ORIGINAL ARTICLE

Genomewide scans of red cell indices suggest linkage on chromosome 6q23

A Iliadou, D M Evans, G Zhu, D L Duffy, I H Frazer, G W Montgomery, N G Martin

See end of article for authors' affiliations

....

Correspondence to:
Dr D M Evans, The Welcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK; davide@well.ox.ac.uk

Background: The red cell indices quantify the size, number and oxygen-carrying ability of erythrocytes. Although the genetic basis of many monogenic forms of anaemia is well understood, comparatively little is known about the genes responsible for variation in the red cell indices among healthy participants.

Objective: To identify quantitative trait loci (QTLs) responsible for normal variation in the red cell indices of 391 pairs of dizygotic twins who were measured longitudinally at 12, 14 and 16 years of age.

Results: Evidence suggesting linkage of red cell indices to haemoglobin concentration (LOD = 3.03) and haematocrit (LOD = 2.95) on chromosome 6q23, a region previously identified as possibly harbouring a QTL for haematocrit, was found. Evidence for linkage to several other regions of the genome, including chromosome 4q32 for red cell count and 7q for mean cell volume, was also found. In contrast, there was little evidence of linkage to the chromosomal regions containing the genes for erythropoietin (7q21) and its receptor (19p13.2), nor to the regions containing the genes for the haemoglobin α (16p13.3) and β chains (11p15.5).

Conclusion: Findings provide additional evidence for a QTL affecting haemoglobin and haematocrit on chromosome 6q23. In contrast, polymorphisms in the genes coding for erythropoietin, its receptor and the haemoglobin α and β chains do not appear to contribute substantially to variation in the red cell indices between healthy persons.

The red cell indices describe the size, number and oxygen-carrying capacity of erythrocytes. Haemoglobin concentration indicates the amount of oxygen-carrying protein haemoglobin in a given volume of blood; red blood cell count (RBC), the concentration of erythrocytes; mean corpuscular volume (MCV), the average volume of each red cell; and haematocrit (HCT), the proportion of blood that consists of erythrocytes. Together, these indices assist in the differential diagnosis of anaemia and are risk factors for a number of clinical conditions. For example, a high haematocrit is associated with increased risk of cerebrovascular[1][2] and coronary artery diseases.[1]

Approximately 7% of the world's population are carriers for different inherited disorders of haemoglobin, making them one of the most prevalent Mendelian diseases.[3] Although the genetic basis of many of these haemoglobinopathies is well understood, comparatively little is known regarding the genetic causes of variation in the normal range of red cell values. Studies on twins have shown that a substantial proportion of the variation between persons is due to genetic factors with heritability estimates ranging from 20% to 96%.[4–9] A recent genomewide linkage study by Lin et al[7] identified a locus on chromosome 6q23–24 linked to HCT, and a pleiotropic locus affecting haemoglobin and HCT on chromosome 9q.[7] Interestingly, the authors found no evidence of linkage to regions containing the genes for the α and β haemoglobin chains (on chromosomes 16p13.3 and 11p15.5), nor erythropoietin and its receptor (on chromosomes 7q21 and 19p13.2), which have all been implicated in Mendelian forms of anaemia. The implication is that the quantitative trait loci (QTLs) responsible for normal variation in the red cell indices might differ from the loci which cause monogenic forms of anaemia.

In a previous study, we reported that variation in the red cell indices was highly heritable, with genetic factors explaining between 61% and 96% of the phenotypic variance in our sample of adolescent twins.[5] In this paper, we extend these results by performing a genomewide linkage scan of the 391 dizygotic twin pairs from that study. We measured the red cell indices of twins at 12, 14 and 16 years of age and performed univariate multipoint sib-pair linkage analyses across the genome. Our study is the first to report linkage results for RBC and MCV, and will hopefully constitute the first stage in the identification and subsequent positional cloning of QTLs responsible for normal variation in the red cell indices.

METHODS

Participants

Twins were recruited as part of an ongoing study concerned with the development of melanocytic naevi (moles), the clinical protocol of which has been described in detail elsewhere.[10–12] Twins were enlisted by contacting the principals of primary schools in the greater Brisbane area, media appeals and by word of mouth. Informed consent was obtained from all participants and parents before testing. The results reported here are for data collected from May 1992 to June 1999. Twins were tested as closely as possible to their 12th, 14th and 16th birthdays. Data were obtained from 706 pairs of twins comprising 154 monozygotic females, 161 monozygotic males, 91 dizygotic females, 104 dizygotic males and 196 dizygotic twin pairs of opposite sex (including 91 pairs in which the females were born first and 105 pairs in which the males were born first). Not all twins were tested across all three measurement occasions. Table 1 shows a breakdown of these data. For example, the first row in table 1 indicates that 85 twin pairs with complete phenotypic and genotypic information were tested at age 12 years only (ie, not at ages 14 or 16 years). No attempt was made to exclude subjects having illness,

....

Abbreviations: HCT, haematocrit; ibd, identical-by-descent; LOD, logarithm of odds ratio; MCV, mean corpuscular volume; QTL, quantitative trait locus; RBC, red blood cell count

Table 1 Breakdown of participation showing the number of complete twin pairs for whom all the red cell indices and genotype information (in the case of dizygotic pairs) were available

12 years	14 years	16 years	MZF	MZM	DZF	DZM	DZOS	Total
√			19	28	9	11	18	85
	√		8	5	2	4	7	26
		√	46	54	14	11	40	165
√	√		28	26	26	37	60	177
√		√	4	3	4	1	7	43
	√	√	8	10	5	7	13	19
√	√	√	41	35	31	33	51	191
Total			154	161	91	104	196	706
Total at 12 years			92	69	71	82	136	450
Total at 14 years			85	76	64	81	131	437
Total at 16 years			99	102	54	52	111	418

DZF, dizygotic females; DZM, dizygotic males; DZOS, dizygotic twin pairs of opposite sex; MZF, monozygotic females; MZM, monozygotic males.

although a few such cases were subsequently excluded as outliers.[13]

Venous blood was collected into a 5-ml EDTA tube. Total blood haemoglobin (g/l), RBC ($\times 10^{12}$/l) and MCV (fl) were measured using a Coulter Model STKS blood counter. From these values, haematocrit was calculated by multiplying RBC count with MCV. Dizygotic twin pairs were selected for linkage analysis on the basis of previous zygosity testing.[14]

Genotypes

DNA was extracted from buffy coats using a modification of the "salt method".[15] For twin pairs of the same sex, zygosity was determined by typing nine independent DNA microsatellite polymorphisms and the X/Y amelogenin marker for sex determination by polymerase chain reaction (ABI Profiler system). All twins were also typed for ABO, Rh and MNS blood groups (zygosity was subsequently confirmed via the genome scans).

The genome scan consisted of 726 highly polymorphic autosomal microsatellite markers at an average spacing of ~5 cM in 539 families (2360 participants). Markers on the X chromosome were also typed, but linkage to these is not reported here. The microsatellites consisted of a combination of markers from the ABI-Prism and CIDR genotyping sets. Overlapping parts of the sample received either a 10 cM scan using the ABI-2 marker set (400 markers) at the Australian Genome Research Facility (Melbourne, Australia) or a 10 cM scan using the Weber marker set at the Center for Inherited Disease Research (Baltimore, Maryland, USA), or both. Only 30 markers were common to both marker sets and were used for quality control; the remaining markers intercalated to form a scan at approximately 5 cM spacing. The only families to receive one scan had both parents genotyped, and so had high information content. The average heterozygosity of markers was 0.78 and the mean information content was 0.77. Although genome scan data were available from parents, twins and siblings, phenotype data (ie, blood cell counts) were available only from twins. Full details of the genome scan are provided elsewhere.[14]

Linkage analyses

Univariate multipoint variance components linkage analysis was used to test for linkage between marker loci and blood cell phenotypes.[16–19] Variance components were estimated by maximum-likelihood analysis of the raw data[20] as implemented in the software package MERLIN[21] along with fixed effects for sex and age (ie, although every effort was made to measure twins at their 12th, 14th and 16th birthdays, some twins were slightly older or younger than this). As both circadian and seasonal effects have been reported for erythrocytes, linear, quadratic and sinusoidal fixed effects were included for the time of day

and month from which blood was sampled.[22] Univariate multipoint linkage analyses were performed for each marker at each age. Only phenotypic data from dizygotic pairs were included in the analyses because monozygotic twins share all their identical-by-descent (IBD) genes across the genome, and are thus uninformative for linkage. Note also that although red cell indices were measured only in twins, parental and sibling genotypes still helped determine IBD sharing between the dizygotic twin pairs.

The null hypothesis that the additive genetic variance in a trait caused by a QTL linked to a given marker is zero (ie, $\sigma_q^2 = 0$) was tested by comparing the likelihood of a reduced model in which σ_q^2 was constrained to zero with the likelihood of a model in which the genetic variance due to the QTL (σ_q^2) was estimated. Twice the difference in natural log-likelihood between these models is distributed asymptotically as a 1/2:1/2 mixture of χ_1^2 and a point mass at zero,[23] whereas the difference between the two $\log_{10}$ likelihoods produces a LOD score equivalent to the classical LOD score of parametric linkage analysis.[24]

Multivariate analyses

Several groups have shown that multivariate methods can increase the power of QTL linkage analysis.[25–29] The rationale is that information on the QTL comes not only from each variable's variance but also from the covariation between the different measures. Multivariate analysis takes advantage of this fact and models both the variation and the covariation between the variables in terms of the underlying QTL. Given the high correlation between the haemoglobin and RBC variables, as well as the moderate negative correlation between the RBC and MCV variables (see Results section), we used a multivariate procedure in an attempt to increase evidence for linkage at the most promising chromosomal regions. Specifically, we performed multivariate analyses of the haemoglobin, RBC and MCV variables on chromosome 6 after the univariate analyses had indicated the possible existence of a QTL affecting these variables in 14-year-old twins (see Results section). We did not perform longitudinal genetic analyses on any of the variables because of the small number of dizygotic twins measured across all occasions (table 1), and did not include HCT in any of the multivariate models as this measure was derived from the RBC and MCV variables.

Multivariate QTL linkage analyses were performed using structural equation modelling as implemented in the computer package Mx[30–32] using data from both monozygotic and dizygotic pairs. The QTL was modelled as a single latent factor which pleiotropically affected the phenotypes. In this model, the correlation between the QTL effects was set to one for monozygotic twins (as monozygotic twins are genetically identical), and to $\hat{\pi}$, the estimated proportion of

shared IBD at the marker locus for dizygotic twins. The probabilities of sharing zero (p_0), one (p_1) or two (p_2) marker alleles IBD for each dizygotic twin pair were calculated in a multipoint fashion using the Lander–Green algorithm as implemented in MERLIN.[21] These probabilities were then used to obtain $\hat{\pi} = p_2 + \frac{1}{2}p_1$ for each dizygotic pair. Residual sources of variation were modelled using saturated Cholesky structures. As phenotypic information was available from monozygotic twins, it was possible to partition the residual familial resemblance into separate variance components owing to additive genetic and common environmental sources of variation. Multivariate analyses included the same fixed effects as in the univariate models. We compared the fit of the full multivariate model with a model where the latent factor containing the QTL effects was absent. Twice the difference in log-likelihoods between the models was evaluated against a χ^2 distribution with 3 degrees of freedom (df).[31]

RESULTS

Table 2 presents the range, means and standard deviations for each of the red cell indices. All variables approximated the normal distribution and displayed minimal skew and kurtosis. We draw the reader's attention to the large inter-individual variation that exists between these healthy, Australian twins.

Figures 1–4 show the results from the genomewide tests of linkage. Each plot displays the LOD scores for all three ages so that it is possible to compare the consistency of results across measurement occasions. Suggestive evidence of linkage occurred at age 14 years on chromosome 6q23 for the highly correlated haemoglobin (LOD = 3.03; fig 1) and haematocrit (LOD = 2.95; fig 2) measures. Some minor evidence was also present at ages 12 and 14 years for RBC on chromosome 6 (fig 3), although none of the red cell indices showed any evidence of linkage in these regions at age 16 years. Another major peak occurred at age 16 years on chromosome 4q32 for RBC (LOD = 3.07; fig 3), with some evidence of linkage at age 12 years but not at age 14 years.

Table 3 summarises several other regions of suggestive linkage (ie, defined here as LOD score >1.5). These included regions which were present either at more than one age or for more than one variable. The most promising example of this was MCV, where there was suggestive evidence of linkage on chromosome 7q at all three ages (fig 4), and chromosome 17, which contained peaks for haemoglobin, RBC, HCT and MCV in similar regions at age 14 years. We found little evidence of linkage on chromosome 16p13.3 or 11p15.5, which contain the genes for the haemoglobin α and β chains (all LODs <1). There was also no evidence of linkage on chromosome 7q21 or 19p13.2, which contain the genes for erythropoietin and its receptor (all LODs <1 except for MCV at age 12 years, where LOD = 1.02 at marker D7S221).

Table 4 displays the correlations between the different red cell indices. In each age group there was a high positive correlation between haemoglobin, HCT and RBC, as well as a moderate negative correlation between RBC and MCV, but little correlation between MCV and the other red cell indices. We attempted to take advantage of this correlation structure by performing a multivariate analysis of haemoglobin, RBC and MCV after the univariate analyses had indicated the possible existence of a QTL on chromosome 6 in 14-year-old twins. Figure 5 displays the results of these analyses. Unfortunately, multivariate analysis of these variables failed to increase evidence of linkage in this region of the genome (a similar analysis performed on just haemoglobin and RBC also produced similar results—results not shown).

Table 2 Minimum and maximum values, means and standard deviations (SD) in the red cell indices for males and females at 12 years, 14 years and 16 years of age

		Females						Males					
	Units	Age	Min	Max	Mean	SD	Skew	Kurt	Min	Max	Mean	SD	Skew	Kurt
Hb	g/l	12	107	164	132.53	8.79	0.172	0.205	105	158	133.19	8.11	0.086	0.338
		14	107	156	131.67	8.43	0.004	−0.146	115	166	139.07	8.88	−0.048	−0.156
		16	110	156	132.57	8.37	0.046	−0.280	124	172	148.55	8.78	0.037	−0.178
HCT	l/l	12	0.318	0.481	0.396	0.025	0.144	0.274	0.329	0.462	0.396	0.023	0.003	0.116
		14	0.330	0.457	0.395	0.023	0.069	−0.316	0.330	0.487	0.415	0.026	−0.122	−0.130
		16	0.340	0.480	0.399	0.024	0.256	−0.083	0.370	0.514	0.444	0.026	−0.102	−0.189
RBC	Cells/pl	12	3.56	5.70	4.63	0.325	−0.034	0.344	3.59	5.84	4.68	0.324	0.230	0.583
		14	3.73	5.45	4.52	0.296	0.318	0.084	3.99	5.72	4.83	0.313	0.114	−0.098
		16	3.78	5.39	4.49	0.283	0.116	0.054	3.99	5.84	5.02	0.301	−0.045	−0.185
MCV	fl	12	74	97	85.57	3.45	−0.009	0.456	75	96	84.67	3.34	0.130	0.074
		14	75	98	87.46	3.64	0.015	0.420	75	99	86.11	3.62	0.394	0.705
		16	77	100	88.92	3.67	0.057	0.330	77	101	88.55	3.47	0.284	0.819

Hb, haemoglobin; HCT, haematocrit; Kurt, kurtosis; Max, maximum; MCV, mean corpuscular volume; Min, minimum; RBC, red blood cell.
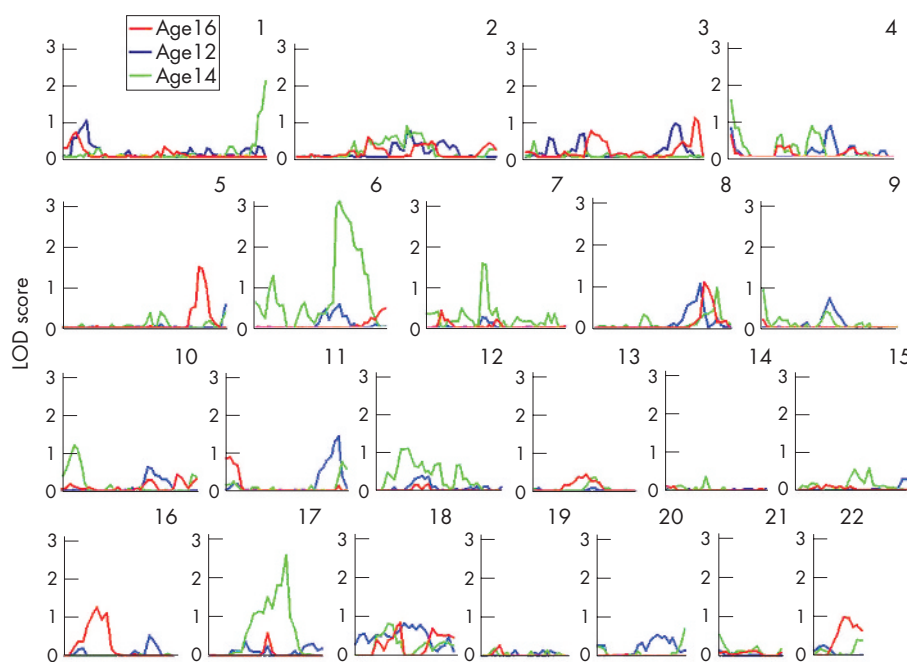
Figure 1 Genome scan for haemoglobin. LOD, limit of detection. Age in years.

DISCUSSION

The most promising result from our study was the large linkage peak on chromosome 6q23 for the haemoglobin and HCT variables in 14-year-old twins. This is the same region identified by Lin et al[7] as part of the Framingham Heart Study, although interestingly, Lin et al[7] found linkage only for HCT and not for the haemoglobin variable. In our study, the haemoglobin and HCT variables were very highly correlated (r = 0.91–0.93 across all sexes and ages) so it was not surprising that the same chromosomal region was implicated by both variables. Chromosome 6q23 contains a number of promising candidate genes including EBP41L2, which codes for protein

4.1G, a member of the erythrocyte cytoskeletal protein 4.1R gene family (defects in erythrocyte cytoskeletal proteins are known to produce anaemia and other haematological anomalies), and HEBP2, which codes for a putative haem-binding protein. A form of hereditary persistence of fetal haemoglobin has also been mapped to this region.[33]

Multivariate analysis of this region in 14-year-old twins failed to increase evidence for linkage. This could be due to a number of factors. Firstly, the distribution of the multivariate test statistic is a complicated mixture of χ^2 distributions which has yet to be well characterised in the literature.[23–31] Evaluating the test against a χ^2 distribution of 3df is slightly conservative
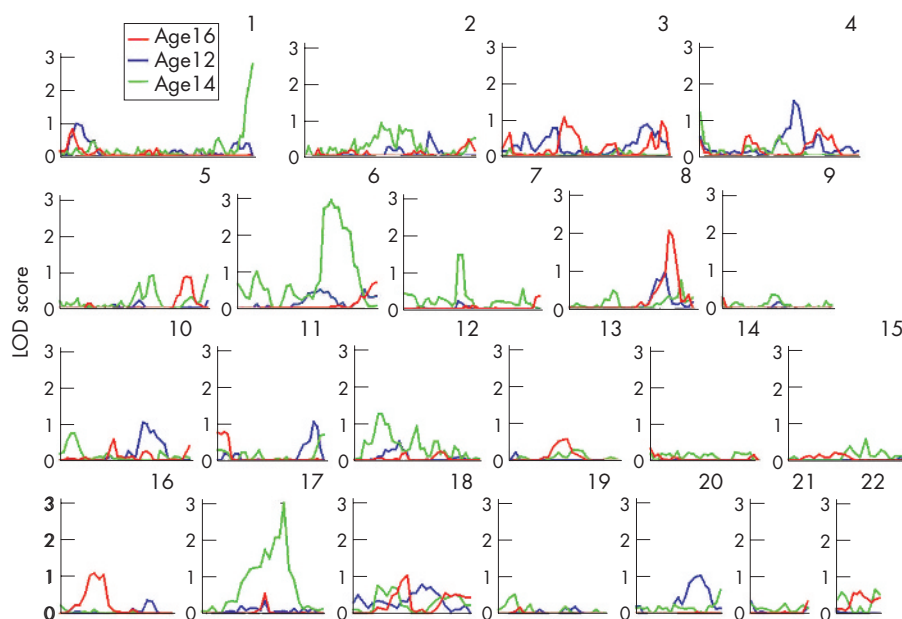

Figure 2 Genome scan for haematocrit. LOD, limit of detection. Age in years.
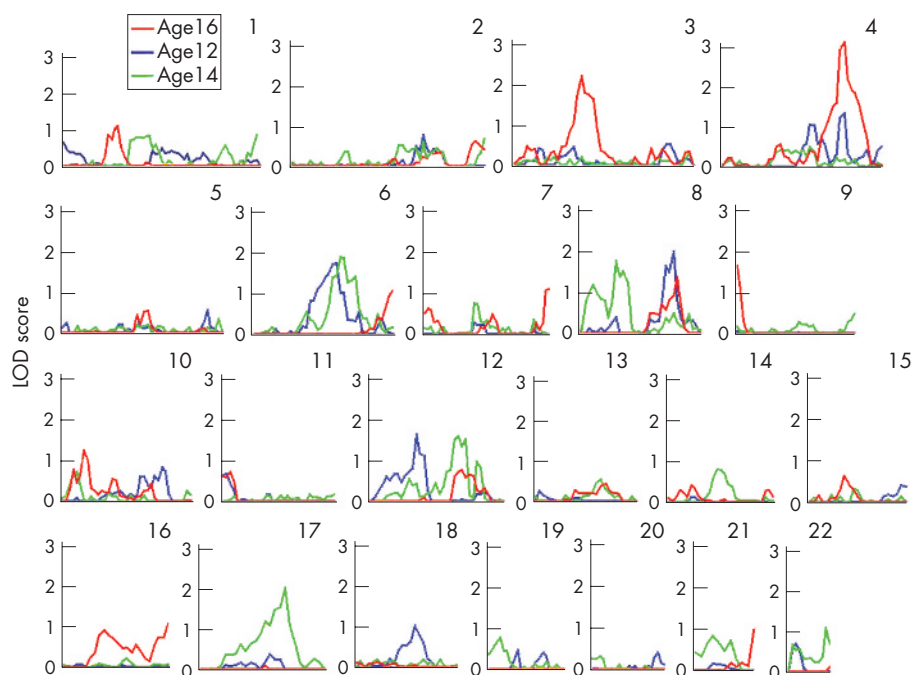
Figure 3 Genome scan for red blood cell count. LOD, limit of detection. Age in years.

and may be partially responsible for the apparent loss in power. Secondly, multivariate linkage analysis increases power to detect linkage only if the QTL pleiotropically affects more than one variable in the analysis. Possibly, the peaks identified in the univariate analyses represent different QTLs which affect the variables separately. In this case, a multivariate analysis of the phenotypes would be less powerful than the univariate analyses, as the multivariate test has more df. Finally, and most cynically, one or more of the univariate results might simply reflect random fluctuation and type I error.

A major strength of our study is the availability of multiple measurements on many of the same persons across time. Although these results do not formally constitute replication (because the data are not independent), they do indicate the extent to which the results are robust with respect to measurement error and temporal changes in the phenotypes. In this vein, it is interesting to note that although there was some evidence of linkage at ages of 12 and 14 years on chromosome 6q23 for the haemoglobin and HCT measures, there was little evidence of linkage in 16-year olds. This is
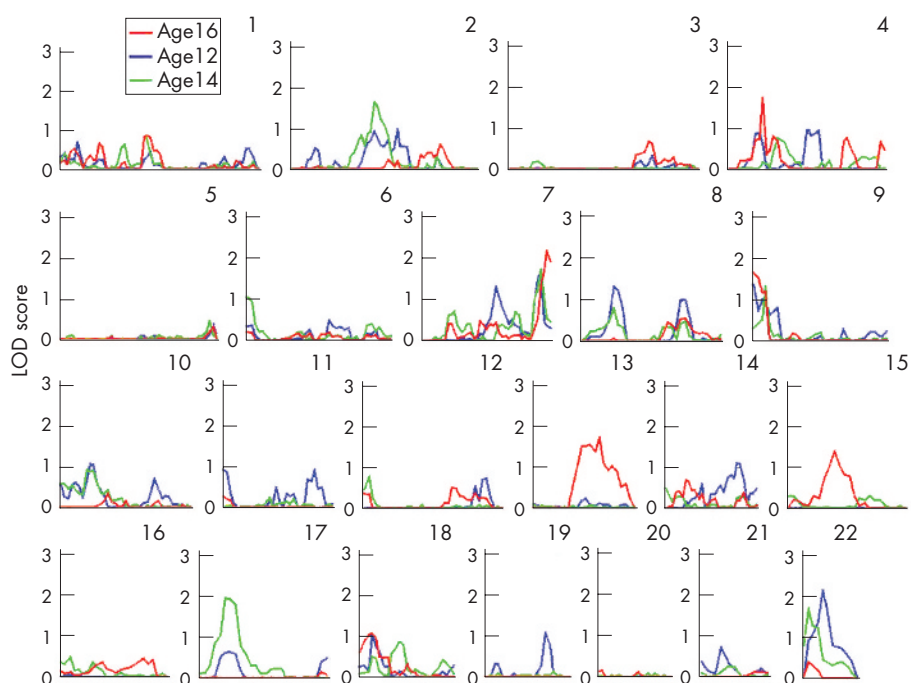

Figure 4 Genome scan for mean corpuscular volume. LOD, limit of detection. Age in years.

Table 3 Regions with multipoint LOD >1.5 from the genome scan of the red cell indices

Chromosome	Phenotype	Age (years)	Region	cM range (Haldane function)	Peak LOD
1	HCT	14	D1S2842 – D1S2836	278.255–288.930	2.72
1	Hb	14	D1S2836	288.930	2.04
2	MCV	14	D2S2216–D2S2972	118.540–123.406	1.63
3	RBC	16	D3S1300–D3S1566	87.870–100.954	2.2
4	MCV	16	D4S2632	59.532	1.73
4	RBC	16	D4S1625–D4S1539	150.163–182.648	3.07
6	RBC	12	D6S474–D6S287	126.170–128.834	1.73
6	RBC	14	D6S1040–D6S262	140.503–141.125	1.88
6	HCT	14	D6S1040–D6S2436	140.503–173.905	2.95
6	Hb	14	D6S1040–D6S2436	140.503–173.905	3.03
7	Hb	14	D7S502–D7S3046	86.776–87.999	1.55
7	MCV	12	D7S798	181.821	1.69
7	MCV	14	D7S798	181.821	1.71
7	MCV	16	D7S3058–D7S559	188.581–195.614	2.14
8	RBC	14	D8S1477	56.517	1.52
8	RBC	12	D8S1784–D8S592	120.456–127.25	1.99
8	HCT	16	D8S514–D8S1179	133.297–138.228	2.02
9	RBC	16	D9S288	8.753	1.62
9	MCV	16	D9S288	8.753	1.61
12	RBC	12	D12S85	64.059	1.63
12	RBC	14	D12S393–D12S346	116.824–118.039	1.61
13	MCV	16	D13S800–D13S265	73.939–87.569	1.74
17	MCV	14	D17S974–D17S1303	32.351–33.751	1.95
17	Hb	14	D17S798–D17S787	59.258–87.946	2.61
17	HCT	14	D17S798–D17S787	59.258–87.946	2.98
17	RBC	14	D17S1868–D17S787	80.324–87.946	2.05
22	MCV	12	D22S689–D22S280	35.746–40.024	2.16
22	MCV	14	D22S539	16.129	1.7

Hb, haemoglobin; HCT, haematocrit; LOD, limit of detection; MCV, mean corpuscular volume; RBC, red blood cell.

particularly interesting given that the phenotypic correlation across measurement occasions was quite high (table 4), and previous longitudinal genetic analyses have suggested that most genetic variance from ages 12 and 14 years is transmitted to 16 years.[13] A possible explanation is that there were fewer dizygotic twins measured at age 16 years than at ages 12 or 14 years (table 1), and hence the power to detect linkage at 16 years was not as great as that at 12 or 14 years. Alternatively, the positive results might simply reflect type I error and stochastic variation in the linkage signal.

Our study also idenitified several other regions of suggestive linkage. Most prominent among these was the large peak on chromosome 4q32 for RBC at age 16 years, and the peak on chromosome 7q36 which displayed linkage to MCV at all ages.
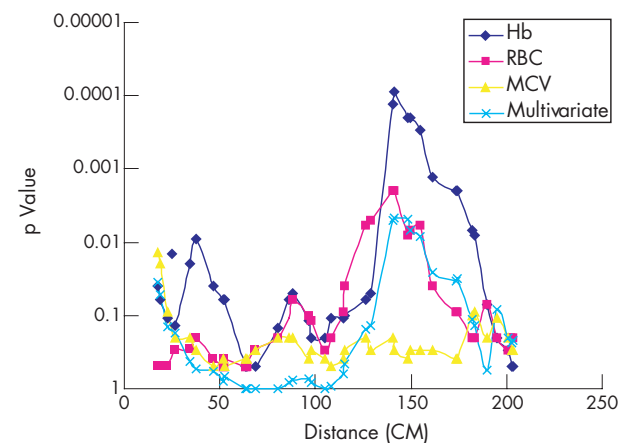

Figure 5 Multivariate analysis of haemoglobin (Hb), red blood cell (RBC) and mean corpuscular volume (MCV) on chromosome 6 for 14-year-old twins. Also shown are the univariate results for each of these variables.

Although there are no obvious candidate genes in these regions, the 7q36 region has previously been implicated in a linkage study involving a large family with heterocellular hereditary persistence of fetal haemoglobin.[34] Although some of these regions may harbour QTLs, other peaks will be the result of random fluctuation and type I error. As always, the key to assessing the relevance of these findings is replication using independent datasets.

Similar to Lin *et al*,[7] we found no evidence of linkage to regions containing the genes for the α and β haemoglobin chains (chromosomes 16p13.3 and 11p15.5), nor to the regions containing the genes for erythropoietin and its receptor (chromosomes 7q21 and 19p13.2). Polymorphisms in all of these genes can produce Mendelian forms of anaemia. The results from this study add weight to the notion that the QTLs responsible for normal variation in the red cell indices differ from the loci responsible for monogenic forms of anaemia, although it should be noted that the power of sib-pair linkage analysis is low, and so QTLs accounting for small proportions of the phenotypic variance could go undetected in our study.

In conclusion, we have found suggestive evidence for a QTL influencing HCT and haemoglobin on chromosome 6q23. This is the same region that was previously identified by Lin *et al*[7] as containing a QTL for HCT. We intend to fine map this region (ie, increasing marker density) and perform association analysis of candidate genes in the area (eg, EBP41L2, HEBP2). This study might represent the first step in the eventual identification of genes, which can increase our understanding of the quantitative genetics of normal variation in the red cell indices.

. . . .

Authors' affiliations
A Iliadou, D M Evans, G Zhu, D L Duffy, G W Montgomery, N G Martin, Queensland Institute of Medical Research, Brisbane, Queensland, Australia
I H Frazer, Centre for Immunology and Cancer Research, University of Queensland, Princess Alexandra Hospital, Brisbane, Queensland, Australia

Table 4 Pearson's correlations between the different red cell indices at 12, 14 and 16 years of age

Males	Females											
	Hb12	RBC12	HCT12	MCV12	Hb14	RBC14	HCT14	MCV14	Hb16	RBC16	HCT16	MCV16
Hb12		0.82	0.95	0.07	0.57	0.39	0.50	0.15	0.48	0.37	0.46	0.13
RBC12	0.80		0.83	−0.44	0.48	0.56	0.43	−0.25	0.34	0.49	0.29	−0.30
HCT12	0.92	0.82		0.13	0.54	0.39	0.51	0.16	0.46	0.38	0.44	0.10
MCV12	−0.04	−0.52	0.04		0.04	−0.34	0.07	0.69	0.11	−0.25	0.16	0.63
Hb14	0.56	0.41	0.51	0.04		0.78	0.91	0.10	0.61	0.35	0.52	0.27
RBC14	0.44	0.63	0.44	−0.44	0.77		0.79	−0.43	0.51	0.56	0.44	−0.19
HCT14	0.52	0.42	0.52	0.03	0.91	0.81		0.20	0.58	0.34	0.53	0.29
MCV14	0.15	−0.31	0.14	0.75	0.22	−0.32	0.30		0.03	−0.35	0.07	0.66
Hb16	0.41	0.21	0.28	0.04	0.68	0.52	0.64	0.13		0.75	0.92	0.29
RBC16	0.38	0.48	0.32	−0.32	0.44	0.62	0.45	−0.30	0.74		0.79	−0.31
HCT16	0.44	0.25	0.36	0.10	0.62	0.51	0.65	0.17	0.93	0.80		0.33
MCV16	0.08	−0.37	0.08	0.66	0.30	−0.17	0.31	0.73	0.31	−0.29	0.34	

Hb, haemoglobin; HCT, haematocrit; MCV, mean corpuscular volume; RBC, red blood cell.

REFERENCES

1 **Gagnon DR**, Zhang TJ, 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.
2 **Toghi H**, Yamanouchi H, Murakami M, Kameyama, M. Importance of the hematocrit as a risk factor in cerebral infarction. *Stroke* 1978;**9**:369–74.
3 **WHO/UNICEF/UNU**. *Iron deficiency anaemia: assessment, prevention, and control*. Geneva: World Health Organization (WHO/NHD/01, 3), 2001.
4 **Dal Colletto GM**, Fulker DW, Barretto OC, Kolya M. Genetic and environmental effects on blood cells. *Acta Genet Med Gemellol* (Roma), 1993;**42**:245–52.
5 **Evans DM**, Frazer IH, Martin NG. Genetic and environmental causes of variation in basal levels of blood cells. *Twin Res* 1999;**2**:250–7.
6 **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.
7 **Lin JP**, O'Donnell CJ, Levy D, Cupples A. Evidence for a gene influencing haematocrit on chromosome 6q23–24: genomewide scan in the Framingham Heart Study. *J Med Genet* 2005;**42**:75–9.
8 **Lindemann JP**, Kang KW, Christian JC. Genetic variance of erythrocyte parameters in adult male twins. *Clin Genet* 1977;**12**:73–6.
9 **Whitfield JB**, Martin NG. Genetic and environmental influences on the size and number of cells in the blood. *Genet Epidemiol* 1985;**2**:133–44.
10 **Aitken JF**, Pfitzner J, Battistutta D, O'Rourke PK, Green AC, Martin NG. Reliability of computer image analysis of pigmented skin lesions of Australian adolescents. *Cancer* 1996;**78**:252–7.
11 **McGregor B**, Pfitzner J, Zhu G, Grace M, Eldridge A, Pearson J, Mayne C, Aitken JF. Genetic and environmental contributions to size, color, shape, and other characteristics of melanocytic naevi in a sample of adolescent twins. *Genet Epidemiol* 1999;**16**:40–53.
12 **Zhu G**, Duffy DL, Eldridge A, Grace M, Mayne C, O'Gorman L, Aitken JF, Ne MC, Hayward NK, Green AC, Martin NG. A major quantitative-trait locus for mole density is linked to the familial melanoma gene CDKN2A: a maximum-likelihood combined linkage and association analysis in twins and their sibs. *Am J Hum Genet* 1999;**65**:483–92.
13 **Evans DM**, Frazer IH, Boomsma DI, Martin NG. Developmental genetics of red cell indices during puberty: a longitudinal twin study. *Int J Hum Genet* 2001;**1**:41–53.
14 **Zhu G**, Evans DM, Duffy DL, Montgomery GW, Medland SE, Gillespie NA, Ewen KR, Jewell M, Liew YW, Hayward NK, Sturm RA, Trent JM, Martin NG. A genome scan for eye color in 502 twin families: most variation is due to a QTL on chromosome 15q. *Twin Res* 2004;**7**:197–210.
15 **Miller SA**, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988;**16**:1215.
16 **Amos CI**. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet* 1994;**54**:535–43.
17 **Almasy L**, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998;**62**:1198–211.
18 **Eaves LJ**, Neale MC, Maes H. Multivariate multipoint linkage analysis of quantitative trait loci. *Behav Genet* 1996;**26**:519–25.
19 **Fulker DW**, Cherny SS. An improved multipoint sib-pair analysis of quantitative traits. *Behav Genet* 1996;**26**:527–32.
20 **Lange K**, Westlake J, Spence MA. Extensions to pedigree analysis. III. Variance components by the scoring method. *Ann Hum Genet* 1976;**39**:485–91.
21 **Abecasis GR**, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002;**30**:97–101.
22 **Touitou Y**, Touitou C, Bogdan A, Reinberg A, Auzeby A, Beck H, Guillet P. Differences between young and elderly subjects in seasonal and circadian variations of total plasma proteins and blood volume as reflected by hemoglobin, hematocrit, and erythrocyte counts. *Clin Chem* 1986;**32**:801–4.
23 **Self SG**, Liang K. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* 1987;**82**:605–10.
24 **Williams JT**, Blangero J. Power of variance component linkage analysis to detect quantitative trait loci. *Ann Hum Genet* 1999;**63**(Pt 6):545–63.
25 **Allison DB**, Thiel B, St Jean P, Elston RC, Infante MC, Schork NJ. Multiple phenotype modeling in gene-mapping studies of quantitative traits: power advantages. *Am J Hum Genet* 1998;**63**:1190–201.
26 **Amos CI**, Laing AE. A comparison of univariate and multivariate tests for genetic linkage. *Genet Epidemiol* 1993;**10**:671–6.
27 **Boomsma DI**. Using multivariate genetic modeling to detect pleiotropic quantitative trait loci. *Behav Genet* 1996;**26**:161–6.
28 **Evans DM**. The power of multivariate quantitative-trait loci linkage analysis is influenced by the correlation between variables. *Am J Hum Genet* 2002;**70**:1599–602.
29 **Martin N**, Boomsma D, Machin G. A twin-pronged attack on complex traits. *Nat Genet* 1997;**17**:387–92.
30 **Eaves LJ**, Neale MC, Maes H. Multivariate multipoint linkage analysis of quantitative trait loci. *Behav Genet* 1996;**26**:519–25.
31 **Marlow AJ**, Fisher SE, Francks C, MacPhie IL, Cherny SS, Richardson AJ, Talcott JB, Stein JF, Monaco AP, Cardon LR. Use of multivariate linkage analysis for dissection of a complex cognitive trait. *Am J Hum Genet* 2003;**72**:561–70.
32 **Neale MC**. *Statistical modelling*, 4th edn. Richmond, VA: Department of Psychiatry, Medical College of Virginia, 1997.
33 **Craig JE**, Rochette J, Fisher CA, Weatherall DJ, Marc S, Lathrop GM, Demenais F, Thein S. Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. *Nat Genet* 1996;**12**:58–64.
34 **Sampietro MT**, Thein SL. Heterocellular hereditary persistence of fetal haemoglobin: strategy for linkage analysis. *Cytogenet Cell Genet*, 1991;**58**:1924–5.