# Supplemental Material

## Regulatory Elements in *SEM1–DLX5–DLX6* (7q21.3) Locus Contribute to Genetic Control of Coronal Nonsyndromic Craniosynostosis and Bone Density-Related Traits

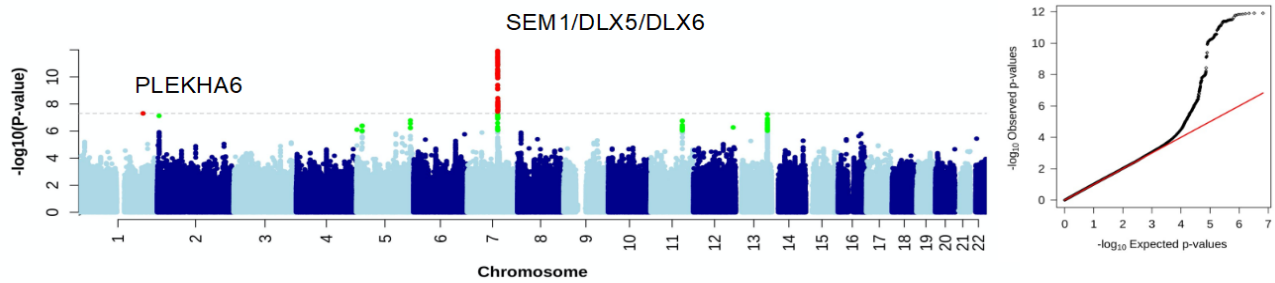
Paola Nicoletti, Samreen Zafer, Lital Matok, Inbar Irron, Meidva Patrick, Rotem Haklai, John Erol Evangelista, Giacomo B Marino, Avi Ma'ayan, Anshuman Sewda, Greg Holmes, Sierra R. Britton, Won Jun Lee, Meng Wu, Ying Ru, Eric Arnaud, Lorenzo Botto, Lawrence C. Brody, Jo C. Byren, Michele Caggana, Suzan L. Carmichael, Deirdre Cilliers, Kristin Conway, Karen Crawford, Araceli Cuellar, Federico Di Rocco, Michael Engel, Jeffrey Fearon, Marcia L. Feldkamp, Richard Finnell, Sarah Fisher, Christian Freudlsperger, Gemma Garcia-Fructuoso, Rhinda Hagge, Yann Heuzé, Raymond J Harshbarger, Charlotte Hobbs, Meredith Howley, Mary M Jenkins, David Johnson, Cristina M. Justice, Alex Kane, Denise Kay, Arun Kumar Gosain, Peter Langlois, Laurence Legal-Mallet, Angela E. Lin, James L. Mills, Jenny E. V. Morton, Peter Noons, Andrew Olshan, John Persing, Julie M. Phipps, Richard Redett, Jennita Reefhuis, Elias Rizk, Thomas D Samson, Gary M. Shaw, Robert Sicko, Nataliya Smith, David Staffenberg, Joan Stoler, Elizabeth Sweeney, Peter J. Taub, Andrew T. Timberlake, Jolanta Topczewska, Steven A. Wall, Alexander F. Wilson, Louise C. Wilson, Simeon A. Boyadjiev, Andrew O. M. Wilkie, Joan T. Richtsmeier, Ethylin Wang Jabs, Paul A. Romitti, David Karasik, Ramon Y. Birnbaum, Inga Peter

# Table of Content

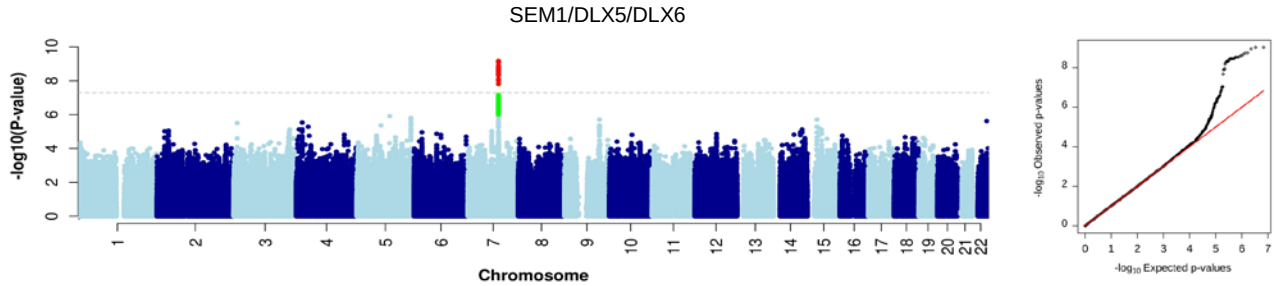
## Figure S1: Manhattan plots and quantile-quantile plots.

Panel A shows the Manhattan plot and QQ plots of the discovery European case-control GWAS study. Panel B shows the Manhattan plot and QQ plots of the multiethnic family-based study (241 European probands overlapped in both analyses); In the Manhattan plots, SNPs shown in green have a significance P-value < $5x10^{-6}$ and those shown in red have a P-value < $5x10^{-8}$. The genome-wide P-value threshold is indicated with a dotted horizontal line at $5x10^{-8}$. QQ plots show the distribution of the negative log transformed observed and expected P-values and the excess of strong associations, without any genomic inflation.

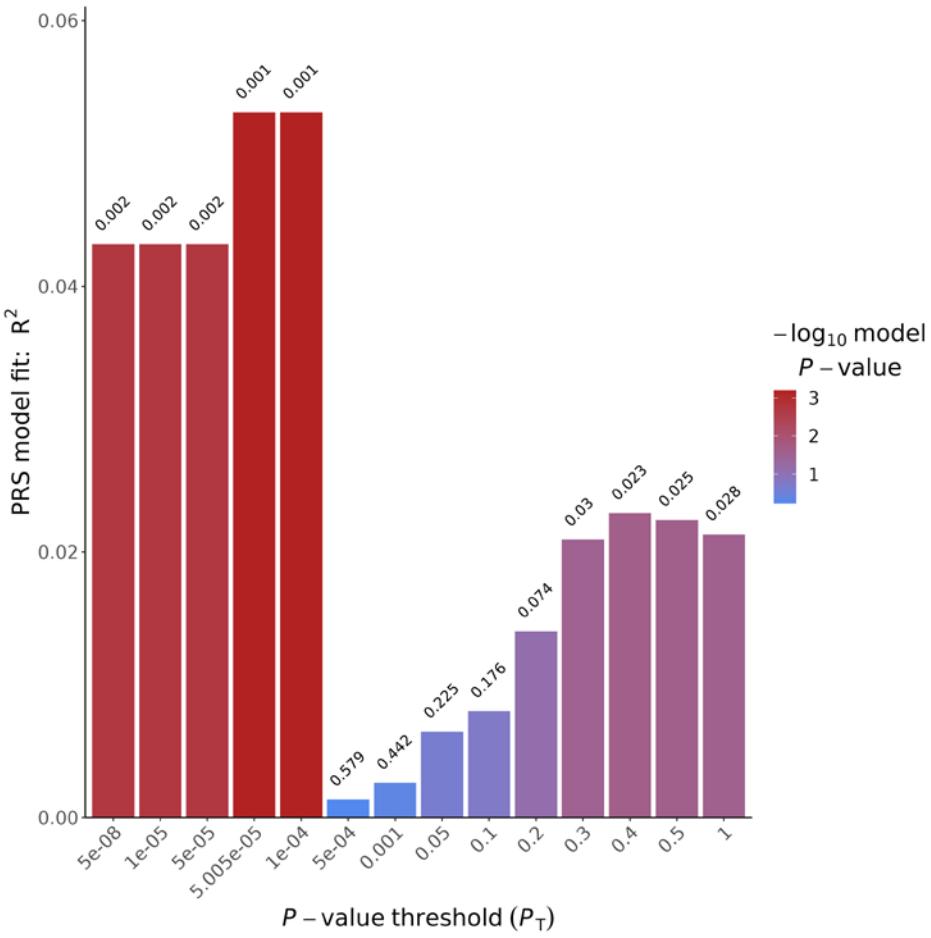
(A)




(B)



## Figure S2: Bar plot of PRS for cNCS predicting cNCS risk at broad P-value thresholds.

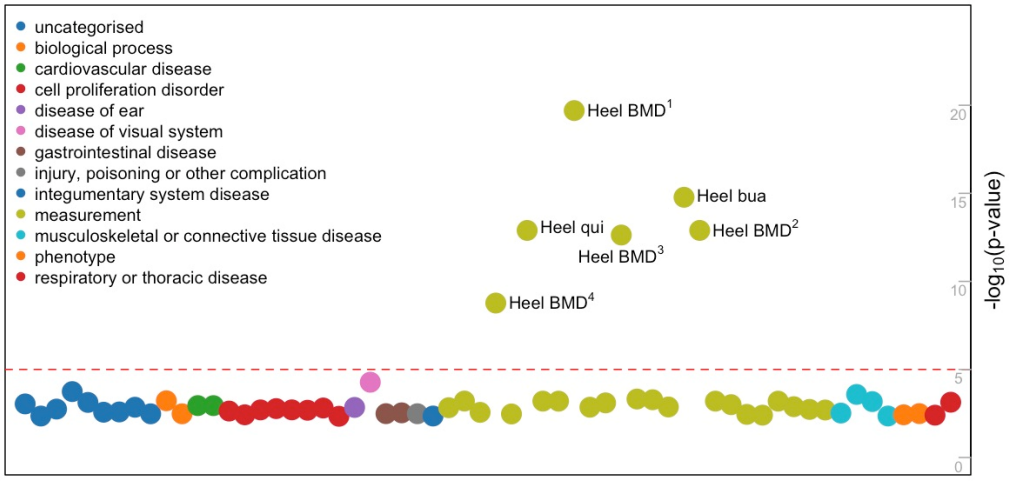



The prediction bar plot generated by PRSice2 displays different GWAS p-value thresholds ranging from 1 to 5x10-8 from European discovery summary statistics (X-axis) and model-fit Nagelkerke's R-squared (Y-axis). The P-values on the top of the bars represent statistical significance for that threshold in the European validation cohort.

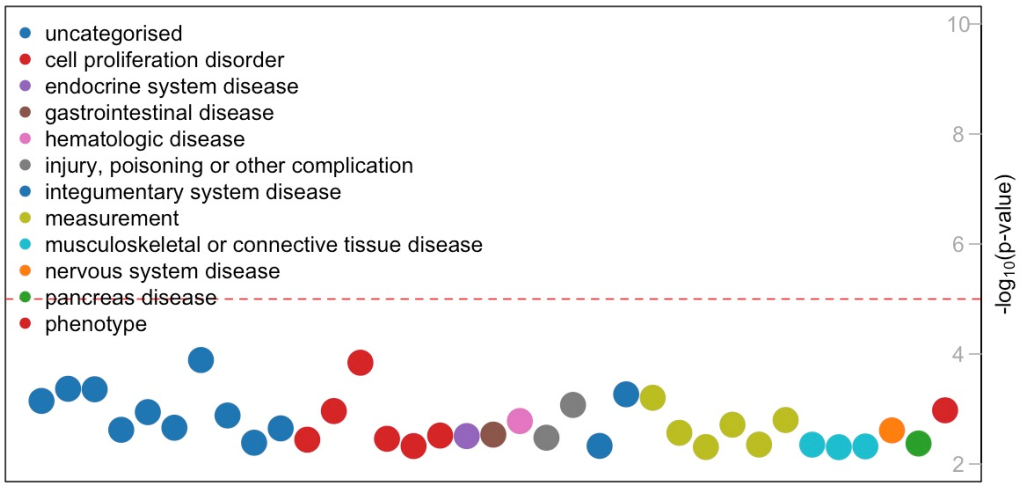
## Figure S3: Phenotype-wide association analysis of complex traits with the top cNSC association variants.

Phenotype-wide association analysis of complex traits associated with **A.** rs17656761, **B.** rs17656761, **C.** rs78353978, **D.** rs33863 and **E.** rs7981517.
Summary statistics from the UK Biobank, FinnGen, and GWAS Catalog repositories were downloaded from Open Target (OT, https://genetics.opentargets.org/). Only traits with p-value < 0.005 are shown in the diagram. Y-axis shows the variant's p-value of association to each trait. The circles are color-coded by the trait category (legend inset) as reported in OT website. The red dashed line shows the significance threshold corrected for the number of traits shown. In panel A, heel bone mineral density appears multiple times since the association was reported in many independent studies/publications as following: Heel bone mineral density (Heel BMD): [1]GCST006979, [2]NEALE2_78_raw (t-score automated), [3]NEALE2_3148_raw, [4]GCST006288. Other heel measurements are also seen - Heel broadband ultrasound attenuation, direct entry (Heel bua): NEALE2_3144_raw and Heel quantitative ultrasound index, direct entry (Heel qui): NEALE2_3147_raw.

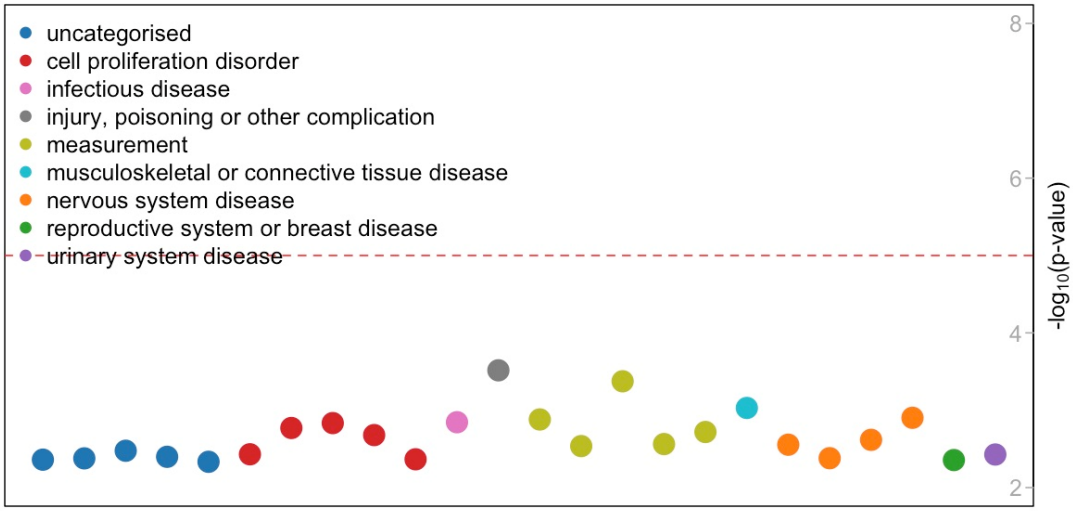
**A**



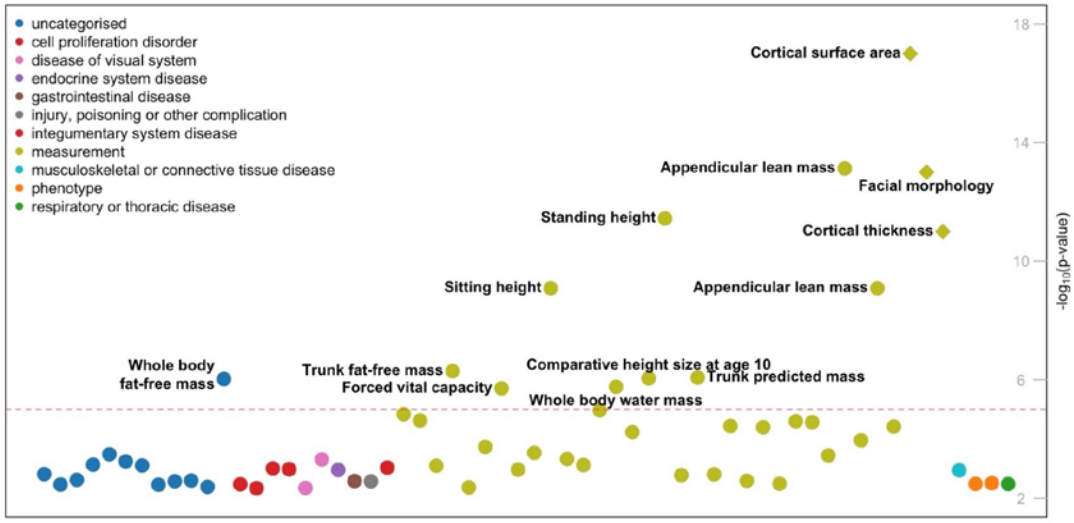
**B**


uncategorised
cell proliferation disorder
endocrine system disease
gastrointestinal disease
hematologic disease
injury, poisoning or other complication
integumentary system disease
measurement
musculoskeletal or connective tissue disease
nervous system disease
pancreas disease
phenotype
10
8
6
4
2
-log10(p-value)

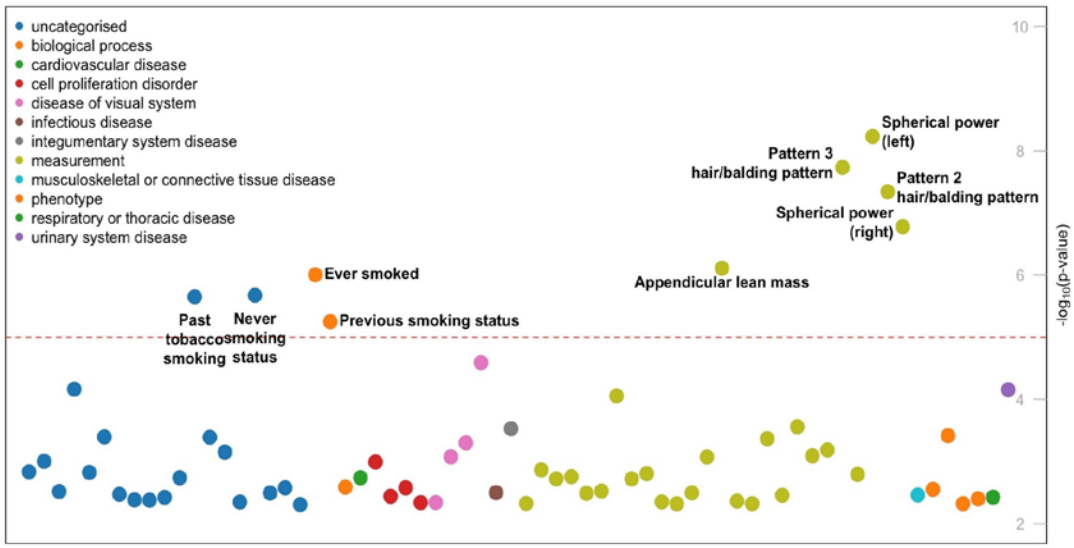
**C**


uncategorised
cell proliferation disorder
infectious disease
injury, poisoning or other complication
measurement
musculoskeletal or connective tissue disease
nervous system disease
reproductive system or breast disease
urinary system disease
8
6
4
2
-log10(p-value)

**D**

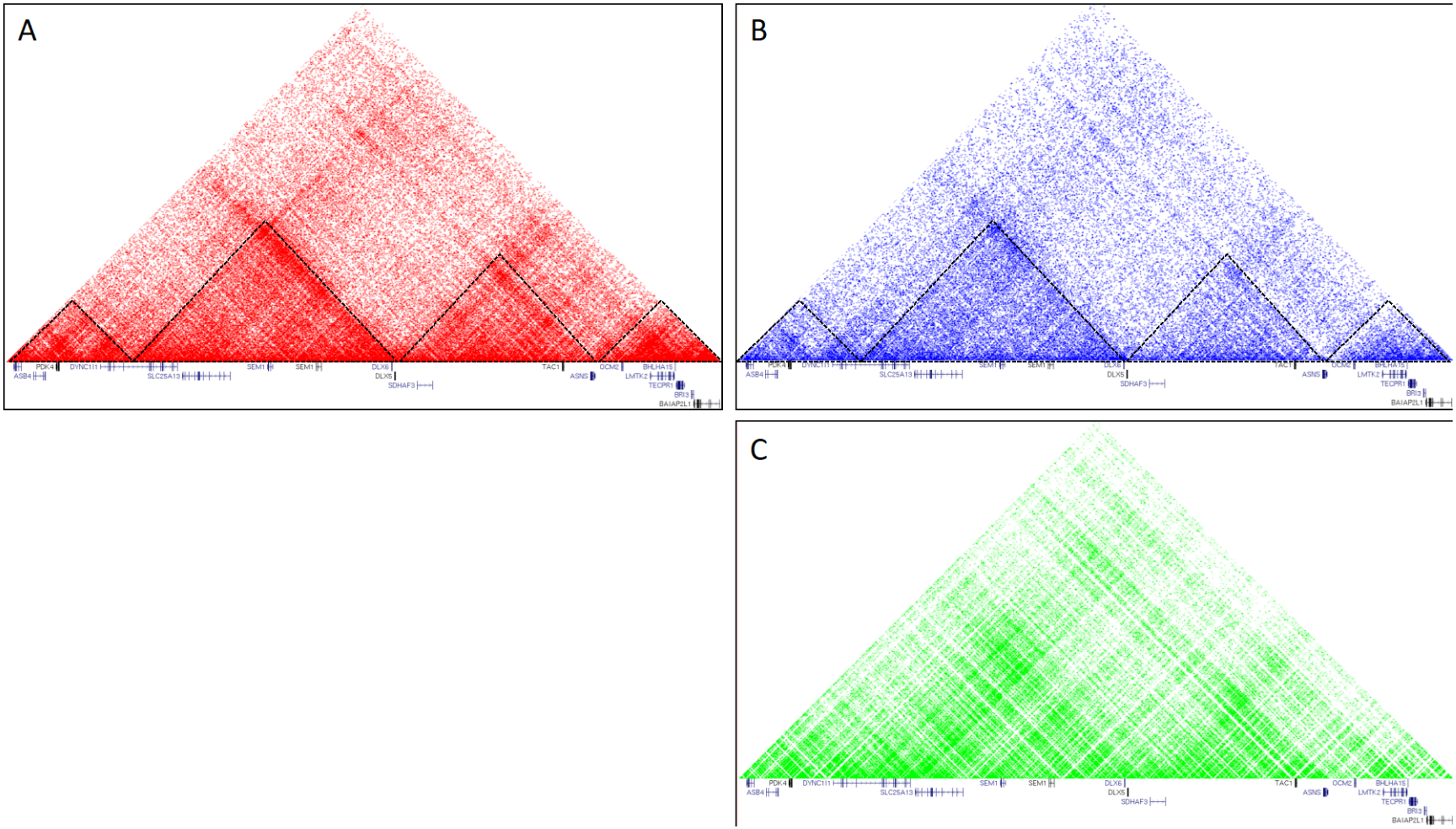

uncategorised
cell proliferation disorder
disease of visual system
endocrine system disease
gastrointestinal disease
injury, poisoning or other complication
integumentary system disease
measurement
musculoskeletal or connective tissue disease
phenotype
respiratory or thoracic disease
Cortical surface area
Appendicular lean mass
Facial morphology
Standing height
Cortical thickness
Sitting height
Appendicular lean mass
Whole body fat-free mass
Trunk fat-free mass
Forced vital capacity
Comparative height size at age 10
Trunk predicted mass
Whole body water mass
-log10(p-value)
18
14
10
6
2

**E**


uncategorised
biological process
cardiovascular disease
cell proliferation disorder
disease of visual system
infectious disease
integumentary system disease
measurement
musculoskeletal or connective tissue disease
phenotype
respiratory or thoracic disease
urinary system disease
Spherical power (left)
Pattern 3 hair/balding pattern
Pattern 2 hair/balding pattern
Spherical power (right)
Ever smoked
Appendicular lean mass
Past tobacco smoking
Never smoking status
Previous smoking status
-log10(p-value)
10
8
6
4
2

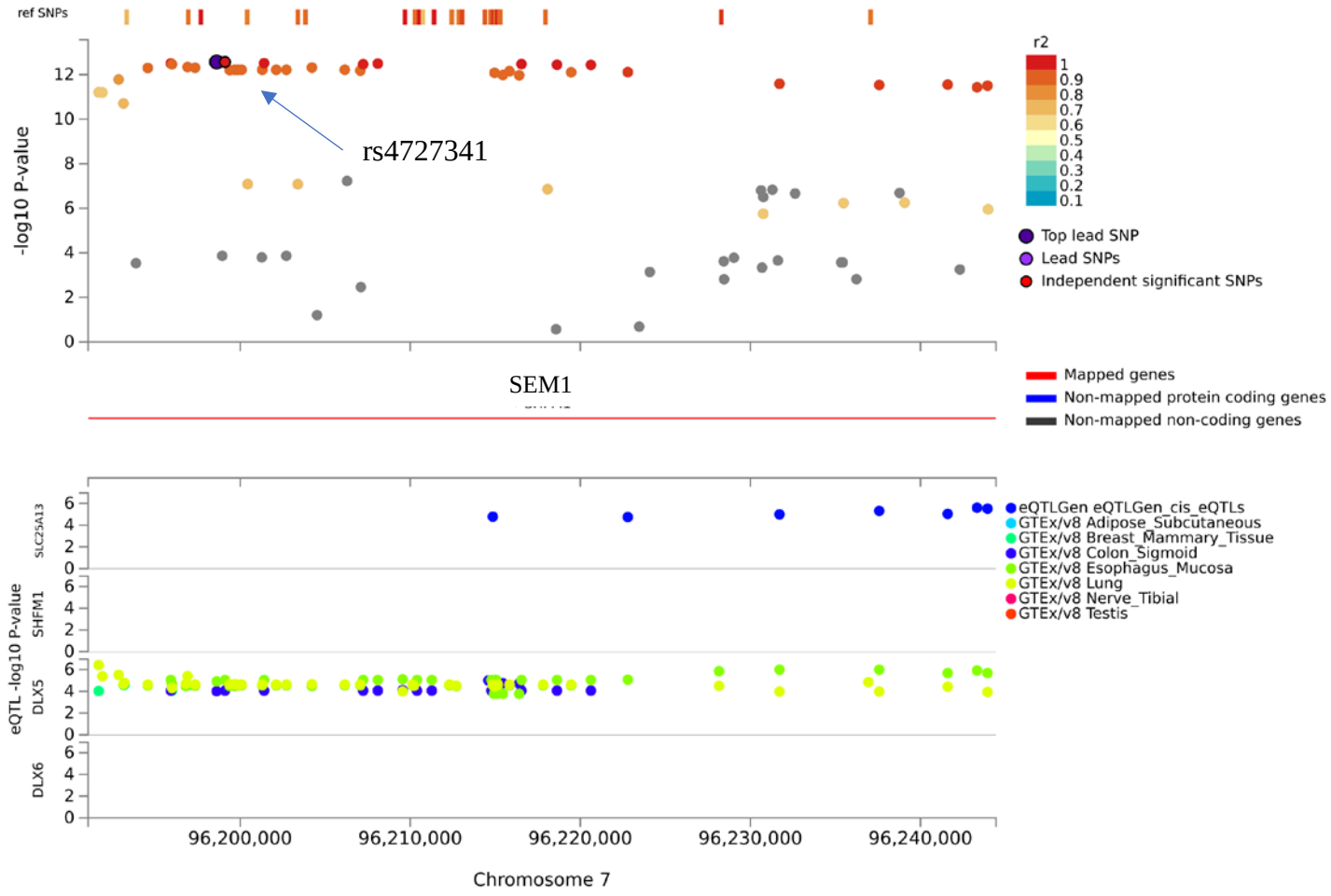
## Figure S4: Topologically Associating Domains (TAD) organization at the SEM1-DLX5-DLX6 locus.

**A.** Hi-C data of human cultured cranial neural crest cells and **B.** human Carnegie stage 17 demonstrate that SEM, DLX5 and DLX6 are in the same TAD, while Hi-C data of **C.** H1 derived mesenchymal stem cells show different TAD organization. Black triangles marked the TADs in the locus. Data was extracted from Wilderman et al.[1]



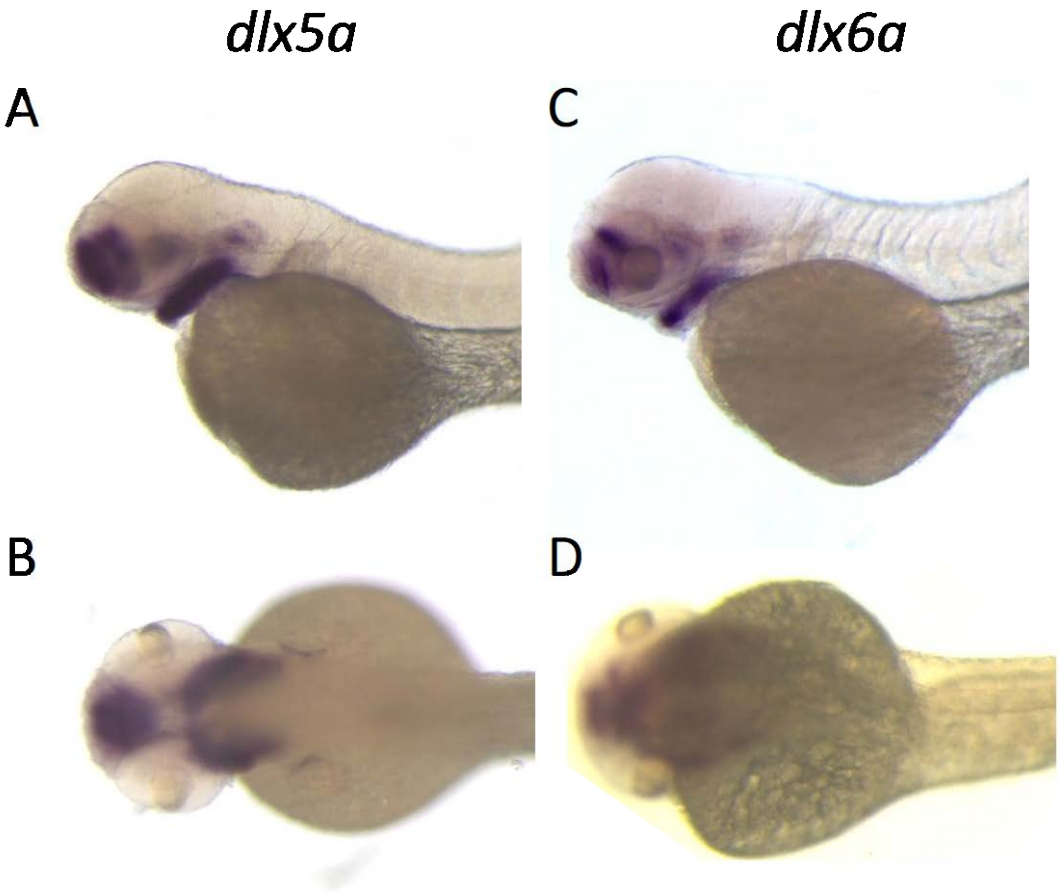
## Figure S5: Publicly available eQTL regulatory annotation of rs4727341 and related variants

Upper panel displays regional association plot of rs4727341 (purple dot) and its high linkage disequilibrium (LD) region ($r^2 > 0.9$). The x-axis represents genomic region of the rs4727341 and of LD related variants while the y-axis shows −log10 P values for individual SNPs from European meta-analysis. Pairwise LD ($r^2$) with the lead variant based on 1000 Genomes phase 3 European reference samples and is described using the color scale given. The bottom panel reports the eQTL variants located in the genomic interval colored by databases as reported in the legend on right side. Each dot is a SNP which corresponds to a SNP reported in upper panel. The x-axis represents genomic location of each SNP, and the y-axis shows the eQTL *P*-value correlated to the expression of the gene reported in each row. The figure was generated in FUMA.



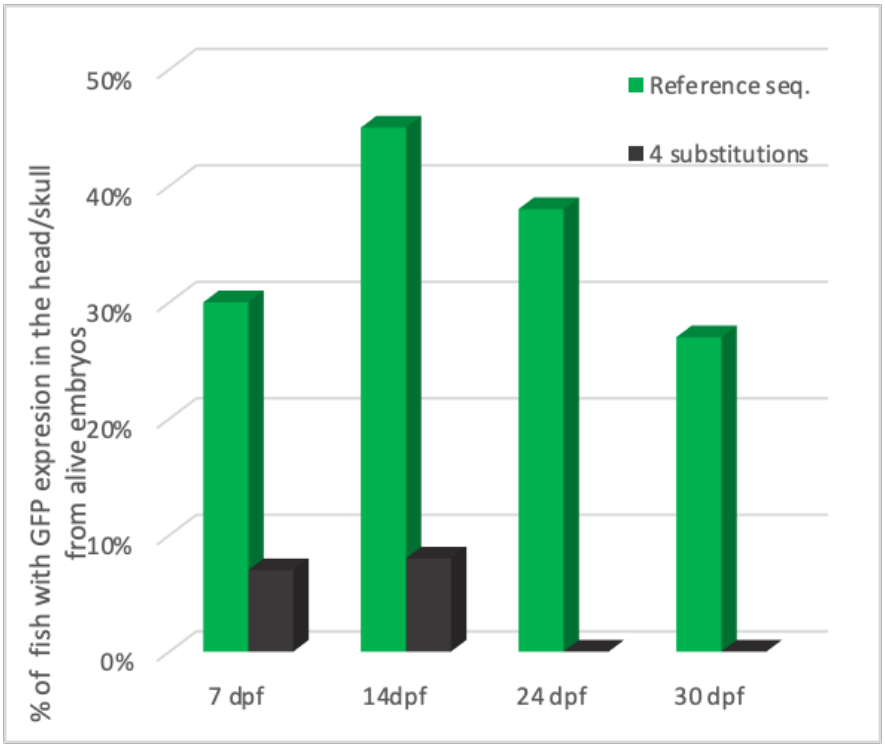
## Figure S6: Zebrafish whole-mount in situ hybridization of dlx5a and dlx6a at 3 days post fertilization.

**A, B** dlx5a is expressed in the pectoral fin, branchial arch, brain, olfactory bulb and otic vesicle.
**C, D** dlx6a is expressed in the branchial arch, brain, olfactory bulb and otic vesicle.



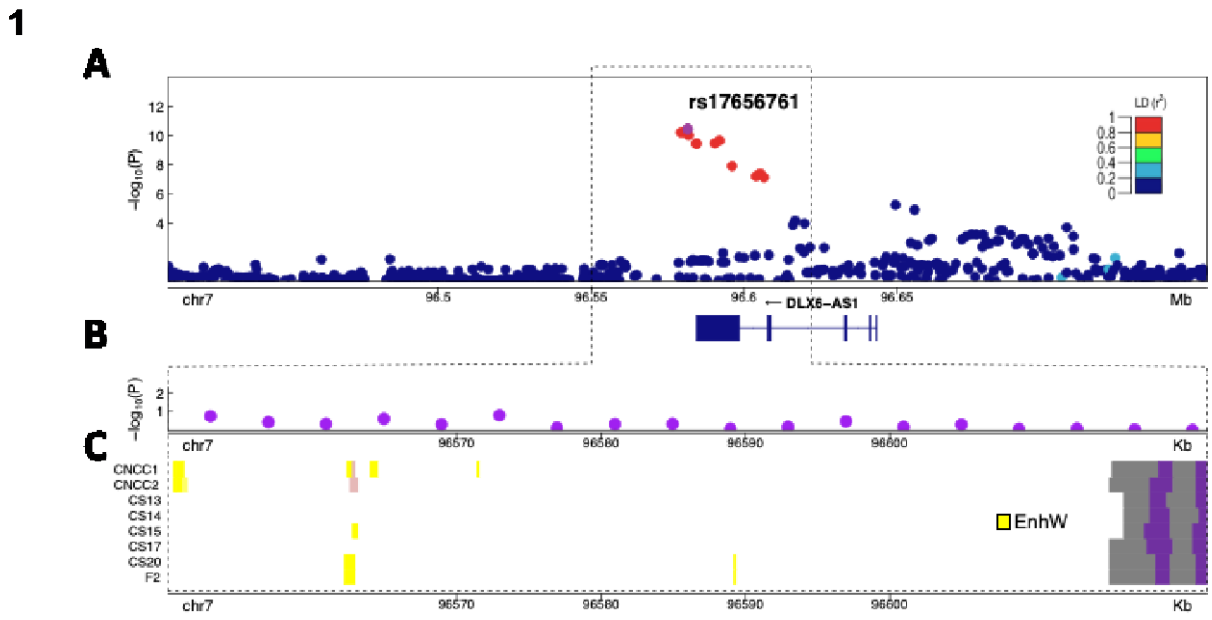
## Figure S7: In vivo activity of eDlx36 control and mutant sequences in zebrafish enhancer assay

Histogram shows the percentage of fish with green florescence protein (GFP) expression in the head/skull in the mutated eDlx36 embryos compared to reference sequence at different developmental stages (days post fertilization, dpf).

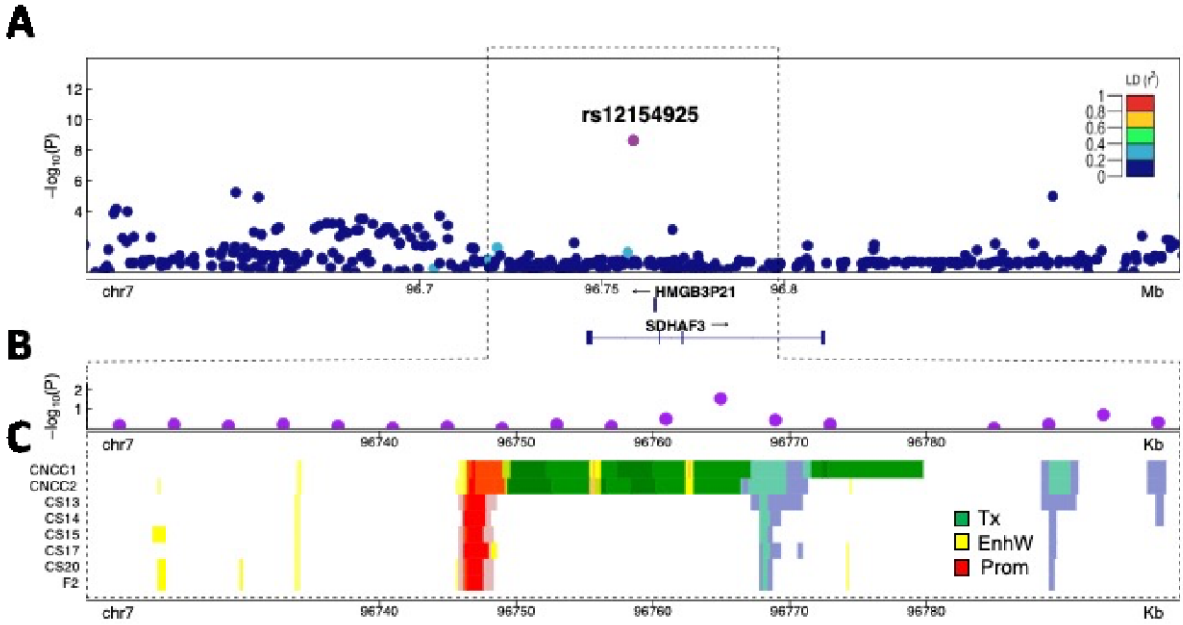


## Figure S8: Annotation of putative craniofacial-specific regulatory elements around rs17656761 (1), rs12154925 (2), rs78353978 (3), rs33863 (4), rs78353978 (5).

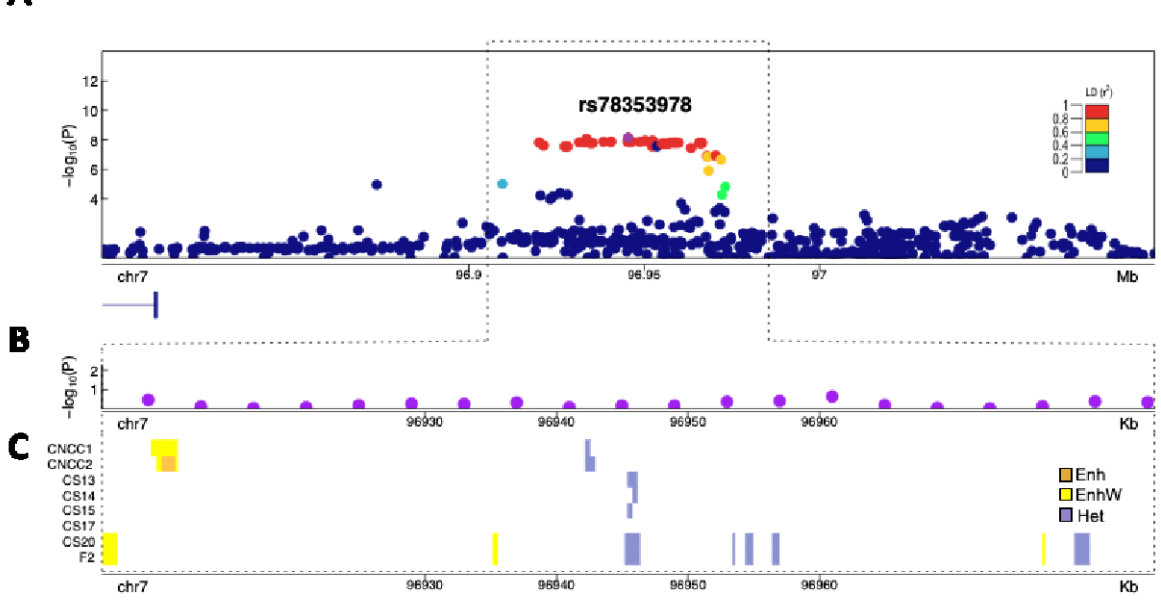
For each signal, the figure shows: ***Panel A*** displays a regional Manhattan plot around each lead variant: The x-axis represents the genomic coordinates of the variants annotated with nearby genes and the y-axis shows −log10 P values for individual SNPs from the European meta-analysis. The lead variant is represented by a purple dot. Pairwise LD (r2) with the lead variant is based on 1000 Genomes phase 3 European reference samples and is described using the color scale given in the legend. We then selected the zoomed region to better evaluate the functional annotation under the high LD block of each signal (gray dashed box). The zoomed-in genomic region is annotated with: ***Panel B***: The y-axis represents the −log10 *P*-values for individual 6kb sliding window from rv-TDT aggregate analysis of rare variants in family-based-study (light purple dots). Each light purple dot is located at the start of each 6kb window. ***Panel C*** shows the craniofacial predicted regulatory elements from Wilderman et al[1] spanning the zoomed genomic region. When present, predicted regulatory elements track the eQTL variants were also displayed and colored by databases as reported in the legend on the right side. Each dot is a SNP which corresponds to a SNP in the zoomed region in *Panel A*. The x-axis represents the genomic location of each SNP, and the y-axis shows the eQTL *P*-value correlated to the expression of the gene reported in each row.



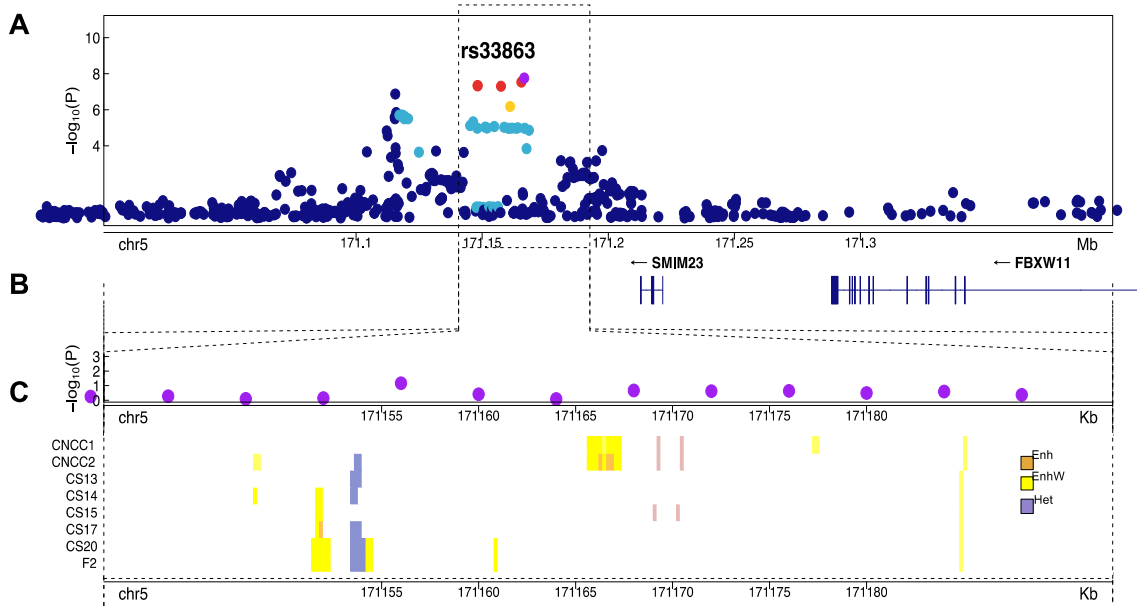
2


A
rs12154925
LD (r²)
1
0.8
0.6
0.4
0.2
0
−log10(P)
12
10
8
6
4
chr7
96.7
96.75
HMGB3P21
96.8
Mb
SDHAF3
B
−log10(P)
2
1
C
chr7
96740
96750
96760
96770
96780
Kb
CNCC1
CNCC2
CS13
CS14
CS15
CS17
CS20
F2
Tx
EnhW
Prom

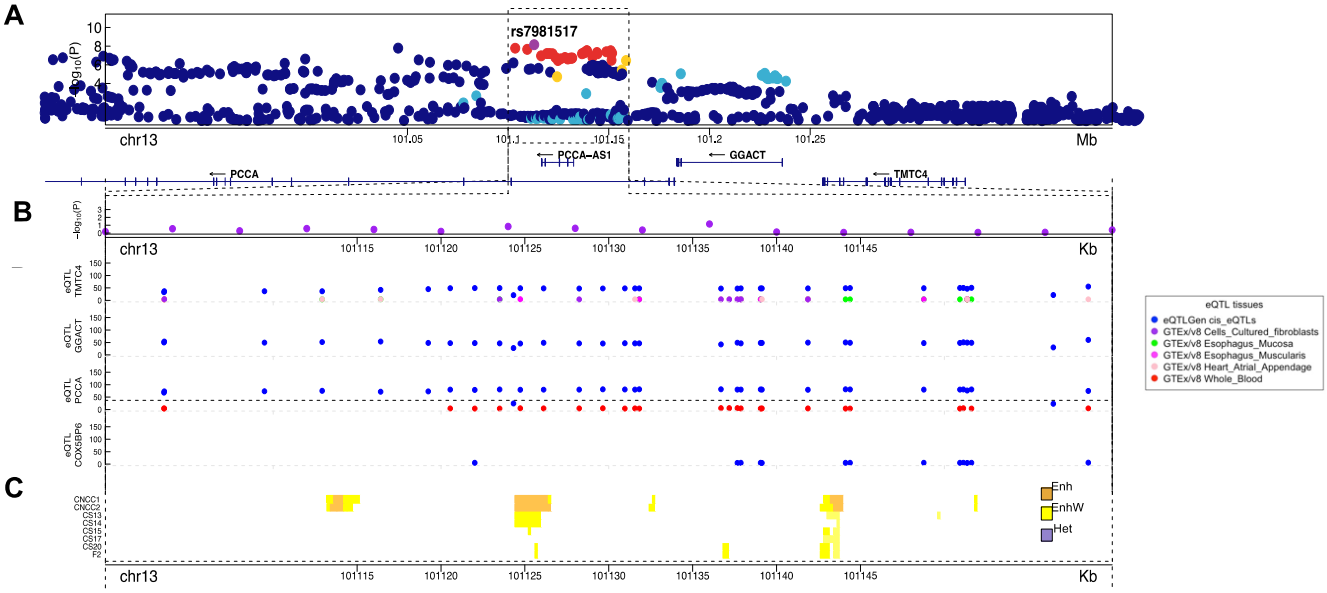
3


A
rs78353978
LD (r²)
1
0.8
0.6
0.4
0.2
0
−log10(P)
12
10
8
6
4
chr7
96.9
96.95
97
Mb
B
−log10(P)
2
1
C
chr7
96930
96940
96950
96960
Kb
CNCC1
CNCC2
CS13
CS14
CS15
CS17
CS20
F2
Enh
EnhW
Het

4

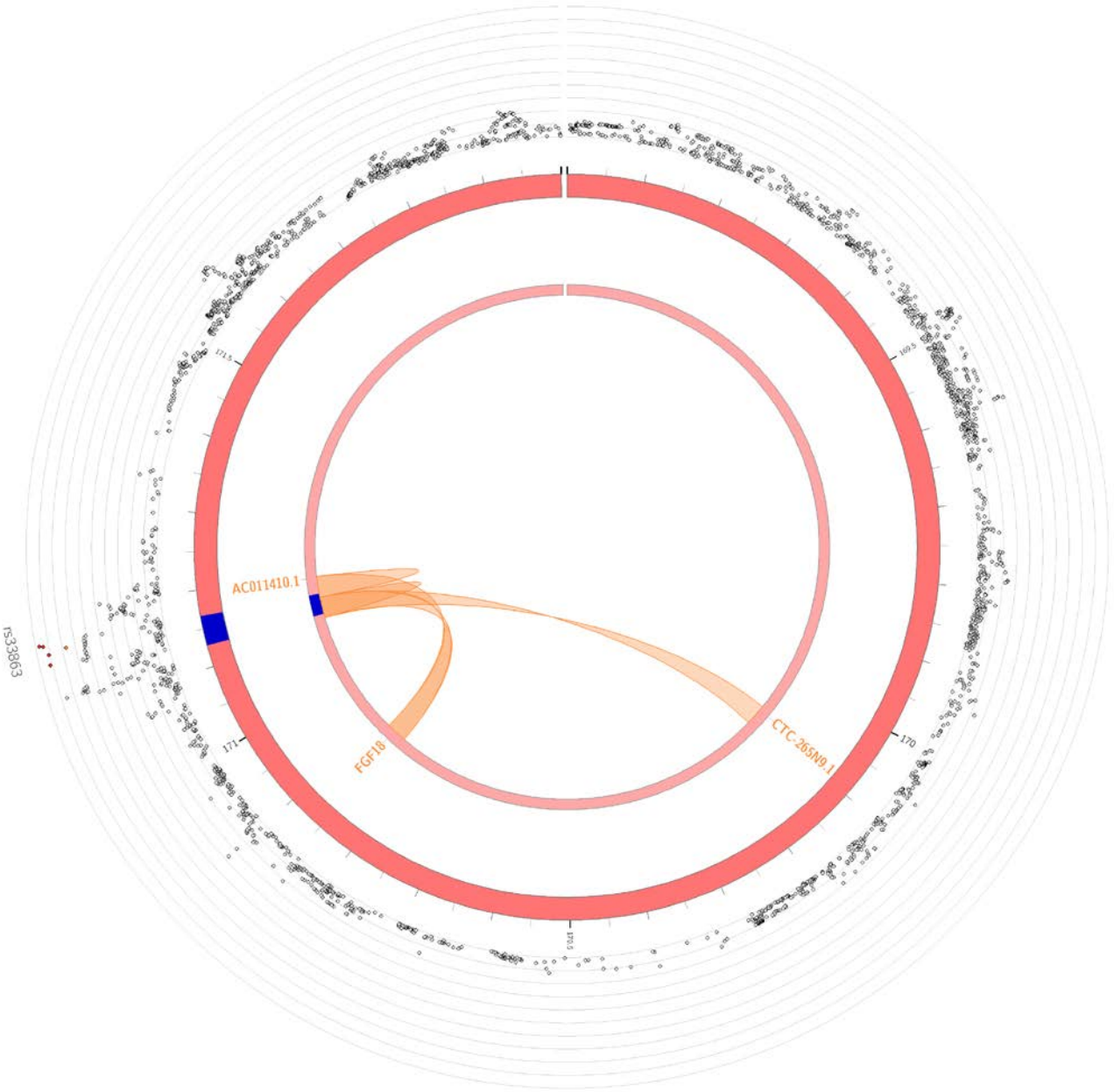

A
rs33863
−log10(P)
chr5
171.1
171.15
171.2
171.25
171.3
Mb
B
← SMIM23
← FBXW11
C
−log10(P)
chr5
171155
171160
171165
171170
171175
171180
Kb
CNCC1
CNCC2
CS13
CS14
CS15
CS17
CS20
F2
Enh
EnhW
Het

5


A
rs7981517
−log10(P)
chr13
101.05
101.1
101.15
101.2
101.25
Mb
← PCCA
← PCCA-AS1
← GGACT
← TMTC4
B
chr13
101115
101120
101125
101130
101135
101140
101145
Kb
eQTL TMTC4
eQTL GGACT
eQTL PCCA
eQTL COX5BP6
C
CNCC1
CNCC2
CS13
CS14
CS15
CS17
CS20
F2
Enh
EnhW
Het
eQTL tissues
eQTLGen cis_eQTLs
GTEx/v8 Cells_Cultured_fibroblasts
GTEx/v8 Esophagus_Mucosa
GTEx/v8 Esophagus_Muscularis
GTEx/v8 Heart_Atrial_Appendage
GTEx/v8 Whole_Blood

## Figure S9: Chromatin interaction analysis of the top hit on chromosome 5 by FUMA.

Circus plot displaying significant chromatin interactions between the target region and other genomic regions (orange arches) as predicted in mesenchymal stem cell lines. Figure was generated in FUMA.



## Table S1: List of sites with the number of contributed samples.

| Site | N Probands | N Parents | N Controls |
|---|---|---|---|
| Department of Pediatrics, UC Davis, CA, USA | 153 | 188 | - |
| The National Birth Defects Prevention Study, USA | 95 | 68 | - |
| MRC Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford, UK | 64 | 88 | - |
| Hospital of Sant Joan de Deu - Universitates de Barcelona, Barcelona, Spain | 49 | 82 | - |
| New York State Congenital Malformations Registry, USA | 33 | - | - |
| Hospital Necker-Enfants maladies, Paris, France | 25 | 20 | - |
| Great Ormond Street Hospital for Children, London, UK | 19 | 26 | - |
| Department of Surgery, Yale University School of Medicine, New Haven, CT, USA | 19 | 28 | - |
| Univ. Hospital Heidelberg, Heidelberg, Germany | 19 | 32 | - |
| University of Utah, Salt Lake City, Utah, USA | 18 | 32 | - |
| Milton S. Hershey Medical Center, Penn State College of Medicine, Philadelphia, PE, USA | 13 | 10 | - |
| Alder Hey Children's Hospital, Liverpool, UK | 10 | 14 | - |
| Birmingham Women's and Children's Hospital, Birmingham, UK | 8 | 14 | - |
| Boston Children's Hospital, Boston, MA, USA | 9 | 8 | - |
| The Craniofacial Center, Medical City Dallas, Dallas, TX, USA | 7 | 6 | - |
| University of Texas Southwestern Medical Center, TX, USA | 5 | 4 | - |
| Lurie Children's Hospital of Chicago, IL, USA | 2 | 4 | - |
| Johns Hopkins University, Baltimore, Maryland, USA | 2 | - | - |
| Iowa Pyloric Stenosis Study, USA | - | - | 166 |
| New York State Birth Defects Surveillance Program, USA | - | - | 112 |
| Icahn School of Medicine, Mount Sinai, NY, USA | - | - | 11 |
| Total | 550 | 624 | 289 |
| Grand Total | | | 1463 |

## Table S2: Clinical characteristics of the study cohorts.

| | | Genetically predicted ethnicity | | | | Proband gender | | Affected suture | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Cohort** | **N** | **AA** | **HISP** | **EUR** | **Others** | **% (N) Male** | **% (N) Female** | **UNK** | **UC** | **UC-L** | **UC-R** | **BC** |
| **Discovery cohort** | | | | | | | | | | | | |
| *Affected probands* | 460 | 5 | 59 | 376 | 20 | 35% (160) | 65% (300) | 19 | 33 | 148 | 204 | 56 |
| Probands with two parents (trio) | 301 | 3 | 37 | 241 | 20 | 35% (106) | 65% (195) | 17 | 11 | 96 | 141 | 36 |
| Probands with one parent (duo) | 85 | 0 | 16 | 69 | - | 32% (27) | 68% (58) | 1 | 4 | 31 | 36 | 13 |
| Single probands | 74 | 2 | 6 | 66 | - | 36% (27) | 64% (47) | 1 | 18 | 21 | 27 | 7 |
| *Unrelated unaffected individuals* | 3376 | - | - | 3376 | - | 85% (2877) | 15% (499) | | | | | |
| **Replication cohort** | | | | | | | | | | | | |
| *Affected probands* | 81 | 4 | 18 | 59 | - | 25% (20) | 75% (61) | 10 | 2 | 26 | 32 | 11 |
| *Unrelated unaffected individuals* | 289 | - | - | 289 | - | 67% (195) | 33% (94) | | | | | |
| **Genome sequenced cohort** | | | | | | | | | | | | |
| *Affected probands** | 89 | 4 | 6 | 77 | 2 | 34% (30) | 66% (59) | 8 | 9 | 24 | 36 | 12 |

*94% overlap with the trios in the discovery cohort. AA = African Americans; HISP = Hispanics; EUR = Europeans; UNK = Unknown; UC = Unicoronal; L= Left; R= Right; BC = Bicoronal.

Table S3: Zebrafish enhancer assay for eDlx34/35/36 at 3 days post fertilization.

| Enhancer | Injected Embryos (N) | Live Embryos (N) | Number (%) embryos positive for Green florescent protein in selected tissues | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Front-nasal | | Branchial arch | | Somatic muscles | | Tail | | Heart | | Notochord | |
| eDLX34 | 150 | 98 | - | - | 3 | 3% | 93 | 95% | - | - | 34 | 35% | - | - |
| eDLX35 | 150 | 107 | 8 | 7% | 34 | 32% | - | - | - | - | 7 | 7% | 36 | 34% |
| eDLX36 | 227 | 151 | 88 | 58% | - | - | - | - | 20 | 13% | - | - | - | - |

## Table S4: Summary statistics of the lead variants under logistic regression model comparing unilateral vs bilateral cases in discovery and replication cohorts.

| Variants | European discovery cohort | | | European replication cohort | | |
|---|---|---|---|---|---|---|
| | P | AF Bicoronal | AF Unicoronal | P | AF Bicoronal | AF Unicoronal |
| rs114264214* | 0.03 | 0.09 | 0.04 | 0.66 | 0.06 | 0.02 |
| rs7981517 | 0.36 | 0.35 | 0.30 | 0.20 | 0.50 | 0.32 |
| rs33863 | 0.36 | 0.49 | 0.43 | 0.11 | 0.31 | 0.50 |
| rs4727341(A) | 0.27 | 0.22 | 0.17 | 0.22 | 0.28 | 0.16 |
| rs17656761 | 0.43 | 0.18 | 0.23 | 0.13 | 0.25 | 0.11 |
| rs12154925 | 0.29 | 0.15 | 0.20 | 0.86 | 0.19 | 0.14 |
| rs78353978 | 0.78 | 0.06 | 0.08 | 0.88 | 0.06 | 0.05 |

AF= allele Frequency; P = logistic regression *P*-value; *rs114264214 is a genome-wide significant variant in the discovery genome-wide association study, but not in the European meta-analysis. Statistical significance threshold after adjustment for multiple testing is *P*-value<0.007.

## Table S5: Summary statistics of the lead variants identified by the European meta-analysis under recessive and dominant models

| Variant | model | Discovery European cohort | | Replication European cohort | |
|---|---|---|---|---|---|
| | | OR (95%CI) | P | OR (95%CI) | P |
| **rs4727341** | *DOM* | 0.45 (0.36,0.57) | 1.90E-11 | 0.51 (0.27,0.93) | 0.03 |
| | *REC* | 0.24 (0.12,0.48) | 5.20E-05 | 0.42 (0.12,1.52) | 0.19 |
| **rs17656761** | *DOM* | 2.12 (1.69,2.67) | 1.10E-10 | 1.37 (0.67,2.78) | 0.39 |
| | *REC* | 2.88 (1.64,5.06) | 2.20E-04 | 1.22 (0.12,12.27) | 0.87 |
| **rs12154925** | *DOM* | 1.86 (1.46,2.37) | 6.00E-07 | 1.96 (0.98,3.92) | 0.06 |
| | *REC* | 4.34 (2.30,8.16) | 5.50E-06 | 2.01 (0.2,20.34) | 0.56 |
| **rs78353978** | *DOM* | 2.73 (1.94,3.83) | 8.20E-09 | 1.82 (0.63,5.25) | 0.27 |
| | *REC* | 4.25 (0.86,21.11) | 7.70E-02 | - | |
| **rs7981517** | *DOM* | 1.68 (1.34,2.11) | 5.90E-06 | 1.85 (1.00,3.43) | 0.05 |
| | *REC* | 2.23 (1.5,3.31) | 7.60E-05 | 4.13 (1.58,10.8) | 0.004 |
| **rs33863** | *DOM* | 1.77 (1.39,2.25) | 3.30E-06 | 2.86 (1.36,6.03) | 0.01 |
| | *REC* | 1.75 (1.31,2.34) | 1.60E-04 | 1.38 (0.64,2.99) | 0.41 |
| **rs114264214*** | *DOM* | 3.57 (2.26,5.65) | 5.00E-08 | 1.97 (0.47,8.20) | 0.35 |
| | *REC* | - | | - | |

OR = Odds Ratio, 95% CI = 95% Confidence Interval; P = *P*-value. -, no homozygote carriers available. *rs114264214 is a genome-wide significant variant in the discovery genome-wide association study, but not in the European meta-analysis.

## Table S6: List of the most enriched pathways in the DisGeNET database identified by MAGMA using the European meta-analysis summary statistics.

| Name | Num Genes | P | P FDRadj |
|---|---|---|---|
| Hypoplastic fingernail | 12 | 7.28E-08 | 0.0007 |
| Serum albumin measurement | 6 | 5.28E-07 | 0.003 |
| Triangular mouth | 6 | 9.19E-07 | 0.003 |
| Bifid tongue | 11 | 3.91E-06 | 0.01 |
| Cleft tongue | 11 | 3.91E-06 | 0.01 |
| Aplasia/Hypoplasia involving the metacarpal bones | 5 | 4.38E-06 | 0.01 |
| Delayed eruption of permanent teeth | 8 | 8.90E-06 | 0.01 |
| Broad hallux phalanx | 10 | 1.22E-05 | 0.02 |
| Accessory kidney | 5 | 3.35E-05 | 0.04 |
| Obesity | 1750 | 4.81E-05 | 0.05 |
| Neurological disability | 6 | 5.02E-05 | 0.05 |
| Short middle phalanx of finger | 15 | 6.36E-05 | 0.05 |
| Brachydactyly type A3 | 12 | 7.25E-05 | 0.05 |
| Short middle phalanx of the 5th finger | 12 | 7.25E-05 | 0.05 |
| Pseudohypoparathyroidism | 20 | 7.60E-05 | 0.05 |

Name = Pathway name; Num Genes = Number of genes overlapping between GWAS dataset and pathway gene set, P = Enrichment P-value; P FDRadj = false discovery rate-adjusted P-value <0.05.

## Table S7: Summary statistics of the Transmission Disequilibrium Test in 301 multi-ethnic trios for the lead variants identified in the European meta-analysis.

| SNP (Allele) | Nearest Gene | TDT (N=301 Trios) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | OR (95%CI) | P | AF AFF | AF UNAFF | T:U_PAT | P_PAT | T:U_MAT | P_MAT | P_POO |
| rs4727341 (A) | *SEM1* | 0.45 (0.33-0.6) | 2.08E-08 | 0.18 | 0.25 | 29.5:78.5 | 2.42E-06 | 36.5:69.5 | 0.001 | 0.26 |
| rs17656761 (A) | *DLX6-AS1* | 1.61 (1.15-2.24) | 4.80E-03 | 0.18 | 0.16 | 40:26 | 0.085 | 50:30 | 0.025 | 0.81 |
| rs12154925 (T) | *SDHAF3* | 1.6 (1.16-2.23) | 5.50E-03 | 0.17 | 0.155 | 46.5:29.5 | 0.05 | 46.5:28.5 | 0.037 | 0.92 |
| rs78353978 (A) | *DLX5/TAC1* | 3.0 (1.64-5.49) | 1.80E-04 | 0.07 | 0.046 | 24:6 | 0.001 | 18:8 | 0.049 | 0.36 |
| rs7981517 (A) | *PCCA* | 1.48 (1.11-1.96) | 6.90E-03 | 0.27 | 0.24 | 50.5:39.5 | 0.246 | 67.5:40.5 | 0.009 | 0.36 |
| rs33863 (A) | *SM1M23 / FGF18* | 1.76 (1.37-2.27) | 9.50E-06 | 0.45 | 0.38 | 81:46 | 0.002 | 83:47 | 0.001 | 0.99 |
| rs114264214* (A) | *PLEKHA6* | 2.86 (1.21-6.76) | 0.01 | 0.035 | 0.024 | 10.5:2.5 | 0.027 | 9.5:4.5 | 0.181 | 0.45 |

SNP = Single Nucleotide Polymorphism; AF AFF = Allele Frequency in affected probands; AF UNAFF = Allele Frequency in unaffected controls;; TDT = transmission disequilibrium test; OR = Odds Ratio, 95% CI = 95% Confidence Interval; T:U-PAT = Paternal transmitted: un-transmitted counts; P_PAT = Paternal chi-squared test; T:U-MAT = Maternal transmitted: un-transmitted counts; P_MAT = Maternal chi-squared test; P_POO = Asymptotic p-value for parent-of-origin test. *rs114264214 is a genome-wide significant variant in the discovery genome-wide association study, but not in the European meta-analysis.

## Table S8: Summary statistics of the trans-ethnic meta-analysis for the lead variants identified in the European meta-analysis.

| SNP | Gene | European meta-analysis | | Hispanic (#case=77, # gnomAD controls=838) | | | | Trans-ethnic meta-analysis | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | OR (95%CI) | *P* | OR (95%CI) | *P* | AF AFF | AF UNAFF | OR (95%CI) | *P* | HP |
| rs4727341 | *SEM1* | 0.49 (0.29-0.69) | 2.69E-13 | 0.64 (0.4-0.98) | 0.04 | 0.20 | 0.29 | 0.50 (0.31-0.68) | 1.82E-13 | 0.72 |
| rs17656761 | *DLX6-AS1* | 1.89 (1.69-2.09) | 3.75E-11 | 1.53 (0.78-2.84) | 0.17 | 0.09 | 0.07 | 1.88 (1.69-2.06) | 3.15E-11 | 0.46 |
| rs12154925 | *SDHAF3* | 1.86 (1.66-2.06) | 2.41E-09 | NA | NA | NA | 0.09 | NA | NA | NA |
| rs78353978 | *DLX5/TAC1* | 2.50 (2.19-2.81) | 6.99E-09 | 1.96 (0.62-5.31) | 0.15 | 0.039 | 0.02 | 2.49 (2.18-2.80) | 5.81E-09 | 0.8 |
| rs7981517 | *PCCA* | 1.64 (1.47-1.79) | 7.13E-09 | 1.47 (0.98-2.19) | 0.05 | 0.29 | 0.22 | 1.63 (1.46-1.79) | 4.43E-09 | 0.75 |
| rs33863 | *SM1M23 / FGF18* | 1.55 (1.39 1.71) | 1.76E-08 | 0.99 (0.69-1.44) | 1 | 0.38 | 0.38 | 1.52 (1.37-1.67) | 3.67E-08 | 0.45 |
| rs114264214* | *PLEKHA6* | 3.37 (2.92-3.82) | 7.32E-08 | 1.00 (0.11-4.66) | 1 | 0.01 | 0.01 | 3.31 (2.87-3.75) | 9.14E-08 | 0.59 |

SNP = Single Nucleotide Polymorphism; OR = Odds Ratio; 95% CI = 95% Confidence Interval; P = *P*-value; AF AFF = Allele Frequency in affected probands; AF UNAFF = Allele Frequency in unaffected controls; HP = heterogeneity *P*-value; *rs114264214 is a genome-wide significant variant in the discovery genome-wide association study, but not in the European meta-analysis.

## Table S9: Multivariate analysis for the lead variants identified in the European meta-analysis.

| SNP | OR (95%CI) | *P* |
|---|---|---|
| rs4727341 | 1.93 (1.58-2.35) | 6.40E-11 |
| rs17656761 | 1.7 (1.4-2.08) | 1.60E-07 |
| rs78353978 | 1.98 (1.39-2.81) | 1.30E-04 |
| rs12154925 | 1.46 (1.16-1.83) | 1.10E-03 |
| rs33863 | 1.59 (1.36-1.85) | 7.60E-09 |
| rs7981517 | 1.49 (1.26-1.77) | 3.90E-06 |

For this analysis, European discovery and validation samples were jointly analyzed in a single case/control cohort. SNP = Single Nucleotide Polymorphism; OR = Odds Ratio; OR = Odds Ratio; 95% CI = 95% Confidence Interval; P = *P*-value.

## Table S10: List of SNPs and weights to calculate the PRS

| Chromosome | Position | A1 | A2 | OR | SE |
|---|---|---|---|---|---|
| 1 | 19108549 | T | C | 1.677 | 0.1146 |
| 1 | 92434589 | G | A | 2.46 | 0.2177 |
| 1 | 147825769 | A | G | 1.869 | 0.1492 |
| 1 | 204310366 | A | G | 3.572 | 0.2335 |
| 1 | 218760614 | T | C | 1.472 | 0.08722 |
| 2 | 2733090 | A | G | 2.92 | 0.2441 |
| 2 | 6097879 | C | A | 1.867 | 0.1161 |
| 2 | 121999108 | T | C | 0.5521 | 0.1364 |
| 2 | 211765267 | C | A | 1.479 | 0.09018 |
| 3 | 39654502 | T | G | 1.445 | 0.09008 |
| 3 | 54034087 | G | A | 1.823 | 0.1467 |
| 3 | 147950334 | C | T | 0.6374 | 0.1033 |
| 3 | 181938437 | C | A | 0.6637 | 0.09127 |
| 4 | 2719538 | A | G | 1.399 | 0.0828 |
| 4 | 44308206 | G | A | 1.759 | 0.1376 |
| 4 | 95784853 | T | C | 1.568 | 0.1062 |
| 4 | 180655625 | T | C | 1.886 | 0.1495 |
| 5 | 1763681 | G | A | 1.498 | 0.08191 |
| 5 | 16543741 | T | C | 2.324 | 0.1992 |
| 5 | 18215931 | C | T | 3.16 | 0.2273 |
| 5 | 125008619 | C | T | 0.6608 | 0.08624 |
| 5 | 145152278 | A | G | 1.78 | 0.1235 |
| 5 | 159940172 | A | G | 2.932 | 0.2631 |
| 5 | 171166685 | A | G | 1.532 | 0.08154 |
| 5 | 173816283 | G | T | 0.6913 | 0.08627 |
| 5 | 177753869 | C | T | 2.324 | 0.207 |
| 6 | 14991293 | T | C | 1.447 | 0.089 |
| 6 | 37800653 | A | G | 3.615 | 0.2795 |
| 6 | 70823287 | C | T | 1.877 | 0.1478 |
| 6 | 133666904 | C | A | 0.6505 | 0.09449 |
| 6 | 133739242 | G | A | 2.585 | 0.2236 |
| 6 | 133788110 | A | C | 1.865 | 0.1468 |
| 6 | 162243368 | C | T | 1.734 | 0.1151 |
| 7 | 18622929 | G | T | 1.397 | 0.08071 |
| 7 | 27757332 | T | C | 2.857 | 0.2547 |
| 7 | 45105133 | G | A | 1.605 | 0.1145 |
| 7 | 46311048 | A | G | 3.128 | 0.2356 |

| | | | | | |
|---|---|---|---|---|---|
| 7 | 95889706 | C | T | 1.386 | 0.07956 |
| 7 | 96198615 | A | G | 0.4809 | 0.1031 |
| 7 | 96581553 | A | G | 1.957 | 0.09893 |
| 7 | 96649522 | A | C | 1.449 | 0.08101 |
| 7 | 96758550 | T | G | 1.869 | 0.1085 |
| 7 | 96928091 | A | G | 1.382 | 0.07882 |
| 7 | 96945446 | A | G | 2.58 | 0.1629 |
| 7 | 101226020 | C | T | 1.424 | 0.08167 |
| 8 | 265337 | C | T | 0.6986 | 0.0879 |
| 8 | 5126290 | C | T | 1.388 | 0.08027 |
| 8 | 11067792 | A | C | 1.925 | 0.1354 |
| 8 | 36531654 | C | T | 3.312 | 0.2801 |
| 8 | 117828881 | A | G | 2.63 | 0.2315 |
| 8 | 136212232 | C | T | 1.414 | 0.08216 |
| 9 | 12052324 | T | C | 2.72 | 0.2344 |
| 9 | 94525596 | T | C | 2.075 | 0.1789 |
| 9 | 95689628 | A | G | 1.429 | 0.07974 |
| 9 | 106627955 | A | C | 1.609 | 0.1042 |
| 9 | 109529338 | A | G | 1.552 | 0.106 |
| 10 | 12591916 | G | A | 1.415 | 0.08137 |
| 10 | 65346300 | G | A | 1.831 | 0.1424 |
| 10 | 78587491 | A | G | 1.717 | 0.1258 |
| 10 | 79893854 | T | G | 2.947 | 0.2664 |
| 11 | 19954112 | G | A | 0.7046 | 0.08497 |
| 11 | 44354066 | T | C | 2.12 | 0.1808 |
| 11 | 99871070 | A | C | 2.884 | 0.2028 |
| 11 | 132752029 | C | T | 0.6535 | 0.1007 |
| 12 | 3129932 | T | C | 2.482 | 0.216 |
| 12 | 14096896 | C | T | 2.977 | 0.2605 |
| 12 | 16466341 | A | C | 2.847 | 0.2443 |
| 12 | 32210238 | A | G | 2.347 | 0.2078 |
| 12 | 69869704 | T | C | 2.034 | 0.1716 |
| 12 | 82027471 | A | G | 1.739 | 0.1355 |
| 12 | 94647066 | A | G | 1.427 | 0.08026 |
| 12 | 113269597 | C | T | 1.995 | 0.1631 |
| 12 | 120271337 | C | T | 2.569 | 0.2233 |
| 12 | 126174476 | T | C | 2.69 | 0.1975 |
| 12 | 128276516 | A | G | 1.608 | 0.1101 |
| 12 | 130907481 | T | C | 2.078 | 0.1721 |

| | | | | | |
|---|---|---|---|---|---|
| 13 | 48123136 | A | G | 2.92 | 0.2356 |
| 13 | 98735995 | C | T | 1.395 | 0.0814 |
| 13 | 100798391 | T | C | 1.507 | 0.08088 |
| 13 | 101055584 | T | C | 0.597 | 0.09756 |
| 13 | 101156493 | A | G | 1.428 | 0.08311 |
| 14 | 42446967 | G | A | 0.6832 | 0.09084 |
| 14 | 52324354 | A | G | 1.48 | 0.09519 |
| 14 | 98888376 | C | T | 1.501 | 0.0891 |
| 15 | 41855403 | T | C | 2.428 | 0.2138 |
| 15 | 56276463 | T | C | 2.166 | 0.1812 |
| 15 | 91054215 | A | G | 1.588 | 0.1115 |
| 15 | 100665168 | A | G | 3.245 | 0.2864 |
| 16 | 26137926 | G | A | 1.728 | 0.1187 |
| 16 | 65700767 | C | T | 2.173 | 0.1645 |
| 16 | 72631201 | T | C | 2.201 | 0.1938 |
| 16 | 72916878 | G | T | 1.887 | 0.1324 |
| 16 | 73030532 | T | G | 1.908 | 0.1506 |
| 16 | 73202685 | T | C | 2.96 | 0.2427 |
| 16 | 80665731 | G | A | 2.139 | 0.1793 |
| 16 | 81951937 | T | G | 0.6874 | 0.08654 |
| 17 | 37101380 | T | C | 1.723 | 0.1241 |
| 17 | 58760784 | T | C | 1.925 | 0.1612 |
| 17 | 76969183 | A | G | 2.556 | 0.2249 |
| 19 | 42764352 | C | T | 1.707 | 0.1302 |
| 20 | 7118592 | C | T | 1.394 | 0.08171 |
| 20 | 61639750 | G | A | 2.752 | 0.2386 |
| 21 | 39714640 | A | G | 2.405 | 0.2111 |
| 21 | 45106294 | A | C | 2.444 | 0.2163 |
| 21 | 47491264 | A | G | 1.5 | 0.09724 |
| 22 | 19956863 | A | G | 2.465 | 0.1948 |

A1 = text allele; A2= reference allele; OR = Odds Ratio; SE=Standard Error

## Table S11: Relationship between the top associated genes examined in the network analysis

| Term | Library | Gene 1 | Gene 2 | Gene 3 |
|---|---|---|---|---|
| Isolated split hand/foot malformation ORPHA:2440 | Orphanet | *SEM1* | *DLX5* | *DLX6* |
| S-(3-hydroxypropyl)cysteine N-acetate | PheGenI | *PCCA* | *SMIM23* | |
| Heel bone mineral density | GWAS Catalog | *SEM1* | *DLX5* | *DLX6-AS1* |
| Oligodactyly | DisGeNET | *SEM1* | *DLX5* | *DLX6* |
| Congenital Foot Deformity | DisGeNET | *SEM1* | *DLX5* | *DLX6* |
| Split hand foot malformation | Rare Diseases GeneRIF | *SEM1* | *DLX5* | *DLX6* |
| POU3F4 | ARCHS4 TFs Coexp | *PCCA* | *DLX5* | |
| YY1 in ECC-1 | ENCODE | *PCCA* | *DLX5* | |
| ZBTB7A in ECC-1 | ENCODE | *PCCA* | *DLX5* | |
| BRCA1 in H1-hESC | ENCODE | *PCCA* | *DLX5* | |
| CEBPB in H1-hESC | ENCODE | *PCCA* | *DLX5* | |
| TCF12 in ECC-1 | ENCODE | *PCCA* | *DLX6* | |

Table S12: Number of variants (single nucleotide variants and insertions/deletions) from the genome-sequencing analysis of 89 probands located within the susceptibility genes on chromosome 7.

| Variant type | Chromosome 7 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **SLC25A13** | **C7orf76** | **RP11-682N22.1** | **SEM1** | **MARK2P10** | **DLX5** | **DLX6** | **DLX6-AS1** | **SDHAF3** | **HMGB3P21** |
| **Total #variants** | **459** | **28** | **53** | **455** | **0** | **40** | **1** | **140** | **103** | **0** |
| Exonic | 1 | 1 | 0 | 2 | | 2 | 1 | 5 | 2 | |
| missense/LOF | 0 | | na | 1 | | 1 | 1 | na | 2 | |
| missense/LOF & damaging | 0 | | na | 0 | | 0 | 0 | na | 0 | |
| splicing | 0 | | 0 | 0 | | 0 | 0 | 0 | 0 | |
| UTR3 | 2 | 3 | 0 | 3 | | 1 | 0 | 0 | 3 | |
| UTR5 | 1 | 0 | 0 | 1 | | 1 | 0 | 0 | 2 | |
| intronic | 453 | 24 | 53 | 446 | | 16 | 0 | 133 | 93 | |
| upstream | 2 | 0 | 0 | 1 | | 10 | 0 | 1 | 1 | |
| downstream | 0 | 0 | 0 | 2 | | 10 | 0 | 1 | 2 | |
| **Novel #variants** | **42** | **1** | **7** | **46** | **0** | **7** | **0** | **20** | **10** | **0** |
| Exonic | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | |
| missense/LOF | na | na | na | na | na | na | na | na | na | na |
| missense/LOF & damaging | na | na | na | na | na | na | na | na | na | na |
| UTR3 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | |
| UTR5 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | |
| intronic | 41 | 1 | 7 | 46 | | 5 | 0 | 20 | 10 | |
| upstream | 1 | 0 | 0 | 0 | | 1 | 0 | 0 | 0 | |
| downstream | 0 | 0 | 0 | 0 | | 1 | | | 0 | |
| gene length (Kb) | 200 | 15 | 42 | 220 | 1 | 4 | 5 | 63 | 64 | 0.5 |

The reported functional classes of the variants were based on annotation performed by ANNOVAR; LOF' is defined as a loss of function variant. 'Splicing' is defined for a variant that is within 2-bp away from an exon/intron boundary. 'Upstream' and 'downstream' are defined as 1-Kb away from transcription start site or transcription end site. na = not present.

Table S13: List of exonic variants identified by genome sequencing in the genes within the lead loci in 89 probands.

| SNP ID | Chr | Position | R | A | AC | Functional annotation based on refGene | | | Prediction of damaging effect | | | | TDT test | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Gene name | Location | Exonic function | REVEL score | MetaSVM score | MetaSVM prediction | CADD Dscore | AF parents | AF probands | P-value |
| rs150911562 | chr5 | 171456729 | C | T | 1 | FGF18 | exon | nonsynonymous SNV | 0.106 | -0.855 | T | 20.8 | 0.003 | 0.006 | 0.32 |
| rs34347344 | chr5 | 171456730 | G | A | 10 | FGF18 | exon | synonymous SNV | . | . | . | . | 0.062 | 0.056 | 0.67 |
| rs10037031 | chr5 | 171790231 | C | T | 1 | SMIM23 | exon | nonsynonymous SNV | 0.054 | -1.074 | T | 25 | 0.008 | 0.006 | 0.56 |
| rs61739670 | chr5 | 171868665 | G | C | 1 | FBXW11 | exon | synonymous SNV | . | . | . | . | 0.003 | 0.006 | 0.32 |
| rs10475991 | chr5 | 171878034 | A | G | 3 | FBXW11 | exon | synonymous SNV | . | . | . | . | 0.014 | 0.017 | 0.65 |
| rs2301629 | chr7 | 96171508 | T | C | 83 | SLC25A13 | exon | synonymous SNV | . | . | . | . | 0.449 | 0.466 | 0.51 |
| rs78670506 | chr7 | 96486315 | G | C | 3 | SEM1 | exon | nonsynonymous SNV | 0.072 | -1.032 | T | 1.69 | 0.025 | 0.017 | 0.32 |
| rs4733 | chr7 | 96709749 | C | T | 2 | SEM1 | exon | synonymous SNV | . | . | . | . | 0.020 | 0.011 | 0.26 |
| rs147783529 | chr7 | 96968909 | A | T | 1 | DLX6-AS1 | ncRNA exon | . | . | . | . | . | 0.003 | 0.006 | 0.32 |
| rs2189772 | chr7 | 96969056 | G | A | 102 | DLX6-AS1 | ncRNA exon | . | . | . | . | . | 0.475 | 0.427 | 0.08 |
| rs75626318 | chr7 | 96978416 | G | A | 1 | DLX6-AS1 | ncRNA exon | . | . | . | . | . | 0.003 | 0.006 | 0.32 |
| rs2214644 | chr7 | 96978757 | C | T | 100 | DLX6-AS1 | ncRNA exon | . | . | . | . | . | 0.480 | 0.438 | 0.13 |
| rs569298699 | chr7 | 96978979 | AGCTCCCCT | A | 2 | DLX6-AS1 | ncRNA exon | . | . | . | . | . | na | na | na |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | TTCCTAT | | | | | | | | | | | | |
| rs201772433 | chr7 | 97009914 | C | T | 1 | DLX6 | exon | nonsynonymous SNV | 0.4 | 0.464 | D | 25.8 | 0.003 | 0.006 | 0.32 |
| rs35273378 | chr7 | 97020904 | G | T | 1 | DLX5 | exon | nonsynonymous SNV | 0.282 | -0.449 | T | 15.67 | 0.006 | 0.006 | 1 |
| rs61753628 | chr7 | 97024372 | G | C | 1 | DLX5 | exon | synonymous SNV | . | . | . | . | 0.003 | 0.006 | 0.32 |
| rs199949264 | chr7 | 97117733 | C | T | 1 | SDHAF3 | exon | nonsynonymous SNV | 0.07 | -1.053 | T | 23.6 | 0.003 | 0.006 | 0.32 |
| rs62624461 | chr7 | 97117880 | T | C | 5 | SDHAF3 | exon | nonsynonymous SNV | 0.538 | -0.735 | T | 34 | 0.028 | 0.028 | 1 |
| rs538229 | chr13 | 100235868 | A | G | 25 | PCCA | exon | synonymous SNV | . | . | . | . | 0.143 | 0.140 | 0.88 |
| rs41281120 | chr13 | 100302950 | A | G | 1 | PCCA | exon | synonymous SNV | . | . | . | . | 0.003 | 0.006 | 0.32 |
| rs35719359 | chr13 | 100309902 | A | G | 8 | PCCA | exon | nonsynonymous SNV | 0.399 | -0.558 | T | 19.1 | 0.051 | 0.045 | 0.64 |
| rs61749895 | chr13 | 100368479 | G | T | 3 | PCCA | exon | nonsynonymous SNV | 0.495 | -0.954 | T | 0.18 | 0.011 | 0.017 | 0.32 |
| rs192063381 | chr13 | 100466146 | A | G | 1 | PCCA-AS1 | ncRNA exon | . | . | . | . | . | 0.003 | 0.006 | 0.32 |
| rs755192518 | chr13 | 100473215 | C | T | 1 | PCCA-AS1 | ncRNA exon | . | . | . | . | . | 0.003 | 0.006 | 0.32 |
| rs3759477 | chr13 | 100477390 | A | G | 53 | PCCA-AS1 | ncRNA exon | . | . | . | . | . | 0.250 | 0.298 | 0.03 |
| rs925574891 | chr13 | 100477395 | G | C | 1 | PCCA-AS1 | ncRNA exon | . | . | . | . | . | 0.003 | 0.006 | 0.32 |
| rs61730956 | chr13 | 100606369 | T | C | 1 | TMTC4 | exon | nonsynonymous SNV | 0.652 | -0.751 | T | 24 | 0.006 | 0.006 | 1 |
| rs55952539 | chr13 | 100614392 | C | T | 2 | TMTC4 | exon | synonymous SNV | . | . | . | . | 0.008 | 0.011 | 0.56 |
| rs749518908 | chr13 | 100625620 | C | A | 1 | TMTC4 | exon | nonsynonymous SNV | 0.241 | -0.637 | T | 26.8 | 0.003 | 0.006 | 0.32 |
| rs946837 | chr13 | 100635086 | C | T | 49 | TMTC4 | exon | nonsynonymous SNV | 0.02 | -0.928 | T | 2.07 | 0.287 | 0.275 | 0.64 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs946838 | chr13 | 100635111 | T | C | 143 | TMTC4 | exon | synonymous SNV | . | . | . | . | 0.202 | 0.197 | 0.8 |
| rs2297943 | chr13 | 100635150 | G | A | 57 | TMTC4 | exon | synonymous SNV | . | . | . | . | 0.309 | 0.320 | 0.63 |
| rs141440152 | chr13 | 100637657 | G | A | 1 | TMTC4 | exon | synonymous SNV | . | . | . | . | 0.003 | 0.006 | 0.32 |
| rs17579147 | chr13 | 100656400 | G | A | 3 | TMTC4 | exon | synonymous SNV | . | . | . | . | 0.020 | 0.017 | 0.71 |
| rs61746911 | chr13 | 100668725 | T | C | 1 | TMTC4 | exon | nonsynonymous SNV | 0.035 | -1.028 | T | 0.002 | 0.003 | 0.006 | 0.32 |
| rs374326272 | chr13 | 100668767 | C | T | 1 | TMTC4 | exon | nonsynonymous SNV | 0.03 | -1.037 | T | 2.517 | 0.003 | 0.006 | 0.32 |

R = reference allele, A= alternative allele; AC= allele count; AF = allele frequency; TDT, transmission disequilibrium test.

## Table S14: Number of variants (single nucleotide variants and insertions/deletions) from the 89 genome-sequenced probands across multiple ethnicities located within the susceptibility loci on chromosome 5 and 13.

| | chr13 | | | | chr5 | | |
|---|---|---|---|---|---|---|---|
| **Variants** | **PCCA** | **PCCA-AS1** | **GGACT** | **TMTC4** | **SMIM23** | **FBXW11** | **FGF18** |
| **Total #variants** | **854** | **52** | **134** | **303** | **14** | **293** | **113** |
| Exonic | 4 | 4 | 0 | 10 | 1 | 2 | 2 |
| *missense/LOF* | 2 | na | na | 5 | 1 | 0 | 1 |
| *missense/LOF & damaging* | 0 | na | na | 0 | 0 | 0 | 0 |
| *splicing* | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| UTR3 | 3 | 0 | 12 | 2 | 0 | 4 | 2 |
| UTR5 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| intronic | 847 | 48 | 116 | 273 | 8 | 277 | 100 |
| upstream | 0 | 0 | 5 | 12 | 2 | 1 | 7 |
| downstream | 0 | 0 | 0 | 4 | 3 | 8 | 1 |
| **Novel #variants** | **96** | **3** | **9** | **29** | **1** | **42** | **9** |
| Exonic | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| missense/LOF | na | na | na | na | na | na | na |
| missense/LOF & damaging | na | na | na | na | na | na | na |
| UTR3 | 0 | 0 | 2 | 0 | 0 | 2 | 1 |
| UTR5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| intronic | 96 | 3 | 7 | 26 | 0 | 40 | 8 |
| upstream | 0 | 0 | 0 | 2 | 1 | 0 | 0 |
| downstream | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| gene length (Kb) | 440 | 17 | 59 | 71 | 24 | 142 | 37 |

The reported functional classes of the variants were based on annotation performed by ANNOVAR; LOF' is defined as a loss of function variant. 'Splicing' is defined for a variant that is within 2-bp away from an exon/intron boundary. 'Upstream' and 'downstream' are defined as 1-Kb away from transcription start site or transcription end site. na = not present.

**Table S15**: List of exonic variants located in *TCF12* and identified by genome sequencing of 89 probands.

| Chr | Position | Ref allele | Alt allele | SNPID | Exonic Function | MetaSVM | REVEL | CADD |
|---|---|---|---|---|---|---|---|---|
| chr15 | 57232784 | G | A | rs12442879 | nonsynonymous SNV | tolerated | 0.24 | 10 |
| chr15 | 57251406 | AACTC | A | . | frameshift deletion | . | | |
| chr15 | 57253416 | A | G | rs753036829 | nonsynonymous SNV | tolerated | 0.02 | 1.6 |
| chr15 | 57262146 | T | G | rs36060670 | nonsynonymous SNV | tolerated | 0.22 | 33 |
| chr15 | 57263138 | GT | G | . | frameshift deletion | . | | |
| chr15 | 57273115 | C | T | rs754118933 | nonsynonymous SNV | damaging | 0.97 | 35 |
| chr15 | 57273215 | T | C | . | nonsynonymous SNV | damaging | 0.97 | 28 |
| chr15 | 57273235 | A | T | . | nonsynonymous SNV | damaging | 0.96 | 29 |

## Table S16: Craniofacial specific enhancer candidates in the *SEM1-DLX5/6* locus.

| Elements | | hg 19 | | | | human CF tissue - enhancers | | | | | | Conservation level |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Num.** | **Name** | **Chr** | **Start** | **End** | **Length** | **CS 13** | **CS 14** | **CS 15** | **CS17** | **CS20** | **CNCC** | |
| 1 | eDlx30 | chr7 | 96,733,796 | 96,734,203 | 408 | Dnase | Dnase | Dnase | Dnase | Dnase | | Mammalian |
| 2 | eDlx31 | chr7 | 96,694,201 | 96,694,400 | 200 | Dnase | Dnase | Dnase | Dnase | Dnase | | Mammalian |
| 3 | eDlx32 | chr7 | 96,415,220 | 96,416,581 | 1362 | Enh | Enh | Enh | Enh | Enh | Enh | Mammalian |
| 4 | eDlx33 | chr7 | 96,337,200 | 96,338,000 | 801 | TxEhnW | TxEhnW | TxEhnW | TxEhnW | TxEhnW | TxEhnW | Mammalian |
| 5 | eDlx34 | chr7 | 96,227,969 | 96,229,202 | 1,234 | Enh | Enh | Enh | | | Enh | Mammalian |
| 6 | eDlx35 | chr7 | 96,225,003 | 96,226,583 | 1,581 | Enh | Enh | Enh | Enh | Enh | Enh | Fish |
| 7 | eDlx36 | chr7 | 96,219,791 | 96,222,002 | 2212 | | | Enh | PromP | | Enh | Sarcopterygii |
| 8 | eDlx37 | chr7 | 96,176,396 | 96,177,200 | 805 | | | Enh | | | Enh | Sarcopterygii |
| 9 | eDlx38 | chr7 | 96,174,799 | 96,175,603 | 805 | | | | | | Enh | Birds |
| 10 | eDlx39 | chr7 | 96,155,601 | 96,156,406 | 806 | Enh | | Enh | | Enh | Enh | Birds |
| 11 | eDlx40 | chr7 | 96,149,956 | 96,150,999 | 1044 | Enh | | Enh | | | | Mammalian |
| 12 | eDlx41 | chr7 | 96,124,390 | 96,125,204 | 815 | Enh | | Enh | PromP | Enh | | Fish |
| 13 | eDlx42 | chr7 | 96,120,596 | 96,121,003 | 408 | | | Enh | | | | Mammalian |
| 14 | eDlx43 | chr7 | 96,055,303 | 96,056,480 | 1178 | | | Enh | | | Enh | Mammalian |
| 15 | eDlx44 | chr7 | 96,044,185 | 96,046,606 | 2,422 | Enh | Enh | Enh | Enh | Enh | Enh | Birds |
| 16 | eDlx45 | chr7 | 96,017,511 | 96,018,611 | 1101 | Enh | | Dnase | Enh | Enh | Enh | Mammalian |

Abbreviations: Carnegie stages 13-20 (CS13-20); CNCC = Cranial Neural Crest Cells; DNase hypersensitivity site (Dnase); histone marks of promoter (PromP); histone marks of active enhancer (Enh); histone marks of weak enhancer (TxEnhW).

## Table S17: Genes whose expression levels are affected by rs7981517 (*PCCA-AS1*) and rs4727341 (*SEM1*) according to FUMA annotation.

| Gene | Symbol | Type | eQTL minP | eQTL datasets | eQTL direction | Ind eQTL SNPs |
|---|---|---|---|---|---|---|
| ENSG00000237082 | *COX5BP6* | Pseudogene | 4.01E-12 | eQTLGen_cis_eQTLs | - | rs2254579; rs7981517 |
| ENSG00000134864 | *GGACT* | Protein coding | 5.49E-61 | eQTLGen_cis_eQTLs | - | rs2254579; rs7981517 |
| ENSG00000175198 | *PCCA* | Protein coding | 8.73E-135 | eQTLGen_cis_eQTLs | - | rs2254579; rs7981517 |
| ENSG00000125247 | *TMTC4* | Protein coding | 3.75E-57 | eQTLGen_cis_eQTLs: Cells_Cultured_fibroblasts* | - | rs2254579; rs7981517 |
| ENSG00000004864 | *SLC25A13* | Protein coding | 2.50E-06 | eQTLGen_cis_eQTLs | + | rs1524919 and others; |
| ENSG00000105880 | *DLX5* | Protein coding | 9.7E-05 | Colon_Sigmoid* | - | rs4727341; and others |
| ENSG00000105880 | *DLX5* | Protein coding | 1.20E-05 | Esophagus_Mucosa* | + | rs4727341; and others |

eQTL = expression quantitative trait locus; Dir = eQTL direction; *from GTExV8.

## Table S18: Transmission Disequilibrium Test single marker association results for the variants located within the eDlx36 regulatory element.

| SNP | Location | AF probands | AF GnomAD |
|---|---|---|---|
| rs28404011 | 7:96220294:C:T | 0.01 (AC = 2) | 0.01 |
| rs975539563 | 7:96220605:A:C | 0.005 (AC = 1) | 0.0003 |
| rs76382010 | 7:96220853:A:G | 0.005 (AC = 1) | 0.009 |
| rs955389935 | 7:96221416:G:A | 0.01 (AC = 2) | 0.0003 |

SNP= single neucleotide polymorphism; Location = chromosome: base pairs: reference allele: alternative allele; AF = allele frequency.

# Reference

[1]Wilderman A, VanOudenhove J, Kron J, Noonan JP, Cotney J. High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Rep* **23**, 1581-1597 (2018).