

Detection and replication of epistasis influencing transcription in humans

Gibran Hemani^{1,2}, Konstantin Shakhbazov^{1,2}, Harm-Jan Westra³, Tonu Esko^{4,5,6}, Anjali K. Henders⁷, Allan F. McRae^{1,2}, Jian Yang¹, Greg Gibson⁸, Nicholas G. Martin⁷, Andres Metspalu⁴, Lude Franke³, Grant W. Montgomery^{7*}, Peter M. Visscher^{1,2*} & Joseph E. Powell^{1,2*}

Epistasis is the phenomenon whereby one polymorphism's effect on a trait depends on other polymorphisms present in the genome. The extent to which epistasis influences complex traits¹ and contributes to their variation^{2,3} is a fundamental question in evolution and human genetics. Although often demonstrated in artificial gene manipulation studies in model organisms^{4,5}, and some examples have been reported in other species⁶, few examples exist for epistasis among natural polymorphisms in human traits^{7,8}. Its absence from empirical findings may simply be due to low incidence in the genetic control of complex traits^{2,3}, but an alternative view is that it has previously been too technically challenging to detect owing to statistical and computational issues⁹. Here we show, using advanced computation¹⁰ and a gene expression study design, that many instances of epistasis are found between common single nucleotide polymorphisms (SNPs). In a cohort of 846 individuals with 7,339 gene expression levels measured in peripheral blood, we found 501 significant pairwise interactions between common SNPs influencing the expression of 238 genes ($P < 2.91 \times 10^{-16}$). Replication of these interactions in two independent data sets^{11,12} showed both concordance of direction of epistatic effects ($P = 5.56 \times 10^{-31}$) and enrichment of interaction P values, with 30 being significant at a conservative threshold of $P < 9.98 \times 10^{-5}$. Forty-four of the genetic interactions are located within 5 megabases of regions of known physical chromosome interactions¹³ ($P = 1.8 \times 10^{-10}$). Epistatic networks of three SNPs or more influence the expression levels of 129 genes, whereby one *cis*-acting SNP is modulated by several *trans*-acting SNPs. For example, *MBNL1* is influenced by an additive effect at rs13069559, which itself is masked by *trans*-SNPs on 14 different chromosomes, with nearly identical genotype-phenotype maps for each *cis-trans* interaction. This study presents the first evidence, to our knowledge, for many instances of segregating common polymorphisms interacting to influence human traits.

In the genetic analysis of complex traits it is usual for SNP effects to be estimated using an additive model in which they are assumed to contribute independently and cumulatively to the mean of a trait. This framework has been successful in identifying thousands of associations¹⁴. But to date, although its contribution to phenotypic variance is frequently the subject of debate¹⁻³, there is little empirical exploration of the role that epistasis has in the architecture of complex traits in humans^{7,8}. Beyond the prism of human association studies there is evidence for epistasis, not only at the molecular scale from artificially induced mutations⁴ but also at the evolutionary scale in fitness adaptation¹⁵ and speciation¹⁶.

Methods are now available to overcome the computational problems involved in searching for epistasis, but its detection still remains problematic owing to reduced statistical power. For example, increased dependence on linkage disequilibrium between causal SNPs and observed SNPs^{17,18},

increased model complexity in fitting interaction terms¹⁹, and more extreme significance thresholds to account for increased multiple testing⁹ all make it more difficult to detect epistasis in comparison to additive effects. Thus, with small genetic effect sizes, as is expected in most complex traits of interest¹⁴, the power to detect epistasis diminishes rapidly. There are two simple ways to overcome this problem. One is by using extremely large sample sizes²⁰; another is by analysing traits that are likely to have large effect sizes among common variants. Because our focus was to ascertain the extent to which instances of epistasis arises from natural genetic variation, we designed a study around the latter approach and searched for epistatic genetic effects that influence gene expression levels. Transcription levels can be measured for thousands of genes and like most complex diseases, these expression traits are typically heritable²¹. But unlike complex diseases, genetic associations with gene expression commonly have very large effect sizes that explain large proportions of the genetic variance²², making them good candidates to search for epistasis, should it exist.

In our discovery data set (Brisbane Systems Genetics Study, BSGS²³) of 846 individuals genotyped at 528,509 SNPs, we used a two-stage approach to identify genetic interactions. First, we exhaustively tested every pair of SNPs for pairwise effects against each of 7,339 expression traits in peripheral blood (1.03×10^{15} statistical tests, family-wise error rate of 5% corresponding to a significance threshold of $P < 2.91 \times 10^{-16}$, Methods). Second, we filtered the SNP pairs from stage 1 on linkage disequilibrium and genotype class counts, and tested the remaining pairwise effects for significant interaction terms and used a Bonferroni correction for multiple testing (estimated type 1 error rate $0.05 \leq \alpha \leq 0.14$, Methods and Supplementary Fig. 1). Using this design we identified 501 putative genetic interactions influencing the expression levels of 238 genes (Supplementary Table 1). We used strict quality control measures to avoid statistical associations being driven by technical artefacts (Methods). However, it remains possible that unexplained technical artefacts may have led to the significant discovery interactions. Of the 501 discovery interactions, 434 had available data and passed filtering (Methods) in two independent replication data sets, Fehrmann¹² and the Estonian Genomics Centre University of Tartu (EGCUT)¹¹, in which we saw convincing evidence for replication. We used the summary statistics from the replication data sets to perform a meta-analysis to obtain an independent P value for the putative interactions, and 30 were significant after applying a Bonferroni correction for multiple testing (5% significance threshold $P < 9.98 \times 10^{-5}$, Table 1). To quantify the similarity of genotype-phenotype maps between the independent data sets (Fig. 1) we decomposed the genetic effects of each of the SNP pairs into orthogonal additive, dominance and epistatic effects ($A1, A2, D1, D2, A1 \times A2, A1 \times D2, D1 \times A2, D1 \times D2$) and tested for concordance of the sign of the most statistically significant effect (Supplementary

¹Queensland Brain Institute, University of Queensland, Brisbane, Queensland 4072, Australia. ²University of Queensland Diamantina Institute, University of Queensland, Princess Alexandra Hospital, Brisbane, Queensland 4072, Australia. ³Department of Genetics, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9700 RB Groningen, The Netherlands. ⁴Estonian Genome Center, University of Tartu, Tartu 51010, Estonia. ⁵Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA. ⁶Divisions of Endocrinology, Children's Hospital, Boston, Massachusetts 02115, USA. ⁷Queensland Institute of Medical Research, Brisbane, Queensland 4006, Australia. ⁸School of Biology and Centre for Integrative Genomics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.

*These authors contributed equally to this work.

Table 1 | Epistatic interactions significant at the Bonferroni level in two replication sets

	Gene (chr.)	SNP 1 (chr.)	SNP 2 (chr.)	BSGS†	Fehrmann‡	EGCUT‡	Meta§
1	<i>ADK</i> (10)	rs2395095 (10)	rs10824092 (10)	6.69*	18.33*	21.21*	39.82*
2	<i>ATP13A1</i> (19)	rs4284750 (19)	rs873870 (19)	5.30	12.18	3.25	14.23
3	<i>C21ORF57</i> (21)	rs9978658 (21)	rs11701361 (21)	9.42	6.08	16.36	21.67
4	<i>CSTB</i> (21)	rs9979356 (21)	rs3761385 (21)	11.99	25.20	16.72	42.27
5	<i>CTSC</i> (11)	rs7930237 (11)	rs556895 (11)	7.16	18.76	15.06	33.53
6	<i>FN3KRP</i> (17)	rs898095 (17)	rs9892064 (17)	16.16	28.24	29.39	59.95
7	<i>GAA</i> (17)	rs11150847 (17)	rs12602462 (17)	13.91	19.98	12.99	32.60
8	<i>HNRPH1</i> (5)	rs6894268 (5)	rs4700810 (5)	15.38	8.55	3.01	10.37
9	<i>LAX1</i> (1)	rs1891432 (1)	rs10900520 (1)	19.16	18.60	11.22	29.24
10	<i>MBNL1</i> (3)	rs16864367 (3)	rs13079208 (3)	13.49	16.25	24.74	41.56
11	<i>MBNL1</i> (3)	rs7710738 (5)	rs13069559 (3)	7.92	2.55	7.89	9.28
12	<i>MBNL1</i> (3)	rs2030926 (6)	rs13069559 (3)	7.10	0.91	5.80	5.53
13	<i>MBNL1</i> (3)	rs2614467 (14)	rs13069559 (3)	5.74	4.13	2.22	5.30
14	<i>MBNL1</i> (3)	rs218671 (17)	rs13069559 (3)	7.63	0.62	5.82	5.23
15	<i>MBNL1</i> (3)	rs11981513 (7)	rs13069559 (3)	7.71	0.43	5.36	4.58
16	<i>MBP</i> (18)	rs8092433 (18)	rs4890876 (18)	5.40	7.06	21.91	28.73
17	<i>NAPRT1</i> (8)	rs2123758 (8)	rs3889129 (8)	8.45	15.12	16.08	30.77
18	<i>NCL</i> (2)	rs7563453 (2)	rs4973397 (2)	7.31	7.51	6.33	12.70
19	<i>PRMT2</i> (21)	rs2839372 (21)	rs11701058 (21)	4.81	0.69	4.47	4.06
20	<i>RPL13</i> (16)	rs352935 (16)	rs2965817 (16)	4.98	3.79	14.41	17.24
21	<i>SNORD14A</i> (11)	rs2634462 (11)	rs6486334 (11)	7.31	13.11	10.96	23.22
22	<i>TMEM149</i> (19)	rs807491 (19)	rs7254601 (19)	12.16	81.55	45.78	145.78
23	<i>TMEM149</i> (19)	rs8106959 (19)	rs6926382 (6)	5.80	3.06	8.80	10.72
24	<i>TMEM149</i> (19)	rs8106959 (19)	rs914940 (1)	6.22	3.36	6.96	9.20
25	<i>TMEM149</i> (19)	rs8106959 (19)	rs2351458 (4)	7.30	0.04	9.61	8.00
26	<i>TMEM149</i> (19)	rs8106959 (19)	rs6718480 (2)	8.55	3.31	5.15	7.36
27	<i>TMEM149</i> (19)	rs8106959 (19)	rs1843357 (8)	6.21	3.72	3.33	6.00
28	<i>TMEM149</i> (19)	rs8106959 (19)	rs9509428 (13)	9.44	0.10	5.75	4.47
29	<i>TRA2A</i> (7)	rs7776572 (7)	rs11770192 (7)	8.23	3.19	1.89	4.09
30	<i>VASP</i> (19)	rs1264226 (19)	rs2276470 (19)	5.09	0.94	5.14	4.95

* $-\log_{10} P$ values for 4 d.f. interaction tests

† Discovery data set

‡ Independent replication data set

§ Meta-analysis of interaction terms between replication data sets only

Table 3 and Methods). Sign concordance between the discovery and both replication data sets was observed in 22 out of the 30 significantly replicated interactions (expected value = 7.5 under the null hypothesis of no interactions, $P = 3.76 \times 10^{-8}$).

In addition, using the meta-analysis from the replication samples only, we observed that 316 of the remaining 404 discovery SNP pairs had replication interaction P values more extreme than the 2.5% confidence interval of the quantile–quantile plot against the null hypothesis of no interactions in which P values are assumed to be uniformly distributed ($P < 1.0 \times 10^{-16}$, Fig. 2 and Supplementary Fig. 2). Concordance of the direction of the effect of the largest variance component was also highly significant ($P = 5.71 \times 10^{-31}$, Supplementary Table 3). The congruence of the epistatic networks in discovery and replication data sets is shown in Fig. 3, demonstrating that these complex genetic patterns are common even across independent data sets. A further replication was attempted using the Centre for Health Discovery and Wellbeing (CHDWB) data set²⁴, but only 20 of the SNP pairs passed filtering because the sample size was small ($n = 139$), and probably because of insufficient power we found no evidence for replication (Supplementary Fig. 6). It should be noted that although it is a necessary step to establish the veracity of the interactions from the discovery set, replication of epistatic effects in independent samples is difficult in practice due to linkage disequilibrium (Methods).

Although seldom the focus of association studies, SNPs with known main effects are often tested for $A \times A$ genetic interactions⁹, but our analysis suggests that this is unlikely to be the best strategy for its detection. The majority of our discovery interactions were composed of one SNP that was significantly associated with the gene expression level in the discovery data set, and one SNP that had no previous association²² (439 out of 501, Methods). Only nine interactions were between SNPs that both had known main effects, whereas 64 were between SNPs that had no known main effects. Additionally, we observed that the largest epistatic variance component for the 501 interactions was equally divided amongst $A \times A$, $A \times D$, $D \times A$ and $D \times D$ at the

discovery stage ($P = 0.22$ for departure from expectation). This is not surprising because these patterns of epistasis used for statistical decomposition are simply convenient orthogonal parameterizations of a two-locus model, and are not intended to model biological function²⁵.

Of the discovery interactions, 26 were *cis-cis* acting (within 1 megabase (Mb) of the transcription start site, mean distance between SNPs was 0.53 Mb), 462 were *cis-trans*-acting and 13 were *trans-trans*-acting. We observed a wide range of significant genotype–phenotype maps (Fig. 1) but the most common pattern of epistasis that we detected involved a *trans*-SNP masking the effect of an additive *cis*-SNP. For example, *MBNL1* (involved in RNA modification and regulation of splicing²⁶) has a *cis* effect at rs13069559 which in turn is controlled by 13 *trans*-SNPs and one *cis*-SNP that each exhibit a masking pattern, such that when the *trans*-SNP is homozygous for the masking allele the decreasing allele of the *cis*-SNP no longer has an effect (Supplementary Fig. 10). Each of these interactions has evidence for replication in at least one data set and six are significantly replicated at the Bonferroni level (Supplementary Fig. 3). We see similar epistatic networks involving multiple (eight or more) *trans*-acting SNPs for other gene expression levels too, for example *TMEM149* (also known as *IGFLR1*) (Supplementary Fig. 11), *NAPRT1* (Supplementary Fig. 12), *TRAPPC5* (Supplementary Fig. 13) and *CAST* (Supplementary Fig. 14). We observed from pedigree analysis that these five gene expression phenotypes had non-additive variance component estimates within the 95th percentile of the 17,994 gene expression phenotypes that were analysed previously²² (Supplementary Table 2 and Methods).

In total the 501 interactions comprised 781 unique SNPs, which we analysed for functional enrichment (Methods). We tested the SNPs for cell-type specific overlap with transcriptionally active chromatin regions, tagged by histone-3-lysine-4, tri-methylation (H3K4me3) chromatin marks, in 34 cell types²⁷ (Supplementary Fig. 5). There was significant enrichment for *cis*-acting SNPs in haematopoietic cell types only ($P < 1 \times 10^{-4}$ for the three tissues with the strongest enrichment after adjusting for multiple testing). However *trans*-acting SNPs did not

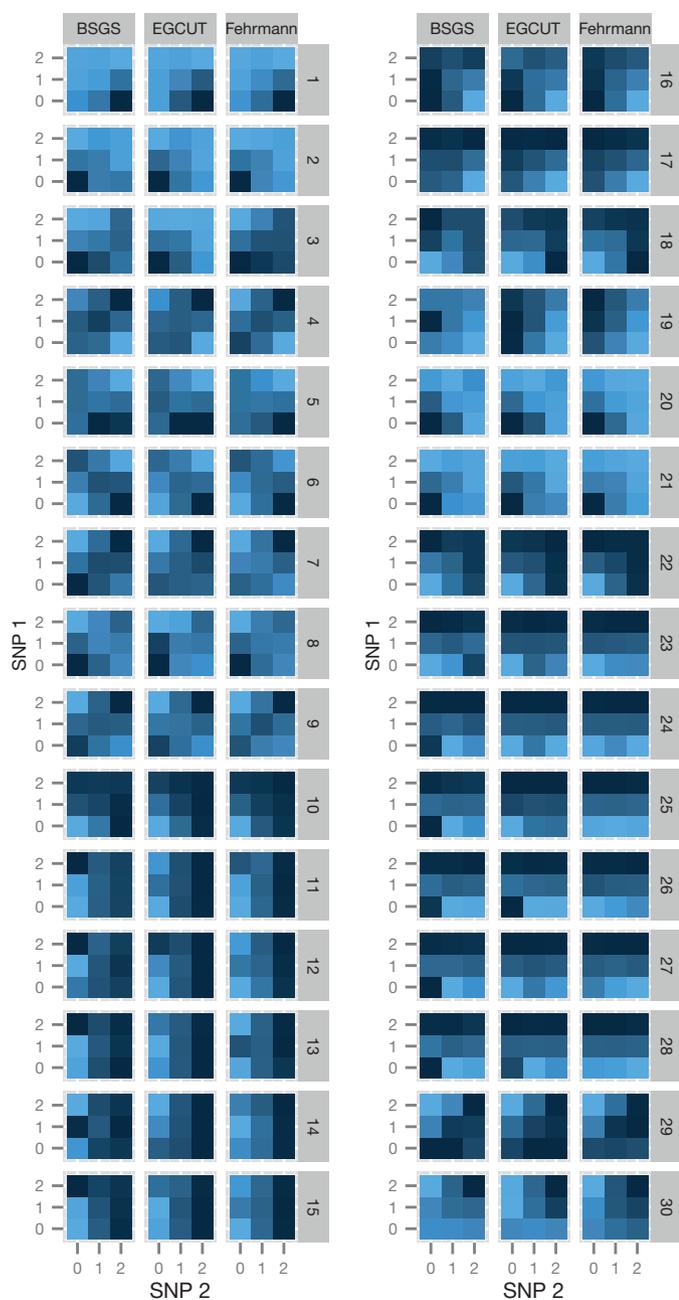


Figure 1 | Replication of genotype-phenotype maps in two independent populations. The genotype-phenotype maps for each epistatic interaction that is significant at the Bonferroni level in both replication data sets are shown. Each genotype-phenotype map consists of nine tiles where each tile represents the expression level for that two-locus genotype class. Phenotypes are for gene transcript levels (dark coloured tiles = high expression, light coloured tiles = low expression). Columns of genotype-phenotype maps are for each independent data set. Rows of genotype-phenotype maps are for each of 30 significantly replicated interactions at the Bonferroni level, corresponding to the rows in Table 1. There is a clear trend of the genotype-phenotype maps replicating across all three data sets.

show any tissue specific enrichment ($P > 0.1$ for all tissues). This difference between *cis* and *trans* SNPs suggests different roles in epistatic interactions where tissue specificity is provided by the *cis*-SNPs. There is also enrichment for *cis*-SNPs to be localized in regions with regulatory genomic features as measured by chromatin states²⁸ (Supplementary Fig. 4).

We also demonstrate physical organization of interacting loci within the cell, suggesting a mechanism by which biological function can lead to

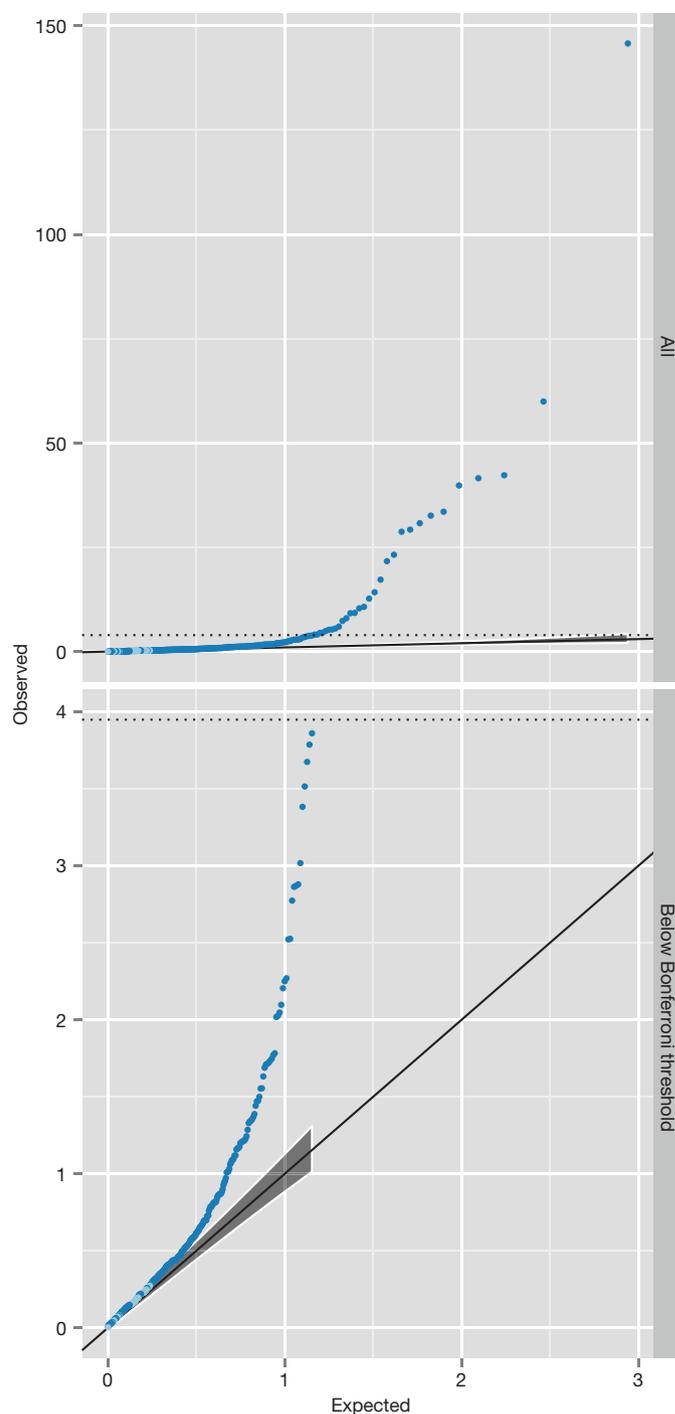


Figure 2 | Q-Q plots of interaction P values from replication data sets.

The top panel shows all 434 discovery SNPs that were tested for interactions. Observed P values (y axis, $-\log_{10}$ scale) are plotted against the expected P values (x axis, $-\log_{10}$ scale). The multiple testing correction threshold for significance following Bonferroni correction is denoted by a dotted line. The bottom panel shows the same data as the top panel but excluding the 30 interactions that were significant at the Bonferroni level in the replication data sets. The shaded grey area represents the 5% confidence interval for the expected distribution of P values. Dark blue points represent P values that exceed the confidence interval, light blue are within the confidence interval.

epistatic genetic variance. It has been shown that different chromosomal regions spatially co-localize in the cell through chromatin interactions¹³. We cross-referenced our epistatic SNPs with a map of chromosome interacting regions ($n = 96,139$) in K562 blood cell lines²⁹ (Methods) and found that 44 epistatic interactions mapped to within 5 Mb ($P < 1.8 \times 10^{-10}$),

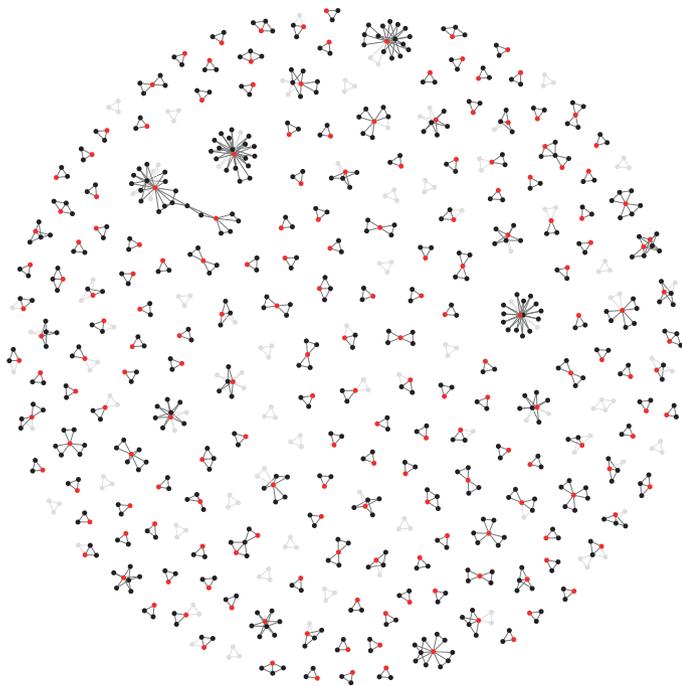


Figure 3 | Discovery and replication of epistatic networks. All 434 putative genetic interactions (edges) with data common to discovery and replication sets is shown, where black nodes represent SNPs and red nodes represent traits (gene expression probes). Three hundred and forty-five interactions had P values exceeding the 2.5% confidence interval following meta-analysis of the replication data. The remaining 89 interactions that did not replicate are depicted in grey. It is evident that a large proportion of the complex networks identified in the discovery set also exist in independent populations. An interactive version of this graph can be found at (http://kn3in.github.io/detecting_epi/).

(Supplementary Fig. 15). Interaction of distant loci may occur through physical proximity in transcriptional factories that organize across different chromosome regions and can regulate transcription of related genes³⁰.

Quantifying the importance of epistasis in complex traits in humans remains an open question. Here we are able to identify 238 gene expression traits with at least one significant interaction given our experiment-wide threshold, where the minimum estimated variance explained by the epistatic effects of any interaction was 2.1% of phenotypic variance. Taking results from our previously published expression quantitative trait loci (eQTL)²³, we calculated that 1,848 of the 7,339 gene expression levels analysed were influenced by additive effects where the estimated additive variance of a locus was 2.1% or greater. Thus, we can infer that the number of instances of large additive effects is substantially greater than the number of instances of large epistatic effects.

In terms of their contribution to complex traits, a more important metric might be the proportion of the variance that the epistatic loci explain². Taking all additive effects detected in ref. 23 that have additive variance explaining 2.1% or greater of phenotypic variance, we calculated that the proportion of total phenotypic variance of all 7,339 gene expression levels explained by additive effects alone was 2.16%. By contrast, the estimated epistatic variance from the interacting SNPs detected in this study on average explain a total of 0.22% of phenotypic variance, approximately ten times lower than the estimated additive variance. There are several caveats to this comparison which we discuss in the Methods.

Overall, we have demonstrated that it is possible to identify and replicate epistasis in complex traits amongst common human variants, despite the relative contribution of pairwise epistasis to phenotypic variation being small. The bioinformatic analyses of the significant epistatic loci indicate that there are a large number of possible mechanisms that can lead to non-additive genetic variation. Further research into such

epistatic effects may provide a useful framework for understanding molecular mechanisms and complex trait variation in greater detail. With computational techniques and data now widely available, the search for epistasis in larger data sets for traits of broader interest is warranted.

METHODS SUMMARY

We searched for pairwise epistasis exhaustively in the BSGS discovery data set²³, which comprises 846 individuals who are genotyped at 528,509 autosomal SNPs. Each individual had gene expression levels measured in peripheral blood at 7,339 probes representing 6,158 RefSeq genes (significant expression in $\geq 90\%$ of individuals). SNP pairs were modelled for full genetic effects, including marginal additive and dominance at both SNPs plus four interaction terms. We used permutation analysis to calculate an experiment-wide significance threshold of $T_e = 2.91 \times 10^{-16}$ at the 5% family-wise error rate (FWER). All SNP pairs with linkage disequilibrium $r^2 > 0.1$ and $D'^2 > 0.1$ were removed to minimise the possibility of haplotype effects. All SNP pairs were required to have at least five data points in all nine genotype classes. If multiple SNP pairs were present on the same chromosomes for a particular expression trait then only the sentinel SNP pair was retained. Finally, a nested test contrasting the full genetic model against the marginal additive and dominance model was performed for each remaining SNP pair. The 501 significant SNP pairs were carried forward for replication in two independent data sets that used the same expression assays for analysing transcription in peripheral blood Fehrmann¹² ($n = 1,240$) and EGCUT¹¹ ($n = 891$). A meta-analysis on the interaction P values from each replication data set was performed to provide an overall replication statistic for each putative interaction.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 September 2013; accepted 7 January 2014.

Published online 26 February 2014.

- Carlberg, O. & Haley, C. S. Epistasis: too often neglected in complex trait studies? *Nature Rev. Genetics* **5**, 618–625 (2004).
- Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4**, e1000008 (2008).
- Crow, J. F. On epistasis: why it is unimportant in polygenic directional selection. *Phil. Trans. R. Soc. B* **365**, 1241–1244 (2010).
- Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
- Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T.-L. V. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **234–237** (2013).
- Carlberg, O., Jacobsson, L., Ahgren, P., Siegel, P. & Andersson, L. Epistasis and the release of genetic variation during long-term selection. *Nature Genetics* **38**, 418–420 (2006).
- Strange, A. *et al.* A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between *HLA-C* and *ERAP1*. *Nature Genetics* **42**, 985–990 (2010).
- Evans, D. M. *et al.* Interaction between *ERAP1* and *HLA-B27* in ankylosing spondylitis implicates peptide handling in the mechanism for *HLA-B27* in disease susceptibility. *Nature Genetics* **43**, (2011).
- Cordell, H. J. Detecting gene–gene interactions that underlie human diseases. *Nature Rev. Genetics* **10**, 392–404 (2009).
- Hemani, G., Theocharidis, A., Wei, W. & Haley, C. EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics* **27**, 1462–1465 (2011).
- Metspalu, A. The Estonian Genome Project. *Drug Dev. Res.* **62**, 97–101 (2004).
- Fehrmann, R. S. N. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the *HLA*. *PLoS Genetics* **7**, e1002197 (2011).
- Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
- Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
- Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. & Kondrashov, F. A. Epistasis as the primary factor in molecular evolution. *Nature* **490**, 535–538 (2012).
- Weir, B. S. Linkage disequilibrium and association mapping. *Annu. Rev. Genomics Hum. Genet.* **9**, 129–142 (2008).
- Hemani, G., Knott, S. & Haley, C. An evolutionary perspective on epistasis and the missing heritability. *PLoS Genet.* **9**, e1003295 (2013).
- Marchini, J., Donnelly, P. & Cardon, L. R. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature Genet.* **37**, 413–417 (2005).
- Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
- Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003).

22. Powell, J. E. *et al.* Congruence of additive and non-additive effects on gene expression estimated from pedigree and SNP data. *PLoS Genet.* **9**, e1003502 (2013).
23. Powell, J. E. *et al.* The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS ONE* **7**, e35430 (2012).
24. Preiner, M. *et al.* Blood-informative transcripts define nine common axes of peripheral blood gene expression. *PLoS Genet.* **9**, e1003362 (2013).
25. Cockerham, C. C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**, 859–882 (1954).
26. Ho, T. H. *et al.* Muscleblind proteins regulate alternative splicing. *EMBO J.* **23**, 3103–3112 (2004).
27. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature Genet.* **45**, 124–130 (2013).
28. Hoffman, M. M., Buske, O., Wang, J. & Weng, Z. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nature Methods* **9**, 473–476 (2012).
29. Lan, X. *et al.* Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic Acids Res.* **40**, 7690–7704 (2012).
30. Rieder, D., Trajanoski, Z. & McNally, J. G. Transcription factories. *Front. Genet.* **3**, 221 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to the volunteers for their participation in these studies. We thank B. Hill, C. Haley and L. Ronnegard for discussions and comments. This work could not have been completed without access to high performance GPGPU compute clusters. We acknowledge iVEC for the use of advanced computing resources located at iVEC@UWA (<http://www.ivec.org>), and the Multi-modal Australian Sciences Imaging and Visualisation Environment (MASSIVE) (<http://www.massive.org.au>). We also thank J. Carroll and I. Porebski from the Queensland Brain Institute Information Technology Group for HPC support. The University of Queensland group is supported by the Australian National Health and Medical Research Council (NHMRC) grants 389892, 496667, 613601, 1010374 and 1046880, the Australian Research Council (ARC) grant (DE130100691), and by National Institutes of Health (NIH) grants GM057091 and GM099568. The QIMR researchers acknowledge funding from the

Australian National Health and Medical Research Council (grants 241944, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 496688 and 552485), and the National Institutes of Health (grants AA07535, AA10248, AA014041, AA13320, AA13321, AA13326 and DA12854). We thank A. Caracella and L. Bowdler for technical assistance with the micro-array hybridisations. The CHDWB study funding support from the Georgia Institute of Technology Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The Fehrmann study was supported by grants from the Celiac Disease Consortium (an innovative cluster approved by the Netherlands Genomics Initiative and partly funded by the Dutch Government (grant BSIK03009)), the Netherlands Organization for Scientific Research (NWO-VICI grant 918.66.620, NWO-VENI grant 916.10.135 to L.F.), the Dutch Digestive Disease Foundation (MLDS W011-30), and a Horizon Breakthrough grant from the Netherlands Genomics Initiative (grant 92519031 to L.F.). This project was supported by the Prinses Beatrix Fonds, VSB fonds, H. Kersten and M. Kersten (Kersten Foundation), The Netherlands ALS Foundation, and J.R. van Dijk and the Adessium Foundation. The research leading to these results has received funding from the European Community's Health Seventh Framework Programme (FP7/2007-2013) under grant agreement 259867. The EGCUT study received targeted financing from the Estonian Government SF0180142s08, Center of Excellence in Genomics (EXCEGEN) and University of Tartu (SP1GVARENG). We acknowledge EGCUT technical personnel, especially V. Soo and S. Smit. Data analyses were carried out in part in the High Performance Computing Center of University of Tartu.

Author Contributions G.H., J.E.P., P.M.V. and G.W.M. conceived and designed the study. G.H., J.E.P., K.S., H.-J.W. and J.Y. performed the analysis. T.E. and A.M. provided the EGCUT data. A.K.H., A.F.M., G.W.M., N.G.M. and J.E.P. provided the BSGS data. G.G. provided the CHDWB data. H.-J.W. and L.F. provided the Fehrmann data. G.H. and J.E.P. wrote the manuscript with the participation of all authors.

Author Information Gene expression data is available at the Gene Expression Omnibus under accession code GSE53195. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.H. (g.hemani@uq.edu.au).

METHODS

Discovery data. The Brisbane Systems Genetics Study (BSGS) comprises 846 individuals of European descent from 274 independent families²³. DNA samples from each individual were genotyped on the Illumina 610-Quad Beadchip by the Scientific Services Division at deCODE Genetics Iceland. Full details of genotyping procedures are given in ref. 31. Standard quality control (QC) filters were applied and the remaining 528,509 autosomal SNPs were carried forward for further analysis.

Gene expression profiles were generated from peripheral blood collected with PAXgene TM tubes (Qiagen, Valencia, CA) using Illumina HT12-v4.0 bead arrays. The Illumina HT-12 v4.0 chip contains 47,323 probes, although some probes are not assigned to RefSeq genes. We removed any probes that did not match the following criteria: contained a SNP within the probe sequence with MAF >0.05 within 1,000 genomes data; did not map to a listed RefSeq gene; were not significantly expressed (based on a detection P value <0.05) in at least 90% of samples. After this stringent QC, 7,339 probes remained for 2D-eQTL mapping. These data are accessible through GEO accession number GSE53195.

Normalization: gene expression profiles were normalized and adjusted for batch and polygenic effects. Profiles were first adjusted for raw background expression in each sample. Expression levels were then adjusted using quantile and \log_2 transformation to standardise distributions between samples. Batch and polygenic effects were adjusted using the linear model:

$$y = \mu + \beta_1 c + \beta_2 p + \beta_3 s + \beta_4 a + g + e \quad (1)$$

where μ is the population mean expression levels, c , p , s and a are vectors of chip, chip position, sex and generation respectively, fitted as fixed effects; and g is a random additive polygenic effect with a variance covariance matrix

$$G_{jk} = \begin{cases} \sigma_a^2 & j=k \\ 2\phi_{jk}\sigma_a^2 & j \neq k \end{cases} \quad (2)$$

The parameter σ_a^2 is the variance component for additive background genetic. Here we are using family based pedigree information rather than SNP based IBD to account for relationships between individuals and so ϕ_{jk} is the kinship coefficient between individuals j and k . The residual, e , from equation (1) is assumed to follow a multivariate normal distribution with a mean of zero. Residuals were normalized by rank transformation and used as the adjusted phenotype for the pairwise epistasis scan to remove any skewness and avoid results being driven by outliers. The GenABEL package for R was used to perform the normalization³².

Exhaustive 2D-eQTL analysis. Two-stage search: we used epiGPU¹⁰ software to perform an exhaustive scan for pairwise interactions, such that each SNP is tested against all other SNPs for statistical association with the expression values for each of the 7,339 probes. This uses the massively parallel computational architecture of graphical processing units (GPUs) to speed up the exhaustive search. For each SNP pair there are 9 possible genotype classes. We treat each genotype class as a fixed effect and fit an 8 d.f. F -test to test the following hypotheses:

$$H_0 : \sum_{i=1}^3 \sum_{j=1}^3 (\bar{x}_{ij} - \mu)^2 = 0; \quad (3)$$

$$H_1 : \sum_{i=1}^3 \sum_{j=1}^3 (\bar{x}_{ij} - \mu)^2 = 0; \quad (4)$$

where μ is the mean expression level and \bar{x}_{ij} is the pairwise genotype class mean for genotype i at SNP 1 and genotype j at SNP 2. This type of test does not parameterize for specific types of epistasis, rather it tests for the joint genetic effects at two loci. This has been demonstrated to be statistically more efficient when searching for a wide range of epistatic patterns, although will also include any marginal effects of SNPs which must be dealt with post-hoc¹⁸.

Stage 1: the complete exhaustive scan for 7,339 probes comprises 1.03×10^{15} F -tests. We used permutation analysis to estimate an appropriate significance threshold for the study. To do this we performed a further 1,600 exhaustive 2D scans on permuted phenotypes to generate a null distribution of the extreme P values expected to be obtained from this number of multiple tests given the correlation structure between the SNPs. We took the most extreme P value from each of the 1,600 scans and set the 5% family-wise error rate (FWER) to be the 95% most extreme of these P values, $T_* = 2.13 \times 10^{-12}$. The effective number of tests in one 2D scan being performed is therefore $N_* = 0.05/T_* \approx 2.33 \times 10^{10}$. To correct for the testing of multiple traits we established an experiment wide threshold of $T_e = 0.05/(N_* \times 7339) = 2.91 \times 10^{-16}$. This is likely to be conservative as it assumes independence between probes.

Filtering: we used two approaches to filter SNPs from stage 1 to be tested for significant interaction effects in stage 2.

Filter 1: after keeping SNP pairs that surpassed the 2.91×10^{-16} threshold in stage 1 only SNP pairs with at least 5 data points in all 9 genotype classes were kept. We then calculated the linkage disequilibrium between interacting SNPs (amongst unrelated individuals within the discovery sample and also from 1,000 genomes data) and removed any pairs with $r^2 > 0.1$ or $D'^2 > 0.1$ to avoid the inclusion of haplotype effects and to increase the accuracy of genetic variance decomposition. If multiple SNP pairs were present on the same chromosomes for a particular expression trait then only the sentinel SNP pair was retained, that is, if a probe had multiple SNP pairs that were on chromosomes one and two then only the SNP pair with the most significant P value was retained. At this stage, 6,404 filtered SNP pairs remained.

Filter 2: we also performed a second filtering screen applied to the list of SNP pairs from stage 1 that was identical to filter 1 but an additional step was included in which any SNPs that had previously been shown to have a significant additive or dominant effect ($P < 1.29 \times 10^{-11}$) were removed²², creating a second set of 4,751 unique filtered SNP pairs.

Stage 2: to ensure that interacting SNPs were driven by epistasis and not marginal effects, we performed a nested ANOVA on each pair in the filtered set to test if the interaction terms were significant. We did this by contrasting the full genetic model (8 d.f.) against the reduced marginal effects model which included the additive and dominance terms at both SNPs (4 d.f.). Thus, a 4 d.f. F -test was performed on the residual genetic variation, representing the contribution of epistatic variance. Significance of epistasis was determined using a Bonferroni threshold of $0.05/(6404 + 4751) = 4.48 \times 10^{-6}$. This resulted in 406 and 95 SNP pairs with significant interaction terms from filters 1 and 2, respectively.

Type 1 error rate. Using a Bonferroni correction of 0.05 in the second stage of the two stage discovery scan implies a type 1 error rate of $\alpha = 0.05$. However, this could be underestimated because the number tests performed in the second stage depends on the number of tests in the first stage, and this depends on statistical power and model choice. We performed simulations to estimate the type 1 error rate of this study design.

We assumed a null model where there was one true additive effect and 7 other terms with no effect. To simulate a test statistic we simulated 8 z -scores, $z_1 \sim N(\sqrt{NCP}, 1)$ and $z_{2,8} \sim N(0, 1)$. Thus $z_{\text{full}} = \sum_{i=1}^8 z_i \sim \chi_8^2$ (representing the 8 d.f. test) and $z_{\text{int}} = \sum_{i=5}^8 z_i \sim \chi_4^2$ (representing the 4 d.f. test in which the null hypothesis of no epistasis is true). For a particular value of NCP we simulated 100,000 z values, and calculated the P_{full} value for the z_{full} test statistic. The n_{int} test statistics with $P_{\text{full}} < 2.31 \times 10^{-16}$ were kept for the second stage, where the type 1 error rate of stage 2 was calculated as the proportion of $P_{\text{int}} < 0.05/n_{\text{int}}$. The power at stage 1 was calculated as $n_{\text{int}}/100,000$. This procedure was performed for a range of NCP parameters that represented power ranging from ~ 0 to ~ 1 .

Population stratification. We ruled out population stratification as a possible cause of inflated test statistics. To test for cryptic relatedness driving the interaction terms we tested for increased linkage disequilibrium among the SNPs³³. We calculated the mean of the off-diagonal elements of the correlation matrix of all unique SNPs from the 501 interactions (731 SNPs) using only unrelated individuals, $\bar{r}^2 = 0.0039$. This is not significantly different from the null hypothesis of zero (sampling error = $1/n_{\text{unrelated}} = 0.0039$).

Probe mapping. To avoid possibility that epistatic signals might arise due to expression probes hybridizing in multiple locations we verified that probe sequences for genes with significant interactions mapped to only a single location. As an initial verification we performed a BLAST search of the full probe sequence against 1,000 genomes phase 1 version 3 human genome reference and ensured that only one genomic location aligned significantly ($P < 0.05$). As a second step, to mitigate the possibility of weak hybridization elsewhere in the genome we divided the probe sequence into three sections (1–25 bp, 13–37 bp, 26–50 bp) and performed a BLAST search of these probe sequence fragments. No probe sequences or probe sequence fragments mapped to positions other than the single expected genomic target ($P < 0.05$).

Replication. Data description: we attempted replication of the 501 significant interactions from the discovery set using three independent cohorts; Fehrmann, EGCUT, and CHDWB. It was required that linkage disequilibrium $r^2 < 0.1$ and $D'^2 < 0.1$ between interacting SNPs (as measured in the replication sample directly), and all nine genotype classes had at least 5 individuals present in order to proceed with statistical testing for replication in both data sets. We also excluded any putative SNPs that had discordant allele frequencies in any of the data sets. Details of the cohorts are as follows.

Fehrmann: the Fehrmann data set¹² consists of peripheral blood samples of $n = 1,240$ unrelated individuals from the UK and the Netherlands. Some of these individuals are patients and others are healthy controls.

Individuals were genotyped using the Illumina HumanHap300, Illumina HumanHap370CNV, and Illumina 610 Quad platforms. RNA levels were quantified using

the Illumina HT-12 V3.0 platform. These data are accessible through GEO accession numbers GSE20332 and GSE20142.

EGCUT: the Estonian Genome Center of the University of Tartu (EGCUT) study¹¹ consists of peripheral blood samples of $n = 891$ unrelated individuals from Estonia. They were genotyped using the Illumina HumanHap370CNV platform. RNA levels were quantified using the Illumina HT-12 V3.0 platform. These data are accessible through GEO accession number GSE48348.

CDHWB: the Center for Health Discovery and Well Being (CDHWB) Study²⁴ is a population based cohort consisting of $n = 139$ individuals of European descent collected in Atlanta, USA. Gene expression profiles were generated with Illumina HT-12 V3.0 arrays from peripheral blood collected from Tempus tubes that preserve RNA. Whole genome genotypes were measured using Illumina OmniQuad arrays. Owing to the small sample size, most SNP pairs did not pass filtering in this data set (20 SNP pairs remained) and so we have excluded it from the rest of the analysis.

Meta-analysis. The 4 d.f. interaction P values for each independent replication data set were calculated using the same statistical test as was performed in the discovery data set. We then took the interaction P values from EGCUT and Fehrmann and calculated a joint P value using Fisher's method of combining P values for a meta-analysis as $-2 \ln p_1 - 2 \ln p_2 \sim \chi^2_{4 \text{d.f.}}$. As in the discovery analysis, all gene expression levels were normalized using rank transformation to avoid skew or outliers in the distribution³⁴.

Concordance of direction of effects. We used four methods to calculate the concordance of the direction of effects between the discovery and replication data sets.

Test 1: here we test whether the most significant epistatic effect in the discovery set is in the same direction as the same epistatic effect in the replication sets. We decomposed the genetic variance into 8 orthogonal effects, four of which are epistatic ($A \times A$, $A \times D$, $D \times A$, $D \times D$). The sign of the epistatic effect that had the largest variance in the discovery was recorded, and then was compared to the same epistatic effect in the two replication data sets (regardless of whether or not the same epistatic effect was the largest in the replication data sets). The probability of the sign being the same in one data set is $1/2$. The probability of the sign being the same in two is $1/4$.

Test 2: here we test whether the most significant epistatic effect in the discovery is the same as the largest epistatic effect in the replication set with the sign being concordant. As in Test 1, but this time we required that the largest effect was the same in the discovery and the replication, and that they had the same sign (for example, if the largest effect in the discovery is $A \times A$, with a positive effect, then concordance is achieved if the same is true in the replication). The probability of one replication data set being concordant by chance is $1/8$, and concordance in both is $1/64$.

Test 3: here we test if the epistatic effects that are significant at nominal $P < 0.05$ in the discovery have the same direction of effect as in the replication. We count all the epistatic variance components in the discovery that have $P < 0.05$ (1,133 amongst the 434 discovery SNP pairs, that is, each SNP pair has at least 1 and at most 4 significant epistatic variance components). Then we compare the direction of the effect in the replication data set. The probability of the sign being the same in one data set for any one significant effect is $1/2$. The probability of the sign being the same in two is $1/4$.

Test 4: if we count how many of the 4 epistatic effects are concordant between the discovery and replication data for each interaction then is this significant from what we expect by chance? There can be either 0, 1, 2, 3 or 4 concordant signs at each interaction, each with expectation of $P = 1/16, 4/16, 6/16, 4/16, 1/16$ under the null hypothesis, respectively. Observed counts are multinomially distributed, and we tested whether the observed proportions were statistically different from the expected proportions using an approximation of the multinomial test³⁵.

The probability of observing the number of concordant signs in tests 1–3 is calculated using a binomial test. All variance decompositions were calculated using the NOIA method³⁶.

Effects of linkage disequilibrium on detection and replication. The power to detect genetic effects, when the observed markers are in linkage disequilibrium with the causal variants, is proportional to r^x . For additive effects $x = 2$, but for non-additive effects x is larger, that is, $x = 4$ for dominance or $A \times A$, $x = 6$ for $A \times D$ or $D \times A$, and $x = 8$ for $D \times D$. Many biologically realistic genotype-phenotype maps may be comprised of all 8 variance components¹⁸.

This is important for both detection and for replication of epistasis. For detection, if the epistatic effect includes the $D \times D$ term then if the two causal variants are tagged by observed markers that are each in linkage disequilibrium $r = 0.9$, then if the true variance is V_t then the observed variance V_o at the markers will be $0.9^8 V_t = 0.43 V_t$. Therefore, it is important to consider the sampling variation of \hat{r}^x in a sample given some true population value of r .

Simulation 1: for some values of fixed population parameters, p_1 (minor allele frequency at observed marker), q_1 (minor allele frequency at causal variant), and r

(linkage disequilibrium between marker and causal variant), the expected haplotype frequencies are:

$$h_{11} = r\sqrt{p_1 q_1 p_2 q_2} + p_1 q_1 \quad (5)$$

$$h_{12} = p_1 q_2 - r\sqrt{p_1 q_1 p_2 q_2} \quad (6)$$

$$h_{21} = p_2 q_1 - r\sqrt{p_1 q_1 p_2 q_2} \quad (7)$$

$$h_{22} = r\sqrt{p_1 q_1 p_2 q_2} + p_2 q_2 \quad (8)$$

where $p_2 = 1 - p_1$ and $q_2 = 1 - q_1$. For a range of population parameters we randomly sampled $2n$ haplotypes where the expected haplotype frequencies were $h_{11}, h_{12}, h_{21}, h_{22}$. From the sample haplotype frequencies we then calculated sample estimates of \hat{r} where:

$$\hat{r} = \frac{\hat{h}_{11} - \hat{p}_1 \hat{q}_1}{\sqrt{\hat{p}_1 \hat{q}_1 \hat{p}_2 \hat{q}_2}} \quad (9)$$

For each value of combination of the parameters p_1, q_1, r and n , 1,000 simulations were performed and the sampling mean and sampling standard deviation of $\hat{r}, \hat{r}^2, \hat{r}^4, \hat{r}^6, \hat{r}^8$ were recorded. It was observed that sampling variance increases for increasing x in \hat{r}^x .

Simulation 2: we assume that the discovery SNP pairs are ascertained (from a very large number of tests) to have high \hat{r} between observed SNPs and causal variants because otherwise power of detection would be low. We can speculate that the distribution of \hat{r} in this ascertained sample will be a mixture of r that is high and r that is lower but with ascertained higher values from sampling. Therefore, we would expect those with truly high r to have a higher replication rate in independent data sets, and those with ascertained high \hat{r} to have lower replication because resampling is unlikely to result in the same extreme ascertainment. To obtain empirical estimates of \hat{r} in discovery and replication data sets we conducted the following simulation: using 1,000 genomes data (phase 1, version 3, 379 European samples) we selected the 528,509 'markers' used in the original discovery analysis, plus 100,000 randomly chosen 'causal variants' (CVs) with minor allele frequency > 0.05 ; the 379 individuals were split into discovery (190) and replication (189) sets; for each CV, the marker with the maximum \hat{r}_D^2 from the marker panel was recorded in the discovery set (this marker was known as the 'discovery marker' (DM)); the \hat{r}_R^2 for each CV/DM pair was then calculated in the replication set where the discovery linkage disequilibrium was ascertained to be high, such that $\hat{r}_D^2 > 0.9$.

We observed that there was an average decrease in \hat{r}_R^x relative to \hat{r}_D^x , and that this decrease was larger with increasing x . We observed that $(\hat{r}_R^2 - \hat{r}_D^2)/\hat{r}_D^2 = 0.029$. Whereas $(\hat{r}_R^8 - \hat{r}_D^8)/\hat{r}_D^8 = 0.092$. The average drop in replication \hat{r}^8 was 3 times higher than the drop in \hat{r}^2 .

Interpretation: simulation 1 shows that sampling variance of r^x increases as x increases. Detection of epistasis is highly dependent upon high \hat{r} . Amongst the discovery SNPs, there will be a mixture of interactions where observed SNPs are either in true high linkage disequilibrium with causal variants or will have highly inflated sample \hat{r}^x compared to the population r^x . Simulation 2 shows that as x gets larger, the average decrease in \hat{r}^x between discovery and replication becomes larger, probably a result of ascertained high \hat{r} in the discovery and increased sampling variance with increasing x in the replication. These results demonstrate that if all else is equal, the impact of sampling variance of r alone will reduce the replication rate of epistatic effects compared to additive effects.

Additive and non-additive variance estimation. Fixed effects: to compare the relative contribution to the phenotypic variance of gene expression levels between additive and epistatic effects we are constrained by the problem that non-additive variance components for a phenotype cannot be calculated directly. Here, we only have SNP pairs that exceed a threshold of $P < 2.91 \times 10^{-16} = T_c$. A strong conclusion cannot be made about the genome-wide variance contribution, but we can compare the variance explained by SNP effects at this threshold for additive scans and epistatic scans.

In ref. 23, an expression quantitative trait locus (eQTL) study was performed searching for additive effects in the same BSGS data set as was used for the discovery here. Using the threshold T_c for the additive eQTL study, 453 of the 7,339 probes analysed here had at least one significant additive effect. Assuming that the phenotypic variance for each of the probes is normalized to 1, the total phenotypic variance of all 7,339 explained by the significant additive effects was 1.73%.

Following the same procedure, at the threshold T_c there were 238 gene expression probes with at least one significant pairwise epistatic interaction out of the 7,339 tested. In total the proportion of the phenotypic variance explained by the epistatic effects at these SNP pairs was 0.25%.

Limitations of this type of comparison: although it is useful to compare the relative variances of epistatic and additive effects, it must be stressed that our

results here are approximations that are very limited by the study design. We estimate that additive effects explain approximately 10 times more variance than epistatic effects, but this could be an overestimate or an underestimate owing to a number of different caveats. First, the ratio of additive to epistatic variance may differ at different minimum variance thresholds, and our estimate is determined by the threshold used. Second, the power of a 1 d.f. test exceeds that of an 8 d.f. test. Third, the non-additive variance at causal variants is expected to be underestimated by observed SNPs in comparison to estimates for additive variance. Last, the extent of winner's curse in estimation of effect sizes may differ between the two studies.

Pedigree estimates. The gene expression levels for *MBNL1*, *TMEM149*, *NAPRT1*, *TRAPPC5* and *CAST* are influenced by large *cis-trans* epistatic networks (eight interactions or more). Though it is not possible to orthogonally estimate the non-additive genetic variance for non-clonal populations, an approximation of a component of non-additive variance can be estimated using pedigree information. The BSGS data are comprised of some related individuals and standard quantitative genetic analysis was used to calculate the additive and dominance variance components for each gene expression phenotype in ref. 22. The dominance effect is likely to capture additive \times additive genetic variance plus some fraction of other epistatic variance components. We found that the aforementioned genes had dominance variance component estimates within the top 5% of all 17,994 gene expression probes that were analysed in ref. 22.

Functional enrichment analysis. Tissue specific transcriptionally active regions: we employed a recently published method (<http://www.broadinstitute.org/mpg/epigwas/>)²⁷ that tests for cell-type-specific enrichment of active chromatin, measured through H3K4me3 chromatin marks²⁷ in regions surrounding the 731 SNPs that comprise the 501 discovery interactions. The exact method used to perform this analysis has been described previously²⁷. Briefly, we tested the hypothesis that the 731 SNPs were more likely to be in transcriptionally active regions (as measured by chromatin marks) than a random set of SNPs selected from the same SNP chip. This hypothesis was tested for 34 cell types across four broad tissue types (haematopoietic, gastrointestinal, musculoskeletal and endocrine, and brain).

Chromosome interactions: it has been shown¹³ that different regions on different chromosomes or within chromosomes spatially co-localize within the cell. We shall refer to the co-localization of two chromosome regions as a chromosome interaction. A map of pairwise chromosome interactions for K562 blood cell lines was recently produced²⁹, and we speculated that part of the underlying biological mechanism behind some of the 501 epistatic interactions may arise from chromosome interactions. We found that 44 of the putative epistatic interactions were amongst SNPs that were within 5 Mb of known chromosome interactions. This means that SNP A was no more than 2.5 Mb from the focal point of the chromosome interaction on chromosome A, and SNP B was no more than 2.5 Mb from the focal point on chromosome B.

We performed simulations to test how extreme the observation of 44 epistatic interactions overlapping with chromosome interactions is compared to chance. Chromosome interactions fall within functional genomic regions^{13,29}, and the SNPs in our epistatic interactions are enriched for functional genomic regions. Therefore, we designed the simulations to ensure that the null distribution was of chromosome interactions between SNPs enriched for functional genomic regions but with no known epistatic interactions. To do this we used the 731 SNPs that form the 501 putative epistatic interactions and randomly shuffled them to create new sets of 501 pairs, disallowing any SNP combinations that were in the original set. Therefore, each new random set was enriched for functional regions but had no genetic interactions. We scanned the map of chromosome interactions for overlaps with the new sets and then repeated the random shuffling process. We performed 1,000 such permutations to generate a null distribution of chromosome interaction overlaps. We repeated this process, searching for overlaps within 1 Mb, 250 kb and 10 kb.

SNP co-localization with genomic features: we tested for enrichment of genomic features for the 687 IndexSNPs that comprise the 434 epistatic interactions with data present in discovery and replication data sets. For each of the 687 IndexSNPs we calculated linkage disequilibrium with all regional SNPs within a radius of 0.5 Mb

and kept all regional SNPs with linkage disequilibrium $r^2 > 0.8$. We then cross-referenced the remaining regional SNPs with the annotated chromatin structure reference²⁸) querying whether the regional SNPs fell in predicted promoter region including TSS (TSS), predicted promoter flanking region (PF), predicted enhancer (E), predicted weak enhancer or open chromatin *cis* regulatory element (WE), CTCF enriched element (CTCF), predicted transcribed region (T) or predicted repressed or low activity region (R) positions. Therefore a particular IndexSNP might cover multiple genomic features through linkage disequilibrium.

We then performed the whole querying process for each of the 528,509 SNPs present in the SNP chip used in the scan and used the results from this second analysis to establish a null distribution for the expected proportion of SNPs for each genomic feature. We calculated *P* values for enrichment of each of the seven genomic features independently, and for *cis*- and *trans*-SNPs separately, using a binomial test. For each genomic feature we used the expected proportion of SNPs as the expected probability of 'success' (*p*). Here, a success is defined as an IndexSNP residing in a region that includes the genomic feature. The observed number of successes for each IndexSNP (*k*) out of the total count of IndexSNPs (*n*) was then modelled as $\Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Transcription factor enrichment: to test for enrichment of transcription factor binding sites (TFBS) we followed a procedure similar to that described in the previous section. For each of the 687 IndexSNPs we extracted regional SNPs as previously described. We then used the PWMEnrich package in Bioconductor (<http://www.bioconductor.org/packages/2.12/bioc/html/PWMEnrich.html>) to identify which TFBSs each of the regional SNPs for one IndexSNP falls in (within a radius of 250 bp). Thus, the number of occurrences of a particular TFBS was counted for each IndexSNP. We used the 'threshold-free affinity' method for identifying TFBSs³⁸.

We constructed a null distribution of expected TFBS occurrences based on the same null hypothesis as described in the previous section. The probability of an IndexSNP covering a particular TFBS is identical to any of the 528,509 SNPs in the discovery SNP chip. To do this, we performed the same procedure for each SNP in the discovery SNP chip as was performed for each IndexSNP to obtain an expected probability of covering a particular TFBS. We then tested the IndexSNPs for enrichment of each TFBS independently, and for *cis*- and *trans*-SNPs separately. *P* values were obtained using *z*-scores, calculated by using a normal approximation to the sum of binomial random variables representing motif hits along the sequence³⁹.

Defining SNP associations identified previously: the discovery data set (BSGS) had previously been analysed for additive and dominant marginal effects for all gene expression levels^{22,23}. To define SNPs that had been previously detected to have effects for a particular gene expression level we used a significance threshold accounting for multiple testing across SNPs and expression probes, $T_m = 0.05 / (528509 \times 7339) = 1.29 \times 10^{-11}$. From this, we found that only nine of the 501 discovery interactions had known main effects, 64 were between SNPs that had no known marginal effects, and 439 were between a SNP with a known marginal effect and a SNP with no known marginal effect.

31. Medland, S. E. *et al.* Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am. J. Hum. Genet.* **85**, 750–755 (2009).
32. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).
33. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genet.* **43**, 519–525 (2011).
34. Westra, H.-J. *et al.* MixupMapper: correcting sample mix-ups in genome-wide datasets increases power to detect small genetic effects. *Bioinformatics* **27**, 2104–2111 (2011).
35. Williams, D. A. Improved likelihood ratio tests for complete contingency tables. *Biometrika* **63**, 33–37 (1976).
36. Álvarez-Castro, J. M., Le Rouzic, A. & Carlborg, O. How to perform meaningful estimates of genetic effects. *PLoS Genet.* **4**, e1000062 (2008).
37. Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467–1471 (2013).
38. Stormo, G. D. DNA binding sites: representation and discovery. *Bioinformatics* **16**, 16–23 (2000).
39. Ho Sui, S. J. *et al.* oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic Acids Res.* **33**, 3154–3164 (2005).