

## Unravelling the basis of variability in cobalamin levels in the general population

Toby Andrew<sup>1,2</sup>, Raj Gill<sup>1</sup>, Irina Gillham-Naseny<sup>1</sup> and Kourosh R. Ahmadi<sup>1,3\*</sup>

<sup>1</sup>Department of Twin Research and Genetic Epidemiology, King's College London, London, UK

<sup>2</sup>Genomics of Common Disease, Imperial College London, London, UK

<sup>3</sup>Department of Nutrition and Metabolism, Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK

(Submitted 28 May 2012 – Final revision received 22 February 2013 – Accepted 22 February 2013 – First published online 29 April 2013)

### Abstract

Cobalamin (Cbl) is an essential B vitamin involved in the normal functioning of the nervous system, the formation of key components of blood, DNA synthesis and methylation, and energy production. Physiological levels of Cbl vary greatly within populations, although the basis for this variability remains largely unknown. We conducted a twin study to characterise the basis of variation in plasma Cbl levels and to test whether common genetic polymorphisms in genes known to cause defects in inborn errors of Cbl metabolism and transport are also associated with mean plasma Cbl levels in the general population. The present results showed that plasma levels of Cbl were heritable, with genetic and phenotypic variance increasing with age, and levels significantly correlated with age, BMI, exercise, alcohol consumption, smoking status, social class and folate levels, which collectively accounted for up to 15% of Cbl variation. Of eight genes responsible for the defects of the Cbl metabolic pathway (cblA–G and mut), *MMAA*, *MMACHC*, *MTRR* and *MUT* harboured polymorphisms that showed evidence of association with Cbl levels. Characterisation of the heritable component of variation in Cbl levels can facilitate the early diagnosis and prognosis of Cbl insufficiency/deficiency in individuals at a higher risk of associated diseases.

**Key words:** Cobalamin: Inborn errors of metabolism: Twin studies: Genetics

Ever since its discovery as a cure for pernicious anaemia, plasma levels of Cbl have been consistently identified as a modifiable endophenotype associated with an increased risk of a number of common, multifactorial diseases, including CVD, neurodegenerative disorders and different forms of cancer<sup>(1)</sup>.

Population levels of cobalamin (Cbl) vary greatly within and between populations. This variability is explained in part by the differences in dietary intake, but also through inter-individual variability in molecular mechanisms responsible for the absorption, distribution, metabolism and elimination of Cbl. In turn, each of these processes is underpinned by the interplay of a myriad of genetic and environmental factors that remain largely unknown. By far the best studied of the molecular processes affecting Cbl disposition is Cbl metabolism, which has led to the identification of a number of genetic causes of altered Cbl metabolism including rare, severe disorders, such as inborn errors of intracellular Cbl metabolism<sup>(2)</sup>. In addition, a more common, subtle variation in Cbl levels in the population can arise as a result of intra-genic polymorphisms affecting Cbl-associated pathway(s) or intergenic regions with few obviously functional links to

normal Cbl disposition, as identified recently through genome-wide association studies<sup>(3,4)</sup>. What has become increasingly apparent is that Cbl variability in the general population has a complex molecular underpinning, which remains largely uncharacterised.

In 1931, Garrod<sup>(5)</sup> hypothesised that rare inborn errors of metabolism were merely extreme examples of what is commonly observed as more minor variations in chemical behaviour, which, in contemporary terms, would be described as conferring protection or susceptibility to diseases. So far, eight specific defects of the Cbl metabolic pathway have been identified (cblA–G and mut) and all but one (cblH) of the causative genes have been identified<sup>(6,7)</sup>.

In line with Garrod's prediction, we hypothesise that common, less penetrant genetic variants at loci/genes that underlie inborn errors of Cbl metabolism are also associated with the large population variability in plasma levels of Cbl. In order to test this hypothesis and improve our understanding of the main factors responsible for population variation in Cbl levels, we carried out a classical twin study with two specific aims: (1) to characterise and quantify the genetic and environmental basis for inter-individual variation in constitutive

**Abbreviations:** Cbl, cobalamin; DZ, dizygotic twins; MZ, monozygotic twins.

\* **Corresponding author:** Dr K. R. Ahmadi, email k.ahmadi@surrey.co.uk

plasma levels of Cbl and (2) to test Garrod's prediction by examining whether common genetic variations (SNP) in genes that are known to cause inborn errors of Cbl metabolism, are also candidate genes for non-pathological variation in plasma Cbl levels.

**Materials and methods**

*Subjects*

Plasma Cbl was determined in a total of 2424 twins from the TwinsUK registry (Table 1)<sup>(8)</sup>. All participants were of North European descent (Caucasians) and between 18 and 79 years old. We excluded twins from the study who were on any medication(s) affecting Cbl status, with a history of chronic CVD, hypertension or diabetes, or on any dietary supplements in the 3 months before participation. Prescription and over-the-counter drug and supplement use was recorded at an interview by a trained nurse. Weight and height were recorded using scales (to the nearest gram) and a stadiometer (in cm) with the subject wearing no shoes and light clothing.

The present study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving human subjects/patients were approved by the St Thomas' Hospital Research Ethics Committee (EC04/015 TwinsUK). Written informed consent was obtained from all subjects.

*Biochemical assays*

Fasting blood samples were collected from each participant and stored on ice. Whole blood was fractionated into serum and plasma within 2 h of collection using standard procedures. Plasma samples were stored at -20°C until assayed. Plasma Cbl and folate levels were determined by a competitive immunoassay using a direct, chemiluminescent technology<sup>(9)</sup>. Total homocysteine (tHcy) levels were measured using isotope dilution liquid chromatography-tandem MS<sup>(10)</sup>.

*Preliminary statistical analysis*

*Equality of means and variance by zygosity.* Plasma Cbl, folate, age and BMI were recorded as continuous variables. We formally tested the equality of means and standard deviations for identical or monozygotic (MZ) and non-identical

or dizygotic (DZ) twins in order to determine whether the data conform to one of the assumptions of the classical twin model, i.e. there is no significant phenotypic mean or variance difference by zygosity status for the measure of interest. For this purpose, regression analyses and the equality of variance tests were conducted for age, BMI, folate and Cbl levels by zygosity to test for potentially significant differences. Observed differences by zygosity were controlled for by a quantile-normal transformation before analysis (see the 'Genetic model fitting' section). All preliminary analyses were performed using STATA (Stata Corporation)<sup>(11)</sup>.

*Covariates of cobalamin.* Robust regression analyses (robust to outliers and heteroscedasticity) were used to test the effects of age, BMI, exercise, alcohol consumption, smoking status, social class and folate on Cbl levels using STATA's 'regress' and 'rreg' regression procedures. Twin-pair relatedness was accounted for in these analyses using the 'cluster' option in STATA.

*Testing for heteroscedasticity – the relationship between cobalamin variance and age.* To test for changes in Cbl variance with age, we used a mixed linear model for the regression of the between-twin variance (derived as a function of the square of the twin-pair mean<sup>(12)</sup>) *v.* the fixed effects of age and age-squared. The age-squared quadratic term was included along with age to test whether the relationship was curvilinear or linear. Additional random effects were fitted for between-family variability (i.e. the intercept term specifies the family identifier) and the age slope term to estimate the population variation in both these estimates. A negative correlation between the random-effect intercept and age slope terms was modelled using STATA's 'unstructured' option.

*Genetic model fitting*

The classical twin model can be used to partition the phenotypic variance and thereby quantify the relative influence of genetic and environmental factors upon the individual variation in a population mean. The model assumes no epistasis, gene-environment correlation or interactions, and these shared environmental factors are not confounded by zygosity (the equal environments assumption). MZ twins generally share identical genes, while DZ twins share, on average, half their segregating genes. To infer heritability, therefore, greater similarity in MZ twin pairs compared with DZ pairs is attributed to genetic factors.

For estimating heritability, two standard methods are used: variance components analysis implemented using maximum likelihood<sup>(13)</sup> and regression-based methods that employ the regression of the twin *v.* co-twin phenotype to infer variance components<sup>(14)</sup>. Each method has its own advantages; thus, we used both analytical approaches in the present study. To implement a maximum-likelihood heritability model, we used the computer package Mx (www.vcu.edu/mx/), designed for the analysis of twin and family data, to partition the observed total Cbl variance ( $V_p$ ) into additive polygenic ( $V_A$ ) and dominant ( $V_D$ ) genetic variance component estimates and environmental components shared by both twins ( $V_C$ ) and unique to each twin ( $V_E$ ). For the analysis, we used

**Table 1.** Characteristics of the study population (Mean values and standard deviations for a total of 2110 female twins)

	MZ		DZ	
	Mean	SD	Mean	SD
<i>n</i>	524		1586	
Age (years)	48.1	13.3	48.7	12.6
Weight (kg)	66.2	11.1	67.6	11
Height (m)	1.6	0.1	1.6	0.1
BMI (kg/m <sup>2</sup> )	25.2	4.2	25.7	4.0
Cobalamin (pmol/l)	442.8	198.8	414.7	157.9
Homocysteine (μmol/l)	6.1	2.9	6.3	2.3
Folate (nmol/l)	27.9	11.6	28.1	11.1

MZ, monozygotic twins; DZ, dizygotic twins.

quantile-normal-transformed Cbl values stratified by zygosity to address the sensitivity of variance components analysis to deviations from a phenotypic normal distribution<sup>(13)</sup>. Although regression-based DeFries & Fulker<sup>(14)</sup> heritability analyses are robust to phenotypic deviations from normality, we used the same quantile-normal-transformed Cbl phenotype both for consistency and to control for potential differences observed by zygosity. We also repeated the analyses, taking Cbl residuals for BMI before the quantile-normal transformation to check for potential confounding by zygosity.

### Candidate gene study

**Genotype data.** Genetic data for the twins were used for the candidate gene study based on the genotypes that were already measured using the Illumina Human317-Quad for the one-half of the TwinsUK sample and the Illumina Human610-Quad Bead Chip for the other half. Illumina SNP for each candidate gene were identified for the analysis by including all intragenic SNP plus  $\pm 20$  kb (NCBI build 36) from the start and stop codons of each gene to represent core and distal regulatory regions.

**Quality control.** For quality control, two levels were used<sup>(15)</sup>. First, subjects were excluded based on the genotype data when (1) the SNP call rate was  $< 95\%$  (missingness) and (2) heterozygosity was  $> 37$  or  $< 33\%$  across all SNP, which is a measure of the expected genetic variation in a population in the absence of any systematic genotyping or sample-handling errors. The samples were tested for population stratification using multidimensional scaling, by comparison with three HapMap phase 2 reference populations (CEU (Utah residents with Northern and Western European ancestry), YRI (Yoruban) and CHB/JPT (Han Chinese in Beijing, China/Japanese)). Ethnic outliers were identified and excluded. Second, SNP were removed if they deviated from Hardy–Weinberg equilibrium at  $P \leq 1 \times 10^{-4}$ , a minor allele frequency  $\leq 1\%$  and a SNP call rate  $\leq 95\%$ .

**Selection of SNP in candidate gene regions.** We used the ‘tagger’ tool in Haploview software<sup>(16)</sup> to quantitatively assess how well the available genotype data represented, i.e. tagged, all known common genetic variations in each gene using a pairwise  $r^2$  metric threshold  $\geq 0.8$  (referred to as tagging efficiency). The degree of linkage disequilibrium – or the extent to which small genomic regions historically co-segregate (and are associated) with one another – in each gene region was also assessed by plotting the pairwise correlation for all the genotyped SNP for each gene (data not shown).

**Genetic association analysis.** Tests of the genetic association between Cbl levels and each SNP were conducted using GenABEL implemented within the statistical package R<sup>(17)</sup>. Briefly, each SNP was tested for the association with Cbl levels using an additive genetic model<sup>(18)</sup> in which each SNP is coded 0, 1 or 2, reflecting a count of the number of copies of the minor allele observed at each diploid locus. For each test, we used a likelihood ratio test to compare a model that included a SNP with a null model in which the SNP is removed from the regression. The likelihood ratio test provides a statistical  $\chi^2$  test of significance with one df

for the additive genetic model. Relatedness among twins was accounted for using a weighted analysis based on a pairwise kinship coefficient matrix for all the study subjects calculated using the genome-wide Illumina panel of SNP.

**Multiple testing.** The type I error rate was controlled for using a Bonferroni correction with the adjusted statistical threshold of significance  $\alpha' = \alpha/n$ , where  $\alpha$  is the nominal significance threshold for a single test (conventionally  $\alpha = 0.05$ ) and  $n$  equals the total number of genotyped SNP tested. The total number of genotyped SNP included in the present study was thirty-eight with a total of eight genes tested, which corresponds to a conservative significance threshold of  $\alpha' = 0.05/38 = 0.001$ .

## Results

Characteristics and blood measurements of the study participants are summarised in Table 1. We excluded the data from male twins ( $n$  150) due to inadequate power as well as twins who suffered from any chronic diseases ( $n$  164). A total of 2110 female twins (262 MZ, 784 DZ twin pairs and nineteen unknown zygosity singletons) were included in the study. None of the participants was treated with drugs known to interfere with Cbl uptake, including anticonvulsants, peptic ulcer drugs, antimetabolites, trimethoprim, hydrazides and mestoranol, or with B-complex supplements.

For each participant, we had data for age, BMI, self-report questionnaire data on alcohol use, smoking status and employment (used to derive social class using the UK General Registrar Classification), Cbl, tHcy and folate levels. These twins represented about one-third of the females on the TwinsUK database. Although they had a mean age that is approximately 5 years older than the remaining TwinsUK database volunteers, they were comparable for mean weight, prevalence of smoking, levels of alcohol consumption as well as a number of other demographic and lifestyle variables. Furthermore, the Cbl sample had a slightly lower prevalence of smoking and alcohol consumption compared with other age-matched populations<sup>(19)</sup>.

### Relationship between age, BMI, exercise, alcohol consumption, smoking status, social class, folate, total homocysteine and cobalamin levels

We identified a total of thirty individuals who were putatively Cbl deficient (Cbl  $< 150$  pmol/l), despite this group having the same mean age and age range compared with the rest of the sample (18–79 years old). Altogether, 88% of the participants had a Cbl value within the normal range (150–660 pmol/l). As elevated levels of Cbl ( $> 660$  pmol/l) are uncommon and may be due to liver disease, we checked liver function test protein levels (alanine transaminase,  $\gamma$ -glutamyl transpeptidase and alkaline phosphatase) for 241 women with Cbl values above 660 pmol/l and found them to be all normal and within the normal range (data not shown).

Small differences in mean and variance for Cbl levels between MZ and DZ twin pairs were identified and controlled for in the heritability analyses by transforming Cbl values

using a quantile-normal function stratified by zygosity. Age ( $\beta = 7.5$ ) and age-squared ( $\beta = -0.1$ ,  $P = 7 \times 10^{-7}$ ,  $R^2 < 0.02$ ) and BMI ( $\beta = -2.3$ ,  $P = 3 \times 10^{-4}$ ,  $R^2 < 0.006$ ) were shown to have a small, but significant effect on Cbl levels with BMI showing a negative and age a positive relationship with Cbl levels (Table 2). Women who reported taking regular physical exercise in the previous year (11% of all female twins), on average, had Cbl levels that were 30 pmol/l higher than those who reported no regular exercise ( $P = 0.02$ ,  $R^2 = 0.004$ ). Similarly for alcohol consumption, women who self-reported either as never or drinking less than 5 units/week (73%) had plasma Cbl levels that were, on average, 32 pmol/l higher than those who reported drinking between 5 and 40 units/week ( $P = 0.02$ ,  $R^2 = 0.005$ ). Women who reported never having smoked (55%) had Cbl levels that were on average 35 and 21 pmol/l higher than those who were current (18%) and ex-smokers (28%), respectively ( $P = 0.09$ ,  $R^2 = 0.005$ ).

Based on their employment status, women who were classified (UK General Registrar Classification) as being social class D/E (semi-skilled or unskilled) had Cbl levels that were on average 48 and 72 pmol/l lower than the social class groups A/B (senior management/professional) and C1/C2 (skilled non-manual/skilled manual,  $P = 0.02$ ,  $R^2 < 0.01$ ), respectively. Folate levels were strongly and positively associated with the variation in Cbl levels ( $\beta = 8.9$ ,  $P = 8 \times 10^{-4}$ ,  $R^2 < 0.14$ ; Table 2). Multiple regression analysis confirmed that age, BMI and folate remained associated with Cbl levels, collectively accounting for approximately 15% of the variation in Cbl levels (data not shown).

### Effect of age on cobalamin levels

Previous studies have shown a clear increase in Cbl deficiency and insufficiency in individuals over the age of 60 years due to the decreased absorption of protein-bound Cbl in the elderly<sup>(20)</sup>. We sought to verify the relationship between Cbl levels and age in the present data. Scatter plots illustrated the positive curvilinear relationship between plasma Cbl levels and age and the large variation in Cbl levels for all ages in this population-based sample of female twins. Despite the large variation, average plasma Cbl levels (fitted line) were seen to increase modestly from the age of 18 years until approximately the mid-40s, plateau and then decrease from the age of 60 years (data not shown).

To assess heteroscedasticity (i.e. the change in Cbl variance across the range of an explanatory variable), we initially plotted the standard deviation of Cbl against arbitrarily defined age bins, and observed a positive increase in phenotypic variance with age (data not shown). When we formally tested this relationship using a mixed model to regress between-twin Cbl variance upon age and age-squared, we confirmed a significant ( $\chi^2 = 12$ ,  $P = 0.003$ ) curvilinear relationship with the variance in Cbl, declining from 18 years to mid-40s and increasing again after the age of 50 years.

### Genetic model fitting and estimation of heritability

The results of genetic model fitting for the quantile-normalised Cbl data are shown in Table 3. The most parsimonious and best-fitting model included an additive polygenic and unique environmental variance component (AE model), with up to

**Table 2.** Association of plasma cobalamin (Cbl) levels with age, BMI, self-reported current physical activity, current alcohol consumption, smoking status, social class and plasma folate levels

	Cbl			Model		
	$\beta$	SE	P	n	R <sup>2</sup>	P
Age (years)	7.5	1.6	$7 \times 10^{-7}$	2210	0.02	$7.4 \times 10^{-7}$
Age + age <sup>2</sup> * (years)	-0.1	0.02	$2 \times 10^{-4}$			
BMI (kg/m <sup>2</sup> )	-2.3	0.64	$3 \times 10^{-4}$	2210	0.006	$3 \times 10^{-4}$
Physical activity†	30.1	13.3	$2 \times 10^{-2}$	1225	0.004	$2 \times 10^{-2}$
Alcohol consumption‡	-31.6	13.2	$2 \times 10^{-2}$	1225	0.005	$2 \times 10^{-2}$
Smoking status§						
Never	-	-	-	2210	0.005	$9 \times 10^{-2}$
Current	-35.2	15.4	$2 \times 10^{-2}$			
Ex-smoker	-21.2	14.5	0.14			
Social class						
D/E	-	-	-	1225	0.01	$2 \times 10^{-2}$
A/B	64.6	28.2	$2 \times 10^{-2}$			
C1/C2	83.5	30.4	$6 \times 10^{-3}$			
Folate (nmol/l)	8.9	0.47	$8 \times 10^{-4}$	2210	0.14	$8 \times 10^{-4}$

$\beta$ , Regression coefficient; SE, robust standard error; R<sup>2</sup>, regression coefficient of determination (i.e. proportion of the Cbl variance explained by the model).

\* The model of Cbl regression with age includes a quadratic age term (age<sup>2</sup>) to account for the curvilinear relationship between Cbl levels and age in the population.

† Physical activity is recorded as self-reported (yes/no) regular activity in the previous year.

‡ Alcohol consumption is modelled as 'never' or '1-5 units/week' v. '> 5 (5-40) units/week'.

§ Smoking status is modelled as 'never smoked' v. 'current' and 'ex-smoker' categories.

|| Social class is based on the UK General Registrar Classification with the categories of senior management (A) or professional (B), skilled non-manual (C1) or skilled manual (C2) v. semi- (D) or unskilled worker (E).

**Table 3.** Heritability of plasma cobalamin levels

Model*	$\chi^2$	df	$\Delta\chi^2$	$\Delta$ df	<i>P</i>	A	95 % CI	C/D	95 % CI	E	95 % CI
ACE	9.3	3	–	–	–	0.41	0.21, 0.57	0.10	0.00, 0.24	0.49	0.42, 0.57
ADE	11.4	3	–	–	–	0.53	0.39, 0.59	0.00	0.00, 0.15	0.46	0.40, 0.53
AE	11.4	4	2.1	1	0.15	0.53	0.47, 0.59	–	–	0.46	0.40, 0.53
CE	24.8	4	15.5	1	$8 \times 10^{-5}$	–	–	0.37	0.32, 0.42	0.63	0.58, 0.68
E	189.6	5	164.8	1	$1 \times 10^{-37}$	–	–	–	–	1	–

$\Delta\chi^2$ , change in  $\chi^2$  fit statistic or likelihood ratio test (LRT); A, additive polygenic variance component; C, shared familial or common environment; D, dominant genetic variance; E, unique environmental effects (including measurement error) that are specific to the individual.

\* Five models are presented, each with best-fit model statistics and standardised variance component estimates for A, C, D and E that add up to one for each model. The most parsimonious and best-fit model (AE) is highlighted, with a heritability estimate of 0.53 (95 % CI 0.47, 0.59) and the remaining variance (0.47) attributed to environmental factors specific to the individual. Model-fit comparisons are based on a LRT, where the model-fit contribution for each parameter (A, D and C) is assessed by individually fixing the value of each to zero to contrast the full model (ACE or ADE) with a nested sub-model (AE, CE and E). If the LRT is significant (here assessed by the change in  $\chi^2$  with one df), the variance component contributes to the model fit and cannot be dropped. Common environment (C) and dominant genetic (D) effects are confounded using twin heritability models, so these components cannot be estimated in the same model.

53% of the variation in plasma Cbl levels accounted for by additive genetic factors (i.e. heritability). The remainder of the variance was accounted for by environmental factors unique (non-shared) to each twin. Given the relative lack of power to detect shared familial effects in twin studies *per se* (21), however, the present data also provide evidence that up to 10% of the variance in plasma Cbl levels is attributable to shared familial environmental factors, when an ACE model is fitted, with the ACE model including additive polygenic (A), common (C) and unique (E) environmental effects.

Given the observation that the mean and variance for the Cbl increase with age, we also fitted heritability models stratified by three age groups (under 40, 40–55 and over 55 years). The DeFries–Fulker regression-based models suggest that the genetic variance of Cbl increases with age with heritability for each group being 18% (ACE model), 48% (AE model) and 58% (AE model), for the under 40-, 40–55- and over 55-year age groups, respectively (Table 4). It should be noted that only the under 40-year age group provided good evidence for a model that includes a large shared familial component (ACE) estimated to account for 30% of the total phenotypic variation in plasma Cbl for this age group. The middle-aged and elderly

group best-fit models did not provide evidence for shared familial components (with the best-fit model for both these age groups being the AE models).

### Impact of inborn errors of cobalamin metabolism genes on constitutive plasma cobalamin levels

**Tagging efficiency.** We formally quantified how well HapMap common SNP, with a minor allele frequency  $\geq 10\%$ , were captured and represented by our sample genotyped SNP based on the degree of correlation ( $r^2$ ) between each SNP pair (namely tagged and tagging SNP, respectively). Based on a minimum  $r^2$  of 0.8, Table 5 shows that our genotyped SNP provided a reasonable representation for most of the genes with the percentage of common SNP tagged ranging from 40 to 92%. For gene regions *MMACHC* and *MMAA* (methylmalonic aciduria cblC type, with homocystinuria and methylmalonic aciduria (cobalamin deficient) cblA type), the proportion of SNP not represented was high with half or more of the common variants not adequately captured. Excluding these two gene regions, the mean proportion of common variants captured by our panel of genotyped markers was close to 80%.

**Table 4.** Change in the heritability of plasma cobalamin levels by age

Age group*	<i>n</i>	A	95 % CI	C/D	95 % CI	Fit statistic (adj $R^2$ )
<40 years	767					
ACE		0.18	–0.24, 0.61	0.30	0.02, 0.58	0.174
CE				0.42	0.33, 0.51	0.173
AE		0.62	0.48, 0.76			0.165
40–55 years	778					
ACE		0.68	0.20, 1.15	–0.13	–0.43, 0.17	0.092
CE				0.28	0.18, 0.37	0.074
AE		0.48	0.33, 0.63			0.091
>55 years	702					
ACE		0.52	0.06, 0.98	0.04	–0.26, 0.34	0.143
CE				0.37	0.27, 0.47	0.132
AE		0.58	0.43, 0.73			0.144

*n*, Number of participants; A, additive polygenic effects; C, shared familial or common environmental effects; D, dominant genetic variance; E, unique environmental effects (including measurement error) that are specific to the individual; ACE, model including variance component (VC) estimates for additive polygenic, shared familial and unique environmental effects; CE, model including only VC for shared familial and unique environmental effects; AE, model including additive polygenic and unique environmental effects; adj  $R^2$ , adjusted  $R^2$  model best-fit statistic.

\* Three VC models are presented for each age group, each with best-fit statistics and VC estimates that add up to one. The best-fit model (based on both the adjusted  $R^2$  fit statistic and the model providing interpretable, non-negative VC estimates) is ACE for the under 40-year age group and AE for the middle-aged and older age groups. Heritability estimates are based on DeFries–Fulker regression analyses of twin data (14), in which unlike VC analyses, regression-based VC estimates are not constrained to lie between 0 and 1, but are freely estimated.

**Table 5.** SNP marker coverage for each of the eight candidate genes

Defects	Genes	Location	Start position	Gene size (kb)	No. of SNP genotyped*	No. SNP tagged ( $r^2 \geq 0.8$ )†	No. of SNP missed	% SNP tagged
cbIA	<i>MMAA</i>	4q31	146,539,415	42	2	4	4	0.50
cbIB	<i>MMAB</i>	12q24	109,991,520	20	4	12	1	0.92
cbIC	<i>MMACHC</i>	1p34	45,965,725	11	1	2	3	0.40
cbID	<i>MMADHC</i>	2q23	150,426,147	18	1	7	4	0.64
cbIE	<i>MTRR</i>	5p15	7,851,299	55	9	18	9	0.67
cbIF	<i>LMBRD1</i>	6q13	70,385,641	121	6	78	29	0.73
cbIG	<i>MTR</i>	1q43	236,958,581	109	8	112	32	0.78
mut	<i>MUT</i>	6p21	49,398,073	33	3	22	6	0.79

cbI, Cobalamin.

\* The number of genotyped SNP for each gene region.

† The degree to which these markers are pairwise correlated ( $r^2$ ) with all common SNP documented by the HapMap project for these regions.

**Genetic association analysis.** We tested whether common SNP in genes known to cause Mendelian defects in inborn errors of Cbl metabolism and transport are associated with plasma Cbl levels. A total of thirty-eight genotyped SNP – tagging a total of 255 common (minor allele frequency  $\geq 0.10$ ) known SNP (minimum  $r^2 \geq 0.8$ ) documented by HapMap – were identified in eight gene regions and tested for the association with plasma levels of Cbl in female Caucasian twins from the TwinsUK registry. It should be noted that for these analyses, we used only one individual from each MZ twin pair with the most complete phenotype data.

The results of the analyses are shown in Table 6. Common polymorphisms in the genes *MMAA*, *MMACHC*, *MTRR* (5-methyltetrahydrofolate-homocysteine methyltransferase) and *MUT* (methylmalonyl coenzyme A mutase) were observed to be nominally associated ( $\alpha = 0.05$ ) with plasma Cbl levels. *MMACHC* ( $P=0.001$ ) and *MUT* ( $P=0.0002$ ) both contained SNP associated with Cbl that passed the conservative Bonferroni significance threshold adjusted for multiple testing ( $\alpha' = 0.001$ ). Furthermore, multiple variants in two of the four genes, *MTRR* and *MUT*, showed evidence for the association with Cbl levels, strengthening the case for the importance of these genes in potentially driving population variability in Cbl levels. We failed to show the association between common variants in the genes *MMAB*, *MMADHC*, *LMBRD1* and *MTR* (methylmalonic aciduria (cobalamin deficiency) cbIB type, methylmalonic aciduria cbID type, with homocystinuria, LMBR1 domain containing 1 and methyltransferase reductase) and plasma levels of Cbl.

## Discussion

We aimed to gain a better understanding of the main genetic and environmental factors that affect the population variation in plasma Cbl levels using a classical twin model. We also tested Garrod's hypothesis that common, less penetrant genetic polymorphisms in genes that underlie inborn errors of Cbl metabolism and transport may be associated with normal population variability in plasma Cbl levels.

The present results verify that variations in plasma levels of Cbl are highly heritable. To our knowledge, this is the first study to report heritability for Cbl levels (estimated to be approximately 53%) for a female population. The only other report is from a much smaller study of ninety-six MZ and

120 DZ twin pairs conducted in elderly Swedish twins (age  $\geq 82$  years), which reported a heritability of 62% for serum Cbl levels<sup>(22)</sup>.

The difference between the heritability estimates between the present study and the Swedish study is likely to be attributable to the large, non-overlapping difference in age for the two samples, with all study subjects under 80 years of age

**Table 6.** Results from the candidate gene association study

Genes	Defect	Location	RefSNP ID	$\chi^2$	$P^*$			
<i>MMAA</i>	cbIA	4q31	rs4835011	0.63	0.43			
			rs4835012	4.71	0.03			
			rs11728906	3.08	0.08			
<i>MMAB</i>	cbIB	12q24	rs4835014	0.44	0.50			
			rs2287182	0	0.83			
			rs7134594	0.16	0.68			
			rs12314392	0.11	0.74			
			rs3759387	0	0.83			
			rs7957619	0.97	0.32			
<i>MMACHC</i>	cbIC	1p34	rs10789465	10.36	1E-03			
			<i>MMADHC</i>	cbID	2q23	rs7580915	0.05	0.82
						<i>MTRR</i>	cbIE	5p15
rs326120	2.74	0.10						
rs1532268	7.87	0.01						
			rs6555501	1.27	0.26			
			rs162031	2.11	0.14			
			rs162036	4.3	0.04			
			rs162040	0.81	0.37			
			rs162048	3.9	0.05			
			rs3776455	9.36	2E-03			
			rs10380	3.03	0.08			
<i>LMBRD1</i>	cbIF	6q13	rs6455338	0	0.83			
			rs3799105	0	0.83			
			rs3778241	1.46	0.22			
			rs9294851	0.02	0.88			
			rs991974	0.23	0.63			
			rs1457498	0.96	0.32			
<i>MTR</i>	cbIG	1q43	rs10733118	0.15	0.68			
			rs12060264	1.07	0.30			
			rs11800413	0.36	0.55			
			rs2789352	0.07	0.79			
			rs2275568	0.29	0.59			
			rs3768142	0.02	0.88			
			rs10925257	0.39	0.53			
			rs1805087	0.36	0.55			
<i>MUT</i>	mut	6p21	rs9381784	4.71	0.03			
			rs6458687	3.44	0.06			
			rs6458690	13.62	2E-04			

RefSNP ID, reference SNP ID number;  $\chi^2$ , chi-squared statistic; cbI, cobalamin.

\* Significant association with Cbl levels was based on a conservative Bonferroni corrected  $P=0.001$ .

for the TwinsUK and over 80 years of age for the Swedish study. The reported mean Cbl values for the two studies are also appreciably different with the TwinsUK value (overall female mean 422 pmol/l) up to 20% higher than those reported by the Swedish study (302 and 344 pmol/l for men and women, respectively). It is worth highlighting that the (1) difference in mean plasma Cbl levels between the two studies could be in part due to differences in the assays used to measure Cbl levels in each study, and (2) heritability estimates between the two studies diminished considerably when comparing the Swedish study with the results from the TwinsUK participants aged over 55 years. For this age group, the heritability estimate is 58%, which is comparable with the Swedish Cbl heritability estimate of 62%.

The present findings support data from the literature showing that plasma cbl levels are significantly correlated with age, BMI, alcohol use, smoking status, social class and folate levels, which in our cohort collectively explained up to 15% of the variation in Cbl levels. Importantly, the present results lend further support to the observational link between the increasing Cbl deficiencies in the elderly, which is commonly attributed to the increased prevalence of atrophic gastritis in this age group and shows both the prominence of genetic factors in driving increased variation in Cbl levels with age and the importance of shared environmental factors – for example, similar dietary habits in earlier life – for younger age groups. This highlights the need to better manage Cbl deficiencies in the elderly through an intervention programme that more specifically targets this stratum of the population.

Finally, the results from the present candidate gene study show the potential for studying genes that underlie rare, Mendelian diseases of Cbl metabolism and transport in the context of normal population variability in plasma Cbl levels. Out of the eight genes investigated, four – including *MMAA*, *MMACHC*, *MTRR* and *MUT* – harbour common polymorphisms that were either suggestive or significantly associated with Cbl levels. This underlines the continuing importance of formulating study designs to test hypotheses where a strong body of evidence already exists – for example, Mendelian conditions with pathophysiological markers – since the observed associations for *MMAA*, *MMACHC* and *MTRR* common polymorphisms in the present study would have probably been overlooked by Genome-wide Association Study (GWAS) studies. In the case of *MUT*, the strongest association in the present study is with SNP rs6458690 ( $P=0.0002$ ), which is in perfect linkage disequilibrium with two SNP ( $r^2=1$ ), rs9473558 and rs9473555. Both of these SNP in *MUT* have previously been shown to be strongly associated with plasma levels of Cbl<sup>(4)</sup> and tHcy levels<sup>(23)</sup>. Moreover, power constraints may have contributed to our failure to show the association between *MTR* and Cbl levels, as SNP rs2275565 in *MTR* is observed to be strongly associated with tHcy levels in larger studies (JBJ van Meurs, G Pare, SM Schwartz, *et al.* unpublished results). Although this SNP was not included in the genotype panel, it is estimated to be in strong pairwise linkage disequilibrium ( $r^2=0.86$ , HapMap CEU data) with two *MTR* SNP that are genotyped for the present study, rs10925257 and rs1805087.

The present results lend support to Garrod's prediction that rare, inborn errors of metabolism are extreme examples of what is commonly observed as minor, healthy variations in chemical behaviour. Inborn errors represent a large number of rare metabolic diseases that affect every major organ system with diverse manifestations. It has been estimated that inborn errors of metabolism can account for up to 15% of all single gene disorders in the population, representing a significant number of genes to be interrogated with suitable biochemical phenotypes, as illustrated in the present study. Once identified, this approach might facilitate the remedy for micronutrient imbalances through, for example, appropriate supplementation of the cognate marker<sup>(24)</sup>.

### Potential limitations

The present study has a few potential limitations, which warrant comment. First, although we aimed to include twins in the study who were deemed to be largely healthy (assessment through general health questionnaires, multiple clinical visits, not taking prescribed medication for chronic diseases), the present results may have nevertheless been influenced to an unknown extent by undocumented circumstances such as irregular diet, mild infection, early stages of a disease or a short course of an over-the-counter medication.

Another limitation was the sex constituency of the study participants. Due to historical reasons, TwinsUK is predominantly made up of female twin volunteers. Although we had data on Cbl levels for up to 200 male twins, we felt that the small sample size and the unbalanced design severely limited our ability to reliably comment either on men or sex differences, so as a result, we did not include the available males with Cbl measures in the present study. This is important in light of well-documented sex differences in constitutive tHcy levels, a marker that is directly influenced by Cbl levels.

The present results are also constrained by the power of the study. The present study had high power to detect low heritability and to test and quantify the effects of potential covariates consistent with previously published results in the literature. Nevertheless, the present genetic association study was underpowered for some of the candidate genes with poor marker coverage (e.g. for regions of low linkage disequilibrium including *MMAA*, *MMACHC*, *MMADHC* and *MTRR*) or the candidate genes harbouring associated variants, but with low penetrance (regardless of reasonable marker coverage). Similarly, the lack of a replication study/cohort is a limitation, which would reduce the potential for type I errors. Nevertheless, the fact that some of the present results replicate the findings from other studies adds confidence to our observations.

Finally, given the findings in the older age group, it would also be important to test the influence of the higher prevalence of impaired renal function and/or Cbl absorption on the present results. Similarly, it is important to validate the results observed for the present study using other biomarkers of Cbl status such as methylmalonic acid and holotranscobalamin.

In conclusion, the present results show that the observed population variability in Cbl levels is driven by a large



number of genetic and environmental factors that contribute differentially as the population ages. Interestingly, we showed that genetic factors are responsible for the increase in plasma Cbl variance with age. This novel observation warrants further investigation, as it may be important to know what genetic factors are driving the increase in phenotypic variance as the population ages. Finally, in line with Garrod's prediction, we observed that common genetic variants in genes that underpin rare Mendelian forms of inborn errors of the Cbl metabolic pathway are also associated with and contribute to the genetic basis for the variation in Cbl levels. While the importance of Cbl metabolic pathway genes in determining Cbl deficiency in old age remains to be determined, the present study does illustrate that at least some of these genes play a role in regulating Cbl levels in the general population.

### Acknowledgements

The authors would like to thank all the twins who participated in the study and acknowledge the dedication and support of the personnel at the Department of Twin Research and Genetic Epidemiology. T. A. acknowledges support by the Medical Research Council UK (Investigator Award 91993).

The authors acknowledge support from the Chronic Diseases Research Foundation, the Medical Research Council and the Wellcome Trust.

K. R. A. conceived the study. K. R. A. and T. A. designed and supervised the study. T. A. and K. R. A. conducted the research. R. G. and I. G.-N. were responsible for the recruitment and phenotypic data queries and management. T. A. analysed the data. K. R. A. and T. A. wrote the paper and had primary responsibility for the final content. All authors read and approved the final manuscript.

There are no conflicts of interest to declare.

### References

- Green R & Miller JW (2007) Vitamin B<sub>12</sub>. In *Handbook of Vitamin*, chapter 13 [J Zempleni, RB Rucker, DB McCormick and JW Suttie, editors]. Boca Raton, FL: CRC Press.
- Rosenblatt DS & Fenton WA (2001) Inherited disorders of folate and cobalamin transport and metabolism. In *The Metabolic and Molecular Basis of Inherited Metabolic Disease*, 8th ed., chapter 155, pp. 3897–3933 [CR, Scriver, AL, Beaudet, WS, Sly and D Valle, editors]. New York: McGraw-Hill.
- Hazra A, Kraft P, Selhub J, *et al.* (2008) Common variants of FUT2 are associated with plasma vitamin B<sub>12</sub> levels. *Nat Genet* **40**, 1160–1162.
- Hazra A, Kraft P, Lazarus R, *et al.* (2009) Genome-wide significant predictors of metabolites in the one-carbon metabolism pathway. *Hum Mol Genet* **18**, 4677–4687.
- Garrod A (1931) *The Inborn Factors in Disease. Facsimile of 1931 Edition*. New York: Oxford University Press.
- Coelho D, Suormala T, Stucki M, *et al.* (2008) Gene identification for the cblD defect of vitamin B<sub>12</sub> metabolism. *N Engl J Med* **358**, 1454–1464.
- Froese DS & Gravel RA (2011) Genetic disorders of vitamin B<sub>12</sub> metabolism: eight complementation groups—eight genes. *Expert Rev Mol Med* **29**, e37.
- Moayyeri A, Hammond CJ, Valdes AM, *et al.* (2012) Cohort Profile: TwinsUK and Healthy Ageing Twin Study. *Int J Epidemiol* **41**, 1–10.
- Beggs MJ, Sohn LJ, Herrmann RJ, *et al.* (1995) *Folate Immunoassay Utilizing Folate Binding Protein in a Multiclonal Antibody Format*. Abbott Park, IL: United States Abbott Laboratories.
- Ubbink J, Hayward Vermaak WJ & Bissbort S (1991) Rapid high performance liquid chromatographic assay for total homocysteine levels in human serum. *J Chromatogr* **565**, 441–446.
- Stata (2009) *Stata/MP 11 for Mac*. College Station, TX: StataCorp LP.
- Barber MJ, Cordell HJ, MacGregor AJ, *et al.* (2004) Gamma regression improves Haseman–Elston and variance components linkage analysis for sib-pairs. *Genet Epidemiol* **26**, 97–107.
- Neale M & Cardon L (1992) *Methodology for Genetic Studies of Twins and Families*. Dordrecht: Kluwer Academic Publishers.
- DeFries J & Fulker D (1985) Multiple regression analysis of twin data. *Behav Genet* **15**, 467–473.
- Weale ME (2010) Quality control for genome-wide association studies. *Methods Mol Biol* **628**, 341–372.
- Barrett JC, Fry B, Maller J, *et al.* (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265.
- Aulchenko YS, Ripke S, Isaacs A, *et al.* (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296.
- Falconer DS & Mackay TFC (1996) *Introduction to Quantitative Genetics*, 4th ed. Harlow, Essex: Addison Wesley Longman.
- Andrew T, Hart DJ, Snieder H, *et al.* (2001) Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res* **4**, 464–477.
- Baik HW & Russell RM (1999) Vitamin B<sub>12</sub> deficiency in the elderly. *Annu Rev Nutr* **19**, 357–377.
- Hopper JL (1999) Why “common environmental effects” are so uncommon in the literature. In *Advances in Twin and Sib-pair Analysis*, pp. 151–166. London: Oxford University Press.
- Nilsson SE, Read S, Berg S, *et al.* (2009) Heritabilities for fifteen routine biochemical values: findings in 215 Swedish twin pairs 82 years of age or older. *Scand J Clin Lab Invest* **69**, 562–569.
- Paré G, Chasman DI, Parker AN, *et al.* (2009) Novel associations of CPS1, MUT, NOX4, and DPEP1 with plasma homocysteine in a healthy population: a genome-wide evaluation of 13,974 participants in the Women's Genome Health Study. *Circ Cardiovasc Genet* **2**, 142–150.
- Ames BN, Elson-Schwab I & Silver EA (2002) High-dose vitamin therapy stimulates variant enzymes with decreased coenzyme binding affinity (increased K(m)): relevance to genetic disease and polymorphisms. *Am J Clin Nutr* **75**, 616–658.