Science Translational Medicine
AAAS

www.sciencetranslationalmedicine.org/cgi/content/full/8/364/364ra151/DC1

Supplementary Materials for

Aggregate penetrance of genomic variants for actionable disorders in European and African Americans

Pradeep Natarajan, Nina B. Gold, Alexander G. Bick, Heather McLaughlin, Peter Kraft, Heidi L. Rehm, Gina M. Peloso, James G. Wilson, Adolfo Correa, Jonathan G. Seidman, Christine E. Seidman, Sekar Kathiresan, Robert C. Green*

*Corresponding author. Email: rcgreen@bwh.harvard.edu

Published 9 November 2016, *Sci. Transl. Med.* **8**, 364ra151 (2016)
DOI: 10.1126/scitranslmed.aag2367

This PDF file includes:

Supplementary Materials

Table S1. ACMG incidental findings genes and transcripts analyzed.

Gene	Transcript
ACTA2	NM_001613.2
ACTC1	NM_005159.4
APC	NM_000038.4
APOB	NM_000384.2
BRCA1	NM_007294.3 NM_007300.3 NM_007297.3 NM_007298.3 NM_007299.3
BRCA2	NM_000059.3
CACNA1S	NM_000069.2
COL3A1	NM_000090.3
DSC2	NM_024422.3 NM_004949.3
DSG2	NM_001943.3
DSP	NM_004415.2
FBN1	NM_000138.4
GLA	NM_000169.2
KCNH2	NM_000238.3 NM_172057.2 NM_172056.2
KCNQ1	NM_000218.2 NM_181798.1
LDLR	NM_000527.4
LMNA	NM_170707.2
MEN1	NM_130799.2
MLH1	NM_000249.3
MSH2	NM_000251.1
MSH6	NM_000179.2
MUTYH	NM_012222.2
MYBPC3	NM_000256.3
MYH11	NM_001040113.1 NM_001040114.1
MYH7	NM_000257.2
MYL2	NM_000432.3
MYL3	NM_000258.2
MYLK	NM_053025.3
NF2	NM_181831.2 NM_181825.2 NM_181832.2 NM_181830.2 NM_181829.2 NM_181828.2 NM_016418.5 NM_000268.3 NM_181833.2
PCSK9	NM_174936.3

PKP2	NM_004572.3
PMS2	NM_000535.5
PRKAG2	NM_016203.3
PTEN	NM_000314.4
RB1	NM_000321.2
RET	NM_020975.4 NM_020630.4
RYR1	NM_001042723.1 NM_000540.2
RYR2	NM_001035.2
SCN5A	NM_198056.2 NM_001099404.1 NM_001160160.1
SDHAF2	NM_017841.2
SDHB	NM_003000.2
SDHC	NM_001035511.1 NM_003001.3
SDHD	NM_003002.2
SMAD3	NM_005902.3 NM_001145103.1
STK11	NM_000455.4
TGFBR1	NM_004612.2
TGFBR2	NM_003242.5 NM_001024847.2
TMEM43	NM_024334.2
TNNI3	NM_000363.4
TNNT2	NM_001001430.1 NM_000364.2
TP53	NM_000546.4
TPM1	NM_001018020.1 NM_001018008.1 NM_001018005.1 NM_000366.5
TSC1	NM_000368.4
TSC2	NM_000548.3
VHL	NM_000551.2
WT1	NM_001198552.1 NM_024426.4 NM_001198551.1 NM_000378.4 NM_024424.3

Table S2. Classification evidence for PVs and LPVs from FHS and JHS participants.

Gene	Variant and Transcript	Amino acid	Associated phenotype	Classification evidence
			PATHOGENIC VARIANTS	
APOB	c.10580G>A NM_000384.2	p.Arg3527Gln	HCL	The p.Arg3527Gln (also referred to as p.Arg3500Gln) variant is a well-established pathogenic variant in apolipoprotein B dysfunction and has been reported in more than 50 families (*67-73*). In vitro functional studies provide evidence that the p.Arg3527Gln variant may impact protein function (*68*). In summary, this variant meets our criteria to be classified as pathogenic for apolipoprotein B dysfunction in an autosomal dominant manner based upon segregation studies and functional evidence.
MLH1	c.2059C>T NM_000249.3	p.Arg687Trp	CC	The p.Arg687Trp variant in *MLH1* has been reported in several individuals with colorectal cancer, disrupts mismatch repair in yeast (*74*) and is considered "Pathogenic" by the ClinGen-approved expert panel InSiGHT variant interpretation committee (*75*). In summary, this variant meets our criteria to be classified as pathogenic for colorectal cancer in an autosomal dominant manner.
SCN5A	c.3214G>T NM_198056.2 NM_001099404.1 NM_001160160.1	p.Glu1072*	Brugada	The p.Glu1072* variant in *SCN5A* has not been previously reported in individuals with disease and was absent from large population studies. This nonsense variant leads to a premature termination codon at position 1072 which is predicted to lead to a truncated or absent protein. Heterozygous loss of function of the *SCN5A* gene is an established disease mechanism in Brugada syndrome. In summary, this variant meets our criteria to be classified as pathogenic for Brugada syndrome in an autosomal dominant based upon absence from controls and predicted impact on the protein.
MYL3	c.170C>G NM_000258.2	p.Ala57Gly	HCM	The p.Ala57Gly variant in *MYL3* has been identified previously by the Partners LMM in 2 individuals with HCM, and in 2 individuals with family history of HCM. This variant was also reported in two Korean families and two Japanese individuals with HCM (*76-78*). The variant was shown to segregate with disease in 5 affected members of the two Korean families. *In vitro* and in vivo functional studies provide some evidence that the p.Ala57Gly variant impacts protein function resulting in a high level of fibrosis and hypertrophy in an animal model (*79, 80*). This variant has been identified in 6/67654 European chromosomes and in 4/8758 East Asian chromosomes by the Exome Aggregation Consortium (*81*). Although this variant has been seen in the general population, its frequency is not high enough to rule out a pathogenic role. In summary, this variant meets our criteria to be classified as pathogenic based upon case observations, segregation studies and functional evidence.
KCNQ1	c.613G>A NM_000218.2	p.Val205Met	LQTS	The p.Val205Met variant in *KCNQ1* has been reported in at least 2 presumably unrelated Gitxsan individuals with LQTS and segregated with the disease in 12 affected family members (*82*). Four individuals have been reported to be homozygous for this variant and present with a clinically more severe phenotype (*83*). This variant has been identified only in 0.023% (1/4402) of African American chromosomes by the NHLBI Exome Sequencing Project (*84*) (dbSNP rs151344631). Computational prediction tools and conservation analysis suggest that the p.Val205Met variant may impact the protein, though this information is not predictive enough to determine pathogenicity. *In vitro* functional studies provide some evidence that the p.Val205Met variant may impact protein function (*82, 85*). In

				summary, this variant meets our criteria to be classified as pathogenic, with incomplete penetrance, for LQTS in an autosomal dominant manner based upon segregation studies, case studies and functional evidence.
KCNQ1	c.1552C>T NM_000218.2	p.Arg518*	LQTS	The p.Arg518* variant in *KCNQ1* is a well-established pathogenic variant associated with LQTS and Jervell and Lange-Nielsen Syndrome (JLNS) (*86*). This variant has been identified in 16/66,544 European chromosomes by the Exome Aggregation Consortium (*81*) (dbSNP rs17215500). This nonsense variant leads to a premature termination codon at position 518 which is predicted to lead to a truncated or absent protein. In summary, this variant meets our criteria to be classified as pathogenic for LQTS in an autosomal dominant manner and for JLNS in an autosomal recessive manner.
MYBPC3	c.26-2A>G NM_000256.3	p.?	HCM	The c.26-2A>G variant in *MYBPC3* has been reported in at least 9 individuals with HCM and segregated with disease in 2 affected relatives from 1 family (*87-91*). In addition, this variant has been identified by the Partners LMM in 3 individuals with HCM and segregated with disease in 2 affected relatives from 2 families. This variant has also been identified in 4/45416 European chromosomes by the Exome Aggregation Consortium (*81*) (dbSNP rs376395543); however, for diseases with clinical variability and reduced penetrance, pathogenic variants may be present at a low frequency in the general population. Finally, this occurs in the invariant region (+/- 1,2) of the splice consensus sequence and is predicted to cause altered splicing leading to an abnormal or absent protein. Heterozygous splice variants in *MYBPC3* are prevalent in cases of HCM. In summary, this variant meets our criteria to be classified as pathogenic for HCM in an autosomal dominant manner based upon predicted variant impact, case observations and segregation studies.
MYBPC3	c.1504C>T NM_000256.3	p.Arg502Trp	HCM	The p.Arg502Trp variant in *MYBPC3* has been well reported in multiple individuals across multiple studies and is known to be pathogenic for HCM. This variant meets our criteria for pathogenicity based upon extensive segregation studies and functional evidence (*90, 92-98*). It is also the most common pathogenic HCM variant.
PKP2	c.1689-1G>C NM_004572.3	p.?	ARVD/ C	The c.1689-1G>C variant in *PKP2* has been reported in 2 individuals with arrhythmogenic right ventricular cardiomyopathy (*99, 100*) and was absent from large population studies. This variant occurs in the invariant region (+/- 1,2) of the splice consensus sequence and is predicted to cause altered splicing leading to an abnormal or absent protein. In summary, this variant meets our criteria to be classified as pathogenic for arrhythmogenic right ventricular cardiomyopathy in an autosomal dominant manner based upon predicted impact on the protein, case observations and absence from controls.
PKP2	c.1237C>T NM_004572.3	p.Arg413*	ARVD/C	The p.Arg413* variant in *PKP2* has been identified in >10 individuals with ARVD/C and segregated with disease in at least 3 affected relatives from one family (*100-105*). Overexpression of this variant in mice increased right ventricular size and shortened ventricular action potential durations (*105*), though this assay may not accurately represent the biological disease state. This nonsense variant leads to a premature termination codon at position 413, which is predicted to lead to a truncated or absent protein. In summary, this variant meets our criteria to be classified as pathogenic for ARVC in an autosomal dominant manner based upon segregation studies, case studies, absence from controls, and predicted gene impact.
BRCA2	c.658_659del NM_000059.3	p.Val220Ilefs*4	BC; OC	The p.Val220Ilefs*4 variant in *BRCA2* has been reported in the literature in numerous individuals with hereditary breast and ovarian cancer (*106-108*) and has been reported in 37 individuals

				with breast and/or ovarian cancer in the Breast Cancer Information Core (BIC) database. This p.Val220fs variant has also been identified in individuals with Fanconi anemia, Wilms tumor, glioblastoma, or medulloblastoma; however all these individuals also carry a second pathogenic *BRCA2* variant (*109-111*). This variant has been identified in 4/51064 European chromosomes by the Exome Aggregation Consortium (*81*) (dbSNP rs80359604). This variant is predicted to cause a frameshift, which alters the protein's amino acid sequence beginning at position 220 and leads to a premature termination codon 4 amino acids downstream. This alteration is then predicted to lead to a truncated or absent protein. In summary, this variant meets our criteria to be classified as pathogenic for *BRCA2*-associated cancers in an autosomal dominant manner based upon genetic studies and the predicted impact to the protein.
BRCA2	c.4398_4402del NM_000059.3	p.Leu1466Phefs*2	BC; OC	The p.Leu1466Phefs*2 variant in *BRCA2* has been reported in at least 1 individual with prostate cancer (*112*). This variant has been identified in 0.036% (3/8240) of European American chromosomes and 0.094% (4/4258) of African American chromosomes by the NHLBI Exome Sequencing Project (*84*) (dbSNP rs80359444). This variant is predicted to cause a frameshift, which alters the protein's amino acid sequence beginning at position 1466 and leads to a premature termination codon 2 amino acids downstream. This alteration is then predicted to lead to a truncated or absent protein. Heterozygous loss of function of the *BRCA2* gene is an established disease mechanism in *BRCA2*-associated cancers. In summary, this variant meets our criteria to be classified as pathogenic for *BRCA2*-associated cancers in an autosomal dominant manner based upon the predicted impact to the protein.
BRCA2	c.5213_5216del NM_000059.3	p.Thr1738Ilefs*2	BC; OC	The p.Thr1738Ilefs*2 variant in *BRCA2* has been reported in at least 2 individuals with breast and/or ovarian cancer (*113, 114*). This variant has been identified in 1/66156 of European chromosomes by the Exome Aggregation Consortium (*81*) (dbSNP rs80359493). This variant is predicted to cause a frameshift, which alters the protein's amino acid sequence beginning at position 1738 and leads to a premature termination codon 2 amino acids downstream. This alteration is then predicted to lead to a truncated or absent protein. Heterozygous loss of function of the *BRCA2* gene is an established disease mechanism in *BRCA2* associated cancers. In summary, this variant meets our criteria to be classified as pathogenic for *BRCA2*-associated cancers in an autosomal dominant manner based upon the predicted impact to the protein.
BRCA2	c.5611_5615del NM_000059.3	p. Lys1872Asnfs*2	BC; OC	The p.Lys1872Asnfs*2 variant in *BRCA2* has been reported in literature in at least 1 individual with breast cancer (*115*). It has also been reported in 3 individuals with breast and/or ovarian cancer in the UMD database and is found at very low frequency (2/10144 African chromosomes) by the Exome Aggregation Consortium (*81*) (dbSNP rs80359525). This variant is predicted to cause a frameshift, which alters the protein's amino acid sequence beginning at position 1872 and leads to a premature termination codon 2 amino acids downstream. This alteration is then predicted to lead to a truncated or absent protein. Heterozygous loss of function of the *BRCA2* gene is an established disease mechanism in *BRCA2*-associated cancers. In summary, this variant meets our criteria to be classified as pathogenic for *BRCA2*-associated cancers in an autosomal dominant manner based upon the predicted impact to the protein and case observations.

BRCA2	c.5855T>A NM_000059.3	p.Leu1952*	BC; OC	The p.Leu1952* variant in *BRCA2* has not been reported in the literature and is at very low frequency (1/4406 African American chromosomes) in the NHLBI Exome Sequencing Project (*84*) (dbSNP rs375064902). This nonsense variant leads to a premature termination codon at position 1952 which is predicted to lead to a truncated or absent protein. In summary, this variant meets our criteria to be classified as pathogenic for *BRCA2*-associated cancers in an autosomal dominant manner based upon the predicted impact to the protein.
BRCA2	c.9382C>T NM_000059.3	p.Arg3128*	BC; OC	The p.Arg3128* variant in *BRCA2* has been reported in the literature in at least 2 individuals with prostate cancer and in one individual with breast cancer (*116-118*). This variant was observed at very low frequency (2/10406 African chromosomes) by the Exome Aggregation Consortium (*81*) (dbSNP rs80359212). This nonsense variant leads to a premature termination codon at position 3128 which is predicted to lead to a truncated or absent protein. In summary, this variant meets our criteria to be classified as pathogenic for *BRCA2*-associated cancers in an autosomal dominant manner based upon the predicted impact to the protein.
MYH7	c.2389G>A NM_000257.2	p.Ala797Thr	HCM	The p.Ala797Thr variant in *MYH7* has been identified in >30 individuals with HCM and segregated with disease in >10 affected family members (*119-124*). This variant has been identified in 4/121372 chromosomes from various ethnicities by the Exome Aggregation Consortium (*81*) (dbSNP rs3218716). Please note that, for diseases with clinical variability and reduced penetrance, pathogenic variants may be present at a low frequency in the general population. In summary, this variant meets our criteria to be classified as pathogenic for HCM in an autosomal dominant manner based upon case observations and segregation studies.
TP53	c.818G>A NM_000546.4	p.Arg273His	LFS	The p.Arg273His variant in *TP53* has been reported in numerous individuals with various types of Li Fraumeni-associated cancers, including sarcomas, gastric carcinoma, breast cancer, uterine serous cancer, rhabdomyosarcoma-associated renal cell carcinoma, and acute lymphoblastic leukemia (*125-132*). *In vitro* functional studies provide some evidence that the p.Arg273His variant may impact protein function (*131, 133-140*). However, these types of assays may not accurately represent biological function. In summary, this variant meets our criteria to be classified as pathogenic for Li-Fraumeni syndrome in an autosomal dominant manner based upon case studies, absence from controls, and functional evidence.
BRCA1	c.5177_5180del NM_007300.3	p.Arg1726Lysfs*3	BC; OC	The p.Arg1726Lysfs*3 variant in *BRCA1* has been reported in at least 5 individuals with breast and ovarian cancer syndrome (*114, 141-143*) and was absent from large population studies (has only been identified in 1/4264 African American chromosomes by the NHLBI Exome Sequencing Project (*84*) (dbSNP rs80357975). This variant is predicted to cause a frameshift, which alters the protein's amino acid sequence beginning at position 1726 and leads to a premature termination codon 3 amino acids downstream. This alteration is then predicted to lead to a truncated or absent protein. Heterozygous loss of function of the *BRCA1* gene is an established disease mechanism in breast and ovarian cancer syndrome. In summary, this variant meets our criteria to be classified as pathogenic for breast and ovarian cancer syndrome in an autosomal dominant manner (*144*).
BRCA1	c.3607C>T NM_007300.3 NM_007294.3	p.Arg1203*	BC; OC	The p.Arg1203* variant in *BRCA1* has been reported in several individuals with breast and ovarian cancer syndrome (*144, 145*) and was absent from large population studies (has only been identified in 1/4406 African American chromosomes by the NHLBI Exome Sequencing Project (*84*) (dbSNP rs62625308).

				This nonsense variant leads to a premature termination codon at position 1203 which is predicted to lead to a truncated or absent protein. Heterozygous loss of function of the *BRCA1* gene is an established disease mechanism in breast and ovarian cancer syndrome. In summary, this variant meets our criteria to be classified as pathogenic for breast and ovarian cancer syndrome in an autosomal dominant manner.
LDLR	c.429C>A NM_000527.4	p.Cys143*	HCL	The p.Cys143* variant in *LDLR* has not been previously reported in individuals with disease and has been identified in 3/66347 of European chromosomes by the Exome Aggregation Consortium (*81*) (dbSNP rs199774121). This nonsense variant leads to a premature termination codon at position 143 which is predicted to lead to a truncated or absent protein. Heterozygous loss of function of the *LDLR* gene is an established disease mechanism in hypercholesterolemia. In summary, this variant meets our criteria to be classified as pathogenic for hypercholesterolemia in an autosomal dominant manner based upon its predicted effect on the protein.
LDLR	c.2054C>T NM_000527.4	p.Pro685Leu	HCL	The p.Pro685Leu variant in *LDLR* has been reported in at least 29 individuals with clinical features of hypercholesterolemia and segregated with disease in their families (*67*, *146*). This variant has been identified in 3/8600 of European American chromosomes by the NHLBI Exome Sequencing Project (*84*) (dbSNP rs28942084). *In vitro* functional studies provide evidence that the p.Pro685Leu variant may impact protein function (*147*). In summary, this variant meets our criteria to be classified as pathogenic for hypercholesterolemia in an autosomal dominant manner based upon segregation studies and functional evidence.
RYR1	c.7300G>A NM_001042723.1 NM_000540.2	p.Gly2434Arg	MHS	The p.Gly2434Arg variant in *RYR1* has been reported in more than 100 individuals with malignant hyperthermia and segregated with the disease in several families (*148-154*). *In vitro* functional studies provide some evidence that the p.Gly2434Arg variant may impact protein function (*152, 155*). This variant has only been identified in 2/66466 of European and 1/10366 African chromosomes by the Exome Aggregation Consortium (*81*) (dbSNP rs121918593). In summary, this variant meets our criteria to be classified as pathogenic for malignant hyperthermia in an autosomal dominant based upon segregation studies, absence from controls, functional evidence.
		LIKELY PATHOGENIC VARIANTS		
MLH1	c.1153C>T NM_000249.3	p.Arg385Cys	CC	The p.Arg385Cys variant in *MLH1* has been reported in 3 individuals with colorectal cancer (*156-158*) and identified in 0.03% (3/8,648) of East Asian chromosomes by the Exome Aggregation Consortium (*81*) (dbSNP rs63750760). The variant is considered "Likely Pathogenic" by the ClinGen-approved expert panel InSiGHT variant interpretation committee (*75*). In summary, although additional studies are required to fully establish its clinical significance, the p.Arg385Cys variant is likely pathogenic.
DSP	c.3865C>T NM_004415.2	p.Gln1289*	ARVD/C	The p.Gln1289* variant in *DSP* has not been previously reported in individuals with disease and had only been identified in 1/10188 of African chromosomes by the Exome Aggregation Consortium (*81*). This nonsense variant leads to a premature termination codon at position 1289 which is predicted to lead to a truncated or absent protein. Frameshift and nonsense variants in *DSP* have been well reported in patients with ARVD/C (*159*), but recent evidence supports that they can also cause DCM (*160*). In summary, although additional studies are required to fully establish its clinical significance, the p.Gln1289* variant is likely pathogenic.

DSP	c.4180C>T NM_004415.2	p.Gln1394*	ARVD/C	The p.Gln1394* variant in *DSP* has not been reported in individuals with disease and has only been identified in 1/66536 of European chromosomes by the Exome Aggregation Consortium (*81*) (dbSNP rs140474226). This nonsense variant leads to a premature termination codon at position 1394 which is predicted to lead to a truncated or absent protein. Frameshift and nonsense variants in DSP have been well reported in patients with ARVD/C (*159*), but recent evidence supports that they can also cause DCM (*160*). In summary, although additional studies are required to fully establish its clinical significance, the p.Gln1394* variant is likely pathogenic.
KCNQ1	c.535G>A NM_000218.2	p.Gly179Ser	LQTS	The p.Gly179Ser variant in *KCNQ1* has been reported in 4 probands (one homozygous, one compound heterozygous and 2 of unreported status) with LQTS as well as segregating in 3 additional family members with LQTS (*161-163*). An additional 3 family members had borderline QT intervals and 3 others were considered normal. This variant has only been identified in 0.023% (1/4,400) of African American chromosomes by the NHLBI Exome Sequencing Project (*84*) (dbSNP rs199473394). Computational prediction tools and conservation analysis suggest that the p.Gly179Ser variant may impact the protein, though this information is not predictive enough to determine pathogenicity. *In vitro* functional studies provide some evidence that the p.Gly179Ser variant impacts protein function in a homozygous state though no effect was seen when mixed with normal protein (*164*). In summary, the p.Gly179Ser variant is likely pathogenic though additional studies are required to confirm its clinical significance as well as evaluate penetrance which is likely incomplete in a heterozygous state.
KCNQ1	c.1085A>G NM_000218.2	p.Lys362Arg	LQTS	The p.Lys362Arg variant in *KCNQ1* has been reported in 3 heterozygous individuals with LQTS and in 2 compound heterozygous individuals with Lange-Nielsen syndrome without auditory phenotype (*165-168*). This variant has been identified in 1/10356 of African chromosomes and in 1/66474 European chromosomes by the Exome Aggregation Consortium (*81*) (dbSNP rs12720458). Although this variant has been seen in the general population, its frequency is not high enough to rule out a pathogenic role. Computational prediction tools and conservation analysis suggest that the p.Lys362Arg variant may impact the protein, though this information is not predictive enough to determine pathogenicity. In summary, the p.Lys362Arg variant is likely pathogenic though additional studies are required to confirm its clinical significance as well as evaluate penetrance which is likely incomplete in a heterozygous state.
GLA	c.335G>A NM_000169.2	p.Arg112His	Fabry	The p.Arg112His variant in *GLA* has been previously identified in 3 individuals with Fabry disease (*169-171*). This variant has been identified in 1/6728 European American chromosomes (female) by the NHLBI Exome Sequencing Project (*84*). Functional studies indicate this variant results in reduced *GLA* activity (*169, 171*). In summary, this variant is likely pathogenic, though additional studies are required to fully establish its clinical significance.

Abbreviations: P, pathogenic; LP, likely pathogenic; LOF, loss-of-function; HCL, hypercholesterolemia; CC, colorectal cancer; LQTS, long QT syndrome; HCM, hypertrophic cardiomyopathy; ARVD/C, arrhythmogenic right ventricular dysplasia / cardiomyopathy; BC, breast cancer; OC, ovarian cancer; LFS, Li-Fraumeni syndrome