Human Molecular Genetics, 2013, Vol. 22, No. 19 **3998–4006**
doi:10.1093/hmg/ddt239
Advance Access published on May 29, 2013

Genome-wide association study identifies loci affecting blood copper, selenium and zinc

David M. Evans[1,2,†], Gu Zhu[4,†], Veronica Dy[6], Andrew C. Heath[7], Pamela A. F. Madden[7], John P. Kemp[1,2], George McMahon[1,2], Beate St Pourcain[2], Nicholas J. Timpson[1,2], Jean Golding[1,2], Debbie A. Lawlor[1,2], Colin Steer[3], Grant W. Montgomery[5], Nicholas G. Martin[4], George Davey Smith[1,2,†] and John B. Whitfield[4,†,*]

[1]MRC Centre for Causal Analyses in Translational Epidemiology, [2]School of Social and Community Medicine and [3]Centre for Child and Adolescent Health, School of Social and Community Medicine, University of Bristol, Bristol, UK [4]Queensland Institute of Medical Research, Genetic Epidemiology, Locked Bag 2000 and [5]Queensland Institute of Medical Research, Molecular Epidemiology, Locked Bag 2000, Royal Brisbane Hospital, Herston, QLD 4029, Australia [6]Royal Prince Alfred Hospital, Sydney, Australia [7]Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA

Received January 9, 2013; Revised May 6, 2013; Accepted May 21, 2013

Genetic variation affecting absorption, distribution or excretion of essential trace elements may lead to health effects related to sub-clinical deficiency. We have tested for allelic effects of single-nucleotide polymorphisms (SNPs) on blood copper, selenium and zinc in a genome-wide association study using two adult cohorts from Australia and the UK. Participants were recruited in Australia from twins and their families and in the UK from pregnant women. We measured erythrocyte Cu, Se and Zn (Australian samples) or whole blood Se (UK samples) using inductively coupled plasma mass spectrometry. Genotyping was performed with Illumina chips and >2.5 m SNPs were imputed from HapMap data. Genome-wide significant associations were found for each element. For Cu, there were two loci on chromosome 1 (most significant SNPs rs1175550, $P = 5.03 \times 10^{-10}$, and rs2769264, $P = 2.63 \times 10^{-20}$); for Se, a locus on chromosome 5 was significant in both cohorts (combined $P = 9.40 \times 10^{-28}$ at rs921943); and for Zn three loci on chromosomes 8, 15 and X showed significant results (rs1532423, $P = 6.40 \times 10^{-12}$; rs2120019, $P = 1.55 \times 10^{-18}$; and rs4826508, $P = 1.40 \times 10^{-12}$, respectively). The Se locus covers three genes involved in metabolism of sulphur-containing amino acids and potentially of the analogous Se compounds; the chromosome 8 locus for Zn contains multiple genes for the Zn-containing enzyme carbonic anhydrase. Where potentially relevant genes were identified, they relate to metabolism of the element (Se) or to the presence at high concentration of a metal-containing protein (Cu).

INTRODUCTION

Trace amounts of many metallic elements are essential, while others are toxic; some can be both, depending on concentration. In either case, the reasons for variation between people are important. Because essential elements are obtained from the diet, their variation has usually been seen from an environmental perspective. However, in addition to differences in diet and other environmental exposure, genetic variation in absorption, metabolic transformation or storage can be important in determining individuals' risk of deficiency or toxicity.

Our previous twin studies have shown evidence for heritable variation in humans for the essential elements copper (Cu), selenium (Se) and zinc (Zn) and for the toxic ones arsenic, cadmium, lead and mercury. Linkage analysis identified chromosomal regions which may contain genes affecting concentrations of these elements in blood (1,2). We now extend the search for gene polymorphisms affecting essential or toxic

*To whom correspondence should be addressed at: Genetic Epidemiology, Queensland Institute of Medical Research, Locked Bag 2000, Royal Brisbane Hospital, Herston, QLD 4029, Australia. Tel: +61 733620229; Fax: +61 733620101; Email: john.whitfield@qimr.edu.au
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors and that the last two authors should be regarded as joint Last Authors.

element concentrations using genome-wide association study (GWAS). This paper concentrates on Cu, Se and Zn.

For Cu, intestinal absorption, tissue and subcellular distribution, and excretion in the bile are each subject to regulation (3,4). Incorporation of Cu into enzymes and transporters requires chaperone proteins (5,6). Excretion requires efflux transporters, and major defects in the Cu transporters ATP7A and ATP7B lead to deficiency (Menkes' disease, OMIM #309400) and overload (Wilson's disease, OMIM #277900), respectively. Experimental deficiency states are associated with fetal malformation and defects in neuronal development (7,8), while human studies have associated variation in Cu status with neurodegeneration (9,10) including a specific myelinopathy improved by Cu supplementation (11).

Control of absorption of Se is less critical because excretion occurs readily in both urine and faeces. In tracer studies, a high proportion of labelled Se is excreted within a few hours of administration (12). After absorption, selenoaminoacids, selenate and selenite are converted to selenide, which in turn gives rise to selenophosphate, selenocysteine-tRNA and selenoproteins. Excess Se is excreted as selenosugars, dimethylselenide or trimethylselonium ion (13). There is evidence for a number of diseases being associated with Se deficiency, including prostate cancer (14), cardiovascular disease (15), infertility and obstetric complications (16), and HIV progression and viral evolution (17,18).

Intestinal Zn absorption requires specific transporters, both for uptake into enterocytes and release to the circulation, and transporter expression is modified by Zn status (19,20). Zn is a constituent of a large proportion of enzymes, and free Zn is an important intracellular signalling molecule subject to subcellular compartmentalization maintained by specific transporters (21,22). Zn deficiency promotes oxidative stress and has been associated with cardiovascular disease and diabetes (23,24), immune dysfunction and infectious diseases (25,26), and male infertility (27). There is evidence for Zn affecting the processing and aggregation of amyloid precursor protein (28,29). Severe human Zn deficiency is found in association with genetic defects of *SLC39A4* encoding the transporter Zip4 (acrodermatitis enteropathica, OMIM #201100) or in extreme lack of available Zn in foods (30). Relative Zn deficiency may be more common, and the beneficial effects of Zn supplementation in human controlled trials (31) support this concept.

In summary, there are many gene products where variation in structure or expression could affect Cu, Se or Zn status, and thus affect health or risk of disease. We apply the hypothesis-free approach of genome-wide association to identify loci which affect Cu, Se or Zn status, using data from adult participants in community-based studies in Australia and the UK. This is, to our knowledge, the first GWAS of blood Cu, Se and Zn concentrations.

RESULTS

Overall results for the genome-wide association analyses are summarized in Figure 1. Details for genome-wide significant loci ($P < 5 \times 10^{-8}$) showing the single-nucleotide polymorphism (SNP) with the lowest *P*-value at each locus (in either cohort or in the combined data), are given in Table 1 and a list of suggestive loci (with lowest *P*-values between 5×10^{-6} and 5×10^{-8}) in Supplementary Material, Table S1. Association data for all SNPs can be downloaded from a link provided in the Supplementary Information. There were two significant loci for Cu (both on chromosome 1), one for Se on chromosome 5 (which was identified as significant in both the Australian and UK studies) and three for Zn (on chromosomes 8, 15 and X). Collectively, these significant loci accounted for 5% of the phenotypic variance for Cu, 4% for Se and 8% for Zn. Regional plots for each significant locus are shown in Figure 2.

Significant SNP associations for Se at the chromosome 5 locus extended over several genes, with $P < 10^{-10}$ in both the Australian (QIMR) and UK (ALSPAC) data across *DMGDH*, *BHMT2* and *BHMT*. The location of the most significant SNP differed between the cohorts (Supplementary Material, Fig. S1), being in *DMGDH* (rs921943) for ALSPAC and *BHMT* (rs7700970) for QIMR. Meta-analysis gave the strongest association for rs921943, and repeating the analysis with this SNP included as a covariate revealed a second significant effect peaking at rs506500 ($P = 4.51 \times 10^{-10}$ in ALSPAC, $P = 0.00057$ in QIMR, $P = 7.69 \times 10^{-12}$ combined). The two independent and significant effects at this locus are illustrated in Supplementary Material, Figure S2.

None of the genome-wide-significant SNPs listed in Table 1 for Cu, Se or Zn showed evidence of association with more than one of these elements, even at $P < 0.05$. However, the suggestive SNPs showed one locus which may affect both Cu and Zn. This was at *EPHA6* (EPH receptor A6), where two uncommon SNPs showed a suggestive result ($P < 5 \times 10^{-6}$) for one element and a *P*-value far lower than would be expected by chance for the other. Details are given in Supplementary Material, Table S2 and Figure S3.

Gene-based analysis showed no additional significant loci beyond those already identified by allelic association analysis of the SNP data.

DISCUSSION

The overall result of our GWAS approach to blood or erythrocyte concentrations of Cu, Se and Zn was the identification of significant loci for each. For Se, where we had data from both studies, the results are mutually reinforcing. This adds evidence to our previous conclusion (2), based on the classical twin method, that genetic variation plays a role in availability of these elements to the tissues for their physiological functions and potentially affects the risk of deficiency. The findings for each element will be considered in turn.

Copper

Two regions on chromosome 1 were found to have significant effects on Cu concentration in erythrocytes, as shown in Figures 1 and 2. At ~3.6 Mbp, a region of significant association included the genes *CCDC27*, *LOC388588* and *LRRC47*, with the regions of linkage disequilibrium extending to cover *KIAA0562* and *DFFB*. None of these genes codes for proteins with known functions or disease associations relevant to Cu metabolism. The second chromosome 1 locus ~149.6 Mbp contains *SELENBP1*, with significant associations extending (probably because of linkage disequilibrium) across *PSMB4* and *POGZ*.

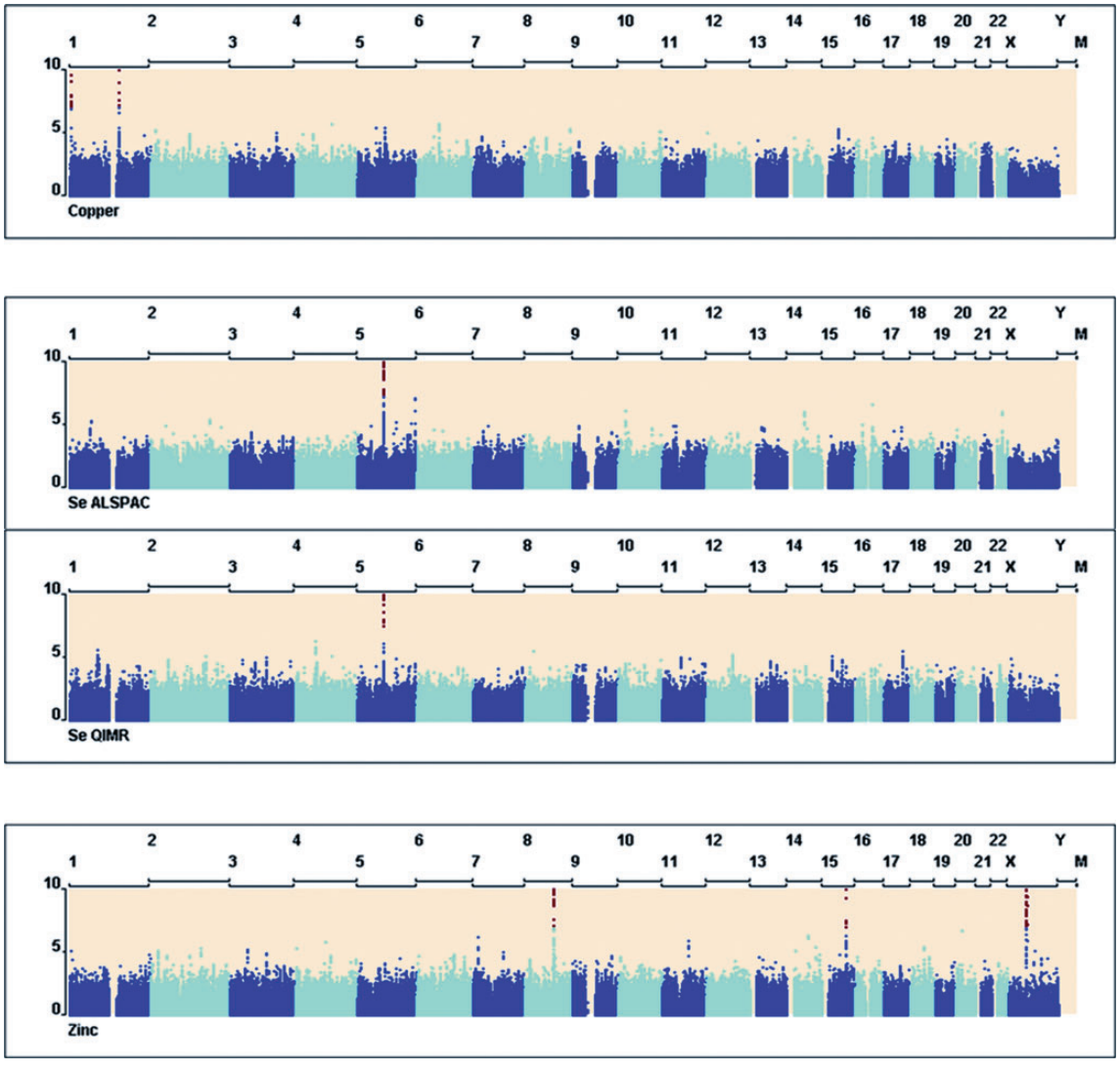

Figure 1. Manhattan plots for Cu, Se and Zn. Se results are shown for both QIMR and ALSPAC data, Cu and Zn for QIMR results only.

The latter two genes have no obvious connection with Cu status, while *SELENBP1* codes for selenium binding protein 1. This protein has no previously documented relationships with Cu, having been studied mainly as a tumour suppressor gene, but the highly significant association with erythrocyte Cu in our data ($P = 2.6 \times 10^{-20}$) makes it unlikely that this is a false positive result and the location of the most strongly associated SNP within an intron of *SELENBP1* makes this the prime candidate for mediating the association with Cu. Each of these chromosome 1 loci for Cu shows indications of regulatory activity in the ENCODE browser, including an H3K27Ac mark in the K562 erythroleukemia cell line. H3K27Ac marks (acetylation of lysine 27 of the H3 histone protein) may indicate sites of enhanced transcription, and the fact that this is found in the erythroleukemia cell line but not the others tested may mean that this is specific to the erythrocyte lineage. However, details of possible links between this regulatory function and erythrocyte Cu uptake and content are unknown.

Selenium

Significant associations between Se and a locus on chromosome 5 (Fig. 2) were present in both the Australian and UK data, although the location of the most significant SNP differed slightly (Supplementary Material, Fig. S1). The most significant SNP for Se in the meta-analysis was in *DMGDH*, coding for dimethylglycine dehydrogenase, although the region of significant hits also covered *BHMT2* and *BHMT* (coding for betaine homocysteine methyl transferases). A second, independent effect was revealed by conditional analysis and the most significant SNP was located within *BHMT*. While we cannot confidently ascribe these two effects to *DMGDH* and *BHMT*, respectively, the enzymes coded by the genes at this locus are all involved in metabolism of sulphur-containing amino acids (conversion of homocysteine to methionine) and may therefore play a role in selenoaminoacid metabolism. These enzymes are also involved in methylation and, as noted above, dimethylselenide and trimethylselonium

Table 1. Genome-wide-significant results ($P < 5.0 \times 10^{-8}$), from association analysis of data from 2603 Australian and 2874 British study participants

	SNP	Chr	BP[a]	Alleles		QIMR			ALSPAC[b]			Meta-analysis		
				Effect	Other	Beta	SE	*P*-value	Beta	SE	*P*-value	Beta	SE	*P*-value
Cu	rs1175550	1	3 681 388	A	G	−0.198	0.032	5.03×10^{-10}	–	–	–	–	–	–
	rs2769264	1	149 611 365	G	T	0.313	0.034	2.63×10^{-20}	–	–	–	–	–	–
Se[c]	rs921943	5	78 352 232	T	C	0.222	0.034	9.33×10^{-11}	0.264	0.030	1.43×10^{-18}	0.246	0.023	9.40×10^{-28}
	rs7700970	5	78 447 080	T	C	0.265	0.037	7.17×10^{-13}	0.173	0.032	4.11×10^{-8}	0.212	0.024	1.72×10^{-18}
Zn	rs1532423	8	86 455 565	A	G	0.178	0.026	6.40×10^{-12}	–	–	–	–	–	–
	rs2120019	15	73 121 237	C	T	−0.287	0.033	1.55×10^{-18}	–	–	–	–	–	–
	rs4826508	X	56 828 420	T	C	0.210	0.030	1.40×10^{-12}	–	–	–	–	–	–

The most significant SNP from each locus which shows $P < 5 \times 10^{-8}$, from QIMR, ALSPAC or meta-analysis of both, is shown.
[a]HapMap data Release 28 Phases II and III, August 2010, on NCBI B36 assembly, dbSNP b126.
[b]Cu and Zn were not measured in the ALSPAC samples.
[c]The most significant SNP for Se differed between the two primary analyses so results are given for both. The linkage disequilibrium between rs921943 and rs7700970 is d′ = 0.51, $r^2 = 0.241$ (from HapMap CEU data).

are two of the excreted Se compounds. In addition, a SNP in *CBS* (coding for cystathionine beta synthase), another enzyme involved in sulphur-containing amino acid metabolism, showed suggestive association with Se (rs6586282, $P = 4.6 \times 10^{-6}$, see Supplementary Material, Table S1). The relevant part of the metabolic pathway of sulphur-containing amino acids is summarized in Supplementary Material, Figure S4.

No previous and relevant genetic studies could be found, and attempts to define the relevant gene by searching published information for effects on gene expression were not helpful. At least one SNP in this region, rs921945, is an eQTL for *BHMT2*, but this SNP did not show significant association with Se in our data ($P = 0.055$). Although we cannot define whether the significant SNPs in this region affect BHMT, BHMT2 or DMGDH activity (or perhaps all three), it seems probable that variation at this locus affects selenoaminoacid metabolism and hence the concentration of Se in erythrocytes.

Zinc

Three regions showed significant associations with Zn concentrations. The effect ∼86.5 Mbp on chromosome 8 can most readily be explained by the location of genes coding for carbonic anhydrases (*CA1, CA2, CA3, CA13*) in this region. Carbonic anhydrases are Zn-containing enzymes which catalyse the formation and dissociation of carbonic acid from carbon dioxide and water, and play an important role in carbon dioxide transport. It is known that carbonic anhydrases account for nearly all the Zn in erythrocytes (32,33) so presumably the significant SNPs in this region either control, or are associated with untyped variants which control, expression of carbonic anhydrases. In this connection, we note that SNPs in this region affect the expression of *CA2, CA3* and *CA13* in transformed lymphocyte lines (34) and that the SNPs which affect expression of *CA2* also showed suggestive associations with Zn concentrations in our data (see Supplementary Material, Table S3). In addition, there is evidence that expression of *CA3* in the liver is affected by rs7004871 [$P = 4.8 \times 10^{-8}$ for *CA3* expression (35)]; the *P*-value for this SNP for Zn concentration in our results is $P = 0.0028$ but nearby SNPs show $P < 5 \times 10^{-8}$.

The chromosome 15 locus covers a number of genes including *SCAMP5* and *PPCDC*. The former has no obvious connection with Zn but *PPCDC*, coding for phosphopantothenoylcysteine decarboxylase, might affect Zn status through effects on vitamin B5 (pantothenate) metabolism. A neurological syndrome of pantothenate kinase-associated neurodegeneration (OMIM #234200, Neurodegeneration with Brain Iron Accumulation 1; NBIA1) is due to enzyme deficiency at an early step of the CoA synthesis pathway. Study of the effects of pantothenate kinase deficiency in *Drosophila* showed that heterozygosity for either of two mutations in that gene was associated with substantially increased whole-body Zn concentration (36). Flies heterozygous for pantothenate kinase deficiency had upregulated transcription of *foi*, the *Drosophila* homolog of the mammalian enterocyte Zn transporter Zip4. By analogy, we speculate that other genetic variation in the CoA synthesis pathway, specifically in *PPCDC*, could also lead to variation in Zn metabolism and hence to our allelic association finding. Examination of the ENCODE data for the chromosome 15 locus showed evidence for regulatory regions close to the lead SNP rs2120019 in

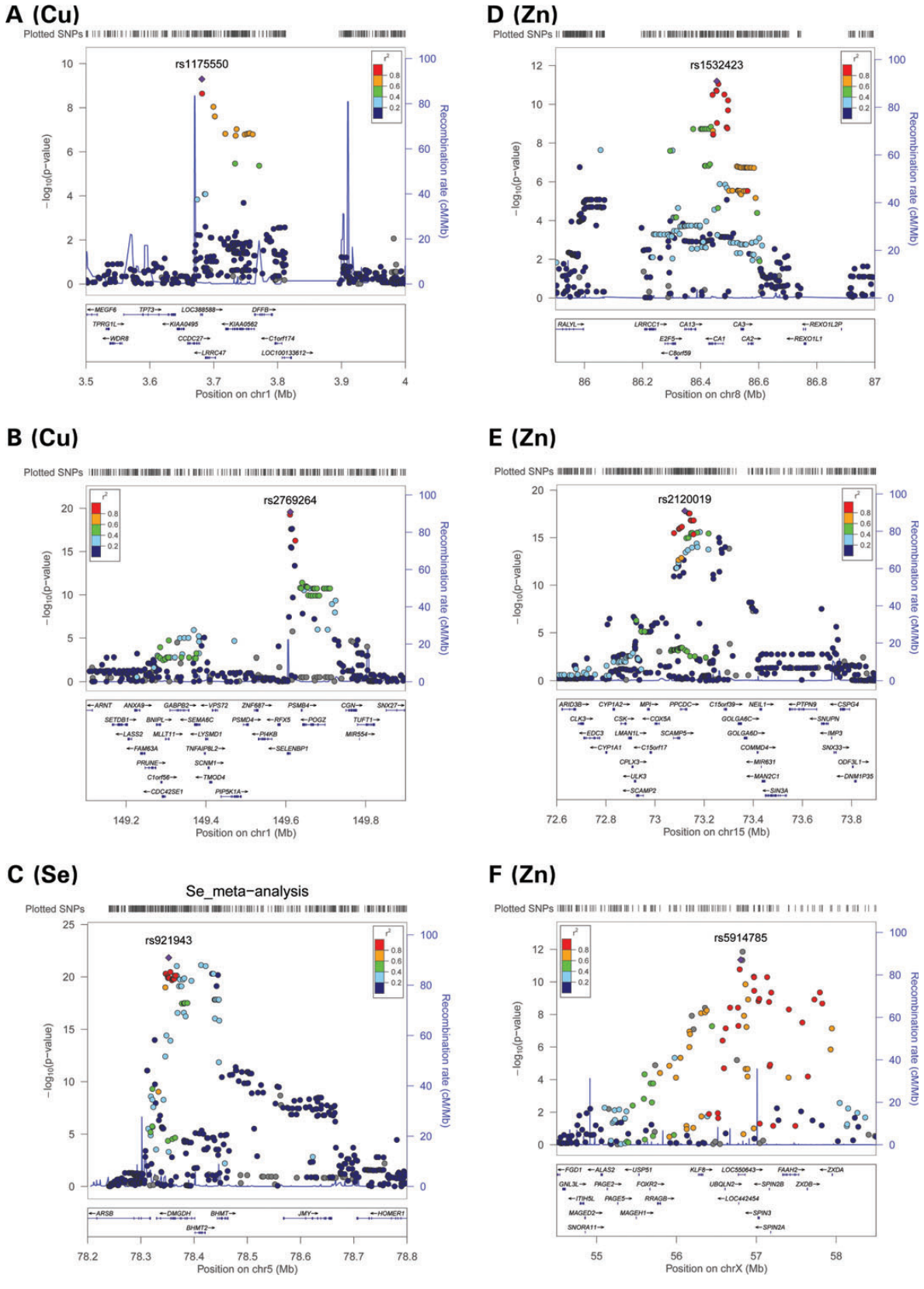

Figure 2. Regional association plots for Cu, Se and Zn. (**A** and **B**) Cu, showing the two significant regions on chromosome 1; (**C**) Se, showing the significant region on chromosome 5 (from meta-analysis of QIMR and ALSPAC data); (**D**–**F**) Zn, showing the significant regions on chromosomes 8, 15 and X. Note that the most significant SNP on chromosome X (rs4826508, $P = 1.40 \times 10^{-12}$) lacks LD data so the second most significant SNP (rs5914785, $P = 4.40 \times 10^{-12}$) is used as the index SNP instead.

multiple cell types (GM12878, HSMM, HUVEC, K562 and NHLF), which would be consistent with a widespread requirement for pantothenate and for regulation of *PPCDC*; but this cannot be taken further with our current data.

The SNPs on the X chromosome showing significant associations with Zn concentrations cover a wide region. Genes within this locus include those for the Zn-finger proteins *KLF8, ZXDA* and *ZXDB*, so genetic variation in this region may affect Zn concentration through effects on the expression of Zn-containing proteins in a similar way to that which we suggest for the carbonic anhydrase genes. Multiple SNPs within this region (listed in Supplementary Material, Table S4) affect expression of *KLF8* in lymphoblastoid cell lines (34), although the exact locations of the peak effects on *KLF8* expression and Zn concentration do not coincide. No similar effect could be found for *ZXDA* or *ZXDB*.

Comparison with genetic linkage results

Data from the Australian cohort were previously used in a genetic linkage design (2), in which suggestive results (LOD > 1.6) were found for Se on chromosomes 4 and 8 (peak LOD scores 2.46 at 130 cM, approximately 128 Mbp, and 2.05 at 25 cM, approximately 11 Mbp, respectively) and for Zn on chromosome 2 (peak LOD score 2.73 at 220 cM, approximately 219 Mbp). In our current study, we found no suggestive ($P < 5 \times 10^{-6}$) allelic associations for these elements within 10 Mbp of these three regions. As polygenic effects will not be detected by sib-pair linkage with feasible numbers of subjects unless they account for 10–20% of variance, the linkage results are not supported by the allelic association data. The exception would be if there were many different, family-specific, variants at one locus; so far, there is little evidence for this type of variation affecting quantitative traits in the general population, but it cannot be excluded.

Limitations

The study design and data availability impose some limitations on our results and the conclusions which can be drawn from them. The elemental analyses were conducted on blood or blood cells and the factors affecting concentrations in other tissues may well differ. Lack of speciation data means that no differentiation can be made between inorganic, organic and protein-bound forms of the measured elements. Functional studies, which are beyond the scope of this article, will be needed to address the mechanisms by which gene variation produces variation in the phenotype. In this article, we have a focus on the general population rather than groups at particularly high risk of deficiency and this aspect too requires further study.

Conclusions

Application of the genome-wide association approach to data on essential element concentrations has shown multiple loci where genetic variation affects these phenotypes, and this provides examples supporting the proposition that genetic differences between people can affect response to nutritional or other environmental factors. Definition of such allelic effects may identify subgroups who are more susceptible to deficiency, and allow selection of subjects likely to benefit either in trials of supplementation or (where the therapeutic window is narrow) in assessment of individual risks and benefits of treatment. As with other genetic studies on quantitative characteristics, study of larger cohorts or meta-analysis of data from multiple studies is expected to reveal additional loci and provide additional insight into the biological processes affecting uptake, excretion and distribution of essential elements in humans.

MATERIALS AND METHODS

Australia (QIMR)

Our analysis is based on the results for 2603 adults with phenotype and genotype data who participated in one or both of two studies run from the Queensland Institute of Medical Research (QIMR).

The first of these studies recruited twins born before 1964 who were enrolled in a volunteer registry (the Australian Twin Registry). Subjects and methods for this study are described in ref. (2). Briefly, participants completed a postal questionnaire in 1989 and a telephone interview in 1993–1994, and provided a blood sample in 1993–1996. We initially determined zygosity from responses to questions about physical similarity, but this has now been updated for those with data included in this article using SNP-typing results. Participants gave written informed consent, and the studies were approved by the appropriate ethics committees. Blood was collected from 1134 men and 2241 women.

As blood samples had been fractionated to provide plasma, buffy coat for DNA extraction, and erythrocytes, we used erythrocytes rather than whole blood for elemental analysis. Samples were stored at -80°C. Before analysis, the erythrocytes were thawed at room temperature and diluted 1:20 in ammonia/EDTA solution containing rhodium as an internal standard. Cu, Se and Zn concentrations were measured by inductively coupled plasma mass spectrometry (ICP-MS) on a Perkin-Elmer Elan 5000 mass spectrometer (PerkinElmer, Inc., Wellesley, MA, USA) or a Varian UltraMass (Varian Inc., Palo Alto, CA, USA). Haemoglobin concentration was then measured on the diluted samples using the cyanmethemoglobin method.

Results were log-transformed and analysis batch, haemoglobin concentration in the thawed sample and analytical quality control data were used as covariates in preliminary steps which generated standardized residuals for subsequent analysis, as previously described (1). Of the 2926 participants from this 1993–96 Twin Study with measurements of element concentrations in erythrocytes, 1570 had genome-wide SNP genotyping data. Genotypes were determined using Illumina chips; methods, quality control steps and imputation of untyped HapMap 2 SNPs were as previously described (37).

The second of the Australian studies which generated data on trace element concentrations took place in 2001–05 and was designed to characterize loci affecting alcohol or nicotine dependence through a genome-wide association approach (38). This was a twin-family design, in which relatives of participants in our earlier twin studies were also recruited. Data were again obtained through telephone interviews, and blood samples were obtained from 8396 people. Again, element concentrations

were determined on erythrocyte fractions by ICP-MS, but with an Agilent 7500 system (Agilent Technologies, Inc., Santa Clara, CA, USA). Covariate adjustment and generation of standardized residuals were carried out in the same way as for the previous study. One thousand one hundred and four people from the nicotine-alcohol studies had both phenotypic (element) data and genome-wide SNP data.

After allowing for overlap of 71 people who had data from participation in both the earlier and later Australian studies, there were 2603 valid sets of phenotype and genotype results. Genome-wide allelic association analysis was carried out on the standardized residuals for each element, further adjusted for sex and age, using an additive model accounting for within-family relatedness in Merlin (39) (http://www.sph.umich.edu/csg/abecasis/Merlin/, accessed 7 June 2012).

UK (ALSPAC)

The Avon Longitudinal Study of Parents and Children (ALSPAC), also known as Children of the Nineties, was designed to understand the ways in which the physical and social environments interact over time with genetic inheritance to affect health, behaviour and development in infancy, childhood and then into adulthood (40,41). The study area (the county of Bristol, formerly known as Avon) is an area bordering the Severn estuary, with a total population of 1 million, and includes Bristol, a major city of population 0.5 million, and surrounding areas which include small towns, villages and farming communities. Eligible women were those who were pregnant, resident in the study area and had an expected date of delivery between 1 April 1991 and 31 December 1992. They were recruited as early in pregnancy as possible. Of all mothers who were interested in taking part, an estimated 80% of the eligible population were included and answered at least one questionnaire. Maternal blood samples were collected in acid washed vacutainers by midwives on the first occasion on which they saw the pregnant women. Samples were kept as whole blood in the original tubes stored at 4°C.

Trace metal analysis was performed by the Center for Disease Control, Atlanta, GA in 2009–10. Clotted whole blood was digested by adding concentrated nitric acid and heating in a microwave at a controlled temperature and time. Addition of rhenium prior to heating allowed the correction of results for any loss due to evaporation. The digestion matrix was diluted 1:9 by volume using internal standards (Ir and Te) at a constant concentration. Diluted liquid samples were introduced as an aerosol into the inductively coupled plasma dynamic reaction cell mass spectrometry (ICP-DRC-MS) through a nebulizer and spray chamber carried by a flowing argon stream. For the analysis of Se, the cell was pressurized with methane allowing collision or reaction with the incoming ions to reduce interference.

A total of 10 015 women (mothers from the ALSPAC cohort) were genotyped using the Illumina 660 quad SNP chip which contains 557 124 SNP markers. Markers with minor allele frequency $< 1\%$, SNPs with $>5\%$ missing genotypes and any markers that failed an exact test of Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$) were excluded from further analyses. Genome-wide identity by state sharing was calculated for each pair of individuals in the cohort to identify cryptic relatedness. In order to identify individuals who might have ancestries other than Western European, we merged data from both cohorts with the 60 western European (CEU) founder, 60 Nigerian (YRI) founder and 90 Japanese (JPT) and Han Chinese (CHB) individuals from the International HapMap Project. Genome-wide IBS distances for each pair of individuals were calculated on markers shared between the HapMap and the Illumina 660K SNP chip, and then the multidimensional scaling option in R was used to generate a two-dimensional plot based upon individuals' scores on the first two principal coordinates from this analysis. Samples that did not cluster with the CEU individuals were excluded from subsequent analyses. In addition, we plotted the proportion of missing data for each individual against their genome-wide heterozygosity. Any individual, who did not cluster with others, was removed from further analyses. Samples were also excluded from analyses in the case of excessive missingness ($>5\%$), unusual genome-wide or X chromosome heterozygosity, as well as one individual from each pair of putatively related individuals (genome-wide IBD $>10\%$). After data cleaning, 8340 individuals and 526688 SNPs were left in the genome-wide data set.

We then conducted imputation using the MACH Markov Chain Haplotyping software with CEU individuals from phase 2 of the HapMap project as a reference set (release 22). The final imputed data set consisted of 8340 individuals, each with 2 594 390 imputed markers. Only imputed genotypes with minor allele frequencies $\geq 1\%$ and R-sqr ≥ 0.3 were considered for association. Of these 8340 with genetic data, 2874 mothers also had phenotype data available.

Metal concentrations were $\log_{10}$ transformed to approximate normality. Batch was included as a random effect and standardized residuals were derived. We then performed genome-wide association analysis on these residuals using the software package mach2qtl (42).

Further analysis

Results from the Australian and UK data for Se were compared and combined by meta-analysis using METAL (43) (http://www.sph.umich.edu/csg/abecasis/Metal/, accessed 6 June 2012). Results were visualized using WGAViewer (44) (http://compute1.lsrc.duke.edu/softwares/WGAViewer/, accessed 6 June 2012) and LocusZoom (45) (http://csg.sph.umich.edu/locuszoom/, accessed 6 June 2012). Conditional analysis, in which the most significant SNP at each locus was included as a covariate in order to detect independent effects at significant loci, was carried out on each of the data sets and results were combined by meta-analysis. Gene-based analysis with VEGAS (46) (http://gump.qimr.edu.au/VEGAS/, accessed 30 October 2012) was used to check whether any genes showed a significant over-representation of nominally significant SNPs, as might occur if several variants in a gene, not in linkage disequilibrium with each other, affect the phenotype. Regions around SNPs which showed genome-wide significance were checked for expression QTLs using 'eQTL.Chicago.edu!' (http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/, accessed 5 November 2012) and 'mRNA by SNP Browser' (http://www.sph.umich.edu/csg/liang/asthma/, accessed 9 April 2013); and for regulatory elements using the ENCODE browser (47) (http://genome.ucsc.edu/cgi-bin/hgTracks, accessed 9 April 2013).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

QIMR Study: We thank the twins and other members of their families for their participation, and the staff of the Genetic Epidemiology and Molecular Epidemiology Units, Queensland Institute of Medical Research, for project management, subject recruitment and interviews, sample collection and sample processing. Trace element measurements were made using the facilities of the Department of Clinical Biochemistry, Royal Prince Alfred Hospital, Sydney, Australia. *ALSPAC Study*: We are extremely grateful to all the individuals who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. We thank the Centre National de Genotypage for generating the ALSPAC GWA data. The assays of the maternal blood samples were carried out at the Centers for Disease Control and Prevention with funding from the US National Oceanic and Atmospheric Administration (NOAA).

Conflict of Interest statement. None declared.

FUNDING

Sample collection, and the recruitment and interviewing of participants, was funded by grants AA007535, AA013320, AA013321, AA013326 and DA012854 from the US National Institutes of Health to A.C.H., N.G.M., P.A.F.M., and the late Richard Todd, MD, PhD. Biomarker measurement was supported by AA014041 to J.B.W. G.W.M. is supported by the National Health and Medical Research Council of Australia Fellowship Scheme. The UK Medical Research Council, the Wellcome Trust (grant ref: 092731), and the University of Bristol currently provide core support for the Avon Longitudinal Study of Parents and their Children. J.P.K. was funded by a Wellcome Trust 4-year PhD studentship in molecular, genetic, and life course epidemiology (WT083431MA).

REFERENCES

1. Whitfield, J.B., Dy, V., McQuilty, R., Zhu, G., Montgomery, G.W., Ferreira, M.A., Duffy, D.L., Neale, M.C., Heijmans, B.T., Heath, A.C. *et al.* (2007) Evidence of genetic effects on blood lead concentration. *Environ. Health Perspect.*, **115**, 1224–1230.
2. Whitfield, J.B., Dy, V., McQuilty, R., Zhu, G., Heath, A.C., Montgomery, G.W. and Martin, N.G. (2010) Genetic effects on toxic and essential elements in humans: arsenic, cadmium, copper, lead, mercury, selenium and zinc in erythrocytes. *Environ. Health Perspect.*, **118**, 776–782.
3. Prohaska, J.R. (2008) Role of copper transporters in copper homeostasis. *Am. J. Clin. Nutr.*, **88**, 826S–829S.
4. van den Berghe, P.V. and Klomp, L.W. (2009) New developments in the regulation of intestinal copper absorption. *Nutr. Rev.*, **67**, 658–672.
5. O'Halloran, T.V. and Culotta, V.C. (2000) Metallochaperones, an intracellular shuttle service for metal ions. *J. Biol. Chem.*, **275**, 25057–25060.
6. Prohaska, J.R. and Gybina, A.A. (2004) Intracellular copper transport in mammals. *J. Nutr.*, **134**, 1003–1006.
7. Uriu-Adams, J.Y. and Keen, C.L. (2005) Copper, oxidative stress, and human health. *Mol. Aspects Med.*, **26**, 268–298.
8. Madsen, E. and Gitlin, J.D. (2007) Copper deficiency. *Curr. Opin. Gastroenterol.*, **23**, 187–192.
9. Rossi, L., Lombardo, M.F., Ciriolo, M.R. and Rotilio, G. (2004) Mitochondrial dysfunction in neurodegenerative diseases associated with copper imbalance. *Neurochem. Res.*, **29**, 493–504.
10. Bayer, T.A. and Multhaup, G. (2005) Involvement of amyloid beta precursor protein (AbetaPP) modulated copper homeostasis in Alzheimer's disease. *J. Alzheimers Dis.*, **8**, 201–206; discussion 209–215.
11. Jaiser, S.R. and Winston, G.P. (2010) Copper deficiency myelopathy. *J. Neurol.*, **257**, 869–881.
12. Kuehnelt, D., Kienzl, N., Traar, P., Le, N.H., Francesconi, K.A. and Ochi, T. (2005) Selenium metabolites in human urine after ingestion of selenite, L-selenomethionine, or DL-selenomethionine: a quantitative case study by HPLC/ICPMS. *Anal. Bioanal. Chem.*, **383**, 235–246.
13. Fairweather-Tait, S.J., Bao, Y., Broadley, M.R., Collings, R., Ford, D., Hesketh, J.E. and Hurst, R. (2011) Selenium in human health and disease. *Antioxid. Redox Signal.*, **14**, 1337–1383.
14. Brinkman, M., Reulen, R.C., Kellen, E., Buntinx, F. and Zeegers, M.P. (2006) Are men with low selenium levels at increased risk of prostate cancer? *Eur. J. Cancer*, **42**, 2463–2471.
15. Flores-Mateo, G., Navas-Acien, A., Pastor-Barriuso, R. and Guallar, E. (2006) Selenium and coronary heart disease: a meta-analysis. *Am. J. Clin. Nutr.*, **84**, 762–773.
16. Mistry, H.D., Pipkin, F.B., Redman, C.W. and Poston, L. (2012) Selenium in reproductive health. *Am. J. Obstet. Gynecol.*, **206**, 21–30.
17. Stone, C.A., Kawai, K., Kupka, R. and Fawzi, W.W. (2010) Role of selenium in HIV infection. *Nutr. Rev.*, **68**, 671–681.
18. Harthill, M. (2011) Review: micronutrient selenium deficiency influences evolution of some viral infectious diseases. *Biol. Trace Elem. Res.*, **143**, 1325–1336.
19. Hunt, J.R., Beiseigel, J.M. and Johnson, L.K. (2008) Adaptation in human zinc absorption as influenced by dietary zinc and bioavailability. *Am. J. Clin. Nutr.*, **87**, 1336–1345.
20. Liuzzi, J.P., Guo, L., Chang, S.M. and Cousins, R.J. (2009) Kruppel-like factor 4 regulates adaptive expression of the zinc transporter Zip4 in mouse small intestine. *Am. J. Physiol. Gastrointest. Liver Physiol.*, **296**, G517–G523.
21. Sekler, I., Sensi, S.L., Hershfinkel, M. and Silverman, W.F. (2007) Mechanism and regulation of cellular zinc transport. *Mol. Med.*, **13**, 337–343.
22. Kambe, T. (2011) An overview of a wide range of functions of ZnT and Zip zinc transporters in the secretory pathway. *Biosci. Biotechnol. Biochem.*, **75**, 1036–1043.
23. Little, P.J., Bhattacharya, R., Moreyra, A.E. and Korichneva, I.L. (2010) Zinc and cardiovascular disease. *Nutrition*, **26**, 1050–1057.
24. Rutter, G.A. (2010) Think zinc: New roles for zinc in the control of insulin secretion. *Islets*, **2**, 49–50.
25. Black, R.E. (2003) Zinc deficiency, infectious disease and mortality in the developing world. *J. Nutr.*, **133**, 1485S–1489S.
26. Tuerk, M.J. and Fazel, N. (2009) Zinc deficiency. *Curr. Opin. Gastroenterol.*, **25**, 136–143.
27. Croxford, T.P., McCormick, N.H. and Kelleher, S.L. (2011) Moderate zinc deficiency reduces testicular Zip6 and Zip10 abundance and impairs spermatogenesis in mice. *J. Nutr.*, **141**, 359–365.
28. Mackenzie, G.G., Zago, M.P., Aimo, L. and Oteiza, P.I. (2007) Zinc deficiency in neuronal biology. *IUBMB Life*, **59**, 299–307.
29. Lovell, M.A. (2009) A potential role for alterations of zinc and zinc transport proteins in the progression of Alzheimer's disease. *J. Alzheimers Dis.*, **16**, 471–483.
30. Sandstead, H.H., Prasad, A.S., Schulert, A.R., Farid, Z., Miale, A. Jr., Bassilly, S. and Darby, W.J. (1967) Human zinc deficiency, endocrine manifestations and response to treatment. *Am. J. Clin. Nutr.*, **20**, 422–442.
31. Prasad, A.S. (2009) Zinc: role in immunity, oxidative stress and chronic inflammation. *Curr. Opin. Clin. Nutr. Metab. Care*, **12**, 646–652.
32. Ohno, H., Doi, R., Yamamura, K., Yamashita, K., Iizuka, S. and Taniguchi, N. (1985) A study of zinc distribution in erythrocytes of normal humans. *Blut*, **50**, 113–116.
33. Gardiner, P.E., Gessner, H., Bratter, P., Stoeppler, M. and Nurnberg, H.W. (1984) The distribution of zinc in human erythrocytes. *J. Clin. Chem. Clin. Biochem.*, **22**, 159–163.

34. Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M. *et al.* (2007) A genome-wide association study of global gene expression. *Nat. Genet.*, **39**, 1202–1207.
35. Schadt, E.E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P.Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C. *et al.* (2008) Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.*, **6**, e107.
36. Gutierrez, L., Sabaratnam, N., Aktar, R., Bettedi, L., Mandilaras, K. and Missirlis, F. (2010) Zinc accumulation in heterozygous mutants of fumble, the pantothenate kinase homologue of Drosophila. *FEBS Lett.*, **584**, 2942–2946.
37. Medland, S.E., Nyholt, D.R., Painter, J.N., McEvoy, B.P., McRae, A.F., Zhu, G., Gordon, S.D., Ferreira, M.A., Wright, M.J., Henders, A.K. *et al.* (2009) Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am. J. Hum. Genet.*, **85**, 750–755.
38. Heath, A.C., Whitfield, J.B., Martin, N.G., Pergadia, M.L., Goate, A.M., Lind, P.A., McEvoy, B.P., Schrage, A.J., Grant, J.D., Chou, Y.L. *et al.* (2011) A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications. *Biol. Psychiatry*, **70**, 513–518.
39. Abecasis, G.R., Cherny, S.S., Cookson, W.O. and Cardon, L.R. (2002) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.*, **30**, 97–101.
40. Boyd, A., Golding, J., Macleod, J., Lawlor, D.A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S. and Davey Smith, G. (2013) Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.*, **42**, 111–127.
41. Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A. *et al.* (2013) Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int. J. Epidemiol.*, **42**, 97–110.
42. Li, Y., Willer, C.J., Ding, J., Scheet, P. and Abecasis, G.R. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.
43. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
44. Ge, D., Zhang, K., Need, A.C., Martin, O., Fellay, J., Urban, T.J., Telenti, A. and Goldstein, D.B. (2008) WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res.*, **18**, 640–643.
45. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.
46. Liu, J.Z., McRae, A.F., Nyholt, D.R., Medland, S.E., Wray, N.R., Brown, K.M., Hayward, N.K., Montgomery, G.W., Visscher, P.M., Martin, N.G. *et al.* (2010) A versatile gene-based test for genome-wide association studies. *Am. J. Hum. Genet.*, **87**, 139–145.
47. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R. *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.