Supplementary results

Trans-meQTL stability

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) (Extended Data Table 7,8), 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. (Extended Data Table 6). 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

The Rotterdam Study is supported by the Erasmus MC University Medical Center and Erasmus University Rotterdam; The Netherlands Organisation for Scientific Research (NWO); The Netherlands Organisation for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); The Netherlands Genomics Initiative (NGI); the Ministry of Education, Culture and Science; the Ministry of Health, Welfare and Sports; the European Commission (DG XII); and the Municipality of Rotterdam. The contribution of inhabitants, general practitioners and pharmacists of the Ommoord district to the Rotterdam Study is gratefully acknowledged.

BIOS Consortium

(Biobank-based Integrative Omics Study)

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

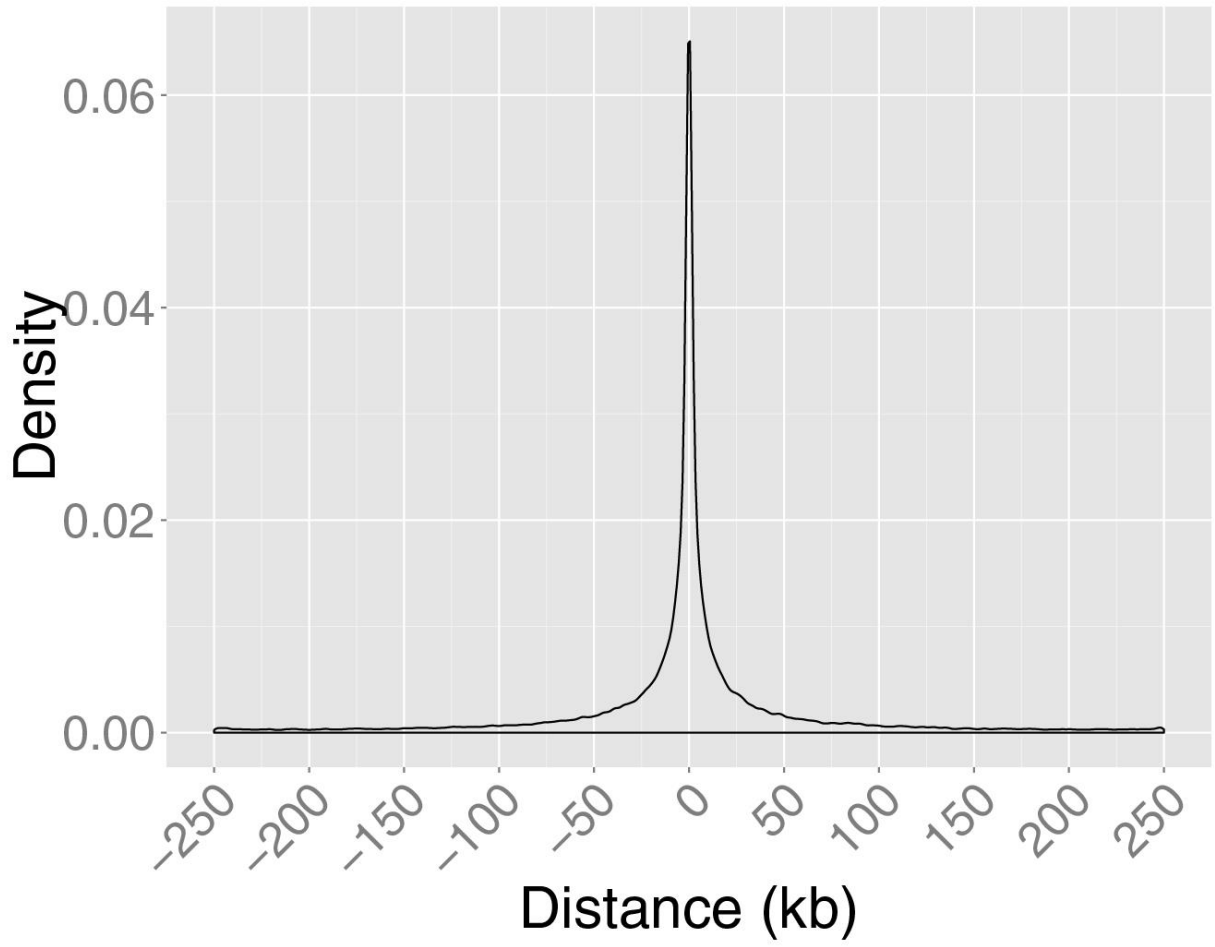
12. Sequence Analysis Support Core, Leiden University Medical Center, Leiden, The Netherlands

13. SURFsara, Amsterdam, the Netherlands

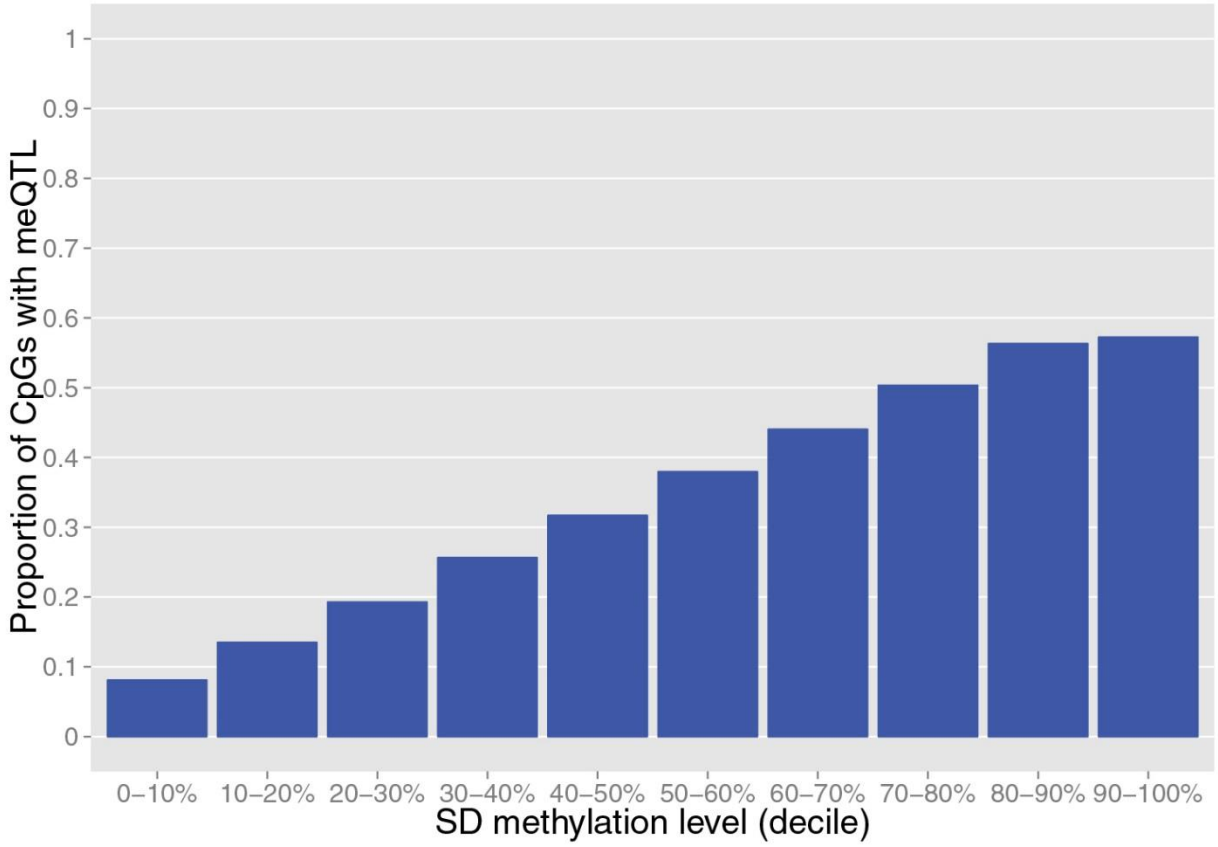
14. Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

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

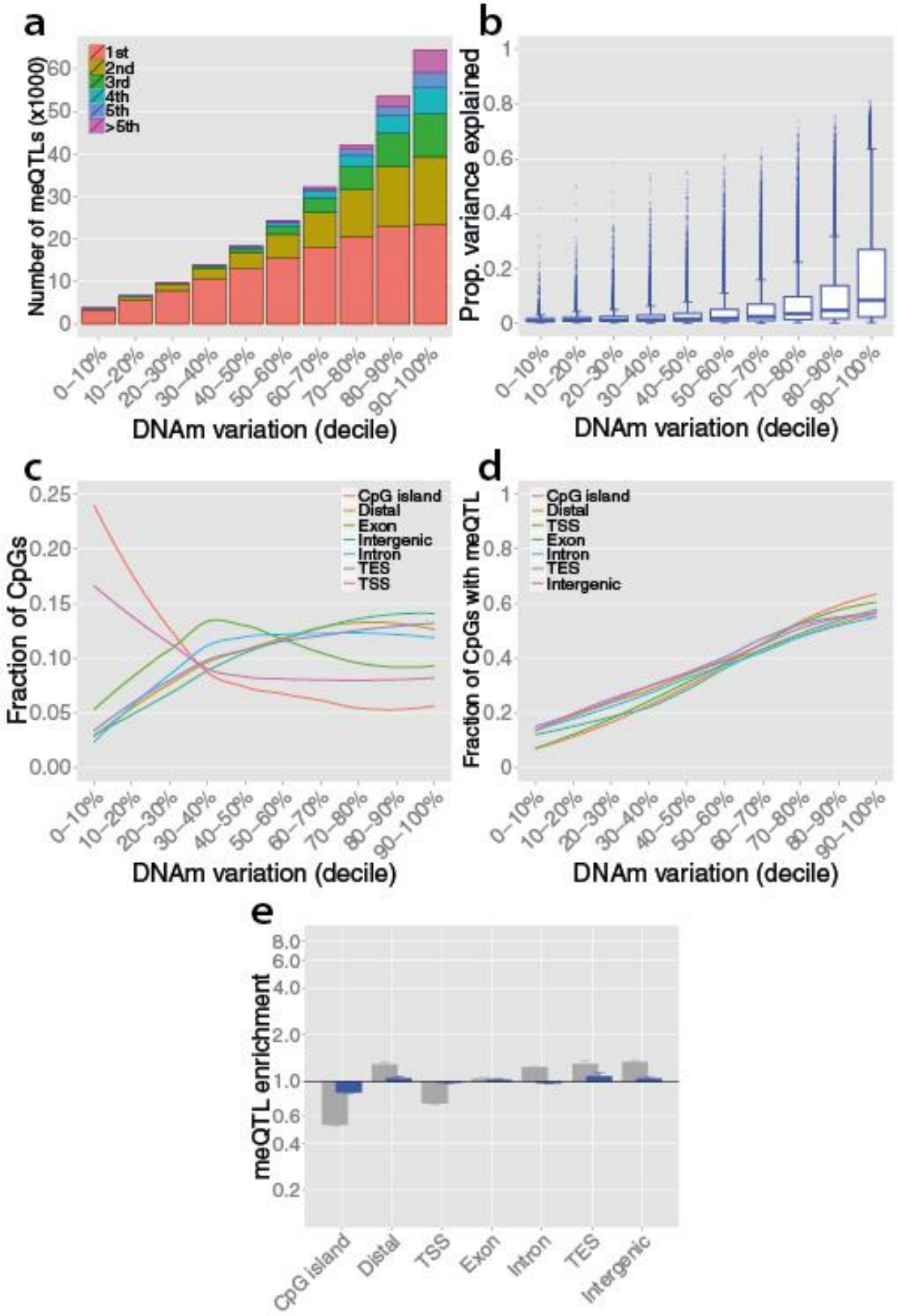
Extended Data Figures



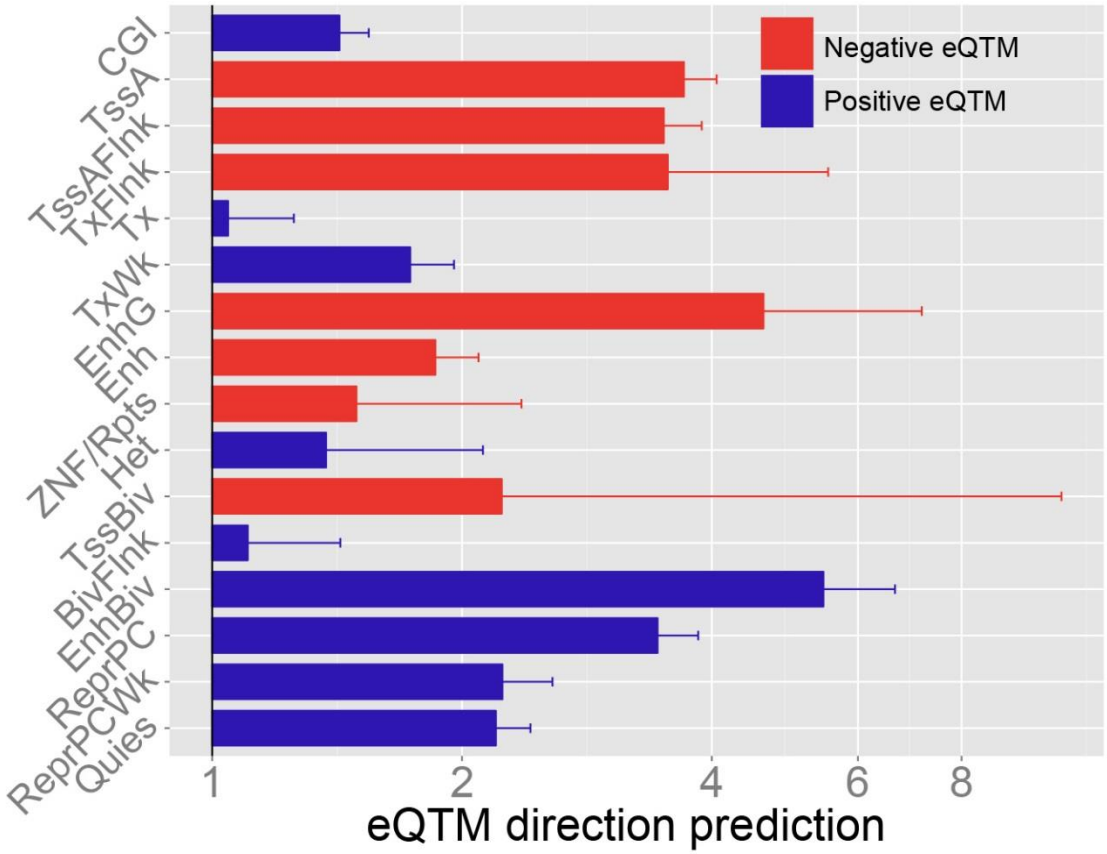
Extended Data Fig. 1. Density plot of the distance between the 139,566 CpGs harboring a *cis*-meQTL and its strongest associated. Most SNP-CpG pairs are in close proximity (median distance: 10 kb), as indicated by the narrow peak around zero.



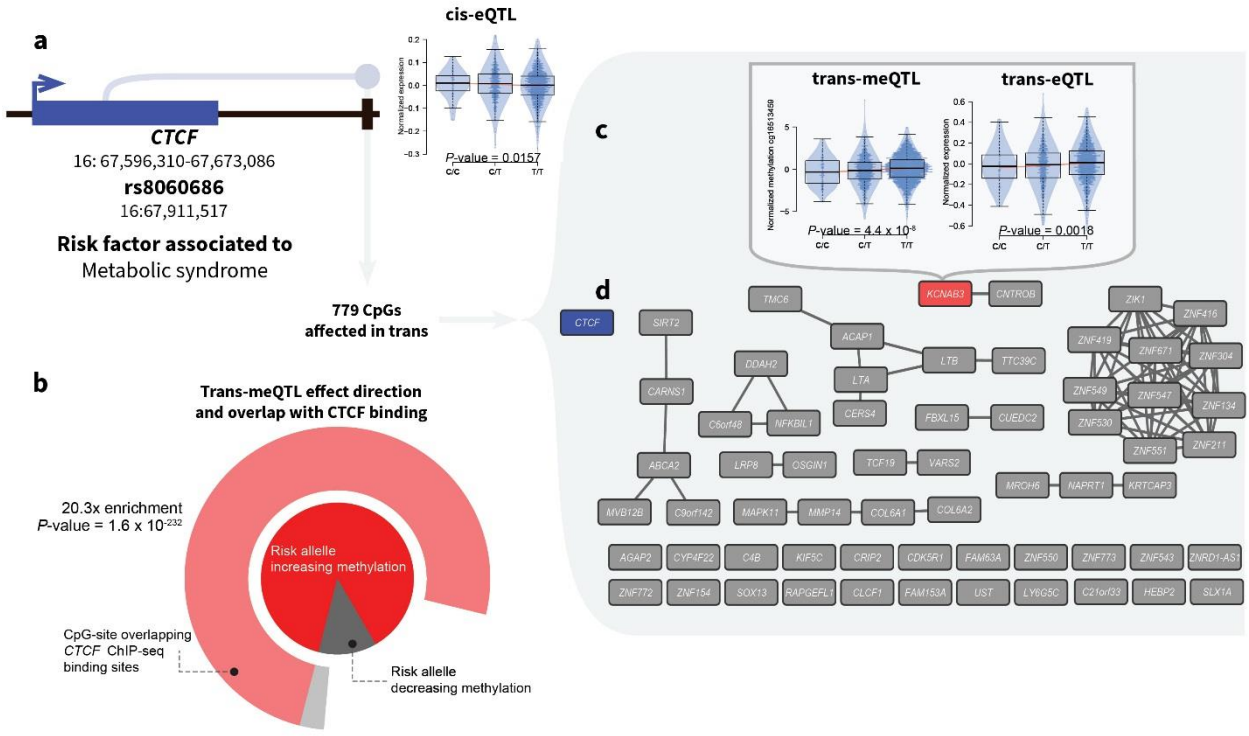
Extended Data Fig. 2. The proportion of CpGs harboring an identified *trans*-meQTL increases with increasing DNA methylation variability. The proportion of CpGs with evidence of a *trans*-meQTL are calculated per decile of methylation variability (x-axis).


a
Number of meQTLs (x1000)
1st
2nd
3rd
4th
5th
>5th
0–10%
10–20%
20–30%
30–40%
40–50%
50–60%
60–70%
70–80%
80–90%
90–100%
DNAm variation (decile)
b
Prop. variance explained
DNAm variation (decile)
c
Fraction of CpGs
CpG island
Distal
Exon
Intergenic
Intron
TES
TSS
DNAm variation (decile)
d
Fraction of CpGs with meQTL
CpG island
Distal
TSS
Exon
Intron
TES
Intergenic
DNAm variation (decile)
e
meQTL enrichment
CpG island
Distal
TSS
Exon
Intron
TES
Intergenic

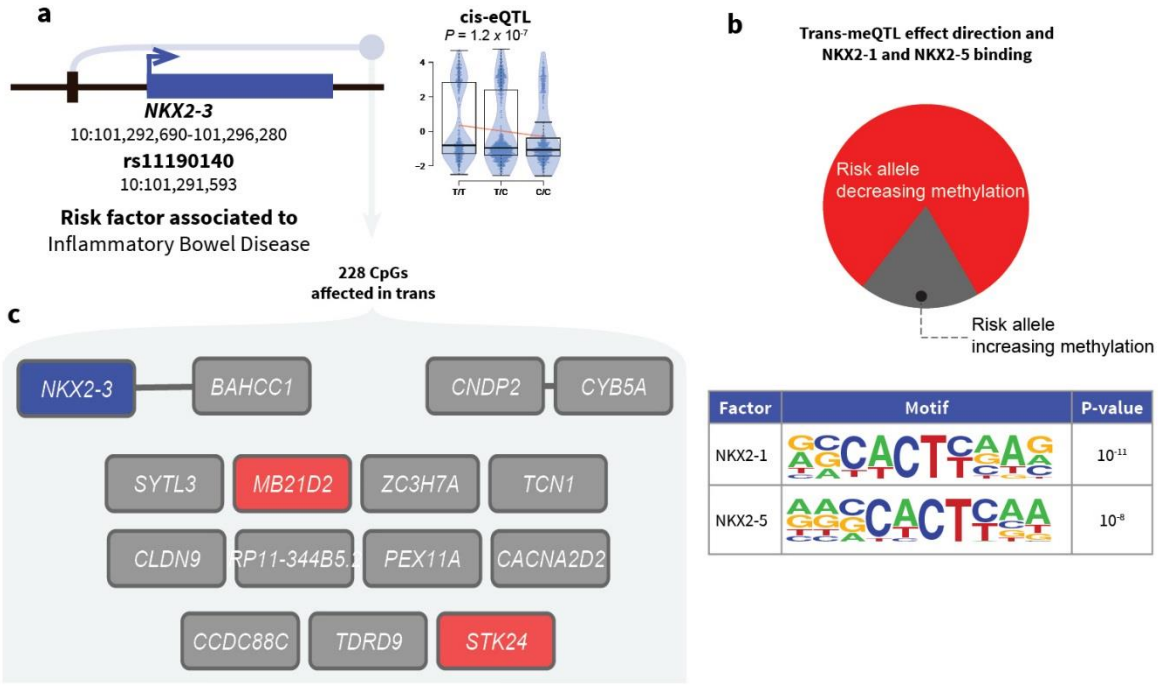
Extended Data Fig. 3. **a,** The number of *cis*-meQTLs found is strongly dependent on the variability of DNA methylation at the CpG site. Variances for 405,709 CpGs interrogated in the analyses were calculated using the 3,841 samples for which 450k data was available. Next, the CpGs were divided in deciles and the number of effect counted per decile. The different stacked colors indicate the 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 interrogated 405,709 CpGs. 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 increased statistical power contributes to, but does not fully explain the increased number of *cis*-meQTLs identified. **c,** DNA methylation variability differs between genomic contexts. Each line represents the proportion of 405,709 used CpGs present in each genomic region. This clearly shows some CpGs on the array are overrepresented in certain genomic contexts. For example, lowly variable CpGs (0-10%) are overrepresented 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. Grey bars show uncorrected odds ratios. Blue bars show odds ratios corrected for methylation variability, and the distance to the nearest SNP.



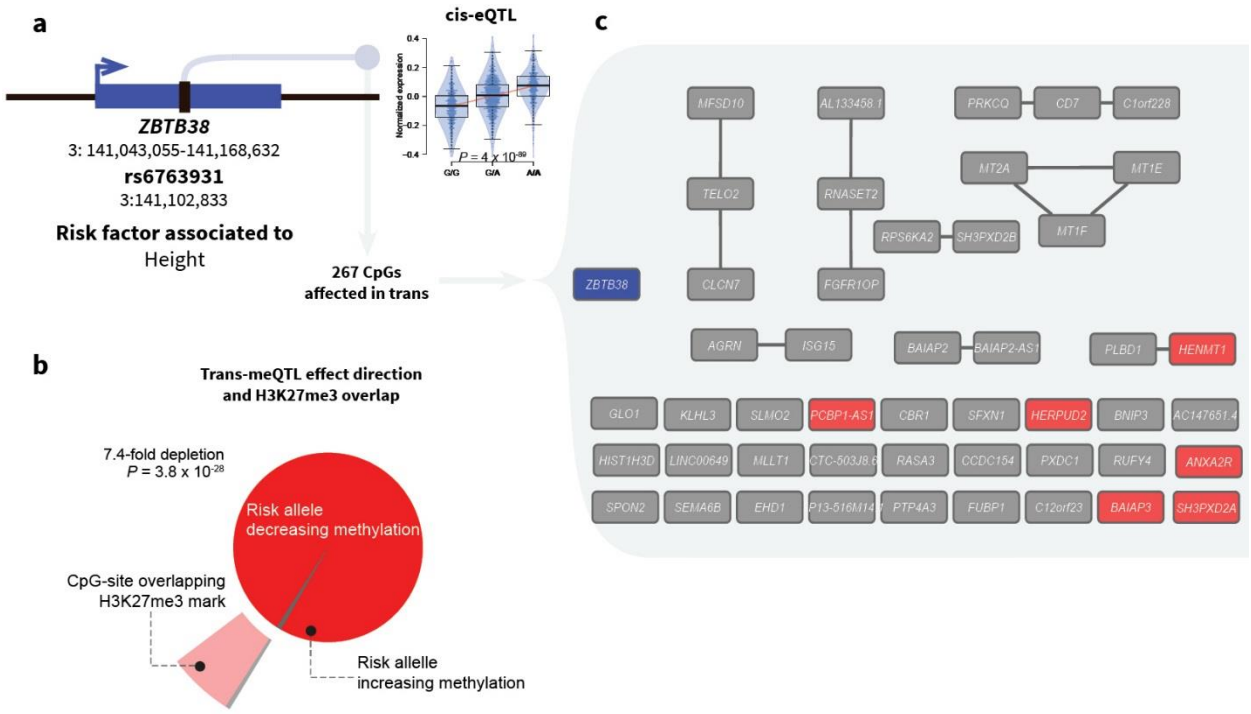
Extended Data Fig. 4. Overrepresentation of positive (blue bars) and negative e-CpGs (red bars) in CpG islands and predicted chromatin states. The x-axis shows this overrepresentation in terms of odds ratios and error bars (95% confidence interval). e-CpGs with negative associations are overrepresented in active regions (e.g., Active TSS and Enhancers), while e-CpGs with positive association are often found in repressed regions (e.g. Quitescent). CGI: CpG island; TssA: Active TSS; TssAFlnk: Flanking active TSS; TxFlnk, Transcribed at gene 5' and 3'; 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.



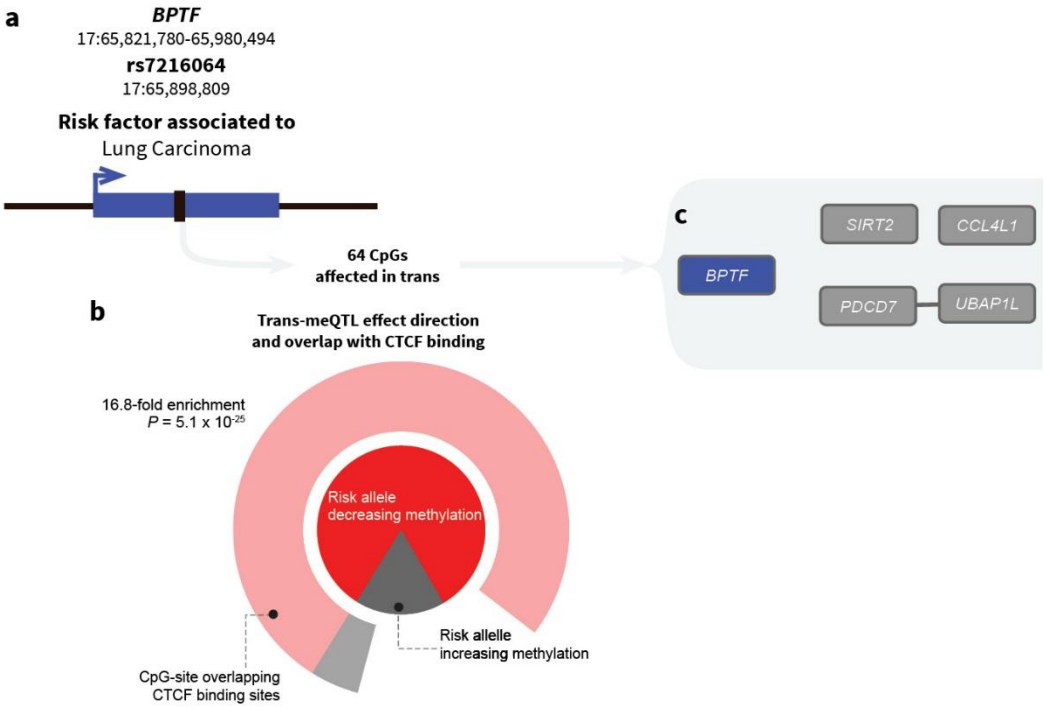
Extended Data Fig. 5. **a**, Depiction of the *CTCF* gene and rs8060686, associated with metabolic syndrome. The plot shows an increased expression of *NFKB1* for the risk allele C. **b**, In addition to influencing *CTCF* expression, rs8060686 also influences DNA methylation at 779 CpGs *in trans*, increasing methylation levels at 87.7% of affected CpG sites (dark grey). In addition, many of the CpG sites (77.4%) overlap with CTCF binding sites (20.3-fold enrichment, *P*-value = 1.6 *x* 10^{-232}), shown in the outer chart. **c**, Illustrations of meQTL (left plot) and eQTL effects (right plot) of rs8060686 *in trans*. Only SNP-gene combinations were tested where the gene was associated with one of the 779 CpGs with a *trans*-meQTL. **d**, Gene network of the genes associated with 60 of the 779 CpGs (7.7%) with a *trans*-meQTL.



Extended Data Fig. 6 Depiction of the *NKX2-3* gene and rs11190140, associated with inflammatory bowl syndrome. The plot shows an increased expression of *NKX2-3* for the risk allele T. **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 affected CpG sites (red). In addition, many of the CpG sites overlap with motifs of *NKX2-1* and *NKX2-5* (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, in blue the cis-eQTL effected gene is shown and in red the genes associated both in methylation and in expression.



Extended Data Fig. 7 Depiction of the *ZBTB38* gene and rs6763931, associated with heigth. The plot shows an increased expression of *ZBTB38* for the risk allele T. **b**, In addition to influencing *ZBTB38* expression, rs6763931 also influences DNA methylation at 267 CpGs *in trans*, decreasing methylation levels at 99.2% of affected CpG sites (red). In addition, a depletion of overlap with H3K27me3 is observed (7.4-fold depletion, *P*-value = 3.8 *x* 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, in blue the cis-eQTL effected gene is shown and in red the genes associated both in methylation and in expression.



Extended Data Fig. 8 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 affected CpG sites (red). In addition, many of the CpG sites (81.3%) overlap with CTCF binding sites (16.8-fold enrichment, *P*-value = 5.1 *x* 10^{-25}), shown in the outer chart. **C** Genes associated to altered methylation caused by rs7216064, in blue BPTF is included.