European Journal of Human Genetics (2012) 20, 97–101
© 2012 Macmillan Publishers Limited All rights reserved 1018-4813/12
www.nature.com/ejhg

ARTICLE

A systematic eQTL study of *cis–trans* epistasis in 210 HapMap individuals

Jessica Becker[1,2,4], Jens R Wendland[3,4], Britta Haenisch[1,2], Markus M Nöthen[1,2] and Johannes Schumacher*[,2]

We aimed at identifying transcripts whose expression is regulated by a SNP–SNP interaction. Out of 47 294 expression phenotypes we used 3107 transcripts that survived an extensive quality control and 86 613 linkage disequilibrium-pruned SNP markers that have been genotyped in 210 individuals. For each transcript we defined *cis*-SNPs, tested them for epistasis with all *trans*-SNPs, and corrected all observed *cis–trans*-regulated expression effects for multiple testing. We determined that the expression of about 15% of all included transcripts is regulated by a significant two-locus interaction, which is more than expected ($P=2.86\times10^{-144}$). Our findings suggest further that *cis*-markers with so called 'marginal effects' are more likely to be involved in two-locus gene regulation than expected ($P=8.27\times10^{-05}$), although the majority of interacting *cis*-markers showed no one-locus regulation. Furthermore, we found evidence that gene-mediated *trans*-effects are not a major source of epistasis, as no enrichment of genes has been found in close vicinity of *trans*-SNPs. In addition, our data support the notion that neither chromosomal regions nor cellular processes are enriched in epistatic interactions. Finally, some of the *cis–trans* regulated genes have been found in genome-wide association studies, which might be interesting for follow-up studies of the corresponding disorders. In summary, our results provide novel insights into the complex genome-transcriptome regulation.
European Journal of Human Genetics (2012) **20,** 97–101; doi:10.1038/ejhg.2011.156; published online 17 August 2011

Keywords: eQTLs; epistasis; interaction; *cis*-regulation; *trans*-regulation

INTRODUCTION

Mapping studies of gene expression phenotypes have successfully lead to the identification of regulatory variants and networks across the genome.[1–11] In these expression quantitative trait locus (eQTL) analyses, genes have been identified whose expression are regulated by SNP markers, which are either in close proximity to (*cis*-acting SNPs) or at greater distances from the gene locus (*trans*-acting SNPs).[12] Although the nature of *cis*-regulation is influenced by factors such as 5′ promoter- or 3′ transcript-variants, the mechanisms involved in *trans*-regulation include gene-mediated (eg, transcription factors) or sterical interactions such as 'chromosome cross-talk'.[13–16] However, at many gene loci it must be assumed that both, *cis*- and *trans*-effects are involved simultaneously in the regulation of expression. Furthermore, it is possible that expression at certain gene loci is regulated by a more complex process that involves epistasis (eg, *cis–trans* interaction). Unfortunately, these regulatory effects are not detected in one-locus eQTL studies where genetic variants are examined solely. There are two main reasons why two-locus or interaction eQTL mappings have not been applied to existing data. First, potential two-locus effects are difficult to identify and interpret, as substantial correction for multiple testing is required if the interaction was analyzed in a genome-wide fashion. In a genome-wide 100K SNP set, for example, the *P*-value of an observed interaction would have to be in the range of $P=5\times10^{-12}$ per transcript before being considered significant. Second, systematic two-locus eQTL mappings require substantial computational resources, although this limitation has recently been overcome by the introduction of novel biostatistical methods.[17–19]

In the present study we tried to circumvent some of the limitations associated to interaction scans and performed a systematic two-locus eQTL study for epistasis. Out of three possible two-locus interaction models (ie, *cis–cis*, *cis–trans*, *trans–trans*), we restricted our analysis only to *cis–trans* epistasis. We used the expression data of 3107 high-quality transcripts and 86 613 linkage disequilibrium (LD)-pruned SNP markers obtained from 210 HapMap founders. For each transcript, we tested whether expression levels showed statistical epistasis between a locus-specific *cis*- and an interacting *trans*-SNP located elsewhere in the genome. Although other interaction effects may be involved in gene regulation, *cis–trans* interacting effects were investigated as these may be easier to interpret. For example, it is difficult to control for intermarker LD in *cis–cis* or for multiple testing in *trans–trans* interaction studies. A further aim of the study was to characterize identified *cis–trans* interaction effects, for example, to determine whether SNP markers involved in epistatic gene regulation also represent significant one-locus eQTLs.

MATERIALS AND METHODS

Expression data and study sample

For our genome-transcriptome eQTL analysis we used the expression phenotypes that have been generated by The Wellcome Trust Sanger Institute Cambridge (GENEVAR, http://ftp://ftp.sanger.ac.uk/pub/genevar/) from human lymphoblastoid cell lines (LCLs) of all 210 founders in the four International HapMap II populations (http://snp.cshl.org/).[8,9] The sample includes 60 Caucasian individuals (CEU, of northern and western European ancestry), 90 Asian individuals (45 Han Chinese, CHB; and 45 Japanese, JPT), as well as 60 African individuals (YRI, from Nigeria). Although this strategy

[1]Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany; [2]Institute of Human Genetics, University of Bonn, Bonn, Germany; [3]Pharma Research and Early Development, F. Hoffmann-La Roche Ltd, Basel, Switzerland
*Correspondence: Dr J Schumacher, Institute of Human Genetics, University of Bonn, Sigmund-Freud Street 25, Bonn D-53127, Germany. Tel: +49 228 287 51028; Fax: +49 228 287 51011; E-mail: johannes.schumacher@uni-bonn.de
[4]These authors contributed equally to this work.
Received 15 February 2011; revised 30 May 2011; accepted 7 July 2011; published online 17 August 2011

cannot detect interaction effects on gene regulation that are restricted to one particular population, use of the combined sample provides improved statistical power for the detection of epistasis and has been successfully used in previous one-locus eQTL studies.[8,9] In this sample, we used only expression phenotypes for transcripts that were filtered through a detailed and extensive quality control. Of the 47 294 transcripts analyzed using Illumina's human whole genome expression (WG-6 version 1) array (Illumina Inc., San Diego, CA, USA), only those probes that have shown an Illumina detection score of >0.99 in each of the four hybridization experiments conducted across all 210 HapMap individuals were used. These scores were obtained from the Sanger Institute website ('gene_profile-files' at http://ftp://ftp.sanger.ac.uk/pub/genevar/) and reduced the number of transcripts included in the present study to 7978 probes. The respective transcripts could be expected to be robustly expressed in human LCLs. In a subsequent step, the presence of SNPs in the hybridization probes was excluded using the web-based program ReMOAT (version March 2009, http://www.compbio.group.cam.ac.uk/Resources/Annotation/index.html)[20] and the dbSNP 126 database (http://www.ncbi.nlm.nih.gov/projects/SNP/). Although there is a current debate in the field as to whether this step is necessary and other studies have included SNP-containing probes, we decided to exclude them as they possibly might influence the true expression quantity. However, the removal of probes with known coding SNPs did not substantially reduce the number of included transcripts to 6226 probes. Furthermore, we used ReMOAT for the inclusion of probes that are located on autosomes only and mapped over the full length (50 bp) to a contiguous genomic location (ie, no intron-spanning probes). We decided to use exon-specific probes only in order to avoid any inaccurate expression signals, which could be caused by insufficient hybridization to different isoforms of the gene (eg, due to exon-skipping or -incorporation). This step reduced the number of included probes to 5237. Next, the uniqueness of genomic hits for each probe was determined using nuID (https://prod.bioinformatics.northwestern.edu/nuID/), which represents a probe identifier for microarray experiments. This reduced the number of included probes further to 4418 showing a nuID uniqueness score of 100. Only these probes could be specifically mapped to a single Entrez GeneID. Entrez Gene is a repository from the National Center for Biotechnology Information (NCBI) for gene-specific information. In final steps, we filtered for probes whose corresponding transcripts were annotated as 'reviewed' or 'validated' using NM*N*=3124). The RefSeq database provides a collection of annotated sequences including transcripts. When multiple probes hybridized to the same RefSeq NM_ transcript, only one randomly selected probe was included in the analyses. In the final filtering step, the UCSC Browser version HG18 (http://genome.ucsc.edu/cgi-bin/hgGateway) was used to identify probes with defined transcription start and end sites. Exact matches were found for a total of 3107 transcripts, and these were included in the two-locus eQTL analysis. The expression data for each of these 3107 probes were subjected to inverse quantile normalization according to the procedure described by Veyrieras *et al*[10] and the normalized data were saved as PLINK[21] alternate phenotype files. PLINK represents the program that was used for the interaction analysis (see below).

Genotyping data

SNP genotypes of each of the 210 founder individuals were obtained from HapMap release 23 using PLINK.[21] A total of 3.95 million SNPs were available for each individual after exclusion of SNPs with Mendel errors. The Mendel check was performed in the 30 CEU and 30 YRI trios analyzed in the HapMap Project. Next, only SNPs were selected, which were located on autosomes, which had no HWE deviation ($P>0.05$), and which had allele frequencies between 0.2–0.8 as well as a per-SNP genotyping missingness cutoff of 0.02. Although this filtering procedure was done in each of the four populations separately, an LD-pruning step was restricted to the YRI acknowledging the lowest LD structure in this population. Here, a pairwise SNP-SNP-r^2 of 0.8 was used as a pruning criterion. The filtering process resulted in N=86 613 SNPs, which were saved as PLINK binary file for inclusion in the analyses.

Interaction analysis

The two-locus interaction eQTL analysis was performed using the PLINK –epistasis command. For every transcript that corresponded to an included probe, *cis*-SNPs were defined as being variants located within the transcript or <1 Mb apart from the transcription start and end site. Each *cis*-SNP of a transcript was then tested for epistasis with all remaining SNPs, which were defined as *trans*-SNPs (ie, 86 613 SNPs minus the number of *cis*-SNPs per transcript). For the interaction eQTL mapping, the four different HapMap populations were used as categorical co-variates. To determine the significance of our findings, we finally corrected for each transcript all *cis–trans* interaction results by multiplying the number of analyzed *cis*-variants with the number of included *trans*-SNPs. This resulted in transcript-wise Bonferroni-adjusted *P*-values between 5.77×10^{-07} (1 *cis*-SNP and 86 612 *trans*-SNPs for *DNAJA2*, *NETO2* and *ORC6L*) and 2.84×10^{-09} (204 *cis*-SNPs and 86 409 *trans*-SNPs for *CHD8* and *SUPT16H*). Under the null hypothesis of no enrichment for transcripts showing *cis–trans* interactions 0.05*3107=155 transcripts would be expected to have at least one significant *cis–trans* interaction following a transcript-wise Bonferroni's correction. The applied correction procedure is also given in detail in Supplementary Table 1.

RESULTS

Of all 3107 included probes we identified 440 transcripts whose expression was – transcript-wise Bonferroni-adjusted – regulated by a *cis–trans* interaction (Supplementary Table 2). The significant two-locus eQTL *P*-values ranged between 4.69×10^{-08} and 2.82×10^{-12}. The observed interactions showed a significant ($P=2.86\times10^{-144}$) and almost threefold enrichment compared with the number of SNP pairs expected under the null hypothesis, ie 5% of all probes (N=155) would be associated by chance. Table 1 lists the top-16 interaction findings, which were all associated with *P*-values of $<10^{-10}$. Importantly, as an LD-pruning step was applied, all of the 440 *cis–trans* SNP combinations were independent and not the result of LD between *cis*- or *trans*-markers.

To elucidate the nature of the epistasis, an analysis was performed to determine whether SNPs, which are involved in gene regulation via one-locus eQTL effects, mainly contributed to the interactions. At present there is no consensus on whether SNPs with so-called 'marginal effects' are more likely to be involved in epistasis and should be prioritized for SNP–SNP interaction scans. An analysis was therefore performed to determine whether the 440 *cis*- and *trans*-SNPs involved in epistasis also have regulatory effects on gene expression without their interacting markers, that is, in a one-locus fashion. This proved to be true for the *cis*-markers: a total of 40 of the 440 *cis*-SNPs (9.09%) also showed regulatory effects in the one-locus analysis at an uncorrected significance level of $P\leq0.05$. This was significant compared with the expected number of SNPs with marginal effects (N=22, $P=8.27\times10^{-05}$) (Supplementary Table 3). However, it is notable that the majority of *cis*-markers ($>$ 90%) were not involved in gene regulation at the one-locus level.

In contrast, only 16 of the 440 two-locus *trans*-SNPs (3.63%) were involved in gene regulation on the one-locus level. This was not significant compared with the number of expected markers (N=22, P=0.187, Supplementary Table 3) and points to more independent mechanisms involved in the one- and two-locus regulation.

As the mechanisms involved in *trans*-regulation and -epistasis are complex and not well understood, we tried to characterize them in more detail. We analyzed whether the *trans*-epistasis is gene or pathway mediated rather than the result of other regulatory mechanisms and tested at each trans-locus if there are more genes in close vicinity to the marker than expected. Of all 440 *trans*-markers, 198 SNPs (45.10%) were closely located to at least one gene according to the program SNPper (http://snpper.chip.org/bio/snpper-enter), that is, the SNP is located within a distance of ≤10 kb to a corresponding gene (Supplementary Table 2). However, the number of observed genes involved in *trans*-epistasis was not significantly increased compared with the number of all potentially involved genes tagged by all included *trans*-SNPs using SNPper (N=35 731, 41.35%, P=0.112).

Table 1 Column 1 lists the top-16 *cis–trans* interacting transcripts; column 2 shows the number of tested *cis*-SNPs for each transcript; column 3 shows the number of *cis-trans* tests; column 4 list shows the Bonferroni-adjusted *P*-values necessary for a 'significant' finding; column 5 shows the uncorrected *P*-value per transcript obtained in the two-locus interaction analysis; the next columns provide information about the *cis*- and *trans*-SNPs including their eQTL effects under a one-locus model

	No. of	*No. of*			*Top cis-acting SNP*				*Top trans-acting SNP*				
	tested	*epistasis*	*Bonferroni*	*Top two-locus*				*One-locus*				*One-locus*	
Transcript[a]	*cis-SNPs*	*tests*	P-*value*	P-*value*	*rs*	*Chr*	*Position*	P-*value*	*rs*	*Chr*	*Position*	P-*value*	*RefSeq genes*
TRIM4	9	779 436	6.41E-08	2.82E-12	rs1121592	7	99 361 567	2.40E-06	rs457414	3	10 177 884	1.47E-01	*VHL*, *IRAK2*
PNPLA6	77	6 663 272	7.50E-09	5.99E-12	rs608773	19	7 743 306	8.73E-01	rs1794066	2	113 602 821	7.19E-01	*IL1RN*
ARNT	27	2 337 822	2.14E-08	8.26E-12	rs7532008	1	149 226 974	8.68E-01	rs2937504	5	11 015 227	5.14E-01	*CTNND2*, *DAP*
MANBA	51	4 414 662	1.13E-08	1.70E-11	rs4698863	4	103 764 896	1.81E-04	rs13171027	5	4 031 902	5.97E-01	*IRX1*
PHF11	46	3 982 082	1.26E-08	2.08E-11	rs2181539	13	48 569 216	6.52E-01	rs7571794	2	67 969 620	7.46E-01	*ETAA1*
C17orf70	47	4 068 602	1.23E-08	5.10E-11	rs7207933	17	77 131 682	3.75E-07	rs35060330	5	150 818 278	8.38E-01	*SLC36A1*
UEVLD	58	5 020 190	9.96E-09	5.66E-11	rs6483561	11	18 966 071	7.05E-01	rs5743404	8	6 724 531	4.68E-01	*DEFB1*
GMDS	138	11 933 550	4.19E-09	6.06E-11	rs932409	6	1 396 521	6.63E-01	rs2143980	14	32 277 657	3.21E-01	*AKAP6*
CCDC28A	65	5 625 620	8.89E-09	6.74E-11	rs12190319	6	138 316 778	2.29E-01	rs1391285	1	215 628 091	4.90E-01	*ESRRG*, *GPATCH2*
UBTD2	92	7 959 932	6.28E-09	6.76E-11	rs17074786	5	171 791 185	3.32E-01	rs4776794	15	64 659 320	8.57E-02	*LCTL*, *SMAD6*
RNF40	15	1 298 970	3.85E-08	6.77E-11	rs4788213	16	29 942 025	1.23E-01	rs638286	19	55 397 668	2.20E-01	*MYH14*
CCDC88C	83	7 181 990	6.96E-09	7.26E-11	rs2430363	14	91 434 804	2.78E-01	rs2748992	6	52 704 534	3.12E-01	—
GEMIN5	98	8 478 470	5.90E-09	7.35E-11	rs7732085	5	153 693 955	2.72E-01	rs1562797	16	52 900 570	4.55E-01	*IRX3*
EZH2	85	7 354 880	6.80E-09	7.82E-11	rs851704	7	147 169 364	3.97E-01	rs1957190	14	45 567 885	5.58E-01	*RPL10L*
TGDS	70	6 058 010	8.25E-09	9.00E-11	rs7993213	13	94 886 853	6.40E-01	rs13392004	2	48 495 333	2.92E-01	*FOXN2*, *CCDC128*
CEBPZ	87	7 527 762	6.64E-09	9.16E-11	rs12052952	2	36 842 683	4.59E-02	rs807018	10	102 763 001	5.41E-01	*PDZD7*

[a]Illumina probe Ids are available upon request.

Previous one-locus eQTL studies have reported an enrichment of certain chromosomal regions involved in the regulation of gene expression. We adapted the approach of Morley *et al*[6] and analyzed our data for evidence for so-called 'master regulator' SNP-regions on a two-locus interaction level. Master regulator-regions are chromosomal regions that contain more SNPs involved in epistasis than expected by chance. All 86 613 SNPs were used, and the entire autosomal genome was divided into 444 non-overlapping bins, each containing 200 neighboring SNPs. We estimated that a bin, which comprises more than 4 of the 440 *trans*-SNPs, would be a master regulator region. However, correcting this number by a factor of 444, which corresponds to the number of analyzed bins, more than six *trans*-SNPs per bin are necessary for defining a significant master regulator region. Only for bins at the end of chromosomes did we adapt our approach to account for the number of SNPs within these regions. For example, if 100 neighboring SNPs were located within the last bin of a chromosome, more than three *trans*-SNPs were necessary to fulfill the criterion of a significant master regulator region. Although we found 8 out of the 444 bins harboring four *trans*-SNPs, which are nominally significant (P=0.019), no bin fulfilled the criterion of a significant master regulator region after the correction procedure. In addition, our data provide no evidence for superordinated mechanisms involved in epistasis by analyzing whether certain chromosomal 'hotspot' regions harbor more regulated transcripts than expected. We used all 3107 transcripts, divided the autosomal genome into 321 bins, each containing 10 neighboring transcripts, and estimated that a bin with more than 6 of the 440 identified transcripts would be a significant hit. After a correction for the number of analyzed bins (factor 321) no hotspot could be identified, although one bin harbored six transcripts and 12 further bins harbored four transcripts (uncorrected P=0.001 and P=0.041, respectively).

On the functional level, we tested whether certain cellular processes are particularly regulated by epistatic effects. We used all 440 genes that were identified as being *cis–trans* regulated and performed an analysis for enriched cellular functions using Ingenuity Pathways Analysis (IPA, version 8.6, http://www.ingenuity.com). IPA is a web-based interface that provides computational algorithms to identify biological processes and networks on the basis of functional annotation and molecular interactions. The top biological category was 'gene expression', including 69 transcripts. However, the most enriched subcategory 'transcription of chromosome components' (P=0.046 after Benjamini–Hochberg correction) was defined by only 4 of all 440 included transcripts (*CREBBP*, *EP300*, *SRC* and *TBP*). Finally, an analysis was performed to determine whether any of the two-locus regulated genes are implicated in complex disorders. Complex disorders were considered, as genome-wide association studies (GWAS) of a number of diseases have failed to identify any one-locus variants, which are associated with a strong genetic effect size. Two-locus regulation may therefore have an impact on the respective phenotypes. Furthermore, the functional consequence of many top GWAS-SNPs is unknown, which suggests that expression differences may be disease-relevant mechanisms. In total, we identified 25 *cis–trans* regulated genes that have been implicated in complex disorders using the web tool GWAS Catalog (http://genome.gov/26525384). For example (Table 2), we identified a two-locus interaction between a *trans*-SNP 5.9 kb upstream of *CCL4* (MIM 182284) and a *cis*-SNP of *BLK* (MIM 191305) influencing its expression. *BLK* is one of the strongest risk genes for rheumatoid arthritis and systemic lupus erythematosus and *CCL4* encodes a chemokine ligand involved in immune activation.[22–26] However, the connection between *BLK* and *CCL4* remains speculative, as it is unclear whether the close proximity of the *trans*-SNP to *CCL4* reflects a gene- or pathway-mediated mechanism, or whether other interaction mechanisms that do not involve *CCL4* exist. Unfortunately, we could not test the effect of the *trans*-SNP on the expression of *CCL4* because no probe for *CCL4* has been included in our analysis. Another interesting finding concerns *STAT2* (MIM 600556). Its expression was found to be *cis–trans* regulated, and the corresponding *trans*-SNP is located 31.1 kb upstream of *IL23R* (MIM 607562) (Table 2). Again, we could not test whether this SNP is involved in the expression of *IL23R* due to a missing probe, but it is noteworthy

Table 2 Column 1 lists the 25 *cis–trans* interacting transcripts listed in GWAS catalog; column 3 lists the observed two-locus *P*-values; the remaining columns provide information concerning the *cis*- and *trans*-SNPs

		Two-locus	Top cis-acting SNP			Top trans-acting SNP			
Transcript[a]	Disease (GWAS catalog)	P-value	rs	Chr	Position	rs	Chr	Position	RefSeq genes
ELMO1	QT interval[26]	1.09E-10	rs10259008	7	36 799 785	rs776692	15	40 134 818	*PLA2G4D, PLA2G4E*
NIN	Cognitive performance[30]	1.13E-10	rs11850904	14	51 130 033	rs6836445	4	29 296 175	—
ZFP64	Amyotrophic lateral sclerosis[31]	1.90E-09	rs4811201	20	49 629 361	rs6561342	13	46 458 623	*HTR2A, GNG5P5*
GORASP2	Cognitive performance[32]	2.15E-09	rs10930438	2	171 315 117	rs17081840	4	55 718 936	*KDR*
VRK2	Schizophrenia[33]	2.16E-09	rs10178765	2	58 363 538	rs4950076	1	95 349 885	*ALG14, TMEM56*
SYNE1	Blood pressure[34]	2.23E-09	rs1856057	6	152 109 562	rs6445296	3	62 678 612	*CADPS*
C6orf106	Height[35,36]	2.52E-09	rs3800341	6	33 972 976	rs17105347	14	36 335 202	*SLC25A21*
JAK2	Inflammatory bowel disease[37,38]	3.40E-09	rs10974793	9	4 793 651	rs12475354	2	77 441 949	*LRRTM4*
WDR1	Serum urate (cardiovascular disease)[39,40]	3.53E-09	rs7660895	4	9 594 543	rs10085762	7	135 220 728	—
CXXC1	Chronic lymphocytic leukemia[41]	3.63E-09	rs1705521	18	45 955 763	rs11836262	12	8 772 935	*FAM80B*
AP1B1	Carotid atherosclerosis[42]	3.83E-09	rs4822998	22	27 690 297	rs2753596	14	38 712 591	*TRAPPC6B*
ST6GAL1	Drug-induced liver injury[43]	4.11E-09	rs3872724	3	188 223 915	rs1959205	14	43 877 663	*YWHAZP1*
PEX1	Height[44]	4.66E-09	rs2285504	7	92 825 257	rs7034789	9	6 935 423	*JMJD2C*
EXT1	Height[45]	4.95E-09	rs7006088	8	119 720 982	rs6696976	1	97 701 564	*DPYD*
BLK	Systemic lupus erythematosus[22,24,25], rheumatoid arthritis[23]	5.30E-09	rs1293320	8	11 729 348	rs1634506	17	31 449 476	*CCL3, CCL4*
WDR36	Plasma eosinophil count (asthma)[46]	5.47E-09	rs27409	5	111 459 912	rs9504183	6	4 605 997	—
FNTB	Mean corpuscular volume[47]	5.96E-09	rs1679880	14	64 723 379	rs7165654	15	56 627 331	*LIPC*
TSR1	Aortic root size[48]	6.70E-09	rs1109303	17	1 350 227	rs1334751	10	29 057 579	*BAMBI*
PRDM1	Systemic lupus erythematosus[24]	7.23E-09	rs1891720	6	107 259 564	rs2993312	13	112 731 466	*MCF2L*
MBD1	Chronic lymphocytic leukemia[41]	8.44E-09	rs1705521	18	45 955 763	rs11836262	12	8 772 935	*FAM80B*
METTL1	Multiple sclerosis[49]	8.68E-09	rs1908536	12	57 124 955	rs4833611	4	120 366 908	*USP53*
LSP1	Breast cancer[50]	8.91E-09	rs2301160	11	1 053 767	rs10930873	2	152 549 752	*CACNB4*
LDLR	Myocardial infarction[51], LDL cholesterol[52–54]	9.08E-09	rs11085720	19	10 178 763	rs6445704	3	54 614 308	*CACNA2D3*
STAT2	Psoriasis[27]	1.20E-08	rs4495925	12	55 554 383	rs10489631	1	67 373 703	*IL23R*
UBE2L3	Systemic lupus erythematosus[24]	4.29E-08	rs165846	22	19 254 028	rs5751963	22	23 462 498	*PIWIL3*

[a]Illumina probe Ids are available upon request.

that both genes have an important role in the innate immune system and have been implicated in the development of psoriasis in a recently published GWAS.[27–29]

DISCUSSION

Genes function through a complex mechanism that involves multiple genetic factors. These effects are missed if genetic factors are examined in isolation without taking potential interactions with other genetic factors into account. The aim of the present study was to elucidate the genetic architecture of gene expression through the performance of a systematic *cis–trans* interaction analysis. Out of 47 294 expression phenotypes, we used 3107 transcripts that survived a stringent quality control procedure and 86 613 LD-pruned SNP markers, which were in linkage equilibrium and have been genotyped in the 210 HapMap founder individuals. Using a conservative correction procedure, we identified that the expression of about 15% of all included transcripts (N=440) is regulated by a two-locus interaction, which is far more than expected by chance ($P=2.86\times10^{-144}$). The results of the present study confirm that epistasis has an important role in the genetic architecture of complex phenotypes and imply that this approach may be of relevance to other eQTL and GWAS data sets. Such studies could also benefit from samples that are ethnically more homogeneous. Although we have used four different populations as categorical covariates, we cannot completely rule out that our results are to a certain degree inflated by the heterogeneity of the present sample.

The present findings also indicate that regulatory one-locus *cis*-markers are more likely to be involved in two-locus gene regulation than would be expected by chance alone ($P=8.27\times10^{-05}$). This suggests that there is a correlation between the mechanisms, which underlie one- and two-locus gene regulation. However, as the majority of *cis*-markers involved in epistasis showed no 'marginal effects', our findings imply that most epistasis effects would be missed if interaction studies were focused on *cis*-markers with marginal effects only.

Furthermore, the present results indicate that gene- or pathway-mediated *trans*-effects were not the major source of epistasis, as *trans*-SNPs were not more likely to be located in or in close proximity to an annotated gene or transcript (P=0.112). Therefore, other regulatory mechanisms, such as non-coding sequence-mediated effects (eg, RNA) and intra- or interchromosomal cross-talk, seem to be of equal importance in *trans*-epistatic regulation.

Our analyses as to whether particular chromosomal regions are involved in epistasis produced negative results ($P>0.05$ for master regulators and hotspots). This implies that *cis–trans* epistasis is not 'topographically' organized throughout the genome. In addition, the IPA analysis revealed that only one functional category (involving only four transcripts) was enriched for epistatic effects (P=0.046 for the subcategory 'transcription of chromosome components' within the high-level category 'gene expression'). This suggests that multiple cellular processes are regulated by two-locus interactions rather than specific ones. Furthermore, 25 of all *cis–trans*-regulated genes have been found to be associated with complex diseases through GWAS. The *trans*-markers and -genes identified in the present study may therefore represent interesting candidates for epistatic tests in the respective GWAS data.

In conclusion, the present *cis–trans* interaction approach identified transcripts, which are potentially influenced by a two-locus epistasis, and yielded certain characteristics of the complex process of genome-transcriptome regulation. Furthermore, the approach may represent a solution for overcoming the problem of multiple testing in interaction scans, and it may thus be worthwhile to apply this approach to other eQTL data. A limitation of this approach, however, is that it is only able to detect *cis–trans* epistasis and cannot be used to detect other regulation mechanisms such as *cis–cis*, *trans–trans* or higher-order interactions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

JS was supported by a NIH/DFG Research Career Transition Award, and MMN was supported by the Alfried Krupp von Bohlen und Halbach-Stiftung. We are grateful to all of the scientists at The Wellcome Trust Sanger Institute in Cambridge who were involved in generating the expression data, and to all of the scientists from the HapMap Consortium who were involved in generating the genotypic data used in the present study.

1 Dimas AS, Deutsch S, Stranger BE *et al*: Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 2009; **325**: 1246–1250.
2 Dixon AL, Liang L, Moffatt MF *et al*: A genome-wide association study of global gene expression. *Nat Genet* 2007; **39**: 1202–1207.
3 Dombroski BA, Nayak RR, Ewens KG, Ankener W, Cheung VG, Spielman RS: Gene expression and genetic variation in response to endoplasmic reticulum stress in human cells. *Am J Hum Genet* 2010; **86**: 719–729.
4 Goring HH, Curran JE, Johnson MP *et al*: Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet* 2007; **39**: 1208–1216.
5 Idaghdour Y, Czika W, Shianna KV *et al*: Geographical genomics of human leukocyte gene expression variation in southern Morocco. *Nat Genet* 2010; **42**: 62–67.
6 Morley M, Molony CM, Weber TM *et al*: Genetic analysis of genome-wide variation in human gene expression. *Nature* 2004; **430**: 743–747.
7 Myers AJ, Gibbs JR, Webster JA *et al*: A survey of genetic human cortical gene expression. *Nat Genet* 2007; **39**: 1494–1499.
8 Stranger BE, Forrest MS, Dunning M *et al*: Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007; **315**: 848–853.
9 Stranger BE, Nica AC, Forrest MS *et al*: Population genomics of human gene expression. *Nat Genet* 2007; **39**: 1217–1224.
10 Veyrieras JB, Kudaravalli S, Kim SY *et al*: High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet* 2008; **4**: e1000214.
11 Cheung VG, Nayak RR, Wang IX *et al*: Polymorphic cis- and trans-regulation of human gene expression. *PLoS Bio* 2010; **8**.
12 Cheung VG, Spielman RS: Genetics of human gene expression: mapping DNA variants that influence gene expression. *Nat rev* 2009; **10**: 595–604.
13 Fraser P, Bickmore W: Nuclear organization of the genome and the potential for gene regulation. *Nature* 2007; **447**: 413–417.
14 Gondor A, Ohlsson R: Chromosome crosstalk in three dimensions. *Nature* 2009; **461**: 212–217.
15 Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA: Interchromosomal associations between alternatively expressed loci. *Nature* 2005; **435**: 637–645.
16 Williams A, Spilianakis CG, Flavell RA: Interchromosomal association and gene regulation in trans. *Trends Genet* 2010; **26**: 188–197.
17 Herold C, Steffens M, Brockschmidt FF, Baur MP, Becker T: INTERSNP: genome-wide interaction analysis guided by *a priori* information. *Bioinformatics* 2009; **25**: 3275–3281.
18 Steffens M, Becker T, Sander T *et al*: Feasible and successful: genome-wide interaction analysis involving all $1.9\times10(11)$ pair-wise interaction tests. *Hum Hered* 2010; **69**: 268–284.
19 Wan X, Yang C, Yang Q *et al*: BOOST: A fast approach to detecting gene-gene interactions in genome-wide case-control studies. *Am J Hum Genet* 2010; **87**: 325–340.
20 Barbosa-Morais NL, Dunning MJ, Samarajiwa SA *et al*: A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Res* 2010; **38**: e17.
21 Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
22 Graham RR, Cotsapas C, Davies L *et al*: Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet* 2008; **40**: 1059–1061.
23 Gregersen PK, Amos CI, Lee AT *et al*: REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat Genet* 2009; **41**: 820–823.
24 Han JW, Zheng HF, Cui Y *et al*: Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet* 2009; **41**: 1234–1237.
25 Hom G, Graham RR, Modrek B *et al*: Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med* 2008; **358**: 900–909.
26 Marroni F, Pfeufer A, Aulchenko YS *et al*: A genome-wide association scan of RR and QT interval duration in 3 European genetically isolated populations: the EUROSPAN project. *Circ Cardiovasc Genet* 2009; **2**: 322–328.
27 Nair RP, Duffin KC, Helms C *et al*: Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. *Nat Genet* 2009; **41**: 199–204.
28 Zhernakova A, van Diemen CC, Wijmenga C: Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat Rev* 2009; **10**: 43–55.
29 Cordell HJ: Detecting gene-gene interactions that underlie human diseases. *Nat Rev* 2009; **10**: 392–404.
30 Cirulli ET, Kasperaviciute D, Attix DK *et al*: Common genetic variation and performance on standardized cognitive tests. *Eur J Hum Genet* 2010; **18**: 815–820.
31 Schymick JC, Scholz SW, Fung HC *et al*: Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol* 2007; **6**: 322–328.
32 Need AC, Attix DK, McEvoy JM *et al*: A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. *Hum Mol Genet* 2009; **18**: 4650–4661.
33 Stefansson H, Ophoff RA, Steinberg S *et al*: Common variants conferring risk of schizophrenia. *Nature* 2009; **460**: 744–747.
34 Levy D, Larson MG, Benjamin EJ *et al*: Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. *BMC Med Genet* 2007; **8**(Suppl 1): S3.
35 Soranzo N, Rivadeneira F, Chinappen-Horsley U *et al*: Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS Genet* 2009; **5**: e1000445.
36 Weedon MN, Lango H, Lindgren CM *et al*: Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 2008; **40**: 575–583.
37 Asano K, Matsushita T, Umeno J *et al*: A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the Japanese population. *Nat Genet* 2009; **41**: 1325–1329.
38 Barrett JC, Hansoul S, Nicolae DL *et al*: Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008; **40**: 955–962.
39 McArdle PF, Parsa A, Chang YP *et al*: Association of a common nonsynonymous variant in GLUT9 with serum uric acid levels in old order amish. *Arthritis Rheum* 2008; **58**: 2874–2881.
40 Wallace C, Newhouse SJ, Braund P *et al*: Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am J Hum Genet* 2008; **82**: 139–149.
41 Crowther-Swanepoel D, Broderick P, Di Bernardo MC *et al*: Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. *Nat Genet* 2010; **42**: 132–136.
42 Shrestha S, Irvin MR, Taylor KD *et al*: A genome-wide association study of carotid atherosclerosis in HIV-infected men. *Aids* 2010; **24**: 583–592.
43 Daly AK, Donaldson PT, Bhatnagar P *et al*: HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat Genet* 2009; **41**: 816–819.
44 Gudbjartsson DF, Walters GB, Thorleifsson G *et al*: Many sequence variants affecting diversity of adult human height. *Nat Genet* 2008; **40**: 609–615.
45 Kim JJ, Lee HI, Park T *et al*: Identification of 15 loci influencing height in a Korean population. *J Hum Genet* 2010; **55**: 27–31.
46 Gudbjartsson DF, Bjornsdottir US, Halapi E *et al*: Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet* 2009; **41**: 342–347.
47 Ganesh SK, Zakai NA, van Rooij FJ *et al*: Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet* 2009; **41**: 1191–1198.
48 Vasan RS, Glazer NL, Felix JF *et al*: Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *JAMA* 2009; **302**: 168–178.
49 MS-consortium: Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat Genet* 2009; **41**: 824–828.
50 Easton DF, Pooley KA, Dunning AM *et al*: Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007; **447**: 1087–1093.
51 Kathiresan S, Voight BF, Purcell S *et al*: Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 2009; **41**: 334–341.
52 Kathiresan S, Melander O, Guiducci C *et al*: Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 2008; **40**: 189–197.
53 Sabatti C, Service SK, Hartikainen AL *et al*: Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* 2009; **41**: 35–46.
54 Willer CJ, Sanna S, Jackson AU *et al*: Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008; **40**: 161–169.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)