Int J Mol Epidemiol Genet 2011;2(4):316-327
www.ijmeg.org /ISSN1948-1756/IJMEG1107004

Original Article

The causal roles of vitamin B_{12} and transcobalamin in prostate cancer: can Mendelian randomization analysis provide definitive answers?

Simon M Collin[1], Chris Metcalfe[1], Tom M Palmer[1,2], Helga Refsum[3,4], Sarah J Lewis[1], George Davey Smith[1,2], Angela Cox[5], Michael Davis[1], Gemma Marsden[6], Carole Johnston[3], J Athene Lane[1], Jenny L Donovan[1], David E Neal[7], Freddie C Hamdy[6], A David Smith[3], Richard M Martin[1,2]

[1]School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, UK; [2]MRC Centre for Causal Analysis in Translational Epidemiology (CAiTE), School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK; [3]Department of Pharmacology, University of Oxford, Le Gros Clark Building, South Parks Road, Oxford, OX1 3QX, UK; [4]Department of Nutrition, Institute of Basic Medical Sciences, Faculty of Medicine, University of Oslo, Oslo, Norway; [5]Institute for Cancer Studies, University of Sheffield, G Floor Medical School, Beech Hill Road, Sheffield, S10 2RX, UK; [6]Nuffield Department of Surgical Sciences, University of Oxford, John Radcliffe Hospital, Headley Way, Headington, Oxford, OX3 9DU, UK; [7]Department of Oncology, University of Cambridge, Box 279 (S4), Addenbrooke's Hospital, Cambridge, CB2 0QQ, UK

Received July 23, 2011; accepted August 23, 2011; Epub November 28, 2011; Published December 15, 2011

Abstract: Circulating vitamin B_{12} (cobalamin/B_{12}) and total transcobalamin (tTC) have been associated with increased and reduced risk, respectively, of prostate cancer. Mendelian randomization has the potential to determine whether these are causal associations. We estimated associations of single nucleotide polymorphisms in B_{12}-related genes (MTR, MTRR, FUT2, TCN2, TCN1, CUBN, and MUT) with plasma concentrations of B_{12}, tTC, holo-transcobalamin, holo-haptocorrin, folate, and homocysteine and with prostate cancer risk in a case-control study (913 cases, 895 controls) nested within the UK-wide population-based ProtecT study of prostate cancer in men age 45-69 years. Instrumental variable (IV) analysis was used to estimate odds ratios for effects of B_{12} and tTC on prostate cancer. We observed that B_{12} was lower in men with FUT2 204G>A (rs492602), CUBN 758C>T (rs1801222) and MUT 1595G>A (rs1141321) alleles (P_{trend}<0.001); tTC was lower in men with the TCN2 776C>G (rs1801198) allele (P_{trend}<0.001). FUT2 204G>A and CUBN 758C>T were selected as instruments for B_{12}; TCN2 776C>G for tTC. Conventional and IV estimates for the association of $\log_e(B_{12})$ with prostate cancer were: OR=1.17 (95% CI 0.90-1.51), P=0.2 and OR=0.60 (0.16-2.15), P=0.4, respectively. Conventional and IV estimates for the association of $\log_e$(tTC) with prostate cancer were: OR=0.81 (0.54-1.20), P=0.3 and OR=0.41 (0.13-1.32), P=0.1, respectively. Confidence intervals around the IV estimates in our study were too wide to allow robust inference. Sample size estimates based on our data indicated that Mendelian randomization in this context requires much larger studies or multiple genetic variants that explain all of the variance in the intermediate phenotype.

Keywords: Vitamin B_{12}/cobalamin, transcobalamin, prostate cancer, Mendelian Randomization

Introduction

Circulating concentrations of vitamin B_{12} have been associated with an increased risk of prostate cancer in a meta-analysis based on 6 studies, including a case-control study nested within the UK-wide, population-based Prostate testing for cancer and Treatment (ProtecT) study of PSA -detected prostate cancer (pooled OR=1.10; 95% CI 1.01, 1.19 per 100 pmol/L increase in B_{12}; P=0.002) [1]. Haptocorrin and transcobalamin are B_{12} transport proteins to which total circulating B_{12} is bound, as holo-haptocorrin and holo-transcobalamin respectively, in an approximate 80:20 ratio [2]. Holo-transcobalamin is the fraction of circulating B_{12} which is available for cellular uptake. Data from ProtecT revealed a positive association of holo-haptocorrin with prostate cancer risk (OR=1.21; 95% CI 1.01, 1.44; P=0.04), but no detectable evidence of an association of holo-transcobalamin with prostate cancer risk (OR=0.99; 95% CI 0.86, 1.14;

P=0.9) [1]. This suggests that the positive association of total B_{12} with prostate cancer risk could be due to reverse causation or confounding, rather than reflecting the action of bioavailable B_{12}. ProtecT data also suggested an inverse association of total transcobalamin (tTC) with prostate cancer risk (OR=0.76; 95% CI 0.55, 1.05; P=0.10).

Mendelian randomization (MR) uses genotypes as instrumental variables (IVs) to test and estimate the causal effect of phenotypes on disease-related outcomes, and can overcome unmeasured confounding [3, 4]. To be valid instruments the genotypes should fulfill the IV assumptions that: (a) the genotype is robustly associated with the phenotype; (b) the genotype is independent of factors confounding the association between the phenotype and outcome, which is in general justified by Mendel's laws [5]; (c) the genotype is associated with the outcome only through the intermediate phenotype, does not have a direct effect on the outcome and does not modify the effect of the phenotype on the outcome [3]. MR studies can require large samples to overcome imprecision in IV estimates arising from the combined variances of the genotype-phenotype and phenotype-outcome associations [6]. Here we use data from ProtecT to assess single-nucleotide polymorphisms (SNPs) in B_{12} and tTC -related genes as potential instrumental variables for determining causal associations of circulating concentrations of vitamin B_{12} and tTC with the risk of PSA-detected prostate cancer.

Materials and methods

The study population, ethical approval, blood sample handling, biochemical analyses, DNA extraction and genotyping are as previously described [1,7]. In brief, ProtecT (ISRCTN20141297) is a UK-wide, population-based study of PSA testing for prostate cancer with an embedded randomized controlled trial of treatments for localized prostate cancer. Between 1999 and 2009, men aged 45-69 years from 337 general practices around nine centres were invited to have a PSA test [8]. Participants with a PSA level ≥ 3.0ng/mL were invited for digital rectal examination and 10-core transrectal ultrasound-guided biopsy. A diagnosis of localized prostate cancer was defined as a positive biopsy, clinical stage T1-T2, NX, M0; advanced prostate cancer was defined as positive biopsy, clinical stage T3-T4 or N1 or M1.

The present study is based upon two case-control series: in the first (genotype series), blood samples from 1,607 cases (1,319 (82%) localized, 288 (18%) advanced) and 3,058 controls (including 1,203 'low PSA' (< 0.5ng/ml) controls that were selected for a genome-wide association study [9]) were genotyped for 13 folate and B_{12}-pathway SNPs [7]; in the second (phenotype series), blood samples from 1,461 cases (1,298 (89%) localized, 163 (11%) advanced) and 1,507 controls were assayed for plasma concentrations of folate, B_{12}, tTC, holotranscobalamin (hTC) and total homocysteine (tHcy), as described previously [1]. Holohaptocorrin (hHC) concentration was calculated by subtracting hTC concentration from total B_{12} concentration, hence does not include cobalamin analogues. In both case-control series, cases were selected at random from among all men diagnosed with localized or advanced cancer who had consented to a blood sample for research. Eligible controls were randomly-selected from men who had a PSA level < 3 ng/mL, or who had a PSA level ≥ 3 ng/mL and a negative biopsy result, and who had consented to provide a research blood sample. Controls were stratum-matched to cases by five-year age group and by the primary care practice from which they were recruited. Our MR analysis is based on 913 cases (809 (89%) localized, 104 (11%) advanced) and 895 controls with genotype and phenotype data (**Figure 1**). The B_{12} and tTC-related genes use in this analysis are described in **Table 1**.

Genotype-disease and genotype-phenotype associations

We tested whether genotype frequencies in controls in the genotype case-control series were in Hardy-Weinberg Equilibrium (HWE) using the Pearson Chi-squared test. We assessed genotype-disease associations in the genotype case-control series by estimating the odds of prostate cancer using logistic regression in an additive (per minor allele) effect model in which genotype was coded A/A=0 (reference), a/A=1, a/a=2. We used an additive effect model with a quadratic term for the minor allele to test whether there was any departure from a linear (additive) effect by observing the P-value and odds ratio (OR) for the quadratic term (OR<1 indicating a dominant effect; OR>1 a recessive

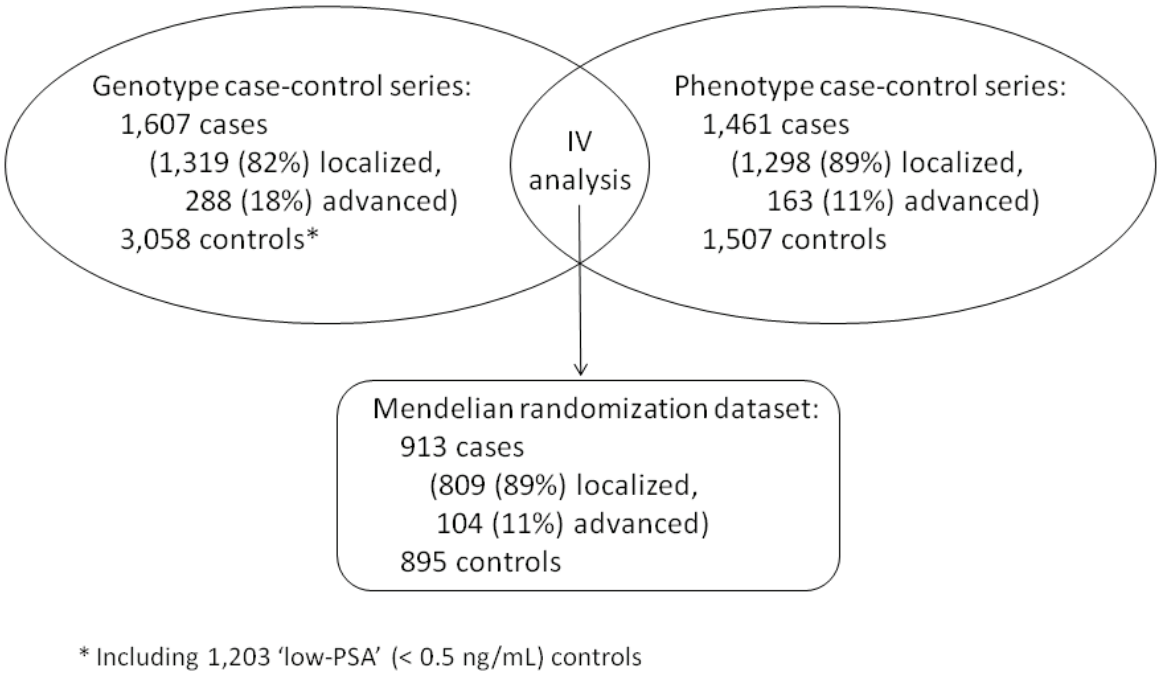

Figure 1. Selection of cases and controls from the ProtecT study for use in our Mendelian randomization analysis.

effect). Odds ratios for associations with advanced and localized cancer were compared using a multinomial logistic regression model. This model provides a statistical test for heterogeneity in odds ratios comparing associations of each genotype with localized vs advanced prostate cancers. We assessed genotype-phenotype associations by estimating per allele (additive) genotypic effects on median plasma vitamin and metabolite concentrations (and PSA levels) in controls using a non-parametric test-for-trend of medians.

IV analysis

F-statistics indicating the strength of a genotype as a potential instrument were estimated by linear regression of $\log_e$-transformed plasma concentrations of B_{12} and tTC in cases and controls against genotype in a weighted additive model. Weights were created by 5-year age group (45-49, 50-54, 55-59, 60-64, 65-69) and study centre to map cases and controls back to their proportions in the population. IV estimates based on weak instruments ($F < 10$) will be biased towards the conventional (confounded) phenotype-outcome association [6]. We assessed the suitability of each SNP as an instrumental variable by: i) considering the F-statistic for each SNP (to ensure a sufficiently strong instrument); ii) eliminating SNPs which were strongly associated with folate or homocysteine (to avoid SNPs with effects on prostate cancer via pathways independent of B_{12} and tTC (horizontal pleitropy); iii) testing for interaction between each SNP and plasma B_{12} or tTC in logistic regression models with prostate cancer as the outcome. IV analyses of the effects of B_{12} and tTC on prostate cancer risk were performed using two consecutive stages of regression [10, 11]. In the first stage, we obtained residuals from the weighted regression of $\log_e$-transformed circulating B_{12} or tTC concentration on the genotype. The residuals estimate the variation in levels of circulating B_{12} or tTC attributable to unmeasured, and potentially confounding, factors. In the second stage, we used a logistic regression model incorporating $\log_e$-transformed B_{12} or tTC concentration and the respective first-stage residual to estimate (with robust standard errors) causal odds ratios for prostate cancer per $\log_e$ B_{12} or tTC concentration.

We estimated the IV analysis sample size required to achieve 90% power (5% level of signifi-

cance) to detect causal associations of B_{12} or tTC (exposure, denoted "X") with prostate cancer (outcome, denoted "Y") using a single instrument (genotype, denoted "G", coded as a recessive or dominant binary effect) using the following method:

1) calculate genotype-outcome effect with **Equation 1;**

$$OR_{GY} = OR_{XY}^{\delta/\Delta} \qquad \textbf{Eq. 1}$$

2) calculate genotype frequency in cases with **Equation 2**:

$$f_1 = f_0 OR_{GY} \big/ \left(1 - f_0 + f_0 OR_{GY}\right) \qquad \textbf{Eq. 2}$$

3) use f_0 and f_1 to estimate sample size using the Stata command sampsi (where OR_{XY} is the odds ratio for a given increase Δ in exposure, δ is the difference in mean exposure between the two genotypes, and f_0 is the frequency of the rarer genotype in controls). This method is based on the Wald-type estimate of the causal odds ratio [12] and assumes that, for a sufficiently strong instrument, uncertainty in the genotype-phenotype association can be ignored relative to uncertainty in the genotype-outcome association. All statistical analyses were performed using Stata Release 11 (StataCorp. 2009, College Station, TX).

Results

Associations of B_{12}-related polymorphisms with plasma concentrations of total B_{12}, total transcobalamin, holo-transcobalamin, holo-haptocorrin, folate, and homocysteine

Three B_{12}-related SNPs were strongly associated (P_{trend}<0.001) with lower plasma concentrations of B_{12}: FUT2 204G>A (rs492602); CUBN 758C>T (rs1801222); and MUT 1595G>A (rs1141321) (**Table 2**). FUT2 204G>A (rs492602) was strongly associated with reduced hHC (P_{trend}<0.001) but had no detectable effect on hTC, whereas CUBN 758C>T and MUT 1595G>A were strongly associated (P_{trend}=0.004) with reduced hHC and reduced hTC. FUT2 204G>A was weakly associated with increased folate (P_{trend}=0.05) and reduced tTC (P_{trend}=0.05); MUT 1595G>A was strongly associated with increased tHcy (P_{trend}=0.003).

TCN2 776C>G (rs1801198) was strongly associated with reduced plasma tTC and hTC levels (P_{trend}<0.001) and TCN1 372T>C (rs526934) was strongly associated with reduced plasma hHC levels (P_{trend}<0.001) and weakly associated with reduced total B_{12} (P_{trend} = 0.03). MTR 2756A>G (rs1805087) was weakly associated with reduced tHcy (P_{trend}=0.04); this SNP did not affect levels of B_{12}, but it was weakly associated with increased hTC (P_{trend}=0.07). None of the polymorphisms were associated with PSA levels in controls.

Associations of B_{12}-related polymorphisms with prostate cancer risk

All SNPs were in HWE in controls. None of the SNPs had an effect on risk of prostate cancer (**Table 3**), and there were no departures from a linear (per minor allele) effect. There were no differences in associations of any of the SNPs with advanced vs localized prostate cancer other than borderline heterogeneity ($P_{heterogeneity}$=0.05) for MUT 1595G>A ($OR_{localized}$=1.04; 95% CI 0.95, 1.15; P=0.4 vs $OR_{advanced}$=0.86; 95% CI 0.72, 1.04; P=0.1).

Assessment of IV assumptions

Of the three SNPs associated with B_{12}, FUT2 204G>A (F=29.8) and CUBN 758C>T (F=13.1) were deemed eligible as instrumental variables. MUT 1595G>A (F=13.9) was strongly associated with increased tHcy (P_{trend}=0.003) which could in turn be causally associated with prostate cancer on a pathway independently from that directly linking B_{12} with prostate cancer. Hence, we decided that MUT 1595G>A could not be used as an instrument for the association of B_{12} with prostate cancer. TCN2 776C>G was the only SNP strongly associated with plasma tTC (F=122.1) and was not associated with any other phenotype except hTC (F=16.2). There were no interactions between FUT2 204G>A and B_{12} ($P_{heterogeneity}$=0.5), CUBN 758C>T and B_{12} ($P_{heterogeneity}$=0.4), or TCN2 776C>G and tTC ($P_{heterogeneity}$=0.9).

IV estimates

IV estimates for the causal effects of B_{12} and tTC on prostate cancer risk using FUT2 204G>A,

Table 1. Single nucleotide polymorphisms investigated as potential instrumental variables for this study

Gene	Name	Variant	SNP ID	Chr	Residue change	Phenotype & disease associations
MTR	5-methyltetrahydrofolate-homocysteine methyltransferase	*2756A>G*	rs1805087	1q43	D919G	Previously named "methionine synthase", this enzyme catalyzes the de-methylation of 5-methyl-tetrahydrofolate to tetrahydrofolate and the re-methylation of homocysteine to methionine, using methyl-cobalamin(III) as an intermediate carrier of the methyl group. The *2756A>G* polymorphism occurs in the activation domain of *MTR* which contains the binding site for S-adenosylmethionine (AdoMet), in close proximity to the adjacent cobalamin-binding domain [29]. Inconsistent evidence of associations of *MTR A2756G* with phenotypes and cancers [30].
MTRR	5-methyltetrahydrofolate-homocysteine methyltransferase reductase	*66A>G*	rs1801394	5p15.3	I22M, I49M	Previously named "methionine synthase reductase", this enzyme re-methylates inactive cobalamin (II) to the active methyl-cobalamin(III) form using a methyl group donated by AdoMet. Inconsistent evidence of associations of *MTRR A66G* with phenotypes and cancers.
FUT2	Fucosyltransferase 2	*204A>G*	rs492602	19q13	A68A	One of two genes encoding the galactoside 2-L-fucosyltransferase enzyme, a Golgi stack membrane protein involved in the creation of a precursor of the H antigen, which is required for the final step in the soluble A and B antigen synthesis pathway. The *204G* allele is associated with higher circulating concentrations of B_{12}, possibly due to its strong linkage disequilibrium with the FUT2 W154X polymorphism that determines gastrointestinal mucosal ABO antigen secretor status [16,25].
TCN2	Transcobalamin II	*776C>G*	rs1801198	22q11	R259P, R232P	Encodes transcobalamin (TC, TCII), a transport protein to which approximately 20% of B_{12} is bound in circulation, representing the fraction of B_{12} that is available for uptake by cells. Although TCN2 is widely expressed, there is some evidence that the vascular endothelium is a primary source [31]. The *776G* allele is associated with lower circulating concentrations of TC [32].
TCN1	Transcobalamin I	*372C>T*	rs526934	11q11		Encodes haptocorrin (HC, R-binder, TCI), a glycoprotein secreted by salivary glands and parietal cells to which approximately 80% of B_{12} is bound in circulation [33]. *TCN1 372C>T* has been associated with lower circulating concentrations of HC and B_{12} [25,34].
CUBN	Cubilin	*758C>T*	rs1801222	10p12	F253S	CUBN functions as a receptor of B_{12} bound to intrinsic factor (IF), facilitating endocytic absorption of the B_{12}-IF complex by intestinal ileal mucosal cells [27]. *CUBN 758C>T* has been associated with lower circulating concentrations of B_{12} [25].
MUT	Methylmalonyl Co-enzyme A mutase	*1595G>A*	rs1141321 (rs9473558)	6p21	R532H	MUT is a mitcochondrial enzyme that converts methylmalonyl Co-enzyme A (CoA) to succinyl-CoA using adenosylcobalamin as co-factor. *MUT 1595G>A* has been associated with lower circulating concentrations of B_{12} [25] and with higher levels of tHcy [28,25].

Table 2. Total B_{12}, bound B_{12}, total transcobalamin, folate and total homocysteine concentrations [median (5^{th} - 95^{th} percentile)] in co

Gene	Variant	SNP	Allele	Allele frequency (%)	Total B_{12} (pmol/L)	Holo-haptocorrin (pmol/L)	Holo-transcobalamin (pmol/L)	Total-transcobalamin (pmol/L)	Folate (nmol/L)	Total homocysteine (µmol/L)
MTR	*2756A>G*	rs1805087	A/A	65.6%	300 (161 – 538)	245 (121 - 446)	56 (22 – 131)	871 (602 – 1274)	16.6 (6.6 – 53.3)	11.1 (7.7 – 17.9)
			A/a	30.4%	305 (171 – 494)	247 (136 - 423)	58 (25 – 117)	907 (614 – 1298)	15.8 (6.2 – 50.8)	10.8 (7.7 – 17.3)
			a/a	4.0%	300 (169 – 631)	248 (122 - 401)	63 (38 – 151)	852 (594 – 1384)	18.5 (5.8 – 44.4)	9.5 (7.3 – 21.0)
			P*	-	0.3	0.7	0.07	0.2	0.7	0.04
MTRR	*66A>G*	rs1801394	A/A	31.9%	320 (161 – 535)	252 (122 - 443)	57 (24 – 125)	870 (599 – 1317)	17.2 (6.5 – 53.6)	11.0 (7.7 – 17.5)
			A/a	49.9%	294 (162 – 525)	245 (124 - 457)	56 (22 – 124)	899 (614 – 1298)	15.0 (6.1 – 50.8)	11.2 (7.7 – 18.7)
			a/a	18.2%	292 (165 – 554)	246 (125 - 430)	59 (20 – 135)	865 (590 – 1256)	18.8 (6.4 – 53.4)	10.6 (7.1 – 17.5)
			P*	-	0.2	0.4	0.8	0.7	0.8	0.6
FUT2	*204G>A*	rs492602	A/A	25.5%	329 (194 – 576)	286 (150 - 513)	57 (22 – 130)	886 (614 – 1274)	16.5 (6.5 – 45.0)	11.1 (7.8 – 17.4)
			A/a	49.6%	300 (163 – 514)	240 (123 - 417)	57 (23 – 125)	893 (612 – 1317)	16.8 (6.5 – 52.0)	10.9 (7.5 – 17.9)
			a/a	24.9%	286 (159 – 473)	230 (119 - 382)	56 (22 – 119)	864 (595 – 1238)	17.8 (6.4 – 54.3)	11.0 (7.8 – 17.4)
			P*	-	<0.001	<0.001	0.8	0.05	0.05	0.5
TCN2	*776C>G*	rs1801198	A/A	30.7%	296 (162 - 516)	231 (125 - 410)	63 (24 - 128)	983 (685 - 1462)	16.3 (6.4 - 51.5)	11.1 (7.9 - 17.7)
			A/a	49.7%	313 (164 - 536)	250 (126 - 446)	56 (23 - 124)	859 (610 - 1243)	15.7 (6.0 - 50.3)	11.2 (7.7 - 18.3)
			a/a	19.6%	315 (158 - 541)	258 (121 - 463)	53 (18 - 101)	782 (564 - 1095)	16.7 (6.7 - 54.2)	11.0 (7.7 - 17.5)
			P*	-	0.2	<0.001	<0.001	<0.001	0.5	0.8
TCN1	*372C>T*	rs526934	A/A	49.9%	308 (173 - 536)	250 (138 - 452)	56 (24 - 116)	873 (599 - 1278)	16.7 (6.4 - 51.5)	11.1 (7.7 - 17.9)
			A/a	43.0%	301 (160 - 508)	235 (125 - 411)	60 (24 - 131)	886 (619 - 1308)	16.5 (6.4 - 52.2)	10.8 (7.7 - 17.4)
			a/a	7.1%	283 (151 - 548)	214 (110 - 417)	61 (16 - 123)	880 (602 - 1215)	16.6 (6.1 - 56.4)	11.1 (7.7 - 18.7)
			P*	-	0.03	<0.001	0.10	0.9	0.5	0.8
CUBN	*758C>T*	rs1801222	A/A	40.3%	323 (164 - 550)	254 (125 - 479)	59 (22 - 131)	876 (615 - 1314)	16.3 (6.1 - 53.1)	11.1 (7.7 - 17.6)
			A/a	47.2%	304 (165 - 528)	243 (130 - 441)	57 (22 - 116)	869 (609 - 1294)	16.0 (6.4 - 48.7)	11.1 (7.7 - 18.0)
			a/a	12.5%	291 (143 - 496)	231 (110 - 406)	50 (21 - 100)	875 (612 - 1288)	16.9 (6.1 - 54.2)	11.5 (7.8 - 18.5)
			P*	-	<0.001	0.003	<0.001	1.0	0.9	0.4
MUT	*1595G>A*	rs1141321	A/A	41.1%	323 (164 - 550)	256 (127 - 452)	60 (24 - 124)	881 (609 - 1293)	16.8 (6.1 - 51.3)	11.0 (7.8 - 17.6)
			A/a	45.4%	301 (159 - 524)	242 (122 - 430)	56 (22 - 118)	869 (612 - 1312)	16.2 (6.2 - 51.2)	11.1 (7.7 - 18.0)
			a/a	13.4%	293 (151 - 522)	236 (128 - 437)	54 (20 - 121)	875 (597 - 1272)	14.8 (5.9 - 56.1)	11.8 (8.2 - 17.9)
			P*	-	<0.001	0.004	<0.001	0.2	0.06	0.003

*non-parametric test-for-trend

Table 3. Odds ratios for prostate cancer by genotype (additive effect model)*

Gene	Variant	SNP	All cases *vs* controls	Localized cases *vs* controls	Advanced cases *vs* controls	$P_{heterogeneity}$**
MTR	*2756A>G*	rs1805087	1.01 (0.90, 1.12), P=0.9	1.00 (0.89, 1.12), P=1.0	1.04 (0.84, 1.29), P=0.7	0.8
MTRR	*66A>G*	rs1801394	0.94 (0.86, 1.03), P=0.2	0.93 (0.84, 1.02), P=0.1	1.01 (0.85, 1.21), P=0.9	0.3
FUT2	*204A>G*	rs492602	0.98 (0.90, 1.07), P=0.7	0.98 (0.89, 1.07), P=0.6	1.01 (0.85, 1.20), P=0.9	0.7
TCN2	*776C>G*	rs1801198	1.06 (0.98, 1.16), P=0.2	1.07 (0.98, 1.17), P=0.2	1.08 (0.91, 1.29), P=0.4	0.8
TCN1	*372C>T*	rs526934	0.95 (0.86, 1.05), P=0.3	0.96 (0.87, 1.07), P=0.5	0.92 (0.75, 1.12), P=0.4	0.7
CUBN	*758C>T*	rs1801222	0.97 (0.89, 1.06), P=0.5	0.96 (0.87, 1.05), P=0.4	1.03 (0.86, 1.24), P=0.7	0.4
MUT	*1595G>A*	rs1141321	1.01 (0.92, 1.10), P=0.9	1.04 (0.95, 1.15), P=0.4	0.86 (0.72, 1.04), P=0.1	0.05

*Genotype dataset comprises 1,607 cases (1,319 localised, 288 advanced) and 3,058 controls (including 1,203 low PSA controls)

**Wald test comparing coefficients for localized and advanced prostate cancer estimated by multinomial logistic regression

Table 4. Odds ratios (OR) for prostate cancer per unit increase in log_e-transformed plasma concentrations of vitamin B_{12} and total transcobalamin by conventional and instrumental variable (IV) analyses

Phenotype	Analysis	OR (95% CI)	P-value
Vitamin B_{12}	Conventional*	1.17 (0.90, 1.51)	0.2
	IV using *FUT2 204G>A*	0.96 (0.21, 4.42)	1.0
	IV using *CUBN 758C>T*	0.24 (0.03, 2.23)	0.2
	IV using both SNPs	0.60 (0.16, 2.15)	0.4
Total transcobalamin	Conventional*	0.81 (0.54, 1.20)	0.3
	IV using *TCN2 776C>G*	0.41 (0.13, 1.32)	0.1

*Conditional logistic regression matching on 5-year age group and recruiting centre, further adjusted for exact age

CUBN 758C>T and TCN2 776C>G, respectively, as instrumental variables are shown in **Table 4**. The effect sizes obtained by conventional logistic regression for the associations of B_{12} and tTC with risk of prostate cancer (OR=1.17; 95% CI 0.90, 1.51; P=0.2 and OR=0.81; 95% CI 0.54, 1.20; P=0.3) were similar in magnitude to the previously reported associations (OR=1.19; 95% CI 0.98, 1.43; P=0.08 and OR=0.76; 95% CI 0.55, 1.05; P=0.1) observed in a larger dataset [1], but are less precisely estimated. The IV estimates for B_{12} using FUT2 204G>A (OR=0.96; 95% CI 0.21, 4.42; P=1.0), CUBN 758C>T (OR=0.24; 95% CI 0.03, 2.23; P=0.2), and both SNPs together in a multiple IV analysis (OR=0.60; 95% CI 0.16, 2.15; P=0.4) are consistent with no causal effect (OR=1) and with the estimate obtained by the conventional analysis (OR=1.17). The IV estimate for tTC (OR=0.41; 95% CI 0.13, 1.32; P=0.1) is also consistent with the null and with the conventional estimate (OR=0.81).

Sample size estimates based on the pooled odds ratio from our previously reported meta-analysis of B_{12}-prostate cancer associations (OR_{XY}=1.10, 95% CI 1.01-1.19 per 100 pmol/L increase in B_{12}) [1] using a recessive effect of FUT2 204G>A on B_{12} (F_{stat}=14.2, δ=43.3 pmol/L, Δ=100 pmol/L, f_0=0.255) indicated that approximately 65,000 cases and the same number of controls would be needed to detect with 90% power (5% level of significance) the genotype-prostate cancer association that would be expected if the B_{12}-prostate cancer association were a true (unconfounded) causal effect. To detect an effect consistent with the upper bound of the 95% confidence interval (OR_{XY}=1.19) would require approximately 20,000 cases and 20,000 controls. The sample size in our study (approximately 1,000 cases and 1,000 controls) would have enabled us to detect with 90% power a doubling in the true causal odds of prostate cancer risk (OR=2.1 per 100 pmol/L increase in B_{12}) given the observed difference in mean B_{12} levels between genotypes of 43.3 pmol/L. In ProtecT controls, such an increase is roughly equivalent to 1 standard deviation (SD) of B_{12} levels (mean=315 pmol/L, SD=119 pmol/L). Sample size estimates based on the previously reported association of tTC with prostate cancer in ProtecT (OR_{XY}=0.91; 95% CI 0.83, 1.00 per SD) [1], using a dominant effect of TCN2 776C>G (F_{stat}=171.7, δ=-144 pmol/L, Δ=308 pmol/L, f_0=0.693), indicated that sample sizes for 90% power of approximately 51,000 cases and 51,000 controls would be required to substantiate the point estimate, and 13,000 cases and 13,000 controls to substantiate the lower bound of the 95% confidence interval.

Discussion

We have replicated several strong genotype-phenotype effects which may be of importance in future research into associations of blood levels of B_{12} and tTC with diseases in humans. Although the results of our IV analysis were insufficiently precise to provide a definitive answer as to the roles of B_{12} and tTC in prostate cancer, we were able to estimate the sample sizes that would be required to obtain precise IV estimates based on observed genotype-phenotype and phenotype-disease associations. As previously reported [13, 6, 14], these tend to be very large unless the phenotype-disease association is of a clinically significant magnitude, e.g. a doubling in risk of prostate cancer per 100 pmol/L increase in circulating B_{12}, or multiple genetic markers that explain all of the variance in the intermediate phenotype are used [15].

The precise mechanism by which genetic variation in FUT2 affects circulating concentrations of B_{12} has not been established. The FUT2 W154X (rs601338) SNP with which FUT2 204G>A (rs492602) is in perfect linkage disequilibrium [16] determines gastrointestinal mucosal ABO antigen secretor status [17, 18]. Secretor vs non-secretor status has been associated with increased susceptibility to Helicobacter pylori infection [19] and Crohn's Disease [20] possibly leading to gastric atrophy and reduced intestinal absorption of B_{12}. The weak associations of FUT2 204G>A with increased plasma folate and reduced total transcobalamin are unlikely to alter the validity of this polymorphism as an instrument for circulating concentrations of B_{12}. We excluded MUT 1595G>A as an instrument for B_{12} on the basis of its strong positive association with tHcy, although this association is plausibly on the same genotype-phenotype-cancer pathway as B_{12} (vertical pleitropy) rather than being on an alternative pathway (horizontal pleitropy). Adding MUT 1595G>A to the two-SNP IV analysis made little difference to the IV estimate (OR=0.48; 95% CI 0.16, 1.44; P=0.2).

Our IV analysis suggested an inverse association of tTC with risk of prostate cancer, although the confidence interval did not exclude a null effect. We have reported evidence of a recessive positive association of the TCN2 776G allele with prostate cancer risk (OR=1.15; 95% CI 0.99, 1.34; P=0.06) [7]. In the present analysis, this effect was not substantiated as a departure from a linear (per allele) effect (quadratic term OR=1.07; 95% CI 0.95, 1.21; P=0.3). However, it is consistent with our previously-reported inverse association of tTC with prostate cancer risk (OR=0.71; 95% CI 0.50, 0.99 per $\log_e$ concentration; P=0.05) [1], because circulating concentrations of tTC are lower with the TCN2 776G allele. We have also reported a positive association of the TCN2 776G allele with PSA velocity in men in the ProtecT study who were under active monitoring following a diagnosis of localized prostate cancer (per allele change in mean PSA velocity=0.61; 95% CI 0.08, 1.14 ng/mL/y; P=0.02) [21], which is consistent with our findings for TCN2 776G, tTC and risk of prostate cancer. The TCN2 776G allele was not associated with PSA levels in controls, hence we do not suspect bias due to men with this allele being more likely to proceed to biopsy in the ProtecT study. The TCN2 776G allele has been associated with higher gene promoter CpG island methylation in colorectal cancer [22] and with increased risk of colorectal adenoma [23], but not colorectal cancer [24]. However, tTC has rarely been studied in relation to cancer risk, and no plausible causal mechanisms linking transcobalamin with carcinogenesis were forthcoming from the literature. Replication of our finding in a much larger study is needed.

Our findings that CUBN 758C>T (rs1801222), MUT 1595G>A (rs1141321) and (less strongly) TCN1 372C>T (rs526934) were inversely associated with levels of B_{12} are the first reported replications of the original findings for these SNPs [25]; for FUT2 204G>A (rs492602) the second replication of its reported association with lower concentrations of circulating B_{12} [16,25]. We have already discussed the possible mechanism linking FUT2 with B_{12} absorption. That FUT2 204G>A was inversely associated with tTC may also be a consequence of gastric atrophy due to H pylori infection in ABO antigen secretors [26], but its positive association with folate is the opposite of observed effects in patients infected with H pylori [26]. CUBN functions as a receptor of B_{12} bound to intrinsic factor (IF), facilitating endocytic absorption of the B_{12}-IF complex by intestinal ileal mucosal cells [27]. The effect on this function of the CUBN 758C>T polymorphism was less dramatic than the effect of the FUT2 204G>A polymorphism and CUBN 758C>T was of less utility as an instrument for circulating B_{12}. We found novel associations of CUBN 758C>T, MUT 1595G>A and TCN2 776C>G with holo-haptocorrin and holo-transcobalamin, and of FUT2 204G>A and TCN1 372C>T with holo-haptocorrin. These findings are unsurprising given the various associations of these SNPs with B_{12}, transcobalamin and haptocorrin. The association of MUT 1595G>A with increased tHcy has been reported previously [28, 25]. MUT is a mitcochondrial enzyme that converts methylmalonyl Co-enzyme A (CoA) to succinyl-CoA using adenosylcobalamin as co-factor, and our findings suggest a mechanism whereby MUT 1595G>A causes depletion of circulating B_{12} leading to increased levels of tHcy, whereas CUBN 758C>T and FUT2 204G>A are associated with lower total B_{12} but do not affect tHcy. An inverse association of MTR 2756A>G with tHcy has been consistently reported in the literature; its positive association with holo-transcobalamin, although weak, is a novel finding.

Some caveats to Mendelian randomization must be considered, namely: unknown linkage disequilibrium between our SNPs and other causal variants; unknown horizontal pleitropy; and developmental changes that compensate for genetic variation (canalization) [3]. Our Mendelian randomization approach to investigating associations of plasma B_{12} and tTC with risk of prostate cancer was inconclusive. Our sample size estimates highlight the need for this approach to be based on much larger studies when effects are of modest magnitude or multiple genetic variants that explain all of the variance in the intermediate phenotype. The ProtecT genome-wide association study could be used as a first stage in identifying such variants, using levels of B_{12} and tTC as quantitative traits, although a replication sample would be needed. With these caveats in mind, we have identified three promising gene variants for Mendelian randomization studies of the role of B_{12} and tTC in prostate cancer aetiology.

Acknowledgements

This research was funded by World Cancer Research Fund (WCRF) UK (grant numbers

2004/18 and 2007/07). The National Cancer Research Institute (administered by the Medical Research Council) provided support for the development of the ProtecT epidemiological database through the ProMPT (Prostate Mechanisms of Progression and Treatment) collaborative. The ProtecT study is supported by the UK NIHR Health Technology Assessment Programme (projects 96/20/06, 96/20/99). Support for the ProtecT biorepository in Cambridge is provided by NIHR through the Biomedical Research Centre. DNA sample preparation in ProtecT was part-funded by a USA Department of Defense award to Angela Cox (award number W81XWH-04-1-0280). TP was supported by MRC Project Grant number G0601625. GDS and TP work within a centre, CAiTE, supported by the MRC (G0600705) and the University of Bristol. The authors would like to acknowledge the tremendous contribution of all members of the ProtecT study research group, and especially the following who were involved in this research: Prasad Bollina, Sue Bonnington, Deborah Cooper, Andrew Doble, Alan Doherty, Emma Elliott, David Gillatt, Phillipa Herbert, Peter Holding, Joanne Howson, Liz Down, Mandy Jones, Roger Kockelbergh, Howard Kynaston, Teresa Lennon, Norma Lyons, Hilary Moody, Philip Powell, Stephen Prescott, Liz Salter, Pauline Thompson. We would like to thank Cynthia Prendergast (University of Oxford) for her invaluable help with the biochemical analyses, Frank Windmeijer (University of Bristol) for his expert advice on instrumental variable analysis, and Roger Harbord (University of Bristol) for developing the method for calculating sample sizes in Mendelian randomization studies with dichotomous outcomes.

Address correspondence to: Dr. Collin SM, School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, UK Tel: 0117 3313323; Fax: 0117 3314088; E-mail: simon.collin@bristol.ac.uk

References

[1] Collin SM, Metcalfe C, Refsum H, Lewis SJ, Zuccolo L, Smith GD, Chen L, Harris R, Davis M, Marsden G, Johnston C, Lane JA, Ebbing M, Bonaa KH, Nygard O, Ueland PM, Grau MV, Baron JA, Donovan JL, Neal DE, Hamdy FC, Smith AD and Martin RM. Circulating folate, vitamin B12, homocysteine, vitamin B12 transport proteins, and risk of prostate cancer: a case-control study, systematic review, and meta-analysis. Cancer Epidemiol Biomarkers Prev 2010; 19: 1632-1642.

[2] Seetharam B, Yammani RR. Cobalamin transport proteins and their cell-surface receptors. Expert Rev Mol Med 2003; 5: 1-18.

[3] Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003; 32: 1-22.

[4] Lawlor DA, Harbord RM, Sterne JA, Timpson N and Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 2008; 27: 1133-1163.

[5] Davey Smith G. Mendelian Randomization for Strengthening Causal Inference in Observational Studies. Perspectives on Psychological Science 2010; 5: 527-545.

[6] Pierce BL, Ahsan H and Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. Int J Epidemiol 2011; 40: 740-752.

[7] Collin SM, Metcalfe C, Zuccolo L, Lewis SJ, Chen L, Cox A, Davis M, Lane JA, Donovan J, Smith GD, Neal DE, Hamdy FC, Gudmundsson J, Sulem P, Rafnar T, Benediktsdottir KR, Eeles RA, Guy M, Kote-Jarai Z, Morrison J, Al Olama AA, Stefansson K, Easton DF, and Martin RM. Association of Folate-Pathway Gene Polymorphisms with the Risk of Prostate Cancer: a Population-Based Nested Case-Control Study, Systematic Review, and Meta-analysis. Cancer Epidemiol Biomarkers Prev 2009; 18: 2528-2539.

[8] Donovan JL, Lane JA, Peters TJ, Brindle L, Salter E, Gillatt D, Powell P, Bollina P, Neal DE, and Hamdy FC. Development of a complex intervention improved randomization and informed consent in a randomized controlled trial. J Clin Epidemiol 2009; 62: 29-36.

[9] Eeles RA, Kote Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, Field HI, Southey MC, Severi G, Donovan JL, Hamdy FC, Dearnaley DP, Muir KR, Smith C, Bagnato M, rdern Jones AT, Hall AL, O'Brien LT, Gehr Swain BN, Wilkinson RA, Cox A, Lewis S, Brown PM, Jhavar SG, Tymrakiewicz M, Lophatananon A, Bryant SL, Horwich A, Huddart RA, Khoo VS, Parker CC, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Fisher C, Jamieson C, Cooper CS, English DR, Hopper JL, Neal DE and Easton DF. Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet 2008; 40: 316-321.

[10] Palmer TM, Thompson JR, Tobin MD, Sheehan NA and Burton PR. Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. Int J Epidemiol 2008; 37: 1161-1168.

[11] Blundell RW, Powell JL. Endogeneity in non-

parametric and semiparametric regression models. 2003; 312-357.
[12] Didelez V, Meng S and Sheehan NA. Assumptions of IV methods for observational epidemiology. Statist Sci 2010; 25: 22-40.
[13] Timpson NJ, Brennan P, Gaborieau V, Moore L, Zaridze D, Matveev V, Szeszenia-Dabrowska N, Lissowska J, Mates D, Bencko V, Foretova L, Janout V, Chow WH, Rothman N, Boffetta P, Harbord RM and Smith GD. Can lactase persistence genotype be used to reassess the relationship between renal cell carcinoma and milk drinking? Potentials and problems in the application of Mendelian randomization. Cancer Epidemiol Biomarkers Prev 2010; 19: 1341-1348.
[14] Schatzkin A, Abnet CC, Cross AJ, Gunter M, Pfeiffer R, Gail M, Lim U and Davey Smith G. Mendelian randomization: how it can-and cannot-help confirm causal relations between nutrition and cancer. Cancer Prev Res (Phila Pa) 2009; 2: 104-113.
[15] Davey Smith G. Use of genetic markers and gene-diet interactions for interrogating population-level causal influences of diet on health. Genes & Nutrition 2010; 1-17.
[16] Hazra A, Kraft P, Selhub J, Giovannucci EL, Thomas G, Hoover RN, Chanock SJ and Hunter DJ. Common variants of FUT2 are associated with plasma vitamin B12 levels. Nat Genet 2008; 40: 1160-1162.
[17] Ferrer Admetlla A, Sikora M, Laayouni H, Esteve A, Roubinet F, Blancher A, Calafell F, Bertranpetit J and Casals F. A natural history of FUT2 polymorphism in humans. Mol Biol Evol 2009; 26: 1993-2003.
[18] Kelly RJ, Rouquier S, Giorgi D, Lennon GG and Lowe JB. Sequence and expression of a candidate for the human Secretor blood group alpha (1, 2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. J Biol Chem 1995; 270: 4640-4649.
[19] Dierkes J, Ebert M, Malfertheiner P and Luley C. Helicobacter pylori infection, vitamin B12 and homocysteine. A review. Dig Dis 2003; 21: 237-244.
[20] McGovern DP, Jones MR, Taylor KD, Marciante K, Yan X, Dubinsky M, Ippoliti A, Vasiliauskas E, Berel D, Derkowski C, Dutridge D, Fleshner P, Shih DQ, Melmed G, Mengesha E, King L, Pressman S, Haritunians T, Guo X, Targan SR, and Rotter JI. Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. Hum Mol Genet 2010; 19: 3468-3476.
[21] Collin SM, Metcalfe C, Refsum H, Lewis SJ, Davey Smith G, Cox A, Davis M, Marsden G, Johnston C, Lane JA, Donovan J, Neal DE, Hamdy FC, Smith D, and Martin RM. Associations of folate, vitamin B12, homocysteine, and folate-pathway polymorphisms with prostate-specific antigen velocity in men with localized prostate cancer. Cancer Epidemiol Biomarkers Prev 2010; 19: 2833-2838.
[22] Hazra A, Fuchs CS, Kawasaki T, Kirkner GJ, Hunter DJ and Ogino S. Germline polymorphisms in the one-carbon metabolism pathway and DNA methylation in colorectal cancer. Cancer Causes Control 2010; 21: 331-345.
[23] Hazra A, Wu K, Kraft P, Fuchs CS, Giovannucci EL and Hunter DJ. Twenty-four non-synonymous polymorphisms in the one-carbon metabolic pathway and risk of colorectal adenoma in the Nurses' Health Study. Carcinogenesis 2007; 28: 1510-1519.
[24] Koushik A, Kraft P, Fuchs CS, Hankinson SE, Willett WC, Giovannucci EL and Hunter DJ. Non-synonymous polymorphisms in genes in the one-carbon metabolism pathway and associations with colorectal cancer. Cancer Epidemiol Biomarkers Prev 2006; 15: 2408-2417.
[25] Hazra A, Kraft P, Lazarus R, Chen C, Chanock SJ, Jacques P, Selhub J and Hunter DJ. Genome-wide significant predictors of metabolites in the one-carbon metabolism pathway. Hum Mol Genet 2009; 18: 4677-4687.
[26] Tamura A, Fujioka T and Nasu M. Relation of Helicobacter pylori infection to plasma vitamin B12, folic acid, and homocysteine levels in patients who underwent diagnostic coronary arteriography. Am J Gastroenterol 2002; 97: 861-866.
[27] Christensen EI, Birn H. Megalin and cubilin: multifunctional endocytic receptors. Nat Rev Mol Cell Biol 2002; 3: 256-266.
[28] Pare G, Chasman DI, Parker AN, Zee RR, Malarstig A, Seedorf U, Collins R, Watkins H, Hamsten A, Miletich JP and Ridker PM. 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 2009; 2: 142-150.
[29] Matthews RG, Sheppard C and Goulding C. Methylenetetrahydrofolate reductase and methionine synthase: biochemistry and molecular biology. Eur J Pediatr 1998; 157 Suppl 2: S54-9: S54-S59.
[30] Yu K, Zhang J, Zhang J, Dou C, Gu S, Xie Y, Mao Y and Ji C. Methionine synthase A2756G polymorphism and cancer risk: a meta-analysis. Eur J Hum Genet 2010; 18: 370-378.
[31] Regec A, Quadros EV, Platica O and Rothenberg SP. The cloning and characterization of the human transcobalamin II gene. Blood 1995; 85: 2711-2719.
[32] Gueant JL, Chabi NW, Gueant Rodriguez RM, Mutchinick OM, Debard R, Payet C, Lu X, Villaume C, Bronowicki JP, Quadros EV, Sanni A, Amouzou E, Xia B, Chen M, Anello G, Bosco P, Romano C, Arrieta HR, Sanchez BE, Romano A, Herbeth B, Anwar W and Namour F. Environmental influence on the worldwide prevalence

of a 776C->G variant in the transcobalamin gene (TCN2). J Med Genet 2007; 44: 363-367.
[33] Carmel R, Parker J and Kelman Z. Genomic mutations associated with mild and severe deficiencies of transcobalamin I (haptocorrin) that cause mildly and severely low serum cobalamin levels. Br J Haematol 2009; 147: 386-391.
[34] Tanaka T, Scheet P, Giusti B, Bandinelli S, Piras MG, Usala G, Lai S, Mulas A, Corsi AM, Vestrini A, Sofi F, Gori AM, Abbate R, Guralnik J, Singleton A, Abecasis GR, Schlessinger D, Uda M and Ferrucci L. Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations. Am J Hum Genet 2009; 84: 477-482.