Gdf11 signalling, Oct4 and the control of vertebrate trunk length

Molecular interactions regulating the process of axial extension during vertebrate development

Rita Aires

Dissertation presented to obtain the Ph.D degree in Developmental Biology

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,
July, 2016

UNIVERSIDADE NOVA DE LISBOA

Gdf11 signalling, Oct4 and the control of vertebrate trunk length

Molecular interactions regulating the process of axial extension during vertebrate development

Rita Aires

Dissertation presented to obtain the Ph.D degree in Developmental Biology

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Research work coordinated by: INSTITUTO GULBENKIAN DE CIÊNCIA

Oeiras, July,2016

UNIVERSIDADE NOVA DE LISBOA

"A sheep, a drum, and a snake fall off a cliff."

"Ba-Dumm-Tss!"

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

The work presented in this thesis was supported by FCT and FSE, grant SFRH/BD/51876/2012.

Apoio financeiro da FCT e do FSE no âmbito do Quadro comunitário de apoio, BD n.º SFRH/BD/51876/2012.

Table of Contents

List of Abbreviations

ActRIB/Acvr1b	Activin Receptor IB
ActRIIA/Acvr2a	Activin Receptor IIA
Alk4	Activin Receptor-like Kinase 4. Same as ActRIB/Acvr1b
AP	Antero-Posterior
AVE	Anterior Visceral Endoderm
BAC	Bacterial Artificial Chromosome
Bapx1	*Bagpipe Homeobox 1*
BMP	Bone Morphogenetic Protein
Bmpr1a	Bone Morphogenetic Protein Receptor, type 1a
Bp	Base pair
Cas9	CRISPR associated protein 9
Cchcr1	*Coiled-Coil Alpha-Helical Rod Protein 1*
Cdh1	E-cadherin
Cdx	*Caudal-type Homeobox*
Cer1	Cerebrus-like protein 1
ChIP	Chromatin Immunoprecipitation
CNH	Chordoneural Hinge
CR	Conserved Region
Cre^{ERT}	Tamoxifen inducible Cre recombinase
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DE	Distal Enhancer
DIG	Digoxigenin
Dkk1	Dickkopf homologue 1
DNA	Deoxyribonucleic acid
DVE	Distal Visceral Endoderm
E	Embryonic day
e.g.	*Exempli gratia*
ECM	Extracelular Matrix
EMT	Epithelial-to-Mesenchymal Transition

emVE	Embryonic Visceral Endoderm
Eomes	*Eomesodermin homolog (Xenopus laevis)*
EPC	Ectoplacental Cone
EpiSCs	Epiblast Stem Cells
ExEc	Extraembryonic Ectoderm
exVE	Extraembryonic Visceral endoderm
Fgf	Fibroblast Growth Factor
FgfR	Fibroblast Growth Factor Receptor
Foxa2	*Forkhead Factor 2*
Gata	GATA binding protein
Gdf	growth and differentiation factor
Hand2	*Heart and neural crest derivatives expressed transcript 2*
Hnf4	*Hepatocyte nuclear factor 4*
Hox	*Homeobox*
ICM	Inner Cell Mass
iPS	Induced Pluripotent Stem Cells
kB	Kilobase
Lefty1	*Left-right determinant factor 1*
Lsm2	*LSM2 Homolog, U6 Small Nuclear RNA And mRNA Degradation Associated*
Mesp	*Mesoderm posterior*
MET	Mesenchymal-to-Epithelial Transition
MMP-1	*Matrix Metalloproteinase-1*
Myf	*Myogenic factor*
Nkx1.2/Sax1	*NK1 Homeobox 2/ Spastic ataxia 1*
NMPs	Neuromesodermal progenitors
No	Notochord
Npdc1	*Neural proliferation, differentiation and control 1*
NSB	Node-Streak Border
NT	Neural Tube
OE	Olfactory Epithelium
Otx2	*Orthodenticle homolog 2 (Drosophila melanogaster)*
Pax	Paired box

PBS	Phosphate Buffered Saline
PBT	Phosphate Buffered Saline, containing 0.1% Tween-20
PCR	Polymerase Chain Reaction
PD	Proximo-Distal
PE	Proximal Enhancer
PEΔSBE	Deletion of Smad Binding Elements in *Oct4* Proximal Enhancer
PEΔTIE	Deletion of TGF-β Inhibitor Element in *Oct4* Proximal enhancer
PFA	Paraformaldheyde
PG	Paralogue Group
Pou5f1/Oct4	*POU domain, class 5, transcription factor 1/ Octamer-binding transcription factor 4*
PrE	Primitive Endoderm
PS	Primitive Streak
PSM	Presomitic Mesoderm
pSmad2/3	Phosphorylated Smad2/3
qPCR	Quantitative Polymerase Chain Reaction
RA	Retinoic Acid
Raldh2	*Retinaldheyde Dehydrogenase 2*
RAR	Retinoid Acid Receptor
RNA	Ribonucleic Acid
SBEs	Smad Binding Elements
Scx	*Scleraxis*
Shh	*Sonic hedgehog*
Snail	*Snail homolog 1 (Drosophila)*
Sox2	*SRY-related HMG box-containing 2*
SV40	Simian vacuolating virus 40
T	Thoracic
TBE	Tris Borate EDTA Buffer
TBM	Tail Bud Mesenchyme
TBST	Tris Buffered Saline, containing 0.1% Tween-20
Tbx6	*T-box 6*
Tcf19	*Transcription Factor 19*
TE	Trophectoderm

TE buffer	Tris EDTA buffer
TGCs	Trophectoderm Giant Cells
TGF-β	Transforming Growth Factor-β
TgfβRI	TGF-β Receptor I
TIE	TGF-β inhibiting element
TREEs	Tissue Regeneration Enhancer Elements
Vars	*Valyl-TRNA Synthetase*
VER	Ventral Epidermal Ridge
Wnt	Wingless-type MMTV Integration site family
β-gal	β –galactosidase

List of Figures

List of Tables

Acknowledgments

Agradecimentos

"No duck is an island"
- In *Tiny Toon Adventures (1990-1994)*

Há crianças que, quando forem grandes, querem ser médicos, enfermeiros ou veterinários. Ou mesmo bombeiros, gestores ou artistas. E depois há aquelas que descobrem a Física sob a forma de chá a ferver em queda livre e, mesmo assim, querem ter um doutoramento. Este, caro leitor, foi o meu caso; só não sabia em quê na altura. Até que um dia vi um embrião de galinha pela primeira vez e o resto é história. E, realmente, porque nenhum pato é uma ilha e eu fui incrivelmente sortuda por ter tido pessoas tão boas, importantes, e que admiro sempre à minha volta durante este tempo todo….

… É tempo de agradecer à Academia!

À malta do Mallo Lab, a começar por *Nosso Senhor O Chefe*. Obrigada, Moisés, por teres aceitado ser o meu orientador. Sem a tua ajuda, supervisão e, especialmente, enorme entusiasmo sobre tudo o que é Ciência, e Desenvolvimento em particular, nada disto teria sido possível. Mas, sobretudo, gostei mesmo de ter partilhado este projecto contigo: como companheiro de bancada, de escrita e como editor *extraordinaire*. Muitas e muitas gracias!!

Ao Arnon Dias Jurberg (aka Arnonzinho), que é parte integrante da *Dream Team* (Dream Team FTW©). Começámos isto juntos e, de certa forma, acabámos isto juntos. Eu podia falar aqui das importantes discussões científicas que tivemos,

mas prefiro recordar as vezes em que íamos à praia beber uma imperial ao final da tarde. Enquanto tínhamos importantes discussões científicas – sempre! À Ana Casaca (aka Casaquinha), o meu grande pilar no laboratório. Realmente não há palavras: sem ti, estes anos não teriam sido a mesma coisa. Obrigada não só por toda a ajuda (és a minha guru #2 para tudo o que é sequências e biologia molecular) mas também por, de vez em quando, *'converteres para PCR'.* E pelos nossas canções partilhadas, pelos cafés e jornal clubs de terça-feira. E falando nestes jornal clubs! Obrigada, Tiago Carneiro, por seres *O* membro honorário do laboratório. Nunca as conversas à hora do café/almoço/ lanche foram tão ecléticas – desde o Direito Português e a definição de personalidade jurídica, passando pela cozinha, até à vida numa certa casa na Venda do Pinheiro. E por compreenderes todas as minhas piadas parvas e por te deixares contaminar pela minha selecção musical do dia, sempre de qualidade!... duvidosa. Juntos, conseguimos anerdalhar o laboratório; a seguir: o Mundo! (Novo hastag: #anerdalharoMundo). À Ana Nóvoa, por ser a mais estupenda mestre de transgénicos e a pessoa ideal com quem ir a um workshop de ponto de cruz subversivo. A um Sábado à tarde. Numa tasca no Cais do Sodré *(True story, caro leitor!)*. À Luisinha, por estar connosco, por partilhar da loucura que é manter uma colónia de *Gdf11* (e infinitas variações) em heterozigotia e por toda a ajuda com a estatística. Ao André, por ter sido o meu primeiro aluno mais a sério, mesmo que só por três semanas. Mega super proud of you! À Irma, pelo bom companheirismo no laboratório e, especialmente, pela ajuda enorme nas genotipagens. E um obrigada muito especial aos velhos e novos membros do lab: Aybuke, Sofia, André... E à Inês Domingues, fonte inesgotável de alegria, entusiasmo e, de vez em quando, bolachinhas de limão com sementes de papoila.

To my thesis committee, Florence Janody e Domingos Henrique, for being available at all times when I needed you. Thank you so much!

Ao PhD Programme in Integrative Biomedical Sciences (PIBS) do IGC e aos seus directores: Thiago Carvalho e Élio Sucena. Em especial ao Élio, por organizares um primeiro ano do mestrado BED na FCUL que faz com que um doutoramento pareça (quase) um passeio no parque. E à Manuela Cordeiro, pela competência e ocasionais puxões de orelhas. Obrigada por tudo! To all members of PIBS 2011: Jarek, Rafal, Rômulo, Oz, Sandra, Sara, Marta, Ana Stankovic, Ana Martins, Inês, Irma, you guys rock! You were the best people to twilight with on the wee hours of some classes! Em especial à Sandra Tavares: a melhor DJ, companheira de AmeeGuS e de gabinete; mulher de gosto musical impecável, e de uma força, honestidade e integridade inabalável. Foi incrível! Ao Rómulo Areal, pelas sessões de discussão metafísicas nas tardes de segunda-feira – na realidade, pelas sessões de discussão metafísicas que aconteciam *de cada vez que nos encontrávamos* por aí. Nunca motivações tão reais de personagens fictícias foram dissecadas de forma tão apaixonante durante tanto tempo! Obrigada por me teres feito descobrir as alegrias do Kindle, do *The name of the wind* e pelas discussões de *Game of thrones* e *Black Sails*. To Ana Stankovic: the best gym buddy, the sharpest scientific mind, and the most amazing friend a gal can have. For telling me five years ago (and even now) that I didn't have to go through stuff alone. You possess the uncanny ability to make me spill the beans and man up in a way I didn't think it was possible. Ao IGC, e a toda a gente que faz deste instituto o melhor sítio do mundo para trabalhar. Às meninas da histologia, Joana Rodrigues e Marta Pinto, um agradecimento especial: não só são talentosas e competentes, como tornaram as minhas tardes passadas no criostato e no vibrótomo tão mais divertidas! Aos membros do meu laboratório de mestrado, o Organogenesis lab: Joaquín Rodríguez-Léon, Catarina Certal, Raquel Tomás, Nando, Diana Chapela. Mas especialmente à Joana Monteiro, que me ensinou as bases da vida no laboratório, desde a primeira *in situ* à última amputação de peixes.

À Isabel Guerreiro, a minha bestie for ever and ever and ever. Realmente, palavras não são mesmo suficientes para ti. Obrigada por toda a tua ajuda, no laboratório e fora dele. Por seres uma força motriz. E por seres a melhor condutora que uma copiloto pode desejar! E ao Tomás, por partilhar da minha nerdice a todos os níveis. It is really a great time to be alive and a nerd contigo! À minha querida Rita Félix, a pessoa com quem quero sempre ir viajar, comer gelados e ouvir música pimba. À Filipa Nunes, o ser mais determinado, focado, motivado e, de vez em quando, azarado à face da Terra. *Mi casa es su casa*, and I really mean it. À Catarina Dourado, parte crucial dos jantares de quinta-feira na Pizza Hut: és grande, moça!

À malta do secundário, por continuarmos a ser uma família. Em especial à Ari e à Ana, as criaturas mais livres e corajosas que tive a sorte de conhecer e de me aturarem. Ao Miguel Moutinho, o meu irmão de alma, par científico, e gémeo musical. Não imagino outra pessoa com quem conduzir sem destino só pelo prazer de ouvir a música, ir beber café ao aeroporto à 00h30, e fazer rotundas ao contrário à 1 da manhã *(outra true story, caro leitor)*.

E finalmente, aos meus pais. A culpa disto da Ciência, penso, foi do meu pai, que um dia me mostrou o porquê da Lua ter fases usando uma laranja + luz da janela como modelo. E por me ensinar que é possível fazer um portão automático usando o motor de uma betoneira. Obrigada, pai e mãe (ou, em bom anerdalhanço de geneticista, à minha F0), por tanto amor e compreensão, fé e por estarem sempre lá. E por nunca terem parado essa criança estranha que lia sem parar de brincar com os cogumelos venenosos no quintal e de fazer misturas estranhas com eles (isto olhando em restrospectiva, foi um bocado arriscado, hein?). À minha irmã, que será sempre das pessoas mais importantes da minha vida e meu exemplo de vida (sim, leste bem, oh minha bimba). Ao resto da família por fazer parte da minha vida, em especial à Gorete, Guidinha, Belinha e Madrinha *(que lindo que até rima)*. Vocês alimentam-me o corpo e a i'álma. À

Dani, que fez de mim quem sou quando me deu a ler *Mafalda*, *Calvin and Hobbes* e *Artemis Fowl*. E às minhas meninas, por terem mantido a minha sanidade mental na maior parte dos últimos quinze anos.

Estes cinco anos passaram depressa demais, e realmente não é justo. But time does fly when you're having fun, não é?

Abstract

Axial elongation is a conserved, fundamental mechanism in vertebrate development. This process consists in the gradual addition of tissue to the posterior-most part of the embryo, resulting in a progressive assembling of the embryonic body in a rostro-caudal sequence. Axial elongation relies on the activity of a dedicated population of cells located in the caudal part of the embryo, the axial progenitors. These progenitors are a highly dynamic, self-renewing, multipotent pool of cells, whose properties vary with the progression of the axial elongation. During the initial stages of development, they reside in the epiblast and primitive streak and include precursors for all three embryonic germ layers – ectoderm, mesoderm and endoderm. These will subsequently differentiate and interact amongst themselves and with other tissues in order to produce the different organs and body structures of the neck and trunk regions. Later in development, the caudal part of the embryo undergoes profound reorganization involving the disappearance of the epiblast and PS and the emergence of the tailbud. This process is associated with major changes in the axial progenitors as well. In particular, the progenitors for the lateral and intermediate mesoderm that are involved in the formation of the trunk organs undergo a process of terminal differentiation resulting in the formation of the hindlimbs and the organization of the embryonic cloaca. Concurrently, the remaining major subset of axial progenitors - the neuromesodermal progenitors (NMPs), which contribute to the generation of the neural tube, axial skeleton and associated muscles - relocate from the epiblast to the tailbud, where they continue with the remaining body axis elongation. Thus, the specific control of axial progenitor types, numbers, and balance between precursors and their derivatives will determine a particular species' final body length, as well as allocation of the post-cranial body into neck,

trunk and tail. As such, axial elongation is a key mechanism involved in the generation of the broad diversity of body shapes and sizes within vertebrates.

In the first part of this thesis, we performed an in-depth analysis of the tail defects observed in *Gdf11*$^{-/-}$ embryos, which were associated with the previously described delayed trunk to tail transition occurring in these mutants. We show that the tail abnormalities are already clear at mid-gestation stages shortly after this transition and seem to result from an expanded population of axial progenitors, combined with alterations in the tissue reorganization associated with the transition from primitive streak-driven to tailbud-dependent axial growth. In particular, morphological, lineage tracing and molecular analyses indicated that *Gdf11* mutant tails contain an excess of axial progenitors, most of them residing in an ectopic ventral structure composed of an epithelium enclosing a mass of mesenchymal cells. Importantly, we discovered that a subset of cells in this ectopic structure expressed *Oct4*, which suggested that the epithelial component of the ectopic ventral tissue likely represents incomplete resolution of the epiblast. We thus demonstrate that Gdf11 signalling is an integral part of the mechanisms involved in epiblast extinction and regulation of the axial progenitor pool size during the trunk to tail transition. These most likely involve functional interactions with Oct4, which probably include direct transcriptional regulation of this gene's expression by downstream components of the Gdf11 signalling pathway. As a result of all these complex processes, Gdf11 activity ensures a proper transition from trunk to tail-forming mechanisms, as well as an adequate, gradual axis termination.

The second part of this thesis aimed at the analysis of the contribution of *Oct4*'s ectopic expression for the *Gdf11* mutant phenotype. Using a transgenic approach, we proved that most axial phenotypes found in *Gdf11*$^{-/-}$ embryos derive from abnormally extended *Oct4* activity in axial progenitor regions during axis elongation. These studies also revealed that sustained Oct4 activity in these

regions was able to keep axial progenitors in a trunk-forming configuration, delaying the trunk to tail transition. Interestingly, this delay was associated with a concomitant caudal shift in the axial level of activation of posterior *Hox* genes, thus linking global distribution of the vertebrate body into trunk or tail regions with the patterning of the axial structures associated with these main areas.

These findings led us to analyze *Oct4* expression in snake embryos. Our research demonstrated that this gene likely suffered heterochronic shifts in its regulation, as its expression seems to be maintained for longer developmental periods in snakes relative to mouse embryos. Genomic analyses indicated that these temporal changes in *Oct4* expression seem to have originated from dramatic genomic rearrangements during the evolutionary trajectories of mammals, lizards and snakes, which could have altered the Oct4 regulatory landscape in these different vertebrate clades. Indeed, transgenic reporter analyses in mice identified the existence of potential regulatory sequences upstream of the *Oct4* gene in squamates that are not shared by its mammalian counterpart. Together, our observations suggest that these genomic and regulatory changes involving Oct4 might have been essential components of the mechanisms originating vertebrate body diversity and the emergence of the snake body plan.

Overall, we show a new role for the pluripotency factor *Oct4* as a key regulator of vertebrate trunk length, and establish the balance between Oct4 and Gdf11 activities as a major regulator of the body allocation into trunk and tail regions during axial extension in vertebrate embryos. Our results not only provide important insights into the different developmental events involved in the making of these two body regions, but also hint at possible ways of generating evolutionary novelty, ultimately contributing to the generation of the wide diversity of body plans observed among members of the vertebrate clade.

Sumário

A extensão axial ou alongamento do eixo corporal é um mecanismo fundamental e evolutivamente bem conservado no desenvolvimento embrionário dos vertebrados. Este processo consiste na adição progressiva de tecido à zona mais posterior do embrião, numa sucessão rostro-caudal que resulta da actividade de um conjunto especializado de células - os progenitores axiais. Os progenitores axiais são uma população celular altamente dinâmica, multipotente e com capacidade de auto-renovação, cujas propriedades se modificam à medida que a extensão axial ocorre. Nas fases iniciais do desenvolvimento embrionário, estas células residem no epiblasto e linha primitiva do embrião e incluem precursores para os três folhetos germinativos embrionários – ectoderme, mesoderme e endoderme. Estes percursores, por sua vez, experimentam processos de diferenciação e interacção com os tecidos circundantes, dando origem a todos os órgãos e restantes estruturas pertencentes às regiões do pescoço e tronco. Contudo, em fases mais tardias do desenvolvimento, a parte posterior do embrião sofre uma complexa reorganização que envolve o desaparecimento do epiblasto e linha primitiva e o início da formação da cauda. A população de progenitores axiais também é restruturada durante este processo. Enquanto os progenitores de mesoderme lateral e intermédia são submetidos a um processo de diferenciação terminal, que resulta na formação dos membros inferiores e na organização da cloaca do embrião, uma outra parte significativa da população – os progenitores neuro-mesodérmicos, que contribuem para a formação do tubo neural, esqueleto axial e músculos associados – é realojada no recém-formado botão da cauda e aí continua o processo de extensão axial do corpo até ao seu final. Deste modo, o controlo adequado do número e tipo de percursores axiais, assim como o balanço entre a quantidade de progenitores e

respectivas formas diferenciadas, é fundamental para a determinação do comprimento final do corpo de cada espécie e distribuição deste pelas regiões do pescoço, tronco e cauda. A extensão axial é, portanto, um processo chave na geração da grande diversidade de formas e tamanhos corporais observados entre vertebrados.

Na primeira parte desta tese foi realizada uma análise profunda dos defeitos observados nas caudas de embriões *Gdf11*$^{-/-}$. Estas anormalidades desde logo aparentaram estar relacionadas com o atraso na transição entre o tronco e a cauda já previamente descrito nestes mutantes. Neste estudo, mostrou-se não só que os defeitos observados nas caudas mutantes surgem logo após o período de transição entre o tronco e a cauda, como também que estas malformações parecem resultar simultaneamente de uma expansão da população de progenitores axiais e de alterações na reorganização dos tecidos associadas à passagem da extensão axial dependente da linha primitiva para a dependente dos progenitores residentes na cauda. As análises morfológicas e moleculares realizadas, assim como os dados obtidos em experiências de seguimento de linhagens celulares, indicaram que a cauda dos embriões mutantes para *Gdf11* apresenta, de facto, um excesso de progenitores axiais. A maioria destes aparenta estar concentrada numa estrutura ectópica, localizada na parte mais ventral da cauda, constituída por uma massa de células mesenquimatosas que se encontram envolvidas por um epitélio. O facto de algumas destas células inesperadamente também expressarem *Oct4* sugeriu que a componente epitelial desta estrutura ectópica poderia ter tido origem em fragmentos residuais do epiblasto. Assim, neste capítulo demonstrou-se que a sinalização Gdf11 é parte integrante do conjunto de mecanismos que regula tanto a extinção do epiblasto como a regulação do número de progenitores axiais durante a transição entre o tronco e a cauda, possivelmente através de interacções funcionais com Oct4. Estas

parecem incluir a regulação directa da expressão de *Oct4* durante a transição, provavelmente através de factores resultantes da via de sinalização Gdf11. Ao coordenar todos estes processos intrincados, a actividade de Gdf11 assegura assim uma correcta transição entre os mecanismos de formação do tronco e os de formação da cauda, tal como uma terminação progressiva do eixo corporal.

A segunda parte desta tese teve como objectivo inicial a análise da contribuição da expressão ectópica de *Oct4* para o fenótipo observado nos embriões *Gdf11*$^{-/-}$. Fazendo uso de uma abordagem experimental baseada em ratinhos transgénicos, verificou-se que grande parte das alterações axiais encontradas em embriões mutantes para *Gdf11* resulta da persistência anormal da expressão de *Oct4* em regiões que incluem progenitores axiais durante o processo de extensão axial. Estes ensaios também revelaram que a manutenção da actividade de Oct4 nestas regiões possibilita a retenção dos progenitores axiais numa configuração favorável à formação do tronco, provocando, dessa forma, um atraso na transição entre o tronco e a cauda. O atraso no início da transição revelou estar associado a uma concomitante posteriorização no nível axial da activação da expressão dos genes *Hox*, sugerindo que a distribuição do corpo nas regiões do tronco e cauda em vertebrados está estreitamente coordenada com a padronização das estruturas axiais associadas a estas zonas.

O facto destes embriões transgénicos apresentarem troncos longos, uma característica específica das cobras, instigou o estudo da expressão de *Oct4* em embriões deste grupo de organismos. Efectivamente, a expressão de *Oct4* em embriões de cobra pareceu ser mantida durante um maior período no desenvolvimento relativamente a embriões de ratinho, o que sugeriu que a regulação da expressão de *Oct4* possivelmente teria estado sujeita a alterações heterocrónicas durante a evolução da linhagem das cobras. Análises genómicas indicaram que as alterações na expressão deste gene podem ter tido origem em rearranjos genómicos dramáticos nas regiões a 5' de *Oct4* durante as trajectórias

evolutivas de mamíferos, lagartos, e cobras. Estas modificações terão resultado numa alteração do ambiente genómico responsável pela regulação da expressão de *Oct4* nos diferentes grupos de vertebrados. De facto, a análise de embriões de ratinho transgénicos contendo genes repórteres sob a influência destas potenciais sequências regulatórias de cobra, permitiu identificar elementos reguladores de *Oct4* partilhados exclusivamente por cobras e lagartos. No geral, estas observações sugerem que as alterações genómicas nas regiões contíguas a *Oct4*, nas quais provavelmente se inseriam as zonas regulatórias deste gene, podem ter sido elementos essenciais nos mecanismos que originaram a diversidade das formas corporais em vertebrados, particularmente no aparecimento do plano corporal das cobras.

Em suma, os resultados descritos nesta tese atribuem uma nova função para o factor de pluripotência *Oct4*, que actua como um gene chave na regulação do tamanho do tronco em vertebrados. Para além disso, estabeleceu-se que as actividades de Oct4 e de Gdf11 constituem os principais componentes envolvidos no controlo da distribuição do corpo em tronco e cauda durante a extensão axial em vertebrados. As descobertas aqui descritas, para além de contribuem para uma maior compreensão dos fenómenos envolvidos na formação destas duas regiões corporais, também sugerem possíveis formas de gerar novidades evolutivas capazes de assegurar a vasta diversidade de planos corporais observados em vertebrados.

Chapter I:
Introduction

"I'm a leaf on the wind – watch how I soar."

- In *Serenity* (2005)

Development of a whole multicellular complex organism from a single cell is not only an evolutionary triumph, but also the most daunting and formidable of tasks. The organism's entire body plan has to be laid down in a series of intricate and interconnected events that comprise various levels of organization, from intracellular processes to vast morphogenetic tissue movements. This means that the embryo's early symmetries must be gradually broken and that most of the initial cell potency needs to be progressively surrendered so that the body can increase in complexity and, ultimately, achieve its final form. Yet, the minutest of mistakes can be either fatal or represent a huge evolutionary opportunity.

This introductory chapter tells the story of this progression, mainly focusing in the mouse embryo. It is intended as a "crash course" on early mouse development, describing its most important events and the major players that take part during this process.

I.1 The mouse as a model organism for development

From all model organisms, the mouse stands out as one of the most popular. As a fellow mammalian vertebrate, the mouse shares many physiological and pathological features with humans (Rosenthal and Brown, 2007). Also, mouse and human genomes show a high degree of evolutionary conservation, even if the two species have diverged more than 96 million years ago (Nei et al., 2001; Nguyen and Xu, 2008).

Small size, ready availability, easy maintenance and husbandry, as well as docility are some of the advantages that mice have compared with other mammalian models (Nguyen and Xu, 2008; Rosenthal and Brown, 2007; Wolpert et al., 1998). For developmental biologists in particular, their relatively short generation time and high fertility are also very convenient characteristics. Even considering its limitations regarding embryo accessibility for grafting and other direct surgical manipulations during embryonic development, the mouse presents

unparalleled opportunities for research. In fact, the sheer amount of available phenotypic data, genomic resources and genetic tools developed for more than 80 years – many of them used in the present work – makes the mouse one of the most powerful model organisms in the pursuing of fundamental questions in mammalian biology and disease (Schofield et al., 2012).

I.2 The early mouse embryo and the first cell fate decisions

Mouse development can be broadly divided into pre- and postimplantation (Lawson and Wilson, 2016; Wilson and Lawson, 2016; Wolpert et al., 1998). Preimplantation development takes approximately four and a half days and comprises all stages between oocyte fertilization and embryonic implantation in the uterine wall. Preimplantation mouse development is highly regulative, which means that the embryo can adapt and compensate for perturbations either in number and/or position of cells (Ziomek et al., 1982). This shows that cells in the early embryo have a high developmental potential and are still quite flexible in terms of cell fate. Yet, cell labelling and lineage tracing experiments show that there might be differences in blastomere developmental properties and a bias towards particular fates as early as the 2-cell stage (Piotrowska-Nitsche et al., 2005; Tabansky et al., 2013).

I.2.1 First cell fate decision: Trophectoderm *vs* Inner Cell Mass

After fertilization the zygote undergoes successive rounds of cell division without growth. This process of cleavage generates small cells with little cytoplasm – the blastomeres (Wolpert et al., 1998). As the embryo reaches the 8-cell stage, it undergoes a process of compaction and becomes a morula (Fig I.1) (Ducibella and Anderson, 1975; Wolpert et al., 1998). Compaction involves changes in blastomere shape, assembly of intercellular adhesion complexes and development of strong apical-basal polarization (Johnson and Ziomek, 1981b).

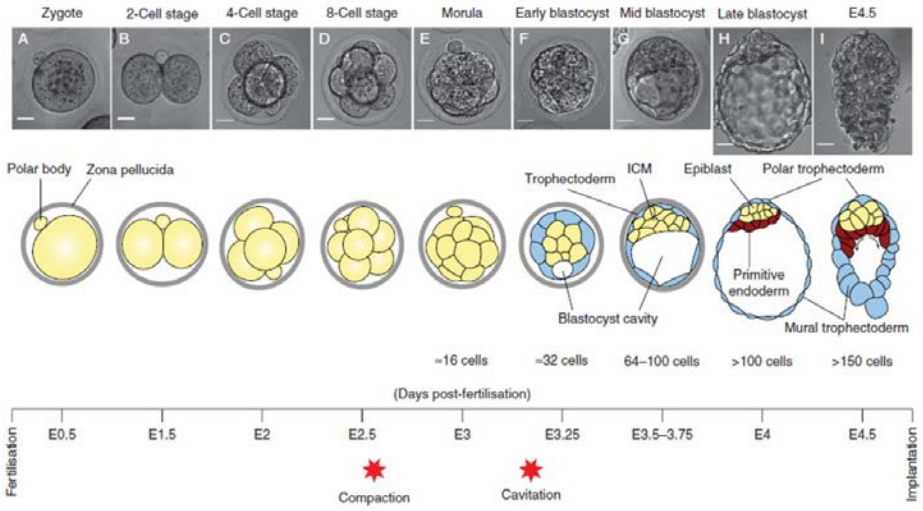

Fig I.1 Preimplantation mouse development and the first cell fate decisions. At the 8-cell stage the embryo undergoes compaction, which generates differences both in position and intercellular contacts among morula cells. This triggers the first cell fate decision whereby cells in the periphery become trophectoderm (TE), whereas internal cells become inner cell mass (ICM). The second lineage decision occurs within cells of the ICM after the blastocyst cavity (or blastocoel) is formed by cavitation. Asymmetries in cell division generate asymmetries in cell signalling, which ultimately result in the acquisition of primitive endoderm (PrE) or epiblast fate. Arrows in the E4.5 scheme denote PrE migration over TE cells. Timeline indicates time elapsed since fertilization in embryonic days (E). Adapted from Saiz and Plusa, 2013.

This polarization allocates specific proteins to apical and basal domains, which will eventually give rise to asymmetries after cell division (reviewed in Bedzhov et al., 2014; Saiz and Plusa, 2013).

The first clear cell fate decision in the mouse embryo, which occurs around embryonic day (E)2.5, has its origins precisely on these subtle molecular differences. In particular, two rounds of asymmetric cell divisions generate an embryo composed of small, non-polarized, inside cells – the Inner Cell Mass (ICM) – that are enclosed within larger, highly polarized, outer cells that make up the trophectoderm (TE) (Johnson and Ziomek, 1981a; Wilson and Lawson, 2016; Wolpert et al., 1998). The TE constitutes the first epithelium formed during embryo development. As such, its cells are highly polarized and strongly connected by intercellular junctions that provide extensive cell-to-cell

communication and cohesion. In contrast, the ICM is mainly composed of cells that have lost polarity and which are exposed to uniform cell-to-cell contacts. By this time (around E3.5) the embryo starts to develop a fluid-filled cavity, the blastocoel, which causes the TE to expand and the ICM to become confined to one side of this vesicle (Wolpert et al., 1998). The resulting embryo is the blastocyst (Fig I.1). TE cells up-regulate *caudal-type homeobox-2* (*Cdx2*) (Beck et al., 1995; Dietrich and Hiiragi, 2007; Niwa et al., 2005; Ralston and Rossant, 2008; Strumpf et al., 2005). Even though not necessary for TE specification *per se*, *Cdx2* is crucial for epithelial integrity and tissue maturation, in such a way that *Cdx2* mutant embryos fail to implant (Strumpf et al., 2005). On the other hand, ICM cells express *Oct4* - also known as *Pou5f1* and a member of the POU domain family of octamer-binding transcription factors - which is absolutely required for the formation and maintenance of the ICM (Nichols et al., 1998; Ryan and Rosenfeld, 1997). Interestingly, both *Cdx2* and *Oct4* are expressed in every blastomere at early cleavage stages. However, as development progresses, these genes become gradually restricted to the TE and ICM, respectively, in part as a consequence of mutual repressive activities (Dietrich and Hiiragi, 2007; Downs, 2008; Niwa et al., 2005; Palmieri et al., 1994; Schöler et al., 1989). Thus, TE and ICM seem to be mutually exclusive identities characterized by complementary *Cdx2* and *Oct4* expression (Niwa et al., 2000; Niwa et al., 2005). In fact, in the absence of *Oct4* all morula cells are diverted into TE fate and lose pluripotency (Nichols et al., 1998). *Oct4* later becomes a crucial factor in the specification and maintenance of the epiblast and primordial germ cell survival (Kehler et al., 2004; Nichols et al., 1998; Niwa et al., 2000).

I.2.2 Second cell fate decision: Primitive Endoderm *vs* Epiblast

The second lineage fate decision occurs within the ICM at approximately E3.5. It results in the formation of the epiblast, which will give rise to the embryo proper,

and the primitive endoderm (PrE), an epithelium that covers the epiblast and separates it from the blastocyst cavity (Fig I.1 and Fig I.2A). Once again, subtle heterogeneities between cells seem to trigger different genetic programs. Biases in the internalization of cells into the deep ICM layers during asymmetric divisions generate variation in expression levels of *Fibroblast Growth Factor 4* (*Fgf4*) and its receptor, *Fibroblast Growth Factor Receptor 2* (*FgfR2*), among ICM cells (Krupa et al., 2014; Morris et al., 2013). Cells that express high levels of *Fgf4* down-regulate *FgfR2* and will become *Nanog*-expressing epiblast cells, whereas cells expressing high levels of *FgfR2* down-regulate *Fgf4*, activate *Gata6* and other PrE markers and enter a PrE cell fate (Chambers et al., 2003; Chazaud et al., 2006; Guo et al., 2010; Mitsui et al., 2003; Yamanaka et al., 2010). This results in a "salt-and-pepper" pattern of epiblast and PrE progenitors that will be subsequently sorted into their final positions by mechanisms such as active cell migration, differential adhesion and selective apoptosis (Chazaud et al., 2006; Meilhac et al., 2009; Plusa et al., 2008). The epiblast and PrE cell lineages will then become stabilized through cross-regulatory processes. Nanog directly represses *Gata6* and promotes Fgf4 secretion that, in turn, activates signalling through FgfR2 in adjacent cells and stabilizes *Gata6* expression to keep their PrE identity (Frankenberg et al., 2011; Schrode et al., 2014; Singh et al., 2007). Interestingly, *Oct4, Cdx2*, *Nanog* and *Gata6* are all coexpressed in every blastomere until the 64-cell blastocyst stage (Guo et al., 2010). This phenomenon might explain why preimplantation mouse development is so characteristically regulative, since strict lineage specifications are delayed until implantation and cells maintain a high potential to change fate due the presence of all these fundamental lineage markers.

Another important epiblast marker is the SRY-related HMG box-containing transcription factor *Sox2*. This gene is up-regulated specifically in epiblast cells and is crucial for epiblast maintenance. In fact, *Sox2* mutant embryos completely fail

to develop while still being able to implant and form extraembryonic structures (Avilion et al., 2003). *Nanog*, *Oct4* and *Sox2* therefore make up the core of the pluripotency network in the epiblast as the absence of any of these factors is enough to disrupt its formation and maintenance. *Oct4* and *Sox2* are also two of the four factors that compose the genetic cocktail able to reprogram somatic cells back into an embryonic stem cell-like state, generating the so-called induced pluripotent stem cells (iPS) (Takahashi and Yamanaka, 2006).

As a result of all these events, at the end of preimplantation development the embryo consists of three different cell lineages: epiblast, TE and PrE. The embryo derives exclusively from epiblast cells, whereas the TE and PrE will give rise to extra-embryonic structures such as the placenta and the yolk sac. These lineages are not only crucial for intra-uterine development, but also play important roles as signalling centres essential for patterning of the embryo proper. By E4.5 the embryo is ready to implant.

I.3 Postimplantation development and early axis specification

I.3.1 The egg cylinder stage

The first terminally differentiated cell type arises upon implantation, when the mural TE cells – TE cells surrounding the blastocoel – go through rounds of endo-reduplication and differentiate into TE giant cells (TGCs). These cells invade the uterine tissues and induce extensive vasculature remodeling and angiogenesis, important to mediate the embryo´s nutrient uptake, waste removal and gas exchanges (Bedzhov et al., 2014). On the other hand, TE cells closer to the epiblast (designated polar TE) proliferate due to the presence of TE progenitors and generate the ectoplacental cone (EPC) and the extraembryonic ectoderm (ExEc) that will become part of the placenta (Bedzhov et al., 2014; Wolpert et al., 1998). The PrE also expands, diversifying into parietal endoderm, embryonic visceral

endoderm (emVE) and extraembryonic visceral endoderm (exVE) (Fig I.2A) (Rivera-Pérez and Hadjantonakis, 2015; Wolpert et al., 1998).

Rapid cell division with growth by both epiblast and extraembryonary tissues causes the embryo to expand, elongating into the blastocyst cavity. At the same time, a second lumen – the proamniotic cavity – is created amidst the epiblast by a process of hollowing. From a ball of non-polarized cells, the epiblast turns into a highly organized rosette-like structure by the establishment of a strong apical-basal polarization, epithelialization and consequent changes in cell shape. This self-organization of epiblast cells seems to be mediated by deposition of extracellular matrix components (ECM) and by activation of β1-integrin receptors (Bedzhov and Zernicka-Goetz, 2014). A similar process is thought to occur at the level of ExEc so that, ultimately, epiblast and ExEc become two contiguous, although distinct, epithelia (Arnold and Robertson, 2009; Bedzhov and Zernicka-Goetz, 2014; Rivera-Pérez and Hadjantonakis, 2015). The embryo thus enters the egg cylinder stage, acquiring a hollow, cylindrical conformation, with a proximal-distal (PD) axis. The site of connection with the uterine tissue becomes the proximal pole, whereas the tip of the cup shaped epiblast represents the distal-most part of the axis (Bedzhov et al., 2014; Lawson and Wilson, 2016).

I.3.2 Proximo-distal axis and formation of the DVE

The establishment of the PD axis in the conceptus (embryo and supporting structures) at E5.0 is the first step towards the generation of the antero-posterior (AP) embryonic axis. The key to a correct PD patterning in the egg cylinder is the establishment of a robust PD gradient of Nodal (Brennan et al., 2001; Kumar et al., 2014). Nodal is a member of the transforming growth factor beta (TGF-β) superfamily of growth factors and, as such, requires two types of serine-threonine kinase receptors – normally known as type I and type II receptors - for signalling transduction. Nodal binds to the type I receptor Alk4 (ActRIB/Acvr1b), which

promotes recruitment of the type II receptors ActRII (ActRIIA/Acvr2a) or ActIIB (Acvr2b) that, in turn, trans-phosphorylate and thus fully activate the type I receptor (Kumar et al., 2001; Shen, 2007). Activated Alk4 phosphorylates cytoplasmatic Smad2 and/or Smad3 that then form a complex with Smad4, enters the nucleus and ultimately regulates expression of target genes (Morikawa et al., 2013). Nodal activity also requires the presence of EGF-CFC co-receptors such as Cripto that confer binding specificity for Alk4 (Yeo and Whitman, 2001). Nodal is initially expressed as an immature ligand (proNodal) by epiblast cells (Conlon et al., 1994; Varlet et al., 1997). ProNodal activates expression of *Furin* and *PACE4* convertases in ExEc cells, which will cleave the propeptide, thus producing mature Nodal in the proximal epiblast (Beck et al., 2002; Ben-Haim et al., 2006). Mature Nodal is then able to activate a positive autoregulatory loop, stimulating its own expression and therefore creating a high Nodal concentration in the proximal epiblast region. Nodal is also capable of inducing Bmp4 production by ExEc cells, which enhances *Wnt3* expression in the proximal egg cylinder that, in turn, promotes Nodal production as well (Ben-Haim et al., 2006; Winnier et al., 1995). Both mature Nodal and Bmp4 can likewise induce expression of the co-receptor Cripto, which is essential for Nodal signalling transduction and, thus, for a correct PD and AP specification (Beck et al., 2002; Ding et al., 1998). Mature Nodal protein is essential for correct visceral endoderm specification as well, since it represses ExVE genes (*Gata4*, *Hnf4*, etc) while maintaining expression of emVE related genes like *Fgf8*, *Fgf5*, *Bmp2*, *Otx2* and *Foxa2* (Mesnard et al., 2006).

Nodal activity is also involved in promoting expression of its inhibitors in the distal-most emVE, which becomes a local epithelial thickening known as the Distal Visceral Endoderm (DVE) (Fig I.2B) (Brennan et al., 2001; Meno et al., 1999; Rivera-Pérez et al., 2003; Takaoka et al., 2006). The DVE constitutes an important signalling center in the embryo that secretes inhibitors of the Nodal [like Cerebrus-like protein 1 (Cer1) and Left-right determinant factor 1 (Lefty1)] and

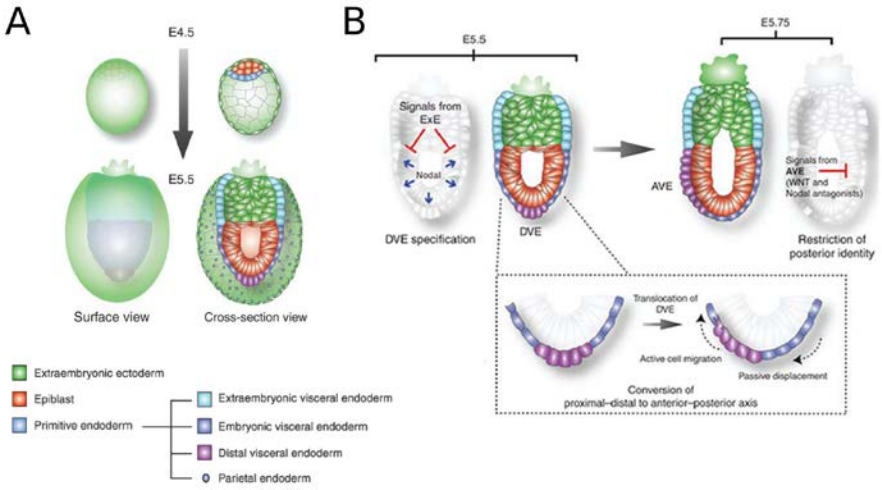

Fig I.2 Proximo-Distal (PD) and Antero-Posterior (AP) axis formation. A. Schematic representation of cell lineages and their spatial relationship in an implanting blastocyst (E4.5) and early egg cylinder stage embryo (E5.5). **B.** AP axis formation. The egg cylinder's radial symmetry is broken when cells belonging to the distal visceral endoderm (DVE) start migrating towards the embryo's future anterior pole. Establishment of the anterior visceral endoderm (AVE) results in a repositioning of the Nodal and Wnt antagonists source, confining the activity of these pathways to the embryo's posterior side. Adapted from Rivera-Pérez and Hadjantonakis, 2015.

Wnt [such as Dickkopf homologue 1 (Dkk1)] signalling pathways (Pfister et al., 2007; Takaoka et al., 2006). The overall result of all these processes is generation of high Nodal/Wnt activities in the proximal epiblast (Richardson et al., 2006; Rodriguez et al., 2005), and their attenuation at the distal epiblast, which eventually lead to correct patterning of the embryo proper.

I.3.3 Antero-posterior axis and formation of the AVE

At E5.75, soon after the PD axis is established, the egg cylinder's radial symmetry is broken when DVE cells start migrating towards the prospective anterior side of the embryo. This process leads to the formation of the Anterior Visceral Endoderm (AVE) near the now anterior ExEc/epiblast boundary (Fig I.2B). The AVE is a distinct visceral endoderm population composed by cells with a tall columnar morphology that express the Nodal and Wnt signalling antagonists *Lefty1*, *Cer1*

and *Dkk1*, in addition to a number of transcription factors (Pfister et al., 2007; Rivera-Pérez et al., 2003). Similarly to the DVE, the AVE's primary role is to maintain a Nodal and Wnt signalling gradient throughout the tissues, this time as an AP gradient. This structure also constitutes an important source of signals for the correct specification of anterior structures while inhibiting the activity of posterior genes in the anterior epiblast (Kimura et al., 2000; Perea-Gomez et al., 2001; Thomas and Beddington, 1996; Yamamoto et al., 2004).

Despite sharing many morphological and molecular features, clonal analysis studies have shown that the AVE does not entirely derive from DVE cells, but that a significant part of it is formed *de novo* from emVE cells that acquire AVE expression markers (Srinivas et al., 2004; Takaoka et al., 2011; Torres-Padilla et al., 2007). Yet, ablation experiments indicate that the DVE is indeed crucial for proper AVE positioning. In fact, DVE displacement seems to be the triggering event leading to the overall movement of emVE towards the future anterior pole of the embryo, which will include the cells that will be part of the future AVE (Miura and Mishina, 2007; Takaoka et al., 2011). This means that the AVE is not static or homogeneous and that its cellular composition varies as it moves anteriorly.

The repositioning of Nodal and Wnt antagonist sources, following the establishment of the AVE, results in a conversion of the PD axis into the AP axis. Hence, posteriorizing factors stay confined to the pole opposite to the AVE, which will play an essential role in gastrulation. At this stage (~E6.0) the cross section of the egg cylinder is not perfectly circular but ellipsoid instead. When first specified, the AP axis is aligned with the short axis of the oval embryo. However, within few hours after AVE settlement, AP axis is shifted towards the long axis by progressive tissue remodelling (Mesnard et al., 2004; Perea-Gomez et al., 2004). This process seems to be dependent on Wnt3 and Fgf8 activity as mutant embryos for these two factors fail to undergo reshaping (Barrow et al., 2007; Guo and Li, 2007). Yet, gastrulation can be induced even in the absence of reshaping, indicating that axis

realignment might serve as a way to maximize the distance between both poles, thus minimizing possible interferences amidst anterior and posterior signals.

I.4 Gastrulation and embryonic germ layer formation

Once the AP axis is settled, the embryo is ready to go through gastrulation. This is a process whereby concerted cell proliferation, migration, differentiation and changes in cell shape and adhesion properties, among other morphogenetic events, creates the organism's basic body plan (Wolpert et al., 1998). As such, gastrulation encompasses dramatic changes in the global structure of the embryo. The most important – and striking – rearrangement is the conversion of the two-layered embryo into a more complex structure composed by three embryonic germ layers: ectoderm, mesoderm and definitive endoderm (Wolpert et al., 1998). Each of these layers gives rise to specific types of tissues. The ectoderm or the "outer layer" will generate the animal's epidermis and nervous system, whereas the mesoderm (or the "middle layer") will provide the skeleto-muscular system, connective tissues and contribute to different extents to the formation of internal organs such as the heart, the kidneys or the muscular layers of the intestine. Finally, the inner-most tissue, the endoderm, will produce the epithelial lining of the gut and respiratory system, besides playing a major part in the formation of digestive organs such as the liver and the pancreas.

I.4.1 Primitive streak specification and positioning

In the mouse, gastrulation begins around E6.5 with the formation of a transient, specialized structure in the proximal posterior epiblast designated as the Primitive Streak (PS). Once formed, the PS is a stable structure and is present throughout gastrulation. However, shortly after it reaches its maximum length at the distal tip of the embryo, it begins to regress progressively, disappearing completely by early organogenesis (around E9.5) (Fig I.3A).

The PS begins as a local epithelial deformation and becomes a cellular discontinuity generated by the progressive initiation of an epithelial-to-mesenchymal transition (EMT) in epiblast cells (Perea-Gomez et al., 2004; Williams et al., 2012; Wolpert et al., 1998). Ingression through the PS, followed by a change from an epithelial to a mesenchymal state, allows cells to insert themselves and migrate between the epiblast and the visceral endoderm, or to intercalate among the visceral endoderm becoming epithelial again (Acloque et al., 2009). This way, transiting cells can become mesoderm or definitive endoderm, respectively, whereas cells that do not ingress through the PS and remain in the epiblast will be part of the neurectoderm and surface ectoderm (Lawson et al., 1991) (Fig I.3A).

Signalling activity from the AVE restricts the posteriorizing activity of the Nodal, BMP and Wnt pathways to the proximal posterior epiblast, eventually leading to PS formation. In fact, Nodal inhibitors are essential for a correct PS positioning since embryos lacking both *Cer1* and *Lefty1* develop multiple, ectopic, primitive streaks throughout the epiblast (Ben-Haim et al., 2006; Brennan et al., 2001; Mishina et al., 1995; Perea-Gomez et al., 2002; Rodriguez et al., 2005). Proper PS formation requires high levels of Smad2- and Smad3-dependent Nodal activity (Ben-Haim et al., 2006; Chu et al., 2005; Conlon et al., 1994; Ding et al., 1998; Dunn et al., 2004; Vincent et al., 2003), as well as ExEc-secreted Bmp4 acting on its receptor Bmpr1a in epiblast cells (Mishina et al., 1995; Winnier et al., 1995). Besides Nodal and Bmp4, several studies have shown that Wnt3 signalling through β-catenin and its co-receptors LRP5 and LRP6 is also indispensable for PS initiation and maintenance (Huelsken et al., 2000; Kelly et al., 2004; Liu et al., 1999; Mohamed et al., 2004; Tortelote et al., 2013; Yoon et al., 2015). In fact, absence of any of the aforementioned factors results in severe PS abnormalities and deficient or non-existing mesoderm production. The resulting PS is thus characterized by a gene expression profile composed of genes belonging or

responding to these signalling pathways, including *Nodal*, *Wnt3*, *Wnt3a, Axin2*, *Lefty2*, *Fgf8* and *T/Brachyury*, and various others like *Snail* and *Sp5* (Ben-Haim et al., 2006; Pfister et al., 2007; Robb and Tam, 2004; Tam and Loebel, 2007; Tortelote et al., 2013).

I.4.2 Primitive Streak morphogenesis and mechanisms of ingression

Contrary to what has been described in other vertebrates like chicken or rabbit, PS morphogenesis in the mouse does not involve large-scale cellular movements to position PS precursor cells or convergence and extension mechanisms for its elongation (Halacheva et al., 2011; Lawson and Schoenwolf, 2001; Viebahn et al., 2002; Voiculescu et al., 2007). Instead, PS morphogenesis in the mouse seems to be a positional phenomenon that occurs by *in situ* EMT in three consecutive phases: basement membrane loss, cell ingression and streak elongation (Williams et al., 2012). Basement membrane loss is an important first step required to break the major physical barrier to cell movement represented by extracellular matrix (ECM) constituents. Localized disaggregation of this structure most likely occurs by down-regulation of genes such as laminin and type IV collagen in PS-forming cells or by active degradation of its components. The next step involves major changes in epiblast cells, so they can leave the epithelial sheet and ingress through the PS. This requires not only cell shape changes (like apical constriction), but also the disassembling of intercellular adhesion complexes and down-regulation of epithelial junctional and polarity proteins (such as E-cadherin, Occludin and β-catenin) so that cells can delaminate from the epithelial sheet and migrate away. Finally, cells in the vicinity of the PS are gradually recruited in a posterior to anterior direction, resulting in an elongation of the PS all the way to the distal tip of the embryo that is complete by E7.0 (Williams et al., 2012). The anterior-most tip of the PS is occupied by the node, which is populated by a specialized group of ciliated, columnar cells that function as an important

signalling centre for patterning (Balmer et al., 2016; Lawson and Wilson, 2016; Yamanaka et al., 2007).

Gastrulation is a continuous process that occurs simultaneously with PS formation. It encompasses the inactivation of epiblast genes, such as *Oct4* and *Sox2*, and the acquisition of mesoderm and endoderm-specific factors (Pfister et al., 2007). The gradual and sequential loss of cells from the epiblast sheet as cells ingress through the PS is compensated by the high proliferation rates observed during these stages. In the end, the overall net result is the generation of a force that passively pulls lateral epiblast cells towards the PS, while its overall epithelial integrity is maintained (Williams et al., 2012).

Several genes and signalling pathways have been identified as having an important role in ingression and EMT through the PS (reviewed in Acloque et al., 2009). FGF signalling is among the best-studied cases in mouse gastrulation and seems to be particularly relevant in cell movement and mesoderm layer formation. In the absence of *Fgf8* or its receptor, *FgfR1*, cells fail to migrate away from the PS as they are unable to down-regulate E-cadherin (also known as *Cdh1*) (Ciruna and Rossant, 2001; Ciruna et al., 1997; Sun et al., 1999). E-cadherin is an integral part of intercellular adherent junctions and is essential to maintain epithelial integrity; thus, its down-regulation is absolutely crucial for EMT. In fact, perturbing E-cadherin function with a blocking antibody is sufficient to trigger the conversion of epiblast into mesenchymal cells (Burdsal et al., 1993). E-cadherin regulation by FGF signalling is mediated by the zinc-finger transcription factor Snail, which has the ability to bind directly to the *Cdh1* gene promoter region and repress its expression (Batlle et al., 2000; Cano et al., 2000; Ciruna and Rossant, 2001). *Snail* mutant embryos are actually capable of forming a "mesoderm" expressing the right set of mesodermal markers; however these cells retain epithelial characteristics such as apical-basal polarity and adherent junctions, as well as maintaining a robust E-cadherin expression (Carver et al., 2001). The T-box

transcription factor *Eomesodermin* (*Eomes*) appears to have a role in E-cadherin regulation as well. Conditional abrogation of *Eomes* in epiblast cells leads to a failure in efficient down-regulation of both E-cadherin transcripts and protein. These embryos also have a thickened PS that likely result from impaired cell delamination and mesoderm migration. However, E-cadherin regulation by *Eomes* seems to be indirect (Arnold et al., 2008). Other transcription factors like *Mesp1* and *Mesp2* have likewise been shown to be essential for nascent mesoderm migration, but whether or not these proteins have any influence upon E-cadherin regulation is unknown (Kitajima et al., 2000; Saga et al., 1999). Besides *Snail*, T-box transcription factors *Tbx6* and *T/Brachyury* were found to be important for mesoderm formation and are regulated by both FGF and Wnt signalling (Chapman and Papaioannou, 1998; Ciruna and Rossant, 2001; Galceran et al., 2001; Wilson et al., 1995; Yamaguchi et al., 1999). In particular, *Tbx6* seems to repress *Sox2* expression, ensuring the complete suppression of the neural transcription program in mesoderm-fated cells (Chapman and Papaioannou, 1998; Takemoto et al., 2011). *Brachyury/T*, in turn, is absolutely required for mesoderm specification and notochord morphogenesis (Herrmann, 1991; Stott et al., 1993; Wilson and Beddington, 1997).

I.4.3 Lineage allocation during gastrulation

During gastrulation cells are continuously recruited from the epiblast to undergo ingression and EMT through the PS (Williams et al., 2012). Fate mapping studies have revealed that not only all epiblast cells are competent to be part of any of the three germ layers, but that distinct mesodermal lineages are specified depending on the time and site of ingression through the PS (Fig I.3B) (Kinder et al., 1999; Lawson et al., 1991; Smith et al., 1994; Wymeersch et al., 2016).

At early stages of gastrulation or early streak stages, the most caudal part of the PS will give rise to the extraembryonic mesoderm. This tissue will provide the

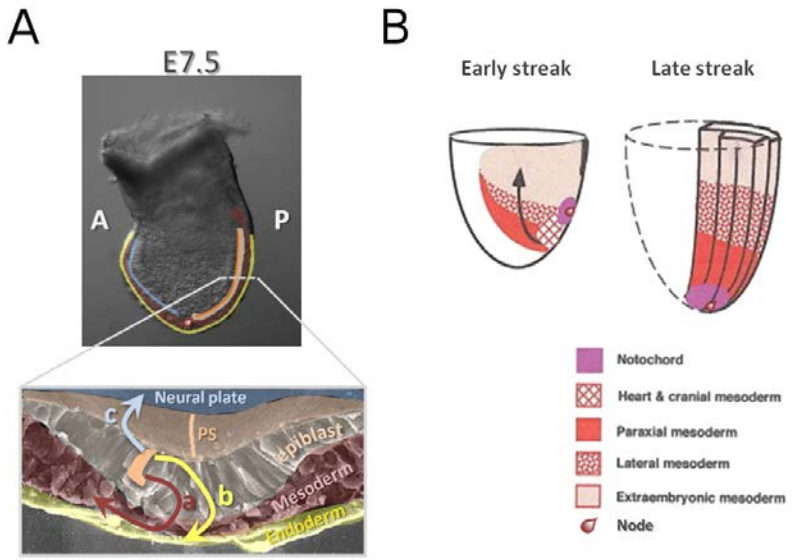

Fig I.3 Ingression of epiblast cells through the primitive streak (PS) and mesodermal lineage allocation during mouse gastrulation. A. Gastrulation begins with the formation of a specialized transient structure, the primitive streak (PS), located in the embryo's posterior side. Epiblast cells (orange) approach the PS region and undergo an epithelial to mesenchymal transition (EMT). These cells can then migrate away from the PS and insert themselves between the epiblast layer and the visceral endoderm (VE), becoming mesoderm (a). Alternatively, ingressing cells can become epithelial once more by a process of mesenchymal to epithelial transistion (MET) and intercalate between cells of the VE, becoming definitive endoderm (b). Cells that do not ingress become part of the ectoderm, giving rise to the neuroctoderm and surface ectoderm (c). **B.** Different mesodermal lineages are specified according to their time and place of ingression within the PS. Extraembryonic mesoderm ingresses through the most posterior part of the PS, whereas the cardiac mesoderm and head mesoderm enter the PS through increasingly more anterior regions. Finally, the anterior-most end of the PS and the node will generate mainly midline axial mesendodermal tissues. By the late streak stage, heart and head mesoderm have ingressed completely, whereas the remaining lateral and paraxial mesoderm continue to be produced. B is adapted from Tam and Behringer, 1997.

mesodermal component of extraembryonic supportive tissues (such as the chorion and visceral yolk sac) and is the origin of the first hematopoietic tissue, the blood islands. Intermediate levels of the PS will be the ingression site of cardiac mesoderm, whereas its anterior part will generate the paraxial mesoderm of the head. Finally, the rostral-most region of the PS and the node will supply the first midline axial mesendodermal tissues: the prechordal plate (the head mesoderm) and cells for the anterior definitive endoderm. By late streak stages,

extraembryonic, cardiac, and head mesoderm have ingressed completely. Caudal and intermediate PS regions will then provide the remainder paraxial and lateral plate mesoderm; while anterior PS and node start laying down first the anterior head process (head notochord) and then trunk notochord (Kinder et al., 1999; Kinder et al., 2001; Parameswaran and Tam, 1995; Smith et al., 1994). Most embryonic definitive endoderm also derives from the epiblast exiting the PS through its most anterior part (Kwon et al., 2008; Lawson et al., 1991; Tam et al., 2003; Wilson and Beddington, 1996). This requires that cells fated to endodermal lineages, after a period of anterior or lateral migration, undergo a mesenchymal-to-epithelial transition (MET). This way, they reacquire epithelial traits and insert themselves into the emVE layer as single cells. This generates a "salt-and-pepper" pattern of cells with different origins that results from dispersion of emVE cells, followed by extensive mixing between emVE and epiblast-derived definitive endoderm cells (Kwon et al., 2008).

Once again, BMP, Wnt and Nodal signalling participate in the regulation of important morphogenetic events, but this time by having crucial roles in mesodermal lineage allocation. In particular, genetic analyses indicate that *Bmp4* seems to play an essential role in the development of extraembryonic structures and in the formation of blood islands (Winnier et al., 1995), whereas *Wnt3a* is essential for trunk paraxial mesoderm development (Yoshikawa et al., 1997). Furthermore, dose-dependent Nodal activity in the epiblast seems to control specification of all the different mesoderm lineages. Analysis of phenotypes resulting from the gradual lowering of Nodal signalling through multiple combinations of *Smad2* and *Smad3* mutations indicated that high levels Nodal activity are required for anterior definitive endoderm and prechordal plate formation (Dunn et al., 2004; Vincent et al., 2003). Conversely, lower levels of Nodal signalling are required to produce tissues originating from the intermediate regions of the PS, such as the paraxial and lateral plate mesoderm (Dunn et al.,

2004). These experiments further demonstrate the high degree of regionalization in cell fate in and around the PS, as well as their dependence in BMP, Wnt and Nodal signalling pathways.

I.5 Axis extension in Vertebrates

Aside from extensive mesoderm and endoderm formation, the first stages of gastrulation also include substantial growth. However, the overall shape of the embryo remains quite unchanged. The first major large-scale tissue reorganization starts around E7.75 with the thickening, flattening and folding of the anterior half of the embryo to build the head folds (Sutherland, 2016). In the meantime, new tissue is continuously added at the embryo's posterior end, generating all the remainder necessary body structures (Wilson et al., 2009). This process of axial extension appears to be conserved in most vertebrates and depends on the proliferation of specialized progenitors that first reside in the PS and later in the tailbud. In the mouse embryo, axial elongation lasts until E13.5, ceasing shortly before the last somites are formed (Stern et al., 2006; Tam and Tan, 1992).

I.5.1 Neuromesodermal progenitors (NMPs)

Lineage tracing and cell grafting experiments showed that the three germ layers are not completely segregated during gastrulation. Instead, some axial precursors actually reside within the PS or in the tailbud during long periods of time, participating continuously in axial elongation and contributing to tissues derived from different germ layers (Cambray and Wilson, 2002; Cambray and Wilson, 2007; Tam and Beddington, 1987; Tam and Tan, 1992; Tzouanacou et al., 2009; Wilson and Beddington, 1996). In fact, surface ectoderm and endoderm are the only lineages fully separated during early stages of PS ingression. All other germ layers appear to segregate gradually as gastrulation progresses, except for the neuroectoderm and mesoderm which share a common precursor during the

entire body axis elongation process (Tzouanacou et al., 2009). These are the neuromesodermal progenitors (NMPs): a multipotent population of precursors that lays down the neural tube and paraxial mesoderm of the post-cranial axis in a progressive, rostral-to-caudal sequence until embryo elongation is complete.

I.5.2 Node-Streak Border vs Chordoneural Hinge

NMPs first occupy the Node-Streak Border (NSB), which corresponds to the junction between the node and the anterior-most region of the PS (Fig I.4A). NSB progenitors are an epithelial population that dwell in this region for the entire PS-dependent period of gastrulation (Cambray and Wilson, 2007; Tam and Tan, 1992; Wilson and Beddington, 1996). However, as the PS starts to disappear between E9.0-9.5, NMPs are relocated into a region in the emerging tail bud that comprises the posterior neural plate and the distal end of the notochord. This area is known as the chordoneural hinge (CNH) (Fig I.4B). NMPs within the CNH then take over the axial extension process throughout the tail, generating ventral neural tube, paraxial mesoderm and notochord tissue until the end of axial elongation (Cambray and Wilson, 2002).

Besides multipotency, represented by their contribution to both neural and mesodermal tissues, NMPs also display self-renewal properties. This means that NMPs are able to contribute tissues to all axial levels, even if normally fated to produce late/posterior-most axial structures. In fact, groups of CNH-derived cells up to E13.5 can not only be serially transplanted through multiple PS without any significant loss of potency, but are also capable of contributing to long axial distances while still contributing to the CNH progenitor pool (Cambray and Wilson, 2002; Tam and Tan, 1992). The same is true for NSB derived explants (Cambray and Wilson, 2007).

I.5.3 Molecular characterization of axial progenitors

The tissue continuity between NSB and CNH is also demonstrated by the extensive similarities in gene expression. Genes expressed in the NSB at E8.5 such as *Wnt3a*, *Fgf8*, *T/Brachyury*, *Foxa2*, *Cdx2* and *Sox2* are likewise expressed in the CNH area and in the same relative positions for most of the remaining axial elongation period (Cambray and Wilson, 2007). FGF, Notch and Wnt signalling all seem to act cooperatively in maintaining NMP proliferation and protecting the less-differentiated end tissues of the tail from terminal differentiation. Mutations or conditional inactivation of genes in these pathways generate axial deformities or even truncations (reviewed in Henrique et al., 2015; Wilson et al., 2009). Particularly, mutants for *Wnt3a* or components of the canonical Wnt signalling pathway show severe axial truncations and expansion of neural tissue at the expense of paraxial mesoderm (Garriock et al., 2015; Takada et al., 1994; Yamaguchi et al., 1999; Yoshikawa et al., 1997). Actually, the formation of neural instead of mesodermal tissues in these mutants indicates that Wnt/β-catenin signalling not only seems to be crucial for progenitor survival and expansion during axial extension, but also appears to have an important role in NMP fate choice by regulating the balance between paraxial mesoderm and neuroectoderm production. This idea is further supported by experiments using *in vitro* produced NMPs (Garriock et al., 2015; Martin and Kimelman, 2012; Olivera-Martinez and Storey, 2007; Tsakiridis et al., 2014; Wymeersch et al., 2016).

Despite the ever-growing body of information regarding axial progenitors and axial progenitor-containing regions, in-depth analyses have been hindered by the lack of specific markers. This probably results from the fact that the PS and axial progenitors residing therein are a dynamic population whose cellular composition changes through time (Jurberg et al., 2013; Williams et al., 2012; Wilson and Beddington, 1996). In fact, fate mapping experiments have shown that epiblast cells belonging to regions that flank the rostral PS – the caudal lateral epiblast or

CLE – are also able to move and reside in the NSB for some time, contributing to both neural and mesodermal lineages (Cambray and Wilson, 2007). Yet, it seems that NMPs in mouse and other vertebrate model organisms can be identified by coexpression of *Sox2* and *T/Brachyury* (Martin and Kimelman, 2012; Olivera-Martinez et al., 2012; Tsakiridis et al., 2014; Wymeersch et al., 2016). Not only is there a close association between coexpression of these markers and NMP-containing areas (NSB, rostral CLE and CNH) during the entire period of axis extension, but cells expressing both genes also disappear shortly before the end of axial extension or in a context of axial truncations (Wymeersch et al., 2016). Additionally, *in vitro* derived $Sox2^{+}T^{+}$ cells grafted into E8.5 NSBs are capable of contributing to both neural and mesodermal lineages (Tsakiridis et al., 2014).

I.5.4 Cessation of body axis extension

Termination of axis extension in vertebrates is associated with a progressive loss of mechanisms that protect axial progenitors from differentiation. Retinoic Acid (RA) signalling appears to have a key role in this process, promoting the expression of differentiation markers in axial tissues (Diez del Corral et al., 2003; Shum et al., 1999). Exposure to exogenous RA generates axial truncations and loss of both *Wnt3a* and *Fgf8* expression (Diez del Corral et al., 2003; Kessel, 1992; Shum et al., 1999). Conversely, FGF signals protect CNH progenitor cells from differentiation by promoting expression of the RA catabolizing enzyme *Cyp26a* and repressing expression of both retinoic acid receptor (*RAR*) *β* and the retinoid synthesizing enzyme retinaldehyde dehydrogenase 2 (*Raldh2*) (Diez del Corral et al., 2003; Olivera-Martinez and Storey, 2007; Wahl et al., 2007).

There is still much to learn about the processes regulating axial extension termination. However, experiments conducted in chick suggest that, as axial extension approaches its end, the distance between RA producing tissues (the somites) and the CNH is progressively shortened. This likely triggers the RA-

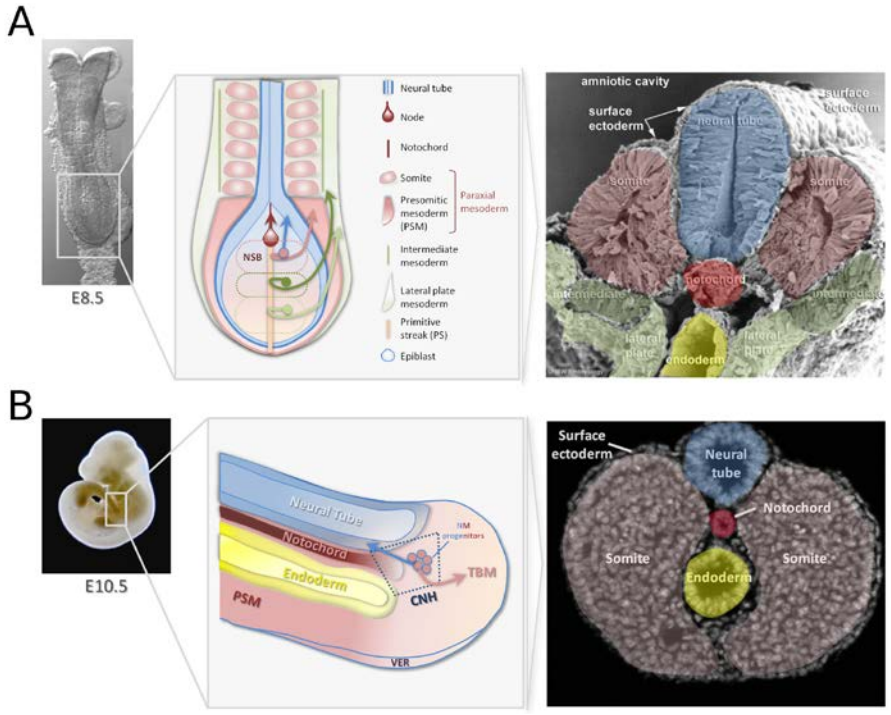

Fig I.4 Axial extension during early organogenesis (A) and tail bud stages (B). Body axis extension occurs progressively by the sequential addition of new tissue produced by axial progenitors to the embryonic caudal end. **A.** Axial progenitors or neuro-mesodermal progenitors (NMPs) first reside in the Node-Streak Border (NSB), which comprises the region between the node and the most anterior part of the primitive streak (PS). These are a self-renewable and multipotent cell population that contributes to both paraxial mesoderm and neural tube. Epiblast cells are still ingressing through the PS and node at this developmental stage, which results in a continuous production of definitive endoderm, notochord and intermediate and lateral plate mesoderm. These tissues will generate all the necessary organs and other structures that constitute the embryo's trunk. **B.** As the PS and epiblast disappear, intermediate and lateral plate mesoderm cease to be generated and NMPs are relocated to the chordo-neural hinge (CNH). This region of the emerging tail bud comprises the posterior neural plate and the distal end of the notochord. NMPs will continue with the elongation of the tail, supplying neural tube and components for the tail bud mesoderm (TBM) that will later be incorporated into the presomitic mesoderm (PSM).

mediated down-regulation of *Wnt3a*, *Fgf8*, *Fgf4* and *Brachyury/T* in the axial progenitor territory, leaving them gradually unprotected from RA influence until all axial progenitors end up disappearing by differentiation and/or apoptosis (Olivera-Martinez et al., 2012; Tam, 1981; Wilson et al., 2009). This seems to be the case for mouse embryos as well. In fact, *Wnt3a*, *Fgf8*, *Brachyury/T* and *Cdx2*

start being down-regulated 48h prior to extension arrest in mouse tail buds, having disappeared completely as the last somites are formed (Cambray and Wilson, 2007). $Sox2^{+}T^{+}$ cells are likewise lost once axial elongation is completed (Wymeersch et al., 2016).

I.5.5 Axial extension and the Vertebrate body plan

A look into the vertebrate clade uncovers a wide variety of body shapes and sizes within organisms belonging to this group. Yet, all vertebrates share a similar spatial tissue organization that starts being defined during gastrulation and is continuously generated for most of the axial extension process (Stern et al., 2006; Wilson et al., 2009). The neural tube, as precursor of the spinal cord, is positioned in the dorsal midline and overlies other midline structures, including the notochord – a transient, rod-like mesodermal formation that runs the entire AP length of the vertebrate embryo (Balmer et al., 2016) – and the embryonic gut. These midline structures are then flanked laterally by the different mesodermal compartments, which, as described previously, derive from cells ingressing at different AP levels of the PS. Paraxial mesoderm lies adjacent to the neural tube, in two parallel stripes of tissue that become compartmentalized into regular segments known as the somites. Further lateral, the vertebrate embryo contains the intermediate mesoderm (precursors of kidney and genital duct), and the lateral plate mesoderm. The latter not only will contribute to the formation of several internal organs and their respective blood supply, but is also the origin of the limb skeleton (Gilbert, 1997; Wolpert et al., 1998).

The vertebrate body can be broadly distributed into three main regions: a complex head with most sensory organs, an organ-containing trunk and a post-anal tail (Irie and Kuratani, 2014; Stern et al., 2006). The trunk and tail regions, as discussed above, are specified and laid down during the process of axial extension. As such, transitions between these compartments not only imply

quantitative and qualitative changes in both the molecular and cellular dynamics of the tissue types generated during axial elongation, but also a strict coordination between them (Jurberg et al., 2013; Wilson et al., 2009). Head to trunk transition seems to encompass, among other things, the finishing of the head and neck structures, emergence of the forelimb and the beginning of internal organ formation. In contrast, trunk to tail transition is associated with the termination of intermediate and lateral plate mesoderm production (which is closely related to hindlimb induction), morphogenesis of the gut-derived cloaca, and also with a progenitor reallocation from the NSB to the CNH, that will keep generating neural tube and paraxial mesoderm during tail formation (Jurberg et al., 2013). This transition entails an important switch in neural tube formation as well. Head and trunk neural tube is generated by the so-called primary neurulation, whereby the neural tissue is formed as a plate that then folds upon itself to form a closed tube. Tail neural tube, however, is produced through tail bud mesenchyme MET followed by cavitation, in a process designated as secondary neurulation (Beck, 2015; Schoenwolf, 1984; Wilson et al., 2009). These differences highlight the remarkable divergence in the extension mechanisms operating in the trunk *vs* tail regions.

I.6 Axial patterning

I.6.1 Somitogenesis or segment formation

The paraxial mesoderm is the origin of a substantial part of the body's structures. These include the skin dermis, the entire body and limb musculature and the most fundamental trait in the vertebrate body: the axial skeleton, which includes vertebrae and ribs (Tam, 1981; Wolpert et al., 1998). As with the rest of the body, paraxial mesoderm is also constantly growing at the posterior embryonic end by the addition of new tissue that, in this case, derives specifically from NMP activity. This tissue then undergoes a process of progressive

differentiation that eventually leads to the formation of their mature derivatives. The most posterior end of the paraxial mesoderm is constituted by non-segmented mesenchymal tissue, a region known as presomitic mesoderm (PSM). The first maturation step consists in the formation of transient blocks of epithelial cells, the somites, that are released from the anterior PSM at a pace matching that of the extension of the PSM at its posterior end. Somites will then undergo a complex differentiation process to form their various derivatives. Somite differentiation is, therefore, not simultaneous, but depends on their position within the AP axis: anterior somites are developmentally more mature than those located at more posterior axial levels since these are generated at later developmental times (Mallo, 2016).

Somitogenesis, or the process of somite formation, occurs at regular intervals of time with a species-specific periodicity: each somite pair is formed approximately every 120 minutes in mouse embryos, whereas it takes about 90 minutes in the chick and 30 minutes in zebrafish (Kimmel et al., 1995; Tam, 1981). The most referenced mechanistic explanation for the process of somitogenesis is the Clock-and-Wavefront model, first proposed by Cooke and Zeeman (1976) (Fig I.5). Briefly, the periodicity of somite formation is regulated by the clock component of the system, whereas the *wavefront* triggers an abrupt change in cellular properties that ultimately results in the formation of a new intersomitic boundary, thus liberating a somite pair (reviewed in Dequéant and Pourquié, 2008; Hubaud and Pourquié, 2014).

The segmentation clock is a molecular oscillator that results mainly from negative feedback loops that produce cyclic expression of genes belonging to the Notch, Wnt and FGF signalling pathways (Aulehla et al., 2003; Bessho et al., 2003; Dequéant et al., 2006). The synchronized cyclic transcription in neighbouring cells generates waves of gene expression that travel through the PSM in a caudal to rostral progression (Fig I.5, blue). Their periodicity closely matches the rhythm of

segmentation, in such a way that a segment is created every time a wave arrives in the anterior PSM (Fig I.5B, D, F) (Aulehla et al., 2008; Masamizu et al., 2006; Palmeirim et al., 1997). The activity of the signalling pathways showing oscillation behaviour, most particularly those triggered by Notch signals, is responsible for activating the program that produces a new segment. This activity, however, remains silent in most PSM, becoming activated only in its anterior-most part (Fig I.5C, G). It is the second component of the model – the wavefront, also known as the determination front – that establishes where cells become competent to respond to Notch signals and make an intersomitic boundary (Fig I.5, blue bracket). The determination front is thought to result from the combination of two main gradients: a posterior to anterior gradient of FGF and Wnt signalling that blocks segmentation (Fig I.5A, red triangle) (Aulehla et al., 2008; Dubrulle and Pourquié, 2004; Dubrulle et al., 2001) and an anterior to posterior gradient of RA that allows the segmentation program to be activated (Fig I.5A, green triangle) (although the involvement of a RA gradient in this process has been recently questioned, see Mallo, 2015). This results in an activity threshold of Wnt, FGF and RA signalling that grants anterior PSM competence to respond to the periodic signals from the molecular clock (Fig I.5, blue bracket). As the embryonic axis elongates, PSM cells at a particular axial level become progressively more anterior, which means that they become gradually exposed to lower FGF and Wnt activity levels while falling more and more under the influence of RA. When FGF and Wnt signalling fall below a certain threshold, cells become competent to respond to clock signals and to activate the segmental and MET programs that allow somite detachment and epithelialisation (Fig I.5B, C, F, G) (Dale et al., 2006; Dequéant and Pourquié, 2008; Duband et al., 1987; Hubaud and Pourquié, 2014; Saga et al., 1997; Tam, 1981). The distance travelled by the determination front during one single oscillation cycle of the segmentation clock thus determines the size of each somite. As such, the speed of the determination front displacement,

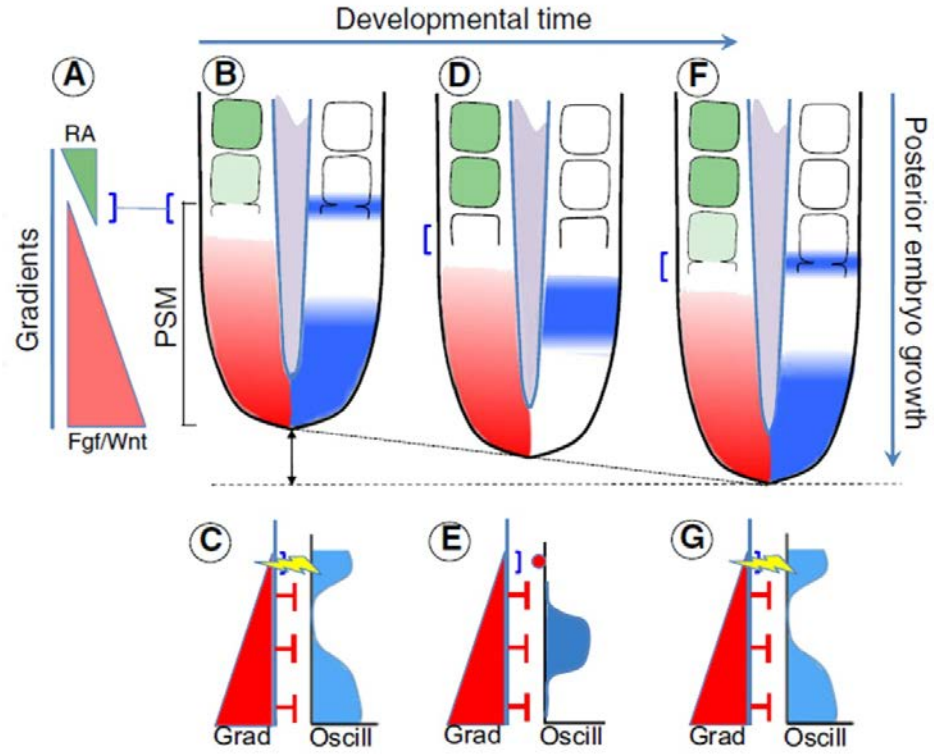

Fig I.5 Clock-and-Wavefront model of somitogenesis. As development progresses, the presomitic mesoderm (PSM) becomes compartmentalized in a rostral to caudal sequence into epithelial segments – the somites. The clock-and-wavefront model proposes that the periodicity of somite formation is regulated by the cyclic transcription of genes belonging to the Notch, FGF and Wnt pathways (the clock component, in blue). This generates periodic waves of gene expression that travel through the PSM in a posterior to anterior progression (**B-G**). The wavefront component, also known as the determination front (represented as a blue bracket), is found in the anterior PSM and results from the combination of two opposing signalling gradients: a caudal to rostral gradient of Wnt and FGF signalling (red triangle in **A**) and a rostral to caudal gradient of RA (green triangle in **A**). The determination front is a threshold level of signalling activity that provides tissues with the necessary competence to respond to the segmentation clock signals. When a wave of cyclic gene expression reaches the determination front, the segmentation program is activated and a new intersomitic border is formed (yellow ray in **C** and **D**). As embryonic axis extension progresses, the Wnt and FGF sources are progressively driven away from the last somite formed, and the determination front moves posteriorly (**D**). However, a new somite will arise only when the next wave of cyclic segmentation clock signalling reaches the determination front in its new position (**F** and **G**). The distance travelled by the determination front during one single oscillation cycle of the segmentation clock determines the size of each somite (adapted from Mallo, 2015).

which ultimately derives from the rate at which new tissue is added to the caudal end of the PSM, will determine the speed at which somitogenesis proceeds along the AP axis. This ensures that axial extension and somite formation are coordinated (Dequéant and Pourquié, 2008; Hubaud and Pourquié, 2014; Martin, 2015). This process goes on until axial extension ceases and all the PSM is consumed. In the end, about 65 somites are formed in the mouse embryo (Tam, 1981). After detaching from the PSM, each new epithelial somite undergoes a progressive differentiation process that leads to the production of mature muscle-skeletal tissues. The first phase of maturation comprises the specification of three major compartments within the somite: the sclerotome, dermomyotome and syndetome. The sclerotome corresponds to the ventromedial part of the somite and will give rise to the entire axial skeleton (reviewed in Christ et al., 2004; Mallo, 2016; Monsoro-Burq, 2005). Sclerotome development is initiated by signals from the adjacent notochord, which secretes Sonic hedgehog (Shh) and the BMP antagonists Gremlin and Noggin (Fan and Tessier-Lavigne, 1994; Stafford et al., 2011). These factors induce prospective sclerotomal cells to undergo an EMT and to express specific early sclerotomal markers like *Pax1*, *Pax9* and *Bapx1*(Christ et al., 2004; Fleming et al., 2015; Monsoro-Burq, 2005; Stafford et al., 2011). These cells will then migrate to different locations around the neural tube and notochord, where they establish discrete subdomains with distinct molecular profiles associated with each specific anatomical component of the vertebrae. The sclerotome-derived vertebral primordia will later undergo a process of endochondral ossification, thus generating an ossified axial skeleton (Christ et al., 2007; Mallo, 2016).

While the ventromedial somite forms the sclerotome, its dorsolateral part keeps its epithelial characteristics to form the dermomyotome. This will later give rise to the dermatome and the myotome, which provide progenitors for the dermis and

skeletal muscle of body and limbs, respectively (Bentzinger et al., 2012). Dermomyotome cells typically express *Pax3* and *Pax7* and their specification is dependent on Wnt signals coming from the surface ectoderm and dorsal neural tube, as well as Shh secreted by the notochord and relative levels of BMP found in the tissue (Borycki et al., 1999; Goulding et al., 1991; Ikeya and Takada, 1998; Jostes et al., 1990; Parr et al., 1993; Reshef et al., 1998). As development progresses, cells at the dorsomedial and ventrolateral borders of the dermomyotome progressively undergo EMT, detach from the epithelium and populate the region between dermomyotome and the sclerotome, becoming the myotome. This compartment characteristically expresses the early muscle markers *Myf5* and *Myf6*, as well as *Myogenin*, *Desmin* and *MyHC* later in development (reviewed in Buckingham and Rigby, 2014; Pownall et al., 2002). The myotome can form either a congruent layer that will end up surrounding the body wall or, at limb levels, constitute a migratory population of individual cells that colonize the limb buds. As the dermomyotome disaggregates and loses its epithelial characteristics, progenitors of the dermis as well as those of the brown adipose tissue are produced and the dermatome is formed (Mallo, 2016).

Finally, a new compartment has been more recently identified, the syndetome, which is specified from *Scleraxis* (*Scx*)-expressing sclerotomal cells adjacent to the myotome. This region generates progenitor cells for the tendons that makes up the essential connective tissue of the musculo-skeletal system (Brent et al., 2003).

Interestingly, there is not a strict coincidence between each individual somite and the final, mature, vertebrae corresponding to that position. Instead, each vertebra is formed by the fusion of the posterior and anterior halves of two adjacent somites, a process designated as re-segmentation (Aoyama and Asamoto, 2000; Christ and Wilting, 1992; Huang et al., 2000). AP polarity within the somite is specified very early in the PSM, even before proper intersomitic

border formation (Mallo, 2016; Saga and Takeda, 2001). This is demonstrated by the existence of specific gene expression profiles in each domain. The most common markers used to identify rostral and caudal somite regions are *Tbx18* and *Uncx4.1*, respectively; yet many other genes, particularly those belonging to or downstream of the Notch signalling pathway, are also differently expressed (Bussen et al., 2004; Kraus et al., 2001; Neidhardt et al., 1997; Saga and Takeda, 2001). Re-segmentation occurs only in the sclerotomal compartment of the somite and it is crucial to ensure the proper movement of the vertebral column by intercalating vertebrae and muscles (Aoyama and Asamoto, 2000; Christ and Wilting, 1992; Huang et al., 2000). This way, each vertebral element can be mobilized by the action of muscles that are generated by the dermomyotome.

I.6.2 Segment identity specification

Organisms belonging to the Tetrapod clade, which includes all land vertebrates, characteristically have skeletons with strong AP regionalization. Typically, the different vertebrae can be grouped into five major domains, cervical, thoracic, lumbar, sacral, and caudal, based on shared morphological characteristics (Burke et al., 1995). Cervical vertebrae constitute the neck skeleton and are those located between the skull and the first rib-bearing vertebra, whereas thoracic vertebrae are defined on the basis of their association with ribs. On the other hand, lumbar vertebrae are generally the largest and densest, and are those between the last thoracic and the first sacral element. Sacral vertebrae are distinguished by the presence of lateral fusions between them that provide the site for pelvic attachment. Finally, caudal vertebrae are those posterior to the sacrum and are typically characterized by a progressive loss of morphological complexity that, in the mouse, is illustrated by the absence of neural arches after the fourth or fifth vertebra (Mallo, 2016; Mallo et al., 2010). The distribution of vertebral elements among the above mentioned regions is known as the vertebral formula. In the

mouse there are normally 7 cervical, 13 thoracic, 6 lumbar, 4 sacral and between 20 to 30 caudal vertebrae (Burke et al., 1995).

Although somites produce structures with specific anatomical characteristics depending on their position in the AP body axis, they are created equal, at least morphologically. Hence, somites must undergo specific differentiation routes, somehow acquiring specific anatomical attributes as a consequence of the activity of a variety of factors regulating their positional identity.

Hox genes

Hox genes are among the most relevant of those factors. They constitute a large family of genes encoding homeodomain-containing transcription factors. Originally discovered in the fruit fly (*Drosophila melanogaster*), these genes are evolutionarily very conserved, and homologous to *Drosophila Hox* genes have since been found in all bilaterian organisms (Duboule, 2007; Garcia-Fernàndez, 2005; Lewis, 1978). Mammalian genomes typically include 39 *Hox* genes, which are organized in four chromosomal clusters (identified from A to D) (Fig I.6, top). Sequence comparisons indicated that each gene in one given cluster has equivalent genes, or paralogues, in one or more of the other clusters, occupying relative similar positions within them. *Hox* genes have 13 such groups, designated as paralogue groups (PGs). This kind of organization likely originated from two consecutive duplication events from a single ancestral complex, followed by divergence and gene loss during the course of evolution (Duboule, 2007; Garcia-Fernàndez, 2005).

A hallmark of *Hox* genes is their characteristic temporal and spatial activation in the three germ layers, in a sequence that follows their order within the clusters – a phenomenon known as colinearity (Duboule and Dollé, 1989; Kmita and Duboule, 2003). In vertebrates, *Hox* gene activation starts with components of *HoxPG1* to *HoxPG4*, which are first expressed in the early PS (Forlani et al., 2003;

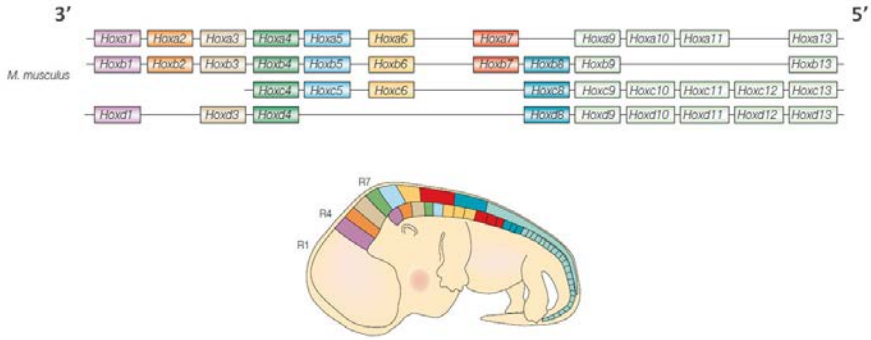

Fig I.6 *Hox* gene expression and genomic organization in the mouse embryo. *Top:* Mammalian genomes typically include 39 *Hox* genes, which are organized in four chromosomal clusters (A-D). Genes that share extensive sequence similarities and occupy similar relative positions within the clusters are further distributed into 13 paralogue groups (PGs). *Bottom:* Hox genes are activated in all three germ layers in a characteristic spatial and temporal sequence, following their order within the clusters. This colinearity of expression results in a unique combination of active *Hox* genes for each axial level – the "Hox code". Adapted from Pearson et al., 2005.

Iimura and Pourquié, 2006). Genes belonging to *Hox*PGs located more 5' within the clusters are then successively activated in the embryo's posterior structures as they are produced by axial progenitors. This temporal and spatial colinearity of expression results in a unique combination of active *Hox* genes for each axial level – the "*Hox* code" – that will play a central role in specifying the final morphology for the corresponding segments along the AP axis (Fig I.6, bottom) (Deschamps and van Nes, 2005; Kmita and Duboule, 2003; Mallo et al., 2009).

The role of *Hox* genes in conferring segment identity has been shown by several loss- and gain-of-function studies. *Hox* gene inactivation typically causes homeotic transformations, whereby one body segment is converted into the identity of another (Lewis, 1978). In vertebrates the study of *Hox* gene function is hindered by the presence of a strong functional redundancy among PG members (Mallo et al., 2009; Wellik and Capecchi, 2003). However, the genesis of mice carrying inactivating mutations in all members of a given PG revealed the essential role that *Hox* genes play in a variety of axial patterning processes. For instance,

complete abrogation of all *HoxPG10* genes resulted in the acquisition of thoracic-like characteristics (as defined by the presence of ribs) by the prospective lumbosacral region, whereas complete inactivation of *HoxPG11* resulted in a transformation of sacral into lumbar-like vertebrae (Wellik and Capecchi, 2003). Gain-of-function experiments have spectacularly confirmed the rib-suppressing activity of *HoxPG10* genes, as premature *Hoxa10* expression in the PSM resulted in completely ribless mice (Carapuço et al., 2005). A similar approach also showed that *Hoxa11* over-expression is able to "sacralize" or even confer caudal-like phenotypes to prospective thoracic and lumbar vertebrae (Carapuço et al., 2005). Conversely, *HoxPG6* genes seem to be the ones responsible for conferring thoracic identity by triggering the rib-promoting program. Indeed, transgenic mice ectopically expressing *Hoxb6* in the PSM developed ribs associated with vertebrae in the prospective neck and lumbar areas (Vinagre et al., 2010).

The Cdx family

Cdx genes are a family of transcription factors closely related to *Hox* genes that also have an important role in the AP patterning of the vertebrate axial skeleton (Deschamps and van Nes, 2005). In mammals, this group is composed of three genes, *Cdx1*, *Cdx2* and *Cdx4*. Inactivation of these genes in the mouse generated homeotic transformations, albeit somewhat less extensive than those observed in mutants for Hox genes (Chawengsaksophak et al., 1997; Chawengsaksophak et al., 2004; Subramanian et al., 1995; van den Akker et al., 2002). Also similarly to *Hox* genes, there is a certain degree of redundancy among them (van den Akker et al., 2002; van Nes et al., 2006; van Rooijen et al., 2012; Young et al., 2009). Yet, it is possible that the homeotic transformations observed in *Cdx* mutant mice derive from alterations in *Hox* gene expression, since it has been shown that Cdx proteins play important roles in the regulation of *Hox* gene expression (van den

Akker et al., 2002). Interestingly, *Cdx* genes seem likewise essential in the process of axial extension. In fact, not only all three *Cdx* genes are expressed in axial progenitor regions, but changes in the allelic composition of *Cdx2*, particularly in the absence of any other Cdx family members, originate axial truncations that can be rescued by *Hoxa5* and *Hoxb8* (van den Akker et al., 2002; Young et al., 2009). This also shows that certain "central" *Hox* genes may still possess the potential to stimulate trunk axial extension.

Retinoic acid, FGF and Wnt signalling

Other factors are known to also have a crucial role in the control of segmental identity. RA was among the first factors shown to be associated with homeotic transformations, which seem to have resulted from changes in *Hox* gene expression (Abu-abed et al., 2001; Kessel, 1992; Kessel and Gruss, 1991). The FGF and Wnt signalling pathways are also important for AP patterning through their involvement in the regulation of *Hox* gene expression. In fact, studies of several hypomorph mutants have shown both anterior and posterior transformations of vertebral identities, which were connected to corresponding alterations in *Hox* gene expression (Partanen et al., 1998; reviewed in Deschamps and van Nes, 2005; Mallo et al., 2009).

Gdf11 signalling

Gdf11 (or Bmp11) signalling is likewise implicated in AP patterning processes and will be one of the main topics in this work. Gdf11 is a secreted factor member of the TGF-β superfamily of signalling factors. As such, it requires the formation of heterodimeric type I and type II receptor complexes upon its binding, in particular those comprised of Alk5 (also known as TGFβRI) and activin type II receptors (ActRIIA or ActRIIB). Alk5 will then transphosphorylate Smad2 and/or Smad3 that, in turn, will activate gene expression upon transduction to the nucleus (Andersson et al., 2006; Hinck, 2012; Oh and Li, 1997; Oh et al., 2002). *Gdf11* expression is

first detected at low levels throughout the whole embryo at ~E7.5; whereas later on, at E8.5, it is observed in the dorsal neural folds and throughout the entire mesenchyme of the posterior embryonic region. After E9.5, *Gdf11* expression gets restricted to the posterior dorsal neural tube and tail bud mesoderm, where it is maintained until well beyond E12.5 (Nakashima et al., 1999). Complete inactivation of *Gdf11* in mice delays the trunk to tail transition up to six somites, as seen by the corresponding posterior displacement of cloaca and hindlimbs (Jurberg et al., 2013; McPherron et al., 1999). This is accompanied by extensive anterior transformations starting at the mid-thoracic level, with mutant mice generally displaying 18 thoracic and 8 lumbar vertebrae, as opposed to the 13 thoracic and 6 lumbar vertebrae found in wild type animals (Lee et al., 2010; McPherron et al., 1999; McPherron et al., 2009). These changes are concomitantly associated with caudal shifts in *Hox* gene expression, particularly of posterior *Hox* genes like *HoxPG10* and *HoxPG11* (McPherron et al., 1999). In fact, it has been shown that signalling through Gdf11/Smad can activate directly *Hoxd11* expression (Gaunt et al., 2013). Genetic experiments have also shown that mutations in Gdf11 receptor ActRIIB display multiple patterning defects, including mild homeotic transformations resembling those observed in *Gdf11* mutant embryos (Oh and Li, 1997). Both severity and penetrance of these phenotypes are increased when combined with a lower allelic dosage of ActRIIA, demonstrating that both receptors cooperatively mediate Gdf11 signalling and can compensate for each other (Oh et al., 2002). Additionally, even though Gdf11 has been shown to interact with both Alk4 and Alk7 in *in vitro* assays, only *Alk5* mutations in a ActRIIB mutant background were shown to reproduce *Gdf11*$^{-/-}$ phenotypes in axial patterning, indicating a genetic interaction between these two receptors (Andersson et al., 2006).

The powerful connection observed between gene expression, segment identity and final vertebral morphology suggests that *Hox* genes in particular may have

played an important role in the evolution of the vast diversity in vertebrate body shapes. In fact, interspecific comparative analyses have revealed that *Hox* genes' anterior expression boundaries are fundamentally associated with body landmarks and anatomical transitions such as the forelimb and hindlimb. These modifications in *Hox* gene expression thus seem to have accompanied the evolutionary changes in vertebrate bodies, even the most extreme ones (Burke et al., 1995; Di-Poï et al., 2010).

Thesis Aims

General Aim:

To investigate the mechanisms controlling vertebrate axial extension and the trunk to tail transition.

Specific aims:

1) To determine the role of Gdf11 signalling in the regulation of the axial progenitor pool size and activity during the trunk to tail transition.
2) To investigate the role of Gdf11 signalling and Oct4 activities during mouse embryonic axial extension.
3) To analyse the contribution that changes in *Oct4* regulation might have had in the evolution of the vertebrate body plan and the emergence of the wide diversity in body shape and size observed among different vertebrate species, with particular emphasis in the origin of the snake body plan.

Chapter II:

Gdf11 signalling and the control of axial progenitor population during the mouse trunk to tail transition

"Trust the hair, Mulch. Trust the hair."

- In *Artemis Fowl and the Time Paradox,* by Eoin Colfer

Author contributions: Rita Aires and Moisés Mallo designed experiments. Rita Aires and Arnon Dias Jurberg performed experiments; Ana Nóvoa performed pronuclear microinjection of DNA constructs. Rita Aires and Moisés Mallo analyzed data. Rita Aires wrote this chapter.

II.1 Summary

A look into the vertebrate clade reveals the great variety of body shapes and sizes within this group. However, despite this diversity, all vertebrate embryos share core developmental processes. One of these is axial elongation, wherein the embryo body gradually extends in a rostral to caudal progression by the continuous proliferation of posterior axial progenitors. The new tissue is then subsequently patterned in a way that all main body components are properly formed and positioned. Gdf11, a member of the TGF-β family of factors, has an important role in axial progenitor behaviour, particularly in the temporal and spatial control of the transition from trunk to tail-forming processes. We found that *Gdf11* mutants have severe tail abnormalities that seem to have originated from the existence of not one, but several axial progenitor-containing sites. Most progenitors seem to be kept in an ectopic ventral epithelial pocket containing an undifferentiated mass of cells, as these tissues express molecular markers normally associated with axial progenitors. Surprisingly, we found that Oct4 was also present in a small subset of cells within the *Gdf11*$^{-/-}$ tail. Biochemical and genetic studies indicate that Gdf11 signalling is likely to negatively regulate *Oct4* expression directly. Overall, our data suggests that Gdf11 acts as a negative modulator of growth, having an important role in the extinction of epiblast-residing axial progenitors during the trunk to tail transition and in the control of the NMP population size during the final stages of axial extension. This way, Gdf11 signalling ensures a gradual termination of the axis and a correct tail specification in the embryo.

II.2 Background

The remarkable variation in body shape and size observed in the vertebrate clade can be traced back to conserved, fundamental developmental processes.

Axial extension, or the embryonic body's elongation, is one such process. It consists in the sequential, rostral-to-caudal, addition of tissue by a group of dedicated cells known as axial progenitors. These are a self-renewing and multipotent population of precursors that generate most tissues in the body, including all mesodermal components, neural tube and part of the definitive endoderm (Kwon et al., 2008; Stern et al., 2006; Tam and Beddington, 1987; Tzouanacou et al., 2009; Wilson and Beddington, 1996; Wilson et al., 2009). Yet, a specific group of cells among these progenitors are capable of producing neural tube and paraxial mesoderm throughout the entire axial elongation period. These are the Neuromesodermal progenitors (NMPs) that, in the mouse embryo, first reside in the epiblast, in the region comprising the caudal node and anterior primitive streak, also known as the node-streak border (NSB) (Cambray and Wilson, 2002; Cambray and Wilson, 2007; Tzouanacou et al., 2009). As the primitive streak disappears, NMPs exit the epiblast and are relocated to the chordoneural hinge (CNH) of the emerging tail bud, which comprises the posterior ventral neural tube and caudal end of the notochord. They then take over the axial extension process until the end of axial elongation (Cambray and Wilson, 2002; Cambray and Wilson, 2007). Although NMP repositioning entails significant changes in their characteristics, particularly a putative epithelial to mesenchymal transition as they move from the epithelial NSB to the mesenchymal CNH, there is a strong continuity in gene expression between the two NMP-containing regions (Cambray and Wilson, 2007; Wilson et al., 2009). Vertebrate NMPs can be identified by co-expression of *Sox2* and *T/Brachyury* (Martin and Kimelman, 2012; Olivera-Martinez et al., 2012; Tsakiridis et al., 2014; Wymeersch et al., 2016), yet further detailed analyses of this cell population have been hindered by the lack of truly NMP-specific markers.

An important functional aspect of axial extension is the distribution of the post-cranial body into its three main regions - neck, trunk and tail – and the transitions between them. The trunk is laid down by epiblast-residing axial progenitors and contains most vital and reproductive organs, which require an extensive contribution from all three germ layers: ectoderm, mesoderm and endoderm (Wolpert et al., 1998). In contrast, CNH-dependent tail elongation is a much simpler process, since this structure is mainly constituted by surface ectoderm, neural tube and paraxial mesoderm. The transition between these two compartments is associated with hindlimb induction and morphogenesis of the gut-derived cloaca, which result from the terminal differentiation of intermediate and lateral plate mesoderm progenitors, and also with NMP reallocation from the NSB to the CNH (Jurberg et al., 2013). Also, the transition entails an important switch in the mechanisms of neural tube formation, from primary neurulation in the trunk to secondary neurulation in the tail (Beck, 2015; Schoenwolf, 1984; Wilson et al., 2009). As such, a correctly patterned vertebrate body requires not only a fine balance between maintenance of the NMP pool and continuous production of tissue, but also a strict control of the amount and types of tissues produced for each main body compartment and for each particular axial level.

Genetic data has shown that Gdf11, a member of the TGF-β family of signalling factors, is a key regulator of the trunk to tail transition and axial progenitor behaviour during mouse axial extension (Jurberg et al., 2013). Also known as Bmp11, this gene is expressed in regions containing axial progenitors during most of the elongation process (Nakashima et al., 1999). *Gdf11* mutant mice show longer trunks due to the delayed onset of this transition (Jurberg et al., 2013; McPherron et al., 1999). Tail development is also seriously compromised in these mutants, since these embryos display severe truncations and vertebrae fusions.

In this chapter we analyzed in detail the tails of *Gdf11*$^{-/-}$ embryos. We show that mutant tails display anterior homeotic transformations, in which the post-sacral vertebrae display a "lumbar-like" morphology with well developed neural arches that bear neural tube until the very end of the truncated tails. Tail defects are already clear at mid-gestation stages, since E10.5 *Gdf11* mutant embryos exhibit enlarged neural tubes and often tail bifurcations. The observed defects seem to stem from the existence of an ectopic ventral pocket of cells that likely represent incomplete resolution of the epiblast during the trunk to tail transition. This structure seems to be the place of residence of an expanded population of axial progenitors and expresses several NMP-associated markers, including *Oct4*. Chromatin immunoprecipitation (ChIP) and genetic analyses suggest that Gdf11 signalling can directly regulate *Oct4* expression through the interaction between existing Smad binding elements (SBEs) and a putative TGF-β inhibitor element (TIE) present in this gene's proximal enhancer region.

II.3 Materials and Methods

II.3.1 Embryos

Gdf11 mutant embryos were produced from intercrosses between *Gdf11*$^{+/-}$ mice (McPherron et al., 1999). For lineage tracing experiments, *Gdf11* heterozygous male mice containing the *Cdx2-Cre*^{ERT} transgene (*Gdf11*$^{+/-}$*::Cdx2-Cre*$^{ERT+/+}$) were crossed with *Gdf11* heterozygous females that additionally carried a knock-in Cre-inducible *LacZ* allele in the ROSA26 locus (*Gdf11*$^{+/-}$*::ROSA26r-β-gal*$^{+/0}$*,* originating from crosses between *Gdf11*$^{+/-}$ and *ROSA26r-β-gal*$^{+/+}$ [FVB.129S4(B6)-Gt(ROSA)26Sor$^{<tm1Sor>}$/J, stock No: 09427, Jackson] mice) to generate *Gdf11*$^{-/-}$ *::Cdx2- Cre*$^{ERT+/0}$*::ROSA26r-β-gal*$^{+/0}$ embryos. CRISPR/Cas9-edited embryos were produced by pronuclear injection (Hogan et al., 1994), using *Cas9* mRNA and guide RNAs Oligo#1 +Oligo#2 or Oligo#1 + Oligo#4 for *PEΔSBE* or *PEΔTIE* deletions, respectively. Oligonucleotide sequences used to produce guide RNAs are listed in

Table II.1. Mouse embryos were recovered by caesarean section at different developmental stages and processed for skeletal analysis, whole mount *in situ* hybridization, immunofluorescence or X-gal staining (see below). All experiments conducted on animals followed the Portuguese (Portaria 1005/92) and European (Directive 2010/63/EU) legislations, concerning housing, husbandry, and welfare. The project was reviewed and approved by the Ethics Committee of "Instituto Gulbenkian de Ciência" and by the Portuguese National Entity, "Direcção Geral de Alimentação Veterinária" (license reference: 014308).

II.3.2 Genotyping

Pups and adult mice were genotyped from tail biopsies. Samples were incubated overnight at 50º C in PBND buffer (50 mM KCl, 10 mM Tris-HCl pH8.3, 2.5 mM $MgCl_2$, 0.1 mg/ml gelatin, 0.45% Nonidet P40 (NP40), 0.45% Tween 20) containing 200 µg/ml proteinase K (Roche). Lysates were then heat-inactivated for 15 minutes at 95ºC. Embryos and fetuses were genotyped by PCR using genomic DNA extracted from yolk sacs and intestines or skin, respectively. Yolk sacs were incubated overnight at 50º C in yolk sac lysis buffer (50 mM KCl, 10 mM Tris-HCl pH8.3, 2 mM $MgCl_2$, 0.45% Tween-20, 0.45% NP40) containing 200 µg/ml of proteinase K). Lysates were heat-inactivated as above. Intestine and skin samples were incubated overnight at 50º C under agitation in Laird's buffer (100 mM Tris-HCl pH8.5, 5 mM EDTA, 0.2% SDS, 200 mM NaCl) supplemented with 100 µg/ml of proteinase K. Genomic DNA was then precipitated with isopropanol (1:1 vol:vol) and transferred to TE buffer (1 mM EDTA, 10 mM Tris-HCl pH8.0). PCR was performed using 1µl of the genomic DNA solution. Primers used for genotyping are summarized in Table II.1.

Genotyping of CRISPR/Cas9-edited embryos was performed as above, using the yolk sac as sample tissue. Edited alleles were analyzed as following. The PCR product was run in a 15% resolving polyacrylamide gel in TBE buffer (TBE 1x, 89

mM Tris, 89 mM Boric acid, 2 mM EDTA) and bands were excised using a scalpel. They were then macerated, incubated in TE for 2 hours and used as sample for a new PCR using the same set of primers. After this amplification step, the PCR product was sequenced with each of the primers used for genotyping.

II.3.3 Phenotypic analysis

Skeletal analyses

Skeletal preparations were performed at E18.5 by alcian blue/alizarin red staining as previously described (Mallo and Brändlin, 1997). Briefly, fetuses were skinned, eviscerated and fixed in 100% ethanol at room temperature. They were stained with alcian blue 8 GX (150 mg/l in 20% acetic acid, 80% ethanol) at room temperature for 12 hours and postfixed in 100% ethanol for 12 hours. Fetuses were cleared in 2% KOH for 8 hours at room temperature, stained with alizarin red S (50 mg/l in 2% KOH) for 2 hours and further cleared in 2% KOH for 10 to 16 hours at room temperature. The reaction was stopped with 25% glycerol and the stained carcasses stored in the same solution.

Histological analysis

Embryos were collected by caesarean section, dissected in PBS 1x (Dulbecco's Phosphate Buffered Saline, without calcium and without Magnesium, Biowest: 1.8 mM KH_2PO_4, 2.7 mM KCl, 10 mM Na_2HPO_4, 137 mM NaCl) and fixed in Bouin's fixative (Sigma) for one overnight (E10.5) or two days (E18.5). They were then dehydrated thoroughly in EtOH 100%, washed extensively in toluol and embedded in paraffin. Embryos were sectioned in 2 µm slices using a microtome and stained with haematoxylin and eosin.

Table II. 1 Oligonucleotides used in this chapter for genotyping, RT-PCR, qPCR , and CRISPR/Cas9 deletions.

Genotyping	*LacZ*	Fw	AGCAGTTTTTCCAGTTCCGTTTATC	T_a= 60ºC
		Rv	AGCGGCGTCAGCAGTTGTTTTTTAT	
	Cre^{ERT}	Fw	CGAGTGATGAGGTTCGCAAG	
		Rv	CCTGATCCTGGCAATTTCGGCT	
	Gdf11	Fw	GCATCCTTTCATGGAGCTTCG	
		wt_Rv	CTGGCCGGAGCAGTAGTTGG	
		mut_Rv	AGTAGAAGGTGGCGCGAAGG	
	PEΔSBE and PEΔTIE	Fw	CTGGGGACCAGGATTGTCCAG	
		Rv	TTGGGGACGTCTGGACAGGAC	
RT-PCR	*Oct4*	Fw	TGAAAGCCCTGCAGAAGGAG	
		Rv	GGTGTCCCTGTAGCCTCATACTC	
	Actin	Fw	ATGAAGATCCTGACCGAGCG	
		Rv	TACTTGCGCTCAGGAGGAGC	
qPCR	Control	Fw	AGTGCTGTCTAGGCCTTAGAG	
		Rv	CAGGGCAGAGCTATCATGCAC	
	SBE-rich	Fw	ACACAAGATGGAATACTGTGC	
		Rv	TGCACCTCCAGTCCTCCAGAG	
CRISPR/Cas9 editing	CRISPR/Cas9 Oligo #1		TAGTGTCTAATCTACCAACC	
	CRISPR/Cas9 Oligo #2		AGGGGATGGAGCCTGGGTGC	
	CRISPR /Cas9 Oligo #4		GCTGAGTGGGCTGTAAGGAC	

Whole mount in situ *hybridization*

Expression analyses were performed by whole mount *in situ* hybridization (Kanzler et al., 1998) using *in vitro* transcribed digoxigenin-labelled antisense RNA probes. Briefly, embryos were dissected out in PBS and fixed with 4% paraformaldehyde (Sigma) (in PBS) at 4ºC overnight. Embryos were washed in PBT (PBS containing 0.1% Tween-20), dehydrated with methanol and rehydrated with PBT. They were then treated with proteinase K (10 µg/ml in PBT) at room temperature for 9 minutes, the reaction was stopped with glycine (2 mg/ml in PBT) and embryos were postfixed with 4% paraformaldehyde, 0.2% glutaraldehyde (Sigma). Hybridization was performed at 70ºC overnight in hybridization solution (50% formamide, 1.3 x SSC [3M NaCl, 300 mM sodium citrate] (pH 5.5), 5 mM EDTA, 0.2 % Tween 20, 50 µg/ml yeast tRNA, 100 µg/ml heparin) containing the RNA probe, followed by three washes at 70ºC in hybridization solution without RNA probe, tRNA and heparin. Embryos were then washed in TBST (25 mM Tris.HCl, pH 8.0, 140 mM NaCl, 2.7 mM KCl, 0.1% Tween 20), equilibrated with MABT (100 mM Maleic acid, 150 mM NaCl, 0.1 % Tween-20, pH 7.5), blocked with MABT/Block [MABT containing 1% blocking reagent (Roche #11096176001) and 10% sheep serum] and incubated with a 1:2000 dilution of alkaline phosphatase-conjugated anti-digoxigenin antibody (Roche #11093274910) in MABT/Block at 4ºC overnight. Embryos were washed extensively with MABT at room temperature, equilibrated in NTMT (100 mM Tris HCl, pH 9.5, 50 mM $MgCl_2$, 100 mM NaCl, 0.1% Tween-20) and developed at room temperature with an NBT/BCIP solution (Roche #11681451001) diluted in NTMT or BM Purple (Roche #11442074001). Reactions were stopped with PBT, fixed with 4% paraformaldehyde in PBS and stored in PBT. The mouse probes used in this chapter are summarized in Table II.2.

Table II. 2 *In situ* hybridization probes used in this chapter.

Gene	Probe	Vector	Linerization	Polimerase	Probe size (bps)	Hydrolysis? (min/sec)
Bmp2		pKS	NotI	T3	580	No
Bmp5	Fragment containing the full ORF.	pKS	XhoI	T7	1300	2' 6''
T/Brachyury	Fragment containing the whole ORF.	pKS	NotI	T3	1800	4'
Sox2	Fragment containing part of the ORF and of the 3'UTR.	pSKII	XhoI	T3	700	45''
Nkx1.2 (Sax1)	Fragment spanning the homeobox.	pKS	HindIII	T7	552	No
Fgf8	Fragment containing part of the ORF.	pKS	ClaI	T3	785	No
Foxa2	Fragment containing part of the ORF.		Asp700	T7		No
Oct4	Fragment spanning exons 2-4 and part of exon 5.	pCRII-TOPO	NotI	Sp6	496	No
Wnt3a	Fragment containing the 3' part of CDS and part of the 3' UTR.	pKXM	MluI	T3	400	No

Immunofluorescence

Immunofluorescence analyses were performed either on whole mount embryos or cryostat sections using antibodies against Oct4 (sc-8628 Santa Cruz Biotechnology, 1:250), Sox2 (ab92494 Abcam, 1:250) and Brachyury (AF2085 R&D Systems, 1:250). After overnight fixation with 4% PFA in PBS, embryos were thoroughly washed in PBS or PBS-0.1% Triton-X (PBST) and processed for sectioning or directly to whole mount immunofluorescence. For sectioning, embryos were incubated overnight at 4ºC in 15% sucrose (Sigma) (in PBS), equilibrated with 15% sucrose/7% gelatin (Sigma) solution (in PBS) for 2-6 hours

at 37ºC, transferred to appropriate moulds and frozen in liquid nitrogen. The embryos were sectioned sagittally at 12 µm intervals in duplicate slides using a cryostat. For immunofluorescence, the material was permeabilized with 0.1% (sections) or 0.5% (whole mount) Triton-X in PBS for 20 minutes (sections) or 1 hour (whole mount), incubated with 1M glycine in PBS pH7.2 for 15 minutes (sections) or 30 minutes (whole mount) and washed several times with PBS or PBST. Embryos were blocked in 3% donkey serum (Biowest S2170-100)/1% BSA (Roche #10735086001) solution in PBST for 1 hour at room temperature (sections) or overnight at 4ºC (whole mount). In both cases, primary antibodies were diluted in blocking solution and incubated overnight at 4ºC. Secondary antibodies (Alexa 488 conjugated donkey anti-goat (A-11055) and Alexa 568-conjugated donkey anti-rabbit (A-10042), 1:500) were diluted in blocking buffer and incubated for 3-4h at room temperature (sections) or overnight at 4ºC (whole mount). After extensive washes in PBS or PBST, material was stained with DAPI (1:1000 in PBS) for 15 minutes (sections) or 1 hour (whole mount). Sections were mounted using VectaShield® (Vector Laboratories, H-1000). Whole mount embryos were dehydrated in series of methanol, cleared in crescent series of methylsalicilate (Sigma) in methanol and mounted in 100% methylsalicilate. Confocal Z-series stacks of immunostained material were acquired on a Leica SP5 confocal microscope.

Post-staining embryo sectioning

To section whole mount-stained embryos, they were fixed in 4 % paraformaldehyde, 0.2 % glutaraldehyde (in PBS) for 15 minutes at room temperature, and included in gelatin/albumin (0.45 % gelatin, 270 g/l bovine serum albumin (Roche), 180 g/l sucrose in PBS, jellified with 1.75 % glutaraldehyde). Sections were cut at 35 µm with a vibratome and mounted with an aqueous mounting solution (Aquatex, Merck).

II.3.4 Lineage tracing analysis and β-galactosidase staining

Recombination was induced at E7.5 by intraperitoneal injection of 200 µl of a 1 mg/ml solution of tamoxifen (Sigma) (in corn oil [Sigma]). Embryos were collected at E10.5 by caesarean section, dissected in PBS and processed for β-galactosidase staining. Briefly, embryos were fixed with Mirky's fixative (National Diagnostics) overnight at 4ºC. After 3 washes in washing solution (0.02% NP40, 0.02% Tween 20, in PBS), embryos were incubated in X-gal staining solution (5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6 \cdot 3H_2O$, 2 mM $MgCl_2$, 0.02% NP40, 0.02% Tween 20, in PBS, including 0.4 mg/ml X-gal [Promega]) and the reaction was monitored regularly. After developing, the embryos were extensively washed in washing solution, fixed overnight in PFA and sectioned in the vibratome.

II.3.5 *Gdf11*$^{-/-}$ tail explant cultures

Gdf11 mutant embryos were collected at E10.5, dissected in PBS and cultured as previously described (Tesar et al., 2007). Briefly, tails were severed from the embryo and incubated in PBS containing 0.5% trypsin (SAFC biosciences) and 2.5% pancreatin (Sigma) for 30 minutes. After three washes with EpiSC medium (DMEM-F12 (Gibco), 20% Knock-Out serum replacement (KSR) (Gibco), 0.1 mM 2-mercaptoethanol (Sigma), 2 mM L-glutamine, and 1x non-essential amino acids (both from Invitrogen)), tails were mechanically dissociated to obtain 10-100 cell clumps. These were plated onto mitomycin C (Sigma)- treated mouse embryonic fibroblasts (MEFs) in 24-well plates, with EpiSCs medium supplemented with 5 ng/ml FGF2 (R&D Systems) and with or without 20 µM Alk-receptor inhibitor SB431542 (Abcam). The cell medium was changed once a day and cells were allowed to grow for three days before being passaged by incubation with 1x accutase (Sigma, A6964).

II.3.6 RT-PCR analysis

RNA extraction was performed with TriZOL (Sigma), according to the manufacturer's instructions. Retrotranscription reaction was carried out with NZY Reverse Transcriptase (MB12401, NZYtech) using random hexamer primers. PCR was performed using 2 µl of RT reaction for *Oct4* and *Actin*, with the primers specified in Table II.1.

II.3.7 Chromatin immunoprecipitation and quantitative PCR (qPCR)

Chromatin immunoprecipitation (ChIP) assays were performed using the posterior region of 100 E8.5 and E9.0 wild type embryos. Briefly, embryos were dissected in PBS and the tissue was crosslinked in 1% formaldehyde for 15 minutes at room temperature, under agitation. The reaction was stopped by adding glycine (NZYtech) to a final concentration of 142 mM and incubating for 5 minutes agitating at room temperature. Tissues were thoroughly washed with cold PBS at 4ºC. Samples were then lysated in lysis buffer (1 mM EDTA, 0.5 mM EGTA, 0.2% SDS, 0.1% sodium deoxycholate, 10 mM Tris-HCl pH8.0) containing protease inhibitors for 45 minutes at 4º C, under agitation. Samples were sonicated to 200bp-1kb fragments and centrifuged at maximum speed for 10 minutes at 8ºC. Supernatants were diluted 1:10 in ChIP dilution buffer (2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH8.1, 1% Triton X-100) and immunoprecipitated overnight at 4ºC using anti-phosphorylated Smad2/3 (#8685, Cell Signaling Technology) and control rabbit IgG (#2729, Cell Signaling Technology) antibodies, pre-bound to Dynabeads Protein A (Invitrogen) in ChIP dilution buffer, rotating. Dynabeads and bound chromatin were repeatedly washed with buffers of increasing stringency (ChIP wash 1: 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH8.0, 1% Triton X-100; ChIP wash 2: 2 mM EDTA, 500 mM NaCl, 20 mM Tris-HCl pH8.0, 1% Triton X-100; ChIP wash 3: 0.5% IPGAL, 250 mM LiCl, 1 mM EDTA, 0.5% sodium deoxycholate, 10 mM Tris-HCl pH8.0) and then

twice in TE buffer pH8.0. The DNA was eluted from the beads with an elution buffer (1% SDS, 10 mM $NaHCO_3$) and incubated for 6 to 16 hours at 65º C. The immunoprecipitate was purified by phenol:chloroform extraction, followed by a chlorophorm extraction and a standard overnight NaOAc/EtOH precipitation. qPCR was performed using 2 μl of sample with iTaq™ Universal SYBR Green Supermix (#1725121, BioRad) in a CFX 384 thermocycler (BioRad), using primers in Table II.1.

II.3.8 Data analysis

Neural tube diameter measurements were done using ImageJ. Data was analyzed and plotted using GraphPad Prism7. Statistical analysis was assessed using an unpaired two-tailed student's t-test, and p-value < 0.05 was considered significant.

II.4 Results

II.4.1 *Gdf11*$^{-/-}$ embryos present severe tail defects and enlarged neural tubes

At E18.5, *Gdf11* mutants exhibited tail truncations after the formation of a variable number of post-sacral vertebrae (Fig II.1A, a-b, McPherron et al., 1999). Unlike wild type embryos, in which the axial skeleton finishes progressively and the neural tube ends at the level of the 5th caudal vertebrae (Fig II.1A, c, f black arrowhead), *Gdf11*$^{-/-}$ fetuses showed vertebrae with fully developed neural arches that accommodated neural tube until the very end of the truncated tail (Fig II.1A, b, c, black arrowhead). Tail abnormalities in *Gdf11* mutants were already clear by mid-gestation stages (E10.5), after the trunk to tail transition period (Fig II.1B). *Gdf11*$^{-/-}$ embryos displayed enlarged and sometimes bifurcating tails that did not show the typical gradual width reduction seen in their wild type counterparts (Fig II.1B, a, f). Defects were also evident in the internal structure of the *Gdf11* mutant

tails. Instead of one intact CNH region, the posterior notochord appeared to be physically separated from the ventral posterior neural tube in *Gdf11*$^{-/-}$ embryos by tail mesenchyme. Also, these tails contained an ectopic pocket of cells located ventrally to the neural tube and notochord (Fig II.1B, white asterisk). This tissue was composed of two main elements: an epithelial sheet continuous with the tail-associated gut, projecting inward towards the gut's lumen, and a mass of undifferentiated mesenchymal cells mostly confined within the epithelial pocket (Fig II.1B, b-d, g-I, white asterisk). Besides their overall thickness/width, *Gdf11*$^{-/-}$ E10.5 tails also displayed enlarged neural tubes that were twice as wide as those found in wild type embryos (Fig II.1C). These findings indicate that the internal and external abnormalities observed in *Gdf11* mutant tails were already clear at mid-gestation stages and that the observed defects may arise from the existence of an ectopic epithelial growth located ventrally to the neural tube and the notochord.

II.4.2 *Gdf11* mutant tails show abnormal gene expression patterns

The striking defects observed in the tails of *Gdf11*$^{-/-}$ embryos at mid-gestation stages lead us to perform a detailed gene expression analysis in this structure at E10.5 (Fig II.2). Molecular markers typically present in undifferentiated tail tissues and NMP-containing regions were found to have expanded expression domains in *Gdf11* mutant tails, which included the whole ectopic ventral epithelial pocket. Particularly, transcripts for *T/Brachyury* (Fig II.2, A-A'') and *Sox2* (Fig II.2, C-C''), two markers that, when together, are associated with NMP-containing regions (Wymeersch et al., 2016), were found both in its epithelial and mesenchymal components. Other genes characteristic of immature tail tissues and axial progenitor areas such as *Nkx1.2* (Fig II.2, E-E''), *Fgf8* (Fig II.2, G-G''), and *Wnt3a* (Fig II.2, K-K'') were likewise present in the two components of the ectopic ventral tail structure.

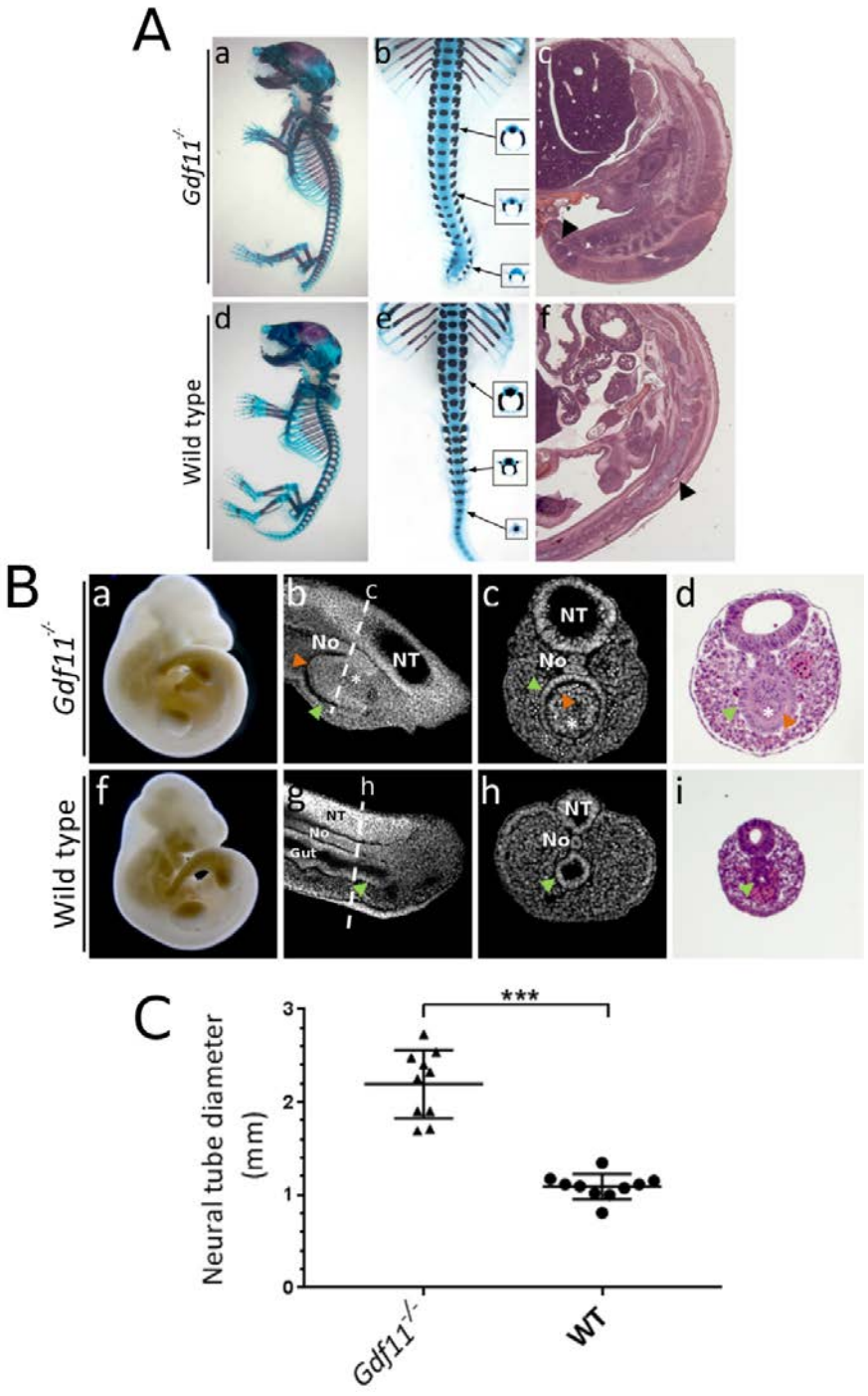
A
a
b
c
Gdf11-/-
d
e
f
Wild type
B
a
b
c
d
Gdf11-/-
No
NT
C
NT
No
f
g
h
i
Wild type
NT
No
Gut
C
Neural tube diameter (mm)
***
3
2
1
0
Gdf11-/-
WT

Fig II. 1

As its epithelial component seems to be continuous with the associated gut, we also tested the gut endodermal marker *Foxa2*. Expression of this gene was weak, scattered, and mostly restricted to the epithelial component of the ectopic tissue, as opposed to the strong signal observed in the epithelium surrounding the entire ectopic pocket that represents the bona fide gut endoderm (Fig II.2, I-I''). As a whole, these expression patterns contrast with the ones seen in wild type embryonic tails, in which these markers were not detected in the gut endoderm or were confined to its posterior end at this stage of development (Cambray and Wilson, 2007) (Fig II.2, B-B'', D-D'', F-F'', H-H'' and L-L''). Further morphological tail abnormalities were also evident by analyzing the expression of ventral epidermal ridge (VER)-associated genes, like *Bmp2* and *Bmp5*. The VER constitutes an epithelial structure located in the posterior ventral portion of the tail that is thought to result from the termination of the primitive streak (Goldman et al., 2000; Ohta et al., 2007). Whilst these genes were expressed in the entire VER epithelium of wild type tails (Fig II.2, N-N'', P-P''), the same markers were found in a spotty, non-continuous manner in *Gdf11* mutant tails (Fig II.2, M-M'', O-O'', black arrowheads).

Fig II.1 *Gdf11* mutants present severe tail defects and enlarged neural tubes (*previous page*). A. Alcian blue/alizarin red skeletal preparations of *Gdf11*$^{-/-}$. (a, b) and wild type (d, e) E18.5 fetuses. Insets in b and e show the morphology of the different post-sacral vertebral elements. c, f: haematoxylin and eosin staining of sagittal sections in *Gdf11*$^{-/-}$ (c) and wild type (f) E18.5 fetuses. Black arrowhead indicates the finishing of the neural tube. **B.** Comparison of *Gdf11*$^{-/-}$ (a-d) and wild type (e-i) tail morphology at mid-gestation stages. a, f: E10.5 whole embryos. Sagittal sections (b, g) and (c, h) transversal sections of *Gdf11*$^{-/-}$ (b, c) and wild type (g, h) tails stained with DAPI. d, i: haematoxylin and eosin staining of E10.5 transversal tail sections. NT, neural tube; no, notochord; gut, embryonic tail gut. Dashed lines represent the level of the transversal section found in the next panel. Green and orange arrowheads indicate the gut epithelium and the epithelial lining of the ectopic ventral mass, respectively. **C.** Neural tube diameter in *Gdf11*$^{-/-}$ *vs* wild type tails at E10.5. Neural tubes were measured from left to right, using transversal sections of ten different embryonic tails per condition. Statistical analysis was assessed using an unpaired two-tailed student's t-test, and p-value < 0.05 was considered significant.

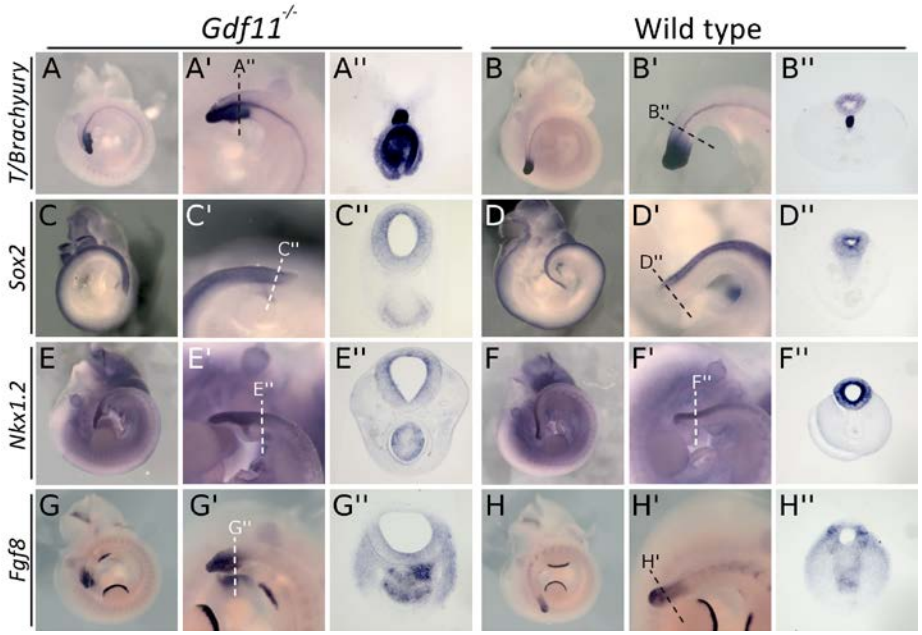

Fig II.2 *Gdf11*$^{-/-}$ tails show abnormal gene expression patterns *(part I)*. Genes associated to undifferentiated tail tissues and NMP-containing regions were analyzed in *Gdf11*$^{-/-}$ (**A-A''**, **C-C''**, **E-E''**, **G-G''**) and wild type (**B-B''**, **D-D''**, **F-F''**, **H-H''**) E10.5 tails: *T/Brachyury* (**A-A''**, **B-B''**), *Sox2* (**C-C''**, **D-D''**), *Nkx1.2* (**E-E''**, **F-F''**), *Fgf8* (**G-G''**, **H-H''**). **A**, **B**, **C**, **D**, **E**, **F**, **G**, **H** show a wholemount view; **A'**, **B'**, **C'**, **D'**, **E'**, **F'**, **G'**, **H'**, a close-up; **A''**, **B''**, **C''**, **D''**, **E''**, **F''**, **G''**, **H''** display transversal sections at the level indicated by the dashed lines.

This suggests that the VER is not properly specified in *Gdf11*$^{-/-}$ embryos, which could account for some of the patterning defects observed in mutant tails.

Overall, these results show that *Gdf11*$^{-/-}$ tails display a global expansion of the expression domains of genes associated with immature tail tissues and NMP-containing regions. These domains specifically encompass the ectopic ventral mass of cells, in such a way that this structure appears to be molecularly similar to the axial progenitor-containing regions and undifferentiated tail tissue found in wild type tails.

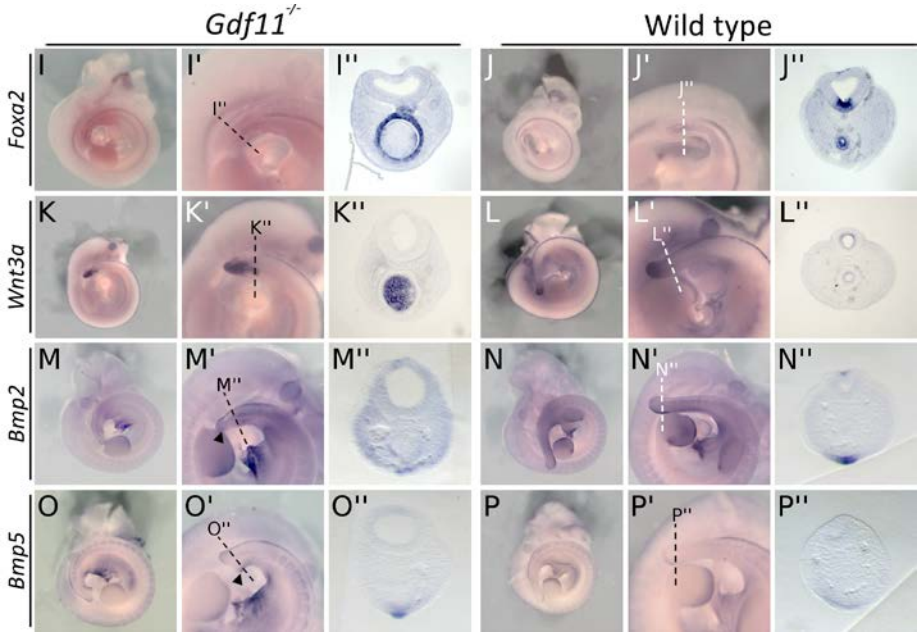

Fig II.2 *Gdf11*$^{-/-}$ tails show abnormal gene expression patterns *(part II)*. Genes associated to undifferentiated tail tissues and NMP-containing regions were analyzed in *Gdf11*$^{-/-}$ (**I-I'', K-K'', M-M'', O-O''**) and wild type (**J-J'', L-L'', N-N'', P-P''**) E10.5 tails: *Foxa2* (**I-I'', J-J''**), *Wnt3a* (**K-K'', L-L''**), *Bmp2* (**M-M'', N-N''**), *Bmp5* (**O-O'', P-P''**). **I**, **J**, **K**, **L**, **M**, **N**, **O**, **P** show a wholemount view; **I'**, **J'**, **K'**, **L'**, **M'**, **N'**, **O'**, **P'**, a close-up; **I''**, **J''**, **K''**, **L''**, **M''**, **N''**, **O''**, **P''** display transversal sections at the level indicated by the dashed lines. Black arrowheads in **O'** and **M'** indicate the location of *Bmp2* and *Bmp5* expression, respectively, in *Gdf11*$^{-/-}$ tails.

II.4.3 *Gdf11* mutants have an expanded axial progenitor population

The presence of enlarged neural tubes, together with a global expansion of NMP-associated molecular markers in *Gdf11* mutant tails, suggested the existence of an increased axial progenitor population in these embryos. Double immunohistochemistry for Sox2 and T/Brachyury in embryonic tail tissues showed that cells expressing both factors were primarily found in the CNH area of wild type embryos (Fig II.3A, c-c'', d-d'') as previously reported (Cambray and Wilson, 2007; Wymeersch et al., 2016). In *Gdf11* mutant tails, however, double positive cells were found not only at the posterior-most end of the notochord and ventral neural tube, but also in most of the epithelial component belonging to the ectopic ventral mass (Fig II.3A, a-a'', b-b'', white boxes). Surprisingly,

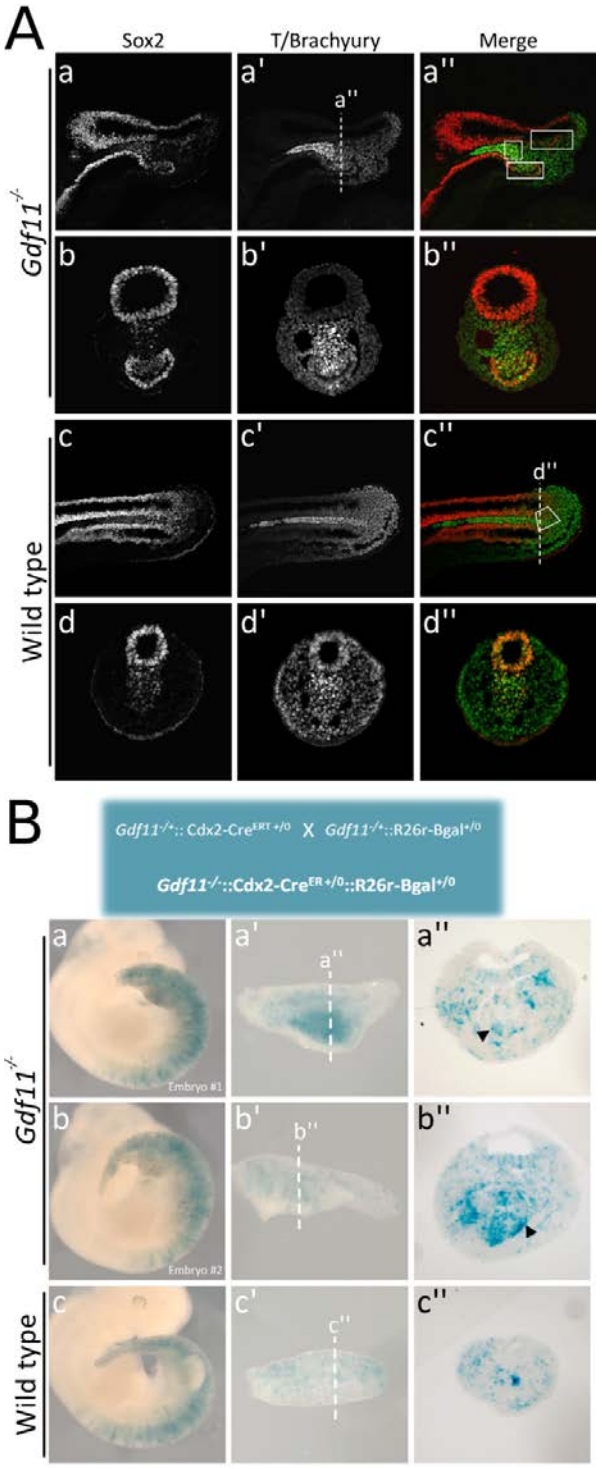

Fig II.3 *Gdf11* mutants have an expanded axial progenitor population. A. Double immunohistochemistry for Sox2 (a, b, c, d and red in a'', b'', c'', d'') and T/Brachyury (a', b', c', d' and green in a'', b'', c'', d'') in *Gdf11*$^{-/-}$ (a-a'', b-b'') and wild type (c, c'', d-d') E10.5 tails. White outlined boxes indicate presence of both factors and possible NMP containing regions in *Gdf11*$^{-/-}$ tails. White outlined trapeze in c'' indicates the location of the CNH in wild type tails. **B.** Lineage tracing experiment using tamoxifen-inducible Cre^{ERT} recombinase (*Cdx2-CreERT*), in combination with an ubiquitously expressed floxed β-galactosidase knock-in allele (*Rosa26r-β-gal*). *Top:* mice crossing strategy. *Bottom: Gdf11*$^{-/-}$:: *Cdx2-CreERT*$^{+/0}$:: *Rosa26r- β-gal*$^{+/0}$ (a-a'', b-b'') embryos and control littermates (*Gdf11*$^{+/+}$ :: *Cdx2-CreERT*$^{+/0}$:: *Rosa26r- β-gal*$^{+/0}$) (c-c''). a, b, c, show wholemount views; a', b', c', tail close-ups; and a'', b'', c'', display transversal sections at the level indicated by the dashed lines. Black arrowheads indicate epithelial β-gal^{+} cells in the ectopical ventral mass of *Gdf11* mutant tails.

Sox2 immunoreactivity was also observed in the gut endoderm of both wild type and *Gdf11* mutant embryos, which contrasts with the absence of *Sox2* transcripts in this structure (Fig II.2 D''; Fig II.3A, a, b, c and d). These data point to the existence of not one, but several CNH-like regions in *Gdf11* mutant tails.

In order to assess location and potency of the NMP population in *Gdf11*$^{-/-}$ tails, we performed lineage-tracing experiments using a Cre^{ERT}/loxP-recombination approach. In particular, we used transgenic mice expressing the tamoxifen-inducible CreERT recombinase under the control of the *Cdx2* enhancer (*Cdx2-Cre*^{ERT}), which has been shown to drive gene expression in axial progenitors throughout development (Benahmed et al., 2008; Jurberg et al., 2013; Jurberg et al., 2014), in combination with a line containing the Cre-reporter *ROSA-β-gal* knock-in allele (*ROSA26r-β-gal*). (Fig II.3B, top; see Methods). In *Cdx2-Cre*$^{ERT+/0}$*::ROSA26r-β-gal*$^{+/0}$ embryos a transient pulse of CreERT–mediated recombination early in development (E7.5) using a single administration of a low dose of tamoxifen results in the labelling of a reduced number or even individual NMPs with active β-gal (Fig II.3B). NMPs can then be identified later in development (E10.5) among the pulse-labelled cells both by their ability to contribute descendants to the neural and mesodermal layers at different axial levels and by their presence in the CNH. This approach was first validated in wild type embryos, in which β-gal-producing cells were found not only in neural and mesodermal tissues all over the trunk and tail, but also in their CNH (Fig II.3B, c-c''). This cell tracing system yielded a similar pattern in *Gdf11* mutant embryos. Yet, there was a strong accumulation of β-gal-labelled cells in the ectopic ventral mass of *Gdf11*$^{-/-}$ tails, as revealed by the presence of intense X-gal staining (Fig II.3B, a-a'', b-b''). Interestingly, these cells were located both in the epithelial (Fig II.3B, a-a'', b-b'', black arrowheads) and mesenchymal components of the ectopic structure. This indicates that *Gdf11* mutants might still possess epithelially-located NMPs, a characteristic of epiblast-dependent stages of axial elongation.

Together, these results show that *Gdf11* mutant embryos have an expanded or even multiple CNH-like regions that appear to occupy three main locations in the tail: the posterior notochord, ventral neural tube and most of the ectopic ventral cellular pocket. This might account not only for the observed tail bifurcations, but also for the enlarged neural tubes and increased global thickness of *Gdf11*$^{-/-}$ tails: NMPs located in two or three foci would likely produce two separate growth domains, while generating more progeny that would contribute to neural and mesodermal lineages. Moreover, these data also suggest that the ectopic ventral mass in mutant tails might still contain epithelial axial progenitors. In fact, the overall gene expression patterns found in the ectopic mass of *Gdf11*$^{-/-}$ embryos resemble to some extent typical features of the NMP-containing epiblast (Fig II.2, Cambray and Wilson, 2007). This suggests that the epiblast in *Gdf11* mutant embryos was probably not completely extinguished during the trunk to tail transition.

II.4.4 Ectopic *Oct4* expression in *Gdf11*$^{-/-}$ tails at E10.5

From the known factors involved in the regulation of axial progenitor activity (Wilson et al., 2009), *Oct4* seems to be unique in being required during trunk formation, yet dispensable at tail bud stages (DeVeale et al., 2013). Interestingly, in about half of the analyzed cases, premature activation of Gdf11/Alk5 signalling in axial progenitors at the beginning of the trunk-formation stage produced embryos with strong morphological resemblance to those where *Oct4* was inactivated at late primitive streak stages (DeVeale et al., 2013; Jurberg et al., 2013). These observations suggest that the epiblast-like structure in the tails of *Gdf11* mutants might be somehow associated with persistent Oct4 activity. Analysis of Oct4 expression was consistent with this hypothesis. While Oct4 expression could not be detected in the tail bud of wild type embryos at E10.5 (DeVeale et al., 2013; Downs, 2008; Osorno et al., 2012) (Fig II.4A, b-b''; Fig II.4B),

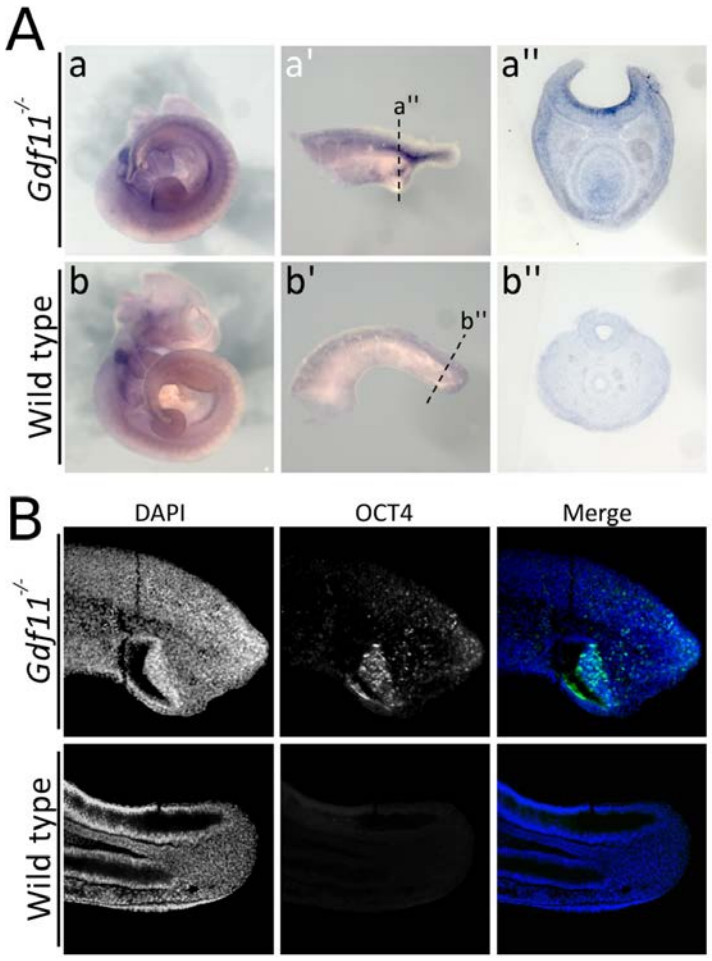

Fig II.4 *Gdf11*$^{-/-}$ tails show ectopic *Oct4* expression at E10.5. A. Wholemount *in situ* hybridization for mouse *Oct4*. a, b, wholemount view; a', b', close up; and a'', b'', transversal sections of *Gdf11*$^{-/-}$ (a-a'') and wild type (b-b'') E10.5 tails at the level indicated by the dashed lines. **B.** Immunofluorescence for Oct4 (green in merge panel) in sagittal sections of *Gdf11*$^{-/-}$ and wild type E10.5 tails.

it was observed in the tail tip, the ventral posterior neural tube and the ectopic ventral tissue mass of similarly staged *Gdf11* mutant embryos (Fig II.4A, a, b-b', Fig II.4B). In this latter structure, Oct4 was found in both its epithelial and mesenchymal components, partially overlapping with T and Sox2 (Fig II.5, A-A'', C-C'', E-E'' and G-G''). This observation further supports the hypothesis that the epithelium in the ectopic ventral epithelial pocket is likely to be an epiblast remnant and is consistent with the role of Gdf11 in promoting the switch in axial extension from epiblast to tailbud-dependent growth (Jurberg et al., 2013).

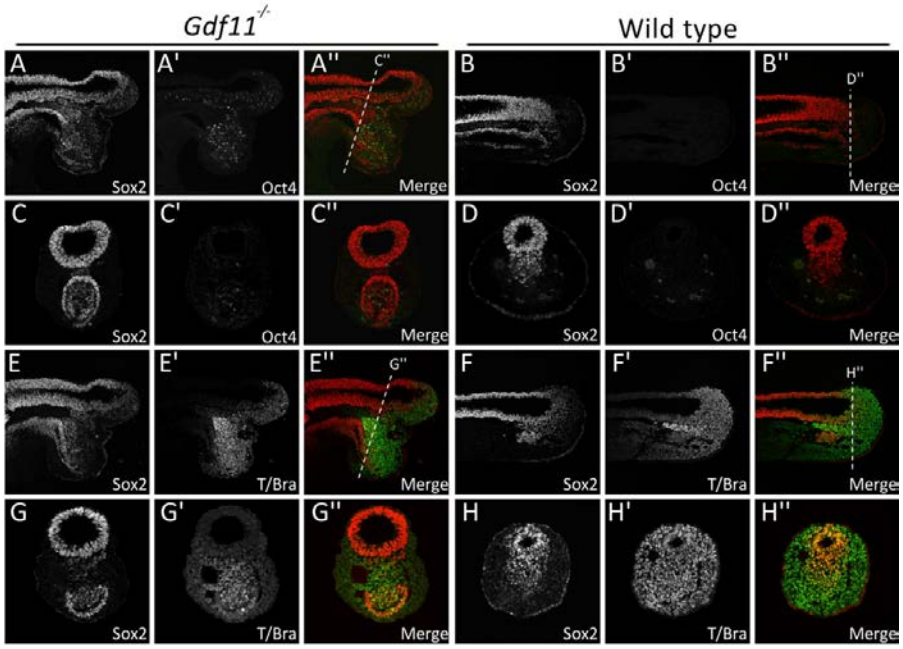

Fig II.5 Oct4 overlaps with Sox2 and T/Brachyury proteins in E10.5 *Gdf11*$^{-/-}$ tails. Double immunohistochemistry for Sox2 (**A**, **B**, **C**, **D**, **E**, **F**, **G**, **H**; red in **A''**, **B''**, **C''**, **D''**, **E''**, **F''**, **G''** and **H''**) and Oct4 (**A'**, **B'**, **C'**, **D'**; green in **A''**, **B''**, **C''** and **D''**) or Sox2 and T/Brachyury (**E'**, **F'**, **G'**, **H'**; green in **E''**, **F''**, **G''** and **H''**) in *Gdf11*$^{-/-}$ (**A-A''**, **C-C''**, **E-E''**, **G-G''**) and wild type (**B-B''**, **D-D''**, **F-F''**, **H-H''**) E10.5 tails. **A-A''**, **B-B''**, **E-E''** and **F-F''** show sagittal sections. Dashed lines represent the level of the transversal section in **C-C''**, **D-D''**, **G-G''** and **H-H''**.

Given the presence of an ectopic epithelium expressing epiblast-associated markers such as Oct4, Sox2 and T/Brachyury in *Gdf11*$^{-/-}$ embryos, we tested whether we could derive epiblast stem cells (EpiScs) – which represent the pluripotent epiblast (Chenoweth et al., 2010; Najm et al., 2011; Tesar et al., 2007) from *Gdf11* mutant tails. These were cultured on MEF-coated dishes in EpiSC medium supplemented either with FGF2 alone or with a combination of FGF2 and the Alk5-inhibitor SB431542. The latter was used so that culture conditions would mimic the Gdf11 signalling-null environment occurring in *Gdf11* mutant embryos. Preliminary results indicated that, in contrast to wild type controls, cells originating from *Gdf11* mutant tails could be grown in both culture conditions, originating compact, epithelial colonies that appeared morphologically similar to EpiSCs colonies one day after plating (Fig II.6A). Colonies maintained this

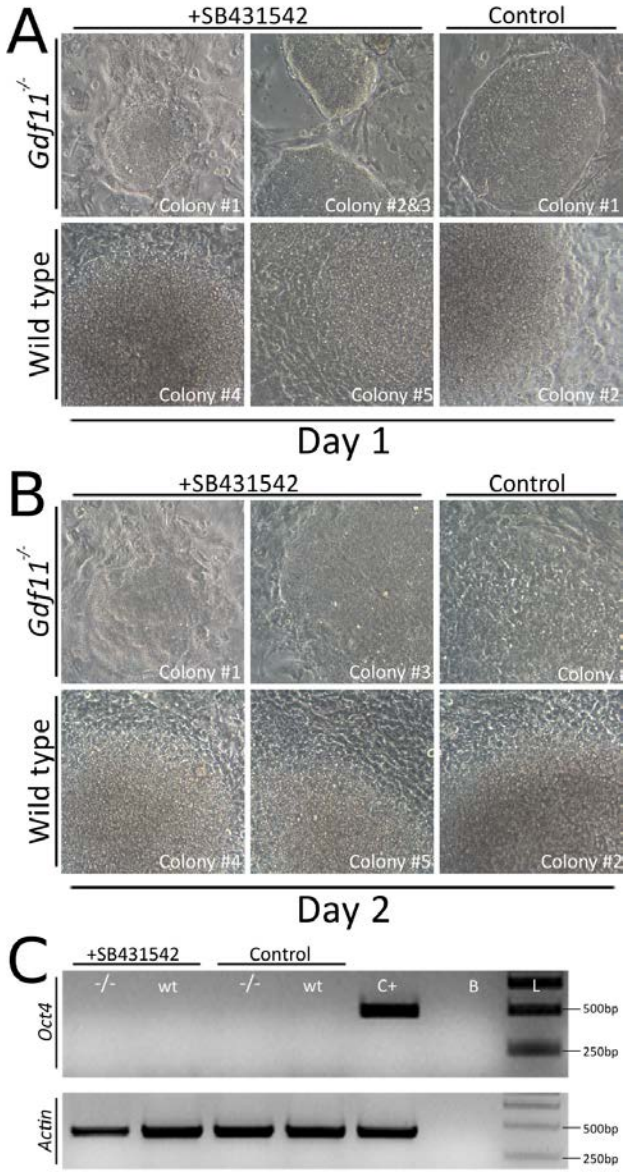

Fig II. 6 *Gdf11*$^{-/-}$ tails are not able to generate true EpiSCs. Colonies derived from *Gdf11*$^{-/-}$ and wild type E10.5 tails after one (**A**) or two days (**B**) in EpiSCs culture conditions, with the Alk5 inhibitor SB431542, or without it (control). **C.** RT-PCR for *Oct4* (top) and *Actin* (bottom, loading control) in colonies cultivated for three days in the two previous conditions, before the first passage. -/-: *Gdf11*$^{-/-}$; wt: wild type; C+: positive control (total cDNA from E7.5 embryos); B: blank; L: ladder (O'GeneRuler 1 kb DNA Ladder, Thermo Scientific).

morphology two days after plating when incubated with the inhibitor-containing medium, but not with medium supplemented with FGF2 alone (Fig II.6B). Furthermore, unlike colonies from wild type tails, *Gdf11*$^{-/-}$ cells could grow and withstand up to 3 passages in both culture conditions. Yet, RT-PCR analyses were unable to detect *Oct4* transcripts three days after plating even before the first passage in any of the culture conditions, either in wild type or mutant cells (Fig II.6C).

These findings demonstrate that even though some important epiblast-associated factors are co-expressed in Gdf11-/- tails, these cells were not capable of generating true EpiSCs when in culture. Nevertheless, they seemed to maintain some self-renewing properties as they could be serially passaged. These mutant cells could thus represent a more advanced epiblast stage, in which the tissue still expresses most of adequate markers but has lost a large part of its competence. Additional work will be needed to further characterize in detail cells obtained from *Gdf11* mutant tails and clarify these issues.

II.4.5 *Oct4* expression appears to be directly regulated by Gdf11 signalling

Given that *Oct4* is ectopically expressed in the absence of *Gdf11*, we decided to investigate whether *Oct4* could be directly regulated by this factor. Gdf11 signals by promoting formation of heterodimeric receptor complexes between Alk5 (also known as TGFβRI) and activin type II receptors such as ActRIIA or ActRIIB. Within these complexes, Alk5 phosphorylates Smad2 and/or Smad3 that, in turn, activate gene expression upon forming a complex with Smad4, entering the nucleus and interacting with Smad-binding elements or SBEs, which contain the core sequence GTCT (or its reverse complement sequence AGAC) (Andersson et al., 2006; Hinck, 2012; Massagué et al., 2005; Oh and Li, 1997; Oh et al., 2002). In mammals, *Oct4* has two main regulatory regions (Fig II.7A). The distal enhancer (DE) is responsible for driving *Oct4* expression both in the pre-implantation embryo and in primordial

germ cells, whereas the proximal enhancer (PE) regulates its expression in the epiblast (Fig II.7A, top panel, green and orange boxes, respectively) (Nordhoff et al., 2001; Yeom et al., 1996). Sequence analyses suggested that *Oct4* could actually be a target of Gdf11 signalling, since several putative SBEs were identified within the PE (Fig II.7B and C, blue boxes), particularly downstream of conserved region 2 (CR2) (Nordhoff et al., 2001) (Fig II.7B and C, pink box). If Gdf11 signalling were indeed necessary for the complete silencing of *Oct4* during the trunk to tail transition, we would expect Smad2/3 to bind *Oct4* regulatory regions, particularly the PE, around this period.

Preliminary chromatin immunoprecipitation (ChIP) experiments from the posterior region of E8.5 and E9.0 embryos detected a considerable enrichment for the SBE-rich region when using an antibody recognizing phosphorylated Smad2/3. (Fig II.7D). To evaluate if this region was indeed responsible for *Oct4* silencing, we used CRISPR/Cas9-mediated genome editing to specifically delete the SBE-rich region within the PE (*PEΔSBE* embryos) (Fig II.7C, green boxes). While most embryos containing the genomic deletion were indistinguishable from wild type littermates (n=11), one exhibited strong ectopic *Oct4* expression in the posterior neural tube and tail bud (Fig II.7E, c-c', black arrowhead). DNA sequencing showed that, while one of the *Oct4* loci of this embryo carried the predicted mutation, the other included a deletion encompassing both the SBE-rich region and the 111 adjacent base pairs 5' to that region. A closer look revealed that those 111bps contained a sequence fitting the 5'-GNNTTGGNGN-3' consensus of a putative TGF-β inhibiting element (TIE) (Fig II.7C, yellow box). TIEs have been described to mediate TGF-β repressive effects, down regulating expression of genes such as *c-myc* and *matrix metalloproteinase-1* (*MMP-1*) that contain these motifs within their promoters (Frederick et al., 2004; White et al., 2000). This way, the sequence 5'-GGGTTGGGGG-3' found upstream the SBE-rich region of the PE could have an

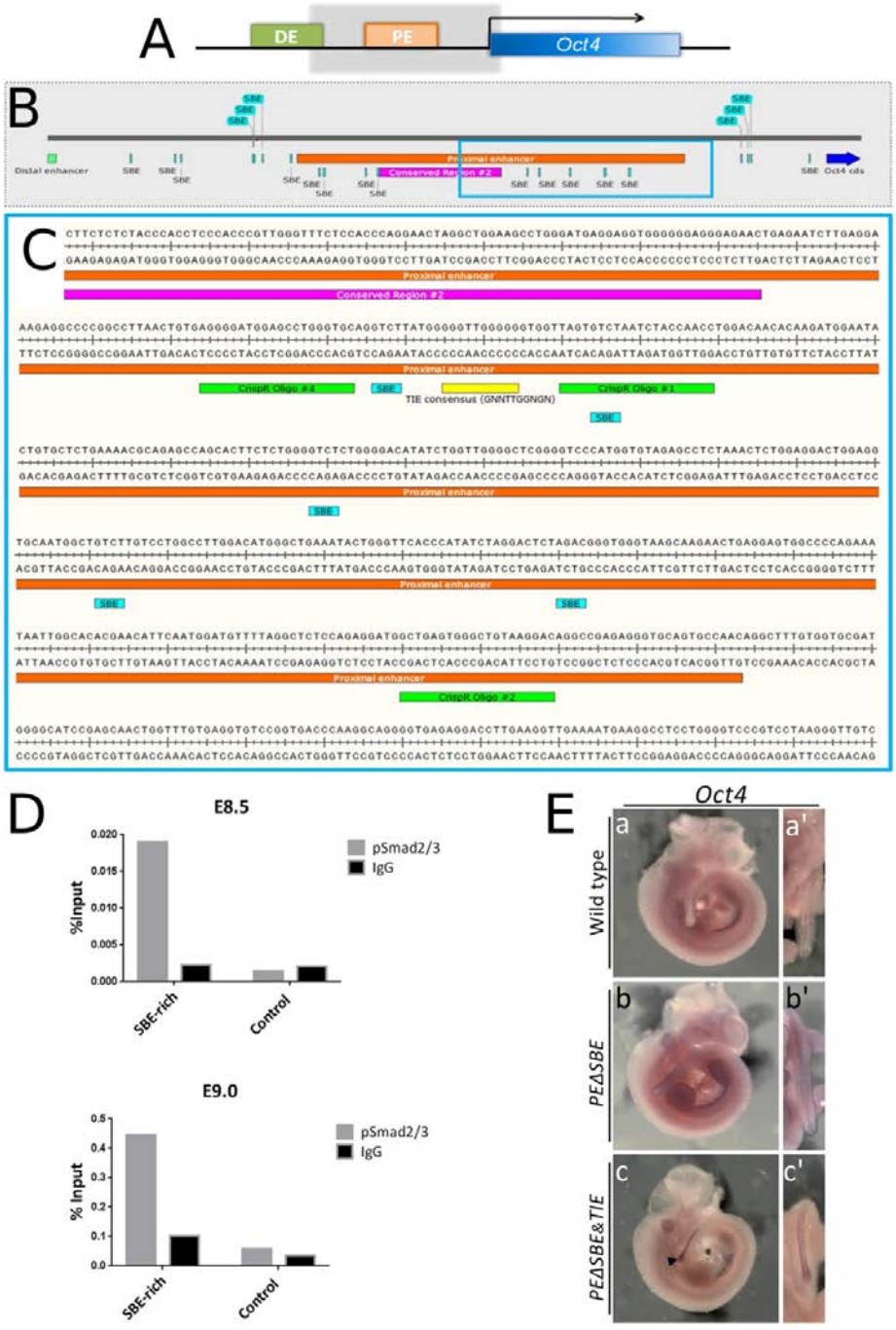
A
DE
PE
Oct4
B
Distal enhancer
Proximal enhancer
Conserved Region #2
Oct4 cds
C
Proximal enhancer
Conserved Region #2
CrispR Oligo #4
CrispR Oligo #1
TIE consensus (GNNTTGGNGN)
CrispR Oligo #2
SBE
D
E8.5
pSmad2/3
IgG
%Input
SBE-rich
Control
E9.0
% Input
E
Oct4
Wild type
PEΔSBE
PEΔSBE&TIE

Fig II.7

important role in *Oct4* silencing during the trunk to tail transition. However, specific deletion of the TIE sequence alone did not lead to de-repression of *Oct4* expression (data not shown). We thus conclude that Gdf11 signalling might be involved in *Oct4* silencing during the trunk to tail transition, possibly through the direct binding of activated Smad2/3 to the TIE and downstream SBEs and/or through interaction with other sequences contained in the *Oct4* regulatory region.

II.5 Discussion

In this chapter we analyzed the tail abnormalities found in *Gdf11* mutant embryos. These defects, which included truncations and vertebral malformations in E18.5 fetuses, were already patent at mid-gestation stages of development in the form of thickened, shortened, and, occasionally, bifurcated tails. Morphological and molecular analyses of *Gdf11*$^{-/-}$ embryos revealed that the observed tail phenotypes mostly stemmed from the existence of an ectopic epithelial pocket of undifferentiated cells, which probably originated from a defective extinction of epiblast-associated axial progenitors during the trunk to

Fig II.7 *Oct4* expression appears to be directly regulated by Gdf11 signalling (*previous page*). A. Schematic view of *Oct4* regulatory regions and locus. *Right: Oct4* has two main regulatory regions: the distal enhancer (DE, green box) and proximal enhancer (PE, orange box). The blue box represents *Oct4* coding region. The grey shaded box represents the region analyzed in detail in this section. Left: close-up of the shaded region of the previous panel. The green box represents the 3' ending of the DE, orange box represents the PE, pink box shows the location of the conserved region 2 (CR2), according to Nordhoff et al., 2001. Light blue boxes indicate Smad-binding elements (SBE) present in this genomic region. Dark blue arrow denotes the beginning of *Oct4*'s coding sequence. **B.** Sequence of the genomic region spanning the 3' region of CR2 (pink box) within the PE until the end of the PE (orange box). The light blue boxes represent SBEs, the yellow box indicate the putative TFG-β inhibitory element (TIE) and the green boxes show the location of the oligos (1, 2 and 4) used for CrispR/Cas9-mediated genome editing. **C.** Chromatin immunoprecipitation (ChIP) experiments with the posterior region of E8.5 and E9.0 embryos, using an antibody recognizing phosphorylated Smad2/3. Relative enrichments are shown as percentage of input. **D.** Wholemount *in situ* hybridization for *Oct4* in embryos subjected to CrispR/Cas9-mediated deletion of the SBE-rich region, using Oligos 1+2. a-a', non-edited embryo (wild type); b-b'', homozygous embryo for SBE-rich region deletion (*PEΔSBE*); c-c'', heterozygous embryo containing one deleted allele (*PEΔSBE*) and other allele carrying an extra 111bps 5' deletion (*PEΔSBE&TIE*). Black arrowhead in c indicates ectopic *Oct4* expression.

tail transition. In fact, previous studies had already identified Gdf11 as a key factor controlling this transition and axial progenitor behavior during axial extension (Jurberg et al., 2014; McPherron et al., 1999). Here we show that absence of Gdf11 generates a miss-regulation of axial progenitor numbers and the presence of multiple NMP-containing regions, instead of one intact CNH. Interestingly, most of the supernumerary NMPs found in mutant tails seemed to be kept in the aforementioned ectopic ventral epithelial pocket. This structure seems to display both CNH- and epiblast-like progenitor characteristics, as shown by lineage tracing experiments and co-expression of several axial progenitor-associated markers, particularly Sox2 and T/Brachyury, both in its epithelial and mesenchymal compartments. The existence of this ectopic mass not only can account for the increased thickness of mutant tails, but also for their frequently observed bifurcations. Indeed, its emergence and expansion would likely impose significant mechanical constraints to tail growth, particularly when combined with the apparent segregation of the CNH into two different, tissue-producing, regions. This would result in two separate foci of tail growth and, consequently, to tail bifurcations.

The presence of Oct4 in a subset of cells in *Gdf11* mutant tails, both in the epithelial and mesenchymal components of the ectopic cellular mass, further supported the hypothesis that the observed ectopic epithelial structure has indeed epiblast-associated progenitor characteristics. In fact, some $Oct4^{+}$ cells also expressed Sox2, T/Brachyury, or even both. This is consistent with the view that epiblast/primitive streak fragments might still be present in $Gdf11^{-/-}$ tails, since the three markers are found together in this structure until *Oct4* disappears around E8.5 (Henrique et al., 2015; Osorno et al., 2012; Wymeersch et al., 2016). It should be noted that the presence of Oct4 did not necessarily correlate with the highest Sox2 levels. This observation is interesting considering that NMPs have been characterized as a cell population expressing low levels of Sox2 and T

(Wymeersch et al., 2016), thus further supporting that the Oct4^{+} cells in the ectopic epithelial ventral pocket might be NSB-associated axial progenitors. One way to verify this hypothesis would be performing a double immunofluorescence for both Oct4 and β-galactosidase proteins in E10.5 *Gdf11$^{-/-}$::Cdx2-Cre$^{ERT+/0}$:: Rosa26r-β-gal$^{+/0}$* embryos with progenitors labelled using our lineage tracing protocol. This way, if Oct4^{+} cells were indeed axial progenitors, we would expect co-localization of the two proteins in the NMP-containing regions of *Gdf11* mutant tails. Yet, *Oct4* expression in *Gdf11$^{-/-}$* tails appeared to be transient, as we could not detect it either in E9.5 or E11.5 mutant embryos (data not shown). This ectopic expression could have resulted from its incomplete silencing during the trunk to tail transition, followed by its upregulation in a small subset of cells at E10.5. Alternatively, this gene could have been re-activated due to the specific molecular environment established in mutant tails, like the one produced by the combination of absent active inhibition (e.g. promoted by Gdf11 signalling, as discussed below) and a molecular context promoting Oct4 activation. The latter could be provided by Sox2 activity, since this factor is known to be part of the positive feedback loop maintaining *Oct4* expression (Chew et al., 2006; Masui et al., 2007). Some support for this possibility is provided by the observation that in the *PEΔSBE&TIE* embryo ectopic *Oct4* activation was mostly seen in Sox2-positive tissues. Whether or not this *Oct4* expression is due to its maintenance or re-activation could be assessed by live imaging experiments using *Gdf11* mutant embryos containing fluorescent reporters for the endogenous Oct4 protein while they undergo the trunk to tail transition.

The fact that Gdf11 signalling can negatively modulate *Oct4* expression, most likely through direct binding of its effectors to SBEs and/or to TIEs, is consistent with its important role in the trunk to tail transition. The presence of both TGF-β activator and repressor sequences in the regulatory regions of a given gene allows activated Smad2/3 to function in a context-dependent manner, acting as

transcriptional activators or repressors depending on existing co-factors and on the promoter sequence to which it is recruited. In the case of Oct4, this configuration would allow Nodal-dependent activation in the epiblast during early development (Mesnard et al., 2006), and Gdf11-dependent repression at later developmental stages. This bimodal "switch" in TGF-β signalling has also been observed in the regulation of other genes such as *c-myc* and *MMP-1* (Frederick et al., 2004; White et al., 2000). Indeed, given that *Oct4* expression is crucial for epiblast and primitive streak maintenance (DeVeale et al., 2013; Mesnard et al., 2006; Nichols et al., 1998), its silencing would be an integral part of the trunk to tail transition mechanisms.

Yet, the observed ectopic *Oct4* expression in the *PEΔSBE&TIE* embryo did not generate any *Gdf11*$^{-/-}$-like tail phenotype. Therefore, *Oct4* might not be involved in the generation of the tail abnormalities characteristic of *Gdf11* mutant embryos or, alternatively, it is not the sole element responsible for those defects. On the other hand, the observed levels of expression might not be sufficient to originate malformations when in the presence of an otherwise active Gdf11 signalling able to regulate additional downstream targets, in contrast to *Gdf11* mutant tails where this signalling is inactive. In the next chapter we analyze the contribution of the ectopic *Oct4* expression to the *Gdf11*$^{-/-}$ phenotype using a transgenic approach to ectopically express *Oct4* in the NSB and CNH throughout the entire axial extension process.

Interestingly, unlike explants from wild type tails, those obtained from *Gdf11*$^{-/-}$ embryos still possessed cells preserving some self-renewing properties, as they could be serially passaged up to 3 times when cultured under EpiSCs conditions supplemented with SB431542. This demonstrates that cells obtained from mutant tails might represent an intermediate potency stage that has lost a large part of its proliferative competence. This is also consistent with the observation that NMPs in mutant tails, although seemingly capable of producing more tissue, particularly

neural tube, eventually stop proliferating and finish axial extension even in the absence of Gdf11 signalling.

Overall, the results presented in this chapter show that Gdf11 signalling seems to be involved in the control of the axial progenitor pool size during the final stages of body axis elongation. Interestingly, Gdf11 signalling has been shown to negatively regulate other progenitor populations, as it is the case of those involved in neurogenesis in the olfactory epithelium of mice (Wu et al., 2003). These data further support *Gdf11*'s role as a chalone, or negative growth modulator (Gamer et al., 2003). Therefore, during the axial extension process, Gdf11 might play a dual role: ensuring a correct trunk to tail transition by promoting the transition from epiblast to tailbud-dependent growth; and regulating NMP population size within the tail, thus contributing to a gradual axis termination.

III.6 Acknowledgements

We would like to thank members of the Mallo lab for useful comments during the course of this project. This work has been supported by grants PTDC/BEX-BID/0899/2014 (FCT, Portugal) and SCML-MC-60-2014 (from Santa Casa da Misericordia de Lisboa, Portugal) to MM. RA is supported by a PhD fellowship (SFRH/BD/51876/2012, from FCT, Portugal).

Chapter III:

Oct4 is a key regulator of vertebrate trunk length

"If I'm gonna die, I'm gonna die historic on the Fury Road."

- In *Mad Max: Fury Road (2015)*

Part of Chapter II and Chapter III are included in:

Aires, R., Jurberg, A. D., Leal, F., Nóvoa, A., Cohn, M. J.; and Mallo, M (2016). Oct4 is a key regulator of vertebrate trunk length diversity. *Dev Cell* (*in press*)

Author contributions: Rita Aires and Moisés Mallo designed experiments. Rita Aires performed experiments; Ana Nóvoa performed pronuclear microinjection of DNA constructs. VISTA analyses and wholemount *in situ* hybridizations in snake embryos were performed by Francisca Leal. Rita Aires and Moisés Mallo analyzed data. Rita Aires and Moisés Mallo wrote this chapter.

III.1 Summary

Vertebrates exhibit a remarkably broad variation in trunk and tail lengths. However, the evolutionary and developmental origins of this diversity remain largely unknown. Posterior *Hox* genes were proposed to be major players in trunk length diversification in vertebrates but functional studies have so far failed to support this view. Here we identify the pluripotency factor *Oct4* as a key regulator of trunk length in vertebrate embryos. Maintaining high *Oct4* levels in axial progenitors throughout development was sufficient to extend trunk length in mouse embryos. *Oct4* also shifted posterior *Hox* gene expression boundaries in the extended trunks, thus providing a link between activation of these genes and the transition into tail development. Furthermore, we show that the exceptionally long trunks of snakes are likely to result from heterochronic changes in *Oct4* activity during body axis extension, which may have derived from differential genomic rearrangements at the *Oct4* locus during vertebrate evolution.

III.2 Background

The Vertebrate clade encompasses remarkable diversity of body shapes and sizes, yet the causes of this variation are poorly understood. Research into this topic has been particularly focused in the wide differences in body length and in regional patterns along the main body axis of vertebrates (Burke et al., 1995; Cohn and Tickle, 1999; Di-Poï et al., 2010; Gomez et al., 2008; Head and Polly, 2015; Woltering, 2012; Woltering et al., 2009). These studies often combined the analysis of animals with extreme variants of specific features (like the long necks of birds or the large rib numbers of snakes) with functional approaches in tractable animal models, allowing for a direct experimental evaluation of the factors potentially involved in generating those patterns. *Hox* genes were identified this way as having a central role in the generation of the different

vertebral patterns observed along the body's anterior-posterior axis and in its variations among species (Di-Poï et al., 2010; Head and Polly, 2015; Woltering et al., 2009). However, the mechanisms regulating other global features of the vertebrate body plan are much less understood. A particularly interesting case is regionalization of the body into head, trunk and tail regions, given that their relative proportions vary widely among vertebrates. Snake bodies provide an extreme example of an uneven distribution of these regions, since these animals are mostly composed of very long, organ-filled, trunks. Snake trunks are not just a simple consequence of their body length, since other vertebrates with remarkably long bodies, like some lizards, are mostly dominated by a long tail and have their organs confined to a relatively small trunk. Also, although long trunks are closely associated with rib-bearing vertebrae, functional studies indicate that *Hox* genes play no major role in the regulation of overall trunk length (Carapuço et al., 2005; Jurberg et al., 2013; Mallo et al., 2010; Vinagre et al., 2010; Wellik and Capecchi, 2003), even if they do determine the identity of its segments. This shows that *Hox* genes are downstream effectors of a still unknown mechanism controlling body region distribution.

Body region allocation and the transition between regions have important morphological, physiological and evolutionary consequences. The partitioning of the body into morphologically discrete regions is defined during embryonic development through the process of axial extension. During this process, the vertebrate body is built progressively from head to tail by the sequential addition of new tissue produced by dedicated axial progenitors located at its posterior end (Wilson et al., 2009; Wymeersch et al., 2016). Embryological studies indicate that, while continuous, axial extension relies on different mechanisms to generate trunk or tail structures. Axial elongation through the trunk requires activity of various types of progenitors located within the epiblast, an epithelial sheet at the posterior embryonic end (Stern et al., 2006; Wilson et al., 2009; Wymeersch et al.,

2016). These include neuromesodermal progenitors (NMPs) that elongate the neural tube and lay down the paraxial mesoderm that will form the musculoskeletal case of the trunk (Cambray and Wilson, 2007; Henrique et al., 2015; Stern et al., 2006; Tzouanacou et al., 2009), and progenitors for the intermediate and lateral mesoderm, which together with the endoderm will build the trunk organs involved in digestive, excretory and reproductive functions (Stern et al., 2006). In contrast, tail development derives almost exclusively from the activity of NMPs that at this stage are embedded within the tail bud mesenchyme (Wilson et al., 2009). Transition from trunk to tail development thus entails a number of morphological and functional changes, including the terminal differentiation of the progenitors for the lateral and intermediate mesoderm and relocation of NMPs to the tail bud (Jurberg et al., 2013).

Genetic data indicate that Gdf11 signalling is a key regulator of the trunk to tail transition. *Gdf11* mutant mice have longer trunks due to the delayed onset of this transition (Jurberg et al., 2013; McPherron et al., 1999). Conversely, premature activation of Gdf11 signalling produces embryos with smaller or even absent trunks as the result of an early transition into tail-producing mechanisms (Jurberg et al., 2013). The intensification of the trunk length phenotype of *Gdf11* mutants when *Gdf8* (*Myostatin*) is simultaneously inactivated indicates partial compensation of Gdf11 activity by *Gdf8* (McPherron et al., 2009). The extended trunks in *Gdf11* mutants suggest that Gdf11 counteracts the activity of hitherto non-identified trunk-promoting factors as part of the program that activates tail development. The identification of these factors might provide important insights into the mechanisms responsible for trunk length diversity among vertebrates.

Here we show that the pluripotency factor *Oct4* is a key regulator of vertebrate trunk length. Sustained *Oct4* expression in axial progenitors during embryonic elongation using a transgenic approach in mice resulted in extended trunks. This factor was also able to coordinate other aspects of anterior-posterior body

patterning, most notably the activation of posterior *Hox* genes. Our results also indicate that persistent *Oct4* expression could be in the origin of the extreme length of snake trunks, since its expression in snake embryos seems to remain active for a longer developmental period when compared to mouse embryos. These heterochronic changes in *Oct4* expression could have resulted from extensive genomic rearrangements during the evolution of vertebrates, generating entirely different genomic configurations in squamates and mammals 5' from the *Oct4* gene. *Oct4* would thereby serve as a link between the overall trunk length and patterning of the paraxial mesoderm, coordinating axial extension with the production of the appropriate skeletal structures for the trunk and tail regions through *Hox* genes.

III.3 Materials and Methods

II.3.1 Embryos

The *Cdx2-Oct4* transgenic construct contained the mouse *Oct4* cDNA (Osorno et al., 2012) cloned between the *Cdx2* enhancer (Benahmed et al., 2008) and the SV40 polyadenylation signal. The *Cdx2-Alk5*^{CA} transgenic construct has been previously described (Jurberg et al., 2013). All transgenics were produced by pronuclear injection (Hogan et al., 1994). Mouse embryos were recovered by cesarean section at different developmental stages and processed for whole mount *in situ* hybridization, β-galactosidase staining, histology or skeletal analysis. *Python regius* eggs were obtained from the python breeding colony at University of Florida. Eggs from *Pantherophis guttatus* (corn snake) were obtained from local breeders. Eggs were collected immediately after (corn snake) or slightly before (python) oviposition and embryos were dissected from the eggs and processed for *in situ* hybridization or RNA/DNA extraction. All experiments conducted on animals followed the Portuguese (Portaria 1005/92) and European (Directive 2010/63/EU) legislations, concerning housing, husbandry, and welfare. The

project was reviewed and approved by the Ethics Committee of "Instituto Gulbenkian de Ciência" and by the Portuguese National Entity, "Direcção Geral de Alimentação Veterinária" (license reference: 014308).

II.3.2 Genotyping

Embryos and fetuses were genotyped by PCR from genomic DNA extracted from yolk sacs and intestines or skin, respectively. Yolk sacs were incubated overnight at 50º C in yolk sac lysis buffer containing proteinase K (50 mM KCl, 10 mM Tris-HCl pH8.3, 2 mM $MgCl_2$, 0.45% Tween-20, 0.45% NP40, with 200 µg/mL of proteinase K). Lysates were heat-inactivated as above. Intestine and skin samples were incubated overnight at 50º C under agitation in Laird's buffer (100 mM Tris-HCl pH8.5, 5 mM EDTA, 0.2% SDS, 200 mM NaCl) supplemented with 100 µg/mL of proteinase K. Genomic DNA was then precipitated with isopropanol (1:1 vol:vol) and transferred to TE buffer (1 mM EDTA, 10 mM Tris-HCl pH8.0). PCR was performed using 1µl of the genomic DNA solution. Primers used for genotyping are summarized in Table III.1.

II.3.3 Phenotypic analysis

Skeletal analyses

Skeletal preparations were performed at E18.5 by alcian blue/alizarin red staining as previously described (Mallo and Brändlin, 1997). Briefly, fetuses were skinned, eviscerated and fixed in 100% ethanol at room temperature. They were stained with alcian blue 8 GX (150 mg/l in 20% acetic acid, 80% ethanol) at room temperature for 12 hours and postfixed in 100% ethanol for12 hours. Fetuses were cleared in 2% KOH for 8 hours at room temperature, stained with alizarin red S (50 mg/l in 2% KOH) for 2 hours and further cleared in 2% KOH for 10 to 16 hours at room temperature. The reaction was stopped with 25% glycerol and the stained carcasses stored in the same solution.

Histological analysis

Embryos were collected by cesarean section, dissected in PBS 1x (Dulbecco's Phosphate Buffered Saline, without calcium and without Magnesium, Biowest: 1.8 mM KH_2PO_4, 2.7 mM KCl, 10 mM Na_2HPO_4, 137 mM NaCl) and fixed in Bouin's fixative (Sigma) for one overnight (E10.5) or two days (E18.5). They were then dehydrated thoroughly in EtOH 100%, washed extensively in toluol and embedded in paraffin. Embryos were sectioned in 2 µm slices using a microtome and stained with hematoxilin and eosin.

Whole mount in situ *hybridization*

Expression analyses were performed by whole mount *in situ* hybridization (Kanzler et al., 1998) using *in vitro* transcribed digoxigenin-labelled antisense RNA probes. Briefly, embryos were dissected out in PBS and fixed with 4% paraformaldehyde (Sigma) (in PBS) at 4ºC overnight. Embryos were washed in PBT (PBS containing 0.1% Tween-20), dehydrated with methanol and rehydrated with PBT. They were then treated with proteinase K (10 µg/ml in PBT) at room temperature for 9 min, the reaction was stopped with glycine (2 mg/ml in PBT) and embryos were postfixed with 4% paraformaldehyde, 0.2% glutaraldehyde (Sigma). Hybridization was performed at 70ºC overnight in hybridization solution (50% formamide, 1.3 x SSC [3M NaCl, 300 mM sodium citrate] (pH 5.5), 5 mM EDTA, 0.2 % Tween 20, 50 µg/ml yeast tRNA, 100 µg/ml heparin) containing the RNA probe, followed by three washes at 70ºC in hybridization solution without RNA probe, tRNA and heparin. Embryos were then washed in TBST (25 mM Tris.HCl, pH 8.0, 140 mM NaCl, 2.7 mM KCl, 0.1% Tween 20), equilibrated with MABT (100 mM Maleic acid, 150 mM NaCl, 0.1 % Tween-20, pH 7.5), blocked with MABT/Block (MABT containing 1% blocking reagent (Roche #11096176001) and

Table III. 1 Oligonucleotides used in this chapter for genotyping, RT-PCR, corn snake Oct4 in situ hybridization probe cloning and python Oct4 upstream region sequencing.

Genotyping	*LacZ*	Fw	AGCAGTTTTTCCAGTTCCGTTTATC
		Rv	AGCGGCGTCAGCAGTTGTTTTTTAT
	Oct4	Fw	TGAAAGCCCTGCAGAAGGAG
		Rv	GGTGTCCCTGTAGCCTCATACTC
	Gdf11	Fw	GCATCCTTTCATGGAGCTTCG
		wt_Rv	CTGGCCGGAGCAGTAGTTGG
		mut_Rv	AGTAGAAGGTGGCGCGAAGG
	Alk5^{CA}	Fw	ACGTTCATGGTTCCGAGAGGC
		Rv	ATCATGTCTCACAGCAAGTCC
RT-PCR	*Py-Npdc1*	Py-Npdc1-RT-F1	GATGGCATACACAGCTACTCG
		Py-Npdc1-RT-R1	ACACCGTATATGAATGTCCTG
	Py-Npdc1/Oct4	Py-Npdc1-RT-F2	CATTCATATACGGTGTAGTCG
		Py-Oct4-RT-R1	CACGCTCCACTTTGATGTCAGGC
	Py-Oct4 5'	Py-Oct4-RT-F1	GCCTGACATCAAAGTGGAGCGTG
		Py-Oct4-RT-R2	TGTTCCAACTCCACTGAGGTG
	Py-Oct4 3'	Py-Oct4-RT-F2	GGCTTCACTCAGGCAGATGTG
		Py-Oct4-RT-R3	GGCCGACTCCATGGAGCACAG
	Py-Npdc1-8th exon	Py-Npdc1-RT-F2	CATTCATATACGGTGTAGTCG
		Py-Npdc1-RT-R2	CAGAAGATGACAATGTAGATGAG
Corn Snake Oct4 *in situ* hybridization probe	Exon 3	CS-Ex3-F	CTGTCGACGGAAGATGTTTAGTCAAACC
		CS-Ex3-R	CTGAATTCTCCTGCAAATTCTCATTACT
	Exon 4	CS-Ex4-F	CTGAATTCAACTGTGCTCCATGGAGTC
		CS-Ex4-R	GTACTAGTATCCTTCTCCAGGCTGAGGTC
	Exon 5	CS-Ex5-F	GTACTAGTGTCTGGTTTTGCAACCG
		CS-Ex5-R	CTGCGGCCGCAGCTTGAATGCATCGGGTG
Corn Snake *Gdf11 in situ* hybridization probe	Exon 2	Snk-Gdf11-ISH-F1	CAGTCGACGGACCCAGTGGTACAGATCG
		Snk-Gdf11-ISH-R1	CATCTAGACAGGCCTTCAGCACCAGGCCC
	Exon 3	Snk-Gdf11-ISH-F2	CATCTAGACACCCCTTCATGGAGCTGCGTG
		Snk-Gdf11-ISH-R2	GTGCGGCCGCTTAAGAGCATCCACACCTGTC
Oct4 5' sequencing	Py-Oct4-gen-F		GACTCGAGCAGCACGAGCCTTCCGAGAGG
	Py-Oct4-gen-R		CTGGATCCGGAAAGGGGTACCAGCTGTGAG

10% sheep serum) and incubated with a 1:2000 dilution of alkaline phosphatase-conjugated anti-digoxigenin antibody (Roche #11093274910) in MABT/Block at 4ºC overnight. Embryos were washed extensively with MABT at room temperature, equilibrated in NTMT (100 mM Tris HCl, pH 9.5, 50 mM MgCl2, 100 mM NaCl, 0.1% Tween-20) and developed at room temperature with an NBT/BCIP solution (Roche #11681451001) diluted in NTMT or BM Purple (Roche #11442074001). Reactions were stopped with PBT, fixed with 4% paraformaldehyde in PBS and stored in PBT. The mouse probes used in this work are summarized in Table III.2.

Post-staining embryo sectioning

To section whole mount stained embryos, they were fixed in 4 % paraformaldehyde, 0.2 % glutaraldehyde (in PBS) for 15 min at room temperature, and included in gelatin/albumin (0.45 % gelatin, 270 g/l bovine serum albumin (Roche), 180 g/l sucrose in PBS, jellified with 1.75 % glutaraldehyde). Sections were cut at 35 µm with a vibratome and mounted with an aqueous mounting solution (Aquatex, Merck).

II.3.4 β-galactosidase reporter analysis.

For reporter analyses, six different highly conserved regions 5' from the snake *Oct4* gene were amplified by PCR from python genomic DNA (the primers are the "Tra" series of those specified on Table S1) and cloned upstream of a cassette containing the adenovirus2 minimal late promoter, the *β-galactosidase* cDNA and the polyadenylation signal from SV40 (Jurberg et al., 2013). Embryos were collected at E10.5 by cesarean section, dissected in PBS and processed for β-galactosidase staining. Briefly, embryos were fixed with Mirky's fixative (National Diagnostics) overnight at 4ºC. After 3 washes in washing solution (0.02% NP40,

Table III. 2 *In situ* hybridization probes used in Chapter III.

Gene	Probe	Vector	Linerization	Polimerase	Probe size	Hydrolysis? (min/sec)
Corn snake (cs) Oct4	Fragment containing exons 3, 4 and 5 of corn snake *Oct4*.	pKS	SalI	T7	570	No
Corn snake (cs) Gdf11	Fragment containing two last exons of corn snake *Gdf11*.	pKS			700	No
Fgf8		pKs	ClaI	T3	785	No
Hand2		pcDNA1	EcoRI	Sp6	1100	50''
Hoxa10	IMAGE 6511608	pCMV-SPORT 6.1	EcoRI	T7	2600	5'35''
Hoxb6	Fragment containing full CDS.	pKS	XhoI	T7	675	No
Hoxc10	cDNA	pKS	EcoRI	T3	1091	50''
Hoxd10	IMAGE 6516538	pCMV-SPORT 6.1	EcoRI	T7	1331	2'6''
Hoxd11	cDNA fragment containing part of exon 1 and exon 2. From Denis Duboule's lab.	pSN	EcoRI	T7	706	No
Hoxd13	cDNA fragment containing part of exon 2 and 3'UTR. From Denis Duboule's lab.	pGEM	PvuII	T7	1520	3' 7''
Ptx1		pKS	BamHI	T7	577	No
Nkx1.2 (Sax1)	From cDNA. Fragment spanning the homeobox.	pKS	HindIII	T7	552	No
T/Brachyury	1.8 kb fragment containing the whole open reading frame (ORF).	pKS	NotI	T3	1800	4'
Tbx4			XbaI	SP6		2'
Uncx4.1	Fragment containing the entire CDS, including 5' and 3' UTR.	pSV-SPORT	SalI	T7	1500	3'2''

0.02% Tween 20, in PBS), embryos were incubated in X-gal staining solution (5mM $K_3Fe(CN)_6$, 5mM $K_4Fe(CN)_6 \cdot 3H_2O$, 2mM $MgCl_2$, 0.02% NP40, 0.02% Tween 20, in PBS, including 0.4mg/ml X-gal [Promega]) and the reaction was monitored regularly. After developing, the embryos were extensively washed in washing solution, fixated overnight in PFA and sectioned in the vibratome.

II.3.5 BAC (Bacterial Artificial Chromosome) transgenics

The BAC containing the Oct4 locus was obtained from a corn snake genomic library (kindly provided by Isabel Guerreiro and Denis Duboule). A BAC clone that included the *Npdc1* and *Oct4* genes was linearized with lambda terminase and used to produce transgenic mouse embryos.

II.3.6 RT-PCR analysis

Total RNA was extracted from python embryonic tissues using Trizol (Sigma-Aldrich). cDNAs were produced by reverse transcription using oligo dT priming and the presence of specific transcripts was in vestigated by PCR. The position of the primers is shown on Fig. 4 and listed in Table III.1. The identity of the PCR products was confirmed by sequencing.

II.3.7 Genomic analysis

Python, king cobra and mouse genomic sequences were obtained from public databases. The mouse genome corresponded to assembly GRCm38/mm10. Python sequences were obtained from contig NW_006534040 and the king cobra genome from contig AZIM01002363. Gaps in the published genomic sequence of the area around the 5' end of the first *Oct4* coding exon of python were filled sequencing of PCR-amplified fragments using primers Py-Oct4-gen-F and Py-Oct4-gen-R (Table III.1). Sequence comparisons were performed using ClustalW2.

We computed long global alignments of the *Oct4* regions (100Kb) from multiple species and represented sequence similarity by curve-based visualization using

the software mVISTA (Frazer et al., 2004) (http://genome.lbl.gov /vista/index .shtml). We used the default parameters from the VISTA browser plot to calculate conserved regions and to display VISTA graphs. The genomic regions of interest were extracted from NCBI genome assemblies using standalone Blast. The *Boa constrictor* genome assembly was downloaded from GigaDB repository (http://gigadb.org/site/index). We identified the scaffolds containing the genomic region where *Oct4* was located using tblastn with the anole and mouse *Oct4* (anole, XP_008120168; mouse, NP_038661) and *Tcf19* (anole, XP_008120167; mouse, AAH04617) as queries. Then we used blastdbcmd in order to parse 100-200kb from the syntenic regions containing *Oct4* and neighboring genes from each species (*Npdc1*, *Tcf19* and *Cchrc1*). Gene annotations from the corresponding scaffolds were extracted from NCBI and used for the VISTA analysis in order to identify coding and non-coding regions. Before using Python and King cobra annotations in VISTA, *Npdc1* and *Oct4* exon annotations were manually curated as described in the results section.

Genomic regions of interest and their corresponding gene annotation for *Anolis carolinensis* and *Pogona vitticeps* were downloaded using the equivalent tools available on their genome browsers. For *Pogona*: https://genomics. canberra.edu.au/gbrowse/gbrowse/pogona_pvi1.1/. For *Anolis*: http://www. ensembl.org/Anolis_carolinensis/Info/Index.

III.4 Results

III.4.1 Sustained Oct4 activity in axial progenitors extends the trunk

In the previous chapter, we showed that *Oct4* was ectopically expressed in a subset of cells in *Gdf11* mutant tails. As conditional inactivation of *Oct4* suggested a role for this gene in the maintenance of the primitive streak (DeVeale et al., 2013), we hypothesized that the longer trunks of *Gdf11* mutants could result from persistent Oct4 activity, which would keep axial growth in a trunk-generating

configuration. To test this, we produced transgenic embryos expressing *Oct4* under the control of the *Cdx2* enhancer (Benahmed et al., 2008) (*Cdx2-Oct4* transgenics), thus overcoming the normal progressive down-regulation of *Oct4* expression in the epiblast (Osorno et al., 2012). From the ten *Cdx2-Oct4* transgenic fetuses recovered at E18.5 nine had abnormal phenotypes that could be divided in two groups. The first group (4 embryos) showed a variable increase in the number of thoracic and lumbar segments in their axial skeletons. The most affected fetus of this group had 17 instead of 13 rib-containing vertebrae and 8 instead of 6 lumbar segments (Fig III.1A, C). These transgenic skeletons phenocopy to a large extent the main axial features observed in *Gdf11* mutant fetuses (McPherron et al., 1999) (Fig III.1B), showing that the longer trunks of *Gdf11* mutants could indeed result from an extended period of Oct4 activity.

The second group of *Cdx2-Oct4* transgenic embryos (5 embryos) was readily identified by the presence of a sacrococcygeal teratoma of variable sizes (Fig III.2, A-C). The axial skeleton in this group of transgenics was characterized by an abnormally large number of ribs (up to about 30 in the most strongly affected specimen), covering most of the body length and associated with the absence of recognizable sacral or caudal structures (Fig III.1D). Histological analyses of this type of transgenic revealed that their neural tubes also extended further posteriorly than in wild type fetuses (Fig III.2, E-H), thus fitting with the posterior extension of thoracic characteristics observed in *Cdx2-Oct4* transgenic fetuses. Remarkably, even in the most strongly affected fetuses, the neck and anterior thoracic segments seemed mostly normal (Fig III.1D). These observations indicate that while *Oct4* over-expression had little or no effect in its normal domain of activity, it interfered dramatically with the development of areas formed after the switch into tailbud-dependent extension. Overall, these anatomical patterns are consistent with a partial or total conversion of the posterior body into a trunk. In addition to teratomas and extended rib cages, this group of *Cdx2-Oct4* transgenics

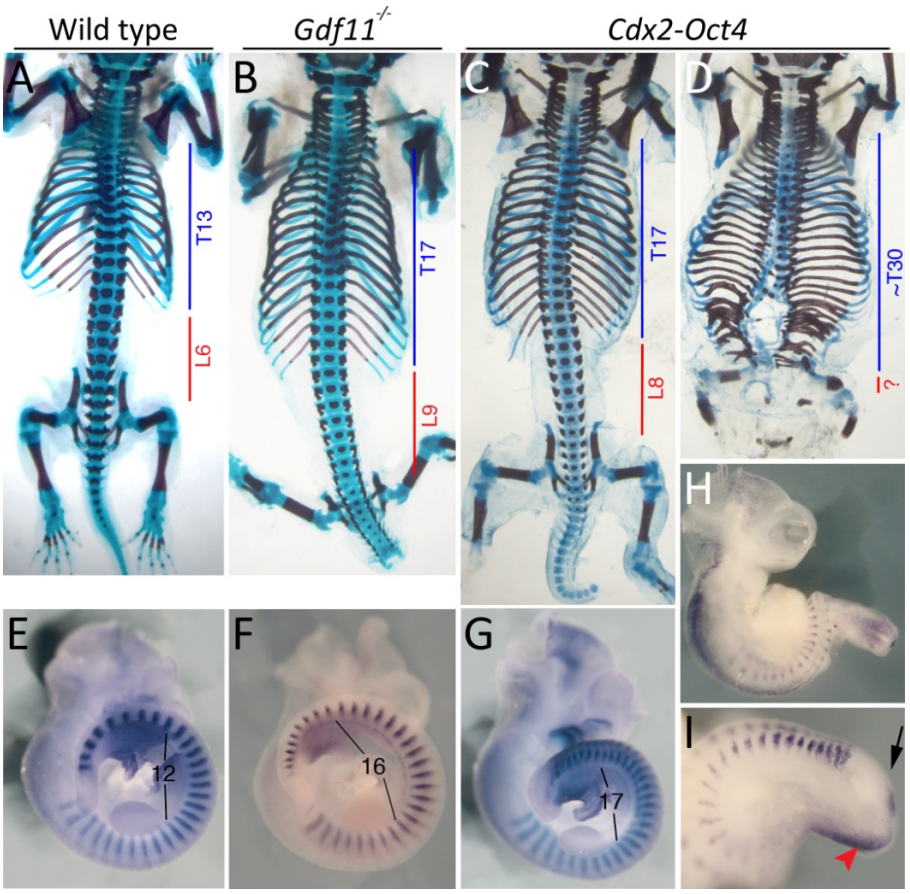

Fig III.1 Sustained *Oct4* expression in axial progenitors extends the trunks in mouse embryos. A-D. Skeletal analysis of a wild type (**A**), a *Gdf11*$^{-/-}$ (**B**) and two *Cdx2-Oct4* transgenic (**C**, **D**) embryos at E18.5. The number of thoracic (T) and lumbar (L) vertebrae is shown. **E-I.** Analysis of a wild type (**E**), *Gdf11*$^{-/-}$ (**F**) and two *Cdx2-Oct4* transgenic (**G-I**) embryos at E10.5 stained simultaneously for *Uncx4.1* and *Ptx1* (**E-G**) or with *Uncx4.1* and *Tbx4* (**H**, **I**). The number of interlimb somites is indicated. **I** shows a close up of the caudal end of the embryo in **H**. The embryo in **G** is expected to produce a skeleton similar to the one shown in **C** and the embryo in **H** and **I** is expected to produce a skeleton similar to that in **D**. The arrow in **I** indicates the posterior embryonic end showing the absence of a tailbud and the red arrowhead the position of the hindlimb/ventral mesoderm.

also had variable malformations in their hindlimbs. In all analyzed specimens, hindlimbs were found on the ventral side of the body with no connection with the

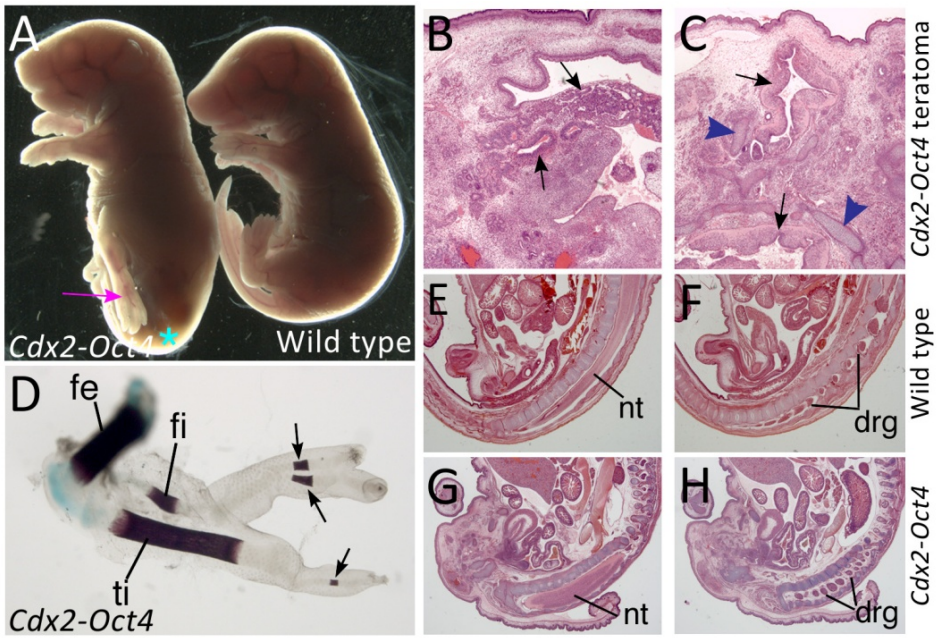

Fig III.2 Additional phenotypes of *Cdx2-Oct4* transgenics. A. External morphology of a strongly affected transgenic fetus at E18.5 (left). A wild type littermate is shown for comparison. The asterisk indicates the position of the sacrococcygeal teratocarcinoma and the pink arrow indicates a hindlimb. **B**, **C**. Two histological sections of a teratocarcinoma found in a *Cdx2-Oct4* transgenic embryo, stained with hematoxylin and eosin. In **B** the arrows show glandular tissue. In **C** the arrows indicate neural tissues and the blue arrowheads cartilages. **D.** A hindlimb of a strongly affected *Cdx2-Oct4* transgenic fetus. The femur (fe) is fairly well formed. The tibia (ti) and fibula (fi) are more affected. Also, there are just a couple of digit-like structures (arrows) connecting directly to the tibia and fibula. **E-H**. Sagittal sections through the caudal part of a wild type (**E**, **F**) or a *Cdx2-Oct4* transgenic fetus (**G**, **H**). In the wild type embryo, the neural tube (nt) finishes around vertebra 34, whereas in the transgenic the neural tube fills the vertebral canal until the very posterior end of the axial skeleton (in this embryo around vertebra 38). In both wild type and *Cdx2-Oct4* transgenics the neural tube is associated with dorsal root ganglia (drg).

axial skeleton (Fig III.1D and Fig III.2A). Skeletal hindlimb morphologies varied from almost normal to strongly malformed, most notably in the distal structures, where some of the bones were either missing or reduced in size (Fig III.1D and Fig III.2D).

At mid gestation stages (E10.5), *Cdx2-Oct4* transgenic embryos had malformations that, consistent with the skeletal phenotypes, were restricted to their posterior region. In these embryos, the distance between the fore and

hindlimb buds was increased to different extents, in accordance with the abnormally large numbers of thoracic vertebrae observed at E18.5 (Fig III.1, G-I). Some of these embryos resembled *Gdf11* mutants, further supporting the inverse functional link between Gdf11 and Oct4 activities. In addition to the posterior displacement of the hindlimbs, mildly affected *Cdx2-Oct4* transgenic embryos often displayed semi-bifurcated or truncated tails (Fig III.1F and G; Fig III.3, A-A', E-E', G-G', I-I', M-M'; Fig III.4B and E) and expanded expression domains of CNH-associated markers such as *T/Brachyury* (Fig III.3, A-A''), *Sox2* (Fig III.3, C-C''), *Fgf8* (Fig III.3, E-E''), *Foxa2* (Fig. III.3, G-G'), *Nkx1.2* (Fig III.3, I-I'), and *Wnt3a* (Fig III.3, K-K'). Furthermore, similarly to what was described in the previous chapter for *Gdf11*$^{-/-}$ embryos, some *Cdx2-Oct4* transgenic tails appeared to contain an ectopic ventral epithelial pocket that strongly expressed *T/Brachyury* (Fig III.3A'', white asterisk).

Another distinctive feature of *Cdx2-Oct4* transgenic embryos was the abnormal morphology of their posterior embryonic end, a trait most clearly observed at stages when their littermates had reached tail bud stages. Strongly affected embryos, such as those containing exceptionally long interlimb regions, lacked a recognizable tail bud and failed to close at the posterior end altogether, leaving a dorsally exposed epithelium that resembled the epiblast of younger wild type embryos (Fig III.1I; Fig III.3, K-K' and Fig. III.4, A-C, F-H). The epiblast-like nature of this epithelium was further evidenced by analysis of molecular markers such as *Fgf8, T/Brachyury* and *Nkx1.2* that at E10.5 were expressed in the open epithelium in patterns resembling those typical of the epiblast of E8.5-E9.0 wild type embryos (Fig III.4 A-A', B-B', C-C', F-F', G-G', H-H'). These embryos also produced lateral mesoderm up to the very posterior end of the main body axis, as revealed by *Hand2* expression. This contrasts with the absence of this mesodermal compartment in the tail region of wild type embryos (Fig III.4D, I).

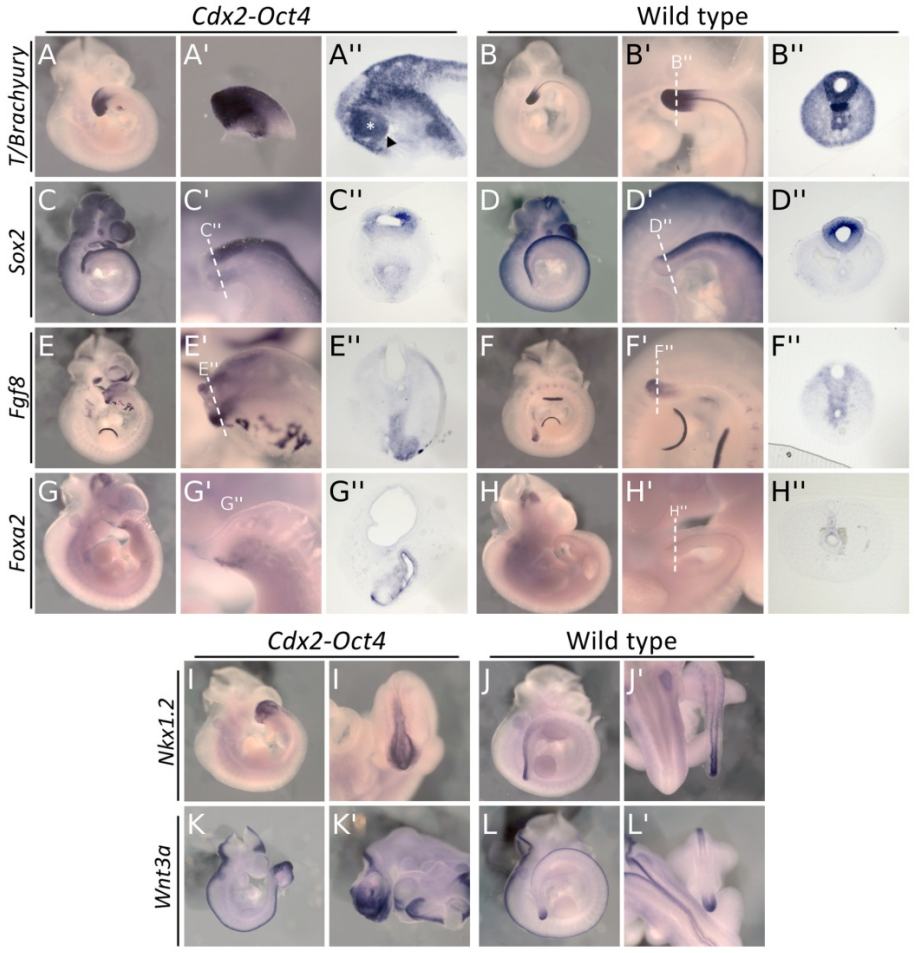

Fig III.3 Molecular characterization of *Cdx2-Oct4* transgenics. Genes associated to axial progenitor-containing regions were analyzed in *Cdx2-Oct4* (**A-A''**, **C-C''**, **E-E''**, **G-G''**, **I-I'**, **K-K'**) and wild type (**B-B''**, **D-D''**, **F-F''**, **H-H''**, **J-J'**, **L-L'**) E10.5 tails: *T/Brachyury* (**A-A''**, **B-B''**), *Sox2* (**C-C''**, **D-D''**), *Fgf8* (**E-E''**, **F-F''**), *Foxa2* (**G-G''**, **H-H''**), *Nkx1.2* (**I-I'**, **J-J'**) and *Wnt3a* (**K-K'**, **L-L'**). **A**, **B**, **C**, **D**, **E**, **F**, **G**, **H**, **I**, **J**, **K**, **L** show a wholemount view; **A'**, **B'**, **C'**, **D'**, **E'**, **F'**, **G'**, **H'**, **I'**, **J'**, **K'**, **L'** , a close-up; **A''** is a sagittal section through the tail in **A**; **B''**, **C''**, **D''**, **E''**, **F''**, **G''**, **H''**, **I''**, **J''**, **K''**, **L''** display transversal sections at the level indicated by the dashed lines. White asterisk and black arrowhead indicate the ectopic ventral epithelial pocket and the epithelial lining of this structure, respectively.

Taken together, skeletal, morphological and gene expression patterns in *Cdx2-Oct4* transgenics indicate that persistent Oct4 activity in the posterior epiblast is sufficient to keep embryonic extension of the main body axis in a trunk

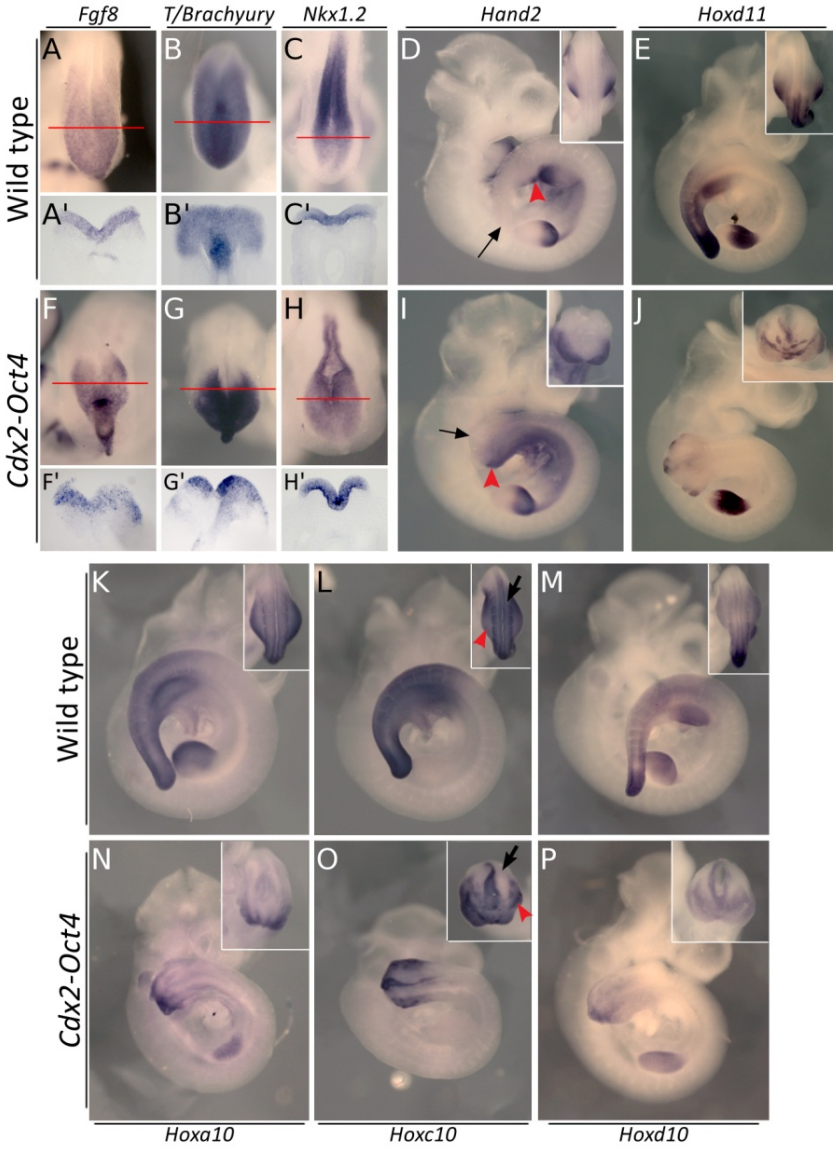

Fig III.4 Molecular characterization of *Cdx2-Oct4* transgenics *(cont)*. E10.5 wild type (**A-E** and **K-M**) and *Cdx2-Oct4* transgenic (**F-J** and **N-P**) embryos were analyzed by *in situ* hybridization with probes for *Fgf8* (**A-A'**, **F-F'**), *T/Brachyury* (**B-B'**, **G-G'**), *Nkx1.2* (**C-C'**, **H-H'**), *Hand2* (**D**, **I**), *Hoxd11* (**E**, **J**), *Hoxa10* (**K**, **N**), *Hoxc10* (**L**, **O**) and *Hoxd10* (**M**, **P**). **A**, **B**, **C**, **F**, **G** and **H** show dorsal close ups of the posterior end of the embryos. **A'**, **B'**, **C'**, **F'**, **G'** and **H'** show transverse sections through the areas indicated in the corresponding whole mount stained embryo. The black arrow in **D**, **I** indicates the posterior end of the embryos, whereas the posterior end of the lateral mesoderm is indicated with a red arrowhead. In all images involving *Hox* genes the inset shows a dorsal view of the posterior embryonic end to show that the paraxial mesoderm of the transgenics is mostly negative for posterior *Hox* gene expression. The arrows in **L** and **O** indicate the paraxial mesoderm and the red arrowhead the hindlimb.

-generating mode for longer developmental periods. Also, *Oct4* miss-expression seems to be the main factor contributing to the *Gdf11* mutant phenotype. This way, Gdf11 signalling seems to be required to counteract Oct4 activity as part of its transition-triggering program into a "tail mode" of development.

III.4.2 *Oct4* expression delays activation of posterior *Hox* genes

The remarkable increase in rib number observed in *Cdx2-Oct4* skeletons suggested delayed activation of posterior *Hox* genes (Carapuço et al., 2005; Mallo et al., 2010; Wellik and Capecchi, 2003). Expression analyses confirmed that *Hox* genes of the paralog group 10 were activated at more posterior levels than in wild type embryos, following the posterior shift in hindlimb location (Fig III.4, K-P; Fig III.5B, E, H). A closer look at these expression patterns indicated that activation of *Hox* group 10 genes was particularly delayed in the paraxial mesoderm of *Cdx2-Oct4* transgenics. Indeed, in these transgenic embryos, transcripts for *Hox* group 10 genes were barely detectable in the somites adjacent to the hindlimb, whereas in wild type embryos, the anterior expression limit of these genes in the somitic mesoderm roughly coincided with the anterior hindlimb border (Fig III.5, A-A', B-B'). In strongly affected *Cdx2-Oct4* embryos, activation of *Hox* group 10 genes was barely detectable in the paraxial mesoderm, albeit some expression was apparent in the emerging hindlimb buds and neural tube (Fig III.4, K-P). Expression of *Hox* genes belonging to more posterior paralog groups, like *Hoxd11* or *Hoxd13,* was also affected, following patterns similar to those observed for *Hox* group 10 genes (Fig III.4E, J; Fig III.5 E, H). Conversely, activation of more anterior *Hox* genes, like *Hoxb6*, was essentially normal (Fig III.5I, J). In general, these expression patterns resembled the delayed activation of posterior *Hox* genes observed in *Gdf11* mutants (Jurberg et al., 2013; McPherron et al., 1999; Szumska et al., 2008) and were complementary to the early activation of the same *Hox* genes in *Cdx2-Alk5*^{CA} embryos that presented a premature induction of the trunk to tail transition

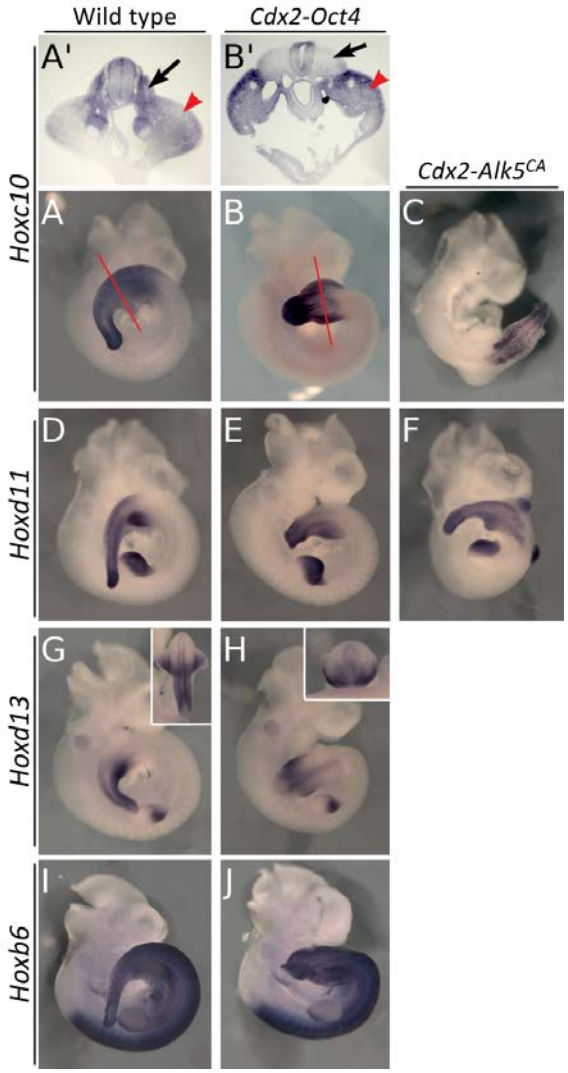

Fig III.5 *Hox* gene expression in *Cdx2-Oct4* transgenic embryos. Expression of *Hoxc10* (**A'**, **B'**, **A-C**), *Hoxd11* (**D-F**), *Hoxd13* (**G**, **H**) and *Hoxb6* (**I**, **J**) was analyzed in E10.5 wild type (**A**, **A'**, **D**, **G**, **I**), *Cdx2-Oct4* (**B**, **B'**, **E**, **H**, **J**) or *Cdx2-Alk5*^{CA} (**C**, **F**) embryos by whole mount *in situ* hybridization. **A-F** show a comparison of two posterior Hox genes, *Hoxc10* and *Hoxd11*, in *Cdx2-Oct4* embryos with a milder phenotype (more similar to those of *Gdf11* mutant embryos, with a posterior displacement of the hindlimb position by about 4 somites) and in *Cdx2-Alk5*^{CA} embryos, in which these genes are activated more anteriorly following the premature trunk to tail transition. **A'** and **B'** show sections through the indicated region in **A** or **B** to indicate the presence of *Hoxa10* expression in somites (arrows) adjacent to the hindlimb (red arrowheads) in wild type embryos and its absence in *Cdx2-Oct4* transgenics. Expression of *Hoxd13* was shifted posteriorly following the position of the hindlimbs. In the paraxial mesoderm, expression was almost undetectable. The anterior expression border of *Hoxb6* expression is not altered in *Cdx2-Oct4* transgenic embryos.

(Jurberg et al., 2013) (Fig III.5C, F). These data further support an inverse functional connection between Gdf11 signalling and Oct4.

III.4.3 Oct4 expression is maintained for longer developmental times in snake embryos

Long rib cages, extended production of lateral mesoderm and delayed activation of posterior *Hox* genes, as observed in *Cdx2-Oct4* transgenic embryos, are also the very same traits found in snakes (Di-Poï et al., 2010; Woltering, 2012; Woltering et al., 2009), suggesting that their elongated trunks might result from sustained Oct4 activity for longer developmental times. To test this hypothesis, we investigated *Oct4* expression by *in situ* hybridization in corn snake embryos and found that *Oct4* continued to be transcribed in the posterior part of the embryonic trunk region at a stage when the tailbud starts becoming evident (Fig III.6A, B), in sharp contrast with equivalently staged mouse embryos where *Oct4* expression had already disappeared completely (Osorno et al., 2012). These results were confirmed by RT-PCR, using python trunk cDNA as template (Fig III.6H, I). This indicates that *Oct4* expression persists for longer developmental times in snake embryos than in mammalian embryos and that heterochronic changes in gene expression could thus be involved in production of the extended, organ-filled trunks characteristic of snakes.

III.4.4 Genomic organization of the mouse and snake Oct4 loci

To understand the possible origin of the differences in snake and mammalian *Oct4* expression, we compared the chromosomal environment of *Oct4* in the genomes of a basal snake, python (Castoe et al., 2013), and the mouse (the mouse genome will be used as reference, but other mammals have similar configurations in this area) (Fig III.6C, D). The chromosomal organization downstream of the *Oct4* locus was fairly similar in python and mouse, with the presence of the *Tcf19* and *Cchcr1* genes. However, the python and mammalian genomes seemed to have

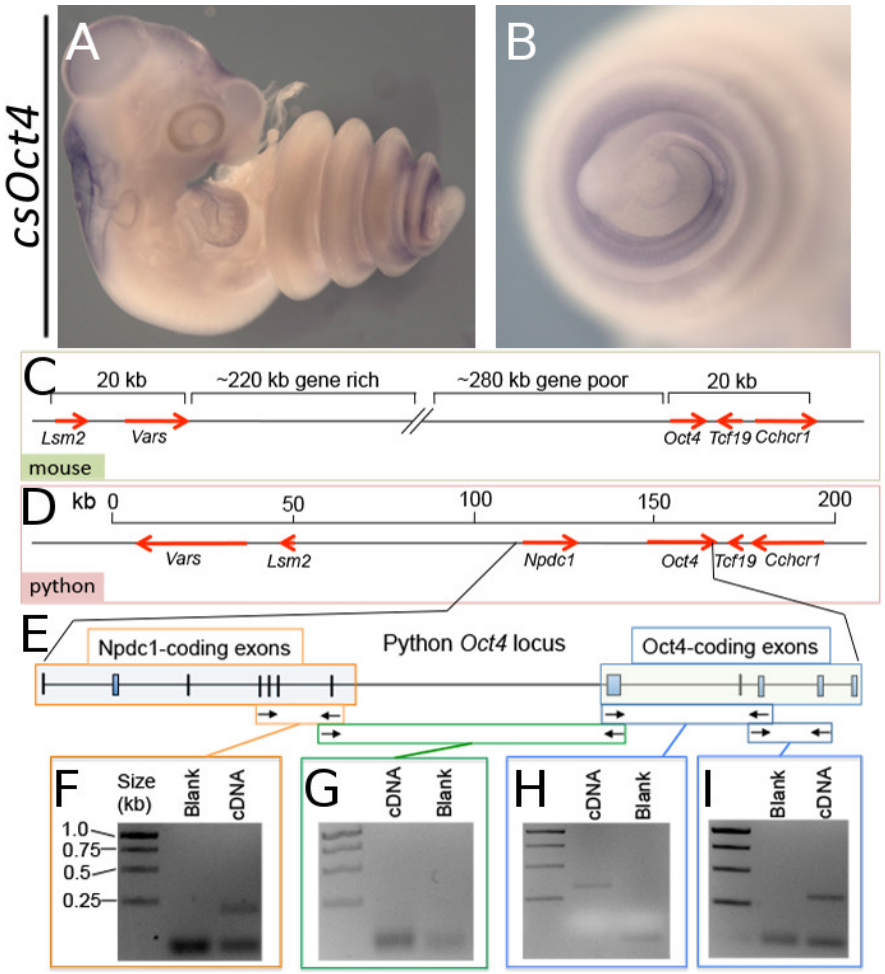

Fig III.6 *Oct4* expression and genomic environment in snake embryos. A, B. *Oct4* expression in a corn snake embryo shortly after it underwent trunk to tail transition. **B.** close up of the lower trunk/tail region. **C.** Structure of the genomic area surrounding the *Oct4* locus in mouse. **D.** Structure of the genomic area surrounding *Oct4* locus in python. In **C** and **D** the coding regions and their transcriptional orientation are indicated with red arrows. **E.** Close up of the region containing *Npdc1* and *Oct4*-coding exons in the python. Exons are represented as boxes. **F-I.** RT-PCR analysis of transcripts synthesized from the python *Npdc1* and *Oct4* transcription units using primer sets Py-Npdc-RT-F1 and Py-Npdc-RT-R1 (**F**), Py-Npdc-RT-F2 and Py-Oct4-RT-R1 (**G**), Py-Oct4-RT-F1 and Py-Oct4-RT-R2 (**H**) and Py-Oct4-RT-F2 and Py-Oct4-RT-R3 (**I**). The position of the primers is indicated in **E**.

lost synteny upstream of the *Oct4* locus. The only existing similarity between these two genomes in this area was the presence of the *Lsm2* and *Vars* genes, located about 500 kb upstream of *Oct4* in the mouse and about 60 kb in the python. Between these genes and *Oct4,* the mouse genome has a 220 kb gene

rich region next to *Lsm2*, followed by an almost gene-free 280 kb region containing just a few scattered histocompatibility complex genes. In python, the region between the *Vars* and *Oct4 loci* contains only exons coding for a protein with high homology with *Npdc1*, some 20 kb upstream of the exons homolog to *Oct4*. The presence of *Npdc1* next to *Oct4* might be a general characteristic of snakes, as it is also present in the king cobra genome (Vonk et al., 2013). In mammals, *Npdc1* and *Oct4* are located in different chromosomes (e.g. mouse chromosomes 2 and 17 for *Npdc1* and *Oct4*, respectively). This indicates that the area 5' from the *Oct4 locus* underwent divergent reorganization after the divergence of snakes and mammals, which might have influenced *Oct4* regulation or even activity in these two taxa.

The annotated *Npdc1* region of python (Castoe et al., 2013) suggested that this gene had lost the last coding exon, which was also not identified in the annotated king cobra genome (Vonk et al., 2013). Actually, in the king cobra *Npdc1* annotation, an additional exon was missing but we could identify it upon comparison with the python genome (Fig III.7A, B). The absence of the last *Npdc1* exon could have promoted an exon-shuffling event, bringing the *Oct4*-coding exons into the *Npdc1* transcript and creating a fusion between the two genes/proteins that could have affected *Oct4* regulation and/or activity. Such a chimeric product was suggested in the python annotation (Castoe et al., 2013). We directly assessed this possibility by searching for the different transcripts potentially produced from this genomic region. By RT-PCR, we detected mRNAs derived from exons with homology to *Npdc1* or to *Oct4* (Fig III.6F, H, I). However, we were unable to amplify PCR products compatible with the existence of the hypothetical chimeric transcript containing *Npdc1* and *Oct4* exons (Fig III.6G). Yet, the free donor and acceptor splice sites in *Npdc1* and *Oct4* exons suggested in the python annotation indicated either incomplete annotation or the existence of

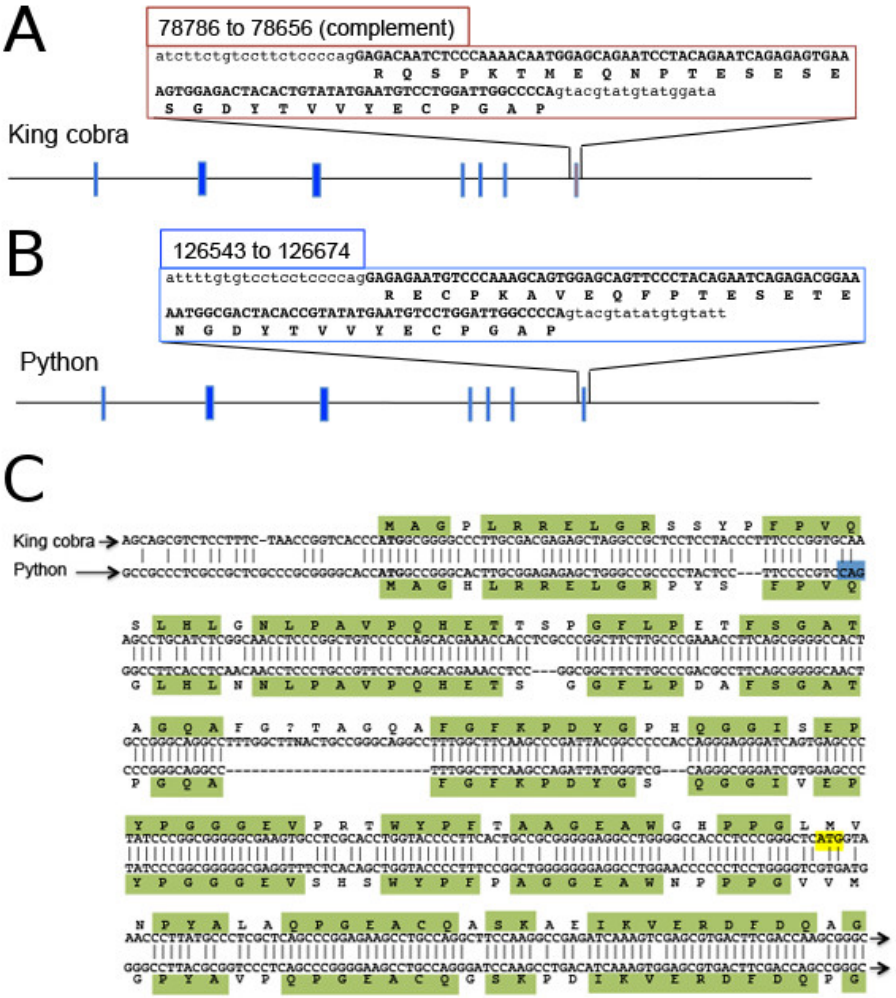

Fig III.7 Comparison of king cobra and python *Npdc1* and *Oct4* sequences. A. Sequence of the seventh *Npdc1* exon from python (bold). The coordinates of the fragment in contig NW_006534040 and the sequence of the protein fragment encoded by this exon are also shown. **B.** Identification of the "missing" seventh *Npdc1* exon from king cobra through alignment with the python genome (coordinates in contig AZIM01002363). **C.** Identification of the N-terminal end of snake Oct4 proteins. This area of the python genome contained sequence gaps that were curated by amplification and sequencing. The resulting sequence provided an extension of the open reading frame with high homology with the corresponding area of the king cobra. For the alignment, a small gap (suggested 4 nucleotide-long) in the published king cobra sequence was disregarded to avoid a frame change. Nucleotide sequence similarity was lost upstream of the ATG. Indicated are the starting ATG of king cobra as described in the genome annotation (yellow). Also indicated is the position of the acceptor site (blue) according to the python annotation.

additional exons that might have compromised our search for the chimeric transcript. Comparison of the python and king cobra genomes revealed the presence of a ~0.8 kb-long highly conserved region about 3 kb downstream of the last annotated *Npdc1* exon (Fig III.8B). This sequence contained a potential splice acceptor that, when spliced to the last *Npdc1* exon, would extend the open reading frame by 21 amino acids with high homology to the corresponding area of the mouse protein (Fig III.8B, D). We confirmed that this region is a bona fide *Npdc1* exon by RT-PCR (Fig III.8C). Therefore, the structure of the python and king cobra *Npdc1* resembles that of its mammalian homolog.

Annotation of the king cobra *Oct4* differed from that of python in two main aspects: no splice acceptor was identified at the 5' end of the first exon and the Oct4 protein was given a start site about 250bp downstream of the splice acceptor annotated in python (Fig III.7C). However, the king cobra sequence upstream of the mapped start codon showed strong homology with the corresponding python sequence both in the nucleotide sequence and predicted translation product (Fig III.7C), suggesting that it is part of the *Oct4*-coding region. Consistent with this, comparison of the two sequences after filling the gaps in the published sequences revealed the presence of an in frame ATG that marks the start of the conservation between the two snake sequences (Fig III.7C). The protein encoded from this start codon bares almost no homology with the mouse protein other than the first five amino acids (Fig III.8E). This contrasts with the high conservation exhibited by parts of the protein encoded by further downstream exons. We confirmed by RT-PCR that this region is a genuine exon of the *Oct4* transcript (Fig III.6H). This indicates that the differential rearrangements at the 5' end of the *Oct4* gene led to a different selection of the first exon(s). Considering that the enhancers controlling *Oct4* expression in the mouse are relatively close to its transcription start site (Yeom et al., 1996), the differences in

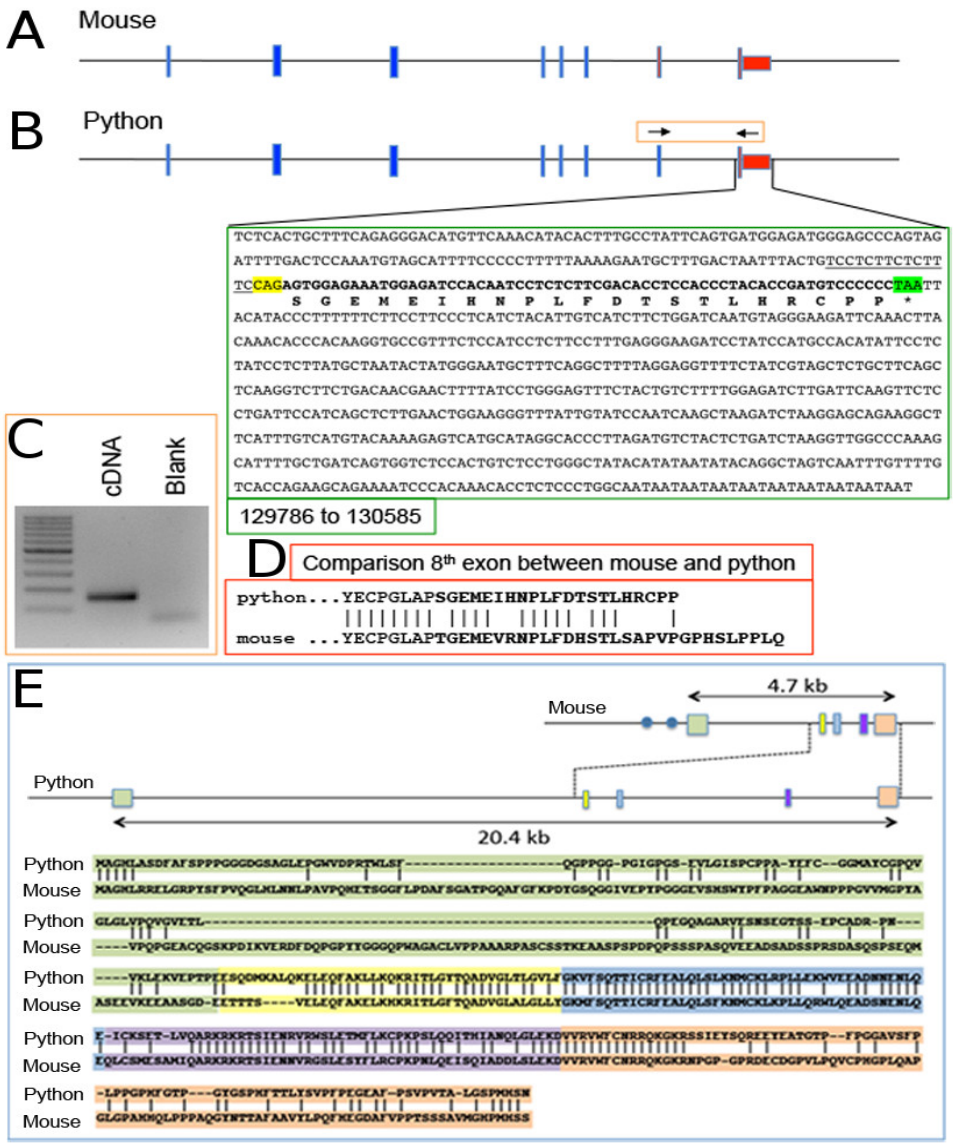

Fig III.8 Comparison of last *Npdc1* exons and first *Oct4* exons between python and mouse. A. Schematic structure of the mouse *Npdc1* gene. **B.** Eighth python *Npdc1* exon. The sequence includes the end of the open reading frame (bold) and the 3' UTR. The king cobra has a similar sequence but is not shown in this figure. The splice acceptor site (in yellow), the termination codon (in green) and translation products are indicated, together with the coordinates of the exon in contig NW_006534040. **C.** RT-PCR analysis showing the existence of the eighth *Npdc1* exon in python using oligonucleotides form the regions indicated in the orange box in B. **D.** Comparison of the protein fragments encoded by the last exons of the mouse and python *Npdc1* genes. **E.** Comparison of the mouse and python Oct4 proteins. Regions encoded by the different exons are color-coded. Note the striking difference in exon1-encoded peptides, which contrasts with the high similarity in the rest of the molecules.

the genomic region 5' of *Oct4* might have also impacted its regulation. Indeed, we were unable to detect any regions in the python genome with homology to the first *Oct4* exon of mammals or to mammalian *Oct4* regulatory regions (Fig III.9B). Similarly, a BLAST search with the sequence of the first python *Oct4* exon failed to identify significant homology in mouse databases.

III.4.5 Organization of the Oct4 locus in lizards

To further explore the role that the genomic organization at the *Oct4* locus might have played in *Oct4* expression during snake embryonic development, we first analyzed the extent of structural and sequence conservation around the *Oct4* locus between lizards and snakes. Analysis of the *Anolis* genome was inconclusive; although in the available annotated sequence there was no reference to *Npdc1* next to *Oct4*, the sequence contains a 30 kb gap where *Npdc1* might be positioned according to the snake genome. However, a VISTA analysis using available genomic sequence data from gecko and glass lizard revealed the presence of *Npdc1*-coding exons in the area 5' from *Oct4*, indicating that the rearrangement placing *Npdc1* upstream of *Oct4* occurred before the snake/lizard divergence (Fig III.9A). Interestingly, addition of other snake and lizard species to the analysis showed substantial differences in this area between snakes and lizards. In particular, while homology among representatives of both squamate groups was mostly restricted to coding regions, the genome of all snake species displayed extensive homology in non-coding regions as well. A similar VISTA analysis performed taking gecko genomic sequence as the reference, revealed that homology among lizards in non-coding regions was not substantially higher than between gecko and python (Fig III.9C). These results indicate that while the general gene structure around the *Oct4* locus is similar in snakes and lizards, non-coding regions within this genomic area were subject to differential evolutionary constraints in these two squamate lineages. Also, the extensive sequence

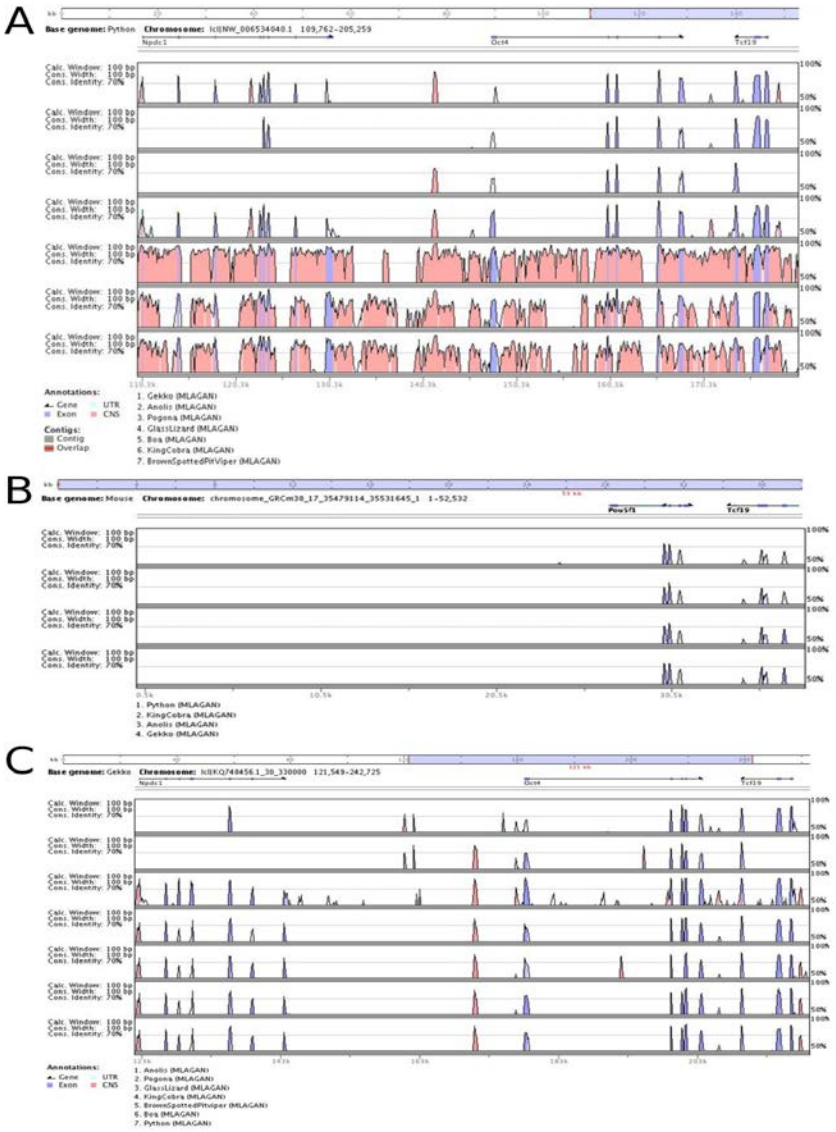

Fig III.9 Comparison of *Oct4*-containing genomic regions of different species. A. Sequence comparison in the vicinity of the *Oct4* locus in snakes and lizards. The genomic sequences from the area including *Oct4* from gecko, *Anolis*, *Pogona*, glasslizard, boa, king cobra and brown spotted pit viper were plotted against the corresponding area of python using a VISTA software. Represented in blue are homologies within coding exons and in red conservation in non-coding regions. **B.** The mouse genomic region was compared with those of python, king cobra, Anolis and Gekko using VISTA. Homologies seem to be reduced to the coding exons, with the exception of the first *Oct4* exon. **C.** Comparison of the Gekko sequence with those of other lizards (Anolis, pogona, glasslizard) and snakes (king cobra, brown spotted pitviper, boa and python). Homology outside the coding regions was much reduced.

homology in non-coding regions around *Oct4* in snakes suggests the existence of highly conserved regulatory information.

III.4. 6 Gdf11 in squamates

Our data from mice indicate functional interactions between Oct4 and Gdf11 signalling during axial extension and, therefore, we explored whether modifications affecting Gdf11 could also have contributed to the snake body plan. Comparison of the *Gdf11*-containing genomic region in mice and snakes revealed that this region underwent differential rearrangement involving an inversion that affected the genomic context around *Gdf11* (Fig III.10A). Despite this rearrangement, *Gdf11* expression in snake embryos kept features compatible with the existence of a balance between Oct4 and Gdf11 signalling activities similar to that observed in mice, as it was restricted to the embryonic area posterior to the *Oct4* expression domain (Fig III.10B).

As with the *Oct4* genomic region, lizards also shared the global genomic structure with snakes at the *Gdf11* locus. Interestingly, a VISTA analysis revealed a high degree of similarity among squamates in this genomic area, including both coding and non-coding regions (Fig III.10C). This clearly contrasted with the strong differential conservation in non-coding regions observed between snakes and lizards at the *Oct4* locus, which highlights the possible importance of the conserved regions at the *Oct4* locus for the production of the snake-like pattern of *Oct4* expression.

III.4.7 Regulation of Oct4 expression in snakes

In an attempt to understand the regulatory potential of the region upstream of the snake *Oct4* locus we tested several highly conserved regions for their ability to activate reporter expression in transgenic mouse embryos. Only two of the tested regions (see methods) seemed to be active in our assay. Interestingly, both

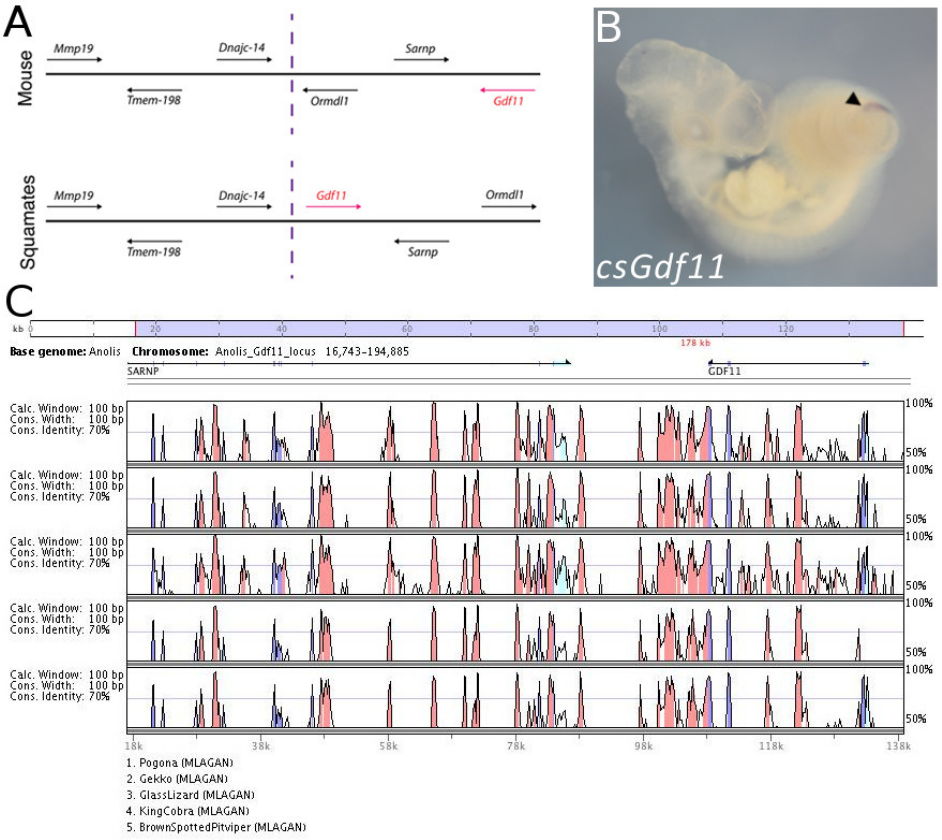

Fig III.10 *Gdf11* in squamates. A. Gene structure around the *Gdf11* locus in mammals and squamates, showing the inverted orientation of the region containing the *Gdf11*, *Sarnp* and *Ormdl1* genes. **B.** Whole mount *in situ* hybridization with a probe against *Gdf11* in corn snake embryos, showing expression in the tailbud (black arrowhead). **C.** Sequence comparison in the vicinity of the *Gdf11* locus in snakes and lizards. Compared are sequences from *Anolis*, *Pogona*, gecko, glasslizard, king cobra and brown spotted pit viper using the VISTA software. Represented in blue are homologies within coding exons and in red conservation in non-coding regions.

regions included the only non-coding fragment (~250bp) showing significant homology between snakes and lizards according to the VISTA analysis.

On its own, this element was very active as an enhancer in mouse embryos, inducing reporter expression in neural tissues and neural crest along most of the AP axis, including head and body structures (Fig III.11A, a-a'', b-b''). This pattern was consistently observed in five different transgenic embryos, only with slightly different levels of intensity, indicating that it likely reflects the element's

regulatory potential in the context of a mouse embryo. Clear expression in the paraxial mesoderm was also observed in the most posterior part of one embryo, but this expression was much reduced in the other embryos (Fig III.11A, a'-a'', b'-b''). The lateral plate mesoderm was negative in all cases.

In snakes, this 250bp fragment is part of a larger segment of very high sequence conservation (about 1.2 kb). This larger fragment was also active in transgenic mice but in a much reduced spatial domain than the 250bp element alone (Fig III.11A, c-c''). It also activated reporter expression in the neural tube and neural crest, but this was restricted to specific areas of the hindbrain and to the spinal cord corresponding to the posterior part of the embryo. This pattern was consistent in four different embryos, indicating that it represents the regulatory potential of this element in mouse embryos. The more restricted activity of the 1.2kb fragment suggests that the full potential of the 250bp sequence might be affected by surrounding conserved snake sequences. We therefore explored the expression patterns produced by the 250bp element when embedded in a more complete snake-like genomic context by generating transgenic embryos with a BAC containing a corn snake genomic region spanning over the *Oct4* and *Npdc1* loci. None of the six transgenic embryos analyzed by in situ hybridization gave a signal for the corn snake *Oct4* gene above what was observed in non-transgenic mouse embryos (Fig III.11B). Together, these results indicate that the region around *Oct4* might contain discrete enhancer elements able to activate transcription in the progenitors of axial structures but that their activity might be under the control of additional regulatory elements.

III.5 Discussion

Oct4 has been extensively studied in the context of its function as a major regulator of pluripotency (Shi and Jin, 2010). We show here that this gene is also a key regulator of trunk length during vertebrate development. It has been

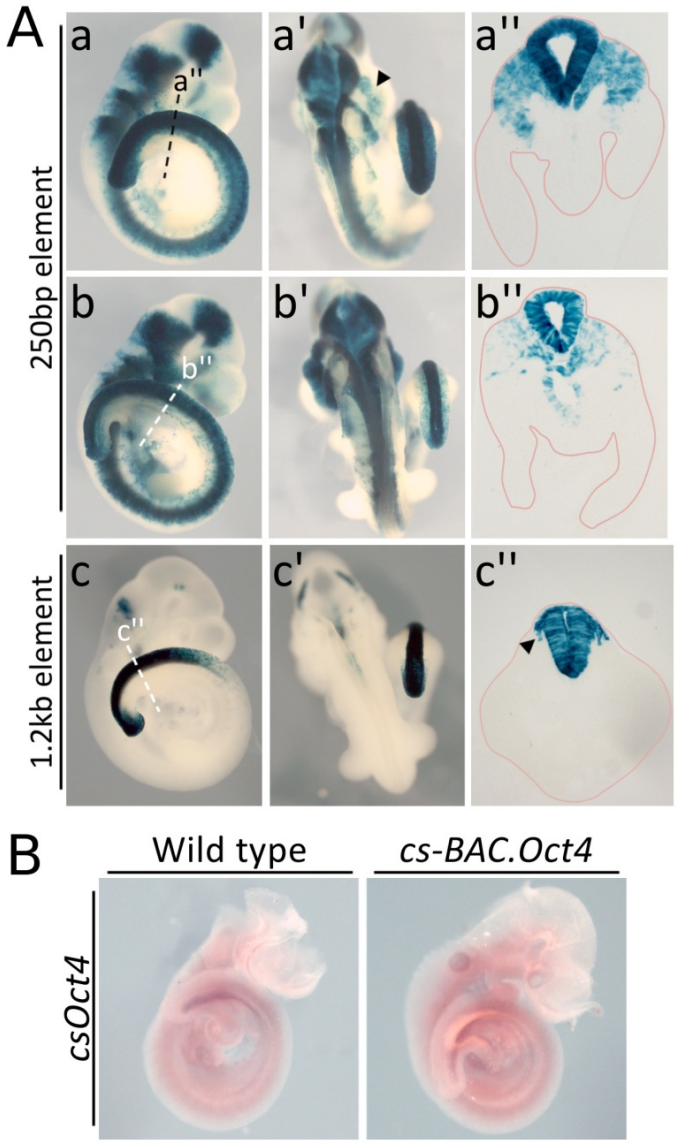

Fig III.11 Regulatory activity within the snake *Oct4* locus. A. β-galactosidase reporter assay to explore the capacity of the conserved 250 bp region (a-a'', b-b'') or the larger 1.2 kb fragment (c-c'') from the genomic region upstream of the snake *Oct4* gene. Lateral (a, b and c) and dorsal (a', b' and c') views of a transgenic embryo stained for β-galactosidase activity. a'', b'' and c'' show sections through the areas indicated in a-c, respectively. The black arrowhead indicates neural crest cells. The outline of the tissue is contoured in red. **B.** Whole mount *in situ* hybridization with a probe against corn-snake *Oct4* mRNA, on wild type or transgenic mice embryos for a BAC containing the *Oct4* locus (*cs-BAC.Oct4*).

previously shown that Gdf11 activity promotes the transition from trunk forming to tail forming mechanisms (Jurberg et al., 2013; Liu, 2006; McPherron et al., 1999; McPherron et al., 2009). These observations, together with the data presented in this chapter, suggest that the balance between Gdf11 and Oct4 activities is a determining factor that regulates regionalization of the vertebrate body into trunk and tail domains. This is consistent with the observation that

experimental changes in Oct4 or Gdf11 activities produced complementary shifts in the trunk to tail transition. The mechanistic details of this interaction are currently unknown. However, it has been shown that in embryonic stem cells Smad3 is directed to its targets in the genome through interactions with Oct4 (Mullen et al., 2011). It is therefore possible that Gdf11 signalling might generate a Smad or Smad-like product that counterbalances Oct4 activity instead of promoting a positive functional interaction. Whether or not this is the case remains to be determined.

One of the most interesting consequences of our findings is the possibility that changes in *Oct4* regulation might have played a central role in the evolution of the vertebrate body plan. Our results indicate that long trunks could be a consequence of different *Oct4* regulatory mechanisms operating in snake taxa that would keep its expression at such levels as to maintain axial growth in a trunk-forming mode for extended developmental periods. Genomic analyses suggest that this resulted from genomic rearrangements involving an extensive region upstream of the *Oct4* locus. These changes might have occurred in sequential steps. A first phase, which seemed to have occurred at an early stage of squamate evolution, involved major rearrangements resulting in the general gene structure observed in this particular chromosomal region. This rearrangement might have had significant impact on the basic *Oct4* regulatory landscape, to the extent that it seemed to have resulted in a different selection of the first *Oct4* exon in squamates and mammals. After this initial major event, the *Oct4* locus diverged considerably in snakes and lizards as estimated by sequence comparisons involving this genomic region. The observation that snakes exhibit remarkable sequence conservation in the non-coding regions around Oct4, which is not shared with lizards, is particularly relevant. Considering that conserved non-coding regions are often part of regulatory processes, it is possible that this area contains elements associated with a common snake trait or function, which could

include regulation of *Oct4* expression during axial extension. The finding that one of these conserved regions was indeed able to activate transcription in transgenic mouse embryos in a pattern compatible with activity in a subset of axial progenitors provides some support to this idea. Interestingly, the most active sequence in the mouse transgenic assay was the only non-coding region upstream of *Oct4* that seemed to have significant homology between snakes and lizards. This could indicate that the element is part of an ancestral regulatory network resulting from the initial recombination event, whose activity was later differentially modulated by additional regulatory elements. Our observation that the activity of this element was substantially different when tested alone or embedded within a larger snake-derived genomic context is consistent with this hypothesis, at least with respect to the snake locus. It is actually somewhat surprising that given the strong intrinsic activity of the 250bp element, the BAC containing the whole genomic region seemed to be unable to activate the snake *Oct4* in mouse embryos. This could indicate that proper control of this and/or other relevant elements requires factors not present in the appropriate combination or in the right spatial-temporal pattern in mouse embryos. This possibility would fit with the extreme molecular adaptations that seem to have occurred in snakes (Castoe et al., 2013; Vonk et al., 2013). If this is the case, the mouse transgenic approach will be rather limited to understand *Oct4* regulation in snakes; a direct analysis of chromatin interactions and histone modification profiles within the *Oct4* locus obtained from snake embryos might be a more appropriate strategy to identify key elements of the snake *Oct4* regulation landscape.

If the 250bp element is involved in the snake-like type of *Oct4* expression, its presence in lizards opens additional questions. One possibility is that this element plays no role in *Oct4* expression in snakes, but is required for another function shared by snakes and lizards. An alternative possibility is that the genomic

environment brought this element into a different function in lizards. One such function could be associated with the tail regeneration capacity characteristic of many lizard species. Regeneration processes typically involve reactivation of progenitors and/or dedifferentiation of somatic cells into proliferative cells (Foglia and Poss, 2016; Poss, 2010) and a factor with the characteristics of Oct4 might have been co-opted for those types of processes. If this would indeed be the case, the 250bp element might fall into the recently described category of regeneration enhancer elements (Kang et al., 2016).

In addition to sustained *Oct4* expression, our experiments also suggest that successful trunk extension in snakes must have required substantial changes in its tail-promoting mechanisms. In particular, the phenotypes of *Cdx2-Oct4* transgenics show that areas of the main embryonic axis that naturally form trunk structures tolerate increased levels of *Oct4* and develop without major problems. However, as soon as the embryo starts laying down the caudal-most areas, tail-promoting factors seem to start taking hold of morphogenesis of these regions. As a consequence, although *Oct4* is still able to promote recognizable trunk structures, these become progressively more disorganized and cells eventually reach a developmental dead-end in large sacrococcygeal teratomas. The presence of such tumours has also been described in other experimental settings upon ectopic *Oct4* expression (Economou et al., 2015). These observations highlight the need of a tight control of *Oct4* expression and of its interactions with other patterning factors to guide the differentiation potential of progenitor cells towards proper physiological routes. A variety of studies have shown a close association between activation of posterior *Hox* genes and the position of the trunk to tail transition (Carapuço et al., 2005; Di-Poï et al., 2010; Jurberg et al., 2013; McPherron et al., 1999; Szumska et al., 2008; Woltering et al., 2009). Given the role that these genes play in anterior-posterior patterning processes (Pearson et al., 2005), these observations suggested a possible role for *Hox* genes in the

evolution of vertebrate trunk length. However, functional assays in the mouse failed to support this hypothesis, as loss and gain of function experiments involving *Hox* genes consistently failed to produce a significant change in the hindlimb position (a proxy for the position of the trunk to tail transition) (Carapuço et al., 2005; Jurberg et al., 2013; Vinagre et al., 2010; Wellik and Capecchi, 2003). Nonetheless, what those experiments clearly showed was that *Hox* genes are crucial in the specification of skeletal features typically associated with the different body sections (Mallo et al., 2010), indicating that both processes must be coordinated. The finding that *Oct4* suppresses posterior *Hox* gene activation when promoting trunk formation, whereas *Gdf11* stimulates their expression when inducing tail development, suggests that the Oct4/Gdf11 system is the primary determinant of global trunk or tail growth modes and that *Hox* genes are then used to transmit patterning information into the mesodermal and neural derivatives of axial progenitors. This, in turn, ensures that all appropriate neural and skeletal elements for the trunk or tail bud-derived regions of the body are properly formed, positioned and coordinated during axial extension. The close association of growth modes with subsequent skeletal patterning increases not only the system's robustness, but also creates developmental modules that could be readily acted upon by selection during the course of evolution.

The finding that both Oct4 and Gdf11 signalling are able to modulate expression of posterior *Hox* genes is also interesting from a regulatory perspective. It will be important to understand how these two activities fit within known regulatory landscapes of *Hox* genes (Darbellay and Duboule, 2016) and whether their regulatory capacity is implemented by direct interaction with discrete elements within or around the *Hox* clusters or indirectly through the control of additional factors that will then modulate *Hox* cluster activity.

III.6 Acknowledgements

We would like to thank Val Wilson for providing *Oct4* cDNA, Jose Belo, Denis Duboule, Andreas Kispert, Andy McMahon, Erik Olson and Cliff Tabin for *in situ* probes, Denis Duboule and Isabel Guerreiro for the corn snake BAC library and Monica Dias, Florence Janody, Ana Stankovic, Miguel Manzanares and members of the Mallo lab for useful comments during the course of this project. This work has been supported by grants PTDC/BEX-BID/0899/2014 (FCT, Portugal) and SCML-MC-60-2014 (from Santa Casa da Misericordia de Lisboa, Portugal) to MM and by Howard Hughes Medical Institute funding to MJC. RA is supported by a PhD fellowship (SFRH/BD/51876/2012, from FCT, Portugal) and FL is supported by a HHMI international graduate student research fellowship.

Chapter IV: General Discussion

Mycroft Holmes: "Oh, Sherlock, what do we say about coincidence?"

Sherlock Holmes: "The Universe is rarely so lazy."

- In *Sherlock (2010): The sign of Three.*

Axial extension is a core developmental process conserved in all vertebrates. It relies on a specific population of cells - the axial progenitors - located in the posterior-most region of the embryo and able to proliferate during the entire process of axial elongation, generating almost all the tissues composing the vertebrate body. The specific control of axial extension processes not only determines the final body length of a particular species, but also defines the post-cranial body allocation into its main regions - neck, trunk, and tail -, as well as the relative proportions among them. As such, axial elongation is one of the key developmental processes responsible for the wide variety in body shapes and sizes observed within the vertebrate clade, determining each species' particular body plan. An appropriate control of the axial progenitor cell population is, therefore, essential for correct axis specification and proper axial termination, having important functional, as well as evolutionary, consequences.

IV.1 *Gdf11* controls the axial progenitor pool size during the trunk to tail transition

In Chapter II, we showed that absence of Gdf11 signalling results in an expansion of the axial progenitor population after the trunk to tail transition. These progenitors in excess are found in three main regions within *Gdf11*$^{-/-}$ tails instead of residing within a single intact CNH. The miss-regulation in axial progenitor numbers is very likely the cause of most of the severe tail abnormalities found in *Gdf11* mutant embryos from mid-gestation stages, which include tail truncations, enlarged neural tubes and frequent tail bifurcations.

Conversion from trunk- to tail-forming mechanisms requires both changes in progenitor type and localization within the embryo, as well as an extensive cellular re-organization of the NMP pool, particularly regarding its size. Retrospective clonal analyses have shown that the axial progenitor number varies throughout development, specifically during the trunk to tail transition period

when a significant increase in clonal complexity was observed at axial levels corresponding to the finishing of the trunk and start of the tail (Tzouanacou et al., 2009). Interestingly, the anterior expression limit of the Gdf11 receptor *Alk5*, also coincides with this axial level (Andersson et al., 2006). This observation, together with the results presented in this thesis, indicates that Gdf11 signalling might have an important role in regulating the extent of NMP pool expansion during the trunk to tail transition, keeping the progenitor population at appropriate numbers. Indeed, a number of studies have shown that Gdf11 and its closely related Gdf family member, Gdf8/Myostatin, have important roles in tissue homeostasis in several biological contexts, acting as negative growth regulators or chalones (reviewed in Walker et al., 2016). *Gdf11* in particular has been demonstrated to be essential for the control of neural progenitor populations in the mouse developing spinal cord and olfactory epithelium (OE)(Shi and Liu, 2011; Wu et al., 2003). Loss of Gdf11 signalling prevents neuronal precursors from exiting the cell cycle, keeping them in a proliferative state. This leads to an increase in both neural progenitor numbers and differentiated neurons in the OE, and to a slower neurogenesis rate in the spinal cord (Shi and Liu, 2011; Wu et al., 2003). These observations are consistent with the excess of NMPs found in *Gdf11*$^{-/-}$ embryos and also help to explain the presence of enlarged neural tubes and truncations in mutant tails. An expanded population of axial progenitors would indeed likely generate more neurons, thus greatly contributing to the wide neural tubes observed in *Gdf11*$^{-/-}$ tails. However, a slower neurogenesis rate would also result in a slower axial tail growth and, consequently, in the maintenance of a large part of that progenitor pool in an undifferentiated state. As our tail explants culture experiments demonstrated, the proliferative abilities of *Gdf11*$^{-/-}$ progenitors, though high, they are still limited. This means that increased numbers of axial progenitors in *Gdf11*$^{-/-}$ tails can overall produce more tissue but, as tail growth could be impaired by slow differentiation rates, they would not be able to

contribute adequately to tail extension and, thus, tail truncations would ultimately result from the exhaustion of the progenitor pool through their miss-differentiation into other tissues. The dissociation of the CNH into several NMP-containing regions would also lead to several dispersed and disarticulated growth foci in *Gdf11* mutant tails. A significant portion of these surplus progenitors seems to be epiblast-associated precursors that are contained in the ectopic ventral pocket; both in its epithelial and mesenchymal components (see next section). Out of their natural context and without proper differentiation cues, these cells most probably continue to proliferate, perturbing tail tissue integrity and imposing significant mechanical constraints to axial tail growth until the end of embryonic development. This cell population probably contributes to the thickening of *Gdf11* mutant mid-gestation tails and will likely become the ectopic neural tissue structure found next to the cloaca of *Gdf11*$^{-/-}$ fetuses (Szumska et al., 2008).

Overall, the results shown in Chapter II indicate that Gdf11 signalling is an important pathway in the control of trunk to tail transition and axial extension cessation, essential for a correct epiblast resolution and NMP pool size regulation during this process. This way, Gdf11 ensures a proper, gradual, embryonic axis termination. A complete, in-depth *in vivo* and *in vitro* characterization of these *Gdf11* mutant progenitors would be interesting, given their proliferative abilities and their propensity towards differentiating into a neural fate. We are currently addressing this issue, with a particular focus on possible biomedical applications in spinal cord regeneration. It is also interesting to speculate whether the recent role in tissue rejuvenation that has been suggested for Gdf11 and Gdf11 signalling is related to this ability to negatively modulate progenitor populations (Loffredo et al., 2013; Sinha et al., 2014; reviewed in Walker et al., 2016). Systemic presence of Gdf11 could maybe stimulate the differentiation of latent progenitor populations throughout the body, promoting adult tissue renewal. Alternatively,

Gdf11 signalling could have had acquired a new function as a positive regulator in adult tissue precursor pools, in contrast its effect during embryonic development. Nevertheless, the role of Gdf11 in adult tissues is still highly debated (Egerman et al., 2016; Loffredo et al., 2013; Sinha et al., 2014; reviewed in Walker et al., 2016). Further work will be needed to clarify these issues and settle these apparently conflicting results.

IV.2 Gdf11 signalling *vs* Oct4 activity: a molecular tug-of-war in the control of vertebrate trunk length

One intriguing finding during the characterization of *Gdf11*$^{-/-}$ tail defects was the presence of ectopic *Oct4* expression in a subset of cells, which indicated that Gdf11 signalling could act as a negative modulator of *Oct4* expression. Previous studies had already suggested a possible functional connection between these two factors. In particular, conditional inactivation experiments showed that *Oct4* is required during trunk formation but not at tail bud stages, as inactivation of this gene at late primitive streak stages generated embryos with dramatic trunk shortenings, yet perfectly specified tails (DeVeale et al., 2013). This phenotype not only was very similar to the one observed in about half of the transgenic embryos obtained when prematurely activating Gdf11/Alk5 signalling in axial progenitors at the beginning of the trunk-forming stage (Jurberg et al., 2013), but also was somewhat complementary to the extended trunks observed in *Gdf11* mutant embryos (Jurberg et al., 2013; McPherron et al., 1999). Moreover, SB431542, an inhibitor of the Alk4/Alk5/Alk7 receptors, can under certain conditions substitute for Oct4 during iPSC reprogramming, mainly by upregulating the endogenous *Oct4* gene (Tan et al., 2015), thus connecting inhibition of TGF-β signalling with *Oct4* activation. Despite this clear functional interaction between Gdf11 signalling and Oct4 *in vivo* and *in vitro*, it is still unclear at which level this interaction occurs. ChIP assays using an antibody for one of Gdf11 signalling effectors and genome

editing experiments showed that part of these interactions might be mediated by direct regulation of *Oct4* expression by the Gdf11 pathway. However, this might be only part of the regulatory interactions between these two factors. The fact that no major phenotypic effects were observed upon ectopic *Oct4* expression when it resulted from a deletion in *Oct4*'s proximal enhancer that potentially hindered Gdf11 activity indicates that a large part of the Gdf11/Oct4 interactions might occur at the level of many different target loci throughout the genome. We are currently trying to address this idea experimentally.

Detection of ectopic *Oct4* expression in the tail tissues of *Gdf11*$^{-/-}$ embryos raised interesting questions about its relative contribution to the different phenotypic traits observed in these embryos. The presence of *Oct4*-expressing cells both in the epithelial and mesenchymal components of the ventral ectopic cellular mass suggested that this factor could account for the formation of these abnormal structures and, hence, for the tail abnormalities found in *Gdf11* mutant embryos. Molecular analyses implied that this ectopic structure likely consisted of epiblast remnants still containing epiblast-associated axial progenitors. These observations, together with the genetic studies involving *Oct4* in the maintenance of the primitive streak (DeVeale et al., 2013) and the recovery of EpiSC characteristics of the tail bud upon *Oct4* over-expression (Economou et al., 2015), indicated that this epiblast-like structure could have indeed derived from incomplete *Oct4* silencing, which would have resulted in an incomplete epiblast extinction during the trunk to tail transition.

Our transgenic experiments further supported the premise that *Oct4* miss-expression in axial progenitor-containing regions has a role in the development of *Gdf11*$^{-/-}$ axial phenotypes, since mildly affected *Cdx2-Oct4* transgenics recapitulated most of the morphological and molecular alterations found in the posterior axial areas of *Gdf11* mutant embryos. Overall, the results from these experiments were consistent with the hypothesis that a complete silencing of

Oct4 expression/activity by Gdf11 signalling is essential for a correct trunk to tail transition and for a proper specification of tail structures. Presence of ectopic Oct4 activity is then sufficient to delay this transition and is seemingly a key contributor to the observed expansion of the axial progenitor population size in *Gdf11*$^{-/-}$ embryos, which results in severe morphological abnormalities in embryonic tails.

Furthermore, our *Oct4* over-expressing experiments confirmed this factor's major role in the regulation of trunk length in mouse axial extension. These transgenic embryos displayed a remarkable increase in the number of somites between the fore and hindlimb buds at E10.5 and an abnormally large number of thoracic vertebrae observed at E18.5, both traits indicating an increase in the number of trunk segments. An interesting characteristic of these embryos was the apparent presence of an epiblast at stages of development when this structure should have already disappeared. This observation is quite relevant, as it links the presence of the epiblast and trunk extension with Oct4 activity. The association between this gene and epiblast-driven axial extension was already inferred from both the normal timing of *Oct4* expression during mouse development (Downs, 2008; Osorno et al., 2012) and the genetic studies indicating an *Oct4* requirement at stages when the epiblast is active (DeVeale et al., 2013). What the *Cdx2-Oct4* transgenics now demonstrate is that Oct4 seems to be itself sufficient to maintain an active epiblast. In fact, it has been suggested that one fundamental role of Oct4 is the preservation of the epithelial integrity as well as EMT inhibition, since most of its evolutionarily conserved targets are factors involved in cell-to-cell adhesion (Livigni et al., 2013; Redmer et al., 2011). This way, sustained Oct4 activity could be involved in maintaining the epiblast for longer developmental times, suppressing the EMT associated with the transition to a tail bud-dependent axial extension and thereby keeping axial progenitors in a trunk-forming configuration. Interestingly, pluripotency seems to be associated with the

epithelial state, as some of these Oct4 targets involved in cell-to-cell adhesion can actually rescue pluripotency to ES cells lacking *Oct4* (Livigni et al., 2013; Redmer et al., 2011). Formation of the trunk region requires a relatively high level of differentiation potency of its progenitors, as their descendants are involved in the formation of a wide variety of structures. Thus, the presence of an epiblast during trunk development might reflect the high potency state promoted by Oct4, which is necessary for axial extension through the embryonic trunk. Moreover, these findings also suggest that Oct4 activity is intrinsically associated to the trunk-elongating mechanisms and that extending that activity during longer developmental periods delays the trunk to tail transition. As such, we would expect an extended *Oct4* expression, with concomitant longer epiblast maintenance, in *Gdf11* mutant embryos. This analysis is currently underway.

Expression of genes belonging to the 5'-most *Hox*PGs followed the re-distribution of the trunk and tail domains in the *Cdx2-Oct4* transgenics. Particularly, activation of 5' *Hox*PG genes, which typically correlates with post-thoracic vertebral identities and with the trunk to tail transition (Burke et al., 1995; Carapuço et al., 2005; Di-Poï et al., 2010; Gaunt, 2000; Jurberg et al., 2013), was posteriorized in *Cdx2-Oct4* transgenics, generally accompanying the hindlimb displacement. This posteriorization was more dramatic in the paraxial mesoderm, where 5' *Hox*PG expression was almost absent in the most affected specimens. This shift in posterior *Hox* gene activation is complementary to that observed upon premature activation of Gdf11/Alk5 signalling, further supporting the opposing functional connection between this signalling pathway and Oct4. Our results, together with all functional assays using *Hox* genes showing that their miss-expression solely affects segmental identities in the axial skeleton and never the axial position of the hindlimb (Carapuço et al., 2005; Jurberg et al., 2013; Vinagre et al., 2010; Wellik and Capecchi, 2003), favour the role of *Hox* genes as

downstream effectors of the Oct4/Gdf11-regulated program in the paraxial mesoderm.

It should be noted, however, that while all *Cdx2-Oct4* embryos exhibited seemingly normal anterior thoracic regions, all analysed strongly affected transgenic fetuses generated progressively more abnormal structures at their caudal-most levels. Yet, all of these embryos appeared to have undergone at least some aspects of the trunk to tail transition, albeit terminating with the development of large sacrococcygeal teratomas. These defects are likely to result from a still active Gdf11 signalling in these embryos. The generation of these tumours in other contexts associated with Oct4 ectopic activity (Economou et al., 2015) is, once again, consistent with the important role of Gdf11 signalling in the proper axial extension termination by negatively regulating *Oct4* expression/activity, while providing correct differentiation cues to axial progenitors and their derivatives.

Interestingly, double mutants for *Gdf11* and *Gdf8/Myostatin* originate embryos sharing many phenotypic traits with strongly affected *Cdx2-Oct4* transgenics, since these also have delayed trunk to tail transitions, with extended trunks containing up to 20 rib-bearing thoracic segments, and severely truncated tails (McPherron et al., 2009). The synergistic effect of the two mutations demonstrates the redundant roles both genes have in the regulation of the onset of the trunk to tail transition. This opens the possibility that other TGF-β pathways might be contributing to this process as well. Analysis of *Oct4* expression dynamics during the trunk to tail transition in these double mutant embryos would be interesting, as it might provide insights into additional mechanisms involved in Oct4 silencing (or lack of thereof). It would be also interesting to investigate the effect of extended Oct4 activity in the absence of *Gdf11*, *Gdf8*, or both. If Oct4 activity is indeed the main trunk-extending factor, we would expect embryos unable to undergo a trunk to tail transition.

In summary, the data presented in this thesis suggests that the molecular "tug-of-war" between Gdf11 (possibly together with Gdf8) and Oct4 activities is a determining factor regulating the regionalization of the vertebrate body into trunk and tail domains. This is consistent with the observation that experimental changes in Oct4 or Gdf11 activities produced complementary shifts in the trunk to tail transition.

IV.3 *Oct4*, *Gdf11* and the evolution of the snake body plan

Extended organ-filled trunks and delayed activation of posterior *Hox* genes are also distinctive features of the snake taxa, thus suggesting that Oct4 activity could be in the origin of the snake body plan. Indeed, *Oct4* expression in the corn snake seemed to have undergone a heterochronic shift relative to mouse, being present during longer developmental times in this species. Changes in the relative timing of developmental events, particularly shifts in the timeframe of activation or silencing of particular genes, are thought to be key events underlying major phenotypic changes during evolution. These types of heterochronic changes have been well documented in the evolution of structures such as the vertebrate limb and craniofacial development in marsupials (reviewed in Keyte and Smith, 2014; Smith, 2003). As such, heterochronic shifts maintaining a robust *Oct4* expression in axial progenitors, in combination with a putative delay on the onset of Gdf11 activity, would keep axial extension in a trunk-generating mode during longer developmental times, thereby generating the extended trunks typical of snake taxa. Whether this activity as a trunk-mode inducing signal is likewise associated with the maintenance of a longer lasting epiblast (as in *Cdx2-Oct4* transgenics) remains to be investigated, since collecting snake embryos in the adequate stage of development is quite challenging. However, as the reprogramming ability of the vertebrate Oct4 protein appears to be an evolutionarily ancient trait, snake Oct4

could likely function in a similar way to its mouse, platypus and axolotl counterparts (Hammachi et al., 2012; Niwa et al., 2008; Tapia et al., 2012).

Our analyses showed that the global gene structure around the *Oct4* locus is quite different in squamates and mammals, particularly regarding syntenic associations, which was consistent with a previous report by Frankenberg and Renfree (2013). However, the origin of these differential rearrangements is not clear. Frankenberg and Renfree suggested that these differences emerged from a genomic duplication of a region containing the ancestral versions of *Npdc1*, *Oct4* and *Fut7* genes in a sarcopteryngian ancestor that was soon followed by gene divergence of one homologue of *Fut7* into *Tcf19* in one of the duplicated sequences. In eutherians, the ancestral version of *Oct4* and one of the *Npdc1* homologues were lost, whereas the entire ancestral duplicated region was deleted in squamates shortly after their lineage divergence. This process ultimately resulted in the loss of synteny between *Npdc1* and *Oct4* observed in mammals and in the flanking of *Oct4* by *Npdc1* and *Tcf19* in all squamates (Frankenberg and Renfree, 2013; reviewed in Frankenberg et al., 2014). However, the lack of similarity between *Tcf19* and *Fut7* together with the observation that the synteny 3' from Oct4 extends further than the Tcf19 gene casts some doubts on this hypothesis.

Regardless of the mechanism, these differential recombination events in mammals and squamates seemed to have dramatically affected the Oct4 regulatory landscape. While Oct4 regulation in mammals is relatively well studied (Liao et al., 2013; Yeom et al., 1996), regulatory processes in snakes are only starting to be understood. The extensive sequence conservation in both coding and non-coding regions around the *Oct4* locus in snakes, together with the absence of a similar conservation in non-coding regions in lizards indicates the existence of regulatory elements within this area. Current technological possibilities hinder a direct experimental evaluation of these elements in snake

embryos. However, the 250bp highly conserved fragment among squamates might be a central element in the regulatory landscape. Indeed, this fragment on its own showed very strong enhancer activity in a transgenic reporter assay in mouse embryos, with a pattern compatible with its expression in axial progenitors. However, when tested embedded within a larger snake genomic context, the activity of this enhancer was reduced, suggesting that the 250bp region's potential is indeed regulated by adjacent sequences conserved exclusively in snake taxa. Interestingly, the 250bp fragment activates reporter expression mainly in Sox2-positive areas such as the neural tube. This suggests that squamates might share some *Oct4* activating mechanisms with mammals, possibly through Sox-responsive elements present in these squamate *Oct4* enhancers. Sequence analyses focusing on the identification of these motifs, followed by multi-species sequence comparisons to assess the functional conservation of these Sox-responsive sites, could shed light on this matter. Nevertheless, the *Oct4* regulatory mechanisms in snakes cannot be properly controlled by the molecular milieu present in the mouse embryo, as a BAC containing the snake *Oct4* genomic region was unable to activate snake *Oct4* expression in transgenic embryos.

The presence of the 250bp fragment also in lizard genomes is quite interesting. It would be important to evaluate if this sequence together with lizard-specific *Oct4* flanking regions can function as tissue regeneration enhancer elements (TREEs, Kang et al., 2016), since regenerative processes generally involve progenitor reactivation and/or cellular reprogramming events (Poss, 2010). Specifically, there seems to be a peak of sequence conservation in all analysed lizard species approximately 10kb upstream from the 250bp *Oct4* regulatory element that is not conserved in snakes. A study of these sequences' activity in a tail injury context could provide insights into the striking divergence and sparse

conservation observed in the squamate *Oct4* 5' regions between regenerating lizards and non-regenerating snakes.

Finally, the extensive sequence conservation in snake *Oct4*-adjacent regions demonstrates that these rearrangements most likely occurred shortly after the divergence from the lizard branch and were actively maintained during the evolution of the snake taxa. Indeed, a recent report describing a new fossil snake species occupying a basal position in the snake phylogeny revealed that the elongated snake body plan, particularly the presence of over 150 pre-sacral vertebrae typical of this group, evolved very early after the lizard to snake transition (Martill et al., 2015). As such, this substantial sequence conservation might have resulted from the action of a strong selective pressure over *Oct4* expression during the adaptation to a fossorial lifestyle, which favoured long trunks for burrowing and prey constriction.

IV.5 Final considerations and future directions

In this work, we analysed the roles of Gdf11 signalling and Oct4 activity in the embryonic axial extension process and in the establishment of the vertebrate body plan. We found that the interaction between these two factors could have contributed to the generation of evolutionary novelty that, ultimately, might have resulted in the wide diversity of body plans observed among vertebrates.

However, important questions regarding the generation of elongated bodies still remain, particularly about the factors involved in the control of axial progenitor potency and proliferation. In fact, ectopic Oct4 expression, whilst increasing absolute trunk length, did not increase the total somite numbers in *Cdx2-Oct4* transgenic embryos relative to equivalently staged mouse embryos. This likely stems from the existence of a functional Gdf11 signalling acting upon this *Oct4*-expressing, highly proliferative, pluripotent axial progenitor population. In the wrong environment and subjected to progressively conflicting signals, these

progenitors would gradually stop giving rise to trunk structures and start producing large teratomas at posterior axial levels where Gdf11 signalling is strongest. On the other hand, although Gdf11 signalling was shown to have an important role in proper axis termination, its absence was not enough to keep axial progenitors in a proliferative state indefinitely possibly due to the transient nature of *Oct4* ectopic expression in *Gdf11* mutant tails. However, over-expression of *Oct4* in a Gdf11-null context has not been tested. Whether these conditions are able to generate embryos with even longer trunks and no trunk to tail transition remains to be determined.

Nevertheless, these results imply that snakes and other long-bodied vertebrates have particularly adapted axial extension processes, containing axial progenitor populations with higher or extended proliferative abilities during longer periods of development. A full characterization of axial progenitor populations in mouse, snake and lizards, coupled with extensive multi-species sequence comparisons, would provide both developmental and evolutionary insights into the evolution of the process of axial extension in each taxa. We are currently addressing this issue by using the lineage tracing approach described in Chapter II for axial progenitor isolation and transcriptomic analyses in mouse embryos. However, the limitations of snakes and lizards as models for developmental biology and experimental evolution studies are yet to be overcome, particularly regarding genome manipulation. Nonetheless, studies in these and other groups belonging to different phylogenetic branches are crucial to fully understand the mechanisms and evolutionary origins of core developmental processes, thus providing insights into new ways of harnessing their full potential and of applying their underlying principles in new ways.

Chapter V:
References

"*Every man has but one destiny."*

- In *The Godfather,* by Mario Puzo

Abu-abed, S., Dolle, P., Metzger, D., Beckett, B., Chambon, P. and Petkovich, M. (2001). The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. *Genes Dev.* **15**, 226–240.

Acloque, H., Adams, M. S., Fishwick, K., Bronner-fraser, M. and Nieto, M. A. (2009). Epithelial-mesenchymal transitions : the importance of changing cell state in development and disease. *J. Clin. Invest.* **119**, 1438–1449.

Andersson, O., Reissmann, E. and Ibáñez, C. F. (2006). Growth differentiation factor 11 signals through the transforming growth factor-beta receptor ALK5 to regionalize the anterior-posterior axis. *EMBO Rep.* **7**, 831–7.

Aoyama, H. and Asamoto, K. (2000). The developmental fate of the rostral/caudal half of a somite for vertebra and rib formation: Experimental confirmation of the resegmentation theory using chick-quail chimeras. *Mech. Dev.* **99**, 71–82.

Arnold, S. J. and Robertson, E. J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* **10**, 91–103.

Arnold, S. J., Hofmann, U. K., Bikoff, E. K. and Robertson, E. J. (2008). Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development* **135**, 501–511.

Aulehla, A., Wehrle, C., Brand-Saberi, B., Kemler, R., Gossler, A., Kanzler, B. and Herrmann, B. G. (2003). Wnt3a plays a major role in the segmentation clock controlling somitogenesis. *Dev. Cell* **4**, 395–406.

Aulehla, A., Wiegraebe, W., Baubet, V., Wahl, M. B., Deng, C., Taketo, M., Lewandoski, M. and Pourquié, O. (2008). A beta-catenin gradient links the clock and wavefront systems in mouse embryo segmentation. *Nat. Cell Biol.* **10**, 186–193.

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N. and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126–40.

Balmer, S., Nowotschin, S., Hadjantonakis, A.-K., Piccirillo, R., Demontis, F., Perrimon, N. and Goldberg, A. (2016). Notochord morphogenesis in mice: Current understanding & open questions. *Dev. Dyn.* 1–47.

Barrow, J. R., Howell, W. D., Rule, M., Hayashi, S., Thomas, K. R., Capecchi, M. R. and McMahon, A. P. (2007). Wnt3 signaling in the epiblast is required for proper orientation of the anteroposterior axis. *Dev. Biol.* **312**, 312–320.

Batlle, E., Sancho, E., Francí, C., Domínguez, D., Monfar, M., Baulida, J. and García De Herreros, A. (2000). The transcription factor Snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat. Cell Biol.* **2**, 84–89.

Beck, C. W. (2015). Development of the vertebrate tailbud. *Wiley Interdiscip. Rev. Dev. Biol.* **4**, 33–44.

Beck, F., Erler, T., Russell, a and James, R. (1995). Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. *Dev. Dyn.* **204**, 219–227.

Beck, S., Le Good, J. A., Guzman, M., Ben Haim, N., Roy, K., Beermann, F. and Constam, D. B. (2002). Extraembryonic proteases regulate Nodal signalling during gastrulation. *Nat. Cell Biol.* **4**, 981–985.

Bedzhov, I. and Zernicka-Goetz, M. (2014). Self-organizing properties of mouse pluripotent cells initiate morphogenesis upon implantation. *Cell* **156**, 1032–1044.

Bedzhov, I., Graham, S. J. L., Leung, C. Y., Zernicka-goetz, M. and Zernicka-goetz, M. (2014). Developmental plasticity , cell fate specification and morphogenesis in the early mouse embryo.

Benahmed, F., Gross, I., Gaunt, S. J., Beck, F., Jehan, F., Domon–Dell, C., Martin, E., Kedinger, M., Freund, J. and Duluc, I. (2008). Multiple Regulatory Regions Control the Complex Expression Pattern of the Mouse Cdx2 Homeobox Gene. *Gastroenterology* **135**, 1238–1247.e3.

Ben-Haim, N., Lu, C., Guzman-Ayala, M., Pescatore, L., Mesnard, D., Bischofberger, M., Naef, F., Robertson, E. J. and Constam, D. B. (2006). The Nodal Precursor Acting via Activin Receptors Induces Mesoderm by Maintaining a Source of Its Convertases and BMP4. *Dev. Cell* **11**, 313–323.

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

Bessho, Y., Hirata, H., Masamizu, Y. and Kageyama, R. (2003). Periodic repression by the bHLH factor Hes7 is an essential mechanism for the somite segmentation clock. *Genes Dev.* **17**, 1451–1456.

Borycki, A. G., Brunk, B., Tajbakhsh, S., Buckingham, M., Chiang, C. and Emerson,

C. P. (1999). Sonic hedgehog controls epaxial muscle determination through Myf5 activation. *Development* **126**, 4053–63.

Brennan, J., Lu, C. C., Norris, D. P., Rodriguez, T. A., Beddington, R. S. P. and Robertson, E. J. (2001). Nodal signalling in the epiblast patterns the early mouse embryo. *Nature* **411**, 965–969.

Brent, A. E., Schweitzer, R. and Tabin, C. J. (2003). A somitic compartment of tendon progenitors. *Cell* **113**, 235–248.

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

Burdsal, C. A., Damsky, C. H. and Pedersen, R. A. (1993). The role of E-cadherin and integrins in mesoderm differentiation and migration at the mammalian primitive streak. *Development* **118**, 829–844.

Burke, A. C., Nelson, C. E., Morgan, B. A. and Tabin, C. J. (1995). Hox genes and the evolution of vertebrate axial morphology. *Development* **121**, 333–46.

Bussen, M., Petry, M., Schuster-Gossler, K., Leitges, M., Gossler, A. and Kispert, A. (2004). The T-box transcription factor Tbx18 maintains the separation of anterior and posterior somite compartments. *Genes Dev.* **18**, 1209–1221.

Cambray, N. and Wilson, V. (2002). Axial progenitors with extensive potency are localised to the mouse chordoneural hinge. *Development* **129**, 4855–4866.

Cambray, N. and Wilson, V. (2007). Two distinct sources for a population of maturing axial progenitors. *Development* **134**, 2829–2840.

Cano, A., Pérez-Moreno, M. A., Rodrigo, I., Locascio, A., Blanco, M. J., del Barrio, M. G., Portillo, F. and Nieto, M. A. (2000). The transcription factor Snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat. Cell Biol.* **2**, 76–83.

Carapuço, M., Nóvoa, A., Bobola, N. and Mallo, M. (2005). Hox genes specify vertebral types in the presomitic mesoderm. *Genes Dev.* **19**, 2116–21.

Carver, E. A., Jiang, R., Lan, Y., Oram, K. F. and Gridley, T. (2001). The mouse Snail gene encodes a key regulator of the epithelial-mesenchymal transition. *Mol Cell Biol* **21**, 8184–8188.

Castoe, T. A., de Koning, A. P. J., Hall, K. T., Card, D. C., Schield, D. R., Fujita, M. K., Ruggiero, R. P., Degner, J. F., Daza, J. M., Gu, W., et al. (2013). The Burmese python genome reveals the molecular basis for extreme adaptation

in snakes. *Proc. Natl. Acad. Sci.* **110**, 20645–20650.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655.

Chapman, D. L. and Papaioannou, V. E. (1998). Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. *Nature* **391**, 695–697.

Chawengsaksophak, K., James, R., Hammond, V. E., Köntgen, F. and Beck, F. (1997). Homeosis and intestinal tumours in Cdx2 mutant mice. *Nature* **386**, 84–87.

Chawengsaksophak, K., de Graaff, W., Rossant, J., Deschamps, J. and Beck, F. (2004). Cdx2 is essential for axial elongation in mouse development. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7641–7645.

Chazaud, C., Yamanaka, Y., Pawson, T. and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev. Cell* **10**, 615–24.

Chenoweth, J. G., McKay, R. D. G. and Tesar, P. J. (2010). Epiblast stem cells contribute new insight into pluripotency and gastrulation. *Dev. Growth Differ.* **52**, 293–301.

Chew, J., Loh, Y., Zhang, W., Chen, X., Tam, W., Yeap, L., Li, P., Ang, Y., Robson, P., Ng, H., et al. (2006). Reciprocal Transcriptional Regulation of Complex in Embryonic Stem Cells Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4 / Sox2 Complex in Embryonic Stem Cells. *Mol. Cell. Biol.* **25**, 6031–6046.

Christ, B. and Wilting, J. (1992). From somites to vertebral column. *Ann. Anat. - Anat. Anzeiger* **174**, 23–32.

Christ, B., Huang, R. and Scaal, M. (2004). Formation and differentiation of the avian sclerotome. *Anat. Embryol. (Berl).* **208**, 333–350.

Christ, B., Huang, R. and Scaal, M. (2007). Amniote somite derivatives. *Dev. Dyn.* **236**, 2382–2396.

Chu, J., Ding, J., Jeays-Ward, K., Price, S. M., Placzek, M. and Shen, M. M. (2005). Non-cell-autonomous role for Cripto in axial midline formation during vertebrate embryogenesis. *Development* **132**, 5539–51.

Ciruna, B. and Rossant, J. (2001). FGF Signaling Regulates Mesoderm Cell Fate

Specification and Morphogenetic Movement at the Primitive Streak. *Dev. Cell* **1**, 37–49.

Ciruna, B. G., Schwartz, L., Harpal, K., Yamaguchi, T. P. and Rossant, J. (1997). Chimeric analysis of fibroblast growth factor receptor-1 (Fgfr1) function: a role for FGFR1 in morphogenetic movement through the primitive streak. *Development* **124**, 2829–2841.

Cohn, M. J. and Tickle, C. (1999). Developmental basis of limblessness and axial patterning in snakes. *Nature* **399**, 474–479.

Conlon, F. L., Lyons, K. M., Takaesu, N., Barth, K. S., Kispert, A., Herrmann, B. and Robertson, E. J. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* **120**, 1919–1928.

Cooke, J. and Zeeman, E. C. (1976). A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *J. Theor. Biol.* **58**, 455–476.

Dale, J. K., Malapert, P., Chal, J., Vilhais-Neto, G., Maroto, M., Johnson, T., Jayasinghe, S., Trainor, P., Herrmann, B. and Pourquié, O. (2006). Oscillations of the snail genes in the presomitic mesoderm coordinate segmental patterning and morphogenesis in vertebrate somitogenesis. *Dev. Cell* **10**, 355–366.

Darbellay, F. and Duboule, D. (2016). Chapter Sixteen - Topological Domains, Metagenes, and the Emergence of Pleiotropic Regulations at Hox Loci. In *Essays on Developmental Biology, Part A* (ed. Biology, P. M. W. B. T.-C. T. in D.), pp. 299–314. Academic Press.

Dequéant, M.-L. and Pourquié, O. (2008). Segmental patterning of the vertebrate embryonic axis. *Nat. Rev. Genet.* **9**, 370–382.

Dequéant, M.-L., Glynn, E., Gaudenz, K., Wahl, M., Chen, J., Mushegian, A. and Pourquié, O. (2006). A complex oscillating network of signaling genes underlies the mouse segmentation clock. *Science* **314**, 1595–1598.

Deschamps, J. and van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development* **132**, 2931–2942.

DeVeale, B., Brokhman, I., Mohseni, P., Babak, T., Yoon, C., Lin, A., Onishi, K., Tomilin, A., Pevny, L., Zandstra, P. W., et al. (2013). Oct4 Is Required ~E7.5 for Proliferation in the Primitive Streak. *PLoS Genet.* **9**, e1003957.

Dietrich, J.-E. and Hiiragi, T. (2007). Stochastic patterning in the mouse pre-implantation embryo. *Development* **134**, 4219–4231.

Diez del Corral, R., Olivera-Martinez, I., Goriely, A., Gale, E., Maden, M. and Storey, K. (2003). Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron* **40**, 65–79.

Ding, J., Yang, L., Yan, Y.-T., Chen, A., Desai, N., Wynshaw-Boris, A. and Shen, M. M. (1998). Cripto is required for correct orientation of the anterior-posterior axis in the mouse embryo. *Nature* **395**, 702–707.

Di-Poï, N., Montoya-Burgos, J. I., Miller, H., Pourquié, O., Milinkovitch, M. C. and Duboule, D. (2010). Changes in Hox genes' structure and function during the evolution of the squamate body plan. *Nature* **464**, 99–103.

Downs, K. M. (2008). Systematic localization of Oct-3/4 to the gastrulating mouse conceptus suggests manifold roles in mammalian development. *Dev. Dyn.* **237**, 464–475.

Duband, J. L., Dufour, S., Hatta, K., Takeichi, M., Edelman, G. M. and Thiery, J. P. (1987). Adhesion molecules during somitogenesis in the avian embryo. *J. Cell Biol.* **104**, 1361–1374.

Duboule, D. (2007). The rise and fall of Hox gene clusters. *Development* **134**, 2549–2560.

Duboule, D. and Dollé, P. (1989). The structural and functional organization of the murine Hox gene family resembles that of Drosophila homeotic genes. *Embo J.* **8**, 1497–1505.

Dubrulle, J. and Pourquié, O. (2004). Fgf8 mRNA decay establishes a gradient that couples axial elongation to patterning in the vertebrate embryo. *Nature* **427**, 419–422.

Dubrulle, J., McGrew, M. J. and Pourquié, O. (2001). FGF signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal Hox gene activation. *Cell* **106**, 219–232.

Ducibella, T. and Anderson, E. (1975). Cell shape and membrane changes in the eight-cell mouse embryo: prerequisites for morphogenesis of the blastocyst. *Dev. Biol.* **47**, 45–58.

Dunn, N. R., Vincent, S. D., Oxburgh, L., Robertson, E. J. and Bikoff, E. K. (2004). Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation

and patterning in the mouse embryo. *Development* **131**, 1717–1728.

Economou, C., Tsakiridis, A., Wymeersch, F. J., Gordon-Keylock, S., Dewhurst, R. E., Fisher, D., Medvinsky, A., Smith, A. J. and Wilson, V. (2015). Intrinsic factors and the embryonic environment influence the formation of extragonadal teratomas during gestation. *BMC Dev. Biol.* **15**, 35.

Egerman, M. A., Cadena, S. M., Gilbert, J. A., Meyer, A., Nelson, H. N., Swalley, S. E., Mallozzi, C., Jacobi, C., Jennings, L. L., Clay, I., et al. (2016). GDF11 Increases with Age and Inhibits Skeletal Muscle Regeneration. *Cell Metab.* **22**, 164–174.

Fan, C. M. and Tessier-Lavigne, M. (1994). Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. *Cell* **79**, 1175–1186.

Fleming, A., Kishida, M. G., Kimmel, C. B. and Keynes, R. J. (2015). Building the backbone: the development and evolution of vertebral patterning. *Development* **142**, 1733–44.

Foglia, M. J. and Poss, K. D. (2016). Building and re-building the heart by cardiomyocyte proliferation. *Development* **143**, 729–740.

Forlani, S., Lawson, K. A. and Deschamps, J. (2003). Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. *Development* **130**, 3807–3819.

Frankenberg, S. and Renfree, M. B. (2013). On the origin of POU5F1. *BMC Biol.* **11**, 56.

Frankenberg, S., Gerbe, F., Bessonnard, S., Belville, C., Pouchin, P., Bardot, O. and Chazaud, C. (2011). Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling. *Dev. Cell* **21**, 1005–1013.

Frankenberg, S. R., Frank, D., Harland, R., Johnson, A. D., Nichols, J., Niwa, H., Schöler, H. R., Tanaka, E., Wylie, C. and Brickman, J. M. (2014). The POU-er of gene nomenclature. *Development* **141**, 2921–3.

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. and Dubchak, I. (2004). VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **32**, 273–279.

Frederick, J. P., Liberati, N. T., Waddell, D. S., Shi, Y. and Wang, X. (2004). Transforming growth factor beta-mediated transcriptional repression of c-myc is dependent on direct binding of Smad3 to a novel repressive Smad

binding element. *Mol. Cell. Biol.* **24**, 2546–59.

Galceran, J., Hsu, S. C. and Grosschedl, R. (2001). Rescue of a Wnt mutation by an activated form of LEF-1: regulation of maintenance but not initiation of Brachyury expression. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8668–73.

Gamer, L. W., Nove, J. and Rosen, V. (2003). Return of the Chalones Members of the TGF NLsuperfamily play many roles in IRE1 : A Role in UPREgulation. **4**, 143–144.

Garcia-Fernàndez, J. (2005). Hox, ParaHox, ProtoHox: facts and guesses. *Heredity (Edinb).* **94**, 145–152.

Garriock, R. J., Chalamalasetty, R. B., Kennedy, M. W., Canizales, L. C., Lewandoski, M. and Yamaguchi, T. P. (2015). Lineage tracing of neuromesodermal progenitors reveals novel Wnt-dependent roles in trunk progenitor cell maintenance and differentiation. *Development* **142**, 1628–1638.

Gaunt, S. J. (2000). Evolutionary shifts of vertebrate structures and Hox expression up and down the axial series of segments: A consideration of possible mechanisms. *Int. J. Dev. Biol.* **44**, 109–117.

Gaunt, S. J., George, M. and Paul, Y. L. (2013). Direct activation of a mouse Hoxd11 axial expression enhancer by Gdf11/Smad signalling. *Dev. Biol.* **383**, 52–60.

Gilbert, S. F. (1997). *Developmental Biology*. 5th ed. Sunderland, Massachusetts: Sinauer Associates.

Goldman, D. C., Martin, G. R. and Tam, P. P. L. (2000). Fate and function of the ventral ectodermal ridge during mouse tail development. *Development* **127**, 2113–2123.

Gomez, C., Ozbudak, E. M., Wunderlich, J., Baumann, D., Lewis, J. and Pourquie, O. (2008). Control of segment number in vertebrate embryos. *Nature* **454**, 335–339.

Goulding, M. D., Chalepakis, G., Deutsch, U., Erselius, J. R. and Gruss, P. (1991). Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *EMBO J.* **10**, 1135–1147.

Guo, Q. and Li, J. Y. H. (2007). Distinct functions of the major Fgf8 spliceform, Fgf8b, before and during mouse gastrulation. *Development* **134**, 2251–2260.

Guo, G., Huss, M., Tong, G. Q., Wang, C., Li Sun, L., Clarke, N. D. and Robson, P. (2010). Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst. *Dev. Cell* **18**, 675–685.

Halacheva, V., Fuchs, M., Dönitz, J., Reupke, T., Püschel, B. and Viebahn, C. (2011). Planar cell movements and oriented cell division during early primitive streak formation in the mammalian embryo. *Dev. Dyn.* **240**, 1905–1916.

Hammachi, F., Morrison, G. M., Sharov, A. A., Livigni, A., Narayan, S., Papapetrou, E. P., O'Malley, J., Kaji, K., Ko, M. S. H., Ptashne, M., et al. (2012). Transcriptional Activation by Oct4 Is Sufficient for the Maintenance and Induction of Pluripotency. *Cell Rep.* **1**, 99–109.

Head, J. J. and Polly, P. D. (2015). Evolution of the snake body form reveals homoplasy in amniote Hox gene function. *Nature* **520**, 86–89.

Henrique, D., Abranches, E., Verrier, L. and Storey, K. G. (2015). Neuromesodermal progenitors and the making of the spinal cord. *Development* **142**, 2864–2875.

Herrmann, B. G. (1991). Expression pattern of the Brachyury gene in whole-mount TWis/TWis mutant embryos. *Development* **113**, 913–917.

Hinck, A. P. (2012). Structural studies of the TGF-betas and their receptors - Insights into evolution of the TGF-beta superfamily. *FEBS Lett.* **586**, 1860–1870.

Hogan, B., Beddington, R., Constantini, F. and Lacy, E. (1994). *Manipulating the mouse embryo: a laboratory manual*. 2nd ed. Cold Spring Harbour Laboratory Press.

Huang, R., Zhi, Q., Brand-Saberi, B. and Christ, B. (2000). New experimental evidence for somite resegmentation. *Anat. Embryol. (Berl).* **202**, 195–200.

Hubaud, A. and Pourquié, O. (2014). Signalling dynamics in vertebrate segmentation. *Nat. Rev. Mol. Cell Biol.* **15**, 709–721.

Huelsken, J., Vogel, R., Brinkmann, V., Erdmann, B., Birchmeier, C. and Birchmeier, W. (2000). Requirement for beta-catenin in anterior-posterior axis formation in mice. *J. Cell Biol.* **148**, 567–78.

Iimura, T. and Pourquié, O. (2006). Collinear activation of Hoxb genes during gastrulation is linked to mesoderm cell ingression. *Nature* **442**, 568–571.

Ikeya, M. and Takada, S. (1998). Wnt signaling from the dorsal neural tube is required for the formation of the medial dermomyotome. *Development* **125**, 4969–4976.

Irie, N. and Kuratani, S. (2014). The developmental hourglass model: a predictor of the basic body plan? *Development* **141**, 4649–55.

Johnson, M. H. and Ziomek, C. A. (1981a). Induction of polarity in mouse 8-cell blastomeres: Specificity, geometry, and stability. *J. Cell Biol.* **91**, 303–308.

Johnson, M. H. and Ziomek, C. A. (1981b). The foundation of two distinct cell lineages within the mouse morula. *Cell* **24**, 71–80.

Jostes, B., Walther, C. and Gruss, P. (1990). The murine paired box gene, Pax7, is expressed specifically during the development of the nervous and muscular system. *Mech. Dev.* **33**, 27–37.

Jurberg, A. D., Aires, R., Varela-Lasheras, I., Nóvoa, A. and Mallo, M. (2013). Switching Axial Progenitors from Producing Trunk to Tail Tissues in Vertebrate Embryos. *Dev. Cell* **25**, 451–462.

Jurberg, A. D., Aires, R., Nóvoa, A., Rowland, J. E. and Mallo, M. (2014). Compartment-dependent activities of Wnt3a/β-catenin signaling during vertebrate axial extension. *Dev. Biol.* **394**, 253–263.

Kang, J., Karra, R., Dickson, A. L., Nachtrab, G. and Goldman, J. A. (2016). Modulation of tissue repair by regeneration enhancer elements.

Kanzler, B., Kuschert, S. J., Liu, Y. H. and Mallo, M. (1998). Hoxa-2 restricts the chondrogenic domain and inhibits bone formation during development of the branchial area. *Development* **125**, 2587–2597.

Kehler, J., Tolkunova, E., Koschorz, B., Pesce, M., Gentile, L., Boiani, M., Lomelí, H., Nagy, A., McLaughlin, K. J., Schöler, H. R., et al. (2004). Oct4 is required for primordial germ cell survival. *EMBO Rep.* **5**, 1078–1083.

Kelly, O. G., Pinson, K. I. and Skarnes, W. C. (2004). The Wnt co-receptors Lrp5 and Lrp6 are essential for gastrulation in mice. *Development* **131**, 2803–2815.

Kessel, M. (1992). Respecification of vertebral identities by retinoic acid. *Development* **115**, 487–501.

Kessel, M. and Gruss, P. (1991). Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid. *Cell* **67**,

89–104.

Keyte, A. L. and Smith, K. K. (2014). Heterochrony and developmental timing mechanisms: Changing ontogenies in evolution. *Semin. Cell Dev. Biol.* **34**, 99–107.

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. and Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Dev. Dynam.* **203**, 253–310.

Kimura, C., Yoshinaga, K., Tian, E., Suzuki, M., Aizawa, S. and Matsuo, I. (2000). Visceral endoderm mediates forebrain development by suppressing posteriorizing signals. *Dev. Biol.* **225**, 304–21.

Kinder, S. J., Tsang, T. E., Quinlan, G. A., Hadjantonakis, A.-K., Nagy, A. and Tam, P. P. L. (1999). The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development* **126**, 4691–4701.

Kinder, S. J., Tsang, T. E., Wakamiya, M., Sasaki, H., Behringer, R. R., Nagy, a and Tam, P. P. (2001). The organizer of the mouse gastrula is composed of a dynamic population of progenitor cells for the axial mesoderm. *Development* **128**, 3623–3634.

Kitajima, S., Takagi, A., Inoue, T. and Saga, Y. (2000). MesP1 and MesP2 are essential for the development of cardiac mesoderm. *Development* **127**, 3215–3226.

Kmita, M. and Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. *Science (80-.).* **301**, 331–333.

Kraus, F., Haenig, B. and Kispert, A. (2001). Cloning and expression analysis of the mouse T-box gene Tbx18. *Mech. Dev.* **100**, 83–86.

Krupa, M., Mazur, E., Szczepańska, K., Filimonow, K., Maleszewski, M. and Suwińska, A. (2014). Allocation of inner cells to epiblast vs primitive endoderm in the mouse embryo is biased but not determined by the round of asymmetric divisions (8→16- and 16→32-cells). *Dev. Biol.* **385**, 136–148.

Kumar, A., Novoselov, V., Celeste, A. J., Wolfman, N. M., Ten Dijke, P. and Kuehn, M. R. (2001). Nodal signaling uses activin and transforming growth factor-beta receptor-regulated Smads. *J. Biol. Chem.* **276**, 656–661.

Kumar, A., Lualdi, M., Lyozin, G. T., Sharma, P., Loncarek, J., Fu, X.-Y. Y. and Kuehn, M. R. (2014). Nodal signaling from the visceral endoderm is required

to maintain Nodal gene expression in the epiblast and drive DVE/AVE migration. *Dev. Biol.* **400**, 1–9.

Kwon, G. S., Viotti, M. and Hadjantonakis, A.-K. (2008). The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. *Dev. Cell* **15**, 509–520.

Lawson, A. and Schoenwolf, G. C. (2001). Cell populations and morphogenetic movements underlying formation of the avian primitive streak and organizer. *Genesis* **29**, 188–195.

Lawson, K. A. and Wilson, V. (2016). A Revised Staging of Mouse Development Before Organogenesis. In *Kaufman's Atlas of Mouse Development Supplement* (ed. Baldock, R.), Bard, J.), Davidson, D. R.), and Morriss-Kay, G.), pp. 51–64. London: Academic Press.

Lawson, K. A., Meneses, J. J. and Pedersen, R. A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **911**, 891–911.

Lee, Y. J., McPherron, A., Choe, S., Sakai, Y., Chandraratna, R. a., Lee, S.-J. and Oh, S. P. (2010). Growth differentiation factor 11 signaling controls retinoic acid activity for axial vertebral development. *Dev. Biol.* **347**, 195–203.

Lewis, E. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **277**, 565–570.

Liao, J., He, Y. and Szabó, P. E. (2013). The Pou5f1 distal enhancer is sufficient to drive Pou5f1 promoter-EGFP expression in embryonic stem cells. *Int. J. Dev. Biol.* **57**, 725–9.

Liu, J.-P. (2006). The function of growth/differentiation factor 11 (Gdf11) in rostrocaudal patterning of the developing spinal cord. *Development* **133**, 2865–2874.

Liu, P., Wakamiya, M., Shea, M. J., Albrecht, U., Behringer, R. R. and Bradley, A. (1999). Requirement for Wnt3 in vertebrate axis formation. *Nat. Genet.* **22**, 361–365.

Livigni, A., Peradziryi, H., Sharov, A. A., Chia, G., Hammachi, F., Migueles, R. P., Sukparangsi, W., Pernagallo, S., Bradley, M., Nichols, J., et al. (2013). A Conserved Oct4/POUV-Dependent Network Links Adhesion and Migration to Progenitor Maintenance. *Curr. Biol.* **23**, 2233–2244.

Loffredo, F. S., Steinhauser, M. L., Jay, S. M., Gannon, J., Pancoast, J. R.,

Yalamanchi, P., Sinha, M., Dall'Osso, C., Khong, D., Shadrach, J. L., et al. (2013). Growth differentiation factor 11 is a circulating factor that reverses age-related cardiac hypertrophy. *Cell* **153**, 828–839.

Mallo, M. (2015). Revisiting the involvement of signaling gradients in somitogenesis. *FEBS J.* 1–8.

Mallo, M. (2016). The Axial Musculoskeletal System. In *Kaufman's Atlas of Mouse Development Supplement* (ed. Baldock, R.), Bard, J.), Davidson, D. R.), and Morriss-Kay, G.), pp. 165–175. London: Academic Press.

Mallo, M. and Brändlin, I. (1997). Segmental identity can change independently in the hindbrain and rhombencephalic neural crest. *Dev. Dyn.* **210**, 146–56.

Mallo, M., Vinagre, T. and Carapuço, M. (2009). The road to the vertebral formula. *Int. J. Dev. Biol.* **53**, 1469–81.

Mallo, M., Wellik, D. M. and Deschamps, J. (2010). Hox genes and regional patterning of the vertebrate body plan. *Dev. Biol.* **344**, 7–15.

Martill, D. M., Tischlinger, H. and Longrich, N. R. (2015). A four-legged snake from the early Cretaceous of Gondwana. *Science (80-.).* **349**, 416–419.

Martin, B. L. (2015). Factors that coordinate mesoderm specification from neuromesodermal progenitors with segmentation during vertebrate axial extension. *Semin. Cell Dev. Biol.* **49**, 59–67.

Martin, B. L. and Kimelman, D. (2012). Canonical Wnt signaling dynamically controls multiple stem cell fate decisions during vertebrate body formation. *Dev. Cell* **22**, 223–232.

Masamizu, Y., Ohtsuka, T., Takashima, Y., Nagahara, H., Takenaka, Y., Yoshikawa, K., Okamura, H. and Kageyama, R. (2006). Real-time imaging of the somite segmentation clock: revelation of unstable oscillators in the individual presomitic mesoderm cells. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 1313–8.

Massagué, J., Seoane, J. and Wotton, D. (2005). Smad transcription factors Smad transcription factors. 2783–2810.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* **9**, 625–635.

McPherron, A. C., Lawler, A. M. and Lee, S.-J. J. (1999). Regulation of anterior/posterior patterning of the axial skeleton by growth/differentiation factor 11. *Nat. Genet.* **22**, 260–264.

McPherron, A. C., Huynh, T. V and Lee, S.-J. (2009). Redundancy of myostatin and growth/differentiation factor 11 function. *BMC Dev. Biol.* **9**, 24.

Meilhac, S. M., Adams, R. J., Morris, S. A., Danckaert, A., Le Garrec, J. F. and Zernicka-Goetz, M. (2009). Active cell movements coupled to positional induction are involved in lineage segregation in the mouse blastocyst. *Dev. Biol.* **331**, 210–221.

Meno, C., Gritsman, K., Ohishi, S., Ohfuji, Y., Heckscher, E., Mochida, K., Shimono, A., Kondoh, H., Talbot, W. S., Robertson, E. J., et al. (1999). Mouse lefty2 and zebrafish antivin are feedback inhibitors of nodal signaling during vertebrate gastrulation. *Mol. Cell* **4**, 287–298.

Mesnard, D., Filipe, M., Belo, J. A. and Zernicka-Goetz, M. (2004). The anterior-posterior axis emerges respecting the morphology of the mouse embryo that changes and aligns with the uterus before gastrulation. *Curr. Biol.* **14**, 184–196.

Mesnard, D., Guzman-Ayala, M. and Constam, D. B. (2006). Nodal specifies embryonic visceral endoderm and sustains pluripotent cells in the epiblast before overt axial patterning. *Development* **133**, 2497–2505.

Mishina, Y., Suzuki, A., Ueno, N. and Behringer, R. R. (1995). Bmpr encodes a type I bone morphogenetic protein receptor that is essential for gastrulation during mouse embryogenesis. *Genes Dev.* **9**, 3027–3037.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642.

Miura, S. and Mishina, Y. (2007). The DVE changes distal epiblast fate from definitive endoderm to neurectoderm by antagonizing nodal signaling. *Dev. Dyn.* **236**, 1602–1610.

Mohamed, O. A., Clarke, H. J. and Dufort, D. (2004). Beta-Catenin Signaling Marks the Prospective Site of Primitive Streak Formation in the Mouse Embryo. *Dev. Dyn.* **231**, 416–424.

Monsoro-Burq, A. H. (2005). Sclerotome development and morphogenesis: when

experimental embryology meets genetics. *Int. J. Dev. Biol.* **49**, 301–308.

Morikawa, M., Koinuma, D., Miyazono, K. and Heldin, C.-H. (2013). Genome-wide mechanisms of Smad binding. *Oncogene* **32**, 1609–1615.

Morris, S. A., Graham, S. J. L., Jedrusik, A. and Zernicka-Goetz, M. (2013). The differential response to Fgf signalling in cells internalized at different times influences lineage segregation in preimplantation mouse embryos. *Open Biol.* **3**, 130104.

Mullen, A. C., Orlando, D. A., Newman, J. J., Lovén, J., Kumar, R. M., Bilodeau, S., Reddy, J., Guenther, M. G., DeKoter, R. P. and Young, R. A. (2011). Master Transcription Factors Determine Cell-Type-Specific Responses to TGF-β Signaling. *Cell* **147**, 565–576.

Najm, F. J., Chenoweth, J. G., Anderson, P. D., Nadeau, J. H., Redline, R. W., McKay, R. D. G. and Tesar, P. J. (2011). Isolation of Epiblast Stem Cells from Preimplantation Mouse Embryos. *Cell Stem Cell* **8**, 318–325.

Nakashima, M., Toyono, T., Akamine, A. and Joyner, A. (1999). Expression of growth/differentiation factor 11, a new member of the BMP/TGF-beta superfamily during mouse embryogenesis. *Mech. Dev.* **80**, 185–189.

Nei, M., Xu, P. and Glazko, G. (2001). Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 2497–2502.

Neidhardt, L. M., Kispert, A. and Herrmann, B. G. (1997). A mouse gene of the paired-related homeobox class expressed in the caudal somite compartment and in the developing vertebral column, kidney and nervous system. *Dev. Genes Evol.* **207**, 330–339.

Nguyen, D. and Xu, T. (2008). The expanding role of mouse genetics for understanding human biology and disease. *Dis. Model. Mech.* **1**, 56–66.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H. and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391.

Niwa, H., Miyazaki, J. and Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24**, 372–376.

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R. and

Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917–929.

Niwa, H., Sekita, Y., Tsend-Ayush, E. and Grützner, F. (2008). Platypus Pou5f1 reveals the first steps in the evolution of trophectoderm differentiation and pluripotency in mammals. *Evol. Dev.* **10**, 671–682.

Nordhoff, V., Hübner, K., Bauer, A., Orlova, I., Malapetsa, A. and Schöler, H. R. (2001). Comparative analysis of human, bovine, and murine Oct-4 upstream promoter sequences. *Mamm. Genome* **12**, 309–317.

Oh, S. P. and Li, E. (1997). The signaling pathway mediated by the type IIB activin receptor controls axial patterning and lateral asymmetry in the mouse. *Genes Dev.* **11**, 1812–1826.

Oh, S. P., Yeo, C. Y., Lee, Y., Schrewe, H., Whitman, M. and Li, E. (2002). Activin type IIA and IIB receptors mediate Gdf11 signaling in axial vertebral patterning. *Genes Dev.* **16**, 2749–2754.

Ohta, S., Suzuki, K., Tachibana, K., Tanaka, H. and Yamada, G. (2007). Cessation of gastrulation is mediated by suppression of epithelial-mesenchymal transition at the ventral ectodermal ridge. *Dev.* **134** , 4315–4324.

Olivera-Martinez, I. and Storey, K. G. (2007). Wnt signals provide a timing mechanism for the FGF-retinoid differentiation switch during vertebrate body axis extension. *Development* **134**, 2125–2135.

Olivera-Martinez, I., Harada, H., Halley, P. A. and Storey, K. G. (2012). Loss of FGF-Dependent Mesoderm Identity and Rise of Endogenous Retinoid Signalling Determine Cessation of Body Axis Elongation. *PLoS Biol.* **10**, e1001415.

Osorno, R., Tsakiridis, A., Wong, F., Cambray, N., Economou, C., Wilkie, R., Blin, G., Scotting, P. J., Chambers, I. and Wilson, V. (2012). The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. *Development* **139**, 2288–2298.

Palmeirim, I., Henrique, D., Ish-Horowicz, D. and Pourquié, O. (1997). Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell* **91**, 639–648.

Palmieri, S. L., Peter, W., Hess, H. and Schöler, H. R. (1994). Oct-4 Transcription Factor Is Differentially Expressed in the Mouse Embryo during Establishment of the First Two Extraembryonic Cell Lineages Involved in Implantation. *Dev.*

Biol. **166**, 259–267.

Parameswaran, M. and Tam, P. P. L. (1995). Regionalisation of cell fate and morphogenetic movement of the mesoderm during mouse gastrulation. *Dev. Genet.* **17**, 16–28.

Parr, B. A., Shea, M. J., Vassileva, G. and McMahon, A. P. (1993). Mouse Wnt genes exhibit discrete domains of expression in the early embryonic CNS and limb buds. *Development* **119**, 247–261.

Partanen, J., Schwartz, L. and Rossant, J. (1998). Opposite phenotypes of hypomorphic and Y766 phosphorylation site mutations reveal a function for Fgfr1 in anteroposterior patterning of mouse embryos. *Genes Dev.* **12**, 2332–2344.

Pearson, J. C., Lemons, D. and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. *Nat. Rev. Genet.* **6**, 893–904.

Perea-Gomez, a, Lawson, K. a, Rhinn, M., Zakin, L., Brûlet, P., Mazan, S. and Ang, S. L. (2001). Otx2 is required for visceral endoderm movement and for the restriction of posterior signals in the epiblast of the mouse embryo. *Development* **128**, 753–65.

Perea-Gomez, A., Vella, F. D. J., Shawlot, W., Oulad-Abdelghani, M., Chazaud, C., Meno, C., Pfister, V., Chen, L., Robertson, E., Hamada, H., et al. (2002). Nodal Antagonists in the Anterior Visceral Endoderm Prevent the Formation of Multiple Primitive …. *Dev. Cell* **3**, 745–756.

Perea-Gomez, A., Camus, A., Moreau, A., Grieve, K., Moneron, G., Dubois, A., Cibert, C. and Collignon, J. (2004). Initiation of gastrulation in the mouse embryo is preceded by an apparent shift in the orientation of the anterior-posterior axis. *Curr. Biol.* **14**, 197–207.

Pfister, S., Steiner, K. A. and Tam, P. P. L. (2007). Gene expression pattern and progression of embryogenesis in the immediate post-implantation period of mouse development. *Gene Expr. Patterns* **7**, 558–573.

Piotrowska-Nitsche, K., Perea-Gomez, A., Haraguchi, S. and Zernicka-Goetz, M. (2005). Four-cell stage mouse blastomeres have different developmental properties. *Development* **132**, 479–490.

Plusa, B., Piliszek, A., Frankenberg, S., Artus, J. and Hadjantonakis, A.-K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development* **135**, 3081–91.

Poss, K. D. (2010). Advances in understanding tissue regenerative capacity and mechanisms in animals. *Nat. Rev. Genet.* **11**, 710–722.

Pownall, M. E., Gustafsson, M. K. and Emerson, C. P. (2002). Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos. *Annu. Rev. Cell Dev. Biol.* **18**, 747–783.

Ralston, A. and Rossant, J. (2008). Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Dev. Biol.* **313**, 614–629.

Redmer, T., Diecke, S., Grigoryan, T., Quiroga-Negreira, A., Birchmeier, W. and Besser, D. (2011). E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. *EMBO Rep.* **12**, 720–6.

Reshef, R., Maroto, M. and Lassar, A. B. (1998). Regulation of dorsal somitic cell fates: BMPs and Noggin control the timing and pattern of myogenic regulator expression. *Genes Dev.* **12**, 290–303.

Richardson, L., Torres-Padilla, M. E. and Zernicka-Goetz, M. (2006). Regionalised signalling within the extraembryonic ectoderm regulates anterior visceral endoderm positioning in the mouse embryo. *Mech. Dev.* **123**, 288–296.

Rivera-Pérez, J. A. and Hadjantonakis, A.-K. (2015). The dynamics of morphogenesis in the early mouse embryo. *Cold Spring Harb. Perspect. Biol.* **7**, a015867.

Rivera-Pérez, J. A., Mager, J. and Magnuson, T. (2003). Dynamic morphogenetic events characterize the mouse visceral endoderm. *Dev. Biol.* **261**, 470–487.

Robb, L. and Tam, P. P. . L. (2004). Gastrula organiser and embryonic patterning in the mouse. *Semin. Cell Dev. Biol.* **15**, 543–554.

Rodriguez, T. A., Srinivas, S., Clements, M. P., Smith, J. C. and Beddington, R. S. P. (2005). Induction and migration of the anterior visceral endoderm is regulated by the extra-embryonic ectoderm. *Development* **132**, 2513–2520.

Rosenthal, N. and Brown, S. (2007). The mouse ascending: perspectives for human-disease models. *Nat. Cell Biol.* **9**, 993–999.

Ryan, A. K. and Rosenfeld, M. G. (1997). POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev.* **11**, 1207–1225.

Saga, Y. and Takeda, H. (2001). The making of the somite: molecular events in

vertebrate segmentation. *Nat. Rev. Genet.* **2**, 835–845.

Saga, Y., Hata, N., Koseki, H. and Taketo, M. M. (1997). Mesp2: A novel mouse gene expressed in the presegmented mesoderm and essential for segmentation initiation. *Genes Dev.* **11**, 1827–1839.

Saga, Y., Miyagawa-Tomita, S., Takagi, A., Kitajima, S., Miyazaki, J. and 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.

Saiz, N. and Plusa, B. (2013). Early cell fate decisions in the mouse embryo. *Reproduction* **145**, R65–R80.

Schoenwolf, G. C. (1984). Histological and ultrastructural studies of secondary neurulation in mouse embryos. *Am. J. Anat.* **169**, 361–376.

Schofield, P. N., Hoehndorf, R. and Gkoutos, G. V (2012). Mouse genetic and phenotypic resources for human genetics. *Hum. Mutat.* **33**, 826–36.

Schöler, H. R., Hatzopoulos, A. K., Balling, R., Suzuki, N. and Gruss, P. (1989). A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *EMBO J.* **8**, 2543–50.

Schrode, N., Saiz, N., Di Talia, S. and Hadjantonakis, A.-K. (2014). GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. *Dev. Cell* **29**, 454–467.

Shen, M. M. (2007). Nodal signaling: developmental roles and regulation. *Development* **134**, 1023–1034.

Shi, G. and Jin, Y. (2010). Role of Oct4 in maintaining and regaining stem cell pluripotency. *Stem Cell Res. Ther.* **1**, 39.

Shi, Y. and Liu, J.-P. (2011). Gdf11 facilitates temporal progression of neurogenesis in the developing spinal cord. *J. Neurosci.* **31**, 883–893.

Shum, A. S. W., Poon, L. L. M., Tang, W. W. T., Koide, T., Chan, B. W. H., Leung, Y. C. G., Shiroishi, T. and Copp, A. J. (1999). Retinoic acid induces down-regulation of Wnt-3a, apoptosis and diversion of tail bud cells to a neural fate in the mouse embryo. *Mech. Dev.* **84**, 17–30.

Singh, A. M., Hamazaki, T., Hankowski, K. E. and Terada, N. (2007). A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells* **25**, 2534–2542.

Sinha, M., Jang, Y. C., Oh, J., Khong, D., Wu, E. Y., Manohar, R., Miller, C., Regalado, S. G., Loffredo, F. S., Pancoast, J. R., et al. (2014). Restoring systemic GDF11 levels reverses age-related dysfunction in mouse skeletal muscle. *Science (80-.).* **344**, 649–652.

Smith, K. K. (2003). Time's arrow: Heterochrony and the evolution of development. *Int. J. Dev. Biol.* **47**, 613–621.

Smith, J. L., Gesteland, K. M. and Schoenwolf, G. C. (1994). Prospective fate map of the mouse primitive streak at 7.5 days of gestation. *Dev. Dyn.* **201**, 279–89.

Srinivas, S., Rodriguez, T., Clements, M., Smith, J. C. and Beddington, R. S. P. (2004). Active cell migration drives the unilateral movements of the anterior visceral endoderm. *Development* **131**, 1157–1164.

Stafford, D. A., Brunet, L. J., Khokha, M. K., Economides, A. N. and Harland, R. M. (2011). Cooperative activity of noggin and gremlin 1 in axial skeleton development. *Development* **138**, 1005–14.

Stern, C. D., Charité, J., Deschamps, J., Duboule, D., Durston, A. J., Kmita, M., Nicolas, J.-F., Palmeirim, I., Smith, J. C. and Wolpert, L. (2006). Head-tail patterning of the vertebrate embryo: one, two or many unresolved problems? *Int. J. Dev. Biol.* **50**, 3–15.

Stott, D., Kispert, A. and Herrmann, B. G. (1993). Rescue of the tail defect of Brachyury mice. *Genes Dev.* **7**, 197–203.

Strumpf, D., Mao, C.-A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F. and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132**, 2093–2102.

Subramanian, V., Meyer, B. I. and Gruss, P. (1995). Disruption of the murine homeobox gene Cdx1 affects axial skeletal identities by altering the mesodermal expression domains of Hox genes. *Cell* **83**, 641–653.

Sun, X., Meyers, E. N., Lewandoski, M. and Martin, G. R. (1999). Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes Dev.* **13**, 1834–1846.

Sutherland, A. E. (2016). Tissue morphodynamics shaping the early mouse embryo. *Semin. Cell Dev. Biol.* 1–10.

Szumska, D., Pieles, G., Essalmani, R., Bilski, M., Mesnard, D., Kaur, K., Franklyn,

A., El Omari, K., Jefferis, J., Bentham, J., et al. (2008). VACTERL/caudal regression/Currarino syndrome-like malformations in mice with mutation in the proprotein convertase Pcsk5. *Genes Dev.* **22**, 1465–1477.

Tabansky, I., Lenarcic, A., Draft, R. W., Loulier, K., Keskin, D. B., Rosains, J., Rivera-Feliciano, J., Lichtman, J. W., Livet, J., Stern, J. N. H., et al. (2013). Developmental bias in cleavage-stage mouse blastomeres. *Curr. Biol.* **23**, 21–31.

Takada, S., Stark, K. L., Shea, M. J., Vassileva, G., McMahon, J. A. and McMahon, A. P. (1994). Wnt-3a regulates somites and tailbud formation in the mouse embryo. *Genes Dev.* **8**, 174–189.

Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676.

Takaoka, K., Yamamoto, M., Shiratori, H., Meno, C., Rossant, J., Saijoh, Y. and Hamada, H. (2006). The mouse embryo autonomously acquires anterior-posterior polarity at implantation. *Dev. Cell* **10**, 451–459.

Takaoka, K., Yamamoto, M. and Hamada, H. (2011). Origin and role of distal visceral endoderm, a group of cells that determines anterior-posterior polarity of the mouse embryo. *Nat. Cell Biol.* **13**, 743–752.

Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D. M., Lovell-Badge, R., Papaioannou, V. E. and Kondoh, H. (2011). Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature* **470**, 394–398.

Tam, P. P. L. (1981). The control of somitogenesis in mouse embryos. *J. Embryol. Exp. Morphol.* **65 Suppl**, 103–128.

Tam, P. P. L. and Beddington, R. S. P. (1987). The formation of mesodermal tissues in the mouse embryo during gastrulation and early organogenesis. *Development* **99**, 109–26.

Tam, P. P. and Behringer, R. R. (1997). Mouse gastrulation: the formation of a mammalian body plan. *Mech. Dev.* **68**, 3–25.

Tam, P. P. L. and Loebel, D. A. F. (2007). Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* **8**, 368–81.

Tam, P. P. L. and Tan, S.-S. S. (1992). The somitogenetic potential of cells in the primitive streak and the tail bud of the organogenesis-stage mouse embryo.

Development **115**, 703–15.

Tam, P. P. L., Kanai-Azuma, M. and Kanai, Y. (2003). Early endoderm development in vertebrates: Lineage differentiation and morphogenetic function. *Curr. Opin. Genet. Dev.* **13**, 393–400.

Tan, F., Qian, C., Tang, K., Abd-Allah, S. M. and Jing, N. (2015). Inhibition of Transforming Growth Factor β (TGF-β) Signaling can Substitute for Oct4 Protein in Reprogramming and Maintain Pluripotency. *J. Biol. Chem.* **290**, 4500–4511.

Tapia, N., Reinhardt, P., Duemmler, A., Wu, G., Araúzo-Bravo, M. J., Esch, D., Greber, B., Cojocaru, V., Rascon, C. A., Tazaki, A., et al. (2012). Reprogramming to pluripotency is an ancient trait of vertebrate Oct4 and Pou2 proteins. *Nat. Commun.* **3**, 1279.

Tesar, P. J., Chenoweth, J. G., Brook, F. a, Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L. and McKay, R. D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199.

Thomas, P. and Beddington, R. (1996). Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo. *Curr. Biol.* **6**, 1487–1496.

Torres-Padilla, M. E., Richardson, L., Kolasinska, P., Meilhac, S. M., Luetke-Eversloh, M. V. and Zernicka-Goetz, M. (2007). The anterior visceral endoderm of the mouse embryo is established from both preimplantation precursor cells and by de novo gene expression after implantation. *Dev. Biol.* **309**, 97–112.

Tortelote, G. G., Hernández-Hernández, J. M., Quaresma, A. J. C., Nickerson, J. A., Imbalzano, A. N. and Rivera-Pérez, J. A. (2013). Wnt3 function in the epiblast is required for the maintenance but not the initiation of gastrulation in mice. *Dev. Biol.* **374**, 164–173.

Tsakiridis, A., Huang, Y., Blin, G., Skylaki, S., Wymeersch, F., Osorno, R. R., Economou, C., Karagianni, E., Zhao, S., Lowell, S., et al. (2014). Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. *Development* **141**, 1209–1221.

Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V. and Nicolas, J.-F. (2009). Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Dev. Cell* **17**, 365–76.

van den Akker, E., Forlani, S., Chawengsaksophak, K., Graaff, W. De and Beck, F. (2002). Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation. **2193**, 2181–2193.

van Nes, J., de Graaff, W., Lebrin, F., Gerhard, M., Beck, F. and Deschamps, J. (2006). The Cdx4 mutation affects axial development and reveals an essential role of Cdx genes in the ontogenesis of the placental labyrinth in mice. *Development* **133**, 419–428.

van Rooijen, C., Simmini, S., Bialecka, M., Neijts, R., van de Ven, C., Beck, F. and Deschamps, J. (2012). Evolutionarily conserved requirement of Cdx for post-occipital tissue emergence. *Development* **139**, 2576–2583.

Varlet, I., Collignon, J. and Robertson, E. J. (1997). Nodal expression in the primitive endoderm is required for specification of the anterior axis during mouse gastrulation. *Development* **124**, 1033–1044.

Viebahn, C., Stortz, C., Mitchell, S. A. and Blum, M. (2002). Low proliferative and high migratory activity in the area of Brachyury expressing mesoderm progenitor cells in the gastrulating rabbit embryo. *Development* **129**, 2355–2365.

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**, 655–61.

Vincent, S. D., Dunn, N. R., Hayashi, S., Norris, D. P. and Robertson, E. J. (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. *Genes Dev.* **17**, 1646–1662.

Voiculescu, O., Bertocchini, F., Wolpert, L., Keller, R. E. and Stern, C. D. (2007). The amniote primitive streak is defined by epithelial cell intercalation before gastrulation. *Nature* **449**, 1049–52.

Vonk, F. J., Casewell, N. R., Henkel, C. V, Heimberg, A. M., Jansen, H. J., McCleary, R. J. R., Kerkkamp, H. M. E., Vos, R. A., Guerreiro, I., Calvete, J. J., et al. (2013). The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20651–6.

Wahl, M. B., Deng, C., Lewandoski, M. and Pourquié, O. (2007). FGF signaling acts upstream of the NOTCH and WNT signaling pathways to control segmentation clock oscillations in mouse somitogenesis. *Development* **134**,

4033–4041.

Walker, R. G., Poggioli, T., Katsimpardi, L., Buchanan, S. M., Oh, J., Wattrus, S., Heidecker, B., Fong, Y. W., Rubin, L. L., Ganz, P., et al. (2016). Biochemistry and Biology of GDF11 and MyostatinResponse to Walker et al. *Circ. Res.* **118**, 1125–1142.

Wellik, D. M. and Capecchi, M. R. (2003). Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. *Science* **301**, 363–367.

White, L. A., Mitchell, T. I. and Brinckerhoff, C. E. (2000). Transforming growth factor β inhibitory element in the rabbit matrix metalloproteinase-1 (collagenase-1) gene functions as a repressor of constitutive transcription. *Biochim. Biophys. Acta - Gene Struct. Expr.* **1490**, 259–268.

Williams, M., Burdsal, C., Periasamy, A., Lewandoski, M. and Sutherland, A. (2012). Mouse primitive streak forms in situ by initiation of epithelial to mesenchymal transition without migration of a cell population. *Dev. Dyn.* **241**, 270–283.

Wilson, V. and Beddington, R. S. P. (1996). Cell fate and morphogenetic movement in the late mouse primitive streak. *Mech. Dev.* **55**, 79–89.

Wilson, V. and Beddington, R. S. P. (1997). Expression of T protein in the primitive streak is necessary and sufficient for posterior mesoderm movement and somite differentiation. *Dev. Biol.* **192**, 45–58.

Wilson, V. and Lawson, K. A. (2016). Development of Early Embryonic Lineages. In *Kaufman's Atlas of Mouse Development Supplement* (ed. Baldock, R.), Bard, J.), Davidson, D. R.), and Morriss-Kay, G.), pp. 65–76. London: Academic Press.

Wilson, V., Manson, L., Skarnes, W. C. and Beddington, R. S. P. (1995). The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. *Development* **121**, 877–86.

Wilson, V., Olivera-Martínez, I., Storey, K. G., Olivera-Martinez, I., Storey, K. G., Olivera-Martínez, I., Storey, K. G., Olivera-Martinez, I. and Storey, K. G. (2009). Stem cells, signals and vertebrate body axis extension. *Development* **136**, 2133–2133.

Winnier, G., Blessing, M., Labosky, P. and Hogan, B. L. M. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev.* **9**, 2105–2116.

Wolpert, L., Beddington, R. S. P., Brockes, J., Jessell, T. M., Lawrence, P. and Meyerowitz, E. (1998). *Principles of Development*. 1st ed. London: Oxford University Press.

Woltering, J. M. (2012). From Lizard to Snake; Behind the Evolution of an Extreme Body Plan. *Curr. Genomics* **13**, 289–299.

Woltering, J. M., Vonk, F. J., Müller, H., Bardine, N., Tuduce, I. L., de Bakker, M. a. G., Knöchel, W., Sirbu, I. O., Durston, A. J. and Richardson, M. K. (2009). Axial patterning in snakes and caecilians: Evidence for an alternative interpretation of the Hox code. *Dev. Biol.* **332**, 82–89.

Wu, H. H., Ivkovic, S., Murray, R. C., Jaramillo, S., Lyons, K. M., Johnson, J. E. and Calof, A. L. (2003). Autoregulation of neurogenesis by GDF11. *Neuron* **37**, 197–207.

Wymeersch, F. J., Huang, Y., Blin, G., Cambray, N., Wilkie, R., Wong, F. C. and Wilson, V. (2016). Position-dependent plasticity of distinct progenitor types in the primitive streak. *Elife* **5**,.

Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N. and McMahon, A. P. (1999). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes Dev.* **13**, 3185–3190.

Yamamoto, M., Saijoh, Y., Perea-Gomez, A., Shawlot, W., Behringer, R. R., Ang, S.-L., Hamada, H. and Meno, C. (2004). Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo. *Nature* **428**, 387–392.

Yamanaka, Y., Tamplin, O. J., Beckers, A., Gossler, A. and Rossant, J. (2007). Live imaging and genetic analysis of mouse notochord formation reveals regional morphogenetic mechanisms. *Dev. Cell* **13**, 884–896.

Yamanaka, Y., Lanner, F. and Rossant, J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715–724.

Yeo, C. Y. and Whitman, M. (2001). Nodal signals to Smads through Cripto-dependent and Cripto-independent mechanisms. *Mol. Cell* **7**, 949–957.

Yeom, Y. I., Fuhrmann, G., Ovitt, C. E., Brehm, A., Ohbo, K., Gross, M., Hubner, K. and Scholer, H. R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881–894.

Yoon, Y., Huang, T., Tortelote, G. G., Wakamiya, M., Hadjantonakis, A.-K.,

Behringer, R. R. and Rivera-Pérez, J. A. (2015). Extra-embryonic Wnt3 regulates the establishment of the primitive streak in mice. *Dev. Biol.* **403**, 80–88.

Yoshikawa, Y., Fujimori, T., McMahon, a P. and Takada, S. (1997). Evidence that absence of Wnt-3a signaling promotes neuralization instead of paraxial mesoderm development in the mouse. *Dev. Biol.* **183**, 234–242.

Young, T., Rowland, J. E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J.-N. N., et al. (2009). Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Dev. Cell* **17**, 516–26.

Ziomek, C. A., Johnson, M. H. and Handyside, A. H. (1982). The developmental potential of mouse 16-cell blastomeres. *J. Exp. Zool.* **221**, 345–355.

"And either way you turn, I'll be there
Open up your skull, I'll be there
Climbing up the walls"

Radiohead –Ok Computer - *Climbing up the walls*

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt