Mixture distribution analysis of phenotypic markers reflecting HFE gene mutations

Christine E. McLaren, Kuo-Tung Li, Chad P. Garner, Ernest Beutler, and Victor R. Gordeuk

The goal of this study was to determine whether statistical modeling of population data for a phenotypic marker could reflect a major locus gene defect. Identifying mutations in the HFE gene makes it possible to assess the association between transferrin saturation (TS) subpopulations and HFE mutations. Data were analyzed from 27,895 white patients who attended a health appraisal clinic and who had TS and common mutations of HFE determined. Mixture distribution modeling of TS was performed, and the proportion of HFE mutations in TS subpopulations was assessed on a probability basis. Three subpopulations of TS were identified, consistent with Hardy-Weinberg conditions for major locus effects. For men, 72% of the subpopulation with the highest mean TS had HFE gene mutations; they were primarily homozygotes or compound heterozygotes. Seventy-three percent of the subpopulation with moderate mean TS also had HFE gene mutations; they were predominantly simple heterozygotes. Thirty-seven percent of the subpopulation with the lowest mean TS were wild-type homozygotes. Similar results were observed for women. These results suggest that statistical modeling of population clinical laboratory test data can reveal the influence of a major locus gene defect and perhaps can be applied to other aspects of body metabolism than iron. (Blood. 2003;102:4563-4566)

Introduction

Homozygous hemochromatosis, a common inherited susceptibility to iron overload, occurs in 0.3% to 0.5% of white persons of western European descent. Most cases of hemochromatosis in white persons are attributable to common missense mutations of the HFE gene. The 2 common missense mutations are C282Y (exon 4, nt 845G>A) and H63D (exon 2, nt 187C>G), and 80% to 90% of patients of northern European ancestry with typical hemochromatosis phenotype are C282Y homozygotes. Transferrin saturation (TS) generally is regarded as the best single screening test for hemochromatosis.

We previously used mixture modeling of TS measured in the second and third National Health and Nutrition Examination Surveys in the United States (NHANES II, NHANES III) and in population data from Australia to study possible genetic influences on iron metabolism in white populations. In each study, we found TS subpopulations consistent with Hardy-Weinberg predictions for a major hemochromatosis locus that leads to altered TS values in affected homozygotes and in heterozygotes. In one study, serum ferritin concentrations were measured simultaneously, permitting the determination of 3 subpopulations of TS with progressively increasing mean age-adjusted serum ferritin concentration values, consistent with increasing body iron stores.

For our previous analyses of population data, HFE gene mutation status was unavailable. In this new study, we analyzed a large data set from patients attending a health appraisal clinic. Transferrin saturation, serum ferritin, and HFE mutations were measured for each patient. The goal was to verify that statistical distribution methodology could be used to analyze laboratory test values to predict the presence of major locus mutations that affect body metabolism.

Patients, materials, and methods

Sources of data

The primary source of data was a study conducted at the Kaiser Permanente San Diego Health Appraisal Clinic from 1998 to 2001. All patients older than 20 years of age who were registered with the clinic were apprised of a research project in which DNA analysis for HFE mutations and measurements of serum ferritin concentration would be added to the tests usually performed. We analyzed data from the 30,966 non-Hispanic white men and women, at least 20 years of age, who consented to participate in the Kaiser Permanente study and for whom HFE genotype data and health and demographic data were obtained. The term white is used in this paper to designate non-Hispanic patients who selected the category white to indicate their sole ancestry. For the purposes of the present study, we did not analyze data from Asian, black, or Hispanic clinic patients because HFE mutations are less common in these populations than among white persons. Details of the laboratory analysis of samples for HFE C282Y and H63D mutations, hematologic measurements, serum iron concentration, total iron-binding capacity, transferrin saturation, and serum ferritin concentration have been described previously.
Table 1. Exclusions from the total sample of white participants at least 20 years of age

<table>
<thead>
<tr>
<th>Exclusion</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sample</td>
<td>15,415</td>
<td>15,294</td>
</tr>
<tr>
<td>Samples excluded on the basis of low Hb or MCV &lt;80 fL</td>
<td>1,258</td>
<td>909</td>
</tr>
<tr>
<td>Transferrin saturation not measured</td>
<td>352</td>
<td>295</td>
</tr>
<tr>
<td>Sample for transferrin saturation modeling</td>
<td>13,805</td>
<td>14,090</td>
</tr>
<tr>
<td>Serum ferritin not measured</td>
<td>533</td>
<td>585</td>
</tr>
<tr>
<td>Sample for serum ferritin analysis</td>
<td>13,272</td>
<td>13,505</td>
</tr>
</tbody>
</table>

Hemoglobin (Hb) exclusions: men, Hb < 13.5 g/dL; women, Hb < 12.0 g/dL.

Statistical modeling of transferrin saturations

For analysis, we selected transferrin saturation and serum ferritin values from 15,294 nonpregnant white women and 15,415 white men at least 20 years of age. Patients were excluded if they had abnormally low values for hemoglobin or mean corpuscular volume (MCV) because associated conditions might have altered transferrin saturation and serum ferritin values.15-17 We did not exclude patients with MCV values above the reference range because hemochromatosis probands have mean MCV values significantly higher than wild-type patients.18 Data sets for transferrin saturation modeling consisted of 13,805 men and 14,090 women (Table 1).

The research hypothesis was that HFE gene mutations influence the distribution of transferrin saturation, resulting in genetically based subpopulations. Details of mixture-modeling techniques as applied to transferrin saturation distributions, including parameter estimation, statistical tests, and confidence intervals for proportions, have been described elsewhere.5,19-21

Comparison of mean serum ferritin concentrations within transferrin saturation subpopulations

In a previous investigation of population data, we found evidence of subpopulations of patients, determined on the basis of transferrin saturation, with significantly different iron stores based on mean serum ferritin concentration. Confirmatory analyses were performed in this new study. Considering the transferrin saturation data sets used for mixture modeling, 533 men and 585 women did not have measurements of serum ferritin concentration. After removing values for these patients, the data sets to analyze serum ferritin concentration consisted of 13,272 men and 13,505 women. Because the distributions of serum ferritin concentration were markedly skewed, the square root transformation was applied to each data set. In addition, because serum ferritin concentration tends to increase with age,24,25 we applied methods for age standardization to age 60 years. The probability that an individual serum ferritin concentration value belonged to each subpopulation according to the expected proportions within transferrin saturation subpopulations.8 Given the total sample size, the number of expected observations within each transferrin saturation interval was calculated. As described previously,7 within a given transferrin saturation interval, each observation was then assigned a probability, and these probabilities were used to assign transferrin saturation values to a given subgroup according to the expected proportions within transferrin saturation subpopulations.

Genetic analysis of the genotype data

With rare exceptions27,28 the C282Y and H63D mutations occur in trans; therefore, we observed 3 gametes, defined by the presence or absence of the mutations, and 6 genotypes in the sample. The 6 genotypes are denoted by the mutation that is present or as wild type (wt) if neither mutation exists on the chromosome, giving wt/wt, wt/H63D, wt/C282Y, H63D/C282Y, H63D, and C282Y/C282Y. The absence of chromosomes carrying both mutations confirms that the mutations are in virtually complete linkage disequilibrium.

Genotype data were analyzed to determine whether the observed genotype frequencies fit the expectations under the Hardy-Weinberg equilibrium model. The Hardy-Weinberg model assumes that the 2 alleles (or mutations) in a genotype are independent and that the genotype frequencies are the product of their constituent allele or mutation frequencies. Deviations from the Hardy-Weinberg model can result from an underlying structure in the population (eg, inbreeding or selection bias) or, more commonly, errors in the data. The C282Y and H63D mutation frequencies in the population were estimated separately from the data for 13,805 men and 14,090 women, and expected genotype frequencies were then calculated from the mutation frequencies, assuming Hardy-Weinberg equilibrium. The Kolmogorov-Smirnov goodness-of-fit test was applied to compare observed and expected genotype frequencies.

Association of modeled TS subpopulations and HFE genotypes

For each sex, the frequency and proportion of genotypes occurring within transferrin saturation subpopulations was estimated. Given the total sample size, the number of expected observations within each transferrin saturation interval was calculated. As described previously,5 within a given transferrin saturation interval, each observation was then assigned a probability, and these probabilities were used to assign transferrin saturation values to a given subgroup according to the expected proportions within transferrin saturation subpopulations.

Results

Statistical modeling of transferrin saturations

The primary analysis was performed on transferrin saturations for 27,895 patients. Results of statistical mixture modeling indicated that the fit of the data to a mixture of 3 normal populations with unequal variances was significantly better than the fit to either a mixture of 2 normal populations or a single normal population for white men (likelihood ratio statistics, 157 [P < .01] and 1704 [P < .01], respectively) and women (likelihood ratio statistics, 120 [P < .01] and 649 [P < .01], respectively). Figure 1 shows the distribution of transferrin saturation for men and women. The fitted subpopulations are superimposed over the histograms of the observed data. Table 2 gives mixture-model parameter estimates for transferrin saturation subpopulations and shows that for each sex, transferrin saturation subpopulations were identified with increasing means for men and women (Trend test, P < .001 for each). We also compared the age-standardized mean serum ferritin concentrations for the TS subpopulations (Table 2). Trend tests demonstrated an increase in mean serum ferritin concentration.

Figure 1. Observed and modeled distributions of transferrin saturation values are shown for each ethnicity and sex. Distribution of transferrin saturation values: (A) 13,805 white men and (B) 14,090 white women. A histogram of the observed data is shown. Dashed lines represent the fitted normal distributions representing each subpopulation. Solid line represents the overall fitted mixture distribution.
standardized to age 60 years, with increasing mean transferrin saturation for men ($P < .0001$) and women ($P < .0001$).

**Genetic analysis of the genotype data**

The observed frequencies of the H63D and C282Y mutations were 0.152 and 0.062, respectively, in men and 0.149 and 0.064, respectively, in women. The data show strong agreement with the Hardy-Weinberg equilibrium model, which would be expected for a large, randomly selected sample from a large, randomly mating population. There were no significant differences between observed and expected genotype frequencies (Kolmogorov-Smirnov statistic = 0.167; $P = 1.0$ for men and women).

**Association of modeled TS subpopulations and $HFE$ genotypes**

The frequency of genotypes within each TS subpopulation is given in Table 3. For patients younger than 60 years, serum ferritin concentration values were standardized to those expected at age 60.

The context of iron overload in the white population. Before the $HFE$ gene was identified, we conducted a series of studies based on the postulate that the hemochromatosis mutation in white patients would influence the distribution of transferrin saturation in the population and would permit the identification of distinct subpopulations based on hemochromatosis genotype. We amassed considerable evidence consistent with the hypothesis but had not had the opportunity to verify our hypothesis by studying a population in which transferrin saturations and $HFE$ genotypes were determined simultaneously in all participants. The present study represents our first opportunity to attempt to verify our hypothesis in a large population in whom phenotyping and genotyping were performed.

Under the hypothesis that the distribution of transferrin saturation reflects several populations based on individual genotype for hemochromatosis, in the present study we applied a statistical distribution methodology to transferrin saturations in a large data set and examined the relationship between the model results and the actual genotyping of $HFE$ mutations. Using this approach, 3 transferrin saturation subpopulations were identified in white men and women. The proportions of these subpopulations are consistent with Hardy-Weinberg criteria for a major, common genetic influence on iron metabolism. For men and women, the subpopulation with higher mean TS values consisted predominantly of those with $HFE$ gene mutations. In the subpopulation with the highest TS, the $HFE$ mutations were predominantly homozygote or compound heterozygote. In the subpopulations with moderate TS, the $HFE$ mutations were predominantly simple heterozygote. In contrast, most of the patients in the subpopulation with the lowest TS were predominantly $HFE$ wild-type homozygote.

Our study has several limitations. We were unable to exclude patients with elevations in erythrocyte protoporphyrin or liver function tests or those with positive serum hepatitis B surface antigen or serum hepatitis C antibody because these laboratory tests were not performed for all participants. Patients with abnormalities in these test results might have had elevations in TS not attributable to a genetic defect and might have been part of the proportion with wild-type $HFE$ found in the subpopulation with the highest TS and moderate TS. The first 10 000 patients were tested for the S65C $HFE$ mutation. Because not all patients were tested, we did not exclude those with other $HFE$ mutations, such as S65C, who might also have been classified as having wild-type $HFE$ in the upper TS subpopulation.

<table>
<thead>
<tr>
<th>Subpopulation</th>
<th>Genotype</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>wt/wt</td>
<td>8016</td>
<td>485</td>
<td>33</td>
<td>8534</td>
</tr>
<tr>
<td></td>
<td>H63D/wt</td>
<td>2941</td>
<td>321</td>
<td>27</td>
<td>3289</td>
</tr>
<tr>
<td></td>
<td>C282Y/wt</td>
<td>724</td>
<td>610</td>
<td>7</td>
<td>1341</td>
</tr>
<tr>
<td></td>
<td>H63D/H63D</td>
<td>139</td>
<td>195</td>
<td>2</td>
<td>336</td>
</tr>
<tr>
<td></td>
<td>C282Y/H63D</td>
<td>68</td>
<td>169</td>
<td>9</td>
<td>246</td>
</tr>
<tr>
<td></td>
<td>C282Y/C282Y</td>
<td>7</td>
<td>11</td>
<td>41</td>
<td>59</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>11 895</td>
<td>1791</td>
<td>119</td>
<td>13 805</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subpopulation</th>
<th>Genotype</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women</td>
<td>wt/wt</td>
<td>8271</td>
<td>445</td>
<td>43</td>
<td>8759</td>
</tr>
<tr>
<td></td>
<td>H63D/wt</td>
<td>2934</td>
<td>277</td>
<td>23</td>
<td>3234</td>
</tr>
<tr>
<td></td>
<td>C282Y/wt</td>
<td>709</td>
<td>706</td>
<td>6</td>
<td>1424</td>
</tr>
<tr>
<td></td>
<td>H63D/H63D</td>
<td>146</td>
<td>195</td>
<td>4</td>
<td>345</td>
</tr>
<tr>
<td></td>
<td>C282Y/H63D</td>
<td>81</td>
<td>182</td>
<td>4</td>
<td>267</td>
</tr>
<tr>
<td></td>
<td>C282Y/C282Y</td>
<td>13</td>
<td>14</td>
<td>34</td>
<td>61</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>12 154</td>
<td>1822</td>
<td>114</td>
<td>14 090</td>
</tr>
</tbody>
</table>

**Discussion**

As we gain understanding of the genetic influences on the body’s metabolic processes, it is likely that we will find that routine laboratory measurements performed in clinical medicine are influenced by genetic mutations and polymorphisms in the population. It is also possible that studying the distributions of the results of laboratory measurements in populations can provide evidence for such as yet unidentified major mutations that influence metabolic pathways of the body. An example of this approach is a community-based prevalence study of hypertension, in which values for angiotensinogen (AGT) collected from Nigerian families were analyzed. Although the AGT genetic mutation status of family members was unknown, a mixture of 2 or 3 distributions fit the AGT values significantly better than a single distribution, and a mixture of 3 distributions provided the best fit to the data, suggesting a major genetic effect. We have examined mixture modeling of a phenotypic marker to reveal a major locus effect in

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Subpopulation</th>
</tr>
</thead>
</table>
Even with these limitations, however, modeling demonstrated subpopulations of TS corresponding to distributions of HFE genotypes in this population. Our findings thus provide confirmation that mutations in a major locus influence the distribution of transferrin saturation in the US white population. Our laboratory tests, including those measuring metals, electrolytes, enzyme activities, metabolic breakdown products, and components of blood and plasma.

Acknowledgment

This is manuscript no. 15507-MEM from The Scripps Research Institute.

References

Mixture distribution analysis of phenotypic markers reflecting HFE gene mutations

Christine E. McLaren, Kuo-Tung Li, Chad P. Garner, Ernest Beutler and Victor R. Gordeuk