Lmx1b-Targeted cis-Regulatory Modules Involved in Limb Dorsalization

Endika Haro  
Department of Pathology and Human Anatomy  
Loma Linda University  
Loma Linda, CA  
eharo@llu.edu

Billy A. Watson  
Department of Pathology and Human Anatomy  
Loma Linda University  
Loma Linda, CA  
bawatson@llu.edu

Jennifer M. Feenstra  
Department of Pathology and Human Anatomy  
Loma Linda University  
Loma Linda, CA  
jennifer.feenstra@ki.se

Luke Tegeler  
Department of Pathology and Human Anatomy  
Loma Linda University  
Loma Linda, CA  
ltegeler@westernu.edu

Charmaine U. Pira  
Department of Pathology and Human Anatomy  
Loma Linda University  
Loma Linda, CA  
cpira@llu.edu

Subburaman Mohan  
Musculoskeletal Disease Center  
Loma Linda VA HealthCare System  
Loma Linda, CA  
Subburaman.Mohan@va.gov

Kerby C. Oberg  
Department of Pathology and Human Anatomy  
Loma Linda University  
Loma Linda, CA  
koberg@llu.edu

Corresponding Author: Kerby C. Oberg, MD, PhD

Key words: Lmx1b, Dorsal-ventral patterning, enhancer, limb, dorsalization
SUMMARY

In this study, genome-wide sites of Lmx1b binding were correlated with Lmx1b-regulated genes during limb development (E12.5) to uncover cis-regulatory modules and their gene targets involved in limb dorsalization.

ABSTRACT

Lmx1b is a homeodomain transcription factor responsible for limb dorsalization. Despite striking double-ventral (loss-of-function) and double-dorsal (gain-of-function) limb phenotypes, no direct gene targets in the limb have been confirmed. To determine direct targets, we performed a chromatin immunoprecipitation against Lmx1b at E12.5 followed by next generation sequencing (ChIP-seq). Nearly 84% (n=617) of the Lmx1b-bound genomic intervals (LBIs) identified overlap with chromatin regulatory marks indicative of potential cis-regulatory modules (PCRs). In addition, 73 LBIs mapped to known CRMs active during limb development. We compared Lmx1b-bound PCRs to genes differentially expressed by Lmx1b and found 292 PCRs within 1 Mb of 254 Lmx1b-regulated genes. Gene ontologic analysis suggests that Lmx1b targets extracellular matrix production, bone/joint formation, axonal guidance, vascular development, cell proliferation and cell movement. We validated the functional activity of a PCR associated with joint-related Gdf5 that provides a mechanism for Lmx1b-mediated joint modification and a PCR associated with Lmx1b that suggests a role in autoregulation. This is the first report to describe genome-wide Lmx1b binding during limb development, directly linking Lmx1b to targets that accomplish limb dorsalization.
INTRODUCTION

The presence of morphologically distinguishable dorsal and ventral (DV) limb asymmetry reflects the polarized pattern accomplished during development. In mouse limbs, hair only appears on the dorsal surface of the autopod and nails on the dorsal aspect of digit tips, whereas footpads and eccrine glands are restricted ventrally. Internal differences are also present. Muscles develop as dorsal extensors and ventral flexors with disparate attachments, bones show subtle asymmetry, and joint morphology supports ventral flexion.

The polarity of the limb ectoderm is achieved by the restricted expression domains of Wingless-type MMTV integration site family member 7a (Wnt7a) and Engrailed 1 (En1) (Parr and McMahon, 1995, Loomis et al., 1996, Cygan et al., 1997, Loomis et al., 1998). As the limb begins to emerge, Wnt7a expression wraps around the limb bud apex (Bell et al., 1998). Bmp signals from the lateral plate mesoderm trigger the activation of En1 in the ventral limb ectoderm (Pizette et al., 2001). En1 expression expands in the ventral ectoderm, restricting Wnt7a expression to the dorsal ectoderm (Loomis et al., 1996, Cygan et al., 1997, Loomis et al., 1998). The restricted dorsal secretion of Wnt7a imparts polarity to the underlying limb mesoderm by triggering the expression of Lmx1b, a LIM homeodomain transcription factor that is ultimately responsible for limb dorsalization (Chen et al., 1998, Parr and McMahon, 1995, Riddle et al., 1995, Vogel et al., 1995). Mice lacking functional Lmx1b develop a ventral-ventral limb phenotype, whereas ectopic ventral expression of Lmx1b leads to a dorsal-dorsal limb phenotype (Chen et al., 1998, Cygan et al., 1997, Vogel et al., 1995). In humans, haploinsufficiency of LMX1B is associated with Nail-Patella Syndrome characterized by nail dysplasia, absent or hypoplastic patellae, progressive renal disease and decreased bone mineral density (Chen et al., 1998, Towers et al., 2004, Dreyer et al., 2000).

Despite the striking effect of Lmx1b on dorsal limb morphology, direct targets in the limb have been elusive. A number of potential targets have been suggested by comparative gene arrays between wild type and Lmx1b knockout mice, but none have been confirmed (Feenstra et al., 2012, Gu and Kania, 2010, Krawchuk and Kania, 2008). Therefore, in order to
elucidate direct limb targets for Lmx1b, we performed chromatin immunoprecipitation against Lmx1b followed by massive parallel sequencing (ChIP-seq) of mouse limbs during dorsalization (E12.5) and compared these sites of Lmx1b binding with genes regulated by Lmx1b. In this report, we describe the identification of Lmx1b-bound potential cis-regulatory modules (PCRM s) associated with Lmx1b-regulated genes and the predicted functional pathways affected. These data suggest that Lmx1b exerts its dorsalizing effects through the targeted regulation of genes involved in bone/joint development, extracellular matrix composition, axon tracking and cell proliferation. Intriguingly, we also identified an Lmx1b-bound PCRM upstream of the Lmx1b gene that provides the capacity for autoregulation.
RESULTS

Validation and genome-wide characterization of the Lmx1b-ChIP-seq

Feenstra and coworkers (Feenstra et al., 2012) found more limb bud genes differentially expressed in the presence of Lmx1b at E12.5 than in either E11.5 or E13.5. Therefore, we performed ChIP-seq analyses on E12.5 mouse limbs. After peak calling analysis (MACS algorithm v1.4.2, cutoff p<1e-5), we identified 735 Lmx1b-bound genomic fragments or intervals (LBIs) common to ChIP-seq replicates with an average length of 470 bp (Table S1).

Comparison of LBIs to the annotated mouse genome showed that intronic (36%) or intergenic (54%) regions were the most common sites for Lmx1b binding (Fig. 1A). Of the regions mapped, 1% were within coding regions, 2% were found within the 5’UTR and 1% within the 3’UTR. Finally, only 6% of the intervals localized to potential promoters (-2500 and +500bp from the transcription start sites (TSS) based on UCSC gene annotation).

MEME-ChIP (Bailey et al., 2015) analysis for de novo motif discovery retrieved TMATWA (M = C or A, W = T or A) as the most common motif found in the Lmx1b-bound genomic regions (p = 3.5e-29) (Fig. 1B). The TMATWA motif includes the published Lmx1b binding site (TAATTA) (Morello et al., 2001) and the distribution of both motifs within intervals was similar (p = 4.5e-24), consistent with Lmx1b as the ChIPed transcription factor.

The biologic functions of promoter-associated intervals were predominantly related to genes with general cell functions as determined by GREAT analysis (Fig. 1C) (McLean et al., 2010). The intronic-associated intervals correlated with genes involved in growth and organ development including limb morphogenesis. Remarkably, the intergenic-associated genomic regions were enriched near genes related to limb and skeletal development, which are functions that are anticipated to be modified by Lmx1b.
Lmx1b-binding is associated with active regulation

Although the annotated location and GREAT analysis give some insight into the potential regulatory activity of the LBIs, cis-regulatory modules (CRM) are characterized by histone modifying co-factors and bound proteins, collectively called chromatin regulatory marks, that affect genome accessibility (Hardison and Taylor, 2012). p300 is a ubiquitous phosphoprotein with intrinsic histone acetyltransferase activity that predicts regulatory activity in a tissue specific manner (Visel et al., 2009). Since Lmx1b is also expressed in other tissues including forebrain, midbrain, hindbrain, spinal cord, kidney and the eye (Chen et al., 1998, Asbreuk et al., 2002), we evaluated whether the distribution of the 735 LBIs conformed to a limb-specific pattern. We compared the 735 LBIs to p300 ChIP-seq data in available embryonic tissues i.e., limb, forebrain and midbrain (Visel et al., 2009) and confirmed that more LBIs co-localized with p300 from the limb (288, ~39%), than from either the forebrain (25, ~3%) or midbrain (1, ~0.1%) (Fig. S1A), consistent with a limb-specific pattern of regulatory activity.

Overlap of an LBI with multiple chromatin regulatory marks increases confidence in its role as a regulatory sequence (Hardison and Taylor, 2012). Thus, in addition to p300, we compared the LBIs to available ChIP-seq data for chromatin regulatory marks associated to enhancer elements (H3K27Ac and H3K4Me2) (Heintzman et al., 2009) and regulatory regions undergoing active transcription (RNA Pol II and Med12) (Kagey et al., 2010) available in E12.5 and E11.5 limbs. We recognize that regulation during these stages is dynamic and what is active at E11.5 might not be active at E12.5, however, this data would still highlight a potential cis-regulatory module (PCRM). After peak calling analyses, the Lmx1b-ChIP-seq reads exhibited a 7-fold enrichment in tagged sequences within the 735 LBIs when compared to the input DNA (Fig. 1D). We next analyzed the distribution of tagged sequences within LBIs that overlap chromatin regulatory marks. We found that Lmx1b ChIPed DNA was enriched within enhancer-associated chromatin marks (H3K27Ac and H3K4me2) and exhibited a 3-fold
enrichment over input DNA, whereas the enrichment within actively transcribed regulatory regions was 5-fold (RNA Pol II and Med12) (Fig. S1B). Overall, Lmx1b-ChIP-seq DNA sequences show a 4-fold enrichment within chromatin regulatory marks, consistent with a role for Lmx1b as a factor involved in active regulation (Fig. 1D and Fig. S3). Since both CRMs and promoters are active regulatory regions, we examined the TSSs for genes regulated by Lmx1b (Feenstra et al., 2012). Similar to the genomic distribution of LBIs, no TSS enrichment was found in the distribution of tagged sequences between Lmx1b-ChIPed DNA and input DNA (Fig. 1D) supporting the concept that Lmx1b-mediated regulation favors CRMs over promoters.

We required at least two active regulatory marks for an interval to be categorized as a PCRM. Based on this criterion, we determined that 617 of the 735 LBIs were PCRM (Fig. 2A and Table S2). The number of PCRM was 30 times higher than random genomic regions with comparable characteristics (n=5 groups, each with 735 random genomic intervals, One sample t-test p<1e-4) (Fig. S2A). The PCRM were dispersed throughout the genome with the largest number (51) being found in chromosome 1, the largest chromosome. Interestingly, no PCRM were identified in the X chromosome, although it is nearly as large as chromosome 2 and 28 Lmx1b-regulated genes are found on the X chromosome (Fig. S3). LBIs that overlapped repressive regulatory marks (H3K27me3) were found in 50 intervals, 41 of which were also associated with 2 or more active regulatory marks (Table S2). Intervals associated to repressive regulatory marks alone or with less than two active regulatory marks (n=9) were considered inactive PCRM or nonspecific Lmx1b binding (Table S2).
**Lmx1b-bound PCRM s are associated with Lmx1b-regulated genes**

In order to determine direct limb targets, we compared Lmx1b-bound PCRM s to previously reported genes regulated by Lmx1b at E12.5 (Feenstra et al., 2012). We found 292 PCRM s within 1 Mb of 254 Lmx1b-regulated genes (Fig. 2A and Table S3). In contrast, when we compared random non-Lmx1b-bound regulatory sequences (n=5 groups, each with 292 H3K27Ac positive intervals) to the 254 Lmx1b-regulated genes, only 60 (std dev=7) regulatory sequences on average associated with 75 (Std Dev=11) genes. This indicates that Lmx1b-PCRM s are enriched for their associated gene targets 5-fold over random regulatory sequences (one sample T-test analysis, p< 1e-4) (Fig. S2B). We also found that some PCRM s were associated to more than one gene and some genes were associated to multiple PCRM s (Fig. 2A). Enrichment of Lmx1b-bound gene-associated PCRM s was validated by qPCR using two independently generated ChIP samples. Nine PCRM s were selected and each exhibited at least a 4-fold enrichment when compared to an unbound control region (Fig. 2B).

Ingenuity Pathway Analysis (IPA) of the 254 genes associated with PCRM s demonstrated a significant effect on 8 functional categories relevant to limb development. These categories include Connective Tissue Development/Function, Skeletal and Muscle System Development/Function and Cellular movement, with some genes present in more than one functional category (Fig. 2C, Table S4). Within the functional categories affected, genes were further subcategorized into limb–related annotated functions (Fig. 2D, Table S4) including 17 that mapped to the annotated biological process termed “limb development”. IPA analyses also predicted that canonical pathways involved in movement, proliferation, cell adhesion, cytoskeletal regulation, and vascular development were targeted (Fig. S4).

Within the connective tissue development category, we found several genes involved with extracellular matrix (ECM) composition that are associated to Lmx1b-bound PCRM s including *Col1a2, Col11a2, Ker, Lum, Dcn, Matn1/4, Epyc, and Has3* (Table S4). In addition, a number of genes associated to bone/joint differentiation *Osr2, Gdf5, Runx2, Sox11, and Trps1* (Table S4) were present within the Skeletal and Muscle Systems Development/Function
category.

**PCRM-associated to Lmx1b-regulated genes are active during limb development**

We performed comparative analyses using the VISTA browser (Frazer et al., 2004) to identify conserved PCRM that were more likely to be functional across species. A high degree of conservation (greater than 70% homology) was found in 289 LBIs. Approximately 90% of the conserved LBIs are PCRM with 105 associated to 94 Lmx1b-regulated genes (Table S2 & S3).

Further functional validation was performed on 2 conserved Lmx1b bound PCRM. One of the PCRM (LBI407) is 66 kb upstream of *Lmx1b* on murine Chromosome 2 (Fig. 3A and Table S3). In silico analysis for transcription factor binding sites in LBI407 confirmed two potential sites for Lmx1b as well as several other transcription factors (Fig. 3B). LBI407 (1224 bp) was isolated from murine genomic DNA, linked to a GFP reporter and transfected into chick presumptive limb mesoderm by electroporation. LBI407 showed robust enhancer activity in the limb 48 hours after transfection (Fig. 4A-C and E-G). Moreover, its activity was restricted to the dorsal mesoderm coincident with *LMX1B* expression (Fig. 4D and H) (Vogel et al., 1995, Riddle et al., 1995, Dreyer et al., 2000).

We validated the activity of another PCRM (LBI443 from Table S3) located 82 Kb downstream of *Growth differentiation factor 5 (Gdf5)* (Fig. 5A), a factor known to be involved in joint development (Settle et al., 2003). This PCRM contained multiple potential binding sites for Lmx1b, Sox, and Osr2 transcription factors (Fig. 5B). A GFP reporter construct containing the LBI443 (867 bp) isolated from murine DNA was transfected into chick presumptive limb mesoderm. LBI443 exhibited enhancer activity in the elbow, wrist, and digit joints overlapping *GDF5* expression (Fig. 6C). Furthermore, using section in situ hybridization we demonstrate that *LMX1B* expression overlaps both LBI443 activity and *GDF5* expression (Fig. 6E-G) dorsally in the developing elbow joint consistent with a role for LMX1B-mediated dorsalization.
An additional 91 Lmx1b-bound PCRs correspond to previously published conserved non-coding DNA sequences available in the Vista Enhancer Browser, 73 of which have confirmed activity during limb development (Table S5) (Visel et al., 2007). We associated 27 of these PCRs to 34 genes differentially expressed in the presence of Lmx1b (Table 1). Based on the expression patterns available in the Mouse Genomic Informatics database, we found that the activity of the Lmx1b-bound PCRs associated with Osr2, Jag1, Wnt5a, Shox2, and Cbln4 genes overlapped their respective limb mRNA expression patterns, which also overlaps Lmx1b’s expression (Loomis et al., 1998, Lan et al., 2001, Witte et al., 2009, Cobb et al., 2006, Haddick et al., 2014).
DISCUSSION

Characteristics of Lmx1b-bound DNA in mouse limbs

We mapped the genome-wide distribution of Lmx1b binding in mouse limbs during limb dorsalization (E12.5). Most of the Lmx1b-bound genomic fragments or intervals (LBIs) mapped to intergenic or intronic regions (Fig. 1A) and associate to active enhancers rather than promoters (Fig. 1D and Fig. S1A-B), a feature shared by other development-related transcription factors (Sheth et al., 2016, McAninch and Thomas, 2014). Based on chromatin regulatory marks (chromatin modifications and bound proteins), nearly 84% (n=617) of the LBIs were categorized as potential cis-regulatory modules (PCRM).

Typically, the distribution of either repressor (H3k27me3) or active chromatin regulatory marks (H3K27Ac, RNAPII or Med12) is mutually exclusive (Ram et al., 2011, Pasini et al., 2010); remarkably, a small population of Lmx1b-bound PCRM, such as LBI407, aligned to both active and repressor marks (Fig. 3A and Table S2). This may reflect differences in regulation within the tightly restricted dorsal and ventral compartments of the developing limb, with potentially different genomic landscapes (Arques et al., 2007, Cotney et al., 2013, Andrey et al., 2017) and the capacity for different transcription factors such as Lmx1b to modify activity of these PCRM. Since gene regulation during limb development is a temporally dynamic process, another interpretation could be that the availability/activity of the regulatory sequence is in transition.

Regulatory sequences that are foundational to the development of the vertebrate body plan show high degrees of conservation across species (de Laat and Duboule, 2013). Thus, the conserved PCRM identified via comparative genomic analysis (n=257) likely play critical roles in Lmx1b-regulated development (Supplementary Table 2). We validated enhancer activity in two of these conserved PCRM (LBI407 and 443, Figs. 4 and 6 respectively) and further confirmed 80% of the PCRM that correspond to conserved cis-regulatory modules (CRMs) have limb activity (Vista Enhancer Browser database (Visel et al., 2007)) (Table S5).
Candidate genes for direct Lmx1b regulation

The distance between a regulatory sequence and its target gene can be highly variable and can associate with targets beyond the nearest gene (Marinic et al., 2013, Carter et al., 2002). Genomic territories with enhanced chromatin interactions called topologically associating domains can extend over roughly 1 Mb in vertebrates (Dixon et al., 2012). Nearly half of our PCRM (n=292) were within 1 Mb of an Lmx1b regulated gene (Fig 2A and Table S3). The two functionally active PCRM (LBI407 and 443) validated were associated with Lmx1b-regulated genes (Lmx1b and Gdf5, respectively) with enhancer activity overlapping associated gene expression (Fig. 4 & 6). Twenty-seven of the PCRM confirmed via the Vista Enhancer Browser (Visel et al., 2007) were also linked to Lmx1b-regulated genes (n=34). Five had readily available limb expression patterns that overlapped the associated PCRM CRM activity (Osr2, Jag1, Wnt5a, Shox2 and Cbln4) (Fig. 7) (Lan et al., 2001, Loomis et al., 1998, McGlinn et al., 2005, Witte et al., 2009, Cobb and Duboule, 2005, Haddick et al., 2014).

PCRM that were not associated to an Lmx1b-regulated gene at E12.5 are likely accessible, but may not be active at this stage. Because limb development is dynamic, stage-specific factors, in addition to Lmx1b binding, may also be required for PCRM activity. Correspondingly, Feenstra and colleagues demonstrated variation in Lmx1b-regulated genes in progressive limb stages (E11.5, E12.5 and E13.5), with less than 10% present across all three stages (Feenstra et al., 2012). Some PCRM-gene associations may also have been missed because of our 1 Mb cut-off for PCRM-target interactions. Although the average CRM-promoter distance is 120 kb (de Laat and Duboule, 2013), interactions have been reported up to a distance of 1.44 Mb (de Laat and Duboule, 2013, Benko et al., 2009).
Predicted Lmx1b-regulated processes

Gene ontologic analysis of candidate PCRM-associated genes predicted target pathways, and tissue systems present in both dorsal and ventral aspects of the limb. Differential compartment-specific regulation of CRMs common to limb tissues is a likely mechanism to refine dorsal asymmetry. Joint development is asymmetric along the dorsal-ventral axis. *Growth differentiation factor 5 (Gdf5)* is a well-established marker for joint development (Settle et al., 2003) with expression spanning both the dorsal and ventral limb compartments. Lmx1b-dependent regulation of the *Gdf5*-associated CRM (LBI443), an enhancer reported to drive *Gdf5* expression in the developing limb joints (Chen et al., 2016), favors a model in which Lmx1b modifies dorsal *Gdf5* enhancer activity.

Lmx1b-dependent regulation of CRMs associated with *Wnt5a* (LBI252) and extracellular matrix (ECM) genes such as *Keratocan, Lumican, Decorin*, and *Epyphican* (LBI89 and 91) (Table S3) are additional mechanisms that could alter limb asymmetry. *Wnt5a* and ECM components can affect cell growth, survival, differentiation, migration, and morphogenesis that could be regulated differentially along the dorsal-ventral axis (Gros et al., 2010, Rozario and DeSimone, 2010, Koohestani et al., 2013). A role for Lmx1b in ECM regulation has already been demonstrated in the kidney with direct regulation of glomerular basement membrane collagens (Morello et al., 2001).

A role for Lmx1b in axonal guidance was proposed with the identification of targets such as *Cbln4* and *Ntn1* (Krawchuk and Kania, 2008, Feenstra et al., 2012). We demonstrate an Lmx1b-bound CRM (LBI456) associated to *Cbln4* that mimics the pattern of *Cbln4* expression (Visel et al., 2007, Haddick et al., 2014) and 2 Lmx1b-bound PCRs (LBI122 and 123) that flank *Ntn1* (Table S3) collectively supporting this hypothesis.
Nail-Patella Syndrome

In humans, haploinsufficiency of LMX1B results in Nail-Patella Syndrome (NPS) characterized by nail dysplasia, absent or hypoplastic patellae, bone fragility and premature osteoarthritis (Sweeney et al., 2003). We have identified several Lmx1b-bound PCRM associated with Gdf5, Sox11 and several ECM-related genes (Tables S3) linked to osteoarthritis (Kan et al., 2013, Syddall et al., 2013, Stefansson et al., 2003, Miyamoto et al., 2007).

Another interesting finding suggested by our study is the positive autoregulation of Lmx1b expression. In the Lmx1b loss-of-function mutant, the Lmx1b transcript is reduced 5-6 fold (Feenstra et al., 2012). Functional validation of LBI407, located 66 kb upstream of Lmx1b, showed dorsally restricted enhancer activity in limb mesoderm coincident with Lmx1b expression (Vogel et al., 1995, Riddle et al., 1995, Dreyer et al., 2000). Positive autoregulation provides a mechanism for maintenance or amplification of Lmx1b expression following initial activation. Autoregulation during development reinforces or stabilizes a transcriptional pattern of differentiation (Crews and Pearson, 2009). This may also be of clinical importance since a population of NPS patients with mutations linked to the LMX1B locus fail to demonstrate mutations in the LMX1B coding sequence (Ghoumid et al., 2016). The fact that NPS results from haploinsufficiency of LMX1B signifies that the functional level of LMX1B is critical to normal development. Disruption of the LMX1B autoregulatory system, via mutations in the LMX1B CRMs, could therefore account for NPS in families that lack mutations in the coding sequence.

In summary, we have generated a genomic data set of Lmx1b-bound PCRM during limb dorsalization. Moreover, we have validated and linked Lmx1b-bound cis-regulatory modules to genes differentially expressed by Lxm1b, highlighting the processes regulated. These data will allow us to characterize the different mechanisms used by Lmx1b to accomplish limb dorsalization.
DATA AVAILABILITY

The Lmx1b-ChIP-seq data sets reported in the paper are available through GEO as GSE84064.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Ralph Witzgall for the kind gift of the Lmx1b antibody (BMO8), Drs. Nadav Ahituv, Thomas Girke, and Paul Labhart for assistance and suggestions regarding data analysis and standardization and Drs. Salva Soriano and Brendon Gongol for critical review of the manuscript. This work was supported in part by grants from the National Organization of Rare Diseases (KCO) and the Loma Linda University Pathology Research Endowment Fund (KCO).

CONTRIBUTIONS

The experimental design was planned by EH, JMF, KCO and SM. The initial ChIP-seq analysis was carried out by LT, EH and KCO. The detailed bioinformatics analysis, regulatory mark comparisons and PCRM calling was performed by EH. The Ingenuity Pathway Analysis was performed by KCO. The qPCR data was performed by BAW. The functional validation experiments were performed by EH and CP. The manuscript was written by EH and KCO with detailed review and editorial changes from BAW, CP, and SM. None of the authors have financial, personal or professional conflicts of interest regarding the research presented.
MATERIALS AND METHODS

Mouse strains

All animal procedures were performed in accordance with protocols established by the Institutional Animal Care and Use Committee of Loma Linda University. The genetic background of the mouse strain used for this study was C57BL/6. The sex was not determined in the embryos used.

Lmx1b-ChIP-seq

Limb tissue from ten E12.5 embryos was pooled for each of the biological replicates (n=2) and submerged in PBS + 1% formaldehyde for 15 minutes. After disruption with a Dounce homogenizer, lysates were sonicated and the DNA sheared to an average length of 300-500 bp. Chromatin (30 ug) was precleared with protein A agarose beads and incubated with 20 ul antibody against Lmx1b (BMO8) (Suleiman et al., 2007) kindly provided Dr. Witzgall. The protein-DNA complex was reverse-crosslinked with an overnight incubation in Proteinase K at 65 °C. ChIPed DNA was purified by phenol-chloroform extraction and ethanol precipitation.

DNA libraries were quantified and sequenced on Illumina’s NextSeq 500. Reads were aligned to the mouse reference genome (mm10) using the default settings for Bowtie algorithm (Li and Durbin, 2009). Lmx1b peak locations were determined using the Model based Analysis for ChIP-seq (MACS) (v1.4.2) (Zhang et al., 2008) with a cutoff of p-value = 1e-5 (empiric false discovery rate = 6-16%).

ChIP Validation by qPCR

Chromatin immunoprecipitation for qPCR was performed using ChIP-it express Kit (Active Motif, USA) following manufacturers recommendations with minor modifications. Lysed cells were sonicated using Epishear Probe sonicator (Active Motif, USA) to obtain
fragments ranging between 300 and 600 bp in length. The sonicated DNA, Lmx1b-specific rabbit polyclonal antiserum, and 25ul of protein G beads were incubated overnight at 4°C. DNA elution and de-crosslink was performed following manufacturer’s recommendations. DNA was purified using the QIAquick PCR purification kit (Qiagen, USA).

qPCR (BIO-RAD, CFX96) validation was performed using SYBR-Green (BIO-RAD 172-5270). Validation was performed in triplicates using two independently generated ChIP samples with the Lmx1b antibody. Primers used for validation of ChIP-qPCR were assayed for primer efficiency and are listed in a 5’ to 3’ orientation in Table S6. Target enrichment was determined by calculating the percentage of target precipitated over the input DNA (% input) and adjusted for primer efficiency. Subsequently, fold enrichment was determined by comparing % input of specific targets over mouse negative control provided in the ChIP-it qPCR Analysis kit (ActiveMotif, USA).

Motif Discovery

Motif discovery analyses of Lmx1b-ChIP-seq retrieved genomic intervals were performed using the online tool MEME-ChIP 4.11.0 version (http://meme-suite.org/tools/meme-chip) as described in Ma et al. (Ma et al., 2014). Input sequences were centered within summit regions of recovered intervals and extended 250 bp in each direction. MEME runs were performed with random subsampling and retrieved motifs between 6 and 10 bp in length with an E-value cut off of >0.5 for the discovery of enriched motifs.
**Genomic Regions Enrichment of Annotations Tool**

Association of genomic regions to genes was performed using the online tool Genomic Regions Enrichment of Annotations Tool (GREAT, http://bejerano.stanford.edu/great/public-2.0.2/html/) (McLean et al., 2010). Parameters were set so that regulatory domains for genes extend in both directions 1 Mb from the midpoint of the gene’s transcription start site (TSS).

**Published ChIP-seq Data**

Limb ChIP-seq data on H3K27Ac was obtained at the National Center for Biotechnology Information (NCBI) from the Gene Expression Omnibus database (GEO, http://www.ncbi.nlm.nih.gov/geo/) under the accession numbers GSE30641 and GSE42413, p300 under accession number GSE13845 and both, H3K27me and H3K4me, under the accession number GSE42237. RNA Pol II and Med12 ChIP-seq data was obtained from Berlivet et al. (Berlivet et al., 2013). The data for comparison was converted to the mouse build mm10 using the UCSC liftover tool. Lmx1b-ChIP-seq retrieved genomic intervals were extended 250 bp in both directions for comparison with published ChIP-seq data.

**Comparative Analyses**

Comparative analysis of Lmx1b-ChIP-seq identified intervals was performed by pairwise alignment (Vista browser, http://genome.lbl.gov/vista/). Species selected for the pairwise alignment comparison were mouse, human, horse, and chicken. An interval was considered conserved when it exhibited at least a 70% homology.
**Gene Ontology:**

Differentially expressed genes in the presence of Lmx1b associated to Lmx1b-bound intervals (LBIs) were classified according to Gene Ontology terms using Ingenuity Pathway Analysis (IPA) software and database (Qiagen, USA).

**Isolation and Cloning of Potential Regulatory Regions:**

The two LBIs analyzed were isolated from mouse genomic DNA by polymerase chain reaction (PCR) using the following primer pairs:

LBI407 (1226 bp fragment), 5’-GGGGACCAGGAGAAATATTACAGTGTG-3’ and 5’-CAGAATCCCCCAGAGATAGATGC-3’;

LBI443 (867 bp fragment), 5’-CTACAGCTCAGTCTCCTTCAGGCTACAC-3’ and 5’-CCATACATACTGAGCCACCACATGG-3’.

The PCR products were cloned into pCR-II TOPO vector (Qiagen, USA) for subsequent subcloning into a thymidine kinase (tk) promoter-driven GFP reporter construct for functional analyses (Uchikawa et al., 2003).

**Functional Enhancer Assays**

Minimal promoter driven GFP reporter constructs bearing the potential regulatory region of interest were delivered into presumptive limb region of Hamburger and Hamilton stage (HH) 14 chicken embryos as previously described by Pira and colleagues (Pira et al., 2008). Transfection efficiency was assessed by co-electroporation of a β-actin promoter-driven RFP construct. Electroporation was performed using the CUY21 electroporation station (Protech International, USA). Depending on the construct of interest, embryos were incubated for 2-5 days before harvesting for visualization of GFP activity and digital image acquisition (Sony DKC-5000).
In Situ Hybridization and Probe Generation;

Whole-mount *in situ* hybridization was performed as described in Yamada et al. (Yamada et al., 1999). Section *in situ* hybridization was performed as described (Moorman et al., 2001). *GDF5* and *LMX1B* probes were generated by RT-PCR as described in Merino et al. (Merino et al., 1998) using the following primer pairs: cGDF5 5’-GTAAGGACGGTGACTCCAAAGG-3’ and 5’-CCTTGCTTCAGGTTCTTACTG-3’; cLMX1B 5’-GGATCGCTTTCTGATGAGG-3’ and 5’-GATGTCATCATTCCCTCCATTG-3’.
Figure 1. Validation and characterization of Lmx1b-bound intervals

A) Genome-wide distribution of LBIs showing a high percentage of intervals in intergenic and intronic regions. B) Distribution of the de novo identified Lmx1b-ChIPed motif
(TMATWA) and predicted distribution of the published TAATTA binding motif (Morello et al., 2001) for Lmx1b in the intervals retrieved from ChIP-seq experiments. C) GREAT analysis of annotated genes within 1 MB of LBIs showing limb development related genes associated to intergenic intervals. D) Heatmaps (left) and summarized averages plots (right) for input and Lmx1b-ChIPed DNA according to fraction of reads aligned to Lmx1b bound intervals (left panel); genomic regions associated with cis-regulatory activity (H3K27Ac, H3K4me2, p300, RNA Pol II and/or Med12) that overlapped Lmx1b targeted intervals (Visel et al., 2009, Berlivet et al., 2013, Cotney et al., 2013, DeMare et al., 2013) (middle); and TSS of Lmb1 regulated genes (right). Lmx1b preferentially binds to regulatory regions rather than promoters.
Figure 2. Functional Categories of PCRM-associated Lmx1b-regulated Genes.
A) Bubble diagram depicting the associations between Lmx1b-bound PCRM (blue) and genes differentially regulated in the presence of Lmx1b: upregulated (red) and downregulated (green). More than one PCRM can associate with a gene and more than one gene can associate with a PCRM. B) Enrichment of Lmx1b-bound PCRM by ChIP-qPCR is indicated as fold change of % input over the unbound control (for tested LBIs, p<0.001, n=6). Associated genes are shown in parentheses for each LBI. C) Predicted functional categories of PCRM-associated Lmx1b-regulated genes using Ingenuity Pathway Analysis (Table S4). B) Annotated Functions within the Functional Categories listed in “C”. The numbers of affected Lmx1b-bound genes are listed on the left (Table S4). The associated genes for the Limb Development Category are listed on the right in the shaded box. Note: a gene can be classified in more than one functional category.
Figure 3. LBI407 is a highly conserved PCRM upstream of *Lmx1b*

A) Annotated browser image depicting chromosomal location of LBI407, 66 kb upstream of *Lmx1b*. Chromatin associated marks (Visel et al., 2009, Berlivet et al., 2013, Cotney et al., 2013, DeMare et al., 2013), location of intervals bound by the transcription factor Lmx1b in E12.5 mouse limbs, and conservation obtained from the UCSC genome browser (top to bottom).
LBI407 (highlighted by a vertical green shaded bar) is highly conserved across vertebrate species and overlaps with 4 chromatin-associated marks (both active and repressor). B) In silico analysis of potential transcription factor binding sites within LBI407 exhibiting two potential binding sites for Lmx1b (yellow).
Figure 4. LBI407 is a dorsally restricted enhancer coincident with *LMX1B* expression.

(A-C) Dorsal and (E-G) transverse views of chicken wing buds 48 h after electroporation at Hamburger-Hamilton stage (HH) 14. (D and H) *In situ* hybridization for *LMX1B* at HH24 chicken wing buds. (B and F) Transfection efficiency is determined by β-actin promoter-driven RFP. (C and G) Enhancer activity of LBI407 drives GFP expression in the dorsal limb mesoderm coincident with the pattern of *LMX1B* expression (D and H). The dorsal and transverse fluorescence views are overlaid with light field to define limb boundaries.
Figure 5. LBI443 is a highly conserved PCRM downstream of Gdf5

A) Annotated browser image depicting chromosomal location of LBI443, 82 kb downstream of Gdf5. Chromatin associated marks (Visel et al., 2009, Berlivet et al., 2013, Cotney et al., 2013, DeMare et al., 2013), Lmx1b-bound interval from E12.5 mouse limbs, and conservation obtained from the UCSC genome browser (top to bottom). LBI443 (highlighted by a vertical
green shaded bar), is highly conserved across vertebrate species and overlaps 5 active chromatin-associated marks. B) In silico prediction of transcription factors binding sites in LBI443 identified 5 potential binding sites for Lmx1b.
Figure 6. LBI443 is active in developing GDF5-positive joints, with the dorsal aspect extending into the Lmx1b expression domain.

(A-C) Dorsal view of HH33 chicken wings 120 h after electroporation (at HH14) showing: A) morphology with normal light, B) transfection efficiency using a β-actin promoter-driven RFP plasmid, and C) functional activity (GFP expression) of the Gdf5-associated LBI443 within the elbow and wrist joints. D) In situ hybridization for GDF5 in HH33 chicken wings shows a pattern of expression within the elbow and wrist that overlaps LBI443 activity. Magnified regions (white boxes) are shown in adjacent panels for the elbow (A’-G’) and wrist regions (A’’-C’’). Activity was also present in joint digits (data not shown). The fluorescence images (RFP and GFP) of dorsal and transverse views are overlaid with light field images to denote limb boundaries. E-G) Section in situ hybridization was also performed in HH27 chicken wings 24h after electroporation to show overlapping LMX1B (E & E’), GDF5 (F & F’) and GFP (representing LBI443 activity) expression. A yellow dashed line in the magnified elbow regions highlights the joint space.
Figure 7. Lmx1b-Bound PCRs Correspond to VISTA Enhancer Browser CRMs

Correlation of 9 Lmx1b bound PCRs with Vista Enhancer Browser CRMs showing activity in transgenic E11.5 mice (used by with permission from the Vista Enhancer Browser database (Visel et al., 2007)). Each illustration includes a screen shot of the UCSC browser showing ~20 kb genomic region with the Lmx1b ChIP-seq tracks (replicate1, purple; replicate2, black), input track (blue), the basewise conservation by phylop (Consrv), the called Lmx1b-bound interval (LBI, orange) and the corresponding VISTA element (black). The top panel has the y-axis labeled and is enlarged to demonstrate the layout for each interval.
Table 1. Gene Associations to Lmx1b-bound PCRs with Known Enhancer Activity

List of the 27 Lmx1b-bound potential cis-regulatory modules (PCRs) with functionally validated limb activity (Visel et al., 2007). Genomic coordinates for the Lmx1b bound intervals and the corresponding element ID of the validated CRM available from the Vista Enhancer Browser (Visel et al., 2007) are shown with the predicted Lmx1b-regulated targets.

<table>
<thead>
<tr>
<th>LBI</th>
<th>LBI-chr</th>
<th>LBI-start</th>
<th>LBI-end</th>
<th>Element</th>
<th>Lmx1b Regulated Gene(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>chr1</td>
<td>75833091</td>
<td>75833611</td>
<td>hs1635</td>
<td>Obs1</td>
</tr>
<tr>
<td>73</td>
<td>chr10</td>
<td>45230026</td>
<td>45230416</td>
<td>mm87</td>
<td>Prdm1</td>
</tr>
<tr>
<td>98</td>
<td>chr10</td>
<td>119761840</td>
<td>119762124</td>
<td>hs1498</td>
<td>Hmga2</td>
</tr>
<tr>
<td>125</td>
<td>chr11</td>
<td>75049646</td>
<td>75050306</td>
<td>hs1445</td>
<td>Pafah1b1/Srr/Tlcd2</td>
</tr>
<tr>
<td>133</td>
<td>chr11</td>
<td>87422998</td>
<td>87423326</td>
<td>mm264</td>
<td>Mir301</td>
</tr>
<tr>
<td>159</td>
<td>chr12</td>
<td>25891661</td>
<td>25892106</td>
<td>hs388</td>
<td>Id2</td>
</tr>
<tr>
<td>214</td>
<td>chr13</td>
<td>56159816</td>
<td>56160413</td>
<td>hs1473</td>
<td>Nsd1</td>
</tr>
<tr>
<td>225</td>
<td>chr13</td>
<td>91187223</td>
<td>91188005</td>
<td>hs1432</td>
<td>Ssbp2</td>
</tr>
<tr>
<td>252</td>
<td>chr14</td>
<td>28972666</td>
<td>28972977</td>
<td>hs1436</td>
<td>Wnt5a</td>
</tr>
<tr>
<td>285</td>
<td>chr15</td>
<td>35533210</td>
<td>35533643</td>
<td>mm703</td>
<td>Osr2</td>
</tr>
<tr>
<td>304</td>
<td>chr16</td>
<td>54991738</td>
<td>54992351</td>
<td>hs1469</td>
<td>Nlkbiz</td>
</tr>
<tr>
<td>333</td>
<td>chr17</td>
<td>47989063</td>
<td>47989453</td>
<td>mm751</td>
<td>Apobec2</td>
</tr>
<tr>
<td>369</td>
<td>chr18</td>
<td>68900908</td>
<td>68910229</td>
<td>mm708</td>
<td>Tcf4</td>
</tr>
<tr>
<td>370</td>
<td>chr18</td>
<td>68910647</td>
<td>68911246</td>
<td>mm708</td>
<td>Tcf4</td>
</tr>
<tr>
<td>398</td>
<td>chr19</td>
<td>60015935</td>
<td>60016244</td>
<td>mm449</td>
<td>Emx2/E330013P04Rik/Rab11fip2</td>
</tr>
<tr>
<td>438</td>
<td>chr2</td>
<td>137247962</td>
<td>137248210</td>
<td>hs1278</td>
<td>Jag1</td>
</tr>
<tr>
<td>439</td>
<td>chr2</td>
<td>137853613</td>
<td>137853963</td>
<td>mm671</td>
<td>Jag1</td>
</tr>
<tr>
<td>456</td>
<td>chr2</td>
<td>171734144</td>
<td>171735524</td>
<td>hs1448</td>
<td>Cbln4</td>
</tr>
<tr>
<td>463</td>
<td>chr3</td>
<td>30104643</td>
<td>30105483</td>
<td>hs1433</td>
<td>Sec62</td>
</tr>
<tr>
<td>464</td>
<td>chr3</td>
<td>30108049</td>
<td>30108319</td>
<td>hs1433</td>
<td>Sec62</td>
</tr>
<tr>
<td>481</td>
<td>chr3</td>
<td>67070452</td>
<td>67071271</td>
<td>hs741</td>
<td>Shox2/Rsdc1/Ptx3</td>
</tr>
<tr>
<td>495</td>
<td>chr3</td>
<td>99665987</td>
<td>99666408</td>
<td>hs1428</td>
<td>Gm42869</td>
</tr>
<tr>
<td>529</td>
<td>chr4</td>
<td>98223273</td>
<td>98223749</td>
<td>hs1484</td>
<td>Atg4c/Dock7/Nfia</td>
</tr>
<tr>
<td>569</td>
<td>chr5</td>
<td>99368833</td>
<td>99369455</td>
<td>mm678</td>
<td>Prkg2</td>
</tr>
<tr>
<td>582</td>
<td>chr6</td>
<td>4553349</td>
<td>4553603</td>
<td>mm400</td>
<td>Col1a2</td>
</tr>
<tr>
<td>643</td>
<td>chr7</td>
<td>112826169</td>
<td>112826588</td>
<td>hs1314</td>
<td>Usp47</td>
</tr>
<tr>
<td>720</td>
<td>chr9</td>
<td>71976098</td>
<td>71976463</td>
<td>hs357</td>
<td>Gm37879/Prtg/Tcf12</td>
</tr>
</tbody>
</table>
References


Supplemental Information

Table S1. Lmx1b-ChIP-seq intervals in E12.5 mouse limb.
Tabulated format of the genomic intervals bound by Lmx1b (LBI) that were identified in mouse limbs (E12.5) in both ChIP-seq replicates. Corresponding chromosome (chr), genomic location in mouse (mm10 reference genome) (LBI-Start and LBI-end), interval length, peak summit, peak value, average value (Avg Val) and bin count are listed (1).

Click here to Download Table S1

Table S2. Correlation between Lmx1b-bound Intervals and Chromatin Regulatory Marks.
Comparative analysis of Lmx1b-bound (LBI) intervals to p300, H3K27Ac, H3K4me2, RNA Pol II, Med12, and H3K27me3 ChIP-seq data. Conservation of each Lmx1b-bound interval is also indicated. The data is sorted by the number of chromatin regulatory marks (# of Marks) and by whether both active and repressor marks were present (Both marks). The shaded columns were not used to determine the number of regulatory marks. A potential regulatory modules (PCRM) is an Active-PCRM if has at least two chromatin regulatory marks associated.

Click here to Download Table S2
**Table S3. Target genes of Lmx1b-bound potential regulatory regions**

Comparison of Lmx1b-bound potential regulatory modules (PCRM) and genes differentially expressed in the presence of Lmx1b. A total of 292 PCRM are associated to 254 genes (Assoc gene). Note that there are multiple PCRM associated to different genes and several genes are associated to multiple PCRM. LBI corresponds to Lmx1b-bound interval number from table S1 for reference. The PCRM are categorized based on their chromatin regulatory marks (Reg Marks). If both active and repressor chromatin regulatory marks are present the PCRM is classified as Both-PCRM (See S2 for reference). The background color in this column indicates whether the associated genes are upregulated (red) downregulated (green) or whether the PCRM has both upregulated and downregulated (yellow) gene associations. The distance from the gene to the PCRM (Distance) is included with the fold change (Fold) and p-values from the published gene array data (Feenstra et al., 2012).

Click here to Download Table S3

**Table S4. Lmx1b-PCRM-associated genes present within functional categories**

PCRM associated genes classified according to functional categories and annotated functions as outlined in Figure 2C & D. P-value, number and names of molecules in each assigned category are specified if available.

Click here to Download Table S4
Table S5. *Lmx1b* bound Potential Cis Regulatory Modules (PCRM)s that correspond to functionally validated elements from the VISTA enhancer browser.

List of 91 *Lmx1b* bound PCRM$s$ and corresponding element ID of the VISTA elements - (Visel et al., 2007), followed by the candidate target gene ID and enhancer activity of the tested element. Note that 71 *Lmx1b* bound PCRM$s$ are functionally active according to the VISTA elements available form the VISTA enhancer browser.

Click here to Download Table S5

Table S6. Primers used for ChIP-qPCR Validation

*Lmx1b* bound Interval (LBI) number and primers sequence for the specified potential target validation by ChIP-qPCR. Primers are in a 5’ to 3’ orientation.

Click here to Download Table S6
Figure S1. Lmx1b bound intervals display a distribution that corresponded to a limb specific pattern and are enriched in genomic regions associated to active regulation.

A) Heatmap (top) and summarized average plots (bottom) showing the distribution of tagged sequences retrieved from the Lmx1b ChIP-seq and input DNA from limb tissue around p300 ChIP-seq intervals in limb, forebrain and midbrain (Visel et al., 2009). More Lmx1b tagged sequences overlap with p300 intervals in the limb than in the forebrain or midbrain with a 3-fold enrichment for the Lmx1b ChIP-seq retrieved tagged sequences over input DNA. B) Distribution of Lmx1b ChIP-seq and input retrieved tagged sequences around regions associated to active regulation determined by ChIP-seq (H3k27Ac, H3Kme2, RNA Pol II and Med12) (Visel et al., 2009, Berlivet et al., 2013, Cotney et al., 2013, DeMare et al., 2013) that overlapped Lmx1b bound intervals (LBI). Lmx1b ChIP-seq tagged sequences are enriched (~3-fold) within genomic regions associated to chromatin regulatory marks (H3K27Ac, H3K4me2) in comparison to input DNA and it is greater (4-fold) around regulatory regions undergoing active transcription (RNA Pol II, Med12). C) Bar graph depicting the number of LBIs that overlap with the different marks associated to cis-regulatory activity, where the number the percentage of LBIs overlapping chromatin regulatory marks. Note that the colors for each of the marks matched those used above for the heatmaps.
Figure S2. Enrichment of marks associated to potential cis-regulatory modules and Lmx1b regulated genes in Lmx1b bound intervals potential cis regulatory modules.

A) Overlap with at least 2 chromatin regulatory marks yields a 30 times higher number of potential cis-regulatory modules (PCRM) identified in the Lmx1b ChIP-seq dataset (617) in comparison to randomly selected genomic regions (n=5 groups, each with 735 random genomic intervals, One sample t-test p<1e-4, mean 19.6 ±6.6). B) Lmx1b bound PCRM are enriched within Lmx1b regulated genes. The number of PCRM associated to Lmx1b regulated genes is ~5 times higher within Lmx1b identified PCRM (292) compared to randomly selected enhancer regions based on H3K27Ac (Cotney et al., 2013) (n=5 groups, each with 292 random genomic intervals, One sample t-test p<1e-4, mean 60.6 ±7.1) and ~3 times higher for the number of Lmx1b regulated genes associated to a PCRM (n=5 groups, each with 292 random genomic intervals, One sample t-test p<1e-4, mean 75.8 ±10.8).
Figure S3. Distribution of Lmx1b-bound PCRM and associated Lmx1b-regulated genes

The genomic distribution of Lmx1b bound intervals (LBIs) is represented on mouse chromosomes. Yellow marks along the chromosome indicate potential *cis*-regulatory modules (PCRM), while light grey marks indicate LBIs that do not meet our criterion of a PCRM (≥ 2 chromatin regulatory marks). Location of Lmx1b-regulated genes at E12.5
(Feenstra et al., 2012) are indicated beside each chromosome, blue indicates association with a PCRM and dark grey indicates non-associated genes. A summary of the PCRM and gene distribution is shown above the chromosomes. No Lmx1b-bound PCRM are identified within the X chromosome.
Figure S4. Prediction of canonical pathways affected

Bar-chart representation of pathways affected according to PCRM-associated genes. Orange bars correspond to an overall upregulation of indicated pathways whereas blue is designated for downregulated ones.
References


