Supplementary Information

Supplementary tables

Table S1. Exome sequencing capture statistics

Sample ID	Total Reads	PostQC Reads	Mapped	Duplicates	On Target	Data in Gb	bases at 1x	bases at 4x	bases at 8x	bases at 20x	Average Coverage
Family A											
I-1	134702157	133756269	99.30%	17.00%	75.66%	13.51	99.90%	99.81%	99.68%	98.90%	139.48
I-2	147618262	146756580	99.42%	13.00%	74.04%	14.82	99.72%	99.60%	99.48%	98.89%	149.31
II-1	113057031	112720329	99.70%	10.00%	73.95%	11.38	99.67%	99.53%	99.33%	98.21%	114.28
II-2	124971214	124486763	99.61%	13.00%	75.45%	12.57	99.70%	99.58%	99.43%	98.61%	129.38
Family B											
Sample ID	**Total Reads**	**PostQC Reads**	**Mapped**	**Duplicates**	**On Target**	**Data in Gb**	**bases at 1x**	**bases at 4x**	**bases at 8x**	**bases at 20x**	**Average Coverage**
II-1	105671388	104827500	99.20%	11.00%	79.24%	10.59	99.82%	99.61%	99.24%	97.34%	136.4
Family C											
Sample ID	**Total Reads**	**PostQC Reads**	**Mapped**	**Duplicates**	**On Target**	**On Target +/- 100nt**	**bases at 1x**	**bases at 10x**	**bases at 20x**	**bases at 50x**	**Average Coverage**
III-1	133200258	132889362	87.90%	43.43%	79.36%	86.14%	98.22%	96.94%	95.69%	86.27%	87.69
IV-1	103423684	103183336	87.79%	43.53%	79.43%	86.09%	98.11%	96.61%	94.77%	74.30%	67.81

IV-3	119460904	119149033	88.21%	39.79%	79.93%	87.48%	98.18%	96.81%	95.43%	84.26%	84.48
IV-4	119744314	119442406	87.73%	41.03%	79.02%	86.53%	98.33%	97.09%	95.65%	81.62%	81.44
IV-7	108627526	108396013	88.95%	42.56%	80.42%	87.20%	98.12%	96.64%	94.97%	79.10%	73.80
IV-11	136178810	135839398	87.83%	42.23%	79.28%	86.60%	98.27%	97.10%	95.97%	87.96%	91.41
Family D											
Sample ID	**Total Reads**	**Mapped**	**Duplicates**	**On Target**	**On Target +/- 100nt**	**bases at 1x**	**bases at 10x**	**bases at 20x**	**bases at 50x**	**Average Coverage**	
I-1	90014486	99.58%	8.35%	69.90%	87.55%	98.62%	96.51%	90.32%	56.98%	64.42	
I-2	89540700	99.52%	5.85%	70.88%	87.76%	98.47%	96.86%	92.10%	59.28%	66.00	
II-1	92561584	99.60%	6.50%	70.17%	87.64%	98.50%	97.04%	92.94%	60.97%	68.05	

Table S2. Recessive protein-altering variants (minor allele frequency < 1%) identified in probands

Gene	OMIM disease annotation	Genomic position	cDNA variant	Protein variant	Mode	Sanger* sequencing confirmation	CADD score**	HGMD entry	Frequency gnomAD (homozy-gotes)***	Score according to ACMG criteria
Family A										
SLC44A5	no	chr1:75708628	NM_152697: c.414C>G	NP_689910: p.Asp138Glu	compound heterozygous	ND	12	not found	988/281344 (6)	benign
SLC44A5	no	chr1:75688101	NM_152697: c.1030A>C	NP_689910: p.Ile344Leu	compound heterozygous	ND	23	not found	2088/282226 (12)	benign
ZNF142	**no**	**chr2:219513813_219513814**	**NM_001105537: c.817_818delAA**	**NP_001099007: p.Lys273Glufs*32**	**compound heterozygous**	**yes**	**33**	**not found**	**not found**	**Likely pathogenic**
ZNF142	**no**	**chr2:219511053**	**NM_001105537: c.1292delG**	**NP_001099007: p.Cys431Leufs*11**	**compound heterozygous**	**yes**	**35**	**not found**	**not found**	**Likely pathogenic**
Family B										
ZNF142	**no**	**chr2:219508064**	**NM_001105537: c.3175C>T**	**NP_001099007: p.Arg1059***	**homozygous**	**yes**	**37**	**not found**	**2/249310 (0)**	**Likely pathogenic**
SLC19A3	MIM: 607483 (thiamine metabolism dysfunction syndrome 2, autosomal-recessive)	chr2:228563479	NM_025243: c.952G>A	NP_079519: p.Ala318Thr	homozygous	yes	28	not found	not found	VUS
Family C										
ZNF142	**no**	**chr2:219507056, chr2:219507054**	**NM_001105537: c.4183delC+4185G>A**	**NP_001099007: p.Leu1395***	**homozygous**	**yes**	35	**not found**	**not found**	**Likely pathogenic**
RUFY4	no	chr2:218938589	NM_198483.3: c.581A>T	NP_940885.2: p.Asn194Ile	homozygous	ND	4.5	**not found**	**320/280382 (03)**	benign
Family D										
ZNF142	**no**	**chr2:219507541**	**NM_001105537: c.3698G>T**	**NP_001099007: p.Cys1233Phe**	**compound heterozygous**	**yes**	**31**	**not found**	**not found**	**VUS**
ZNF142	**no**	**chr2:219505483**	**NM_001105537: c.4498C>T**	**NP_001099007: p.Arg1500Trp**	**compound heterozygous**	**yes**	**26**	**not found**	**not found**	**VUS**

*ND, not done. **CADD model: GRCh37-v1.4; PHRED scores are listed. ****ZNF142* variants were absent from our in-house dataset of 10,000 exomes, which includes >300 of Czech and Slovakian descent, and >500 individuals of Turkish origin.

Table S3. Heterozygous protein-altering variants (minor allele frequency < 1%) identified in probands

Gene	OMIM disease annotation	Genomic position	cDNA variant	Protein variant	Mode	Sanger* sequencing confirmation	CADD score**	HGMD entry	Frequency gnomAD	Significance according to ACMG
HGMD-annotated variants (heterozygous) in neurologically relevant disease genes identified in proband B-II-1 (family B)										
GMPPB	MIM: 615350-52 (muscular dystrophy-dystroglycanopathy types 14 A-C, recessive)	chr3:49759776	NM_013334: c.656T>C	NP_037466: p.Ile219Thr	heterozygous	ND	24	CM146305	17/282458	pathogenic
MUT	MIM: 251000 (methylmalonic aciduria mut(0) type, autosomal-recessive)	chr6:49419420	NM_000255: c.1091A>C	NP_000246: p.Tyr364Ser	heterozygous	ND	26	CM074361	11/281422	pathogenic
Heterozygous variants identified in proband II-2 (Family D)										
MDGA2	no	chr14:47341163	uc010ani.3 c.1226G>A	uc010ani.3: p.Gly409Asp	de novo heterozygous	ND	9	not found	not found	VUS

*ND, not done. **CADD model: GRCh37-v1.4; PHRED scores are listed. VUS, variant of uncertain significance.

Table S4. C_2H_2-type zinc finger proteins with clinical synopses. The 720 genes in the C_2H_2 family were cross-referenced with OMIM clinical synopses and classified as involved in neurodevelopmental processes (NDD) or other processes (OT). The NDDM designation was assigned to genes with a 'neurologic' and/or 'muscle, soft tissue' clinical synopsis, and also includes 'head and neck' clinical synopses only when they comprise macrocephaly or microcephaly.

C_2H_2 family gene names	Protein Accession	Neurodevelopmental involvement	OMIM clinical synopses
CTCF	P49711	NDD	MENTAL RETARDATION, AUTOSOMAL DOMINANT 21; MRD21
EGR2	P11161	NDD	HYPERTROPHIC NEUROPATHY OF DEJERINE-SOTTAS
GFI1B	Q5VTD9	OT	BLEEDING DISORDER, PLATELET-TYPE, 17; BDPLT17
GLI1		OT	POLYDACTYLY, POSTAXIAL, TYPE A8; PAPA8
GLI2		NDD	CULLER-JONES SYNDROME; CJS
GLI3	P10071	NDD	POLYDACTYLY, POSTAXIAL, TYPE A1; PAPA1
IKZF1		OT	IMMUNODEFICIENCY, COMMON VARIABLE, 13; CVID13
KLF1	Q13351	OT	ANEMIA, CONGENITAL DYSERYTHROPOIETIC, TYPE IV; CDAN4
MECOM	Q03112	OT	RADIOULNAR SYNOSTOSIS WITH AMEGAKARYOCYTIC THROMBOCYTOPENIA 2; RUSAT2
PRDM5	Q9NQX1	NDD	BRITTLE CORNEA SYNDROME 2; BCS2
PRDM6	Q9NQX0	OT	PATENT DUCTUS ARTERIOSUS 3; PDA3
PRDM12	Q9H4Q4	NDD	NEUROPATHY, HEREDITARY SENSORY AND AUTONOMIC, TYPE VIII; HSAN8
PRDM16		OT	LEFT VENTRICULAR NONCOMPACTION 8; LVNC8
SNAI1	O95863	OT	WAARDENBURG SYNDROME, TYPE 2D; WS2D
SP7	Q8TDD2	NDD	OSTEOGENESIS IMPERFECTA, TYPE XII; OI12
TRPS1	Q9UHF7	NDD	TRICHORHINOPHALANGEAL SYNDROME, TYPE I; TRPS1
TSHZ1	Q6ZSZ6	OT	AURAL ATRESIA, CONGENITAL; CAA
WT1	P19544	NDD	WILMS TUMOR 1; WT1
YY1		NDD	GABRIELE-DE VRIES SYNDROME; GADEVS
ZEB2		NDD	MOWAT-WILSON SYNDROME; MOWS
ZIC2	O95409	NDD	HOLOPROSENCEPHALY 5; HPE5
HIVEP2	P31629	NDD	MENTAL RETARDATION, AUTOSOMAL DOMINANT 43; MRD43
ZNF41	P51814	NDD	MENTAL RETARDATION, X-LINKED 89; MRX89
BCL6	P41182	NDD	MICROPHTHALMIA, SYNDROMIC 2; MCOPS2
ZNF57	Q68EA5	OT	DIABETES MELLITUS, TRANSIENT NEONATAL, 1
ZNF81	P51508	NDD	MENTAL RETARDATION, X-LINKED 45; MRX45
ZFPM2	Q8WW38	NDD	46,XY SEX REVERSAL 9; SRXY9
ZNF141		OT	POLYDACTYLY, POSTAXIAL, TYPE A6; PAPA6
ZBTB16	Q05516	NDD	SKELETAL DEFECTS, GENITAL HYPOPLASIA, AND MENTAL RETARDATION
ZNF148	Q9UQR1	NDD	GLOBAL DEVELOPMENTAL DELAY, ABSENT OR HYPOPLASTIC CORPUS CALLOSUM, AND DYSMORPHIC FACIES; GDACCF
GFI1	Q99684	OT	NEUTROPENIA, NONIMMUNE CHRONIC IDIOPATHIC, OF ADULTS
ZIC1		NDD	CRANIOSYNOSTOSIS 6; CRS6
ZIC3	O60481	NDD	HETEROTAXY, VISCERAL, 1, X-LINKED; HTX1
EGR1	P18146	NDD	PONTOCEREBELLAR HYPOPLASIA, TYPE 7; PCH7

C_2H_2 family gene names	Protein Accession	Neurodevelopmental involvement	OMIM clinical synopses
ZBTB18	Q99592	NDD	MENTAL RETARDATION, AUTOSOMAL DOMINANT 22; MRD22
ZBTB20	Q9HC78	NDD	PRIMROSE SYNDROME; PRIMS
SCAPER	Q9BY12	NDD	INTELLECTUAL DEVELOPMENTAL DISORDER AND RETINITIS PIGMENTOSA; IDDRP
FEZF1	A0PJY2	NDD	HYPOGONADOTROPIC HYPOGONADISM 22 WITH OR WITHOUT ANOSMIA; HH22
ZNF335	Q9H4Z2	NDD	MICROCEPHALY 10, PRIMARY, AUTOSOMAL RECESSIVE; MCPH10
GZF1	Q9H116	OT	JOINT LAXITY, SHORT STATURE, AND MYOPIA; JLSM
OVOL2	Q9BRP0	OT	CORNEAL DYSTROPHY, POSTERIOR POLYMORPHOUS, 1; PPCD1
ZNF341	Q9BYN7	NDD	HYPER-IgE RECURRENT INFECTION SYNDROME 3, AUTOSOMAL RECESSIVE; HIES3
ZNF408	Q9H9D4	OT	EXUDATIVE VITREORETINOPATHY 6; EVR6
ZFHX2	Q9C0A1	NDD	MARSILI SYNDROME; MARSIS
ZNF423	Q2M1K9	ND	NEPHRONOPHTHISIS 14; NPHP14
ZBTB24	O43167	NDD	IMMUNODEFICIENCY-CENTROMERIC INSTABILITY-FACIAL ANOMALIES SYNDROME 2; ICF2
ZNF469		NDD	BRITTLE CORNEA SYNDROME 1; BCS1
ZNF513	Q8N8E2	OT	RETINITIS PIGMENTOSA 58; RP58
GLIS3	Q8NEA6	NDD	DIABETES MELLITUS, NEONATAL, WITH CONGENITAL HYPOTHYROIDISM; NDH
ZNF711	Q9Y462	NDD	MENTAL RETARDATION, X-LINKED 97; MRX97
SALL1	Q9NSC2	NDD	TOWNES-BROCKS SYNDROME 1; TBS1
SALL2	Q9Y467	OT	COLOBOMA, OCULAR, AUTOSOMAL RECESSIVE
SALL4	Q9UJQ4	NDD	IVIC SYNDROME
CHAMP1	Q96JM3	NDD	MENTAL RETARDATION, AUTOSOMAL DOMINANT 40; MRD40
BCL11A		NDD	INTELLECTUAL DEVELOPMENTAL DISORDER WITH PERSISTENCE OF FETAL HEMOGLOBIN
BCL11B	Q9C0K0	NDD	IMMUNODEFICIENCY 49; IMD49
PLAG1	Q6DJT9	OT	SALIVARY GLAND ADENOMA, PLEOMORPHIC
ZBTB42	B2RXF5	NDD	LETHAL CONGENITAL CONTRACTURE SYNDROME 6; LCCS6

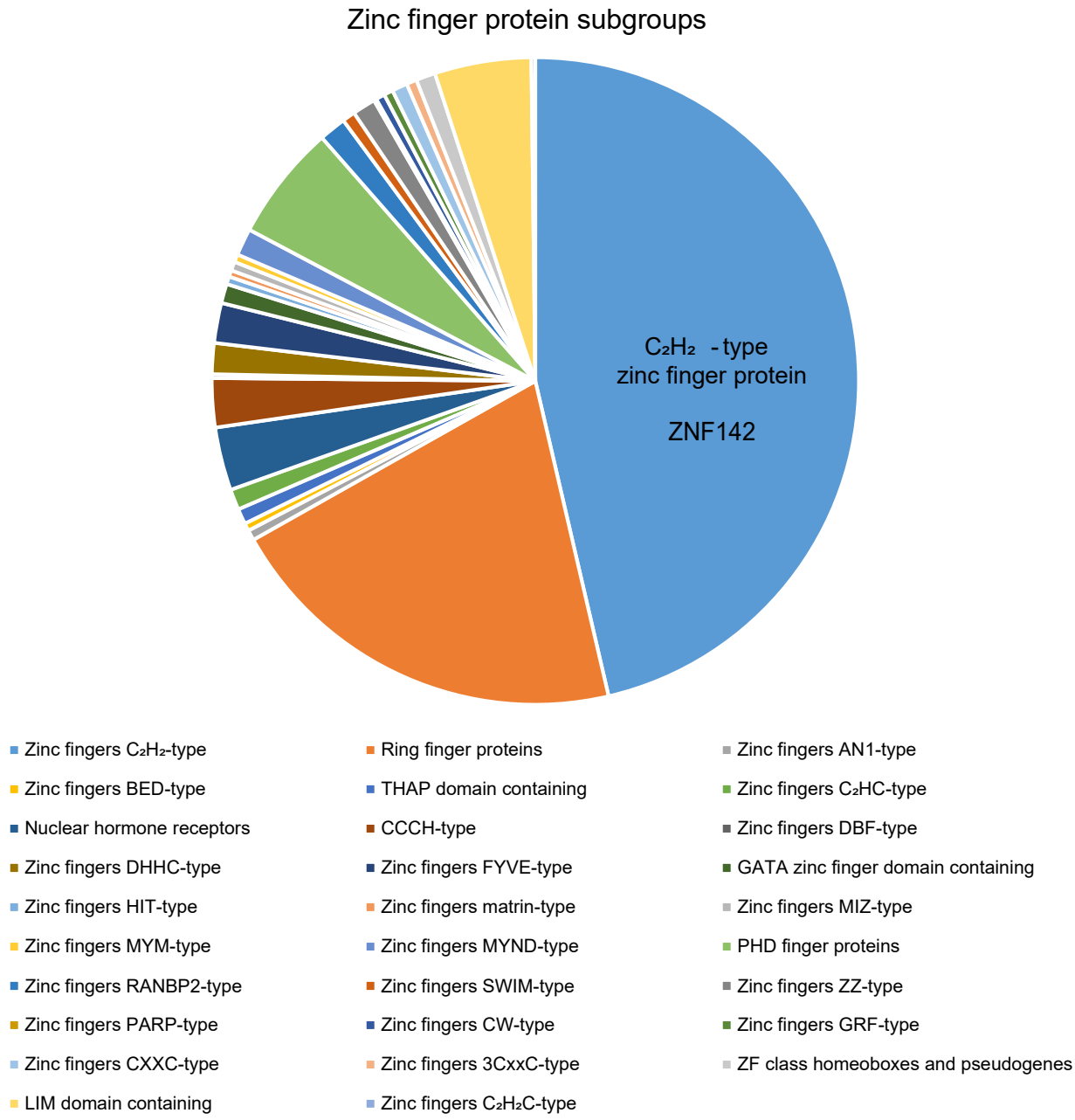

Zinc finger protein subgroups
C2H2 - type zinc finger protein
ZNF142
Zinc fingers C2H2-type
Ring finger proteins
Zinc fingers AN1-type
Zinc fingers BED-type
THAP domain containing
Zinc fingers C2HC-type
Nuclear hormone receptors
CCCH-type
Zinc fingers DBF-type
Zinc fingers DHHC-type
Zinc fingers FYVE-type
GATA zinc finger domain containing
Zinc fingers HIT-type
Zinc fingers matrin-type
Zinc fingers MIZ-type
Zinc fingers MYM-type
Zinc fingers MYND-type
PHD finger proteins
Zinc fingers RANBP2-type
Zinc fingers SWIM-type
Zinc fingers ZZ-type
Zinc fingers PARP-type
Zinc fingers CW-type
Zinc fingers GRF-type
Zinc fingers CXXC-type
Zinc fingers 3CxxC-type
ZF class homeoboxes and pseudogenes
LIM domain containing
Zinc fingers C2H2C-type

Figure S1

Fig. S1: Zinc finger protein (ZNF/zfps) subgroups in human. ZNF/zfps represent a large class of proteins with 1593 genes described in the HUGO Gene Nomenclature committee (HGNC) database (https://www.genenames.org/). The ZNF/zfp family subdivides into 29 subgroups based on folded protein structures. ZNF142 belongs to the largest subgroup, whose 720 members have a C_2H_2 domain. Pie chart shows the size comparison of the various ZNF/zfp subfamilies represented in the HGNC database, with ZNF142 denoted within the C_2H_2 subfamily.

Family C

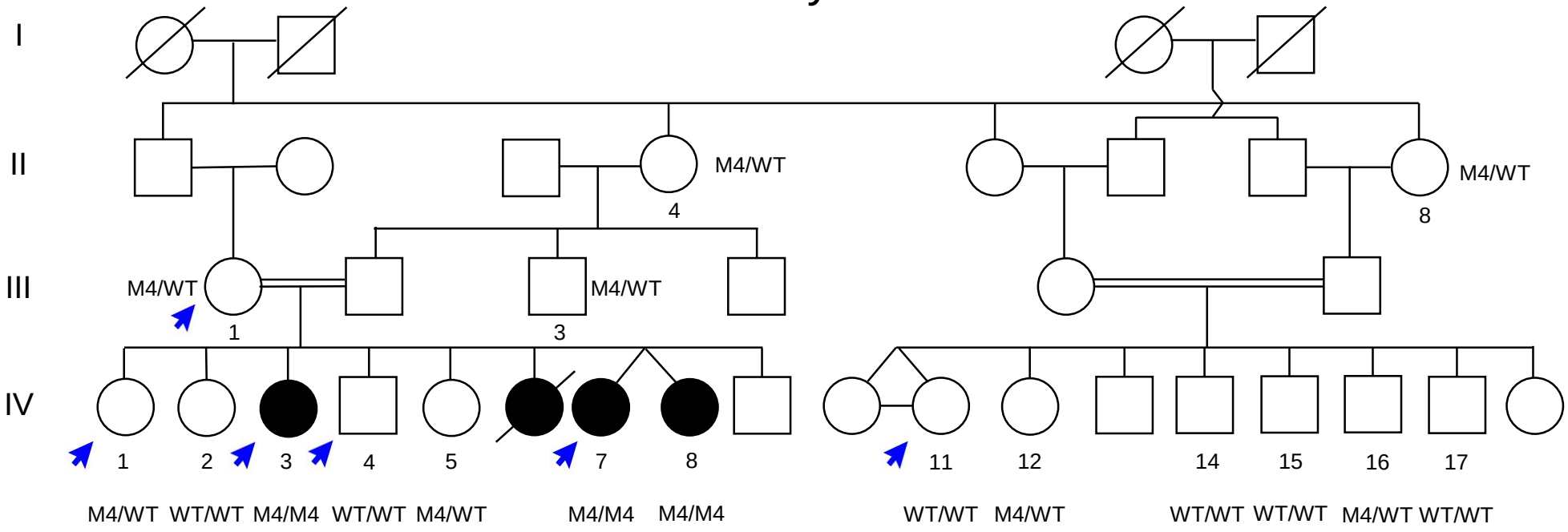
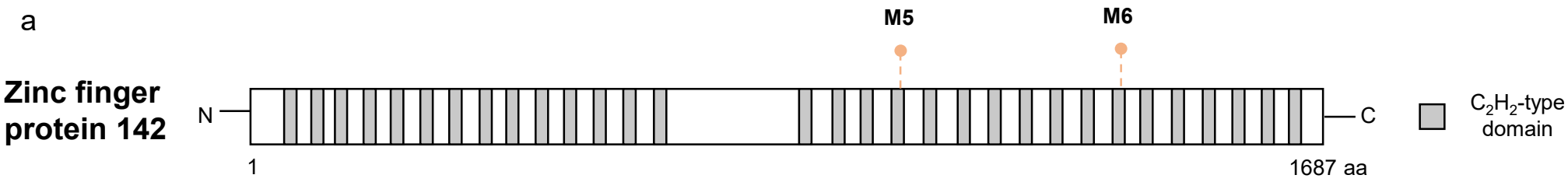


Figure S2

Fig. S2: Extended pedigree of Family C indicating a recessive pattern of inheritance. Family C has two consanguineous loops represented by double horizontal lines. The unaffected individuals are indicated by unfilled circles (females) and squares (males) whereas the affected individuals are represented by black filled circles. The blue arrows indicate individuals subjected to ES. Genotypes are labeled under each individual (WT/WT for wild type, M4/WT for heterozygous and M4/M4 for homozygous). M4; c. 4183delC+4185G>A; p. Leu1395*.

a


Zinc finger protein 142
M5
M6
N
C
1
1687 aa
C_2H_2-type domain

b

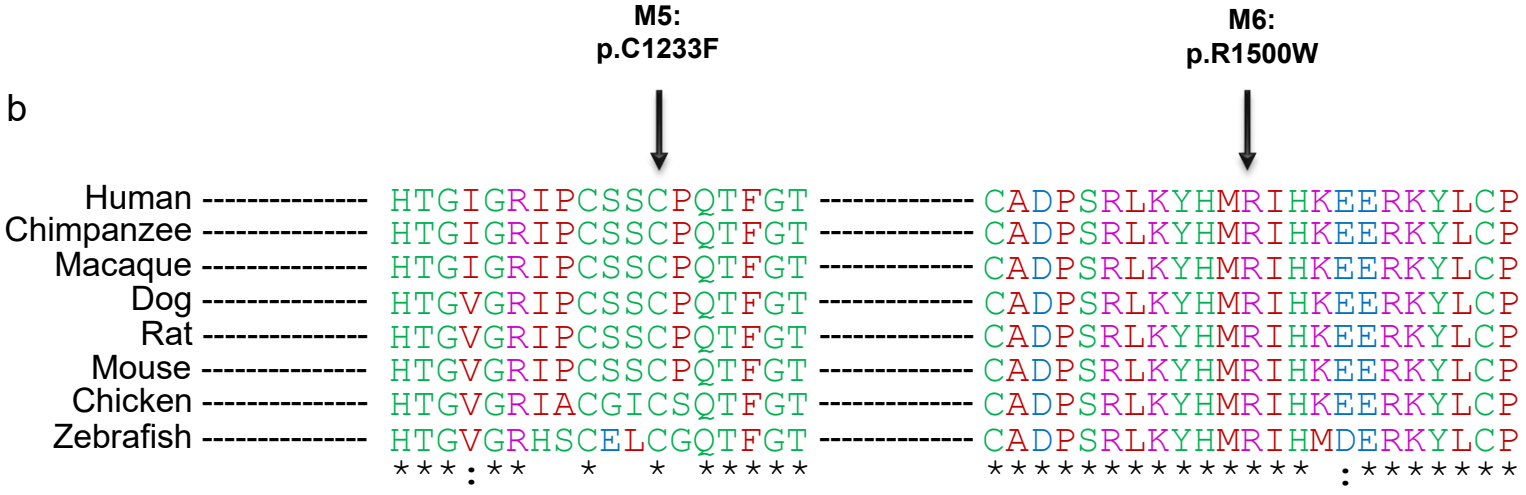

M5: p.C1233F
M6: p.R1500W
Human ---------- HTGIGRIPCSSCPQTFGT ---------- CADPSRLKYHMRIHKEERKYLCP
Chimpanzee ---------- HTGIGRIPCSSCPQTFGT ---------- CADPSRLKYHMRIHKEERKYLCP
Macaque ---------- HTGIGRIPCSSCPQTFGT ---------- CADPSRLKYHMRIHKEERKYLCP
Dog ---------- HTGVGRIPCSSCPQTFGT ---------- CADPSRLKYHMRIHKEERKYLCP
Rat ---------- HTGVGRIPCSSCPQTFGT ---------- CADPSRLKYHMRIHKEERKYLCP
Mouse ---------- HTGVGRIPCSSCPQTFGT ---------- CADPSRLKYHMRIHKEERKYLCP
Chicken ---------- HTGVGRIACGICSQTFGT ---------- CADPSRLKYHMRIHKEERKYLCP
Zebrafish ---------- HTGVGRHSCELCGQTFGT ---------- CADPSRLKYHMRIHMDERKYLCP

Figure S3

Fig. S3: Missense *ZNF142* variants in family D are highly conserved.

(a) Schematic representation of ZNF142 protein; gray colored rectangles represent predicted C_2H_2-type domains. Missense variants are indicated with salmon-colored lollipops.

(b) Multiple sequence alignment of regions surrounding the two missense variants in ZNF142 generated with Clustal Omega. Amino acid color scheme: Red- hydrophobic or aromatic (AVFPMILWY); blue- acidic (DE); magenta- basic (RHK); green- hydroxyl, amine, basic and glutamine (STYHCNGQ). Consensus symbols, an asterisk (*) indicates positions which have a single, fully conserved residue; a colon (:) indicates conservation between groups of strongly similar properties scoring > 0.5 in the Gonnet PAM 250 matrix.

Supplementary Methods

ES and variant annotation

Families A and B. We performed quartet-based ES (Family A, Fig. 1a); or ES of the simplex case (Family B, Fig. 1b). We constructed exome-enriched indexed libraries with the SureSelect All Exon system (Agilent Technologies) per manufacturer`s recommendations. Paired-end (101 base pair [bp]) sequencing was carried out on either a HiSeq4000 (family A) or HiSeq2500 (family B) platform (Illumina). We processed raw data with an in-house bioinformatics pipeline with established methods[1,2]. In brief, read-mapping to the reference genome (hg19) was done with the Burrows-Wheeler Aligner (BWA) tool[3] or GATK. Single-nucleotide variants and small insertions and deletions were called with SAMtools and PINDEL and we mandated that variants satisfied the criteria of a Phred-scaled genotype quality ≥30 and a read depth ≥10. ExomeDepth was used to query exomes for copy-number variants. Aligned sequencing reads were inspected with the Integrative Genomics Viewer (IGV, Broad Institute).

Family C. We performed ES on DNA obtained from peripheral blood of two affected females (C-IV-3 and C-IV-7), their healthy mother (C-III-1), two healthy siblings (C-IV-1, C-IV-4), and a healthy cousin (C-IV-11; Fig. 1c; Fig. S2). We performed library capture with the Nimblegen Exome capture kit and generated 150 bp paired-end reads on an Illumina HiSeq 4000 instrument as described.[4] Sequencing reads were processed initially with the TrimGalore toolkit, to remove any Illumina adapter sequences or low-quality base calls from the 3' end of the reads. Reads were then aligned to the human genome (hg19) with BWA[3]. Alignment processing and variant calls were performed using GATK[5,6], following the Broad Institute's Best Practices workflow[7]. Variants were then annotated with the Variant Effect Predictor[8] toolkit to identify the potential functional impact of each variant according to the human transcriptome (GRCh37r75)[8]. We confirmed filtered variants visually using IGV.

Family D. Exome sequencing was performed on three individuals from Family D (D-I-1, D-I-2, D-II-2; Fig. 1d). Genomic DNA was sonicated to approximately 200 bp fragments and adaptor-ligated to make a library for paired-end sequencing. Following amplification and barcoding, the libraries were hybridized to biotinylated complementary RNA oligonucleotide baits from the SureSelect XT Human All Exon +UTR v5 75Mb Kit (Agilent Technologies, Santa Clara, CA) and purified using streptavidin-bound magnetic beads. Amplification was performed prior to sequencing on the Illumina HiSeq 2000 system (San Diego, CA). Exome reads were aligned with Novoalign (v3.02.00) and genome reads with bwa mem (v0.7.15) to the human reference genome assembly (hg19 dbSNP132-masked, UCSC Genome Browser). PCR duplicates were removed using Mark Duplicates from Picard (http://picard.sourceforge.net). SamTools and PicardTools were used to further process the SAM/BAM files. Variant calling was performed with GATK Unified Genotyper and variant annotation using ANNOVAR.

Variant filtering

We filtered data to retain functional variants exclusive to affected individuals (predicted to alter mRNA splicing or amino acid sequence) with a minor allele frequency (MAF) of ≤1% in public SNP databases (dbSNP142, 1000 Genomes, Exome Aggregation Consortium [ExAC], Genome aggregation database [gnomAD]) that segregated with disease in the pedigree. We also considered phenotype assessment, pathogenic variant database searches (ClinVar, Human Gene Mutation Database [HGMD], and the DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources [DECIPHER]), and American College of Medical Genetics categorization[9] as described[1]. We used SAMtools varFilter, custom scripts and our in-house database of controls (Munich, n=10,000 exomes; families A and B); Enlis Genome Research software (Enlis Genomics; family C); and SamTools and PicardTools (family D).

References

1. Zech M, Boesch S, Jochim A, et al. Clinical exome sequencing in early-onset generalized dystonia and large-scale resequencing follow-up. *Mov Disord.* 2017;32(4):549-559.
2. Zech M, Boesch S, Maier EM, et al. Haploinsufficiency of KMT2B, Encoding the Lysine-Specific Histone Methyltransferase 2B, Results in Early-Onset Generalized Dystonia. *Am J Hum Genet.* 2016;99(6):1377-1387.
3. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.
4. Villanueva A, Willer JR, Bryois J, Dermitzakis ET, Katsanis N, Davis EE. Whole exome sequencing of a dominant retinitis pigmentosa family identifies a novel deletion in PRPF31. *Invest Ophthalmol Vis Sci.* 2014;55(4):2121-2129.
5. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303.
6. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491-498.
7. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013;43:11 10 11-33.
8. Cunningham F, Amode MR, Barrell D, et al. Ensembl 2015. *Nucleic Acids Res.* 2015;43(Database issue):D662-669.
9. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-424.