UNIVERSITY OF CALIFORNIA,
IRVINE

An Evaluation of Variants Associated with 176 Conditions Identified through Exome Sequencing in an Armenian Sample

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Genetic Counseling

by

Aida Natalie Akopyan

Thesis Committee:
Professor Robert K. Moyzis, PhD., Chair
Adjunct Professor Pamela Flodman, MSc, MS
Assistant Adjunct Professor Elizabeth C. Chao, M.D.

2019

DEDICATION

To

my grandmother, Hilda

in recognition of her endless love and encouragement

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would first like to thank my thesis committee: Professor Robert K. Moyzis, Adjunct Professor Pamela Flodman, and Assistant Adjunct Professor Elizabeth C. Chao. Thank you for your assistance in the development of this project and guidance throughout the entire process. This thesis would not have been possible without your tireless efforts.

I would like to give special thanks to my program director, Pamela Flodman, for her constant encouragement and kindness. Her dedication to her students is truly remarkable, and my experience at UCI would not have been the same without her guidance and support. Thank you so much, Pam.

I would also like to acknowledge Dr. Wayne Grody and Dr. Salpy Akaragian of UCLA for allowing me to be involved in the ArmGenia project and for sharing this data set with me.

I extend my gratitude to Dr. Moyra Smith for providing me with the annotation tool that I used in this study, VarSeq, as well as a quiet place to work in her office.

The last two years of my education would not have been possible without the help of my fellow classmates at UCI. I know I have made lifelong friends at UCI, and I look forward to being your colleague. Thank you.

Finally, I must express my very profound gratitude to my friends, my family and my fiancé for providing me with unfailing support and continuous encouragement throughout my years of study. This accomplishment would not have been possible without them. Thank you.

ABSTRACT OF THE THESIS

An Evaluation of Variants Associated with 176 Conditions Identified through Exome Sequencing in an Armenian Sample

By

Aida Natalie Akopyan

Master of Science in Genetic Counseling

University of California, Irvine, 2019

Professor Robert K. Moyzis, Chair

This study aims to shed light on genetic variation among Armenians, an understudied population. Given the increasing uptake of genetic testing in various settings, the lack of published research poses a significant barrier to providing precise and thorough genetic counseling to Armenian individuals. By analyzing exome sequence from 44 Armenian individuals for 176 genes related to conditions commonly tested for in expanded carrier screening, expanded carrier screening positive rates (ECS-PRs) were calculated for individuals in this study. Fifty-two percent of individuals in this cohort were found to be a carrier of a pathogenic variant for at least one recessive condition. A total of seven out of 44 individuals were found to be carriers for familial Mediterranean fever; however, when FMF was excluded from ECS-PR calculation, the percentage of individuals who were found to be a carrier of a pathogenic variant for at least one recessive condition only decreased by 4%, yielding an ECS-PR of 48%. There were also two individuals with the same *CFTR* pathogenic variant, four individuals with four distinct *PAH* variants, and two individuals with two distinct *PMM2* variants. This study provides data to a recommendation of expanded carrier screening in Armenian individuals as compared to targeted testing for FMF only. Results of this study will improve genetic counseling provided to Armenian individuals, particularly in prenatal and pre-

conceptional settings, and promote diversity and equity in the field of genetics to improve the overall quality of services for providers and patients alike.

I. INTRODUCTION

I.I. *The Importance of Population Genetics and a Human Reference Genome*

The study of human population genetics has been a focus for scientists for many years, well before the advent of complex sequencing technologies or the establishment of a human reference genome. Population genetic studies aim to identify genetic differences within and between different populations, and by doing so, provide insight into evolution, natural selection, genetic drift, and clinical interpretation of genetic variants. Ancient populations that have remained isolated from other populations for many years can be a desirable study cohort because of their higher level of genetic homogeneity, which in turn allows for a smaller effective sample size in characterizing genetic variability [Hatzikotoulas et al., 2014]. In addition to this, the persistence of some variants in an isolated population may demonstrate that a variant has an adaptive advantage, especially since isolated populations tend to be exposed to the same environmental and cultural influences [Gibbs et al., 2015]. Isolated populations also typically show longer regions of non-random association of genetic factors, otherwise known as linkage disequilibrium, which aid in clarifying disease association loci. When two or more genetic factors are inherited together more often than expected, these form haplotype blocks that can be tracked over multiple generations. However, with longer haplotype blocks, it is more difficult to pinpoint the causal variant for a condition of interest. Therefore, the optimal method to identify disease related loci through association studies is to perform analyses in multiple different ethnic populations, including isolated populations, in order to narrow in on regions of interest for various genetic conditions [Hatzikotoulas et al., 2014].

In order to better understand why it is important to study the genetics of isolated, non-European white populations, it is helpful to reflect on the development of the human reference

genome. For centuries, people have been fascinated by what makes humans different from other species and from one another, but it wasn't until the early-mid 20th century that greater advances were made towards discovering the basis of these differences. Researchers James Watson and Francis Crick first proposed the double helix as the structure of DNA in 1953, and as DNA sequencing eventually became automated, the Human Genome Project was officially initiated in 1990 [Pray, 2008; International Human Genome Sequencing Consortium, 2004]. Many of the formative years of the project were focused on improving efficiency of sequencing and other methodologies. Even prior to the project's formal proposal, scientists in 20 sequencing centers from six different countries began collaborating on this international project, which they believed would change the scientific community and healthcare practices forever [What a long, strange trip it's been. . ., 2001]. Volunteers for DNA collection were sought out through advertisements placed in neighborhoods around the sequencing centers, and ultimately many more samples were collected than were actually used in the project. Thus the identities of the individuals' genomes analyzed remained anonymous, even to the volunteers themselves. In February of 2001, 90% of the Human Genome project was published as a draft, and it eventually revealed that there are approximately 20,000 genes in the human genome rather than 100,000 as was previously estimated [International Human Genome Sequencing Consortium, 2004]. The completed sequence was released in 2004 and covered approximately 99% of the euchromatic genome. This included the majority of gene-rich portions and transcriptionally active regions of the genome, with only 341 sequence gaps and an error rate of about 1 per 100,000 base pairs. This was greatly reduced from the draft sequence, which initially contained approximately 150,000 gaps [International Human Genome Sequencing Consortium, 2004]. Gap resolution was difficult due to regions of the genome where there are substantial sequence similarities, such as with

segmental duplications and other repetitive sequences. In order to tackle these challenges, the project utilized PCR and traditional sequencing methods to obtain a sequence that bordered a gap. Research scientists in the project then used the ends of these sequences as the primers for the gap region and completed long-range sequencing into the gap regions, a method known as primer walking. These methods were largely effective in continuing sequencing into the gaps and resulted in the vast majority of them being resolved in the finished sequence. With the majority of artifactual duplications being resolved, further studies of known duplicated segments such as the 22q DiGeorge region, the 17p Charcot-Marie-Tooth region, and the Y chromosome became possible. In addition to this, the development of a reference genome has aided in the identification and annotation of all protein coding regions of the genome, which has proven essential for improving the understanding of disease mechanism and etiology. Lastly, the presence of the sequence has provided a gateway into exploring genetic adaptation through the birth of new genes and creation of pseudogenes through gene death [International Human Genome Sequencing Consortium, 2004]. These are just a few examples of the clinical utility of a near-complete human reference sequence.

The establishment of a human reference genome also allowed for the development of multiple testing methodologies such as gene chip arrays; however, limitations of the reference quickly became apparent. Given that the majority of the individuals whose genomes were used during the establishment of the human reference genome were of European ancestry, many unique SNPs that are seen only in some populations were likely not discovered, which resulted in a reference sequence that was not necessarily representative of all different ethnicities. Probes that were created to match certain regions of the genome for sequencing and amplification have proven difficult to design due to genetic variation between individuals across different

populations. Single nucleotide polymorphisms within the probe regions hinder proper sequencing, and the presence of these SNPs varies across individuals and populations [Rosenfeld et al., 2012]. The HapMap Consortium and the 1000 Genomes project have attempted to address this disparity to a large extent.

The HapMap project, which was completed in two phases, aimed to characterize a large proportion of SNPs found in 270 individuals from geographically distinct areas. Phase II of the project was able to capture over 3.1 million common SNPs and provide insight on linkage disequilibrium and the distribution and frequency of the changes. The study also found that 10-30% of individuals from the same population share at least one region of extended genetic identity [Frazer et al., 2007]. By creating a more robust archive of common SNPs and learning more about haplotypes and recombination events, the project allowed for genome-wide association studies to become a more widely available tool for studying common genetic conditions such as cancer and cardiovascular disease.

The 1000 Genomes project was designed to increase the number of variants reported in human genomes by analyzing 2,504 individual genomes from 26 different populations [Gibbs et al., 2015]. The project identified more than 80 million variants, the vast majority of which were single nucleotide polymorphisms (SNPs). The study has estimated that it includes >99% of SNPs with a frequency of greater than 1% across multiple ancestries. Interestingly, while sequences from individuals of African ancestry showed greater levels of diversity, this did not correlate to the number of disease-related variants identified in these individuals. The highest yield for pathogenic alleles was seen in individuals of European ancestry, likely due to the fact that these populations have been studied in greater detail than non-European populations, and therefore alleles in these individuals are detected and interpreted more readily [Gibbs et al., 2015]. For this

reason, it is of the utmost importance to explore both common and rare genetic variants across all populations, in order to better characterize the clinical significance of variants for both Mendelian and common disorders.

While many population studies such as these have been performed, the findings of these studies have only emphasized the importance of continuing this work and how it is essential to study as many different populations as possible. While international genome-wide association studies may prove beneficial in encouraging collaboration and sample diversity, the benefit of studying homogeneous populations remains a constant. Certain polymorphisms and pathogenic changes have likely arisen in isolated populations, and without performing more detailed studies in populations across the globe, the extent of genetic diversity cannot be fully known.

As the emphasis on personalized medicine has grown over recent years, the importance of understanding genetic variation across populations has only increased. Although sequencing technologies have improved drastically, recent studies have shown that the technologies are not without flaws. The detection of low to moderate frequency variants (<5%) in population databases poses the question of whether these variants are truly present in individuals or if they represent some sequencing error in the technology. While many of these variants may be due to DNA polymerase sequencing errors, recent evidence has shown that variants introduced into the sample prior to PCR amplification are a cause for some of these low to moderate frequency sequencing calls, likely due to the introduction of changes early on in sequence amplification [Chen et al., 2017]. The presence of these variants results in confusion with interpretation of variants. Many widely used databases have incorrectly reported some of these variants in their repositories, particularly in G-to-T variant reads [Chen et al., 2017].

Exome sequencing has facilitated deeper analyses of genetic variation across different populations, although the interpretation of rare variants identified in select populations still remains a challenge. With exome sequencing, nearly all of an individual's protein-coding information is sequenced, often yielding never-before-seen sequence variants, which can provide further insight on genetic diversity. Additionally, the prevalence of alleles which are associated with a variety of conditions, such as cystic fibrosis and phenylketonuria, can be elucidated with exome sequencing, which can help in guiding better risk assessment and counseling of patients. Recognition of the limits of our understanding of variation in the human genome across populations has sparked the interest of many. Individuals of some ethnic groups have even begun studying genetic variation of different populations in order to better characterize all variants. Until very recently, such studies have not been carried out in individuals of Armenian ancestry.

I.II. *Armenian Genetic Evolution and Migration Patterns*

The Republic of Armenia is a small nation located just below the Caucasus Mountain range, in a region historically referred to as the Near East. Armenia has an ancient history which dates back to before the Bronze age, and evidence has shown that Armenian farmers of the Near East admixed with other populations who migrated to Europe during the Neolithic [Haber et al., 2015]. A 2015 study collected a set of 173 Armenian samples from Armenia and Lebanon and performed genome-wide analysis to compare Armenian genetic variation with 78 other populations worldwide. Results showed that Armenians cluster with other ethnic groups of (1) European, (2) Near Eastern, and (3) Caucasus origins. In particular, Armenians were genetically closest within these groups to (1) Spaniards, Italians, Romanians, (2) Lebanese, Jews, Druze, Cypriots, (3) Georgians, and Abkhazians [Haber et al., 2015]. Multiple mixture events were

shown to have occurred, initially around 3800 BCE between sub-Saharan Africans and Western Europeans, then Sardinians and Caucasus groups, and lastly multiple Eurasian peoples. However, after approximately 1200 BCE, Armenian admixture with other populations seems to have ceased, creating a population that has been genetically isolated for the last 3000 years [Haber et al., 2015].

While Armenia is a land-locked country, it once spanned from sea to sea, reaching the Mediterranean, and encompassing much of what is now called Eastern Turkey. The shifts in its territorial holdings, migration patterns, and genetic admixture have repeatedly been influenced by conflicts with other nations. Similar to Lebanese Christians and Sephardic Jews, cultural, religious, and ethno-linguistic differences between Armenians and their neighbors resulted in the creation of a genetically isolated population. In 1915 CE, these differences greatly contributed to what became known as the Armenian Genocide, the extermination of nearly 1.5 million Armenians. In the midst of World War I, Armenians were displaced from their ancestral homes in current-day Eastern Turkey, and mass deportations resulted in many moving to the United States, Canada, France, Germany, Lebanon, Syria, Iran, and Egypt [Chaloyan, 2017]. Following these events, Armenia was taken over by the USSR and remained as a member of the Soviet Union for the majority of the 20th century, until it regained independence in 1991 [Chaloyan, 2017]. As the Armenian diaspora grew and living conditions suffered under new leadership, families encouraged their loved ones to also emigrate to countries like the U.S. This ultimately resulted in another wave of migration with 610,000 people exiting the country between 1991-1995.

Currently, reports claim that approximately 3.5 million Armenians remain in the Republic of Armenia, while there are between 7.5-8 million Armenians in the diaspora

[Chaloyan, 2017]. Many of these individuals reside in California, particularly Los Angeles, and other cities along the East Coast. The formation of different diaspora populations has also contributed to a slight genetic divergence even among Armenians. Armenians who predominantly left the region following the events of 1915, sometimes referred to as Western Armenians, can be differentiated from Armenians who resided in Eastern, or modern day Armenia [Haber et al., 2015]. However, the Armenian clusters were significantly more similar in genetic structure to one another than they were to other populations, reaffirming that the population has been genetically isolated for many years. Research in recent years has demonstrated the value of studying isolated populations in order to learn more about rare variant-phenotype associations, gene-environment interactions, and genetic diversity between and among populations [Haber et al., 2015]. This study will aim to do just that, by characterizing a small, healthy Armenian population's genetic variation. With regards to Armenians, one cannot speak of genetic conditions without discussing familial Mediterranean fever.

I.III. *Familial Mediterranean Fever - the most common hereditary condition in Armenians*

Familial Mediterranean fever (FMF) is a genetic condition that is found at a higher frequency in individuals of Armenian ancestry. FMF is an autosomal recessive condition that is most often associated with recurrent episodes of high fever, inflammation and severe abdominal pain [Shohat and Halpern, 2011]. The condition can be categorized into two types: type 1 and type 2. In type 1, febrile and inflammatory episodes usually begin in early childhood or adolescence. In adulthood, there is an increased risk for amyloidosis, a build-up of abnormal protein in an organ, and eventual end-stage renal disease, particularly if individuals are untreated. In individuals with type 2 FMF, amyloidosis in adulthood is often the presenting feature of the

condition. Individuals with FMF are typically treated with colchicine, which greatly reduces the likelihood of febrile and inflammatory episodes, as well as amyloidosis [Shohat and Halpern, 2011]. For individuals who seek treatment early in life and continue management into adulthood, the prognosis for FMF is quite good. Interestingly, some individuals who are diagnosed with FMF only have one identified mutation in the associated gene, *MEFV*. One study found that up to 25% of individuals diagnosed with FMF may be heterozygotes; therefore, it is difficult to determine if these individuals are truly affected due to a single mutation or if other modifying factors play a role as well [Marek-Yagel et al., 2009]. The most common variants seen in FMF carriers are p.M694V, p.M694I, p.V726A, p.E148Q, and p.M680I. The carrier frequency of FMF in Armenian individuals is approximately 1 in 5; therefore, the reproductive risk for Armenian couples is quite high and warrants consideration of carrier testing for these couples [Sarkisian et al., 2008]. While there has been negligible interest in pregnancy termination for the condition, some couples may find that the knowledge that they are at risk to have a child with FMF can help them to identify the early signs of the condition and therefore implement treatment earlier on.

In addition to FMF, Armenians are at risk for other recessive conditions, just as individuals of all populations are, regardless of ancestry. Many of these conditions could pose greater health risks to an individual than FMF typically does. With continued advancements in sequencing and variant analysis, the availability of carrier screening panels has grown tremendously, and the uptake of expanded carrier screening is only increasing. In this study, the predicted performance of one of the most common commercially available expanded carrier screening panels which tests for 176 conditions will be evaluated in an Armenian population.

The following section will review how carrier screening has evolved over the years and how expanded carrier screening has become widely used in the clinical setting.

I.IV. *Expanded Carrier Screening*

Carrier screening is genetic testing that is performed to determine if an individual is at an increased risk to have a child with a certain genetic condition. The majority of carrier screening is for autosomal recessive and X-linked recessive conditions, in which carriers typically do not express any features of the condition. If two individuals who mate are carriers for pathogenic variants in the same gene related to a genetic condition, there is a 25% chance to have a child affected with the condition [Bajaj and Gross, 2014]. Traditionally, carrier screening was performed for individuals who were expected to be at high risk for certain genetic conditions, either due to their ancestry or family history. For example, individuals of Ashkenazi Jewish ancestry might seek counseling and screening for conditions such as Tay-Sachs disease, and African Americans might choose to have screening for sickle cell anemia. However, many ethical dilemmas arose throughout the 20th century due to targeted ethnic carrier screening. Individuals were often discriminated against by employers, the military, or society in general, and relations between individuals of certain populations became tense [Fulda and Lykens, 2006]. Eventually, carrier screening practices changed by making testing options more universal to all individuals, regardless of *a priori* risk. Both the American College of Medical Genetics and the American College of Obstetricians and Gynecologists now recommend offering carrier testing for two recessive conditions, cystic fibrosis (CF) and spinal muscular atrophy (SMA), to all women during pregnancy or preconceptionally [Grody et al., 2013; Romero et al., 2017]. In addition to these two, carrier screening for inherited forms of anemia, known as

hemoglobinopathies, is also offered to most individuals who are from a high-risk ethnicity group. Although targeted carrier screening is effective to an extent, given the major advances in sequencing technologies and the plummeting costs of testing, expanded carrier screening for all individuals, regardless of *a priori* risk, is becoming quite popular [Henneman et al., 2016].

Expanded carrier screening (ECS) allows for an individual to determine their carrier status for multiple, often over one hundred, genetic conditions simultaneously. Much like traditional carrier screening, ECS typically screens for the most common autosomal recessive and X-linked conditions that are seen across populations. The American College of Medical Genetics (ACMG) released recommendations in 2013 to aid genetic testing laboratories in the selection of conditions to be screened for on ECS panels [Grody et al., 2013]. The exact language used in the recommendations is as follows:

1. Disorders should be of a nature that most at-risk patients and their partners identified in the screening program would consider having a prenatal diagnosis to facilitate making decisions surrounding reproduction.
2. When adult-onset disorders (disorders that could affect the offspring of the individual undergoing carrier screening once the offspring reaches adult life) are included in screening panels, patients must provide consent to screening for these conditions, especially when there may be implications for the health of the individual being screened or other family members.
3. For each disorder, the causative gene(s), mutations, and mutation frequencies should be known in the population being tested, so that meaningful residual risk in individuals who test negative can be assessed.

4. There must be validated clinical association between the mutation(s) detected and the severity of the disorder.
5. Compliance with the American College of Medical Genetics and Genomics Standards and Guidelines for Clinical Genetics Laboratories, including quality control and proficiency testing.

Given major advancements in technology and the increasing knowledge of genetic testing in society, largely due to the propagation of direct-to-consumer genetic testing, general population screening for multiple conditions is likely to become the new standard of practice. While there are many benefits to ECS, such as providing couples with significantly more information on their risks for certain genetic conditions, it is important to remain cautious about how ECS is implemented clinically. As the ACMG recommended, couples should be allowed to decide whether or not they would like to be informed about risks for conditions that have variable expressivity or reduced penetrance, particularly if the condition is typically mild [Grody et al., 2013]. Interestingly, FMF is one such condition in which the utility of inclusion on ECS panels has been questioned, particularly because individuals at risk to have a child with FMF often do not consider prenatal testing and the condition does not always present in childhood. However, given that the condition can be life threatening if amyloidosis develops and given the high carrier frequency in some populations, the condition has consistently been included on many ECS panels. Accurate and comprehensive information regarding the residual risks for the conditions is also of the utmost importance when carrier screening results are reported out to individuals since ECS is a form of screening and will not detect all carriers for any given condition. Two recent American studies showed that between 25%-30% of individuals who underwent carrier

screening using an ECS panel which included approximately 400 known pathogenic variants were found to be a carrier for at least one condition [Peyser et al., 2018; Lazarin et al., 2013]. Two other studies which were performed in Spain reported values as high as 56%-84%, although these two studies included screening of 200 and 549 genes, respectively [Abulí et al., 2016; Martin et al., 2015]. Thus, the number of genes analyzed in the ECS panels and the methodologies used in these studies, which ranged from targeted genotyping of common variants to next-generation sequencing (NGS), varied greatly and likely impacted the number of carriers that were identified in each of the studies. Estimates of the yield of ECS panels may continue to increase as ECS panels become larger. ACMG recommends genetic counseling for all couples that are interested in ECS, as well as more detailed post-test counseling in the event of positive results [Grody et al., 2013].

The American College of Obstetricians and Gynecologists (ACOG) has also issued multiple position statements regarding carrier screening. In position statement 690, ACOG's recommendations for including a condition on an ECS panel are that the condition have a carrier frequency of 1 in 100 or greater, that the condition have a well-defined phenotype, that the condition have a detrimental effect on quality of life for most affected individuals, that the condition possibly cause cognitive or physical impairment, that the condition typically require medical or surgical intervention, that the condition often have a childhood onset, and that there be prenatal diagnosis available for the condition [Rink et al., 2017]. In position statement 691, ACOG presents guidelines for testing considerations based on the patient's ethnicity [Romero et al., 2017]. While there is significant overlap between the recommendations from ACMG and ACOG, there are a few differences.

Notably, ACOG's 1 in 100 carrier frequency criterion from position statement 690 is limiting in the number of conditions that should be included in ECS, as shown by a recent study [Ben-Shachar et al., 2019]. The study was conducted using data from couples that underwent ECS for 176 conditions offered through a widely used commercial laboratory. Carrier rates and at-risk couple rates were elucidated based on three different types of panels recommended by ACOG. The first panel consisted of 172 of the 176 conditions originally on the ECS panel since these 172 met at least 6 of the 7 ACOG 690 criteria. The second panel that was considered was based on the ethnicity-specific recommendations highlighted in ACOG's committee opinion 691 [Romero et al., 2017]. The last panel included testing for only cystic fibrosis and spinal muscular atrophy, since these two conditions are recommended for testing in all ethnicities [Ben-Shachar et al., 2019]. With a 172 condition panel, carrier identification and at-risk couple detection each decreased by 3% as compared to the 176 condition panel. When the ACOG 691 panel was used, this decreased carrier identification by 77% and at-risk couple detection by 66% as compared to the 176 condition panel. Lastly, with CF and SMA only, carrier identification decreased by 88% and at-risk couple detection decreased by 84% as compared to the 176 condition panel [Ben-Shachar et al., 2019].

One can see that although the guidelines were aimed to improve overall clinical practices, and some may argue that they do, this study showed that the number of at-risk couples that would be identified if such stringent criteria were used is lower when using some of ACOG's recommended criteria. With regard to the 1 in 100 criteria, the study considered three different definitions for how this could be interpreted: a 1 in 100 carrier rate in any ethnicity, a 1 in 100 US-weighted carrier rate, or a 1 in 100 carrier rate in all ethnicities. Using one of these definitions, carrier screening panels would test between 3 and 38 conditions, which would reduce

carrier detection rates by 36%-79% and at-risk couple detection rates by 11%-92%. As previously stated, the benefit of ECS is that it identifies a greater number of at-risk couples. However, identifying additional carriers may incur greater costs for insurance companies and patients since some individuals might use knowledge of their carrier status in order to pursue further testing, such as carrier screening in their partner, *in vitro* fertilization (IVF), or preimplantation genetic testing for monogenic/single-gene disorders (PGT-M). The study performed a cost-benefit analysis by comparing the number of at-risk couples identified, which is the benefit of ECS, to the number of carriers identified, which is one of the costs of ECS and was the focus in this paper [Ben-Shachar et al., 2019]. They found that adding rarer conditions to the panel helped to increase the number of at-risk couples identified and that the relationship between the number of at-risk couples identified to the number of carriers identified was roughly linear until a condition carrier rate of approximately 1 in 1000 [Ben-Shachar et al., 2019]. Thus, Ben-Shachar et al. infer that it may be worthwhile to include conditions on ECS panels with a carrier rate of 1 in 1000 or more frequent, since adding more conditions tends to increase the number of at-risk couples detected, which is the presumed benefit of performing ECS, without disproportionately increasing insurance-related costs. However, for very rare conditions where many variants identified would not be able to be classified as pathogenic or benign, the clinical utility would be very low and would not provide as much useful information to couples or inform them of their residual risk [Ben-Shachar et al., 2019]. This concept provides yet another reason why it is critical to perform more research in understudied populations, since this would help to increase the clinical detection for individuals of these populations.

Given the increasingly complex considerations that must be discussed regarding ECS, genetic counseling has become more imperative than ever. Genetic counselors frequently

confront difficult discussions with patients due to the growing number of testing options, and therefore it is important to provide counselors with as many tools as possible to aid in counseling individuals of different backgrounds. Although ethnicity is often a key consideration when counselors discuss the availability of different types of prenatal and preconception testing, there are many ethnicities in which specific carrier frequencies or even estimates are not well known, including Armenians. Thus, further discussion and studies are necessary in order to increase the amount of information available to genetics professionals in order to provide patients with more accurate and culturally sensitive information.

IV. *Genetic Counseling*

The National Society of Genetic Counselors defines genetic counseling as "the process of helping people understand and adapt to the medical, psychological and familial implications of genetic contributions to disease" [Resta et al., 2006]. Indeed, genetic counseling is a process in that it requires an extensive knowledge of a wide array of genetic conditions and the ability to interpret family and medical histories to assess the likelihood of a given condition. Genetic counselors also provide education, resources, and support to patients and other clinicians. At the core of genetic counseling values is the belief that all individuals should be given the opportunity to make informed decisions regarding their own medical care, and it is the genetic counselor's job to aid in this decision-making. Over the years, genetic counseling has expanded as a profession, especially in the last two decades following major advancements in molecular technologies. While the majority of genetic counselors in practice are women of European ancestry, there are a growing number of individuals from different cultural backgrounds who are seeking genetic counseling services [Mittman and Downs, 2008]. Recent U.S. Census data, as

well as studies from the 2010 Census, have shown that there are a growing number of individuals in the U.S. reporting non-white and multiracial backgrounds [U.S. Census Bureau, 2012]. This is primarily due to marriage between individuals of different cultures, which will undoubtedly continue to happen at an increasing rate in America.

Given these trends, many professionals are arguing for the use of expanded carrier screening rather than ethnicity-based screening in all women of reproductive age, given that the latter would fail to detect many carriers and at risk couples that would not meet recommended criteria for testing [Peyser et al., 2018]. The American College of Medical Genetics and Genomics, American College of Obstetricians and Gynecologists, National Society of Genetic Counselors, Perinatal Quality Foundation, and Society for Maternal-Fetal Medicine released a joint statement in 2015 providing recommendations for counseling about ECS to patients [Edwards et al., 2015]. The statement recommends that patients should be counseled on basic information about the conditions being screened for and the detection rates for these, although it would not be possible to discuss each condition in detail. The statement also provided insight on future needs to improve ECS, which includes development of a repository of curated variants, which would be made possible through extensive pan-ethnic genotyping. As genetic counselors, it is important to be aware of what testing is appropriate to offer to individuals, not only based on published society recommendations, but also on a patient-specific basis. Different individuals have differing comfort levels with knowing more or less information, so it is important to discuss both the benefits and limitations of ECS and other carrier screening options with patients. Additionally, as previously mentioned, the clinical utility of ECS for some ethnicities is lower since it can be more difficult to interpret variants identified in understudied populations. Therefore, it is critical to emphasize that ECS is not a diagnostic tool, but rather a screening tool

that is designed to have a high detection rate for the majority of conditions tested in as many ethnicities as possible, but that ultimately it is not designed to detect all possible changes or conditions that a person could carry. This truth should especially be emphasized in individuals of non-European ancestry, at least until more studies are completed to better identify disease-causing and benign variation across all populations.

I.VI. *Significance and Aims of this Study*

Given the lack of published research on the Armenian community and the increasing uptake of expanded carrier screening and other forms of genetic testing, this study will aim to better understand genetic variation among Armenians. Studies have shown that the percentage of individuals that are detected to be a carrier for at least one condition when expanded carrier screening is completed can be quite high in some populations [Peyser et al., 2018; Abulí et al., 2016; Martin et al., 2015; Lazarin et al., 2013]. By analyzing exome sequences from 44 Armenian individuals for 176 conditions that are included in one commercially available expanded carrier screen (Appendix B), the expanded carrier screening positive rate (ECS-PR) can be estimated for individuals of this ethnicity. Additionally, this score can be compared to estimates proposed by other studies in the literature, primarily in individuals of European ancestry. Since it is already known that there is a high carrier frequency for FMF in Armenians, this study will also aim to identify the likelihood of receiving a positive carrier screen result for other conditions. I hypothesize that the ECS-PR in Armenians will be similar to previously reported rates in other populations; however, if FMF were to be removed from analyses to evaluate for the ECS-PR without the condition, I hypothesize that the ECS-PR in Armenian individuals will be lower than previously reported rates in other populations.

Another goal of this study is to investigate rare variants that may be identified in this sample, some of which have likely been previously reported in databases as being of uncertain significance. This study may also identify variants that have not been seen in other populations thus far. I hypothesize that all of the individuals will have variants that are of unknown significance or that have not been seen before. It is likely that some of the genes in the expanded screen are better studied than others, and may have fewer never-before-seen variants. However, given that Armenians are an understudied population, I hypothesize that there will be many novel variants identified. To our knowledge, this study is the first of its kind in the Armenian population. Overall, the goal of this study is to provide more information to the public and medical professionals about Armenian genetic variation. More comprehensive studies are necessary in this population in order to improve public health and medical care in both Armenia and for Armenians in the diaspora. The hope is that this information will improve genetic counseling provided to individuals of Armenian ancestry, particularly in prenatal and pre-conceptional settings, and to encourage others to pursue similar studies in other ethnic populations. Knowledge of carrier status is an important piece of information for many couples that are planning a family, and as information grows about genetic variation in the Armenian population specifically, testing may become more available and attractive to individuals of the population. This would promote diversity and equity in the field of genetics, and improve the overall quality of services for providers and patients alike.

II. METHODS

II.I. *The ArmGenia Project*

In 2015, a group of Armenian investigators received seed funding from Sara Chitijian, a UCLA alumna and daughter of a Genocide survivor, in order to begin an Armenian Reference Genome project. Drs. Sevak Avagyan and Evgeni Sokurenko, co-founders of the ArmGenia Research Charitable Trust, teamed up with the UCLA Center for World Health and Dr. Wayne Grody, professor of pathology and laboratory medicine, pediatrics, and human genetics at the David Geffen School of Medicine at UCLA in order to pursue the ArmGenia project. Their primary goal was to shed light on the genetic origins of the Armenian people, in order to increase knowledge and improve healthcare for Armenians, especially for FMF [Djourabchi, https://www.worldhealth.med.ucla.edu/index.php/ucla-armgenia-mou-signing-mapping-of-the-armenian-genome/, accessed April 5 2019].

II.II. *IRB Approval*

The ArmGenia project has Institutional Review Board (IRB) approval in Yerevan, Armenia. An agreement was drafted with UCLA in order to have data primarily analyzed at UCLA. Dr. Wayne Grody and Dr. Salpy Akaragian, former director of the UCLA International Nursing Program and primary liaison to the ArmGenia Research Charitable Trust, have determined that no further IRB approval is necessary since the data that was utilized in this study is completely de-identified. Additionally, a Request for Determination of Human Studies Research was submitted to the UCI IRB, and it was confirmed that this project "does not qualify as human subjects research because the activities do not involve human subjects given that all

data will be de-identified. Therefore, the activities are not subject to UCI IRB review and approval." This email statement was received on 11/28/2018 from the UCI IRB office.

II.III. *Data Collection and Sample Information*

The initial set of samples of the ArmGenia project was collected from 25 unrelated and healthy Armenian individuals who were at least 60 years old at the time that they enrolled in the project. At a later date, a second set of samples was collected from 19 individuals who were also unrelated and healthy Armenian individuals and at least 60 years old at the time that they enrolled in the project. The 44 individuals originated from multiple regions across Armenia, Nagorno-Karabakh, and what was formerly known as Western Armenia. No additional information is known about the individuals whose exomes were sequenced.

II.IV. *Exome Sequencing*

Exome sequencing and analysis was performed in two separates batches at UCLA using the following identical pipeline for each. Raw FASTQ files were aligned in a cloud-based pipeline using BWA-MEM software. Raw alignments were checked for quality using GATK HaplotypeCaller by calculating the Genotype Quality (GQ). The GQ is calculated using Phred-scaled likelihood (PL) scores, which are defined as $PL = -10 * \log P(\text{Genotype}|\text{Data})$. The GQ is the difference between the PL of the second most likely genotype, and the PL of the most likely genotype, and the highest possible confidence value is 99. Variant calling was performed using GATK HaplotypeCaller using joint variant calling, meaning that a set of previously sequenced samples were combined with the novel samples of interest in order to reduce calling bias and false positivity. The two batches (25 samples and 19 samples, respectively) were each processed

with a similar-sized sample of non-Armenian individuals. Data from these non-Armenian individuals were later omitted during data analysis. Variant calls were recalibrated and recalled for short indel detection. Two separate VCF files with the 25 and 19 samples were created and included single nucleotide variants (SNVs) and short indels. Within the respective VCF files the chromosome position, genotype, allelic depth, read depths, genotype quality, and variant allele frequency were included for each of the individual samples. Variants were annotated using VarSeq® version 2.1.0 (Golden Helix, Inc., Bozeman, MT), as is outlined in the following section.

II.V. *Variant Annotation and Filtration*

VarSeq® is a software that can be used to annotate large variant data sets. VarSeq® imports variant annotation information directly from a number of different sources. This study utilized annotations provided by RefSeq Genes 105 Interim v1, NCBI (https://www.ncbi.nlm.nih.gov/refseq/rsg/) and ClinVar (data frozen from 2017-09-05, NCBI; https://www.ncbi.nlm.nih.gov/clinvar/) [Pruitt et al., 2012; Landrum et al., 2016]. VCF files of the UCLA exome data were imported into VarSeq® for analysis.

Variants were filtered to include only those in the genes associated with the 176 conditions tested for in the Foresight (Myriad Women's Health, South San Francisco, CA) expanded carrier screen, henceforward to be referred to as the ECS panel, which is available commercially for clinical genetic testing. This panel is used frequently in prenatal and adult clinical genetics settings, and was selected by this study in order to assess the yield of this panel in the Armenian population as compared to other populations from previous studies. A list of the genes and conditions included in the ECS panel can be found in Appendix B, as well as

information regarding ACOG and ACMG testing recommendations for some of the genes. This process was completed separately for each of the batches (25 and 19 samples, respectively). The variants were exported for each of the batches as a spreadsheet, and the spreadsheets were manipulated to include only the 25 and 19 Armenian samples, respectively. A combined data set of the 44 Armenian samples and all variants observed in either or both of the data sets was created. Data management to merge the data from the two batches was accomplished using SPSS v.25 (IBM Corp., Armonk, NY). Genotypes were excluded from further analysis if they had a GQ score of <90, a minor allele frequency (MAF) of the alternate allele of <.35, or if there were low or inconsistent quality scores in multiple individuals for the variant.

II.VI. *Data Analysis*

In the annotated data set, variants were assessed based on their clinical significance reported in ClinVar. Using the "Clinical Significance" column from ClinVar in the exported spreadsheets, variants were examined if they had at least one assertion that was pathogenic or likely pathogenic, or if the variant had no associated ClinVar record. These variants were subsequently reviewed for additional or updated classifications in ClinVar (April to May, 2019) and any newly available assertions were imported into the data set. Variants are included as pathogenic or likely pathogenic in this study based on the consensus ClinVar classification (2015 - May 2019). For the remaining number of no-call variants not reported on ClinVar as of May 2019, Varsome (Saphetor SA, Lausanne, Switzerland; https://varsome.com/) was accessed between April 2019-May 2019 and was used for further analysis.

Varsome is a free online variant analysis tool which compiles information from more than 30 cross-referenced public data resources, including ClinVar, RefSeq, GnomAD

exomes/genomes (https://gnomad.broadinstitute.org/), ExAC (http://exac.broadinstitute.org/), publications, structural variant tools, and *in silico* prediction models [Kopanos et al., 2018]. Varsome uses ACMG guidelines for variant classification [Richards et al., 2015], in order to automatically predict the ACMG classification for each variant, even if the variant is novel [Kopanos et al., 2018]. Variants that had an ACMG predicted classification of benign, likely benign, or VUS were not included in the list of pathogenic and likely pathogenic variants reported in this study. Two variants that were not reported in ClinVar were included as pathogenic or likely pathogenic because they had an ACMG predicted classification of pathogenic or likely pathogenic in Varsome with the following supporting lines of evidence: PVS1 (premature truncation) and PM2 (rarity) [Richards et al., 2015].

The total number of variants identified in any of the Armenian individuals was quantified with SPSS using the merged data set that combined all variants seen in the initial 25 samples, the 19 samples, or both. The total number of pathogenic and likely pathogenic variants and the number of pathogenic and likely pathogenic variants identified in each sample were counted. Expanded carrier screening positive rates (ECS-PRs) were calculated in this population when all of the 176 conditions were included and also when FMF was excluded. Confidence intervals (95% CI) were calculated for each of the carrier frequencies associated with the conditions in which there were multiple carriers identified. A 95% confidence interval was also calculated for each of the ECS-PRs. Allele frequencies were collected from GnomAD (date accessed May 2019) for the variants associated with conditions in which there were multiple carriers identified. Both the ECS-PRs and the allele frequencies were compared between the findings in this study and findings that have previously been reported.

III. RESULTS

III.I. *Characterization of Variants and Assessment of Associated Conditions*

This study aimed to characterize the spectrum of genetic variation related to autosomal recessive disorders seen in Armenian individuals in a subset of clinically relevant genes. The vast majority of the variants reported were missense and synonymous variants. A breakdown of how variants were analyzed and classified is displayed in Figure 1.

The total number of variants present in the genes associated with the ECS panel in the VCF file was 2085 (Figure 1; I.I). Of these 2085 variants, 1307 variants were detected in at least one of the 44 Armenian individuals (Figure 1; II.I). Therefore, 778 variants were only present in the non-Armenian individuals and were omitted from further analyses (Figure 1; II.II). Of the 1307 variants identified in the Armenian cohort, 913 were annotated as benign, likely benign, variant of uncertain significance (VUS), or other according to ClinVar (freeze 2017 09 05) (Figure 1; III.II). No further analyses were performed with these variants. These variants are included in the list of all other variants called in the VCF files in Appendix A.

In the 44 individuals, a total of 66 variants had at least one pathogenic or likely pathogenic assertion, or had conflicting interpretations of pathogenicity, as annotated by ClinVar (freeze 2017 09 05) (Figure 1; III.I). Three of these 66 variants were omitted because of low GQ (<90), low MAF (<.35), or inconsistency in the quality of the call across individuals for that variant (Figure 1; IV.II). Of the 63 that were evaluated on ClinVar (April to May 2019), 21 variants were classified as pathogenic or likely pathogenic (Figure 1; V.I). Of the variants that initially had at least one pathogenic, likely pathogenic, or conflicting interpretations of pathogenicity assertion, 42 were reported as benign, likely benign or VUS based on the consensus on ClinVar (Figure 1; V.II).

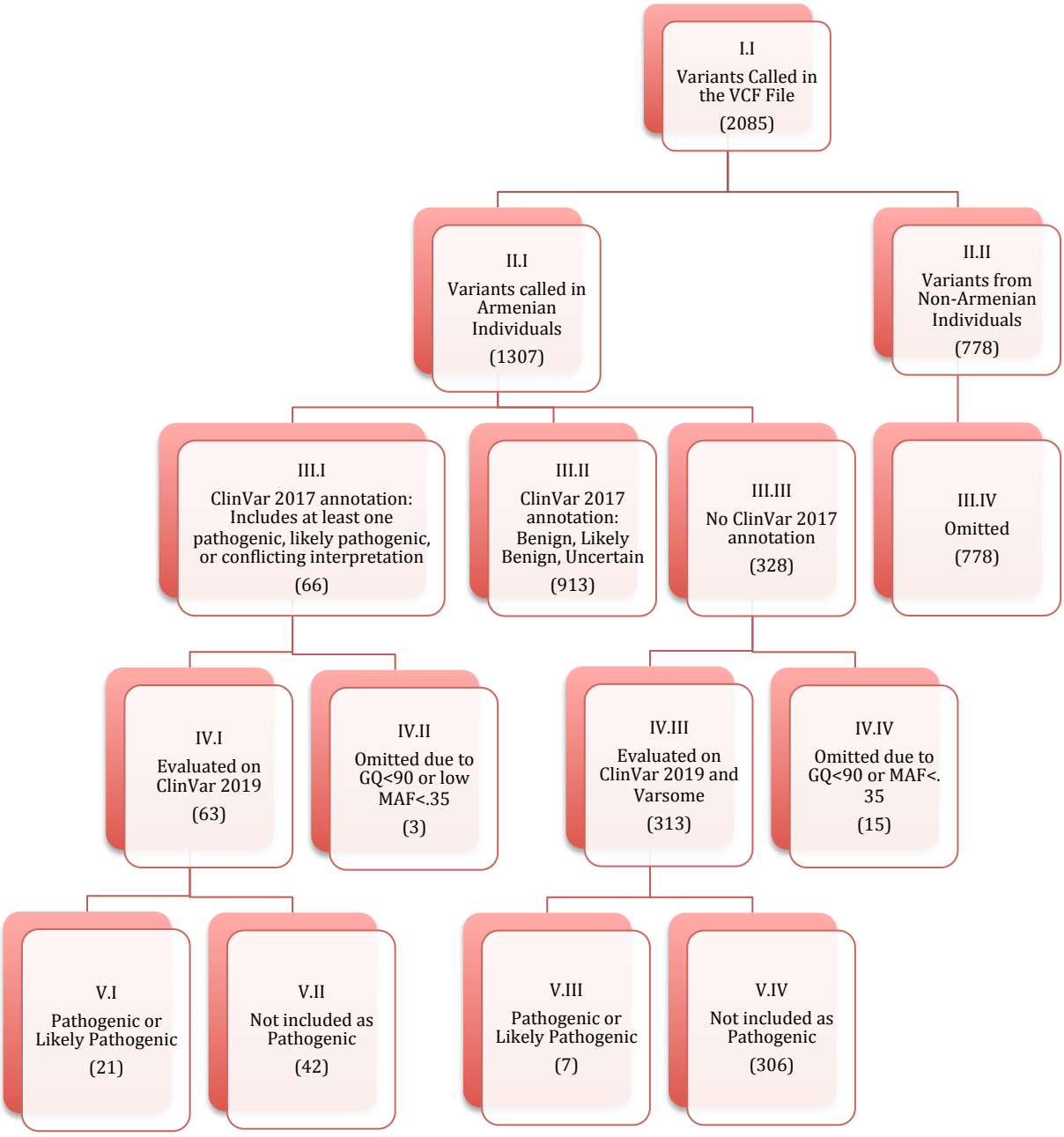

Figure 1. Variant Curation and Carrier Identification Flowchart: The total number of variants (I.I) was divided into those in the Armenians (II.I) and those not in the Armenians. III.I, III.II, and III.III represent different annotations provided by ClinVar (freeze 2017 09 05). V.I and V.III represent variants with a ClinVar (April-May 2019) consensus of pathogenic or likely pathogenic or a Varsome pathogenic or likely pathogenic classification (as outlined by ACMG, Richards et al., 2015). V.II and V.IV represent variants that were not reported as pathogenic or likely pathogenic after checking ClinVar and Varsome.

A total of 328 variants had no annotation from ClinVar (Figure 1; III.III). Fifteen of these 328 variants were omitted because of low GQ (<90), low MAF (<.35), or inconsistency in the quality of the call across individuals for that variant (Figure 1; IV.IV). Of the 313 that were reviewed for updated records from ClinVar and Varsome, 7 were identified as pathogenic or likely pathogenic variants (Figure 1; V.III). The remaining 306 variants for which there was no ClinVar annotation (Figure 1; V.IV) were included in the list of all other variants called in the VCF files in Appendix A. All of the pathogenic and likely pathogenic variants (N=28) reported in at least one of the Armenian individuals are listed in Table 1.

Table 1. Pathogenic and Likely Pathogenic Variants Identified in the 44 Individuals

Gene	DNA Sequence Change	Protein Sequence Change	Type of Variant	Associated Condition
ACADS	c.973C>T	p.Arg325Trp	Missense	Deficiency of butyryl-CoA dehydrogenase
ALDOB	c.448G>C	p.Ala150Pro	Missense	Hereditary fructose intolerance
ALG6	c.1128-2A>C	–	Splice Acceptor	Congenital disorder of glycosylation type 1C
ASS1	c.470G>A	p.Arg157His	Missense	Citrullinemia, type 1
ATM	c.7671_7674delGTTT	p.Phe2558Leufs	Deletion	Ataxia-telangiectasia syndrome
CBS	c.434C>T	p.Pro145Leu	Missense	Homocystinuria, pyridoxine responsive
CFTR	c.3154T>G	p.Phe1052Val	Missense	Cystic Fibrosis
DHCR7	c.964-1G>C	–	Splice Acceptor	Smith-Lemli-Opitz syndrome
FAH	c.554-1G>T	–	Splice Acceptor	Tyrosinemia, type 1
GALC	c.956A>G	p.Tyr319Cys	Missense	Krabbe disease
GALT	c.881T>A	p.Phe294Tyr	Missense	Galactosemia
GJB2	c.35delG	p.Gly12Valfs	Deletion	*GJB2*-related Non-syndromic Hearing Loss
HEXA	c.1511G>A	p.Arg504His	Missense	Tay-Sachs disease

IDUA	c.510_511insAGTTCCA	p.His171Serfs	Insertion	Mucopolysaccharidosis type I
LAMA2	c.2451-2A>G	–	Splice Acceptor	LAMA2-related muscular dystrophy
MEFV	c.2282G>A	p.Arg761His	Missense	Familial Mediterrannean Fever
MEFV	c.2177T>C	p.Val726Ala	Missense	Familial Mediterrannean Fever
MEFV	c.2080A>G	p.Met694Val	Missense	Familial Mediterrannean Fever
MEFV	c.2040G>C	p.Met680Ile	Missense	Familial Mediterrannean Fever
NPC1	c.352_353delAG	p.Gln119Valfs	Deletion	Niemann-Pick disease type C
PAH	c.782G>A	p.Arg261Gln	Missense	Phenylketonuria
PAH	c.842C>T	p.Pro281Leu	Missense	Phenylketonuria
PAH	c.898G>T	p.Ala300Ser	Missense	Phenylketonuria
PAH	c.1208C>T	p.Ala403Val	Missense	Phenylketonuria
PKHD1	c.4870C>T	p.Arg1624Trp	Missense	*PKHD1*-Related Autosomal Recessive Polycystic Kidney Disease
PMM2	c.367C>T	p.Arg123Ter	Nonsense	Congenital Disorder of Glycosylation, Type Ia
PMM2	c.422G>A	p.Arg141His	Missense	Congenital Disorder of Glycosylation, Type Ia
SLC26A2	c.235C>T	p.Gln79Ter	Nonsense	Sulfate Transporter-Related Osteochondrodysplasia

The genes and conditions for which at least one individual in this study was found to be a carrier of a pathogenic or likely pathogenic variant are listed in Table 1. Individuals were carriers of pathogenic or likely pathogenic variants in genes related to 21 conditions. There were four conditions for which there were multiple carriers, as shown in Table 3. These were cystic fibrosis, familial Mediterranean fever, phenylketonuria, and congenital disorder of glycosylation, type Ia. Two different individuals (5%, 95% CI: <0.1%-11.4%) were carriers for the same *CFTR* variant, c.3154T>G.

Table 2. Conditions for which Multiple Carriers were Identified

Gene	DNA Sequence Change	Protein Sequence Change	Associated Condition	Number of Carriers
CFTR	c.3154T>G	p.Phe1052Val	Cystic Fibrosis	2
MEFV	c.2040G>C c.2080A>G c.2177T>C c.2282G>A	p.Met680Ile p.Met694Val p.Val726Ala p.Arg761His	Familial Mediterranean Fever	1 2 3 1
PAH	c.782G>A c.898G>T c.1208C>T c.842C>T	p.Arg261Gln p.Ala300Ser p.Ala403Val p.Pro218Leu	Phenylketonuria	1 1 1 1
PMM2	c.367C>T c.422G>A	p.Arg123Ter p.Arg141His	Congenital Disorder of Glycosylation, Type Ia	1 1

There were a total of seven carriers (16%, 95% CI: 5.2%-26.8%) for pathogenic variant associated with FMF in this healthy Armenian cohort. Four different *MEFV* variants were seen in the seven individuals. Two individuals were carriers for c.2080A>G, and three individuals were carriers for c.2177T>C. In addition to the four known pathogenic variants, five individuals were carriers for the *MEFV* c.442G>C variant (Appendix A), which was excluded from further analysis due to uncertain clinical significance [Shohat and Halpern, 2011]. Four individuals (9%, 95% CI: 0.5%-17.5%) were carriers for four distinct *PAH* pathogenic variants in this cohort. Two individuals (5%, 95% CI: <0.1%-11.4%) were carrier for two distinct *PMM2* pathogenic variants in this cohort.

The allele frequencies from GnomAD are reported for the variants in the conditions in which there were multiple carriers identified (Tables 3, 4, 5).

Table 3. GnomAD Allele Frequencies for *MEFV* Pathogenic Variants

	MEFV **c.2282G>A p.Arg761His**	***MEFV*** **c.2177T>C p.Val726Ala**	***MEFV*** **c.2080A>G p.Met694Val**	***MEFV*** **c.2040G>C p.Met680Ile**
Ashkenazi Jewish	0 / 10370 (0%)	407 / 10368 (3.9%)	0 / 10370 (0%)	0 / 10080 (0%)
European (non-Finnish)	10 / 129148 (0.008%)	115 / 129180 (0.09%)	59 / 129184 (0.05%)	2 / 113758 (0.002%)
European (Finnish)	0 / 25116 (0%)	0 / 25122 (0%)	0 / 25120 (0%)	0 / 21644 (0%)
Latino	3 / 35438 (0.008%)	8 / 35440 (0.02%)	9 / 35438 (0.03%)	0 / 34588 (0%)
South Asian	4 / 30616 (0.01%)	3 / 30616 (0.01%)	0 / 30616 (0%)	0 / 30616 (0%)
East Asian	38 / 19954 (0.2%)	0 / 19950 (0%)	0 / 19954 (0%)	0 / 18394 (0%)
African	2 / 24964 (0.008%)	0 / 24968 (0%)	2 / 24968 (0.008%)	0 / 16256 (0%)
Other	1 / 7222 (0.01%)	28 / 7226 (0.4%)	7 / 7226 (0.1%)	0 / 6138 (0%)
Total	58 / 282828 (0.02%)	561 / 282870 (0.2%)	77 / 282876 (0.03%)	2 / 251474 (0.0008%)

Table 4. GnomAD Allele Frequencies for *PAH* Pathogenic Variants

	PAH c.782G>A	*PAH* c.842C>T	*PAH* c.898G>T	*PAH* c.1208C>T
Ashkenazi Jewish	3 / 10368 (0.3%)	0 / 10366 (0%)	71 / 10350 (0.7%)	53 / 10368 (0.5%)
European (non-Finnish)	46 / 129078 (0.04%)	26 / 128978 (0.02%)	58 / 128846 (0.05%)	96 / 129168 (0.07%)
European (Finnish)	0 / 25124 (0%)	2 / 25124 (0.008%)	0 / 25100 (0%)	0 / 25124 (0%)
Latino	4 / 35432 (0.01%)	0 / 35434 (0%)	3 / 35408 (0.008%)	8 / 35394 (0.02%)
South Asian	0 / 30616 (0%)	0 / 30616 (0%)	15 / 30608 (0.05%)	1 / 30614 (0.003%)
East Asian	2 / 19948 (0.01%)	0 / 19952 (0%)	0 / 19936 (0%)	0 / 19950 (0%)
African	1 / 24960 (0.004%)	1 / 24964 (0.004%)	0 / 24948 (0%)	3 / 24958 (0.01%)
Other	5 / 7222 (0.07%)	0 / 7218 (0%)	5 / 7214 (0.07%)	3 / 7228 (0.04%)
Total	61 / 282748 (0.02%)	2 / 251474 (0.0008%)	152 / 282410 (0.05%)	164 / 282804 (0.06%)

Table 5. GnomAD Allele Frequencies for *CFTR* and *PMM2* Pathogenic Variants

	CFTR c.3154T>G	*PMM2* c.367C>T	*PMM2* c.422G>A
Ashkenazi Jewish	21 / 10348 (0.2%)	0 / 9374 (0%)	71 / 9376 (0.8%)
European (non-Finnish)	117 / 128594 (0.09%)	3 / 93654 (0.003%)	511 / 94084 (0.5%)
European (Finnish)	6 / 25038 (0.02%)	0 / 20984 (0%)	177 / 21196 (0.8%)
Latino	22 / 35218 (0.06%)	3 / 30562 (0.01%)	64 / 30676 (0.2%)
South Asian	2 / 30532 (0.007%)	0 / 25432 (0%)	24 / 25492 (0.09%)
East Asian	0 / 19924 (0%)	0 / 16514 (0%)	2 / 16632 (0.01%)
African	0 / 24932 (0%)	0 / 20560 (0%)	15 / 20700 (0.07%)
Other	9 / 7174 (0.1%)	1 / 6152 (0.02%)	27 / 6220 (0.4%)
Total	177 / 281760 (0.06%)	7 / 223232 (0.003%)	891 / 224376 (0.4%)

III.II. *Expanded Carrier Screening Positive Rates*

Expanded carrier screening positive rates (ECS-PRs) were calculated for this cohort in order to estimate the proportion of individuals that would be identified as a carrier for at least one of the conditions on a commonly used expanded carrier screening panel in the clinical setting (Table 6). The ECS-PR for when all of the conditions on the ECS panel were included was 52% (95% CI: 37%-67%).

Table 6. Expanded carrier screening positive rates (ECS-PRs)

	One Condition	Two Conditions	At Least One Condition
N	14	9	23
%	**32%**	**20%**	**52%**

This study also calculated ECS-PRs when FMF was excluded from analyses (Table 7). The ECS-PR for when FMF was excluded was 48% (95% CI: 33%-63%).

Table 7. Expanded carrier screening positive rates (ECS-PRs) when FMF is excluded

	One Condition	Two Conditions	At Least One Condition
N	17	4	21
%	**39%**	**9%**	**48%**

IV. DISCUSSION

The two overarching aims of this study were to gather additional information about genetic variation in individuals of Armenian ancestry and to provide clinicians, such as genetic counselors, with a better understanding of considerations to make when counseling individuals of Armenian ancestry. Additional studies are highly recommended in order to increase breadth and generalizability of the results to the Armenian population as a whole.

IV.I. *Increasing Knowledge of Human Genetic Variation*

As the first study of its kind in the Armenian population, this project aimed to identify and categorize pathogenic and likely pathogenic variants found in the genes associated with 176 conditions that are commonly tested for in expanded carrier screening. While this is a small cohort study, this information will hopefully help build to the general knowledge about genetic variants observed in Armenians and humans in general.

Previous studies have shown that the carrier frequency of familial Mediterranean fever in Armenians is approximately 1 in 7 [Rogers et al., 1989]. The carrier frequency of 16% found in this cohort is comparable to reported frequencies in previous studies [Rogers et al., 1989; Ong et al., 2013].

The *MEFV* variants that were identified in the individuals in this cohort have all been reported in Armenian individuals in previous studies [Ong et al., 2013]. Three out of four of the pathogenic changes that were found, c.2080A>G, c.2177T>C, and c.2040G>C, are some of the most common *MEFV* changes seen in individuals of Mediterranean and Middle Eastern ancestry [Shohat and Halpern, 2011]. The c.2080A>G variant was seen in two of the individuals in this cohort, and it is seen most commonly in individuals of European (non-Finnish; NFE) and Latino

ancestries (GnomAD). More specifically, studies have shown that this variant is seen at a higher frequency in northern European, Iranian, and Azeri Turkish populations [Shohat and Halpern, 2011]. A retrospective study performed at UCLA in individuals with FMF, many of whom were Armenian-Americans, found a high frequency of c.2080A>G homozygotes in Armenians (12/34), as well as a significant number of c.2080A>G compound heterozygotes (17/34) [Ong et al., 2013]. This was significantly higher than the number of c.2080A>G carriers seen in other ethnicities. Individuals with the c.2080A>G variant have been found to be at a much higher risk for AA amyloidosis, and therefore a poorer clinical outcome. Homozygous individuals such as these would have likely been missed in this study during recruitment, since criteria for enrollment in the study required that the individuals be healthy and greater than 60 years old. Another *MEFV* variant, c.2177T>C was identified in three individuals in this cohort and is seen at a higher frequency in individuals of Ashkenazi Jewish ancestry (GnomAD). The c.2282G>A variant was seen at the highest frequency in individuals of East Asian ancestry, and the c.2040G>C variant was only reported in individuals of NFE ancestry (GnomAD). FMF patients with these three pathogenic variants have been shown to be less likely to develop AA amyloidosis, and may be associated with a milder clinical phenotype as compared to c.2080A>G [Ong et al., 2013]. In many cases, knowledge of a person's genotype is actually quite important for determining prognosis and prevention of life-threatening complications that can arise from the disease.

In this cohort, five individuals were carriers for the VUS, c.442G>C, which is also considered to be a founder variant in individuals of Mediterranean ancestry and has been seen in individuals who have FMF [Shohat and Halpern, 2011; Ong et al., 2013]. Ong et al. reported that the c.442G>C variant has also been seen in individuals of Asian ancestry, some of whom were

affected individuals in a homozygous or even heterozygous state. There is a single assertion submitted in ClinVar from the year 2000 which reports this variant as pathogenic when found in a haplotype block with another variant, c.2082G>A [Booth et al., 2000]. Conversely, many laboratories on ClinVar still classify this variant as benign or VUS. Therefore, the assertion provided by a single source was not sufficient for this variant to be included as pathogenic in this study. If this study had included c.442G>C as a pathogenic variant as in Ong et al., there would have been a total of 12 (27%) carriers of pathogenic *MEFV* variants instead of seven (16%). Although this study did not include the c.442G>C variant as a pathogenic variant, some studies continue to assert that it may play a role in disease development [Ong et al., 2013; Booth et al., 2000]. Thus, this variant could benefit from further analyses in order to determine a consensus regarding its classification.

In addition to the four *MEFV* variants, population frequencies were assessed for the variants that were associated with other conditions in which there were multiple carriers identified in this cohort. The *CFTR* variant that was identified in two individuals of this cohort, c.3154T>G, is seen at the highest frequencies in individuals of Ashkenazi Jewish and NFE ancestries. Of the two *PMM2* variants, the c.367C>T variant is more rare than the c.422G>A variant. The c.422G>A variant is seen at higher frequencies in European (Finnish), Ashkenazi Jewish, and NFE ancestries. Three of the four *PAH* variants seen in this cohort were seen at the highest frequencies in individuals of Ashkenazi Jewish and NFE ancestries.

It is known that the incidence of PKU varies greatly across different populations (1:4,000 to 1:2,600 in Turkey, 1:4,500 in Ireland, 1:10,000 in Northern Europeans, 1:200,000 in Finland) [Kostandyan et al., 2011]. However, as was noted in Kostandyan et al., the incidence among Armenians is unknown. Their study estimated the incidence could be 1 in 8,300, which would

correspond to a carrier frequency of 1 in 46 (2.2%), although they did believe this was an underestimate [Kostandyan et al., 2011]. In the current study, four out of 44 individuals (9%) were found to be carriers for PKU. This frequency is quite high compared to reported frequencies in other populations, such as in Turkish individuals who have one of the highest reported carrier frequencies (1 in 26; 3.8%) for the condition [Özalp et al., 2001; Kostandyan et al., 2011]. However, given the small sample size, the 95% confidence interval is wide, and this may be an overestimate of the true carrier frequency. Regardless, this finding does add to what limited research has been done in this population to date. Three of the four *PAH* pathogenic variants seen in this cohort, c.782G>A, c.842C>T, and c.1208C>T, were also observed in Kostandyan et al., and the c.782G>A and c.842C>T were two of the most common variants seen. The fourth variant present in this cohort, c. 898G>T, was not seen in the previous study. Perhaps the greatest benefit of these analyses is the impact that this information could have on patient care in the clinical setting. If the carrier frequency for PKU in Armenians in the general population is as high as what was seen in this cohort, it will be important to continue to expand public health initiatives in Armenia. Knowledge of the carrier frequency may help to motivate health professionals. Armenia initiated its neonatal screening program for PKU on April 15, 2008 [Kostandyan et al., 2011]. However, there are no publications that highlight how uniformly neonatal screening is carried out in Armenia. As medical advancements in the country continue to be made, particularly outside of major metropolitan areas, it will be of the utmost importance to grow public health initiatives and to improve the overall quality of medical care for individuals with rare genetic disorders. In addition to newborn screening, it will be important to also begin offering expanded carrier screening to patients in order to provide parents with more options regarding family planning and the health of their children.

IV.II. *Utility of Expanded Carrier Screening Across Populations*

This study explored the value of expanded carrier screening in an underserved population. While many individuals in this cohort would have been identified as FMF carriers had they had ECS performed clinically, it is noteworthy that a great many more were identified as carriers for conditions other than FMF. Only two of 23 carrier individuals were carriers for FMF alone. Multiple studies have been completed in individuals of different ethnicities and populations that have shown a wide range of results regarding the percentage of individuals that are identified as a carrier after undergoing ECS. Two previous studies performed in separate ethnically diverse populations have shown that approximately 25-30% of individuals who underwent expanded carrier screening were identified as a carrier for at least one condition [Peyser et al., 2018; Lazarin et al., 2013]. However, both of these studies utilized targeted genotyping of about 400 common variants associated with approximately 100 genetic conditions [Peyser et al., 2018; Lazarin et al., 2013]. Another large study in Barcelona that performed targeted genotyping as well as NGS on a portion of the genes in the ECS panel reported that 56% of the individuals in their cohort were carriers for at least one condition [Abulí et al., 2016]. Estimates of the yield of ECS have been reported to be as high as 84% in another Spanish study, which was completed in a population of gamete donors and couples seeking assisted reproductive technologies (ART) in Valencia [Martin et al., 2015]. The testing methodology that was used in Martin et al. involved NGS for 549 autosomal and X-linked recessive genes, making it the largest ECS screen of these four studies [Martin et al., 2015]. Therefore, studies have shown a large range in the approximations of the yield of different ECS panels. The ECS-PRs calculated in this study are within the range of those reported in the literature, given the large

confidence intervals on this estimate. [Peyser et al., 2018; Lazarin et al., 2013; Abulí et al., 2016; Martin et al., 2015]. Additionally, the ECS-PR (52%) in this study was not only due to the high frequency of FMF in this population, since the ECS-PR was nearly as high (48%) when FMF was not included in the calculation. However, we recognize that the small cohort size may affect the accuracy of the ECS-PRs, as demonstrated by the wide confidence intervals.

These findings provide support for the utility of expanded carrier screening in Armenian individuals, since indeed many of the individuals whose samples were sequenced as part of this study were identified to be carriers for a variety of different conditions. There were three conditions other than FMF in which there were multiple individuals identified as carriers: cystic fibrosis, PKU, and congenital disorder of glycosylation, type 1a. If these individuals had only been tested for FMF based on their high-risk ethnicity status, only seven individuals would have been identified as carriers, resulting in 16 missed carriers that would have been identified by ECS.

The results of this study may encourage the growth and implementation of expanded carrier screening in Armenia, since currently expanded carrier screening is not routinely offered to individuals in Armenia. When designing expanded carrier screening panels, it would be prudent for laboratories to strive to have the highest detection rate for conditions, rather than having more conditions, in order to provide patients with the highest possible negative predictive values (NPV) [Beauchamp et al., 2018]. While some experts in the field have suggested that FMF not be included on ECS panels since it does not typically have a severe outcome in affected individuals, carrier testing for FMF can be worthwhile in a number of Non-European ancestries, such as Armenians, Iranians, and Northern-African Jews, since the condition can be severe in some cases and has management guidelines for treatment of affected individuals [Romero et al.,

2017; Shohat and Halpern, 2011]. These ethnicities have a higher carrier frequency for FMF than many other conditions that are included on nearly all ECS panels, such as cystic fibrosis [Shohat and Halpern, 2011; Rink et al., 2017]. As individuals from a variety of ethnic backgrounds consider having expanded carrier screening, it will be important for genetic counselors and other medical professionals to discuss the specific conditions on an ECS panel that have the highest *a priori* risk for any given individual, whether based on ethnicity or family history.

IV.III. *Genetic Counseling Considerations*

Genetic counseling is a crucial component in the process of obtaining informed consent and ordering genetic testing, such as an ECS panel. It is very important for genetic counselors to have an understanding about their clients' cultural background, in addition to their genetic risks, in order to build good rapport with and instill trust in their patients. As an ethnic population that experienced great trauma in the early 20th century and has taken over a century to recover from these events, it is the job of the genetic counselor to be aware of this history and to be cautious with how genetic risk information is presented. As has been seen with other ethnicities that have undergone such events, many individuals in the population may be particularly sensitive regarding any possible genetic differences they may have, even though it is known that individuals of all ancestries carry pathogenic genetic variants. Special care should be taken to emphasize that all individuals are carriers for certain genetic conditions, and that individuals have no choice regarding the genetic information that they inherit or that they pass down to their children. As always, it is also important to stress the possibility of variants of uncertain significance, particularly since Armenians remain an understudied population. It would be helpful to provide Armenian individuals with an estimate of the chance that they will be found to

be a carrier for at least one condition on an ECS panel, such as the one used for gene selection in this study. Although the exact ECS-PR (52%) is only an estimate based on a small sample of Armenian individuals, genetic counselors and other health professionals may consider using an estimate based on previous studies in other populations, and noting that the chance of receiving a positive result may be a little higher in the Armenian population.

IV.IV. Limitations

The following limitations were present in this study. First, given the small cohort size, one cannot make generalizations regarding carrier frequencies for the majority of the conditions screened. This study methodology may have under-sampled homozygotes, and so this sample would be expected to result in an underestimate of the frequency of pathogenic and likely pathogenic alleles, although this would not be expected to significantly impact the carrier frequencies for rare conditions since homozygotes are rare in the population. However, for a common condition like FMF, the under-sampling of homozygotes could theoretically impact estimates of the carrier frequency.

The type of testing performed on the individuals in this study presented a different set of limitations. Although exome sequencing provided a wide breadth of coverage for genes to study, it would not have detected all of the types of variants that may have been picked up on a specially designed expanded carrier screen. The majority of the genes selected for this study are screened using full gene sequencing and copy number analysis. While this exome sequencing assay was able to detect sequencing changes as well as small insertions and deletions, it could not detect large copy number changes. Additionally, some of the genes on the panel require additional testing methodologies in order to detect variants commonly seen in the condition and

avoid calling errors, as described in Beauchamp et al. [Beauchamp et al., 2018]. A list of all of the conditions in the ECS panel is included in Appendix B, which also indicates whether a gene typically requires supplemental analyses for specific variant detection beyond standard NGS. For example, analyses performed for alpha thalassemia include analysis of homologous regions and 13 common variants, including deletions that are seen more commonly in individuals of Asian ancestry. Therefore, exome sequencing would have had a very low yield for pathogenic changes related to alpha thalassemia, since sequence changes are not the most common cause for the condition. Spinal muscular atrophy is another such condition in which targeted carrier screening utilizes non-NGS technology. For this condition, the primary screening assay is for copy number variants, such as deletions, which also would not have been detected by exome sequencing. Lastly, some of the genes on the panel such as *CYP21A2*, which is associated with congenital adrenal hyperplasia, are notoriously difficult to sequence due to obstacles such as pseudogenes [Xu et al., 2013; Beauchamp et al., 2018]. A clinically utilized expanded carrier screen would likely account for these difficulties and would perform further analyses for completion and to ensure proper coverage; however, this exome sequencing assay did not. Indeed, at least one *CYP21A2* variant was identified in this cohort that was reported as pathogenic on ClinVar, but had consistently low genotype quality scores in the individuals sequenced in this study, possibly due to pseudogene interference, and was therefore omitted from the results.

IV.V. Conclusions and Recommendations for Future Studies

This study reported information regarding pathogenic, likely pathogenic, and other variants seen in this population, which can provide a starting point for additional studies. Further analysis of the variants identified in this cohort that did not have any ClinVar annotation may be

valuable. It may also be instructive to compare what the anticipated performance would be for a larger ECS panel as compared to the one used in this study. Variation in the 59 genes currently recommended by the ACMG for reporting of secondary findings [Kalia et al., 2017] could also be evaluated for the Armenian population, although this particular sample would not be the best cohort for such a study. It may be less informative to study autosomal dominant conditions in a group of healthy, older individuals since individuals with high and moderately penetrant pathogenic variants may have a reduced life-span or not be considered healthy.

If the ArmGenia project or another research group were able to collect additional samples of affected individuals for various conditions, such as cancer, then this would allow for further investigation of various genetic conditions in Armenian individuals. In general, many more studies need to be done in Armenian individuals in order to provide greater insights for public health initiatives both in the U.S., where there is a relatively large Armenian population, and in Armenia, where public health is beginning to make new and important advancements.

This study reaffirms that there are other conditions in addition to FMF that Armenian individuals should be screened for in order to provide accurate reproductive risks to couples. This study provides evidence in support of expanded carrier screening in Armenian individuals as compared to targeted testing for FMF only. Results of this study may help to improve genetic counseling provided to Armenian individuals, particularly in prenatal and pre-conceptional settings, and promote diversity and equity in the field of genetics to improve the overall quality of services for providers and patients alike.

REFERENCES

Abulí A, Boada M, Rodríguez-Santiago B, Coroleu B, Veiga A, Armengol L, Barri PN, Pérez-Jurado LA, Estivill X. 2016. NGS-Based Assay for the Identification of Individuals Carrying Recessive Genetic Mutations in Reproductive Medicine. Hum. Mutat.

Bajaj K, Gross S. 2014. Carrier Screening: Past, Present, and Future. J. Clin. Med. 3: 1033–1042.

Beauchamp KA, Muzzey D, Wong KK, Hogan GJ, Karimi K, Candille SI, Mehta N, Mar-Heyming R, Kaseniit KE, Kang HP, Evans EA, Goldberg JD, Lazarin GA, Haque IS. 2018. Systematic design and comparison of expanded carrier screening panels. Genet. Med. 20: 55–63.

Ben-Shachar R, Svenson A, Goldberg JD, Muzzey D. 2019. A data-driven evaluation of the size and content of expanded carrier screening panels. Genet. Med. 0: 1–9.

Booth DR, Gillmore JD, Lachmann HJ, Booth SE, Bybee A, Soytürk M, Akar S, Pepys MB, Tunca M, Hawkins PN. 2000. The genetic basis of autosomal dominant familial Mediterranean fever. QJM.

Chaloyan A. 2017. Historical Background of Armenian Migration, and Formation of the Armenian Diaspora. Manifestations of the Strategic Nationalism BT - Fluctuating Transnationalism: Social Formation and Reproduction among Armenians in Germany. In: Chaloyan A, editor. Wiesbaden: Springer Fachmedien Wiesbaden, p 63–98.

Chen L, Liu P, Evans Jr. TC, Ettwiller LM. 2017. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. Science (80-.). 355: 752–756.

Djourabchi F. UCLA-ARMGENIA MOU Signing: Mapping of the Armenian Genome. UCLA Cent. World Heal.

Edwards JG, Feldman G, Goldberg J, Gregg AR, Norton ME, Rose NC, Schneider A, Stoll K, Wapner R, Watson MS. 2015. Expanded carrier screening in reproductive medicine-points to consider. Obstet. Gynecol. 125: 653–662.

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MMY, Tsui SKW, Xue H, Wong JTF, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD,

Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, You QS, Tam PKH, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449: 851–861.

Fulda KG, Lykens K. 2006. Ethical issues in predictive genetic testing: A public health perspective. J. Med. Ethics 32: 143–147.

Gibbs RA, Boerwinkle E, Doddapaneni H, Han Y, Korchina V, Kovar C, Lee S, Muzny D, Reid JG, Zhu Y, Wang J, Chang Y, Feng Q, Fang X, Guo X, Jian M, Jiang H, Jin X, Lan T, Li G, Li J, Li Y, Liu S, Liu X, Lu Y, Ma X, Tang M, Wang B, Wang G, Wu H, Wu R, Xu X, Yin Y, Zhang D, Zhang W, Zhao J, Zhao M, Zheng X, Lander ES, Altshuler DM, Gabriel SB, Gupta N, Gharani N, Toji LH, Gerry NP, Resch AM, Flicek P, Barker J, Clarke L, Gil L, Hunt SE, Kelman G, Kulesha E, Leinonen R, McLaren WM, Radhakrishnan R, Roa A, Smirnov D, Smith RE, Streeter I, Thormann A, Toneva I, Vaughan B, Zheng-Bradley X, Bentley DR, Grocock R, Humphray S, James T, Kingsbury Z, Lehrach H, Sudbrak R, Albrecht MW, Amstislavskiy VS, Borodina TA, Lienhard M, Mertes F, Sultan M, Timmermann B, Yaspo M-L, Mardis ER, Wilson RK, Fulton L, Fulton R, Sherry ST, Ananiev V, Belaia Z, Beloslyudtsev D, Bouk N, Chen C, Church D, Cohen R, Cook C, Garner J, Hefferon T, Kimelman M, Liu C, Lopez J, Meric P, et al. 2015. A global reference for human genetic variation. Nature 526: 68–74.

Grody WW, Thompson BH, Gregg AR, Bean LH, Monaghan KG, Schneider A, Lebo R V. 2013. ACMG position statement on prenatal/preconception expanded carrier screening. Genet. Med. 15: 482–483.

Haber M, Mezzavilla M, Xue Y, Comas D, Gasparini P, Zalloua P, Tyler-Smith C. 2015. Genetic evidence for an origin of the Armenians from Bronze Age mixing of multiple populations. Eur. J. Hum. Genet. 24: 931.

Hatzikotoulas K, Gilly A, Zeggini E. 2014. Using population isolates in genetic association studies. Brief. Funct. Genomics 13: 371–377.

Henneman L, Borry P, Chokoshvili D, Cornel MC, van El CG, Forzano F, Hall A, Howard HC, Janssens S, Kayserili H, Lakeman P, Lucassen A, Metcalfe SA, Vidmar L, de Wert G, Dondorp WJ, Peterlin B. 2016. Responsible implementation of expanded carrier screening. Eur. J. Hum. Genet. 24: e1–e12.

International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. Nature 431: 931–945.

Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, McKelvey KD, Ormond KE, Richards CS, Vlangos CN, Watson M, Martin CL, Miller DT. 2017. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): A policy statement of the American College of Medical Genetics and Genomics. Genet. Med. 19: 249–255.

Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, Massouras A. 2018. VarSome: the human genomic variant search engine. Bioinformatics: 1–3.

Kostandyan N, Britschgi C, Matevosyan A, Oganezova A, Davtyan A, Blau N, Steinmann B, Thöny B. 2011. The spectrum of phenylketonuria genotypes in the Armenian population: Identification of three novel mutant PAH alleles. Mol. Genet. Metab. 104: 93–97.

Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR. 2016. ClinVar: Public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 44: D862–D868.

Lazarin GA, Haque IS, Nazareth S, Iori K, Patterson AS, Jacobson JL, Marshall JR, Seltzer WK, Patrizio P, Evans EA, Srinivasan BS. 2013. An empirical estimate of carrier frequencies for 400+ causal Mendelian variants: Results from an ethnically diverse clinical sample of 23,453 individuals. Genet. Med. 15: 178–186.

Marek-Yagel D, Berkun Y, Padeh S, Abu A, Reznik-Wolf H, Livneh A, Pras M, Pras E. 2009. Clinical disease among patients heterozygous for familial Mediterranean fever. Arthritis Rheum. 60: 1862–1866.

Martin J, Asan, Yi Y, Alberola T, Rodríguez-Iglesias B, Jiménez-Almazán J, Li Q, Du H, Alama P, Ruiz A, Bosch E, Garrido N, Simon C. 2015. Comprehensive carrier genetic test using next-generation deoxyribonucleic acid sequencing in infertile couples wishing to conceive through assisted reproductive technology. Fertil. Steril.

Mittman IS, Downs K. 2008. Diversity in genetic counseling: Past, present and future. J. Genet. Couns. 17: 301–313.

Ong F, Vakil H, Xue Y, Kuo J, Shah K, Lee R, Bernstein K, Rimoin D, Getzug T, Das K, Deignan J, Rotter J, Grody WW. 2013. The M694V mutation in Armenian-Americans: A 10-year retrospective study of MEFV mutation testing for familial Mediterranean fever at UCLA. Clin. Genet. 84: 55–59.

Özalp I, Coşkun T, Tokatli A, Kalkanoğlu HS, Dursun A, Tokol S, Köksal G, Özgüç M, Köse R. 2001. Newborn PKU screening in Turkey: At present and organization for future. Turk. J. Pediatr.

Peyser A, Singer T, Mullin C, Bristow SL, Gamma A, Onel K, Hershlag A. 2018. Comparing ethnicity-based and expanded carrier screening methods at a single fertility center reveals significant differences in carrier rates and carrier couple rates. Genet. Med. 0.

Pray LA. 2008. Discovery of DNA Structure and Function: Watson and Crick. Nat. Educ. 1: 100.

Pruitt KD, Tatusova T, Brown GR, Maglott DR. 2012. NCBI Reference Sequences (RefSeq): Current status, new features and genome annotation policy. Nucleic Acids Res. 40: 130–

135.

Resta R, Biesecker BB, Bennett RL, Blum S, Hahn SE, Strecker MN, Williams JL. 2006. A New Definition of Genetic Counseling : National Society of Genetic Counselors ' Task Force Report The National Society of Genetic Counselors ' Definition Task Force : 15.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. 2015. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. 17: 405–424.

Rink B, Jr JRB, Jr DNS. 2017. Carrier screening in the age of genomic medicine. Obstet. Gynecol.

Rogers DB, Shohat M, Petersen GM, Bickal J, Congleton J, Schwabe AD, Rotter JI. 1989. Familial Mediterranean fever in Armenians: Autosomal recessive inheritance with high gene frequency. Am. J. Med. Genet.

Romero S, Jr JRB, Saller DN, Giardine R. 2017. ACOG 2017 Carrier screening for genetic conditions. Obstet. Gynecol. 129: e41–e55.

Rosenfeld JA, Mason CE, Smith TM. 2012. Limitations of the human reference genome for personalized genomics. PLoS One 7: 1–10.

Sarkisian T, Ajrapetian H, Beglarian A, Shahsuvarian G, Egiazarian A. 2008. Familial Mediterranean Fever in Armenian population. Georgian Med. News: 105—111.

Shohat M, Halpern GJ. 2011. Familial Mediterranean fever—A review. Genet. Med. 13: 487–498.

U.S. Census Bureau. 2012. The Two or More Races Population: 2010 (2010 Census Briefs, No. C2010BR-13). U.S. Census Bur.: 1–23.

What a long, strange trip it's been. . . 2001. Nature 409: 756.

Xu Z, Chen W, Merke DP, McDonnell NB. 2013. Comprehensive mutation analysis of the CYP21A2 gene: An efficient multistep approach to the molecular diagnosis of congenital adrenal hyperplasia. J. Mol. Diagnostics 15: 745–753.

APPENDIX A: List of All Other Variants Called in this Cohort in the VCF Files

Gene	Variant	Gene	Variant
ABCC8	c.4714G>A	LAMA2	c.8124T>A
ABCC8	c.4105G>T	LAMA3	c.2225G>A
ABCC8	c.3819G>A	LAMA3	c.3980G>A
ABCC8	c.2485C>T	LAMA3	c.1217G>A
ABCC8	c.2277C>T	LAMA3	c.373C>T
ABCC8	c.1947G>A	LAMA3	c.661A>G
ABCC8	c.330C>T	LAMA3	c.2764G>A
ABCC8	c.207T>C	LAMA3	c.2809C>T
ABCD1	c.2019C>T	LAMA3	c.342T>C
ACADM	c.127G>A	LAMA3	c.695G>A
ACADM	c.416A>G	LAMA3	c.1806G>A
ACADM	c.286+178G>A	LAMA3	c.2387C>A
ACADM	c.1161A>G	LAMA3	c.2901T>C
ACADS	c.1108A>G	LAMA3	c.3066C>G
ACADS	c.321T>C	LAMA3	c.3617T>C
ACADS	c.360C>T	LAMA3	c.3622C>A
ACADS	c.625-99T>C	LAMA3	c.4110G>A
ACADS	c.625G>A	LAMA3	c.4260G>C
ACADS	c.990C>T	LAMA3	c.4530C>T
ACADVL	c.99C>A	LAMA3	c.110T>G
ACADVL	c.957G>A	LAMA3	c.495C>T
ACADVL	c.1824C>T	LAMA3	c.1320C>G
ADA	c.508A>G	LAMA3	c.3618C>A
ADA	c.534A>G	LAMA3	c.3673A>G
ADA	c.390G>A	LAMA3	c.3904C>T
ADA	c.239A>G	LAMB3	c.3404G>A
ADA	c.36G>A	LAMB3	c.773G>A
AGA	c.34G>T	LAMB3	c.1873C>T
AGA	c.446C>G	LAMB3	c.927C>T
AGL	c.686A>G	LAMB3	c.417C>T
AGL	c.461G>A	LAMB3	c.3432A>G
AGL	c.4331A>G	LAMB3	c.2777C>A
AGL	c.4575T>A	LAMB3	c.2673A>G
AGL	c.894C>T	LAMB3	c.2554A>T
AGL	c.3199C>T	LAMB3	c.2124T>C
AGL	c.3343G>A	LAMB3	c.2069A>G
AGXT	c.332G>A	LAMB3	c.1764T>C
AGXT	c.1030G>A	LAMB3	c.1716T>C
AGXT	c.32C>T	LAMB3	c.1579G>A
AGXT	c.264C>T	LAMB3	c.1149G>A
AGXT	c.654G>A	LAMB3	c.1051G>A

AGXT	c.1020A>G	LAMB3	c.1015T>C
AIRE	c.329G>A	LAMB3	c.541A>G
AIRE	c.588C>T	LAMB3	c.384C>T
AIRE	c.681C>T	LAMB3	c.291A>C
AIRE	c.834C>G	LAMB3	c.138C>T
AIRE	c.1197T>C	LAMC2	c.566G>T
AIRE	c.1578T>C	LAMC2	c.1899G>C
ALDH3A2	c.328G>A	LAMC2	c.2387C>T
ALDH3A2	c.1446A>T	LAMC2	c.344G>A
ALG6	c.482A>G	LAMC2	c.3448C>T
ALG6	c.727G>A	LAMC2	c.3546A>G
ALG6	c.1323T>C	LAMC2	c.297C>T
ALG6	c.391T>C	LAMC2	c.371C>T
ALG6	c.911C>T	LAMC2	c.483C>T
ALMS1	c.281C>T	LAMC2	c.741C>A
ALMS1	c.1868A>G	LAMC2	c.798T>G
ALMS1	c.3392A>G	LAMC2	c.2198G>C
ALMS1	c.7089G>A	LAMC2	c.2688G>A
ALMS1	c.9457A>T	LIPA	c.67G>A
ALMS1	c.12001C>T	LIPA	c.46A>C
ALMS1	c.1267G>A	LRFN4	c.960G>A
ALMS1	c.1571_1573dupCTC	LRFN4	c.1019T>C
ALMS1	c.4948G>A	LRPPRC	c.2481A>G
ALMS1	c.4993T>G	LRPPRC	c.2562A>G
ALMS1	c.10863A>G	LRPPRC	c.1068A>G
ALMS1	c.1174C>T	LRPPRC	c.246G>A
ALMS1	c.2012T>G	MAN2B1	c.1521G>T
ALMS1	c.2187C>T	MAN2B1	c.1068C>G
ALMS1	c.4176A>G	MAN2B1	c.922C>T
ALMS1	c.4241G>C	MAN2B1	c.2310C>T
ALMS1	c.5623A>G	MAN2B1	c.1437G>T
ALMS1	c.6209T>C	MAN2B1	c.1441G>T
ALMS1	c.6299C>T	MAN2B1	c.1238A>G
ALMS1	c.6333T>A	MAN2B1	c.1010G>A
ALMS1	c.6851G>C	MAN2B1	c.935C>T
ALMS1	c.7721G>A	MAN2B1	c.832C>G
ALMS1	c.8014G>C	MCOLN1	c.966A>C
ALMS1	c.8478G>T	MCOLN1	c.984C>T
ALMS1	c.8567A>G	MEFV	c.2118G>A
ALMS1	c.9558C>T	MEFV	c.1760-5C>T
ALMS1	c.12086G>A	MEFV	c.1518C>T
ALMS1	c.12172C>T	MEFV	c.1421_1422insACA
ALPL	c.455G>A	MEFV	c.443A>T
ALPL	c.787T>C	MEFV	c.1764G>A

ALPL	c.876A>G	MEFV	c.1760-30T>A
ALPL	c.1565T>C	MEFV	c.1530T>C
AMT	c.354G>A	MEFV	c.1503C>T
AMT	c.954G>A	MEFV	c.1428A>G
AMT	c.631G>A	MEFV	c.1422G>A
ARG1,MED23	c.270C>T	MEFV	c.1223G>A
ARSA	c.1442G>A	MEFV	c.1105C>T
ARSA	c.1149C>T	MEFV	c.942C>T
ARSA	c.636C>T	MEFV	c.442G>C
ARSA	c.1493G>A	MEFV	c.414A>G
ARSA	c.1178C>G	MEFV	c.306T>C
ARSA	c.1055A>G	MESP2	c.803A>T
ARSA	c.1002C>T	MKS1	c.1349T>C
ARSA	c.585G>T	MKS1	c.404C>A
ARSA	c.459C>T	MKS1	c.1671G>C
ASL	c.392C>T	MLC1	c.279G>A
ASPA	c.230A>G	MLC1	c.597A>G
ASPA	c.693C>T	MLC1	c.594C>T
ASS1	c.1211G>A	MLC1	c.512G>T
ASS1	c.501C>T	MMAA	c.747G>A
ASS1	c.876T>C	MMAA	c.1089G>C
ATM	c.2119T>C	MMAB	c.398C>T
ATM	c.3277A>G	MMAB	c.716T>A
ATM	c.4101C>T	MMACHC	c.641G>A
ATM	c.1744T>C	MMACHC	c.78C>A
ATM	c.3161C>G	MMACHC	c.182G>A
ATM	c.4578C>T	MMACHC	c.766_771delGCCCCC
ATM	c.5557G>A	MMACHC	c.321G>A
ATP7A	c.2205T>A	MPI	c.684C>T
ATP7A	c.4356G>C	MPI	c.1131A>G
ATP7A	c.2299G>C	MUT	c.1629C>T
AXDND1	c.3032-3141G>A	MUT	c.1342G>A
BBS1	c.24T>C	MUT	c.2011A>G
BBS1	c.1634A>G	MUT	c.1991C>T
BBS1	c.378G>A	MUT	c.1595G>A
BBS10	c.1970C>T	MUT	c.1495G>A
BBS10	c.1631A>G	MUT	c.636G>A
BBS10	c.1616C>T	MYO7A	c.1606G>A
BBS12	c.69G>A	MYO7A	c.1872G>A
BBS12	c.116T>C	MYO7A	c.2886G>C
BBS12	c.1782A>G	MYO7A	c.4074C>T
BBS12	c.1157G>A	MYO7A	c.4742C>G
BBS12	c.1380G>C	MYO7A	c.6203T>G
BBS12	c.1398C>T	MYO7A	c.6246C>T

BBS12	c.1399G>A	MYO7A	c.895G>A
BBS12	c.1410C>T	MYO7A	c.4577G>A
BBS12	c.1872A>G	MYO7A	c.4697C>T
BBS2	c.98C>A	MYO7A	c.5487C>T
BBS2	c.740T>G	MYO7A	c.47T>C
BBS2	c.566G>A	MYO7A	c.288G>A
BBS2	c.367A>G	MYO7A	c.468C>T
BBS2	c.209G>A	MYO7A	c.510G>A
BCKDHA	c.34C>A	MYO7A	c.783T>C
BCKDHA	c.116C>A	MYO7A	c.4755C>T
BCKDHA	c.452C>T	MYO7A	c.4996A>T
BCKDHA	c.972C>T	MYO7A	c.5156A>G
BCKDHA	c.1221A>G	MYO7A	c.5715A>G
BCKDHB	c.1069A>G	MYO7A	c.6240C>T
BCKDHB	c.987C>T	MYO7A	c.6318G>A
BCS1L	c.64C>T	MYO7A	c.6519C>T
BCS1L	c.201C>T	NAGLU	c.2209C>G
BCS1L	c.628G>A	NAGLU	c.2209C>G
BCS1L	c.996C>T	NAGLU	c.1435G>A
BCS1L	c.1017T>C	NAGLU	c.2157G>A
BLM	c.2268A>G	NAGLU	c.423T>C
BLM	c.2119C>T	NAGLU	:c.2209C>A
BLM	c.1122T>C	NBN	c.2016A>G
BLM	c.1722A>G	NBN	c.1197T>C
BLM	c.2603C>T	NBN	c.102G>A
BLM	c.3102G>A	NEB	c.24220G>T
BLM	c.3531C>A	NEB	c.22222C>T
BLM	c.3945C>T	NEB	c.22068A>G
BLM	c.3961G>A	NEB	c.21487T>C
CAPN3	c.1290A>G	NEB	c.21044C>G
CAPN3	c.1813G>C	NEB	c.19705A>C
CAPN3	c.96T>C	NEB	c.18555G>A
CAPN3	c.318C>T	NEB	c.18431A>G
CAPN3	c.606T>C	NEB	c.9346G>A
CAPN3	c.706G>A	NEB	c.8519A>G
CBS	c.133C>T	NEB	c.6159G>A
CBS	c.833T>C	NEB	c.5987C>A
CBS	c.699C>T	NEB	c.5680A>G
CDKL5	c.2714-1385A>G	NEB	c.3416C>T
CFTR	c.224G>A	NEB	c.877C>T
CFTR	c.1516A>G	NEB	c.25472C>T
CFTR	c.2991G>C	NEB	c.22500C>T
CFTR	c.1001G>A	NEB	c.22075A>C
CFTR	c.1163C>T	NEB	c.21825C>T

CFTR	c.1251C>A	NEB	c.20078C>T
CFTR	c.3897A>G	NEB	c.11729A>G
CFTR	c.4025G>T	NEB	c.11148G>C
CFTR	c.443T>C	NEB	c.10005G>A
CFTR	c.1408G>A	NEB	c.9467T>A
CFTR	c.1727G>C	NEB	c.7343G>A
CFTR	c.2002C>T	NEB	c.5633T>A
CFTR	c.2562T>G	NEB	c.5420C>T
CFTR	c.3870A>G	NEB	c.3623T>C
CFTR	c.4389G>A	NEB	c.581A>G
CLN3	c.1158C>T	NEB	c.194C>T
CLN3	c.1211A>G	NEB	c.25500T>G
CLN3	c.242C>T	NEB	c.25309A>G
CLN5	c.726C>A	NEB	c.25140C>T
CLN5	c.528T>G	NEB	c.24538G>C
CLN5	c.1103A>G	NEB	c.21690A>G
CLRN1	c.57A>T	NEB	c.21341G>A
COL4A3	c.3047G>A	NEB	c.20766C>T
COL4A3	c.222G>A	NEB	c.20598C>G
COL4A3	c.1668G>A	NEB	c.20192A>T
COL4A3	c.2501A>G	NEB	c.18305G>C
COL4A3	c.127G>C	NEB	c.18294T>C
COL4A3	c.422T>C	NEB	c.16911A>G
COL4A3	c.1195C>T	NEB	c.16762T>A
COL4A3	c.1223G>A	NEB	c.16544A>C
COL4A3	c.1352A>G	NEB	c.16542C>G
COL4A3	c.1452G>A	NEB	c.10809G>C
COL4A3	c.1721C>T	NEB	c.10707G>A
COL4A3	c.3258G>A	NEB	c.9363T>C
COL4A3	c.3807C>A	NEB	c.9124T>C
COL4A3	c.4484A>G	NEB	c.8734T>C
COL4A4	c.4207T>C	NEB	c.8466C>T
COL4A4	c.3684G>A	NEB	c.8318G>A
COL4A4	c.680G>A	NEB	c.7839G>C
COL4A4	c.4932C>T	NEB	c.5971C>T
COL4A4	c.4080G>A	NEB	c.4471G>A
COL4A4	c.3011C>T	NEB	c.4435G>A
COL4A4	c.2996G>A	NEB	c.4407G>C
COL4A4	c.1634G>C	NEB	c.3901T>C
COL4A4	c.1444C>T	NEB	c.3593A>G
COL4A5	c.1992G>T	NEB	c.3081A>T
CPS1	c.3643A>G	NEB	c.771T>C
CPS1	c.3683G>A	NEB	c.571G>C
CPS1	c.4252C>G	NPC1	c.3797G>A

CPS1	c.-4_-3insTTC	NPC1	c.2793C>T
CPS1	c.1030A>G	NPC1	c.2572A>G
CPS1	c.1032C>T	NPC1	c.2073G>A
CPS1	c.2448C>T	NPC1	c.1926G>C
CPS1	c.2679C>G	NPC1	c.387T>C
CPS1	c.4217C>A	NPC2	c.88G>A
CPT1A	c.2235G>A	NPHS1	c.1596G>A
CPT1A	c.1296G>A	NPHS1	c.658T>G
CPT1A	c.266G>A	NPHS1	c.563A>T
CPT1A	c.1251T>C	NPHS1	c.2970C>T
CPT1A	c.963G>A	NPHS1	c.1110T>C
CPT1A	c.823G>A	NPHS1	c.881C>T
CPT2	c.1025T>C	NPHS1	c.3315G>A
CPT2	c.236A>C	NPHS1	c.2971G>C
CPT2	c.1806T>C	NPHS1	c.2289C>T
CPT2	c.1102G>A	NPHS1	c.1320C>T
CPT2	c.1939A>G	NPHS1	c.1223G>A
CTNS	c.154G>A	NPHS1	c.349G>A
CTNS	c.504G>A	NPHS2	c.87C>G
CTNS	c.779C>T	NPHS2	c.686G>A
CTNS	c.*300C>G	NR0B1	c.498G>A
CTSK	c.831A>G	NR0B1	c.114C>T
CYP11B1	c.394C>T	OPA3	c.231T>C
CYP11B1	c.259G>C	OTC	c.137A>G
CYP11B1	c.1086G>C	OTC	c.809A>G
CYP11B1	c.873G>A	PAH	c.609C>T
CYP11B1	c.246C>T	PAH	c.397A>C
CYP11B1	c.225A>G	PAH	c.1242C>T
CYP11B1	c.128G>A	PAH	c.529G>A
CYP21A2	c.983C>T	PAH	c.1155C>G
CYP21A2	c.318G>C	PAH	c.735G>A
CYP21A2	c.308G>A	PAH	c.696G>A
CYP27A1	c.674G>A	PC	c.3029C>T
CYP27A1	c.888A>G	PC	c.1317C>T
CYP27A1	c.1151C>T	PC	c.2619C>T
DBT	c.1150A>G	PCCA	c.762C>T
DHCR7	c.1272C>T	PCCA	c.604G>A
DHCR7	c.1158T>C	PCCA	c.1342A>G
DHCR7	c.438T>C	PCCA	c.627A>G
DHCR7	c.231C>T	PCCA	c.1423A>G
DHCR7	c.207T>C	PCCA	c.1651G>T
DHCR7	c.189G>A	PCCB	c.815G>A
DLD	c.543A>T	PCCB	c.911C>T
DMD	c.8762A>G	PCDH15	c.5734G>C

DMD	c.7183G>A	PCDH15	c.4925G>C
DMD	c.6787C>T	PCDH15	c.3709G>A
DMD	c.2391T>G	PCDH15	c.1891G>T
DMD	c.2343C>A	PCDH15	c.92-523T>C
DMD	c.1122G>C	PCDH15	c.4596C>T
DMD	c.8810G>A	PCDH15	c.5254_5256delCCT
DMD	c.7728T>C	PCDH15	c.5104C>G
DMD	c.5234G>A	PCDH15	c.4254A>G
DMD	c.3734C>T	PCDH15	c.3817C>A
DMD	c.3021G>A	PCDH15	c.2581G>A
DMD	c.2645A>G	PCDH15	c.2291G>A
DMD	c.1635A>G	PCDH15	c.960A>G
DYSF	c.1351A>G	PCDH15	c.4982A>C
DYSF	c.2802G>A	PCDH15	c.4853A>C
DYSF	c.3131G>A	PCDH15	c.4487G>A
DYSF	c.356T>C	PCDH15	c.4581C>A
DYSF	c.1120G>C	PCDH15	c.2786G>A
DYSF	c.3191_3196dupCGGAGG	PCDH15	c.1910A>G
DYSF	c.3641C>T	PCDH15	c.1304A>C
DYSF	c.5373G>A	PCDH15	c.1263T>C
DYSF	c.5757T>G	PCDH15	c.1138G>A
DYSF	c.383G>A	PCDH15	c.55T>G
DYSF	c.393C>T	PEX1	c.1509A>G
DYSF	c.509C>A	PEX1	c.627G>A
DYSF	c.942C>T	PEX1	c.3103G>A
DYSF	c.1369G>A	PEX1	c.1579A>G
DYSF	c.1827T>C	PEX1	c.2442C>T
DYSF	c.1860C>T	PEX1	c.2331C>A
DYSF	c.2583A>T	PEX1	c.2088A>G
DYSF	c.3972C>T	PEX10	c.880A>G
DYSF	c.4008C>A	PEX10	c.291A>G
DYSF	c.4068C>T	PEX10	c.279C>T
DYSF	c.4504C>T	PEX2	c.748T>C
DYSF	c.5859A>C	PEX2	c.18G>A
ERCC6	c.352C>T	PEX2	c.550T>C
ERCC6	c.631G>A	PEX6	c.2644G>A
ERCC6	c.4238A>G	PEX6	c.1387A>G
ERCC6	c.3689G>C	PEX6	c.2816C>A
ERCC6	c.3637A>G	PEX6	c.2814G>A
ERCC6	c.3289A>G	PEX6	c.2426C>T
ERCC6	c.3010C>T	PEX6	c.2364G>A
ERCC6	c.2751C>T	PKHD1	c.11878G>A
ERCC6	c.1397+8395C>T	PKHD1	c.11832C>A
ERCC6	c.1397+8063G>A	PKHD1	c.4438A>G

ERCC6	c.1397+7671C>T	PKHD1	c.9415G>T
ERCC6	c.1196G>A	PKHD1	c.8848G>A
ERCC6	c.528A>G	PKHD1	c.12143A>G
ERCC6	c.411G>A	PKHD1	c.11714T>A
ERCC6	c.135C>G	PKHD1	c.11696A>G
ERCC8	c.435T>C	PKHD1	c.11525G>T
ERCC8	c.363T>C	PKHD1	c.10521C>T
EVC	c.276G>A	PKHD1	c.9402G>A
EVC	c.1528G>A	PKHD1	c.9237G>A
EVC	c.1971C>T	PKHD1	c.7764A>G
EVC	c.2626G>A	PKHD1	c.7587G>A
EVC	c.569G>T	PKHD1	c.6950T>C
EVC	c.1304G>A	PKHD1	c.5896C>T
EVC	c.1653G>A	PKHD1	c.5608T>G
EVC	c.221A>C	PKHD1	c.3756G>C
EVC	c.249A>G	PKHD1	c.2489A>G
EVC	c.341C>T	PKHD1	c.2278C>T
EVC	c.969T>C	PKHD1	c.2046A>C
EVC	c.1026G>C	PKHD1	c.1736C>T
EVC	c.1068A>G	PKHD1	c.1587T>C
EVC	c.1115C>T	PKHD1	c.1185T>C
EVC	c.1320T>A	PKHD1	c.234C>T
EVC	c.1346C>A	PKHD1	c.214C>T
EVC	c.1727G>A	PMM2	c.590A>C
EVC	c.2279G>A	PMM2	c.712C>T
EVC	c.2505G>A	PMM2	c.324G>A
EVC2	c.913G>T	POMGNT1	c.1867A>G
EVC2	c.904T>A	POMGNT1	c.1666G>A
EVC2	c.2405G>T	POMGNT1	c.681A>G
EVC2	c.2350A>G	POMGNT1	c.178A>C
EVC2	c.2017A>G	POMGNT1	c.620A>G
EVC2	c.1730T>C	PPT1	c.309T>C
EVC2	c.870G>A	PPT1	c.401T>C
EVC2	c.864C>T	PROP1	c.59A>G
EVC2	c.575C>G	PROP1	c.27T>C
EVC2	c.258G>A	RTEL1	c.1403C>G
EVC2	c.3253C>T	RTEL1	c.1905C>T
EVC2	c.2235A>G	RTEL1	c.2019C>T
EVC2	c.2095A>G	RTEL1	c.2223T>C
EVC2	c.2061T>C	RTEL1	c.3254G>C
EVC2	c.1437A>G	RTEL1	c.3327C>T
EVC2	c.688A>G	RTEL1	c.2945C>T
FAH	c.267G>C	RTEL1	c.3173C>A
FAH	c.1056C>T	RTEL1	c.371A>G

FANCA	c.2267G>C	RTEL1	c.654T>C
FANCA	c.*385C>G	RTEL1	c.1008G>A
FANCA	c.3807G>C	RTEL1	c.2064T>C
FANCA	c.3654A>G	RTEL1	c.2184C>T
FANCA	c.3263C>T	RTEL1	c.2346G>A
FANCA	c.2901C>T	RTEL1	c.2856C>T
FANCA	c.2574C>G	RTEL1	c.2857G>A
FANCA	c.2426G>A	RTEL1	c.3198A>C
FANCA	c.2151G>T	SACS	c.2580A>G
FANCA	c.1927C>G	SACS	c.9682A>G
FANCA	c.1235C>T	SACS	c.9666G>T
FANCA	c.1143G>T	SACS	c.9372T>A
FANCA	c.796A>G	SACS	c.2146C>T
FANCC	c.77C>T	SACS	c.1373C>T
FANCC	c.1200C>T	SACS	c.124A>G
FANCC	c.626G>A	SACS	c.12304T>C
FANCC	c.584A>T	SACS	c.11032C>G
FKRP	c.1411C>T	SACS	c.10338G>A
FKRP	c.135C>T	SACS	c.10291G>C
FKTN	c.166C>T	SACS	c.10106T>C
FKTN	c.608G>A	SACS	c.9981T>C
FKTN	c.1026C>A	SACS	c.8853T>C
FKTN	c.1336A>G	SACS	c.6195T>C
FMR1	c.414G>A	SACS	c.4466A>G
GAA	c.286A>G	SACS	c.2080G>A
GAA	c.395G>C	SACS	c.1656A>G
GAA	c.348G>T	SACS	c.696T>A
GAA	c.1830C>T	SGCA	c.933C>T
GAA	c.2417C>T	SGCB	c.799C>T
GAA	c.324T>C	SGCD	c.756G>A
GAA	c.596A>G	SGCD	c.15G>C
GAA	c.642C>T	SGCD	c.775_777delAAG
GAA	c.668G>A	SGCD	c.84T>C
GAA	c.921A>T	SGCD	c.290G>A
GAA	c.1203G>A	SGCG	c.267A>G
GAA	c.1374C>T	SGCG	c.228T>C
GAA	c.1581G>A	SGCG	c.312T>G
GAA	c.2133A>G	SGCG	c.347G>A
GALC	c.550C>T	SGCG	c.705T>C
GALC	c.537T>C	SGCG	c.860A>G
GALC	c.41C>G	SGSH	c.1367G>A
GALC	c.1921A>G	SGSH	c.1159G>A
GALC	c.1698A>T	SGSH	c.1081G>A
GALC	c.1685T>C	SLC12A6	c.260A>T

GALC	c.1620A>G	SLC12A6	c.1551G>C
GALC	c.1350C>T	SLC12A6	c.1236G>A
GALC	c.984G>A	SLC12A6	c.475C>T
GALC	c.742G>A	SLC17A5	c.899C>T
GALC	c.330C>T	SLC17A5	c.886G>A
GALC	c.75C>A	SLC17A5	c.606A>G
GALC	c.61G>C	SLC17A5	c.246G>A
GALC	c.42G>C	SLC22A5	c.810C>T
GALC	c.49G>A	SLC22A5	c.285T>C
GALT	c.652C>T	SLC22A5	c.807A>G
GALT	c.940A>G	SLC26A2	c.1474C>T
GCDH	c.861C>T	SLC26A2	c.1721T>C
GCDH	c.1173G>T	SLC26A2	c.2065A>T
GJB2	c.457G>A	SLC26A4	c.209C>T
GLA	c.235G>A	SLC26A4	c.1195T>C
GLB1	c.1306C>T	SLC37A4	c.527+1delG
GLB1	c.469G>T	SLC37A4	c.467C>T
GLB1	c.1594A>G	SMPD1	c.89T>G
GLB1	c.1561T>C	SMPD1	c.1643C>G
GLB1	c.34T>C	SMPD1	c.138_143delGCTGGC
GLB1	c.29C>T	SMPD1	c.719G>A
GLDC	c.2858C>G	SMPD1	c.729C>T
GLDC	c.1545G>A	SMPD1	c.1460C>T
GNE	c.1863C>T	SMPD1	c.636T>C
GNPTAB	c.3598G>A	SMPD1	c.1071C>T
GNPTAB	c.1931C>T	SMPD1	c.1522G>A
GNPTAB	c.1932A>G	SMPD1	c.1749G>A
GNPTAB	c.18G>A	STAR	c.361C>T
GRHPR	c.14G>A	STAR	c.189A>G
GRHPR	c.343G>A	TAT	c.92C>T
GRHPR	c.579A>G	TAT	c.309G>A
GRHPR	c.963G>A	TAT	c.43C>T
HADHA	c.652G>C	TCIRG1	c.1245G>C
HADHA	c.474C>T	TGM1	c.270T>A
HBA1	c.341T>A	TGM1	c.1146C>A
HBB	c.180G>A	TGM1	c.726G>A
HBB	c.9T>C	TH	c.1348G>A
HEXA	c.1306A>G	TH	c.303T>C
HEXA	c.9C>T	TH	c.174C>T
HEXB	c.185T>C	TH	c.813G>A
HEXB	c.362A>G	TH	c.360G>A
HEXB	c.619A>G	TH	c.334G>A
HGSNAT	c.1839C>T	TMEM216	c.28C>T
HGSNAT	c.1749T>C	TMEM216	c.432-11_432-10insA

HLCS	c.1920C>T	TMEM216	c.264G>A
HLCS	c.971G>A	TMEM216	c.432-1G>C
HLCS	c.834C>T	TPP1	c.1615G>T
HLCS	c.286G>A	TPP1	c.1044C>T
HLCS	c.126G>T	TPP1	c.293C>T
HMGCL	c.795C>G	TTPA	c.359-3delT
HMGCL	c.654A>G	USH1C	c.1414G>A
HOGA1	c.363G>C	USH1C	c.403G>A
HOGA1	c.474A>G	USH1C	c.1557G>C
HOGA1	c.396G>A	USH1C	c.1440C>T
HOGA1	c.912C>A	USH1C	c.1188A>G
HSD17B4	c.420A>T	USH2A	c.13364C>T
HSD17B4	c.1174G>A	USH2A	c.10666G>A
HSD17B4	c.1876T>G	USH2A	c.8827G>A
HSD17B4	c.317G>A	USH2A	c.7685T>C
HSD17B4	c.1531T>C	USH2A	c.6713A>C
HYLS1	c.468A>G	USH2A	c.3945T>C
HYLS1	c.-80-3126C>G	USH2A	c.13297G>T
HYLS1	c.-81+2379C>G	USH2A	c.11815G>A
HYLS1	c.-80-3471G>A	USH2A	c.9650A>T
HYLS1	c.-80-2955C>A	USH2A	c.9008T>C
HYLS1	c.-80-2826A>C	USH2A	c.7718G>A
HYLS1	c.91T>C	USH2A	c.6875G>A
IDUA	c.299+3564C>T	USH2A	c.2522C>A
IDUA	c.1174C>T	USH2A	c.14074G>A
IDUA	c.299+2336C>T	USH2A	c.14050T>C
IDUA	c.299+2351A>G	USH2A	c.12666A>G
IDUA	c.299+3221G>A	USH2A	c.12612A>G
IDUA	c.299+3476C>A	USH2A	c.11946G>A
IDUA	c.299+3632G>C	USH2A	c.11602A>G
IDUA	c.510G>A	USH2A	c.11504C>T
IDUA	c.299+1115G>A	USH2A	c.10232A>C
IDUA	c.299+1277G>C	USH2A	c.9595A>G
IDUA	c.299+1323T>C	USH2A	c.9430G>A
IDUA	c.299+2072C>T	USH2A	c.9343A>G
IDUA	c.543T>C	USH2A	c.9296A>G
IDUA	c.891C>T	USH2A	c.8656C>T
IDUA	c.942G>C	USH2A	c.8624G>A
IDUA	c.1081G>A	USH2A	c.7506G>A
IDUA	c.1100C>T	USH2A	c.6506T>C
IDUA	c.1164G>C	USH2A	c.6317T>C
IKBKAP	c.2543C>A	USH2A	c.5013C>A
IKBKAP	c.1911T>C	USH2A	c.4994T>C
IKBKAP	c.934G>A	USH2A	c.4457G>A

IKBKAP	c.751A>G	USH2A	c.2109T>C
IKBKAP	c.189C>T	USH2A	c.1931A>T
IKBKAP	c.3876T>G	USH2A	c.1419C>T
IKBKAP	c.1073G>A	USH2A	c.504A>G
IKBKAP	c.948G>A	USH2A	c.373G>A
IKBKAP	c.3473C>T	VPS13B	c.2471C>T
IKBKAP	c.3214T>A	VPS13B	c.3203C>T
IKBKAP	c.3069G>C	VPS13B	c.4881C>T
IKBKAP	c.2855A>T	VPS13B	c.6907C>T
IKBKAP	c.2490A>G	VPS13B	c.7920G>A
IKBKAP	c.1965C>T	VPS13B	c.2473G>A
IKBKAP	c.1926G>A	VPS13B	c.4847G>A
IKBKAP	c.1574G>A	VPS13B	c.6025G>A
IKBKAP	c.819C>T	VPS13B	c.7227G>A
IKBKAP	c.441G>A	VPS13B	c.7239C>T
IVD	c.970T>C	VPS13B	c.8978A>G
IVD	c.732C>T	VPS13B	c.9567T>C
KCNJ11	c.801C>G	VPS13B	c.10124C>T
KCNJ11	c.762G>A	VPS13B	c.10140G>T
KCNJ11	c.678C>T	VPS13B	c.10294G>A
KCNJ11	c.1154C>G	XPC	c.102G>A
KCNJ11	c.1143G>A	XPC	c.2815C>A
KCNJ11	c.1009G>A	XPC	c.2061G>A
KCNJ11	c.808C>G	XPC	c.1881T>A
KCNJ11	c.570C>T	XPC	c.1496C>T
KCNJ11	c.561G>A	XPC	c.1475G>A
KCNJ11	c.67A>G	ZFYVE26	c.7055C>T
LAMA2	c.1798G>A	ZFYVE26	c.6918A>T
LAMA2	c.2462C>T	ZFYVE26	c.5784T>C
LAMA2	c.2670A>C	ZFYVE26	c.3722G>A
LAMA2	c.1562C>T	ZFYVE26	c.2887G>C
LAMA2	c.8586T>C	ZFYVE26	c.677G>A
LAMA2	c.156C>T	ZFYVE26	c.3334G>T
LAMA2	c.381C>A	ZFYVE26	c.2826G>A
LAMA2	c.1491T>C	ZFYVE26	c.1184G>T
LAMA2	c.1856G>A	ZFYVE26	c.1153A>T
LAMA2	c.2799A>G	ZFYVE26	c.987C>T
LAMA2	c.3412G>A	ZFYVE26	c.5678G>T
LAMA2	c.4750G>A	ZFYVE26	c.5672A>G
LAMA2	c.4956C>G	ZFYVE26	c.5612G>A
LAMA2	c.5466A>G	ZFYVE26	c.4370G>A
LAMA2	c.5502G>A	ZFYVE26	c.3308C>T
LAMA2	c.6237G>A	ZFYVE26	c.3210C>G
LAMA2	c.6246G>A	ZFYVE26	c.3118T>A

LAMA2	c.7760C>T	ZFYVE26	c.2692A>T
LAMA2	c.7830G>C	ZFYVE26	c.2559G>A
LAMA2	c.7845G>A	ZFYVE26	c.2112T>C
LAMA2	c.7906A>G		
ABCC8	c.3525C>T	IDUA	c.1467C>T
ABCC8	c.2545G>A	IDUA	c.299+1657C>T
ABCC8	c.1686C>T	IDUA	c.24C>A
ABCD1	c.1683C>T	IDUA	c.60G>A
ABCD1	c.1699C>T	IDUA	c.99T>G
ABCD1	c.1700A>G	IDUA	c.299+1888C>T
ABCD1	c.1744G>A	IKBKAP	c.2446A>C
ABCD1	c.1748T>A	IKBKAP	c.2294G>A
ABCD1	c.1816T>C	IKBKAP	c.1758T>G
ABCD1	c.1823G>A	LAMB3	c.67C>T
ABCD1	c.1548G>A	LAMC2	c.599C>A
ACADS	c.511C>T	LRFN4	c.9G>A
ADA	c.22G>A	LRFN4	c.1225G>T
AGL	c.3884G>A	LRFN4	c.1880C>T
AGL	c.83-33C>T	LRFN4	c.1725C>T
AGXT	c.26C>A	MEFV	c.605G>A
AIRE	c.901G>A	MEFV	c.495C>A
AIRE	c.99T>C	MESP2	c.531G>A
ALMS1	c.72_74dupGGA	MESP2	c.561delG
ALMS1	c.66_74delGGAGGAGGA	MESP2	c.558G>A
ALMS1	c.69_74delGGAGGA	MESP2	c.558_581del
ALMS1	c.70_71insCGG	MESP2	c.558_569delGGGGCAGGGGCA
ALMS1	c.72_74delGGA	MESP2	c.558_581del
ALMS1	c.72_74dupGGA	MESP2	c.561G>A
ALPL	c.330T>C	MESP2	c.573G>A
ATM	c.5948A>G	MESP2	c.585A>G
ATP7A	c.4048G>A	MESP2	c.597A>G
ATP7A	c.-21-2239C>T	MLC1	c.1053T>C
BBS1	c.1413C>T	MLC1	c.1052_1053ins(33)
BBS10	c.1442A>G	MLC1	c.1031A>G
CBS	c.1080C>T	MLC1	c.996T>C
CBS	c.573G>A	MLC1	c.978C>T
CDKL5	c.3003C>T	MMAB	c.57C>A
CDKL5	c.3084G>A	MMAB	c.56G>A
CFTR	c.1234_1238delGCAAA	MPI	c.10C>T
CFTR	c.1265C>T	MUT	c.711A>G
CFTR	c.1312A>G	MYO7A	c.3503+12_3503+33del
CFTR	c.1360_1362delTTG	MYO7A	c.5860C>A
CFTR	c.1360_1362delTTG	NAGLU	c.1788C>T

CFTR	c.1365G>T	NBN	c.553G>C
CFTR	c.1584G>A	NEB	c.15821A>T
CLN5	c.4C>T	NEB	c.17635-3dupT
CLN5	c.72A>G	NEB	c.16021T>C
CLN5	c.223T>C	NEB	c.14370G>A
CLN5	c.234C>G	NEB	c.13353C>T
CLN6	c.34G>A	NEB	c.11077C>T
COL4A3	c.1099G>A	NEB	c.16450G>A
COL4A3	c.485A>G	NEB	c.14734A>G
COL4A3	c.976G>T	NEB	c.5370G>A
COL4A3	c.3325C>T	NPC1	c.644A>G
COL4A4	c.4548A>G	NPHS1	c.1926A>G
COL4A4	c.4207T>C	NPHS1	c.3230A>G
COL4A4	c.3979G>A	NPHS1	c.791C>G
COL4A4	c.3684G>A	NPHS1	c.492C>T
COL4A4	c.3594G>A	NPHS2	c.102A>G
COL4A4	c.1833T>C	PEX10	c.6C>G
COL4A5	c.247C>T	PEX6	c.330C>G
CYP11B1	c.1157C>T	PEX6	c.302C>G
CYP21A2	c.29_31delTGC	PEX6	c.399G>T
CYP21A2	c.118C>T	PKHD1	c.11340T>C
CYP21A2	c.138C>A	PKHD1	c.3785C>T
CYP21A2	c.188A>T	POMGNT1	c.1257G>A
CYP21A2	c.747C>G	PROP1	c.640A>C
CYP21A2	c.803C>T	PROP1	c.636T>C
CYP21A2	c.806G>C	PROP1	c.630A>C
CYP21A2	c.844G>T	PROP1	c.629_630insCCCCCCCCCCCC
CYP21A2	c.1125C>T	PROP1	c.624_625insCCCCCCCCCCCC
CYP21A2	c.1174G>A	PROP1	c.425C>T
DMD	c.8489G>A	PROP1	c.424G>A
DMD	c.4072-295T>G	RTEL1	c.1800C>T
DMD	c.7096A>C	RTEL1	c.1833G>A
DMD	c.4275A>G	RTEL1	c.2618G>A
DYSF	c.2105C>T	RTEL1	c.3047C>T
DYSF	c.570T>C	RTEL1	c.1050G>A
ERCC8	c.1123-3delT	RTEL1	c.2123G>A
EVC	c.769C>T	RTEL1	c.2616T>C
EVC	c.772T>C	RTEL1	c.*61A>G
EVC	c.1854C>T	SGCB	c.21_23delGGC
EVC2	c.3507C>T	SGSH	c.957G>A
EVC2	c.2077G>A	SLC12A6	c.3003C>T
EVC2	c.52C>T	SLC26A4	c.1790T>C
FANCA	c.12G>A	SLC37A4	c.984+247C>T
FANCA	c.1501G>A	SLC37A4	c.857G>A

FANCA	c.17T>A	SLC37A4	c.528delT
FANCA	c.3982A>G	SLC37A4	c.526delT
FKRP	c.341C>G	SLC37A4	c.1224G>A
GAA	c.2065G>A	SLC37A4	c.1062C>T
GAA	c.2338G>A	SMPD1	c.107_118delTGCTGGCGCTGG
GAA	c.2553G>A	SMPD1	c.564dupC
GCDH	c.*161G>A	SMPD1	c.567dupA
GCDH	c.*165A>G	SMPD1	c.575G>C
GLB1	c.76-4426C>T	SMPD1	c.573T>C
GLDC	c.578A>G	TCIRG1	c.479G>A
GLDC	c.2113G>A	TCIRG1	c.1985G>A
GLDC	c.498T>C	TCIRG1	c.166C>T
GLDC	c.190G>T	TCIRG1	c.384C>T
GLDC	c.52G>T	TH	c.687G>T
GLDC	c.501G>A	TPP1	c.1542A>T
GLDC	c.249G>A	USH1C	c.1285-7531C>T
GNPTG	c.67G>C	USH1C	c.1285-7619C>T
HEXA	c.1518A>G	USH2A	c.9262G>A
HSD17B4	c.2182A>G	USH2A	c.13191G>A
HSD17B4	c.1675A>G	VPS13B	c.5980A>G
HSD17B4	c.2199C>T	VPS13B	c.9407A>T
IDS	c.937C>T	VPS13B	c.1206+33T>G
IDS	c.438C>T	VPS13B	c.9424A>G
IDUA	c.314G>A	ZFYVE26	c.879G>A
IDUA	c.352C>T	ZFYVE26	c.3365C>T
IDUA	c.1230C>G	ZFYVE26	c.1844C>T
IDUA	c.1360G>A		

*The variants above the black bar had a GQ score>/=90 in all of the 44 Armenian individuals. The variants below the black bar are variants in which there was a GQ score<90 in at least one of the 44 Armenian individuals. The majority of these variants were likely present in at least one of the 44 Armenian individuals; however, some of the variants below the black bar may have been incorrect calls and may in fact not be present in this cohort.

APPENDIX B: List of the 176 Conditions in the Foresight Expanded Carrier Screen

CONDITION	*GENE*	CONDITION	*GENE*
11-Beta-Hydroxylase-Deficient Congenital Adrenal Hyperplasia	*CYP11B1*	Hydrolethalus Syndrome	*HYLS1*
21-Hydroxylase-Deficient Congenital Adrenal Hyperplasia	*CYP21A2**	Hypophosphatasia, Autosomal Recessive	*ALPL*
6-Pyruvoyl-Tetrahydropterin Synthase Deficiency	*PTS*	Inclusion Body Myopathy 2	*GNE*
ABCC8-Related Hyperinsulinism	*ABCC8*	Isovaleric Acidemia	*IVD*
Adenosine Deaminase Deficiency	*ADA*	Joubert Syndrome 2	*TMEM216*
Adrenoleukodystrophy: X-Linked	*ABCD1*	KCNJ11-Related Familial Hyperinsulinism	*KCNJ11*
Alpha Thalassemia	*HBA1/HBA2**	Krabbe Disease	*GALC*
Alpha-Mannosidosis	*MAN2B1*	LAMA2-Related Muscular Dystrophy	*LAMA2*
Alpha-Sarcoglycanopathy (including Limb-Girdle Muscular Dystrophy, Type 2D)	*SGCA*	Leigh Syndrome, FrenchCanadian Type	*LRPPRC*
Alport Syndrome, X-Linked	*COL4A5*	Lipoamide Dehydrogenase Deficiency	*DLD*
Alstrom Syndrome	*ALMS1*	Lipoid Congenital Adrenal Hyperplasia	*STAR*
AMT-Related Glycine Encephalopathy	*AMT*	Lysosomal Acid Lipase Deficiency	*LIPA*
Andermann Syndrome	*SLC12A6*	Maple Syrup Urine Disease, Type Ia	*BCKDHA*
Argininemia	*ARG1*	Maple Syrup Urine Disease, Type IB	*BCKDHB*
Argininosuccinic Aciduria	*ASL*	Maple Syrup Urine Disease, Type II	*DBT*
ARSACS	*SACS*	Medium Chain Acyl-CoA Dehydrogenase Deficiency	*ACADM*
Aspartylglycosaminuria	*AGA*	Megalencephalic Leukoencephalopathy with Subcortical Cysts	*MLC1*

Ataxia with Vitamin E Deficiency	*TTPA*	Metachromatic Leukodystrophy	*ARSA*
Ataxia-Telangiectasia	*ATM*	Methylmalonic Acidemia, cblA Type	*MMAA*
ATP7A-Related Disorders	*ATP7A*	Methylmalonic Acidemia, cblB Type	*MMAB*
Autosomal Recessive Osteopetrosis, Type 1	*TCIRG1*	Methylmalonic Aciduria and Homocystinuria, cblC Type	*MMACHC*
Bardet-Biedl Syndrome, BBS1-Related	*BBS1*	MKS1-Related Disorders	*MKS1*
Bardet-Biedl Syndrome, BBS10-Related	*BBS10*	Mucolipidosis III Gamma	*GNPTG*
Bardet-Biedl Syndrome, BBS12-Related	*BBS12*	Mucolipidosis IV	*MCOLN1*
Bardet-Biedl Syndrome, BBS2-Related	*BBS2*	Mucopolysaccharidosis, Type I (including Hurler Syndrome)	*IDUA*
Beta-Sarcoglycanopathy (including Limb-Girdle Muscular Dystrophy, Type 2E)	*SGCB*	Mucopolysaccharidosis, Type II	*IDS*
Biotinidase Deficiency	*BTD*	Mucopolysaccharidosis, Type IIIA	*SGSH*
Bloom Syndrome	*BLM*	Mucopolysaccharidosis, Type IIIB	*NAGLU*
Calpainopathy	*CAPN3*	Mucopolysaccharidosis, Type IIIC	*HGSNAT*
Canavan Disease	*ASPA*	Muscle-Eye-Brain Disease	*POMGNT1*
Carbamoylphosphate Synthetase I Deficiency	*CPS1*	MUT-Related Methylmalonic Acidemia	*MUT*
Carnitine Palmitoyltransferase IA Deficiency	*CPT1A*	MYO7A-Related Disorders	*MYO7A*
Carnitine Palmitoyltransferase II Deficiency	*CPT2*	NEB-Related Nemaline Myopathy	*NEB*
Cartilage-Hair Hypoplasia	*RMRP*	Niemann-Pick Disease, Type C	*NPC1*
Cerebrotendinous Xanthomatosis	*CYP27A1*	Niemann-Pick Disease, Type C2	*NPC2*
Citrullinemia, Type 1	*ASS1*	Niemann-Pick Disease, SMPD1-Associated	*SMPD1*

CLN3-Related Neuronal Ceroid Lipofuscinosis	*CLN3*	Nijmegen Breakage Syndrome	*NBN*
CLN5-Related Neuronal Ceroid Lipofuscinosis	*CLN5*	Northern Epilepsy	*CLN8*
CLN6-Neuronal Ceroid Lipofuscinosis, Type 6	*CLN6*	Ornithine Transcarbamylase Deficiency	*OTC*
Cohen Syndrome	*VPS13B*	PCCA-Related Propionic Acidemia	*PCCA*
COL4A3-Related Alport Syndrome	*COL4A3*	PCCB-Related Propionic Acidemia	*PCCB*
COL4A4-Related Alport Syndrome	*COL4A4*	PCDH15-Related Disorders (including Usher Syndrome, Type 1F)	*PCDH15*
Congenital Disorder of Glycosylation, Type Ia	*PMM2*	Pendred Syndrome	*SLC26A4*
Congenital Disorder of Glycosylation, Type Ib	*MPI*	Peroxisome Biogenesis Disorder, Type 3	*PEX12*
Congenital Disorder of Glycosylation, Type Ic	*ALG6*	Peroxisome Biogenesis Disorder, Type 4	*PEX6*
Congenital Finnish Nephrosis	*NPHS1*	Peroxisome Biogenesis Disorder, Type 5	*PEX2*
Costeff Optic Atrophy Syndrome	*OPA3*	Peroxisome Biogenesis Disorder, Type 6	*PEX10*
Cystic Fibrosis	*CFTR*	PEX1-Related Zellweger Syndrome Spectrum	*PEX1*
Cystinosis	*CTNS*	Phenylalanine Hydroxylase Deficiency	*PAH*
D-Bifunctional Protein Deficiency	*HSD17B4*	PKHD1-Related Autosomal Recessive Polycystic Kidney Disease	*PKHD1*
Delta-Sarcoglycanopathy	*SGCD*	Polyglandular Autoimmune Syndrome, Type 1	*AIRE*
Dysferlinopathy	*DYSF*	Pompe Disease	*GAA*
Dystrophinopathies (including Duchenne/Becker Muscular Dystrophy)	*DMD*	PPT1-Related Neuronal Ceroid Lipofuscinosis	*PPT1*
ERCC6-Related Disorders	*ERCC6*	Primary Carnitine Deficiency	*SLC22A5*
ERCC8-Related Disorders	*ERCC8*	Primary Hyperoxaluria, Type 1	*AGXT*

EVC-Related Ellis-Van Creveld Syndrome	*EVC*	Primary Hyperoxaluria, Type 2	*GRHPR*
EVC2-Related Ellis-Van Creveld Syndrome	*EVC2*	Primary Hyperoxaluria, Type 3	*HOGA1*
Fabry Disease	*GLA*	PROP1-Related Combined Pituitary Hormone Deficiency	*PROP1*
Familial Dysautonomia	*IKBKAP*	Pycnodysostosis	*CTSK*
Familial Mediterranean Fever	*MEFV*	Pyruvate Carboxylase Deficiency	*PC*
Fanconi Anemia Complementation, Group A	*FANCA*	Rhizomelic Chondrodysplasia Punctata, Type 1	*PEX7*
Fanconi Anemia, Type C	*FANCC*	RTEL1-Related Disorders	*RTEL1*
FKRP-Related Disorders	*FKRP*	Salla Disease	*SLC17A5*
FKTN-Related Disorders (including Walker-Warburg Syndrome)	*FKTN*	Sandhoff Disease	*HEXB*
Fragile X Syndrome	*FMR1**	Segawa Syndrome	*TH*
Galactokinase Deficiency	*GALK1*	Short Chain Acyl-CoA Dehydrogenase Deficiency	*ACADS*
Galactosemia	*GALT*	Sjogren-Larsson Syndrome	*ALDH3A2*
Gamma-Sarcoglycanopathy	*SGCG*	Smith-Lemli-Opitz Syndrome	*DHCR7*
Gaucher Disease	*GBA**	Spastic Paraplegia, Type 15	*ZFYVE26*
GJB2-Related DFNB1 Nonsyndromic Hearing Loss and Deafness (including two GJB6 deletions)	*GJB2*	Spinal Muscular Atrophy	*SMN1**
GLB1-Related Disorders	*GLB1*	Spondylothoracic Dysostosis	*MESP2*
GLDC-Related Glycine Encephalopathy	*GLDC*	Steroid-Resistant Nephrotic Syndrome	*NPHS2*
Glutaric Acidemia, Type 1	*GCDH*	Sulfate Transporter-Related Osteochondrodysplasia	*SLC26A2*
Glycogen Storage Disease, Type Ia	*G6PC*	TGM1-Related Autosomal Recessive Congenital Ichthyosis	*TGM1*
Glycogen Storage Disease, Type Ib	*SLC37A4*	TPP1-Related Neuronal Ceroid Lipofuscinosis	*TPP1*
Glycogen Storage Disease, Type III	*AGL*	Tyrosinemia, Type I	*FAH*

GNPTAB-Related Disorders	*GNPTAB*	Tyrosinemia, Type II	*TAT*
GRACILE Syndrome	*BCS1L*	USH1C-Related Disorders	*USH1C*
HADHA-Related Disorders (including Long Chain 3-Hydroxyacyl-CoA Dehydrogenase Deficiency)	*HADHA*	USH2A-Related Disorders	*USH2A*
Hb Beta Chain-Related Hemoglobinopathy (including Beta Thalassemia and Sickle Cell Disease)	*HBB*	Usher Syndrome, Type 3	*CLRN1*
Hereditary Fructose Intolerance	*ALDOB*	Very Long Chain Acyl-CoA Dehydrogenase Deficiency	*ACADVL*
Herlitz Junctional Epidermolysis Bullosa, LAMA3-Related	*LAMA3*	Wilson Disease	*ATP7B*
Herlitz Junctional Epidermolysis Bullosa, LAMB3-Related	*LAMB3*	X-Linked Congenital Adrenal Hypoplasia	*NR0B1*
Herlitz Junctional Epidermolysis Bullosa, LAMC2-related	*LAMC2*	X-Linked Juvenile Retinoschisis	*RS1*
Hexosaminidase A Deficiency (including Tay-Sachs Disease)	*HEXA*	X-Linked Myotubular Myopathy	*MTM1*
HMG-CoA Lyase Deficiency	*HMGCL*	X-Linked Severe Combined Immunodeficiency	*IL2RG*
Holocarboxylase Synthetase Deficiency	*HLCS*	Xeroderma Pigmentosum, Group A	*XPA*
Homocystinuria caused by Cystathionine Beta-Synthase Deficiency	*CBS*	Xeroderma Pigmentosum, Group C	*XPC*

* Indicates that the gene is typically analyzed using a custom assay by the commercial laboratory (differs from this study, where exome sequencing was performed, as highlighted above in METHODS)

Text Green: Indicates that testing is recommended by ACOG

Text Purple: Indicates that testing is recommended by ACMG

Text Blue: Indicates that testing is recommended by ACOG and ACMG