

## Actionable genomic variants in 6045 participants from the Qatar Genome Program

Amal Elfatih, Borbala Mifsud, Najeeb Syed, Ramin Badii, Hamdi Mbarek, Fatemeh Abbaszadeh, The Qatar Genome Program Research Consortium, Xavier Estivill

### Item type

Journal Contribution

### Terms of use

This work is licensed under a [CC BY-NC 4.0](#) license

### This version is available at

[https://manara.qnl.qa/articles/journal\\_contribution/Actionable\\_genomic\\_variants\\_in\\_6045\\_participants\\_from\\_the\\_Qatar\\_Genome\\_Program](https://manara.qnl.qa/articles/journal_contribution/Actionable_genomic_variants_in_6045_participants_from_the_Qatar_Genome_Program)



Access the item on Manara for more information about usage details and recommended citation.

Posted on Manara – Qatar Research Repository on

2021-08-24

## RESEARCH ARTICLE

# Actionable genomic variants in 6045 participants from the Qatar Genome Program

Amal Elfatih<sup>1</sup>  | Borbala Mifsud<sup>1,2</sup> | Najeeb Syed<sup>3</sup> | Ramin Badii<sup>4</sup> |  
Hamdi Mbarek<sup>5</sup> | Fatemeh Abbaszadeh<sup>6</sup>  | The Qatar Genome Program Research  
Consortium | Xavier Estivill<sup>7</sup>

<sup>1</sup>Genomics and Precision Medicine, College of Health and Life Science, Hamad Bin Khalifa University, Doha, Qatar

<sup>2</sup>William Harvey Research Institute, Queen Mary University London, London, UK

<sup>3</sup>Applied Bioinformatics Core, Integrated Genomics Services, Research Branch, Sidra Medicine, Doha, Qatar

<sup>4</sup>Molecular Genetics Laboratory, Hamad Medical Corporation, Doha, Qatar

<sup>5</sup>Qatar Genome Program, Qatar Foundation Research, Development and Innovation, Qatar Foundation, Doha, Qatar

<sup>6</sup>Diagnostic Genomic Division, Hamad Medical Corporation, Doha, Qatar

<sup>7</sup>Research Department, T'havia Quantitative Genomics Laboratories (qGenomics), Barcelona, Spain

## Correspondence

Amal Elfatih, Genomics and Precision Medicine, College of Health and Life Science, Hamad Bin Khalifa University, Doha, Qatar.  
Email: [abdeltameed@hbku.edu.qa](mailto:abdeltameed@hbku.edu.qa)

Xavier Estivill, Research Department, T'havia Quantitative Genomics Laboratories (qGenomics), Barcelona, Spain.  
Email: [estivill@me.com](mailto:estivill@me.com)

## Abstract

In a clinical setting, DNA sequencing can uncover findings unrelated to the purpose of genetic evaluation. The American College of Medical Genetics and Genomics (ACMG) recommends the evaluation and reporting of 59 genes from clinic genomic sequencing. While the prevalence of secondary findings is available from large population studies, these data lack Arab and other Middle Eastern populations. The Qatar Genome Program (QGP) generates whole-genome sequencing (WGS) data and combines it with phenotypic information to create a comprehensive database for studying the Qatari and wider Arab and Middle Eastern populations at the molecular level. This study identified and analyzed medically actionable variants in the 59 ACMG genes using WGS data from 6045 QGP participants. Our results identified a total of 60 pathogenic and likely pathogenic variants in 25 ACMG genes in 141 unique individuals. Overall, 2.3% of the QGP sequenced participants carried a pathogenic or likely pathogenic variant in one of the 59 ACMG genes. We evaluated the QGP phenotype-genotype association of additional nonpathogenic ACMG variants. These variants were found in patients from the Hamad Medical Corporation or reported incidental findings data in Qatar. We found a significant phenotype association for two variants, c.313+3A>C in LDLR, and c.58C>T (p.Gln20\*) in the TPM1.

## KEYWORDS

ACMG, Biobank, exome sequencing, genome sequencing, medically actionable, Qatar

## 1 | INTRODUCTION

Technological developments and reduction in costs have led to genome/exome sequencing becoming fully incorporated in both research and clinical settings to understand and diagnose genetic diseases. Occasionally, genome/exome sequencing can uncover secondary findings that are not related to the primary purpose for

collecting the genomic data. Up until now, there is no consensus on whether to return these secondary findings or not. Some recommend the disclosure of secondary findings to patients when there is a pathogenic (P) variant that could lead to a disease with evidence that strongly supports early intervention benefits (Green et al., 2013). Another point of view considers the patient perspective in which receiving secondary findings can cause distress rather than being of

This is an open access article under the terms of the Creative Commons Attribution NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Human Mutation* published by Wiley Periodicals LLC.

any clinical benefit, leading to health insurance issues and/or employment discrimination (Hegde et al., 2015).

American College of Medical Genetics and Genomics (ACMG) recommends clinical sequencing laboratories to report P variants in 59 genes even when these are unrelated to the primary reason of genetic analysis. These 59 genes are associated with highly penetrant mutations leading to diseases. Individuals carrying P variants in these genes are at high risk to develop diseases in the future that could be treated or prevented to reduce morbidity and/or mortality (Kalia et al., 2017). The ACMG has recently released an updated policy statement for reporting secondary findings. A total of 73 genes have been included in the SF v3.0 list for reporting of secondary findings, with an intent to update the gene list annually (Miller, Lee, Chung, et al., 2021; Miller, Lee, Gordon, et al., 2021). While the prevalence and epidemiology of medically actionable variants are available from large population studies, there is a lack of data on genetic variants prevalent in the Arab and Middle Eastern population, where consanguineous marriages are common (Al Thani et al., 2019; Scott et al., 2016; Tadmouri et al., 2009).

The Qatar Genome Program (QGP) is a population-based project launched by the Qatar Foundation to generate a large whole-genome sequencing (WGS) data set. In combination with the comprehensive phenotypic information collected by the Qatar Biobank (QBB) (Al Thani et al., 2019), this will be used to study the genetic architecture of the Qatari population, generate a Qatari reference genome, and implement precision medicine (<https://qatargenome.org.qa>). To this date, the program has sequenced more than 10,000 whole genomes from the local population. The combination of WGS and health data, at the individual level, confer the possibility of understanding the diverse phenotypic consequences and the penetration of genetic variations in disease-causing genes.

Our goal in this study was to identify and analyze medically actionable variants in the 59 ACMG genes in WGS data from 6045 participants from the QGP first phase. In summary, the study is a systematic attempt to identify clinically relevant findings to make clinical prediction feasible for the individual's health. To the best of our knowledge, this study is the largest and most comprehensive analysis to estimate medically actionable variants in the Arab population of the Middle East.

## 2 | MATERIALS AND METHODS

### 2.1 | ACMG 59 genes recommended for reporting secondary findings

Genes studied here are the 59 medically actionable genes (Table S1), as recommended by the ACMG SF v2.0, to return in any clinical genomic sequencing laboratory, even when unrelated to the primary medical reason for testing since they are associated with highly penetrant mutations causing disease phenotypes (Kalia et al., 2017).

### 2.2 | Data set of Qatari population genomes

QBB targets to recruit 60,000 subjects aged 18 years or more with follow-up every 5 years. The current analysis includes data for 6045 QGP participants. Detailed information about the cohort, phenotypic data, and collection of samples by QBB have been described elsewhere (Al Thani et al., 2019). The QBB comprises phenotypic data, including sociodemographic information, medical history, imaging, and physical and clinical measurements. Moreover, biological samples (blood, urine, saliva, DNA, RNA, and viable cells) of recruited participants are also collected and stored.

The WGS data set encompassed 88,191,239 genetic variants, of which 74,991,446 were single nucleotide variants and 13,199,792 insertions and deletions, obtained from 6045 participants in Qatar. Details on the collection and processing of WGS data from the QGP have been described previously (Al Thani et al., 2019; Thareja et al., 2021). Briefly, genomic DNA was extracted from whole peripheral blood using the automated QIASymphony SP instrument according to the Qiagen MIDI kit protocol's recommendations (Qiagen). Sequencing was conducted on the Illumina HiSeq X Ten (Illumina) platform with an average of  $\times 30$  coverage. FASTQ (v0.11.2) files (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) were aligned to the reference genome GRCh37 (hs37d53). Data were processed using the bwakit (v0.7.11) (<https://github.com/lh3/bwa/tree/master/bwakit>) tool. Variants calling were performed using GATK (v3.4) following best practice for variant calling (<https://software.broadinstitute.org/gatk/documentation/article?id=3238>). Joint calling was performed on individual gVCF files to generate a joint multi-samples VCF file for all the samples. After performing the GATK VQSR filtering steps, only variants with PASS filter were considered for further downstream analysis.

A total of 148,285 variants were retrieved from the QGP data set corresponding to the 59 ACMG genes. These variants were annotated using SnpEff/SnpSift (v4.3t) (<https://pcingola.github.io/SnpEff/>) in which different databases, including dbSNP (build 151) (Sherry et al., 2001), dbNSFP (v2.9) (<https://sites.google.com/site/jpopgen/dbNSFP>), GWAS catalog (<https://www.ebi.ac.uk/gwas/>), and msigDBdb (v5.0) (<https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>) were used for an integrated annotation. We further performed annotation using different clinical genome reference databases, such as Human Genome Mutation Database (HGMD) (v2018.2) (Stenson et al., 2009), ClinVar (v20190211) (Landrum et al., 2016), and Online Mendelian Inheritance in Man (OMIM). Allele frequencies across global populations were added using frequency databases, such as gnomAD Exome and WGS (r2.0.2), Exac (r0.3.1), 1000 Genome Project (Altshuler et al., 2010), and the Greater Middle East data (GME) (Scott et al., 2016).

### 2.3 | Ethics statement

QGP data set is anonymized. All phenotypic data, including clinical, biological, physical, demographic, and family history, were obtained from the QBB (<http://www.qatarbiobank.org.qa>). All QBB participants were requested to sign a generic consent form before their

participation, allowing their samples and data to be used anonymously for research (Al Thani et al., 2019). Hamad Medical Corporation (HMC) Ethics Committee approved the QBB study protocol, and it is renewed annually by QBB Institutional Review Board (IRB). A request was made to access the QGP and QBB data (<https://www.qatarbiobank.org.qa/research/how-to-apply>), which was approved by the QBB IRB (IRB protocol number, QF-QGP-RES-PUB-002).

## 2.4 | Curation of Broad, Strict, and likely pathogenic (LP) loss of function (LOF) lists

We used the same criteria as the UK Biobank (UKB) study (Van Hout et al., 2020) to generate a high confidence Strict list and a comprehensive Broad list of potentially P variants. We compiled the variant data of 6045 WGS from the QGP data set to generate the Strict list of variants identified as P or LP, with no conflicts in interpretation and having  $\geq 2$  stars (multiple submitters provided the pathogenicity of the variant, or the variant was tested using a genetic panel) according to the review and clinical significance criteria on the NCBI ClinVar database. The comprehensive Broad list contains variants reported in the HGMD and ClinVar database, identified as disease-causing (DM) and P/LP according to HGMD and ClinVar database, respectively. Variants with a conflicting interpretation of pathogenicity between HGMD and ClinVar databases were removed from the list. All variants that fulfilled the Broad or the Strict lists classification criteria were described as P variants.

In addition to P variants, for both the Strict and Broad lists, we also included expected to be P LOF variants in the 59 ACMG genes, where according to ACMG recommendations (Green et al., 2013; Kalia et al., 2017; Richards et al., 2015), LOF is known to be the disease-causing mechanism. This list includes predicted LOF variants (stop gained, start lost, stop lost, splice donor, splice acceptor, and frameshift) that are unreported in dbSNP, ClinVar, HGMD, 1000 Genome Project, gnomAD, and GME, the sequence variations are expected to be P and have MAF  $\leq 0.001$  in QGP. We classified variants in this list as LP.

## 2.5 | Analysis of phenotypic data from QBB

All QGP participants who were recruited by QBB went into three main visits that included (1) participant questionnaires conducted by trained staff nurses and computerized self-administered questions, (2) physical and clinical measurements, and (3) collection and storing of biological samples (blood, urine, and saliva).

From the QBB (Al Thani et al., 2019), we extracted phenotypic data related to cardiovascular diseases, cancer, and familial hypercholesterolemia associated with the ACMG 59 genes (Table S2), to identify genotype-positive participant (GPP) who carry a unique variant in one of 59 ACMG genes and have supporting phenotypic information.

We have used the term "variant" to describe a unique variant in one of 59 ACMG genes and GPPs to indicate the total number of individuals (at least one) who carry a unique variant in any one of 59 ACMG genes.

## 2.6 | ACMG variants in an independent cohort from the clinic

HMC exome sequencing data set (unpublished data), from an independent cohort of clinical samples from HMC, contains 139 variants found in patients with different clinical indications for genetic testing. We extracted the variants in the 59 ACMG genes and found eight variants in seven genes, c.3920T>A (p.Ile1307Lys) in *APC*; c.650C>T (p.Thr217Ile) in *DSG2*; c.313+3A>c in *LDLR*; c.1376C>G (p.Ser459\*) in *PMS2*; c.883G>A (p.Glu295Lys) in *SCN5A*; c.340+5G>C and c.233A>G in *VHL*; and c.4065\_4068delTCAA (p.Asn1355fs) in *BRCA1*. We analyzed the phenotypic data of QGP participants who are genotype positive for the same seven variants that were found in patients from the HMC cohort.

## 2.7 | Medically actionable variants in 59 ACMG genes from reported data in Qatar

We performed an online search for the frequency of medically actionable variants in Qatar. The Jain et al. study is the only available published estimation of the frequency of medically actionable variants in the Qatari population (Jain et al., 2018). One thousand and five whole exomes and whole genomes from unrelated healthy individuals in Qatar (Fakhro et al., 2016) have been used to calculate the frequency of actionable secondary findings. Jain et al. presented a frequency of 0.59% medically actionable variants in Qatar. Four P and LP variants were identified, and three LOF variants were classified as variants of unknown significance (VUS). We compared the classification of these variants (4 P/LP and the 3 VUS) with our data and analyzed the phenotypic data of the QGP-GPPs.

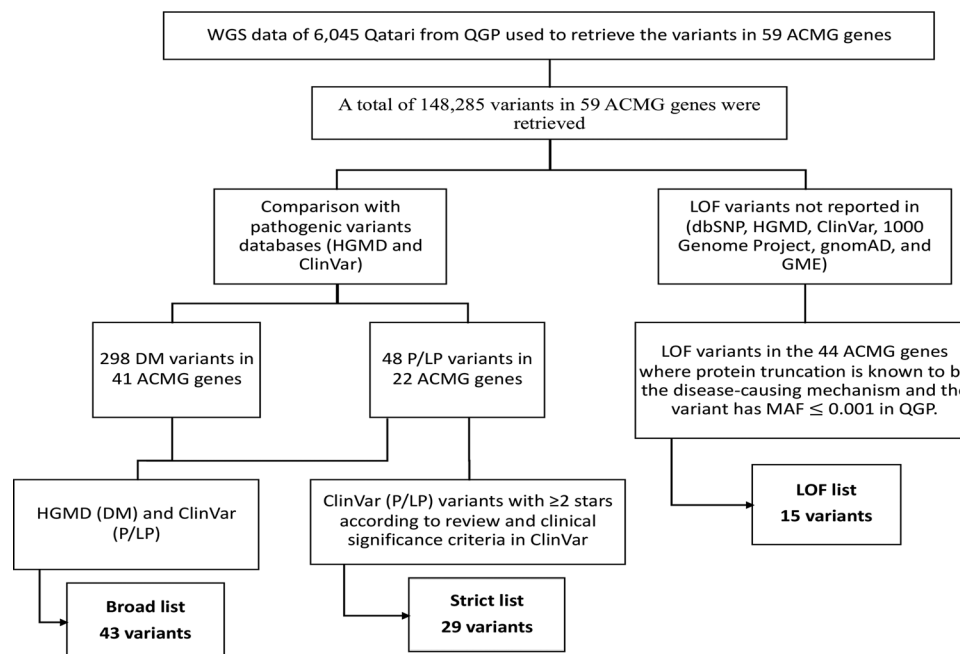
## 2.8 | Statistical analysis

To test the genotype-phenotype association of variants identified in the HMC cohort and in the Jain et al. study, we performed Fisher's exact test using GraphPad Prism (version 8.4.3). The association was described with an odds ratio (OR) and the level of significance was expressed using *p* values: \**p* < .05, \*\**p* < .01, \*\*\**p* < .001, \*\*\*\**p* < .0001.

# 3 | RESULTS

## 3.1 | Actionable P and LP variants in QGP data

We found 298 variants in 41 of the 59 ACMG genes, annotated as DM mutations in HGMD, with a median of five variants per gene.



**FIGURE 1** Flow chart of pathogenic and likely pathogenic variants classification in 59 ACMG genes identified in 6045 WGS data from the Qatar Genome Program. DM, disease-causing; GME, Great Middle East; HGMD, Human Gene Mutation Database; LOF, loss of function; LP, likely pathogenic; MAF, minor allele frequency; QGP, Qatar Genome Program; P, pathogenic

Forty-eight variants in our data set were annotated as P or LP in 22 of 59 ACMG genes, according to the review and clinical significance criteria of the ClinVar database. Thirty-eight LOF variants that were not reported before in HGMD, ClinVar, dbSNP, gnomAD Exome and WGS, 1000 Genome Project, and GME were also found in 59 ACMG genes. Out of these 38 not reported LOF variants, 15 variants were found in genes where protein truncation is known to be the disease-causing mechanism (*MUTYH*, *TNNT2*, *KCNQ1*, *WT1*, *PKP2*, *BRCA1*, *BRCA2*, *DSC2*, *LDLR*, and *COL3A1*). Furthermore, these LOF variants had  $MAF \leq 0.001$  in QGP (Table S3). Using our variant classification criteria, we identified 29 P variants in the Strict list, 43 P variants in the Broad list, and 15 LP variants in the LOF list (Figure 1 and Table 1).

Of note, two P variants (c.2449C>T [p.Arg817Trp] in *MYBPC3*, and c.1127G>A [p.Arg376His] in *SCN5A*) included in the Strict list, were not found in the Broad list, meaning that they were not classified as DM mutations in HGMD.

P and LP variants in *LDLR* were the most frequent (8 variants and 12 GPPs), followed by *ATP7B* (6 variants and 12 GPPs), *MUTYH* (5 variants and 24 GPPs), *BRCA1* (5 variants and 15 GPPs), and *BRCA2* (4 variants and 9 GPPs).

### 3.2 | P and LP cancer variants

P and LP variants associated with different cancer types were the most frequent medically actionable variants in QGP data (35%) (Figure 2a). *MUTYH* (5 variants and 24 GPPs), *BRCA1* (5 variants and 15 GPPs), *BRCA2* (4 variants and 9 GPPs), *WT1* (2 variants and

3 GPPs), *PMS2* (1 variant and 3 GPPs), *VHL* (1 variant and 2 GPPs), *MLH1* (1 variant and 1 GPP), *RET* (1 variant and 1 GPP), and *SDHB* (1 variant and 1 GPP).

Ten DM variants were identified in *BRCA2* (Table S4), of which two variants, c.9382C>T (p.Arg3128\*), and c.4211\_4215delCAAAT (p.Ser1404\*), were classified as P in the Broad and Strict lists.

We found four QGP subjects carrying the frameshift variant c.4211\_4215delCAAAT (p.Ser1404\*) in *BRCA2*, which results in an absent or disrupted protein product. Although this variant has been reported as P, we evaluated the four QGP subjects, and none of them had any personal or parent history of cancer (Table S5). Two unreported variants in *BRCA2* were classified as LP in the LOF list, c.-39-1G>C, and c.2711\_2714dupGAAA (p.Asn905Lysfs\*4). We analyzed the phenotypic data and found that the two GPPs (both males) carrying c.2711\_2714dupGAAA (p.Asn905Lysfs\*4) had a family history of cancer (one participant reported a history of a father with cancer “not specified” and the other participant reported a history of a mother with breast cancer).

We also found four QGP subjects carry the mutation c.4850C>A (p.Ser1617\*) in *BRCA1*. This variant is known to be associated with familial breast/ovarian cancer and with the hereditary cancer-predisposing syndrome (Al Hannan et al., 2019). Out of four QGP subjects, three have a family history of cancer (mother with breast cancer, a mother with endometrium cancer, and a father with bowel cancer). The variant has not been detected in the gnomAD or ExAC datasets. We found the c.4065\_4068delTCAA (p.Asn1355fs) in *BRCA1*, known to be associated with familial breast and ovarian cancer (Maxwell et al., 2017), in a clinical case of breast cancer evaluated at HMC. We found one QGP subject (carrier frequency of

TABLE 1 Medically actionable pathogenic and likely pathogenic variants identified in 6045 whole-genome sequencing data from the Qatar Genome Program

Gene	Gene ID	CHR	Locus	rs ID	Transcript	cDNA name	Protein name	QGP AF	GnomAD MAF	HGMD class	ClinVar significance	ClinVar review status	Variant list
APOB	ENS-G00000-084674	2	21229040	rs368278927	ENST00000233242	c.10700C>T	p.Thr3567Met	0.00074	0.00013	DM	Pathogenic	No assertion criteria provided	Broad
ATP7B	ENS-G00000-123191	13	52511706	rs1308732238rs121907990	ENST00000242839	c.3809A>G	p.Asn1270Ser	8.27E-05	6.6E-05	DM	Pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
ATP7B	ENS-G00000-123191	13	52511412	rs587783317	ENST00000242839	c.4021G>A	p.Gly1341Ser	8.27E-05	NA	DM	Pathogenic	Criteria provided, single submitter	Broad
ATP7B	ENS-G00000-123191	13	52532470	rs137853284	ENST00000242839	c.2332C>T	p.Arg778Trp	0.00016	NA	DM	Pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
ATP7B	ENS-G00000-123191	13	52534334	rs121908001	ENST00000242839	c.2071G>A	p.Gly691Arg	0.00016	NA	DM	Likely pathogenic	Criteria provided, single submitter	Broad
ATP7B	ENS-G00000-123191	13	52535995	rs72552285	ENST00000242839	c.1924G>C	p.Asp642His	0.00025	NA	DM	Likely pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
ATP7B	ENS-G00000-123191	13	52515322	rs755554442	ENST00000242839	c.3451C>T	p.Arg1151Cys	0.00025	NA	DM	Likely pathogenic	Criteria provided, single submitter	Broad
BRCA1	ENS-G00000-012048	17	41243479	rs80357508	ENST00000471181	c.4065_4068-delTCAA	p.Asn1355fs	8.27E-05	NA	DM	Pathogenic	Reviewed by expert panel	Strict and Broad
BRCA1	ENS-G00000-012048	17	41223144	rs80357429	ENST00000471181	c.4850C>A	p.Ser1617*	0.00033	NA	DM	Pathogenic	Reviewed by expert panel	Strict and Broad

TABLE 1 (Continued)

Gene	Gene ID	CHR	Locus	rs ID	Transcript	cDNA name	Protein name	QGP AF	GnomAD MAF	HGMD class	ClinVar significance	ClinVar review status	Variant list
BRCA1	ENS- G00000- 012048	17	41231418	NA	ENST00000471181	c.4358-2A>T	NA	0.00025	NA	NA	NA	NA	LP LOF
BRCA1	ENS- G00000- 012048	17	41243451	NA	ENST00000471181	c.4096+1G>C	NA	0.00025	NA	NA	NA	NA	LP LOF
BRCA1	ENS- G00000- 012048	17	41246182	NA	ENST00000471181	c.1365dupT	p.Ile456Tyrfs*2	0.00033	NA	NA	NA	NA	LP LOF
BRCA2	ENS- G00000- 139618	13	32968951	rs80359212	ENST00000380152	c.9382C>T	p.Arg3128*	8.27E-05	0.00023	DM	Pathogenic	Reviewed by expert panel	Strict and Broad
BRCA2	ENS- G00000- 139618	13	32912701	rs786203340	ENST00000380152	c.4211_4215- delCAAAAT	p.Ser1404*	0.00033	NA	DM	Pathogenic	Reviewed by expert panel	Strict and Broad
BRCA2	ENS- G00000- 139618	13	32890558	NA	ENST00000380152	c.-39-1G>C	NA	0.00016	NA	NA	NA	NA	LP LOF
BRCA2	ENS- G00000- 139618	13	32911202	NA	ENST00000380152	c.2711_2714- dupGAAA	p.Asn905- Lysfs*4	0.00016	NA	NA	NA	NA	LP LOF
COL3- A1	ENS- G00000- 168542	2	1898589- 82	NA	ENST00000404636	c.1218delT	p.Pro407- Leufs*3	8.27E-05	NA	NA	NA	NA	LP LOF
DSC2	ENS- G00000- 134755	18	28669534	NA	ENST00000280904	c.497delA	p.Asn166Thrfs*8	0.00016	NA	NA	NA	NA	LP LOF
FBN1	ENS- G00000- 166147	15	48725078	rs779749926	ENST00000316623	c.6724C>T	p.Arg2242Cys	8.27E-05	NA	DM	Likely pathogenic	Criteria provided, single submitter	Broad
KCNH2	ENS- G00000- 055118	7	1506525- 69	rs972201049	ENST00000330883	c.23C>T	p.Ala8Val	8.27E-05	NA	DM	Likely pathogenic	Criteria provided, single submitter	Broad

(Continues)



TABLE 1 (Continued)

Gene	Gene ID	CHR	Locus	rs ID	Transcript	cDNA name	Protein name	QGP AF	GnomAD MAF	HGMD class	ClinVar significance	ClinVar review status	Variant list
KCNQ1	ENS-G00000-053918	11	2591949	rs120074178	ENST00000155840	c.569G>A	p.Arg190Gln	8.27E-05	NA	DM	Pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
KCNQ1	ENS-G00000-053918	11	2594209	rs120074186	ENST00000155840	c.914G>T	p.Trp305Leu	0.00041	NA	DM	Pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
KCNQ1	ENS-G00000-053918	11	2482945	NA	ENST000000335475	c.5+1G>A	NA	0.00066	NA	NA	NA	NA	LP LOF
LDLR	ENS-G00000-130164	19	11200090	rs879254375	ENST00000557933	c.-135C>G	NA	8.27E-05	NA	DM	Pathogenic/likely pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
LDLR	ENS-G00000-130164	19	11213418	rs771019366	ENST00000557933	c.269A>G	p.Asp90Gly	0.00025	NA	DM	Pathogenic/likely pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
LDLR	ENS-G00000-130164	19	11226874	rs758194385	ENST00000557933	c.1691A>G	p.Asn564Ser	8.27E-05	NA	DM	Likely pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
LDLR	ENS-G00000-130164	19	11222274	rs752951310	ENST00000557933	c.1145G>T	p.Gly382Val	8.27E-05	NA	DM	Likely pathogenic	Criteria provided, single submitter	Broad
LDLR	ENS-G00000-130164	19	11222283	rs879254809	ENST00000557933	c.1154T>G	p.Leu385Arg	8.27E-05	NA	DM	Likely pathogenic	Criteria provided, single submitter	Broad
LDLR		19	11227603	rs763147599	ENST00000557933	c.1774G>A	p.Gly592Arg	0.00016	NA	DM			



TABLE 1 (Continued)

Gene	Gene ID	CHR	Locus	rs ID	Transcript	cDNA name	Protein name	QGP AF	GnomAD MAF	HGMD class	ClinVar significance	ClinVar review status	Variant list
LDLR	ENS- G00000-130164	19	11210922	rs776421777	ENST00000557933	c.91G>A	p.Glu31Lys	8.27E-05	6.66E-05	DM	Likely pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
LDLR	ENS- G00000-130164	19	11201365	NA	ENST00000252444	c.67+1074C>-A	NA	0.00016	NA	NA	NA	NA	LP LOF
LMNA	ENS- G00000-160789	1	1561067-75	rs57920071	ENST00000368300	c.1444C>T	p.Arg482Trp	8.27E-05	NA	DM	Pathogenic	Criteria provided, single submitter	Broad
MLH1	ENS- G00000-076242	3	37042544	rs63751665	ENST00000231790	c.306G>T	p.Glu102Asp	8.27E-05	NA	DM	Likely pathogenic	Reviewed by expert panel	Strict and Broad
MUTY-H	ENS- G00000-132781	1	45797348	rs587783057	ENST00000450313	c.1171C>T	p.Gln391*	0.00016	NA	DM	Pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
MUTY-H	ENS- G00000-132781	1	45800165	rs587780088	ENST00000450313	c.55C>T	p.Arg19*	0.00016	NA	DM	Pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
MUTY-H	ENS- G00000-132781	1	45798475	rs34612342	ENST00000450313	c.536A>G	p.Tyr179Cys	0.00033	0.0023	DM	Pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
MUTY-H	ENS- G00000-132781	1	45796890	rs587778541	ENST00000450313	c.1437_1439-delGGA	p.Glu480del	0.0012	0.00011	DM	Pathogenic	Criteria provided, multiple	Strict and Broad

(Continues)

TABLE 1 (Continued)

Gene	Gene ID	CHR	Locus	rs ID	Transcript	cDNA name	Protein name	QGP AF	GnomAD MAF	HGMD class	ClinVar significance	ClinVar review status	Variant list
MUTY-H	ENS-G00000-132781	1	45799181	NA	ENST00000450313	c.252C>A	p.Tyr84*	8.27E-05	NA	NA	NA	submitters, no conflicts	LP LOF
MYBP-C3	ENS-G00000-134571	11	47359095	rs727503188	ENST00000399249	c.2449C>T	p.Arg817Trp	8.27E-05	NA	DM?	Pathogenic/ likely pathogenic	Criteria provided, multiple submitters, no conflicts	Strict
MYBP-C3	ENS-G00000-134571	11	47359343	rs1313107874r-s371488302	ENST00000399249	c.2311G>A	p.Val771Met	8.27E-05	NA	DM	Likely pathogenic	Criteria provided, single submitter	Broad
MYBP-C3	ENS-G00000-134571	11	47371575	rs730880619	ENST00000256993	c.495G>C	p.Glu165Asp	0.00041	6.67E-05	DM	Likely pathogenic	Criteria provided, single submitter	Broad
MYH7	ENS-G00000-092054	14	23887522	rs727503246	ENST00000355349	c.4066G>A	p.Glu1356Lys	8.27E-05	NA	DM	Likely pathogenic	Reviewed by expert panel	Strict and Broad
MYH7	ENS-G00000-092054	14	23894049	rs138049878	ENST00000355349	c.2608C>T	p.Arg870Cys	8.27E-05	NA	DM	Likely pathogenic	Reviewed by expert panel	Strict and Broad
PKP2	ENS-G00000-057294	12	33049514	rs397516997	ENST00000070846	c.148_151del-IACAG	p.Thr50Serf-s*61	0.00041	NA	DM	Pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
PKP2	ENS-G00000-057294	12	33030878	NA	ENST00000070846	c.925_935del-GCCGTG-GATTC	p.Ala309Glnf-s*23	8.27E-05	NA	NA	NA	LP LOF	LP LOF
PMS2	ENS-G00000-122512	7	6027020	rs587780724	ENST00000265849	c.1376C>G	p.Ser459*	0.00025	NA	DM	Pathogenic	Criteria provided, multiple	Strict and Broad

TABLE 1 (Continued)

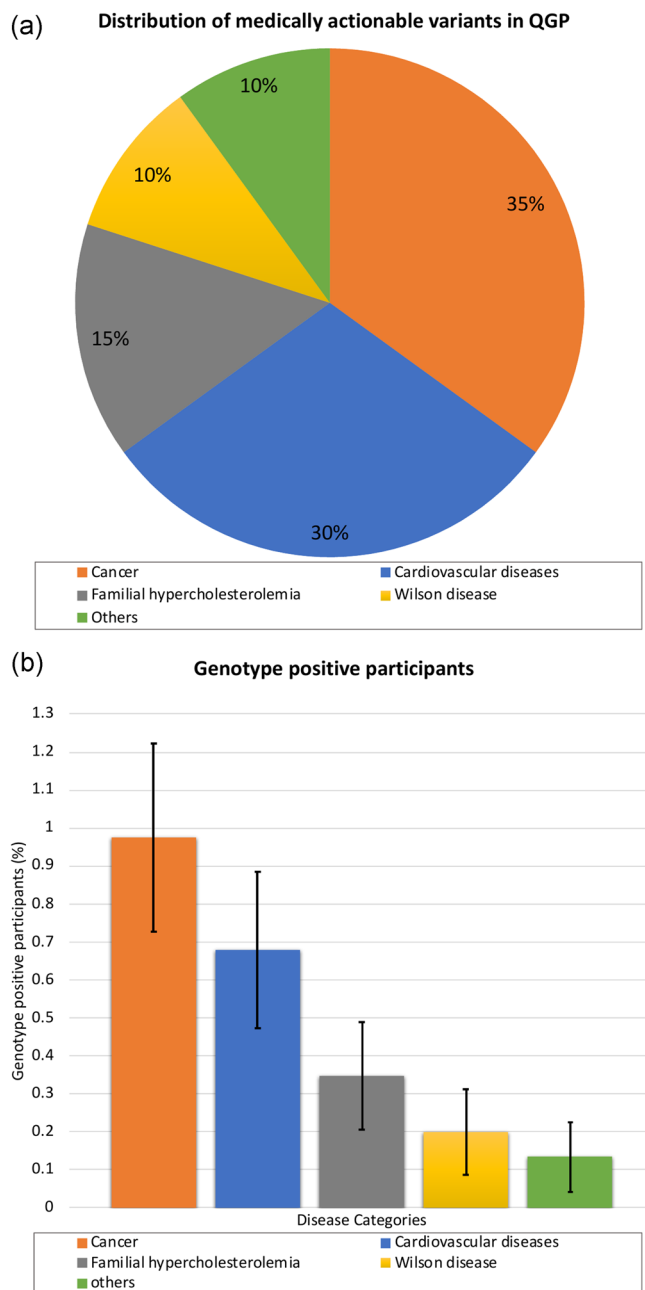
Gene	Gene ID	CHR	Locus	rs ID	Transcript	cDNA name	Protein name	QGP AF	GnomAD MAF	HGMD class	ClinVar significance	ClinVar review status	Variant list
RET	ENS-G00000-165731	10	43614996	rs79658334	ENST000000355710	c.2410G>A	NA	8.27E-05	0.0012	DM	Pathogenic/likely pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
RYR1	ENS-G00000-196218	19	38946103	rs111888148	ENST000000359596	c.1589G>A	p.Arg530His	8.27E-05	6.68E-05	DM	Pathogenic/likely pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
RYR1	ENS-G00000-196218	19	38958433	rs756138074	ENST000000359596	c.3362A>G	p.Tyr1121Cys	8.27E-05	NA	DM	Likely pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
RYR1	ENS-G00000-196218	19	38990601	rs118192174	ENST000000359596	c.7268T>A	p.Met2423Lys	8.27E-05	NA	DM	Pathogenic	No assertion criteria provided	Broad
RYR1	ENS-G00000-196218	19	39013851	rs193922837	ENST000000359596	c.10348-6C>G	NA	0.00025	0.00046	DM	Pathogenic/likely pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
SCN5A	ENS-G00000-183873	3	38622757	rs1045414916 r- s199473180	ENST000000333535	c.2893C>T	p.Arg965Cys	0.00016	NA	DM	Pathogenic	Criteria provided, single submitter	Broad
SCN5A	ENS-G00000-183873	3	38648173	rs199473101	ENST000000333535	c.1127G>A	p.Arg376His	0.00025	6.67E-05	DM?	Pathogenic/likely pathogenic	Criteria provided, multiple submitters, no conflicts	Strict
SDHB	ENS-G00000-117118	1	17371313	rs202101384	ENST000000375499	c.143A>T	p.Asp48Val	8.27E-05	NA	DM	Likely pathogenic	Criteria provided,	Broad

(Continues)

TABLE 1 (Continued)

Gene	Gene ID	CHR	Locus	rs ID	Transcript	cDNA name	Protein name	QGP AF	GnomAD MAF	HGMD class	ClinVar significance	ClinVar review status	Variant list
single submitter													
TNNI3	ENS-G00000-129991	19	55665525	rs942457087 r- s3977516347	ENST000000344887	c.422G>A	p.Arg141Gln	8.27E-05	0.00011-457	DM	Pathogenic/ likely pathogenic	Criteria provided, multiple submitters, no conflicts	Strict and Broad
TNNI2	ENS-G00000-118194	1	2013311-47	NA	ENST00000458432	c.610G>T	p.Glu204*	8.27E-05	NA	NA	NA	NA	LP LOF
TNNI2	ENS-G00000-118194	1	2013359-64	NA	ENST00000458432	c.239+2T>C	NA	8.27E-05	NA	NA	NA	NA	LP LOF
VHL	ENS-G00000-134086	3	10191555	rs5030823	ENST00000256474	c.548C>G	p.Ser183Trp	0.00016	NA	DM	Likely pathogenic	No assertion criteria provided	Broad
WT1	ENS-G00000-184937	11	32438088	NA	ENST00000332351	c.951-2A>G	NA	8.27E-05	NA	NA	NA	NA	LP LOF
WT1	ENS-G00000-184937	11	32452075	NA	ENST00000379079	c.10+1G>A	NA	0.00016	NA	NA	NA	NA	LP LOF

Abbreviations: AF, allele frequency; CHR, chromosome; DM, disease-causing mutation; DM?, likely disease-causing mutation; HGMD, Human Gene Mutation Database; LP LOF, likely pathogenic loss of function list; QGP, Qatar Genome Program; WGS, whole-genome sequence; NA, not applicable; MAF, maximum allele frequency.



**FIGURE 2** The distribution of ACMG pathogenic and likely pathogenic medically actionable variants in Qatar genome data. (a) Frequency of pathogenic and likely pathogenic variants in 59 ACMG genes among different diseases categories. (b) Frequency of QGP participants who carry pathogenic and likely pathogenic variants based on different diseases categories. Error bars represent a 95% confidence interval. Cardiovascular diseases include arrhythmogenic right ventricular cardiomyopathy type 9, Brugada syndrome 1, Dilated cardiomyopathy 1A, Familial hypertrophic cardiomyopathy 1 and 7, Long QT syndrome 1 and 2. Cancer includes Lynch syndrome, Multiple endocrine neoplasia, type 2a and 2b, MYH-associated polyposis, Paragangliomas 4, Von Hippel-Lindau syndrome, Breast-ovarian cancer, familial 1 and 2. Others include: Malignant hyperthermia, Ehlers-Danlos syndrome, and Marfan's syndrome. ACMG, American College of Medical Genetics and Genomics; QGP, Qatar Genome Program

0.008%) carries c.4065\_4068delTCAA (p.Asn1355fs) in *BRCA1*. This subject, had a maternal history of endometrium cancer.

We found three QGP subjects carrying the variant c.1376C>G (p.Ser459\*) in the *PMS2* gene. This variant is expected to result in an absent or disrupted *PMS2* protein product. This variant is not present in population databases (ExAC and gnomAD). c.1376C>G (p.Ser459\*) has been reported in individuals affected with T-cell lymphoblastic lymphoma (Oberg et al., 2016). Out of three QGP subjects, we found that two heterozygote subjects have a positive family history of lung cancer (Table S5).

### 3.3 | P and LP cardiovascular variants

Thirty percent of the identified P and LP variants were found in genes related to cardiovascular diseases, *KCNQ1* (3 variants and 14 GPPs), *MYBPC3* (3 variants and 7 GPPs), *PKP2* (2 variant and 6 GPPs), *SCN5A* (2 variants and 5 GPPs), *MYH7* (2 variants and 2 GPPs), *TNNT2* (2 variants and 2 GPPs), *DSC2* (1 variant and 2 GPPs), *LMNA* (1 variant and 1 GPP), *KCNH2* (1 variant and 1 GPP), and *TNNI3* (1 variant and 1 GPP).

DM variants in *SCN5A* were the most prevalent HGMD DM variants in QGP data (37 DM variants) (Table S4). Only two variants in *SCN5A* were reported in our lists, the c.2893C>T (p.Arg965Cys) in the Broad list and c.1127G>A (p.Arg376His) in the Strict list.

DM variants in *MYBPC3* and *KCNQ1* were also frequent in the HGMD DM variants list in QGP data. We identified 18 DM variants in *MYBPC3*. Three P variants, c.2311G>A (p.Val771Met), c.495G>C (p.Glu165Asp), and c.2449C>T (p.Arg817Trp) in *MYBPC3*, associated with familial hypertrophic cardiomyopathy, were reported as P variants in our lists. We found that five GPPs carried c.495G>C (p.Glu165Asp) in *MYBPC3*. By analyzing the phenotypic data of GPPs for c.495G>C (p.Glu165Asp), we found that one QGP subject, aged 33 years, had an abnormal electrocardiogram (ECG) report (Sinus rhythm with first degree AV block), two subjects had borderline ECG, and one of them had parents with a history of heart disease (both, mother and father) (Table S6).

We identified 14 DM variants in the *KCNQ1* gene in the QGP participants; two of them, c.569G>A (p.Arg190Gln) and c.914G>T (p.Trp305Leu), were reported in the pathogenic Strict and Broad lists. Detailed phenotypic analysis of six GPP who carried these two variants is described in Table S6. One previously not reported variant c.5+1G>A in *KCNQ1* was reported as LP in the LOF list. Out of the eight QGP subjects who carried this variant, one had an abnormal ECG (Sinus rhythm with occasional ventricular premature complex), one had a borderline ECG, and one subject had a history of sibling diagnosed with a heart attack (Table S6).

We found five QGP subjects carrying the c.148\_151delACAG (p.Thr50Serfs\*61) mutation in the *PKP2* gene in the Strict list. We evaluated the phenotypic data of the five heterozygous QGP subjects who carried this variant. We found one of them had abnormal ECG (right bundle branch block), one had borderline ECG, and one with

parents with a history of heart disease (Table S6). One GPP (31 years old) was found to carry the previously unreported frameshift variant c.925\_935delGCCGTGGATTC (p.Ala309Glnfs\*23) in *PKP2*. We evaluated the phenotypic data of this GPP and found an abnormal ECG report (Sinus bradycardia with arrhythmia and marked right axis deviation) and self-reported history of a father with heart disease.

### 3.4 | P and LP variants associated with familial hypercholesterolemia

Seven P variants and one LP variant were identified in the *LDLR* gene associated with familial hypercholesterolemia. In total, 12 GPPs carried P and LP variants in the *LDLR* gene. Ten out of 12 GPPs who had P and LP variants in *LDLR* had either abnormal lipid profiles or self-reported medical history of high cholesterol (Table S7).

We identified one P variant in the *APOB* gene, c.10700C>T (p.Thr3567Met), in nine heterozygous participants. Three out of these nine GPPs were found to have self-reported that they had been on treatment for high cholesterol.

### 3.5 | Other diseases

We identified P and LP variants in genes associated with other diseases, such as Wilson disease (MIM# 277900) (6 variants in *ATP7B*), Malignant hyperthermia (MIM# 145600) (4 variants in *RYR1*), Ehlers-Danlos syndrome (MIM# 130050) (1 variant in *COL3A1*), and Marfan's syndrome (MIM# 154700) (1 variant in *FBN1*), in 20 GPPs in QGP; however, we did not find related phenotypic data for these diseases in QBB.

### 3.6 | ACMG variants in HMC exome data and QGP phenotypes

Out of the eight variants in 59 ACMG genes found in HMC exome sequencing data, seven variants were found in QGP data. We analyzed the phenotype-genotype association of QGP participants who are genotype positive for the same seven variants that have been found in patients from HMC. We found 509 QGP subjects (489 heterozygotes and 20 homozygotes), (QGP carrier frequency of 4.3%) had the c.3920T>A (p.Ile1307Lys) variant in *APC*. The HMC patient carrying this variant had a personal and a family history of colon cancer. We found 67 QGP participants had personal or family clinical histories of different cancer types ( $p = .2799$ , OR = 1.160). Out of these 67 QGP subjects, 19 GPPs reported personal or familial histories of colorectal and gastric cancers ( $p = .7016$ , OR = 1.103). c.3920T>A (p.Ile1307Lys) has been detected in gnomAD at an allele frequency of 3.6% in Ashkenazi Jewish individuals but only 0.06% in other population groups. c.3920T>A (p.Ile1307Lys) was initially identified in Ashkenazi Jewish subjects but has also been shown to be present in other population groups (Shtoyerman-Chen et al., 2001). This variant is associated with a 2.5-fold risk of colon cancer and other cancers

in several population groups (Leshno et al., 2016). The high risk of cancer in subjects carrying this mutation is thought to be due to an increase in the length of a polyA tract in the risk allele, which makes it prone to polymerase slippage and to the generation of frameshift variants in dividing cells (Ghatak et al., 2017). In our cohort, we found slightly increased OR, but it is not statistically significant. Therefore we could not confirm the pathogenicity of this variant.

The c.883G>A (p.Glu295Lys) variant in *SCN5A* was detected in homozygosity in one clinical case with Brugada syndrome and abnormal cardiac rhythm in the HMC exome sequencing data set. We found nine GPPs carrying c.883G>A (p.Glu295Lys). Out of these, two had abnormal ECG (one subject had sinus tachycardia with possible left atrial enlargement, and another subject had ST elevation with normally inflected T wave), five had borderline ECG ( $p = .1794$ , OR = 3.495), and three had a parent with a history of heart disease ( $p = >.9999$ , OR = 0.7975). The variant was detected only once in gnomAD ( $T = 0.00000$  (1/245,968)).

We have found that six QGP subjects (carrier frequency of 0.05%) carried the c.313+3A>C variant in *LDLR*. This mutation is predicted to cause abnormal splicing by destroying the splice donor site in intron three. The variant was not observed in the NHLBI Exome Sequencing Project, indicating it is not a common benign variant in European and African American populations (Do et al., 2015). This variant has been detected in two patients at HMC that presented with familial hypercholesterolemia. We found six GPPs carrying c.313+3A>C in heterozygosity in QGP data. All six QGP participants had a significant self-reported history of high cholesterol ( $***p = .0009$ , OR = Inf.), and five participants also reported treatment for high cholesterol (one participant on a diet only and four participants on diet and medication) ( $*p = .0120$ , OR = 11.37).

c.650C>T (p.Thr217Ile) in *DSG2* was identified in heterozygosity in a clinical case at HMC that presented with recurrent cardiopulmonary arrest, developmental delay, hypotonia, and thinning of the corpus callosum. The parents of this case were unaffected. We found 34 QGP-GPPs carrying c.650C>T (p.Thr217Ile) without any personal or family history of heart disease.

Two HMC clinical cases of Von Hippel Lindau syndrome were found to carry c.233A>G and c.340+5G>C variants in *VHL*. We could not identify any QGP-GPP for c.233A>G; however, we found 74 QGP subjects (carrier frequency of 0.6%) carrying c.340+5G>C. This intron splice mutation in the first intron of *VHL* was not expected to have clinical significance because it was identified in 4.5% (9/200) Southern Han Chinese chromosomes by the 1000 Genomes Project (Olfson et al., 2015). It has been detected in gnomAD at an allele frequency of 0.0021. Among the 74 QGP subjects carrying c.340+5G>C, 16 had a personal or family history of cancer. However, the results were not statistically significant ( $p = .1995$ , OR = 1.460), and it is not expected to have clinical significance.

Finally, we found one HMC clinical case carrying the P variant c.1376C>G (p.Ser459\*) in *PMS2* and one HMC clinical case carry the P variant c.4065\_4068delTCAA (p.Asn1355fs) in *BRCA1*. These two variants were reported as P variants in our Broad and Strict lists, and we have provided phenotypic data for these in the cancer section (Table S5).



### 3.7 | Reported medically actionable data in Qatar and QGP phenotypes

Four P and LP variants were reported in Jain et al. study in 1005 whole exomes and genomes from Qatar (Jain et al., 2018). These variants are c.14524G>A (p.Val4842Met) in *RYR1*; c.619G>A (p.Val207Met) in *KCNQ1*; c.4018G>A (p.Val1340Ile) in *SCN5A*; and c.536A>G (p.Tyr179Cys) in *MUTYH*. Out of four P/LP variants in Jain et al.'s study, only c.536A>G (p.Tyr179Cys) in *MUTYH* were reported in our data as a P variant.

We found three heterozygous QGP subjects carrying the c.619G>A (p.Val207Met) variant in *KCNQ1*, which was annotated as LP in 1005 whole exomes and genomes from Qatar (Jain et al., 2018). This variant is classified as likely benign in the ClinVar database. All three subjects carrying *KCNQ1* c.619G>A (p.Val207Met) had normal ECG.

We found two heterozygous GPP carrying the variant c.4018G>A (p.Val1340Ile) in *SCN5A*. This variant was annotated as LP in 1005 whole exomes and genomes from Qatar (Jain et al., 2018). The two QGP subjects carrying this variant had borderline ECG alterations (20-year-old female with marked sinus arrhythmia and 32-year-old male with sinus bradycardia) ( $p = .1239$ , OR = Infinity).

Five QGP-GPPs carried the variant c.14524G>A (p.Val4842Met) in *RYR1* in heterozygosity, which is classified as P in 1005 whole exomes and genomes from Qatar. We did not find related phenotypic data for Malignant hyperthermia associated with *RYR1* mutations in the QBB.

Three LOF variants in Jain et al.'s study classified as VUS were observed, c.2037C>A (p.Cys679\*) in *PCSK9*, c.9976A>T (p.Lys3326\*) in *BRCA2*, and c.58C>T (p.Gln20\*) in *TPM1*.

We have found 40 QGP subjects (MAF of 0.003 in QGP) carrying the stop-gain variant c.9976A>T (p.Lys3326\*) in *BRCA2*. The variant was annotated as VUS in the data of 1005 whole exomes and genomes from Qatar (Jain et al., 2018). We found one genotype-positive QGP subject carrying c.9976A>T (p.Lys3326\*) had breast cancer, and 9 QGP subjects having parents with upper aerodigestive cancers ( $p = .1279$ , OR = 1.763).

We found that 15 QGP subjects carry the stop gain mutation c.58C>T (p.Gln20\*) in the *TPM1* gene. c.58C>T (p.Gln20\*) was annotated as VUS in the 1005 whole exomes and genomes from Qatar due to the absence of clinical information (Jain et al., 2018). We have found here that 8 of these 15 QGP subjects carrying c.58C>T (p.Gln20\*) have abnormal ECG data (\*\* $p = .0006$ , OR = 6.592), five have fathers with a history of heart disease, and four have mothers with heart disease ( $p = .1971$ , OR = 2.055). Interestingly, c.58C>T (p.Gln20\*) has rarely been identified in large population data sets (1000 Genomes, and ExAC), with an allele frequency of 0.0003.

Finally, four heterozygous QGP-GPPs carrying c.2037C>A (p.Cys679\*) in *PCSK9*. c.2037C>A (p.Cys679\*) was reported as VUS in the 1005 whole exomes and genomes from Qatar. We evaluated the phenotypic data of the four QGP-GPPs, and none of them had abnormal lipid profiles or self-reported histories of hypercholesterolemia.

## 4 | DISCUSSION

The availability of a large WGS data set from the QGP first phase prompted us to study the frequency of medically actionable variants in the Arabic population, which is mostly uncharacterized due to the lack of large population-based genomic data from the Middle East. We examined 6045 WGS samples from QGP to apply stringent criteria to identify Broad and Strict lists of P variants, as well as LP LOF list. Our results identified a total of 60 P and LP variants in 25 ACMG genes in 141 unique individuals. Overall, 2.3% of the QGP sequenced participants carried a P or LP variant in one of the 59 ACMG genes. The 2.3% frequency represents nonredundant GPPs for P and LP variants in Broad, Strict, and LOF lists. However, one QGP participant was found to carry the two P variants c.3362A>G (p.Tyr1121Cys) in *RYR1* and c.1437\_1439delGGA (p.Glu480del) in *MUTYH* in heterozygosity.

Only six P variants in our analysis (c.148\_151delACAG (p.Thr50Serfs\*61) in *PKP2*; c.2608C>T (p.Arg870Cys) in *MYH7*; c.4065\_4068delTCAA (p.Asn1355fs) in *BRCA1*; c.1691A>G (p.Asn564Ser) in *LDLR*; c.3362A>G (p.Tyr1121Cys) in *RYR1*; and c.422G>A (p.Arg141Gln) in *TNNI3* overlap with P variants in the UKB study. Contrary to expectation, no P or LP variant in the *ATP7B* gene was identified in the UKB study. However, heterozygous GPPs for P variants in *ATP7B*, associated with Wilson disease, were the second most frequent in the QGP data set. This result emphasizes the importance of evaluating the population specificity for medically actionable variants.

All P and LP variants were identified in genes responsible for dominant conditions, meaning that heterozygous GPPs are at risk, except five P and LP variants in *MUTYH* inherited as autosomal recessive were found in 24 heterozygous GPPs.

Among the 60 P and LP variants identified in our lists, variants associated with cancer were the most frequent (35%), which is consistent with the high prevalence of this disease in the Qatari population (Azamjah et al., 2019; Bener et al., 2008; Chaabna et al., 2018; Nair et al., 2008).

A total of 59 QGP-GPPs were found to have P or LP variants in cancer-related genes. One of the findings in this study is the variant c.3920T>A (p.Ile1307Lys) in the *APC* gene. This variant has been reported at high frequency and was classified as a risk allele for colorectal cancer among Ashkenazi Jewish (Shtoyerman-Chen et al., 2001; Woodage et al., 1998). Other studies of c.3920T>A (p.Ile1307Lys) in different populations did not report an increased risk of the hereditary cancer-predisposing syndrome (Liang et al., 2013). We have found a clinical case of colon cancer evaluated at HMC carrying c.3920T>A (p.Ile1307Lys). However, the statistical analysis of GPPs in the QGP showed no significant association with cancer susceptibility in the Qatari population.

c.9976A>T (p.Lys3326\*) in *BRCA2* has been reported to increase the risk of (OR = 2.53) upper aerodigestive tract cancers (Delahaye-Sourdeix et al., 2015). The frequency of c.9976A>T (p.Lys3326\*) in the global population goes against this variant's pathogenicity. The pathogenicity of c.9976A>T (p.Lys3326\*) was reevaluated in a cohort



of breast, ovarian, and prostate cancer cases against non-cancer controls (Meeks et al., 2016). The results confirmed the association of c.9976A>T (p.Lys3326\*) with an increased risk of breast cancer and ovarian cancer.

We detected one GPP carrying the P variant c.143A>T (p.As- p48Val) in gene *SDHB*, which is known to be associated with hereditary paraganglioma-pheochromocytoma syndrome. This syndrome is a very rare condition in which early detection of tumors can minimize complications.

0.68% of QGP participants carry P or LP variants associated with different types of cardiovascular conditions. DM variants in *SCN5A* were the most frequent in our results. However, most variants classified in HGMD as DM in the *SCN5A* gene did not meet our criteria to be reported as P variants. Given that some of these cardiac conditions, such as Brugada syndrome, might only be diagnosed as a result of sudden death, early screening and intervention can be crucial. Brugada syndrome is responsible for 12% of sudden cardiac death cases in the general population (Juang & Huang, 2004), most of the cases are asymptomatic and the age of diagnosis is usually around 40 (Probst et al., 2012). P variants in *SCN5A* are the most common inherited cause of Brugada syndrome (Li et al., 2018; Monasky et al., 2019). In the literature, different clinical findings within the same family could be observed in related patients with Brugada syndrome (Nakajima et al., 2013). This could indicate high genetic heterogeneity and phenotypic diversity associated with Brugada syndrome. The variant c.883G>A (p.Glu295Lys) in *SCN5A* (MAF of 0.0007 in QGP) was detected in homozygosity in a case of Brugada syndrome at HMC. c.883G>A (p.Glu295Lys) has not previously been linked with *SCN5A* associated diseases. In a recent study carried out by Jain et al., genetic variants in 1005 exome sequenced and genome sequenced data from Qatar were studied to analyze secondary findings in 59 ACMG genes (Jain et al., 2018). The dissimilarities in the method make the comparison difficult. However, variant c.4018G>A (p.Val1340Ile) in *SCN5A* was annotated as LP in the Jain study, and we found that the two QGP subjects carrying the variant had borderline ECG.

In our analysis, we identified five QGP subjects carrying the c.148\_151delACAG (p.Thr50Serfs\*61) mutation in the *PKP2*. This known PP mutation is predicted to cause a shift in the reading frame beginning at codon 50 to create a premature stop codon and a truncated protein. In addition, marked desmosome abnormalities and accumulation of lipid droplets in cardiac myocytes have been identified (Caspi et al., 2013). By evaluating the phenotypic data of the five heterozygous QGP subjects carrying c.148\_151delACAG (p.Thr50Serfs\*61), we have found that none had personal or family history of heart disease.

The availability of phenotypic data related to familial hypercholesterolemia in the QBB allowed us to identify QGP participants with abnormal lipid profiles or a medical history of high cholesterol. Twenty-one GPPs carry P or LP variants in *ABOP* and *LDLR*, out of which 13 QGP participants have either abnormal lipid profiles or reported medical history of high cholesterol. Early diagnosis and treatment of individuals suspected to have familial

hypercholesterolemia can decrease coronary heart disease events and mortality associated with familial hypercholesterolemia.

An important limitation of this study is that the QBB informed consent does not allow the return of results to participants, but this policy is under revision (Al Thani et al., 2019). The ability to return secondary findings with potentially lifesaving interventions is indispensable to help the patients, family members, and the clinicians to identify the best interventions and medical follow-up strategies, especially in populations with high consanguinity rates, and to understand the penetrance and expressivity of the identified medically actionable variants in different populations. There are also other limitations consistent with our method of variant prioritization. For example, we filtered QGP data based on HGMD and ClinVar Professional entries, and this may exclude predicted P variants that have not been classified as DM or P/LP in these databases, respectively. Furthermore, interpretation of the variant should be given in a clinical context, but the lack of sufficient phenotypic data or links to participants' medical records precluded the diagnosis of the clinical cases. Consequently, variant calls could not be verified based on clinical phenotypes. All these limitations could lead to an underestimation of the prevalence and frequency of medically actionable variants.

Efforts to share knowledge in the genetics society through centralized databases that include phenotypic data should improve the understanding of secondary findings and provide additional information to assess the role of very rare variants in most populations. Our study is a systematic analysis of the available information from Qatar to identify clinically relevant secondary findings, and this framework can be adjusted or replicated as knowledge of genetic diseases expand and guidelines regarding the return of secondary findings continue to improve.

## ACKNOWLEDGMENTS

The authors are grateful to all the researchers in Qatar Genome Program and Qatar Biobank's team for sharing genome sequence data and phenotypic information, without which this analysis would not have been possible. Qatar Biobank and Qatar Genome Program are both Research, Development and Innovation entities within Qatar Foundation for Education, Science and Community Development. Open Access funding provided by the Qatar National Library.

## The Qatar Genome Program Research (QGPR) Consortium

Qatar Genome Project Management: Said Ismail (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Wadha Al-Muftah (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Radja Badji (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Hamdi Mbarek (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Dima Darwish (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Tasnim Fadl (Qatar Genome Program,

Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Heba Yasin (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Maryem Ennaifar (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Rania Abdel-Atif (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Fatima Alkuwari (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Muhammad Alvi (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Yasser Al Sarraj (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Chadi Saad (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Asmaa Althani (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar; Qatar Biobank for Medical Research, Qatar Foundation, Doha, Qatar).

Biobank and Sample Preparation: Eleni Fethnou (Qatar Biobank for Medical Research, Qatar Foundation, Doha, Qatar), Fatima Qafoud (Qatar Biobank for Medical Research, Qatar Foundation, Doha, Qatar), Eiman Alkhayat (Qatar Biobank for Medical Research, Qatar Foundation, Doha, Qatar), Nahla Afifi (Qatar Biobank for Medical Research, Qatar Foundation, Doha, Qatar).

Sequencing and Genotyping group: Sara Tomei (Integrated Genomics Services, Sidra Medicine, Doha, Qatar), Wei Liu (Integrated Genomics Services, Sidra Medicine, Doha, Qatar), Stephan Lorenz (Integrated Genomics Services, Sidra Medicine, Doha, Qatar).

Applied Bioinformatics Core: Najeeb Syed (Applied Bioinformatics Core, Sidra Medicine, Doha, Qatar), Hakeem Almabrazi (Applied Bioinformatics Core, Sidra Medicine, Doha, Qatar), Fazulur Rehaman Vempalli (Applied Bioinformatics Core, Sidra Medicine, Doha, Qatar), Ramzi Temanni (Applied Bioinformatics Core, Sidra Medicine, Doha, Qatar).

Data Management and Computing Infrastructure group: Tariq Abu Saqri (Biomedical Informatics, Sidra Medicine, Doha, Qatar), Mohammedhusen Khatib (Biomedical Informatics, Sidra Medicine, Doha, Qatar), Mehshad Hamza (Biomedical Informatics, Sidra Medicine, Doha, Qatar), Tariq Abu Zaid (Biomedical Informatics, Sidra Medicine, Doha, Qatar), Ahmed El Khouly (Biomedical Informatics, Sidra Medicine, Doha, Qatar), Tushar Pathare (Biomedical Informatics, Sidra Medicine, Doha, Qatar), Shafeeq Poolat (Biomedical Informatics, Sidra Medicine, Doha, Qatar), Rashid Al-Ali (Biomedical Informatics, Sidra Medicine, Doha, Qatar).

Consortium Lead Principal Investigators: Omar M. E. Albagha (College of Health and Life Sciences, Hamad Bin Khalifa University, Education City, Doha, Qatar; Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK), Souhaila Al-Khodori (Microbiome and Biomarkers Discovery lab, Sidra Medicine, Doha, Qatar), Mashael Alshafai (College of Health Sciences, Qatar University, Doha, Qatar), Ramin Badii (Molecular Genetics Lab, Hamad Medical Corporation, Doha, Qatar), Lotfi Chouchane (Department of Genetic Medicine,

Microbiology and Immunology, Weill Cornell Medicine-Qatar, Doha, Qatar), Xavier Estivill (Research Branch, Sidra Medicine, Doha, Qatar; Quantitative Genomics Laboratories (qGenomics), Barcelona, Spain), Khalid Fakhro (Genomic Medicine Lab, Sidra Medicine, Doha, Qatar), Hamdi Mbarek (Qatar Genome Program, Qatar Foundation Research Development and Innovation, Qatar Foundation, Doha, Qatar), Younes Mokrab (Medical and Population Genomics Lab, Sidra Medicine, Doha, Qatar), Jithesh V. Puthen (College of Health and Life Sciences, Hamad Bin Khalifa University, Education City, Doha, Qatar), Karsten Suhre (Qatar Biobank for Medical Research, Qatar Foundation, Doha, Qatar; Bioinformatics Core, Weill Cornell Medicine-Qatar, Education City, Doha, Qatar; Department of Biophysics and Physiology, Weill Cornell Medicine, New York, NY, USA), Zohreh Tatari (Clinical Research Centre, Sidra Medicine, Doha, Qatar).

## REFERENCES PIPELINES

<https://qatargenome.org.qa>

<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

<https://github.com/lh3/bwa/tree/master/bwakit>

<https://software.broadinstitute.org/gatk/documentation/article?id=3238>

<https://pcingola.github.io/SnpEff/>

<https://sites.google.com/site/jpopgen/dbNSFP>

<https://www.ebi.ac.uk/gwas/>

<https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>

## CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the QGP and QBB, but restrictions apply to the availability of these data, which were used under license QF-QGP-RES-PUB-002 for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the QBB institutional review board.

## ORCID

Amal Elfatih  <http://orcid.org/0000-0002-4783-1384>

Fatemeh Abbaszadeh  <http://orcid.org/0000-0002-0964-9100>

## REFERENCES

- Al Hannan, F., Keogh, M. B., Taha, S., & Al Buainain, L. (2019). Characterization of BRCA1 and BRCA2 genetic variants in a cohort of Bahraini breast cancer patients using next-generation sequencing. *Molecular Genetics and Genomic Medicine*, 7(7), 00771. <https://doi.org/10.1002/mgg3.771>
- Al Thani, A., Fthenou, E., Paparrodopoulos, S., Al Marri, A., Shi, Z., Qafoud, F., & Afifi, N. (2019). Qatar Biobank Cohort Study: Study design and first results. *American Journal of Epidemiology*, 188(8), 1420–1433. <https://doi.org/10.1093/aje/kwz084>
- Altshuler, D. L., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Collins, F. S., De La Vega, F. M., Donnelly, P., Egholm, M., Flisek, P., Gabriel, S. B., Gibbs, R. A., Knoppers, B. M., Lander, E. S., Leirach, H., Mardis, E. R., McVean, G. A., Nickerson, D. A., & Peterson, J. L. (2010). A map of

- human genome variation from population-scale sequencing. *Nature*, 467(7319), 1061–1073. <https://doi.org/10.1038/nature09534>
- Azamjah, N., Soltan-Zadeh, Y., & Zayeri, F. (2019). Global trend of breast cancer mortality rate: A 25-year study. *Asian Pacific Journal of Cancer Prevention*, 20(7), 2015–2020. <https://doi.org/10.31557/APJCP.2019.20.7.2015>
- Bener, A., Ayub, H., & Kakil, R. I. W. (2008). Patterns of cancer incidence among the population of Qatar: A worldwide comparative study. *Asian Pacific Journal of Cancer Prevention*, 9(1), 19–24.
- Caspi, O., Huber, I., Gepstein, A., Arbel, G., Maizels, L., Boulos, M., & Gepstein, L. (2013). Modeling of arrhythmogenic right ventricular cardiomyopathy with human induced pluripotent stem cells. *Circulation: Cardiovascular Genetics*, 6(6), 557–568. <https://doi.org/10.1161/CIRCGENETICS.113.000188>
- Chaabna, K., Cheema, S., Abraham, A., Alrouh, H., & Mamtani, R. (2018). Adult mortality trends in Qatar, 1989–2015: National population versus migrants. *PLOS One*, 13(9), e0203996. <https://doi.org/10.1371/journal.pone.0203996>
- Delahaye-Sourdeix, M., Anantharaman, D., Timofeeva, M. N., Gaborieau, V., Chabrier, A., Vallée, M. P., Lagiou, P., Holcátová, I., Richiardi, L., Kjaerheim, K., Agudo, A., Castellsagué, X., Macfarlane, T. V., Barzan, L., Canova, C., Thakker, N. S., Conway, D. I., Znaor, A., Healy, C. M., ... McKay, J. D. (2015). A rare truncating BRCA2 variant and genetic susceptibility to upper aerodigestive tract cancer. *Journal of the National Cancer Institute*, 107(5), 2–5. <https://doi.org/10.1093/jnci/djv037>
- Do, R., Stitzel, N. O., Won, H. -H., Jørgensen, A. B., Duga, S., Angelica Merlini, P., Kiezun, A., Farrall, M., Goel, A., Zuk, O., Guella, I., Asselta, R., Lange, L. A., Peloso, G. M., Auer, P. L., Girelli, D., Martinelli, N., Farlow, D. N., DePristo, M. A., & Project, N. E. S. (2015). Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature*, 518(7537), 102–106. <https://doi.org/10.1038/nature13917>
- Fakhro, K. A., Staudt, M. R., Ramstetter, M. D., Robay, A., Malek, J. A., Badii, R., Al-Marri, A. A. N., Khalil, C. A., Al-Shakaki, A., Chidiac, O., Stadler, D., Zirie, M., Jayyousi, A., Salit, J., Mezey, J. G., Crystal, R. G., & Rodriguez-Flores, J. L. (2016). The Qatar genome: A population-specific tool for precision medicine in the Middle East. *Human Genome Variation*, 3(1), 1–7. <https://doi.org/10.1038/hgv.2016.16>
- Ghatak, S., Chakraborty, P., Sarkar, S. R., Chowdhury, B., Bhaumik, A., & Kumar, N. S. (2017). Novel APC gene mutations associated with protein alteration in diffuse type gastric cancer. *BMC Medical Genetics*, 18(1), 1–11. <https://doi.org/10.1186/s12881-017-0427-2>
- Green, R. C., Berg, J. S., Grody, W. W., Kalia, S. S., Korf, B. R., Martin, C. L., McGuire, A. L., Nussbaum, R. L., O'Daniel, J. M., Ormond, K. E., Rehm, H. L., Watson, M. S., Williams, M. S., Biesecker, L. G., O'Daniel, J. M., Ormond, K. E., Rehm, H. L., Watson, M. S., Williams, M. S., & Biesecker, L. G. (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genetics in Medicine*, 15(7), 565–574. <https://doi.org/10.1038/GIM.2013.73>
- Hegde, M., Bale, S., Bayrak-Toydemir, P., Gibson, J., Bone Jeng, L. J., Joseph, L., Laser, J., Lubin, I. M., Miller, C. E., Ross, L. F., Rothberg, P. G., Tanner, A. K., Vitazka, P., & Mao, R. (2015). Reporting incidental findings in genomic scale clinical sequencing—A clinical laboratory perspective: A report of the Association for Molecular Pathology. *Journal of Molecular Diagnostics*, 17(2), 107–117. <https://doi.org/10.1016/j.jmoldx.2014.10.004>
- Jain, A., Gandhi, S., Koshy, R., & Scaria, V. (2018). Incidental and clinically actionable genetic variants in 1005 whole exomes and genomes from Qatar. *Molecular Genetics and Genomics*, 293(4), 919–929. <https://doi.org/10.1007/s00438-018-1431-8>
- Juang, J. M., & Huang, S. K. S. (2004). Brugada syndrome—An under-recognized electrical disease in patients with sudden cardiac death. *Cardiology*, 101(4), 157–169. <https://doi.org/10.1159/000076693>
- Kalia, S. S., Adelman, K., Bale, S. J., Chung, W. K., Eng, C., Evans, J. P., Herman, G. E., Hufnagel, S. B., Klein, T. E., Korf, B. R., McKelvey, K. D., Ormond, K. E., Richards, C. S., Vlangos, C. N., Watson, M., Martin, C. L., & Miller, D. T. (2017). Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): A policy statement of the American College of Medical Genetics and Genomics. *Genetics in Medicine*, 19(2), 249–255. <https://doi.org/10.1038/gim.2016.190>
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., Jang, W., Katz, K., Ovetsky, M., Riley, G., Sethi, A., Tully, R., Villamarin-Salomon, R., Rubinstein, W., & Maglott, D. R. (2016). ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Research*, 44(D1), D862–D868. <https://doi.org/10.1093/nar/gkv1222>
- Leshno, A., Shapira, S., Liberman, E., Kraus, S., Srour, M., Harlap-Gat, A., Avivi, D., Galazan, L., David, M., Maharshak, N., Moanis, S., Arber, N., & Moshkowitz, M. (2016). The APC I1307K allele conveys a significant increased risk for cancer. *International Journal of Cancer*, 138(6), 1361–1367. <https://doi.org/10.1002/ijc.29876>
- Li, W., Yin, L., Shen, C., Hu, K., Ge, J., & Sun, A. (2018). SCN5A variants: Association with cardiac disorders. *Frontiers in Physiology*, 9, 1372. <https://doi.org/10.3389/fphys.2018.01372>
- Liang, J., Lin, C., Hu, F., Wang, F., Zhu, L., Yao, X., Wang, Y., & Zhao, Y. (2013). APC polymorphisms and the risk of colorectal neoplasia: A huge review and meta-analysis. *American Journal of Epidemiology*, 177(11), 1169–1179. <https://doi.org/10.1093/aje/kws382>
- Maxwell, K. N., Wubbenhorst, B., Wenz, B. M., De Sloover, D., Pluta, J., Emery, L., Barrett, A., Kraya, A. A., Anastopoulos, I. N., Yu, S., Jiang, Y., Chen, H., Zhang, N. R., Hackman, N., D'andrea, K., Daber, R., Morrisette, J., Mitra, N., Feldman, M., ... Nathanson, K. L. (2017). BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. *Nature Communications*, 8(1), 1–11. <https://doi.org/10.1038/s41467-017-00388-9>
- Meeks, H. D., Song, H., Michailidou, K., Bolla, M. K., Dennis, J., Wang, Q., Barrowdale, D., Frost, D., McGuffog, L., Ellis, S., Feng, B., Buys, S. S., Hopper, J. L., Southey, M. C., Tesoriero, A., James, P. A., Bruinsma, F., Campbell, I. G., Broeks, A., & Goldgar, D. E. (2016). BRCA2 polymorphic stop codon K3326X and the risk of breast, prostate, and ovarian cancers. *Journal of the National Cancer Institute*, 108(2), djv315. <https://doi.org/10.1093/jnci/djv315>
- Miller, D. T., Lee, K., Chung, W. K., Gordon, A. S., Herman, G. E., Klein, T. E., Stewart, D. R., Amendola, L. M., Adelman, K., Bale, S. J., Gollob, M. H., Harrison, S. M., Hershberger, R. E., McKelvey, K., Richards, C. S., Vlangos, C. N., Watson, M. S., Martin, C. L., & Group, A. S. F. W. (2021). ACMG SF v3.0 list for reporting of secondary findings in clinical exome and genome sequencing: a policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 23(8), 1381–1390. <https://doi.org/10.1038/s41436-021-01172-3>
- Miller, D. T., Lee, K., Gordon, A. S., Amendola, L. M., Adelman, K., Bale, S. J., Chung, W. K., Gollob, M. H., Harrison, S. M., Herman, G. E., Hershberger, R. E., Klein, T. E., McKelvey, K., Richards, C. S., Vlangos, C. N., Stewart, D. R., Watson, M. S., Martin, C. L., & Group, A. S. F. W. (2021). Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2021 update: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 23(8), 1391–1398. <https://doi.org/10.1038/s41436-021-01171-4>
- Monasky, M. M., Micaglio, E., Cicone, G., Benedetti, S., Di Resta, C., Vicedomini, G., Borrelli, V., Ghiroldi, A., Piccoli, M., Anastasia, L., Santinelli, V., Ferrari, M., & Pappone, C. (2019). Genotype/phenotype relationship in a consanguineal family with Brugada syndrome harboring the R1632C missense variant in the SCN5A gene. *Frontiers in Physiology*, 10, 666. <https://doi.org/10.3389/fphys.2019.00666>

- Nair, P., Obeid, T., & Tadmouri, G. O. (2008). Genetic disorders in the Arab world. *BMJ*, 333, 831.
- Nakajima, S., Makiyama, T., Hanazawa, K., Kaitani, K., Amano, M., Hayama, Y., Onishi, N., Tamaki, Y., Miyake, M., Tamura, T., Kondo, H., Motooka, M., Izumi, C., Nakagawa, Y., & Horie, M. (2013). A novel SCN5A mutation demonstrating a variety of clinical phenotypes in familial sick sinus syndrome 2013. *Internal Medicine*, 52(16), 1805–1808. <https://doi.org/10.2169/internalmedicine.52.0085>
- Oberg, J. A., Glade Bender, J. L., Sulis, M. L., Pendrick, D., Sireci, A. N., Hsiao, S. J., Turk, A. T., Dela Cruz, F. S., Hibshoosh, H., Remotti, H., Zylber, R. J., Pang, J., Diolaiti, D., Koval, C., Andrews, S. J., Garvin, J. H., Yamashiro, D. J., Chung, W. K., Emerson, S. G., ... Kung, A. L. (2016). Implementation of next generation sequencing into pediatric hematology-oncology practice: Moving beyond actionable alterations. *Genome Medicine*, 8(1), 1–19. <https://doi.org/10.1186/s13073-016-0389-6>
- Olfson, E., Cottrell, C. E., Davidson, N. O., Gurnett, C. A., Heusel, J. W., Stitzel, N. O., Chen, L. S., Hartz, S., Nagarajan, R., Saccone, N. L., & Bierut, L. J. (2015). Identification of medically actionable secondary findings in the 1000 genomes. *PLOS One*, 10(9), 1–18. <https://doi.org/10.1371/journal.pone.0135193>
- Probst, V., Chatel, S., Gourraud, J., & Marec, H. L. (2012). Clinical arrhythmias risk stratification and therapeutic approach in Brugada syndrome clinical arrhythmias. *Arrhythmia & Electrophysiology Review*, 1(1), 9–12. <https://doi.org/10.15420/aer.2012.1.1.17>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Reh, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405–424. <https://doi.org/10.1038/gim.2015.30>
- Scott, E. M., Halees, A., Itan, Y., Spencer, E. G., He, Y., Azab, M. A., Gabriel, S. B., Belkadi, A., Boisson, B., Abel, L., Clark, A. G., Greater Middle East Variome Consortium, Alkuraya, F. S., Casanova, J. L., & Gleeson, J. G. (2016). Characterization of greater middle eastern genetic variation for enhanced disease gene discovery. *Nature Genetics*, 48(9), 1071–1079. <https://doi.org/10.1038/ng.3592>
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, 29(1), 308–311. <https://doi.org/10.1093/nar/29.1.308>
- Shtoyerman-Chen, R., Friedman, E., Figer, A., Carmel, M., Patael, Y., Rath, P., Fidler, H. H., Bar-Meir, S., & Theodor, L. (2001). The I1307K APC polymorphism: Prevalence in non-Ashkenazi Jews and evidence for a founder effect. *Genetic Testing*, 5(2), 141–146. <https://doi.org/10.1089/109065701753145628>
- Stenson, P. D., Mort, M., Ball, E. V., Howells, K., Phillips, A. D., Cooper, D. N., & Thomas, N. S. T. (2009). The human gene mutation database: 2008 update. *Genome Medicine*, 1(1), 13. <https://doi.org/10.1186/gm13>
- Tadmouri, G. O., Nair, P., Obeid, T., Al Ali, M. T., Al Khaja, N., & Hamamy, H. A. (2009). Consanguinity and reproductive health among Arabs. *Reproductive health*, 6(1), 17. <https://doi.org/10.1186/1742-4755-6-17>
- Thareja, G., Al-Sarraj, Y., Belkadi, A., Almotawa, M., Suhre, K., & Albagha, O. M. E. (2021). Whole genome sequencing in the Middle Eastern Qatari population identifies genetic associations with 45 clinically relevant traits. *Nature Communications*, 12(1), 1250. <https://doi.org/10.1038/s41467-021-21381-3>
- Van Hout, C. V., Tachmazidou, I., Backman, J. D., Hoffman, J. D., Liu, D., Pandey, A. K., Gonzaga-Jauregui, C., Khalid, S., Ye, B., Banerjee, N., Li, A. H., O'Dushlaine, C., Marcketta, A., Staples, J., Schurmann, C., Hawes, A., Maxwell, E., Barnard, L., Lopez, A., & Center, R. G. (2020). Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature*, 586(7831), 749–756. <https://doi.org/10.1038/s41586-020-2853-0>
- Woodage, T., King, S. M., Wacholder, S., Hartge, P., Struwing, J. P., McAdams, M., Laken, S. J., Tucker, M. A., & Brody, L. C. (1998). The APC I1307K allele and cancer risk in a community-based study of Ashkenazi Jews. *Nature Genetics*, 20(1), 62–65. <https://doi.org/10.1038/1722>

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Elfatih, A., Mifsud, B., Syed, N., Badii, R., Mbarek, H., Abbaszadeh, F., & Estivill, X. (2021). Actionable genomic variants in 6045 participants from the Qatar Genome Program. *Human Mutation*, 42, 1584–1601. <https://doi.org/10.1002/humu.24278>