Title

De novo, deleterious sequence variants that alter the transcriptional activity of the homeoprotein PBX1 are associated with intellectual disability and pleiotropic developmental defects

Journal

Human Molecular Genetics, 26(24)

ISSN

0964-6906

Authors

Slavotinek, Anne
Risolino, Maurizio
Losa, Marta
et al.

Publication Date

2017-12-15

DOI

10.1093/hmg/ddx363

Peer reviewed

Human Molecular Genetics, 2017, Vol. 26, No. 24 4849–4860

doi: 10.1093/hmg/ddx363
Advance Access Publication Date: 22 September 2017
Original Article

ORIGINAL ARTICLE

De novo, deleterious sequence variants that alter the transcriptional activity of the homeoprotein PBX1 are associated with intellectual disability and pleiotropic developmental defects

Anne Slavotinek[1,2,*,†], Maurizio Risolino[2,3,†], Marta Losa[2,3], Megan T. Cho[4], Kristin G. Monaghan[4], Dina Schneidman-Duhovny[5,6], Sarah Parisotto[7], Johanna C. Herkert[8], Alexander P.A. Stegmann[9,10], Kathryn Miller[11], Natasha Shur[11], Jacqueline Chui[12], Eric Muller[12], Suzanne DeBrosse[13], Justin O. Szot[14,15], Gavin Chapman[14,15], Nicholas S. Pachter[16,17], David S. Winlaw[18,19], Bryce A. Mendelsohn[1,2], Joline Dalton[20], Kyriakie Sarafoglou[21], Peter I. Karachunski[22], Jane M. Lewis[23], Helio Pedro[7], Sally L. Dunwoodie[14,15], Licia Selleri[2,3,‡] and Joseph Shieh[1,2,‡]

[1]Division of Medical Genetics, Department of Pediatrics, [2]Institute of Human Genetics and [3]Program in Craniofacial Biology, Departments of Orofacial Sciences and Anatomy, University of California San Francisco, San Francisco, CA, USA, [4]GeneDx, Gaithersburg, MD, USA, [5]School of Computer Science and Engineering and [6]Department of Biochemistry, Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel, [7]Division of Genetics, Department of Pediatrics, Hackensack University Medical Center, Hackensack, NJ, USA, [8]Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, [9]Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands, [10]Department of Genetics, Radboud University Medical Center (RUMC), Nijmegen, The Netherlands, [11]Department of Pediatrics, Albany Medical Center, Albany, NY, USA, [12]Clinical Genetics, Stanford Children's Health at CPMC, San Francisco, CA, USA, [13]Center for Human Genetics, University Hospitals Cleveland Medical Center, Cleveland, OH, USA, [14]Developmental and Stem Cell Biology Division, Victor Chang Cardiac Research Institute, Sydney, NSW, Australia, [15]University of New South Wales, Sydney, NSW, Australia, [16]Genetic Services of Western Australia, King Edward Memorial Hospital, Perth, WA, Australia, [17]School of Paediatrics and Child Health, University of Western Australia, Perth, WA, Australia, [18]University of Sydney, Medical School, Sydney, NSW, Australia, [19]Heart Centre for Children, Children's Hospital at Westmead, Sydney, NSW, Australia, [20]Paul and Shelia Wellstone Muscular Dystrophy Center, University of Minnesota, Minneapolis, MN, USA, [21]Department of Pediatrics, University of Minnesota Masonic Children's Hospital, Minneapolis, MN, USA, [22]Department of Neurology, University of Minnesota, Minneapolis,

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
[‡]The authors wish it to be known that, in their opinion, the last two authors should be regarded as joint Senior Authors.
Received: June 13, 2017. **Revised:** August 25, 2017. **Accepted:** September 15, 2017

MN, USA and [23]Department of Urology, University of Minnesota Masonic Children's Hospital, Minneapolis, MN, USA

*To whom correspondence should be addressed at: Division of Medical Genetics, Department of Pediatrics, University of California San Francisco, RH384C, 1550 4th St, San Francisco, CA 94143-2711, USA. Tel: +1 4155141783; Fax: +1 4154769305; Email: anne.slavotinek@ucsf.edu

Abstract

We present eight patients with *de novo*, deleterious sequence variants in the *PBX1* gene. *PBX1* encodes a three amino acid loop extension (TALE) homeodomain transcription factor that forms multimeric complexes with TALE and HOX proteins to regulate target gene transcription during development. As previously reported, *Pbx1* homozygous mutant mice ($Pbx1^{-/-}$) develop malformations and hypoplasia or aplasia of multiple organs, including the craniofacial skeleton, ear, branchial arches, heart, lungs, diaphragm, gut, kidneys, and gonads. Clinical findings similar to those in *Pbx* mutant mice were observed in all patients with varying expressivity and severity, including external ear anomalies, abnormal branchial arch derivatives, heart malformations, diaphragmatic hernia, renal hypoplasia and ambiguous genitalia. All patients but one had developmental delays. Previously reported patients with congenital anomalies affecting the kidney and urinary tract exhibited deletions and loss of function variants in *PBX1*. The sequence variants in our cases included missense substitutions adjacent to the PBX1 homeodomain (p.Arg184Pro, p.Met224Lys, and p.Arg227Pro) or within the homeodomain (p.Arg234Pro, and p.Arg235Gln), whereas p.Ser262Gln*fs**2, and p.Arg288* yielded truncated PBX1 proteins. Functional studies on five *PBX1* sequence variants revealed perturbation of intrinsic, PBX-dependent transactivation ability and altered nuclear translocation, suggesting abnormal interactions between mutant PBX1 proteins and wild-type TALE or HOX cofactors. It is likely that the mutations directly affect the transcription of PBX1 target genes to impact embryonic development. We conclude that deleterious sequence variants in *PBX1* cause intellectual disability and pleiotropic malformations resembling those in *Pbx1* mutant mice, arguing for strong conservation of gene function between these two species.

Introduction

PBX1, the gene encoding the Pre-B-cell leukemia homeobox (PBX) transcription factor (1), is the mammalian homologue of the extradenticle (*exd*) gene in *Drosophila melanogaster* (2). *PBX1* shares 71% identity to *exd* and encodes a three amino acid loop extension (TALE) homeodomain transcription factor (2,3). PBX1 and its related family members, PBX2–4 (3), dimerize with other TALE class homeodomain proteins from the Myeloid Ecotropic Integration Site (MEIS) and PBX regulating protein (PREP) families through a PBC domain to form nuclear complexes that enhance the binding specificity of HOX proteins to DNA, thus regulating target genes that control segment identity and organ patterning during embryogenesis (3–6). We previously reported that *Pbx1* is widely transcribed throughout the developing murine embryo and *Pbx1*-deficient mice that are homozygous for a null allele ($Pbx1^{-/-}$) develop pleiotropic developmental defects. These comprise malformations, severe hypoplasia, or aplasia of multiple organs, including craniofacial skeleton, ear pinnae, branchial arch-derived structures, heart, lungs, diaphragm, liver, stomach, gut, pancreas, kidneys and gonads (4,7–9). Heterozygous mice are viable and fertile, but are smaller in size than wildtype (WT) mice, with a 30% reduction in weight prior to eight weeks of age (4). $Pbx1^{-/-}$ embryos also exhibit abnormal derivatives of the second branchial arch, with elongated and deformed hyoid bones, dysplastic styloid processes and lack of stapes and oval windows (4).

Recently, eight patients were reported with congenital anomalies affecting the kidney and urinary tract (CAKUT), and interstitial deletions of varying sizes that all encompassed *PBX1* at chromosome 1q23.3q24.1 (10). All patients manifested renal hypoplasia, with or without renal dysplasia that presented as a loss of corticomedullary differentiation and renal cysts (10) As expression of the *PBX1* gene was detected in the fetal and adult kidneys in humans, it was concluded that *PBX1* was a new causative gene for CAKUT (10). Extrarenal manifestations in those patients were described as developmental delays and autism spectrum disorder, outer ear anomalies, hearing impairment, and cardiac defects. However, the report could not unequivocally associate the observed abnormalities with haploinsufficiency of *PBX1* due to the inclusion of additional genes in the large interstitial deletions (10). A recent study also examined 204 patients with CAKUT syndrome for variants in 330 genes (11) The authors identified three, de novo, heterozygous *PBX1* variants and two deletions for a detection frequency of 2.5%. Out of those cases, a female with hearing loss and renal hypoplasia had a frameshift variant in *PBX1* (p.Asn143Thrfs*37), an 11-year-old female with renal cystic dysplasia, developmental and growth delays and a long and narrow face, had a nonsense variant (p.Arg184*) and a fetus with oligohydramnios and renal hypoplasia had a splice site variant (c.511–2 A > G). In addition, two patients with *PBX1* deletions manifested renal dysplasia with hearing impairment and renal agenesis/dysplasia, developmental differences, microcephaly together with facial anomalies, respectively (11).

We now present eight patients who are heterozygous for *de novo*, deleterious sequence variants in *PBX1* and establish pathogenic mutations of this gene as causative in the etiology of a pleiotropic intellectual disability syndrome with ear, branchial arch, cardiac, pulmonary, diaphragmatic, renal, and gonadal malformations. Importantly, functional assays testing five of the sequence variants demonstrated overall abnormal transactivation capabilities for the mutant PBX1 proteins.

Results

Clinical features of the eight patients with *de novo*, *PBX1* sequence variants

The clinical findings for the eight patients, ordered according to the position of their *PBX1* sequence variants (patient 1–7; while

Table 1. Clinical findings in eight patients with *de novo* sequence variants in *PBX1*

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8
Aural findings								
Dysplastic ears	−	−	−	−	+	+	+	−
EAC stenosis[a]	−	−	−	−	−	+	+	−
Microtia	−	−	−	−	+	+	−	−
Hearing loss	−		−	−	−	+; unilat.[b]	−	−
Branchial findings								
Cartilaginous rests	−	−	+	−	−	−	−	−
Cardiac/Pulmonary								
Ebstein anomaly	−	−				−	+	−
PDA[c]	−	+[d]		+	+	−		
Tetralogy of Fallot[e]	+				+; unilat.	−		−
Eventration diaphragm	−	+; unilat.				−		−
Lung hypoplasia	−	+/−[f]			+; unilat.	−		−
Renal								
Renal hypoplasia	−[g]	−	−	−	−	+	+; UTIs[h]	−
Pyelocaliectasis	−[g]	−	−	+; unilat.	−	−	−	−
Increased echogenicity	−[g]	−	−	+	−	−	−	−
Dilated fetal ureters	−[g]	−	−	−	+	−	−	−
Craniofacial								
Dysmorphism	−	+		−	+[i]	+	+[j]	−
Micrognathia	–	+			+	+		−
Wide neck/nuchal fold	−	−		+	+			−
Other								
Brachydactyly[k]	−	−			+	−	+	−
Cryptorchidism	−	+[l]	+	NA	+	NA	NA	+[m]
PBX1 variant	p.Arg184Pro	p.Met224Lys	p.Arg227Pro	p.Arg234Pro	p.Arg235Gln	p.Ser262Glnfs*2	p.Arg288*	p.Arg235Gln

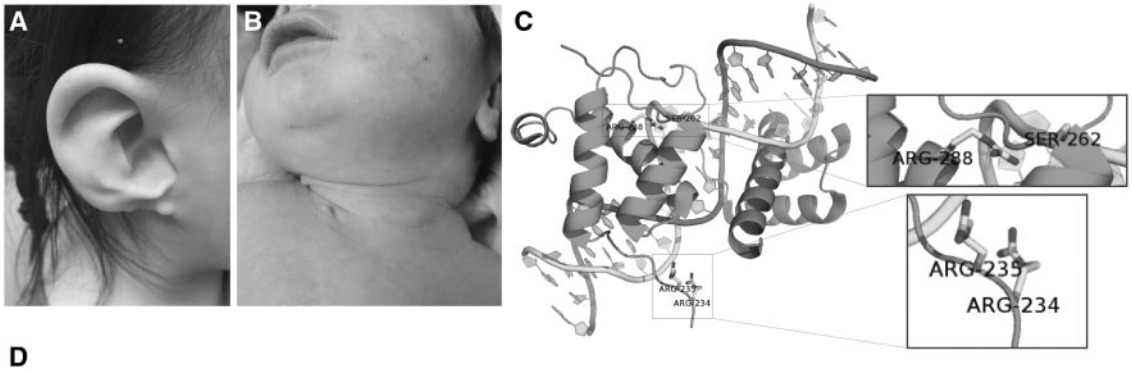
EAC stenosis[a] = external auditory canal stenosis; unilat.[b] = unilateral; PDA[c] = patent ductus arteriosus; +[d] = PDA and secundum ASD; Tetralogy of Fallot[e] = Tetralogy of Fallot with absent pulmonary valve, severe pulmonary stenosis, ventricular septal defect; +/−[f] = probable; UTI[h] = urinary tract infections; −[g] = no renal ultrasound scan was performed in Patient 1; +[i] = blue sclera, ear microtia with prominent lobules, prominent nuchal folds, and micrognathia; +[j] = Low anterior/posterior hairlines; coarse hair, deep-set eyes, short philtrum, an exaggerated Cupid's bow, bifid uvula, prominent mandible, brachydactyly type D, pes planovalgus and genu valgum; absent axillary hair; Brachydactyly[k] = brachydactyly and/or nail hypoplasia; +[l] = microphallus, rugated tissue superior and lateral to phallus, intra-abdominal testicular tissue and persistence of the Müllerian ducts; +[m] = undervirilized external genitalia with a micropenis (2.2 x 0.8 cm), labial folds without palpable gonads and two perineal openings comprising a vaginal introitus and urethra meatus. Pelvic ultrasound showed uterus didelphys.

patient 8 carries an identical mutation to patient 5), have been summarized in Table 1. Further clinical descriptions have been provided as Supplementary Files.

The patients ranged from 3 months to 27 years of age. During the pregnancies, oligohydramnios, increased nuchal folds, and unilateral diaphragmatic eventration were each present in two patients. Developmental delays were present in all but two children (patient 1 and patient 8) who were 3 and 12 months of age respectively. In those with delays, the age of independent walking ranged from 18–36 months and single words were first uttered at around 2 years of age. Only one patient was severely delayed and carried the additional diagnoses of autism and obsessive/compulsive disorder. Hypotonia was present in four patients. All patients except one had unremarkable growth parameters at birth, but postnatal short stature and failure to thrive developed in two, with weight and height measurements less than the fifth centile. Microtia and/or dysplastic ear helices were noted in three patients (Fig. 1A), with absent tragus and prominent lobes and attached earlobes. One patient had unilateral, mild to moderate conductive hearing loss with bilateral stenosis of the external auditory canals. Craniofacial dysmorphic findings, apart from ear anomalies and micrognathia, did not delineate a recognizable facial phenotype. One male had cartilaginous neck rests, a rare finding caused by aberrant development of the branchial arches (Fig. 1B) (12). Cardiac findings comprised Ebstein anomaly, Tetralogy of Fallot and patent ductus arteriosus that required surgery in two patients. One male had right lung hypoplasia with right diaphragm eventration/paralysis and another patient with a male karyotype and atypical sexual development also had a unilateral diaphragmatic eventration. Renal anomalies comprised hypoplasia, left pyelocaliectasis and dilated fetal ureters in seven patients; patient 1 did not undergo a renal ultrasound scan. The oldest patient also presented with recurrent urinary tract infections. Four males with karyotype 46, XY had cryptorchidism and two males had atypical sexual development: one with intra-abdominal testes and retained Müllerian structures and the other with micropenis and undervirilized male external genitalia, a vaginal introitus and uterus didelphys.

Identification of *PBX1* sequence variants using whole exome sequencing

For Patient 1, whole exome sequencing (WES) identified a *de novo*, heterozygous variant in *PBX1*: c.551 G > C, predicting (p.Arg184Pro) that was the only potential disease-causing candidate using a combination of *in silico* approaches. This sequence variant, which was confirmed by Sanger sequencing, affects the PBC-B domain that is required for interaction with MEINOX proteins (3). The variant was identified in a single trio from a cohort comprising a total of 120 individuals (65 females and 55 males) from 30 families affected by different types of sporadic



D

Species	Gene	Alignments for p.(Arg227Pro), p.(Arg234Pro) and p.(Arg235Gln)
H.sapiens - wildtype	ENST00000394412	K Q S T C E A V M I L **R** S R F L D A **R** **R** K R R
H.sapiens - mutated	Mutated	K Q S T C E A V M I L **P** S R F L D A **P** **Q** K R R
P.troglodytes	ENSPTRG00000001611	K Q S T C E A V M I L **R** S R F L D A **R** **R** K R R
M.mulatta	ENSMMUG00000019108	K Q S T C E A V M I L **R** S R F L D A **R** **R** K R R
F.catus	ENSMMUG00000019108	K Q S T C E A V M I L **R** S R F L D A **R** **R** K R R
M.musculus	ENSMUSG00000052534	K Q S T C E A V M I L **R** S R F L D A **R** **R** K R R
G.gallus	ENSGALG00000003429	K Q S T C E A V M I L **R** S R F L D A **R** **R** K R R
T.rubripes	ENSTRUG00000007203	K Q S T C E A V M I L **R** S R F L D A **R** **R** K R R
D.rerio	ENSDARG00000089550	K Q S T C E A V M I L **R** S R F L D A **R** **R** K R R
D.melanogaster	FBgn0000611	K Q S T C E A V M I L **R** S R F L D A **R** **R** K
C.elegans	F31E3.1	K Q S T C E A V M I L **R** S R F L D A **R** **R** K R R
X.tropicalis	ENSXETG00000010464	K Q S T C E A V M I L **R** S R F L D A **R** **R** K R R

Figure 1. Abnormalities in ear and branchial arch-derived structures in patients with heterozygous, *de novo*, *PBX1* sequence variants. (**A**) Right ear from Patient 6, showing a dysplastic helix with absence of the tragus and attached earlobe. (**B**) Cartilaginous neck rests seen in Patient 3. (**C**) Mapping of mutations onto the homeodomain structure. (**D**) Protein sequence analysis showing sites of altered residues in PBX1 in relation to an interactive DNA strand. (wildtype; top row)

or inherited congenital heart disease (CHD) for a total of 57 affected individuals and 63 unaffected controls.

For Patient 2, WES revealed a *de novo*, heterozygous variant in *PBX1*: c.671 T > A, predicting (p.Met224Lys). The mean depth of coverage was 150x with a quality threshold of 97.8%. This variant is a non-conservative amino-acid substitution, occurring at a position that is conserved across species. In view of the anomalies in the ear, renal, and branchial arch derivatives in these patients, we examined the genomic regions corresponding to the genes involved in branchio-oto-renal (BOR) syndrome (13) and to *HNF1B* that is mutated in CAKUT. No mutations were detected in these genes or any other CAKUT gene, despite high sequence coverage (*EYA1* 100%, *SIX5* 99.1%, *SIX1* 100%, and *HNF1B* 100%).

Patient 3 had a *de novo*, heterozygous sequence variant, c.680 G > C, predicting (p.Arg227Pro) in *PBX1*. The variant GATK QUAL score was 673 (>500 considered high quality) and coverage was >80x. The Arg227 residue is proximal to the predicted DNA-binding homeodomain of PBX1 (amino acids 233–295; Fig. 1C). Arg227 is in a region of high sequence identity among PBX family members, at a residue with complete conservation across species (Fig. 1D). No mutations were detected in the genes involved in BOR syndrome and *HNF1B*, despite high sequence coverage (*EYA1* 100%, *SIX5* 99.1%, *SIX1* 100%, and *HNF1B* 100%).

In Patient 4, WES identified a *de novo*, heterozygous variant, c.701 G > C, predicting (p.Arg234Pro), in the homeodomain of PBX1 that exhibits complete conservation across species (Fig. 1C and D). No mutations were detected in three genes associated with BOR syndrome and *HNF1B*, all of which had coverage of 100% (*EYA1*, *SIX5*, *SIX1*, and *HNF1B*).

For Patient 5, WES demonstrated a *de novo*, heterozygous variant, c.704 G > A, predicting (p.Arg235Gln), in *PBX1*. This variant is in a genomic region of high conservation (Fig. 1C and D). Despite good coverage, no mutations were identified in the BOR genes (*EYA1* 100%, *SIX5* 91.5%, *SIX1* 100% and *HNF1B* 100%).

In Patient 6, mean depth of coverage was 109x and quality threshold (percentage of captured region covered by at least 10 sequence reads) was 96.1%. The results showed heterozygosity for a *de novo*, sequence variant in *PBX1*, c.783dupC, predicting (p.Ser262Glnfs*2). This frameshift mutation is expected to lead to nonsense-mediated decay and to result in premature protein truncation. Coverage of the genes involved in BOR syndrome was high (*EYA1* 100%, *SIX5* 99.8%, *SIX1* 100% and *HNF1B* 100%) and no mutations were detected.

In Patient 7, clinical WES was performed with a mean depth of coverage of 95x and a heterozygous, *de novo*, *PBX1* (p.Arg288*) sequence variant with a GATK score of 814 and a read depth of 84x was detected. The variant was confirmed by Sanger sequencing and results in a premature protein truncation.

In Patient 8, whole exome sequencing showed a de novo, heterozygous variant in *PBX1*, c.704 G > A, predicting p.Arg235Gln, that was identical to the variant identified in Patient 5. Patient 8 was also heterozygous for a variant in the steroidogenic factor-1 *(NR5A1)* gene, c.1379 A > G, predicting (p.Gln460Arg). However, six individuals were reported in the Exome Aggregation Consortium (ExAC) database to have heterozygous missense variants at the Gln460 residue in *NR5A1*: five with (p.Gln460Pro) and one with (p.Gln460Leu). In addition, the above substitutions occur at a position that is not conserved across species. Lastly, PolyPhen-2 predicted that the *NR5A1* variant identified in Patient 8 is likely to be benign (PolyPhen-2 score of 0). Given all of the above, the *NR5A1* variant was considered of unknown significance and the *PBX1* variant deemed as the causative mutation in Patient 8.

Table 2. Nucleotide variants in *PBX1* (NM_002585.3)

Chromosome	Nucleotide	Protein	Exon	ExAC[a] /dbSNP[b]	SIFT[c]	Poly-Phen-2[d]	Mutation Taster	CADD[e] score	GERP[f] score
chr1(GRCh37): g.164768976G>C	c.551G>C	(p.Arg184Pro)	4	–/–[g]	Damaging	Prob damaging	DC[h]	34	5.76[i]
chr1(GRCh37): g.164769096T>A	c.671T>A	(p.Met224Lys)	4	–/–	Damaging	Poss damaging	DC	31	5.64
chr1(GRCh37): g.164769105G>C	c.680G>C	(p.Arg227Pro)	4	–/–	Damaging	Prob damaging	DC	34	5.64
chr1(GRCh37): g.164769126G>C	c.701G>C	(p.Arg234Pro)	4–5[j]	–/–	Damaging	Prob damaging	DC	34	5.64
chr1(GRCh37): g.164776781G>A	c.704G>A	(p.Arg235Gln)	5	–/–	Damaging	Prob damaging	DC	26.4	4.16[k]
chr1(GRCh37): g.164776860_ 164776861insC	c.783dupC	(p.Ser262Gln *fs**2)	5	–/–	nd[l]	nd	DC	nd	5.6
chr1(GRCh37): g.164781251C>T	c.862C>T	(p.Arg288*)	6	–/–	nd	nd	DC	nd	3.7

ExAC[a] = Exome aggregation consortium browser; dbSNP[b] = Single nucleotide polymorphism database; SIFT[c] = Sorting tolerant from intolerant; Poly-Phen-2[d] = Polymorphism Phenotyping v2; CADD[e] = Combined Annotation Dependent Depletion; GERP[f] = Genomic evolutionary rate profiling; –/–[g] = absent in both ExAC browser and dbSNP; DC[h] = predicted to be disease-causing; 5.76[i]—Exon 4 average GERP score was 4.19; 4–5[j] = codon crosses a splice site; 4.16[k]—Exon 5 average GERP score 4.71; nd[l] = no data.

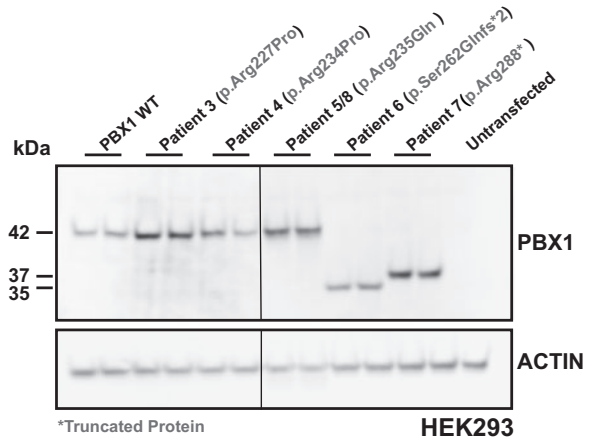


Figure 2. Analysis of ectopic expression of wildtype and mutant *PBX1* cDNAs in HEK293 cells. Human HEK293 cells were transfected in duplicate with either *PBX1* wildtype (WT) or *PBX1*-Mutant cDNAs. Two separate western blots (divided by vertical line) detect the five different *PBX1* mutant proteins identified in the six patients (Patient 3 and 4, left blot; Patient 5/8, 6 and 7, right blot). The blots were revealed with the PBX1 Ab. PBX1 WT and untransfected HEK293 cells were used as controls in both experiments. Exposure of the western blots was reduced to a few seconds in order to visualize the ectopic proteins without saturation of the signal; this did not allow concomitant detection of the endogenous PBX1 protein, which is present in HEK293 cells as we previously reported (16). Anti-ACTIN was used to normalize for protein loading. Protein sizes are indicated on the left of the blots (kiloDaltons; kDa).

All of the *PBX1* sequence variants were heterozygous and *de novo*, and none of them were found in 1000 genomes or the ExAC browser databases. Furthermore, all variants were predicted to be pathogenic and the majority had high Genomic Evolutionary Rate Profiling (GERP) scores (Table 2). All patients carried only sequence variants in *PBX1* except for patient 8, as discussed above, and no additional variants in genes associated with CAKUT were identified.

Prediction of protein structure perturbations

Structural mapping of the sites of the four, most downstream (3'), human sequence variants in *PBX1* showed that the homeodomain region was disordered and charged, two characteristic properties of DNA-binding proteins that can contribute to specific, DNA sequence selectivity for DNA interactions (14) The mutations affecting Arg234, Arg235, and probably Arg227 are hypothesized to affect the ability of PBX1 to bind DNA, whereas Arg288 comes into direct contact with the DNA and may directly affect the mode of this interaction (Fig. 1C). The amino acids that were substituted in the PBX1 homeodomain of these patients were conserved in many species (Fig. 1D).

Functional assessment of the *PBX1* sequence variants identified in six patients

The *PBX1* sequence variants identified in six out of the eight patients described in this study (Patients 3–8; Table 1) were cloned in a pSG5 expression vector and confirmed by Sanger sequencing. After transient transfection in HEK293 cells in culture, western blot analysis showed that all of the constructs ectopically expressed PBX1 proteins of various sizes, according to the mutation introduced (Fig. 2). Mutations present in Patients 3, 4 and 5/8 resulted in the production of PBX1 protein of the same size as the WT (42 kDa). In contrast, mutations found in Patients 6 and 7 yielded truncated proteins of shorter size than WT (approximately 35 and 37 kDa, respectively). These results demonstrated that all of the *PBX1* mutations identified in these six patients do not affect the production of the relative proteins.

To assess the transactivation capabilities of the PBX1 proteins carrying the sequence variants described above, we performed luciferase reporter assays following our established protocols (15,16) We employed a regulatory element present on human chromosome 1 within the *PBX1* promoter (chr1: 164, 527, 984–164, 528, 440; hg19), which is highly conserved within vertebrates (thereafter named Pbx1 Prom) (Fig. 3A). Using chromatin immunoprecipitation coupled with quantitative polymerase chain reaction (ChIP-qPCR) on mouse embryonic hindlimbs and midfaces at embryonic day (E)10.5, we unequivocally demonstrated that the orthologous element within the mouse *Pbx1* promoter (chr1: 170, 362, 511–170, 362, 668; mm9) is strongly bound by Pbx1 itself, suggesting the presence of an autoregulatory loop (Fig. 3B). *PBX1* Prom, which contains two putative PBX1 binding sites, was cloned upstream of the luciferase reporter gene (Fig. 3C) within a pGL3 vector (pGL3-*PBX1* Prom). We then used the human HEK293 cell line, which expresses endogenous PBX1 protein (16), for luciferase reporter assays. These revealed that when WT PBX1 and its cofactor PREP1 (5) were co-transfected in these cells, luciferase activity driven by the pGL3-*PBX1* Prom was increased by almost two fold *versus* the pGL3 empty vector (Fig. 4A). These results are consistent with the

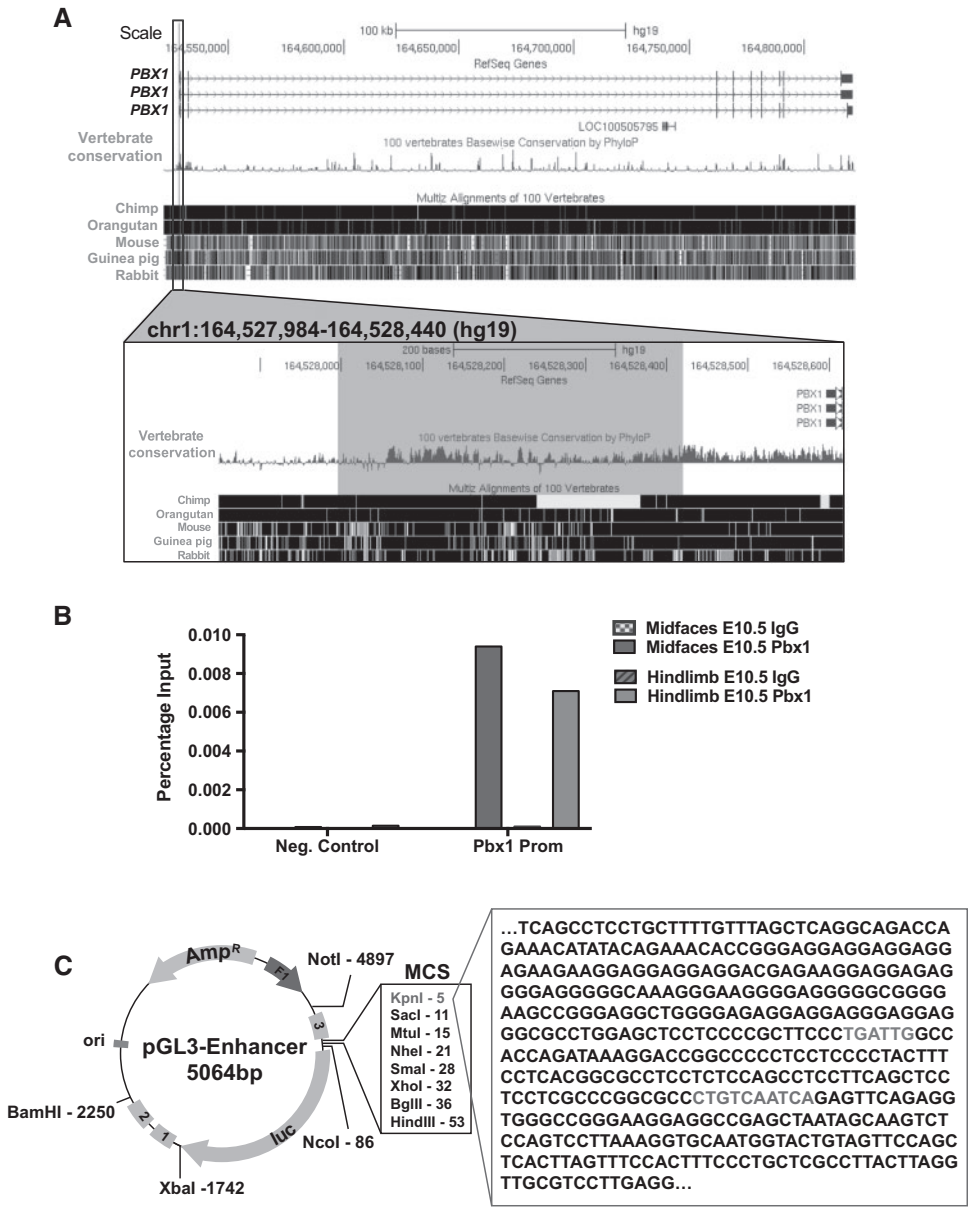


Figure 3. Identification of a *PBX1* regulatory element within *PBX1* own promoter. (**A**) Adapted, UCSC genome browser view of the human *PBX1* locus (hg19). Regulatory element present on chromosome 1 within the human *PBX1* promoter (chr1: 164, 527, 984–164, 528, 440) is highlighted in grey. This region displays high conservation within vertebrates. The light-colored bar reflects uncertainty in the relationship between the DNA of human and chimp, due to lack of sequence in relevant portions of the alignments. (**B**) Murine Pbx1 binding to *Pbx1* Prom by ChIP-qPCR. Histograms represent the enrichment of each region following immunoprecipitation with the Pbx1 Ab and the negative control Ab (IgG) calculated as percentage of input. *Pbx1* Prom is strongly bound by Pbx1 itself in both E10.5 midfaces and hindlimbs. An unbound region was employed as a negative control (Neg. control; chr11: 103, 623, 380–103, 623, 491, mm9). (**C**) A 457 bp fragment within the human *PBX1* promoter was cloned into the pGL3-TK vector using the *KpnI* restriction site. Left, vector map; Right, *PBX1* promoter sequence containing two putative PBX binding sites: TGATTG and the canonical decameric Pbx-Prep motif CTGTCAATCA (27).

moderate, but reproducible and statistically significant, transactivation activity of PBX1 with its cofactor PREP1 versus empty vector in similar reporter assays in cell culture, as we previously reported (15). In contrast, when we co-transfected the *PBX1* mutant cDNAs together with the cDNA for PREP1 in HEK293 cells, luciferase activity driven by the pGL3-*PBX1* Prom was significantly reduced for all five *PBX1* mutations analysed (Fig. 4A). Furthermore, in order to evaluate the intrinsic transactivation capability of the *PBX1* mutations described above without the interference of endogenous PBX1 protein, we generated a murine cell line with *Pbx1* loss-of-function (LOF). We used an available primary mesenchymal cell line derived from E15.5 mouse embryos (15) and targeted exon 3 of *Pbx1* (mc4-Pbx1 LOF) via clustered regularly interspaced short palindromic repeats (CRISPR)/Cas-mediated genome editing (17). After targeting of exon 3, western blot demonstrated a striking reduction of Pbx1 protein [both Pbx1a and Pbx1b (16)] in the clone with insertions/deletions (indels) affecting both alleles (*Pbx1* LOF), as compared with WT and clones with heterozygous indels (Fig. 4B). Interestingly, reporter assays performed as described above using Pbx1 LOF mc4 cells revealed that luciferase activity driven by the pGL3-*PBX1* Prom was significantly reduced only for the

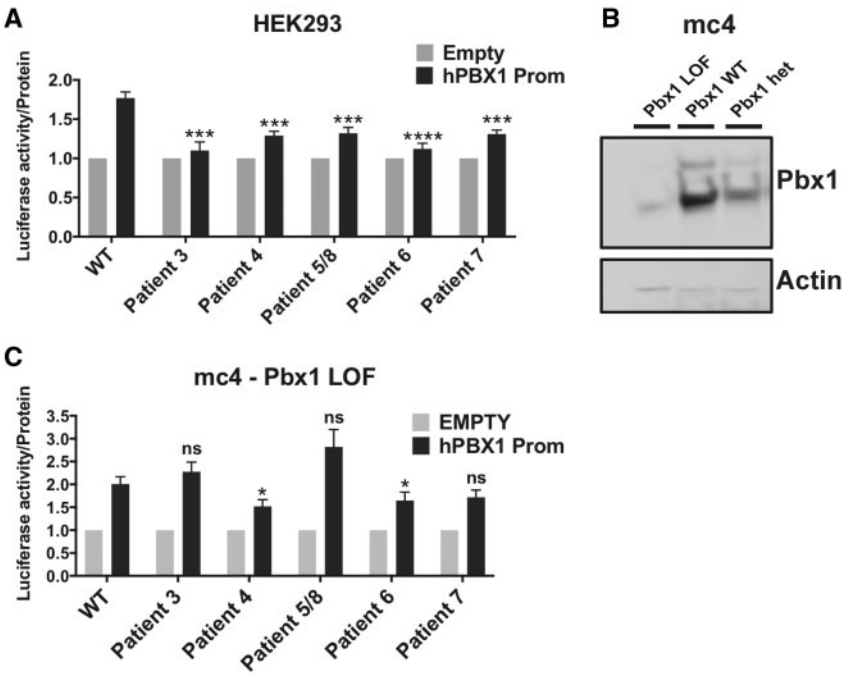


Figure 4. Functional assessment of the transactivation capability of PBX1 mutant proteins carrying the mutations identified in six patients. (**A**) Luciferase activity driven by *PBX1* Prom in human HEK293 cells transfected with PBX1 and PREP1 expression constructs. Values represent fold activation of pGL3-hPBX1 Prom over basal promoter activity (pGL3 empty vector, set to 1) normalized by total amount of protein. Each value is obtained from the average of at least five independent experiments, each performed in triplicate. Error bars represent the standard error of the mean (SEM). Asterisks represent *P*-values: (***P-value < 0.001), (****P-value < 0.0001). (**B**) Western blot analysis of murine mc4 cells subjected to CRISPR/Cas-mediated gene targeting of *Pbx1*. The blot was revealed with the PBX1 Ab. Anti-Actin was used to normalize for protein loading. Loss-of-function (LOF); presence of indels in both *Pbx1* alleles. WT; wildtype. Het; presence of indel in one *Pbx1* allele. (**C**) Luciferase assays were conducted as described in (A) using the murine cell line with Pbx1 LOF obtained via CRISPR/Cas-mediated genome editing (mc4-Pbx1 LOF). Error bars represent the SEM. Asterisks represent *P*-values: (*P-value < 0.05). ns; not significant.

mutations present in Patients 4 and 6 (Fig. 4C). These results establish that, in a cellular system with markedly reduced levels of endogenous PBX1, only the PBX1 proteins carrying the mutations identified in Patients 4 and 6 intrinsically alter the transactivation capability of the *PBX1* transcription factor. These findings also strongly suggest that the endogenous WT PBX1 protein present in HEK293 cells, which resembles human haploinsufficiency, interacts with the ectopic mutant proteins. This abnormal interaction of WT and mutant PBX1 could adversely affect target gene transactivation.

It is also worthy of note that to execute their developmental programs on specific target genes, PBX homeodomain proteins must enter the nucleus, a process that is concomitantly regulated by the presence of two nuclear localization signals (NLS) within the homeodomain (18) and by PBX dimerization with TALE and HOX cofactors (5). To assess the ability of the PBX1 mutant proteins to translocate into the cell nucleus, we further analysed the cytoplasmic and nuclear fractions of HEK293 cells transfected with cDNAs carrying the *PBX1* mutations identified in the six patients together with the PREP1 cofactor. Notably, levels of the PBX1 proteins carrying the mutations present in Patients 6 and 7 (with truncations that delete the homeodomain 3' region comprising the second NLS) were remarkably reduced in the nuclear fraction as compared with WT control (Fig. 5). However, in these mutants, the PBX1 cytoplasmic protein fraction was not reduced, ruling out confounding effects due to degradation or efficiency of transfection. These results support a scenario whereby the altered PBX1-dependent transactivation capabilities observed in Patients 6 and 7 can result, at least in part, from defective nuclear translocation.

Discussion

We have described eight individuals with intellectual disability, developmental delays, and variable abnormalities (including craniofacial, ear, branchial arch, cardiac, pulmonary, diaphragmatic, renal, and genital defects) carrying *de novo*, heterozygous, deleterious sequence variants in the *PBX1* gene. The clinical findings in individual patients overlapped sufficiently to delineate a recognizable syndrome. The pleiotropic developmental phenotype is consistent with the broad expression of *Pbx1* during murine embryogenesis, with predominant localization to neural tissues [brain, spinal cord, and ganglia (6,19)] as well as head and olfactory epithelium, heart, lung, gut, spleen, pancreas, kidney, and gonads (4,8,15). Interestingly, in view of the dysplastic ear helices observed in these patients, murine *Pbx1* craniofacial expression at E13.5 is highest in the auricular pinna (4). *Pbx1* is also detectable in the ectoderm of the second branchial arch and around the first pharyngeal groove at E11.5 (4). The human phenotype closely recapitulates the abnormalities previously described in *Pbx1*$^{-/-}$ mouse embryos that included cardiac outflow tract defects with absence of the ductus arteriosus at E14.5; diaphragmatic hernia at E15.5; bilateral or unilateral renal hypoplasia; axial malpositioning of the kidneys; unilateral renal agenesis and impaired renal function (4,7–9,16,20). *Pbx1*$^{-/-}$ mouse embryos also failed to develop typical branchial arches and, at E10.5, the pharyngeal groove separating arches 3 and 4 was absent, and the fourth branchial arch was small compared with WT embryos (8), resulting in absence or hypoplasia of organs derived from the caudal pharyngeal region (21). Conditional mutant mice with loss of *Pbx1* function in the

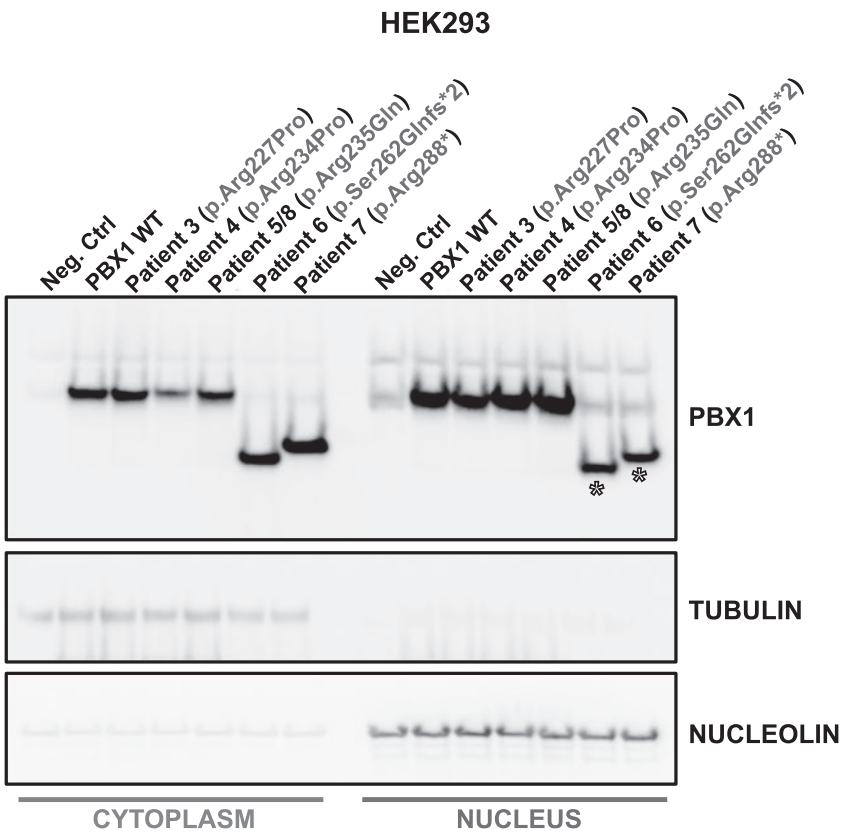


Figure 5. Mutations identified in Patients 6 and 7 result in lower levels of PBX1 proteins in the nucleus. Western blot analysis of PBX1 protein levels in cytoplasmic (left) and nuclear (right) extracts from human HEK293 cells transfected for 24 h with either *PBX1* wildtype (WT) or *PBX1*-Mutant cDNAs together with the cofactor PREP1. Cytoplasmic and nuclear fractions separated by empty lane. Patients 6 and 7 show remarkably reduced PBX1 levels in the nuclear fraction (black empty asterisks). Equal loading and nucleo-cytoplasmic fractionation were monitored with anti-Nucleolin and anti-Tubulin antibodies. Minor nuclear contamination of the cytoplasmic fraction is detectable. Neg. Ctrl, Negative control: untransfected HEK293 cells.

ventricular zone and in postmitotic neurons demonstrated abnormal laminar patterning of the neocortex, with hypoplasia and dyslamination of the frontal cortex (19) Accordingly, the developmental delays observed in our patients are likely to result from impaired PBX1-dependent transcription of critical target genes in the developing brain. The ear and branchial arch anomalies noted in our patients also bear some resemblance to those present in BOR syndrome (13). However, despite high sequence coverage, no mutations were found in any of the genes implicated in BOR syndrome (*EYA1*, *SIX5* and *SIX1*) in our patients. Interestingly, the ear anomalies, including stenosis of the external auditory canals, and the renal hypoplasia observed in the two patients with truncating mutations of *PBX1* closely resemble the phenotypes of previously reported cases with large interstitial chromosome deletions encompassing *PBX1* (10). Recently, mutations in *PBX1* were also identified in five cases out of a cohort of 204 patients with CAKUT syndrome (11). Clinical findings in the five patients included renal hypoplasia, renal dysplasia with a horseshoe kidney, deafness, developmental delays, growth delays with microcephaly and facial dysmorphic features (11). Three of the reported cases exhibited *PBX1* variants, while two had deletions encompassing *PBX1*. However, no functional studies were provided for any of the patients.

One striking observation concerning the patients described in this study is the high variability of their phenotypes, with disparate clinical findings including branchial arch rests, diaphragmatic eventration, and ambiguous genitalia (Table 1). Possible explanations for the phenotypic variability observed in patients with *PBX1* mutations may include allelic heterogeneity. Indeed, it would appear that loss-of-function alleles (deletions, frameshift and stop variants) may result in CAKUT, hearing impairment, and developmental differences, whereas missense variants may result in more complex phenotypes that include cardiac defects, diaphragmatic eventration, and ambiguous genitalia. However, the regulatory function of PBX1 as a *HOX* gene cofactor is also likely to result in substantial phenotypic heterogeneity. It is worthy of note that two patients with phenotypes of different complexity and severity carry the same *PBX1* variant (Patients 5 and 8), suggesting that interactions of PBX1 with TALE or HOX cofactors can be perturbed in a different fashion and with varied outcomes. Lastly, the presence of different genetic modifiers interacting with *PBX1* or of mutations in non-coding regulatory elements of the genome is also likely to play a role in driving genetic heterogeneity.

In humans, PBX1 encodes a 430 amino acid protein with a homeobox domain at residues 233–305 that serves as a eukaryotic DNA-binding motif. Four of the sequence variants described in our patients consisted of substitutions involving closely spaced residues within the homeodomain; in addition one variant involving p.Arg227 was immediately adjacent to the homeodomain and within the PBC domain, which is conserved among Pbx, Exd and Ceh-20, and interacts with Meis/Prep cofactors (3). This clustering of mutations is striking and argues for a critical role of the homeodomain residues for PBX1 function in humans. In support of a phenotype associated with

haploinsufficiency for *PBX1* is the dearth of loss-of-function variants listed in the ExAC browser and the pLI score (0.91) of the gene (22). Interestingly, population missense variants are not only rare in PBX1, but we noted that PBX1 has a ~120 amino-acid missense-depleted region (MDR) encompassing the homeodomain (23). Notably, the *PBX1* missense mutations identified in our patients are located within or just proximal to this MDR, further supporting their pathogenicity. It will be of interest to further ascertain why this region of PBX1 is sensitive to missense variation. It is also unclear if certain missense substitutions may act as dominant negative mutations, given the propensity of PBX1 complexes to form multimers with both TALE and non-TALE cofactors (3). Indeed, as PBX1 acts in a transcription complex with MEIS1 (3,24) and murine Meis1 is expressed in the second branchial arch (25), aberrant function of MEIS1 may also prove to be an important contributor to this aspect of the *PBX1* phenotype. It is worthy of note that two of the five *PBX1* mutations we analysed resulted in premature protein truncation due to the generation of early in-frame stop codons, but none of the identified mutations significantly perturbed protein production, as shown by our western blot analyses.

The identification of a strong auto-regulatory element within the conserved *PBX1* promoter, which we characterized *in vivo* by ChIP analysis in different embryonic tissues, provided a powerful reagent for functional assessment of the *PBX1* mutations under analysis. Parallel evaluation of the transactivation capability of PBX1 mutant proteins carrying the five different mutations in a system containing WT PBX1 (human HEK293 cells) versus a system with marked reduction of endogenous Pbx1 protein (murine mc4-Pbx1 LOF), afforded an assessment of two distinct scenarios. The former cellular assay reflects the transactivation capability of the different PBX1 mutations in the presence of WT protein, which resembles *PBX1* haploinsufficiency in the patients. The latter assay, in contrast, allows the evaluation of the intrinsic transactivation capability of the different PBX1 mutations in a system with low levels of endogenous WT PBX1 protein. Interestingly, in the presence of WT PBX1, all five mutant proteins exhibited a significant decrease in transactivation capability, despite the different locations of the mutations within the protein domains (homeodomain or PBC) (3,24). In this system, the mutant proteins might either be directly responsible for the decrease of transactivation activity observed or might affect the capability of the endogenous WT protein to physiologically transactivate target genes, a scenario that closely resembles the human disease with *PBX1* haploinsufficiency. Parallel evaluation of the PBX1 mutant proteins in the system with marked reduction of WT Pbx1 demonstrated that only two of the five mutant proteins (Patients 4 and 6) exhibited significantly diminished transactivation activity. These findings strongly suggest that these two mutations can directly affect the intrinsic capability of PBX1 to transactivate downstream transcriptional targets. Interestingly, while the mutation in Patient 7 shows a similar trend, but without reaching statistical significance (P-value = 0.08), mutations in Patients 3 and 5/8 do not exhibit this behavior. It is noteworthy that, while Patient 3 carries a mutation outside of the homeodomain, Patient 5/8 bears a mutation inside it, like Patients 4, 6 and 7, making it difficult to directly link the transactivation abnormalities of the mutated proteins solely to the presence of an abnormal homeodomain. Furthermore, the mutations identified in Patients 6 and 7 were associated with lower levels of PBX1 proteins in the cell nucleus, as compared with WT. These results highlight that in these cases altered transactivation capabilities can result, at least in part, from abnormal nuclear translocation of PBX1 and the consequent impaired formation of complexes comprising PBX1 and its cofactors, which is similar to the reported behavior of EYA1 (26). Taken together, our experiments suggest that the phenotypes observed in *PBX1* mutant patients likely result from a combination of perturbed biological processes including: 1) intrinsic alterations of transactivation ability of the PBX1 mutant proteins, at least in Patients 4 and 6; 2) broadly abnormal interactions between mutant PBX1 and WT proteins (including PBX1 itself and other TALE or HOX cofactors) (24,27); 3) defective translocation of mutant PBX1 proteins to the nucleus in Patients 6 and 7. Overall, these scenarios can affect physiological transcription of PBX1 target genes, thus causing aberrations of normal embryonic development.

We demonstrate for the first time that *de novo*, deleterious *PBX1* sequence variants can cause a syndromic form of intellectual disability with pleiotropic developmental defects that involve multiple organ systems. The novel phenotype we report here is highly variable in expressivity and severity between patients, consistent with findings we previously reported in mouse embryos with constitutive or conditional loss of *Pbx1*, suggesting high functional conservation of this essential developmental gene between these two species. It is likely that the precise definition of the human syndrome with *PBX1* mutations will be refined as an increasing numbers of cases are reported and the mechanisms of action for the sequence variants are further understood.

Materials and Methods

Case reports

Clinicians were contacted after matching in GeneMatcher (28) and informed consent for sequencing was obtained from all patients. Medical history and phenotypic features were extracted from clinical records as per standard practice.

Whole exome sequencing

Exome sequencing was performed with a trio approach on probands (patients 1–8) and biological parents for all families. All sequence variants are described with reference to *PBX1* transcript NM_ 002585.3.

For Patient 1, whole exome sequencing (WES) was performed as a research test. Exomes libraries were prepared using the Nextera Rapid Capture Exome kit (Illumina, CA). Exome libraries were sequenced on Illumina HiSeq2000 and HiSeq2500 instruments typically multiplexing 12 barcoded libraries per sequencing lane. Sequencing coverage analysis was performed using Bedtools (29) in conjunction with custom scripts.

Following exome sequencing, DNA sequence reads were aligned to the human reference genome build hg19 using the Burrows-Wheeler Aligner (30) to create Binary Alignment Map (BAM) files. Single Nucleotide Variant (SNV) calling was performed on the BAM files using Genome Analysis Toolkit (GATK) software (31) with the option of HaplotypeCaller1 (32). This generated a genomic Variant Call Format (VCF) file for each sample. VCF files from multiple samples were then processed together via a Joint Genotyping call with GATK to produce one multi-sample VCF file for all DNA variants. Variants in the VCF were annotated with various metrics using ANNOVAR (33). The VarSifter tool (34) was used to filter variants based on genomic location, allele frequency, inheritance pattern and predicted pathogenicity. Given the absent family history for congenital

heart disease (CHD), only rare variants at a Minor Allele Frequency (MAF) < 0.01 with respect to the ExAC database, fitting *de novo*, recessive, or compound heterozygous inheritance models were considered. Variants passing the population frequency thresholds were checked with the Integrative Genomics Viewer (IGV) tool (35) to confirm read quality. Variants were subsequently considered pathogenic if they exceeded both *in silico* pathogenicity prediction thresholds: Combined Annotation-Dependent Depletion (CADD) > 15 (36), and PolyPhen2-HVAR ≥ 0.446 (37).

For Patients 2–6 and 8, who underwent WES as a clinical test, sequencing technology and the variant interpretation protocol have been previously reported (38).

For Patient 7, WES and variant calling were performed as previously described (39). Briefly, the exome was captured using the Agilent SureSelect v4 kit (Agilent, Santa Clara, CA, USA). Exome libraries were sequenced on an Illumina HiSeq2000 instrument (Illumina, San Diego, CA, USA) with 101 bp paired-end reads at a median coverage of 75x. Sequence reads were aligned to the hg19 reference genome using BWA version 0.5.9-r16. Variants were subsequently called by the GATK unified genotyper, version 3.2–2 and annotated using a custom diagnostic annotation pipeline.

Protein modeling studies

We mapped the deleterious sequence variants in *PBX1* onto the PBX1 homeodomain structure (PDB code 1puf). This structure starts with residue 233, and thus does not include the Arg184, Met224 or Arg227 residues of PBX1. The mapped region is disordered and charged (14).

Site-directed mutagenesis

We generated expression vectors containing five *PBX1* cDNAs carrying the sequence variants predicting p.Arg227Pro, p.Arg234Pro, p.Arg235Gln, p.Ser262Glnfs*2 and p.Arg288* that were identified in six of the eight patients described in this study. We utilized a polymerase chain reaction (PCR)-based approach using specific primers for each single mutation (Supplementary Material, Table S1). The pSG5-PBX1b (40) expression vector was used as a template for the PCR reaction. High-fidelity Taq polymerase enzyme Q5 (New England BioLabs, MA) was employed. After purifying the PCR product from agarose gel, T4 Polynucleotide Kinase (New England BioLabs, MA) was used in order to add a 5'-phosphate to the DNA, following manufacturer's instructions. 50 ng of DNA was used in a ligation reaction with T4 ligase (New England BioLabs, MA), followed by bacterial transformation. Positive colonies carrying the desired sequence variants were identified with Sanger sequencing.

Western blot analysis

Total protein lysates, as well as nuclear and cytoplasmatic fractions, were resolved by 10% SDS-PAGE and blotted to polyvinylidene difluoride (PVDF) membranes (Millipore, CA) as previously described (41). Membranes were first incubated with the primary antibody (Ab) overnight and then with a peroxidase-conjugated secondary Ab. Peroxidase activity was measured by using an enhanced chemiluminescence (ECL) kit (Pierce/Thermo Fisher Scientific), following the manufacturer's instructions. The Abs used in this study were: Pbx1 (Cell Signaling, 4342), Actin (Proteintech, IL), Tubulin (Proteintech, IL), and Nucleolin (Santa Cruz Biotechnology, CA).

Chromatin immunoprecipitation (ChIP)-quantitative (q)PCR

Embryonic tissues (midfaces and hindlimbs) were isolated from embryonic day (E)10.5 mouse embryos after timed-matings of Swiss-Webster mice. The embryonic tissues were immediately crosslinked for 10 min in 1% formaldehyde (Electron Microscopy Sciences, #15710). ChIP assays were performed as previously described (42). The crosslinked material was sonicated to 200–500 basepair (bp) fragments (Diagenode Bioruptor). Immunoprecipitation (IP) was performed from 12–14 midfaces or 24–26 hindlimbs from E10.5 embryos, by overnight incubation at 4 °C with 1 µg of anti-Pbx1 Ab or control IgG (Cell Signalling, 2729), followed by Dynabeads protein A (Invitrogen). After washing, IP and input DNA was de-crosslinked and purified using QIAquick PCR purification kit (Qiagen, 28106). qPCR was performed to compare enrichment of specific genomic locations after immunoprecipitation on a QuantStudio 6 Flex Real-Time PCR System (Life Technologies). Enrichment was calculated as percentage of the input. Primers used for qPCR amplification (40 cycles) were:

Pbx1 Prom forward 5'- GCCCCCTTCTATGATTGGCTAGT-3';

reverse 5'- TGGGTCATAGCTCCGCCTCC-3';

Negative control forward 5'- GCCAGCCTGAGCTATATGG A-3';

reverse 5'-TGGACGCTGGGAACTAAATC-3'.

Cell culture and transfections

The human HEK293 cell line and a primary murine mesenchymal cell line derived from E15.5 embryonic spleens (mc4) were cultured in DMEM (Dulbecco's modified Eagle's medium; Corning) supplemented with 10% fetal bovine serum (FBS; Gibco) and penicillin (100 U/ml)/streptomycin (100 U/ml; Lonza) at 37 °C, under 5% CO_2. Cell transfections were carried out using Lipofectamine 2000 (Thermo Fisher Scientific), following the manufacturer's instructions.

Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas-mediated genome editing of *Pbx1* in mc4

The following guide RNAs (gRNAs; CRs) were designed to target exon 3 of murine *Pbx1* by employing the CRISPR design software (http://tools.genome-engineering.org): CR1: AGACCCCCAGCTCA TGCGAC, CR2: TATTCGGGGAGCCCAAGAAG. These 2 CRs were independently cloned in the piCRg vector, which contains Cas9 and the tracrRNA (17). The targeting vectors were then delivered to the primary mesenchymal cells by lipofection. Single clones were picked and Sanger sequencing was used to identify the presence of the mutations introduced by CRISPR-Cas.

Luciferase assays

A 457 bp fragment encompassing two of the *PBX1* binding sites within the *PBX1* promoter was amplified from human genomic DNA using the following primers:

Forward 5'-GGGGTACCCCTCAGCCTCCTGCTTTTGTTT-3'

Reverse 5'- GGGGTACCCCCCTCAAGGACGCAACCTAAG-3'.

KpnI restriction sites (underlined) were used to clone the fragment into the pGL3-TK vector (Promega). Transfections were conducted in 24-well plates, as previously reported (15,16). A total amount of 350 ng DNA per well was transfected into the cells and comprised 50 ng of the reporter vector (either pGL3-TK-PBX1 promoter or empty pGL3-TK), 150 ng of the pSG5-PREP1 construct expressing the cDNA of the PBX1 cofactor *PREP1*, and 150 ng of the pSG5-PBX1 construct carrying either WT *PBX1* cDNA, or one of the five *PBX1* cDNAs containing the human mutations. For all of these experiments, the short isoform of WT PBX1 (PBX1b) (40) was used. 48 h after transfection, cells were harvested and lysed in Passive Lysis Buffer (Promega). Lysates were analysed by using the Luciferase Assay System (Promega) in a Glomax Explorer System (Promega). Normalization was performed after measuring the protein content for each sample using a Bradford protein assay (Bio-Rad Laboratory). The results represent an average of five replicate experiments, with each experiment performed in triplicate. An unpaired t-test was conducted to determine statistical significance.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

We are grateful to the patients and families for their participation and to the UCLA Clinical Genomics Center for their help with sequencing Patient 3. L.S. wants to personally thank Mr. Robert Aho for assistance with preparation of the figures.

Conflict of Interest statement. Megan Cho and Kristin Monaghan are employees of GeneDx.

Funding

The National Institutes of Health [R01 grant DE024745 to L.S.], the Dr. Caroline H. Damsky Award to M.R., and the Australian National Health and Medical Research Council Fellowship [ID1042002 to S.L.D].

References

1. Nourse, J., Mellentin, J.D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S.D. and Cleary, M.L. (1990) Chromosomal translocation t(1; 19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. *Cell*, **60**, 535–545.
2. Rauskolb, C., Peifer, M. and Wieschaus, E. (1993) extradenticle, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene pbx1. *Cell*, **74**, 1101–1112.
3. Moens, C.B. and Selleri, L. (2006) Hox cofactors in vertebrate development. *Dev. Biol.*, **291**, 193–206.
4. Selleri, L., Depew, M.J., Jacobs, Y., Chanda, S.K., Tsang, K.Y., Cheah, K.S., Rubenstein, J.L., O'Gorman, S. and Cleary, M.L. (2001) Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development*, **128**, 3543–3557.
5. Berthelsen, J., Kilstrup-Nielsen, C., Blasi, F., Mavilio, F. and Zappavigna, V. (1999) The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH. *Genes Dev.*, **13**, 946–953.
6. Hanley, O., Zewdu, R., Cohen, L.J., Jung, H., Lacombe, J., Philippidou, P., Lee, D.H., Selleri, L. and Dasen, J.S. (2016) Parallel Pbx-dependent pathways govern the coalescence and fate of motor columns. *Neuron*, **91**, 1005–1020.
7. Schnabel, C.A., Godin, R.E. and Cleary, M.L. (2003) Pbx1 regulates nephrogenesis and ureteric branching in the developing kidney. *Dev. Biol.*, **254**, 262–276.
8. Chang, C.P., Stankunas, K., Shang, C., Kao, S.C., Twu, K.Y. and Cleary, M.L. (2008) Pbx1 functions in distinct regulatory networks to pattern the great arteries and cardiac outflow tract. *Development*, **135**, 3577–3586.
9. Stankunas, K., Shang, C., Twu, K.Y., Kao, S.C., Jenkins, N.A., Copeland, N.G., Sanyal, M., Selleri, L., Cleary, M.L. and Chang, C.P. (2008) Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease. *Circ. Res.*, **103**, 702–709.
10. Le Tanno, P., Breton, J., Bidart, M., Satre, V., Harbuz, R., Ray, P.F., Bosson, C., Dieterich, K., Jaillard, S., Odent, S. *et al.* (2017) PBX1 haploinsufficiency leads to syndromic congenital anomalies of the kidney and urinary tract (CAKUT) in humans. *J. Med. Genet.*, **54**, 502–510.
11. Heidet, L., Morinière, V., Henry, C., De Tomasi, L., Reilly M.L., Humbert, C., Alibeu, O., Fourrage, C., Bole-Feysot, C., Nitschké, P. *et al.* (2017) Targeted exome sequencing identifies *PBX1* as involved in monogenic congenital anomalies of the kidney and urinary tract. *J. Am. Soc. Nephrol.*, May 31. pii: ASN.2017010043.
12. Coras, B., Hafner, C., Roesch, A., Vogt, T., Landthaler, M. and Hohenleutner, U. (2005) Congenital cartilaginous rests of the neck (wattles). *Dermatol. Surg.*, **31**, 1349–1350.
13. Morisada, N., Nozu, K. and Iijima, K. (2014) Branchio-oto-renal syndrome: comprehensive review based on nationwide surveillance in Japan. *Pediatr. Int.*, **56**, 309–314.
14. Vuzman, D. and Levy, Y. (2010) DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proc. Natl. Acad. Sci. USA*, **107**, 21004–21009.
15. Koss, M., Bolze, A., Brendolan, A., Saggese, M., Capellini, T.D., Bojilova, E., Boisson, B., Prall, O.W.J., Elliott, D.A., Solloway, M. *et al.* (2012) Congenital asplenia in mice and humans with mutations in a Pbx/Nkx2-5/p15 module. *Developmental Cell*, **22**, 913–926.
16. Hurtado, R., Zewdu, R., Mtui, J., Liang, C., Aho, R., Kurylo, C., Selleri, L. and Herzlinger, D. (2015) Pbx1-dependent control of VMC differentiation kinetics underlies gross renal vascular patterning. *Development*, **142**, 2653–2664.
17. González, F., Zhu, Z., Shi, Z.D., Lelli, K., Verma, N., Li, Q.V. and Huangfu, D. (2014) An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell*, **15**, 215–226.
18. Saleh, M., Huang, H., Green, N.C. and Featherstone, M.S. (2000) A conformational change in PBX1A is necessary for its nuclear localization. *Exp. Cell Res.*, **260**, 105–115.
19. Golonzhka, O., Nord, A., Tang, P.L., Lindtner, S., Ypsilanti, A.R., Ferretti, E., Visel, A., Selleri, L. and Rubenstein, J.L. (2015) Pbx regulates patterning of the cerebral cortex in progenitors and postmitotic neurons. *Neuron*, **88**, 1192–1207.
20. Russell, M.K., Longoni, M., Wells, J., Maalouf, F.I., Tracy, A.A., Loscertales, M., Ackerman, K.G., Pober, B.R., Lage, K., Bult,

C.J. et al. (2012) Congenital diaphragmatic hernia candidate genes derived from embryonic transcriptomes. *Proc. Natl. Acad. Sci. USA*, **109**, 2978–2983.
21. Manley, N.R., Selleri, L., Brendolan, A., Gordon, J. and Cleary, M.L. (2004) Abnormalities of caudal pharyngeal pouch development in Pbx1 knockout mice mimic loss of Hox3 paralogs. *Dev. Biol.*, **276**, 301–312.
22. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B. *et al.* (2016) Analysis of protein-coding genetic variation in 60, 706 humans. *Nature*, **536**, 285–291.
23. Ge, X., Gong, H., Dumas, K., Litwin, J., Phillips, J., Waisfisz, Q., Weiss, M., Hendriks, Y., Suurman, K., Nelson, S. *et al.* (2016). Missense-depleted regions in population exomes implicate ras superfamily nucleotide-binding protein alteration in patients with brain malformation. *NPJ Genom. Med.*, pii: 16036. doi: 10.1038/npjgenmed.2016.36.
24. Longobardi, E., Penkov, D., Mateos, D., De Florian, G., Torres, M. and Blasi, F. (2014) Biochemistry of the tale transcription factors PREP, MEIS, and PBX in vertebrates. *Dev. Dyn*, **243**, 59–75.
25. González-Lázaro, M., Roselló-Díez, A., Delgado, I., Carramolino, L., Sanguino, M.A., Giovinazzo, G. and Torres, M. (2014) Two new targeted alleles for the comprehensive analysis of Meis1 functions in the mouse. *Genesis*, **52**, 967–975.
26. Ohto, H., Kamada, S., Tago, K., Tominaga, S.I., Ozaki, H., Sato, S. and Kawakami, K. (1999) Cooperation of six and eya in activation of their target genes through nuclear translocation of Eya. *Mol. Cell Biol.*, **19**, 6815–6824.
27. Penkov, D., Mateos San Martín, D., Fernandez-Díaz, L.C., Rosselló, C.A., Torroja, C., Sánchez-Cabo, F., Warnatz, H.J., Sultan, M., Yaspo, M.L., Gabrieli, A. *et al.* (2013) Analysis of the DNA-binding profile and function of TALE homeoproteins reveals their specialization and specific interactions with Hox genes/proteins. *Cell Rep.*, **3**, 1321–1333.
28. Sobreira, N., Schiettecatte, F., Valle, D. and Hamosh, A. (2015) GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.*, **36**, 928–930.
29. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
30. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
31. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, **20**, 1297–1303.
32. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M. *et al.* (2011) A framework for variation discovery and genotyping using nextgeneration DNA sequencing data. *Nat. Genet.*, **43**, 491–498.
33. Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data. *Nucl. Acids Res.*, **38**, e164.
34. Teer, J.K., Green, E.D., Mullikin, J.C. and Biesecker, L.G. (2012) VarSifter: visualizing and analyzing exome-scale sequence variation data on a desktop computer. *Bioinformatics*, **28**, 599–600.
35. Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*, **14**, 178–192.
36. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M. and Shendure, J. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.*, **46**, 310–315.
37. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
38. Tanaka, A.J., Cho, M.T., Millan, F., Juusola, J., Retterer, K., Joshi, C., Niyazov, D., Garnica, A., Gratz, E., Deardorff, M. *et al.* (2015) Mutations in SPATA5 are associated with microcephaly, intellectual disability, seizures, and hearing loss. *Am. J. Hum. Genet.*, **97**, 457–464.
39. de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C. *et al.* (2012) Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.*, **367**, 1921–1929.
40. Schnabel, C.A., Selleri, L., Jacobs, Y., Warnke, R. and Cleary, M.L. (2001) Expression of Pbx1b during mammalian organogenesis. *Mech. Dev*, **100**, 131–135.
41. Risolino, M., Mandia, N., Iavarone, F., Dardaei, L., Longobardi, E., Fernandez, S., Talotta, F., Bianchi, F., Pisati, F., Spaggiari, L. *et al.* (2014) Transcription factor PREP1 induces EMT and metastasis by controlling the TGF-β-SMAD3 pathway in non-small cell lung adenocarcinoma. *Proc. Natl. Acad. Sci. U S A*, **111**, E3775–E3784.
42. Amin, S. and Bobola, N. (2014) Chromatin immunoprecipitation and chromatin immunoprecipitation with massively parallel sequencing on mouse embryonic tissue. *Methods Mol. Biol.*, **1196**, 231–239.