Available online at www.sciencedirect.com

SCIENCE DIRECT

Genomics 84 (2004) 565–576

GENOMICS

www.elsevier.com/locate/ygeno

The human LMX1B gene: transcription unit, promoter, and pathogenic mutations

Jennifer A. Dunston[a], Jeanette D. Hamlington[a,1], Jayshree Zaveri[a], Elizabeth Sweeney[b], Julie Sibbring[c], Catherine Tran[a], Maria Malbroux[a], John P. O'Neill[a], Roger Mountford[c], Iain McIntosh[a,*]

[a]*McKusick-Nathans Institute of Genetic Medicine, 733 N. Broadway/BRB 407, Johns Hopkins University, Baltimore, MD 21287, USA*
[b]*Merseyside and Cheshire Clinical Genetics Service, Royal Liverpool Children's Hospital, Alder Hey, Liverpool, L12 2AP, UK*
[c]*Merseyside and Cheshire Molecular Genetics Laboratory, Liverpool Women's Hospital, Liverpool, L8 7SS, UK*

Received 29 April 2004; accepted 3 June 2004

Abstract

LMX1B is a LIM-homeodomain transcription factor required for the normal development of dorsal limb structures, the glomerular basement membrane, the anterior segment of the eye, and dopaminergic and serotonergic neurons. Heterozygous loss-of-function mutations in LMX1B cause nail patella syndrome (NPS). To further understand LMX1B gene regulation and to identify pathogenic mutations within the coding region, a detailed analysis of LMX1B gene structure was undertaken. 5′-RACE and primer extension identified a long 5′-untranslated region of 1.3 kb that contains two upstream open-reading frames (uORFs). Transient transfection assays showed that sequences required for basal promoter activity extend no further than 112 bp upstream. An additional 47 mutations have been identified in the coding region, as well as nine deletions of large portions of the gene, but not in the promoter or highly conserved intronic sequences. The range of mutations and the identification of uORFs suggest further complexity in the regulation of LMX1B expression.

Keywords: LMX1B; Nail patella syndrome; 5′-UTR; Upstream open-reading frame; Transient transfection; Mutation; Conserved sequence; Real-time PCR

Introduction

LMX1B is a member of the LIM-homeodomain (LIM-HD) family of transcription factors that play a variety of roles during development to determine body pattern in vertebrates and invertebrates [1,2]. LIM domains comprise two zinc-finger motifs and participate in protein-protein interaction [3,4]. The homeodomains of LIM-HD proteins contain a characteristic signature that has been conserved through evolution; this signature is the amino acid sequence TGL at positions 38–40 in the homeodomain [5]. The HD of LMX1B is known to bind the FLAT element (TTAA-TAATCTAATTA) in electrophoretic mobility-shift assays (EMSA) [6–8].

The LMX1B gene is composed of eight exons spliced to give a ~7-kb mRNA [7,9]. Alternative splicing of the last 21 nucleotides of exon 7 has been reported [10], resulting in predicted proteins of 372 and 379 amino acids. The functional difference of the two isoforms is unknown. The second and third exons of the gene each encode a LIM domain, and the coding sequence for the homeodomain is divided among exons 4, 5, and 6.

Analyses of *lmx1b*$^{-/-}$ mice have shown that Lmx1b is required for a wide range of developmental processes including dorso-ventral patterning of the limb, differentiation of dopaminergic and serotonergic neurons, patterning

* Corresponding author. Fax: (410) 502 5677.
E-mail address: mcintosh@jhmi.edu (I. McIntosh).
[1] Present address: University of Texas Health Science Center San Antonio, San Antonio, TX 78229, USA.

0888-7543/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2004.06.002

of the skull, and normal development of the kidney and eye [10–16]. Heterozygous loss-of-function mutations in LMX1B result in the pleiotropic phenotype, nail patella syndrome (NPS, OMIM 161200; [7–10,17–19]). Although classically regarded as a disorder of connective tissue with characteristic defects in the nails, patellae, elbows, and pelvis [20], the description of nephropathy and glaucoma in affected individuals [21,22], together with knowledge of the role LMX1B plays in development, suggests a broader phenotype [23]. Missense mutations are concentrated within the homeodomain and the canonical residues of the LIM domains [17–19], with frameshift and nonsense mutations more widely distributed throughout the coding region. Direct sequencing of the exons and flanking introns identifies approximately 85% of NPS mutations [17]. Additionally, analysis of intragenic polymorphisms and Southern blotting reveal substantial deletions of the coding sequence [17]. Balanced translocations, presumably disrupting LMX1B, have also been reported in NPS patients [24,25].

A detailed analysis of LMX1B gene structure was undertaken to identify potential regulatory elements and pathogenic mutations in individuals with NPS that lack coding or splice sequence mutations. Functional elements of the LMX1B gene, including the promoter, poly(A) cleavage signal sequence and putative regulatory regions, were identified and screened for mutations. Deletions encompassing various portions of the gene, including a whole gene deletion, were identified in 9 unrelated families. Sequencing the coding region identified an additional 47 novel mutations. No pathogenic mutations were identified in the promoter, poly(A) cleavage signal sequence, or putative regulatory regions; however, sequence variants were identified. A subset of these variants was tested for a functional effect on basal promoter activity.

Results

Genomic sequence of LMX1B

Prior to the completion of the human genome sequence, BAC RPCI-11-265A2 was identified as containing the LMX1B gene [26]. Long-range PCR was employed to amplify the sequence between exons 3 and 4, and 6 and 7 for sequence determination. The sequence 5′ of the translation initiation codon, and 3′ of the termination codon (NM_002316), was determined by sequencing of subclones of BAC RPCI-11-265A2. A total of 4.0 kb 5′ and 5.9 kb 3′ were determined. This sequence was in agreement with that determined subsequently by Celera and the Human Genome Project [27,28]. The latter sources were employed to obtain the sequence of the 75.3-kb second intron.

Transcription start-site mapping

HEK 293 cells and primary human fibroblasts were tested for LMX1B expression by RT-PCR. Both cell lines express LMX1B, although expression was more robust in HEK 293 cells (data not shown). For this reason, RNA from HEK 293 cells was chosen as a source of RNA for transcription start-site mapping.

Initial attempts at mapping the transcription start site using 5′RACE suggested a 5′-UTR of about 100 nt, consistent with the average size of a 5′-UTR of a eukaryotic gene [29]. However, this transcription start site could not be confirmed by another method such as primer extension. Analysis of the genomic sequence indicated the presence of four CpG islands at the 5′ end of the gene (Fig. 1). The GC-rich nature of the 5′-UTR probably allowed the formation of stable secondary structure, leading to the premature termination of reverse transcription during the RACE reactions.

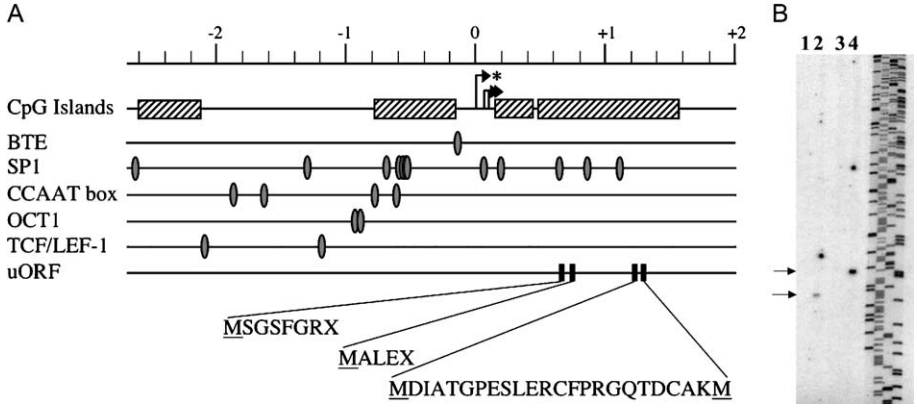


Fig. 1. The LMX1B promoter. (A) The transcription start sites determined by 5-RACE (asterisk) and PE are indicated by arrows in the context of the LMX1B locus. The scale indicates the distance in kb from the start site determined by 5-RACE. CpG islands are indicated by boxes (diagonal lines). Potential transcription factor binding sites are indicated by ovals, and uORFs are indicated by black boxes with the translation drawn below. (B) Primer extension reactions of HEK 293 poly(A)$^+$ RNA (lanes 2 and 4) and control RNA (lanes 1 and 3). Separate primer extension reactions were performed with primers LMX-5′Rc (lanes 1 and 2) and LMX-5′Ri (lanes 3 and 4), which are 1051 and 1027 nt, respectively, upstream of the translation start site. The products are indicated by arrows with apparent sizes of 226 and 237 nt for LMX-5′Rc and LMX-5′Ri, respectively. A dideoxy sequencing ladder was run as a size marker.

To circumvent the difficulties commonly associated with GC-rich genes [30], overlapping RT-PCR was used to "walk out" from the translation start site. Primers within regions known to be transcribed were then used for 5′-RACE, allowing for successful identification of the transcription start site by 5′-RACE and primer extension (Fig. 1). A single transcription start site was identified by sequencing 5′-RACE PCR product directly or by sequencing clones (n=3). The 5′-UTR of the LMX1B gene is 1.3 kb upstream of the translation start site defined by Dreyer et al. [7]. The transcription start site is located within a cluster of CpG islands (Fig. 1) and is not associated with a consensus TATA box. Genes lacking a TATA box tend to have multiple transcription start sites [30]. Although a single product was obtained with 5′-RACE, products of two different sizes were obtained with primer extension (Fig. 1). For clarity, the transcription start site identified by 5′-RACE will be designated position 1 in the transcript and referred to as the transcription start site (TSS); yet, the use of multiple start sites in vivo cannot be excluded. As indicated in Fig. 1, numerous sequence elements such as GC box, CCAAT box, and initiator sequence were identified in the vicinity of the transcription start site.

Inspection of the 5′-UTR sequence revealed the presence of two upstream open-reading frames (ORF) and an upstream start codon in-frame with the published ORF. The presence of an upstream start codon is consistent with the orthologous sequence reported for *Xenopus* [31]. Examination of the genomic sequence of other species, chick (AY163158) mouse (NT_039206.), and rat (NW_047652), revealed an additional 23 amino acids in frame with the reported start codon in all species. These 23 amino acids are identical in human, mouse, and rat. Only two differences occur between mammals and chick, and four occur between mammals and frog. The additional coding sequence was not present in the available genomic sequence for LMX1A genes from these species. The extension of the open-reading frame predicts a protein of 395 or 402 amino acids, depending on exon 7 splice-site selection.

3′ boundary of the LMX1B transcriptional unit

The 3′ boundary of the gene was determined by 3′-RACE using human fetal kidney poly(A)$^+$ RNA as template. Further evidence of the 3′ boundary of the LMX1B gene was obtained by searching the human EST GenBank database. Three IMAGE clones {1604421 (AA988020 and BX107616), 3085162 (BF507905), and 6586150 (BU557932)} mapped to the 3′ boundary of the gene. The clones were derived from a neuroendocrine carcinoid library (Lu5), a mammary gland adenocarcinoma library, and a mixed tissue library, respectively, and included a poly(A) cleavage signal sequence and a poly(A) tail that did not align with the genomic sequence. An EST (BM687469) that aligned with the 3′-UTR of the LMX1B gene, but did not include a poly(A) tail, was identified from a human fetal eye library. Lmx1b expression has been reported in ocular tissues of the mouse [13]. Each of these ESTs and the 3′-RACE product utilized the only consensus poly(A) cleavage signal sequence, AAUAAA, within 8.5 kb of the translation termination codon. These results indicate that the 3′-UTR of the gene is approximately 4.6 kb. Together with the 5′-UTR of 1.3 kb and a coding region of 1206 nucleotides, the total predicted transcript length is in agreement with the ~7-kb transcript observed by Northern blot analysis [7].

Identification of cis-acting regulatory regions

The ability of LMX1B 5′-flanking sequence to direct transcription of a reporter gene in primary human fibroblast cells was studied. The initial reporter construct tested, −2699/+807, directed transcription of the reporter gene at a level 9-fold greater than background. Since this reporter construct contained 2699 bp upstream of the transcription start site identified in HEK 293 cells by 5′-RACE, a 5′-deletion series of this construct was generated to allow for localization of basal promoter activity (Fig. 2A). Deletion of sequence from −2699 to −442 resulted in a slight, but significant ($P < 0.01$), increase in reporter activity. Further deletion to −379 resulted in a significant increase in reporter activity relative to either the −2699/+807 or −442/+807 construct ($P < 0.001$). The smallest region of 5′-flanking sequence tested was −112/+807 construct that contained 112 bp of 5′-flanking sequence. The level of reporter activity from this construct was equivalent to that seen by the −379 construct ($P > 0.05$). These results suggest that there may be a relatively weak repressor between −2699 and −442 and a strong repressor between −442 and −379. Sequences between −379 and −112 had no effect on reporter activity; therefore, the sequences required for basal transcription appear to be contained within the −112/+807 construct.

To test for enhancer activity, larger constructs were tested for reporter activity and compared to the −2699/+807 construct. Inclusion of additional sequence to −9900 resulted in a significant decrease in reporter activity relative to the −2699/+807 construct (one-way ANOVA, Tukey posttest, $P < 0.05$; Fig. 2A). This decrease is likely to be related to the relative mass of the test constructs, since an equal mass of test construct was utilized for each assay. When the size of the insert and vector is taken into account, the −9900 construct is approximately twice as large as the −2699 construct, which corresponds to the approximate 50% decrease in reporter activity. Enhancers were not identified in this region either because they are not active in fibroblasts or because they are located in regions that have not been tested.

Sequence comparison was used to identify putative regulatory regions within the LMX1B gene from 98.5 kb 5′ of the transcription start site to 113.6 kb 3′ of the transcript boundary. Local alignments of genomic sequence were determined for the *LMX1B* genes from human (NT_008470, build 32) and mouse (NT_039205, build 30). Comparison of

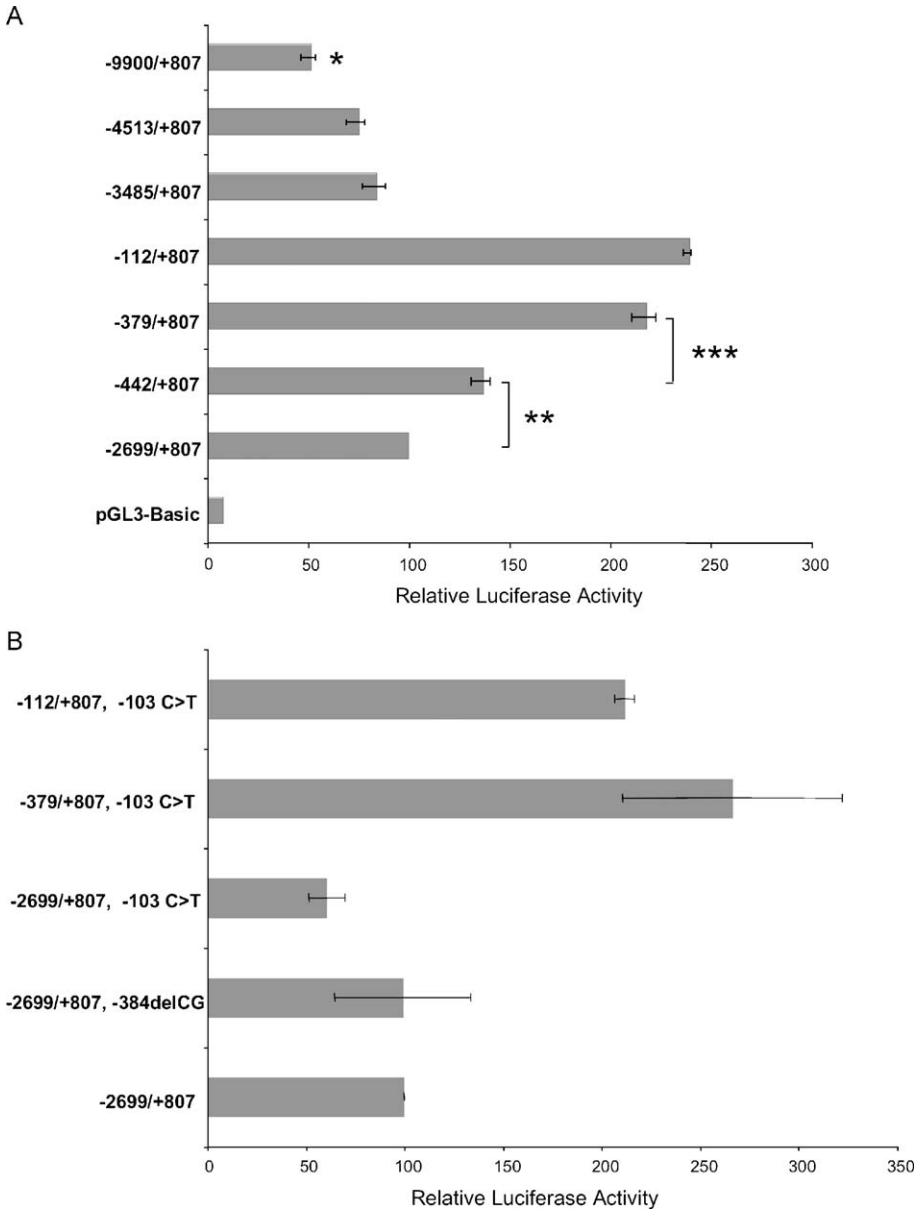


Fig. 2. The activity of LMX1B promoter and enhancer regions in human fibroblasts. Relative luciferase activity is plotted (with standard error of 3 replicates in at least 2 experiments), with the activity of the −2699/+807 construct set at 100, for (A) a 5′-deletion series to localize the LMX1B promoter, and additional 5′-flanking sequence to identify enhancers; and (B) sequence variants identified within the promoter region. Reporter constructs are numbered relative to the transcription start site identified in HEK 293 cells (Fig. 1). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

genomic sequence (data not shown) revealed strong conservation in coding sequence regions and widespread conservation of noncoding sequence. Within intron 2 of the gene, 5 regions longer than 250 bp with greater than 85% identity were found to be conserved between human and mouse. The degree of conservation within these regions is consistent with a functional role, possibly as gene expression regulatory regions [32,33].

Mutation screening

Sequencing of the exons and immediate flanking intron sequence in 105 unrelated individuals identified 47 novel mutations (Table 1), and 42 recurrences of mutations described previously [17,19]. Each mutation was shown to be present in other affected family members, and absent in unaffected parents of sporadic cases, using the method indicated in Table 1. No pathogenic mutations were identified in 16 individuals by direct sequencing. By a combination of Southern blot analysis, real-time PCR assays, and segregation analyses (in instances when DNA was available from multiple family members), deletions of substantial portions of the coding region were identified in 9 of 28 families (comprising the 16 index cases in which no mutation was identified by sequencing and an additional 12 in which sequencing had previously failed to identify a

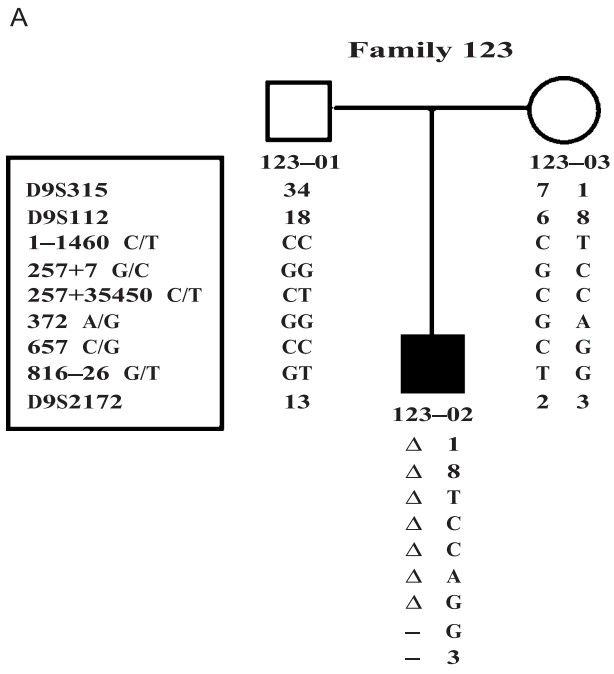
Table 1
NPS mutations identified within the coding region of LMX1B

Mutation	Nucleotide	Domain	Detection	Putative effect
p.G38fs	c.113-114delGGinsT	Leader	−*Sma*I	Frameshift, PTC
p.C56S	c.167T>A	LIM-A	SSCP	Disrupt Zn-finger
p.E57X	c.169G>T	LIM-A	+*Bfa*I	PTC
p.Q60X	c.177C>T	LIM-A	SSCP	PTC
p.N72fs	c.214_321del	LIM-A	SSCP	Frameshift, PTC
p.W76R	c.226T>C	LIM-A	+*Msp*I	Disrupt Zn-finger
p.W76X	c.227G>A	LIM-A	+*Mae*I	PTC
p.E78X	c.232G>T	LIM-A	+*Mae*I	PTC
p.C94W	c.282C>G	LIM-A	SSCP	Disrupt Zn-finger
p.K104fs	c.310_312del1	LIM-A	Sequence	Frameshift, PTC
p.Q105fs	c.313del	LIM-A	Sequence	Frameshift, PTC
p.Y107fs	c.320del	LIM-A	Sequence	Frameshift, PTC
p.Q108fs	c.322del	LIM-A	SSCP	Frameshift, PTC
c.326+1G>A	c.326+1G>A	LIM-A	SSCP	Loss of exon 2, framshift, PTC
c.326+3T>C	c.326+3T>C	LIM-A	SSCP	Loss of exon 2, framshift, PTC
c.327-2_334del	c.327-2_334el	LIM-B	SSCP	Loss of exon 3, framshift, PTC
p.C115R	c.343T>C	LIM-B	+*Bst*UI	Disrupt Zn-finger
p.E120fs	c.356_357del1	LIM-B	Sequence	Frameshift, PTC
p.H137R	c.410A>G	LIM-B	Sequence	Disrupt Zn-finger
p.R148P	c.443G>C	LIM-B	SSCP	Disrupt Zn-finger
p.V187_K247del	c.560–1848_741+3del	Linker, HD	Southern	Loss of exon 4, Frameshift, PTC
p.R221fs	c.661del	HD	SSCP	Frameshift, PTC
p.P222fs	c.662_664insA	HD	Sequence	Frameshift, PTC
p.S242X	c.725 C>A	HD	−*Taq*I	PTC
p.K243fs	c.727_728del1	HD	Sequence	Frameshift, PTC
p.C245X	c.735_736CC>AA	HD	Sequence	PTC
p.V248_Q265del	c.742-9_785del51	HD	Gel	DNA binding
c.742-2A>G	c.742-2A>G	HD	+*Msp*I	Loss of exon 5
c.742-1G>A	c.742-1G>A	HD	SSCP	Loss of exon 5
c.742-1G>C	c.742-1G>C	HD	SSCP	Loss of exon 5
p.R249G	c.745C>G	HD	Sequence	DNA binding
p.T256R	c.767C>G	HD	Sequence	DNA binding
p.L258F	c.772C>T	HD	−*Hae*III	DNA binding
p.S259fs	c.774_775del	HD	Sequence	Frameshift, PTC
p.R261C	c.781C>T	HD	+*Hha*I	DNA binding
p.V263A	c.788T>C	HD	−*Ava*II	DNA binding
p.V265F	c.793G>T	HD	−*Bst*NI	DNA binding
p.V265L	c.793G>C	HD	SSCP	DNA binding
p.F267L	c.799T>C	HD	Sequence	DNA binding
p.F267I	c.799T>A	HD	+*Hin*fI	DNA binding
p.Q268H	c.804 G>T	HD	Sequence	DNA binding
c.819+2T>G	c.819+2T>G	HD	SSCP	Loss of exon 5
c.820-2del	c.820-2del	HD	SSCP	Loss of exon 6, frameshift, PTC
c.820-1G>A	c.820-1G>A	HD	−*Pst*I	Loss of exon 6, frameshift, PTC
p.L277fs	c.829del	C-term	+*Ava*II	Frameshift, PTC
p.R280fs	c.838_847del, 856_857ins (845_856,845_876,853_856,854_856)	C-term	Gel	Frameshift, PTC

The domain, means of detection, and putative effect are indicated.

mutation) (Fig. 3). The mutation designated as p.V187_K247del was identified as an aberrantly migrating fragment on a Southern blot of *Hin*cII/*Hin*dIII-digested genomic DNA in an affected mother and son. Subsequent long-range PCR and TA cloning determined the precise breakpoints. No other deletions generated aberrantly migrating fragments under these conditions (data not shown).

Of the families studied, no pathogenic mutations were identified in the 5′-flanking sequence, poly(A) cleavage signal sequence, or putative regulatory regions within intron 2. The sequence in or around the poly(A) cleavage signal sequence was invariant. Both rare variants and polymorphisms, defined as mutations found at a frequency greater than 0.01, were identified in the 5′-flanking sequence and intron 2 within the patient population (Table 2). Within the promoter region, a two-base deletion was found in a sporadic case (-384delCG, numbered relative to the transcription start site, Fig. 1), but was also found in the proband's unaffected father, although not in 100 control chromosomes, and is thus a rare neutral variant (see below).



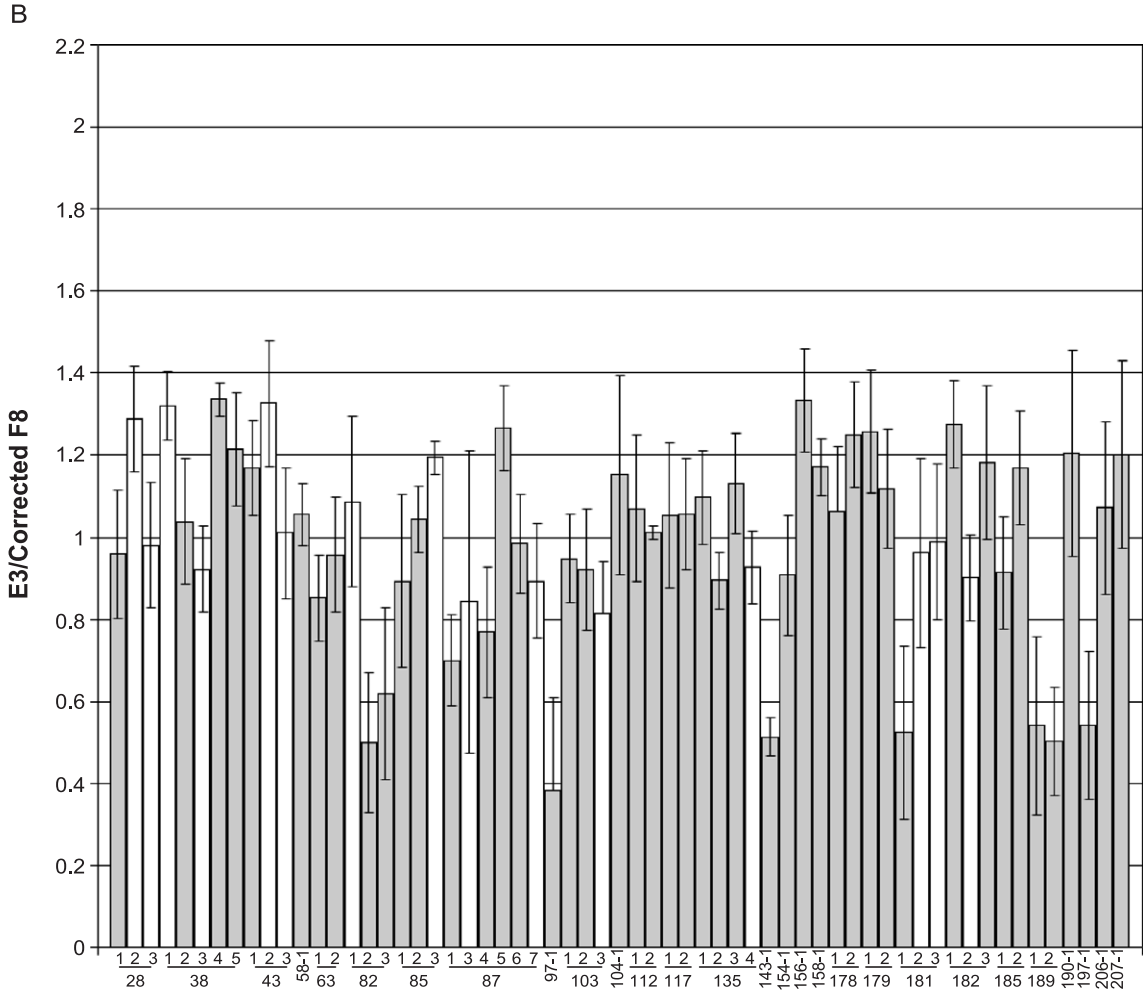


Fig. 3. Identification of LMX1B deletion mutations in N PS patients. (A) A large deletion is inferred in the proband since no paternal markers can be identified at least 100 kb upstream of LMX1B (D9S315) and up to exon 4 of the LMX1B gene. Markers in IVS6 and 3′ of the gene are uninformative. Paternity was confirmed using other autosomal markers. (B) Quantification of LMX1B copy number by real-time PCR relative to Factor VIII (adjusted for gender). Affected individuals are represented by the shaded bars, unaffected family members are white.

Table 2
Summary of sequence variants identified

Region	Variant	Frequency[c]
Promoter	−384delCG[a]	*n* = 1
Promoter	−103C/T[a]	0.61/0,39
IVS2c	262G/A[b]	*n* = 2
IVS2c	350C/T[b]	0.71/0.29
IVS2e	188C/T[b]	*n* = 1

[a] Bases are numbered relative to the transcription start site identified in HEK 293 cells.
[b] Bases are numbered relative to their position in the PCR product, with the first base of the forward primer called +1. The IVS2 regions were amplified with the primers listed in Table 3. No variants were identified in IVS2a,b,d,f.
[c] For allele frequencies the rare allele is indicated with its frequency (determined in >100 chromosomes). For rare variants, the number of times the variant was observed during sequencing of 10 individuals is shown.

A second variant identified in the promoter region (−103 T) was found in multiple individuals and shown to be in Hardy-Weinberg equilibrium in >100 control chromosomes. Three variants were identified in the intron 2 sequence studied: two in multiple unrelated individuals, indicating that they are likely to be polymorphic, and a third in a single affected individual, but not in their affected son.

Functional analysis of sequence variants

The functional consequences of the two variants identified in the 5′-flanking sequence, −384delCG and −103 T alleles, were tested in the transient transfection assay (Fig. 2B). The −384delCG allele was tested in the context of the −2699/+807 construct, and the −103 T allele was tested in the context of the −2699/+807, −379/+807, and −112/+807 constructs. The statistical test for functional consequences of the two variants located in the 5′-flanking sequence consisted of three groups of comparisons: 1, −2699/+807: WT; −103 C > T, and −384delCG; 2, −379/+807, WT and −103 C > T; and 3, −112/+807: WT and −103 C > T. One-way ANOVA indicated that there is no significant difference between the means of any of the −2699/+807 constructs. A two-tailed *t* test was used to test for a difference between the means of the WT and −103 T allele in the −379/+807 and −112/+807 groups. No significant difference was found.

Discussion

The transcription start site of the LMX1B gene was identified to define the LMX1B transcriptional unit and localize the promoter. The TSS is located within a cluster of four CpG islands and is not associated with a TATA box. Typically, genes associated with CpG islands and lacking a TATA box are thought to be regulated by the transcription factor Sp1 [30]. While Sp1-binding sites were found, none were located within the 112 bp directly upstream of the transcription start site, which was capable of driving expression of a reporter gene in transient transfection assays. However, the presence of Sp1-binding sites within the 5′-UTR of the gene may influence the level of transcription.

Features of the 5′-UTR suggest that LMX1B expression is controlled beyond the level of transcription initiation. The average length of a 5′-UTR in a vertebrate mRNA is reported to be between 90 and 210 nucleotides [34]. The 5′-UTR of the *LMX1B* gene is much longer (1.3 kb), has a high GC content, and contains uORFs. These features are found more commonly in transcripts that encode low-abundance proteins [35] and suggest that the levels of LMX1B protein may be regulated posttranscriptionally. The translation efficiency of an ORF is determined, in part, by the sequence context of the start codon. Sequences that conform to the Kozak consensus sequence, ccRcc*ATG*G, are more efficiently translated (reviewed by Kozak [36]). The italicised ATG represents the start codon, with positions −3 and +4, relative to the position of the A in the start codon, being most important in determining translation efficiency. In the LMX1B transcript, the uORF at position −639, relative to the published translation start site (NP_002307), is found within a weak Kozak consensus sequence with both positions −3 and +4 differing from the consensus. This ORF is probably bypassed by the scanning ribosome by a method referred to as leaky scanning [36]. In contrast, the context of the initiator codon for the second ORF at position −553 is found within a strong Kozak sequence. If this uORF were translated, the LMX1B ORF would be translated by a translation reinitiation method [36]. The next start codon at position −69 is within an adequate Kozak consensus sequence, with only one discrepancy from the consensus at position −3. The start codon for the published ORF is also within a stronger consensus sequence with a discrepancy at position +4. However, translation of the sequence beginning at position −69 yields an ORF that is continuous with the LMX1B open-reading frame, suggesting that the full-length protein is 395 or 402 amino acids (depending on the splicing of exon 7). This argument is supported by the conservation of this sequence among human, rodents, chicken, and Xenopus, and the lack of conservation immediately upstream. The nucleotide sequence is absolutely identical among human, mouse, and rat, and 84% identical between mammals and chick. In contrast, the 69 bp immediately upstream is 12–13% divergent between human and rodents, and 43% diverged between human and chick.

The uORFs may play a role in the translational control of *LMX1B*. Although the transcription start sites have not been identified in other species, uORFs can be identified within 1 kb upstream of the initiation codon in mouse and chick. Of interest, the short ORF initiating at −484 (relative to the ATG in human) is absolutely conserved in mouse, but the ORF initiating at −570 is 121 amino acids in length but bears no homology to any reported protein.

Since genetic evidence suggests that the variants identified upstream of the TSS are not pathogenic, it is not surprising that the activity of the reporter construct was not greatly reduced (Fig. 2B). In the context of this assay, these variants did not have any significant effect on reporter activity, but the possibility that these variants affect the basal level of LXM1B transcript in vivo cannot be excluded. The presence of a long 5′-UTR containing canonical Sp1 sites (Fig. 1A) raises the possibility of downstream transcription initiation sites, as observed for the PDGF-B and PTEN genes [37,38]. The identification of the transcription initiation sites utilized in vivo is complicated by the restriction of postnatal expression of LMX1B to the kidney and central nervous system [7,14].

The novel pathogenic mutations identified (Table 1) show a similar distribution to that described previously [17–19], being concentrated in the LIM- and homeodomains. For the first time, however, mutations have been identified in exon 1 (p.G38fs) upstream of the LIM-A domain, and in exon 6 (p.L277fs, p.R280fs), downstream of the homeodomain. Each of these mutations is predicted to introduce a premature stop codon. Of further interest, three deletions covering splice junctions (c327-2_334del, p.V248_Q265del, and c.820-2del) may allow for the maintenance of the reading frame via cryptic splicing; however, the predicted loss of conserved amino acids in the LIM- or homeodomains remains consistent with these being loss-of-function alleles. Confirmation of the effects of these mutations on RNA splicing requires the availability of tissue samples expressing LMX1B. The nomenclature for all mutations described previously has been amended to agree with current standards [39] and to accommodate the additional coding sequence identified (Fig. 1). A complete list is provided in Supplementary Table 1.

Although genomic deletions may be detected by Southern blotting or haploinsufficincy for intragenic markers [17], these methods require either substantial amounts of genomic DNA or samples from multiple family members. An example of the latter in Fig. 3A illustrates identification of a deletion extending from at least 100 kb upstream to at least exon 4 of the LMX1B gene in a sporadic case of NPS, with a 50% probability that the deletion extends into the 3′-flanking region. This affected individual presented with patellar hypoplasia, elbow dysplasia, proteinuria, iliac horns, and talipes, but only mild nail dysplasia, suggesting that the phenotype resulting from absence of one copy of LMX1B is no different from that seen in individuals with a range of point mutations [8].

Real-time PCR represents an alternative methodology for identification of genomic deletions that is especially useful in the absence of additional family information. Fig. 3B shows the use of a real-time PCR probe to exon 3 of LMX1B and identification of a deletion in 6 of 29 unrelated NPS families in whom DNA sequencing did not reveal a pathogenic mutation. Unfortunately, real-time PCR appears sensitive to the GC content of the target sequence as 4 independent sets of probe and primers failed to unambiguously identify known deletions of the GC-rich exon 2 (data not shown). Overall, deletions of all or part of LMX1B represent ~5% of all NPS mutations ([17], Fig. 3, data not shown).

A total of 137 NPS mutations have now been identified within the LMX1B coding sequence (Supplementary Table 1), excluding genomic deletions covering large portions of the gene. Of particular note, no missense mutations have been identified downstream of the homeodomain, suggesting that alteration of this domain would result in a phenotype distinct from NPS. The amino acid sequence of this domain is absolutely conserved between human and rat, and only a single difference between human and mouse. Furthermore, no amino acid substitutions that are not pathogenic mutations have been identified in sequencing over 200 unrelated individuals [8,17,19]. The observations that the same range of NPS symptoms is observed in patients with missense, nonsense, or deletion mutations, and that the range and severity of symptoms can vary as much within as between families, support the hypothesis that NPS is the result of haploinsufficiency for LMX1B [8,17,19,23]. This argument is also supported by the lack of any dominant-negative effects observed in vitro when expressing constructs with missense or truncation mutations [40]. It has become generally accepted that mutations resulting in premature termination codons (PTCs) result in the degradation of mutant transcripts via nonsense-mediated decay (NMD) to prevent synthesis of truncated proteins [41]. However, the presence of uORFs in the LMX1B 5′-UTR suggests that either LMX1B transcripts can escape NMD, or that the NMD pathway controls the stability of normal LMX1B transcripts [42]. The former scenario suggests that some NPS mutations may act in a dominant-negative manner, but that any effect on phenotype is too subtle to distinguish their effects from deletion mutations. The latter supports a model in which the level of LMX1B protein synthesis is under extremely tight control and any reduction in the amount of functional protein has an obvious downstream effect, namely NPS. These alternatives are not mutually exclusive but further work is required to dissect the regulation of uORF-containing transcripts in mammalian cells.

Methods

LMX1B sequencing

Southern blotting of BAC RPCI-11-265A2 DNA, with oligonucleotide probes derived from the LMX1B coding sequence, confirmed the presence of all 8 LMX1B exons within the BAC and was used to generate a restriction map of the BAC. Restriction bands containing exons 1 and 8 were isolated and ligated into pBluescript II KS+ vector. Clones were sequenced using fluorescent (Johns Hopkins

Medicine, Genetic Resources Core Facility, DNA Analysis Facility on an ABI 3700 Sequencer) or ^{33}P-labeled chain terminators (Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit, USB).

RT-PCR

Total RNA was isolated from normal primary human fibroblasts or human embryonic kidney (HEK) 293 cells using Trizol reagent (Invitrogen). Poly(A)$^+$ RNA was isolated from total RNA with the mRNA Separator Kit (Clontech) and reverse-transcribed with random hexamers using the ThermoScript RT-PCR System for First-Strand cDNA Synthesis (Invitrogen), with the addition of PCRx Enhancer (Invitrogen) to a final concentration of 2X. To test for gene expression in cell lines, exons 1 to 3 of the *LMX1B* gene were amplified using primers LMX-1F* and 3-ANTI. Following a 5-min denaturation at 95°C, 35 cycles of 95°C, 30 s, 61°C, 30 s, 72°C, 30 s were performed, followed by a 10-min extension at 72°C. Overlapping RT-PCR products were generated for the 5′-UTR using the following sets of primers: LMX-5′Fh and LMX-GSP2 (annealing temperature of 56°C, 3X PCRx Enhancer), LMX-5′Fr and LMX-GSP2c (annealing temperature of 60°C, 1X PCRx Enhancer), and LMX-5′Fu and LMX-5′Ri (annealing temperature of 60°C, 1X PCRx Enhancer). PCR products were sequenced with the same primers used for PCR using ^{33}P-labeled dideoxy terminators (Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit, USB). Sequences for all primers are provided in Table 3.

Transcription start-site mapping

The 5′ boundary of the human LMX1B transcriptional unit was mapped using 5′-rapid amplification of cDNA ends (RACE) and primer extension (PE). Poly(A)$^+$ RNA from HEK 293 cells, treated with DnaseI (DNA-free, Ambion), was used as a template for all reactions. The GeneRacer Kit (Invitrogen) was employed for the 5′-RACE reactions according to the manufacturer's instructions with some exceptions. The reverse transcription reactions were performed with the ThermoScript RT-PCR System for First-Strand cDNA Synthesis (Invitrogen) using a gene-specific primer (LMX-5′Ri). Nested PCR were performed using the PCR$_x$ Enhancer System (Invitrogen) with the following sets of primers: GR-5′ and LMX-5′Rc followed by GR′5′b and LMX-5′Rd. The reactions were denatured for 5 min at 95°C and amplified for 35 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 30 s. The purified PCR product was sequenced with the LMX-5′Rd primer using ^{33}P-labeled dideoxy terminators (Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit, USB). Additionally, the product was cloned into the pCR2.1-TOPO vector (Invitrogen). Resulting clones were sequenced with T7 primer at the Johns Hopkins Medicine, Genetic Resources Core Facility, DNA Analysis Facility on an ABI 3700 Sequencer.

For PE, 2 μg of HEK 293 poly(A)$^+$ RNA or control RNA (2 μg yeast total RNA) was reversetranscribed with LMX-5′Ri and LMX-5′Rc primers labeled with [γ-^{32}P]ATP (NEN Life Science). Labeled primer was mixed with the RNA and heated to 65°C for 5 min, to denature secondary structure, and immediately transferred to ice. The ThermoScript RT-PCR System for First-Strand cDNA Synthesis (Invitrogen) was used to transcribe RNA at 64°C to allow for gene-specific annealing of the primer and reverse transcription. Primer extension products were denatured for 10 min at 90°C before being loaded on a 6% denaturing polyacrylamide gel. Sequencing reactions of pUC18 plasmid with –40F primer (USB) and plasmid X6.6, a BAC 265A2 subclone, with LMX-5′Ri primer (Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit, USB) were loaded at the same time as the primer extension products. The gel was run at 60 W and exposed to Kodak XAR film with an intensifying screen at −80°C overnight.

3′-End mapping

Human fetal kidney poly(A)$^+$ RNA (50 ng, Clontech) was denatured at 70°C and transferred to 50°C to prevent the formation of secondary structure. The RNA was primed with the AP primer and reverse-transcribed at 42°C with Superscript II RT (Invitrogen). RT was omitted from one reaction to serve as a negative control. The target cDNA was amplified using the Expand Long Template PCR System (Roche) with the LMX-3′Fd and UAP primers. After incubating the reactions for 2 min at 94°C, 30 cycles of the following were initiated: 30 s at 95°C, 30 s at 63°C, and 4 min at 68°C. The product was sequenced with primers LMX-3′Fd, LMX-3′Fe, LMX-3′Ff, LMX-3′Fh, LMX-3′Rd (Supplementary Table 2), and AUAP (Invitrogen) using ^{33}P dideoxy terminators (Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit, USB).

Transient transfection of primary human fibroblasts

The Dual-Luciferase Reporter Assay System (Promega) was used to determine the relative promoter activity of a series of test plasmids containing varying amounts of 5′-flanking sequence of the *LMX1B* gene. Each of the test plasmids was constructed in the pGL3-Basic Vector using standard subcloning procedures. Inserts were derived from subclones of BAC 265A2. For each set of transfection experiments, a series of control transfections were performed. These included transfection of no DNA as a negative control, the empty pGL3-Basic vector to measure background activity of the firefly luciferase, the pGL3-Control as a positive control for firefly luciferase, the pRL-null plasmid as a positive control for Renilla luciferase, and each test plasmid alone to allow detection of promoter cross-talk in cotransfections. To quantitate promoter activity of each test plasmid, the test plasmids were cotransfected with

Table 3

Name	Strand	5′ base in LMX1B Sequence NT_008470.15	Sequence (5′→3′)
GR-5′	NA	NA	CTGGAGCACGAGGAGACTGA
GR-5′b	NA	NA	CTGACATGGACTGAAGGAGTA
LMX-5′Fu	F	31,198,231	TCCAGGGCCAGCCGTCCTC
LMX-5′Fr	F	31,198,373	AGTCGGGGGTTCCCAAGCG
LMX-GSP2c	R	31,198,868	TGAGCCGCCGGGTTTAGAG
LMX-5′Fh	F	31,198,727	GGACGGCACTATTTGACGTGG
LMX-5′Ft	F	31,197,827	CTTGATCCGCGCGTCCCAG
LMX-5′Rc	R	31,198,374	CTGTCGGCCGAGGTAGGTC
LMX-5′Rd	R	31,198,239	GGCCCTGGACTCGTGCAAG
LMX-5′Ri	R	31,198,399	GGAGTGGCCGCTTGGGAAC
LMX-GSP2	R	31,198,438	CAACATCTTGGCGCAGTCCGT
LMX-1F*	F	31,199,413	GGCAGACGGACTGCGCC
IVS2a_F	F	31,223,085	GAGCCCCTCTCTGCTGGTG
IVS2a_R	R	31,223,565	TGCTAAGCACTGGCTGTCTCC
IVS2b_F	F	31,223,498	CCGCTTGGCAGTTTTTAATC
IVS2b_R	R	31,223,984	CCTTCCCAACTCTCCTTTCC
IVS2c_F	F	31,235,584	CCCCTCTCCCTTCCTGAG
IVS2c_R	R	31,236,025	CAGGCCCTTCTCCTCCAC
IVS2d_F	F	31,243,373	CACATTCCCAGCTCCTGATT
IVS2d_R	R	31,243,776	AGCAAACCAGGCCTGATACA
IVS2e_F	F	31,261,723	CTGGGTAAATGCTGGGAGAA
IVS2e_R	R	31,262,097	GACCACATCCCTGGGTCTC
IVS2f_F	F	31,224,590	TGCAAGCTGTTAGTGACTGG
IVS2f_R	R	31,224,987	GCTCTAAAATCTGGGCCATC
3-ANTI	R	31,275,846	GAAGCAGCCCAGGTGGTAC
LMX-3′Fd	F	31,283,578	CACCACCCTAAAGTCTGCCTG
LMX-3′Fe	F	31,283,765	CATGTGGCCTCGTGTCAC
LMX-3′Ff	F	31,284,083	GAAGCGGTCATCTCATCTGC
LMX-3′Fh	F	31,285,116	CAGACCAAGGTGCCAGCAC
LMX-3′Rd	R	31,285,134	GTGCTGGCACCTTGGTCTG
LMX-3′Fm	F	31,285,726	CCTTGTACAGACAGATTCTC
LMX-3′Rg	R	31,286,049	GTGGCCATCAACAAATCTG

Primers were designed with the aid of the Primer3 program (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi).

pRL-null in triplicate. Briefly, primary human fibroblasts were seeded at 1.5×10^5 cells per well in a 6-well plate and grown overnight to approximately 70% confluency. For each well, 8 μl of Lipofectamine (Invitrogen) was used to transfect 2.5 μg of test plasmid and 0.25 μg of pRL-null plasmid. After 47 h, cells were processed for the Dual-Luciferase Reporter Assay System (Promega) and read in a BD Monolight 2010 Luminomter. The level of Renilla luciferase activity was used to normalize firefly luciferase activity for transfection efficiency. The luciferase data were converted into relative firefly luciferase activity per microgram of protein, as determined by the BioRad protein assay. The average value for the three trials is reported as a percentage of −2699/+807 activity. One-way analysis of variance (ANOVA) followed by a Tukey posttest comparing all pairs of columns (GraphPad Prism 4 statistical software, GraphPad) was used to test for significant differences between test plasmids.

Analysis of genomic sequence

Repetitive elements in genomic sequences were identified with the RepeatMasker program [43]. Percentage identity plots (PIPs) were created with the AdvancedPipMaker program [44]. Analysis of CpG islands was performed with the EMBOSS CpG plot [45]. The default settings require a minimum observed/expected ration of CpG dinucleotides of 0.6, a minimum percentage GC content of 50%, and a minimum length of 200 bp over a 100-bp window with 1-bp steps.

Putative transcription factor binding sites were identified using the TRANSFAC database and the MatInspector program [45,46]. The following matrix families were searched with a core similarity threshold of 1.00 an optimized matrix similarity: V$SP1F, V$CEBP, V$ECAT, V$PCAT, V$RCAT, V$OCT1, V$TBPF, and V$LEFF.

Mutation identification

The LMX1B coding sequence was screened for mutations as described [8]. Genomic deletions were identified by Southern blotting, genotyping flanking markers, and SNPs within the coding region [17,47] or by quantitative real-time PCR [48]. Real-time PCR was performed in triplicate with primers (forward 5′-TGCTGCTGCGTGTGTGA, reverse 5′-CTTCTCGTAGT-

CACCCTTGCA) and probe (5′-FAM-CCTTGAGCAC-GAATTC) specific for LMX1B of exon 3 using an ABI Prism 7900. Primers for the Factor VIII were employed as an internal control [48]. The ratio of exon 3 to factor VIII was plotted, after correcting for the gender of the subject, with standard errors calculated according to the supplier's instructions (ABI).

Acknowledgments

The authors are grateful to patients and families for their participation in this study, to Britt Silkey for assistance on obtaining blood samples, Hal Dietz, for the use of tissue culture facilities, Emily Niemitz for advice on real-time PCR, and Andrew McCallion and Joshua Mendell for helpful discussions. This work was supported, in part, by the Greenberg Center for Skeletal Dysplasias and grants from the National Institutes of Health (AR44702, I.M.), the Birth Defects Foundation, and the Skeletal Dysplasia Group (E.S).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2004.06.002.

References

[1] I.B. Dawid, R. Toyama, M. Taira, LIM domain proteins, CR. Acad. Sci. III 318 (1995) 295–306.
[2] J. Curtiss, J.S. Heilig, DeLIMiting development, BioEssays 20 (1998) 58–69.
[3] J.D. Johnson, W. Zhang, A. Rudnick, W.J. Rutter, M.S. German, Transcriptional synergy between LIM-homeodomain proteins and basic helix-loop-helix proteins: the LIM2 domain determines specificity, Mol. Cell. Biol. 17 (1997) 3488–3496.
[4] L.W. Jurata, G.N. Gill, Functional analysis of the nuclear LIM domain interactor NL1, Mol. Cell. Biol. 17 (1997) 5688–5698.
[5] S. Banerjee-Basu, A.D. Baxevanis, Molecular evolution of the homeodomain family of transcription factors, Nucleic Acids Res. 29 (2001) 3258–3269.
[6] M.S. German, J. Wang, R.B. Chadwick, W.J. Rutter, Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: building a functional insulin mini-enhancer complex, Genes Dev. 6 (1992) 2165–2176.
[7] S.D. Dreyer, et al., Mutations in *LMX1B* cause abnormal skeletal patterning and renal dysplasia in nail patella syndrome, Nature Genet. 19 (1998) 47–50.
[8] I. McIntosh, et al., Mutation analysis of LMX1B gene in nail-patella syndrome patients, Am. J. Hum. Genet. 63 (1998) 1651–1658.
[9] D. Vollrath, et al., Loss of function mutations in the LIM homeodomain gene, LMX1B, in nail-patella syndrome, Hum. Mol. Genet. 7 (1998) 1091–1098.
[10] M. Seri, et al., Identification of LMX1B gene point mutations in Italian patients affected with nail-patella syndrome, Int. J. Mol. Med. 4 (1999) 285–290.
[11] H. Chen, et al., Limb and kidney defects in *Lmx1b* mice suggest an involvement of *LMX1B* in human nail-patella syndrome, Nature Genet. 19 (1998) 51–55.
[12] H. Chen, et al., Multiple calvarial defects in *lmx1b* mutant mice, Dev. Genet. 22 (1998) 314–320.
[13] C.L. Pressman, H. Chen, R.L. Johnson, Lmx1b, a LIM homeodomain class transcription factor, is necessary for normal development of multiple tissues in the anterior segment of the murine eye, Genesis 26 (2000) 15–25.
[14] M.P. Smidt, C.H. Asbreuk, J.J. Cox, H. Chen, R.L. Johnson, J.P. Burbach, A second independent pathway for development of mesencephalic dopaminergic neurons requires Lmx1b, Nature Neurosci. 3 (2000) 337–341.
[15] Y.Q. Ding, et al., Lmx1b is essential for the development of serotonergic neurons, Nature Neurosci. 6 (2003) 933–938.
[16] R. Morello, et al., Regulation of glomerular basement membrane collagen expression by LMX1B contributes to renal disease in nail patella syndrome, Nature Genet. 27 (2001) 205–208.
[17] M.V. Clough, J.D. Hamlington, I. McIntosh, Restricted distribution of loss-of-function mutations within the LMX1B genes of nail patella syndrome patients, Hum. Mutat. 14 (1999) 459–465.
[18] E.M. Bongers, M.C. Gubler, N.V. Knoers, Nail-patella syndrome. Overview on clinical and molecular findings, Pediatr. Nephrol. 17 (2002) 703–712.
[19] J.D. Hamlington, C. Jones, I. McIntosh, Twenty-two novel LMX1B mutations identified in nail patella syndrome (NPS) patients, Hum. Mutat. 18 (2001) 458.
[20] R.K. Beals, A.L. Eckhardt, Hereditary onycho-osteodysplasia (nail-patella syndrome), J. Bone Joint Surg. 51A (1969) 505–515.
[21] M.C. Browning, N. Weidner, W.B. Lorentz Jr., Renal histopathology of the nail-patella syndrome in a two-year-old boy, Clin. Nephrol. 29 (1988) 210–213.
[22] P.R. Lichter, J.E. Richards, C.A. Downs, H.M. Stringham, M. Boehnke, F.A. Farley, Cosegregation of open-angle glaucoma and the nail-patella syndrome, Am. J. Ophthalmol. 124 (1997) 506–515.
[23] E. Sweeney, A. Fryer, R. Mountford, A. Green, I. McIntosh, Nail patella syndrome: a review of the phenotype aided by developmental biology, J. Med. Genet. 40 9 (2003) 153–162.
[24] H.C. Duba, M. Erdel, J. Loffler, J. Wirth, B. Utermann, G. Utermann, Nail patella syndrome in a cytogenetically balanced t(9;17)(q34.1;q25) carrier, Eur. J. Hum. Genet. 6 (1998) 75–79.
[25] Silahtaroglu, et al., Molecular cytogenetic detection of 9q34 breakpoints associated with nail patella syndrome, Eur. J. Hum. Genet. 7 (1999) 68–76.
[26] W.M. Eyaid, et al., Physical mapping of the nail patella syndrome interval at 9q34: ordering of STSs and ESTs, Hum. Genet. 103 (1998) 525–526.
[27] J.C. Venter, et al., The sequence of the human genome, Science 291 (2001) 1304–1351.
[28] E.S. Lander, et al., Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.
[29] M. Kozak, An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs, Nucleic Acids Res. 15 (1987) 8125–8184.
[30] M. Carey, S.T. Smale, Transcriptional Regulation in Eukaryotes: Concepts, Strategies and Techniques, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000, pp. 97–135.
[31] C.E. Haldin, S. Nijjar, K. Masse, M.W. Barnett, E.A. Jones, Isolation and growth factor inducibility of the Xenopus laevis Lmx1b gene, Int. J. Dev. Biol. 47 (2003) 253–262.
[32] R.C. Hardison, Conserved noncoding sequences are reliable guides to regulatory elements, Trends Genet. 16 (2000) 69–72.
[33] S. Levy, S. Hannenhalli, C. Workman, Enrichment of regulatory signals in conserved non-coding genomic sequence, Bioinformatics 17 (2001) 871–877.
[34] H.A. Meijer, A.A. Thomas, Control of eukaryotic protein synthesis by upstream open reading frames in the 5′-untranslated region of an mRNA, Biochem. J. 367 (2002) 1–11.

[35] A.V. Kochetov, et al., Eukaryotic mRNAs encoding abundant and scarce proteins are statistically dissimilar in many structural features, FEBS Lett. 440 (1998) 351.
[36] M. Kozak, Pushing the limits of the scanning mechanism for initiation of translation, Gene 299 (2002) 1–34.
[37] B. Han, Z. Dong, J.T. Zhang, Tight control of platelet-derived growth factor B/c-sis expression by interplay between the 5′-untranslated region sequence and the major upstream promoter, J. Biol. Chem. 278 (2003) 46983–46993.
[38] B. Han, Z. Dong, Y. Liu, Q. Chen, K. Hashimoto, J.T. Zhang, Regulation of constitutive expression of mouse PTEN by the 5′-untranslated region, Oncogene 22 (2003) 5325–5337.
[39] J.T. den Dunnen, S.E. Antonarakis, Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion, Hum. Mutat. 15 (2000) 7–12.
[40] S.D. Dreyer, et al., LMX1B transactivation and expression in nail-patella syndrome, Hum. Mol. Genet. 9 (2000) 1067–1074.
[41] P.A. Frischmeyer, H.C. Dietz, Nonsense-mediated mRNA decay in health and disease, Hum. Mol. Genet. 8 (1999) 1893–1900.
[42] M.J. Ruiz-Echevarria, S.W. Peltz, The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames, Cell 101 (2000) 741–751.
[43] A.F.A. Smit, P. Green, RepeatMasker at http://ftp.genome.washington.edu/RM/RepeatMasker.html.
[44] S. Schwartz, PipMaker-a web server for aligning two genomic DNA sequences, Genome Res. 10 (2000) 577–586, http://bio.cse.psu.edu/pipmaker.
[45] F. Larsen, G. Gundersen, R. Loez, H. Prydz, CpG islands as gene markers in the human genome, Genomics 13 (1992) 1095–1107, http://www.ebi.ac.uk/emboss/cpgplot/#.
[46] E. Wingender, et al., TRANSFAC: an integrated system for gene expression regulation, Nucleic Acids Res. 28 (2000) 316–319, www.genomatix.de.
[47] I. McIntosh, et al., Fine mapping of the nail-patella syndrome locus at 9q34, Am. J. Hum. Genet. 60 (1997) 133–142.
[48] K. Wilke, B. Duman, J. Horst, Diagnosis of haploidy and triploidy based on measurement of gene copy number by real-time PCR, Hum. Mutat. 16 (2000) 431–436.