Sex- and age-interacting eQTLs in human complex diseases

Chen Yao^{1,2,3}, Roby Joehanes^{1,2,3}, Andrew D. Johnson^{1,2}, Tianxiao Huan^{1,2,3}, Tõnu Esko^{8,9,10}, Saixia Ying⁴, Jane E. Freedman⁵, Joanne Murabito^{1,6}, Kathryn L. Lunetta⁷, Andres Metspalu^{8,11}, Peter J. Munson⁴ and Daniel Levy^{1,2,3,*}

¹National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA, ²Division of Intramural Research, National Heart and ³Population Research Branch, National Institutes of Health, National Heart, Lung, and Blood Institute, Bethesda, MD, USA, ⁴Mathematical and Statistical Computing Laboratory, National Institutes of Health, Bethesda, MD, USA, ⁵Department of Medicine, University of Massachusetts Medical School, Worcester, MA, USA, ⁶Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA, ⁷Department of Biostatistics, Boston University, Boston, MA, USA, ⁸Estonian Genome Center, University of Tartu, Tartu, Estonia, ⁹Division of Endocrinology, Children's Hospital Boston, Boston, MA, USA, ¹⁰Broad Institute, Cambridge, MA, USA and ¹¹Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

Received May 16, 2013; Revised October 31, 2013; Accepted November 12, 2013

Many complex human diseases exhibit sex or age differences in gene expression. However, the presence and the extent of genotype-specific variations in gene regulation are largely unknown. Here, we report results of a comprehensive analysis of expression regulation of genetic variation related to 11 672 complex disease-associated SNPs as a function of sex and age in whole-blood-derived RNA from 5254 individuals. At false discovery rate <0.05, we identified 14 sex- and 10 age-interacting expression quantitative trait loci (eQTLs). We show that these eQTLs are also associated with many sex- or age-associated traits. These findings provide important context regarding the regulation of phenotypes by genotype–environment interaction.

INTRODUCTION

Most common complex diseases are a consequence of genetic factors and environmental exposures (1,2). Sex and age can be considered 'environmental' exposures that might interact with a subset of genetic variants in a manner similar to extrinsic environmental factors. Many human diseases display sex and age differences in prevalence, clinical features and prognosis (3). For example, atherosclerotic cardiovascular disease is primarily a condition of middle aged and older adults and it has an earlier age of onset in men than in women. In contrast, many autoimmune diseases can begin in childhood or adolescence and are more prevalent in women than in men.

Several studies have explored genotype–phenotype associations and identified sex or age differences (4–10). For example, genetic variation at the *LYPLAL1* locus was found to have a sex-interaction in relation to obesity traits (11). A polymorphism in the angiotensin-converting enzyme (ACE) was

found to be associated with hypertension and with diastolic blood pressure among men but not women (12). Joubert *et al.* (9) identified two SNPs with age-dependent genetic effects on blood pressure. Fritsche *et al.* identified various loci associated with age-related macular degeneration (AMD) (13). However, the mechanisms by which genetic variants affect gene expression are largely unknown. One approach for discovering genotype–phenotype interactions in the context of gene regulation is to study gene expression as a quantitative phenotype and identify genetic variation that is associated with gene expression levels (14). Interrogating quantitative trait locus (QTL) mapping with gene expression profiling allows for the identification of expression quantitative trait loci (eQTLs) among GWAS results (15). An eQTL can be classified as *cis*- or *trans*-acting, based on local or distant regulation of gene expression, respectively. Among genes that display sex-specific or age-related changes in expression, a subset may exhibit genotype interaction with sex or age. While some sex-specific (16,17) and numerous

*To whom correspondence should be addressed at: 73 Mt. Wayte Avenue, Framingham Heart Study, Framingham, MA 01702 USA. Tel: +1 5089353458; Fax: +1 5086261262; Email: levyd@nih.gov

age-related expression changes (18–21) have been identified, few studies have yet considered sex- and age-interactions in eQTL analysis (22,23).

In this paper, we explored the effect of genotype-by-sex and genotype-by-age interaction for human complex disease traits based on whole blood gene expression. To this end, we examined 11 672 trait-associated SNPs from the National Human Genome Research Institute (NHGRI) GWAS catalogue (24) and the dbGaP database (25) in relation to gene expression in 5254 participants in the Framingham Heart Study (FHS). The large sample sizes allowed to formally test for the interaction of eQTL genotypes with sex and age and their association with gene expression. Many of the eQTLs that we identified have been implicated in GWAS of sex- and age-related diseases.

RESULTS

SNPs associated with gene expression as a function of sex or age

Whole blood samples from 5254 FHS participants (2833 women, 2421 men, age range 24–92, mean 55 years) were analysed for expression levels of 17 873 genes. Individual transcript intensities were corrected for experimental variation and normalized as described in the Methods section. The expression values of the 17 873 transcripts were treated as quantitative traits and fitted to linear regression models to compute interaction eQTLs. Among the 11 672 eligible trait-associated SNPs, we identified 14 eQTLs (13 autosomal *cis*-eQTLs, 1 long-range *cis*-eQTL on the X chromosome) with significant sex interactions (Table 1) and 10 eQTLs with significant age interactions (all *cis*; Table 2) at false discovery rate (FDR) < 0.05 (permutation FDR < 0.02). No allele frequency differences were found among sex-interaction eQTLs.

Genotype–sex interaction results display different effect models

After identifying genotype–sex interactions (i.e. genetic variation that is associated with different expression levels in men and women), the cohort was stratified by sex and eQTLs were considered in sex-specific analysis (with adjustment for age). We identified four eQTLs (at FDR < 0.05 for the interaction test and $P < 0.05$ in either sex) conforming to Model A: a genetic variant has the same direction of effect but differences in the magnitude of effect in both sexes; one eQTL conforming to Model B: genotypic effects are present only in one sex; and nine eQTLs conforming to Model C: genotypic effects are present in both sexes but with opposite directions of effect (see Supplementary Material, Methods and Fig. 1 for details).

Model A results

As shown in Figure 2A, for rs9302752, genotype is associated with *NOD2* gene expression in both men and women, but there is a greater effect on expression in women than in men (interaction $P = 9.03 \times 10^{-13}$). Zhang *et al.* (26) found that each copy of the G allele of rs9302752 (near *NOD2*) conferred ~1.6-fold higher odds of leprosy among exposed individuals. They also suggested that the association of leprosy with rs9302752 reflects the effects of regulatory variants on *NOD2*

expression. Our results are not only consistent with this assumption about regulation of *NOD2*, but in addition, we found that the AA genotype of rs9302752 confers a higher expression of *NOD2* in men than in women (t -test, $P = 6.0 \times 10^{-6}$), while for the GG genotype, *NOD2* expression is higher in women than in men (t -test $P = 0.0075$). Leprosy is more prevalent in men than women (27). A recent study found that the expression of *NOD2* correlated with the clinical presentation of leprosy and was greater in patients with limited than progressive disease (28).

Model B results

rs8047080 (and other nearby SNPs) has been reported to be associated with high-density lipoprotein (HDL) cholesterol (29). We found that the association of this SNP with the expression of *TSNAXIP1* is male specific. As shown in Figure 2B, each copy of the G allele of rs8047080 is associated with an increased expression of *TSNAXIP1* in men ($P = 7.03 \times 10^{-05}$), but not in women ($P = 0.1$). This gene is mainly expressed in the testis, but the function of *TSNAXIP1* and the mechanism of its association with HDL are unknown.

Model C results

rs6867983 in *C5orf35* has been shown to be associated with plasma triglyceride levels (30). We found that this SNP has opposite effects on the expression of *IL6ST* in men and women. As shown in Figure 2C, each copy of the T allele is associated with an increased expression of *IL6ST* in women ($P = 0.0006$), but decreased expression in men ($P = 0.05$). *IL6ST* is a ubiquitously expressed intermediate of the interleukin-6 signalling pathway and is associated with metabolic syndrome and breast cancer (31,32). We also found that the expression of *IL6ST* is higher in women than in men (t -test $P = 1.07 \times 10^{-49}$). Another SNP, rs5991441 on the X chromosome, has been shown to be associated with lipids (33). We found that this SNP and other nearby SNPs [rs5991573 is associated with blood pressure (34), and rs5991545 is associated with diabetes mellitus (35)] have opposite effects on the expression of *CXorf23* in men and women. As shown in Figure 2D, each copy of the T allele of rs5991441 is associated with an increased expression of *CXorf23* in women ($P = 3.26 \times 10^{-5}$), but decreased expression in men ($P = 3.67 \times 10^{-5}$). Notably, because *CXorf23* is on the X chromosome and 20 megabases away from rs5991545, this eQTL can be defined as long-range *cis*-regulating (36).

Genotype–age interactions reflecting genotype effects on human ageing

For age-related eQTLs, different genetic variants showed different effects on gene expression throughout the age span. We found many age-related traits with genotype–age interaction in association with the expression of *SLC44A4*. rs659445, an intronic-variant of *EHMT2*, is associated with macular degeneration (25). As shown in Figure 3A, each copy of the C allele of this SNP not only increases the expression of *SLC44A4* ($P < 2.2 \times 10^{-16}$), but also alters its age-dependency ($P = 3.41 \times 10^{-08}$ for the CC genotype, $P = 0.2$ for the TT genotype). Another SNP rs494620 causes a synonymous change in *SLC44A4*; this SNP is associated with age of menopause (37). As shown in Figure 3B, each copy of the A allele of this eQTL not only decreases the expression of *SLC44A4* ($P < 2.2 \times$

Table 1. Sex-interaction eQTLs in human complex traits

eQTL	Chromosome: Position	Trait/disease	Expressed gene	<i>P</i> -value (interaction)	<i>P</i> -value (main effect)	β (men)	<i>P</i> -value (men)	β (women)	<i>P</i> -value (women)	<i>P</i> -value (differential expression)
rs9302752	16 : 50 719 103	Leprosy	NOD2	8.15E-10	2.41E-35	0.23	4.21E-233	0.29	<1E-300	NS
rs9270986 (rs3129889 rs3135388 rs9271366)	6 : 32 574 060	Multiple sclerosis	HLA-DRB5	4.33E-09	1.29E-218	2.49	<1E-300	2.35	<1E-300	NS
rs3129860	6 : 32 401 079	Diabetes mellitus, Type 1	HLA-DRB5	2.57E-05	1.99E-121	2.45	<1.0E-300	2.33	<1.0E-300	NS
rs1335515	14 : 58 385 365	Attention deficit disorder with hyperactivity	KIAA0586	4.00E-06	NS	0.01	0.03	−0.016	0.00014	1.69E-10
rs12145451	1 : 212 425 291	Heart failure	PPP2R5A	3.35E-06	1.27E-15	0.14	8.21E-74	0.19	2.30E-152	NS
rs8047080 (rs3868142 rs3868143 rs11860295)	16 : 67 402 588	Lipoproteins, HDL	TSNAXIP1	6.08E-06	NS	0.03	7.03E-05	−0.011	NS	NS
rs599634	6 : 50 013 663	Body mass index	MUT	8.35E-06	NS	0.04	9.13E-06	−0.01	0.16	NS
rs846793	6 : 101 305 983	Memory	GRIK2	8.6E-06	NS	−0.03	0.001	−0.024	0.002	NS
rs11634397	15 : 80 432 222	Diabetes mellitus, Type 2	C15orf37	9.60E-06	NS	−0.026	0.002	0.018	0.015	0.013
rs17291650	12 : 51 213 433	Death, sudden, cardiac	LIMA1	1.66E-05	NS	0.023	0.026	−0.031	0.00081	2.02E-13
rs6867983	5 : 55 854 153	Triglycerides	IL6ST	2.67E-05	NS	−0.023	0.05	0.038	0.0006	1.07E-49
rs3132610	6 : 30 544 401	Lupus erythematosus, systemic	HCG8	3.71E-05	NS	−0.036	NS	0.06	0.00011	0.01
rs644148	19 : 44 970 935	Personality	BLOC1S3	3.93E-05	NS	0.013	0.016	−0.015	0.0014	NS
rs10503734	8 : 23 528 230	Blood pressure	NKX3-1	5.07E-05	NS	0.023	0.0034	0.066	2.42E-18	1.91E-06
X-Chromosome eQTL										
rs5991441	X : 42 930 414	Lipids	CXorf23 (X: 19 930 978–	2.70E-10	NS	−0.13	3.67E-05	0.18	3.26E-05	0.0012
rs5991573	X : 42 930 110	Blood pressure	19 988 416)	2.73E-10	NS	−0.13	3.68E-05	0.181	3.27E-05	0.0012
rs5991545	X : 42 889 004	Diabetes mellitus		2.17E-09	NS	−0.12	9.41E-05	0.17	0.00011	0.0012

() denotes SNPs associated with the same traits and in the LD ($R^2 > 0.8$).

β denotes change in expression (in RMA units) per allele.

NS denotes SNP is not significantly associated with gene expression in all samples or within specific gender ($P > 0.05$); or gene not significantly different expressed between man and woman (FDR > 0.05).

Table 2. Age-interaction eQTLs in human complex traits

eQTL	Chromosome: position	Disease/trait	Expressed gene	P-value (interaction)	P-value (main effect)	β (major allele)	β (minor allele)	P-value (differential expression)
rs659445 (rs535586 rs652888)	6 : 31 864 304	Macular degeneration	SLC44A4	6.33E-16	<1.0E-300	0.00044	−0.0046	2.02E-07
rs486416 (rs760293 rs644045)	6 : 31 856 070	Multiple sclerosis	SLC44A4	6.35E-16	<1.0E-300	0.00044	−0.0046	2.02E-07
rs494620	6 : 31 838 713	Menopause (age at onset)	SLC44A4	7.91E-08	1.34E-197	−0.0026	9.67E-05	2.02E-07
rs2736428	6 : 31 843 924	Telomere length	SLC44A4	6.11E-07	1.81E-131	−0.00233	0.000805	2.02E-07
rs805297 (rs2844479)	6 : 31 622 606	Rheumatoid arthritis	SLC44A4	6.27E-07	1.82E-94	−0.0026	0.000461	2.02E-07
rs437179 (rs389883)	6 : 31 929 014	Lupus erythematosus, systemic	SLC44A4	4.51E-07	1.38E-57	−0.0076	0.0023	2.02E-07
rs11643520	16 : 24 531 718	Brain	PRKCB	2.82E-06	1.07E-05	0.0014	−0.0080	7.41E-05
rs6701037	1 : 175 120 079	Alcoholism	KIAA0040	3.10E-06	1.42E-30	0.0028	0.00031	2.88E-14
rs2647087 (rs2858333)	6 : 32 681 049	Psoriasis	HLA-DQA2	5.30E-06	8.21E-08	−0.0027	−0.0089	NS
rs7745656	6 : 32 680 970	Diabetes mellitus, Type 1	HLA-DQA2	5.30E-06	8.21E-08	−0.0027	−0.0089	NS
rs2182114 (rs1245058)	1 : 111 784 227	Cholesterol, HDL	CHI3L2	1.32E-05	NS	−0.00046	−0.0037	5.72E-09
rs10181051	2 : 242 800	Neuroblastoma	SH3YL1	1.68E-05	1.76E-19	−0.054	−0.0014	7.53E-13
rs236918 (rs7112513)	11 : 117 091 609	Iron-regulatory proteins	TAGLN	2.30E-05	NS	0.0017	−0.022	0.00031
rs9635963	18 : 21 749 615	Sex hormone-binding globulin	CABLES1	2.71E-05	3.37E-05	−0.011	0.0040	NS
rs6733301	2 : 25 276 284	Body height	ADCY3	5.05E-05	5.88E-5	1.64E-05	0.0023	NS
rs10782001	16 : 30 942 625	Psoriasis	ZNF747	5.13E-05	NS	0.00031	−0.0019	NS

β denotes change in expression (in RMA units) per allele (year/age).
NS denotes SNP is not significantly associated with gene expression in whole samples ($P > 0.05$); or gene not significantly different expressed with age (FDR > 0.05).

10^{-16}), but also alters its age-dependent expression ($P = 2.9 \times 10^{-07}$ for the GG genotype, $P = 0.6$ for the AA genotype). Nearby SNPs rs486416 [associated with multiple sclerosis (38)] and rs2736428 [associated with leucocyte telomere length (39)] also show strong age interaction. Upon observing that alleles of rs659445 and rs494620 ($R^2 < 0.5$) have opposite effects on the expression of *SLC44A4*, we then tested for SNP $\times$ SNP interaction. As shown in Figure 4, these two SNPs display significant interaction ($P = 4.16 \times 10^{-6}$). A haplotype with the rs659445 C and rs494620 G alleles has a higher expression of *SLC44A4*. Of note, an association between the age of menopause with AMD was recently reported (40). To explore the relation between *SLC44A4* expression and age, we tested SNP $\times$ age interaction for each exon probesets of *SLC44A4*. We found strong SNP $\times$ age interactions for three exons of *SLC44A4* (Affymetrix probesets 2949503, 2949504 and 2949505; chr6:31941714-31944981, hg19). These *SLC44A4* exon-specific eQTLs are associated with many age-related traits (Supplementary Material, Table S1). These findings may be due to alternative splicing of *SLC44A4* (see section ‘Discussion’ for details).

Sex and age-interaction genes show sex-specific or age-related expression

Of 17 873 gene level transcripts, we identified 5866 genes that are differentially expressed in men versus women (including

236 on the X chromosome and 15 on the Y chromosome), and 6034 genes (including 198 on the X chromosome and 14 on the Y chromosome) that are differentially expressed as a function of age (at FDR < 0.05). For SNPs associated with complex traits, we found seven out of 14 eQTL-gene pairs that were differentially expressed in men and women (Table 1, last column). For the other eQTL-gene pairs that were not differentially expressed, one reason is that the genotypes may have opposite sex effects on expression (e.g. *NOD2*). A higher proportion of age-related eQTLs also show differential expression with age (6 out of 10 eQTLs, Table 2).

Internal and external validation

To perform internal validation, we split the samples by pedigrees into independent discovery (3507 samples) and replication sets (1694 samples). Using linear regression at a P -value threshold < 0.05 , we found a high concordance between discovery and replication. From the 14 identified eQTLs with interaction, 2 (one sex interaction and one age interaction) were significant in discovery, but not in replication (Supplementary Material, Table S2). In addition, all 14 eQTLs displayed directionally consistent associations in the discovery and internal validation sets. Notably, the probability by chance of obtaining P -values < 0.05 in both the discovery and replication sets is 0.05×0.05 , which was less than the Bonferroni FDR threshold ($0.05/14$ for sex-interaction eQTLs and $0.05/10$ for age-interaction eQTLs).

Figure 1. Models of genotype–sex interactions effects on gene expression that differ between men and women. For gene expression (y -axis), the genetic variant has the same direction of effect but differences are present in the magnitude of effect in both sexes (A); present only in one sex (B) or present in both sexes but with opposite directions of effect (C). M represents men, W represents women.

Figure 2. Effects of four sex-interaction eQTLs. For log₂ intensity of gene expression (y -axis), the genotype effects on expression might be in the same direction but show differences in magnitude of effect (A) or be present only in women (B) or present in both sexes but with opposite directions of effect (C and D). Red lines display expression values by the genotype in men, blue lines represent expression values by the genotype in women.

A total of 899 samples [mean age 37 (standard deviation 16.6) years; 50% females] from the Estonian Gene Expression Cohort (41) were used for external validation. After probeset matching and quality control, eight sex-interaction eQTLs and six age-interaction eQTLs had both genotype and gene expression data available in the Estonian samples. Using linear regression at a P -value threshold <0.05 , we replicated two sex-interactions (rs9302752 associated with leprosy and rs599634 associated with BMI) and three age-interaction (rs10782001 with psoriasis, rs2182114 with HDL cholesterol and rs659445 of macular degeneration) eQTLs (Supplementary Material, Table S3). Failure to replicate the other eQTLs may be due to the small sample size and much younger age of validation samples. For example, rs10503734, rs12145451 and rs9270986 are significant

sex-interaction eQTLs (model 1) in our data. Although the differences in effect sizes (betas) between men and women for these eQTLs are similar in our data and the Estonian data, these SNPs did not demonstrate statistically significant interaction with sex in the Estonian data. Seven of eight sex-interaction eQTLs and five of six age-interaction eQTLs that were present in the FHS and Estonian result had the same directions of effect.

DISCUSSION

We identified eQTLs that are associated with altered gene expression as a function of sex and age. These findings required formal testing of interaction and could not be identified

Figure 3. Effects of two age-interaction eQTLs. Genotype effects on expression in relation to two traits macular degeneration (A), and menopause (B). The x-axis represents the age of individuals (in years) and the y-axis represents the log2 intensity expression of *SLC44A4*. Red, green and blue lines represent the three genotypes.

Figure 4. SNP–SNP interaction for the expression of *SLC44A4* as the function of age. SNP–SNP interaction for the expression of *SLC44A4* involving rs659445 and rs694620. In each segment, the x-axis represents the age of individuals (in years) and the y-axis represents the expression of *SLC44A4*.

otherwise. These interactions may account for part of the phenotypic differences between men and women or as a function of age. Numerous instances of sex differences in genetic associations have been reported in autoimmune diseases and metabolic traits. Travis *et al.* (42) found that rs1270942 and rs3131379

demonstrate significant sex interaction in systemic lupus erythematosus. We also found evidence that these two SNPs are sex-interaction eQTLs ($P = 0.042$ for rs1270942, $P = 0.046$ for rs3131379, see Table 3 for details). Other examples of sex-interaction eQTLs include rs2605100 at the *LYPLAL1* locus, which was previously reported (11) to show sex interaction in relation to obesity traits, and SNPs (rs4343 and rs4329) at the *ACE* locus previously reported (12) sex interaction in relation to blood pressure. We also found evidence of two sex-interaction eQTLs (rs10401969 and rs17750998) in the loci that were previously reported to show heterogeneity for lipid traits between men and women (43). In addition, among the previously reported genetic variants associated with AMD (13), we found two loci (rs2230199 and rs4420638) that showed evidence of age-interaction eQTLs.

SLC44A4 is a choline transporter-like protein that is expressed in prostate, pancreatic and other cancers (44). SNPs in the *SLC44A4* region have been found to be associated with type 1 diabetes mellitus (25), vitiligo (45), age at menopause (37), leucocyte telomere length (39) and multiple sclerosis (46). But the association between the expression of *SLC44A4* and human ageing has not been reported previously. Three transcript variants encoding different isoforms have been found for this gene. We found that the expression of three exon probesets in two exons (Affymetrix exon cluster 395 343 and 395 344) was strongly affected by SNP $\times$ age interactions (Supplementary Material, Table S1). In addition, these two exons are located in the alternative splicing region of *SLC44A4* containing rs11966200 (chr6:31944795-31945295, hg19), which is associated with vitiligo (Fig. 5). We hypothesize that alternative splicing of *SLC44A4* is associated with advancing age and is affected by genotypes of nearby SNPs.

Table 3. Sex- and age-interaction eQTLs in sex- and age-dependent loci

eQTL	Chromosome: position	Disease/trait	Expressed gene	P-value (interaction)
rs1270942	6 : 31 918 860	Diabetes mellitus, Type 1; Lupus erythematosus, systemic	NEU1	0.042
rs3131379	6 : 31 721 033	Diabetes mellitus, Type 1; lupus erythematosus, systemic	NEU1	0.046
rs2605100	1 : 219 644 224	Adiposity	BPNT1	0.0009
rs4343	17 : 61 566 031	Angiotensin-converting enzyme inhibitors	PECAM1	0.0015
rs4329	17 : 61 563 458	Metabolism	PECAM1	0.0015
rs10401969	19 : 19 407 718	Cholesterol; triglycerides; cholesterol, LDL	ELL	0.02
rs17750998	19 : 19 388 446	Iron	MAU2	0.0021
rs2230199	19 : 6 718 387	Macular degeneration	XAB2	0.032
rs4420638	19 : 45 422 946	Cholesterol, LDL; Alzheimer's disease	CD3EAP	0.009

β denotes change in expression (in RMA units) per allele.

Figure 5. Genome browser view of two exons of *SLC44A4*. Two red arrows denote the locus of two exons: Affymetrix Exon cluster 395 343 and 395 344, respectively. Purple arrow bar denotes the strange splicing region of *SLC44A4*. SNPs in green colour are those reported in NHGRI GWAS catalogue.

The X chromosome was thought to be responsible for sex-specific immune responses (47). We found only one long-range *cis*-eQTL on the X chromosome. The relation between *CXorf23* (regulated by genotypes of rs5991441) and metabolic traits deserves future experimental confirmation. The lack of significant *trans*-eQTLs between SNPs on the X chromosome and autosomal genes suggests that regulation between the X chromosome and autosomal genes could be epigenetic (48). Ten per cent of all microRNAs are on X chromosome (49). In addition, recent studies have demonstrated that genes on the sex chromosomes can regulate sex differences in somatic gene expression even without hormonal differences. For example, *SRY* regulates autosomal gene expression by binding directly to DNA (50). Dosage differences of X-linked genes can induce phenotypic sex differences that are likely to be mediated by autosomal genes (51).

Recent studies by Dimas *et al.* (23) and Kent *et al.* (22) formally tested genotype $\times$ sex (or genotype $\times$ age) interaction in 379 lymphoblastoid cell lines and 1240 peripheral mononuclear cell samples, respectively. We were unable to replicate their findings of sex interaction of rs167769 and rs2872507 or age interaction for *UBQLNL*. One possible reason is that eQTLs may be tissue specific (52). Another reason is that detecting interactions may require almost a 10-fold increase in sample size compared with sample sizes needed to study main effects (53,54). Luan *et al.* (55) calculated the power to detect gene–environment interaction and suggested that >2000 observations are needed to detect a moderate interaction (interaction effect size = 0.25

and main effect beta = 0.5 for the common allele), at an allele frequency of 5%. With a sample size of over 5000 people, our data reveal the existence of sex and age interactions of eQTL genotypes with gene expression. We limited our analysis to 11 672 disease-related SNPs from GWAS. Genome-wide analysis of sex and age eQTL interactions may identify stronger effect sizes of genotypes on expression. Our analysis was also restricted to common variants identified in GWAS of complex traits, which are typically affected by multiple genes and environmental factors. It is possible that analysis of rare variants with large effects on disease may show stronger interaction effects of age and sex.

In conclusion, we report results of a comprehensive analysis of genotype $\times$ sex and genotype $\times$ age interactions on gene expression in a large study of 5254 individuals. We identified age or sex interactions for a number of eQTLs that were previously reported to be associated with complex diseases. These interacting SNPs may alter gene expression and influence disease susceptibility or progression in population subgroups. The lack of sex interaction between SNPs on the X chromosome with the expression of autosomal genes is unexpected, suggesting that the effect of sex-specific distant genetic regulation may be weak and epigenetic regulation may play an important role. Given the large number of human diseases that exhibits sex and age differences in onset, prevalence or natural history, this study may provide new insight into the genomic basis of human disease.

MATERIALS AND METHODS

Study sample

The FHS has three generations (cohorts) of participants including an initial cohort of participants (original cohort), their offspring and offspring spouses (offspring cohort) and adult grandchildren (third generation cohort) of the original cohort participants. The three cohorts were recruited and examined beginning in 1948, 1971 and 2002, respectively (56). Collection of blood samples and RNA preparation were described in previous study (57). A total of 5254 participants from the offspring and third generation cohorts had both genome-wide genotyping and gene expression profiling and were included in this study.

Trait-associated and X chromosome SNPs

The NHGRI GWAS catalogue (24) and dbGaP database (25) (accessed 13 Jan 2013) were queried for GWAS SNPs that overlapped with SNPs genotyped or imputed in the FHS. The NHGRI GWAS catalogue used published GWAS studies with at least 100 000 SNPs in the initial stage. The dbGaP archives and distributes the results of studies that have investigated the interaction of the genotype and phenotype, whether published or not. From these two databases, we derived a total of 30 493 SNP-trait associations with 20 565 SNPs and 6856 traits. There were 11 672 trait-associated SNPs at P -values $< 1.0 \times 10^{-5}$ (383 on the X chromosome). Genotyping and quality control methods were described in a previous study (7). These SNPs were tested for eQTLs status and for interaction with age and sex in 5254 FHS participants.

Gene expression

Gene expression in whole blood from FHS participants was assessed with the Affymetrix GeneChip Human Exon 1.0 ST Array, which contains > 5.5 million probes to monitor the expression of $\sim 287\,000$ exons and 18 000 genes. Our analysis focused on a core annotation set of 17 873 genes (17 195 autosomal, 652 on chromosome X and 26 on chromosome Y). Sex-chromosome mapped genes were defined based on Affymetrix gene annotations. All data from the gene expression experiment were normalized using the RMA (robust multi-array average) package to remove any technical or spurious background variation. Linear regression was used to adjust for technical covariates (batches, positive and negative control probes etc.). The first principle component was included in the regression model to adjust for population stratification (58). The complete blood cell count was included in the regression model to adjust cell proportion.

Statistical analysis of eQTLs

Association analysis was conducted using the MatrxQL (59) and kinship package (<http://cran.r-project.org/web/packages/coxme.60>) in R. Testing for association of SNPs with gene expression was conducted using a linear model with interaction terms. SNPs mapping within 1 Mb upstream and downstream of the transcription start site of genes were defined as *cis*-eQTLs. Considering gene expression, denoted by y , the linear model that relates variation in y to SNPs and other covariates is given

by the general formula:

$$y = \mu + \beta_1(\text{SNP}) + \beta_2(\text{Sex}) + \beta_3(\text{Age}) + \beta_4(\text{Sex} \times \text{SNP}) + \beta_5(\text{Age} \times \text{SNP}) + e_f + e_i,$$

where β_1 , β_2 and β_3 are the regression coefficients of the SNP (additive model), sex, and age parameters, respectively. β_4 and β_5 are the regression coefficients of sex- and age-interaction parameters, respectively. We included the correlation matrix e_f of kinship package to account for familial correlation.

The Benjamini–Hochberg FDR (60) was applied to account for multiple testing. For multiple SNPs in strong LD [linkage disequilibrium $R^2 > 0.8$ (61)], we used the SNP with the most significant interaction P -value to represent the SNP cluster. In addition, we also conducted permutation testing. For each significant sex- or age-interaction eQTL (no LD filter), we permuted the sex or age findings for the samples 1000 times and re-ran the interaction test. The permutation FDR was calculated from the proportion of permutation P -values that were less than original P -values. For eQTLs with significant interactions with sex (FDR < 0.05), we then tested association of SNPs with gene expression in each gender separately. For each significant sex interaction, the genotypic effects were grouped into one of three models: Model A: a genetic variation has the same direction of effect but differences are present in the magnitude of effect in both sexes; Model B: genotypic effects are present only in one sex; Model C: genotypic effects are present in both sexes but with opposite directions of effect (Fig. 1). To evaluate sampling effect, we further test the allele frequency difference between man and woman for each eQTLs by the Chi-squared test.

For eQTLs with significant interactions with age, we tested association of gene expression with age in each genotype separately.

Differential gene expression analysis without respect to genotypes

The relations between gene expression and sex or age were tested using a linear regression model with the gene expression level as the dependent variable, sex or age as an explanatory variable. Separate regression models were fitted for each of the full set of 17 873 genes:

$$y = \mu + \beta_1(\text{Sex}) + \beta_2(\text{Age}) + e_f + e_i,$$

where β_1 , and β_2 are the regression coefficients for sex, and age parameters, respectively, and e_f is the correlation matrix from the kinship package to account for familial correlation.

The FDR < 0.05 threshold was used to determine whether a gene was differentially expressed with sex or age.

Internal and external validation

Internal validation was performed by splitting FHS samples into independent discovery and replication sets according to pedigree. There were 3507 discovery samples (1889 women, 1618 men, age range 24–92, mean 54.9 years) and 1694 replication samples (922 women, 772 men, age range 24–90, mean 55.1

years) with no overlapping families in the discovery and replication sets. For significant sex- or age-interaction eQTLs, we tested for significance in both the discovery and replication sets.

Eight hundred and ninety-nine unrelated samples (50% females; age range 18–87, mean age 37 years) from the Estonian gene expression cohort (41) were used for external validation. The sample represents a random population-based cohort from the Estonian Genome Centre, University of Tartu (EGCUT) biobank. Genotyping was performed using Illumina Human 370 CNV array (Illumina, Inc., San Diego, USA). Whole-genome gene-expression levels were obtained using the Illumina Human HT12v3 array (Illumina, Inc., San Diego, USA) (see Supplementary Material, Methods for details of genotype imputation and gene expression normalization).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank all of the study participants who helped to create this valuable resource and support of this work. We thank Chunyu Liu and Brian Chen for advice and discussion. We thank EGCUT personnel, especially Ms M. Hass and Mr V. Soo. EGCUT data analyses were carried out in part in the High Performance Computing Center of the University of Tartu.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by National Institutes of Health (NIH) contract N01-HC-25195. EGCUT received financing by FP7 grants (201 413, 245 536), also received targeted financing from the Estonian Government (SF0180142s08) and direct funding from the Ministries of Research and Science and Social Affairs. EGCUT studies are funded by the University of Tartu in the framework of the Center of Translational Genomics and by the European Union through the European Regional Development Fund, in the framework of the Centre of Excellence in Genomics.

REFERENCES

- Clayton, D.G. (2009) Prediction and interaction in complex disease genetics: experience in type 1 diabetes. *PLoS Genet.*, **5**, e1000540.
- Hunter, D.J. (2005) Gene-environment interactions in human diseases. *Nat. Rev. Genet.*, **6**, 287–298.
- Jousilahti, P., Vartiainen, E., Tuomilehto, J. and Puska, P. (1999) Sex, age, cardiovascular risk factors, and coronary heart disease: a prospective follow-up study of 14 786 middle-aged men and women in Finland. *Circulation*, **99**, 1165–1172.
- Ober, C., Loisel, D.A. and Gilad, Y. (2008) Sex-specific genetic architecture of human disease. *Nat. Rev. Genet.*, **9**, 911–922.
- Wang, C., Cheng, Y., Liu, T., Li, Q., Fillingim, R.B., Wallace, M.R., Staud, R., Kaplan, L. and Wu, R. (2008) A computational model for sex-specific genetic architecture of complex traits in humans: implications for mapping pain sensitivity. *Mol. Pain*, **4**, 13.
- Weiss, L.A., Pan, L., Abney, M. and Ober, C. (2006) The sex-specific genetic architecture of quantitative traits in humans. *Nat. Genet.*, **38**, 218–222.
- Benjamin, A.M., Suchindran, S., Pearce, K., Rowell, J., Lien, L.F., Guyton, J.R. and McCarthy, J.J. (2011) Gene by sex interaction for measures of obesity in the Framingham heart study. *J. Obes.*, **2011**, 329038.
- Heid, I.M., Jackson, A.U., Randall, J.C., Winkler, T.W., Qi, L., Steinthorsdottir, V., Thorleifsson, G., Zillikens, M.C., Speliotes, E.K., Magi, R. *et al.* (2010) Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.*, **42**, 949–960.
- Joubert, B.R., Diao, G., Lin, D., North, K.E. and Franceschini, N. (2009) Longitudinal age-dependent effect on systolic blood pressure. *BMC Proc.*, **3**(Suppl. 7), S87.
- Randall, J.C., Winkler, T.W., Kutalik, Z., Berndt, S.I., Jackson, A.U., Monda, K.L., Kilpelainen, T.O., Esko, T., Magi, R., Li, S. *et al.* (2013) Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.*, **9**, e1003500.
- Lindgren, C.M., Heid, I.M., Randall, J.C., Lamina, C., Steinthorsdottir, V., Qi, L., Speliotes, E.K., Thorleifsson, G., Willer, C.J., Herrera, B.M. *et al.* (2009) Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet.*, **5**, e1000508.
- O'Donnell, C.J., Lindpaintner, K., Larson, M.G., Rao, V.S., Ordovas, J.M., Schaefer, E.J., Myers, R.H. and Levy, D. (1998) Evidence for association and genetic linkage of the angiotensin-converting enzyme locus with hypertension and blood pressure in men but not women in the Framingham Heart Study. *Circulation*, **97**, 1766–1772.
- Fritsche, L.G., Chen, W., Schu, M., Yaspan, B.L., Yu, Y., Thorleifsson, G., Zack, D.J., Arakawa, S., Cipriani, V., Ripke, S. *et al.* (2013) Seven new loci associated with age-related macular degeneration. *Nat. Genet.*, **45**, 433–439, e431–432.
- Rockman, M.V. and Kruglyak, L. (2006) Genetics of global gene expression. *Nat. Rev. Genet.*, **7**, 862–872.
- Farber, C.R. and Lusis, A.J. (2008) Integrating global gene expression analysis and genetics. *Adv. Genet.*, **60**, 571–601.
- Zhang, W., Huang, R.S., Duan, S. and Dolan, M.E. (2009) Gene set enrichment analyses revealed differences in gene expression patterns between males and females. *In Silico Biol.*, **9**, 55–63.
- Yang, X., Schadt, E.E., Wang, S., Wang, H., Arnold, A.P., Ingram-Drake, L., Drake, T.A. and Lusis, A.J. (2006) Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.*, **16**, 995–1004.
- Virts, E.L. and Thoman, M.L. (2010) Age-associated changes in miRNA expression profiles in thymopoiesis. *Mech. Ageing Dev.*, **131**, 743–748.
- Harries, L.W., Hernandez, D., Henley, W., Wood, A.R., Holly, A.C., Bradley-Smith, R.M., Yaghootkar, H., Dutta, A., Murray, A., Frayling, T.M. *et al.* (2011) Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. *Ageing Cell*, **10**, 868–878.
- Sebastiani, P., Solovieff, N., Dewan, A.T., Walsh, K.M., Puca, A., Hartley, S.W., Melista, E., Andersen, S., Dworkis, D.A., Wilk, J.B. *et al.* (2012) Genetic signatures of exceptional longevity in humans. *PLoS One*, **7**, e29848.
- Vinuela, A., Snoek, L.B., Riksen, J.A. and Kammenga, J.E. (2010) Genome-wide gene expression regulation as a function of genotype and age in *C. elegans*. *Genome Res.*, **20**, 929–937.
- Kent, J.W. Jr, Goring, H.H., Charlesworth, J.C., Drigalenko, E., Diego, V.P., Curran, J.E., Johnson, M.P., Dyer, T.D., Cole, S.A., Jowett, J.B. *et al.* (2012) Genotypic interaction in human transcriptional ageing. *Mech. Ageing Dev.*, **133**, 581–590.
- Dimas, A.S., Nica, A.C., Montgomery, S.B., Stranger, B.E., Raj, T., Buil, A., Giger, T., Lappalainen, T., Gutierrez-Arcelus, M., McCarthy, M.I. *et al.* (2012) Sex-biased genetic effects on gene regulation in humans. *Genome Res.*, **22**, 2368–2375.
- Hindorf, L.A., Sethupathy, P., Jenkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA*, **106**, 9362–9367.
- Mailman, M.D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L. *et al.* (2007) The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.*, **39**, 1181–1186.
- Zhang, F.R., Huang, W., Chen, S.M., Sun, L.D., Liu, H., Li, Y., Cui, Y., Yan, X.X., Yang, H.T., Yang, R.D. *et al.* (2009) Genomewide association study of leprosy. *N. Engl. J. Med.*, **361**, 2609–2618.
- Britton, W.J. and Lockwood, D.N. (2004) Leprosy. *Lancet*, **363**, 1209–1219.

28. Schenk, M., Krutzik, S.R., Sieling, P.A., Lee, D.J., Teles, R.M., Ochoa, M.T., Komisopoulou, E., Sarno, E.N., Rea, T.H., Graeber, T.G. *et al.* (2012) NOD2 triggers an interleukin-32-dependent human dendritic cell program in leprosy. *Nat Med*, **18**, 555–563.
29. Sabatti, C., Service, S.K., Hartikainen, A.L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C.G., Zaitlen, N.A., Varilo, T., Kaakinen, M. *et al.* (2009) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat. Genet.*, **41**, 35–46.
30. Waterworth, D.M., Ricketts, S.L., Song, K., Chen, L., Zhao, J.H., Ripatti, S., Aulchenko, Y.S., Zhang, W., Yuan, X., Lim, N. *et al.* (2010) Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.*, **30**, 2264–2276.
31. Gottardo, L., De Cosmo, S., Zhang, Y.Y., Powers, C., Prudente, S., Marescotti, M.C., Trischitta, V., Avogaro, A. and Doria, A. (2008) A polymorphism at the IL6ST (gp130) locus is associated with traits of the metabolic syndrome. *Obesity (Silver Spring)*, **16**, 205–210.
32. Garcia-Tunon, I., Ricote, M., Ruiz, A., Fraile, B., Paniagua, R. and Royuela, M. (2005) IL-6, its receptors and its relationship with bcl-2 and bax proteins in infiltrating and *in situ* human breast carcinoma. *Histopathology*, **47**, 82–89.
33. Kathiresan, S., Manning, A.K., Demissie, S., D'Agostino, R.B., Surti, A., Guiducci, C., Gianniny, L., Burt, N.P., Melander, O., Orho-Melander, M. *et al.* (2007) A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med. Genet.*, **8**(Suppl. 1), S17.
34. Levy, D., Larson, M.G., Benjamin, E.J., Newton-Cheh, C., Wang, T.J., Hwang, S.J., Vasan, R.S. and Mitchell, G.F. (2007) Framingham Heart Study 100 K Project: genome-wide associations for blood pressure and arterial stiffness. *BMC Med. Genet.*, **8**(Suppl. 1), S3.
35. Meigs, J.B., Manning, A.K., Fox, C.S., Florez, J.C., Liu, C., Cupples, L.A. and Dupuis, J. (2007) Genome-wide association with diabetes-related traits in the Framingham Heart Study. *BMC Med. Genet.*, **8**(Suppl. 1), S16.
36. Gilad, Y., Rifkin, S.A. and Pritchard, J.K. (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.*, **24**, 408–415.
37. Stolk, L., Zhai, G., van Meurs, J.B., Verbiest, M.M., Visser, J.A., Estrada, K., Rivadeneira, F., Williams, F.M., Cherkas, L., Deloukas, P. *et al.* (2009) Loci at chromosomes 13, 19 and 20 influence age at natural menopause. *Nat. Genet.*, **41**, 645–647.
38. Hafler, D.A., Compston, A., Sawcer, S., Lander, E.S., Daly, M.J., De Jager, P.L., de Bakker, P.I., Gabriel, S.B., Mirel, D.B., Iverson, A.J. *et al.* (2007) Risk alleles for multiple sclerosis identified by a genomewide study. *N. Engl. J. Med.*, **357**, 851–862.
39. Levy, D., Neuhausen, S.L., Hunt, S.C., Kimura, M., Hwang, S.J., Chen, W., Bis, J.C., Fitzpatrick, A.L., Smith, E., Johnson, A.D. *et al.* (2010) Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. *Proc. Natl. Acad. Sci. USA*, **107**, 9293–9298.
40. Feskanich, D., Cho, E., Schaumberg, D.A., Colditz, G.A. and Hankinson, S.E. (2008) Menopausal and reproductive factors and risk of age-related macular degeneration. *Arch. Ophthalmol.*, **126**, 519–524.
41. Nelis, M., Esko, T., Magi, R., Zimprich, F., Zimprich, A., Toncheva, D., Karachanak, S., Piskackova, T., Balascak, I., Peltonen, L. *et al.* (2009) Genetic structure of Europeans: a view from the North-East. *PLoS ONE*, **4**, e5472.
42. Hughes, T., Adler, A., Merrill, J.T., Kelly, J.A., Kaufman, K.M., Williams, A., Langefeld, C.D., Gilkeson, G.S., Sanchez, E., Martin, J. *et al.* (2012) Analysis of autosomal genes reveals gene-sex interactions and higher total genetic risk in men with systemic lupus erythematosus. *Ann. Rheum. Dis.*, **71**, 694–699.
43. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J. *et al.* (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, **466**, 707–713.
44. Mosca, E., Alfieri, R., Merelli, I., Viti, F., Calabria, A. and Milanesi, L. (2010) A multilevel data integration resource for breast cancer study. *BMC Syst. Biol.*, **4**, 76.
45. Quan, C., Ren, Y.Q., Xiang, L.H., Sun, L.D., Xu, A.E., Gao, X.H., Chen, H.D., Pu, X.M., Wu, R.N., Liang, C.Z. *et al.* (2010) Genome-wide association study for vitiligo identifies susceptibility loci at 6q27 and the MHC. *Nat. Genet.*, **42**, 614–618.
46. Baranzini, S.E., Wang, J., Gibson, R.A., Galwey, N., Naegelin, Y., Barkhof, F., Radue, E.W., Lindberg, R.L., Uitdehaag, B.M., Johnson, M.R. *et al.* (2009) Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum. Mol. Genet.*, **18**, 767–778.
47. Spolarics, Z. (2007) The X-files of inflammation: cellular mosaicism of X-linked polymorphic genes and the female advantage in the host response to injury and infection. *Shock*, **27**, 597–604.
48. Wijchers, P.J. and Festenstein, R.J. (2011) Epigenetic regulation of autosomal gene expression by sex chromosomes. *Trends Genet.*, **27**, 132–140.
49. Pinheiro, I., De Jager, L. and Libert, C. (2011) X-chromosome-located microRNAs in immunity: might they explain male/female differences? The X chromosome-genomic context may affect X-located miRNAs and downstream signaling, thereby contributing to the enhanced immune response of females. *Bioessays*, **33**, 791–802.
50. Sekido, R. and Lovell-Badge, R. (2009) Sex determination and SRY: down to a wink and a nudge? *Trends Genet.*, **25**, 19–29.
51. Heard, E. and Distech, C.M. (2006) Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev.*, **20**, 1848–1867.
52. Hernandez, D.G., Nalls, M.A., Moore, M., Chong, S., Dillman, A., Trabzuni, D., Gibbs, J.R., Ryten, M., Arepalli, S., Weale, M.E. *et al.* (2012) Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. *Neurobiol. Dis.*, **47**, 20–28.
53. Brookes, S.T., Whitley, E., Peters, T.J., Mulheran, P.A., Egger, M. and Davey Smith, G. (2001) Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol. Assess.*, **5**, 1–56.
54. Patsopoulos, N.A., Tatsioni, A. and Ioannidis, J.P. (2007) Claims of sex differences: an empirical assessment in genetic associations. *JAMA*, **298**, 880–893.
55. Luan, J.A., Wong, M.Y., Day, N.E. and Wareham, N.J. (2001) Sample size determination for studies of gene-environment interaction. *Int. J. Epidemiol.*, **30**, 1035–1040.
56. Splansky, G.L., Corey, D., Yang, Q., Atwood, L.D., Cupples, L.A., Benjamin, E.J., D'Agostino, R.B. Sr, Fox, C.S., Larson, M.G., Murabito, J.M. *et al.* (2007) The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am. J. Epidemiol.*, **165**, 1328–1335.
57. Joehanes, R., Johnson, A.D., Barb, J.J., Raghavachari, N., Liu, P., Woodhouse, K.A., O'Donnell, C.J., Munson, P.J. and Levy, D. (2012) Gene expression analysis of whole blood, peripheral blood mononuclear cells, and lymphoblastoid cell lines from the Framingham Heart Study. *Physiol. Genomics*, **44**, 59–75.
58. Zhang, Y., Guan, W. and Pan, W. (2013) Adjustment for population stratification via principal components in association analysis of rare variants. *Genet. Epidemiol.*, **37**, 99–109.
59. Shabalin, A.A. (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, **28**, 1353–1358.
60. Hochberg, Y. and Benjamini, Y. (1990) More powerful procedures for multiple significance testing. *Stat. Med.*, **9**, 811–818.
61. Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J. and de Bakker, P.I. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, **24**, 2938–2939.