**, 603–12.

**Bonfanti, C., Rossi, G., Tedesco, F. S., Giannotta, M., Benedetti, S., Tonlorenzi, R., Antonini, S., Marazzi, G., Dejana, E., Sassoon, D., et al.** (2015). PW1/Peg3 expression regulates key properties that determine mesoangioblast stem cell competence. *Nat Commun* **6**, 6364.

**Bothe, I. and Dietrich, S.** (2006). The molecular setup of the avian head mesoderm and its implication for craniofacial myogenesis. *Developmental Dynamics* **235**, 2845–2860.

**Brack, A. S., Conboy, M. J., Roy, S., Lee, M., Kuo, C. J., Keller, C. and Rando, T. A.** (2007). Increased Wnt Signaling During Aging Alters Muscle Stem Cell Fate and Increases Fibrosis. *Science* **317**, 807–810.

**Braun, T., Bober, E., Winter, B., Rosenthal, N. and Arnold, H. H.** (1990). Myf-6, a new member of the human gene family of myogenic determination factors: evidence for a gene cluster on chromosome 12. *Embo J* **9**, 821–31.

**Brent, A. E. and Tabin, C. J.** (2004). FGF acts directly on the somitic tendon progenitors through the Ets transcription factors Pea3 and Erm to regulate scleraxis expression. *Development* **131**, 3885–3896.

**Brent, A. E., Schweitzer, R. and Tabin, C. J.** (2003). A Somitic Compartment of Tendon Progenitors. *Cell* **113**, 235–248.

**Brent, A. E., Braun, T. and Tabin, C. J.** (2005). Genetic analysis of interactions between the somitic muscle, cartilage and tendon cell lineages during mouse development. *Development* **132**, 515–528.

**Brown, C. B., Engleka, K. A., Wenning, J., Lu, M. M. and Epstein, J. A.** (2005). Identification of a hypaxial somite enhancer element regulating Pax3 expression in migrating myoblasts and characterization of hypaxial muscle Cre transgenic mice. *Genesis* **41**, 202–209.

**Buckingham, M.** (2017). Gene regulatory networks and cell lineages that underlie the formation of skeletal muscle. *Proc National Acad Sci* **114**, 5830–5837.

**Buckingham, M. and Relaix, F.** (2007). The Role of Pax Genes in the Development of Tissues and Organs: Pax3 and Pax7 Regulate Muscle Progenitor Cell Functions. *Cell Dev Biology* **23**, 645–673.

**Buckingham, M. and Rigby, P. W. J.** (2014). Gene Regulatory Networks and Transcriptional Mechanisms that Control Myogenesis. *Dev Cell* **28**, 225–238.

**Burke, A. C.** (1999). 6 Hox Genes and the Global Patterning of the Somitic Mesoderm. *Curr Top Dev Biol* **47**, 155–181.

**Burke, A. C. and Nowicki, J. L.** (2003). A New View of Patterning Domains in the Vertebrate Mesoderm. *Dev Cell* **4**, 159–165.

**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. *Developmental Cell* **5**, 877–889.

**Carlson, B. M.** (2014). *Human Embryology and Developmental Biology 5th Edition*. (ed. Elsevier).

**Carvajal, J. J., Cox, D., Summerbell, D. and Rigby, P. W.** (2001). A BAC transgenic analysis of the Mrf4/Myf5 locus reveals interdigitated elements that control activation and maintenance of gene expression during muscle development. *Development (Cambridge, England)* **128**, 1857–68.

**Carvajal, J. J., Keith, A. and Rigby, P. W. J.** (2008). Global transcriptional regulation of the locus encoding the skeletal muscle determination genes Mrf4 and Myf5. *Gene Dev* **22**, 265–276.

**Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A. and Bozzoni, I.** (2011). A Long Noncoding RNA Controls Muscle Differentiation by Functioning as a Competing Endogenous RNA. *Cell* **147**, 358–369.

**Chai, Y. and Maxson, R. E.** (2006). Recent advances in craniofacial morphogenesis. *Developmental Dynamics* **235**, 2353–2375.

**Chang, T. H.-T., Primig, M., Hadchouel, J., Tajbakhsh, S., Rocancourt, D., Fernandez, A., Kappler, R., Scherthan, H. and Buckingham, M.** (2004). An enhancer directs differential expression of the linked Mrf4 and Myf5 myogenic regulatory genes in the mouse. *Dev Biol* **269**, 595–608.

**Cheng, G., Merriam, A. P., Gong, B., Leahy, P., Khanna, S. and Porter, J. D.** (2004a). Conserved and muscle-group-specific gene expression patterns shape postnatal development of the novel extraocular muscle phenotype. *Physiol Genomics* **18**, 184–195.

**Cheng, L., Alvares, L. E., Ahmed, M. U., El-Hanfy, A. S. and Dietrich, S.** (2004b). The epaxial–hypaxial subdivision of the avian somite. *Dev Biol* **274**, 348–369.

**Christ, B., Jacob, H. J. and Jacob, M.** (1977). Experimental analysis of the origin of the wing musculature in avian embryos. *Anat Embryol* **150**, 171–186.

**Christ, B., Huang, R. and Scaal, M.** (2007). Amniote somite derivatives. *Dev Dynam* **236**, 2382–2396.

**Cobourne, M. T., Iseki, S., Birjandi, A. A., Al-Lami, H. A., Thauvin-Robinet, C., Xavier, G. M. and Liu, K. J.** (2018). How to make a tongue: Cellular and molecular regulation of muscle and connective tissue formation during mammalian tongue development. *Semin Cell Dev Biol* **91**, 45–54.

**Colasanto, M. P., Eyal, S., Mohassel, P., Bamshad, M., Bonnemann, C. G., Zelzer, E., Moon, A. M. and Kardon, G.** (2016). Development of a subset of forelimb muscles and their attachment sites requires the ulnar-mammary syndrome geneTbx3. *Dis Model Mech* **9**, 1257–1269.

**Comai, G. and Tajbakhsh, S.** (2014). Current Topics in Developmental Biology. *Curr Top Dev Biol* **110**, 1–73.

**Comai, G., Heude, E., Mella, S., Paisant, S., Pala, F., Gallardo, M., Langa, F., Kardon, G., Gopalakrishnan, S. and Tajbakhsh, S.** (2019). A distinct cardiopharyngeal mesoderm genetic hierarchy establishes antero-posterior patterning of esophagus striated muscle. *Elife* **8**, e47460.

**Comai, G., Tesarova, M., Dupé, V., Rhinn, M., Garcia, P. V., Silva, F. da, Feret, B., Exelby, K., Dollé, P., Carlsson, L., et al.** (2020). Local retinoic acid directs emergence of the extraocular muscle functional unit. *Biorxiv* 2020.01.07.897694.

**Conerly, M. L., Yao, Z., Zhong, J. W., Groudine, M. and Tapscott, S. J.** (2016). Distinct Activities of Myf5 and MyoD Indicate Separate Roles in Skeletal Muscle Lineage Specification and Differentiation. *Dev Cell* **36**, 375–85.

**Cossu, G., Angelis, L. D., Borello, U., Berarducci, B., Buffa, V., Sonnino, C., Coletta, M., Vivarelli, E., Bouche, M., Lattanzi, L., et al.** (2000). Determination, diversification and multipotency of mammalian myogenic cells. *Int J Dev Biology* **44**, 699–706.

**Couly, G., Grapin-Botton, A., Coltey, P., Ruhin, B. and Douarin, N. M. L.** (1998). Determination of the identity of the derivatives of the cephalic neural crest: incompatibility between Hox gene expression and lower jaw development. *Dev Camb Engl* **125**, 3445–59.

**Crist, C. G., Montarras, D. and Buckingham, M.** (2012). Muscle Satellite Cells Are Primed for Myogenesis but Maintain Quiescence with Sequestration of Myf5 mRNA Targeted by microRNA-31 in mRNP Granules. *Cell Stem Cell* **11**, 118–126.

**Dastjerdi, A., Robson, L., Walker, R., Hadley, J., Zhang, Z., Rodriguez-Niedenführ, M., Ataliotis, P., Baldini, A., Scambler, P. and Francis-West, P.** (2007). Tbx1 regulation of myogenic differentiation in the limb and cranial mesoderm. *Dev Dynam* **236**, 353–363.

**Daubas, P., Tajbakhsh, S., Hadchouel, J., Primig, M. and Buckingham, M.** (2000). Myf5 is a novel early axonal marker in the mouse brain and is subjected to post-transcriptional regulation in neurons. *Dev Camb Engl* **127**, 319–31.

**Davies, T., Kim, H. X., Spica, N. R., Lesea-Pringle, B. J., Dumont, J., Shirasu-Hiza, M. and Canman, J. C.** (2018). Cell-intrinsic and -extrinsic mechanisms promote cell-type-specific cytokinetic diversity. *Elife* **7**, e36204.

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

**Day, K., Shefer, G., Richardson, J. B., Enikolopov, G. and Yablonka-Reuveni, Z.** (2007). Nestin-GFP reporter expression defines the quiescent state of skeletal muscle satellite cells. *Dev Biol* **304**, 246–259.

**Debbache, J., Parfejevs, V. and Sommer, L.** (2018). Cre-driver lines used for genetic fate mapping of neural crest cells in the mouse: An overview. *Genesis* **56**, e23105.

**Dellavalle, A., Sampaolesi, M., Tonlorenzi, R., Tagliafico, E., Sacchetti, B., Perani, L., Innocenzi, A., Galvez, B. G., Messina, G., Morosetti, R., et al.** (2007). Pericytes of human skeletal muscle are myogenic precursors distinct from satellite cells. *Nat Cell Biol* **9**, 255–267.

**Dennis, J.** (2019). Revisiting the Pharyngeal Apparatus: Arches, Pouches, & Grooves—Oh My! *Faseb J* **33**, 12.2-12.2.

**Depew, M. J., Lufkin, T. and Rubenstein, J. L. R.** (2002). Specification of Jaw Subdivisions by Dlx Genes. *Science* **298**, 381–385.

**Depew, M. J., Simpson, C. A., Morasso, M. and Rubenstein, J. L. R.** (2005). Reassessing the Dlx code: the genetic regulation of branchial arch skeletal pattern and development: Dlx regulation of branchial arch development, M. J. Depew et al. *J Anat* **207**, 501–561.

**Deries, M. and Thorsteinsdóttir, S.** (2016). Axial and limb muscle development: dialogue with the neighbourhood. *Cell Mol Life Sci* **73**, 4415–4431.

**Deries, M., Schweitzer, R. and Duxson, M. J.** (2010). Developmental fate of the mammalian myotome. *Dev Dynam* **239**, 2898–2910.

**Deries, M., Gonçalves, A. B., Vaz, R., Martins, G. G., Rodrigues, G. and Thorsteinsdóttir, S.** (2011). Extracellular matrix remodeling accompanies axial muscle development and morphogenesis in the mouse. *Dev Dynam* **241**, 350–364.

**Dietrich, S., Schubert, F. R., Healy, C., Sharpe, P. T. and Lumsden, A.** (1998). Specification of the hypaxial musculature. *Dev Camb Engl* **125**, 2235–49.

**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 (Cambridge, England)* **126**, 1621–9.

**Diogo, R., Kelly, R. G., Christiaen, L., Levine, M., Ziermann, J. M., Molnar, J. L., Noden, D. M. and Tzahor, E.** (2015). A new heart for a new head in vertebrate cardiopharyngeal evolution. *Nature* **520**, 466–73.

**Dodou, E., Verzi, M. P., Anderson, J. P., Xu, S.-M. and Black, B. L.** (2004). Mef2c is a direct transcriptional target of ISL1 and GATA factors in the anterior heart field during mouse embryonic development. *Development* **131**, 3931–3942.

**Donghui, C., Shicai, C., Wei, W., Fei, L., Jianjun, J., Gang, C. and Hongliang, Z.** (2009). Functional modulation of satellite cells in long-term denervated human laryngeal muscle: SCs in Denervated Human Laryngeal Muscle. *Laryngoscope* **120**, 353–358.

**Donoghue, M. J., Morris-Valero, R., Johnson, Y. R., Merlie, J. P. and Sanes, J. R.** (1992). Mammalian muscle cells bear a cell-autonomous, heritable memory of their rostrocaudal position. *Cell* **69**, 67–77.

**Douarin, N. L. and Kalcheim, C.** (1999). The Neural Crest.

**Dumont, N. A., Wang, Y. X. and Rudnicki, M. A.** (2015). Intrinsic and extrinsic mechanisms regulating satellite cell function. *Development* **142**, 1572–1581.

**Durland, J. L., Sferlazzo, M., Logan, M. and Burke, A. C.** (2008). Visualizing the lateral somitic frontier in the Prx1Cre transgenic mouse. *J Anat* **212**, 590–602.

**Duxson, M. J., Usson, Y. and Harris, A. J.** (1989). The origin of secondary myotubes in mammalian skeletal muscles: ultrastructural studies. *Dev Camb Engl* **107**, 743–50.

**Eade, K. T., Fancher, H. A., Ridyard, M. S. and Allan, D. W.** (2012). Developmental Transcriptional Networks Are Required to Maintain Neuronal Subtype Identity in the Mature Nervous System. *Plos Genet* **8**, e1002501.

**Ehehalt, F., Wang, B., Christ, B., Patel, K. and Huang, R.** (2004). Intrinsic cartilage-forming potential of dermomyotomal cells requires ectodermal signals for the development of the scapula blade. *Anat Embryol* **208**, 431–437.

**Eichmann, A., Corbel, C., Nataf, V., Vaigot, P., Breant, C. and Douarin, N. M. L.** (1997). Ligand-dependent development of the endothelial and hemopoietic lineages from embryonic mesodermal cells expressing vascular endothelial growth factor receptor 2. *Proc National Acad Sci* **94**, 5141–5146.

**Eng, D., Ma, H.-Y., Xu, J., Shih, H.-P., Gross, M. K., Kioussi, C. and Kiouss, C.** (2012). Loss of Abdominal Muscle in Pitx2 Mutants Associated with Altered Axial Specification of Lateral Plate Mesoderm. *Plos One* **7**, e42228.

**Engleka, K. A., Gitler, A. D., Zhang, M., Zhou, D. D., High, F. A. and Epstein, J. A.** (2005). Insertion of Cre into the Pax3 locus creates a new allele of Splotch and identifies unexpected Pax3 derivatives. *Dev Biol* **280**, 396–406.

**Ericsson, R., Cerny, R., Falck, P. and Olsson, L.** (2004). Role of cranial neural crest cells in visceral arch muscle positioning and morphogenesis in the Mexican axolotl,Ambystoma mexicanum. *Dev Dynam* **231**, 237–247.

**Esner, M., Meilhac, S. M. M., Relaix, F., Nicolas, J.-F. F., Cossu, G. and Buckingham, M. E.** (2006). Smooth muscle of the dorsal aorta shares a common clonal origin with skeletal muscle of the myotome. *Development (Cambridge, England)* **133**, 737–49.

**Espinosa-Soto, C.** (2018). On the role of sparseness in the evolution of modularity in gene regulatory networks. *PLoS computational biology* **14**, e1006172.

**Evano, B. and Tajbakhsh, S.** (2018). Skeletal muscle stem cells in comfort and stress. *Npj Regen Medicine* **3**, 24.

**Evano, B., Gill, D., Hernando-Herraez, I., Comai, G., Stubbs, T. M., Commere, P.-H., Reik, W. and Tajbakhsh, S.** (2020). Transcriptome and epigenome diversity and plasticity of muscle stem cells following transplantation. *Biorxiv* 2020.05.20.107219.

**Evans, A. L. and Gage, P. J.** (2005). Expression of the homeobox gene Pitx2 in neural crest is required for optic stalk and ocular anterior segment development. *Hum Mol Genet* **14**, 3347–3359.

**Evans, D. J. and Noden, D. M.** (2006). Spatial relations between avian craniofacial neural crest and paraxial mesoderm cells. *Developmental dynamics* **235**, 1310–25.

**Fetcho, J. R.** (1987). A review of the organization and evolution of motoneurons innervating the axial musculature of vertebrates. *Brain Res Rev* **12**, 243–280.

**Fong, A. P. and Tapscott, S. J.** (2013). Skeletal muscle programming and re-programming. *Curr Opin Genet Dev* **23**, 568–573.

**Forcales, S. V., Albini, S., Giordani, L., Malecova, B., Cignolo, L., Chernov, A., Coutinho, P., Saccone, V., Consalvi, S., Williams, R., et al.** (2011). Signal-dependent incorporation of MyoD-BAF60c into Brg1-based SWI/SNF chromatin-remodelling complex: BAF60c-MyoD marks chromatin for SWI/SNF recruitment. *Embo J* **31**, 301–316.

**Formicola, L., Marazzi, G. and Sassoon, D. A.** (2014). The extraocular muscle stem cell niche is resistant to ageing and disease. *Front Aging Neurosci* **6**, 328.

**Francis, N. J. and Kingston, R. E.** (2001). Mechanisms of transcriptional memory. *Nat Rev Mol Cell Bio* **2**, 409–421.

**Franz, T.** (1989). Persistent truncus arteriosus in the Splotch mutant mouse. *Anat Embryol* **180**, 457–464.

**Frisdal, A. and Trainor, P. A.** (2014). Development and evolution of the pharyngeal apparatus. *Wiley Interdiscip Rev Dev Biology* **3**, 403–18.

**Gage, P. J., Suh, H. and Camper, S. A.** (1999a). The bicoid -related Pitx gene family in development. *Mamm Genome* **10**, 197–200.

**Gage, P. J., Suh, H. and Camper, S. A.** (1999b). Dosage requirement of Pitx2 for development of multiple organs. *Dev Camb Engl* **126**, 4643–51.

**Gans, C. and Northcutt, R. G.** (1983). Neural Crest and the Origin of Vertebrates: A New Head. *Science* **220**, 268–273.

**Giordani, J., Bajard, L., Demignon, J., Daubas, P., Buckingham, M. and Maire, P.** (2007). Six proteins regulate the activation of Myf5 expression in embryonic mouse limbs. *Proc National Acad Sci* **104**, 11310–11315.

**Giordani, L., He, G. J., Negroni, E., Sakai, H., Law, J. Y. C., Siu, M. M., Wan, R., Corneau, A., Tajbakhsh, S., Cheung, T. H., et al.** (2019). High-Dimensional Single-Cell Cartography Reveals Novel Skeletal Muscle-Resident Cell Populations. *Mol Cell* **74**, 609-621.e6.

**Goding, G. S., Al-Sharif, K. I. and McLoon, L. K.** (2005). Myonuclear Addition to Uninjured Laryngeal Myofibers in Adult Rabbits. *Ann Otology Rhinology Laryngology* **114**, 552–557.

**Gohring, W., Sasaki, T., Heldin, C.-H. and Timpl, R.** (1998). Mapping of the binding of platelet-derived growth factor to distinct domains of the basement membrane proteins BM-40 and perlecan and distinction from the BM-40 collagen-binding epitope. *Eur J Biochem* **255**, 60–66.

**Gopalakrishnan, S., Comai, G., Sambasivan, R., Francou, A., Kelly, R. G. and Tajbakhsh, S.** (2015). A Cranial Mesoderm Origin for Esophagus Striated Muscles. *Developmental cell* **34**, 694–704.

**Graham, A.** (2001). The development and evolution of the pharyngeal arches. *J Anat* **199**, 133–141.

**Graham, A., Poopalasundaram, S., Shone, V. and Kiecker, C.** (2019). A reappraisal and revision of the numbering of the pharyngeal arches. *J Anat* **235**, 1019–1023.

**Grand, F. L., Grifone, R., Mourikis, P., Houbron, C., Gigaud, C., Pujol, J., Maillet, M., Pagès, G., Rudnicki, M., Tajbakhsh, S., et al.** (2012). Six1 regulates stem cell repair potential and self-renewal during skeletal muscle regenerationSix1 and muscle stem cell self-renewal. *J Cell Biology* **198**, 815–832.

**Green, S. A., Simoes-Costa, M. and Bronner, M. E.** (2015). Evolution of vertebrates as viewed from the crest. *Nature* **520**, 474–482.

**Grieshammer, U., Sassoon, D. and Rosenthal, N.** (1992). A transgene target for positional regulators marks early rostrocaudal specification of myogenic lineages. *Cell* **69**, 79–93.

**Grifone, R., Demignon, J., Houbron, C., Souil, E., Niro, C., Seller, M. J., Hamard, G. and Maire, P.** (2005). Six1 and Six4 homeoproteins are required for Pax3 and Mrf expression during myogenesis in the mouse embryo. *Development* **132**, 2235–2249.

**Grifone, R., Demignon, J., Giordani, J., Niro, C., Souil, E., Bertin, F., Laclef, C., Xu, P.-X. and Maire, P.** (2007). Eya1 and Eya2 proteins are required for hypaxial somitic myogenesis in the mouse embryo. *Dev Biol* **302**, 602–616.

**Grifone, R., Jarry, T., Dandonneau, M., Grenier, J., Duprez, D. and Kelly, R. G.** (2008). Properties of branchiomeric and somite-derived muscle development in Tbx1 mutant embryos. *Developmental dynamics* **237**, 3071–8.

**Grim, M. and Wachtler, F.** (1991). Muscle morphogenesis in the absence of myogenic cells. *Anat Embryol* **183**, 67–70.

**Grimaldi, A., Parada, C. and Chai, Y.** (2015). A Comprehensive Study of Soft Palate Development in Mice. *PloS one* **10**, e0145018.

**Gros, J., Manceau, M., Thomé, V. and Marcelle, C.** (2005). A common somitic origin for embryonic muscle progenitors and satellite cells. *Nature* **435**, 954–958.

**Gross, M. K., Moran-Rivard, L., Velasquez, T., Nakatsu, M. N., Jagla, K. and Goulding, M.** (2000). Lbx1 is required for muscle precursor migration along a lateral pathway into the limb. *Dev Camb Engl* **127**, 413–24.

**Groves, J. A., Hammond, C. L. and Hughes, S. M.** (2005). Fgf8 drives myogenic progression of a novel lateral fast muscle fibre population in zebrafish. *Development* **132**, 4211–4222.

**Günther, S., Kim, J., Kostin, S., Lepper, C., Fan, C.-M. and Braun, T.** (2013). Myf5-Positive Satellite Cells Contribute to Pax7-Dependent Long-Term Maintenance of Adult Muscle Stem Cells. *Cell Stem Cell* **13**, 590–601.

**Guo, C., Sun, Y., Zhou, B., Adam, R. M., Li, X., Pu, W. T., Morrow, B. E., Moon, A. and Li, X.** (2011). A Tbx1-Six1/Eya1-Fgf8 genetic pathway controls mammalian cardiovascular and craniofacial morphogenesis. *J Clin Invest* **121**, 1585–1595.

**Gustafsson, M. K., Pan, H., Pinney, D. F., Liu, Y., Lewandowski, A., Epstein, D. J. and Emerson, C. P.** (2002). Myf5 is a direct target of long-range Shh signaling and Gli regulation for muscle specification. *Gene Dev* **16**, 114–126.

**Hadchouel, J., Carvajal, J. J., Daubas, P., Bajard, L., Chang, T., Rocancourt, D., Cox, D., Summerbell, D., Tajbakhsh, S., Rigby, P. W. J., et al.** (2003). Analysis of a key regulatory region upstream of the Myf5 gene reveals multiple phases of myogenesis, orchestrated at each site by a combination of elements dispersed throughout the locus. *Development* **130**, 3415–3426.

**Haldar, M., Karan, G., Tvrdik, P. and Capecchi, M. R.** (2008). Two cell lineages, myf5 and myf5-independent, participate in mouse skeletal myogenesis. *Dev Cell* **14**, 437–45.

**Han, D., Zhao, H., Parada, C., Hacia, J. G., Bringas, P. and Chai, Y.** (2012). A TGFβ-Smad4-Fgf6 signaling cascade controls myogenic differentiation and myoblast fusion during tongue development. *Dev Camb Engl* **139**, 1640–50.

**Han, A., Zhao, H., Li, J., Pelikan, R. and Chai, Y.** (2014). ALK5-mediated transforming growth factor β signaling in neural crest cells controls craniofacial muscle development via tissue-tissue interactions. *Mol Cell Biol* **34**, 3120–31.

**Harel, I., Nathan, E., Tirosh-Finkel, L., Zigdon, H., Guimarães-Camboa, N., Evans, S. M. and Tzahor, E.** (2009). Distinct Origins and Genetic Programs of Head Muscle Satellite Cells. *Dev Cell* **16**, 822–832.

**Harel, I., Maezawa, Y., Avraham, R., Rinon, A., Ma, H.-Y., Cross, J. W., Leviatan, N., Hegesh, J., Roy, A., Jacob-Hirsch, J., et al.** (2012). Pharyngeal mesoderm regulatory network controls cardiac and head muscle morphogenesis. *Proceedings of the National Academy of Sciences* **109**, 18839–18844.

**Hasson, P., DeLaurier, A., Bennett, M., Grigorieva, E., Naiche, L. A., Papaioannou, V. E., Mohun, T. J. and Logan, M. P. O.** (2010). Tbx4 and Tbx5 Acting in Connective Tissue Are Required for Limb Muscle and Tendon Patterning. *Dev Cell* **18**, 148–156.

**Hasty, P., Bradley, A., Morris, J. H., Edmondson, D. G., Venuti, J. M., Olson, E. N. and Klein, W. H.** (1993). Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature* **364**, 501–506.

**Havis, E., Coumailleau, P., Bonnet, A., Bismuth, K., Bonnin, M.-A., Johnson, R., Fan, C.-M., Relaix, F., Shi, D.-L. and Duprez, D.** (2012). Sim2 prevents entry into the myogenic program by

repressing MyoD transcription during limb embryonic myogenesis. *Development* **139**, 1910–1920.

**Helmbacher, F.** (2018). Tissue-specific activities of the Fat1 cadherin cooperate to control neuromuscular morphogenesis. *Plos Biol* **16**, e2004734.

**Helmbacher, F. and Stricker, S.** (2020). Tissue cross talks governing limb muscle development and regeneration. *Semin Cell Dev Biol* **104**, 14–30.

**Hernández-Hernández, J. M., García-González, E. G., Brun, C. E. and Rudnicki, M. A.** (2017). The myogenic regulatory factors, determinants of muscle development, cell identity and regeneration. *Semin Cell Dev Biol* **72**, 10–18.

**Hernandez-Torres, F., Rodríguez-Outeiriño, L., Franco, D. and Aranega, A. E.** (2017). Pitx2 in Embryonic and Adult Myogenesis. *Frontiers Cell Dev Biology* **5**, 46.

**Heude, E., Bouhali, K., Kurihara, Y., Kurihara, H., Couly, G., Janvier, P. and Levi, G.** (2010). Jaw muscularization requires Dlx expression by cranial neural crest cells. *Proc National Acad Sci* **107**, 11441–11446.

**Heude, É., Rivals, I., Couly, G. and Levi, G.** (2011). Masticatory muscle defects in hemifacial microsomia: A new embryological concept. *Am J Med Genet A* **155**, 1991–1995.

**Heude, E., Tesarova, M., Sefton, E. M., Jullian, E., Adachi, N., Grimaldi, A., Zikmund, T., Kaiser, J., Kardon, G., Kelly, R. G., et al.** (2018). Unique morphogenetic signatures define mammalian neck muscles and associated connective tissues. *eLife* **7**,.

**Heymann, S., Koudrova, M., Arnold, H.-H., Köster, M. and Braun, T.** (1996). Regulation and Function of SF/HGF during Migration of Limb Muscle Precursor Cells in Chicken. *Dev Biol* **180**, 566–578.

**Hiruma, T., Nakajima, Y. and Nakamura, H.** (2002). Development of pharyngeal arch arteries in early mouse embryo. *J Anat* **201**, 15–29.

**Hosokawa, R., Oka, K., Yamaza, T., Iwata, J., Urata, M., Xu, X., Bringas, P., Nonaka, K. and Chai, Y.** (2010). TGF-β mediated FGF10 signaling in cranial neural crest cells controls development of myogenic progenitor cells through tissue–tissue interactions during tongue morphogenesis. *Dev Biol* **341**, 186–195.

**Hu, P., Geles, K. G., Paik, J.-H., DePinho, R. A. and Tjian, R.** (2008). Codependent Activators Direct Myoblast-Specific MyoD Transcription. *Dev Cell* **15**, 534–546.

**Huang, R., Zhi, Q., Izpisua-Belmonte, J.-C., Christ, B. and Patel, K.** (1999). Origin and development of the avian tongue muscles. *Anat Embryol* **200**, 137–152.

**Huber, T. L., Kouskoff, V., Fehling, H. J., Palis, J. and Keller, G.** (2004). Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. *Nature* **432**, 625–630.

**Huynh, T., Chen, L., Terrell, P. and Baldini, A.** (2007). A fate map of Tbx1 expressing cells reveals heterogeneity in the second cardiac field. *genesis* **45**, 470–475.

**Ippolito, J., Arpke, R. W., Haider, K. T., Zhang, J. and Kyba, M.** (2012). Satellite cell heterogeneity revealed by G-Tool, an open algorithm to quantify myogenesis through colony-forming assays. *Skelet Muscle* **2**, 13.

**Ishii, M., Han, J., Yen, H.-Y., Sucov, H. M., Chai, Y. and Maxson, R. E.** (2005). Combined deficiencies of Msx1 and Msx2 cause impaired patterning and survival of the cranial neural crest. *Development* **132**, 4937–4950.

**Ishii, K., Suzuki, N., Mabuchi, Y., Ito, N., Kikura, N., Fukada, S., Okano, H., Takeda, S. and Akazawa, C.** (2015). Muscle Satellite Cell Protein Teneurin-4 Regulates Differentiation During Muscle Regeneration: Teneurin-4 Regulates Muscle Regeneration. *Stem Cells* **33**, 3017–3027.

**Iwata, J., Suzuki, A., Pelikan, R. C., Ho, T.-V. and Chai, Y.** (2013). Noncanonical transforming growth factor β (TGFβ) signaling in cranial neural crest cells causes tongue muscle developmental defects. *J Biological Chem* **288**, 29760–70.

**Jacob, M., Christ, B. and Jacob, H. J.** (1975). [Regional determination of the paraxial mesoderm in young chick embryos]. *Verhandlungen Der Anatomischen Gesellschaft* **69**, 263–9.

**Jen, Y., Weintraub, H. and Benezra, R.** (1992). Overexpression of Id protein inhibits the muscle differentiation program: in vivo association of Id with E2A proteins. *Gene Dev* **6**, 1466–1479.

**Jerome, L. A. and Papaioannou, V. E.** (2001). DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nature genetics* **27**, 286–91.

**Jiang, X., Iseki, S., Maxson, R. E., Sucov, H. M. and Morriss-Kay, G. M.** (2002). Tissue Origins and Interactions in the Mammalian Skull Vault. *Dev Biol* **241**, 106–116.

**Jimenez, M. A., Åkerblad, P., Sigvardsson, M. and Rosen, E. D.** (2006). Critical Role for Ebf1 and Ebf2 in the Adipogenic Transcriptional Cascade ▿ †. *Mol Cell Biol* **27**, 743–757.

**Joe, A. W. B., Yi, L., Natarajan, A., Grand, F. L., So, L., Wang, J., Rudnicki, M. A. and Rossi, F. M. V.** (2010). Muscle injury activates resident fibro/adipogenic progenitors that facilitate myogenesis. *Nat Cell Biol* **12**, 153–163.

**Kablar, B., Krastel, K., Ying, C., Asakura, A., Tapscott, S. J. and Rudnicki, M. A.** (1997). MyoD and Myf-5 differentially regulate the development of limb versus trunk skeletal muscle. *Dev Camb Engl* **124**, 4729–38.

**Kalcheim, C., Cinnamon, Y. and Kahane, N.** (1999). Myotome formation: a multistage process. *Cell Tissue Res* **296**, 161–173.

**Kaminski, H. J., Richmonds, C. R., Kusner, L. L. and Mitsumoto, H.** (2002). Differential Susceptibility of the Ocular Motor System to Disease. *Ann Ny Acad Sci* **956**, 42–54.

**Kardon, G., Harfe, B. D. and Tabin, C. J.** (2003). A Tcf4-Positive Mesodermal Population Provides a Prepattern for Vertebrate Limb Muscle Patterning. *Dev Cell* **5**, 937–944.

**Kassar-Duchossoy, L., Gayraud-Morel, B., Gomès, D., Rocancourt, D., Buckingham, M., Shinin, V. and Tajbakhsh, S.** (2004). Mrf4 determines skeletal muscle identity in Myf5:Myod double-mutant mice. *Nature* **431**, 466–471.

**Kassar-Duchossoy, L., Giacone, E., Gayraud-Morel, B., Jory, A., Gomès, D. and Tajbakhsh, S.** (2005). Pax3/Pax7 mark a novel population of primitive myogenic cells during development. *Gene Dev* **19**, 1426–1431.

**Keefe, A. C., Lawson, J. A., Flygare, S. D., Fox, Z. D., Colasanto, M. P., Mathew, S. J., Yandell, M. and Kardon, G.** (2015). Muscle stem cells contribute to myofibres in sedentary adult mice. *Nat Commun* **6**, 7087.

**Kelly, R. G.** (2013). Stem Cells in Craniofacial Development and Regeneration. 87–107.

**Kelly, R. G. and Buckingham, M. E.** (2000). Modular regulation of theMLC1F/3F gene and striated muscle diversity. *Microsc Res Techniq* **50**, 510–521.

**Kelly, A. M. and Rubinstein, N. A.** (1980). Why are fetal muscles slow? *Nature* **288**, 266–269.

**Kelly, R. G., Jerome-Majewska, L. A. and Papaioannou, V. E.** (2004). The del22q11.2 candidate gene Tbx1 regulates branchiomeric myogenesis. *Human molecular genetics* **13**, 2829–40.

**Kieny, M., Mauger, A. and Sengel, P.** (1972). Early regionalization of the somitic mesoderm as studied by the development of the axial skeleton of the chick embryo. *Dev Biol* **28**, 142–161.

**Kim, E., Wu, F., Wu, X. and Choo, H. J.** (2020). Generation of craniofacial myogenic progenitor cells from human induced pluripotent stem cells for skeletal muscle tissue regeneration. *Biomaterials* **248**, 119995.

**Kingsley, E. P., Eliason, C. M., Riede, T., Li, Z., Hiscock, T. W., Farnsworth, M., Thomson, S. L., Goller, F., Tabin, C. J. and Clarke, J. A.** (2018). Identity and novelty in the avian syrinx. *P Natl Acad Sci Usa* **115**, 201804586.

**Kioussi, C., Briata, P., Baek, S. H., Rose, D. W., Hamblet, N. S., Herman, T., Ohgi, K. A., Lin, C., Gleiberman, A., Wang, J., et al.** (2002). Identification of a Wnt/Dvl/β-Catenin → Pitx2 Pathway Mediating Cell-Type-Specific Proliferation during Development. *Cell* **111**, 673–685.

**Kong, P., Racedo, S. E., Macchiarulo, S., Hu, Z., Carpenter, C., Guo, T., Wang, T., Zheng, D. and Morrow, B. E.** (2014). Tbx1 is required autonomously for cell survival and fate in the pharyngeal core mesoderm to form the muscles of mastication. *Human molecular genetics* **23**, 4215–31.

**Konieczny, P., Swiderski, K. and Chamberlain, J. S.** (2013). Gene and cell-mediated therapies for muscular dystrophy: Therapy for Muscular Dystrophy. *Muscle Nerve* **47**, 649–663.

**Korrapati, S., Taukulis, I., Olszewski, R., Pyle, M., Gu, S., Singh, R., Griffiths, C., Martin, D., Boger, E., Morell, R. J., et al.** (2019). Single Cell and Single Nucleus RNA-Seq Reveal Cellular Heterogeneity and Homeostatic Regulatory Networks in Adult Mouse Stria Vascularis. *Front Mol Neurosci* **12**, 316.

**Kostallari, E., Baba-Amer, Y., Alonso-Martin, S., Ngoh, P., Relaix, F., Lafuste, P. and Gherardi, R. K.** (2015). Pericytes in the myovascular niche promote post-natal myofiber growth and satellite cell quiescence. *Development* **142**, 1242–1253.

**Krumlauf, R.** (1994). Hox genes in vertebrate development. *Cell* **78**, 191–201.

**Kuratani, S.** (2007). Neural crest and evolution of the vertebrate body plan. *Faseb J* **21**, A88–A88.

**Lagha, M., Kormish, J. D., Rocancourt, D., Manceau, M., Epstein, J. A., Zaret, K. S., Relaix, F. and Buckingham, M. E.** (2008). Pax3 regulation of FGF signaling affects the progression of embryonic progenitor cells into the myogenic program. *Gene Dev* **22**, 1828–37.

**Lagha, M., Brunelli, S., Messina, G., Cumano, A., Kume, T., Relaix, F. and Buckingham, M. E.** (2009). Pax3:Foxc2 Reciprocal Repression in the Somite Modulates Muscular versus Vascular Cell Fate Choice in Multipotent Progenitors. *Dev Cell* **17**, 892–899.

**Lagha, M., Sato, T., Regnault, B., Cumano, A., Zuniga, A., Licht, J., Relaix, F. and Buckingham, M.** (2010). Transcriptome analyses based on genetic screens for Pax3 myogenic targets in the mouse embryo. *Bmc Genomics* **11**, 696.

**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**, 305–315.

**Lemos, D. R., Paylor, B., Chang, C., Sampaio, A., Underhill, T. M. and Rossi, F. M. V.** (2012). Functionally Convergent White Adipogenic Progenitors of Different Lineages Participate in a Diffused System Supporting Tissue Regeneration. *Stem Cells* **30**, 1152–1162.

**Lemos, D. R., Babaeijandaghi, F., Low, M., Chang, C.-K., Lee, S. T., Fiore, D., Zhang, R.-H., Natarajan, A., Nedospasov, S. A. and Rossi, F. M. V.** (2015). Nilotinib reduces muscle fibrosis in chronic muscle injury by promoting TNF-mediated apoptosis of fibro/adipogenic progenitors. *Nat Med* **21**, 786–794.

**Lepper, C., Conway, S. J. and Fan, C.-M.** (2009). Adult satellite cells and embryonic muscle progenitors have distinct genetic requirements. *Nature* **460**, 627–631.

**Lepper, C., Partridge, T. A. and Fan, C.-M.** (2011). An absolute requirement for Pax7-positive satellite cells in acute injury-induced skeletal muscle regeneration. *Development* **138**, 3639–3646.

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

**Lescroart, F., Hamou, W., Francou, A., Théveniau-Ruissy, M., Kelly, R. G. and Buckingham, M.** (2015). Clonal analysis reveals a common origin between nonsomite-derived neck muscles and heart myocardium. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 1446–51.

**L'Honoré, A., Ouimette, J.-F., Lavertu-Jolin, M. and Drouin, J.** (2010). Pitx2 defines alternate pathways acting through MyoD during limb and somitic myogenesis. *Development* **137**, 3847–3856.

**Lima, J. E. de, Blavet, C., Bonnin, M.-A., Hirsinger, E., Comai, G., Yvernogeau, L., Bellenger, L., Mella, S., Nassari, S., Robin, C., et al.** (2020). BMP signalling directs a fibroblast-to-myoblast conversion at the connective tissue/muscle interface to pattern limb muscles. *Biorxiv* 2020.07.20.211342.

**Lin, C.-Y., Chen, W.-T., Lee, H.-C., Yang, P.-H., Yang, H.-J. and Tsai, H.-J.** (2009). The transcription factor Six1a plays an essential role in the craniofacial myogenesis of zebrafish. *Dev Biol* **331**, 152–166.

**Liu, N., Williams, A. H., Kim, Y., McAnally, J., Bezprozvannaya, S., Sutherland, L. B., Richardson, J. A., Bassel-Duby, R. and Olson, E. N.** (2007). An intragenic MEF2-dependent enhancer directs muscle-specific expression of microRNAs 1 and 133. *Proc National Acad Sci* **104**, 20844–20849.

**Liu, N., Garry, G. A., Li, S., Bezprozvannaya, S., Sanchez-Ortiz, E., Chen, B., Shelton, J. M., Jaichander, P., Bassel-Duby, R. and Olson, E. N.** (2017). A Twist2-dependent progenitor cell contributes to adult skeletal muscle. *Nat Cell Biol* **19**, 202–213.

**Logan, M., Martin, J. F., Nagy, A., Lobe, C., Olson, E. N. and Tabin, C. J.** (2002). Expression of Cre recombinase in the developing mouse limb bud driven by aPrxl enhancer. *Genesis* **33**, 77–80.

**Lukjanenko, L., Karaz, S., Stuelsatz, P., Gurriaran-Rodriguez, U., Michaud, J., Dammone, G., Sizzano, F., Mashinchian, O., Ancel, S., Migliavacca, E., et al.** (2019). Aging Disrupts Muscle Stem Cell Function by Impairing Matricellular WISP1 Secretion from Fibro-Adipogenic Progenitors. *Cell Stem Cell* **24**, 433-446.e7.

**MacQuarrie, K. L., Yao, Z., Fong, A. P., Diede, S. J., Rudzinski, E. R., Hawkins, D. S. and Tapscott, S. J.** (2012). Comparison of Genome-Wide Binding of MyoD in Normal Human Myogenic Cells and Rhabdomyosarcomas Identifies Regional and Local Suppression of Promyogenic Transcription Factors. *Mol Cell Biol* **33**, 773–784.

**Maire, P., Santos, M. D., Madani, R., Sakakibara, I., Viaut, C. and Wurmser, M.** (2020). Myogenesis control by SIX transcriptional complexes. *Semin Cell Dev Biol* **104**, 51–64.

**Maltzahn, J. von, Jones, A. E., Parks, R. J. and Rudnicki, M. A.** (2013). Pax7 is critical for the normal function of satellite cells in adult skeletal muscle. *P Natl Acad Sci Usa* **110**, 16474–9.

**Man, C. Y. Y. W., Chinnery, P. F. and Griffiths, P. G.** (2005). Extraocular muscles have fundamentally distinct properties that make them selectively vulnerable to certain disorders. *Neuromuscular Disord* **15**, 17–23.

**Mankoo, B. S., Collins, N. S., Ashby, P., Grigorieva, E., Pevny, L. H., Candia, A., Wright, C. V. E., Rigby, P. W. J. and Pachnis, V.** (1999). Mox2 is a component of the genetic hierarchy controlling limb muscle development. *Nature* **400**, 69–73.

**Mansouri, A. and Gruss, P.** (1998). Pax3 and Pax7 are expressed in commissural neurons and restrict ventral neuronal identity in the spinal cord. *Mech Develop* **78**, 171–178.

**Martik, M. L., Gandhi, S., Uy, B. R., Gillis, J. A., Green, S. A., Simoes-Costa, M. and Bronner, M. E.** (2019). Evolution of the new head by gradual acquisition of neural crest regulatory circuits. *Nature* **574**, 675–678.

**Mashinchian, O., Pisconti, A., Moal, E. L. and Bentzinger, C. F.** (2018). Chapter Two The Muscle Stem Cell Niche in Health and Disease. *Curr Top Dev Biol* **126**, 23–65.

**Mathew, S. J., Hansen, J. M., Merrell, A. J., Murphy, M. M., Lawson, J. A., Hutcheson, D. A., Hansen, M. S., Angus-Hill, M. and Kardon, G.** (2011). Connective tissue fibroblasts and Tcf4 regulate myogenesis. *Dev Camb Engl* **138**, 371–84.

**Matsuoka, T., Ahlberg, P. E., Kessaris, N., Iannarelli, P., Dennehy, U., Richardson, W. D., McMahon, A. P. and Koentges, G.** (2005). Neural crest origins of the neck and shoulder. *Nature* **436**, 347.

**Maurange, C. and Paro, R.** (2002). A cellular memory module conveys epigenetic inheritance of hedgehog expression during Drosophila wing imaginal disc development. *Gene Dev* **16**, 2672–2683.

**Meadows, E., Cho, J.-H., Flynn, J. M. and Klein, W. H.** (2008). Myogenin regulates a distinct genetic program in adult muscle stem cells. *Dev Biol* **322**, 406–414.

**Megeney, L. A. and Rudnicki, M. A.** (1995). Determination versus differentiation and the MyoD family of transcription factors. *Biochem Cell Biol* **73**, 723–732.

**Meilhac, S. M. and Buckingham, M. E.** (2018). The deployment of cell lineages that form the mammalian heart. *Nat Rev Cardiol* **15**, 705–724.

**Mercuri, E. and Muntoni, F.** (2013). Muscular dystrophies. *Lancet* **381**, 845–860.

**Merrell, A. J. and Kardon, G.** (2013). Development of the diaphragm - a skeletal muscle essential for mammalian respiration. *Febs J* **280**, 4026–4035.

**Messina, G. and Cossu, G.** (2009). The origin of embryonic and fetal myoblasts: a role of Pax3 and Pax7. *Gene Dev* **23**, 902–905.

**Messina, G., Biressi, S., Monteverde, S., Magli, A., Cassano, M., Perani, L., Roncaglia, E., Tagliafico, E., Starnes, L., Campbell, C. E., et al.** (2010). Nfix Regulates Fetal-Specific Transcription in Developing Skeletal Muscle. *Cell* **140**, 554–566.

**Michailovici, I., Eigler, T. and Tzahor, E.** (2015). Current Topics in Developmental Biology. *Part 1: Craniofacial Morphogenesis and Regeneration: From Cells to Tissues to Organs* **115**, 3–30.

**Millay, D. P., O'Rourke, J. R., Sutherland, L. B., Bezprozvannaya, S., Shelton, J. M., Bassel-Duby, R. and Olson, E. N.** (2013). Myomaker is a membrane activator of myoblast fusion and muscle formation. *Nature* **499**, 301–5.

**Miller, R. G., Sharma, K. R., Pavlath, G. K., Gussoni, E., Mynhier, M., Yu, P., Lanctot, A. M., Greco, C. M., Steinman, L. and Blau, H. M.** (1997). Myoblast implantation in Duchenne muscular dystrophy: The San Francisco study. *Muscle Nerve* **20**, 469–478.

**Miller, K. A., Barrow, J., Collinson, J. M., Davidson, S., Lear, M., Hill, R. E. and MacKenzie, A.** (2007). A highly conserved Wnt-dependent TCF4 binding site within the proximal enhancer of the anti-myogenic Msx1 gene supports expression within Pax3-expressing limb bud muscle precursor cells. *Dev Biol* **311**, 665–678.

**Millington, G., Elliott, K. H., Chang, Y.-T., Chang, C.-F., Dlugosz, A. and Brugmann, S. A.** (2017). Cilia-dependent GLI processing in neural crest cells is required for tongue development. *Dev Biol* **424**, 124–137.

**Minoux, M. and Rijli, F. M.** (2010). Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. *Development (Cambridge, England)* **137**, 2605–21.

**Mitchell, K. J., Pannérec, A., Cadot, B., Parlakian, A., Besson, V., Gomes, E. R., Marazzi, G. and Sassoon, D. A.** (2010). Identification and characterization of a non-satellite cell muscle resident progenitor during postnatal development. *Nat Cell Biol* **12**, 257–266.

**Miyashita, T.** (2016). Fishing for jaws in early vertebrate evolution: a new hypothesis of mandibular confinement. *Biological reviews of the Cambridge Philosophical Society* **91**, 611–57.

**Moncaut, N., Cross, J. W., Siligan, C., Keith, A., Taylor, K., Rigby, P. W. J. and Carvajal, J. J.** (2012). Musculin and TCF21 coordinate the maintenance of myogenic regulatory factor expression levels during mouse craniofacial development. *Development* **139**, 958–967.

**Moncaut, N., Rigby, P. W. J. and Carvajal, J. J.** (2013). Dial M(RF) for myogenesis. *Febs J* **280**, 3980–3990.

**Monsoro-Burq, A. H.** (2015). PAX transcription factors in neural crest development. *Semin Cell Dev Biol* **44**, 87–96.

**Montarras, D., L'honoré, A. and Buckingham, M.** (2013). Lying low but ready for action: the quiescent muscle satellite cell. *Febs J* **280**, 4036–50.

**Mourikis, P. and Tajbakhsh, S.** (2014). Distinct contextual roles for Notch signalling in skeletal muscle stem cells. *Bmc Dev Biol* **14**, 2.

**Mourikis, P., Gopalakrishnan, S., Sambasivan, R. and Tajbakhsh, S.** (2012). Cell-autonomous Notch activity maintains the temporal specification potential of skeletal muscle stem cells. *Development* **139**, 4536–4548.

**Mozzetta, C., Consalvi, S., Saccone, V., Tierney, M., Diamantini, A., Mitchell, K. J., Marazzi, G., Borsellino, G., Battistini, L., Sassoon, D., et al.** (2013). Fibroadipogenic progenitors mediate the ability of HDAC inhibitors to promote regeneration in dystrophic muscles of young, but not old Mdx mice. *Embo Mol Med* **5**, 626–39.

**Muhl, L., Genové, G., Leptidis, S., Liu, J., He, L., Mocci, G., Sun, Y., Gustafsson, S., Buyandelger, B., Chivukula, I. V., et al.** (2020). Single-cell analysis uncovers fibroblast

heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat Commun* **11**, 3953.

**Murakami, G. and Nakamura, H.** (1991). Somites and the Pattern Formation of Trunk Muscles: A Study in Quail-Chick Chimera. *Arch Histol Cytol* **54**, 249–258.

**Murphy, M. M., Lawson, J. A., Mathew, S. J., Hutcheson, D. A. and Kardon, G.** (2011). Satellite cells, connective tissue fibroblasts and their interactions are crucial for muscle regeneration. *Development* **138**, 3625–3637.

**Nabeshima, Y., Hanaoka, K., Hayasaka, M., Esuml, E., Li, S., Nonaka, I. and Nabeshima, Y.** (1993). Myogenin gene disruption results in perinatal lethality because of severe muscle defect. *Nature* **364**, 532–535.

**Nassari, S., Duprez, D. and Fournier-Thibault, C.** (2017). Non-myogenic Contribution to Muscle Development and Homeostasis: The Role of Connective Tissues. *Frontiers in Cell and Developmental Biology* **5**, 22.

**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.

**Naumann, B., Warth, P., Olsson, L. and Konstantinidis, P.** (2017). The development of the cucullaris muscle and the branchial musculature in the Longnose Gar, (Lepisosteus osseus, Lepisosteiformes, Actinopterygii) and its implications for the evolution and development of the head/trunk interface in vertebrates. *Evolution & Development* **19**, 263–276.

**Niro, C., Demignon, J., Vincent, S., Liu, Y., Giordani, J., Sgarioto, N., Favier, M., Guillet-Deniau, I., Blais, A. and Maire, P.** (2010). Six1 and Six4 gene expression is necessary to activate the fast-type muscle gene program in the mouse primary myotome. *Dev Biol* **338**, 168–182.

**Noden, D. M.** (1983). The role of the neural crest in patterning of avian cranial skeletal, connective, and muscle tissues. *Dev Biol* **96**, 144–165.

**Noden, D. M. and Francis-West, P.** (2006). The differentiation and morphogenesis of craniofacial muscles. *Dev Dynam* **235**, 1194–1218.

**Noden, D. M. and Trainor, P. A.** (2005). Relations and interactions between cranial mesoderm and neural crest populations. *Journal of anatomy* **207**, 575–601.

**Nogueira, J. M., Hawrot, K., Sharpe, C., Noble, A., Wood, W. M., Jorge, E. C., Goldhamer, D. J., Kardon, G. and Dietrich, S.** (2015). The emergence of Pax7-expressing muscle stem cells during vertebrate head muscle development. *Front Aging Neurosci* **7**, 62.

**Nowicki, J. L. and Burke, A. C.** (2000). Hox genes and morphological identity: axial versus lateral patterning in the vertebrate mesoderm. *Dev Camb Engl* **127**, 4265–75.

**Nowicki, J. L., Takimoto, R. and Burke, A. C.** (2003). The lateral somitic frontier: dorso-ventral aspects of anterio-posterior regionalization in avian embryos. *Mech Develop* **120**, 227–240.

**Okano, J., Sakai, Y. and Shiota, K.** (2008). Retinoic acid down-regulates Tbx1 expression and induces abnormal differentiation of tongue muscles in fetal mice. *Dev Dyn* **237**, 3059–70.

**Ono, Y., Boldrin, L., Knopp, P., Morgan, J. E. and Zammit, P. S.** (2009). Muscle satellite cells are a functionally heterogeneous population in both somite-derived and branchiomeric muscles. *Dev Biol* **337**, 29–41.

**Ontell, M. and Kozeka, K.** (1984). The organogenesis of murine striated muscle: A cytoarchitectural study. *Am J Anat* **171**, 133–148.

**Palacios, D., Mozzetta, C., Consalvi, S., Caretti, G., Saccone, V., Proserpio, V., Marquez, V. E., Valente, S., Mai, A., Forcales, S. V., et al.** (2010). TNF/p38α/Polycomb Signaling to Pax7 Locus in Satellite Cells Links Inflammation to the Epigenetic Control of Muscle Regeneration. *Cell Stem Cell* **7**, 455–469.

**Pan, H., Gustafsson, M. K., Aruga, J., Tiedken, J. J., Chen, J. C. J. and Emerson, C. P.** (2011). A role for Zic1 and Zic2 in Myf5 regulation and somite myogenesis. *Dev Biol* **351**, 120–7.

**Pannerec, A., Formicola, L., Besson, V., Marazzi, G. and Sassoon, D. A.** (2013). Defining skeletal muscle resident progenitors and their cell fate potentials. *Development* **140**, 2879–2891.

**Parada, C. and Chai, Y.** (2015). Mandible and Tongue Development. *Current topics in developmental biology* **115**, 31–58.

**Parada, C., Han, D. and Chai, Y.** (2012). Molecular and cellular regulatory mechanisms of tongue myogenesis. *Journal of dental research* **91**, 528–35.

**Passerieux, E., Rossignol, R., Letellier, T. and Delage, J. P.** (2007). Physical continuity of the perimysium from myofibers to tendons: Involvement in lateral force transmission in skeletal muscle. *J Struct Biol* **159**, 19–28.

**Patapoutian, A., Wold, B. J. and Wagner, R. A.** (1995). Evidence for Developmentally Programmed Transdifferentiation in Mouse Esophageal Muscle. *Science* **270**, 1818–1821.

**Pavlath, G. K., Thaloor, D., Rando, T. A., Cheong, M., English, A. W. and Zheng, B.** (1998). Heterogeneity among muscle precursor cells in adult skeletal muscles with differing regenerative capacities. *Developmental Dynamics* **212**, 495–508.

**Pearse, R. V., Scherz, P. J., Campbell, J. K. and Tabin, C. J.** (2007). A cellular lineage analysis of the chick limb bud. *Dev Biol* **310**, 388–400.

**Poopalasundaram, S., Richardson, J., Scott, A., Donovan, A., Liu, K. and Graham, A.** (2019). Diminution of pharyngeal segmentation and the evolution of the amniotes. *Zoological Lett* **5**, 6.

**Potthoff, M. J. and Olson, E. N.** (2007). MEF2: a central regulator of diverse developmental programs. *Development* **134**, 4131–4140.

**Prummel, K. D., Nieuwenhuize, S. and Mosimann, C.** (2020). The lateral plate mesoderm. *Dev Camb Engl* **147**, dev175059.

**Prunotto, C., Crepaldi, T., Forni, P. E., Ieraci, A., Kelly, R. G., Tajbakhsh, S., Buckingham, M. and Ponzetto, C.** (2004). Analysis of Mlc-lacZ Met mutants highlights the essential function of Met for migratory precursors of hypaxial muscles and reveals a role for Met in the development of hyoid arch-derived facial muscles. *Developmental dynamics* **231**, 582–91.

**Randolph, M. E., Luo, Q., Ho, J., Vest, K. E., Sokoloff, A. J. and Pavlath, G. K.** (2014). Ageing and muscular dystrophy differentially affect murine pharyngeal muscles in a region-dependent manner. *J Physiology* **592**, 5301–5315.

**Randolph, M. E., Phillips, B. L., Choo, H., Vest, K. E., Vera, Y. and Pavlath, G. K.** (2015). Pharyngeal Satellite Cells Undergo Myogenesis Under Basal Conditions and Are Required for Pharyngeal Muscle Maintenance. *STEM CELLS* **33**, 3581–3595.

**Rayment, I., Holden, H., Whittaker, M., Yohn, C., Lorenz, M., Holmes, K. and Milligan, R.** (1993). Structure of the actin-myosin complex and its implications for muscle contraction. *Science* **261**, 58–65.

**Relaix, F. and Zammit, P. S.** (2012). Satellite cells are essential for skeletal muscle regeneration: the cell on the edge returns centre stage. *Development* **139**, 2845–2856.

**Relaix, F., Rocancourt, D., Mansouri, A. and Buckingham, M.** (2004). Divergent functions of murine Pax3 and Pax7 in limb muscle development. *Gene Dev* **18**, 1088–1105.

**Relaix, F., Rocancourt, D., Mansouri, A. and Buckingham, M.** (2005). A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. *Nature* **435**, 948–953.

**Relaix, F., Montarras, D., Zaffran, S., Gayraud-Morel, B., Rocancourt, D., Tajbakhsh, S., Mansouri, A., Cumano, A. and Buckingham, M.** (2006). Pax3 and Pax7 have distinct and overlapping functions in adult muscle progenitor cells. *The Journal of Cell Biology* **172**, 91–102.

**Relaix, F., Demignon, J., Laclef, C., Pujol, J., Santolini, M., Niro, C., Lagha, M., Rocancourt, D., Buckingham, M. and Maire, P.** (2013). Six Homeoproteins Directly Activate Myod Expression in the Gene Regulatory Networks That Control Early Myogenesis. *Plos Genet* **9**, e1003425.

**Ribas, R., Moncaut, N., Siligan, C., Taylor, K., Cross, J. W., Rigby, P. W. J. and Carvajal, J. J.** (2011). Members of the TEAD family of transcription factors regulate the expression of Myf5 in ventral somitic compartments. *Dev Biol* **355**, 372–380.

**Rinon, A., Lazar, S., Marshall, H., Buchmann-Moller, S., Neufeld, A., Elhanany-Tamir, H., Taketo, M. M., Sommer, L., Krumlauf, R. and Tzahor, E.** (2007). Cranial neural crest cells regulate head muscle patterning and differentiation during vertebrate embryogenesis. *Development* **134**, 3065–3075.

**Rishniw, M., Xin, H., Deng, K. and Kotlikoff, M. I.** (2003). Skeletal myogenesis in the mouse esophagus does not occur through transdifferentiation. *genesis* **36**, 81–82.

**Robson, L. G. and Hughes, S. M.** (1996). The distal limb environment regulates MyoD accumulation and muscle differentiation in mouse-chick chimæric limbs. *Dev Camb Engl* **122**, 3899–910.

**Rubenstein, A. B., Smith, G. R., Raue, U., Begue, G., Minchev, K., Ruf-Zamojski, F., Nair, V. D., Wang, X., Zhou, L., Zaslavsky, E., et al.** (2020). Single-cell transcriptional profiles in human skeletal muscle. *Sci Rep-uk* **10**, 229.

**Rudnicki, M. A., Braun, T., Hinuma, S. and Jaenisch, R.** (1992). Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. *Cell* **71**, 383–90.

**Rudnicki, M. A., Schnegelsberg, P. N. J., Stead, R. H., Braun, T., Arnold, H.-H. and Jaenisch, R.** (1993). MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75**, 1351–1359.

**Saberi, M., Pu, Q., Valasek, P., Norizadeh-Abbariki, T., Patel, K. and Huang, R.** (2017). The hypaxial origin of the epaxially located rhomboid muscles. *Ann Anat - Anatomischer Anzeiger* **214**, 15–20.

**Sacchetti, B., Funari, A., Remoli, C., Giannicola, G., Kogler, G., Liedtke, S., Cossu, G., Serafini, M., Sampaolesi, M., Tagliafico, E., et al.** (2016). No Identical "Mesenchymal Stem Cells" at Different Times and Sites: Human Committed Progenitors of Distinct Origin and Differentiation Potential Are Incorporated as Adventitial Cells in Microvessels. *Stem Cell Rep* **6**, 897–913.

**Sagai, T., Amano, T., Maeno, A., Kimura, T., Nakamoto, M., Takehana, Y., Naruse, K., Okada, N., Kiyonari, H. and Shiroishi, T.** (2017). Evolution of Shh endoderm enhancers during morphological transition from ventral lungs to dorsal gas bladder. *Nat Commun* **8**, 14300.

**Sambasivan, R., Gayraud-Morel, B., Dumas, G., Cimper, C., Paisant, S., Kelly, R. G., Kelly, R. and Tajbakhsh, S.** (2009). Distinct regulatory cascades govern extraocular and pharyngeal arch muscle progenitor cell fates. *Developmental cell* **16**, 810–21.

**Sambasivan, R., Yao, R., Kissenpfennig, A., Wittenberghe, L. V., Paldi, A., Gayraud-Morel, B., Guenou, H., Malissen, B., Tajbakhsh, S. and Galy, A.** (2011a). Pax7-expressing satellite cells are indispensable for adult skeletal muscle regeneration. *Development* **138**, 3647–3656.

**Sambasivan, R., Kuratani, S. and Tajbakhsh, S.** (2011b). An eye on the head: the development and evolution of craniofacial muscles. *Development* **138**, 2401–2415.

**Sampath, S. C., Sampath, S. C., Ho, A. T. V., Corbel, S. Y., Millstone, J. D., Lamb, J., Walker, J., Kinzel, B., Schmedt, C. and Blau, H. M.** (2018). Induction of muscle stem cell quiescence by the secreted niche factor Oncostatin M. *Nat Commun* **9**, 1531.

**Santagati, F. and Rijli, F. M.** (2003). Cranial neural crest and the building of the vertebrate head. *Nat Rev Neurosci* **4**, 806–818.

**Sato, T., Rocancourt, D., Marques, L., Thorsteinsdóttir, S. and Buckingham, M.** (2010). A Pax3/Dmrt2/Myf5 Regulatory Cascade Functions at the Onset of Myogenesis. *Plos Genet* **6**, e1000897.

**Schäfer, K. and Braun, T.** (1999). Early specification of limb muscle precursor cells by the homeobox gene Lbx1h. *Nat Genet* **23**, 213–216.

**Schofield, R.** (1978). The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells* **4**, 7–25.

**Schoser, B. G. H. and Pongratz, D.** (2006). Extraocular Mitochondrial Myopathies and their Differential Diagnoses. *Strabismus* **14**, 107–113.

**Schubert, F. R., Singh, A. J., Afoyalan, O., Kioussi, C. and Dietrich, S.** (2018). To roll the eyes and snap a bite - function, development and evolution of craniofacial muscles. *Seminars in cell & developmental biology*.

**Schuster-Gossler, K., Cordes, R. and Gossler, A.** (2006). Premature myogenic differentiation and depletion of progenitor cells cause severe muscle hypotrophy in Delta1 mutants. *Proc National Acad Sci* **104**, 537–542.

**Seale, P., Bjork, B., Yang, W., Kajimura, S., Chin, S., Kuang, S., Scimè, A., Devarakonda, S., Conroe, H. M., Erdjument-Bromage, H., et al.** (2008). PRDM16 controls a brown fat/skeletal muscle switch. *Nature* **454**, 961–967.

**Sefton, E. M. and Kardon, G.** (2019). Connecting muscle development, birth defects, and evolution: An essential role for muscle connective tissue. *Curr Top Dev Biol* **132**, 137–176.

**Sefton, E. M., Gallardo, M. and Kardon, G.** (2018). Developmental origin and morphogenesis of the diaphragm, an essential mammalian muscle. *Dev Biol* **440**, 64–73.

**Shan, T., Liang, X., Bi, P., Zhang, P., Liu, W. and Kuang, S.** (2013). Distinct populations of adipogenic and myogenic Myf5-lineage progenitors in white adipose tissues. *J Lipid Res* **54**, 2214–24.

**Shapira, S. N., Lim, H.-W., Rajakumari, S., Sakers, A. P., Ishibashi, J., Harms, M. J., Won, K.-J. and Seale, P.** (2017). EBF2 transcriptionally regulates brown adipogenesis via the histone reader DPF3 and the BAF chromatin remodeling complex. *Gene Dev* **31**, 660–673.

**Shefer, G., Wleklinski-Lee, M. and Yablonka-Reuveni, Z.** (2004). Skeletal muscle satellite cells can spontaneously enter an alternative mesenchymal pathway. *J Cell Sci* **117**, 5393–5404.

**Shih, H., Gross, M. K. and Kioussi, C.** (2007). Cranial muscle defects of Pitx2 mutants result from specification defects in the first branchial arch. *Proceedings of the National Academy of Sciences* **104**, 5907–5912.

**Shih, H. P., Gross, M. K. and Kioussi, C.** (2008). Muscle development: Forming the head and trunk muscles. *Acta Histochem* **110**, 97–108.

**Siles, L., Sánchez-Tilló, E., Lim, J.-W., Darling, D. S., Kroll, K. L. and Postigo, A.** (2013). ZEB1 Imposes a Temporary Stage-Dependent Inhibition of Muscle Gene Expression and Differentiation via CtBP-Mediated Transcriptional Repression. *Mol Cell Biol* **33**, 1368–1382.

**Smith, K. K.** (1992). The evolution of the mammalian pharynx. *Zool J Linn Soc-lond* **104**, 313–349.

**Soleimani, V. D., Yin, H., Jahani-Asl, A., Ming, H., Kockx, C. E. M., Ijcken, W. F. J. van, Grosveld, F. and Rudnicki, M. A.** (2012a). Snail regulates MyoD binding-site occupancy to direct enhancer switching and differentiation-specific transcription in myogenesis. *Mol Cell* **47**, 457–68.

**Soleimani, V. D., Punch, V. G., Kawabe, Y., Jones, A. E., Palidwor, G. A., Porter, C. J., Cross, J. W., Carvajal, J. J., Kockx, C. E. M., van IJcken, W. F. J., et al.** (2012b). Transcriptional Dominance of Pax7 in Adult Myogenesis Is Due to High-Affinity Recognition of Homeodomain Motifs. *Dev Cell* **22**, 1208–1220.

**Solnica-Krezel, L. and Sepich, D. S.** (2012). Gastrulation: Making and Shaping Germ Layers. *Annu Rev Cell Dev Bi* **28**, 687–717.

**Spencer, R. F. and Porter, J. D.** (2006). Biological organization of the extraocular muscles. *Prog Brain Res* **151**, 43–80.

**Spitz, F., Demignon, J., Porteu, A., Kahn, A., Concordet, J.-P., Daegelen, D. and Maire, P.** (1998). Expression of myogenin during embryogenesis is controlled by Six/sine oculis homeoproteins through a conserved MEF3 binding site. *Proc National Acad Sci* **95**, 14220–14225.

**Stuelsatz, P., Keire, P., Almuly, R. and Yablonka-Reuveni, Z.** (2012). A Contemporary Atlas of the Mouse Diaphragm: Myogenicity, Vascularity, and the Pax3 Connection. *J Histochem Cytochem* **60**, 638–657.

**Stuelsatz, P., Shearer, A., Li, Y., Muir, L. A., Ieronimakis, N., Shen, Q. W., Kirillova, I. and Yablonka-Reuveni, Z.** (2015). Extraocular muscle satellite cells are high performance myo-engines retaining efficient regenerative capacity in dystrophin deficiency. *Dev Biol* **397**, 31–44.

**Sumiyama, K. and Tanave, A.** (2020). The regulatory landscape of the Dlx gene system in branchial arches: shared characteristics among Dlx bigene clusters and evolution. *Dev Growth Differ* **62**, 355–362.

**Summerbell, D., Ashby, P. R., Coutelle, O., Cox, D., Yee, S. and Rigby, P. W.** (2000). The expression of Myf5 in the developing mouse embryo is controlled by discrete and dispersed enhancers specific for particular populations of skeletal muscle precursors. *Dev Camb Engl* **127**, 3745–57.

**Swartz, M. E., Eberhart, J., Pasquale, E. B. and Krull, C. E.** (2001). EphA4/ephrin-A5 interactions in muscle precursor cell migration in the avian forelimb. *Dev Camb Engl* **128**, 4669–80.

**Swinehart, I. T., Schlientz, A. J., Quintanilla, C. A., Mortlock, D. P. and Wellik, D. M.** (2013). Hox11 genes are required for regional patterning and integration of muscle, tendon and bone. *Development* **140**, 4574–4582.

**Tabler, J. M., Rigney, M. M., Berman, G. J., Gopalakrishnan, S., Heude, E., Al-Lami, H. A., Yannakoudakis, B. Z., Fitch, R. D., Carter, C. M., Vokes, S. A., et al.** (2017). Cilia-mediated Hedgehog signaling controls form and function in the mammalian larynx. *Elife* **6**, e19153.

**Tajbakhsh, S.** (2009). Skeletal muscle stem cells in developmental versus regenerative myogenesis. *Journal of Internal Medicine* **266**, 372–389.

**Tajbakhsh, S., Rocancourt, D. and Buckingham, M.** (1996a). Muscle progenitor cells failing to respond to positional cues adopt non-myogenic fates in myf-5 null mice. *Nature* **384**, 266–270.

**Tajbakhsh, S., Bober, E., Babinet, C., Pournin, S., Arnold, H. and Buckingham, M.** (1996b). Gene targeting the myf-5 locus with nlacZ reveals expression of this myogenic factor in mature skeletal muscle fibres as well as early embryonic muscle. *Developmental dynamics* **206**, 291–300.

**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**, 127–138.

**Takahashi, K. and Yamanaka, S.** (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–76.

**Tapscott, S. J.** (2005). The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* **132**, 2685–2695.

**Teboul, L., Hadchouel, J., Daubas, P., Summerbell, D., Buckingham, M. and Rigby, P. W. J.** (2002). The early epaxial enhancer is essential for the initial expression of the skeletal muscle determination gene Myf5 but not for subsequent, multiple phases of somitic myogenesis. *Dev Camb Engl* **129**, 4571–80.

**Tierney, M. T. and Sacco, A.** (2016). Satellite Cell Heterogeneity in Skeletal Muscle Homeostasis. *Trends Cell Biol* **26**, 434–44.

**Tierney, M. T., Gromova, A., Sesillo, F. B., Sala, D., Spenlé, C., Orend, G. and Sacco, A.** (2016). Autonomous Extracellular Matrix Remodeling Controls a Progressive Adaptation in Muscle Stem Cell Regenerative Capacity during Development. *Cell Reports* **14**, 1940–1952.

**Tokita, M. and Schneider, R. A.** (2009). Developmental origins of species-specific muscle pattern. *Dev Biol* **331**, 311–325.

**Tokita, M., Nakayama, T., Schneider, R. A. and Agata, K.** (2013). Molecular and cellular changes associated with the evolution of novel jaw muscles in parrots. *Proc Royal Soc B Biological Sci* **280**, 20122319.

**Tosney, K. W., Dehnbostel, D. B. and Erickson, C. A.** (1994). Neural Crest Cells Prefer the Myotome's Basal Lamina over the Sclerotome as a Substratum. *Dev Biol* **163**, 389–406.

**Tremblay, J. P., Malouin, F., Roy, R., Huard, J., Bouchard, J. P., Satoh, A. and Richards, C. L.** (1993). Results of a Triple Blind Clinical Study of Myoblast Transplantations without Immunosuppressive Treatment in Young Boys with Duchenne Muscular Dystrophy. *Cell Transplant* **2**, 99–112.

**Tzahor, E.** (2015). Head muscle development. *Results Problems Cell Differ* **56**, 123–42.

**Tzahor, E. and Evans, S. M.** (2011). Pharyngeal mesoderm development during embryogenesis: implications for both heart and head myogenesis. *Cardiovascular Research* **91**, 196–202.

**Tzahor, E., Kempf, H., Mootoosamy, R. C., Poon, A. C., Abzhanov, A., Tabin, C. J., Dietrich, S. and Lassar, A. B.** (2003). Antagonists of Wnt and BMP signaling promote the formation of vertebrate head muscle. *Gene Dev* **17**, 3087–3099.

**Uezumi, A., Fukada, S., Yamamoto, N., Takeda, S. and Tsuchida, K.** (2010). Mesenchymal progenitors distinct from satellite cells contribute to ectopic fat cell formation in skeletal muscle. *Nat Cell Biol* **12**, 143–152.

**Uezumi, A., Ito, T., Morikawa, D., Shimizu, N., Yoneda, T., Segawa, M., Yamaguchi, M., Ogawa, R., Matev, M. M., Miyagoe-Suzuki, Y., et al.** (2011). Fibrosis and adipogenesis originate from a common mesenchymal progenitor in skeletal muscle. *J Cell Sci* **124**, 3654–64.

**Uezumi, A., Ikemoto-Uezumi, M. and Tsuchida, K.** (2014). Roles of nonmyogenic mesenchymal progenitors in pathogenesis and regeneration of skeletal muscle. *Front Physiol* **5**, 68.

**Valdez, G., Tapia, J. C., Lichtman, J. W., Fox, M. A. and Sanes, J. R.** (2012). Shared Resistance to Aging and ALS in Neuromuscular Junctions of Specific Muscles. *Plos One* **7**, e34640.

**Vallecillo-García, P., Orgeur, M., Hofe-Schneider, S. vom, Stumm, J., Kappert, V., Ibrahim, D. M., Börno, S. T., Hayashi, S., Relaix, F., Hildebrandt, K., et al.** (2017). Odd skipped-related 1 identifies a population of embryonic fibro-adipogenic progenitors regulating myogenesis during limb development. *Nat Commun* **8**, 1218.

**Vasyutina, E., Stebler, J., Brand-Saberi, B., Schulz, S., Raz, E. and Birchmeier, C.** (2005). CXCR4 and Gab1 cooperate to control the development of migrating muscle progenitor cells. *Gene Dev* **19**, 2187–2198.

**Vasyutina, E., Lenhard, D. C., Wende, H., Erdmann, B., Epstein, J. A. and Birchmeier, C.** (2007). RBP-J (Rbpsuh) is essential to maintain muscle progenitor cells and to generate satellite cells. *Proc National Acad Sci* **104**, 4443–4448.

**Vega-Lopez, G. A., Cerrizuela, S. and Aybar, M. J.** (2017). Trunk neural crest cells: formation, migration and beyond. *Int J Dev Biology* **61**, 5–15.

**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 Biology* **128**, 563–576.

**Vinagre, T., Moncaut, N., Carapuço, M., Nóvoa, A., Bom, J. and Mallo, M.** (2010). Evidence for a Myotomal Hox/Myf Cascade Governing Nonautonomous Control of Rib Specification within Global Vertebral Domains. *Dev Cell* **18**, 655–661.

**Vitelli, F., Morishima, M., Taddei, I., Lindsay, E. A. and Baldini, A.** (2002). Tbx1 mutation causes multiple cardiovascular defects and disrupts neural crest and cranial nerve migratory pathways. *Human molecular genetics* **11**, 915–22.

**Wada, M. R., Inagawa-Ogashiwa, M., Shimizu, S., Yasumoto, S. and Hashimoto, N.** (2002). Generation of different fates from multipotent muscle stem cells. *Dev Camb Engl* **129**, 2987–95.

**Walz, P. C., Hiatt, K. K., Naidu, M. and Halum, S. L.** (2008). Characterization of laryngeal muscle stem cell survival and proliferation. *Laryngoscope* **118**, 1422–6.

**Wang, J., Kumar, R. M., Biggs, V. J., Lee, H., Chen, Y., Kagey, M. H., Young, R. A. and Abate-Shen, C.** (2011). The Msx1 Homeoprotein Recruits Polycomb to the Nuclear Periphery during Development. *Dev Cell* **21**, 575–588.

**Wang, H., Holland, P. W. H. W. and Takahashi, T.** (2019). Gene profiling of head mesoderm in early zebrafish development: insights into the evolution of cranial mesoderm. *EvoDevo* **10**, 14.

**White, R. B., Biérinx, A.-S., Gnocchi, V. F. and Zammit, P. S.** (2010). Dynamics of muscle fibre growth during postnatal mouse development. *Bmc Dev Biol* **10**, 21.

**Whitesell, T. R., Chrystal, P. W., Ryu, J.-R. R., Munsie, N., Grosse, A., French, C. R., Workentine, M. L., Li, R., Zhu, L. J., Waskiewicz, A., et al.** (2019). foxc1 is required for embryonic head vascular smooth muscle differentiation in zebrafish. *Developmental biology* **453**, 34–47.

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

**Wilschut, K. J., Ling, V. B. and Bernstein, H. S.** (2012). Concise Review: Stem Cell Therapy for Muscular Dystrophies. *Stem Cell Transl Med* **1**, 833–842.

**Winter, B., Braun, T. and Arnold, H. H.** (1992). Co-operativity of functional domains in the muscle-specific transcription factor Myf-5. *Embo J* **11**, 1843–55.

**Wotton, K. R., Schubert, F. R. and Dietrich, S.** (2014). Hypaxial Muscle: Controversial Classification and Controversial Data? *Results Problems Cell Differ* **56**, 25–48.

**Wu, W., Huang, R., Wu, Q., Li, P., Chen, J., Li, B. and Liu, H.** (2014). The role of Six1 in the genesis of muscle cell and skeletal muscle development. *Int J Biol Sci* **10**, 983–9.

**Wurmser, M., Chaverot, N., Madani, R., Sakai, H., Negroni, E., Demignon, J., Saint-Pierre, B., Mouly, V., Amthor, H., Tapscott, S., et al.** (2020). SIX1 and SIX4 homeoproteins regulate PAX7+ progenitor cell properties during fetal epaxial myogenesis. *Dev Camb Engl* dev.185975.

**Xian, X., Gopal, S. and Couchman, J. R.** (2009). Syndecans as receptors and organizers of the extracellular matrix. *Cell Tissue Res* **339**, 31.

**Xu, M., Chen, X., Chen, D., Yu, B., Li, M., He, J. and Huang, Z.** (2019). Regulation of skeletal myogenesis by microRNAs. *J Cell Physiol* **235**, 87–104.

**Yahya, I., Böing, M., Pu, Q., Puchert, M., Oedemis, V., Engele, J., Brand-Saberi, B. and Morosan-Puopolo, G.** (2020). Cxcr4 and Sdf-1 are critically involved in the formation of facial and non-somitic neck muscles. *Sci Rep-uk* **10**, 5049.

**Yamaguchi, M., Watanabe, Y., Ohtani, T., Uezumi, A., Mikami, N., Nakamura, M., Sato, T., Ikawa, M., Hoshino, M., Tsuchida, K., et al.** (2015). Calcitonin Receptor Signaling Inhibits Muscle Stem Cells from Escaping the Quiescent State and the Niche. *Cell Reports* **13**, 302–314.

**Yaseen-Badarneh, W., Kraft-Sheleg, O., Zaffryar-Eilot, S., Melamed, S., Sun, C., Millay, D. P. and Hasson, P.** (2020). Fibroblast fusion to the muscle fiber regulates myotendinous junction formation. *Biorxiv* 2020.07.20.213199.

**Yin, H., Pasut, A., Soleimani, V. D., Bentzinger, C. F., Antoun, G., Thorn, S., Seale, P., Fernando, P., van IJcken, W., Grosveld, F., et al.** (2013). MicroRNA-133 Controls Brown Adipose Determination in Skeletal Muscle Satellite Cells by Targeting Prdm16. *Cell Metab* **17**, 210–224.

**Yoon, J. K., Olson, E. N., Arnold, H.-H. and Wold, B. J.** (1997). Different MRF4 Knockout Alleles Differentially Disrupt Myf-5 Expression: cis-Regulatory Interactions at the MRF4/Myf-5 Locus. *Dev Biol* **188**, 349–362.

**Yoshioka, K., Nagahisa, H., Miura, F., Araki, H., Kamei, Y., Kitajima, Y., Seko, D., Nogami, J., Tsuchiya, Y., Okazaki, N., et al.** (2020). Hoxa10 mediates positional memory to govern stem cell function in adult skeletal muscle. *Biorxiv* 2020.07.16.207654.

**Yutzey, K. E. and Konieczny, S. F.** (1992). Different E-box regulatory sequences are functionally distinct when placed within the context of the troponin I enhancer. *Nucleic Acids Res* **20**, 5105–5113.

**Yvernogeau, L., Auda-Boucher, G. and Fontaine-Perus, J.** (2011). Limb bud colonization by somite-derived angioblasts is a crucial step for myoblast emigration. *Development* **139**, 277–287.

**Zacharias, A. L., Lewandoski, M., Rudnicki, M. A. and Gage, P. J.** (2010). Pitx2 is an upstream activator of extraocular myogenesis and survival. *Dev Biol* **349**, 395–405.

**Zammit, P. S., Partridge, T. A. and Yablonka-Reuveni, Z.** (2006). The Skeletal Muscle Satellite Cell: The Stem Cell That Came in From the Cold. *J Histochem Cytochem* **54**, 1177–1191.

**Zhang, Z., Huynh, T. and Baldini, A.** (2006). Mesodermal expression of Tbx1 is necessary and sufficient for pharyngeal arch and cardiac outflow tract development. *Development (Cambridge, England)* **133**, 3587–95.

**Zhang, Q., Vashisht, A. A., O'Rourke, J., Corbel, S. Y., Moran, R., Romero, A., Miraglia, L., Zhang, J., Durrant, E., Schmedt, C., et al.** (2017). The microprotein Minion controls cell fusion and muscle formation. *Nat Commun* **8**, 15664.

**Zhou, Y., Gong, B. and Kaminski, H. J.** (2012). Genomic Profiling Reveals Pitx2 Controls Expression of Mature Extraocular Muscle Contraction–Related Genes. *Investigative Opthalmology Vis Sci* **53**, 1821.

**Ziermann, J. M., Diogo, R. and Noden, D. M.** (2018). Neural crest and the patterning of vertebrate craniofacial muscles. *Genesis (New York, N.Y. : 2000)* **56**, e23097.