

Ashkenazi Jewish genomic variants: integrating data from the Israeli National Genetic Database and gnomAD

Joël Zlotogora, MD, PhD¹, George P. Patrinos, PhD^{2,3} and Vardiella Meiner, MD^{1,4}

Purpose: The aim of the study was to compare the data for mutations related to clinical disorders reported among Ashkenazi Jewish patients in the Israeli National Genetic Database (INGD) with variants included in the Genome Aggregation Database (gnomAD).

Methods: We extracted data for mutations claimed to cause disorders reported among Ashkenazi Jews from the INGD and searched gnomAD for each of them. We compared the allele frequency of each variant in Ashkenazi Jews with that of other delineated populations.

Results: Of the 58 INGD-reported mutations related to autosomal-dominant disorders, 19 were present in gnomAD (32.8%). Of the 309 mutations related to autosomal-recessive disorders, 240 (77.7%) were variants found in gnomAD. Of these variants, 202

(84.2%) were documented among one or more Ashkenazi individuals. At this point in the INGD, there are 168 Ashkenazi assumed founder mutations in 128 different genes corresponding to 111 autosomal-recessive disorders.

Conclusion: Integration of information on mutations among Ashkenazi Jews extracted from the INGD with their population frequency recorded in gnomAD is important for effective straightforward molecular diagnosis as well as for targeted carrier screening either for reproductive decision-making or for implementation of disease-modifying behavior.

Genet Med advance online publication 16 November 2017

Key Words: Ashkenazi Jews; founder mutations; gnomAD; screening

INTRODUCTION

With approximately 14,412,200 individuals as of the beginning of 2016, the world Jewish population includes a majority of Ashkenazi origin.¹ The Ashkenazi Jewish population that lived mainly in central and Eastern Europe maintained a genetic isolation, separated from its neighbors by religious and cultural practices as well as consanguinity. The main evidence for this isolation of the Ashkenazi Jewish population is the existence of genetic characteristics, including a high prevalence of autosomal-recessive diseases and a relatively high frequency of alleles that confer a risk to common disorders such as breast and ovarian cancer, or variants associated with inflammatory bowel disease or for Parkinson disease.² The Israeli National Genetic Database (INGD) (<http://server.goldenhelix.org/israeli>), launched in 2006, is a freely available online resource for medical genetic caregivers.³ The database includes mutations related to clinical disorders characterized among Jews and Israeli Arabs from medical publications in the scientific literature or from personal communications associating variants with clinical phenotype.

In 2016, we reviewed the INGD and curated it according to entries into the ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar>) and ExAC (<http://exac.broadinstitute.org>) databases.⁴ Later in 2016, the Genome Aggregation Database (gnomAD;

<http://gnomad.broadinstitute.org>) became available online, and included the Ashkenazi Jews as a distinct group of 5,685 individuals representing more than 8% of the total number of individuals in the database.⁵

The availability of the molecular data associated with genetic disorders in the INGD, and the population distribution of variants in gnomAD, allowed comparing the two data sets with reference to incidence of disease-causing variants among Ashkenazi Jews, and the results are presented.

MATERIALS AND METHODS

Extraction data from the databases

INGD

Details on the INGD were previously published.^{3,4} The database includes all mutations for recessive disorders based on published and on personal communication of clinical genetic reports. The criteria for being listed in the INGD for mutations in dominant or X-linked disorders is being reported in several families, or if associated with a frequent disorder.

We extracted data for mutations claimed to cause disorders reported among Ashkenazi Jews from the INGD.

Since the INGD is a biomedical literature-derived database, which relies mostly on published information, the classification of Ashkenazi origin relies on the issued report without

¹Faculty of Medicine, Hebrew University, Jerusalem, Israel; ²Department of Pharmacy, University of Patras School of Health Sciences, Patras, Greece; ³Department of Bioinformatics, Faculty of Medicine and Health Sciences, Erasmus University Medical Center, Rotterdam, The Netherlands; ⁴Department of Genetics and Metabolic Diseases, Hadassah-Hebrew University Medical Center, Jerusalem, Israel. Correspondence: Joël Zlotogora (zlotogora@gmail.com)

Submitted 10 June 2017; accepted 3 October 2017; advance online publication 16 November 2017. doi:10.1038/gim.2017.193

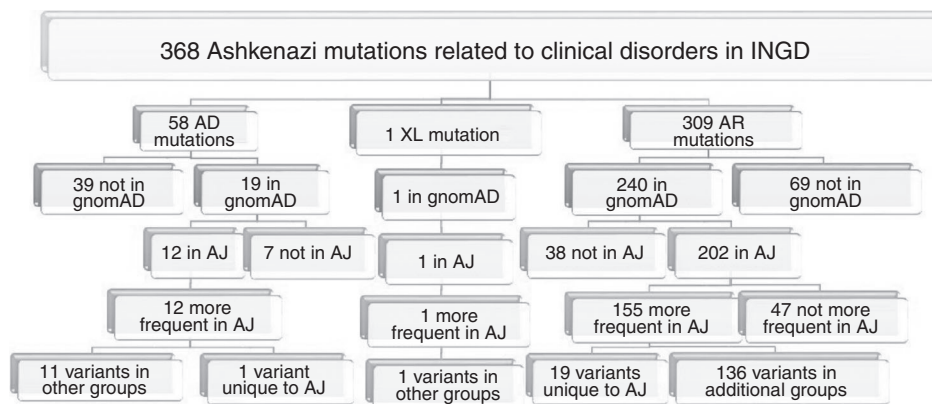


Figure 1 Study design and results. AD, autosomal dominant; AJ, Ashkenazi Jews; AR, autosomal recessive; gnomAD, Genome Aggregation Database; INGD, Israeli National Genetic Database.

specific knowledge on population metrics or grandparental ancestries.

GnomAD

The data set provided in gnomAD spans 126,216 exome sequences and 15,136 whole-genome sequences extracted from a variety of large-scale sequencing projects.⁵ The Ashkenazi Jews represent more than 8% of the total number of individuals included in gnomAD and the data originated from 2,641 inflammatory bowel disease and 3,044 non-inflammatory bowel disease samples (<https://ibd.broadinstitute.org>).

Comparisons between gnomAD variants and the INGD-derived disease-causing mutations

We queried gnomAD for each of the disease or disease-risk causative mutations present among the Ashkenazi Jews recorded in the INGD. The mutations that initially were overlooked, mostly due to the use of an outdated misleading nomenclature, were followed by search in the Human Gene Mutation Database (<http://www.hgmd.org>).

For each single variant found in gnomAD, we compared the allele frequency in Ashkenazi Jews with that documented for the other delineated populations, and measured the differences using Pearson's chi-squared test.

RESULTS

On 1 June 2017, the INGD comprised 368 entries reported specifically in Ashkenazi Jews of the total of 887 (41.5%) entries among Jews. There were 309 (84%) mutations related to clinical disorders located in 167 genes corresponding to 140 different autosomal-recessive disorders recorded in the INGD that have been reported among Ashkenazi Jewish patients (**Figure 1**).

There were 59 (16%) mutations related to clinical disorders located in 11 genes corresponding to 7 autosomal-dominant disorders and 1 X-linked recessive disorder (**Figure 1**).

Comparison between the curated Ashkenazi mutations related to clinical disorders in the INGD and variants in gnomAD

Autosomal-recessive disorders

Of the 309 mutations related to clinical disorders recorded in the INGD, 240 (77.7%) were found in gnomAD. Of these, 202 (84.2%) variants in gnomAD were documented among one or more Ashkenazi individuals and the remaining 38 were reported solely in non-Ashkenazi Jewish groups (**Figure 1**).

Nineteen of the 202 variants were present in gnomAD uniquely among Ashkenazi Jews ("Ashkenazi-only"), with an allele frequency of 1/10,000–1/1,000 (range 1–10 alleles) (**Supplementary Table 1 online**). None of these variants were present in the INGD among non-Ashkenazi Jews.

Of the mutations related to clinical disorders in INGD, 38 gnomAD-present variants were missing in the Ashkenazi Jewish group and recorded only in the other ethnic groups. Most of these variants were reported among non-Finnish Europeans (32/38; 84%) but also in additional groups. Four of these 38 variants were present in INGD among occasional non-Ashkenazi Jewish patients.

We identified 183 INGD-derived mutations as gnomAD variants in Ashkenazi Jews that were also present in other population groups. Of these, 50 variants were only present in one additional ethnic group in gnomAD; in most cases non-Finnish Europeans (39/50; 78%). From the 39 variants unique to Ashkenazi Jews and non-Finnish Europeans, 38 were more frequent in Ashkenazi Jews.

Altogether, in 136 of the 183 variants (74.3%), the allele frequency for each variant, calculated independently, was significantly higher in Ashkenazi Jews than in any of the other population groups ($P < 0.005$) ("Ashkenazi-prevalent"). Eleven of these 136 variants were present in INGD among non-Ashkenazi Jewish patients.

Autosomal-dominant disorders

Of the 58 mutations related to autosomal-dominant disorders, 19 were present in gnomAD including 12 (20.7%) in one or more Ashkenazi individuals (**Figure 1**).

One variant (*MSH6*, c.3984_3987dupGTCA; p.Leu1330-ValfsTer12) was present in gnomAD uniquely among Ashkenazi Jews (**Supplementary Table 1**). One *LDLR* variant (c.2479G > A; p.Val827Ile) was reported in gnomAD in homozygosity (**Supplementary Table 2**).

X-linked disorder

The X-linked putative disease-causing mutation c.5030G > A; p.Arg1683Gln in *COL4A5* was reported in gnomAD in 3 of 7,291 Ashkenazi Jewish alleles (all females).

Mutations related to autosomal-recessive disorders in INGD reported in more than two nonrelated, nonconsanguineous families

In the INGD, 113 mutations related to autosomal-recessive disorders were reported in more than 2 nonrelated non-consanguineous families. These mutations were located in 81 different genes corresponding to 65 autosomal-recessive disorders (**Supplementary Table 3**).

One hundred and five of these 113 mutations (92.9%) were present in gnomAD. The other 8 mutations were mostly large deletions or duplications, missing as expected for data derived from deep sequencing reported in gnomAD.

Interestingly, 55 INGD mutations reported in less than 2 nonrelated families were present in gnomAD with a significantly increased allele frequency among Ashkenazi Jews (**Supplementary Table 4**).

Ashkenazi Jewish founder mutations related to autosomal-recessive disorders

There were 160 mutations related to clinical autosomal-recessive disorders deposited in INGD and found in gnomAD either uniquely among Ashkenazi Jews or in a significantly increased frequency (105 reported in more than 2 nonrelated nonconsanguineous families and 55 reported in occasional patients). Those 160 mutations combined with the 8 mutations not identified in gnomAD and reported in more than two nonrelated families in INGD may be assumed Ashkenazi founder recessive mutations. These 168 Ashkenazi assumed founder mutations were present in 128 different genes corresponding to 111 autosomal-recessive disorders.

In 27 of the 128 different genes (21.1%), there was more than 1 assumed Ashkenazi founder mutation (**Supplementary Tables 3 and 4**). In 19 of the 27 genes 2 assumed founder mutations are known, and in 6 genes 3 assumed founder mutations are known. There were 4 assumed founder mutations in one gene (*GBA*), and *PAH* encoding for phenylalanine hydroxylase harbored 7 assumed founder mutations.

DISCUSSION

Comparison of the information on published mutations related to clinical disorders in Ashkenazi Jews deposited in INGD with sequence variants obtained from whole-exome and whole-genome sequencing in gnomAD reveals interesting aspects with relevance both to population genetics and to clinical practice. An upgraded version of the INGD is

currently pursued to include additional data, particularly cross-referencing it with gnomAD.⁴

A major part of the clinically related Ashkenazi mutations in the INGD was found in gnomAD (260/368; 70.7%), while a previous similar search in ExAC identified only 38%.⁴ The increased yield upon searching gnomAD compared with our previous report stems from the fact that gnomAD comprises a large amount of available data with the possibility to perform a more accurate search for the variants and to focus specifically on the Ashkenazi subgroup. In addition, we were able to cross-check the mutation nomenclature with its genomic details as presented in the Human Gene Mutation Database.

Out of the 240 INGD Ashkenazi autosomal-recessive mutations related to clinical disorders that we identified in gnomAD, 202 were present among Ashkenazi Jews. Of these, 19 (9.4%) were neither found in non-Ashkenazi Jews nor in other populations groups, probably representing mutations that first occurred in Ashkenazi Jewish individuals and are referred to as “Ashkenazi-only” mutations. The rest were reported also in additional populations; however, a vast majority (74.3%) were significantly more frequent among Ashkenazi Jews and hence are referred to as “Ashkenazi-prevalent.” Not surprisingly, non-Finnish Europeans were the group in which these INGD-derived variants were most frequently present.

The 168 Ashkenazi assumed founder mutations included previously known founder mutations and mutations related to clinical disorders that were either “Ashkenazi-only” or “Ashkenazi-prevalent” for which the option for recurrent mutational events is still a possibility because haplotype analysis was not performed. While for the disease-causing mutations that are part of the currently available screening programs, estimations of carrier frequencies are available from a large number of healthy controls,⁶ data for the other INGD-deposited mutations were nonexistent or based on small samples. The availability of a large number of Ashkenazi Jewish sequences in gnomAD allowed for an accurate estimation of the frequency of the mutations related to clinical disorders reported in this population. Previously, disorders such as Tay–Sachs disease, Gaucher disease, Niemann–Pick type A disease, mucopolysaccharidosis IV, familial dysautonomia, Bloom syndrome, Canavan disease, factor XI deficiency, or torsion dystonia were characterized as “Ashkenazi Jewish diseases” on clinical grounds.⁷ Advancement in molecular studies offered the opportunity to characterize founder disease-causing mutations in these disorders as well as in others reported in several families. Nowadays, with the availability of the data for large numbers of healthy Ashkenazi Jews carrying variants associated with either recessive or dominant disorders, it is possible to determine their frequency. For instance, two recently added variants to the INGD were in the *IDH3A* gene. These variants were reported in compound heterozygosity in an Ashkenazi patient affected with retinitis pigmentosa with pseudocoloboma.⁸ One of these *IDH3A* variants was present in gnomAD with a carrier

frequency of 0.91% among Ashkenazi Jews, significantly higher than in other populations and thus was classified as an Ashkenazi Jewish assumed founder mutation (**Supplementary Table 4**).

Recently Rivas *et al.*⁹ published a direct analysis of the data from the network of exome sequencing that developed the data set of Ashkenazi Jews included in gnomAD. This work identified 49 ClinVar alleles with a frequency of at least 0.002 enriched among Ashkenazi Jews. The methods of the analysis and the criteria of inclusions were different from those in the present study and 7 of these 49 variants were not included in the INGD. Of the missing INGD variants only 1 was classified as a pathogenic/likely pathogenic variant (*CNGA3*; c.101+1G>A). The other 6 included 1 that is not related to a disease (*PRB3*, p.Arg49Cys), 3 risk-associated variants (*NOD2*, p.Ala612Thr; *CHEK2*, p.Ser428Phe; and *CFTR*, p.Leu997Phe), and 2 that were reported in ClinVar as “conflicting interpretations of pathogenicity” (*USH2A*, p.Arg4192His and *HGSNAT*, p.Ala615Thr).

Another significant observation made because of the availability of a large sequencing database of healthy individuals is that some of the mutations related to clinical recessive disorders found among Ashkenazi Jews at a frequency significantly higher than among other population were observed in homozygosity among presumably healthy individuals. This phenomenon has been reported previously and allowed either reclassification of variants that were initially thought to be pathogenic or identification of rare individuals who are resilient to the corresponding disease.^{10,11} A recent analysis of the EXAC data set identified 113 variants with sufficient support for pathogenicity among healthy individuals for whom a childhood disorder was anticipated based on concordance with the disease inheritance.¹² Sixteen of these variants were found among Ashkenazi Jews at a frequency significantly higher than among other populations mostly in disorders in which the homozygous individuals are either undiagnosed or if diagnosed consider themselves healthy. This was the case for variants related to pentosuria, factor XI deficiency, familial hypercholesterolemia, familial Mediterranean fever, foveal hypoplasia, *GJB2*-related hearing defect, Gaucher disease, hyperphenylalaninemia, and short-chain acyl CoA dehydrogenase deficiency. However, there were five other disorders in which the variant was considered fully penetrant yet a homozygous individual was reported (dehydrolipoamide dehydrogenase deficiency, carnitine palmitoyl transferase deficiency type II, Usher syndrome type IIIA, glycogen storage disease type 1a, and oculocutaneous albinism). For instance, the variant [c.247C>T; p.Arg83Cys] in *G6PC* that has been reported in patients with glycogen storage disease type 1a results in no enzymatic activity¹³ and was considered as fully penetrant. This variant was reported in 151 alleles in gnomAD, for which 68 of 10,152 Ashkenazi Jews included the only homozygote (**Supplementary Results**). Although one may consider the possibilities of sequencing errors or underdiagnosis, this finding raises the possibility of protective factors that should be further investigated.^{10,12}

There were more than one assumed founder mutations in 27 of the 128 genes (21.1%) in which the 168 Ashkenazi assumed founder mutations were located (**Supplementary Tables 3 and 4**). For almost all these 27 genes, the existence of multiple founder mutations was well known and the reason for their co-occurrence has been the subject of debates, in particular concerning the genes related to lysosomal disorders. It has been argued that the high frequency of Tay–Sachs, Niemann–Pick, mucopolipidosis IV, and Gaucher as well as the existence of more than one founder mutation in each of the diseases is secondary to a past heterozygote selective advantage.^{2,14} Another possibility favored by others is random genetic drift to explain the primary determinant of disease mutations in the Ashkenazi population.^{2,15} A new observation made in the comparison between the INGD and gnomAD is that 7 of the 14 *PAH* variants are putative founder mutations with a total carrier frequency of 5.23%. Phenylketonuria is very rare among Ashkenazi Jews and the 14 *PAH* variants reported among them are related mainly to hyperphenylalaninemia.¹⁶ High frequencies of variants in *PAH* as well of a large number of assumed founder variants have been observed in various regions of the world and a heterozygous advantage was suggested.^{17,18} The carrier frequency of *PAH* variants among Ashkenazi Jews as well as the number of founding variants are also in favor of a selective advantage to carriers of *PAH* variants and cannot be explained by random genetic drift.

The information on the 168 Ashkenazi assumed founder mutations and their frequency in the population is important for effective straightforward diagnosing of affected individuals as well as for targeted carrier screening for reproductive decision-making.^{19,20} One issue that needs additional consideration upon planning such a comprehensive carrier-screening program is its clinical utility in particular regarding the severity and frequency of the disorders. Ashkenazi mutations with a low prevalence will have a diminished usefulness for improving health outcomes.¹⁹

The integration of data from gnomAD and the INGD together with the clinical impact of each of the variants may be the platform for building two clinically distinct screening programs for the Ashkenazi population: one that focuses on disorders within the context of reproductive medicine, and a second one relevant for personal disease-modifying behavior.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

ACKNOWLEDGMENTS

Part of our work has received funding from the European Commission (GEN2PHEN project, FP7-200754; RD-Connect project, FP7-305444) and Golden Helix Foundation grants to G.P.P.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. DellaPergola S. World Jewish population, 2016. In: Dashefsky A, Sheskin IM (eds). *American Jewish Year Book*. Springer: Dordrecht, The Netherlands, 2016:253–332.
2. Ostrer H, Skorecki K. The population genetics of the Jewish people. *Hum Genet* 2013;132:119–127.
3. Zlotogora J, van Baal S, Patrinos GP. Documentation of inherited disorders and mutation frequencies in the different religious communities in Israel in the Israeli National Genetic Database. *Hum Mutat* 2007;28:944–949.
4. Zlotogora J, Patrinos GP. The Israeli National Genetic Database: a 10-year experience. *Hum Genomics* 2017;11:5.
5. Lek M, Karczewski KJ, Minikel EV et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–291.
6. Shi L, Webb BD, Birch AH, et al. Comprehensive population screening in the Ashkenazi Jewish population for recurrent disease-causing variants. *Clin Genet* 2016;9:599–604.
7. Goodman, RM. *Genetic Disorders Among the Jewish People*. Johns Hopkins University Press: Baltimore, MD, 1979.
8. Pierrache LHM, Kimchi A, Ratnapriya R, et al. Whole-exome sequencing identifies biallelic IDH3A variants as a cause of retinitis pigmentosa accompanied by pseudocoloboma. *Ophthalmology* 2017;124:992–1003.
9. Rivas MA, Koskela J, Huang H, et al. Insights into the genetic epidemiology of Crohn's and rare diseases in the Ashkenazi Jewish population. *bioRxiv* 2016;077180.
10. Chen R, Shi L, Hakenberg J, et al. Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nat Biotechnol* 2016;34:531–538.
11. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet* 2013;132:1077–1130.
12. Tarailo-Graovac M, Zhu JYA, Matthews A, van Karnebeek CDM, Wasserman WW. Assessment of the ExAC data set for the presence of individuals with pathogenic genotypes implicated in severe Mendelian pediatric disorders. *Genet Med*; e-pub ahead of print 4 May 2017.
13. Lei KJ, Pan CJ, Shelly LL, Liu JL, Chou JY. Identification of mutations in the gene for glucose-6-phosphatase, the enzyme deficient in glycogen storage disease type 1a. *J Clin Invest* 1994;93:1994–1999.
14. Zlotogora J, Bach G. The possibility of a selection process in the Ashkenazi Jewish population. *Am J Hum Genet* 2003;73:438–440.
15. Risch N, Tang H, Katzenstein H, Ekstein J. Geographic distribution of disease mutations in the Ashkenazi Jewish population supports genetic drift over selection. *Am J Hum Genet* 2003;72:812–822.
16. Bercovich D, Elimelech A, Yardeni T, et al. A mutation analysis of the phenylalanine hydroxylase (PAH) gene in the Israeli population. *Ann Hum Genet* 2008;72:305–309.
17. Carter KC, Byck S, Waters PJ, et al. Mutation at the phenylalanine hydroxylase gene (PAH) and its use to document population genetic variation: the Quebec experience. *Eur J Hum Genet* 1998;6:61–70.
18. Krawczak M, Zschocke J. A role for overdominant selection in phenylketonuria? Evidence from molecular data. *Hum Mutat* 2003;21:394–397.
19. Baskovich B, Hiraki S, Upadhyay K, et al. Expanded genetic screening panel for the Ashkenazi Jewish population. *Genet Med* 2016;18:522–528.
20. Zlotogora J, Meiner V. Ashkenazi carrier screening for reproductive planning: is this what we planned for? *Genet Med* 2016;18:529.