Supplemental Material

Supplemental methods:

Genome sequencing and bioinformatics

Genome sequencing was performed on Individuals III:1, III:2, III:3, III:4, IV:1, IV:2, IV:3, and IV:4 from Family 1 and IV:1 from Family 2 using Illumina X Ten. Reads were mapped to the human reference genome (NCBI build37/hg 19 version) and variants were called according to standard protocols (1-3). CNVnator was used for copy number variant detection (4). Candidate variants were confirmed with Sanger sequencing and deletion specific PCR. Primers utilized are listed in **Supplemental Table 7**.

iPSCs

We derived induced pluripotent stem cells iPSCs from dermal fibroblasts. Fibroblasts from both affected and control individuals were treated with the non-integrating Sendai virus vector carrying the Yamanaka reprogramming factors (MYC, KLF4, OCT4, and SOX2) using the Cytotune® iPS 2.0 Sendai reprogramming Kit (ThermoFisher Scientific) according to the manufacturer's protocol. Seven days post-transduction, we plated the cells on Matrigel® Membrane Matrix (Corning) coated culture dishes and switched from fibroblast media to iPSCs medium (*e.g.*, mTESR1 (StemCell Technologies)). Once colonies emerged, we manually picked individual iPSC colonies based on the similarity of their morphology to embryonic stem cell colonies and expression of the pluripotency marker TRA-1-60 using Human Pluripotent Stem Cell Live Cell Imaging Kit (R&D Systems). We expanded these clonal derived lines and tested the individual clones for pluripotency by ICC and qRT-PCR for OCT4, SOX2, NANOG, TRA-1-60, and TRA-1-81, as well as G-band karyotype to ensure that no gross chromosomal abnormality had arisen during the reprogramming process or expansion of the lines. Loss of the Sendai virus expression construct, which typically occurs within 40 days of transduction, was assessed using TaqMan RT-PCR assays to monitor the expression of Sendai viral vector genes (5-7).

Early otic lineage cell differentiation

We generated early otic lineage cells following the protocol of Matsuoka et al (7, 8). We plated iPSCs at low density in Matrigel (Corning Life Sciences) coated plates (10,000 cells/cm^2) and cultured them in mTESR1 stem cell growth media (StemCell Technologies). On day 3, we switched the media to N2B27-CDM (DMEM/F12 (1:1) containing 1x N2, 1x B27 supplements, 1x non-essential amino acids, 1x Glutamax, 100 uM β-mercaptoethanol) supplemented with 10 ng/mL FGF2, 10 ng/mL BMP4 (Peprotech), and 1 uM of the TGF beta inhibitor SB431542 (Biogem). At this stage, cells were positive for the NNE markers DLX3 and AP2α. On day 6, we changed the supplements to the medium with 10 ng/mL FGF2 (Peprotech), 1 uM SB431542 (Biogem), 2 uM of the WNT inhibitor IWP-2 (Biogem), and 100 nM of the BMP inhibitor LDN193189 (Biogem) to drive formation of the PPE that is positive for EYA1 and p75, with reduced expression of SOX2. On day 11, we harvested the cells by treatment with Accutase™ to form a single cell suspension. Cells were negatively selected for expression of the

pluripotency marker TRA-1-81 and positively for expression of p75 using CD271 magnetic beads, and plated at 25,000 cells/cm^2.

RNA-Seq analysis

RNA extraction and purification were performed using TRIzol (Invitrogen) and RNAeasy Minikit (Qiagen), respectively. We sequenced the transcriptomes of iPSCs and otic progenitor cells derived from patient fibroblasts from the proband (IV:4). We used ASE9203 and CD2 as control cells. We performed RNA-Seq on two clones of the proband and ASE9203 and CD2 as control cells in technical duplicates on days 3, 6 and 11.

Total RNA was quantified and qualified using the Agilent Bioanalyzer to have an RNA integrity score (RIN) set to 9. For RNA preparations, 1000 ng of total RNA was used as input for the Illumina TruSeq Total RNA Library Prep Kit with Ribo-Zero. Each sample had an incorporated unique barcode to allow for multiplexing. For total RNA libraries, ~50 million raw reads were generated in a 2 x 100 paired-end sequencing run.

Raw sequence data from the Illumina NovaSeq 6000 were processed by the on-instrument Real Time Analysis software to basecall files. These were converted to demultiplexed FASTQ files with the Illumina supplied scripts in the BCL2FASTQ software (v1.8.4). Quality of the reads was determined with FASTQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) for per base sequence quality, duplication rates, and overrepresented k-mers. Illumina adapters were trimmed from the ends of the reads using the Trim Galore! Package (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Reads were aligned to the human reference genome (hg19) with the STAR aligner (v2.5.0a) (9). Gene count quantification for total RNA was performed using the GeneCounts function within STAR against the GENCODE v19 transcript file. We run RSEM with the STAR aligner to create Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values.

TaqMan assays for *GDF6* (Hs01377663_m1; labeled with FAM) and *GAPDH* (Hs02758991_g1; labeled with VIC) were used to confirm gene expression levels.

RT-PCR

To evaluate expression of *Gdf6*, different tissues including lung, kidney, liver, spleen, hippocampus, cortex, kidney, and cochlea were dissected from P15 wildtype mice. In addition, cochlear expression of *Gdf6* was assessed in embryos of 17.5 dpc. Total RNA was isolated with TRIzol Reagent (Invitrogen) according to the manufacturer's instructions. Prior to reverse transcription, RNA samples were treated with rDNAse I (DNA-free kit, Applied Biosystems). cDNA was synthetized using qScript XLT cDNA SuperMix (Quanta Biosciences). Primers indicated in **Supplemental Table 7** were used to amplify *Gdf6* and *Gapdh*.

Animals

C57Bl/6:129-Gdf6 $^{tm1Kng/J}$ mice (10) were cryo-recovered at the Jackson Laboratory (JAX stock #024648). For colony expansion and stock maintenance two backcrosses with wildtype C57Bl/6 mice were done at the University of Miami. At weaning age, mice were genotyped and housed 2-4 per cage in a room with a 12 hour light:dark cycle (lights on at 7 AM, off at 7 PM) with access to food and water *ad*

libitum. F2 heterozygous animals were mated in order to obtain the E20 embryos that were utilized for phenotypic characterization.

Inner ear morphology

To determine whether *Gdf6* mutant mice showed any morphological change in the inner ear, specimens from *Gdf6* wildtype and *Gdf6* mutant mice at P0 were harvested and fixed overnight in Bodian's fixative. Organs were then dehydrated in ethanol and cleared in methyl salicylate. Membranous labyrinths were visualized by injecting 0.1% white latex paint in methyl salicylate as previously described (11).

Supplemental References:

1. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491-8.
2. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-303.
3. Li H, and Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010;26(5):589-95.
4. Abyzov A, Urban AE, Snyder M, and Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 2011;21(6):974-84.
5. Marion RM, Strati K, Li H, Tejera A, Schoeftner S, Ortega S, et al. Telomeres acquire embryonic stem cell characteristics in induced pluripotent stem cells. *Cell Stem Cell.* 2009;4(2):141-54.
6. Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol.* 2010;28(8):848-55.
7. Huber LJ, and Chao MV. Mesenchymal and neuronal cell expression of the p75 neurotrophin receptor gene occur by different mechanisms. *Dev Biol.* 1995;167(1):227-38.
8. Matsuoka AJ, Morrissey ZD, Zhang C, Homma K, Belmadani A, Miller CA, et al. Directed Differentiation of Human Embryonic Stem Cells Toward Placode-Derived Spiral Ganglion-Like Sensory Neurons. *Stem Cells Transl Med.* 2017;6(3):923-36.
9. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21.
10. Settle SH, Rountree RB, Sinha A, Thacker A, Higgins K, and Kingsley DM. Multiple joint and skeletal patterning defects caused by single and double mutations in the mouse Gdf6 and Gdf5 genes. *Developmental Biology.* 2003;254(1):116-30.
11. Diaz-Horta O, Abad C, Sennaroglu L, Foster J, 2nd, DeSmidt A, Bademci G, et al. ROR1 is essential for proper innervation of auditory hair cells and hearing in humans and mice. *Proc Natl Acad Sci U S A.* 2016;113(21):5993-8.

Supplemental Table 1: Genome sequencing coverage statistics of the study subjects.

Family	Individual	Average sequencing depth (X)	Coverage 1X (%)	Coverage 4X (%)	Coverage 10X (%)	Coverage 20X (%)
Family 1	III:1	38.89	99.6	99.6	99.4	96.1
Family 1	III:2	34.85	98.9	98.8	98.7	97.4
Family 1	III:3	33.37	99.7	99.6	99.2	93.9
Family 1	III:4	40.68	99.5	99	98.7	98.3
Family 1	IV:1	39.60	99.6	99.6	99.4	96.3
Family 1	IV:2	39.40	99	98.8	98.7	98.1
Family 1	IV:3	34.65	98.9	98.8	98.7	97.4
Family 1	IV:4	37.55	99.6	99.6	99.3	95.7
Family 2	IV:1	37.07	99.6	99.6	99.3	95.3

Supplemental Table 2: Parameters used for detecting homozygous runs (Enlis software; https://www.enlis.com/).

Parameter	Genome Sequencing
Variation types to test	SNPs only
Ignore rare or common variations	<5% and >95%
Ignore variations that have a quality score less than	20
Consolidate nearby regions	checked
Minimum region length	50,000 bp
Window scan size	50 variations
Maximum number heterozygous variants in window scan	0
Percent of windows for homozygous position call	5%
Number of consecutive positions to call a homozygous region	100
Maximum gap between variations in a region	100,000 bp

Supplemental Table 3: Coverage of homozygous regions (>2 MB) in two affected individuals of Family 1 via genome sequencing. Overlapping regions are in red.

Family 1, Individual IV:4					Family 1, Individual IV:1				
Chromosome # & Position (hg19)	Region Length	Coverage		Average sequencing depth (X)	Chromosome # & Position (hg19)	Region Length	Coverage		Average sequencing depth (X)
		1X (%)	10X (%)				1X (%)	10X (%)	
13:63659389-95376449	31717061	99.5	99.4	38.34	7:9486339-32425158	22938820	100	99.9	40.11
20:30324255-52445502	22121248	99.7	99.7	37.96	9:91105090-103764630	12659541	98	97.9	40
8:92775372-113875795	21100424	99	98.9	38.04	12:115679329-127901084	12221756	99.6	99.5	40.2
8:23888781-43757648	19868868	99.1	99	38.15	8:86849171-98195883	11346713	98.2	98	39.32
3:164435296-180680013	16244718	100	99.9	38.33	15:34719415-45094667	10375253	100	99.9	40.69
13:52263103-63598906	11335804	99.8	99.8	38.66	5:34543129-43502947	8959819	99.9	99.9	40.17
8:46889559-57926726	11037168	99.8	99.8	38.46	9:81749354-89653533	7904180	100	99.9	40.13
13:100635377-110965373	10329997	100	99.9	38.4	16:1309074-8928135	7619062	99.3	99.2	40.69
14:78924499-87983256	9058758	99.9	99.9	38.59	15:45337503-52586967	7249465	99.9	99.9	40.38
14:90679910-97658940	6979031	100	99.9	38.18	13:106756228-113082906	6326679	97.5	97.3	39.21
3:157474574-164059123	6584550	100	99.9	38.42	4:70299575-76317295	6017721	99.5	99.3	39.51
11:61017935-67000950	5983016	100	99.9	37.63	12:42514923-47626581	5111659	99.9	99.8	40.33
11:71281860-76902595	5620736	99.9	99.8	38.11	8:81958306-86546030	4587725	100	99.9	40.1
13:95386766-100567591	5180826	99.9	99.9	38.24	5:115380055-119664937	4284883	99.9	99.8	39.74
8:58150948-62661193	4510246	100	99.9	38.53	4:99707017-103640758	3933742	100	99.9	40.31
11:56482758-60957072	4474315	100	99.9	38.43	1:215222115-219055281	3833167	100	99.9	39.57
13:40302811-44726453	4423643	100	99.8	38.5	2:6366461-10086506	3720046	100	99.9	40.53
11:79480228-83837570	4357343	100	99.9	38.6	6:35820544-39212618	3392075	100	99.9	40.85
1:209304875-213638762	4333888	100	99.9	38.28	1:236659617-239925109	3265493	100	99.9	39.75
4:94719474-99026609	4307136	100	99.9	38.5	2:3185407-6267084	3081678	95.1	95	37.95
9:84336588-88442462	4105875	100	99.9	38.49	4:76480880-79513215	3032336	100	99.9	40.13
14:97774442-101420572	3646131	100	99.9	38.1	4:64237289-67116175	2878887	99.4	99.2	39.58
13:44732330-48237322	3504993	99.9	99.8	38.4	5:43513126-46365857	2852732	99.9	99.9	41.92
13:48389582-51744914	3355333	99.8	99.7	38.5	15:55825945-58666679	2840735	100	99.9	40.46
11:68302101-71243244	2941144	96.6	96.5	36.43	21:29425578-32242849	2817272	100	99.9	40.42
10:13170888-16061958	2891071	100	99.9	38.05	19:56358567-59097229	2738663	100	99.9	41.57
7:155923631-158706407	2782777	99.9	99.4	37.15	7:68277408-70990854	2713447	100	99.9	39.19
14:87998907-90670994	2672088	100	99.9	38.31	4:96836048-99528486	2692439	99.8	99.7	39.76
11:45693984-48334369	2640386	99.8	99.7	37.9	16:8937378-11500301	2562924	100	99.8	40.41
20:52489424-54974993	2485570	100	99.9	38.39	5:49625088-52175563	2550476	100	99.9	40.19
13:37998714-40295129	2296416	99.9	99.9	38.63	6:33252115-35753559	2501445	99.9	99.9	41.31
7:114240116-116312439	2072324	100	100	38.55	12:48342098-50652526	2310429	100	99.9	41.03

Supplemental Table 4: Shared autozygous regions between individuals IV:4 and IV:1 in Family 1.

Chromosome # & Position (hg19)	Shared Region Length	Family 1, Individual IV:4			Family 1, Individual IV:1		
		Coverage 1X (%)	Coverage 10X (%)	Average sequencing depth (X)	Coverage 1X (%)	Coverage 10X (%)	Average sequencing depth (X)
8: 92,775,372-98,195,883	5,420,511	96.1	95.9	36.88	96.3	95.9	38.45
9: 84,336,588-89,653,533	5,316,945	100	99.9	38.52	100	99.9	40.16
13:106,756,228-110,965,373	4,209,145	100	99.9	38.34	99.9	99.8	40.02
4: 96,836,048-99,026,609	2,190,561	100	100	38.56	99.8	99.7	39.69

Supplemental Table 5: Rare variants segregating within autozygous runs in Family 1.

Chr #	Position (hg19)	Reference	Alt	Type	ConsDetail	GeneName	mamPhCons	verPhCons	CADD* PHRED
13	108048554	A	T	SNV	intron	*FAM155A*	0.001	0.001	1.321
13	108059787	C	T	SNV	intron	*FAM155A*	0.086	0.121	0.555
13	108077707	A	C	SNV	intron	*FAM155A*	0.001	0.002	3.716
8	93022265	C	A	SNV	intron	*RUNX1T1*	0	0	3.127
8	93333064	C	T	SNV	intergenic	NA	0	0	11.81
8	93453821	A	T	SNV	intergenic	NA	0.002	0.002	2.972
8	93471560	A	G	SNV	intergenic	NA	0.11	0.221	9.094
8	93484030	G	A	SNV	intergenic	NA	0	0	3.732
8	93512618	T	C	SNV	downstream	*RP11-587H10.1*	0.003	0.004	7.659
8	93842879	T	C	SNV	intron,non_coding	*RP11-100L22.2*	0	0	2.729
8	94101844	T	A	SNV	intergenic	NA	0.001	0.001	4.4
8	94144787	T	A	SNV	downstream	*C8orf87*	0.002	0.005	3.328
8	94320734	C	T	SNV	intron,non_coding	*LINC00535*	0.001	0.001	1.324
8	94333785	T	C	SNV	intron,non_coding	*LINC00535*	0	0	1.162
8	94429249	T	A	SNV	intron,non_coding	*LINC00535*	0	0	3.157
8	95220647	C	G	SNV	intron	*CDH17*	0.004	0.005	4.876
8	95222961	A	G	SNV	upstream	*CDH17*	0.019	0.025	3.514
8	95249876	A	G	SNV	intergenic	NA	0	0	5.025
8	95492684	C	T	SNV	intergenic	NA	0.009	0.01	0.83
8	96060679	A	T	SNV	intron	*NDUFAF6*	0.145	0	11.46
8	96111765	G	A	SNV	intron,non_coding	*NDUFAF6*	0.006	0.007	2.358
8	96573358	C	G	SNV	intron,non_coding	*KB-1047C11.2*	0.001	0.001	1.237
8	97191568	T	C	SNV	intergenic	NA	0.007	0.008	11.67
8	97298334	C	T	SNV	intron	*PTDSS1*	0	0	4.176
8	97335444	G	A	SNV	intron	*PTDSS1*	0.001	0.001	0.046
8	97481231	C	T	SNV	intergenic	NA	0.001	0.001	3.655
8	97791652	A	G	SNV	intron	*CPQ*	0.001	0.001	3.702
8	98002717	T	C	SNV	intron	*CPQ*	0.199	0.268	8.29
8	98077066	T	C	SNV	intron	*CPQ*	0	0	4.648
8	98176496	C	T	SNV	intergenic	NA	0.146	0.178	0.162
8	96023063		G	INS	intron	*NDUFAF6*	0.002	0.002	NA
8	97061381	A		DEL	intergenic	NA	0.206	0.331	NA

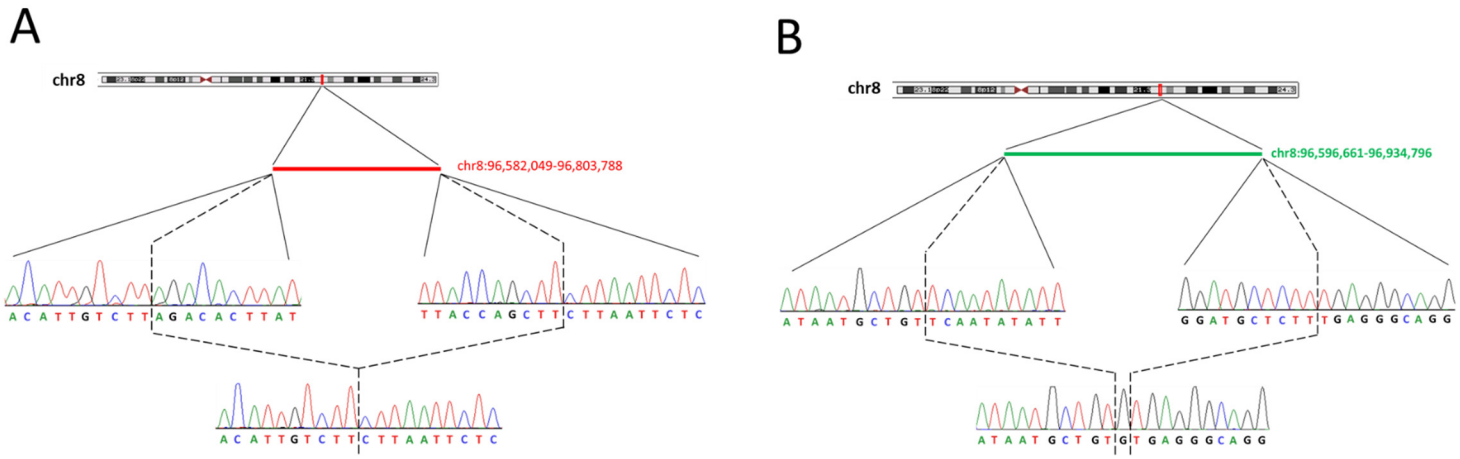
*CADD: Combined Annotation Dependent Depletion. CADD score is <15 for all variants.

Supplemental Table 6: Outcome of variant filtering in Family 1.

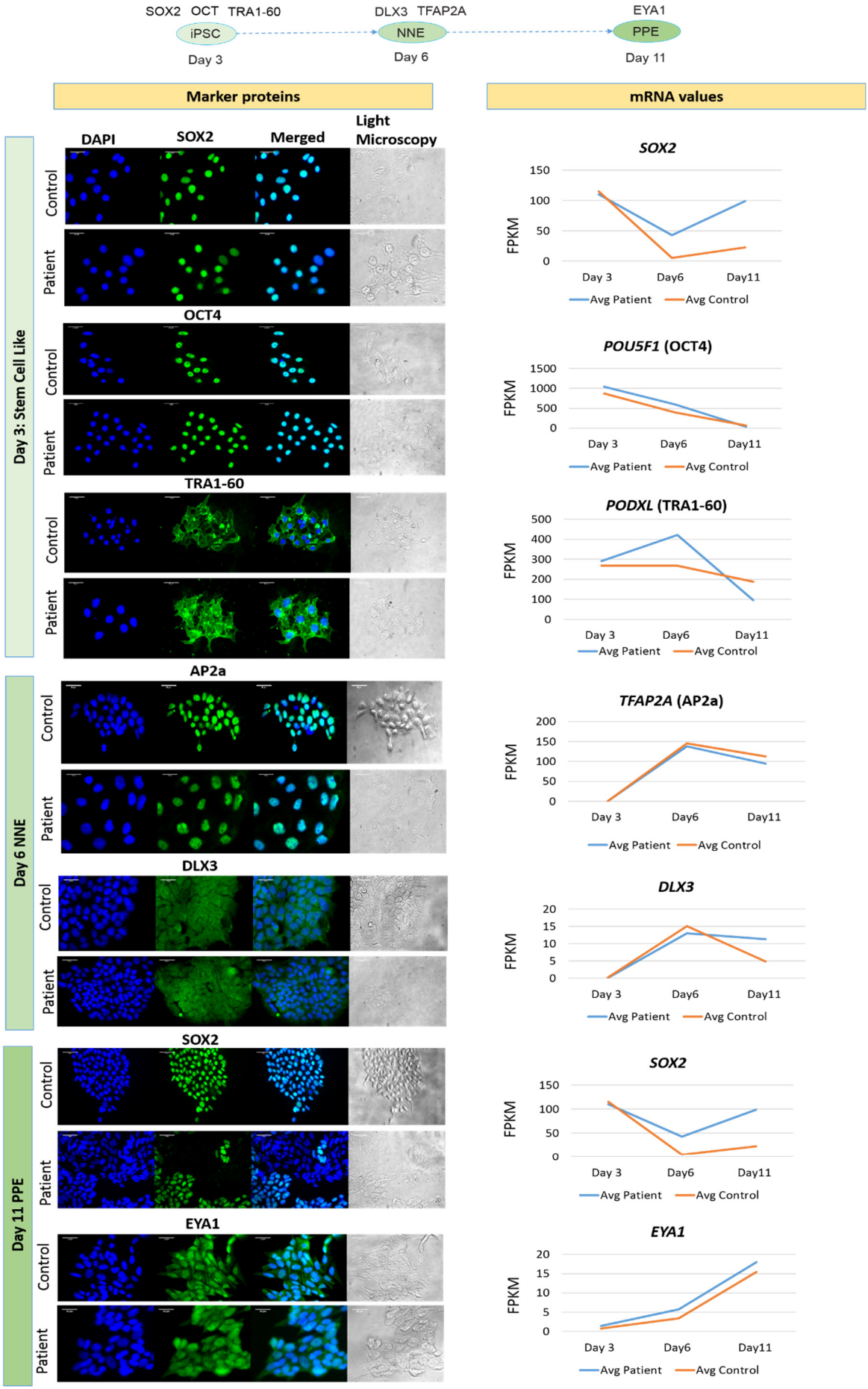
	SNVs		CNVs	
	Individual IV:1	**Individual IV:4**	**Individual IV:1**	**Individual IV:4**
Total Number of Variants	6,580,399	6,567,417	4,782 (1,543 del CNV)	4,571 (1,529 del CNV)
# of homozygous variants in shared homozygous regions (MAF <0.0007 gnomAD)	911	789	2	2
# of variants not in internal controls and segregating in the family	32		1	
# of conserved variants mamPhCons and verPhCons >0.4	0		-	

Supplemental Table 7: Sequences of primers used for analysis.

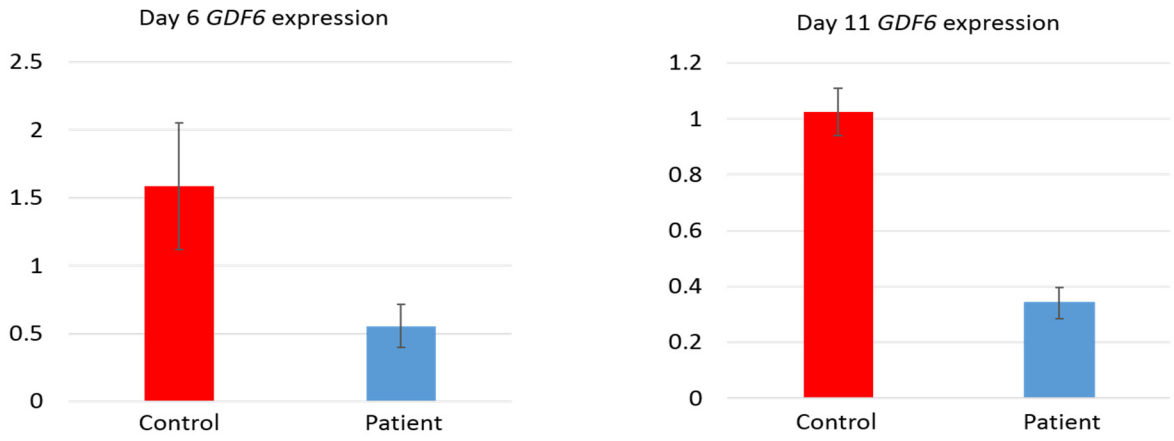
Primer Name	**Sequence 5' to 3'**	**Amplicon size (bp)**	**Description**
m_Gdf6_F	GCGCTTCACCTTACCTCAAG	198	Primers used for gene expression in mice
m_Gdf6_R	GGACGAAATCTTTCCTGTCG		
m_Gapdh_F	AGGTCGGTGTGAACGGATTTG	129	Primers used for gene expression in mice
m_Gapdh_R	TGTAGACCATGTAGTTGAGGTCA		
Fam1_brk_F	TTGGAGGTTTTCCCTCCTCT	425	Family 1 deletion-specific primers
Fam1_brk_R	AATGTCAACCCCTGTCCATC		
Fam1_brk_crt1_F	AGGTTATCCAATTTGTTGTCCTGT	330	Family 1 deletion breakpoint wildtype 1
Fam1_brk_crt1_R	CAAAGTGAGCAGGAGGAAGG		
Fam1_brk_crt2_F	AGGGAGGACATGACAACAGTCT	309	Family 1 deletion breakpoint wildtype 2
Fam1_brk_crt2_R	TCAAGCTAATTGGGGCAAAC		
Del_wildtype_F	TCCACCTTGACAAGATCAACC	307	Primers amplifying a region inside the overlapping (Family 1 and Family 2) deletion
Del_wildtype_R	AAAAATGCACCAGGGACAAG		
Fam2_brk_F	TGCAGCTGAGACGAGAGAGA	269	Family 2 deletion-specific primers
Fam2_brk_R	ATCAGTCCCAGGCTTGCTAA		
Fam2_brk_crt1_F	CTGACCAAGAGGTCAAATGGA	592	Family 2 deletion breakpoint wildtype 1
Fam2_brk_crt1_R	AACAAGGAGGCATTCACACC		
Fam2_brk_crt2_F	TGCCTGTCTCATGGTCAGAA	654	Family 2 deletion breakpoint wildtype 2
Fam2_brk_crt2_R	ATCAGTCCCAGGCTTGCTAA		



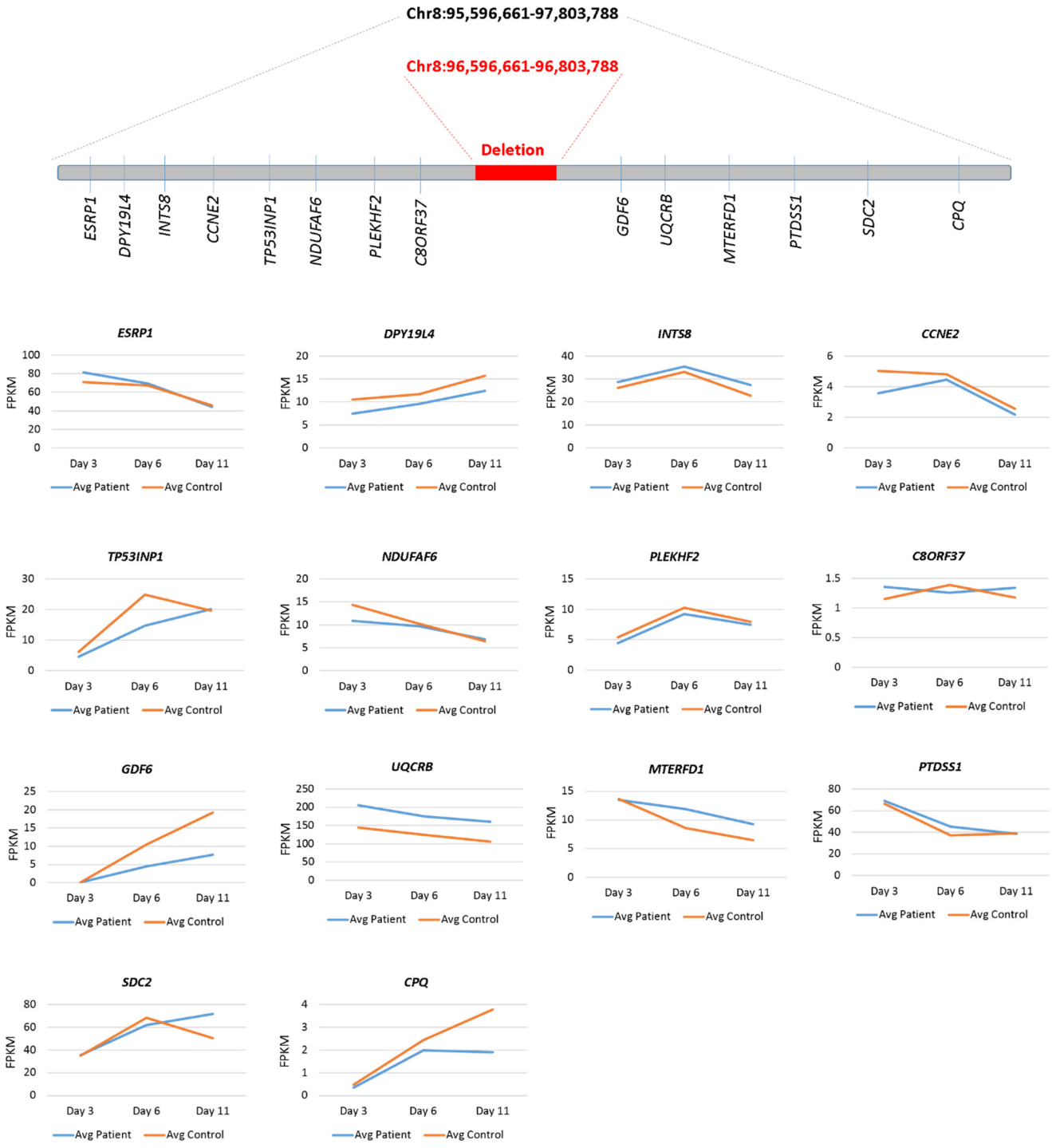
Supplemental Figure 1: Sanger sequencing confirmation of breakpoints. **A**) Electropherogram shows Family 1 seq[GRCh37] del(8)(q22.1(96,582,048)::q22.1(96,803,789)) deletion breakpoints. **B**) Electropherogram shows Family 2 seq[GRCh37] del(8)(q22.1(96,596,660)::G::q22.1(96,934,797)) deletion breakpoints.



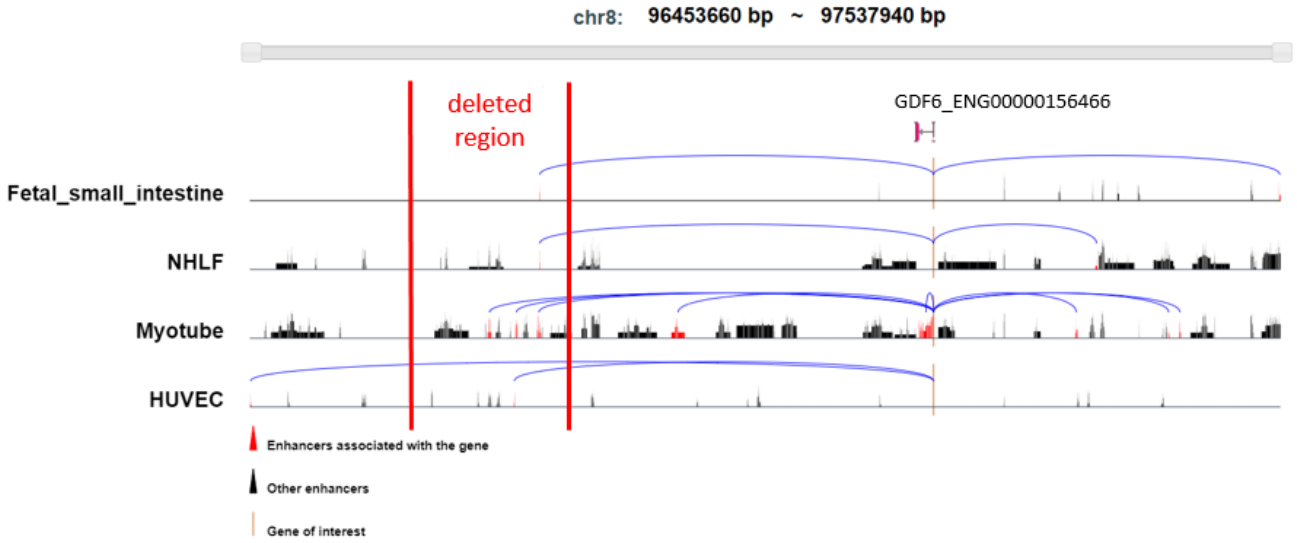
Supplemental Figure 2: Summary of expression levels of markers specific for otic progenitor cell differentiation steps. RNA-seq for expression of each gene is represented by using FPKM values on Y axis.



Supplemental Figure 3: Confirmatory *GDF6* expression levels on Days 6 and 11 performed by TaqMan probes. There is statistical difference for expression levels on Day 11 between control and patient cells (n= 8 cases; n=8 controls, $p<0.001$ via independent samples Mann Whitney U test).

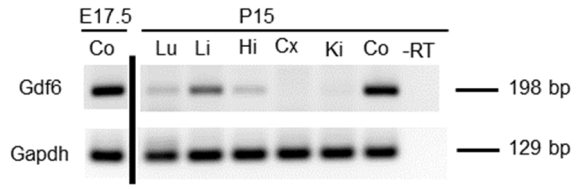


Supplemental Figure 4: Expression values for genes located in 1 MB flanking region of the homozygous deletions (Y axis = FPKM, X axis= Days). There is no statistical difference for any gene except for *GDF6* between patient and control cells.

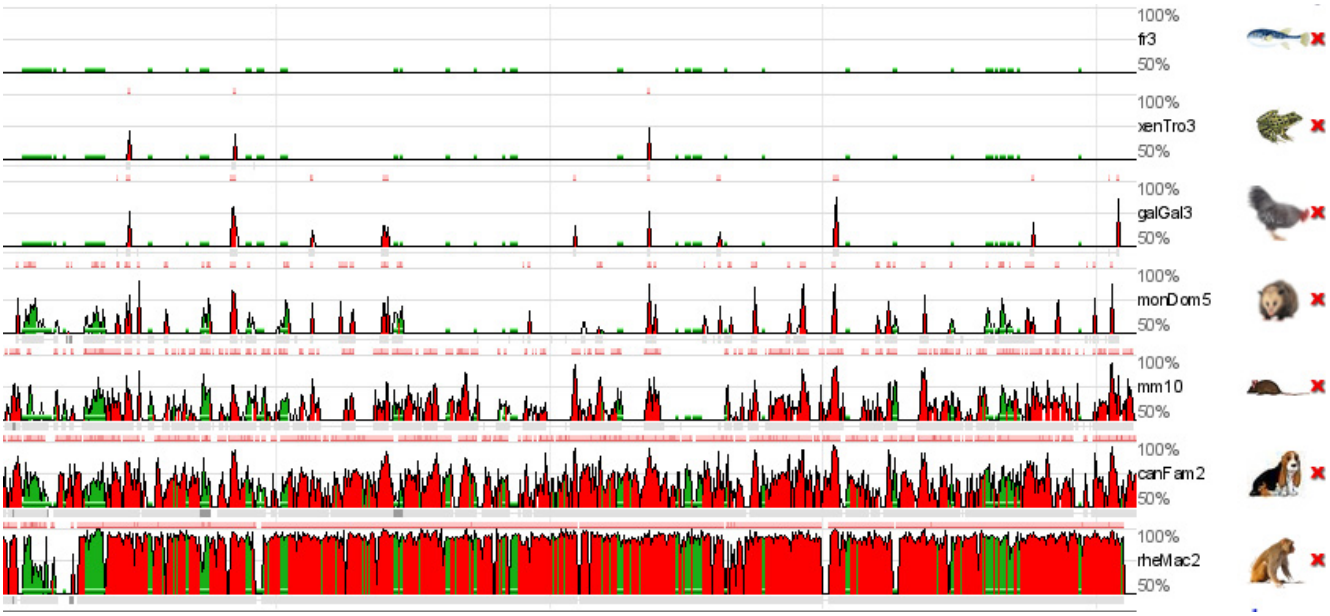


Cell type	Enhancer ID	Enhancer Position (hg19)	Transcript ID	Confidence Score
Fetal_small_intestine	E19-12413	chr8:96758350-96759000	ENST00000287020	0.856
NHLF	E08-10027	chr8:96758290-96758960	ENST00000287020	0.856
Myotube	E24-17395	chr8:96703960-96707630	ENST00000287020	0.732
Myotube	E24-17396	chr8:96733040-96735310	ENST00000287020	0.836
Myotube	E24-17398	chr8:96755600-96760000	ENST00000287020	0.864
HUVEC	E03-17997	chr8:96731420-96732470	ENST00000287020	0.866

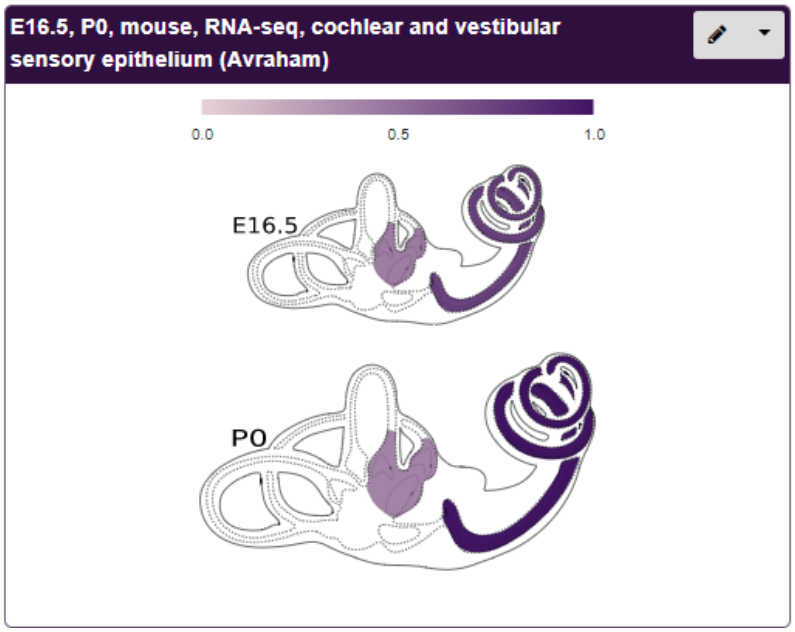
Supplemental Figure 5: Representation of the predicted interactions between *GDF6* promoter and enhancers in deleted region by using EnhancerAtlas (http://www.enhanceratlas.org/). NHLF: Normal human lung fibroblasts; HUVEC: Human Umbilical Vein Cells.



Supplemental Figure 6: Expression and localization of *Gdf6* in mouse tissues and cochlea. Expression of *Gdf6* in the mouse was confirmed by RT-PCR. Expression in cochlea at E17.5 and P15 is demonstrated by amplification of a 198 bp fragment with specific primers for *Gdf6*. *Gapdh* amplification was utilized as positive control. At P15, expression of *Gdf6* is present in lung (Lu), liver (Li), hippocampus (Hi), cortex (Cx), kidney (Ki), and cochlea (Co) with high expression. Lanes were run on the same gel but were noncontiguous.



Supplemental Figure 7: Figure represents Evolutionary Conserved Regions (ECRs) in the overlapping deleted genomic region (chr8:96,596,661-96,803,788) through species by using ECR Browser (https://ecrbrowser.dcode.org/). Human hg19 genome used as a base reference for comparison.



Supplemental Figure 8: Figure represents Cochlear and vestibular expression of *Gdf6* (data from Rudnicki et. al., 2014) by using gene Expression Analysis Resource (gEAR; https://umgear.org/).