The American Journal of Human Genetics, Volume *102*

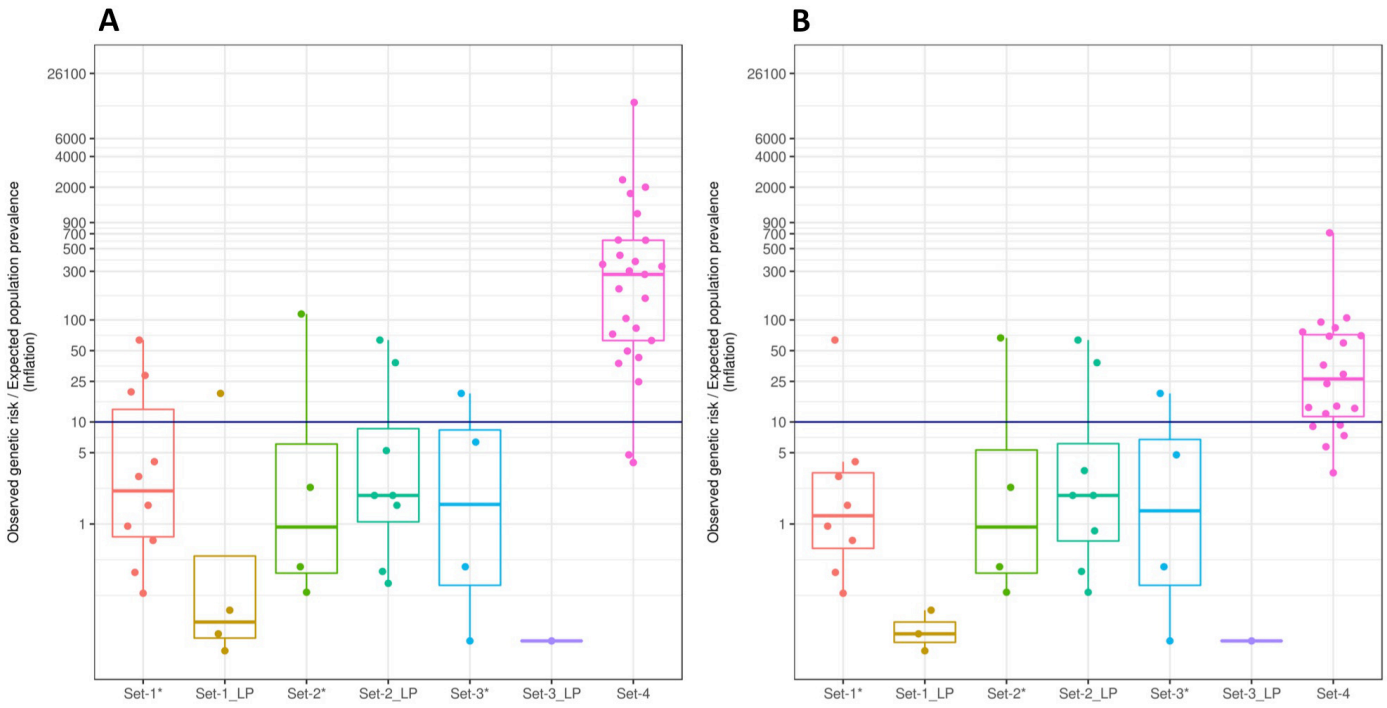
Supplemental Data

Identification of Misclassified ClinVar Variants via Disease Population Prevalence

Naisha Shah, Ying-Chen Claire Hou, Hung-Chun Yu, Rachana Sainger, C. Thomas Caskey, J. Craig Venter, and Amalio Telenti

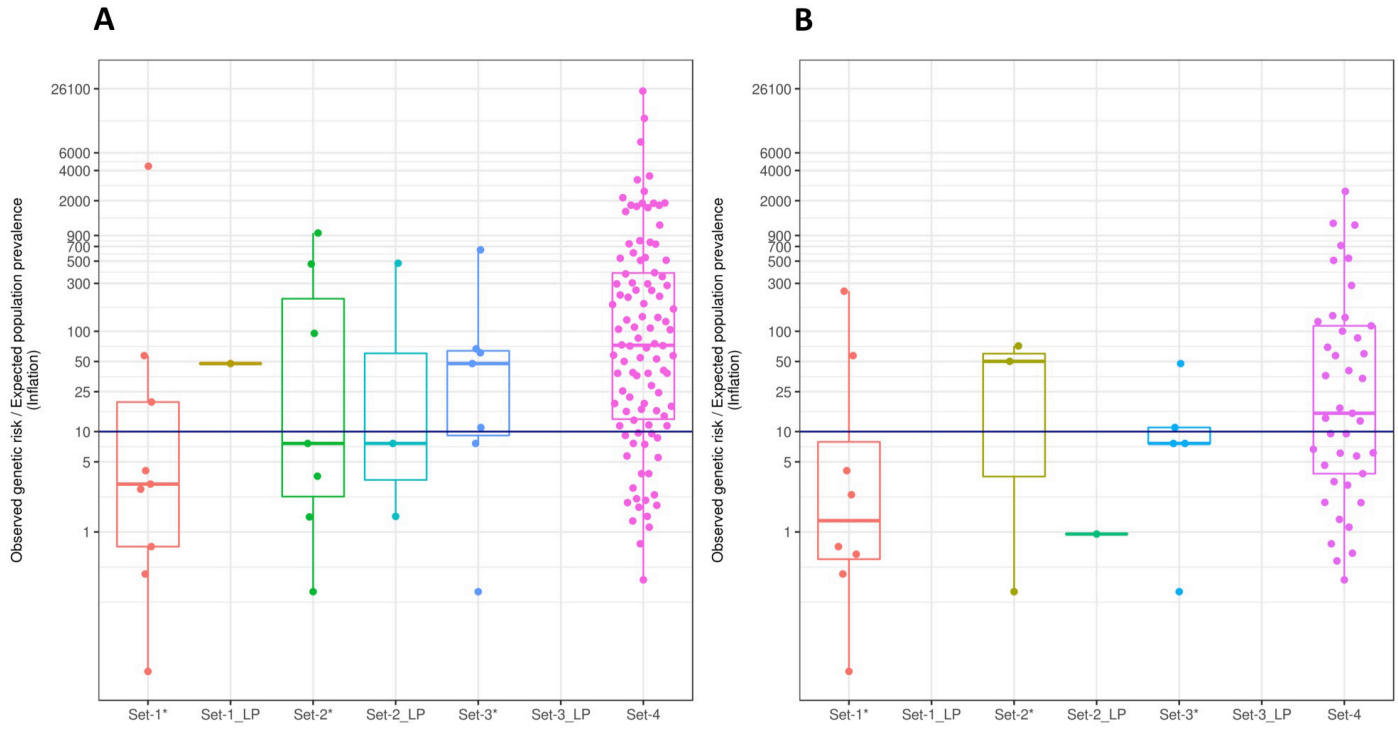
SUPPLEMENTAL FIGURES

Figure S1: Genetic risk in ACMG-59 conditions with additional P and LP sets



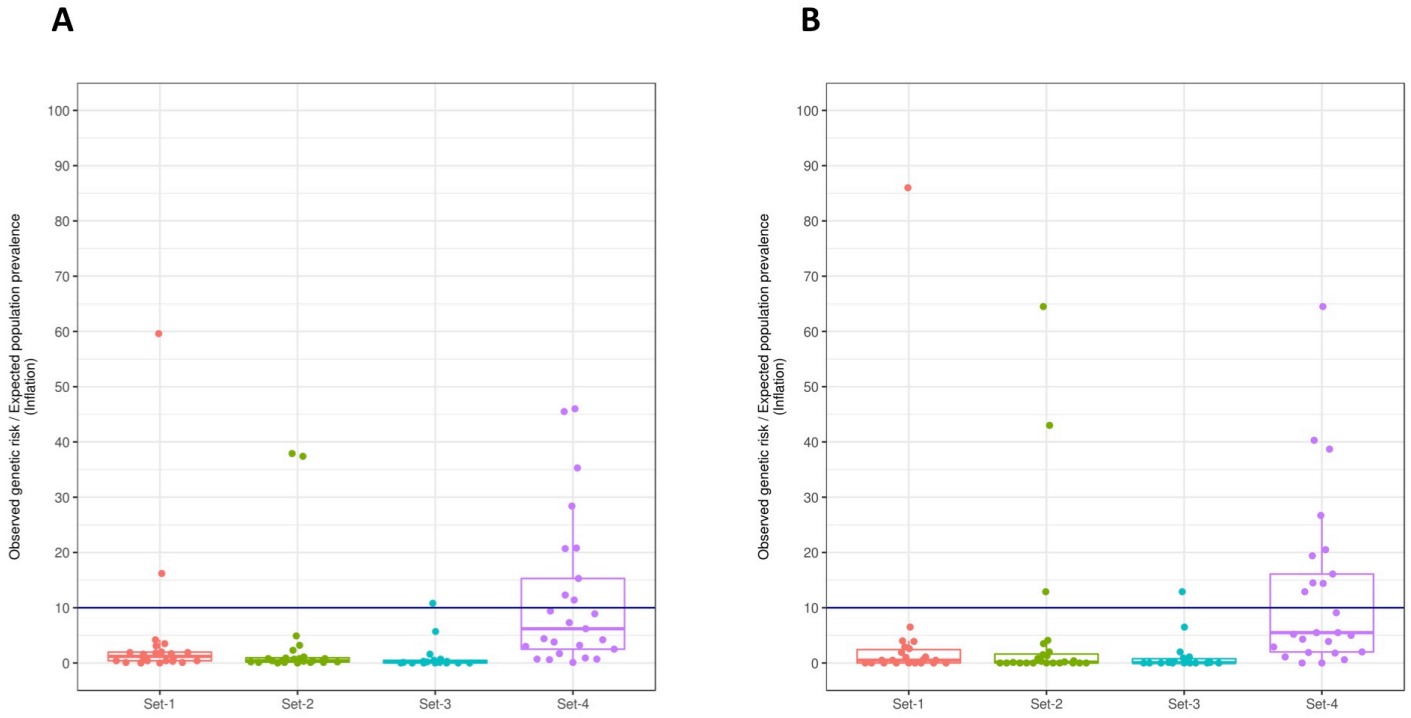
Fold-change of observed genetic risk over expected population prevalence using ClinVar variant sets for the ACMG-59 conditions. The observed genetic risk was calculated using the study population. Each point represents a condition; each condition may be represented in more than one set. The navy-blue line at a fold-change of 10 (i.e. inflation) indicates a theoretical penetrance of 10%. Observations above this line are highly suggestive of misclassified variants. A) Fold-change was calculated using variants per variant set: Set-1* consists of pathogenic (P) variants with 2 or more ClinVar review stars (i.e. two or more submitters with assertion criteria, expert panel and practice guideline); Similarly, Set-1_LP consists of LP variants. Set-2* consists of P variants with 1 star (i.e. one submitter with assertion criteria); Similarly, Set-2_LP consists of LP variants. Set-3* consists of P variants with 0 star (i.e. submitter with no assertion criteria submitted in ClinVar); Similarly, Set-3_LP consists of LP variants. Set-4 consists of variants with conflicting interpretations of pathogenicity. B) Fold-change was re-calculated after variants were filtered for disease-specific minor allele frequency thresholds.

Figure S2: Genetic risk in Orphanet conditions with additional P and LP sets



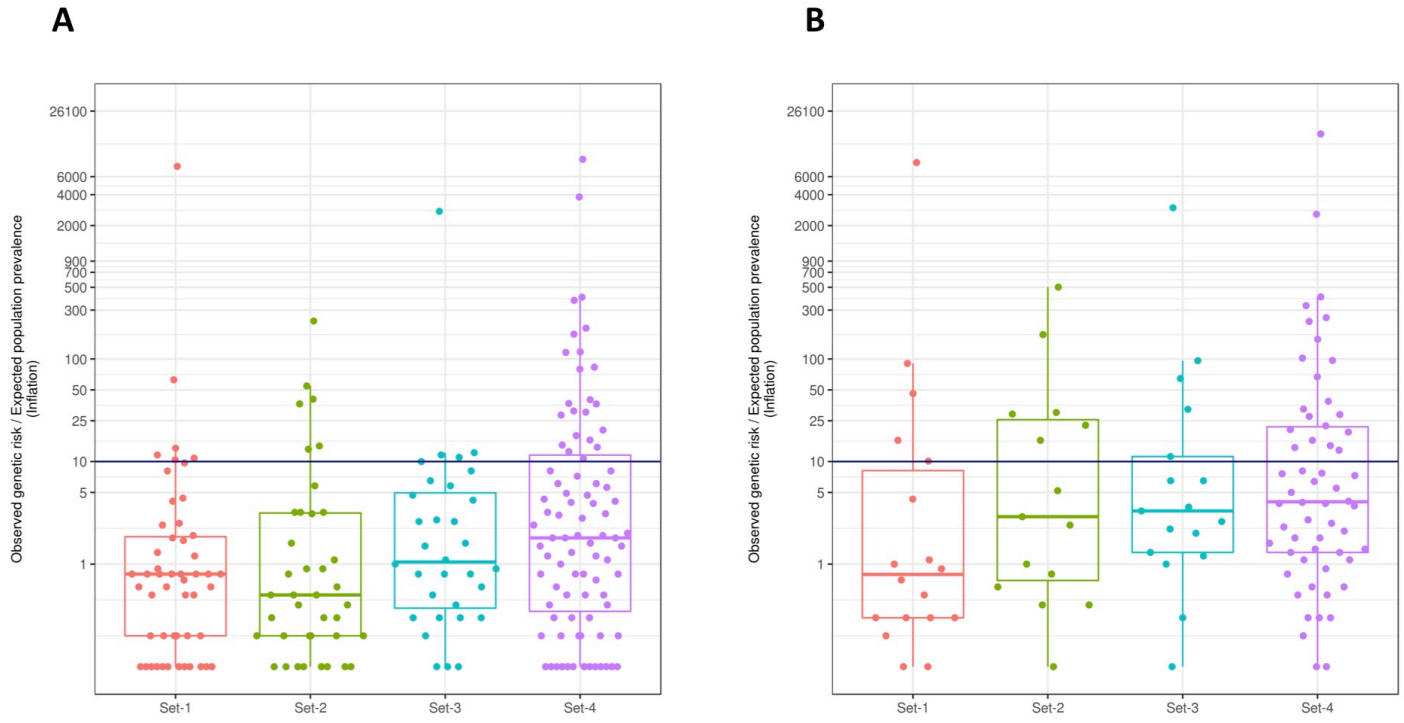
Fold-change of observed genetic risk over expected population prevalence using ClinVar variant sets for the Orphanet conditions. The observed genetic risk was calculated using the study population. Each point represents a condition; each condition may be represented in more than one set. The navy-blue line at a fold-change of 10 (i.e. inflation) indicates a theoretical penetrance of 10%. Observations above this line are highly suggestive of misclassified variants. A) Fold-change was calculated using variants per variant set: Set-1* consists of pathogenic (P) variants with 2 or more ClinVar review stars (i.e. two or more submitters with assertion criteria, expert panel and practice guideline); Similarly, Set-1_LP consists of LP variants. Set-2* consists of P variants with 1 star (i.e. one submitter with assertion criteria); Similarly, Set-2_LP consists of LP variants. Set-3* consists of P variants with 0 star (i.e. submitter with no assertion criteria submitted in ClinVar); Similarly, Set-3_LP consists of LP variants. Set-4 consists of variants with conflicting interpretations of pathogenicity. B) Fold-change was re-calculated after variants were filtered for disease-specific minor allele frequency thresholds.

Figure S3: Genetic risk in ACMG-59 conditions using gnomAD



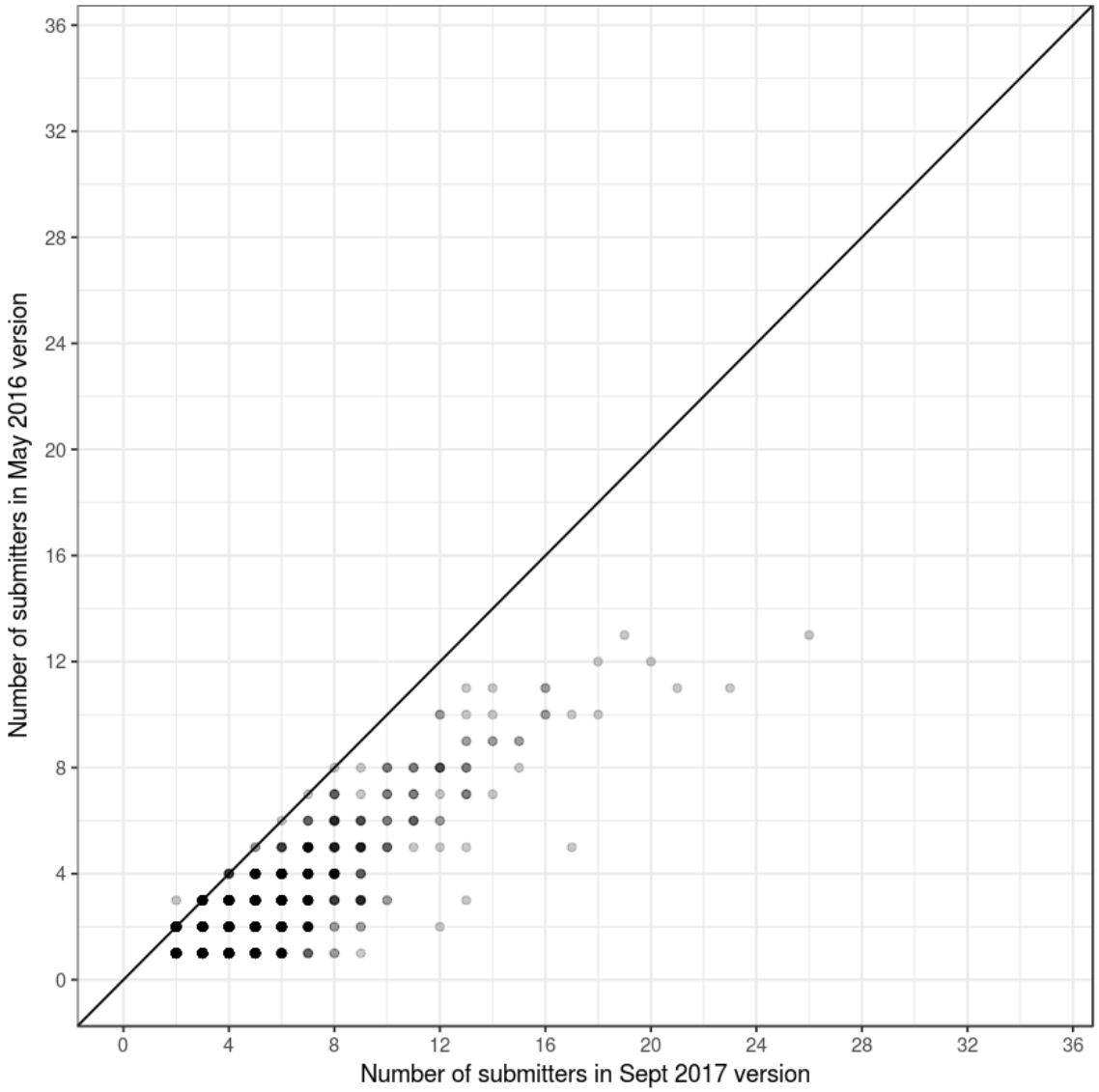
Fold-change of observed genetic risk over expected population prevalence using ClinVar variant sets for the ACMG-59 conditions. Each point represents a condition; each condition may be represented in more than one set. The navy-blue line at a fold-change of 10 (i.e. inflation) indicates a theoretical penetrance of 10%. Observations above this line are highly suggestive of misclassified variants. Fold-change was calculated using variants after disease-specific minor allele frequency filtering per variant set: Set-1 consists of variants with 2 or more ClinVar review stars (i.e. two or more submitters with assertion criteria, expert panel and practice guideline); Set-2 consists of variants with 1 star (i.e. one submitter with assertion criteria); Set-3 consists of variants with 0 star (i.e. submitter with no assertion criteria submitted in ClinVar); Set-4 consists of variants with conflicting interpretations of pathogenicity. A) The observed genetic risk was calculated using gnomAD exome data. B) The observed genetic risk was calculated using gnomAD genome data.

Figure S4: Genetic risk in Orphanet conditions using gnomAD



Fold-change of observed genetic risk over expected population prevalence using ClinVar variant sets for the Orphanet conditions. Each point represents a condition; each condition may be represented in more than one set. The navy-blue line at a fold-change of 10 (i.e. inflation) indicates a theoretical penetrance of 10%. Observations above this line are highly suggestive of misclassified variants. Fold-change was calculated using variants after disease-specific minor allele frequency filtering per variant set: Set-1 consists of variants with 2 or more ClinVar review stars (i.e. two or more submitters with assertion criteria, expert panel and practice guideline); Set-2 consists of variants with 1 star (i.e. one submitter with assertion criteria); Set-3 consists of variants with 0 star (i.e. submitter with no assertion criteria submitted in ClinVar); Set-4 consists of variants with conflicting interpretations of pathogenicity. A) The observed genetic risk was calculated using gnomAD exome data. B) The observed genetic risk was calculated using gnomAD genome data.

Figure S5: Number of ClinVar submitters for variants reclassified to conflicting interpretations of pathogenicity



For 855 P/LP, 2525 VUS, and 2487 B/LB variants that changed its classification to conflicting interpretations of pathogenicity from May 2016 version of ClinVar to September 2017 version, the plot shows the number of ClinVar submitters.

SUPPLEMENTAL TABLES

Table S1: ClinVar variant sets

A list of ClinVar variant sets that was used in the study. The chromosomal positions are in GRCh38 human reference built. Provided as a separate excel file.

Table S2: Genetic risk in ACMG-59 conditions.

Conditions	Estimated Population Prevalence	Mode of Inheritance	Genes	Observed Genetic risk	Fold Change	Observed Genetic Risk (dMAF)	Fold Change (dMAF)
Lynch Syndrome	227.27	Autosomal dominant	MLH1 MSH2 MSH6 PMS2	105	0.5	105	0.5
Familial hypercholesterolemia	500	Autosomal dominant	APOB LDLR PCSK9	314	0.6	314	0.6
Ehlers-Danlos syndrome, vascular type	1	Autosomal dominant	COL3A1	0	-	0	-
Familial adenomatous polyposis	3.2	Autosomal dominant	APC	0	-	0	-
Catecholaminergic polymorphic ventricular tachycardia	10	Autosomal dominant	RYR2	10	-	10	-
Ornithine transcarbamylase deficiency	7.14	X-linked recessive	OTC	10	-	10	-
Hypertrophic cardiomyopathy, Dilated cardiomyopathy	400	Autosomal dominant; X-linked recessive	MYBPC3 MYH7 TNNT2 TNNI3 TPM1 MYL3 ACTC1 PRKAG2 MYL2 LMNA GLA	581	1.5	562	1.4
WT1-related Wilms tumor	10	Autosomal dominant	WT1	19	1.9	19	1.9
Arrhythmogenic right ventricular cardiomyopathy	100	Autosomal dominant	DSC2 DSG2 DSP PKP2 TMEM43	295	3.0	229	2.3
Juvenile polyposis	6.25	Autosomal dominant	BMPR1A SMAD4	0	-	0	-
Hereditary Breast and Ovarian Cancer	250	Autosomal dominant	BRCA1 BRCA2	448	1.8	448	1.8

MYH-Associated Polyposis; Adenomas, multiple colorectal, FAP type 2; Colorectal adenomatous polyposis,autosomal recessive,with pilomatricomas	2.5	Autosomal recessive	MUTYH	10	-	10	-
Marfan Syndrome, Loeys-Dietz Syndromes, and Familial Thoracic Aortic Aneurysms and Dissections	20	Autosomal dominant	ACTA2 FBN1 MYH11 SMAD3 TGFBR1 TGFBR2	114	5.7	76	3.8
Romano-Ward Long QT Syndromes Types 1, 2, and 3, Brugada Syndrome	50	Autosomal dominant	KCNH2 KCNQ1 SCN5A	305	6.1	305	6.1
Multiple Endocrine Neoplasia Type 1	3.3	Autosomal dominant	MEN1	0	-	0	-
Li-Fraumeni Syndrome	7	Autosomal dominant	TP53	48	6.8	48	6.8
Retinoblastoma	6	Autosomal dominant	RB1	48	7.9	38	6.4
Neurofibromatosis type 2	1.78	Autosomal dominant	NF2	0	-	0	-
Familial Medullary Thyroid Cancer (FMTC);Multiple Endocrine Neoplasia Type 2	2.9	Autosomal dominant	RET	57	19.7	10	3.3
Peutz-Jeghers Syndrome	2.2	Autosomal dominant	STK11	0	-	0	-
PTEN Hamartoma Tumor Syndrome	0.5	Autosomal dominant	PTEN	0	-	0	-
Hereditary Paraganglioma-Pheochromocytoma Syndrome	0.3	Autosomal dominant	SDHAF2 SDHB SDHC SDHD	48	158.8	48	158.8
Tuberous Sclerosis Complex	17.2	Autosomal dominant	TSC1 TSC2	0	-	0	-
Von Hippel Lindau syndrome	2.3	Autosomal dominant	VHL	0	-	0	-
Wilson disease	10	Autosomal recessive	ATP7B	0	-	0	-
Malignant hyperthermia susceptibility	1	Autosomal dominant	CACNA1S RYR1	219	219.2	133	133.4

A list of ACMG-59 conditions with at least one P/LP variant from set-1, set-2 or set-3 observed in the study. The last two columns with “dMAF” suffix (“Observed Genetic risk (dMAF)” and “Fold Change (dMAF)”) are observed genetic risk and fold change calculated after filtering variants using disease-specific minor allele frequency (dMAF) filter. The population prevalence and genetic risk are calculated per 100,000. Fold change was not calculated if only one individual of the 10,495 samples was identified with genetic risk of the disease condition.

Table S3: Genetic risk in Orphanet conditions.

Orphanet ID	Condition	Estimated Population Prevalence	Mode of Inheritance	Genes	Observed Genetic risk	Fold Change	Observed Genetic Risk (dMAF)	Fold Change (dMAF)
48	Congenital bilateral absence of vas deferens	50	Autosomal recessive	CFTR	10	-	10	-
55	Oculocutaneous albinism	45	Autosomal recessive	OCA2	10	-	10	-
60	Alpha-1-antitrypsin deficiency	63.5	Autosomal recessive	SERPINA1	191	3.00	38	0.60
122	Birt-Hogg-Dubé syndrome	0.5	Autosomal dominant	FLCN	29	57.17	29	57.17
130	Brugada syndrome	75	Autosomal dominant	CACNA1C-AS1;CACNA1C;CACNA1C-AS2;CACNA1C;CACNA1C;CACNA2D1;CACNB2;GPD1L;KCNE3;NSUN6;CACNB2;SCN10A;SCN3B;SCN5A;TRPM4	38	0.51	38	0.51
145	Hereditary breast and ovarian cancer syndrome	250	Autosomal dominant	BRCA1;BRCA2	400	1.60	353	1.41
212	Cystathioninuria	7.1	Autosomal recessive	CTH	10	-	0	-
232	Sickle cell anemia	467.3	Autosomal recessive	HBB	19	0.04	19	0.04
268	Autosomal recessive limb-girdle muscular dystrophy type 2B	0.13	Autosomal recessive	DYSF	10	-	0	-
282	Frontotemporal dementia	3	Autosomal dominant	CHMP2B;POU1F1;MAPT	10	-	10	-
287	Ehlers-Danlos syndrome, classic type	5	Autosomal dominant	COL5A1;COL5A2;LOC101448202;COL5A1	10	-	10	-
324	Fabry disease	1.11	X-linked recessive	RPL36A-HNRNPH2;GLA	10	-	10	-
325	Congenital factor II deficiency	0.05	Autosomal recessive	F2	48	952.83	0	-
377	Gorlin syndrome	5.3	Autosomal dominant	LOC100507346;PTCH1;PTCH1;PTCH2;SUFU	19	3.60	0	-

429	Hypochondroplasia	3.3	Autosomal dominant	FGFR3	10	-	10	-
524	Li-Fraumeni syndrome	7	Autosomal dominant	TP53	29	4.08	29	4.08
558	Marfan syndrome	20	Autosomal dominant	FBN1;TGFBR2	29	1.43	19	0.95
565	Menkes disease	2.5	X-linked recessive	ATP7A	19	7.62	0	-
586	Cystic fibrosis	111	Autosomal recessive	CFTR	10	-	10	-
597	Central core disease	0.4	Autosomal dominant	RYR1	86	214.39	57	142.93
636	Neurofibromatosis type 1	50	Autosomal dominant	NF1	19	0.38	19	0.38
652	Multiple endocrine neoplasia type 1	3.3	Autosomal dominant	MEN1;RET	10	-	10	-
653	Multiple endocrine neoplasia type 2	2.9	Autosomal dominant	RET	57	19.71	10	-
661	Ondine syndrome	0.5	Autosomal dominant	BDNF-AS;BDNF;GDNF;RET	324	647.93	0	-
676	Hereditary chronic pancreatitis	0.57	Autosomal dominant	CFTR;CTRC;SPINK1	2820	4948.05	172	300.90
758	Pseudoxanthoma elasticum	2.5	Autosomal recessive	ABCC6	152	60.98	19	7.62
759	Central precocious puberty	20	Autosomal dominant	KISS1R	152	7.62	0	-
790	Retinoblastoma	6	Autosomal dominant	RB1	10	-	10	-
882	Tyrosinemia type 1	54	Autosomal recessive	FAH	76	1.41	0	-
1243	Best vitelliform macular dystrophy	20	Autosomal dominant	BEST1	19	0.95	19	0.95
2152	Mowat-Wilson syndrome	1.7	Autosomal dominant	ZEB2	10	-	10	-
2337	Non-epidermolytic palmoplantar keratoderma	2.5	Autosomal dominant	AQP5	19	7.62	19	7.62
2686	Cyclic neutropenia	0.1	Autosomal dominant	ELANE	10	-	10	-
3193	Supravalvular aortic stenosis	13.3	Autosomal dominant	ELN	10	-	0	-
32960	Tumor necrosis factor receptor 1 associated periodic syndrome	0.1	Autosomal dominant	TNFRSF1A	48	476.42	0	-
44890	Gastrointestinal stromal tumor	14.5	Autosomal dominant	KIT;PDGFRA;SDHB;SDHC	19	1.31	19	1.31
79241	Biotinidase deficiency	5	Autosomal recessive	BTD	10	-	10	-

79432	Oculocutaneous albinism type 2	46.15	Autosomal recessive	OCA2	10	-	10	-
98672	Autosomal dominant optic atrophy	83	Autosomal dominant	OPA1	10	-	10	-
98878	Hemophilia A	19.3	X-linked recessive	F8	10	-	10	-
98879	Hemophilia B	4	X-linked recessive	F9	10	-	10	-
100985	Autosomal dominant spastic paraplegia type 4	0.91	Autosomal dominant	SPAST	10	-	10	-
101016	Romano-Ward syndrome	40	Autosomal dominant	KCNQ1	29	0.71	29	0.71
182090	Pulmonary arterial hypertension	5.2	Autosomal dominant	BMPR2;ENG;LOC102723566;ENG;SMAD9	57	10.99	57	10.99

A list of Orphanet conditions with at least one P/LP variant from set-1, set-2 or set-3 observed in the study. The last two columns with "dMAF" suffix ("Observed Genetic risk (dMAF)" and "Fold Change (dMAF)") are observed genetic risk and fold change calculated after filtering variants using disease-specific minor allele frequency (dMAF) filter. The population prevalence and genetic risk are calculated per 100,000. Fold change was not calculated if only one individual of the 10,495 samples was identified with genetic risk of the disease condition.