HGG Advances

ARTICLE

Burden of heterozygote carriers for autosomal recessive conditions in the Middle East: A study of 14,392 genomes

Esther Nkrumah,[1,12] Hajer Al-Mulla,[1,12] Aryan Bashar,[2,12] Vishal Soman,[3] Chadi Saad,[4,5] John Wang,[6] Helia Darabi,[7] Svetlana A. Yatsenko,[8] Aleksandar Rajkovic,[9,10] Uma Chandran,[3] Hamdi Mbarek,[4,5] and Mahmoud Aarabi[7,11,13,*]

Summary

The American College of Medical Genetics and Genomics (ACMG) recommends Tier-3 reproductive carrier screening for 97 genes associated with autosomal recessive (AR) conditions (AR genes). Gene selection for screening is based on a gene carrier frequency (GCF) of ≥1/200 in the Genome Aggregation Database (gnomAD)v2 populations and disease severity. The utility of ACMG Tier 3 in the Middle Eastern population is unclear as this ancestral group was not represented in the gnomADv2. We utilized genome data from 14,392 individuals in the Qatar Genome Project to estimate the carrier frequency of AR conditions in the Middle Eastern population. The frequency of 136,624 pathogenic/likely pathogenic variants in 2,987 AR genes from ClinVar was analyzed to estimate the GCF in the Qatari cohort. Genes with GCF ≥ 1/200 were curated by an expert panel for the severity of corresponding conditions. We identified 69 genes with GCF ≥ 1/200 associated with moderate to profound clinical presentations, 53 of which were unique to the Middle Eastern population. Common variants were observed with high frequency for individual genes such as *IL2RA* that suggested the presence of founder effects not described previously. Simulation studies predicted that inclusion of the ancestry-specific genes increased the chance of detecting at-risk couples from 3.85% to 8.15% in the Middle Eastern population. This study highlights the limitations of relying on global datasets to accurately identify the candidate genes for carrier screening in different populations. Our findings provide a framework for targeted, population-specific carrier-screening programs in regions not represented in the large genome datasets.

Introduction

Carrier screening is performed preconceptionally or prenatally to identify individuals at risk of having a child with autosomal recessive (AR) or X-linked (XL) conditions.[1] At-risk individuals and couples have choices that support their reproductive autonomy, including preimplantation genetic testing for monogenic disorders (PGT-M), use of donor gametes, prenatal diagnosis, termination of pregnancy, and adoption.[2] Historically, carrier screening has been done in specific populations, such as screening for Tay-Sachs disease in the Ashkenazi Jewish population and sickle cell disease (SCD) in the African/Mediterranean population.[3,4] Aligning screening with population risks provided couples with cost-effective and targeted testing options.[1] However, this approach was challenged by the increasing number of reproductive partners from different ancestries and the fact that individuals may not be aware of their genetic ancestry to determine the type of targeted screening needed.[5] In addition, development of genetic technologies, such as next-generation sequencing (NGS), was a driving force behind expanding the carrier screening to include multiple conditions[1] in a single test known as expanded carrier screening.[6]

In 2021, the American College of Medical Genetics and Genomics (ACMG) released its carrier-screening clinical practice resource for AR and XL genetic conditions and recommended offering carrier screening to all individuals during pregnancy and preconception regardless of their ancestry.[7] ACMG's practice resource included a tiered model for carrier screening.[7] Tier-1 screening consisted of a pan-ethnic approach that included two conditions, cystic fibrosis (CF) and spinal muscular atrophy (SMA).[8,9] Tier 2 was based on recommendations from the American College of Obstetricians and Gynecologists (ACOG) that included conditions with a well-defined phenotype and a carrier frequency of ≥1/100 in population databases.[3,10] ACMG recommended the Tier-3 carrier screening that included conditions with a carrier frequency ≥1/200 and moderate to profound clinical presentations, comprising

[1]Joan H. Marks Graduate Program in Human Genetics, Sarah Lawrence College, Bronxville, NY 10708, USA; [2]Data Science Graduate Program, University of Colorado Boulder, Boulder, CO, USA; [3]Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; [4]Qatar Genome, Qatar Precision Health Institute, Qatar Foundation, Doha 5825, Qatar; [5]Development and Innovation, Qatar Foundation, Doha, Qatar; [6]Division of Genetic and Genomic Medicine, Department of Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; [7]Medical Genetics and Genomics Laboratories, University of Pittsburgh Medical Center, Pittsburgh, PA, USA; [8]Department of Pathology, Stanford University, Stanford, CA, USA; [9]Department of Pathology, University of California, San Francisco, San Francisco, CA, USA; [10]Institute of Human Genetics, University of California, San Francisco, San Francisco, CA, USA; [11]Departments of Pathology, and Obstetrics, Gynecology, and Reproductive Sciences, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
[12]These authors contributed equally
[13]Lead contact
*Correspondence: mahmoud.aarabi@pitt.edu
https://doi.org/10.1016/j.xhgg.2026.100625.

genes associated with 97 AR and 16 XL conditions. ACMG Tier-4 carrier screening had no limit on carrier frequency, thereby screening for rare genetic conditions that may be more pertinent to consanguineous relationships or specific ancestral populations.[7]

An important element in the selection of genes recommended as the ACMG Tier 3 for carrier screening was a carrier frequency of at least 1/200 for pathogenic or likely pathogenic variants based on the data obtained from the Genome Aggregation Database (gnomAD) v2.0.2.[11] This database contained exome and genome data from approximately 125,000 unrelated individuals with no history of severe pediatric disease, categorized into seven ancestral populations described by gnomAD as European (non-Finnish), European (Finnish), African/African American, admixed American, Ashkenazi Jewish, East Asian, and South Asian.[12] The Middle Eastern population was not represented in gnomADv2.0.2. In gnomADv3, released in late 2020, approximately 76,000 genomes were included from the above ancestral populations along with 158 genomes from the Middle Eastern population.[13] The newest version of gnomAD, v4.1 released in 2024, consisted of an unprecedented number of exomes (over 700,000) including 2,884 from the Middle Eastern population. Our previous analysis of gnomAD v3 and v4 demonstrated dynamic changes in carrier frequency of AR conditions as data from new populations became available.[14,15] Our studies also revealed the need for larger genome databases from various populations to better understand the burden of heterozygote carriers for AR conditions. To date, there is no information on such burdens in the Middle Eastern population due to limited genome information. Here, we analyzed the largest Middle Eastern dataset from the Qatar Genome Program (QGP), comprising genome data from 14,392 individuals, to estimate the carrier frequency of pathogenic/likely pathogenic (P/LP) variants in AR genes. We aimed to develop and curate a list of potential genes for carrier screening based on carrier frequency and disease severity to compare with the ACMG Tier-3 list and data from gnomAD.

Subjects and methods

Population and genome sequencing data

The Qatar genomes consisted of a cohort of 14,392 Qatari participants from the population-based QGP.[16] Ethical approval was provided by the institutional review boards of the Qatar Biobank (QBB; QF-QGP-RES-PUB-002) and the University of Pittsburgh (STUDY24070109). A detailed description of the recruitment process and collection of phenotypic data by the QGP and the QBB was previously reported.[16,17] A signed consent was obtained from all included participants in the QBB, and genome sequencing (GS) data used in the current study was obtained from the QGP as previously described.[17] All QGP samples were sequenced on Illumina HiSeq X instruments (Illumina, San Diego, CA, USA) to a target average depth of coverage of 30×. Quality control of the produced FastQ files was performed using FastQC (v0.11.2). Read mapping, variant calling, and joint variant calling were performed using Sentieon's DNASeq pipeline v201808.03 (Sentieon, San Jose, CA, USA), following the BWA-GATK Best Practice Workflow and using the GRCh38/hg38 reference genome. The produced genome VCFs were jointly called to produce one multi-sample variant carrier frequency (VCF) file (msVCF) for the whole cohort.

Bioinformatic analysis of ClinVar and Qatar genome data

Bioinformatic analysis was performed using Python version 3.11.9 and bash version 4.4.19. A VCF file containing the complete list of variants mapped to the GRCh38 reference genome was downloaded from the ClinVar ftp site on October 7, 2023. Autosomal variants that had P/LP clinical significance without conflicting evidence were extracted from the VCF file. Variants in 2,987 genes associated with AR conditions were selected for further study (Figure 1A). Based on filtering criteria, a list was made consisting of variants from the ClinVar VCF file, which were stored in a TSV file used for downstream analysis. Estimation of carrier frequency was performed as described previously.[11,14,15] Briefly, the filtered ClinVar variant list was queried against the Qatar multi-sample VCF file for matching purposes. Matching variants were identified using chromosome number, position, and reference/alternate alleles for each ClinVar variant. Allele count (AC), allele number (AN), and number of homozygotes (nhomalt) were extracted and stored in TSV files sorted by chromosome numbers. AC for each variant and total ANs were used to estimate the allele frequency after removal of the number of homozygotes. VCF was calculated as the proportion of individuals in a population that carry the variant of interest, as shown in Equation 1.

$$VCF = \frac{AC - nhomalt}{0.5 \times AN} \quad \text{(Equation 1)}$$

VCF values were then used to calculate the gene carrier frequency (GCF) for each of the AR genes in the population. GCF was defined as the proportion of individuals in a population heterozygous for a P/LP variant in a given gene, as described previously,[11,14,15] and as shown in Equation 2, where g represents the gene and v is the total number of P/LP variants of that gene used in the analysis.

$$GCF_g = 1 - \prod_{i=1}^{v}(1 - VCF_i) \quad \text{(Equation 2)}$$

Carrier-frequency data for VCF and GCF were concatenated into TSV files.

Curation for clinical severity of AR conditions

We took a stepwise approach to analyze the clinical severity of AR genes with GCF $\geq$ 1/200 in the Qatar population (Figure 1B). First, AR genes with GCF $\geq$ 1/200 that were part of the ACMG Tier-3 list were included. Second, we included genes with GCF $\geq$ 1/200 that were previously curated and determined to be associated with moderate, severe, or profound clinical presentation.[11,18] The remaining genes were annotated and reviewed by two genetic counselors as well as specialists in laboratory and clinical genetics, pediatrics, and maternal-fetal medicine. Various resources, including Clinical Genome Resource (ClinGen), MedGen, UniProt, OMIM, GeneReviews, NORD, and MedlinePlus, were used to assess disease characteristics and severity based on the algorithm previously described.[18] Genes with GCF $\geq$ 1/200 and association with moderate, severe, and profound clinical presentations were selected. Genes associated with mild conditions or

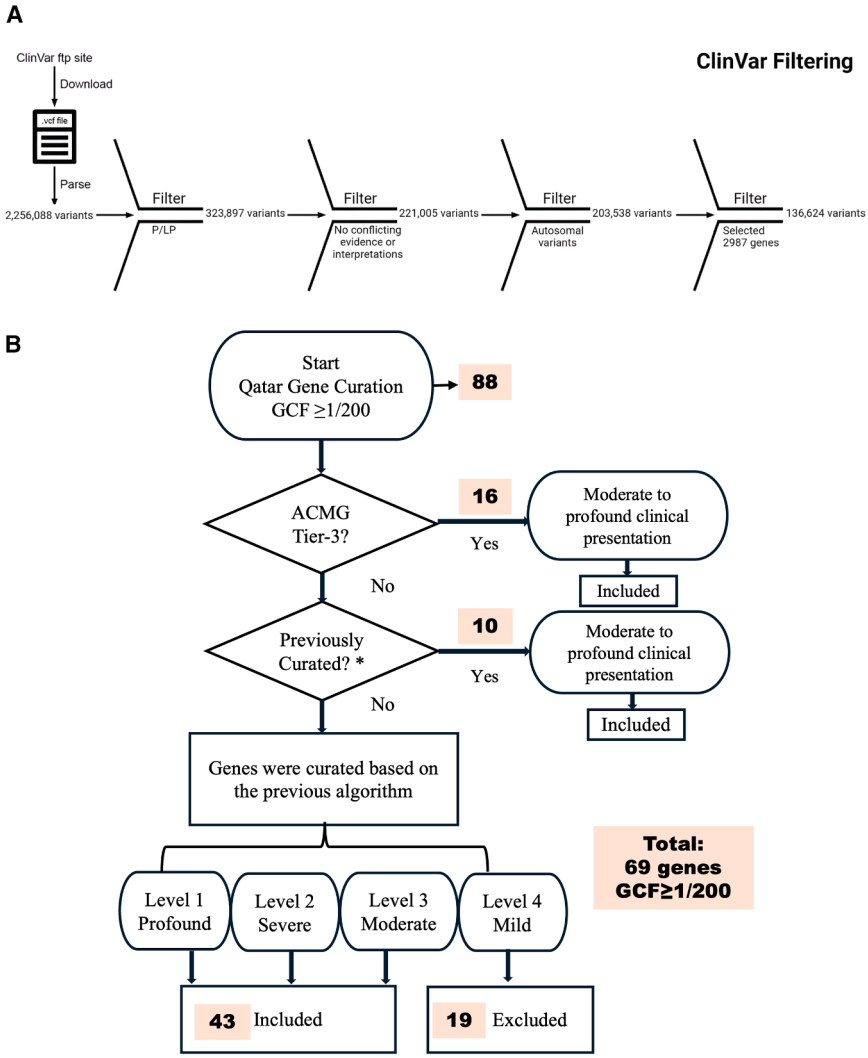

Figure 1. Analysis and curation of genes associated with AR conditions (AR genes) with a carrier frequency of ≥1/200 in the Qatar genomes

(A) Filtration of pathogenic/likely pathogenic (P/LP) variants from ClinVar in 2,987 AR genes.

(B) Curation of genes based on disease severity, as described by Lazarin et al. Only genes associated with moderate, severe, or profound conditions were included.

adult-onset diseases were excluded; a summary of justification for assigning the mild classification is provided (Table S1).

Estimating the burden of heterozygote carriers and homozygote/compound heterozygote affected individuals

The probability that an individual and a couple would be heterozygote carriers for any P/LP variants in at least one of the AR genes with a GCF ≥ 1/200 was assessed through use of Equations 3 and 4, where n is the total number of genes included.

$$P_{Carrier\ of\ P/LP\ for\ any\ gene} = 1 - \prod_{i=1}^{n}(1 - GCF_i) \quad \text{(Equation 3)}$$

$$P_{Both\ carriers\ of\ P/LP\ for\ any\ gene} = 1 - \prod_{i=1}^{n}\left[1 - (GCF_i)^2\right] \quad \text{(Equation 4)}$$

The probability of a heterozygote couple conceiving a homozygote/compound heterozygote child was assessed based on Equation 5.

$$P_{Offspring\ with\ two\ P/LP\ variants\ at\ least\ one\ gene} = 1 - \prod_{i=1}^{n}\left[1 - 0.25 \times (GCF_i)^2\right] \quad \text{(Equation 5)}$$

The estimated percentages were made based on GCF values for individuals from the Qatar population. The AR genes from the ACMG Tier-3 list were used for comparison. The GCF of the QGP population was used for the shared genes between QGP and ACMG Tier 3.

Results

Curation of genes with GCF ≥ 1/200 in Qatari population

Analysis of ClinVar revealed 136,624 P/LP variants with no conflicting interpretations among the 2,987 genes associated with AR conditions (AR genes; Figure 1A). The list of variants was utilized to determine the GCF in 14,392 genomes from the Qatari population. We identified 88 genes from the Qatari population with a GCF ≥ 1/200 that were subject for annotation of clinical severity by our expert panel (Figure 1B). Of the 88 genes with GCF ≥ 1/200, 16 genes belonged to the ACMG Tier-3 gene list (Table 1). In addition, 10 genes with GCF ≥ 1/200 (Table 2) were previously determined to be associated with moderate, severe, or profound conditions1[11,18] and, thus, were included in our final list. The remaining 62 genes were annotated by our group for association with moderate to profound presentations, resulting in inclusion of 43 genes (Table 3); 19 genes were excluded due to mild presentations and/or adult-onset conditions (Table S1). Our analysis thus identified 69 genes with GCF ≥ 1/200 associated with moderate to profound presentations (43 newly curated genes, 10 genes previously characterized, and 16 genes in the ACMG Tier-3 gene list; Figure 1B).

Comparison of Qatar genomes with gnomAD Middle Eastern population

We then asked whether the number and content of genes with GCF ≥ 1/200 is different between the Qatar genomes and the Middle Eastern data in gnomAD. Carrier

Table 1. Genes with a GCF ≥ 1/200 shared between the Qatar genomes and the ACMG Tier-3 gene list ($n = 16$)

Gene	MIM#	GCF
USH2A	608400	0.08968
CYP21A2	613815	0.06554
SMPD1	607608	0.05231
MCCC2	609014	0.03513
HBB	141900	0.03317
CFTR	602421	0.02947
PAH	612349	0.01301
HBA2	141850	0.01284
GJB2	121011	0.01256
DHCR7	602858	0.0092
NEB	161650	0.00858
ACADM	607008	0.0079
SLC26A4	605646	0.00748
HBA1	141800	0.00613
DLD	238331	0.00587
PCDH15	605514	0.0047

GCF, gene carrier frequency.

Table 2. Genes with GCF ≥ 1/200 in Qatar genomes that were previously identified to be associated with moderate to profound clinical presentations ($n = 10$)

Gene	MIM#	GCF
CYP1B1	601771	0.02122
CEP152	613529	0.01238
DNAH11	603339	0.00871
POLR1C	610060	0.00619
GALC	606890	0.00593
LAMA2	156225	0.00511
LTBP4	604710	0.00504
DPYD	612779	0.00477
F11	264900	0.00456
RYR1	180901	0.00456

GCF, gene carrier frequency.

frequency used for curation of the ACMG Tier-3 list was derived from gnomADv2.0.2 that did not have representation of the Middle Eastern participants (Figure 2A). Only 16 genes with GCF ≥ 1/200 were shared between ACMG Tier-3 and Qatar genomes, while 81 ACMG Tier-3 genes did not meet the GCF ≥ 1/200 requirement for screening in the Qatari population (Figure 2B; Table 1). The total number of genes with GCF ≥ 1/200 in 158 Middle Eastern participants in gnomADv3.1.1 were 59, of which 12 were shared with Qatar genomes[14] (Table S2). Representation of the Middle Eastern population increased to 2,873 in gnomADv4.1.0 with 59 genes having GCF ≥ 1/200, as we reported previously (Table S2).[15] Our analysis showed that 35 genes, with GCF ≥ 1/200 and associated with moderate to profound presentation, were shared between the Qatar genomes and the Middle Eastern population in gnomADv4.1.0 (Figures 2B and 2C; Table S2). Interestingly, 25 genes had GCF ≥ 1/200 only in gnomADv4.1.0 Middle Eastern population, but not in Qatar genomes, while 31 of 69 genes with GCF ≥ 1/200 in Qatar genomes were below the cutoff of 1/200 in the gnomAD Middle Eastern population (Figures 2B and 2C; Table S2). This finding likely demonstrates the heterogeneous Middle Eastern population in gnomADv4.1.0, which might be different from Qatari genomes due to genomic variations. The gene with the highest GCF in Qatar population (*IL2RA*, MIM: 147730; GCF = 0.16477) was unique to this database with one variant that was not reported previously in the gnomAD Middle Eastern population (Figure 2D; Tables S3 and S4), suggestive of a founder variant.

Number of P/LP variants contributing to GCF ≥ 1/200 in Qatar genomes

The number of P/LP variants from ClinVar for genes with GCF ≥ 1/200 ranged from 1 to 38 variants (Figure 3A). Most genes had fewer than five P/LP variants contributing to the observed GCF ≥ 1/200 (Figure 3B). Information on all P/LP variants and their respective carrier frequency, categorized by genes, is provided in Tables S3 and S4.

Simulating the clinical utility of carrier-screening panels for Qatari population

To understand the utility of carrier-screening panels in detection of at-risk couples, we used a simulation model to predict the probabilities based on the ACMG Tier 3 or the Qatar genome data.

The probability that an individual and a couple would be heterozygote carriers for any P/LP variants in at least one of the AR genes with a GCF ≥ 1/200 was estimated, followed by the risk of heterozygote reproductive partners to conceive a homozygote/compound-heterozygote child. If tested for the AR genes included in the ancestry-neutral ACMG Tier-3 list, an individual from the Middle Eastern/Qatari population would have up to a 73.42% chance of being a heterozygote carrier for P/LP variants in at least one of the Tier-3 genes (Table 4). Couples would have as high as a 3.85% risk of being carriers of a P/LP variant of the same gene(s) and up to an estimated 0.97% risk of having a homozygote or compound heterozygote child. When screened for only the list of 69 AR genes derived from the current study, the chance of finding at-risk couples increased from 3.85% to 6.63%, and the probability of detecting a homozygote or compound heterozygote child increased from 0.97% to 1.69% (Table 4). We then examined a model where individuals receive screening for genes with GCF ≥ 1/200 from the current study and the ACMG Tier-3 genes. The combined model would detect as high as 8.15% at-risk couples among the

Table 3. Genes with GCF ≥ 1/200 in Qatar genomes associated with moderate to profound presentations based on the annotation in the current study ($n = 43$)

Gene	MIM#	GCF
IL2RA	147730	0.16477
ATG7	608760	0.12668
RBM8A	605313	0.03953
MPL	159530	0.01943
NCF1	608512	0.01885
SLC2A10	606145	0.01554
SERPINH1	600943	0.01541
EYS	612424	0.01443
DCAF17	612515	0.01432
PIK3CG	601232	0.01418
CYP4F22	611495	0.01124
RAB27A	603868	0.01036
VSX2	142993	0.01023
SLC24A1	603617	0.00927
MYO9A	604875	0.00913
KCNE1	176261	0.00858
MICU1	605084	0.00825
DAAM2	606627	0.00817
IFNG	147570	0.00804
POLRMT	601778	0.00676
PKLR	609712	0.006
KANK2	614610	0.00593
CYP2R1	608713	0.00586
FLG	135940	0.00585
ABCB4	171060	0.00579
STRC	606440	0.00578
PROC	612283	0.00559
SLC22A5	603377	0.00552
SPG7	602783	0.00552
C12orf57	615140	0.00538
GRM6	604096	0.00524
FARS2	611592	0.00518
ZNF335	610827	0.00518
ABCA4	601691	0.00503
KIF26A	613231	0.00498
ATAD3A	612316	0.00498
FCSK	608675	0.00491
SERPINA1	107400	0.0049
C6	217050	0.00477
EPCAM	185535	0.0047
MEFV	608107	0.0047
GJB6	604418	0.00464
LAMA5	601033	0.00463
CYP2U1	610670	0.00457

GCF, gene carrier frequency.

individuals screened, resulting in the probability of a homozygote or compound heterozygote child at 2.09% (Table 4).

Discussion

This study identified high frequency of heterozygote carriers for a unique set of genes in the Middle Eastern population that are distinct from other ancestries in large genome datasets. Variants in genes such as *IL2RA* were found to have a very high carrier frequency suggesting founder effects for conditions not identified previously. The genes with high carrier frequency can be utilized to direct the carrier-screening efforts in the Middle Eastern population as a customized regional solution. Based on our simulation, we recommend that a Tier-4 screening may have the highest benefit encompassing the ACMG Tier-3 panel and the genes specific to the Middle Eastern population as determined here.

Reproductive carrier screening is accessible in many countries, although there are differences in the number and types of conditions screened and variations in the cost structures underlying the screening program, with some being user paid and others government funded.[19] In several Middle Eastern countries, including Qatar, Iran, Turkey, Iraqi Kurdistan, Saudi Arabia, Jordan, Bahrain, and the United Arab Emirates, carrier screening is called premarital screening (PMS).[19,20] PMS is offered free of charge and is mandatory in most of the Middle Eastern countries mentioned above before couples receive their marriage certificates.[19] Current PMS programs in Middle Eastern countries do not comprehensively screen a broad spectrum of AR conditions,[20] nor do they meet the ACMG recommended Tier-3 panel for carrier screening.[7] In Qatar, PMS is currently offered by primary healthcare centers and selected private hospitals[21] for SCD, thalassemia, homocystinuria, CF, and an optional screening for SMA.[20] Conditions were selected for screening due to higher disease frequencies observed in the Qatari population.[21] In this study, we asked whether the Middle Eastern population would benefit from following the current guidelines and adopting the ACMG Tier-3 list. We identified 69 genes with GCF ≥ 1/200 and association with moderate, severe, or profound clinical presentations in the current cohort of the Qatari genomes. The limited overlap of only 16 genes between

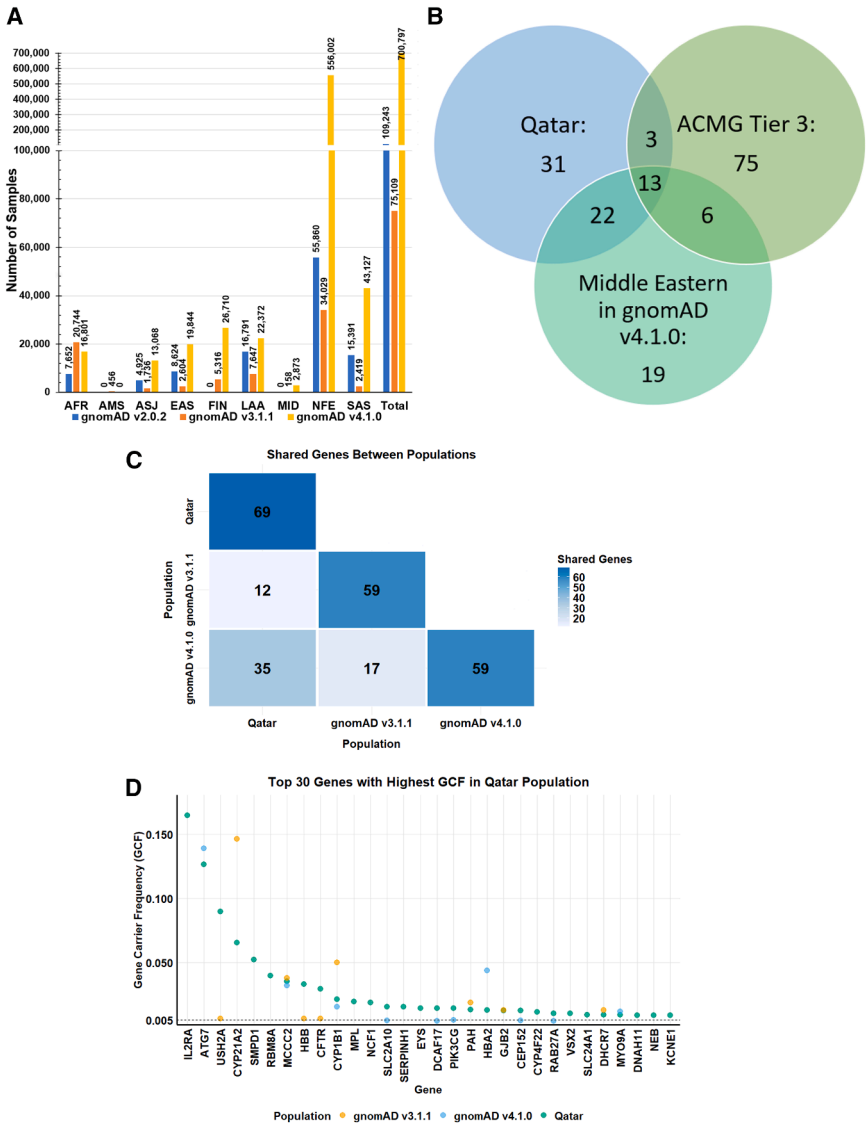

Figure 2. Comparative analysis of Qatar genomes and gnomAD Middle Eastern population

(A) Representation of ancestral populations in gnomAD v2, v3, and v4.

(B) Comparison of genes with gene carrier frequency (GCF) ≥1/200 in Qatar genomes, ACMG Tier-3 carrier-screening list (derived from gnomAD v2.0.2), and gnomAD v4.1.0.

(C) Shared genes between Qatar genomes and Middle Eastern population in gnomAD v3 and v4.

(D) GCF of the 30 most common genes in gnomAD v3, v4, and Qatar population. AFR, African/African American; AMS, Amish; ASJ, Ashkenazi Jewish; EAS, East Asian; FIN, Finnish; LAA, Latino/admixed American; MID, Middle Eastern; NFE, non-Finnish European, SAS, South Asian.

the Qatari list and the ACMG Tier-3 list highlights the varied utility of professional guidelines in different populations. Version 2 of gnomAD, used for ACMG Tier-3 development, contained no data from the Middle Eastern populations. We propose that limited representation of genomic data can explain the disparity in the number of genes with GCF ≥ 1/200 between Qatari genomes and the ACMG Tier-3 genes derived from gnomADv2. Our previous studies have shown that genome data from approximately 9,000 individuals in a given ancestral population are needed for an accurate estimation of carrier frequency and selection of genes with GCF ≥ 1/200.[15] The current study highlights the need for more research among populations with low representation in gnomAD and other global genomic datasets.

The highest GCF (0.165) was observed in the *IL2RA* gene, associated with immunodeficiency due to CD25 deficiency (MIM: 606367). This immune condition presents in early infancy with skin manifestations, immunodeficiency, and recurrent infections. There is no report on the prevalence of this condition among the Qatari population. However, a report of CD25 deficiency in a child with a similar presentation in Saudi Arabia attributed the condition to the same variant in the *IL2RA* gene.[22] This, together with the high number of individuals carrying the same variant in the Qatar sample population, is suggestive of a founder effect, although other reasons such as selection or variant non-pathogenicity should be considered. The *IL2RA* gene was not detected as a gene with GCF ≥ 1/200 in our previous study of the gnomADv4 population.[15] This could be due to a smaller number of Middle Eastern individuals in gnomADv4 or the high frequency of heterozygous carriers limited to the Qatari/Saudi population but not the rest of Middle Eastern population. More research is needed to determine the founder effect by haplotype analysis and age estimation and to study the frequency of CD25 deficiency in Qatar and neighboring countries in the region. Another gene with a high carrier frequency, *CYP21A2*, is associated with congenital adrenal hyperplasia (CAH) with several contributing variants in the Qatari population. CAH is prevalent in the Middle East2[23–25] and is included in the newborn screening programs across the region, including Qatar.[26] *CYP21A2*, along with *ATG7*, *USH2A*, and *SMPD1*, were among the genes with high GCF that were also detected in our study of the Middle Eastern population in gnomADv4 previously.[15] High frequency of heterozygous carriers for known founder variants or previously reported pathogenic variants in the Qatari population was also observed in *DCAF17* (MIM: 612515; GCF = 0.01432) associated with Woodhouse-Sakati

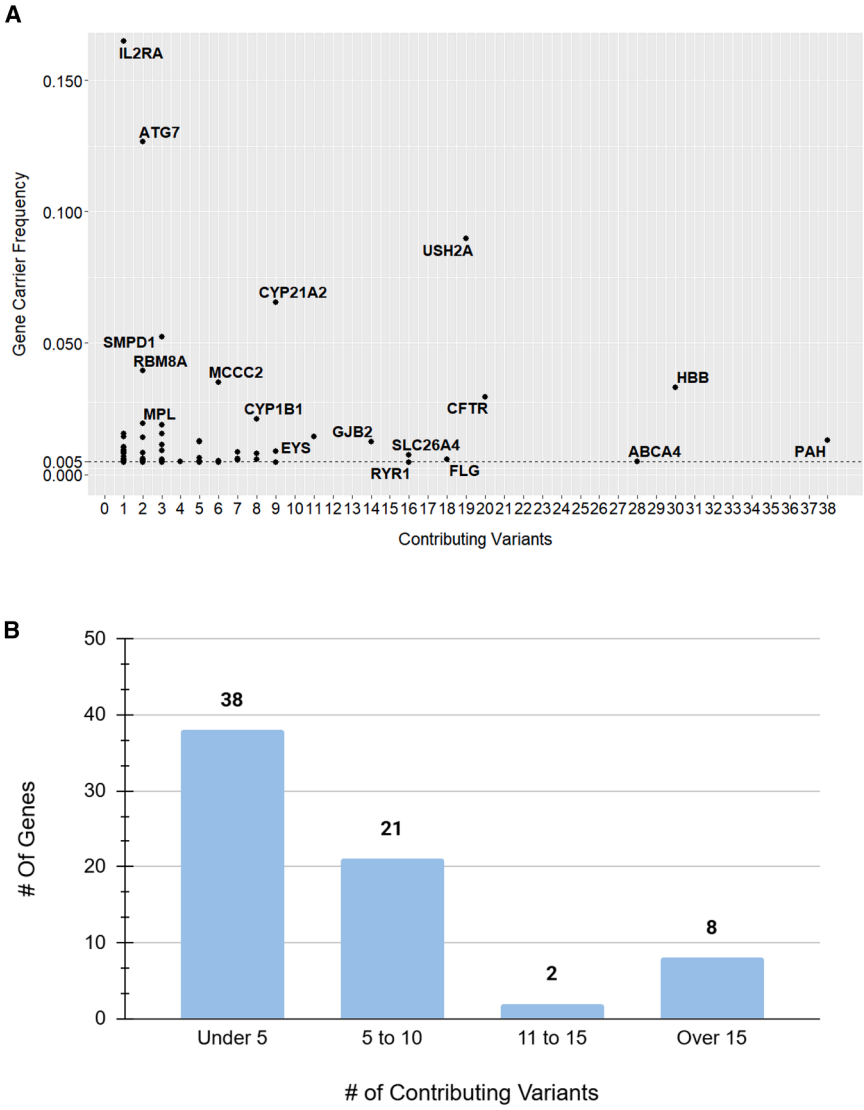

Figure 3. Number of P/LP variants contributing to GCF ≥ 1/200 in Qatar genomes

(A) Scatterplot demonstrates the GCF and the number of contributing variants for 69 genes with GCF ≥ 1/200 in Qatar genomes. Each dot represents the number of contributing variants for the gene. The horizontal line at GCF = 0.005 indicates the threshold used.

(B) Bar chart illustrating the distribution of genes by number of contributing variants. Most genes had fewer than five contributing variants, while a small number had more than 15 contributing variants.

syndrome (MIM: 241080), *RAB27* (MIM: 603868; GCF = 0.01036) associated with Griscelli syndrome, type 2 (MIM: 607624), *MPL* (MIM: 159530; GCF: 0.01943) associated with amegakaryocytic thrombocytopenia, congenital, 1 (MIM: 604498), and *VSX2* (MIM: 142993; GCF = 0.01023) associated with Microphthalmia/coloboma 3 (MIM: 610092).[16,27–30] Notably, none of these genes are currently included in the Qatar's PMS program, and there are no published data on their frequency in Qatar, warranting further investigation on the status of these genetic conditions in the Middle Eastern population.

In populations with high rates of consanguinity or founder effects combined with endogamy, the frequency of AR conditions is increased. Qatar, for example, has been found to have a consanguinity rate of approximately 54%,[21] which reemphasizes a need for comprehensive screening strategies.[31] The current study demonstrated the limited utility of ACMG Tier 3 for carrier screening of individuals or couples from Qatar/Middle Eastern populations. We showed that the probability of finding couples at risk almost doubles using a panel of 69 genes identified in this study as compared to a panel of ACMG Tier-3 genes. The highest probability (up to 8.15%) was observed when the ACMG Tier 3 was combined with the common genes in the Qatar genomes in one panel. It should be noted that our analysis assumes independence among variants, which may not hold in populations with assortative mating or inbreeding. Positive correlation between variants would likely cause our approach to overestimate the probability of carrying at least one variant. Based on our findings, we recommended defining a Tier-4 gene panel for the Qatari, and the Middle Eastern population, that encompasses ACMG Tier-3 genes and the Qatari genes for the maximal detection of at-risk couples. Defining a national or regional Tier-4 screening panel can be adapted by other regions as more genome data become available from the populations with small representation in gnomAD. In addition, periodic review of the screening programs should be considered to ensure that screening efforts remain aligned with emerging scientific evidence, changes in healthcare needs, advancement in technology, as well as ethical, social, and cultural factors in any given population.

Our analysis is derived from the genome data of the Qatari population, with origins that trace back mainly to general Arabs, peninsular Arabs, Arabs of Western Eurasia and Persia, and admixed Arabs.[16] While a rich genomic diversity was captured in Qatari genomes with relevance across the broader Middle East, this analysis may not be completely representative of other Middle Eastern countries that could be heterogeneous with some genomic variations.[32] Comparison of our findings with the future large studies on genomes from Arab and non-Arab Middle

Table 4. Comparison of probabilities for detection of heterozygote carriers and risk of homozygote/compound heterozygote individuals using ACMG Tier 3 or the Qatar genome data

Population	Number of genes	P Max individual carrier	P Max couples at risk	P Max affected child
ACMG	97	73.42%	3.85%	0.97%
QGP	69	68.75%	6.63%	1.69%
ACMG and QGP	150	87.62%	8.15%	2.09%

Columns demonstrates the probabilities for the indicated outcome based on the screening panel used. P Max represents the maximum probability of the specified event based on GCF $\geq$ 1/200. Couples at risk are couples where both individuals have any P/LP variant affecting the same gene. This includes situations where couples carry the same or different P/LP variants.

Eastern populations is needed to determine the concordance of the genes with GCF $\geq$ 1/200. Our bioinformatic pipeline can be utilized for analysis of carrier burden once the new data become available. Besides the ancestral differences, variations in sequencing platforms, sequencing depth, and population sizes may also affect the carrier frequency estimations.[14,33] Another limitation of the current study, and any estimation of carrier burden of ClinVar P/LP variants associated with AR conditions, is the limited representation of different ancestries in the ClinVar database due to abundance of submitters from US and European laboratories. As a result, P/LP variants from underrepresented ancestries may not be present in ClinVar or may be present but classified with conflicting interpretations. Availability of a list of P/LP variants from the Middle Eastern population in ClinVar and non-ClinVar resources may necessitate reanalysis of carrier burdens in future studies. Finally, our gene curation may not be comprehensive in evaluating the severity of conditions associated with the genes with GCF $\geq$ 1/200 in the Qatari population. Incomplete penetrance and variable expressivity are factors that may limit a subjective disease-severity evaluation. We suggest a thorough review of the clinical severity of these genes before considering them for any carrier-screening program as well as future research on the prevalence of these conditions in the Middle East.

The threshold of carrier frequency as $\geq$1/200 can be a limiting factor in selection of genes for carrier screening. Genes that are less common may be excluded while they are associated with severe presentations. An alternative approach in reproductive carrier screening focuses on disease severity without using a strict GCF cutoff of 1/200.[34–36] Depending on the cost to the healthcare system, public health policies, and the size of population screened, this approach may be considered by some countries. Beyond the gene selection for a carrier-screening program, the clinical utility of such programs and the impact on couples' reproductive decision-making are important considerations. Reproductive carrier screening has been shown to result in informed decisions being made by at-risk couples in various regions.[37–39] Access to reproductive options is among the major factors that determine the clinical utility of carrier screening.[40] Importantly, the clinical utility may be affected by the level of acceptability of carrier screening in various regions. In Qatar, the rates of conditions included in the premarital screening program remain elevated,[41] likely due to consanguineous marriages.[19,41] Couples may not change their view on marriage after learning their carrier status, as commitments to the relationship often outweigh considerations related to genetic factors.[42] Despite these challenges, a comprehensive screening program may be beneficial as it determines the risk of having an affected child and prepares the medical team for clinical management during pregnancy and upon the birth of the affected neonates. Concurrent pre-/post-test genetic counseling and cultural-oriented educational materials may help in raising awareness among couples opting for carrier screening.

In summary, this study determines the burden of heterozygous carriers for recessive conditions in a large cohort of Middle Eastern genomes from the QGP. Our comprehensive analysis provides a list of 69 genes with high carrier frequency associated with moderate, severe, or profound clinical presentations that, together with the ACMG Tier-3 screening panel, could be considered for reproductive carrier screening in the Middle East. This study uncovered a number of genes with high carrier frequency for single variants that warrant further research on the presence of founder effect in the Qatari and Middle Eastern population.

Data and code availability

- The codes used in the bioinformatics pipeline and the parsed ClinVar and genome data is available at GitHub repository (https://github.com/chadisaad/AR-genes-database_query-2023) and Zenodo (https://zenodo.org/records/18255554).
- The data collected or generated by the QBB are made available to *bona fide* researchers employed within or otherwise contractually bound to public and private institutions that conduct scientific research and that meet the requirements detailed in the QBB Research Access policy.
- The raw, unfiltered, individual-level calls have been deposited with QBB, now part of the Qatar Precision Health Institute (QPHI).
- GS data and population-level multi-sample SV VCF generated in this study can be accessed through the project accession ID (QF-QGP-PUB-002) upon submission and

approval of a data-access request. Researchers may apply for access via the QPHI portal at https://www.qphi.org.qa/research/how-to-apply.

Acknowledgments

This project was supported by the National Institutes of Health through grant number UL1TR001857, and by the Research Pilot Project Award to M.A. from the Department of Pathology at the University of Pittsburgh School of Medicine.

Author contributions

Conceptualization, E.N., H.A.-M., S.A.Y., A.R., U.C., H.M., and M.A.; data curation and formal analysis, E.N., H.A.-M., A.B., V.S., C.S., U.C., H.M., and M.A.; verification, E.N., H.A.-M., A.B., V.S., and C.S.; investigation, methodology, and result interpretation, E.N., H.A.-M., J.W., S.A.Y., A.R., U.C., H.M., and M.A.; project administration and supervision, U.C., H.M., and M.A.; visualization, A.B. and V.S. ; writing – original draft, E.N., H.A.-M., A.B., V.S., and M.A.; writing – review and editing, C.S., J.W., A.R., S.A.Y., U.C., H.M., and M.A.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2026.100625.

Received: October 20, 2025
Accepted: May 9, 2026

References

1. Kraft, S.A., Duenas, D., Wilfond, B.S., and Goddard, K.A.B. (2019). The evolving landscape of expanded carrier screening: challenges and opportunities. Genet. Med. *21*, 790–797. https://doi.org/10.1038/s41436-018-0273-4.
2. Goldberg, J.D., Pierson, S., and Johansen Taber, K. (2023). Expanded carrier screening: What conditions should we screen for? Prenat. Diagn. *43*, 496–505. https://doi.org/10.1002/pd.6306.
3. (2017). Committee Opinion No. 691: Carrier Screening for Genetic Conditions. Obstet. Gynecol. *129*, e41–e55. https://doi.org/10.1097/AOG.0000000000001952.
4. Gross, S.J., Pletcher, B.A., Monaghan, K.G., Professional Practice and Guidelines Committee, and Guidelines, C. (2008). Carrier screening in individuals of Ashkenazi Jewish descent. Genet. Med. *10*, 54–56. https://doi.org/10.1097/GIM.0b013e31815f247c.
5. Westemeyer, M., Saucier, J., Wallace, J., Prins, S.A., Shetty, A., Malhotra, M., Demko, Z.P., Eng, C.M., Weckstein, L., Boostanfar, R., et al. (2020). Clinical experience with carrier screening in a general population: support for a comprehensive pan-ethnic approach. Genetics in medicine. Genet. Med. *22*, 1320–1328. https://doi.org/10.1038/s41436-020-0807-4.
6. Bajaj, K., and Gross, S. (2014). Carrier Screening: Past, Present, and Future. J. Clin. Med. *3*, 1033–1042. https://doi.org/10.3390/jcm3031033.
7. Gregg, A.R., Aarabi, M., Klugman, S., Leach, N.T., Bashford, M.T., Goldwaser, T., Chen, E., Sparks, T.N., Reddi, H.V., Rajkovic, A., et al. (2021). Screening for autosomal recessive and X-linked conditions during pregnancy and preconception: a practice resource of the American College of Medical Genetics and Genomics (ACMG). Genet. Med. *23*, 1793–1806. https://doi.org/10.1038/s41436-021-01203-z.
8. Grody, W.W., Cutting, G.R., Klinger, K.W., Richards, C.S., Watson, M.S., Desnick, R.J.; and Subcommittee on Cystic Fibrosis Screening Accreditation of Genetic Services Committee ACMG American College of Medical Genetics (2001). Laboratory standards and guidelines for population-based cystic fibrosis carrier screening. Genet. Med. *3*, 149–154. https://doi.org/10.1097/00125817-200103000-00010.
9. Prior, T.W., Professional Practice and Guidelines Committee, and Guidelines, C. (2008). Carrier screening for spinal muscular atrophy. Genet. Med. *10*, 840–842. https://doi.org/10.1097/GIM.0b013e318188d069.
10. Committee Opinion No (2017). 690: Carrier Screening in the Age of Genomic Medicine. Obstet. Gynecol. *129*, e35–e40. https://doi.org/10.1097/AOG.0000000000001951.
11. Guo, M.H., and Gregg, A.R. (2019). Estimating yields of prenatal carrier screening and implications for design of expanded carrier screening panels. Genet. Med. *21*, 1940–1947. https://doi.org/10.1038/s41436-019-0472-7.
12. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.
13. Chen, S., Francioli, L.C., Goodrich, J.K., Collins, R.L., Kanai, M., Wang, Q., Alföldi, J., Watts, N.A., Vittal, C., Gauthier, L.D., et al. (2024). A genomic mutational constraint map using variation in 76,156 human genomes. Nature *625*, 92–100. https://doi.org/10.1038/s41586-023-06045-0.
14. Schmitz, M.J., Aarabi, M., Bashar, A., Rajkovic, A., Gregg, A.R., and Yatsenko, S.A. (2022). Carrier frequency of autosomal recessive genetic conditions in diverse populations: Lessons learned from the genome aggregation database. Clin. Genet. *102*, 87–97. https://doi.org/10.1111/cge.14148.
15. Schmitz, M.J., Bashar, A., Soman, V., Nkrumah, E.A.F., Al Mulla, H., Darabi, H., Wang, J., Kiehl, P., Sethi, R., Dungan, J., et al. (2025). Leveraging diverse genomic data to guide equitable carrier screening: Insights from gnomAD v.4.1.0. Am. J. Hum. Genet. *112*, 181–195. https://doi.org/10.1016/j.ajhg.2024.11.004.
16. Mbarek, H., Devadoss Gandhi, G., Selvaraj, S., Al-Muftah, W., Badji, R., Al-Sarraj, Y., Saad, C., Darwish, D., Alvi, M., Fadl, T., et al. (2022). Qatar genome: Insights on genomics from the Middle East. Hum. Mutat. *43*, 499–510. https://doi.org/10.1002/humu.24336.
17. Al Thani, A., Fthenou, E., Paparrodopoulos, S., Al Marri, A., Shi, Z., Qafoud, F., and Afifi, N. (2019). Qatar Biobank Cohort Study: Study Design and First Results. Am. J. Epidemiol. *188*, 1420–1433. https://doi.org/10.1093/aje/kwz084.
18. Lazarin, G.A., Hawthorne, F., Collins, N.S., Platt, E.A., Evans, E.A., and Haque, I.S. (2014). Systematic Classification of Disease Severity for Evaluation of Expanded Carrier Screening

Panels. PLoS One *9*, e114391. https://doi.org/10.1371/journal.pone.0114391.
19. Saffi, M., and Howard, N. (2015). Exploring the Effectiveness of Mandatory Premarital Screening and Genetic Counselling Programmes for beta-Thalassaemia in the Middle East: A Scoping Review. Public Health Genom. *18*, 193–203. https://doi.org/10.1159/000430837.
20. Alhosain, A. (2018). Premarital Screening Programs in the Middle East, from a Human Right's Perspective. Diversity and Equality in Health and Care *15*, 41–45. https://doi.org/10.21767/2049-5471.1000154.
21. Al-Dewik, N., Al-Mureikhi, M., Shahbeck, N., Ali, R., Al-Mesaifri, F., Mahmoud, L., Othman, A., AlMulla, M., Sulaiman, R.A., Musa, S., et al. (2018). Clinical genetics and genomic medicine in Qatar. Mol. Genet. Genomic Med. *6*, 702–712. https://doi.org/10.1002/mgg3.474.
22. Alaifan, M.A., Abusharifah, O., Bokhary, R.Y., Banaganapalli, B., Shaik, N.A., Kamal, N.M., and Saadah, O.I. (2022). Granulomatous hepatitis in a Saudi child with IL2RA defect: a case report and literature review. Ther. Adv. Chronic Dis. *13*, 20406223221116798. https://doi.org/10.1177/20406223221116798.
23. Sack, J., Front, H., Kaiserman, I., and Schreiber, M. (1997). 21-Hydroxylase Deficiency: Screening and Incidence in Israel. Horm. Res. *48*, 115–119. https://doi.org/10.1159/000185500.
24. Al-Obaidi, R.G.Y., Al-Musawi, B.M.S., Al-Zubaidi, M.A.K., Oberkanins, C., Németh, S., and Al-Obaidi, Y.G.Y. (2016). Molecular Analysis of CYP21A2 Gene Mutations among Iraqi Patients with Congenital Adrenal Hyperplasia. Enzyme Res. *2016*, 9040616. https://doi.org/10.1155/2016/9040616.
25. Al Hosani, H., Salah, M., Osman, H.M., Farag, H.M., El Assiouty, L., Saade, D., and Hertecant, J. (2014). Expanding the comprehensive national neonatal screening programme in the United Arab Emirates from 1995 to 2011. East. Mediterr. Health J. *20*, 17–23. https://doi.org/10.26719/2014.20.1.17.
26. Skrinska, V., Khneisser, I., Schielen, P., and Loeber, G. (2020). Introducing and Expanding Newborn Screening in the MENA Region. Int. J. Neonatal Screen. *6*, 12. https://doi.org/10.3390/ijns6010012.
27. Al-Sulaiman, R., Othman, A., El-Akouri, K., Fareed, S., AlMulla, H., Sukik, A., Al-Mureikhi, M., Shahbeck, N., Ali, R., Al-Mesaifri, F., et al. (2020). A founder RAB27A variant causes Griscelli syndrome type 2 with phenotypic heterogeneity in Qatari families. Am. J. Med. Genet. *182*, 2570–2580. https://doi.org/10.1002/ajmg.a.61829.
28. Ali, R., Al-Dewik, N., Mohammed, S., Elfituri, M., Agouba, S., Musa, S., Mahmoud, L., Almulla, M., El-Akouri, K., Mohd, H., et al. (2022). Expanding on the phenotypic spectrum of Woodhouse-Sakati syndrome due to founder pathogenic variant in DCAF17: Report of 58 additional patients from Qatar and literature review. Am. J. Med. Genet. *188*, 116–129. https://doi.org/10.1002/ajmg.a.62501.
29. Faiyaz-Ul-Haque, M., Zaidi, S.H.E., Al-Mureikhi, M.S., Peltekova, I., Tsui, L.C., and Teebi, A.S. (2007). Mutations in the CHX10 gene in non-syndromic microphthalmia/anophthalmia patients from Qatar. Clin. Genet. *72*, 164–166. https://doi.org/10.1111/j.1399-0004.2007.00846.x.
30. El-Harith, E.H.A., Roesl, C., Ballmaier, M., Germeshausen, M., Frye-Boukhriss, H., von Neuhoff, N., Becker, C., Nürnberg, G., Nürnberg, P., Ahmed, M.A.M., et al. (2009). Familial thrombocytosis caused by the novel germ-line mutation p.Pro106Leu in the MPL gene. Br. J. Haematol. *144*, 185–194. https://doi.org/10.1111/j.1365-2141.2008.07430.x.
31. Xiao, Q., and Lauschke, V.M. (2021). The prevalence, genetic complexity and population-specific founder effects of human autosomal recessive disorders. npj Genom. Med. *6*, 41. https://doi.org/10.1038/s41525-021-00203-x.
32. Almarri, M.A., Haber, M., Lootah, R.A., Hallast, P., Al Turki, S., Martin, H.C., Xue, Y., and Tyler-Smith, C. (2021). The genomic history of the Middle East. Cell *184*, 4612–4625.e14. https://doi.org/10.1016/j.cell.2021.07.013.
33. Davidson, A.L., Leonard, C., Koufariotis, L.T., Parsons, M.T., Hollway, G.E., Pearson, J.V., Newell, F., Waddell, N., and Spurdle, A.B. (2021). Considerations for using population frequency data in germline variant interpretation: Cancer syndrome genes as a model. Hum. Mutat. *42*, 530–536. https://doi.org/10.1002/humu.24183.
34. Ben-Shachar, R., Svenson, A., Goldberg, J.D., and Muzzey, D. (2019). A data-driven evaluation of the size and content of expanded carrier screening panels. Genet. Med. *21*, 1931–1939. https://doi.org/10.1038/s41436-019-0466-5.
35. Kirk, E.P., Ong, R., Boggs, K., Hardy, T., Righetti, S., Kamien, B., Roscioli, T., Amor, D.J., Bakshi, M., Chung, C.W.T., et al. (2021). Gene selection for the Australian Reproductive Genetic Carrier Screening Project ("Mackenzie's Mission"). Eur. J. Hum. Genet. *29*, 79–87. https://doi.org/10.1038/s41431-020-0685-x.
36. Gruzin, M.J., Hobbs, M., Ellsworth, R.E., Poll, S., Aguilar, S., Knezovich, J., Faulkner, N., Olsen, N., Aradhya, S., and Burnett, L. (2025). Optimizing gene panels for equitable reproductive carrier screening: The Goldilocks approach. Genet. Med. *27*, 101387. https://doi.org/10.1016/j.gim.2025.101387.
37. Ghiossi, C.E., Goldberg, J.D., Haque, I.S., Lazarin, G.A., and Wong, K.K. (2018). Clinical Utility of Expanded Carrier Screening: Reproductive Behaviors of At-Risk Couples. J. Genet. Couns. *27*, 616–625. https://doi.org/10.1007/s10897-017-0160-1.
38. Capalbo, A., Fabiani, M., Caroselli, S., Poli, M., Girardi, L., Patassini, C., Favero, F., Cimadomo, D., Vaiarelli, A., Simon, C., et al. (2021). Clinical validity and utility of preconception expanded carrier screening for the management of reproductive genetic risk in IVF and general population. Hum. Reprod. *36*, 2050–2061. https://doi.org/10.1093/humrep/deab087.
39. Wang, T., Kiss, D., McFadden, K., Byrnes, J., Scuffham, P., Delatycki, M.B., and Downes, M. (2023). Clinical utility of reproductive carrier screening for preconception and pregnant couples for multiple genetic conditions: a systematic review and meta-analysis. Expert Rev. Mol. Diagn. *23*, 419–429. https://doi.org/10.1080/14737159.2023.2206519.
40. Dive, L., Laberge, A.-M., Freeman, L., and Bunnik, E.M. (2025). Beyond severity: utility as a criterion for setting the scope of RGCS. Eur. J. Hum. Genet. *33*, 226–230. https://doi.org/10.1038/s41431-024-01640-9.
41. Abdu, Y., Ahmed, K., Ibrahim, M.I.M., Abdou, M., Ali, A., Alsiddig, H., Selim, N.A., and Yassin, M.A. (2023). Perception of consanguineous marriage among the qatari population. Front. Public Health *11*, 1228010. https://doi.org/10.3389/fpubh.2023.1228010.
42. Alswaidi, F.M., and O'Brien, S.J. (2009). Premarital screening programmes for haemoglobinopathies, HIV and hepatitis viruses: review and factors affecting their success. J. Med. Screen *16*, 22–28. https://doi.org/10.1258/jms.2008.008029.