Genetics inMedicine | ORIGINAL RESEARCH ARTICLE

Open

An empirical estimate of carrier frequencies for 400+ causal Mendelian variants: results from an ethnically diverse clinical sample of 23,453 individuals

Gabriel A. Lazarin, MS[1], Imran S. Haque, PhD[1], Shivani Nazareth, MS[1], Kevin Iori, BS[1], A. Scott Patterson, MA[1], Jessica L. Jacobson, MD[1,2], John R. Marshall, MD[1,3], William K. Seltzer, PhD, FACMG[1], Pasquale Patrizio, MD[4], Eric A. Evans, PhD[1] and Balaji S. Srinivasan, PhD[1,5,6]

Purpose: Recent developments in genomics have led to expanded carrier screening panels capable of assessing hundreds of causal mutations for genetic disease. This new technology enables simultaneous measurement of carrier frequencies for many diseases. As the resultant rank-ordering of carrier frequencies impacts the design and prioritization of screening programs, the accuracy of this ranking is a public health concern.

Methods: A total of 23,453 individuals from many obstetric, genetics, and infertility clinics were referred for routine recessive disease carrier screening. Multiplex carrier screening was performed and results were aggregated for this study.

Results: Twenty-four percent of individuals were identified as carriers for at least one of 108 disorders, and 5.2% were carriers for multiple disorders. We report tabulations of carrier frequency by self-identified ethnicity and disease.

Conclusion: To our knowledge, this study of a large, ethnically diverse clinical sample provides the most accurate measurements to date of carrier frequencies for hundreds of recessive alleles. The study also yields information on the clinical considerations associated with routine use of expanded panels and provides support for a pan-ethnic screening paradigm that minimizes the use of "racial" categories by the physician, as recommended by recent guidelines.

Genet Med 2013:15(3):178–186

Key Words: carrier frequency; carrier screening; genetic testing; pan-ethnic; recessive disease

INTRODUCTION

Carrier screening is the practice of diagnosing individuals and couples at risk of conceiving children affected by recessive diseases. This practice is widespread and growing in importance, as carrier couples identified before conception may decide to pursue a variety of preventive measures, including *in vitro* fertilization with preimplantation genetic diagnosis, use of donor gametes, adoption, or early prenatal diagnostic testing.

In 2001 cystic fibrosis (CF) became the first disease recommended for routine carrier screening in the United States when the American Congress of Obstetrics and Gynecology (ACOG) and the American College of Medical Genetics (ACMG) jointly recommended that CF screening be offered to all individuals of Caucasian and Ashkenazi Jewish ethnicity and that it be "made available" to individuals of all other ethnic groups.[1] In 2005, in recognition of increasing population ethnic admixture, the ACOG Committee on Genetics stated it was reasonable to offer CF screening pan-ethnically.[2]

Routine carrier screening for spinal muscular atrophy (SMA) has become more common since 2008 when ACMG recommended that screening be offered pan-ethnically.[3] However, SMA screening is still not performed as routinely as CF testing,[4,5] in part because of a conflicting ACOG statement that SMA should not be offered to the general population.[6]

In the absence of family history, carrier screening for diseases beyond CF and SMA is typically offered based on stated ethnicity. For example, hemoglobinopathy carrier screening is offered to individuals of African, Southeast Asian, or Mediterranean ancestry,[7] whereas a panel of 4–16 diseases is routinely offered to individuals of Ashkenazi Jewish ancestry.[8–10]

However, the advent of modern genomics has changed the health economic calculus around genetic screening, with costs-per-megabase for DNA sequencing falling at a rate faster than Moore's Law would indicate.[11] Consequently, the amortized cost of carrier screening across the population for hundreds of genetic diseases is now far less than the cost of treating an affected child.[12,13,14] Genomics has also called into question the reification of "racial" categories in clinical guidelines, as high-throughput assays now allow a physician to prescribe the same test for every patient, regardless of ethnicity, with little or no increase in cost.

The first two authors contributed equally to this work. The last two authors co-advised this work.

[1]Department of Genetics, Counsyl, South San Francisco, California, USA; [2]Department of Pathology, New York University, New York, New York, USA; [3]Department of Obstetrics and Gynecology, David Geffen School of Medicine at UCLA, Los Angeles, California, USA; [4]Department of Obstetrics, Gynecology, and Reproductive Sciences, Yale University School of Medicine, New Haven, Connecticut, USA; [5]Department of Computer Science, Stanford University, Stanford, California, USA; [6]Department of Statistics, Stanford University, Stanford, California, USA. Correspondence: Balaji Srinivasan (balajis@counsyl.com) and Eric Evans (eaevans@@counsyl.com)

Submitted 3 February 2012; accepted 1 August 2012; advance online publication 13 September 2012. doi:10.1038/gim.2012.114

Here we present clinical data from a large-scale deployment of expanded carrier screening across dozens of practices. In addition to providing perhaps the best estimates to date of a wide range of Mendelian allele frequencies, the deployment provides a concrete example of how genomics may reduce the role of "race" in biomedicine. It also identifies a number of severe, frequent, and easily detected disease alleles that have been omitted to date from population screening programs but that are easily included in expanded carrier testing panels.

MATERIALS AND METHODS

Multiplex platform

The screening platform (Universal Genetic Test, Counsyl, South San Francisco, CA) uses high-throughput genotyping to identify disease-causing variants and corresponding wild-type alleles. In total, 417 disease-causing mutations associated with 108 recessive diseases were assayed and interpreted via fluorescent sequences. We consider a subset (listed in **Supplementary Table S1** online) of the diseases identifiable via the platform. The analytical performance has been reported previously and compares favorably with traditional single-gene testing methods.[12] Our reported disease carrier frequencies are lower bounds, as there are variants for each disease that the assay platform does not test.

Study population

This study sample included 23,453 individuals. Referral sources included family practitioners, geneticists, genetic counselors, obstetricians, perinatologists, and reproductive endocrinologists. Documented informed consent, including consent to research, was required of all patients and is on record at our facility. Institutional review board exemption is applicable due to de-identification of the data presented (45 CFR part 46.101(b)(4)). Genetic counseling was made available at no cost to all tested individuals.

Funding was performed as fee-for-service, typically reimbursed through third-party payers.

Disease panels

Diseases identifiable via the test platform range in severity from mild to incompatible with life. Nearly every disease can be considered severe (associated with infant or child mortality), significant (associated with progressive disease and reduced life span), or requiring significant intervention or treatment. The remainder is composed of milder conditions characterized by high allele frequencies and reduced penetrances, e.g., factor V Leiden thrombophilia (OMIM 188055), glucose-6-phosphate dehydrogenase deficiency (OMIM 305900), *HFE*-associated hereditary hemochromatosis (OMIM 235200), prothombin thrombophilia (OMIM 176930), pseudocholinesterase deficiency (OMIM 177400), and two common *MTHFR* mutations (OMIM 607093).

As the testing platform permits customization of disease panels, some physicians chose to prescribe testing for a subset of the full 417-mutation panel. In addition, ongoing improvements to the platform resulted in additions and refinements of various disease mutations. The sum effect of these changes is that the number of individuals tested for a given mutation varies moderately, as shown in the first column of **Supplementary Table S1** online. All OMIM numbers are also noted in **Supplementary Table S1** online.

RESULTS

Data were available from 23,453 screened patients. Routine screening for a possible carrier state was the indication in all individuals. Screening for other indications, including potential gamete donor, infertility, and personal or family history, was also performed. Individuals with those indications were excluded from this study.

Population demographics

Self-reported ethnicity is summarized in **Table 1**. Caucasians constituted 60.6%, and 75.0% were female. Most samples came from within the United States; a minority came from Australia, Canada, England, or New Zealand.

The median age was 33.0 years and mean age was 33.63 years. Most (n = 17,865) of the individuals were from 21–39 years of age. A small percentage was under 21 years (1.6%) or over 45 years (2.8%).

Disease carrier states

Alleles associated with 96 recessive diseases (excluding mild conditions) were identified in our 23,453 patients (**Supplementary Table S1** online). Among the mild conditions, the most common was *MTHFR* deficiency, with a frequency of 1 in 1.9. Because these conditions typically have limited reproductive decision-making significance, they will not be considered further in this analysis. All subsequent analysis considers only those diseases listed in **Supplementary Table S1** online.

Of the total sample, 24.0% of individuals (n = 5,633) were heterozygous for at least one non-mild condition. A total of 7,067 heterozygous states were identified. Carrier statistics are fully reported in **Supplementary Table S1** online.

Seventy-eight individuals were identified as homozygotes or compound heterozygotes for the following conditions: α-1-antitrypsin deficiency (n = 38), cystic fibrosis (n = 9), *GJB2*-related DFNB1 nonsyndromic hearing loss and deafness (n = 6), factor XI deficiency (n = 5), Gaucher disease (n = 4), familial Mediterranean fever (n = 3), carnitine palmitoyltransferase II deficiency (n = 2), medium chain acyl-CoA dehydrogenase deficiency (n = 2), sickle cell disease (n = 2), short chain acyl-CoA dehydrogenase deficiency (n = 2), achromatopsia (n = 1), β-thalassemia (n = 1), hexosaminidase A deficiency (n = 1), familial dysautonomia (n = 1), lipoamide dehydrogenase deficiency (n = 1), Niemann–Pick disease type C (n = 1), Pompe disease (n = 1), and spinal muscular atrophy (n = 1). The published specificity of the testing method suggests that these are "true positives" and merit further examination of clinical correlations.[12] Review of clinical notes found at least two individuals with previously known diagnoses, for Gaucher disease and deafness. Another individual reported a history of sickle cell disease but did not specify whether this was familial or personal. All others did not report diagnosis information in their test requisitions and selected the routine carrier screening indication. For asymptomatic

Table 1 Carrier statistics categorized by self-reported ethnicity

Ethnicity	Screened, *n*	Ethnicity as % of total group (%)	Carrier frequency		
			n	Percentage (%)	1 in ___
All	23,453	100.0	5,633	24.0	4.2
African or African-American	1,193	5.1	188	15.8	6.3
Ashkenazi Jewish	2,410	10.3	1,050	43.6	2.3
East Asian	1,121	4.8	95	8.5	11.8
Finnish	34	0.1	16	47.1	2.1
French Canadian or Cajun	193	0.8	63	32.6	3.1
Hispanic	2,302	9.8	423	18.4	5.4
Middle Eastern	512	2.2	76	14.8	6.7
Mixed or other Caucasian	6,744	28.8	1,650	24.5	4.1
Native American	48	0.2	3	6.3	16.0
Northern European	6,171	26.3	1,630	26.4	3.8
Pacific Islander	77	0.3	8	10.4	9.6
South Asian	1,123	4.8	144	12.8	7.8
Southeast Asian	462	2.0	59	12.8	7.8
Southern European	1,063	4.5	228	21.4	4.7

individuals with conditions such as α-1-antitrypsin deficiency, results could be used to guide appropriate surveillance.

Carrier rate variability by ethnicity

Frequencies of positive results varied by ethnic group (Table 1). On average, 24.0% of individuals were positive for at least one condition. When stratified by self-reported ethnicity, this frequency ranged from 43.6% of Ashkenazi Jewish individuals to 8.5% of East Asians. For ethnic groups like the Ashkenazi Jewish, this frequency is unsurprising given the availability of screening for many conditions found in this group.

Multiple-disease carriers

Some individuals were heterozygous for multiple disorders (Table 2), with ~5.2% (n = 1,210) found to be carriers of two or more disorders. Most were heterozygous for only two conditions (4.3% of all screenees and 83.9% of multiple-disease carriers), although a small number were carriers of three (0.7 and 13.8%) or more than three conditions (0.1 and 2.3%). Ashkenazi Jewish individuals were most frequently identified as multiple carriers, with 13.3% of all tested Ashkenazi Jews carrying more than one genetic disorder. These values are unsurprising, as population geneticists have long known that individuals carry on average 4–5 recessive lethal alleles.[10,15] Put another way, the average number of positive results for recessive lethals in this panel is only $(0 \times 0.694 + 1 \times 0.244 + 2 \times 0.053 + 3 \times 0.008 + 4 \times 0.001) = 0.378$; sequencing the entire genome of each patient would reveal ~10 times as many lethal recessives on average.

Carrier couples

Some pairs of patients identified themselves as couples. In this set, we found several "carrier couples" in which both partners were heterozygous carriers for the same condition. Table 3 lists disorders identified in carrier couples at elevated risk of passing the carried disorder to their children.

Comparison to traditional screening guidelines

Among carrier states we detected, 76.7% were for diseases not included in ACOG carrier screening guidelines, whereas 69.0% were for diseases not included in ACMG guidelines. The ACMG's inclusion of spinal muscular atrophy accounted for most of the discrepancy.

We detected 433 individuals who would not have been identified as disease carriers in accordance with conventional ethnicity-based screening paradigms (Table 4). For example, a non-Jewish carrier for familial dysautonomia would not be identified if following guidelines that recommend screening only for those with Jewish ancestry—26.3% of familial dysautonomia carriers in our dataset did not report Jewish ancestry. Rates of non-Jewish carriers are even higher, an average of 40.7%, for diseases recommended by the ACMG. Furthermore, in Table 5, we show the 10 diseases with the highest carrier frequency for each population with more than 500 tested individuals, along with ACOG/ACMG screening recommendations. In general, many prevalent diseases in each population are not currently recommended for screening by either the ACOG or the ACMG.

DISCUSSION

Data quality and limitations

This study represents the first large-scale analysis of patients undergoing carrier screening for an extensive list of recessive disorders in a clinical setting. As such, the data provide the first report of carrier frequencies for many rare disorders in multiple ethnic groups and are an important resource for guiding diagnosis, treatment, and prevention of Mendelian disease.

Because our assay was based on targeted genotyping, we were able to find only carriers of particular disease-causing mutations, not all carriers of each disease. (Patient reports explain that genotyping can only establish risk reduction, not risk elimination.) Consequently, the disease frequencies we report should be considered lower bounds, as there are other

Table 2 Carriers of multiple diseases categorized by self-reported ethnicity

	Carrier, one disease		Carrier, two diseases		Carrier, three diseases		Carrier, four or more diseases		Carrier, two or more diseases		
	n	% Total[a]	*n*	% Total[a]	*n*	% Total[a]	*n*	% Total[a]	*n*	% Total[a]	% Carriers
All	4,423	18.9	1,015	4.3	167	0.7	28	0.1	1,210	5.2	21.5
Mixed or other Caucasian	1,315	19.5	289	4.3	39	0.6	7	0.1	335	5.0	20.3
Northern European	1,320	21.4	266	4.3	42	0.7	2	0.0	310	5.0	19.0
Ashkenazi Jewish	729	30.2	244	10.1	61	2.5	16	0.7	321	13.3	30.6
Hispanic	367	15.9	48	2.1	8	0.3	0	0.0	56	2.4	13.2
African or African-American	144	12.1	41	3.4	2	0.2	1	0.1	44	3.7	23.4
South Asian	109	9.7	31	2.8	3	0.3	1	0.1	35	3.1	24.3
East Asian	79	7.0	13	1.2	2	0.2	1	0.1	16	1.4	16.8
Southern European	180	16.9	43	4.0	5	0.5	0	0.0	48	4.5	21.1
Middle Eastern	62	12.1	11	2.1	3	0.6	0	0.0	14	2.7	18.4
Southeast Asian	44	9.5	14	3.0	1	0.2	0	0.0	15	3.2	25.4
French Canadian or Cajun	54	28.0	8	4.1	1	0.5	0	0.0	9	4.7	14.3
Pacific Islander	7	9.1	1	1.3	0	0.0	0	0.0	1	1.3	12.5
Native American	2	4.2	1	2.1	0	0.0	0	0.0	1	2.1	33.3
Finnish	11	32.4	5	14.7	0	0.0	0	0.0	5	14.7	31.3

[a]The "% Total" refers to the percentage of individuals in that row's population carrying the specified number of diseases. For example, of all screened Northern Europeans, 1,320 or 21.4% carried one disease, 266 or 4.3% carried two diseases, and 5.0% carried two or more diseases. Of all carriers who were Northern European, 19.0% carried two or more diseases.

Table 3 Carrier couples identified

Disease	Number of couples
α-1-Antitrypsin deficiency	47
Cystic fibrosis	27
Spinal muscular atrophy	15
Sickle cell disease	10
Familial Mediterranean fever	5
GJB2-related DFNB1 nonsyndrome hearing loss and deafness	5
Smith–Lemli–Opitz syndrome	3
Factor XI deficiency	3
Familial dysautonomia	2
β-Thalassemia	2
Gaucher disease	2
Tay–Sachs disease	1
Sulfate transporter-related osteochondrodysplasia	1
Medium chain acyl-CoA dehydrogenase deficiency	1
Short chain acyl-CoA dehydrogenase deficiency	1
Carnitine palmitoyltransferase II deficiency	1
Mucolipidosis IV	1

Carrier couples are defined as pairs of individuals identified as partners by self-report, both sharing heterozygosity for at least one disease. This table likely represents a subset of carrier couples we have identified, because most were tested as individuals and we therefore did not know who the partner was.

disease-causing alleles not surveyed here. Further examination of each individual disease on an allelic basis will help elucidate important data points such as overall carrier frequency, disease penetrance, and clinical sensitivity.

Although our targeted genotyping data is not comprehensive, it is a significant improvement on carrier rate estimation based on disease incidence rates. Such estimation suffers from statistical challenges such as underreporting of mild phenotypes and embryonic/child mortality affecting the assumptions of Hardy–Weinberg equilibrium.[24,29,30] In contrast, our data come from directly assaying carrier genotypes rather than relying on mapping from phenotypic incidence to genotype.

Comparison to previously published carrier frequencies

Table 5 shows, for each ethnic group with $n > 500$ in our data set, the 10 most commonly detected diseases, along with our estimates and literature estimates for carrier frequency. For a handful of disease/ethnicity pairings, we were unable to find a previously published estimate. For most diseases, particularly in the well-studied Northwestern European and Ashkenazi Jewish populations, our estimates are similar to previously published estimates of carrier frequency. However, there are some notable outliers.

We observed significantly higher carrier frequencies than expected for severe diseases in several populations. In particular, we find a cystic fibrosis carrier frequency of 1 in 40 among south Asians (95% confidence interval 1 in 29–64), in marked

Table 4 Carrier statistics among ACOG- and ACMG-targeted populations

	Targeted ethnic group statistics[a]		Pan-ethnic statistics[b]		
		Carrier frequency in targeted population	Carrier frequency in all populations	Carriers not from targeted ethnic group	
Disease	**Population**	***n* Carriers/ *n* total**	***n* Carriers/ *n* tested**	***n* Nontargeted carriers/ *n* total carriers**	**%**
Sickle cell disease (Hb S mutation)[c]	African-American	89/1,121	145/21,360	56/145	38.6
β-Thalassemia[c]	African-American	27/1,121	163/21,443	72/163	44.2
	Southern European	17/999			
	South Asian	33/1,068			
	Southeast Asian	14/417			
Canavan disease[c,d]	AJ[e]	43/2,384	71/21,581	28/71	39.4
Familial dysautonomia[c,d]	AJ	56/2,385	76/21,592	20/76	26.3
Tay–Sachs disease/ Hex A deficiency[c,d]	AJ	90/2,386	151/21,985	61/151	40.4
Fanconi anemia group C[d]	AJ	24/2,385	44/21,476	20/44	45.5
Niemann–Pick type A[d]	AJ	25/2,385	33/21,476	8/33	24.2
Bloom syndrome[d]	AJ	23/2,384	35/21,466	12/35	34.3
Mucolipidosis IV[d]	AJ	18/2,385	36/21,476	18/36	50.0
Gaucher disease[d]	AJ	142/2,384	280/21,473	138/280	49.3
Cystic fibrosis[c,d]	Caucasian	521/14,155			
	AJ	90/2,402			
	African-American	9/1,188	N/A—pan-ethnic screen		
	Asian	16/2,777			
	Hispanic	33/2,294			
Spinal muscular atrophy[d]	Caucasian	285/14,005			
	AJ	40/2,371			
	African-American	10/1,178	N/A— pan-ethnic screen		
	Asian	34/2,745			
	Hispanic	28/2,274			

ACMG, American College of Medical Genetics; ACOG, American Congress of Obstetrics and Gynecology; AJ, Ashkenazi Jewish.
[a]Statistics for ethnic groups targeted for screening by ACOG and ACMG guidelines. [b]Statistics for all 23,453 individuals. [c]Conditions included in ACOG guidelines. [d]Conditions included in ACMG guidelines. [e]Cystic fibrosis analysis here accounts for only 23 mutations recommended by ACMG. Our total panel includes 100 mutations and these data are reported in **Supplementary Table S1** online.

contrast to the published 1 in 118 rate.[21] This corroborates reports that cystic fibrosis is under-reported in the South Asian population.[22] Another notable outlier in both Ashkenazi Jews and East Asians is carnitine palmitoyltransferase II deficiency, which appears as a notably frequent disease (1 in 43 in Ashkenazi Jews and 1 in 378 in East Asians), although the literature marks it as very rare.[23] Finally, we detect Smith–Lemli–Opitz syndrome (SLOS) at a significantly higher than expected rate in all European populations, in Hispanics, and Middle Easterners (1 in 123 from literature vs. 1 in 68 here; 95% confidence interval 1 in 60–78). Prior work suggests that a higher carrier frequency for SLOS than that computed from its birth prevalence is reasonable, because mutations in *DHCR7*, the causative gene for SLOS, may cause significant fetal mortality.[24]

In the African-American population, we observe significantly reduced carrier frequencies relative to literature statistics for 7 of the top 10 diseases, other than sickle cell disease, α-1-antitrypsin deficiency, and cystic fibrosis. Comparison to prior data in this population is confounded by significantly varying levels of ethnic admixture in the African-American population; it is possible that our study population is genetically distinct at these loci relative to prior studies.[16] Although we have not attempted to estimate admixture proportions in our study patients, it may be an interesting avenue for future research.

Similarly, in the East Asian population, we find much lower carrier frequencies for Pendred syndrome (1 in 252 vs. 1 in 51) and Pompe disease (1 in 366.5 vs. 1 in 112). We suspect two independent causes for the discrepancy. First, general-population screening data are not available for these diseases in this population; consequently, statistics are often derived from targeted study populations enriched in the mutations in question.[25] Second, the East Asian population displays sufficient genetic diversity (e.g., population structure among Han Chinese, Koreans, and Japanese) to have confounded results

Table 5 Top 10 most common carrier frequencies by population

Disease	Counsyl frequency (1 in)	Counsyl 95% CI (1 in)	Literature frequency (1 in)	# Tested	Screening?
All populations (*N* = 23,453)					
α-1-Antitrypsin deficiency	13.1	12–14	11.5	15,484	×
Cystic fibrosis	27.8	26–30	31.7	23,369	Y
DFNB1	42.6	39–47	42.8	15,799	×
Spinal muscular atrophy	57.1	52–63	54	23,127	1
Familial Mediterranean fever	64.2	57–73	Unknown	15,854	×
Smith–Lemli–Opitz syndrome	68.2	60–78	123	15,825	×
Sickle cell disease/β-thalassemia	69.6	63–78	158	21,360	×
Gaucher disease	76.7	69–87	200	21,473	×
Factor XI deficiency	92.0	80–108	Rare	15,724	×
Achromatopsia	97.5	85–115	123	15,798	×
Northwestern Europe (*N* = 12,915)					
α-1-Antitrypsin deficiency	10.2	10–11	11.4	8,570	×
Cystic fibrosis	22.9	21–25	28	12,870	Y
DFNB1	46.0	40–54	33.2	8,735	×
Spinal muscular atrophy	49.9	45–57	47	12,730	1
Smith–Lemli–Opitz syndrome	50.3	44–59	123.9	8,750	×
MCAD deficiency	73.5	62–90	61.7	8,749	×
Hereditary fructose intolerance	90.1	75–113	81.1	8,744	×
Achromatopsia	91.0	76–114	123	8,734	×
Hereditary thymine-uraciluria	96.4	80–122	33	8,680	×
Familial Mediterranean fever	99.5	82–126	Unknown	8,756	×
Ashkenazi Jewish (*N* = 2,410)					
Factor XI deficiency	12.6	11–15	8	1,501	×
Familial Mediterranean fever	13.2	11–16	10.5	1,513	×
Gaucher disease	16.8	14–20	17	2,384	1
DFNB1	21.3	17–28	21	1,509	×
Cystic fibrosis	21.6	18–27	29	2,402	Y
α-1-Antitrypsin deficiency	24.2	19–32	16	1,452	×
Hexosaminidase A deficiency	26.9	22–34	27.4	2,366	Y
SCAD deficiency	29.1	23–40	Unknown	1,511	×
Familial dysautonomia	41.8	33–57	31	2,385	Y
CPT II deficiency	43.2	33–65	Rare/too few patients worldwide	1,512	×
Hispanic (*N* = 2,302)					
α-1-Antitrypsin deficiency	9.1	8–11	9.2	1,608	×
Cystic fibrosis	52.1	40–74	59	2,294	Y
DFNB1	67.6	48–114	100	1,623	×
Spinal muscular atrophy	81.2	59–130	68	2,274	1
Sickle cell disease/β-thalassemia	83.1	60–138	128	2,077	×
Pompe disease	134.7	85–326	100	1,616	×
CDG-Ia	135.4	86–328	Unknown	1,625	×
Smith–Lemli–Opitz syndrome	135.5	86–328	No reliable data	1,626	×
Familial Mediterranean fever	136.6	86–331	Unknown	1,639	×
Phenylalanine hydroxylase deficiency	162.8	99–462	Unknown	1,628	×

Confidence intervals computed as Wilson 95% score interval. Some listed diseases overlap in definition in terms of genetic variants. Specific variants tested for each disease listed here are given in Supplementary Table B. Please refer to Supplementary Table C for citations to previous literature. CDG-Ia, congenital disorder of glycosylation type Ia; CI, confidence intervals; CPT II deficiency, carnitine palmitoyltransferase II deficiency; DFNB1, *GJB2*-related DFNB 1 nonsyndromic hearing loss and deafness; MCAD deficiency, medium chain acyl-CoA dehydrogenase deficiency; MLC, megalencephalic leukoencephalopathy with subcortical cysts; SCAD deficiency, short chain acyl-CoA dehydrogenase deficiency; Y, both ACMG and ACOG recommend offering screening to this population; 1, one of ACMG/ACOG recommend offering screening to this population; ×, neither ACMG/ACOG recommend offering screening to this population.

Table 5 Continued on next page

Table 5 Continued

Disease	Counsyl frequency (1 in)	Counsyl 95% CI (1 in)	Literature frequency (1 in)	# Tested	Screening?
African American (*N* = 1,193)					
Sickle cell disease/β-thalassemia	9.7	8–12	10	1,121	1
α-1-Antitrypsin deficiency	29.2	21–50	37.3	672	×
Cystic fibrosis	74.2	50–149	84	1,188	Y
Pompe disease	112.5	61–800	59.7	675	×
Spinal muscular atrophy	117.8	72–334	72	1,178	1
Galactosemia	169.5	82–∞	94	678	×
Gaucher disease	213.2	109–4,770	34.6	1,066	×
DFNB1	226.0	99–∞	25.3	678	×
Hereditary fructose intolerance	226.0	99–∞	Unknown	678	×
Smith–Lemli–Opitz syndrome	339.0	127–∞	137.8	678	×
South Asia (*N* = 1,123)					
Achromatopsia	23.5	18–35	123	798	×
Sickle cell disease/β-thalassemia	30.1	23–45	25	1,055	1
Cystic fibrosis	40.1	29–64	118	1,122	Y
Spinal muscular atrophy	73.3	48–153	52	1,099	1
Hereditary thymine-uraciluria	99.8	58–371	33	798	×
DFNB1	99.9	58–372	148	799	×
Citrullinemia type 1	212.0	93–∞	Unknown	636	×
MLC	212.0	93–∞	Unknown	636	×
Biotinidase deficiency	266.3	116–∞	123	799	×
Tyrosinemia type I	266.3	116–∞	173	799	×
Eastern Asia (*N* = 1,121)					
DFNB1	22.2	17–34	33–200	756	×
Sickle cell disease/β-thalassemia	78.0	50 –178	50	1,014	1
Spinal muscular atrophy	85.5	55–195	59	1,111	1
Gaucher disease	96.9	59–275	Unknown	969	×
Achromatopsia	189.0	91–∞	123	756	×
Cystic fibrosis	223.6	115–5011	242	1118	Y
α-1-Antitrypsin deficiency	249.0	109–∞	450	747	×
Pendred syndrome	252.0	110–∞	51	756	×
Pompe disease	366.5	137–∞	112	733	×
CPT II deficiency	378.0	142–∞	Rare/too few patients worldwide	756	×
Southern Europe (*N* = 1,063)					
α-1-Antitrypsin deficiency	17.0	13–24	13.1	715	×
Cystic fibrosis	35.3	26–55	28	1,059	Y
DFNB1	37.2	26–67	33.2	745	×
Spinal muscular atrophy	40.3	29–66	47	1,049	1
Sickle cell disease/β-thalassemia	52.4	36–97	50	996	1
Familial Mediterranean fever	74.5	46–211	Unknown	745	×
Phenylalanine hydroxylase deficiency	74.6	46–211	50.5	746	×
Pompe disease	81.9	49–260	100	737	×
Smith–Lemli–Opitz syndrome	82.9	49–263	Unknown	746	×
Hereditary thymine-uraciluria	105.1	59–495	33	736	×

Confidence intervals computed as Wilson 95% score interval. Some listed diseases overlap in definition in terms of genetic variants. Specific variants tested for each disease listed here are given in Supplementary Table B. Please refer to Supplementary Table C for citations to previous literature. CDG-Ia, congenital disorder of glycosylation type Ia; CI, confidence intervals; CPT II deficiency, carnitine palmitoyltransferase II deficiency; DFNB1, *GJB2*-related DFNB 1 nonsyndromic hearing loss and deafness; MCAD deficiency, medium chain acyl-CoA dehydrogenase deficiency; MLC, megalencephalic leukoencephalopathy with subcortical cysts; SCAD deficiency, short chain acyl-CoA dehydrogenase deficiency; Y, both ACMG and ACOG recommend offering screening to this population; 1, one of ACMG/ACOG recommend offering screening to this population; ×, neither ACMG/ACOG recommend offering screening to this population.

Table 5 Continued on next page

Table 5 Continued

Disease	Counsyl frequency (1 in)	Counsyl 95% CI (1 in)	Literature frequency (1 in)	# Tested	Screening?
Middle East (*N* = 512)					
Familial Mediterranean fever	24.5	17–49	Variable (1/10–1/20)	392	×
Sickle cell disease/β-thalassemia	5.1	31–165	30 (but variable)	469	×
Hereditary thymine-uraciluria	54.9	31–256	33	384	×
Achromatopsia	55.1	31–257	123	386	×
Cystic fibrosis	63.1	37–234	91	505	Y
Spinal muscular atrophy	72.4	41–340	25	507	1
DFNB1	77.6	40–1627	83	388	×
Inclusion body myopathy 2	96.5	47–∞	15 (Iranian Jews), unknown in others	386	×
Hereditary fructose intolerance	97.0	47–∞	Unknown	388	×
Smith–Lemli–Opitz syndrome	129.3	57–∞	Rare	388	×

Confidence intervals computed as Wilson 95% score interval. Some listed diseases overlap in definition in terms of genetic variants. Specific variants tested for each disease listed here are given in Supplementary Table B. Please refer to Supplementary Table C for citations to previous literature. CDG-Ia, congenital disorder of glycosylation type Ia; CI, confidence intervals; CPT II deficiency, carnitine palmitoyltransferase II deficiency; DFNB1, *GJB2*-related DFNB 1 nonsyndromic hearing loss and deafness; MCAD deficiency, medium chain acyl-CoA dehydrogenase deficiency; MLC, megalencephalic leukoencephalopathy with subcortical cysts; SCAD deficiency, short chain acyl-CoA dehydrogenase deficiency; Y, both ACMG and ACOG recommend offering screening to this population; 1, one of ACMG/ACOG recommend offering screening to this population; ×, neither ACMG/ACOG recommend offering screening to this population.

in earlier genetic studies.[26] Our self-reported ethnicity data do not have sufficient detail to identify this structure.

We note that direct comparison between our results and the published literature can be difficult due to previously mentioned statistical considerations and differing population substructures. As an example, carrier frequencies in Middle Eastern subpopulations reported in previous literature can vary widely and could also be substantially different from frequencies in Americans of Middle Eastern origin. Therefore, the references in **Table 5** are not meant to present an exhaustive review but to provide context for comparison. The specific ethnic distributions merit further examination because robust data are unavailable for many diseases.

Comparison to ACOG/ACMG guidelines

ACOG and ACMG guidelines aim to identify couples with a minimum a priori carrier risk, based on ethnic background or general-population disease prevalence. The ACOG recommends at least a subset of patients be offered carrier screening for CF, hemoglobinopathies, Tay–Sachs disease, familial dysautonomia, and Canavan disease.[1,2,6–8] ACMG recommendations add SMA, Fanconi anemia Group C, Niemann–Pick disease type A, Bloom syndrome, mucolipidosis IV, and Gaucher disease type 1.

For their intended purposes, the ACOG and ACMG guidelines serve well. However, our results demonstrate that most of the heterozygous states we identified fall outside these guidelines (**Tables 4** and **5**). Because the current thought is that the average individual is heterozygous for at least five recessive diseases, this is not surprising.[15] By comparison, screening most Caucasians only for CF, as is recommended by the ACOG, yields a 4% positive rate. Focusing on a smaller number of conditions and on targeted ethnic groups means that most individuals at elevated risk are not identified.

In particular, the data show that a number of severe Mendelian disorders are more prevalent than commonly understood (**Table 5**), are often present outside their characteristic ethnic groups (**Table 4**), and are not covered by current screening guidelines (**Table 5**). For example, we find a SLOS carrier frequency of 1 in 68.2 in our overall population, almost twice the previously reported 1 in 123 carrier frequency.[27] Neither the ACOG nor the ACMG currently recommend offering screening for this severe condition. In contrast, the ACMG does recommend screening for SMA, with a similar worldwide carrier frequency of 1 in 57 (our data) or 1 in 54 (literature).[28] Indeed, in the Northwestern European population, we find indistinguishable carrier frequencies for SLOS and SMA: 1 in 50.3 and 1 in 49.9, respectively. Carnitine palmitoyltransferase II deficiency presents another illustration of a disease with a frequency worthy of consideration of wider screening. In our Ashkenazi Jewish population, the carrier rate was higher than several other conditions currently included in screening guidelines (**Table 5**).

Ethnicity-based screening model vs. universal screening

Historically, screening has focused on a limited disorder list primarily determined by self-reported ethnic group. Others[1,2,6–8,17] have noted that this model will soon be inadequate, given: (i) 2010 US Census data that demonstrate sharp increases in individuals reporting mixed racial ancestry, particularly among the younger population approaching reproductive age;[18] (ii) limitations of, and patient preferences against, use of racial and ethnic categorization in medicine;[19,20] (iii) the possibility of unknown or unreported ancestry, due to limited family history knowledge, adoption, or other factors; and (iv) the decreasing cost of panethnic screening due to advances in genomics.[11] Indeed, most participating clinics applied the same testing panel regardless of

ethnicity, phasing out the ethnicity-based screening paradigm in favor of a universal screening model.

We have presented perhaps the most accurate measurements to date of carrier frequencies for hundreds of recessive Mendelian alleles using a large, ethnically diverse clinical sample. In contrast, prior data suffer from the statistical irregularities inherent in aggregating the results of multiple studies conducted under different protocols. Those irregularities result not from mistakes by the authors. Instead, the relative ordering of disease frequencies can only be determined with confidence when they are all assessed on the same dataset and under the same conditions. This, to our knowledge, has never been accomplished on the scale of this study, where we have applied multiplex carrier testing to sample a large population at a large number of genetic loci simultaneously.

Implications of widespread expanded carrier screening merit further study. These may include increased partner testing via screening or sequencing methods, higher demand for genetic counseling services and genetics education for nongenetics providers, cost-effectiveness, and patient interest and anxiety.

Our data demonstrate that current ethnicity-based approaches of genetic screening are not optimally aligned with the real distribution of carrier frequencies for severe genetic disease. In particular, we find that numerous conditions, not currently suggested for screening by the ACOG and the ACMG, are prevalent; and that other diseases currently recommended for screening only in certain populations, such as sickle cell disease and Canavan disease, are in fact widely distributed outside their "home" populations. We believe that these facts should be considered alongside the rapidly decreasing cost of multilocus genetic testing in the design of future carrier screening guidelines.

SUPPLEMENTARY MATERIAL

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

DISCLOSURE

P.P., J.R.M., and J.L.J. are advisors to Counsyl. The remaining authors are employees at Counsyl.

REFERENCES

1. American College of Obstetricians and Gynecologists & American College of Medical Genetics. *Preconception and Prenatal Carrier Screening for Cystic Fibrosis. Clinical and Laboratory Guidelines*. American College of Obstetricians and Gynecologists: Washington, DC, 2001.
2. American College of Obstetricians and Gynecologists. Committee opinion no. 325: update on carrier screening for cystic fibrosis. *Obstet Gynecol* 2005;106:1465–1468.
3. Prior TW; Professional Practice and Guidelines Committee. Carrier screening for spinal muscular atrophy. *Genet Med* 2008;10:840–842.
4. Samons MS. *Early Experiences of Prenatal Genetic Counselors With Spinal Muscular Atrophy Carrier Testing [Thesis]*. Brandeis University: Waltham, MA, 2010.
5. Sims CA, Callum P, Ray M, Iger J, Falk RE. Genetic testing of sperm donors: survey of current practices. *Fertil Steril* 2010;94:126–129.
6. American College of Obstetricians and Gynecologists. Committee opinion no. 432: spinal muscular atrophy. *Obstet Gynecol* 2009;113:1194–1196.
7. American College of Obstetricians and Gynecologists. Practice bulletin no. 78: hemoglobinopathies in pregnancy. *Obstet Gynecol* 2007;109:229–237.
8. American College of Obstetricians and Gynecologists. Committee opinion no. 442: preconception and prenatal carrier screening for genetic diseases in individuals of Eastern European Jewish descent. *Obstet Gynecol* 2009;14:950–953.
9. Gross SJ, Pletcher BA, Monaghan KG; Professional Practice and Guidelines Committee. Carrier screening in individuals of Ashkenazi Jewish descent. *Genet Med* 2008;10:54–56.
10. Scott SA, Edelmann L, Liu L, Luo M, Desnick RJ, Kornreich R. Experience with carrier screening and prenatal diagnosis for 16 Ashkenazi Jewish genetic diseases. *Hum Mutat* 2010;31:1240–1250.
11. National Human Genome Research Institute. DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program. http://www.genome.gov/sequencingcosts. Accessed 2 September 2011.
12. Srinivasan BS, Evans EA, Flannick J, et al. A universal carrier test for the long tail of Mendelian disease. *Reprod Biomed Online* 2010;21:537–551.
13. Wei S, Quigg MH, Monaghan KG. Is cystic fibrosis carrier screening cost effective? *Community Genet* 2007;10:103–109.
14. Musci TJ, Caughey AB. Cost-effectiveness analysis of prenatal population-based fragile X carrier screening. *Am J Obstet Gynecol* 2005;192:1905–12; discussion 1912.
15. McConkey E. *Human Genetics: The Molecular Revolution*, 1st ed. Jones and Bartlett: Sudbury, MA, 1993.
16. Parra EJ, Marcini A, Akey J, et al. Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 1998;63:1839–1851.
17. Lee SS, Mountain J, Koenig B, et al. The ethics of characterizing difference: guiding principles on using racial categories in human genetics. *Genome Biol* 2008;9:404.
18. U.S. Census Bureau. *Statistical Abstract of the United States: 2011*, 130th ed. Washington, DC, 2010.
19. Condit C, Templeton A, Bates BR, Bevan JL, Harris TM. Attitudinal barriers to delivery of race-targeted pharmacogenomics among informed lay persons. *Genet Med* 2003;5:385–392.
20. De Marco M, Cykert S, Coad N, et al. Views on personalized medicine: do the attitudes of African American and white prescription drug consumers differ? *Public Health Genomics* 2010;13:276–283.
21. Rohlfs EM, Zhou Z, Heim RA, et al. Cystic fibrosis carrier testing in an ethnically diverse US population. *Clin Chem* 2011;57:841–848.
22. Mei-Zahav M, Durie P, Zielenski J, et al. The prevalence and clinical characteristics of cystic fibrosis in South Asian Canadian immigrants. *Arch Dis Child* 2005;90:675–679.
23. Taroni F, Verderio E, Dworzak F, Willems PJ, Cavadini P, DiDonato S. Identification of a common mutation in the carnitine palmitoyltransferase II gene in familial recurrent myoglobinuria patients. *Nat Genet* 1993;4:314–320.
24. Nowaczyk MJ, Waye JS, Douketis JD. DHCR7 mutation carrier rates and prevalence of the RSH/Smith-Lemli-Opitz syndrome: where are the patients? *Am J Med Genet A* 2006;140:2057–2062.
25. Park HJ, Shaukat S, Liu XZ, et al. Origins and frequencies of SLC26A4 (PDS) mutations in east and south Asians: global implications for the epidemiology of deafness. *J Med Genet* 2003;40:242–248.
26. Sawai H, Nishida N, Mbarek H, et al. No association for Chinese HBV-related hepatocellular carcinoma susceptibility SNP in other East Asian populations. *BMC Med Genet* 2012;13:47.
27. Kelley RI, Hennekam RC. The Smith-Lemli-Opitz syndrome. *J Med Genet* 2000;37:321–335.
28. Sugarman EA, Nagan N, Zhu H, et al. Pan-ethnic carrier screening and prenatal diagnosis for spinal muscular atrophy: clinical laboratory analysis of >72,400 specimens. *Eur J Hum Genet* 2012;20:27–32.
29. Lacassie Y, Storment JM, Lazarin GA. Serendipitous diagnosis of mild recessive multiple epiphyseal dysplasia through parental-targeted screening test. *Am J Med Genet A* 2011;155A:3136–3138.
30. Unger S, Bonafé L, Superti-Furga A. Multiple epiphyseal dysplasia: clinical and radiographic features, differential diagnosis and molecular basis. *Best Pract Res Clin Rheumatol* 2008;22:19–32.