Precision medicine of monogenic disorders: Lessons learned from the Saudi human genome

Malak Abedalthagafi¹,²

¹Saudi Human Genome Project, King Fahad Medical City and King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia, ²Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA

TABLE OF CONTENTS

1. Abstract
2. Introduction
   2.1. National human genome projects
   2.2. SHGP sequencing approaches
   2.3. Intellectual disability
   2.4. Other CNS diseases
   2.5. Autoimmune disease
   2.6. Kidney disease
   2.7. Hereditary cancer syndromes
   2.8. Congenital anomalies
   2.9. Muscle, joint and developmental limb disorders
   2.10. Other diseases
3. Summary
4. Acknowledgements
5. References

1. ABSTRACT

Congenital and genetic disorders cause many diseases in Arab countries due to large family sizes and high levels of inbreeding. Saudi Arabia (SA) has the highest consanguinity rates among Middle Eastern countries (~60% of all marriages) and is burdened by the highest number of genetic diseases. Genetic diseases can be life-threatening, often manifesting early in life. Approximately 8% of births in SA are affected, and more common genetic diseases, such as metabolic disease and cancer, manifest later in life in up to 20% of the population. This represents a massive healthcare burden to SA hospitals. The number of genetic disorders in the human population ranges from 7000 to 8000, over 3000 of which are caused by unknown mutations. In 2013, SA initiated the Saudi Human Genome Program (SHGP), which aims to sequence over 100,000 human genomes, with the goal of identifying strategies to discover, prevent, diagnose and treat genetic disorders through precision therapy. High-technology genomics and informatic-based centers that exploit next-generation sequencing (NGS) have now identified mutations underlying many unexplained diseases.

2. INTRODUCTION

The Saudi Human Genome Program (SHGP) was designed to identify mutations in genes that predispose individuals to disease (1-3). The abundance of genetic diseases in SA combined with large family sizes makes it easier to identify the genes/mutations that underlie a particular disease (Figure 1) (4). With this knowledge, preventative counseling can be planned, or a regimen of rational precision-based therapies can be formulated (5, 6). In addition, the growing catalogue of disease-causing mutations allows premarital DNA tests for prospective partners, reducing the rates of afflicted offspring (Figure 1). In many cases, the disease-causing genes and
Genomic basis of diseases and future targeted therapies in Saudi patients

<table>
<thead>
<tr>
<th>Generation</th>
<th>Genetic relationship</th>
<th>Inbreeding coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Unrelated</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>Siblings</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>First Cousins (28.4%)</td>
<td>$1/12$</td>
</tr>
<tr>
<td>IV</td>
<td>Second Cousins (14.6%)</td>
<td>$1/64$</td>
</tr>
<tr>
<td>V</td>
<td>Third Cousins (15.2%)</td>
<td>$1/256$</td>
</tr>
</tbody>
</table>

Figure 1. Genetic relationships and the rates of consanguinity in SA. Studies suggest that 57.7% of the families screened in SA are consanguineous (9, 55). The most frequent consanguineous marriages are between first cousins (28.4%), followed by distant relatives (15.2%) and second cousins (14.6%). These rates can differ among provinces. The offspring of consanguineous unions will be at an increased risk of genetic disorders (red dots) because of the expression of autosomal recessive gene mutations inherited from a common ancestor. The inbreeding coefficient denotes the probability that an individual is homozygous for an ancestral allele by inheritance.

gene variants identified in the SA population can be used to verify results obtained by research groups outside of SA that draw on conclusions from fewer case studies (7). Thus, despite its national nature, the project has the potential to benefit the global fight against inherited genetic diseases, but further improvements are still required (Figure 2).

The SHGP was launched in December 2013 in the city of Riyadh, at the location of the King Abdulaziz City for Science and Technology (KACST). The SHGP was funded and organized by the KACST, and it involves a national network of seven genome centers that were designed to recruit subjects and to carry out sequencing. The core technology in genome centers is next-generation sequencing (NGS), which has enabled efficient and cost-effective readings of entire human genomes. Over the past decade, NGS has exponentially increased our understanding of genetic Mendelian diseases by permitting the analysis of multiple genomic regions in a single reaction (8). NGS has been utilized with success in the clinic, with a reported diagnostic yield as high as ~25% (9). The use of NGS enables the generation of large volumes of sequencing data in hours, representing a large advance over the human genome project that required years to be completed at a cost of 3 billion US dollars. NGS has proven to be a powerful application for the detection of disease-causing mutations, particularly in patients with monogenic disorders. While NGS has benefited the field of human genetics as a whole, the study of rare genetic diseases has perhaps witnessed the greatest advances (8). This permitted the genetic diagnosis of disorders for which the molecular mechanisms leading to disease
were unknown. As an example, the focus on mitochondrial proteins helped identify ACAD9 in cases of complex I deficiency (10). The datasets produced from NGS sequence reads are extensive; as such, advanced computing infrastructure to process

Figure 2. Human genomics in global health. (A) Red dots indicate the Arab countries with the highest rates of inbred marriages and thus genetic diseases. (B) Colored dots highlight several global projects in addition to the Saudi Human Genome Project now underway to sequence inherited disease. These projects aim to share information and to develop innovative approaches in the field of human genetics and genome-based therapy.
the genomic data has been established. A centralized knowledge base at the KACST was also developed to store information on population variations, including those causing disease, and to make this information available to clinicians in SA to enable future diagnostic and screening efforts (Figure 3) (7). This success of the SHGP requires the ability to collect an array of subjects that fully represent normal and disease states in the Saudi population. The genomes must be sequenced, the data correctly interpreted and validated, and the results stored in an accessible knowledge-base for future use.

Figure 3. Workflow highlighting how data from the SHGP can benefit disease diagnosis and revolutionize gene- and pathway-targeted therapeutics. The SHGP will allow more rational, efficient, and effective drug development pipelines centered on disease-specific genetic abnormalities. This progress provides opportunities for identifying new and effective treatment approaches and rapid disease diagnoses in SA and other Arab countries.
Over the last five years, the project has helped identify the genetic basis of different inherited diseases. To date, over 99 publications charter its success. The project has advanced treatment plans using stem cell and gene editing, in which a defective gene is manipulated by silencing or reintroduced using gene therapy. The KACST plans to support initiatives from genetic data to develop further technologies and treatment regimens for afflicted SA patients. To date, the SHGP is the largest study on the mutational spectrum of genetic diseases in the SA and Arab populations. The non-selective nature of tested families and its representation of all regions of SA allows inference of important patterns of national genetic diseases (11). The highly consanguineous population of SA differs from outbred populations (such as those in Western countries) and thus impacts the landscape of disease-causing mutations (12). The consanguineous nature biases the occurrence of homozygous recessive mutations in diseases that were characterized as dominant, particularly for mutations leading to intellectual disabilities (13). These characteristics increase the sensitivity of NGS assays, allowing the genetic basis of many diseases to be defined.

Using new genomic data, the SHGP has identified a number of diseases that are common or endemic in SA and have a strong genetic component, including intellectual disorders, CNS disorders, metabolic diseases, developmental diseases, hereditary cancers, autoimmune diseases and cardiovascular diseases (11). Some common limitations are shared by the studies published from the SHGP, including a lack of functional follow-up studies, the absence of a replication cohort for gene discovery studies and the differing analysis pipelines. Herein, we will discuss the new genetic knowledge of these diseases and the development of personalized medicine that may arise from this new knowledge.

2.1. National human genome projects

The SHGP is one of an array of national genome projects that have aimed to describe the genetic background of specific populations (Figure 2). The projects share common aims, namely, (1) the development of infrastructure to create a network of state-of-the-art human genome sequencing centers; (2) training to create the skills, capabilities and capacity for human genome sequencing; and (3) the ability to solve disease genetics and identify traits or disorders within a specific population that can facilitate the genetic distribution of disease within a country. The projects also share common drawbacks, including the incorrect interpretation of a pathogenic variant, resulting from an incomplete human reference genome; an incomplete exon being represented in public databases; and an incorrect annotation of transcripts/exons owing to differing tissue-specific expression patterns (14). The use of genetics to guide therapy is also not without critique. Genomic medicine, through the ability to control the basic genes of a human, may lead to a loss of human diversity, the development of “designer” humans, and a genetic racism for inherited diseases.

An array of national genome projects based on NGS approaches is well established, including projects in Iceland (deCODE), the United Kingdom, the Netherlands, Finland and Sweden (reviewed in (15)). The scale of interest in national genomic screening is highlighted by the emergence of several new sequencing programs, including the initiative on rare and undiagnosed diseases in Japan that was launched in 2015 through collaborations with more than 400 hospitals, including 34 clinical centers; the 100,000 genomes project in China that launched a five-year project in December 2017 and will sequence the genomes of 100,000 individuals from different Chinese ethnic backgrounds; the genome program in Qatar that aims to establish a Qatari reference genome map through the sequencing of 3,000 genomes (~1% of the Qatari population); the genomic health futures mission in Australia that launched in May 2018; the “all of us” research program in the United States that aims to compile data from ≥ 1 million individuals beginning with its inception in May 2018; the
personalized medicine program in Estonia that launched in 2016 from an initial cohort of 51,535 gene donors; the genomic program (génomique) in France that was launched in 2016 that will encompass a network of 12 genomic sequencing centers; the Dubai genomics program in the UAE that is currently focusing on building genomic medicine infrastructure and initiating large-scale whole-genome sequencing; and the Turkish genome project in Turkey that plans to sequence 1 million genomes by 2023. When compared to the scale of these studies, the SA population of ~32 million may appear low, but the SHGP has developed ≥13 gene panels covering ≥ 5,000 inherited diseases, identifying ≥ 2,000 variants underlying disease (1-3). Over 500 of these mutations are common and have been represented in multiple patients. An underrepresentation of specific ethnic groups in open access international DNA databases occurs, which highlights the importance of having DNA databases from ethnically matching healthy individuals. This is a feature that nation-specific human genome sequencing efforts such as the SHGP have achieved. However, nation-specific studies such as the SHGP are independent and do not form part of a global consortium or open access multipopulation DNA database, limiting the availability and global applications of the study findings. This differs from projects such as the 1000 Genomes Project that allows access to sequencing data from ≥ 26 countries to be globally available.

2.2. SHGP sequencing approaches

The initial aim of the SHGP was to read the genomes of affected individuals to identify the casual disease genes and their mutations. If tens of thousands of datasets are correctly analyzed, the major disease risk factors in the Saudi population can be identified (1-5). Entire genomes are sequenced when genes that cause human disease are to be identified in a Mendelian fashion (9). Sequencing of the coding regions using whole-exome sequencing (WES) has been successful and has clinical utility, particularly in the case of neurological disorders. However, WES typically uncovers a number of variants from which identifying causal variants can be challenging. Several tools have been developed to prioritize WES variants, but these require expert analysis. A leading concern of WES in clinical genomics is secondary findings, defined as variants that are potentially medical relevant but are missed as they are unrelated to the medical rationale for the test. This inability of WES to identify incidental/secondary findings is a commonly encountered problem (9, 16).

An alternative to WES is the use of gene panels, which are when a range of genes relevant to a particular phenotype are sequenced in patients presenting that phenotype. This approach provides a focused analysis of clinically relevant genes minimizing many WES challenges. Difficulties in this process arise from the design of an appropriate gene list for a given phenotype, as many diseases have variable clinical presentation. Indeed, gene panels developed across laboratories often differ for the same disease. This can be overcome by relaxing the clinical indications and the integration of supporting evidence to classify diseases into five categories as follows: (1) pathogenic, (2) likely to be pathogenic, (3) uncertain significance of variant, (4) likely benign, or (5) benign. These guidelines have improved disease classification; however, in practice, they remain challenging to standardize. In addition, a shortcoming of gene panels is their need to be updated with the discovery of new gene-disease associations. This is challenging as further disease mutants accumulate from large international datasets.

The Saudi Mendeliome Group defined 13 broad clinical themes in which ~3000 Mendelian genes were distributed (17). The power of this approach was highlighted by the success of the gene panel on the diagnosis of 2300 patients presenting with a range of medical and surgical diseases. Al-Mousa and coworkers also displayed the benefits of targeted NGS panels for primary immunodeficiency diseases, which were more sensitive and cost-effective
than WES approaches, particularly in patients with an atypical presentation of known primary immunodeficiency genes (18). Mustafa and coworkers advanced these studies, demonstrating the suitability of a targeted AmpliSeq Inherited Disease Panel (IDP) (consisting of 328 genes underlying more than 700 inherited diseases) for the first-line screening of genetic diseases following clinical validation (19). It is therefore clear that the sequencing approach is dependent on the type of disorder and degree of genetic or phenotypic heterogeneity. WES has utility for gene discovery studies when exploring the genetic landscape of the population and in diseases with genetic or phenotypic heterogeneity. For those diseases in which little genetic information is available, gene panels have utility, particularly for clinical conditions with distinct and homogenous phenotypes, for which there is sufficient knowledge of their genetic basis.

Despite the obvious progress, knowledge gaps and areas of concern remain. A leading challenge remains that NGS technology is underutilized. Clinical genomics are primarily employed to detect SNVs and indels, but its application in the screening of CNVs, RNA or methylation status remains limited (16). NGS also fails to detect repeat expansion mutations or to quantify alleles. Experts have also expressed concerns over the small numbers of clinical geneticists and/or genetic counselors, as well as the lack of genomic knowledge among active clinicians. This results in a loss of confidence in the application of NGS data to clinical decision making.

2.3. Intellectual disability

Intellectual disability (ID) is a common developmental disorder characterized by congenital limitations in intellectual function. Advances in high-throughput whole-genome sequencing (WGS) and single-cell sequencing have increased the number of causative genes identified for this human disease. As the clinical features of ID display heterogeneity regarding their genetic causative factors, the characterization of ID has benefited from these advances. In SA, the most common cause of ID is genetic and due to the marriage of relatives (1). The SHGP has expanded the locus and allelic heterogeneity of ID in recent years, and the culmination of these studies demonstrates the power of positional mapping to reveal new and unusual mutational mechanisms (20-25). Anazi et al. (21) described the phenotypic and genetic findings of 105 patients originating from 68 families with ID, revealing several new disease variants. These genes included TRAK1, GTF3C3, SPTBN4 and NKX6-2 as well as novel variants in 14 other genes, including ANKH4D1, ASTN2, ATP13A1, FMO4, MADD, MFSDF11, NCKAP1, NFASC, PCDHGA10, PPP1R21, SLC12A2, SLK, STK32C and ZFAT (21). In particular, MADD and PCDHGA10 were compelling candidates as biallelic deleterious variants in two independent ID families were discovered (21). In other studies, Anazi et al. (20) prospectively used molecular karyotyping, multigene panels and whole-exome sequencing (WES) in a cohort of over 300 ID subjects and compared the data with the clinical evaluations of the patients. WES identified independent mutations in three new candidate ID genes, DENND5A, NEMF and DNHD1, and all patients harboring these mutations displayed comparable phenotypes. In addition, de novo and recessive variants in 32 other genes were identified, namely, MAMDC2, TUBAL3, CPNE6, KLHL24, USP2, PIP5K1A, UBE4A, TP53TG5, ATOH1, C16ORF90, SLC39A14, TRERF1, RGL1, CDH11, SYDE2, HIRA, FEZF2, PROC1, PIANP, PLK2, QRFPR, AP3B2, NUDT2, UFC1, BTN3A2, TADA1, ARFGEF3, FAM160B1, ZMYM5, SLC45A1, ARHGAP33 and CAPS2 (20). Causal variants from previously published ID genes, including ASTN1, HELZ, THOC6, WDR45B, ADRA2B and CLIP1, were also confirmed, strengthening the accuracy of the study. The very recent identification of WDR45B as an ID gene in two large cohorts of affected individuals was revealed in independent studies (25). Additionally, KIF14 dysfunction was revealed as a cause of autosomal recessive primary microcephaly.
Genomic basis of diseases and future targeted therapies in Saudi patients

... (an extremely rare condition characterized by a reduced cerebral cortex accompanied by ID) (26). KIF14 is a mitotic motor protein that is required for spindle localization of the mitotic citron rho-interacting kinase, CIT, highlighting a cellular mechanism of disease progression.

Genetic sequencing has also expanded our knowledge of ID formation. Patel and colleagues demonstrated variable expressivity in three consanguineous families linked to novel 3’ UTR mutations in SLC4A4, a gene known to be mutated in a syndromic form of ID (27). The 3’ UTR motif mediates posttranscriptional control of many genes, and a marked reduction in the transcript level of SLC4A4 was observed in cells from ID patients. This novel mutational mechanism expanded our knowledge of the variables that underlie phenotypic expressivity in human disease (27).

2.4. Other CNS diseases

For many years, knowledge of the mutations in genes that lead to neurological disorders in SA has been lacking. The gaps in this knowledge have been filled using WES in multiple consanguineous families presenting diseases associated with impaired brain function. Alazami et al. (28) identified a total of 33 genes (SPDL1, TUBA3E, INO80, NID1, TSEN15, DMBX1, CLHC1, C12orf4, WDR93, ST7, MATN4, SEC24D, PCDHB4, PTPN23, TAF6, TBCK, FAM177A1, KIAA1109, MTSS1L, XIRP1, KCTD3, CHAF1B, ARV1, ISCA2, PTRH2, GEMIN4, MYOCD, PDPR, DPH1, NUP107, TMEM92, EPB41L4A, and FAM120AOS) as mutations that may impair brain activity, but in vitro and in vivo data that support a direct link between the identified mutations and impaired brain activity are required. Sowada et al. (29) described new forms of developmental and epileptic encephalopathies (DEE) with deleterious biallelic variants in PTPN23, a tyrosine kinase strongly expressed in neuronal tissue. The phenotype was characterized by early onset drug-resistant epilepsy, developmental delay, microcephaly, and on occasion premature death. Ramadan et al. (30) identified that recessive mutations in SCN1B can cause severe epilepsy in humans in addition to the known dominant SCN1B mutations, which were previously reported. This highlighted the need to consider recessive mutations in the interpretation of variants in typically dominant genes.

Shamseldin et al. (31) identified other candidate genes involved in a distinct neurodevelopmental phenotype in a multiplex consanguineous family. Mutations in MICU2, a major component of the mitochondrial calcium uniporter complex, lead to impaired mitochondrial Ca

2+
homeostasis and severe cognitive impairment, spasticity, and white matter involvement (31). Al Mutairi and colleagues also identified mutations in the short-chain enoyl-CoA hydratase (SCEH), a mitochondrial enzyme involved in the oxidation of fatty acids and valine catabolism in early childhood Leigh syndrome (32). When these mutations were accompanied by a missense mutation in ECHS1, a lethal phenotype was observed.

The phenotypic and molecular spectrum of Aicardi-Goutières Syndrome (a rare genetic neurological disorder) was recently noted by Al-Mutairi and coworkers to be caused by homozygous mutations in RNASEH2B, RNASEH2A RNASEH2C, SAMHD1, and TREX1 as well as heterozygous mutations in IFIH1 in Arab pediatric patients (13). Additionally, in cases of spastic ataxia and hypomyelination, homozygosity mapping and WES revealed missense mutations in NKX6-2, which are known to encode a transcriptional repressor with high CNS expression (33). Accordingly, when NKX6-2 was silenced in mice, hypomyelination was observed (15, 33).

2.5. Autoimmune disease

Studies assessing systemic lupus erythematosus (SLE) in Arab countries suggest its high prevalence in the Arab world (34-36). Using NGS, Carbonella et al. (37) identified a homozygous 2 bp deletion in DNASE1L3 that leads to autosomal recessive autoimmune disease (AID), which mimics SLE. The same
mutations were reported in three siblings from consanguineous parents who presented with hypocomplementemic urticarial vasculitis syndrome (HUVS). As approximately 50% of individuals with HUVS develop SLE, whether SLE is a subphenotype or a separate condition was not clear. Studies to identify the cause of an autosomal-recessive form of systemic juvenile idiopathic arthritis (JIA), which displays both autoimmune- and auto-inflammatory-etiologies, in SA patients identified a homoallelic missense mutation in \textit{LACC1} (38). This gene encodes the enzyme laccase, which is a multicopper oxidoreductase. Missense mutations were identified using WES and confirmed by Sanger sequencing. The mutations were present in all investigated cases of JIA disease based on an autosomal-recessive pattern of inheritance. Given the known association of LACC1 with Crohn’s disease and leprosy, this further highlighted this enzyme as a drug target in genetic autoinflammatory disorders (38). This originates from data demonstrating that leprosy hijacks the immune system, manipulating defense mechanisms and causing them to attack neurons.

2.6. Kidney disease

Inherited cystic kidney disorders commonly cause end-stage renal disease. To define the phenotype and genotype of cystic kidney disease in fetuses and neonates, Al-Hamed \textit{et al.} (39) correlated antenatal- and postnatal-renal-ultrasound examination with targeted exon sequencing using a defined renal gene panel. The disease phenotypes observed were severe, with 36 cases of stillbirth or perinatal death. Renal gene panel testing identified causative mutations in up to 60% of the families, and mutations were found in 12 genes, including an inferred novel variant in \textit{NEK8} (39). Mutations in \textit{CC2D2A} were the most common cause of antenatal cystic kidney disease and suspected ciliopathy. The renal gene panel from this study holds promise for a rapid molecular diagnosis for this disease in the future.

Meckel-Gruber syndrome (MKS) is characterized by occipital encephalocele, polydactyly and polycystic kidneys and is genetically heterogeneous with mutations in up to twelve genes known to date. Using a gene panel and NGS in 25 MKS families, Shaheen \textit{et al.} (40) identified a homozygous splice variant in \textit{TMEM107}, which encodes for a protein that localizes to the transition zone at the proximal region of the ciliary axoneme. As \textit{TMEM107}-deficient mice are known to display typical ciliopathy phenotypes, \textit{TMEM107} mutations are now accepted as a factor contributing to human MKS and thus represent a novel drug target for this disorder. In a large population of patients with phenotypes that span the entire ciliopathy spectrum, loss of function mutations in \textit{TXNDC15} encoding a thiol isomerase causing MKS were identified (41). Other novel candidate genes for ciliopathies were also discovered in these studies, including \textit{TRAPPC3, EXOC3L2, FAM98C, C17orf61, LRRCC1, NEK4, and CELSR2}.

2.7. Hereditary cancer syndromes

Hereditary cancer syndromes account for approximately 5% of all malignancies. Hereditary cancer syndromes occur due to an inherited mutation increasing the risk of tumor development, which typically occurs at an early age (42). In the majority of hereditary malignant cancers, the elevated risk is due to the mutation of a single gene (monogenic hereditary disease). The ability to identify germline variants in hereditary cancer cases has been a challenge. A history of incomplete cataloguing of cancer-relevant genes combined with a lack of agreement on the patients who should be tested has posed significant barriers. In a recent study from SA, Siraj \textit{et al.} (43) designed a hereditary oncogenesis predisposition evaluation (HOPE) that consisted of many of the genes with known associations with cancer and assessed the effectiveness of HOPE on ≥ 1300 tumor/blood samples from patients with ovarian, breast, colorectal and thyroid cancer. Most notably, pathogenic alleles in DNA repair/genomic instability genes other than \textit{BRCA2}, \textit{ATM} and \textit{PALB2} accounted for at least 11.1, 16.8, 50 and 45.5% of mutation-positive ovarian, breast, thyroid and colorectal cancer.
Genomic basis of diseases and future targeted therapies in Saudi patients

patients (CRC), respectively (43). A family history was absent in many of these mutation-positive cases, suggesting a high contribution of germline mutations to cancer predisposition and extending our knowledge beyond “classic” hereditary cancer genes.

2.8. Congenital anomalies

Embryonic lethality is recognized as a phenotypic result of individual gene mutations. However, identifying embryonically lethal genes in humans is challenging, particularly when the phenotype manifests at the preimplantation stage. To catalogue recessively acting embryonic lethal genes in SA, Alazami et al. (44) identified two families with a female-limited infertility phenotype. Using autozygosity mapping and WES, the phenotype was matched to a single mutation in TLE6, a maternal-effect gene that encodes a member of the subcortical maternal complex (44). Female patients homozygous for TLE6 mutations were found to not undergo early cleavage, leading to sterility. The human mutation abrogates TLE6 phosphorylation, which is reported to be critical for oocyte meiosis (44). This was the first report to identify a human defect in the subcortical maternal complex. Maddirevula et al. (45) performed positional mapping and WES and revealed two homozygous deleterious variants in PATL2, which is also necessary for female meiosis and fertility. PATL2 encodes a highly conserved oocyte-specific mRNP repressor of translation, again identifying a novel disease mechanism in humans. The diagnosis of these mutations in IVF clinics may thus hold great promise for the identification of fetal abnormalities.

Orofacial clefting is among the most prevalent type of birth defect. Mutations in the HYAL2 gene encoding hyaluronidase 2 (which degrades extracellular hyaluronan) were identified as a critical component of the developing heart and palatal shelf matrix (46). In SA studies, Harms et al. (47) identified a homozygous truncating variant in CDH11 associated with Elsahy–Waters syndrome (EWS), also known as branchial–skeletal–genital syndrome. Included in the clinical features of CDH11, mutation-positive individuals had upper eyelid coloboma, which is a new phenotype of this disease that may aid in its diagnosis.

In other studies, Shaheen et al. (48) identified loss-of-function mutations in SMG9 that lead to a loss of nonsense-mediated decay (NMD), which is an important process to degrade transcripts containing premature stop codons that removes their potentially harmful consequences. SMG9 was shown to be required for normal human development, most likely through this transcriptional regulatory role.

Congenital hydrocephalus is an important birth defect that results in children born with an excessive accumulation of cerebrospinal fluid in the brain. WES combined with positional mapping identified causal mutations in 16 genes; interestingly, none of these mutations were X-linked. Ciliopathies and dystroglycanopathies were the most common etiologies of this congenital hydrocephalus cohort. In a single family with four affected members, a homozygous truncating variant in EML1 was identified (49). Recessive mutations in WDR81, previously linked to cerebellar ataxia, mental retardation, and disequilibrium syndrome, were also found. Other previously identified candidates, including MPDZ, were identified, thus highlighting the importance of recessive mutations in this disease condition (49).

2.9. Muscle, joint and developmental limb disorders

Congenital disorders of glycosylation are frequently associated with muscle weakness in the Arab population. Using WES, Monies and coworkers reported clinical and pathological features resulting from a homozygous mutation of ALG2 in an extended family (50). Mutations of ALG2 manifested as limb and muscle wasting, with defects at both the neuromuscular junction and sarcomere evident. The same group developed a first-line diagnostic assay for limb-girdle muscular dystrophy and myopathies
by evaluating a panel of 759 OMIM genes associated with neurological disorders in Saudi patients presenting with muscle weakness (51).

Larsen syndrome (LS) is characterized by the dislocation of large joints with heterozygous FLNB mutations, accounting for the majority of cases. However, biallelic mutations in \textit{CHST3} and \textit{B4GALT7} were recently identified by Patel \textit{et al.} (52), confirming recessive forms of the disease. In a multiplex consanguineous Saudi family affected by severe and recurrent large joint dislocation and severe myopia, a homozygous truncating variant was identified in \textit{GZF1} through a combined autozygome and exome approach (52). The same approach identified a second homozygous truncating \textit{GZF1} variant in a second multiplex consanguineous family affected by severe myopia and milder skeletal involvement (53). \textit{GZF1} encodes the GDNF-inducible zinc finger protein 1, which is a transcription factor of unknown developmental function, and it is expressed in the eyes and limbs of developing mice. These studies reveal new functionality of this protein and its contribution to human disease when dysfunctional.

The spectrum of genetic muscle diseases also extends to diseases of the heart muscle. Dilated cardiomyopathy (DCM) is a common cardiomyopathy that leads to systolic dysfunction and heart failure due to the heart becoming enlarged, thick and rigid. Despite knowledge of over 30 known genes for DCM (primarily in sarcomere and cytoskeletal proteins), Al-Yacoub \textit{et al.} (54) identified mutations in \textit{FBXO32} as a novel DCM-causing locus. The protein encoded by \textit{FBXO32} (also known as atrogin-1) is an E3 ligase that is expressed selectively in skeletal muscle and cardiomyocytes.

2.10. Other diseases

Molecular genetic studies are of increasing importance in the diagnosis and classification of several other diseases. Causative mutations in \textit{SKIV2L} and \textit{TTC37} that cause congenital diarrheal disorders in SA patients were identified by Monies and coworkers (55). Congenital cafe-au-lait spots on the pelvis and lower limbs were a unique and consistent clinical feature of these patients and will aid the differential diagnosis of congenital diarrheal disorders in the future (55).

Ectodermal dysplasia is a highly heterogeneous group of disorders that affect skin, hair, nails and teeth. In an autosomal dominant form of ectodermal dysplasia, Shamseldin \textit{et al.} (53) identified a novel variant in \textit{KDF1} in a Saudi family. The recapitulation of the phenotype was observed in \textit{KDF1} knockout mice, suggesting a causal role played by the \textit{KDF1} variant (56).

Diabetic retinopathy (DR) is a common clinical expression of diabetes mellitus-induced vasculopathy and is a major cause of vision loss. In a cohort of SA diabetic individuals who did not develop DR 10 years after diagnosis compared to SA diabetic individuals with DR, WES identified three genes (\textit{NME3}, \textit{LOC728699}, and \textit{FASTK}) whose rare variant burden protects against DR, highlighting them as attractive candidate drug targets (57). In syndromic RD cases, WES revealed \textit{AGBL5} and \textit{CDH16} as likely gene candidates for the disease (58). A homozygous truncating mutation in \textit{DNAJC17} in a family with retinitis pigmentosa and hypogammaglobulinemia was also identified, thus expanding the allelic spectrum of known RD genes. Patel \textit{et al.} (56), when investigating the genetic basis of pediatric cataracts, applied a multigene panel as well as WES and identified 15 novel genes, including \textit{GEMIN4}, which was mutated in families with cataracts and global developmental delay. Mutations in \textit{RIC1} were identified in patients with cataracts, brain atrophy and microcephaly. Among others, two further candidates that were biallelically inactivated in single cataract families were reported, including \textit{TAF1A} (cataract with global developmental delay) and \textit{WDR87} (nonsyndromic cataract) (56). This expanded the allelic and locus heterogeneity of pediatric cataracts for future diagnosis/treatment.
Advanced cholestatic liver disease is a leading referral to pediatric liver transplant centers in Arab countries. Using an NGS panel of pediatric patients with advanced cholestatic liver disease, causal mutations were identified in the majority of cases and included the novel identification of TJP2 and VIPAS39 (59). Due to the alarmingly high carrier frequency of founder mutations identified in this cohort, primary prevention through carrier screening was suggested as the most effective method of minimizing this disease incidence.

3. SUMMARY

It is clear that the SHGP has made strides to improve our knowledge of genetic diseases. From the findings, the “one-size-fits-all” approach to medicine, which is based on broad population averages, is not the solution for future therapies in SA. Our enhanced knowledge of the genetic basis for disease offers the prospect of precision medicine. Personalized treatments can allow the individualization of therapy tailored to the molecular profile of a patient’s disease. In the case of hereditary cancer, for example, the identification of tumor-specific alterations, such as point mutations and neoantigen production, can determine a patient’s prognosis and the aggressiveness of the tumor and can predict how well the patient will respond to specific chemotherapy treatments or anti-cancer vaccines (60-65). In the diagnosis, management and treatment of genetic errors of metabolism, disorders may be treatable with stem cell transplantation or with enzyme replacement therapy using CRISPR Cas-9, which was designed to compensate for the identified defective gene. Such approaches have proven effective in the successful correction of inherited liver diseases (66).

Armed with disease-related genetic knowledge, it is also important that genetic counseling services and disease prevention programs run in parallel. Genetic counselors can provide advice for a wide range of genetic disorders with a focus on their prevention through intensive family education and preventive reproductive options. Prenatal diagnosis and preimplantation genetic diagnosis are of further benefit to recognize genetically inflicted siblings at an early stage. From this information, it is recommended that SA and other Arab governments establish programs that demonstrate the benefits of marriage outside of families (thus reducing inbred genetic diseases) and offer embryonic counseling services for those already afflicted. The combination of efficient diagnosis and precision-based therapy should reduce the occurrence of these genetic defects while improving patient care in the future.

From a sequencing perspective, the advent of NGS technologies has increased its availability. This is highlighted by platforms such as Illumina that have led to a surge in screening programs. NGS, however, has limitations. NGS relies on short ≤ 1 kb reads; thus, millions of fragments must be scaled in parallel. This is not conducive to de novo assembly as the data require reference genomes. Deep sequencing can compensate for this, but repetitive regions remain challenging. Sample preparation is also time-intensive. This area has been addressed with an array of amplification kits that simplify DNA purification, but further advances in this area are required. In this regard, third-generation sequencers can read kbs of sequence without clonal amplification, subverting the preparation steps required for NGS. This simplifies de novo assembly and the reading of repetitive regions. The drawback is that single-molecule sequencing can lack accuracy, but technology is improving, and error rates are decreasing. Regarding human genome projects, further harnessing of the power of NGS is required. For example, NGS now allows the quantification of RNA modifications at a genome-wide scale, revolutionizing our understanding of diverse RNA modifications and their association with disease. The advances in this field require simultaneous advances in chemical reagents, specific antibodies against RNA modifications, and the application of single-molecule RNA sequencing technologies. It is therefore clear that as NGS develops, a step change in both
ease-of-use and speed is predicted. Future technologies should simplify the workflow of technical preparation steps, making it more available for national genome sequencing programs that mirror the SHGP in new countries and territories. The SHGP should now look to form collaborative efforts with other national sequencing programs to increase data accessibility and our understanding of genetic diseases on a global level.

4. ACKNOWLEDGEMENTS

The author thanks King Abdulaziz City for Science and Technology and the Saudi Human Genome Project for technical support.

5. REFERENCES


Genomic basis of diseases and future targeted therapies in Saudi patients

Ion Torrent (TM) inherited disease panel with the PGM (TM) sequencing platform for rapid and comprehensive mutation detection. Genes 9 (2018). DOI: 10.3390/genes9050267
DOI: 10.3390/genes9050267


*Ann Neurol* 82, 562-577 (2017)  
DOI: 10.1002/ana.25044  
PMid:28892560

*Genome Biol* 18, 144 (2017)  
DOI: 10.1093/gene/awx237  
PMid:29053821

DOI: 10.1016/j.celrep.2014.12.015  
PMid:25558065

*Hum Genet* 136, 1455-1461 (2017)  
DOI: 10.1007/s00439-017-1850-3  
PMid:29090338

*Clin Genet* 92, 327-331 (2017)  
DOI: 10.1111/cge.12999  
PMid:28218389

*Brain* 140, 2806-2813 (2017)  
DOI: 10.1093/brain/awx237  
PMid:29053821

*Clin Genet* 91, 629-633 (2017)  
DOI: 10.1111/cge.12891  
PMid:27905109

*Am J Hum Genet* 100, 969-977 (2017)  
DOI: 10.1016/j.ajhg.2017.05.009  
PMid:28575651  
PMCID:PMC5473715

34. A. M. ALDhanhani, M. Agarwal, Y. S. Othman, O. Bakoush: Incidence and prevalence of systemic lupus erythematosus among the native Arab population in UAE. 
*Lupus* 26, 664-669 (2017)  
DOI: 10.1177/0961203316678677  
PMid:27831539
Genomic basis of diseases and future targeted therapies in Saudi patients


53. H. E. Shamseldin, O. Khalifa, Y. M. Binamer, A. Almutawa, S. T. Arold, H. Zaidan, F. S. Alkuraya: KDF1, encoding keratinocyte differentiation factor 1, is mutated in a multigenerational family with ectodermal dysplasia. *Hum Genet* 136,


**Key Words:** Genetics, Congenital, Consanguinity, Saudi human genome project, Precision medicine, NGS, Review

**Send correspondence to:** Malak Abedalthagafi, King Fahad Medical City and King Abdulaziz City for Science and Technology, P.O. Box 59046, Riyadh 11525, Kingdom of Saudi Arabia, Tel: 966 11 4813202, Fax: 966 1 4813230, E-mail: malthagafi@kacst.edu.sa