

The importance of population specific sequence variants as control to investigate the causality of rare sequence variants in human diseases in Jordans

Abstract

The majority of Mendelian diseases are caused by mutations affecting the coding segments of a gene. Therefore, Whole Exome Sequences (WES) based on Next Generation Sequencing (NGS) technology is obtained for patients affected by rare Mendelian diseases. The generated sequence file of NGS technology harbors huge numbers of variants compared to the reference genome. In order to identify the best candidate causative variant (mutation), filtration processes should take place. Mendelian diseases are rare hence making their causative variants rare. Variants are basically filtered based on their frequency within controls. The current publicly available controls are stored in gnomAD, EXAC, EVS and the 1000 genome. None of these repositories contain genetic data for Middle Eastern populations. This hinders the differentiation between real disease causing variants and the local polymorphism of rare variants especially in populations with high consanguinity such as Jordan. Population specific catalogs of genetic variation is mandatory to identify the rare variants causing rare diseases, and for drug pharmacokinetics treatment efficacy and adverse drug reactions. This review highlights the importance of having a Jordanian specific genetic variant control cohort to achieve the goal of variant identification implicitly.

Keywords: population genetics, consanguineous, MAF, OR, population stratification

Volume 6 Issue 3 - 2017

Tawfiq Froukh,¹ Saja Q Froukh²
¹Department of Biotechnology and Genetic Engineering, Philadelphia University, Jordan

²Center for Research on Health and Aging, University of Illinois, USA

Correspondence: Tawfiq Froukh, Department of Biotechnology and Genetic Engineering Philadelphia University - Jerash Road, Amman (11118) Jordan, Tel +962 6 4799000, Fax +962 6 4799040, Email tfroukh@philadelphia.edu.jo

Received: September 18, 2017 | **Published:** October 23, 2017

Human diseases

Human diseases are caused by environmental factors (infections, malnutrition, poisons, or injuries), or genetic factors including single gene/Mendelian diseases, complex diseases (defects in multiple genes), and genomic diseases (chromosomal abnormalities). The focus is on Mendelian diseases which are characterized by: (1) rare alleles (<0.5% MAF, minor allele frequency) or very rare alleles (<0.1%MAF), and (2) high effect size (OR>3; Odds Ratio of genetic variant per disease expression).¹ Much of what is known about the relationship between gene function and phenotype is based on the identification of rare variants causing Mendelian diseases. Such identifications have developed new diagnostic, therapeutic, and preventative strategies.²

Mendelian diseases and next generation sequencing (NGS)

Diagnosing many Mendelian diseases by phenotypic features and conventional diagnostic testing is challenging in most cases.³ According to the National Institute of Health diagnosed program, a general clinical geneticist diagnostic rate is ~34% for adults and ~11% for children.⁴ Moreover, the diagnoses time is prolonged. For example, in a survey for time needed to diagnose 8 rare diseases including fragile X syndrome and cystic fibrosis: 25% of the families waited between 5 and 30 years for the diagnosis, and 40% of the families were with wrong initial diagnosis.⁵

The NGS is used to obtain whole genome sequence (WGS) or whole exome sequence (WES). The human WGS and WES are useful to detect DNA variants in patients with rare disorders.⁶ WES has been

useful to identify variants that cause Mendelian diseases. The majority of these diseases are caused by variants in the coding region of a gene which composes around 1% of the human genome (~60Megabase).⁷ The availability of clinical WES testing promises better diagnostic yield, and importantly, studies of the diagnostic efficacy of clinical WES show that the diagnostic successes depend on the discovery of disease causing genes.⁸ This highlights the value of continued research into the genetic basis of Mendelian diseases. In addition, the diagnostic rates will continue to increase as the work continued toward a more complete catalog of the disease-causing genes and the disease-causing variants.

Sequence variants as control

One powerful approach to discover disease causing genes and variants is to study diseases in populations with high rates of parental consanguinity where recessive forms of diseases are enriched. In Jordan, 39.7% of marriages are consanguineous. The lowest percentage is in the capital Amman where 25.5% of marriages are consanguineous, and the highest is in Irbid-north east- with 52.1% consanguineous marriages (www.consang.net). In Amman before 1980, first cousin marriages comprised ~30% of all marriages.⁹ Such high percentages of consanguinity in Jordan result in an increased risk for recessive Mendelian diseases. Among couples with a child genetically diagnosed with a recessive genetic disease in Amman, 69% were offspring of couples who were first degree cousins, compared to 14% from non-consanguineous marriages.¹⁰

Recently NGS technique was used in Jordan to identify the variants causing intellectual disability (ID) and other forms of neurodevelopmental disorders in consanguineous families. These

studies were within frame of joint projects as follows:

1. Seven consanguineous families with nonspecific ID were investigated. Causative variants were identified in genes previously implicated with ID (two families), variants of uncertain significance were identified in genes previously implicated in association with ID (two families), and candidate variants were identified in genes not previously implicated with ID (two families). By matching the phenotypes in the seventh family with other families from Egypt and Pakistan pathogenic variants were identified in the gene MBOAT7 as causative.^{11,12}
2. A project between Philadelphia University/Jordan and the institute of medical genetics and applied genomics/Tuebingen University/Germany funded by the German Academic Exchange Service (DAAD) aims to identify the genetic causes of ID in 103 consanguineous families in Jordan. While sequencing and variant filtration are progressing, likely pathogenic mutations were identified in 15 families and strong candidates were identified in an additional 25 families. In one of the families from Jordan, the phenotypes of the proband found to overlap with one of the patients in Saudi Arabia. Joint collaboration was established and a new gene implicated with epileptic encephalopathy was investigated and published.¹³
3. Genome-wide genetic analyses on ten Jordanian families using a combination of whole exome sequencing and homozygosity mapping were conducted at Columbia University Medical Center in New York. The likely causative variants were revealed in 50% of the families, including four with recessive disease-causing mutations, and strong candidates in an additional case.¹⁴
4. Furthermore, variants in the three genes that were previously implicated with ID were identified in three consanguineous families from Jordan.¹⁵

Sequence variants as control

Interpretation of variants obtained by NGS requires large scale reference data sets of human genetic variation. EVS¹⁶ and the 1000 genome project¹⁷ are publicly available datasets that contain DNA data for 6,503 exomes and 2,504 individuals respectively. EXAC is a larger dataset that harbors data for 60,706 individuals.¹⁸ The represented populations in the EXAC dataset are African/African American (5,203), Latino (5,789), East Asian (4,327), Finnish (3,307), Non-Finnish European (33,370), South Asian (8,256), and other (454).^{18,19} The largest publicly available dataset of DNA variants is gnomAD which spans 123,136 exome sequences and 15,496 whole-genome sequences from unrelated individuals sequenced as part of various disease-specific and population genetic studies.¹⁸ The represented populations in the gnomAD dataset are African/African American (7,652 exomes and 4,368 genomes), Latino (16,791 exomes and 419 genomes), Ashkenazi Jewish (4,925 exomes and 151 genomes), East Asian (8,624 exomes and 811 genomes), Finnish (11,150 exomes and 1,747 genomes), Non-Finnish European (55,860 exomes and 7,509 genomes), South Asian (15,391 exomes and 0 genomes), and other (2,743 exomes and 491 genomes). The limitation in these large datasets is the lack of Middle Eastern populations.¹⁸

The controls that were used in the exhibited projects in Jordan are the public datasets (EVS, 1000 genome project, EXAC and gnomAD) and the internal control cohort in Germany (Erlangen and Tuebingen) or in the U.S. (New York). No controls are available from Jordan and/

or the Middle East. Because of the population stratification of rare variants which show stronger geographic clustering than common variants, control datasets should match closely with the patient's ancestry.¹⁹ Therefore, population specific catalogs of genetic variation are mandatory not only to identify the rare variants causing the rare diseases, but also for drug pharmacokinetics treatment efficacy and adverse drug reactions. Lack of population-specific genetic variation hinders the differentiation between real disease-causing variants and the local polymorphism of rare variants.²⁰ Therefore, establishing population specific DNA variant datasets in Jordan is very important in order to proceed identifying the rare genetic variants that cause rare Mendelian diseases in Jordan.

Acknowledgements

The authors would like to thank the Jordanian families willing to participate as control.

Conflict of interest

The author declares no conflict of interest.

References

1. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747–753.
2. Bainbridge MN, Wiszniewski W, Murdock DR, et al. Whole-genome sequencing for optimized patient management. *Sci Transl Med*. 2011;3(87):87re3.
3. Shashi V, McConkie-Rosell A, Rosell B, et al. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genet Med*. 2014;16(2):176–182.
4. Gahl WA, Markello TC, Toro C, et al. The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. *Genet Med*. 2012;14(1):51–59.
5. http://www.eurordis.org/IMG/pdf/Fact_Sheet_Eurordiscare2.pdf
6. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014;508(7497):469–476.
7. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*. 2011;12(11):745–755.
8. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*. 2014;312(18):1870–1879.
9. Khoury SA, Massad D. Consanguineous marriage in Jordan. *Am J Med Genet*. 1992;43(5):769–775.
10. Hamamy HA, Masri AT, Al-Hadidy AM, et al. Consanguinity and Genetic disorders. *Saudi Med J*. 2007;28(7):1015–1017.
11. Reuter MS, Tawamie H, Buchert R, et al. Diagnostic Yield and Novel Candidate Genes by Exome Sequencing in 152 Consanguineous Families with Neurodevelopmental Disorders. *JAMA Psychiatry*. 2017;74(3):293–299.
12. Johansen A, Rosti RO, Musaev D, et al. Mutations in MBOAT7, Encoding Lysophosphatidylinositol Acyltransferase I, Lead to Intellectual Disability Accompanied by Epilepsy and Autistic Features. *Am J Hum Genet*. 2016;99(4):912–916.
13. Han C, Alkhatir R, Froukh T, et al. Epileptic Encephalopathy Caused by Mutations in the Guanine Nucleotide Exchange Factor DENND5A. *Am J Hum Genet*. 2016;99(6):1359–1367.

14. Tawfiq Froukh, Xiaolin Zhu, Vandana Shashi, et al. Genetic basis of intellectual disability in consanguineous families from Jordan. Submitted.
15. Alkhateeb AM, Aburahma SK, Habbab W, et al. Novel mutations in WWOX, RARS2, and C10orf2 genes in consanguineous Arab families with intellectual disability. *Metab Brain Dis*. 2016;31(4):901–907.
16. <http://evs.gs.washington.edu/EVS/>
17. <http://browser.1000genomes.org/index.html>
18. <http://exac.broadinstitute.org>
19. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285–291.
20. Goldstein DB, Allen A, Keebler J, et al. Sequencing studies in human genetics: design and interpretation. *Nat Rev Genet*. 2013;14(7):460–470.