The global prevalence of *HFE* and non-*HFE* hemochromatosis estimated from analysis of next-generation sequencing data

Daniel F. Wallace, PhD^{1,2} and V. Nathan Subramaniam, PhD^{1,2}

Purpose: The prevalence of *HFE*-related hereditary hemochromatosis (HH) among European populations has been well studied. There are no prevalence data for atypical forms of HH caused by mutations in *HFE2*, *HAMP*, *TFR2*, or *SLC40A1*. The purpose of this study was to estimate the population prevalence of these non-*HFE* forms of HH.

Methods: A list of HH pathogenic variants in publically available next-generation sequence (NGS) databases was compiled and allele frequencies were determined.

Results: Of 161 variants previously associated with HH, 43 were represented among the NGS data sets; an additional 40 unreported functional variants also were identified. The predicted prevalence of *HFE* HH and the p.Cys282Tyr mutation closely matched previous estimates from similar populations. Of the non-*HFE* forms of iron

overload, *TFR2*-, *HFE2*-, and *HAMP*-related forms are predicted to be rare, with pathogenic allele frequencies in the range of 0.00007 to 0.0005. Significantly, *SLC40A1* variants that have been previously associated with autosomal-dominant ferroportin disease were identified in several populations (pathogenic allele frequency 0.0004), being most prevalent among Africans.

Conclusion: We have, for the first time, estimated the population prevalence of non-*HFE* HH. This methodology could be applied to estimate the population prevalence of a wide variety of genetic disorders.

Genet Med advance online publication 3 December 2015

Key Words: hemochromatosis; mutations; non-*HFE* hemochromatosis; next-generation sequencing; prevalence

INTRODUCTION

Hereditary hemochromatosis (HH) is caused by mutations in genes involved in the regulation of systemic iron homeostasis.¹ Dysregulation of iron homeostasis in HH leads to the absorption and storage of more iron than is required, leading to tissue damage and disease. Five genes have been associated with various subtypes of the disorder.² The most common gene implicated in HH is *HFE* (OMIM 235200); homozygosity for a single missense substitution (p.Cys282Tyr) is responsible for between 60 and 100% of cases among European populations.^{3,4} The p.Cys282Tyr mutation has an allele frequency of around 5% in the European population, with a decreasing gradient from north to south.⁵ It is thought to have originated in a Celtic population several thousand years ago, although a Viking origin for the mutation also has been proposed.⁵ The p.Cys282Tyr founder mutation has since spread throughout Europe, perhaps as a result of a selective advantage to those carrying the mutation, who may be protected from the development of iron-deficiency anemia.⁵ The p.His63Asp variant in *HFE* is more common than p.Cys282Tyr, having an allele frequency of around 14% among Europeans. It also has a much wider geographical distribution, being present at varying allele frequencies in most of the world's populations.⁶ p.His63Asp is generally not considered to be a pathogenic mutation because the majority of people who are homozygous show no evidence

of iron overload. Individuals who are compound heterozygous for p.Cys282Tyr and p.His63Asp can develop iron overload, and this variant is enriched among individuals with a diagnosis of HH.⁷ However, the penetrance of this genotype is low, and other cofactors that contribute to iron overload are usually present.⁸

Much of the knowledge we have gained regarding the molecular mechanisms regulating iron homeostasis has stemmed from the identification and characterization of genes implicated in non-*HFE* forms of iron overload. Central to the regulation of systemic iron homeostasis is the hepcidin-ferroportin axis. Hepcidin is a negative regulator of iron absorption and recycling, and functions to limit iron release from cells by binding to the iron exporter ferroportin, causing its internalization and degradation.⁹

All forms of HH reported so far result from defective function or regulation of hepcidin or ferroportin. Iron overload in all autosomal-recessive forms of HH results from either inappropriately low or no production of hepcidin. For example, *HFE* is an upstream regulator of hepcidin, and patients with *HFE* HH have low concentrations of hepcidin relative to iron stores.¹⁰ One form of juvenile HH (type 2A; OMIM 602390) is caused by mutations in an upstream regulator of hepcidin, hemojuvelin (encoded by the *HFE2* gene).¹¹ A second form of juvenile HH (type 2B; OMIM 613313) is caused by mutations

¹Membrane Transport Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, Australia; ²Faculty of Medicine and Biomedical Sciences, The University of Queensland, Brisbane, Queensland, Australia. Correspondence: Daniel F. Wallace (Daniel.Wallace@qimrberghofer.edu.au)

Submitted 10 June 2015; accepted 27 August 2015; advance online publication 3 December 2015. doi:10.1038/gim.2015.140

in the hepcidin gene (*HAMP*) itself.¹² Patients with type 3 HH (OMIM 604250) have mutations in the gene encoding another regulator of hepcidin, transferrin receptor 2 (*TFR2*).¹³ An autosomal-dominant form of HH (type 4; OMIM 606069) is caused by heterozygous mutations in the ferroportin gene (*SLC40A1*) and can result in one of two phenotypes caused by either loss of iron transport ability or hepcidin insensitivity.¹⁴

The population prevalence of *HFE* HH and the *HFE* variants p.Cys282Tyr and p.His63Asp have been studied extensively. As far as we are aware, the population prevalence of non-*HFE* forms of HH have not been determined. This is likely because of the low frequency of these conditions relative to *HFE* HH among European populations, the low frequency of HH among non-European populations, and the paucity of non-*HFE* mutations identified in the screening of control groups. With the advent of next-generation sequencing (NGS) methods that can sequence whole genomes or exomes relatively quickly and cost-effectively, it is now possible to search relatively large population groups for variants in virtually any gene in the human genome. We hypothesized that pathogenic alleles known to cause non-*HFE* forms of iron overload would be detected in large genomic sequence databases and that data derived from these databases could be used to estimate the population prevalence of these disorders. By analyzing data from the 1000 Genomes Project (1000G),¹⁵ Exome Sequencing Project (ESP; Exome Variant Server, NHLBI GO ESP, Seattle, WA; <http://evs.gs.washington.edu/EVS/>), and Exome Aggregation Consortium (ExAC; Cambridge, MA; <http://exac.broadinstitute.org>), we have attempted to estimate the population prevalence of all forms of HH. Our analysis has produced a first estimate for the population prevalence of the various forms of non-*HFE* HH; this study forms a valuable resource for informing the medical and research community on the incidence of these atypical iron overload disorders and variants that are potential genetic modifiers of disease.

MATERIALS AND METHODS

Identification and classification of HH-causing variants

An extensive literature review, which included a comprehensive search of the National Center for Biotechnology Information PubMed database (<http://www.ncbi.nlm.nih.gov/pubmed>) for articles published up until July 2015, was performed to identify all variants in the *HFE* (NM_000410.3), *HFE2* (NM_213653.3), *HAMP* (NM_021175.2), *TFR2* (NM_003227.3), and *SLC40A1* (NM_014585.5) genes, which are reported to be associated with HH. Variants were classified into those that seemed to be pathogenic and those that were possibly pathogenic. Variants were classified as pathogenic if they were reportedly associated with iron overload, defined by elevated serum ferritin and/or transferrin saturation, in patients or families with HH and in the homozygous or compound heterozygous state for *HFE*, *HFE2*, *HAMP*, or *TFR2* or in the heterozygous state for *SLC40A1*. Variants were classified as possibly pathogenic if they were reported only in the heterozygous state for *HFE*, *HFE2*, *HAMP*, or *TFR2* genes, if there were conflicting reports of whether they were associated with iron overload, or if they were reported only

as modifiers of the *HFE* HH iron overload phenotype. For example, the *HFE* p.Cys282Tyr mutation was classified as pathogenic, whereas other common variants, such as *HFE* p.His63Asp, *HFE* p.Ser65Cys, and *SLC40A1* Gln248His, for which penetrance is known to be low, were classified as possibly pathogenic. Many of the variants have been reported only in single cases or families, and pathogenicity in these cases was assigned based on the phenotypic data available and the opinions of the study authors.

Identification of major functional variants of HH genes

The 1000G, ESP, and ExAC databases were searched for the presence of variants in the *HFE*, *HFE2*, *HAMP*, *TFR2*, and *SLC40A1* genes that would be expected to have major functional or structural effects on the encoded protein but had not previously been reported in patients with HH. These included frameshift, premature stop codon, initiator codon, splice donor, and splice acceptor variants.

Determination of HH gene variant allele frequencies

The coding sequence nucleotide changes for each variant were identified and converted to human genome coordinates (hg19) using the Mutalyzer program (<https://mutalyzer.nl/>). The genome coordinates and nucleotide changes were converted into ANNOVAR format and run through the wANNOVAR program¹⁶ to generate annotations for each variant, including allele frequencies in the 1000G, ESP, and ExAC data sets. Allele frequencies for previously reported pathogenic variants and major functional variants were used to estimate the population prevalence of homozygous and heterozygous HH genotypes using the Hardy-Weinberg equation. When more than five alleles were detected in the ExAC data set, a binomial proportion 95% confidence interval (CI) was calculated using the Wilson score method.

RESULTS

Identification of HH-causing variants

An extensive literature review was conducted to identify pathogenic and possibly pathogenic variants in the five genes associated with HH. This process yielded a total of 215 variants across the 5 HH genes: 38 in *HFE*, 55 in *HFE2*, 13 in *HAMP*, 49 in *TFR2*, and 60 in *SLC40A1*; these were reported in 153 publications (**Supplementary Table S1** online). Of the 215 variants identified, 161 were classified as pathogenic and 54 were classified as possibly pathogenic.

Frequency of previously reported HH gene pathogenic variants

The frequency of these variants in the 1000G, ESP, and ExAC data sets was determined using wANNOVAR.¹⁶ Of the 161 pathogenic variants tested, 43 (27%) were present in either the 1000G, ESP, or ExAC data sets (**Table 1**). All 5 HH genes were represented among these 43 variants: 12 in *HFE*, 13 in *HFE2*, 1 in *HAMP*, 9 in *TFR2*, and 8 in *SLC40A1*. As expected, the most common pathogenic HH variant present in all data sets was the *HFE* p.Cys282Tyr variant, which was present with allele

Table 1 Frequencies of previously reported hereditary hemochromatosis gene pathogenic variants in the 1000 Genomes Project, Exome Sequencing Project, and Exome Aggregation Consortium data sets

Genomic position (hg19)	Gene	Nucleotide change	Amino acid change	SNP number	1000G (allele frequency)	ESP6500 (allele frequency)	ExAC (allele frequency)
chr6:26091188C>T	HFE	c.196C>T	p.Arg66Cys				0.0000247
chr6:26091191C>T	HFE	c.199C>T	p.Arg67Cys				0.0000165
chr6:26091203C>T	HFE	c.211C>T	p.Arg71*				0.0000247
chr6:26091269delG	HFE	c.279del	p.Trp94Glyfs*117				0.00000824
chr6:26091679delC	HFE	c.480del	p.Arg161Glyfs*50				0.00000826
chr6:26091749T>C	HFE	c.548T>C	p.Leu183Pro	rs199916850	0.0002		0.0000083
chr6:26091818G>T	HFE	c.616+1G>T	Splicing				0.00000853
chr6:26092967G>A	HFE	c.671G>A	p.Arg224Gln	rs62625346	0.0002	0.0002	0.0001
chr6:26092987_26092989delTAC	HFE	c.691_693del	p.Tyr231del				0.000033
chr6:26092993C>T	HFE	c.697C>T	p.Gln233*				0.00000824
chr6:26093141G>A	HFE	c.845G>A	p.Cys282Tyr	rs1800562	0.013	0.048	0.0324
chr6:26093461G>A	HFE	c.1006+1G>A	Splicing	rs573745685	0.0006		0.0000665
chr1:145415419T>C	HFE2	c.238T>C	p.Cys80Arg	rs28940586			0.0000085
chr1:145415483T>C	HFE2	c.302T>C	p.Leu101Pro	rs74315327			0.00000865
chr1:145415675T>A	HFE2	c.494T>A	p.Leu165*				0.00000836
chr1:145415697C>G	HFE2	c.516C>G	p.Asp172Glu				0.00000835
chr1:145415762T>C	HFE2	c.581T>C	p.Leu194Pro				0.00000834
chr1:145416320T>A	HFE2	c.665T>A	p.Ile222Asn	rs74315325		0.0001	0.00000824
chr1:145416517C>T	HFE2	c.862C>T	p.Arg288Trp				0.0000165
chr1:145416609_145416610insG	HFE2	c.960dup	p.Cys321Valfs*21			0.00024	
chr1:145416614G>T	HFE2	c.959G>T	p.Gly320Val	rs74315323		0.0003	0.0002
chr1:145416617_145416618delinsAA	HFE2	c.962_963delG CinsAA	p.Cys321*				0.00000824
chr1:145416618C>G	HFE2	c.963C>G	p.Cys321Trp				0.0000165
chr1:145416631C>T	HFE2	c.976C>T	p.Arg326*	rs74315324		0.0001	0.0000165
chr1:145416808C>T	HFE2	c.1153C>T	p.Arg385*				0.00000824
chr7:100238469G>A	TFR2	c.313C>T	p.Arg105*	rs80338878			0.0000166
chr7:100231138A>T	TFR2	c.515T>A	p.Met172Lys	rs80338879			0.00000898
chr7:100231035T>C	TFR2	c.614+4A>G	Splicing				0.00001
chr7:100228596G>A	TFR2	c.1186C>T	p.Arg396*	rs80338882			0.00000855
chr7:100226978C>T	TFR2	c.1288G>A	p.Gly430Arg				0.00000825
chr7:100225599T>C	TFR2	c.1538-2A>G	Splicing	rs371347559		0.0001	0.0000165
chr7:100218667G>A	TFR2	c.2219C>T	p.Thr740Met	rs150806077	0.0004		0.0000244
chr7:100218543C>T	TFR2	c.2343G>A	p.Trp781*				0.00000835
chr7:100218512C>T	TFR2	c.2374G>A	p.Gly792Arg	rs80338891		0.0002	
chr19:35775913C>T	HAMP	c.223C>T	p.Arg75*	rs200488037	0.0002		0.0000165
chr2:190445231_190445245del	SLC40A1	c.-59_-45del	5'UTR				0.0000413
chr2:190430301A>G	SLC40A1	c.539T>C	p.Ile180Thr				0.0000661

1000G, 1000 Genomes Project; ESP, Exome Sequencing Project; ExAC, Exome Aggregation Consortium; HH, hereditary hemochromatosis; SNP, single-nucleotide polymorphism.

Table 1 Continued on next page

Table 1 Continued

Genomic position (hg19)	Gene	Nucleotide change	Amino acid change	SNP number	1000G (allele frequency)	ESP6500 (allele frequency)	ExAC (allele frequency)
chr2:190430230C>T	<i>SLC40A1</i>	c.610G>A	p.Gly204Ser	rs387907377			0.0000165
chr2:190428912C>T	<i>SLC40A1</i>	c.800G>A	p.Gly267Asp	rs104893664			0.0000247
chr2:190428903T>A	<i>SLC40A1</i>	c.809A>T	p.Asp270Val	rs368420430		0.0003	0.0000742
chr2:190428601G>A	<i>SLC40A1</i>	c.1111C>T	p.Arg371Trp	rs146240138	0.0008	0.0006	0.0001
chr2:190428600C>T	<i>SLC40A1</i>	c.1112G>A	p.Arg371Gln	rs387907378			0.00000825
chr2:190428310C>T	<i>SLC40A1</i>	c.1402G>A	p.Gly468Ser				0.00000844
Combined pathogenic allele frequencies for each HH gene							
	<i>HFE</i>				0.014	0.0482	0.0327 (0.0319–0.0339)
	<i>HFE (C282Y)</i>				0.013	0.048	0.0324 (0.0315–0.0334)
	<i>HFE (non-C282Y)</i>				0.001	0.0002	0.000307 (0.000236–0.000441)
	<i>HFE2</i>					0.00074	0.000316 (0.000209–0.000405)
	<i>TFR2</i>				0.0004	0.0003	0.000102 (0.000051–0.000173)
	<i>HAMP</i>				0.0002		0.0000165
	<i>SLC40A1</i>				0.0008	0.0009	0.00034 (0.000292–0.000517)

1000G, 1000 Genomes Project; ESP, Exome Sequencing Project; ExAC, Exome Aggregation Consortium; HH, hereditary hemochromatosis; SNP, single-nucleotide polymorphism.

frequencies of 1.3, 4.8, and 3.24% in the 1000G, ESP, and ExAC data sets, respectively (Table 1). When broken down into subpopulations, the frequency of *HFE* p.Cys282Tyr was highest in the European-American cohort of ESP (6.4%), the non-Finnish European cohort of ExAC (5.14%), and the European cohort of 1000G (4.3%) (Supplementary Table S2 online). The allele frequencies for *HFE* p.Cys282Tyr are in good agreement with those observed in other population studies of European individuals.⁵ The *HFE* p.Cys282Tyr mutation also was present with lower and varied allele frequencies among other subpopulations (Supplementary Table S2 online); it was lowest among East Asians (0% in 1000G; 0.01% in ExAC). Of the 43 pathogenic variants identified, nearly all ($n = 41$) were present in the larger ExAC data set, 10 in ESP, and 7 in 1000G; only 3 were present in all 3 data sets (Supplementary Table S2 online). Of the non-*HFE* p.Cys282Tyr variants, all had low allele frequencies relative to p.Cys282Tyr. For each of the five HH genes, we summed the allele frequencies for all previously reported pathogenic variants to determine the combined allele frequencies of all pathogenic variants for each gene and in each data set (Table 1) and subpopulation (Supplementary Table S2 online).

Frequency of unreported HH gene major functional variants

In addition to variants that have been previously associated with HH, we identified variants that would be predicted to have major functional or structural effects on the encoded proteins and are therefore likely pathogenic but not previously reported among patients with HH. These included frameshift, premature stop

codon, initiator codon, splice donor, and splice acceptor variants. We identified 40 such variants (Table 2 and Supplementary Table S3 online): 37 were present in the ExAC data set and 5 in the ESP data set; 2 of these were present in both the ExAC and ESP data sets. None were found in the 1000G data set.

Predicted frequency of pathogenic HH genotypes

Using the pathogenic allele frequencies from the previously reported HH variants (Table 1 and Supplementary Table S2 online), we estimated the heterozygous and homozygous pathogenic genotype frequencies for each gene using the Hardy-Weinberg equation (Table 3). The affected genotype carrier rates were calculated from the homozygous genotype frequencies for the *HFE*, *HFE2*, *HAMP*, and *TFR2* genes and from the heterozygous genotype frequencies for *SLC40A1*; these were expressed as 1 per “ n ” of the population that would be predicted to carry HH pathogenic genotypes (Table 3). This analysis was performed for the total 1000G, ESP, and ExAC data sets (Table 3), and also for the different subpopulations of the 1000G (Supplementary Table S4 online), ESP (Supplementary Table S5 online), and ExAC (Supplementary Table S6 online) data sets. The same analysis was performed by combining the allele frequencies derived from the previously reported pathogenic HH variants (Table 1 and Supplementary Table S2 online) with those derived from the unreported major functional variants (Table 2 and Supplementary Table S3 online). The major functional variants are highly likely to alter protein structure and function and hence to be pathogenic; they were included in the analysis

Table 2 Frequencies of unreported hereditary hemochromatosis gene major functional variants in the Exome Sequencing Project and Exome Aggregation Consortium data sets

Genomic position (hg19)	Gene	Nucleotide change	Amino acid change	SNP number	ESP (allele frequency)	ExAC (allele frequency)
chr6:26087746T>C	HFE	c.76+2T>C	Splicing			0.0000085
chr6:26091775C>T	HFE	c.574C>T	p.Gln192*	rs140957442	0.000077	0.0000083
chr6:26093026C>T	HFE	c.730C>T	p.Gln244*			0.0000082
chr6:26093088C>G	HFE	c.792C>G	p.Tyr264*			0.0000082
chr6:26093186G>A	HFE	c.890G>A	p.Trp297*			0.0000082
chr6:26093188G>T	HFE	c.892G>T	p.Glu298*			0.0000082
chr6:26093353C>A	HFE	c.899C>A	p.Ser300*			0.0000082
chr6:26091747_26091748delCC	HFE	c.546_547del	p.Leu183Glyfs*60			0.0000249
chr6:26092961_26092962delCT	HFE	c.667_668del	p.Leu223Thrfs*20			0.0000082
chr6:26093421delG	HFE	c.968del	p.Gly323Glufs*5		0.00008	0.0000082
chr6:26094422_26094423insG	HFE	c.1015_1016insG	p.Met339Serfs*8			0.0000330
chr6:26087670T>G	HFE	c.2T>G	initiator codon			0.0000084
chr1:145416640C>T	HFE2	c.985C>T	p.Arg329*			0.0000082
chr1:145414789delA	HFE2	c.8del	p.Glu3Glyfs*18			0.0000082
chr1:145414839_145414840insT	HFE2	c.59dup	p.Ser21Lysfs*25			0.0000824
chr1:145415347_145415348delGG	HFE2	c.169_170del	p.Gly57Phefs*92			0.0000083
chr1:145415645delC	HFE2	c.466del	p.Arg156Glyfs*90			0.0000085
chr1:145415798delC	HFE2	c.621del	p.Met208Trpfs*38			0.0000083
chr1:145416838delC	HFE2	c.1183del	p.His395Ilefs*17			0.0000082
chr1:145414782A>T	HFE2	c.1A>T	initiator codon			0.0000082
chr1:145416419delGT	HFE2	c.765_766del	p.Asp256Profs*13		0.00008	
chr7:100229427A>C	TFR2	c.1106+2T>G	Splicing			0.0000084
chr7:100225445T>C	TFR2	c.1606-2A>G	Splicing			0.0000165
chr7:100238497T>C	TFR2	c.287-2A>G	Splicing			0.0000083
chr7:100228650G>A	TFR2	c.1132C>T	p.Gln378*			0.0000373
chr7:100238457G>A	TFR2	c.325C>T	p.Gln109*			0.0000083
chr7:100238479G>T	TFR2	c.303C>A	p.Tyr101*			0.0000083
chr7:100238848G>A	TFR2	c.37C>T	p.Gln13*			0.0000846
chr7:100225922delC	TFR2	c.1398del	p.Arg468Alafs*23			0.0000999
chr7:100228650delG	TFR2	c.1132del	p.Gln378Lysfs*8			0.0000124
chr7:100229524delG	TFR2	c.1011del	p.Phe338Serfs*31			0.0000083
chr7:100231157delC	TFR2	c.496del	p.Val166Trpfs*10			0.0000090
chr7:100225543delCT	TFR2	c.1591_1592del	p.Ser531Cysfs*17		0.00008	
chr19:35775691G>A	HAMP	c.91-1G>A	Splicing			0.0000082
chr19:35775840G>A	HAMP	c.151-1G>A	Splicing			0.0000330
chr19:35773482T>A	HAMP	c.2T>A	initiator codon			0.0000165
chr19:35775905delG	HAMP	c.215del	p.Cys72Serfs*48		0.00008	
chr2:190444538C>T	SLC40A1	c.111+1G>A	Splicing			0.0000082
chr2:190426702G>A	SLC40A1	c.1618C>T	p.Arg540*			0.0000082
chr2:190428442_190428443insT	SLC40A1	c.1269dup	p.Pro424Thrfs*2			0.0000082

ESP, Exome Sequencing Project; ExAC, Exome Aggregation Consortium; SNP, single-nucleotide polymorphism.

in an attempt to obtain a more accurate picture of the population prevalence of HH genotypes. The heterozygous and homozygous genotype frequencies and the affected genotype carrier rates for the combined analysis are presented in

Table 4, and also are broken down into ESP and ExAC sub-populations (**Supplementary Tables S7 and S8** online).

Because the ExAC data set is significantly larger than the ESP and 1000G data sets (>60,000 unrelated individuals compared

Table 3 Previously reported hereditary hemochromatosis (HH) gene pathogenic variant heterozygote frequency, homozygote frequency, and predicted HH genotype carrier rates in the 1000 Genomes Project, Exome Sequencing Project, and Exome Aggregation Consortium data sets

Gene	1000G				ESP				ExAC			
	Het	Hom	HH genotype carrier rate	HH genotype carrier rate	Het	Hom	HH genotype carrier rate	HH genotype carrier rate	Het	Hom	HH genotype carrier rate	HH genotype carrier rate
<i>HFE</i>	0.0276	0.000196	5,102	0.0918	0.00232	0.00232	430	0.0633 (0.0618–0.0655)	0.0633 (0.0618–0.0655)	0.00107 (0.00102–0.00115)	935 (869–982)	935 (869–982)
<i>HFE-C282Y</i>	0.0257	0.000169	5,917	0.0914	0.0023	0.0023	434	0.0627 (0.061–0.0646)	0.0627 (0.061–0.0646)	0.00105 (0.000992–0.00112)	953 (896–1,008)	953 (896–1,008)
<i>HFE-non-C282Y</i>	0.002	0.000001	1,000,000	0.0004	0.00000004	0.00000004	25,000,000	0.000614 (0.000472–0.000882)	0.000614 (0.000472–0.000882)	0.00000009 (0.000000056–0.000000194)	10,618,353 (5,141,890–17,954,611)	10,618,353 (5,141,890–17,954,611)
<i>HFE2</i>	0	0	N/A	0.00148	0.00000055	0.00000055	1,828,044	0.000632 (0.000418–0.00081)	0.000632 (0.000418–0.00081)	0.00000001 (0.000000044–0.000000164)	9,993,371 (6,096,632–22,893,249)	9,993,371 (6,096,632–22,893,249)
<i>TFR2</i>	0.0008	0.00000016	6,250,000	0.0006	0.00000009	0.00000009	11,111,111	0.000203 (0.000102–0.000346)	0.000203 (0.000102–0.000346)	0.00000001 (0.000000003–0.000000003)	96,755,165 (33,412,409–384,467,512)	96,755,165 (33,412,409–384,467,512)
<i>HAMP</i>	0.0004	0.00000004	25,000,000	0	0	0	N/A	0.000329	0.000329	0.000000003	3,686,487,806	3,686,487,806
<i>SLC40A1</i>	0.0016	0.00000064	626	0.0018	0.00000081	0.00000081	556	0.000679 (0.000584–0.00103)	0.000679 (0.000584–0.00103)	0.000000115 (0.000000085–0.000000267)	1,473 (968–1,713)	1,473 (968–1,713)

95% Confidence intervals are shown in parentheses.

1000G, 1000 Genomes Project; ESP, Exome Sequencing Project; ExAC, Exome Aggregation Consortium; Het, heterozygote frequency; Hom, homozygote frequency.

with >6,500 and >2,500, respectively) and also includes data from ESP and 1000G, we present more detailed results only for the ExAC data set in the following sections.

Predicted frequency of *HFE* pathogenic genotypes

Our analysis revealed that, based on the ExAC cohort, 1 of 927 (95% CI: 860–970) individuals would be predicted to carry a pathogenic genotype that could cause *HFE* HH (Table 4). We also broke down the *HFE* pathogenic genotypes into the common p.Cys282Tyr homozygous genotype and pathogenic genotypes that do not carry the p.Cys282Tyr variant. As expected, the p.Cys282Tyr homozygous genotype was vastly more frequent than non-p.Cys282Tyr pathogenic genotypes, with carrier rates of 1 in 953 (95% CI: 896–1,008) and 1 in 4,991,817 (95% CI: 2,759,351–7,846,276), respectively (Table 4). However, this does not take into account compound heterozygous genotypes containing p.Cys282Tyr and another pathogenic *HFE* variant. Based on the genotype frequencies of the combined *HFE* pathogenic genotypes and the p.Cys282Tyr homozygous genotype, we can predict that approximately 1 of every 37 patients with *HFE* HH—or around 1 in 34,000 of the population—will be compound heterozygous for p.Cys282Tyr and another pathogenic *HFE* variant. The same analysis was performed on the different ethnic subpopulations of ExAC (Supplementary Table S8 online). This analysis revealed that the highest prevalence of *HFE* pathogenic genotypes is in the non-Finnish European population, with a predicted pathogenic genotype carrier rate of 1 in 373. This is largely made up of p.Cys282Tyr homozygotes (1 in 379). The lowest prevalence of *HFE* pathogenic genotypes was observed in the East Asian population (pathogenic genotype carrier rate of 1 in 286,530) and, unlike all other populations, *HFE* HH genotypes were largely made up of non-p.Cys282Tyr variants (1 in 319,857), with only 1 in 100,000,000 predicted to be p.Cys282Tyr homozygous and only 1 in 10 *HFE* pathogenic genotypes predicted to contain a p.Cys282Tyr allele.

Predicted frequency of non-*HFE* pathogenic genotypes

Analysis of variants of the genes causing recessive forms of HH revealed that pathogenic variants causing these forms of iron overload are rare (Table 4). Based on the ExAC data set, homozygous *HFE2* pathogenic genotypes were predicted to be present in 1 in 4,791,406 (95% CI: 3,121,527–9,238,643) of the population, with the highest prevalence predicted in the South Asian population (1 in 1,611,664; Supplementary Table S8 online). Homozygous *TFR2* pathogenic genotypes have a predicted carrier rate of approximately 1 in 6,164,757 (95% CI: 3,429,355–12,846,700) within the overall ExAC data set (Table 4) and are most frequent among the non-Finnish European population (1 in 753,395; Supplementary Table S8 online). Homozygous *HAMP* pathogenic genotypes were predicted to be very rare based on the ExAC data set (1 in 181,700,486; 95% CI: 43,282,548–493,827,160). This very low prevalence is likely due to only one previously reported pathogenic allele (p.Arg75*) being observed on two occasions among the entire ExAC cohort of >60,000. One of these was in the

Table 4 Hereditary hemochromatosis (HH) gene combined previously reported pathogenic and major functional variant heterozygote frequency, homozygote frequency, and predicted HH genotype carrier rates in the Exome Sequencing Project and Exome Aggregation Consortium data sets

Gene	ESP			ExAC		
	Combined allele frequency	Het	Hom	HH genotype carrier rate	Het	Hom
<i>HFE</i>	0.0484	0.092	0.00234	428	0.0635 (0.0321–0.0341)	0.00108 (0.00103–0.00116)
<i>HFE</i> -C282Y	0.048	0.09139	0.00230	434	0.0627 (0.0315–0.0334)	0.00105 (0.000992–0.00112)
<i>HFE</i> -non-C282Y	0.000357	0.000714	0.00000013	7,846,276	0.000895 (0.000714–0.0012)	0.0000002 (0.000000127–0.000000362)
<i>HFE2</i>	0.00082	0.00164	0.000000067	1,488,602	0.000913 (0.000658–0.00113)	0.00000021 (0.000000108–0.000000302)
<i>TFR2</i>	0.00038	0.00076	0.000000014	610,352	0.000805 (0.000279–0.00054)	0.00000016 (0.000000078–0.000000292)
<i>HAMP</i>	0.00008	0.00016	0.000000006	156,250,000	0.000148 (0.000045–0.000152)	0.000000006 (0.000000002–0.000000023)
<i>SLC40A1</i>	0.0009	0.0018	0.000000081	556	0.000728 (0.000313–0.000544)	0.00000013 (0.000000098–0.000000296)

95% Confidence intervals are shown in parentheses.

ESP, Exome Sequencing Project; ExAC, Exome Aggregation Consortium; het, heterozygote frequency; hom, homozygote frequency.

East Asian population and one in the non-Finnish European population. Four additional predicted functional variants also were observed (Table 2, Supplementary Table S3 online), with the highest frequency among the American population (1 in 4,680,732; Supplementary Table S8 online).

Analysis of variants in the *SLC40A1* gene, causing autosomal-dominant ferroportin disease, revealed allele frequencies for eight previously reported pathogenic variants and three additional major functional variants. The combined pathogenic allele frequencies were 0.0364% (95% CI: 0.000313–0.000544) for the overall ExAC data set, giving a predicted pathogenic genotype carrier rate of 1 in 1,373 (95% CI: 920–1,598). This figure approaches the frequency of *HFE* HH and is largely due to relatively high allele frequencies for two *SLC40A1* variants (p.Asp270Val and p.Arg371Trp) in the African population. The predicted *SLC40A1* pathogenic genotype carrier rate of these two variants is 1 in 197 among the African population.

DISCUSSION

The high prevalence of HH among European populations is almost entirely due to the high frequency of the *HFE* p.Cys282Tyr mutation. Many studies have determined the population prevalence of this mutation, including large population screening studies such as the North American Hemochromatosis and Iron Overload Screening (HEIRS) study¹⁷ and the Melbourne HealthIron study.¹⁸ Southern European populations have lower carrier rates for p.Cys282Tyr,^{19,20} and the proportion of HH that is related to non-*HFE* genes is higher in southern Europe and other parts of the world where the p.Cys282Tyr mutation is less common.²⁰ In Asian countries the majority of HH cases are related to non-*HFE* genes.²¹ Despite the wealth of publications on the various causes of non-*HFE* HH, to our knowledge no systematic study has attempted to estimate the prevalence of these atypical forms of iron overload. The frequency of individual non-*HFE* mutations among the general population have only been estimated to be rare; no other quantitative measure of their frequency is available.

The 1000G,¹⁵ ESP, and ExAC data sets provide a huge resource of genomic information that can be mined for a wide variety of applications. We have taken advantage of these data sets to obtain a first estimate of the prevalence of *HFE* and non-*HFE* disease-causing variants. By far the most common pathogenic variant was *HFE* p.Cys282Tyr, which is present at the highest frequency among non-Finnish Europeans (allele frequency 5.14%), similar to that reported in various European populations (allele frequencies ranging from 0 to 14%).^{5,17,22} An additional 42 HH-causing variants were identified among the three data sets, and 40 unreported variants that are predicted to have major functional or structural effects on the encoded proteins were also used in our analysis. Using the allele frequencies of these variants, we could determine that all the recessively inherited forms of non-*HFE* HH are predicted to be rare. Pathogenic *HFE2* and *TFR2* variants were predicted to cause iron overload in approximately 1 in 5 to 6 million people. Pathogenic *HAMP* variants were even rarer. Given that heterozygosity for

mutations in *SLC40A1* lead to ferroportin disease, we were surprised to find relatively high frequencies of *SLC40A1* variants classified as pathogenic. These variants were found at the highest frequencies among African populations (0.25%) but were also present in the American (0.039%), East Asian (0.033%), and non-Finnish European (0.03%) populations. Two *SLC40A1* variants—p.Asp270Val and p.Arg371Trp—solely contributed to the high allele frequency among the African populations. The *SLC40A1* p.Asp270Val variant was originally reported in a 27-year-old black South African female with iron overload.²³ The p.Asp270Val variant was also reported in an African-American male reported to have mild iron overload and coexistent hepatitis C virus infection.²⁴ He also had stainable iron in hepatocytes and Kupffer cells, a pattern consistent with ferroportin disease. In this study the p.Asp270Val variant also was detected in one of 258 African-American controls; however, no clinical data from this individual were available.²⁴ One study determined that the p.Asp270Val variant had no effect on the hepcidin sensitivity of ferroportin; however, its effect on iron transport was not measured.²⁵ The p.Arg371Trp variant was described as part of a series of new *SLC40A1* variants in a family with iron overload.²⁶ In the same publication, another variant affecting the same amino acid (p.Arg371Gln), similarly detected at a low frequency in the ExAC non-Finnish European population, was also reportedly associated with iron overload.²⁶ The observations linking the *SLC40A1* p.Asp270Val and p.Arg371Trp variants to iron overload and their relatively high allele frequencies point to a potential role for these variants as a cause of iron overload in African populations. However, further functional studies or larger population studies are required to unequivocally link these variants to impaired ferroportin function and iron overload in individuals carrying these variants. The other six *SLC40A1* variants identified in the ExAC data set (c.-59_-45del, p.Ile180Thr, p.Gly204Ser, p.Gly267Asp, p.Arg371Gln, and p.Gly468Ser) have all been associated with iron overload in patients and/or families,^{26–33} although ferroportin containing the p.Ile180Thr variant seemed to behave similarly to the wild type when analyzed in functional assays.³⁰ The *SLC40A1* p.Val162del variant has been reported as a cause of ferroportin disease more frequently (in nine publications) than any other *SLC40A1* mutation (**Supplementary Table S1** online); however, we did not detect it in any of the genomic sequence databases. It is widely believed that this 3-bp deletion has occurred multiple times in various populations, and there is now evidence for this, with the identification of a de novo p.Val162del variant occurring in an isolated case of ferroportin disease.³⁴ These data suggest that p.Val162del is more frequently identified than other *SLC40A1* mutations because it has occurred multiple times in isolated populations rather than occurring once and spreading to different populations, and these factors may explain why we did not detect it in this study.

We used freely available NGS data in an attempt to estimate the population prevalence of non-*HFE* HH. While we arrived at estimates for the prevalence of non-*HFE* HH using three data sets, there are some limitations to this study. (i) The size

of the populations studied may be enough to determine the prevalence of more common inherited disorders but may be less accurate for rarer inherited disorders. We believe the larger ExAC data set is likely to provide more accurate estimates. (ii) The geographic and ethnic origins of the populations studied are diverse but are still not fully representative of the world's population because they are more biased toward populations with a European background. It is possible that geographic and ethnic differences in the prevalence of non-*HFE* HH have not been captured by our analysis. For example, non-*HFE* HH seems to be more prevalent in southern Europe compared with northern Europe.²⁰ (iii) Our analysis relied on variant data collated from published reports, together with the addition of variants that would be expected to have major functional effects on the encoded proteins; it is possible that other HH-causing variants have been either not identified or not reported in the literature and hence would not have been captured in our study. (iv) The accuracy of the publically available NGS data is another limitation. While on the whole the variant calls for the data sets are likely to be highly accurate, the depth of coverage across samples could be variable, and it has been shown that not all variant calls across NGS platforms can be validated using more traditional approaches such as Sanger sequencing.³⁵ (v) Our study does not take into account potential differences in the penetrance of variants or the possibility that iron overload can be caused by digenic inheritance of variants in *HFE* and other non-*HFE* genes.³⁶ For example, the p.His63Asp and p.Ser65Cys variants in *HFE* are known to contribute to some cases of iron overload but have a much lower penetrance than p.Cys282Tyr.^{7,8} We decided to exclude these variants from our analysis so as not to over inflate population prevalence estimates.

The method we used to estimate the prevalence of non-*HFE* HH is very straightforward and could also be applied to estimate the prevalence of a wide variety of other inherited disorders. Indeed, similar methods have been recently used to determine the prevalence of McArdle disease.³⁷ However, the composition of the population data sets should be taken into account when applying this technique to other diseases. While the 1000G data set represents control individuals from geographically and ethnically diverse populations, the larger ESP and ExAC data sets are made up predominantly of European patient and control cohorts that have been used to study a variety of diseases. Hence it would not be advisable to use these data to study the prevalence of variants that may increase the risk of diseases included in these data sets, such as diabetes and heart and lung disorders. For example, the ESP data set includes a cohort of patients that has been used to study modifier genes in cystic fibrosis.³⁸ Hence, *CFTR* pathogenic variants are over-represented in the ESP data set. We believe that it is unlikely that the makeup of the ESP and ExAC data sets has greatly affected our estimation of the prevalence of HH. While iron accumulation in the pancreas can cause diabetes and in the heart can lead to cardiomyopathy, there is little evidence that HH variants increase the risk of these diseases among the general population.^{39,40} In support of this, the prevalence of the *HFE* p.Cys282Tyr variant within

the ESP data set is very similar to that observed in large population studies from North America,^{17,22} indicating that there is no enrichment of this mutation.

The current explosion in NGS has the potential to greatly increase the available data from genomic sequencing projects. In the future, variant data are likely to be available from a larger number of human genomes and from more ethnically diverse populations, allowing more reliable and accurate measures of variant frequency and predicted population prevalence of non-HFE HH and a wide variety of other inherited diseases.

SUPPLEMENTARY MATERIAL

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

ACKNOWLEDGMENTS

The authors thank Jonathan Ellis for assistance with bioinformatic analysis and Cameron McDonald for useful discussions. D.F.W. was supported by a Gastroenterological Society of Australia Senior Research Fellowship. V.N.S. is a National Health and Medical Research Council Senior Research Fellow (APP1024672).

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

- Darshan D, Frazer DM, Anderson GJ. Molecular basis of iron-loading disorders. *Expert Rev Mol Med* 2010;12:e36.
- Wallace DF, Subramaniam VN. Non-HFE haemochromatosis. *World J Gastroenterol* 2007;13:4690–4698.
- Feder JN, Gnirke A, Thomas W, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet* 1996;13:399–408.
- Bacon BR, Powell LW, Adams PC, Kresina TF, Hoofnagle JH. Molecular medicine and hemochromatosis: at the crossroads. *Gastroenterology* 1999;116:193–207.
- Lucotte G, Dieterlen F. A European allele map of the C282Y mutation of hemochromatosis: Celtic versus Viking origin of the mutation? *Blood Cells Mol Dis* 2003;31:262–267.
- Merryweather-Clarke AT, Pointon JJ, Shearman JD, Robson KJ. Global prevalence of putative haemochromatosis mutations. *J Med Genet* 1997;34:275–278.
- Beutler E. The significance of the 187G (H63D) mutation in hemochromatosis. *Am J Hum Genet* 1997;61:762–764.
- Walsh A, Dixon JL, Ramm GA, et al. The clinical relevance of compound heterozygosity for the C282Y and H63D substitutions in hemochromatosis. *Clin Gastroenterol Hepatol* 2006;4:1403–1410.
- Nemeth E, Tuttle MS, Powelson J, et al. Heparin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. *Science* 2004;306:2090–2093.
- Bridle KR, Frazer DM, Wilkins SJ, et al. Disrupted hepcidin regulation in HFE-associated haemochromatosis and the liver as a regulator of body iron homeostasis. *Lancet* 2003;361:669–673.
- Papanikolaou G, Samuels ME, Ludwig EH, et al. Mutations in HFE2 cause iron overload in chromosome 1q-linked juvenile hemochromatosis. *Nat Genet* 2004;36:77–82.
- Roetto A, Papanikolaou G, Politou M, et al. Mutant antimicrobial peptide hepcidin is associated with severe juvenile hemochromatosis. *Nat Genet* 2003;33:21–22.
- Camaschella C, Roetto A, Cali A, et al. The gene TFR2 is mutated in a new type of haemochromatosis mapping to 7q22. *Nat Genet* 2000;25:14–15.
- Mayr R, Janecke AR, Schranz M, et al. Ferroportin disease: a systematic meta-analysis of clinical and molecular findings. *J Hepatol* 2010;53:941–949.
- The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–1073.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
- Adams PC, Reboussin DM, Barton JC, et al.; Hemochromatosis and Iron Overload Screening (HEIRS) Study Research Investigators. Hemochromatosis and iron-overload screening in a racially diverse population. *N Engl J Med* 2005;352:1769–1778.
- Allen KJ, Gurrin LC, Constantine CC, et al. Iron-overload-related disease in HFE hereditary hemochromatosis. *N Engl J Med* 2008;358:221–230.
- Sánchez M, Villa M, Ingelmo M, et al. Population screening for hemochromatosis: a study in 5370 Spanish blood donors. *J Hepatol* 2003;38:745–750.
- Piperno A, Sampietro M, Pietrangelo A, et al. Heterogeneity of hemochromatosis in Italy. *Gastroenterology* 1998;114:996–1002.
- McDonald CJ, Wallace DF, Crawford DH, Subramaniam VN. Iron storage disease in Asia-Pacific populations: the importance of non-HFE mutations. *J Gastroenterol Hepatol* 2013;28:1087–1094.
- Beutler E, Felitti VJ, Koziol JA, Ho NJ, Gelbart T. Penetrance of 845G→A (C282Y) HFE hereditary haemochromatosis mutation in the USA. *Lancet* 2002;359:211–218.
- Zaahl MG, Merryweather-Clarke AT, Kotze MJ, van der Merwe S, Warnich L, Robson KJ. Analysis of genes implicated in iron regulation in individuals presenting with primary iron overload. *Hum Genet* 2004;115:409–417.
- Lee PL, Gaasterland T, Barton JC. Mild iron overload in an African American man with SLC40A1 D270V. *Acta Haematol* 2012;128:28–32.
- Rice AE, Mendez MJ, Hokanson CA, Rees DC, Björkman PJ. Investigation of the biophysical and cell biological properties of ferroportin, a multipass integral membrane protein iron exporter. *J Mol Biol* 2009;386:717–732.
- Le Lan C, Mosser A, Ropert M, et al. Sex and acquired cofactors determine phenotypes of ferroportin disease. *Gastroenterology* 2011;140:1199–1207.e1.
- Cunat S, Giansily-Blaizot M, Bismuth M, et al.; CHU Montpellier AOI 2004 Working Group. Global sequencing approach for characterizing the molecular background of hereditary iron disorders. *Clin Chem* 2007;53:2060–2069.
- Del-Castillo-Rueda A, Moreno-Carralero MI, Cuadrado-Grande N, et al. Mutations in the HFE, TFR2, and SLC40A1 genes in patients with hemochromatosis. *Gene* 2012;508:15–20.
- Bach V, Remacha A, Altés A, Barceló MJ, Molina MA, Baiget M. Autosomal dominant hereditary hemochromatosis associated with two novel Ferroportin 1 mutations in Spain. *Blood Cells Mol Dis* 2006;36:41–45.
- Callebaut I, Joubrel R, Pissard S, et al. Comprehensive functional annotation of 18 missense mutations found in suspected hemochromatosis type 4 patients. *Hum Mol Genet* 2014;23:4479–4490.
- Santos PC, Cançado RD, Pereira AC, et al. Hereditary hemochromatosis: mutations in genes involved in iron homeostasis in Brazilian patients. *Blood Cells Mol Dis* 2011;46:302–307.
- Cremonesi L, Forni GL, Soriani N, et al. Genetic and clinical heterogeneity of ferroportin disease. *Br J Haematol*. 2005;131:663–670.
- Lee PL, Gelbart T, West C, Barton JC. SLC40A1 c.1402G→A results in aberrant splicing, ferroportin truncation after glycine 330, and an autosomal dominant hemochromatosis phenotype. *Acta Haematol* 2007;118:237–241.
- Unal S, Piperno A, Gumruk F. Iron chelation with deferasirox in a patient with de-novo ferroportin mutation. *J Trace Elem Med Biol* 2015;30:1–3.
- Lam HY, Clark MJ, Chen R, et al. Performance comparison of whole-genome sequencing platforms. *Nat Biotechnol* 2012;30:78–82.
- Merryweather-Clarke AT, Cadet E, Bomford A, et al. Digenic inheritance of mutations in HAMP and HFE results in different types of haemochromatosis. *Hum Mol Genet* 2003;12:2241–2247.
- De Castro M, Johnston J, Biesecker L. Determining the prevalence of McArdle disease from gene frequency by analysis of next-generation sequencing data. *Genet Med* 2015; e-pub-ahead of print 5 March 2015.
- Emond MJ, Louie T, Emerson J, et al.; National Heart, Lung, and Blood Institute (NHLBI) GO Exome Sequencing Project; Lung GO. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. *Nat Genet* 2012;44:886–889.
- Hetet G, Elbaz A, Gariepy J, et al. Association studies between hemochromatosis gene mutations and the risk of cardiovascular diseases. *Eur J Clin Invest* 2001;31:382–388.
- Rong Y, Bao W, Rong S, et al. Hemochromatosis gene (HFE) polymorphisms and risk of type 2 diabetes mellitus: a meta-analysis. *Am J Epidemiol* 2012;176:461–472.