



OPEN ACCESS



Open Access
Scan to access more
free content

ORIGINAL ARTICLE

A targeted next-generation sequencing assay for the molecular diagnosis of genetic disorders with orodental involvement

Megana K Prasad,¹ Véronique Geoffroy,¹ Serge Vicaire,² Bernard Jost,² Michael Dumas,² Stéphanie Le Gras,² Marzena Switala,^{3,4} Barbara Gasse,⁵ Virginie Laugel-Haushalter,⁶ Marie Paschaki,^{1,6} Bruno Leheup,⁷ Dominique Droz,⁸ Amelie Dalstein,⁸ Adeline Loing,³ Bruno Grollemund,⁴ Michèle Muller-Bolla,^{9,10} Sérena Lopez-Cazaux,¹¹ Maryline Minoux,^{3,4} Sophie Jung,^{3,4} Frédéric Obry,^{3,4} Vincent Vogt,^{3,4} Jean-Luc Davideau,⁴ Tiphaine Davit-Beal,^{5,12} Anne-Sophie Kaiser,¹³ Ute Moog,¹³ Béatrice Richard,¹⁴ Jean-Jacques Morrier,¹⁴ Jean-Pierre Duprez,¹⁴ Sylvie Odent,¹⁵ Isabelle Bailleul-Forestier,¹⁶ Monique Marie Rousset,¹⁷ Laure Merametdijan,¹⁸ Annick Toutain,¹⁹ Clara Joseph,²⁰ Fabienne Giuliano,²¹ Jean-Christophe Dahlet,³ Aymeric Courval,²² Mustapha El Alloussi,²³ Samir Laouina,²³ Sylvie Soskin,²⁴ Nathalie Guffon,²⁵ Anne Dieux,²⁶ Bérénice Doray,²⁷ Stephanie Feierabend,²⁸ Emmanuelle Ginglinger,²⁹ Benjamin Fournier,^{30,31} Muriel de la Dure Molla,^{30,31} Yves Alembik,²⁷ Corinne Tardieu,³² François Clauss,^{3,4} Ariane Berdal,^{30,31} Corinne Stoetzel,¹ Marie Cécile Manière,^{3,4} Hélène Dollfus,^{1,33} Agnès Bloch-Zupan^{3,4,6}

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jmedgenet-2015-103302>).

For numbered affiliations see end of article.

Correspondence to

Professor Agnès Bloch-Zupan,
Faculty of Dentistry,
University of Strasbourg,
8 rue St Elisabeth,
Strasbourg 67000, France;
agnes.bloch-zupan@unistra.fr

Received 6 June 2015
Revised 8 September 2015
Accepted 24 September 2015
Published Online First
26 October 2015

ABSTRACT

Background Orodental diseases include several clinically and genetically heterogeneous disorders that can present in isolation or as part of a genetic syndrome. Due to the vast number of genes implicated in these disorders, establishing a molecular diagnosis can be challenging. We aimed to develop a targeted next-generation sequencing (NGS) assay to diagnose mutations and potentially identify novel genes mutated in this group of disorders.

Methods We designed an NGS gene panel that targets 585 known and candidate genes in orodental disease. We screened a cohort of 101 unrelated patients without a molecular diagnosis referred to the Reference Centre for Oro-Dental Manifestations of Rare Diseases, Strasbourg, France, for a variety of orodental disorders including isolated and syndromic amelogenesis imperfecta (AI), isolated and syndromic selective tooth agenesis (STHAG), isolated and syndromic dentinogenesis imperfecta, isolated dentin dysplasia, otodental dysplasia and primary failure of tooth eruption.

Results We discovered 21 novel pathogenic variants and identified the causative mutation in 39 unrelated patients in known genes (overall diagnostic rate: 39%). Among the largest subcohorts of patients with isolated AI (50 unrelated patients) and isolated STHAG (21 unrelated patients), we had a definitive diagnosis in 14 (27%) and 15 cases (71%), respectively. Surprisingly, *COL17A1* mutations accounted for the majority of autosomal-dominant AI cases.

Conclusions We have developed a novel targeted NGS assay for the efficient molecular diagnosis of a wide

variety of orodental diseases. Furthermore, our panel will contribute to better understanding the contribution of these genes to orodental disease.

Trial registration numbers NCT01746121 and NCT02397824.

INTRODUCTION

Orodental disorders encompass a number and variety of diseases that affect the teeth and oral cavity. Broadly, these disorders can be classified into anomalies of tooth number, shape and size (eg, hypo/oligo/ano-dontia (collectively selective tooth agenesis (STHAG)), microdontia, globodontia), anomalies of tooth structure (eg, amelogenesis imperfecta (AI), hereditary dentin disorders) and anomalies of tooth eruption. The prevalence of these disorders varies from relatively common (4.2% for hypodontia in the Caucasian population)¹ to extremely rare (1 in 100 000 for dentin dysplasia (DD) type 1).²

Orodental disorders can have a genetic, environmental or multifactorial basis.^{3–4} Although evidence demonstrates a role for environmental pollutants such as dioxins and fluoride in developmental enamel defects,^{5–6} a number of studies have also demonstrated a strong genetic aetiology for several orodental diseases (reviewed in refs. 7–9). Among the >5000 known genetic syndromes, >900 have orodental/craniofacial features.¹⁰ Even in the case of isolated orodental diseases, significant genetic heterogeneity exists, with several of the same genes being involved in isolated and syndromic forms of



CrossMark

To cite: Prasad MK, Geoffroy V, Vicaire S, et al. *J Med Genet* 2016;53:98–110. doi:10.1136/jmedgenet-2015-103302
98–110.

disease. For instance, mutations in eight genes have been implicated in STHAG (*PAX9*, *MSX1*, *LTPB3*, *AXIN2*, *WNT10A*, *EDA*, *EDARADD* and *EDAR*),^{11–17} of which several (*MSX1*, *WNT10A*, *EDA*, *EDARADD* and *EDAR*) have also been linked with forms of ectodermal dysplasia.^{18–20} Similarly, mutations in a number of genes have been implicated thus far in AI, of which 10 cause an exclusively dental phenotype (*ENAM*, *WDR72*, *KLK4*, *AMELX*, *MMP20*, *FAM83H*, *AMBN*, *ITGB6*, *SLC24A4* and *c4orf26*),^{21–30} some cause syndromic disease with AI (*LTPB3*, *FAM20A*, *CNNM4*, *ROGDI*, *STIM1* and *FAM20C*),^{31–36} and yet others account for isolated and syndromic AI (*COL17A1*, *LAMA3*, *LAMB3* and *DLX3*).^{37–40} Indeed, the pattern of inheritance and penetrance associated with each gene also varies. This wide range of heterogeneity can render genetic diagnosis challenging.

Yet, the early molecular diagnosis of orodental disorders is important as it can improve patient care. For instance, mutations in *AXIN2* that cause STHAG have been shown to predispose carriers to colorectal cancer.¹² Early diagnosis of *AXIN2* mutations can hence alert clinicians to counsel patients to have regular colonoscopies. Similarly, the identification of mutations in *FAM20A* in patients presenting with AI can prompt a renal investigation for the management of nephrocalcinosis.⁴¹

Targeted next-generation sequencing (NGS) has proven extremely beneficial clinically for the molecular diagnosis of a number of genetically heterogeneous disorders, such as hearing loss, mitochondrial disease, intellectual disability (ID), neuromuscular disorders and Bardet-Biedl syndrome.^{42–46} Better coverage, lower cost and relative ease of data interpretation have made it more commonplace for routine clinical use than whole-exome sequencing (WES) and whole-genome sequencing (WGS).⁴⁷

We have developed the first targeted NGS panel for the molecular diagnosis of isolated and syndromic orodental disorders. We demonstrate the utility of this panel in the molecular diagnosis of a variety of orodental disorders. In a cohort of 103 patients (101 unrelated) without a known molecular diagnosis referred to the Reference Centre for Rare Diseases with Oro-dental Manifestations, we were able to provide a definitive molecular diagnosis in 39% of patients in known disease-causing genes. The identification of mutations in genes underlying syndromic forms of orodental disease highlights the potential benefits of a complete oral investigation in the diagnosis of rare genetic diseases. The aim of this article is to demonstrate the utility of this targeted NGS assay for the diagnosis of mutations in known orodental disease genes. However, the potential for the discovery of novel genes is addressed in the discussion.

METHODS

Patients

Patients were referred to the Reference Centre for Rare Diseases with Oro-dental Manifestations (Strasbourg, France) by dentists, paediatricians and geneticists from 13 hospitals and 32 private practices in France, Germany and Morocco. The inclusion criterion for the study was the presence of an orodental anomaly, defined as an anomaly of the mouth, including teeth and surrounding structures such as the periodontium (alveolar bone, ligament and gingivae), as well as defects of lip and palate formation. Patients with known mutations used for the validation assay were previously diagnosed either in clinical (Laboratory for Genetic Diagnosis, Strasbourg University Hospital) or research laboratories. DNA was obtained from peripheral blood or saliva samples (Oragene DNA, DNA Genotek, Canada).

Patient phenotype was recorded using D[4]/Phenodent (<http://www.phenodent.org>).

Gene selection and targeted capture design

Genes were selected based on their involvement in human diseases with orodental phenotypes, mutation in animal models presenting orodental disorders,^{48–49} expression in the developing mouse tooth⁵⁰ and known role in tooth development. Two versions of the gene panel were developed, v1.0 and v2.0. v1.0 was used for patients V1.01–V1.16, whereas v2.0 was used for patients V2.01–V2.95. Complementary RNA capture probes were designed against all coding exons and 25 bp of flanking intronic sequence in order to cover splice junctions of these genes using the SureDesign portal (<https://array.chem.agilent.com/suredesign>, Agilent, USA).

Library preparation, sequencing and data analysis

Targeted regions were captured using a Custom SureSelectXT2 in-solution target enrichment kit (Agilent) and libraries were prepared for sequencing (2×100 bp) on the HiSeq2500 (Illumina, USA) following the manufacturer's instructions. For v1.0, 16 samples were multiplexed per lane for sequencing, whereas for v2.0, 32 samples were multiplexed per flow-cell lane. Read alignment, and variant calling and annotation were performed using standard methods. Briefly, reads were aligned to the GRCh37 reference genome using Burrow-Wheeler aligner (v0.7.5a)⁵¹ ensuring tagging of multi-mapped reads, and duplicates were marked with Picardv1.102 (<http://picard.sourceforge.net>). Indel realignment, base quality score recalibration and variant calling were performed with the GATK Toolkit v3.1 using hard-filtering parameters.^{52–54} Variants were annotated using snpEffv3.4.⁵⁵ Variant frequencies were compared with an internal exome database and prioritised using VARRank.⁵⁶ Variants were prioritised by allele frequency (<1% in the Single Nucleotide Polymorphism database (dbSNP137), 1000 Genomes database,⁵⁷ Exome Variant Server (EVS) database⁵⁸ and our internal database, except for non-syndromic STHAG, for which we used a cut-off of <4%) and predicted functional effect (frameshift, invariant splice sites, non-synonymous and splice affecting mutations). Missense mutations were evaluated for pathogenicity bioinformatically using Sorting Intolerant from Tolerant (SIFT), PolyPhen, MutationTaster and amino acid conservation.^{59–61} Splice affecting mutations were evaluated bioinformatically using Human Splice Finder, MaxEntScan, NNSplice, Gene Splicer, SSF, Rescue ESE and ESE Finder.^{62–67} CNVs were detected as previously described.^{44–45} Variant pathogenicity was interpreted according to the American College of Medical Genetics guidelines.⁶⁸ Evidence used to establish pathogenicity is provided in online supplementary figures S1–S45.

Mutation validation

All mutations were validated by Sanger sequencing (GATC, Germany). Segregation analyses were performed whenever DNA was available for additional family members. Single-molecule PCR followed by Sanger sequencing was used to phase biallelic mutations when parental DNA was unavailable.^{69–70} Deletions were validated by qPCR. The region of interest and an internal control region (*RPPH1*) were amplified from 2 ng of genomic DNA from the patient and a control using the iQ SYBR Green Supermix (170-8880, Biorad, USA) on a CFX96 Real-Time System (Biorad). Data were analysed using the Pfaffl method.⁷¹

RESULTS

Validation of v1.0 of the NGS panel

As a proof of principle, we created a primary version (v1.0) of a custom NGS panel that targets 560 known and candidate genes (2.36 Mb) in orodental disorders. Of these genes, probes for 175 known and strong candidate genes (0.81 Mb) were designed to have superior coverage (diagnosis subpanel), whereas the remaining candidate genes constituted a 'Discovery subpanel' (see online supplementary tables S1 and S2). We validated v1.0 in a cohort of 16 patients: patients V1.01–V1.08 with isolated and/or syndromic orodental diseases with known mutations as determined by previous candidate gene Sanger sequencing, WES, or array CGH (see online supplementary table S3) and patients V1.09–V1.16 without a known molecular diagnosis. Patients V1.01–V1.08 were selected to include a variety of heterozygous, homozygous and hemizygous mutations (substitutions, indels, a large heterozygous deletion) at eight different loci in order to test the sensitivity of the gene panel to different types of genetic alterations. Using our computational pipeline, 10 of 10 mutations, including a large heterozygous deletion on the X chromosome (see online supplementary table S3), were detected in a blinded manner by a bioinformatician who was unaware of the molecular alteration but was aware of the clinical diagnosis for each patient. The bioinformatician was also blinded to the mode of inheritance in order to simulate a diagnostic scenario where such information is often unavailable. Samples V1.09–V1.16 were selected to cover a variety of orodental phenotypes (AI, dentinogenesis imperfecta (DGI), DD, STHAG) to determine the potential to identify unknown mutations with our panel. The molecular results from samples V1.09–V1.16 are presented and discussed in tables 1–3, together with the results from v2.0 of the gene panel presented below. Briefly, pathogenic or likely pathogenic mutations were identified in six of eight samples among samples V1.09–V1.16 (tables 2 and 3).

On average, in this validation cohort, by multiplexing 16 samples per lane of a sequencing flow cell, 2.2 Gb of sequence was generated per sample, giving a mean coverage of 365 \times , with 95.31% nucleotides covered at $\geq 50\times$ (see online supplementary table S4). The mean coverage of the diagnostic subpanel was 404 \times , with 98.94% of nucleotides covered at $\geq 20\times$, permitting confident diagnosis of mutations in these genes, for which we set a minimum coverage threshold of 20 \times (see online supplementary table S4). A small percentage of targeted regions (33.3 kb, 1.4%) had an average coverage $< 20\times$ across samples due to high guanine-cytosine (GC) content or pseudogenes with highly similar sequences (see online supplementary table S5). The majority of these regions were in the discovery subpanel (28.1 kb). Under-representation of high GC content and pseudogenic regions is a known issue with probe-based targeting strategies. However, these gaps can be filled in by Sanger sequencing for regions considered to be relevant on a case-by-case basis. On average, we detected ~2648 variants/sample, of which ~87 variants/sample were rare (<1% in dbSNP137, 1000Genomes, EVS and an in-house database). Among these rare variants, we identified on average per sample ~20 missense changes, 0.38 nonsense mutations, 0.25 splice-site changes and 0.56 frameshift-inducing indels (see online supplementary table S6).

Development of v2.0 of the NGS panel

Having validated v1.0 of our NGS gene panel, we included 25 additional genes to the discovery subpanel to create a second

version of the panel (v2.0) that included 585 genes (2.47 Mb). These additional genes were included due to their implication in animal models of orodental disease that was unknown when designing v1.0 of the panel. v2.0 of the panel was used to sequence 95 patients (V2.01–V2.95). Furthermore, since the average coverage achieved with v1.0 was more than sufficient for confident molecular diagnosis, with v2.0, 32 samples were multiplexed per well of a flow cell lane in order to reduce sequencing costs while ensuring a minimum average coverage of 100 \times per sample. Sequencing output with v2.0 is shown in online supplementary table S7. Briefly, we achieved a mean coverage of 179 \times overall, with 97.2% of the targeted region covered at $\geq 20\times$. Furthermore, the diagnostic panel had an average coverage of 211 \times with 97% of bases covered at $\geq 50\times$.

Screening a cohort of patients with diseases with orodental involvement

A description of the final cohort of 103 patients (101 unrelated) without a known mutation sequenced with v1.0 (V1.09–V1.16) and v2.0 (V2.01–2.95) and the diagnostic yield by disease category is shown in table 1. Isolated AI was the most common disorder in the cohort (50%), followed by isolated STHAG (20%). We also included several patients with syndromic forms of AI, ranging from well-defined syndromes, such as Enamel Renal Syndrome, to undefined and suspected syndromes, hoping that a molecular diagnosis may aid difficult clinical diagnosis. The remaining cases consisted of patients with isolated or syndromic dentin disorders, syndromic STHAG, suspected otodental dysplasia and primary failure of tooth eruption. Since AI and STHAG can be inherited in an autosomal-dominant (AD),

Table 1 Cohort description and diagnostic yield per disease category

Disease	Number of patients	Diagnostic yield, N (%)
Isolated AI	52 (51*)	14 (27%)
Of which confirmed AI	51 (50*)	15 (29%, 30%*)
Of which suspected AI	1	0 (0%)
Syndromic AI	14	1 (7%)
Enamel renal syndrome	1	1 (100%)
Mucopolysaccharidosis IV A	1	0 (0%)†
Kohlschutter Tonz (suspected)	1	0 (0%)
Osteogenesis imperfecta (suspected)	1	0 (0%)
Spondyloepiphyseal dysplasia	1	0 (0%)†
Undefined syndrome	9	0 (0%)
Isolated STHAG	21	15 (71%)
Syndromic STHAG	4	2 (50%)
Ectodermal dysplasia	3	1 (33%)
Intellectual disability with STHAG	1	1 (100%)
Isolated DGI	5 (4*)	5, 4* (100%)
Syndromic DGI	2	1 (50%)
Goldblatt syndrome (suspected)	1	0 (0%)
Osteogenesis imperfecta	1	1 (100%)
Isolated DD	2	1 (50%)
Otodental syndrome	1	1 (100%)
Primary failure of tooth eruption	2	0 (0%)
Total	103 (101*)	40, 39* (39%)

*Number of unrelated patients.

†Likely pathogenic mutations were identified in *GALNS* in these patients.

AI, amelogenesis imperfecta; DD, dentin dysplasia; DGI, dentinogenesis imperfecta; STHAG, selective tooth agenesis.

Table 2 Pathogenic mutations identified in patients without a known mutation

Patient ID	Age (years)	Sex	Mode	Clinical features	Gene	Transcript	c.	p.	Inheritance	Reference	Figure
<i>Pathogenic variants</i>											
<i>Isolated AI</i>											
V1.14	M	25	?	Hypoplastic, hypomineralised	<i>WDR72</i>	NM_182758.3	c.[182A>G];[815G>A]	p.[H61R];[W272*]	Compound heterozygous	Novel	S1
V2.05	M	6	AR	N/A	<i>CNNM4</i>	NM_020184.3	c.[1495G>A];[1495G>A]	p.[V499M];[V499M]	Homozygous (C)	Novel	S2
V2.08	M	12	AR	Hypomineralised	<i>SLC24A4</i>	NM_153646.3	c.[(1537+1_1538-1)_(*)67_?del];[(1537+1_1538-1)_(*)67_?del]	p.[0?];[0?]	Homozygous (C)	72	S3
V2.09	F	16	AD	Hypoplastic	<i>COL17A1</i>	NM_000494.3	c.[2407G>T]; [=]	p.[G803*];[=]	Maternal	73	S4
V2.18	M	13	?	Hypoplastic	<i>LAMB3</i>	NM_000228.2	c.[124C>T]; [=]	p.[R42*];[=]	?	74	S5
V2.20	F	18	?	Hypomature	<i>FAM83H</i>	NM_198488.3	c.[1289C>A];[=]	p.[S430*];[=]	?	75	S6
V2.26	M	13	AD	Hypomineralised	<i>FAM83H</i>	NM_198488.3	c.[1282C>T];[=]	p.[Q428*];[=]	Paternal	Novel	S7
V2.28	M	14	?	Hypoplastic	<i>AMELX</i>	NM_182680.1	c.[155C>T];[=]	p.[P52L];[=]	De novo	Novel	S8
V2.29	M	14	AD	N/A	<i>ENAM</i>	NM_031889.2	c.[123+1G>A];[=]	p.[0?];[=]	Paternal	Novel	S9
V2.48	F	20	AD	Hypoplastic	<i>COL17A1</i>	NM_000494.3	c.[1646G>A];[=]	p.[W549*];[=]	Maternal	Novel	S10
V2.53	F	19	?	Hypoplastic	<i>AMBN</i>	NM_016519.5	c.[532-1G>C];[532-1G>C]	p.[0?];[(0?)]	Homozygous	Novel	S11
V2.63	M	10	?	N/A	<i>FAM83H</i>	NM_198488.3	c.[2029C>T];[=]	p.[Q677*];[=]	?	76	S12
V2.79	M	10	?	Hypoplastic, hypomature	<i>COL17A1</i>	NM_000494.3	c.[1873C>T];[=]	p.[R625*];[=]	?	Novel	S13
V2.82	F	14	AD	Hypoplastic	<i>COL17A1</i> ; <i>LAMA3</i>	NM_000494.3; NM_198129.1	COL17A1:c.[1141+1G>A]; <i>LAMA3</i> :c.[6477_6486del]	COL17A1:p.[0?]; <i>LAMA3</i> :p.[I2159Mfs*46]	Maternal and ?	Novel	2
<i>Syndromic AI</i>											
V2.06	F	14	?	Enamel Renal Syndrome	<i>FAM20A</i>	NM_017565.3	c.[1106_1107del]; [c.1006_1107del]	p.[E2316Gfs*10]; [E2316Gfs*10]	Homozygous	Novel	S14
<i>Isolated STHAG</i>											
V1.11	M	21	?	Ag 12, 13, 14, 15, 18, 22, 23, 24, 25, 28, 31, 34, 35, 38, 44, 45, 48	<i>WNT10A</i>	NM_025216.2	c.[383G>A];[=]	p.[R128Q];[=]	Maternal	77	S15
V1.15	F	11	?	Ag 12, 15, 17, 22, 25, 27, 28, 31, 35, 37, 38, 41, 44, 45, 47, 48	<i>WNT10A</i>	NM_025216.2	c.[343A>C](.)[682T>A]	p.[K115Q](.)[F228I]	?	Novel and 78	S16
V2.55	F	15	?	Ag 11, 12, 13, 14, 17, 18, 21, 22, 23, 24, 25, 27, 28, 31, 32, 33, 34, 37, 38, 41, 42, 43, 44, 45, 47, 48	<i>WNT10A</i>	NM_025216.2	c.[321C>A];[321C>A]	p.[C107*];[C107*]	Homozygous	78	S17
V2.65	F	17	?	Ag 12, 13, 15, 17, 18, 22, 23, 25, 27, 28, 31, 32, 35, 37, 38, 41, 42, 45, 48	<i>WNT10A</i>	NM_025216.2	c.[682T>A];[682T>A]	p.[F228I];[F228I]	Homozygous	78	S18
V2.66	M	37	?	Ag 12, 13, 18, 22, 28, 31, 32, 37, 38, 41, 42, 47, 48	<i>WNT10A</i>	NM_025216.2	c.[682T>A];[682T>A]	p.[F228I];[F228I]	Homozygous	78	S19
V2.67	M	11	AD	Ag 15, 16, 17, 18, 25, 26, 27, 28, 36, 37, 38, 45, 46, 47, 48	<i>PAX9</i>	NM_006194.3	c.[(?-115)_(*62_?)del];[=]	p.[0?];[=]	Paternal	78	S20
V2.69	M	18	?	Ag 13, 14, 18, 22, 23, 24, 28, 31, 41, 45	<i>WNT10A</i>	NM_025216.2	c.[637G>A];[=]	p.[G213S];[=]	?	79	S21
V2.71	M	11	?	Ag 12, 13, 15, 17, 18, 22, 23, 25, 27, 28, 31, 32, 33, 34, 35, 37, 38, 41, 42, 43, 44, 45, 47, 48	<i>WNT10A</i>	NM_025216.2	c.[682T>A];[321C>A]	p.[F228I];[C107*]	Compound heterozygous	78	S22
V2.72	M	33	?	Ag 17, 18, 22, 28, 31, 32, 37, 38, 41, 42, 47, 48	<i>WNT10A</i>	NM_025216.2	c.[682T>A](.)[321C>A]	p.[F228I](.)[C107*]	?	78	S23
V2.74	M	26	?	Ag 13, 15, 18, 23, 24, 25, 28, 31, 33, 34, 35, 37, 38, 41, 42, 43, 44, 45, 47, 48	<i>WNT10A</i>	NM_025216.2	c.[682T>A];[682T>A]	p.[F228I];[F228I]	Homozygous	78	S24

Continued

Table 2 Continued

Patient ID	Sex	Age (years)	Mode	Clinical features	Gene	Transcript	c.	p.	Inheritance	Reference	Figure
V2.76	M	15	?	Ag 12, 14, 15, 18, 22, 24, 25, 28, 31, 34, 35, 37, 38, 41, 42, 43, 44, 45, 47, 48	<i>WNT10A</i>	NM_025216.2	c.[682T>A];[682T>A]	p.[F228I];[F228I]	Homozygous	78	S25
V2.78	F	9	?	Ag 12, 14, 22, 31, 32, 35, 41	<i>EDA</i>	NM_001399.4	c.[467G>A];[=]	p.[R156H];[=]	De novo	17	S26
V2.91	F	12	AD	Ag 14, 15, 24, 25 35, 36, 45, 46	<i>MSX1</i>	NM_002448.3	c.[249del];[=]	p.[E84Rfs*76];[=]	Paternal	Novel	S27
V2.92	M	9	?	Ag 12, 17, 22, 23, 24, 25, 27, 35, 37, 33, 31, 41, 43, 45, 47	<i>WNT10A</i>	NM_025216.2	c.[682T>A];[=]	p.[F228I];[=]	Maternal	78	S28
V2.93	M	28	?	Ag 14, 15, 18, 25, 27, 28, 32, 34, 35, 37, 38, 44, 45, 47, 48	<i>WNT10A</i>	NM_025216.2	c.[682T>A];[682T>A]	p.[F228I];[F228I]	Homozygous	78	S29
Syndromic STHAG											
V2.54	F	49	?	ED	<i>WNT10A</i>	NM_025216.2	c.[682T>A];[c.416C>T]	p.[F228I];[p.A139V]	Compound heterozygous	78 and novel	S30
V2.87	F	3	?	Intellectual disability	<i>CTNNB1</i>	NM_001904.3	c.[998dup];[=]	p.[Tyr333*];[=]	?	Novel	S31
Isolated DGI/DD											
V1.09	F	10	AD	DD*	<i>DSPP</i>	NM_014208.3	c.[3480del];[=]	p.[S1160Rfs*154];[=]	?	Novel	S32
V1.10	M	9	AD	DGI*	<i>DSPP</i>	NM_014208.3	c.[3480del];[=]	p.[S1160Rfs*154];[=]	Maternal	Novel	S32
V2.36	M	45	AD	DGI	<i>DSPP</i>	NM_014208.3	c.[3533_3534insTA];[=]	p.[N1179Tfs*136];[=]	?	Novel	S33
V2.55	F	12	AD	DGI	<i>DSPP</i>	NM_014208.3	c.[52G>T];[=]	p.[V18F];[=]	Maternal	80	S34
V2.57	F	44	AD	DGI	<i>DSPP</i>	NM_014208.3	c.[3480del];[=]	p.[S1160Rfs*154];[=]	?	Novel	S35
V2.59	F	41	?	DGI	<i>DSPP</i>	NM_014208.3	c.[3682_3686del];[=]	p.[S1228*];[=]	?	Novel	S36
Syndromic DGI											
V2.58	M	47	AD	Osteogenesis imperfecta	<i>COL1A1</i>	NM_000088.3	c.[3837_3840del];[=]	p.[N1279Lfs*51];[=]	Paternal	Novel	S37
Otodental syndrome											
V2.03	M	13	AD		<i>FGF3</i>	NM_005247.2	c.[(?-3)_220+1_221-1]del);[=]	p.[0?];[=]	Paternal	Novel	S38

*Related individuals.

AD, autosomal dominant; AI, amelogenesis imperfecta; AR, autosomal recessive; C, consanguineous parents; DD, dentin dysplasia; DGI, dentinogenesis imperfecta; ED, ectodermal dysplasia; ID, intellectual disability; STHAG, selective tooth agenesis.

Table 3 Likely pathogenic variants and other variants identified in patients without a known mutation

Patient ID	Sex	Age (y)	Mode	Clinical features	Gene	Transcript	c.	p.	Inheritance	Reference	Figure
<i>Likely pathogenic variants</i>											
V1.12	M	11	?	Ectodermal dysplasia	EDA	NM_001399.4	c.[396+5G>A];[=]	p.[0?];[=]	Maternal	Novel	S39
V2.07	M	14	?	AI hypomineralised, hypoplastic	MMP20	NM_004771.3	c.[954-2A>T];[126+6T>G]	p.[0?];[0?]	Compound heterozygous	25 and novel	S40
V2.13	F	17	?	AI hypoplastic, hypomature	MMP20	NM_004771.3	c.[954-2A>T];[c.103A>C]	p.[0?];[R35R]	?	25 and novel	S41
V2.15	M	20	?	Spondyloepiphyseal dysplasia	GALNS	NM_000512.4	c.[121-31];[935C>G]	p.[0?];[T32S]	Compound heterozygous	Novel and 81	S42
V2.32	F	26	?	AI hypomineralised, hypoplastic	FAM20A	NM_017565.3	c.[590-2A>G;590-3C>A];[c.1294G>A]	p.[0?];[c.[AA32T]	Maternal and de novo	82 and novel	S43
V2.86	F	6	?	AI hypoplastic	LAMB3	NM_000228.2	c.[1903C>T];[=]	p.[R635*];[=]	Maternal, asymptomatic	83	S44
Cases where only a single pathogenic variant was identified											
V2.49	F	4	?	Mucopolysaccharidosis IVA	GAI NS	NM_000512.4	c.[115C>T];[?]	p.[R386C];[?]	de novo	84	S45

AI, amelogenesis imperfecta; ID, intellectual disability.

autosomal-recessive (AR) or X-linked fashion, the mode of transmission was unclear for the majority of cases. Sixty-six cases showed sporadic disease, whereas 35 showed familial transmission. The mode of inheritance could be confidently inferred only in 18 cases, 15 as AD due to the presence of other affected family members, 2 as AR due to the presence of consanguinity between the parents of the index case and 1 as X-linked. Due to the large number of cases without a known mode of inheritance of disease, a common situation in diagnostic cohorts, variant filtration was performed for all samples for all modes of inheritance. Overall, we had a definitive molecular diagnosis in known genes in 39 cases (39%) and discovered 21 novel pathogenic variants. All pathogenic mutations identified in this cohort are listed in **table 2**. Likely pathogenic mutations and other variants are shown in **table 3**. The distribution of pathogenic variants identified with respect to the disease category and the number of cases corresponding to each mutated gene is shown in **figure 1**.

Isolated AI: *COL17A1* mutations are a frequent cause of AI. Of the 50 unrelated cases with confirmed isolated AI, we identified the causative mutation in 14 cases (27%). Surprisingly, the most frequently mutated gene in our cohort was *COL17A1*, with four independent patients segregating pathogenic mutations in this gene (of which three are novel). In the biallelic state, mutations in *COL17A1*, encoding a structural component of hemidesmosomes, cause junctional, non-Herlitz-type epidermolysis bullosa (EB), a skin blistering disorder with associated dental enamel defects.⁸⁵ However, rare cases of isolated ADAI in heterozygous carriers have been reported.^{37 86} Similarly, heterozygous mutations in *LAMA3* and *LAMB3*, which also encode hemidesmosomal components, can cause ADAI, whereas biallelic mutations in these genes cause EB.^{38 39} We found one family segregating a known heterozygous *LAMB3* nonsense mutation. These patients showed a similar enamel phenotype characterised by a hypoplastic form of AI with thin enamel and a characteristic pitting of the enamel that varied in severity (**figures 2A–H** and **3C**). Depending on the extent of pitting, the enamel surface was sometimes rough. Secondary extrinsic colouration of the pits rendered them more visible. Interestingly, patient V2.82 had two unlinked mutations, one maternally inherited novel *COL17A1* splice mutation (c.1141 +1G>A) and a novel *LAMA3* frameshift-inducing deletion (p.I2159Mfs*46) that is absent in her mother (**figure 2I**). The father of the patient was unavailable for testing but was described as unaffected at the time of ascertainment. The c.1141 +1G>A mutation is predicted to cause either in-frame skipping of exon 14 (which encodes a part of the intracellular domain) or retention of intron 14 that would subsequently introduce a premature termination codon (PTC) and likely nonsense-mediated decay (NMD) of the transcript. Patient V2.82's enamel phenotype is more severe than that of her mother, suggesting digenic inheritance in this individual. Patient V2.82 has hypoplastic AI with clearly visible pitted enamel in the vestibular, lingual and palatal surfaces of premolars and molars and a white colouration following Retzius striae visible in the permanent incisors (**figure 2A, B**). Her mother's dentition, in contrast, shows more discrete signs of AI and fewer pits are visible on the buccal surfaces of premolars (**figure 2C, D**). Digenic mutations in *LAMB3* and *COL17A1* that modify the severity of the EB phenotype have been previously reported.⁸⁷ However, this is the first report of seemingly digenic inheritance in ADAI.

Methods

Figure 1 Distribution of pathogenic variants identified by disease category. The genes in which pathogenic mutations were identified are shown with respect to the disease in which they were identified. In parentheses next to each gene is the number of independent patients in whom mutations were identified in the gene. This figure includes results from patients V1.09–V1.16 and V2.01–V2.95. DD dentin dysplasia; DGI, dentinogenesis imperfecta; STHAG selective tooth agenesis.

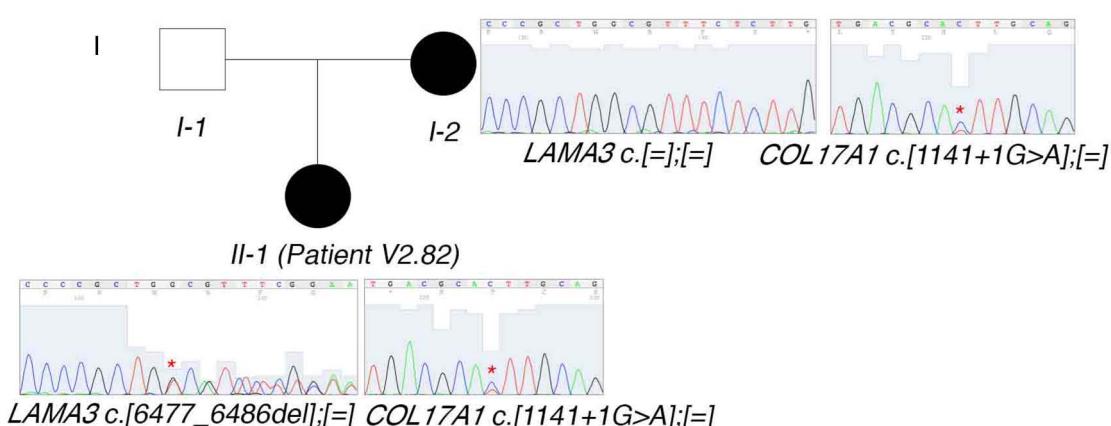
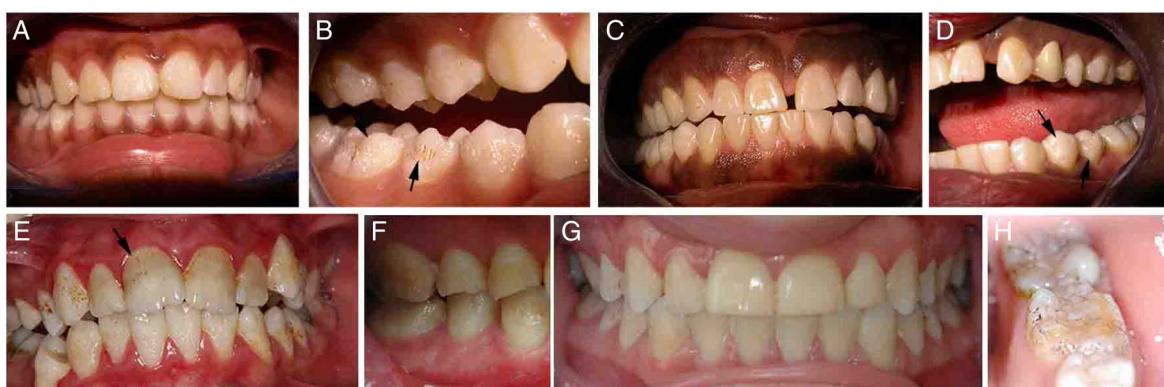
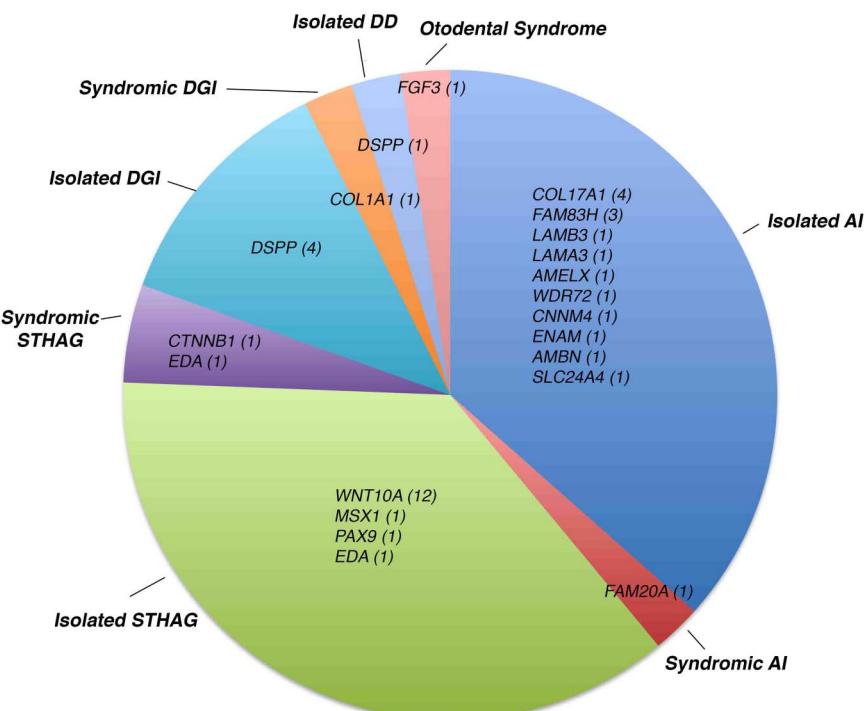


Figure 2 *COL17A1* mutations show a characteristic enamel phenotype in autosomal-dominant amelogenesis imperfecta. (A–H) Photos of the enamel phenotypes of individuals with *COL17A1* mutations. (A, B) Patient V2.82. (C, D) Mother of patient V2.82. (E, F) Patient V2.09. (G, H) Patient V2.48. Arrows mark pits in the enamel that are characteristic of *COL17A1* mutations. Extrinsic colouration makes these pits more visible in (B), (E) and (H). (I) Seemingly digenic inheritance of amelogenesis imperfecta in patient V2.82. Asterisks on the sequence chromatograms mark the mutated nucleotides.

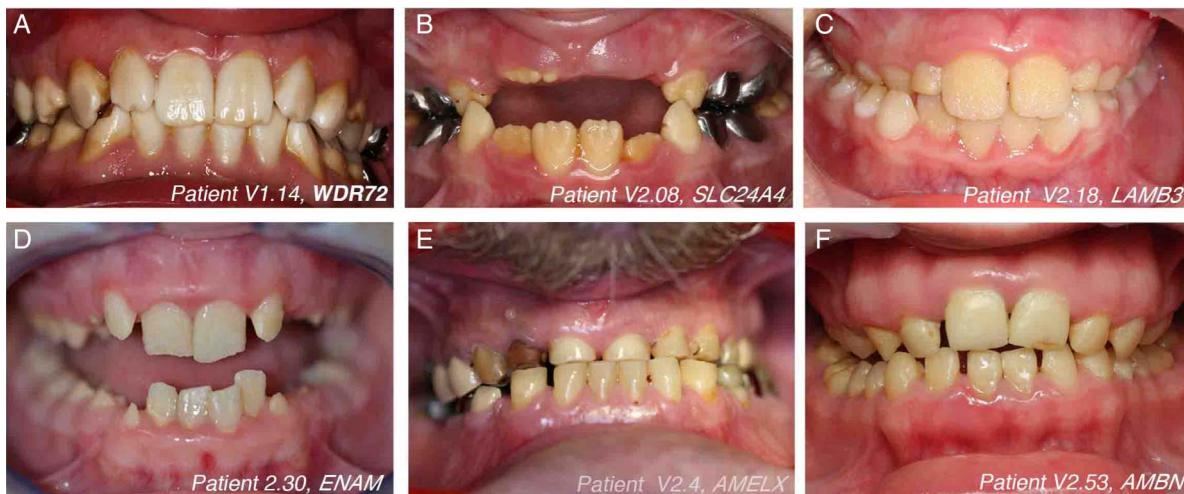


Figure 3 The variety of isolated amelogenesis imperfecta phenotypes seen in this cohort. Representative images of the enamel phenotype associated with mutations in different genes. The mutated gene and patient number are indicated in each panel.

We also identified a family with a novel homozygous splice-site mutation (*c.532-1G>C*) in *AMBN*. To date, only a single family has been reported to carry mutations in this gene.²⁷ Similar to the patients in the initial study, our patient too exhibited hypoplastic AI with very limited enamel as seen in the unrestored yellow-tinged premolars and permanent molars. However, the tooth surfaces did not seem as pitted as described previously (figure 3F).²⁷ This mutation is predicted to alter mRNA splicing by inducing either the skipping of exon 7 or the retention of intron 6. Exon skipping would lead to in-frame deletion of exon 7, which encodes a domain involved in heparin and fibronectin interaction that is thought to be important for the interaction of *AMBN* with dental epithelial cells.^{88 89} Intron retention would introduce a PTC that is likely to cause NMD of the mRNA. Additionally, a cryptic acceptor splice-site is predicted bioinformatically to be activated in exon 7 that would cause partial out-of-frame skipping of exon 7 and a subsequent PTC (see online supplementary figure S11). Thus, this mutation is predicted to cause at least a partial loss of *AMBN* function. We also identified novel mutations in *WDR72*, *AMELX* and *ENAM*, and a known deletion of the last three exons of *SLC24A4* (table 1). Representative images of the enamel phenotype associated with mutations in each gene are shown in figure 3. Similar to previous reports, *WDR72* and *SLC24A4* mutations caused hypomineralised, hypomature AI with brownish/yellowish discoloured teeth.^{22 29} The enamel showed a lack of contrast with the underlying dentin upon X-ray imaging (data not shown). The softer enamel seen in the patient with *WDR72* mutations was subject to wear, whereas the enamel in the patient with the *SLC24A4* mutation was opaque. Patients with *AMELX*, *LAMB3* and *ENAM* mutations had hypoplastic AI with yellow discolouration and smaller teeth due to thinner enamel.

Interestingly, we identified a novel homozygous missense mutation (*p.V499M*) in *CNNM4* in patient V2.05 who was born to first-cousins and was referred for isolated AI. This mutation affects a highly conserved residue in a functional cystathione beta-synthase domain of the protein and is predicted bioinformatically by three algorithms (SIFT, PolyPhen and MutationTaster) to be deleterious (see online supplementary figure S2). However, mutations in this gene are known to cause Jalili syndrome, which is characterised by a combination of AI and cone-rod dystrophy.^{33 90} A subsequent full-field

electroretinography revealed a marked loss of cone response and a less severe loss of rod response in the patient, thus confirming the Jalili phenotype (see online supplementary figure S2E, F).

Syndromic AI

We confirmed the molecular diagnosis in one case of Enamel Renal Syndrome (patient V2.06). This patient was homozygous for a novel frameshift-inducing deletion in *FAM20A* (*c.1106_1107del*).

In patient V2.15, referred for the management of AI in the context of suspected spondyloepiphyseal dysplasia (SED), we identified a known pathogenic missense mutation (*p.T312S*) and a likely pathogenic intronic variant (*c.121-31*) in *GALNS* (mutated in mucopolysaccharidosis (MPS) IVA) predicted to create an intronic splice enhancer (see online supplementary figure S42). However, the biological effect of this intronic mutation warrants further investigation. No deleterious mutations were identified in *CHST3* or *COL2A1* and a differential diagnosis of mucolipidosis type III was eliminated due to normal hexosaminidase, alpha-L-fucosidase and beta-D-glucuronidase levels (data not shown). Similar clinical presentation of SED and MPS IVA has been previously demonstrated;⁹¹ therefore, our results suggest a potential differential diagnosis of MPS IVA in this patient. In the case of patient V2.49, referred for the management of AI linked to MPS IVA,^{92 93} we identified only a single de novo known pathogenic mutation in *GALNS* (*p.R386C*), suggesting the presence of a second mutation in intronic or regulatory regions, or a structural mutation missed by our pipeline. The presence of an undetected second mutation is consistent with the virtually absent *GALNS* activity in the patient's leucocytes (1 nmol/h/mg protein) (data not shown). Both these patients exhibited hypoplastic AI with thin striated and pitted enamel and subsequently flattened buccal surfaces (see online supplementary figure S45) consistent with previous reports of the enamel phenotype seen in MPS IVA,⁹⁴ further supporting a diagnosis of MPS IVA in both patients. However, additional tests are necessary to confirm this diagnosis.

Isolated and syndromic STHAG

We identified the causative mutation in 15/21 (71%) cases of isolated STHAG. The majority carried mutations in *WNT10A* either in a heterozygous, homozygous or compound

heterozygous state as has been previously reported.^{77–79} Fortunately, we were also able to provide a molecular diagnosis for a patient presenting with ID who was referred for the management of hypodontia. The patient (V2.87) had agenesis of five incisors, two maxillary incisors and three mandibular permanent incisors (agenesis 12, 22, 41, 31 and 32) and had no family history of hypodontia or ID. We identified a novel heterozygous nonsense mutation in *CTNNB1* (p.Tyr333*), which encodes beta-catenin, a member of the WNT signalling pathway, which is essential during multiple stages of tooth development.⁹⁵ Recent reports have established a role for heterozygous *CTNNB1* mutations in ID characterised by mild to severe ID, autism spectrum disorder childhood hypotonia with progressive spastic diplegia, microcephaly and significant additional craniofacial and brain abnormalities, which is compatible with this patient's phenotype (data not shown).^{96,97} However, the orodental phenotype in this syndrome has not been studied in detail. Our report suggests that patients with ID due to *CTNNB1* mutations may require a dental examination and management of hypodontia. Furthermore, this case highlights the benefit of an interdisciplinary approach to patient care and the potential utility of a full dental examination in syndromes with craniofacial involvement.

Dentinogenesis imperfecta and dentin dysplasia

Among the six independent cases of isolated dentin disorders, we identified the causative mutation in five. All patients carried mutations in *DSPP*, the only gene implicated thus far in isolated DD/DGI.² The majority of mutations were present in the repeat-rich exon 5 that is often refractory to Sanger sequencing and hence until now excluded from routine diagnostic sequencing.⁹⁸ We achieved an average coverage of 771× and 408× over exon 5 of *DSPP* with v1.0 and v2.0, respectively. Hence, our assay provides an effective means to screen for mutations in the last exon of *DSPP*. Our bioinformatic pipeline detected the majority of mutations; however, one mutation (c.3682_3686del) was missed due to the difficulty of calling variants in structurally complex regions (personal communication, MKP, Broad Institute GATK Team), but was detected upon visual inspection of the reads. We also identified a novel frameshift-inducing deletion in *COL1A1* in a family segregating AD osteogenesis imperfecta with DGI.

Otodental dysplasia

We identified a heterozygous deletion of the first exon of *FGF3* in a family segregating suspected AD otodental dysplasia, which is characterised by globodontia and hearing loss, sometimes segregating with colobomas, most likely due to the deletion of the adjacent *FADD* gene.⁹⁹ This deletion was validated by qPCR on genomic DNA, which showed the presence of the deletion in patient V2.03 as well as in his affected father (see online supplementary figure S38).

DISCUSSION

We have developed a targeted NGS assay for the diagnosis and discovery of mutations underlying rare genetic disorders with orodental manifestations. We achieved high sequencing coverage in the targeted regions, 179× average coverage with 97.2% of the targeted region covered at ≥20× with V2.0. In a cohort of 101 unrelated patients with a variety of orodental genetic disorders of unknown genetic aetiology, we were able to detect the underlying pathogenic mutation in 39 cases (39%) in known genes.

Targeted NGS gene panels are being widely used for the diagnosis of a variety of genetic disorders, including ID, mitochondrial disorders and neuromuscular diseases.^{42–44} However, to the best of our knowledge, this is the first report of a targeted NGS gene panel for orodental disorders. Therefore, this diagnostic tool responds to a veritable need within the dental genetics community. Furthermore, two concrete examples from our cohort suggest the utility of this tool to the wider medical genetics community. First, we were able to diagnose a mutation in *CTNNB1* in a patient with ID and tooth agenesis, suggesting the potential utility of our assay in screening for mutations in ID cases with orodental involvement. Second, the detection of a *CNNM4* mutation in a patient referred for isolated AI led to an ophthalmological investigation for cone-rod dystrophy and the confirmation of Jalili syndrome. These cases highlight how the oral consultation can be a port of entry for the diagnosis and management of patients with rare genetic diseases, especially in light of the involvement of the same genes in syndromic and isolated forms of orodental disease and the sometimes non-evident extra-oral signs associated with some diseases.

Our targeted NGS panel targets fewer genes than WES/WGS and may thus exclude potential disease-causing genes. Yet, it has several advantages in a clinical setting. First, it provides higher sequencing coverage in the targeted regions than WES/WGS. A comparison of the coverage of the same regions achieved with v2.0 of our gene panel and a commercially available WES kit demonstrates that the coverage from WES falls short of the requirements for confident molecular diagnosis, that is, only 89.09% of targets covered at ≥20× with WES vs 97.23% of targets covered at ≥20× with our panel (see online supplementary figure S46 and table S8). Second, due to the smaller number of variants identified by targeted sequencing (~2600 variants with our panel vs ~30 000 with WES), variant interpretation is relatively simpler. This is also true for CNVs—we were able to detect and validate a single exon deletion in *FGF3*. Finally, the cost, amount of time for data analysis and interpretation, and data storage requirements are lower with targeted NGS than with WES/WGS approaches, making it an attractive alternative for clinical implementation. Alternatively, our assay could serve as a primary mutation-screening tool to exclude mutations in known genes before performing WES/WGS.

We expected that the inclusion of candidate genes selected based on their expression in developing mouse teeth⁵⁰ and their implication in animal models of orodental disorders would permit the identification of novel genes mutated in these disorders. In order to identify potentially novel genes in patients lacking mutations in known genes, we performed a preliminary analysis focusing on rare loss-of-function variants (nonsense, invariant splice site and frameshift variants) that were present in the same gene in at least two independent families with the same phenotype. Such a strategy was previously used successfully to identify a novel gene, *SETD5*, in ID.¹⁰⁰ However, this preliminary analysis did not yield any positive findings. Indeed, this analysis was limited in that it did not consider missense mutations whose effect is less clear and was complicated by the lack of information regarding the mode of inheritance in the majority of cases. Also, in this analysis patients were regrouped by disease category (all mutation-negative AI patients, all mutation-negative STHAG patients) without subclassification based on fine phenotype (eg, hypoplastic vs hypomineralised AI and severity of STHAG), which may be important to include in order to overcome the difficulties posed by genetic heterogeneity.¹⁰¹ Furthermore, the rarity of mutations in some genes may also necessitate much larger cohorts. Indeed, future analyses of this cohort will try to address these limitations.

Among the 50 unrelated patients with confirmed isolated AI in our cohort, we had a molecular diagnosis rate of 27%. This is slightly lower than the diagnostic rates reported by other studies that performed candidate gene sequencing in large AI cohorts. By performing Sanger sequencing of six genes (*FAM83H*, *ENAM*, *AMELX*, *MMP20*, *KLK4* and *WDR72*) in 71 families, Wright *et al*⁷⁵ identified mutations in 26 families (37%). By sequencing the same six genes in an independent cohort of 39 AI kindreds, Chan *et al*¹⁰² identified mutations in 19 kindreds (49%). In the latter study, the diagnostic rate was shown to be dependent on the mode of transmission: X-linked>AD>AR>simplex cases. In our study, the majority of cases were simplex cases (31 cases), which may explain to a degree our lower diagnostic rate despite the inclusion of a larger number of known AI genes. Furthermore, the 50 patients selected for this study belong to a larger AI cohort of 71 patients, a part of which (including patients in this study) had been previously sequenced by candidate gene Sanger sequencing and in which mutations were identified in 11 patients¹⁰³ (online supplementary table S3 and unpublished data). Therefore, the overall diagnostic rate in the 71-patient cohort is 35% (25 patients), which is closer to published diagnostic rates. In further contrast to the above-mentioned studies where *FAM83H* was the most frequently mutated gene in ADAI, *COL17A1* was the most frequently mutated gene in ADAI in our cohort, with 8% of our AI cohort carrying deleterious mutations in *COL17A1*. Additionally, we demonstrate that the presence of unlinked mutations in *COL17A1* and *LAMA3* can modify the severity of AI, suggesting a digenic mode of AI inheritance associated with mutations in the hemidesmosomal components. Therefore, our findings are novel in demonstrating the frequency of *COL17A1* mutations in ADAI. Finally, the large number of AI patients without mutations in known genes suggests that the genetic and allelic heterogeneity underlying AI is yet to be fully unravelled.

The diagnostic rate in syndromic AI was low in our cohort (7%). This is likely because of the inclusion of patients without a clear clinical diagnosis. These cases likely represent novel phenotypes with mutations in novel genes. Thus, WES may be better suited for such cases. Yet, the identification of a single de novo pathogenic mutation in *GALNS* in patient V2.49 can inform genetic counselling in this family. Similarly, the identification of one pathogenic and one potentially pathogenic mutation in *GALNS* in patient V2.15 permits an evidence-based investigation of MPS IVA by assaying leucocyte *GALNS* activity.¹⁰⁴ Therefore, although sequencing did not provide a definitive molecular diagnosis in these cases, the results of this assay can contribute to orienting clinical investigations/counselling.

Among the 21 cases of isolated STHAG, we identified the causative mutation in 15 (71%). This diagnostic rate is comparable to that reported in the literature. By screening *WNT10A*, *MSX1*, *PAX9*, *IRF6* and *AXIN2* by Sanger sequencing in 34 patients with isolated hypodontia, van den Boogaard *et al*⁷⁷ were able to identify the underlying mutation in 71% of cases. Similar to other reports, *WNT10A* accounted for the majority of mutations in our cohort (52%) and were also associated with ectodermal dysplasia (patient V2.54), with F228I, similar to other reports, being the most commonly mutated allele in our cohort, consistent with its population frequency (2.4% in European-Americans) and the prevalence of STHAG in this population (4.2%).^{77–79 105 106}

In conclusion, we have developed the first targeted NGS gene panel for the diagnosis and discovery of mutations in genetic disorders with orodental involvement. This panel can be reliably

used for the molecular diagnosis in known genes of a variety of genetic disorders and can serve as a primary screening tool before the application of WES/WGS. Additionally, this panel also provides the potential for the discovery of novel genes mutated in orodental disorders.

Author affiliations

- ¹Laboratoire de Génétique Médicale, INSERM U1112, Institut de génétique médicale d'Alsace, FMTS, Université de Strasbourg, Strasbourg, France
- ²Plateforme de Biopuces et Séquençage, Institut de Génétique et de Biologie Moléculaire et Cellulaire-Centre Européen de Recherche en Biologie et en Médecine, CNRS UMR7104, INSERM U964, Université de Strasbourg, Illkirch, France
- ³Centre de Référence des Manifestations Odontologiques des Maladies Rares, Pôle de Médecine et Chirurgie Bucco-dentaires, Hôpitaux Universitaires de Strasbourg (HUS), Strasbourg, France
- ⁴Faculté de Chirurgie Dentaire, Université de Strasbourg, Strasbourg, France
- ⁵Évolution et Développement du Squelette-EDS, UMR7138-SAE, Université Pierre et Marie Curie, Paris, France
- ⁶Institut de Génétique et de Biologie Moléculaire and Cellulaire-Centre Européen de Recherche en Biologie et en Médecine, CNRS UMR7104, INSERM U964 Université de Strasbourg, Illkirch, France
- ⁷Faculté de Médecine, CHU de Nancy, Université de Lorraine, Vandoeuvre-Les-Nancy, France
- ⁸CHU de Nancy, Service d'Odontologie, Nancy, France
- ⁹Département d'Odontologie Pédiatrique, UFR d'Odontologie, Université de Nice Sophia-Antipolis, CHU de Nice, Nice, France
- ¹⁰URB2i—EA 4462, Paris Descartes, Paris, France
- ¹¹Faculté de Chirurgie Dentaire, Département d'Odontologie Pédiatrique, CHU Hotel Dieu, Service d'odontologie conservatrice et pédiatrique, Nantes, France
- ¹²Faculté de Chirurgie Dentaire, Département d'Odontologie Pédiatrique, Université Paris Descartes, Montrouge, France
- ¹³Institute of Human Genetics, Heidelberg University, Heidelberg, Germany
- ¹⁴Service de Consultations et Traitements Dentaires, Hospices Civils de Lyon, Faculté d'Odontologie, Université Claude Bernard Lyon1, Lyon, France
- ¹⁵Service de Génétique Clinique, CHU de Rennes, Rennes, France
- ¹⁶Faculté de Chirurgie Dentaire, CHU de Toulouse, Odontologie Pédiatrique, Université Paul Sabatier, Toulouse, France
- ¹⁷Unité Fonctionnelle d'Odontologie pédiatrique, Service d'odontologie, CHRU de Lille, Lille, France
- ¹⁸Faculté de Chirurgie Dentaire, Service d'Odontologie Conservatrice et Endodontie, CHU Nantes, Université de Nantes, France
- ¹⁹Service de Génétique, CHU Tours, Tours, France
- ²⁰Département d'Odontologie Pédiatrique, Université de Nice Sophia-Antipolis, CHU Nice, Nice, France
- ²¹Service de Génétique Médicale, CHU Nice, Nice, France
- ²²Pôle de Médecine et de Chirurgie Bucco-dentaire, Hôpital Civil, HUS, Strasbourg, France
- ²³Faculty of Dental Medicine, Department of Pediatric Dentistry, University Mohammed V Rabat, Morocco
- ²⁴Pédiatrie 1, Hôpitaux Universitaires de Strasbourg, Strasbourg, France
- ²⁵Département de Pédiatrie, CHU de Lyon, Lyon, France
- ²⁶Service de génétique clinique Guy Fontaine, Centre Hospitalier Régionale Universitaire (CHRU) de Lille, Lille, France
- ²⁷Service de Génétique Médicale, CHU de Strasbourg, Strasbourg, France
- ²⁸Klinik für Zahnerhaltungskunde und Parodontologie, Universitäts Klinikum, Freiburg, Germany
- ²⁹Service de Génétique, Centre Hospitalier de Mulhouse, Mulhouse, France
- ³⁰Laboratoire de Physiopathologie Orale Moléculaire INSERM UMR S1138, Centre de Recherche des Cordeliers, Universités Paris-Diderot et Paris-Descartes, Paris, France
- ³¹Centre de Référence des Malformations Rares de la Face et de la Cavité Buccale MAFACE, Hôpital Rothschild, Pôle d'Odontologie, Paris, France
- ³²Aix-Marseille Université, UMR 7268 ADES/EFS/CNRS, APHM, Hôpital Timone, Service Odontologie, Marseille, France
- ³³Service de Génétique Médicale, Centre de Référence pour les Affections Rares en Génétique Ophtalmologique, HUS, Strasbourg, France

Acknowledgements The authors would like to thank all patients and their families for participating in this study. They also thank Ms Vanessa Stoehr for administrative help, Dr Mathilde Huckert and Jorel Salomon for technical assistance, Drs Elise Schaefer and Sophie Scheidecker for help with obtaining clinical data, and Dr Yaumara Perdomo-Trujillo for the electroretinography examination. They would like to thank Drs Alain Verloes, Jean-Yves Sire, Jean-Pierre Clavert, Béatrice Walter, Patrick Lutz, André Nirrengarten, Isabelle Rian, Thibault Sibert, Laurence Lustig-Grimm, Daniel Jasik, James Lespinasse, Jacques Hassid, Smail Hadj-Rabia, Gabriel Dominici, Laurent Riguet, Dan Lipsker, Jean Luc Alessandri, Florence Rousselet, Philippe Schoenlaub, François Serres, Etienne Mornet, Thomas Edouard,

Methods

Bertrand Baumann, Klaus Dietrich, Yves Bolender, Sandra Mercier, Elisabeth Steichen-Gersdorf, Jean-Gabriel Chillès, Hadrien Bonomi Dunoyer, Philippe Jonveaux, Supawich Morkumed, Patimaporn Pungchanchaikul, Sameh Mabrouk, Noora Zouari, Ilhame Mamouni, Alicia Rossa, Caroline Delfosse, Sylvie Goldmann, Esther Levy, Pierre Wahl, Jean-Claude Schoeffler, François Charton, Pascale Meng, Pierre-Emile Spaeth, Marie-Claude Addor, René Serfaty, Anthony Kurtz, Renaud Rinkenbach, Alain Duret, Jahan Figier, Denys Chaigne, Marc Hildwein, Florence Jouanet, Olivier Grison, Marie-Noëlle Duffard, Gautier Jouvenet Gilquin, François Undreiner, Michel Kretz, Cyrille Marc, Olivier Baver, Isabelle Zumsteg, Clémence Vorpahl, Alain Burtscher, Michel Kretz, Cyrille Marc, Jean-Pierre Strauss, Elise Pilavyan, Pierre Bonnaure, and Céline Benarroch-Leininger, Rémy Mathis, William Bacon, Marie-Rose Javier, Bilal Ahmed and Elia Sfeir for referring patients. Sequencing was performed by the IGBMC Microarray and Sequencing platform, a member of the 'France Génomique' consortium (ANR-10-INBS-0009).

Contributors Study concept and design: MKP, HD and AB-Z. Patient inclusion: MS, BL, DD, AD, AL, BG, MM-B, SLC, MM, SJ, FO, VV, J-LD, TD-B, A-SK, UM, BR, J-JM, J-PD, SO, IB-F, MMR, LM, AT, CJ, FG, J-CD, AC, MEA, SL, SS, NG, AD, BD, SF, EG, BF, MDM, YA, CT, FC, AB, MCM, HD and AB-Z. Data generation, analysis and interpretation: MKP, VG, SV, BJ, MD, SLG, VL-H, BG, MP, CS and AB-Z. Drafting of the manuscript: MKP and AB-Z. Critical reading of the manuscript: UM, AB and HD.

Funding This work was supported by grants from the French Ministry of Health (National Program for Clinical Research, PHRC 2008 N°4266 Amelogenesis imperfecta), the University Hospital of Strasbourg (API, 2009–2012, 'Development of the oral cavity: from gene to clinical phenotype in Human') and the EU-funded project (ERDF) A27 'Oro-dental manifestations of rare diseases,' supported by the RMT-TMO Offensive Sciences initiative, INTERREG IV Upper Rhine programme. <http://www.genosmile.eu>

Competing interests None declared.

Patient consent Obtained.

Ethics approval French Ministry of Higher Education and Research (MESR).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All PCR primers and protocols are available upon request from the corresponding author.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- 1 Polder BJ, Van't Hof MA, Van Der Linden FP, Kuijpers-Jagtman AM. A meta-analysis of the prevalence of dental agenesis of permanent teeth. *Community Dent Oral Epidemiol* 2004;32:217–26.
- 2 de La Dure-Molla M, Philippe Fournier B, Berdal A. Isolated dentinogenesis imperfecta and dentin dysplasia: revision of the classification. *Eur J Hum Genet* 2015;23:445–51.
- 3 Brook AH. A unifying aetiological explanation for anomalies of human tooth number and size. *Arch Oral Biol* 1984;29:373–8.
- 4 Brook AH, Smith JM. The aetiology of developmental defects of enamel: a prevalence and family study in East London, UK. *Connect Tissue Res* 1998;39:151–6; discussion 87–94.
- 5 Alaluusua S, Lukinmaa PL. Developmental dental toxicity of dioxin and related compounds—a review. *Int Dent J* 2006;56:323–31.
- 6 Koch G. Prevalence of enamel mineralisation disturbances in an area with 1–1.2 ppm F in drinking water. Review and summary of a report published in Sweden in 1981. *Eur J Paediatr Dent* 2003;4:127–8.
- 7 Bloch-Zupan A, Sedano H, Scully C. *Dento/oro/craniofacial anomalies and genetics*. 1st edn. Boston, MA: Elsevier, 2012.
- 8 Cobourne MT, Sharpe PT. Diseases of the tooth: the genetic and molecular basis of inherited anomalies affecting the dentition. *Wiley Interdiscip Rev Dev Biol* 2013;2:183–212.
- 9 Crawford PJ, Aldred M, Bloch-Zupan A. Amelogenesis imperfecta. *Orphanet J Rare Dis* 2007;2:17.
- 10 Hennekam RCM, Allanson JE, Krantz ID, Gorlin RJ. *Gorlin's syndromes of the head and neck*. 5th edn. Oxford; New York: Oxford University Press, 2010.
- 11 Bergendal B, Klar J, Stecksén-Blicks C, Norderyd J, Dahl N. Isolated oligodontia associated with mutations in EDARADD, AXIN2, MSX1, and PAX9 genes. *Am J Med Genet A* 2011;155A:1616–22.
- 12 Lammi L, Arte S, Somer M, Järvinen H, Lahermo P, Thesleff I, Pirinen S, Nieminen P. Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer. *Am J Hum Genet* 2004;74:1043–50.
- 13 Noor A, Windpassinger C, Vitcu I, Orlic M, Arshad Rafeq M, Khalid M, Nasir Malik M, Ayub M, Alman B, Vincent JB. Oligodontia is caused by mutation in LTBP3, the gene encoding latent TGF-beta binding protein 3. *Am J Hum Genet* 2009;84:519–23.
- 14 Stockton DW, Das P, Goldenberg M, D'souza RN, Patel PI. Mutation of PAX9 is associated with oligodontia. *Nat Genet* 2000;24:18–19.
- 15 Tao R, Jin B, Guo SZ, Qing W, Feng FY, Brooks DG, Liu L, Xu J, Li T, Yan Y, He L. A novel missense mutation of the EDA gene in a Mongolian family with congenital hypodontia. *J Hum Genet* 2006;51:498–502.
- 16 Vastardis H, Karimbu N, Guthua SW, Seidman JG, Seidman CE. A human MSX1 homeodomain missense mutation causes selective tooth agenesis. *Nat Genet* 1996;13:417–21.
- 17 Montreal AW, Zonana J, Ferguson B. Identification of a new splice form of the EDA1 gene permits detection of nearly all X-linked hypohidrotic ectodermal dysplasia mutations. *Am J Hum Genet* 1998;63:380–9.
- 18 Cluzeau C, Hadj-Rabia S, Jambou M, Mansour S, Guigue P, Masmoudi S, Bal E, Chassaing N, Vincent M-C, Viot G, Clauss F, Manière M-C, Toupenay S, Le Merrer M, Lyonnet S, Cormier-Daire V, Amiel J, Faivre L, De Prost Y, Munnoch A, Bonnefont J-P, Bodemer C, Smahi A. Only four genes (EDA1, EDAR, EDARADD, and WNT10A) account for 90% of hypohidrotic/anhidrotic ectodermal dysplasia cases. *Hum Mutat* 2011;32:70–2.
- 19 Jumlongras D, Bei M, Stimson JM, Wang W-F, Depalma SR, Seidman CE, Felbor U, Maas R, Seidman JG, Olsen BR. A nonsense mutation in MSX1 causes Witkop syndrome. *Am J Hum Genet* 2001;69:67–74.
- 20 Montreal AW, Ferguson BM, Headon DJ, Street SL, Overbeek PA, Zonana J. Mutations in the human homologue of mouse dl cause autosomal recessive and dominant hypohidrotic ectodermal dysplasia. *Nat Genet* 1999;22:366–9.
- 21 Rajpar MH, Harley K, Laing C, Davies RM, Dixon MJ. Mutation of the gene encoding the enamel-specific protein, enamelin, causes autosomal-dominant amelogenesis imperfecta. *Hum Mol Genet* 2001;10:1673–7.
- 22 El-Sayed W, Parry DA, Shore RC, Ahmed M, Jafri H, Rashid Y, Al-Bahlani S, Al Harasi S, Kirkham J, Inglehearn CF, Mighell AJ. Mutations in the beta propeller WDR72 cause autosomal-recessive hypomaturational amelogenesis imperfecta. *Am J Hum Genet* 2009;85:699–705.
- 23 Hart PS, Hart TC, Michalec MD, Ryu OH, Simmons D, Hong S, Wright JT. Mutation in kallikrein 4 causes autosomal recessive hypomaturational amelogenesis imperfecta. *J Med Genet* 2004;41:545–9.
- 24 Lagerström M, Dahl N, Nakahori Y, Nakagome Y, Bäckman B, Landegren U, Pettersson U. A deletion in the amelogenin gene (AMG) causes X-linked amelogenesis imperfecta (AIH1). *Genomics* 1991;10:971–5.
- 25 Kim JW, Simmer JP, Hart TC, Hart PS, Ramaswami MD, Bartlett JD, Hu JC. MMP-20 mutation in autosomal recessive pigmented hypomaturational amelogenesis imperfecta. *J Med Genet* 2005;42:271–5.
- 26 Kim JW, Lee SK, Lee ZH, Park J-C, Lee K-E, Lee M-H, Park J-T, Seo B-M, Hu JC-C, Simmer JP. FAM83H mutations in families with autosomal-dominant hypocalcified amelogenesis imperfecta. *Am J Hum Genet* 2008;82:489–94.
- 27 Poupter JA, Murillo G, Brookes SJ, Smith CEL, Parry DA, Silva S, Kirkham J, Inglehearn CF, Mighell AJ. Deletion of ameloblastin exon 6 is associated with amelogenesis imperfecta. *Hum Mol Genet* 2014;23:5317–24.
- 28 Poupter JA, Brookes SJ, Shore RC, Smith CEL, Abi Farraj L, Kirkham J, Inglehearn CF, Mighell AJ. A missense mutation in ITGB6 causes pitted hypomineralized amelogenesis imperfecta. *Hum Mol Genet* 2014;23:2189–97.
- 29 Parry DA, Poupter JA, Logan CV, Brookes SJ, Jafri H, Ferguson CH, Anvari BM, Rashid Y, Zhao H, Johnson CA, Inglehearn CF, Mighell AJ. Identification of mutations in SLC24A4, encoding a potassium-dependent sodium/calcium exchanger, as a cause of amelogenesis imperfecta. *Am J Hum Genet* 2013;92:307–12.
- 30 Parry DA, Brookes SJ, Logan CV, Poupter JA, El-Sayed W, Al-Bahlani S, Al Harasi S, Sayed J, Raïf EM, Shore RC, Dashash M, Barron M, Morgan JE, Carr IM, Taylor GR, Johnson CA, Aldred MJ, Dixon MJ, Wright JT, Kirkham J, Inglehearn CF, Mighell AJ. Mutations in C4orf26, encoding a peptide with in vitro hydroxyapatite crystal nucleation and growth activity, cause amelogenesis imperfecta. *Am J Hum Genet* 2012;91:565–71.
- 31 Huckert M, Stoetzel C, Morkmued S, Laugel-Haushalter V, Geoffroy V, Muller J, Clauss F, Prasad MK, Obry F, Raymond JL, Switala M, Alembik Y, Soskin S, Mathieu E, Hemmerle J, Weickert J-L, Dabovic BB, Rifkin DB, Dheedene A, Boudin E, Caluseriu O, Cholette M-C, Mcleod R, Antequera R, Gelle M-P, Coeuriat J-L, Jacquelain L-F, Bailleul-Forestier I, Manière M-C, Van Hul W, Bertola D, Dolle P, Verloes A, Mortier G, Dollfus H, Bloch-Zupan A. Mutations in the Latent TGF-beta Binding Protein 3 (LTBP3) gene cause brachyolmia with amelogenesis imperfecta. *Hum Mol Genet* 2015;24:3038–49.
- 32 O'Sullivan J, Bitu CC, Daly SB, Urquhart JE, Barron MJ, Bhaskar SS, Martelli-Júnior H, dos Santos Neto PE, Mansilla MA, Murray JC, Coletta RD, Black GCM, Dixon MJ. Whole-Exome sequencing identifies FAM20A mutations as a cause of amelogenesis imperfecta and gingival hyperplasia syndrome. *Am J Hum Genet* 2011;88:616–20.
- 33 Parry DA, Mighell AJ, El-Sayed W, Shore RC, Jalili IK, Dollfus H, Bloch-Zupan A, Carlos R, Carr IM, Downey LM, Blain KM, Mansfield DC, Shahrabi M, Heidari M,

- Aref P, Abbasi M, Michaelides M, Moore AT, Kirkham J, Inglehearn CF. Mutations in CNNM4 cause Jalili syndrome, consisting of autosomal-recessive cone-rod dystrophy and amelogenesis imperfecta. *Am J Hum Genet* 2009;84:266–73.
- 34 Schossig A, Wolf NI, Fischer C, Fischer M, Stocker G, Pabinger S, Dander A, Steiner B, Tönnz O, Kotzot D, Haberlandt E, Amberger A, Burwinkel B, Wimmer K, Fauth C, Grond-Ginsbach C, Koch MJ, Deichmann A, von Kalle C, Bartram CR, Kohlschütter A, Trajanoski Z, Zschocke J. Mutations in ROGDI Cause Kohlschütter-Tönz Syndrome. *Am J Hum Genet* 2012;90:701–7.
- 35 Acevedo AC, Poulter JA, Alves PG, De Lima C, Castro L, Yamaguti P, Paula LM, Parry DA, Logan CV, Smith CEL, Johnson CA, Inglehearn CF, Michell AJ. Variability of systemic and oro-dental phenotype in two families with non-lethal Raine syndrome with FAM20C mutations. *BMC Med Genet* 2015;16:8.
- 36 Wang S, Choi M, Richardson AS, Reid BM, Seymen F, Yildirim M, Tuna E, Gencay K, Simmer JP, Hu JC. STIM1 and SLC24A4 Are Critical for Enamel Maturation. *J Dent Res* 2014;93(7 Suppl):94S–100S.
- 37 McGrath JA, Gatalica B, Li K, Dunnill MG, McMillan JR, Christiano AM, Eady RA, Utton J. Compound heterozygosity for a dominant glycine substitution and a recessive internal duplication mutation in the type XVII collagen gene results in junctional epidermolysis bullosa and abnormal dentition. *Am J Pathol* 1996;148:1787–96.
- 38 Poulter JA, El-Sayed W, Shore RC, Kirkham J, Inglehearn CF, Michell AJ. Whole-exome sequencing, without prior linkage, identifies a mutation in LAMB3 as a cause of dominant hypoplastic amelogenesis imperfecta. *Eur J Hum Genet* 2014;22:132–5.
- 39 Yuen WY, Pasmooy AM, Stellingsma C, Jonkman MF. Enamel defects in carriers of a novel LAMA3 mutation underlying epidermolysis bullosa. *Acta Derm Venereol* 2012;92:695–6.
- 40 Dong J, Amor D, Aldred MJ, Gu T, Escamilla M, Macdougall M. DLX3 mutation associated with autosomal dominant amelogenesis imperfecta with taurodontism. *Am J Med Genet A* 2005;133A:138–41.
- 41 de la Dure-Molla M, Quentri M, Yamaguti PM, Acevedo A-C, Michell AJ, Viikula M, Huckert M, Berdal A, Bloch-Zupan A. Pathognomonic oral profile of Enamel Renal Syndrome (ERS) caused by recessive FAM20A mutations. *Orphanet J Rare Dis* 2014;9:84.
- 42 Chae JH, Vasta V, Cho A, Lim BC, Zhang Q, Eun SH, Hahn SH. Utility of next generation sequencing in genetic diagnosis of early onset neuromuscular disorders. *J Med Genet* 2015;52:208–16.
- 43 DaRe JT, Vasta V, Penn J, Tran NT, Hahn SH. Targeted exome sequencing for mitochondrial disorders reveals high genetic heterogeneity. *BMC Med Genet* 2013;14:118.
- 44 Redin C, Gérard B, Lauer J, Herenger Y, Muller J, Quartier A, Masurel-Pautel A, Willems M, Lesca G, El-Chehadeh S, Le Gras S, Vicaire S, Phillips M, Dumas M, Geoffroy V, Feger C, Haumesser N, Alembik Y, Barth M, Bonneau D, Colin E, Dollfus H, Doray B, Delrue M-A, Drouin-Garraud V, Flori E, Fradin M, Francannet C, Goldenberg A, Lumbruso S, Mathieu-Dramard M, Martin-Coignard D, Lacombe D, Morin G, Polge A, Sukno S, Thauvin-Robinet C, Thevenon J, Doco-Fenzy M, Genevieve D, Sarda P, Edery P, Isidor B, Jost B, Olivier-Faivre L, Mandel J-L, Piton A. Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *J Med Genet* 2014;51:724–36.
- 45 Redin C, Le Gras S, Mhamdi O, Geoffroy V, Stoetzel C, Vincent M-C, Chiurazzi P, Lacombe D, Ouertani I, Petit F, Till M, Verloes A, Jost B, Chaabouni HB, Dollfus H, Mandel J-L, Muller J. Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alström syndromes. *J Med Genet* 2012;49:502–12.
- 46 Shearer AE, DeLuca AP, Hildebrand MS, Taylor KR, Gurrola J, Scherer S, Scheetz TE, Smith RJH. Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. *Proc Natl Acad Sci USA* 2010;107:21104–9.
- 47 Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. *Nat Rev Genet* 2013;14:295–300.
- 48 Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE, Mouse Genome Database G. The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic Acids Res* 2015;43(Database issue):D726–36.
- 49 Bradford Y, Conlin T, Dunn N, Fashena D, Frazer K, Howe DG, Knight J, Mani P, Martin R, Moxon SAT, Paddock H, Pich C, Ramachandran S, Ruef BJ, Ruzicka L, Bauer Schaper H, Schaper K, Shao X, Singer A, Sprague B, Van Slyke C, Westerfield M. ZFIN: enhancements and updates to the Zebrafish Model Organism Database. *Nucleic Acids Res* 2011;39(Database issue):D822–9.
- 50 Laugel-Hauschalter V, Paschaki M, Thibault-Carpentier C, Dembelé D, Dollé P, Bloch-Zupan A. Molars and incisors: show your microarray IDs. *BMC Res Notes* 2013;6:113.
- 51 Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–95.
- 52 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.
- 53 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
- 54 Van Der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Rozenz D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013;11:11 1–11 33.
- 55 Cingolani P, Platts A, Wang Le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 2012;6:80–92.
- 56 Geoffroy V, Pizot C, Redin C, Piton A, Vasli N, Stoetzel C, Blavier A, Laporte J, Muller J. VaRank: a simple and powerful tool for ranking genetic variants. *PeerJ* 2015;3:e796.
- 57 Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA, 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.
- 58 Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA. <http://evs.gs.washington.edu/EVS/> (accessed Feb 2015).
- 59 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
- 60 Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4:1073–81.
- 61 Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 2014;11:1361–2.
- 62 Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 2003;31:3568–71.
- 63 Desmet FO, Hamroun D, Lalonde M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009;37:e67.
- 64 Fairbrother WG, Yeh RF, Sharp PA, Burge CB. Predictive identification of exonic splicing enhancers in human genes. *Science* 2002;297:1007–13.
- 65 Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 2001;29:1185–90.
- 66 Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol* 1997;4:311–23.
- 67 Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004;11:377–94.
- 68 Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. *Genet Med* 2008;10:294–300.
- 69 Kravtsberg Y, Nekhaeva E, Chang C, Ebralidse K, Khrapko K. Analysis of somatic mutations via long-distance single molecule PCR. *DNA Amplif* 2004;97–110.
- 70 Ruano G, Kidd KK, Stephens JC. Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules. *Proc Natl Acad Sci USA* 1990;87:6296–300.
- 71 Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 2001;29:e45.
- 72 Seymen F, Lee KE, Tran Le CG, Yildirim M, Gencay K, Lee ZH, Kim J-W. Exonal deletion of SLC24A4 causes hypomaturation amelogenesis imperfecta. *J Dent Res* 2014;93:366–70.
- 73 Darling TN, McGrath JA, Yee C, Gatalica B, Hametner R, Bauer JV, Pohla-Gubo G, Christiano AM, Utton J, Hintner H, Yancey KB. Premature termination codons are present on both alleles of the bullous pemphigoid antigen 2/type XVII collagen gene in five Austrian families with generalized atrophic benign epidermolysis bullosa. *J Invest Dermatol* 1997;108:463–8.
- 74 McGrath JA, Pulkkinen L, Christiano AM, Leigh IM, Eady RA, Utton J. Altered laminin 5 expression due to mutations in the gene encoding the beta3 chain (LAMB3) in generalized atrophic benign epidermolysis bullosa. *J Invest Dermatol* 1995;104:467–74.
- 75 Wright JT, Torain M, Long K, Seow K, Crawford P, Aldred MJ, Hart PS, Hart TC. Amelogenesis imperfecta: genotype-phenotype studies in 71 families. *Cells Tissues Organs* 2011;194:279–83.
- 76 Lee SK, Hu JC, Bartlett JD, Lee K-E, Lin BP-J, Simmer JP, Kim J-W. Mutational spectrum of FAM83H: the C-terminal portion is required for tooth enamel calcification. *Hum Mutat* 2008;29:E95–9.
- 77 van den Boogaard MJ, Crétion M, Bronkhorst Y, Van Der Hout A, Hennekam E, Lindhout D, Cune M, Ploos Van Amstel HK. Mutations in WNT10A are present in more than half of isolated hypodontia cases. *J Med Genet* 2012;49:327–31.
- 78 Bohring A, Stamm T, Spaich C, Haase C, Spree K, Hehr U, Hoffmann M, Ledig S, Sel S, Wieacker P, Röpke A. WNT10A mutations are a frequent cause of a broad spectrum of ectodermal dysplasias with sex-biased manifestation pattern in heterozygotes. *Am J Hum Genet* 2009;85:97–105.

Methods

- 79 Plaisancié J, Bailleul-Forestier I, Gaston V, Vaysse F, Lacombe D, Holder-Espinasse M, Abramowicz M, Coubes C, Plessis G, Faivre L, Demeer B, Vincent-Delorme C, Dollfus H, Sigaudy S, Guillén-Navarro E, Verloes A, Jonveaux P, Martin-Coignard D, Colin E, Bieth E, Calvas P, Chassaing N. Mutations in WNT10A are frequently involved in oligodontia associated with minor signs of ectodermal dysplasia. *Am J Med Genet A* 2013;161:671–8.
- 80 Xiao S, Yu C, Chou X, Yuan W, Wang Y, Bu L, Fu G, Qian M, Yang J, Shi Y, Hu L, Han B, Wang Z, Huang W, Liu J, Chen Z, Zhao G, Kong X. Dentinogenesis imperfecta 1 with or without progressive hearing loss is associated with distinct mutations in DSPP. *Nat Genet* 2001;27:201–4.
- 81 Wang Z, Zhang W, Wang Y, Meng Y, Su L, Shi H, Huang S. Mucopolysaccharidosis IVA mutations in Chinese patients: 16 novel mutations. *J Hum Genet* 2010;55:534–40.
- 82 Cho SH, Seymour F, Lee KE, Lee S-K, Kweon Y-S, Kim KJ, Jung S-E, Song SJ, Yildirim M, Bayram M, Tuna EB, Gencay K, Kim J-W. Novel FAM20A mutations in hypoplastic amelogenesis imperfecta. *Hum Mutat* 2012;33:91–4.
- 83 Kivirikko S, McGrath JA, Pulkkinen L, Uitto J, Christiano AM. Mutational hotspots in the LAMB3 gene in the lethal (Herlitz) type of junctional epidermolysis bullosa. *Hum Mol Genet* 1996;5:231–7.
- 84 Ogawa T, Tomatsu S, Fukuda S, Yamagishi A, Rezvi GMM, Sukegawa K, Kondo N, Suzuki Y, Shimozawa N, Orii T. Mucopolysaccharidosis IVA: screening and identification of mutations of the N-acetylgalactosamine-6-sulfate sulfatase gene. *Hum Mol Genet* 1995;4:341–9.
- 85 Pasmanij AM, Pas HH, Jansen GH, Lemmink HH, Jonkman MF. Localized and generalized forms of blistering in junctional epidermolysis bullosa due to COL17A1 mutations in the Netherlands. *Br J Dermatol* 2007;156:861–70.
- 86 Murrell DF, Pasmanij AM, Pas HH, Marr P, Klingberg S, Pfendner E, Uitto J, Sadowski S, Collins F, Widmer R, Jonkman MF. Retrospective diagnosis of fatal BP180-deficient non-Herlitz junctional epidermolysis bullosa suggested by immunofluorescence (IF) antigen-mapping of parental carriers bearing enamel defects. *J Invest Dermatol* 2007;127:1772–5.
- 87 Floeth M, Bruckner-Tuderman L. Digenic junctional epidermolysis bullosa: mutations in COL17A1 and LAMB3 genes. *Am J Hum Genet* 1999;65:1530–7.
- 88 Beyeler M, Schild C, Lutz R, Chiquet M, Trueb B. Identification of a fibronectin interaction site in the extracellular matrix protein ameloblastin. *Exp Cell Res* 2010;316:1202–12.
- 89 Zhang X, Diekwich TG, Luan X. Structure and function of ameloblastin as an extracellular matrix protein: adhesion, calcium binding, and CD63 interaction in human and mouse. *Eur J Oral Sci* 2011;119(Suppl 1):270–9.
- 90 Polok B, Escher P, Ambresin A, Chouery E, Bolay S, Meunier I, Nan F, Hamel C, Munier FL, Thilo B, Mégarbané A, Schorderet DF. Mutations in CNNM4 cause recessive cone-rod dystrophy with amelogenesis imperfecta. *Am J Hum Genet* 2009;84:259–65.
- 91 Mendelsohn NJ, Wood T, Olson RA, Temme R, Hale S, Zhang H, Read L, White KK. Spondyloepiphyseal dysplasias and bilateral legg-calv  -perthes disease: diagnostic considerations for mucopolysaccharidoses. *JIMD Rep* 2013;11:125–32.
- 92 Gardner DG. The dental manifestations of the Morquio syndrome (mucopolysaccharidosis type IV). A diagnostic aid. *Am J Dis Child* 1975;129:1445–8.
- 93 Levin LS, Jorgenson RJ, Salinas CF. Oral findings in the Morquio syndrome (mucopolysaccharidosis IV). *Oral Surg Oral Med Oral Pathol* 1975;39:390–5.
- 94 Rølling I, Clausen N, Nyvad B, Sindet-Pedersen S. Dental findings in three siblings with Morquio's syndrome. *Int J Paediatr Dent* 1999;9:219–24.
- 95 Liu F, Chu EY, Watt B, Zhang Y, Gallant NM, Andl T, Yang SH, Lu M-M, Piccolo S, Schmidt-Ullrich R, Taketo MM, Morrissey EE, Atit R, Dlugosz AA, Millar SE. Wnt/beta-catenin signaling directs multiple stages of tooth morphogenesis. *Dev Biol* 2008;313:210–24.
- 96 de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, Vulto-Van Silfhout AT, Koolen DA, De Vries P, Gilissen C, Del Rosario M, Hoischen A, Scheffer H, De Vries BBA, Brunner HG, Veltman JA, Vissers LELM. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* 2012;367:1921–9.
- 97 Tucci V, Kleefstra T, Hardy A, Heise I, Maggi S, Willemsen MH, Hilton H, Esapa C, Simon M, Buenavista M-T, McGuffin LJ, Vizor L, Dodero L, Tsafaris S, Romero R, Nillesen WN, Vissers LELM, Kempers MJ, Vulto-Van Silfhout AT, Igbal Z, Orlando M, Maccione A, Lassi G, Farisello P, Contestabile A, Tinarelli F, Nieuw T, Raimondi A, Greco B, Cantatore D, Gasparini L, Berdonini L, Bifone A, Gozzi A, Wells S, Nolan PM. Dominant beta-catenin mutations cause intellectual disability with recognizable syndromic features. *J Clin Invest* 2014;124:1468–82.
- 98 McKnight DA, Suzanne Hart P, Hart TC, Hartsfield JK, Wilson A, Wright JT, Fisher LW. A comprehensive analysis of normal variation and disease-causing mutations in the human DSPP gene. *Hum Mutat* 2008;29:1392–404.
- 99 Gregory-Evans CY, Moosajee M, Hodges MD, Mackay DS, Game L, Vargesson N, Bloch-Zupan A, Ruschendorf F, Santos-Pinto L, Wackens G, Gregory-Evans K. SNP genome scanning localizes oto-dental syndrome to chromosome 11q13 and microdeletions at this locus implicate FGF3 in dental and inner-ear disease and FADD in ocular coloboma. *Hum Mol Genet* 2007;16:2482–93.
- 100 Grozeva D, Carsi K, Spasic-Boskovic O, Parker MJ, Archer H, Firth HV, Park S-M, Canham N, Holder SE, Wilson M, Hackett A, Field M, Floyd JAB, Hurles M, Raymond FL. De novo loss-of-function mutations in SETD5, encoding a methyltransferase in a 3p25 microdeletion syndrome critical region, cause intellectual disability. *Am J Hum Genet* 2014;94:618–24.
- 101 Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010;42:790–3.
- 102 Chan HC, Estrella NM, Milkovich RN, Kim JW, Simmer JP, Hu JC. Target gene analyses of 39 amelogenesis imperfecta kindreds. *Eur J Oral Sci* 2011;119(Suppl 1):311–23.
- 103 Gasse B, Karayigit E, Mathieu E, Jung S, Garret A, Huckert M, Morkmued S, Schneider C, Vidal L, Hemmerle J, Sire J-Y, Bloch-Zupan A. Homozygous and compound heterozygous MMP20 mutations in amelogenesis imperfecta. *J Dent Res* 2013;92:598–603.
- 104 Wood TC, Harvey K, Beck M, Burin MG, Chien Y-H, Church HJ, D'Almeida V, van Diggelen OP, Fietz M, Giugliani R, Harmatz P, Hawley SM, Hwu W-L, Ketteridge D, Lukacs Z, Miller N, Pasquali M, Schenone A, Thompson JN, Tylee K, Yu C, Hendriksz CJ. Diagnosing mucopolysaccharidosis IVA. *J Inher Metab Dis* 2013;36:293–307.
- 105 Arzoo PS, Klar J, Bergendal B, Norderyd J, Dahl N. WNT10A mutations account for (1/4) of population-based isolated oligodontia and show phenotypic correlations. *Am J Med Genet A* 2014;164A:353–9.
- 106 Mostowska A, Biedziak B, Zadurska M, Dunin-Wilczynska I, Lianeri M, Jagodzinski PP. Nucleotide variants of genes encoding components of the Wnt signalling pathway and the risk of non-syndromic tooth agenesis. *Clin Genet* 2013;84:429–40.

Supplementary Table 1. HUGO Gene Nomenclature Committee (HGNC) gene symbols of genes included in the targeted next-generation sequencing assay

Genes included in the diagnostic sub-panel

AIRE ALPL ALX3 AMBN AMELX AMELY AMTN ANKRD11 ANTXR1 APC ATR AXIN2 BCOR BLM BMP4 C4ORF26 CA2 CCBE1 CCDC8 CEP152 CFDP1 CHD7 CLCN7 CLDN1 CNM4 COL17A1 COL1A1 COL1A2 COL5A1 COL5A2 COL7A1 COL9A1 COL9A2 COX7B CREBBP CRTAP CTNNB1 CTSC CTSK CUL7 CYP27B1 DLX3 DSP DSPP EDA EDAR EDARADD ELN ENAM EP300 EVC EVC2 EXT1 FAM20A FAM83H FERMT3 FGF10 FGF23 FGF3 FGF8 FGFR1 FGFR2 FGFR3 FKBP10 FOXC1 GALNT3 GAS1 GDF5 GJA1 GLA GLI2 GNAS GTF2I GTF2IRD1 IFT43 IKBKG IL11RA IRF6 KAL1 KANSL1 KDM6A KLK4 KMT2D KRT14 KRT5 LAMA3 LAMB3 LAMC2 LEF1 LEMD3 LEPRE1 LIMK1 LRP5 LTBP3 MID1 MMP1 MMP20 MSX1 MSX2 MUTYH NFKBIA NHS NSD1 ODSL1 ODAM OFD1 ORAI1 OSTM1 PAX9 PCNT PHEX PITX2 PLEC PLEKHM1 PLXNB2 POLR1C POLR1D POLR3A POLR3B PORCN PPIB PRKAR1A PROK2 PROKR2 PTCH1 PTCH2 PTH1R PVRL1 PVRL4 RAB23 RAI1 RASGRP2 RBBP8 RECQL4 RFC2 ROGDI ROR2 RUNX2 SALL4 SAT1 SATB2 SEC23A SERPINF1 SERPINH1 SH3BP2 SHH SIX3 SLC24A4 SLC26A2 SMARCAL1 SMOC2 SP6 SP7 SPARC SPARCL1 SUFU SUMO1 TBX22 TCIRG1 TCOF1 TGIF1 TNFRSF11A TNFRSF11B TP63 TRPS1 TSC1 TSC2 TUFT1 UBR1 VDR WDR72 WNT10A WNT5A WRN ZIC2

Genes included in the discovery sub-panel

ACVR1 ACVR2A ACVR2B ADAMTS10 ADAMTS2 ADARA AGPAT2 AGXT AHCY AIP ALDH3A2 ALKBH1 AMER1 ANKH AP2B1 AP3B1 APAF1 APCDD1 ARHGAP6 ARSB ASXL1 ATP6V0A2 ATP7A ATRIP ATRX B3GALTL B3GAT3 B4GALT7 BANF1 BARX1 BCL11B BGLAP BGN BMP2 BMP7 BMPR1A BNC2 BOC BRAF CACNA1C CARD9 CASK CASP7 CCL2 CD96 CDC42 CDC6 CDH1 CDH23 CDH3 CDKN1A CDKN1C CDON CDSN CENPJ CHRD CHST3 CHSY1 CHUK CIB2 CLCN5 CLEC7A COL10A1 COL11A1 COL11A2 COL2A1 COL3A1 COX4I2 CRISPLD2 CRK CSF1 CTGF DCAF17 DCN DFNB31 DHCR24 DHCR7 DHODH DKC1 DKK1 DKK4 DLG1 DLL1 DLX1 DLX2 DLX4 DLX5 DLX6 DMP1 DPYSL4 DSC3 DSG4 EDNRB EFNB1 EHMT1 EXT2 EXTL3 EYA1 FAM20C FBN1 FERMT1 FGD1 FGF13 FGF18 FGF20 FGF4 FGF9 FKBP6 FLNA FLNB FMOD FOS FOXE1 FOXF2 FOXO1 FRAS1 FREM2 FRZB FST FUZ FZD1 FZD2 FZD6 G6PC3 GAB1 GAD1 GALNS GDNF GJB6 GLB1 GLI1 GLI3 GNPTAB GORAB GPC3 GRB2 GSC GSK3B GUSB HAND1 HAND2 HCCS HDAC4 HGSNAT HHAT HOXC13 HOXD13 HR HRAS HYAL IBSP ICAM1 IDS IDUA IFT122 IFT88 IGF1 IHH IL17F IL17RA INHBA INHBB INPP5E INSR ISL1 ITGA11 ITGA6 ITGAV ITGB2 ITGB4 ITGB6 ITGB8 JAG1 JAG2 KAT6B KAZN KCNJ2 KIF7 KISS1 KISS1R KL KRAS KRT1 KRT10 KRT16 KRT17 KRT6A KRT6B KRT74 KRT83 KRT9 LAMA5 LHX6 LHX8 LIFR LIPH LMNA LOR LPAR6 LRP4 LRP6 LTBP2 LUM LUZP1 MAP2K1 MAP2K2 MAP3K11 MASP1 MBTPS2 MED12 MED25 MEPE MGP MITF MMP14 MMP16 MMP2 MMP3 MMP9 MN1 MNT MSC MVP MY07A NAGLU NCOA2 NCOR1 NELF NF2 NFE2L2 NFIC NIPBL NKX2-3 NKX3-2 NLRP1 NOG NOP10 NOTCH1 NOTCH2 NOTCH3 OCRL ORC1 OSR2 PAF1 PCDH15 PDGFA PDGFRA PDS5A PDS5B PHC1 PIGL PITX1 PKDCC PKP1 PLCD1 PLG PLOD1 PLOD3 PLXNA1 PLXNA2 PLXNA3 PLXNA4 PLXNB1 PLXNB3 PLXNC1 PLXND1 POLD1 POSTN POU1F1 PRDM1 PRDM16 PRKCI PRRX1 PRRX2 PTHLH PTPN11 PTPRF PTPRS PVRL3 RAPSN RBL1 RBL2 RBM28 RFNG RGS2 RIN2 RMRP RPS6KA3 RSP02 RSP04 SCARF2 SFN SGSH SH3PXD2B SHOX2 SIM2 SLC20A2 SLC32A1 SLC34A2 SLC35B2 SLC35C1 SLC39A13 SLC4A2 SLC4A4 SLC4A5 SMAD2 SMAD3 SMAD5 SMG1 SMO SMPD3 SNAI1 SNAI2 SOS1 SOST SOSTDC1 SOX10 SOX11 SOX18 SOX2 SOX3 SOX5 SP3 SPP1 SPRY2 SPRY4 SQSTM1 SSTR5 ST14 STAG1 STAT1 STAT3 STIM1 SUOX TAB2 TACR3 TBCE TBX1 TBX10 TBX15 TBX2 TBX3 TCF21 TCP1 TERC TERT TFAP2A TFIP11 TGFA TGFB2 TGFB3 TGFBR1 TGFBR2 THRA TINF2 TMCO1 TMEM107 TNFRSF19 TNFSF11 TRAF6 TRIM37 TRIP11 TRPV3 TSHZ1 TWIST1 TWIST2 UBB USH1C USH1G USH2A WDR19 WDR35 WHSC1 WNT1 WNT10B WNT3 WNT4 WNT6 WNT7B ZEB1 ZEB2 ZMPSTE24 ZNF469

Genes added to the discovery sub-panel in the screening assay

ADAM10 AKAP9 ANTXR2 BAZ1B BBX BMP1 CHPF CSR2BP FAM111A FAM111B FAM73B GALC HMX3 IFITM5 IFT20 ITGA3 ITGB1 KDM4B LTBP1 NSUN2 NTRK1 RHOBTB3 SLC25A21 TMEM38B UBE3B

Supplementary Table 2. Details of the targeted capture design

		v1.0	v2.0
Number of loci targeted	Total	560	585
	Diagnosis sub-panel	175	175
	Discovery sub-panel	385	410
Size of targeted regions (Mb)	Total	2.36	2.47
	Diagnosis sub-panel	0.81	0.81
	Discovery sub-panel	1.5	1.66

Supplementary Table 3. Known variants in samples V1.01-V1.08 identified with v1.0 of the panel

ID	Gene	Transcript	c.	p.	Inheritance	Reference
V1.01	<i>LTBP3</i>	NM_001130144.2	c.[1531+1G>A];[421C>T]	p.[Q141*];[p.0?]	AR	[1]
V1.02	<i>FAM20A</i>	NM_017565.3	c.[34_35del];[612del]	p.[L12Afs*67];[L205Cfs*11]	AR	[2]
V1.03	<i>DSPP</i>	NM_014208.3	c.[3676del];[=]	p.[S1226Afs*88];[=]	AD	-
V1.04	<i>AMELX</i>	NM_182680.1	c.[11G>A];[0]	p.[W4*];[0]	X-linked	-
V1.05	<i>ENAM</i>	NM_031889.2	c.[588+1del]	p.[N197Ifs*81]	AD	-
V1.06	<i>CNNM4</i>	NM_020184.3	c.[586T>C];[586T>C]	p.[S196P];[S196P]	AR	[3]
V1.07	<i>MMP20</i>	NM_004771.3	c.[954-2A>T];[954-2A>T]	p.[0?];[0?]	AR	-
V1.08	Contiguous gene deletion		hg19 chrX:(?_185958)_(12815845-?)del [*hg19 chrX:10417360-8699939del]	p.[0?];[=]	AD	-

Supplementary Table 4. Sequencing statistics from v1.0 of the gene panel on samples V1.01-V1.16

ID	Total sequenced nucleotides	Total passed filter aligned nucleotides	Total passed filter aligned nucleotides in targeted regions ^o	Mean coverage*			% bases ≥ 1X*			% bases ≥ 20X*			% bases ≥ 50X*			% bases ≥ 100X*		
	Total	Diag. sub-panel	Disc. sub-panel	Total	Diag. sub-panel	Disc. sub-panel	Total	Diag. sub-panel	Disc. sub-panel	Total	Diag. sub-panel	Disc. sub-panel	Total	Diag. sub-panel	Disc. sub-panel	Total	Diag. sub-panel	Disc. sub-panel
V1.01	2529393600	1935516105	973675975	410.57	443.57	393.18	99.5	99.5	99.6	97.5	98.8	96.8	94.5	97.8	92.8	88.3	95	84.8
V1.02	2441063200	1869714700	937342410	395.32	433.5	375.19	99.6	99.3	99.7	97.9	98.8	97.5	95.6	98.2	94.3	90.1	95.9	87
V1.03	2152958600	1631544311	855687555	360.71	400.9	339.54	99.7	99.6	99.8	98.2	99.1	97.7	95.7	98.4	94.3	89.6	96.2	86.2
V1.04	1703053800	1303350887	686090778	289.28	325.36	270.27	99.6	99.5	99.7	97.7	98.9	97.1	94.4	98	92.5	86.2	94.4	82
V1.05	1641523000	1195222257	637282471	268.63	298.28	253.01	99.7	99.5	99.7	97.6	98.9	97	93.8	97.9	91.7	84.3	93.2	79.7
V1.06	2100703400	1579542394	853771624	360.02	398.63	339.67	99.6	99.3	99.8	98.1	98.9	97.6	95.7	98.2	94.4	89.9	96	86.7
V1.07	2869433600	2167359316	1072784699	452.21	503.56	425.16	99.7	99.4	99.8	98.5	99	98.2	96.9	98.6	96	93.2	97.5	90.9
V1.08	2550773800	1933898916	1026048161	432.66	472.66	411.58	99.6	99.4	99.7	98.1	98.9	97.7	95.9	98.3	94.7	90.7	96.3	87.8
V1.09	2354231600	1774468518	909271752	383.38	425.66	361.1	99.7	99.4	99.8	98.2	99	97.9	96.2	98.5	94.9	90.9	96.6	87.9
V1.10	1713496800	1283336880	691905675	291.67	321.11	276.16	99.7	99.5	99.8	97.6	99	96.9	93.9	97.9	91.8	85.2	93.8	80.7
V1.11	2530456400	1906056269	1009463211	425.6	476.94	398.54	99.7	99.5	99.8	98.5	99.1	98.1	96.7	98.6	95.6	92.2	97.2	89.6
V1.12	2486134200	1883955509	1032670398	435.41	491.12	406.05	99.7	99.5	99.8	98.4	99.1	98.1	96.7	98.7	95.7	92.5	97.3	89.9
V1.13	2182365400	1613504161	835260345	352.15	390.72	331.83	99.7	99.5	99.8	98.1	99	97.6	95.4	98.3	93.8	88.9	95.8	85.3
V1.14	2284365400	1691327624	866816596	365.33	407.21	343.26	99.7	99.5	99.8	98.3	99.1	97.8	95.9	98.5	94.6	90.2	96.5	86.8
V1.15	2281142800	1687807698	912197621	384.7	423.73	364.13	99.5	99.3	99.6	97.7	98.7	97.2	95.2	97.9	93.8	89.5	95.3	86.4
V1.16	1250639000	944935101	538571957	227.08	251.73	214.1	99.6	99.4	99.7	97.2	98.7	96.4	92.5	97.1	90	81.1	90.9	76
Mean	2191983413	1650096290	864927577	364.67	404.04	343.92	99.64	99.44	99.74	97.98	98.94	97.48	95.31	98.18	93.81	88.93	95.49	85.48

^otargeted regions = baited regions + 15 bp on either side, *includes only reads with MQ > 30, Diag. = Diagnostic, Disc. = Discovery

Supplementary Table 5. Regions with an average per sample coverage <20X in v1.0 of the panel

Chr	Region Start	Region End	%GC	Region Length	Average Coverage Per Individual	Gene	Position
Diagnostic sub-panel							
7	74150814	74151053	42.26%	239	0	<i>GTF2I</i>	exon18
7	74163577	74163816	45.61%	239	0	<i>GTF2I</i>	exon26
7	74168150	74168389	48.54%	239	0	<i>GTF2I</i>	exon30
X	153786686	153786925	61.51%	239	0	<i>IKBKG</i>	exon4
X	153788578	153788817	73.22%	239	0	<i>IKBKG</i>	exon5
X	153789831	153790070	59.41%	239	0	<i>IKBKG</i>	exon6
X	153790977	153791216	62.76%	239	0	<i>IKBKG</i>	exon7
X	153791725	153791964	63.18%	239	0	<i>IKBKG</i>	exon8
X	153792115	153792294	63.13%	179	0	<i>IKBKG</i>	exon9
7	74163305	74163484	37.99%	179	0.033170391	<i>GTF2I</i>	exon25
7	74165631	74165810	39.11%	179	0.065293296	<i>GTF2I</i>	exon27
7	74169765	74169944	32.96%	179	0.065293296	<i>GTF2I</i>	exon31
X	153792485	153792724	64.02%	239	0.132583682	<i>IKBKG</i>	exon10/3'UTR
7	74143059	74143298	47.28%	239	0.484309623	<i>GTF2I</i>	exon13
7	74144510	74144689	35.75%	179	0.689944134	<i>GTF2I</i>	exon14
18	59992470	59992709	83.68%	239	0.966788703	<i>TNFRSF11A</i>	5'UTR/exon1
4	2819914	2820153	85.36%	239	1.021705021	<i>SH3BP2</i>	5'UTR/exon1
4	5713075	5713314	82.43%	239	2.111140167	<i>EVC</i>	5'UTR/exon1
11	68080108	68080347	82.01%	239	3.053870293	<i>LRP5</i>	5'UTR/exon1
17	17713200	17713379	73.74%	179	7.438896648	<i>RAI1</i>	exon6/3'UTR
9	137533968	137534207	82.01%	239	10.91082636	<i>COL5A1</i>	5'UTR/exon1
2	220417548	220417787	75.73%	239	13.43122385	<i>OBSL1</i>	exon17
1	110612916	110613275	78.33%	359	19.7992688	<i>ALX3</i>	5'UTR/exon1
Discovery sub-panel							
9	140513401	140513580	87.15%	179	0.061103352	<i>EHMT1</i>	5'UTR/exon1
14	105634635	105634814	86.59%	179	0.274790503	<i>JAG2</i>	5'UTR/exon1
8	42995507	42995806	81.27%	299	0.327759197	<i>HGSNAT</i>	5'UTR/exon1
1	15250815	15251114	83.61%	299	0.35451505	<i>KAZN</i>	5'UTR/exon1
18	72923762	72923941	82.12%	179	0.423882682	<i>TSHZ1</i>	intron1
9	139440118	139440297	78.21%	179	0.468924581	<i>NOTCH1</i>	5'UTR
9	101867416	101867655	84.10%	239	0.486401674	<i>TGFBR1</i>	5'UTR/exon1
X	153784336	153784635	62.54%	299	0.525710702	<i>IKBKG</i>	exon3
12	112856833	112857012	77.10%	179	0.808659218	<i>PTPN11</i>	5'UTR/exon1
7	150760182	150760361	75.42%	179	0.992667598	<i>SLC4A2</i>	5'UTR/exon1
5	177027147	177027326	81.56%	179	1.301326816	<i>B4GALT7</i>	5'UTR/exon1
9	132428096	132428455	82.73%	359	1.313022284	<i>PRRX2</i>	5'UTR/exon1
9	140353399	140353638	81.17%	239	1.360355649	<i>NSMF</i>	5'UTR/exon1
1	3310967	3311146	69.83%	179	1.392458101	<i>PRDM16</i>	intron4
21	38119940	38120419	81.42%	479	1.4375	<i>SIM2</i>	exon11/3'UTR
3	38495750	38495929	81.56%	179	1.516061453	<i>ACVR2B</i>	5'UTR/exon1
1	22469257	22469496	82.01%	239	1.531380753	<i>WNT4</i>	5'UTR/exon1
9	96717167	96717466	78.60%	299	1.567934783	<i>BARX1</i>	5'UTR/exon1
1	110453575	110453754	83.80%	179	1.930865922	<i>CSF1</i>	5'UTR/exon1
5	134369367	134369606	76.99%	239	1.947437238	<i>PITX1</i>	5'UTR/exon1
6	44225052	44225231	78.77%	179	2.208100559	<i>SLC35B2</i>	5'UTR/exon1
18	10454568	10454747	78.77%	179	2.311452514	<i>APCDD1</i>	5'UTR
1	120534005	120534184	41.90%	179	2.470670391	<i>NOTCH2</i>	intron4

Chr	Region Start	Region End	%GC	Region Length	Average Coverage Per Individual	Gene	Position
17	78193949	78194188	77.82%	239	2.544979079	<i>SGSH</i>	5'UTR/exon1
6	132272190	132272369	79.33%	179	2.602304469	<i>CTGF</i>	5'UTR
20	10654018	10654257	77.41%	239	2.612186192	<i>JAG1</i>	5'UTR/exon1
4	980832	981071	79.50%	239	2.720188285	<i>IDUA</i>	5'UTR/exon1
11	130029795	130030034	78.66%	239	2.83289749	<i>ST14</i>	5'UTR/exon1
17	80009329	80009688	81.34%	359	2.934192201	<i>RFNG</i>	5'UTR/exon1
5	178772140	178772379	80.75%	239	2.951621339	<i>ADAMTS2</i>	5'UTR/exon1
11	46939815	46939994	78.21%	179	2.981494413	<i>LRP4</i>	5'UTR/exon2
19	7293733	7293972	78.66%	239	2.993723849	<i>INSR</i>	5'UTR/exon1
X	128674346	128674525	76.54%	179	3.069832402	<i>OCRL</i>	5'UTR/exon1
5	179248469	179248648	70.95%	179	3.455656425	<i>SQSTM1</i>	intron1
19	5229449	5229748	80.27%	299	3.777801003	<i>PTPRS</i>	exon15
X	20284626	20284805	77.65%	179	4.146997207	<i>RPS6KA3</i>	5'UTR
6	10412805	10413044	74.48%	239	4.197960251	<i>TFAP2A</i>	intron1
7	558517	558696	75.98%	179	4.499301676	<i>PDGFA</i>	5'UTR/exon1
17	73749801	73750100	78.93%	299	4.509824415	<i>ITGB4</i>	exon33
20	60941973	60942332	78.27%	359	4.547179666	<i>LAMA5</i>	5'UTR/exon1
19	15311541	15311780	80.33%	239	4.950052301	<i>NOTCH3</i>	5'UTR/exon1
7	74162283	74162522	35.98%	239	4.967573222	<i>GTF2I</i>	exon24
3	184098061	184098300	79.50%	239	5.02248954	<i>CHRD</i>	5'UTR/exon1
7	559571	559870	77.59%	299	5.138586957	-	upstream
15	68724290	68724469	76.54%	179	5.259078212	<i>ITGA11</i>	5'UTR/exon1
7	140624315	140624554	77.41%	239	5.379445607	<i>BRAF</i>	5'UTR/exon1
9	124990858	124991097	77.82%	239	5.442730126	<i>LHX6</i>	5'UTR/exon1
22	17565931	17566170	77.82%	239	5.686453975	<i>IL17RA</i>	5'UTR/exon1
17	40688242	40688721	79.12%	479	6.322286013	<i>NAGLU</i>	5'UTR/exon1
20	30946517	30946696	76.54%	179	6.405726257	<i>ASXL1</i>	5'UTR/exon1
19	41836916	41837155	74.06%	239	6.517259414	<i>TGFB1</i>	5'UTR/exon1
3	110790825	110791064	73.64%	239	6.552824268	<i>PVRL3</i>	5'UTR/exon1
13	31774167	31774346	77.65%	179	6.819832402	<i>B3GALT1</i>	5'UTR/exon1
7	74157704	74157943	34.73%	239	6.993200837	<i>GTF2I</i>	exon20
14	95235965	95236384	76.85%	419	7.16900358	<i>GSC</i>	5'UTR/exon1
5	1294845	1295144	79.60%	299	7.27132107	<i>TERT</i>	5'UTR/exon1
16	18900651	18900950	40.13%	299	7.477842809	<i>SMG1</i>	exon6
16	68771253	68771432	76.54%	179	7.535963687	<i>CDH1</i>	5'UTR/exon1
2	172291031	172291270	78.66%	239	7.769089958	<i>DCAF17</i>	5'UTR/exon1
2	174829087	174829326	73.22%	239	8.390167364	<i>SP3</i>	exon2
10	134000497	134000676	77.65%	179	8.491620112	<i>DPYSL4</i>	5'UTR/exon1
10	31608103	31608282	68.72%	179	8.50349162	<i>ZEB1</i>	5'UTR/exon1
6	132271871	132272170	74.58%	299	8.639423077	<i>CTGF</i>	exon2
2	174828430	174828669	76.99%	239	8.67834728	<i>SP3</i>	exon3
19	4123707	4123946	77.41%	239	8.754445607	<i>MAP2K2</i>	5'UTR/exon1
7	100860381	100860620	74.90%	239	8.780857741	<i>PLOD3</i>	5'UTR/exon1
11	65381392	65381571	73.18%	179	9.057960894	<i>MAP3K11</i>	5'UTR
19	8649739	8649978	72.38%	239	9.0666841	<i>ADAMTS10</i>	exon25
9	124989604	124989843	73.22%	239	9.203713389	<i>LHX6</i>	exon2
3	38070924	38071103	73.74%	179	9.818435754	<i>PLCD1</i>	5'UTR/exon1
20	60884342	60884581	69.87%	239	9.882060669	<i>LAMA5</i>	exon80
16	88923106	88923345	76.99%	239	9.935669456	<i>GALNS</i>	5'UTR/exon1
7	557977	558216	75.73%	239	10.41082636	<i>PDGFA</i>	intron1

Chr	Region Start	Region End	%GC	Region Length	Average Coverage Per Individual	Gene	Position
2	121103719	121104258	75.88%	539	10.43552876	<i>INHBB</i>	5'UTR/exon1
11	69589473	69589892	77.57%	419	10.58412888	<i>FGF4</i>	5'UTR/exon1
16	68679476	68679715	69.87%	239	10.76281381	<i>CDH3</i>	exon2
2	178129192	178129371	74.86%	179	10.77513966	<i>NFE2L2</i>	5'UTR/exon1
5	78280706	78281125	76.37%	419	10.91139618	<i>ARSB</i>	5'UTR/exon1
22	20791835	20792074	79.92%	239	11.20737448	<i>SCARF2</i>	5'UTR/exon1
2	219724681	219724920	76.15%	239	11.67651674	<i>WNT6</i>	5'UTR/exon1
16	53468439	53468738	73.91%	299	12.00585284	<i>RBL2</i>	5'UTR/exon1
22	20779637	20780596	76.12%	959	12.44584202	<i>SCARF2</i>	exon11/3'UTR
3	129159104	129159283	65.36%	179	13.49965084	<i>IFT122</i>	5'UTR/exon1
7	128828948	128829367	77.33%	419	13.60739857	<i>SMO</i>	5'UTR/exon1
20	60897644	60897883	69.87%	239	13.80334728	<i>LAMA5</i>	exon46
X	153991159	153991338	72.07%	179	13.82087989	<i>DKC1</i>	5'UTR/exon1
20	62680481	62680900	77.33%	419	14.07741647	<i>SOX18</i>	5'UTR/exon1
7	556959	557198	74.48%	239	14.16971757	<i>PDGFA</i>	exon2
4	996476	997004	78.03%	528	14.34824811	<i>IDUA</i>	exon9/10
22	46367956	46368195	75.31%	239	14.42024059	<i>WNT7B</i>	intron1
20	62679467	62680366	75.97%	899	15.1881257	<i>SOX18</i>	exon2/3'UTR
X	54521532	54521891	73.61%	359	17.47423398	<i>FGD1</i>	5'UTR/exon1
17	78187524	78187763	70.29%	239	17.77301255	<i>SGSH</i>	exon6
1	11994755	11994994	74.48%	239	17.78242678	<i>PLOD1</i>	5'UTR/exon1
18	10454840	10455079	75.00%	239	17.79550209	<i>APCDD1</i>	5'UTR/exon1
2	42275299	42276018	77.78%	719	17.81389082	<i>PKDCC</i>	5'UTR/exon1
5	52776544	52776783	72.50%	239	18.40115063	<i>FST</i>	5'UTR/exon1
19	5250930	5251109	58.89%	179	18.53840782	<i>PTPRS</i>	intron 9
1	2985752	2985931	70.00%	179	18.89385475	<i>PRDM16</i>	5'UTR/exon1
11	2905860	2906759	77.44%	899	18.92575083	<i>CDKN1C</i>	5'UTR/exon1
19	3366560	3366739	72.22%	179	19.40851955	<i>NFIC</i>	5'UTR/exon1
11	65374713	65375012	73.33%	299	19.76066054	<i>MAP3K11</i>	exon5

Supplementary Table 6. Variant detection statistics in samples V1.01-V1.16 with v1.0 of the gene panel

ID	Number of variants identified	Number of SNPs	Number of INDELS	Number of filtered*					Number of filtered variants annotated as splice mutations
				variants with allele frequency < 1% in dbSNP137, 1000G, EVS and in-house database	Number of filtered variants annotated as missense	Number of filtered variants annotated as frameshift	Number of filtered variants annotated as nonsense		
V1.01	2547	2138	409	74	22	0	1	1	
V1.02	2572	2127	445	64	16	3	0	0	
V1.03	3059	2560	499	180	40	1	0	0	
V1.04	2704	2260	444	68	19	0	1	1	
V1.05	3077	2582	495	189	31	3	0	0	
V1.06	2671	2225	446	79	28	0	0	0	
V1.07	2553	2134	419	77	13	1	0	1	
V1.08	2553	2144	409	72	20	0	0	1	
V1.09	2622	2186	436	70	13	0	1	0	
V1.10	2441	2018	423	69	22	0	0	0	
V1.11	2701	2263	438	81	21	1	1	0	
V1.12	2563	2130	433	69	16	0	0	0	
V1.13	2581	2149	432	90	22	0	0	0	
V1.14	2538	2091	447	76	18	0	1	0	
V1.15	2588	2158	430	59	15	0	0	0	
V1.16	2599	2188	411	72	14	0	1	0	
Mean	2648.06	2209.56	438.50	86.81	20.63	0.56	0.38	0.25	

* filtered: variants that passed the following criteria: allele frequency < 1% in dbSNP137, 1000 genomes, Exome Variant Server, and in-house database; °includes only reads with MQ > 30

Supplementary Table 7. Sequencing statistics for patients v2.01-v2.95 sequenced with v2.0 of the gene panel

ID	Total sequenced nucleotides	Total passed filter aligned nucleotides	Total passed filter aligned nucleotides in targeted regions	Mean coverage	% bases ≥ 1X°	% bases ≥ 20X°	% bases ≥ 50X°	# variants	# SNP	# Indel	# filtered* variants	# filtered missense	# filtered frameshift	# filtered nonsense	# filtered splice
V2.01	848050800	736028984	277592817	107.8	99.8	96	81.5	2780	2262	518	70	13	0	0	1
V2.02	960169200	860522137	327213166	127.05	99.8	96.4	85	2876	2364	512	83	14	3	0	0
V2.03	747160000	667348632	261857870	101.64	99.8	95.1	78.5	2642	2167	475	65	11	1	0	1
V2.04	800411200	725089453	282525424	109.54	99.8	95.4	80.3	2869	2349	520	74	11	0	0	0
V2.05	746819200	684586911	269867229	104.51	99.8	94.2	77.3	2670	2215	455	110	23	3	0	0
V2.06	692747000	634750100	264984723	102.61	99.7	93.5	75.7	2874	2357	517	128	19	5	1	0
V2.07	887134800	756202602	284776259	110.64	99.8	95.7	81.5	2726	2208	518	71	12	3	0	2
V2.08	922939800	823410379	320339971	124.17	99.8	96.2	83.5	2789	2284	505	86	12	0	0	0
V2.09	776923600	651158111	299181653	115.81	99.8	96	83.1	2741	2227	514	73	6	3	1	1
V2.10	644774400	545336576	262728853	101.82	99.7	95	79.1	2769	2278	491	72	15	0	0	2
V2.11	757023200	659217579	313145842	121.23	99.8	96.4	84.7	2790	2280	510	78	13	0	0	1
V2.12	855912600	779255895	359977159	139.45	99.9	96.9	87.4	2823	2316	507	41	2	0	0	1
V2.13	1502519800	1282612834	528188469	204.98	99.9	98.2	94	2846	2328	518	82	18	3	0	2
V2.14	1149269200	1015333953	438537896	169.99	99.8	97.8	91.6	2738	2234	504	77	15	0	0	2
V2.15	1004121800	895602006	407007571	157.78	99.9	97.6	90.3	2766	2265	501	71	13	3	0	1
V2.16	958794200	806152575	341959344	132.41	99.8	96.5	85.4	2759	2260	499	90	23	2	0	1
V2.17	2123494400	1875903407	809214774	313.75	99.9	98.9	96.4	3745	3122	623	128	26	0	0	2
V2.18	1048638000	926159642	403679470	156.47	99.9	97.5	89.4	2766	2271	495	70	19	0	1	0
V2.19	1059056800	945324906	402670360	156.13	99.8	97.1	88.6	2889	2366	523	71	14	0	0	1
V2.20	895689600	814345529	372856997	144.44	99.8	97.2	88.7	2802	2296	506	64	13	4	1	1
V2.21	2664350600	2397448572	1007063992	390.46	99.9	99.1	97.4	3725	3090	635	169	34	3	1	1
V2.22	1935887200	1685866367	758461285	293.93	99.9	98.9	96.7	3641	3018	623	101	17	2	0	1
V2.23	1188179000	1077109416	459677384	178.31	99.9	97.5	90.7	2792	2278	514	66	10	3	0	1
V2.24	1398400200	1213462486	510731305	198.47	99.9	98.3	94.2	2801	2288	513	87	19	3	0	1
V2.25	1495761200	1364107809	591255764	229.51	99.9	98.4	94.9	2722	2204	518	85	9	0	0	1
V2.26	988928400	870596072	399228912	154.73	99.9	97.5	89.6	2805	2286	519	73	9	0	1	3
V2.27	568525800	457626650	227965737	88.22	99.8	93.9	74.1	2673	2194	479	68	10	0	0	1
V2.28	1046138800	946606434	446714024	172.78	99.9	97.9	91.7	2781	2273	508	78	10	3	1	3
V2.29	574351000	513661968	244780596	94.79	99.8	95.2	78	2744	2251	493	70	11	3	0	1
V2.30	2313111600	2007376602	880783599	340.92	100	99	97	3792	3153	639	173	38	0	1	2
V2.31	773530600	681983651	334742156	129.78	99.8	97.1	88.2	2824	2294	530	56	8	0	0	1

ID	Total sequenced nucleotides	Total passed filter aligned nucleotides	Total passed filter aligned nucleotides in targeted regions	Mean coverage	% bases ≥ 1X°	% bases ≥ 20X°	% bases ≥ 50X°	# variants	# SNP	# Indel	# filtered* variants	# filtered missense	# filtered frame-shift	# filtered nonsense	# filtered splice
V2.32	1107243800	952873442	432500622	167.53	99.8	97.7	91.3	2773	2251	522	78	15	3	0	2
V2.33	1262394400	1126406753	433938620	168.63	99.9	97.6	90.8	2868	2357	511	76	20	3	0	1
V2.34	729711600	644240049	262838572	102.14	99.8	95	78.8	2770	2256	514	77	21	0	0	1
V2.35	1088653800	995478858	399149335	154.88	99.8	97.2	89.1	2660	2153	507	66	14	1	0	1
V2.36	1402971000	1257523339	482424435	187.65	99.9	98	92.4	2942	2410	532	114	22	4	0	1
V2.37	1432402000	1286637761	483808260	188.06	99.8	97.9	92.5	2753	2249	504	78	13	3	0	1
V2.38	1185778400	1072343725	422060643	163.82	99.8	97.5	90.4	2842	2339	503	84	8	0	2	1
V2.39	798479800	710469922	311345177	120.66	99.9	96.3	83.8	2763	2267	496	65	11	1	0	1
V2.40	1501506200	1374930437	542011393	210.45	99.9	98.2	93.8	2825	2289	536	74	7	0	0	1
V2.41	755312000	668403887	283829802	110.19	99.7	94.9	80.1	2803	2277	526	76	15	0	1	1
V2.42	955305400	817563816	354413159	137.59	99.8	97.2	88.1	2894	2359	535	74	10	1	0	1
V2.43	960359000	871557783	371634147	144.17	99.8	97.1	88.1	2725	2243	482	63	9	0	0	1
V2.44	1504276200	1303262139	515070393	200.06	99.9	98.2	93.2	2772	2268	504	77	13	3	0	2
V2.45	1061382400	924381441	402600639	156.27	99.8	97.5	90	2695	2181	514	80	16	0	0	1
V2.46	1040252400	925302762	380872541	148.02	99.8	97.5	89.9	2768	2261	507	74	7	4	1	2
V2.47	1291747800	1145317180	447396501	173.76	99.8	97.8	91.7	2817	2303	514	91	15	3	0	4
V2.48	1333586800	1140100684	481154521	185.91	99.8	97.7	92.1	2784	2269	515	72	15	3	1	1
V2.49	1343604200	1196406787	480490318	186	99.8	97.9	92.7	2754	2230	524	79	15	0	1	1
V2.50	1026321600	910914613	390514110	151.06	99.9	97.3	89.2	2756	2253	503	88	17	1	0	2
V2.51	1084525800	921482358	387228819	149.66	99.8	97.4	89.9	2743	2238	505	84	21	2	0	1
V2.52	893392400	766979835	339596355	131.15	99.8	96.2	85.3	2748	2289	459	69	9	0	0	3
V2.53	1397405400	1283250012	519754347	201.22	99.8	97.6	92.2	2905	2383	522	73	16	0	0	1
V2.54	1619633800	1496266236	613038295	236.97	99.9	98.4	94.9	2771	2260	511	79	18	0	0	1
V2.55	1525109600	1409329796	565809153	218.73	99.8	98	93.7	2793	2301	492	69	10	1	0	1
V2.56	1080181600	963545575	407219662	157.32	99.8	97.4	90.2	2873	2386	487	62	12	0	0	1
V2.57	1719283000	1516114039	573781112	222.16	99.8	98.2	94.2	2784	2267	517	87	16	3	0	1
V2.58	1291547200	1099723443	446787037	172.49	99.9	97.8	91.5	2691	2220	471	81	14	1	1	1
V2.59	1056774800	958828432	402546525	155.55	99.8	97.5	90.6	2665	2172	493	68	16	2	0	1
V2.60	1386210000	1231985409	485313286	187.5	99.8	97.9	92.7	3061	2524	537	111	21	0	0	0
V2.61	1434162600	1325855516	508545342	196.55	99.8	97.6	92.5	3292	2734	558	230	40	3	0	1
V2.62	1311658400	1164036368	448507877	173.73	99.8	97.8	92.2	2842	2338	504	78	12	0	0	2
V2.63	784783800	612917519	276014895	106.25	99.8	93.8	77	2662	2214	448	85	23	0	1	0
V2.64	1477661800	1260260874	548548734	211.77	99.9	97.7	92.8	2822	2330	492	108	26	2	0	1
V2.65	1507305000	1331063062	558019554	216.29	99.8	97.8	93.2	2599	2161	438	56	10	2	1	1

ID	Total sequenced nucleotides	Total passed filter aligned nucleotides	Total passed filter aligned nucleotides in targeted regions	Mean coverage	% bases ≥ 1X°	% bases ≥ 20X°	% bases ≥ 50X°	# variants	# SNP	# Indel	# filtered* variants	# filtered missense	# filtered frame-shift	# filtered nonsense	# filtered splice
V2.66	1561032200	1301266158	553784231	213.92	99.8	97.5	92.3	2749	2285	464	42	9	2	0	1
V2.67	1845095000	1618692423	627804999	243.04	99.9	98	93.7	2801	2340	461	61	17	2	0	1
V2.68	2211562400	1953015892	735168165	285.02	99.9	98.5	95.8	2729	2229	500	55	4	3	0	1
V2.69	1365493200	1159701682	507692366	196.28	99.9	97.8	92.1	2641	2165	476	80	16	2	0	1
V2.70	2452899800	2205738567	871452465	337.39	99.9	98.3	95.7	2768	2294	474	62	9	2	1	1
V2.71	2757139800	2381482082	995840560	384.73	99.9	98.6	96.2	2808	2336	472	58	14	0	0	3
V2.72	722424400	644206852	295660531	114.65	99.8	95.3	81.4	2838	2370	468	78	18	2	0	1
V2.73	1041247000	935138556	401389849	155.98	99.8	97.2	89.4	2675	2219	456	49	12	1	0	1
V2.74	817505800	747780422	330902440	128.37	99.8	96.3	85	2667	2197	470	57	8	2	0	1
V2.75	1044084000	932546469	410852215	159.53	99.8	97	89.1	2677	2232	445	79	20	2	0	1
V2.76	1097412400	1007856644	411589046	159.73	99.8	97.4	89.6	2811	2330	481	63	14	2	0	1
V2.77	887949600	819098704	369027076	143.13	99.8	96.8	87.7	2725	2265	460	50	9	2	0	1
V2.78	1149476000	1036810860	452316293	175.16	99.8	97.4	90.9	2660	2206	454	56	14	0	0	1
V2.79	553319000	493066518	241368038	93.42	99.8	94.1	75.6	2647	2200	447	59	15	1	1	1
V2.80	1209526800	999910832	436988783	169.13	99.8	97.4	90.6	2568	2129	439	39	2	2	0	1
V2.81	848366200	739270269	340597891	132	99.8	96.9	87.2	2893	2410	483	72	15	1	2	2
V2.82	952025200	842805360	391452746	151.82	99.7	97	89.2	3263	2741	522	125	18	6	0	2
V2.83	1281135400	1159643658	485024670	188.02	99.8	97.4	91.2	2576	2143	433	49	16	3	0	1
V2.84	1642417600	1457576294	591501793	229.36	99.9	98.4	94.8	2755	2256	499	57	17	2	0	2
V2.85	1391188600	1239369834	534937296	207.52	99.8	98.1	93.9	2802	2334	468	80	11	0	0	1
V2.86	1498139000	1280633352	550946548	213.44	99.8	98	93.9	2780	2312	468	83	23	2	1	1
V2.87	2070727600	1871514233	738939121	286.94	99.9	98.6	95.9	2733	2255	478	63	13	3	1	1
V2.88	2083217200	1871302407	781453864	302.93	99.9	98.6	96.2	2825	2358	467	87	16	2	0	1
V2.89	1101851800	939454646	425625286	165.26	99.8	97.9	91.8	2835	2328	507	79	16	3	0	1
V2.90	1384881400	1233825394	538855687	208.95	99.9	98.2	93.7	2676	2220	456	68	18	2	1	2
V2.91	1933795600	1759986293	759672348	294.62	99.8	98.5	95.9	2598	2160	438	62	20	4	0	1
V2.92	1122398800	999746160	452635262	175.43	99.9	97.9	91.7	2783	2313	470	55	12	0	0	1
V2.93	901247800	798025190	378901767	146.81	99.9	97.1	88.4	2728	2259	469	64	15	3	1	1
V2.94	1181614200	1040532507	455503452	176.51	99.9	97.7	91.3	2703	2228	475	47	12	2	0	1
V2.95	1935740600	1749394743	743850497	288.45	99.9	98.6	96	2811	2317	494	52	11	0	0	1

*filtered: variants that passed the following criteria: allele frequency < 1% in dbSNP137, 1000 genomes, Exome Variant Server, and in-house database; °includes only reads with MQ > 30

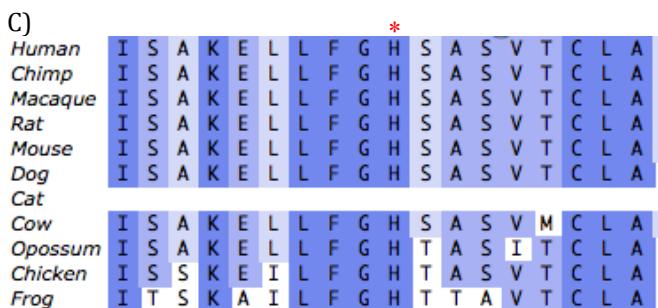
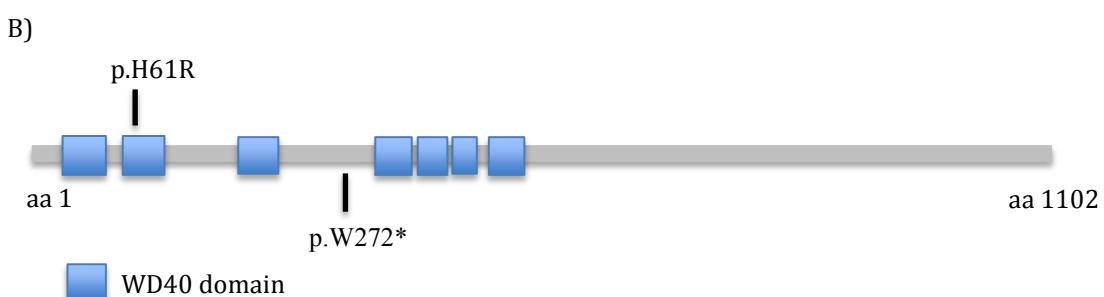
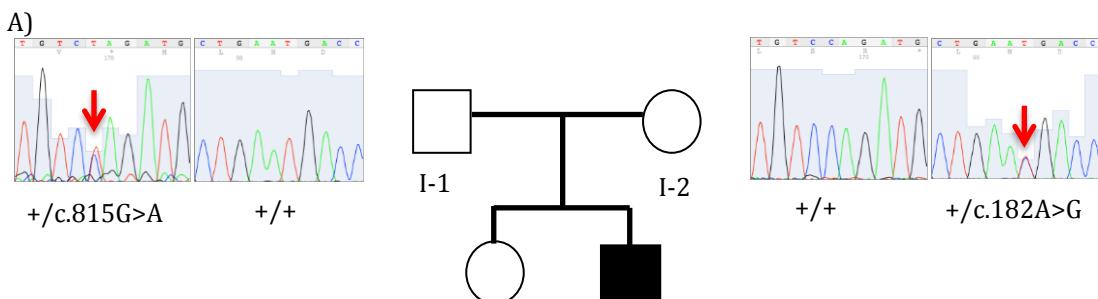
Supplementary Table 8. Comparison of sequence coverage of the targeted regions achieved with v2.0 of our NGS gene panel and that of the same regions achieved by whole exome sequencing in 8 random samples with a similar capture/sequencing protocol and bioinformatics pipeline

		Targeted NGS gene panel (v2.0)	V5
Mean coverage (X)	Overall	178.90	100.84
	Diagnosis sub-panel	211.17	102.62
	Discovery sub-panel	175.35	99.97
% reads >= 20X	Overall	97.23	89.09
	Diagnosis sub-panel	99.41	89.14
	Discovery sub-panel	96.65	89.06
% reads >= 50X	Overall	89.39	70.25
	Diagnosis sub-panel	97.01	70.96
	Discovery sub-panel	87.63	69.91

Coverage calculations include only high-quality mapped reads (MQ > 30); V5: Agilent SureSelect V5 (Agilent, Santa Clara, USA)

REFERENCES

1. Huckert M, Stoetzel C, Morkmued S, et al. Mutations in the Latent TGF-beta Binding Protein 3 (LTBP3) gene cause brachyolmia with amelogenesis imperfecta. *Human molecular genetics* 2015 doi: 10.1093/hmg/ddv053 [published Online First: 2015/02/12].
2. Jaureguierry G, De la Dure-Molla M, Parry D, et al. Nephrocalcinosis (enamel renal syndrome) caused by autosomal recessive FAM20A mutations. *Nephron. Physiology* 2012;122(1-2):1-6 doi: 10.1159/000349989 [published Online First: 2013/02/26].
3. Parry DA, Mighell AJ, El-Sayed W, et al. Mutations in CNNM4 cause Jalili syndrome, consisting of autosomal-recessive cone-rod dystrophy and amelogenesis imperfecta. *American journal of human genetics* 2009;84(2):266-73 doi: 10.1016/j.ajhg.2009.01.009 [published Online First: 2009/02/10].



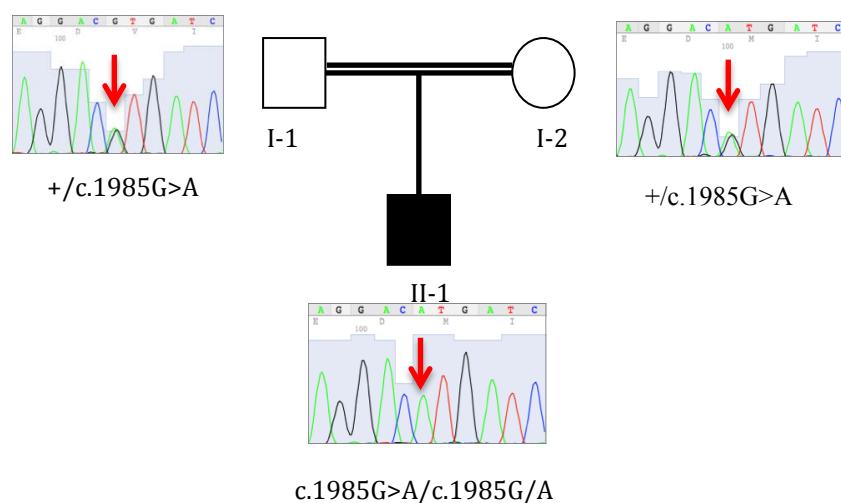
D)

Mutation	PolypHEN[1]	SIFT[2]	MutationTaster[3]	1000G[4]	EVS[5]	ExAC[6]	Grantham Score
p.H61R	Probably damaging	Deleterious	Disease-causing	Absent	Absent	Absent	29
p.W272X	-	-	-	Absent	Absent	Absent	-

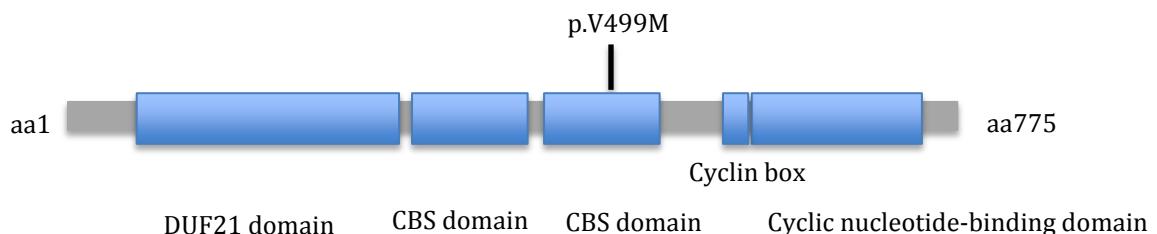
Supplementary Figure 1. Two novel mutations in *WDR72* in patient V1.14

A) Family pedigree showing segregation of the two mutations. The mutated nucleotides are indicated by a red arrow. B) Location of the two mutations in the protein. Image adapted from [7]. C) Multi-species sequence conservation around amino acid 61 (Alamut v2.6.2). The p.H61R (*) mutation affects a conserved nucleotide in a functional WD40 domain. D) Table showing the predicted effects of the missense mutation and the frequency of both mutations in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



B)

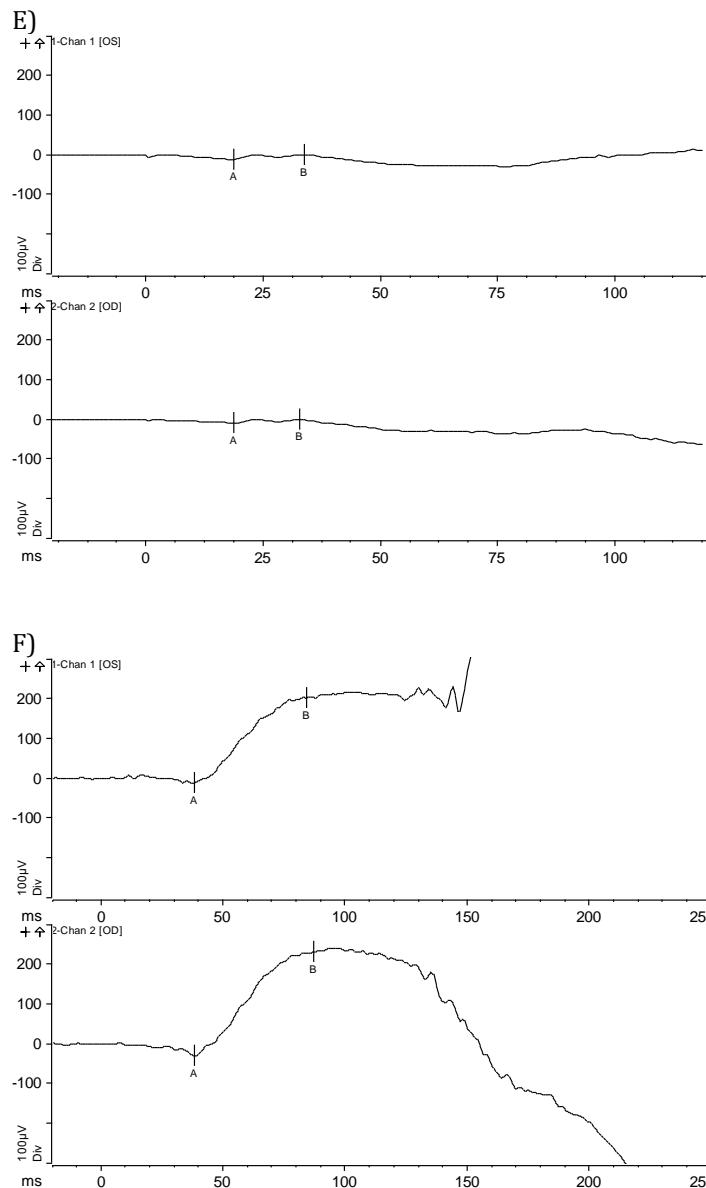


C)

	V	T	L	E	D	V	I	E	E	I	I	K	S
<i>Human</i>	V	T	L	E	D	V	I	E	E	I	I	K	S
<i>Chimp</i>	V	T	L	E	D	V	I	E	E	I	I	K	S
<i>Macaque</i>	V	T	L	E	D	V	I	E	E	I	I	K	S
<i>Rat</i>	V	T	L	E	D	V	I	E	E	I	I	K	S
<i>Mouse</i>	V	T	L	E	D	V	I	E	E	I	I	K	S
<i>Dog</i>	V	T	L	E	D	V	I	E	E	I	I	K	S
<i>Frog</i>	V	T	L	E	D	V	I	E	E	I	I	K	S
<i>Tetraodon</i>	V	T	L	E	D	V	I	E	E	I	I	K	S
<i>Fruitfly</i>	V	T	L	E	D	V	I	E	E	L	I	Q	A
<i>C. elegans</i>	V	T	L	E	D	I	V	E	E	I	L	Q	A
<i>Baker's yeast</i>	L	T	L	E	D	V	I	E	E	L	I	G	E

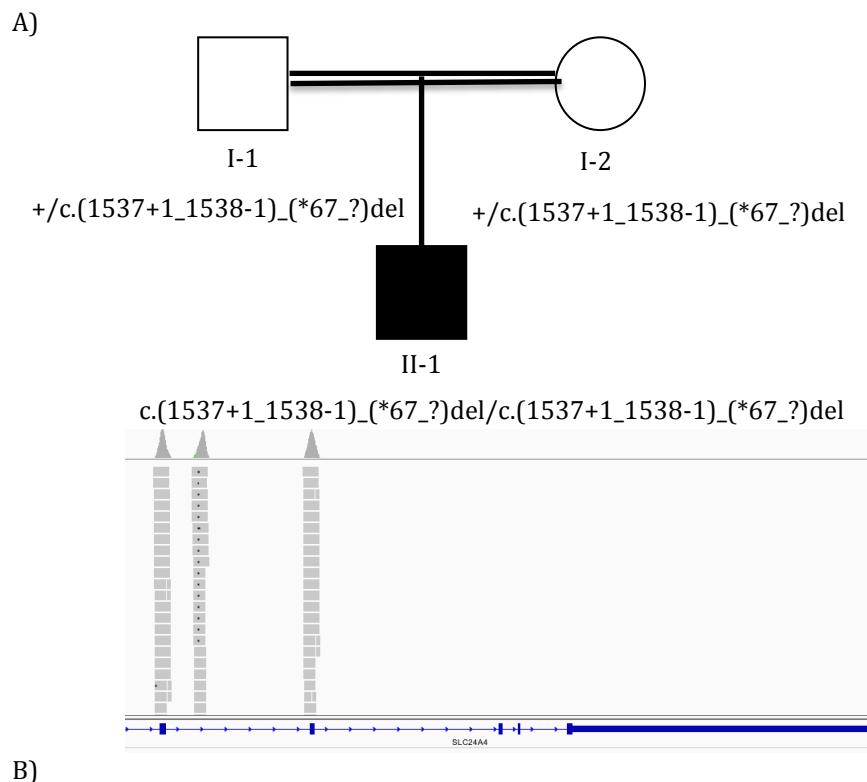
D)

Mutation	Polphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham Score
p.V499M	Deleterious	Deleterious	Disease-causing	Absent	Absent	Absent	21

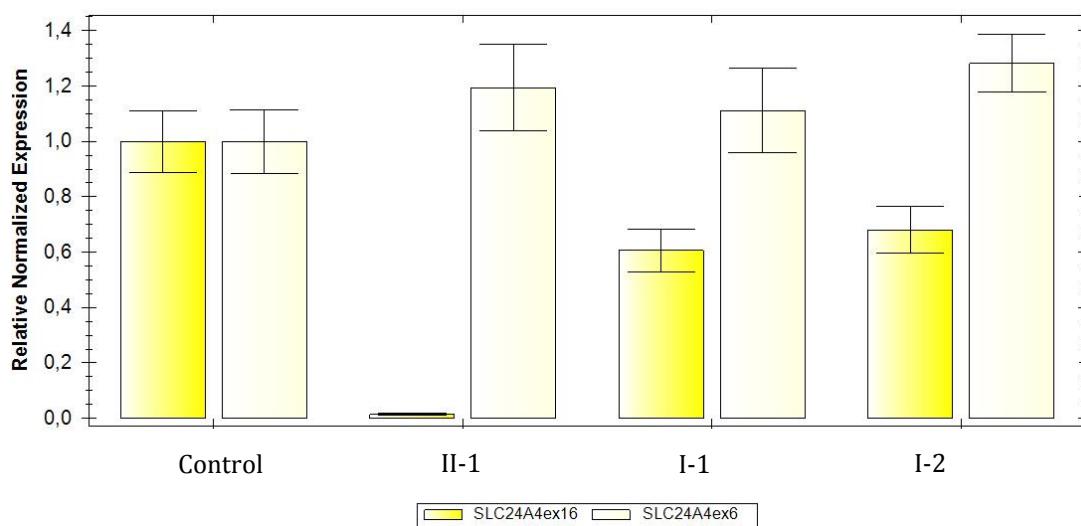


Supplementary Figure 2. Novel homozygous mutation in *CNNM4* in patient V2.05

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated by a red arrow. B) Location of the mutation in the protein. Image adapted from [8]. C) Multi-species sequence conservation of amino acid 499 (Alamut v2.6.2). The p.V499M (*) mutation affects a conserved nucleotide in a core functional CBS domain. CBS: cystathionine beta-synthase, core domain; DUF21: domain of unknown function DUF21. D) Table showing the predicted effects of the missense mutation and the frequency of the mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server. E-F) Results from a full-field electroretinography (ERG) exam[9] in patient V2.05; E) Results from the light-adapted 3 ERG protocol showing a marked loss of cone response; F) Results from the dark-adapted 0.01 protocol showing a less severe reduction in rod response.



B)

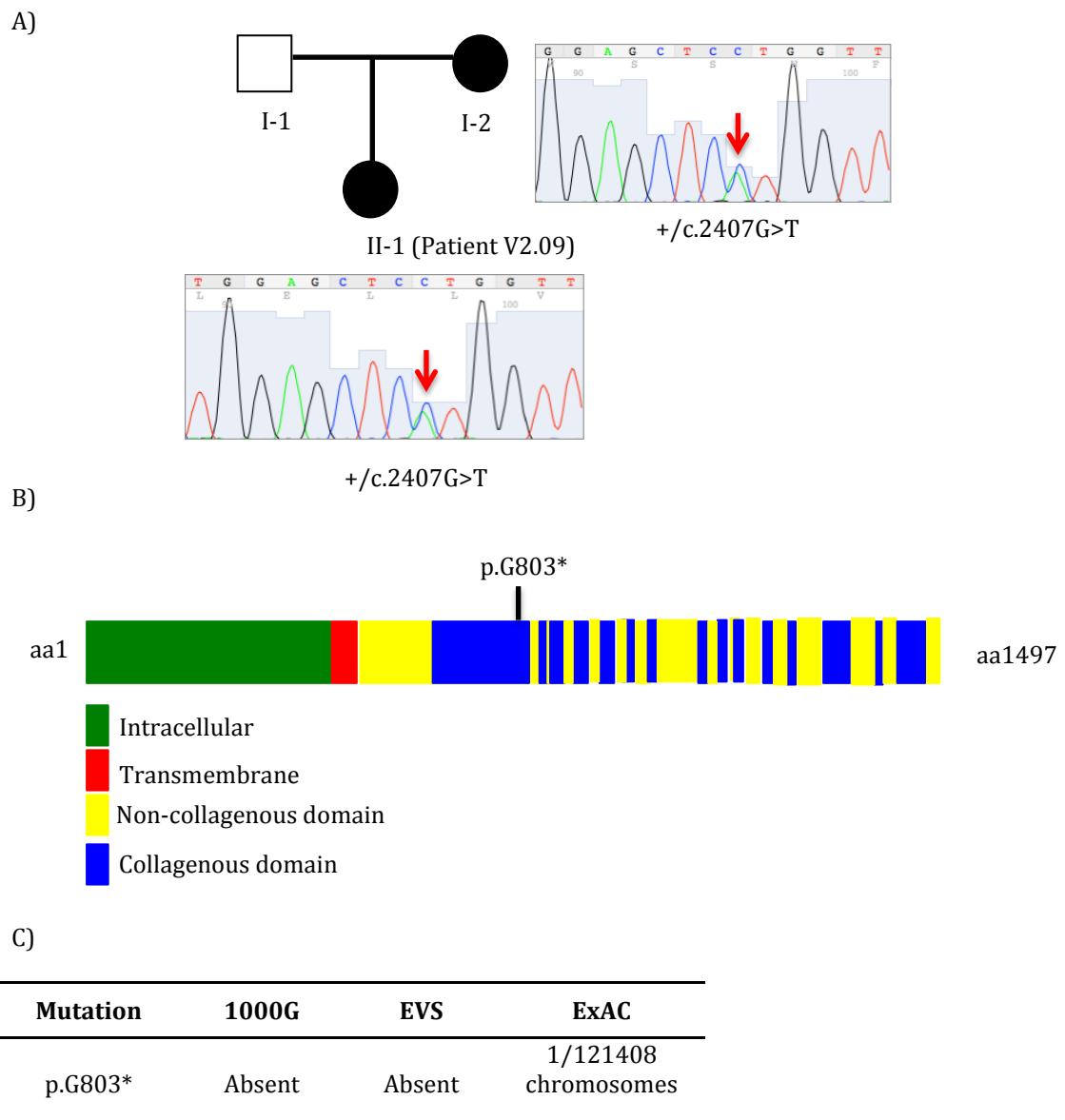


C)

Mutation	Database of Genomic Variants[10]
c.(1537+1_1538-1)_(*67_?)del	Absent

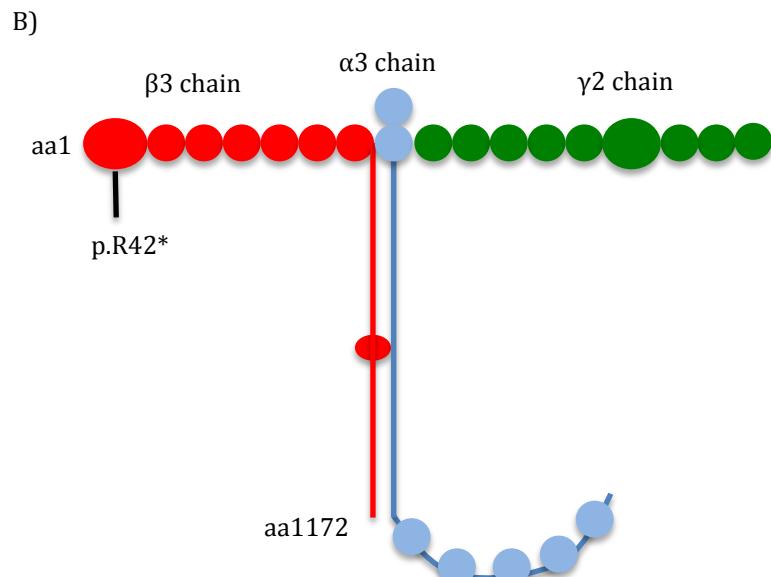
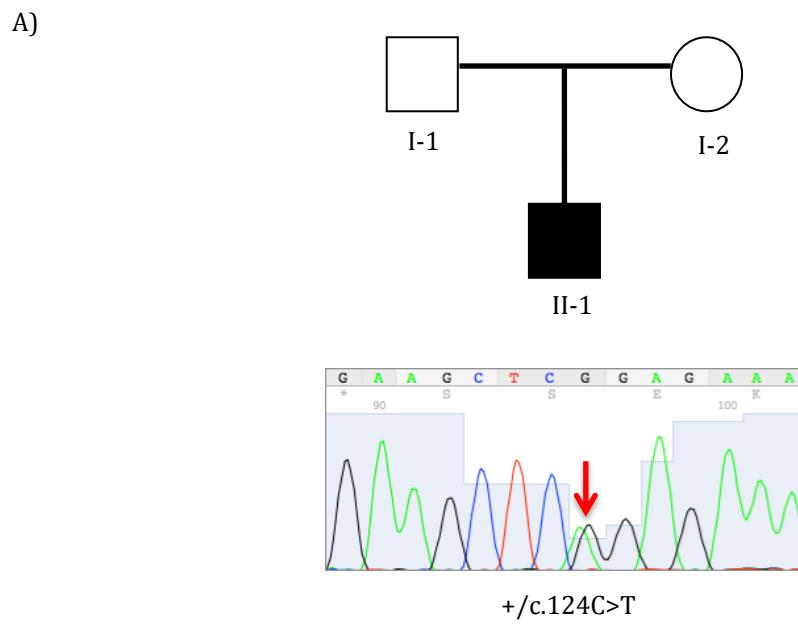
Supplementary Figure 3. Homozygous deletion in *SLC24A4* in patient V2.08

A) Family pedigree showing segregation of the deletions. Shown is a screen shot of the bam file as visualised by Integrated Genomics Viewer. There is a lack of reads overlapping the last 3 exons of the gene. B) Quantitative PCR (qPCR) validation of the homozygous deletion in the patient and the heterozygous deletion in each parent.



Supplementary Figure 4. Heterozygous nonsense mutation in *COL17A1* in patient V2.09

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutated residue in the protein. Image adapted from [11]. C) Table showing the frequency of the mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

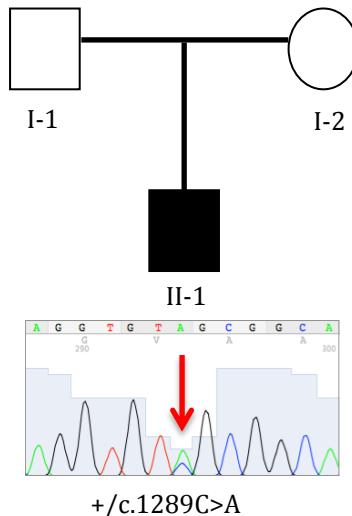


C)

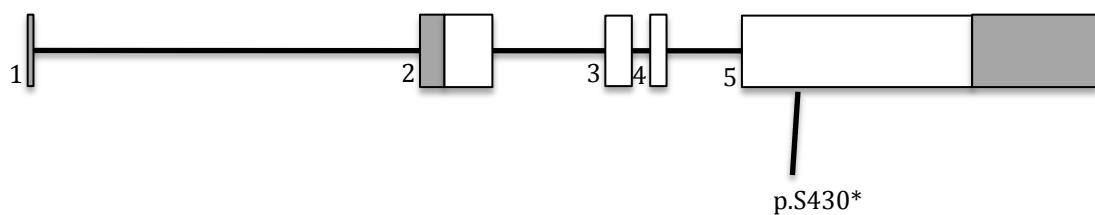
Mutation	1000G	EVS	ExAC
p.R42*	Absent	Absent	2/121238 chromosomes

Supplementary Figure 5. Heterozygous nonsense mutation in *LAMB3* in patient V2.18
A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutated residue in the protein. Image adapted from [12]. C) Table showing the frequency of the mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



B)



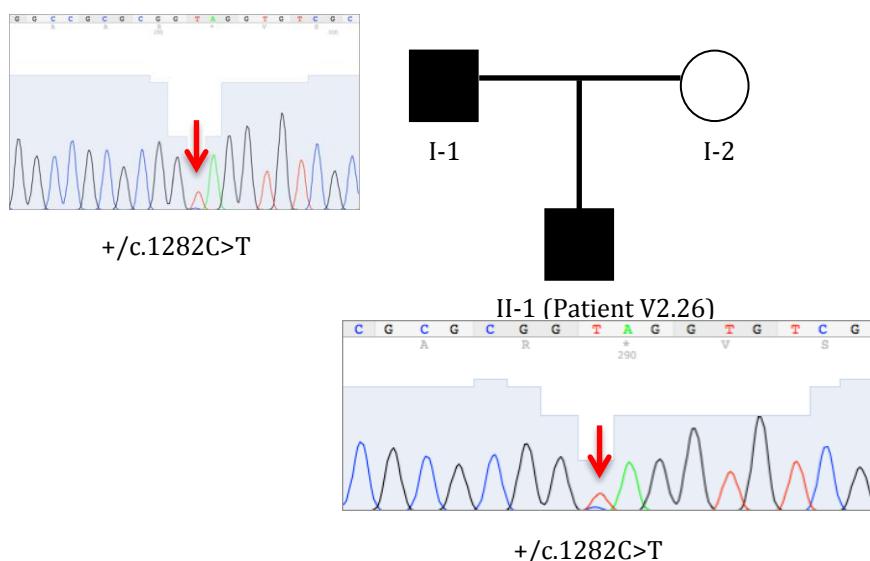
C)

Mutation	1000G	EVS	ExAC
p.S430*	Absent	Absent	Absent

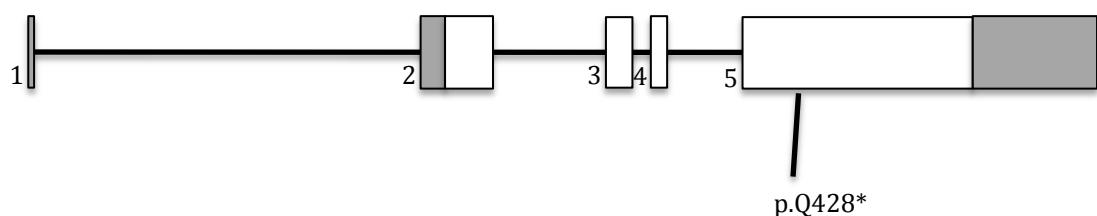
Supplementary Figure 6. Heterozygous nonsense mutation in *FAM83H* in patient V2.20

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutated residue in the gene. Coding exons are represented as white rectangles whereas grey rectangles represent non-coding exons. The exons are numbered from the 5' end of the gene. Image adapted from [13]. C) Table showing the frequency of the mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



B)



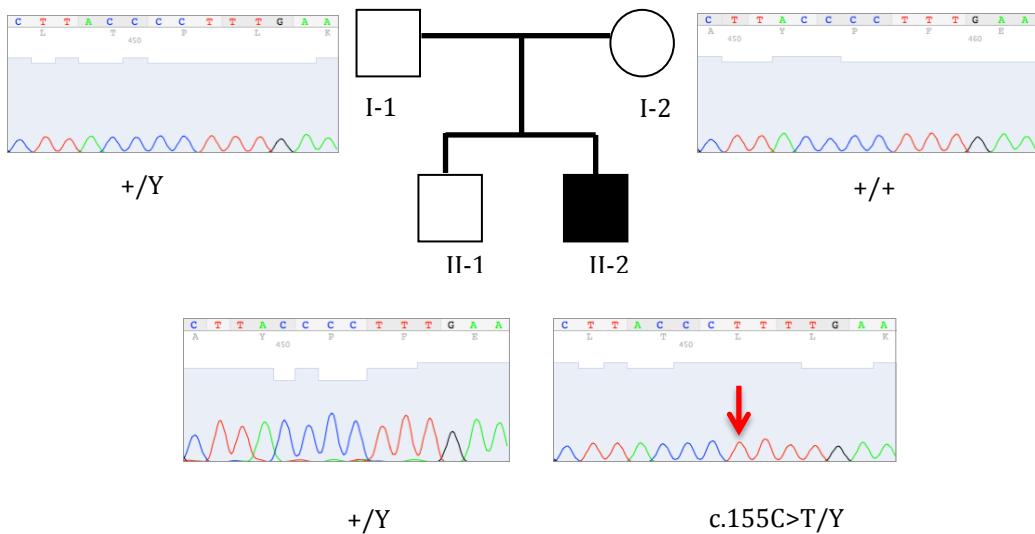
C)

Mutation	1000G	EVS	ExAC
p.Q428*	Absent	Absent	Absent

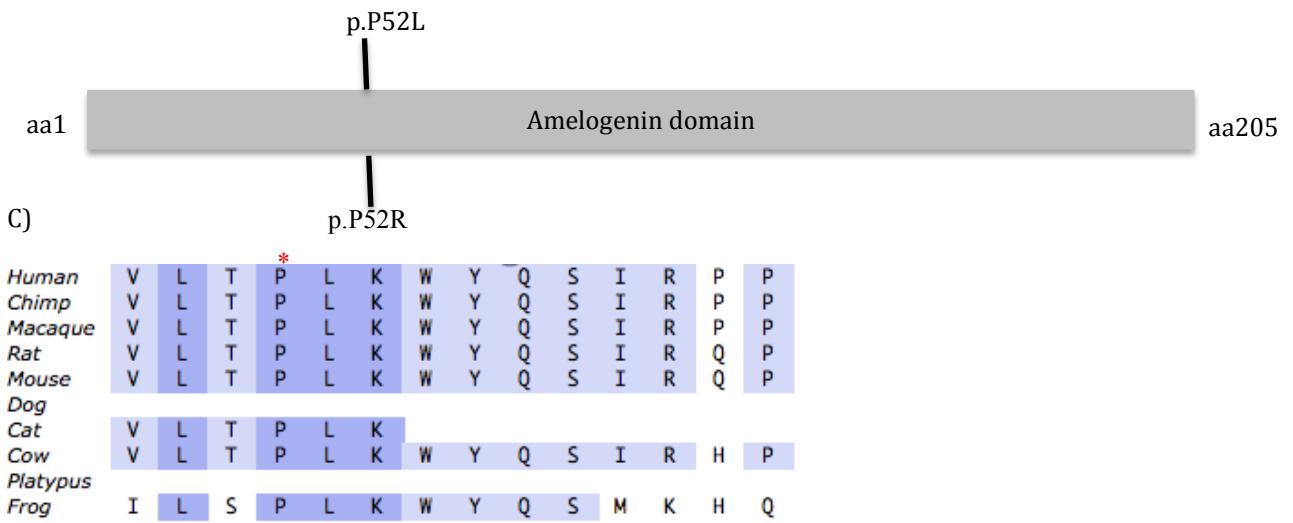
Supplementary Figure 7. Heterozygous nonsense mutation in *FAM83H* in patient V2.26

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutated residue in the gene. Coding exons are represented as white rectangles whereas grey rectangles represent non-coding exons. The exons are numbered from the 5' end of the gene. Image adapted from [13]. C) Table showing the frequency of the mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



B)

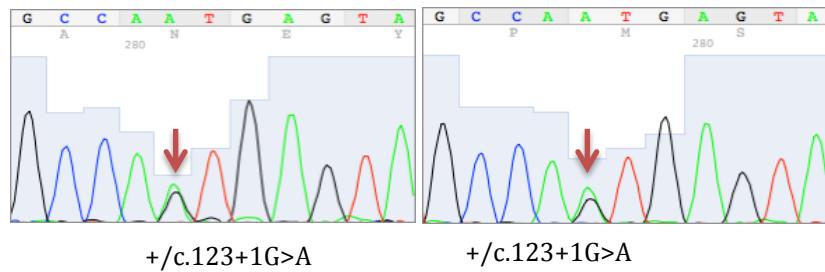
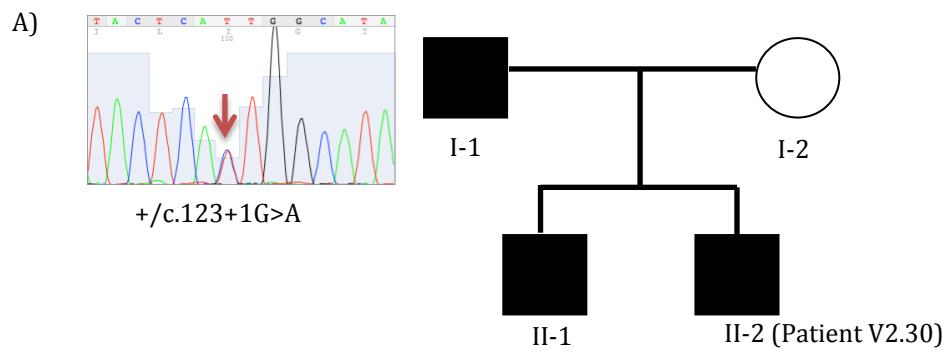


C)

Mutation	Polphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham score
p.P52L	Deleterious	Deleterious	Disease-causing	Absent	Absent	Absent	98

Supplementary Figure 8. Heterozygous nonsense mutation in *AMELX* in patient V2.28

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutated residue in the protein. Also shown is the p.P52R mutation previously reported by [14] that affects the same residue. C) Multi-species sequence conservation around amino acid 52 (Alamut v2.6.2). The p.P52R (*) mutation affects a conserved nucleotide in a functional domain. D) Table showing the frequency of the mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

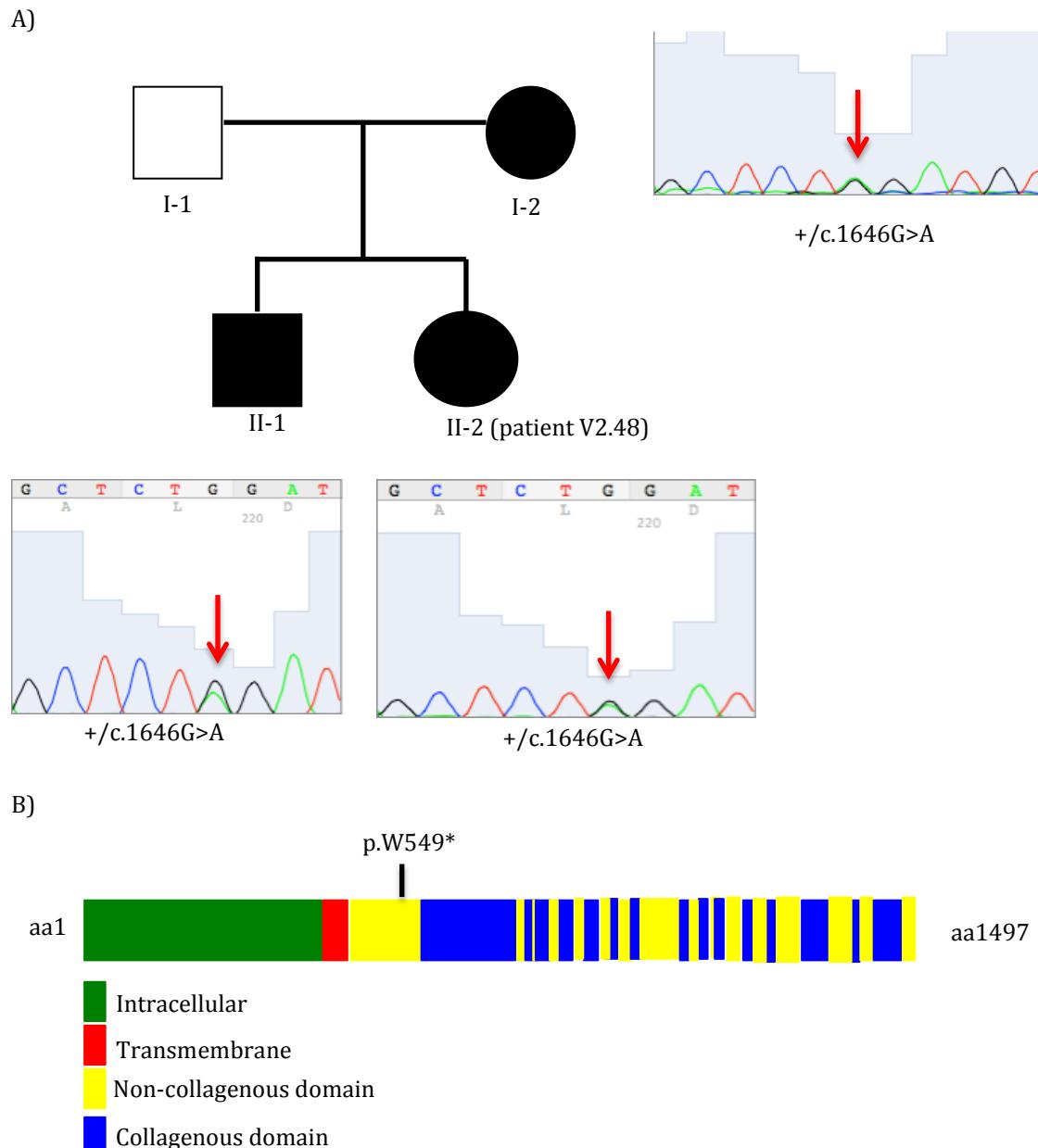


B)

Mutation	SSF	NNSplice	MaxEntScan	1000G	EVS	ExAC
c.123+1G>A	-100%	-100%	-100%	Absent	Absent	Absent

Supplementary Figure 9. Heterozygous splice donor mutation in ENAM in patient V2.30

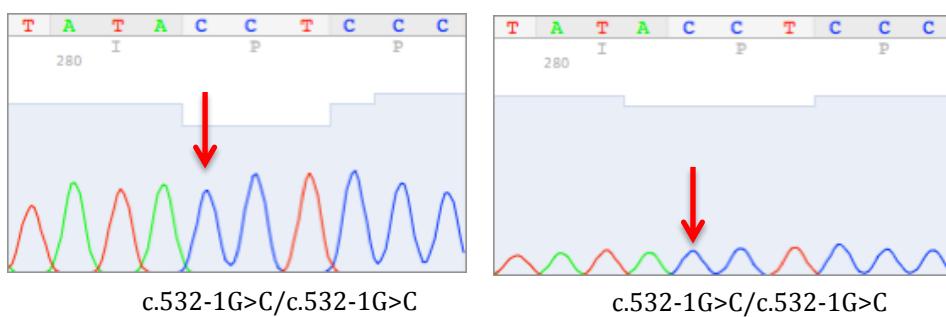
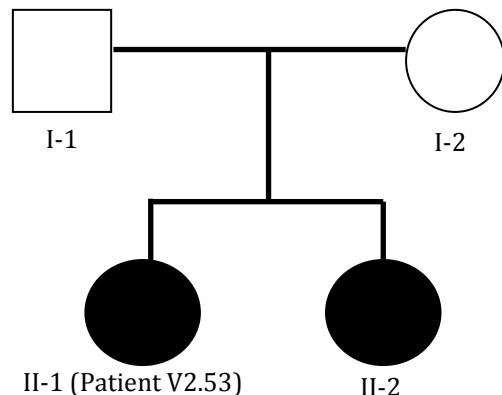
A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Table showing bioinformatically predicted effect on splicing and the frequency of the mutation in public databases. SSF: Splice Site Finder; EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.



Supplementary Figure 10. Heterozygous nonsense mutation in *COL17A1* in patient V2.48

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutated residue in the protein. Image adapted from [11]. C) Table showing the frequency of both mutations in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)

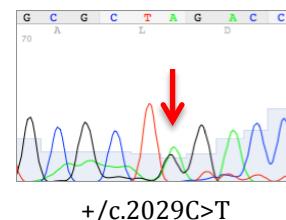
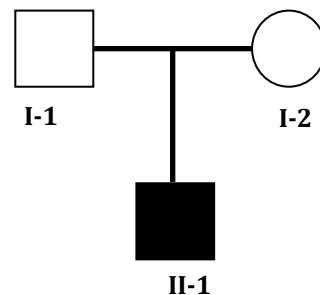


B)

Mutation	SSF[15]	NNSplice[16]	MaxEntScan[17]	1000G	EVS	ExAC
c.532-1G>C	-100%	-100%	-100%	Absent	1/12950 chromosomes	1/118926 chromosomes
c.539	+80.33	+6.11	-			

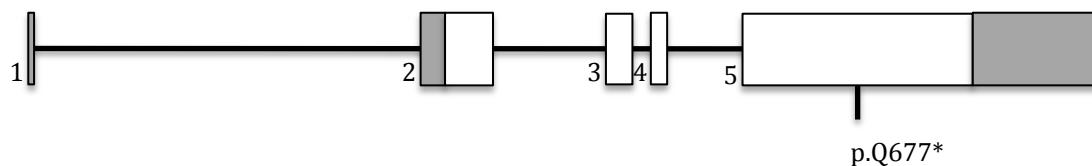
Supplementary Figure 11. Homozygous splice acceptor mutation in *AMBN* in patient V2.53
 A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Table showing bioinformatically predicted effect on splicing and the frequency of the mutation in public databases. SSF: Splice Site Finder; EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



+/c.2029C>T

B)



p.Q677*

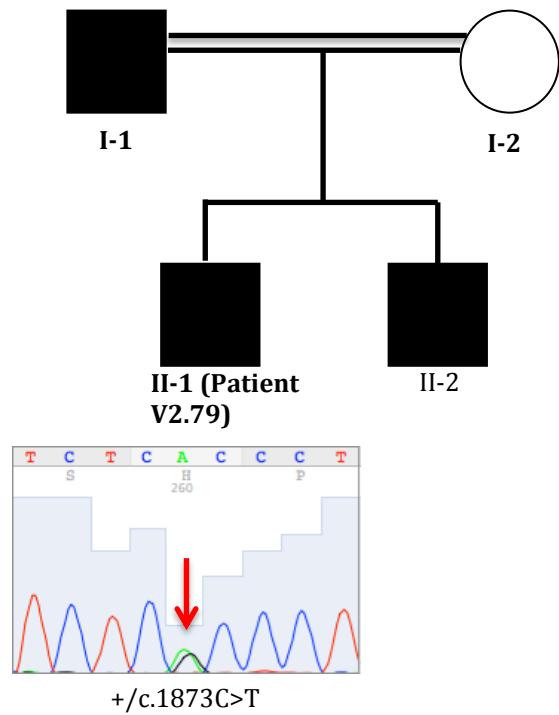
C)

Mutation	1000G	EVS	ExAC
p.Q677*	Absent	Absent	Absent

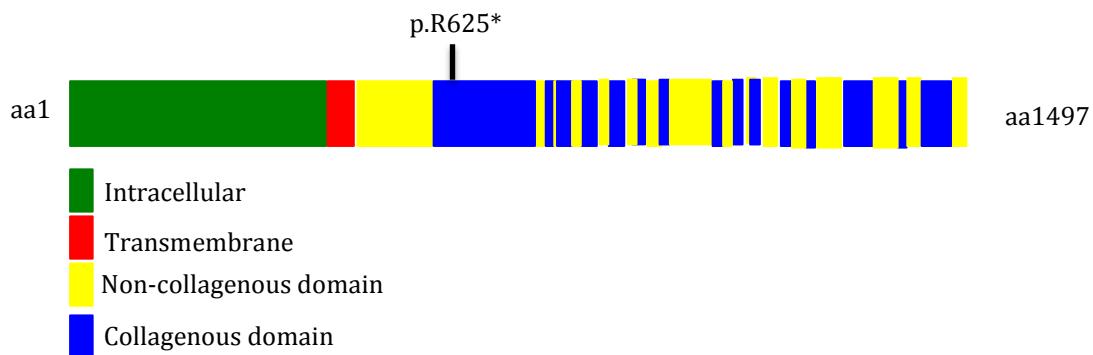
Supplementary Figure 12. Heterozygous nonsense mutation in *FAM83H* in patient V2.63

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutated residue in the gene. Coding exons are represented as white rectangles whereas grey rectangles represent non-coding exons. The exons are numbered from the 5' end of the gene. Image adapted from [13]. C) Table showing the frequency of the mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



B)

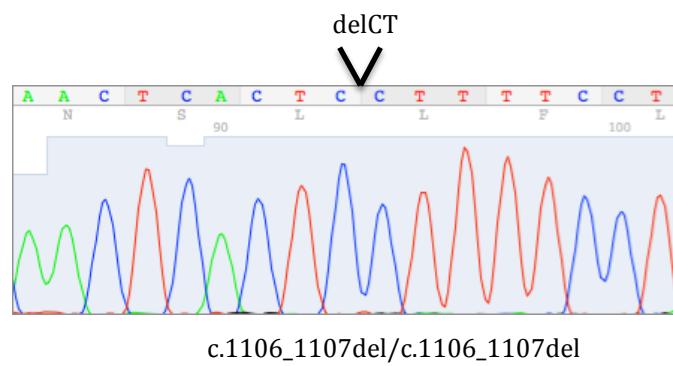
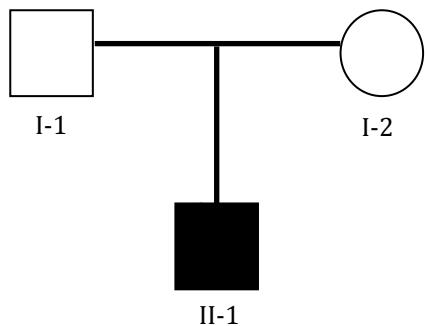


C)

Mutation	1000G	EVS	ExAC
p.R625*	Absent	Absent	1 / 121206 chromosomes

Supplementary Figure 13. Heterozygous nonsense mutation in *COL17A1* in patient V2.79
 A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. DNA was not available from additional members of the family. B) Location of the mutated residue in the protein. Image adapted from [11]. C) Table showing the frequency of both mutations in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



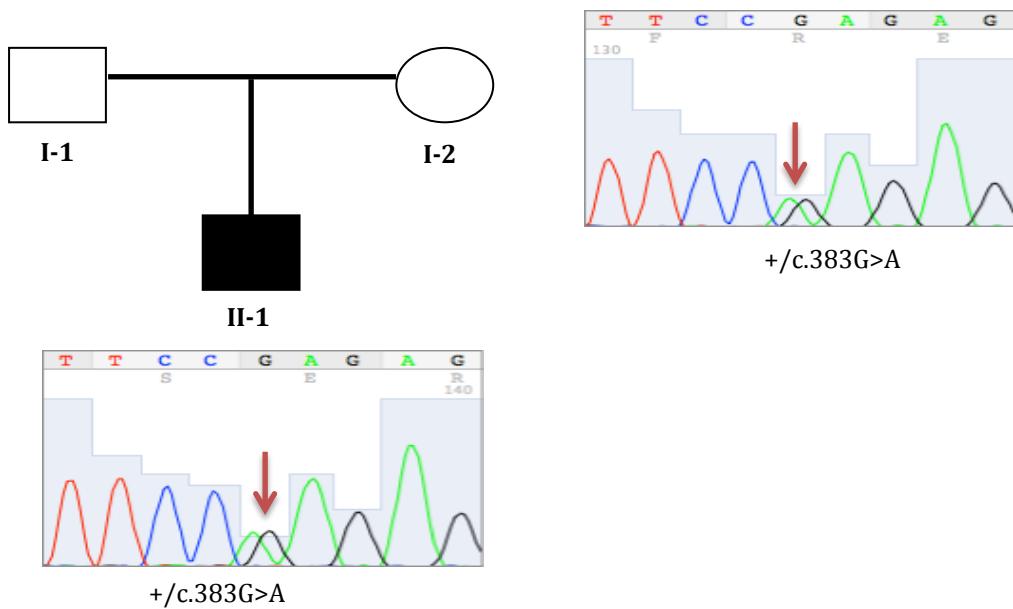
B)

Mutation	1000G	EVS	ExAC
p.E2316Gfs*10	Absent	Absent	Absent

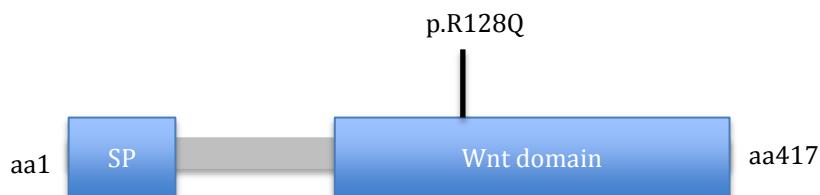
Supplementary Figure 14. Homozygous frameshift-causing dinucleotide deletion in *FAM20A* in patient V2.06

A) Family pedigree showing Sanger chromatogram of the mutation. The deletion is indicated. B) Table showing the frequency of the mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



B)



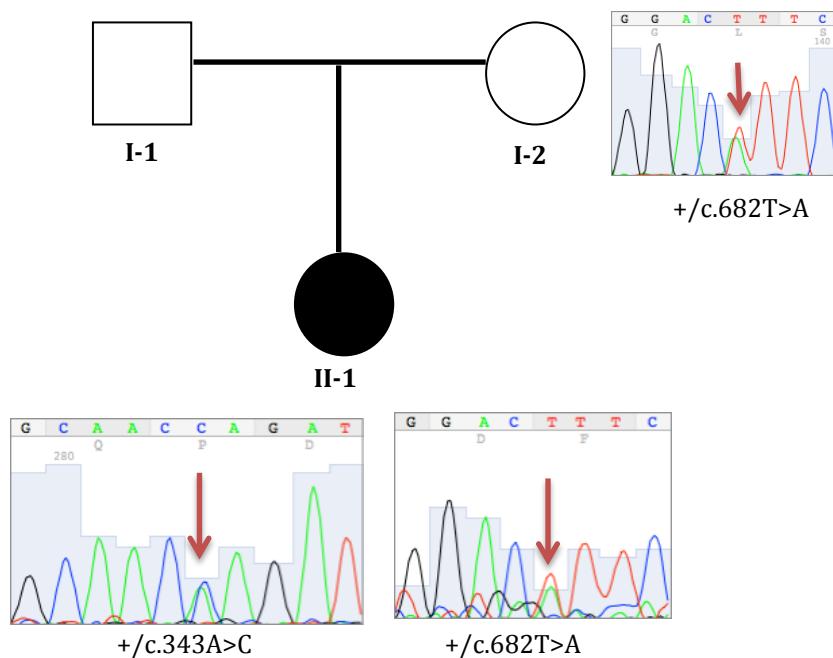
C)

Mutation	Polyphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham Score
p.R128Q	Probably damaging	Deleterious	Disease-causing	Absent	1/13006 chromosomes	1/121340 chromosomes	43

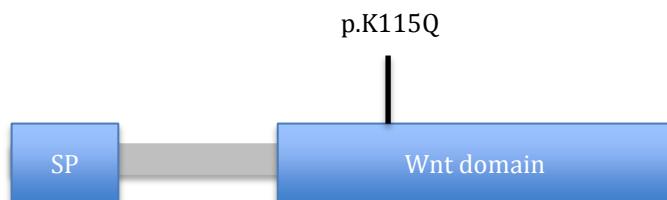
Supplementary Figure 15. Heterozygous missense mutation in *WNT10A* in patient V1.11

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the missense mutation in the protein. SP: Signal peptide. Image based on [18]. C) Table showing the bioinformatically predicted effect and the frequency of the missense mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

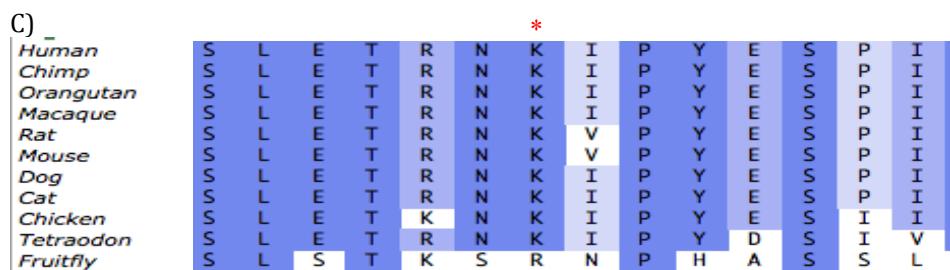
A)



B)



C)

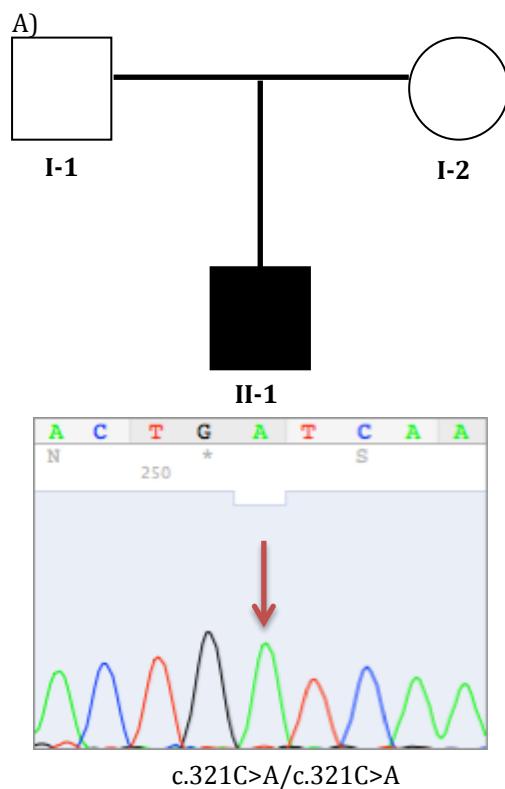


D)

Mutation	Polyphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham Score
p.F228I	Deleterious	Deleterious	Disease-causing	30/4978 chromosomes	241/13000 chromosomes	1499/117764 chromosomes	21
p.K115Q	Benign	Deleterious	Disease-causing	Absent	Absent	Absent	53

Supplementary Figure 16. Two heterozygous missense mutations in *WNT10A* in patient V1.15

A) Family pedigree showing segregation of the mutations. The mutated nucleotide is indicated with an arrow. B) Location of the novel mutation in the protein. C) Sequence alignment around amino acid 115. The K115Q (*) mutation affects a conserved amino acid in a functional domain. D) Table showing the bioinformatically predicted effect and the frequency of the missense mutations in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.



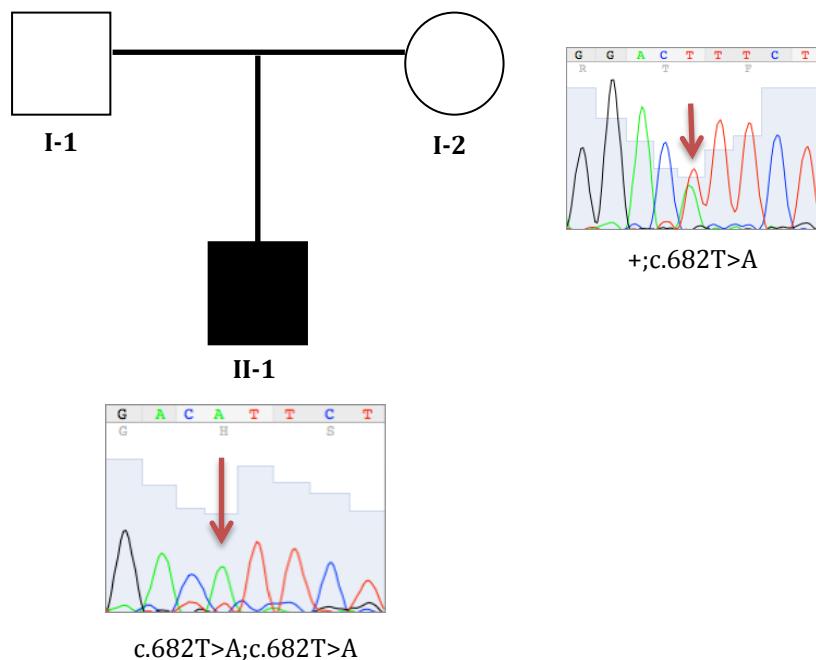
B)

Mutation	1000G	EVS	ExAC
p.C107*	Absent	13/130006 chromosomes	71/121352 chromosomes

Supplementary Figure 17. Homozygous missense mutation in *WNT10A* in patient V2.55

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Table showing the frequency of the mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



B)



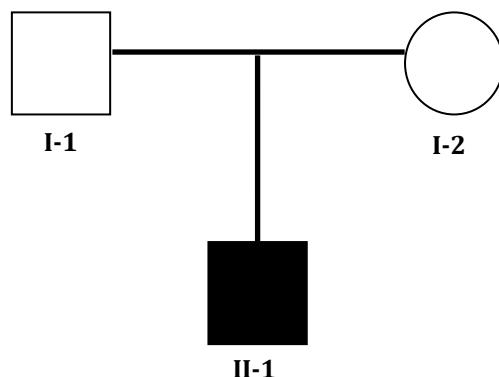
C)

Mutation	Polyphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham score
p.F228I	Deleterious	Deleterious	Disease-causing	30/4978 chromosomes	241/13000 chromosomes	1499/117764 chromosomes	21

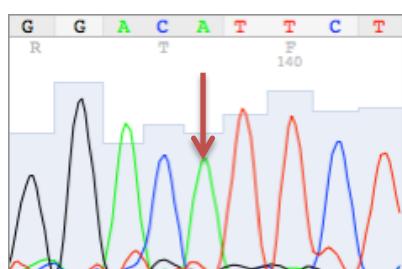
Supplementary Figure 18. Homozygous missense mutation in *WNT10A* in patient V2.65

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutation in the protein. SP: Signal peptide. Image based on [18]. C) Table showing the bioinformatically predicted effect and the frequency of the missense mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



B)



c.682T>A;c.682T>A

p.F228I



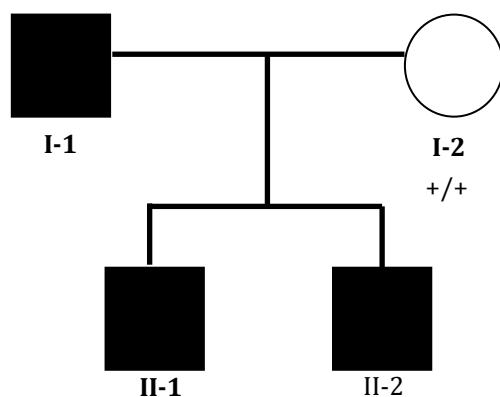
C)

Mutation	Polyphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham score
p.F228I	Deleterious	Deleterious	Disease-causing	30/4978 chromosomes	241/13000 chromosomes	1499/117764 chromosomes	21

Supplementary Figure S19. Homozygous missense mutation in *WNT10A* in patient V2.66

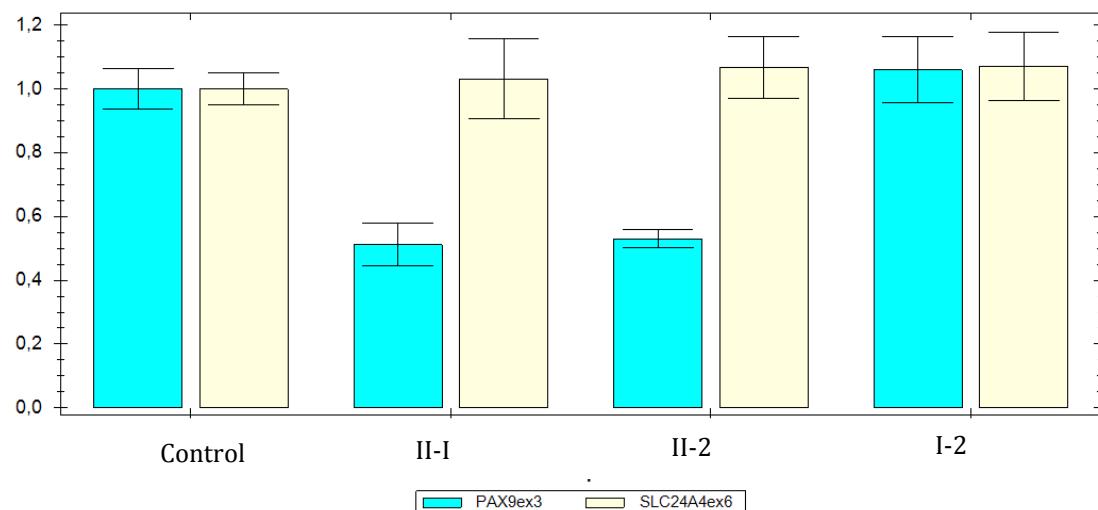
A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutation in the protein. SP: Signal peptide. Image based on [18]. C) Table showing the bioinformatically predicted effect and the frequency of the missense mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



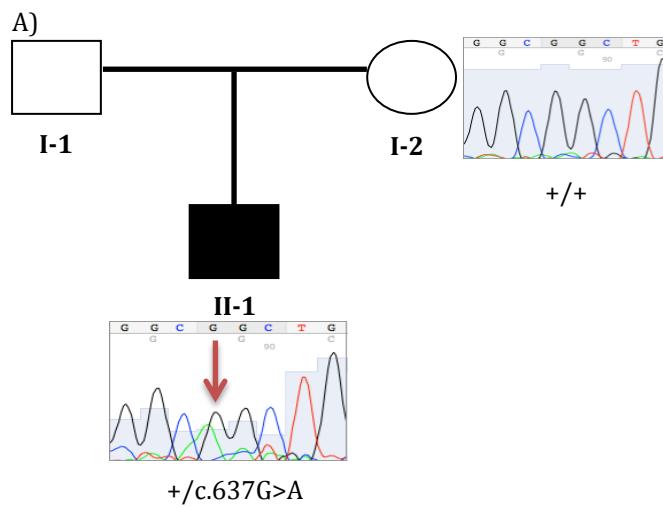
+/c.[(?-115)_(*62_?)del] +/c.[(?-115)_(*62_?)del]

B)



Supplementary Figure 20. Heterozygous deletion in *PAX9* in patient V2.67

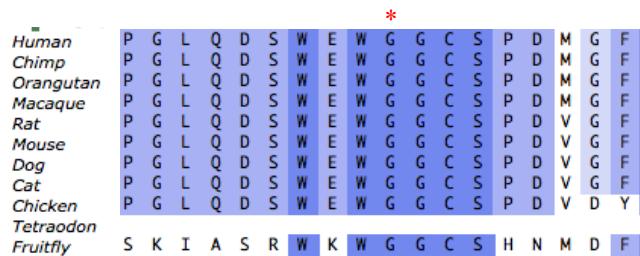
A) Family pedigree. B) Quantitative PCR (qPCR) validation of the deletion in the family.



B)



C)

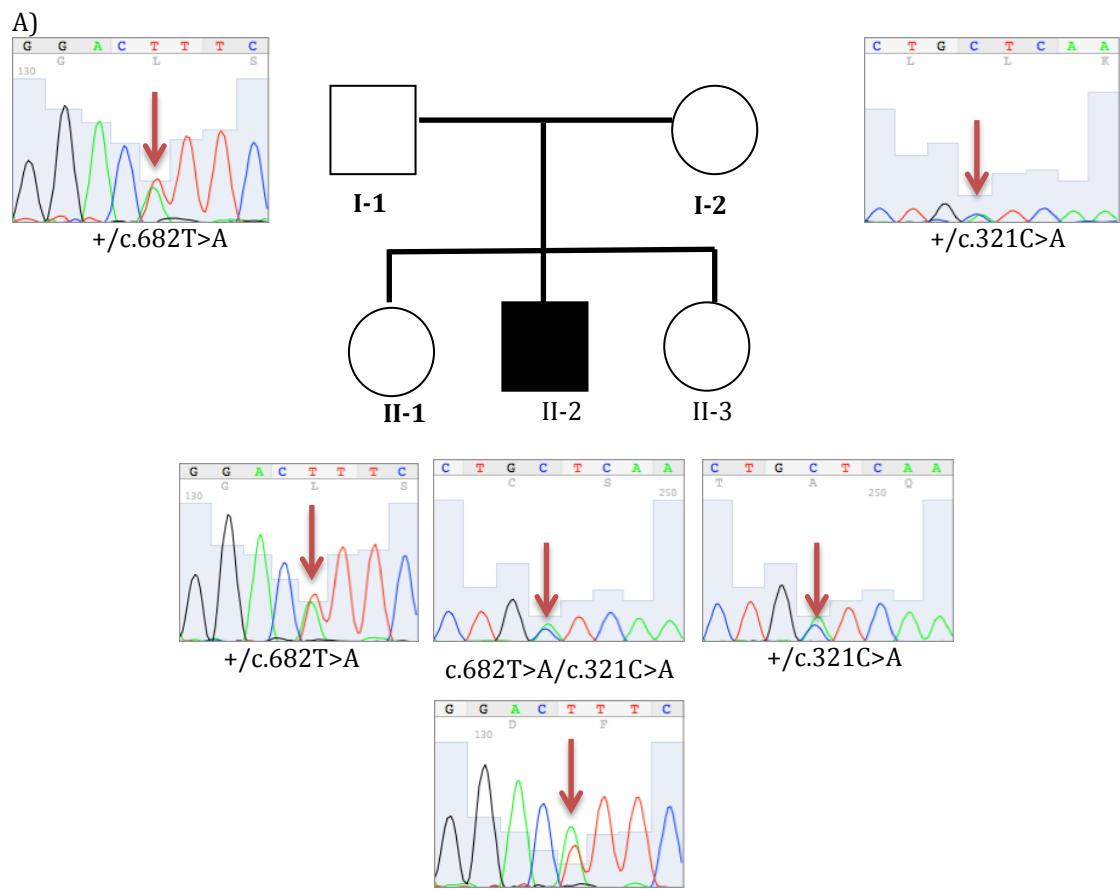


D)

Mutation	Polyphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham score
p.G213S	Deleterious	Deleterious	Disease-causing	10 chromosomes	1 chromosome	208 chromosomes	56

Supplementary Figure 21. Heterozygous missense mutation in *WNT10A* in patient V2.69

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutation in the protein. C) Sequence alignment around amino acid 213. The G213S (*) mutation affects a conserved amino acid in a functional domain. D) Table showing the bioinformatically predicted effect and the frequency of the missense mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

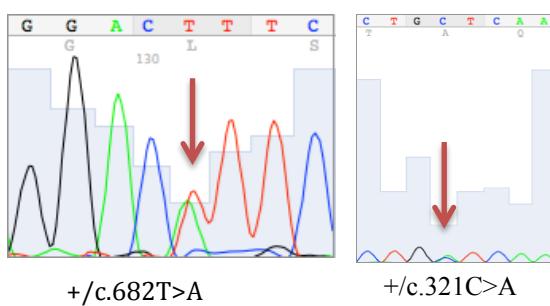
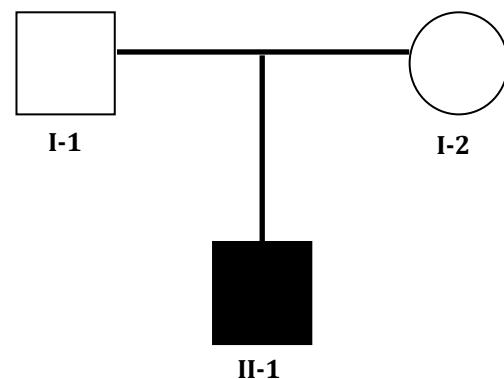


B)

Mutation	Polyphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham score
p.F228I	Deleterious	Deleterious	Disease-causing	30/4978 chromosomes	241/13000 chromosomes	1499/117764 chromosomes	21
p.C107*	-	-	-	Absent	13/130006 chromosomes	71/121352 chromosomes	-

Supplementary Figure 22. Complex heterozygous missense mutation in *WNT10A* in patient V.71

A) Family pedigree showing segregation of the mutation. The mutated nucleotides are indicated with arrows. B) Table showing the bioinformatically predicted effect and the frequency of the mutations in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

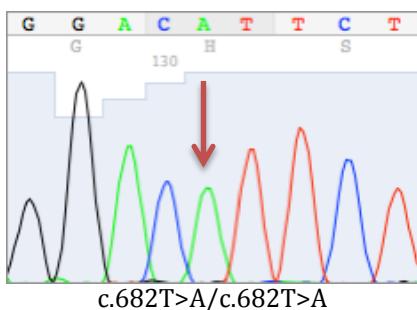
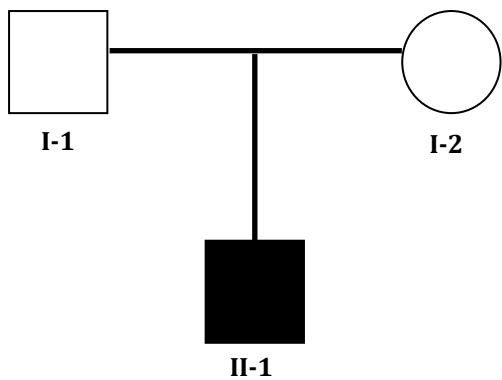


B)

Mutation	Polyphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham score
p.F228I	Deleterious	Deleterious	Disease-causing	30/4978 chromosomes	241/13000 chromosomes	1499/117764 chromosomes	21
p.C107*	-	-	-	Absent	13/130006 chromosomes	71/121352 chromosomes	-

Supplementary Figure 23. Two heterozygous missense mutations in *WNT10A* in patient V.72

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. Both mutations are known pathogenic mutations. B) Table showing the bioinformatically predicted effect and the frequency of the mutations in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

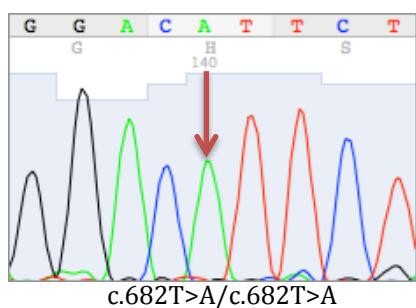
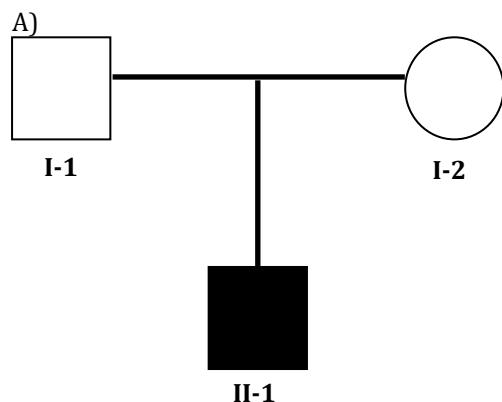


B)

Mutation	PolypHEN	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham score
p.F228I	Deleterious	Deleterious	Disease-causing	30/4978 chromosomes	241/13000 chromosomes	1499/117764 chromosomes	21

Supplementary Figure 24. Homozygous missense mutation in *WNT10A* in patient V2.74

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. The mutation is known pathogenic. B) Table showing the bioinformatically predicted effect and the frequency of the mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

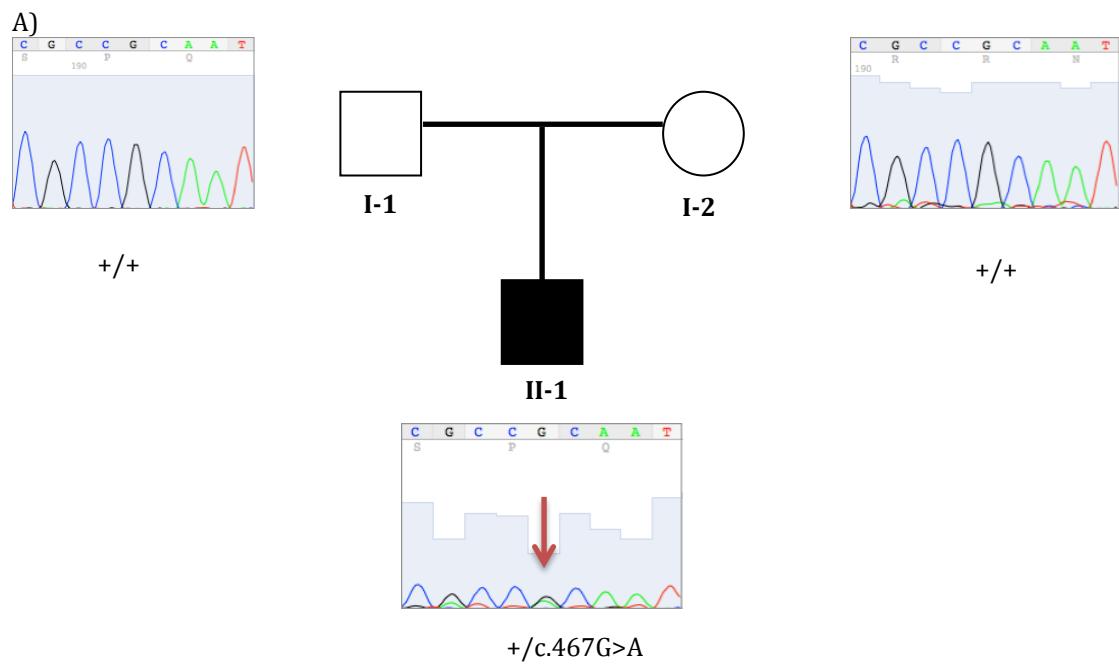


B)

Mutation	Polyphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham score
p.F228I	Deleterious	Deleterious	Disease-causing	30/4978 chromosomes	241/13000 chromosomes	1499/117764 chromosomes	21

Supplementary Figure 25. Homozygous missense mutation in *WNT10A* in patient V2.76

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Table showing the bioinformatically predicted effect and the frequency of the mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

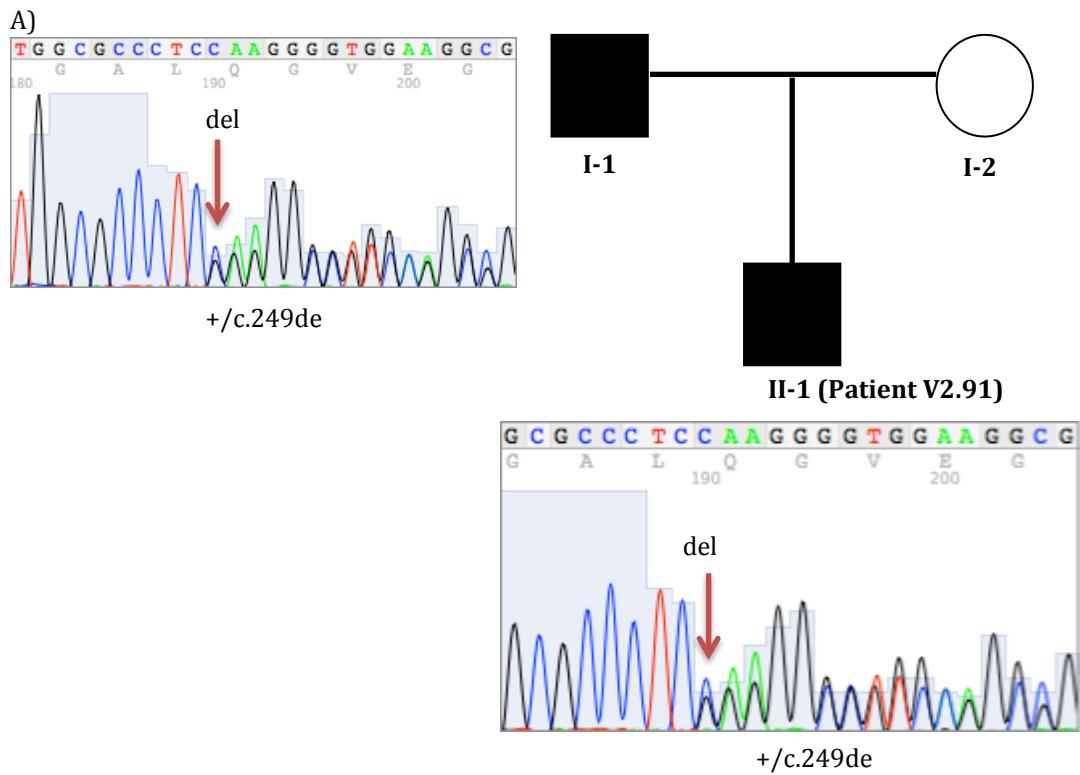


B)

Mutation	Polyphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham Score
p.R156H	Deleterious	Deleterious	Disease-causing	Absent	Absent	Absent	29

Supplementary Figure 26. Heterozygous de novo mutation in *EDA* in patient V2.78

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Table indicating the bioinformatically predicted effect of the missense mutation and its frequency in publicly available databases. EVS: Exome Variant Server; ExAC: Exome Aggregation Consortium.



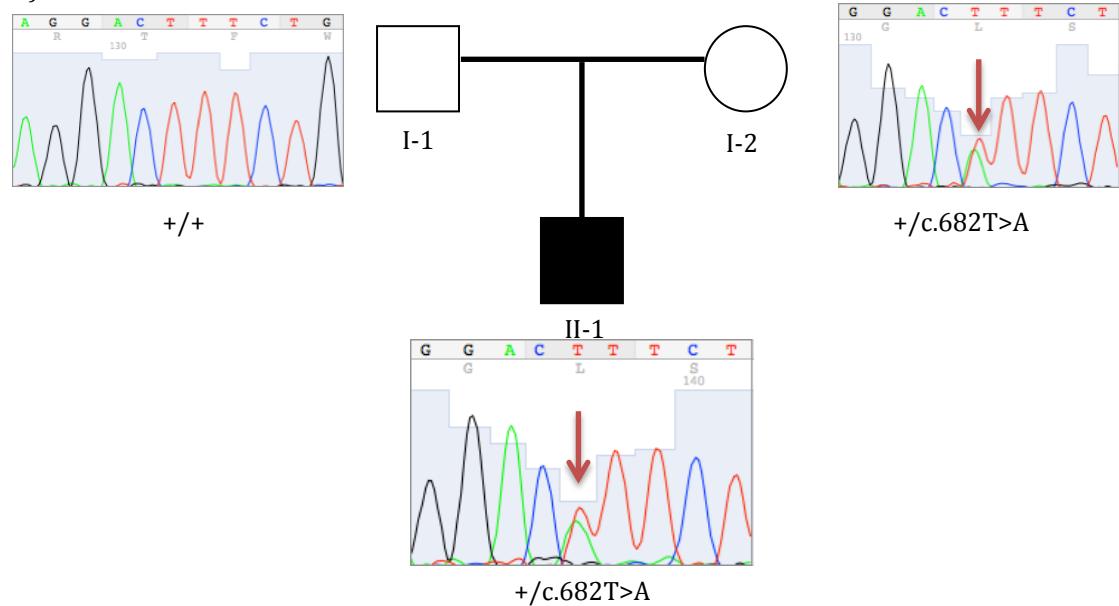
B)

Mutation	1000G	EVS	ExAC
E84Rfs*76	Absent	Absent	Absent

Supplementary Figure 27. Heterozygous frameshift inducing deletion mutation in *MSX1* in patient V2.91

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Table indicating the frequency of the mutation in publicly available databases. EVS: Exome Variant Server; ExAC: Exome Aggregation Consortium.

A)

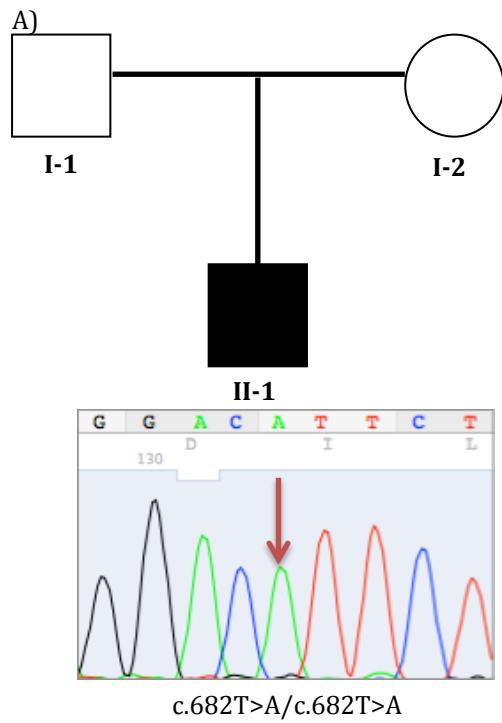


B)

Mutation	Polphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham score
p.F228I	Deleterious	Deleterious	Disease-causing	30/4978 chromosomes	241/13000 chromosomes	1499/117764 chromosomes	21

Supplementary Figure 28. Heterozygous missense mutation in *WNT10A* in patient V2.92

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. This mutation is known to show sex-biased manifestations in males in the heterozygous state [18].



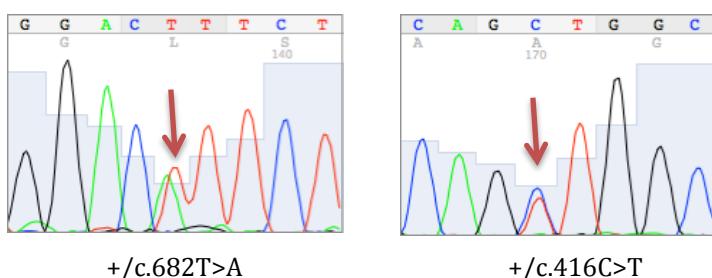
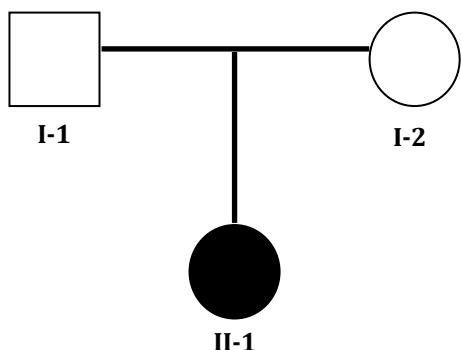
B)

Mutation	Polyphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham score
p.F228I	Deleterious	Deleterious	Disease-causing	30/4978 chromosomes	241/13000 chromosomes	1499/117764 chromosomes	21

Supplementary Figure 29. Homozygous missense mutation in *WNT10A* in patient V2.93

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow.

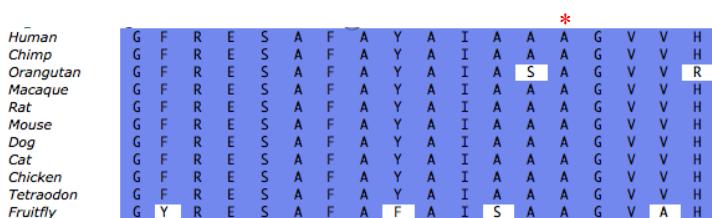
A)



B)



C)



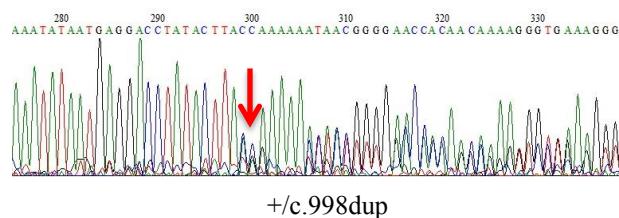
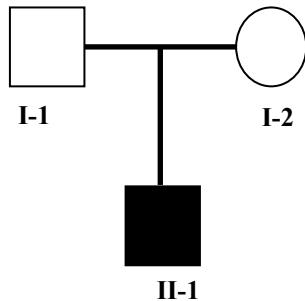
D)

Mutation	Polyphen	SIFT	MutationTaster	1000G	EVS	ExAC	Grantham Score
p.F228I	Deleterious	Delerterious	Disease-causing	30/4978 chromosomes	241/13000 chromosomes	1499/117764 chromosomes	21
p.A139V	Deleterious	Deleterious	Disease-causing	Absent	Absent	Absent	64

Supplementary Figure 30. Two heterozygous missense mutations in *WNT10A* in patient 2.54

A) Family pedigree showing segregation of the mutations. The mutated nucleotides are indicated with an arrow. B) Location of the novel mutation in the protein. C) Sequence alignment around amino acid 139 (Alamut v2.6.2). The A139V (*) mutation affects a well-conserved amino acid in a functional domain. D) Table showing the bioinformatically predicted effect and the frequency of the missense mutations in public databases. EVS: Exome Variant Server; ExAC: Exome

A)



B)

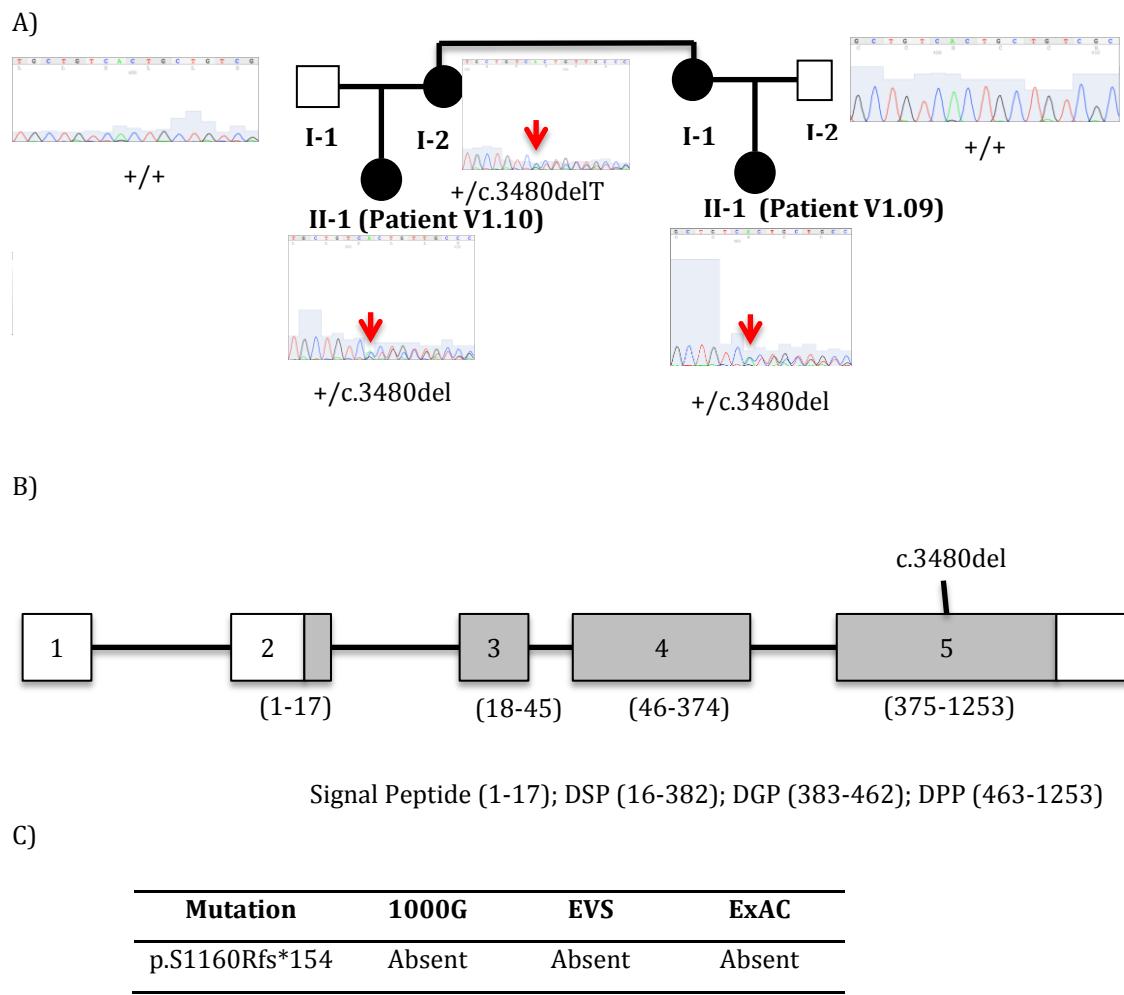


C)

Mutation	1000G	EVS	ExAC
p.Tyr333*	Absent	Absent	Absent

Supplementary Figure 31. Heterozygous single nucleotide insertion in *CTNNB1* in patient V2.87

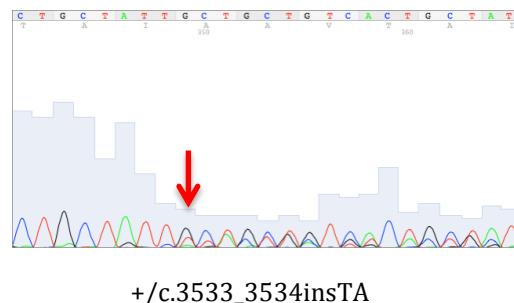
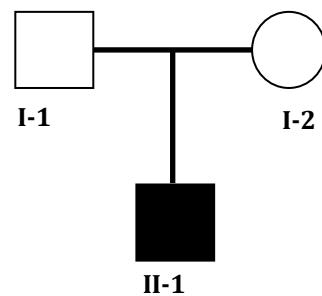
A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutation in the protein. Blue boxes represent armadillo repeats. Image adapted from [19]. C) Table showing the frequency of the missense mutations in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.



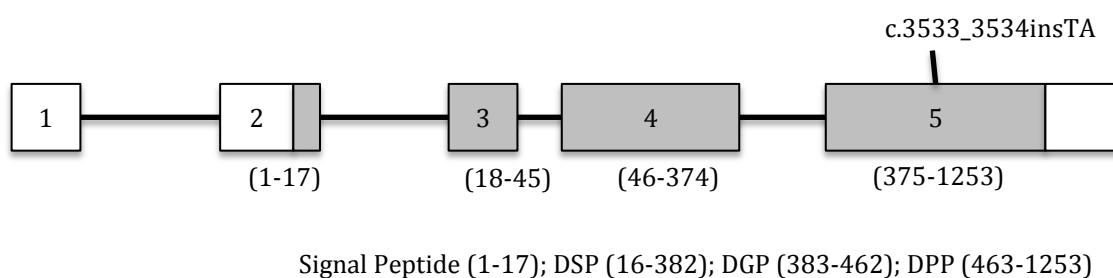
Supplementary Figure 32. Heterozygous single nucleotide deletion in *DSPP* in patients V1.09 and V1.10

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutation in the gene. White boxes represent non-coding exons whereas grey boxes represent coding sequences. The exons are numbered from 5' to 3'. Below, the amino acids that contribute to the different proteins encoded by the transcript are shown. DSP: dentin sialoprotein; DGP: Dentin glycoprotein; DPP: Dentin phosphoprotein. C) Table showing the frequency of the missense mutations in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



B)

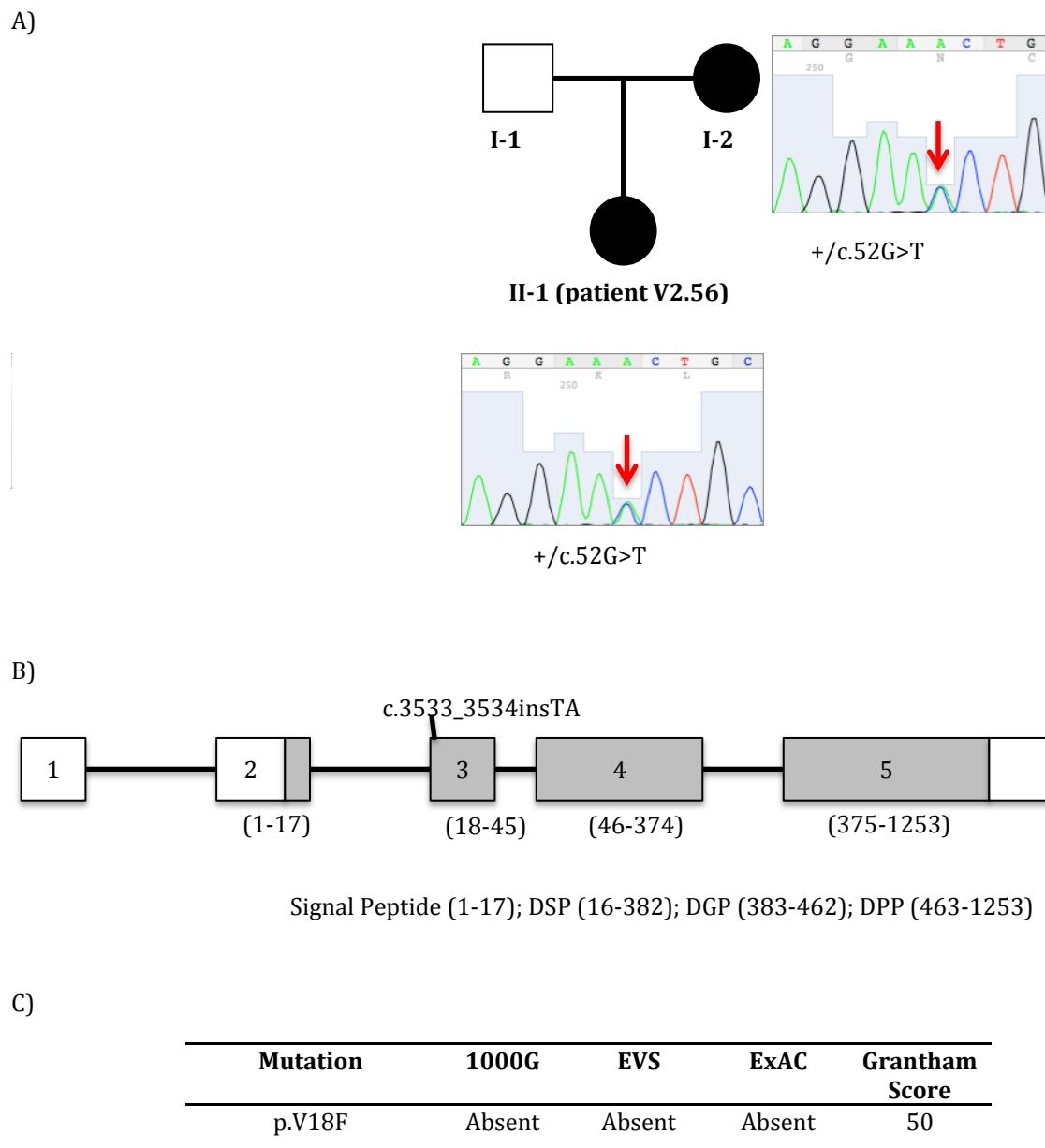


C)

Mutation	1000G	EVS	ExAC
p.Asn1179Thrfs*136	Absent	Absent	Absent

Supplementary Figure 33. Heterozygous dinucleotide deletion in *DSPP* in patient V.36

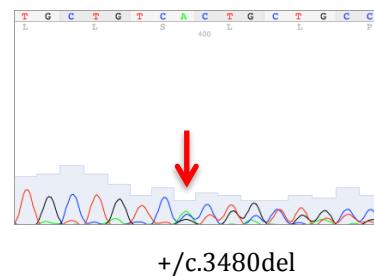
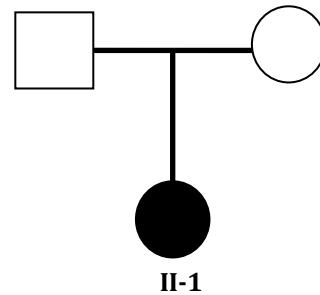
A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutation in the gene. White boxes represent non-coding exons whereas grey boxes represent coding sequences. The exons are numbered from 5' to 3'. Below, the amino acids that contribute to the different proteins encoded by the transcript are shown. DSP: dentin sialoprotein; DGP: Dentin glycoprotein; DPP: Dentin phosphoprotein. C) Table showing the frequency of the missense mutations in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.



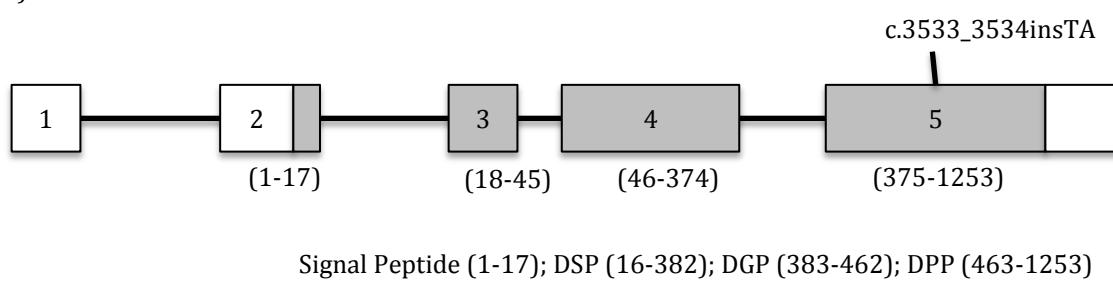
Supplementary Figure 34. Heterozygous missense mutation in *DSPP* in patient V2.55

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutation in the gene. White boxes represent non-coding exons whereas grey boxes represent coding sequences. The exons are numbered from 5' to 3'. Below, the amino acids that contribute to the different proteins encoded by the transcript are shown. DSP: dentin sialoprotein; DGP: Dentin glycoprotein; DPP: Dentin phosphoprotein. C) Table showing the frequency of the missense mutations in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



B)

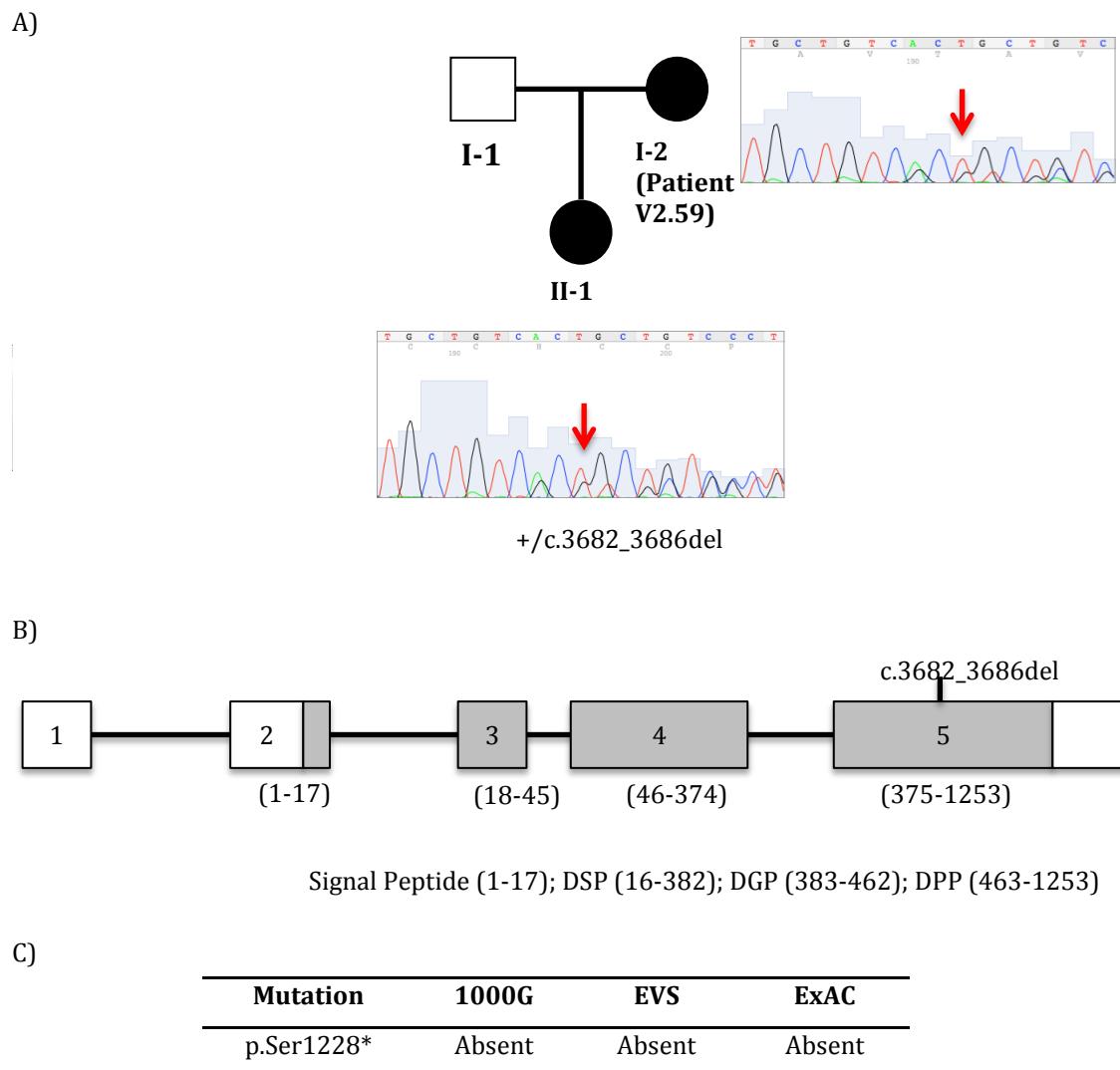


C)

Mutation	1000G	EVS	ExAC
p.Asn1179Thrf*136	Absent	Absent	Absent

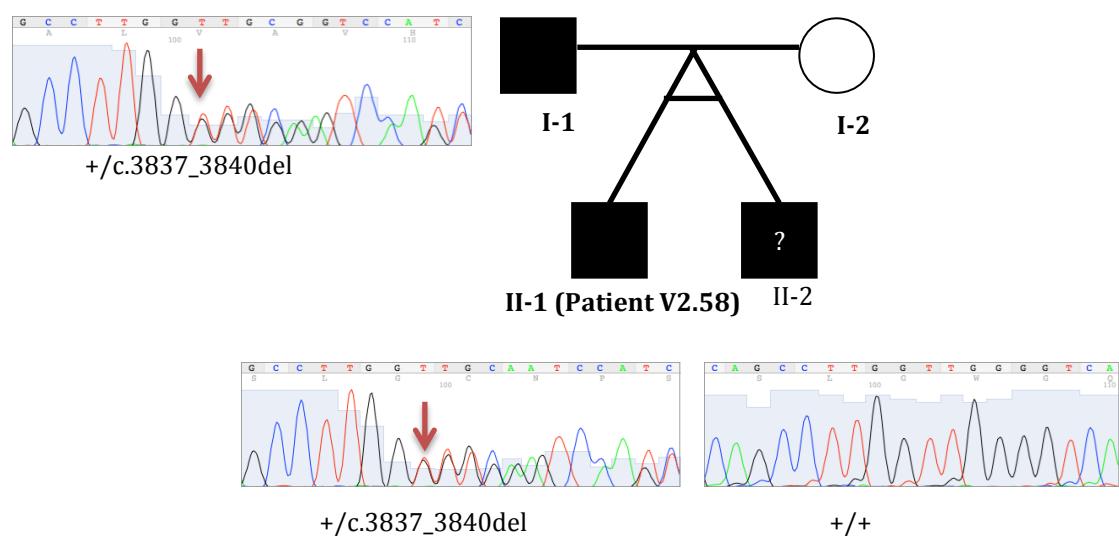
Supplementary Figure 35. Heterozygous dinucleotide deletion in *DSPP* in patient V2.57

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutation in the gene. White boxes represent non-coding exons whereas grey boxes represent coding sequences. The exons are numbered from 5' to 3'. Below, the amino acids that contribute to the different proteins encoded by the transcript are shown. DSP: dentin sialoprotein; DGP: Dentin glycoprotein; DPP: Dentin phosphoprotein. C) Table showing the frequency of the missense mutations in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.



Supplementary Figure 36. Heterozygous dinucleotide deletion in *DSPP* in patient V2.59
A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutation in the gene. White boxes represent non-coding exons whereas grey boxes represent coding sequences. The exons are numbered from 5' to 3'. Below, the amino acids that contribute to the different proteins encoded by the transcript are shown. DSP: dentin sialoprotein; DGP: Dentin glycoprotein; DPP: Dentin phosphoprotein. C) Table showing the frequency of the missense mutations in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



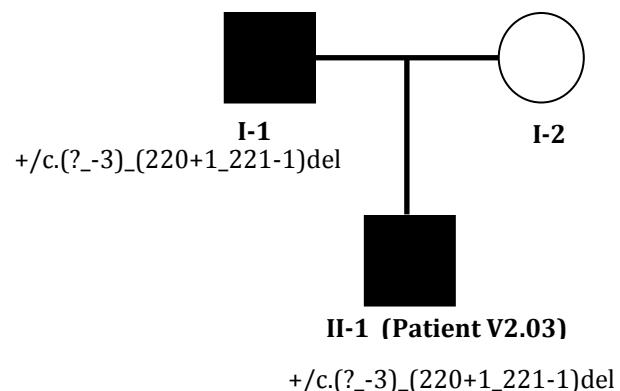
B)

Mutation	1000G	EVS	ExAC
p.N1279Lfs*51	Absent	Absent	Absent

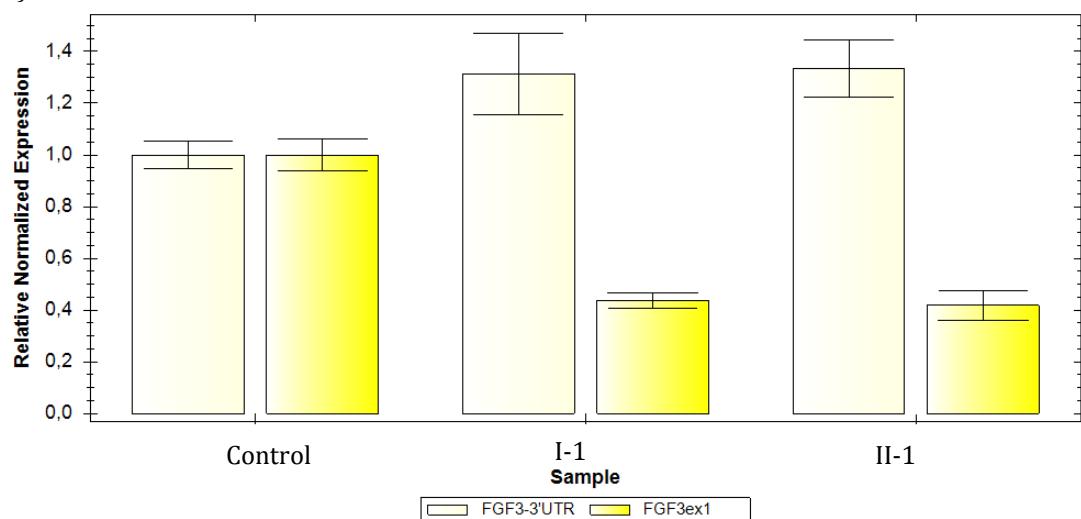
Supplementary Figure 37. Heterozygous deletion in *COL1A1* in patient V2.58

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. The phenotype in II-2 is less severe than that of I-1 and II-1 and thus may have a distinct aetiology. B) Table showing the frequency of the mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



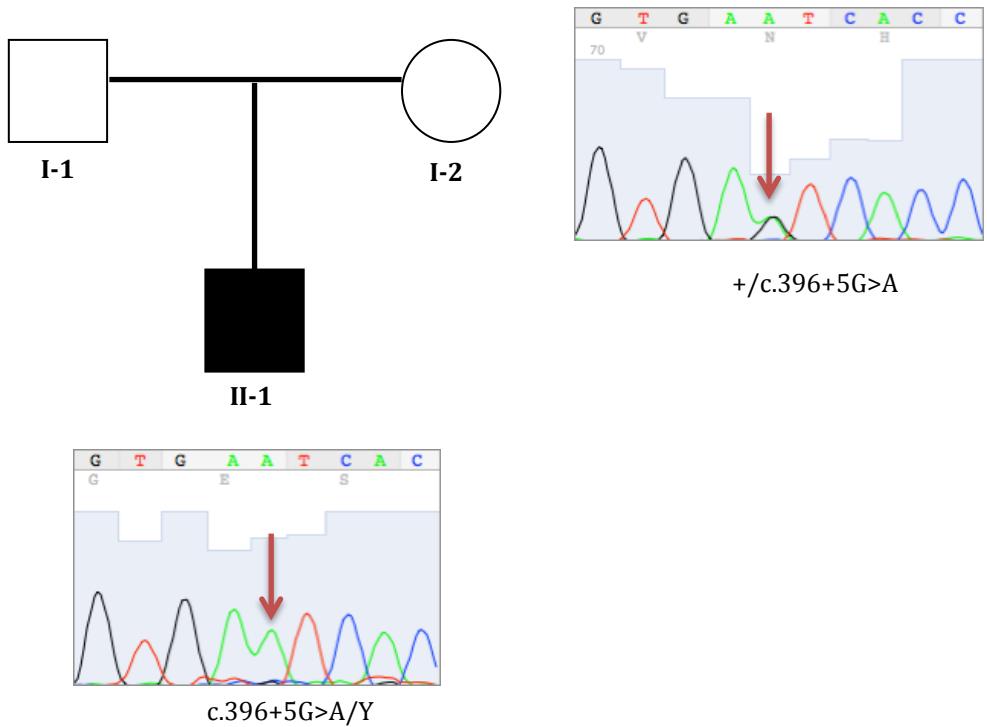
B)



Supplementary Figure 38. Heterozygous deletion in *FGF3* in patient V2.03

A) Family pedigree. B) Quantitative PCR (qPCR) validation of the deletion in the family.

A)

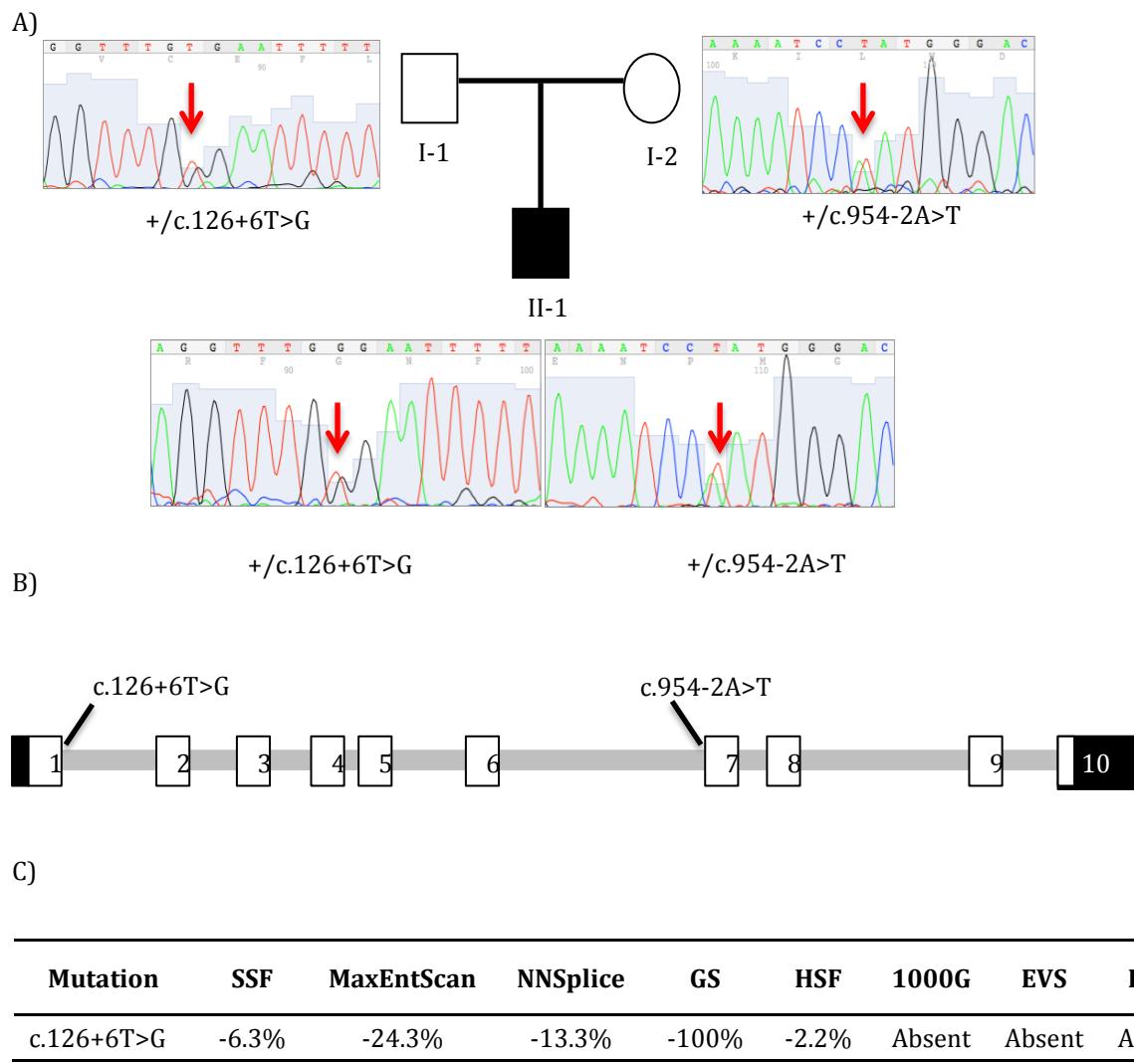


B)

Mutation	SSF	MaxEntScan	NNSplice	GS[20]	HSF[15]	1000G	EVS	ExAc
c.396+5G>A	-12.7%	-38.2%	-2.8%	-36.9%	-12.3%	Absent	Absent	Absent

Supplementary Figure 39. Hemizygous missense mutation in *EDA* in patient V1.12

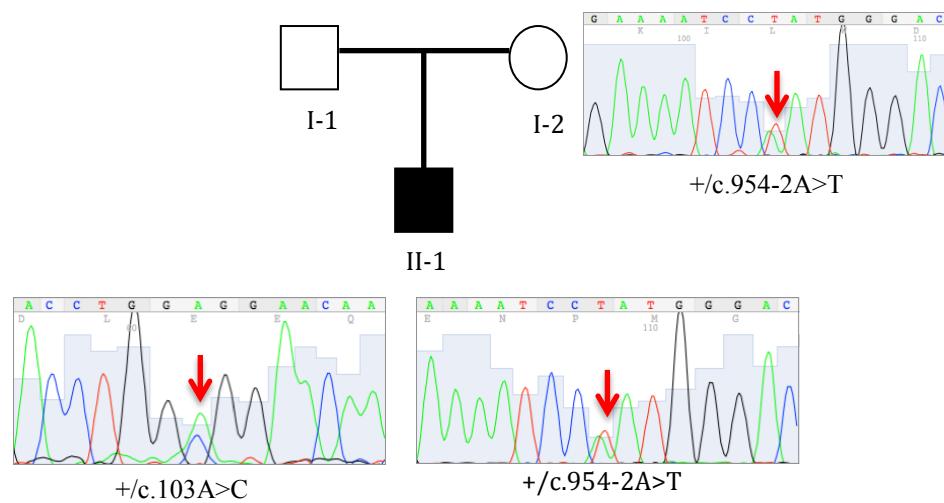
A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Table showing the bioinformatically predicted effect on splicing and the frequency of the missense mutation in public databases. SSF: Splice Site Finder; GS: Gene Splicer; HSF: Human Splice Finder; EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.



Supplementary Figure 40. Complex heterozygous mutations in MMP20 in patient V2.07

A) Family pedigree showing segregation of the mutations. The mutated nucleotides are indicated with an arrow. B) Location of the two mutations in the gene. White rectangles represent coding exons whereas black rectangles represent non-coding regions. The exons are numbered from 5' to 3'. C) Table showing bioinformatically predicted effect on splicing and the frequency of the c.126+6T>G mutation in public databases. SSF: Splice Site Finder; GS: GeneSplicer; HSF: Human Splice Finder; EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.

A)



B)

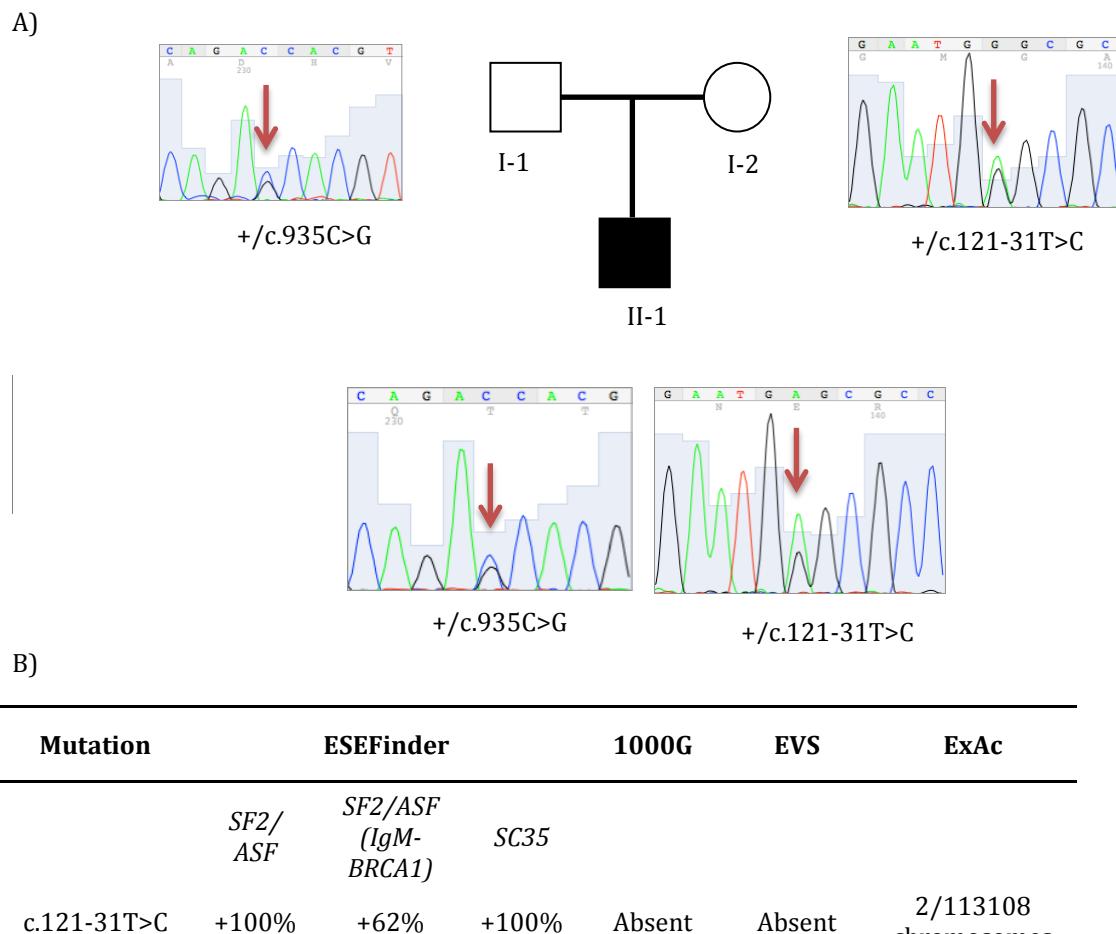


C)

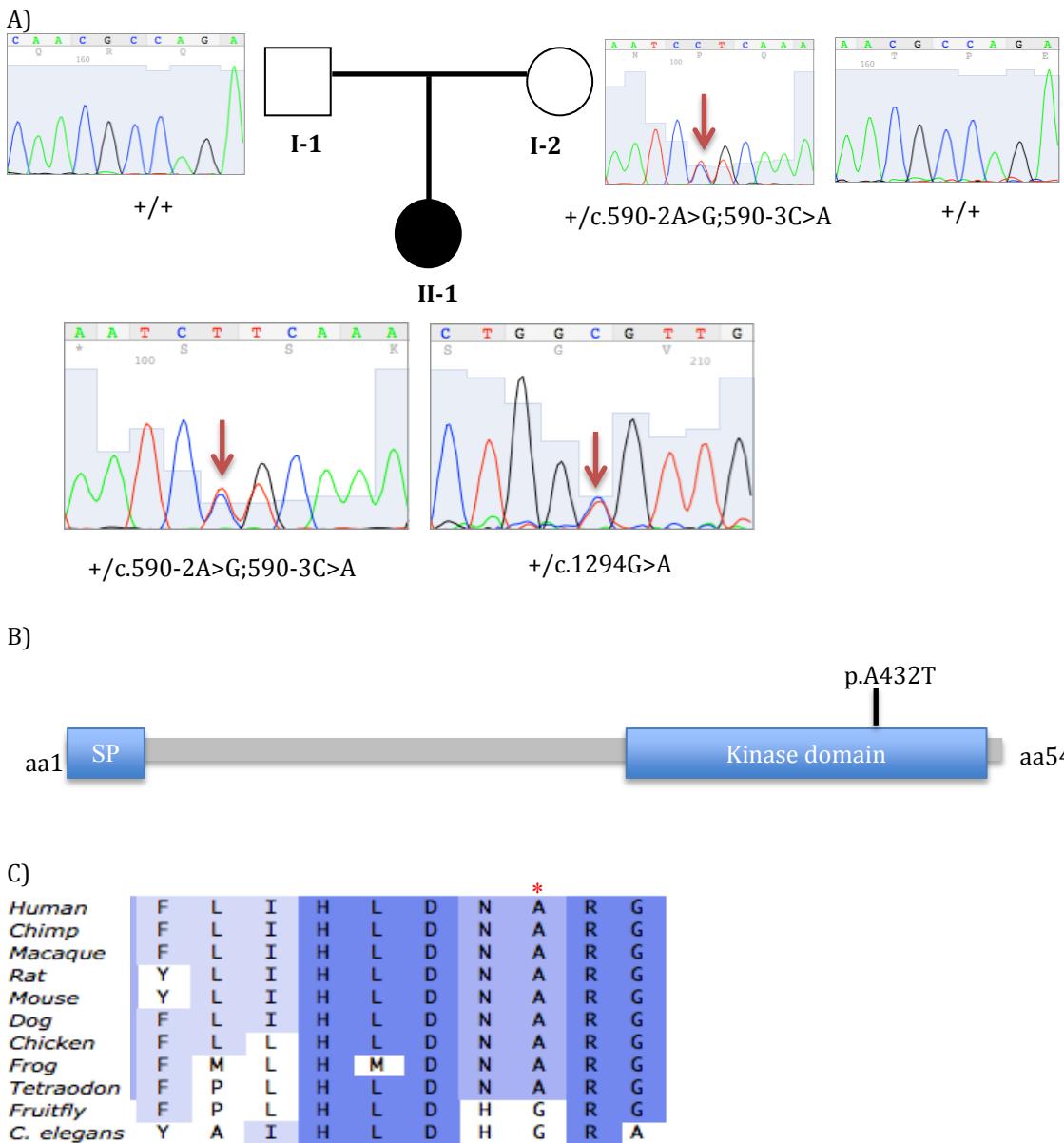
Mutation	ESEFinder[21]	Rescue ESE[22]	1000G	EVS	ExAc
<i>SF2/ASF</i> (<i>IgM-BRCA1</i>)	<i>SF2/ASF</i> (<i>IgM-BRCA1</i>)	<i>SC35</i>			
c.103A>C	-100%	-100%	=	-100%	Absent Absent 20/121092 chromosomes

Supplementary Figure 41. Two heterozygous mutations in *MMP20* in patient V2.13

A) Family pedigree showing segregation of the mutations. The mutated nucleotides are indicated with an arrow. B) Location of the two mutations in the gene. White rectangles represent coding exons whereas black rectangles represent non-coding regions. The exons are numbered from 5' to 3'. C) Table showing bioinformatically predicted effect on splicing and the frequency of the mutation in public databases. The synonymous variant is predicted to destroy an exonic splice enhancer. ESE: exonic splice enhancer; EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.



Supplementary Figure 42. Complex heterozygous mutations in *GALNS* in patient V2.15
A) Family pedigree showing segregation of the mutations. The mutated nucleotides are indicated with an arrow. B) Table showing bioinformatically predicted effect on splicing and the frequency of the intronic mutation in public databases. The intronic variant is predicted to create an intronic splice enhancer that may disrupt normal splicing. ESE: exonic splice enhancer; EVS: Exome Variant Server; ExAC: Exome aggregation consortium server.



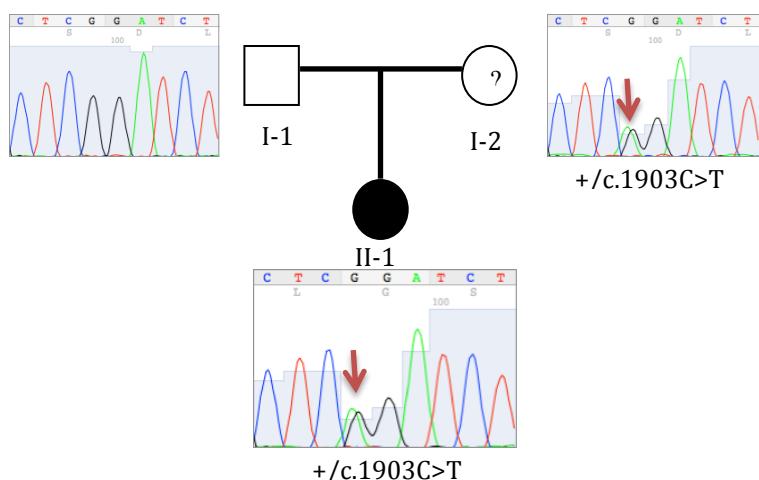
D)

Mutation	Polyphen	SIFT	MutationTaster	1000G	EVS	ExAC
c.590-3C>A	-	-	-	Absent	Absent	1/120086 chromosome
p.A432T	Deleterious	Deleterious	Disease-causing	Absent	Absent	5/119942 chromosomes

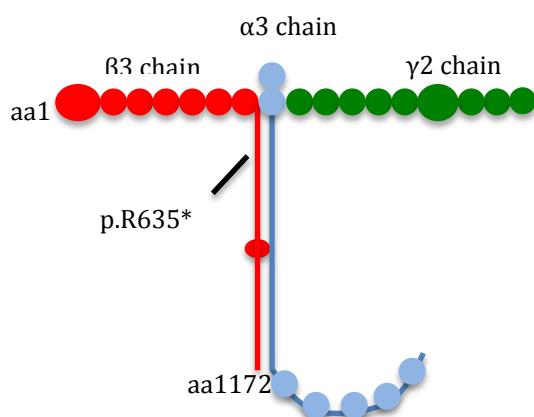
Supplementary Figure 43. Three heterozygous mutations in FAM20A in patient V2.32

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the missense mutation in the protein. Image adapted from [23]. C) Multi-species sequence conservation around amino acid 432 (Alamut v2.6.2). The p.A432T mutation affects a well-conserved residue (marked by an asterisk) in a functional WD40 domain. D) Table showing the bioinformatically predicted effect of and the frequency of the missense mutation in public databases. SSF: Splice Site Finder; EVS: Exome Variant Server; ExAC: Exome aggregation consortium server. The phase of the mutations could not be established as the c.1294G>A mutation is *de novo* and is 16 kb away from the splice site mutations.

A)



B)



C)

Mutation	1000G	EVS	ExAC
p.R635*	Absent	7/130006 chromosomes	92/121034 chromosomes

Supplementary Figure 44. Heterozygous nonsense mutation in *LAMB3* in patient V2.86 – Incomplete penetrance or variable expressivity in the mother?

A) Family pedigree showing segregation of the mutation. The mutated nucleotide is indicated with an arrow. B) Location of the mutated residue in the protein. Image adapted from [12]. C) Table showing the frequency of the mutation in public databases. EVS: Exome Variant Server; ExAC: Exome aggregation consortium server. The mother was noted as unaffected in the preliminary dental examination. She was not available for examination after obtaining the genotyping results. However, there is evidence of variable expressivity of *LAMB3* mutations.[24] Given the known role of heterozygous *LAMB3* nonsense mutations in bacteria and the frequency of this mutation in the general population [Freq = .00076 in ExAC, .0014 < freq of AI < .00007[25]], it is likely that this mutation is associated with a variably expressed enamel phenotype. Additional investigations such as a re-analysis of the mother's dental phenotype, inclusion of additional family members who may be carriers of the mutation, and ascertainment and dental examination of additional families segregating this mutation will be necessary in order to conclusively determine the pathogenicity of this variant.

A)

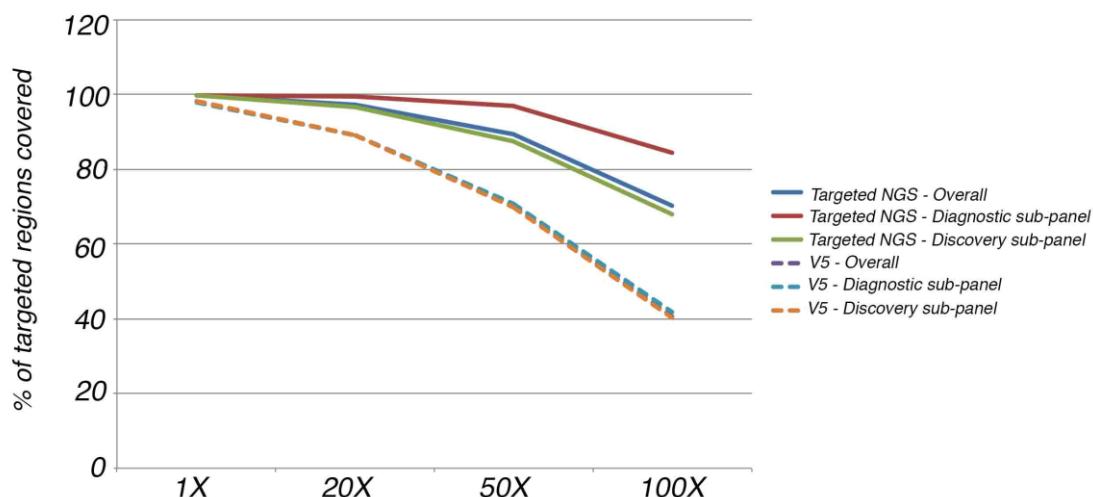


B)



Supplementary Figure 45. Enamel phenotype seen in patients V2.15 and V2.49 in the context of *GALNS* mutations.

A) Dental photograph and panoramic radiograph of patient V2.15. B) Dental photograph of patient V2.49.



Supplementary Figure 46. Comparison of sequencing coverage of the targeted regions achieved with v2.0 of the gene panel and with whole exome sequencing

The percentage of the targeted regions in the v2.0 panel as well as the two sub-panels covered at different thresholds is shown. V5: Agilent SureSelect V5. Coverage calculations included only high-quality uniquely mapped reads ($\text{MQ} > 30$). The values for the V5-Overall and V5-Discovery sub-panels are close, such that the two lines overlap.

REFERENCES

1. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nature methods* 2010;7(4):248-9 doi: 10.1038/nmeth0410-248 [published Online First: 2010/04/01].
2. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* 2009;4(7):1073-81 doi: 10.1038/nprot.2009.86 [published Online First: 2009/06/30].
3. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature methods* 2014;11(4):361-2 doi: 10.1038/nmeth.2890 [published Online First: 2014/04/01].
4. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491(7422):56-65 doi: 10.1038/nature11632 [published Online First: 2012/11/07].
5. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA. URL: <http://evs.gs.washington.edu/EVS/> (accessed 2/2015).
6. Exome Aggregation Consortium (ExAC). <http://exac.broadinstitute.org> (accessed 02/2015).
7. El-Sayed W, Parry DA, Shore RC, et al. Mutations in the beta propeller WDR72 cause autosomal-recessive hypomaturation amelogenesis imperfecta. *American journal of human genetics* 2009;85(5):699-705 doi: 10.1016/j.ajhg.2009.09.014 [published Online First: 2009/10/27].
8. Parry DA, Mighell AJ, El-Sayed W, et al. Mutations in CNM4 cause Jalili syndrome, consisting of autosomal-recessive cone-rod dystrophy and amelogenesis imperfecta. *American journal of human genetics* 2009;84(2):266-73 doi: 10.1016/j.ajhg.2009.01.009 [published Online First: 2009/02/10].
9. McCulloch DL, Marmor MF, Brigell MG, et al. ISCEV Standard for full-field clinical electroretinography (2015 update). *Documenta ophthalmologica. Advances in ophthalmology* 2015;130(1):1-12 doi: 10.1007/s10633-014-9473-7 [published Online First: 2014/12/17].
10. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic acids research* 2014;42(Database issue):D986-92 doi: 10.1093/nar/gkt958 [published Online First: 2013/11/01].
11. Pasmooij AM, Pas HH, Jansen GH, Lemmink HH, Jonkman MF. Localized and generalized forms of blistering in junctional epidermolysis bullosa due to COL17A1 mutations in the Netherlands. *The British journal of dermatology* 2007;156(5):861-70 doi: 10.1111/j.1365-2133.2006.07730.x [published Online First: 2007/02/01].
12. Nakano A, Chao SC, Pulkkinen L, et al. Laminin 5 mutations in junctional epidermolysis bullosa: molecular basis of Herlitz vs. non-Herlitz phenotypes. *Human genetics* 2002;110(1):41-51 doi: 10.1007/s00439-001-0630-1 [published Online First: 2002/01/26].
13. Wright JT, Frazier-Bowers S, Simmons D, et al. Phenotypic variation in FAM83H-associated amelogenesis imperfecta. *Journal of dental research* 2009;88(4):356-60 doi: 10.1177/0022034509333822 [published Online First: 2009/05/02].

14. Kida M, Sakiyama Y, Matsuda A, et al. A novel missense mutation (p.P52R) in amelogenin gene causing X-linked amelogenesis imperfecta. *Journal of dental research* 2007;86(1):69-72 [published Online First: 2006/12/26].
15. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research* 2009;37(9):e67 doi: 10.1093/nar/gkp215 [published Online First: 2009/04/03].
16. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *Journal of computational biology : a journal of computational molecular cell biology* 1997;4(3):311-23 [published Online First: 1997/10/01].
17. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of computational biology : a journal of computational molecular cell biology* 2004;11(2-3):377-94 doi: 10.1089/1066527041410418 [published Online First: 2004/08/03].
18. Bohring A, Stamm T, Spaich C, et al. WNT10A mutations are a frequent cause of a broad spectrum of ectodermal dysplasias with sex-biased manifestation pattern in heterozygotes. *American journal of human genetics* 2009;85(1):97-105 doi: 10.1016/j.ajhg.2009.06.001 [published Online First: 2009/06/30].
19. Kuechler A, Willemsen MH, Albrecht B, et al. De novo mutations in beta-catenin (CTNNB1) appear to be a frequent cause of intellectual disability: expanding the mutational and clinical spectrum. *Human genetics* 2015;134(1):97-109 doi: 10.1007/s00439-014-1498-1 [published Online First: 2014/10/20].
20. Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic acids research* 2001;29(5):1185-90 [published Online First: 2001/02/27].
21. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic acids research* 2003;31(13):3568-71 [published Online First: 2003/06/26].
22. Fairbrother WG, Yeh RF, Sharp PA, Burge CB. Predictive identification of exonic splicing enhancers in human genes. *Science* 2002;297(5583):1007-13 doi: 10.1126/science.1073774 [published Online First: 2002/07/13].
23. Tagliabracci VS, Pinna LA, Dixon JE. Secreted protein kinases. *Trends in biochemical sciences* 2013;38(3):121-30 doi: 10.1016/j.tibs.2012.11.008 [published Online First: 2013/01/02].
24. Lee KE, Ko J, Le CG, et al. Novel LAMB3 mutations cause non-syndromic amelogenesis imperfecta with variable expressivity. *Clinical genetics* 2015;87(1):90-2 doi: 10.1111/cge.12340 [published Online First: 2014/02/06].
25. Crawford PJ, Aldred M, Bloch-Zupan A. Amelogenesis imperfecta. *Orphanet journal of rare diseases* 2007;2:17 doi: 10.1186/1750-1172-2-17 [published Online First: 2007/04/06].