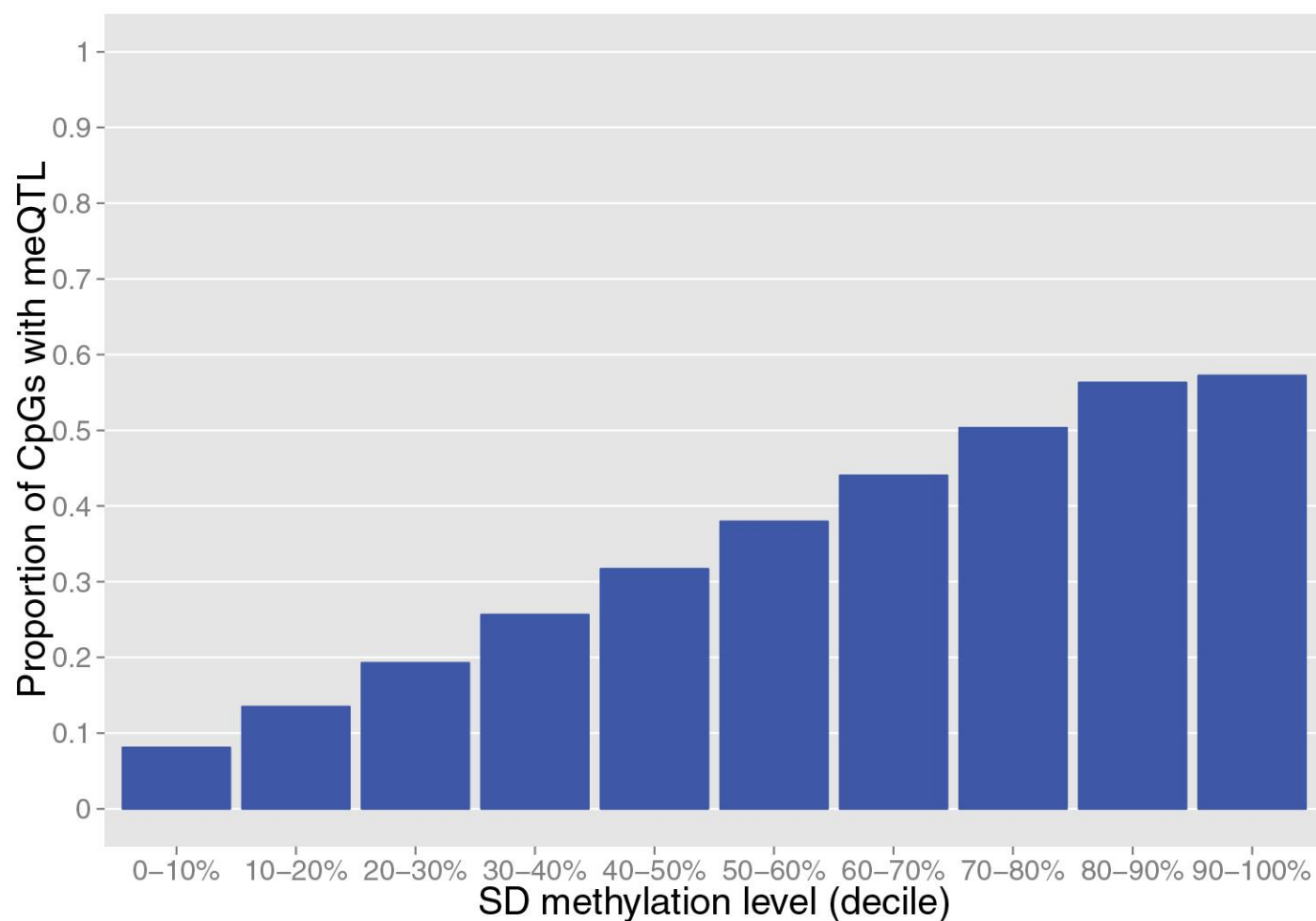**Supplementary Figure 1**

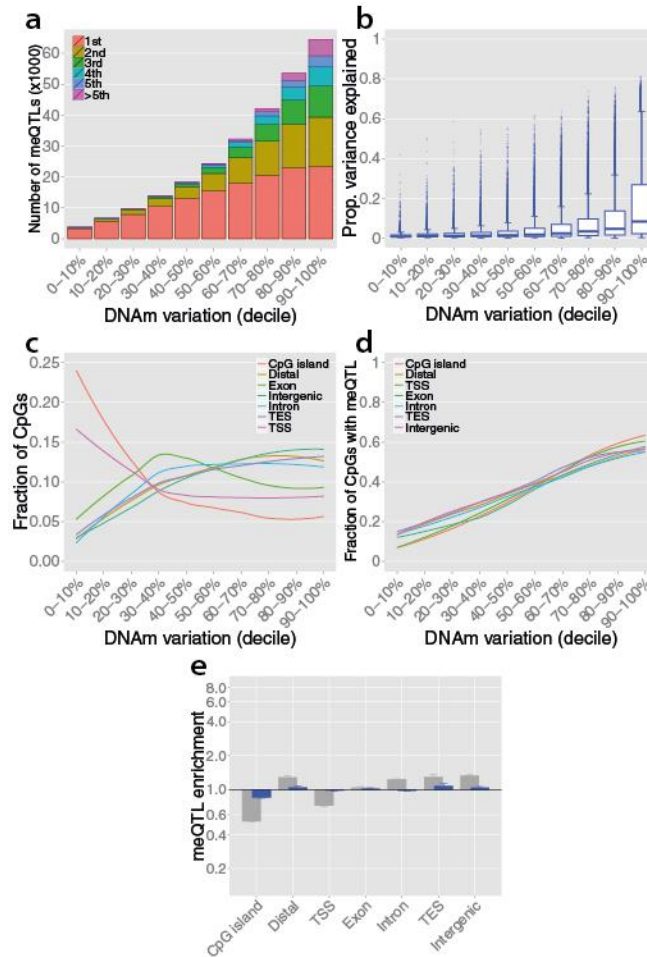**Density of distances between CpG sites and the most strongly associated meQTL SNP.**

Density plot of the distances between the 139,566 CpGs harboring a *cis*-meQTL and the most strongly associated SNP. Most SNP–CpG pairs are in close proximity (median distance = 10 kb), as indicated by the narrow peak around zero.

Supplementary Figure 2

Relationship between methylation variation and meQTL-associated CpGs.
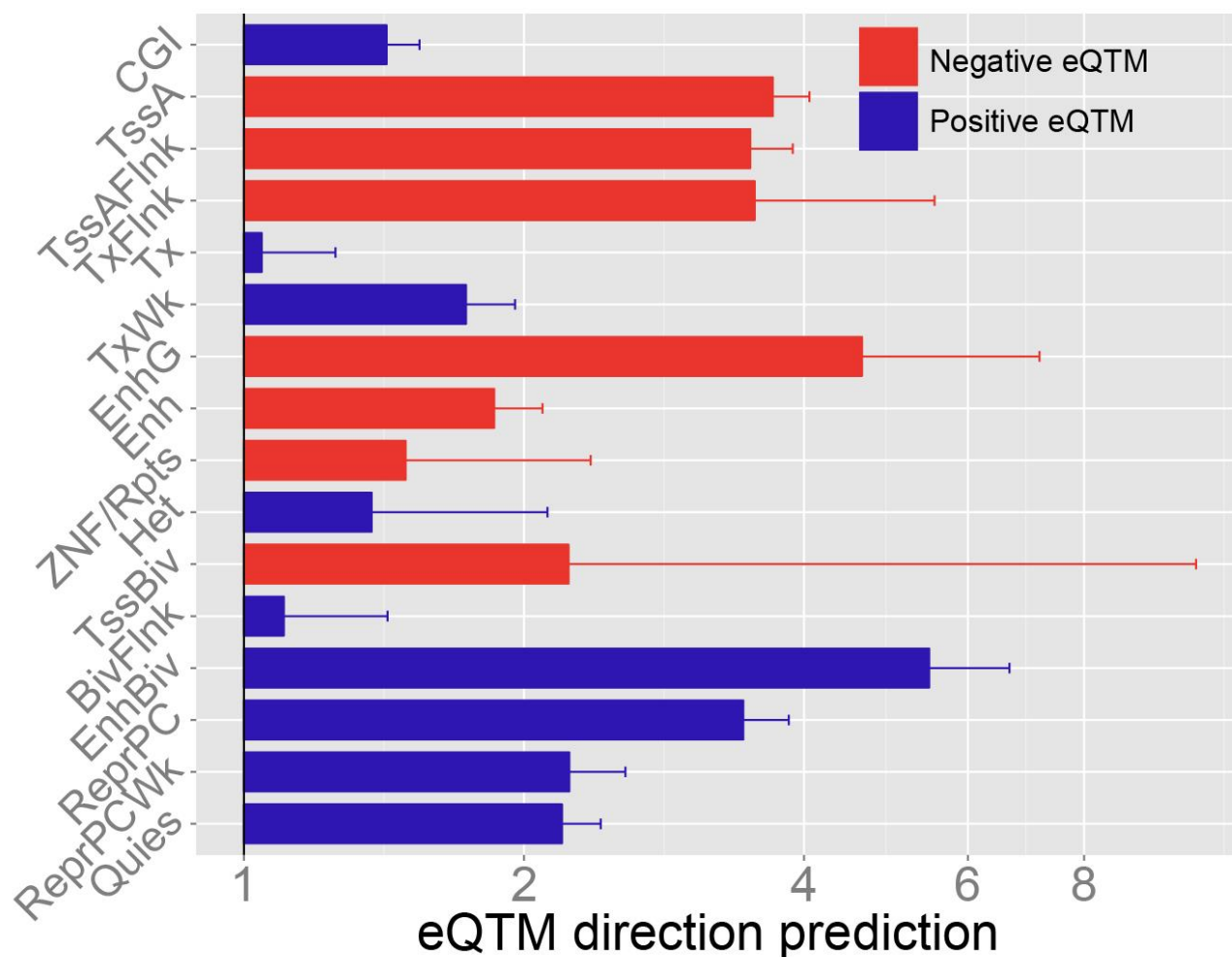
The proportion of CpGs harboring an identified *trans*-meQTL increases with increasing variability in DNA methylation. The proportion of CpGs with evidence of a *trans*-meQTL is calculated per decile of variability in methylation (*x* axis).

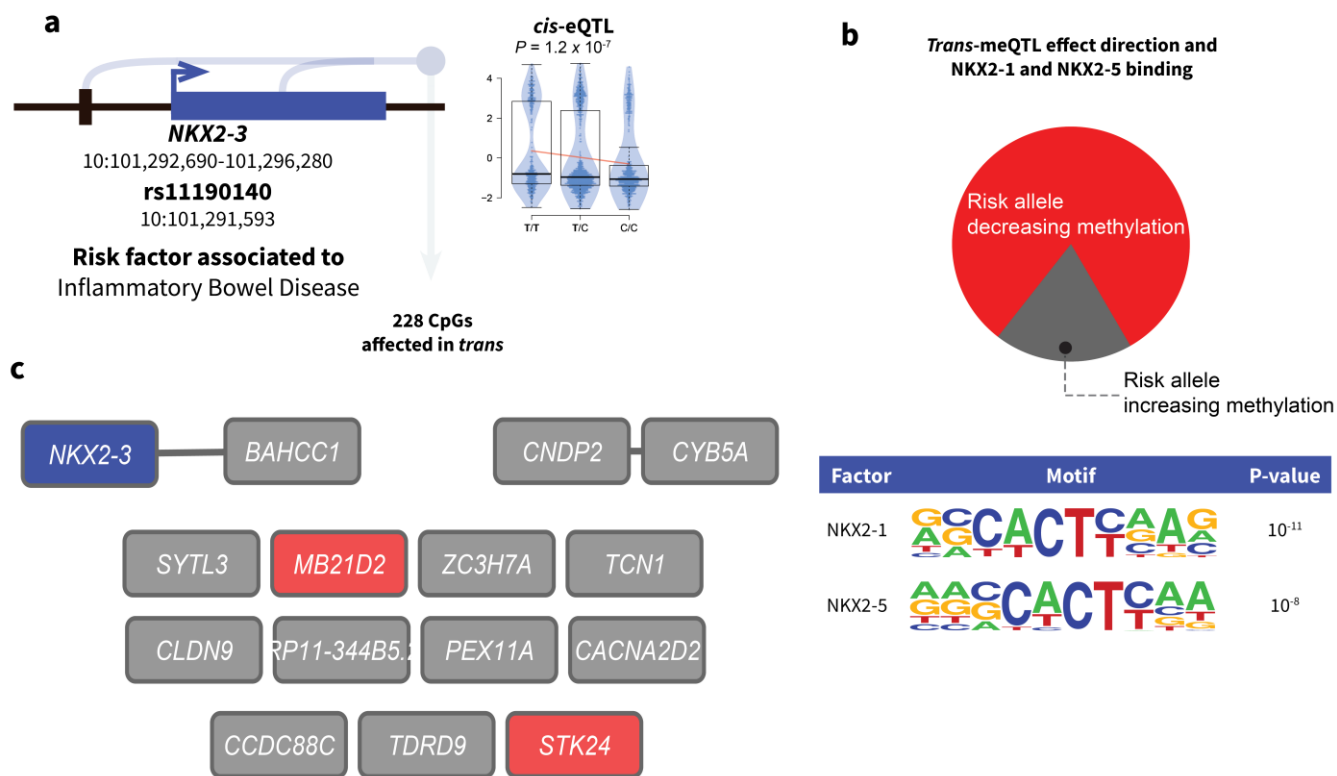**Supplementary Figure 3**

Characterization of cis-meQTLs.

(**a**) The number of *cis*-meQTLs found is strongly dependent on the variability in DNA methylation at a CpG site. Variances for 405,709 CpGs interrogated in the analyses were calculated using the 3,841 samples for which 450K data were available. Next, the CpGs were divided into deciles and the number of effects was counted for each decile. The different stacked colors correspond to primary, secondary, etc., effects. (**b**) The proportion of variance explained remains limited, even for highly variable CpG sites. The *x* axis shows the variances calculated for the 405,709 CpGs interrogated. The *y* axis shows the proportion of that variance explained by our identified *cis*-meQTLs. The limited proportion of variance explained, even for highly variable probes, suggests that increased statistical power contributes to but does not fully explain the increased number of *cis*-meQTLs identified. (**c**) DNA methylation variability differs across genomic contexts. Each line represents the proportion of the 405,709 CpGs used present in each genomic region. This clearly shows that some CpGs on the array are over-represented in certain genomic contexts. For example, weakly variable CpGs (0–10%) are over-represented in CpG islands. This may confound any enrichment analyses if variability in DNA methylation is influencing the likelihood of a given CpG harboring a meQTL. (**d**) DNA methylation variability seems to be the driving factor for identifying *cis*-meQTLs, even within genomic contexts. Each line again represents a distinct genomic context. (**e**) Reported enrichments of *cis*-meQTL effects for certain genomic contexts are strongly attenuated after accounting for the differential variability in DNA methylation between those genomic regions. Gray bars show uncorrected odds ratios. Blue bars show odds ratios corrected for methylation variability and the distance to the nearest SNP.

Supplementary Figure 4

Characterization of *cis*-eQTMs in relation to the direction of the eQTM effect.
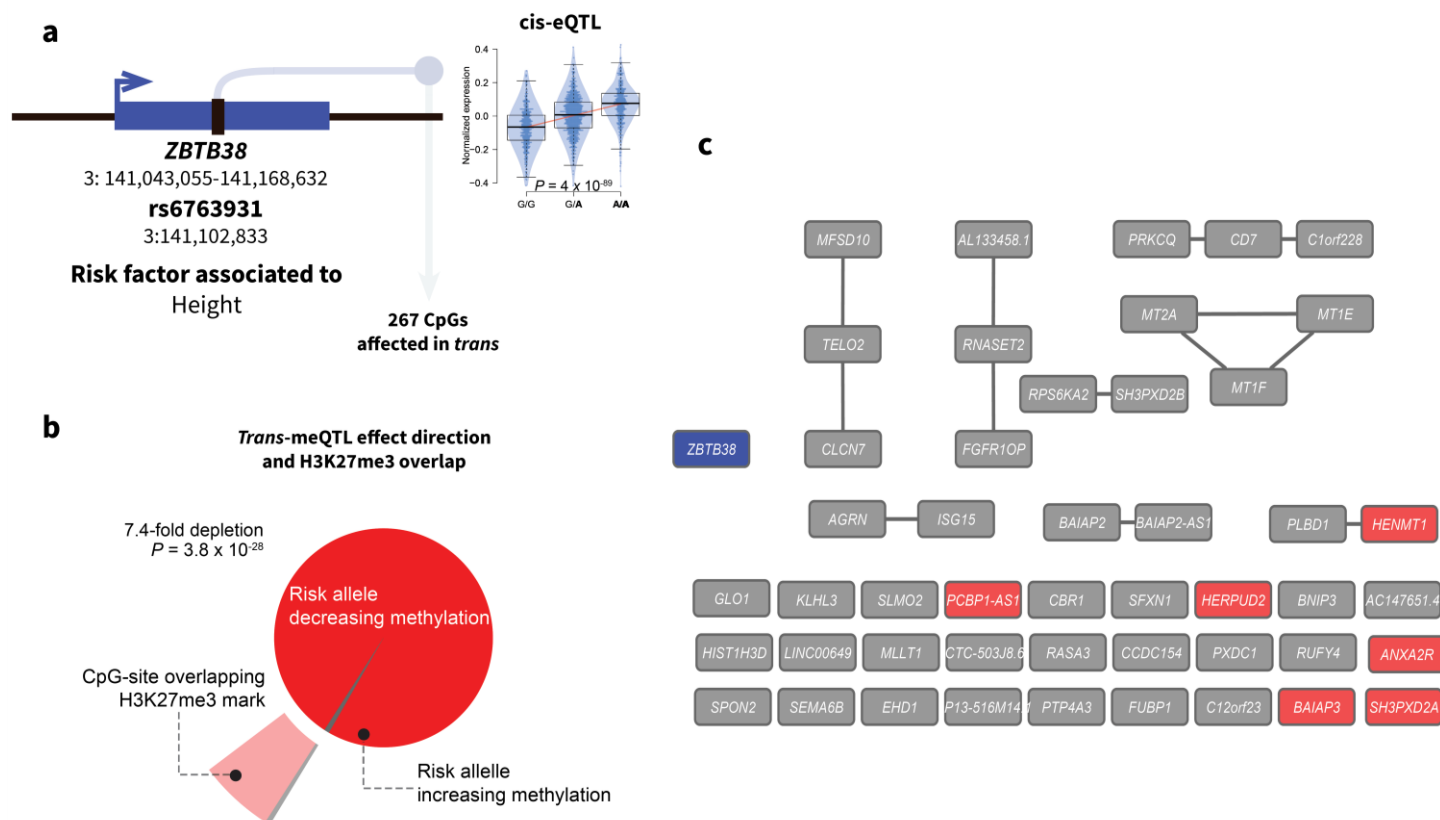
Over-representation of positive (blue bars) and negative (red bars) e-CpGs in CpG islands and predicted chromatin states. The *x* axis shows this over-representation in terms of odds ratios and error bars (95% confidence intervals). e-CpGs with negative associations are over-represented in active regions (for example, active TSSs and enhancers), whereas e-CpGs with positive associations are often found in repressed regions (for example, quiescent regions). CGI, CpG island; TssA, active TSS; TssAFlnk, flanking active TSS; TxFlnk, transcribed at gene 5′ or 3′ end; Tx, strong transcription; TxWk, weak transcription; EnhG, genic enhancer; Enh, enhancer; ZNF/Rpts, ZNF genes and repeats; Het, heterochromatin; TssBiv, bivalent/poised TSS; BivFlnk, flanking bivalent TSS/enhancer; EnhBiv: bivalent enhancer.

Supplementary Figure 5

***Trans*-meQTLs identified for a risk factor for inflammatory bowel disease, rs11190140, and the overlap with *NKX2-3*.**
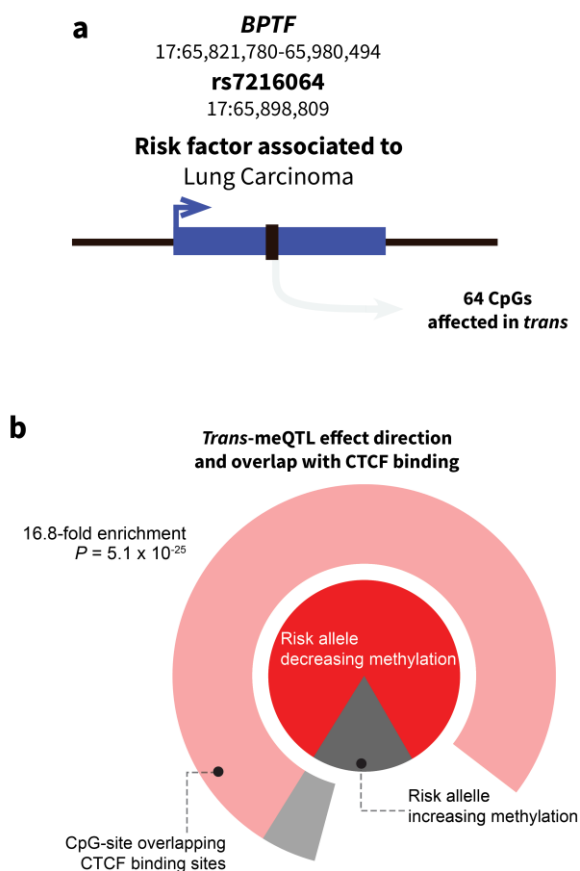
(**a**) Depiction of the *NKX2-3* gene and rs11190140, associated with inflammatory bowel disease. The plot shows increased expression of *NKX2-3* for the T risk allele. (**b**) In addition to influencing *NKX2-3* expression, rs11190140 also influences DNA methylation at 228 CpGs in *trans*, decreasing methylation levels at 81.1% of the affected CpG sites (red). In addition, many of the CpG sites overlap with NKX2-1 and NKX2-5 motifs (there is no NKX2-3 motif or ChIP–seq data available). (**c**) Gene network of the genes associated with 15 of the 228 CpGs (6.6%) with a *trans*-meQTL: blue, *cis*-eQTL-affected gene; red, genes associated both in methylation and expression.

Supplementary Figure 6

***Trans*-meQTLs identified for a risk factor for height, rs6763931, and the overlap with *ZBTB38*.**

(**a**) Depiction of the *ZBTB38* gene and rs6763931, associated with height. The plot shows increased expression of *ZBTB38* for the T risk allele. (**b**) In addition to influencing *ZBTB38* expression, rs6763931 also influences DNA methylation at 267 CpGs in *trans*, decreasing methylation levels at 99.2% of the affected CpG sites (red). In addition, depletion of overlap with H3K27me3 is observed (7.4-fold depletion, $P = 3.8 \times 10^{-28}$), shown in the outer chart. (**c**) Gene network of the genes associated with 60 of the 779 CpGs (7.7%) with a *trans*-meQTL: blue, *cis*-eQTL effected gene; red, genes associated both in methylation and expression.

Supplementary Figure 7

***Trans*-meQTLs identified for a risk factor related to lung carcinoma, rs7216064, and overlap with *BPTF*.**

(**a**) Depiction of the *BPTF* gene and rs7216064, associated with lung carcinoma. (**b**) rs7216064 influences DNA methylation at 64 CpGs in *trans*, decreasing methylation levels at 82.8% of the affected CpG sites (red). In addition, many of the CpG sites (81.3%) overlap with CTCF-binding sites (16.8-fold enrichment, $P = 5.1 \times 10^{-25}$), shown in the outer chart.

Supplementary note

Supplementary Results

Trans-meQTL stability

To ascertain stability our *trans*-meQTLs, we performed a replication analysis in a the set of 1,748 lymphocyte samples18: of the 18,764 overlapping *trans*-meQTLs between the datasets that could be tested, 94.9% had a consistent allelic direction (Figure 1E). 12,098 *trans*-meQTLs were nominally significant (unadjusted P < 0.05), of which 99.87% had a consistent allelic direction. This indicates that the identified *trans*-meQTLs are robust and not caused by differences in cell-type composition (Supplementary Table 4).

To further exclude the possibility of confounding by cellular heterogeneity, we performed our *trans*-meQTL mapping on uncorrected methylation data and data corrected for known cell type proportions (Neutrophil, Lymphocyte, Monocyte, Eosinophil and Basophil percentage). These analyses led to significantly less *trans*-meQTLs (17,704 and 19,625, respectively) (SupplementaryTable 5,6), suggesting cellular heterogeneity does not confound our results. Of the 17,704 *trans*-meQTLs that are identified in the uncorrected data 82% are shared with the final *trans*-meQTL mapping, all in the same allelic direction. For the 19,625 *trans*-meQTLs we identified after correcting for cell-type information, 80% of the *trans*-meQTLs are shared with the final *trans*-meQTL analysis, again all in the same allelic direction.

Furthermore, *trans*-meQTL mapping only using SNPs known to influence cell proportions[1,2] in blood revealed that most of these SNPs have no or very few *trans*-meQTLs, whereas widespread *trans*-meQTL effects were to be expected if our analysis had not properly controlled for blood cell composition (SupplementaryTable 7). 153 of these 261 SNPs affect a single CpG site in trans only, thus contrasting the reviewer's prediction. The SNP (rs9932319, reported to be affecting kir+ NK cells) that is affecting

methylation in trans most (altering 486 CpG sites) maps in close proximity to the CTCF transcription factor. Since the *trans*-meQTL CpG sites are strongly enriched for CTCF binding, we conclude that these *trans*-meQTLs are true positives and not false-positive findings due to differences in cell-type proportions.

Lastly, we linked our GWAS SNPs to the SNPs known to influence cell proportions and found that only 0.6% of the GWAS SNPs are in high LD with SNPs known to influence cell proportions.

References

1.   Orrù, V. *et al.* Genetic variants regulating immune cell levels in health and disease. *Cell* **155,** 242–56 (2013).

2.   Roederer, M. *et al.* The genetic architecture of the human immune system: A bioresource for autoimmunity and disease pathogenesis. *Cell* **161,** 387–403 (2015).

Supplementary acknowledgements

Biobank-based Integrative Omics Study Consortium membership

Management Team

Bastiaan T. Heijmans (chair)[1], Peter A.C. 't Hoen[2], Joyce van Meurs[3], Aaron Isaacs[4], Rick Jansen[5], Lude Franke[6].

Cohort collection

Dorret I. Boomsma[7], René Pool[7], Jenny van Dongen[7], Jouke J. Hottenga[7] (Netherlands Twin Register); Marleen MJ van Greevenbroek[8], Coen D.A. Stehouwer[8], Carla J.H. van der Kallen[8], Casper G. Schalkwijk[8] (Cohort study on Diabetes and Atherosclerosis Maastricht); Cisca Wijmenga[6], Lude Franke[6], Alexandra Zhernakova[6], Ettje F. Tigchelaar[6] (LifeLines Deep); P. Eline Slagboom[1], Marian Beekman[1], Joris Deelen[1], Diana van Heemst[9] (Leiden Longevity Study); Jan H. Veldink[10], Leonard H. van den Berg[10] (Prospective ALS Study Netherlands); Cornelia M. van Duijn[4], Bert A. Hofman[11], Aaron Isaacs[4], André G. Uitterlinden[3] (Rotterdam Study).

Data Generation

Joyce van Meurs (Chair)[3], P. Mila Jhamai[3], Michael Verbiest[3], H. Eka D. Suchiman[1], Marijn Verkerk[3], Ruud van der Breggen[1], Jeroen van Rooij[3], Nico Lakenberg[1].

Data management and computational infrastructure

Hailiang Mei (Chair)[12], Maarten van Iterson[1], Michiel van Galen[2], Jan Bot[13], Daria V. Zhernakova[6], Rick Jansen[5], Peter van 't Hof[12], Patrick Deelen[6], Irene Nooren[13], Peter A.C. 't Hoen[2], Bastiaan T. Heijmans[1], Matthijs Moed[1].

Data Analysis Group

Lude Franke (Co-Chair)[6], Martijn Vermaat[2], Daria V. Zhernakova[6], René Luijk[1], Marc Jan Bonder[6], Maarten van Iterson[1], Patrick Deelen[6], Freerk van Dijk[14], Michiel van Galen[2], Wibowo Arindrarto[12], Szymon M. Kielbasa[15], Morris A. Swertz[14], Erik. W van Zwet[15], Rick Jansen[5], Peter A.C. 't Hoen (Co-Chair)[2], Bastiaan T. Heijmans (Co-Chair)[1].

1. Molecular Epidemiology Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands

2. Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

3. Department of Internal Medicine, ErasmusMC, Rotterdam, The Netherlands

4. Department of Genetic Epidemiology, ErasmusMC, Rotterdam, The Netherlands

5. Department of Psychiatry, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands

6. Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands

7. Department of Biological Psychology, VU University Amsterdam, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands

8. Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, Maastricht, The Netherlands

9. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands

10. Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands

11. Department of Epidemiology, ErasmusMC, Rotterdam, The Netherlands

12. Sequence Analysis Support Core, Leiden University Medical Center, Leiden, The Netherlands

13. SURFsara, Amsterdam, the Netherlands

14. Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

15. Medical Statistics Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands