The American Journal of Human Genetics, Volume *103*

Supplemental Data

The Genetic Landscape of Diamond-Blackfan Anemia

Jacob C. Ulirsch, Jeffrey M. Verboon, Shideh Kazerounian, Michael H. Guo, Daniel Yuan, Leif S. Ludwig, Robert E. Handsaker, Nour J. Abdulhay, Claudia Fiorini, Giulio Genovese, Elaine T. Lim, Aaron Cheng, Beryl B. Cummings, Katherine R. Chao, Alan H. Beggs, Casie A. Genetti, Colin A. Sieff, Peter E. Newburger, Edyta Niewiadomska, Michal Matysiak, Adrianna Vlachos, Jeffrey M. Lipton, Eva Atsidaftos, Bertil Glader, Anupama Narla, Pierre-Emmanuel Gleizes, Marie-Françoise O'Donohue, Nathalie Montel-Lehry, David J. Amor, Steven A. McCarroll, Anne H. O'Donnell-Luria, Namrata Gupta, Stacey B. Gabriel, Daniel G. MacArthur, Eric S. Lander, Monkol Lek, Lydie Da Costa, David G. Nathan, Andrei A. Korostelev, Ron Do, Vijay G. Sankaran, and Hanna T. Gazda

SUPPLEMENTAL FIGURE AND TABLE LEGENDS

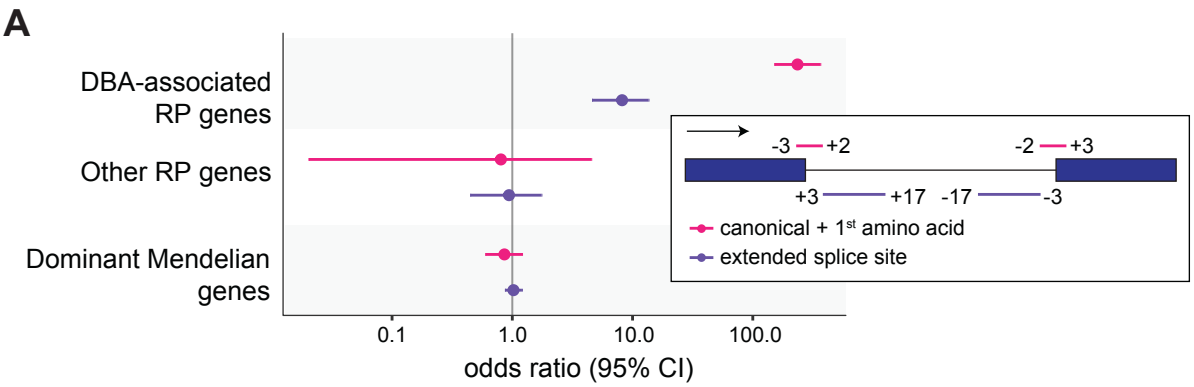
A
DBA-associated RP genes
Other RP genes
Dominant Mendelian genes
0.1
1.0
10.0
100.0
odds ratio (95% CI)
-3
+2
-2
+3
+3
+17
-17
-3
canonical + 1st amino acid
extended splice site

Figure S1. Splicing Mutations at Known DBA-associated Genes. (**A**) Both canonical and extended splicing mutations are enriched for known DBA genes. No enrichment was observed for other RP genes or for other dominant Mendelian genes.

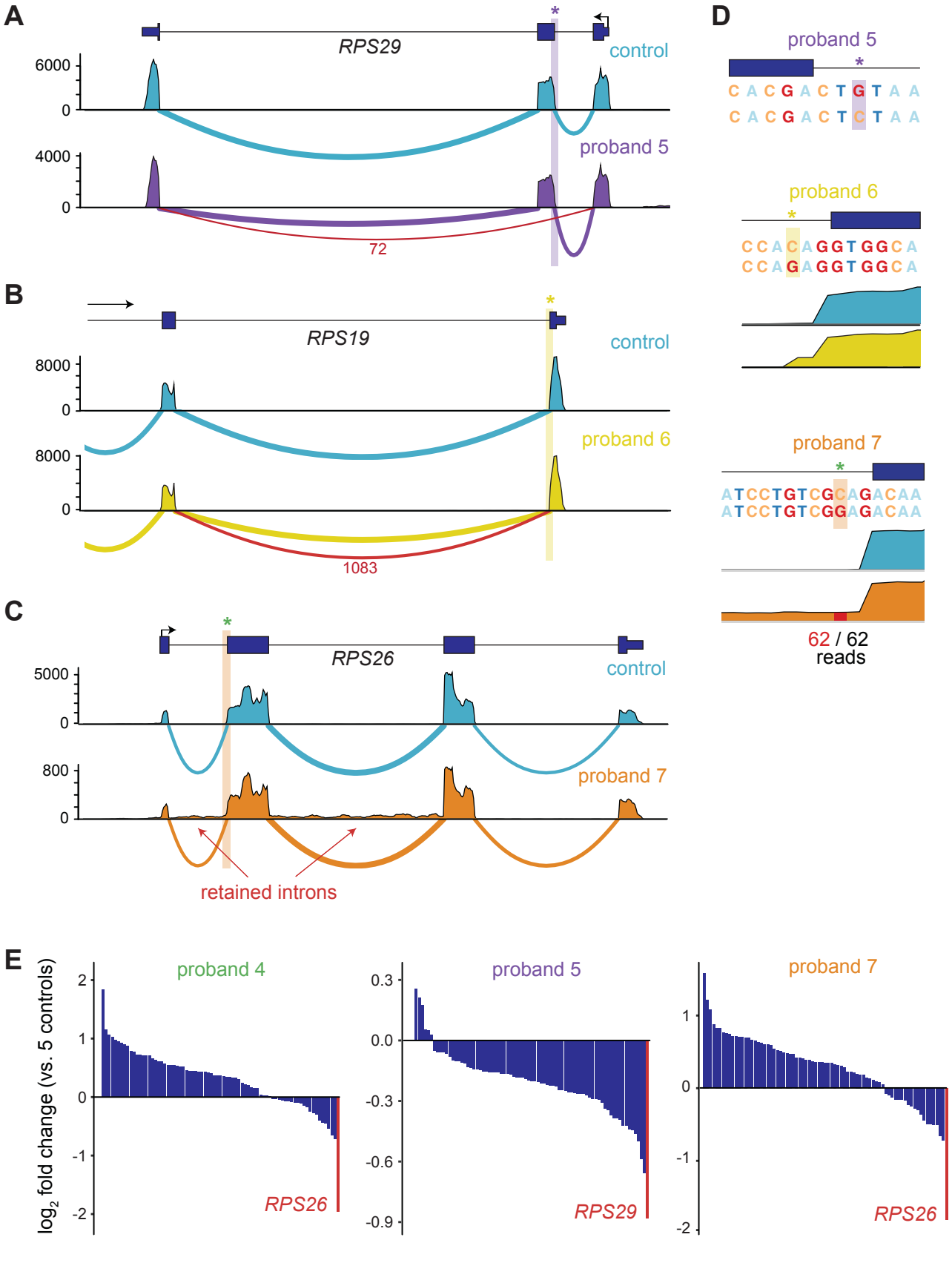
A
RPS29
control
6000
0
4000
0
proband 5
72
B
RPS19
control
8000
0
8000
0
proband 6
1083
C
RPS26
control
5000
0
800
0
proband 7
retained introns
D
proband 5
C A C G A C T G T A A
C A C G A C T C T A A
proband 6
C C A C A G G T G G C A
C C A G A G G T G G C A
proband 7
A T C C T G T C G C A G A C A A
A T C C T G T C G G A G A C A A
62 / 62
reads
E
log2 fold change (vs. 5 controls)
proband 4
2
1
0
-1
-2
RPS26
proband 5
0.3
0.0
-0.3
-0.6
-0.9
RPS29
proband 7
1
0
-1
-2
RPS26

Figure S2. Additional Non-Canonical Splice Variants in Known DBA-associated Genes.

(**A-C**) Sashimi plots of non-canonical splice mutants and a representative control are shown. The number of reads spanning each junction is indicated by the size of the sashimi plot curve. Novel junctions due to the mutation indicated in (**A-B**) are highlighted in red. (**D**) Location and consequence of each mutation is shown, in addition to coverage plots for the exon extension mutant (**B**) and intron retention mutant (**C**). (**E**) Log_2 fold change in transcripts per million across annotated RP genes for the indicated proband vs. 5 control LCLs.

Figure S3

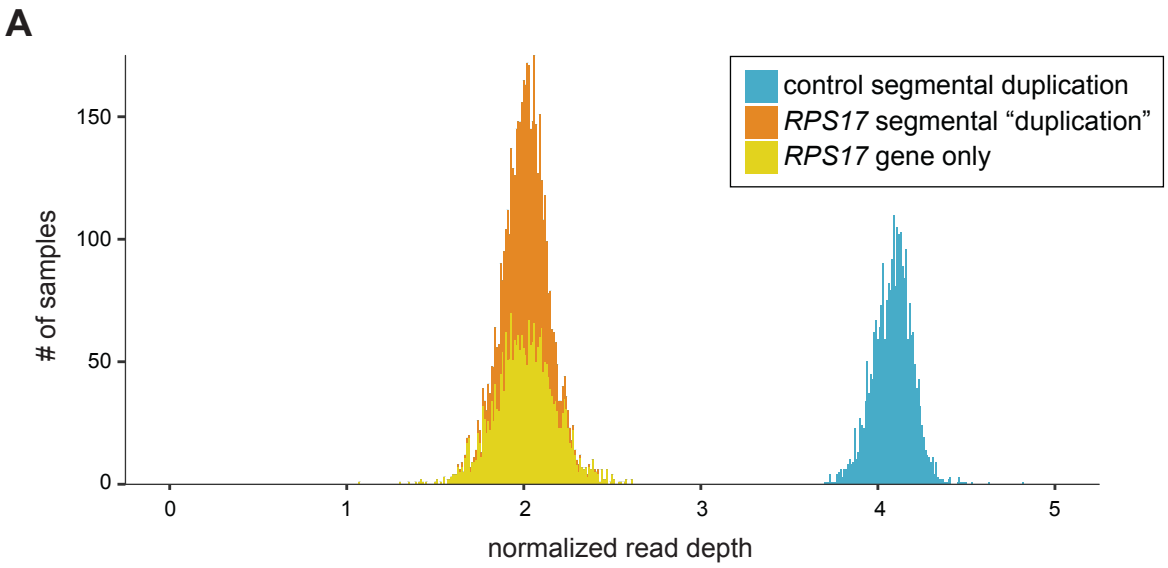
A
control segmental duplication
RPS17 segmental "duplication"
RPS17 gene only
# of samples
150
100
50
0
0
1
2
3
4
5
normalized read depth

Figure S3. Re-evaluations of *RPS17* Copy Numbers using WGS.
(**A**) Using 2,535 individual samples that underwent WGS from the 1000 Genomes Project, we estimated the copy number of the annotated segmental duplication containing *RPS17* (and *RPS17L*) in hg19. Nearly all samples have coverage indicative of only 2 copies for the *RPS17* "segmental duplication", rather than 4 copies, as can be observed in the control segmental duplication of approximately the same size.

Figure S4

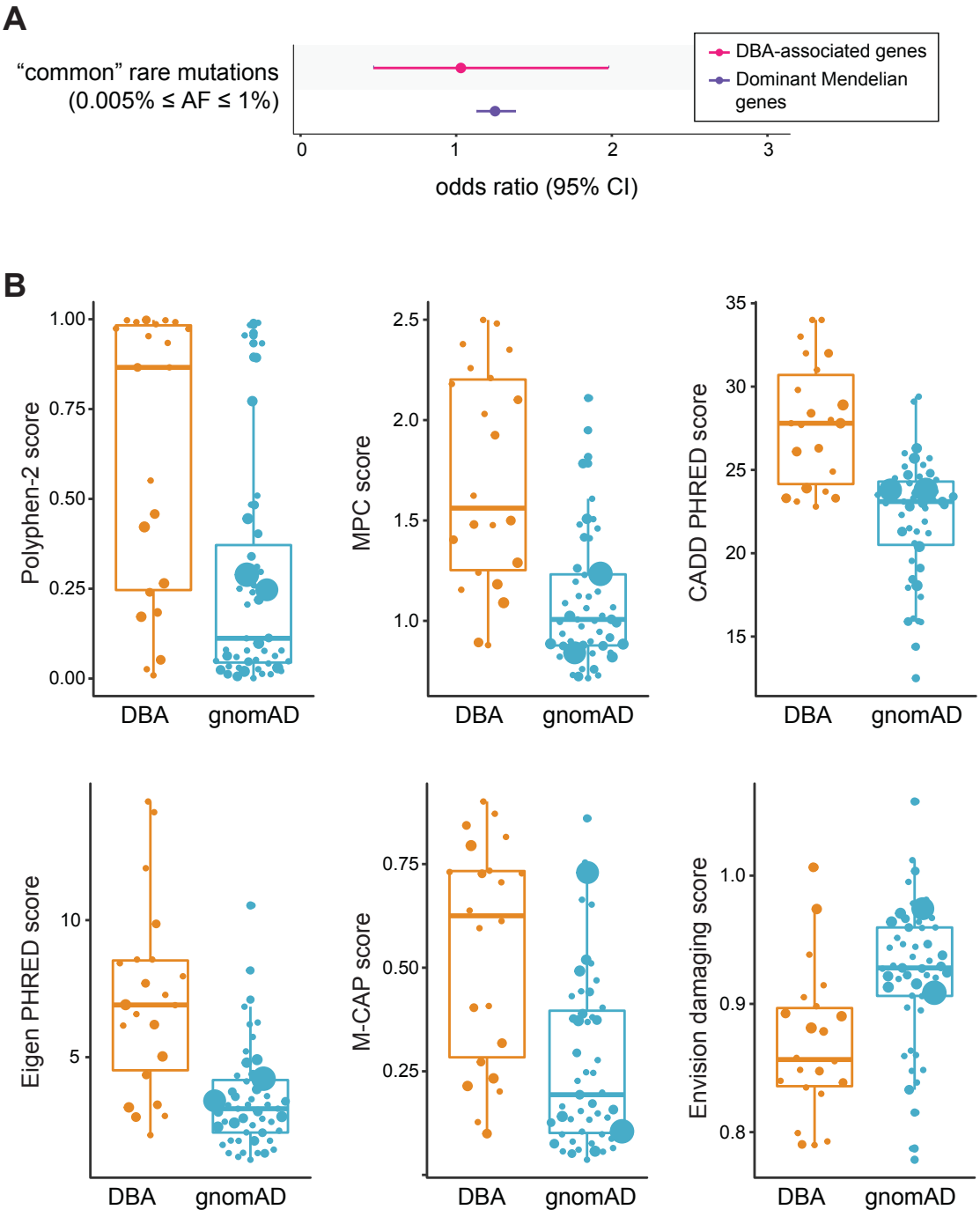
A
"common" rare mutations
(0.005% ≤ AF ≤ 1%)
DBA-associated genes
Dominant Mendelian genes
0
1
2
3
odds ratio (95% CI)
B
Polyphen-2 score
1.00
0.75
0.50
0.25
0.00
DBA
gnomAD
MPC score
2.5
2.0
1.5
1.0
CADD PHRED score
35
30
25
20
15
Eigen PHRED score
10
5
M-CAP score
0.75
0.50
0.25
Envision damaging score
1.0
0.9
0.8

Figure S4. No Evidence for More Common DBA Mutations.
(**A**) No enrichment is observed for more common mutations in known DBA-associated genes. There is a small but significant enrichment observed for other dominant Mendelian genes, indicating a possible mismatch in variant quality filtering or in population stratification that could result in unmodeled confounding. (**B**) Results from 6 well known missense variant effect predictors indicate that DBA *RPS19* mutations are more damaging than gnomAD *RPS19* mutations (no allele frequency filter was required for gnomAD mutations). However, no predictor can perfectly separate the two groups.

Figure S5

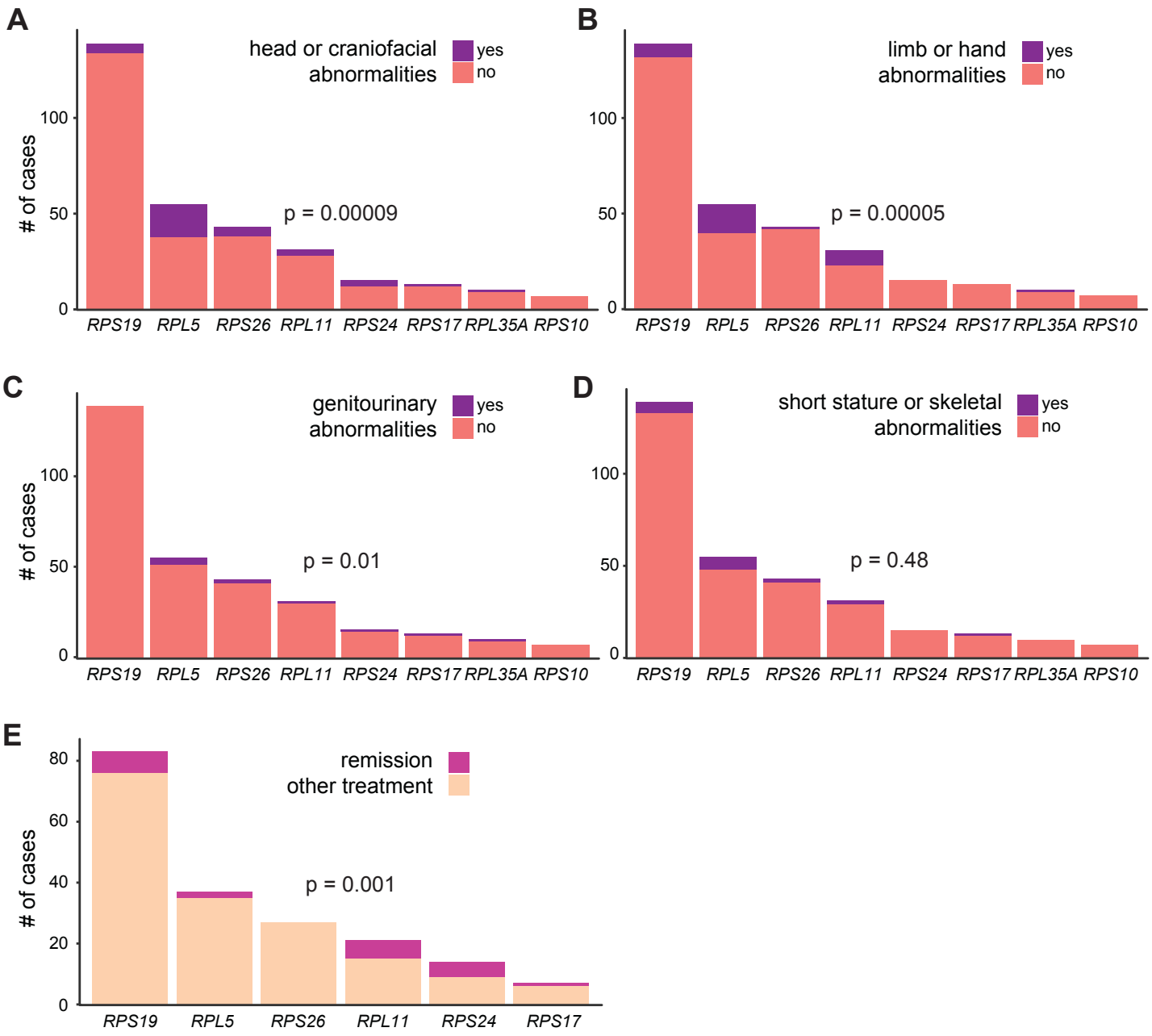
A
head or craniofacial abnormalities
yes
no
# of cases
100
50
0
p = 0.00009
RPS19 RPL5 RPS26 RPL11 RPS24 RPS17 RPL35A RPS10
B
limb or hand abnormalities
yes
no
100
50
0
p = 0.00005
RPS19 RPL5 RPS26 RPL11 RPS24 RPS17 RPL35A RPS10
C
genitourinary abnormalities
yes
no
# of cases
100
50
0
p = 0.01
RPS19 RPL5 RPS26 RPL11 RPS24 RPS17 RPL35A RPS10
D
short stature or skeletal abnormalities
yes
no
100
50
0
p = 0.48
RPS19 RPL5 RPS26 RPL11 RPS24 RPS17 RPL35A RPS10
E
remission
other treatment
# of cases
80
60
40
20
0
p = 0.001
RPS19 RPL5 RPS26 RPL11 RPS24 RPS17

Figure S5. Additional Phenotypic Associations.
Differences in (**A**) head or craniofacial abnormalities, (**B**) limb or hand abnormalities, (**C**) genitourinary abnormalities, (**D**) short stature or skeletal abnormalities, and (**E**) remission status were observed between different RP genes. A χ^2 test was used to test the hypothesis that there were differences in proportion of the outcome between RP genes.

Figure S6

A

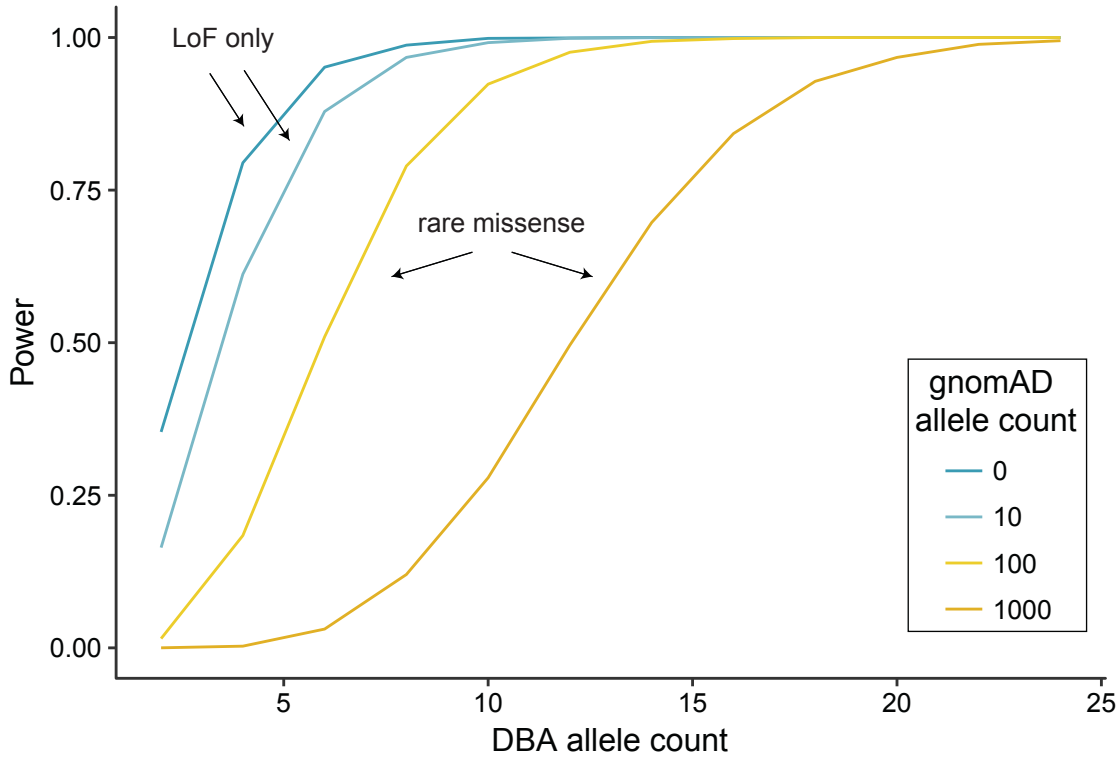
LoF only
rare missense
Power
1.00
0.75
0.50
0.25
0.00
5
10
15
20
25
DBA allele count
gnomAD
allele count
0
10
100
1000

Figure S6. Gene Burden Power Analysis.
(**A**) Power for different allele count scenarios was calculated using a 1-sided Fisher's exact test. Within the quality normalized gnomAD dataset, we observe that the top 10 and 25% of constrained genes have fewer than 4 or 10 rare LoF allele counts in gnomAD, respectively. Thus, we consider genes with 0-10 rare LoF as "constrained" genes, similar to the RP genes already implicated in DBA (median of 0 rare LoF alleles in gnomAD). In this scenario, the gene burden tests were well powered (> 80%) to detect an exome-wide significant gene association with as few as 4 to 6 individuals with LoF mutations (corresponding to 1-1.5% of DBA incidence). Similarly, we observe that the top 25% of constrained genes have 165 missense mutations and 53 predicted damaging missense mutations (median of 59 and 8 for RP genes implicated in DBA, respectively). Thus, we consider 100 counts in gnomAD as a reasonable number of rare missense alleles for a "constrained" gene, but we also consider an extreme scenario of up to 1,000 allele counts. In both scenarios, gene burden tests were well powered (> 80%) to detect an exome-wide significant gene association with as few as 8 to 16 individuals with missense mutations (corresponding to 2-4% of DBA incidence). Thus, we are theoretically well powered to detect mutations between 1-4% of total DBA incidence. However, after conservatively adjusting the variant quality threshold in our DBA cohort specifically for burden analysis, we were unable to detect exome-wide significant associations for RPL35A (2%) and RPS24 (3%) as several validated variants were filtered due to lower quality scores in WES, but we could detect exome-wide significant associations for RPS10 (1%) and RPL11 (7%). Given these considerations, we believe that a more appropriate, if slightly conservative, estimate of our true power lies closer to ≈ 5%.

Table S1. Frequency of Cohort Recruitment by Country and Center. Table S2. Summary of Identified Mutations.

Table S3. Solve Rate Across Families and Singletons.

Table S4. Variants Missed By WES But Identified By Sanger Sequencing. Table S5. Putative CNVs Identified by WES.

Table S6. WES-based CNV Validation.

Table S7. DBA-associated Genes Are De-enriched For LoF And Missense Variants.

Table S8. Predicted Structural Impacts of *RPS19* Missense Mutations.

Table S1. Frequency of Cohort Recruitment by Country and Center.

Registries	no.	%
DBA Registry North America	112	23.7%
French DBA Registry	73	15.5%
Hematology centers		
USA	161	34.1%
Poland	67	14.2%
Turkey	16	3.4%
UK	6	1.3%
Singapore	4	0.8%
Germany	3	0.6%
The Netherlands	3	0.6%
Australia	3	0.6%
UAE	2	0.4%
Bahrain	1	0.2%
Canada	1	0.2%
Greece	1	0.2%
Hungary	1	0.2%
Iceland	1	0.2%
India	1	0.2%
Mexico	1	0.2%
Direct Contact		
USA	15	3.2%

Table S3. Solve Rate Across Families and Singletons.

Category	Solved	Total	%
Unaffected parents	17	27	63%
Affected 1st/2nd/3rd degree relative	17	23	74%
Singleton	276	352	78%

Table S4. Variants Missed By WES But Identified By Sanger Sequencing.

Gene	Mutation	Type	Issue
RPL5	chr1:93306123:A>AAGATGTATA	inframe	long indel
*RPL5**	chr1:93301919:CTGTGG>C + 39 bp insertion	frameshift	long indel
*RPL5**	chr1:93301919:CTGTGG>C + 39 bp insertion	frameshift	long indel
RPS10	chr6:34392996:C>T	start site	low coverage
RPS17	chr15:82823347:CTC>C	frameshift	duplicated gene
RPS17	chr15:82824477:GCG>AA	frameshift	duplicated gene
RPS24	—	frameshift	long indel

*cousins

Table S6. WES-based CNV Validation.

Sample	Gene	Validated Exon #
DBA_del_1	*RPL11*	2
DBA_del_2	*RPL15*	—**
DBA_del_3	*RPL35A*	3
DBA_del_4	*RPL35A*	3
DBA_del_5	*RPS17*	1,2,3,4,5
DBA_del_6	*RPS17*	1,2,3,4,5
DBA_del_7	*RPS17*	1,2,3,4,5
DBA_del_8	*RPS17*	1,2,3,4,5
DBA_del_9	*RPS17*	1,2,3,4,5
DBA_del_10	*RPS17*	1,2,3,4,5
DBA_del_11	*RPS17*	1,2,3,4,5
DBA_del_12	*RPS17*	1,2,3,4,5
DBA_del_13	*RPS17*	1,2,3,4,5
DBA_del_14	*RPS19*	2
DBA_del_15	*RPS19*	2
DBA_del_16	*RPS19*	2
DBA_del_17	*RPS19*	6
DBA_del_18	*RPS19*	6
DBA_del_19	*RPS19*	2,4,5,6
DBA_del_20	*RPS19*	2,4,5,6
DBA_del_21	*RPS19*	2,4,5,6
DBA_del_22	*RPS19*	4,5
DBA_del_23	*RPS24*	2
DBA_del_24	*RPS24*	2
DBA_del_25	*RPS24*	2
DBA_del_26	*RPS24*	5
DBA_del_27	*RPS26*	2
DBA_del_28	*RPS26*	2
DBA_del_29	*RPS26*	2
DBA_del_30	*RPS26*	2
DBA_del_31	*RPS26*	—**

**validated by array CGH

Table S7. DBA-associated Genes Are De-enriched For LoF And Missense Variants.

Gene	pLi (LoF)	o/e LoF (90% CI)	o/e mis (90%) CI	z (missense)
RPL11	0.94	0.00 (0.00-0.34)	0.50 (0.40-0.63)	1.82
RPL15	0.97	0.00 (0.00-0.27)	0.53 (0.44-0.65)	1.92
RPL18	0.96	0.00 (0.00-0.29)	0.57 (0.47-0.69)	2.20
RPL26	0.92	0.00 (0.00-0.37)	0.52 (0.41-0.67)	1.62
RPL27	0.86	0.00 (0.00-0.47)	0.64 (0.52-0.81)	1.18
RPL31	0.86	0.00 (0.00-0.47)	0.49 (0.38-64)	1.66
RPL35	0.56	0.15 (0.05-0.72)	0.62 (0.05-0.79)	1.23
RPL35A	0.91	0.00 (0.00-0.40)	0.52 (0.39-0.69)	1.41
RPL5	0.99	0.00 (0.00-0.20)	0.59 (0.50-0.69)	1.96
RPS10	0.94	0.00 (0.00-0.33)	0.65 (0.53-0.80)	1.27
RPS15A	0.81	0.00 (0.00-0.55)	0.26 (0.18-0.38)	2.32
RPS17	—*	—*	—*	—*
RPS19	0.82	0.00 (0.00-0.53)	0.60 (0.48-0.76)	1.34
RPS24	0.69	0.12 (0.04-0.58)	0.74 (0.63-0.86)	1.17
RPS26	0.83	0.00 (0.00-0.52)	0.36 (0.26-0.50)	1.95
RPS27	0.77	0.00 (0.00-0.63)	0.56 (0.40-0.78)	1.05
RPS28	0.69	0.00 (0.00-0.79)	0.31 (0.20-0.51)	1.54
RPS29	0.31	0.25 (0.09-1.17)	0.49 (0.34-0.72)	1.96
RPS7	0.96	0.00 (0.00-0.30)	0.56 (0.45-0.69)	1.72
*TSR2***	0.81	0.00 (0.00-0.56)	0.66 (0.51-0.85)	1.01
*GATA1***	0.95	0.00 (0.00-0.32)	0.75 (0.65-0.87)	0.77

pLI, probability of loss of function intolerance; o/e observed/expected; CI, confidence interval;

*hg19 duplicate genes not included

**X-chromosomal genes