Supporting Information

Towards improved genomic decision with structure-assisted variant interpretation

Nathália P.D. de Nigro[1,2 §], Paula F.C. Franklin[1,2 §], Lucas A. Machado[3,4], Daniel C.P. Câmara[5], Roberta Soares Faccion[4,6 *], Zilton Vasconcelos[4 *] and Mauricio G.S. Costa[1,2 *]

1. Programa de Computação Científica, Vice-Presidência de Educação, Informação e Comunicação, Fundação Oswaldo Cruz. Av.Brasil 4365, 21040-900, Rio de Janeiro, Brazil.
2. Programa de Pós-Graduação em Biologia Computacional e Sistemas, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz. Av.Brasil 4365, 21040-900, Rio de Janeiro, Brazil.
3. Instituto de Tecnologia em Imunobiológicos (Biomanguinhos) Fundação Oswaldo Cruz. Av.Brasil 4365, 21040-900, Rio de Janeiro, Brazil
4. Laboratório de Alta Complexidade (LACIFF), Unidade de Pesquisa Clínica, Instituto Nacional de Saúde da Mulher, da Criança e do Adolescente Fernandes Figueira (IFF), Fundação Oswaldo Cruz (Fiocruz), Av. Rui Barbosa, 716 – Flamengo, 22250-020, Rio de Janeiro, RJ, Brazil.
5. Brazilian Medical Unit - BRAMU, Médicos Sem Fronteiras Brasil, Rio de Janeiro, Brazil
6. Laboratório de Imunofisiologia, Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Av. Carlos Chagas Filho 373 – Centro de Ciências da Saúde – Cidade Universitária, 21941-902, Rio de Janeiro, RJ, Brazil.

Supplementary Table 1: Information related to the obtaining of curated models

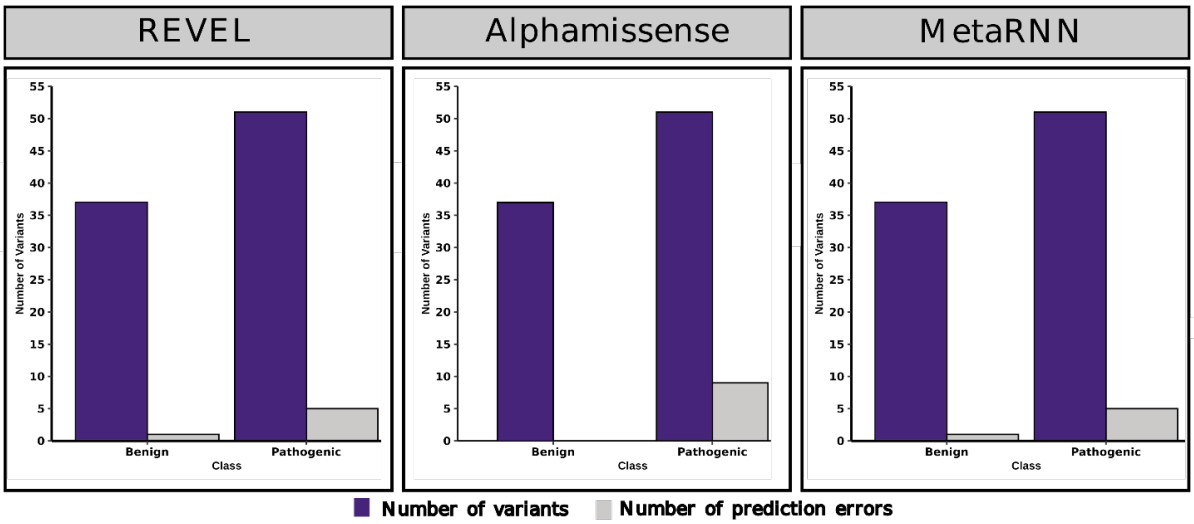
	ADA2	**CDC42**	**STING1**	**TBK1**	**WDR1**
Uniprot code	Q9NZK5	P60953	Q86WV6	Q9UHD2	O75083
Template PDB code	3LGD	2NGR	6NT5	6NT9 / 6RSR	-
Biological assembly	Homodimer	Monomer	Homodimer, forms high-order oligomers	Homodimer	Monomer
Template AlphaFold dB	-	-	AF-Q86WV6-F1-v4	AF-Q9UHD2-F1-v4	AF-O75083-F1-v4
Protein length	511	191	379	729	606
Model length	482	191	343	658	606

Supplementary Table 2: Description of ACMG/AMP criteria considered in this study

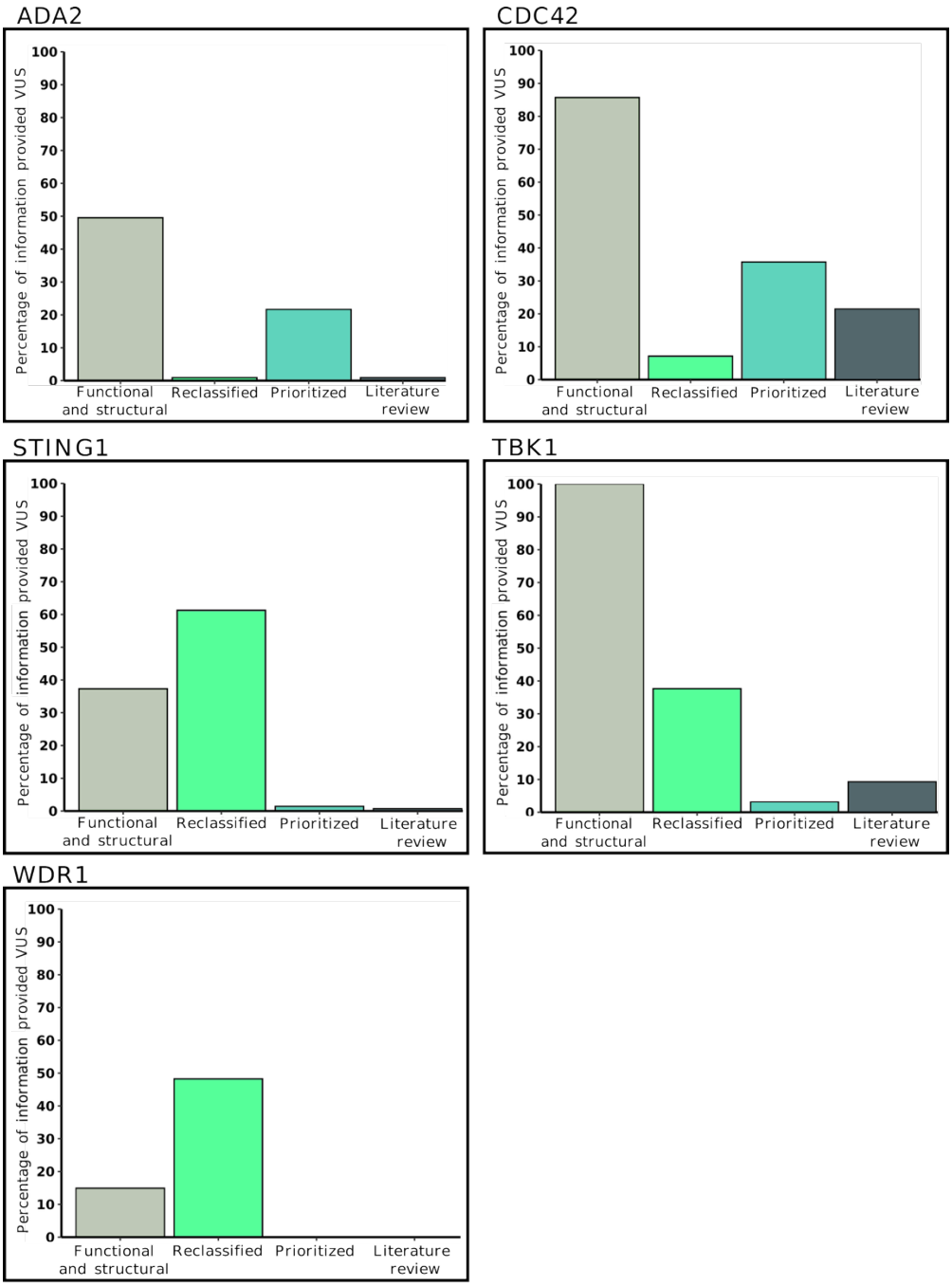
Criteria	Description	Class	Evidence Strength	Points	Reference
BA1	Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium.	benign	stand alone	NA	ACMG/AMP [9]
BS1	Allele frequency is greater than expected for disorder.	benign	strong	-4	VarSome [15]
PM2	Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium.	pathogenic	supporting	+1	ClinGen [12]
BP1	Missense variant in a gene for which primarily truncating variants are known to cause disease.	benign	supporting	+1	VarSome [15]
PP2	Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease.	pathogenic	supporting	+1	VarSome [15]
PM1	Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation.	pathogenic	moderate	+2	ACMG/AMP [9]
PM5	Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before.	pathogenic	moderate	+2	ACMG/AMP [9]
PP3	Multiple lines of computational evidence support a deleterious effect on the gene or gene product.	pathogenic	supporting, to strong	+1 to +4	ClinGen [12]
BP4	Multiple lines of computational evidence suggest no impact on gene or gene product.	benign	supporting, to strong	-1 to -4	ClinGen [12]
PP5	Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation.	pathogenic	supporting	+1	ACMG/AMP [9]
BP6	Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation.	benign	supporting	-1	ACMG/AMP [9]

Supplementary Table 3: Thresholds adopted to adjust the strength of in silico predictors

	REVEL	AlphaMissense	MetaRNN
Benign Strong	< 0.016	<= 0.0853	<= 0.108
Benign Moderate	0.016 to 0.183	> 0.0853 to 0.166	> 0.108 to 0.267
Benign Supporting	> 0.183 to 0.29	> 0.166 to 0.316	> 0.267 to 0.43
Ambiguous	> 0.29 to 0.644	>0.316 to 0.787	> 0.43 to < 0.748
Pathogenic Supporting	> 0.644 to 0.773	> 0.787 to 0.956	>= 0.748 to 0.841
Pathogenic Moderate	> 0.773 to 0.932	> 0.956 to 0.994	>= 0.841 to 0.939
Pathogenic Strong	> 0.932	> 0.994	>= 0.939



Supplementary Figure 1: Divergence between REVEL, AlphaMissense, and MetaRNN predictions and clinical evidence from ClinVar. Bar plots show the total number of variants labeled as Benign or Pathogenic in ClinVar for the genes in our dataset (purple) and the number of classification errors made by each predictor (gray). Errors occur when the prediction conflicts with the clinical classification (for example, a pathogenic variant predicted as benign).



Supplementary Figure 2: SAVI performance statistics per gene. Percentage of VUS for which SAVI provided information and/or reclassification.