Heterogeneity in biomarkers, mitogenome and genetic disorders of Arab population with special emphasis on large-scale whole-exome sequencing

Type
Review paper (After Acceptance of EiC)

Keywords
Exome, Arab world, Mendelian Genes, Intellectual disability, Medical Genetics, Genetic disorders, Pathogenic Mutation

Abstract
More than 25 million DNA variations were discovered as novel including major alleles from Arab population. Exome studies on Arabs discovered >3000 novel nucleotide variants associated with >1200 rare genetic disorders. Reclassification of many pathogenic variant into benign through the Arab database enhance building a detailed and comprehensive map of Arab morbid genome. Intellectual disability stands first with the combined and observed carrier frequency. Genome studies and advanced computational biology discovered interesting novel candidate disease marker variations in many genes from consanguineous families with intellectual disability, neurogenetic disorders, blood and bleeding disorder and rare genetic diseases. Pathogenic variants in C12orf57 gene are prominently associated with the etiology of developmental delay/intellectual impairment. Arab mitogenome exposed hundreds of variations in mtDNA genome and its association with obesity. Further study is needed in genomics to fully comprehend the molecular abnormalities and associated pathogenesis that cause inherited disorders in Arab ancestries.
Heterogeneity in biomarkers, mitogenome and genetic disorders of Arab population with special emphasis on large-scale whole-exome sequencing

J Francis Borgio*

Department of Genetic Research, Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam 31441, Saudi Arabia. Email: fbalexander@iau.edu.sa; borgiomicro@gmail.com

Short title: Genomics and Biomarkers of Arab ancestries

Abstract word count : 307
Total word count : 6915
Total number of figures : 6
Total number of Tables : 1
Number of references : 159

* Corresponding authors:

Dr. J. Francis Borgio, Associate professor of Molecular Genetics, Department of Genetic Research, Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia.

Phone: 00966 13 3330864.

ORCID ID: https://orcid.org/0000-0001-7199-1540

Email: fbalexander@iau.edu.sa, borgiomicro@gmail.com
Heterogeneity in biomarkers, mitogenome and genetic disorders of Arab population with special emphasis on large-scale whole-exome sequencing

Abstract

More than 25 million DNA variations were discovered as novel including major alleles from Arab population. Exome studies on Saudi genome discovered >3000 novel nucleotide variants associated with >1200 rare genetic disorders. Reclassification of many pathogenic variants in Human Gene Mutation Database and ClinVar Database into benign through the Arab database enhance building a detailed and comprehensive map of human morbid genome. Intellectual disability stands first with the combined and observed carrier frequency of 0.06779 among Saudi Arabians; retinal dystrophy is the next highest. Genome studies discovered interesting novel candidate disease marker variations in many genes from consanguineous families. More than 7 pathogenic variants in C12orf57 gene are prominently associated with the etiology of developmental delay/intellectual impairment in Arab ancestries. Advancement in the large-scale genome studies opens a new outlook on Mendelian genes and disorders. In the past half-dozen years, candidate genes of intellectual disability, neurogenetic disorders, blood and bleeding disorder and rare genetic diseases are well documented through genomic medicine studies in combination with advanced computational biology applications. Arab mitogenome exposed hundreds of variations in mtDNA genome and ancestral sharing with Africa, the Near East and East Asia and its association with obesity. These recent discoveries on disease markers and molecular genetics of Arab population will have a positive impact towards supporting the genetic counsellors on reaching consanguineous families to manage stress linked to genetics and precision medicine. This narrative review summarises the advancements of molecular medical genetics and recent discoveries on pathogenic variants. Despite the fact that these initiatives are targeting the genetics and genomics of disorders prevalent in Arab populations, a lack of complete cooperation across the projects needed to be revisited to uncover the Arab population's prominent disease markers. This shows that further study is needed in genomics to fully comprehend the molecular abnormalities and associated pathogenesis that cause inherited disorders in Arab ancestries.

Keywords: Arab world; Mendelian Genes; Intellectual disability; Medical Genetics; Genetic disorders; Pathogenic Mutation; Exome
**Introduction**

The Arab population or people speaking the Arabic language are reported with a significantly high incidence of genetic abnormalities due to high rate of first-cousin marriages [1]. Increased level of inbreeding, large family sizes, high mother and father ages characterize the Arab population. In earlier days not much familiar public health efforts are aimed at preventing genetic diseases, but recently there are country-wide initiatives for screening and preventive measures have been initiated. In the recent past, so many genetic disorders have given importance in the field of medical genetics to identify disease-causing mutations using next-generation sequencing techniques in Arab region. Vast deserts, suffocating heat, Oil wells, throngs of pilgrims, and prosperous-looking cities continue to dominate popular opinions of Arab population especially Saudi Arabia. This view is now expanded to incorporate advanced initiative that is claiming a global spotlight as one of the top human genomics projects: the Saudi Human Genome Program (SHGP), Kuwait Genome Project (KGP), Qatar Genome Programme (QGP), Emirati Genome Program (EGP), and Bahrain Genome Project (BGP). Arab countries such as Kingdom of Saudi Arabia, United Arab Emirates and Qatar have invested a lot of money and efforts in the next-generation genomics to reveal the molecular basis of diseases and disorders prevalence in the Arab population. The post-Arab genome programs are a golden era in molecular genetics and disease biomarkers of Arabian genomic studies. These Arab genome programs have overcome the obstacles and limitations on collecting samples. The Program SHGP, KGP, QGP, EGP and BGP are visionary projects on human genomics with the vision to detect and prevent genetic diseases in the Kingdom of Saudi Arabia (KSA), the State of Kuwait, the State of Qatar, the United Arab Emirates (UAE) and the Kingdom of Bahrain, respectively to establish world-class facility on genomic research in the Arab ancestry [2–6]. The SHGP was implemented by King Abdulaziz City for Science and Technology, on its way to reach sequence 100,000 human genomes in Saudi Land to achieve gold standard findings in the Saudi population on genomics-based biomedical research [2]. The central laboratory of SHGP in 2018, is one of the first national initiatives and the largest genomic project in the Middle East; details can be obtained from the freely available database (https://shgp.kacst.edu.sa/index.en.html) [7]. The SHGP initiative revealed >3000 new nucleotide variants to the literature, that are associated with >1200 rare genetic disorders [7], [8], [9]. Earlier reviews on population structure and prevalent genetic disorders can help to understand the initial studies on Mendelian genes and
associated traditions in the high consanguineous community of Kingdom of Saudi Arabia and other Arab populations [10–12]. The State of Qatar, which is located in the Arabian Gulf established the QGP (https://www.qatargenome.org.qa/), a Qatari population-based genome initiative aimed at sequencing the complete genomes of a large part of the Qatari population, began its pilot phase in 2015 [3]. The Kuwait Genome Project (KGP) is a high-coverage effort to sequence Kuwaiti genomes from genetically different subgroups of Kuwaitis [4].

Although genome researches have the potential to transform medical approaches related to prevention of disorders, developing advanced diagnosis, and formulating treatment strategies, it is not without its obstacles and limitations [3]. The willingness of individuals to volunteer samples and use genetic testing services is critical to achieve the goals of the national genome initiatives like KGP, SHGP, QGP, BGP and EGP. Basic genetic knowledge, a genetic history of families with genetic disorders, and prior experience were all found to be significant predictors of willingness to Arab participate. The findings from 837 adult Qatari show widespread acceptance, but they also the study point to the need for greater education and personalized counselling to explain the procedure, obstacles, and advantages of genetic testing [3]. These Arab genome initiatives have surmounted the challenges and constraints of sample collection. Whole exome sequencing (WES) on evaluating 48 Qatari (30 families) patients with inherited Mendelian disorders from the "Qatar Mendelian Disease pilot program," revealed 25 unique disease-causing pathogenic alleles that are important to premarital and newborn screening Arab populations [13]. Next-generation sequencing method like WES plays important role in molecular pathogenesis [14,15]. The narrative review reviews the advancements throughout the Arab genome programme period, including major breakthroughs in pathogenic disease associated variations and advances in molecular genetics from large-scale whole-exome sequencing data.

**Novel variants and reclassifying pathogenic variants into benign**

Initial study on Qatari revealed 58.37 % of novel variants among the identified 5,452,613 variants [16], significant differences were observed on the initial studies on Saudi genome [17], and Kuwaiti genome [4,18]. Still more in-depth genome-wide studies are needed to overcome the understudied [19]. A recent findings from QGP phase 1, which sequenced the whole genomes of 6,045 Qatari. They found almost 88 million DNA variations. Interestingly, with 24,620,313 of them being novel to various genome databases [20]. The study discovered that many uncommon deletion variations were more common
and prevalent in the Qatari population. Most of the observed novel variants appear to give disease protection in the Qatari population and have changed the genetic architecture, which is consistent with the region's high consanguinity. Five non-admixed groupings were discovered in the genetic makeup of the Qatari population. The study also described the heritability of genetic disease marker correlations for 45 clinical phenotypes [20]. Recent study on reclassification of pathogenic variant [16 Human Gene Mutation Database (1 disease mutation and 15 disease mutation) variants and 10 ClinVar (6 pathogenic, 4 variant of uncertain significance) variants] into benign by the team of Saudi Human Genome Program (SHGP) database was possible because of the world's biggest collection of genetic variations from Arabs through the establishment of SHGP database [8]. SHGP database using 3070 disease associated genes (including 2000 whole-exome sequencing WES) covering >4000 Mendelian disorders of 5849 non-overlapping individuals can reveal Arab-enriched and Arab-specific common variants in the Arab population in general and Saudi population in particular [8]. It was expected at least some reported “genetic disease associated variants” might characterize Arab-enriched or Arab-specific common mutations that might not be detected as such by utilizing these genomic data as most of the disease associated genes and variants were described for the first time in Arab subjects with disease phenotype due to the high level of consanguinity. Also these factors enables positional mapping in genome, and since the ethnicity is not much represented in publicly accessible variant databases (Figure 1). The study reveals that 16 HGMD variants and 10 ClinVar variants had a Saudi ancestry frequency (in population) of 5% in the database (SHGP) but only <5% in publicly available databases, allowing all these 26 pathogenic variants be reliably reclassify as benign. Autozygosity is rich in SHGP database; based on Saudi population frequency (>1%) and occurrence of homozygosity genotype in Saudi individuals who lack the phenotype described already, the study was able to reclassify 484 ClinVar variants [184 Pathogenic, 25 Likely Pathogenic, 275 Variant of uncertain significance] and 607 HGMD variants [103 disease mutation and 504 disease mutation] as benign [8]. Due to the founder effect, QGP discovered certain rare pathogenic variants mutations with high MAF in the Qatari population [20]. Variants, rs750046020 in MPL gene, specifically linked with thrombocytosis, appear at MAF of 0.009 in Qatari Arab population [20]. Similarly, at a MAF of 0.007, variations in the genes KRT5 (rs267607448) and CBS (rs398123151) linked to Epidermolysis Bullosa and homocystinuria, respectively, are found in the Qatari population [20]. Ten novel mutations were discovered in ten known genes associated with myeloproliferative neoplasms, and seven potential candidate genes were discovered with
seven novel variants from Qatari patients with myeloproliferative neoplasms [21]. Total of 1,790,171 annotated variants and 8,462 structural variants were found novel variations across the Emirathí genome, were found while constructing the United Arab Emirates Reference Genome [22]. Kuwaiti genome studies discovered 58,186 SNPs, 32,686 indels as novel variants [4,18]. Most of the coding novel variants discovered in the Kuwaiti genome studies were found to be associated with autosomal recessive disorders in the Arab population [18].

**Figure 1:** Reclassified variants in HGMD (Human Gene Mutation Database) and ClinVar Database through SHGP database observations.

**Carrier frequency of genetic disorders**

The combined and observed carrier frequency of genetic disorders among Saudi Arabians was estimated for the first time by utilizing a large sampling (n=7,101 patients) using targeted gene panel sequencing and whole-exome sequencing [7]. The study reveals the most prevalent carrier status of various diseases (**Figure 2**): Intellectual disability stands first with the combined and observed carrier frequency of 0.06779, followed by Retinal dystrophy, Glaucoma, Inborn errors of metabolism, Sickle cell disease/thalassemia, Deafness, Dysmorphic/dysplasia, Ataxia, Myopathy/muscular dystrophy, Polycystic kidney disease/nephronophthisis, Joubert syndrome/Meckel Gruber syndrome, Carbonic anhydrase II deficiency, Cystic Fibrosis, Bardet–Biedl syndrome, and Cataract [7] (**Figure 2**). Furthermore, Study on 2357 Saudi patients through gene panels, revealed 433 novel disease alleles and 355 existing variants, which is the largest notable number of novel disease alleles submitted to the HGMD in a single genomic study. Third, the 433 novel disease variants identified in the genome study by Saudi Mendeliome Group represent an exceptional resource on the Arab variome (whole set of genetic variations) as most of the study participants were Arab origin. Hence these observations and additions to the HGMD by the Saudi genome studies is an invaluable resource for the interpretation of clinical and molecular genomic analysis on Mendelian genes in Arab patients [23]. These novel disease markers will help address the uncertainty in the identification of many Arab-specific disease markers. All these studies reveal a high prevalence of variants and carriers, as the degree of consanguinity is high among the Saudi ancestry, hence studies were able to identify a lot of autozygosity-related variations in homozygosity [7,23]. The large-scale exome studies on the carrier frequency of genetic disorders among the Arab population changes the previous observations of the most prevalent hemoglobinopathies [7,11]. Carrier frequency of the intellectual disability is three times
more than that of sickle cell disease and thalassemia among the Arab population with 25-60% of consanguinity rates [7,11]. In-depth functional analysis of observed intellectual disability genetic markers can help for the early identification of the intellectual disability genetic markers and designing novel treatment strategies. The most prevalent recessive alleles discovered in the Qatari population are those associated with developmental disorders and structural deformities [20]. Nothing surprising that the large number of Arabs in Saudi Arabia share similar and common carrier frequency with Qatar’s native population with minor differences in some alleles (DCAF17: c.436delC) [20].

**Figure 2**: Combined and observed carrier frequency of genetic disorders in Arab ancestries with an example from Saudi Arabian population (n=7,101).

**Clinically relevant traits and precision medicine**

Foundation stone was laid studies on the Qatari population on the clinically relevant traits and loci associated genetics along with earlier studies [16,24]. Genetics import important role on the output and diversity in the clinical laboratory tests results and decision-making in the precision medicine [24]. Furthermore, very little is known about the genetic variability of Arab population compare to the other ancestries. A detailed study presented the results of genome-wide association research on forty-five clinically relevant features among Qatari population, conducted by means of WES technique of 6218 subjects [24,25]. They reported Arab (Qatari) and European populations with similar trait heritability (r = 0.81). Moreover, Arab (Qatari) ancestries are less similar with African ancestries (r = 0.44) on trait heritability. Authors found 281 unique and significant associations between the variants and clinical traits through GWAS analysis, which coincide with previously discovered associations with other populations mainly Europeans. The relationships between allele frequencies for repeated loci are lower in Japanese (r = 0.80) and African (r = 0.85) while higher in European (r = 0.94) populations [26–29]. Authors presented 17 novel and Arab-dominant signals based on Qatars that shed light on the molecular pathways that control the clinical traits [24]. The polygenic scores developed from European data have lower predictive performance in the Qatar based Arab population. Findings from the large-scale analysis on the clinical traits of the Arab population are expected to have an impact on its future use in precision medicine and genetic architecture of complex diseases prevalent in Arab population. Finally, for the first time in a Arab population, this study on heritability analysis and GWAS-based genetic association analysis on clinically relevant variables precisely identified shared genetic loci with variation in linkage disequilibrium patterns [24]. Metabolic quantitative trait loci identification analysis revealed genetic loci associated with pharmacological/drug targets on the consanguineous Arab population, which is important to precision
medicine in the Arab world [30]. Constructing specific references sequences for a subpopulation of Arab ancestry is also important in precision medicine, the initiation of the UAE reference panel is a good initiative [22,31,32].

**High genetic heterogeneity**

Arab population is known for its high genetic heterogeneity on the diversity of phenotypes and clinical outcomes [1,33–38]. The Arab World is an ethnocultural melting pot located geographically as a land bridging Eurasia and Africa, it functioned as a gateway for the migration of human from African region to the rest of the civilized world. The wide range of pathogenic gene mutations linked to various genetic diseases in the Arabs perhaps best exemplifies the heterogeneity of Arabs [1]. The high consanguinity of Arab people is well-known. Because of the complexity of consanguinity and heterogeneity, understanding the causes of Arab population-specific genetic disorders in the region is challenging. The Arab region is populated by genetically diverse populations, as evidenced by previous genetic studies [1,33–36]. Inter-regional genetic variability in Arab ancestry is also highlighted in a recent GWAS study [36]. The prevalence of heterogeneity in genetics of highly consanguineous Arab families makes the interpretation for the genetic basis and etiology of numerous inherited disorders most common in Arab populations. In Arab countries, more than 1200 distinct genetic disorders have been documented, 60% of which are autosomal recessive disorders and more than 40% are confined to a certain demographic or geographical location [1,10–12,33]. High-resolution population genetic studies on the Arab population were carried out to understand the genetic diversity on inherited diseases and the degree of heterogeneity among Kuwaiti Arabs using SNP microarray (n=583) [1]. Within the newly established groupings, the study infers the primary sources of ancestry among Kuwaiti Arabs. Various findings reinforce the genetic imprints and regional genetic heterogeneity in Arab people on the blood disorders, metabolic disorders, disorders of the circulatory system, breast cancer, ovarian cancer, colorectal carcinoma and prostate cancer [1].

**Neurogenetic disorders and candidate genes**

The most common referral request for WES analysis was for neurological diseases in a Arab country (Lebanon) [39]. WES analysis of pre-screened multiplex consanguineous families with neurogenetic disorders reveals candidate genes and disease makers [40]. The 33 genes (observed phenotype) such as SPDL1 (primary microcephaly and neonatal death), TUBA3E (microlissencephaly and global developmental delay), INO80, TSEN15, and PCDHB4 (global developmental delay or GDD and primary microcephaly), DMBX1 (epilepsy, GDD, and poor
weight gain), CLHC1 (myopathy), NID1 (muscle weakness, hydrocephalus, and global developmental delay), C12orf4 (global developmental delay), WDR93 (autism spectrum disorder), ST7 (brain atrophy and global developmental delay), MATN4 (holoprosencephaly), SEC24D, and MYOCD (intellectual disability and epilepsy), KIAA1109 (Dandy-Walker malformation, club feet, flexed deformity, hydrocephalus, pleural effusion and micrognathia), PTPN23 (global developmental delay and brain atrophy), TAF6 (global developmental delay and dysmorphism), TBCK (hypotonia, GDD, dysmorphism, epilepsy, and VSD), FAM177A1 (macrocephaly, dolichocephaly, intellectual disability, and mild obesity), MTSS1L (neurodegeneration and brain iron accumulation), ARV1 (neurodegenerative disease), XIRP1 (primary microcephaly), CHAF1B (global developmental delay and ADHD), KCTD3 (cerebellar hypoplasia, seizure, and severe psychomotor retardation), ISCA2 (consistent with mitochondrial encephalopathy and neurodegeneration with high lactate peak in the brain and marked white matter changes.), PDPR (typical Joubert syndrome and GDD), PTRH2 (hearing loss, and ataxia also GDD), GEMIN4 (severe osteopenia congenital cataract, tubulopathy, and GDD), DPH1 (Dandy-Walker malformation, developmental delay, cerebellar vermis hypoplasia, and hydrocephalus.), NUP107 (early onset focal segmental glomerulosclerosis, light complexion, and global developmental delay), TMEM92 (cerebellar atrophy, hydrocephalus; cognitive, speech, and motor delay), EPB41L4A (spastic paraplegia, failure to thrive), and FAM120AOS (GERD, chronic lung disease, coarse facial features, skin laxity, scoliosis, hypotonia, undescended testicles and pectus excavatum,) were identified among the pre-screened multiplex consanguineous families with neurogenetic disorders. Missense (n=22) variations were common among the disease variations in the patients with neurogenetic disorders (Figure 3) [40]. RTTN and ASPM genes were associated with primary microcephaly apart from the above listed [40–42]. Cerebellar ataxia in sibling WES study showed novel variant in CWF19L1 gene [43] and GBA2 gene in consanguineous Saudi family [44]. A study from Qatari subject revealed the association of novel pathogenic PGAP3 variant with global developmental delay and neuromuscular abnormalities and brain anomalies [45]. Studies from Egyptian subject revealed the association of pathogenic SCN10A variation in patients with neuromuscular disease and epileptic encephalopathy [46], SLC39A8 variation in patients with intellectual disability, developmental delay, cerebellar atrophy, hypotonia, strabismus, and variable short stature [47]. PQBP1 variation in patients with intellectual disability [48] biotin-thiamine responsive encephalopathy [49] and TTC5 variation in patients with intellectual disability syndromes [50]. PUS3 gene with frameshift variant was observed as candidate gene for intellectual disability in the study with 103
families from Jordan [51]. A family WES study reveals variations in the genes ASPA and ARSA in patients with leukodystrophy in Jordan [52]. Trio Exome Sequencing of four families from Jordan identified the multi-exon deletion in VPS13B gene in the patients with Cohen syndrome [53]. Consanguineous families from Jordan with neuromuscular disorders identified multiple variants WES in DYSF, MPV17, SLC25A46 and SGCG genes [54]. Lebanese subjects with neurodevelopmental and neuromuscular disorders from 200 WES analysis and other studies revealed various genes with pathogenic variations [55–57]. In order to identify the associated pathways and biological processes associated with the genes identified as the neurogenetic disorders candidate genes of Arab population, protein-protein interaction analysis was carried out using STRING 10 [58]. The analysis revealed that the genes are linked in terms of text mining, no significant associated pathways were identified, this indicates that the impact of cellular processes and pathways are needed to be studied to reveal the neurogenetic disorders candidate genes on the neurogenetic disorders. Incorporation of proteomics and pathway analysis may reveal the impact of individual genes on neurogenetic disorders in Arab population and network of genes associated pathways on neurogenetic disorders.

Figure 3: Neurogenetic disorders and candidate genes from WES analysis of pre-screened multiplex consanguineous families of Arab origin. Protein network indicates the protein-protein interaction analysis of genes of neurogenetic disorders candidate genes of Arab population.

Blood and bleeding disorder

Molecular defects blood disorders, beta thalassemia, sickle cell disorder, alpha thalassemia and G6PD (Glucose-6-Phosphate Dehydrogenase) deficiency are the most common in the Arab population are extensively listed and reviewed earlier [1,59–63]. The large-scale (~5000) WES analysis by the SHGP on the 1285 cases with 17 blood and bleeding disorders reveals more novel variants (n=140) than previously reported pathogenic variants (n=98) (Table 1) among the 821 variants [9]. VWF, F8, F5, G6PD, F2, F7, F10, F13B, FGA, and HBA2 genes observed with novel variants. Authors have prioritised the list of genes, and novel variants in their study, which will definitely be a source of genetic data and will aid in the development of molecular variants screening and to enhance the efficiency of the preventive programs for the blood and bleeding disorders prevalent in the Kingdom [9]. This
molecular data will also make a significant impact on developing molecular diagnostic tools for the clinically overlapping disorders with blood and bleeding disorders and to designed treatment strategies. VWF, F10, HBB, F7, and F5 genes observed with notable carrier frequency on the variants associated with Saudi Ancestry with blood and bleeding diseases. VWF, G6PD and F8 genes are considered as the important genes of interest on the prevailing blood and bleeding disorders [9]. The study very precisely ranks the variants based on novelty, pathogenicity and existence in the genomic databases. The study suggested for prioritising autosomal recessive variants in HBB, F7, F13A1, F11 and VWF genes to screen the carrier for the blood and bleeding disorders prevalent in Saudi population [9].

Table 1: Novel variants observed and its significance in Saudi population.

<table>
<thead>
<tr>
<th>Study description</th>
<th>Previously reported pathogenic variants</th>
<th>Novel variants</th>
<th>Prominent Genes with Novel variants</th>
<th>Novel variants and notable phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood and bleeding disorder</td>
<td>98 (stop loss =1, synonymous =2, splice and UTR =11, truncating =12, variants of unknown significance = 19, benign or likely benign = 27, re-classified pathogenic or likely pathogenic = 52, CADD score &gt;20 = 56, non-synonymous = 72)</td>
<td>140 (Nonsynonymous = 117, CADD score &gt;20 = 71, pathogenic = 32, splice = 6, Truncating = 4, frameshifts = 13, likely benign = 3, variants of unknown significance = 105)</td>
<td>VWF, F8, F5, G6PD, F2, F7, F10, F13B, FGA, HBA2</td>
<td>3 truncating variants observed in 2 cases of aplastic anemia, and one patients with excessive bleeding</td>
<td>[9]</td>
</tr>
<tr>
<td>Familial transthyretin (TTR) amyloidosis (ATTR)</td>
<td>158 TTR variants. (coding or flanking regions = 28, missense TTR variants = 12, uncertain significance = 3, negatively impact TTR function resulting in amyloidosis = 2;)</td>
<td>3 novel TTR c.404C&gt;T (p.S135F), c.298A&gt;G (p.K100E) and c.428C&gt;T (p.T143I,) (Nonsynonymous = 3, interface in the middle of the TTR tetramer and thyroxine =1; interface in the middle of two TTR molecules and RBP4 = 1)</td>
<td>TTR</td>
<td>only 1 allele of each novel TTR variant being observed</td>
<td>[64]</td>
</tr>
<tr>
<td>Antenatal cystic kidney disease and ciliopathy</td>
<td>8 (CC2D2A, c.3084delG p.Arg1028Rfs<em>3; CC2D2A, c.3364C&gt;T p.Pro1122Ser; CC2D2A, p.Trp1511Arg c.1413-2A&gt;G; CEP290, c.5668G&gt;T p.Gly1890</em>; PKHD1, c.4870C&gt;T</td>
<td>13 novel pathogenic variants (B9D1, c.508_510delCTCT p.L170del; MKS1, c.417+1G&gt;A; INVS, c.1760delA p.Q587Rfs<em>2; MKS1, c.1066C&gt;T p.Q356</em>; TMEM67, c.457T&gt;G p.C153G; TCTN2, c.1852C&gt;T p.Q618*; TMEM67, c.1413-2A&gt;G; MKS1, c.1066C&gt;T p.Q356*; CEP290, c.3777_3778delAG p.R1259del*16; PKHD1, c.3539G&gt;A p.I1180E;</td>
<td>B9D1, CC2D2A, INVS, MKS1, MKS1, TCTN2, TMEM67, CEP290, NEK8, NPHP3PKHD1, TCTN2</td>
<td>Fetal death (B9D1, CC2D2A, INVS, MKS1); Perinatal death (INVS, TCTN2); Ascites (CC2D2A, c.4437+1G&gt;A); Pericardial effusion and congenital heart malformation in Saudi Ancestry (TMEM67, c.1413-2A&gt;G);</td>
<td>[65]</td>
</tr>
</tbody>
</table>
Familial transthyretin amyloidosis

Autosomal dominant disease, Familial transthyretin (TTR) amyloidosis (ATTR) Personalized for its genetic predisposition using 13906 Saudi exomes. Authors have identified many amino acid substitutions variants in TTR gene and revealed the potential association with amyloidosis (Figure 4) [64]. A total of 158 TTR mutations identified in the study population, coming back 28 were in the coding and franchising region including 12 nonsynonymous mutations during the analysis of 13906 Saudi exome. Among the 12 missense TTR variants: three [c.239C>T:p.T80I; c.424G>A:p.V142I and c.238A>G:p.T80A] with known impact on TTR functions; 3 [c.368G>A:p.R123H, c.385G>A:p.A129T and c.370C>T:p.R124C] with the unknown significance; three [c.76G>A:p.G26S, c.140A>G:p.N47S and c.328C>A:p.H110N] mutations are benign and likely benign and the remaining 3 [c.298A>G:p.K100E, c.404C>T:p.S135F, and c.428C>T:p.T143I] novel variants [64]. Among the 12 missense TTR variants, 2 mutations [less frequent c.238A>G:p.T80A (0.00004), most prevalent c.424G>A:p.V142I (0.001)] associated with amyloidosis, that with well-known pathogenic impact on the function of the TTR protein [64]. This study is the first large-scale study on analysing the TTR mutations from the Saudi population. Even though...
three novel variants were reported in this study on Saudi with systemic amyloidosis, the known and previously reported mutation, TTR:c.424G>A(p.V142I) is the variant observed most frequently in the Arabian ancestry [64]. This particular mutation is frequently identified in African descendants and African American population. Common Variant observed in Ireland and United Kingdom, TTR:c.238A>G(p.T80A) is rare in the Arabian population [64]. Many variants of TTR gene reported from Western Europe, Italian and Denmark populations were not observed in the Saudi ancestry, furthermore, the most common mutation observed in TTR gene from Chinese and Mexican populations was also not observed in the Saudi ancestry including the 5 regions of Saudi Arabia. This study has a major limitation accessing the clinical data related to the novel TTR variants and the variance observed in TTR gene, this very clearly indicates that more studies are needed on clinical and genetic studies in order to reveal the clinical manifestations of the observed mutations in the Saudi ancestry.

**Figure 4:** Novel variants observed in TTR gene from Saudi population with Familial transthyretin (TTR) amyloidosis (ATTR) or carrier. Modified from Abouelhoda et al. [64]. PDB ID: 2rox used for thyroxin-bound TTR and the PDB ID: 2wqa used for RBP-bound TTR.

**Fetal death and perinatal death**

A comprehensive genomic analysis on 34 families with fetal death, perinatal death and renal phenotype (Enlarged, echogenic with cysts, Cystic) revealed genes and novel variants associated with fetal death (B9D1, CC2D2A, and MKSI) and Perinatal death (INVS and TCTN2) [65]. The study revealed 13 novel pathogenic variants [(Frame shift: INVS:c.1760delA; and CEP290:c.3777_3778delAG); Deletions: B9D1:c.508_510delCTC; NPHP3:c.2694-1_2delAG; and TCTN2:c.252_253delTG; CC2D2A:c.4437+1G>A; MKSI:c.417+1G>A; MKSI:c.1066C>T; TCTN2:c.1852C>T; NEK8:c.1401G>A; TMEM67:c.457T>G; TMEM67:c.1413-2A>G; MKSI:c.1066C>T; PKHD1:c.3539G>A), this clearly indicates the genetic heterogeneity. In additions to renal defects, Ascites (CC2D2A), pericardial effusion and congenital heart malformation (TMEM67), and Hepatomegaly (PKHD1) were also reported [65]. The most prevalent cause of an antenatal ciliopathy is due to CC2D2A gene variation among the families studied. Variations in FRAS1 and FREM2 genes were identified in Six consanguineous Saudi Arabian families with antenatal/perinatal death [68].

**Intellectual disability**
Intellectual disability is the most prevalent genetic disorder in terms of carrier frequency in the Saudi population (Figure 2) [7]. Temtamy syndrome is a type of intellectual impairment with epilepsy, ocular involvement and corpus callosum dysgenesis. A detailed study on temtamy syndrome exposed a high carrier frequency of start loss mutation [c.1A>G; p.M1?] in Chromosome 12 Open Reading Frame 57 (C12orf57) gene in Saudi population, it was the first observation of the syndrome to C12orf57. The study showed variable phenotypic in the cases with intellectual disability. The study conclusively reported the C12orf57 gene variants in the pathology of intellectual disability / developmental delay in Arab population. Interestingly, the observations on the absence of typical syndromic features in Saudi ancestry are associated with the intellectual disability, which is common and recessive. The start loss mutation [c.1A>G; p.M1?] in C12orf57 gene is the most frequently prevalent (80.3% of all cases studied) disease maker in the Saudi ancestry (Figure 5) [69]. Another 6 variants in addition to start loss variation in the C12orf57 gene were reported (Figure 5) [69]. The seven variations, in the C12orf57 gene is Founder mutation c.1A>G, p.(Met1?), c.53-2A>G, c.184C>T, p.(Gln62*), p.(Gln15*), c.229+2T>C, c.43C>T, and c.152T>A, p.(Leu51Gln) accumulated in the exon 1 and 2 (Figure 5) [69]. C12orf57 gene was also associated with microphthalmia in the Saudi population (Table 1) [66]. Large-Scale Exome Sequencing and other studies reported the association between variations in AKAP6, ALKBH8, CLSTN2, EIF2A, IQSEC3, PLXNA2, PLXNA3, SMG8, BEAN1, HELZ, SLC4A4, SLC4A10 and TNIK and intellectual disability [70–75]. Intellectual disability patients (n=105) from 68 families identified intellectual disability candidate genes such as ASTN2, ANKHD1, FMO4, ATP13A1, GTF3C3, MADD, NCKAP1, STK32C, NFASC, MFSD11, NKX6-2, PCDHGA10, SLC12A2, PPP1R21, SLK, SPTBN4, TRAK1, and ZFAT [76]. Expression analysis confirmed the intellectual disability candidate gene, NCKAP1 [76]. A study from Qatari subject revealed the association of novel pathogenic PGAP3 variant with hyperphosphatasia with mental retardation, global developmental delay and neuromuscular abnormalities and brain anomalies [45]. Protein-protein interaction analysis of C12orf57 gene with literature using STRING revealed the high interaction with HLA genes and associated pathways (Figure 5B) [58]. Studies from Egyptians and Jordanians subject revealed the association of pathogenic SLC39A8, PQBP1, TTC5 and PUS3 variation in patients with intellectual disability [47], [48], [50], [51]. WES analysis from Labanese subject with intellectual disability identified the association of pathogenic variations in COQ8A and MED25 [77]. Further, the protein-protein interaction analysis (STRING [58]) of genes of intellectual disability candidate genes of Arab population from various studies have revealed
(Figure 5C) the most significant (p value<0.05) biological processes, neuron differentiation (C12orf57, SPTBN4, TRAK1, NFASC, NCKAP1, PLXNA2, NKX6-2, PLXNA3, SLC4A10, and TNIK), neuron development (C12orf57, SPTBN4, TRAK1, NFASC, NCKAP1, PLXNA2, PLXNA3, SLC4A10, and TNIK), neuron projection morphogenesis (C12orf57, SPTBN4, TRAK1, NFASC, NCKAP1, PLXNA2, PLXNA3, and TNIK), central nervous system neuron differentiation (C12orf57, SPTBN4, NKX6-2, PLXNA3, and SLC4A10), central nervous system neuron development (C12orf57, SPTBN4, PLXNA3, and SLC4A10) and generation of neuron (C12orf57, SPTBN4, TRAK1, NFASC, NCKAP1, ASTN2, PLXNA2, NKX6-2, PLXNA3, SLC4A10, and TNIK).

Figure 5: A: Disease markers of intellectual disability in C12orf57 gene. Variants and their percentage of contribution on the presence of intellectual disability in Saudi population. B: Protein-protein interaction analysis of C12orf57 gene with literature using STRING[58]. C: Protein-protein interaction analysis of genes of intellectual disability candidate genes of Arab population.

Microphthalmia
Genetic investigation using multi-gene panel and WES on 93 families with, developmental eye defect (microphthalmia) revealed 55 point variations in 24 disease genes, including 15 novel variations [66]. In addition, the study discovered interesting candidate variations in two genes such as MYO10 and ZNF219 that have not been related to human disorders, which were notable novel possibilities for microphthalmia in the study, most of them were Saudis, there were 14 Egyptian and Lebanese [66]. SLC18A2, PAX6, DSC3 and CNKSR1 genes are very rarely linked in the families with microphthalmia.

Metabolic traits
Although concerted efforts at a national level to promote awareness about the dangers of a sedentary lifestyle and consuming fast food, the people in the Arab regions continue to be at risk for metabolic diseases. There are no well-known genetic risk factors for metabolic diseases in the Arab population like European population, and the existing genetic risk loci for metabolic traits to the Arab population were not shown adequately. A recent GWAS study on Arab (Kuwaiti) population to identify genetic risk factors for quantitative characteristics including anthropometry, insulin resistance, lipid profile, and levels of blood pressure failed to discover associated recessive variations or unable to replicate the identified loci [78]. This is mainly due to the complicated gene–environment interactions in Arab
ancestries. However, another study on the same population identified the TNKS haplotype associated with hypertension [36].

**Other Genetic diseases**

Variants in the Mendeliome on Saudi ancestry: APC-related Cenani–Lenz syndrome, Steel syndrome, syndromic cataract, oral–facial–digital syndrome, CHARGE-like presentation, epileptic encephalopathy, Ehlers–Danlos-like syndrome, and congenital hydrocephalus were reported either with compatible phenotypes (homozygous variant in 30 genes) or phenotypes different from the original reports (homozygous mutations in 18 candidate genes) [79]. Studies on systemic juvenile idiopathic arthritis (LACCI gene) [80,81]; recurrent pregnancy loss (ASIC5 gene) [82]; tricho-hepato-enteric syndrome (SKIV2L and TTC37 genes) [83]; STING-associated vasculopathy of infantile-onset (STING1 Gene) [84]; multiple congenital anomaly syndrome (SMG9 gene) [85]; diabetic retinopathy (NME3, LOC728699, and FASTK genes) [86]; congenital neutropenia with inflammatory bowel disease (G6PC3 gene) [87]; skeletal dysplasia (XYLT1) [88,89]; Wolf–Hirschhorn syndrome (WHSC1 gene) [90]; lymphatic dysplasia with nonimmune hydrops fetalis (PIEZO1) [89]; Cohen syndrome (VPS13B gene) [91]; severe combined immunodeficiency disease (AK2, JAK3, and MTHFD1 genes) [92]; celiac disease (CPED1 gene) [93]; hereditary spherocytosis type 3 (SPTA1 gene) [89]; developmental delay, cerebellar hypoplasia, and myoclonic seizures (KCNMA1 gene) [74]; Cenani-Lenz syndrome (APC gene) [74]; Sjogren-Larsson syndrome (ELOVL4 gene) [74]; autism spectrum disorder (multigene) [94]; congenital heart disease (PRKDI gene) [95]; ciliopathies [96]; Parkinsonism (PLA2G6 gene) [97]; retinal dystrophies (CLRN1, ABCA4, CERKL, AGBL5, CDH16, and DNAJC17 genes) [98–100]; Pediatric Asthma [101]; cardiovascular genetic diseases (LDLR gene) [102]; enteroendocrine dysfunction (PCSK1) [103]; tricho-hepato-enteric syndrome (TTC37 and SKIV2L) [104]; Wolcott–Rallison syndrome (EIF2AK3) [105], Fanconi–Bickel syndrome (SLC2A2) [105] and Alström syndrome (Exon 19 Skipping in ALMS1 Gene) [106,107]; using whole exome or whole genome analysis revealed disease markers in Saudi population. Families with autosomal recessive retinal dystrophies from various ethnicity including Saudis were analysed for candidate genes using WES, 45 unique deleterious variants including 18 novel variants observed [108]. UGCG (UDP-glucose ceramide glucosyltransferase) gene related ichthyosis was reported from a single Saudi family, caused by NM_003358:exon2:c.142dupA variations in the UGCG gene [109]. Epileptic encephalopathy in a consanguineous family was exome analysed and identified pathogenic variation in FRRSIL [c.961C>T: p.(Gln321*)] gene [110].
**NOTCH4** gene with truncating and a splice variant was observed in 2 Saudi patients with Parkinson’s disease [111]. Furthermore, the study reported 18 genes (121 mutations) on Parkinson’s disease in 60 Saudis, most of the variants are missense (n=90) and nonsense (n=11) [111]. Gene specific pathogenic variants analysis from a Arab population of Qatars (n= 6045 whole-genome sequencing) revealed the highest pathogenic and likely pathogenic variants in *SCN5A* (n=37 variants) gene followed by *ATP7B* (n=26 variants), *LDLR* (n=22 variants) and *RYR1* (n=20 variants) (Figure 6) [112]. The study revealed the associated genes on Li-Fraumeni syndrome (*TP53* gene); Peutz-Jeghers syndrome (*STK11* gene); Lynch syndrome (*MSH6, MLH1, PMS2, and MSH2* genes); Familial adenomatous polyposis (*APC* gene); Juvenile polyposis (*SMAD4* and *BMPRIA* genes); Von Hippel–Lindau syndrome (*VHL* gene); Multiple endocrine neoplasia type 1 (*MEN1* gene); Familial medullary thyroid cancer (*RET* gene); Multiple endocrine neoplasia type 2 (*RET* gene); PTEN hamartoma tumor syndrome (*PTEN* gene); WT1-related Wilms tumor (*WT1* gene); Retinoblastoma (*RB1* gene); Hereditary paraganglioma pheochromocytoma syndrome (*SDHAF2, SDHD, SDHB* and *SDHC* genes); Neurofibromatosis type 2 (*NF2* gene); Loey-Dietz syndromes, Marfan syndrome, and familial thoracic aortic aneurysms and dissections (*TGFBR2, FBN1, ACTA2, TGFBR1, MYH11* and *SMAD3* genes); Tuberous sclerosis complex (*TSC1* and *TSC2* genes); Ehlers-Danlos syndrome (*COL3A1* gene); Wilson disease (*ATP7B* gene); Dilated cardiomyopathy, Hypertrophic cardiomyopathy (*MYH7, MYBPC3, TPM1, MYL3, TNNT2, TNNI3, LMNA, ACTC1, MYL2, PRKAG2* and *GLA* genes) Arrhythmogenic right ventricular cardiomyopathy (*DSC2, TMEM43, PKP2, DSP*, and *DSG2* genes); atecholaminergic polymorphic ventricular tachycardia (*RYR2* gene); Long QT and Brugada Syndrome (*SCN5A, KCNH2, and KCNQ1* genes); Malignant hyperthermia susceptibility (*RYR1* and *CACNA1S* genes); Familial hypercholesterolemia (*PCSK9, APOB, and LDLR* genes); and Ornithine transcarbamylase deficiency (*OTC* gene) diseases in Arab population (n= 6045 Qataris whole-genome sequencing) [112]. Low bone mineral density associated variants were found in non-coding RNA *MALAT1* (antisense *TALAM1*), *FASLG, SAG, LSAMP* and *FAM189A2* from Qatari based Arab population with 3000 WGS [113] and *IL1RL1* gene variant with familial Mediterranean fever [114].

**Figure 6:** Number of disease-causing, pathogenic and likely pathogenic variants per gene through 6045 WGS of Qatars from QGP.
Studies (WES) in Egyptian individuals revealed associated genes of primary hyperoxaluria type I (AGXT gene), infantile hypercalcemia/hypophosphatemia/nephrolithiasis (SLC34A1 gene) [115], severe combined immunodeficiency (JAK3 gene) [116], propionic acidemia (PCCA gene) [117], sulfite oxidase deficiency (SUOX gene), molybdenum cofactor deficiency (MOCS2 gene) [118], primary hereditary microcephaly (ASPM gene) [119], familial Mediterranean fever (MEFV gene) [120], thiamine-responsive megaloblastic anemia (SLC19A2 gene) [121], cerebellar atrophy and developmental delay (PLA2G6, KIF1A and MOCS2A genes) [122], micro-/anophthalmia (VSX2, SOX2, and FOXE3 genes) [123], RAG-deficiency (RAG1 and RAG2 genes) [124], auriculocondylar syndrome (PLCB4, GNAI3, and EDNI genes) [125], ectodermal dysplasia (EDA, EDAR, and EDARADD genes) [126], 3-phosphoglycerate dehydrogenase deficiency (PHGDH gene) [127], carpenter syndrome (RAB23 gene) [128], disorders/differences of sex development (NR5A1, CYP19A1, AMH, AMHR2, WT1, HHAT, and FANCA and in the X-linked genes KDM6A and ARX genes) [129] and autosomal recessive polycystic kidney disease (PKHD1 gene) [130]. WES studies in Iraqi people revealed associated genes of epileptic encephalopathy (SLC13A5 gene) [131], juvenile neuronal ceroid lipofuscinoses (CLN3 gene) [132], inherited thrombocytopenias (FYB gene) [133], microcephaly (TMX2 gene) [134], non-syndromic retinal dystrophy (POCI1B gene) [135], progressive pseudorheumatoid dysplasia (WISP3 gene) [136], dedicator of cytokinesis 8 deficiency (DOCK8 gene) [137], and developmental delay (MED27 gene) [138]. Jordanian patient with hypotrichosis-lymphedema-telangiectasia syndrome was analysed using WES for causative genetic factors, the study uncovered SOX18 mutation associated with the syndrome [139]. WES analysis of four consanguineous Jordanian families detected causative genetic variant responsible for unknown gastrointestinal related diseases [140]. Across the five consanguineous families, four variations in the RP1 and RLBP1 genes were identified as disease-causing variant in the patients with retinitis pigmentosa [141]. Two novel pathogenic variants in DYSF gene were identified in the patients with muscular dystrophies from consanguineous Jordanian families [142]. Eight Jordanian consanguineous families with multiple keratoconic individuals analysed using WES, that revealed two variants in the genes MYOF and STX2, and one variant in the genes: COL6A5, ZNF676 and ZNF765 [143]. WES analysis of Lebanese subjects with various genetic disorders (renal dysfunction, PKHD1, GRHPR and NPHS2 genes) revealed associated genes with the overall success rate of 56% [55]. Other studies on Lebanese with cytochrome c oxidase deficiency (PET100 gene) [144], discoid lupus erythematosus (TRAF3IP2 gene) [145], Bardet-Biedl and usher syndromes (BBS9, ARL6, BBS12 and BBS5 genes) [146]and Basel-Vanagaite-
Smirin-Yosef syndrome (MED25 gene) [147] revealed the causative genes. Genetic polymorphisms in Arab individuals with obesity were discussed earlier [148–150].

Genetic analysis on Saudi patients (n=98) with cholestatic liver disease observed 37 variations including 15 novel mutations in ABCB11, ABCB4, ABCC2, AGL, ATP8B1, CYP7B1, FAH, TJP2, and VIPAS39 [67]. ABCB11 gene was reported with the most common candidate disease marker gene among 25% of cases studied [67]. Saudi patients (n=261) with primary immunodeficiency NGS sequencing platform through targeted gene panel (162 PID genes) revealed 89 mutations (45 mutations and 44 novels) [151]. Sickle cell disease is the 5th prevalent genetic disorder on carrier frequency in the Saudi population [7]. The high fetal haemoglobin (HbF) in sickle cell patients is common among Saudis, whole genome studies may reveal candidate HbF-modulating genes and ‘missing heritability’ in the Arab ancestries. WES based Trio (n=16) analysis of attention deficit hyperactivity disorder (ADHD) reported 32 rare variants in 31 candidate genes; exclusively 5 genes (PSRC1, PTP4A3, NEK4, NLE1, and TMEM183A) first observed among the Saudi ancestry [152]. Genomic studies on congenital microcephaly proposed BPTF, CCNH, MAP1B, and PPFIBP1 genes as candidate genes and confirmed ANKLE2, THG1L, YARS, and FRMD4A, genes as congenital microcephaly candidate genes [153]. Targeted sequencing on Saudi patients (14 families) with Glanzmann thrombasthenia discovered 17 variants in ITGA2B and ITGB3 genes including 6 novel mutations [154]. The largest Arab study on 80 patients from 72 families (mostly consanguineous) with peroxisomal disorders (molecularly characterized) identified disease-causing variants (n=43; 50% novel) and estimated the disease burden (~1:30 000) [155]. Exome sequencing and immunophenotyping of CARMIL2-Deficient T cells confirmed the candidature of novel 13bp frameshift deletion (c.2536_2548del) and substitution p.R50T in CARMIL2 gene for combined immunodeficiencies [156].

**Mitogenome**

Sequencing and analysis of the entire mitochondrial DNA (mtDNA) genomes of Emirate females (n=232) and comparison with African ancestries was carried out to characterize the genetic landscape of the Arab population from United Arab Emirates [157]. Mitochondrial Genome of Arab (Emirati) population from the United Arab Emirates revealed the prevalence of haplogroups, heteroplasmy, genetic variation, and demography. The study revealed 968 variations in mtDNA genome in 15 haplogroups. Another study on the mitogenome of
Kuwaiti population (288 whole-exome) identified 1,241 mtDNA SNPs [158]. Qatari (n=864) population based mitogenome analysis 1831 mtDNA variants including 56 indels and 1775 SNPs [159]. The report from Emirati population showed that the Arab population received sufficient gene flow from African ancestries. Furthermore, the study showed that the Arab population received sufficient gene flow from Africa ancestries and a demographic bottleneck was observed that occurs around the time of Western European contact [157]. Since the earliest modern human emerged from Africa, multiple maternal lineage migratory events favoring African–Asian corridors have occurred, according to the research based on UAE mitogenomes [157]. Ancestral sharing with Africa, the Near East and East Asia was also revealed from the study [157]. Complete mitochondrial haplogroups analysis in Arab populations from Qatar and Kuwait based investigations revealed the relationship of mtDNA variations with obesity [158,159]. The missense variant identified in MT-ND2 gene, MT:5460G > A in Kuwaiti population is the most significant mtDNA variation associated with obesity [158].

**Summary and Future Directions**

Even though, the KGP, SHGP, QGP, BGP and EGP are revisiting the genetics and genomics of Arab populations’ ancestries lack of complete coordination between the initiatives is a major limitation on revealing the real disease makers of the Arab population [6]. Recent genomic studies identified more than 3000 new nucleotide variations linked to more than 1200 uncommon genetic diseases and numerous pathogenic variants HGMD and ClinVar Database have been reclassified as benign. However, more root level studies are needed on validating these changes from pathogenic to benign using Sanger sequencing and expression levels in the Arab population. More attentions are needed on intellectual impairment as they are the highest among Saudi Arabians from consanguineous families. Large-scale genomic research have opened up a fresh perspective on Mendelian genes and disorders. Arab research institutions must make a concerted effort to develop better strategies for doing more large-scale genotype-phenotype association studies in hereditary disorders prevalent in the region. Whole genome sequencing in the large-scale control subjects including patients with inherited diseases can discover the molecular basis in the unsolved cases in the Arabian studies. Transcriptome studies in Saudi population shall reveal the functional and molecular impact on the disease development by the candidate genes and pathogenic variations. Novel variants observed in inherited diseases will be the best resource for the developing disease
maker panel for the Arab ancestry. Complicated gene–environment interactions in Arab ancestries with various diseases will open real networking factors on complex diseases prevalent in the Arab population. National coordination between regional research centres in the country and in Arab region can reveal in depth gene networks on diseases and associated pathways.

**Declaration of conflicting interests**

The author declares that he has no competing interests.

**Acknowledgments**

The author thank the Dean, Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia, for her continuous support and encouragement. Author thank Dr Sayed AbdulAzeez for critical comments. Author thank The Deanship of Scientific Research, Imam Abdulrahman Bin Faisal University (Grant No: 2017-100-IRMC). The funders had no role in the study design, and decision to publish, or preparation of the manuscript.
References


**Legend**

**Figure 1:** Reclassified variants in HGMD (Human Gene Mutation Database) and ClinVar Database through SHGP database observations.

**Figure 2:** Combined and observed carrier frequency of genetic disorders in Arab ancestries with an example from Saudi Arabian population (n=7,101).

**Figure 3:** Neurogenetic disorders and candidate genes from WES analysis of pre-screened multiplex consanguineous families of Arab origin. Protein network indicates the protein-protein interaction analysis of genes of neurogenetic disorders candidate genes of Arab population.

**Figure 4:** Novel variants observed in TTR gene from Saudi population with Familial transthyretin (TTR) amyloidosis (ATTR) or carrier. Modified from Abouelhoda et al. [64]. PDB ID: 2rox used for thyroxin-bound TTR and the PDB ID: 2wqa used for RBP-bound TTR.

**Figure 5:** A: Disease markers of intellectual disability in C12orf57 gene. Variants and their percentage of contribution on the presence of intellectual disability in Saudi population. B: Protein-protein interaction analysis of C12orf57 gene with literature using STRING[58]. C: Protein-protein interaction analysis of genes of intellectual disability candidate genes of Arab population.

**Figure 6:** Number of disease-causing, pathogenic & likely pathogenic variants per gene through 6045 WGS of Qataris from QGP.

**Table 1:** Novel variants observed and its significance in Saudi population.
Table 1: Novel variants observed and its significance in Saudi population.

<table>
<thead>
<tr>
<th>Study description</th>
<th>Previously reported pathogenic variants</th>
<th>Novel variants</th>
<th>Prominent Genes with Novel variants</th>
<th>Novel variants and notable phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood and bleeding disorder</td>
<td>98 (stop loss =1, synonymous =2, splice and UTR =11, truncating =12, variants of unknown significance = 19, benign or likely benign = 27, re-classified pathogenic or likely pathogenic = 52, CADD score &gt;20 = 56, nonsynonymous = 72)</td>
<td>140 (Nonsynonymous = 117, CADD score &gt;20 = 71, pathogenic = 32, splice = 6, Truncating = 4, frameshifts = 13, likely benign = 3, variants of unknown significance = 105)</td>
<td>VWF, F8, F5, G6PD, F2, F7, F10, F13B, FGA, HBA2</td>
<td>3 truncating variants observed in 2 cases of aplastic anemia, and one patients with excessive bleeding</td>
<td>[47]</td>
</tr>
<tr>
<td>Familial transthyretin (TTR) amyloidosis (ATTR)</td>
<td>158 TTR variants. (coding or flanking regions = 28, missense TTR variants = 12, uncertain significance = 3, negatively impact TTR function resulting in amyloidosis = 2; )</td>
<td>3 novel TTR c.404C&gt;T (p.S135F), c.298A&gt;G (p.K100E) and c.428C&gt;T (p.T143I), (Nonsynonymous = 3, interface in the middle of the TTR tetramer and thyroxine =1; interface in the middle of two TTR molecules and RBP4 = 1)</td>
<td>TTR</td>
<td>only 1 allele of each novel TTR variant being observed</td>
<td>[48]</td>
</tr>
<tr>
<td>Mutation Details</td>
<td>Mutations Characterized</td>
<td>Novel Mutations</td>
<td>Literature Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>----------------</td>
<td>----------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microphthalmia</td>
<td>ALDH1A3, C12orf57, GDF3, CRYAA, DSC3, MAB21L2, OTX2, MFRP, MYO10, PAX6, PRSS56, PXDN, RAB3GAP2, SLC18A2, STRA6, ZNF219</td>
<td>Variation in PRSS56 or MFRP cause posterior microphthalmos/ nanophthalmos</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cholestatic liver disease</td>
<td>ABCB11, ABCB4, ABCC2, AGL, ATP8B1, CYP7B1, FAH, TJP2, VIPAS39</td>
<td>15 novel</td>
<td>[51]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ABCC2, c.2273G&gt;T with Dubin–Johnson Syndrome; CYP7B1, c.1057G&gt;T: p.E353X with bile acid synthesis defect</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Reclassified variants in HGMD (Human Gene Mutation Database) and ClinVar Database through SHGP database observations.
Figure 2: Combined and observed carrier frequency of genetic disorders in Arab ancestries with an example from Saudi Arabian population (n=7,101).
Figure 3: Neurogenetic disorders and candidate genes from WES analysis of pre-screened multiplex consanguineous families of Arab origin. Protein network indicates the protein-protein interaction analysis of genes of neurogenetic disorders candidate genes of Arab population.
Figure 4: Novel variants observed in TTR gene from Saudi population with Familial transthyretin (TTR) amyloidosis (ATTR) or carrier. Modified from Abouelhoda et al. [48]. PDB ID: 2rox used for thyroxin-bound TTR and the PDB ID: 2wqa used for RBP-bound TTR.
Figure 5: A: Disease markers of intellectual disability in C12orf57 gene. Variants and their percentage of contribution on the presence of intellectual disability in Saudi population. B: Protein-protein interaction analysis of C12orf57 gene with literature using STRING[42]. C: Protein-protein interaction analysis of genes of intellectual disability candidate genes of Arab population.
Figure 6: Number of disease-causing, pathogenic & likely pathogenic variants per gene through 6045 WGS of Qataris from QGP.