CONSERVED TARGETS OF ISL1 DURING HINDLIMB INITIATION AND OUTGROWTH OF THE EXTERNAL GENITALIA

by

SERGIO G. MINCHEY

(Under the Direction of Douglas B. Menke)

ABSTRACT

Despite the incredible morphological diversity found within vertebrates, embryonic development is often regulated by shared foundational mechanism maintained from a common ancestor. For instance, the tetrapod limb displays tremendous variation in size and shape but largely develops according to conserved gene regulatory networks, including signaling pathways, transcription factors, and enhancers. The phallus of amniotes is also thought to share a basic homology originating from their last common ancestor over 300 million years ago. The genital tubercle (GT) of mammals – the embryonic precursor to the penis and clitoris – shows similarities with the external genitalia of other amniotes prior to sexual differentiation. Many genes and enhancers show a common role in both the limbs and the external genitalia, and disruption of these loci can cause aberrant phenotypes in both structures. Among these shared limb-genital genes is *Isl1*. The *Isl1* gene encodes a homeodomain transcription factor that is required for initiation of the hindlimb bud and for normal GT development. This dissertation uses a combination of RNA-seq, ChIP-seq, and comparative genomics to identify potential targets of ISL1 during both hindlimb initiation and GT outgrowth. ChIP-seq data from mouse, chick, and other species reveals targets that are likely conserved from the last common ancestor

of amniotes. A subset of ISL1 targets are shared between the hindlimb and the external genitalia, supporting the hypothesis that aspects of phallus development were coopted from the limb gene regulatory network. This systematic investigation of ISL1 transcriptional targets expands our knowledge of *Isl1*'s role in appendage development and generates further testable hypotheses concerning the establishment of these structures.

INDEX WORDS: ISL1, Islet1, Hindlimb, Limb Development, Genital Tubercle, Amniote Phallus, ChIP-seq, RNA-seq, Enhancers

# CONSERVED TARGETS OF ISL1 DURING HINDLIMB INITIATION AND OUTGROWTH OF THE EXTERNAL GENITALIA

by

SERGIO G. MINCHEY

B.S., University of Georgia, 2012

M.S., Georgia College & State University, 2015

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2022

© 2022

Sergio G. Minchey

All Rights Reserved

# CONSERVED TARGETS OF ISL1 DURING HINDLIMB INITIATION AND OUTGROWTH OF THE EXTERNAL GENITALIA

by

SERGIO G. MINCHEY

Major Professor: Douglas B. Menke

Committee: Brian G. Condie
James D. Lauderdale
Robert J. Schmitz
Michael A. White

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2022

ACKNOWLEDGEMENTS

Current and past members of the Menke lab have helped me with aspects of my project. Shana Pau introduced me to embryonic dissections in mice and chicks. Ashley Rasys provided help with lizard and alligator dissections and with imaging of embryos. Shirley Wang and Sungdae Park aided with analyzing sequence data. Sungdae answered countless questions about lab protocols and troubleshooting, maintained mouse lines that I used to generate *Isl1* knockouts, and generated the ISL1 ChIP-seq data in the mouse GT. Megh Mehta compiled the multi-species alignments shown in Figs. S2.3 and S3.8C. Aaron Alcala and I have carried out similar experiments and our cooperation has been valuable. My mentor Doug Menke has provided excellent guidance on my research and a supportive environment for conducting additional experiments. Doug's interest in studying appendage development outside of the mouse model led to an intriguing multi-species comparative approach facilitated with the help of collaborators.

Chicken eggs were provided by UGA Poultry Science, including Robert Beckstead, Jason Payne, and Jeanna Wilson. Alligator eggs were provided by Ben Parrott of UGA and the Savannah River Ecology Lab. Turtle samples were dissected by Marty Cohn's lab at the University of Florida, including Brooke Armfield, Rajee Rajakumar, and Darius Ramkhalawan. Other members of our lab have provided help in lab, feedback on my talks, and work to maintain our lizard colony, including Sukhada Samudra, Megan Meany, Christina Sabin, Anna Iouchmanov, Rida Osman, and Amber Rittgers. I wish continued success to those working to establish this emerging model organism. I also thank my committee for their guidance: Jim Lauderdale, Bob Schmitz, Mike White, and Brian Condie.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

# CHAPTER 1

# INTRODUCTION AND LITERATURE REVIEW

## Mechanisms of Limb Development

The tetrapod limb has long been studied as a model of embryonic development, morphological evolution, and in more recent years of gene regulatory networks (Zuniga and Zeller, 2020). The limb emerges from lateral plate mesoderm surrounded by ectoderm, with signaling centers that coordinate development along each of its three axes: proximal-distal, anterior-posterior, and dorsal-ventral. The organizing centers of the apical ectodermal ridge (AER) and the zone of polarizing activity (ZPA) were described by the pioneering work of John Saunders using the chicken embryo as a model system (Tickle, 2017). Saunders (1948) discovered the importance of the AER by removing it at different stages, with earlier removal resulting in greater truncation of the wing. Later grafting experiments identified the ZPA as capable of inducing posteriorization of the anterior wing bud, resulting in a mirror duplication of the digits (Maccabe et al., 1973).

Researchers would later search for the candidate molecular factors underlying these organizing centers. Placement of retinoic acid (RA) in the anterior wing bud had a similar effect as the ZPA grafts (Summerbell, 1983; Tickle et al., 1982). Riddle et al. (1993) showed that RA is induced by the gene *Sonic hedgehog* (*Shh*), which encodes a secreted ligand protein. They found that implanting cells expressing *Shh* is sufficient to induce similar mirror duplications as found in ZPA grafts, thus identifying *Shh* as the crucial ZPA morphogen. Mouse knockout models reinforce the necessity of *Shh*, showing severe truncations of the limbs (Chiang et al., 1996).

Vogel and Tickle (1993) found that removal of the posterior AER disrupts the ZPA, suggesting interaction between the two signaling centers.

Fibroblast growth factor 2 (FGF2) or FGF4-soaked beads are able to replace the function of the AER, pointing towards FGF proteins as important factors in limb development (Fallon et al., 1994; Niswander et al., 1993; Vogel and Tickle, 1993). FGF1, FGF2, and FGF4 can induce ectopic limb buds when implanted in the flank of chick embryos (Cohn et al., 1995; Ohuchi et al., 1995). These are capable of developing into complete limbs with wing or leg identity depending on the placement of the signal along the antero-posterior axis. The discovery that *Fgf8* is expressed throughout the AER in mouse beginning at early stages of limb initiation made it the best candidate for the major regulator of the AER (Crossley and Martin, 1995; Heikinheimo et al., 1994; Mahmood et al., 1995; Ohuchi et al., 1994). This is also seen in chicks, where FGF8 is capable of inducing ectopic limb buds, rescuing loss of the AER and sustaining *Shh* expression (Crossley et al., 1996; Mahmood et al., 1995; Vogel et al., 1996).

Saunders and Reuss (1974) demonstrated the inductive power of the early chick wing bud mesoderm. Stripped of ectoderm and placed under the flank ectoderm, this tissue results in the formation of ectopic limbs with an AER. The lack of *Fgf8* in the limb mesenchyme suggests another FGF factor functions upstream of AER formation. *Fgf10* becomes restricted to the chick limb mesenchyme, preceding *Fgf8* expression in the AER, and FGF10 can induce ectopic limbs in the flank (Ohuchi et al., 1997). Furthermore, deletion of *Fgf10* in mice results in total disruption of forelimb and hindlimb formation (Min et al., 1998; Sekine et al., 1999).

*Fgf10* is no longer able to activate *Fgf8* in the AER when deleting the FGF transmembrane receptor gene *Fgfr2*, thus ablating limb outgrowth and reducing mesenchymal expression of *Fgf10* (Xu et al., 1998). This identifies the mechanism by which mesenchymal

*Fgf10* and ectodermal *Fgf8* mutually reinforce each other in a positive feedback loop facilitating outgrowth. Prior to limb initiation and the establishment of this feedback loop, the somatopleure undergoes a transition from epithelial to mesenchymal identity at the level of the forelimb and hindlimb field (Gros and Tabin, 2014). By reanalyzing the cellular dynamics of FGF10-induced ectopic limbs in the chick and *Fgf10*-null mice, Gros and Tabin found that *Fgf10* plays a role in promoting this transition to a mesenchymal state.

The bone morphogenetic protein (BMP) signaling pathway also contributes to the establishment of the AER during limb initiation. BMP ligand genes are expressed in the limb field of both mouse and chick from early stages, with *Bmp4* showing a broad mesenchymal expression (Robert, 2007). BMP4 signals via the BMP receptor IA, and early deletion of *Bmpr1a* in the ectoderm can cause agenesis of the hindlimb or limbs that develop a defective AER (Ahn et al., 2001; Pajni-Underwood et al., 2007). The transcription factor SMAD4 acts as the effector of this pathway in the AER, and its deletion in these cells disrupts *Fgf8* expression and causes defective AER formation in the hindlimb (Benazet and Zeller, 2013). These studies show that the effect of BMP signaling on the AER varies by time, as it inhibits *Fgf8* during limb outgrowth once the ZPA is established. *Shh* then activates the BMP inhibitor *Gremlin1*, thus maintaining the AER which in turn maintains *Shh* (Bénazet et al., 2009; Zúñiga et al., 1999).

Retinoic acid (RA) is another signaling pathway found to play a crucial role in limb initiation (Feneck and Logan, 2020). Classic embryonic chick experiments demonstrated the inductive potential of the paraxial/somitic mesoderm taken from the level of the forelimb or hindlimb field (Murillo-Ferrol, 1965; Pinot, 1970). Murillo-Ferrol (1965) also described that placement of an impermeable barrier between the paraxial and lateral plate mesoderm could disrupt limb development, suggesting the “presence of a factor arising from the somites” which

can diffuse into the limb. RA is found in the chick limb bud, and its inhibition using disulphiram disrupts forelimb initiation (Stratford et al., 1996). Deletion of *Raldh2* in mouse, which encodes an enzyme for RA synthesis, disrupts forelimb development before causing embryonic lethality (Niederreither et al., 1999).

Nishimoto and colleagues found that blocking RA with an impermeable barrier in chick disrupts hindlimb as well as forelimb initiation, preventing expression of *Fgf10, Fgf8,* and *Shh* (Nishimoto et al., 2015). Both FGF and RA soaked beads are capable of rescuing limb initiation in either the forelimb or hindlimb, and RA induces *Fgf10* expression in the LPM. *Raldh2* is expressed normally in the forelimb LMP following barrier insertion, suggesting RA produced locally in the limb is insufficient. RA secreted from the somites thus appears to be the essential paraxial signal. RA regulates gene expression by binding to RA receptors (RAR) that bind to RA response elements (RARE) on the genome (Berenguer and Duester, 2021). RA is degraded by CYP26 enzymes, and the *Cyp26b1* gene is expressed in the distal limb bud following initiation (Yashiro et al., 2004). RA promotes *Meis1/2* expression, and thus *Cyp26b1* indirectly represses these genes and restricts them to the proximal limb bud. *Meis1/2* encode transcription factors that are required early for hindlimb initiation and later for the development of proximal structures (Delgado et al., 2020, 2021).

Nishimoto and coauthors additionally demonstrated that removal of the WNT-signaling factor β-catenin in mouse results in truncated forelimbs (Nishimoto et al., 2015). At the posterior level, β-catenin removal causes a more severe hindlimb agenesis phenotype (Kawakami et al., 2011). These results are consistent with the finding that particular WNT genes are sufficient to induce ectopic budding and *Fgf10* expression in the chick (Kawakami et al., 2001). *Lef1* and *Tcf7* encode transcription factors that interact with β-catenin in response to WNT signaling.

Compound deletion of these genes in mouse does not prevent limb initiation, but causes defective outgrowth and the absence of the AER (Galceran et al., 1999).

## Factors Specific to the Forelimb or Hindlimb

As revealed by studies of these crucial signaling pathways, the serially homologous structures of the forelimb and hindlimb share various major mechanisms of development (Duboc and Logan, 2011). Hamburger (1938) found that transplants of forelimb or hindlimb primordia onto the flank maintained the limb identity from the donor cells. Peebles (1911) is credited as the first to attempt heterotopic grafts, placing the forelimb bud in place of the hindlimb and vice-versa, albeit to inconclusive results. Saunders et al. (1959) reported that early hindlimb bud mesenchyme grafted under the forelimb AER differentiated into leg parts. These experiments suggest that the factors underlying limb identity are within the mesenchyme of the limb primordia starting from early stages of development.

When researchers later examined expression patterns of T-box transcription factors, they found that *Tbx5* and *Tbx4* are expressed from early stages in the forelimb and hindlimb mesenchyme, respectively, in both mouse (Gibson-Brown et al., 1996) and chick (Gibson-Brown et al., 1998; Isaac et al., 1998; Logan et al., 1998). Ectopic limbs induced in the flank express the T-box factor associated with the type of limb formed (Gibson-Brown et al., 1998; Isaac et al., 1998; Logan et al., 1998; Ohuchi et al., 1998). Heterotopic grafts between limbs maintain expression of the T-box factor expressed from the donor graft (Gibson-Brown et al., 1998; Isaac et al., 1998). Deletion of *Tbx5* in the mouse prevents development of the forelimb and initiation of *Fgf10* expression (Agarwal et al., 2003; Rallis et al., 2003). Blocking RA from the somites,

however, shows that *Tbx5* is not sufficient for forelimb initiation or *Fgf10* activation without RA signaling (Nishimoto et al., 2015).

*Tbx4* is also not sufficient for hindlimb initiation in the absence of somitic RA secretion. Furthermore, deletion of *Tbx4* in mice shows that it is not strictly necessary for the initiation of hindlimb budding or *Fgf10* expression, though development is severely ablated and *Fgf10* is not maintained (Naiche and Papaioannou, 2003). Later conditional deletion does not result in the same *Fgf10* deficiency, showing a crucial early role (Naiche and Papaioannou, 2007). *Tbx4* can replace *Tbx5* function and rescue forelimb development, suggesting the differing roles of these paralogs may not be intrinsic to their structure, but rather because of differences in the regulatory network of forelimbs and hindlimbs (Minguillon et al., 2005). Contrary results were found in the chick, where misexpression of *Tbx4/5* can alter limb identity (Rodriguez-Esteban et al., 1999; Takeuchi et al., 1999), and ectopic limbs induced in the flank by *Tbx4* or *Tbx5* differ in their morphology (Takeuchi et al., 2003). *Cis*-regulatory changes near *Tbx5* are associated with feathered feet in certain breeds of pigeons and chickens, and these birds exhibit ectopic *Tbx5* expression in the embryonic hindlimb (Domyan et al., 2016). These findings suggests that *Tbx4* and *Tbx5* may have different roles in limb identity between birds and mammals.

PITX1 is another transcription factor with a hindlimb-restricted expression domain beginning prior to hindlimb initiation in both mouse and chick (Lanctôt et al., 1997; Szeto et al., 1996). Deletion of *Pitx1* in mouse shows that it is not required for hindlimb initiation, though these hindlimbs exhibit lower *Tbx4* expression and deficiencies in skeletal development (Lanctôt et al., 1999; Szeto et al., 1999). In addition, misexpression of *Pitx1* in the chick forelimb bud results in morphological changes consistent with a partial shift towards hindlimb identity, along with increased *Tbx4* expression (Logan and Tabin, 1999; Szeto et al., 1999). Overexpression in

the mouse forelimb also causes a shift towards hindlimb identity, increased *Tbx4* expression, reduction of *Shh* expression and loss of posterior digits (DeLaurier et al., 2006). Together, these results suggest *Pitx1* is upstream of *Tbx4* and promotes hindlimb identity.

Ouimette and colleagues confirmed the loss of hindlimb identity and *Tbx4* expression in *Pitx1*-null mice (Ouimette et al., 2010). Restoring normal *Tbx4* levels in these mutants rescues some features of hindlimb skeletal and muscular development, while driving *Tbx5* expression does not. Using a luciferase reporter construct with T-box binding elements, they found that *Tbx5* acts as an activator while *Tbx4* can act as a repressor in vitro. They identified a repressor domain specific to *Tbx4* that may account for its ability to partially induce hindlimb identity. A recent study by Duboc and colleagues nonetheless shows that *Tbx5* can completely rescue hindlimb development in *Tbx4* knockouts (Duboc et al., 2021). Thus, in the mouse, while *Tbx4* appears to have some ability to control limb identity, this does not appear essential in the presence of *Pitx1*. This study also finds that compound deletion of *Tbx4* and *Pitx1* completely ablates hindlimb budding, showing that they act in parallel during hindlimb initiation.

These genes also play a role during paired fin development in jawed fish, whose last common ancestor with birds and mammals lived over 400 million years ago (Brazeau and Friedman, 2015). Knockdown or mutations of the *tbx5a* gene in zebrafish show its requirement for pectoral fin development (Ahn et al., 2002; Garrity et al., 2002; Ng et al., 2002), and a naturally-occurring mutation in the nuclear localization signal of *Tbx4* in zebrafish causes loss of pelvic fins (Don et al., 2016). Populations of threespine stickleback lacking a pelvic spine show a loss of pelvic *Tbx4* and *Pitx1* expression (Cole et al., 2003). Trait mapping points towards a deletion of a *Pitx1* enhancer underlying this morphological evolution (Chan et al., 2010; Shapiro et al., 2004). While functional tests have not been carried out in the more distantly-related

cartilaginous fish, similar expression domains of *Tbx4* and *Tbx5* are found in the shark (Tanaka et al., 2002).

Thus some mechanisms of tetrapod limb development – including differences between the forelimb and hindlimb – appear to have arisen during the evolution of paired fins. Phylogenetic analysis suggests that genome duplication resulted in pairs of T-box genes including *Tbx5* and *Tbx4*, which arose from a single ancestral gene still found in the finless *Amphioxus* (Ruvinsky et al., 2000). This *AmphiTbx4/5* gene can rescue forelimb development in *Tbx5*-null mice (Minguillon et al., 2009) and acts as an activator in vitro like *Tbx5* (Ouimette et al., 2010). During the evolution of paired fins, these genes may have acquired differences in their *cis*-regulatory landscapes allowing for specific functions in the pelvic or pectoral fins, along with the gain of a repressor domain in *Tbx4*. Conserved domains of activity are also seen with some limb enhancers like the *Tbx4* enhancer HLEB, which shows hindlimb-specific activity in reporter assays using the mouse or stickleback sequence (Menke et al., 2008).

## Transcriptional Regulation by Enhancers

Enhancers are a type of *cis*-regulatory element that are essential for controlling transcription in different cell types during development (Long et al., 2016). Enhancers can be activated in particular cells by pioneer factors that bind to closed DNA wrapped in nucleosomes. These pioneer factors can both facilitate the eviction of nucleosomes and the recruitment of histone modifying enzymes, thus opening chromatin and allowing binding of non-pioneer transcription factors (TFs). These TFs bind preferentially to particular short sequence motifs found within the open DNA and cooperate with coactivators which can then influence the activity of RNA polymerase. Active enhancers are characterized by clustered binding of multiple

TFs and are surrounded by the histone modifications H3K27ac and H3K4me1. H3K27ac in particular distinguishes active enhancers from poised enhancers marked by H3K4me1 alone (Creyghton et al., 2010). Distal, non-promoter H3K27ac enrichment is highly variable between cell types and is positively associated with transcription of proximal genes.

Chromatin looping in the nucleus allows these distal enhancers to regulate genes that appear distant in the linear genome. Looping of DNA was hypothesized in the 1980s as an explanation for the ability of proteins to seemingly regulate genes from a distance in bacteria (Ptashne, 1986). Researchers from this era also elucidated enhancer function in animal cells, showing they can function in different orientations or distances, upstream or downstream of a gene (Schaffner, 2015). The dynamics of eukaryotic 3D chromatin architecture in particular contexts can now be probed using various chromosome conformation capture techniques (Yu and Ren, 2017). This reveals that chromatin loops occur within large chromosomal compartments called topologically associating domains (TADs), which are widely conserved between mouse and human. Some are also known to be conserved beyond mammals, as with the TADs regulating the *Hoxd* locus in both mouse and chick (Yakushiji-Kaminatsui et al., 2018). Looping is facilitated by the DNA-binding protein CTCF and the cohesin protein complex. Chromatin looping allows for very long-range interactions between enhancers and promoters, as with the *Shh* limb enhancer ZRS located ~1 Mb from *Shh*'s promoter (Lettice et al., 2003).

ZRS also provides an example of enhancer conservation, as multi-species alignments of this locus show orthology across jawed vertebrates (Kvon et al., 2016; Lettice et al., 2003). ZRS sequences from different species show similar activity in mouse transgenic assays except from snakes, suggesting enhancer decay following limb loss (Kvon et al., 2016). The *Tbx4* hindlimb enhancer HLEB shows a similar pattern of vertebrate conservation along with snake-specific

divergence in activity (Infante et al., 2015; Menke et al., 2008). Following the accumulation of whole genome sequences, conservation has become a criterion in searching for functional elements due to negative selection applying more constraint to these loci (Birney et al., 2007).

While enhancers are typically more conserved than the background non-functional sequence, they are still generally much more evolutionarily labile than protein-coding regions. This relates to the concept of enhancer modularity, whereby tissue-specific enhancers are thought to be a major source of phenotypic evolution, while TFs are more pleiotropic and thus constrained in evolution by their multi-faceted necessity during embryonic development. Genes can also be regulated by multiple enhancers, providing redundancy and buffering the effects of mutations (Kvon et al., 2021). The role of coding changes in evolution is a subject of debate, and TFs can exhibit their own forms of modularity (Cheatle Jarvela and Hinman, 2015). Additionally, examples of enhancer pleiotropy have been identified in both flies and vertebrates, thus complicating the traditional paradigm of enhancer modularity (Sabarís et al., 2019).

The advent of high-throughput sequencing and thus ChIP-seq has allowed genome-wide binding profiles of TFs to be identified within certain cells or tissues (Johnson et al., 2007; Robertson et al., 2007). ChIP-seq on two liver TFs in different species of mammals and chicken shows that most peaks are not conserved, suggesting widespread turnover of TF binding sites (Schmidt et al., 2010). Known functional targets of these factors were however enriched for conserved binding events. ChIP-seq data from PITX1 in the mouse hindlimb shows binding at putative limb enhancers indicated by H3K27ac (Infante et al., 2013). HLEA is a *Tbx4* hindlimb enhancer conserved in mammals, though not across vertebrates like HLEB (Menke et al., 2008). PITX1 binds most clearly to HLEA, and possibly also acts via HLEB. Moreover, ChIP-seq on

PITX1 in the hindlimb of the *Anolis* lizard reveals conserved binding at HLEB, as well as other non-conserved binding loci near *Tbx4* (Wang et al., 2018).

Combining ChIP-seq and RNA-seq suggests that PITX1 directly targets genes involves in chondrogenesis and myogenesis (Wang et al., 2018). Two RNA-seq analyses confirm the previous finding that *Tbx4* is downstream of PITX1 (Nemec et al., 2017; Wang et al., 2018). These results suggest that a conserved function of PITX1 across amniotes is the direct activation of *Tbx4*. ChIP-seq data from TBX4 and HOXC10 in the hindlimb reveals hundreds of putative instances of co-binding at limb enhancers, identifying a composite T-box-Hox motif enriched at these sites (Jain et al., 2018). Most of these TBX4-HOXC10 sites also overlap with PITX1 peaks. MEIS1 ChIP-seq in the mouse forelimbs and hindlimbs also finds that it often binds alongside T-box and Hox binding loci (Delgado et al., 2021).

ChIP-seq data from histone modifications can survey enhancers independent of particular TFs. Andrey and colleagues combined ChIP-seq on multiple histone marks with Capture-C in order to identify over 1000 putative limb enhancers at three different stages (Andrey et al., 2017). H3K27ac and H3K4me1 mark putative enhancers, while Capture-C associates promoters of hundreds of limb genes with these enhancers. This method identifies the spatially-restricted ZRS enhancer and many other loci that previously showed limb activity in transgenic mouse assays. Extensive overlap in gene expression and enhancer activity is seen in forelimbs and hindlimbs at the same stage, while greater differences are seen at different stages.

ATAC-seq is a technique that more directly identifies open chromatin genome-wide, allowing putative active enhancers to be systematically surveyed (Buenrostro and Giresi, 2013; Buenrostro et al., 2015). Young and colleagues used this to find enhancers with differential activity in the early emu forelimb bud as compared to the chick to identify possible mechanisms

of emu wing reduction (Young et al., 2019). The many thousands of ATAC-seq peaks found in these datasets were filtered to find only dozens of loci active in the hindlimbs of both species, the forelimb of chick, but not the forelimb of the emu or the flank of either species. Two of these loci occur near *Fgf10* and *Sall1*, genes which are underexpressed in the emu forelimb as shown by RNA-seq, while the *Sall1* locus drives reporter activity in the chick forelimb bud.

A recent analysis used a combination of RNA-seq and ATAC-seq to characterize gene expression and chromatin accessibility at three different stages of forelimb development in both mouse and chick (Jhanwar et al., 2021). This allows for the identification of differentially expressed genes (DEG) and differentially accessible chromatin (DAC) modules, which show significant correspondence with each other in mouse. For example, loci with a more open chromatin state at E9.75 than at E10.5 or E11.5 tend to occur near genes that also show a higher expression at E9.75. The same clear correspondence between DEG and DAC modules is not seen in the chick, although limb development genes show similar temporal dynamics between the two species, including genes in the Hox family and the BMP and FGF pathways.

## *Isl1* in Hindlimb Initiation

The *Isl1* gene – also called *Islet1* – was first identified in cDNA from an insulin-producing cell line derived from rat pancreatic islet cells (Karlsson et al., 1990). Electrophoretic mobility shift assays suggest its ability to bind insulin gene enhancers via the homeodomain motif TAAT. ISL1 has a cysteine-rich LIM domain, which is a protein-protein interaction domain named after *Isl1* and two other genes identified in *C. elegans* that share this domain – ***L****in11*, ***I****sl1*, and ***M****ec3* (Freyd et al., 1990). *Isl1* is expressed during motor neuron differentiation (Ericson et al., 1992), and *Isl1*-null mice show a lack of motor neuron development prior to

embryonic lethality (Pfaff et al., 1996). *Isl1* and its paralog *Isl2* are among the 12 LIM-homeodomain genes identified in the mammalian genome, including *Lhx* and *Lmx* genes (Hunter and Rhodes, 2005).

Combinations of these proteins – along with the *Ldb* genes that lack DNA-binding domains – are thought to constitute a "LIM code" which can specify different cell fates (Gadd et al., 2011). Expression of *Ngn2* with *Isl1* and *Lhx3* is sufficient to drive spinal motor neuron differentiation in embryonic stem cells, while replacement of *Lhx3* with *Phox2a* causes cranial motor neuron fate (Mazzoni et al., 2013). ISL1 forms complexes with LHX3 and possibly the non-LIM homeodomain factor PHOX2A in these contexts, and ChIP-seq indicates that many ISL1 binding loci depend on the presence of these partners. These context-dependent binding loci are associated with differentially expressed genes between the two fates, as well as somewhat different sequence motifs (Mazzoni et al., 2013). In the context of cardiac differentiation, ISL1 can be found in complexes with the histone acetyltransferase p300 (Yu et al., 2013), the H3K27 demethylase JMJD3 (Wang et al., 2016), and the chromatin remodeling BRG1-BAF60C-based SWI/SNF complex (Gao et al., 2019).

Shiga and colleagues found ISL1 protein (though not ISL2) in a transient expression domain of the chick hindlimb field mesenchyme, while it does not appear in the forelimb mesenchyme (Shiga et al., 2002). This hindlimb expression is found during limb initiation at Hamburger-Hamilton (HH) stage 17, persists in the HH20 hindlimb bud, but largely is not found by HH24. This analysis also found ISL1 expressing cells in the AER of both forelimbs and hindlimbs, which persists through HH24 after mesenchymal expression diminished. A hindlimb-restricted domain of *Isl1* transcript is also found in mouse, and *Isl1*-Cre shows that *Isl1*-expressing cells broadly contribute to the hindlimb mesenchyme, though not to the ectoderm

(Yang et al., 2006). *Isl1*-null mice show heart defects and arrest by E9.5 (Cai et al., 2003; Pfaff et al., 1996), thus precluding their use for studying appendage development. Use of a *T*-Cre allele allows for conditional knockout (cKO) of *Isl1* around E7.5, resulting in a complete failure to initiate hindlimb budding (Kawakami et al., 2011; Narkis et al., 2012).

In situ hybridizations performed in these studies show a dramatic loss of *Fgf10/8* expression, more modest reduction in *Tbx4* expression, but no apparent disruption of *Pitx1*. Deletion of the LIM co-regulators *Ldb1/2* also completely prevents hindlimb initiation and *Fgf10* expression (Narkis et al., 2012). These *Ldb* mutants survive longer than the *Isl1 T*-Cre cKOs and show that hindlimbs fail to emerge by E14.5, suggesting this is not merely caused by a delay in development. Kawakami et al. (2011) also found that these *Isl1* cKOs show a disruption of nuclear localization of β-catenin, suggesting a role upstream of the WNT signaling pathway. Conditional deletion of *Ctnnb1* – the gene which encodes β-catenin – reveals a similar requirement for hindlimb initiation and *Fgf10/8*, though not for *Pitx1*. The *Isl1* and *Ctnnb1* mutants both show a significant reduction in cell proliferation. While β-catenin is not required for initial *Isl1* expression, it appears to play a role in maintaining its expression.

Itou and colleagues used *Hoxb6*-Cre mice to conditionally delete *Isl1* later, thus allowing hindlimb buds to form but eliminating ISL1 to a variable degree by E9.25-E9.5 (Itou et al., 2012). This causes varied loss of the posterior-most digits, as well as zeugopod and pelvic bones (Fig. 1.1). There is a corresponding loss of *Shh* expression, variable reduction of its downstream target markers *Gli1* and *Ptch1*, reduced posterior cell cycle gene expression, and expansion of the anterior genes *Alx4* and *Irx3*. *Hand2* expression is also reduced, which encodes a TF that activates *Shh* expression in the ZPA (Galli et al., 2010), suggesting a potential link through

which ISL1 activates *Shh*. In addition, this mutant model shows reduction of genes involved in pelvic development like *Twist1*, *Prrx1*, and *Pax1*.

Deletion of *Pitx1*, *Tbx4*, or compound deletion of both does not disrupt *Isl1* expression in the hindlimb field (Duboc et al., 2021). This suggests that *Isl1* is not downstream of either of these factors, but acts in parallel with them and possibly upstream of *Tbx4*. *Gdf11* encodes a secreted TGFβ factor that is involved in the trunk-to-tail transition and is a strong candidate for an upstream activator of *Isl1*. *Gdf11* deletion in mouse causes the hindlimbs to emerge in a substantially more posterior position, whereas precocious activation of GDF signaling using a transgene results in anteriorization of the hindlimbs and truncated axial development (Jurberg et al., 2013). Using the same transgene to precociously drive *Isl1* expression causes a similarly anteriorization phenotype.

Putative binding sites of SMAD2 – a downstream effector of GDF11 signaling – are found in the *Isl1* enhancer CR2 that was initially identified in cardiac cells (Jurberg et al., 2013). This enhancer drives expression in the lateral plate mesoderm of the presumptive hindlimb field, though its activity is abolished upon deletion of multiple SMAD motifs. This indicates a mechanism by which *Gdf11* from the posterior paraxial mesoderm activates *Isl1* in the surrounding lateral plate mesoderm, thus terminating axial elongation and initiating development of the hindlimbs and external genitalia at this juncture. This relationship between *Gdf11* and *Isl1* in the hindlimb field is conserved in chicks, and modulation of *Gdf11* provides a mechanism for tetrapods to evolve different axial lengths (Matsubara et al., 2017b).

## Genetics of External Genital Development

The genital tubercle (GT) is the embryonic precursor to the penis and clitoris, which is found not only in mammals but in other groups of amniotes (Fig. 1.2). The mouse GT forms by day E11.5 and undergoes the early phase of ambisexual development (Georgas et al., 2015). Sexual dimorphism does not become anatomically evident until E16.5 (Suzuki et al., 2002), although the male-specific pattern of androgen receptor (AR) distribution becomes apparent by E15.5 (Zheng et al., 2015). In males, AR responds to androgen hormones by localizing to the nucleus to regulate gene expression, thus inducing masculinization and penile development. Like the limb, the GT emerges as an outgrowth consisting of mesoderm surrounded by ectoderm, although additionally the urethral plate is derived from the cloacal endoderm (Cohn, 2011; Seifert et al., 2008). Many genes and signaling pathways involved in limb development also appear to be involved in development of the GT, including the hedgehog, WNT, BMP, and FGF pathways (Tarulli et al., 2021).

Compound deletion of *Hoxa13* and *Hoxd13* in mice causes a failure to form both the digits and the GT (Kondo et al., 1997; Warot et al., 1997), while human mutations in each gene are associated with limb and penile defects (Roux et al., 2019; Tüzel et al., 2007). *Wnt5a*-null mice show truncation of the limbs and GT with variable penetrance (Seifert et al., 2009a; Yamaguchi et al., 1999), while human mutations are associated with Robinow syndrome which includes limb and genital defects (Person et al., 2010). *Fgf8* is expressed in the distal urethral epithelium (DUE), somewhat akin to its narrow epithelial domain in the AER, although in a different germ layer (Haraguchi et al., 2000). *Fgf10* is expressed in the mesenchyme and *Fgfr2* in both layers. In contrast to the limbs, deletion of *Fgf10* does not prevent GT outgrowth or DUE

*Fgf8* expression, though it causes more subtle defects later. Removal of the DUE or inhibition of FGF8 inhibits outgrowth in cultured GTs, while addition of FGF8 can restore growth.

Canonical WNT/β-catenin signaling also functions in the DUE, and conditional inactivation in these cells results in a failure of proper GT formation along with increased apoptosis and decreased proliferation (Lin et al., 2008). This also results in loss of *Fgf8* in the DUE, while overexpression causes both increased growth and *Fgf8* expression. Lin and colleagues also performed comparable loss and gain-of-function experiments in the limb ectoderm, showing a similar dependence of AER *Fgf8* on β-catenin signaling. The DEU signaling center appears to be upstream of *Bmp4* and *Wnt5a* in the surrounding mesenchyme, as well as *Shh* in the urethral endoderm. Disruption of β-catenin in the mesenchyme or ectoderm also results in defective GT development. Miyagawa and colleagues also conditionally removed β-catenin in the mesenchyme or endoderm, showing its requirement for GT outgrowth and for *Shh* and *Fgf8* expression, thus reinforcing its importance in GT development (Miyagawa et al., 2009).

*Shh* expression occurs in the cloacal endoderm prior to GT formation, persisting throughout the urethral plate endoderm during GT outgrowth (Haraguchi et al., 2001; Perriton et al., 2002). The gene plays a crucial role in promoting cell proliferation and inhibiting apoptosis. *Shh*-null mice show severe ablation of GT development and suggest that the pathway lies upstream of *Fgf10, Fgf8*, *Bmp2*, *Bmp4*, and *Wnt5a*. Conditional deletion at a later stage of the sexually indifferent phase allows the GT to develop with less severe defects, confirming *Shh*'s role upstream of these pathways and also showing underexpression of *Hoxa13* and *Hoxd13* (Lin et al., 2009). β-catenin signaling is also severely disrupted in the endoderm of *Shh* mutants (Miyagawa et al., 2009). Manipulation of the hedgehog pathway receptor smoothened using

conditional Cre drivers specific to the mesenchyme or endoderm suggests that *Shh* signals largely via the GT mesenchyme (Lin et al., 2009). Disrupting smoothened in the ectoderm also shows a role of *Shh* signaling to this layer to promote urethral tube closure (Seifert et al., 2009b).

From these disruption experiments, the β-catenin/*Fgf8* DUE domain and the broader underlying urethral plate *Shh* domain appear to mutually reinforce one another, as with the interaction of the ZPA and AER in the limb. In contrast to the earlier experiments inhibiting FGF8 protein, genetic removal of *Fgf8* in the DUE does not disrupt GT outgrowth or gene expression, suggesting that it is dispensable (Miyagawa et al., 2009; Seifert et al., 2009a). Furthermore, while *Fgf8* transcript is detectible in the DUE of mouse, pig, and opossum, it is not seen in in the turtle or alligator (Seifert et al., 2009a). To address possible redundancy of Fgf ligands, Lin and colleagues inactivated the receptor genes *Fgfr1* and *Fgfr2* (Lin et al., 2013). Inactivation throughout the GT causes substantial underdevelopment and disruption of *Shh*, while removal from the urethral plate results in normal overall outgrowth.

Furthermore, this study found that overexpression of *Fgf8* in the urethral endoderm and in the AER causes overgrowth of the GT and the limb, respectively. Restoring *Fgf8* expression also rescues GT development from agenesis following β-catenin removal, and can play a similar role in partially rescuing the limb. By contrast, *Fgf8* is not capable of rescuing loss of *Shh* in either appendage. These results suggest that both appendages share a common β-catenin-*Fgf8* pathway and a *Shh* signaling center, and while these pathways reinforce each other, they have non-redundant functions. The TF-encoding gene *Sp8* also appears to play a common role in the DUE and AER, partially mediating the activation of *Fgf8* downstream of β-catenin.

While the mouse model has been enormously useful for probing the genetics of GT development (Haller and Ma, 2019; Hashimoto et al., 2019), the external genitalia are highly

diverse across amniotes. For example, non-mammalian amniotes do not undergo urethral tube closure, squamates develop paired hemipenes rather than a midline phallus, while the tuatara and most birds lack an intromittent phallus altogether following development (Gredler et al., 2014). This led to different hypotheses about the evolutionary origin of the phallus, including a common origin in the last common ancestor or separate evolution in mammals, squamates, and the clade containing turtles and archosaurs. In recent years, the accumulation of evidence based on early embryonic initiation of genital budding and gene expression suggests a common origin from the last common ancestor of amniotes over 300 million years ago (Gredler, 2016).

Genital initiation begins with paired genital swellings in all amniote lineages, including even the tuatara which lacks a phallus (Sanger et al., 2015). While these swellings remain separate and grow into paired hemipenes in squamates, in other groups they merge at the midline to form the GT. Furthermore, many genes show similar expression domains in different amniote clades, which is consistent with shared ancestral functions. In situ hybridization shows *Shh* expression in the urethral endoderm and *Hoxd13* broadly in the GT of turtles (Larkins and Cohn, 2015), alligators (Gredler et al., 2015a), and birds (Herrera et al., 2013). Birds like the chicken, however, later undergo phallic regression as a result of increased apoptosis. While *Hoxa13* and *Hoxd13* are both found in the hemipenes of snakes (Leal and Cohn, 2014), *Shh* is initially found only in the cloacal region of snakes and *Anolis* lizards (Gredler et al., 2015b). The sulcus of each hemipene is derived from the ectoderm and not from cloacal endoderm like the urethra of other amniotes, raising questions of how the *Shh* center regulates hemiphallus development.

In addition to comparisons of gene expression, analyses of enhancer activity can show overlapping functions between these appendages. Lonfat et al. (2014) identified digit and GT enhancers of the *Hoxa13* and *Hoxd13* regulatory landscapes using 4C and transgenic reporter

assays, finding both pleiotropic and tissue-specific enhancers. Guerreiro et al. (2016) tested the activity of some of these *Hoxd13* enhancer sequences from snake, lizard, and chick, finding both conservation and divergence of their expression domains. For example, both the mouse and snake version of the enhancer GT2 show genital-specific activity, consistent with a conserved function from their last common ancestor. By contrast, while the mouse Island-I enhancer shows digit-specific activity, the lizard sequence drives limb and genital activity, suggesting a possible gain of function. While the mouse Prox enhancer shows limb-GT activity, the snake version shows only GT activity, consistent with a loss of function following limb loss.

Infante and colleagues used H3K27ac ChIP-seq to systematically compare the enhancer landscapes of mouse limb and genital buds across the genome (Infante et al., 2015). By associating these regions with the nearest genes and performing gene ontology analysis, the active GT enhancers appear enriched near genes involved in limb development. Furthermore, there is extensive overlap between active limb and GT enhancers as compared to other tissues. These enhancers can thus be divided into limb-only, limb-GT, and GT-only groups. In addition to being enriched near limb genes, the limb-GT and GT-only enhancers are both enriched near genes expressed in mouse penis development. *De novo* motif analysis of the GT-only enhancers shows that one of the most enriched motifs is somewhat similar to the known ISL1 motif.

## *Isl1* in Genital Development

An *Isl1* expression domain occurs in the developing genitalia starting from early stages (Ching et al., 2018; Draaken et al., 2015; Kaku et al., 2013). In contrast to the transient expression in the limb, *Isl1* expression in the GT persists throughout the sexually indifferent phase of development. It appears prior to GT formation in the peri-cloacal mesenchyme and

diminishes around the onset of sexual differentiation (Ching et al., 2018; Draaken et al., 2015). Tschopp and colleagues used RNA-seq to find transcriptomic differences between the hindlimb and genitalia of mouse and *Anolis* lizard, finding genes with higher expression in either appendage (Tschopp et al., 2014). In mouse, this was based on hindlimb buds at E10.0 and E11.5 – shortly after *Isl1* is known to diminish (Itou et al., 2012) – and the GT at E11.5 and E13.5.

From these and comparable stages in the lizard, *Isl1* shows significantly greater expression in the genitalia than the hindlimb of both species, singling it out as a marker of genital development. These results also incidentally suggest that *Isl1* expression likely diminishes rapidly in lizard hindlimb development as in mouse, while persisting in the hemiphallus through the ambisexual phase. Furthermore, these researchers carried out experiments in the chick in which they grafted cloacal endoderm to the hindlimb bud or the tail. This results in ectopic budding and induces ISL1 expression at the site of the graft, suggesting that the *Shh*-expressing cloacal endoderm acts as a signaling center upstream of mesenchymal *Isl1*. The fact that *Isl1* expression occurs not only in the genitalia of mice, but also of chickens and lizards suggests that this is a conserved expression domain among amniotes.

Conditional deletion of *Isl1* using *Hoxa3*-IRESCre (Zhang et al., 2017) or with *Tbx4*-Cre (Ching et al., 2018) results in substantial hypoplasia of the GT. *Hoxb6*-Cre *Isl1* cKOs were analyzed for their kidney defects and also appeared to have severe GT reduction, although further examination of this phenotype was not discussed (Kaku et al., 2013). Ching and colleagues performed functional tests of *Isl1* in the mouse GT using a combination of cKOs, ChIP-seq, and in situ hybridization (Ching et al., 2018). The *Tbx4*-Cre allele ablates *Isl1* in mesenchymal cells. In addition to overall hypoplasia, these mutants show patterning defects including enlarged urethral openings and reduced preputial swellings. In situ hybridizations

suggest that *Fgf10, Wnt5a,* and *Bmp4* are underexpressed in the *Isl1* cKOs, and ChIP-seq shows that ISL1 binding events occur near these genes, suggesting they could be direct targets.

There is also a marked reduction of apoptosis in the GT mesenchyme of *Isl1* mutants, consistent with *Bmp4*'s role in apoptosis. This research did not, however, detect significant changes in the proportion of proliferating cells, although it is possible that subtle changes were not detected by this analysis. While the bulk of *Isl1* expression in the GT occurs in the more abundant mesenchymal cells, there is also an expression domain within the endodermally-derived urethral epithelial cells (Ching et al., 2018; Su et al., 2019). *Isl1* deletion in this layer using a tamoxifen-inducible Cre causes urethral hypoplasia (Su et al., 2019). Single-cell RNA-seq analysis of the GT at E14.5 clustered the mesenchyme into 10 groups, showing *Isl1* enriched in a cluster that also shows enrichment of *Ptch1* and *Hoxd13* (Armfield and Cohn, 2021). Epithelial cells were also subdivided into 4 clusters, with *Isl1* showing enrichment in the presumptive preputial gland epithelial cluster alongside *Wnt5a* and *Tbx4*.

Figures

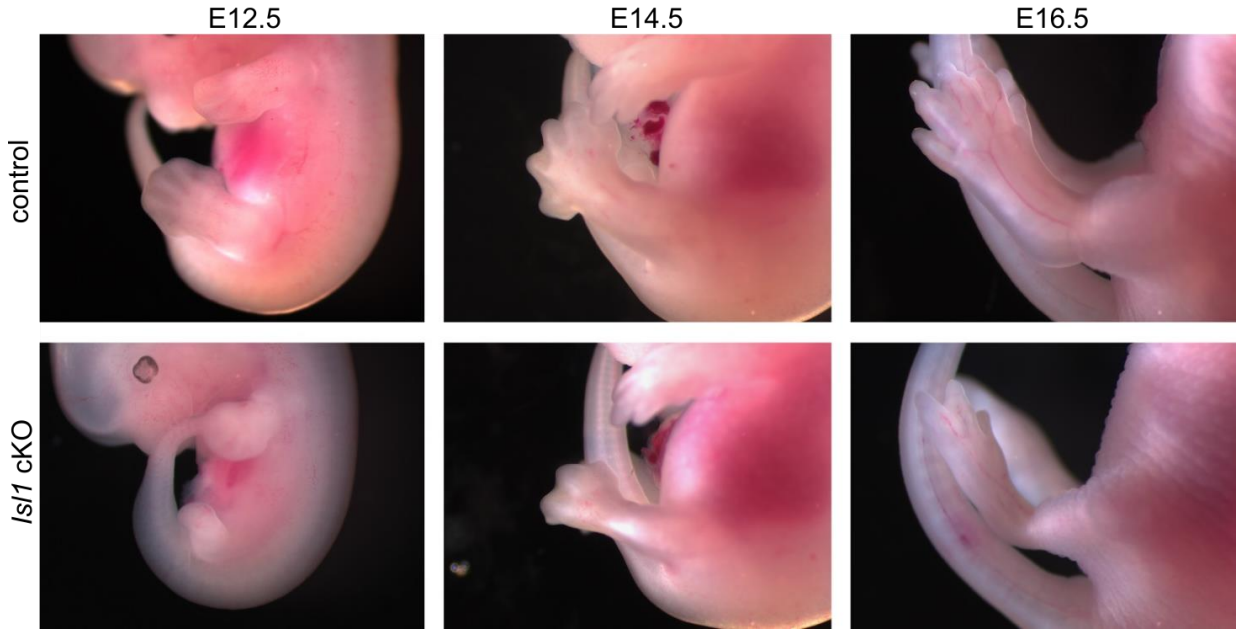


**Figure 1.1.** Hindlimb defects in *Isl1* conditional knockouts using *Hoxb6*-Cre. Embryos collected at three stages of development. Recapitulation of the phenotype described in greater detail by Itou et al. (2012) using the same knockout method.

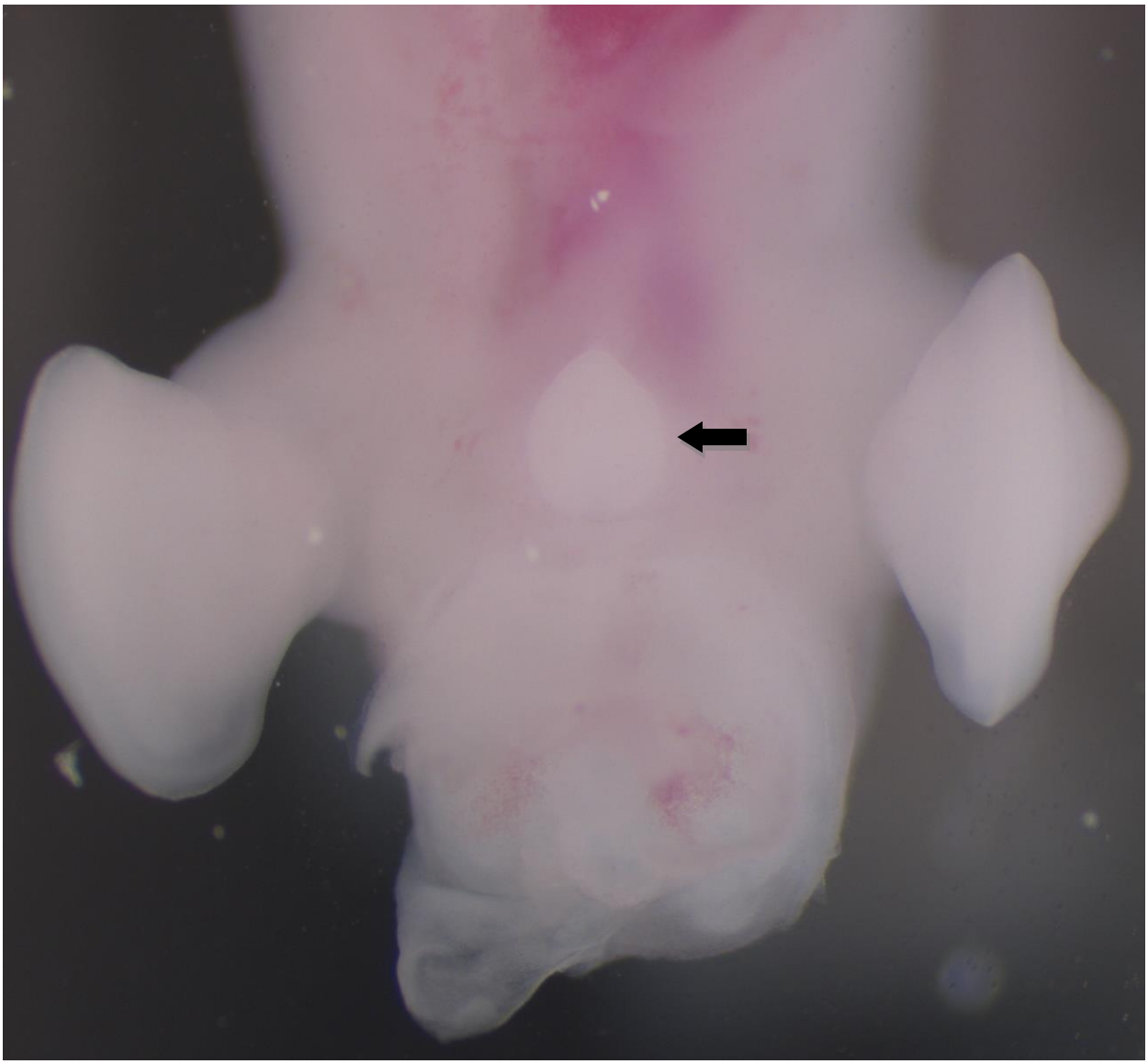

**Figure 1.2.** Genital tubercle of the American alligator prior to sexual dimorphism. Ventral view of the posterior portion of the embryo with the tail removed and the GT marked by the arrow.

# CHAPTER 2

# ISL1 DIRECTLY ACTIVATES KEY LIMB GENES DURING HINDLIMB INITIATION IN MOUSE AND CHICK [1]

[1] Minchey S.G., et al., to be submitted to *Development*.

Abstract

The initiation of limb outgrowth from the lateral plate mesoderm involves the interaction of various transcription factors and signaling molecules within a gene regulatory network. Some key aspects of this limb initiation network are shared between the forelimbs and hindlimbs, such as the establishment of an *Fgf10-Fgf8* feedback loop and later the *Shh*-expressing zone of polarizing activity. Other factors are restricted to the forelimbs or hindlimbs, such as *Tbx5* in the forelimb and *Tbx4*, *Pitx1*, and *Isl1* in the hindlimb. While *Tbx5* is required for forelimb initiation and *Fgf10* expression, *Isl1* is notable for being the only hindlimb specific factor with this strict requirement. To expand our understanding of its role in hindlimb development, we disrupted *Isl1* expression in the mouse hindlimb field and performed RNA-seq to find downstream genes with perturbed expression. We combined this with ChIP-seq in early mouse and chick hindlimb fields to find which genes are potential direct targets of ISL1. Our results not only confirm the previous findings that *Tbx4* is partially dependent on *Isl1*, but suggest this is due to direct binding at the enhancer HLEB in both mouse and chick. We also find potential conserved roles in the activation of *Bmp4*, *Hoxd10*, and other pathways with known functions in limb development. Some ISL1 targets also appear to be shared during its role in the development of the external genitalia.

## Introduction

The emergence of forelimb and hindlimb buds from the lateral plate mesoderm involves many common mechanisms (Royle et al., 2021). For example, *Fgf10* is capable of initiating budding in the embryonic chick flank (Ohuchi et al., 1997) and is required for formation of the forelimbs and hindlimbs in mice (Min et al., 1998; Sekine et al., 1999). Its roles include the induction of an epithelial-to-mesenchymal transition in the flank (Gros and Tabin, 2014) and the establishment of a positive feedback loop with *Fgf8* in the overlying apical ectodermal ridge (AER) (Ohuchi et al., 1997; Xu et al., 1998). Following initiation, the forelimbs and hindlimbs each rely on *Shh* in the posterior zone of polarizing activity (ZPA) to control both limb outgrowth and patterning of the anterior-posterior axis (Chiang et al., 1996; Riddle et al., 1993; Yang et al., 1997).

Differences between the mechanisms of forelimb and hindlimb development are apparent upstream of these signaling pathways. *Tbx5* encodes a forelimb-specific transcription factor which is required for forelimb bud initiation and *Fgf10* expression (Agarwal et al., 2003; Rallis et al., 2003). The paralogous *Tbx4* gene is expressed specifically in the hindlimb, though it is not strictly required for hindlimb initiation, and small limb buds with lower *Fgf10* expression still form in its absence (Naiche and Papaioannou, 2007). The hindlimb-specific factor *Pitx1* is also not required for hindlimb initiation, though it appears to activate *Tbx4* expression in both the mouse and chick (Lanctôt et al., 1999; Logan and Tabin, 1999; Szeto et al., 1999; Wang et al., 2018). A recent compound knockout of *Tbx4* and *Pitx1* resulted in the absence of hindlimbs and *Fgf10* expression, though did not have any effect on *Isl1* expression (Duboc et al., 2021).

*Isl1* is a LIM-Homeodomain transcription factor (TF) that is expressed transiently in the hindlimb but not the forelimb of both mouse and chick (Shiga et al., 2002; Yang et al., 2006).

Early conditional deletion of *Isl1* using a *T*-Cre allele has shown its requirement for hindlimb initiation and for normal *Tbx4* and *Fgf10* expression (Kawakami et al., 2011; Narkis et al., 2012). These mutants also show disruption of the Wnt pathway, with β-catenin failing to localize to the nucleus (Kawakami et al., 2011). Later deletion using a *Hoxb6*-Cre allele allows the hindlimb to develop, but disrupts the ZPA genes *Hand2* and *Shh*, causing a loss of posterior skeletal elements. *Isl1* also appears to be restricted to the posterior hindlimb by *Gata6*, which possibly represses its anterior expression via binding to the *cis*-regulatory elements CR1 and CR2 (Tahara et al., 2018).

Neither of these *Isl1* knockout models show diminished *Pitx1* expression, suggesting that *Isl1* and *Pitx1* act in parallel to activate *Tbx4*. Hindlimb expression of *Tbx4* in the mouse is regulated by the enhancers HLEA and HLEB, with HLEB showing high conservation across tetrapods (Menke et al., 2008). PITX1 hindlimb ChIP-seq shows binding to both HLEA and HLEB in the mouse and HLEB in the *Anolis* lizard (Infante et al., 2013; Wang et al., 2018). In the mouse, deletion of HLEB causes reduced *Tbx4* expression during the early hindlimb stages E9.5 and E10.5, while expression recovers at later stages (Infante et al., 2015). HLEA deletion affects transcription most dramatically from E10.5 to later stages.

This timing of HLEB's transcriptional impact aligns well with the early transient expression of *Isl1*, which in mouse shows sharply diminishing expression between E9.75 and E10.0 as the hindlimb bud emerges (Itou et al., 2012). Expression in the chick seems to persist somewhat later, with ISL1 protein showing clear expression at Hamburger-Hamilton (HH) stage 20, after hindlimb buds have formed (Shiga et al., 2002). To comprehensively find the downstream roles of *Isl1* in directing hindlimb initiation, we performed RNA-seq in the E9.75 hindlimb fields of *Hoxb6*-Cre conditional knockout (cKO) mice. We also performed ChIP-seq on

ISL1 in both the mouse and chick hindlimb fields/buds to find what enhancers and genes may be directly targeted by ISL1. We find evidence of conserved binding to the HLEB enhancer, as well as conserved putative enhancers of *Bmp4* and *Hoxd10*. Many ISL1 targets also appear to be shared during its role in the development of the genital tubercle (GT).

## Results

### *Isl1 cKOs show differential expression of limb development genes in the hindlimb field*

We used the *Hoxb6*-Cre allele (Lowe et al., 2000) to conditionally delete exon 2 of *Isl1* in the hindlimb field and performed RNA-seq to find differentially expressed genes (Fig. S2.1). We found 149 genes that were underexpressed in the cKOs and 77 that were overexpressed based on an FDR-adjusted p-value cutoff of 0.05 (Fig. 2.1, Table S2.1). We used DAVID to perform gene ontology analysis and found that the underexpressed set is enriched for genes known to be involved in limb development. Consistent with published in situ hybridization findings (Kawakami et al., 2011; Narkis et al., 2012), we see an underexpression of *Tbx4, Fgf10,* and *Fgf8,* while *Pitx1* and *Hoxc10* are not differentially expressed. Some underexpressed genes may be downstream consequences of disrupting the FGF feedback loop. This could include *Spry1/2/4,* which coincide with and negatively regulate FGF-signaling (Minowada et al., 1999), and *Etv4/5,* which are activated by FGF-signaling and repress *Shh* in the anterior limb bud (Lettice et al., 2012; Mao et al., 2009; Zhang et al., 2009).

Another underexpressed gene which was not previously reported in *Isl1* cKOs is *Bmp4,* which plays a role in establishing the AER by signaling through the BMP receptor IA and the TF SMAD4 (Ahn et al., 2001; Benazet and Zeller, 2013; Pajni-Underwood et al., 2007). We do not see underexpression of *Smad4,* although this gene is expressed broadly in the mesenchyme

(Bénazet et al., 2012) and thus we cannot disentangle its AER-specific expression levels. Our ability to detect reduced *Fgf8* expression, however, suggests sufficient sensitivity to detect some ectodermally-restricted genes. Examples of this include the ectodermally-expressed *Fgf20* (Hajihosseini and Heath, 2002), and *Dlx2*/*5* which are restricted to the AER (Ferrari et al., 1999; Robledo et al., 2002; Thomas et al., 1998).

Other underexpressed genes known to be involved in limb development include *Hoxd10*, *Cyp26b1*, *Wnt5a*, *Sall1/3*, and *Zbtb16* (also called *Plzf*). We do not find underexpression of *Hand2* at this stage in contrast to a prior in situ hybridization experiment suggesting *Hand2* was underexpressed as early as E9.75 (Itou et al., 2012). *Shh* transcripts were not detected in either the cKO or control embryos, and thus this stage may be too early to detect alterations in the ZPA. *Gli1* was also not differentially expressed at this stage, although we did detect underexpression of *Ptch1*. We find underexpression of *Ets*2, which was previously observed in *Hand2* mutants (Osterwalder et al., 2014). These ETS TFs are thought to activate *Shh* expression by binding to the ZRS enhancer (Kvon et al., 2016; Lettice et al., 2012).

In addition to the *Tbx4-Fgf* pathway and the *Hand2-Shh* pathway, another important aspect of *Isl1*'s requirement in the hindlimb field appears to be its role upstream of WNT/β-catenin signaling (Kawakami et al., 2011). We do not find differential expression of *Ctnnb1* itself, which encodes the β-catenin protein that lies downstream of *Isl1*. Although Wnt-signaling GO terms are not among the most enriched for the underexpressed genes (Fig. 2.1), there are several genes from this pathway that are underexpressed. This includes the WNT-ligand gene *Wnt5a*, the membrane receptor gene *Frzb*, and the TF genes *Lef1* and *Tcf7*.

As indicated by the top gene ontology terms (Fig. 2.1), WNT-pathway genes are also found among the overepxressed set, including the ligand genes *Wnt2* and *Wnt4*. In contrast to the

underexpressed FGF ligands, we see an overexpression of the FGF-receptor genes *Fgfr2* and *Fgfr3* and the membrane glycoprotein gene *Gas1* which positively regulates *Fgf10* expression (Liu et al., 2002). We also see overexpression of the TF genes *Irx3/5*, which repress the ZPA possibly by activating *Gli3* (Li et al., 2014), and of *Alx4* which similarly plays a role in anterior limb identity (Kuijper et al., 2005; Matsubara et al., 2017a). As with the underexpression of posterior factors, overexpression of these anterior genes was previously detected in *Hand2* mutants (Galli et al., 2010; Osterwalder et al., 2014).

*ISL1 binds near limb genes in the hindlimb field of mouse and chick*

We performed ChIP-seq on ISL1 in the mouse hindlimb field at E9.75 and in the hindlimb field/buds of the chick from HH18-20 using the same antibody. This resulted in 1,485 reproducible peaks in the mouse and 5,509 in the chick (Table S2.2). We also performed ChIP-seq on H3K27ac in the mouse hindlimb field at the same stage and retreived a published H3K27ac ChIP-seq dataset in the chick HH19 hindlimb (Yakushiji-Kaminatsui et al., 2018). We used HOMER (Heinz et al., 2010) to determine the top *de novo* motifs within the central 100 bp around ISL1 peak summits and used GREAT (McLean et al., 2010) to associate peaks with nearby genes and perform gene ontology analysis (Fig. 2.2 and 2.3).

In the mouse, we find that ISL1 binding sites tend to be flanked by H3K27ac (Fig. 2.2A) and to occur far from transcription start sites (Fig. 2.2C), which is characteristic of binding to active, distal enhancers. In the mouse E11.5 hindlimb (Infante et al., 2015) we find apprently reduced activity of these enhancers, suggesting a possible transient role during hindlimb initiation (Fig. 2.2A). In the chick we find a similar enhancer binding pattern (Fig. 2.3A,C), although these enhancers seem to exhibit a similar activity during later limb development. In

mouse, the top *de novo* motif resembles the known ISL1 motif (Fig. 2.2B). The top motif in the chick shows the highest match score with a different homeobox factor, though this motif appears very similar to the one found in mouse (Fig. 2.3B).

In both species we find enrichment of a HOX13-like motif (Fig. 2.2B and 2.3B), suggesting that ISL1 binding sites may occur near those of HOX factors binding at the same developmental stage or in other contexts. We also find similarities between the two species when looking at gene ontology enrichement terms for the mouse peaks (Fig. 2.2D) and the chick peaks that occur on orthologous loci converted to the mouse genome (Fig. 2.3D). Both suggest that ISL1 tends to occur nearest genes expressed during mouse limb development and genes whose human orthologs are implicated in limb disorders. In addition, we find a subset of 70 peaks that are conserved in both species, occuring on orthologous sequence and showing enrichment in both datasets. Depsite being a small sample, these conserved ISL1 binding loci show similar patterns in gene ontology and motif enrichment as the larger peak sets (Fig. S2.2).

*Directly activated targets of ISL1 in the hindlimb field*

Many of the differentially expressed genes are expected to be indirect downstream effects of *Isl1* deletion. We used Cistrome-GO (Li et al., 2019) to integrate the mouse differential expression data with the mouse ChIP-seq data and find what types of genes may be direct targets of ISL1. We find appendage and limb development terms highly enriched among the underexpressed genes, but not the overexpressed genes (Fig. 2.4A,B), similarly to what we found in the RNA-seq data alone. There is also a greater association between the ChIP-seq data and the underexpressed genes than with the overexpressed genes as measured using a precision-recall curve (Fig. 2.4C), consistent with ISL1 acting primarily as an activator.

The Cistrome-GO approach ranks all genes based on the composite strength of ChIP-seq signal and differential expression, but does not delineate a cutoff for genes that are considered putative targets. To produce a more discrete set of target genes, we used the same FDR p-value cutoff that was used for the RNA-seq analysis and associated each ChIP-seq peak with the nearest gene. We did this using both the mouse peaks and the chick peaks, separately for the underexpressed and overexpressed genes (Table S2.3). Using this approach, we identify 44 putative directly activated targets from the mouse ChIP-seq data, 61 from the chick data, and 25 that overlap between the two species. For putative directly repressed genes, we find 15 in mouse, 31 in chick, and 9 genes in common.

The directly activated target list contains notable genes in limb development, including *Tbx4*, *Bmp4*, *Prrx1*, *Wnt5a*, *Cyp26b1*, and *Mecom* (Fig. 2.7). *Prrx1* is considered a marker for the mesenchyme starting from limb initiation (Logan et al., 2002; Nohno et al., 1993), and *Prrx1*-null mice show shortened zeugopods (Martin et al., 1995). *Cyp26b1* encodes an enzyme that degrades retinoic acid (RA), and *Cyp26b1*-null mice show severly truncated limbs (Yashiro et al., 2004). *Wnt5a*-null mice show truncated limbs (Yamaguchi et al., 1999) and *Mecom*-null mice show complete absence of the hindlimb (Hoyt et al., 1997). The activation of *Tbx4* and *Bmp4* both provide possible pahways via which ISL1 can induce the mesenchymal-AER FGF feedback loop.

*ISL1 activates Tbx4, Bmp4, and Hoxd10 via conserved limb enhancers*

Most of the above putative direct targets are not bound via the same conserved enhancers in mouse and chick. Using only the 70 conserved ISL1 peaks, we find that one of these peaks occurs on the known limb enhancer of *Tbx4* HLEB, (Fig. 2.5). PITX1 is known to bind to HLEA

in the hindlimb, an enhancer that is not well-conserved across amniotes (Infante et al., 2013; Menke et al., 2008). We do not see evidence of ISL1 binding to HLEA, which is consistent with the finding that the mouse HLEB affects *Tbx4* expression more during early hindlimb development while HLEA exerts its influence later (Infante et al., 2015). Our H3K27ac data at E9.75 nevertheless suggests that HLEA is active during hindlimb initiation, and thus ISL1 appears to have a conserved function in activating *Tbx4* specifically via HLEB.

Another conserved peak occurs within a locus near *Bmp4* called CONS3.8 (Fig. 2.6) which shows both mesenchymal and ecterdomal activity in the forelimbs and hindlimbs as well as orofacial activity (Jumlongras et al., 2012). Jumlongras and colleagues found that this locus is conserved between human and mouse, found a potential MSX motif, and showed that MSX2 can bind to the enhancer in vitro. They also found that mutation of this motif inhibits the ability for MSX2 to bind to the enhancer and results in a loss of its limb activity. Interestingly, this motif also matches the canonical ISL1 motif and is conserved not just across mammals, but other tetrapods as well (Fig. S2.3A,B).

Distal enhancers frequently do not associate with the nearest gene in the linear genome, thus posing a problem for ChIP-seq peak-to-gene associations. 5C analyses of human cells lines found that fewer than half of distal elements associate with the nearest TSS even when using only actively transcribed genes (Sanyal et al., 2012). An example in limbs is seen in the *Hoxd* cluster which is regulated during early budding by the telomeric domain (T-DOM) in both mouse and chick (Andrey et al., 2013; Yakushiji-Kaminatsui et al., 2018). This domain is divided into two sub-TADs by a chromatin boundary marked by the conserved elements CS38-40. CS39 drives expression in the mouse forelimb and hindlimb fields, and deletion or inversion of the CS38-40 region reduces *Hoxd11* expression in the forelimb (Rodríguez-Carballo et al., 2020).

We find conserved ISL1 binding at CS38 as well as binding at CS39 in the chick during hindlimb initiation (Fig. S2.4). This could be related to the underexpression of *Hoxd10,* and possibly also of *Hoxd1* which is within the T-DOM boundary. This region and others within this TAD appear active at this stage based on H3K27ac, though this activity diminishes in the later hindlimb buds. As with the *Bmp4* enhancer, we find a conserved ISL1 motif within the CS38 peak (Fig. S2.3C). Furthermore, the conserved CS38 peak falls near a lincRNA in mouse which is one of four underexpressed lncRNAs identified in a separate analysis of our RNA-seq data, known as ENSMUSG00000087333.1 or *Gm13652* (Table S2.1). In the chick this peak overlaps with another lncRNA known as LOC107053812. It is thus possible that ISL1 has a conserved function in activating *Hoxd10* and/or uncharacterized lncRNAs in the tetrapod hindlimb.

*Shared targets of ISL1 in the hindlimbs and the genital tubercle*

The genital tubercle (GT) is known to express many of the same genes involved in limb development (Cohn, 2011) and also shows widespread sharing of enhancers (Infante et al., 2015). For example *Hox13* genes are involved in both digit and GT development, and certain enhancers show activity in both appendages (Amândio et al., 2020; Lonfat et al., 2014). We performed an analysis of ISL1 in the GT, finding that that it binds to many known limb enhancers (Chapter 3). Consistent with expectations, we find many shared ISL1 binding loci between the hindlimb field and GT in both mouse and chick (Fig. S2.6). In the mouse, this represents over a third of the hindlimb field peaks, while in chick it is over half, although far more GT-specific peaks were identified in both species.

Common target genes between the two tissues could be activated via these shares binding loci or through different loci. In the mouse, we find 22 genes that are putative directly activated

targets in the hindlimb field and GT, representing 50% of target genes in the former and 5.7% of target genes in the latter (Table S2.4). 7 genes are putative directly repressed targets in the mouse hindlimb field and GT, comprising 47% of targets in the former and 2.7% in the latter. 16 of the 22 shared activated targets are associated with shared peaks, and 4 of 7 repressed targets.

We also find many shared target genes by combining the chick ChIP-seq data with the mouse differentially expressed genes. 28 directly activated targets are shared, representing 45.9% of hindlimb targets and 6.6% of GT targets, and 23 of these are associated with shared peaks. 14 directly repressed genes are shared, comprising 45.2% of hindlimb targets and 4.2% of GT targets, with 12 of these associated with shared peaks. We find 5 genes that are underexpressed in the hindlimb field and GT and associated with peaks that are both conserved and shared – *Unc5c*, *Mab21l1*, *Evx1*, *Tbx4*, and *Bmp4*. Of these genes, *Tbx4* and *Bmp4* are both bound via known limb enhancers that may have been coopted during the evolution of the amniote phallus. The only putative directly repressed gene in both the hindlimb field and GT associated with a shared, conserved peak is *Rarb*.

## Discussion

Prior *Isl1* knockout and in situ hybridization experiments revealed essential roles for *Isl1* during hindlimb initiation, which our work expands upon using high-throughput experiments. By analyzing overall transcriptomic changes in *Isl1* cKOs, our results provide a top-down approach for identifying potential downstream target genes. Our ChIP-seq data then allows us to identify which of these genes may be direct targets, identify enhancers of interest, and find targets that may be conserved across amniotes. Our work both confirms and extends upon our previous understanding of *Isl1*'s role upstream of *Tbx4* and *Fgf10/8*. Consistent with previous findings,

*Tbx4* but not *Pitx1* is underexpressed in *Isl1* cKOs, and our ChIP-seq data suggest the former is controlled through ISL1 binding to HLEB.

This finding supports a model in which ISL1 activates *Tbx4* early in its expression when HLEB is most active, whereas HLEA remains active later in hindlimb development and continues to be bound by PITX1 in mouse (Infante et al., 2013, 2015). HLEA is mostly not conserved outside of mammals, so it is unclear whether this temporal pattern holds in other amniotes. Parallel research on ISL1 in the external genitalia finds conservation of three ISL1 motifs within HLEB across amniotes, except for snakes in which these motifs have undergone progressive decay (Chapter 3, Fig S3.5). The king cobra version of HLEB is no longer able to drive hindlimb expression in transgenic mouse assays, although restoration of multiple mutated bases rescued this domain of hindlimb activity (Fig. 3.6).

Earlier deletion of *Isl1* prior to hindlimb initiation shows that *Tbx4* is reduced though not eliminated, while *Fgf10* expression is more thoroughly ablated (Kawakami et al., 2011; Narkis et al., 2012), suggesting that ISL1 may influence *Fgf10* via other pathways. As in these early *Isl1* cKOs, Duboc et al. (2021) found that compound deletion of *Tbx4* and *Pitx1* causes complete disruption of hindlimb initiation and *Fgf10* expression. These mutants still show normal *Isl1* expression, leading to a model in which *Isl1* and *Tbx4+Pitx1* act in parallel to activate *Fgf10*. While our data strongly indicate that ISL1 is directly upstream of *Tbx4*, it is less clear whether it directly activates *Fgf10*. The chicken ChIP-seq data identify *Fgf10* as a putative direct target gene, while the mouse data do not (Table S2.3). Our mouse ChIP-seq reveals far fewer peaks than in the chick, and weaker signal likely limits our ability to find true binding loci.

Separately, ISL1 may indirectly activate *Fgf10* through the additive effects of multiple target genes and pathways that have not been thoroughly characterized. The WNT/β-catenin

pathway is known to be essential for hindlimb initiation and *Fgf10* expression, and *Isl1* is required for nuclear localization of β-catenin (Kawakami et al., 2011). In the chick, *Wnt2b* in the forelimb and *Wnt8c* in the hindlimb appear to act via β-catenin to induce limb outgrowth and *Fgf10* expression, though this is not found in the mouse (Agarwal et al., 2003; Kawakami et al., 2001). We do not find ISL1 binding near *Wnt8c* (called *Wnt8a* in the galGal6 annotation) during chick hindlimb initiation, though we do find evidence that ISL1 activates *Wnt5a*. Limb outgrowth is inhibited in *Wnt5a* mutants, although limbs still develop with normal AER *Fgf8* expression until E12.5 (Gao et al., 2018; Yamaguchi et al., 1999). This implies that *Wnt5a* is not required for limb initiation or establishment of the *Fgf10-Fgf8* feedback loop.

In addition, *Wnt5a* is commonly though to act via non-canonical WNT pathways (Geetha-Loganathan et al., 2008). During limb outgrowth *Wnt5a* acts via the JNK/planar cell polarity pathway to promote mesenchymal cell elongation and division along the proximo-distal axis (Gros et al., 2010). During digit development, *Wnt5a* is even known to promote β-catenin degradation and inhibit the canonical pathway (Topol et al., 2003), as well as acting via the $Ca^{2+}$ pathway (Farrera-Hernández et al., 2021). We also find evidence that *Lef1* and *Tcf7* are downstream of ISL1 and could be direct targets in mouse based on broader association rules. The TFs encoded by these genes affect canonical signaling by forming complexes with β-catenin, and deletion of both *Lef1* and *Tcf7* results in smaller limb buds (Galceran et al., 1999), though they do not show the complete ablation found in *Ctnnb1* knockouts (Kawakami et al., 2011). It thus remains unclear from our analyses how ISL1 plays its indespensible role in activating nuclear β-catenin translocation.

We find evidence of direct activation of *Bmp4*, possibly via a known limb enhancer conserved across tetrapods. Early signaling through the BMP receptor IA in the ectoderm is

essential for hindlimb outgrowth, AER formation, and maintenance of *Fgf10* expression (Ahn et al., 2001; Pajni-Underwood et al., 2007). This provides another potential pathway via which ISL1 can impact AER *Fgf8* expression and maintenance of the FGF feedback loop. *Bmp7* encodes another BMP ligand that is expressed during early limb development, thus possibly compensating for loss of *Bmp4* (Ahn et al., 2001; Robert, 2007). Accordingly, conditional deletion of *Bmp4* alone using an *Isl1*-Cre does not result in hindlimb agenesis or loss of *Fgf8,* though loss of midline/pelvic tissue results in a hindlimb fusion phenotype (Suzuki et al., 2012).

Another essential role of *Isl1* is its establishment of ZPA *Hand2-Shh* expression. While Itou et al. (2012) found a dramatic loss of *Hand2* by E10.0 using the same *Isl1* cKO as in our study, we do not find significant underexpression at E9.75 in our RNA-seq data. Using our ChIP-seq data from chick however, we find ISL1 binding at putative *Hand2* enhancers, including near a tested limb enhancer mm1689, though we don't see this signal in the mouse (Fig. S2.5). Another uncharacterized putative enhancer shows possible ISL1 binding in both species, although ISL1 enrichment at this locus was not consistent enough between replicates to pass the reproducibility threshold and thus enter our peak set. This locus contains an ISL1 motif conserved across tetrapods (Fig. S2.3D). Thus while *Hand2* is not a candidate for a direct target based on our approach, it is possible that ISL1 exerts a delayed effect on *Hand2* expression, and limitations of our mouse ChIP-seq data prevent us from detecting true binding loci.

*Hoxd10* expression can upregulate *Shh* in the chick limb bud (Matsubara et al., 2017a) and activate the mouse ZRS enhancer via sites deleted in python (Leal and Cohn, 2016). ISL1 potentially activates *Hoxd10* via a conserved enhancer, along with a lincRNA of uncharacterized function. This may contribube to the establishment of the ZPA, though paralogs like *Hoxa/c10* and *Hoxd11* remain normally expressed, and Hox genes are known to be highly redundant

(Wellik and Capecchi, 2003). ISL1 may also repress genes that inhibit the ZPA, such as *Alx4* and *Irx3*, a finding consistent with Itou et al. (2012). Another component of the ZPA that is disrupted are the ETS factors *Ets2* and *Etv4/5*, which regulate the boundaries of the ZPA.

Our results also suggest a possible role of ISL1 in repressing the retinoic acid (RA) pathway, which may be shared in the GT. RA secreted from the somites is essential for hindlimb initiation and *Tbx4/Fgf10* expression (Nishimoto et al., 2015). *Cyp26b1* encodes an enzyme that degrades RA, while *Rarb* encodes a retinoic acid receptor which binds RA, enters the nucleus and regulates gene expression (Berenguer and Duester, 2021). We find both that *Cyp26b1* is a putative directly activated target, while *Rarb* is a directly repressed target, which would thus be expected to reduce RA activity. The effect of RA signaling on posterior mesenchymal *Isl1* expression has not been tested, and it is unknown what role RA plays during outgrowth of the external genitalia.

GDF11 signaling also emanates from the posterior paraxial mesoderm, and is known to activate *Isl1* expression in the hindlimb field and cloacal region (Jurberg et al., 2013; Matsubara et al., 2017b). This suggests that *Isl1* is responsive to paraxial signals in both contexts, though it is unclear why *Isl1* expression persists in the external genitalia while rapidly clearing from the hindlimb bud. A crucial difference may be the link between the GT and the *Shh* expressing cloacal endoderm, which appears to be upstream of *Isl1* unlike the ZPA, and may maintain its expression following GT formation (Tschopp et al., 2014). While we find many common target genes between the hindlimb field and GT, we find far more in the latter, and this may be partly due to the complexity of the GT at this more advanced stage (Armfield and Cohn, 2021). In the limbs, the transcriptome and *cis*-regulatory landscapes vary substantially across different stages (Andrey et al., 2017; Jhanwar et al., 2021). The role of *Isl1* in the cloacal region prior to genital

budding is unexplored, and it is conceivable that additional commonalities between hindlimb and genital initiation might be found during this more primordial state.

## Methods

### *Mouse lines*

*Isl1* cKO mice were generated by crossing *Hoxb6*-Cre mice (Lowe et al., 2000) with *Isl1*+/- mice, followed by crossing *Hoxb6*-Cre; *Isl1*+/- mice with *Isl1*flox/flox on a C57BL/6 background. Embryos with the *Hoxb6*-Cre; *Isl1*-/flox genotype were used as the conditional knockouts, while *Hoxb6*-Cre; *Isl1*+/flox embryos were used as controls.

### *RNA-seq*

Dissections were performed in the late afternoon of the ninth day following observation of a vaginal plug to target E9.75 in the C57BL/6 mice, which tend towards a more delayed development. Embryos that appeared too early (E9.5) or too late (E10.0) on the basis of hindlimb field development and overall embryo size were excluded. Left and right hindlimb fields from the appropriate embryos were carefully dissection and pipetted together using a wide orifice pipette tip directly into mirVana lysis buffer. This was then stored at -80C until cKO and control individuals were determined by genotyping using DNA from the remaining anterior portion of each embryo.

We collected total RNA separately from 6 female mutant and 6 female control mice using the mirVana miRNA Isolation Kit (ThermoFisher Scientific). mRNA libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina), generating 6 biological replicates from each genotype. These were then sequenced at the Georgia Genomics and Bioinformatics

Core to generate 50 bp single-end reads. Reads were aligned to the mm10 genome using STAR (v. 2.7.3a) and transcripts were counted using HTSeq (v. 0.13.5). We used mm10 refGene transcripts for the main analysis, and separately used the GENCODE annotation of lncRNAs. Differentially expressed genes were determined using DESeq2 (Galaxy v. 2.11.40.6).

*Chromatin immunoprecipitation*

Left and right hindlimb fields were dissected from ICR mice in the early afternoon to target E9.75, which tended to occur earlier in these mice. Chicken (Cobb500 breed) hindlimb fields/buds were dissected from Hamburger-Hamilton stage 18 to 20 (Hamburger and Hamilton, 1992). Samples were crosslinked in 1% formaldehyde for 20 minutes at room temperature and frozen at -80C for later use. These were later suspended in lysis buffer and briefly homogenized using a probe sonifier (Branson 450). They were then placed in a bath sonifier (Diagenode Bioruptor 300) and sheared for a total of 45 cycles of 30 seconds on/30 seconds off. Each ChIP experiment was performed in duplicate, except for the mouse ISL1 ChIP in which we performed 3 replicates and later combined reads from two of the replicates before performing IDR analysis.

The chick samples were processed with an older ChIP protocol which used PureProteome Protein G magnetic beads (Millipore) incubated with the antibody and ~500 μg of chromatin for ISL1 ChIP as measured by Nanodrop. The mouse experiments were performed using a modified protocol using Protein G Agarose Columns (Active Motif), targeting at least ~50 μg of chromatin for ISL1 ChIP, and ~25 μg for H3K27ac ChIP. For ISL1 ChIP we used 5 μg of anti-ISL1 rabbit monoclonal antibody (Abcam EP4182). For H3K27ac ChIP we used 3 μg of mouse monoclonal antibody CMA309 (previously from Millipore, subsequently from Cosmo Bio).

Libraries were prepared using the NEBNext Ultra II Library Prep Kit. Library amplification was done for 15 cycles. One-sided size selection was performed to remove primer and adapter fragments using Sera-Mag SpeedBeads. Libraries were sequenced by the Georgia Genomics and Bioinformatics Core (GGBC) to generate single-end 75 bp reads, except for one replicate of the mouse H3K27ac experiment which were sequenced by Genewiz to generate paired-end 150 bp reads.

*ChIP-seq data analysis*

The GGBC provided pre-trimmed reads with adapter sequence removed. The one mouse H3K27ac paired-end replicate was trimmed using cutadapt using the settings –O 1, -m 25 so that even one base of adapter overhang was trimmed. Adapter sequences used were as follows:

AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC

AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

We used the AQUAS Transcription Factor and Histone ChIP-seq processing pipeline starting from trimmed reads (https://github.com/kundajelab/chipseq_pipeline). We used the pipeline's default bwa aligner to align mouse data to GRCM38/mm10 (UCSC) and chick to galGal6 (UCSC). We set the fragment size to 175 (-speak_spp 175 –extsize_macs2 175). For the mouse H3K27ac ChIP, we used the default histone ChIP setting, which calls peaks using MACS2 and performs naïve overlap between the two replicates. The same IDR pipeline was used to analyze the ISL1 ChIP-seq data from the mouse and chick GT. We also performed ISL1 ChIP-seq in the E10.5 mouse forelimb and generated IDR peaks as a control experiment to filter out potential false positives from the other datasets.

For both ISL1 and H3K27ac ChIP, the pipeline uses MACS2 to generate bigWig plots displaying fold-enrichment over the input. The pooled replicate display of these fold-enrichment

plots were used in our figures. We used LiftOver on the UCSC genome browser to convert mouse bed coordinates to the chicken genome using default settings. UCSC Table Browser was then used to find overlapping peaks between mouse and chick.

*Public data used*

Mouse H3K27ac data from E11.5 hindlimbs performed by Infante et al. (2015) under accession GSE64055 was reprocessed in mm10 using fastq files and the above ChIP-seq pipeline. We similarly re-analyzed the raw mouse ISL1 ChIP-seq data from cardiac cells (Wang et al., 2016) under accession GSE79701 and from pancreatic β cells (Ediger et al., 2017) under accession GSE84759. One replicate of chick HH19 hindlimb H3K27ac ChIP-seq data (Yakushiji-Kaminatsui et al., 2018) was taken from accession GSE115563. We used Bowtie2 v. 3.5.1 to map these reads to galGal6, the callpeak function in MACS2 (v. 2.7.1) to call peaks, and the bdgcmp function to generate a fold enrichment plot. Mouse hindlimb PITX1 ChIP-seq data was processed in mm10 by Nemec et al., (2017) and thus processed data was used directly from accession GSE100734.

*Motif analysis*

We used HOMER v4.11 to perform *de novo* motif analysis using the central 100 bp around ISL1 peaks summits, using 100k random genomic sequences as the background (-N 100000), searching for motifs 8 bp in length (-len 8). For the conserved archosaur peaks, we used the summit coordinates from the chicken peaks and performed the motif analysis in that genome.

*Gene ontology analyses*

For RNA-seq data, we used DAVID (v. 6.8). Underexpressed or overexpressed genes were identified with an FDR-adjusted p-value cutoff of 0.05. For ChIP-seq data, mm10 peak coordinates were used as input to GREAT (v. 3.0.0) (http://great.stanford.edu/great/public-3.0.0/html/). We used the default "basal plus extension" association rule, which assigns each gene a regulator domain up to 1000 kb unless it reaches another gene's proximal domain (5 kb upstream, 1 kb downstream). For the Cistrome-GO analysis, we used the mm10 genome and input the mouse ISL1 ChIP-seq peaks as well as the DESeq2 output file with a half-decay distance of 50 kB. For the precision-recall curve, we set DE gene cutoffs at 0 for logFoldChange and 0.05 for FDR.

*Assignment of putative direct target genes*

For mouse, we used the same mm10 TSSs used by GREAT (https://great-help.atlassian.net/wiki/spaces/GREAT/pages/655445/Genes). For chick, we extracted TSSs from the longest transcript isoform of each coding gene in the galGal6 NCBI RefSeq annotation. To reduce false negatives based on naming differences between species, we used gProfiler orth (Galaxy v. 0.1.7) to find synonymous gene names. We then used the bedtools ClosestBed function (Galaxy v. 2.30.0) to assign each mouse or chick ISL1 peak to the nearest TSS in the respective annotations. These genes were then filtered using the mouse RNA-seq results to find genes that are both differentially expressed and have at least one associated peak.

*Heatmap visualization*

For the RNA-seq data, we used the normalized counts output by DESeq2 for differentially expressed genes, sorted the underexpressed genes alphabetically on top and the overexpressed genes below. We used the Galaxy tool heatmap2 (v3.0.1.0.0), scaling the data by row because of extreme differences in expression levels between genes. For ChIP-seq, we used the bed files centered on peak summits and the bigWig signal files as inputs for computeMatrix (Galaxy v. 3.3.2.0.0), using the scale-regions setting with windows of 1 kb and the bin size set to 1 bp. We used plotHeatmap (Galaxy v. 3.3.2.0.1) to generate the heatmaps from the matrix files.

Figures

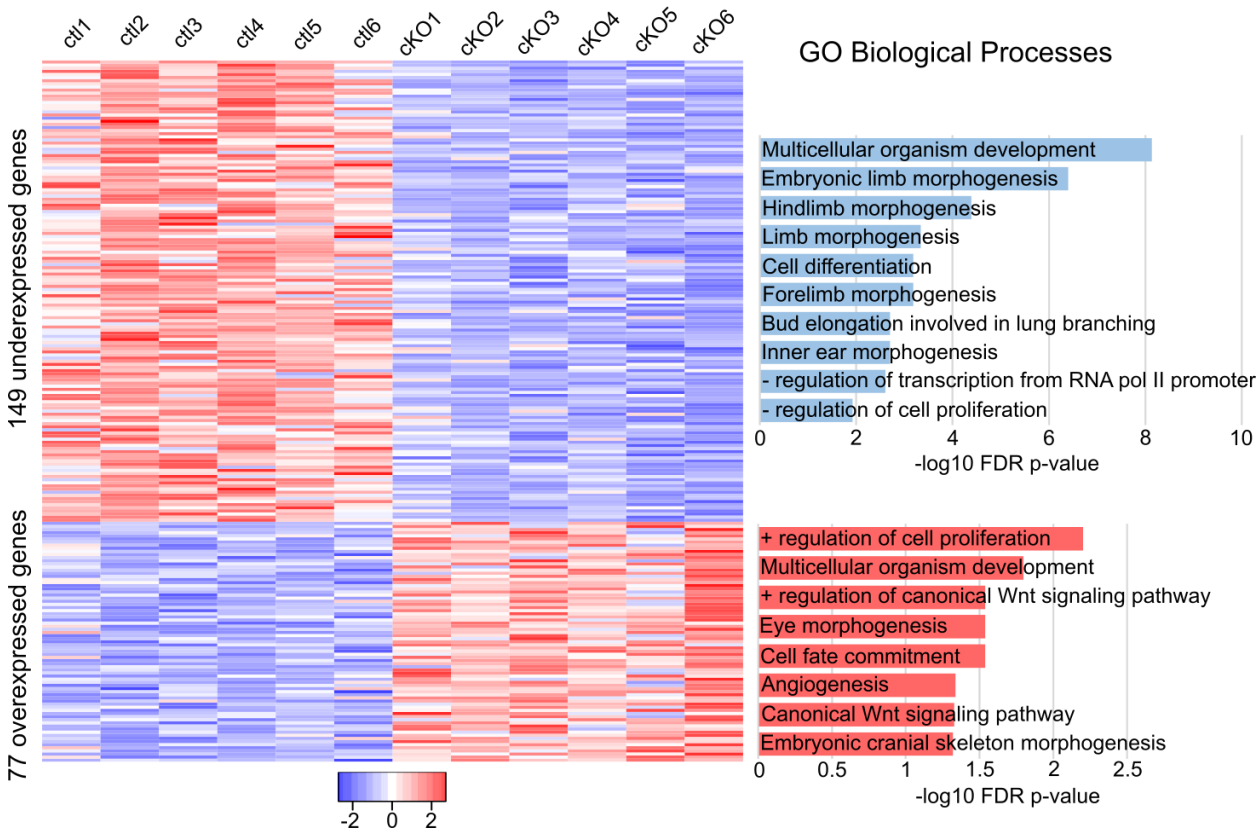


**Figure 2.1.** Differentially expressed genes in the mouse hindlimb field. Comparison of *Isl1* conditional knockout (cKO) and control (ctl) mice. DAVID gene ontology analysis shows the top GO Biological Processes categories enriched in the 149 underexpressed genes (top) and the 77 overexpressed genes (bottom).

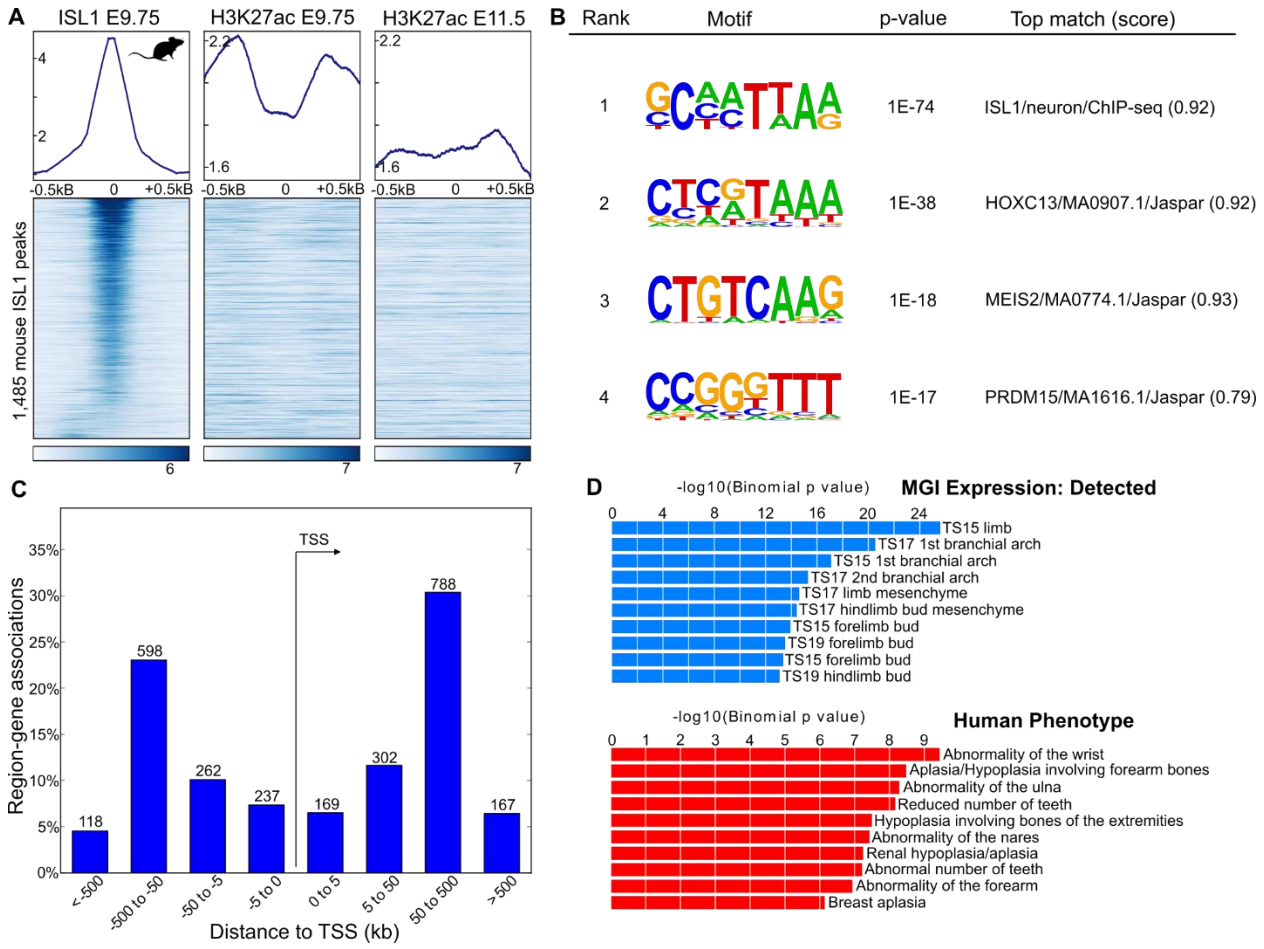


**Figure 2.2.** ISL1 ChIP-seq in the mouse hindlimb field. **(A)** ISL1 and H3K27ac signal in the mouse hindlimb field, as well as E11.5 hindlimb buds (Infante et al., 2015), centered on ISL1 peak summits. **(B)** *De novo* motif analysis showing the top motifs enriched in the central 100 bp around peak summits. **(C)** Distribution of ISL1 peaks relative to the nearest associated genes based on GREAT. **(D)** GO analysis showing top 10 terms based on mouse embryonic gene expression (top) and on human phenotypes (bottom).

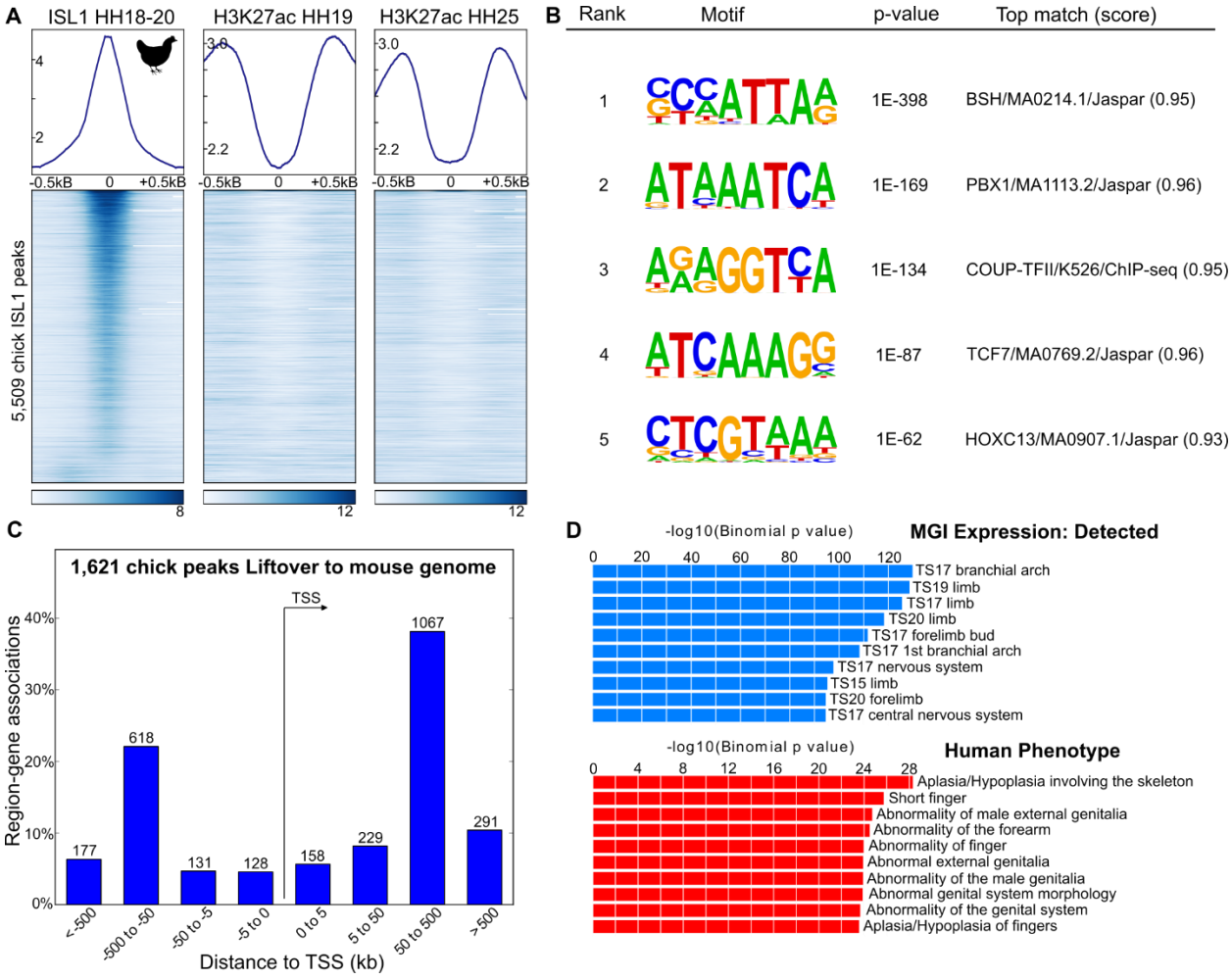


**Figure 2.3.** ISL1 ChIP-seq in the chick hindlimb field. **(A)** ISL1 and H3K27ac (Yakushiji-Kaminatsui et al., 2018) signal in the chick hindlimb field, as well as HH25 hindlimb buds, centered on ISL1 peak summits. **(B)** *De novo* motif analysis showing the top motifs enriched in the central 100 bp around peak summits. **(C)** Distribution of ISL1 peaks converted to the mouse genome relative to the nearest transcription start sites (TSS) based on GREAT. **(D)** GO analysis showing top 10 terms based on mouse embryonic gene expression (top) and on human phenotypes (bottom).

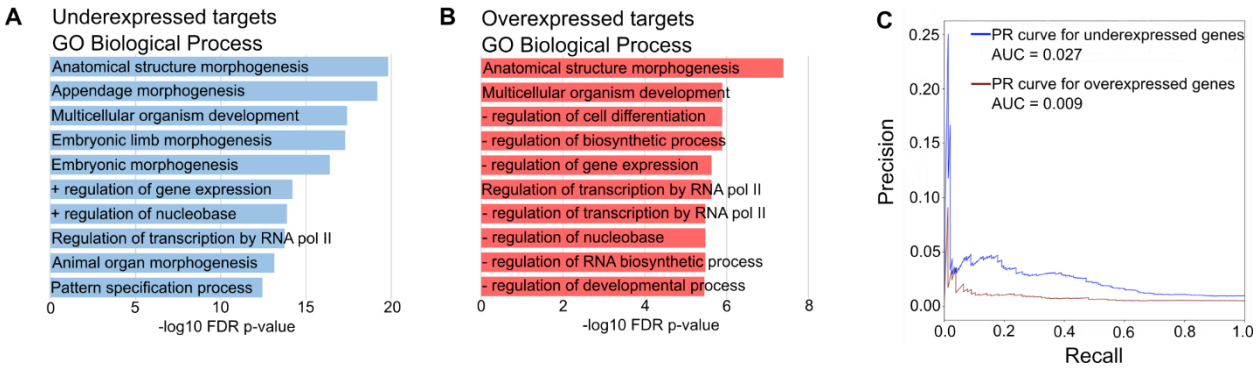


**Figure 2.4.** Putative direct targets of ISL1 in the mouse hindlimb field based on Cistrome-GO. **(A)** Top 10 GO Biological Process terms for underexpressed putative direct targets. **(B)** Top 10 GO Biological Process terms for overexpressed putative direct targets. **(C)** Precision-recall curves showing the association between the ChIP-seq data and underexpressed genes (blue) vs overexpressed genes (red).

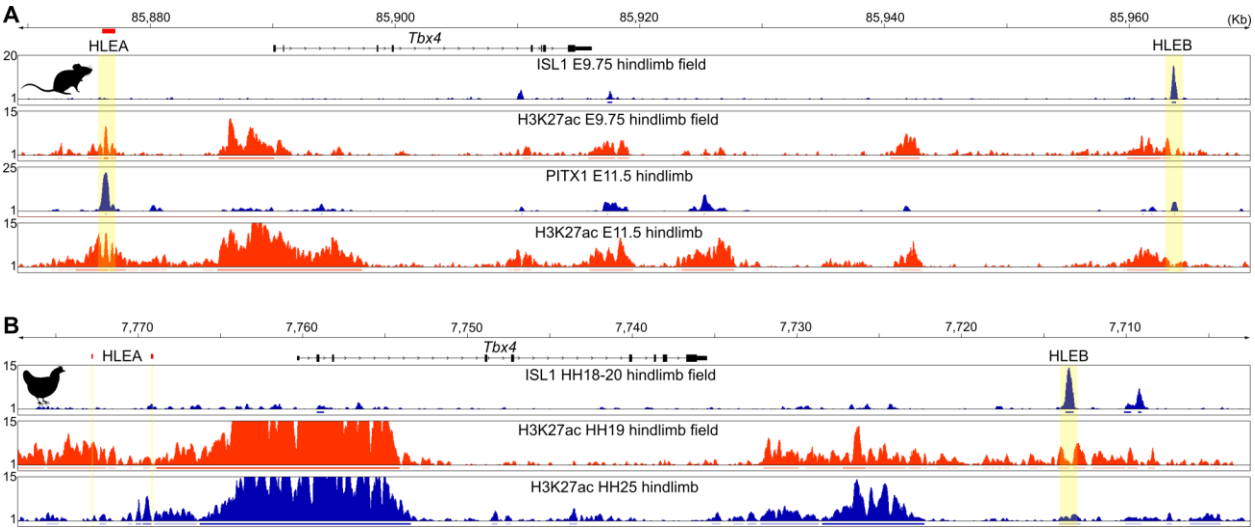


**Figure 2.5.** ISL1 binding at the hindlimb *Tbx4* enhancer HLEB in mouse and chick. **(A)** Mouse ISL1 and H3K27ac signal in the hindlimb field, as well as PITX1 (Nemec et al., 2017) and H3K27ac signal in the later hindlimb bud. mm10 coordinates chr11:85,869,165-85,970,165. **(B)** Chick ISL1 and H3K27ac signal in the hindlimb field, as well as H3K27ac in the later hindlimb bud. galGal6 coordinates chr19:7,702,496-7,777,496.

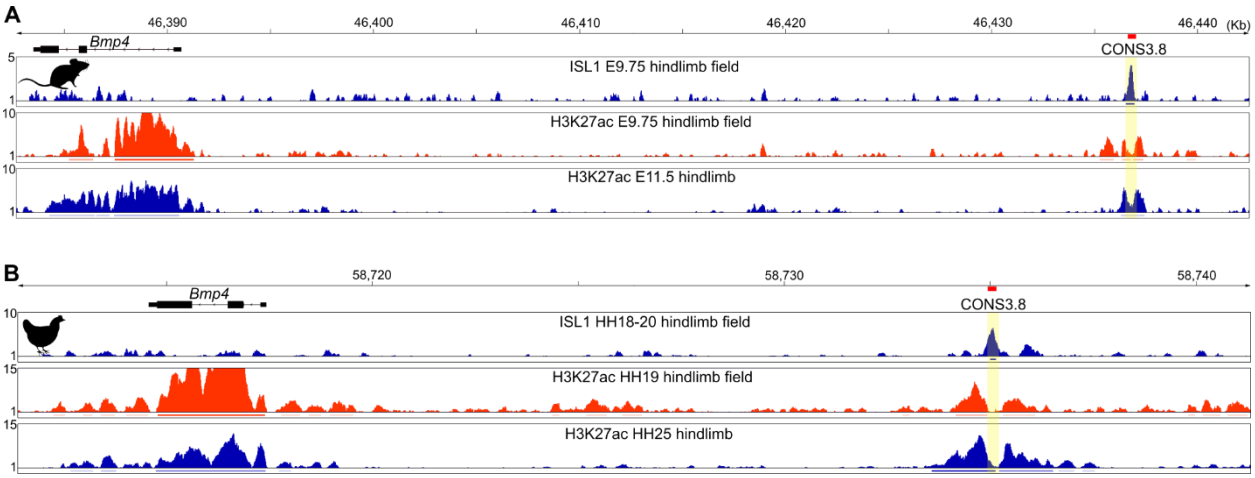


**Figure 2.6.** ISL1 binding at the limb *Bmp4* enhancer CONS3.8 in mouse and chick. **(A)** Mouse ISL1 and H3K27ac signal in the hindlimb field, as well H3K27ac signal in the later hindlimb bud. mm10 coordinates chr14:46,382,664-46,442,664. **(B)** Chick ISL1 and H3K27ac signal in the hindlimb field, as well as H3K27ac in the later hindlimb bud. galGal6 coordinates chr5:58,711,396-58,741,396.

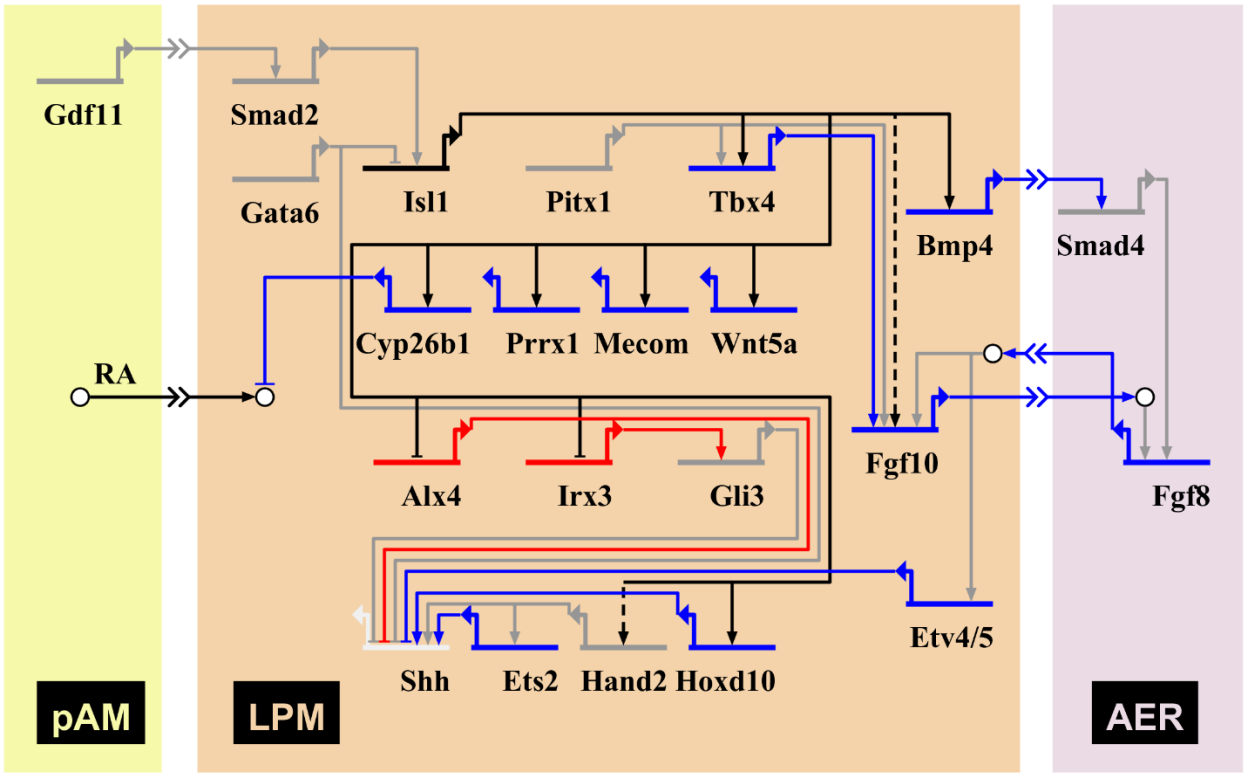


**Figure 2.7.** *Isl1*'s role in a proposed gene regulatory network for hindlimb initiation. Blue genes are underexpressed in *Isl1* cKOs, while red genes are overexpressed. Grey genes are not differentially expressed, while *Shh* transcripts are not detected at this stage. Putative direct targets of ISL1 based on ChIP-seq data in mouse and chick are represented with solid lines, while dashed lines are used for *Fgf10* and *Hand2* based on peaks found specifically in chick. pAM: posterior/paraxial mesoderm; LPM: lateral plate mesoderm; AER: apical ectodermal ridge.

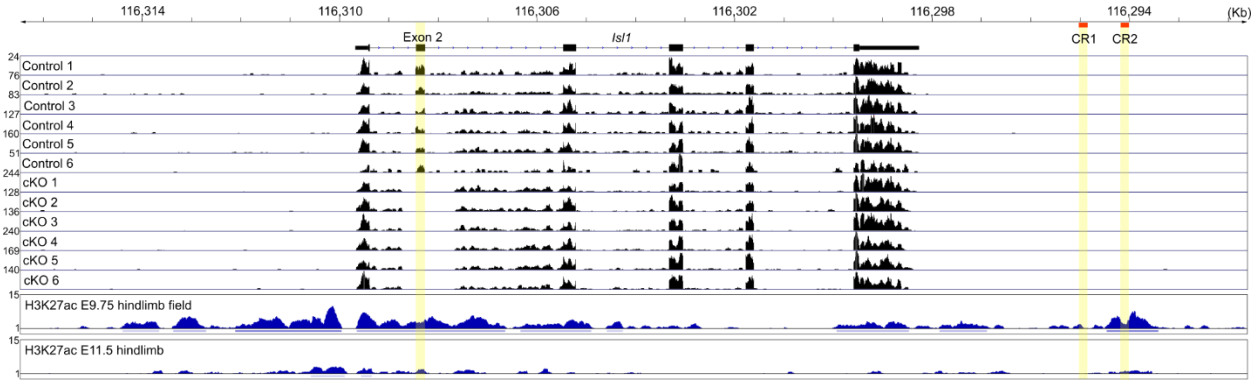


**Figure S2.1.** *Isl1* mRNA transcripts in control and conditional knockout mouse hindlimb fields. H3K27ac ChIP-seq enrichment in the hindlimb field and later hindlimb bud are shown below. mm10 coordinates chr13:116,291,619-116,316,534.

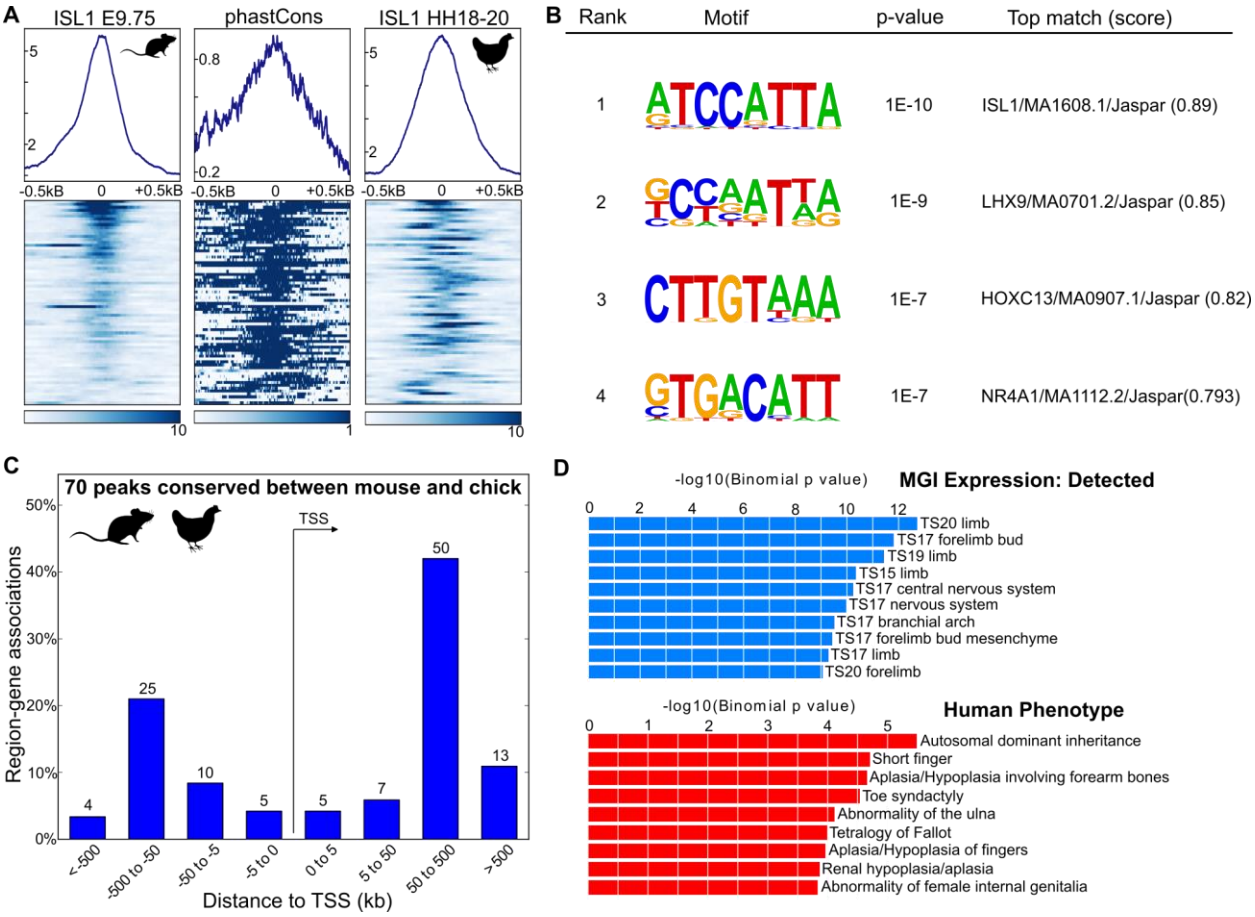


**Figure S2.2.** Conserved ISL1 peaks found in mouse and chick. **(A)** ISL1 signal in mouse, centered on mouse peak summits, as well as phastCons conservation scores, and ISL1 signal in chick at orthologous loci. **(B)** Top *de novo* motifs from conserved peaks. **(C)** Distribution of conserved peaks relative to the nearest associated genes based on GREAT. **(D)** GO analysis showing top 10 terms based on mouse embryonic gene expression (top) and on human phenotypes (bottom).

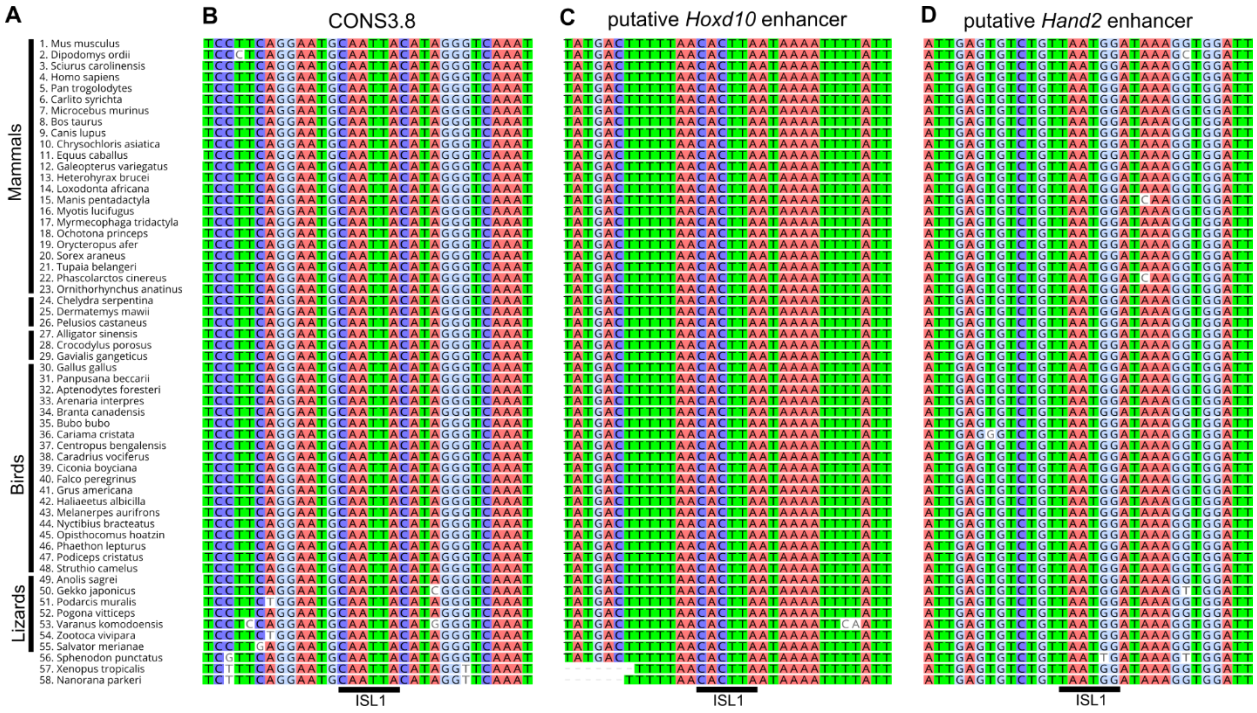


**Figure S2.3.** Multi-species alignments of three putative enhancers bound by ISL1. **(A)** Sample of species representing (from top to bottom) mammals, turtles, crocodilians, birds, lizards, tuatara, and two frog species. **(B)** Potential ISL1 binding site within the CONS3.8 enhancer. **(C)** ISL1 motif within a putative *Hoxd10* enhancer. **(D)** ISL1 motif within a putative *Hand2* enhancer.

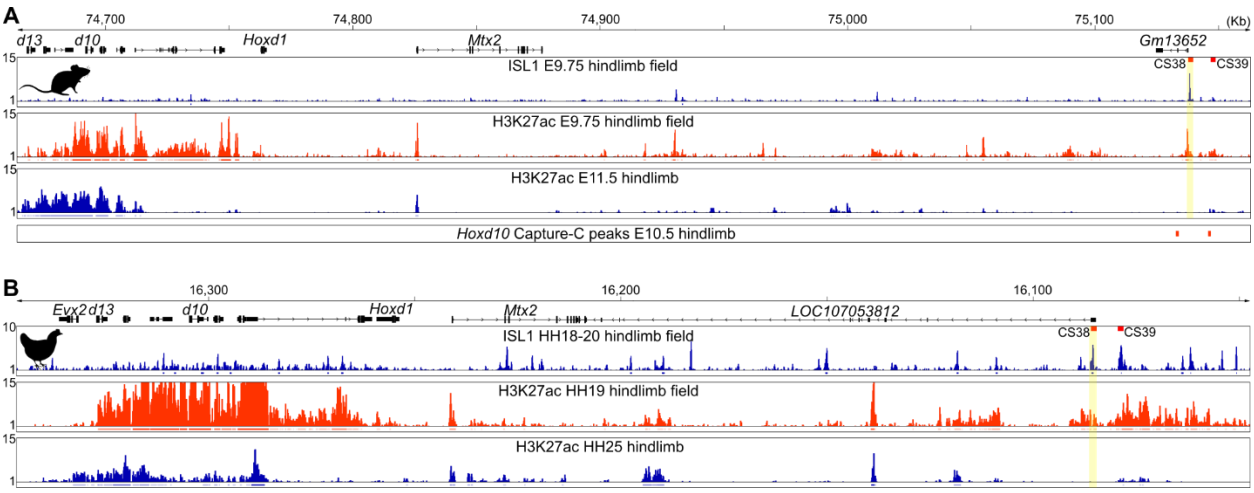


**Figure S2.4.** Conserved ISL1 binding at putative *Hoxd10* enhancer CS38 in mouse and chick. **(A)** Mouse ISL1 and H3K27ac signal in the hindlimb field, as well H3K27ac signal in the later hindlimb bud. mm10 coordinates chr2:74,664,335-75,164,335. **(B)** Chick ISL1 and H3K27ac signal in the hindlimb field, as well as H3K27ac in the later hindlimb bud. galGal6 coordinates chr7:16,047,440-16,347,440.

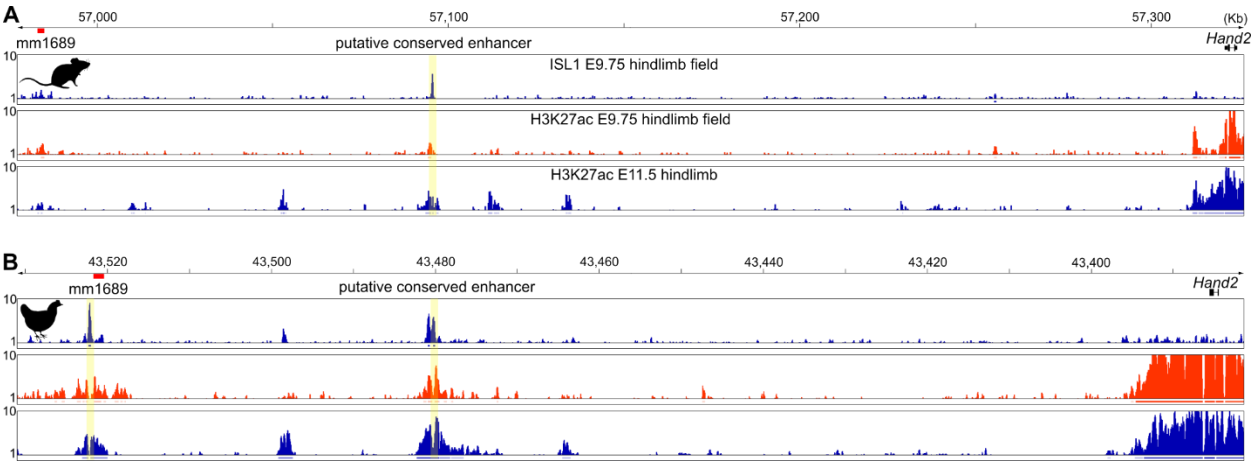


**Figure S2.5.** Equivocal evidence of ISL1 binding at putative *Hand2* enhancers. **(A)** Mouse ISL1 and H3K27ac signal in the hindlimb field, as well H3K27ac signal in the later hindlimb bud. mm10 coordinates chr8:56,977,333-57,327,333. **(B)** Chick ISL1 and H3K27ac signal in the hindlimb field, as well as H3K27ac in the later hindlimb bud. galGal6 coordinates chr4:43,381,409-43,531,409.

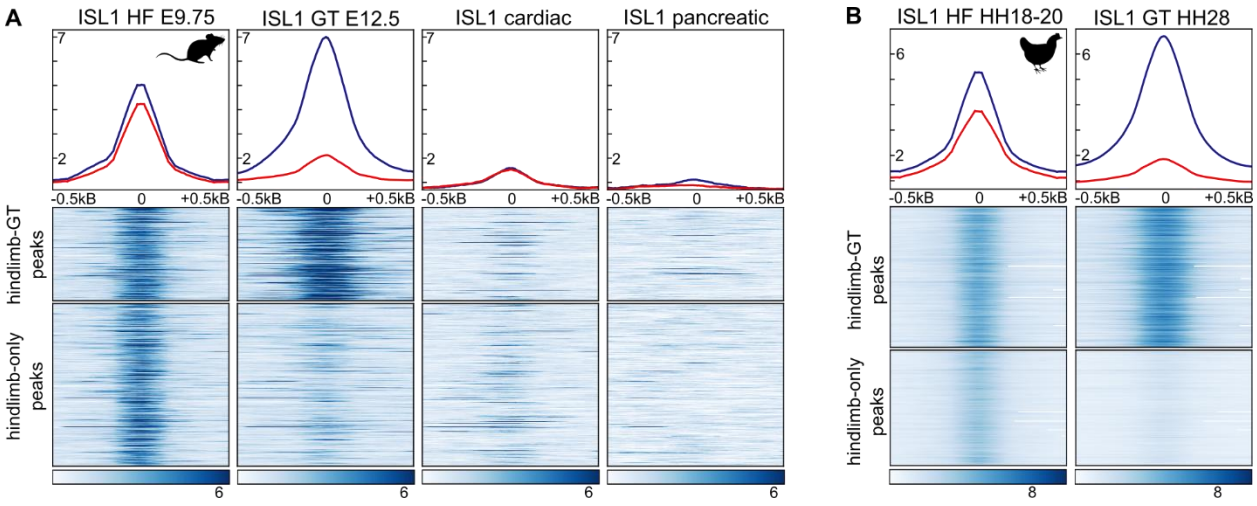


**Figure S2.6.** Shared ISL1 binding loci between the hindlimb field and genital tubercle. **(A)** Mouse ISL1 peaks shared with the GT (top, blue line plot) and specific to the hindlimb field (bottom, red line). Signal from ISL1 ChIP-seq in cardiac (Wang et al., 2016) and pancreatic cells (Ediger et al., 2017) show a lack of enrichment in these shared limb-GT regions. **(B)** Chick ISL1 peaks shared with the GT (top) and specific to the hindlimb field (bottom).

**Table S2.1.** Differentially expressed genes in the mouse hindlimb field. Genes are sorted by adjusted p-values (padj) with a 0.05 cutoff.

*refGene transcripts*

| Gene | baseMean | log2FoldChange | lfcSE | stat | pvalue | padj |
|---|---|---|---|---|---|---|
| 1110032F04Rik | 133.87304 | -1.045099118 | 0.107211 | -9.74804 | 1.88E-22 | 2.65E-18 |
| Galnt5 | 115.84669 | -1.020762393 | 0.108365 | -9.4197 | 4.52E-21 | 3.18E-17 |
| Pdlim3 | 94.395849 | -1.060674491 | 0.116157 | -9.13137 | 6.76E-20 | 3.17E-16 |
| Etv5 | 839.1243 | -0.728260001 | 0.080278 | -9.07168 | 1.17E-19 | 4.13E-16 |
| Prdm1 | 252.85109 | -1.031122624 | 0.115877 | -8.89846 | 5.66E-19 | 1.59E-15 |
| Lrrn3 | 90.942849 | -1.022639052 | 0.115871 | -8.8257 | 1.09E-18 | 2.55E-15 |
| Capn6 | 1618.0742 | -0.933744745 | 0.106852 | -8.73866 | 2.36E-18 | 4.74E-15 |
| Dsc2 | 723.31803 | -0.786798849 | 0.093603 | -8.4057 | 4.25E-17 | 7.49E-14 |
| Popdc2 | 179.38015 | -0.808760358 | 0.098547 | -8.20684 | 2.27E-16 | 3.55E-13 |
| Parm1 | 515.82765 | 0.713838052 | 0.089392 | 7.985497 | 1.40E-15 | 1.79E-12 |
| Tle4 | 1350.6949 | -0.742532946 | 0.092984 | -7.98556 | 1.40E-15 | 1.79E-12 |
| Pdha1 | 2713.7826 | -0.504645928 | 0.063471 | -7.95079 | 1.85E-15 | 2.01E-12 |
| Unc5c | 1016.8794 | -0.857029176 | 0.107754 | -7.95354 | 1.81E-15 | 2.01E-12 |
| Bace2 | 23.191124 | -0.813516161 | 0.111679 | -7.28444 | 3.23E-13 | 3.25E-10 |
| Dusp6 | 2865.9485 | -0.744389103 | 0.102929 | -7.23209 | 4.76E-13 | 4.46E-10 |
| Spry4 | 230.74024 | -0.724220431 | 0.108907 | -6.64988 | 2.93E-11 | 2.58E-08 |
| Dlx2 | 129.56101 | -0.712998813 | 0.108633 | -6.56335 | 5.26E-11 | 4.36E-08 |
| Steap2 | 955.54004 | 0.647304611 | 0.09895 | 6.541764 | 6.08E-11 | 4.76E-08 |
| Ints6l | 822.13159 | -0.603515344 | 0.092899 | -6.49647 | 8.22E-11 | 6.09E-08 |
| Apcdd1 | 2416.0722 | -0.631641592 | 0.098552 | -6.40924 | 1.46E-10 | 1.03E-07 |
| Sall3 | 473.11171 | -0.675327501 | 0.105721 | -6.38784 | 1.68E-10 | 1.13E-07 |
| Angpt2 | 93.150021 | -0.700004968 | 0.110373 | -6.34218 | 2.27E-10 | 1.39E-07 |
| Etv4 | 507.01792 | -0.660480508 | 0.10409 | -6.34531 | 2.22E-10 | 1.39E-07 |
| Adamts5 | 33.378315 | -0.670199133 | 0.108383 | -6.1836 | 6.27E-10 | 3.68E-07 |
| Wnt2 | 922.74079 | 0.625588833 | 0.101339 | 6.17324 | 6.69E-10 | 3.77E-07 |
| Dtx4 | 299.58841 | -0.665587392 | 0.108678 | -6.12437 | 9.10E-10 | 4.93E-07 |
| Chst15 | 224.26477 | -0.588949951 | 0.097008 | -6.07112 | 1.27E-09 | 6.62E-07 |
| Elovl7 | 21.289499 | -0.669202695 | 0.110959 | -6.03108 | 1.63E-09 | 8.19E-07 |
| Frzb | 843.18106 | -0.488590576 | 0.081598 | -5.98777 | 2.13E-09 | 1.03E-06 |
| Nfam1 | 153.80681 | -0.647357957 | 0.10837 | -5.97359 | 2.32E-09 | 1.09E-06 |
| Mab21l1 | 20.678387 | -0.547779816 | 0.092778 | -5.9042 | 3.54E-09 | 1.61E-06 |
| Tcf7 | 1654.5623 | -0.408384109 | 0.0694 | -5.8845 | 3.99E-09 | 1.76E-06 |
| Gask1b | 531.8063 | -0.542062832 | 0.093104 | -5.82214 | 5.81E-09 | 2.48E-06 |
| Tril | 2812.9532 | -0.668665936 | 0.116249 | -5.752 | 8.82E-09 | 3.65E-06 |
| Spry1 | 354.02356 | -0.564128009 | 0.098509 | -5.72668 | 1.02E-08 | 4.12E-06 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Hoxd1 | 244.28914 | -0.651532076 | 0.11485 | -5.67288 | 1.40E-08 | 5.34E-06 |
| Hsd11b2 | 376.76828 | -0.631247701 | 0.111194 | -5.67699 | 1.37E-08 | 5.34E-06 |
| Gas1 | 6691.9331 | 0.502908585 | 0.089012 | 5.649927 | 1.61E-08 | 5.95E-06 |
| Il17rd | 1367.9024 | -0.567964528 | 0.101193 | -5.61267 | 1.99E-08 | 7.19E-06 |
| Dgkk | 351.89357 | -0.640846238 | 0.114367 | -5.60343 | 2.10E-08 | 7.40E-06 |
| B4galt6 | 1026.581 | -0.532929603 | 0.09597 | -5.5531 | 2.81E-08 | 9.41E-06 |
| Lmo7 | 259.81807 | -0.540366699 | 0.097294 | -5.55398 | 2.79E-08 | 9.41E-06 |
| Dpysl3 | 2781.9166 | -0.482776828 | 0.08723 | -5.53454 | 3.12E-08 | 1.02E-05 |
| Rspo1 | 540.39344 | 0.570044844 | 0.103163 | 5.525659 | 3.28E-08 | 1.05E-05 |
| Amph | 278.07447 | -0.598205151 | 0.108562 | -5.51025 | 3.58E-08 | 1.12E-05 |
| Evx2 | 50.518844 | -0.599583009 | 0.109568 | -5.47223 | 4.44E-08 | 1.36E-05 |
| Fmn1 | 151.21129 | -0.606986771 | 0.111177 | -5.45966 | 4.77E-08 | 1.43E-05 |
| Hs3st3b1 | 355.77435 | -0.550165004 | 0.101116 | -5.44091 | 5.30E-08 | 1.55E-05 |
| Nfatc4 | 1910.4342 | 0.422126496 | 0.077652 | 5.436109 | 5.45E-08 | 1.56E-05 |
| Nlrp6 | 143.37069 | -0.628998652 | 0.116359 | -5.40566 | 6.46E-08 | 1.82E-05 |
| Nsg2 | 99.320308 | 0.598243021 | 0.112953 | 5.296378 | 1.18E-07 | 3.26E-05 |
| Rai14 | 2441.4364 | -0.396985351 | 0.075594 | -5.25154 | 1.51E-07 | 4.08E-05 |
| Pcdh7 | 1213.3575 | -0.471733916 | 0.090053 | -5.23842 | 1.62E-07 | 4.30E-05 |
| Wnt5a | 1521.8097 | -0.487122363 | 0.093793 | -5.19358 | 2.06E-07 | 5.38E-05 |
| Irx3 | 2245.4346 | 0.457679209 | 0.088186 | 5.189923 | 2.10E-07 | 5.39E-05 |
| Chrna6 | 13.241619 | -0.442043631 | 0.085715 | -5.15714 | 2.51E-07 | 6.30E-05 |
| Lgi2 | 296.24514 | -0.485650023 | 0.095503 | -5.08518 | 3.67E-07 | 9.07E-05 |
| Slitrk6 | 450.96935 | 0.581107084 | 0.115113 | 5.048156 | 4.46E-07 | 0.000108 |
| Atp1b1 | 112.83292 | -0.559824609 | 0.112156 | -4.99147 | 5.99E-07 | 0.000141 |
| Cnksr2 | 289.81187 | -0.509143693 | 0.101998 | -4.99172 | 5.98E-07 | 0.000141 |
| Lratd2 | 368.12757 | -0.443553641 | 0.089328 | -4.96544 | 6.85E-07 | 0.000158 |
| Cspg4 | 72.931163 | 0.567684106 | 0.114558 | 4.955419 | 7.22E-07 | 0.000164 |
| Fgfr3 | 559.47117 | 0.49478878 | 0.100032 | 4.946314 | 7.56E-07 | 0.000166 |
| Zhx3 | 154.88774 | -0.52938579 | 0.106985 | -4.94824 | 7.49E-07 | 0.000166 |
| Rxfp1 | 9.9633826 | -0.4229302 | 0.085598 | -4.94091 | 7.78E-07 | 0.000168 |
| Jun | 1150.157 | 0.43257736 | 0.088269 | 4.900671 | 9.55E-07 | 0.000204 |
| Lmo2 | 444.5884 | -0.43342798 | 0.088701 | -4.88638 | 1.03E-06 | 0.000216 |
| Zbtb16 | 532.62797 | -0.487868864 | 0.100519 | -4.85349 | 1.21E-06 | 0.000251 |
| 6530402F18Rik | 253.66498 | -0.500489032 | 0.103877 | -4.81811 | 1.45E-06 | 0.000296 |
| Ppp2r2b | 422.89031 | 0.539046681 | 0.112081 | 4.809443 | 1.51E-06 | 0.000304 |
| Bcl9l | 1419.8831 | 0.306020474 | 0.064001 | 4.781481 | 1.74E-06 | 0.000345 |
| Clgn | 32.117546 | -0.53146273 | 0.111341 | -4.7733 | 1.81E-06 | 0.000354 |
| Ncald | 385.21428 | 0.42259802 | 0.08916 | 4.739774 | 2.14E-06 | 0.000411 |
| Prkci | 3926.5171 | -0.358338155 | 0.075673 | -4.73534 | 2.19E-06 | 0.000411 |
| Tspan12 | 313.04999 | -0.492750327 | 0.104019 | -4.73713 | 2.17E-06 | 0.000411 |
| Ets2 | 2603.8719 | -0.489228883 | 0.104092 | -4.69998 | 2.60E-06 | 0.000482 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Col3a1 | 9183.7032 | 0.435396032 | 0.092862 | 4.688647 | 2.75E-06 | 0.000503 |
| Zhx2 | 188.98494 | -0.472207225 | 0.101662 | -4.64486 | 3.40E-06 | 0.000614 |
| Hopx | 37.89513 | -0.483161666 | 0.104405 | -4.62778 | 3.70E-06 | 0.000659 |
| Abca8b | 204.99103 | -0.49069377 | 0.10667 | -4.6001 | 4.22E-06 | 0.000743 |
| Ndnf | 781.70513 | 0.435950929 | 0.095456 | 4.567049 | 4.95E-06 | 0.00086 |
| Tmem30b | 122.06621 | 0.529556544 | 0.116314 | 4.552808 | 5.29E-06 | 0.000909 |
| Prep | 3440.1588 | -0.224222929 | 0.049521 | -4.52784 | 5.96E-06 | 0.001011 |
| Hand1 | 1197.0569 | 0.472217409 | 0.104607 | 4.5142 | 6.36E-06 | 0.001065 |
| Cttnbp2 | 267.38575 | 0.460229576 | 0.102223 | 4.502212 | 6.72E-06 | 0.001114 |
| Igfbp3 | 346.78842 | -0.488924742 | 0.108908 | -4.48935 | 7.14E-06 | 0.00117 |
| Adcy5 | 67.541513 | 0.512140384 | 0.114809 | 4.460813 | 8.16E-06 | 0.001321 |
| 4930481A15Rik | 13.962596 | 0.437954645 | 0.098485 | 4.446909 | 8.71E-06 | 0.001385 |
| Pdgfra | 3940.727 | 0.395750144 | 0.089016 | 4.445822 | 8.76E-06 | 0.001385 |
| Mybpc1 | 231.94783 | 0.516714318 | 0.116351 | 4.440989 | 8.95E-06 | 0.001401 |
| Adamts17 | 158.23889 | -0.479580323 | 0.108407 | -4.42389 | 9.69E-06 | 0.001484 |
| Phldb2 | 2491.6346 | -0.457364365 | 0.103336 | -4.42598 | 9.60E-06 | 0.001484 |
| Vegfa | 1257.7914 | -0.391699231 | 0.088852 | -4.40846 | 1.04E-05 | 0.001576 |
| Galnt4 | 335.0625 | -0.446349434 | 0.10142 | -4.40098 | 1.08E-05 | 0.001614 |
| Dusp4 | 505.05716 | -0.481945526 | 0.109898 | -4.3854 | 1.16E-05 | 0.001716 |
| Lamc3 | 46.361401 | 0.506786585 | 0.115699 | 4.38022 | 1.19E-05 | 0.001739 |
| Mdfic | 184.23539 | -0.454129311 | 0.104317 | -4.35337 | 1.34E-05 | 0.001946 |
| Adam19 | 1280.1995 | -0.376001905 | 0.086805 | -4.33156 | 1.48E-05 | 0.002127 |
| Vegfc | 746.61595 | 0.336016046 | 0.077824 | 4.317616 | 1.58E-05 | 0.002243 |
| Psd2 | 387.54113 | 0.419524449 | 0.097788 | 4.290144 | 1.79E-05 | 0.002514 |
| 0610040J01Rik | 114.6981 | -0.47939625 | 0.112323 | -4.26802 | 1.97E-05 | 0.002641 |
| Col26a1 | 1703.3127 | 0.326696233 | 0.07653 | 4.26888 | 1.96E-05 | 0.002641 |
| Ism1 | 200.24156 | 0.43997421 | 0.102926 | 4.274654 | 1.91E-05 | 0.002641 |
| Mpped2 | 474.07638 | 0.435189125 | 0.101807 | 4.274665 | 1.91E-05 | 0.002641 |
| Rcn3 | 1586.1729 | 0.319252718 | 0.074833 | 4.266227 | 1.99E-05 | 0.002641 |
| Wnt4 | 520.01449 | 0.412193538 | 0.096579 | 4.267939 | 1.97E-05 | 0.002641 |
| Fgf20 | 28.5458 | -0.412722495 | 0.096877 | -4.26029 | 2.04E-05 | 0.002662 |
| Lrig3 | 794.69575 | 0.388138452 | 0.091094 | 4.260877 | 2.04E-05 | 0.002662 |
| Masp1 | 294.53286 | -0.432279756 | 0.101869 | -4.24349 | 2.20E-05 | 0.002843 |
| Gja1 | 4624.6492 | -0.334297703 | 0.079001 | -4.23157 | 2.32E-05 | 0.002943 |
| Hoxd10 | 783.10141 | -0.362634372 | 0.085656 | -4.23362 | 2.30E-05 | 0.002943 |
| Rnf128 | 133.48293 | -0.473031029 | 0.111838 | -4.2296 | 2.34E-05 | 0.002943 |
| Notum | 239.98113 | 0.448868818 | 0.106196 | 4.226813 | 2.37E-05 | 0.002953 |
| Lef1 | 2302.7842 | -0.333216212 | 0.078911 | -4.22268 | 2.41E-05 | 0.002982 |
| Pdzd2 | 82.454868 | -0.480818675 | 0.114105 | -4.21381 | 2.51E-05 | 0.003074 |
| Pth1r | 987.13641 | 0.36541565 | 0.087036 | 4.198429 | 2.69E-05 | 0.003262 |
| Dll4 | 70.118392 | -0.487245678 | 0.116233 | -4.19198 | 2.77E-05 | 0.003328 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Spred2 | 943.19298 | -0.294750111 | 0.070589 | -4.17557 | 2.97E-05 | 0.003547 |
| Bves | 217.86464 | -0.464303891 | 0.111482 | -4.16484 | 3.12E-05 | 0.003687 |
| Bnc1 | 824.06282 | 0.408113087 | 0.098047 | 4.16241 | 3.15E-05 | 0.003695 |
| Evx1os | 228.11891 | -0.466181323 | 0.112765 | -4.13411 | 3.56E-05 | 0.004146 |
| Pink1 | 206.98911 | 0.445175223 | 0.107858 | 4.127435 | 3.67E-05 | 0.004234 |
| Cdon | 1656.0522 | 0.411116556 | 0.10058 | 4.087447 | 4.36E-05 | 0.004952 |
| Sorbs2 | 70.649401 | -0.474658355 | 0.116124 | -4.08752 | 4.36E-05 | 0.004952 |
| Bmp4 | 1185.8359 | -0.37234907 | 0.091199 | -4.08282 | 4.45E-05 | 0.005012 |
| Nptx2 | 53.766573 | -0.471468389 | 0.116121 | -4.06014 | 4.90E-05 | 0.00548 |
| Parp9 | 38.654819 | -0.469822545 | 0.115776 | -4.05802 | 4.95E-05 | 0.005487 |
| Il1rap | 74.300799 | -0.468138449 | 0.116046 | -4.03408 | 5.48E-05 | 0.00603 |
| Col5a1 | 3122.4213 | 0.295347438 | 0.073639 | 4.010761 | 6.05E-05 | 0.006606 |
| Cped1 | 160.49252 | -0.46243353 | 0.115542 | -4.00231 | 6.27E-05 | 0.006794 |
| Ankrd6 | 257.81803 | 0.355425136 | 0.088954 | 3.995595 | 6.45E-05 | 0.006936 |
| Snai2 | 610.00278 | 0.378477965 | 0.095331 | 3.970157 | 7.18E-05 | 0.007661 |
| Pkp2 | 372.86898 | -0.424993499 | 0.107342 | -3.95926 | 7.52E-05 | 0.0079 |
| Steap1 | 271.3553 | 0.460710753 | 0.11635 | 3.959692 | 7.50E-05 | 0.0079 |
| Dok4 | 2600.2375 | -0.343574878 | 0.086972 | -3.9504 | 7.80E-05 | 0.008096 |
| LOC101055843 | 56.127896 | -0.444634722 | 0.11262 | -3.94809 | 7.88E-05 | 0.008096 |
| Spry2 | 240.8192 | -0.396842573 | 0.10048 | -3.94948 | 7.83E-05 | 0.008096 |
| Cyp26b1 | 87.70355 | -0.44219142 | 0.112424 | -3.93324 | 8.38E-05 | 0.008551 |
| Colec12 | 1026.8158 | -0.285188683 | 0.072563 | -3.93022 | 8.49E-05 | 0.008597 |
| Fut8 | 521.03331 | -0.317852845 | 0.081029 | -3.92272 | 8.76E-05 | 0.008806 |
| Ifrd1 | 977.70941 | -0.312973888 | 0.080123 | -3.90617 | 9.38E-05 | 0.009364 |
| Dtx3l | 69.890817 | -0.453754915 | 0.116356 | -3.89971 | 9.63E-05 | 0.00955 |
| Nrp2 | 3142.5835 | 0.288639312 | 0.074063 | 3.897205 | 9.73E-05 | 0.009581 |
| Stx3 | 168.3538 | -0.409785976 | 0.105307 | -3.89133 | 9.97E-05 | 0.009748 |
| Fgf10 | 354.84732 | -0.391422522 | 0.100943 | -3.87765 | 0.000105 | 0.010242 |
| Cd44 | 81.961224 | -0.417634167 | 0.108286 | -3.85677 | 0.000115 | 0.011005 |
| Prrx1 | 3643.154 | -0.398080769 | 0.103213 | -3.85688 | 0.000115 | 0.011005 |
| Cdo1 | 128.10242 | -0.44312991 | 0.115293 | -3.84353 | 0.000121 | 0.011531 |
| Gabra4 | 34.040384 | -0.426866411 | 0.111118 | -3.84157 | 0.000122 | 0.011531 |
| Rarb | 774.05559 | 0.290025667 | 0.07552 | 3.840383 | 0.000123 | 0.011531 |
| Celf3 | 137.98062 | 0.435190783 | 0.11368 | 3.8282 | 0.000129 | 0.012036 |
| Cbx8 | 80.571485 | 0.437361205 | 0.114484 | 3.820271 | 0.000133 | 0.012348 |
| Cldn10 | 51.429196 | -0.437659431 | 0.115173 | -3.80001 | 0.000145 | 0.013229 |
| Cpz | 122.40912 | 0.429098491 | 0.112892 | 3.800976 | 0.000144 | 0.013229 |
| Serinc5 | 441.5734 | -0.394274286 | 0.10387 | -3.79583 | 0.000147 | 0.013367 |
| Col5a2 | 2536.139 | 0.26891533 | 0.070881 | 3.793925 | 0.000148 | 0.013384 |
| Tbx18 | 27.149199 | -0.327782499 | 0.086873 | -3.77312 | 0.000161 | 0.014458 |
| Lins1 | 451.6708 | -0.263141728 | 0.06995 | -3.76183 | 0.000169 | 0.015031 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Nkd1 | 1420.6578 | 0.235877595 | 0.062811 | 3.755352 | 0.000173 | 0.015328 |
| Evx1 | 328.50511 | -0.434387478 | 0.115896 | -3.74808 | 0.000178 | 0.01568 |
| Ptx3 | 105.29977 | 0.427291715 | 0.11409 | 3.745211 | 0.00018 | 0.01568 |
| Tmem200c | 29.569283 | -0.429764527 | 0.114757 | -3.74498 | 0.00018 | 0.01568 |
| Bmf | 641.90768 | 0.386181725 | 0.103286 | 3.738965 | 0.000185 | 0.015864 |
| S1pr3 | 1091.5754 | 0.350923938 | 0.093841 | 3.739565 | 0.000184 | 0.015864 |
| Irx5 | 772.75172 | 0.339966742 | 0.090992 | 3.736217 | 0.000187 | 0.015941 |
| Cd248 | 705.2825 | 0.317630279 | 0.08523 | 3.726754 | 0.000194 | 0.016364 |
| Tbx4 | 1192.8304 | -0.422593981 | 0.1134 | -3.72658 | 0.000194 | 0.016364 |
| Dlx1 | 25.6212 | -0.421694208 | 0.113693 | -3.70906 | 0.000208 | 0.017435 |
| Dlx5 | 77.993093 | -0.421528382 | 0.114032 | -3.69659 | 0.000219 | 0.018206 |
| Sall1 | 2057.0163 | -0.374385841 | 0.10174 | -3.67983 | 0.000233 | 0.019331 |
| Aff2 | 247.13283 | 0.426837929 | 0.116057 | 3.677816 | 0.000235 | 0.019369 |
| Slc22a17 | 416.90463 | 0.37052816 | 0.100909 | 3.671911 | 0.000241 | 0.019707 |
| F2r | 3460.9595 | -0.249213909 | 0.067912 | -3.66967 | 0.000243 | 0.019766 |
| Cdh11 | 1604.0052 | -0.340452663 | 0.092852 | -3.6666 | 0.000246 | 0.01989 |
| Pde4d | 235.32157 | 0.342857252 | 0.093683 | 3.659755 | 0.000252 | 0.020312 |
| Irx4 | 50.204662 | 0.424897027 | 0.116342 | 3.652133 | 0.00026 | 0.020806 |
| Cygb | 88.545637 | -0.421559346 | 0.116095 | -3.63116 | 0.000282 | 0.022445 |
| Slc22a3 | 31.01879 | 0.411088272 | 0.113329 | 3.627385 | 0.000286 | 0.022526 |
| St6gal2 | 193.60578 | -0.418745137 | 0.115442 | -3.62732 | 0.000286 | 0.022526 |
| Cftr | 70.613359 | 0.419788401 | 0.115828 | 3.624228 | 0.00029 | 0.022671 |
| Greb1 | 2419.0423 | -0.383992848 | 0.106187 | -3.61621 | 0.000299 | 0.023128 |
| Zmym3 | 744.51178 | 0.250637374 | 0.069289 | 3.617295 | 0.000298 | 0.023128 |
| Nuak1 | 466.66733 | -0.319483809 | 0.088881 | -3.5945 | 0.000325 | 0.025007 |
| Fam107a | 68.079074 | 0.415260679 | 0.115713 | 3.588723 | 0.000332 | 0.025311 |
| Lin28b | 3307.4931 | -0.351866685 | 0.09812 | -3.58609 | 0.000336 | 0.025311 |
| Nkd2 | 527.78125 | -0.349705889 | 0.097454 | -3.58842 | 0.000333 | 0.025311 |
| Ripor3 | 95.492861 | 0.417150511 | 0.116337 | 3.585711 | 0.000336 | 0.025311 |
| Irs2 | 830.83081 | -0.291281771 | 0.081347 | -3.58072 | 0.000343 | 0.025584 |
| Mtx2 | 1853.3786 | -0.189438036 | 0.052914 | -3.58014 | 0.000343 | 0.025584 |
| Sema5b | 68.646655 | -0.411400182 | 0.115096 | -3.57441 | 0.000351 | 0.026013 |
| Plxna2 | 2222.8342 | 0.374872166 | 0.105431 | 3.555616 | 0.000377 | 0.027798 |
| Tshz2 | 448.85606 | -0.285082563 | 0.0803 | -3.5502 | 0.000385 | 0.028228 |
| Trib2 | 287.59687 | -0.339632922 | 0.095821 | -3.54447 | 0.000393 | 0.0287 |
| Emp2 | 168.0856 | 0.386293398 | 0.109088 | 3.541125 | 0.000398 | 0.028917 |
| Ctsh | 52.925142 | -0.408041851 | 0.115868 | -3.52162 | 0.000429 | 0.03097 |
| Msi2 | 2977.0888 | -0.32448702 | 0.092195 | -3.51959 | 0.000432 | 0.031049 |
| Cpvl | 6.7692152 | -0.273548568 | 0.077919 | -3.51067 | 0.000447 | 0.031946 |
| Gnal | 254.39359 | -0.398941704 | 0.113792 | -3.50587 | 0.000455 | 0.032295 |
| Lrp12 | 597.80026 | -0.224291814 | 0.06399 | -3.5051 | 0.000456 | 0.032295 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Smim10l2a | 94.30305 | -0.397745422 | 0.113685 | -3.49866 | 0.000468 | 0.032919 |
| Pdpn | 430.7314 | 0.332004661 | 0.095158 | 3.488995 | 0.000485 | 0.033963 |
| Slc6a6 | 2954.2743 | -0.194213554 | 0.055794 | -3.48088 | 0.0005 | 0.034835 |
| Sncaip | 248.87206 | 0.35974701 | 0.103488 | 3.476217 | 0.000509 | 0.035218 |
| Sox12 | 2510.9235 | 0.237473286 | 0.068332 | 3.475308 | 0.00051 | 0.035218 |
| Haglr | 113.48567 | -0.398341433 | 0.114859 | -3.4681 | 0.000524 | 0.036 |
| Ptch1 | 1104.3664 | -0.312946052 | 0.090436 | -3.46042 | 0.000539 | 0.036863 |
| Alx4 | 3559.8358 | 0.30267215 | 0.087928 | 3.442292 | 0.000577 | 0.039234 |
| Fgfr2 | 2304.1071 | 0.277384324 | 0.080757 | 3.434806 | 0.000593 | 0.040011 |
| Psat1 | 4094.7248 | -0.203845851 | 0.059355 | -3.43438 | 0.000594 | 0.040011 |
| Lpar6 | 680.45993 | -0.284622546 | 0.082921 | -3.43247 | 0.000598 | 0.040101 |
| Gnpnat1 | 366.71425 | -0.26459191 | 0.077179 | -3.42828 | 0.000607 | 0.040532 |
| Abi3bp | 38.187533 | 0.390693782 | 0.114016 | 3.426656 | 0.000611 | 0.040584 |
| Prrt4 | 407.06657 | 0.355769326 | 0.103893 | 3.424374 | 0.000616 | 0.040734 |
| Mecom | 828.1278 | -0.381473556 | 0.111512 | -3.42092 | 0.000624 | 0.041063 |
| Cdhr1 | 107.95284 | -0.387734215 | 0.113617 | -3.41265 | 0.000643 | 0.042132 |
| Cdh24 | 746.76367 | 0.309541712 | 0.090793 | 3.409303 | 0.000651 | 0.042454 |
| Klhl5 | 536.12011 | -0.299191849 | 0.087932 | -3.40253 | 0.000668 | 0.043122 |
| Pck2 | 322.82624 | 0.335690121 | 0.098639 | 3.403207 | 0.000666 | 0.043122 |
| Col16a1 | 119.2473 | 0.388125603 | 0.114207 | 3.398429 | 0.000678 | 0.043179 |
| Fgf8 | 22.864841 | -0.260908803 | 0.076749 | -3.39951 | 0.000675 | 0.043179 |
| Rubie | 11.372148 | -0.299098657 | 0.087968 | -3.40008 | 0.000674 | 0.043179 |
| Mtss1 | 736.99928 | -0.26588611 | 0.078504 | -3.38691 | 0.000707 | 0.044431 |
| Ntng2 | 69.121277 | -0.393820203 | 0.116273 | -3.38704 | 0.000707 | 0.044431 |
| Runx1t1 | 564.50644 | 0.353950744 | 0.104438 | 3.389099 | 0.000701 | 0.044431 |
| Syt9 | 60.471495 | -0.386738709 | 0.11453 | -3.37675 | 0.000733 | 0.0459 |
| Tvp23a | 327.70002 | 0.269207353 | 0.079876 | 3.370321 | 0.000751 | 0.046776 |

*lncRNA transcripts*

| **Gene** | **baseMean** | **log2FoldChange** | **lfcSE** | **stat** | **pvalue** | **padj** |
|---|---|---|---|---|---|---|
| ENSMUSG00000087333.1 | 150.35096 | -0.9973085 | 0.149446 | -6.6733727 | 2.50E-11 | 1.18E-07 |
| ENSMUSG00000086938.8 | 13.592986 | 1.4008879 | 0.3005525 | 4.6610428 | 3.15E-06 | 0.0073949 |
| ENSMUSG00000100426.1 | 7.2704368 | 1.4353152 | 0.3215319 | 4.4639894 | 8.04E-06 | 0.0126061 |
| ENSMUSG00000079499.9 | 279.80327 | -0.6113139 | 0.1395343 | -4.3810996 | 1.18E-05 | 0.0138776 |
| ENSMUSG00000086126.1 | 223.6862 | -0.6995146 | 0.1720999 | -4.0645858 | 4.81E-05 | 0.0377003 |
| ENSMUSG00000105628.1 | 6.0034272 | -1.3202836 | 0.3222404 | -4.0972009 | 4.18E-05 | 0.0377003 |

**Table S2.2.** Sample of reproducible peaks from ISL1 ChIP-seq in the hindlimb field.

*Mouse*

| Chr | Start | End | -log10pval | summit |
|---|---|---|---|---|
| chr14 | 16575143 | 16575493 | 210.15003 | 173 |
| chr13 | 44730948 | 44731313 | 198.836 | 233 |
| chr5 | 1.04E+08 | 1.04E+08 | 174.49307 | 150 |
| chr14 | 69029168 | 69029489 | 166.81607 | 153 |
| chr12 | 82773858 | 82774104 | 157.42743 | 119 |
| chr1 | 1.69E+08 | 1.69E+08 | 157.16892 | 129 |
| chr15 | 43389458 | 43389805 | 138.43967 | 155 |
| chr5 | 3647059 | 3647313 | 137.39883 | 97 |
| chr11 | 85963520 | 85963864 | 134.98508 | 184 |
| chr11 | 1.1E+08 | 1.1E+08 | 134.32902 | 125 |
| chr2 | 1.81E+08 | 1.81E+08 | 125.6544 | 103 |
| chr1 | 9761333 | 9761573 | 123.97229 | 122 |
| chr5 | 20744418 | 20744711 | 121.21907 | 120 |
| chr1 | 37409254 | 37409539 | 120.18762 | 173 |
| chr11 | 72407390 | 72407628 | 118.64928 | 107 |
| chr4 | 49469887 | 49470097 | 118.19188 | 99 |
| chr2 | 75138395 | 75138625 | 117.64711 | 113 |
| chr1 | 73030863 | 73031074 | 114.92874 | 93 |
| chr4 | 1.48E+08 | 1.48E+08 | 112.74374 | 132 |
| chrX | 7041682 | 7041917 | 111.8393 | 94 |
| chr2 | 50793110 | 50793317 | 110.12545 | 96 |
| chr15 | 84809620 | 84809817 | 109.45888 | 98 |
| chr4 | 94417280 | 94417518 | 108.43145 | 115 |
| chr2 | 1.14E+08 | 1.14E+08 | 107.98111 | 120 |
| chr14 | 22205918 | 22206107 | 107.87766 | 99 |
| chr6 | 53912248 | 53912478 | 107.00833 | 118 |
| chr2 | 1.06E+08 | 1.06E+08 | 106.87687 | 107 |
| chr10 | 27285521 | 27285747 | 106.80633 | 105 |
| chr4 | 1.52E+08 | 1.52E+08 | 105.47782 | 129 |
| chr19 | 14424740 | 14424969 | 104.5949 | 117 |
| chr3 | 1.59E+08 | 1.59E+08 | 103.43814 | 87 |
| chrX | 75750469 | 75750685 | 102.69333 | 94 |
| chr7 | 99272877 | 99273101 | 102.2773 | 115 |
| chr5 | 99522769 | 99522969 | 102.16497 | 117 |
| chr6 | 1.17E+08 | 1.17E+08 | 100.86545 | 126 |
| chr18 | 68194798 | 68195021 | 100.36441 | 128 |
| chr13 | 84063521 | 84063726 | 100.26561 | 91 |

*Chick*

| Chr | Start | End | -log10pval | summit |
|---|---|---|---|---|
| chr14 | 5138608 | 5139136 | 352.6637 | 173 |
| chr19 | 6668036 | 6668372 | 347.4077 | 172 |
| chr6 | 26949226 | 26949580 | 338.3124 | 183 |
| chr15 | 92043 | 92388 | 315.2161 | 171 |
| chr3 | 52401584 | 52401910 | 313.9496 | 164 |
| chr23 | 2097447 | 2097780 | 309.8664 | 168 |
| chr2 | 1.49E+08 | 1.49E+08 | 309.4445 | 173 |
| chr15 | 6816878 | 6817186 | 303.9449 | 152 |
| chr4 | 1863132 | 1863456 | 302.4651 | 157 |
| chr2 | 43628879 | 43629177 | 293.0417 | 140 |
| chr6 | 21954581 | 21954878 | 281.2626 | 142 |
| chr14 | 451125 | 451413 | 273.9812 | 143 |
| chr9 | 19987877 | 19988388 | 265.4523 | 176 |
| chr1 | 48748722 | 48749069 | 259.4798 | 168 |
| chr3 | 31160906 | 31161251 | 254.5021 | 187 |
| chr26 | 4177014 | 4177322 | 248.304 | 161 |
| chr1 | 51128106 | 51128386 | 245.5928 | 144 |
| chr4 | 89132184 | 89132504 | 244.9877 | 170 |
| chr2 | 1.32E+08 | 1.32E+08 | 242.6487 | 221 |
| chr2 | 37693632 | 37693936 | 241.5692 | 133 |
| chr12 | 10777291 | 10777605 | 241.426 | 156 |
| chr7 | 17339407 | 17339687 | 239.6053 | 138 |
| chr3 | 39776025 | 39776316 | 239.3416 | 130 |
| chr19 | 10133209 | 10133625 | 234.9881 | 147 |
| chrUn_NW_020110167v1 | 864776 | 865061 | 232.5759 | 147 |
| chr4 | 77442718 | 77443013 | 229.3742 | 153 |
| chr1 | 2572517 | 2572842 | 226.059 | 165 |
| chrZ | 20211566 | 20211890 | 223.2031 | 190 |
| chr3 | 25456680 | 25456959 | 222.6992 | 144 |
| chr33 | 3315218 | 3315524 | 222.3056 | 157 |
| chr1 | 53897862 | 53898156 | 215.3223 | 168 |
| chr22 | 3763481 | 3763776 | 215.2589 | 150 |
| chr3 | 16246599 | 16246927 | 214.5717 | 164 |
| chr20 | 5019457 | 5019752 | 214.5372 | 121 |
| chr2 | 1.38E+08 | 1.38E+08 | 213.3568 | 134 |
| chr4 | 20575393 | 20575839 | 204.2353 | 242 |
| chrZ | 66692580 | 66692890 | 202.437 | 196 |
| chr7 | 9068146 | 9068489 | 200.992 | 182 |

*Conserved*

| Chr | Start | End | -log10pval | summit |
|---|---|---|---|---|
| chr14 | 16575143 | 16575493 | 210.15 | 173 |
| chr13 | 44730948 | 44731313 | 198.836 | 233 |
| chr14 | 69029168 | 69029489 | 166.8161 | 153 |
| chr11 | 85963520 | 85963864 | 134.9851 | 184 |
| chr2 | 75138395 | 75138625 | 117.6471 | 113 |
| chr6 | 53912248 | 53912478 | 107.0083 | 118 |
| chr13 | 84063521 | 84063726 | 100.2656 | 91 |
| chr3 | 1.41E+08 | 1.41E+08 | 93.25476 | 78 |
| chr2 | 46547679 | 46547857 | 81.36342 | 83 |
| chr13 | 48066886 | 48067031 | 80.89652 | 84 |
| chr1 | 1.21E+08 | 1.21E+08 | 72.09069 | 208 |
| chr1 | 16229889 | 16229993 | 71.646 | 43 |
| chr8 | 1.15E+08 | 1.15E+08 | 63.98542 | 77 |
| chr16 | 74216908 | 74217324 | 62.84598 | 208 |
| chr19 | 55743415 | 55743831 | 62.08653 | 208 |
| chr11 | 1.09E+08 | 1.09E+08 | 60.67002 | 208 |
| chr17 | 51398635 | 51398749 | 59.95253 | 55 |
| chr17 | 62701321 | 62701435 | 55.32519 | 59 |
| chr8 | 26718830 | 26718965 | 54.45256 | 89 |
| chr2 | 1.05E+08 | 1.05E+08 | 53.88262 | 42 |
| chr19 | 59690655 | 59691071 | 53.09065 | 208 |
| chr6 | 99013592 | 99014008 | 51.75037 | 208 |
| chr6 | 51961102 | 51961518 | 49.99922 | 208 |
| chr17 | 25593809 | 25594225 | 48.1189 | 208 |
| chr14 | 46436529 | 46436945 | 46.2499 | 208 |
| chr3 | 67036665 | 67037081 | 45.75337 | 208 |
| chr3 | 30304112 | 30304528 | 45.14953 | 208 |
| chr14 | 22760323 | 22760739 | 44.74737 | 208 |
| chr3 | 55685631 | 55686047 | 44.39775 | 208 |
| chr13 | 78862863 | 78863279 | 43.85344 | 208 |
| chr14 | 72295071 | 72295487 | 42.991 | 208 |
| chr14 | 22794614 | 22795030 | 42.84497 | 208 |
| chr13 | 42709247 | 42709663 | 42.62092 | 208 |
| chr2 | 1.37E+08 | 1.37E+08 | 42.33899 | 208 |
| chr2 | 93738659 | 93739075 | 41.94231 | 208 |
| chr18 | 38483685 | 38484101 | 41.81141 | 208 |
| chr14 | 69029314 | 69029730 | 41.52025 | 208 |
| chr2 | 1.34E+08 | 1.34E+08 | 38.80175 | 208 |

**Table S2.3.** Putative direct targets of ISL1 in the hindlimb field.

*Mouse directly activated*

| Gene | baseMean | log2FoldChange | padj |
|---|---|---|---|
| Lrrn3 | 90.94285 | -1.02264 | 2.55E-15 |
| Dsc2 | 723.318 | -0.7868 | 7.49E-14 |
| Popdc2 | 179.3801 | -0.80876 | 3.55E-13 |
| Tle4 | 1350.695 | -0.74253 | 1.79E-12 |
| Unc5c | 1016.879 | -0.85703 | 2.01E-12 |
| Spry4 | 230.7402 | -0.72422 | 2.58E-08 |
| Angpt2 | 93.15002 | -0.7 | 1.39E-07 |
| Mab21l1 | 20.67839 | -0.54778 | 1.61E-06 |
| Tcf7 | 1654.562 | -0.40838 | 1.76E-06 |
| Dpysl3 | 2781.917 | -0.48278 | 1.02E-05 |
| Fmn1 | 151.2113 | -0.60699 | 1.43E-05 |
| Pcdh7 | 1213.357 | -0.47173 | 4.30E-05 |
| Wnt5a | 1521.81 | -0.48712 | 5.38E-05 |
| Cnksr2 | 289.8119 | -0.50914 | 0.000141 |
| Atp1b1 | 112.8329 | -0.55982 | 0.000141 |
| Rxfp1 | 9.963383 | -0.42293 | 0.000168 |
| Abca8b | 204.991 | -0.49069 | 0.000743 |
| Prep | 3440.159 | -0.22422 | 0.001011 |
| Phldb2 | 2491.635 | -0.45736 | 0.001484 |
| Mdfic | 184.2354 | -0.45413 | 0.001946 |
| Fgf20 | 28.5458 | -0.41272 | 0.002662 |
| Masp1 | 294.5329 | -0.43228 | 0.002843 |
| Gja1 | 4624.649 | -0.3343 | 0.002943 |
| Pdzd2 | 82.45487 | -0.48082 | 0.003074 |
| Dll4 | 70.11839 | -0.48725 | 0.003328 |
| Bmp4 | 1185.836 | -0.37235 | 0.005012 |
| Cyp26b1 | 87.70355 | -0.44219 | 0.008551 |
| Fut8 | 521.0333 | -0.31785 | 0.008806 |
| Prrx1 | 3643.154 | -0.39808 | 0.011005 |
| Cdo1 | 128.1024 | -0.44313 | 0.011531 |
| Evx1 | 328.5051 | -0.43439 | 0.01568 |
| Tmem200c | 29.56928 | -0.42976 | 0.01568 |
| Tbx4 | 1192.83 | -0.42259 | 0.016364 |
| Greb1 | 2419.042 | -0.38399 | 0.023128 |
| Nuak1 | 466.6673 | -0.31948 | 0.025007 |
| Lin28b | 3307.493 | -0.35187 | 0.025311 |
| Mtx2 | 1853.379 | -0.18944 | 0.025584 |

| | | | |
|---|---|---|---|
| Sema5b | 68.64666 | -0.4114 | 0.026013 |
| Cpvl | 6.769215 | -0.27355 | 0.031946 |
| Gnal | 254.3936 | -0.39894 | 0.032295 |
| Psat1 | 4094.725 | -0.20385 | 0.040011 |
| Mecom | 828.1278 | -0.38147 | 0.041063 |
| Cdhr1 | 107.9528 | -0.38773 | 0.042132 |
| Ntng2 | 69.12128 | -0.39382 | 0.044431 |

*Chick directly activated*

| **Gene** | **baseMean** | **log2FoldChange** | **padj** |
|---|---|---|---|
| Etv5 | 839.1243 | -0.72826 | 4.13E-16 |
| Prdm1 | 252.8511 | -1.03112 | 1.59E-15 |
| Tle4 | 1350.695 | -0.74253 | 1.79E-12 |
| Unc5c | 1016.879 | -0.85703 | 2.01E-12 |
| Dusp6 | 2865.949 | -0.74439 | 4.46E-10 |
| Dlx2 | 129.561 | -0.713 | 4.36E-08 |
| Apcdd1 | 2416.072 | -0.63164 | 1.03E-07 |
| Sall3 | 473.1117 | -0.67533 | 1.13E-07 |
| Adamts5 | 33.37832 | -0.6702 | 3.68E-07 |
| Chst15 | 224.2648 | -0.58895 | 6.62E-07 |
| Nfam1 | 153.8068 | -0.64736 | 1.09E-06 |
| Mab21l1 | 20.67839 | -0.54778 | 1.61E-06 |
| Gask1b | 531.8063 | -0.54206 | 2.48E-06 |
| Tril | 2812.953 | -0.66867 | 3.65E-06 |
| Spry1 | 354.0236 | -0.56413 | 4.12E-06 |
| Hoxd1 | 244.2891 | -0.65153 | 5.34E-06 |
| Il17rd | 1367.902 | -0.56796 | 7.19E-06 |
| Fmn1 | 151.2113 | -0.60699 | 1.43E-05 |
| Hs3st3b1 | 355.7743 | -0.55017 | 1.55E-05 |
| Rai14 | 2441.436 | -0.39699 | 4.08E-05 |
| Pcdh7 | 1213.357 | -0.47173 | 4.30E-05 |
| Wnt5a | 1521.81 | -0.48712 | 5.38E-05 |
| Cnksr2 | 289.8119 | -0.50914 | 0.000141 |
| Zbtb16 | 532.628 | -0.48787 | 0.000251 |
| Prkci | 3926.517 | -0.35834 | 0.000411 |
| Zhx2 | 188.9849 | -0.47221 | 0.000614 |
| Prep | 3440.159 | -0.22422 | 0.001011 |
| Igfbp3 | 346.7884 | -0.48892 | 0.00117 |
| Vegfa | 1257.791 | -0.3917 | 0.001576 |
| Mdfic | 184.2354 | -0.45413 | 0.001946 |

| | | | |
|---|---|---|---|
| 0610040J01Rik | 114.6981 | -0.4794 | 0.002641 |
| Fgf20 | 28.5458 | -0.41272 | 0.002662 |
| Gja1 | 4624.649 | -0.3343 | 0.002943 |
| Hoxd10 | 783.1014 | -0.36263 | 0.002943 |
| Pdzd2 | 82.45487 | -0.48082 | 0.003074 |
| Bmp4 | 1185.836 | -0.37235 | 0.005012 |
| Cped1 | 160.4925 | -0.46243 | 0.006794 |
| Pkp2 | 372.869 | -0.42499 | 0.0079 |
| Spry2 | 240.8192 | -0.39684 | 0.008096 |
| Cyp26b1 | 87.70355 | -0.44219 | 0.008551 |
| Colec12 | 1026.816 | -0.28519 | 0.008597 |
| Fgf10 | 354.8473 | -0.39142 | 0.010242 |
| Prrx1 | 3643.154 | -0.39808 | 0.011005 |
| Tbx18 | 27.1492 | -0.32778 | 0.014458 |
| Evx1 | 328.5051 | -0.43439 | 0.01568 |
| Tbx4 | 1192.83 | -0.42259 | 0.016364 |
| Sall1 | 2057.016 | -0.37439 | 0.019331 |
| Cdh11 | 1604.005 | -0.34045 | 0.01989 |
| Greb1 | 2419.042 | -0.38399 | 0.023128 |
| Nuak1 | 466.6673 | -0.31948 | 0.025007 |
| Lin28b | 3307.493 | -0.35187 | 0.025311 |
| Mtx2 | 1853.379 | -0.18944 | 0.025584 |
| Irs2 | 830.8308 | -0.29128 | 0.025584 |
| Sema5b | 68.64666 | -0.4114 | 0.026013 |
| Tshz2 | 448.8561 | -0.28508 | 0.028228 |
| Trib2 | 287.5969 | -0.33963 | 0.0287 |
| Gnal | 254.3936 | -0.39894 | 0.032295 |
| Lrp12 | 597.8003 | -0.22429 | 0.032295 |
| Slc6a6 | 2954.274 | -0.19421 | 0.034835 |
| Psat1 | 4094.725 | -0.20385 | 0.040011 |
| Mecom | 828.1278 | -0.38147 | 0.041063 |

*Mouse + chick directly activated*

| **Gene** | **baseMean** | **log2FoldChange** | **padj** |
|---|---|---|---|
| Tle4 | 1350.695 | -0.74253 | 1.79E-12 |
| Unc5c | 1016.879 | -0.85703 | 2.01E-12 |
| Mab21l1 | 20.67839 | -0.54778 | 1.61E-06 |
| Fmn1 | 151.2113 | -0.60699 | 1.43E-05 |
| Pcdh7 | 1213.357 | -0.47173 | 4.30E-05 |
| Wnt5a | 1521.81 | -0.48712 | 5.38E-05 |

| | | | |
|---|---|---|---|
| Cnksr2 | 289.8119 | -0.50914 | 0.000141 |
| Prep | 3440.159 | -0.22422 | 0.001011 |
| Mdfic | 184.2354 | -0.45413 | 0.001946 |
| Fgf20 | 28.5458 | -0.41272 | 0.002662 |
| Gja1 | 4624.649 | -0.3343 | 0.002943 |
| Pdzd2 | 82.45487 | -0.48082 | 0.003074 |
| Bmp4 | 1185.836 | -0.37235 | 0.005012 |
| Cyp26b1 | 87.70355 | -0.44219 | 0.008551 |
| Prrx1 | 3643.154 | -0.39808 | 0.011005 |
| Evx1 | 328.5051 | -0.43439 | 0.01568 |
| Tbx4 | 1192.83 | -0.42259 | 0.016364 |
| Greb1 | 2419.042 | -0.38399 | 0.023128 |
| Nuak1 | 466.6673 | -0.31948 | 0.025007 |
| Lin28b | 3307.493 | -0.35187 | 0.025311 |
| Mtx2 | 1853.379 | -0.18944 | 0.025584 |
| Sema5b | 68.64666 | -0.4114 | 0.026013 |
| Gnal | 254.3936 | -0.39894 | 0.032295 |
| Psat1 | 4094.725 | -0.20385 | 0.040011 |
| Mecom | 828.1278 | -0.38147 | 0.041063 |

*Mouse directly repressed*

| **Gene** | **baseMean** | **log2FoldChange** | **padj** |
|---|---|---|---|
| Gas1 | 6691.9331 | 0.5029086 | 5.95E-06 |
| Irx3 | 2245.4346 | 0.4576792 | 5.39E-05 |
| Slitrk6 | 450.96935 | 0.5811071 | 0.0001083 |
| Ppp2r2b | 422.89031 | 0.5390467 | 0.0003044 |
| Bcl9l | 1419.8831 | 0.3060205 | 0.0003451 |
| Col3a1 | 9183.7032 | 0.435396 | 0.0005029 |
| Hand1 | 1197.0569 | 0.4722174 | 0.0010653 |
| Mpped2 | 474.07638 | 0.4351891 | 0.0026407 |
| Pth1r | 987.13641 | 0.3654156 | 0.0032624 |
| Nrp2 | 3142.5835 | 0.2886393 | 0.0095812 |
| Rarb | 774.05559 | 0.2900257 | 0.0115308 |
| Ptx3 | 105.29977 | 0.4272917 | 0.0156797 |
| Fam107a | 68.079074 | 0.4152607 | 0.025311 |
| Pdpn | 430.7314 | 0.3320047 | 0.0339629 |
| Alx4 | 3559.8358 | 0.3026721 | 0.0392341 |

*Chick directly repressed*

| Gene | baseMean | log2FoldChange | padj |
|---|---|---|---|
| Wnt2 | 922.74079 | 0.6255888 | 3.77E-07 |
| Gas1 | 6691.9331 | 0.5029086 | 5.95E-06 |
| Irx3 | 2245.4346 | 0.4576792 | 5.39E-05 |
| Fgfr3 | 559.47117 | 0.4947888 | 0.0001664 |
| Bcl9l | 1419.8831 | 0.3060205 | 0.0003451 |
| Col3a1 | 9183.7032 | 0.435396 | 0.0005029 |
| Ndnf | 781.70513 | 0.4359509 | 0.0008598 |
| Hand1 | 1197.0569 | 0.4722174 | 0.0010653 |
| Cttnbp2 | 267.38575 | 0.4602296 | 0.001114 |
| Adcy5 | 67.541513 | 0.5121404 | 0.0013214 |
| Vegfc | 746.61595 | 0.336016 | 0.0022432 |
| Mpped2 | 474.07638 | 0.4351891 | 0.0026407 |
| Ism1 | 200.24156 | 0.4399742 | 0.0026407 |
| Lrig3 | 794.69575 | 0.3881385 | 0.0026617 |
| Bnc1 | 824.06282 | 0.4081131 | 0.0036949 |
| Cdon | 1656.0522 | 0.4111166 | 0.0049524 |
| Snai2 | 610.00278 | 0.378478 | 0.0076614 |
| Steap1 | 271.3553 | 0.4607108 | 0.0078998 |
| Nrp2 | 3142.5835 | 0.2886393 | 0.0095812 |
| Rarb | 774.05559 | 0.2900257 | 0.0115308 |
| Col5a2 | 2536.139 | 0.2689153 | 0.0133836 |
| Nkd1 | 1420.6578 | 0.2358776 | 0.0153284 |
| S1pr3 | 1091.5754 | 0.3509239 | 0.015864 |
| Irx5 | 772.75172 | 0.3399667 | 0.015941 |
| Aff2 | 247.13283 | 0.4268379 | 0.0193694 |
| Irx4 | 50.204662 | 0.424897 | 0.0208057 |
| Alx4 | 3559.8358 | 0.3026721 | 0.0392341 |
| Fgfr2 | 2304.1071 | 0.2773843 | 0.0400114 |
| Abi3bp | 38.187533 | 0.3906938 | 0.0405838 |
| Pck2 | 322.82624 | 0.3356901 | 0.0431221 |
| Runx1t1 | 564.50644 | 0.3539507 | 0.0444306 |

*Mouse + chick directly repressed*

| Gene | baseMean | log2FoldChange | padj |
|---|---|---|---|
| Gas1 | 6691.9331 | 0.5029086 | 5.95E-06 |
| Irx3 | 2245.4346 | 0.4576792 | 5.39E-05 |
| Bcl9l | 1419.8831 | 0.3060205 | 0.0003451 |
| Col3a1 | 9183.7032 | 0.435396 | 0.0005029 |
| Hand1 | 1197.0569 | 0.4722174 | 0.0010653 |
| Mpped2 | 474.07638 | 0.4351891 | 0.0026407 |
| Nrp2 | 3142.5835 | 0.2886393 | 0.0095812 |
| Rarb | 774.05559 | 0.2900257 | 0.0115308 |
| Alx4 | 3559.8358 | 0.3026721 | 0.0392341 |

**Table S2.4.** Shared putative direct targets in the hindlimb field and genital tubercle.

*Mouse directly activated*

| Gene | shared peak(s)? |
|---|---|
| Unc5c | Y |
| Tle4 | Y |
| Mab21l1 | Y |
| Wnt5a | Y |
| Pdzd2 | Y |
| Pcdh7 | Y |
| Cyp26b1 | Y |
| Evx1 | Y |
| Tbx4 | Y |
| Sema5b | Y |
| Greb1 | Y |
| Bmp4 | Y |
| Lrrn3 | Y |
| Dsc2 | Y |
| Dpysl3 | Y |
| Fut8 | Y |
| Fmn1 | |
| Cnksr2 | |
| Prrx1 | |
| Lin28b | |
| Gja1 | |
| Spry4 | |

*Chick directly activated*

| Gene | shared peak(s)? |
|---|---|
| Unc5c | Y |
| Tle4 | Y |
| Fmn1 | Y |
| Mab21l1 | Y |
| Cnksr2 | Y |
| Wnt5a | Y |
| Pdzd2 | Y |
| Pcdh7 | Y |
| Cyp26b1 | Y |
| Evx1 | Y |
| Tbx4 | Y |
| Sema5b | Y |
| Greb1 | Y |
| Bmp4 | Y |
| Gja1 | Y |
| Prdm1 | Y |
| Dusp6 | Y |
| Etv5 | Y |
| Tril | Y |
| Apcdd1 | Y |
| Zhx2 | Y |
| Tbx18 | Y |
| Irs2 | Y |
| Gnal | |
| Prrx1 | |
| Lin28b | |
| Chst15 | |
| Rai14 | |

*Conserved directly activated*

| Gene | shared peak(s)? |
|---|---|
| Unc5c | Y |
| Mab21l1 | Y |
| Evx1 | Y |
| Tbx4 | Y |
| Bmp4 | Y |

*Mouse directly repressed*

| Gene | shared peak(s)? |
|---|---|
| Slitrk6 | Y |
| Ppp2r2b | Y |
| Mpped2 | Y |
| Rarb | Y |
| Gas1 | |
| Irx3 | |
| Col3a1 | |

*Chick directly repressed*

| Gene | shared peak(s)? |
|---|---|
| Gas1 | Y |
| Irx3 | Y |
| Rarb | Y |
| Abi3bp | Y |
| Aff2 | Y |
| Fgfr2 | Y |
| Ism1 | Y |
| Lrig3 | Y |
| Pck2 | Y |
| S1pr3 | Y |
| Snai2 | Y |
| Wnt2 | Y |
| Col3a1 | |
| Mpped2 | |

*Conserved directly repressed*

| Gene | shared peak(s)? |
|---|---|
| Rarb | Y |
| Irx3 | |

# CHAPTER 3

# ISL1 TARGETS CONSERVED APPENDAGE ENHANCERS DURING DEVELOPMENT OF THE AMNIOTE PHALLUS [2]

[2] Minchey, S.G., et al., to be submitted to *Development*.

Abstract

The development of the genital tubercle (GT) – the embryonic precursor to the penis and clitoris – involves expression of many genes that play a role in limb development. Similarities in early genital budding across amniotes suggest a shared evolutionary origin of the external genitalia from their last common ancestor over 300 million years ago. Prior studies have shown that many limb enhancers are active during phallus development, raising the possibility that limb regulatory elements were coopted during the evolution of the phallus. However, patterns of transcription factor binding during GT development remains largely unexplored. The *Isl1* gene encodes a transcription factor required for normal GT outgrowth. Using a combination of RNA-seq and ChIP-seq, we have identified putative direct targets of ISL1 during mouse GT development. Limb genes are overrepresented among these targets, and many appear to be activated via known limb enhancers. Incorporating ISL1 ChIP-seq analyses from embryonic genitalia of other amniotes such as chick and alligator reveals deeply conserved ISL1 binding events at some of these enhancers, possibly highlighting targets of functional importance that have been maintained for hundreds of millions of years. A notable example of a conserved target is *Tbx4* and its enhancer HLEB, which is known to be an active enhancer during hindlimb development. Another important limb-GT gene, *Hoxd13*, appears to be targeted by ISL1 via multiple conserved enhancers, some of which are GT-specific. Our results suggest that many ISL1 target enhancers in the GT are well-conserved, including both enhancers that may have been coopted from limb development and those that do not show evidence of pleiotropy.

## Introduction

While the external genitalia of mammals, reptiles and birds display extensive diversity, similarities in embryonic development suggest that the phallus evolved once in the last common ancestor of amniotes over 300 million years ago (Gredler, 2016). The genital tubercle (GT) has largely been studied in mouse models (Haller and Ma, 2019; Hashimoto et al., 2019), though analyses of morphology and gene expression have been carried out in the embryonic genitalia of birds (Herrera et al., 2013), squamates (Gredler et al., 2015b; Leal and Cohn, 2014), alligators (Gredler et al., 2015a), and turtles (Larkins and Cohn, 2015). These reveal that some of the key GT genes identified in mice show similar patterns of expression in other amniotes. Furthermore, all amniote phallus development begins with the emergence of paired genital swellings, even in the tuatara which lacks a phallus later in development (Sanger et al., 2015). Although these swellings remain separate and develop into paired hemipenes or hemiclitores in squamates, in other groups they merge at the midline to form the GT.

The mouse GT forms by day E11.5 and undergoes an ambisexual phase of development. The androgen receptor shows a similar pattern of expression between males and females at E13.5, but by E15.5 AR shows clear sexual dimorphism, localizing to the nuclei in males to regulate gene expression (Zheng et al., 2015). The hormone-independent ambisexual phase involves the expression of many genes known for their involvement in limb development (Cohn, 2011). For example, *Shh* signaling is required in mice for limb development as well as GT initiation and outgrowth (Haraguchi et al., 2001; Lin et al., 2009; Perriton et al., 2002; Seifert et al., 2009b). Mice that are homozygous double knockouts for *Hoxa13* and *Hoxd13* fail to develop digits or the GT (Kondo et al., 1997; Warot et al., 1997), while human mutations in each gene are associated with limb and penile defects (Roux et al., 2019; Tüzel et al., 2007).

Lonfat et al. (2014) found that not only are these *Hox13* genes expressed in the limbs and GT, but that there are shared limb-genital enhancers active in both appendages. ChIP-seq data from H3K27ac – a histone modification associated with open chromatin – can reveal active promoters and enhancers in a tissue (Creyghton et al., 2010). Infante et al. (2015) used this to compare enhancer activity in mouse limb and genital buds on a genome-wide scale. Enhancers active in the GT tend to occur near genes involved in limb development, and many are active in the limbs and GT but not in other tissues. This includes HLEB – a known hindlimb enhancer of *Tbx4* which is well-conserved across amniotes. HLEB drives expression in the GT, and its deletion in mice results in lower *Tbx4* expression and altered genital morphology.

The *Isl1* gene encodes a LIM-homeodomain transcription factor (TF) that is known to play a role in the GT as well as in limb development. In the context of limbs, *Isl1* is expressed transiently and plays a crucial role during the initiation of hindlimb budding, but is not found in the forelimb (Itou et al., 2012; Kawakami et al., 2011; Narkis et al., 2012; Shiga et al., 2002). By contrast, *Isl1* is expressed for a much longer period of time in the mouse GT, starting in the cloacal region prior to genital budding and persisting through the ambisexual phase of GT development (Ching et al., 2018; Draaken et al., 2015). Zhang et al. (2017) identified *Isl1* as a major susceptibility gene for bladder exstrophy in humans, finding that conditional deletion of *Isl1* using a *Hoxa3*-Cre driver in mouse results in urogenital defects that included GT hypoplasia. Ching et al. (2018) conditionally deleted *Isl1* using a *Tbx4*-Cre driver, also resulting in GT hypoplasia. They found that ISL1 may be upstream of *Bmp4*, *Wnt5a*, and *Fgf10*, noting ISL1 binding events near these genes and reduced expression following *Isl1* deletion. Tschopp et al. (2014) identified *Isl1* as a marker of genital budding in lizards and chicks, suggesting conservation of this expression domain among amniotes.

Here we explore the role of *Isl1* during amniote embryonic phallus development using conditional deletion, transcriptomic profiling, and identification of ISL1 binding loci genome-wide. We generate conditional knockouts (cKOs) of *Isl1*, resulting in severe deficiency of GT development. Using RNA-seq we identify many genes downstream of *Isl1* that display altered transcription levels following *Isl1* deletion. ChIP-seq allows us to identify which of these genes are putative direct targets of ISL1, and extending this across different species of amniotes reveals enhancers that appear conserved from their last common ancestor. Many targets are known appendage genes that are important for GT development. Some are bound via known limb enhancers that may have been coopted during the evolution of the amniote phallus, and we examine an intriguing example of this in HLEB. Other conserved phallus enhancers may have evolved to regulate these genes in a tissue-specific manner.

## Results

### *Isl1 cKOs show severe GT defects and underexpression of appendage genes*

We used the *Hoxb6*-Cre allele (Lowe et al., 2000) to conditionally delete *Isl1* in the genitalia. While external genital budding still occurs in mutants, they show extreme GT hypoplasia (Fig. 3.1A). This is consistent with previously reported *Isl1* deletions using other Cre drivers (Ching et al., 2018; Zhang et al., 2017), although the *Hoxb6*-Cre driver seems to produce more severe hypoplasia than prior mutants. The GT normally forms from paired genital swellings by E11.5, but in these mutants the paired swellings are still apparent at E12.5, suggesting a delay in GT formation.

We used RNA-seq to find differences in gene expression between the genitalia of mutant and control mice at E12.5. We identified 786 underexpressed genes and 687 overexpressed genes

and used DAVID to perform gene ontology analysis (Huang et al., 2009). Genes involved in embryonic limb morphogenesis are enriched among the underexpressed genes, but not the overexpressed group (Fig. 3.1B). Other sets of genes underexpressed in the absence of *Isl1* are genes involved in the cell cycle/cell division, and genes involved in WNT-signaling. There are also sets of underexpressed genes involved in transcriptional regulation, indicating that many transcription factors (TFs) are directly or indirectly activated by ISL1.

We might expect a large portion of these differentially expressed genes to be indirect effects because of the wide gap in time between when *Isl1* was conditionally deleted and when we observed the downstream transcriptomic impact. This corresponds with the substantial phenotypic divergence at E12.5 between the mutant and control genitalia. The Jax Gene Expression Database lists 173 genes whose expression have been identified in Theiler stage 20 GTs (close to E12.5). 49 of these genes are underexpressed in our data, 21 are overexpressed, and the remaining 103 (~60%) are not differentially expressed. Thus, while there are many genes that are differentially expressed, this does not appear to be merely the result of GT genes as a whole being underexpressed as a result of delayed GT formation.

Among the underexpressed genes are *Bmp4* and *Wnt5a*, confirming what was found in E14.5 GTs of *Tbx4*-Cre *Isl1* cKOs (Ching et al., 2018). This study also found an underexpression of *Fgf10* in these mutants, although it is not significantly altered in our data. *Msx1* and *Msx2* are downstream effectors of BMP signaling (Marazzi et al., 1997), and both are underexpressed in our *Isl1* cKO GTs. *Sp6* and *Sp8* are expressed in the urethral epithelium downstream of WNT/β-catenin signaling, and *Sp8* was shown to act upstream of *Fgf8* and *Fgf9* (Lin et al., 2013). Our data show that *Sp6/7/8/9* and *Fgf9* are underexpressed, while *Fgf8* expression levels are too low to detect significant differences. Furthermore, genes in the hedgehog pathway are

underexpressed, which is considered a crucial part of the urethral plate organizing center (Haraguchi et al., 2001; Lin et al., 2009; Perriton et al., 2002; Seifert et al., 2009b). These underexpressed genes include *Shh* itself, *Ptch1/2*, and *Gli1/3*.

The distal *Hox* genes *Hoxa13* and *Hoxd13* play important roles in GT development (Kondo et al., 1997; Warot et al., 1997). While *Hoxa13* is not differentially expressed, *Hoxd13*, as well as *Hoxd11*, *Hoxd12*, and *Evx2* are underexpressed, all of which are at the 5' end of the *Hoxd* cluster. Underexpressed genes of the T-box family include *Tbx2*, *Tbx3*, *Tbx4*, and *Tbx5*. While these genes were previously shown to be expressed in the mouse GT (Douglas et al., 2012), their function has not been well-explored, though *Tbx3* mutations in humans are associated with micropenis as part of Ulnar-Mammary Syndrome (Sasaki et al., 2002). *Dlx5/6* double knockouts exhibit defects in urethral formation (Suzuki et al., 2008), and these are among the most significantly underexpressed genes in our data.

*ISL1 binds near limb genes in the GT and targets conserved enhancers*

To find which genes are potential direct targets of ISL1 in the GT, we performed ChIP-seq on ISL1 in the embryonic genitalia of mouse, chick, American alligator, red-eared slider turtle, and the brown anole lizard. In the mouse, we identify 14,600 reproducible binding loci, which mostly occur >50 kb away from transcription start sites (Fig. 3.2A), consistent with binding to distal *cis*-regulatory elements. We used GREAT to associate these peaks with the nearest genes and perform gene ontology analysis (McLean et al., 2010). This shows that ISL1 tends to bind near genes expressed in the embryonic mouse limb and penis (Fig. 3.2B, top) and genes whose orthologs have been associated with digit defects when mutated in humans (Fig. 3.2B, bottom).

Amândio et al. (2020) performed CUT-and-RUN on HOXA13 in the mouse GT at E13.5 – one day in development after our ChIP-seq data. We reanalyzed these data, finding that roughly 38% of the ISL1 peaks overlap with a HOXA13 peak, and *de novo* motif analysis using HOMER (Heinz et al., 2010) finds that the top motif in these ISL1 sites resembles that of HOXA13 (Fig. 3.2C,D top). Motifs resembling ISL1 are nevertheless among the top 5 *de novo* motifs for the ISL1+HOXA13 peaks. With the remaining ~62% of ISL1 peaks that do not overlap a HOXA13 peak, the top *de novo* motif most resembles ISL1 (Fig. 3.2C,D bottom). Both the ISL1+HOXA13 and the ISL1-only peaks also show enrichment of motifs matching TCF/LEF – TFs in the Wnt-signaling pathway – as well as TWIST1.

Infante et al. (2015) identified enhancers that show limb-specific, GT-specific, and limb-GT activity using H3K27ac ChIP-seq, filtering out enhancers that are not appendage-specific using data from other tissues. Ching et al. (2018) performed ISL1 ChIP-seq in the GT at E14.5, finding that it binds many of the limb-GT and GT-specific enhancers, but much less frequently the limb-specific enhancers. Our ChIP-seq data from earlier in GT development (E12.5) are consistent with these results (Fig S3.1A). Using vertebrate phastCons scores, we visualized the degree of conservation around ISL1 summits within the limb-GT and GT-specific enhancers (Fig. S3.1B). Comparing the central 100 bp around these summits, we find a significantly higher mean conservation score within the limb-GT enhancers compared to the GT-specific ones.

To systematically explore ISL1 binding outside of mammals, we used the ISL1 GT data from chick and alligator. Like most birds, the chicken ultimately lacks an external phallus, undergoing GT reduction during embryonic development (Herrera et al., 2013). We performed ChIP-seq before this occurs to avoid possible differences in the *cis*-regulatory landscape associated with this change. In addition, to reinforce our confidence in ISL1 binding loci, we

used the alligator which undergoes normal GT development (Gredler et al., 2015a). Crocodilians are the most closely-related living outgroup to avians, both being members of the archosaur clade. To identify ISL1 binding regions shared among these archosaurs, we used the UCSC LiftOver tool to convert alligator ISL1 binding loci to the chicken genome. This resulted in 4,897 binding loci that are conserved between the two species (Fig. 3.3A).

As with the mouse, we find enrichment of motifs resembling those of HOXA13, ISL1/2, TWIST1, and TCF7 (Fig 3.3B). The frequency of a HOX13 motif in both the mouse and archosaur peaks raises questions regarding how ISL1 interacts with these sites. We used the MEME-suite program CentriMo (Bailey and Machanick, 2012) to look at the distribution of known motifs from a human and mouse database. In both the mouse and archosaur peaks, the top two centrally-enriched motifs are the human HOXB13 and ISL2 motifs, which are similar to the top motifs we find using HOMER (Fig. S3.2). 62% of the mouse peaks and 60% of the archosaur peaks have one of these two motifs in their central region, but only 15% and 13% of each peak set contain both. This suggests there are many ISL1 peaks that have a central HOX13 motif, but not a central ISL1 motif.

Of the conserved archosaur peaks, 1,790 were converted to the mouse genome, allowing us to analyze them using GREAT. As with the mouse, these peaks show a distal *cis*-regulatory element pattern (Fig. 3.3C). They tend to occur near genes involved in limb development and genes associated with abnormal genital morphology in humans (Fig. 3.3E). We see a similar distribution among the subset of peaks that are conserved across mouse, chick and alligator, which we consider conserved amniote peaks (Fig. 3.3D). These peaks are associated with genes involved in mouse urethral development and genes associated with abnormal scrotum morphology in humans (Fig. 3.3F).

*Conserved directly activated targets of ISL1 in the GT*

To determine which genes are potential direct targets of ISL1, we used the program Cistrome-GO (Li et al., 2019) to integrate the RNA-seq data with the conserved amniote ChIP-seq peaks (Fig. 3.4A). We used only underexpressed genes because they show a more appendage-like signal (Fig. 3.1B) and because they have a stronger association with ChIP-seq peaks than do the overexpressed genes (Fig. 3.4B), consistent with ISL1 primarily acting as an activator in the GT. Putative conserved, directly activated targets are enriched for genes involved in appendage/limb morphogenesis (Fig. 3.4C) and genes involved in the cell-cycle and WNT-signaling (Fig. 3.4D). Among the top-20 candidate genes are *Msx1/2*, *Tbx2/4*, *Isl2*, *Wnt5a*, and *Hoxd13* (Table S3.3).

Given that ISL1 appears to activate known appendage genes in the GT and that many enhancers are known to be shared between the limbs and GT (Infante et al., 2015), we looked for evidence that ISL1 may activate these genes via known limb enhancers. Andrey et al. (2017) characterized enhancers active in embryonic mouse forelimbs and hindlimbs at three developmental stages using a combination of histone ChIP-seq and Capture-C. The Capture-C data allowed them to associate these enhancers with the genes they likely regulate. We find 250 mouse ChIP-seq peaks that occur within these limb enhancers, and 31 of these are conserved amniote peaks (Fig. S3.3). Additionally, we find 42 conserved peaks that occur on putative enhancers active only in the GT identified by Infante et al. (2015) (Fig S3.4).

H3K27ac ChIP-seq signal from the GT and limbs of mouse, chick, and alligator supports the idea that these are shared limb-GT enhancers (Fig. S3.3A-C). This can be contrasted with conserved ISL1 peaks that occur on the GT-specific enhancers (Fig S3.4A-C). While some of these enhancers show limb activity in the chick and alligator, the overall limb activity is much

lower than for the known limb enhancers, suggesting that many conserved enhancers have maintained their tissue expression domains across these species. Combining ISL1 binding to known limb enhancers with our differential expression data allows us to identify strong candidate direct targets of ISL1, assuming the enhancers associate with the same genes in the limbs and GT. Among the genes that are potentially directly regulated via known limb enhancers are *Wnt5a, Hoxd13, Tbx2* and *Tbx4* (Fig. S3.3D).

*ISL1 activates Tbx4 via the limb-GT enhancer HLEB*

The limb enhancer associated with *Tbx4* identified by this analysis is HLEB, which was previously shown to have both hindlimb and GT activity (Infante et al., 2015; Menke et al., 2008). In addition to this ISL1 binding locus found in the mouse, chick, and alligator GT (Fig. 3.5A-C), we also find it in the turtle GT and the lizard hemiphallus (Fig. 3.5D-E). HOXA13 also appears to bind to this enhancer in the mouse GT. We generated a multi-species alignment of the core orthologous sequence of HLEB, including mammals, turtles, crocodilians, lizards and snakes. This reveals three highly conserved ISL1 motifs that may be specific ISL1 binding sites, as well as a HOX13-like motif (Fig. S3.5).

While the HOX13 motif remains intact in snakes, all 3 ISL1 motifs are mutated in at least some species of snakes. The boa constrictor and king cobra HLEB sequence was previously shown to exhibit reduced though not absent GT activity as compared to the mouse and *Anolis* HLEB (Infante et al., 2015). Both of these show a mutation in one of the ISL1 motifs that appears ancestral to all snakes (Fig. S3.5, left), whereas the king cobra (*Ophiophagus hannah*) has an additional mutation that appears in more advanced snakes (right). We tested LacZ constructs containing <1 kb of the wildtype king cobra HLEB sequence and one in which we

restored the two bases associated with the ISL1 sites (Fig. S3.6). We also tested a construct with 26 bases restored that are conserved among our sample of limbed amniotes (excluding birds and tuatara) but mutated in king cobra. The two ISL1 sites are a subset of these 26 bases.

The wildtype construct is able to recapitulate the ventral GT activity previously observed with the cobra HLEB (Infante et al., 2015) in 3 of the 12 embryos (Fig. 3.6A, S3.7A-B, Table S3.4). Another 2 embryos show strong expression throughout the GT, but these embryos also show widespread non-specific activity (Fig. S3.7C-D). Restoration of two ISL1 sites does not appear sufficient to rescue the activity of HLEB previously seen by Infante and colleagues with the mouse or lizard HLEB (Fig. 3.6B). Restoring 26 highly conserved bases, however, does appear to be associated with an expansion of HLEB's GT domain (Fig. 3.6C). Unlike the other two constructs, none of these embryos show a ventrally biased expression domain, and restoration of these bases also appears to result in a rescue of the posterior hindlimb domain that Infante and coathors saw using the mouse HLEB. 4 in 10 of these embryos show both strong staining throughout the GT and posterior hindlimb activity.

*ISL1 shows conserved binding at GT-specific enhancers of Hoxd13*

*Hoxd13* is another target gene showing conserved binding at a known limb enhancer (Fig. S3.3D). In addition to being annotated as a limb enhancer (Andrey et al., 2017), this locus is within a larger region shown to have GT activity by Amândio et al. (2020), which they labeled IIIE. We see evidence of limb and GT activity in this conserved ISL1 binding locus from our H3K27ac data (Fig. 3.7). These researchers also identified another region which shows GT activity called IVE, and here we find another conserved ISL1 binding locus showing no evidence of limb activity. A third conserved ISL1 peak occurs within the known enhancer GT2 which was

previously shown to have GT-specific activity (Amândio et al., 2020; Lonfat et al., 2014). While the IIIE and IVE peaks do not show consistent H3K27ac enrichment in chick and alligator, the GT2 locus exhibits robust activity in the GT of all three species though not in the limbs.

## Discussion

TFs play different roles depending on cellular context, mediated in part by differences in the activity of *cis*-regulatory elements. The cell-specific functions of enhancers are controlled by pioneer TFs that have the ability to open chromatin and allow binding of other TFs. In cardiac progenitor cells, ISL1 appears to play a pioneering role (Gao et al., 2019), although this function remains unexplored in appendage development. Combinatorial activity of TFs provides another layer of control over cell fate. For example, *Isl1* expression combined with *Lhx3* and *Ngn2* leads mouse embryonic stem cells to differentiate into spinal motor neurons, while replacing *Ngn2* with *Phox2a* leads to a cranial motor neuron fate (Mazzoni et al., 2013). Thus exploring multiple TFs in a developmental context is necessary to reveal how they are interconnected. To date there have been few high-throughput explorations of TF activity in the GT.

Though little is known about the gene regulatory network controlling GT development, our results show that ISL1 plays a crucial role upstream of many known GT genes. Although *Hoxb6*-Cre deletion of *Isl1* does not cause complete GT agenesis, it does result in more severe ablation than previously reported using the *Tbx4*-Cre allele (Ching et al., 2018). The *Hoxb6*-Cre allele is active broadly in the posterior mesenchyme at stages prior to GT initiation (Lowe et al., 2000). Thus it may cause a more thorough deletion of *Isl1* throughout the early pericloacal mesenchyme than the *Tbx4*-Cre allele. Low level *Isl1* expression was previously reported in the

urethral plate endoderm (Ching et al., 2018; Su et al., 2019), so there may also be differences in the deletion of *Isl1* in this domain.

*Isl1* appears to be upstream of genes and signaling cascades known to be crucial for GT development, including genes of the Hedgehog, WNT, and BMP pathways. Disruption of these pathways in the GT mesenchyme produces similarly severe phenotypes of the GT (Kajioka et al., 2020; Lin et al., 2008, 2009), and thus they could partly explain why *Isl1* is indispensable for normal development. Given the complex downstream effects of removing genes at the higher levels of a gene regulator network, we expect many of the differentially expressed genes to be indirect rather than direct targets of ISL1. This is especially the case given the extreme phenotypic divergence that occurs between the time of conditional deletion and testing of gene expression. Deletion at later staged using another Cre allele might reduce non-primary effects.

Our ISL1 ChIP-seq data allow us to focus on genes that are more likely to be direct targets of ISL1. Incorporating data from different species of amniotes identifies conserved enhancer loci that are targeted across distantly related amniotes. The prevalence of these conserved GT enhancers is consistent with the current view that the amniote phallus evolved once in their last common ancestor. While focusing on conserved loci necessarily excludes many lineage-specific elements, it also may be expected to enrich for functional elements more than the nonfunctional loci that are frequently detected by ChIP-seq (Sakabe and Nobrega, 2013). Enhancers conserved across these amniote species are also more likely to be conserved in humans and thus may be important for human GT development.

Focusing on conserved ISL1 binding loci also excludes lineage-specific functions that could be of interest, and the *Hoxd* locus provides a possible example. While the conserved ISL1 peaks at the IIIE and IVE enhancers were found in the hemiphallus of the brown anole, the GT2

peak appears completely absent (Fig. S3.8A). Looking outside of this region to other annotated *Hoxd* enhancers, however, we find a possible change in enhancer activity between species at the known digit enhancer Island-I (Fig. S3.8B). While no peak was found in the mouse, there are peaks in the chick, alligator, and lizard Island-I locus. Furthermore, we find an ISL1 motif that appears unique to lizards within a specific region of Island-I that exhibits possible increased ISL1 binding in the lizard (Fig. S3.8C). If this represents a true difference it is notable because the lizard Island-I sequence has been shown to drive some GT expression in mouse embryos, while the mouse version shows limb-specific activity (Guerreiro et al., 2016).

One mechanism by which enhancers can evolve is cooption, where an ancestral *cis*-regulatory sequence is repurposed in a different developmental context (Glassford et al., 2015; Xin et al., 2020). A subset of conserved ISL1 targets are known limb enhancers, consistent with cooption of pre-existing limb enhancers during the evolution of the amniote phallus. These pleiotropic enhancers can combine with TFs expressed in both the limbs and GT, as well as TFs that differ between the two tissues. An example of the former is shown by ISL1 binding to HLEB, an enhancer which is active during early hindlimb development coinciding with *Isl1* expression (Infante et al., 2015; Kawakami et al., 2011).

In mice, *Isl1* appears to play a role in the trunk-to-tail transition, being involved in the terminal differentiation of trunk progenitors (Jurberg et al., 2013). This corresponds to the position of the hindlimbs and cloaca as the major axial landmarks immediately anterior to the tail. The exact position of the cloaca varies across amniotes, but it seems to activate *Isl1* expression in the surrounding mesenchymal tissue across amniotes (Tschopp et al., 2014). If this is an ancestral role among tetrapods, it is possible that the cloacal *Isl1* domain existed prior to the evolution of the amniote phallus, thus allowing it to regulate limb enhancers once they were

coopted for phallus development. More work on *Isl1*'s expression domain outside of amniotes (e.g. amphibians) may shed light on this question.

*Isl1* plays a transient role during hindlimb initiation, and thus many of these coopted limb enhancers are likely not regulated by ISL1 during limb development because of incompatible timing. Enhancers active during digit development are a clear example of this, as digit development takes place far after hindlimb initiation. During limb initiation, the *Hoxd* locus is controlled by the T-DOM topologically associating domain, while C-DOM becomes active later during digit development (Rodríguez-Carballo et al., 2020). We see evidence of ISL1 regulating a shared limb-GT enhancer within the C-DOM and thus this enhancer is presumably not active during hindlimb initiation (Fig. 3.7). Although we cannot say to what extent these ISL1 binding loci are functionally important for these enhancers without additional experiments, they illustrate the potential complexities of enhancer pleiotropy.

Our manipulation of the cobra HLEB sequence is one approach for testing the function of particular enhancers. While HLEB was known to be pleiotropic in the hindlimb and genitalia, our experiments probe particular bases within this enhancer, finding a set that appear necessary for its complete activity in both of these contexts. Two ISL1 sites mutated in king cobra do not appear sufficient to account for the loss of activity in the ventral GT or the hindlimb. Restoring these sites along with another 24 bases across a span of 240 bp, however, does rescue this activity. It is possible that these ISL1 sites are necessary but other TF binding sites are also required to drive expression. The 26 bases restored include those that fall on or near potential MEIS, HOX13, and LEF/TCF sites (Fig. S3.6).

Researchers of *Drosophila* enhancers have distinguished between those that exhibit "site pleiotropy" – shared TF binding sites used in different contexts – and pleiotropic enhancers with

distinct TF binding sites used in different tissues (Preger-Ben Noon et al., 2018; Sabarís et al., 2019). Using this dichotomy, the ISL1 binding sites within HLEB along with the other sites we tested may be site pleiotropic. Other shared limb-GT enhancers like those regulating *Hoxd13* may contain non-pleiotropic ISL1 binding sites used during GT development while remaining unused during limb development at stages when *Isl1* is not expressed. Alternatively, there could be binding sites that are site pleiotropic between the limbs and GT, but are occupied by a different TF in the two tissues, e.g. a binding site occupied by ISL1 in the GT may be occupied by another homeodomain TF in the limbs. These different types of pleiotropy are not mutually exclusive, and enhancers may contain a mixture of TF binding sites of each category.

We also identify ISL1 binding loci within GT-only enhancers that are conserved across amniotes. These presumably are derived from their last common ancestor and may have evolved along with the phallus. This fits the more traditional paradigm of enhancers as providing modular control of gene expression and reducing pleiotropic effects of mutations (Sabarís et al., 2019). While we cannot rule out pleiotropic activity of these enhancers in untested stages of limb development or other cell types, we do see less conservation around ISL1 binding sites within GT-specific mouse enhancers (Fig. S3.1) as consistent with expectations (Sabarís et al., 2019). Exploring other TFs and appendage enhancers in the GT could reveal more of its gene regulatory network and what aspects are shared with the limbs.

## Methods

### *Mouse lines*

*Hoxb6*-Cre mice (Lowe et al., 2000) were crossed with *Isl1*+/- mice to generate *Hoxb6*-Cre; *Isl1*+/- mice on a C57BL/6 background. These were then crossed with *Isl1*flox/flox mice to

generate embryos for both phenotypic and transcriptomic analysis. "*Isl1* cKO" refers to the *Hoxb6*-Cre; *Isl1*flox/- genotype, while the control embryos have the *Hoxb6*-Cre; *Isl1*flox/+ genotype.

*RNA-seq*

The GTs of control mice and paired genital swellings of *Isl1* cKO mice were dissected at E12.5. We collected total RNA separately from the genitalia of 5 *Isl1* cKO females and 5 control females using the mirVana miRNA isolation kit (ThermoFisher Scientific), thus allowing for 5 replicates. mRNA libraries were prepared from 500 ng of total RNA using the TruSeq Stranded mRNA Library Prep Kit (Illumina). These were then sequenced at the Georgia Genomics and Bioinformatics Core to generate 50 bp single-end reads. Reads were aligned to the mm10 genome using STAR (v. 2.7.3a) and transcripts were counted using HTSeq (v. 0.13.5). Differentially expressed genes were determined using DESeq2 (Galaxy v. 2.11.40.6).

*Chromatin immunoprecipitation*

GTs were collected from ICR mice at E12.5, chickens (Cobb500 breed) at Hamburger-Hamilton stage 28 (Hamburger and Hamilton, 1992), American alligators at Ferguson stage 15-16 (Ferguson 1985), red-eared slider turtles at Greenbaum stage 16-17 (Greenbaum, 2002), and from the hemiphallus of Sanger stage 7-9 brown anoles (Sanger et al., 2008). Each biological replicate consisted of tissue from many individual embryos, and thus the samples were of mixed sex for species with chromosomal sex determination (mouse, chick, and lizard), although these stages are prior to sexual differentiation of the phallus. Forelimbs and hindlimbs were collected from chicks at HH25, turtles at stage 12, and alligators at stage 13. With the exception of the

alligator and turtle ISL1 ChIP, each experiments was performed in duplicate, although we only used one replicate of the lizard ISL1 ChIP as this one was performed using a finer dissection technique and generated more robust signal. Samples were crosslinked in 1% formaldehyde for 20 minutes at room temperature and frozen at -80C for later use.

The tissue samples were suspended in lysis buffer and briefly homogenized using a probe sonifier (Branson 450). They were then placed in a bath sonifier (Diagenode Bioruptor 300) and sheared for a total of 45 cycles of 30 seconds on/30 seconds off. The mouse and chick samples were processed using an older ChIP protocol using PureProteome Protein G magnetic beads (Millipore) incubated with the antibody and ~500 μg of chromatin for ISL1 ChIP as measured by Nanodrop. The alligator, turtle, and lizard samples were processed using a modified protocol that uses Protein G Agarose Columns (Active Motif) and ~50 μg of chromatin for ISL1 ChIP.
For ISL1 ChIP we used 5 μg of anti-ISL1 rabbit monoclonal antibody (Abcam EP4182).

For H3K27ac ChIP we used 3 μg of mouse monoclonal antibody CMA309 (previously from Millipore, subsequently from Cosmo Bio), except for the chick forelimb and hindlimbs where we used 3ug of rabbit polyclonal antibody ab4729 (Abcam). Libraries were prepared using the NEBNext Ultra II Library Prep Kit. Library amplification was done for 15 cycles, with additional cycles done in some cases where the final concentration was too low. Libraries were cleaned up using Sera-Mag SpeedBeads. Libraries were sequenced by the Georgia Genomics and Bioinformatics Core (GGBC) to generate single-end 75 bp reads, except for the alligator libraries which were sequenced by Genewiz to generate paired-end 150 bp reads.

*ChIP-seq data analysis*

The GGBC provided pre-trimmed reads with adapter sequence removed. The alligator paired-end data was trimmed using cutadapt with and only forward reads were used. This was done with the settings –O 1, -m 25 so that even one base of adapter overhang was trimmed, reads adapter sequences used were as follows:

AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC

AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

We used the AQUAS Transcription Factor and Histone ChIP-seq processing pipeline starting from trimmed reads (https://github.com/kundajelab/chipseq_pipeline). We used the pipeline's default bwa aligner to align mouse data to GRCM38/mm10 (UCSC), chick to galGal6 (UCSC), alligator to allMis0.2/allMis1 (UCSC), turtle to CAS_Tse_1.0 (NCBI RefSeq), and lizard to AnoSag2.0. We set the fragment size to 175 (-speak_spp 175 –extsize_macs2 175). For mouse and chick ISL1 ChIP-seq, we performed the full IDR analysis using two replicates with the default SPP peak caller. With the alligator, turtle, and lizard, only one replicate was used, so using MACS2 peaks we set a q-value cutoff of 0.01 and fold-enrichment cutoff of 4. The bed files used in our downstream analyses come from these IDR peaks for mouse and chick and threshold cutoff peaks for the other species.

For H3K27ac ChIP we used the default histone ChIP setting, which calls peaks using MACS2 and performs naïve overlap between the two replicates. For both ISL1 and H3K27ac ChIP, the pipeline uses MACS2 to generate bigWig plots displaying fold-enrichment over the input, which was used in our figures. We used LiftOver on the UCSC test browser (https://genome-test.gi.ucsc.edu/cgi-bin/hgLiftOver) to convert alligator bed coordinates to the chicken genome using default settings. We used the same tool on the main UCSC site to convert from chicken to mouse. UCSC Table Browser was then used to find overlapping peaks between

different datasets (e.g. conserved archosaur peaks, conserved amniote peaks, and peaks that overlap published enhancer datasets).

*Gene ontology analyses*

For RNA-seq data, we used DAVID (v. 6.8). We used underexpressed genes or overexpressed genes with an FDR-adjusted p-value threshold of 0.05. For ChIP-seq data, mm10 peak coordinates were input to GREAT (v. 3.0.0). We used the default "basal plus extension" association rule, which assigns each gene a regulator domain up to 1000 kb unless it reaches another genes proximal domain (5 kb upstream, 1 kb downstream). For the Cistrome-GO analysis, we used the mm10 genome and input the conserved amniote ISL1 ChIP-seq peaks as well as the DESeq2 output file with a half-decay distance of 50 kB. For the precision-recall curve, we set DE gene cutoffs at 0 for logFoldChange and 0.05 for FDR.

*Public data used*

Mouse H3K27ac data from E12.5 GT, E11.5 forelimb and hindlimb from Infante et al. (2015) (accession GSE64055) was reprocessed in mm10 using fastq files and the above ChIP-seq pipeline. We converted coordinates of "limb specific", "limb-gt specific", and "gt specific" enhancers in Supplementary Table S4 of that publication from mm9 to mm10 for use in our analyses. Annotated mouse limb enhancers were acquired from Andrey et al. (2017) Supplemental Table S3. We used the UCSC Table Browser to combine the six limb enhancer sets, merging enhancers that overlap. This allowed us to then find overlapping ISL1 peaks and associate them with genes contacted by those enhancers.

Mouse HOXA13 CUT-and-RUN data from E13.5 GT was used from Amândio et al., (2020) (accession GSE138514). We aligned the fastq reads to mm10 using Bowtie2. We called peaks and generated the signal plot using Genrich in ATAC-seq mode (-j) but without Tn5 adjustment (-D), kept unpaired alignments (-y), excluded mitochondrial reads (-e chrM), and set the cut sites to a size of 100 bp (-d 100). This resulted in 13,037 HOXA13 peaks from this one replicate, which we used to find the 5,577 overlapping ISL1 peaks.

*Motif analysis*

We used HOMER to perform *de novo* motif analysis using the central 100 bp around ISL1 peaks summits, using 100k random genomic sequences as the background (-N 100000), and searched for motifs of length 8 bp (-len 8). For the conserved archosaur peaks, we used the summit coordinates from the chicken peaks and performed the motif analysis in that genome. To find the distribution of motifs around peak summits, we used CentriMo (v. 5.4.1). We used the central 500 bp around the peak summits and searched for known motifs in the "Human and Mouse (HOCOMOCO v11 FULL)" database.

*Heatmap visualization*

For the RNA-seq data, we used the normalized counts output by DEseq2 for differentially expressed genes, sorted the underexpressed genes alphabetically on top and the overexpressed genes alphabetically below. We used the Galaxy tool heatmap2 v3.0.1.0.0, scaling the data by row because of extreme differences in expression levels between genes.

For ChIP-seq/CUT-and-RUN data, we used the bed files centered on peak summits and the bigWig signal files as inputs for computeMatrix (Galaxy v. 3.3.2.0.0), using the scale-regions

setting with windows of either 1kb or 2kb. The bin size was set to 1 bp except for Figs. S3.3 and S3.4 where it was set to 20 bp. We used plotHeatmap (Galaxy v. 3.3.2.0.1) to generate the heatmaps from the matrix files.

*HLEB alignment and transgenic assays*

We generated a multispecies alignment of the core conserved portion of HLEB in Geneious. This alignment consist of 88 species of amniotes and a conserved region spanning <500 bp, part of which is shown in Fig. S3.5. The alignment excluded birds and tuatara because of their unique external genital development, but included other limbed amniotes and snakes. We manually curated the alignment to identify conserved bases that are invariant in the sample of limbed amniotes, and from these we identified bases that are uniquely mutated in king cobra and possibly other snakes. There is one 5 bp deletion in snakes which consists of 3 invariant bases and 2 bases that are largely conserved but variable. We used the more common CGGAA found in most non-snake species to restore this deletion, along with another 21 consensus bases mutated in cobra, including 2 bases on putative ISL1 motifs.

We ordered three different gBlock fragments from IDT containing each versions of the king cobra HLEB sequence. These sequences contained either 863 bp of HLEB sequence – or 869 bp in the case of the fragment with 26 bases restored – as well as flanking homology arms for insertion into an Hsp68-LacZ backbone. From this backbone and insert sequence, we used the NEBuilder® HiFi DNA Assembly Cloning Kit to carry out plasmid assembly and cloning. We used the QIAfilter Plasmid Midi Kit from Qiagen to extract and purify plasmid DNA. Generation of transgenic mice, including plasmid linearization, pronuclear injection, breeding, embryo collection, PCR validation, and LacZ staining were carried out by Cyagen.

Figures

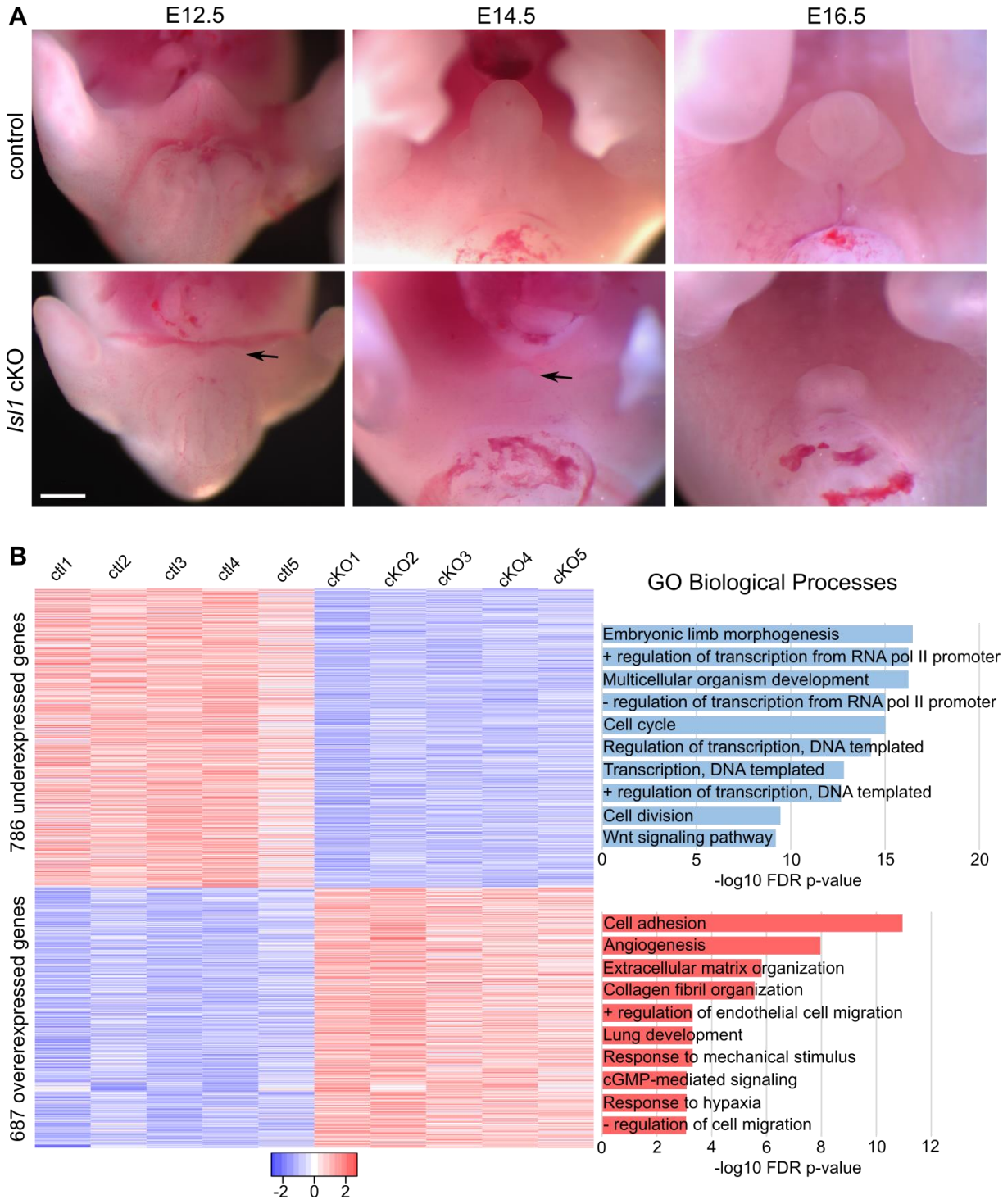
A
E12.5
E14.5
E16.5
control
Isl1 cKO
B
ctl1
ctl2
ctl3
ctl4
ctl5
cKO1
cKO2
cKO3
cKO4
cKO5
786 underexpressed genes
687 overerexpressed genes
-2
0
2
GO Biological Processes
Embryonic limb morphogenesis
+ regulation of transcription from RNA pol II promoter
Multicellular organism development
- regulation of transcription from RNA pol II promoter
Cell cycle
Regulation of transcription, DNA templated
Transcription, DNA templated
+ regulation of transcription, DNA templated
Cell division
Wnt signaling pathway
0
5
10
15
20
-log10 FDR p-value
Cell adhesion
Angiogenesis
Extracellular matrix organization
Collagen fibril organization
+ regulation of endothelial cell migration
Lung development
Response to mechanical stimulus
cGMP-mediated signaling
Response to hypaxia
- regulation of cell migration
0
2
4
6
8
10
12
-log10 FDR p-value

**Figure 3.1.** Phenotypic and transcriptomic effects of *Isl1* deletion on the genital tubercle. **(A)** Light microscopy images of the external genitalia of control (top) and *Isl1* cKO (bottom) mouse embryos from E12.5 to E16.5. Scale bar (bottom left) indicates 500 μm, with all images at the same magnification. **(B)** 786 underexpressed genes (top) and 687 overexpressed genes (bottom) in *Isl1* cKO GTs relative to control GTs. Heatmap is scaled by row to represent relative expression of each gene across 10 samples – 5 control and 5 mutant replicates. Gene ontology analysis shows top 10 "GO Biological Processes" terms from DAVID.

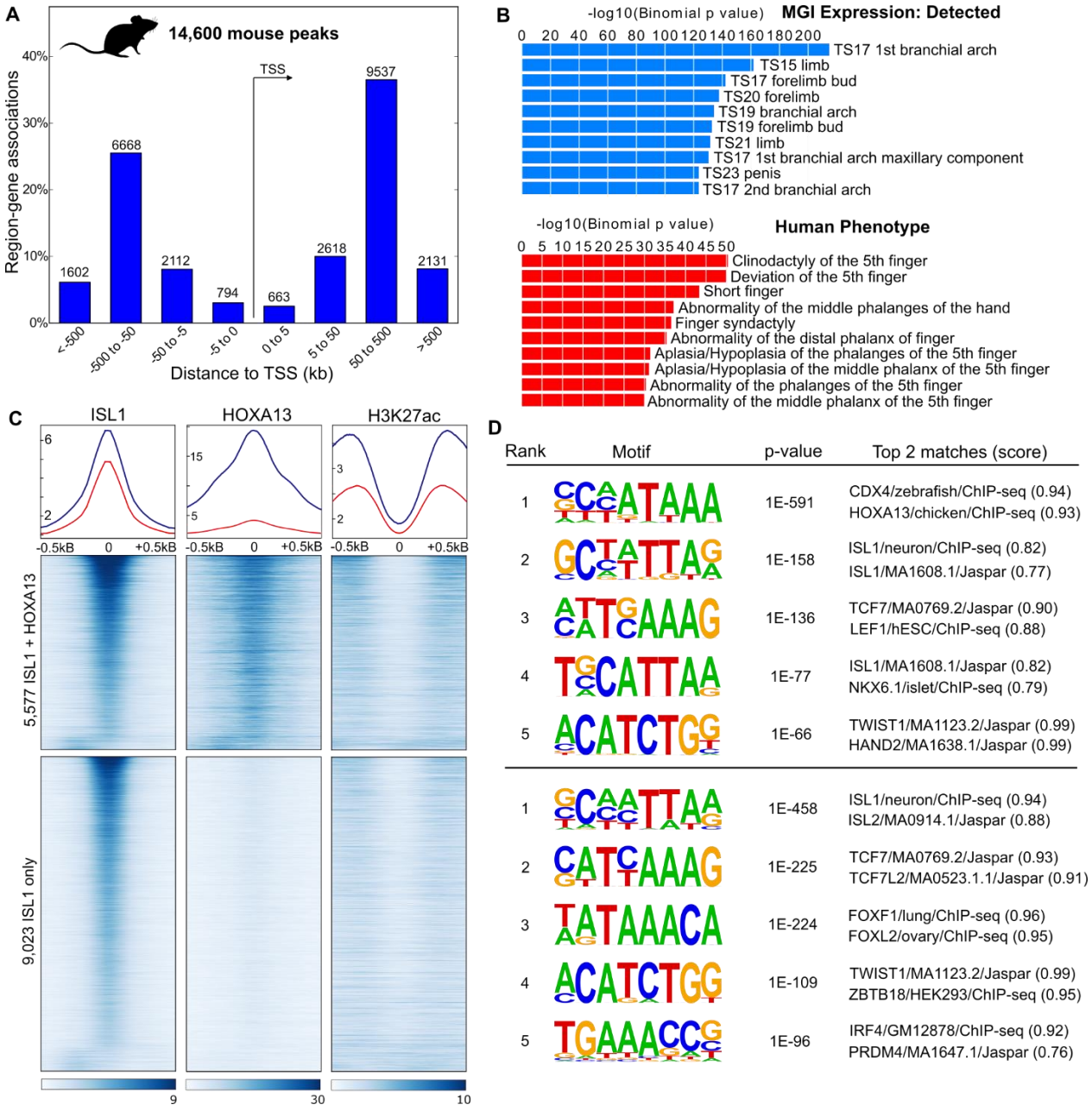


**Figure 3.2.** ISL1 ChIP-seq in the mouse genital tubercle. **(A)** Distribution of ISL1 peaks relative to nearby transcription start sites (TSS) as determined using GREAT. **(B)** GO analysis showing top 10 categories based on mouse gene expression (top) and human phenotypes (bottom). **(C)** ISL1, HOXA13, and H3K27ac signal centered on ISL1 binding loci, split into ISL1+HOXA13 peaks (top, blue line plot) and ISL1 only peaks (bottom, red line). **(D)** *De novo* motif analysis from HOMER showing the top 5 motifs from the central 100 bp of each set of peaks in (C).

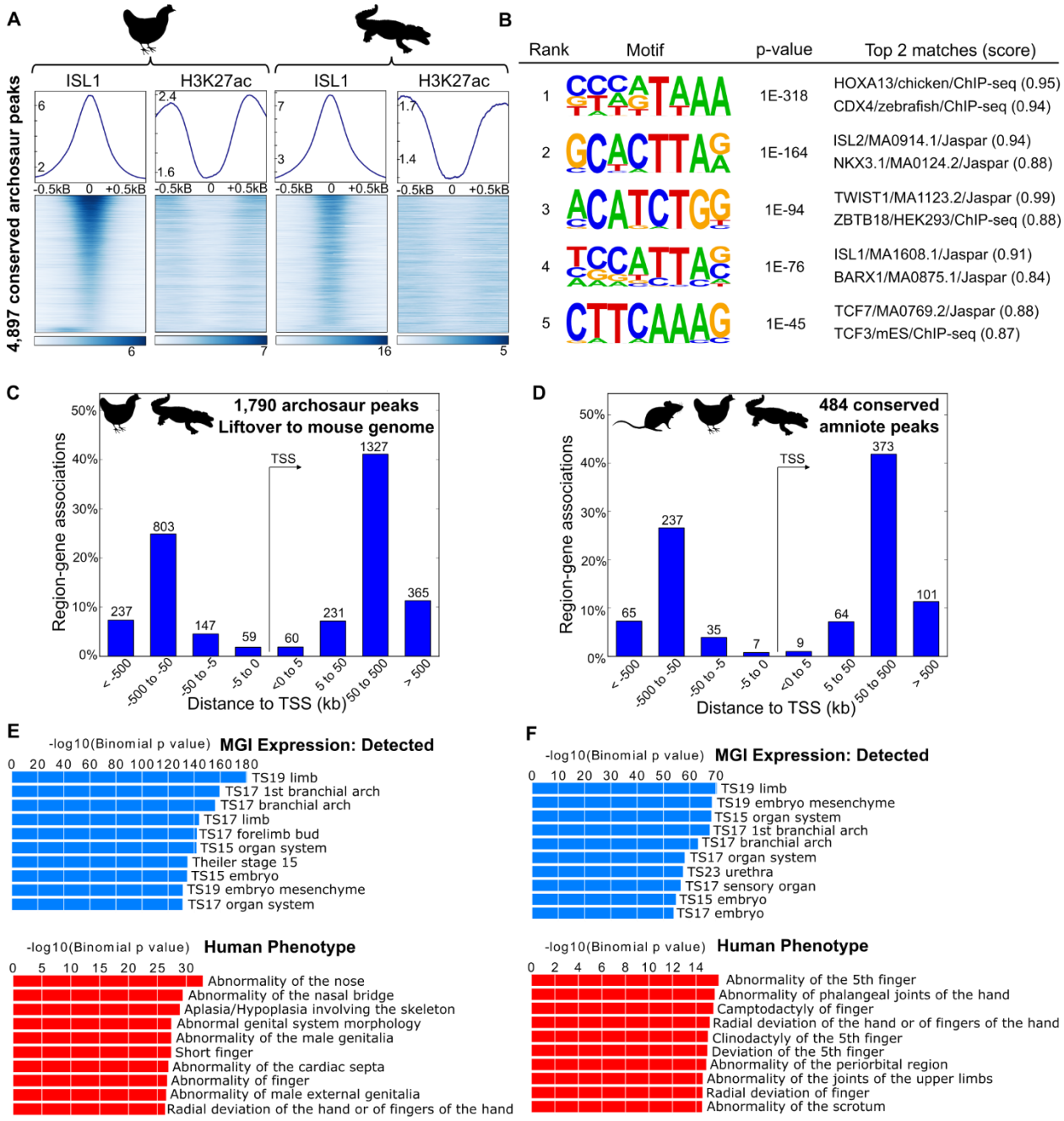
A
ISL1
H3K27ac
ISL1
H3K27ac
4,897 conserved archosaur peaks
-0.5kB 0 +0.5kB
B
Rank
Motif
p-value
Top 2 matches (score)
1
1E-318
HOXA13/chicken/ChIP-seq (0.95)
CDX4/zebrafish/ChIP-seq (0.94)
2
1E-164
ISL2/MA0914.1/Jaspar (0.94)
NKX3.1/MA0124.2/Jaspar (0.88)
3
1E-94
TWIST1/MA1123.2/Jaspar (0.99)
ZBTB18/HEK293/ChIP-seq (0.88)
4
1E-76
ISL1/MA1608.1/Jaspar (0.91)
BARX1/MA0875.1/Jaspar (0.84)
5
1E-45
TCF7/MA0769.2/Jaspar (0.88)
TCF3/mES/ChIP-seq (0.87)
C
1,790 archosaur peaks
Liftover to mouse genome
Region-gene associations
TSS
237
803
147
59
60
231
1327
365
< -500
-500 to -50
-50 to -5
-5 to 0
<0 to 5
5 to 50
50 to 500
> 500
Distance to TSS (kb)
D
484 conserved
amniote peaks
Region-gene associations
TSS
65
237
35
7
9
64
373
101
Distance to TSS (kb)
E
-log10(Binomial p value)
MGI Expression: Detected
TS19 limb
TS17 1st branchial arch
TS17 branchial arch
TS17 limb
TS17 forelimb bud
TS15 organ system
Theiler stage 15
TS15 embryo
TS19 embryo mesenchyme
TS17 organ system
-log10(Binomial p value)
Human Phenotype
Abnormality of the nose
Abnormality of the nasal bridge
Aplasia/Hypoplasia involving the skeleton
Abnormal genital system morphology
Abnormality of the male genitalia
Short finger
Abnormality of the cardiac septa
Abnormality of finger
Abnormality of male external genitalia
Radial deviation of the hand or of fingers of the hand
F
-log10(Binomial p value)
MGI Expression: Detected
TS19 limb
TS19 embryo mesenchyme
TS15 organ system
TS17 1st branchial arch
TS17 branchial arch
TS17 organ system
TS23 urethra
TS17 sensory organ
TS15 embryo
TS17 embryo
-log10(Binomial p value)
Human Phenotype
Abnormality of the 5th finger
Abnormality of phalangeal joints of the hand
Camptodactyly of finger
Radial deviation of the hand or of fingers of the hand
Clinodactyly of the 5th finger
Deviation of the 5th finger
Abnormality of the periorbital region
Abnormality of the joints of the upper limbs
Radial deviation of finger
Abnormality of the scrotum

**Figure 3.3.** ISL1 ChIP-seq in the chicken and alligator genital tubercle. **(A)** ISL1 and H3K27ac signal centered on ISL1 binding loci that are conserved between chick and alligator, or "archosaur peaks". **(B)** *De novo* motif analysis from HOMER showing the top 5 motifs from the central 100 bp of archosaur peaks. **(C)** Distribution of archosaur ISL1 peaks relative to nearby transcription start sites (TSS). **(D)** Distribution of amniote ISL1 peaks (conserved in mouse, chick, and alligator) relative to nearby transcription start sites (TSS). **(E)** GO analysis on archosaur peaks showing top 10 categories based on mouse gene expression (top) and human phenotypes (bottom). **(F)** GO analysis on amniote peaks showing top 10 categories based on mouse gene expression (top) and human phenotypes (bottom).

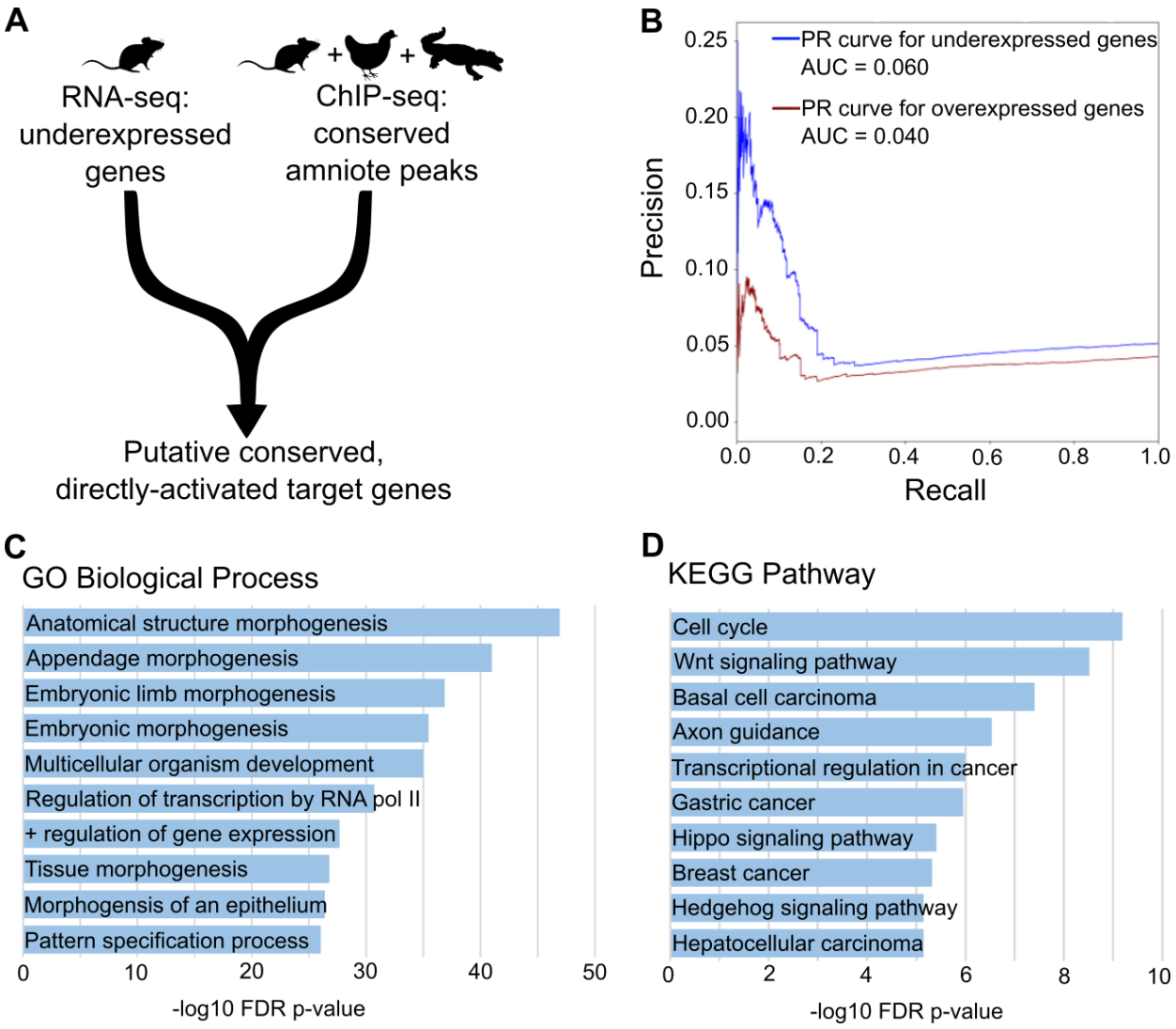


**Figure 3.4.** Conserved, directly activated targets of ISL1 in the genital tubercle. **(A)** Approach for identifying putative direct target genes from underexpressed genes in *Isl1* cKOs with ISL1 peaks conserved across amniotes. **(B)** Precision-recall curves for underexpressed and overexpressed genes using the conserved amniote ISL1 peaks. The area under the curves (AUC) suggests that the ChIP-seq data is more predictive of underexpressed genes. **(C)** GO analysis from Cistrome-GO on putative conserved, directly activated target genes showing the top 10 terms within "GO Biological Process". **(D)** Top 10 terms within "KEGG Pathway".

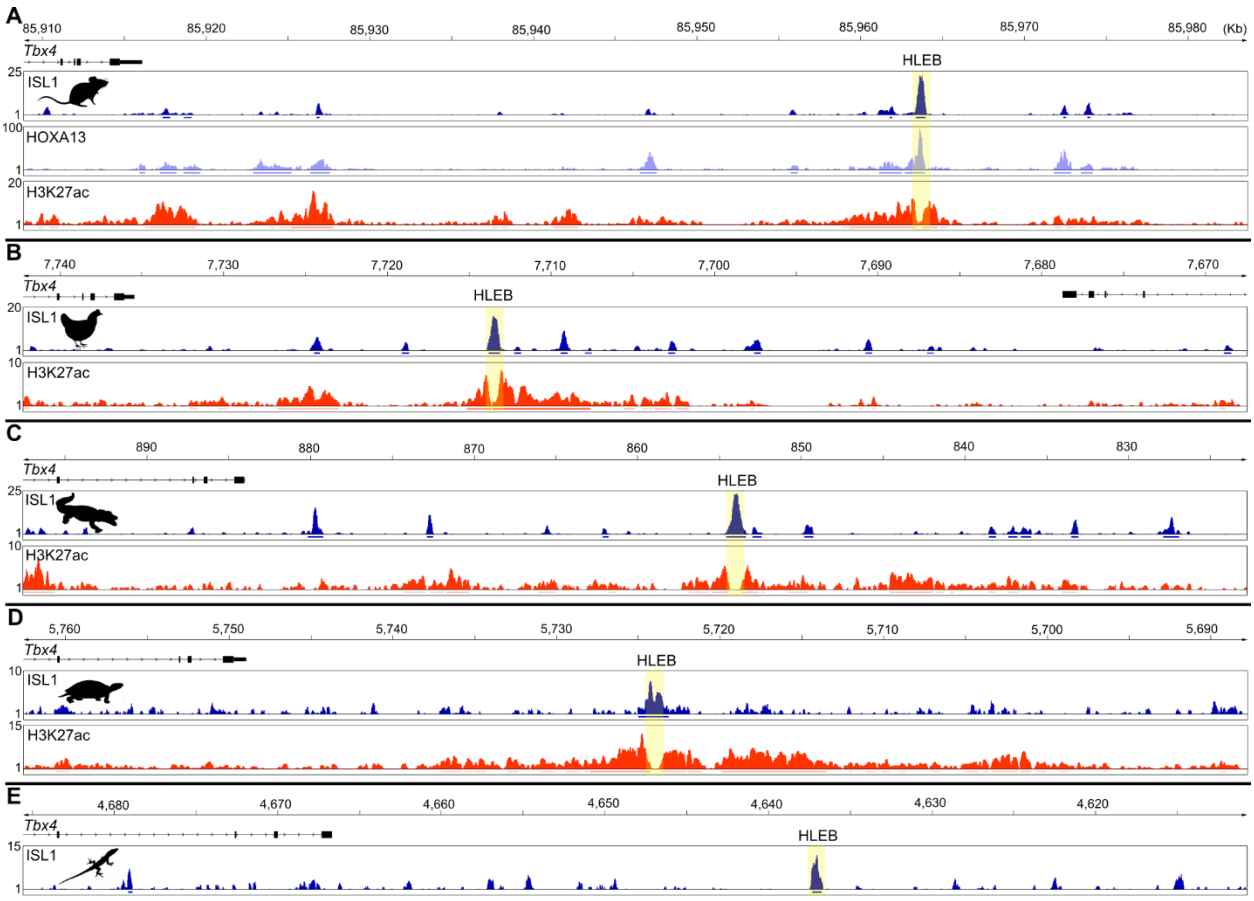


**Figure 3.5.** ISL1 binding at the *Tbx4* enhancer HLEB in the genitalia across amniotes. **(A)** Mouse ISL1 ChIP-seq, HOXA13 CUT-and-RUN (Amândio et al. 2020), and H3K27ac ChIP-seq (Infante et al. 2015) in the genital tubercle. mm10 coordinates chr11:85,908,850-85,983,849. **(B)** Chicken ISL1 and H3K27ac ChIP-seq in the GT. galGal6 coordinates chr19:7,667,461-7,742,460. **(C)** Alligator ISL1 and H3k27ac ChIP-seq in the GT. allMis1 coordinates JH733892:822,775-897,774. **(D)** Turtle ISL1 and H3k27ac ChIP-seq in the GT. CAS_Tse_1.0 coordinates NC_048315:5,687,778-5,762,777. **(E)** Brown anole ISL1 ChIP-seq in the hemiphallus. AnoSag2.0 coordinates scaffold_11:4,610,758-4,685,757.

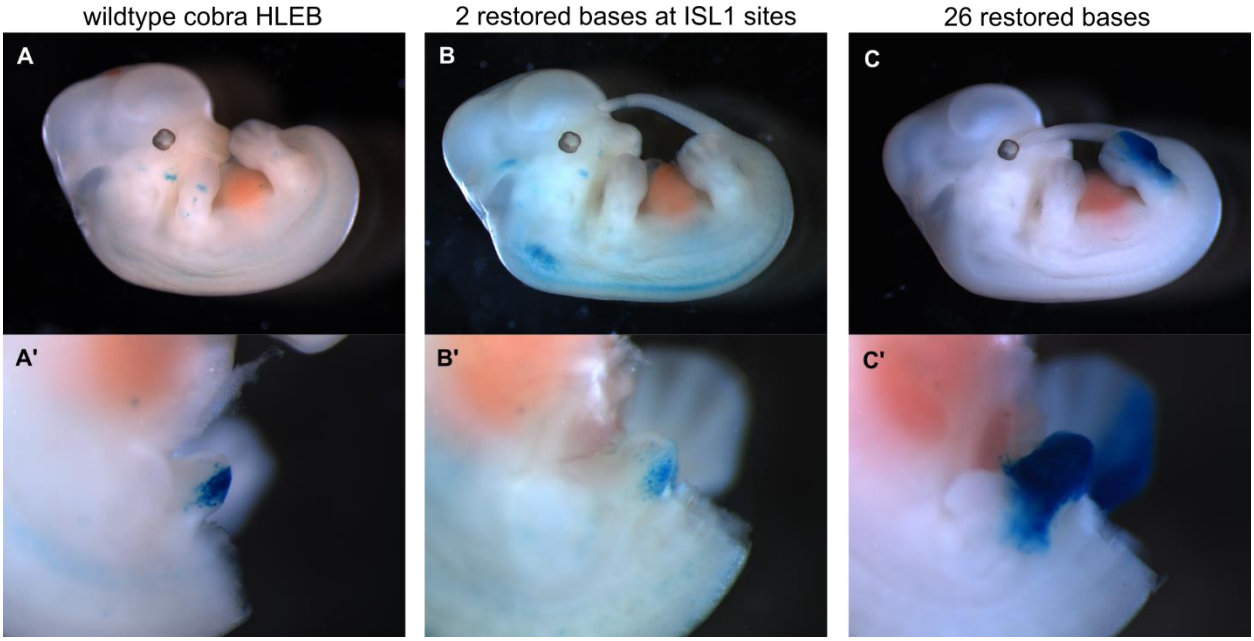


**Figure 3.6.** Activity of cobra HLEB transgenes. **(A)** Example of mouse carrying a transgenic LacZ construct containing the wildtype cobra HLEB. **(B)** Mouse carrying transgenic cobra HLEB construct with 2 bases restored at putative ISL1 sites. **(C)** Mouse carrying transgenic cobra HLEB construct with 26 bases restored. A', B' and C' show a lateral view of the GT with the right hindlimb and tail removed.

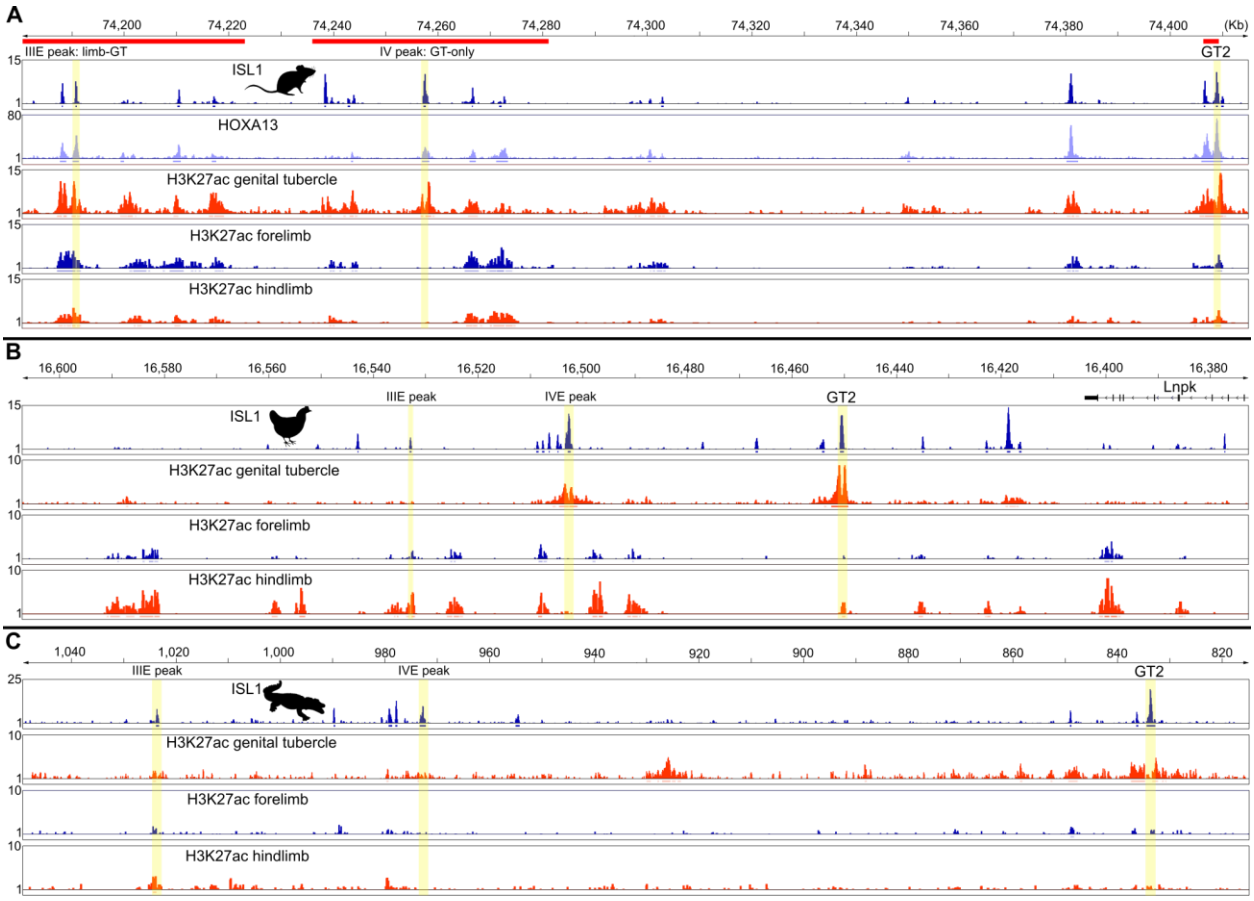


**Figure 3.7.** Conserved ISL1 binding at known enhancers of *Hoxd13*. **(A)** Mouse ISL1 ChIP-seq and HOXA13 CUT-and-RUN (Amândio et al. 2020) in the genital tubercle and H3K27ac ChIP-seq (Infante et al. 2015) in the GT, forelimb, and hindlimb. mm10 coordinates chr2:74,180,518-74,415,518. **(B)** Chicken ISL1 ChIP-seq in the GT and H3K27ac in the GT, forelimb, and hindlimb. galGal6 coordinates chr7:16,372,745-16,607,745. **(C)** Alligator ISL1 ChIP-seq in the GT and H3K27ac in the GT, forelimb, and hindlimb. allMis1 coordinates JH736379:814,909-1,049,909.

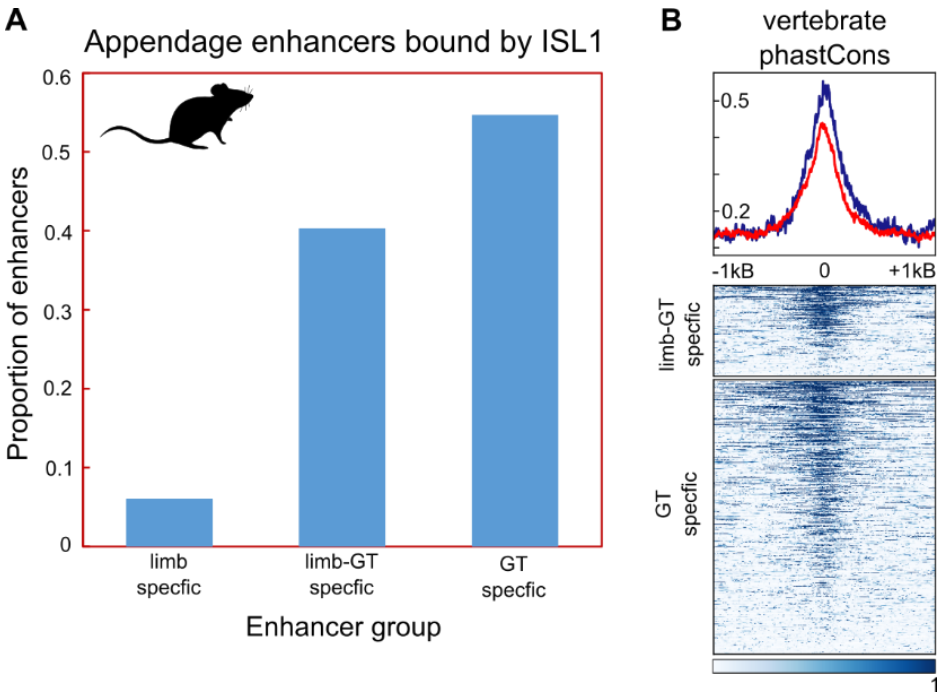


**Figure S3.1.** ISL1 binding to appendage enhancers in the genital tubercle. **(A)** Proportion of limb-specific, limb-GT, and GT-specific enhancers bound by ISL1 in the mouse GT. Enhancer groups are from Infante et al. (2015). **(B)** Vertebrate phastCons scores around ISL1 peak summits that occur within limb-GT (top, blue line plot) or GT-specific enhancers (bottom, red line). Mean phastCons score in central 100 bp is ~0.515 for the limb-GT enhancers and ~0.418 for the GT-specific enhancers (two-tailed T-test p-value = 6.03E-6).

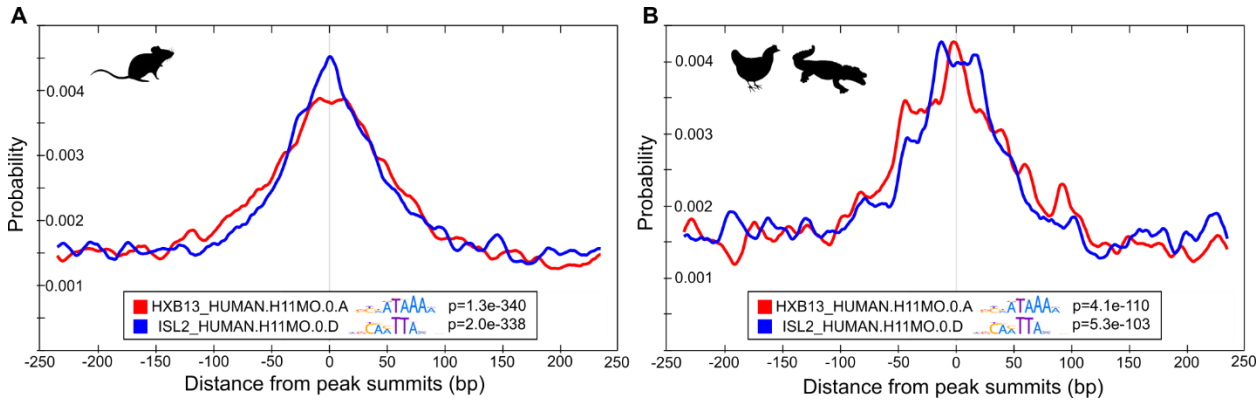


**Figure S3.2.** Enrichment of HOX13 and ISL motifs near ISL1 peak summits. **(A)** CentriMo analysis on mouse ISL1 ChIP-seq peaks centered on peak summits, showing top two most centrally enriched motifs from a database of known human and mouse motifs. **(B)** CentriMo analysis on archosaur ISL1 ChIP-seq peaks centered on peak summits, showing top two most centrally enriched motifs from a database of known human and mouse motifs.

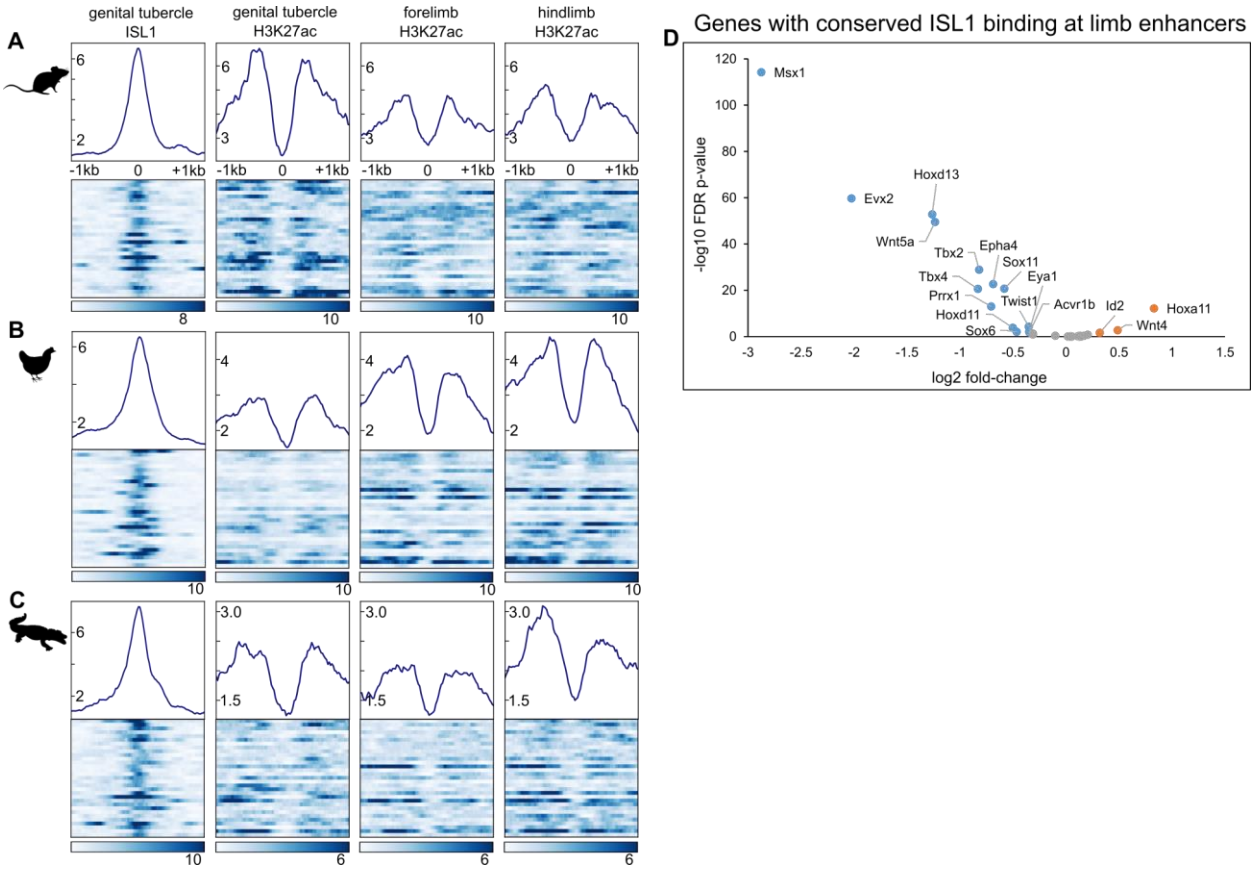


**Figure S3.3.** Annotated limb enhancers with conserved ISL1 binding. **(A)** Conserved ISL1 binding loci in the mouse GT that occur within annotated limb enhancers from Andrey et al. (2017). From left to right, ISL1 in the GT, H3K27ac in the GT, H3K27ac in the forelimb, and H3K27ac in the hindlimb. **(B)** Orthologous chicken regions showing the corresponding ChIP-seq data from (A). **(C)** Orthologous alligator regions. **(D)** Volcano plot showing only genes with conserved ISL1 binding loci at annotated limb enhancers. Blue genes are underexpressed, red are overexpressed.

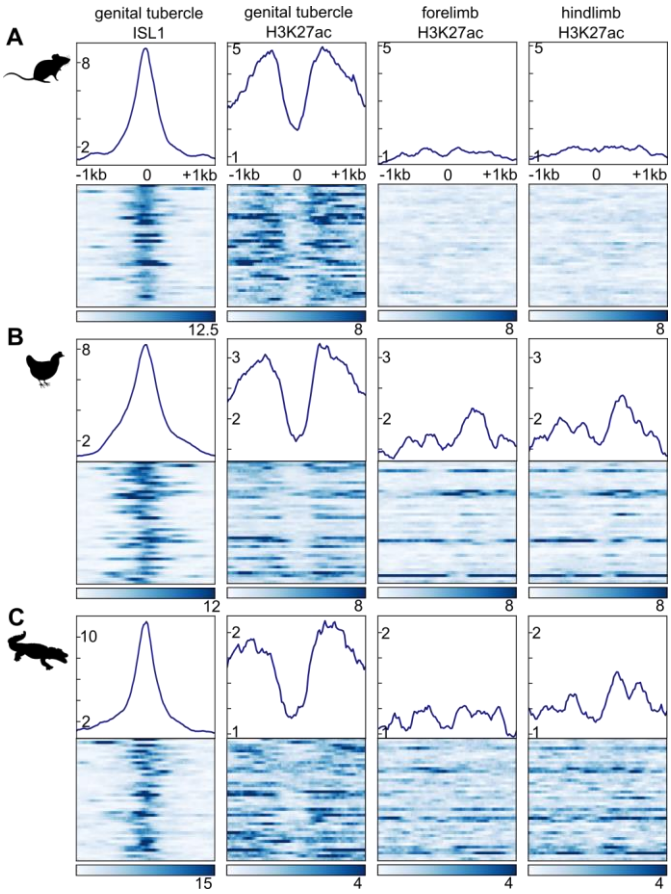


**Figure S3.4.** Genital tubercle-specific enhancers with conserved ISL1 binding. **(A)** Conserved ISL1 binding loci in the mouse genital tubercle that occur within GT-specific enhancers from Infante et al. (2015). From left to right, ISL1 in the GT, H3K27ac in the GT, H3K27ac in the forelimb, and H3K27ac in the hindlimb. **(B)** Orthologous chicken regions showing the corresponding ChIP-seq data from (A). **(C)** Orthologous alligator regions.

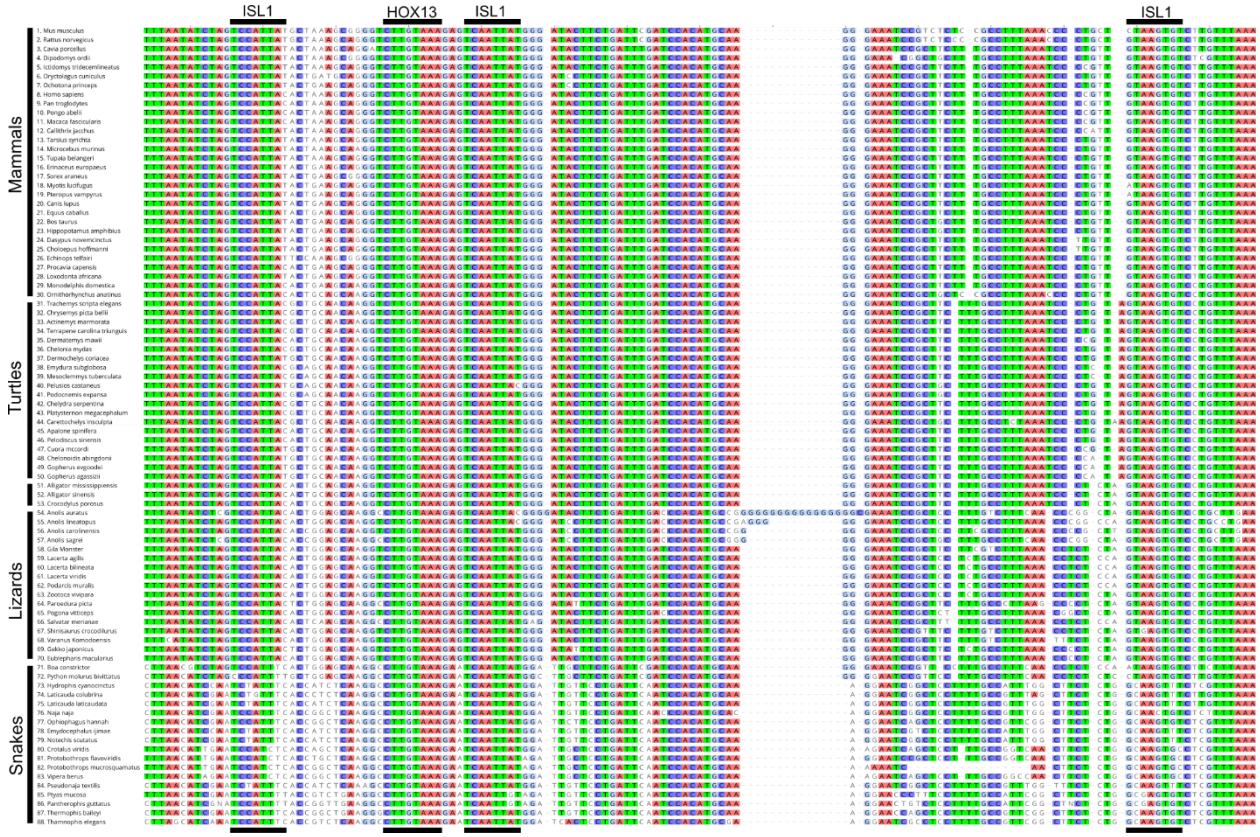


**Figure S3.5.** Highly conserved ISL1 motifs within HLEB. Multi-species alignment showing a sample of amniotes that undergo normal phallus development, including (from top to bottom): mammals, turtles, crocodilians, lizards, and snakes. Three ISL1 motifs are mostly invariant across the limbed amniotes, but show mutations in at least some snake species. King cobra (*Ophiophagus hannah*) has a mutation that appears ancestral to all snakes (A->T, leftmost ISL1 motif) and a second that appears in all Colubroid snakes (T->C, rightmost ISL1 motif).

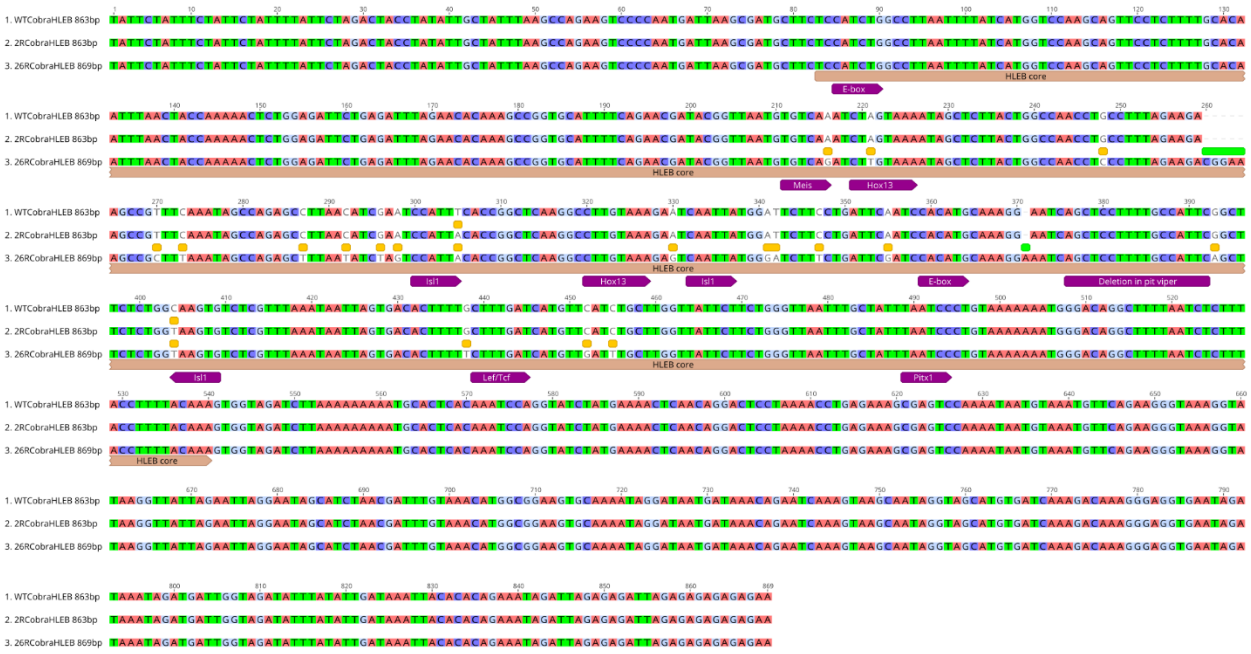


**Figure S3.6.** Transgenic sequences derived from cobra HLEB. Sequence 1 is the wildtype cobra HLEB sequence, sequence 2 has 2 bases restored at ISL1 sites, and sequence 3 has 26 conserved bases restored. Restored bases are indicated above each sequence, while putative transcription factor binding sites are indicated below.

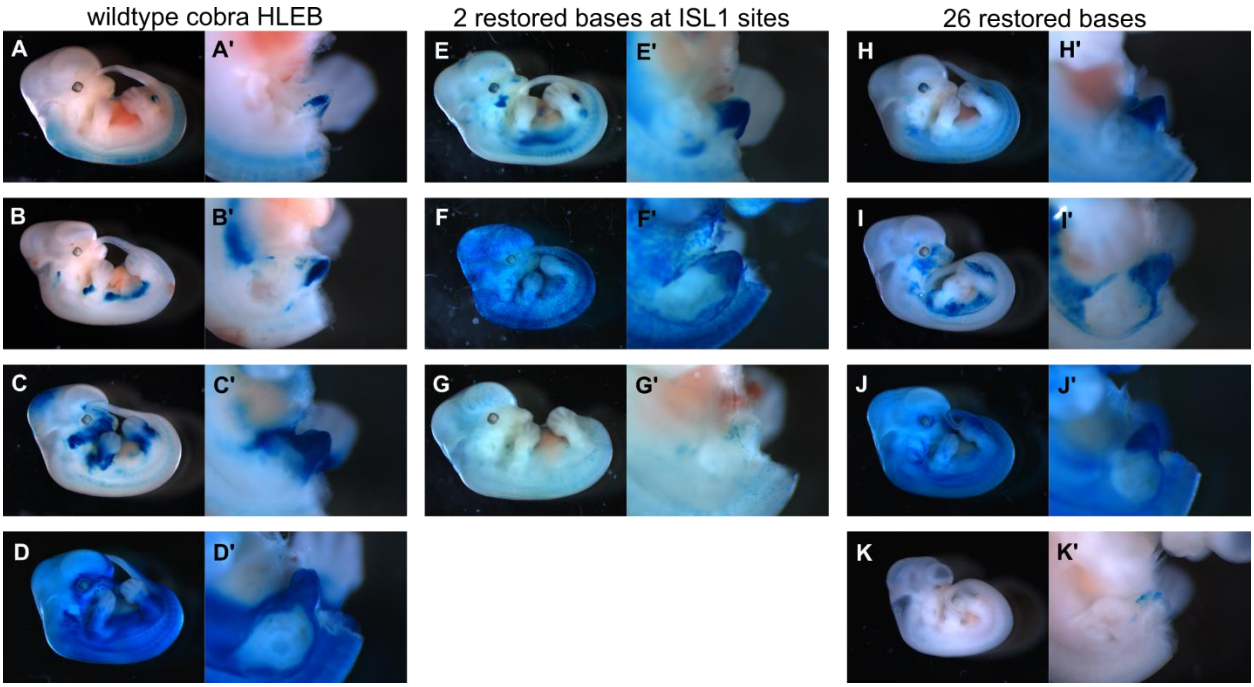


**Figure S3.7.** Other transgenic mice that show LacZ activity. The same genotypes are shown as in Fig. 3.6.

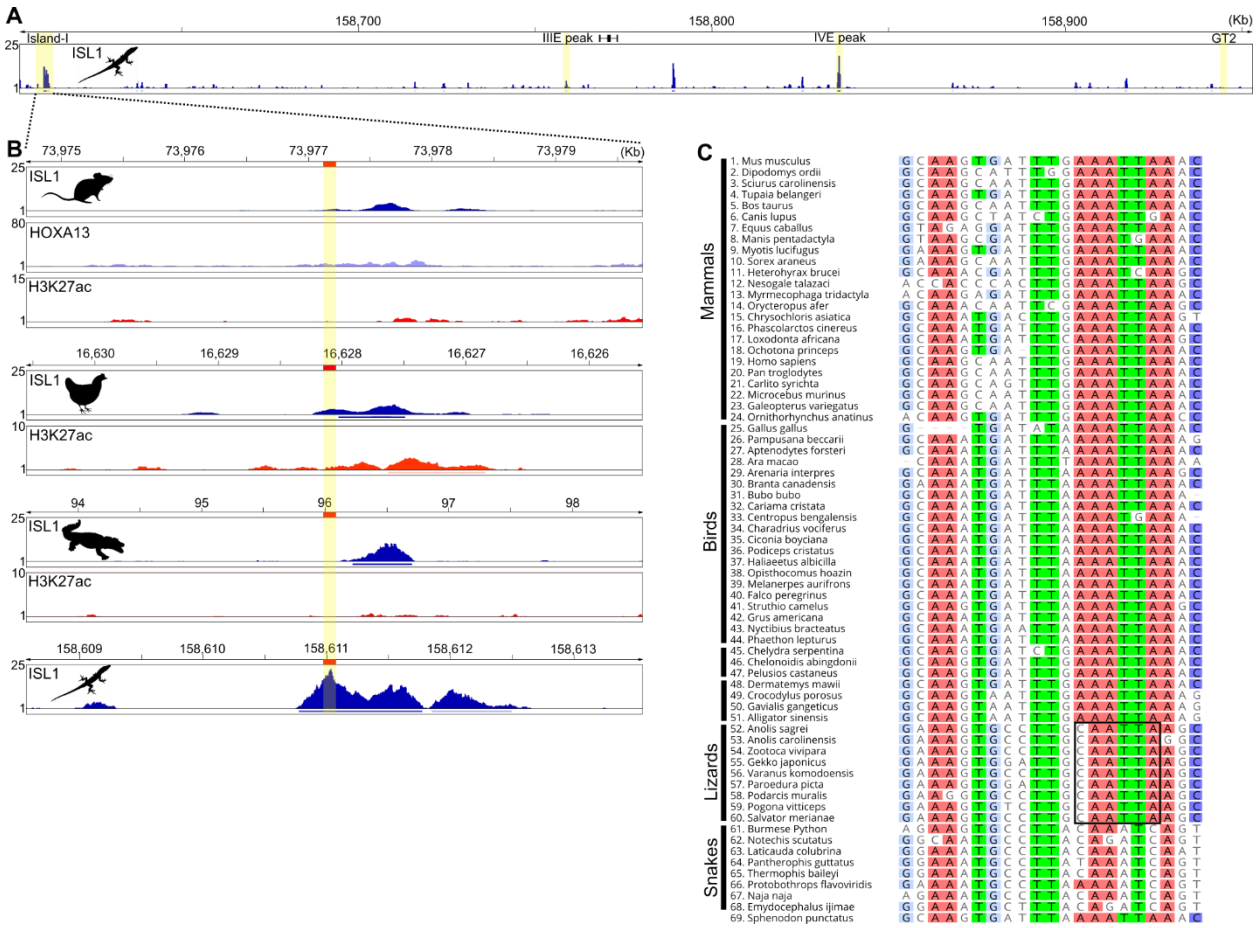


**Figure S3.8.** Possible gain of ISL1 binding site in the *Hoxd13* enhancer Island-I in lizards. **(A)** ISL1 ChIP-seq in the lizard hemiphallus. AnoSag2.0 coordinates scaffold_1:158,604,000-158,954,000. **(B)** ISL1 ChIP-seq in the genitalia of mouse, chick, alligator, and lizard showing a ~5 kb region corresponding to the Island-I region shown in (A). Highlighted region is 100 bp centered on a possible site of mutation in lizards, also corresponding to an apparent gain of ISL1 signal. **(C)** Multi-species alignment including (from top to bottom) mammals, birds, crocodilians, turtles, lizards, snakes, and tuatara. Box shows putative gain of ISL1 binding site in lizards.

**Table S3.1.** Sample of differentially expressed genes in the mouse genital tubercle.

| Gene | baseMean | log2FoldChange | lfcSE | stat | pvalue | padj |
|---|---|---|---|---|---|---|
| Dlx5 | 481.0555 | -3.96832969 | 0.161575 | -24.5603 | 3.36E-133 | 4.68E-129 |
| Msx1 | 2046.688 | -2.868983259 | 0.124721 | -23.0032 | 4.33E-117 | 3.02E-113 |
| Lmo7 | 1803.999 | -2.062518563 | 0.098142 | -21.0156 | 4.72E-98 | 2.19E-94 |
| Dlx6 | 154.7027 | -3.184507646 | 0.176496 | -18.0429 | 8.96E-73 | 3.12E-69 |
| Lix1 | 2666.616 | -1.562681634 | 0.088091 | -17.7394 | 2.08E-70 | 5.80E-67 |
| Pdlim3 | 488.2946 | -2.11805552 | 0.122645 | -17.2698 | 7.94E-67 | 1.84E-63 |
| Msx2 | 755.988 | -2.315779889 | 0.134692 | -17.1931 | 2.99E-66 | 5.95E-63 |
| Evx2 | 897.4076 | -2.024159002 | 0.120203 | -16.8395 | 1.25E-63 | 2.18E-60 |
| Wnt5a | 6059.595 | -1.27254624 | 0.07924 | -16.0594 | 4.91E-58 | 7.60E-55 |
| Hoxd13 | 6809.945 | -1.260118365 | 0.079771 | -15.7968 | 3.27E-56 | 4.56E-53 |
| Sp9 | 134.6407 | -2.667341145 | 0.178894 | -14.9102 | 2.83E-50 | 3.58E-47 |
| Gfra2 | 892.1082 | -1.486387232 | 0.101143 | -14.6958 | 6.85E-49 | 7.96E-46 |
| Epha5 | 522.2822 | -1.823750037 | 0.127418 | -14.3131 | 1.81E-46 | 1.94E-43 |
| Hoxa11os | 2490.783 | -1.026403466 | 0.076887 | -13.3495 | 1.19E-40 | 1.19E-37 |
| Dpysl3 | 5924.248 | -0.964331061 | 0.074079 | -13.0175 | 9.73E-39 | 9.04E-36 |
| Hoxd12 | 2450.459 | -1.14251579 | 0.095558 | -11.9562 | 6.02E-33 | 5.03E-30 |
| Tfap2a | 535.0762 | -1.641431 | 0.137305 | -11.9546 | 6.14E-33 | 5.03E-30 |
| Tbx2 | 4140.621 | -0.819035443 | 0.069027 | -11.8655 | 1.79E-32 | 1.38E-29 |
| Isl2 | 653.6532 | -1.269664415 | 0.109177 | -11.6295 | 2.92E-31 | 2.14E-28 |
| Greb1 | 1390.024 | -1.298082744 | 0.116416 | -11.1504 | 7.13E-29 | 4.97E-26 |
| Scrn1 | 330.6641 | -1.738304316 | 0.160472 | -10.8324 | 2.42E-27 | 1.60E-24 |
| Rai14 | 6413.043 | -0.687053651 | 0.064412 | -10.6666 | 1.46E-26 | 9.24E-24 |
| Epha4 | 5135.488 | -0.687517359 | 0.064671 | -10.6309 | 2.14E-26 | 1.30E-23 |
| Kif26b | 7273.015 | -0.784023124 | 0.077196 | -10.1563 | 3.11E-24 | 1.80E-21 |
| Sox11 | 15343.7 | -0.581771855 | 0.057338 | -10.1464 | 3.44E-24 | 1.92E-21 |
| Tbx4 | 2795.046 | -0.828871729 | 0.082138 | -10.0912 | 6.04E-24 | 3.24E-21 |
| Mllt3 | 2813.124 | -0.721853385 | 0.071664 | -10.0728 | 7.29E-24 | 3.76E-21 |
| Pkdcc | 1501.01 | 1.217306966 | 0.121168 | 10.04646 | 9.52E-24 | 4.74E-21 |
| Hand2 | 1012.313 | -1.278841378 | 0.127651 | -10.0183 | 1.27E-23 | 6.09E-21 |
| Postn | 1420.439 | 1.331742237 | 0.133457 | 9.978819 | 1.89E-23 | 8.76E-21 |
| Gdf6 | 206.1543 | -1.60558255 | 0.161062 | -9.96874 | 2.09E-23 | 9.39E-21 |
| Lrig3 | 2593.922 | 0.789226173 | 0.079568 | 9.918908 | 3.45E-23 | 1.50E-20 |
| Dlx6os1 | 52.24797 | -1.732111556 | 0.174728 | -9.9132 | 3.65E-23 | 1.54E-20 |
| Maged2 | 6851.623 | 0.562913369 | 0.057595 | 9.773597 | 1.46E-22 | 5.99E-20 |
| Msi2 | 4449.844 | -0.582043059 | 0.060055 | -9.69187 | 3.27E-22 | 1.30E-19 |
| Dlx3 | 203.8388 | -1.465177948 | 0.151331 | -9.68193 | 3.60E-22 | 1.39E-19 |
| Islr | 1118.49 | 1.282310002 | 0.132715 | 9.662164 | 4.37E-22 | 1.64E-19 |
| Prdm1 | 424.2776 | -1.437597146 | 0.149783 | -9.59788 | 8.16E-22 | 2.99E-19 |

**Table S3.2.** Sample of peaks from ISL1 ChIP-seq in the genital tubercle.

*Mouse*

| Chr | Start | End | -log10pval | summit |
|---|---|---|---|---|
| chr11 | 85963380 | 85963984 | 427.68387 | 293 |
| chr5 | 20133924 | 20134454 | 418.38012 | 226 |
| chr5 | 119518674 | 119519096 | 392.90984 | 215 |
| chr9 | 42010328 | 42010894 | 388.04964 | 215 |
| chr18 | 69051882 | 69052351 | 375.9791 | 218 |
| chr1 | 13715430 | 13715842 | 367.09368 | 183 |
| chr17 | 17466719 | 17467117 | 366.97632 | 174 |
| chr3 | 70035965 | 70036353 | 355.05466 | 199 |
| chr16 | 77210090 | 77210548 | 352.42358 | 211 |
| chr14 | 60863281 | 60863967 | 351.82278 | 424 |
| chr4 | 29306549 | 29307080 | 351.09205 | 320 |
| chr3 | 79742416 | 79743091 | 350.16163 | 303 |
| chr4 | 57166366 | 57166778 | 345.56151 | 206 |
| chr8 | 50240387 | 50240730 | 344.84166 | 174 |
| chr1 | 59520218 | 59520774 | 338.45888 | 317 |
| chr10 | 29751099 | 29751602 | 337.52104 | 218 |
| chr17 | 48266745 | 48267108 | 335.20677 | 167 |
| chr12 | 31701192 | 31701676 | 329.88426 | 242 |
| chr5 | 84093964 | 84094496 | 328.85056 | 261 |
| chr5 | 133772552 | 133773071 | 328.40748 | 266 |
| chr3 | 65055355 | 65055809 | 325.61764 | 206 |
| chr16 | 90007609 | 90008017 | 324.69704 | 190 |
| chr9 | 55568120 | 55568539 | 324.05896 | 234 |
| chr1 | 22330560 | 22331219 | 320.44611 | 300 |
| chr3 | 56302356 | 56302764 | 318.90327 | 198 |
| chr1 | 179147076 | 179147528 | 313.40852 | 245 |
| chr15 | 46592789 | 46593256 | 305.504 | 198 |
| chr9 | 104123460 | 104123888 | 302.90287 | 199 |
| chr4 | 10808728 | 10809065 | 300.98323 | 174 |
| chr1 | 69319590 | 69319942 | 300.92893 | 173 |
| chr12 | 9997182 | 9997594 | 299.84543 | 234 |
| chr9 | 64341861 | 64342380 | 297.36968 | 294 |
| chr15 | 35702462 | 35703054 | 296.15419 | 359 |
| chr11 | 86118026 | 86118532 | 295.5103 | 247 |
| chr3 | 141475666 | 141476316 | 291.47674 | 307 |
| chr4 | 31263567 | 31264076 | 290.55075 | 188 |

*Chick*

| Chr | Start | End | -log10pval | summit |
|---|---|---|---|---|
| chr17 | 3887292 | 3887791 | 524.25406 | 289 |
| chr4 | 51400601 | 51401225 | 519.14313 | 302 |
| chr21 | 5870821 | 5871403 | 509.41012 | 215 |
| chr2 | 137060572 | 137061247 | 507.98886 | 322 |
| chr12 | 8928372 | 8929141 | 496.88141 | 383 |
| chr1 | 157342315 | 157342993 | 496.40755 | 373 |
| chr1 | 178846288 | 178846849 | 491.08371 | 313 |
| chr2 | 117807949 | 117808456 | 489.577 | 306 |
| chr8 | 3617102 | 3617816 | 489.09027 | 338 |
| chr4 | 22828181 | 22828706 | 487.66363 | 254 |
| chr1 | 62689590 | 62690228 | 486.04583 | 261 |
| chr1 | 134734076 | 134734584 | 482.25003 | 250 |
| chr7 | 33978386 | 33979041 | 478.62443 | 319 |
| chr1 | 156698395 | 156698975 | 478.27777 | 310 |
| chr4 | 82093125 | 82093613 | 475.62484 | 247 |
| chr14 | 11613869 | 11614483 | 474.28003 | 273 |
| chr1 | 192890576 | 192891191 | 474.19955 | 331 |
| chr2 | 47547460 | 47548023 | 473.18235 | 272 |
| chr1 | 126872077 | 126872746 | 471.29037 | 371 |
| chr2 | 81838453 | 81839003 | 466.27631 | 254 |
| chr15 | 12686115 | 12686620 | 464.5356 | 263 |
| chr3 | 41687296 | 41688096 | 463.72082 | 387 |
| chr4 | 7480541 | 7481008 | 463.20584 | 259 |
| chr19 | 7647680 | 7648287 | 463.1738 | 298 |
| chr1 | 19706894 | 19707367 | 462.28501 | 178 |
| chr2 | 126159522 | 126159942 | 461.0161 | 197 |
| chr19 | 7713121 | 7713836 | 460.25726 | 361 |
| chr3 | 92734463 | 92735029 | 459.96462 | 254 |
| chr2 | 45458653 | 45459101 | 459.72817 | 220 |
| chr11 | 4430888 | 4431433 | 459.04981 | 272 |
| chr1 | 188321638 | 188322376 | 455.92011 | 333 |
| chr3 | 49211101 | 49211736 | 454.92627 | 324 |
| chr4 | 76109354 | 76109793 | 454.62333 | 219 |
| chr1 | 193612087 | 193612581 | 454.50759 | 250 |
| chr1 | 175999706 | 176000230 | 454.41336 | 238 |
| chr6 | 14885597 | 14886125 | 454.27041 | 254 |
| chrZ | 9830387 | 9830836 | 453.42853 | 242 |
| chr4 | 69900780 | 69901360 | 453.00329 | 282 |

*Alligator*

| Chr | Start | End | -log10pval | summit |
|---|---|---|---|---|
| JH739738 | 251636 | 252378 | 394.02707 | 492 |
| JH738623 | 158303 | 160120 | 384.89191 | 272 |
| JH737817 | 272403 | 273761 | 376.66153 | 516 |
| JH737817 | 275242 | 276449 | 371.1792 | 539 |
| JH736185 | 61578 | 63776 | 369.20038 | 808 |
| JH735249 | 752961 | 753572 | 366.44571 | 288 |
| JH731326 | 113117 | 114322 | 357.2225 | 686 |
| JH732130 | 260671 | 261333 | 353.05563 | 289 |
| JH734798 | 191193 | 192037 | 348.71359 | 329 |
| JH738290 | 233488 | 234416 | 346.42633 | 550 |
| JH734884 | 599681 | 600498 | 344.3338 | 259 |
| JH735807 | 270167 | 270701 | 341.44882 | 265 |
| JH733614 | 696578 | 697230 | 336.85422 | 305 |
| JH738777 | 1427888 | 1429092 | 336.12537 | 831 |
| JH732000 | 60221 | 60728 | 327.79962 | 273 |
| JH731138 | 81602 | 83577 | 325.54593 | 730 |
| JH739256 | 178826 | 180158 | 325.18631 | 508 |
| JH738492 | 73814 | 75259 | 323.44135 | 523 |
| JH732260 | 35716 | 37329 | 321.70065 | 904 |
| JH736906 | 723271 | 723936 | 321.36734 | 196 |
| JH738987 | 105422 | 106342 | 316.97003 | 566 |
| JH733216 | 177967 | 178582 | 312.61456 | 312 |
| JH736794 | 549540 | 550009 | 310.79153 | 239 |
| JH736231 | 108350 | 108945 | 310.52072 | 336 |
| JH739842 | 62765 | 63385 | 308.80133 | 327 |
| JH738883 | 143001 | 143434 | 306.46902 | 215 |
| JH738392 | 410109 | 410843 | 304.64612 | 482 |
| JH733060 | 307394 | 308434 | 303.90698 | 551 |
| JH735570 | 280213 | 281471 | 303.56747 | 631 |
| JH733970 | 979561 | 980438 | 301.56256 | 592 |
| JH738656 | 60644 | 61377 | 301.13632 | 325 |
| JH732151 | 141047 | 142527 | 299.44525 | 329 |
| JH732980 | 29930 | 30875 | 295.73587 | 549 |
| JH734429 | 100794 | 101645 | 289.34756 | 382 |
| JH738431 | 278316 | 278843 | 288.75919 | 265 |
| JH734059 | 497411 | 498603 | 286.06393 | 470 |
| JH735170 | 169314 | 170424 | 282.72437 | 261 |
| JH738375 | 262619 | 263382 | 281.26813 | 469 |

**Table S3.3.** Top putative conserved, directly activated target genes from Cistrome-GO.

| Gene | PeakNumber | adjRPscore | logFoldChange | DE significance | Rank_by_rank_product |
|---|---|---|---|---|---|
| Msx2 | 3 | 0.461 | -2.318 | 2.42E-63 | 1 |
| Msx1 | 4 | 0.306 | -2.874 | 6.19E-115 | 2 |
| Lix1 | 1 | 0.392 | -1.563 | 3.96E-67 | 3 |
| Tbx4 | 9 | 0.591 | -0.832 | 2.53E-21 | 4 |
| Tbx2 | 9 | 0.489 | -0.82 | 1.35E-29 | 5 |
| Mab21l1 | 5 | 0.821 | -0.67 | 1.56E-06 | 6 |
| Tfap2a | 2 | 0.403 | -1.643 | 3.78E-30 | 7 |
| Alx4 | 2 | 0.479 | -0.638 | 6.12E-11 | 9 |
| Isl2 | 1 | 0.34 | -1.269 | 3.96E-28 | 10 |
| Gja1 | 1 | 0.426 | -0.515 | 5.70E-09 | 11 |
| Prdm1 | 2 | 0.344 | -1.438 | 2.58E-19 | 12 |
| Hoxa11os | 4 | 0.171 | -1.027 | 1.65E-37 | 13 |
| Evx2 | 5 | 0.057 | -2.024 | 1.91E-60 | 14 |
| Unc5c | 1 | 0.369 | -0.798 | 1.11E-12 | 15 |
| Wnt5a | 7 | 0.073 | -1.235 | 2.91E-50 | 16 |
| Hoxd13 | 5 | 0.051 | -1.261 | 1.65E-53 | 17 |
| Lmo4 | 4 | 0.346 | -0.633 | 2.92E-14 | 18 |
| Jag1 | 1 | 0.286 | -1.009 | 1.05E-17 | 20 |
| Hoxd12 | 5 | 0.046 | -1.144 | 2.59E-30 | 22 |
| Ppp4r3a | 1 | 0.425 | -0.266 | 7.29E-04 | 23 |
| Gdf6 | 1 | 0.063 | -1.603 | 1.10E-20 | 24 |
| Lnpk | 5 | 0.173 | -0.808 | 5.95E-12 | 25 |
| Hottip | 4 | 0.154 | -0.985 | 5.81E-11 | 29 |
| Hand2 | 1 | 0.019 | -1.281 | 3.60E-21 | 31 |
| Sall4 | 2 | 0.285 | -0.917 | 4.46E-06 | 32 |
| Evx1os | 4 | 0.153 | -1.034 | 2.65E-09 | 33 |
| Bach2 | 2 | 0.175 | -0.635 | 9.17E-08 | 34 |
| Evx1 | 4 | 0.152 | -1.029 | 1.39E-08 | 37 |
| Casc3 | 1 | 0.342 | -0.296 | 6.81E-04 | 38 |
| Grem1 | 1 | 0.266 | -0.721 | 3.71E-05 | 39 |
| Twist1 | 5 | 0.285 | -0.352 | 5.73E-05 | 40 |
| Bmi1 | 1 | 0.425 | -0.283 | 2.95E-02 | 41 |
| Foxf2 | 2 | 0.016 | -0.822 | 7.46E-18 | 42 |
| Rspo3 | 1 | 0.064 | -0.889 | 5.18E-11 | 43 |
| Tbx3 | 5 | 0.022 | -0.526 | 6.08E-15 | 44 |
| Vax1 | 3 | 0.031 | -1.343 | 3.82E-13 | 46 |
| Tox3 | 4 | 0.289 | -0.635 | 4.17E-04 | 50 |
| Tbx18 | 1 | 0.319 | -0.57 | 2.09E-03 | 51 |

**Table S3.4.** Activity of the three cobra HLEB transgenes.

| | ventral GT | dorsal GT | ventral bias | posterior HL | limb other | non-appendage | figure image |
|---|---|---|---|---|---|---|---|
| **WT-1** | ++ | - | Y | - | ~ | ~ | 3.6A |
| **WT-2** | ++ | - | Y | - | - | + | S3.7A |
| **WT-3** | +++ | + | Y | - | + | ++ | S3.7B |
| **WT-4** | +++ | +++ | N | ++ | ++ | ++ | S3.7C |
| **WT-5** | +++ | +++ | N | ++ | ++ | +++ | S3.7D |
| **WT-6** | - | - | n/a | - | - | - | |
| **WT-7** | - | - | n/a | - | - | - | |
| **WT-8** | - | - | n/a | - | - | - | |
| **WT-9** | - | - | n/a | - | - | - | |
| **WT-10** | - | - | n/a | - | - | - | |
| **WT-11** | - | - | n/a | - | - | - | |
| **WT-12** | - | - | n/a | - | - | - | |
| | | | | | | | |
| **2R-1** | + | - | Y | - | - | ~ | 3.6B |
| **2R-2** | +++ | ++ | Y | - | - | ++ | S3.7E |
| **2R-3** | +++ | +++ | N | ++ | ++ | +++ | S3.7F |
| **2R-4** | ~ | ~ | n/a | ~ | ~ | ~ | S3.7G |
| **2R-5** | - | - | n/a | - | - | - | |
| **2R-6** | - | - | n/a | - | - | - | |
| **2R-7** | - | - | n/a | - | - | - | |
| **2R-8** | - | - | n/a | - | - | - | |
| **2R-9** | - | - | n/a | - | - | - | |
| | | | | | | | |
| **26R-1** | +++ | +++ | N | +++ | - | - | 3.6C |
| **26R-2** | +++ | +++ | N | ++ | - | + | S3.7H |
| **26R-3** | +++ | +++ | N | ++ | - | ++ | S3.7I |
| **26R-4** | +++ | +++ | N | +++ | + | +++ | S3.7J |
| **26R-5** | + | + | N | + | - | - | S3.7K |
| **26R-6** | - | - | n/a | ~ | - | - | |
| **26R-7** | - | - | n/a | - | - | - | |
| **26R-8** | - | - | n/a | - | - | - | |
| **26R-9** | - | - | n/a | - | - | - | |
| **26R-10** | - | - | n/a | - | - | - | |

~ : marginal acvitity. +/++/+++ : intensity and/or pervasiveness of signal

# CHAPTER 4

# CONCLUSIONS AND FUTURE DIRECTIONS

## *Major findings of this dissertation*

This work extends upon our knowledge of *Isl1*'s role in both hindlimb initiation and genital tubercle (GT) outgrowth. *Isl1* evidently occupies a crucial position in these gene regulatory networks based on the severity of the phenotype and widespread perturbation of gene expression upon its removal. While previous publications focused on particular genes that might be downstream of *Isl1*, the experiments presented here systematically identify many putative targets using a combination of RNA-seq and ChIP-seq. This includes genes which were already thought to be downstream of *Isl1*, such as *Tbx4* in the hindlimb field. In this case, ChIP-seq adds to our knowledge by showing that *Tbx4* is most likely a direct target of ISL1 via HLEB, and this function appears to be conserved across amniotes and shared with the GT. Far more genes and enhancers are previously unexplored targets of ISL1, allowing for numerous hypotheses about its role in activating or repressing important transcription factors and signaling pathways.

HLEB provides a paradigmatic example of a GT enhancer regulated by ISL1 that was likely coopted from the limb regulatory network during the evolution of the amniote phallus. While restoration of 26 conserved bases in the cobra HLEB shows restoration of activity in both the hindlimb and GT, restoration of ISL1 sites alone is not sufficient. This suggests that multiple mutated bases are essential for the complete pleiotropic function of this enhancer. The ISL1 sites may still be essential in combination with other sites, such as putative HOX13 sites, and further experiments may clarify this. For example, restoration of 24 of the 26 bases excluding the two

ISL1 sites could show the wildtype cobra pattern of GT expression and/or no hindlimb activity. This would support the idea that ISL1 is necessary but not sufficient to rescue activity. Alternatively, ablation of these ISL1 sites in the mouse HLEB could produce a similar result.

*Further mutant models and genome-wide analyses*

In the hindlimb field, it is possible my experiments deleting *Isl1* using a *Hoxb6*-Cre allele miss some genes downstream of *Isl1* because of the later, variable clearance of ISL1 protein from this knockout model (Itou et al., 2012). This conditional knockout (cKO) model was useful for studying the role of *Isl1* in both the hindlimb and the GT, as the latter develops much later and these mice survive through embryonic development. Earlier deletion using *T*-Cre completely ablates hindlimb initiation (Kawakami et al., 2011; Narkis et al., 2012), which may have the potential to reveal additional differentially expressed genes, while any common results would serve to reinforce my findings. In situ hybridizations could also be used to validate and resolve the expression patterns of differentially expressed genes.

Conversely, using a Cre allele that drives deletion later might be useful for narrowing down genes that are direct targets of ISL1 in the GT. This is because the *Hoxb6*-Cre allele ablates *Isl1* before the GT emerges, thus causing severe phenotypic divergence by E12.5. This likely also results in a greater proportion of indirect transcriptional effects than in the hindlimb field experiment. Research on *Shh* shows a similarly severe phenotype upon complete deletion (Haraguchi et al., 2001; Perriton et al., 2002), which was followed up by conditional deletion at later stages (Lin et al., 2009). A Cre allele that begins its activity around E11.5 may allow the GT to form more normally at E12.5, thus reducing the incidence of non-primary effects.

Deleting *Isl1* at even later stages might allow for further study of its role at E14.5 prior to sexual differentiation. E14.5 GT ISL1 ChIP-seq data from Ching et al. (2018) shows binding at a mammalian penile spine enhancer of *Ar* which is deleted in humans (McLean et al., 2011; Reno et al., 2013). Another approach employed by Lin and colleagues is the use of gain-of-function alleles, which they applied for studying the downstream activity of β-catenin (Lin et al., 2008), *Shh* (Lin et al., 2009) and *Fgf8* (Lin et al., 2013). *Shh* expression appears diminished in the *Isl1* cKO genitalia, which could contribute to the severity of the phenotype. Rescuing hedgehog signaling using a Cre allele might partially rescue the phenotype. Alternatively, it is possible the severity of the *Isl1* cKO phenotype results from the underexpression of myriad genes in the mesenchyme, and thus cannot be rescued through a particular pathway.

Another approach would be to perform RNA-seq in the cloacal mesenchyme prior to GT outgrowth using the *Hoxb6*-Cre allele, with embryos collected at E10.5. This could be coupled with ISL1 ChIP-seq at E10.5 to focus on *Isl1*'s role during genital bud initiation, as well as publically available ATAC-seq, H3K27ac ChIP-seq, and HOXA13 CUT-and-RUN in the E10.5 cloacal region (Amândio et al., 2020). This may address the extent to which enhancers active at E12.5 are already active prior to GT formation, including known appendage enhancers like HLEB. Amândio and coauthors found evidence that the GT2 enhancer is open and bound by HOXA13 at E10.5, but there hasn't been a published genome-wide analysis of these data. Such an analysis of the cloacal region could indicate the extent to which this region prefigures the appendage-like patterns seen later in the GT, or shows similarities with the hindlimb field.

In addition, more work could be done to understand the role of *Hox13* genes and their relationship with *Isl1* in GT development. *Hoxa13* and *Hoxd13* double knockouts show severe ablation of the GT similarly to *Isl1* cKOs (Kondo et al., 1997; Warot et al., 1997). RNA-seq

could be performed in these knockouts and combined with HOXA13 CUT-and-RUN data from a similar stage to find putative direct targets of these genes in the GT. This would also show which genes are common targets of ISL1 and HOX13, including genes associated with shared ISL1+HOX13 binding loci. It would be possible to perform ISL1 ChIP-seq using the genitalia of *Hoxa13* and *Hoxd13* double knockouts at E12.5 when these mice can be easily distinguished phenotypically. This may reveal whether there is a widespread requirement of HOX13 factors for ISL1 to bind to certain sites, particularly at sites containing apparent HOX13 motifs.

In limb cells, different HOX factors appear to have largely similar binding affinities and overlapping binding loci, with differences seemingly mediated in part by the presence of cofactors (Jerković et al., 2017). ATAC-seq in the distal limb buds of wildtype and *Hoxa/d13* null mice reveals sites of HOX13-dependent chromatin accessibility, suggesting a pioneering role for these factors (Desanlis et al., 2020). Of 2,241 HOX13-dependent accessible sites identified in this study, 348 (~15%) overlap with sites occupied by both ISL1 and HOXA13 in the GT. Despite the fact that there are far more ISL1-only peaks identified in the GT (Fig. 3.2), these only overlap with 93 (~4%) of these HOX13-dependent sites from the limb. This is consistent with the idea that HOX13 factors are playing a similar role in allowing for chromatin accessibility at certain enhancers in the GT, some of which overlap with enhancers in the distal limbs, while others are likely GT-specific. Parallel experiments in the GT could test this.

ISL1 itself also has the potential to play a pioneering role which has not been explored in the hindlimb field or the GT. ISL1 is capable of binding to compacted chromatin in vitro, and forms complexes with the BRG1-BAF60C SWI/SNF chromatin remodeling complex in cardiac progenitor cells (Gao et al., 2019). ATAC-seq found that among accessible sites occupied by ISL1, a widespread diminishment of accessibility occurs upon deletion of *Isl1* or of *Brg1*, while

the same does not occur at accessible sites not occupied by ISL1. Co-immunoprecipitation indicates interaction between ISL1 and BRG1. While this study reported a similar number of ISL1 peaks as our GT ChIP-seq (>14k), a very small percentage of these overlap (~2%), suggesting ISL1 overwhelmingly binds to different sites in cardiac and appendage development.

Analogous experiments could be carried out in the hindlimb field or GT to find sites that may depend on ISL1-directed chromatin remodeling in this context. ChIP-seq data from BRG1 might find sites co-bound with ISL1, while co-immunoprecipitation of ISL1 and other factors like BRG1 or HOX proteins may indicate direct interactions in GT or hindlimb field protein extract. ATAC-seq in mutant and control mice might identify ISL1-dependent sites of accessibility. H3K27ac ChIP-seq could potentially be performed in *Isl1* cKO genitalia, though it would be implausible in *Isl1* hindlimb fields given the small amount of tissue and the inability to phenotypically differentiate the mutants at E9.75.

*Research on particular enhancers*

This dissertation focuses on enhancers and genes that lie downstream of *Isl1*. A question that remains open for investigation is what regulates *Isl1* expression in both the hindlimb field and the GT. In the hindlimb field the CR2 enhancer drives expression, and deletion of putative SMAD binding sites within this enhancer ablates this domain of activity (Jurberg et al., 2013). The activity of this enhancer during hindlimb initiation is corroborated by our H3K27ac data (Fig. S2.1). Recent ChIP-seq analyses have been performed on SMAD4 in the forelimb buds using a FLAG epitope tag (Gamart et al., 2021; Malkmus et al., 2021). Repeating this in hindlimb field might validate binding to the CR2 enhancer, while multi-species alignments might reveal specific binding sites that are highly conserved.

This could clarify the connection between paraxial GDF11 signaling and *Isl1* activation in hindlimb initiation. Furthermore, deletion of this enhancer might reveal whether it is required for normal *Isl1* expression and hindlimb development. Based on published ATAC-seq signal tracks in the cloacal region at E10.5 and E13.5 (Amândio et al., 2020), this enhancer also may be active during genital initiation and outgrowth. A challenge in investigating the upstream regulation of *Isl1*, however, is the vast gene desert in which it resides. Roughly 1.2 Mb separates *Isl1* and the next gene *Itga1* in the mouse genome, with another ~0.5 Mb upstream. Within these regions are numerous ATAC-seq peaks as well as ISL1 binding sites in the GT, some of which are conserved across amniotes.

*Isl1* may be involved in an autoregulatory loop to maintain its own GT expression, though this would not explain how it is initially activated downstream of cloacal *Shh*. ChIP-seq on GLI proteins may indicate a mechanism by which hedgehog signaling activates mesenchymal *Isl1*. Following an approach used for studying the regulation of the *Hoxa* and *Hoxd* loci by the Duboule lab (Amândio et al., 2020; Lonfat et al., 2014), 4C could be used with the *Isl1* promoter as a viewpoint. A combination of open chromatin, long-range contacts with the *Isl1* promoter, sequence conservation, and transcription factor binding might narrow down putative enhancers for further testing via transgenic assays.

Downstream of *Isl1*, more work could be done to characterize enhancers with conserved binding across amniotes. An intriguing example that appears in both the hindlimb and GT data is the *Bmp4* enhancer CONS3.8, initially identified as a limb and orofacial enhancer (Jumlongras et al., 2012). Similarly to HLEB, this appears to be both a conserved and shared enhancer between limbs and GT, with E10.5 cloacal ATAC-seq suggesting activity prior to GT outgrowth. Research akin to that previously conducted on HLEB (Infante et al., 2015) might suggest its role

in regulating *Bmp4* transcription. As with the *Isl1* locus, the large genomic space downstream of *Bmp4* suggests the possibility of multiple, partly redundant enhancers involved in its regulation.

*Investigating lineage-specific differences in gene regulation*

Another line of research concerning the *Bmp4* locus focuses on phallus reduction in the chicken. Increased apoptosis in the chick GT relative to the duck GT seems to be related to increased expression of *Bmp4* (Herrera et al., 2013). Members of the Cohn lab provided GT samples from the early and late GT of the chick, duck, and turtle (*T. scripta*). In chick and duck, the early GTs are from HH29 and late GTs from HH35, roughly the stage when phallus reduction begins. The turtle samples come from stage 14 and stage 16-17, along with stage 14 forelimbs and hindlimb. I performed H3K27ac ChIP-seq in duplicate using each of these samples, with the processed data available for further analysis on the Menke lab external drive.

These data could be used to find enhancers at the *Bmp4* locus that show unique activity in the chick GT at HH35 as compared to the other samples. Using the approach of Infante et al. (2015), reproducible H3K27ac peaks within 1 kb of each other could be merged to generate putative active enhancers in each sample. Conversion between chick and duck loci can be done on the UCSC test browser, allowing for the identification of conserved enhancers within birds. The alligator H3K27ac data used in Chapter 3 of this dissertation may also be useful, as crocodilians are the closest outgroup to birds and alligator coordinates can also be lifted over to the chicken genome. I only performed ChIP-seq in the alligator GT at stage 15-16, which is closer to later GT development.

The red-eared slider turtle is a more distantly related outgroup to birds, and its genome is not currently on the USCS genome browser for LiftOver. One possible workflow would be to sequentially find enhancers that fit the following criteria: 1) active in the late but not early chick

GT; 2) conserved in the duck genome and inactive at either stage; 3) conserved in the alligator genome and inactive; 4) conserved in the turtle genome and inactive at either stage. Each step would narrow down the set of loci, and step 4 could be done by using Blast to convert particular loci of interest to the turtle genome. Each step might yield loci of interest near *Bmp4*, though loci that show conservation across these species and pass through these filtration steps might be particularly compelling. This might identify a conserved *Bmp4* enhancer that is coopted during chicken phallus reduction, perhaps an enhancer involved in apoptosis in the digits. One benefit of such a high-throughput approach is that it might reveal other enhancers of interest near genes not initially considered.

Species-specific or lineage-specific functions of ISL1 could be also be probed using the data generated in this dissertation. The ISL1 ChIP-seq data from the genitalia of multiple species of amniotes was primarily used to identify conserved functions of ISL1. These data may also be a resource for studying enhancer evolution, as with the putative ISL1 site found in lizards but not in other amniotes (Fig. S3.8). A more systematic approach might use a whole genome alignment including the brown anole along with chicken and mouse. This would allow for the conversion of, for example, lizard genomic coordinates into other species to find the subset of lizard ISL1 peaks that occur on conserved loci. Conversion can be performed using the halLiftover tool using a HAL alignment file and bed coordinates as input.

Rather than finding conserved peaks, this could then be used to find peaks unique to one species that occur within orthologous sequence. These loci could then be analyzed for the presence of motifs of interest like the ISL1 motif in each species, identifying potential gain or loss of function in a particular lineage. Such sites could then be associated with nearby genes of interest that may be regulated by these enhancers. These enhancers as well as those with

conserved activity could be tested using CRISPR knockout experiments in the brown anole. Given the dramatically different morphology of the paired hemiphallus in squamates, this species provides an intriguing model for further studies along these lines.

# REFERENCES


Agarwal, P., Wylie, J.N., Galceran, J., Arkhitko, O., Li, C., Deng, C., Grosschedl, R., and Bruneau, B.G. (2003). Tbx5 is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo. Development *130*, 623–633.

Ahn, D., Kourakis, M.J., Rohde, L.A., Silver, L.M., and Ho, R.K. (2002). T-box gene tbx5 is essential for formation of the pectoral limb bud. Nature *417*, 754–758.

Ahn, K., Mishina, Y., Hanks, M.C., Behringer, R.R., and Bryan Crenshaw, E. (2001). BMPR-IA signaling is required for the formation of the apical ectodermal ridge and dorsal-ventral patterning of the limb. Development *128*, 4449–4461.

Amândio, A.R., Lopez-Delisle, L., Bolt, C.C., Mascrez, B., and Duboule, D. (2020). A complex regulatory landscape involved in the development of mammalian external genitals. Elife *9*, e52962.

Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., and Duboule, D. (2013). A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science *340*, 1234167.

Andrey, G., Schöpflin, R., Jerkovic, I., Heinrich, V., Ibrahim, D.M., Paliou, C., Hochradel, M., Timmermann, B., Haas, S., Vingron, M., et al. (2017). Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. Genome Res. *27*, 223–233.

Armfield, B.A., and Cohn, M.J. (2021). Single cell transcriptomic analysis of external genitalia reveals complex and sexually dimorphic cell populations in the early genital tubercle.

Dev. Biol. *477*, 145–154.

Bailey, T.L., and Machanick, P. (2012). Inferring direct DNA binding from ChIP-seq. Nucleic Acids Res. *40*, e128.

Benazet, J.D., and Zeller, R. (2013). Dual requirement of ectodermal Smad4 during AER formation and termination of feedback signaling in mouse limb buds. Genesis *51*, 660–666.

Bénazet, J.D., Bischofberger, M., Tiecke, E., Gonçalves, A., Martin, J.F., Zuniga, A., Naef, F., and Zeller, R. (2009). A self-regulatory system of interlinked signaling feedback loops controls mouse limb patterning. Science *323*, 1050–1053.

Bénazet, J.D., Pignatti, E., Nugent, A., Unal, E., Laurent, F., and Zeller, R. (2012). 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.

Berenguer, M., and Duester, G. (2021). Role of retinoic acid signaling, FGF signaling and Meis genes in control of limb development. Biomolecules *11*, 80.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.

Brazeau, M.D., and Friedman, M. (2015). The origin and early phylogenetic history of jawed vertebrates. Nature *520*, 490–497.

Buenrostro, J.D., and Giresi, P.G. (2013). Transposition of native chromatin for multimodal regulatory analysis and personal epigenomics. Nat. Methods *10*, 1213–1218.

Buenrostro, J., Wu, B., Chang, H., and Greenleaf, W. (2015). ATAC-seq: a method for assaying

chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. *109*, 21.29.1-21.29.9.

Cai, C.L., Liang, X., Shi, Y., Chu, P.H., Pfaff, S.L., Chen, J., and Evans, S. (2003). Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. Dev. Cell *5*, 877–889.

Chan, Y.F., Marks, M.E., Jones, F.C., Jr, G.V., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., et al. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science *327*, 302–305.

Cheatle Jarvela, A.M., and Hinman, V.F. (2015). Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks. Evodevo *6*, 3.

Chiang, C., Litingtung, Y., Lee, E., Young, K.E., Corden, J.L., Westphal, H., and Beachy, P. a (1996). Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. Nature *383*, 407–413.

Ching, S.T., Infante, C.R., Du, W., Sharir, A., Park, S., Menke, D.B., and Klein, O.D. (2018). Isl1 mediates mesenchymal expansion in the developing external genitalia via regulation of Bmp4, Fgf10 and Wnt5a. Hum. Mol. Genet. *27*, 107–119.

Cohn, M.J. (2011). Development of the external genitalia: conserved and divergent mechanisms of appendage patterning. Dev. Dyn. *240*, 1108–1115.

Cohn, M.J., Izpisúa-Belmonte, J.C., Abud, H., Heath, J.K., and Tickle, C. (1995). Fibroblast growth factors induce additional limb development from the flank of chick embryos. Cell *80*, 739–746.

Cole, N.J., Tanaka, M., Prescott, A., and Tickle, C. (2003). Expression of limb initiation genes

and clues to the morphological diversification of threespine stickleback. Curr. Biol. *13*, 951–952.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M. a, Frampton, G.M., Sharp, P. a, et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. *107*, 21931–21936.

Crossley, P.H., and Martin, G.R. (1995). The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. Development *121*, 439–451.

Crossley, P.H., Minowada, G., MacArthur, C.A., and Martin, G.R. (1996). Roles for FGF8 in the induction, initiation, and maintenance of chick limb development. Cell *84*, 127–136.

DeLaurier, A., Schweitzer, R., and Logan, M. (2006). Pitx1 determines the morphology of muscle, tendon, and bones of the hindlimb. Dev. Biol. *299*, 22–34.

Delgado, I., López-Delgado, A.C., Alberto, R.D., Giovinazzo, G., Cadenas, V., Fernández-De-Manuel, L., Sánchez-Cabo, F., Anderson, M.J., Lewandoski, M., and Torres, M. (2020). Proximo-distal positional information encoded by an Fgf-regulated gradient of homeodomain transcription factors in the vertebrate limb. Sci. Adv. *6*, eaaz0742.

Delgado, I., Giovinazzo, G., Temiño, S., Gauthier, Y., Balsalobre, A., Drouin, J., and Torres, M. (2021). Control of mouse limb initiation and antero-posterior patterning by Meis transcription factors. Nat. Commun. *12*, 3086.

Desanlis, I., Kherdjemil, Y., Mayran, A., Bouklouch, Y., Gentile, C., Sheth, R., Zeller, R., Drouin, J., and Kmita, M. (2020). HOX13-dependent chromatin accessibility underlies the transition towards the digit development program. Nat. Commun. *11*, 2491.

Domyan, E.T., Kronenberg, Z., Infante, C.R., Vickrey, A.I., Stringham, S.A., Bruders, R., Guernsey, M.W., Park, S., Payne, J., Beckstead, R.B., et al. (2016). Molecular shifts in limb identity underlie development of feathered feet in two domestic avian species. Elife *5*, 1–21.

Don, E.K., de Jong-Curtain, T.A., Doggett, K., Hall, T.E., Heng, B., Badrock, A.P., Winnick, C., Nicholson, G.A., Guillemin, G.J., Currie, P.D., et al. (2016). Genetic basis of hindlimb loss in a naturally occurring vertebrate model. Biol. Open *5*, 359–366.

Douglas, N.C., Heng, K., Sauer, M. V., and Papaioannou, V.E. (2012). Dynamic expression of Tbx2 subfamily genes in development of the mouse reproductive system. Dev. Dyn. *241*, 365–375.

Draaken, M., Knapp, M., Pennimpede, T., Schmidt, J.M., Ebert, A.K., Rösch, W., Stein, R., Utsch, B., Hirsch, K., Boemers, T.M., et al. (2015). Genome-wide association study and meta-analysis identify ISL1 as genome-wide significant susceptibility sene for bladder exstrophy. PLoS Genet. *11*, e1005024.

Duboc, V., and Logan, M.P.O. (2011). Regulation of limb bud initiation and limb-type morphology. Dev. Dyn. *240*, 1017–1027.

Duboc, V., Sulaiman, F., Feneck, E., Kucharska, A., Bell, D., Holder-Espinasse, M., and Logan, M.P.O. (2021). Tbx4 function during hindlimb development reveals a novel mechanism to explain the origins of proximal limb defects. Development *148*, dev199580.

Ediger, B.N., Lim, H., Juliana, C., Groff, D.N., Williams, L.T., Dominguez, G., Liu, J., Taylor, B.L., Walp, E.R., Kameswaran, V., et al. (2017). LIM domain-binding 1 maintains the terminally differentiated state of pancreatic β cells. J. Clin. Invest. *127*, 215–229.

Ericson, J., Thor, S., Edlund, T., Jessell, T.M., and Yamada, T. (1992). Early stages of motor

neuron differentiation revealed by expression of homeobox gene Islet-1. Science *256*, 1555–1560.

Fallon, J.F., López, A., Ros, M.A., Savage, M.P., Olwin, B.B., Simandl, B.K., Fallon, J.F., Lopez, A., Ros, M.A., Savage, M.P., et al. (1994). FGF-2: apical ectodermal ridge growth signal for chick limb development. Science *264*, 104–107.

Farrera-Hernández, A., Marín-Llera, J.C., and Chimal-Monroy, J. (2021). WNT5A-Ca2+-CaN-NFAT signalling plays a permissive role during cartilage differentiation in embryonic chick digit development. Dev. Biol. *469*, 86–95.

Feneck, E., and Logan, M. (2020). The role of retinoic acid in establishing the early limb bud. Biomolecules *10*, 312.

Ferrari, D., Harrington, A., Dealy, C.N., and Kosher, R.A. (1999). Dlx-5 in limb initiation in the chick embryo. Dev. Dyn. *216*, 10–15.

Freyd, G., Kim, S.K., and Horvitz, H.R. (1990). Novel cysteine-rich motif and homeodomain in the product of the Caenorhabditis elegans cell lineage gene lin-11. Nature *344*, 867–879.

Gadd, M.S., Bhati, M., Jeffries, C.M., Langley, D.B., Trewhella, J., Guss, J.M., and Matthews, J.M. (2011). Structural basis for partial redundancy in a class of transcription factors, the LIM homeodomain proteins, in neural cell type specification. J. Biol. Chem. *286*, 42971–42980.

Galceran, J., Fariñas, I., Depew, M.J., Clevers, H., and Grosschedl, R. (1999). Wnt3a(-/-)-like phenotype and limb deficiency in Lef1(-/-)Tcf1(-/-) mice. Genes Dev. *13*, 709–717.

Galli, A., Robay, D., Osterwalder, M., Bao, X., Bénazet, J.D., Tariq, M., Paro, R., Mackem, S., and Zeller, R. (2010). Distinct roles of Hand2 in initiating polarity and posterior Shh expression during the onset of mouse limb bud development. PLoS Genet. *6*, e1000901.

Gamart, J., Barozzi, I., Laurent, F., Reinhardt, R., Martins, L.R., Oberholzer, T., Visel, A., Zeller, R., and Zuniga, A. (2021). SMAD4 target genes are part of a transcriptional network that integrates the response to BMP and SHH signaling during early limb bud patterning. Development *148*, dev200182.

Gao, B., Ajima, R., Yang, W., Li, C., Song, H., Anderson, M.J., Liu, R.R., Lewandoski, M.B., Yamaguchi, T.P., and Yang, Y. (2018). Coordinated directional outgrowth and pattern formation by integration of wnt5a and Fgf signaling in planar cell polarity. Development *145*, dev163824.

Gao, R., Liang, X., Cheedipudi, S., Cordero, J., Jiang, X., Zhang, Q., Caputo, L., Günther, S., Kuenne, C., Ren, Y., et al. (2019). Pioneering function of Isl1 in the epigenetic control of cardiomyocyte cell fate. Cell Res. *29*, 486–501.

Garrity, D.M., Childs, S., and Fishman, M.C. (2002). The heartstrings mutation in zebrafish causes heart/fin Tbx5 deficiency syndrome. Development *129*, 4635–4645.

Geetha-Loganathan, P., Nimmagadda, S., and Scaal, M. (2008). Wnt signaling in limb organogenesis. Organogenesis *4*, 109–115.

Georgas, K.M., Armstrong, J., Keast, J.R., Larkins, C.E., McHugh, K.M., Southard-Smith, E.M., Cohn, M.J., Batourina, E., Dan, H., Schneider, K., et al. (2015). An illustrated anatomical ontology of the developing mouse lower urogenital tract. Dev. *142*, 1893–1908.

Gibson-Brown, J.J., Agulnikb, S.I., Chapmana, D.L., Alexioua, M., Garveyb, N., Silverb, L.M., and Papaioannoua, V.E. (1996). Evidence of a role for T-box genes in the evolution of limb morphogenesis and the specification of forelimb/hindlimb identity. Mech. Dev. *56*, 93–101.

Gibson-Brown, J.J., Agulnik, S.I., Silver, L.M., Niswander, L., and Papaioannou, V.E. (1998).

Involvement of T-box genes Tbx2-Tbx5 in vertebrate limb specification and development. Development *125*, 2499–2509.

Glassford, W.J., Johnson, W.C., Dall, N.R., Smith, S.J., Liu, Y., Boll, W., Noll, M., and Rebeiz, M. (2015). Co-option of an ancestral Hox-regulated network underlies a recently evolved morphological novelty. Dev. Cell *34*, 520–531.

Gredler, M.L. (2016). Developmental and evolutionary origins of the amniote phallus. Integr. Comp. Biol. *56*, 694–704.

Gredler, M.L., Larkins, C.E., Leal, F., Lewis, A.K., Herrera, A.M., Perriton, C.L., Sanger, T.J., and Cohn, M.J. (2014). Evolution of external genitalia: Insights from reptilian development. Sex. Dev. *8*, 311–326.

Gredler, M.L., Seifert, A.W., and Cohn, M.J. (2015a). Morphogenesis and patterning of the phallus and cloaca in the American alligator, Alligator mississippiensis. Sex. Dev. *9*, 53–67.

Gredler, M.L., Sanger, T.J., and Cohn, M.J. (2015b). Development of the cloaca, hemipenes, and hemiclitores in the green anole, Anolis carolinensis. Sex. Dev. *9*, 21–33.

Greenbaum, E. (2002). A standardized series of embryonic stages for the emydid turtle Trachemys scripta. Can. J. Zool. *80*, 1350–1370.

Gros, J., and Tabin, C.J. (2014). Vertebrate limb bud formation is initiated by localized epithelial-to-mesenchymal transition. Science *343*, 1253–1256.

Gros, J., Hu, J.K.H., Vinegoni, C., Feruglio, P.F., Weissleder, R., and Tabin, C.J. (2010). WNT5A/JNK and FGF/MAPK pathways regulate the cellular events shaping the vertebrate limb bud. Curr. Biol. *20*, 1993–2002.

Guerreiro, I., Gitto, S., Novoa, A., Codourey, J., Nguyen Huynh, T.H., Gonzalez, F.,

Milinkovitch, M.C., Mallo, M., and Duboule, D. (2016). Reorganisation of Hoxd regulatory landscapes during the evolution of a snake-like body plan. Elife *5*, e16087.

Hajihosseini, M.K., and Heath, J.K. (2002). Expression patterns of fibroblast growth factors-18 and -20 in mouse embryos is suggestive of novel roles in calvarial and limb development. Mech. Dev. *113*, 79–83.

Haller, M., and Ma, L. (2019). Temporal, spatial, and genetic regulation of external genitalia development. Differentiation *110*, 1–7.

Hamburger, V. (1938). Morphogenetic and axial self-differentiation of transplanted limb primordia of 2-day chick embryos. J. Exp. Zool. *77*, 379–399.

Hamburger, V., and Hamilton, H.L. (1992). A series of normal stages in the development of the chick embryo. Dev. Dyn. *195*, 231–272.

Haraguchi, R., Suzuki, K., Murakami, R., Sakai, M., Kamikawa, M., Kengaku, M., Sekine, K., Kawano, H., Kato, S., Ueno, N., et al. (2000). Molecular analysis of external genitalia formation: the role of fibroblast growth factor (Fgf) genes during genital tubercle formation. Development *127*, 2471–2479.

Haraguchi, R., Mo, R., Hui, C., Motoyama, J., Makino, S., Shiroishi, T., Gaffield, W., and Yamada, G. (2001). Unique functions of Sonic hedgehog signaling during external genitalia development. Development *128*, 4241–4250.

Hashimoto, D., Hyuga, T., Acebedo, A.R., Alcantara, M.C., Suzuki, K., and Yamada, G. (2019). Developmental mutant mouse models for external genitalia formation. Congenit. Anom. (Kyoto). *59*, 74–80.

Heikinheimo, M., Lawshé, A., Shackleford, G.M., Wilson, D.B., and MacArthur, C.A. (1994). Fgf-8 expression in the post-gastrulation mouse suggests roles in the development of the

face, limbs and central nervous system. Mech. Dev. *48*, 129–138.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Herrera, A.M., Shuster, S.G., Perriton, C.L., and Cohn, M.J. (2013). Developmental basis of phallus reduction during bird evolution. Curr. Biol. *23*, 1065–1074.

Hoyt, P.R., Bartholomew, C., Davis, A.J., Yutzey, K., Gamer, L.W., Potter, S.S., Ihle, J.N., and Micenski, M.L. (1997). The Evi1 proto-oncogene is required at midgestation for neural, heart, and paraxial mesenchyme development. Mech. Dev. *65*, 55–70.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Hunter, C.S., and Rhodes, S.J. (2005). LIM-homeodomain genes in mammalian development and human disease. Mol. Biol. Rep. *32*, 67–77.

Infante, C.R., Park, S., Mihala, A.G., Kingsley, D.M., and Menke, D.B. (2013). Pitx1 broadly associates with limb enhancers and is enriched on hindlimb cis-regulatory elements. Dev. Biol. *374*, 234–244.

Infante, C.R., Mihala, A.G., Park, S., Wang, J.S., Johnson, K.K., Lauderdale, J.D., and Menke, D.B. (2015). Shared enhancer activity in the limbs and phallus and functional divergence of a limb-genital cis-regulatory element in snakes. Dev. Cell *35*, 107–119.

Isaac, A., Rodriguez-Esteban, C., Ryan, A., Altabef, M., Tsukui, T., Patel, K., Tickle, C., and Izpisúa-Belmonte, J.C. (1998). Tbx genes and limb identity in chick embryo development. Development *125*, 1867–1875.

Itou, J., Kawakami, H., Quach, T., Osterwalder, M., Evans, S.M., Zeller, R., and Kawakami, Y. (2012). Islet1 regulates establishment of the posterior hindlimb field upstream of the Hand2-Shh morphoregulatory gene network in mouse embryos. Development *139*, 1620–1629.

Jain, D., Nemec, S., Luxey, M., Gauthier, Y., Bemmo, A., Balsalobre, A., and Drouin, J. (2018). Regulatory integration of Hox factor activity with T-box factors in limb development. Development *145*, dev159830.

Jerković, I., Ibrahim, D.M., Andrey, G., Haas, S., Hansen, P., Janetzki, C., González Navarrete, I., Robinson, P.N., Hecht, J., Mundlos, S., et al. (2017). Genome-wide binding of posterior HOXA/D transcription factors reveals subgrouping and association with CTCF. PLOS Genet. *13*, e1006567.

Jhanwar, S., Malkmus, J., Stolte, J., Romashkina, O., Zuniga, A., and Zeller, R. (2021). Conserved and species-specific chromatin remodeling and regulatory dynamics during mouse and chicken limb bud development. Nat. Commun. *12*, 5685.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1503.

Jumlongras, D., Lachke, S.A., Connell, D.J.O., Aboukhalil, A., Li, X., Choe, S.E., Ho, J.W.K., Turbe-doan, A., Robertson, E.A., Olsen, B.R., et al. (2012). An evolutionarily conserved enhancer regulates Bmp4 expression in developing incisor and limb bud. PLoS One *7*, e38568.

Jurberg, A.D., Aires, R., Varela-Lasheras, I., Nóvoa, A., and Mallo, M. (2013). Switching axial progenitors from producing trunk to tail tissues in vertebrate embryos. Dev. Cell *25*, 451–462.

Kajioka, D., Suzuki, K., Nakada, S., Matsushita, S., Miyagawa, S., Takeo, T., Nakagata, N., and Yamada, G. (2020). Bmp4 is an essential growth factor for the initiation of genital tubercle (GT) outgrowth. Congenit. Anom. (Kyoto). *8*, 15–21.

Kaku, Y., Ohmori, T., Kudo, K., Fujimura, S., Suzuki, K., Evans, S.M., Kawakami, Y., and Nishinakamura, R. (2013). Islet1 deletion causes kidney agenesis and hydroureter resembling CAKUT. J. Am. Soc. Nephrol. *24*, 1242–1249.

Karlsson, O., Thor, S., Norberg, T., Ohlsson, H., and Edlund, T. (1990). Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. Nature *344*, 879–882.

Kawakami, Y., Capdevila, J., Bu, D., Itoh, T., Izpisu, J.C., and Jolla, L. (2001). WNT signals control FGF-dependent limb initiation and AER induction in the chick embryo. Cell *104*, 891–900.

Kawakami, Y., Marti, M., Kawakami, H., Itou, J., Quach, T., Johnson, A., Sahara, S., O'Leary, D.D.M., Nakagawa, Y., Lewandoski, M., et al. (2011). Islet1-mediated activation of the β-catenin pathway is necessary for hindlimb initiation in mice. Development *138*, 4465–4473.

Kondo, T., Zákány, J., Innis, J., and Duboule, D. (1997). Of fingers, toes and penises. Nature *390*, 29.

Kuijper, S., Feitsma, H., Sheth, R., Korving, J., Reijnen, M., and Meijlink, F. (2005). Function and regulation of Alx4 in limb development: complex genetic interactions with Gli3 and Shh. Dev. Biol. *285*, 533–544.

Kvon, E.Z., Kamneva, O.K., Melo, U.S., Barozzi, I., Osterwalder, M., Mannion, B.J., Tissi??res, V., Pickle, C.S., Plajzer-Frick, I., Lee, E.A., et al. (2016). Progressive loss of function in

a limb enhancer during snake evolution. Cell *167*, 633–642.

Kvon, E.Z., Waymack, R., Gad, M., and Wunderlich, Z. (2021). Enhancer redundancy in development and disease. Nat. Rev. Genet. *22*, 324–336.

Lanctôt, C., Lamolet, B., and Drouin, J. (1997). The bicoid-related homeoprotein Ptx1 defines the most anterior domain of the embryo and differentiates posterior from anterior lateral mesoderm. Development *124*, 2807–2817.

Lanctôt, C., Moreau, A., Chamberland, M., Tremblay, M.L., and Drouin, J. (1999). Hindlimb patterning and mandible development require the Ptx1 gene. Development *126*, 1805–1810.

Larkins, C.E., and Cohn, M.J. (2015). Phallus development in the turtle Trachemys scripta. Sex. Dev. *9*, 34–42.

Leal, F., and Cohn, M.J. (2014). Development of hemipenes in the ball python snake python regius. Sex. Dev. *9*, 6–20.

Leal, F., and Cohn, M.J. (2016). Loss and re-emergence of legs in snakes by modular evolution of Sonic hedgehog and HOXD enhancers. Curr. Biol. *26*, 2966–2973.

Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. *12*, 1725–1735.

Lettice, L.A., Williamson, I., Wiltshire, J.H., Peluso, S., Devenney, P.S., Hill, A.E., Essafi, A., Hagman, J., Mort, R., Grimes, G., et al. (2012). Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. Dev. Cell *22*, 459–467.

Li, D., Sakuma, R., Vakili, N.A., Mo, R., Puviindran, V., Deimling, S., and Zhang, X. (2014). Formation of proximal and anterior limb skeleton requires early function of Irx3 and Irx5 and is negatively regulated by Shh signaling. Dev. Cell *29*, 233–240.

Li, S., Wan, C., Zheng, R., Fan, J., Dong, X., Meyer, A., and Liu, X.S. (2019). Cistrome-GO: a web server for functional enrichment analysis of transcription factor ChIP-seq peaks. Nucleic Acids Res. *47*, W206–W211.

Lin, C., Yin, Y., Long, F., and Ma, L. (2008). Tissue-specific requirements of β-catenin in external genitalia development. Development *135*, 2815–2825.

Lin, C., Yin, Y., Veith, G.M., Fisher, A. V, Long, F., and Ma, L. (2009). Temporal and spatial dissection of Shh signaling in genital tubercle development. Development *136*, 3959–3967.

Lin, C., Yin, Y., Bell, S.M., Veith, G.M., Chen, H., Huh, S.H., Ornitz, D.M., and Ma, L. (2013). Delineating a conserved genetic cassette promoting outgrowth of body appendages. PLoS Genet. *9*, e1003231.

Liu, Y., Liu, C., Yamada, Y., and Fan, C. (2002). Growth arrest specific gene 1 acts as a region-specific mediator of the Fgf10/Fgf8 regulatory loop in the limb. Development *129*, 5289–5300.

Logan, M., and Tabin, C.J. (1999). Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity. Science *283*, 1736–1739.

Logan, M., Simon, H.G., and Tabin, C. (1998). Differential regulation of T-box and homeobox transcription factors suggests roles in controlling chick limb-type identity. Development *125*, 2825–2835.

Logan, M., Martin, J.F., Nagy, A., Lobe, C., Olson, E.N., and Tabin, C.J. (2002). Expression of

Cre recombinase in the developing mouse limb bud driven by a Prxl enhancer. Genesis *33*, 77–80.

Lonfat, N., Montavon, T., Gitto, S., and Duboule, D. (2014). Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. Science *346*, 1004–1007.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell *167*, 1170–1187.

Lowe, L.A., Yamada, S., and Kuehn, M.R. (2000). Hoxb6-Cre transgenic mice express Cre recombinase in extra-embryonic mesoderm, in lateral plate and limb mesoderm and at the midbrain/hindbrain junction. Genesis *26*, 118–120.

Maccabe, A.B., Gasseling, M.T., and Saunders, J.W. (1973). Spatiotemporal distribution of mechanisms that control outgrowth and anteroposterior polarization of the limb bud in the chick embryo. Mech. Ageing Dev. *2*, 1–12.

Mahmood, R., Bresnick, J., Hornbruch, A., Mahony, C., Morton, N., Colquhoun, K., Martin, P., Lumsden, A., Dickson, C., and Mason, I. (1995). A role for FGF-8 in the initiation and maintenance of vertebrate limb bud outgrowth. Curr. Biol. *5*, 797–806.

Malkmus, J., Ramos Martins, L., Jhanwar, S., Kircher, B., Palacio, V., Sheth, R., Leal, F., Duchesne, A., Lopez-Rios, J., Peterson, K.A., et al. (2021). Spatial regulation by multiple Gremlin1 enhancers provides digit development with cis-regulatory robustness and evolutionary plasticity. Nat. Commun. *12*, 5557.

Mao, J., Mcglinn, E., Huang, P., Tabin, C.J., and Mcmahon, A.P. (2009). Fgf-dependent Etv4/5 activity is required for posterior restriction of Sonic hedgehog and promoting outgrowth of the vertebrate limb. Dev. Cell *16*, 600–606.

Marazzi, G., Wang, Y., and Sassoon, D. (1997). Msx2 is a transcriptional regulator in the Bmp4-

mediated programmed cell death pathway. Dev. Biol. *186*, 127–138.

Martin, J.F., Bradley, A., and Eric, N.O. (1995). The paired-like homeo box gene MHox is required for early events of skeletogenesls in multiple lineages. Genes Dev. *9*, 1237–1249.

Matsubara, H., Saito, D., Abe, G., Yokoyama, H., Suzuki, T., and Tamura, K. (2017a). Upstream regulation for initiation of restricted Shh expression in the chick limb bud. Dev. Dyn. *246*, 417–430.

Matsubara, Y., Hirasawa, T., Egawa, S., Hattori, A., Suganuma, T., Kohara, Y., Nagai, T., Tamura, K., Kuratani, S., Kuroiwa, A., et al. (2017b). Anatomical integration of the sacral-hindlimb unit coordinated by GDF11 underlies variation in hindlimb positioning in tetrapods. Nat. Ecol. Evol. *1*, 1392–1399.

Mazzoni, E.O., Mahony, S., Closser, M., Morrison, C.A., Nedelec, S., Williams, D.J., An, D., Gifford, D.K., and Wichterle, H. (2013). Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. Nat. Neurosci. *16*, 1219–1227.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

McLean, C.Y., Reno, P.L., Pollen, A.A., Bassan, A.I., Capellini, T.D., Guenther, C., Indjeian, V.B., Lim, X., Menke, D.B., Schaar, B.T., et al. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature *471*, 216–219.

Menke, D., Guenther, C., and Kingsley, D. (2008). Dual hindlimb control elements in the Tbx4 gene and region-specific control of bone size in vertebrate limbs. Development *135*, 2543–2553.

Min, H., Danilenko, D.M., Scully, S.A., Bolon, B., Ring, B.D., Tarpley, J.E., DeRose, M., and Simonet, W.S. (1998). Fgf-10 is required for both limb and lung development and exhibits striking functional similarity to Drosophila branchless. Genes Dev. *12*, 3156–3161.

Minguillon, C., Del Buono, J., and Logan, M.P. (2005). Tbx5 and Tbx4 are not sufficient to determine limb-specific morphologies but have common roles in initiating limb outgrowth. Dev. Cell *8*, 75–84.

Minguillon, C., Gibson-Brown, J.J., and Logan, M.P. (2009). Tbx4/5 gene duplication and the origin of vertebrate paired appendages. Proc. Natl. Acad. Sci. U. S. A. *106*, 21726–21730.

Minowada, G., Jarvis, L.A., Chi, C.L., Neubüser, A., Sun, X., Hacohen, N., Krasnow, M.A., and Martin, G.R. (1999). Vertebrate Sprouty genes are induced by FGF signaling and can cause chondrodysplasia when overexpressed. Development *126*, 4465–4475.

Miyagawa, S., Moon, A., Haraguchi, R., Inoue, C., Harada, M., Nakahara, C., Suzuki, K., Matsumaru, D., Kaneko, T., Matsuo, I., et al. (2009). Dosage-dependent hedgehog signals integrated with Wnt/β-catenin signaling regulate external genitalia formation as an appendicular program. Development *136*, 3969–3978.

Murillo-Ferrol, N.L. (1965). Étude causale de la différenciation la plus précoce de l'ébauche morphologique des membres. Analyse expérimentale chez les embryons d'oiseaux. Acta Anat. (Basel). *62*, 80–103.

Naiche, L. a, and Papaioannou, V.E. (2007). Tbx4 is not required for hindlimb identity or post-bud hindlimb outgrowth. Development *134*, 93–103.

Naiche, L.A., and Papaioannou, V.E. (2003). Loss of Tbx4 blocks hindlimb development and

affects vascularization and fusion of the allantois. Development *130*, 2681–2693.

Narkis, G., Tzchori, I., Cohen, T., Holtz, A., Wier, E., and Westphal, H. (2012). Isl1 and Ldb co-regulators of transcription are essential early determinants of mouse limb development. Dev. Dyn. *241*, 787–791.

Nemec, S., Jain, D., Sung, H., Pastinen, T., and Drouin, J. (2017). Pitx1 directly modulates the core limb development program to implement hindlimb identity. Development *144*, 3325–3335.

Ng, J.K., Kawakami, Y., Büscher, D., Raya, Á., Itoh, T., Koth, C.M., Rodríguez Esteban, C., Rodríguez-León, J., Garrity, D.M., Fishman, M.C., et al. (2002). The limb gene Tbx5 promotes limb initiation by interacting with Wnt2b and Fgf10. Development *129*, 5161–5170.

Niederreither, K., Vermot, J., Messaddeq, N., Schuhbaur, B., Chambon, P., and Dollé, P. (1999). Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. Nat. Genet. *21*, 444–448.

Nishimoto, S., Wilde, S.M., Wood, S., and Logan, M.P.O. (2015). RA acts in a coherent feed-forward mechanism with Tbx5 to control limb bud induction and initiation. Cell Rep. *12*, 879–891.

Niswander, L., Tickle, C., Vogel, A., Booth, I., and Martin, G.R. (1993). FGF-4 replaces the apical ectodermal ridge and directs outgrowth and patterning of the limb. Cell *75*, 579–587.

Nohno, T., Koyama, E., Fumio, M., Taniguchi, S., Ohuchi, H., Saito, T., and Noji, S. (1993). A chicken homeobox gene related to Drosophila paired is predominately expressed in the developing limb. Dev. Biol. *158*, 254–264.

Ohuchi, H., Yoshioka, H., Tanaka, A., Kawakami, Y., Nohno, T., and Noji, S. (1994). Involvement of androgen-induced growth factor (Fgf-8) gene in mouse embryogenesis and morphogenesis. Biochem. Biophys. Res. Commun. *204*, 882–888.

Ohuchi, H., Nakagawa, T., Yamauchi, M., Ohata, T., Yoshioka, H., Kuwana, T., Mima, T., Mikawa, T., Nohno, T., and Noji, S. (1995). An additional limb can be induced from the flank of the chick embryo by FGF4. Biochem. Biophys. Res. Commun. *209*, 809–816.

Ohuchi, H., Nakagawa, T., Yamamoto, A., Araga, A., Ohata, T., Ishimaru, Y., Yoshioka, H., Kuwana, T., Nohno, T., Yamasaki, M., et al. (1997). The mesenchymal factor, FGF10, initiates and maintains the outgrowth of the chick limb bud through interaction with FGF8, an apical ectodermal factor. Development *124*, 2235–2244.

Ohuchi, H., Takeuchi, J., Yoshioka, H., Ishimaru, Y., Ogura, K., Takahashi, N., Ogura, T., and Noji, S. (1998). Correlation of wing-leg identity in ectopic FGF-induced chimeric limbs with the differential expression of chick Tbx5 and Tbx4. Development *125*, 51–60.

Osterwalder, M., Speziale, D., Shoukry, M., Mohan, R., Ivanek, R., Kohler, M., Beisel, C., Wen, X., Scales, S.J., Christoffels, V.M., et al. (2014). HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. Dev. Cell *31*, 345–357.

Ouimette, J.-F., Jolin, M.L., L'honoré, A., Gifuni, A., and Drouin, J. (2010). Divergent transcriptional activities determine limb identity. Nat. Commun. *1*, 35.

Pajni-Underwood, S., Wilson, C.P., Elder, C., Mishina, Y., and Lewandoski, M. (2007). BMP signals control limb bud interdigital programmed cell death by regulating FGF signaling. Development *134*, 2359–2368.

Peebles, F. (1911). On the interchange of the limbs of the chick by transplantation. Biol. Bull.

*20*, 14–18.

Perriton, C., Powles, N., Chiang, C., Maconochie, M.M., and Cohn, M.J. (2002). Sonic hedgehog signaling from the urethral epithelium controls external genital development. Dev. Biol. *247*, 26–46.

Person, A.D., Beiraghi, S., Sieben, C.M., Hermanson, S., Neumann, A.N., Robu, M.E., Schleiffarth, J.R., Billington, C.J., Van Bokhoven, H., Hoogeboom, J.M., et al. (2010). WNT5A mutations in patients with autosomal dominant Robinow syndrome. Dev. Dyn. *239*, 327–337.

Pfaff, S.L., Mendelsohn, M., Stewart, C.L., Edlund, T., and Jessell, T.M. (1996). Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. Cell *84*, 309–320.

Pinot, M. (1970). Le rôle du mésoderme somitique dans la morphogenèse précoce des membres de l'embryon de poulet. J. Embryol. Exp. Morphol. *23*, 109–151.

Preger-Ben Noon, E., Sabarís, G., Ortiz, D.M., Sager, J., Liebowitz, A., Stern, D.L., and Frankel, N. (2018). Comprehensive analysis of a cis-regulatory region reveals pleiotropy in enhancer function. Cell Rep. *22*, 3021–3031.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature *322*, 697–701.

Rallis, C., Bruneau, B.G., Del Buono, J., Seidman, C.E., Seidman, J.G., Nissim, S., Tabin, C.J., and Logan, M.P.O. (2003). Tbx5 is required for forelimb bud formation and continued outgrowth. Development *130*, 2741–2751.

Reno, P.L., McLean, C.Y., Hines, J.E., Capellini, T.D., Bejerano, G., and Kingsley, D.M. (2013). A penile spine/vibrissa enhancer sequence is missing in modern and extinct humans but

is retained in multiple primates with penile spines and sensory vibrissae. PLoS One *8*, e84258.

Riddle, R.D., Johnson, R.L., Laufer, E., and Tabin, C. (1993). Sonic hedgehog mediates the polarizing activity of the ZPA. Cell *75*, 1401–1416.

Robert, B. (2007). Bone morphogenetic protein signaling in limb outgrowth and patterning. Dev. Growth Differ. *49*, 455–468.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat. Methods *4*, 651–657.

Robledo, R.F., Rajan, L., Li, X., and Lufkin, T. (2002). The Dlx5 and Dlx6 homeobox genes are essential for craniofacial, axial, and appendicular skeletal development. Genes Dev. *16*, 1089–1101.

Rodríguez-Carballo, E., Lopez-Delisle, L., Willemin, A., Beccari, L., Gitto, S., Mascrez, B., and Duboule, D. (2020). Chromatin topology and the timing of enhancer function at the HoxD locus. Proc. Natl. Acad. Sci. *117*, 31231–31241.

Rodriguez-Esteban, C., Tsukui, T., Yonei, S., Magallon, J., Tamura, K., and Belmonte, J.C.I. (1999). The T-box genes Tbx4 and Tbx5 regulate limb outgrowth and identity. Nature *398*, 814–818.

Roux, M., Bouchard, M., and Kmita, M. (2019). Multifaceted Hoxa13 function in urogenital development underlies the hand-foot-genital syndrome. Hum. Mol. Genet. *28*, 1671–1681.

Royle, S.R., Tabin, C.J., and Young, J.J. (2021). Limb positioning and initiation: an evolutionary

context of pattern and formation. Dev. Dyn. *250*, 1264–1279.

Ruvinsky, I., Silver, L.M., and Gibson-Brown, J.J. (2000). Phylogenetic analysis of T-box genes demonstrates the importance of Amphioxus for understanding evolution of the vertebrate genome. Genetics *156*, 1249–1257.

Sabarís, G., Laiker, I., Preger-Ben Noon, E., and Frankel, N. (2019). Actors with multiple roles: pleiotropic enhancers and the paradigm of enhancer modularity. Trends Genet. *35*, 423–433.

Sakabe, N.J., and Nobrega, M.A. (2013). Beyond the ENCODE project: using genomics and epigenomics strategies to study enhancer evolution. Philos. Trans. R. Soc. B Biol. Sci. *368*, 20130022.

Sanger, T.J., Losos, J.B., and Gibson-brown, J.J. (2008). A developmental staging series for the lizard genus Anolis: a new system for the integration of evolution, development, and ecology. J. Morphol. *269*, 129–137.

Sanger, T.J., Gredler, M.L., and Cohn, M.J. (2015). Resurrecting embryos of the tuatara, Sphenodon punctatus, to resolve vertebrate phallus evolution. Biol. Lett. *11*, 20150694.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113.

Sasaki, G., Ogata, T., Ishii, T., Hasegawa, T., Sato, S., and Matsuo, N. (2002). Novel mutation of TBX3 in a Japanese family with ulnar-mammary syndrome: implication for impaired sex development. Am. J. Med. Genet. *110*, 365–369.

Saunders, J.W. (1948). The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm. J. Exp. Zool. *108*, 363–403.

Saunders, J.W., and Reuss, C. (1974). Inductive and axial properties of prospective wing-bud

mesoderm in the chick embryo. Dev. Biol. *38*, 41–50.

Saunders, J.W., Gasseling, M.T., and Cairns, J.M. (1959). The differentiation of prospective thigh mesoderm grafted beneath the apical ectodermal ridge of the wing bud in the chick embryo. Dev. Biol. *1*, 281–301.

Schaffner, W. (2015). Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. Biol. Chem. *396*, 311–327.

Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., et al. (2010). Five-vertebrate ChlP-seq reveals the evolutionary dynamics of transcription factor binding. Science *328*, 1036–1040.

Seifert, A.W., Harfe, B.D., and Cohn, M.J. (2008). Cell lineage analysis demonstrates an endodermal origin of the distal urethra and perineum. Dev. Biol. *318*, 143–152.

Seifert, A.W., Yamaguchi, T., and Cohn, M.J. (2009a). Functional and phylogenetic analysis shows that Fgf8 is a marker of genital induction in mammals but is not required for external genital development. Development *136*, 2643–2651.

Seifert, A.W., Bouldin, C.M., Choi, K.S., Harfe, B.D., and Cohn, M.J. (2009b). Multiphasic and tissue-specific roles of sonic hedgehog in cloacal septation and external genitalia development. Development *136*, 3949–3957.

Sekine, K., Ohuchi, H., Fujiwara, M., Yamasaki, M., Yoshizawa, T., Sato, T., Yagishita, N., Matsui, D., Koga, Y., Itoh, N., et al. (1999). Fgf10 is essential for limb and lung formation. Nat. Genet. *21*, 138–141.

Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jónsson, B., Schluter, D., and Kingsley, D.M. (2004). Genetic and developmental basis of evolutionary pelvic

reduction in threespine sticklebacks. Nature *428*, 717–723.

Shiga, T., Inoue, K., and Masuda, T. (2002). Differential expression of the islet-1 homeodomain protein in the limb bud of the chick embryo. Anat. Embryol. (Berl). *205*, 223–228.

Stratford, T., Horton, C., and Maden, M. (1996). Retinoic acid is required for the initiation of outgrowth in the chick limb bud. Curr. Biol. *6*, 1124–1133.

Su, T., Liu, H., Zhang, D., Xu, G., Liu, J., Evans, S.M., and Pan, J. (2019). LIM homeodomain transcription factor Isl1 affects urethral epithelium differentiation and apoptosis via Shh. Cell Death Dis. *10*, 713.

Summerbell, D. (1983). The effect of local application of retinoic acid to the anterior margin of the developing chick limb. J. Embryol. Exp. Morphol. *78*, 269–289.

Suzuki, K., Ogino, Y., Murakami, R., Satoh, Y., Bachiller, D., and Yamada, G. (2002). Embryonic development of mouse external genitalia: insights into a unique mode of organogenesis. Evol. Dev. *4*, 133–141.

Suzuki, K., Haraguchi, R., Ogata, T., Barbieri, O., Alegria, O., Vieux-Rochas, M., Nakagata, N., Ito, M., Mills, A.A., Kurita, T., et al. (2008). Abnormal urethra formation in mouse models of split-hand/split-foot malformation type 1 and type 4. Eur. J. Hum. Genet. *16*, 36–44.

Suzuki, K., Adachi, Y., Numata, T., Nakada, S., Yanagita, M., Nakagata, N., Evans, S.M., Graf, D., Economides, A., Haraguchi, R., et al. (2012). Reduced BMP signaling results in hindlimb fusion with lethal pelvic/urogenital organ aplasia: a new mouse model of sirenomelia. PLoS One *7*, e43453.

Szeto, D.P., Rvan, A.K., O'Connell, S.M., and Rosenfeld, M.G. (1996). P-OTX: a PIT-1-interacting homeodomain factor expressed during anterior pituitary gland development.

Proc. Natl. Acad. Sci. U. S. A. *93*, 7706–7710.

Szeto, D.P., Rodriguez-Esteban, C., Ryan, A.K., O'Connell, S.M., Liu, F., Kioussi, C., Gleiberman, A.S., Izpisúa-Belmonte, J.C., and Rosenfeld, M.G. (1999). Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. Genes Dev. *13*, 484–494.

Tahara, N., Akiyama, R., Theisen, J.W.M., Kawakami, H., Garry, D.J., Kawakami, Y., Biology, C., Program, S., Wellstone, S., and Dystrophy, M. (2018). Gata6 restricts Isl1 to the posterior of nascent hindlimb buds through Isl1 cis-regulatory modules. Dev. Biol. *434*, 612–626.

Takeuchi, J.K., Koshiba-Takeuchi, K., Matsumoto, K., Vogel-Höpker, A., Naitoh-Matsuo, M., Ogura, K., Takahashi, N., Yasuda, K., and Ogura, T. (1999). Tbx5 and Tbx4 genes determine the wing/leg identity of limb buds. Nature *398*, 810–814.

Takeuchi, J.K., Koshiba-Takeuchi, K., Suzuki, T., Kamimura, M., Ogura, K., and Ogura, T. (2003). Tbx5 and Tbx4 trigger limb initiation through activation of the Wnt/Fgf signaling cascade. Development *130*, 2729–2739.

Tanaka, M., Munsterberg, A., Anderson, W.G., Prescott, A.R., Hazon, N., and Tickle, C. (2002). Fin development in a cartilaginous fish and the origin of vertebrate limbs. Nature *416*, 527–531.

Tarulli, G.A., Cripps, S.M., Pask, A.J., and Renfree, M.B. (2021). Spatiotemporal map of key signaling factors during early penis development. Dev. Dyn. doi: 10.1002/dvdy.433.

Thomas, T., Kurihara, H., Yamagishi, H., Kurihara, Y., Yazaki, Y., and Olson, E.N. (1998). A signaling cascade involving endothelin-1, dHAND and Msx1 regulates development of neural-crest-derived branchial arch mesenchyme. Development *125*, 3005–3014.

Tickle, C. (2017). An historical perspective on the pioneering experiments of John Saunders. Dev. Biol. *429*, 374–381.

Tickle, C., Alberts, B., Wolpert, L., and Lee, J. (1982). Local application of retinoic acid to the limb bond mimics the action of the polarizing region. Nature *296*, 564–566.

Topol, L., Jiang, X., Choi, H., Garrett-Beal, L., Carolan, P.J., and Yang, Y. (2003). Wnt-5a inhibits the canonical Wnt pathway by promoting GSK-3-independent β-catenin degradation. J. Cell Biol. *162*, 899–908.

Tschopp, P., Sherratt, E., Sanger, T.J., Groner, A.C., Aspiras, A.C., Hu, J.K., Pourquié, O., Gros, J., and Tabin, C.J. (2014). A relative shift in cloacal location repositions external genitalia in amniote evolution. Nature *516*, 391–394.

Tüzel, E., Şamli, H., Kuru, I., Türkmen, S., Demir, Y., Maralcan, G., and Güler, C. (2007). Association of hypospadias with hypoplastic synpolydactyly and role of HOXD13 Gene mutations. Urology *70*, 161–164.

Vogel, A., and Tickle, C. (1993). FGF-4 maintains polarizing activity of posterior limb bud cells in vivo and in vitro. Development *119*, 199–206.

Vogel, A., Rodriguez, C., and Izpisúa-Belmonte, J.C. (1996). Involvement of FGF-8 in initiation, outgrowth and patterning of the vertebrate limb. Development *122*, 1737–1750.

Wang, J.S., Infante, C.R., Park, S., and Menke, D.B. (2018). PITX1 promotes chondrogenesis and myogenesis in mouse hindlimbs through conserved regulatory targets. Dev. Biol. *434*, 186–195.

Wang, Y., Li, Y., Guo, C., Lu, Q., Wang, W., Jia, Z., Chen, P., Ma, K., Reinberg, D., and Zhou, C. (2016). ISL1 and JMJD3 synergistically control cardiac differentiation of embryonic stem cells. Nucleic Acids Res. *44*, 6741–6755.

Warot, X., Fromental-Ramain, C., Fraulob, V., Chambon, P., and Dollé, P. (1997). Gene dosage-dependent effects of the Hoxa-13 and Hoxd-13 mutations on morphogenesis of the terminal parts of the digestive and urogenital tracts. Development *124*, 4781–4791.

Wellik, D.M., and Capecchi, M.R. (2003). Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. Science *301*, 363–366.

Xin, Y., Le Poul, Y., Ling, L., Museridze, M., Mühling, B., Jaenichen, R., Osipova, E., and Gompel, N. (2020). Enhancer evolutionary co-option through shared chromatin accessibility input. Proc. Natl. Acad. Sci. *117*, 20636–20644.

Xu, X., Weinstein, M., Li, C., Naski, M., Cohen, R.I., Ornitz, D.M., Leder, P., and Deng, C. (1998). Fibroblast growth factor receptor 2 (FGFR2)-mediated reciprocal regulation loop between FGF8 and FGF10 is essential for limb induction. Development *125*, 753–765.

Yakushiji-Kaminatsui, N., Lopez-Delisle, L., Bolt, C.C., Andrey, G., Beccari, L., and Duboule, D. (2018). Similarities and differences in the regulation of HoxD genes during chick and mouse limb development. PLoS Biol. *16*, e3000004.

Yamaguchi, T.P., Bradley, A., Mcmahon, A.P., and Jones, S. (1999). A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. Development *126*, 1211–1223.

Yang, L., Cai, C., Lin, L., Qyang, Y., Chung, C., Monteiro, R.M., Christine, L., Fishman, G.I., Cogen, A., and Evans, S. (2006). Isl1Cre reveals a common Bmp pathway in heart and limb development. Development *133*, 1575–1585.

Yang, Y., Drossopoulou, G., Chuang, P.T., Duprez, D., Marti, E., Bumcrot, D., Vargesson, N., Clarke, J., Niswander, L., McMahon, A., et al. (1997). Relationship between dose, distance and time in Sonic Hedgehog-mediated regulation of anteroposterior polarity in

the chick limb. Development *124*, 4393–4404.

Yashiro, K., Zhao, X., Uehara, M., Yamashita, K., Nishijima, M., Nishino, J., Saijoh, Y., Sakai, Y., and Hamada, H. (2004). Regulation of retinoic acid distribution is required for proximodistal patterning and outgrowth of the developing mouse limb. Dev. Cell *6*, 411–422.

Young, J.J., Grayson, P., Edwards, S. V., and Tabin, C.J. (2019). Attenuated Fgf signaling underlies the forelimb heterochrony in the emu Dromaius novaehollandiae. Curr. Biol. *29*, 3681–3691.

Yu, M., and Ren, B. (2017). The three-dimensional organization of mammalian genomes. Annu. Rev. Cell Dev. Biol. *33*, 265–289.

Yu, Z., Kong, J., Pan, B., Sun, H., Lv, T., Zhu, J., Huang, G., and Tian, J. (2013). Islet-1 may function as an assistant factor for histone acetylation and regulation of cardiac development-related transcription factor Mef2c expression. PLoS One *8*, e77690.

Zhang, R., Knapp, M., Suzuki, K., Kajioka, D., Schmidt, J.M., Winkler, J., Yilmaz, Ö., Pleschka, M., Cao, J., Kockum, C.C., et al. (2017). ISL1 is a major susceptibility gene for classic bladder exstrophy and a regulator of urinary tract development. Sci. Rep. *7*, 42170.

Zhang, Z., Verheyden, J.M., Hassell, J.A., and Sun, X. (2009). FGF-regulated Etv genes are essential for repressing Shh expression in mouse limb buds. Dev. Cell *16*, 607–613.

Zheng, Z., Armfield, B.A., and Cohn, M.J. (2015). Timing of androgen receptor disruption and estrogen exposure underlies a spectrum of congenital penile anomalies. Proc. Natl. Acad. Sci. *112*, E7194–E7203.

Zuniga, A., and Zeller, R. (2020). Dynamic and self-regulatory interactions among gene regulatory networks control vertebrate limb bud morphogenesis. Curr. Top. Dev. Biol.

*139*, 61–88.

Zúñiga, A., Haramis, A.P.G., McMahon, A.P., and Zeller, R. (1999). Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds. Nature *401*, 598–602.