

## Supplementary Material

### Ultra-Rare Genetic Variation in Relapsing Polychondritis: A Whole-Exome Sequencing Study

Yiming Luo<sup>1,2</sup>, Marcela A. Ferrada<sup>2</sup>, Keith A. Sikora<sup>3</sup>, Cameron Rankin<sup>2</sup>, Hugh Alessi<sup>2</sup>, Daniel L. Kastner<sup>4</sup>, Zuoming Deng<sup>5</sup>, Mengqi Zhang<sup>6</sup>, Peter A. Merkel<sup>7</sup>, Virginia B. Kraus<sup>8</sup>, Andrew S. Allen<sup>9</sup>, Peter C. Grayson<sup>2</sup>

1. Division of Rheumatology, Department of Medicine, Columbia University Irving Medical Center, New York, NY
2. Vasculitis Translational Research Program, Systemic Autoimmunity Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD
3. Pediatric Translational Research Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD
4. Metabolic, Cardiovascular and Inflammatory Disease Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD
5. Biodata Mining and Discovery Section, Office of Science and Technology, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD
6. Biostatistics and Research Decision Sciences, Merck & Co., Inc., Rahway, NJ 07065, USA
7. Division of Rheumatology, Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA
8. Division of Rheumatology, Department of Medicine, Duke University School of Medicine, Durham, NC
9. Division of Integrative Genomics, Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC

## Supplemental Methods

### Clinical Assessment

All patients underwent a standardized clinical and laboratory assessment, including a detailed medical history, physical exam, Physician Global Assessment (PhGA), audiology, otolaryngologist evaluation, dynamic computed tomography (CT) of the chest, and pulmonary function test (PFT). Patients who were treated with TNF inhibitors (TNFi) (adalimumab, etanercept, infliximab, golimumab and certolizumab) were assessed by the treating physician for therapeutic response based on clinical improvement of inflammatory manifestations. Patients who had improvement in their inflammatory manifestations at the next clinical visit after starting TNFi without the need to further increase steroid therapy were considered responders to TNFi. Patients were followed from January 2017 to June 2022.

### Bioinformatic Processing

The raw sequencing data (fastq format) from RP and controls were processed by the same bioinformatic pipeline following the best practice guideline by Genome Analysis Toolkit (GATK) (Broad Institute, Massachusetts, U.S.). The variants were annotated with ANNOVAR<sup>1</sup> and Ensembl Variant Effect Predictor (VEP)<sup>2</sup>.

Additional quality control steps were performed with variants included only if they met the following criteria: (1) at least 10-fold coverage (2) QUAL at least 50 (3) GQ at least 20 (4) QD at least 2 (5) MQ at least 40 (6) RPRS greater than -3 (7) MQRS greater than -10 (8) FS  $\leq$  200 for indels and  $\leq$  60 for SNVs (9) VQSR tranche of 99% (10) Hardy–Weinberg equilibrium p value  $\geq$  0.001 (11) het alt allele ratio 0.3 to 0.7, ref-ref alt allele ratio  $<$  0.15, hom alt allele ratio  $>$  0.85 (12) genotype call rate  $\geq$  90% in cases and controls, respectively. Most of these criteria were based on a previously published exome-wide rare variant association study<sup>3</sup>.

The coverage of each Consensus Coding Sequence (CCDS) site was obtained for all samples by Samtools<sup>4</sup>. For each CCDS site, the absolute difference of the percentage of samples with coverage  $\geq$  10 between RP and controls was calculated. A cut-off of the absolute difference was selected when the following conditions were met: (1) [(absolute difference cut-off) + (percentage of sites pruned)] was minimized; (2) absolute difference cut-off  $\geq$  5%; (3) percentage of sites retained  $\geq$  90%.

After the above quality control steps, principal component analysis (PCA) was performed using Eigensoft<sup>5</sup> over a subset of uncorrelated ( $r^2 < 0.1$ ) polymorphic markers from the Illumina HumanCore chip. This marker selection was consistent with a previously published study<sup>6</sup>. PC outliers were removed if they were away from the center and exceeded 6 standard deviations (SD), which is the default setting for Eigensoft.

Identity-by-descent analysis was performed using KING<sup>7</sup> inside Plink 2<sup>8</sup> with the same set of polymorphic markers. One member of each pair was removed if the KING kinship coefficient was below 0.094.

### Quantile-quantile plot

Quantile-quantile plot (QQ-plot) was used to examine the distribution of observed p values against p values from a null distribution for each gene. The null distribution was obtained by permutation. In this process, the case and control labels were randomly permuted 1,000 times, and rank-order p values were obtained for each gene in each permutation set using Firth's logistic regression. The expected p value in the null model for each gene was the mean of each rank-ordered p values across the 1,000 permutations. We also obtained the genomic inflation factor Lambda. QQ-plot is a method to examine whether there is inflation of results from systematic bias, such as confounding due to population genetic structure.

### Variant-level association analyses

We used REGENIE v3.0.3 for variant-level association analyses via a two-step procedure as described by Mbatchou et al.<sup>9</sup>. Briefly, the first step fits a regression model for the outcome using the leave one chromosome out (LOCO) scheme, based on the subset of uncorrelated common genetic variants used in the PCA analysis and with minor allele count > 100. The LOCO phenotypic predictions were used as covariates in the second step to perform variant association analyses using an approximate Firth's logistic regression approach described by Mbatchou et al.<sup>9</sup>. Consistent with the gene-level analyses, the variant-level association models in both steps also included sex and first 3 genetic PCs as covariates. The calibration of variant-level association analyses was evaluated using permutation-based QQ plot, as described in the gene-level association analyses.

### Higher criticism test

Higher criticism (HC) test is a modification of the Kolmogorov-Smirnov (KS) test to improve statistical power for weak signal detection by scaling the KS test by

$Var_{H_0}(\widehat{F}(x)) = F_0(x)(1 - F_0(x))/n = x(1 - x)/n$  and restricting the domain over which the supremum is taken to the tail. These modifications of the KS test give rise to HC test as follows:

$$HC = \sup_{x < \alpha} \left\{ \frac{[\widehat{F}(x) - x]}{\sqrt{x(1 - x)/n}} \right\}.$$

To incorporate prior information into the HC framework by applying weight to each gene as follows:

$$HC^* = \sup_{x < \alpha} \left\{ \frac{[\widehat{F}(x^*) - F_0(x^*; w)]}{\sqrt{F_0(x^*; w)(1 - F_0(x^*; w))/n}} \right\},$$

For a given centrality measure, Let  $c_i$  be the centrality for the  $i$ th gene. We

take  $w_{ci} = \lambda(ac_i + bc)^{-1}$ , where  $c$  is the mean centrality across the gene set,  $a$  and  $b$  are 0.95 and 0.05 as used in the original literature, and  $\lambda$  is a scaling factor so that the mean of the weights is one.

HC test can be viewed as a goodness-of-fit test, and we tested whether the distribution of the p values associated with each gene (from the gene-level collapsing analysis) in the gene set was

consistent with the distribution of the test statistics under the global null. We obtained empirical null distributions of our statistics by randomly permuting case/control status 1,000 times. Because each gene may be included in multiple gene sets, leading to correlation between tests, we used the step-down minP algorithm to obtain family-wise error rate (FWER) adjusted p values across all the gene sets analyzed<sup>10</sup>.

In the HC test weighted by eigenvector centrality in the rare damaging model, for the “HALLMARK\_TNFA\_SIGNALING\_VIA\_NFKB” gene set, all the permuted HC statistics were smaller than the observed HC statistics. In this scenario, the p value output, including the multiplicity-adjusted p value, from the “wHC” package would be 0. We instead performed 3,000 permutations for the rare damaging model to obtain non-zero p values. The centrality calculation was based on gene interaction data from the bioGRID database (version 3.4.147)<sup>11</sup>. Protein interaction networks were generated from the bioGRID database for top ranked genes in significant pathways from weighted higher criticism test.

### DCBLD2 protein structure prediction

We predicted the structures of DCBLD2 protein, wild type and with the ultra-rare damaging variants, using AlphaFold2<sup>12</sup>. We accessed AlphaFold2 via ColabFold version 1.5.2, selecting the template mode with PDB100 and the default parameters<sup>13</sup>. The protein structure images were processed and visualized using PyMOL (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC).

### Sanger Sequencing

Custom primers were designed using Primer3, USCS Genome Browser, and NCBI BLAST for three regions of interest in the DCBLD2 gene: Q435fs-F 5'-GGAATAAACTGACATCCGAGCA-3' and Q435fs-R 5'-GCCCTGACCTCTATTGATTGA-3', I514F-F 5'-CTGCAACACAGTGGCCTTCTC-3' and I514F-R 5'-CTCCTCGAACAGCAATGACC-3', and G261V-F and L250X-F 5'-TTCTTAGCCCAGTGCTACCC-3' and G261V-R and L250X-R 5'-GAACTCCAGAGTAGTTAGAGGTGTG-3'. PCR was performed using a master mix of custom forward and reverse primers, OneTaq Hot Start 2x Master Mix with Standard Buffer from NEB, and water. Thermocycling conditions followed 95°C for 3 minutes followed by 34 cycles of 94°C for 30 secs, 60°C for 30 secs, and 68°C for 60 secs, ending with a final step of 68°C for 5 minutes. PCR products were purified using Qiagen QIAquick PCR Purification Kit. PCR products were checked for site specificity by gel electrophoresis, and sequenced by ACGT DNA Sequencing Services (Germantown, MD, USA) to confirm the variant calling validity of the DCBLD2 ultra-rare damaging QVs from WES.

The validation of Q435fs variant was performed after bacterial transformation. The PCR product was ligated into pGEM®-T Easy Vector (ProMega). Plasmids containing the PCR product were transformed into JM109 High Efficiency Competent E. coli using a heat-shock method (ProMega). Bacterial cells were plated on LB/IPTG/X-GAL/Amp plates and incubated overnight at 37°C. After overnight incubation, individual white colonies were transferred to bacterial culture tubes with LB/Amp liquid medium and incubated overnight at 37°C with shaking. Plasmids were then isolated from E. coli via ProMega miniprep kit specifications. After plasmid isolation, samples were sent to ACGT for sequencing using universal primers T7 and SP6.

### Power Calculation

Statistical power was calculated using simulation with different levels of odds ratio (OR) tested given our sample size of 66 cases and 2923 controls. The probability of an individual carrying a QV is  $p_1$  for RP and  $p_2$  for controls, both assumed to follow a binomial distribution. We assumed  $p_2=0.001$ .  $p_1$  will be calculated based on different OR tested. Firth's logistic regression was performed 5,000 times with each simulation for each tested OR. The proportion of p values below  $2.0 \times 10^{-6}$  and  $6.7 \times 10^{-7}$  is the exome-wide and study-wide statistical power for each tested OR, respectively.

The result of power calculation with different levels of OR was as following:

OR=50, power(exome)=0.36, power(study)=0.30  
OR=60, power(exome)=0.48, power(study)=0.42  
OR=70, power(exome)=0.59, power(study)=0.54  
OR=80, power(exome)=0.69, power(study)=0.65  
OR=90, power(exome)=0.78, power(study)=0.74  
OR=100, power(exome)=0.84, power(study)=0.80

### Supplemental Results

#### Variant-level association analyses

We performed variant-level association analyses on all the 529,675 common and rare variants which met our quality control criteria using an approximate Firth's logistic regression schemes implemented in REGENIE v3.0.3. Out of the analyses, 4,774 (0.9%) tests did not converge. Among these, 4,745 (99.3%) had only one individual and 23 had two individuals carrying the tested variant out of a total of 2,989 individuals. There were 6 tests that the tested variant was present in all the 2989 individuals, likely due to sequencing artifacts.

No variant reached exome-wide significance. There were also no additional compelling hypothesis-generating findings. The test with the smallest p value was a common non-frameshift deletion in the *PLIN4* gene, followed with three synonymous variants. Variants with p value < 0.01 are described in the **Supplemental Table 8**. The Manhattan plot and permutation-based QQ plot are shown in the **Supplemental Figure 7**. There was no evidence of systematic bias or poor calibration ( $\Lambda = 1.07$ ).

**Supplemental Table 1** Qualifying variants selection criteria for collapsing analysis

Model names	VAF in GnomAD	Predicted deleterious effects	Others
Ultra-rare model (primary)	< 0.1%	non-synonymous, CADD $\geq$ 20 or PTV	
Rare damaging model	< 0.1%	non-synonymous, CADD $\geq$ 20 or PTV	
Rare PTV model	< 0.1%	non-synonymous, PTV	
Recessive model	< 1%	non-synonymous, CADD $\geq$ 20 or PTV	at least biallelic
Synonymous model (negative control)	< 0.1%	synonymous	

VAF: variant allele frequency; GnomAD: the Genome Aggregation Database; CADD: Combined Annotation Dependent Depletion; PTV: protein truncating variant

**Supplemental Table 2** Percentage of exonic sites with well and balanced coverage in each gene (Table in a separate file)**Supplemental Table 3** DCBLD2 ultra-rare damaging QVs site mean coverage in GnomAD

Variant position	Mean exome sample coverage	Percentage of exome samples with at least 10x coverage	Mean genome sample coverage	Percentage of genome samples with at least 10x coverage
3:98531235-TG-T	71.09	99.83%	32.04	99.93%
3:98530074-T-A	62.54	99.07%	31.49	99.93%
3:98541120-C-A	52.13	99.84%	32.7	100%
3:98541153-A-T	54.06	99.91%	33.36	100%

**Supplemental Table 4** Results of gene-level collapsing analysis (Table in a separate file)

**Supplemental Table 5** All the DCBLD2 variants identified in this study  
(Table in a separate file)

**Supplemental Table 6** Clinical characteristics of 89 RP patients with and without *DCBLD2* ultra-rare damaging QVs

	RP with <i>DCBLD2</i> ultra-rare damaging QVs (n = 7)	RP without <i>DCBLD2</i> ultra-rare damaging QVs (n = 82)	P value
<b>Demographic Features</b>			
Female sex	74%	76%	1.00
Age at the onset of symptoms (years, mean+-SD)	23 +- 17	30 +- 16	0.24
Age at diagnosis (years, mean+-SD)	31 +- 20	37 +- 18	0.49
<b>Clinical Manifestations</b>			
Auricular chondritis	86%	61%	0.25
Nasal chondritis	57%	80%	0.16
Airway chondritis	43%	49%	1.00
Costochondritis	57%	74%	0.38
Ocular inflammation	43%	27%	0.40
Cutaneous inflammation	50%	23%	0.17
Arthritis or arthralgia	86%	87%	1.00
Oral ulcerations	57%	21%	0.05
Genital ulcerations	29%	15%	0.30
Venous thromboembolism <sup>a</sup>	14%	5%	0.34
Raynaud's phenomenon	29%	15%	0.30
Central nervous system involvement	0%	0%	1.00
Myocarditis	0%	0%	1.00
Pericarditis	14%	2%	0.22
Aortitis	14%	1%	0.15
Antinuclear antibody positive	14%	14%	1.00
Elevated c-reactive protein	43%	32%	0.55
Elevated erythrocyte sedimentation rate	14%	17%	0.86
MAGIC syndrome <sup>b</sup>	29%	13%	0.27
<b>Treatment</b>			
High-dose steroid <sup>c</sup>	57%	58%	1.00
Immunosuppressants	57%	83%	0.13
Biological or targeted immunosuppressants	43%	70%	0.21
<b>Complications</b>			
Death	0%	5%	1.00

ICU admission	14%	15%	1.00
Tracheotomy	14%	10%	0.54
Subglottic stenosis	14%	17%	1.00
Tracheomalacia	29%	41%	0.70
Bronchomalacia	29%	20%	0.63

- Venous thromboembolism includes unprovoked deep venous thrombosis and pulmonary embolism.
- Mouth and genital ulcers with inflamed cartilage (MAGIC) syndrome is defined as patients with RP who also met the International Criteria for Behcet's Disease (ICBD) criteria for Behcet's disease
- High-dose steroid means prednisone higher than 60 mg daily or equivalent.

**Supplemental Table 7** Clinical manifestations of patients with RP and DCBLD2 ultra-rare damaging QVs

P-1-1	P-1-2	P-2	P-3
57-year-old European American female with RP duration of 13 years. She had auricular and nasal chondritis, costochondritis, airway chondritis with bronchomalacia, oral ulcers, scleritis, arthralgia and pericarditis. She also had chronic thromboembolic pulmonary hypertension (CTEPH) and restrictive cardiomyopathy on supplemental oxygen at home. She failed methotrexate, azathioprine, cyclophosphamide and is currently on infliximab.	32-year-old European American female with RP duration of 3 years. She had auricular and nasal chondritis, costochondritis, arthralgia, photosensitive skin rash and hearing loss. She had a history of three spontaneous abortions and one episode of provoked deep venous thrombosis after surgery. She is currently on methotrexate and doing well.	53-year-old European American female with RP duration of 24 years. She had auricular chondritis, costochondritis, airway chondritis complicated by tracheo-bronchomalacia and subglottic stenosis requiring multiple tracheostomies, oral and genital ulcers, hearing loss, photosensitive skin rash, Raynaud phenomena and aortitis. She failed tocilizumab, tofacitinib and rituximab.	66-year-old European American male with RP duration of 19 years. He had auricular chondritis, costochondritis, airway chondritis, arthralgia, and inflammatory eye disease. He had severe mitral and tricuspid regurgitation and status post mitral valvuloplasty. He failed methotrexate and sulfasalazine. He is currently on adalimumab.
p-4-1	P-4-2	P-5	

<p>37-year-old European American female with RP duration of 7 years. She had auricular and nasal chondritis, costochondritis, malar rash, oral and genital ulcers and arthralgia. She was not on any immunosuppressive treatment.</p>	<p>9-year-old mixed ancestry (European and Native American) male with RP duration of 7 years. He had auricular chondritis, malar rash and arthralgia. He is on methotrexate and doing well.</p>	<p>10-year-old European American female with RP duration of 7 years. She had auricular chondritis, skin rash and arthralgia. She also had multiple episodes of severe croup and frontal bone lytic lesions. She is not on any immunosuppressive treatment.</p>	
---	---	--	--

**Supplemental Table 8** Variants with p values < 0.01 in the variant-level association analyses (Table in a separate file)

**Supplemental Table 9** Genes favorably weighted in the HALLMARK TNF pathway (Table in a separate file)

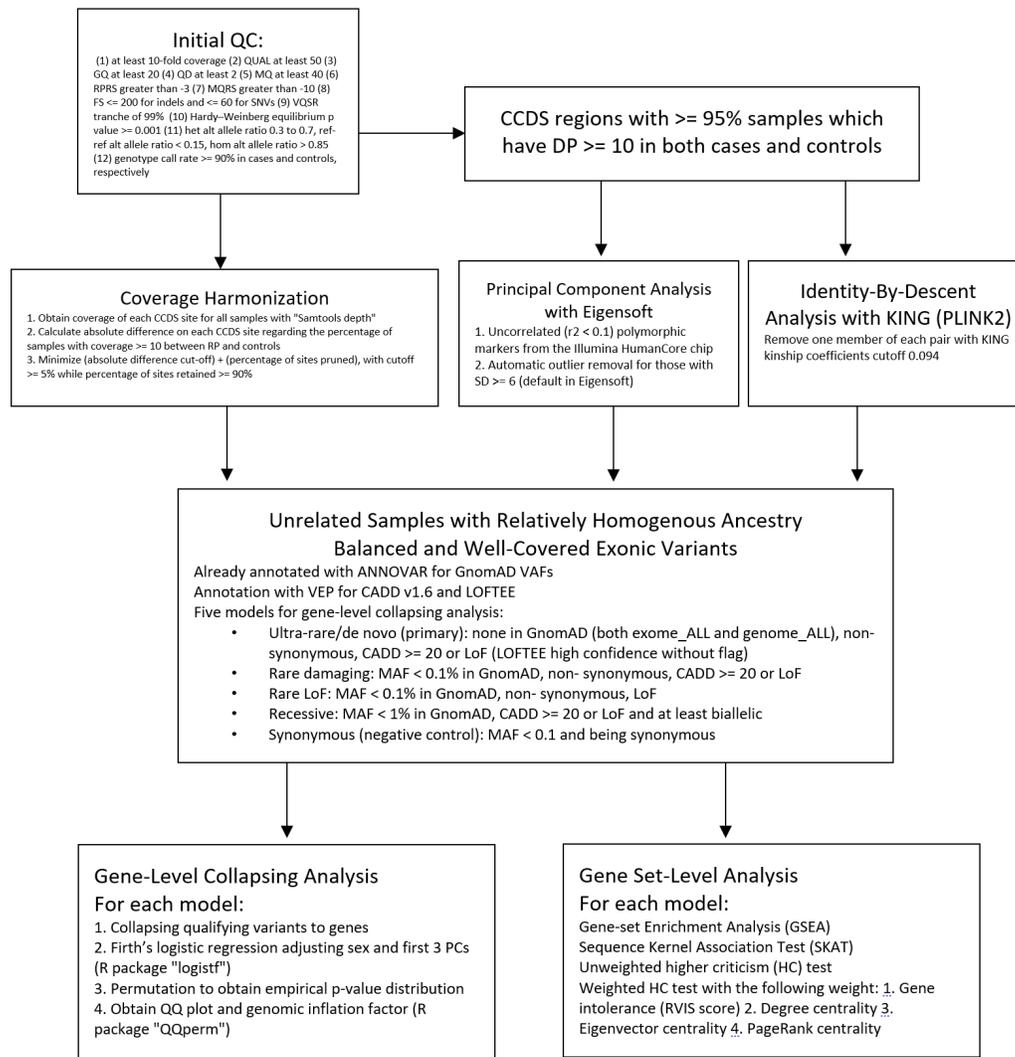
**Supplemental Table 10** Pathway analysis using higher criticism test (Table in a separate file)

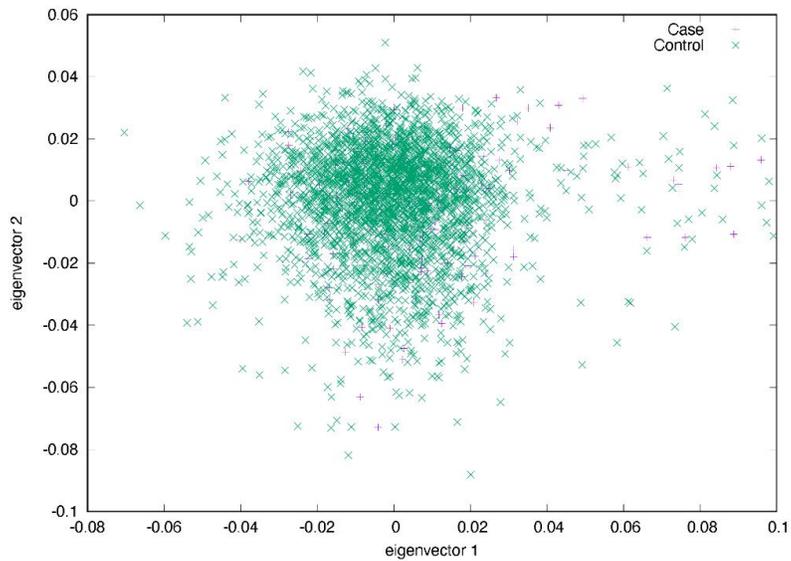
**Supplemental Table 11** Pathway analysis using GSEA (Table in a separate file)

**Supplemental Table 12** Pathway Analysis using SKAT robust test (Table in a separate file)

**Supplemental Table 13** Pathway analysis top hits (Table in a separate file)

**Supplemental Figure 1** Flowchart of bioinformatic processing and statistical methods  
(Table in a separate file)

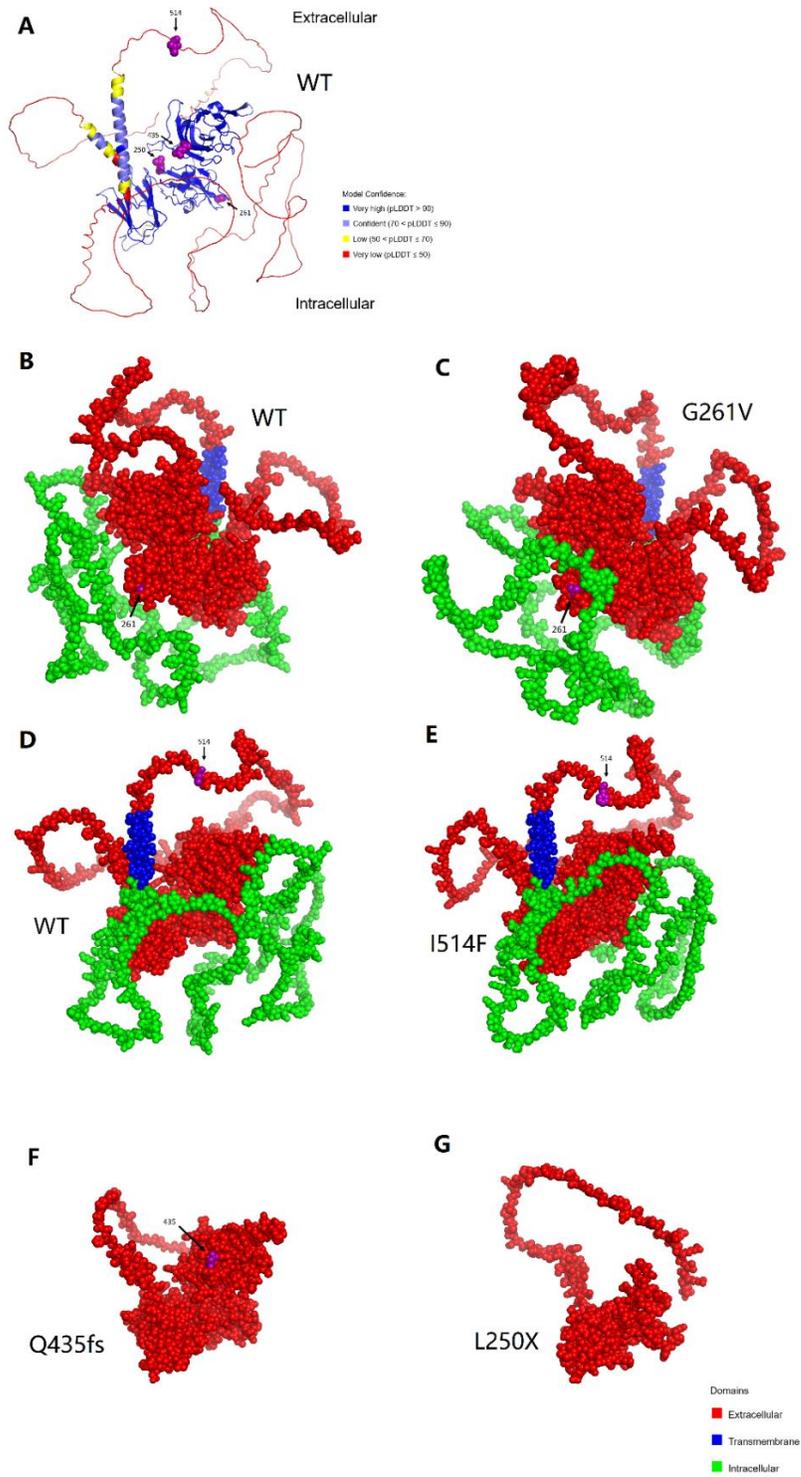


**Supplemental Figure 2** PCA plot of cases and controls

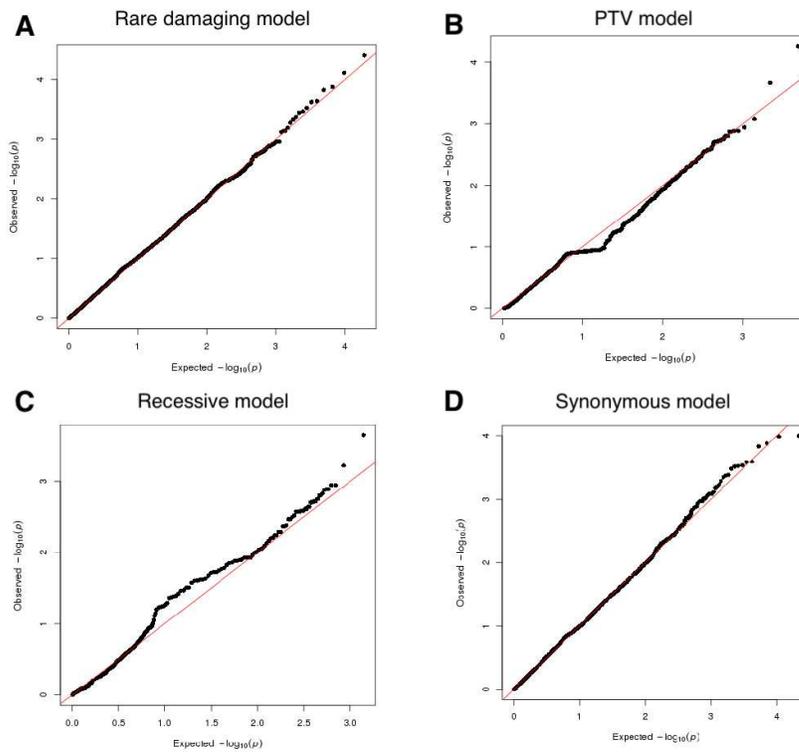
The PCA analysis examines population genetic structure and confirms that the cases and controls are from a homogenous ancestral cluster.



**Supplemental Figure 4 Three-dimensional structures of DCBLD2 protein with the discovered ultra-rare damaging variants predicted by AlphaFold2.** The protein structures of DCBLD2 wild type (WT) and with the ultra-rare damaging variants were predicted using AlphaFold2. **A.** Overview of the DCBLD2 wild type (WT) protein structure and the confidence of prediction. The predicted local distance difference test (pLDDT) is a per-residue estimate of its confidence in the prediction. **B, C.** Predicted changes in the structure of the DBLCD2 protein with the G261V variant (C) compared to WT (B). Glycine at position 261 is substituted by valine, which has a larger side chain and is more hydrophobic compared to glycine. **D, E.** Predicted changes in the structure of the DBLCD2 protein with the I514F variant (D) compared to WT (E). Isoleucine at position 514 is substituted by phenylalanine. Both Isoleucine and phenylalanine are hydrophobic. Isoleucine has an aliphatic side chain, while phenylalanine has a bulkier aromatic side chain. **F.** Predicted changes in the structure of the DBLCD2 protein with the Q435f variant, which results in a truncated protein lacking both the transmembrane and intracellular domains. **G.** Predicted changes in the structure of the DBLCD2 protein with the L250X variant, which results in a truncated protein lacking both the transmembrane and intracellular domains. (The figure can be found on the following page)



**Supplemental Figure 5** QQ plot of rare damaging model, PTV model, recessive model and synonymous model

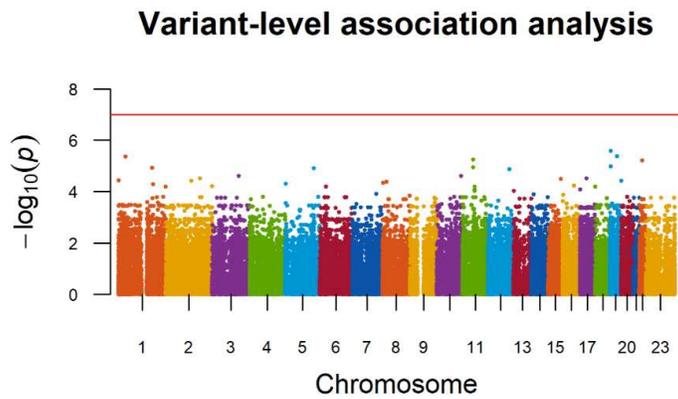


No genes were found associated with RP in the rare damaging model, PTV model, recessive model or synonymous model.

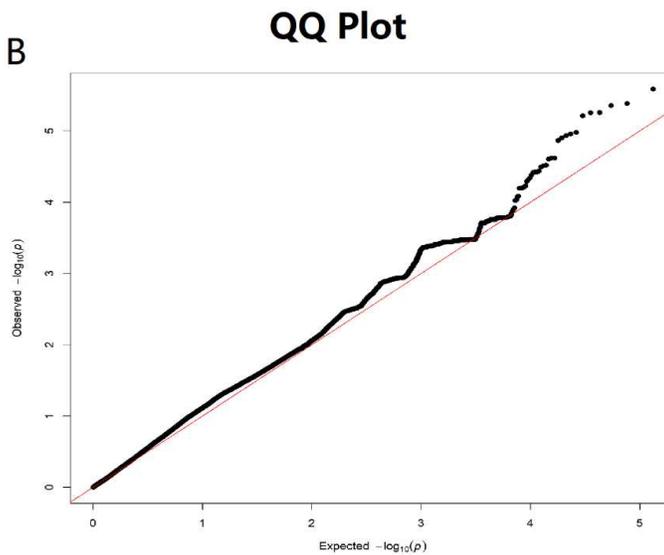
**Supplemental Figure 7** Sanger sequencing chromatograms of the discovered *DCBLD2* ultra-rare damaging QVs

**Supplemental Figure 7 Manhattan plot and permutation-based QQ plot in the variant-level association analyses.** No variant reached exome-wide significance and there were no compelling hypothesis-generating findings. There was no evidence of systematic bias or poor calibration (Lambda = 1.07).

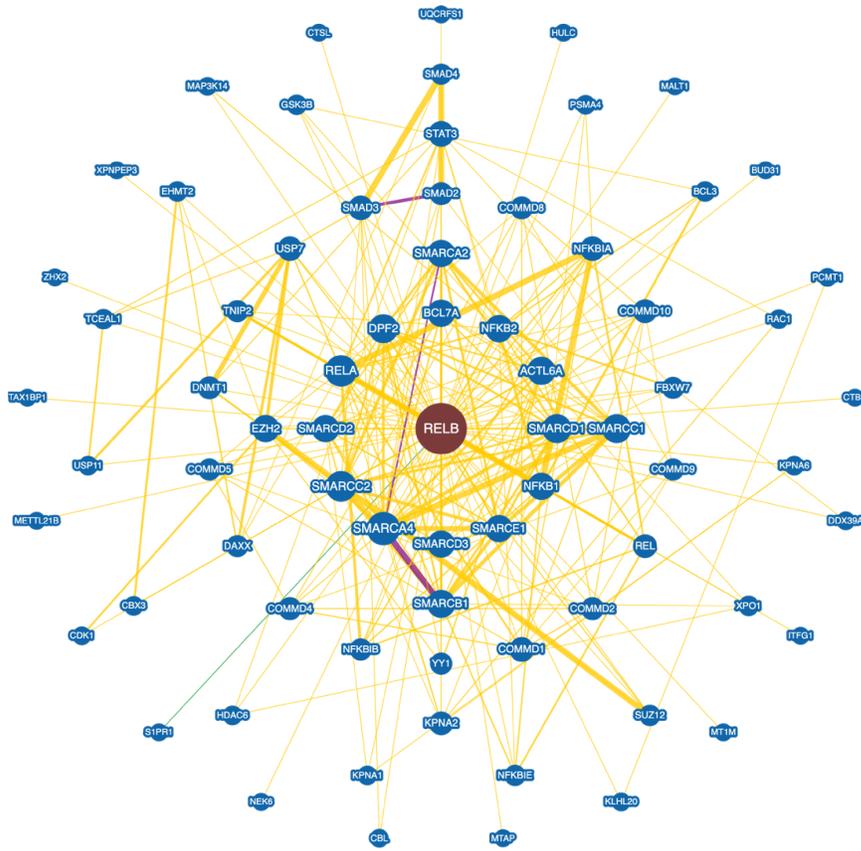
A



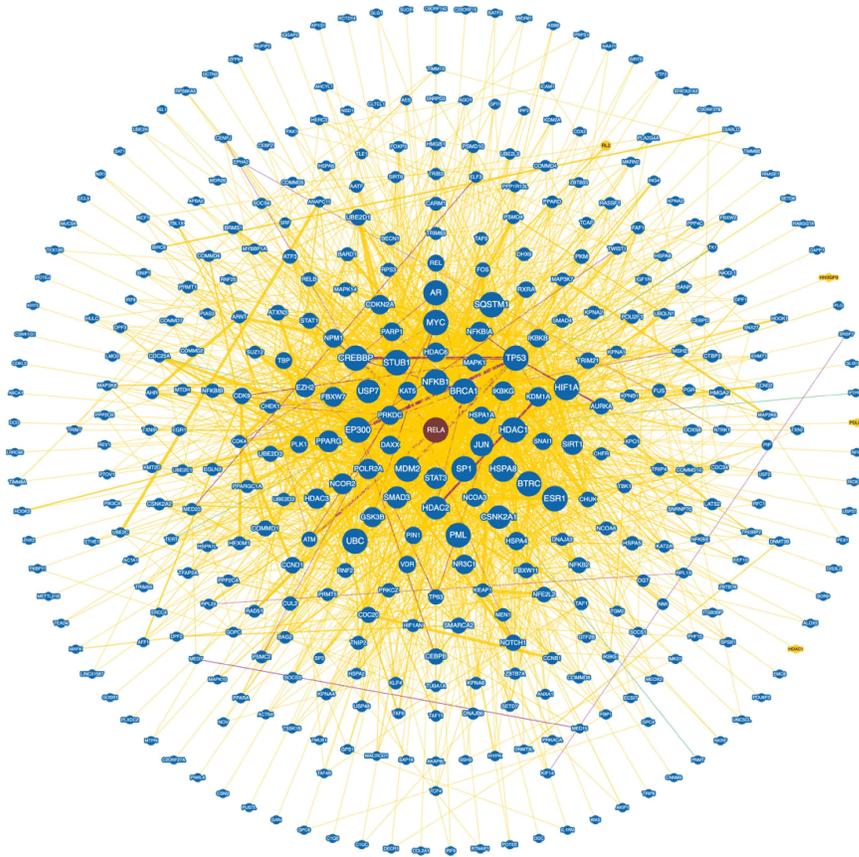
B



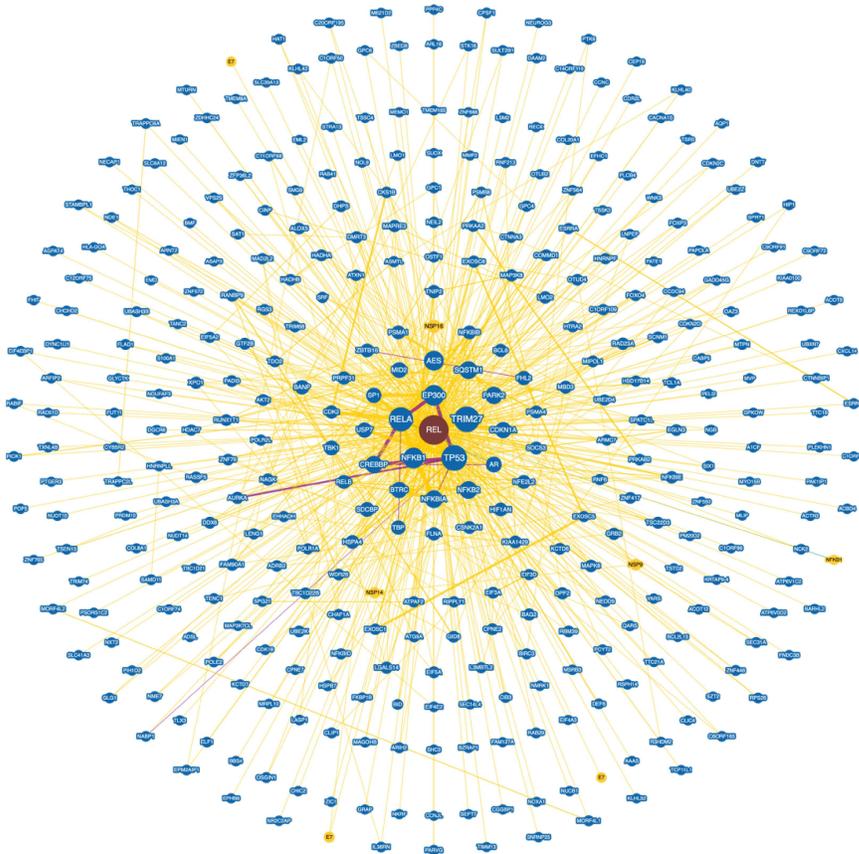
**Supplemental Figure 8** RELB protein interaction network



Supplemental Figure 9 RELA protein interaction network

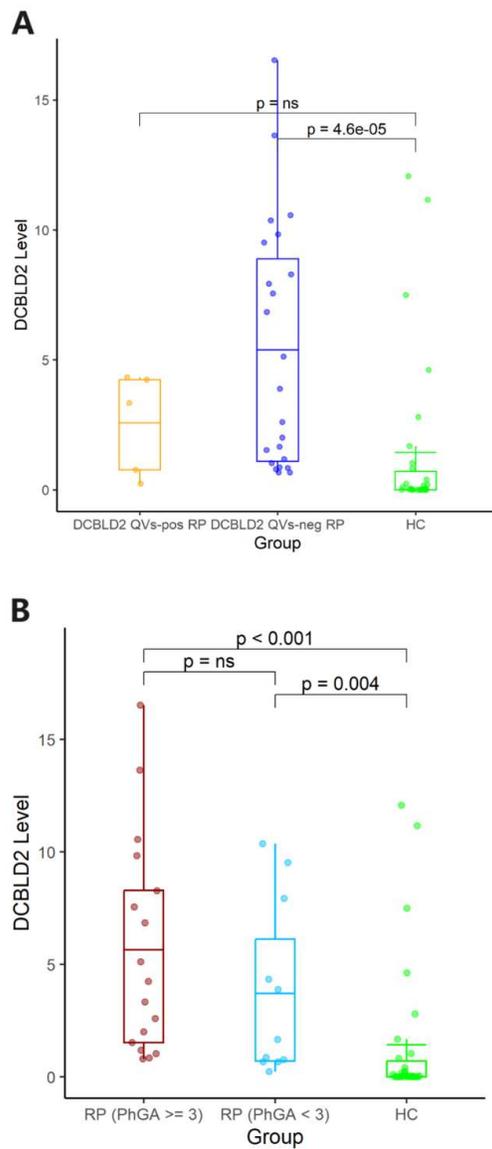


## Supplemental Figure 10 REL protein interaction network



**Supplemental Figure 11 Plasma DCBLD2 protein levels in RP subgroups and healthy controls.**

**A.** Plasma DCBLD2 levels in RP subgroups based on the *DCBLD2* ultra-rare QVs status. The DCBLD2 levels in RP patients with *DCBLD2* ultra-rare QVs (median 3.34 ng/uL) were numerically lower than RP patients without *DCBLD2* ultra-rare QVs (median 5.12 ng/uL), but still numerically higher than in healthy controls (HC) (0.05 ng/uL). **B.** Plasma DCBLD2 levels in RP subgroups stratified based on physician global assessment (PhGA). The DCBLD2 levels in RP patients with PhGA  $\geq 3$  (median 4.68 ng/uL) were numerically higher than RP patients with PhGA  $< 3$  (median 3.88 ng/uL). Notably, both groups had significantly higher levels than the HC group



## REFERENCES

1. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38, e164. 10.1093/nar/gkq603.
2. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol* 17, 122. 10.1186/s13059-016-0974-4.
3. Cameron-Christie, S., Wolock, C.J., Groopman, E., Petrovski, S., Kamalakaran, S., Povysil, G., Vitsios, D., Zhang, M., Fleckner, J., March, R.E., et al. (2019). Exome-Based Rare-Variant Analyses in CKD. *J Am Soc Nephrol* 30, 1109-1122. 10.1681/ASN.2018090909.
4. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079. 10.1093/bioinformatics/btp352.
5. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet* 2, e190. 10.1371/journal.pgen.0020190.
6. Petrovski, S., Todd, J.L., Durham, M.T., Wang, Q., Chien, J.W., Kelly, F.L., Frankel, C., Mebane, C.M., Ren, Z., Bridgers, J., et al. (2017). An Exome Sequencing Study to Assess the Role of Rare Genetic Variation in Pulmonary Fibrosis. *Am J Respir Crit Care Med* 196, 82-93. 10.1164/rccm.201610-2088OC.
7. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867-2873. 10.1093/bioinformatics/btq559.
8. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. 10.1186/s13742-015-0047-8.
9. Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J.A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B., et al. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* 53, 1097-1103. 10.1038/s41588-021-00870-7.
10. Youngchao Ge, S.D., Terence P. Speed (2003). Resampling-based multiple testing for microarray data analysis. *Test* 12, 1-77.
11. Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A., et al. (2017). The BioGRID interaction database: 2017 update. *Nucleic Acids Res* 45, D369-D379. 10.1093/nar/gkw1102.
12. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583-589. 10.1038/s41586-021-03819-2.
13. ColabFold - v1.5.2. (2023). <https://github.com/sokrypton/ColabFold>.

