Published in final edited form as:

Gene Expr Patterns. 2006 August ; 6(6): . doi:10.1016/j.modgep.2005.11.012.

Identification and developmental expression analysis of a novel homeobox gene closely linked to the mouse *Twirler* mutation

Han Liu¹, Wenjin Liu², Kathleen M. Maltby³, Yu Lan³, and Rulang Jiang^{1,3,*}

¹Department of Biology, University of Rochester, Rochester, NY 14627

²Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642

³Center for Oral Biology and Department of Biomedical Genetics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642

Abstract

The *Twirler* mutation arose spontaneously and causes inner ear defects in heterozygous and cleft lip and/or cleft palate in homozygous mutant mice, providing a unique animal model for investigating the molecular mechanisms of inner ear and craniofacial development. Here we report the identification of a novel homeobox gene, *Iroquois-related homeobox like-1* (*Irx11*), from the *Twirler* locus. *Irx11* encodes a TALE-family homeodomain protein with its homeodomain exhibiting the highest amino acid sequence identity (54%) to those of invertebrate Iroquois and vertebrate Irx subfamily members. The putative Irx11 protein lacks the Iro-box, a conserved motif in all known members of the Irx subfamily. Searching the databases showed that *Irx11* orthologs exist in *Xenopus*, chick, and mammals. In situ hybridization analyses of mouse embryos at various developmental stages showed that *Irx11* mRNA is highly expressed in the frontonasal process and palatal mesenchyme during primary and secondary palate development. In addition, *Irx11* mRNA is strongly expressed in mesenchyme surrounding the developing inner ear, in discrete regions of the developing mandible, in the dermamyotome during somite differentiation, and in a subset of muscular structures in late embryonic stages. The developmental expression pattern indicates that *Irx11* is a good candidate gene for the *Twirler* gene.

Keywords

cleft lip; cleft palate; craniofacial development; dermamyotome; homeobox; homeodomain; Hox; intervertebral disc; Iroquois; Irx; Irx11; mandible; palate development; secondary palate; skeletal muscle; somite; TALE; *Twirler* mutation

1. Results and Discussion

Cleft lip and cleft palate (CLP) are common birth defects affecting 1 in 500 to 2000 live births worldwide (Bender, 2000; Wilkie and Morris-Kay, 2001; Mossey and Little, 2002). Mouse models have been widely used to study the etiology of these defects since the morphogenetic and molecular processes of craniofacial development are strikingly similar in the mouse and human. One good mouse model for investigating the genetic and developmental mechanisms of orofacial clefting is the *Twirler* (*Tw*) mutant mouse (Gong et al., 2000). The *Tw* mutation arose spontaneously and causes inner ear defects in

* Author for Correspondence: Rulang Jiang, Ph.D, Center for Oral Biology, University of Rochester Medical Center, 601 Elmwood Avenue, Box 611, Rochester, NY 14642, Telephone: (585)273-1426, Fax: (585)276-0190, Rulang_Jiang@urmc.rochester.edu.

heterozygous and cleft lip and cleft palate in homozygous mutant mice (Lyon, 1958, Gong et al., 2000). Previous genetic linkage studies indicated that *Tw* is closely linked to the classical *ataxia* mutation on proximal chromosome 18 (Lyon, 1958; 1975; Lane, 1981). Recently, the *ataxia* mutation was identified as a loss of function mutation in the *Usp14* gene (Wilson, 2002), providing a genomic anchor point for the hunt of the *Tw* gene. We used an intersubspecific backcross mapping approach to narrow down the genomic region harboring the *Tw* gene. Genotyping of nearly 600 mouse progeny from (*Tw*/CAST × *C57BL/6*) crosses using multiple microsatellite markers from proximal chromosome 18 localized the *Tw* locus to a 4.3 Mb genomic region between *D18Mit19* and *D18Mit166*. According to the latest version of the NCBI mouse genome database, this region contains less than twenty known and predicted genes. A predicted gene located in the middle of this genomic region, named 9430023B20Rik (Genbank accession number **NM_177595**) because the genomic sequence matched the corresponding RIKEN full-length cDNA clone isolated from a mouse E12 cDNA library, caught our attention. We carried out RT-PCR analysis of E11.5 mouse embryonic head mRNA and found that this gene is expressed in the E11.5 mouse head tissues (data not shown). Sequencing of subcloned RT-PCR products confirmed the correctness of the cDNA sequence (data not shown). Sequence analysis showed that it encodes a protein of 353 amino acid residues with a putative homeodomain motif highly similar to those of invertebrate *Iroquois* and vertebrate *Irx* subfamily members of the three-amino-acid-loop-extension (TALE) homeodomain family (Burglin, 1997) (Fig. 1). Thus, we renamed this gene *Iroquois homeobox like-1* (*Irx11*). The *Irx11* homeodomain contains the proline-tyrosine-proline tripeptide residues conserved in the loop between helices 1 and 2 in all TALE homeodomains (Burglin, 1997) (Fig. 1B). The amino acid residue in the *Irx11* homeodomain corresponding to position 50 of the typical homeodomain is alanine, which is characteristic of the *Irx* subfamily (Burglin, 1997) (Fig. 1B). However, the first 15 residues of the *Irx11* homeodomain are highly divergent from the *Irx* subfamily (Fig. 1B). Moreover, the *Irx11* protein lacks the *Iro*-box, a conserved motif carboxyl to the homeodomain in all known members of the *Irx* subfamily (Burglin, 1997; Cavodeassi et al., 2001; Gómez-Skarmeta and Modolell, 2002). Furthermore, the *Irx* subfamily genes are clustered in both *Drosophila* and vertebrates (Cavodeassi et al., 2001; Gómez-Skarmeta and Modolell, 2002), whereas *Irx11* is located in between several non-transcription factor genes (data not shown). Since the vertebrate *Irx* genes are more similar to one another than to the *Drosophila Iroquois* complex and since *Caenorhabditis elegans* has only one *Iroquois* gene (Burglin, 1997; Bosse et al., 2000; Peters et al., 2000), it has been hypothesized that the gene duplications may have occurred independently in *Drosophila* and vertebrates (Cavodeassi et al., 2001). The high sequence similarity in the homeodomains of *Irx11* and of the *Irx* proteins suggests that *Irx11* is a novel divergent member of the vertebrate *Irx* gene subfamily. Blast searches of NCBI databases showed that *Irx11* orthologs exist in *Xenopus*, chick, and several mammalian species (data not shown). The predicted mouse *Irx11* protein shares an overall 87% amino acid sequence identity with the product of the human *C10orf48* gene (Genbank accession number **NM_173576**) (Fig. 1A). Extensive overall sequence similarity was also found between the mouse *Irx11* protein and the products of the *Xenopus* MGC 82772 full-length cDNA (Genbank accession number **BC077517**) and of the chick LOC428415 predicted locus (Genbank accession number **XM_425976**) (Fig. 1B, and data not shown).

To evaluate whether *Irx11* is a good candidate for the *Tw* gene, we analyzed its mRNA expression pattern during mouse embryogenesis by using whole mount and section in situ hybridization assays. During early mouse embryogenesis, *Irx11* mRNA expression was first detected at embryonic day (E) 9.5 (20-somite stage) specifically in the dermamyotome of differentiating somites (Fig. 2G). By E10, the dermamyotome expression domain extended caudally to the forelimb bud level (Fig. 2A, B). Section in situ hybridization detected, in addition to expression in the dermamyotome, moderate levels of *Irx11* mRNA in the

frontonasal process as well as in the neural crest-derived head mesenchyme at this stage (Fig. 2H). The yolk sac also appeared to express low levels of *Irx11* mRNA (Fig. 2H). By E10.75 to E11.5, in addition to expression in the dorsal dermamyotome, *Irx11* mRNA expression was also detected in the ventral domain of dermamyotome in the trunk region between the fore- and hind-limb buds (Fig. 2C, J). High levels of *Irx11* mRNA were also detected in the neural crest-derived head and frontonasal mesenchyme, as well as in the mesenchyme surrounding the otic vesicle (Fig. 2I, J). In addition, high levels of *Irx11* mRNA were detected in the myogenic precursor cells migrating into the limb buds (Fig. 2C, D, K). From E11.5 to E11.75, a strong domain of *Irx11* mRNA expression was detected in the forehead region ventral to the telencephalic vesicles (Fig. 2C, D, E, I). At E12.5, *Irx11* mRNA expression in the limbs was concentrated in the digital mesenchyme (Fig. 2F).

Interestingly, strong *Irx11* mRNA expression was found in the secondary palate throughout its development. At E12.5, the palatal shelves outgrow bilaterally from the oral sides of the maxillary processes. *Irx11* mRNA expression was highly activated throughout the mesenchyme of the outgrowing palate shelves (Fig. 3A). Strong *Irx11* mRNA expression persists in the palatal mesenchyme during palatal growth and fusion from E12.5 to E14.5 (Fig. 3, A–C). *Irx11* mRNA is also highly expressed in the neural crest-derived head mesenchyme surrounding the brain as well as in the preossification mesenchymal cells surrounding the Meckel's cartilage in the mandible at these stages (Fig. 3, A–C). In addition, a subset of craniofacial and body skeletal muscle cells also express high levels of *Irx11* mRNA (Fig. 3, A–E). *Irx11* mRNA expression appears highly regulated during muscle development as highly expressing muscle cells were often found adjacent to non-expressing muscle cells, particularly in the developing tongue musculature (Fig. 3, B–D). Some muscle bundles only express *Irx11* mRNA in terminal regions (Fig. 3D). In addition, high levels of *Irx11* mRNA are also detected in a segmented pattern in the developing intervertebral disc but not in the cartilage primordia of the vertebrae (Fig. 3E). These data indicate that *Irx11* mRNA expression is highly regulated during mouse embryogenesis.

The TALE family of homeodomain proteins often interacts with other transcription factors, particularly the Hox proteins, to regulate specific developmental processes (Bertolino et al., 1995; Rieckhof et al., 1997; Mann and Affolter, 1998; Wang et al., 2001). Both *Drosophila* and vertebrate *Irx* family genes play essential roles in embryonic development and patterning (reviewed by Cavodeassi et al., 2001; Gómez-Skarmeta and Modolell, 2002). The dynamic expression pattern of *Irx11* mRNA during craniofacial and palate development, together with its chromosomal location suggest that it is a good candidate for the *Tw* gene. The abundant and highly regulated *Irx11* mRNA expression during dermamyotome and subsequent skeletal muscle development also warrants further investigation of the role of this putative transcription factor in muscle development.

2. Experimental Procedures

2.1. RT-PCR

Total RNAs were extracted from E11.5 mouse embryos. RT-PCR amplification of *Irx11* cDNA was done with the following primer pairs: RT1F: 5'-GACTCCGAGGCTCTGCCGAA-3', RT1R: 5'-CAGGAGTCGCCATCGCTGCTCA-3', RT2F: 5'-CATGAACACCATCGTCTTCAAC-3', and RT2R: 5'-CATGACATGTCTCAAGGAGTC-3'. These two pairs of primers amplify partially overlapping cDNA fragments that together contain the entire protein-coding region.

2.2. in situ hybridization

Embryos of different stages were dissected and fixed in 4% paraformaldehyde in PBS overnight at 4°C. A subcloned 620 bp RT-PCR product corresponding to the 3' half of the coding region and an expressed sequence tag matching the 3' untranslated region of the *Irx11* gene were used as templates for cRNA probe synthesis. Both probes detected identical *Irx11* mRNA expression patterns during mouse embryogenesis in both whole mount and section in situ hybridization assays. For whole mount in situ hybridization, embryos were hybridized with digoxigenin labeled cRNA probes as describe previously (Jiang et al., 1998). For section in situ hybridization, paraffin sections were processed and probed with ³³P-labeled cRNA probes and data recorded as described by McGrath et al. (1999).

Acknowledgments

This work was supported by NIH/NIDCR grants R01 DE013681 and R01 DE015207 to RJ.

References

- Bender PL. Genetics of cleft lip and palate. *J. Pediat. Nurs.* 2000; 15:242–249.
- Bertolino E, Reimund B, Wildt-Perinic D, Clerc RG. A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. *J. Biol. Chem.* 1995; 270:31178–31188. [PubMed: 8537382]
- Bosse A, Stoykova A, Nieselt-Struwe K, Chowdhury K, Copeland NG, Jenkins NA, Gruss P. Identification of a novel mouse Iroquois homeobox gene, *Irx5*, and chromosomal localization of all members of the mouse Iroquois gene family. *Dev. Dyn.* 2000; 218:160–174. [PubMed: 10822268]
- Burglin TR. Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic Acid Res.* 1997; 25:4173–4180. [PubMed: 9336443]
- Cavodeassi F, Modolell J, Gómez-Skarmeta JL. The Iroquois family of genes: from body building to neural patterning. *Development.* 2001; 128:2847–2855. [PubMed: 11532909]
- Gómez-Skarmeta JL, Modolell J. *Iroquois* genes: genomic organization and function in vertebrate neural development. *Curr. Opin. Genet. Dev.* 2002; 12:403–408. [PubMed: 12100884]
- Gong S-G, White NJ, Sakasegawa AY. The *Twirler* mouse, a model for the study of cleft lip and palate. *Arch. Oral Biol.* 2000; 45:87–94. [PubMed: 10669096]
- Jiang R, Lan Y, Norton CR, Sundberg JP, Gridley T. The *Slug* gene is not essential for mesoderm or neural crest development in mice. *Dev. Biol.* 1998; 198:277–285. [PubMed: 9659933]
- Lane PW, Searle AG, Beechey CV, Eicher EM. Chromosome 18 of the house mouse. *J. Hered.* 1981; 72:409–412. [PubMed: 6949984]
- Lyon MF. *Twirler*: a mutant affecting the inner ear of the house mouse. *J. Embry. Exp. Morph.* 1958; 6:105–116.
- Lyon MF. Recombination between *Tw* and *ax*. *Mouse News Lett.* 1975; 52:36.
- Mann RS, Affolter M. Hox proteins meet more partners. *Curr. Opin. Genet. Dev.* 1998; 8:423–429. [PubMed: 9729718]
- McGrath KE, Koniski AD, Maltby KM, McGann JK, Palis J. Embryonic expression and function of the chemokine SDF-1 and its receptor, CXCR4. *Dev. Biol.* 1999; 213:442–456. [PubMed: 10479460]
- Mossey, PA.; Little, J. Epidemiology of oral clefts: an international perspective. In: Wyszynski, DF., editor. *Cleft Lip & Palate, From Origin to Treatment*. Oxford University Press; 2002. p. 127–158.
- Peters T, Dildrop R, Ausmeier K, Ruther U. Organization of mouse *iroquois* homeobox genes in two clusters suggests a conserved regulation and function in vertebrate development. *Genome Res.* 2000; 10:1453–1462. [PubMed: 11042145]
- Rieckhof GE, Casares F, Ryoo HD, Abu-Shaar M, Mann RS. Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. *Cell.* 1997; 91:171–183. [PubMed: 9346235]

- Wang GF, Nikovits WJ, Bao ZZ, Stockdale FE. Irx4 forms an inhibitory complex with the vitamin D and retinoic X receptors to regulate cardiac chamber-specific slow MyHC3 expression. *J. Biol. Chem.* 2001; 276:28835–28841. [PubMed: 11382777]
- Wilkie AO, Morris-Kay GM. Genetics of craniofacial development and malformation. *Nat. Genet. Rev.* 2001; 2:458–468.
- Wilson SM, Bhattacharyya B, Rachel RA, Coppola V, Tessarollo L, Householder DB, Fletcher CF, Miller RJ, Copeland NG, Jenkins NA. Synaptic defects in *ataxia* mice result from a mutation in *Usp14*, encoding a ubiquitin-specific protease. *Nat. Genet.* 2002; 32:420–425. [PubMed: 12368914]

A

Irx11	MNTIVFNKLGGAVLFEDRGTPDRERGSRTFSGFLDNPHGTGPEVGIPDGPP	50
C10ORF48	..T.....SSA....DGGASE....GRPYS.VLDS.HAR.E..IP..PP	50
MGC82772	..A.....SGQ....ESAN-D....VRPYL.VMES.N--.D..LQ..VS	47
Irx11	<u>LKDNLSLRHRTGARQNGGKVRHKRQALQDMARPLKQWLYKHRDNPYPTK</u>	100
C10ORF48	L...LG.....T.ARQNG.....	100
MGC82772	I...AS.....P.SSRGS.....	97
Irx11	<u>TEKILLALGSQMTLVQVSNWFANARRRLKNTVRQPDLSWALRIKLYNKYV</u>	150
C10ORF48		150
MGC82772		147
Irx11	QGNAERLSVSSDGDSCSEDGENPPRNHMNEEGYSTPAHHTVIKGESSAIK	200
C10ORF48	D-.....NP..T.M..GG.NT.V..P..KS.N.V..	199
MGC82772	E-.....HV..S.N..EA.DK.M..T..TK.S.V..	196
Irx11	AGGRPESRAAEDYVSPPKYKSSLLNRYLNDSLRHVMATSTAMMGKTRRRN	250
C10ORF48	A.V.P..R.S....A.....N.....TNNT.MG...Q..	249
MGC82772	S.V.A..S.S....S.....H.....SNAA.RE...Q..	246
Irx11	HSGSFSSNEFEELVSPSSSETEGTFVYRTDTPDIGSTKGDSAANRRGPS	300
C10ORF48	N.....TLENGSN.GE..A.RK.P.	299
MGC82772	N.....MLESAAN.CQ..V.KE.A.	296
Irx11	KDDTYWKEINAAMALTNLAQKGDEVQGTTSCTIIQKSSHIAEVKTVKPLPVQRF	353
C10ORF48	..D...K.....KL.....A.....VP.V.QF	352
MGC82772	..E...R.....KL.....S.....VA.L.KY	349

B

	helix 1	helix 2	helix 3
Dm ara	LAARRKNATRESTATLKAWLNEHKKNPYPYTKGEKIMLAIITKMTLTQVSTWTFANARRRLKKEN		
Ce 1k114	DGIKKR.....A..P..D..HS.R.....S.AD.V...VG.G.....		
Irx1	DPG.P.....ST.....R.....		
Irx2	DP.Y.....DA..T.....R.....		
Irx3	DPS.P.....ST.....R.....		
Irx4	SGT.....T.ST.....Q..R.....		
Irx5	DP.Y.....DA..T.....R.....		
Irx6	S.G.....SA.....H..R.....		
Irx11	KVRHKRQ.LQDMARP..Q..YK.RD.....T...L..LGSQ...V...N.....NTV		
Dm Antp	RKRG.QTY..YQ.LE.EKEFHFN--R.L.RRRR.EI.HALCL.ER.IKI..Q.R.MKWKKEN		
	10 20 30 40 50 60		

Fig. 1.

Irx11 encodes a new TALE family homeodomain protein in vertebrates. (A) Alignment of the mouse *Irx11* protein sequence with the products of the human *C10orf48* gene and the *Xenopus* MGC82772 full-length cDNA. Numbers to the right represent numbers of the last residue in that line. Dots represent identities to the mouse *Irx11* protein and dashes represent gaps. The homeodomain sequences (underlined) are identical in all three species. The overall high amino acid sequence identity indicates that these are orthologs in different species. (B) Alignment of the homeodomain sequence of the mouse *Irx11* protein with those of *Drosophila* (Dm ara), *Caenorhabditis elegans* (Ce 1k114), and mouse *Irx* proteins. The

residues of the three helices of the homeodomain are overlined and marked on top. For comparison, the typical homeodomain of *Drosophila* Antennapedia protein (Dm Antp) is shown at the bottom. Dots represent identities to the *Drosophila* ara sequence and dashes represent gaps. Numbers in the bottom denote amino acid positions in the typical homeodomain. The TALE family of homeodomain proteins has additional three amino acids between the first and second helices, compared with the typical homeodomain of 60 amino acid residues. The proline-tyrosine-proline tripeptide residues (PYP, underlined) are conserved in all TALE homeodomains (Burglin, 1997).

Fig. 2.

Expression pattern of *Irx11* mRNA in early mouse embryos. mRNA signals detected by whole mount in situ hybridization appear purple (A–F) and those detected by section in situ hybridization are shown in red on blue-counterstained sections (G–K). (A) *Irx11* mRNA was detected at E10.0 in the dermamyotome of rostral somites (arrowheads). (B) Dorsal view of an E10.0 embryo showing strong *Irx11* mRNA expression in the dermamyotome of somites around the forelimb level. (C) At E11.5, *Irx11* mRNA was expressed in the ventral domain of dermamyotome caudal to the forelimb bud (arrow) and in all dorsal dermamyotome. Strong *Irx11* expression was also detected at the base of the forelimb buds. (D) At E11.75, strong *Irx11* mRNA expression was detected in discrete domains in the fore- and hindlimb buds. (E) High magnification lateral view of the forehead region of an E11.75 embryo showing strong expression of *Irx11* mRNA ventral to the telencephalic vesicle (arrow). (F) An E12.5 forelimb bud showing *Irx11* mRNA expression in the digital mesenchyme. (G) In situ hybridization of sagittal sections of E9.5 embryos showed specific *Irx11* mRNA expression in the newly formed dermamyotome of rostral somites (arrow). (H) Sagittal section of an E10.5 embryo showing *Irx11* mRNA expression in the frontonasal process and in the yolk sac, in addition to the dermomyotome. (I, J) Sagittal sections of E11.5 embryos showing *Irx11* mRNA expression in the medial nasal process (arrow in I), head mesenchyme including mesenchyme surrounding the developing otocyst, as well as in the ventral domain of the dermamyotome (arrowhead in J). (K) Frontal section of an E11.5 embryo showing *Irx11* mRNA expression in the medial and lateral nasal processes, the dermamyotome and the myogenic mesenchyme of the forelimb buds. dm, dermamyotome; fl, forelimb bud; fnp, frontonasal process; hl, hindlimb; hm, head mesenchyme; lnp, lateral nasal process; mnp, medial nasal process; ot, otic vesicle; tv, telencephalic vesicle; ys, yolk sac.

Fig. 3.

Dynamic expression of *Irx11* mRNA during craniofacial and skeletal muscle development. mRNA signals are shown in red on blue-counterstained sections. (A) Frontal section of an E12.5 embryo showing strong *Irx11* mRNA expression in the neural crest-derived head mesenchyme surrounding the brain (arrowheads), in the secondary palate, in the mesenchymal precursor cells of the mandibular bone surrounding the Meckel's cartilages, and in a subset of skeletal muscle cells (arrows). (B) Frontal section of an E13.5 embryo showing strong *Irx11* mRNA expression in the mesenchyme of the palate shelves and mandible, periocular muscles (arrows), in a subset of cells in the muscle of the tongue (arrowheads). (C, D) Frontal (C) and sagittal (D) sections of E14.5 embryos showing that strong *Irx11* mRNA expression persists in the palatal shelves after elevation and fusion, as well as in the periocular muscles (arrow in C) and in the mesenchyme surrounding the mandibular ossification centers. (E) Sagittal section of an E14.5 embryo showing highly restricted expression of *Irx11* mRNA in the intervertebral discs (arrowheads) and in the tail muscles (arrows). The asterisk in E marks an air bubble with autofluorescence. No *Irx11* mRNA expression in the liver was detected by in situ hybridization. e, eye; gt, genital tubercle; li, liver; lu, lung; mb, mandible; mc, Meckel's cartilage; mi, mandibular incisor; nc, nasal cavity; oc, ossification center in the mandible; p, palate; po, preossification mesenchyme in the mandible; pp, primary palate; t, tongue; tl, tail.