**Genetic Complexity of Diagnostically Unresolved Ehlers-Danlos Syndrome**


**Supplementary Methods**

**Whole exome sequencing and variant analysis**

Genomic DNA from 89 individuals was processed using the SureSelectXT2 Human All Exon V5 capture kit (Agilent) and sequenced with 75 base paired-end reads on a HiSeq 4000 (Illumina) and from 85 samples with 100 base paired-end reads on a HiSeq 2500 (Illumina). Raw sequence data will be made available on reasonable request to the study’s data access committee, chaired by TJA.

FASTQs were aligned to the human genome reference (GRCh37) using bwa mem (0.7.12). Alignments were post-processed using Picard (v2.1.1) for identification of duplicate reads and the Genome Analysis ToolKit (GATK, 3.5-0-g36282e4) for indel realignment and base recalibration. Genotype likelihoods for each sample were calculated using the GATK HaplotypeCaller and resulting GVCF files were called jointly using GATK’s GenotypeGVCFs function. Functional annotations were added using Ensembl’s Variant Effect Predictor (v90). VASE (v0.1, https://github.com/david-a-parry/vase) was used to perform dominant and recessive segregation filtering of variants. Variants with a frequency greater than 0.1% (for dominant filtering) or 0.5% (for recessive filtering) in gnomAD or dbSNP150 or those not annotated as either high or moderate impact variants or as splice region variants were removed. Splice region variants not overlapping the canonical +/-2 donor/acceptor intron positions were only retained if they had an ada score and rf score from dbscSNV (https://doi.org/10.1093/nar/gku1206) of 0.8 or higher. Genotype calls were filtered if PHRED scale genotype quality scores were below 20, based on fewer than 5 reads or if the ratio of variant reads compared to total depth was below 0.25.
Additionally, variants were processed using the G2P plugin for VEP (https://www.ebi.ac.uk/gene2phenotype/g2p_vep_plugin) and the Genomics England Panel App (Ehlers-Danlos Syndrome(https://panelapp.genomicsengland.co.uk/api/v1/panels/53/?version=2.0).

A further ‘exomiser’ based analysis using all the HPO terms currently identified as clinical criteria in the 2017 EDS nosology. Variants were reviewed for known EDS genes, mendelian disorders with EDS features or symptoms, HTAD, genes abnormally expressed in skin fibroblast from patients with vEDS, cEDS and hEDS. Variant calls were searched for genes associated with the previously linked region for hEDS reported by Syx et al., pelvic organ prolapse, genome wide association studies for GIH, knee pain, rotator cuff injury and pelvic organ prolapse (https://www.ebi.ac.uk/gwas/) .

Database searches and variant assessment

**Mendelian Disorders:** Dominant and autosomal recessive variant datasets were searched using OMIM annotations. Variants with CADD score > 15 were selected for further review to assess for the updated ACMG criteria for pathogenicity using the annotation tool Varsome (https://varsome.com/) and Franklin by Genoox (https://franklin.genoox.com). This included ClinVar reports, functional annotation, previous published reports of specific variants, occurrence of the variant in a specific protein domain and reported allele frequency (https://gnomad.broadinstitute.org/).

A specific search for variants in EDS genes from the 2017 nosology was completed: classical EDS (cEDS): COL5A1, COL5A2, COL1A1, classical like EDS (clEDS): TNXB, cardiac valvular EDS (cvEDS): COL1A2, vascular EDS (vEDS): COL3A1, COL1A1, dermatosparaxis EDS (dEDS): ADAMTS2, kyphoscoliotic EDS (kEDS): PLOD1, FKBP14, Brittle Cornea Syndrome (BCS): PRDM5, ZNF469, spondyloplastic EDS (spEDS): B4GALT7, B3GALT6, SLC39A13, Musculocontractural EDS (mcEDS): CHST14, DSE, myopathic EDS (mEDS): COL12A1, periodontal EDS (pEDS): C1R, C1S.

Further searches were completed for rare variants in disorders associated with EDS like phenotypes: including Ehlers-Danlos syndrome classic-like-2: AEBP1, Bethlem myopathy: COL6A1, COL6A2, COL6A3 and Zimmerman-Laband Syndrome: KCNH1, ATP6V1B2, KCCN3.

We searched for rare variants in Mendelian disorders associated with EDS symptomatology, including dysautonomia: SPTLC1, WNK1 and IBKAP, familial mast cell disorders, TPSAB1, KIT and erythermalgia SCN9A.
We searched for rare variants in Mendelian disorders with multisystem manifestations which are rarely associated with aneurysm: Neurofibromatosis type I (MIM 613113) NF1, Tuberous Sclerosis (MIM 191100) TSC1, TSC2, Birt-Hogg-Dube syndrome (MIM 135150) FLCN and Singleton Merten Syndrome (MIM 182250) IFIH1, DDX58.

We completed a review of rare variants in genes causative for Inborn errors of metabolism with features of hereditary disorders of connective tissue, these may be underdiagnosed: homocystinuria: CBS, Wilson disease: ATP7B, Occipital horn syndrome/ Menke’s disease: ATP7A and hypophosphatasia: ALPL.

We searched for HTAD genes using the ClinGen criteria \(^2\) (https://clinicalgenome.org/docs/clinical-validity-of-genes-for-heritable-thoracic-aortic-aneurysm-and-dissection/) for genes strongly associated with HTAD: ACTA2, COL3A1, FBN1, MYH11, MYLK, SMAD3, TGFBI, TGFBR1, and TGFBR2. Potentially diagnostic: EFEMP2, ELN, FBN2, FLNA, NOTCH1, SLC2A10, SMAD4, and SKI. Gene with limited evidence of causality: COL4A5, CBS, PKD1, and PKD2, genes with no evidence/ experimental data only for causality: ACVR1A, ADAMTS10, B3GAT3, COL1A1, COL1A2, COL4A1, COL5A1, COL5A2, COL9A1, COL9A2, COL11A1, COL18A1, EMILIN1, ENG, GATA5, GJA1, JAG1, MED12, PLOD1, PLOD3, SMAD6, UFP3B, and VCAN. Newly identified genes: BGN, FOXE3, HCN4, MAT2A, MFAP5, SMAD2, and TGFBI.

**Mendelian Disorders awaiting confirmation:** We searched for rare variants in Mendelian entities with EDS like features, awaiting confirmation with autosomal recessive inheritance: PLOD3, ALDH18A1, ATP6V0D2, ATP6V1E1, CAPN3, GORAB, OBSL1, IFT122, PLP1, SPARC and EFEMP15. Similarly, we searched for Mendelian entities with EDS-like features awaiting confirmation: autosomal dominant connective tissue disorder with peripheral neuropathy: EMILIN1, cardiochondroplasty condition: LAMA5, nemaline myopathy: RYR3.

We searched for rare variants in genes reported in association with risk of intracranial aneurysm \(^1\) (family studies reviewed in PMID: 32367296): ADAMTS15, ANGPTL6, ARGHGEF16, LOXL2, PCNT, RNF213, THSD1, TMEM132B, NEK4, EDIL3, EDNRB, DNAH9 and GGA3.

**Genes reported as abnormally expressed in EDS linkage studies:** We searched for rare variants in genes within the linked region for hEDS \(^6\): BMP1, CNOT7, CSGALNACT1, LOXL2, LPL, SLC39A14, HR, NPM2, DOCKS, ADAMDEC1, ADAM7, GNRH1, STC1, ADAM28, FGF17, SORBS3, NKK3-1, SFTPC, NEFL, FGF20, ADAM28, FGL1, ASAHI1, PDLIM2, CCAR2 LZTS1 NKX2-6, NAT1, DOK2, TNFRSF1B DMTN, EGF17, KCTD9, NPM2, PDLIM2, ENTPP4, SLC18A1, SFTPC, ATP6V1B2, PDGFRL, PCMI, PFLIM2, TNFRSF1D, GFRA2, NEFM, SLC7A1, BIN3, POLR3D, VSP37A, C8orf20.
Genes reported as abnormally expressed in skin fibroblast studies: We searched for rare variants (germline) with CADD>15 in genes abnormally expressed in skin fibroblasts from cEDS patients 4: SPP1, POSTN, EDIL3, PAPPA, IGFBP2, C3, DNAJB7, CCGP1, ATG10, SVIP, ALG13, VIPAS39, HIF4A, CDKN1A, CCNE2, ASF1B, CLSNP, DTL, DDIAS.

We searched for rare variants (germline) with CADD>15 in genes abnormally expressed in skin fibroblasts from vEDS patients with confirmed COL3A1 mutations 5: FBN2, TNFAIP6, PTC2, HIST1H4L, ITGA3, HSPG2, MMP24, EDNRA, LOXL3, P4HA2, P4HA3.

We searched for rare variants (germline) with CADD>15 in genes abnormally expressed in skin fibroblasts from hEDS patients 3: CDH11, MMP9, CCN1, CCN2, ITGB3, ILK, PINCH, PARVA, PARVB, PARVG, PXN, AKT1, AKT2, AKT3, GSK3β, NFkB1, CDH1, MMP 2, SNAI1, SNAI2.

Genes reported as associated with features of EDS in GWAS: We reviewed our data for rare variants (MAF<0.1% and CADD>15) in GWAS Loci for one of the diagnostic criteria for hEDS: self-reported Beighton score >S with P < 5 x 10⁻⁸: STON1, (MIM 605357), EFEMP1 (MIM 601548, Doyne honeycomb degeneration of retina #126600), C2orf54 (Not annotated), ABI3BP (MIM 606279), VCAN (MIM 118661, Wagner syndrome #143200), NOTCH4 (MIM 164951), XKR6 (Not annotated), NEDD4 (MIM 602278), PIEZO1 (MIM 611184, Dehydrated hereditary stomatocytosis with or without pseudohypermalemia and/or perinatal edema #3194380, Lymphoedema (AR, LoF).

We reviewed our data for rare variants (MAF<0.1% and CADD>15) in GWAS Loci for pelvic organ prolapse with P < 5 x 10⁻⁸ 9: WNT4, GDF7, EFEMP1, FAT4, IMPDH1, TBX5, SALL1.

We reviewed our data for rare variants (MAF<0.1% and CADD>15) in GWAS Loci for knee pain and rotator cuff injury associated loci (https://www.ebi.ac.uk/gwas/), with P < 5 x 10⁻⁸: COL27A1 (MIM 608461, Steel syndrome), GDF5 (MIM 601146, multiple phenotypes), DENND2C, SASH1, ESRRB, FGFR1, TNC and DEFB1.

Assessment of Candidate genes: We reviewed our data for rare variants expected to result in loss of function, identifying genes with OMIM annotation. For non-annotated genes we reviewed the probability of loss of function intolerance scores (pLi) and biological plausibility, looking for published evidence of expression or impact on the extracellular matrix, collagen synthesis or function, aneurysm formation in human tissue studies and reported EDS or HTAD like phenotypes in animal models https://www.alliancegenome.org/. Similarly, we reviewed our data for novel missense, splice and synonymous variants (gnomAD frequency = 0). Variants with high CADD scores (>20) were selected for further review as above. The entire dataset were reviewed for the same or further rare variants in the same gene.
Genetic burden analysis

Analysis of sequence data where there are systemic differences in coverage between cases and controls typically leads to inflated type I errors, but discarding those samples with insufficient read depth can result in a loss of power. TASER is a program for testing association using sequencing reads without calling genotypes, which is robust to a wide range of differential sequencing qualities between cases and controls. TASER uses the total number of reads mapped to a variant, and the number carrying the minor allele, to calculate a score statistic at each position in a gene of interest, thus providing an assessment of the association of each individual variant with the disease phenotype. A burden statistic is then calculated for each gene as the sum of the score statistics for each of the variants within that gene, allowing identification of genes that have a higher or lower accumulation of rare variants in the cases than might be expected, compared to controls. A bootstrap procedure is used for assessing the significance of the burden statistic. TASER includes a screening procedure to screen-in loci based on allele counts (not on assigned genotypes) where: 1) Alternate allele read count frequency (AACF) in the entire cohort < 0.05 (can be adjusted if required); 2) AACF is not less than 1/(2n) where n is the sample size of the overall cohort tested.

For each of the sequences, we split the DNA sequence into non-overlapping exons, where the gene was the unit of the burden test, in genomic order. Each chromosome was split into 100 gene “processing” blocks based on the GRCh37, resulting in the analysis of 16560 genes in 240 blocks. Only bases called with a quality score >30 were added to the read count at each position within each exon, and only if the resultant read depth was greater than 2. The upper MAF limit for analysis was set at 0.05 in the base population. The top scoring loci from this analysis are shown in Table 2. Since analysis of rare variant burden was performed in 16560 genes, a p value of 0.05/16560 = 3 x 10^-6 would be considered genome-wide evidence for statistical significance. Examination of QQ plots from the overall set of 16560 χ² test statistics derived from the bootstrap p values showed a slight inflation (genomic control inflation factor λ=1.11) so we adjusted the p values by dividing the χ² test statistics by 1.11 and recalculating the implied p values.
References


