Identification of *cis*-regulatory modules in mouse embryonic limb buds and heart using endogenous epitope-tagged transcription factors

*Inauguraldissertation*

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Frédéric Laurent

aus Fey, VD

Basel, 2014
Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von
Prof. Dr. Rolf Zeller (Dissertationsleiter und Fakultätsverantwortlicher),
Prof. Dr. Verdon Taylor (Korreferent)

Basel, den 16. September 2014

Prof. Dr. Jörg Schibler
Dekan
# Table of contents

1. Table of contents .................................................................................................................. 3

2. Summary ............................................................................................................................... 7

3. List of Abbreviations ............................................................................................................ 9

4. Introduction ........................................................................................................................... 11
   4.1. Cis-regulatory modules and transcriptional control of embryonic development.......... 11
   4.2. Morphogenesis of the Heart .......................................................................................... 16
       4.2.1. Composition and function of the adult mammalian heart .................................. 16
       4.2.2. Overview of the main steps of cardiac development in the mouse ................. 19
       4.2.3. From cardiogenic mesoderm to the myocardium and endocardium: determination of the cardiac lineages ................................................................. 21
       4.2.4. Patterning and differentiation of the second heart field ................................. 23
       4.2.5. The T-box genes family and the development of the atrioventricular canal and atrioventricular node ............................................................... 26
       4.2.6. Formation of the cardiac valves: from endothelial-mesenchymal transition to the cardiac cushions............................................................... 29
       4.2.7. Hand2 is a crucial regulator for the development of second heart field derived structures............................................................... 32
   4.3. The BMP signaling pathway during limb bud development and the Grem1 cis-regulatory landscape ......................................................................................... 38

5. Aims of the Thesis ............................................................................................................... 41

6. Results ................................................................................................................................... 43
   6.1. From cis-regulatory modules to trans-acting factors: control of Grem1 genomic landscape during limb bud development and generation of Smad4\(^{3x\text{FLAG}}\) mice by homologous recombination .................................................. 43
      6.1.1. Paper 1: “Smad4 is required to induce digit ray primordia and to initiate the aggregation and differentiation of chondrogenic progenitors in mouse limb buds” .... 43
      6.1.2. Paper 2: “Conserved cis-regulatory regions in a large genomic landscape control SHH and BMP-regulated Gremlin1 expression in mouse limb buds” ............ 55
      6.1.3. The Grem1 cis-regulatory landscape integrates several signaling inputs from multiple CRMs ................................................................................................. 77
      6.1.4. Identification of SMAD4 target CRMs during limb bud development: the need for a new tool ................................................................................................. 79
      6.1.5. Generation of Smad4\(^{3x\text{FLAG}}\) mice by homologous recombination ..................... 80
   6.2. From trans-acting factor to cis-regulatory modules: identification of the direct transcriptional targets of HAND2 during heart morphogenesis ........................................ 85
6.2.1. The Hand2^{2xFLAG} allele is a new genetic tool to study HAND2 distribution during heart development ................................................................. 85

6.2.2. ChIP-Seq analysis using pooled Hand2 expressing tissues identifies the genome-wide HAND2 target regions in association with tissue-specific gene functions .......... 87

6.2.3. HAND2 controls the expression of genes involved in early second heart field morphogenesis .................................................................................................................... 91

6.2.4. HAND2 controls the expression of genes associated with known cardiac-specific enhancers ............................................................................................................................... 93

6.2.5. Hand2-deficient mouse embryos lack the delaminated endocardial cells in the atrioventricular canal .................................................................................................................. 96

6.2.6. HAND2 directly regulates Has2-mediated extracellular matrix deposition in the cardiac jelly ............................................................................................................................ 97

6.2.7. Twist1 but not Sox9 is a direct transcriptional target of HAND2 during cardiac cushion formation .............................................................................................................. 99

6.2.8. Snai1-dependent initiation of endothelial-mesenchymal transition in the atrioventricular canal is regulated by HAND2 ................................................................................................. 102

6.2.9. The Hand2 loss-of-function affects the EndMT upstream signaling but not the patterning of the myocardium .................................................................................................... 104

6.2.10. Tbx2 and Tbx3 are direct targets of HAND2 during cardiac development ...... 106

7. Discussion .......................................................................................................................... 113

7.1. Identification of HAND2 target gene networks that function during heart morphogenesis .............................................................................................. 113

7.1.1. Genome-wide identification of cardiac CRMs ................................................................. 113

7.1.2. HAND2 and its target CRMs in the heart: direct and indirect DNA-binding interactions ............................................................................................................................ 114

7.1.3. HAND2 target genes in the progenitors of the second heart field ............................. 115

7.1.4. Furin and Myocd are direct transcriptional targets of HAND2 ........................................ 116

7.1.5. A causal link between atrioventricular canal defects and embryonic lethality in Hand2-deficient embryos? .................................................................................. 117

7.2. Gene networks controlling endothelial-mesenchymal transition in the atrioventricular canal are directly regulated by HAND2 .................................................................................. 119

7.2.1. Hand2 participates in extracellular matrix deposition within the cardiac jelly ...... 119

7.2.2. HAND2 directly regulates Twist1 and Snai1 expression during formation of the AVC cardiac cushions ................................................................................................................ 120

7.2.3. HAND2 is an intermediate factor in the myocardial/endocardial signaling events that induce the EndMT ........................................................................................................ 124

7.2.4. A role for Hand2 in cardiac valves maturation and remodeling? .............................. 126

7.3. The paralogous Tbx2 and Tbx3 genomic landscapes encode heart and limb bud specific CRMs bound by HAND2 chromatin complexes ........................................... 126

8. Conclusions and Outlook ................................................................................................. 129
9. Materials and Methods ................................................................. 131

9.1. Genetic engineering of plasmids .................................................. 131

9.1.1. General cloning procedures ................................................... 131

9.1.2. Construction of pBSK Smad4\^\text{FLAG} targeting vector for Homologous
Recombination .............................................................................. 132

9.2. Cell culture .................................................................................. 133

9.2.1. Embryonic Fibroblasts (EMFIs) culture ...................................... 133

9.2.2. Embryonic Stem (ES) cell culture ............................................ 134

9.2.3. ES cells electroporation and selection ....................................... 135

9.2.4. ES cell clones screening by Southern Blot .................................. 136

9.3. Mice husbandry and experiments with embryos .......................... 138

9.3.1. Ethic statement ........................................................................ 138

9.3.2. Mouse strains .......................................................................... 138

9.3.3. Generation of LacZ reporter transgenic embryos ...................... 138

9.3.4. Whole-mount LacZ staining of mouse embryos ......................... 139

9.3.5. Whole-mount \textit{in situ} hybridization (WISH) ............................ 139

9.3.6. Preparation of digoxigenin-labelled RNA riboprobe .................. 140

9.4. Molecular Biology ........................................................................ 141

9.4.1. Immunocytochemistry ............................................................. 141

9.4.2. Protein detection by Western Blot (WB) .................................... 141

9.4.3. Immunoprecipitation (IP) ......................................................... 143

9.4.4. Chromatin Immunoprecipitation (ChIP) ................................. 144

9.4.5. Extraction of RNA and cDNA synthesis ................................... 147

9.4.6. Real-time quantitative PCR (RT-qPCR) .................................... 148

9.5. Histology ..................................................................................... 149

9.5.1. Paraffin embedding procedure ................................................. 149

9.5.2. Preparation of frozen sections ................................................. 150

9.5.3. Immunohistochemistry ............................................................. 150

9.5.4. TUNEL assay .......................................................................... 151

9.5.5. Hematoxylin/Eosin staining ..................................................... 152

9.5.6. Alcian Blue staining ................................................................. 152

9.5.7. \textit{In situ} hybridization on sections ......................................... 152

9.6. Genomics online resources ......................................................... 154

9.7. Tables .......................................................................................... 154

9.7.1. Antibodies tables ..................................................................... 154

9.7.2. Primers tables .......................................................................... 155
10. Acknowledgments ................................................................. 159
11. Bibliography .......................................................................... 161
12. Appendixes ............................................................................. 187
  12.1. Generation of Smad4^3xFLAG mice using dRMCE .............. 187
    12.1.1. Testing three epitope-tagged versions of SMAD4 in ES cells ... 187
    12.1.2. Embryonic lethality of Smad4^3xFLAG homozygous embryos .... 191
  12.2. Additional methods relative to the dRMCE targeting .......... 194
    12.2.1. Construction of the pcDNA3 Smad4^3xTag expression vectors ... 194
    12.2.2. Construction of the pDREV S4^3xTag targeting vectors for dRMCE ... 195
    12.2.3. Transfection of HEK-293T cells ....................................... 196
    12.2.4. dRMCE: ES cells electroporation, selection and screening .... 197
    12.2.5. Differentiation of ES cells to embryoid bodies .................. 197
    12.2.6. dRMCE primers table .................................................. 198
  12.3. Manuscripts in preparation ................................................... 199
  12.4. Curriculum vitae ................................................................. 201
2. Summary

The genes that control mouse embryonic development are tightly regulated spatially and temporally by the integration of various signaling pathways. The interaction of downstream transcription factors with cis-regulatory modules located in the genomic landscapes of their target genes is crucial to control gene expression and enable the growth and patterning of embryonic tissues. However, in most of the cases, the genome-wide range of target regions bound by a given trans-acting factor in a tissue-specific manner is not known. In this thesis, I have taken advantage of epitope-tagging essential transcriptional regulators such as SMAD4 and HAND2 to gain mechanistic insight into their functions during limb and heart development.

In particular, the morphogenesis of the limb bud depends on a self-regulatory system of signaling feedback loops in the core of which the BMP antagonist GREMLIN1 plays a critical role. We analyzed the cis-regulatory landscape of Gremlin1 and identified a conserved region that recapitulates several aspects of the dynamic expression of Gremlin1 during limb bud development and requires the BMP signal transducer SMAD4 to be expanded in the distal anterior part of the limb bud. To investigate the Gremlin1 cis-regulatory modules that directly interact with SMAD4-containing chromatin complexes, I have generated a new mouse line that expresses a 3xFLAG epitope-tagged version of SMAD4 from its endogenous locus, with which we will perform ChIP-Seq analysis from developing embryonic tissues, including limb buds.

Following a similar rationale, I have used ChIP-Seq data from a mouse line expressing a Hand2\(^{3xFLAG}\) allele to provide the first identification of the HAND2 target gene networks that function during cardiogenesis, which is severely impaired in the absence of HAND2. This analysis indicates that HAND2 directly controls the expression of the paralogous genes Tbx2 and Tbx3 in the progenitors of the second heart field that give rise to the outflow tract and right ventricle. This regulation is mediated by the interaction of HAND2 with tissue-specific cis-regulatory modules located in Tbx2 and Tbx3 genomic landscapes. Furthermore, we observed in Hand2-
deficient embryos that the cells of the atrioventricular canal do not undergo the endothelial-mesenchymal transition that normally initiates the formation of the cardiac cushions, which are the precursors of the cardiac valves. We established that the expression of *Snai1*, *Twist1* and *Has2*, which are key genes in this process, are directly controlled by HAND2-containing chromatin complexes. Epitope-tagging of transcription factors is thus a powerful tool to identify the range of cis-regulatory modules they bind to, which allows to uncover the transcriptional networks they control during embryonic development.
### 3. List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AER</td>
<td>Apical Ectodermal Ridge</td>
</tr>
<tr>
<td>AVC</td>
<td>Atrioventricular Canal</td>
</tr>
<tr>
<td>AVN</td>
<td>Atrioventricular Node</td>
</tr>
<tr>
<td>BAC</td>
<td>Bacterial Artificial Chromosome</td>
</tr>
<tr>
<td>bHLH</td>
<td>basic Helix-Loop-Helix</td>
</tr>
<tr>
<td>BMP</td>
<td>Bone Morphogenetic Protein</td>
</tr>
<tr>
<td>BRE</td>
<td>BMP Responsive Element</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>CHD</td>
<td>Congenital Heart Disease</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin Immunoprecipitation</td>
</tr>
<tr>
<td>CMV</td>
<td>Cytomegalovirus</td>
</tr>
<tr>
<td>cNCC</td>
<td>cardiac Neural Crest Cells</td>
</tr>
<tr>
<td>Cq</td>
<td>Quantification Cycle</td>
</tr>
<tr>
<td>CRM</td>
<td>Cis-Regulatory Module</td>
</tr>
<tr>
<td>CS</td>
<td>Coding Sequence</td>
</tr>
<tr>
<td>CTCF</td>
<td>CCCTC-binding Factor</td>
</tr>
<tr>
<td>Cx40</td>
<td>Connexin 40</td>
</tr>
<tr>
<td>DBH</td>
<td>Dopamine Beta-Hydroxylase</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethylpyrocarbonate</td>
</tr>
<tr>
<td>DIG</td>
<td>Digoxigenin</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl Sulfoxide</td>
</tr>
<tr>
<td>DNase HS</td>
<td>DNase I Hypersensitive Site</td>
</tr>
<tr>
<td>dRMCE</td>
<td>dual Recombinase-Mediated</td>
</tr>
<tr>
<td>E</td>
<td>Embryonic day</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular Matrix</td>
</tr>
<tr>
<td>EMFI</td>
<td>Embryonic Mouse Fibroblast</td>
</tr>
<tr>
<td>EMT</td>
<td>Epithelial-Mesenchymal Transition</td>
</tr>
<tr>
<td>EndMT</td>
<td>Endothelial-Mesenchymal Transition</td>
</tr>
<tr>
<td>ES cell</td>
<td>Embryonic Stem cell</td>
</tr>
<tr>
<td>EUCOMM</td>
<td>European Conditional Mouse</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal Bovine Serum</td>
</tr>
<tr>
<td>FCS</td>
<td>Fetal Calf Serum</td>
</tr>
<tr>
<td>FGF</td>
<td>Fibroblast Growth Factor</td>
</tr>
<tr>
<td>FLK1</td>
<td>Fetal Liver Kinase 1</td>
</tr>
<tr>
<td>FMN1</td>
<td>Formin1</td>
</tr>
<tr>
<td>GBR</td>
<td>GLI binding region</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescent Protein</td>
</tr>
<tr>
<td>GO</td>
<td>Gene Ontology</td>
</tr>
<tr>
<td>GRE1</td>
<td>GLI Responsive Element 1</td>
</tr>
<tr>
<td>GREAT</td>
<td>Genomic Regions Enrichment Annotation Tool</td>
</tr>
<tr>
<td>GREM1</td>
<td>Gremlin 1</td>
</tr>
<tr>
<td>GRS1</td>
<td>Gremlin Regulatory Sequence 1</td>
</tr>
<tr>
<td>HA</td>
<td>Hyaluronic Acid</td>
</tr>
<tr>
<td>HA-tag</td>
<td>Hemagglutinin tag</td>
</tr>
<tr>
<td>HAND2</td>
<td>Heart, Autonomic nervous system and Neural crest Derivatives 2</td>
</tr>
<tr>
<td>HAS2</td>
<td>Hyaluronan Synthase 2</td>
</tr>
<tr>
<td>HCN4</td>
<td>Hyperpolarization-activated Cyclic Nucleotide-gated potassium channel 4</td>
</tr>
<tr>
<td>HEK</td>
<td>Human Embryonic Kidney cell</td>
</tr>
<tr>
<td>HEY</td>
<td>Hairy/Enhancer-of-split related with YRPW motif</td>
</tr>
<tr>
<td>HMCO</td>
<td>Human Mouse Chicken Opossum</td>
</tr>
<tr>
<td>HR</td>
<td>Homologous Recombination</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish Peroxidase</td>
</tr>
<tr>
<td>ICC</td>
<td>Immunocytochemistry</td>
</tr>
<tr>
<td>ID1</td>
<td>Inhibitor of Differentiation 1</td>
</tr>
<tr>
<td>IHc</td>
<td>Immunohistochemistry</td>
</tr>
<tr>
<td>IKMC</td>
<td>International Knockout Mouse Consortium</td>
</tr>
<tr>
<td>IPP</td>
<td>Immunoprecipitation</td>
</tr>
<tr>
<td>ISL1</td>
<td>Islet LIM homebox 1</td>
</tr>
<tr>
<td>LA</td>
<td>Left Atrium</td>
</tr>
<tr>
<td>LIF</td>
<td>Leukemia Inhibitory Factor</td>
</tr>
<tr>
<td>LV</td>
<td>Left Ventricle</td>
</tr>
<tr>
<td>MCP</td>
<td>Multipotent Cardiac Progenitors</td>
</tr>
<tr>
<td>MEF2C</td>
<td>Myocyte Enhancer Factor 2C</td>
</tr>
<tr>
<td>MESP1</td>
<td>Mesoderm Posterior 1 homolog</td>
</tr>
<tr>
<td>MET</td>
<td>Mesenchymal-Epithelial Transition</td>
</tr>
<tr>
<td>MSX1</td>
<td>Msh homebox 1</td>
</tr>
<tr>
<td>MYOCD</td>
<td>Myocardin</td>
</tr>
<tr>
<td>NDR</td>
<td>Nuclear Dbf2-Related kinases</td>
</tr>
<tr>
<td>Neo</td>
<td>Neomycin</td>
</tr>
<tr>
<td>NFATC1</td>
<td>Nuclear Factor of Activated T-cells, Cytoplasmic 1</td>
</tr>
<tr>
<td>NPPA</td>
<td>Natriuretic Peptide A</td>
</tr>
<tr>
<td>O.C.T</td>
<td>Optimum Cutting Temperature</td>
</tr>
<tr>
<td>OFT</td>
<td>Outflow Tract</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate Buffered Saline</td>
</tr>
<tr>
<td>PBT</td>
<td>PBS with 0.1% Tween 20</td>
</tr>
<tr>
<td>PEO</td>
<td>Proepicardial Organ</td>
</tr>
<tr>
<td>PFA</td>
<td>Paraformaldehyde</td>
</tr>
<tr>
<td>PHF</td>
<td>Primary Heart Field</td>
</tr>
<tr>
<td>polyA</td>
<td>Polyadenylation</td>
</tr>
<tr>
<td>PRX1</td>
<td>Paired Related homeobox 1</td>
</tr>
<tr>
<td>Puro</td>
<td>Puromycin</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>RA</td>
<td>Right Atrium</td>
</tr>
<tr>
<td>RPL19</td>
<td>Ribosomal Protein L19</td>
</tr>
<tr>
<td>rpm</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>RT</td>
<td>Room Temperature</td>
</tr>
<tr>
<td>RV</td>
<td>Right Ventricle</td>
</tr>
<tr>
<td>SAN</td>
<td>Sinoatrial Node</td>
</tr>
<tr>
<td>SB</td>
<td>Southern Blot</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium Dodecyl Sulfate</td>
</tr>
<tr>
<td>SHF</td>
<td>Second Heart Field</td>
</tr>
<tr>
<td>SHH</td>
<td>Sonic Hedgehog</td>
</tr>
<tr>
<td>SNAI1</td>
<td>Snail family zinc finger 1</td>
</tr>
<tr>
<td>SOX9</td>
<td>SRY (sex determining region Y)-box 9</td>
</tr>
<tr>
<td>SV40</td>
<td>Simian Virus 40</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription Factor</td>
</tr>
<tr>
<td>TGFβ2</td>
<td>Transforming Growth Factor Beta 2</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcriptional Start Site</td>
</tr>
<tr>
<td>TUNEL</td>
<td>Terminal deoxynucleotidyl transferase dUTP nick end labeling</td>
</tr>
<tr>
<td>UCSC</td>
<td>University of California Santa Cruz</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated Region</td>
</tr>
<tr>
<td>V5-tag</td>
<td>Simian Virus 5 tag</td>
</tr>
<tr>
<td>VE-Cad</td>
<td>Vascular Endothelial Cadherin</td>
</tr>
<tr>
<td>VEC</td>
<td>Valve Endothelial Cell</td>
</tr>
<tr>
<td>VEGF</td>
<td>Vascular Endothelial Growth Factor</td>
</tr>
<tr>
<td>VIC</td>
<td>Valve Interstitial Cell</td>
</tr>
<tr>
<td>VIM</td>
<td>Vimentin</td>
</tr>
<tr>
<td>WB</td>
<td>Western Blot</td>
</tr>
<tr>
<td>WISH</td>
<td>Whole-mount in situ Hybridization</td>
</tr>
<tr>
<td>WNT</td>
<td>Wingless-type MMTV integration site</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-Type</td>
</tr>
<tr>
<td>βglob</td>
<td>β-globin</td>
</tr>
</tbody>
</table>
4. Introduction

4.1. Cis-regulatory modules and transcriptional control of embryonic development

The question of how complex, multicellular-types tissues and organs arise from undifferentiated progenitor cells during embryonic development is fascinating. Cell fates are controlled by robust spatio-temporal and tissue-specific gene expressions, which progressively shape the body plan of the developing embryo. Growth and patterning are regulated by combinatorial interactions between transcription factors and multiple genetic cis-regulatory modules, whose integrated action determines the quantitative transcriptional output of a given protein-coding gene.

*Cis*-regulatory modules (CRMs) are composed of clusters of binding sites for multiple transcription factors (TFs) and can act as enhancers, silencers or insulators (Fig.1A). An enhancer is able to activate transcription independently of its location or orientation relative to a promoter, often in a tissue-specific manner (Bulger and Groudine, 2011; Ong and Corces, 2011; Levine, 2010) while a silencer will have the opposite effect and repress the expression of a target gene (Petrykowska et al., 2008). Insulators are CRMs that can either block the action of others CRMs on a promoter when located between them, or create barriers to prevent the spreading of heterochromatin compaction (Gaszner and Felsenfeld, 2006). Insulators are typically interacting with the ubiquitously expressed CTCF zinc-finger protein (Bell et al., 1999; Ong and Corces, 2014). Functional redundancy between CRMs acting on the same gene can occur, which confers phenotypic robustness ("shadow enhancers", (Barolo, 2012; Frankel et al., 2010). CRMs can be intergenic, embedded in introns of unrelated neighboring genes, or located in gene deserts (Zeller and Zuniga, 2007; Nobrega et al., 2003). In addition to CRMs proximal to the promoter of their target gene, CRMs usually operate over distances shorter than 50kb but can occasionally act over considerably longer distances (up to 2-3Mb, (Krivega and Dean, 2012)). They are often scattered in large genomic landscapes, or archipelagos, and cooperate to synergistically act on their target genes (Montavon et al., 2011; Marinic et al., 2013).
The physical contact between distant-acting CRMs and the basal transcriptional machinery at the promoter depends on the three-dimensional architecture of the genome and higher-order chromatin interactions, with regions of open euchromatin forming loops that bring together multiple genomic elements (Fig. 1B). The formation of these loops can be initiated de novo following an instructive process or be preformed in a ground state, and depends on CRMs-TFs interactions (de Laat and Duboule, 2013).

Figure 1. Cis-regulatory modules and genomic architecture
(A) Categories of cis-regulatory modules. From top to bottom: basal transcriptional machinery (RNA polymerase II and general transcription factors) bound to the promoter of a gene. Enhancer and silencer sequences mediate positive and negative effects on the transcription of a target gene. Insulator blocks the activity of an enhancer of gene A on gene B. Barrier insulator prevents the spread of chromatin condensation. Adapted from (Noonan and McCallion, 2010).

(B) Schematic representation of the cis-regulatory landscape of a gene. The chromatin is composed of DNA wrapped around histones to form nucleosomes and can be in different compaction states. Open chromatin loops to bring CRMs in physical contact with the promoter and basal transcriptional machinery. Adapted from (Lenhard et al., 2012).

The accessibility of CRMs to TFs during developmental progression depends on the state of chromatin compaction. Active CRMs are located in regions of open chromatin and are therefore sensitive to digestion by DNA nucleases (Crawford et al., 2006). They are also bound by the ubiquitously expressed p300 histone acetyltransferase (Visel et al., 2009a). Indeed, the chromatin signature of active CRMs can also be determined by the presence of specific post-translational modifications on the
histones tails in the nucleosomes flanking these CRMs (Calo and Wysocka, 2013; Spitz and Furlong, 2012). The acetylation of the lysine 27 of Histone 3 (H3K27ac), a high H3K4me1/H3K4me3 ratio and the absence of the repressive H3K27me3 are all typically associated with active CRMs, high levels of H3K4me3 being observed at actively transcribed promoters (Fig.2A) (Heintzman et al., 2007; Bonn et al., 2012; Creyghton et al., 2010; Rada-Iglesias et al., 2011; Djebali et al., 2012).

The best technique to assess the interactions between a TF of interest and its target CRMs is chromatin immunoprecipitation (ChIP, Fig.2B). ChIP relies on the chemical crosslinking of proteins and DNA interactions within a tissue or a living cell population (typically ~10⁷ cells), followed by shearing the chromatin into small fragments (200-300bp) by sonication. An antibody recognizing the TF of interest is then used to selectively immunoprecipitate the chromatin complexes in which this TF is contained. The associated DNA fragments are finally purified and can be used for deep sequencing to identify genome-wide target CRMs (ChIP-Seq, (Park, 2009), or for qPCR to confirm a specific TF-CRM interaction (ChIP-qPCR).

Based on these observations, several methods exist to identify CRMs within the genome that can be applied to specific tissues or developmental stages (Hardison and Taylor, 2012). DNase hypersensitivity profiling and ChIP-Seq using an antibody recognizing p300, H3K27ac or H3K4me1 can all identify candidate CRMs genome-wide (Cotney et al., 2012; May et al., 2012; Blow et al., 2010; Visel et al., 2009a; Shu et al., 2011). Chromosome conformation capture (3C) and derivatives (4C, Hi-C…) can identify all potential chromatin-chromatin interactions and thus all the CRMs that contact a specific promoter (Dekker et al., 2013). In addition to these experimental approaches, in silico comparative genomics can identify conserved non-coding elements between different species, even if sequence conservation does not always implies that the CRM function is conserved (Nelson and Wardle, 2013). Finally, computational approaches to identify the occurrence of TF binding motifs can also identify candidate CRMs (Narlikar et al., 2010).
Figure 2. Chromatin marks and ChIP-Seq procedure

(A) Representation of genome-wide chromatin signatures for a hypothetical cis-regulatory genomic landscape. From top to bottom: transcriptional start sites and exons of active genes (grey arrows and boxes) and associated CRMs (blue ovals). Blue: active CRMs are hypersensitive to DNase and are marked by p300 binding, H3K4me1 and H3K27ac. Orange: repressive H3K27me3 mark. Green: promoter-enriched H3K4me3 mark. Adapted from (Phillips-Cremins and Corces, 2013). (B) Main steps of ChIP-Seq procedure. DNA and interacting proteins from cultured cells or embryonic tissues are crosslinked and sheared in small chromatin fragments. Immunoprecipitation of the protein of interest (red) is achieved using either an antibody recognizing it directly (right) or recognizing an epitope tag (left, yellow circle). Following reverse crosslinking, the purified DNA fragments are sequenced and mapped to a reference genome. Adapted from (Visel et al., 2009b).

Transcription factors usually recognize small degenerate sequence-specific binding motifs of 6-12bp (Loots and Ovcharenko, 2004). The binding motifs of several different TFs can be observed within a CRM, leading to possible combinatorial binding of various TFs at different developmental time points and in different tissues at the same CRM. Furthermore, the binding of a specific TF can be a prerequisite for
the recruitment of additional TFs, cofactors and/or chromatin remodeling complexes that will modulate their function. Together, the cooperative and/or additive interactions between these TFs, or between TFs interacting with different CRMs, will determine the transcriptional output of their target genes (Wilczynski and Furlong, 2010; Yanez-Cuna et al., 2012; Spitz and Furlong, 2012; Farnham, 2009; MacQuarrie et al., 2011). Electrophoretic Mobility Shift Assay (EMSA) is traditionally used to assess the interaction between a TF and a CRM, but ChIP-qPCR is becoming more frequently used. To study the function and spatiotemporal expression pattern driven by a candidate CRM, the standard method is a reporter assays in which the candidate CRM is used to drive the expression of a reporter gene (LacZ, GFP, Luciferase…) in transgenic embryos or transfected cells. Finally, combining reporter assays with the mutation of TFs binding sites within a given CRM allows analyzing the requirement of specific TFs regarding the functionality and enhancing capacity of the considered CRM.

Genome-wide identification of CRMs interacting with a specific TF using ChIP-Seq analysis requires high quality antibodies. Tagging TFs with established epitope-tags represents an alternative to commercial antibodies that often lack specificity for their target protein (Fig.2B). For this, conventional genome editing techniques by homologous recombination or cassette exchange in mouse ES cells allow the introduction of an epitope tag in the endogenous locus coding for a specific TF. Targeting the endogenous locus maintains its transcriptional regulation by native promoter and CRMs. It allows studying the spatio-temporal distribution of the TF and the identification of its target genes in a physiological context, without off-target effects coming from overexpression (Zhang et al., 2008a; Conway et al., 2010). While most of the identified mutations associated with congenital developmental malformations are mapped to coding sequences, only a few have been located in CRMs (Gordon and Lyonnet, 2014; Smemo et al., 2012; Uslu et al., 2014; Lettice et al., 2003). The identification and mapping of CRMs is thus crucial for potential diagnosis of congenital malformations, and to understand how complex gene patterns are regulated during embryonic development.
4.2. Morphogenesis of the Heart

4.2.1. Composition and function of the adult mammalian heart

The heart is the first organ to form during embryonic development. Unlike other embryonic structures and organs, the heart and associated structures of the cardiovascular system are essential for the survival of the embryo. Perturbations in the complex gene regulatory networks controlling the formation of the cardiac structures can lead to various types of congenital heart diseases (CHD), which can be observed in 6 to 75 per 1000 live births and are one of the major cause of mortality in infants and adults (Bruneau, 2008; Khoshnood et al., 2012; Ishikawa et al., 2011; Wren et al., 2012; Brickner et al., 2000). The adult mammalian heart is a muscular organ composed of four distinct chambers, two atria and two ventricles, separated by cardiac valves. Its rhythmic and coordinated contractions driven by the conduction system pump blood to the organs of the body via the blood vessels, a function crucial to maintain the homeostasis of the organism (Fig.3A). The anatomy of the heart has significantly changed over the course of evolution. From a simple linear tube subjected to peristaltic contractions in primitive chordates, through a single atrium and single ventricle in fishes, the vertebrate heart evolved to a more powerful and complex pump by successive additions of new cardiac structures that likely occurred through the acquisition of new regulatory inputs and by gene duplications (Olson, 2006; Jensen et al., 2013). The cardiac chambers are mainly composed of three concentric tissue layers with different developmental origins: the endocardium is the endothelial lining of the heart and is in contact with the blood. The myocardium consists mainly of a muscular layer of contractile cardiomyocytes and conductive cells. Finally, the epicardium is a specialized epithelium mostly composed of connective tissues that envelop and protect the heart (Fig.3A). During embryogenesis, the endocardium is separated from the myocardium by a layer of extracellular matrix, the cardiac jelly. The cells of the myocardium are derived from two different lineages: the primary heart field and the second heart field (Buckingham et al., 2005).
Figure 3. Anatomy of the adult mammalian heart

(A) Blood circulation inside the heart (direction indicated by arrows). Deoxygenated blood (blue) returning from the body through the superior and inferior vena cava enters the right atrium, where cardiac contractions pumps it out through the tricuspid valve into the right ventricle. Delayed contractions pumps the blood into the pulmonary arteries through the pulmonary valve. Left and right pulmonary arteries are connected to the lungs, where the blood gets oxygenated (red) before returning to the heart via the pulmonary veins. The blood enters the left atrium, is pumped through the mitral valve into the left ventricle, and pumped out in the aorta through the aortic valve to be distributed to the remainder of the organism. Image modified from the Texas Heart Institute website. (B) Cardiac conduction system. The electrical impulse is generated in the sinoatrial node and contracts the atria. It is transmitted to the atroioventricular node, which propagates it with a delay to the bundle of His (right and left branches) along the interventricular septum. From the heart apex, the impulse is transmitted to the fast conducting Purkinje fibers, which contracts the ventricles. From Main Line Health website (http://www.mainlinehealth.org/oth/Page.asp?pageID= OTH005458). (C) Above view of a transverse heart section showing the opening and closure of the cardiac valves following cardiac contractions. The mitral (bicuspid) and tricuspid valves are composed of two and three leaflets, respectively. The pulmonary and aortic (semilunar) valves are composed of three leaflets called cusps. Adapted from (http://www.tokresource.org/tok_ classes/biobiobi/biomenu/transport_system/index.htm).

The cardiac contractions depend on an electrical impulse generated in the pacemaker cells of the sinoatrial node. This impulse is transmitted through the myocardium of the atria, leading to their simultaneous contraction, to the atroioventricular node and further to the ventricles to contract them (Fig.3B). A delay in the propagation of the impulse from the atroioventricular node allows non-simultaneous contractions between the two atria and the two ventricles, necessary
for unidirectional blood flow. The heart functions therefore as two serially working atrial and ventricular pumps (Christoffels and Moorman, 2009). In the mouse, this whole cardiac cycle takes less than 200ms for an average of 500 beats per minute (bpm) (Berul et al., 1996; Ho et al., 2011), which is much faster than the average human cardiac rhythm of 70bpm and reflects the effects of the size of an organism on the speed of its metabolism (Kleiber, 1947). Congenital malformations and aging can cause dysfunctions of the pacemaker tissues, resulting in arrhythmias. If the primary pacemaker in the sinoatrial node is impaired or if the electrical impulse is not propagated, the atrioventricular node can act as an accessory pacemaker and initiate an electrical impulse (Christoffels et al., 2010).

The cardiac valves are essential to avoid blood regurgitation between the chambers, and open and close over 700'000 times per day in the mouse. The adult heart contains two atrioventricular valves separating atria and ventricles (tricuspid and mitral (or bicuspid) valves), and two semilunar valves located between the ventricles and the great arteries (pulmonary and aortic valves) (Fig.3A). After the blood is pumped from the atria to the ventricles, the atrioventricular valves close and the semilunar valves open (Fig.3C). The inverse takes place following blood expulsion into the great arteries. Each valve consists of 2-3 leaflets, composed of an outer layer of valve endothelial cells (VEC) enveloping three stratified layers of extracellular matrix (collagen, proteoglycan and elastin) interspersed with valve interstitial cells (VIC). The composition of these extracellular matrix layers is essential for the biomechanical properties of the leaflets to open and close, and the VIC and VEC plays additional roles in their homeostasis (Hinton et al., 2006; Tao et al., 2012). Indeed, in valves diseases the leaflets loose this stratification, leading to improper valve closure, backward blood flow and cardiac insufficiencies (Hinton et al., 2006). Congenital valve diseases are observed in 5% of live human births, accounting for 25-30% of all cardiovascular malformations and are a major cause of mortality in adults (Armstrong and Bischoff, 2004; Combs and Yutzey, 2009). The most common valve diseases (2-4% of the human population) include bicuspid aortic valve (only two leaflets), mitral valve prolapse (increased deposition of extracellular matrix) and valve calcification (osteogenic process leading to extracellular matrix mineralization) (Rajamannan et al., 2003; Tao et al., 2012).
4.2.2. Overview of the main steps of cardiac development in the mouse

The heart of the mouse embryo can first be observed in a crescent-shape structure at embryonic day 7.5 (E7.5) (Fig.4A). The cardiac crescent is composed of myocardial progenitors from both the primary heart field and the second heart field lineages, which will contribute to different structures in the adult heart. These progenitors fuse across the embryo midline, forming a linear tube that starts beating at around E8.0. This heart tube is mainly composed of cells from the primary heart field and serves as a scaffold for second heart field cells that migrate from adjacent splanchnic pharyngeal mesoderm to both the venous and arterial poles of the tube (Fig.4A). The venous pole, located caudally, is the entry point of the blood in the heart tube (inflow tract, prospective atrial chambers). The arterial pole, located cranially, is where the blood exits the heart tube (prospective outflow tract). The outflow tract is a transient structure that connects the right ventricle to the aortic sac, from which the aortic arches arteries arise in the branchial (pharyngeal) arches. At E8.5, the heart undergoes a rightward looping that breaks its bilateral symmetry and progressively moves the venous pole cranially. Myocardial cells rapidly proliferate and expand to shape the future cardiac chambers (Fig.4A). The primary heart field (PHF), marked by the expression of Nkx2.5, contributes mainly to the prospective left ventricle and to both atria, and partially to the prospective right ventricle. In contrast, the cells of the second heart field (SHF), marked by the expression of Isl1, contribute mainly to the prospective right ventricle, outflow tract and prospective atria.

From E9.0-E9.5, cardiac neural crest cells delaminate from the dorsal ectoderm of the neural tube and migrate through the branchial arches to the outflow tract (Fig.4B) (Jiang et al., 2000). The developing heart is in the meantime invaded by another population of extracardiac cells: from E8.5-E9.0, cells from the transient proepicardial organ located at the venous pole detach and migrate to form both the epicardium that envelop the heart and the coronary vasculature that will provide blood supply to the heart itself (Fig.4B) (Reese et al., 2002; Männer et al., 2001).
Figure 4. Main steps of mouse heart development

(A) Cardiac development in the mouse. See details in main text. From left to right: cardiac crescent stage (E7.5), heart tube stage (E8.0), loping heart (E8.5), formation of the chambers (E10.5) and functional four chambers heart (E15.0). Red indicates the primary heart field (PHF) and blue the second heart field (SHF) and their derivative structures. Adapted from (Xin et al., 2013). (B) Contribution of extracardiac cell populations to the developing heart at ~E10.5. In purple, cardiac neural crest cells (cNCC) migrate from the neural tube through the pharyngeal arches (PhA) and invade the outflow tract. In yellow, cells from the proepicardial organ migrate to cover the heart and form the epicardium and coronary vasculature. Adapted from (Laugwitz et al., 2008). (C) Formation of the compact zone and trabeculations in the ventricular myocardium at ~E.10.5. OFT: outflow tract. RV: right ventricle. LV: left ventricle. AVC: atrioventricular canal. RA: right atrium. LV: left atrium. Cell types in the bottom panel are indicated. Adapted from (Misra and Garg, 2013).

Between E10.5 and E12, the proliferation of the cardiomyocytes in the myocardium of the ventricular chambers leads to the formation of a multilayered compact zone that increases the thickness of the ventricular wall to ensure powerful contractions. Some of these cells migrate towards the interior of the ventricles and form finger-like protrusions called trabeculations, which increases the surface of the myocardium to
ensure sufficient oxygenation before the establishment of a functional coronary circulation within the myocardium (Fig. 4C) (Martin-Puig et al., 2008). The conduction system and the cardiac valves are formed during this period of chambers formation (see below) (Christoffels et al., 2010; Combs and Yutzey, 2009). The cardiac valves develop both in the outflow tract and in the atrioventricular canal, which connects the left ventricle to the common atria before they are separated (Fig. 4C). Together with the formation of these valves, the formation of the interventricular, interatrial and atrioventricular septa divide the heart into four functional chambers (Lin et al., 2012). Along with chambers septation, the cardiac neural crest cells contribute to divide the outflow tract into the pulmonary trunk and the aorta that are finally connected to the right and left ventricles, respectively.

4.2.3. From cardiogenic mesoderm to the myocardium and endocardium: determination of the cardiac lineages

The different structures composing the mammalian embryonic heart are derived from progenitor cells of three developmentally distinct origins: the cardiogenic mesoderm, the cardiac neural crest cells (cNCC) and the proepicardial organ (PEO) (Laugwitz et al., 2008). These multipotent progenitors differentiate into several cardiac cell types through a complex lineage diversification involving multiple intermediate progenitor states (Martin-Puig et al., 2008; Vincent and Buckingham, 2010; Lescroart and Meilhac, 2012).

The cardiogenic mesoderm contains the progenitors that will contribute to the earliest cardiac structures derived from both the PHF and SHF (Fig. 5). It is marked by the transient expression of Mesp1, and can first be detected in the anterior part of the primitive streak at the onset of gastrulation (E6.5) (Bondue and Blanpain, 2010). Following gastrulation, these progenitors migrate away from the primitive streak (PS) to anterior-lateral positions under the head folds, forming two groups of expanding lateral plate mesoderm cells on both sides of the midline (Fig. 5). At E7.0-7.5, these heart-forming regions extend across the embryo midline and fuse to form the cardiac crescent. Paracrine signals from adjacent tissues contribute to activate the expression of cardiac transcription factors in Mesp1+ cells of the cardiac crescent (Rana et al., 2013; Harvey, 2002; Zaffran and Frasch, 2002).
Figure 5. Lineage tree of the myocardial and endocardial progenitors

Representation of the main cellular hierarchies and lineages relationships of the cardiac cells with associated genetic marker(s). *Brachyury* (*Bry*) precursors give rise both to the cardiogenic mesoderm expressing *Mesp1* and the hemangioblast cells that are the progenitors of the hematopoietic and endothelial cell lineages. The primary and second heart fields derive from *Isl1*/*Nkx2.5*/*Flk1* multipotent cardiac progenitors (MCP). Following successive transient progenitor states, they give rise to the cardiomyocytes of the atria and ventricles, the cells of the conduction system and smooth muscle cells. The endocardial cells derive from both *Isl1*+ SHF progenitors and vascular progenitors unrelated with the myocardium. The post-gastrulation stages at which the cardiogenic mesoderm and subsequent cardiac crescent can be observed are represented under the corresponding cells they contain. PS: primitive streak. HF: head folds. ML: midline (adapted from (Buckingham et al., 2005)). This scheme is not exhaustive and is adapted from (Laugwitz et al., 2008), with additional data based mainly on (Martin-Puig et al., 2008; Misfeldt et al., 2009; Milgrom-Hoffman et al., 2011).

The multipotent cardiac progenitor cells (MCP) that further express *Isl1*, *Nkx2.5* and *Flk1* are able to differentiate into all major cell types of the heart: cardiomyocytes, smooth muscle cells, endothelial cells and cells of the conduction system (Fig.5) (Moretti et al., 2006; Wu et al., 2006; Sun et al., 2007; Martin-Puig et al., 2008). The two myocardial lineages are determined at the cardiac crescent stage from these MCP cells. Due to their common origin and close proximity in the cardiac crescent,
the distinction between the progenitors of the two lineages is mainly due to the timing of their differentiation (Meilhac et al., 2004; Meilhac et al., 2003; Dyer and Kirby, 2009). The PHF cells, embedded in the lateral plate mesoderm, are the first ones to differentiate and will form the early heart tube and derived left ventricle (LV). The SHF cells, coming from splanchnic pharyngeal mesoderm in the dorsal wall of the pericardial cavity, differentiate later and are progressively added to the elongating heart tube from both poles to contribute to the outflow tract (OFT), right ventricle (RV) and atria (Cai et al., 2003; Zaffran et al., 2004; Meilhac et al., 2004) (Fig.4A). PHF cells are marked mainly by the expression of Nkx2.5, but also of Tbx5 and Hand1. In contrast, Isl1 expression marks the cells of the SHF, which also express Gata4, Foxh1, Tbx1, Fgf8, Fgf10, Mef2c, Hand2 and Nkx2.5 (Buckingham et al., 2005; Laugwitz et al., 2008).

The developmental origin of the endocardium has been controversial regarding the stage at which it segregates from the progenitors of the myocardial lineages (Harris and Black, 2010; Lescroart and Meilhac, 2012; Puceat, 2013). Two distinct sources of endocardial progenitors have been identified (Fig.5). The first population derives from the Isl1+ SHF and shares a common precursor with the myocardium (Ferdous et al., 2009; Misfeldt et al., 2009). The second population is composed of Flk1+ vascular endothelial cells unrelated with the myocardium. This endocardial population migrates from a medial position to the cardiac crescent and enters the heart from the arterial pole (Milgrom-Hoffman et al., 2011). The endocardial cells in the outflow tract and in the atroioventricular canal give rise to the cardiac cushions that are the precursors of the cardiac valves (see below) (Puceat, 2013; von Gise and Pu, 2012). Furthermore, ventricular endocardial cells invade the myocardium in the compact zone and form the endothelium of the coronary arteries between E11.5 and E13.5 (Wu et al., 2012).

4.2.4. Patterning and differentiation of the second heart field

The Isl1+ second heart field is located in the dorsal pericardial wall and is patterned along the embryo antero-posterior (cranio-caudal) axis, with cells coming from distinct but clonally related regions taking different migratory routes for different developmental outcomes (Kelly, 2012; Lescroart and Meilhac, 2012; Lescroart et al.,
This patterning requires retinoic acid upstream of a graded expression of Hox genes, but the boundary between the anterior and the posterior SHF is not clearly defined (Ryckebusch et al., 2008; Bertrand et al., 2011; Kelly, 2012).

The cells in the anterior SHF are closely linked to the mesodermal core of the branchial arches (Tzahor and Evans, 2011). They are mainly characterized by the expression of Fgf10, Fgf8 and Tbx1, and migrate through the arterial pole to contribute to the OFT and RV myocardium (Fig.6A) (Kelly et al., 2001; Mjaatvedt et al., 2001; Xu et al., 2004; Lescroart et al., 2010). The anterior SHF cells are clonally related to the skeletal head muscles, with lateral regionalization that links the progenitors of the aorta and pulmonary trunk to the progenitors of the right and left facial expression muscles, respectively (Lescroart et al., 2010). The anterior SHF should not be mixed up with the “secondary” heart field, which was identified in the chicken and originates from a subdomain of the anterior SHF that gives rise only to the OFT distal myocardium (Waldo et al., 2001; Dyer and Kirby, 2009). In contrast to the anterior SHF, the posterior SHF contributes to the venous pole of the heart with clonal relationship between the cells of the superior vena cava, pulmonary vein and of both atria (Lescroart et al., 2012). Furthermore, the atria are clonally related to the AVC myocardium, and their left/right identity is further determined by asymmetric expression of Pitx2 (Domínguez et al., 2012; Galli et al., 2008).

Two major properties discriminate the SHF from the PHF cells: their sustained proliferation and their differentiation delay, controlled by paracrine Sonic Hedgehog (SHH), Wnt, Fibroblast Growth Factors (FGFs), Bone Morphogenetic Proteins (BMPs) and Notch signaling from the surrounding pharyngeal ectoderm and endoderm and by autocrine signaling from the pharyngeal mesoderm itself (Fig.6B) (Rochais et al., 2009b; Vincent and Buckingham, 2010; Francou et al., 2013). Indeed, SHF cells remain in an undifferentiated progenitor state in the pharyngeal mesoderm at the time the PHF cells form the heart tube and differentiate into cardiomyocytes. It is only after their migration and addition to the heart tube that the exposure to new signaling factors differentiate SHF cells into cardiomyocytes or smooth muscle cells, associated with downregulation of their early markers Fgf8, Fgf10, Tbx1 and Isl1 and...
activation of the expression of Nkx2.5, Gata4, Mef2c and Hand2 (Dyer and Kirby, 2009; Kelly, 2012).

Figure 6. Migration and differentiation of second heart field progenitors
(A) Migration of anterior second heart field progenitors from the pharyngeal mesoderm of the branchial arches (BA) towards the heart tube. Blue: sub-lineage originating in the 1st BA and contributing to the RV myocardium and to the masticatory muscles. Pink: sub-lineage originating in the 2nd BA and giving rise to the OFT myocardium and to the facial expression muscles. From (Lescroart et al., 2010). (B) Signaling interactions controlling the balance between FGF-induced proliferation and BMP-induced differentiation of second heart field progenitors towards cardiomyocytes. See details in main text. From (Francou et al., 2013).

The balance between proliferation and differentiation of SHF cells depends mainly on opposite effects of FGF and BMP signaling (Fig.6B) (Hutson et al., 2010). Fgf10 and Fgf8 functionally overlap and control the survival and progenitor state maintenance of the SHF cells that will contribute to the OFT and RV (Ilagan et al., 2006; Park et al., 2006; Watanabe et al., 2010). The expression of Fgf10 in these progenitors depends on a CRM located in its proximal promoter, which is mainly activated by TBX1 and ISL1 and repressed by NKX2.5 upon myocardium differentiation (Watanabe et al., 2012). In addition, FGF signaling is also activated by Wnt/β-catenin in the anterior SHF (Cohen et al., 2007). In contrast, the non-canonical Wnt11 later promotes differentiation of SHF cells (Cohen et al., 2012; Pandur et al., 2002). BMP signaling from the distal prospective OFT antagonizes the proliferative FGF signals through the activation of Msx1 and Msx2 in the invading cNCC, but the mechanism by which these BMP targets further downregulate FGF ligands and decreases the proliferation of SHF progenitors is unclear (Tirosh-Finkel et al., 2010; Hutson et al., 2006). BMP4
is required for the septation of the OFT and the formation of the cardiac cushions, and its expression in the pharyngeal mesoderm and in the OFT is activated by Wnt signaling (McCulley et al., 2008; Ai et al., 2007). Furthermore, autocrine pharyngeal mesodermal FGF8 signaling is also required for Bmp4 expression (Park et al., 2008). Finally, both Tbx2 and Tbx3 are necessary to activate BMP signaling in the distal OFT (Mesbah et al., 2012). These regulatory crosstalks between proliferative FGF and differentiating BMP signals suggest that their expression levels are critical to reach a signaling threshold necessary to affect cell fate (Vincent and Buckingham, 2010). The differentiation of SHF-derived structures (OFT, RV) depends further on a network of various transcription factors, among which Hand2 plays an essential role (see below).

4.2.5. The T-box genes family and the development of the atrioventricular canal and atrioventricular node

The first signs of an electrical activity in the developing heart can be observed at the heart tube stage, with pacemaker activity at the venous pole generating slow peristaltic contractions (Christoffels and Moorman, 2009; Christoffels et al., 2010). This pacemaker activity, both at early and late stage of cardiac development, depends on the expression of the potassium/sodium channel Hcn4 (Stieber et al., 2003). When cardiac chambers start to differentiate, the transmission of the electrical impulse depends on rapid conduction through Cx40/Cx43 gap-junction channels in atria and ventricles myocardium and on slow conduction through Cx45/Cx30.2 gap-junctions channels in the AVC and OFT, providing a contraction delay in these two compartments (de Jong et al., 1992; Moorman and Christoffels, 2003; Christoffels and Moorman, 2009). The primordia that form later the definitive components of the conduction system start to differentiate around E9.0-E9.5, and can originate both from transdifferentiation of cardiomyocytes or from pre-specified cardiac progenitors (Munshi, 2012; Christoffels and Moorman, 2009).

Several members of the T-box family of transcriptional regulators play an essential role during cardiac development and in the formation of the conduction system in particular, functions that are conserved among vertebrates (Stennard and Harvey,
This family is characterized by a highly conserved “T-box” DNA binding domain and consists of at least 17 members in mammals. Based on sequence homology, two main subfamilies can be identified: the first one contains *Tbx1*, *Tbx18* and *Tbx20*, while the second contains *Tbx2*, *Tbx3*, *Tbx4* and *Tbx5* (Plageman and Yutzey, 2005). The members of this second subfamily are closely related, as they evolved by two rounds of gene duplications from a single ancestral gene that resulted in two separate paralogous gene clusters: the *Tbx3/Tbx5* and *Tbx2/Tbx4* landscapes (Agulnik et al., 1996). *Tbx1* is involved in the proliferation of anterior SHF progenitors (Xu et al., 2004; Watanabe et al., 2012), *Tbx5* is a marker of the PHF involved in the differentiation of cardiomyocytes (Bruneau et al., 1999; Bruneau et al., 2001), *Tbx18* is important for the formation of the epicardium (Wu et al., 2013; Cai et al., 2008), and *Tbx4* is not involved in cardiogenesis. The main T-box genes involved in the formation of the conduction system are the transcriptional repressors *Tbx2* and *Tbx3*. *Tbx2* has been implicated mainly in the formation of the atrioventricular node (AVN) (Aanhaanen et al., 2011; Aanhaanen et al., 2009), while *Tbx3* marks the entire cardiac conduction system with highest expression in the developing and mature sinoatrial node (SAN), AVN and bundle of His (Hoogaars et al., 2004; Hoogaars et al., 2007b; Bakker et al., 2008; Horsthuis et al., 2009). *Tbx2* and *Tbx3* are expressed in the myocardium of the AVC and OFT, but not in the atrial and ventricular chambers. Their patterning and morphogenesis effect on these compartments occurs by repression of the chamber myocardial differentiation gene program, mainly through the direct repression of the differentiation marker *Nppa* (ANF, Atrial Natriuretic Factor) (Habets et al., 2002; Christoffels et al., 2004; Harrelson et al., 2004; Dupays et al., 2009; Mesbah et al., 2008). The AVC and OFT thus retain a low proliferation rate and slow conduction mode in comparison to the differentiating chambers. This progressively replaces the peristaltic contractions of the heart tube by a pattern of serial contractions combining alternate slow conducting and poorly contracted myocardium in AVC and OFT with fast conducting and contracting myocardium in the chambers (Moorman and Christoffels, 2003; Christoffels et al., 2010).
The atrioventricular node (AVN) starts to differentiate at E9.5 from the myocardium of the AVC, which originates from both the PHF and posterior SHF (Aanhaanen et al., 2009; Domínguez et al., 2012). The AVC is specified by the expression of Bmp2 in the myocardium, which mediates the induction of Tbx2 expression to repress the chamber differentiation genes Cx40, Cx43 and Nppa (Fig.7) (Yamada et al., 2000; Ma et al., 2005; Harrelson et al., 2004; Christoffels et al., 2004; Habets et al., 2002). In the cardiomyocytes of the neighboring chambers, Notch signaling activates Hey1 and Hey2, which suppress Tbx2 expression thereby restricting it to the AVC myocardium and creating the boundaries between the AVC and the heart chambers (Rutenberg et al., 2006; Kokubo et al., 2007). Furthermore, Tbx2 expression in chambers myocardium is also repressed by TBX20 (Singh et al., 2005; Cai et al., 2005; Stennard et al., 2005). This repression is not direct but is mediated by interfering with BMP2 signal transduction: TBX20 binds to SMAD1/5 and prevents their interaction with SMAD4, which in turn prevent them to bind to Tbx2 proximal promoter (Fig.7 and Fig.10B) (Singh et al., 2009; Singh and Kispert, 2010). Tbx3 is also expressed in the myocardium of the AVC, where it functions together with Tbx2 to inhibit precocious chamber differentiation (Hoogaars et al., 2004; Singh et al., 2012). These two genes also have non-redundant functions, Tbx2 being more important in the regulation of AVC patterning and Tbx3 in the formation of the AVN itself by controlling the expression of the pacemaker gene program (Munshi, 2012; Frank et al., 2012; Aanhaanen et al., 2011; Horsthuis et al., 2009). By E13.5, the Tbx3/Hcn4/Cx30.2/Cx45 positive cells of the AVN will extend into the left ventricle.
and connect with both the fast conducting Cx40/Cx43 positive trabeculations and the interventricular septum to form the bundle of His (Christoffels et al., 2010).

4.2.6. Formation of the cardiac valves: from endothelial-mesenchymal transition to the cardiac cushions

Epithelial-mesenchymal transition (EMT) is a common developmental mechanism by which epithelial cells are converted into mesenchymal cells by the loss of cell-cell adhesion and cellular polarity, and the gain of motile and invasive characteristics (Fig.8A). For instance, EMT is crucial for the formation of the mesoderm during gastrulation, the delamination of the cNCC from the neural tube, and the formation of the epicardium (Sauka-Spengler and Bronner-Fraser, 2008; Männner et al., 2001; Olivey and Svensson, 2010; von Gise and Pu, 2012). Successive waves of EMT and MET (mesenchymal-epithelial transition) are typically required for the formation of organs (Thiery et al., 2009; Lim and Thiery, 2012; Kovacic et al., 2012).

The formation of the cardiac valves is initiated in the endocardium at E9.5 in the AVC and E10.5 in the OFT by endothelial-mesenchymal transition (EndMT), a particular type of EMT. In response to signals coming from the myocardium, endocardial cells delaminate, transform into mesenchymal cells (Fig.8A,B) and migrate into the cardiac jelly between the two cellular layers, where they proliferate to form the cardiac cushions (Fig.8B). Following proliferative expansion and increased production of ECM, the cardiac cushions are remodeled into cardiac valves primordia (von Gise and Pu, 2012). The formation of the semilunar valves in the OFT also involves the cNCC. After their migration through the branchial arches, these cells invade the mesenchyme of the OFT cardiac cushions and contribute to their fusion to form the aorticopulmonary septum that divides the OFT into aorta and pulmonary trunk, and the associated semilunar valves (Lin et al., 2012). Alterations in the formation of the cardiac cushions or in their fusion with the cardiac septa can lead to atrial-septal and atrioventricular-septal defects, two frequent types of valves-associated CHDs.

In the AVC, the EndMT is initiated by a crosstalk between myocardium and endocardium involving multiple signaling pathways (Fig.8C). Myocardial Bmp2 is crucial for both myocardium patterning (see above) and endocardial EndMT (Ma et
al., 2005; Rivera-Feliciano and Tabin, 2006). BMP2 paracrine signaling induces the expression of Msx1 in the endocardium, which functions to maintain Notch1 expression (Ma et al., 2005; Chen et al., 2008). Then, Notch1 endocardial signaling promotes the EndMT through the indirect non-cell autonomous activation of myocardial Tgfβ2 (Timmerman et al., 2004). Furthermore, BMP2, Notch1 and TGFβ2 all converge to synergically activate Snai1 expression in the endocardium (Timmerman et al., 2004; Niessen et al., 2008; Luna-Zurita et al., 2010). In addition, endocardial Notch1 signaling upregulates myocardial Bmp2 expression via paracrine Wnt4, and BMP2 signaling also induces Tgfβ2 expression in the myocardium via Tbx2 (Wang et al., 2013; Ma et al., 2005; Shirai et al., 2009). This complex integration of BMP2/Notch1/TGFβ2 signaling between myocardium and endocardium drives and restricts the EndMT to the AVC (Garside et al., 2013).

**Figure 8. Formation of the cardiac valves by endothelial-mesenchymal transition**

(A) Mechanism of epithelial-mesenchymal transition (EMT). During EMT, a cell in an epithelial sheet looses its adherent junctions and polarity while gaining migratory and invasive phenotypes. It delaminates or ingress depending on the tissue and becomes a mesenchymal cells. From (von Gise and Pu, 2012). (B) Main steps of endothelial-mesenchymal transition (EndMT) in atrioventricular (left panel) and semilunar (right panel) valves. The atrioventricular valves leaflets are attached to the papillary muscles in the ventricles by chordae tendinae (CT). Adapted from (Lin et al., 2012). (C) AVC gene regulatory network controlling EndMT. See details in main text. Adapted from (von Gise and Pu, 2012).
Furthermore, myocardial TBX20 is also important to pattern the AVC upstream of the EndMT by regulating Bmp2 expression (Cai et al., 2011). In contrast, endocardial TBX20 is not required to initiate the EndMT, but is essential later for cardiac cushion maturation and valves elongation by altering Wnt/β-catenin signaling (Cai et al., 2013). Finally, signaling through the endocardial ErbB3 tyrosine kinase receptor is also necessary to induce EndMT (Camenisch et al., 2002), and requires upstream activation of ErbB3 by GATA4 (Rivera-Feliciano et al., 2006).

Snai1 and the related gene Snai2 are key mediators of EMT in a variety of developmental processes and disease states, such as the initiation of cancer metastasis (Barrallo-Gimeno and Nieto, 2005; Carver et al., 2001). During normal cardiac cushion development, endocardial Notch signaling induces the expression of Snai1, which in turn directly downregulates the expression of the adherent junction marker VE-Cadherin (VE-Cad) and hence disrupts cell contacts in the endocardium (Timmerman et al., 2004; Luna-Zurita et al., 2010). In addition, Snai1 upregulates the expression of the intermediate filament Vimentin (VIM) to reorganize the cytoskeleton and gain cellular motility (Cano et al., 2000), and activates the expression of matrix metalloproteinases such as MMP15 that degrade the basal surface of endocardial cells to facilitate their delamination and migration into the cardiac jelly (Tao et al., 2011; Song et al., 2000).

One of the main components of the cardiac jelly ECM is hyaluronic acid (HA). HA interacts with other ECM components such as the proteoglycans aggrecan and versican, thereby contributing to the architecture of the ECM in the cardiac jelly to provide a support for the migration of the cardiac cushion progenitors (Bernanke and Markwald, 1979). HA also participates in ErbB3 tyrosine kinase receptor-mediated signaling transduction, which activates Ras that is required for AVC morphogenesis (Camenisch et al., 2000; Camenisch et al., 2002; Lakkis and Epstein, 1998). HA is mainly produced by the enzyme Hyaluronan Synthase 2 (HAS2). Has2 expression in the myocardium depends of Bmp2 (Ma et al., 2005), and it is also expressed by the endocardium and the migrating cardiac cushion mesenchymal progenitors (Fig.8C).
Has2-deficient embryos die by E10.0 due to hypoplastic OFT and RV and lack of ventricular trabeculation and cardiac cushions (Camenisch et al., 2000).

In the AVC, the transcriptional regulator Twist1 controls the migration, proliferation and differentiation of cardiac cushion cells and the subsequent remodeling of the cardiac valves (Shelton and Yutzey, 2008; Lee and Yutzey, 2011; Vrijicak et al., 2012; Chakraborty et al., 2010). One target of Twist1 in the cardiac cushions is Sox9 (Vrijicak et al., 2012). Sox9 is not required to initiate the EndMT and migration, but is activated following delamination and regulates cardiac jelly invasion and proliferation of cardiac cushion progenitors (Akiyama et al., 2004). Sox9 is also required at later stages for proper distribution of ECM molecules during valves remodeling (Lincoln et al., 2007), and its downregulation leads to valve calcification phenotypes (Peacock et al., 2010).

Calcineurin/NFATc signaling is also critical for cardiac cushion formation, together with VEGF. Nfatc1 is expressed specifically in endothelial cells of the endocardium but not the remainder of the vasculature, and is required for the formation of the semilunar valves but not for EndMT (Ranger et al., 1998; de la Pompa et al., 1998). At E9.0, Calcineurin/NFATc signaling in the AVC myocardium reduces the production of VEGF, as low levels of VEGF are required for EndMT. Accumulated levels of VEGF ultimately terminate EndMT at E10.5, and a second wave of Calcineurin/NFATc signaling in the endocardium directs valves elongation and remodeling from E11.0 (Chang et al., 2004; Lambrechts and Carmeliet, 2004). During valves remodeling, Nfatc1 prevents a fraction of endocardial cell to undergo EndMT, which maintains them in a proliferative state (Wu et al., 2011).

4.2.7. *Hand2* is a crucial regulator for the development of second heart field derived structures

*Hand2* and its closely related gene *Hand1* encode basic helix-loop-helix (bHLH) proteins belonging to the Twist-family of transcriptional regulators. These proteins are composed of a basic domain followed by two amphiphatic α-helices that are separated by a loop of variable length. While the helix-loop-helix motif (HLH)
mediates protein-protein interactions, the basic domain is responsible for DNA interactions through canonical E-box motifs (CANNTG). In vitro, HAND2 interacts with the highest affinity with the CATCTG E-box motif (Dai and Cserjesi, 2002). The dimerization of two bHLH proteins via their HLH domain is crucial for E-box recognition by two juxtaposed basic domains, and the choice of binding partner directly dictates the function and specificity of the bHLH transcriptional complex (Massari and Murre, 2000; Conway et al., 2010; Vincentz et al., 2011). The phosphorylation status of HAND2 and others Twist-family members at conserved Threonine and Serine residues in helix 1 alters their dimerization affinity for others bHLH proteins and consequently affects their DNA binding specificity (Firulli et al., 2005; Firulli et al., 2007; Barnes and Firulli, 2009).

Hand2 (or dHand) is expressed in the deciduum, heart, autonomic nervous system and neural crest derivatives and is crucial for the morphogenesis of these embryonic structures. In the sympathetic nervous system, Hand2 regulates the expression of the norepinephrine biosynthetic enzymes that impact on the cardiac rhythm (Morikawa et al., 2007; Hendershot et al., 2008; Morikawa and Cserjesi, 2008; VanDusen et al., 2014). In addition, Hand2 also regulates the formation of craniofacial structures derived from neural crest cells (Charité et al., 2001; Yanagisawa et al., 2003; Funato et al., 2009; Barron et al., 2011; Xiong et al., 2009). Hand2 is also essential to activate Shh expression in the posterior forelimb and hindlimb bud mesenchyme, necessary for antero-posterior patterning of the limb (Charité et al., 2000; te Welscher et al., 2002; Galli et al., 2010).

The Hand2 genomic sequence, expression pattern and functions are highly conserved among vertebrate species (Angelo et al., 2000; Srivastava, 1995; Yelon et al., 2000). During mouse cardiac development, Hand2 expression is first detected in SHF progenitors at the cardiac crescent stage (E7.5). It is further maintained in the whole heart tube before being progressively restricted to the RV and OFT myocardium during cardiac looping, with remaining expression in the LV to a lesser extent (E8.5). From the heart tube stage, Hand2 expression can also be detected in the endocardium and in the splanchnic pharyngeal mesoderm in the dorsal pericardial wall. From E9.5, Hand2 is also expressed in the proepicardial organ and
derived epicardium, in the branchial arches and in the migrating cNCC that invade them, and in the developing vasculature. Its expression is gradually restricted during cardiac chambers remodeling to be detected only in the atria by birth (Fig.9A) (Srivastava, 1995; Srivastava et al., 1997; Biben and Harvey, 1997; Barnes et al., 2011; Dirkx et al., 2013). In contrast, Hand1 is expressed mainly in PHF derived structures and becomes restricted to the LV when cardiac looping occurs (Fig.9A) (Srivastava et al., 1997; Biben and Harvey, 1997). HAND2 and HAND1 can form heterodimers, and they function redundantly and in a gene dosage-dependent manner for the differentiation of the PHF-derived LV, while Hand1 does not play a role in the formation of the SHF-derived RV (Firulli et al., 2000; Firulli et al., 2010; McFadden et al., 2005).

Embryos deficient for Hand2 display severe cardiac defects leading to growth retardation and embryonic lethality by E10.5 (Srivastava et al., 1997). The heart of these embryos appears morphologically normal until cardiac looping, but then displays several malformations: an hypoplastic RV that results in an abrupt connection between the OFT and LV, a thin myocardium without trabeculation, a dilated aortic sac, an absence of the aortic arches arteries, and hypoplastic branchial arches (Fig.9B) (Srivastava et al., 1997; Togi et al., 2006). In these mutant embryos, Hand1 expression is not affected (Srivastava et al., 1997). The conditional inactivation of Hand2 in the epicardial lineage using both Hand1-Cre and Wt1-CreERT2 disrupts the secondary epicardial EMT and the formation of the coronary vasculature, showing that Hand2 is also required for the formation and differentiation of these cardiac structures (Barnes et al., 2011). Furthermore, Hand2 conditional deletion in SHF progenitors using Isl1-Cre leads to a hypoplastic RV and to increased cell death in SHF progenitors in the pharyngeal mesoderm, impairing their subsequent migration towards the OFT. This establishes that Hand2 is crucial for the survival of the SHF progenitors and for the subsequent formation of the RV (Tsuchihashi et al., 2011). Conditional inactivation of Hand2 in cNCC using Wnt1-Cre results in mispatterned aortic arches arteries, hypertrabeculation, enlarged RV, septation defects in both the ventricles and the OFT, and double-outlet right ventricle (DORV) in which both the aorta and the pulmonary trunk are connected to the right
ventricle (Holler et al., 2010; Morikawa and Cserjesi, 2008). Interestingly, cNCC-specific deletion of \textit{Hand2} also showed that \textit{Hand2} controls the expression of several genes involved in the proliferation and differentiation of both cNCC and SHF-derived cells in the RV trabecular myocardium (Holler et al., 2010; Morikawa and Cserjesi, 2008). Indeed, the marker of myocardial chamber differentiation \textit{Nppa} is downregulated in these conditional mutant embryos, which, together with their enlarged RV and hypertrabeculation, points to a defect in the control of SHF proliferation. This occurs through an interaction between cNCC and SHF progenitors that takes place before their migration into the PHF heart tube (Morikawa and Cserjesi, 2008; Houweling et al., 2005; Yelbuz et al., 2002).

\textbf{Figure 9.} \textit{Hand2} is a crucial regulator for second heart field differentiation

(A) Schematic representation of the expression of \textit{Hand1} (top panel, blue) and \textit{Hand2} (lower panel, red) during cardiac development. CC: cardiac crescent. V: ventricle. End: endocardium. OFT: outflow tract. LV: left ventricle. See details in main text. Adapted from (Vincentz et al., 2011). (B) Morphology of the developing heart in wild-type (top panels) and \textit{Hand2}-deficient embryos (lower panel). Differences can be observed from the stage of cardiac looping (E8.25). From E8.5, the absence of the right ventricle (rv, black arrowhead in \textit{Hand2}\textsuperscript{+/+}) is obvious in \textit{Hand2}\textsuperscript{−/−} embryos (white arrowhead). a: atria. R/L: right/left. D/V: dorsal/ventral. (C) Model of the transcriptional network active in the second heart field that controls the development of the right ventricle and outflow tract. Solid lines: direct in \textit{vivo} interactions. Dotted lines: genetic or \textit{in vitro} interactions. From (Laugwitz et al., 2008).
The proliferative role of *Hand2* was shown in zebrafish embryos as well, in which it can promote the proliferation of SHF-derived cardiomyocytes. *Hand2* is also re-expressed in adult zebrafish heart following cardiac injury, where it increases cardiomyocytes proliferation thereby enhancing their cardiac regenerative capacities (Schindler et al., 2014; Wills et al., 2008). Similarly, *Hand2* expression is reactivated in the postnatal mouse heart following cardiac stress, leading to cardiomyocytes hypertrophy, cardiac dilation and eventually heart failure (Dirkx et al., 2013). This re-expression is mediated by the transcriptional inhibition of the microRNA miR-125 by Calcineurin/NFATc, as otherwise mirR-125 inhibits *Hand2* translation by interacting with its 3’UTR (Dirkx et al., 2013). Negative regulation of *Hand2* translation is also mediated by interactions between its 3’UTR and the microRNAs miR-1 and miR-133a, which originate from the same bicistronic precursor (Zhao et al., 2005; Vo et al., 2010). Overexpression of miR-1 leads to defects in the proliferation of the cardiomyocytes, which also points to a role of *Hand2* controlling the proliferation of SHF myocardial cells (Zhao et al., 2005).

The RV-specific expression of *Hand2* is controlled by a GATA factors-dependent enhancer located in the *Hand2* proximal promoter (McFadden et al., 2000). In *Gata4*-deficient embryo, the precardiogenic mesoderm fails to fuse to form a heart tube (cardia bifida) (Kuo et al., 1997; Molkentin et al., 1997). In these mutant embryos, the RV-specific CRM of *Hand2* is still able to drive *LacZ* reporter expression in the two primitive hearts with highest intensity in the right one, suggesting that GATA5 or GATA6 could compensate for the loss of GATA4 to regulate *Hand2* expression in the RV (McFadden et al., 2000). Furthermore, GATA factors also regulate the expression of the PHF marker *Nkx2.5* in the RV through specific CRMs (Searcy et al., 1998; Lien et al., 1999; Reecy et al., 1999). Embryos deficient for *Nkx2.5* form a heart tube but fail to undergo cardiac looping. They show a severe reduction of the expression of cardiac differentiation markers, but not of *Hand2* (Lyons et al., 1995; Biben and Harvey, 1997; Yamagishi et al., 2001). Furthermore, the heart of embryos lacking both *Nkx2.5* and *Hand2* do not form any ventricles but only a single atrium, likely originating from the posterior SHF (Yamagishi et al., 2001; Lescroart et al., 2012; Srivastava, 2006).
The differentiation of the RV from anterior SHF progenitors is also impaired in embryos lacking *Isl1*, the early marker of this lineage (Cai et al., 2003). RV hypoplasia can also be observed in embryos deficient for *Mef2c*, which downregulates *Hand2* expression (Lin, 1997). *Mef2c* is a target of several key SHF transcription factors including ISL1, GATA4 and FOXH1 (Dodou et al., 2004; von Both et al., 2004). Furthermore, MEF2C directly activates the expression of the histone methyltransferase *Smyd1* (*Bop1*), which inactivation leads as well to RV hypoplasia and *Hand2* downregulation (Phan et al., 2005; Gottlieb et al., 2002). These and other genetic interactions determine a cascade of transcription factors crucial for the development of the right ventricle and the outflow tract (Fig.9C). *In vitro*, HAND2 directly activates the transcription of *Nppa* synergically with MEF2C, GATA4 or NKX2.5 (Zang et al., 2004; Dai et al., 2002; Thattaliyath et al., 2002), but no direct endogenous target gene of HAND2 in the second heart field or in other cardiac lineages have been identified so far (Vincentz et al., 2011).
4.3. The BMP signaling pathway during limb bud development and the Grem1 cis-regulatory landscape

The developing tetrapod limb bud has been extensively used as a model to study the interactions between signaling networks and their effects on gene regulation (Zeller et al., 2009). It also provides an excellent system to analyze cis-regulatory landscapes and the role of individual CRMs using transgenic approaches (Zeller and Zuniga, 2007; Spitz and Duboule, 2008).

The tetrapod limb buds originate, as the cardiovascular system, from the lateral plate mesoderm and emerge as pockets of mesenchymal cells surrounded by a monolayer of ectodermal cells. Mouse limb buds outgrowth and patterning along the proximo-distal and antero-posterior axis are controlled by two signaling centers: the zone of polarizing activity (ZPA) and the apical ectodermal ridge (AER), which interact in a self-regulatory feedback system involving SHH, BMPs, FGFs and the BMP antagonist GREMLIN1 (GREM1) (Fig.10A) (Panman et al., 2006; Benazet et al., 2009; Zeller et al., 2009).

The BMP signaling pathway occupies a central role during mouse limb bud development (reviewed in (Pignatti et al., 2014)). It is involved in limb bud initiation, outgrowth and patterning and is important for the chondrogenic differentiation of the limb skeletal elements (Kobayashi et al., 2005). BMP ligands belong to the TGFβ superfamily. In the developing limb bud, BMP2, BMP4 and BMP7 are expressed in specific spatio-temporal patterns, and their activity is modulated by various extracellular antagonists. The binding of BMP ligand dimers to serine/threonine receptors triggers the phosphorylation of receptor-associated SMADs (R-SMADs, or SMAD1,5,8), which mediate canonical BMP signal transduction by associating with SMAD4 (co-SMAD). The resulting transcriptional complex is translocated to the nucleus, where it interacts with BMP responsive elements (BRE) in the cis-regulatory landscapes of target genes (Id1, Msx2...) and activates their expression (Fig.10B).

SMAD4 is an essential mediator for both BMP and TGFβ signal transduction in the development of various tissues. It is composed of two evolutionary conserved MAD homology domains (MH) separated by a linker.
Figure 10. Limb bud development and BMP signaling pathway

(A). Interlinked signaling feedback loops control limb bud outgrowth. During the initiation phase, high BMP4 (blue) levels in the mesenchyme maintain AER-FGFs (green) and activate the expression of the BMP antagonist Grem1 (purple). During the propagation phase, increasing GREM1 levels antagonize
BMP4 signaling, which enables FGF signaling from the AER to the underlying mesenchyme for proliferative outgrowth and to the ZPA to maintain Shh expression (red). The SHH/GREM1/AER-FGFs epithelial-mesenchymal feedback loop is self-terminated as Grem1 is repressed by AER-FGFs and as the expanding population of Shh-expressing descendant cells becomes refractory to Grem1 expression, generating a gap that brings the Grem1 expression domain out of reach of long range SHH signaling. This results in an increase in BMP activity, initiating chondrogenic differentiation of mesenchymal progenitors. From (Zeller et al., 2009). (B) BMP signaling pathway. See main text for details. From (Pignatti et al., 2014). (C) The Grem1-Fmn1 cis-regulatory landscape. Point mutations, deletions (Δ), inversions and transgene insertions are indicated above or below the landscape. Mutations indicated in pale blue cause the limb deformity (ld) phenotype due to disruption of Grem1 cis-regulation. The three yellow boxes indicate the location of the conserved HMCOs elements. Adapted from (Zeller and Zuniga, 2007).

The N-terminal MH1 domain binds to DNA, mediates protein-protein interactions and contains the nuclear localization signal. The C-terminal MH2 domain is phosphorylated by type I BMP receptors and mediates heterodimerization with R-SMADs. Embryos deficient for Smad4 die before E7.5 due to gastrulation defects, which preclude the analysis of its specific role during limb bud development (Sirard et al., 1998).

Grem1 is a crucial node during limb bud morphogenesis, as its dynamic expression is controlled by BMP, SHH and FGF signaling pathways during limb bud initiation, expansion and termination, respectively (Fig.10A) (Zuniga et al., 1999; Capdevila et al., 1999; Scherz et al., 2004; Verheyden and Sun, 2008; Benazet et al., 2009). Grem1-deficient mice lack kidneys and ureters and die shortly after birth. They also display limb skeleton patterning defects such as fused ulna and radius and reduction in digit number (Michos et al., 2004; Khokha et al., 2003). Grem1 expression in limb buds depends on a large cis-regulatory landscape that extends at least in the neighboring Formin1 (Fmn1) gene. This region is able to activate reporter expression in the limb bud similarly to Grem1 endogenous expression, and contains three regions highly conserved between Human, Mouse, Chicken and Opossum (HMCOs), that can potentially act as CRMs (Fig.10C) (Zuniga et al., 2004).
5. Aims of the Thesis

During embryonic development, genomic cis-regulatory landscapes integrate inputs from multiple signaling pathways to control cell specification, determination and differentiation. Understanding the spatio-temporal and tissue-specific control of genes expression involves the identification of the cis-regulatory modules necessary for their activity and of the trans-acting factors required to activate or repress these genes.

In my PhD project I became interested in transcriptional gene regulation during mouse limb bud and heart morphogenesis. The first part of this thesis focuses on limb bud development and more specifically on components of the Bone Morphogenetic Protein (BMP) signaling pathway. The aims of this first part were: (1) to analyze the role of the BMP signal transducer SMAD4 during limb bud chondrogenesis using in vitro differentiation systems; (2) to study the role of a highly conserved region in the cis-regulatory landscape of Grem1 by analyzing its reporter activity in transgenic wild-type and mutant embryos (the publications that include the results from these two projects are directly presented in this thesis in their published version); (3) to globally dissect the transcriptional networks regulated by SMAD4 during chondrogenesis and to identify the cis-regulatory modules interacting with SMAD4 in the Grem1 genomic landscape. Due to the lack of suitable anti-SMAD4 antibodies for ChIP-Seq, I used cassette exchange and homologous recombination approaches in embryonic stem cells to generate a new mouse line that expresses an epitope-tagged SMAD4 protein from its endogenous locus.

In the second part of this thesis, my aim was to identify the direct transcriptional targets of the essential transcription factor HAND2 during heart development. Towards this aim I analyzed ChIP-Seq data previously generated in the laboratory using a mouse endogenously expressing an epitope-tagged HAND2 protein. By combining gene and protein expression analysis with cellular resolution, ChIP-qPCR and transgenic reporter studies, I was able to define the regulatory network controlled by HAND2 during second heart field development and in the formation of the cardiac cushions.
6. Results

6.1. From cis-regulatory modules to trans-acting factors: control of Grem1 genomic landscape during limb bud development and generation of Smad4\textsuperscript{3xFLAG} mice by homologous recombination

6.1.1. Paper 1: “Smad4 is required to induce digit ray primordia and to initiate the aggregation and differentiation of chondrogenic progenitors in mouse limb buds”

This publication focuses on the conditional deletion of Smad4 from the limb buds mesenchyme using Prx1-Cre mice. I carried out the analysis of the high-density in-vitro limb mesenchymal cell cultures isolated from wild-type and Smad4-deficient limb buds shown in Figure 6.
Smad4 is required to induce digit ray primordia and to initiate the aggregation and differentiation of chondrogenic progenitors in mouse limb buds

Jean-Denis Bénazet1,*,1, Emanuele Pignatti1,4, Ashleigh Nugent1, Erkan Unal1,2, Frédéric Laurent1 and Rolf Zeller1,5

SUMMARY

SMAD4 is an essential mediator of canonical TGFβ/BMP signal transduction and we inactivated Smad4 in mouse limb buds from early stages onward to study its functions in the mesenchyme. While this Smad4 inactivation did not alter the early Sox9 distribution, prefiguring the chondrogenic primordia of the stylopod and zeugopod, it disrupted formation of all Sox9-positive digit ray primordia. Specific inactivation of Smad4 during handplate development pointed to its differential requirement for posterior and anterior digit ray primordia. At the cellular level, Smad4 deficiency blocked the aggregation of Sox9-positive progenitors, thereby preventing chondrogenic differentiation as revealed by absence of collagen type II. The progressive loss of SOX9 due to disrupting digit ray primordia and chondrogenesis was paralleled by alterations in genes marking other lineages. This pointed to a general loss of tissue organization and diversion of mutant cells toward non-specific connective tissue. Conditional inactivation of Bmp2 and Bmp4 indicated that the loss of digit ray primordia and increase in connective tissue were predominantly a consequence of disrupting SMAD4-mediated BMP signal transduction. In summary, our analysis reveals that SMAD4 is required to initiate: (1) formation of the Sox9-positive digit ray primordia; and (2) aggregation and chondrogenic differentiation of all limb skeletal elements.

KEY WORDS: BMP signaling, Limb bud, Mouse

INTRODUCTION

Bone morphogenetic proteins (BMPs) are ligands belonging to the transforming growth factor beta (TGFβ) superfamily, which control inductive processes in early embryos and during organogenesis (reviewed by Zakin and De Robertis, 2010) and are required for chondrogenesis, growth and ossification of bones in vertebrates (reviewed by Wu et al., 2007). BMP ligands interact with two types of BMP receptors (BMPR1 and BMPR2), which in turn trigger intracellular signal transduction by phosphorylation of the receptor-associated SMAD1, -5 and -8 proteins. These in turn form a complex with SMAD4, an essential mediator of both canonical BMP and TGFβ signal transduction (Yang et al., 2002). These SMAD complexes translocate to the nucleus and regulate the expression of target genes in concert with other transcriptional regulators (Feng and Derynck, 2005). Genetic and functional analysis of mouse and chicken limb bud development has uncovered important morphoregulatory functions of the BMP pathway in the establishment and functioning of the two limb bud signaling centers, namely sonic hedgehog (SHH) signaling by the posterior mesenchymal organizer and fibroblast growth factor (FGF) signaling by the apical ectodermal ridge (AER).

During limb bud outgrowth, BMPs and the BMP antagonist gremlin 1 (Grem1) control the epithelial-mesenchymal interactions that coordinate outgrowth with patterning and determination of digit identities (reviewed by Zeller et al., 2009). During the onset of limb bud development, BMP4 is first required in the mesenchyme and BMPR1A in the ectoderm for formation of the AER, as inactivation of these molecules disrupts establishment of the AER-FGF signalling centre, which results in limb truncations and loss of dorsoventral polarity (Ahn et al., 2001; Pizette et al., 2001; Pajni-Underwood et al., 2007; Bénazet et al., 2009). During initiation of limb bud outgrowth, the SHH signalling centre is established and restricted to the posterior mesenchyme under the influence of BMP signalling (Bastida et al., 2004). Concurrently, BMP4 triggers expression of the BMP antagonist Grem1 in the posterior limb bud mesenchyme, which causes rapid lowering of mesenchymal BMP activity (Nissim et al., 2006; Bénazet et al., 2009). The subsequent distal-anterior expansion of the Grem1 expression domain depends largely on its positive regulation by AER-FGF and SHH signalling. This self-regulatory SHH/GREM1/AER-FGF feedback loop coordinates anteroposterior (AP) with proximodistal (PD) limb bud outgrowth and patterning (Zúñiga et al., 1999; Michos et al., 2004; Panman et al., 2006; Bénazet et al., 2009; Probst et al., 2011).

During limb bud outgrowth, low mesenchymal BMP activity is required to restrict the length of the AER, as inactivation of Bmp4 during this developmental period results in extended AER-FGF signalling and digit polydactylly (Selever et al., 2004; Bénazet et al., 2009). As inactivation of Bmpr1a in the mesenchyme also disrupts limb bud outgrowth (Ovchinnikov et al., 2006), tight control of BMP activity is required for normal progression of limb development (reviewed by Zeller et al., 2009).

These signalling interactions are terminated as the SHH/GREM1/AER-FGF feedback signaling system breaks down owing to progressive inhibition of Grem1 expression by high AER-FGF levels and the increasing separation of the Shh and Grem1 expression domains.
expression domains (Scherz et al., 2004; Verheyden and Sun, 2008). This termination results in a renewed increase in BMP activity during development of the digit ray primordia (giving rise to metacarpals and phalanges) (Bénazet et al., 2009), which probably serves two purposes: first, to control digit identity by signaling from the interdigital mesenchyme to the distal phalanx-forming region (PFR) (Dahn and Fallon, 2000; Suzuki et al., 2008; Witte et al., 2010), and second, to induce apoptosis of the interdigital mesenchyme (Bandyopadhyay et al., 2006; Pajni-Underwood et al., 2007; Maatouk et al., 2009; Wong et al., 2012). In agreement with these dynamic changes in BMP activity, BMPs inhibit chondrogenic differentiation of early limb bud mesenchymal progenitors while promoting chondrogenesis at late stages (Karamboulas et al., 2010). Noggin-mediated BMP antagonism under experimental conditions showed that BMPs are required to initiate compaction and chondrogenic differentiation of Sox9-positive mesenchymal progenitors (Brunet et al., 1998; Pizette and Niswander, 2000; Barna and Niswander, 2007). Genetic inactivation of both Bmp2 and Bmp4 in the limb bud mesenchyme disrupted formation of the two posterior-most digit primordia and the ulna (Bandyopadhyay et al., 2006). Experimental evidence indicated that BMPs participate in activating Sox9 expression in prechondrogenic progenitors, and Sox9 not only marks the chondrogenic lineage but is essential to initiate chondrogenesis during formation of skeletal elements (Bi et al., 1999; Akiyama et al., 2005; Pan et al., 2008).

Genetic analysis of the BMP pathway during chondrogenesis established that BMP signaling and antagonism are required recurrently for the proliferation of prechondrogenic progenitors, chondrocyte differentiation, shaping the skeletal primordia and formation of the endochondral growth plate (Brunet et al., 1998; Kobayashi et al., 2005; Yoon et al., 2005). In parallel, mesenchymal cells receiving different signals are specified as tendons, ligaments and connective tissues. For example, chondrogenic differentiation is inhibited and mesenchymal cells remain in a proliferative, undifferentiated state when they are exposed to both ectodermal WNT and AER-FGF signaling, whereas WNT signaling alone promotes the development of non-specific connective tissue (ten Berge et al., 2008; Gros et al., 2010).

TGFβ ligands are expressed from early limb bud stages onward, and inhibition of TGFβ signal transduction in cultured mesenchymal cells provided evidence that TGFβ signaling functions in early limb buds to alleviate the inhibitory effects of BMPs on initiation of chondrogenic differentiation (Karamboulas et al., 2010). Therefore, brief exposure of mesenchymal progenitors to TGFβ signaling seems necessary to allow subsequent induction of chondrogenesis by BMPs (Roark and Greer, 1994; Karamboulas et al., 2010). Furthermore, implantation of TGFβ1-loaded beads into the interdigit mesenchyme of chicken limb buds is able to induce formation of ectopic digit phalanges, i.e. trigger ectopic chondrogenesis (Merino et al., 1999; Lorda-Diez et al., 2011). Genetic analysis of TGFβ ligands and receptors in mouse embryos has revealed their essential roles during limb long bone, joint and tendon morphogenesis (Sanford et al., 1997; See and Serra, 2007). Although this genetic analysis revealed essential functions during limb skeletal development, it is possible that earlier functions in initiating chondrogenesis were masked by functional complementation among TGFβ and/or BMP ligands and receptors (see e.g. Karamboulas et al., 2010).

As Smad4-deficient (Smad4<sup>−/−</sup>) embryos die during gastrulation (Chu et al., 2004), we used a conditional loss-of-function allele (Yang et al., 2002) to study its requirement in the limb bud mesenchyme. Prxl-Cre-mediated conditional inactivation of Smad4 in limb bud mesenchyme resulted in the clearance of SMAD4 protein during early limb bud outgrowth, and the distribution of Sox9 transcript and SOX9 protein was normal until at least embryonic day (E) 11.0 in the proximal mesenchyme. By contrast, Sox9 expression was not activated in the distal Smad4-deficient limb bud mesenchyme that would normally give rise to the digit ray primordia. The analysis of SOX9 also indicated that with the exception of a small proximal aggregate, the aggregation and condensation of all mesenchymal progenitors to form the cartilage elements was disrupted in Smad4-deficient limb buds. Mutant cells failed to deposit extracellular collagen type II (COL type II) fibres, a hallmark of chondrogenesis, and progressively lost SOX9, i.e. chondrogenic lineage. In parallel, the expression of markers for other mesenchymal lineages was upregulated in Smad4-deficient limb buds without their normal spatial restriction, which provided evidence for a diversion of Smad4-deficient progenitors toward non-specific connective tissue.

MATERIALS AND METHODS

Ethics statement concerning animal experiments and mouse strains

Animal experiments were approved by the legally required regional commission in strict accordance with Swiss law. All studies were classified as grade zero, which implies minimal suffering. The 3R and Basel Declaration principles were implemented. The Prxl-Cre was used to conditionally inactivate Smad4 (Smad4<sup>fl/fl</sup>) (Yang et al., 2002) in the mesenchyme from early limb bud stages onward (Logan et al., 2002), and the Hoxa13<sup>Cre<sup>−/−</sup></sup> knock-in allele was used for autopod specific inactivation (Lopez-Rios et al., 2012; Scotti and Knitta, 2012). Prxl-Cre-mediated inactivation of Bmp2 and Bmp4 was done as described (Bandyopadhyay et al., 2006).

In situ hybridization and skeletal preparations

Whole-mount in situ hybridization was performed using standard protocols. Control and experimental embryos were age-matched by counting somites. To assess the clearance of Smad4 transcript in Smad4-deficient limb buds, a probe against the deleted region, coding exon 8 (Yang et al., 2002) was used for hybridization. Depending on the analysis, Smad4<sup>−/−</sup>, Smad4<sup>fl/fl</sup>, Smad4<sup>fl/fl</sup> or Smad4<sup>−/−</sup> Prxl-Cre<sup>−/−</sup> or Hoxa13<sup>Cre<sup>−/−</sup></sup> heterozygous embryos were used as controls and collectively referred to as ‘wild-type controls’. Alcian Blue and Alizarin Red were used to reveal cartilage and bone.

Optical projection tomography

Optical projection tomography (OPT) imaging (Sharpe et al., 2002) was used to acquire 3D images and prepare optical sections. RNA hybrids were detected using nitro-blue tetrazolium and 5-bromo-4-chloro-3-indolylphosphate (NBT/BCIP, Roche). Whole-mount immunofluorescence analysis was done as follows: embryos were permeabilized in cold acetone (15-30 minutes) and then incubated in blocking solution (PBT with 1% bovine serum albumin, 1% dimethyl sulfoxide and 5% inactivated bovine serum) for 20 minutes. Smad4 was detected using rabbit monoclonal anti-SMAD4 antibodies, 1:50 in blocking solution (Abcam; 12 hour incubation at 4°C). Samples were washed extensively and incubated with secondary antibodies Alexa Fluor 594-conjugated anti-rabbit antibody, 1:250 (Invitrogen; overnight at 4°C). Subsequently, samples were embedded in 1% low melting point agarose (Sigma), dehydrated in 100% methanol and cleared in BABB. Samples were scanned either at high (1024×1024 pixels) or intermediate resolution (512×512 pixels) in the Bioptons OPT scanner using SkyScan software (Bioptons, MRC Technology). The GP1 filter (425/40 nm, 475 nm LP) was used to detect sample anatomy. No filter (bright field) or the TXR filter (560/40 nm, 610 nm LP) was used to image NBT/BCIP signals or fluorescently labeled samples, respectively. OPT scans were reconstructed using NRecon software (SkyScan) and analysed using the Bioptons Viewer. Rendering images were taken using the maximum intensity projection function. For generation of iso-surfaces, the iso-surface editor of the Bioptons viewer program was used at 25%
iso-surface quality and 50% Gaussian smoothing parameters. Iso-surface values were determined and compared with the strength and distribution of the original signal to exclude aberrant iso-surfaces.

Quantification of transcript levels by quantitative real-time PCR
Forelimbs were isolated at E10.5 (35-39 somites) and E11.75 (51-53 somites) and analysed (Bénazet et al., 2009) using the primers listed in supplementary Table S1. At least eight samples per stage and genotype were analysed. All results shown are mean±SD; each dot represents one sample, and the significance of all differences was verified using the two-tailed non-parametric Mann-Whitney U-test.

Immunofluorescence analysis
Embryos or forelimb buds were isolated and fixed in 4% PFA overnight at 4°C. Samples were dehydrated stepwise into 100% ethanol, then transferred to xylene. Subsequently, samples were embedded in Paraplast wax (Sigma) and sectioned following standard procedures. For most preparations of mesenchymal cells that were plated in DMEM/F12 medium containing 0.3% Triton X-100 and 1% BSA (blocking solution; 1 hour at room temperature). Following pre-treatment, sections were incubated with antigen was retrieved by digesting sections with pepsin (1 mg/ml in 10 mM HCl, pH 2.0) for 15 minutes (37°C). Sections were blocked using PBS containing 0.1% DMSO. Media were refreshed every 24 hours. The results shown are representative of eight independent experiments (supplementary material Fig. S1A). Mesenchymal transcripts were undetectable by whole-mount antibody staining in wild-type (29 somites) and Smad4A/M forelimb buds at E9.75 (27 somites). Smad4 levels are also reduced along the primary axis, as only one Smad4 allele is functional in Smad4A/M embryos. The broken line indicates the approximate position of the virtual OPT section shown in the right-most panel. Smad4 is lost from the mutant mesenchyme, whereas AER expression remains (arrowheads). (E)Mx2, a direct transcriptional target of BMP signaling, is downregulated in Smad4A/M forelimb buds by E9.75 (26 somites). 1-5: digits 1-5; f, femur; h, humerus; p, pelvis; p*, rudimentary pelvis; r, radius; s, scapula; t, tibia; u, ulna.

Fig. 1. Prx1-Cre-mediated inactivation of Smad4 in mouse forelimb bud mesenchyme. (A) Skeletal preparations at embryonic day E14.5. Note the absence of all skeletal elements including the scapula in Smad4A/M forelimbs (FL) and formation of only a rudimentary girdle in mutant hindlimbs (HL). (B) Brightfield images of wild-type and Smad4A/M embryos at E14.5 to reveal the paddle-like limb morphology. (C) Smad4 transcript distribution (blue-purple staining) in wild-type, Smad4A/+ and Smad4A/M forelimb buds at E9.5 (23-24 somites). Smad4 transcript levels are reduced in Smad4A/+ forelimb buds and cleared in Smad4A/M forelimb buds. (D) Detection of SMAD4 protein (green fluorescence) by whole-mount antibody staining in wild-type (29 somites) and Smad4A/M forelimb buds at E9.75 (27 somites). Smad4 levels are also reduced along the primary axis, as only one Smad4 allele is functional in Smad4A/M embryos. The broken line indicates the approximate position of the virtual OPT section shown in the right-most panel. Smad4 is lost from the mutant mesenchyme, whereas AER expression remains (arrowheads). (E)Mx2, a direct transcriptional target of BMP signaling, is downregulated in Smad4A/M forelimb buds by E9.75 (26 somites). 1-5: digits 1-5; f, femur; h, humerus; p, pelvis; p*, rudimentary pelvis; r, radius; s, scapula; t, tibia; u, ulna.

Limb mesenchymal cell culture and immunofluorescence
Forelimbs were dissected in ice-cold PBS, and 2% trypsin digestion (25 minutes at 4°C) in combination with gently pipetting resulted in single cell preparations of mesenchymal cells that were plated in DMEM/F12 medium (supplemented with 0.5% penicillin/streptomycin and 10% FBS, Gibco-BRL) at 7.5×10³ cells/300 µl in eight-well chamber slides (Ibidi). After 48 hours’ culturing, samples were washed in PBS and fixed in 4% PFA (30 minutes at room temperature) before permeabilization and antigen detection as described above. BMP signaling was specifically inhibited by adding 10 µM dorsomorphin (Yu et al., 2008) and TGFβ signaling was inhibited by adding 10 µM SB431542 (Inman et al., 2002) in 0.1% DMSO (final concentration in medium). Control cells were cultured in medium containing 0.1% DMSO. Media were refreshed every 24 hours. The results shown are representative of eight independent experiments (supplementary material Fig. S5). Inhibition of p38, which is involved in non-canonical BMP signal transduction, had no effect on cell aggregation and the onset of chondrogenic differentiation (data not shown).

RESULTS
Requirement of mesenchymal Smad4 for propagation of the SHH/GREM1/AER-FGF feedback loop during mouse limb bud development
Prx1-Cre-mediated conditional inactivation of Smad4 in the mesenchyme (Smad4A/M) resulted in fore- and hindlimbs lacking skeletal elements (Fig. 1A). Stunted paddles replaced the autopod with digits at embryonic day E14.5 (Fig. 1B, arrowheads). As this initial analysis pointed to a late onset phenotype, the clearance of Smad4 gene products was assessed (Fig. 1C,D; supplementary material Fig. S1A). Mesenchymal transcripts were undetectable by ~E9.5 (Fig. 1C) and proteins from ~E9.75 onward in Smad4A/M forelimb buds (Fig. 1D). As the Prx1-Cre transgene is activated in
the forelimb bud territory before E9.0 (Logan et al., 2002), the complete functional inactivation of Smad4 in forelimb buds required ~18 hours. In agreement with these inactivation kinetics, the expression of the BMP target Msx2 was reduced from ~E9.75 onward (Fig. 1E; supplementary material Fig. S1B). By contrast, the three main Bmp ligands were either unchanged or slightly increased in mutant limb buds (supplementary material Fig. S2).

Previous analysis provided evidence that activation and initial upregulation of Grem1 expression in the limb bud mesenchyme depended on BMP4 signaling (Nissim et al., 2006; Bénazet et al., 2009). In Smad4ΔΔM forelimb buds, Grem1 expression was reduced, but its posterior domain was maintained at E10.25 (Fig. 2A). To discriminate between SMAD4-independent regulation and the well-characterized positive regulation of Grem1 expression by SHH signaling (Zühiga et al., 1999; Panman et al., 2006), forelimb buds lacking both Shh and Smad4 were analysed. In contrast to the low expression in ShhΔΔ forelimb buds (Fig. 2B, left panel), Grem1 expression was lost from ShhΔΔSmad4ΔΔM forelimb buds (Fig. 2B, right panel). These results showed that Grem1 expression in the limb bud mesenchyme is positively regulated by both SHH and SMAD4-mediated signal transduction. Contrary to Grem1, Shh expression and signal transduction appeared normal (Fig. 2C and data not shown), whereas the AER-Fgf8 expression domain was broadened and levels increased from ~E10.5 onward in Smad4ΔΔM forelimb buds (Fig. 2D). This analysis showed that mesenchymal Smad4 is either not required, or its functional inactivation occurred too late to disrupt establishment of the AER-Fgf8 expression domain (Bénazet et al., 2009), in contrast to its direct requirement in the AER (Bénazet and Zeller, unpublished). As the SHH/GREM1/AER-FGF signaling system terminated in wild-type limb buds, the expression of both Shh (Fig. 2E) and AER-Fgf8 (Fig. 2F) persisted in Smad4ΔΔM forelimb buds. Rather unexpected from previous analysis (Scherz et al., 2004; Verheyden and Sun, 2008), these results reveal that mesenchymal Smad4 is required for correct termination of the epithelial-mesenchymal feedback signaling system. In agreement with the initially normal progression of limb bud development, no striking alterations in cell proliferation were apparent in Smad4ΔΔM forelimb buds (supplementary material Fig. S3A). However, apoptosis of core mesenchymal cells was increased up to ~E11.5, whereas apoptosis in mutant handplates was suppressed because of the absence of the interdigit mesenchyme (supplementary material Fig. S3B; Fig. 3).

The spatiotemporal expression of the 5′-most HoxD and HoxA genes, which regulate limb bud patterning and outgrowth (reviewed by Zakany and Duboule, 2007) was not significantly altered up to E11.5 (data not shown). Fig. 3 shows the expression of select markers for the autopod, digit and interdigit territories. Hoxa13 correctly delineated the presumptive autopod territory in mutant limb buds, but none of the putative digit primordia were outlined (Fig. 3A). Cyp26b1 is normally expressed in the distal part of all digit primordia including the PFRs (Fig. 3B) (Yashiro et al., 2004). Smad4ΔΔM forelimb buds lacked these PFR-like expression domains and Cyp26b1 was largely confined to the anterior and posterior margins (Fig. 3B). Dlx5 is normally expressed by the AER and interdigit mesenchyme in wild-type limb buds (Robledo et al., 2002), but its expression was confined largely to the AER in Smad4ΔΔM forelimb buds (Fig. 3C). These results indicated that
was progressively lowered (Fig. 4A, right-most lower panel). This downregulation was paralleled by a complete loss of the Sox9-positive primordia corresponding to carpals and digit rays (Fig. 4A,B). By E12.25, the characteristic Sox9-positive domains prefiguring all of the chondrogenic limb skeletal elements were largely lost from Smad4ΔΔM limb buds (Fig. 4A, right-most panels). By E13.5, only very low Sox9 expression remained in Smad4ΔΔM limb buds, whereas Sox9 was abundant in the PFRs and phalanges of all wild-type digits (supplementary material Fig. S4A). To gain further insight, OPT (Sharpe et al., 2002) was used in combination with iso-surface rendering for comparative analysis of the Sox9 transcript distribution in wild-type and Smad4ΔΔM forelimb buds. This analysis revealed thinning of the distal arch of Sox9-positive cells (Fig. 4B, brackets), whereas the proximal domain was enlarged in Smad4ΔΔM forelimb buds (Fig. 4B, arrowheads; better seen in supplementary material Movie 1).

The requirement of Smad4 for formation of the Sox9-positive digit rays could be transient, as evidenced by the kinetics of Hoxa13Cre− mediated conditional inactivation of either one (Smad4ΔΔA13) or both Smad4 alleles (Fig. 4C, Smad4ΔΔA13/ΔA13; supplementary material Fig. S4). Inactivation of Smad4 using the Hoxa13Cre knock-in allele (Scotti and Kmita, 2012) resulted in progressive clearance of Smad4 transcripts during autopod development and specific agenesis of the autopod skeleton (supplementary material Fig. S4B,C). The functional loss of Smad4 should be faster in Smad4ΔΔA13 than Smad4ΔΔA13/ΔA13 forelimb buds, as two conditional alleles have to be inactivated and twice as much SMAD4 proteins cleared in the latter (Bénazet et al., 2009). As only one Hoxa13 allele is functional in mouse embryos carrying the Hoxa13Cre knock-in, we first established that formation of digit ray primordia was not altered in Hoxa13Cre− and Hoxa13Cre Smad4ΔΔA13/ΔA13 forelimb buds (Fig. 4C). However, only the two posteriormost digit ray primordia were apparent in Smad4ΔΔA13 forelimb buds, whereas the primordia of digit rays 2-5 formed in all Smad4ΔΔA13/ΔA13 forelimb buds (Fig. 4D). These results indicated that morphogenesis of anterior digit ray primordia could require Smad4 for longer or at a later stage than posterior primordia.

**SOX9-positive mesenchymal progenitors fail to aggregate and initiate chondrogenic differentiation of all cartilage elements in Smad4ΔΔM limb buds**

As whole-mount RNA in situ hybridization analysis did not provide cellular resolution, the SOX9 protein distribution was analysed on serial sections of forelimb buds (Fig. 5, red fluorescence). At E10.5, scattered SOX9-positive mesenchymal cells were detected mainly in the core of both wild-type and Smad4ΔΔM forelimb buds (Fig. 5A,B). By E11.75, SOX9-positive progenitors condensed to initiate formation of the cartilage elements of the humerus and the ulna/radius and the posterior digit 4 in wild-type limb buds (Fig. 5C) (Zhu et al., 2008). In Smad4ΔΔM forelimb buds, a small proximal condensation was apparent (Fig. 5D, arrowhead), but most SOX9-positive cells remained loose and in a pattern reminiscent of the Sox9-positive distal arch (compare Fig. 5D and Fig. 4A). During subsequent development, no organized mesenchymal condensations and cartilage elements formed in Smad4ΔΔM forelimb buds (Fig. 5E-H). Rather, the small proximal condensation disappeared and SOX9-positive cells dispersed throughout the distal mesenchyme (Fig. 5F,H). Taken together, this analysis indicated that aggregation of the SOX9-positive progenitors to initiate chondrogenic differentiation and

**Fig. 3. Mesenchymal Smad4 inactivation disrupts the spatial distribution of genes that mark specific autopod territories.**

(A) Hoxa13 expression at E12.5 (60 somites). In wild-type forelimb buds, the distal parts of the developing digits and phalanges express the highest levels of Hoxa13 transcript. In Smad4ΔΔM forelimb buds, Hoxa13 expression is distally restricted but not localized to specific structures. (B) Gsc26b1 marks the PFRs of all developing digit primordia in wild type, whereas no differential expression is apparent in Smad4ΔΔM forelimb buds. (C) Dlx5 is expressed by the AER, subectodermal mesenchyme, and the interdigit mesenchyme in wild type, which is lacking in Smad4ΔΔM forelimb buds. The wild-type digit ray primordia are labelled 1 (thumb) to 5 (little finger).
formation of cartilage elements was disrupted in Smad4-deficient forelimb buds.

To gain further insight into the underlying cellular defects, wild-type and mutant mesenchymal cells were dissociated from limb buds at E11.5 and high-density cultures were used to assess their chondrogenic differential potential (Barna and Niswander, 2007). A hallmark of chondrogenic differentiation is the transcriptional activation of Col2a1, a direct transcriptional target of SOX9 that encodes the pro-alpha1(II) chain of COL type II, which is cartilage-specific (Bell et al., 1997; Lefebvre et al., 1997). In wild-type mesenchymal cultures, cells aggregated and compacted to form the typical cell-dense condensations often paralleled by reduced SOX9 expression, and the onset of chondrogenesis was revealed by abundant COL type II (Fig. 6A).

By contrast, no COL type II was detected in areas of less densely packed cells (Fig. 6A, enlargements), which revealed that chondrogenesis was initiated in regions with high mesenchymal cell density, i.e. after cells had aggregated. By contrast, slightly lower levels of SOX9 were consistently detected in mesenchymal cells isolated from Smad4-deficient forelimb buds, which failed to aggregate and produce COL type II fibres (Fig. 6B). This failure of Smad4-deficient mesenchymal cells to aggregate appeared more severe than the effects of Noggin-mediated inhibition of BMP activity in limb bud mesenchymal cells, which aggregate but fail to undergo subsequent compaction and initiation of chondrogenesis (Barna and Niswander, 2007).

To gain further insight, we selectively inhibited BMP type I receptors and thereby BMP signal transduction using Dorsomorphin (Yu et al., 2008), which disrupted condensation and COL type II expression of wild-type cells similar to Smad4-deficient cells (compare supplementary material Fig. S5B and Fig. 6B). Inhibiting TGFβ receptors with the selective antagonist SB431542 (Inman et al., 2002) blocked cell aggregation and COL type II deposition to a similar extent (supplementary material Fig. S5B) (Karamboulas et al., 2010). These results indicated that BMP and TGFβ signal transduction were both required for initiating chondrogenesis of mesenchymal progenitors in culture.

As these experiments revealed striking effects on deposition of extracellular COL type II fibres, the extent to which Col2a1 expression might be altered in Smad4-deficient forelimb buds was determined (Fig. 7A). At E11.5, the Col2a1 transcript distributions in wild-type and mutant forelimb buds were similar and comparable to Sox9 at this stage (compare Fig. 7A, left panels and Fig. 4A). As autopod development progressed, Col2a1 was expressed abundantly by all skeletal elements including digit rays in wild-type limb buds (Fig. 7A, upper panels). By contrast, Col2a1 expression remained diffuse and was progressively lost.
from Smad4Δ/ΔM forelimb buds (Fig. 7A, lower panels). Indeed, high levels of COL type II protein were detected in all chondrogenic primordia of the skeletal elements in wild-type limb buds, whereas only few scattered positive clusters were detected in Smad4Δ/ΔM limb buds (Fig. 7B). Taken together, this analysis (Figs 5-7) revealed the disruption of chondrogenesis and absence of all cartilage elements prefiguring the limb skeleton in Smad4Δ/ΔM forelimb buds.

To determine the extent to which Smad4-deficient mesenchymal cells might acquire non-chondrogenic fates, genes marking other mesenchymal lineages were analysed. Collagen type I (COL type I) is expressed by diverse connective tissues such as skin, tendons, ligaments and bones, and muscle-associated connective tissues (see e.g. ten Berge et al., 2008). In wild-type limb buds, Col1a2 was detected first in the proximal mesenchyme (Fig. 7C, E11.5). At later stages, Col1a2 outlined the forming perichondria and tendons (Fig. 7C, arrowheads). In Smad4Δ/ΔM forelimb buds, Col1a2 transcript and protein were detected throughout the mesenchyme (Fig. 7C,D, lower panels). COL type I localized to the perichondrium and developing dermis in wild-type limb buds (Fig. 7D, upper panels), whereas its expression appeared more widespread and disorganized in Smad4Δ/ΔM forelimb buds (Fig. 7D, lower panels).

As decorin (Dcn) regulates the assembly of collagen fibres (Danielson et al., 1997), its expression was analysed. In wild-type limb buds, the initially rather diffuse Dcn expression became excluded from the chondrogenic cores of all developing cartilage elements including digit rays (Fig. 8A, upper panels and data not shown). By contrast, Dcn expression was enhanced and persisted throughout the mesenchyme of Smad4Δ/ΔM forelimb buds (Fig. 8A, lower panels; supplementary material Fig. S6A). Possible effects on Dcn expression were also assessed in forelimb buds lacking both mesenchymal Bmp2 and Bmp4, as the two posteriormost digit ray primordia are not formed (Fig. 8B) (Bandopadhyay et al., 2006). Analogous to the Smad4 deficiency (Fig. 8A), the absence of these two digit ray primordia results in uniformly increased expression of Dcn in the posterior mesenchyme of limb buds lacking both Bmps (Fig. 8B, posterior region indicated by a bracket). The striking similarity of these alterations indicates that formation of the posterior Sox9-positive digit ray primordia and restriction of Dcn expression depends on BMP signal transduction that is mediated by SMAD4.

In addition, the expression of the transcriptional regulator scleraxis (Scx, marking tendon and ligament progenitors) (Schweitzer et al., 2001) was also increased in Smad4Δ/ΔM forelimb buds (supplementary material Fig. S6B). While Scx marked both dorsal and ventral tendons in wild-type limb buds (Fig. 8C, upper panels), its expression was upregulated and diffuse in Smad4Δ/ΔM forelimb buds (Fig. 8C, lower panels). This was paralleled by widespread and diffuse expression of Fijx1, which normally marks the attachment points of the forming tendons and the developing joints (Rock et al., 2005), in Smad4-deficient forelimb buds (supplementary material Fig. S6C). In addition to tendons and ligaments, limb skeletal muscles form in concert with the limb skeletal elements. Analysis of the MyoD distribution (marking all limb muscle primordia) (Francis-West et al., 2003) showed that dysplastic skeletal muscles were present in the flank and proximal part of Smad4-deficient forelimb buds, whereas the distal mesenchyme was devoid of MyoD-expressing cells (supplementary...
DISCUSSION

Our genetic analysis of mesenchymal Smad4 during mouse limb bud development uncovers its dual requirement for initiating: (1) formation of the Sox9-positive digit ray primordia; and (2) aggregation and chondrogenic differentiation of all Sox9-positive prechondrogenic progenitors that give rise to the cartilage elements of the future limb skeleton (Fig. 8D). SMAD4 is essential to induce chondrogenesis because cell aggregation, which is the first step toward chondrogenic differentiation, is disrupted and mutant progenitors remain scattered in Smad4Δ/ΔM forelimb buds (Fig. 8D) (see also Pizette and Niswander, 2000; Barna and Niswander, 2007). By contrast, the expression of Sox9, which prefigures the primordia of the stylopod and zeugopod, is normal up to at least E11.0 in Smad4Δ/ΔM forelimb buds. A role of SMAD4 in activation of Sox9 expression in early limb buds cannot be formally excluded, as the protein is only cleared around the time Sox9 is activated in Smad4Δ/ΔM forelimb buds. Nevertheless, the Smad4-independent phase of Sox9 expression coincides well with the period of low BMP activity in mouse limb buds (Bénazet et al., 2009), which corroborates the proposal that the spatiotemporal regulation of Sox9 expression does not require SMAD4-mediated signal transduction before handplate formation (this study). Indeed, it has been proposed that BMPs inhibit Sox9 expression and initiation of chondrogenic differentiation in the mesenchyme of early limb buds; an effect that can be alleviated by short-term exposure to TGFβ signaling in culture (Karamboulas et al., 2010). However, inhibition of BMP or TGFβ signal transduction in vitro disrupted chondrogenesis in a similar way to the genetic Smad4 deficiency, but did not cause complete loss of SOX9 (this study).

In addition to Sox9, the mesenchymal expression of the early chondrocyte marker Col2a1 is activated normally in Smad4Δ/ΔM forelimb buds. During distal progression of limb bud development, Col2a1 expression is lost concurrent with the failure to refine SOX9 protein distribution, as expected from direct transcriptional regulation of Col2a1 by SOX9 (Bell et al., 1997; Lefebvre et al., 1997). In Smad4Δ/ΔM forelimb buds, the progressive loss of Sox9 and Col2a1 was paralleled by widespread and diffuse expression of markers for other lineages and a general absence of tissue organization and differentiation. Therefore, SMAD4 appears to function in signaling networks that restrict non-chondrogenic progenitors and connective tissue lineages. In agreement with the diffuse Sxc expression in Smad4Δ/ΔM forelimb buds, inhibition of BMP signaling results in ectopic Sxc expression without the formation of additional tendons (Schweitzer et al., 2001). However,
as both Colla1 and Colla2 are positively regulated by SCX (Cserjesi et al., 1995), the ectopic Scx expression in Smad4ΔΔM forelimb buds could contribute to the increased COL type I expression (this study). These molecular changes and the complete lack of tissue organization in Smad4-deficient limb buds (Fig. 8D) indicate that SMAD4 functions in the morphoregulatory networks that pattern diverse limb bud mesenchymal lineages and/or orchestrate coordinated tissue differentiation. Taken together, these studies uncover the recurrent differential requirement of SMAD4-mediated canonical BMP/TGFβ signal transduction during specification and aggregation of SOX9-positive prechondrogenic progenitors and initiation of their chondrogenic differentiation (Fig. 8D).

Although formation of the Sox9-positive primordia of the stylopod and zeugopod appears Smad4-independent, the induction of the Sox9-positive digit ray primordia depends critically on Smad4 and increased BMP activity (Bandyopadhyay et al., 2006; Bénazet et al., 2009; Lopez-Rios et al., 2012; this study). As the proximal Sox9-positive domains are enlarged, the mutant prechondrogenic cells may contribute preferentially to proximal primordia as evidenced by thinning of the distal arch in Smad4ΔΔM forelimb buds (this study). Such a differential requirement of Smad4 might be of evolutionary relevance as the autopod and digits of tetrapod limbs are so-called neomorphic structures, whose development may have relied on co-opting novel gene-regulatory networks (reviewed by Woltering and Duboule, 2010). Our analysis indicates that limb bud development and formation of Sox9-positive digit ray primordia become Smad4 dependent as BMP activity rises owing to self-termination of the SHH/Grem1/AER-FGF signaling system (Scherz et al., 2004; Verheyden and Sun, 2008; Benazet et al., 2009). During this developmental period, differential BMP signaling appears to determine digit identities in both chicken and mouse (Dahn and Fallon, 2000; Suzuki et al., 2008; Witte et al., 2010). As extensive genetic analysis of TGFβ signaling in mouse embryos has not produced similar phenotypes (see e.g. Sanford et al., 1997; Seo and Serra, 2007), SMAD4-mediated BMP signal transduction is likely to be predominant during these advanced stages of digit ray primordia growth and patterning. It has been proposed that the progenitors giving rise to the digit ray primordia are recruited from the distalmost limb bud mesenchyme (Suzuki et al., 2008), which is defective in Smad4ΔΔM forelimb buds as revealed by the increasing gap of Sox9-negative cells in the distal mesenchyme (this study). In addition, temporally controlled Smad4 inactivation in the autopod provided evidence for its possible differential requirement for posterior and anterior digit ray primordia. A region free of the BMP antagonist GREM1 is generated within the posterior limb bud mesenchyme by the population of Shh descendants, which is refractory to Grem1 expression and increases as limb bud development progresses (Scherz et al., 2004). These results together with the differential requirement observed in this study indicate that BMP activity might first rise in the posterior mesenchyme, where Bmp2 and Bmp4 are co-expressed and required for digit ray formation (Bandyopadhyay et al., 2006).
mechanism that controls BMP-dependent formation of the digit ray primordia was recently uncovered (Lopez-Rios et al., 2012). Inactivation of Gli3 in the developing autopod promoted cell cycle entry and delayed the exit of digit progenitors to chondrogenesis due to prolonged GREM1-mediated BMP-antagonism in the anterior limb bud mesenchyme. Taken together, this indicates that two distinct mechanisms regulate digit ray formation in the posterior and anterior handplate via BMP-dependent SMAD4-mediated signal transduction. The spatiotemporal requirements of BMPs and SMAD4 for posterior and anterior digit ray primordia (Bandyopadhyay et al., 2006; Lopez-Rios et al., 2012; this study) is apparently at odds with the genetic analysis of the temporal requirement of Shh during mouse limb bud development (Zhu et al., 2008). This analysis showed that digit ray primordia form with alternating posterior to anterior sequence (starting with d4, d2, d5 and ending with d3), while digit 1 forms in a SHH-independent manner. However, these apparent discrepancies are resolved if BMPs and SMAD4 were transiently required to initiate a posterior and anterior mesenchymal condensation that would subsequently give rise to the primordia of digits 4 and 5 and digits 2 and 3, respectively, following the sequence defined by Zhu et al. (Zhu et al., 2008).

Finally, BMPs are widely used to induce chondrogenesis in ex vivo models of cartilage and bone engineering, but outcomes are often conflicting (Jiang et al., 2010; Sanchez-Adams and Athanasiou, 2012). In light of the complex regulation of chondrogenesis during limb bud development, ex vivo developmental engineering of cartilage and endochondral bone (Scotti et al., 2010) will probably require activation and inhibition of BMP signal transduction in a spatiotemporally controlled manner as revealed by the present study. In particular, this study highlights the temporal kinetics of the SMAD4 requirement during formation of the Sox9-positive digit ray primordia and its more general role in inducing aggregation of the mesenchymal progenitors during the onset of chondrogenesis, which is key to developmentally engineering cartilage and bone.

Acknowledgements
We are grateful to Chucia Deng for providing us with the Smad4 conditional allele and to Georg Hollander for founder mouse, to Marie Kmita and Martina Scotti for providing us with the Hoxa13-Cre knock-in mouse strain, and to Philippa Francis-West and Manfred Gesser for providing plasmids for BMP signaling in the developing mouse embryo. Development 131, 3501-3512.


6.1.2. Paper 2: “Conserved cis-regulatory regions in a large genomic landscape control SHH and BMP-regulated Gremlin1 expression in mouse limb buds”

In this publication, the three conserved HMCO elements located in the Gremlin1 cis-regulatory landscape were tested for their capacity to regulate the expression of Gremlin1 during limb bud development. The analysis focuses on the 9kb region that includes HMCO1, termed Gremlin1 Regulatory Sequence 1 (GRS1). The GRS1 region was used to generate stable transgenic mouse lines with LacZ reporter. My contribution to this project includes the detailed genomic sequence alignment for the HMCOs element in additional species (Figure 1 and Figure S1). I also carried out the analysis of GRS1-βglob-LacZ transgene reporter expression in seven transgenic founders (data not shown), and the subsequent analysis of GRS1-βglob-LacZ reporter expression in stable GRS1tg10/+ and GRS1tg2/+ embryos in wild-type and various mutant backgrounds (Figures 4-6 and Figures S2-S4). Finally, I prepared all figures and genomic schemes of the manuscript.
Conserved cis-regulatory regions in a large genomic landscape control SHH and BMP-regulated Gremlin1 expression in mouse limb buds

Aimée Zuniga1, Frédéric Laurent1, Javier Lopez-Rios1, Christian Klasen2,3, Nicolas Matt1,4 and Rolf Zeller1*

Abstract

Background: Mouse limb bud is a prime model to study the regulatory interactions that control vertebrate organogenesis. Major aspects of limb bud development are controlled by feedback loops that define a self-regulatory signalling system. The SHH/GREM1/AER-FGF feedback loop forms the core of this signalling system that operates between the posterior mesenchymal organiser and the ectodermal signalling centre. The BMP antagonist Gremlin1 (GREM1) is a critical node in this system, whose dynamic expression is controlled by BMP, SHH, and FGF signalling and key to normal progression of limb bud development. Previous analysis identified a distant cis-regulatory landscape within the neighbouring Formin1 (Fnm1) locus that is required for Grem1 expression, reminiscent of the genomic landscapes controlling HoxD and Shh expression in limb buds.

Results: Three highly conserved regions (HMCO1-3) were identified within the previously defined critical genomic region and tested for their ability to regulate Grem1 expression in mouse limb buds. Using a combination of BAC and conventional transgenic approaches, a 9 kb region located ~70 kb downstream of the Grem1 transcription unit was identified. This region, termed Grem1 Regulatory Sequence 1 (GRS1), is able to recapitulate major aspects of Grem1 expression, as it drives expression of a LacZ reporter into the posterior and, to a lesser extent, in the distal-anterior mesenchyme. Crossing the GRS1 transgene into embryos with alterations in the SHH and BMP pathways established that GRS1 depends on SHH and is modulated by BMP signalling, i.e. integrates inputs from these pathways. Chromatin immunoprecipitation revealed interaction of endogenous GLI3 proteins with the core cis-regulatory elements in the GRS1 region. As GLI3 is a mediator of SHH signal transduction, these results indicated that SHH directly controls Grem1 expression through the GRS1 region. Finally, all cis-regulatory regions within the Grem1 genomic landscape locate to the DNase I hypersensitive sites identified in this genomic region by the ENCODE consortium.

Conclusions: This study establishes that distant cis-regulatory regions scattered through a larger genomic landscape control the highly dynamic expression of Grem1, which is key to normal progression of mouse limb bud development.

Background

During embryonic development, the spatio-temporal expression of transcriptional regulators and morphogenetic signals is tightly controlled. In fact, different types of congenital malformations are caused by mutations affecting specific cis-regulatory and non-coding genomic regions that alter the expression of genes or gene clusters in specific tissues (reviewed in ref. [1-3]). Genome-wide functional mapping revealed many of the large genomic landscapes that control the expression of developmental regulator genes in a global and tissue-specific manner [4]. This study and others reveal that the cis-regulatory regions controlling gene expression in a particular tissue are often located several hundred kilobases (kb) or even megabases (Mb) up- or downstream of the transcriptional start sites in other loci, and may control the expression of several genes [5-7]. Recently, it has been shown how multiple cis-regulatory regions interact to control the...
expression of the 5’HoxD gene complex in the presumptive digit territory. The transcriptionally active part of the 5’HoxD gene cluster forms a so-called regulatory archipelago in which dispersed cis-regulatory elements cooperate to control gene expression in the distal limb bud by interacting with proximal promoters over large distances [8]. In fact, regulatory landscapes with distant and dispersed cis-regulatory regions seem to be a recurring theme in the dynamic spatio-temporal regulation of genes that are essential for limb bud development (reviewed in ref. [2,9,10], see also below).

Vertebrate limb bud development is controlled by interactions between two main signalling centres, the apical ectodermal ridge (AER) and the zone of polarizing activity (ZPA) located in the posterior limb mesenchyme. These two signalling centres interact as part of a self-regulatory feedback signalling system involving several signalling pathways [11]. In particular, ZPA cells produce the Sonic Hedgehog (SHH) signal, which together with GREMLIN1-mediated antagonism of Bone Morphogenetic Proteins (BMPs) in the posterior-distal limb bud mesenchyme propagates Fibroblast Growth Factor (FGF) signalling in the AER [11-14]. This SHH/GREM1/AER-FGF feedback-signalling loop promotes distal progression of limb bud outgrowth and formation of the autopod which gives rise to carpals and digit rays (reviewed in ref. [15]). SHH expression in the posterior limb bud mesenchyme is controlled by a cis-regulatory region located about 800 kb upstream of the transcriptional start site. Deletion of this ZPA regulatory region (ZRS) results in loss of distal limb structures [16], while point mutations within the ZRS result in anterior ectopic Shh expression and duplications of the thumb and/or anterior digits in different species [17,18]. In cells actively transcribing Shh, the ZRS loops out of its chromosomal territory to the Shh transcription unit, which reveals how distant cis-regulatory elements control gene expression [19]. Recently, several transcriptional regulators have been identified that control Shh expression by directly interacting with the distant ZRS region. These include HOX proteins, the bHLH transcription factor HAND2, and ETS transcription factors, providing a glimpse of the complex transcriptional regulation of Shh in the posterior limb bud mesenchyme [20-22].

SHH signalling is transduced by the GLI family of transcriptional regulators and inhibits the constitutive processing of the full-length GLI3 activator to its repressor form GLI3R (see e.g. ref. [23]). Of the three GLI transcription factors expressed during limb bud development, only Gli3 is essential on its own (see e.g. ref. [24]). The inactivation of Gli3 alters morphogenetic feedback signalling and results in formation of additional anterior digits [23,25-27]. In particular, Gli3 restraints the proliferation of mesenchymal progenitors in the anterior limb bud mesenchyme and promotes initiation of digit ray chondrogenesis by directly repressing Grem1 expression during handplate formation [26].

We have previously shown that the expression of Grem1 in the limb bud mesenchyme is controlled by a large genomic landscape downstream of its transcription unit [7]. In fact, several of the classical limb deformity mutations that disrupt distal limb bud development and formation of metanephric kidneys in the mouse are caused by deletions affecting this cis-regulatory landscape rather than directly altering the Grem1 gene products [7,13,28]. Molecular and genetic analysis in the mouse identified a 70 kb genomic region located downstream of the Grem1 transcription unit within the neighbouring Formin1 (Fmn1) gene that is required for Grem1 expression in the limb bud mesenchyme [7,29]. Detailed analysis of this Grem1 genomic landscape revealed similarities with the global control region (GCR) that controls the expression of 5’HoxD genes in the limb bud mesenchyme [6], but did not reveal the structural nature and transacting factors and/or signalling pathways controlling these cis-regulatory regions.

To gain further insights into the Grem1 landscape, the potential cis-regulatory activities of the three highest evolutionarily conserved genomic regions within the 70 kb Grem1 critical region were analysed. In addition to its expression in the posterior mesenchyme, emphasis was also given to the coordinated distal-anterior expansion of Grem1 expression, which is key to orderly progression of limb bud development [14,30]. Combining BAC with conventional transgenic approaches, we identified a 9 kb genomic region that is able to recapitulate major but not all aspects of the dynamic spatio-temporal Grem1 expression in the limb bud mesenchyme. This 9 kb region was termed Grem1 Regulatory Sequence 1 (GRS1), and contains a core sequence that is essential to express a LacZ transgene under control of a β-globin minimal promoter in a Grem1-like pattern. The GRS1 region drives gene expression into the posterior and subsequently distal-anterior mesenchyme, i.e. reproduces aspects of the distal-anterior expansion of Grem1 expression during progression of limb bud development. By crossing GRS1 transgenic mice into different mutant contexts, we establish that the GRS1 region is controlled by inputs from both the SHH and BMP signalling pathways in limb buds. In addition, chromatin immunoprecipitation (ChIP) shows that GLI3 interacts with the conserved HMCO1 sequences in the GRS1 region. The functionally relevant cis-regulatory regions identified by the present and two previous studies [31,32] map to the endogenous DNase I hypersensitive sites within the Grem1 genomic landscape recently identified by the ENCODE consortium [33,34]. This indicates that the Grem1 cis-regulatory landscape is composed of at least five active cis-regulatory
regions that control the spatio-temporal expression of Grem1 in mouse embryos and limb buds.

**Results and Discussion**

**Identification of conserved limb bud regulatory regions within the Grem1 genomic landscape**

Using functional genetics in the mouse, we previously identified a ~70 kb region located downstream of Grem1 that is required for its expression in the limb bud mesenchyme (Figure 1A, [7]). This limb bud cis-regulatory region is located between coding exons 19 and 23 of the neighbouring Fmn1 gene. The Evolutionary Conserved Regions (ECR) genome browser (http://ecrbrowser.dcode.org/) was used for multiple sequence alignments to compare the mouse genome with different mammalian and the chicken genomes. This analysis revealed several blocks of highly conserved sequences among mammalian species, but only three of them were also highly conserved in the chicken genomes. This analysis revealed several blocks of highly conserved sequences among mammalian species, but only three of them were also highly conserved in the chicken and termed HMCO1, 2 and 3 (Human-Mouse-Chicken-Opossum conserved sequences 1 to 3, Figure 1A). The most conserved parts of these three HMCO core regions are ~80% identical (for details see Additional file 1 and Additional file 2). Of these three regions only HMCO1 is also present in the orthologous genomic region in frogs (Figure 1A), which express Grem1 during limb bud development [35]. Despite the fact that Grem1 is expressed during fin bud development [36], no HMCO homologies are present in the zebrafish genome (Figure 1A). ClustalW2 alignments (http://www.ebi.ac.uk/Tools/msa/clustalw2/) of the Fmn1-Grem1 regions from different species revealed the absence of HMCO homologies in other ray-finned fish species such as fugu and medaka (Figure 1B). In contrast, all three HMCO regions are present in the genome of coelacanth, a lobe-finned fish closely related to tetrapods (Figure 1B and Additional file 1) [37]. As the Grem1 and Fmn1 loci are linked in all species analysed (Figure 1B), the distance between the orthologous Fmn1 coding exon 22 (located adjacent to HMCO1) and Grem1 coding exon 2 was determined. During evolution, this distance increased as it is smallest in the genomes of ray finned fishes and at least ~10-fold larger in tetrapods and coelacanth (Figure 1B). In addition, the distance between Fmn1 exon 19 (close to HMCO3) and exon 22 correlates well with the presence or absence of the three HMCO regions (Figure 1B and data not shown). In ray-finned fishes and frogs (lacking all or specifically HMCO2, 3, respectively), the distance is ~3-7 kb, while in mammals and coelacanth it ranges from ~50-70 kb. Thus, the increase of intronic and intergenic regions in the Fmn1-Grem1 landscape correlates well with appearance of the three HMCO regions during vertebrate evolution.

To determine the requirement of each of these three HMCO regions for Grem1 expression in mouse limb buds, we used a BAC-based strategy in combination with analysis of transgenic founder embryos (Figure 2A-D and Additional file 3). A 250 kb mouse genomic BAC encoding the critical region and the Grem1 transcription unit was used to fuse the LacZ gene in frame with the Grem1 ORF [7]. We assessed the expression of the LacZ reporter transgene by analysing the spatio-temporal distribution of β-galactosidase activity (Figure 2A-D). Expression of the Grem1-LacZ fusion protein in the forelimb bud mesenchyme was detected in a posterior-distal domain (Figure 2A), mimicking the early endogenous Grem1 expression rather accurately. Therefore, the Grem1-LacZ BAC construct was used to engineer deletions of the three HMCO regions and determine their requirement for Grem1-LacZ expression in the posterior mesenchyme. To control for reproducible generation of expressing BAC transgenic founder embryos, the Grem1-LacZ BAC (Figure 2A) was injected in parallel to the BACs with engineered deletions (Figure 2B-D). While the control Grem1-LacZ BAC was always robustly expressed in the posterior mesenchyme (Figure 2A), deletion of the 520 bp HMCO1 core region resulted in complete loss of LacZ expression from limb buds (Figure 2B, n = 5/7 embryos with LacZ expression, for details see Additional file 3). In contrast, deletion of the 1279 bp HMCO2 core region only caused partial loss of LacZ expression from the posterior limb bud mesenchyme (Figure 2C; Additional file 3). LacZ remained expressed normally in the majority of founder embryos carrying a 924 bp deletion of the HMCO3 region (Figure 2D, n = 3/5 embryos with LacZ expression, for details see Additional file 3). Taken together, this BAC transgenic analysis establishes HMCO1 as most critical for Grem1 expression in the posterior-distal limb bud mesenchyme. As the HMCO1 core region is highly conserved in tetrapods and lobe-finned fish but not ray-finned fish (Figure 1B), it likely represents a cis-regulatory region important to tetrapod evolution. The other two HMCO regions might contribute to robust expression of Grem1-LacZ in the posterior mesenchyme, as in particular the deletion of HMCO2 results in significantly reduced LacZ expression (Figure 2C).

In an attempt to gain further insight into the enhancer potential and possible interactions of HMCO2/3 with HMCO1, conventional transgenic approaches using a minimal human β-globin promoter (ßglob-LacZ) were employed. However, neither individual HMCO regions (data not shown) nor in a combination of all three was able to drive robust expression of the ßglob-LacZ transgene in the posterior limb bud mesenchyme (Figure 3A, compare to Figure 2A). Expression of the ßglob-LacZ transgene under control of all three HMCO core regions resulted in scattered LacZ positive cells in the anterior-distal mesenchyme of forelimb buds (left panel,
Figure 1 (See legend on next page.)
Figure 3A, n = 3/3). In contrast, the transgene was strongly expressed in the posterior embryo including hindlimb buds (right panel, Figure 3A). These results show that a transgene consisting of an array of the HMCO core regions is unable to drive \( \text{LacZ} \) expression into the posterior forelimb bud mesenchyme. This indicates that additional elements of the \( \text{Grem1} \) genomic landscape are required for \( \text{Grem1} \) expression in limb buds.

Next, we assessed to which extent a larger genomic fragment containing HMCO1 and flanking regions could drive \( \text{ßglob-\text{LacZ}} \) expression into the posterior limb bud mesenchyme (Figure 3B). Analysis of transgenic founder embryos revealed that this transgenic construct drives robust \( \text{ßglob-\text{LacZ}} \) expression in the posterior limb bud mesenchyme (Figure 3B, see also Figure 4), strikingly similar to the domain of the parental \( \text{Grem1-LacZ} \) BAC construct (Figure 2A). Taken together, these results indicate that this cis-regulatory region enhances \( \text{Grem1} \) expression in the posterior limb bud mesenchyme, which prompted us to term this 9 kb region \( \text{GRS1} \) Regulatory Sequence 1 (GRS1, Figure 4A). GRS1 encompasses the HMCO1 region next to \( \text{Fmn1} \) coding exon 22 and a further downstream region that is highly conserved in mammals. This region, like HMCO1 overlaps with a previously identified GLI binding region (GGR, [32]). In addition, two potential GLI binding sites were identified by in silico analysis (Figure 4A).

**Spatio-temporal activity of the GRS1 region during mouse limb bud development**

To analyze comparatively the activity of the GRS1 region with respect to the spatio-temporal regulation of \( \text{Grem1} \) expression, transgenic mouse strains expressing the GRS1-\( \text{ßglob-\text{LacZ}} \) reporter construct were established (Figure 4). Seven transgenic founders were obtained. In three independent strains, \( \text{LacZ} \) was expressed in the posterior fore- and hindlimb bud mesenchyme in a pattern comparable to the transgenic founder embryos (Figure 3B and data not shown). Two of these three transgenic strains were analyzed in detail and used to study the spatio-temporal \( \text{LacZ} \) distribution (Figure 4B, C and Additional file 4). Initial analysis revealed that \( \text{ß} \)-galactosidase activity of the \( \text{LacZ} \) transgene was detected earlier in one strain and is significantly higher than in the other strain (Figure 4B, compare to Figure 4C). As levels did not change during subsequent generations (data not shown), this is likely due to differences in transgene copy number and/or integration site. Therefore, the copy number of both strains was determined by real-time qPCR analysis (Additional file 4). This analysis revealed that ten copies of the GRS1-\( \text{ßglob-\text{LacZ}} \) transgene were integrated into the genome of the strain expressing higher \( \text{LacZ} \) levels (GRS1\(^{tg10}\)), while only two copies were detected in the GRS1\(^{tg2}\) strain (Additional file 4). In GRS1\(^{tg2}\) embryos, \( \text{ß} \)-galactosidase activity is first detected at ~E10.25 in the posterior forelimb bud mesenchyme and continuously increases until ~E11.75 (Figure 4B). From ~E11.0 onwards, scattered positive cells were detected in the anterior mesenchyme, and this anterior expression increased during distal limb bud outgrowth to form a crescent in the distal-anterior autopod (right panels, Figure 4B). By E11.75, \( \text{ß} \)-galactosidase activity was rather variable, such that the anterior crescent separated from the posterior domain in some forelimb buds. In GRS1\(^{tg2}\) embryos, the spatio-temporal pattern is similar, but expression levels are significantly reduced due to the lower transgene copy number, resulting in detection of \( \text{ß} \)-galactosidase activity from only ~E10.5 onwards (Figure 4C, arrowhead). In both strains, GRS1 drove expression of the \( \text{ßglob-\text{LacZ}} \) reporter specifically in the limb bud mesenchyme, despite some low \( \text{ß} \)-galactosidase activity detected in the developing eyes of GRS1\(^{tg2}\) transgenic embryos (Additional file 4). This analysis reveals the robust nature of the GRS1 region, which functions in positive regulation of \( \text{Grem1} \) expression in the limb bud mesenchyme. The posterior domain

---

(See figure on previous page.)

**Figure 1 Identification of three highly conserved non-coding regions in tetrapods and a lobe-finned fish.** (A) Sequence alignment of the genomic region critical for \( \text{Grem1} \) expression in mouse limb buds [7] using the ECR browser with the mouse genome release 9 (mm9) as reference genome. The critical genomic region on mouse chromosome 2 is shown in centromeric (CEN) to telomeric (TEL) orientation and is located downstream of the \( \text{Grem1} \) coding exons. As the critical genomic region is part of the \( \text{Fmn1} \) locus, the interspersed \( \text{Fmn1} \) coding exons are indicated as open boxes in the scheme. Three blocks of highly conserved non-coding sequences were identified (HMCO1-3) and are indicated in blue. Conserved coding regions are indicated in black and non-coding regions conserved ≥74% over ≥100 bp are coloured salmon. The peak detected in the region of HMCO1 in the zebrafish genome corresponds to an array of the HMCO core regions next to \( \text{Grem1} \) and flanking regions similar to the domain of the parental mesenchyme (Figure 3B, see also Figure 4), strikingly noticeable. Embryos revealed that this transgenic construct drives \( \text{ßglob-\text{LacZ}} \) expression in the posterior limb bud mesenchyme. The posterior domain is indicated in blue. Conserved linkage between the \( \text{Grem1} \) and \( \text{Fmn1} \) loci in vertebrates. Increased intergenic distances correlate with the presence of HMCO regions in tetrapods and coelacanth in contrast to ray-finned fishes. The phylogenetic tree analysis was done with the UCSC multiple alignment functions to align the 3’ part of the \( \text{Fmn1} \) coding exons 22. Regions consisting of repetitive sequences are shown in green (see Additional file 2). (B) Conserved linkage between the \( \text{Grem1} \) and \( \text{Fmn1} \) loci in vertebrates. Increased intergenic distances correlate with the presence of HMCO regions in tetrapods and coelacanth in contrast to ray-finned fishes. The phylogenetic tree analysis was done with the UCSC multiple alignment functions to align the 3’ part of the \( \text{Fmn1} \) coding exons 22. Regions consisting of repetitive sequences are shown in green (see Additional file 2).
of β-galactosidase activity at E10.5 (Figure 4B) is comparable to the Grem1 transcript distribution in wild-type limb buds at earlier stages (E9.5-E10.25, Figure 4D). This temporal delay is likely due to postponed transcriptional activation and/or up-regulation of the GRS1 transgene as the establishment of a posterior LacZ expression domain is also only apparent at E10.5 in GRS1<sup>tg10/+</sup> transgenic limb buds (Additional file 5). Therefore, additional cis-regulatory regions likely control the temporally correct early onset of Grem1 expression. Similar delays in the onset of LacZ reporter gene expression have been previously observed by analyzing the ZRS cis-regulatory region that controls Shh expression in mouse limb buds [38]. During distal progression of limb bud development, endogenous Grem1 expression expands distal-anterior within the developing handplate and begins to fade by E11.75 (Figure 4D), due to termination of the SHH/GREM1/AER-FGF feedback loop and GLI3-mediated repression in the anterior limb bud mesenchyme [11,26,39,40]. In contrast, β-galactosidase activity remains high in the posterior limb bud mesenchyme of GRS1 transgenic embryos and the distal-anterior expansion of its expression is significantly delayed and only

---

**Figure 2** A BAC-based strategy to assess the requirement of highly conserved non-coding regions (HMCO1-3) for Grem1 expression in mouse limb buds. (A) A 250 kb BAC spanning the genomic region from Fmn1 exon 19 to 150 kb up-stream of the Grem1 transcription is able to drive expression of a LacZ transgene (blue arrow; inserted in-frame into the Grem1 ORF) into the posterior limb bud mesenchyme in transgenic founder embryos. (B) Deletion of the HMCO1 core region (520 bp) abolishes the LacZ expression. (C, D) Deletion of the core HMCO2 region (1279 bp) results in reduced LacZ expression, while deletion of HMCO3 (924 bp) does not alter the LacZ distribution. β-galactosidase activity colours expressing cells blue.
occurs as transcription of the endogenous Grem1 locus is starting to terminate (E11.5 onwards Figure 4B, C, compare to Figure 4D).

Grem1 expression is restricted to dorsal and ventral limb bud mesenchyme and excluded from the chondro-ogenic core mesenchyme (Figure 5A), which is relevant to BMP-mediated induction of chondrogenesis (ref. [26,41] and Benazet et al., submitted). Furthermore, the LIM-homeodomain transcription factors Lhx2 and Lhx9 have been implicated in regulating Grem1 expression predominantly in the ventral limb bud mesenchyme in response to SHH signalling [42]. Therefore, the extent to which the dorso-ventral transcript distribution is maintained by the GRS1 transgene was assessed (Figure 5B). Indeed, the expression of both the high (Figure 5B) and low copy transgenes (Additional file 6) remained excluded from the core mesenchyme throughout limb bud outgrowth and patterning. Similar to Grem1 transcripts, β-galactosidase activity is higher dorsally than ventrally (Figure 5B, compare to Figure 5A). By E11.75, β-galactosidase activity expands anteriorly in both the dorsal and ventral mesenchyme (E11.75, Figure 5B), reminiscent of the crescent of Grem1 transcript (Figure 5A and see before), which indicates that the GRS1 cis-regulatory regions are required to activate Grem1 expression in the posterior mesenchyme. Alternatively, the anterior region of the Grem1 cis-regulatory regions is able to recapitulate major aspects of Grem1 expression in the limb bud mesenchyme. In particular, the GRS1 also recapitulates aspects of the distal-anterior expansion of Grem1 expression in limb buds (Figure 4, 5), which was not observed in previous attempts to identify cis-regulatory regions that control Grem1 expression in limb buds [7,32]. The fact that the GRS1 transgene recapitulates some aspects of this distal-anterior expansion is important, as Grem1 transcription normally expands anteriorly in register with posterior AER-FGFs and allows propagation of SHH/GREM1/AER-FGF signalling [14,30]. However, the delay in activation and lack of termination of the GRS1 transgene (Figure 4) indicates that other regulatory inputs from the Grem1 genomic landscape are required to regulate its dynamic expression in mouse limb buds.

The GRS1 cis-regulatory region integrates inputs from the SHH and BMP signalling pathways

During limb bud initiation, mesenchymal BMP4 signal transduction is likely required to activate Grem1 expression in the posterior mesenchyme, while SHH is primarily required for up-regulation and distal-anterior expansion during progression of limb bud development [11,14,43]. Furthermore, GLI3 in the anterior mesenchyme and AER-FGF signal transduction are required to restrict and eventually terminate Grem1 expression (starting ~ E11.5-11.75) [26,40]. Therefore, GRS1^{βglob-LacZ} embryos lacking key components of both the SHH and BMP signalling pathway in their limb buds were generated to gain insight into the possible direct impact of these main signalling pathways on the GRS1 element (Figure 6). Analysis of Shh^{A/A} GRS1^{βglob-LacZ} embryos revealed the complete absence of β-galactosidase activity in Shh-deficient forelimb buds (Figure 6A). This contrasts with the endogenous Grem1 expression, which is activated but not maintained in Shh-deficient limb buds [11,14]. These results show that activation of the GRS1 region depends on SHH signalling, which indicated that it could also participate in ectopic Grem1 activation due to anterior ectopic SHH signalling in mouse [14] and chicken limb buds [12]. Furthermore, GLI3 proteins interact with specific cis-regulatory regions in the Grem1 genomic landscape [32] and genetic analysis has shown that Gli3 is essential for the spatio-temporally controlled
restriction and subsequent termination of Grem1 expression in the distal anterior mesenchyme [25,26,44]. In Gli3-deficient GRS1 transgenic (Gli3<sup>x2/x2</sup>GRS1<sup>G10/+</sup>) forelimb buds, β-galactosidase activity in the posterior limb buds is comparable to GRS1<sup>G10/+</sup> limb buds, while expression in the anterior mesenchyme is significantly increased by E11.75 (Figure 6B, compare to right-most panels, Figure 4B). This late up-regulation of anterior β-galactosidase activity indicates that GRS1 is required for GLI3-mediated termination of Grem1 expression in the anterior limb bud mesenchyme, which is essential for the spatio-temporally correct initiation of mesenchymal condensations and chondrogenic differentiation of anterior digits [26].

BMPs control Grem1 activation in the posterior limb bud mesenchyme and directly modulate its expression as part of the self-regulatory SHH/GREM1/AER-FGF feedback signalling system. Therefore, a conditional Smad4 loss-of-function allele [45] in combination with the Prx1-Cre recombinase strain [46] was used to inactivate Smad4<sup>x1/SM</sup> and thereby canonical BMP signal transduction in the limb bud mesenchyme. In Smad4<sup>x1/SM</sup> mutant limb buds, endogenous Grem1 expression is down-regulated but not lost, while Shh expression remains...
Figure 5 Dorsal and ventral restriction of GRS1-mediated β-galactosidase activity. (A) Grem1 transcript remains restricted to the dorsal and ventral forelimb bud mesenchyme in wild-type embryos throughout limb bud development. (B) GRS1-mediated expression of βglob-LacZ is able to correctly restrict β-galactosidase activity along the dorso-ventral limb bud axis. D: dorsal; V: ventral.

Figure 6 The GRS1 transgene is controlled by both SHH and BMP activity. (A) GRS1-dependent β-galactosidase activity is completely lost from Shh-deficient limb buds (Shh^{Δ/Δ} GRS1^{tg10/+}). (B) In contrast, inactivation of Gli3 results in anterior up-regulation of β-galactosidase activity at E11.75 (Gli3^{Xt/Xt} GRS1^{tg10/+}). (C) β-galactosidase activity fails to expand anteriorly in Smad4-deficient limb buds (Smad4^{Δ/ΔM} GRS1^{tg10/+}). (D) β-galactosidase activity is up-regulated in Grem1-deficient limb buds from E11.0 onward (Grem1^{Δ/Δ} GRS1^{tg10/+}).
In Smad4Δ/AM;GRI3Δ/Δ mice, β-galactosidase activity appears normal in the posterior limb bud mesenchyme, while distal-anterior expansion fails to occur (Figure 6C, compare to Figure 4B). This indicates that GRS1 activity is modulated by SMAD4-mediated signal transduction during the progression of limb bud development. These results indicate that also the distal-anterior expansion of Grem1 expression depends on BMP activity and agree with the observation that genetic lowering of Bmp4 results in a global reduction of Grem1 expression in the limb bud mesenchyme [11].

Finally, in Grem1-deficient embryos, BMP activity is increased due to reduced BMP antagonism, which results in up-regulation of non-functional Grem1 transcript [13]. As BMP signal transduction modulates GRS1 activity during advanced limb bud development (Figure 6C), we also determined the potential influence of Grem1 deficiency on GRS1 activity. In Grem1Δ/Δ;GRS1Δ/Δ E10.5 limb buds β-galactosidase activity is initially similar to GRS1Δ/Δ limb buds (Figure 6D, compare to Figure 4B). Subsequently, β-galactosidase activity is increased and by E11.75 the entire Grem1Δ/Δ;GRS1Δ/Δ limb bud is positive (Figure 6D). These alterations of β-galactosidase activity in Smad4Δ/AM and Grem1Δ/Δ forelimb buds reveal that GRS1 activity is extensively modulated by changes in BMP activity, in particular also the anterior expansion of its expression. Taken together, this genetic analysis (Figure 6) shows that GRS1 activity critically depends on SHH and is modulated extensively by BMP signal transduction. This analysis identifies the GRS1, located ~70 kb downstream of the Grem1 transcription unit as a cis-regulatory region that integrates inputs by both SHH and BMP signal transduction. However, as previous analysis provided good evidence that Grem1 is activated by mesenchymal BMP signalling upstream of establishing the SHH/FGF feedback loop [11], these studies point to the existence of additional unknown BMP response regions that activate Grem1 expression in the posterior limb bud mesenchyme.

Endogenous GLI3 proteins are part of the chromatin complexes interacting with specific parts of the GRS1 region in mouse limb buds

The observed loss of GRS1 activity in Shh-deficient limb buds (Figure 6A) indicated that the GRS1 region could be regulated directly by GLI proteins. Indeed, two GLI binding regions without consensus GLI binding motif (GBR1 and GBR2, [32]) were previously mapped to the 9 kb GRS1 region using transgene-mediated expression of an epitope-tagged GLI3 transgene in combination with chromatin immunoprecipitation (ChIP). To study the potential interaction of endogenous GLI3 proteins with the GRS1 region in its normal genomic context in wild-type limb buds, ChIP analyses using antibodies that detect both the GLI3 activator and repressor protein isoforms were performed [26,47]. GLI3 ChIP using extracts of wild-type and Gli3-deficient limb buds at E11.5 revealed significant enrichment of one specific region within GRS1 by real-time qPCR analysis (Figure 7A). A significant ~4-fold enrichment of an amplicon located within HMCO1 was observed by comparing wild-type to Gli3-deficient limb buds (amplicon “d”; p = 0.008) and to a less conserved region (amplicon “c”); located outside HMCO1 close to a putative GLI binding site; Figure 7A). In contrast, no significant enrichments of the other amplicons located in GRS1 were detected (Figure 7A). As the amplicons located in HMCO2 (Figure 7B) and HMCO3 (Figure 7C) were also not significantly enriched, the GLI3-containing chromatin complexes appear to interact rather specifically with a region within the highly conserved HMCO1 in wild-type limb buds. These results not only corroborates the identity of previously identified GBR1 [32], but indicate that this region mediates the effects of SHH signal transduction on GRS1 activity in the posterior limb bud mesenchyme.

Spatio-temporal Grem1 expression is regulated by the interaction of multiple cis-regulatory regions far downstream of the transcription unit

Previous genetic analysis showed that Grem1 transcription in limb buds is initiated by BMP signalling and up-regulated under the influence of SHH and AER-FGF signaling as limb bud development progresses [11]. Finally, Grem1 expression is terminated concomitantly with the initiation of digit formation by high levels of FGF signal transduction and GLI3-mediated repression in the anterior mouse limb bud mesenchyme [26,40]. This dynamic spatio-temporal regulation of Grem1 expression indicates that the activity of different cis-regulatory elements may change over time. The spatio-temporal activity of the GRS1 transgene (Figure 4B) indicates that it is primarily regulated by the SHH/GREM1/AER-FGF feedback signalling system as limb bud development progresses. In particular, its activity is first apparent when the SHH/GREM1/AER-FGF feedback is already established, and the expected termination does not occur, as β-galactosidase activity remains in the posterior mesenchyme after the endogenous Grem1 transcripts have been down-regulated (Figure 4B, D). These results indicate that FGF-mediated termination of Grem1 expression and the underlying self-regulatory feedback signalling system [40] does not occur by FGF signal transduction impacting the GRS1 region. In addition, neither the 70 kb critical region of the Grem1 landscape (Grem1-LacZ, Figure 2A) nor transgenes derived from this region (Figures. 2, 3, 4 and ref. [32]) are able to accurately recapitulate the entire spatio-temporal distribution of Grem1 transcript in the
limb bud mesenchyme. Therefore, either the interactions among these elements or additional as yet unknown cis-regulatory regions located outside the critical region must also participate in the regulation of Grem1 expression. Indeed, there is circumstantial evidence for the latter, as deletion of the GC-rich Fmn1 coding exon 9 (located ∼200 kb downstream of the Grem1 transcription unit) results in non-complementation with a Grem1 null allele, causing a hypomorphic Grem1 limb skeletal phenotype, and significantly reduced Grem1 expression (ref. [29] and A.Z and R.Z., unpublished data).

Mapping of DNAse I hypersensitive (HS) sites is used to identify active regulatory regions in chromatin (reviewed in ref. [48]). Recently, the ENCODE consortium has done genome-wide mapping of HS sites in several mouse tissues including limb buds (ENCODE Group, University of Washington) [33,34]. Interestingly, this analysis revealed three HS sites within the GBRs of the GRS1 and HMCO2, while no HS sites mapping to HMCO3 were detected in limb buds at E11.5 (Figure 8). The five HS sites mapped by the ENCODE consortium in the Grem1 genomic landscape overlap with the cis-regulatory elements required for Grem1 expression in mouse limb buds (Figure 8). The HS site mapping to HMCO1 overlaps well with the amplicon enriched by GLI3-ChIP analysis (Figure 7A). In fact, all HS sites in the Grem1 genomic landscape overlap either with the GBR regions interacting with GLI3 [26,32] or CTCF binding sites [31]. Analysis of Ctfc-deficient limb buds revealed its requirement for up-regulation and distal-anterior expansion of Grem1 expression [31]. Interestingly, the HS site overlapping with GBR4 [32] is not conserved between mammals and birds (Figure 8 and data not shown). While the early GBR4 activity is comparable to GRS1 in the posterior mesenchyme (Figure 4B), it does not support distal-anterior expansion of the expression during subsequent development [32]. Thus, the interaction of the GBR4 and GRS1 cis-regulatory regions could provide Grem1 expression with the necessary robustness and evolutionary plasticity, as has been postulated for the transcriptional regulation of HoxD genes by multiple interacting regulatory regions during limb bud development [8]. Taken together, these studies establish that the five HS sites in the Grem1 genomic landscape are functionally relevant for Grem1 expression in mouse limb buds, but not all of them are evolutionary highly

---

**Figure 7 Endogenous GLI3 complexes bind to specific HMCO and GBR regions in the Grem1 genomic landscape.** (A) Several amplicons were designed to detect the conserved and functionally relevant regions within the GRS1 region. The exact genomic locations of all amplicons are listed in Additional file 7. The potential interaction of endogenous GLI3 proteins with these critical amplicons was analysed by ChIP-qPCR analysis of wild-type (open bars) and GLI3<sup>−/+</sup> (black) limb buds at E11.5. Two stars indicate significant enrichment of amplicon "d" in the HMCO1 region (p = 0.008). (B) GLI3 ChIP-qPCR analysis of the HMCO2 region. Amplicons "f" and "g" are located inside HMCO2 and the previously identified GBR3 [32]. (C) GLI3 ChIP-qPCR analysis of HMCO3. All values are shown as mean ± SD.
conserved. Indeed, experimental and comparative evolutionary evidence indicates that alterations in the spatio-temporal expression of Grem1, and thereby the activity of the SHH/GREM1/AER-FGF feedback loop, likely contribute significantly to variations in digit numbers and morphologies (reviewed in ref. [49]).

Conclusions
In this study, we identify GRS1 as a distant cis-regulatory region that encompasses the highly conserved HMCO1 region and show that it controls important aspects of Grem1 expression in the posterior and distal-anterior limb bud mesenchyme. GRS1 activity depends critically on SHH signalling and endogenous GLI3 protein complexes interact with specific regions within HMCO1 in wild-type limb buds. In addition, the anterior expansion of GRS1 activity is also regulated by BMP signal transduction. These results together with previous studies reveal the large genomic architecture composed of several distant cis-regulatory regions that control the highly dynamic Grem1 expression in response to signalling inputs from the BMP, SHH and FGF pathways. It is likely that the interactions among these dispersed cis-regulatory regions provide the dynamic Grem1 expression in limb buds with the necessary robustness, but at the same time allow for evolutionary plasticity of its expression [8].

Methods
Ethics Statement
All the experiments were conducted in strict accordance with Swiss law following the 3R principles and the conduct defined by the Basel Declaration. All studies involving mice were classified as grade zero by the Animal Welfare and Ethics Commission of both cantons of Basel and Argovia, which implies minimal suffering.

Mouse strains
The Shh loss-of-function allele [50], the Gli3 extra-toes-1 allele [51], the Grem1 null allele [13] and the GRS1(+/+) alleles (this study) were maintained in an NMRI background, while the Smad4 conditional allele (Smad4floox, [45]) and the Prxl-Cre transgene [46] were maintained in a C57BL/6J background.

BAC modifications
BAC constructs were engineered using ET cloning as previously described [7]. The deletions ΔHMCO1, ΔHMCO2 and ΔHMCO3 were induced in the Grem1-LacZ BAC clone using a Kanamycin resistance selection cassette. All primer sequences are available upon request. See Additional file 8 for the genomic coordinates of HMCO1-3.

βglob-LacZ transgene constructs
An expression cassette was constructed in pKS-Bluescript using LacZeoPin reporter under control of the minimal human β-globin promoter sequences [52]. Various combinations of HMCO sequences were inserted downstream into this cassette. The HMCO1, -2 and -3 core sequences were amplified by PCR directly from mouse genomic DNA (primers available on request). The 9 kb GRS1 region was initially subcloned as a 9.6 kb Bgl2 fragment from BAC RP23-113 H17 (chr2: 113,611,499-113,847,299 (mm10). BacPac Resources, Children's Hospital, Oakland, USA). All constructs were linearized with Ksp1 before microinjection.
Generation of transgenic founder embryos and transgenic strains
BAC and βglob-LacZ transgenic constructs were injected into the pronucleus of fertilized mouse eggs. Several founder embryos for each construct were scored for β-galactosidase activity (Additional file 3). For the GRS1-βglob-LacZ transgene, seven independent transgenic mouse strains were established by crossing founders into the NMRI background. Transgene copy numbers were determined by real-time qPCR [53] using the Bio-Rad CFX96 Real-Time PCR System in combination with the iQ SYBR Green Supermix (Bio-Rad). For each mouse, 20 ng of genomic DNA were analysed in triplicate. The primers to amplify the DBH genomic region were used for normalization and non-transgenic littermate DNA was used as control. The normalized control levels were set to 2 as the primers also amplified both alleles of the endogenous HMC01 regions. The 2^ΔΔCt formula was used to calculate and normalize transgene copy numbers. In addition to three different F1 males, three embryos from the F2 and three embryos from the F3 generation were analysed for the two transgenic strains shown to ascertain stable transmission of the transgenes. Primers sequences are available on request.

Detection of β-galactosidase activity
Embryos were isolated in ice-cold PBS and staged according to somites numbers, then fixed in 1% formaldehyde, 0.2% glutaraldehyde, 0.02% NP40, 0.01% sodium deoxycholate in 1x PBS for 20–30 minutes at 4°C. Subsequently, embryos were washed three times in 1x PBS for 5 minutes at room temperature and incubated overnight at 37°C and in the dark in 1 mg/mL X-Gal, 0.25 mM K3Fe(CN)6, 0.25 mM K4Fe(CN)6, 0.01% NP40, 0.4 mM MgCl2 and 1% sodium deoxycholate to detect β-galactosidase activity, which colours cells blue. To stop the reaction, embryos were washed three times in 1x PBS for 5 minutes each at room temperature.

Chromatin Immunoprecipitation (ChIP)
Forelimbs and hindlimbs of 10 wild-type or Gli3NC/NC embryos at E11.5 were dissected and processed for ChIP as described [26] using a polyclonal anti-GLI3 antibody (#2676, Genentech, [47]). To compute the level of enrichment of a given region, the Ct values of both input and ChIP samples were compared with those of a negative control amplicon located in the mouse β-actin locus [21]. All results (mean ± SD) were obtained by analysing three completely independent experiments per genotype. The significance of all differences was assessed using the two-tailed, non-parametric Mann–Whitney test. The coordinates of the relevant amplification are shown in Additional file 7.

All genomic sequencealignments of the Fnm1-Grem1 locus region were performed using the ECR browser [54] and the ClustalW2 program [55].

Additional files

**Additional file 1:** Figure S1. Conservation of the HMCO core regions. ClustalW2 multiple sequence alignment of the HMCO1, HMCO2 and HMCO3 core regions of mouse (mm10), human (hg19), chimpanzee (panTro3), dog (canFam2), bovine (bosTaur2), opsum (monDom5), chicken (galGal3), lizard (anoCar2), frog (xenTro2), and coelacanth (LatChal1) genomes. The corresponding genomic coordinates are indicated. HMCO1: 73 of 149 nucleotides were conserved in all species; HMCO2: 182/298 conserved nucleotides; HMCO3: 45/137 conserved nucleotides.

**Additional file 2:** Table S1. Genomic coordinates for the sequence comparisons shown in Figure 1A.

**Additional file 3:** Table S2. Analysis of transgenic founder embryo.

**Additional file 4:** Figure S2. Limb bud mesenchymal expression of GRS1-βglob-LacZ transgene in two independent transgenic mouse strains. (A) Expression of the GRS1-βglob-LacZ transgene is restricted to limb buds, with expression initiating earlier in forelimb than hindlimb buds. Note the differences in β-galactosidase activity in GRS1tg10/− (left panel) and GRS1tg10/−/− right panel transgenic mouse embryos. (B) Using real-time qPCR, the transgene copy numbers in both the GRS1tg10/− and GRS1tg10/−/− mouse strains were determined in comparison to wild-type mice (carrying 2 copies of the endogenous GRS1 regions). This analysis revealed that the GRS1tg10/− strain carries 10 copies and GRS1tg10/−/− 2 copies of the transgene, respectively.

**Additional file 5:** Figure S3. Comparison of the LacZ mRNA and β-galactosidase reporter activity in early limb buds of GRS1tg10/− embryos. Distribution of β-galactosidase activity (A) and LacZ transcripts (B) in forelimb buds of GRS1tg10/− embryos at E10.0 and E10.5. Arrowheads point to the posterior expression domains.

**Additional file 6:** Figure S4. Dorso-ventral distribution of β-galactosidase activity in forelimb buds expressing the GRS1tg10/− transgene. The GRS1tg10/− transgene respects the dorsal and ventral restriction of the mesenchymal expression domains. Note that overall expression is significantly lower than in GRS1tg10/− limb buds (compare to Figure S3B). D: dorsal; V: ventral.

**Additional file 7:** Table S3. qPCR amplioncs for GLI3 ChIP analysis.

**Additional file 8:** Table S4. Genomic coordinates of identified Grem1 regulatory regions.

Authors’ contributions
AZ conceived most of the experiments and wrote the manuscript together with RZ. AZ also carried out most of the genetic and transgenic analysis together with FL. FL also prepared the figures for the manuscript. FL, JLR and AZ also carried out most of the genetic and transgenic analysis together with RZ. All authors read and approved the final manuscript.

Acknowledgements
The authors are grateful to A. Offinger and her staff for excellent animal care. L. d'Amato and Christina Torres are thanked for technical assistance. S. Scales (Genentech) provided the GLI3 antibodies. We thank group members for critical input on the manuscript. The ENCODE Group at University of Washington (J. Stamatoyannopoulos and coworkers) are acknowledged for the genome-wide DNAse I hypersensitive site mapping data made available freely through the ENCODE Data Coordination Center at UCSC (release date on March 29, 2012). NM received a EU Marie Curie postdoctoral fellowship. This research is supported by grants of the Swiss National Science Foundation (grants no. 31003A_130803 to RZ and 310000_122558 to AZ), a
EU reintegration grant (PERG-CA-2009-246576 to JLR) and the University of Basel. The authors declare no conflict of interest.

Author details
1 Developmental Genetics, Department of Biomedicine, University of Basel, Mattenstrasse 28, CH-4038, Basel, Switzerland.
2 Transgenic Service, EMBL, Meyerhofstrasse 1, 69120, Heidelberg, Germany.
3 Present address: Max-Planck-Centrum Berlin, Robert Rohrer-Straße 10, 13313, Berlin, Germany.
4 Present address: Université de Strasbourg, Institut de Biologie Cellulaire et Moléculaire, UPR 9022 CNRS, 15 rue René Descartes, 67084, Strasbourg, France.

Received: 28 March 2012 Accepted: 12 July 2012 Published: 13 August 2012

References


doii:10.1186/1471-213X-12-23

Cite this article as: Zuniga et al: Conserved cis-regulatory regions in a large genomic landscape control SHH and BMP-regulated Gremlin1 expression in mouse limb buds. BMC Developmental Biology 2012 12:23.
HMCQ2 conserved core

Mouse
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:113674550

Human
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
15:33111832

Chimpanzee
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
15:29589602

Dog
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
30:5129417

Bovine
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
10:29908339

Opossum
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:88909582

Chicken
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
5:32836402

Lizard
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:3292976

Coelacanth
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
JH127606.1:131370

**********

Mouse
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:113674609

Human
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
15:33111765

Chimpanzee
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
15:29589535

Dog
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
30:5129476

Bovine
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
10:29908398

Opossum
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:88909642

Chicken
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
5:32836461

Lizard
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:3292977

Coelacanth
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
JH127606.1:131429

**********

Mouse
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:113674688

Human
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
15:33111799

Chimpanzee
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
15:29589476

Dog
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
30:5129535

Bovine
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
10:299083457

Opossum
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:88909701

Chicken
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
5:32836520

Lizard
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:32929817

Coelacanth
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
JH127606.1:131488

**********

Mouse
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:113674727

Human
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
15:33111649

Chimpanzee
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
15:29589417

Dog
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
30:5129594

Bovine
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
10:29908516

Opossum
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:88909758

Chicken
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
5:32836579

Lizard
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:32928957

Coelacanth
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
JH127606.1:131546

**********

Mouse
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:113674784

Human
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
15:33111603

Chimpanzee
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
15:29589360

Dog
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
30:5129651

Bovine
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
10:29908573

Opossum
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:88909815

Chicken
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
5:32836636

Lizard
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:32928989

Coelacanth
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
JH127606.1:131604

**********

Mouse
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:113674844

Human
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
15:33111535

Chimpanzee
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
15:29589300

Dog
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
30:5129697

Bovine
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
10:29908633

Opossum
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:88909867

Chicken
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
5:32836696

Lizard
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
1:32928838

Coelacanth
---------------------------------------------
TGCAAAAGAAGACATGCTTGATCGCCACTAGCTGATGAGAGTA
JH127606.1:131664
Table S1

Genomic coordinates for the sequence comparisons shown in Fig. 1A

<table>
<thead>
<tr>
<th>Region</th>
<th>Species</th>
<th>% ID (threshold 74%)</th>
<th>Length (100bp threshold)</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMCO1</td>
<td>Zebrafish</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HMCO1</td>
<td>Xenopus</td>
<td>74.7%</td>
<td>100</td>
<td>xenTro2 scaffold_37:3687354-3687453</td>
</tr>
<tr>
<td>HMCO1</td>
<td>Chicken</td>
<td>80.2%</td>
<td>381</td>
<td>galGal3 chr5:32849567-32849947</td>
</tr>
<tr>
<td>HMCO1</td>
<td>Opossum</td>
<td>81.3%</td>
<td>430</td>
<td>monDom5 chr1:188935147-188935576</td>
</tr>
<tr>
<td>HMCO1</td>
<td>Dog</td>
<td>80.5%</td>
<td>387</td>
<td>canFam2 chr30:5150126-5150512</td>
</tr>
<tr>
<td>HMCO1</td>
<td>Chimpanzee</td>
<td>81.5%</td>
<td>508</td>
<td>panTro3 chr15:29568244-29568751</td>
</tr>
<tr>
<td>HMCO1</td>
<td>Human</td>
<td>81.5%</td>
<td>508</td>
<td>hg19 chr15:33090437-33090944</td>
</tr>
<tr>
<td>HMCO2</td>
<td>Zebrafish</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HMCO2</td>
<td>Xenopus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HMCO2</td>
<td>Chicken</td>
<td>80.4%</td>
<td>377</td>
<td>galGal3 chr5:32836351-32836727</td>
</tr>
<tr>
<td>HMCO2</td>
<td>Opossum</td>
<td>80.1%</td>
<td>380</td>
<td>monDom5 chr1:188909506-188909885</td>
</tr>
<tr>
<td>HMCO2</td>
<td>Dog</td>
<td>84.2%</td>
<td>494</td>
<td>canFam2 chr30:5129238-5129731</td>
</tr>
<tr>
<td>HMCO2</td>
<td>Chimpanzee</td>
<td>84.6%</td>
<td>818</td>
<td>panTro3 chr15:29588916-29589733</td>
</tr>
<tr>
<td>HMCO2</td>
<td>Human</td>
<td>84.5%</td>
<td>818</td>
<td>hg19 chr15:33111145-33111962</td>
</tr>
<tr>
<td>HMCO3</td>
<td>Zebrafish</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HMCO3</td>
<td>Xenopus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HMCO3</td>
<td>Chicken</td>
<td>74%</td>
<td>272</td>
<td>galGal3 chr5:32819753-32820024</td>
</tr>
<tr>
<td>HMCO3</td>
<td>Opossum</td>
<td>77%</td>
<td>331</td>
<td>monDom5 chr1:188837685-188838015</td>
</tr>
<tr>
<td>HMCO3</td>
<td>Dog</td>
<td>78.3%</td>
<td>842</td>
<td>canFam2 chr30:5078806-5079647</td>
</tr>
<tr>
<td>HMCO3</td>
<td>Chimpanzee</td>
<td>80.7%</td>
<td>940</td>
<td>panTro3 chr15:29642824-29643763</td>
</tr>
<tr>
<td>HMCO3</td>
<td>Human</td>
<td>80.1%</td>
<td>937</td>
<td>hg19 chr15:33164854-33165790</td>
</tr>
</tbody>
</table>
**Table S2**

Analysis of transgenic founder embryo

<table>
<thead>
<tr>
<th>Transgene</th>
<th>Embryos *</th>
<th>Forelimb bud expression</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grem1-LacZ</td>
<td>3</td>
<td>3 posterior</td>
<td>Fig.1B</td>
</tr>
<tr>
<td>Grem1-LacZ ΔHMCO1</td>
<td>7</td>
<td>1 strong</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 very weak</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 no expression</td>
<td>Fig.1C</td>
</tr>
<tr>
<td>Grem1-LacZ ΔHMCO2</td>
<td>3</td>
<td>2 weak posterior</td>
<td>Fig.1D</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 no expression</td>
<td></td>
</tr>
<tr>
<td>Grem1-LacZ ΔHMCO3</td>
<td>5</td>
<td>3 posterior</td>
<td>Fig.1E</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 no expression</td>
<td></td>
</tr>
<tr>
<td>HMCO123-βglob-LacZ</td>
<td>3</td>
<td>3 anterior</td>
<td>Fig.2A</td>
</tr>
<tr>
<td>GRS1-βglob-LacZ</td>
<td>2</td>
<td>2 posterior</td>
<td>Fig.2B</td>
</tr>
</tbody>
</table>

* Founder embryos with β-galactosidase activity anywhere in limb buds and/or trunk

**Fig.S2**

A  

![GRS1<sup>tg10/4</sup>](E11.5)  ![GRS1<sup>tg2/4</sup>](E11.5)

B  

![Bar chart showing protein levels](chart.png)
Fig. S3

A  E10.0  E10.5
β-Gal

B
LacZ

GRS1^10/4

Fig. S4

E11.25  Ventral  E11.75  Ventral
GRS1^2+/+

D  V  D  V  D  V
### Table S3

#### qPCR amplicons for GLI3 ChIP analysis

<table>
<thead>
<tr>
<th>GLI3 ChIP amplicon</th>
<th>Coordinates (mm10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>chr2:113691635-113691714</td>
</tr>
<tr>
<td>b</td>
<td>chr2:113692244-113692320</td>
</tr>
<tr>
<td>c</td>
<td>chr2:113693100-113693179</td>
</tr>
<tr>
<td>d</td>
<td>chr2:113693471-113693578</td>
</tr>
<tr>
<td>e</td>
<td>chr2:113697427-113697518</td>
</tr>
<tr>
<td>f</td>
<td>chr2:113674650-113674732</td>
</tr>
<tr>
<td>g</td>
<td>chr2:113675141-113675273</td>
</tr>
<tr>
<td>h</td>
<td>chr2:113618703-113618814</td>
</tr>
<tr>
<td>i</td>
<td>chr2:113619171-113619319</td>
</tr>
</tbody>
</table>

### Table S4

#### Genomic coordinates of identified *Grem1* regulatory regions

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Length</th>
<th>Coordinates (mm10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMCO1</td>
<td>520 bp</td>
<td>chr2:113693204-113693723</td>
</tr>
<tr>
<td>HMCO2</td>
<td>1279 bp</td>
<td>chr2:113674096-113675374</td>
</tr>
<tr>
<td>HMCO3</td>
<td>924 bp</td>
<td>chr2:113618678-113619601</td>
</tr>
<tr>
<td><em>GRS1</em>-ßglob-LacZ transgene</td>
<td>9'831 bp</td>
<td>chr2:113689791-113699621</td>
</tr>
<tr>
<td>GBR1 (HMCO1, ChIP amplicons c-d) [32]</td>
<td>439 bp</td>
<td>chr2:113693005-113693989</td>
</tr>
<tr>
<td>GBR2 (ChIP amplicons a-b) [32]</td>
<td>700 bp</td>
<td>chr2:113691653-113692352</td>
</tr>
<tr>
<td>GBR3 (HMCO2, ChIP amplicons f-g) [32]</td>
<td>985 bp</td>
<td>chr2:113674455-113675313</td>
</tr>
<tr>
<td>GBR4 (with GLI binding motif) [32]</td>
<td>439 bp</td>
<td>chr2:113640843-113641281</td>
</tr>
<tr>
<td>CTCF binding region [31]</td>
<td>468 bp</td>
<td>chr2:113648107-113648574</td>
</tr>
</tbody>
</table>
6.1.3. The *Grem1* cis-regulatory landscape integrates several signaling inputs from multiple CRMs

In this study, we showed that a 9kb region located in the *cis*-regulatory landscape of *Grem1* regulates aspects of the posterior activation and distal anterior expansion of *Grem1* expression during progression of limb bud outgrowth. This *Gremlin1 Regulatory Sequence 1* (*GRS1*) is activated by the Sonic hedgehog (SHH) pathway and interacts with GLI3 containing protein complexes. Furthermore, *GRS1* expansion towards the distal anterior part of the limb bud is dependent on BMP signaling, as seen by conditionally inactivating the BMP signal transducer SMAD4 in limb bud mesenchyme. Several other studies identified additional CRMs located outside of *GRS1* that contribute to *Grem1* expression and interact with GLI3 or CTCF chromatin complexes (Vokes et al., 2008; Soshnikova et al., 2010). Together, these data reveal that the large *cis*-regulatory landscape of *Grem1* is an archipelago composed of multiple CRMs that integrate signaling inputs from various pathways to regulate the robust and dynamic expression of *Grem1* in limb buds.

As limb bud outgrowth advances, several factors eventually contribute to the termination of *Grem1* expression and feedback signaling: 1) increasing high levels of FGF signaling from the AER repress *Grem1* in the distal mesenchyme (Verheyden and Sun, 2008); 2) GLI3 represses *Grem1* in the anterior limb bud mesenchyme, which enables the rise of BMP activity and BMP-dependent initiation of chondrogenesis (Lopez-Rios et al., 2012); 3) the induction of TBX2 by BMP signaling represses *Grem1* by direct interaction of this transcription factor with the GBR4 element in Shh descendant cells in the posterior limb bud (Farin et al., 2013); and 4) Shh descendant cells become unable to express *Grem1* in the posterior limb bud (Scherz et al., 2004). Our study shows that *GRS1* activity is not terminated with the same kinetics as endogenous *Grem1* expression. In particular, a strong LacZ expression domain remains in the posterior mesenchyme. This domain becomes spatially separated from the distal anterior domain at E12.0. This suggests that the CRM(s) that mediates FGF dependent repression of *Grem1* are not part of *GRS1*. In contrast, the anterior expansion and upregulation of *GRS1*-LacZ domain in *Gli3*-
deficient limb buds show that \textit{GRS1} encodes CRMs involved in GLI3R-mediated repression of \textit{Grem1} expression. The GBR4 CRM interacts both with the GLI3 and TBX2 transcriptional regulators and is able to direct \textit{LacZ} expression in the posterior but not anterior mesenchyme (Lopez-Rios et al., 2012; Farin et al., 2013; Vokes et al., 2008). This region, located \~50kb upstream of \textit{GRS1}, was called GLI Responsive Element 1 (GRE1) in a recent study (Li et al., 2014). GRE1 is activated by SHH and bound by GLI3A chromatin complexes in the posterior limb bud, and its activity is not affected by inhibiting FGF signaling. In addition to its enhancer activity in the posterior limb bud, GRE1 participates in GLI3R-mediated silencing of \textit{Grem1} in the anterior mesenchyme (Li et al., 2014). GRE1 and \textit{GRS1} are both directly regulated by SHH/GLI3 and may thus be functionally redundant and/or are likely required cooperatively for regulating the dynamic expression of \textit{Grem1} in the limb bud mesenchyme.

The activation of \textit{Grem1} expression in the posterior limb bud mesenchyme requires BMP signaling and the HOX transcriptional regulators upstream of SHH signaling. SHH signaling is subsequently required to upregulate and distal-anteriorly expand \textit{Grem1} expression (Zuniga et al., 1999; Panman et al., 2006; Nissim et al., 2006; Benazet et al., 2009; Sheth et al., 2013). Interestingly, the expression of the \textit{GRS1} reporter requires SHH signaling and β-galactosidase activity is first detected only after initiation of \textit{Grem1} endogenous expression in the limb bud mesenchyme. Thus, the BMP Responsive Elements (BRE) and HOX CRMs are not encoded by the \textit{GRS1} and remain to be identified.

In \textit{Smad4}^{\Delta\DeltaM} limb buds, the initial expression of \textit{Grem1} in the posterior mesenchyme is reduced, fails to expands anteriorly, and is lost in \textit{Shh}^{\Delta\Delta};\textit{Smad4}^{\Delta\DeltaM} forelimb buds (Benazet et al., 2012). Thus, \textit{Grem1} initiation depends on positive regulation by both SHH and SMAD4-mediated BMP signal transduction. Moreover, the absence of anterior expansion of \textit{GRS1} expression domain in \textit{Smad4}^{\Delta\DeltaM} limb buds suggests that \textit{GRS1} contains BREs controlling SMAD4-dependant expansion of \textit{Grem1} expression after its initial activation (this study and (Benazet et al., 2012). The distal anterior expansion of \textit{Grem1} expression is also regulated by CTCF and HOX transcriptional regulators (Soshnikova et al., 2010; Sheth et al., 2013).
To summarize the integration of these various signaling inputs by multiple CRMs located in *Grem1* cis-regulatory landscape, we propose the following kinetics for the regulation of *Grem1* expression: 1) An early expression domain is initiated in the posterior limb bud mesenchyme by BMP signaling and HOX transcriptional regulators through CRMs that need to be identified. This posterior domain is maintained by SHH/GLI3A inputs acting on the GLI-binding regions in GRS1 and GRE1 CRMs. 2) The dynamic expansion of *Grem1* expression in the distal anterior mesenchyme depends on BMP/SMAD4 signal transduction impacting on GRS1 and on CRMs regulated by CTCF and HOX complexes. This distal-anterior domain is likely silenced by GLI3R impacting on GRE1 CRM. 3) The termination of the posterior domain may in part be mediated by TBX2 interacting with GRE1. 4) In contrast, termination of the distal-anterior domain likely depends on GLI3R-mediated repression via the GRS1 and GRE1 CRMs and unknown FGF-responsive elements. DNase HS sites mark all the known *Grem1* CRMs. Each of them likely encodes binding sites for the different trans-acting factors that fulfill different activities depending on the specific spatial and temporal developmental context.

GLI3 direct interactions with relevant regions in GRS1 and GRE1 have been well characterized (this study and (Li et al., 2014)). The next major challenge is to understand how *Grem1* cis-regulatory landscape integrates BMP signaling inputs. We are attempting this by identifying the CRMs that interact with SMAD4-containing chromatin complexes.

6.1.4. Identification of SMAD4 target CRMs during limb bud development: the need for a new tool

In addition to the role of SMAD4 in the regulation of GRS1 activity (Zuniga et al., 2012), we showed that Smad4 is crucial for the initiation of digits primordia and chondrogenesis (Benazet et al., 2012). Furthermore, limb bud ectodermal Smad4 is required for AER initiation and for AER-FGF mediated termination of epithelial-mesenchymal signaling (Benazet and Zeller, 2013). Despite the inactivation of Smad4 in limb buds affects the expression of a large number of downstream genes, there is no evidence that these genes are direct SMAD4 transcriptional targets.
We thus decided to initiate a genome-wide analysis of the SMAD4 chromatin complexes in limb buds using ChIP-Seq to: 1) identify SMAD4 interacting CRMs in *Grem1* regulatory landscape and 2) to get further insights into the direct BMP targets that are involved in chondrogenic differentiation.

Several SMAD4 ChIP-Seq studies using cell lines in combination with different types of antibodies have been published (Koinuma et al., 2009; Qin et al., 2009; Fei et al., 2010; Kennedy et al., 2011; Morikawa et al., 2011). Therefore, we tested supposedly ChIP-grade anti-SMAD4 antibodies on a known BRE located in the promoter of the gene *Id1* using wild-type embryonic limb buds, but no SMAD4 chromatin complexes were detected (data not shown). Therefore, we decided to introduce a 3xFLAG epitope tag in the endogenous SMAD4 protein as we have successfully done for HAND2 (see Chapter 6.2).

### 6.1.5. Generation of Smad4^3xFLAG^ mice by homologous recombination

**Targeting and testing ES cells for SMAD4^3xFLAG^ expression**

Two previous attempts were made to produce a mouse line expressing a SMAD4^3xFLAG^ protein using cassette exchange approaches. However, the resulting *Smad4^3xF^dRMCE^* and *Smad4^3xF^-R26^* alleles were expressed at levels too low for genome-wide ChIP-Seq analysis (see Appendix 12.1). Therefore, we decided to introduce the 3xFLAG epitope tag in the endogenous SMAD4 protein using conventional homologous recombination. A targeting vector consisting of two homology arms flanking the 3'-end of *Smad4* coding sequence, in which the 3xFLAG was inserted in frame between the exon 12 and the 3'UTR, was constructed (Fig.11A). A floxed Neomycin selection cassette was inserted downstream of *Smad4* locus in the 3' homology arm. This targeting vector was linearized and electroporated in R1 embryonic stem (ES) cells to generate the *Smad4^2xF-HR^* allele (Fig.11A, thereafter referred as *S4^2xF-HR^*). 662 ES cells clones were picked (together with A.Zuniga) and 424 of them screened for correct recombination by Southern Blot using a probe located outside of the 5' homology arm (Fig.11A and Fig.11B, top panel).
Figure 11. Expression of the Smad4<sup>3xFLAG-HR</sup> allele in mouse ES cells

(A) Homologous recombination strategy. Top scheme: linearized targeting vector containing two homology arms flanking the 3'-end of Smad4 coding sequence. A 3xFLAG epitope tag was fused in frame between coding exon 12 and the 3'UTR. A floxed Neo cassette is inserted downstream of Smad4 locus in the 3' homology arm. Homologous recombination generates the Smad4<sup>3xFLAG-HR</sup> allele (bottom scheme). The restriction sites used for Southern Blot analysis are indicated in italic, and the Southern Blot probes are shown as green boxes. (B) Southern Blot analysis of seven ES cells clones. Wild-type R1 ES cells were used as control. Left panel: Genomic DNA digested with BamHI and blotted with the 5' probe shown in (A). The sizes of the expected wild-type and 5'-recombined bands are indicated on the left. Middle panel: 3' probe blot following BamHI digestion. Right panel: blot probed for the Neo cassette following BamHI digestion. (C) Western Blot detection of the SMAD4<sup>3xFLAG</sup> protein in extracts of several different S4<sup>3xF-HR</sup> ES cells clones. Wild-type R1 ES cells and the negative clone 22B5 were used as controls, while S4<sup>3xF-dRMCE</sup> ES cells were used as a control for a low SMAD4<sup>3xFLAG</sup> expression (see Appendix 12.1). Clones 22B6 and 22D8 express the highest levels of SMAD4 (~63kDa). ACTIN (42kDa) was used as a loading control and is present in similar amounts in all samples. (D) qPCR on Smad4 cDNA from wild-type, S4<sup>3xF-dRMCE</sup> and several clones of the S4<sup>3xF-HR</sup> ES cells. Intermediate expression in clone 22E4 indicates that this is a mixed clone. (E) Immunofluorescence detection of the SMAD4<sup>3xFLAG</sup> protein in undifferentiated wild-type or S4<sup>3xF-HR</sup> ES cells. Green: SMAD4<sup>3xFLAG</sup> protein. Blue: nuclei stained by Hoechst. Scale bar: 50µm
Positively recombined ES cells clones were further screened for correct 3’ recombination (Fig.11A and Fig.11B, middle panel). Finally, single insertion of the Neomycin cassette was assessed to exclude ES cell clones with random integration of the targeting construct (Fig.11A and Fig.11B, bottom panel). In total, 29 of the 424 ES cell clones carried the $S_4^{3xF-HR}$ allele (6.8%, similar to the expected 3-6% of correct targeting on Smad4 locus (Chu et al., 2004; Takaku et al., 1998)). In addition, Western Blot analysis detected the intact SMAD4$^{3xFLAG}$ protein in ES cells (Fig.11C). Western Blot analysis also allowed identifying mixed ES cell clones expressing the tagged protein at lower levels (Fig.11C, compare ES cell clones 22B6, 22D8 to 22E4. Clone 22B5 is a negative control). Finally, the levels of the SMAD4$^{3xFLAG}$ protein in the different ES cells clones carrying the $S_4^{3xF-HR}$ allele was much higher compared to those in ES cells carrying the damaged $S_4^{3xF-dRMCE}$ allele (Fig.11C and Appendix 12.1). These results and the mixed nature of some ES cell clones, such as 22E4, were confirmed by qPCR analysis (Fig.11D). Immunocytochemistry showed that the 3xFLAG epitope tag allows detection of endogenous SMAD4 proteins both in the cytoplasm and nuclei of ES cell clones (Fig.11E). Three of the tested clones were karyotyped and did not show any obvious chromosomal abnormalities (data not shown).

**Smad4$^{3xFLAG}$ allele: a new tool for future ChIP-Seq studies**

Eight of the positive $S_4^{+/3xF-HR}$ ES cells clones were injected in C57BL/6 blastocysts at the Transgenic Mouse Core Facility (TMCF, Basel, see Fig.12A). Chimeric males were obtained from three independent clones and mated to CMV-Cre females (C57BL/6 background) to delete the floxed Neo selection cassette. Four chimeras transmitted the $S_4^{3xF-HR}$ allele to 67-100% of their offspring (Fig.12B). Germline transmission was assessed by coat color (Fig.12A) and confirmed by PCR genotyping. CMV-Cre-mediated deletion of the floxed Neo selection cassette resulted in the $S_4^{3xF-HR\triangle Neo}$ allele (Fig.12C). Specific primers were used to discriminate between the $S_4^{wt}$, $S_4^{3xF-HR}$ and $S_4^{3xF-HR\triangle Neo}$ alleles (Fig.12C, arrows, and Fig.12D). Mice homozygous for the $S_4^{3xF-HR\triangle Neo}$ allele are viable and obtained in the expected Mendelian ratio from two independent ES cell clones (data not shown).
Figure 12. Establishment of Smad43x/3x mouse strain

(A) Schematic representation of the generation of chimeras and F1 offspring from S43xF-HR/+ ES cells (adapted from Pettitt et al., 2009). (B) Table summarizing the percentage of germline transmission observed for each of the chimeras derived from S43xF-HR/+ ES cells clones (22B6, 31A6 and 31E1). (C) Schematic representation of the S4wt (top), S43xF-HR (middle) and S43xF-HR2Neo (bottom) alleles. The S43xF-HR2Neo allele is obtained by CMV-Cre mediated recombination of the loxP sites in the S43xF-HR allele. Arrows indicate primers used for the genotyping of these alleles. (D) PCR genotyping of all three considered alleles. The sizes of the expected bands are indicated on the right. (E) Detection of the SMAD43xFLAG protein in limb buds of S43xF/3xF embryos (E11.75). Protein extracts from wild-type limb buds were used as control. (F) RT-qPCR analysis of E11.75 limb buds from S4+/+ and
$S4^{3xF/3xF}$ embryos. Normalized $Smad4$ transcript level using the junction between exon 8-9 (left panel), exon 12-3'UTR junction (middle panel) and FLAG-3'UTR (right panel) for amplification. (G) The anti-FLAG M2 antibody was used to detect SMAD4$^{3xFLAG}$ proteins in frozen sections of limb buds. From left to right: wild-type forelimb (E10.5), $S4^{3xF/3xF}$ forelimb (E10.5) and $S4^{3xF/3xF}$ forelimb (E11.5). Scale bar: 100µm (50µm in high magnification).

The $Smad4^{3xF}$ strain was established by breeding the $Smad4^{3xF-HR\Delta Neo}$ allele from the ES cell clone 31A6 into a NMRI background (to obtain the most embryos possible for future analysis). The SMAD4$^{3xFLAG}$ protein is detected with high sensitivity and specificity in limb bud tissues from $Smad4^{3xF/+}$ (not shown) and $Smad4^{3xF/3xF}$ embryos (E11.75, Fig.12E). The levels of $Smad4$ transcripts in $Smad4^{3xF/+}$ and $Smad4^{3xF/3xF}$ limb buds were similar to wild-type controls (E11.75, Fig.12F, left panel). In heterozygous embryos, the $S4^{wt}$ allele was expressed as expected at ~50% in comparison to $Smad4^{+/+}$ embryos (Fig.12F, middle panel). Conversely, in heterozygous embryos the $S4^{3xF}$ allele was expressed at ~50% in comparison to $Smad4^{3xF/3xF}$ embryos (Fig.12F, right panel). Altogether, these data establish that the $Smad4^{3xF-HR\Delta Neo}$ allele is expressed like the wild-type allele, which makes this allele a promising tool to study SMAD4 chromatin complexes and protein interactions in vivo.

Finally, we performed immunohistochemistry on frozen limb bud sections to assess SMAD4 in vivo localization (together with J.Leclercq). As seen by whole-mount antibody staining (Benazet et al., 2012), SMAD4 is evenly distributed throughout limb bud mesenchyme at E10.5 and E11.5 (Fig.12G). In most of the limb bud cells, SMAD4 proteins are detected both in the cytoplasm and nuclei (Fig.12G, high magnification).

We will now use the $Smad4^{3xF-HR\Delta Neo}$ allele to determine the genome-wide range of SMAD4 target sequences using ChIP-Seq analysis. To this aim, we have collected limb buds from around 600x E10.75-E11.0 embryos (37-42 somites). The chromatin of these limb buds was crosslinked and snap frozen and is now being processed for ChIP-Seq analysis. The range of SMAD4 interacting sequences will provide us interesting insights into the transcriptional networks regulated by SMAD4 during mouse limb bud development. In addition, identifying the CRMs that interact with SMAD4 chromatin complexes in the $Grem1$ cis-regulatory landscape will be of particular interest to understand how this complex landscape integrates various signaling inputs.
6.2. From \textit{trans}-acting factor to \textit{cis}-regulatory modules: identification of the direct transcriptional targets of HAND2 during heart morphogenesis

6.2.1. The \textit{Hand2}^{3\times\text{FLAG}} allele is a new genetic tool to study HAND2 distribution during heart development

By taking advantage of the straightforward and high efficient strategy to target mouse ES cells using Dual Recombinase-Mediated Cassette Exchange (dRMCE, (Osterwalder et al., 2010a), a mouse line with a 3xFLAG epitope-tag inserted into the endogenous \textit{Hand2} locus was generated by M.Osterwalder. The generation and characterization of this \textit{Hand2}^{3\times\text{FLAG}} allele is described in (Osterwalder et al., in press). Briefly, the 3xFLAG epitope tag was fused in frame the N-terminal part of the \textit{Hand2} coding sequence and used to replace the corresponding region in conditional \textit{Hand2} R1 ES cells by dRMCE (Galli et al., 2010) (Fig.13A). The insertion of this epitope tag does not affect the expression and the function of the \textit{Hand2} gene products, and the tagged HAND2 proteins are readily detected in all \textit{Hand2} expressing tissues (heart, limb buds and branchial arches) with high specificity and sensitivity, both by Western Blot and by immunohistochemistry (Osterwalder et al., in press). Homozygous \textit{Hand2}^{3xF/3xF} develop normally and do not exhibit any phenotype or sign of embryonic lethality as observed in \textit{Hand2}-deficient embryos (Srivastava et al., 1997). The \textit{Hand2}^{3\times\text{FLAG}} allele is thus a new fantastic tool to study the endogenous spatio-temporal distribution of HAND2 proteins and identify its downstream functional interactions during heart, limb buds and branchial arch morphogenesis.

During the onset of heart development, \textit{Hand2} transcripts are mainly expressed in the myocardium of second heart field (SHF) derived structures such as the outflow tract (OFT) and right ventricle (RV), the endocardium of various cardiac compartments, the proepicardial organ (PEO) and derived epicardium, and structures derived from cardiac neural crest cells (cNCC) (Srivastava, 1995; Biben and Harvey, 1997; Srivastava et al., 1997; Barnes et al., 2011). In contrast to these studies, relying on whole-mount or section \textit{in situ}-hybridization, we used
immunohistochemistry to detect the HAND23xFLAG proteins with the M2 anti-FLAG antibody (Sigma). This allowed us to determine the HAND2 proteins distribution with higher single-cell resolution than using conventional in-situ hybridization techniques and allowed us to co-localize HAND2 with other proteins (see below).

Figure 13. The Hand23xFLAG allele is expressed throughout cardiac development
(A) Schematic illustration of the Hand23xFLAG allele (Osterwalder et al., in press). The 3xFLAG epitope tag (yellow box) is fused in frame downstream of the endogenous Hand2 ATG translation start site. (B-F) Distribution of the endogenous HAND23xFLAG protein during heart development at (B) E8.0, (C) E8.5, (D-E) E9.5 and (F) E11.5. Enlargement of the boxed areas are shown in (B'-F'). (B-E) Transverse paraffin sections and (F) sagittal frozen section from the left side of the embryo. The section in (D) is localized more cranially than the section in (E). Scale bars: 50µm. Open arrowhead: splanchnic pharyngeal mesoderm, white arrowhead: myocardium, white arrow: endocardium, white asterisk: mesenchymal cells of the cardiac cushions, open arrow: epicardium. avc: atrioventricular canal, ba: branchial arches, cc: cardiac cushions, la: left atrium, lv: left ventricle, oft: outflow tract, ra: right atrium, rv: right ventricle, v: common ventricular chamber.
We initially wanted to establish HAND2 proteins distribution during key stages of early cardiac development. At E8.0, HAND2\(^{3x\text{FLAG}}\) proteins are detected in the entire primary heart tube in both the myocardial and endocardial layers and in the adjacent splanchnic pharyngeal mesoderm (Fig.13B-B'). During cardiac looping (E8.5), HAND2\(^{3x\text{FLAG}}\) remains in the endocardium of the two developing ventricles. In contrast, HAND2\(^{3x\text{FLAG}}\) proteins remain in the myocardium of the right ventricle while they are absent from the myocardium of the left ventricle (LV) (Fig.13C-C'). At E9.5, HAND2\(^{3x\text{FLAG}}\) proteins are detected in the myocardium and endocardium in a region extending from the outflow tract to the right ventricle, (Fig.13D-D') and in the endocardial cells of the atrioventricular canal (AVC, Fig.13E-E'). At E11.5, HAND2\(^{3x\text{FLAG}}\) proteins are present not only in the endocardium but also in the cells of the cardiac cushions (CC), formed by endothelial-mesenchymal transition (EndMT) in both OFT and AVC (Fig.13F-F'). HAND2\(^{3x\text{FLAG}}\) proteins are also detected in the epicardium at this stage. HAND2\(^{3x\text{FLAG}}\) proteins are nuclear, which agrees with their function as a transcription factor. In summary, our analysis of the cardiac HAND2\(^{3x\text{FLAG}}\) distribution confirms and refines its previously described expression pattern to single cell resolution during early heart development.

6.2.2. ChIP-Seq analysis using pooled Hand2 expressing tissues identifies the genome-wide HAND2 target regions in association with tissue-specific gene functions

Over the past two decades, numerous studies have been published aiming to decipher the functions of HAND2 during cardiogenesis, and genes downregulated during altered heart development in Hand2-deficient embryos had been identified. These putative HAND2 target genes include Gata4, Irx4, Nppa, Plxna2 and Cx40 (Srivastava et al., 1997; Bruneau et al., 2000; Thattaliyath et al., 2002; Morikawa and Cserjesi, 2008; McFadden et al., 2005). However, none of these genes has been confirmed as direct endogenous target of HAND2, and the gene regulatory networks controlled by HAND2 during normal embryonic development remained unknown (Vincentz et al., 2011). Therefore, we used the Hand2\(^{3x\text{FLAG}}\) allele to identify the direct transcriptional targets of HAND2 during embryonic development.
For this purpose, M.Osterwalder and colleagues performed a chromatin immunoprecipitation in combination with deep sequencing (ChIP-Seq) using the \textit{Hand2}^{3x\text{FLAG}} allele and the M2 anti-FLAG antibody. This analysis was performed using pooled \textit{Hand2} expressing embryonic tissues (heart, limb buds and branchial arches, see Fig.14A) from a pool of approximately 150 \textit{Hand2}^{3x\text{FLAG}+/+} and \textit{Hand2}^{3x\text{FLAG}/3x\text{FLAG}} embryos at E10.5. \textit{Hand2} is expressed at high levels in all these tissues, with the expression in cardiac tissues being very similar as observed at E11.5 (Fig.13F-F’ and data not shown). Expressing tissues were pooled to generate the significant amounts of chromatin immunoprecipitates required to assemble the genomic library for deep-sequencing. In addition, this allows the identification of both tissue-specific and shared candidate \textit{HAND2} targets genes.

The \textit{HAND2}^{3x\text{FLAG}} ChIP sample and the input control sample were sequenced using an Illumina Genome Analyzer in the Laboratory of Quantitative Genomics (D-BSSE, ETHZ, Basel). The resulting sequence tags were mapped to the reference mouse genome (NCBI37/mm9) and enrichment of ChIP versus input for each sequence tag were computed by R.Ivanek using MACS-based peak calling analysis (Model-based Analysis of ChIP-Seq data), see (Zhang et al., 2008b). The \textit{HAND2}^{3x\text{FLAG}} ChIP-Seq peaks were visualized using the UCSC browser (Karolchik et al., 2014). A wild-type control ChIP sample was also processed and sequenced in parallel to the \textit{HAND2}^{3x\text{FLAG}} sample to identify non-specific antibody-related false hits. Such false hits were excluded from further analysis, which reveals a major advantage of using epitope-tagged transcription factors for ChIP-Seq experiments. Analysis of the input was a critical control to confirm that the signal enrichment of the sequence tags in the \textit{HAND2}^{3x\text{FLAG}} ChIP is representative of real transcription factor binding and not due to an experimental or computational bias (Rozowsky et al., 2009). To restrain the \textit{HAND2}^{3x\text{FLAG}} ChIP-Seq dataset to the statistically relevant enriched genomic regions, only the 2000 highest enriched peaks according to MACS analysis (Top 2000, Fig.14B, left panel, green) were considered for further analysis using GREAT (Genomic Regions Enrichment Annotations Tool), (McLean et al., 2010). Most of the target regions mapped between 10-100kb from known transcriptional start sites (TSS), and encoded \textit{bona-fide} E-box consensus CRTCTGKHTT sequences that were identified using the Homer software (Heinz et al., 2010).
Figure 14. ChIP-Seq using HAND2\textsuperscript{3xFLAG} embryos reveals HAND2 candidate target genes in developing heart and limb buds

(A) Hand2\textsuperscript{3xFLAG} embryos were used for ChIP-Seq analysis from pooled dissected Hand2 expressing tissues at E10.5 (white arrowhead: heart. Black arrow: forelimb bud. Black arrowhead: branchial arches). For details see (Osterwalder et al., in press). (B) The top 2000 enriched genomic regions (green, left diagram) and their 2402 associated genes (green, right diagram) were intersected with regions bound by the transcriptional co-activator p300 in developing hearts (E11.5, red) and limb buds (E11.5, blue). (C) Gene Ontology analysis shows that genes associated with regions bound by both HAND2 and p300 are involved in cardiac (red, 240 genes) or limb bud (blue, 349 genes) development, respectively.

This consensus sequence matches the CATCTG HAND2 binding site identified previously (Dai and Cserjesi, 2002). These whole-genome data are described in (Osterwalder et al., in press). The 2402 genes associated to the Top 2000 enriched genomic regions represent HAND2 candidate target genes in developing hearts, limb buds and branchial arches. To identify potential cardiac targets of HAND2, we intersected the Top 2000 peaks with regions interacting with the transcriptional co-activator p300 in developing hearts (Blow et al., 2010) (Fig.14B, red), and limb buds (Visel et al., 2009a) (Fig.14B, blue) using the UCSC browser. p300 is a critical histone acetyltransferase that is part of the majority of enhancer-associated protein
complexes (Eckner et al., 1994; Yao et al., 1998; Visel et al., 2009a). Tissue-specific p300 ChIP-Seq datasets (Blow et al., 2010; Visel et al., 2009a) thus provide an additional validation of the range of HAND2 target sequences that act as potential cis-regulatory modules (CRMs) in heart and/or limb bud tissues development. GREAT analysis further identified the genes in proximity to the Top 2000 regions bound by both HAND2 and p300. The weak overlap suggests that most of the regions interacting with HAND2 are not associated with active genes, in agreement with the function of HAND2 as a transcriptional repressor (Fig.14B). The biological relevance of these genes was determined by Gene Ontology (GO) analysis (Ashburner et al., 2000) (Fig.14C). In each tissue, HAND2 interacting regions were associated with genes involved in development (e.g. GO “embryo development” or “embryonic morphogenesis”) and in DNA binding (e.g. GO “regulation of DNA binding”, not shown), relevant for HAND2 function as a transcription factor. Moreover, tissue-specific HAND2 p300 bound regions were associated with either cardiac (Fig.14C, red) or limb bud development GO terms, respectively (Fig.14C, blue). Notably, cardiogenesis related GO terms were representative of heart compartments with known Hand2 functions (e.g. GO “cardiac neural crest cell development involved in outflow tract morphogenesis” and GO “cardiac right ventricle morphogenesis”). In addition to Hand2 itself, several genes relevant to heart development were identified among the 240 genes associated with regions interacting with both HAND2 and p300. Table 1 summarizes the main functions of some of these genes.

Table 1. Genes involved in cardiac development associated with a genomic region interacting with HAND2 and p300

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Putative HAND2 target</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hes1</td>
<td>OFT development</td>
<td>-</td>
<td>(Rochais et al., 2009a)</td>
</tr>
<tr>
<td>Sox4</td>
<td>Ventricular septation and OFT formation</td>
<td>-</td>
<td>(Schilham et al., 1996)</td>
</tr>
<tr>
<td>Tgfβ2</td>
<td>Inducing role in the EndMT process resulting in cardiac cushion formation, and cardiac valves remodeling</td>
<td>-</td>
<td>(Ma et al., 2005; Rivera-Feliciano and Tabin, 2006; Shirai et al., 2009; Azhar et al., 2011)</td>
</tr>
<tr>
<td>Tbx2</td>
<td>Involved at several steps of cardiogenesis (Chambers differentiation, AVC patterning, formation of cardiac cushions, cardiac valves, and conduction system)</td>
<td>-</td>
<td>(Hariri et al., 2012; Greulich et al., 2011; Hoogaars et al., 2007a; Stennard et al., 2005; Plageman and Yutzey, 2005)</td>
</tr>
<tr>
<td>Tbx3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tbx20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plxna2</td>
<td>Cardiac neural crest cells migration</td>
<td>Yes</td>
<td>(Brown et al., 2001; Morikawa and Cserjesi, 2008)</td>
</tr>
<tr>
<td>Cx40</td>
<td>Heart-specific gap-junction coding gene</td>
<td>Yes</td>
<td>(Delorme et al., 1995; Kirchhoff et al., 2000; McFadden et al., 2005)</td>
</tr>
<tr>
<td>(Gja5)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In contrast, the cis-regulatory landscapes of the putative HAND2 targets Gata4 and Nppa contain no associated ChIP-Seq peak. We used a Hi-C dataset representing the range of potential distant chromatin-chromatin interactions (Dixon et al., 2012), but could not detect any distant HAND2\(^{3x\text{FLAG}}\) ChIP-Seq peak associated with Gata4 or Nppa, suggesting that these two genes are indirect targets of HAND2 (data not shown).

Crossing the HAND2\(^{3x\text{FLAG}}\) ChIP-Seq data with the p300 tissue-specific genome-wide targets datasets thus identified tissue-specific targets and likely biological functions for Hand2. Therefore, this Gene Ontology analysis validates pooling embryonic tissues for ChIP-Seq analysis. In addition, it narrows the range of potential HAND2 targets genes involved in each of the relevant developing tissues. A complete and functional analysis of the HAND2 target genes during limb bud development was performed by Osterwalder and colleagues (Osterwalder et al., in press), while I focused on the analysis of potential HAND2 target genes during heart development.

### 6.2.3. HAND2 controls the expression of genes involved in early second heart field morphogenesis

The transcriptional regulators Tbx2 and Tbx3 function during early development of second heart field (SHF) derived structures (Harrelson et al., 2004; Mesbah et al., 2008; Mesbah et al., 2012). Both genes were identified by ChIP-Seq analysis as associated with genomic regions interacting with HAND2\(^{3x\text{FLAG}}\) chromatin complexes (Table 1, see chapter 6.2.10 for the analysis of their cis-regulatory landscapes). SHF progenitors reside initially in the splanchnic pharyngeal mesoderm in the dorsal wall of the pericardial cavity, before progressively migrating to the primary heart tube to contribute to the myocardium of the OFT and RV (Kelly, 2012). These cardiac structures and the pharyngeal mesoderm express high levels of HAND2 (Fig.13B-D), and are hypoplastic in Hand2-deficient embryos (Srivastava et al., 1997). We therefore wanted to gain insights into HAND2 functions in the transcriptional networks governing SHF morphogenesis by looking at the early expression of key cardiac genes. The early marker of the SHF Isil remains expressed normally in the pharyngeal mesoderm of Hand2-deficient embryos (E8.25, Fig.15A-B), indicating that the SHF forms normally.
### Figure 15. Expression of early cardiac development markers in wild-type and Hand2-deficient embryos

Effects of Hand2 inactivation on the expression of cardiac genes at E8.25-E8.5. (A-B) *Isl1* expression is not altered (right side view). (C-D) *Tbx2* expression is ectopically expressed in pharyngeal mesoderm (arrowheads, compare with *Isl1* expression domain in (A)). (E-F) *Tbx3* expression is expanded towards the OFT (arrowheads). (G-J) *Nppa* expression is reduced at E8.25 (G-H, front view) and lost in the RV at E9.5 (I-J, arrowheads). (K-L) *Wnt11* expression is reduced in the OFT (arrowheads). (M-X) The expression of *Tbx1* (M-N), *Tbx5* (O-P), *Nkx2.5* (Q-R), *Mef2c* (S-T), *Bmp4* (U-V), and *Fgf8* (X-Y) are not altered in Hand2-deficient embryos.

Interestingly, *Tbx2* and *Tbx3* are expressed ectopically in the *Isl1* positive cells of the pharyngeal mesoderm (compare Fig.15C with Fig.15A and Fig.15D with Fig.15B), and expands to the OFT (compare Fig.15D,F with Fig.15C,E). In contrast, the expression of *Nppa*, which is repressed by TBX2 and TBX3 and required for cardiac
chambers differentiation (Habets et al., 2002; Hoogaars et al., 2004; Houweling et al., 2005) is reduced in Hand2-deficient embryos (E8.25, Fig.15G-J), and totally absent from the hypoplastic RV (E9.5, Fig.15I-J, arrowheads (Morikawa and Cserjesi, 2008; Thattaliyath et al., 2002). These data indicate that the ectopic expression of Tbx2/Tbx3 in myocardial progenitors of the anterior SHF in Hand2-deficient embryos (Fig.15D,F) accounts for the loss of Nppa expression in the right ventricle, while the right ventricle hypoplasia might be an indirect consequence of these alterations. Wnt11 is also essential for the differentiation of SHF cells in the OFT and RV (Cohen et al., 2012; Zhou et al., 2007; Pandur et al., 2002), and its expression is strongly reduced in the OFT of Hand2-deficient embryos (Fig.15K-L). A region located 58kb upstream of Wnt11 is among the highly enriched candidate CRMs associated to HAND2^3xFLAG chromatin complexes, but this remains to be verified and functionally analyzed further (data not shown).

In contrast, the expression of several other important cardiac genes involved in the balance between proliferation and differentiation of the SHF progenitors was not affected by the loss of Hand2 (Tbx1, Tbx5, Nkx2.5, Mef2c, Bmp4, Fgf8: Fig.15M-X). In conclusion, Hand2 controls distinct aspects of the formation of SHF derived structures through a specific set of cardiac regulators (Tbx2, Tbx3, Wnt11).

6.2.4. HAND2 controls the expression of genes associated with known cardiac-specific enhancers

The VISTA enhancer browser is a database consisting of conserved enhancer elements that were tested for LacZ reporter expression (Visel et al., 2007; Attanasio et al., 2013). It can therefore be used to identify overlaps between known cardiac enhancers and the putative CRMs identified by HAND2^3xFLAG ChIP-Seq analysis. Indeed, this analysis identified Furin and Myocardin (Myocd) as potential direct HAND2 target genes. These two genes have been previously established as essential for cardiac development and their landscapes encode VISTA enhancers that are expressed in the heart and interact with HAND2 and p300 protein complexes.

Furin encodes a proprotein convertase that activate signals from the TGFβ family such as BMP4, TGFβ1 and BMP10 by proteolytic cleavage (Cui et al., 1998; Cui et
al., 2001; Dubois et al., 2001; Susan-Resiga et al., 2011). In the developing heart, both the endocardial and myocardial layers express Furin (Roebroek et al., 1998). Embryos deficient in Furin suffer from cardia bifida, which may still result in formation of a short single heart tube that fails to undergo looping, and eventually die around E10.5 (Roebroek et al., 1998; Constam and Robertson, 2000). Endothelial-specific inactivation of Furin results in ventricular septation defects and cardiac valve thickening, which is paralleled by reduced BMP4 activation processing (Kim et al., 2012).

MYOCD is a transcriptional co-activator of serum response factor (SRF) that modulates the expression of tissue-specific SRF-target genes by functioning as a switch between smooth and cardiac skeletal muscle differentiation (Wang et al., 2001; Li et al., 2003; Wang et al., 2003; Long et al., 2007; Hoofnagle et al., 2011). Myocd expression is restricted to these two types of muscles cells, and its inactivation induces embryonic lethality by E10.5 due to defective smooth muscle vasculature (Li et al., 2003). In adult mice, Myocd function is critical for cardiomyocytes survival (Huang et al., 2009).

We used the UCSC browser to visualize the HAND2^{3xFLAG} ChIP-Seq peaks together with the cardiac p300 ChIP-Seq peaks and the conservation of these regions in placental mammals (Fig.16A,E). All other panels representing ChIP-Seq data are organized in accordance with Fig.16A,E. The Furin cis-regulatory landscape encodes a -43kb region, highly enriched for HAND2 and p300 binding (green bar in Fig.16A). This region overlaps (blue bar in Fig.16A) the VISTA enhancer mm169 that is expressed in the heart (Fig.16B) and abdominal region (data not shown, see VISTA browser). The direct interaction of HAND2 complexes with this enhancer was confirmed by ChIP-qPCR using embryonic hearts (E10.5, Fig.16C. See chapter 9.4.6 for fold enrichments calculations). This analysis established the significant 200-fold enrichment of HAND2 interactions with Furin -43kb enhancer in comparison to an unrelated other region chosen for its absence of enrichment by ChIP-Seq analysis (-50kb, Fig.16A and Fig.16C). To determine if Furin is indeed a direct transcriptional target of HAND2, we compared its expression in wild-type and Hand2-deficient embryos (E9.0-E9.25, Fig.16D). Furin expression is altered in Hand2 mutant hearts, in particular in the endocardium (Fig.16D, right panels, arrows).
Figure 16. *Furin* and *Myocd* are direct targets of HAND2 in the developing heart

UCSC browser windows representing the *Furin* (A) and *Myocd* (E) cis-regulatory landscapes. Distances to TSS are indicated in kb. HAND2<sup>3xFLAG</sup> ChIP-Seq peaks are shown in green and p300 ChIP-Seq peaks in hearts (E11.5) are shown in black, above the placental mammal conservation (Cons) plot (PhyloP, UCSC Genome Browser). Green bars indicate HAND2<sup>3xFLAG</sup> peaks in the top 2000 enriched regions. (B,F) LacZ reporter expression of the VISTA browser enhancers represented by blue bars in (A) and (E). (C,G) ChIP-qPCR establishing binding of HAND2<sup>3xFLAG</sup> chromatin complexes at *Furin* (C) and *Myocd* (G) associated cardiac enhancers in E10.5 hearts. No enrichment was seen in control regions. Mean ±SD (n=3) is shown. (D,H) Expression of *Furin* (D) and *Myocd* (H) in wild-type and *Hand2<sup>−/−</sup>* embryonic hearts (E9.0-9.5) using whole mount RNA in-situ hybridization. In-situ hybridization on paraffin sections shows the specific loss of *Furin* from the endocardium (indicated by arrows).

Similarly, HAND2 containing chromatin complexes are enriched at an intronic region located +60kb upstream of the *Myocd* TSS (Fig.16E, green bar), which overlaps with the heart-specific VISTA enhancer mm67 (blue bar in Fig.16E, expression in Fig.16F). Direct interaction of HAND2 with this CRM was confirmed by ChIP-qPCR (Fig.16F). In agreement, *Myocd* expression is altered in *Hand2*-deficient embryos (Fig.16H). Together, these results indicate that *Hand2* is required for *Furin* and *Myocd* expression, and that these genes are likely direct HAND2 targets whose expression is regulated by cardiac enhancers.
6.2.5. *Hand2*-deficient mouse embryos lack the delaminated endocardial cells in the atrioventricular canal

In response to myocardial signals, endocardial cells in the AVC undergo an endothelial-mesenchymal transition (EndMT) around E9.5. These cells delaminate and migrate into the cardiac jelly, where they proliferate and form the cardiac cushions that will ultimately give rise to the cardiac valves (Armstrong and Bischoff, 2004; von Gise and Pu, 2012; Kovacic et al., 2012; Lim and Thiery, 2012; Lin et al., 2012). As HAND2 is expressed by both the endocardium of the AVC and the mesenchymal cells of the cardiac cushions (Fig.13E-F), we investigated if HAND2 regulates endocardial genes that control EndMT mechanism and formation of the cardiac cushions.

Using sagittal sections, I first studied the morphological alterations of the AVC in *Hand2*-deficient embryos. Compared to wild-type hearts in which AVC endocardial cells delaminate normally (E9.5, Fig.17A, asterisk in top panel), a total absence of delaminated cells was apparent in *Hand2*-deficient embryos (E9.5, Fig.17A, bottom panel). Furthermore, the superior and inferior layers of the endocardium were not in contact, leaving an opening in the AVC lumen (Fig.17A). To confirm that the absence of delaminated cells was not due to increased endocardial cell death, wild-type and *Hand2*-deficient hearts were analyzed by TUNEL (E9.5, Fig.17B). No increase in apoptotic cell number was observed in *Hand2*-deficient hearts, which thus excludes cell death as likely cause of delamination failure. In addition, the expression of the transcription factor NFATC1 (Nuclear Factor of Activated T cells) was analyzed as it is expressed specifically in endothelial cells of the endocardium and not required for EndMT. Its functions in maintaining the endocardial lineage are crucial for valve remodeling at later stages (Ranger et al., 1998; de la Pompa et al., 1998; Wu et al., 2011; Chang et al., 2004). NFATC1 is coexpressed with HAND2 in the endocardium and remains expressed normally in *Hand2*-deficient embryos (Fig.17C). This analysis indicate that the endocardium of *Hand2*-deficient embryos is normally patterned with respect to the fates of cardiac valve progenitors. The absence of delaminated cells is thus not due to an endocardial defect but likely a consequence of defective EndMT and/or migration of the cardiac cushion progenitors.
Figure 17. Absence of delaminated endocardial cells in the atrioventricular canal (AVC) of Hand2-deficient embryos

(A) Hematoxylin-eosin staining of sagittal heart sections in Hand2-deficient embryo (lower panel) compared to wild-type embryos (top panel) (E9.5). Black arrow: endocardium, black asterisk: mesenchymal cells of the cardiac cushions, lv: left ventricle, avc: atrioventricular canal, la: left atrium.

(B) TUNEL assay using sagittal heart sections of Hand2-deficient embryos (lower panel) and wild-type controls (top panel) (E9.5). Apoptotic cells are shown in green, nuclei are counterstained with Hoechst (blue). White arrow: endocardium, white arrowhead: myocardium.

(C) The endocardial marker NFATC1 co-localizes with HAND2 in the AVC (white arrow, top panel) and remains expressed in Hand2-deficient embryos (open arrow, lower panel). Scale bar: 100µm.

6.2.6. HAND2 directly regulates Has2-mediated extracellular matrix deposition in the cardiac jelly

To gain further insight into the defective delamination of AVC endocardial cells, the glycosaminoglycans in the extracellular matrix (ECM) were stained by alcian blue. Glycosaminoglycans are required for the migration of cardiac cushion progenitors into the cardiac jelly. Indeed, the ECM in the cardiac jelly of the AVC appears reduced in Hand2-deficient hearts in comparison to wild-type controls (Fig.18A). This suggests that the endocardial cells are unable to migrate properly, which would explain the lack of cardiac cushion progenitors in the absence of Hand2 (Fig.17A).

The enzyme Hyaluronan Synthase 2 (Has2) produces one major component of the ECM, the hyaluronic acid (HA). Has2 is downregulated by Hand2-deficient embryos, which, together with its requirement for cardiac jelly and cardiac cushion formation, points to it being a potential direct HAND2 target (Tsuchihashi et al., 2011;
Indeed, our ChIP-Seq analysis revealed two HAND2 binding regions within the proximal promoter of Has2 located 2.4kb downstream and 0.9kb upstream of the TSS, respectively (Fig.18B). ChIP-qPCR analysis of E10.5 HAND2\textsuperscript{3xFLAG} hearts established that HAND2 complexes interact with the +2.4kb region but neither to the -0.9kb nor to a control region (Fig.18C). In agreement, the expression of Has2 is altered in the AVC of Hand2-deficient hearts (Fig.18D).

Figure 18. The production of extracellular matrix in AVC cardiac jelly is regulated by HAND2, possibly by direct regulation of Has2 expression. (A) Alcian Blue staining of sagittal heart sections in Hand2-deficient embryos (right panel) and wild-type controls (left panel) (E9.5). (B) UCSC browser windows representing the Has2 proximal promoter region. Green bars indicate HAND2\textsuperscript{3xFLAG} peaks in the top 2000 enriched regions. The light blue bar at -0.9kb represents the location of a published enhancer that is active in limb buds (Liu et al., 2013). (C) ChIP-qPCR establishing direct interaction of HAND2\textsuperscript{3xFLAG} chromatin complexes with the Has2 proximal promoter in E10.5 hearts (E10.5). The -4.3kb region is used as negative control. Mean ±SD (n=3) is shown. (D) Expression of Has2 in wild-type and Hand2\textsuperscript{-/-} embryonic heart by whole mount (left panels) and section (right panels) in-situ hybridization (E9.5). (E) Transgenic embryo (E10.5, n=1) showing LacZ expression under control of the +2.4kb CRM in the Has2 intronic region indicated by a dark blue bar in panel (B). The transgene is not expressed in the heart (E') but in the dorsal and ventral forelimb bud mesenchyme (E'') and in the dorsal hindlimb bud mesenchyme (E''').
These results support the proposal that Has2 is as a direct target of HAND2 in the developing heart. To assess the potential transcription enhancing activity of the +2.4kb region, we used a LacZ reporter assay in transgenic mouse embryos. Preliminary analysis shows that the region encompassing the Has2 +2.4kb HAND2 peak (Fig.18B, dark blue bar) drives reporter expression in the somites, second branchial arch and in the limb buds of a transgenic embryo at E10.5 (n=1, Fig.18E). However, no LacZ reporter expression was detected in the heart (Fig.18E'). In limb buds, the LacZ expression is restricted to the dorsal and ventral forelimb sub-ectodermal mesenchyme (Fig.18E'') and to the dorsal hindlimb (Fig.18E'''). Preliminary ChIP-qPCR analysis confirms that HAND2 complexes bind to this +2.4kb region in developing limb buds (E10.5, data not shown). The lack of reporter expression in the heart may be an effect of the transgene insertion site and needs to be investigated by the generation of additional independent transgenic founders embryos (currently ongoing). This is important as our other findings support a role of HAND2 in regulating Has2 during the formation of the cardiac cushions.

6.2.7. Twist1 but not Sox9 is a direct transcriptional target of HAND2 during cardiac cushion formation

Twist1 and Sox9 are required for the proliferation and differentiation of the mesenchymal cells of the cardiac cushions (Shelton and Yutzey, 2008; Akiyama et al., 2004) and regions interacting with HAND2 complexes were detected in the regulatory landscapes of both genes. We therefore investigated if alterations in these two putative HAND2 targets could contribute to the disruption of cardiac cushion morphogenesis in Hand2-deficient embryos. HAND2 complexes interact with several regions within Twist1 cis-regulatory landscape; the most enriched being located at +10kb and at +130kb from Twist1 TSS (Fig.19A). ChIP-qPCR analysis confirmed that both regions are enriched in HAND2 containing chromatin complexes in developing hearts (E10.5, Fig.19B). Furthermore, preliminary ChIP-qPCR results indicate that the binding region at +130kb is not enriched in limb buds, suggesting that it could be a heart-specific CRM bound by HAND2 complexes (data not shown). Therefore, the corresponding genomic region encoding this peak is currently being used to generate LacZ reporter transgenic embryos (dark blue bar in Fig.19A).
Figure 19. HAND2 directly control Twist1 expression in the cells of the cardiac cushions
(A) UCSC browser windows representing the HAND2 3xFLAG ChIP-Seq peaks within Twist1 cis-regulatory landscape. The green bar at +130kb from Twist1 TSS indicates top2000 enriched region. The dark blue bar at +130kb represents the genomic region used to generate transgenic reporter embryos. The light blue bar at +10kb represents the location of the VISTA enhancer hs1880 not active at E11.5. (B) ChiP-qPCR establishing direct interaction of HAND2 3xFLAG chromatin complexes with the Twist1 +10kb and +130kb regions in hearts (E10.5), but not with the control region located at +135kb. Mean ±SD (n=3) is shown. (C) Expression of Twist1 in wild-type and Hand2 Δ/Δ embryonic heart at E9.5. (D,F) TWIST1 is co-expressed with HAND2 in the mesenchymal cells of the cardiac cushions (white asterisks) but not endocardial cells (white arrow) at E9.5 (D) and E10.5 (F). (E) TWIST1 is not expressed by the AVC of Hand2-deficient embryos (open arrow: endocardium). (G) TWIST1 is also co-expressed with SOX9 in the mesenchymal cells of the cardiac cushions (white asterisks) at E10.5. Scale bar: 100µm

In contrast, the human orthologous region encompassing the putative HAND2 CRM at +10kb does not drive LacZ reporter expression in transgenic reporter embryos (E11.5, VISTA element hs1880, light blue bar in Fig.19A). Twist1 expression is totally lost from the AVC of Hand2-deficient hearts (E9.5, Fig.19C), in agreement with Twist1 being a direct transcriptional target of HAND2. Indeed, the TWIST1 protein is
co-expressed with HAND2 in the progenitors that have delaminated after the initial EndMT (Fig.19D, white asterisks). In Hand2-deficient hearts, no cells expressing TWIST1 are detected in the AVC as expected from the absence of delaminated cells (Fig.19E). By E10.5, TWIST1 is co-expressed in all cardiac cushion progenitors with HAND2 (Fig.19F) and SOX9 (Fig.19G). In summary, this analysis shows that Twist1 expression depends on HAND2 during cardiac cushion formation. The CRM identified 130kb upstream of Twist1 TSS is likely involved in mediating direct transcriptional regulation of Twist1 expression by HAND2 after delamination.

The expression of Sox9 is downregulated in the AVC of Twist1-deficient embryos (Vrljicak et al., 2012). The co-expression of HAND2, TWIST1 and SOX9 by cardiac cushion progenitor cells suggests that Sox9 may also be a direct target of HAND2 in these cells, but no region interacting with HAND2 in Sox9 genomic landscape in developing hearts could be confirmed to date (ChIP-Seq and ChIP-qPCR, data not shown). In wild-type embryos, SOX9 expression initiates on the atrial side of the AVC in cells that have undergone EndMT. In these cardiac cushion progenitors, SOX9 is co-expressed with HAND2 (E9.5, Fig.20A, white asterisks).

**Figure 20.** Sox9-mediated cardiac jelly invasion is not regulated by HAND2
(A) SOX9 is co-expressed with HAND2 in the mesenchymal cells of the cardiac cushions (white asterisks) but not endocardial cells (white arrow) at E9.5. (B) In Hand2-deficient embryos, SOX9 remains expressed in endocardial cells (open arrow). Scale bar: 100µm
The remainder of the AVC endocardium expresses only HAND2 (Fig.20A, white arrow). Interestingly, in Hand2-deficient hearts SOX9 is expressed in endocardial cells located on the atrial side of the AVC (Fig.20B, open arrow), at the atrial-ventricular position where the first delaminated SOX9-positive cells normally would appear. As these cells express SOX9 but fail to delaminate, it suggests that the initial EndMT, prior to Sox9-dependant cardiac jelly invasion, might be impaired in Hand2-deficient hearts.

6.2.8. **Snai1-dependent initiation of endothelial-mesenchymal transition in the atrioventricular canal is regulated by HAND2**

We have shown that Hand2-deficient embryos lack delaminated cardiac cushion progenitors, and have obtained evidence that this phenotype could be caused by 1) deficient Has2-mediated control of ECM production impairing cellular migration; and by 2) a defect in Twist1-mediated regulation of the proliferation of post-EndMT delaminated cardiac cushion progenitors. We therefore wondered whether the initial EndMT was also affected by the Hand2 deficiency, and investigated the expression of the EMT inducing transcription factor SNAI1 and its downstream effectors. Consistent with the hypothesis of a likely underlying EndMT defect in Hand2 mutant hearts, Snai1 expression was totally lost from mutant AVC by E9.5 (Fig.21A). Furthermore, several HAND2 interacting regions were identified within Snai1 genomic landscape (Fig.21B). ChIP-qPCR analysis confirmed the existence of HAND2 binding regions located 45kb, 58kb and 59kb downstream of Snai1 TSS in developing hearts (E10.5, Fig.21C). The genomic region encompassing the peaks at +58/59kb (blue bar in Fig.21B) was used to generate LacZ reporter transgenic mouse embryos (n=5, Fig.21D). This CRM is active in the branchial arches, craniofacial structures (Fig.21D’), forelimbs, and epicardium (Fig.21D’’). The extent of cardiac expression will require histological analysis. The pattern of LacZ reporter expression in these embryonic structures is similar to the Snai1 endogenous expression at an equivalent developmental stage (E10.5, (Sefton et al., 1998)).

To gain further insight into the potential disruption of the EndMT in Hand2-deficient hearts, VE-Cad and VIM distribution were studied. Both these genes are established Snai1 targets during cardiac cushion formation (Timmerman et al., 2004; Cano et al., 2000; Batlle et al., 2000).
Figure 21. Snai1 expression and the downstream endothelial-mesenchymal transition are altered in the AVC of Hand2-deficient embryos

(A) Expression of Snai1 in wild-type and Hand2Δ/Δ embryonic heart at E9.5. (B) UCSC browser windows representing the regions enriched in HAND2�FLAG chromatin complexes in Snai1 cis-regulatory landscape. The green bar at +58-59kb from Snai1 TSS indicates top2000 enriched region. The dark blue bar represents the genomic region used to generate transgenic reporter embryos. (C) ChIP-qPCR establishing direct interaction of HAND2�FLAG chromatin complexes with the Snai1 +45kb, +58kb and +59kb regions in hearts (E10.5). The +64kb region is used as negative control. Mean ±SD (n=3) is shown. (D) Transgenic embryo (E10.5, n=5) showing LacZ expression under control of the CRM located +58/59kb from the Snai1 TSS. The transgene is active mainly in craniofacial structures (D') and in the epicardium (D''). (E) In the wild-type AVC, VE-Cadherin is expressed at intercellular junctions in the endocardium (red, white arrow) but not delaminated cardiac cushion cells (white asterisk). VIM is expressed by both endocardial and cardiac cushion cells. (F) The endocardium of Hand2-deficient embryos expresses VE-Cad but not VIM (red, open arrow) but not VIM. Scale bar: 50µm

In wild-type embryos, VE-Cadherin marks the intercellular junctions between the cells of the endocardium (Fig.21E, white arrow) while it is no longer expressed by the delaminated progenitors (Fig.21E, white asterisk). In contrast, VIM is expressed by both endocardial (Fig.21E, white arrow) and delaminated progenitors (Fig.21E, white asterisk). Remarkably, VE-Cad remains expressed by the endocardium of Hand2-
deficient hearts (Fig.21F, open arrow), while VIM is lost from these cells (Fig.21F, white arrow). Endocardial cells of the AVC in Hand2-deficient embryos lack Snai1 expression and characteristic EndMT changes that normally allow them to initiate delamination and migration. Both SNAI1-mediated disassembly of VE-Cad positive junctions between endocardial cells and the activation of the motile cellular phenotype in VIM positive cells are disrupted in the absence of HAND2. Together, these findings indicate that HAND2 regulates EndMT by directly controlling Snai1 expression.

6.2.9. The Hand2 loss-of-function affects the EndMT upstream signaling but not the patterning of the myocardium

The EndMT of the AVC endocardial progenitors is initiated by BMP2 signaling from the surrounding myocardium and additionally involves myocardial TGFβ2 and endocardial GATA4 and Notch1 (Ma et al., 2005; Rivera-Feliciano et al., 2006; Timmerman et al., 2004; Luna-Zurita et al., 2010). As HAND2 appears to directly regulate the expression of key genes involved in EndMT, we investigated if the upstream myocardial signaling events are also altered by loss of Hand2 functions. In the myocardium of E9.5 Hand2-deficient embryos, Bmp2 is expressed in a smaller subdomain of the AVC than in wild-type (Fig.22A) and the expression of its downstream effector Msx1 is lost from the endocardium (Fig.22B). In addition, in Hand2-deficient embryos Gata4 expression is also lost from the endocardium (Fig.22C) and Tgfβ2 expression from the myocardium (Fig.22D). In contrast, Notch1 expression is not affected by the loss of Hand2 (Fig.22E). These alterations raise the possibility that in addition to a direct role in EndMT, Hand2 also functions in regulating these signaling cascades in the endocardium, possibly downstream of Notch1 and BMP2 and upstream of Gata4. The reduced Bmp2 expression in the AVC myocardium (Fig.22A) could be a consequence of a patterning defect in the AVC, which would impact on myocardium-endocardium signaling. For this reason, we also analyzed the expression of Tbx2 and Tbx3 that are required downstream of Bmp2 for the patterning of the AVC by inhibiting chamber differentiation through Nppa repression (Ma et al., 2005; Christoffels et al., 2004).
In Hand2-deficient embryos, the expression of TBX2 and TBX3 is not altered in the myocardium of the AVC (Fig.23A-10D) and remains complementary to Nppa expression in the left ventricle and atrium (Fig.23E). Finally, the delineation of LV-AVC boundary by Hey2 is also not altered by the inactivation of Hand2 (Fig.23F), which shows that the AVC myocardium is correctly patterned in spite of the more restricted Bmp2 expression domain. Therefore, the defects in the EndMT in Hand2-deficient embryos cannot be a consequence of altered myocardial signaling. The loss of Tgfβ2 and Msx1 expression (Fig.22B,D) is rather caused by an intermediate role of HAND2 between the Notch1/Bmp2/Tgfβ2 endocardium-myocardium feedback loops and the direct regulation of the Snai1 dependent EndMT.
Figure 23. The myocardium of the AVC is normally patterned in Hand2-deficient embryos (A-D) In wild-type embryos (E9.5), TBX2 (A) and TBX3 (C) are expressed in the myocardium of the AVC (white arrows) whereas HAND2 is expressed in the endocardium (white arrowheads). In Hand2-deficient embryos, both TBX2 (B) and TBX3 (D) remain expressed by the myocardium (open arrowheads). Scale bar: 50µm (E) Nppa expression is identical in the LV and LA myocardium of wild-type and Hand2-deficient embryos. (F) The LV-AVC boundary is marked by the limit of Hey2 expression domain and remains in Hand2-deficient embryos.

6.2.10. Tbx2 and Tbx3 are direct targets of HAND2 during cardiac development

We showed that Hand2 deficiency alters Tbx2 and Tbx3 expression in SHF progenitors (E8.25, Fig.15C-F), but not their expression in the myocardium of the AVC (E9.5, Fig.23A-D). Several regions within both Tbx3 and Tbx2 cis-regulatory landscapes were identified by the HAND2\textsuperscript{3xFLAG} ChIP-Seq analysis (Fig.24A and Fig.25C).
Figure 24. The Tbx3 -18kb element is bound by HAND2 chromatin complexes in developing hearts and displays cardiac-specific enhancer activity

(A) UCSC browser windows representing HAND2xFLAG ChIP-Seq peaks in the Tbx3 cis-regulatory landscape. The green bars at -58kb and -18kb from Tbx3 TSS indicate top2000 enriched regions. The dark blue bar represents the genomic regions used to generate transgenic reporter embryos which expression are shown in (C) and (D). (B) ChIP-qPCR from embryonic hearts (E10.5, red, left panel) and limb buds (E10.5, green, right panel) establish that HAND2 chromatin complex interact preferentially with the Tbx3 -18kb CRM in developing hearts and with the -58kb CRM in limb buds. (C) The Tbx3 -18kb CRM shows heart-specific enhancer activity in LacZ-reporter transgenic embryos. The transgene is expressed in the AVC, OFT and interventricular septum (IVS). (D) In contrast, the -58kb CRM drives reporter expression in limb buds and flank mesoderm but not in the developing heart. (E) The Tbx3 -18kb CRM is active in the cells of the prospective atrioventricular node.
Tissue-specific ChIP-qPCR established that the region located -18kb upstream of the *Tbx3* TSS was enriched mostly in HAND2 chromatin complexes isolated from developing hearts (E10.5, Fig.24B, left panel), while the -58kb region was enriched mostly in developing limb buds and flank tissues (E10.5, Fig.24B, right panel). To further determine if these candidate CRMs function as transcriptional enhancers, transgenic mouse embryos using a *LacZ* reporter construct were produced by M.Shoukry in the laboratory of A.Visel (Berkeley, USA). This analysis reveals that the -18kb CRM enhances expression in the AVC, OFT and interventricular septum (Fig.24C). In contrast, the activity of the -58kb CRM is restricted to the posterior-proximal forelimb bud and flank mesenchyme (Fig.24D). The CRM-*LacZ* reporter expression patterns are consistent with the ChIP-qPCR binding profiles (Fig.24B), which reveal tissue-specific functional relevance of the HAND2 interaction with these two *Tbx3* CRMs. The CRM located at -18kb is expressed in the atrioventricular node (AVN) in the cells that express *Tbx3* in the myocardium adjacent to the cardiac cushions (Fig.24E, (Christoffels et al., 2010). Prior to formation of the AVN, no *LacZ* expression driven by this CRM was observed (data not shown). Interestingly, this CRM is located within the 160kb genomic landscape (spanning from -83kb to +77kb) previously identified as regulating *Tbx3* expression in both AVN and limb buds (Horsthuis et al., 2009). Our ChIP-Seq analysis not only identified a discrete AVN-specific CRM within *Tbx3* *cis*-regulatory landscape, but also established this CRM as interacting directly with HAND2 chromatin complexes during the regulation of *Tbx3*-mediated development of the cardiac conduction system.

Next, we determined the TBX3 and HAND2 protein distribution at E11.5, when *Tbx3* -18kb CRM drives reporter expression in the AVN (Fig.24E), but no evidence of cells co-expressing both transcription factors was obtained. While TBX3 is confined to the myocardium (Fig.25A, white arrowheads), HAND2 is expressed in the endocardium (Fig.25A, white arrows) and mesenchymal cells of the cardiac cushions (Fig.25A, white asterisks). One possible explanation is that the *Tbx3* -18kb CRM is repressed in HAND2 expressing cells and is active in the myocardium of the AVN due to the lack of HAND2 chromatin complexes. The enrichment of HAND2 complexes at this CRM in limb buds (Fig.24B, right panel), in combination with lack of *LacZ* reporter
expression (Fig.24C) is in agreement with the proposal that the HAND2 chromatin complexes repress this CRM.

Figure 25. HAND2 directly interacts with a CRM located in Tbx2 cis-regulatory landscape in developing hearts

(A) Expression of TBX3 (red) and HAND2 (green) in the AVC at E11.5. TBX3 is expressed by myocardial cells (white arrowheads) and HAND2 in endocardial cells (white arrows) and mesenchymal cells of the cardiac cushions (white asterisks). (B) Expression of TBX2 (red) and HAND2 (green) in the AVC at E11.5. TBX2 is expressed in myocardial cells (white arrowheads) and co-expressed with HAND2 in mesenchymal cells of the cardiac cushions (white asterisks). Scale bar: 100µm. (C) UCSC browser windows representing HAND2<sup>3xFLAG</sup> ChIP-Seq peaks in the Tbx2 cis-regulatory landscape. The green bars at -112kb and -32kb from Tbx2 TSS indicate top2000 enriched regions. The light blue bar at -112kb indicates a limb bud enhancer (Infante et al., 2013). The dark blue bar at -32kb indicates the genomic region used to generate transgenic reporter embryos. (D) ChIP-qPCR from E10.5 embryonic hearts (red, top panel) and limb buds (green, bottom panel) establishes that HAND2 chromatin complexes interacts preferentially with the Tbx2 -32kb CRM in the heart and with the -112kb CRM in limb bud.
In contrast to TBX3, TBX2 is coexpressed with HAND2 in the cardiac cushion cells at E11.5 (Fig.25B, white asterisks). Of the two candidates CRMs interacting with HAND2 within Tbx2 regulatory landscape, the CRM located 112kb upstream of Tbx2 (Fig.25C, light blue bar) has previously been shown to act as a transcriptional enhancer active in the posterior limb bud mesenchyme (Infante et al., 2013). Accordingly, the enrichment of HAND2 chromatin complexes encompassing this CRM was higher in developing limb buds than hearts (Fig.25D). In contrast, the -32kb candidate CRM is enriched both in developing hearts and limb buds (Fig.25D). Transgenic embryos using this region are currently being generated (dark blue bar in Fig.25C, results not yet available). Given the colocalization of TBX2 and HAND2 in mesenchymal cells of the cardiac cushions (E11.5, Fig.25B), we expect this -32kb CRM to drive LacZ expression in developing cushions. If the enrichment of this CRM in HAND2 chromatin complexes isolated from limb buds is functionally relevant (Fig.25D, lower panel), LacZ expression should also be detected in limb buds.

Another member of the T-box gene family important for cardiac development is Tbx20. While myocardial Tbx20 regulates Bmp2 expression in the AVC, endocardial Tbx20 is not required for EndMT (Cai et al., 2011; Cai et al., 2013). A prominent HAND2 ChIP-Seq peak was identified 200kb upstream of the Tbx20 TSS (Fig.26A, green bar). This region is the orthologous of a human VISTA enhancer element (hs463, Fig.26A, dark blue bar) that drives LacZ reporter expression in the heart of transgenic embryos (Fig.26B). However, no changes in Tbx20 expression were detected in the developing hearts of Hand2-deficient embryos (Fig.26C). This agrees with previous observations indicating that Tbx20 acts upstream of Hand2 (Singh et al., 2005; Takeuchi et al., 2005). Nevertheless, the high enrichment of HAND2 chromatin complexes at the CRM located -200kb from Tbx20 points to a possible involvement of HAND2 at late stages on Tbx20-mediated maturation of the cardiac cushions (Cai et al., 2013).
Figure 26. HAND2 directly interacts with a cardiac CRM located in Tbx20 cis-regulatory landscape

(A) UCSC browser window showing a region from the HAND2\textsuperscript{3xFLAG} ChIP-Seq top 2000 enriched region located 200kb upstream of Tbx20 TSS (green bar). The dark blue bar represents the location of the VISTA enhancer hs463. (B) Tbx20 expression in the endocardium (arrows) and myocardium (arrowheads) of wild-type and Hand2-deficient embryos. (C) LacZ reporter activity driven by the VISTA enhancer hs463.
7. Discussion

7.1. Identification of HAND2 target gene networks that function during heart morphogenesis

The development of the four-chambered mammalian heart depends on complex spatial and temporal coordination of signaling pathways and transcription factors directing the proliferation and commitment of myocardial and endocardial progenitors into multiple cell types (Lescroart and Meilhac, 2012; Vincent and Buckingham, 2010). The wide range of life-threatening congenital heart diseases resulting from perturbations during cardiogenesis reflects the complexity and importance of the tightly regulated mechanisms governing heart development (Bruneau, 2008). It is thus crucial for both diagnosis and potential regenerative therapies to understand how hierarchical gene regulatory networks interact to orchestrate the morphogenesis of cardiac structures (Srivastava and Olson, 2000; Olson, 2006; Srivastava, 2006; Rana et al., 2013; Xin et al., 2013; Wamstad et al., 2014). Transcription factors active in the second heart field have been well studied, but knowledge on the gene networks regulated by HAND2 and its direct targets is limited (Rochais et al., 2009b; Dyer and Kirby, 2009; Kelly, 2012; Vincentz et al., 2011). Mutations of HAND2 have been linked to human congenital heart diseases, which makes the identification of the gene networks controlled by HAND2 of particular interest (Shen et al., 2010; Strehle et al., 2012).

7.1.1. Genome-wide identification of cardiac CRMs

Several studies have determined the genomic target range of transcriptional co-activators and the distribution of different chromatin modifications marks in the heart (May et al., 2012; Blow et al., 2010; Shen et al., 2012). In contrast, so far few ChIP-Seq datasets from cardiac transcription factors (TFs) are available. Many of them are cell type or lineage specific and have been generated using commercial ChIP-grade antibodies in adult tissues. For example, TBX3, which is expressed by the conduction system, has been overexpressed in the adult myocardium to study its target range (van den Boogaard et al., 2012). This approach identified common transcriptional targets of TBX3, NKL2.5 and GATA4, but overexpression likely does not reflects the physiological TF-chromatin complexes and is potentially prone to the identification of
false positive interactions. Expression of several biotinylated cardiac TFs in cultured HL1 cardiomyocytes also led to the identification of cardiac cis-regulatory modules (CRMs), but targets in cultured cells may differ from those in vivo (He et al., 2011; Massie and Mills, 2008). In another study, TBX20 binding regions in adult hearts were identified using a BAC transgenic mouse line expressing a Tbx20 allele tagged by eGFP, but a transgenic approach may again not reveal all endogenous targets (Shen et al., 2011). Altogether, the limitations of these studies show the importance of using endogenous levels of expression of TFs for ChIP-Seq approaches to accurately profile the DNA binding of transcriptional regulators. For this purpose, epitope tagging of TFs for high-throughput genome-wide approaches has become a powerful tool (Zhang et al., 2008a; Kidder et al., 2011; Landt et al., 2012). In particular, mouse alleles that express endogenously epitope-tagged TFs have been successfully used for ChIP-Seq analysis in various embryonic tissues (Attanasio et al., 2014). We have also chosen to insert a 3xFLAG epitope into the HAND2 protein to identify the CRMs interacting with HAND2 in a physiological manner during mouse embryogenesis. As the ChIP-Seq analysis used in this study was done using pooled Hand2-expressing embryonic tissues, we plan to perform a cardiac-specific ChIP-Seq analysis to identify additional heart-specific targets of HAND2.

7.1.2. HAND2 and its target CRMs in the heart: direct and indirect DNA-binding interactions

The genome-wide range of HAND2-containing chromatin complexes does not imply that all candidate CRMs are directly bound by HAND2. For instance, embryos homozygous for a mutant form of HAND2 deficient for DNA-binding exhibit right ventricle (RV) hypoplasia and lethality two to three days later than Hand2 null embryos (Liu et al., 2009). This suggests that the control of gene expression during second heart field (SHF) patterning may be regulated independently of the direct interaction of HAND2 with DNA, while the subsequent growth of the ventricle requires direct binding of HAND2 to its target CRMs (Liu et al., 2009). These mutant embryos also display disorganized and enlarged cardiac cushions by E11.5, which points to a role of Hand2 in the formation of these structures. The DNA binding mutation of Hand2 does not affect the expression of the HAND2 genetic targets Gata4 and Nppa, suggesting that
these genes are regulated by HAND2 in a DNA-binding independent manner (Srivastava et al., 1997; Thattaliyath et al., 2002; Liu et al., 2009). In agreement, our results suggest that Gata4 and Nppa are indirect targets of HAND2 as the ChIP-Seq analysis and Hi-C data did not detect any HAND2 candidate CRMs in their cis-regulatory landscapes (Dixon et al., 2012).

In addition, heterodimerization of HAND2 with other transcriptional regulators as part of chromatin complexes adds another level of complexity to the attempt to identify direct HAND2 target CRMs (Vincentz et al., 2011). Indeed, several studies have shown that HAND2 interacts with the Nppa promoter as part of a complex with MEF2C, GATA4 and/or NKX2.5, and that at least some of these interactions do not require E-box motifs (Zang et al., 2004; Dai et al., 2002; Thattaliyath et al., 2002). It is tempting to speculate that the CRMs that are directly interacting with HAND2 should encode the consensus E-box sequence CANNTG (Dai and Cserjesi, 2002). In contrast, this sequence could be absent from CRMs to which HAND2 binds indirectly. Indeed, 47% of the peaks identified in the HAND23xFLAG ChIP-Seq analysis were enriched in motifs containing a variant of the E-box sequence (Osterwalder et al., in press), showing that in about half of the candidate CRMs HAND2 might not directly bind to DNA. The identification of potential HAND2-interacting proteins within transcriptional complexes could be achieved by searching known transcription factor binding sites motifs in the identified CRMs, or by Mass Spectrometry that takes advantage of the 3xFLAG tag that we have introduced in HAND2.

### 7.1.3. HAND2 target genes in the progenitors of the second heart field

Our analysis mainly focuses on HAND2 transcriptional targets at E9.5, which is the developmental stage at which Hand2-deficient embryos are still alive that is closest to the stages used for ChIP-Seq analysis. The early embryonic lethality could likely be circumvented by conditionally inactivating Hand2 e.g. in the endothelial compartment using the Tie2-Cre or Nfatc1-Cre alleles to study its functions in the endocardium in more details (Kisanuki et al., 2001; Wu et al., 2012).

Our data show that both the myocardium and endocardium of the AVC in Hand2-deficient embryos are not affected by apoptosis at E9.5. However, the apoptosis is increased already at E8.5 in the bulbus cordis that connects the OFT to the RV and by
E9.5 also in the pharyngeal mesoderm and OFT (data not shown and (Yamagishi et al., 2001; Tsuchihashi et al., 2011)). These two studies concluded that Hand2 is required for the survival of SHF cells. In particular, the RV hypoplasia is caused by the defective contribution of SHF progenitors to the heart (Yamagishi et al., 2001; Tsuchihashi et al., 2011). Therefore, we analyzed several markers of the SHF at the stage where these cells start to invade the heart tube (E8.25-E8.5). In Hand2-deficient embryos, we observed ectopic expression of Tbx2 and Tbx3 in the anterior SHF progenitors in the pharyngeal mesoderm (the Tbx2 and Tbx3 cis-regulatory landscapes are discussed below). In another study, early induction of a Tbx2 transgene at E7.5 severely alters cardiac development and results in hypoplasia of the heart tube. This alteration is caused by impaired deployment of the SHF progenitors, reduced myocardial proliferation, and repression of the chamber differentiation markers Nppa and Cx40 (Dupays et al., 2009). As TBX2 directly repress Nppa expression (Habets et al., 2002), and as Nppa expression is lost from the hypoplastic RV in Hand2-deficient embryos, we can establish a signaling cascade with direct (Tbx2) and indirect (Nppa) effects of Hand2 on the development of the RV.

The expression of Nkx2.5 in the pharyngeal mesoderm is necessary for the development of the SHF (Zhang et al., 2014). The ChiP-Seq analysis also identified a CRM highly enriched in HAND2 chromatin complexes in the Nkx2.5 proximal promoter (data no shown). This CRM drives LacZ reporter expression in Hand2-expressing structures such as the OFT and RV (Tanaka et al., 1999). As Nkx2.5 expression is not affected in Hand2-deficient embryos, this suggests a very robust regulation of its expression through additional cardiac CRMs or compensation by other TFs and reflects the central role of Nkx2.5 for cardiac development.

7.1.4. Furin and Myocd are direct transcriptional targets of HAND2
We identified Furin as a new direct target of HAND2 in the endocardium of the AVC. Indeed, the inactivation of Furin in endothelial cells results in malformation of the mitral and tricuspid valves, which are derived from the AVC cardiac cushions and do not form in Hand2-deficient embryos (see below) (Kim et al., 2012; Lin et al., 2012). FURIN is involved in processing of BMP10, and in Bmp10-deficient embryos, the formation of the
cardiac cushions and trabeculations is impaired and the ventricular wall remains hypoplastic (Susan-Resiga et al., 2011; Chen et al., 2004). Bmp10 participates in ventricular chamber formation by activating the expression of Tbx20 (Zhang et al., 2011). In agreement, the inactivation of Tbx20 disrupts trabeculation and cardiac cushion formation (Stennard et al., 2005). Both the acellular AVC cardiac cushions and hypotrabeculation phenotype are reminiscent of the ones observed in Hand2-deficient embryos. In addition, Hand2 expression in the whole ventricular myocardium in transgenic mice promotes the formation of trabeculations and the expression of Bmp10 (Togi et al., 2006). Finally, Bmp10 is transactivated by the MYOCD/SRF complex, which induces cardiomyocytes proliferation and chamber maturation (Huang et al., 2012). This is interesting in light of our observation that Myocd is a direct target of HAND2. Together, these results suggest that the trabeculation and cardiac cushion defects in Hand2-deficient embryos may be partially mediated by impairing downstream BMP10 signaling via Furin or/and Myocd. This will be analyzed further by assessing Bmp10 expression in Hand2 constitutive and/or conditional mutant embryos.

7.1.5. A causal link between atrioventricular canal defects and embryonic lethality in Hand2-deficient embryos?

The AVC of Hand2-deficient embryos displays two major defects compared to wildtypes. First, the cells delaminating from the endocardium to form the cardiac cushions are absent (see below). Second, there is an opening between the superior and inferior layers of the endocardium that leaves the AVC lumen open, in contrast to wild-type hearts. Such defects in AVC closure have only been described for Endoglin-deficient embryos (Bourdeau et al., 1999). Endoglin is expressed by the endocardium and by the endothelium compartment of blood vessels. Its inactivation results in defective EndMT, acellular cardiac cushions and embryonic lethality by E10.5. (Bourdeau et al., 1999). In addition, the inactivation of Bmp2, Has2, Cspg2 or Tbx20 also alter the formation of the cardiac cushions and leads to embryonic lethality prior to E10.5-E11.5, similarly to the Hand2 inactivation (Rivera-Feliciano and Tabin, 2006; Camenisch et al., 2000; Yamamura et al., 1997; Mjaatvedt et al., 1998; Cai et al., 2011). All these mutants also display a failure in the AVC lumen closure, despite this phenotype is not described in these studies. In zebrafish, the two sides of the AVC are brought together by the
contractions of the myocardium (Scherz et al., 2008). Thus, myocardial contraction failure is a possible explanation to the defective closure of the AVC. In physiological conditions, the electrical impulse driving peristaltic cardiac contractions in early embryos at the heart tube stage is generated in the venous pole. Conduction is rapid in atria and ventricles, while in the AVC and OFT the conduction is slow resulting in a contraction delay in these compartments (de Jong et al., 1992; Moorman and Christoffels, 2003; Christoffels and Moorman, 2009). The primordia of the future components of the adult conduction system start to develop at ~E9.5, which is around the onset of embryonic lethality in Hand2-deficient mouse embryos (Christoffels and Moorman, 2009; Christoffels et al., 2010; Srivastava et al., 1997). Unidirectional blood flow without regurgitation is also established at this stage and requires coordinated cardiac contractions and formation of the AVC cardiac cushions (Nishii and Shibata, 2006). Thus, irregular contractions and absence of cardiac cushions results in retrograde blood flow into the atria, which reduces cardiac output and leads to death (Scherz et al., 2008). In Hand2-deficient mouse embryos, cardiac contractions are observed at E9.5 but appear slightly less coordinated in comparison to wild-type controls. This observation would require recording and careful analysis to firmly establish this phenotype (Nishii and Shibata, 2006). Together, these results are in agreement with the proposal that defective AVC closure in Hand2 and other mutants may be causally linked to impaired blood pumping and thus to their early embryonic lethality.

In conclusion, the cardiac failure and lethality of Hand2-deficient embryos is likely multifactorial, including atresia of the aortic arches arteries, massive cell death of SHF progenitors in the branchial arches, hypotrabeculation that reduce the myocardial surface and impairs oxygenation, and absence of cardiac cushions and possibly impaired cardiac contractions ((Srivastava et al., 1997; Aiyer et al., 2005) and this study). This plethora of defects illustrates the complexity of heart morphogenesis and of the regulatory networks controlled by HAND2 in this process.
7.2. Gene networks controlling endothelial-mesenchymal transition in the atrioventricular canal are directly regulated by HAND2

Previous studies have postulated a role for Hand2 in cardiac cushion formation but no direct involvement or mechanism was established (Camenisch et al., 2000; Liu et al., 2009; Holler et al., 2010; Vincentz et al., 2011; VanDusen and Firulli, 2012). The HAND2^3xFLAG ChIP-Seq analysis identified CRMs in the cis-regulatory landscapes of several genes with a role in the EndMT and differentiation of the AVC cardiac cushions. Our in-depth analysis of these interactions uncovered a gene network controlled by HAND2 that functions in the initiation of EndMT and cardiac cushion formation. The block in EndMT initiation has not been analyzed in previous studies and is morphologically strikingly similar to the consequences of inactivating other genes functioning in this process (e.g. Cspg2, which encodes the extracellular matrix protein Versican (Yamamura et al., 1997; Mjaatvedt et al., 1998)).

7.2.1. Hand2 participates in extracellular matrix deposition within the cardiac jelly

The deposition of extracellular matrix (ECM) in the cardiac jelly of the AVC is required for the migration of the mesenchymal cells of the cardiac cushions (Bernanke and Markwald, 1979). Versican and hyaluronic acid (HA) are both ECM components of the cardiac jelly. Inactivation of the HA producing enzyme Has2 also disrupts cardiac cushion formation and results in phenotypes similar to those observed in Hand2-deficient embryos (Camenisch et al., 2000). Our study establishes that Has2 expression in the AVC depends on Hand2, and that Has2 is a direct target of HAND2-containing chromatin complexes (Fig.27A). However, at this stage it is not clear how HAND2 complexes regulate Has2 expression in the developing heart. The CRM we identified in the first intron of Has2 drives reporter expression in the limb buds, suggesting that the previously described limb enhancer of Has2 located in its proximal promoter is part of a larger CRM (Liu et al., 2013).

We showed that the ECM in the cardiac jelly is reduced in Hand2-deficient mouse embryos, and several studies support an involvement of Hand2 in the composition and modeling of ECM components. Conditional deletion of Hand2 within the PEO and epicardium results in disorganization of Fibronectin (Fn1) fibrils in the epicardial ECM.
In Hand2-deficient zebrafish embryos, the Fibronectin network is also disorganized (Le et al., 2005). Detailed analysis suggests that Hand2 negatively regulate Fn1 expression, which allows cardiomyocyte movements required for fusion of the cardiac tube (Garavito-Aguilar et al., 2010). This is likely an indirect effect as Fn1 transcripts are not altered in Hand2-deficient mouse embryos (Barnes et al., 2011), and no evident HAND2-enrichments are detected around the Fn1 locus by ChIP-Seq analysis. In addition, Hand2 regulates the expression of matrix metalloproteinases (MMP), which in turn degrade laminin and enable the migration of lateral plate mesodermal cells and leftward gut looping in zebrafish (Yin et al., 2010). In the mouse, MMPs belonging to the ADAMTS family function in the formation of the cardiac jelly in the AVC by cleaving Versican in the cells that undergo EndMT (Kern et al., 2006). Haploinsufficiency for Adamts9 causes valves malformations (Kern et al., 2010), and interestingly, several members of the ADAMTS family contain one to several regions enriched in the HAND2^3xFLAG^ ChIP-Seq analysis. However, these HAND2 associated regions have not yet been analyzed to establish if HAND2 acts indeed upstream of ADAMTS metalloproteinases in mouse embryos.

Together, these data indicate that Hand2 participates in regulating the ECM composition in vertebrate embryos. In particular, HAND2 may directly participate in remodeling the ECM components of the AVC cardiac jelly, which in turn contribute to the migration of the mesenchymal cells of the cardiac cushions. In addition, it would be interesting to determine if the functions of Hand2 in controlling ECM components/regulators are also relevant in other embryonic and adult tissues or in pathological situations such as cancer (Lu et al., 2012).

7.2.2. HAND2 directly regulates Twist1 and Snai1 expression during formation of the AVC cardiac cushions

In addition to the ECM formation, the secondary EMT of epicardial cells is also affected by the inactivation of Hand2 (Barnes et al., 2011). Therefore, we decided to investigate other potential direct targets of HAND2 that are required for the EndMT in the AVC. In particular, Twist1, Sox9 and Snai1 function at crucial steps of the AVC cardiac cushion morphogenesis (Vrljicak et al., 2012; Akiyama et al., 2004; Timmerman et al., 2004). Our ChIP-Seq analysis identified potential CRMs in all of their cis-regulatory
landscapes. Indeed, the expression of both Twist1 and Snai1 is lost from the AVC in Hand2-deficient embryos, while Sox9 expression remains normal. In agreement, ChIP-qPCR from embryonic hearts confirmed that HAND2 chromatin complexes interact with CRMs within the Twist1 and Snai1 but not Sox9 cis-regulatory landscapes, which allowed us to conclude that Twist1 and Snai1 are direct transcriptional targets of HAND2 (Fig.27A).

TWIST1 is expressed in the AVC upon delamination of endocardial cells, most likely as a consequence of its regulation by HAND2. Previous studies have shown that Twist1 regulates the expression of genes involved in the proliferation of AVC cardiac cushion cells but is, on its own, not required for their formation (Chakraborty et al., 2010; Vrljicak et al., 2012; Vincentz et al., 2008). However, the migration of the cNCC that populate the OFT cardiac cushions is impaired or delayed in Twist1-deficient embryos. These cNCC form nodules that express Hand2 inside the mesenchyme of the cardiac cushions (Vincentz et al., 2008). Interestingly, Hand2 is required for the migration of the cNCC in the OFT (Holler et al., 2010). As our study establishes direct transcriptional control of Twist1 by HAND2, this regulation could also be operational in migratory cNCC. In fact, previous analysis already suggested that the absence of EndMT and cardiac cushion phenotype in Twist1-deficient embryos could be due to functional redundancy with Hand2 (Vincentz et al., 2008). TWIST1 forms heterodimers with HAND2 and other bHLH TFs, and these interactions are regulated by phosphorylation that dictate the DNA binding specificity of the transcriptional complexes (Firulli et al., 2005; Firulli et al., 2007). The Hand2*FLAG allele could be a suitable tool to assess this protein-protein interaction by co-immunoprecipitation of the transcriptional complexes from developing hearts followed my Mass Spectrometry. Finally, we have initiated transgenic reporter analysis to determine the activity of the Twist1 +130kb candidate CRM. We expect this element to be active in the cardiac cushions of the AVC, but if LacZ reporter expression is also detected in the cushions of the OFT, this would support the proposal that Twist1 expression in the cNCC is regulated by HAND2.
Figure 27. A model of the regulatory network controlled by Hand2 during endothelial-mesenchymal transition and cardiac cushion formation in the atrioventricular canal
(A) Model of HAND2 functions during cardiac cushion formation. Snai1, Has2 and Twist1 are newly identified direct targets of HAND2 during initiation of EndMT, migration of delaminated cells and/or cardiac cushion proliferation. (B) Schematic representation of the gene whose expression is altered in Hand2-deficient embryos (grey), and (C) proposed model of HAND2 as an intermediate relay in the signaling interactions between Notch1, Bmp2 and Tgfβ2. Dashed green arrows indicate interactions that remain to be tested or confirmed.

Genetic targets of TWIST1 in the cardiac cushions of the AVC include Bmp2 and Sox9, whose expression is reduced in the AVC of Twist1-deficient embryos (Vrljicak et al., 2012). As Sox9 expression is initiated in endocardial cells of Hand2-deficient embryos, it suggests that TWIST1 upregulates Sox9 expression in cells after delamination for
proper cardiac cushion remodeling but is not necessary for Sox9 activation (Lincoln et al., 2007).

SNAIL1 is a prototypic transcription factor involved in many different physiological and pathological EMT processes by controlling genes that induce changes in cell shape, adhesion and motility (Barrallo-Gimeno and Nieto, 2005; Thiery et al., 2009). In particular, SNAIL1-mediated downregulation of E-Cadherin genes is a key event in the initiation of EMT (Batlle et al., 2000; Cano et al., 2000). The formation of the cardiac cushions is no exception, as SNAIL1-mediated downregulation of the vascular E-Cadherin (VE-Cadherin) is necessary for progression of the endocardial EMT (Timmerman et al., 2004; Vestweber, 2008). In parallel, SNAIL1 upregulates Vimentin expression to increase the motility of the delaminating cells (Cano et al., 2000; Ivaska, 2011; Mendez et al., 2010). In addition to the total absence of Snai1 in the AVC, VIM expression is not activated and VE-Cad junctions are maintained in the endocardium of Hand2-deficient embryos. These results indicate that Hand2 is required for Snai1 activation and initiation of the EndMT. Indeed, HAND2 chromatin complexes interact with two candidate CRMs in the Snai1 cis-regulatory landscape. As Snai1 acts cooperatively with Snai2 to induce the EndMT (Niessen et al., 2008), we also looked into a potential direct regulation of Snai2 by HAND2 but no experimental evidence in favor of such direct regulation were obtained (data not shown).

Despite Snai1 is expressed in the premigratory and migratory neural crest cells, it is not required for their initial delamination (Sefton et al., 1998; Murray and Gridley, 2006). The Snai1 CRM we identified by ChIP-Seq analysis drives LacZ reporter expression in migratory neural crest cells, craniofacial structures, and in the epicardium. Indeed, Snai1 is important both for the development of craniofacial structures and for the EMT of epicardial cells (Cordero et al., 2011; Murray et al., 2007; Tao et al., 2013). Interestingly, Hand2 is also expressed in craniofacial neural crest derivatives and regulates the formation of the palate, mandible and tongue (Charité et al., 2001; Yanagisawa et al., 2003; Funato et al., 2009; Barron et al., 2011; Xiong et al., 2009). Hand2 is also involved in the EMT of epicardial cells and their differentiation into coronary vasculature (Barnes et al., 2011). We have thus identified a CRM active in
several embryonic tissues where both *Snai1* and *Hand2* are expressed, suggesting that *Snai1* might be a universal target of HAND2. The precise role of this *Snai1* CRM during cardiac cushion formation and in other biological contexts remains to be investigated.

To gain further insight into the roles of HAND2 in the EndMT, we will use AVC explant culture, which has been successfully used to determine the role of different TFs in EndMT and cardiac cushion formation (Runyan and Markwald, 1983; Camenisch et al., 2000; Akiyama et al., 2004; Timmerman et al., 2004; Luna-Zurita et al., 2010; Moskowitz et al., 2011). This assay allows dissection of the molecular mechanisms and interactions that control the formation of cardiac cushions by analyzing the migratory and invasive capacities of the endocardial cells. Initially, we will compare wild-type versus *Hand2<sup>Δ/Δ</sup>* AVC explants in culture. In addition, *Hand2* could be specifically inactivated in explants using *Adeno-Cre* viral particles in combination with the *Hand2* conditional allele, and we could also attempt rescue experiments by re-expressing *Hand2* using viral vectors as well. Adding hyaluronic acid to *Hand2*-deficient cultured explants could provide insights into the extent to which Has2-mediated ECM defects contribute to the failure to form the cardiac cushions in these mutants.

### 7.2.3. HAND2 is an intermediate factor in the myocardial/endocardial signaling events that induce the EndMT

To determine the possible involvement of myocardial signals in the disruption of *Snai1*-mediated EndMT in *Hand2*-deficient embryos, we looked at the expression of *Bmp2*, *Tgfβ2*, *Notch1* and other key genes involved in the myocardial-endocardial crosstalk that patterns the AVC and initiates the EndMT. Our results indicate that the AVC myocardium is correctly patterned in the absence of HAND2, and that *Hand2* acts downstream of the relevant signals and upstream of activating *Snai1*. Furthermore, this gene expression analysis allowed us to make some assumptions regarding the exact position and role of HAND2 in the signaling network that control the EndMT in the AVC.

The loss of the BMP2 target *Msx1* in the endocardium shows that the absence of *Hand2* alters BMP2 downstream signaling (Fig.27B). Endothelial deletion of the BMP signal transducer *Smad4* results in defective EndMT and in acellular cardiac cushions,
as in Hand2-deficient embryos (Moskowitz et al., 2011). Furthermore, Smad4 genetically interacts with Gata4 in the endocardium, and both proteins can heterodimerize and cooperatively activate the expression of target genes in the AVC endocardium (Moskowitz et al., 2011). The new Smad4^{3xFLAG} allele that I generated could thus be used to study the impaired signaling activity of BMP2 in the endocardium of Hand2-deficient embryos (Fig.27C). We are currently generating the compound Hand2^{Δv};Smad4^{3xF-HR/+} mice to assess this hypothesis.

In the endocardium, Notch1 signaling represses Bmp2 expression through Hey1 and Hey2, which in turns prevents endocardial cells to acquire a cardiomyocyte cell fate (Luna-Zurita et al., 2010). In Hand2-deficient embryos, Notch1 and Hey2 are expressed normally and there is no ectopic activation of Bmp2 in the endocardium. This indicates that Hand2 is not part of the Notch1-HEY1/2-BMP2 endocardial signaling module.

The myocardial expression of Tgfβ2 depends on both Notch1 signaling through a postulated retrograde endocardial to myocardial signal (Timmerman et al., 2004), and on BMP2 signaling through Tbx2 (Ma et al., 2005; Shirai et al., 2009). In addition, Notch1 promotes paracrine endocardial to myocardial Wnt4 signaling that in turns regulates Bmp2 expression (Wang et al., 2013). In Hand2-deficient embryos, the expression of Tgfβ2 but not of Notch1, Bmp2 and Tbx2 is lost (Fig.27B). This suggests that Tgfβ2 is regulated indirectly by Hand2 downstream of Notch1. In this model, Hand2 would occupy an intermediate signaling role between Notch1 and the other postulated endocardial signal that regulates Tgfβ2 expression, rather than in the Notch1-Wnt4-BMP2-TBX2 signaling module (Fig.27C).

Finally, Snai1 expression in the endocardium and subsequent EndMT depends on the synergic action of Notch1, TGFβ2 and BMP2 signaling (Timmerman et al., 2004; Niessen et al., 2008; Luna-Zurita et al., 2010; Garside et al., 2013). Our data established that HAND2 is also directly required for the activation of Snai1 expression. To determine whether HAND2 acts in parallel or as an intermediate transcriptional relay in these signaling interactions will require further analysis of embryos deficient for some of the other key genes discussed here (Fig.27C).
7.2.4. A role for Hand2 in cardiac valves maturation and remodeling?

The cardiac valves are derived from the cardiac cushions. The ChIP-Seq analysis identified numerous potential CRMs associated with genes that function in cardiac valve remodeling. However, the lethality of Hand2-deficient embryos prior to the stage of valve remodeling precludes the analysis of Hand2 functions during cardiac cushion maturation. Nevertheless, we established that HAND2 chromatin complexes bind a cardiac enhancer of Tbx20, which controls proliferation of the valve endothelial cells during valve elongation (Fig.27A) (Cai et al., 2013). Other likely direct targets of HAND2 with ChIP-Seq peaks in their cis-regulatory landscapes include Col1a1, which is important for valves structure and maintenance, and Scleraxis, required for differentiation and ECM organization during valve remodeling (Peacock et al., 2008; Levay et al., 2008). The conditional removal of Hand2 from the mesenchymal cells of the cardiac cushions from E10.5 does not affect cardiac valves maturation (VanDusen et al., 2014). Therefore, the requirement of Hand2 for EndMT initiation is a critical step, whereas Hand2 appear to be dispensable during cardiac cushion remodeling, maybe due to a compensatory mechanism mediated by Twist1 (see above).

7.3. The paralogous Tbx2 and Tbx3 genomic landscapes encode heart and limb bud specific CRMs bound by HAND2 chromatin complexes

The HAND2-3xFLAG ChIP-Seq analysis identified several candidate CRMs in the cis-regulatory landscapes of the T-box genes Tbx2 and Tbx3. Using these candidate CRMs in LacZ reporter assays, we identified one heart and one limb bud specific CRM within the Tbx3 ~160kb cis-regulatory landscape. In addition to the heart and limb buds, this large landscape is able to activate reporter expression in the eye, snout and mammary glands (Horsthuis et al., 2009). As Hand2 is not expressed in these other structures where the two identified CRMs are also not active, it appears that Tbx3 expression is regulated by an archipelago containing several other tissue-specific CRMs. Indeed, a recent study used chromatin architecture profiling from embryonic hearts and established that the Tbx3 promoter physically contacts the HAND2-interacting cardiac CRM and an additional pan-cardiac enhancer (van Weerd et al.,
Both CRMs synergize to activate Tbx3 expression in the atrioventricular conduction system. In addition, the cardiac-specific CRM that we identified by HAND2^{3xFLAG} ChIP-Seq analysis requires the synergic action of GATA4 and SMADs to be active (van Weerd et al., 2014). This CRM is able to drive LacZ reporter expression in the myocardium of the AVC, where the atrioventricular node will form. As HAND2 is not expressed in these cells, it is possible that this CRM is repressed by HAND2 chromatin complexes. In addition, the ectopic expression of Tbx3 at early stages in Hand2-deficient embryos argues in favor of a general repressive effect of HAND2 on this gene during cardiac development. The additional HAND2-independent CRM(s) accounting for robust expression of Tbx3 could explain why genetic inactivation of Hand2 does not result in ectopic activation of Tbx3 expression at the onset of cardiac conduction system formation, but only at earlier stages.

The Tbx3/Tbx5 and Tbx2/Tbx4 landscapes are likely similar, as they evolved by two rounds of gene duplication from a single ancestral gene (Fig.28A) and (Agulnik et al., 1996). Indeed, we also detected two CRMs interacting with HAND2 chromatin complexes in the Tbx2 landscape. The activity of one of them is currently being analyzed, while the other drives LacZ reporter expression in the posterior limb bud mesenchyme (Infante et al., 2013). An additional HAND2^{3xFLAG} ChIP-Seq peak located closer to Tbx4 was reported to be a Tbx4 hindlimb enhancer (Menke et al., 2008). When we compare the HAND2 ChIP-Seq profiles for the Tbx2/Tbx4 and Tbx3/Tbx5 cis-regulatory landscapes, we observe a similar distribution that argues in favor of similar cardiac and limb bud CRM activities in both landscapes (Fig.28B). However, the functions of Tbx5 and Tbx4 have diverged as they are specifically functioning in forelimb and hindlimb induction, respectively. The amphioxus ancestral Tbx4/5 gene lacks the CRM required for limb initiation but is able to initiate limb formation in mouse transgenic embryos under control of a limb mesenchymal promoter. This indicates that changes in the CRMs rather than gene functions underlie the specialization of Tbx4 and Tbx5 during limb evolution (Minguillon et al., 2009). It would be interesting to study the evolutionary and functional relationship of the cardiac CRMs of Tbx2 and Tbx3, as gene duplications can also result in the acquisition of new structures and functions in the developing heart (Olson, 2006). The orthologous human and chicken sequences of
the CRM of Tbx3 that interacts with HAND2 complexes in the heart display similar reporter activity as their murine counterpart in the AVC. This indicates evolutionary conserved functions of this CRM during Tbx3-mediated formation of the conduction system (van Weerd et al., 2014).

During cardiac development, Tbx2 and Tbx3 are both upregulated in the SHF progenitors of Hand2-deficient embryos, while they are downregulated in limb buds (Osterwalder et al., in press). In contrast, the expression of Tbx4 and Tbx5 is not changed in any of the tissues. This shows that HAND2-mediated regulation of Tbx2 and Tbx3 is common between the two clusters, whereas this is not the case for Tbx4 and Tbx5. Indeed, the chromatin architecture profile of the Tbx3/Tbx5 landscape clearly establishes that these two genes are regulated independently (van Weerd et al., 2014).

In summary, the cis-regulatory mechanisms controlling Tbx2 and Tbx3 expression are shared between cardiac and limb tissues. These mechanisms involve the action of the common trans-acting factor HAND2 on tissue-specific CRMs embedded in cis-regulatory landscapes of likely similar architecture due to a common evolutionary origin.
8. Conclusions and Outlook

The two major aims of this thesis were (1) to generate a new genetic tool to identify the transcriptional networks controlled by SMAD4 during mouse limb bud development, and (2) to analyze the transcriptional networks controlled by HAND2 during mouse cardiac development. Both projects involved the use of endogenously tagged versions of these two transcription factors.

The mouse line expressing the Smad4^{3xFLAG} allele that I generated and characterized is now ready to be used for ChIP-Seq analysis. This new allele will allow us identifying the direct targets of SMAD4 chromatin complexes in developing limb buds and in other tissues such as the embryonic heart. In particular, the dynamics of BMP signaling activity during the different phases of limb bud development and the consequential transcriptional response of its target genes can now be monitored in vivo by assessing the temporal occupancy of SMAD4-containing chromatin complexes. This will provide us more insights into the downstream molecular circuitry that controls initiation of limb bud outgrowth, patterning, and chondrogenic differentiation. Moreover, the question of how BMP signaling regulates the expression of Grem1 can now be investigated at the transcriptional level by the identification of the Grem1 CRMs that interact with SMAD4 complexes. The ChIP-Seq data that this new mouse line will allow us to generate will also be intersected with ChIP-Seq data from additional transcription factors (e.g. GLI3), in order to understand how cis-regulatory landscapes integrate several signaling inputs and how it affects their transcriptional output.

The analysis of the HAND2^{3xFLAG} ChIP-Seq data allowed me to uncover new direct transcriptional targets of HAND2 in the developing heart, such as Furin, Tbx2 and Tbx3. In addition, I identified a defect during the endothelial-mesenchymal transition in the atrioventricular canal of Hand2-deficient embryos and could establish that HAND2 is a key regulator of cardiac cushion formation by directly controlling the expression of Snai1. The functional analysis of the HAND2 target CRMs that I identified in the cis-regulatory landscapes of Snai1, Twist1 and Has2 is still ongoing
and could help us to determine if HAND2 controls EMT in additional developmental/pathological processes. Furthermore, the future use of AVC explant cultures will allow us to study the role of HAND2 during EndMT and cardiac cushion formation at the cellular level, and the conditional deletion of Hand2 in the endothelium using Tie2-Cre will also enable us to analyze further this EndMT phenotype. Finally, a cardiac-specific ChIP-Seq analysis will be an important complement to this study in order to identify additional CRMs that interact with HAND2 specifically during heart development.

The data presented in my thesis constitute the first identification of the HAND2 direct transcriptional targets during cardiogenesis. The use of an endogenously epitope-tagged form of HAND2 illustrates the power of such genetic tools to understand the functions of transcription factors during embryonic development.
9. Materials and Methods

9.1. Genetic engineering of plasmids

9.1.1. General cloning procedures

DNA fragments used to construct plasmid vectors were either recovered from restriction enzyme digested vectors or BACs; or produced by PCR using the Expand High FidelityPLUS PCR System (Roche) with specific primers to introduce the desired restriction enzyme recognition sites at both ends. Following digestion of plasmids/BACs and inserts with compatible restriction enzymes (from Roche or NEB), fragments of interest were recovered from agarose gels and purified using QIAquick Gel Extraction Kit (Qiagen). De-phosphorylation was performed with the rAPid alkaline phosphatase kit (Roche). Ligation of compatible plasmids and inserts was done for 15min-1hr at room temperature (RT) using T4 Ligase (NEB). Then the ligated mixture was diyalised using MF™ nitrocellulose membrane filters (Millipore). Electrocompetent bacteria (XL1-Blue, XL10-Gold or DH5a) were electroporated with ligated vectors in Gene Pulser 0.2cm cuvettes using the Biorad MicroPulser™. Afterwards, bacteria were plated on LB Agar plates containing the appropriate antibiotic selection. Plasmid DNA was purified by alkaline lysis using the Plasmid Midi Kit (Qiagen) or the NucleoBond® Xtra Midi kit (Machery Nagel) from bacteria grown in liquid cultures. Correct ligation and orientation of the insert was assessed by restriction digest or PCR. Alternatively, colony lifts onto nylon membrane were used and positive clones identified using a DIG-dUTP labeled probe. All construct inserts were sequenced (Microsynth).

To generate linkers, 100nM of each oligonucleotide were mixed in 1x annealing buffer (0.1M NaCl, 10mM Tris-HCl pH 7.4) in 50µl and transferred to a water bath at 85-95°C on a floating rack for annealing. Annealed linkers were cooled gradually to RT in H₂O for 3-4hrs. For ligation, serial dilutions of 1µl annealed linkers (1:100, 1:1000 or 1:5000) in 0.5x annealing buffer were used. Plasmids were not de-phosphorylated for ligation of linkers, as oligonucleotides do not harbor a 5’ phosphate.
9.1.2. Construction of pBSK Smad4<sup>3xFLAG</sup> targeting vector for Homologous Recombination

The targeting vector was designed to introduce a sequence encoding the 3xFLAG epitope tag in frame into the 3′ end of Smad4 coding sequence. Additional modifications included a STOP codon after the 3xFLAG, which was followed by the endogenous Smad4 3′UTR and a SV40 polyadenylation site. A Neomycin selection cassette flanked by loxP sites was also inserted 366bp downstream of Smad4 locus as part of the 3′ homology arm. Firstly, 5′ and 3′ homology arms were recovered from pBACe3.6 encoding the bMQ420b13 clone, which covers the Smad4 coding sequence. The 5′ arm was subcloned into pBluescript II KS (pBSK, Stratagene) as an Apal-Apal 4277bp fragment encoding a part of Smad4 locus covering the intron 10/11, exon 11, intron 11/12 and exon 12. The 3′ arm was inserted into the plasmid pcDNA3 (Invitrogen) as a BglII-BglII 6045bp fragment, which encodes a part of the 3′UTR and the downstream intergenic region. The fragment containing the 3′ end of exon 12, the 3xFLAG-STOP and the 5′ part of the 3′UTR was obtained from a previously assembled pXCH-Smad4<sup>3xFLAG</sup> vector (data not shown) as a 295bp Apal-BglII fragment. The floxed Neo cassette was recovered from the pLoxpNeo-2 vector (Michos et al., 2004) as a 2026bp EcoRI-Xhol fragment that was blunted. To generate the targeting vector, these DNA fragments were assembled in pBSK using a linker containing the following restriction sites: Asp718-MluI-Apal-BglII-BamHI-BclI-MluI-Sacl. The Asp718 and Sacl restriction sites in pBSK were destroyed by insertion of the linker. The BamHI site was used for screening purposes. The sequence of this linker is displayed in the Primers Tables in the chapter 9.7.2. DNA fragments were assembled sequentially as follows: first the Apal-FLAG-BglII fragment, then the BglII-3′arm-BglII fragment from pcDNA3 (inserted in the compatible BclI site on the linker), then the Apal-5′arm-Apal insert from pBSK, and finally the blunt Neo cassette in reverse orientation into a BmgBI site located in the 3′arm. For electroporation, the targeting fragment was released from its pBSK backbone as a 12′344bp MluI-MluI fragment (see Fig.29 for a detailed scheme).
Figure 29. Generation of the pBSK-Smad4\(^{3xFLAG}\) targeting vector for homologous recombination

(A) 5’ and 3’ homology arms were recovered from pBACe3.6 clone bMQ420b13. (B) A fragment containing the 3’ end of exon 12, the 3xFLAG-STOP and the 5’ part of the 3’UTR was obtained from a previously assembled pXCH-Smad4\(^{3xFLAG}\) vector. (C) The floxed Neomycin cassette was recovered from the pLoxpNeo2 vector and blunted. The targeting vector was assembled in pBSK template vector using a linker containing the following restriction sites: Asp718-MluI-MluI-ApaI-BglII-BamHI-MluI-SacI. Fragments were cloned in the following order: (D) 3xFLAG containing fragment, (E) 3’arm, (F) 5’arm and (G) Neo cassette in reverse orientation. (H) For electroporation, the targeting fragment was released from pBSK as a MluI-MluI fragment. See main text for details.

9.2. Cell culture

9.2.1. Embryonic Fibroblasts (EMFIs) culture

EMFIs were used as feeder cells for embryonic stem cells and grown at 37°C in 5% CO\(_2\). EMFI medium consists of DMEM containing 4.5g/l Glucose (Gibco 41966-029), 10% FCS (Gibco 10270-106), 100U Penicillin, 0.1mg/ml Streptomycin (Pen Strep, Gibco 15140-122) and 2mM L-Glutamine (Gibco 25030-024). A frozen vial of EMFIs was thawed at 37°C in a water bath and diluted in a 15ml Falcon tube containing 10ml of EMFI medium. After centrifugation at 1200rpm for 5min, the medium was removed by aspiration and the EMFIs resuspended in 3ml prewarmed medium. Cells
were distributed equally to five 10cm culture dishes (BD Falcon). Medium was changed every two days. When EMFIs became confluent, they were split using 0.05% Trypsin EDTA (Gibco 25300-054) onto another plate (for a maximum of two passages) or treated with Mitomycin C (Sigma M-0503) to stop cell division. For trypsinization, EMFIs were rinsed with 4ml prewarmed trypsin and incubated in 3ml trypsin for 5min at 37°C. Single cells were collected in 7ml EMFI medium and spun for 5min at 1200 rpm. Cell pellets were resuspended in a small volume, diluted and seeded at 1:4-1:6 in 10cm dishes. Mitomycin C at 1mg/ml in DPBS (Gibco 14190-094) (stored at 4°C in the dark for ≤ two weeks) was diluted to 10µg/ml in EMFI medium for treatment of cells. EMFIs plates were incubated at 37°C with 5% CO₂ for 2h, washed 3x with PBS and transferred back to the incubator in fresh EMFI medium. Growth-arrested EMFIs were kept in culture for maximally one week and their medium was changed every two days.

9.2.2. Embryonic Stem (ES) cell culture
ES cells were grown at 37°C with 7.5% CO₂ on a layer of growth arrested EMFIs. ESC medium consists of Knockout DMEM (Gibco 10829-018) supplemented with 15% Hyclone FBS (Thermo Scientific), 100U-0.1mg/ml Pen Strep, 2mM L-Glutamine, 1mM β-mercaptoethanol (Gibco 31350-010), 1x Non-essential Amino Acids (Gibco 11140-035), 1mM Sodium Pyruvate (Gibco 11360-039) and 10⁵ units/ml ESGRO LIF (Millipore ESG1106). The medium was changed every day, and the ES cells split every two days using 0.05% Trypsin EDTA. After a brief wash in 4ml trypsin, cells were incubated in 3ml trypsin for 15min at 37°C. A single-cell suspension was obtained by gentle pipetting, and ES cells were separated from EMFIs by incubating the cells at 37°C for 15min. While EMFIs adhere to the surface of the dish, the floating ES cells were carefully recovered and transferred to a 15ml Falcon tube. After 5min centrifugation at 1200rpm, the ES cell pellet was resuspended in ES cells medium and split to 1:3-1:10 on growth arrested EMFIs. Alternatively, ES cells were grown on gelatin instead of EMFIs. A solution of 0.1% gelatin in H₂O (Sigma G-2500) was used to coat culture dishes for 10min at RT. The gelatin solution was removed and the gelatin-coated dishes dried for 10min at RT before use. For freezing, trypsinized ES cell pellets were resuspended in ice-cold filtered freezing medium
(50% ES cells medium 40% FCS, 10% DMSO (Sigma D-8418)). 5-6 precooled Cryovials (NUNC) were filled with 1.5ml of ES cells recovered from a 10cm culture dish. Cryovials were transferred to -80°C in a Styrofoam box and transferred to liquid N₂ the next day.

9.2.3. ES cells electroporation and selection

R1 ES cells (129 background) were thawed 6 days before electroporation and split twice, on day 2 and day 4. On day 6, the ES cells were trypsinized, pooled in 50ml Falcon tubes and centrifuged for 5min at 1200rpm. Then ES cells were resuspended in 3ml medium and counted using a Neubauer chamber. After re-centrifugation, the cells were resuspended in DPBS (Gibco 14190-094) at a final concentration of 1.875 x 10⁷ cells/ml. For electroporation, 35µg of linearized targeting vector (pBSK-Smad4₃xFLAG) were added to Gene Pulser cuvettes (Biorad 165-2088) and then 0.8ml ES cells suspension was added (1.5 x 10⁷ cells). Water was used as a control instead of DNA in some ES cell samples. Electroporation was done using Biorad Gene Pulser II set at 0.24kV and 475µF capacitance. Cuvettes were immediately transferred on ice and incubated for 20min after electroporation. The contents of the cuvettes were then pooled and transferred to a 50ml Falcon tubes containing prewarmed ES cell medium, and the cuvettes rinsed with ES cell medium to recover all cells. Medium was added to a total volume of 20ml, and the ES cell suspension was gently mixed before plating at 1:5 or 1:10 dilutions on gelatin-coated dishes containing ES cell medium. Non-electroporated ES cells were diluted and plated at similar dilutions as control. Medium was changed after 24h, and antibiotic selection started after 48h. For this, 190µg/ml of G-418 Geneticin (Gibco 10131-019) were used in ES cell medium with medium change everyday during 11-13 days. One dish for each condition/dilution (+DNA, -DNA, no electroporation) was maintained in culture medium without G-418 as selection control. In these dishes, the ES cells grew normally. In the dishes of ES cells under selection that were electroporated without DNA, no ES cell clones were observed. Following selection, ES cell clones were picked and transferred on growth arrested EMFIs in 48 well plates. The ES cell selection medium was changed everyday. Clones were passaged after two days. After two additional days, the clones were trypsinized and half of the cells transferred
to a fresh 48 well plate coated with gelatin, while the other half were plated on growth arrested EMFIs. Upon reaching confluency, the cells grown on gelatin were used to extract DNA for screening, while the plates with cells grown on EMFIs were frozen in freezing medium after two additional days and stored at -80°C. Screening of the correctly targeted clones was done by Southern Blotting (see below). Then, the plates containing positive clones were carefully thawed at 37°C and the medium was changed several times to remove all traces of DMSO. The recovery and growth of each clone was then individually monitored and the medium changed every day for 3-4 days until growing colonies became visible. Clones were passaged to larger wells and finally to 10cm dishes to freeze several vials per expanded ES cell clone.

9.2.4. ES cell clones screening by Southern Blot

Extraction of DNA from confluent ES cells grown on gelatin-coated 48 multiwell plates was achieved as follows: each well was gently washed 3 times with 750µl PBS without disturbing the monolayer of cells, and 250µl lysis buffer (100mM Tris-HCl pH 8.5, 5mM EDTA pH 8.0, 0.2% SDS, 200mM NaCl, 100µg Proteinase K) were then added in each well. The plates were incubated overnight at 55°C in a moist chamber. The next day, 500µl of absolute EtOH were added dropwise to the sides of the wells. The plates were incubated overnight at RT, which resulted in appearance of white strands of DNA. The plates were then carefully inverted to remove the EtOH and each well gently washed 3 times with 750µl of 70% EtOH. Finally, plates were dried overnight upside-down and then stored at RT. For Southern Blot screening, such precipitated DNAs were digested overnight in a volume of 100µl with 30 units of the appropriate restriction enzyme. The next day the reaction was spiked with 10 units and digested for additional 7hrs. Half of the content of the well was then loaded on a 0.7% agarose gel containing 5% EtBr and ran in 1xTBE at 60V for ~24h until appropriate separation was achieved. The agarose gel was then depurinated in 0.25M HCl for 10min; rinsed in H2O; washed 2 times 15min in denaturation solution (0.5M NaOH, 1.5M NaCl); rinsed in H2O; washed 2 times 15min in neutralization solution (0.5M Tris-HCl pH 7.5, 1.5M NaCl); rinsed in H2O and finally equilibrated for at least 15min in 20xSSC (3M NaCl, 150mM sodium citrate, pH 7.0). The separated DNA fragments were then transferred onto a positively charged nylon membrane
(Roche) by capillary transfer. The membrane was pre-soaked in H\(_2\)O and equilibrated in 2xSSC, and the transfer was done overnight in 20xSSC. Following the transfer, the membrane was soaked in 2xSSX for 5min and dried for 1h at RT on a sheet of Whatman 3MM paper (Sigma). The membrane was subsequently baked for 2hrs at 80°C in a sheet of 3MM paper, and stored at RT. Prehybridization of the membrane was done using a rolling bottle system with side DNA facing inside at 42°C for ≥1hr in prewarmed prehybridization solution (5xSSC, 50% deionized formamide, 0.1% N-lauroylsarcosine, 0.02% SDS, 2% blocking solution). The blocking solution was prepared as 10% (w/v) Blocking Reagent (Roche) in 1x maleic acid buffer (0.1M C\(_4\)H\(_4\)O\(_4\) 0.15M NaCl, pH 7.5) at 60°C. The prehybridization solution was replaced by prewarmed hybridization solution containing 4µl of DIG-dUTP labeled probes generated using the PCR DIG probe synthesis kit (Roche). Probes detected correct recombination of both the 5' and 3' homology arms. The presence of a single insertion was assessed by detection of the Neomycin cassette (see Fig.11A for the location of the probes). Clones were first screened using the 5' probe, and positive clones further screened with the 3' and Neo probes. The sequences of the primers used to generate these probes are shown in the chapter 9.7.2. Hybridization was done overnight at 42°C in rolling bottles. The next day, the membrane was washed twice for 15min with 2xSSC, 0.1% SDS at RT and twice for 15min with 0.2xSSC, 0.1% SDS at 68°C in rolling bottles. The membrane was then rinsed in maleic acid wash buffer (0.1M C\(_4\)H\(_4\)O\(_4\) 0.15M NaCl, 0.3% Tween-20, pH 7.5) and blocked with 1% blocking solution in maleic acid buffer for 30min-3hrs. Subsequently, the membrane was incubated in 1% blocking solution in maleic acid buffer containing 1:20'000 anti-digoxigenin-AP antibody (Roche) for 30min at RT. The membrane was then transferred to a plastic tray and washed 3 times for 15min in maleic acid wash buffer, and equilibrated in detection buffer (0.1M Tris-HCl, 0.1M NaCl, pH 9.5) for 2min at RT. The specifically hybridized bands were detected as follows: the membrane was placed on an overhead foil and covered with detection buffer that contained 4µl/ml of CDP Star (Roche) for 5min at RT. The membrane was then exposed for 10min to Kodak BioMax MR films or Amersham Hyperfilm™ ECL (GE Healthcare) and developed using the Fujifilm FPM-100A Desktop Processor. To allow re-hybridization to another probe, the membrane was washed extensively with H\(_2\)O in
a plastic tray, transferred back to a rolling bottle and washed 2x for 15min with 0.2M NaOH, 0.1% SDS at 37°C. After rinsing in 2xSSC, the membrane was stored at 4°C in 2xSSC prior re-incubation in pre-hybridization solution.

9.3. Mice husbandry and experiments with embryos

9.3.1. Ethic statement
All experiments conducted with mice were performed with strict respect of the Swiss law, 3R principles and Basel Declaration. They were classified as grade 0, implying minimal suffering.

9.3.2. Mouse strains
The following mouse strains were used for these studies: GRS1 transgene (Zuniga et al., 2012), Grem1 null (Michos et al., 2004), Shh loss-of-function (St-Jacques et al., 1998), Gli3 extra-toes-J (Maynard et al., 2002) and Hand2\textsuperscript{3xF} (Osterwalder et al., in press) alleles were maintained in an NMRI background; the Smad4 conditional allele (Yang et al., 2002), Prx1-Cre transgene (Logan et al., 2002), CMV-Cre transgene (Schwenk et al., 1995) and Dre transgene (Anastassiadis et al., 2009) were maintained in a C57BL/6 background; and the Smad4\textsuperscript{3xF} and Hand2 null alleles (Galli et al., 2010) in a mixed background. All mice and embryos were genotyped by PCR for relevant alleles using primers listed in the chapter 9.7.2.

9.3.3. Generation of \textit{LacZ} reporter transgenic embryos
Candidate CRM regions were either amplified by PCR from mouse genomic DNA (Has2 +2.4kb and Twist1 +130kb CRMs) or recovered as restriction fragment from a BAC (Snai1 +58/59kb CRM, BAC clone RP23-193B17) and cloned into a βglobin promoter-LacZ reporter-SV40 polyA vector. Transgenic embryos were generated by pronuclear injection at the Transgenic Mouse Core Facility (TMCF), Basel (Has2 CRM) or by Cyagen for the Twist1 and Snai1 CRMs. Founder embryos were stained for LacZ activity as described below. The transgenic embryos for the candidate CRMs of Tbx2 and Tbx3 were generated by M.Shoukry and A.Visel (Genomics Division, Lawrence Berkeley National Laboratory) as described by (Osterwalder et
al., in press). See chapter 9.7.2 for the genomic coordinates of the candidate CRM regions and for the primers used for their amplification by PCR.

9.3.4. Whole-mount LacZ staining of mouse embryos

Embryos were isolated in ice-cold PBS and staged by counting their somites numbers, before fixation in 1% formaldehyde, 0.2% glutaraldehyde, 0.02% NP-40, 0.01% sodium deoxycholate in PBS for 20-30 min at 4°C. After washing the embryos 3 times in PBS for 5 min at RT, they were incubated overnight at 37°C in the dark in 1 mg/mL X-Gal in dimethyl formamide, 0.25 mM K3Fe(CN6), 0.25 mM K4Fe(CN6), 0.01% NP-40, 0.4 mM MgCl2, 1% sodium deoxycholate in PBS. β-galactosidase activity results in expressing cells being colored in blue. The reaction was stopped by washing the embryos 3 times in PBS for 5 min at RT.

9.3.5. Whole-mount in situ hybridization (WISH)

Embryos isolated in ice-cold PBS were transferred in 2 ml Eppendorf tubes and fixed overnight in 4% paraformaldehyde (PFA) in PBS at 4°C. The next day they were washed 2 times in PBS-0.1% Tween-20 (PBT) and subsequently dehydrated in a gradient series of 25%, 50%, 75% MetOH/PBT and finally 2 times 100% MetOH. Such dehydrated embryos were stored at -20°C in MetOH. For the first day of the WISH, embryos were rehydrated in a reverse series of 75%, 50%, 25% MetOH/PBT followed by 2 washes in PBT (5 min each). Embryos were then bleached in 6% hydrogen peroxide (H2O2)/PBT for 15 min at RT and washed 3 times in PBT (5 min at RT). This was followed by treatment with 10 µg/ml proteinase K in PBT for 5 min (E8.5) to 15 min (E10.5), depending on the embryonic stage. It is important to not move the embryos during this step. Freshly prepared 2 mg/ml glycine in PBT was used to inactivate the proteinase K (5 min at RT). Then the embryos were washed twice in PBT for 5 min at RT. Embryos were refixed in fresh 0.2% glutaraldehyde, 0.1% Tween-20 in 4% PFA (in PBS) for 20 min at RT and washed twice in PBT. Then they were equilibrated in 2 ml prewarmed prehybridization buffer (50% deionized formamide, 5x SSC pH 4.5, 2% BCl blocking powder, 0.1% Tween-20, 0.5% CHAPS (Sigma), 50 µg/ml yeast RNA (Sigma R8759), 5 mM EDTA, 50 µg/ml heparin (Sigma H5515)) at 70°C for 1 hr or longer. The prehybridization buffer was then replaced by...
1ml of fresh prewarmed hybridization buffer containing 10µl/ml of digoxigenin-labelled RNA riboprobe (see below) and incubated overnight at 70°C.

On the second day, the probe-containing hybridization buffer was replaced by prewarmed prehybridization buffer, followed by a series of 75%, 50%, 25% prehybridization buffer in 2xSSC (0.3M NaCl, 0.03M sodium citrate pH 4.5). All these washing steps were performed for 5min at 70°C. Subsequently, the embryos were washed twice in 2xSSC, 0.1%CHAPS for 30min at 70°C. To remove unbound DNA probes, embryos were treated with 20µg/ml RNase A in 2xSSC,0.1% CHAPS for 45min at 37°C. Embryos were then washed twice with 100mM maleic acid disodium, 150mM NaCl pH 7.5 for 10min at RT and twice for 30min at 70°C. Three washes in freshly prepared TBST (140mM NaCl, 2.7mM KCl, 25mM Tris-HCl pH 7.5, 1% Tween 20) of 5min each at RT were then done. This was followed by blocking in 10% lamb serum/TBST for at least 1hr at RT. This blocking solution was replaced with a solution containing anti-digoxigenin-AP Fab fragments (Roche) conjugated to alkaline phosphatase. These Fab fragments were diluted 1:5000 in 1% lamb serum in TBST. Embryos were incubated overnight at 4°C with gentle rocking.

On the third day, embryos were washed 3 times in TBST for 5min at RT and 5 times in TBST for 1hr-1.5hrs at RT. This was followed by an overnight wash in TBST at 4°C. All washes were performed on a rocking platform. On the fourth day, the embryos were washed 3 times in freshly prepared NTMT (100mM NaCl, 100mM Tris-HCl pH 9.5), 50mM MgCl₂, 1% Tween-20) for 10min at RT and transferred into 1ml BM purple AP substrate (Roche) at RT. Development of the WISH was performed in the dark and monitored every hour until reaching appropriate probe-specific signal intensity. The development was stopped by washing 5 times in PBT (10min) and twice in PBS (5min). Finally, embryos were post-fixed in 4% PFA in PBS and stored at 4°C. WISH pictures were taken using a Leica MZ FLIII stereomicroscope and the Leica Application Suite V3 software.

### 9.3.6. Preparation of digoxigenin-labelled RNA riboprobe

To prepare riboprobes for in-situ hybridization, the plasmid containing the cloned cDNA of interest were linearized and transcribed using the appropriate T3, T7 or SP6 RNA polymerase in combination with the DIG RNA labeling kit (Roche). The resulting
dig-UTP riboprobes were purified by two steps of ethanol precipitation (including linearized polyacrylamide in the first one), resuspended in TE (10mM Tris HCl pH 7.4, 1mM EDTA pH 8.0) and stored at -20°C. Before use, the riboprobes were heated for 5min at 85°C and equilibrated at 70°C in pre-hybridization buffer. Probes in hybridization buffer were stored at -20°C and re-used several times.

9.4. Molecular Biology

9.4.1. Immunocytochemistry
Cultured cells were carefully washed 2 times with prewarmed PBS (Gibco, 37°C) and fixed 30min in 4% PFA in PBS at RT with gentle rocking. After 3 times washes in PBS for 5min each at RT, fixed cells were permeabilized with 0.3% Triton X-100 in PBT and blocked in 10% goat serum, 0.3% Triton X-100, 0.3% BSA (Sigma) in PBT for 1h at RT. Primary antibody incubation was performed overnight at 4°C in 1% goat serum, 0.3% Triton X-100, 0.3% BSA in PBS containing the relevant primary antibody at the appropriate dilution (see Antibodies Tables in chapter 9.7.1). The following day, after three 5min washes in PBT, the cells were incubated with secondary antibodies (see chapter 9.7.1) diluted in 1% goat serum in PBS for 1hr at RT in the dark. Cells were washed 3 times in PBT (5min) and incubated with 1µg/ml Hoechst-33258 (Calbiochem) in PBS for 5min to stain nuclei. After three 5min washes in PBT, the cells were stored in PBS at 4°C in a moist chamber. High-resolution fluorescent images were acquired using the Leica SP5 confocal microscope.

9.4.2. Protein detection by Western Blot (WB)
Proteins were extracted either from cultured cells or from embryonic tissues dissected in ice-cold PBS. Cultured cells were carefully washed with prewarmed PBS (Gibco, 37°C), detached and pelleted by centrifugation (3min, 1000 rpm). The cells were then resuspended in PBS and pelleted again by centrifugation (5min, 5000 rpm). After discarding the supernatant, cell pellets were snap-frozen in liquid N₂ and stored at -80°C before further processing. For embryonic tissues dissected and collected in cold PBS in Eppendorf tubes, snap-freezing in liquid N₂ and -80°C storage was performed after removing as much cold PBS as possible from the
dissected tissues. Cellular lysis of frozen cells or tissues was performed by gently resuspending the frozen sample in an appropriate volume of ice-cold complete RIPA buffer (50mM Tris-HCl pH 7.5, 1% NP-40, 0.25% Na-deoxycholate, 150mM NaCl, 1mM EDTA, 0.4mM Pefabloc, 1x Complete Mini protease inhibitor cocktail (Roche), 1mM activated Na$_3$VO$_4$, 1mM NaF) followed by a 30min incubation on ice. Lysed cells were then centrifuged 15min at 13'000 rpm (4°C) to remove cellular debris, and the supernatant was transferred to a new ice cold Eppendorf tube. The protein concentration was determined using the BiCinchoninic acid Assay (BC Assay, Pierce) and specific amounts of BSA as standard controls. Protein extracts were then aliquoted and stored at -80°C.

For SDS-PAGE, protein extracts were transferred on ice and mixed with 5x protein gel sample buffer (GSB: 0.5M Tris-HCl pH 6.8, 2% SDS, 20% glycerol, 0.0025% bromophenol blue, with freshly added 5% β-mercaptoethanol) to a final concentration of 20% GSB. Protein extracts were denatured for 5min at 98°C and cooled down on ice for 3min before being loaded on a 10% SDS polyacrylamide gel. Proteins were separated by electrophoresis in running buffer (25mM Tris base, 192mM Glycine, 0.1% (w/v) SDS) for 30min at 80V and 2-3hrs at 100-120V. Following SDS-PAGE, the SDS acrylamide gel was equilibrated in transfer buffer (25mM Tris base, 192mM Glycine, 20% (v/v) MetoH) for 15min. The proteins were then transferred from the gel to a PVDV (polyvinylidene fluoride) Immobilon-P transfer membrane (Millipore). Prior to transfer, the membrane was activated for 1min in MetOH and equilibrated in transfer buffer for 15min. Wet transfer was then conducted at 4°C at 100V for 2hrs. Following transfer, the membrane was rinsed in TBST (0.01M Tris-HCl pH 8, 0.05% Tween-20, 0.15M NaCl) and then blocked in 5% (w/v) milk powder in TBST for 1h at RT. The membrane was then incubated with primary antibody diluted in 5% milk powder in TBST (see chapter 9.7.1 for dilutions) overnight at 4°C. The following day, the membrane was rinsed in TBST and washed 3 times for 15min in TBST. Then it was incubated with the appropriate HRP conjugated secondary antibody (see chapter 9.7.1) diluted in 5% milk powder in TBST for 1hr at RT. After three 15min washes in TBST, ElectroChemiLuminescence (ECL) detection of the antibody-labeled proteins was performed for 5min at RT using a 1:1 mixture of the SuperSignal™ West Pico Chemiluminescent Substrate kit (Thermo Scientific) and the Visualizer™ Western
Blot Detection Kit (Millipore). The membrane was then exposed to Kodak BioMax MR film or Amersham Hyperfilm™ ECL (GE Healthcare) and developed in the Fujifilm FPM-100A Desktop Processor. To detect another protein, the membrane was stripped in Restore™ Western Blot Stripping Buffer (Thermo Scientific) for 5min at RT and extensively washed in TBST. After 30min of blocking, the membrane was incubated with primary antibody for 1hr at RT (for anti-actin antibody) or overnight at 4°C. Subsequent steps were performed as described above.

9.4.3. Immunoprecipitation (IP)

Protein extracts from cultured cells or from dissected embryonic tissues were prepared as described above. For each IP reaction, 150µg of total protein extract were used together with 15µl (45µg) of Dynabeads® Protein G (Life Technologies) and 1µg of anti-FLAG M2 antibody (Sigma) in a total volume of 150µl. First, for each IP reaction 150µg of total protein extract (in complete RIPA buffer with 3x Complete Mini protease inhibitor cocktail, see above) were transferred to 0.2ml PCR microtubes (Thermo Scientific) and topped up to 150µl with modified RIPA buffer. 1µg of anti-FLAG M2 antibody was added and the mixture was incubated for 3hrs at 4°C on a rotating wheel. During this incubation time, appropriate amounts of Dynabeads were transferred to a 1.5ml Eppendorf tube containing 1ml of cold RIPA buffer (without proteases and phosphatases inhibitors: 50mM Tris-HCl pH 7.5, 1% NP-40, 0.25% Na-deoxycholate, 150mM NaCl, 1mM EDTA). Dynabeads were washed 3 times with 1ml cold RIPA buffer using a magnetic rack, then resuspended in an appropriate volume of cold RIPA buffer and equally divided into new Eppendorf tubes according to the number of samples and kept on ice. After the 3hrs of incubation, the mixture was used to resuspend the Dynabeads, transferred back in the microtubes and incubated overnight at 4°C on a rotating wheel. The following day, the Dynabeads solution was transferred back to a 1.5ml Eppendorf tube and put on a magnetic rack. The supernatant (=unbound protein fraction) was transferred to a new Eppendorf tube, and the beads were washed 5 times with 0.8ml of ice-cold complete RIPA buffer. Dynabeads were resuspended in 100µl of ice-cold complete RIPA buffer, transferred to a new tube to avoid co-elution of proteins bound to the tube, and finally resuspended in 20µl cold complete RIPA buffer. To elute the
immunoprecipitated proteins from the Dynabeads, 8µl of GSB containing 5% β-mercaptoethanol (see above) were added to each sample. Samples were heated at 98°C for 5min and incubated 3min on ice. The magnetic rack was then used to separate the Dynabeads from the denatured immunoprecipitated protein complexes. The eluted proteins were transferred into a new tube and stored at -20°C if not directly loaded on a SDS-PAGE gel. In addition to the IPP samples, input control (20µg of total protein extract) and unbound protein fraction control (20µl of the 1µl/µg total protein extract recovered following the immunoprecipitation) were also analyzed on SDS-PAGE gels. Immunoblotting using primary biotinylated anti-FLAG M2 antibody and Streptavidin-HRP secondary antibody was performed as described above for regular Western-Blotting.

9.4.4. Chromatin Immunoprecipitation (ChIP)

This ChIP protocol for embryonic tissues is adapted from (Visel et al., 2009a) for tissue collection and crosslinked chromatin preparation, and on (Kim et al., 2007) and (Vokes et al., 2008) for subsequent steps, and has been modified and optimized by Marco Osterwalder and Javier Lopez-Rios. Embryos were collected in ice-cold PBS, and dissected tissues (forelimbs/hindlimbs or hearts) from somites-matched embryos of the same genotype were pooled and collected in ice-cold PBS into 2ml Eppendorf tubes. Dissected tissues were washed 2 times with ice-cold DPBS containing Ca²⁺ and Mg²⁺ (Gibco) and transferred to a 2ml glass douncer (Tissue Grind Tube Size 2ml, Kimble-Chase) on ice. Tissue disaggregation was achieved by applying 25 strokes with pestle A (Tissue Grind Pestle LC 2ml, Kimble-Chase) followed by 25 strokes with pestle B (Tissue Grind Pestle SC 2ml, Kimble-Chase). The solution containing nuclei was transferred back to a 2ml Eppendorf tube, and the douncer was rinsed with 0.3ml cold DPBS w/Ca²⁺Mg²⁺ that was also added to the sample. After centrifugation at 3000 rpm for 3min (4°C), the supernatant was discarded and nuclei were resuspended in 1.5ml RT DPBS w/Ca²⁺Mg²⁺ containing 150µl 11x crosslinking buffer (0.1M NaCl, 1mM EDTA, 0.5mM EGTA, 50mM Hepes pH 8.0; add 11% formaldehyde just before use). The mixture was incubated for 5min at RT with gentle horizontal shaking. To stop crosslinking, 75µl of 2.5M Glycine were added, followed by another incubation of 5min at RT with horizontal shaking. The samples were then
centrifuged at 3000 rpm for 3 min (4°C), supernatants removed, and pellets resuspended in 1.5 ml ice-cold DPBS w/Ca²⁺Mg²⁺. After re-centrifugation at 3000 rpm for 3 min (4°C), the supernatants were discarded and the nuclear pellets snap-frozen in liquid N₂ and stored at -80°C. In order to use sufficient material for the ChIP, crosslinked nuclear pellets from different embryo collections can be pooled in one single sample. Generally, 3 samples containing crosslinked nuclei from similar amounts of dissected embryos were processed for ChIP in parallel in one experiment.

Before performing the ChIP, antibodies coupled to magnetic beads were prepared. For one ChIP sample, 20 µl of Dynabeads® Protein G were used and transferred to a 2 ml Eppendorf tube containing 1 ml freshly prepared ice-cold BSA (5 mg/ml in DPBS w/Ca²⁺Mg²⁺). Beads were washed 6 times with 1 ml cold BSA/PBS using a magnetic rack, resuspended in a volume corresponding to 2.5 x the original volume of beads (50 µl per sample) and transferred to a 2 ml screw cap microtube (Starstedt). 2 µg of anti-FLAG M2 antibody were used for each ChIP sample and incubated overnight on a rotating wheel at 4°C.

The following day, the Dynabeads-antibodies complexes were transferred back to a 2 ml Eppendorf tube and washed 6 times with 1 ml cold BSA/PBS before being resuspended in the original volume of suspended beads (20 µl per sample) and kept on ice. In the meantime, snap-frozen crosslinked nuclear pellets were thawed on ice and resuspended in 6 ml cold Lysis buffer (50 mM Hepes pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100, 1x Complete Mini protease inhibitor cocktail (Roche)). If necessary, several crosslinked nuclear pellets were pooled and resuspended in 6 ml of ice-cold Lysis buffer (see comment above). Lysates were incubated for 10 min at 4°C with gentle rocking, followed by 10 min centrifugation at 2500 rpm (4°C). Supernatants were discarded and the pellets resuspended in 2 ml of RT Protein Extraction Buffer (0.2 M NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris-HCl pH 8.0, 1x Complete Mini protease inhibitor cocktail) to which 3 ml of Protein Extraction Buffer were added. The samples were incubated for 10 min at RT on a rocking platform and the nuclei were pelleted by centrifuging 10 min at 2500 rpm at 4°C. After discarding the supernatant, the nuclei were resuspended in 1 ml ice-cold Chromatin Extraction Buffer (1 mM EDTA pH 8.0,
0.5mM EGTA pH 8.0, 10mM Tris-HCl pH 8.0, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine, 3x Complete Mini protease inhibitor cocktail). Samples were then transferred to Covaris TC 12x12 tubes and sonicated to shear the crosslinked chromatin using the S220 Covaris Ultra-sonicator. Each sample was sonicated for 15 min using the 5% duty cycle, 140 watts Peak Incident Power and 200 Cycles per Burst. The sheared chromatin was then transferred back to 1.5ml Eppendorf tubes, and insoluble debris were pelleted by centrifuging for 10 min at 13’000 rpm (4°C). The supernatants were moved to new 2ml screw cap microtubes, and the volume was topped up to 1.060ml with Chromatin Extraction Buffer. At this stage, 2x 30µl aliquots were taken and separated to new Eppendorf tubes: one for input control and the other to assess the quality of the sonicated chromatin on an agarose gel. Both aliquots were kept overnight at 4°C to be treated identical to the ChIP sample. Then, 300µl of ChIP cocktail Mix (130µl Triton X-100 10%, 3µl Na-Deoxycholate 10%, 26µl Complete protease inhibitor solution (from a 50x tablet dissolved in 1ml mQ H2O), 131µl TE buffer (100mM Tris HCl pH 7.4, 10mM EDTA pH 8.0) and 10µl mQ H2O were added to each 1ml ChIP sample to adjust to RIPA buffer conditions (see below). For each ChIP sample, 20µl of the freshly washed Dynabead-antibody complexes were added and incubated overnight on a rotating wheel (4°C).

The next day, the dynabeads were transferred to 1.5ml Eppendorf tubes containing 1ml of freshly prepared cold RIPA buffer (50mM Hepes pH 8.0, 1mM EDTA pH 8.0, 1% NP-40, 0.7% Na-Deoxycholate, 0.5M LiCl, 1x Complete Mini protease inhibitor cocktail), and subsequently washed 6 times with 1ml cold RIPA buffer on a magnetic rack. The beads were rinsed once with 1ml TE-plus (100mM Tris HCl pH 8.0, 10mM EDTA pH 8.0, 50mM NaCl, 1x Complete Mini protease inhibitor cocktail) and centrifuged for 3min at 1000 rpm (4°C). All residual liquid was removed from the beads using the magnetic rack, and the beads were resuspended in 100µl freshly prepared Elution Buffer (1mM EDTA, 1% SDS, 10mM Tris-HCl pH 8.0). Elution of the crosslinked protein-DNA complexes from the Dynabead-antibody complexes was performed at 65°C for 15min on a heat block shaking at 1300 rpm. Beads were then centrifuged for 1min at 13’000 rpm and retained in the tube using a magnetic rack, while 100µl of supernatant containing the protein-DNA complexes were transferred to a PCR microtube. In parallel, the 30µl input aliquot and the 30µl chromatin aliquot
(kept on ice) were also transferred to PCR microtubes containing 120µl of Elution Buffer. Reverse crosslinking was performed for all three samples by incubating them overnight at 65°C (≥12hrs). The next day, samples were transferred back to 1.5ml Eppendorf tubes and treated with RNase A (from bovine pancreas, Sigma) in TE (100mM Tris HCl pH 8.0, 10mM EDTA pH 8.0) at a final concentration of 0.2µg/µl. After 1hr incubation at 37°C, Proteinase K was added to a final concentration of 0.2µg/µl and proteins were digested for 2hrs at 55°C. Finally, DNA was purified from all three samples using the QIAquick Gel Extraction Kit (Qiagen). Elution was achieved in two steps using 30µl Buffer EB each, and the eluted ChIP and input samples were aliquoted and stored at -20°C before use for qPCR. To assess the quality of the sonication, half of the purified chromatin control samples were loaded on an agarose gel to estimate the range of DNA fragments sizes (between 150bp to ~1.5kb with an maximal intensity at ~300bp, data not shown).

**9.4.5. Extraction of RNA and cDNA synthesis**

RNA was extracted either from cultured ES cells or from embryonic tissues dissected in cold PBS. For cultured ES cells, trypsin was used to detach the cells. These cells were then preplated as described in chapter 9.2.2 to separate them from EMFIs. ES cells were recovered in a 15ml Falcon tube, centrifuged for 5min at 1200 rpm and resuspended in 1.5ml ES cell medium. After re-centrifugation (3min at 5000 rpm), the medium was discarded and the cell pellet covered with 100µl RNAlater® (Ambion) and left overnight at 4°C before storage at -20°C.

For embryonic samples, the dissected tissues were transferred in ice-cold PBS in 1.5ml Eppendorf tubes. After PBS removal, samples were covered with 100µl of RNAlater® and left overnight at 4°C before storage at -20°C. Total RNA was extracted using the RNeasy Micro Kit (Qiagen). The RNA concentration was assessed using a Nanodrop 2000 C (Thermo Scientific), and RNA samples were stored at -20°C.

To synthetize cDNA, 10pg-5µg RNA were mixed with the following reagents: 1µl oligo (dT)\textsubscript{12-18} (500ng); 1µl dNTP mix (10mM each of dATP, dGTP, dCTP and dTTP at neutral pH); sterile water to a total volume of 13µl. The reaction mixture was heated
at 65°C for 5min and incubated on ice for at least 1min. After brief centrifugation, the following reagents were added: 4µl 5x First Strand Buffer, 1µl 0.1M DTT, 1µl RNaseOUT™ 40U/µl, 1µl Superscript™ III RT 200U/µl (all from Invitrogen). The mixture was incubated 1hr at 50°C, and cDNA synthesis was terminated by heating for 15min at 70°C. cDNAs were stored at -20°C before amplification by PCR (to generate templates for in situ hybridization probes or for cloning purposes) or by qPCR (to quantify target genes expression).

9.4.6. Real-time quantitative PCR (RT-qPCR)

Real-time quantitative PCR was done using the Bio-Rad CFX96 Real-Time PCR System in combination with the iQ SYBR Green Supermix (Bio-Rad). qPCR was used to determine transgene copy numbers from genomic DNA (Tesson et al., 2002), to quantify gene expression from embryonic tissues or cultured cells cDNA, or to compute levels of enrichment of genomic regions following ChIP. Each qPCR reaction was done in a 20µl reaction containing 0.3µM of each primer diluted in EB buffer (10mM Tris-HCl pH 8.5), 50% of SYBR green, and DNA template diluted in MiliQ H₂O. All primers sequences are shown in chapter 9.7.2.

- For the GRS1 transgene copy number determination, 20ng of genomic DNA was used as template. Triplicates were analyzed, using primers amplifying the Dopamine Beta-Hydroxylase (DBH) genomic region as normalizers. Transgene relative quantification cycle (Cq) values were normalized to DBH Cq values. Results are reported as mean ± standard deviation (SD) of normalized fold expression levels (2^∆∆Cq) of transgenic individuals (n= 9 or more), relative to non-transgenic control littermates (n=3) encoding the two endogenous copies of the GRS1 region.

- For gene expression quantification, 0.05µl of cDNA were used as template and analyzed in triplicate. Primers detecting the ribosomal protein L19 (RPL19) transcripts were used as internal normalizer. Relative Cq values of target transcripts were normalized to the Cq values of RPL19, and normalized fold expression levels (2^∆∆Cq) are shown as mean ± SD relative to the mean value (set at 100%) of a control sample.
For computation of enrichment levels of a region of interest (ROI) following ChIP, 1µl of immunoprecipitated genomic DNA or input DNA were used as template and analyzed in duplicate. Primers recognizing an unlinked amplicon in the β-actin locus (Galli et al., 2010) were used as normalizing control region (NCR). Fold-change enrichments between input samples and ChIP samples were calculated as follows:

\[
(Cq_{ROI \text{ ChIP}} - Cq_{ROI \text{ input}}) = \Delta Cq_{ROI}
\]
\[
(Cq_{NCR \text{ ChIP}} - Cq_{NCR \text{ input}}) = \Delta Cq_{NCR}
\]

Fold enrichment = \(2^{-(\Delta Cq_{ROI} - \Delta Cq_{NCR})}\)

All results are displayed as mean ± SD and are based on analyzing 3 independent samples (90-100 hearts each for HAND2\(^3\text{xFLAG}\) ChIP). To avoid artifact bias of fold enrichment, a Cq of 32 was defined as minimum background value threshold.

9.5. Histology

9.5.1. Paraffin embedding procedure

For paraffin embedding, embryos were dissected in ice-cold PBS and fixed overnight in 4% PFA in PBS (4°C). After three 5min washes in PBS, the embryos were dehydrated into EtOH using a graded series (25% EtOH/PBS, 50% EtOH/PBS, 75% EtOH/H\(_2\)O, 100% EtOH). Each EtOH incubation was done for 10min to 1hr depending on the size of the embryos, in ice-cold EtOH solution (4°C). Embryos were then transferred to RT and left to warm for 30min. This was followed by two incubations in EtOH for 10min each at RT. Following dehydration, embryos were cleared in Xylene for twice 25min at RT and transferred to a 50:50 (v/v) Xylene/paraffin mixture for 30min at 60°C. Then they were incubated with freshly melted paraffin 5-10 times for at least 30min each at 60°C. Then the samples were embedded in fresh paraffin wax (60°C) using a stereomicroscope for proper positioning. Paraffin blocks were then transferred to a 4°C cooling plate overnight and then stored at RT. The Microm HM 355 microtome was used to cut sections of 7µm (for immunohistochemistry and other histological stainings) or 14µm (for section in-situ hybridization) that were mounted on Superfrost Plus slides (Thermo Scientific) and stored at RT until use.
9.5.2. Preparation of frozen sections

For some section in-situ hybridization and some immunohistochemistry stainings, samples were embedded in Optimum Cutting Temperature (O.C.T) compound (Tissue-Tek) for cryoprotection as follows. Embryos were dissected in ice-cold PBS and fixed for 2hrs at 4°C in 4% PFA in PBS and washed 3 times in PBS (5min). Embryos were then transferred to 10% sucrose/PBS (w/v), 20% sucrose/PBS and 30% sucrose/PBS (1hr each) with gentle shaking at 4°C. Then the embryos were transferred into embedding molds, the 30% sucrose solution carefully removed with a plastic pipette and replaced by 50:50 (v/v) O.C.T/30% sucrose. Embryos were positioned using a stereomicroscope. Freezing was achieved by dipping and holding the bottom of the embedding mold into 2-Methylbutane cooled on dry ice in a Styrofoam box until the embedding medium was completely frozen. Samples were stored at -80°C and sectioned (10µm) using a Microm HM 500 OM Cryostat (-20°C). Frozen sections were mounted on Superfrost Plus slides and stored at -80°C before use.

9.5.3. Immunohistochemistry

IHC on paraffin sections

Sections of embryonic tissues were deparaffinized with twice 10min in Xylene followed by twice 10min in 100% EtOH. Sections were rehydrated in a reverse EtOH/H2O series (70% EtOH/H2O, 50% EtOH/H2O, 30% EtOH/H2O, 5’ each), rinsed in H2O and washed twice 5min in PBS. Antigen retrieval was achieved by autoclaving the sections in 10mM Sodium Citrate (pH 6.0) with 0.05% Tween-20 for 5min at 120°C. The sections were then cooled to RT (30min), washed twice in PBS (5min) and 4 times in PBT (PBS with 0.1% Tween-20, 5min). Sections were blocked in 10% goat serum in PBT for 1hr at RT and incubated overnight in a moist chamber in 1% goat serum in PBT containing primary antibody at the proper dilution (4°C) (see chapter 9.7.1). The following day, sections were washed three times 10min in PBT and incubated in 1% goat serum/PBT containing the appropriate secondary antibody for 1hr at RT (in the dark). The sections were then washed three times for 10min in PBT and twice in PBS (5min). Incubation in 1µg/ml Hoechst-33258/PBS (Calbiochem) for 5min counterstained nuclei. Sections were washed 5 times in PBS.
(5min) and a drop of mounting medium Mowiol 4-88 (Calbiochem) was added. They were covered with a coverslip and the slides were dried overnight at RT in the dark. For immunohistochemistry and for TUNEL assays (see below), the Leica SP5 confocal microscope was used to acquire high-resolution fluorescent images. Slides were then stored at 4°C. For double staining with two mouse primary antibodies, the anti-FLAG M2 antibody was directly labeled using the APEX™ Alexa Fluor® 488 Antibody Labeling Kit (Life Technologies). Immunostaining was performed as follows: 1) overnight incubation with mouse primary antibody (anti-NFATC1 or anti-TWIST1) 2) 1hr incubation with goat anti-mouse Alexa 594 3) overnight incubation with Alexa 488-labelled anti-FLAG M2 4) 1hr incubation with rabbit anti-Alexa488 and 5) 1hr incubation with goat anti-rabbit Alexa 488.

IHC on frozen sections
Sections were washed with PBS 3 times (5min), with 0.2% Triton X-100/PBS for 20min, and again three times with PBS (5min). Sections were then blocked in 1% BSA in PBT for 1hr at RT, followed by incubation with primary antibodies (see chapter 9.7.1) in 1% BSA/PBS overnight (4°C). The second day, the sections were washed 3 times for 10min in PBS and once in PBT (5min). Incubation with the appropriate secondary antibodies diluted in 1% BSA/PBS was done for 1hr at RT in the dark. Sections were washed 3 times for 10min in PBS and once in PBT (5min), counterstained with 1µg/ml Hoechst-33258/PBS (5min), followed by final washes in PBS (≥ 5x 5min). Mowiol 4-88 was used for mounting before addition of a coverslip. Before analysis, slides were dried overnight at RT in the dark and stored at 4°C.

9.5.4. TUNEL assay
Cell death in embryonic tissues was detected on paraffin section using the In situ Cell Death Detection Kit (Fluorescein, Roche). Following deparaffinization, rehydration and two PBS washes (5min), sections were washed with Tris-HCl 10mM pH 7.5 (5min) and permeabilized in 20µg/ml Proteinase K in Tris-HCl 10mM pH 7.5 (20min at RT). After two PBS washes (5min), sections were incubated with the TUNEL reaction mixture for 1hr at 37°C in a moist chamber in the dark. Then, the sections were washed 3 times in PBS (5min), and the nuclei counterstained with 1µg/ml
Hoechst-33258/PBS (5min). Finally, after two more PBS washes (5min), sections were mounted in Mowiol 4-88 and stored as described.

9.5.5. Hematoxylin/Eosin staining
Sections of paraffin embedded samples were deparaffinized and rehydrated as described above, and then washed extensively in tap water for 5min. Histological staining was done by incubating the sections 3min in filtered Hematoxylin solution (Sigma) and washing in tap water for 3min. Then the sections were incubated for 1-3min in filtered 0.15% Eosin/EtOH. Sections were then incubated in 95% EtOH (5min), followed by two incubations in 100% EtOH (5min) and three incubations in Xylene (5min). For mounting, the sections were covered with a drop of Eukitt (Fluka) and a coverslip was added. Slides were stored as described. Pictures were acquired using a Zeiss Axioskop 2 Plus microscope.

9.5.6. Alcian Blue staining
To stain proteoglycans, sections of paraffin embedded samples were deparaffinized, rehydrated and washed in tap water as described above. Sections were washed for 3min in 3% acetic acid in H₂O (v/v), pH 2.5. Staining was performed for minimum 1hr in 1% Alcian Blue (Sigma) in 3% acetic acid solution (w/v), pH 2.5. After extensive washes in tap water, the sections were incubated in Fastred (Sigma) for 5min at RT to counterstain nuclei. Slides were mounted with Mowiol 4-88 and stored as describe. Protocol adapted from (Rivera-Feliciano and Tabin, 2006).

9.5.7. In situ hybridization on sections
For all steps prior to probe hybridization, 0.1% DEPC (v/v) solutions were used. Paraffin sections of embryos were deparaffinized twice in Xylene (5min) and twice in 100% EtOH (5min). Sections were rehydrated as follows: 75% EtOH/H₂O-DEPC, 50% EtOH/H₂O-DEPC, 25% EtOH/H₂O-DEPC, and twice in PBS-DEPC (5min each). Following rehydration, sections were fixed in 4% PFA/PBS-DEPC (10min), rinsed in PBT-DEPC and washed twice in PBT-DEPC (5min). Sections were then incubated in 1µg/ml Proteinase K in PBS-DEPC (10min) and washed twice in PBT-DEPC (5min). For in situ hybridization on frozen sections, the protocol was started after the
Proteinase K incubation. Post-fixation was done in 4% PFA/PBS-DEPC (5min), followed by a rinse in PBT-DEPC and two washes in PBT-DEPC (5min). Sections were blocked with 0.25% (v/v) Acetic Anhydride in 0.1M Triethanolamine (pH 8) for 15min, rinsed in PBT-DEPC and washed three times in PBT-DEPC (5min). Sections were then incubated in 65°C prewarmed pre-hybridization buffer (10mM Tris-HCl pH 7.5, 600mM NaCl, 1mM EDTA, 0.25% SDS, 10% Dextran Sulfate, 200µg/ml yeast tRNA (Gibco), 50% Formamide, 1x Denhartdt’s (0.02% (w/v) Ficoll 400, 0.02% (w/v) Polyvinilpyrrolidone, 0.02% BSA in H₂O-DEPC)), covered with a coverslip and incubated at 65°C for 1hr. For hybridization, sections were incubated in prewarmed hybridization buffer containing 1% (v/v) DIG-labelled RNA probe, covered with a coverslip and incubated at 65°C overnight in a moist chamber. The second day, coverslips were removed from the slides by rinsing them in 5x SSC (20x SSC: 3M NaCl, 0.3M Tri-sodium citrate dehydrate, pH 4.5) prewarmed to 65°C followed by two washes in 5x SSC (5min each) at 65°C. Sections were then incubated in 1xSSC/50% formamide at 65°C (30min) and washed in TNE (50mM Tris-HCl pH 7.5, 2.5M NaCl, 5mM EDTA) at 37°C for 10min. Next, the sections were incubated in 20µg/ml RNase A in TNE at 37°C (30min) and washed in TNE at 37°C (10min). Further washes included 2xSSC at 65°C (20min), two 0.2xSSC washes at 65°C (20min) and two MABT (200mM Maleic Acid, 300mM NaCl, pH 7.5, 0.1% Tween-20 (v/v)) at RT (5min). Sections were blocked in 20% lamb serum in MABT for 1hr at RT and incubated overnight in 2% lamb serum in MABT containing 1:2000 anti-DIG-AP (Roche) in a moist chamber (4°C). The third day, the sections were rinsed and then washed three times in MABT at RT (5min each). Then, the sections were equilibrated in NTMT (100mM NaCl, 100mM Tris-HCl pH 9.5, 50mM MgCl₂, 0.1% Tween-20) at RT for 10min, and the signal developed using BM purple AP substrate (Roche). Development of the in-situ hybridization was performed in the dark at RT for 1hr to 3 days depending on the probe, with regular monitoring using a stereomicroscope. The reaction was stopped by rinsing the sections with NTMT and washing them at least three times in PBS (5min each). The sections were post-fixed in 0.1% glutaraldehyde, 4% PFA/PBS for 20min, washed twice in PBS (5min) and rinsed in H₂O. Sections were mounted in Mowiol 4-88 and stored as described.
9.6. Genomics online resources

All genomic sequences for cloning purposes or sequence alignments were retrieved from ENSEMBL (Flicek et al., 2014) or UCSC browser (Karolchik et al., 2014). Alignments were performed using the ECR browser (Ovcharenko et al., 2004) or the ClustalW2 program (Larkin et al., 2007). Other useful online resources include the VISTA enhancer browser (Visel et al., 2007) and the Mouse Genome Informatics website (http://www.informatics.jax.org).

9.7. Tables

9.7.1. Antibodies table

<table>
<thead>
<tr>
<th>Antigen</th>
<th>Host</th>
<th>Distributor</th>
<th>Use</th>
<th>Dilutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLAG (M2)</td>
<td>monoclonal mouse Sigma, F1804</td>
<td>WB, IP, ICC, IHC, ChiP</td>
<td>1:2000, 1µg Ab/150µg prot, 1:4000, 1:1000, 2µg/sample</td>
<td></td>
</tr>
<tr>
<td>FLAG BioM2-Biotin</td>
<td>monoclonal mouse Sigma, F9291</td>
<td>WB</td>
<td>1:2000</td>
<td></td>
</tr>
<tr>
<td>ACTIN</td>
<td>polyclonal rabbit Sigma, A2066</td>
<td>WB</td>
<td>1:10'000</td>
<td></td>
</tr>
<tr>
<td>HA</td>
<td>polyclonal rabbit Sigma, H6908</td>
<td>WB, IP, ICC</td>
<td>1:2000, 1:2000, 1µg Ab/150µg prot</td>
<td></td>
</tr>
<tr>
<td>V5</td>
<td>monoclonal mouse Invitrogen, R960-25</td>
<td>WB, IP, ICC</td>
<td>1:10'000, 1:2000, 1µg Ab/150µg prot</td>
<td></td>
</tr>
<tr>
<td>SMAD4</td>
<td>monoclonal rabbit Abcam, ab-40759</td>
<td>WB</td>
<td>1:1000</td>
<td></td>
</tr>
<tr>
<td>GLI3</td>
<td>polyclonal rabbit Genentech, 2676</td>
<td>WB, ChiP</td>
<td>1:1000, 2µg/sample</td>
<td></td>
</tr>
<tr>
<td>TBX2</td>
<td>polyclonal mouse Milipore, 07-318</td>
<td>IHC</td>
<td>1:2000</td>
<td></td>
</tr>
<tr>
<td>TBX3</td>
<td>polyclonal goat Santa Cruz, sc-31656</td>
<td>IHC</td>
<td>1:300</td>
<td></td>
</tr>
<tr>
<td>ISL1 (Mab15)</td>
<td>monoclonal mouse DSHB, 40.2D6</td>
<td>IHC</td>
<td>1:100</td>
<td></td>
</tr>
<tr>
<td>SOX9</td>
<td>polyclonal rabbit Milipore, ab-5535</td>
<td>ICC, IHC</td>
<td>1:5000, 1:10'000</td>
<td></td>
</tr>
<tr>
<td>TWIST1</td>
<td>monoclonal mouse Santa Cruz, sc-81417</td>
<td>IHC</td>
<td>1:500</td>
<td></td>
</tr>
<tr>
<td>VE CADHERIN</td>
<td>polyclonal rabbit Abcam, ab-33168</td>
<td>IHC</td>
<td>1:1000</td>
<td></td>
</tr>
<tr>
<td>VIMENTIN</td>
<td>monoclonal mouse Sigma, V2258</td>
<td>IHC</td>
<td>1:500</td>
<td></td>
</tr>
<tr>
<td>NFATC1</td>
<td>monoclonal mouse Santa Cruz, sc-7294</td>
<td>IHC</td>
<td>1:500</td>
<td></td>
</tr>
<tr>
<td>COLLAGEN 2</td>
<td>monoclonal mouse ThermoSci, MS-235-P1</td>
<td>ICC</td>
<td>1:100</td>
<td></td>
</tr>
</tbody>
</table>
### Secondary Antibodies

<table>
<thead>
<tr>
<th>Antigen</th>
<th>Distributor</th>
<th>Use</th>
<th>Dilutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>goat α mouse IgG, HRP conjugate</td>
<td>Millipore, 12-349</td>
<td>WB</td>
<td>1:20'000</td>
</tr>
<tr>
<td>goat α rabbit IgG, HRP conjugate</td>
<td>Millipore, 12-348</td>
<td>WB</td>
<td>1:20'000</td>
</tr>
<tr>
<td>Streptavidin-HRP</td>
<td>Jackson, 016-030-084</td>
<td>WB</td>
<td>1:20'000</td>
</tr>
<tr>
<td>goat α mouse IgG (H+L) Alexa Fluor®488 F(ab')2</td>
<td>Invitrogen, A-11017</td>
<td>ICC, IHC</td>
<td>1:1’000</td>
</tr>
<tr>
<td>goat α mouse IgG (H+L) Alexa Fluor®594 F(ab')2</td>
<td>Invitrogen, A-11020</td>
<td>IHC</td>
<td>1:1’000</td>
</tr>
<tr>
<td>goat α rabbit IgG (H+L) Alexa Fluor®488 F(ab')2</td>
<td>Invitrogen, A-11070</td>
<td>IHC</td>
<td>1:1’000</td>
</tr>
<tr>
<td>goat α rabbit IgG (H+L) Alexa Fluor®594</td>
<td>Invitrogen, A11037</td>
<td>ICC, IHC</td>
<td>1:1’000</td>
</tr>
<tr>
<td>rabbit α goat IgG (H+L) Alexa Fluor®488</td>
<td>Invitrogen, A-11080</td>
<td>IHC</td>
<td>1:1’000</td>
</tr>
<tr>
<td>rabbit α Alexa Fluor®488 IgG</td>
<td>Invitrogen, A-11094</td>
<td>IHC</td>
<td>1:1’000</td>
</tr>
<tr>
<td>Anti-Digoxigenin-AP, Fab fragments from sheep</td>
<td>Roche, 11 093 274 910</td>
<td>WISH, SB</td>
<td>1:5000, 1:20'000</td>
</tr>
</tbody>
</table>

### 9.7.2. Primers tables

<table>
<thead>
<tr>
<th>Locus</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Size</th>
<th>Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRS1 transgene</td>
<td>5'-GCCTCTTCGCTATTACGCCAGCT-3'</td>
<td>5'-AATTAACCCTCACTAAAGGG-3'</td>
<td>300bp</td>
<td>Tg</td>
</tr>
<tr>
<td>Shh</td>
<td>5'-ATGCTTGCTCGCTTGCTTGAA-3'</td>
<td>5'-GAAGAGATCAAGGCAAGCTCTGC-3'</td>
<td>500bp</td>
<td>Wt</td>
</tr>
<tr>
<td></td>
<td>5'-GGACACATTCTATGCAGGG-3'</td>
<td>5'-GAAGAGATCAAGGCAAGCTCTGC-3'</td>
<td>150bp</td>
<td>Δ</td>
</tr>
<tr>
<td>Gli3</td>
<td>5'-GGGTGAACAGCATCAAAATGGAG-3'</td>
<td>5'-ATAAGCCATGTGTCCTGC-3'</td>
<td>600bp</td>
<td>Xlj</td>
</tr>
<tr>
<td></td>
<td>5'-TACCCAGCGAGACTCAGTAG-3'</td>
<td>5'-AAACCCGTCGCTAGAGAAG-3'</td>
<td>500bp</td>
<td>Δ</td>
</tr>
<tr>
<td>Smad4</td>
<td>5'-GGGCAGCGTAGCATATAAGAC-3'</td>
<td>5'-AAGTACCACATTACACCTCC-3'</td>
<td>390bp + 450bp</td>
<td>Wt + Floxed</td>
</tr>
<tr>
<td></td>
<td>5'-AAGAGCCACAGTCAAGCAG-3'</td>
<td>5'-CTGCCACCCAAACGTACCTC-3'</td>
<td>500bp</td>
<td>Δ</td>
</tr>
<tr>
<td>Cre</td>
<td>5'-GCCTCGATTACCGTCTGCAACAGA-3'</td>
<td>5'-GTGGGAGATGGCCGCGGCAACCATT-3'</td>
<td>750bp</td>
<td>Tg</td>
</tr>
<tr>
<td>Grem1</td>
<td>5'-ATGAATCGCACGCACTCAACT-3'</td>
<td>5'-TCAAGTCGATGGAGATG-3'</td>
<td>750bp</td>
<td>Wt</td>
</tr>
<tr>
<td></td>
<td>5'-GGGCACATGGCTCGCTTAGG-3'</td>
<td>5'-AAGGCCTCCCTCCCTCCGGA-3'</td>
<td>550bp</td>
<td>Δ</td>
</tr>
<tr>
<td>Hand2</td>
<td>5'-TGCTACTCGCTCTCTCTCT-3'</td>
<td>5'-CCGATCTGAGCTAGC-3'</td>
<td>330bp</td>
<td>Wt + Floxed</td>
</tr>
<tr>
<td></td>
<td>5'-AGAAGAGGACCTCGCG-3'</td>
<td>5'-CTGCTGTCGTTGCCTGATT-3'</td>
<td>496bp</td>
<td>Δ</td>
</tr>
<tr>
<td></td>
<td>5'-TGGAGGGCCAGGAGAGGAGGATG-3'</td>
<td>5'-GACAGGGCCATCTGTAGTCG-3'</td>
<td>209bp + 284bp</td>
<td>Wt + 3xFlag</td>
</tr>
</tbody>
</table>
### Smad4 Homologous Recombination targeting - cloning primers

<table>
<thead>
<tr>
<th>Linker</th>
<th>Sequence</th>
<th>Restriction sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linker Up</td>
<td>5'‐AGCGCTGGGCCCTTTAATAGATCTGGATCCCTGATCAACGCCTAGCT‐3'</td>
<td>MluI, Apal, spacer, BglII, BamHI, BclI, MluI, X SacI</td>
</tr>
<tr>
<td>Linker Down</td>
<td>3'‐CATGTTGCCACCCCGGAATTACTGACCTAGACTGTTGCCA‐5'</td>
<td>X KpnI, Apal, spacer, BglII, BamHI, BclI, MluI</td>
</tr>
</tbody>
</table>

### Smad4 Homologous Recombination targeting - Southern Blot probes primers

<table>
<thead>
<tr>
<th>Probe</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' probe</td>
<td>5'TTGCTCCACAGGAGAACTAACA‐3'</td>
<td>5'‐TTGCTCCACAGGAGAACTAACA‐3'</td>
<td>580bp</td>
</tr>
<tr>
<td>3' probe</td>
<td>5'‐GCCAGGATTTGATGAGTT‐3'</td>
<td>5'‐GCCAGGATTTGATGAGTT‐3'</td>
<td>515bp</td>
</tr>
<tr>
<td>Neo probe</td>
<td>5'‐GACGTTTCACCTGAGCAGGGAAGG‐3'</td>
<td>5'‐GACGTTTCACCTGAGCAGGGAAGG‐3'</td>
<td>508bp</td>
</tr>
</tbody>
</table>

### Smad4 Homologous Recombination targeting - mice genotyping primers

<table>
<thead>
<tr>
<th>Locus</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Size</th>
<th>Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smad4 3xF‐HR</td>
<td>5'‐ACAGCCTCCACACTTGTGCT‐3'</td>
<td>5'‐TGTCCTGCTAAGAGCAAGGCA‐3'</td>
<td>396bp + 501bp</td>
<td>WT + 3xF HR</td>
</tr>
<tr>
<td>Smad4 3xF‐HR Neo</td>
<td>5'‐GAGCCAGTTGTACCATGCG‐3'</td>
<td>5'‐GCCTGAAGAACGAGATCAGC‐3'</td>
<td>432bp</td>
<td>Neo'</td>
</tr>
<tr>
<td>Smad4 3xF‐HR ∆Neo</td>
<td>5'‐AGGACATTCCCCATGGACACTG‐3'</td>
<td>5'‐AGCAGTGGCCTGGTCAGATGA‐3'</td>
<td>576bp + 610bp</td>
<td>WT + 3xF ∆Neo</td>
</tr>
</tbody>
</table>

### Candidates CRMs for transgenic reporters - cloning primers

<table>
<thead>
<tr>
<th>CRM</th>
<th>Coordinates (mm9)</th>
<th>Forward / Reverse primers</th>
<th>Size</th>
<th>Restr. site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has2 +2.4kb Tg</td>
<td>chr15:56,523,368‐56,525,546</td>
<td>5'‐GAGGAAGCTTGTGCAAGAAGAACGATCTCA‐3'</td>
<td>2218 bp</td>
<td>HindIII</td>
</tr>
<tr>
<td>Tbx2 +32kb Tg</td>
<td>chr11:85,612,756‐85,614,268</td>
<td>5'‐GAGGCCCCCATTTTCTCTGC‐3'</td>
<td>1513 bp</td>
<td></td>
</tr>
<tr>
<td>Tbx3 +58kb Tg</td>
<td>chr5:120,060,938‐120,063,445</td>
<td>5'‐GAATAGCGAGAGCAGACATC‐3'</td>
<td>2508 bp</td>
<td></td>
</tr>
<tr>
<td>Tbx3 +18kb Tg</td>
<td>chr5:120,100,229‐120,102,755</td>
<td>5'‐GGGTGTAACGTGGAGA‐3'</td>
<td>2527 bp</td>
<td></td>
</tr>
<tr>
<td>Twist1 +130kb Tg</td>
<td>chr12:34,771,890‐34,772,660</td>
<td>5'‐GAGGAAGCTTGTGCAAGAAGAACGATCTCA‐3'</td>
<td>753bp</td>
<td>HindIII</td>
</tr>
<tr>
<td>Snai1 +58/59kb Tg</td>
<td>chr2:167,419,933‐167,424,457</td>
<td>Released from BAC RP23-193B17 as HpaI/MfeI fragment</td>
<td>4625 bp</td>
<td></td>
</tr>
</tbody>
</table>

### Candidates CRMs for transgenic reporters - genotyping primers

<table>
<thead>
<tr>
<th>Locus</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has2 +2.4kb Tg</td>
<td>5'‐GAGGAAGTTAAGGATGACAG‐3'</td>
<td>5'‐AATGTGAGCGAGAACACCCG‐3'</td>
<td>916bp</td>
</tr>
</tbody>
</table>
**ChIP qPCR primers**

<table>
<thead>
<tr>
<th>ChIP Amplicon</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Size</th>
<th>Rank in Top2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Id1 BRE -1.1kb</td>
<td>5'-AGAATGCTCCAGCCAGTTT-3'</td>
<td>5'-TGACGTCACCATCATAAAA-3'</td>
<td>93bp</td>
<td>-</td>
</tr>
<tr>
<td>Id1 -3kb (ctrl)</td>
<td>5'-TTCTTCTCTGGTGCCAGTG-3'</td>
<td>5'-AACCAGCCTGATCATGC-3'</td>
<td>140bp</td>
<td>-</td>
</tr>
<tr>
<td>Actb +1.8kb ctrl region</td>
<td>5'-GATCTGAGACATGCAAGAGCTG-3'</td>
<td>5'-GGCCCTGGAGTTGATGATG-3'</td>
<td>115bp</td>
<td>-</td>
</tr>
<tr>
<td>Tbx2 -58kb (ctrl)</td>
<td>5'-TGCTGGCTCTGTCAGTGACCT-3'</td>
<td>5'-CCAGATGGCATTACAGACAG-3'</td>
<td>131bp</td>
<td>216</td>
</tr>
<tr>
<td>Tbx2 -59kb (ctrl)</td>
<td>5'-AACCTCGGTACACTGCAAC-3'</td>
<td>5'-ACAGAAAAATAAGGCTGGCC-3'</td>
<td>134bp</td>
<td>-</td>
</tr>
<tr>
<td>Tbx3 +1.8kb</td>
<td>5'-AGGATTCCAGGGAGCTG-3'</td>
<td>5'-TTGCTCAGTACGATGACATAAAA-3'</td>
<td>110bp</td>
<td>944</td>
</tr>
<tr>
<td>Tbx2 -204kb (ctrl)</td>
<td>5'-CTCCACTGATGGCTGAAAT-3'</td>
<td>5'-CCAGATGGCATTACAGACAG-3'</td>
<td>138bp</td>
<td>164</td>
</tr>
<tr>
<td>Tbx2 -199kb (ctrl)</td>
<td>5'-CTGTTGCTACCTCCGATCA-3'</td>
<td>5'-CACCCTCACCTGAGTTGATG-3'</td>
<td>112bp</td>
<td>-</td>
</tr>
<tr>
<td>Tbx2 +135kb (ctrl)</td>
<td>5'-AGGGAAGAGAGTTAGGGG-3'</td>
<td>5'-CACCAGAGAAAGGGCTCT-3'</td>
<td>90bp</td>
<td>595</td>
</tr>
<tr>
<td>Myocd +60kb</td>
<td>5'-AGGATGTCAGGGGCTGCT-3'</td>
<td>5'-AAAGCTGTGGGAGGACAGT-3'</td>
<td>80bp</td>
<td>11</td>
</tr>
<tr>
<td>Myocd -31kb (ctrl)</td>
<td>5'-AGGAGCAGAGATCAAACACTA-3'</td>
<td>5'-AAAGCTGTGGGAGGACAGT-3'</td>
<td>80bp</td>
<td>11</td>
</tr>
<tr>
<td>Twist1 +10kb</td>
<td>5'-AGGATGTCAGGGGCTGCT-3'</td>
<td>5'-AAAGCTGTGGGAGGACAGT-3'</td>
<td>80bp</td>
<td>11</td>
</tr>
<tr>
<td>Twist1 +233kb</td>
<td>5'-AGGATGTCAGGGGCTGCT-3'</td>
<td>5'-AAAGCTGTGGGAGGACAGT-3'</td>
<td>80bp</td>
<td>11</td>
</tr>
<tr>
<td>Snai1 +64kb (ctrl)</td>
<td>5'-AGGATGTCAGGGGCTGCT-3'</td>
<td>5'-AAAGCTGTGGGAGGACAGT-3'</td>
<td>80bp</td>
<td>11</td>
</tr>
<tr>
<td>Snai1 +59kb</td>
<td>5'-ACCTCCTCCTGAGTTGAG-3'</td>
<td>5'-TTGCTGAGCATGCAAGAAA-3'</td>
<td>73bp</td>
<td>1588</td>
</tr>
<tr>
<td>Snai1 +58kb (ctrl)</td>
<td>5'-ACCTCCTCCTGAGTTGAG-3'</td>
<td>5'-TTGCTGAGCATGCAAGAAA-3'</td>
<td>73bp</td>
<td>1588</td>
</tr>
<tr>
<td>Sox9 +32kb</td>
<td>5'-AGGATGTCAGGGGCTGCT-3'</td>
<td>5'-AAAGCTGTGGGAGGACAGT-3'</td>
<td>80bp</td>
<td>11</td>
</tr>
<tr>
<td>Sox9 +233kb</td>
<td>5'-AGGATGTCAGGGGCTGCT-3'</td>
<td>5'-AAAGCTGTGGGAGGACAGT-3'</td>
<td>80bp</td>
<td>11</td>
</tr>
<tr>
<td>Has2 +4.3kb (ctrl)</td>
<td>5'-AGGATGTCAGGGGCTGCT-3'</td>
<td>5'-AAAGCTGTGGGAGGACAGT-3'</td>
<td>80bp</td>
<td>11</td>
</tr>
<tr>
<td>Has2 -0.9kb</td>
<td>5'-AGGATGTCAGGGGCTGCT-3'</td>
<td>5'-AAAGCTGTGGGAGGACAGT-3'</td>
<td>80bp</td>
<td>11</td>
</tr>
<tr>
<td>Has2 +2.4kb</td>
<td>5'-AGGATGTCAGGGGCTGCT-3'</td>
<td>5'-AAAGCTGTGGGAGGACAGT-3'</td>
<td>80bp</td>
<td>11</td>
</tr>
</tbody>
</table>

**qPCR primers**

<table>
<thead>
<tr>
<th>Locus</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smad4 exon 6-7 / 7-8</td>
<td>5'-CCATCATAAAAGCAGACTCACC -3'</td>
<td>5'-GCCAGTAAATGCAGGATGG-3'</td>
<td>128bp</td>
</tr>
<tr>
<td>Smad4 exon 8-9</td>
<td>5'-TTTCCATCTCGTCTCCGTA-3'</td>
<td>5'-AACAGGCTGCTCTCACCTGAGTA-3'</td>
<td>76bp</td>
</tr>
<tr>
<td>Smad4 exon12-3'UTR</td>
<td>5'-TGCTGGGATAGGATCCAC-3'</td>
<td>5'-CGTGCTGGGATAGTCTCA-3'</td>
<td>108bp</td>
</tr>
<tr>
<td>Smad4 3xFlag3'-UTR</td>
<td>5'-AACAGGCTGCTCTCACCTGAGTA-3'</td>
<td>5'-CGTGCTGGGATAGTCTCA-3'</td>
<td>70bp</td>
</tr>
<tr>
<td>Rpl19</td>
<td>5'-ACCCTGAGCGAAGG-3'</td>
<td>5'-TACCCTGAGCGAAGG-3'</td>
<td>53bp</td>
</tr>
<tr>
<td>GRS1</td>
<td>5'-CCAGGCTGGGATAGTCTCA-3'</td>
<td>5'-CCAGGCTGGGATAGTCTCA-3'</td>
<td>64bp</td>
</tr>
<tr>
<td>Dbh</td>
<td>5'-AGGATGTCAGGGGCTGCT-3'</td>
<td>5'-AAATGGCTTGAGGCCCTC-3'</td>
<td>117bp</td>
</tr>
</tbody>
</table>
10. Acknowledgments

First and foremost, I would like to sincerely thank Prof. Rolf Zeller who gave me the fantastic opportunity to do my PhD in his laboratory. I thank him for the great mentoring and guidance over the years; his advice and ideas were always precious and useful and helped me to improve my scientific thinking and writing.

I am deeply thankful to Javier Lopez-Rios for his invaluable supervision and teaching. It was always a pleasure to work with him and an enriching source of new knowledge, science-related or not. His constant scientific enthusiasm and excitement was one of the best motivation drivers I had during these years.

I am truly indebted to Marco Osterwalder for his great help, support and patience with my numerous questions. Without the massive amount of amazing work he did in the laboratory, a vast majority of what is in this thesis regarding ChIP-Seq experiments and analysis would not have been possible.

It was also a pleasure to collaborate with Aimée Zuniga, and I would like to thank her for everything I learnt while working together in the cell culture room and on the GRS1 project.

I would also like to express my sincere gratitude to Prof. Verdon Taylor and Prof. Markus Affolter for being members of my thesis committee.

Many thanks go to the team of the Transgenic Mouse Core Facility of the University of Basel, especially Daniela Klewe-Nebenius and Liliane Todesco, for the generation of chimeric mice and some of the transgenic founder embryos. I would also like to thank Malak Shoukry and Axel Visel from Lawrence Berkeley National Laboratory, who generated some other transgenic embryos for us. I am also very grateful to Angelika Offinger and Emilia Stanislawa Terszowska for excellent care of our mice.

I am thankful to Isabelle Ginez for her indispensable technical help with sectioning, to Pascal Lorentz for teaching and helping me with the confocal microscope, to Susie Boudebaba who always supplied us with clean and autoclaved glassware and
solutions, and to Chris Müller for all her administrative help. And special thanks got to Debora Schmitz, who was the first one to get me interested in cardiac development.

Furthermore, I would like to warmly thank Emanuele Pignatti, Sumit Jaiswal, Erkan Ünal, Dario Speziale and Ashleigh Nugent for their valuable friendship and all the great moments we shared in the laboratory, but mostly for all the ones we spent outside. In addition, I am especially grateful to Julie Leclercq, who will successfully continue some of the projects I worked on and whom I had a lot of pleasure teaching. The nice atmosphere and constructive discussions we had in the laboratory would not have been possible without all the present and past members of the group, which I thank very much: Gretel Nusspaumer, Nathalie Riesen, Jorge Dorado, Virginie Tissières, Paola Valdivieso, Rosaria Costa, Sandro Nuciforo, Patric Schlenker, Simone Probst, Catherine Vaillant, Amandine Duchesne, Alexandre Gonçalves, Alexandra Schauerte, Jean-Denis Benazet, Adrian Hermann and Dimitri Robay.

Many thanks go to the members of the DBM PhD Club committee, especially the two other founder members Carlos Mayer and Fabrizio Botindari. It was as much fun to organize all these events together, than to participate to them. I am also thankful to my friends Valérie Murset, César Metzger, Vincent Croset and Joanes Grandjean, who also survived their PhD after we graduated from the University of Lausanne, and with whom it was always a pleasure to discuss and share good times.

Last but not least, my deepest and warmest gratitude goes to my parents, to my sister and to Jérémy for their incredible support and patience during these years.
11. Bibliography


12. Appendixes

12.1. Generation of Smad4\textsuperscript{3xFLAG} mice using dRMCE

12.1.1. Testing three epitope-tagged versions of SMAD4 in ES cells

The first attempt to generate a mouse with an epitope tag fused in frame in the Smad4 coding sequence was done using dRMCE. This technique is based on dual recombination of the loxP and FRT sites present in numerous conditional alleles available as ES cell lines, for example from the EUCOMM and IKMC consortium (Osterwalder et al., 2010a). The co-electroporation of the pDIRE vector coding for the iCre and FLPo recombinases together with a pDREV targeting vector is designed to insert an epitope tag, a mutation or a reporter in a specific floxed locus at much higher efficiencies than homologous recombination (Osterwalder et al., 2010a).

Several epitope-tagged versions of SMAD4 had been generated and used mainly for in-vitro expression studies in various cell lines over the past decades, without a consistent position of the epitope tag relative to the N-term or the C-term extremities of the protein. The choice of the epitope tag and its positions are important points to avoid potential interference with the physiological function of the target transcription factor. We thus took advantage of the high targeting efficiency with dRMCE to test three different epitope tags and assess their sensitivity and specificity before deciding which one using to generate a Smad4 epitope-tagged mouse strain. The three different epitope tags (FLAG, HA and V5) were inserted in frame of the C-terminal part of SMAD4 and multimerized (3x) to increase detection sensitivity (3xFLAG, 3xHA and 3xV5). In silico comparison of SMAD4 3D structure with the predicted 3D structures of the three different versions of the SMAD4\textsuperscript{3xTAG} protein using I-TASSER (Roy et al., 2010) did not show significant alteration in the conformation of the epitope tagged protein (data not shown). Overexpression of the three versions of the SMAD4\textsuperscript{3xTAG} protein in HEK293 cells showed that the three epitope tags are efficiently detected with high sensitivity using specific antibodies (data not shown).

For dRMCE mediated cassette exchange, we used an existing floxed Smad4 knockout-first tm1a allele generated by IKMC. This allele encodes a promotorless gene-trap selection cassette flanked by FRT sites, which is followed by loxP sites flanking Smad4 exon 9 (S4\textsuperscript{EC} allele, Fig.30A).
Figure 30. Dual recombinase-mediated cassette exchange in mouse ES cells to engineer a Smad4<sup>3xTAG</sup> allele

(A) dRMCE targeting strategy. From top to bottom: the Smad4 wild-type allele (S<sup>4wt</sup>). The pDREV targeting vector containing a FRT site, Smad4 exons 9-12, 3 different versions of a 3x epitope tag (3xFLAG, 3xHA or 3xV5) fused in frame to Smad4 exon 12, a STOP codon, Smad4 3'UTR, SV40 pA signal, a Puromycin selection cassette flanked by rox sites, and a loxP site. The Smad4 EUCOMM conditional allele with a promoterless cassette flanked by FRT sites and Smad4 exon 9 flanked by loxP sites (S<sup>4EC</sup> ). The Smad4 deleted EUCOMM allele following double iCre and FLPo recombination without insertion of a targeting element (S<sup>4<sup>∆</sup>E</sup>). The Smad4 targeted allele with dRMCE-mediated insertion of the targeting element from pDREV between the FRT and loxP sites (S<sup>4xT-dRMCE</sup>). The primers for PCR screening are indicated with arrows. The rest restriction sites used for Southern Blot are indicated in italics, and the Southern Blot probes are shown as green boxes. (B) PCR screening analysis of all Smad4 alleles. Primers used for each reaction are indicated on the right. For the S<sup>4xT-dRMCE</sup> allele, the screening was performed both on the 3' and on the 5' sides. Clone Φ1C3 is a
positively targeted clone, χ5C3 is a S4\textsuperscript{DE/+} clone, and χ1E4 a parental S4\textsuperscript{EC/+} clone. S4\textsuperscript{+/+} (green) and S4\textsuperscript{EC/+} (blue) ES cells were used as controls. (C) Table showing the number of PCR positive clones in relation to the total number of clones picked. In total, 116 of 168 ES cell clones were screened by PCR. 88 ES cell clones were correctly targeted on both 3' and 5' sides. (D) Southern Blot analysis for three ES cell clones of each replacement. S4\textsuperscript{+/+} (green) and S4\textsuperscript{EC/+} (blue) ES cells were used as controls. Top panel: ES cells DNA was digested with NdeI and blotted with the 5' probe shown in (A). The sizes of the expected S4\textsuperscript{wt}, S4\textsuperscript{EC} or S4\textsuperscript{3xT-dRMCE} bands are indicated on the left. Middle panel: EcoRV digestion in combination with the 3' probe. Lower panel: detection of the Puro cassette after EcoRV digestion.

Two ES cell clones of this EUCOMM conditional allele were used and targeted in parallel (A07 and G07). The targeting vector was designed to replace the promoterless cassette and exon 9 of the S4\textsuperscript{EC} allele with the coding sequence of exon 9-12 with an in frame 3x epitope TAG, a STOP codon, the endogenous Smad4 3'UTR and a polyadenylation site from SV40. These modifications are followed by a Puromycin selection cassette that is flanked by rox sites (Fig.30A). Following electroporation and Puromycin selection, 168 ES cell clones were picked (together with M.Osterwalder). Screening of positively targeted clones was done by PCR to discriminate between the different possible recombination events and the parental allele (Fig.30A,B). For the six conditions (two ES cell clones and three different tags), 76% of the 116 screened ES cell clones were correctly exchanged (Fig.30C). This percentage is very close to the 69% obtained in a previous targeting of the S4\textsuperscript{EC} allele (Osterwalder et al., 2010a). Southern Blot analysis using probes for both the 5' and 3' arms of the exchanged region and a probe detecting the Puro cassette confirmed correct recombination and single insertion of the targeting construct in all tested clones (Fig.30D). For each condition, three clones with normal karyotypes were analyzed further (data not shown).

The three versions of the epitope-tagged SMAD4 proteins were detected in ES cells using specific antibodies. This showed that tagged SMAD4 was best detected by using the 3xFLAG and 3xV5 epitope tags (Fig.31A). As the Hand2\textsuperscript{2xFLAG} allele had produced high quality ChIP-Seq data from limb buds (Osterwalder et al., in press), we selected the Smad4\textsuperscript{3xFLAG} allele for further analysis. As immunoprecipitation is the crucial step in the ChIP-Seq process, we used the anti-FLAG M2 antibody to establish its ability to IPP the SMAD4\textsuperscript{3xFLAG} protein (Fig.31B).
Figure 31. Detection tests of SMAD4<sup>3xTag</sup> in ES cells
(A) Western Blot detection of SMAD4<sup>3xHA</sup> (left), SMAD4<sup>3xV5</sup> (middle) and SMAD4<sup>3xFLAG</sup> (right) in protein extracts of two different S4<sup>3xF-dRMCE/+</sup> ES cell clones each. Wild-type ES cells were included as controls. SMAD4<sup>3xTag</sup> size: ~63kDa.

(B) Immunoprecipitation using the anti-FLAG M2 antibody on two S4<sup>3xF-dRMCE/+</sup> ES cell clones and wild-type cells. The top panel is blotted with a biotinylated anti-FLAG antibody and the lower panel with an anti-Actin antibody (size: 42kDa). On the left, 20µg of protein extracts were loaded as input controls. In the middle, the whole immunoprecipitate from 150µg of total proteins (1µg/µl) was loaded. On the right, 20µl of the unbound fraction after the IPP were loaded. The SMAD4<sup>3xFLAG</sup> protein is detected in the input and IPP samples of the S4<sup>3xF-dRMCE/+</sup> clones. All SMAD4<sup>3xFLAG</sup> protein was immunoprecipitated as revealed by its absence in the unbound fraction (top right panel). ACTIN was used as a loading control and is present in similar amounts in all samples.

(C) Immunofluorescence detection of the SMAD4<sup>3xFLAG</sup> protein in undifferentiated ES cells (left panels) and after twelve days of differentiation into embryoid bodies (right panels). Two different S4<sup>3xF-dRMCE/+</sup> ES cell clones were analyzed together with wild-type controls. Green: SMAD4<sup>3xFLAG</sup>. Blue: Hoechst. Scale bar: 20µm.

Finally, as SMAD4 is shuffling between cytoplasm and nucleus following activation of BMP signal transduction, immunocytochemistry was used to assess its cellular localization. In undifferentiated ES cells, the SMAD4<sup>3xF</sup> protein was detected both in the cytoplasm and in the nuclei (Fig.31C, left panel). Previous studies had shown that BMP signaling and consequential phosphorylation and nuclear translocation of SMAD1/5 are higher in differentiated ES cells (Fei et al., 2010; Galvin et al., 2010; James et al., 2005). Therefore, the SMAD4<sup>3xF</sup> protein distribution was assessed in embryoid bodies after 12 days of differentiation, at which Smad4 transcripts levels are the highest (M.O., unpublished result). Consistent with activation of BMP signal transduction in differentiating ES cells, we observed an increase in nuclear
SMAD4\(^{3x\text{FLAG}}\) proteins in comparison to undifferentiated ES cells (Fig. 31C, right panel). We thus concluded that the SMAD4\(^{3x\text{FLAG}}\) protein was functional.

12.1.2. Embryonic lethality of Smad4\(^{3x\text{FLAG}}\) homozygous embryos

Three different Smad4\(^{3x\text{FLAG}/+}\) ES cells clones (C57BL/6 background) were injected into BALB/c blastocysts at the Transgenic Mouse Core Facility (TMCF, Basel). Only one 80% chimeric male was obtained from clone Φ3A2 and no chimeras were obtained from the other two clones. This chimeric male was mated to C57BL/6 females and germline transmission was observed in 6% of the offspring and confirmed by PCR genotyping (see primers F2/R2 in Fig. 32F). As Smad4 is ubiquitously expressed, we could efficiently detect the SMAD4\(^{3x\text{FLAG}}\) protein from various embryonic tissues by Western Blot analysis of heterozygous embryos (E11.5, Fig. 32A). Anti-FLAG immunoprecipitation from embryonic forelimbs of S4\(^{3x\text{FLAG}/+}\) also established highly efficient and clear IP of SMAD4\(^{3x\text{FLAG}}\) proteins (E11.5, Fig. 32B).

Next, we wanted to assess a candidate direct target of SMAD4 by ChIP-qPCR in limb buds using the Smad4\(^{2x\text{FLAG}}\) allele. A good candidate is the BMP Responsive Element (BRE) located in the proximal promoter of Id1. The expression of Id1 is cleared from Smad4\(^{\Delta/\Delta}\) limb buds (E.P, unpublished result), and SMAD4 interacts in vitro with this BRE (Korchynskyi and ten Dijke, 2002; Lopez-Rovira et al., 2002; Katagiri et al., 2002). This conserved BRE overlaps with a DNase HS region (ENCOD/E UW) and a p300 ChIP-Seq peak in forelimb buds (E11.5) (Visel et al., 2009a) (Fig. 32C, left panel). Therefore, we assessed if the SMAD4\(^{3x\text{FLAG}}\) protein was interacting in vivo with this BRE by ChIP-qPCR from pooled wild-type and Smad4\(^{3x\text{FLAG}/+}\) embryonic limb buds versus wild-type limb buds (E11.75, Fig. 32C, right panel). We observed a small ~3x fold enrichment of Id1 BRE in the FLAG sample.

In parallel, we attempted to generate homozygous Smad4\(^{2x\text{FLAG}/3x\text{FLAG}}\) mice. However, we observed that this was not possible due to early embryonic lethality similar to the lethality of Smad4\(^{\Delta/\Delta}\) embryos at E7.5 (Fig. 32D, (Sirard et al., 1998). The likely explanation for this lethality is that the S4\(^{3x\text{FLAG}-dRMCE}\) allele was somehow damaged and its expression reduced to low levels. Indeed, quantification of the expression of Smad4 by RT-qPCR in embryonic limb buds showed that this was the case (E11.0, Fig. 32E).
Figure 32. Expression of the Smad4<sup>3xFLAG</sup> allele in mouse embryos

(A) Western Blot detection of SMAD4<sup>3xFLAG</sup> (top panel) and Actin (lower panel) from protein extracts of various embryonic tissues from a pool of S<sup>4+/-</sup> and S<sup>4xF/+</sup> embryos at E11.5. Protein extracts from wild-type forelimbs at E11.5 were used as a control (green). (B) Immunoprecipitation analysis from forelimb buds at E11.5. (C) Anti-FLAG ChIP from S<sup>4+/-</sup> and S<sup>4xF/+</sup> E11.75 forelimbs. Left panel: UCSC browser window representing the BMP Responsive Element (BRE) located -1.1kb from I<sup>d1</sup> TSS (black box). DNase HS and p300 ChIP-Seq profile from E11.5 limb buds are displayed above the placental mammal conservation (Cons) plot. Right panel: ChIP-qPCR for I<sup>d1</sup> BRE in S<sup>4+/-</sup> (green) and S<sup>4xF/+</sup> (blue) samples. Mean ±SD (n=2) is shown. (D) Bright-field pictures of a S<sup>4xF/+</sup> embryo at E9.0 (left) and dead S<sup>4xF/3xF</sup> embryo (right, white arrow). Scale bar: 0.5mm. (E) RT-qPCR analysis of Smad4 transcripts in limb buds at E11.0. Top panel: schematic representation of the Smad4<sup>wt</sup> and Smad4<sup>3xFLAG</sup> transcripts with the qPCR primers indicated by arrows. The area indicated by the red box in the Smad4<sup>3xFLAG</sup> transcript corresponds to the targeted region introduced by the pDREV vector. RT-qPCR amplification probed transcripts containing exons 6-8 (top left graph), exons 8-9 (top right graph), exon 12-3'UTR (bottom left graph) and Flag-3'UTR (bottom right graph) from the following genotypes: S<sup>4+/-</sup> (green), S<sup>4D</sup> (red) and S<sup>4xF/+</sup> (blue). The normalized fold expression is indicated
above each bar relative to the \( S4^{3xF/+} \) embryos (bottom right graph) or wild-type control. (F) *Smad4* transcripts levels following removal of the Puro cassette. Top panel: schematic representation of the *Smad4* \( 3xF/dRMCE;\Delta\text{Puro} \) allele obtained by mating \( S4^{3xF/+} \) females with \( Dre^{tg/+} \) males. F2/R2: primers used to genotype *Smad4* \( 3x\text{FLAG} \) versus *Smad4* \( \text{wt} \) allele. \( \Delta\text{PF}/\Delta\text{PR} \): primers used to genotype *Smad4* \( 3xF/dRMCE;\Delta\text{Puro} \) allele. Bottom panel: RT-qPCR in limb buds at E11.75 to detect *Smad4* transcripts encoding exons 6-8 (left) and exons 8-9 (right). The following genotypes were used: \( S4^{+/+} \) (green), \( S4^{3xF/+} \) (blue) and \( S4^{3xF;\Delta\text{P}/+} \) (grey). Normalized fold expression levels are indicated above each bar relative to the wild-type control.

Assessing *Smad4* expression at the junction between exons 6-8 (i.e. outside of the targeted region) and at the exons 8-9 junction (i.e. in the targeted region) showed that *Smad4* \( \text{wt} /+ \) embryos (Fig.32E, red) have \(~50\%\) of transcripts in comparison to *Smad4* \( +/- \) embryos (Fig.32E, green). While the expression of coding exons 6-8 in *Smad4* \( 3x\text{FLAG}/+ \) embryos was reduced to 81\% (Fig.32E, top left panel, blue), the expression of exons 8-9 dropped to 52\% in *Smad4* \( 3x\text{FLAG}/+ \) embryos, similar to *Smad4* \( \text{wt} /+ \) embryos (Fig.32E, top right panel, blue). These results and others established that the *Smad4* \( 3x\text{FLAG}-dRMCE \) allele is severely hypomorphic, as the majority of the *Smad4* \( 3x\text{FLAG}-dRMCE \) transcripts appear truncated in the crucial 3’ part. The small fraction of full-length transcripts is sufficient to produce SMAD4 \( 3x\text{FLAG} \) proteins at levels detectable by Western Blot and IP analysis (Fig.32A and Fig.32B), but not sufficient to sustain the viability of homozygous embryos (Fig.32D).

As the presence of the neighboring Puromycin selection cassette in the *Smad4* \( 3x\text{FLAG}-dRMCE \) allele (Fig.30A) could account for this reduction in transcription (Lewandoski, 2007), we used *Dre*-recombinase expressing mice to remove the \( \text{rox} \)-flanked PGK-Puro cassette (Anastassiadis et al., 2009). However, RT-qPCR analysis showed that the deletion of the Puro cassette did not rescue the expression levels of the *Smad4* \( 3x\text{FLAG}-dRMCE \) allele and embryonic lethality of homozygous embryos (Fig.32F and data not shown).

The reduced transcription level in the targeted region of the locus is thus likely due to a defect in the *Smad4* \( 3x\text{FLAG}-dRMCE \) allele itself. As the whole targeted region was fully sequenced and no mutation was detected, we suspect that the parental IKMC knockout-first conditional allele \( S4^{EC} \) could have already been damaged. The automated, computer-assisted design of the targeting vector used to generate IKMC
alleles could introduce errors that cannot be detected until the knockout-first conditional allele is tested (Skarnes et al., 2011). In particular, we realized that the FRT-promoterless cassette-loxP is inserted in Smad4 intron 8-9 close to the 5’ end of exon 9. Thus, the presence of the remaining FRT site in this intron in the targeted allele may affect the correct splicing of this region. Such incorrectly spliced transcripts could be eliminated by nonsense-mediated RNA decay, resulting in the observed reduction in the amount of transcripts coding exons 8-9 (not shown) and 9-10 (Fig.32E). Alternatively, the presence of the FRT site in intron 8-9 could have generated a cryptic polyA signal, leading to a majority of transcripts truncated after exon 8. These two possibilities have not been investigated further but provide potential explanations for the observed severe hypomorphic phenotype.

As the Smad43xFLAG-dRMCE allele is not expressed at wild-type levels, we tested another cassette exchange strategy. We introduced the whole coding sequence and endogenous promoter of Smad4 with the 3xFLAG epitope tag in the non-essential Rosa26 locus using ES cells previously engineered for efficient RMCE (Tsanov et al., 2012). The resulting Smad42xFLAG-R26 allele was expressed at levels lower than the Smad43xFLAG-dRMCE allele in ES cells (data not shown), and we decided to resort to conventional homologous recombination to generate a normally expressed Smad43xFLAG allele (see chapter 6.1.5).

12.2. Additional methods relative to the dRMCE targeting

12.2.1. Construction of the pcDNA3 Smad43xTag expression vectors

The Smad4 coding sequence (CS) was first amplified by PCR using primers to introduce an Asp718 and NarI restriction sites at the 5’- and 3’- ends, respectively. A Kozak sequence (GCCGCCACC) was introduced as part of the 5’ primer between the Asp718 site and Smad4 ATG codon to enhance translation. This Asp718-Smad4 CS-NarI fragment was inserted into a pBluescript II KS (pBSK, Stratagene) backbone vector previously constructed by J.Lopez-Rios (unpublished data) in frame with an epitope tag followed by a STOP codon. Two versions of this vector were used, one carrying a 3xFLAG and one with a 3xHA epitope tag. The whole Smad4 CS-3xTag-STOP fragments from the pBSK-Smad43xFLAG and pBSK-Smad43xHA plasmids were then transferred into the pcDNA3 (Invitrogen) mammalian expression vector as an
Asp718-EcoRI fragment. This resulted in the pcDNA3 Smad4\(^{3xFLAG}\) and pcDNA3 Smad4\(^{3xHA}\) expression vectors (Fig.33). A pcDNA3 vector expressing Smad4 with 3xV5 epitope tag was also constructed (see below).

**12.2.2. Construction of the pDREV \(S4^{3xTag}\) targeting vectors for dRMCE**

Three versions of the pDREV targeting vector for dRMCE were constructed using the three different epitope tags (3xFLAG, 3x HA and 3xV5). All were designed to target the promoterless cassette of the knockout first Smad4\(^{tm1a(EUCOMM)Wtsi}\) allele in ES cells by dual Recombinase Mediated Cassette Exchange (dRMCE, see (Osterwalder et al., 2010a)). First, a synthetic fragment containing a part of Smad4 intron 9/10, the cDNA of coding exons 9-12, an in frame 3xV5 tag and a STOP codon was synthetized by GeneArt in the pMA vector (pMA-Tag). This fragment was flanked by BstBI restriction site on the 5' end and by a EcoRI-MfeI-Ndel-Pacl polylinker on the 3' end. Subsequently, a BamHI-EcoRI fragment containing part of exon 12 and the 3xV5 tag was released from the pMA-Tag vector and introduced into the pcDNA3 expression vectors containing the Smad4 CS to replace the 3xFLAG tag with the 3xV5 (pcDNA3-Smad4\(^{3xV5}\) expression vector). In addition, BamHI-EcoRI fragments containing part of exon 12 and the 3xFLAG or 3xHA epitope tags were released from the pBSK-Smad4\(^{3xTag}\) vectors and introduced into the pMA-Tag to replace the 3xV5 BamHI-EcoRI fragment. A PCR fragment encoding Smad4 3'UTR was amplified from cDNA with primers introducing MfeI and Ndel restriction sites on the 5'-' and 3'-' ends respectively. The Smad4 3'UTR was then cloned into the three different pMA-Tag vectors as an MfeI-Ndel fragment. Finally, a BstBI-Pacl fragment containing the 3' part of Smad4 coding sequence, the epitope tag and the 3'UTR was released from the three pMA-Tag vectors and transferred into the pDREV-0 targeting vector (Osterwalder et al., 2010a). These pDREV-Smad4\(^{3xTag}\) vectors were the ones used for dRMCE (Fig.33).
Figure 33. Generation of the pDREV-Smad4\textsuperscript{3xTag} targeting vectors for dRMCE and pCDNA3-Smad4\textsuperscript{3xTag} expression vectors

(A) The Smad4 coding sequence was amplified by PCR and (B) inserted in the pBSK vector containing either a 3xFLAG or 3xHA epitope tag. (C) The expression vectors were generated by subcloning Smad4 CS\textsuperscript{3xTag} into pCDNA3. (D) A synthetic fragment containing the 3' end of Smad4 CS and a 3xV5 epitope tag was assembled by GeneArt in the pMA vector. (E) The fragments containing the 3xFLAG or 3xHA epitope tags were released from pBSK Smad4 CS\textsuperscript{3xTag} and exchanged with the equivalent 3xV5 fragment in the pMA vector. (F) The Smad4 CS\textsuperscript{3xV5} expression vector was generated by exchanging the 3xFLAG fragment with the 3xV5 fragment in the pCDNA3 vector. (G) The Smad4 3'UTR was amplified by PCR and (H) inserted in the three different types of pMA-Tag templates. (I) The fragment containing the 3' end of Smad4 CS, the epitope tag and the 3'UTR was recovered from the three versions of pMA-Tag templates and introduced into the pDREV-0 targeting vector by replacing the H2B-T2A Venus cassette. See main text for details.

12.2.3. Transfection of HEK-293T cells

Human Embryonic Kidney 293T cells (HEK-293T) were cultured in DMEM containing 4.5g/l Glucose, 10% FCS, 100U-0.1mg/ml Pen Strep, 2mM L-Glutamine, 1x Non-essential Amino Acids, 1mM Sodium Pyruvate at 37°C with 5% CO₂. The pCDNA3-Smad4\textsuperscript{3xTag} expression vectors were transfected into 60-80% confluent HEK-293T cells grown in 6-well plates using the FuGENE®6 Transfection Reagent (Promega). The pEGFP vector was co-transfected to assess the transfection efficiency, and the
empty pCDNA3 vector was transfected as negative control. HEK-293T cells were collected for protein extraction 24hrs after transfection.

12.2.4. dRMCE: ES cells electroporation, selection and screening
Two clones of the JM8.F6 ES cells carrying the Smad4 knockout first allele tm1a(EUCOMM)Wtsi (EUCOMM project ID 29509, called hereafter Smad4 EUCOMM Conditional (S4EC)) were targeted in parallel with the three versions of the pDREV-Smad4$^{3xTag}$ vector. ES cells were cultured and expanded for electroporation as described in the chapter 9.2.3 (see also ((Osterwalder et al., 2010b)). Briefly, $S4^{EC/+}$ cells were co-electroporated with both a pDREV-Smad4$^{3xTag}$ vector and the pDIRE vector, which encodes the iCre and Flpo recombinases for dRMCE (Osterwalder et al., 2010a). ES cells were selected for 9-11 days in 0.5µg/ml of Puromycin (Sigma P9620). After picking ES cell clones and expanding them in 48 well plates, a part of the ES cells were collected to isolate DNA while the remainder was further expanded. The ES cell clones were screened by PCR and negative or mixed clones were eliminated. Positive clones were expanded into 10cm dishes and frozen in aliquots.

The PCR screening strategy was designed to discriminate between the endogenous Smad4 wild-type allele ($S4^{wt}$, primers F1/R5), the Smad4 EUCOMM conditional allele ($S4^{EC}$, primers βgal/neo), the deleted EUCOMM allele ($S4^{DE}$, primers F1/R5) and positively targeted Smad4$^{3xTag}$-dRMCE allele both on both sides of the cassette exchange (primers PGK/R5 and F2/Rb). Correct targeting of the Smad4 locus was confirmed by Southern Blot analysis using both 5’ and 3’ probes and the Puromycin selection cassette. See Fig.30A for the location of the screening primers and Southern Blot probes; and chapter 12.2.6 for their sequences.

12.2.5. Differentiation of ES cells to embryoid bodies
$Smad4^{3xFLAG-dRMCE/+}$ ES cells into embryo bodies were differentiated using the protocol developed by (Kurosawa et al., 2003). 2x10$^4$ ES cells were resuspended in ES cell medium containing no LIF in order to induce differentiation and incubated at 37°C in low-adherence conical 1.5ml screw cap micro tube (Starstedt) for 5 days. The embryoid bodies were transferred to μ-Slide 8 well (Ibidi) and cultured for up to
two weeks. The medium was changed every second day, and differentiated embryoid bodies were fixed for immunocytochemical analysis.

12.2.6. dRMCE primers table

### Smad4 dRMCE targeting - cloning primers

<table>
<thead>
<tr>
<th>Target</th>
<th>Primer</th>
<th>Used for</th>
</tr>
</thead>
<tbody>
<tr>
<td>S4_CS_Asp718</td>
<td>5'-ATAAGTACCAGCCGCACCATGGCAATATGCTATAAC-3'</td>
<td>Forward for S4 CS with 5' Asp718, KozaK, ATG</td>
</tr>
<tr>
<td>S4_CS_Narl</td>
<td>5'-TATGGGCCGCTCTAAAGGCTGGGCTCC-3'</td>
<td>Reverse for S4 CS with 3' Narl</td>
</tr>
<tr>
<td>S4_3UTR_MfeI</td>
<td>5'-TAATCAATTGGATCTCACACCACCGG-3'</td>
<td>Forward for S4 3'UTR with 5' MfeI</td>
</tr>
<tr>
<td>S4_3UTR_Ndel</td>
<td>5'-TGGCATATGTTGATAGTGACTCAG-3'</td>
<td>Reverse for S4 3'UTR with 3' Ndel</td>
</tr>
</tbody>
</table>

### Smad4 dRMCE targeting - screening primers

<table>
<thead>
<tr>
<th>Locus</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Size</th>
<th>Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smad4 F1-R5</td>
<td>5'-ACTAACTCTGTGTCTACAGGCCGCCTG-3'</td>
<td>5'-GCTGCCAATCATTAGGCA-3'</td>
<td>1265bp</td>
<td>Smad4 Wt</td>
</tr>
<tr>
<td>Smad4 F1-R5</td>
<td>5'-ACTAACTCTGTGTCTACAGGCCGCCTG-3'</td>
<td>5'-GCTGCCAATCATTAGGCA-3'</td>
<td>559bp</td>
<td>Smad4 ΔEUCOMM</td>
</tr>
<tr>
<td>Smad4 PGK-R5</td>
<td>5'-GCAAAACAAAATTAAGGCGCA-3'</td>
<td>5'-GCTGCCAATCATTAGGCA-3'</td>
<td>456bp</td>
<td>Smad4 ΔEUCOMM-3 rec</td>
</tr>
<tr>
<td>Smad4 F2-Rb</td>
<td>5'-GCAATCCAACCAAGGCTTCAGG-3'</td>
<td>5'-CAGACAGACTGATGGCTGG-3'</td>
<td>1395bp</td>
<td>Smad4 -3xTag 5' rec</td>
</tr>
<tr>
<td>bgal-neo</td>
<td>5'-AGCAGAGCGGTGGAACCTGG-3'</td>
<td>5'-GCATCAGAGCAGCGGATTCCT-3'</td>
<td>557bp</td>
<td>Smad4 EUCOMM conditional</td>
</tr>
</tbody>
</table>

### Smad4 dRMCE targeting - Southern Blot probes primers

<table>
<thead>
<tr>
<th>Probe</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' probe</td>
<td>5'-ATGCAAAAAAAAGCTTTAAGGG-3'</td>
<td>5'-TTATAGGGAAGGCTCAAATGCAAG-3'</td>
<td>475bp</td>
</tr>
<tr>
<td>3' probe</td>
<td>5'-GAGCACCACCAAGCGCAAAGTT-3'</td>
<td>5'-TGTTTCTGCTAGGCTAAAGA-3'</td>
<td>487bp</td>
</tr>
<tr>
<td>Puro probe</td>
<td>5'-GCAAAACAAAATTAAGGCGCA-3'</td>
<td>5'-GCAGTTACATCGGCTG-3'</td>
<td>334bp</td>
</tr>
</tbody>
</table>

### Smad4 dRMCE - mice genotyping primers

<table>
<thead>
<tr>
<th>Locus</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Size</th>
<th>Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smad4</td>
<td>5'-AATTAGTACCAGCCGCACCATGGCAATAGTCTATAAC-3'</td>
<td>5'-CAGGGATTATGGAATCTGAGATG-3'</td>
<td>429bp + 536bp</td>
<td>Wt + 3xFlag</td>
</tr>
<tr>
<td>Dre</td>
<td>5'-TGCTGTCCTCCTATCCAC-3'</td>
<td>5'-CGGAATCTACGCTTAGG-3'</td>
<td>300bp</td>
<td>Tg</td>
</tr>
<tr>
<td>Puromycin</td>
<td>5'-GCAAAACAAAATTAAGGCGCA-3'</td>
<td>5'-GCGCCGAAAAATGCAAG-3'</td>
<td>456bp</td>
<td>Puro'</td>
</tr>
<tr>
<td>S4-ΔPuro</td>
<td>5'-GGTTTGCCAAACCATCATACT-3'</td>
<td>5'-GCTAAACACACGCTTGAAC-3'</td>
<td>498bp</td>
<td>ΔPuro</td>
</tr>
</tbody>
</table>
12.3. Manuscripts in preparation

“Mammalian NDR kinases are essential for cardiac development and impact on the Notch pathway”

Debora Schmitz-Rohmer¹, Simone Probst², Alexander Hergovich¹,⁴, Zhong-Zhou Yang³, Frédéric Laurent², Rolf Zeller² and Brian A. Hemmings¹

¹ Friedrich Miescher Institut for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland
² Department Biomedizin, Mattenstrasse 28, CH-4058 Basel, Switzerland
³ Model Animal Research Center of NanJing University, 12 Xue-Fu Road, Pukou District, NanJing, P.R. China 210061
⁴ current address: UCL Cancer Institut, University College London, London WC1E 6BT, United Kingdom

The aim of this project is to understand the roles of NDR1 and NDR2 kinases during mouse embryonic development by generating double knockout animals. Among other phenotypes, the heart tube of Ndr-deficient embryos does not loop and is filled with cells that obstruct the cardiac lumen and prevent blood flow. Ndr-deficient embryos die by E10.0 from cardiac insufficiency. I contributed to thy project by performing Optical Projection Tomography (OPT) analysis of wild-type and Ndr-deficient embryos at E7.75-E8.0. Analysis of virtual sections from 3D reconstructed embryos allowed better visualization of the cardiac obstruction phenotype observed in Ndr-deficient embryos.
“Generation of a bone organ in vivo through endochondral ossification by adipose-derived stromal cells”

Rik Osinga¹,², Nunzia di Maggio², Nima Allafi¹, Andrea Barbero², Frédéric Laurent³, Dirk J. Schaefer¹, Ivan Martin², Arnaud Scherberich²

¹ Department of Plastic, Reconstructive, Aesthetic and Hand Surgery, University Hospital of Basel
² Laboratory of Tissue Engineering, Department of Surgery, University Hospital of Basel and Department of Biomedicine, University of Basel, Basel, Switzerland
³ Department of Biomedizin, Mattenstrasse 28, CH-4058 Basel, Switzerland

In this project, adipose-derived stromal cells (ASC) from human tissues were cultured in micromasses or on collagen scaffolds and implanted subcutaneously in nude mice to investigate their potential to form bone tissues by endochondral ossification. These pellets and scaffolds were harvested after a few weeks and sectioned for histological analysis. I performed in situ hybridization on some of these sections for the chondrocyte differentiation marker Indian Hedgehog.
12.4. Curriculum vitae

FREDERIC LAURENT

Date and place of birth 17th December 1984, Meyrin, Switzerland
Origin Fey (VD), Switzerland
Nationality Swiss, French and Italian
Address Florastrasse 25, 4057 Basel, Switzerland
Phone number +41 (0)79 718 55 61
E-mail frederic.laurent@unibas.ch

EDUCATION AND DIPLOMAS

2010 – 2014 University of Basel, Department of Biomedicine, Developmental Genetics group
PhD in Genetics

2007 – 2009 University of Lausanne, Faculty of Biology and Medicine, School of Biology
Master of Science in Genomics and Experimental Biology

2004 – 2007 University of Lausanne, Faculty of Biology and Medicine, School of Biology
Bachelor of Science in Biology

2000 – 2003 High school of Nyon – Matura School in Biology and Chemistry
Matura work: “Can we learn by role playing games?” – Grade excellent
Bachelor and Matura Certificate

RESEARCH EXPERIENCE

06/2010 – 09/2014 PhD in Genetics
Supervisor: Prof. Dr Rolf Zeller
Department of Biomedicine, University of Basel
“Identification of cis-regulatory modules in mouse embryonic limb buds and heart using endogenous epitope-tagged transcription factors”

02/2008 – 01/2009 Master Thesis
Supervisor: Dr Elena Buetti
Institute of Microbiology, University Hospital of the canton of Vaud (CHUV)
“Analysis of signaling pathways involved in the transcription of the Mouse Mammary Tumor Virus using mutant PKCδ and Src kinases”

02/2008 – 05/2008 Assistant-student for a scientific forest inventory (“Bois de Chênes” reserve)
Institute of Terrestrial Ecosystems of Swiss Federal Institute of Technology, Zürich (ETH)

Supervisor: Dr Liliane Michalik
Center for Integrative Genomics, University of Lausanne
“Cell protection during skin repair: do Nrf2 and PPARβ combine their efforts?”

Supervisor: Prof. Winship Herr
Center for Integrative Genomics, University of Lausanne
“Characterization of CeHCF-1 in RNAi in Caenorhabditis elegans”
PUBLICATIONS


JD Benazet, E Pignatti, A Nugent, E Unal, F Laurent, R Zeller. Smad4 is required to induce digit ray primordia and to initiate the aggregation and differentiation of chondrogenic progenitors in mouse limb buds. Development 139, 4250-4260 (2012)

CONFERENCES PARTICIPATION

2014 Weinstein Cardiovascular Development Conference, Madrid, Spain – Poster
2013 26th Mouse Molecular Genetics Meeting, Hinxton, UK – Selected speaker
2007 Life Science Summer Research Program, EPFL, Lausanne, Switzerland – Poster

TECHNICAL SKILLS

Animal handling
Mouse care, husbandry and colony management. Embryo collection, in vivo wound healing experiments

Molecular biology
DNA and RNA extraction, PCR, RT-qPCR, molecular cloning with plasmids and BACs, chromatin immunoprecipitation, whole-mount in situ hybridization, Southern Blot, Western Blot

Cell culture
HEK293, mouse embryonic stem cells. Gene targeting by cassette exchange and homologous recombination, immunocytochemistry, karyotyping, transfection, viral transduction, luciferase reporter assays

Histology
Tissue preparation and sectioning with microtome and cryostat, immunohistochemistry, section in situ hybridization, TUNEL assay, hematoxylin/eosin and alcian blue stainings

Microscopy
Light, fluorescent and confocal microscopy, optical projection tomography

PROFESSIONALS AND SOFT SKILLS

• Creativity • Problems solving skills
• Initiative • Team building and leadership
• Flexibility/adaptability • Project management
• Highly organized • Event planning
• Work well under pressure

ASSOCIATIONS MEMBERSHIP

Department of Biomedicine PhD Club, Basel
2011 – 2014 Co-founder. Organization of scientific retreats, career days and social events for the PhD, MD-PhD and Master students of the Department

Association of Biology students, University of Lausanne
2006 – 2008 Responsible for events and animation for Bachelor and Master students

Scout group « Groupe de la Harpe », Rolle-Aubonne (VD)
1994 – 2008 Organization of weekly outdoors activities, weekends and summer camps for 7-11 kids
TEACHING EXPERIENCE

2011-2013   Assistant in Developmental Biology practical for medicine students, University of Basel
2014        Supervision of a 1st year PhD student regarding mouse colony management and molecular biology

SWISS MILITARY SERVICE

<table>
<thead>
<tr>
<th>Year</th>
<th>Rank</th>
<th>Corps</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>First Lieutenant</td>
<td>Cp EM hôp 2</td>
<td>School of Logistics Officers 1/10 (Bern)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Developing personal skills in management and leadership. Learning military techniques and tactics specifically applied to the army medical and logistics corps</td>
</tr>
<tr>
<td>2009–2010</td>
<td>Practical job as platoon leader, School of Hospital Recruits 41/3 (Moudon, VD)</td>
<td></td>
<td>Organizing and teaching basic military training for a platoon of medical soldiers</td>
</tr>
<tr>
<td>2004</td>
<td>First Lieutenant</td>
<td></td>
<td>School of Hospital Recruits 41/1 (Moudon, VD)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Basic military training as a medical soldier with specialization in emergency room</td>
</tr>
</tbody>
</table>

ADDITIONAL TRAININGS AND DIPLOMAS

2013   Leadership Level 1 Diploma – Swiss Managers Organization
        Group animation and dynamics, adult pedagogy, giving feedback, stress management, oral communication techniques, work techniques, ethics, co-workers motivation
2011   Optical Projection Tomography training in MRC Technology, Edinburgh, Scotland
2010   Certificate “Managing my time at the group/team level” – Swiss association for the Training of Executives (SVF-ASFC)
2009   Military sport instructor certificate
2007   Laboratory animal science course – Swiss Veterinary Office of canton of Vaud
        Legislation, ethical principles, biology and handling of laboratory animals, anesthesia, blood sampling and microsurgery
2004   Diploma of Health Care Assistant – Swiss Red Cross
2003   JV base training
        Health promotion and addictions prevention in youth associations of canton of Vaud
2002   Instructor license in trekking and sports camps – Youth + Sports (J+S), Swiss Federal Office of Sports
2002   Rescuer Certificate

AWARDS AND FELLOWSHIPS

2009   Honorary Certificate – Prof. Nicole Galland, School of Biology Director, University of Lausanne
        For active involvement as member of the committee of the association of students
2007   Summer Research Fellowship – School of Life Sciences, EPFL
        For the participation to the Summer Research Program in Life Sciences

LANGUAGES

<table>
<thead>
<tr>
<th>Language</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>Mother tongue</td>
</tr>
<tr>
<td>English</td>
<td>Fluent</td>
</tr>
<tr>
<td>German</td>
<td>Scholar level</td>
</tr>
</tbody>
</table>