The power of large numbers

Frequencies of rare pathogenic mutations in the 23andMe database

E.R. Chang, D.A. Hinds, N. Eriksson, J.M. Macpherson, C.B. Do, B.T. Naughton

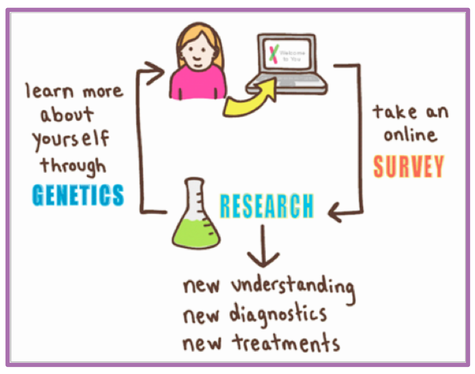
23andMe, Inc., Mountain View, CA

Introduction

23andMe, with over 180,000 genotyped customers, is in a unique position to accurately estimate the carrier frequencies of rare pathogenic mutations. Accurate, population-specific estimates of mutation frequencies facilitate effective and efficient carrier screening. We calculated seven population-specific mutation frequencies for each of 191 rare, autosomal recessive mutations; of the 1337 population-specific mutation frequencies calculated, nine are particular interest and are reported here. We found that several mutations that were thought to be associated only with certain populations are also present in additional populations at comparable or higher frequencies.

23andMe Research

23andMe is a unique research platform that uses both genetic (from customers' saliva) and phenotypic data (from customers' survey responses) to make new scientific discoveries. We are also committed to sharing these new findings with the customers and with the broader research community.



Methods

Participants were 23andMe customers who had consented to use of their data for research. With the exception of about 13,000 individuals recruited for disease studies, the 23andMe database is unselected for any particular disease or condition.

We used genetically-determined ancestry to place customers into one of eight categories: southern European, northern European, Latino, African American, South Asian, eastern Asian, Ashkenazi Jewish and other. No customer was included in more than one category.

To create a collection of relatively independent samples, we filtered out close relatives. For every pair of relatives more closely related than second cousins, we removed one individual. The resulting collection of study participants totaled 108,300 customers. The number of customers falling into each ancestry category is given in Table 1.

Table 1. Genetically-inferred ancestry of the 23andMe customers included in this study.

Ancestry category	Approximate number of customers included in study
Northern European	72,900
Southern European	6,000
Ashkenazi Jewish	7,200
Latino	4,700
African American	8,400
South Asian	3,000
Eastern Asian	6,100

For each of the 191 mutations covered in 23andMe's 48 carrier status reports, the population specific frequencies were calculated. These observed frequencies were compared to frequencies reported in the literature.

Results

For mutations predominately reported in a given population, customers carrying the mutation typically had the expected ancestry. There were, however, some surprises. Occasionally a mutation thought to affect people of one ancestry was observed in customers of another ancestry. Sometimes a mutation frequency in our database differed from the frequency previously reported for a given population (see Table 2).

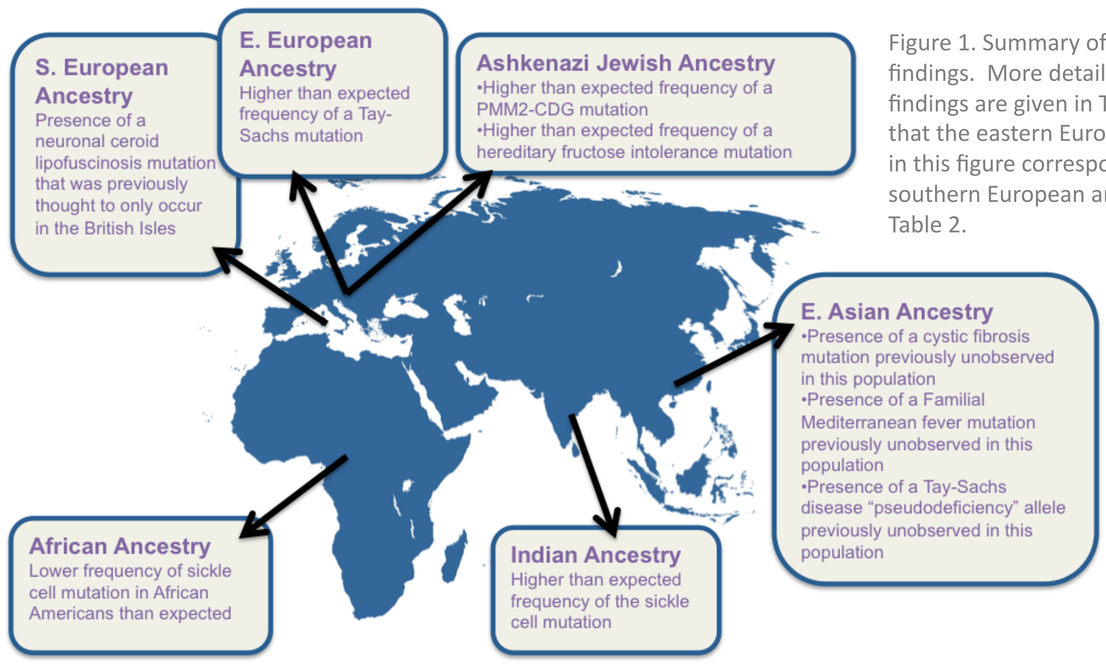


Figure 1. Summary of unexpected findings. More details on these findings are given in Table 2. Note that the eastern European ancestry in this figure corresponds to southern European ancestry in Table 2.

Table 2. Mutations and frequencies referred to in Figure 1. The symbol † denotes that the observed frequency was higher than previously reported, the symbol * means that the frequency was lower than previously reported, and the symbol ‡ means that this is the first observation of the given mutation the specified population (as far as we can tell from the literature).

Disease	Mutation	S. European	N. European	Latino	African	S. Asian	E. Asian	Ashkenazi Jewish
Sickle cell anemia	Hb S mutation in the HBB gene	0.05% (3 carriers)	0.02% (16 carriers)	0.44% (21 carriers)	**6.82%* (576 carriers)**	**0.43%† (13 carriers)**	0.00% (0 carriers)	0.01% (1 carrier)
Cystic fibrosis	2184delA mutation in the CFTR gene	0.02% (1 carrier)	0.01% (6 carriers)	0.02% (1 carrier)	0.01% (1 carrier)	0.00% (0 carriers)	**0.03%‡ (2 carriers)**	0.00% (0 carriers)
Congenital disorder of glycosylation, type 1a	R141H mutation in the PMM2 gene	1.03% (43 carriers)	1.09% (532 carriers)	0.83% (28 carriers)	0.34% (25 carriers)	0.10% (2 carriers)	0.00% (0 carriers)	**1.41%† (65 carriers)**
Familial Mediterranean fever	R761H mutation in the MEFV gene	0.07% (4 carriers)	0.01% (8 carriers)	0.02% (1 carrier)	0.00% (0 carriers)	0.00% (0 carriers)	**0.25%‡ (15 carriers)**	0.01% (1 carrier)
Hereditary fructose intolerance	delta4E4 mutation in the ALDOB gene	0.14% (6 carriers)	0.06% (29 carriers)	0.06% (2 carriers)	0.01% (1 carrier)	0.00% (0 carriers)	0.00% (0 carriers)	**0.37%† (17 carriers)**
Neuronal ceroid lipofuscinosis	T75P mutation in the PPT1 gene	**0.05% ‡ (2 carriers)**	0.03% (14 carriers)	0.00% (0 carriers)	0.00% (0 carriers)	0.00% (0 carriers)	0.00% (0 carriers)	0.00% (0 carriers)
Tay-Sachs disease	G269S mutation in the HEXA gene	**0.21%† (12 carriers)**	0.06% (43 carriers)	0.00% (0 carriers)	0.01% (1 carrier)	0.00% (0 carriers)	0.00% (0 carriers)	0.19% (13 carriers)
	R249W mutation in the HEXA gene ("pseudodeficiency" allele)	0.05% (3 carriers)	0.06% (40 carriers)	0.02% (1 carrier)	0.08% (7 carriers)	0.00% (0 carriers)	**0.22%‡ (13 carriers)**	0.00% (0 carriers)

Discussion

Using the power of a large database, we were able to discover rare pathogenic mutations in unexpected populations.

More research is needed to establish why we observed these mutations in new populations or at higher or lower than reported frequencies. There are several possible sources for these discrepancies:

- Unaccounted for ancestry from the population where the mutation is typically found
- Differences between self-reported and genetically-inferred ancestry
- Small size of previous studies
- Differences between ancestry of 23andMe customers and of participants in previous studies

We will further investigate some of these possibilities using (1) local ancestry inference at the site of the mutations and (2) surveys comparing self-reported and genetic-based ancestry.

Regardless of the origin of these new occurrences, the implication for carrier screening remains the same. **Broad carrier screening panels, rather than narrow population-specific carrier screening panels, are the best way to ensure that all carriers of rare pathogenic mutations are alerted to their carrier status**.

Acknowledgments

We thank 23andMe's customers who consented to participate in research for enabling this study. We also thank the employees of 23andMe who contributed to the development of the infrastructure that made this research possible.

References and Resources

1. Akin H, Onay H, Turker E, Cogulu O, Ozkinay F. MEFV mutations in patients with Familial Mediterranean Fever from the Aegean region of Turkey. Mol Biol Rep. 2010;37(1):93-8.
2. Ashley-koch A, Yang Q, Olney RS. Sickle hemoglobin (HbS) allele and sickle cell disease: a HuGE review. Am J Epidemiol. 2000;151(9):839-45.
3. Bender, MA, Hobbs W. (Updated May 17, 2012). Sickle Cell Disease. In: GeneReviews (database online). Available at http://www.ncbi.nlm.nih.gov/books/NBK1377/. Accessed October 19, 2012.
4. Booth DR, Lachmann HJ, Gillmore JD, Booth SE, Hawkins PN. Prevalence and significance of the familial Mediterranean fever gene mutation encoding pyrin Q148. QJM. 2001;94(10):527-31.
5. Cao Z, Natowicz MR, Kaback MM, et al. A second mutation associated with apparent beta-hexosaminidase A pseudodeficiency: identification and frequency estimation. Am J Hum Genet. 1993;53(6):1198-205.
6. Cao Z, Petroulakis E, Salo T, Triggs-raine B. Benign HEXA mutations, C739T(R247W) and C745T (R249W), cause beta-hexosaminidase A pseudodeficiency by reducing the alpha-subunit protein levels. J Biol Chem. 1997;272(23):14975-82.
7. Coffee EM, Yerkes L, Ewen EP, Zee T, Tolan DR. Increased prevalence of mutant null alleles that cause hereditary fructose intolerance in the American population. J Inherit Metab Dis. 2010;33(1): 33-42.
8. Das AK, Becerra CH, Yi W, et al. Molecular genetics of palmitoyl-protein thioesterase deficiency in the U.S. J Clin Invest. 1998;102(2):361-70.
9. Lazarin GA, Haque IS, Nazareth S, et al. An empirical estimate of carrier frequencies for 400+ causal Mendelian variants: results from an ethnically diverse clinical sample of 23,453 individuals. Genet Med. 2012.
10. Matthijs G, Schollen E, Bjursell C, et al. Mutations in PMM2 that cause congenital disorders of glycosylation, type Ia (CDG-Ia). Hum Mutat. 2000;16(5):386-94.
11. Navon R, Kolodny EH, Mitsumoto H, Thomas GH, Proia RL. Ashkenazi-Jewish and non-Jewish adult GM2 gangliosidosis patients share a common genetic defect. Am J Hum Genet. 1990;46(4): 817-21.
12. Santer R, Rischewski J, Von weihe M, et al. The spectrum of aldolase B (ALDOB) mutations and the prevalence of hereditary fructose intolerance in Central Europe. Hum Mutat. 2005;25(6):594.
13. Shukla RN, Solanki BR, Parande AS. Sickle cell disease in India. Blood. 1958;13(6):552-8.
14. Watson MS, Cutting GR, Desnick RJ, et al. Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel. Genet Med. 6(5):387-91.