Interaction Between Two Quantitative Trait Loci Affects Fetal Haemoglobin Expression

C. Garner1,2,∗, S. Menzel1, C. Martin3, N. Silver1, S. Best3, T. D. Spector4 and S. L. Thein3

1Epidemiology Division, Department of Medicine, University of California, Irvine, California,
2Department of Environmental Health, Science, and Policy, University of California, Irvine, California,
3Department of Haematological Medicine, Guy’s, King’s and St. Thomas’ School of Medicine, London, United Kingdom,
4Twin Research and Genetic Epidemiology Unit, St Thomas’ Hospital, London, United Kingdom

Summary

The biological mechanisms controlling complex quantitative traits are likely to be affected by interactions between genetic factors, sometimes referred to as epistasis. The identification of interacting loci through genetic analysis faces many challenges, and few examples of replicated findings of interaction exist for humans and model system organisms. The replication of an interaction, or the non-independence, of two quantitative trait loci (QTL) affecting the developmental switch from the expression of fetal to adult haemoglobin is reported here. Fetal haemoglobin expression in adults is a highly heritable, yet complex, phenotype. Using a sample of 874 dizygotic twin pairs of European descent, we found linkage to a QTL on chromosome 8 to be conditional on the twin pairs’ genotypes at a polymorphism in the β-globin complex; an interaction originally identified in a large Asian Indian kindred. The β-globin polymorphism has been previously shown to be associated with fetal haemoglobin levels in adults. This study reports the first known replication of a genetic interaction between QTLs influencing a complex human trait.

Keywords: fetal haemoglobin, linkage analysis, epistasis, sib pairs, quantitative trait loci.

Introduction

The term epistasis means different things to different geneticists. One common theme among the definitions of epistasis is genetic interaction, or the non-independence of genetic factors, acting on a trait (Frankel & Schork, 1996; Cordell, 2002). The difficulty in defining epistasis comes from the fact that a genetic interaction defined in a statistical model can represent a large number of biological mechanisms; the statistical model of epistasis does not uniquely identify a specific type of biological process. The lack of correspondence between the statistical model and biological reality has led geneticists to refer to the statistically measured non-independence of genetic factors as statistical interaction (Cordell, 2002), realizing that further experimentation is required to determine the nature of the corresponding biological interaction. Statistical interaction becomes more meaningful when it validates a biological hypothesis. However, using statistical models to test existing biological hypotheses can also be problematic because the complex biological mechanism may not be amenable to modelling with realistic data sets, nor measurable with an intuitive statistical model. Selective breeding of model organisms can provide powerful study designs for mapping interacting loci in non-human species. The existence of epistasis is known, but the detection and characterization of biological interactions using human genetic analyses remains a difficult challenge.

During human development, the switch from fetal haemoglobin (Hb F, α2γ2) to adult haemoglobin (Hb A, α2β2) synthesis occurs gradually towards the end of gestation. At birth, Hb F comprises approximately 70% of the total Hb, and declines to adult levels of less than 1% by two years of age. Residual fetal Hb synthesis
continues throughout adult life; the Hb F is restricted to a subset of erythrocytes termed F cells (FC) (Boyer et al. 1975). The baseline levels of Hb F and FC are highly correlated, and vary considerably, by at least 20 fold, in otherwise normal subjects. The heritability of FC levels was estimated to be 0.89 in the European population (Garner et al. 2000b), indicating that the expression of the \( \gamma \) globin gene in adults is under strong genetic control. FC is influenced by age (Rutland et al. 1983), sex (Miyoshi et al. 1988) and a common DNA sequence variant (C to T) at position –158 upstream of the \( \gamma \)-globin gene, referred to as the XmnI-\( \gamma \) polymorphism (Gilman & Huisman, 1985; Sampietro et al. 1992; Garner et al. 2000a). The ‘T’ variant that creates a cleavage site for XmnI restriction has been found to be common in multiple populations; its frequency is approximately 0.33 and the genotype accounts for 13–32% of the total phenotypic variance in FC in the European population (Garner et al. 2000a). The \( \gamma \)-158 T variant is associated with increased Hb F production in adult life, particularly in conditions of erythroid stress such as \( \beta \) thalassemia (Labie et al. 1985; Thein et al. 1987; Ho et al. 1998) and sickle cell anemia (Labie et al. 1985). Unlike rare mutations in the \( \gamma \) globin promoter that are associated with Hb F of 10–35% in heterozygotes (Wood, 2001), and show clear Mendelian segregation, the role of the \( \gamma \)-158 variant is unclear. It has been implicated as a possible cause of slightly elevated Hb F levels in some families, but even within families the association is not consistent, suggesting that the effect of the XmnI-\( \gamma \) site may be modulated by the presence of an intermediary factor(s). Though increased levels of Hb F and FC in adulthood have no clinical importance on their own, there is evidence that high levels of the trait can ameliorate the effects of \( \beta \) thalassemia and sickle cell anemia. Hence, knowledge of the genetic factors contributing to adult Hb F levels could lead to improved therapeutic strategies for Hb F augmentation that could benefit patients with severe haemoglobinopathies.

A statistical interaction between the XmnI-\( \gamma \) polymorphism in the \( \beta \)-globin complex and a quantitative trait locus (QTL) on chromosome 8 was previously shown to influence the expression of fetal haemoglobin in a large Asian Indian kindred (Garner et al. 2002). Model-based linkage analysis detected the effect of the chromosome 8 QTL and showed it to be conditional on the individual’s XmnI-\( \gamma \) genotype. Linkage to the QTL was later detected using a sample of 319 dizygotic twin pairs of European descent from the St. Thomas’ Twin Registry (TwinsUK) (Garner et al. 2004). However, using a statistical interaction test described below no statistical interaction was detected, presumably due to the small sample size. An additional sample of European dizygotic twin pairs from the same UK registry was subsequently phenotyped and genotyped for markers in the chromosome 8 candidate interval, increasing the sample size to 874 dizygotic twins pairs with genotype and phenotype data. Evidence for statistical interaction between the XmnI-\( \gamma \) polymorphism and the QTL on chromosome 8 in the larger sample of 874 sib pairs of European descent is reported here. The existence of a genetic interaction is consistent with our knowledge of the biological system. However, the results do not show a clearly intuitive pattern, suggesting a more complex underlying biological mechanism.

Materials and Methods

The genotyping and phenotyping were carried out using the same protocols described in the earlier linkage analysis of dizygotic twin pairs from the St. Thomas’ UK Adult Twin (TwinsUK) Registry (Spector & MacGregor, 2002). Because dizygotic siblings are genetically equivalent to sibling pairs (with the advantage of perfect age matching), the twin pairs will henceforth be referred to as sib- sibs. A 21 cM region spanning the centromere of chromosome 8q was analyzed. Identity-by-descent (IBD) probabilities across the region were calculated every cM for the first and second data sets separately, using the MAPMAKER/SIBS program (Kruglyak & Lander, 1995). IBD was estimated separately for the two data sets because the genotype data was not complete between the two sets. Six markers were used for IBD computation in each of the data sets, with three markers being common between both sets and acting as anchors for the genetic map; the genetic maps are shown in Table 1.

FC levels were log transformed for the linkage analysis. Multipoint linkage analysis was carried out using a linear regression model originally described by Carey & Williamson (1991) and Fulker et al. (1991) then subsequently extended to accommodate other study
Table 1: Maps used for IBD computation.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Location</th>
<th>Set 1</th>
<th>Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S1722</td>
<td>62.47</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>D8S255</td>
<td>64.6</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>D8S509</td>
<td>69.4</td>
<td>6.9</td>
<td></td>
</tr>
<tr>
<td>D8S1828</td>
<td>71.0</td>
<td></td>
<td>7.5</td>
</tr>
<tr>
<td>D8S1723</td>
<td>75.39</td>
<td>12.9</td>
<td></td>
</tr>
<tr>
<td>D8S507</td>
<td>75.39</td>
<td>11.7</td>
<td></td>
</tr>
<tr>
<td>D8S1812</td>
<td>78.78</td>
<td>16.3</td>
<td></td>
</tr>
<tr>
<td>D8S260</td>
<td>79.36</td>
<td>15.9</td>
<td></td>
</tr>
<tr>
<td>D8S1767</td>
<td>83.51</td>
<td>21.0</td>
<td>21.0</td>
</tr>
</tbody>
</table>

Table 2: Six possible unordered dizygotic twin pair genotypes and identity-by-state variable definition, with number of pairs and statistical characteristics of groups.

<table>
<thead>
<tr>
<th>Twin pair Genotypes</th>
<th>Pairs</th>
<th>mean (SD)</th>
<th>correlation (p-value)</th>
<th>Allele C i.b.s.</th>
<th>Allele T i.b.s.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C/C, C/C</td>
<td>239</td>
<td>0.90 (0.61)</td>
<td>0.49 (&lt;0.001)</td>
<td>2</td>
<td>0.49 (&lt;0.001)</td>
</tr>
<tr>
<td>C/C, C/T</td>
<td>271</td>
<td>1.10 (0.62)</td>
<td>0.15 (0.014)</td>
<td>1</td>
<td>0.28 (&lt;0.001)</td>
</tr>
<tr>
<td>C/T, C/T</td>
<td>207</td>
<td>1.32 (0.65)</td>
<td>0.39 (&lt;0.001)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C/C, T/T</td>
<td>40</td>
<td>1.27 (0.67)</td>
<td>-0.03 (0.88)</td>
<td>0</td>
<td>0.24 (0.003)</td>
</tr>
<tr>
<td>C/T, T/T</td>
<td>87</td>
<td>1.49 (0.61)</td>
<td>0.27 (0.012)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>T/T, T/T</td>
<td>30</td>
<td>1.60 (0.63)</td>
<td>0.58 (0.001)</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 1 Multipoint analysis of linkage and interaction, and simulation results. In the legend “Linkage” refers to the test of $\beta_3$ given in equation (1) and “Linkage $\times$ ibsC” refers to the test of $\beta_7$ given in equation (2). The simulation results are indicated by (Sim.) in the legend. Error bars for the simulation plot represent 1.64 standard deviations in the distribution of statistics computed from the simulated data.

Results

Figure 1 shows the $-\log_{10}(p$-values) for the $t$ statistic of the regression parameter $\beta_3$ (in equation (1) above) or $\beta_7$ (in equation (2) above), computed at one cM intervals across the 21 cM region of chromosome 8. The most significant evidence for linkage occurs at the 5–6 cM position with a $-\log_{10}(p$-values) statistic for the $\beta_3$ parameter of 2.12, corresponding to a $p$-value of 0.0075. No evidence for interaction between the XmnI-GY genotype of the pair and linkage to chromosome 8 region was observed when the ibsT variable was used in the model, with the smallest $p$-value being 0.7 for the $\beta_7$ parameter. Figure 1 shows that the $\beta_7$ parameter was significant for the ibsC variable in a portion of the region tested. The greatest statistical evidence for interaction occurred at the 22 cM position, with a statistic of 2.48 corresponding to a $p$-value of 0.0033. At position 5, the peak of linkage without XmnI-GY interaction, the $p$-value for $\beta_7$ was 0.051. The results of the analysis of covariance show that a significant relationship exists between the sibling pairs’ allele sharing at the XmnI-GY polymorphism and the chromosome 8 QTL in determining the extent to which the pairs’ FC levels are predictive of each other.

To investigate the nature of the interaction, the data were stratified on the ibsC variable and linkage analysis was carried out in each of the sub-samples separately, using the linkage model given in equation (1) above. There were 478 pairs with $ibsC=1$, 239 with $ibsC=2$, and 157 with $ibsC=0$. Sibling pairs that share one C allele IBS showed significant linkage to chromosome 8 (Figure 2b) and pairs sharing either zero or two C alleles showed no evidence for linkage (Figures 2a and 2c). The results for the stratified sample clearly show why the interaction term, $\beta_7$, was significant for the ibsC and chromosome 8 interaction.

There are two possible explanations why the measured interaction may not be purely biological. First, the $ibsC=1$ category has twice as many pairs as the $ibsC=2$, and three times as many pairs as the $ibsC=0$ category. A sample size of 478 sibling pairs will have
considerably more power than a sample size of 239 or 157, and would be expected to yield a more significant statistic if linkage exists. Second, pairs that share more alleles IBS at the XmnI-γ locus will tend to have more highly correlated FC levels and a smaller FC variance because of the association between the XmnI-γ polymorphism and FC. As more of the sib correlation and trait variance is explained by the polymorphism and FC, there will be less power to detect linkage to loci affecting the correlation of the siblings’ trait values. These two possible explanations were tested through simulation.

The sample mean and standard deviation of the log FC levels were 1.16 and 0.66, respectively. Table 2, columns 3 and 4, show the mean and standard deviation of the log FC levels within the six possible unordered sibling pair XmnI-γ genotypes classes, as well as the observed correlation coefficient between the siblings’ trait levels. The trend of higher mean FC levels for pairs sharing more T alleles is consistent with the association of the allele with increased levels of the trait. The standard deviations within each group range from 0.61 to 0.67. There is a clear relationship between the XmnI-γ genotypes of the pair and the observed trait correlation, reflecting the association of the polymorphism with increased FC. Higher correlations are observed for pairs having concordant XmnI-γ genotypes. Columns 6 and 8 of Table 2 show that the sibling pair correlations vary within each IBS class variable as well.

A simulation experiment was carried out to evaluate the probability of the observed results. In each iteration of the simulation the 874 pairs were randomly ordered, then the top 157 were assigned $ibsC=0$, the second 478 were assigned $ibsC=1$, and the last 239 were assigned $ibsC=2$. One thousand replicated samples were taken at the 1, 8, 15 and 22 cM positions of the genetic interval. The full sample was analyzed for interaction by testing the significance of the $\beta_7$ parameter in equation (2), and linkage was tested within the three sub-samples stratified by the $ibsC$ variable by computing the significance of the $\beta_3$ parameter in equation (1). The sib correlation within each sub-sample was computed for each replicate.

The simulated distribution of $-\log(p$-value) statistics for the interaction term, $\beta_7$, are shown in Figure 1. The points represent the average value of the $-\log10(p$-value) and the bars represent 1.64 standard deviations in the simulated distribution, corresponding to a one-sided probability of 0.05. The results show that the multipoint statistics computed from the real data were outside the 1.64 standard deviation range of those computed from simulated data across nearly the entire 21 cM interval. The distribution of the linkage statistics, $\beta_3$, computed for each sub-sample are shown in Figures 2a–2c. Figures 2a and 2c show that the multipoint plots computed from the real data are within the 1.64 standard deviation range of the simulated distribution for sample sizes of 157 and 239, respectively. Figure 2b shows that the multipoint statistics for the $ibsC=1$ category computed from the real data were outside the 1.64 standard deviations range of the simulated distribution.

The difference in the simulated statistics between the three sample sizes provides an estimate of their relative power to detect linkage. A comparison of the simulation results presented in Figures 2b and 2c shows an increase in power is achieved by doubling the sample size from 239 to 478 sibling pairs, with the average statistics being approximately 0.75 and 1.25, respectively. However, the difference in the real $ibsC=1$ and $ibsC=2$ multipoint linkage curves far exceeds the expected increase in power given by the increase in sample size. The difference between the real $-\log10(p$-value) statistics of the $ibsC=1$ and $ibsC=2$ groups at the 15 cM position was 2.80. The average difference between the statistics calculated from the simulated samples corresponding to the two $ibsC$ groups was 0.49 ($SD = 1.24$). The observed difference between linkage results for the $ibsC=1$
and $ibsC=2$ samples was nearly two standard deviations greater than the expectation calculated from the simulation.

A simple linear regression analysis of the simulated data was carried out to determine the extent to which the sibling correlation within each of the three subsamples predicts the $-\log_{10}(p$-value) linkage statistic. The replicates simulated at the 15 cM position were used in the analysis. For the sample size of 478 pairs there was no relationship between sibling correlation and the linkage statistics. For sample sizes of 239 and 157 pairs there was a weak, but statistically significant ($p$-values < 0.05), relationship between sibling correlation and the $-\log_{10}(p$-value) linkage statistic. Higher sibling correlations were associated with lower $-\log_{10}(p$-value) statistics; however, sibling correlation only accounted for 2% of the variance in the linkage statistics for both sample sizes. The difference in the sib correlation between the 478 and 239 pair groups only explained 1% of the variation in the difference of the $-\log_{10}(p$-value) statistics between the two groups. The average $\log_{10}(p$-value) statistic for the replicate samples of 478 pairs that had a sibling correlation greater than 0.45 was 1.68 (S.D. = 0.53), which is only slightly higher than the correlation calculated from all simulated replicates.

The variance in FC levels of the first sibling was 0.42 in both the $ibsC=0$ and $ibsC=1$ groups and 0.36 in the $ibsC=2$ group. The difference in variance between the 478 and 239 pair groups only explained 1% of the variation in the difference of the $-\log_{10}(p$-value) statistics between the two groups. The average $-\log_{10}(p$-value) statistic among simulated samples of 478 pairs that had a variance in the second sibs’ FC levels greater than 0.42 was 1.39 (S.D. = 0.73). The average $-\log_{10}(p$-value) statistic among samples of 239 pairs that had a variance in the second sibs’ FC levels less than 0.36 was 0.68 (S.D. = 0.46). The results provide no evidence that the variation in linkage results observed across $ibsC$ classes was the result of the different sib correlations or variances within the samples.

**Discussion**

An interacting QTL was originally mapped to chromosome 8q in a large Asian Indian kindred using a two-locus linkage model (Garner et al. 2002). The specific purpose of the original study was to identify loci that were interacting with, or modifying, the effect of the $XmnI-Gy$ polymorphism on adult FC levels. Moreover, the chromosome 8 locus was not detected fortuitously using a single-locus approach. Linkage of FC levels was subsequently detected for chromosome 8 in a relatively small sample of European twin pairs under a single locus linkage model (Garner et al. 2004); and the $ibsC=1$ category was large enough that the linkage signal could be detected in the full sample. However, the sample size of 319 pairs did not provide enough power to detect the interaction in an analysis of covariance. The analysis of 874 twin pairs described here shows that a statistical interaction between the $XmnI-Gy$ polymorphism and the chromosome 8 QTL exists in Europeans sampled from the United Kingdom. An earlier study by Garner et al. (2002) showed that the statistical interaction was present in Asian Indians from Gujerat. These results do not imply that a genetic variant is common between the two populations. However, it does show that genetic variation in a gene on chromosome 8 affects FC levels in both populations, through an interaction with a genetic variant of the $\beta$-globin complex.

Detection of the statistical interaction does not provide much information about the nature of the underlying biological mechanism. Linkage was observed in dizygotic twin pairs that share one C allele identically-by-state. The results suggest that the heterozygote $XmnI-Gy$ genotype is related to the effects of the chromosome 8 QTL, which implies that an over-dominance mechanism, whereby the dominance effect associated with the heterozygote exceeds the additive effect of the homozygote, may be involved. However, adding the 87 twin pairs with one C/T heterozygote and one T/T homozygote resulted in a reduction of the significance of the result, and no evidence for linkage was observed within the sample of 87 individuals (results not shown). The $XmnI-Gy$ polymorphism has been shown to be associated with FC levels and to account for all of the linkage observed for the $\beta$-globin complex (Garner et al. 2000a). It is possible that another variant site in the complex that is in linkage disequilibrium with the $XmnI-Gy$ site is interacting with the chromosome 8 QTL but does not have a detectable marginal effect. It is not intuitive how linkage disequilibrium with another variant that is etiological could explain the lack of relationship.
between the linkage and \( ibsC \) variable. For example, if the interacting variant was in coupling linkage disequilibrium with the C allele, one would expect to observe an effect when both twins were homozygous for the C allele, unless an over-dominance mechanism existed. The sequence in the \(-158 \) region of the \( \gamma \) promoter is not a recognized binding motif for any known transcription factors; however, the mechanism of increased \( \gamma \) globin expression is likely to involve a set of transcription factors and co-activators functioning within multiprotein complexes (Kadonaga, 1998).

It is likely that genetic interactions are a common mechanism underlying complex traits. However, published findings of genetic interaction are not common and replication is rare for humans or any other organism. There are numerous reasons why statistical analyses of genetic interactions are not routinely carried out, including; the large sample sizes required to test for statistical interactions; the lack of simple genetic analysis software for testing epistasis models; the lack of consensus on what epistasis is and how it should be modelled; the difficulties of interpreting the results; and no strong biological knowledge on which to base an analysis of interactions. The results presented here were achieved with a relatively large sample of dizygotic twin pairs and a strong biological hypothesis; however, the complete underlying biological mechanism that gives rise to the measured statistical interaction cannot be deduced from studies of this kind. Intermediate factors may exist between the chromosome 8q gene product and the \( \beta \) globin complex that lack the heritable variation necessary for detection by linkage analysis. Nonetheless, this study demonstrates the importance and advantage of extending genetic analysis beyond simple single-locus, major gene models if one wishes to penetrate the mechanisms underlying complex traits. The wealth of clinical and genetic data that has been collected at great public expense should be analyzed as exhaustively as possible, and in such a way that incorporates the biological knowledge underlying the trait of interest.

**Acknowledgements**

The work is supported by a grant (5 R01 HL69259-03) from the National Institutes of Health (NIH/NHLBI) to Swee Lay Thein and Chad Garner, and in part, by a grant (G0000111) from the Medical Research Council UK, to Swee Lay Thein. The TwinsUK cohort is supported by the Wellcome Trust.

**References**


© University College London 2005


Received: 7 December 2004
Accepted: 11 March 2005