Supporting Appendix for
"A Public Resource Facilitating Clinical Use of Genomes"

Madeleine P. Ball*, Joseph V. Thakuria*, Alexander Wait Zaranek*, Tom Clegg, Abraham M. Rosenbaum, Xiaodi Wu, Misha Angrist, Jong Bhak, Jason Bobe, Matthew J. Callow, Carlos Cano, Michael F. Chou, Wendy K. Chung, Shawn M . Douglas, Preston W. Estep III, Athurva Gore, Peter Hulick, Alberto Labarga, Je-Hyuk Lee, Jeantine Lunshof, Byung Chul Kim, Jong-Il Kim, Zhe Li, Michael F. Murray, Geoffrey B. Nilsen, Brock A. Peters, Anugraha M. Raman, Hugh Y. Rienhoff, Kimberly Robasky, Matthew T. Wheeler, Ward Vandewege, Dan Vorhaus, Joyce L. Yang, Luhan Yang, John Aach, Euan A. Ashley, Radoje Drmanac, Seong-Jin Kim, Jin Billy Li, Leonid Peshkin, Christine E. Seidman, Jeong-Sun Seo, Kun Zhang, Heidi L. Rehm, George M. Church

* These authors contributed equally

Index

Item	Page
Supporting Materials and Methods	S2
Dataset S1: PGP Participants and associated health conditions (description only; file provided separately)	S6
Figure S1: PGP account status	S7
Figure S2: Demographics of PGP participants	S8
Figure S3: Histogram of call rates, split by call status in first genome	S9
Figure S4: Venn diagrams of shared calls, split by variant type	S10
Figure S5: MYL2-A13T pedigrees	S12
Table S1: PGP-10 participants and associated cell lines	S14
Table S2: PGP-10 genome data statistics	S15
Table S3: Prioritization score calculation	S16
Table S4: Variants reported or predicted to have significant phenotypic consequences	S17
Table S5: Heterozygous variants with reported or potential severe recessive effects	S18
Table S6: PGP-10 trait questionnaire and responses	S19
Table S7: GET-Evidence: Variant evidence and clinical importance scoring	S21
Table S8: GET-Evidence: Assessment of strength of evidence	S22
Table S9: GET-Evidence: Assessment of clinical importance	S23
Table S10: GET-Evidence information regarding variants from Table S5	S24

Supporting Materials and Methods

Enrollment process and open consent

Pre-enrollment screening requires that volunteers (1) be at least 21 years of age, (2) be a citizen or permanent resident of the United States, and (3) not be subject to undue influence or coercion by the Principle Investigator of this study. Volunteers with a monozygotic twin must also have their twin complete enrollment to be eligible. Individuals are asked to name two designated proxies who may be contacted in the event of death or incapacitation.

An ongoing relationship with participants exists, allowing recontact and continuing follow-up. To monitor the project for potential negative outcomes, participants are required to respond to private quarterly questionnaires where they are asked to report changes in safety, well-being, or interactions with others due to their participation in the project. If a participant has not updated their safety questionnaire at least three times in the past twelve months their account is considered lapsed, and they must update the safety questionnaire before adding data or modifying their account. Participants are invited at any point to withdraw from the project (although we do not guarantee that data, which has been public, is removable from all sources). If this recontact process fails, designated proxies are contacted, where available, to determine whether the participant is deceased or incapacitated -- depending on the decision made by the proxies, the account may be closed and removed, or updated and remain public.

Genetic data, when produced by the PGP, is initially provided privately to participants along with our preliminary interpretation and becomes public after 30 days. All genome and other public participant data are linked to the participant ID and published in the public domain under the Creative Commons CC0 waiver (1). Current and historical copies of our consent forms are provided publicly at http://www.personalgenomes.org/consent/. This enables reuse or customization, e.g. for international use. The enrollment exam is focused to assess understanding of PGP protocols, risks and benefits as well as a basic understanding of genetics as it pertains to informing family members.

CCR-formatted health record data

Health record data to date has been collected using interfaces with Google Health and Microsoft HealthVault to import Continuity of Care Record (CCR) format health record data. Specific data regarding conditions, medications, and procedures are pulled from these data and are published publicly on participant profiles. SI Dataset S1 was constructed using data from 1,021 recent Google Health records for participants, downloaded on October 27, 2011. Health condition descriptions and codes were parsed from the records and matched to PGP participant ID. Entries were pooled based on their ICD9 code or, if not present, by Google code.

Cell lines, samples, genome sequencing and quality assessment

EBV-transformed lymphocyte cell lines were derived from whole blood and fibroblast cell lines were derived from 3mm skin punch biopsies by the Harvard Medical School Cytogenetic Core Facility. Induced pluripotent cell lines were derived from these fibroblast cell lines according to the method described in Lee et al. (2). DNA was extracted from these cell lines and sent to Complete Genomics for whole genome sequencing.

Prioritization score assessment using disease-specific mutation databases

Genetic variant lists were downloaded from five publicly available disease-specific databases (all September 2011): the Albinism Database (3), the ALS Online Genetics Database (4), the Cardiogenomics Sarcomere Protein Gene Mutation Database (5), the Connexins and deafness Homepage (6), and the Autosomal Dominant Polycystic Kidney Disease Mutation Database (7). Where appropriate, amino acid numbering in databases was adjusted to match positions predicted by Polyphen 2 and GET-Evidence's genome analysis, which both use the canonical transcripts in the UCSC Known Genes annotation.

GET-Evidence data processing and editing platform

Imported databases were downloaded, parsed, and entered into our MySQL database and are used for variant prioritization. Databases used in our analyses include GeneTests (4,253 genes, January 2010), Online Mendelian Inheritance in Man (OMIM) (9,142 nonsynonymous substitutions extracted through custom scripts from June 2009), HuGENavigator (2,298 dbSNP IDs, January 2010), and PharmGKB (2,488 dbSNP IDs, April 2010). In addition, all nonsynonymous predictions found in the PGP genomes are entered into our database, which is then regularly updated with Polyphen 2 predictions based on the whole human proteome sequence space (downloaded July 2011). We also used web-search results, generated by the Yahoo API, to help find additional literature. Web hits are first generated when a variant is added to the database and are periodically updated. These data sources provide us with a rich set of disease, pharmacogenetic and literature predictions, and are available in the public domain or under non-commercial terms.

Genome data processing was written in Python, performed with a series of modules and originally developed as the Trait-o-matic software used in previous genome publications (8–11) and tested on other public genomes (12–21).

Processing steps involve:

1. Genome data file format is automatically detected and, if it is in CGI var file format, VCF, or 23andme genotyping data format, it is automatically translated to the GFF file format used internally by the system. Regions which are only partially called (hemizygously no-call) in the CGI var file are treated as a no-call region. Interpretation of both build 36 and build 37 genome data is supported.
2. IDs for dbSNP locations are added to variants, if not already present, based on matching positions in the latest dbSNP data.
3. Nonsynonymous amino acid change predictions are made using the knownCanonical transcripts listed in the UCSC Known Genes transcript annotations (22). This is done by predicting the variant and reference transcript nucleotide sequences, predicting amino acid sequences from these, then detecting if differences occur between variant and reference versions. Nonsynonymous predictions include multiple base substitutions and frameshift and in-frame length changing variants, in addition to single amino acid substitutions and nonsense mutations.
4. Nonsynonymous changes and dbSNP IDs are searched against a pre-loaded list of existing GET-Evidence entries (which is populated by the databases mentioned previously, previously seen public variants, and any manually added entries). If no GET-Evidence match is found, prioritization score is predicted based on existing computational and gene-specific data.
5. All nonsynonymous variants, or other variants matching existing GET-Evidence entries, are reported in GET-Evidence reports.

High variant frequency in the general population is useful for interpretation as it generally indicates that a variant is unlikely to have severe clinical consequences. To acquire this data, we downloaded and extracted allele frequency information from Exome Variant Server and 1000 Genomes data. In addition, we calculated frequencies for all variants in our combined set of 64 genomes (data for the PGP-10 and the 54 unrelated public genomes released by Complete Genomics). Exome Variant Server data was downloaded May 2012, using the ESP5400 data released in December 2011 (23). 1000 Genomes data was downloaded May 2012, using the phase 1 integrated release version 3 data released in April 2012 (24).

Users are identified in GET-Evidence using OpenID (25). All user edits are stored in the GET-Evidence MySQL database in a manner that retains records of all previous edits made to the page. These data are released under a CC0 license at http://evidence.personalgenomes.org/download (26) and the GET-Evidence software is released under the GNU AGPL version 3.0 or any later version (27).

Annotation of article abstracts was implemented using a standalone installation of BioNotate (28). When a user clicks the "annotate" button, GET-Evidence passes the variant name and the relevant article's PubMed identifier to the BioNotate instance, along with a user-identifying token. BioNotate retrieves the article abstract using NCBI's

API and allows the user to indicate the significance of words and phrases within the abstract. Upon clicking "submit", the user returns to the GET-Evidence site. GET-Evidence retrieves the newly annotated abstract from BioNotate in XML format and saves it in the edit history for the relevant variant page.

References

1. CC0 1.0 Universal (CC0 1.0) Public Domain Dedication *Creative Commons*. Available at: http://creativecommons.org/publicdomain/zero/1.0/ [Accessed December 2, 2011].

2. Lee J-H et al. (2009) A Robust Approach to Identifying Tissue-Specific Gene Expression Regulatory Variants Using Personalized Human Induced Pluripotent Stem Cells. *PLoS Genet* 5:e1000718.

3. Albinism Database *downloaded September 2011*. Available at: http://albinismdb.med.umn.edu/ [Accessed December 1, 2011].

4. Lill CM, Abel O, Bertram L, Al-Chalabi A (2011) Keeping up with genetic discoveries in amyotrophic lateral sclerosis: the ALSoD and ALSGene databases. *Amyotroph Lateral Scler* 12:238–249.

5. Genomics of Cardiovascular Development, Adaptation, and Remodeling. *NHLBI Program for Genomic Applications, Harvard Medical School Accessed September 2011*. Available at: http://www.cardiogenomics.org [Accessed May 26, 2010].

6. Ballana E, Ventayol M, Rabionet R, Gasparini P, Estivill X Connexins and deafness Hopepage. *downloaded September 2011*. Available at: http://davinci.crg.es/deafness/ [Accessed December 1, 2011].

7. The Autosomal Dominant Polycystic Kidney Disease Mutation Database *downloaded September 2011*. Available at: http://pkdb.mayo.edu/ [Accessed December 1, 2011].

8. Drmanac R et al. (2010) Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays. *Science* 327:78–81.

9. Kim J-I et al. (2009) A highly annotated whole-genome sequence of a Korean individual. *Nature* 460:1011–1015.

10. Ashley EA et al. (2010) Clinical assessment incorporating a personal genome. *The Lancet* 375:1525–1535.

11. Dewey FE et al. (2011) Phased Whole-Genome Genetic Risk in a Family Quartet Using a Major Allele Reference Sequence. *PLoS Genet* 7:e1002280.

12. Levy S et al. (2007) The Diploid Genome Sequence of an Individual Human. *PLoS Biol* 5:e254.

13. Wheeler DA et al. (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452:872–876.

14. Ng PC et al. (2008) Genetic Variation in an Individual Human Exome. *PLoS Genet* 4:e1000160.

15. Bentley DR et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59.

16. Wang J et al. (2008) The diploid genome sequence of an Asian individual. *Nature* 456:60–65.

17. Ahn S-M et al. (2009) The first Korean genome sequence and analysis: Full genome sequencing for a socio-ethnic group. *Genome Res* 19:1622 –1629.

18. McKernan KJ et al. (2009) Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res* 19:1527 –1541.

19. Pushkarev D, Neff NF, Quake SR (2009) Single-molecule sequencing of an individual human genome. *Nat Biotechnol* 27:847–850.

20. Ng SB et al. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461:272–276.

21. Schuster SC et al. (2010) Complete Khoisan and Bantu genomes from southern Africa. *Nature* 463:943–947.

22. Hsu F et al. (2006) The UCSC Known Genes. *Bioinformatics* 22:1036 –1046.

23. Exome Variant Server, NHLBI Exome Sequencing Project (ESP), Seattle, WA. Available at: http://evs.gs.washington.edu/EVS/ [Accessed May 15, 2012].

24. Consortium T1000 GP (2010) A map of human genome variation from population-scale sequencing. *Nature* 467:1061–1073.

25. Recordon D, Reed D (2006) in *Proceedings of the second ACM workshop on Digital identity management*, DIM '06. (ACM, New York, NY, USA), pp 11–16. Available at: http://doi.acm.org/10.1145/1179529.1179532 [Accessed December 14, 2011].

26. CC0 1.0 Universal (CC0 1.0) Public Domain Dedication *Creative Commons*. Available at: http://creativecommons.org/publicdomain/zero/1.0/ [Accessed December 2, 2011].

27. GNU Affero General Public License Available at: http://www.gnu.org/licenses/agpl.html [Accessed December 14, 2011].

28. Cano C, Monaghan T, Blanco A, Wall DP, Peshkin L (2009) Collaborative text-annotation resource for disease-centered relation extraction from biomedical text. *J Biomed Inform* 42:967–977.

Dataset S1: PGP participants and associated health conditions

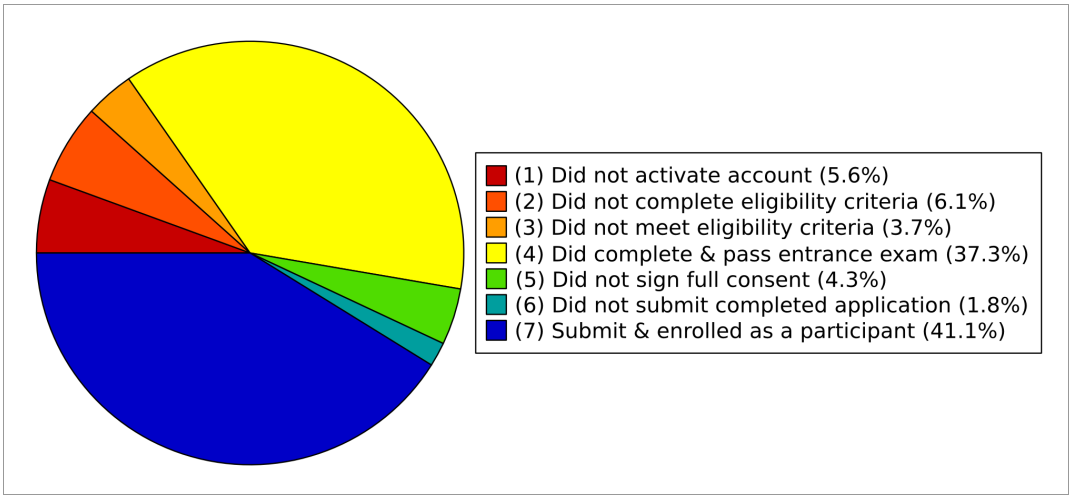
Description

Column 1: Total number of participant health records reporting this condition
Column 2: Description of condition
Column 3: ICD9 code, if available
Column 4: Google code(s) matching ICD9 code
Column 5: Participant IDs

Data from 1,021 recent Google Health records for participants was downloaded on October 27 2011. Health condition descriptions and codes were parsed from the records and matched to PGP participant ID. Entries were pooled based on their ICD9 code or, if not present, by Google code.

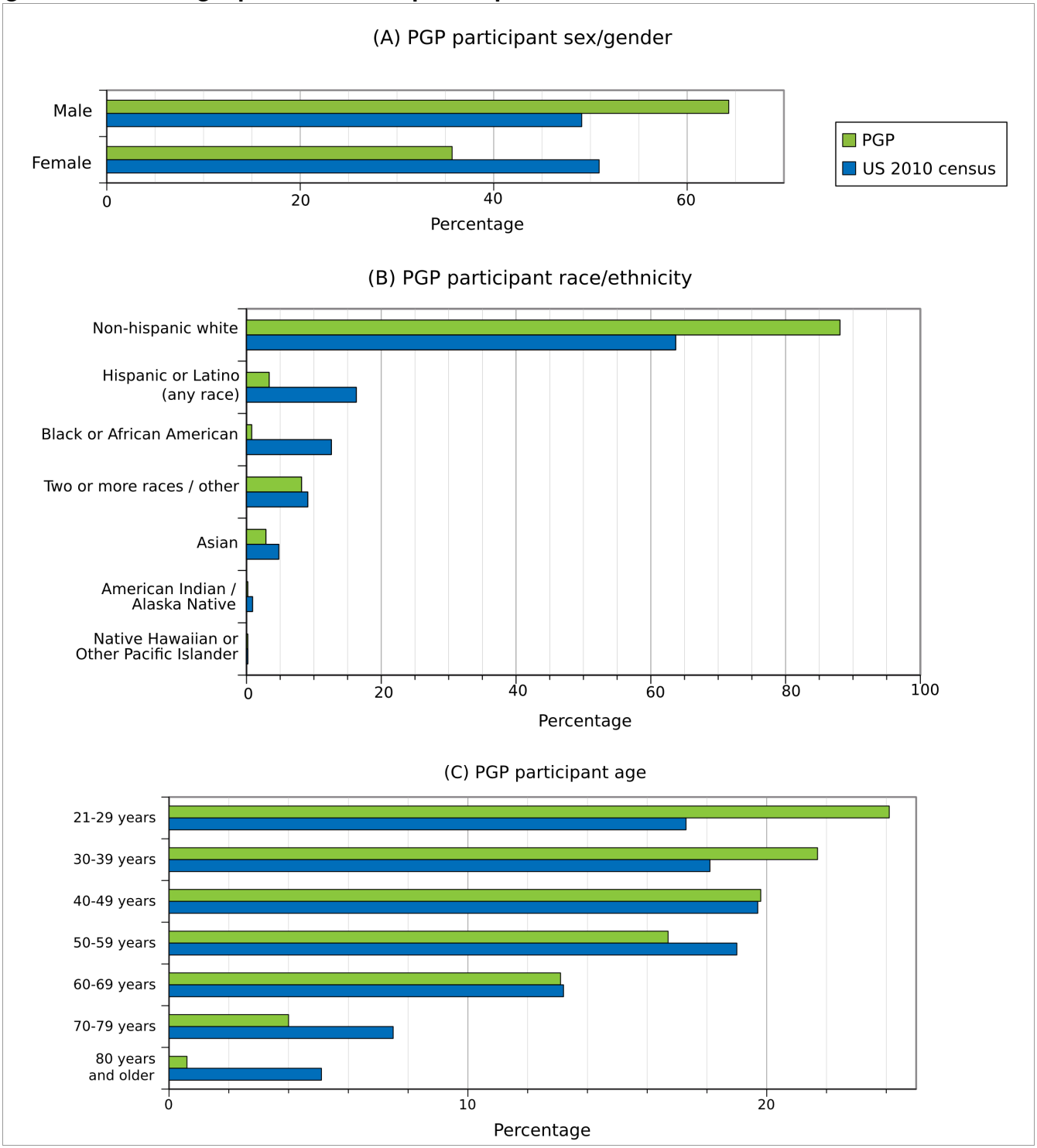
Please see supporting files for this table.

Figure S1: PGP account status



In an analysis of accounts created through the Personal Genome Project website, 41% of users complete all enrollment steps to become fully enrolled participants. Of those that do not complete all steps, the largest fraction (37%) fail to complete and pass the entrance exam. The entrance exam asks participants to demonstrate an understanding of basic genetics concepts and of the risks and potential outcomes that may result from publicly donating genome data and tissue samples. These statistics were calculated from 1,138 user log records spanning December 2010 to December 2011.

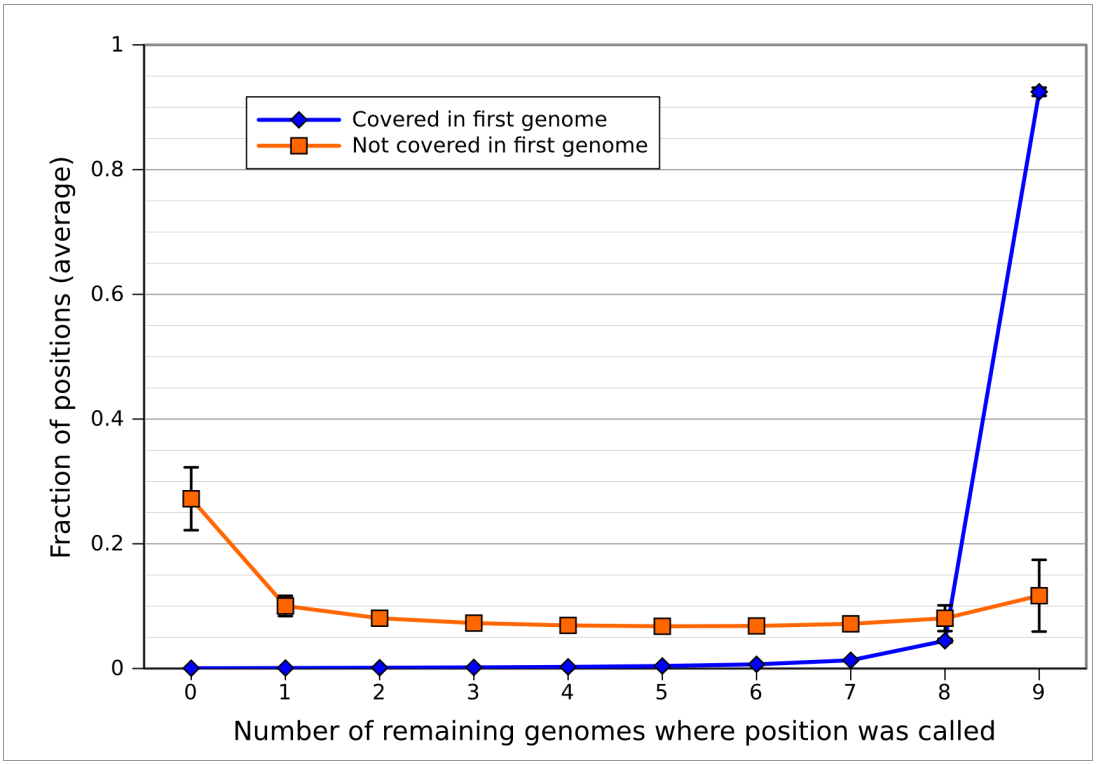
Figure S2: Demographics of PGP participants



Because participation is self-selecting and occurs through an online enrollment mechanism, the demographics of PGP participants is expected to differ from the United States population (2010 census data). Based on self-reported survey data from over 1000 participants, **(A)** Males are overrepresented and females are underrepresented, **(B)** Non-Hispanic whites are over-represented and minority groups are underrepresented, **(C)** Younger ages are overrepresented and older ages are underrepresented.

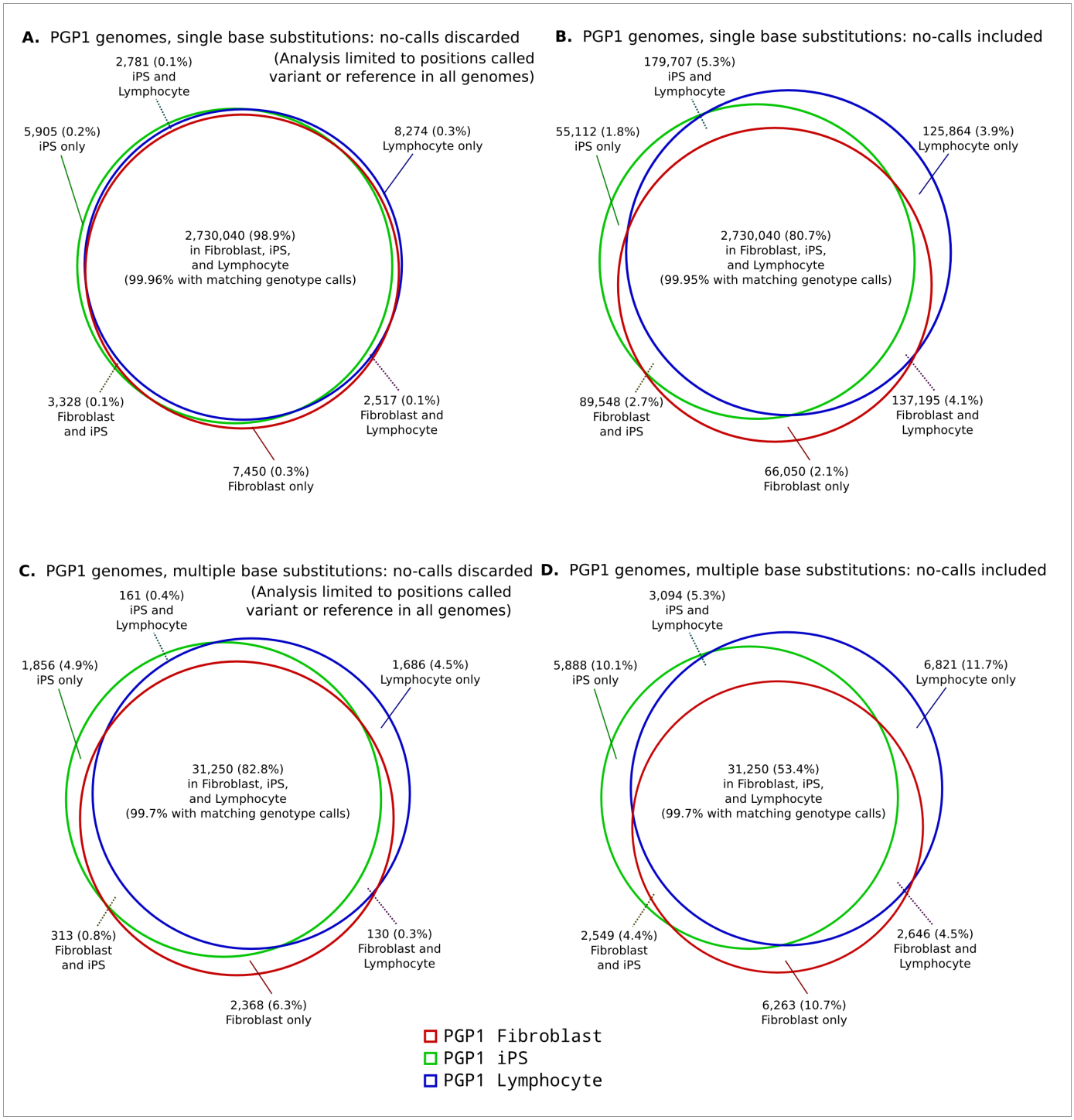
Note: For US census data, Figure 2C reports the percentages of the population within the 21 and older age range, which are the ages eligible for PGP enrollment.

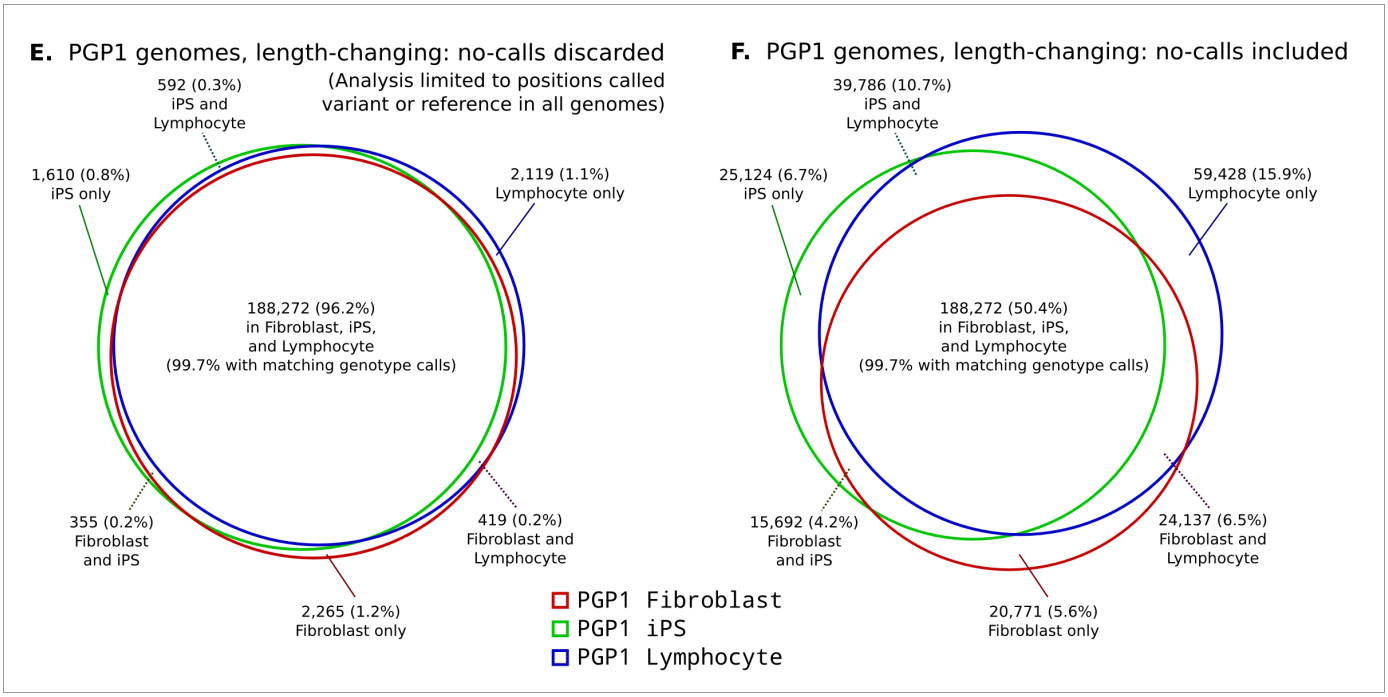
Figure S3: Histogram of call rates, split by call status in first genome



A histogram of how many of the remaining nine genomes a position was called in, split by its call status in the first genome examined. Positions which have their genotype called are highly correlated between genome data. Sites which were called in a given genome were much more likely to be called in all other genomes (blue line) – on average 92% of positions that were called in one genome were also called in the remaining nine genomes. Similarly, sites not called in a given genome were more likely to be not called in other genomes (orange line) – on average 27% of positions not called in one genome were also not called in the remaining nine genomes. Data represent the average when each of the ten genomes is used as the “first”; error bars are the standard deviation of this data.

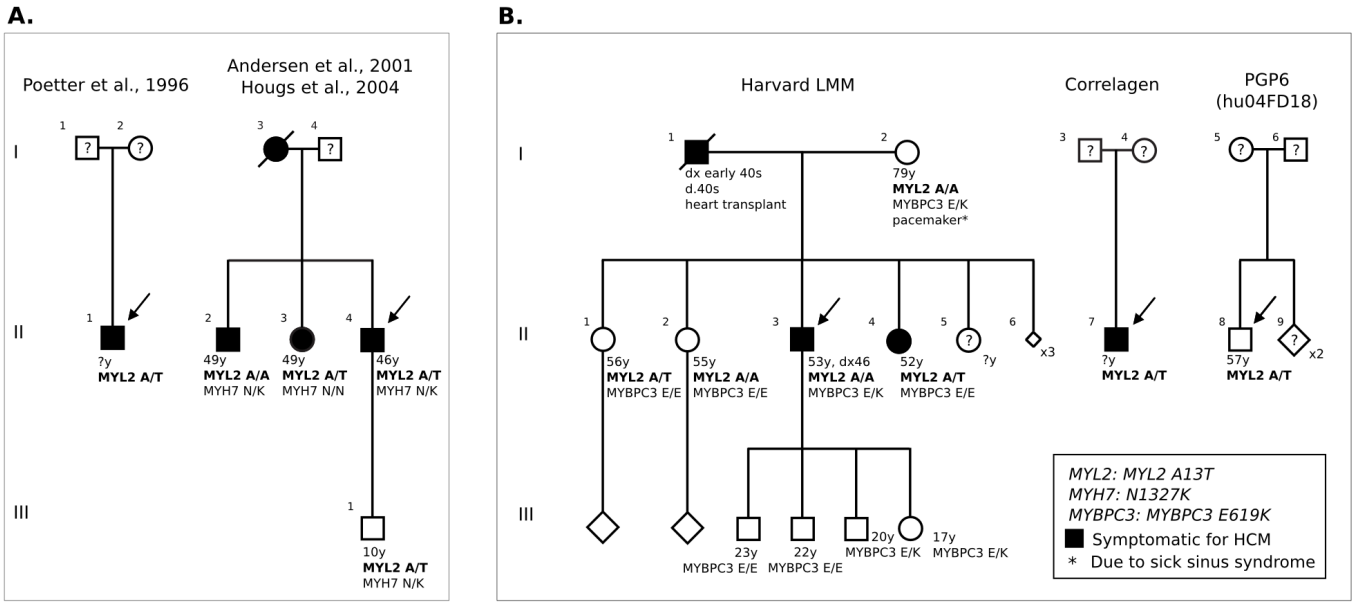
Figure S4: Venn diagrams of shared calls, split by variant type





Analysis of PGP1 genome variant calls from three different tissues (as in Figure 2), split by variant type. **(A)** Overlap of all single-base substitution variant calls, limited to positions explicitly called as variant or reference in all three genomes. On average, 99.7% called variant in one genome are called variant in at least two out of three. **(B)** Overlap of all single-base substitution variant calls, including uncalled positions lacking explicit reference or variant calls. **(C)** Overlap of all multi-base substitution variant calls, limited to positions explicitly called as variant or reference in all three genomes. On average, 94.1% called variant in one genome are called variant in at least two out of three. **(D)** Overlap of all multi-base substitution variant calls, including uncalled positions lacking explicit reference or variant calls. **(E)** Overlap of all length-changing variant calls, limited to positions explicitly called as variant or reference in all three genomes. On average, 99.0% called variant in one genome are called variant in at least two out of three. **(F)** Overlap of all length-changing variant calls, including uncalled positions lacking explicit reference or variant calls.

Figure S5: MYL2-A13T pedigrees



(A) MYL2-A13T has been implicated in causing hypertrophic cardiomyopathy (HCM) in a dominant fashion. This variant was initially reported in a study which implicated nonsynonymous variants in MYL2 and MYL3 (myosin essential and regulatory light chains) as causing HCM (1). The A13T variant was seen in one of four HCM cases with nonsynonymous variants in MYL2 in a screen of 399 unrelated cases. All four variants were reported to have strong evolutionary conservation. The association of this variant with HCM was later reported on in a case with two out of three affected siblings—the third sibling was initially thought to be a phenocopy due to concurrent obesity and hypertension (2) and then later suspected to have disease due to another variant identified in the MYH7 gene (3). Functional studies also reported that the product of MYL2-A13T bound calcium significantly differently to wild-type (4).

(B) To check for additional unpublished clinical data, we contacted all four laboratories in the United States (Harvard-Partners Laboratory for Molecular Medicine (LMM), Correlagen, GeneDx, PGxHealth) offering CLIA–approved diagnostic sequencing of MYL2 for cardiomyopathy. Only two had observed this variant. Correlagen reported finding the variant in one patient with HCM , though no further clinical or family history data was available. The LMM studied a family with two siblings and their father with HCM; the variant was found in only one of the two affected siblings (although the father was deceased in his early 40's from cardiomyopathy treated with heart transplant, the absence of the variant in the mother indicates that the father was likely positive). Another variant, MYBPC3 Glu619Lys, was initially considered causal in the other sibling, but their 79-year-old mother (who also carries this variant) had a normal echocardiogram.

Additionally, we noted that PGP6 is Ashekenazi Jewish (AJ). The LMM family is also AJ and, when contacted, P. Andersen reported that the Andersen/Hougs pedigree was AJ. This raises the possibility that the variant is a polymorphism within the AJ population. To test this we screened an AJ DNA panel and did not detect the variant in any of the 116 controls individuals we examined.

References

1. Poetter K et al. (1996) Mutations in either the essential or regulatory light chains of myosin are associated with a rare myopathy in human heart and skeletal muscle. *Nat Genet* 13:63-69.

2. Andersen PS et al. (2001) Myosin light chain mutations in familial hypertrophic cardiomyopathy: phenotypic presentation and frequency in Danish and South African populations. *Journal of Medical Genetics* 38::e43.

3. Hougs L et al. (2004) One third of Danish hypertrophic cardiomyopathy patients have mutations in MYH7 rod region. *Eur J Hum Genet* 13:161-165.

4. Szczesna D et al. (2001) Familial Hypertrophic Cardiomyopathy Mutations in the Regulatory Light Chains of Myosin Affect Their Structure, Ca2+Binding, and Phosphorylation. *Journal of Biological Chemistry* 276:7086 -7092.

Table S1: PGP-10 participants and associated cell lines

Participant ID	PGP Nickname	Full name	Coriell repository ID	Cell type
hu43860C	PGP1	George M. Church	GM20431	EBV-transformed lymphocyte
			GM23248	Fibroblast
huC30901	PGP2	John D. Halamka	GM21070	EBV-transformed lymphocyte
huBEDA0B	PGP3	Esther Dyson	GM21660	EBV-transformed lymphocyte
huE80E3D	PGP4	Misha Angrist	GM21667	EBV-transfomed lymphocyte
			GM23249	Fibroblast
hu9385BA	PGP5	Kirk Michael Maxey	GM21687	EBV-transformed lymphocyte
			GM23250	Fibroblast
hu04FD18	PGP6	Steven Pinker	GM21730	EBV-transformed lymphocyte
hu0D879F	PGP7	Keith F. Batchelder	GM21731	EBV-transformed lymphocyte
huAE6220	PGP8	Stanley N. Lapidus	GM21781	EBV-transformed lymphocyte
hu034DB1	PGP9	Rosalynn D. Gill	GM21833	EBV-transformed lymphocyte
			GM23251	Fibroblast
hu604D39	PGP10	James Louis Sherley	GM21846	EBV-transformed lymphocyte

Table S2: PGP-10 genome data statistics

Individual (huID)	Ungapped build 37 locations called homozygously	# of substitution variants vs. build 37 reference	# of short insertions and deletions vs. build 37 reference
PGP1 (hu43860C)	96.7%	3,216,092	310,621
PGP2 (huC30901)	96.5%	3,212,647	315,289
PGP3 (huBEDA0B)	95.9%	3,082,457	272,251
PGP4 (huE80E3D)	96.0%	3,148,580	277,661
PGP5 (hu9385BA)	96.8%	3,259,173	321,731
PGP6 (hu04FD18)	96.1%	3,161,062	279,947
PGP7 (hu0D879F)	97.2%	3,318,280	352,538
PGP8 (huAE6220)	97.2%	3,328,192	347,134
PGP9 (hu034DB1)	95.5%	3,057,821	257,667
PGP10 (hu604D39)	95.2%	3,611,748	284,696
Average:	**96.5%**	**3,239,605**	**301,954**

Ungapped build lengths were taken from:
http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/data/

Table S3: Prioritization score calculation

Each variant can score a maximum of two points in each category, for a total of up to six points.

Category	Points	Criteria
Computational	2 points	* Variant has an allele frequency of < 5% and Polyphen 2 predicts "probably damaging" (score > 0.85) * Variant has an allele frequency of < 5% and is predicted to have a severely disruptive effect on protein sequence (nonsense or frameshift mutation)
	1 point	* Variant is predicted to be "probably damaging" by Polyphen 2, but allele frequency is >= 5% * Variant is predicted to cause a frameshift or nonsense mutation, but allele frequency is >= 5% * Variant is nonsynonymous, has an allele frequency < 5%, and Polyphen 2 score is unknown or predicted to be "possibly damaging"
	0 points	* Variant is synonymous * Variant is nonsynonymous, but does not meet above criteria
Variant-specific databases	2 points	* Variant is seen in OMIM * Variant is seen in any two of the following lists: -- PharmGKB -- HuGENet -- confirmed or unevaluated online web page hits
	1 point	* Variant is seen in any one of the following lists: -- PharmGKB -- HuGENet -- confirmed or unevaluated online web page hits
	0 points	* Variant is not matched to any databases and has no confirmed or unevaluated online web page hits.
Gene-specific databases	2 points	* Variant is nonsynonymous and occurs in a gene with clinical testing available (as recorded by the GeneTests database) and an associated GeneReviews article.
	1 point	* Variant is nonsynonmyous and occurs in a gene with clinical testing available
	0 points	* Variant is synonymous or occurs within a gene which is not clinically tested.

Table S4:
Variants reported or predicted to have significant phenotypic consequences

Participant	**Variant** (heterozygous unless otherwise noted)	**Predicted phenotype**	**Supporting publications** (PMIDs)	**Confirmed by participant phenotype?**
PGP1 (hu43860C)	SERPINA1-E366K/ SERPINA1-E288V (compound heterozygous)	Moderate alpha-1 antitrypsin deficiency (increased susceptibility to liver and lung disease – the latter generally for emphysema rather than infections)	6976856, 8970361, 18565211	Participant reports frequent lung infections, but no diagnosis of COPD and no history of smoking.*
PGP1 (hu43860C)	WFS1-C426Y	Familial depression**	11244483	Participant reports mild sypmtoms.*
PGP2 (huC30901)	FLG-S761fs	Palmar hyperlinearity and keratosis pilaris	None variant specific. [11244483] predicts mild phenotype in carriers of a similar Ichthyosis Vulgaris mutation.	Participant reports not having this phenotype.
PGP5 (hu9385BA)	PKD1-R4276W	Autosomal dominant polycystic kidney disease	10200984	Participant reports no personal or family history of this disease or associated symptoms.
PGP6 (hu04FD18)	MYL2-A13T	Hypertrophic cardiomyopathy	8673105, 11102452, 11748309, 12668451, 14594949	Participant evaluated, echocardiogram was normal. Family history is ambiguous (parents healthy, but siblings of both parents had early mortality attributed to cardiac disease).
PGP9 (hu034DB1)	SCN5A-G615E	Long-QT Syndrome	11997281, 15840476, 18071069, 19716085	Participant reports no personal or family history of this disease. An unrelated EKG examination in 2010 produced normal results.
PGP10 (hu604D39)	PKD2-S804N	Autosomal dominant polycystic kidney disease	17582161	Participant reports no personal or family history of this disease or associated symptoms.
PGP10 (hu604D39)	SLC9A3R1-R153Q	Kidney stones	18784102	Participant reports no personal or family history of these symptoms.
PGP10 (hu604D39)	RHO-G51A	Autosomal dominant retinitis pigmentosa	8317502	Participant reports no personal or family history of this disease and associated symptoms.
PGP10 (hu604D39)	EVC-R443Q***	Ellis-van Creveld syndrome or related symptoms	10700184	Participant phenotype not consistent with this prediction.

* Participant reported traits after return of personal results reporting the genetic trait and putative associated phenotype.
** Although most reports for WFS1 involve it causing Wolfram syndrome in a recessive manner, this publication suggested that rare substitution variants in the gene carried heterozygously (including the one listed here) may be associated with increased risk for psychiatric disease.
*** Although Ellis-van Creveld syndrome is generally recessive, this publication reported a father–daughter pair with symptoms similar to Ellis-van Creveld syndrome. Both were heterozygous for this variant, implying that it was acting in a dominant manner.

Table S5:
Heterozygous variants with reported or potential severe recessive effects

Participant	Variant	Predicted phenotype (a "?" denotes a variant with no supporting published findings)	Supporting publications (Listed as PMID when available. Only variant-specific publications are noted)	Computational evidence
PGP1 (hu43860C)	RYR2-G1885E	Arrhythmogenic right ventricular cardiomyopathy (when compound het. with G1886S)	16769042, 16769042	Rare*, PPH2: unknown, BLOSUM100 predicts Gly to Glu is disruptive.
PGP1 (hu43860C)	FIG4-K278fs	Charcot-Marie-Tooth Disease Type 4J?	None	Rare*, Frameshift
PGP2 (huC30901)	RP1-T373I	Retinitis pigmentosa	11095597, 11095597	PPH2: Prob damaging
PGP2 (huC30901)	FLG-S761fs	Ichthyosis vulgaris?	None	Rare*, Frameshift
PGP2 (huC30901)	TGM1-Y312fs	Congenital ichthyosis?	None	Rare*, Frameshift
PGP2 (huC30901)	SLC4A1-E40K	Hemolytic anemia	8471774	Rare*, PPH2: Benign
PGP3 (huBEDA0B)	ABCA4-P1780A	Stargardt disease	10746567	Rare*, PPH2: Prob damaging
PGP3 (huBEDA0B)	LRP5-V667M	Osteoporosis-pseudoglioma syndrome	11719191	Rare*, PPH2: Prob damaging
PGP4 (huE80E3D)	SPG11-K1013E	Spastic paraplegia	Mention in meeting abstract: Boukhris et al., 19th Meeting of the European Neurological Society (2009)	Rare*, PPH2: unknown
PGP4 (huE80E3D)	CPT2-S113L	Late-onset carnitine palmitoyltransferase deficiency	8358442	Rare*, PPH2: Prob damaging
PGP4 (huE80E3D)	SLC6A5-T425M	Hyperekplexia	16751771	Rare*, PPH2: Prob damaging
PGP5 (hu9385BA)	SERPINA1-R247C	Alpha-1 antitrypsin deficiency	CCHMC Molecular Genetics Laboratory Mutation DB (online)	Rare*, PPH2: Prob damaging
PGP5 (hu9385BA)	CBS-T460M	Homocystinuria	Mention in meeting abstract: Redonnet-Vernhet et al., Annual Symposium of the Society for the Study of Inborn Errors of Metabolism (2010)	Rare*, PPH2: Prob damaging
PGP5 (hu9385BA)	ACADVL-R385W	Very long chain acyl-coenzyme A dehydrogenase deficiency	CCHMC Molecular Genetics Laboratory Mutation Database (online)	Rare*, PPH2: Prob damaging
PGP5 (hu9385BA)	TGM1-E520G	Lamellar ichthyosis	11348475, 19241467	Rare*, PPH2: Prob damaging
PGP5 (hu9385BA)	SLC7A9-A182T	Cystinuria	10471498, 11157794	Rare*, PPH2: Benign
PGP5 (hu9385BA)	SLX4-G1396fs	Fanconi Anemia (complementation group P)?	None	Rare*, Frameshift
PGP10 (hu604D39)	SPG7-G199fs	Hereditary spastic paraplegia?	None	Rare*, Frameshift
PGP10 (hu604D39)	SLC26A4-I300L	Pendred Syndrome	http://www.healthcare.uiowa.edu/labs/pendredandbor/slcMutations.htm	Rare*, PPH2: Prob damaging
PGP2 & PGP10	PEX1-I696M	Peroxisome biogenesis disorders	11389485	PPH2: Benign

* Rare variants were only seen once in 64 PGP10 + public CGI genomes

Table S6: PGP-10 trait questionnaire and responses

Predicted amino acid change: Potential phenotype (dominance) [individual w/ variant]	Prioritization reasons (PMIDs are listed for published findings)	Question(s)	Participant responses
SLC9A3R1-R153Q: Kidney stones (dom) [PGP10 het]	Rare*, PPH2: prob damaging, OMIM, Published findings (18784102)	Have you ever had kidney stones? Have any of your first degree relatives (parents, siblings, or children) had kidney stones?	All PGP10 reported "no" to the first question. For the second question: PGP1 reported gallstones in his mother, all others (including PGP10) reported "no".
PKD1-G3300R: Polycystic kidney disease (dom) [PGP7: het]	Rare*, PPH2: prob damaging, Clinically tested gene	Have you or a relative been diagnosed with polycystic kidney disease?	All PGP10 reported "no".
PKD2-S804N: Polycystic kidney disease (dom) [PGP10: het]	Rare*, PPH2: prob damaging, Published findings (17582161, 20881056), Clinically tested gene	Have you or a relative been diagnosed with polycystic kidney disease?	All PGP10 reported "no".
AARS-K967M: Charcot-Marie Neuropathy (dom) [PGP4: het]	Rare*, PPH2: prob damaging, Clinically tested gene	Have you or a relative been diagnosed with Charcot-Marie Neuropathy?	All PGP10 reported "no".
MYO1A-S797F: Hearing loss (dom) [PGP1: het]	Rare*, PPH2: poss damaging, Published findings (12736868)	Do you have profound hearing loss/deafness or use hearing aids?	All PGP10 reported "no".
SCN5A-G615E: Long QT syndrome (dom) [PGP9: het]	Rare*, Published findings (11997281, 15840476, 18071069, 19716085, 20486126), Clinically tested gene	Have you or a relative been diagnosed with long-QT syndrome? Do you have a relative who has died suddenly due to cardiac failure at an unusually young age?	For the first question: PGP3 reported she may have an affected relative, all others (including PGP9) reported "no". For the second question: PGP5 reported that his paternal grandfather died of what was believed to be MI at the age of 58, while PGP4 and PGP6 also reported "yes". All others (including PGP9) reported "no".
MYL2-A13T: Hypertrophic cardiomyopathy (dom) [PGP6: het]	Rare*, OMIM, Published findings (8673105, 11102452, 11748309, 12668451, 14594949, 15483641), Clinically tested gene	Have you or a relative been diagnosed with hypertrophic cardiomyopathy? Have you or a first-degree relative been diagnosed with cardiovascular disease before the age of 50? Do you have a relative who has died suddenly due to cardiac failure at an unusually young age?	All PGP10 reported "no" to the first question. For the second question: PGP9 reported that she had a first degree relative affected by cardiovascular disease, and PGP4 reported grandparents affected. All others (including PGP6) reported "no". For the third question: PGP4 and PGP6 reported "yes". PGP6 reported both maternal and paternal uncles who died of myocardial infarctions at ages of 41, 56, and 59 and were believed to have had coronary artery disease.
LDLR-V827I: Hypercholesterolemia (dom) [PGP6: het]	Rare*, PPH2: Prob damaging, Clinically tested gene	Have you or a relative been diagnosed with hypercholesterolemia (high cholesterol)?	PGP1 and PGP5 reported high cholesterol, PGP4 and PGP6 reported borderline high.
PCSK9-R237W: Hypocholesterolemia (dom) [PGP9: het]	Rare*, PPH2: Prob damaging, Published findings (15358785, 16424354, 16571601, 17765244), Clinically tested gene	Have you or a relative been diagnosed with hypocholesterolemia (abnormally low cholesterol)?	All PGP10 reported "no".

FLG-S761fs: Palmar hyperlinearity / Keratosis pilaris** (dom) [PGP2: het]	Rare*, Frameshift predicted, Clinically tested gene	Some subtle skin phenotypes can be caused by heterozygous variants would would cause severe skin disorder if homozygous. These can include palmar hyperlinearity (causing a hand to look unusually old). Do you have palmar hyperlinearity? Some subtle skin phenotypes can be caused by heterozygous variants would would cause severe skin disorder if homozygous. These can include keratosis pelaris (bumps on the skin on the upper arms, cheeks, or thighs), or fine scale on the skin. Do you have keratosis pilaris?	All PGP10 reported "no" to the first question. For the second question: PGP4 reported "maybe", all others (including PGP2) reported "no".
NF1-Q2721R: Neurofibromatosis 1 (dom) [PGP8: het]	Rare*, PPH2: Prob damaging, Clinically tested gene	Do you have cafe au lait spots (light brown birthmarks)? If so, please describe how many and whether they are larger than 15mm in any direction (a dime is 17mm).	PGP4 reported one spot that is 25mm at its longest. PGP5 reported a son with cafe au lait spots. All others (including PGP8) reported "no".
ALK-L1033P: Neuroblastoma (dom) [PGP3: het]	Rare*, PPH2: Prob damaging, Clinically tested gene	Have you or a relative been diagnosed with neuroblastoma?	All PGP10 reported "no".
WFS1-C426Y: Psychiatric disease (dom) [PGP1: het]	Rare*, PPH2: Poss damaging, Published findings (11244483), Clinically tested (but for unrelated disease)	Have you been diagnosed with any of the following psychiatric diseases? • Major depression • Bipolar disorder • Schizoaffective disorder • Schizophrenia	PGP4 reported depression/anxiety, no others reported psychiatric disease.
KCNQ3-R777Q: Benign neonatal seizures (dom) [PGP5: het]	Rare*, PPH2: Prob damaging, Clinically tested	Did you or a relative have benign seizures when an infant, during the first month of life, that went away?	PGP1 reported a second-degree relative with this condition. All others (including PGP5) reported "no".
SEPT9-R355W: Neuralgic amyotrophy (dom) [PGP6: het]	Rare*, PPH2: Prob damaging, Clinically tested	Neuralgic amyotrophy is a rare disease characterized by sudden onset of severe pain in shoulder or upper limbs, and subsequent muscle atrophy. Have you or a relative been diagnosed with neuralgic amyotrophy?	All PGP10 reported "no".
RHO-G51A: Retinitis pigmentosa (dom) [PGP10: het]	Rare*, PPH2: Prob damaging, OMIM, Published findings (8317502, 9380676, 16962629), Clinically tested gene	Autosomal dominant retinitis pigmentosa is characterized by progressive late onset vision loss, beginning with loss of night vision and peripheral vision. Do you or a relative have retinitis pigmentosa or similar symptoms?	All PGP10 reported "no".

* Variants are called "Rare" if only seen once in 64 PGP10 & public individuals (at least 100 chromosomes).
** Disruptive variants in this gene are reported to cause ichthyosis vulgaris in a recessive manner. Some literature implicates these genes in causing mild phenotypes (palmar hyperlinearity and keratosis pilaris) when heterozygous (see Table 2).

Table S7: GET-Evidence: Variant evidence and clinical importance scoring

Variant evidence: Computational	Add points for every consistent prediction, subtract points for contradicting evidence. * Other reports for this gene implicate it in same disease: +1 * Polyphen 2 prediction: +1 * SIFT prediction: +1 * Presence in conserved domain: +1 * Disruptive amino acid substitution (BLOSUM100 score): +1 * Nonsense or frameshift mutation: +2 -1 point total if overall evidence contradicts proposed effect
Variant evidence: Functional	Add points for each different functional observation. * Change in enzyme activity: +1 * Change in binding affinity: +1 * Change in cellular localization: +1 * Change in gene expression: +1 * Change in protein expression: +1 * Phenotype effect in animal models: +2
Variant evidence: Case/control	Significance scores for case/control data should derive from a single publication thought to be best representative of the variant's effect. Allele frequencies from other studies should only be used when an extremely high discordance contradicts the paper's hypothesis. -1 point total if case/control data and/or allele frequency strongly contradicts predicted effect 0 points if no evidence or significance > 0.1 1 point if significance < 0.1 2 points if significance < 0.05 3 points if significance < 0.025 4 points if significance < 0.01 5 points if significance < 0.0001
Variant evidence: Familial	-1 point total if familial data strongly contradicts predicted effect 0 points if no familial data or LOD < 0.5 1 point if LOD >= 0.5 2 points if LOD >= 1.0 3 points if LOD >= 1.5 and seen in at least 2 unrelated individuals 4 points if LOD >= 3 and seen in at least 2 unrelated individuals 5 points if LOD >= 5 and seen in at least 2 unrelated individuals
Clinical importance: Severity	0 points for benign 1 point for rarely having any effect on health (e.g. small increased susceptibility to infections -- either choose this or a low penetrance score, not both) 2 points for mild effect on quality of life and/or usually not symptomatic (Cystinuria) 3 points for moderate effect on quality of life (e.g., Familial Mediterranean Fever) 4 points for severe effect: causes serious disability or reduces life expectancy (e.g., Sickle-cell, Stargardt's disease) 5 points for very severe effect, lethal by early adulthood (e.g., Lethal junctional epidermolysis bullosa, Adrenoleukodystrophy)
Clinical importance: Treatability	0 points for no clinical evidence supporting intervention (e.g., PAF acetylhydrolase deficiency) 1 point for incurable: Treatment only to alleviate symptoms 2 points for potentially treatable: Treatment is in development or controversial 3 points for somewhat treatable: Standard treatment, but only a small or moderate improvement of mortality/morbidity 4 points for treatable: Standard treatment significantly reduces the amount of mortality/morbidity, but does not eliminate it 5 points for extremely treatable: Well-established treatment essentially eliminates the effect of the disease (e.g., PKU)
Clinical importance: Penetrance	0 points if < 0.1% attributable risk (extremely low penetrance) 1 point if ≥ 0.1% attributable risk (very low penetrance) 2 points if ≥ 1% attributable risk (low penetrance) 3 points if ≥ 5% attributable risk (moderate penetrance) 4 points if ≥ 20% attributable risk (moderately high penetrance) 5 points if ≥ 50% attributable risk (complete or highly penetrant)

Table S8: GET-Evidence: Assessment of strength of evidence

well-established	* At least 4 points in either “Case/control evidence” or “Familial evidence” **and** * At least eight points total in evidence categories
likely	* At least 3 points in either “Case/control evidence” or “Familial evidence” **and** * At least five points total in evidence categories
uncertain	Any variants which do not meet the above requirements.

Table S9: GET-Evidence: Assessment of clinical importance

high clinical importance	* At least 4 points in penetrance (high-moderate penetrance / >= 20% attributable risk) **and either**: * At least 3 stars in severity and at least 4 stars in treatability **or** * At least 4 stars in severity
moderate clinical importance	* At least 3 points in penetrance (high-moderate penetrance / >= 5% attributable risk) **and either**: * At least 2 stars in severity and at least 4 stars in treatability **or** * At least 3 stars in severity
low clinical importance	Any variants which do not meet the above requirements.

Table S10: GET-Evidence information regarding variants from Table S5

Variant	Predicted phenotype (a "?" denotes a variant with no supporting published findings)	Allele frequency	Prioritization score	Evidence assessment in GET-Evidence	Clinical importance assessment in GET-Evidence
RYR2-G1885E	Arrhythmogenic right ventricular cardiomyopathy (when compound het. with G1886S)	1.8%	4	Uncertain	High
FIG4-K278fs	Charcot-Marie-Tooth Disease Type 4J?	unknown	3	Uncertain	Moderate
RP1-T373I	Retinitis pigmentosa	1.2%	5	Uncertain	High
FLG-S761fs	Ichthyosis vulgaris?	unknown	4	Uncertain	Moderate
TGM1-Y312fs	Congenital ichthyosis?	unknown	4	Uncertain	Moderate
SLC4A1-E40K	Hemolytic anemia	1.2%	3	Uncertain	Moderate
ABCA4-P1780A	Stargardt disease	0.04%	5	Uncertain	High
LRP5-V667M	Osteoporosis-pseudoglioma syndrome	4.15%	6	Uncertain	High
SPG11-K1013E	Spastic paraplegia	1.0%	4	Uncertain	High
CPT2-S113L	Late-onset carnitine palmitoyltransferase deficiency	0.1%	6	Well-established	High
SLC6A5-T425M	Hyperekplexia	0.01%	5	Uncertain	Moderate
SERPINA1-R247C	Alpha-1 antitrypsin deficiency	0.3%	4	Uncertain	High
CBS-T460M	Homocystinuria	unknown	4	Uncertain	High
ACADVL-R385W	Very Long Chain Acyl-Coenzyme A Dehydrogenase Deficiency	unknown	4	Uncertain	High
TGM1-E520G	Lamellar ichthyosis	0.6%	5	Uncertain	Moderate
SLC7A9-A182T	Cystinuria	0.3%	4	Uncertain	Moderate
SLX4-G1396fs	Fanconi Anemia (complementation group P)?	unknown	4	Uncertain	High
SPG7-G199fs	Hereditary spastic paraplegia?	unknown	4	Uncertain	High
SLC26A4-I300L	Pendred Syndrome	0.4%	5	Uncertain	Moderate
PEX1-I696M	Peroxisome biogenesis disorders	2.7%	4	Uncertain	High