Supplemental Information

Figure S1. Pedigrees and localization of *PITX1* variants

A. Localization of *PITX1* variants identified in the study along the protein. Yellow motif represents homeobox domain. **B.** Pedigrees of *LMX1B* families

Figure S2. Protein sequence alignment of *PITX1*

This multiple sequence alignment of PITX1 protein across different species show conservation and the localization of *PITX1* variants identified in this study.

Figure S3. Protein sequence alignment of *HOXD12*

This multiple sequence alignment of HOXD12 protein across different species show conservation and the localization of *HOXD12* variants identified in this study.

Figure S4. Pedigrees of segregated *HOX* families

Figure S5. Pedigrees of *LMX1B* families

Figure S6. Pedigrees of *COL9A3* families

Figure S7. Pedigrees of *COL15A1* families

Table S1. Filters used in gene burden analysis

NFE: non-Finish European; AC: allele count; AF: allele frequency.

Table S2. Detailed Information for Enriched genes in burden analysis

AR: autosomal recessive; AD: autosomal dominant.

Inheritance and musculoskeletal phenotypes: a brief phenotype description from OMIM and MGI (Mouse Genome Informatics) databases. If the genes are not OMIM disease genes, we provide skeletal or limb/digit phenotypes from mouse studies if there is any.

Related: gene invovled in limb-related phenotypes/functions. This information is used to prioritize our enriched genes. “Yes” indicates the gene has limb-related phenotypes/functions, “No” indicates no related phenotype/function, and “Maybe” indicates gene with skeletal related phenotype, such as osteogenesis imperfecta, but does not have strong link to limb phenotype.

Table S3. Candidate SNVs in prioritized enriched genes in clubfoot cohort

Relative.Degree: if there are relatives with clubfoot. If there are first degree relatives with clubfoot, the information is marked as Yes/Fisrt Degree

AF_nfe: allele frequency from gnomAD non-Finish European data

limb_samples: sample(s) carry this variant in the cohort

CLNSIG: reported clinical significance in ClinVar

CLNDN: reported clinical phenotype reported in ClinVar

InterVar_automated: predicted clinical significance in InterVar

ACMG: the category of variants based on ACMG guideline

Segregation: segregation results. "NA" means no other family members to test segregation; "Potentially" means inherited from the parent with family history; "TRUE" means the variant

segregate with the phenotype; "De novo" means both parents do not have phenotype and do not have the variant

Dz.model: disease model (HET: heterozygous; HOM: homozygous)

TRAPD_public: whether the variant is included in 857 clubfoot patients in the TRAPD assay using gnomAD non-Finish European data. "TRUE" means the variant is included in the burden analysis. "SCREEN" means the variant is obtained by screening throughout the whole cohort.

TRAPD_inhouse: whether the variant is included in 857 clubfoot patients in the TRAPD assay using in-house Caucasian controls. "TRUE" means the variant is included in the burden analysis. "SCREEN" means the variant is obtained by screening throughout the whole cohort.

PATH: pathogenic

VUS: Variant of Uncertain Significance

self-reported Ethnicity: NA means not reported

other genetic factors: whether there are additional possibly causative genetic findings in the exome for the indicated person. “NA” means we did not identify additional possibly causative genetic factors.

Table S4. Candidate SNVs in other posterior *HOX* genes in clubfoot cohort

Relative.Degree: if there are relatives with clubfoot. If there are first degree relatives with clubfoot, the information is marked as Yes/Fisrt Degree

AF_nfe: allele frequency from gnomAD non-Finish European data

limb_samples: sample(s) carry this variant in the cohort

CLNSIG: reported clinical significance in ClinVar

CLNDN: reported clinical phenotype reported in ClinVar

InterVar_automated: predicted clinical significance in InterVar

ACMG: the category of variants based on ACMG guideline

Segregation: segregation results. "NA" means no other family members to test segregation; "Potentially" means inherited from the parent with family history; "TRUE" means the variant segregate with the phenotype; "De novo" means both parents do not have phenotype and do not have the variant

Dz.model: disease model (HET: heterozygous; HOM: homozygous)

TRAPD_public: whether the variant is included in 857 clubfoot patients in the TRAPD assay using gnomAD non-Finish European data. "TRUE" means the variant is included in the burden analysis. "SCREEN" means the variant is obtained by screening throughout the whole cohort.

TRAPD_inhouse: whether the variant is included in 857 clubfoot patients in the TRAPD assay using in-house Caucasian controls. "TRUE" means the variant is included in the burden analysis. "SCREEN" means the variant is obtained by screening throughout the whole cohort.

PATH: pathogenic

VUS: Variant of Uncertain Significance

self-reported Ethnicity: NA means not reported

other genetic factors: whether there are additional possibly causative genetic findings in the exome for the indicated person. “NA” means we did not identify additional possibly causative genetic factors.

Table S5. Gene and region burden results for *COL12A1*

gnomAD TRAPD: TRAPD assay using gnomAD non-Finish European data

in-house TRAPD: TRAPD assay using in-house Caucasian controls

P_DOM: p value in dominant mode of TRAPD

CASE: count in cases

CONTROL: count in controls

COL12A1_whole: the whole COL12A1 region, including both isoforms

COL12A1_diff: the unique region encoding the long isoform

COL12A1_share: the shared regions between the long and short isoforms of COL12A1

Table S6. Candidate SNVs in *COL15A1* in clubfoot cohort

Relative.Degree: if there are relatives with clubfoot. If there are first degree relatives with clubfoot, the information is marked as Yes/Fisrt Degree

AF_nfe: allele frequency from gnomAD non-Finish European data

limb_samples: sample(s) carry this variant in the cohort

CLNSIG: reported clinical significance in ClinVar

CLNDN: reported clinical phenotype reported in ClinVar

InterVar_automated: predicted clinical significance in InterVar

ACMG: the category of variants based on ACMG guideline

Segregation: segregation results. "NA" means no other family members to test segregation; "Potentially" means inherited from the parent with family history; "TRUE" means the variant segregate with the phenotype; "De novo" means both parents do not have phenotype and do not have the variant

Dz.model: disease model (HET: heterozygous; HOM: homozygous)

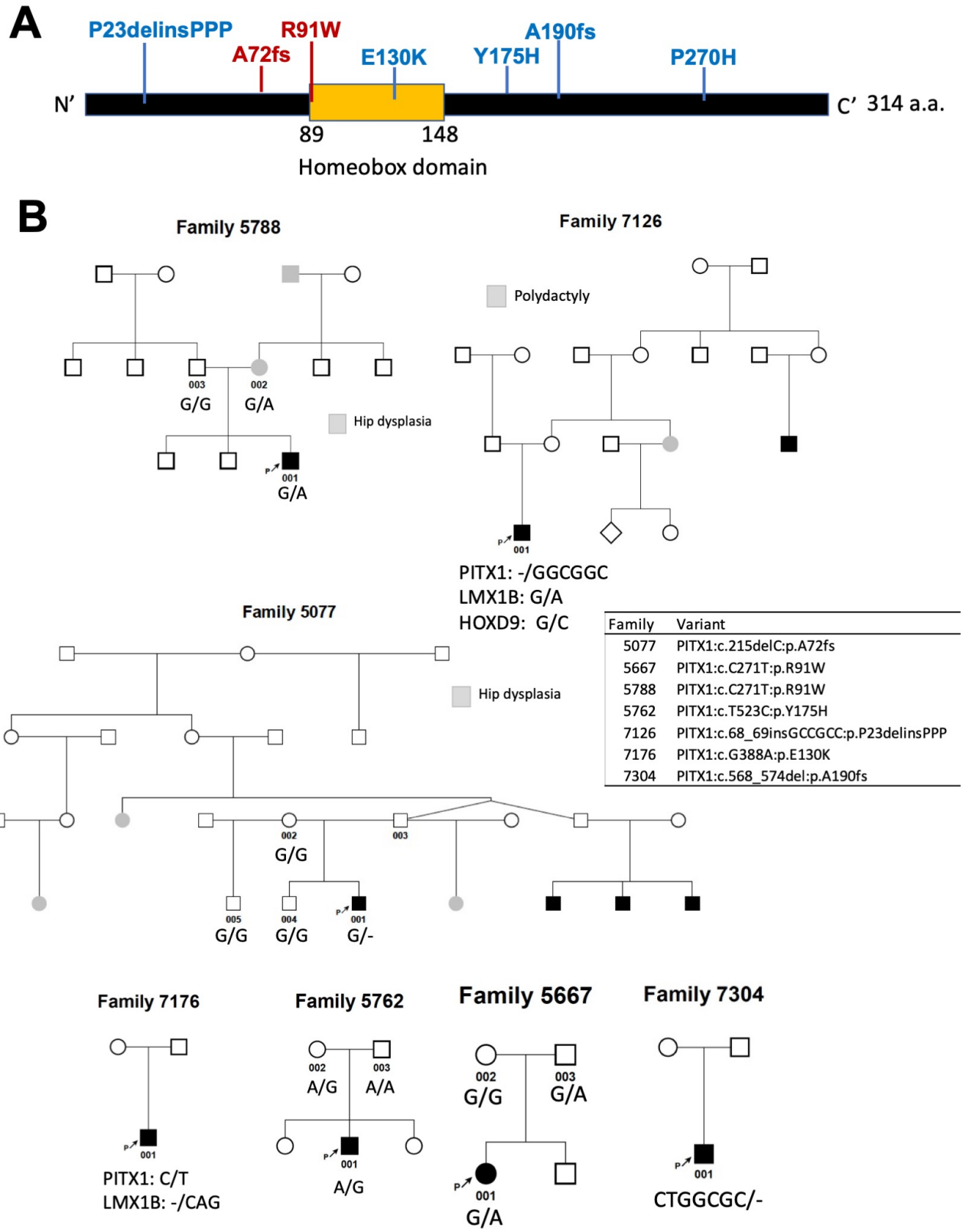
TRAPD_public: whether the variant is included in 857 clubfoot patients in the TRAPD assay using gnomAD non-Finish European data. "TRUE" means the variant is included in the burden analysis. "SCREEN" means the variant is obtained by screening throughout the whole cohort.

TRAPD_inhouse: whether the variant is included in 857 clubfoot patients in the TRAPD assay using in-house Caucasian controls. "TRUE" means the variant is included in the burden analysis. "SCREEN" means the variant is obtained by screening throughout the whole cohort.

PATH: pathogenic

VUS: Variant of Uncertain Significance

other genetic factors: whether there are additional possibly causative genetic findings in the exome for the indicated person. “NA” means we did not identify additional possibly causative genetic factors.



Family	Variant
5077	PITX1:c.215delC:p.A72fs
5667	PITX1:c.C271T:p.R91W
5788	PITX1:c.C271T:p.R91W
5762	PITX1:c.T523C:p.Y175H
7126	PITX1:c.68_69insGCCGCC:p.P23delinsPPP
7176	PITX1:c.G388A:p.E130K
7304	PITX1:c.568_574del:p.A190fs

Figure S1. Pedigrees and localization of *PITX1* variants

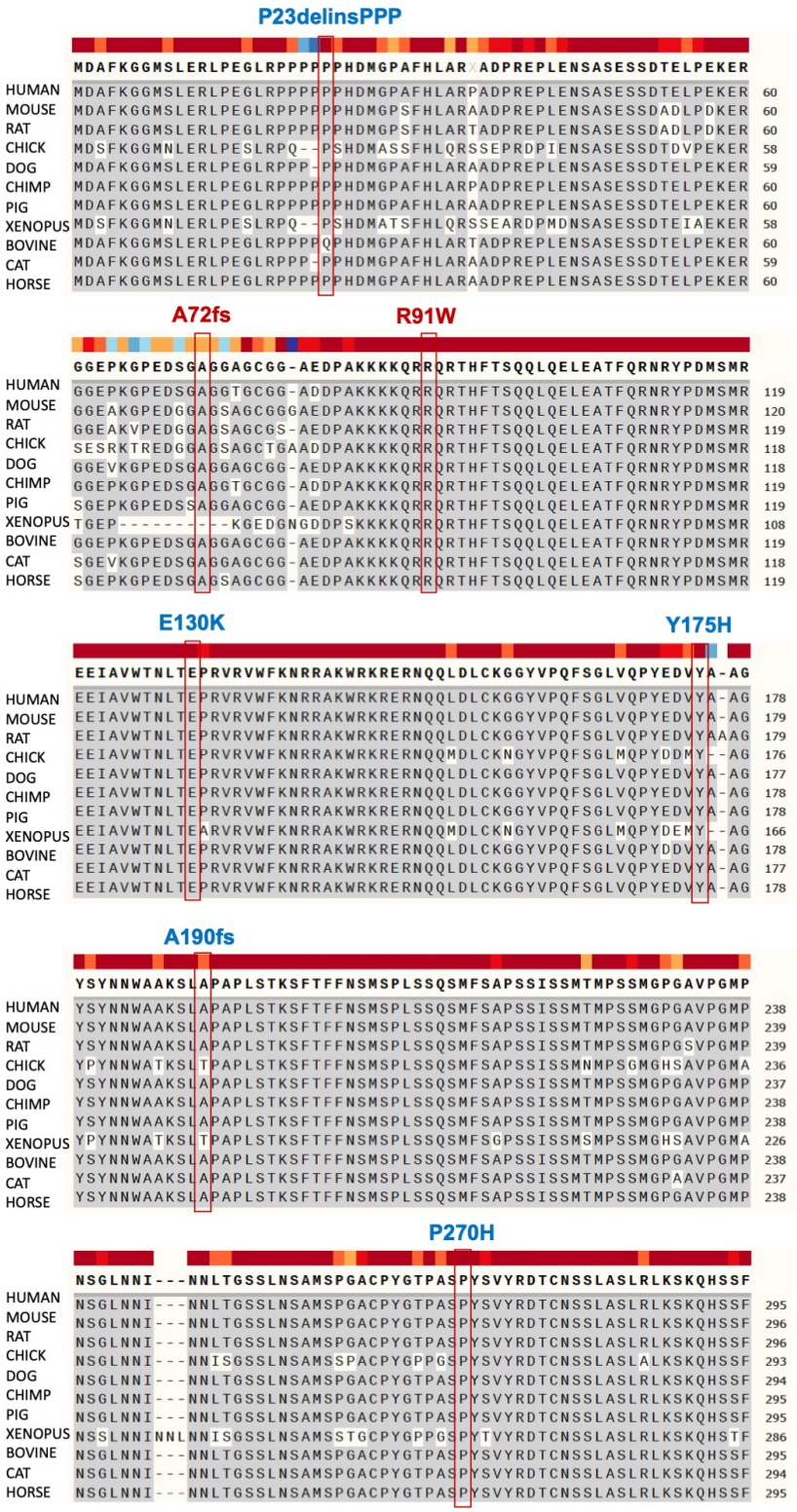


Figure S2. Protein sequence alignment of PITX1

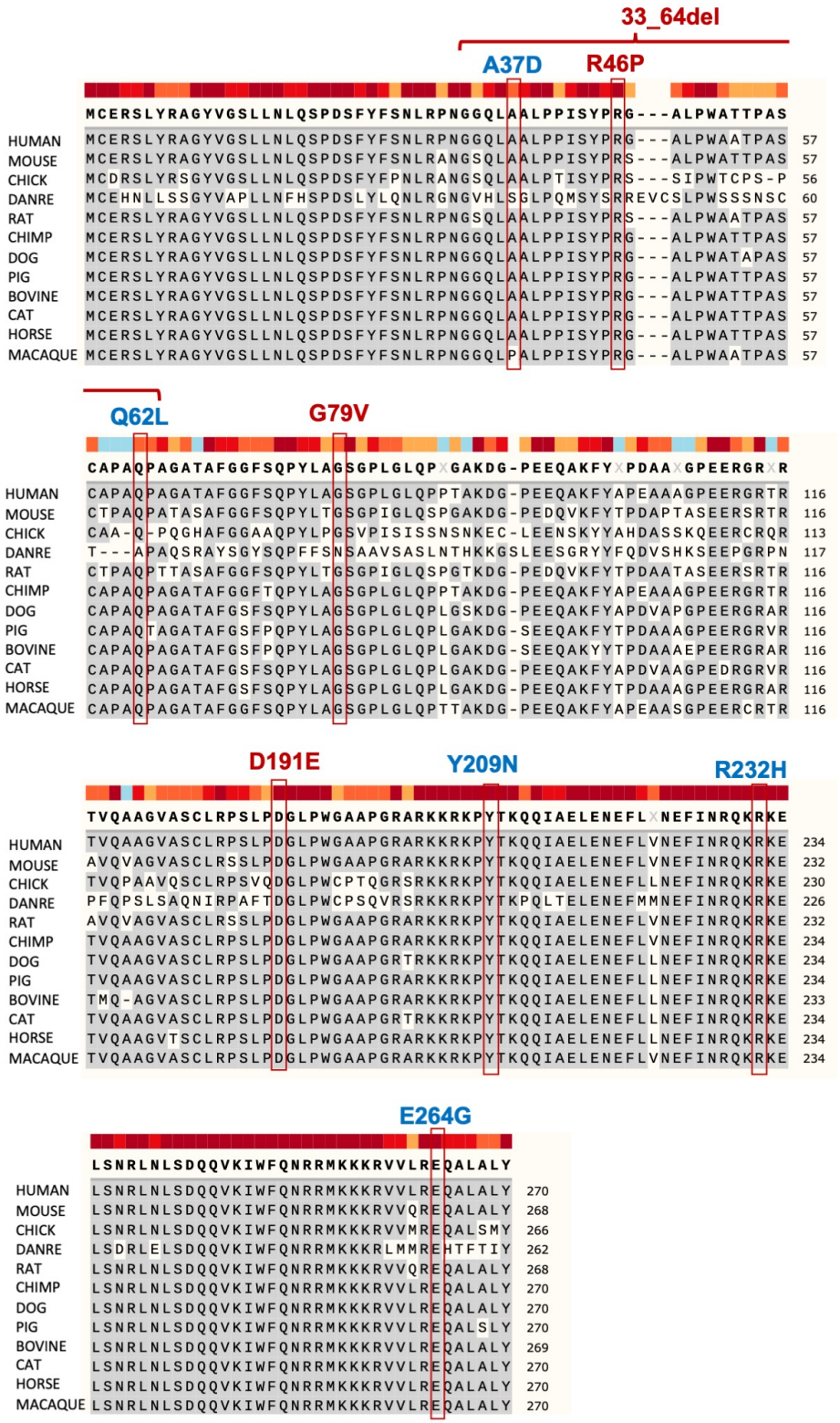
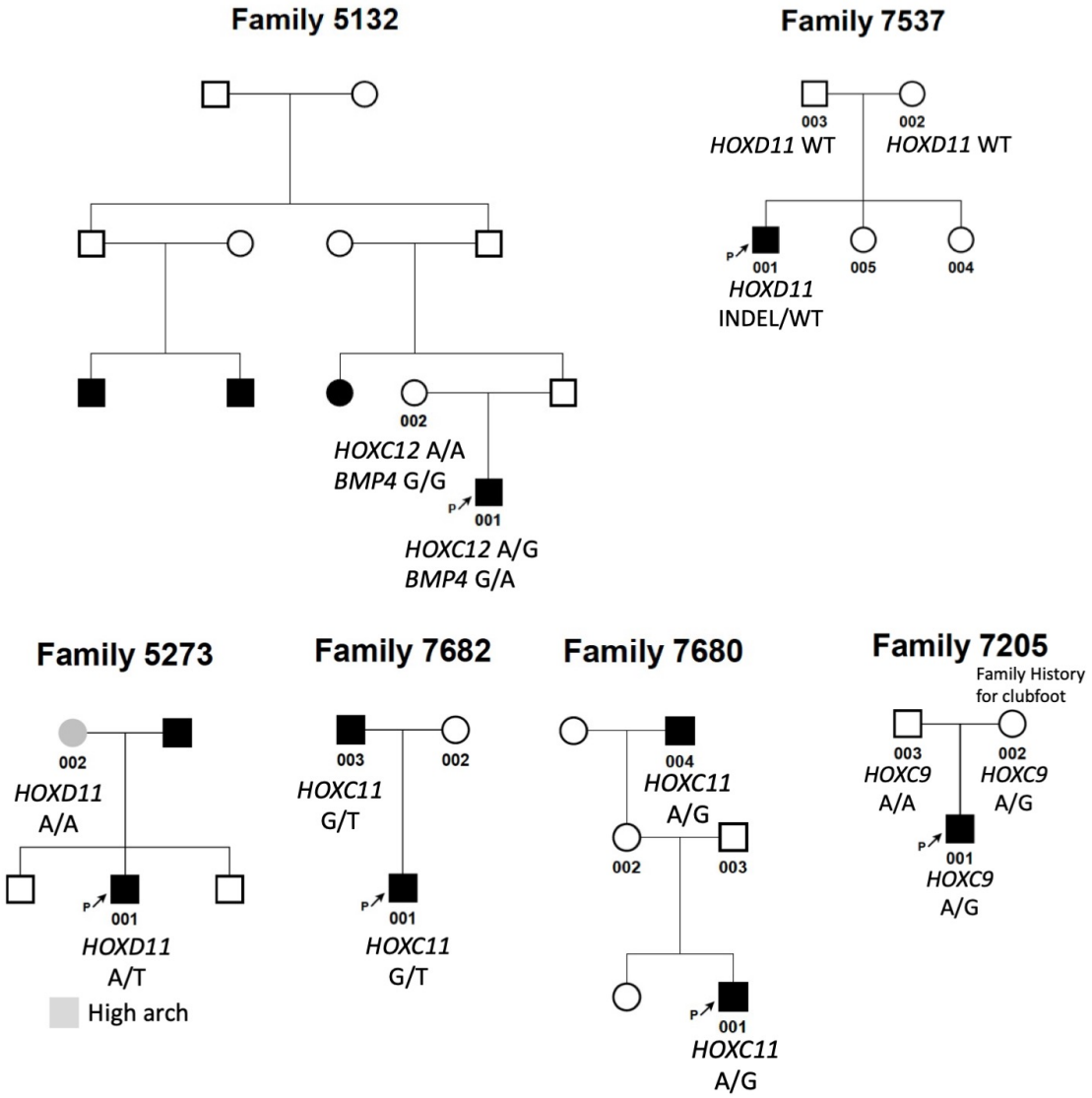


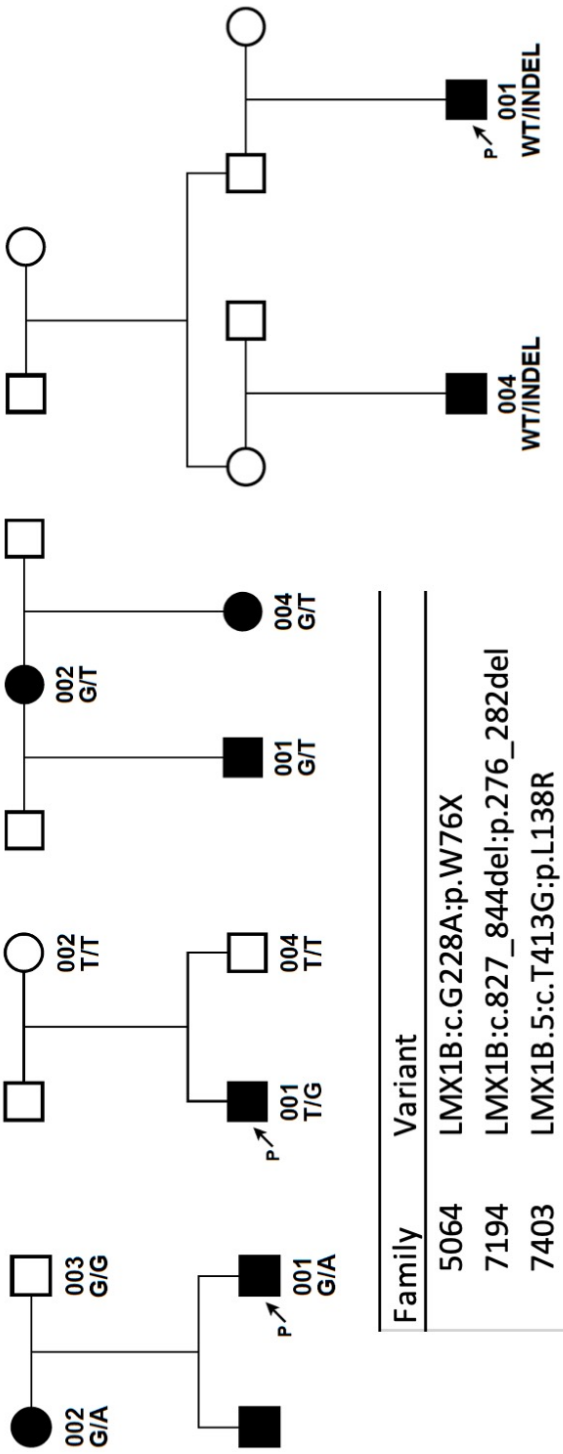
Figure S3. Protein sequence alignment of HOXD12



Family	Variant
7537	HOXD11:c.266_280del:p.89_94del
5132	HOXC12:c.A1G:p.M1V
7682	HOXC11:c.G70T:p.E24X
7680	HOXC11:c.A860G:p.E287G
5273	HOXD11:c.A1T:p.M1L
7205	HOXC9:c.A748G:p.M250V

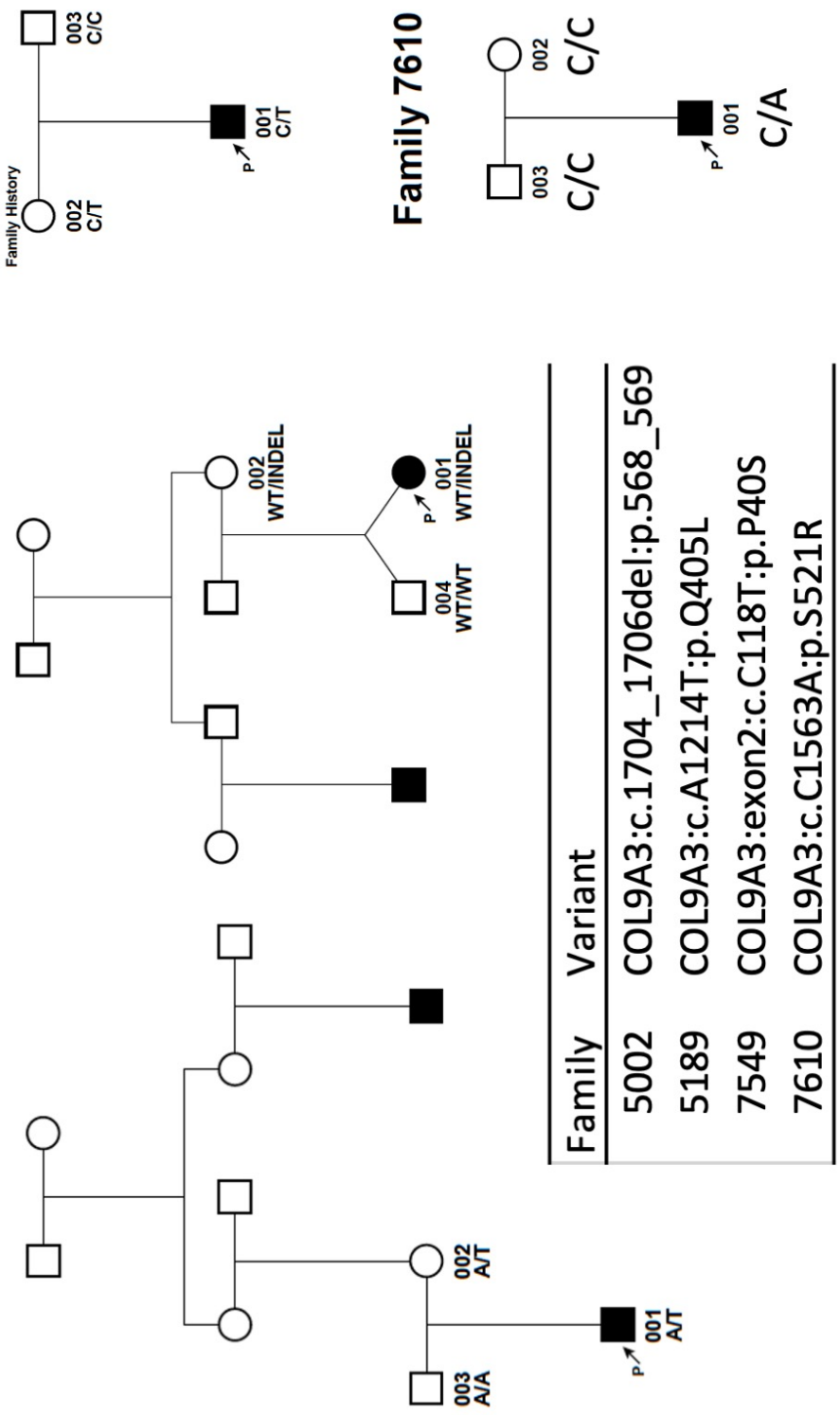
Figure S4. Pedigrees of segregated HOX families

Figure S5. Pedigrees of *LMX1B* families

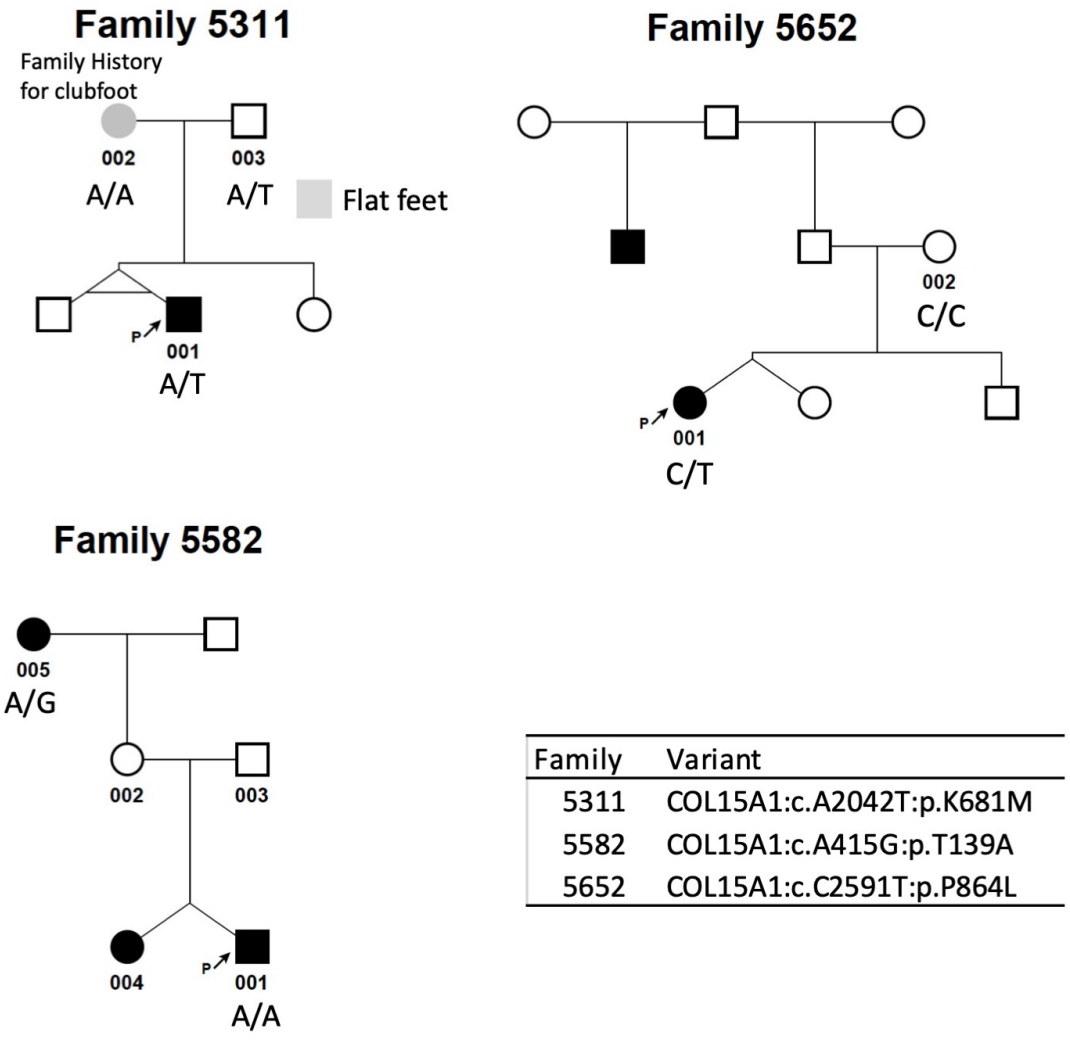


Family	Variant
5064	LMX1B:c.G228A:p.W76X
7194	LMX1B:c.827_844del:p.276_282del
7403	LMX1B.5:c.T413G:p.L138R
7706	LMX1B:c.G344T:p.C115F

Figure S6. Pedigrees of *COL9A3* families



Family	Variant
5002	COL9A3:c.1704_1706del:p.568_569
5189	COL9A3:c.A1214T:p.Q405L
7549	COL9A3:exon2:c.C118T:p.P40S
7610	COL9A3:c.C1563A:p.S521R



Family	Variant
5311	COL15A1:c.A2042T:p.K681M
5582	COL15A1:c.A415G:p.T139A
5652	COL15A1:c.C2591T:p.P864L

Figure S7. Pedigrees of *COL15A1* families