HHS Public Access

Author manuscript

Clin Genet. Author manuscript; available in PMC 2016 June 01.

Published in final edited form as:

Clin Genet. 2015 June ; 87(6): 570–575. doi:10.1111/cge.12425.

Determination of the Allelic Frequency in Smith-Lemli-Opitz Syndrome by Analysis of Massively Parallel Sequencing Data Sets

Joanna L. Cross[*,1], **James Iben**[*,1], **Claire Simpson**[2], **Audrey Thurm**[3], **Susan Swedo**[3], **Elaine Tierney**[4], **Joan Bailey-Wilson**[2], **Leslie G. Biesecker**[5], **Forbes D. Porter**[1], and **Christopher A. Wassif**[1]

[1]Program in Developmental Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA

[2]Statistical Genetics Section, National Human Genome Research Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA

[3]Pediatrics & Developmental Neuroscience Branch, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA

[4]Department of Psychiatry, Kennedy Krieger Institute, Baltimore, MD, USA

[5]Human Development Section, National Human Genome Research Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA

Abstract

Data from massively parallel sequencing or “Next Generation Sequencing” of the human exome has reached a critical mass in both public and private databases, in that these collections now allow researchers to critically evaluate population genetics in a manner that was not feasible a decade ago. The ability to determine pathogenic allele frequencies by evaluation of the full coding sequences and not merely a single SNP or series of SNPs will lead to more accurate estimations of incidence. For demonstrative purposes we analyzed the causative gene for the disorder Smith-Lemli-Opitz Syndrome (SLOS), the 7-dehydrocholesterol reductase (*DHCR7*) gene and determined both the carrier frequency for *DHCR7* mutations, and predicted an expected incidence of the disorder. Estimations of the incidence of SLOS have ranged widely from 1:10,000 to 1:70,000 while the carrier frequency has been reported as high as 1 in 30. Using four exome data sets with a total of 17,836 chromosomes, we ascertained a carrier frequency of pathogenic *DHRC7* mutations of 1.01%, and predict a SLOS disease incidence of 1/39,215 conceptions. This approach highlights yet another valuable aspect of the exome sequencing databases, to inform clinical and health policy decisions related to genetic counseling, prenatal testing and newborn screening.

Corresponding Author: Christopher A. Wassif M.S., Candidate for Doctorate in Philosophy, University of Oxford, Research Associate, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, 9000 Rockville Pike Building CRC Room 1E-3288, Bethesda, MD 20892-1830, (301) 496-6664.

[*]Authors contributed equally to this work

Conflict of interest statement: LGB is an uncompensated advisor to the Illumina Corporation and receives royalties from the Genentech Corporation; all other authors report no conflict of interest.

Introduction

Mutations in the 7-dehydrocholesterol reductase gene (*DHCR7*) cause the autosomal recessive disorder, Smith-Lemli-Opitz Syndrome (SLOS), a multiple malformation/cognitive impairment syndrome. Perturbations of DHCR7 function result in the accumulation of 7-dehydrocholesterol (7DHC) and reduction in cholesterol levels. This alteration of sterol composition results in a plethora of congenital malformations accompanied by various degrees of cognitive and behavioral impairments. The incidence of SLOS has been reported between 1:10,000 to 1:70,000 depending on the population being studied (1), while the carrier frequency has been reported as high as 1 in 30 (2). Efforts to determine the allelic frequency of mutations has been limited to a few common mutations but are inadequate for predicting overall carrier frequencies or predicting the rate of SLOS pregnancies given the large number of pathogenic mutations found in *DHCR7*. To date 164 mutations have been reported (3-8). Although SLOS has been thought of predominantly as a disorder of Northern European decent (9), there have been reported cases in nearly all ethnicities. Even within Europe there are frequency differences and gradients for individual mutations. For example, the most common mutant allele, c.964-1G>C, likely originated in the British Isles but has a decreasing prevalence moving eastward across Northern Europe (9). c.964-1G>C accounts for approximately one third of all mutant alleles reported, with a reported carrier frequency of approximately 1% in North American populations (1). Reported estimations of the clinical incidence of SLOS may be erroneous due to lack of ascertainment at both ends of the clinical spectrum. Many severely affected patients are likely lost *in utero* (10), while mildly affected patients with minimal malformations may fail to prompt biochemical testing and are therefore potentially missed. Due to this potentially biased ascertainment, it is difficult to discern both the full prevalence of the disorder or the allelic frequency. In this report we sought to determine the full allelic frequency of SLOS utilizing exome sequencing.

Materials and Methods

Generation of the variant call data presented in this report used the following analysis. Variant calls from the V3 release of the 1000 Genomes Project (11) were pulled for regions overlapping a defined gene, in this case *DHCR7*, by Perl script. Each individual call of the 1092 samples was evaluated for each allele (2184 calls) at every base of the coding exons or occurring within 5 base pairs of an exon. This was done for all the data sets, with the exception of the ClinSeq® data that was evaluated within 2 base pairs of an exon. Variant calls were tabulated by defined ethnicity of the sample, allowing for mutation frequencies to be calculated for both the overall set and for individual ethnic groups at each mutant position. Mutations were annotated using SNPnexus (12) using Refseq (13) annotations and Polyphen-2 severity of effect predictions (14), variations detected within 5 bases of intron exon boundaries were analyzed by MaxEntScan (15) Annotated synonymous and UTR mutations were dropped prior to calculating an overall mutation rate for the gene. Separately, data was compiled from ClinSeq® (16), NHLBI GO Exome Sequencing Project (ESP) (17), and a NIH inter-institute collaboration on autism for the gene of interest using a similar scripted pipeline.

Classification of the variant as pathogenic or nonpathogenic was accomplished by comparing the variants found in the data sets against all published mutations in literature as well as performing a gene search of the Human Gene Mutation Database (HGMD®) (8). For variants that had not been previously associated with SLOS, we classified those determined by Polyphen-2 as “Probably Damaging”, or “Possibly Damaging” as pathogenic. The splice variant, c.99(-4)G>A, was tested by isolation of mRNA from a heterozygous cell line. Reverse transcription PCR using SuperScript III® (Life Technologies) and Sanger sequencing were performed.

ESP contributed a maximum allele number of 12,988, 1000 genome contributed 2,184 alleles, ClinSeq® contributed 1,902 alleles and the NIH inter-institute collaboration on Autism project contributed 762 alleles. This totaled to a maximum of 17,836 alleles that we analyzed. None of these datasets included any patients ascertained for the presence of SLOS, so we considered them to be unbiased with respect to variation in the *DHCR7* gene.

Results

Comparison of the human reference sequence to exome sequencing data available from the four databases revealed 515 nonsynonymous sequence variants, in *DHCR7* from 17,836 chromosomes, indicating an overall incidence of variation of 2.89% (515 variant alleles/ 17,836 total alleles). Further analysis of the 515 variants found 74 distinct changes; 71 coding single nucleotide base variants, 2 splice mutations, and 1 insertion. The majority of these changes were unique in the data set (43 of the 74 variants), whereas 9 variants were noted 10 or more times and accounted for 408 of the total 515 variants detected (79%) (Table 1). In order to determine the rate of pathogenic alleles we conducted a literature search and queried HGMD® and found 25 of 74 distinct variants published in individuals diagnosed with SLOS. This included 23 coding single nucleotide base variants and 2 splice mutations. We further performed *in silico* analysis on all the coding single nucleotide base variants using Polyphen-2, which provided a predictive assignment of “benign”, “possibly damaging”, or “probably damaging”. Of the 23 coding single nucleotide base variants identified previously in SLOS, Polyphen-2 classified 18 as probably damaging, 4 as possibly damaging, and 1 as benign. Of the remaining 48 coding single nucleotide base variants Polyphen-2 predicted 17 as probably damaging, 7 as possibly damaging, and 24 as benign. For this study we classified variants that have not been previously associated with SLOS as pathogenic if they had Polyphen-2 prediction of “Probably Damaging”, or “Possibly Damaging”.

The 2 splice variants identified in this study were c.964-1G>C and c.99-4G>A. The c. 964-1G>C mutation is the most common mutation found in SLOS patients and has previously been shown to have a carrier frequency of approximately 1% (2). Restricting analysis of 1000 Genome Project data to 988 alleles from Northern European populations (CEU, CLM, FIN, GBR, IBS, TSI, PUR) we found an incidence of 0.81% consistent with published results (Table 2). The second splice variant, c.99-4G>A, presented with a combined frequency of 0.17%. Although this variant has recently been published as a causative mutation in SLOS (3), the relatively high carrier frequency combined with only being reported once led us to hypothesize that it is non-pathogenic. *In silico* analysis with

MaxEntScan predicted this change to be neutral. To verify the nonpathological nature of this variant, mRNA from a cell line heterozygous for this polymorphism, was reverse transcribed and sequenced. No evidence of alternative splicing was detected and sequencing identified two distinct normal alleles (data not shown). We conclude that this variant is benign. In addition, we identified one 5-nucleotide insertion, c.849_856CTTCT, that predicts a frameshift, p.F284S*10, and early termination that we consider pathogenic.

Combining published pathogenic alleles, with the alleles identified by Polyphen-2 as possibly damaging, probably damaging, and the one likely pathogenic insertion resulted in a total of 191 pathogenic alleles. Thus, these data predict a carrier frequency for pathogenic *DHRC7* mutations of 1.07%. Polyphen-2 provided a false discovery rate (FDR) for each of the 71 coding single nucleotide base variants. The average FDR was 12% for all 3 categories, where as the average FDR of the combined probably damaging, or possibly damaging was 5%. Taking into account the FDR of the probably damaging, or possibly damaging variants we are left with 181 pathogenic alleles, and a carrier frequency for pathogenic *DHRC7* mutations of 1.01% (181 variant alleles/17,836 total alleles). Based on these carrier frequencies we are able to predict a SLOS disease incidence of1/34,937 to 1/39,215 conceptions, respectively. This estimate is in good concordance, especially when one takes into account *in utero* demise, with the estimated incidence of 1/40,000 reported by Lowry and Yong, and Nowaczyk et al. (18, 19) and an incidence of 1/50,000 based on the biochemical diagnosis of approximately 40 cases of SLOS per year in the US between 1995 and 1998 (20).

Discussion

While this data set is extremely powerful, it is not without limitations. We detected the most common mutation reported in SLOS, c.964-1G>C, as well as many of the other common mutations. However, of the 164 known mutations we only detected 25, or 15.2% of the known pathogenic alleles in SLOS. A number of factors could influence both the accuracy of our allelic prediction and the number of variants detected. First and foremost are the quality of the sequencing and the depth of coverage. Exome capture is still not perfect and routinely fails to capture the whole exome, as noted in our supplemental tables we did not have sequence at every variant detected for all 17,836 chromosomes. Numerous alignment and calling algorithms exist that can impact the number of variants called per sample. The depth of coverage is important to consider as the 1000 genomes project is sequencing to an average depth of 4×. As such the project predicts they will only detect a minor variant allele that occurs at a frequency of 1%. Many exome sequencing projects, including the 3 other data sets presented here, are sequencing to a greater depth of coverage, which should permit increased detection of rare variants. Second, is our interpretation of these data. We utilized the *in silico* predictions of Polyphen to predict deleterious mutations, while in good agreement with the published pathogenic alleles there was one variant, c.400G>T, p. V134L that Polyphen predicted as benign, but has been reported as deleterious in an SLOS patient (3). However, no functional analysis has been performed on this mutation, and it is possible that it is a benign variant. Conversely it is possible that the overall incidence maybe slightly higher due to our exclusion of coding single nucleotide base variants deemed by polyphen-2 to be “Benign” if subsequent studies find them to be causative.

Greater ethnicity data in association with the exome data sets will enhance the ability to accurately assess frequencies of rare variants within any population. This was most evident in our study of the common c.964-1G>C mutation in these data sets. When we were able to directly identify individuals of northern European decent we found a rate similar to published literature. The lack of ethnicity data by no means diminishes the power of these data sets; the inclusion of ethnicity would simply increase the versatility of the data to perform greater population genetic analysis. As the cost of exome sequencing continues to decrease these data sets are likely to continue growing, becoming even more powerful resources.

This study predicts the carrier frequency of *DHCR7* mutations based on 17,836 *DHCR7* alleles. This approach has general utility for rare diseases in that an accurate estimate of carrier frequency can be used for prediction of disease incidence, allow determination of more accurate genetic risk assessments, and guide public health decisions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

JLC, JI, CS, AT, SS, BBW, LGB, FDP, and CAW are all funded by the NIH intramural research program, DHHS. ET is funded by Autism Speaks.

References

1. Porter FD, Herman GE. Malformation syndromes caused by disorders of cholesterol synthesis. J Lipid Res. 2011; 52(1):6–34. [PubMed: 20929975]
2. Battaile KP, Battaile BC, Merkens LS, Maslen CL, Steiner RD. Carrier frequency of the common mutation IVS8-1G>C in DHCR7 and estimate of the expected incidence of Smith-Lemli-Opitz syndrome. Molecular genetics and metabolism. 2001; 72(1):67–71. [PubMed: 11161831]
3. Waterham HR, Hennekam RC. Mutational spectrum of Smith-Lemli-Opitz syndrome. Am J Med Genet C Semin Med Genet. 2012; 160C(4):263–84. [PubMed: 23042628]
4. Balogh I, Koczok K, Szabo GP, et al. Mutational spectrum of smith-lemli-opitz syndrome patients in hungary. Mol Syndromol. 2012; 3(5):215–22. [PubMed: 23293579]
5. Ellingson MS, Wick MJ, White WM, et al. Pregnancy in an individual with mild Smith-Lemli-Opitz syndrome. Clin Genet. 2013
6. Sparks SE, Wassif CA, Goodwin H, et al. Decreased cerebral spinal fluid neurotransmitter levels in Smith-Lemli-Opitz syndrome. J Inherit Metab Dis. 2014
7. Olah AV, Szabo GP, Varga J, et al. Relation between biomarkers and clinical severity in patients with Smith-Lemli-Opitz syndrome. Eur J Pediatr. 2013; 172(5):623–30. [PubMed: 23319240]
8. Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: 2008 update. Genome Med. 2009; 1(1):13. [PubMed: 19348700]
9. Witsch-Baumgartner M, Ciara E, Loffler J, et al. Frequency gradients of DHCR7 mutations in patients with Smith-Lemli-Opitz syndrome in Europe: evidence for different origins of common mutations. Eur J Hum Genet. 2001; 9(1):45–50. [PubMed: 11175299]
10. Putnam AR, Szakacs JG, Opitz JM, Byrne JL. Prenatal death in Smith-Lemli-Opitz/RSH syndrome. Am J Med Genet A. 2005; 138(1):61–5. [PubMed: 15940686]
11. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491(7422):56–65. [PubMed: 23128226]

12. Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. Bioinformatics. 2009; 25(5):655–61. [PubMed: 19098027]
13. Haas D, Garbade SF, Vohwinkel C, et al. Effects of cholesterol and simvastatin treatment in patients with Smith-Lemli-Opitz syndrome (SLOS). J Inherit Metab Dis. 2007; 30(3):375–87. [PubMed: 17497248]
14. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7(4):248–9. [PubMed: 20354512]
15. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004; 11(2-3):377–94. [PubMed: 15285897]
16. Biesecker LG, Mullikin JC, Facio FM, et al. The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine. Genome Res. 2009; 19(9):1665–74. [PubMed: 19602640]
17. [cited 2014 January] Exome Variant Server NHLBI GO Exome Sequencing Project (ESP). Available from: http://evs.gs.washington.edu/EVS/
18. Lowry RB, Yong SL. Borderline normal intelligence in the Smith-Lemli-Opitz (RSH) syndrome. Am J Med Genet. 1980; 5(2):137–43. [PubMed: 7395908]
19. Nowaczyk MJ, Waye JS, Douketis JD. DHCR7 mutation carrier rates and prevalence of the RSH/Smith-Lemli-Opitz syndrome: where are the patients? American journal of medical genetics Part A. 2006; 140(19):2057–62. [PubMed: 16906538]
20. Kelley RI, Hennekam RC. The Smith-Lemli-Opitz syndrome. J Med Genet. 2000; 37(5):321–35. [PubMed: 10807690]

Table 1

Composite Summary Table.

This table summarizes the 74 distinct variants analyzed in this report. Each variant has a corresponding cDNA number; protein change and RS number when available and the majority have been assigned either a Polyphen or MaxEntScan score, as well variants that have been previously published are noted. The number of alleles analyzed for each variant and the total number of times the variant was seen out of these alleles is noted in conjunction with the frequency of each variant in each of the four data sets and the carrier rate for each variant. The variants considered pathogenic have no shading while those considered non-pathogenic are shaded grey.

cDNA	Protein	rs#	Polyphen/Published/MaxEntScan	Total Alleles	Total Variants	1000 Genome	ESP	Clinseq	Autism	Rate
c.3G>A	p.M1I	rs121909767	POSSIBLY DAMAGING/Published	17836	1	0	1	0	0	0.01%
c.14C>T	p.S5L	rs1127869	BENIGN	17836	2	2	0	0	0	0.01%
c.25A>T	p.I9F	rs115595829	BENIGN	17502	171	13	153	3	2	0.98%
c.28C>G	p.P10A	rs139166382	BENIGN	17836	3	0	3	0	0	0.02%
c.70G>T	p.A24S	rs146867923	BENIGN	17736	43	4	37	0	2	0.24%
c.77A>C	p.Q26P	rs199815542	BENIGN	17836	1	0	0	1	0	0.01%
c.89G>C	p.G30A	rs200334114	PROBABLY DAMAGING/Published	17836	1	1	0	0	0	0.01%
c.91C>T	p.R31C	rs367585401	PROBABLY DAMAGING/Published	17836	1	0	1	0	0	0.01%
c.92G>A	p.R31H	rs370307688	PROBABLY DAMAGING	17836	2	0	2	0	0	0.01%
c.99-4G>A	intronic	rs140748737	Neutral/Published	15834	27	19	0	0	8	0.17%
c.127G>A	p.V43I	rs200984695	BENIGN	17836	1	0	1	0	0	0.01%
c.131T>C	p.I44T	rs142897396	BENIGN	17836	5	0	5	0	0	0.03%
c.165C>G	p.Y55X	rs374941029	PROBABLY DAMAGING	17836	1	0	1	0	0	0.01%
c.199G>A	p.A67T	rs143999854	BENIGN	17746	12	1	10	1	0	0.07%
c.208G>A	p.G70S	rs144512551	BENIGN	17772	22	2	19	1	0	0.12%
c.221A>C	p.D74A	rs374874199	BENIGN	17836	1	0	1	0	0	0.01%
c.223A>T	p.I75F	rs370748173	BENIGN	17836	2	0	2	0	0	0.01%
c.226G>A	p.V76I		BENIGN	17502	1	0	0	0	1	0.01%
c.232G>A	p.G78R	rs373352413	PROBABLY DAMAGING	17836	1	0	1	0	0	0.01%
c.265A>T	p.T89S	rs375997113	POSSIBLY DAMAGING	17836	1	0	1	0	0	0.01%
c.278C>T	p.T93M	rs80338853	PROBABLY DAMAGING/Published	17836	1	0	1	0	0	0.01%
c.289G>A	p.A97T	rs150563256	BENIGN	17836	5	4	1	0	0	0.03%

cDNA	Protein	rs#	Polyphen/Published/MaxEntScan	Total Alleles	Total Variants	1000 Genome	ESP	Clinseq	Autism	Rate
c.292C>T	p.Q98X	rs104886039	PROBABLY DAMAGING/Published	17822	2	0	1	1	0	0.01%
c.321G>C	p.Q107H	rs104886040	PROBABLY DAMAGING/Published	17836	1	0	1	0	0	0.01%
c.376G>A	p.V126I	rs143587828	PROBABLY DAMAGING	17836	2	0	2	0	0	0.01%
c.385A>G	p.I129V	rs138215017	BENIGN	17836	1	0	1	0	0	0.01%
c.400G>T	p.V134L	rs201466849	BENIGN/Published	17308	3	1	1	0	1	0.02%
c.418G>A	p.V140M	rs373908315	BENIGN	17836	1	0	1	0	0	0.01%
c.452G>A	p.W151X	rs11555217	PROBABLY DAMAGING/Published	17422	15	0	11	4	0	0.09%
c.461C>G	p.T154R	rs143312232	PROBABLY DAMAGING/Published	17494	10	0	5	0	5	0.06%
c.506C>T	p.S169L	rs80338855	POSSIBLY DAMAGING/Published	17836	1	0	1	0	0	0.01%
c.509C>G	p.P170R	rs150459687	PROBABLY DAMAGING	17836	1	0	1	0	0	0.01%
c.523G>A	p.D175N	rs368269558	POSSIBLY DAMAGING	17836	1	0	1	0	0	0.01%
c.583G>A	p.A195T	rs151170252	BENIGN	17836	3	0	3	0	0	0.02%
c.603C>A	p.F201L	rs377727130	BENIGN	17836	1	0	1	0	0	0.01%
c.658A>C	p.M220L	rs200659923	POSSIBLY DAMAGING	17836	1	1	0	0	0	0.01%
c.670G>A	p.E224K	rs373121544	PROBABLY DAMAGING/Published	17836	1	0	1	0	0	0.01%
c.683G>A	p.R228Q	rs201556114	POSSIBLY DAMAGING	17836	1	1	0	0	0	0.01%
c.719A>G	p.N240S	rs148609143	PROBABLY DAMAGING	17836	5	0	5	0	0	0.03%
c.724C>T	p.R242C	rs80338856	PROBABLY DAMAGING/Published	17836	2	0	2	0	0	0.01%
c.822C>A	p.N274K	rs139787408	POSSIBLY DAMAGING/Published	17836	1	0	1	0	0	0.01%
c.849_856insCTTCT				17824	1	0	0	1	0	0.01%
c.852C>A	p.F284L	rs184297154	PROBABLY DAMAGING/Published	17836	1	1	0	0	0	0.01%
c.883A>G	p.I295V	rs201574502	PROBABLY DAMAGING	17836	1	1	0	0	0	0.01%
c.907G>A	p.G303R	rs142808899	PROBABLY DAMAGING/Published	17836	4	0	4	0	0	0.02%
c.928G>A	p.D310N	rs370955781	PROBABLY DAMAGING	17836	1	0	1	0	0	0.01%
c.964-1G>C	intronic	rs138659167	Negative/Published	16736	94	10	70	9	5	0.56%
c.988G>A	p.V330M	rs139724817	POSSIBLY DAMAGING/Published	17812	8	2	5	0	1	0.04%
c.1004C>G	p.P335R	rs77762671	BENIGN	17836	2	2	0	0	0	0.01%
c.1012G>A	p.V338M	rs72954276	BENIGN	17568	14	4	9	1	0	0.08%
c.1018G>A	p.V340I	rs148081697	BENIGN	17830	1	0	1	0	0	0.01%

Clin Genet. Author manuscript; available in PMC 2016 June 01.

cDNA	Protein	rs#	Polyphen/Published/MaxEntScan	Total Alleles	Total Variants	1000 Genome	ESP	Clinseq	Autism	Rate
c.1025T>C	p.L342P	rs199957106	POSSIBLY DAMAGING	17836	1	1	0	0	0	0.01%
c.1084C>T	p.R362C	rs371302153	PROBABLY DAMAGING/Published	17832	1	0	1	0	0	0.01%
c.1085G>A	p.R362H	rs142213147	PROBABLY DAMAGING	17836	2	0	2	0	0	0.01%
c.1088G>A	p.R363H	rs200539324	POSSIBLY DAMAGING	17836	1	1	0	0	0	0.01%
c.1138T>C	p.C380R	rs373306653	PROBABLY DAMAGING/Published	17836	1	0	1	0	0	0.01%
c.1156G>A	p.D386N	rs369837196	PROBABLY DAMAGING	17836	1	0	1	0	0	0.01%
c.1204G>A	p.V402M	rs200099137	BENIGN	17822	1	0	1	0	0	0.01%
c.1210C>T	p.R404C	rs61757582	PROBABLY DAMAGING/Published	17574	1	0	0	1	0	0.01%
c.1228G>C	p.G410R		PROBABLY DAMAGING/Published	17502	1	0	0	0	1	0.01%
c.1231G>A	p.D411N	rs372055524	PROBABLY DAMAGING	17828	2	0	2	0	0	0.01%
c.1270G>T	p.G424C	rs368150818	PROBABLY DAMAGING	17810	1	0	1	0	0	0.01%
c.1270G>A	p.G424S		BENIGN	17442	1	0	0	1	0	0.01%
c.1290C>G	p.Y430X	rs140791666	PROBABLY DAMAGING	17828	1	0	1	0	0	0.01%
c.1300A>G	p.I434V	rs375187933	BENIGN	17832	1	0	1	0	0	0.01%
c.1336C>T	p.R446W	rs145043679	PROBABLY DAMAGING/Published	17834	2	0	2	0	0	0.01%
c.1342G>A	p.E448K	rs80338864	PROBABLY DAMAGING/Published	17572	1	0	0	1	0	0.01%
c.1354G>A	p.A452T	rs140400648	BENIGN	17836	2	2	0	0	0	0.01%
c.1366G>A	p.G456S	rs201847193	PROBABLY DAMAGING	17836	1	1	0	0	0	0.01%
c.1369C>T	p.R457W	rs371873032	PROBABLY DAMAGING	17832	2	0	2	0	0	0.01%
c.1381C>T	p.R461C	rs199506852	POSSIBLY DAMAGING	17836	2	2	0	0	0	0.01%
c.1384T>C	p.Y462H	rs201270451	PROBABLY DAMAGING/Published	17630	1	0	0	1	0	0.01%
c.1405C>T	p.R469C	rs148660993	PROBABLY DAMAGING	17818	2	0	2	0	0	0.01%
c.1406G>A	p.R469H	rs201150384	PROBABLY DAMAGING	17836	1	1	0	0	0	0.01%

Table 2

Summary of the common c.964-1G>C variant.

This table summarizes the data found on the common c.964-1G>C variant. This variant was found 10 times in the 1000 Genome database, 70 times in ESP, 9 times in ClinSeq and 5 times in the Autism database. These equate to carrier rates of 0.46%, 0.54%, 0.97% and 0.66% respectively.

	1000 Genome	NE populations*	ESP	ClinSeq	Autism
c.964-1G>C variant	10	8	70	9	5
Total alleles analyzed	2184	988	12864	926	762
Rate	0.46%	0.81%	0.54%	0.97%	0.66%

* The 1000 Genome project defined the populations analyzed and therefore if only Northern European populations were taken into consideration (CEU, CLM, FIN, GBR, IBS, TSI, PUR, Supplemental Table 1), a carrier rate of 0.81% was found. This variation of the ethnic diversity emphasizes the importance of considering the population studied.