

# Evidence for a gene influencing haematocrit on chromosome 6q23–24: genomewide scan in the Framingham Heart Study

J-P Lin, C J O'Donnell, D Levy, L A Cupples

*J Med Genet* 2005;42:75–79. doi: 10.1136/jmg.2004.021097

For more than 40 years, a number of studies have revealed that high haematocrit (HCT) levels are associated with increased risk for cerebrovascular disease,<sup>1–3</sup> cardiovascular disease (CVD),<sup>4–6</sup> peripheral vascular disease,<sup>7–9</sup> and all cause mortality.<sup>5, 10–11</sup> During 34 years of follow up in more than 5200 individuals, a Framingham investigation demonstrated that increased HCT was significantly associated with increased risks for CVD, coronary heart disease, and myocardial infarction in both men and women.<sup>5</sup> A significant increase in all cause mortality for individuals with very low or high HCT was also observed.<sup>5</sup> Although HCT levels were related to other vascular risk factors, the risk associated with an elevated HCT persisted after accounting for other risk factors for cardiovascular and cerebrovascular events, and for all cause mortality.<sup>5, 12</sup>

HCT is the percentage of whole blood that is comprised of red blood cells, and is a compound measure of red blood cell number and size. HGB is an abundant protein within red blood cells and serves as the main oxygen carrying component of red blood cells, therefore HCT and HGB are strongly correlated. From the rheological viewpoint, blood viscosity depends largely on HCT value. There is an inverse relationship between viscosity and vascular blood flow;<sup>13</sup> high HCT hampers organ perfusion.

Twin studies in healthy humans have suggested that HCT variation is partly determined by genetic factors with heritability estimated at 40–65%.<sup>14–16</sup> A number of gene products are known to be involved in erythropoiesis, most notably erythropoietin. However, the genes that determine an individual's normal HCT level in the general population are unknown. A genome scan to map genes controlling HCT in the spontaneously hypertensive rat indicated a significant association between a marker on chromosome 4 and the observed variability of HCT.<sup>17</sup> No association was found between HCT and erythropoietin, which was mapped to chromosome 12 in rat.<sup>17</sup> So far, no linkage analysis of HCT in humans has been reported. We thus report one of the first linkage studies of HCT in the Framingham Heart Study, with the goal of identifying chromosomal regions that may contain quantitative trait loci (QTL) involved in controlling HCT. Because HCT and HGB are strongly correlated, we also carried out a genome scan on HGB. Finally, we conducted a bivariate linkage analysis of HCT and HGB.

## SUBJECTS AND METHODS

The Framingham Heart Study, a population based study, began in 1948 with the recruitment of 5209 residents aged 28–62 years (mean age 44.1) from Framingham, Massachusetts.<sup>18</sup> The participants have undergone biennial examinations since the study began. In 1971, the Framingham Offspring Study<sup>19</sup> was started, in part to evaluate the genetic components of cardiovascular disease aetiology. In total, there were 5124 subjects aged 5–70 years

## Key points

- Elevated haematocrit (HCT) levels are associated with increased risk for vascular diseases. We carried out genome scans for quantitative trait loci (QTL) on HCT and on haemoglobin (HGB), a correlated trait.
- The heritabilities were estimated as 41% for HCT and 45% for HGB.
- The genomewide linkage analysis revealed evidence of significant linkage for HCT to chromosome 6q23–24, with a lod score of 3.4 at location 136 cM. Only one other region in the genome, chromosome 1, produced a multipoint lod score >1.5 (lod for this region was 1.7).
- The results of the HGB genome scan were different from those of HCT; there was no evidence of linkage for HGB to the same chromosome 6q region.
- Bivariate linkage analysis also did not support QTL pleiotropy in this chromosome area; however, bivariate analyses provided evidence of QTL pleiotropy to chromosome 9q, with a lod score of 3.1 at location 149 cM.
- We conclude that chromosome 6q may harbour a gene that is specific to HCT but not HGB, whereas a shared gene for both traits may lie on chromosome 9q.

(mean age 36.3) including the offspring of the original cohort and the spouses of the offspring. The offspring cohort has been examined every 4 years (except the first two examinations, with 8 years intervening). Within the study, the 330 largest extended families were selected for a 10 cM density genomewide scan (399 markers). The number of subjects genotyped was 1702. We used measurements from offspring cohort examination 1 and original cohort examination 12 for our genome scan. Both examinations were conducted in the early 1970s. As HGB was measured at offspring cohort examination 1 but not at original cohort examination 12, we could only carry out a genome scan on HGB using the offspring cohort. For a better comparison, we also carried out a genome scan on HCT using the offspring cohort only. Finally, we carried out a bivariate genomewide linkage analysis of HCT and HGB in the offspring cohort only.

HCT was measured by the Wintrobe method. Blood was collected and spun at 5000 rpm for 20 minutes in a balanced oxalate tube. The percentage of total blood volume due to red blood cells was determined visually against a calibrated scale. Subjects were weighed in light clothing and with shoes removed. The average number of cigarettes smoked per day over the prior year was based on self reports. Alcohol

consumption was reported by subjects as their usual number of drinks per day and converted to fluid ounces/week for analysis. Laboratory measurements were made on 12 hour fasting venous blood samples that were collected in tubes containing 0.1% EDTA. Lipid determinations were performed at the Framingham Heart Study laboratory, which participates in the Standardization Program of the Centers for Disease Control. All subjects provided informed consent prior to each clinic visit and the examination protocol was approved by the Institutional Review Board at Boston Medical Center, Massachusetts. The clinical and laboratory methods have been detailed elsewhere.<sup>18</sup>

Genomic DNA was isolated from nucleated blood cells. DNA samples were sent to the Marshfield Mammalian Genotyping Service (<http://research.marshfieldclinic.org/genetics/>). At an average 10 cM density, 399 microsatellite markers (screening set 9)<sup>20</sup> covered the genome, with an average marker heterozygosity of 0.77. The genotyping data were cleaned with two steps. Firstly, the sibling kin program in Aspex (<ftp://lahmed.standord.edu/pub/aspex/index.html>) was used to verify family relationships based on all markers available. Secondly, the GENTEST program, as a precursor of INFER, created by Southwest Foundation for Biomedical Research (<http://www.sfbr.org/sfbr/public/software/software.html>) was used to identify and eliminate additional genotype inconsistencies. When inconsistencies were found, the genotyping values in all members of the nuclear family were set to missing.

Variation in HCT from known factors was identified and removed by regression modelling incorporated in SOLAR, to enhance the ability of linkage analysis to detect genetically determined variation using a maximum likelihood based variance decomposition method.<sup>21, 22</sup> The covariates selected ( $p < 0.05$ ) and incorporated into both the heritability estimation and the linkage analyses were age, sex, weight, total cholesterol, high density lipoprotein (HDL) cholesterol, triglyceride, diabetes, smoking (number/day), and alcohol intake (ounces/week). After adjustment for covariates, the variance component method used the residual variation for heritability estimation and linkage analyses. The same covariates were applied for HCT and HGB using the offspring cohort examination 1 genome scan data and bivariate genome scan on HCT and HGB.

An estimate of heritability was obtained using the variance component method. Residual heritability is the proportion of total phenotypic variation due to additive genetic effects, after removing the variation attributable to covariates. Variance component linkage analysis was used for the linkage analysis between random DNA markers covering the entire genome and HCT or HGB (or HCT and HGB for bivariate analysis), adjusted for known covariates. This approach made use of all information in pedigrees of any size and structure. Marker allele frequencies were estimated from the study participants and then used to estimate the proportion of shared alleles that were identical by descent (IBD) among all relative pairs. A likelihood ratio test was used to evaluate linkage by comparing a purely polygenic model (without consideration of genetic marker information) to a model that incorporates IBD information at the marker. The lod score was the  $\log_{10}$  of the ratio of the likelihoods of two models, one purely polygenic versus one that also included IBD information at the marker. Because bivariate linkage analysis may improve both power and localisation of a shared gene for correlated quantitative traits,<sup>23</sup> we carried out a bivariate genome scan on HCT and HGB. For bivariate analysis, lod scores are reported with one degree of freedom; equivalent lod scores are comparable to univariate lod scores. For evaluation of significance, we used the genomewide significant  $p$  values suggested by Lander and Kruglyak.<sup>24</sup> The

total phenotypic correlation between HCT and HGB was estimated taking the family structure into account:

$$(r_p = (h_1 \times h_2 \times r_g) + [(1-h_1)^{1/2}(1-h_2)^{1/2} \times r_e])$$

where  $r_p$ ,  $r_g$ , and  $r_e$  are total phenotypic, genetic, and environmental correlations, and  $h_1$  and  $h_2$  are the square roots of heritabilities of HCT and HGB.

## RESULTS

The total number of individuals with measured HCT and with all covariates (age, sex, weight, smoking, alcohol consumption, total cholesterol, high density lipoprotein cholesterol, triglycerides, and diabetes) used for the heritability estimates and linkage analysis in the original cohort and offspring was 2278 (the actual number of individuals considered in linkage analysis). The total number of individuals in the offspring cohort only with measured HCT and HGB and all covariates was 1444. The mean values of the clinical covariate of those individuals (about 50% male) are displayed in table 1. Of the 2278 individuals from the original and offspring cohorts, 1524 had marker genotypes; these included 1323 full sibling pairs, 52 half sibling pairs, 645 cousin pairs, and 354 avuncular pairs. Among the 1444 individuals in the offspring cohort, 1213 (1124) individuals had HCT (HGB) and all covariates measured and had genotypic data available, including 1245 (1109) full sibling pairs, 52 (50) half sibling pairs, 631 (522) cousin pairs, and 78 (62) avuncular pairs.

The mean (SD) estimates of skewness and kurtosis of the HCT distribution (including offspring cohort examination 1 and original cohort examination 12) were 0.033 (0.028) and 0.283 (0.055), respectively. As a rough measure, normality could not be rejected.<sup>25</sup> Bivariate genetic analysis indicated that HCT was both genetically and environmentally correlated with HGB ( $r_g = 0.85$ ,  $p = 0.02$ ;  $r_e = 0.97$ ,  $p < 0.01$ ). The genetic correlation,  $r_g$ , was significantly different from 1 ( $p < 0.01$ ). The total phenotypic correlation coefficient between HCT and HGB (offspring examination 1 only) was 0.95 ( $p < 0.01$ ).

The heritability estimate for HCT, after adjusting for the covariates, was 41%, indicating that a substantial portion of the variation in HCT was attributable to additive genetic factors. The proportion of variance due to all covariates included in the model (age, sex, weight, total cholesterol, HDL cholesterol, triglyceride, diabetes, cigarettes per day, and alcohol intake) was approximately 40%. The heritability estimate for HGB, after adjusting for the same covariates, was 45%, and the proportion of variance due to the covariates was 49%.

For multipoint linkage analysis of HCT, a maximum lod score of 3.4 (genomic  $p = 0.02$ ) was observed on chromosome 6q23–24, with the peak location at 136 cM (fig 1). Other than this, there was only one region in the genome, chromosome 1, at 132 cM, with a maximum multipoint lod score of 1.5 or higher (lod score = 1.7) (table 2). From a simulation study, the power to detect a QTL effect size of 20% for lod scores of 2.0, 1.5, and 1.0 or higher was ~67%, ~80%, and ~90%, respectively. Therefore, we decided to report the results with a lod score  $> 1.5$ .

Evidence for HCT linkage on chromosome 6 occurred in the region bounded by markers GATA23F08 and GATA32B03. The lod-1 supporting interval (the region corresponding to maximum lod score minus 1) for the HCT QTL spans a 11 cM interval flanked by markers GATA31 and GATA184A08, which are 27 cM apart. When the HCT genome scan was restricted to offspring cohort examination 1 only, the results were similar to those obtained for the combined samples of both cohorts (table 2).

For the HGB genome scan at offspring cohort examination 1, there was little evidence for linkage of HGB to the

**Table 1** Characteristics of the 2278 individuals with HCT and all covariates measured (offspring and original cohort) and 1444 individuals with HCT, HGB, and all covariates measured (offspring only) used in the linkage analysis

Variable	Offspring and cohort, (n=2278)		(Offspring only, n=1444)	
	Mean (SD)	Range	Mean (SD)	Range
HCT (%)	43.75 (3.86)	28 to 62	43.05 (3.77)	28 to 62
HGB (g/l)	NA	NA	145.1 (14.2)	80 to 200
Age (years)	41.62 (16.17)	10 to 82	32.75 (10.46)	10 to 64
Weight (pounds)	157.24 (32.95)	54 to 283	156.33 (34.43)	54 to 281
Total cholesterol (mg/l)	2029.8 (435.9)	960 to 4130	1907.3 (382.4)	960 to 3460
HDL cholesterol (mg/l)	507.2 (144.7)	160 to 1290	504.1 (141.1)	160 to 1180
Triglycerides (mg/l)	1031.8 (738.1)	120 to 7950	918.5 (695.1)	120 to 7950
Smoking (no/day)	10.62 (14.54)	0 to 90	13.00 (15.42)	0 to 90
Alcohol (ounces/week)	3.60 (5.08)	0 to 51	3.56 (4.90)	0 to 43
Diabetes (%)	2 (15)%		1 (9)%	

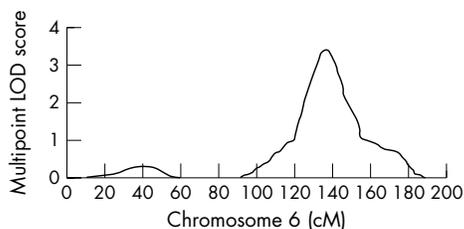
chromosome 6q region (136 cM, lod = 0.37). There were several regions with lod scores  $\geq 1.5$  that did not overlap with the findings of the HCT genome scan, except on chromosome 1 (table 2).

In addition to the regions on chromosomes 1 and 6, a bivariate genome scan for HCT and HGB also identified regions on chromosomes 7, 9, 14, and 19. The first three regions were identified in the HGB univariate genome scan also (table 2). While the evidence suggests that the QTL on chromosome 6 may be unique to HCT, we cannot exclude the possibility of QTL pleiotropy in this region using a comparison of the likelihood of the model restricting the QTL correlation to zero to the model in which the QTL correlation was estimated (data not shown). In contrast, the bivariate analysis revealed a substantial increase in the lod score on chromosome 9 from 1.6 to 3.1 (genomic  $p = 0.045$ ) with a change in position from 136 cM to 149 cM, suggesting pleiotropy in this region for HCT and HGB. The QTL pleiotropy test on 9q region was significant ( $p > 0.05$ , data not shown).

## DISCUSSION

Our results revealed significant evidence for linkage of HCT to chromosome 6q23–24 with a lod score of 3.4 using the variance component method, after adjustment for covariates. These results provide evidence of a possible HCT QTL on chromosome 6q23–24.

In the genome scan on HGB, there was minimal evidence for linkage of HGB to the chromosome 6q area. While HCT and HGB are strongly correlated genetically and environmentally, the only common locus in the results of the two genome scans that examined these measures individually was on a chromosome 1 area (132 cM) at which both traits had a lod score suggestive of linkage.



**Figure 1** Multipoint lod scores for HCT on chromosome 6: x axis values are cM; y axis values are multipoint lod scores. Linkage analyses were conducted using multivariable residuals, adjusted for age, sex, weight, total cholesterol, high density lipoprotein cholesterol, smoking and alcohol intake, triglycerides, and diabetes.

Of import, the lod score in the bivariate analysis increased substantially in the chromosome 9q region with a significant genomewide  $p$  value. Therefore, the region on chromosome 9 may harbour a gene related to both HCT and HGB, while the one on chromosome 6 appears to be more related to HCT. The lod scores in the remaining locations did not reach genomewide significant levels. Our results support the hypotheses that there may be both a major locus on chromosome 9 and polygenic pleiotropy for HCT and HGB.

Within the region of chromosome 6q there are a few candidate genes, including EBP41L2, coding for protein 4.1G, which is a member of the erythrocyte membrane skeletal protein 4.1R (EPB41) gene family,<sup>26</sup> and HEBP2, which codes for a putative haem binding protein (138 cM). One form of hereditary persistence of fetal haemoglobin (HPFH) has also been mapped to this area.<sup>27–28</sup>

Erythropoiesis, the production of red cells by bone marrow, is regulated by the hormone erythropoietin, which binds to the erythropoietin receptor to regulate bone marrow erythroid cell proliferation. The erythropoietin gene maps to the human chromosome 7q21<sup>29</sup> while the erythropoietin receptor maps to chromosome 19p13.<sup>30</sup> We did not find evidence suggestive of linkage to these regions.

The HCT is the proportion of the blood that consists of red blood cells, and the principal components of the red cells are

**Table 2** Chromosome regions in genome scans with multipoint lod  $\geq 1.5$ 

Genome scan study sample	n*	Chrom.	cM	lod score	QTL specific heritability (%)	
					HCT	HGB
HCT offspring and original cohort	1524	1	132	1.7		23
		6	136	3.4		38
HCT offspring	1213	1	131	2.2		26
		6	137	3.3		36
HGB offspring	1124	1	132	1.5	24	
		7	20	1.8	26	
		9	136	1.6	27	
		14	111	2.0	29	
		15	89	1.6	25	
		17	104	2.1	33	
HCT and HGB offspring (bivariate)	1124	1	140	1.6	25	24
		6	136	3.2	37	24
		7	24	1.7	14	23
		9	149	3.1	4	19
		14	111	1.7	19	23
		19	100	1.5	8	1

\*The total number of individuals who were both genotyped and phenotyped. Chrom., chromosome

HGB and membrane. The HGB tetramer consists of two pairs of globin polypeptide chains coded by globin genes located in two clusters:  $\alpha$ -like genes mapping to the chromosome 16pter-p13<sup>31</sup> and  $\beta$ -like genes to the chromosome 11p15.5.<sup>32</sup> A single molecule of haem is attached to each polypeptide chain. The major haemoglobin of intrauterine life is fetal HGB, or HbF. In adult blood, fetal haemoglobin is present in only a very small number of red cells, the F cells. The predominant HGB of postnatal or adult life is haemoglobin A, which consists of two  $\alpha$ -globin and two  $\beta$ -globin chains. Mutations in  $\alpha$ -globin and  $\beta$ -globin genes cause thalassaemias, sickle cell anaemia, methaemoglobinaemias, and erythraemias. No evidence of linkage was found in these two globin gene cluster areas.

Red cell membrane is composed of a lipid bilayer anchored to a network of proteins, the membrane skeleton, which is important for maintaining red cell shape and regulating membrane properties of deformability. Mutations in genes coding the membrane skeleton cause red cell morphological disorders, such as hereditary spherocytosis and hereditary elliptocytosis. Those genes include spectrin  $\alpha$ -chain located on the human chromosome 1q21,<sup>33</sup> spectrin  $\beta$ -chain 14q23–24,<sup>34</sup> ankyrin 8p11,<sup>35</sup> band 3 17q21–22,<sup>36</sup> protein 4.1R 1p36–34,<sup>37</sup> and protein 4.2 15q15.<sup>38</sup> We further confirmed that there was no evidence suggestive of linkage in these chromosome regions.

Protein 4.1R plays an important role in maintaining erythrocyte shape and membrane mechanical properties. Mutations in the protein 4.1R gene cause hereditary elliptocytosis. Recently, another protein 4.1 gene, protein 4.1G, a close homologue of protein 4.1R, was mapped to the human chromosome 6q23,<sup>25</sup> where the peak maximum lod score of our HCT genome scan is located. This gene is widely expressed in human tissues. A study indicated that the two proteins, 4.1R and 4.1G, may play fundamental analogous roles but at different intracellular sites.<sup>26</sup> The detailed function of the protein 4.1G has not yet been reported. Further studies may be warranted to explore whether variation in these and other candidate gene in the linkage region are associated with HCT and related phenotypes.

As there was minimal evidence for linkage of HGB to the chromosome 6q23–24 area, this may imply that the 6q area harbours a gene specific to HCT but not HGB. In addition to HGB, the key component of HCT is red cell membrane. It may be worthwhile to pay special attention to membrane genes as playing a role in our genome scan findings.

A limitation of our study was that our cohort is largely white. Therefore, caution is advised in extrapolating our results to other ethnic groups.

Overall, we found significant evidence for linkage of HCT to chromosome 6q23–24 that harbours several positional candidate genes and significant evidence for joint linkage of HCT and HGB to chromosome 9q34, although we did not find an important candidate gene in this 9q region. Our results support pursuit of fine mapping and association studies between HCT (HGB) and positional candidate genes in these chromosomal regions.

## ACKNOWLEDGEMENT

We thank Dr T Tao (National Library of Medicine) for his help in carrying out the candidate gene search.

## Authors' affiliations

J-P Lin, DECA/NHLBI/NIH, 6701 Rockledge Dr, Suite 8110, Bethesda MD 20892, USA

C J O'Donnell, D Levy, NHLBI/Framingham Heart Study, 73 Mount Wayte Avenue, Suite 2, Framingham, MA 01702, USA

C J O'Donnell, Cardiology Division, Massachusetts General Hospital, Boston, MA 02114, USA

D Levy, Cardiology Division, Boston University School of Medicine, Boston, MA 02115, USA

L A Cupples, Department of Biostatistics, Boston University School of Public Health, 715 Albany Street, TE422, Boston, MA 02118, USA

This work was supported by the National Heart, Lung, and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195).

Competing interests: none declared

Correspondence to: Dr J-P Lin, NHLBI/NIH, 6701 Rockledge Dr, Suite 8110, Bethesda, MD 20892–7938; USA; linj@nhlbi.nih.gov

Received 8 April 2004

Revised 30 July 2004

Accepted 24 August 2004

## REFERENCES

- Kannel WB, Gordon T, Wolf PA, McNamara P. Hemoglobin and the risk of cerebral infarction: the Framingham study. *Stroke* 1972;**3**:409–20.
- Tohgi H, Yamanouchi H, Murakami M, Kameyama M. Importance of the hematocrit as a risk factor in cerebral infarction. *Stroke* 1978;**9**:369–74.
- Giroud M, Creisson E, Fayolle H, Andre N, Becker F, Martin D, Dumas R. Risk factors for primary cerebral hemorrhage: a population-based study—the Stroke Registry of Dijon. *Neuroepidemiology* 1978;**14**:20–6.
- Carter C, McGee D, Reed D, Yano K, Stemermann G. Hematocrit and the risk of coronary heart disease. *Am Heart J* 1983;**105**:674–9.
- Gagnon DR, Zhang T-J, Brand FN, Kannel WB. Hematocrit and the risk of cardiovascular disease—the Framingham study: a 34-year follow-up. *Am Heart J* 1994;**127**:674–82.
- Lippi G, Manzato F, Franchini M, Guidi G. Relationship between hematocrit, primary hemostasis, and cardiovascular disease in athletes. *Am Heart J* 2002;**144**:14.
- Dormandy JA, Hoare E, Colley J, Arrowsmith DE, Dormandy TL. Clinical, haemodynamic, rheological, and biochemical findings in 126 patients with intermittent claudication. *BMJ* 1973;**4**:576–81.
- Dormandy JM, Gutteridge JM, Hoare E, Dormandy TL. Effect of clofibrate on blood viscosity in intermittent claudication. *BMJ* 1974;**4**:259–62.
- Low GDO, Forbes CD. Blood rheology and thrombosis. *Clin Haematol* 1981;**10**:343–67.
- Elwood PC, Waters WE, Benjamin IT, Sweetnam PM. Mortality and anaemia in women. *Lancet* 1974;**1**:891–4.
- Sorlie PD, Garcia-Palmieri MR, Costas R Jr, Havlik RJ, Juan S. Hematocrit and risk of coronary heart disease: the Puerto Rico Heart Health Program. *Am Heart J* 1974;**101**:456–61.
- Wannamethee G, Perry UJ, Shaper AG. Hematocrit, hypertension and risk of stroke. *J Intern Med* 1994;**235**:163–8.
- Oh EO, Lechner H, Aranibar A. High blood viscosity syndrome in cerebral infarction. *Stroke* 1974;**5**:330–3.
- Whitfield JB, Martin NG. Genetic and environmental influences on the size and number of cells in the blood. *Genet Epidemiol* 1985;**2**:133–44.
- Evans DM, Frazer IH, Martin NG. Genetic and environmental causes of variation in basal levels of blood cells. *Twin Res* 1999;**2**:250–7.
- Garner C, Mitchell J, Hatzis T, Reittie J, Farrall M, Thein SL. Haplotype mapping of a major quantitative-trait locus for fetal hemoglobin production, on chromosome 6q23. *Am J Hum Genet* 1998;**62**:1468–74.
- Pravenec M, Zidek V, Zdobinska M, Kren V, Krenova D, Bottger A, van Zutphen LFM, Wang J-M, St Lezin E. Mapping genes controlling hematocrit in the spontaneously hypertensive rat. *Mamm Genome* 1997;**8**:387–9.
- Dawber TR. *The Framingham Heart Study: the epidemiology of atherosclerotic diseases*. Cambridge, MA: Harvard University Press, 1980.
- Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families—the Framingham offspring study. *Am J Epidemiol* 1979;**110**:281–90.
- Yuan B, Vaske D, Weber JL, Beck J, Sheffield VC. Improved set of short-tandem-repeat polymorphisms for screening the human genome. *Am J Hum Genet* 1997;**60**:459–60.
- Almasy L, Blangero J. Multipoint quantitative—trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998;**62**:198–211.
- Blangero J, Almasy L. Multipoint oligogenic linkage analysis of quantitative traits. *Genet Epidemiol* 1997;**14**:959–64.
- Almasy L, Dyer TD, Blangero J. Bivariate quantitative trait linkage analysis: Pleiotropy versus co-incident linkages. *Genet Epidemiol* 1997;**14**:953–8.
- Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995;**11**:271–7.
- Bliss CI. *Statistics in biology*, vol 1. New York: McGraw-Hill, 1967.
- Parra M, Gascard P, Walensky LD, Snyder SH, Mohandas N, Conboy JG. Cloning and characterization of 4.1G (EPB41L2), a new member of the skeletal protein 4.1 (EPB41) gene family. *Genomics* 1998;**49**:298–306.
- Craig JE, Rochette J, Fisher CA, Weatherall DL, Marc S, Lathrop GM, Demenais F, Thein S. Dissecting the loci controlling fetal haemoglobin production on chromosome 11p and 6q by the regressive approach. *Nat Genet* 1996;**12**:58–64.
- Garner C, Tatu T, Reittie JE, Littlewood T, Darley J, Cervino S, Farrall M, Kelly P, Spector TD, Thein SL. Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood* 2000;**95**:342–6.
- Lee-Huang S. Cloning and expression of human erythropoietin cDNA in *Escherichia coli*. *Proc Natl Acad Sci USA* 1984;**81**:2708–12.

- 30 **Jones SS**, D'Andrea AD, Haines LL, Wong GG. Human erythropoietin receptor: cloning, expression, and biologic characterization. *Blood* 1990;**76**:31–5.
- 31 **Deisseroth A**, Nienhuis A, Turner P, Velez R, Anderson WF, Ruddle FH, Lawrence J, Creagan RP, Kucherlapati RS. Localization of the human alpha globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridization assay. *Cell* 1977;**12**:205–18.
- 32 **Deisseroth A**, Nienhuis AW, Lawrence J, Giles RE, Turner P, Ruddle FH. Chromosomal localization of the human beta globin gene to human chromosome 11 in somatic cell hybrids. *Proc Natl Acad Sci USA* 1978;**75**:1456–60.
- 33 **Huebner K**, Palumbo AP, Isobe M, Kozak CA, Monaco S, Rovera G, Croce CM, Curtis PJ. The alpha-spectrin gene is on chromosome 1 in mouse and man. *Proc Natl Acad Sci USA* 1985;**82**:3790–3.
- 34 **Prchal JT**, Morley BJ, Yoon S-H, Coetzer TL, Palek J, Conboy JG, Kan YW. Isolation and characterization of cDNA clones for human erythrocyte beta-spectrin. *Proc Natl Acad Sci USA* 1987;**84**:7468–72.
- 35 **Tse WT**, Menninger JC, Yang-Feng TL, Francke U, Sahr KE, Lux SE, Ward DC, Forget BG. Isolation and chromosomal localization of a novel non-erythroid ankyrin gene. *Genomics* 1991;**10**:858–66.
- 36 **Showe LC**, Ballantine M, Huebner K. Localization of the gene for the erythroid anion exchange protein, band 3 (EMPB3), to human chromosome 17. *Genomics* 1987;**1**:71–6.
- 37 **Conboy J**, Mohandas N, Tchernia G, Kan YW. Molecular basis of hereditary elliptocytosis due to protein 4.1 deficiency. *New Engl J Med* 1986;**315**:680–5.
- 38 **Sung LA**, Fan YS, Lambert K, Chien S, Lin C. Mapping of gene for human erythrocyte protein 4.2 to chromosomal region 15q15-q21. *Cytogenet Cell Genet* 1991;**58**:1996.