Developmental Genetics of Red Cell Indices During Puberty: A Longitudinal Twin Study

David M. Evans[1], Ian H. Frazer[2], Dorret I. Boomsma[3] and Nicholas G. Martin[1]

[1]*Queensland Institute of Medical Research and Joint Genetics Program, University of Queensland, Brisbane, Australia*
[2]*Centre for Immunology and Cancer Research, University of Queensland, Princess Alexandra Hospital, Brisbane, Australia*
[3]*Department of Experimental Psychology, Free University, Amsterdam, The Netherlands*

KEY WORDS Haemoglobin; red cell count; mean corpuscular volume; twins; genetics.

ABSTRACT Red cell number and size increase during puberty, particularly in males. The aim of the present study was to determine whether expression of genes affecting red cell indices varied with age and sex. Haemoglobin, red cell count, and mean cellular volume were measured longitudinally on 578 pairs of twins at twelve, fourteen and sixteen years of age. Data were analysed using a structural equation modeling approach, in which a variety of univariate and longitudinal simplex models were fitted to the data. Significant heritability was demonstrated for all variables across all ages. The genes involved did not differ between the sexes, although there was evidence for sex limitation in the case of haemoglobin at age twelve. Longitudinal analyses indicated that new genes affecting red cell indices were expressed at different stages of puberty. Some of these genes affected the different red cell indices pleiotropically, while others had effects specific to one variable only.

INTRODUCTION

In healthy adults, haemoglobin, red cell count, haematocrit and mean cell volume change little over time (Costongs et al. 1985; Dot et al. 1992; Fraser et al. 1989; Ross et al. 1988; Statland et al. 1978). From childhood to adulthood, however, there is an increase in these and other indices of erythrocyte number and size. This increase is greater in males than females and is probably due to the combined action of several hormones including testosterone and insulin-like growth factor (IGF-1) that are secreted in increased amounts during adolescence (Anttila et al. 1994; Thomsen et al. 1986).

Several twin studies have demonstrated that part of the variation in red cell indices between

Address for correspondence: Prof. N. G. Martin, Queensland Institute of Medical Research, PO Royal Brisbane Hospital, Brisbane 4029, Australia, *Phone:* +61 7 3362 0278, *Fax:* +61 7 3362 0101, *Email:* nickM@qimr.edu.au

individuals is due to genetic factors (Dal Colletto et al. 1993; Evans et al. 1999; Garner et al. 2000; Lindemann et al. 1977; Whitfield and Martin 1985; Yokoyama and Akiyama 1995). No study to date, however, has examined whether different genes influence red cell indices at different ages. For example, one set of genes might affect erythrocyte numbers before puberty, while another set of genes may be responsible for variation during adulthood. Thomsen et al. (1986) found that a steep rise in haemoglobin concentration in pubescent boys took place during a short period in adolescence, soon after an increase in testosterone production. In contrast, Anttila et al. (1994) in a study of younger boys, failed to find an association between serum testosterone and haemoglobin levels, but did find an association between serum IGF-1 and haemoglobin. These results imply that different systems and consequently different sets of genes may regulate red cell indices at different stages of puberty.

The aim of the present study was to investigate whether expression of genes affecting red cell indices varied with age. Changes in the magnitude of genetic and environmental effects over time can be addressed by cross-sectional studies that measure subjects of different ages. However, in order to assess whether the same genes affect the trait over time, longitudinal data are required. Twins were therefore measured on haemoglobin (HB), red cell count (RBC), and mean corpuscular volume (MCV) over three occasions- at twelve, fourteen and sixteen years of age, corresponding roughly to early, middle and late stages of puberty. It was hypothesised that genes responsible for variation between individuals in the early stages of puberty (ie. twelve

years) would differ from the genes operating at later stages of puberty (ie. fourteen and sixteen years).

MATERIALS AND METHODS

Subjects and Measures

Data were collected as part of ongoing studies into the development of melanocytic naevi (moles) at ages twelve and fourteen, and cognition at age sixteen, the clinical protocols of which have been described in detail elsewhere (Aitken et al. 1996; Evans et al. 1999; McGregor et al. 1999; Wright et al. Submitted; Zhu et al. 1999). 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 prior to testing. The results reported here concern data collected from May 1992 to June 1999. Twins were tested as closely as possible to their twelfth, fourteenth and sixteenth birthdays. Data were obtained from 578 pairs of twins comprising 108 monozygotic female (MZF), 111 monozygotic male (MZM), 91 dizygotic female (DZF), 96 male (DZM) and 172 opposite sex (DZOS) twin pairs (including 85 pairs where the female was born first and 87 pairs where the male was born first). Not all twins were tested across all three measurement occasions (see Table 1 for a breakdown of these data). No attempt was made to exclude subjects suffering from illness, although a few such cases were subsequently excluded as outliers (see below). Venous blood was collected into a 5ml EDTA tube. Total blood haemoglobin g/L (HB), red blood cell count x

10^{12}/L (RBC), and mean corpuscular volume fL (MCV) were measured using a Coulter Model STKS blood counter. Zygosity in same sex pairs was determined by typing 8 highly polymorphic DNA microsatellite markers and three blood groups (ABO, MNS, Rh). In 50 pairs of twins where DNA was not available, zygosity was judged by similarity of appearance.

Statistical Analyses

In order to allow use of all available data and avoid listwise deletion of cases, structural equation modeling using maximum-likelihood analysis of individual observations was performed using the computer package Mx (Neale 1997). The structural equation model comprised two parts: (1) a model for the expected values of observations in terms of fixed effects such as age and sex (the "model for the means"), and (2) a model for the covariance matrix of residuals after removal of these fixed effects. Significant twin correlations establish the presence of familial aggregation for the measure of interest, but do not distinguish between the possible mechanisms through which this similarity arises. Phenotypic variation may be conceptualised as the net result of four sources of variation: additive genetic (A), non-additive genetic (D), and common environmental (C) sources of variation which make siblings more alike, and random environmental variation (E) which make siblings different. The goal of structural equation modeling is to determine which combination of A, C, D and E provides the most parsimonious fit to the observed data. The task is further complicated by the fact that C and D are negatively

Table 1: Breakdown of participation data showing number of twin pairs who were tested at each age

Measurement Occasion			Zygosity					
12	14	16	MZF	MZM	DZF	DZM	DZOS	Total
Ö			20	23	23	30	50	146
	Ö		7	7	4	5	9	32
		Ö	17	22	18	9	25	91
Ö	Ö		19	18	23	24	42	126
Ö		Ö	3	4	2	1	5	15
	Ö	Ö	8	8	5	6	8	35
Ö	Ö	Ö	34	29	16	21	33	133
Total in simple model			108	111	91	96	172	578
Total at 12			76	74	64	76	130	420
Total at 14			68	62	48	56	92	326
Total at 16			62	63	41	37	71	274

confounded so that they cannot be estimated concurrently in a study of MZ and DZ twins reared together. This does not imply that C and D cannot both contribute to the phenotypic variance of a trait, rather they cannot be estimated simultaneously with data from twins reared together.

Univariate Analyses

We began by fitting a general sex-limitation model to each variable at each age (see Fig. 1). The model for the means consisted of a sex deviation, as well as regression terms for age and for the time of day which the blood was sampled. It is well known that blood cell number is subject to circadian influences and follows a sinusoidal pattern over the 24-hour period. Within the narrow range of sampling times in the present study, the relationship between time of sampling and blood cell number seemed to be approximated adequately by a linear function, and this was included in the means model along with age and sex. The model for the covariances consisted of different variance components for males and females (ie. different estimates for A, C and E), as well as an unconstrained additive genetic cor-

relation between opposite sex twin pairs (Neale and Cardon 1992). This model was compared against more simplified sub-models. First the additive genetic correlation between opposite sex twin pairs was constrained to 0.5 (equivalent to testing whether different sets of genes affected the trait in males and females i.e. "non-scalar sex-limitation"), and then male and female variance components were equated (equivalent to testing whether the magnitude of genetic and environmental effects differed between males and females). Finally, the significance of each variance component was tested by dropping it from the model. All comparisons between models were assessed by difference in log-likelihood $D c^2$- twice the difference in log-likelihood being distributed as c^2 with the degrees of freedom (df) for this test equal to the difference in df between the full and sub-models.

Longitudinal Analyses

A cholesky decomposition was performed on the time series data from each variable (Neale and Cardon 1992). In the cholesky model the expected variance-covariance matrix is parameterised in terms of n factors (where n is
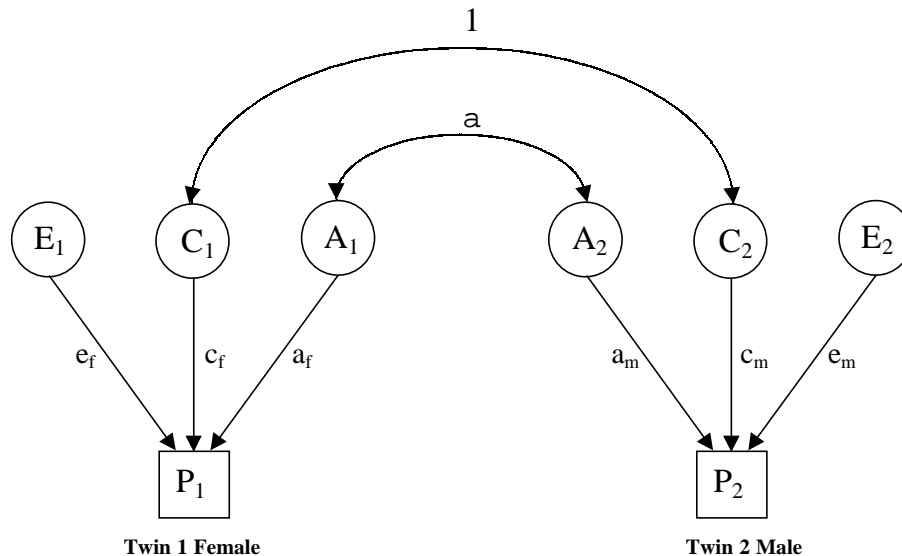

Fig. 1. Path diagram for a univariate general sex-limitation model for opposite sex twin pairs. Note that the additive genetic correlation between these pairs is unconstrained, and that males and females have different path coefficients

the number of variables), where all variables load on the first factor, n-1 variables load on the second factor and so on, until the final variable loads on the nth factor only. Each source of phenotypic variation (ie. A, C or D, and E) is parameterised in such a way (see Fig. 2). The model for the means for these analyses consisted of eight parameters: age deviations at fourteen and sixteen years of age, sex deviations for ages twelve, fourteen and sixteen and finally regression terms for the time of day at which blood was sampled on each of the three measurement occasions.
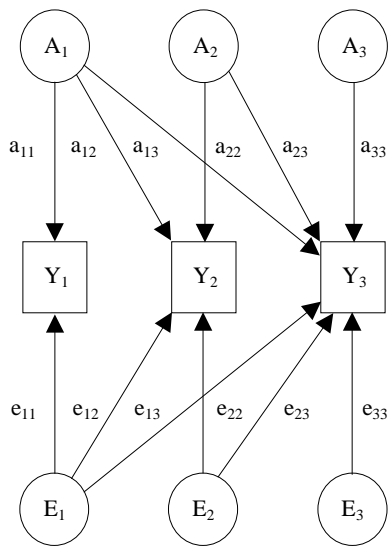

Fig. 2. A Cholesky decomposition. The observed variance-covariance matrix is decomposed into additive genetic, common environmental (or non-additive genetic) and unique environmental sources of variation and covariation. Each of these sources is parameterised in terms of n factors (where n is the number of variables) where all variables load on the first factor, n-1 variables load on the second factor and so on, until the final variable loads on the nth factor only

As pointed out by Boomsma et al. (1989), the problem with using a cholesky design to analyse longitudinal data is that a cholesky structure does not take full advantage of the time series nature of the data (ie. that causation is unidirectional throughout time). We therefore fitted a simplex model - a model that explicitly takes into account the longitudinal nature of the data - to each of the red cell variables (Fig. 3). Simplex models are autoregressive models where the latent variable at time (i) is causally related to the immediately preceding latent variable (i − 1) through a linear relation ie.

$$h_i = b_i h_{i-1} + z_i$$

where h_i is the latent variable at time (i), b_i is the linear regression of the latent factor on the previous latent variable, and z_i represents a new input (innovation) at time (i) which is uncorrelated with h_{i-1}. When using data from MZ and DZ twin pairs, structural equations of this type can be expressed for additive genetic sources of variation (A), common environmental (C) (or non-additive genetic sources of variation (D)) and unique environmental sources of variation (E). Also part of the model is a structural equation relating the observed phenotypes to the latent factors:

$$y_i = l_i h_i + e_i$$

where l_i is the factor loading of the observed phenotype on the latent variable at time (i), and e_i is a measurement error term which affects the phenotype, but is uncorrelated with h_i. In order for the model to be identified, either the factor loadings of the observed on latent variables must be set to unity and the variance of the innovations estimated, or alternatively, the variance of the innovation terms standardised to one and the factor loadings estimated. In the present study, the factor loadings of the observed variables on the latent factors (ie. the ls) were set to one and the variance of the innovation terms was estimated.

A final point concerns the distinction between innovations of latent factors (z_i) and the measurement errors of observed variables (e_i). The innovations are that part of the latent factor at time (i) that is not caused by the latent factor at time (i-1), but are part of every subsequent time point. In contrast, measurement errors are terms that do not influence observed variables at subsequent time points. As such, the simplex design allows discrimination of transient factors that affect measurement at one time point only, and factors that are continuously present or exert a long term influence throughout the time

series (Boomsma et al. 1989; Neale and Cardon 1992). Note also that the variance of the measurement error terms affecting the first and last measurement occasions must be constrained equal in order for the model to be identified. This is because error variance at these occasions would otherwise be indistinguishable from innovation variance. In the present study, the variance of all measurement error terms was constrained equal.

The simplex model in the present study consisted of sixteen parameters: the variances of the three innovations for each of the three sources of variation (ie. A, C and E sources for HB and RBC; A, D and E sources for MCV), two transmission coefficients for each source of variation, and finally the variance of the measurement error term that affected phenotypes at each age but was not transmitted to subsequent ages. The

model for the means was the same as that used for the cholesky analysis.

RESULTS

Of the 578 pairs of twins tested, data were not available (eg. blood specimen unsuitable for analysis, insufficient amount of blood taken etc.) from nineteen individuals at age twelve, twelve individuals at age fourteen, and six individuals at age sixteen. As this study was concerned with the causes of variation within the normal range (as opposed to pathological causes of extreme values) one twin's data were excluded from all analyses because of aplastic anaemia. Scores greater than 3.5 standard deviations from the mean would be excluded from further analyses. Two subjects' scores were excluded from HB at fourteen (166 g/L) and sixteen (98 g/L) years of
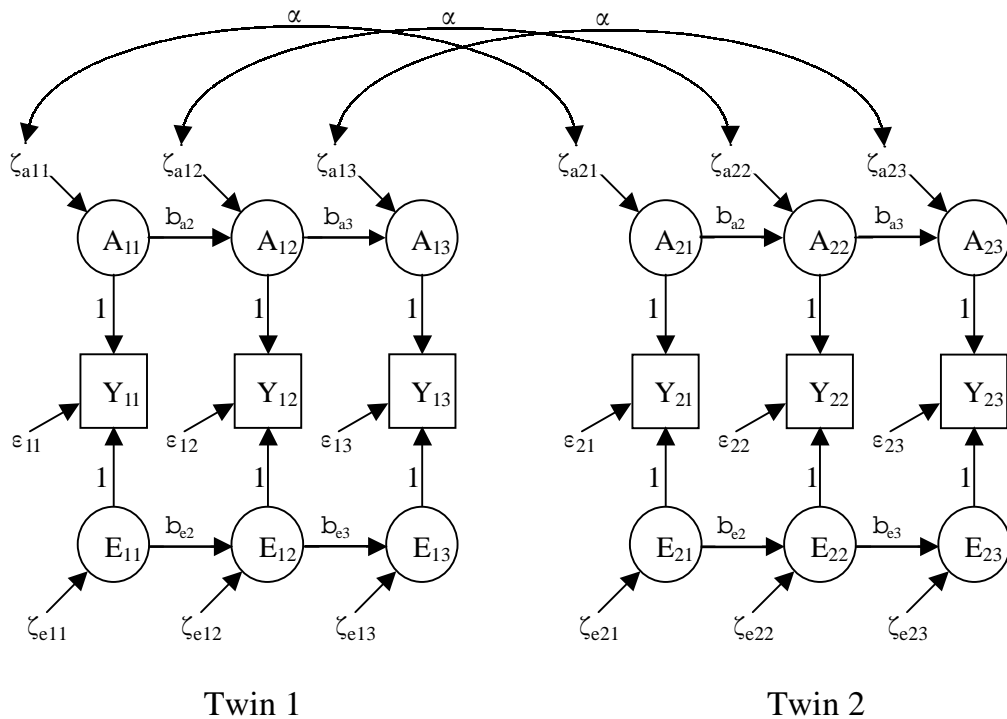
Fig. 3. Genetic simplex model for a single variable measured across three time points. The loadings of the observed variables (Y) on the latent factors are set to unity, and the variance of the innovation terms (z) are estimated. A measurement error term (e) also influences the observed phenotype at each occasion. The variance of this error term is equated across twins and measurement occasions. For MZ twins, a = 1, for DZ twins a = 0.5. Only additive genetic (A) and environmental sources (E) of variation are included in this path diagram. Similar structures may be written for common environmental sources (C) or non-additive genetic (D) sources of variation

age. Two individuals were excluded from analyses involving RBC (12yrs: 6.30, 6.25 cells/L x 10^{-12}; 14yrs: 6.22, 6.12 cells/L x 10^{-12}; 16 yrs: 6.60, 6.38 cells/L x 10^{-12}) and MCV (12yrs: 62, 64 fL; 14yrs: 66, 68 fL; 16 yrs: 65.9, 68.8 fL) at all ages, and a further two subjects were excluded from analyses involving MCV at fourteen (75fL) and sixteen (71 fL) years respectively.

Multivariate outliers were identified by fitting an empirical model to the data where males and females had separate means and variances. The square root of the Mahalanobis distance for each pedigree (i) was then calculated using the formula:

$$Q_i = (x_i - m_i)' S_i^{-1} (x_i - m_i)$$

where S_i^{-1} is the maximum likelihood estimated covariance matrix, m_i the estimated mean and x_i the observed trait value of the pedigree (i). Q_i was then transformed to approximate the standard normal distribution using the function:

$$Z_i = ((Q_i/n_i)^{1/3} - 1 + 2/(9n_i))(9n_i/2)^{1/2}$$

where n_i is the number of individuals in pedigree (i) (Hopper and Matthews 1982). Pedigrees with Z scores in excess of 3.5 were excluded from further analyses for that variable. This included a pair of twins for MCV at twelve (86.3, 80 fL), a pair at fourteen (83, 78 fL) and one pair who was excluded from analyses involving MCV because of extreme values at sixteen years of age (78, 79 fL).

The means and standard deviations of all blood measurements for males and females are listed in table 2. In females, there was little change in HB from twelve to sixteen years, while RBC decreased and MCV increased. In males, all red cell indices increased with age.

Correlations between twin pairs are presented in table 3 for the different zygosity groups. The correlation between MZ twin pairs was very high for all variables at all ages, particularly in the case of MCV. The corresponding DZ correlations were substantial, but lower, suggesting that genetic factors contributed to variation in all indices at all ages. The correlation between DZ

Table 2: Red cell means and standard deviations for male and female subjects ignoring their twin relatedness

Variable	Age	Range	Females			Males		
			N	Mean	SD	N	Mean	SD
HB (g/L)								
	12	107 - 164	404	133.61	8.81	416	134.74	7.76
	14	109 - 166	318	133.87	7.96	320	140.70	8.55
	16	113 - 172	272	133.91	8.32	268	149.92	9.03
RBC (Cells/L x 10^{-12})	12	3.56 - 5.84	404	4.66	.32	414	4.72	.32
	14	3.73 - 5.72	319	4.57	.30	318	4.87	.32
	16	3.78 - 5.84	273	4.49	.29	266	5.01	.32
MCV (fL)								
	12	74 - 97	404	85.37	3.37	412	84.50	3.32
	14	78 - 98	316	87.56	3.31	318	85.65	3.33
	16	71 - 99	269	89.17	3.32	266	88.73	3.44

Table 3: Correlations between twin pairs at 12, 14 and 16 years for HB, RBC and MCV[a]

Age	HB			RBC			MCV		
	12	14	16	12	14	16	12	14	16
Zygosity									
MZF	.86	.74	.81	.87	.84	.87	.96	.94	.88
MZM	.82	.74	.84	.90	.81	.86	.94	.96	.96
DZF	.62	.59	.33	.63	.62	.50	.41	.18	.45
DZM	.58	.47	.64	.65	.60	.64	.56	.49	.57
DZOS	.48	.35	.27	.49	.46	.17	.43	.37	.38
Pooled									
MZ	.84	.77	.91	.88	.86	.91	.95	.95	.92
DZ	.53	.46	.42	.57	.51	.42	.45	.32	.44

[a]Number of pairs for each age and zygosity is approximately as shown at the foot of table 1

male twins remained fairly constant across all ages. In contrast, the correlation between female DZ twin pairs decreased noticeably at age fourteen in the case of MCV and at age sixteen for HB and RBC. The correlation between opposite sex twin pairs decreased from age twelve to sixteen across all variables, most noticeably for HB and RBC suggesting that gene expression may differ between the sexes.

Univariate Analyses

Table 4 displays the standardised variance components for the best fitting univariate models, the -2 log-likelihood of the general sex limitation models, and the difference in c^2 between the general sex limitation models and the best fitting models. Genetic factors contributed significantly to all variables at all ages accounting for between 31 and 96 percent of the total variance. Heritability increased from age twelve to

sixteen in the case of HB, but stayed relatively constant for RBC and MCV. Non-additive genetic factors were significant for MCV at age fourteen. Common environmental factors were only significant at age twelve for HB. The remaining variables were best explained by AE models at all ages. Sex-limitation was significant in the case of HB at twelve, and for MCV at sixteen years of age. The additive genetic correlation between opposite sex twin pairs could not be constrained to 0.5 in twelve year olds for HB ($Dc_1^2 = 3.866$, p = .049), suggesting that different sets of genes affected the trait in males and females. Males had higher heritabilities than females for MCV.

Longitudinal Analyses

Cholesky Model

Standardised results of the cholesky analyses

Table 4: Univariate genetic modeling for red cell indices at each age. Parameter estimates for the best fitting model only are shown. The general model specifies regression terms for sex, age, time of blood sampling, has different variance components for males and females, and an unconstrained additive genetic correlation between opposite sex twins (a)

| Variable | Age | Males | | | Females | | | a | General Model | | Best Model | |
		A	C	E	A	C	E		-2LL	df	Dc^2	df
HB	12	.63	.19	.18	.63	.19	.18	.16	5497.426	809	6.567	4
HB	14	.73	-	.27	.73	-	.27	.5	4329.536	627	6.023	5
HB	16	.81	-	.19	.81	-	.19	.5	3697.125	529	6.718	5
RBC	12	.86	-	.14	.86	-	.14	.5	101.709	807	8.995	5
RBC	14	.80	-	.20	.80	-	.20	.5	62.031	626	8.070	5
RBC	16	.83	-	.17	.83	-	.17	.5	50.333	528	9.155	4
MCV	12	.96	-	.04	.96	-	.04	.5	3828.078	805	4.572	5
MCV	14	.42	.53[a]	.05	.42	.53[a]	.05	.5	2979.060	623	2.377	4
MCV	16	.95	-	.05	.88	-	.12	.5	2550.165	524	2.641	3

[a]refers to a non-additive genetic variance component

Table 5: Cholesky decompositions of covariance between measures at ages twelve, fourteen and sixteen[a]. Results are standardised to unit variance at each age. The non-additive genetic part of the model could be dropped for MCV without significantly affecting the model's fit ($Dc_6^2 = 8.009$, p = .237)

	A1	A2	A3	C1	C2	C3	E1	E2	E3
HB_{12}	.80			.44			.41		
HB_{14}	.60	.52		-.26	.18		.05	.52	
HB_{16}	.61	.30	.57	-.01	.13	.00	.03	.06	.44
RBC_{12}	.80			.47			.38		
RBC_{14}	.61	.44		.10	.46		.02	.46	
RBC_{16}	.59	.32	.45	.33	.28	.00	.05	.08	.40
MCV_{12}	.98			-			.21		
MCV_{14}	.85	.47		-	-		.01	.23	
MCV_{16}	.78	.30	.46	-	-	-	.03	.11	.27

[a] HB: -2LL = 13324.822, df = 1971; RBC: -2LL = -115.558, df = 1967; MCV: -2LL = 8583.612, df = 1958

are presented in table 5. Note especially the large loadings of all variables on the second and third genetic factors. If the same set of genes were responsible for variation across all three measurement occasions, all the measurements would have loaded highly on the first genetic factor, and the factor loadings on the second and third genetic factors would have been zero. This was clearly not the case for any of the variables, suggesting that new genetic input was affecting the red cell indices at measurement occasions two and three.

Univariate Simplex Models

Results of fitting the simplex models to the three time points for each variable are presented in figure 4. Comparison of the log-likelihoods between the full cholesky and simplex models revealed that in all cases the simplex models provided an adequate fit to the observed data (HB: $Dc_2^2 = 1.019$, p = .601; RBC: $Dc_2^2 = 5.339$, p = .069; MCV: $Dc_2^2 = 5.537$, p = .063). Dropping the common environmental part of the simplex model resulted in a significant deterioration in fit for HB (HB: $Dc_5^2 = 28.591$, p = .000) and RBC ($Dc_5^2 = 30.441$, p = .000). However, the unique environmental simplex structures could be dropped without significantly affect-

ing the fit of the models (HB: $Dc_5^2 = 9.742$, p = .083; RBC: $Dc_5^2 = 6.182$, p = .289). For MCV, the non-additive genetic simplex structure (D) could be dropped without significantly affecting the model ($Dc_5^2 = 8.613$, p = .126), but not the unique environmental part of the model ($Dc_5^2 = 37.109$, p = .000), implying that stable unique environmental influences influenced the trait throughout puberty.

Examination of the maximum likelihood parameter estimates revealed that the most important variance component of each model was the large genetic innovation influencing the first measurement occasion (Fig. 4). Note also the large additive genetic transmission coefficients which indicate that most of this genetic variance was transmitted to subsequent ages. Smaller genetic innovations were also present at fourteen and sixteen years. At age fourteen, 14.6%, 14.6% and 21.9% of the total variance for HB, RBC, and MCV respectively were due to genetic innovation, whereas at the same age 53.0%, 57.0%, and 72.3% of the total variance were due to genetic influences transmitted from age twelve. At age sixteen, 29.1%, 15.4%, and 21.7% of the total variance for HB, RBC, and MCV respectively were due to genetic innovation, whereas at the same age, 51.4%, 67.9%, and 70.3% of
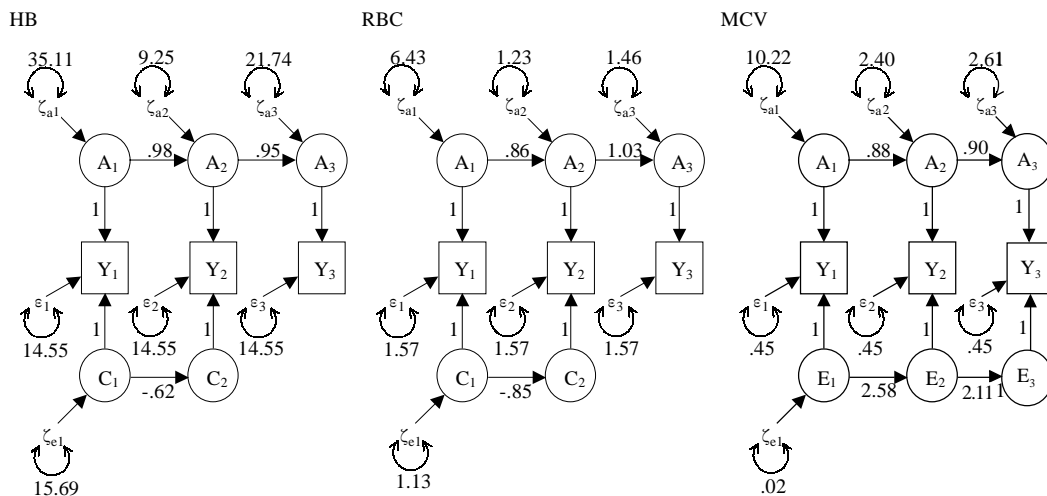

Fig. 4. Maximum likelihood parameter estimates for HB, RBC and MCV simplex models. Some of the environmental components could be dropped without significantly affecting the fit of the models. The variance components for the RBC model have been multiplied by 100

the total variance were due to genetic influences transmitted from previous ages.

If genetic variance at the second and third measurement occasions was due solely to the amplification of existing genetic variance then we should be able to drop the genetic innovations at the second and third measurement occasions without significantly affecting the fit of the model. Dropping the genetic innovation at fourteen years resulted in a significant deterioration in fit for RBC and MCV, but not HB (HB: Dc_1^2 = 2.214, p = .137; RBC: Dc_1^2 = 5.313, p = .021; MCV: Dc_1^2 = 107.111, p = .000). However, dropping the genetic innovation at sixteen resulted in a significant deterioration in fit for all the variables (HB: Dc_1^2 = 13.325, p = .000; RBC: Dc_1^2 = 7.186, p = .007; MCV: Dc_1^2 = 83.888, p = .000). These results indicate that part of the genetic variance at these ages is due to new genes being expressed over and above amplification of existing genetic variance.

Bivariate Analyses

The previous results suggested that the correlations between successive red cell measurements were explained adequately by underlying simplex processes. As the different red cell indices were correlated, the same simplex process might underlie the covariance between different indices across measurement occasions. In particular, the correlation between HB and RBC was high across all time points, while RBC and MCV had a moderate negative correlation (Table 6). To investigate this possibility we first fitted two cholesky models - one to the HB and RBC data, and the other to the RBC and MCV data, incorporating the same means model to each variable as described previously. RBC was rescaled to cells/L x 10^{-11} to aid optimisation of the maximum likelihood function.

The pattern of factors obtained was very similar for both the HB-RBC (-2LL = 19932.794, df = 3911) and RBC-MCV (-2LL = 17371.622, df = 3898) analyses (Table 7). All variables loaded highly on the first genetic factor suggesting the existence of a common set of genes affecting both variables across measurement occasions. However, for both analyses, the variables

Table 6: Pearson correlations between the different red cell indices at ages 12, 14 and 16

		HB_{12}	HB_{14}	HB_{16}	RBC_{12}	RBC_{14}	RBC_{16}	MCV_{12}	MCV_{14}	MCV_{16}
					Females					
HB_{12}		1.00	.36	.52	.80	.24	.35	.09	.13	.13
HB_{14}		.35	1.00	.50	.31	.77	.34	.01	.03	.08
HB_{16}	M	.40	.66	1.00	.45	.50	.77	.01	-.06	.13
RBC_{12}	A	.80	.28	.37	1.00	.51	.61	-.46	-.38	-.35
RBC_{14}	L	.27	.79	.59	.56	1.00	.67	-.45	-.53	-.46
RBC_{16}	E	.37	.47	.80	.63	.74	1.00	-.45	-.50	-.42
MCV_{12}	S	-.03	-.05	-.17	-.57	-.55	-.58	1.00	.80	.75
MCV_{14}		.07	.11	.07	-.46	-.46	-.45	.86	1.00	.81
MCV_{16}		-.06	.16	.15	-.54	-.35	-.41	.86	.85	1.00

Table 7: Cholesky decompositions of HB-RBC data and of RBC-MCV data (standardised results)

	A1	A2	A3	A4	A5	A6	C1	C2	C3	C4	C5	C6	E1	E2	E3	E4	E5	E6
HB_{12}	.77						.48						.42					
HB_{14}	.56	.48					-.12	.42					.06	.52				
HB_{16}	.56	.28	.53				.10	.35	.00				.02	.06	.44			
RBC_{12}	.58	-.09	.05	.54			.40	.15	.11	.17			.33	-.03	.06	.18		
RBC_{14}	.36	.26	.18	.53	.17		-.05	.49	.00	.00	.00		.00	.42	.00	.03	.18	
RBC_{16}	.41	-.02	.48	.45	.02	.19	.14	.39	.08	.11	.00	.00	.00	.07	.34	.02	.09	.20
RBC_{12}	.80						.47						.38					
RBC_{14}	.62	.42					.09	.47					.02	.46				
RBC_{16}	.63	.24	.44				.28	.32	.00				.04	.09	.39			
MCV_{12}	-.48	-.08	-.09	.78			-.15	-.24	.13	.07			-.03	-.01	-.01	.23		
MCV_{14}	-.47	-.23	-.12	.71	.37		-.07	-.11	-.05	.03	.00		.00	-.01	-.04	.00	.23	
MCV_{16}	-.37	-.12	-.12	.62	.37	.36	-.20	-.13	.16	.08	.00	.00	.00	-.03	-.05	.03	.09	.27

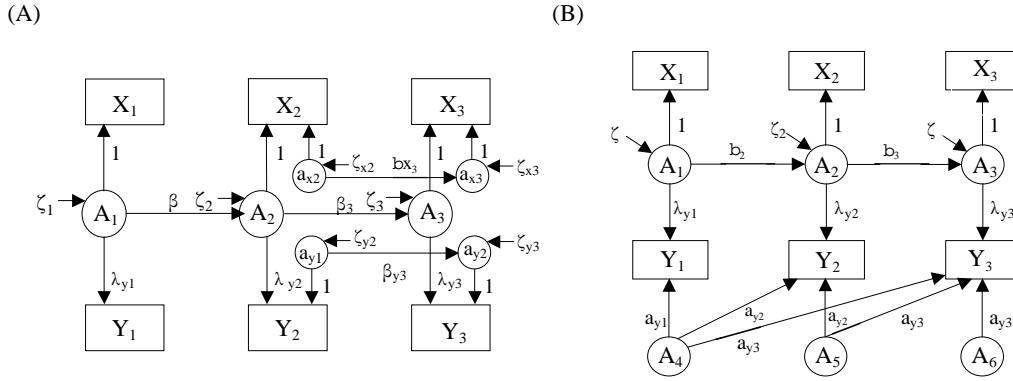(A)   (B)

Fig. 5. Bivariate models for the covariation between red cell indices across measurement occasions. In Model (A), genetic covariance is the result of a common simplex process and two variable specific processes that begin at the second measurement occasion. In Model (B) genetic covariance between the two variables is again the result of a common simplex process. A cholesky type structure also permits genetic covariance across measurement occasions within the second variable only. Both models also include full cholesky structures for common and unique environmental sources of variation (not shown)

measured on occasion two loaded on a second genetic factor, and the variables measured on occasion three loaded on a third genetic factor, suggesting that new sets of genes which affect both traits became operational at fourteen and sixteen years of age. Finally, there was a fourth, fifth and a sixth genetic factor on which the second variable loaded, indicating the expression of variable specific genes across all measurement occasions.

We then fitted a hybrid model, similar to that used in Boomsma et al. (1989), to the HB and

RBC data, and to the RBC and MCV data. These models contained a genetic simplex process common to both variables (Fig. 5A), and cholesky type structures for C and E. In order for the model to be identified, the factor loadings of one of the variables on the common genetic process were fixed to unity and the factor loadings of the other variable were estimated. The model also contained genetic simplex structures unique to each variable. These variable specific structures began at the second measurement occasion in order to separate their effects

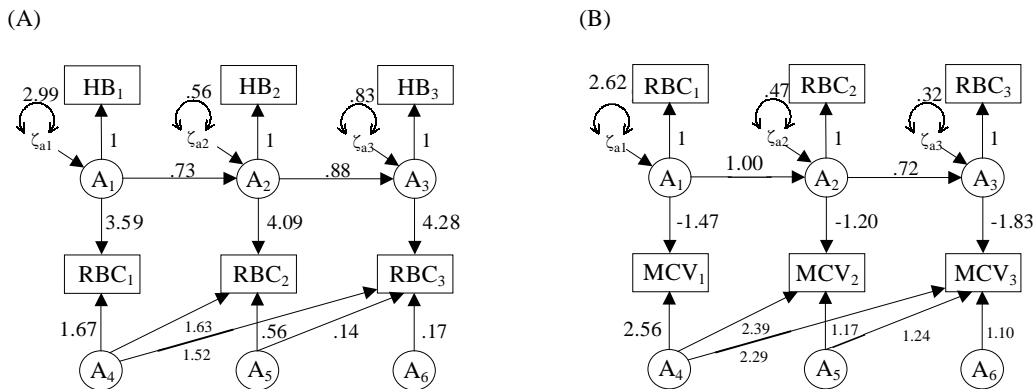(A)   (B)

Fig. 6. Bivariate models for (A) HB and RBC (-2LL = 19949.111, df = 3918), and for (B) RBC and MCV (-2LL = 17377.771, df = 3905). For simplicity, common environmental and unique environmental cholesky structures are not shown

from the structure common to both variables, and thus for the model to be identified (Boomsma et al. 1989).

Both models fitted the data poorly compared with the full cholesky model (HB-RBC model: $Dc_7^2 = 85.910$, p = .000; RBC-MCV: $Dc_7^2 = 57.654$, p = .000), suggesting that additional processes contributed to genetic covariation across measurement occasions. The inadequacy of the model was presumably because genetic covariance between the first measurement occasion and subsequent occasions independent of the common simplex process (ie. as indicated by the fourth genetic factor in the cholesky analysis) could not be accounted for by the model. We therefore replaced both variable specific simplex structures with a single genetic cholesky structure in each of the models (see Fig. 5B). Both these models were not significantly different in fit from the bivariate cholesky models (HB-RBC model: $Dc_7^2 = 12.223$, p = .093; RBC-MCV: $Dc_7^2 = 6.149$, p = .522). Figure 6 displays the significant maximum likelihood parameter estimates for these models. Note the significant factor loadings on the common simplex process for both models, indicating that the same simplex process underlay much of the genetic covariance between the variables. Note also the significant factor loadings on the variable specific cholesky structures suggesting that genes which are variable specific are also switched on during adolescence.

DISCUSSION

This study demonstrates that new genes affecting red cell indices are expressed during adolescence. While most of the genetic variance in each of the traits was transmitted from previous ages, a significant portion of variance was introduced by genetic innovation at sixteen years, and innovation at fourteen years in the case of RBC and MCV. The challenge now is to identify the genes responsible for this variation.

The bivariate results indicate that the same simplex process underlies much of the genetic covariation between HB and RBC, and also between RBC and MCV. This implies that not only do a common set of genes affect these variables, but also "new" genes which are expressed dur-

ing puberty affect both traits. The negative factor loadings of the MCV variables on the RBC-MCV common process indicate that the genes which affect both variables have opposite effects. This is not surprising as there is a well documented negative correlation between red cell size and volume (Evans et al. 1999; Whitfield and Martin 1985).

Both bivariate analyses also suggested the existence of variable specific genetic processes in addition to the common process. This process was modeled by a cholesky structure which contained three genetic factors. It therefore seems likely that genes which are variable specific are also switched on during adolescence. In the present study this variable specific process was not approximated well by simplex structures. As alluded to earlier, this was perhaps because (in order for the model to be identified) the variable specific simplex structures began from the second measurement occasion and thus did not permit covariation between the first measurement occasion and subsequent time points.

For the most part, variation in males appeared to be due to the same genetic influences as in females. Even though the correlations between opposite sex twins were substantially lower than the correlations between same sex DZ twins, particularly at age sixteen, these sex-limitation effects did not approach statistical significance. The exception was HB at age twelve where the correlation between additive genetic effects in males and females was 0.16. This result may reflect type I error brought about by the large number of comparisons performed in this study. Alternatively, the effect might be genuine, in which case partly different genes determinine haemoglobin levels in males and females. If this is the case, then it is interesting that the effect was not replicated in fourteen and sixteen year olds- particularly as the correlation between opposite sex twins was lower in both these age groups. This could be because of lower numbers of twins tested at these ages and the associated lack of power to detect such an effect. Previous studies have failed to find evidence for different sets of genes influencing red cell indices in males and females, although these studies used smaller sample sizes than in the present experiment (Dal Colletto et al. 1993; Evans et

al. 1999; Garner et al. 2000; Lindemann et al. 1977; Whitfield and Martin 1985; Yokoyama and Akiyama 1995). We also found that additive genetic factors were more important in determining MCV in sixteen year old males than females. This difference might have been due to menstruation producing increased differences between sisters. Clearly the question of sex-limitation deserves further testing in samples of greater size.

Common environmental factors were only significant in the case of HB at age twelve. This was in contrast to Whitfield and Martin (1985) who failed to find significant evidence for common environmental effects on HB and RBC, but found that common environmental effects influenced MCV. Because young cotwins live together, they are likely to share very similar dietary iron intakes and exercise patterns (which affect red cell counts),(Boyadjiev and Taralov 2000) which may contribute significantly to common environmental variation in red cell indices at this age. As twins get older, and presumably become more independent of their parents and each other, they are less likely to share the same diet, exercise routine etc. In older twins, therefore, influences such as diet and exercise still contribute to variation, but are specific rather than common environmental influences. Similarly, it was unsurprising that common environmental influences did not contribute significantly to MCV, as many young co-twins are unlikely to engage together in activities which affect mean corpuscular volume (eg. smoking or consuming alcohol with each other).

Interestingly the common environmental transmission (beta) coefficients for the simplex analyses of HB and RBC were negative (Fig. 4). The sign of these coefficients does not affect the expected variances of the model, but does affect the expected covariances- the negative sign serving to decrease the expected covariance between measurement occasions. The negative transmission coefficients then were presumably estimated as a result of the low observed correlation between ages twelve and fourteen for these variables (Table 6). These low correlations may be a result of differences in the timing and speed of maturation between subjects (Loesch et al. 1995). For example, some subjects may mature early and therefore exhibit a high degree of change between ages twelve and fourteen, whereas others may mature later and exhibit only a small change between twelve and fourteen. In this way, trait values at twelve may not be predictive of trait values at fourteen. Similar effects would also explain why some of the correlations between twin pairs were abnormally low (eg. DZ females for MCV at fourteen years of age). If twins differed slightly in their timing of puberty, and this coincided with the time they were measured, the correlation between twin pairs would be decreased. This low correlation between DZ females for MCV at fourteen years was probably responsible for the significance of non-additive genetic factors in the univariate analysis at this age.

In conclusion, the present study demonstrates that new genes affecting red cell indices are expressed during adolescence over and above those that are expressed at age twelve. Some of these genes affect the different indices pleiotropically, while others have variable specific effects. These genes appear to be largely the same in males and females. We are now performing a genome scan of the DZ twins and siblings from this study with the aim of identifying quantitative trait loci underlying regulation of the red cell indices and other blood cell types. We intend using some of the multivariate models specified here to increase the power of our linkage analysis (Boomsma 1996; Martin et al. 1997).

ACKNOWLEDGEMENTS

Diseases. The haematological work for this study was funded by Queensland Health Pathology Services.

REFERENCES

Aitken JF, Pfitzner J, Battistutta D, O'Rourke, PK, Green AC, Martin NG 1996. Reliability of computer image analysis of pigmented skin lesions of Australian adolescents. *Cancer,* **78(2)**: 252-7.

Anttila R, Koistinen R, Seppala M, Koistinen H, Siimes MA 1994. Insulin-like growth factor I and insulin-like growth factor binding protein 3 as determinants of blood hemoglobin concentration in healthy subjects. *Pediatr Res,* **36(6)**: 745-8.

Boomsma DI 1996. Using multivariate genetic modeling to detect pleiotropic quantitative trait loci. *Behav Genet,* **26(2)**: 161-6.

Boomsma DI, Martin NG, Molenaar PC 1989. Factor and simplex models for repeated measures: application to two psychomotor measures of alcohol sensitivity in twins. *Behav Genet,* **19(1)**: 79-96.

Boyadjiev N, Taralov Z 2000. Red blood cell variables in highly trained pubescent athletes: a comparative analysis. *Br J Sports Med,* **34(3)**: 200-4.

Costongs GM, Janson PC, Bas BM, Hermans J, Brombacher PJ, van Wersch JW 1985. Short-term and long-term intra-individual variations and critical differences of haematological laboratory parameters. *J Clin Chem Clin Biochem,* **23(2)**: 69-76.

Dal Colletto GM, Fulker DW, Barretto OC, Kolya M 1993. Genetic and environmental effects on blood cells. *Acta Genet Med Gemellol (Roma),* **42(3-4)**: 245-52.

Dot D, Miro J, Fuentes-Arderiu, X 1992. Within-subject biological variation of hematological quantities and analytical goals. *Arch Pathol Lab Med,* **116(8)**: 825-6.

Evans DM, Frazer IH, Martin NG 1999. Genetic and environmental causes of variation in basal levels of blood cells. *Twin Res,* **2(4)**: 250-7.

Fraser CG, Wilkinson SP, Neville RG, Knox JD, King JF, MacWalter RS 1989. Biologic variation of common hematologic laboratory quantities in the elderly. *Am J Clin Pathol,* **92(4)**: 465-70.

Garner C, Tatu T, Reittie JE, Littlewood T, Darley J, Cervino S, Farrall M, Kelly P, Spector TD, Thein SL 2000. Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood,* **95(1)**: 342-6.

Hopper JL, Matthews JD 1982. Extensions to multivariate normal models for pedigree analysis. *Ann Hum Genet,* **46**: 373 - 383.

Lindemann JP, Kang KW, Christian JC 1977. Genetic variance of erythrocyte parameters in adult male twins. *Clin Genet,* **12(2)**: 73-6.

Loesch DZ, Hopper JL, Rogucka E, Huggins RM 1995. Timing and genetic rapport between growth in skeletal maturity and height around puberty: similarities and differences between girls and boys. *Am J Hum Genet,* **56(3)**: 753-9.

Martin N, Boomsma D, Machin G 1997. A twin-pronged attack on complex traits. *Nat Genet,* **17(4)**: 387-92.

McGregor B, Pfitzner J, Zhu G, Grace M, Eldridge A, Pearson J, Mayne C, Aitken JF, Green AC, Martin NG 1999. Genetic and environmental contributions to size, color, shape, and other characteristics of melanocytic naevi in a sample of adolescent twins. *Genet Epidemiol,* **16(1)**: 40-53.

Neale MC 1997. *Statistical Modelling.* (4th edition ed.). Richmond, VA: Department of Psychiatry, Medical College of Virginia.

Neale MC, Cardon LR 1992. *Methodology for Genetic Studies of Twins and Families.* Dordrecht, The Netherlands: Kluwer Academic Publisers.

Ross DW, Ayscue LH, Watson J, Bentley SA 1988. Stability of hematologic parameters in healthy subjects. Intraindividual versus interindividual variation. *Am J Clin Pathol,* **90(3)**: 262-7.

Statland BE, Winkel P, Harris SC, Burdsall MJ, Saunders AM 1978. Evaluation of biologic sources of variation of leukocyte counts and other hematologic quantities using very precise automated analyzers. *Am J Clin Pathol,* **69(1)**: 48-54.

Thomsen K, Riis B, Krabbe S, Christiansen C 1986. Testosterone regulates the haemoglobin concentration in male puberty. *Acta Paediatr Scand,* **75(5)**: 793-6.

Whitfield JB, Martin NG 1985. Genetic and environmental influences on the size and number of cells in the blood. *Genet Epidemiol,* **2(2)**: 133-44.

Wright MJ, Boomsma D, Geffen G, Geffen L, Ando J, De Geus E, Posthuma D, Van Baal C, Hansell N, Luciano M, Smith G, Hasegawa T, Hiraishi K, Ono Y, Martin N Submitted. Genetics of cognition: outline of collaborative twin study. *Twin Research.*

Yokoyama Y, Akiyama T 1995. Intrapair differences of the blood cell components and lymphocyte subsets in monozygotic and dizygotic twins. *Acta Genet Med Gemellol (Roma),* **44(3-4)**: 203-14.

Zhu G, Duffy DL, Eldridge A, Grace M, Mayne C, O'Gorman L, Aitken JF, Neale MC, Hayward NK, Green AC., Martin NG 1999. 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,* **65(2)**: 483-92.