

## Report

# Evidence of Genetic Interaction between the $\beta$ -Globin Complex and Chromosome 8q in the Expression of Fetal Hemoglobin

Chad P. Garner,<sup>1</sup> Thanusak Tatu,<sup>2</sup> Steve Best,<sup>2</sup> Lisa Creary,<sup>2</sup> and Swee Lay Thein<sup>2</sup>

<sup>1</sup>Department of Integrative Biology, University of California, Berkeley, Berkeley; and <sup>2</sup>Department of Haematological Medicine, Guys', King's, and St. Thomas' School of Medicine, London

During human development, the switch from fetal to adult hemoglobin (Hb) is not complete with the residual  $\gamma$ -globin expression being restricted to a subset of erythrocytes termed "F cells" (FC). Statistical analyses have shown the FC trait to be influenced by a common sequence variant (C→T) at position –158 upstream of the G $\gamma$ -globin gene, termed the "XmnI-G $\gamma$  polymorphism." The XmnI-G $\gamma$  site is believed to be involved in the expression of the G $\gamma$ -globin gene through interaction with transcription factors, and polymorphisms in the transcription factors could be influencing fetal Hb expression, conditional on the XmnI-G $\gamma$  site. Using a two-locus model, in which the second locus was the known quantitative-trait locus (QTL) at the XmnI-G $\gamma$  site, we showed suggestive linkage to chromosome 8q. A maximum single-point LOD score of 4.33 and a multipoint LOD score of 4.75 were found in a 15–20 cM region of chromosome 8q. A single-locus analysis failed to show linkage of FC to the region when the XmnI-G $\gamma$  site was accounted for by removing its effects from the data or including it as a covariate. Results of the single-locus analysis were significant when the effects of the XmnI-G $\gamma$  site were not accounted for in any way. The results of analysis in a large Indian kindred indicate that there is an interaction between the XmnI-G $\gamma$  site and a QTL on chromosome 8q that is influencing the production of fetal Hb.

The switch from fetal ( $\alpha_2\gamma_2$ ) to adult ( $\alpha_2\beta_2$ ) hemoglobin (Hb) synthesis is not complete, in that small amounts of fetal Hb (Hb F) continue to be synthesized throughout adult life. This residual  $\gamma$ -globin expression in adults is restricted to a minority of erythrocytes termed "F cells" (FC) (Boyer et al. 1975). Surveys of normal blood donors have shown that the levels of Hb F and FC vary considerably within populations (>20-fold) and that the distribution of the trait is substantially positively skewed (Zago et al. 1979; Miyoshi et al. 1988; Economou et al. 1991; Sampietro et al. 1992). Between 10% and 15% of normal adults have Hb F levels  $\geq 0.6\%$  of the total Hb corresponding to  $\geq 4.0\%$  FC.

As an initial step toward dissection of the genetic basis of the control of Hb F and FC production in

adults, samples of MZ and DZ twins (in a healthy European population) were used to demonstrate that the trait is highly heritable (heritability = 0.89) (Garner et al. 2000b). Statistical analyses have shown the trait to be influenced by several factors, including age, sex, and a common sequence variant (C→T) at position –158 upstream of the G $\gamma$ -globin gene, a variant commonly referred to as the "XmnI-G $\gamma$  polymorphism" (Gilman and Huisman 1985; Garner et al. 2000a). The XmnI-G $\gamma$  site has been shown to be associated with high FC levels and to account for 13%–32% of the total phenotypic variance of the trait (Garner et al. 2000a). On the basis of the results of a combined linkage and association analysis, we conclude that the XmnI-G $\gamma$  site is influencing—either directly or in strong linkage disequilibrium with another locus—G $\gamma$ -globin gene expression on the  $\beta$ -globin complex (Garner et al. 2000a). Linkage to adult Hb F and FC levels has been identified in two other regions of the genome. One locus has been mapped to chromosome 6q23 in an extensive Asian Indian kindred (see below) with  $\beta$ -thalassemia and high FC levels (in the range defined as heterocellular hereditary persistence of fetal hemoglo-

Received October 18, 2001; accepted for publication December 10, 2001; electronically published January 30, 2002.

Address for correspondence and reprints: Dr. Chad P. Garner, Department of Integrative Biology, University of California, Berkeley, Berkeley, CA 94720-3140. E-mail: cgarner@socrates.berkeley.edu

© 2002 by The American Society of Human Genetics. All rights reserved. 0002-9297/2002/7003-0025\$15.00

**Table 1****Results of Two-Locus Linkage Analysis for 40 Markers on Chromosome 8q**

MARKER	LOCATION <sup>a</sup>	HETEROZYGOSITY	UNRESTRICTED MODEL		ADDITIVE MODEL	
			LOD Score	$\theta_2$	LOD Score	$\theta_2$
D8S540	60	.61	2.44	<.001	1.75	<.001
D8S1770	59.4	.41	.00	.49	.00	.49
D8S1810	59.4	.66	3.14	<.001	.29	.30
D8S278	60	.69	2.53	<.001	2.31	.05
D8S259	60	.79	1.26	.13	.67	.23
D8S1758	60	.66	3.70	<.001	3.39	.04
D8S513	60	.68	.45	.29	.50	.27
D8S1709	60	.70	1.48	.025	2.34	.06
D8S1845	60	.64	.64	.18	.25	.27
D8S283	60	.73	1.75	<.001	3.15	<.001
D8S535	60	.74	1.47	<.001	.67	<.001
D8S505	60	.64	.11	.29	.00	>.50
D8S1700	60.6	.66	1.22	.13	.85	.16
D8S1766	60.6	.49	.79	<.001	.40	>.50
D8S1750	60.6	.77	1.18	.13	2.12	.09
D8S1803	60.7	.75	2.95	<.001	2.91	<.001
D8S1722	60.7	.85	2.41	.08	2.63	.07
D8S536	61.5	.68	.23	.33	.27	.28
D8S1791	62.3	.79	.31	.23	.17	.25
D8S1821	62.3	.76	.86	.20	.83	.21
D8S255	64	.72	1.68	.11	1.64	.10
D8S515	64.6	.66	1.08	.10	1.10	.13
D8S532	64.6	.78	1.52	<.001	3.15	<.001
D8S1817	64.6	.72	.86	.12	.15	>.50
D8S268	64.6	.23	1.05	<.001	.29	.06
D8S531	65.7	.61	1.79	<.001	1.51	<.001
D8S1745	65.8	.80	1.52	<.001	1.99	<.001
D8S1831	66.9	.72	.54	.26	.15	.33
D8S538	66.9	.74	3.14	<.001	3.53	.03
D8S1773	66.9	.66	2.44	<.001	2.75	<.001
D8S517	66.9	.20	.01	>.50	.01	.42
D8S524	66.9	.66	1.89	<.001	1.93	.04
D8S1815	66.9	.67	1.05	<.001	.00	<.001
D8S1833	66.9	.66	4.33	<.001	2.53	<.001
D8S509	69.2	.61	.48	.17	.10	.33
D8S1816	70.6	.48	1.38	<.001	2.45	<.001
D8S1828	70.6	.89	2.06	<.001	1.14	.08
D8S285	70.6	.64	1.46	.003	.10	.33
D8S1723	74.9	.49	1.58	<.001	1.48	<.001
D8S507	75	.82	1.40	.076	.69	.16

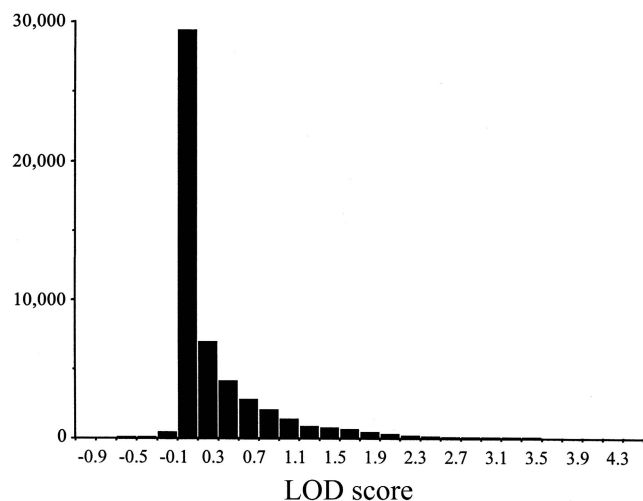
<sup>a</sup> The locations of the markers were obtained from the published Whitehead YAC physical order (Hudson et al. 1995) and from the Généthon genetic map (Dib et al. 1996).

bin) (Craig et al. 1996); the other locus, which is associated with FC variation in sickle cell disease, has been mapped to chromosome Xp22.2-p22.3 (Dover et al. 1992).

In earlier studies, as part of a systematic search for loci that regulate  $\gamma$ -globin gene expression in  $\beta$ -thalassemia, a large, Asian Indian kindred was identified that included individuals with  $\beta$ -thalassemia and elevated FC levels (Thein and Weatherall 1989). The extended family included 210 individuals spanning seven generations. With the use of a genomewide quantitative-trait linkage-mapping strategy that accounted for

the covariate effects of known genetic and nongenetic factors, a gene modifying Hb F and FC production was mapped to chromosome 6q23 in this kindred (Thein et al. 1994; Craig et al. 1996). A genomewide linkage analysis similar to the one performed by Craig et al. (1996), now accounting for the chromosome 6q23 locus and *XmnI*-G $\gamma$  site, as well as the other covariates, failed to identify additional loci influencing FC levels in the kindred; however, significant familial correlations of FC levels were observed.

Clinical and epidemiological studies suggest that the G $\gamma$ -158 C $\rightarrow$ T substitution predisposes toward increased



**Figure 1** Null distribution of LOD scores computed under the unrestricted two-locus models. The distribution was generated from 50,000 replicates, in which a simulated marker was unlinked to QTL2. The pedigree structure, trait values, and *XmnI*-G $\gamma$  genotypes were identical to the analyzed data. A gene-dropping algorithm written by C.G. generated the replicate marker data.

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). Although the sequence in this region of the G $\gamma$  promoter is not a recognized binding motif for any of the known transcription factors (Wood 2001), the mechanism of increased  $\gamma$ -globin expression is likely to involve a network of transcription factors and their coactivators, which function within multiprotein complexes (Kadonaga 1998). In such a scenario, altered expression of the G $\gamma$ -globin gene could arise from an alteration of the *cis*-binding site itself or from a change in one of the transcription factors in the multiprotein network. The latter could involve a protein-protein interaction or a DNA-protein interaction. Thus, the *XmnI*-G $\gamma$  site could be fine-tuning the expression of the G $\gamma$ -globin gene through its interactions with transcription factors, and polymorphisms in the transcription factors—or subunits thereof—could be influencing Hb F expression, conditional on the *XmnI*-G $\gamma$  site. To test this hypothesis, a reanalysis of genomewide marker data in the Indian kindred was performed, under a two-locus genetic model, with one of the loci being the *XmnI*-G $\gamma$  site. Given sufficient power, this model could detect linkage to loci that have an effect on FC levels that is conditional upon the presence of the *XmnI*-G $\gamma$  site. We have identified a locus on chromosome 8q that shows suggestive linkage to the FC trait, and the observed linkage is conditional upon the effects associated with the *XmnI*-G $\gamma$  site.

The kindred, which was described elsewhere in detail (Thein et al. 1994), was analyzed as six independent pedigrees, to break several inbreeding relationships, as described by Craig et al. (1996). FC assays and determination of the FC values were described elsewhere (Craig et al. 1996). DNA was extracted from peripheral blood leukocytes, using standard procedures. The T $\rightarrow$ C polymorphism at position  $-158$  of the G $\gamma$ -globin gene and the  $\beta$ -thalassemia mutations were determined, as described by Craig et al. (1993). The marker data were the same as those used by Craig et al. (Craig et al. 1996) in the previous genome screen of the Indian kindred, with an additional 45 markers that were genotyped to follow up on results from the original study. Additional markers were typed using an automated DNA sequencer (ABI 377 or ABI 3100).

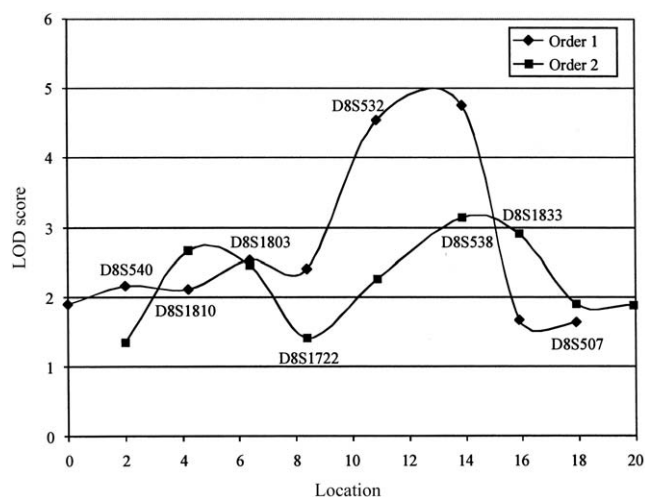
The effects of the covariates—age, sex,  $\beta$ -thalassemia, and the chromosome 6q23 locus—on the variance in FC levels were estimated by linear regression and removed from the data prior to the linkage analysis. The effect of the *XmnI*-G $\gamma$  polymorphism was similarly removed prior to analysis, where indicated. Allele frequencies of the 266 markers analyzed in the study were estimated by counting the alleles in the founders of the kindred. Genotype errors were identified using the program PedCheck (O’Connell and Weeks 1998) and were either reassigned on the basis of a reevaluation of the data or excluded from the data.

The two-locus probability models used are extensions of the single, major-locus model for quantitative-trait data in families. A statistical description of the models is given in the Appendix. Likelihoods were computed using the program VITESSE (O’Connell and Weeks 1995). The genomewide analysis was performed under the unrestricted two-locus model, in which the nine joint-genotype means are independent. The parameters

**Table 2**

**Results of Single-Locus Linkage Analysis for 10 Markers on Chromosome 8q**

MARKER	SCORES FROM SINGLE-LOCUS MODELS WITH					
	No Accounting for <i>XmnI</i> -G $\gamma$		Preadjustment for <i>XmnI</i> -G $\gamma$		<i>XmnI</i> -G $\gamma$ as a Covariate	
	LOD	$\theta$	LOD	$\theta$	LOD	$\theta$
D8S540	1.32	.05	1.37	.06	1.31	.08
D8S1810	.00	>.50	.29	>.50	.00	>.50
D8S278	.78	.21	1.17	.24	.90	.24
D8S1758	1.09	.13	.68	.22	.44	.22
D8S1803	1.70	<.001	.17	.35	.11	.38
D8S1722	2.49	<.001	.25	.32	.20	.30
D8S538	3.52	<.001	.66	.27	.65	.24
D8S1773	1.32	<.001	.39	.18	.72	.11
D8S1833	1.68	<.001	1.14	.18	1.07	.17
D8S1828	.00	.48	.16	.37	.01	.47



**Figure 2** Multipoint LOD scores for two groups of four markers (orders 1 and 2) computed under the unrestricted two-locus model. The two marker orders were chosen so that when combined, multipoint LOD scores would be computed approximately every 2 cM across the chromosome 8q region. Marker names are shown above the line for order 1 and below the line for order 2. Locations are given in centimorgans, starting from the most centromeric location tested. The distances between the markers were estimated using the program CRIMAP (Green et al. 1990) and by comparison with the published Génethon map (Dib et al. 1996). Likelihoods were computed using the program VITESSE (O'Connell and Weeks 1995).

of the unrestricted, two-locus model include the following: (1) allele frequencies of the two QTLs (QTL1 and QTL2), (2) allele frequencies of the *XmnI*-G $\gamma$  site and the test marker, (3) nine joint-genotype means, (4) a single within-genotype variance, and (5) the recombination rates between the *XmnI*-G $\gamma$  site and QTL1 and between the test marker and QTL2. Markers that were not linked to the *XmnI*-G $\gamma$  site were the only ones tested according to the assumption that the two QTLs segregate independently. Forty markers on chromosome 8q were tested using an additional two-locus model, termed the additive model. In the additive model, the nine joint-genotype means are represented by five parameters: a single mean for the low-trait joint-genotype mean of the two QTLs and four parameters measuring allelic and dominance effects that were constrained to have positive values (see Appendix).

For both the unrestricted and the additive models, the marker and *XmnI*-G $\gamma$  allele frequencies were fixed at the values estimated from the data, and the allele frequency of QTL1 was fixed to be the same as that for the *XmnI*-G $\gamma$  site. The polymorphism marked by the *XmnI*-G $\gamma$  site is common, having a frequency of  $\sim 0.35$  in the founders of the kindred. The allele frequencies of QTL1 and the *XmnI*-G $\gamma$  polymorphism were assumed to be equal.

The recombination rate between the *XmnI*-G $\gamma$  site and QTL1 was fixed at 0.001. Likelihoods of the unrestricted model were maximized on 12 parameters: the allele frequency of QTL2, nine joint-genotype means, a single within-genotype variance, and the recombination fraction between QTL2 and the test marker. Likelihoods of the additive model were maximized on eight parameters, because the joint-genotype means were restricted to a five-parameter model, as described above and in the Appendix. LOD scores were calculated as the difference between the  $\log_{10}$  of the maximum likelihood and the  $\log_{10}$  of the likelihood of that model, with the recombination fraction between the test marker and QTL2 being fixed at 0.50.

Likelihoods were computed for 10 markers under the single-locus mixed model (Morton and MacLean 1974) using the program PAP (Hasstedt 1994). The effect from the *XmnI*-G $\gamma$  genotype was ignored, estimated, or removed from the data by linear-regression analysis prior to the linkage analysis or was included in the model as a covariate. The likelihoods were maximized on the basis of six or seven parameters, depending on whether the covariate effect of the *XmnI*-G $\gamma$  site was included in the model. The parameters included the allele frequency of the QTL, three genotype means, one within-genotype variance, the recombination fraction between the marker and the QTL, and the coefficient of covariance for the *XmnI*-G $\gamma$  site. LOD scores were calculated as described above.

LOD scores greater than an arbitrarily set threshold of 1.5 were observed for four microsatellites, namely, D2S2216, D6S283, D7S531, and D8S283; these markers and the linked flanking markers were genotyped in the full kindred of 210 individuals. The additional genotyping did not increase the evidence of linkage to chromosomes 2, 6, and 7; however, LOD scores of 3.14, 1.75, and 2.41 were observed at markers D8S1810, D8S283, and D8S1722, respectively. Thirty-seven additional markers across  $\sim 15$  cM of chromosome 8q were genotyped and analyzed for linkage under the unrestricted and additive two-locus models (results shown in table 1). The average heterozygosity of the markers was 0.66, ranging from 0.20 (for marker D8S517) to 0.89 (for marker D8S1828). The average LOD scores of the 40 markers were 1.50 and 1.30, under the unrestricted and

**Table 3**

**Joint-Genotype Means for the Unrestricted Two-Locus Model**

Genotype	BB	Bb	bb
AA	$\mu_1$	$\mu_4$	$\mu_7$
Aa	$\mu_2$	$\mu_5$	$\mu_8$
aa	$\mu_3$	$\mu_6$	$\mu_9$

**Table 4**  
**Joint-Genotype Means for the Additive Two-Locus Model**

Genotype	BB	Bb	bb
AA	$\mu_1$	$\mu_1 + b_1$	$\mu_1 + b_1 + b_2$
Aa	$\mu_1 + a_1$	$\mu_1 + a_1 + b_1$	$\mu_1 + a_1 + b_1 + b_2$
aa	$\mu_1 + a_1 + a_2$	$\mu_1 + a_1 + a_2 + b_1$	$\mu_1 + a_1 + a_2 + b_1 + b_2$

additive models, respectively. The correlation between the LOD scores computed under the two models was 0.72 ( $P < .0001$ ). There was a weak correlation between the LOD scores and heterozygosities of the markers under both models (unrestricted:  $r = .28$ ,  $P = .08$ ; additive:  $r = .30$ ,  $P = .06$ ). Six markers had LOD scores  $>3.0$ , with marker D8S1833 having the highest LOD score (4.33 under the unrestricted model). Two markers (D7S1758 and D8S538) showed LOD scores  $>3.0$  under both the unrestricted and additive models. Two markers (D8S283 and D8S532) showed LOD scores  $>3.0$  under the additive model only, and two markers (D8S1810 and D8S1833) did under the unrestricted model only.

To assess the fit of the model to the data and the significance of the results, a sample from the null distribution of LOD scores was generated for the unrestricted model. Figure 1 shows the simulated distribution of LOD scores. The following probabilities were calculated from the distribution:  $\Pr_{(LOD > 1.0)} = .095$ ;  $\Pr_{(LOD > 2.0)} = .018$ ;  $\Pr_{(LOD > 3.0)} = .0037$ ; and  $\Pr_{(LOD > 4.0)} = .0008$ .

On the basis of these results—and without consideration of multiple testing—the probability of observing the LOD score of 4.33 at marker D8S1833 is  $<.0008$ . A Bonferroni multiple-test correction of this  $P$  value for the 40 markers on chromosome 8q, tested under the unrestricted model gives an adjusted  $P$  value of .032; a genomewide correction for 305 tests gives a  $P$  value of .24. These adjusted significance levels are highly conservative, given that the analysis was of a single kindred and that linkage (nonindependence) extends over large genetic distances. The significance levels found from simulation are not consistent with the typical assumption that a LOD score  $>3.0$  roughly corresponds to a  $P$  value  $<.0001$  (Lander and Kruglyak 1995). The observed LOD score of 4.33 is a suggestive finding. Negative LOD scores were observed for a very small proportion of the simulated replicates (fig. 1), suggesting a problem with the maximum-likelihood computation in these tests. Maximum likelihoods of the data were computed twice, from different starting values, to confirm that the results were consistent. Figure 2 shows the results of the multipoint linkage analysis. The maximum multipoint LOD score was at the location of marker D8S538 for both orders; LOD = 4.75 for order 1 and LOD = 3.13 for order 2 (fig 2). A one-LOD-unit confidence interval around the peak of the multipoint LOD score

under order 1 extends  $\sim 4$  cM from marker D8S1722 to D8S507.

Table 2 shows the results from the single-locus analyses. When the *Xmnl*-G $\gamma$  genotype is not accounted for in any way, 7 of the 10 markers showed LOD scores  $>1.0$ , with the markers D8S1722 and D8S538 having LOD scores of 2.49 and 3.52, respectively. When the data are preadjusted for the genotype, the LOD scores show a substantial reduction; markers D8S1833, D8S278, and D8S540 show LOD scores of 1.14, 1.17, and 1.37, respectively. A similar effect is observed when the genotype is modeled as a covariate; markers D8S1833 and D8S540 show LOD scores of 1.07 and 1.31, respectively. These results suggest that the linkage signal observed between chromosome 8q and FC levels is conditional on the *Xmnl*-G $\gamma$  effect.

The maximum-likelihood estimates of the unrestricted and additive model parameters does not reveal any consistent pattern that would allow one to speculate on the mechanism by which the putative QTL on chromosome 8q is interacting with the known QTL at the  $\beta$ -globin complex. The simplest mechanism would be a purely additive one, with no dependence between the two loci. Under a purely additive model, one would expect that accounting for the covariate effects of the *Xmnl*-G $\gamma$  genotype would cause the single-locus LOD scores at chromosome 8q to increase, because taking those effects into account would reduce the variance within each of the chromosome 8q QTL genotypes. The opposite effect was observed. Accounting for the *Xmnl*-G $\gamma$  genotype effect as a covariate or by removing its effect from the FC distribution gave substantially lower LOD scores than did the single-locus analysis of the data with the effects unaccounted for. The two-locus models tested here almost certainly do not represent the true mechanism by which the locus on chromosome 8q and the *Xmnl*-G $\gamma$  site are influencing FC levels; rather, the two-locus models are simply more powerful than a single-locus model, given the underlying true mechanism. The results indicate that part of this complex mechanism involves a conditional interaction between the chromosome 8q locus and either the *Xmnl*-G $\gamma$  site or another polymorphism nearby.

The 8q QTL could encode for a regulatory factor, or a subunit thereof, which could act either as a repressor or a transcriptional activator that binds directly to the *Xmnl*-G $\gamma$  site or another binding site nearby. Alterna-

tively, the 8q protein could act as a molecular bridge in a protein-protein interaction. In these situations, the effects of any changes in the *trans*-regulatory protein would be conditional on the presence of the *cis*-binding site. There is evidence that coinheritance of some of the genetic determinants of high levels of Hb F may play a significant role in modifying the phenotype of  $\beta$ -thalassemia and sickle cell anemia. Hence, knowledge of the different genetic factors that contribute to the levels of Hb F and FCs could lead to improved therapeutic strat-

egies for Hb F augmentation that would benefit patients with severe hemoglobinopathies.

## Acknowledgments

This work was supported by Medical Research Council grant G00001113 ID51640. C.P.G. was supported by National Institutes of Health grant GM-40282 to Montgomery Slatkin. T.T. was supported by the Royal Thai Government.

## Appendix

The two-locus models are described in terms of the following components:  $P(g)$  = the probability distribution of genotypes in the population;  $P(g_O|g_F, g_M)$  = the probability that an offspring has genotype  $g_O$ , given that the father and mother have genotypes  $g_F$  and  $g_M$ , respectively;  $P(Y|g)$  = the probability that an individual has phenotype  $Y$ , given that he or she has genotype  $g$  (i.e., the penetrance probability); and  $\theta$  = the recombination rate.

The two-locus model is an extension of the single-locus model for QTL genotype  $g$  to the joint-genotypes  $G$  of QTL1 and QTL2. QTL1 and QTL2 have two alleles,  $A, a$  and  $B, b$ , respectively. The frequencies of  $A$  and  $a$  in the population are  $p$  and  $(1-p)$ , respectively, and of  $B$  and  $b$  are  $q$  and  $(1-q)$ , respectively. The product of the Hardy-Weinberg equilibrium frequencies of the two loci gives the probability distribution of joint-genotypes,  $P(G)$ . For example, the expected population frequency of the  $AA, Bb$  joint-genotype is  $2p^2q(1-q)$ .

Let  $\tau_{g1} = P$  (a parent with genotype  $g$  at QTL1 transmits allele  $A$  to a child) and  $\tau_{g2} = P$  (a parent with genotype  $g$  at QTL2 transmits allele  $B$  to a child), then, with Mendelian inheritance,  $\tau_{AA} = 1$ ,  $\tau_{Aa} = 0.5$ , and  $\tau_{aa} = 0$ , and  $\tau_{BB} = 1$ ,  $\tau_{Bb} = 0.5$ , and  $\tau_{bb} = 0$ . The conditional probabilities that an offspring has joint genotypes  $G_O$ , given that the father and mother have joint genotypes  $G_F$  and  $G_M$ , respectively, for three offspring joint genotypes, are as follows:

$$P(G_O = AABb|G_F = st, G_M = uv) = \tau_s \tau_t \tau_u \tau_v ;$$

$$P(G_O = AaBb|G_F = st, G_M = uv) = \tau_s(1 - \tau_t)\tau_u(1 - \tau_v) ;$$

$$P(G_O = aabb|G_F = st, G_M = uv) = (1 - \tau_s)(1 - \tau_t)(1 - \tau_u)(1 - \tau_v) .$$

The remaining six probabilities are found similarly. Given a normally distributed trait, with mean  $\mu$  and variance  $\sigma^2$ , the phenotype  $Y$  at point  $X$  of the normal distribution is denoted by  $\varphi(X - \mu, \sigma^2) = (2\pi)^{-1/2}(\sigma^2)^{-1/2}e^{-1/2(X-\mu)^2/\sigma^2}$ , then the joint-genotype penetrance function for the quantitative trait is the probability-density function  $f(Y|G) = \varphi(Y - \mu_G, \sigma^2)$ .

The means of the joint-genotype ( $\mu_G$ ), under the unrestricted and additive models are shown in tables 3 and 4. Under the additive model, the heterozygote means were constrained to be between the homozygous means so that the model does not allow for overdominance. There is a single within-genotype standard deviation,  $\sigma^2$ .

The probability of an offspring's marker (QTL2 and *XmnI-G $\gamma$* —QTL1) oligogenotypes is expressed in terms of the product of the recombination fraction,  $\theta$ , and the penetrance function. When the gametic phase of the parents is known, the probability of the offspring's phenotype, given their genotypes at the marker and the *XmnI-G $\gamma$*  polymorphism, is the sum of the probabilities of all of the offspring's possible oligogenotypes. When the parental phase is unknown, there is a sum over all possible parental phases that are informative for linkage. The likelihood of a family of arbitrary size is:

$$L \propto \sum \dots \sum_{\text{genotypes}} [\Pi_{\text{founders}} P_i(G) \times \Pi_{\text{nonfounders}} P(G_j|G_F G_M) \times \Pi_{\text{all individuals}} P(Y_k|G_k)] .$$

## References

- Boyer SH, Belding TK, Margolet L, Noyes AN (1975) Fetal hemoglobin restriction to a few erythrocytes (F cells) in normal human adults. *Science* 188:361–363
- Craig JE, Rochette J, Fisher CA, Weatherall DJ, Marc S, Lathrop GM, Demenais F, Thein SL (1996) Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. *Nat Genet* 12:58–64
- Craig JE, Sheerin SM, Barnetson R, Thein SL (1993) The molecular basis of HPFH in a British family identified by heteroduplex formation. *Br J Haematol* 84:106–110
- Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152–154
- Dover GJ, Smith KD, Chang YC, Purvis S, Mays A, Meyers DA, Sheils C, Serjeant G (1992) Fetal hemoglobin levels in sickle cell disease and normal individuals are partially controlled by an X-linked gene located at Xp22.2. *Blood* 80:816–824
- Economou EP, Antonarakis SE, Kazazian HHJ, Serjeant GR, Dover GJ (1991) Variation in hemoglobin F production among normal and sickle cell adults is not related to nucleotide substitutions in the gamma promoter regions. *Blood* 77:174–177
- Garner C, Tatu T, Game L, Cardon LR, Spector TD, Farrall M, Thein SL (2000a) A candidate gene study of F cell levels in sibling pairs with joint linkage and association analysis. *GeneScreen* 1:9–14
- Garner C, Tatu T, Reittie J, Littlewood T, Darley J, Cervino S, Farrall M, Kelly P, Spector TD, Thein SL (2000b) Genetic influences on F cells and other hematological variables: a twin heritability study. *Blood* 95:342–346
- Gilman JG, Huisman TH (1985) DNA sequence variation associated with elevated fetal  $\gamma$ -globin production. *Blood* 66:783–787
- Green P, Falls K, Crooks S (1990) Documentation for CRIMAP, version 2.4. Department of Genetics, School of Medicine, Washington University, St Louis
- Hasstedt SJ (1994) Pedigree Analysis Package. Version 4.0. Department of Human Genetics, University of Utah, Salt Lake City
- Ho PJ, Hall GW, Luo LY, Weatherall DJ, Thein SL (1998) Beta-thalassemia intermedia: is it possible to consistently predict phenotype from genotype? *Br J Haematol* 100:70–78
- Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu S-H, Hu X, Colbert AME, Rosenberg M, Reeve-Daly MP, Rozen S, Hui L, Wu X, Vestergaard C, Wilson KM, Bae JS, Miatra S, Ganiatsas S, Evans CA, DeAngelis MM, Ingalls KA (1995) An STS-based map of the human genome. *Science* 270:1945–1954
- Kadonaga JT (1998) Eukaryotic transcription: an interlaced network of transcription factors and chromatic-modifying machines. *Cell* 92:307–313
- Labie D, Pagnier J, Lapoumeroulie C, Rouabhi F, Dunda-Belkhouja O, Chardin P, Beldjord C, Wajcman H, Fabry ME, Nagel RL (1985) Common haplotype dependency of high  $\gamma$ -globin gene expression and high Hb F levels in  $\beta$ -thalassaemia and sickle cell anemia patients. *Proc Natl Acad Sci USA* 82:2111–2114
- Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241–247
- Miyoshi K, Kaneto Y, Kawai H, Ohchi H, Niki S, Hasegawa K, Shirakami A, Yamano T (1988) X-linked dominant control of F-cells in normal adult life: characterization of the Swiss type as hereditary persistence of fetal hemoglobin regulated dominantly by gene(s) on X chromosome. *Blood* 72:1854–1860
- Morton NE, MacLean CJ (1974) Analysis of family resemblance. III. Complex segregation analysis of complex traits. *Am J Hum Genet* 26:489–503
- O'Connell JR, Weeks DE (1995) The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. *Nat Genet* 11:402–408
- (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63:259–266
- Sampietro M, Thein SL, Contreras M, Pazmany L (1992) Variation of HbF and F-cell number with the G-gamma *Xmn* I (C-T) polymorphism in normal individuals. *Blood* 79:832–833
- Thein SL, Wainscoat JS, Sampietro M, Old JM, Cappellini D, Fiorelli G, Modell B, Weatherall DJ (1987) Association of thalassaemia intermedia with a beta-globin gene haplotype. *Br J Haematol* 65:367–373
- Thein SL, Weatherall DJ (1989) A non-deletion hereditary persistence of fetal hemoglobin (HPFH) determinant not linked to the  $\beta$ -globin gene complex. In: Stamatoyannopoulos G, Nienhuis AW (eds) Hemoglobin switching. Part B. Cellular and molecular mechanisms. Alan R Liss, New York, pp 97–111
- Thein SL, Sampietro M, Rohde K, Rochette J, Weatherall DJ, Lathrop GM, Demenais F (1994) Detection of a major gene for heterocellular hereditary persistence of fetal hemoglobin after accounting for genetic modifiers. *Am J Hum Genet* 54:214–228
- Wood WG (2001) Hereditary persistence of fetal hemoglobin and  $\delta\beta$ -thalassaemia. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL (eds) Disorders of hemoglobin: genetics, pathophysiology, and clinical management. Cambridge University Press, Cambridge, pp 356–388
- Zago MA, Wood WG, Clegg JB, Weatherall DJ, O'Sullivan M, Gunson H (1979) Genetic control of F cells in human adults. *Blood* 53:977–986