

Genetic Complexity of Diagnostically Unresolved Ehlers-Danlos Syndrome

Anthony Vandersteen^{1,2}, Ruwan A. Weerakkody^{2,3}, David Parry², Christina Kanonidou⁴, Daniel J. Toddie-Moore², Jana Vandrovcova⁵, Rebecca Darlay⁶, Javier Santoyo-Lopez⁷, Alison Meynert⁸, NIHR BioResource⁹, Hanadi Kazkaz¹⁰, Rodney Grahame¹⁰, Carole Cummings¹¹, Marion Bartlett¹¹, Neeti Ghali¹¹, Angela F. Brady¹¹, F. Michael Pope¹¹, Fleur S. Van Dijk¹¹, Heather J. Cordell⁶, Timothy J. Aitman².

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 dbSNV (<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 ¹. Variants were reviewed for known EDS genes ¹, mendelian disorders with EDS features or symptoms, HTAD ², genes abnormally expressed in skin fibroblast from patients with vEDS, cEDS and hEDS ³⁻⁵. Variant calls were searched for genes associated with the previously linked region for hEDS reported by Syx et al ⁶, pelvic organ prolapse ⁷, genome wide association studies for GJH, knee pain, rotator cuff injury and pelvic organ prolapse (<https://www.ebi.ac.uk/gwas/>) ^{8,9}.

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 ¹⁰⁻¹³ using the annotation tool Varsome ¹⁴: (<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 ¹ 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*, spondylodysplastic 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 erythralgia *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*, *TGFB2*, *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: *ACVRL1*, *ADAMTS10*, *B3GAT3*, *COL1A1*, *COL1A2*, *COL4A1*, *COL5A1*, *COL5A2*, *COL9A1*, *COL9A2*, *COL11A1*, *COL18A1*, *EMILIN1*, *ENG*, *GATA5*, *GJA1*, *JAG1*, *MED12*, *PLOD1*, *PLOD3*, *SMAD6*, *UPF3B*, and *VCAN*. Newly identified genes: *BGN*, *FOXE3*, *HCN4*, *MAT2A*, *MFAP5*, *SMAD2*, and *TGFB3*.

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 *EFEMP* ^{15,16}.

Similarly, we searched for Mendelian entities with EDS-like features awaiting confirmation: autosomal dominant connective tissue disorder with peripheral neuropathy: *EMILIN1*, cardio-spondylocarpofacial syndrome: *MAP3K7*, multisystem connective tissue disorder: *LAMA5*, nemaline myopathy *RYR3*.

We searched for rare variants in genes reported in association with risk of intracranial aneurysm ¹⁷ (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*, *DOCK5*, *ADAMDEC1*, *ADAM7*, *GNRH1*, *STC1*, *ADAM28*, *FGF17*, *SORBS3*, *NKX3-1*, *SFTPC*, *NEFL*, *FGF20*, *ADAM28*, *FGL1*, *ASAH1* *PDLIM2*, *CCAR2* *LZTS1* *NKX2-6*, *NAT1*, *DOK2*, *TNFRSF10B* *DMTN*, *EGF17*, *KTCD9*, *NPM2*, *PDLIM2*, *ENTPP4*, *SLC18A1*, *SFTPC*, *ATP6V1B2*, *PDGFRL*, *PCM1*, *PFLIM2*, *TNFRSF10D*, *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*, *CCPG1*, *ATG10*, *SVIP*, *ALG13*, *VIPAS39*, *HIF4A*, *CDKN1A*, *CCNE2*, *ASF1B*, *CLSPN*, *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*, *PTCH2*, *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*, *GSK3B*, *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 >5 with $P < 5 \times 10^{-8}$: *STON1* (MIM 605357), *EFEMP1* (MIM 601548, Doyme 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 pseudohyperkalemia 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 $P < 5 \times 10^{-8}$ 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 $P < 5 \times 10^{-8}$: *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 tested¹⁸.

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 $0.05/16560 = 3 \times 10^{-6}$ would be considered genome-wide evidence for statistical significance. Examination of QQ plots from the overall set of 16560 χ^2 test statistics derived from the bootstrap p values showed a slight inflation (genomic control inflation factor $\lambda=1.11$) so we adjusted the p values by dividing the χ^2 test statistics by 1.11 and recalculating the implied p values.

References

1. Malfait F, Francomano C, Byers P, et al. The 2017 international classification of the Ehlers-Danlos syndromes. *Am J Med Genet C Semin Med Genet* 2017;175(1):8-26. doi: 10.1002/ajmg.c.31552
2. Renard M, Francis C, Ghosh R, et al. Clinical Validity of Genes for Heritable Thoracic Aortic Aneurysm and Dissection. *J Am Coll Cardiol* 2018;72(6):605-15. doi: 10.1016/j.jacc.2018.04.089
3. Zoppi N, Chiarelli N, Binetti S, et al. Dermal fibroblast-to-myofibroblast transition sustained by α v β 3 integrin-ILK-Snail1/Slug signaling is a common feature for hypermobile Ehlers-Danlos syndrome and hypermobility spectrum disorders. *Biochim Biophys Acta Mol Basis Dis* 2018;1864(4 Pt A):1010-23. doi: 10.1016/j.bbadis.2018.01.005 [published Online First: 20180105]
4. Chiarelli N, Carini G, Zoppi N, et al. Molecular insights in the pathogenesis of classical Ehlers-Danlos syndrome from transcriptome-wide expression profiling of patients' skin fibroblasts. *PLoS One* 2019;14(2):e0211647. doi: 10.1371/journal.pone.0211647 [published Online First: 20190204]
5. Chiarelli N, Carini G, Zoppi N, et al. Transcriptome analysis of skin fibroblasts with dominant negative COL3A1 mutations provides molecular insights into the etiopathology of vascular Ehlers-Danlos syndrome. *PLoS One* 2018;13(1):e0191220. doi: 10.1371/journal.pone.0191220 [published Online First: 20180118]
6. Syx D, Symoens S, Steyaert W, et al. Ehlers-Danlos Syndrome, Hypermobility Type, Is Linked to Chromosome 8p22-8p21.1 in an Extended Belgian Family. *Dis Markers* 2015;2015:828970. doi: 10.1155/2015/828970 [published Online First: 20151004]
7. Nikolova G, Lee H, Berkovitz S, et al. Sequence variant in the laminin gamma1 (LAMC1) gene associated with familial pelvic organ prolapse. *Hum Genet* 2007;120(6):847-56. doi: 10.1007/s00439-006-0267-1 [published Online First: 20061005]
8. Pickrell JK, Berisa T, Liu JZ, et al. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* 2016;48(7):709-17. doi: 10.1038/ng.3570 [published Online First: 20160516]
9. Olafsdottir T, Thorleifsson G, Sulem P, et al. Genome-wide association identifies seven loci for pelvic organ prolapse in Iceland and the UK Biobank. *Commun Biol* 2020;3(1):129. doi: 10.1038/s42003-020-0857-9 [published Online First: 20200317]
10. 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):405-24. doi: 10.1038/gim.2015.30 [published Online First: 20150305]
11. Abou Tayoun AN, Pesaran T, DiStefano MT, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat* 2018;39(11):1517-24. doi: 10.1002/humu.23626 [published Online First: 20180907]
12. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med* 2019;12(1):3. doi: 10.1186/s13073-019-0690-2 [published Online First: 20191231]
13. Pejaver V, Byrne AB, Feng BJ, et al. Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. *Am J Hum Genet* 2022;109(12):2163-77. doi: 10.1016/j.ajhg.2022.10.013 [published Online First: 20221121]
14. Kopanos C, Tsiolkas V, Kouris A, et al. VarSome: the human genomic variant search engine. *Bioinformatics* 2019;35(11):1978-80. doi: 10.1093/bioinformatics/bty897
15. Alazami AM, Al-Qattan SM, Faqeih E, et al. Expanding the clinical and genetic heterogeneity of hereditary disorders of connective tissue. *Hum Genet* 2016;135(5):525-40. doi: 10.1007/s00439-016-1660-z [published Online First: 20160329]

16. Driver SGW, Jackson MR, Richter K, et al. Biallelic variants in EFEMP1 in a man with a pronounced connective tissue phenotype. *Eur J Hum Genet* 2020;28(4):445-52. doi: 10.1038/s41431-019-0546-7 [published Online First: 20191202]
17. Sauvigny T, Alawi M, Krause L, et al. Exome sequencing in 38 patients with intracranial aneurysms and subarachnoid hemorrhage. *J Neurol* 2020;267(9):2533-45. doi: 10.1007/s00415-020-09865-6 [published Online First: 20200504]
18. Hu YJ, Liao P, Johnston HR, et al. Testing Rare-Variant Association without Calling Genotypes Allows for Systematic Differences in Sequencing between Cases and Controls. *PLoS Genet* 2016;12(5):e1006040. doi: 10.1371/journal.pgen.1006040 [published Online First: 20160506]