Comprehensive mechanistic characterization of mQTLs in an Asian population

Sijia Wang (✉ wangsijia@picb.ac.cn)

CAS Key Laboratory of Computational Biology, Shanghai Institutes of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences <https://orcid.org/0000-0001-6961-7867>

Qianqian Peng

Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences <https://orcid.org/0000-0002-2018-5706>

Xinxuan Liu

Key Laboratory of Genomic and Precision Medicine, China Gastrointestinal Cancer Research Center, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101 <https://orcid.org/0000-0002-1653-1434>

Han Jing

CAS Key Laboratory of Computational Biology, Shanghai Institutes of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences

Wenran Li

CAS Key Laboratory of Computational Biology, Shanghai Institutes of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences

Xingjian Gao

Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences

Qi Luo

CAS-SINH

Charles Breeze

National Cancer Institute <https://orcid.org/0000-0002-5294-915X>

Jiarui Li

CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences <https://orcid.org/0000-0003-0581-5850>

Siyu Pan

Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences

Guochao Li

Key Laboratory of Genomic and Precision Medicine, China Gastrointestinal Cancer Research Center, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101

Jiaqiang Qian

CAS Key Laboratory of Computational Biology, Shanghai Institutes of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences

Liyun Yuan

Bio-Med Big Data Center, CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences <https://orcid.org/0000-0001-9122-6782>

Na Yuan

Beijing Institute of Genomics

Chenglong You

CAS Key Laboratory of Computational Biology, Shanghai Institutes of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences

Siyuan Du

CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences <https://orcid.org/0000-0003-1602-1669>

Yuanning Zheng

Fudan University

Ziyu Yuan

Taizhou Institute of Health Sciences, Fudan University

Jingze Tan

Fudan University

Jiucun Wang

Fudan University <https://orcid.org/0000-0003-2765-0620>

Guoqing Zhang

CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences

Xianping Lu

Shenzhen Chipscreen Biosciences Co. Ltd.

Leming Shi

State Key Laboratory of Genetic Engineering, Human Phenome Institute, School of Life Sciences and Shanghai Cancer Center, Fudan University, Shanghai 200438, China <https://orcid.org/0000-0002-2981-4150>

Shicheng Guo

University of Wisconsin-Madison

Yun Liu

Fudan University <https://orcid.org/0000-0001-5753-1265>

Ting Ni

Collaborative Innovation Center of Genetics and Development, Human Phenome Institute, School of Life Sciences, Fudan University <https://orcid.org/0000-0001-7007-1072>

Bo Wen

Fudan University <https://orcid.org/0000-0003-0394-7366>

Li Jin

Fudan University

Andrew Teschendorff

CAS-MPG Partner Institute for Computational Biology <https://orcid.org/0000-0001-7410-6527>

Fan Liu

CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, University of Chinese Academy of Sciences, Chinese Academy of Sciences <https://orcid.org/0000-0001-9241-8161>

Article

Keywords: methylation quantitative trait loci (mQTLs), characterization, Asian population

Posted Date: November 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1088676/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Comprehensive mechanistic characterization of mQTLs in an Asian population

Qianqian Peng^{1*}, Xinxuan Liu^{2,3,4*}, Han Jing^{1*}, Wenran Li^{1*}, Xingjian Gao^{2,3}, Qi Luo¹, Charles E. Breeze⁵, Jiarui Li¹, Siyu Pan^{2,3}, Guochao Li^{2,3}, Jiaqiang Qian¹, Liyun Yuan¹, Na Yuan^{2,3}, Chenglong You¹, Siyuan Du¹, Yuanting Zheng^{6,7}, Ziyu Yuan⁸, Jingze Tan^{6,7}, Jiucun Wang^{6,7,8}, Guoqing Zhang^{1,8}, Xianping Lu⁹, Leming Shi^{6,7,8}, Shicheng Guo^{10,11}, Yun Liu¹², Ting Ni¹³, Bo Wen^{6,14}, Li Jin^{6,7,8}, Andrew E. Teschendorff^{1**}, Fan Liu^{2,3,4**}, Sijia Wang^{1,8,15**}

¹CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China.

²Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China.

³China National Center for Bioinformation, Beijing, China.

⁴School of Future Technology, University of Chinese Academy of Sciences, Beijing, China.

⁵Altius Institute for Biomedical Sciences, Seattle, WA, USA.

⁶State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, and Human Phenome Institute, Fudan University, Shanghai, China.

⁷Ministry of Education Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Sciences, Fudan University, Shanghai, China.

⁸Taizhou Institute of Health Sciences, Fudan University, Taizhou, Jiangsu, China.

⁹Shenzhen Chipscreen Biosciences Co. Ltd., Shenzhen, Guangdong, China.

¹⁰Department of Medical Genetics, School of Medicine and Public Health, University of Wisconsin-

24 Madison, Madison, WI, USA.

25 ¹¹Center for Precision Medicine Research, Marshfield Clinic Research Institute, Marshfield, WI, USA.

26 ¹²MOE Key Laboratory of Metabolism and Molecular Medicine, Department of Biochemistry and

27 Molecular Biology, School of Basic Medical Sciences, Fudan University, Shanghai, China.

28 ¹³State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and

29 Development, Human Phenome Institute, School of Life Sciences and Huashan Hospital, Fudan

30 University, Shanghai, China.

31 ¹⁴The Fifth People's Hospital of Shanghai and Institutes of Biomedical Sciences, Fudan University,

32 Shanghai, China;

33 ¹⁵Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming,

34 China.

35

36 * These authors contributed equally: Qianqian Peng, Xinxuan Liu, Han Jing, Wenran Li.

37 ** Corresponding authors:

38 Andrew E. Teschendorff: andrew@picb.ac.cn

39 Fan Liu: liufan@big.ac.cn

40 Sijia Wang: wangsijia@picb.ac.cn

41

Abstract

Identification and characterization of methylation quantitative trait loci (mQTLs) can help elucidate the role of DNA methylation changes as a potential mediator of genetic risk loci. However, mQTLs remain poorly characterized: they have not yet been mapped in the largest ethnic populations, their cell-type specific nature has not been resolved, and the proportion of mQTLs attributed to different molecular mechanisms is unknown. Here we perform the first mQTL-mapping study in a large Han Chinese population, demonstrating that over 80% of mQTLs are shared with those identified in White Caucasians. We further estimate that over 90% of mQTLs are shared between different blood cell-lineages. mQTLs demonstrate a strong enrichment for variants influencing chromatin accessibility. We identify a number of GWAS-linked transcription factor trans-mQTL hotspots associated with eosinophilia, ulcerative colitis and body mass index, and a subset of trans-mQTLs within the NFkB-pathway that may mediate the risk of obesity. In summary, this study significantly expands our understanding of mQTLs and their potential role in mediating disease risk, whilst also contributing the first mQTL-database in an Asian population.

Introduction

Recent mQTL studies performed in predominantly White Caucasian populations have demonstrated that a significant proportion of the DNA methylome is under genetic influence¹⁻¹³. These studies^{1,2,8,11} have shown the critical importance of single nucleotide polymorphism (SNP)-mediated transcription factor (TF) modulation in explaining trans-mQTLs, whilst also demonstrating that many cis-mQTLs co-locate with chromatin interaction domains, thus highlighting some of the potential mechanisms that can give rise to mQTLs. However, our current understanding of the mQTL landscape is incomplete. First of all, it is unclear how many of the reported mQTLs in White Caucasians are also present in East Asians, the largest ethnic group. To establish this is critically important for a better understanding of how epigenetically mediated genetic predisposition to disease could preferentially affect specific ethnicities. Second, it is still unclear if mQTLs detected from a complex tissue like whole blood reflect underlying DNA methylation (DNAm) changes in all of the blood cell subtypes, or in just a few subsets. Cell-type specific differential DNAm changes have led to novel insights in the context of smoking and cancer¹⁴⁻¹⁷, hence it is important to establish if DNAm changes driven by SNPs display cell-type specificity or not, as such knowledge can lead to a much deeper and improved understanding of the role of DNAm changes in disease predisposition. Third, recent studies have highlighted the important role SNPs play in modulating chromatin accessibility¹⁸⁻²⁰. Thus, it is critically important to dissect mQTLs into the various mechanistic models that can give rise to them. Fourth, previous mQTL studies have used the DNAm 450K beadchip, a two-fold lower resolution than the current 850K chip, particularly in relation to enhancer regions^{21,22}.

Here we take a first step towards building a more complete picture of the mQTL landscape, by performing a comprehensive mechanistic characterization of mQTLs in a well sized East Asian population using the 850K beadchip, at cell-type resolution, and by integration with an algorithm that predicts tissue and cell-type specific causal variants for chromatin accessibility.

Results

mQTL mapping and validation in East Asians reveals strong similarities with those in Caucasians

To identify mQTLs in East Asians we measured DNAm at over 850K CpGs and genotypes at over 8M SNPs in whole blood from 3,523 Han Chinese (median age = 52, age range = 18-83) (the “NSPT” cohort) (**Methods; Fig. 1a**). After preprocessing and quality control, we retained 811,876 CpGs and 7.58M SNPs. To efficiently perform over 10^{12} mQTL regressions, we developed FastQTLmapping²³, a linear regression solver in C++ for exhaustive regression analysis between two extraordinarily large matrices allowing for covariates (**Methods**). Compared to MatrixEQTL²⁴, the current state-of-the-art algorithm, fastQTLmapping was up to 4 and 11 times faster in the single-thread and 32 CPU threads settings, respectively (**Fig. 1b**). Overall, fastQTLmapping completed the mQTL mapping (8.6M SNPs $\times$ 812K CpGs $\times$ 3,523 samples) in just 6.79 hours running with 32-threads and 321GB peak memory, whilst with MatrixEQTL it would have taken approximately two weeks.

Following previous studies^{2,5,10}, we defined three categories of mQTLs: cis-mQTLs (SNP-CpG distance within 1Mbp), long-cis (lcis)-mQTLs (SNP-CpG pair within a 1~5Mbp window), and trans-mQTLs for all others. We identified about 62.92M genome-wide significant mQTLs (**Fig. 1a&c**), including 56.29M cis-, 2.27M lcis- and 4.36M trans-mQTLs (Bonferroni adjusted $P_{\text{cis}} < 1.06 \times 10^{-11}$, $P_{\text{lcis}} < 2.86 \times 10^{-12}$ and $P_{\text{trans}} < 8.16 \times 10^{-15}$). The mQTL SNPs (mSNPs) covered more than 2/3 of tested SNPs (5.56M), while mQTL CpGs (mCpGs) covered 1/3 of tested CpGs (284,128). Most lcis-mCpGs (88.91%) were also significantly associated with mSNPs within 1Mbp, whilst only 41.25% of the trans-mCpGs showed significant association with mSNPs within 1Mbp (**Extended Fig. 1a&b**). The minor allele frequency (MAF) distributions of cis- and lcis-mSNPs were skewed towards lower frequency compared to trans-mSNPs (**Fig. S1a**). The distributions of methylation beta values, variation levels, and heritability estimates were largely similar between cis-, lcis-, and trans-mCpGs (**Fig. S1b-d**). Consistent with previous findings^{2,25}, the explained mCpG variance (R^2) decreased with its distance

108 from a cis-mSNP (**Extended Fig. 2a&b**), which is likely explained by a reduced effect size due to
 109 linkage disequilibrium (LD) with another causal and physically closer cis-mSNP (**Extended Fig. 2c**).
 110 In an independent Han Chinese cohort (CGZ, N = 798)²⁶⁻²⁸, we successfully replicated 87.05% (False
 111 Discovery Rate or FDR < 0.05) of the significant findings in our discovery cohort, with high directional
 112 consistency (99.79% directionally consistent association, $r = 0.98$, P-value < 1.00×10^{-323} , **Fig. 1d**).
 113 Next, we compared our East Asian (EAS) derived mQTLs with those reported in a European (EUR)
 114 ancestry cohort, using data from the Framingham Heart Study (FHS, N = 4170)². Out of 16.07M
 115 mQTLs with genome-wide significance in FHS and representation in our dataset, a total of 13.03M
 116 (81.09%) were also significant (FDR < 0.05) in our EAS cohort. For these pan-ethnic mQTLs, we
 117 observed a fairly strong correlation between EAS and FHS effect sizes ($r = 0.79$, P-value < 1.00×10^{-323}),
 118 with the vast majority (12.08M, 92.67%) of effects being in the same direction (**Fig. 1e**). The
 119 allele frequency differences for pan-ethnic mSNPs were significantly smaller when compared to the
 120 18.91% mSNPs that were only significant in FHS (one-tailed T-test, P-value < 1.00×10^{-323} , **Extended**
 121 **Fig. 3**). This demonstrates that a substantial proportion of mQTLs are shared between EAS and EUR
 122 populations, and that the differences between ethnic groups are driven mainly by differences in allele
 123 frequency.
 124 mSNPs were enriched in genomic functional regions such as promoters and exons, and this pattern
 125 was more pronounced for trans-mSNPs than cis-mSNPs (**Fig. 1f**). Similar with Caucasian studies^{2,29},
 126 cis-mCpGs were enriched in enhancer and CpG island shores, while trans-mCpGs were relatively more
 127 enriched in TSS200 and CpG islands compared to cis-mCpGs (**Fig. 1f**). Genes near cis-mCpGs were
 128 more enriched for basic biological functions, components and processes, in contrast to trans-mCpGs
 129 which were more enriched for immune-related ontologies and pathways (**Fig. S2**). We further found
 130 that the CpG and SNP defining an mQTL were more likely to both map to transcription factor binding
 131 sites (TFBS), to promoters, and to CTCF binding sites (**Fig. 1g, Extended Fig. 4c-e**). Further
 132 supporting a functional role of mQTLs, we observed enrichment of mSNPs in the GWAS Catalog³⁰

133 and eQTL databases, and enrichment of mCpGs in the EWAS Atlas³¹ (Fig. 1h, Extended Fig. 5, Fig.
134 S3).

135
136 **Fig. 1 Identification and validation of mQTLs in East Asians.**

137 **a**, Flowchart of mQTL study in Han Chinese. **b**, Comparing the efficiency of fastQTLmapping with MatrixEQTL²⁴.
138 Every test is repeated five times to take an average. **c**, Atlas of the significant mQTLs in Han Chinese (NSPT), red,
139 cis-; orange, lcis-; blue, trans-. Left, the distribution of the genome-wide significant mQTLs (x-axis: SNPs, y-axis:
140 CpGs), point size indicates significance; Middle, zoom in chromosome 6; Right, an example of a single trans-mQTL,
141 y-axis shows the methylation levels (beta values) of cg19268652 (body of *HIVEP1*) and x-axis shows the three
142 genotypes of rs2233437 (intronic of *NFKBIE*). **d**, Comparing the scaled effect sizes of 2.14M mQTL pairs with an
143 independent Han Chinese dataset (CGZ). **e**, Comparing the scaled effect sizes between Han Chinese (NSPT) and
144 Europeans (FHS)² for 13.03M mQTL associations. **f**, Enrichment of mSNPs and mCpGs in different genomic regions.
145 The y-axis indicates the fold enrichment compared with all tested SNPs or CpGs. The P-value of hypergeometric test
146 is adjusted by Bonferroni method (P.bfadjust). **g**, Enrichment of mQTL pairs in functional elements (CTCF: CTCF-
147 enriched elements, E: enhancers, P: promoters, PF: promoter flanking regions, RE: regulatory elements, TFBS: TF
148 binding sites). Heatmap shows the fold changes (see **Methods**) of SNP-CpG pairs in all combinations of functional
149 categories. **h**, Enrichment of mSNPs and mCpGs in phenotype-related sites from GWAS Catalog³⁰ and EWAS Atlas³¹
150 compared with permuted sites (1×10^4). The y-axis shows the percentage of GWAS or EWAS significant signals
151 overlapped with mQTLs or permuted sites.

153 **Most mQTLs are not cell-type or cell-lineage specific**

154 An important as yet unresolved question relates to whether identified mQTLs are present in all
155 underlying blood cell types or only in specific subsets. To explore this, we applied CellDMC¹⁴, an
156 algorithm designed to detect cell-type specific differential DNAm, in order to identify cell-type
157 specific mQTLs among the mQTLs detected in NSPT (**Methods**). CellDMC was first run at the
158 resolution of two cell-lineages (lymphoid and myeloid) and separately again at the higher resolution
159 of six cell-types (B cells, CD4⁺ and CD8⁺ T cells, natural killer (NK), monocytes (Mono) and
160 neutrophils (Neut)), in both cases identifying mQTLs that were present in each lineage or cell-type
161 (**Fig. 2a**). Using computed sensitivities from simulation¹⁵, we estimate that approximately 90% of
162 mQTLs are shared between lineages (**Fig. 2b**), consistent with previous lower bound estimates (i.e.,
163 70%) obtained on BLUEPRINT data^{7,32}. Cell-lineage/cell-type specific mQTLs were validated in the

independent Han Chinese cohort (CGZ) (**Fig. 2c&d**), as well as in the BLUEPRINT data where DNAm had been measured in purified neutrophils, monocytes and CD4⁺ T-cells (**Fig. 2e**). Importantly, the effect-size of myeloid mQTLs was significantly larger than that of lymphoid mQTLs for both cis and trans-mQTL categories (**Fig. 2f**), consistent with myeloid cells being the dominant component in blood (on average, a sample is composed of 69% myeloid cells and 31% lymphoid cells). Finally, using eFORGE2^{33,34} we asked if myeloid and lymphoid mCpGs displayed enrichment for DNase Hypersensitive Sites (DHSs), as assessed using DHS profiles obtained in myeloid and lymphoid cell-types from ENCODE³⁵, BLUEPRINT³⁶ and the Epigenomic Roadmap³⁷. Interestingly, while myeloid-mCpGs displayed clear specificity, displaying much stronger enrichment of DHSs as defined in myeloid cells, lymphoid mCpGs displayed strong enrichment in both lineages (**Fig. 2g and Extended Fig. 6**).

Fig. 2 Cell-type and cell-lineage specific mQTLs and validation.

a, Number of mQTLs found in each cell-type (top-panel: red=lymphoid cell-types, blue=myeloid cell-types) and cell-lineage (lower panel), by application of CellDMC. **b**, The estimated proportion of shared and specific mQTLs in the two main cell lineages; **c&d**, Validation of cell-type and cell-lineage specific mQTLs in the independent Chinese

181 Han cohort (CGZ). Scatterplots display the t-statistics of association between genotype and DNAm in the discovery
182 cohort (x-axis), vs the corresponding statistics in the validation cohort (y-axis). Odds Ratio and one-tailed Fisher's
183 test P-value are given. **e**, Validation of cell-lineage specific mQTLs in BLUEPRINT (BP) data. Scatterplots display
184 the signed statistical significance of myeloid or lymphoid mQTLs (x-axis) vs the corresponding values in the purified
185 samples from BP (y-axis). P-value of a linear regression is given. Barplots compare the overlap of myeloid/lymphoid
186 mQTLs with those identified in the BP data (BP summary statistics only included index-SNPs and not other SNPs in
187 LD), against the expected number under the null. Binomial P-value is given. Neut = neutrophil, Mono = monocytes,
188 CD4T = CD4⁺ T-cells. **f**, Violin plots of effect-size for cell-lineage specific mQTLs. P-value is from a Wilcoxon rank
189 sum test. **g**, eFORGE analysis of myeloid and lymphoid mCpGs. Significance levels of enrichment (y-axis) of
190 myeloid and lymphoid mCpGs for DHSs as derived in lymphoid and myeloid cell-types (x-axis) from BLUEPRINT,
191 ENCODE, Consolidated Roadmap and original Roadmap. Color of cell-type indicates its source.

193 **Enrichment of mQTLs within TADs and active enhancer domains**

194 We mapped cis-mQTL pairs to chromatin loops derived from PCHi-C (promoter capture Hi-C), and
195 H3K27ac HiChIP (Hi-C chromatin immunoprecipitation) data of blood cells^{38,39}. Approximately 72.3%
196 of cis-mQTL pairs were located within the same topologically associating domain (TAD) defined by
197 PCHi-C data, a 6.5-fold enrichment over genomic background and a 1.7-fold enrichment over
198 distance-matched SNP-CpG pairs (**Fig. 3a, Extended Fig. 4b**). cis-mQTLs were also enriched for
199 loops derived from PCHi-C and H3K27ac HiChIP data (15.6- and 7.1-fold, respectively; **Fig. 3b,**
200 **Extended Fig. 4a**). In total, 26.5% of PCHi-C loops and 13.8% of HiChIP loops colocalized with cis-
201 mQTLs, accounting for 8,488,883 and 177,949 mQTLs, respectively.

202 Next, we collected DNase-seq data of T cells from ENCODE⁴⁰ (such data are not available for
203 neutrophils, the dominant component in blood, **Fig. S4**) to check the degree of co-opening of cis-
204 mQTLs. We found that cis-mQTLs are significantly more co-opening than SNP-CpG pairs derived
205 from the genomic background or from distance-matched pairs (2.2- and 1.2-fold, respectively; **Fig.**
206 **3c**). A similar enrichment was obtained by using DNase-seq data of B cells (2.2- and 1.3-fold,

207 respectively). In addition, we collected H3K27ac ChIP-seq data of neutrophils⁴⁰ and checked the co-
208 activity of cis-mQTLs, finding that CpG-SNP pairs from cis-mQTLs are more co-active than SNP-
209 CpG pairs derived from genomic background and from distance-matched pairs (3.7- and 2.1-fold,
210 respectively; **Fig. 3c**).

211

212 **Most cis/lcis mQTLs arise due to changes in chromatin accessibility**

213 SNPs are known to influence chromatin state^{18,41}, and we hypothesized that a significant proportion of
214 mQTLs could be driven by such variants¹⁹. In order to estimate this proportion, we considered two
215 separate models for how mQTLs could arise (**Fig. 3d&e**). In one model (M1) a SNP located in a
216 regulatory element (RE) could affect the binding of a pioneer factor, altering chromatin accessibility.
217 This in turn can influence DNAm at a distal CpG through 3D interactions (**Fig. 3d**). We considered an
218 mQTL pair to be explained by model M1 if it simultaneously satisfied the following constraints: (1)
219 both SNP and CpG are located within open regions and within a TAD; (2) the SNP in the RE modulates
220 chromatin accessibility. In the competing model (M2), a RE does not necessarily change the
221 accessibility of the region but can still influence DNAm at a distal CpG by disturbing the binding
222 affinity of co-factors or non-pioneer TFs. For an mQTL to be explained by M2, we required it to satisfy
223 the following: (1) the SNP and CpG are located at the binding sites of TFs; (2) the mQTL is located in
224 a RE that has functional interactions with its regulating TF, and where the TF-RE interactions are
225 derived from a gene regulatory network inference model PECA⁴²; (3) the SNP and CpG are located
226 within a TAD. Since for both models mQTLs are required to be within a TAD (**Extended Data Fig.**
227 **4b**), these capture cis/lcis-mQTLs.

228 In order to identify mQTLs explained by M1, we used our OpenCausal tool¹⁹ in conjunction with TF-
229 expression levels in blood to identify SNPs in REs that may influence chromatin accessibility in blood
230 tissue. We checked the enrichment of our mSNPs in such potential open-causal SNPs, finding a highly
231 significant 2-fold enrichment (Fisher's test: P-value = 1.6×10^{-16} ; **Fig. 3f**). Using this approach, we

232 estimated that 38.1% of mQTLs are explained by model M1. In contrast, we estimated about 7.9% of
233 mQTLs to be explainable via model M2. Thus, combining models M1 and M2 we can explain about
234 40.4% of mQTLs (**Fig. 3g**). By collecting RNA-Seq data of NK and Neut⁴³ and using these data as
235 input for the OpenCausal algorithm, we observed that 96.9% of M1 mQTLs are shared between NK
236 and Neut, suggesting that most of the M1 mQTLs are independent of cell type (**Fig. 3h**).

Fig. 3 Mechanistic interpretation of cis- and lcis-mQTLs.

a, The schema of mQTL pairs located within/across/outside a TAD (above); Comparison of the number of SNP-CpG pairs within/across/outside TADs for mQTLs and two control groups (below). **b**, The schema of mQTL pairs validated by a chromatin loop (above); Comparison of the number of SNP-CpG pairs validated by PCHi-C or HiChIP loops

(below). **c**, The schema of mQTL pairs being co-opening or co-active; Comparison of the number of co-opening/co-active SNP-CpG pairs (below). **d**, Schematic overviews of the M1 regulatory process. **e**, Schematic overviews of the M2 regulatory process. **f**, Enrichment of mQTLs in OpenCausal variants of blood tissue derived from an external database (Fisher's test: $P\text{-value} = 1.6 \times 10^{-16}$, 2.02-fold); **g**, Proportion of mQTLs interpreted by the two mechanisms M1 and M2; **h**, The proportion of potential opening-causal mQTLs overlapped/specific in the two main cell lineages, myeloid and lymphoid.

248

249 **Trans-mQTL hotspots are mediated by nearby TFs and map to super-enhancers**

250 Previous mQTL studies have indicated a role of transcription factors (TFs) in the formation of trans-
 251 mQTLs^{1,8,44-48}. In line with this, our trans-mSNPs were more likely to be surrounded by TFs (Fisher's
 252 exact test, $P\text{-value} < 1 \times 10^{-323}$, OR = 1.53, **Fig. 4a**). Furthermore, we found that the number of trans-
 253 mSNPs on each chromosome was highly correlated with TF density (**Extended Fig. 7, Fig. S5**), even
 254 after adjusting for gene density (**Fig. 4b, Table S1**). Next, we defined a trans-mQTL hotspot^{1,8,44-50} as
 255 a trans-mQTL network where the number of unlinked trans-mCpGs (**Methods**) associated with the
 256 index SNP (the "hotness-index") is particularly large (**Fig. 4c, Fig. S6, Table S2**). With hotness-index
 257 increasing, the proportion of mSNPs in hotspots surrounded by TFs was monotonically increasing (**Fig.**
 258 **4c**). Of note, the number of trans-mQTL hotspots on chr19 was much larger than other chromosomes
 259 (**Fig. 4d&e**). These hotspots often fell in regions of high TF density (**Fig. 4d**), a pattern observed on
 260 most chromosomes (**Fig. S7-S27**), indicating a close relationship between hotspots and TF density.
 261 Among the top 1% hotspots ($n = 16$, indexed from H1 to H16, hotness-index > 97, **Fig. 4e, Table 1**),
 262 15 were surrounded by at least one TF ($\pm 1\text{Mbp}$ of the index SNP), and the remaining one (H9) mapped
 263 to *CHD7*, a DNA binding protein (DBP) coding gene. For 10 of the 15 TF-surrounding hotspots, their
 264 trans-mCpGs were significantly enriched for the corresponding TF binding motifs, as assessed using
 265 two different motif enrichment tools (TFmotifView⁵¹ and PWMEnrich⁵², **Methods, Extended Fig. 8**).
 266 Most of the associated trans-mCpGs displayed directional skews in their DNAm change, e.g., trans-

267 mCpGs associated with rs4666078 (H2) predominantly displayed hypomethylation with the G allele
268 (**Table 1, Extended Fig. 8**), which is consistent with these DNAm changes arising because of
269 differential TF binding. These lines of evidence further emphasize the important role of TFs in the
270 formation of trans-mQTL hotspots. Incidentally, half of these hotspots (8/16) overlapped with the top-
271 ranked hotspots reported in the FHS cohort (EUR)², indicating that they are pan-ethnic (**Fig. 4e**).
272 Previous studies have shown that mSNPs in trans-mQTL hotspots are likely also cis-eQTLs affecting
273 nearby TF expression^{1,2}. However, of the 16 hotspots identified here, only five contained cis-eQTLs
274 linked to a nearby TF (**Table S3**). Even at a nominal level, there were four hotspots which did not
275 contain cis-eQTLs of a nearby TF (**Table S4**). On the other hand, all 16 hotspots identified in our data
276 contained super enhancers (**Fig. 4f, Fig. S28, Table S5**). Given that super-enhancers play key roles in
277 specifying cell-type function⁵³, the trans-mQTL hotspots identified here could reflect modulation of
278 cell-type function and identity mediated by the corresponding TFs and DBP. Of the 16 hotspots, 13
279 localized to promoter-promoter interaction regions, and 16 colocalized to promoter-other interaction
280 regions. Moreover, all 16 hotspots contained at least one mSNPs with significant OpenCausal scores
281 and 13 hotspots were significantly enriched for such potential open-causal variants when compared to
282 non-hotspot mSNPs (**Fig. 4f, Fig. S28, Table S6**). These findings are consistent with super-enhancers
283 containing many DNA binding events and with their sensitivity to perturbations⁵⁴, as indeed their role
284 is to maintain and modulate the high expression of TFs.

285

Fig. 4 trans-mQTL hotspots mediated by TFs and super enhancers.

a, Relation between trans-mQTL SNPs and near-TFs (± 1 Mbp) SNPs. y-axis shows the proportion of trans-mSNPs in the SNPs sets. P-value and odds ratio (OR) were from Fisher's exact test. **b**, The relation between density (count/Mbp) of trans-mSNPs (x-axis) and TFs after correcting PCGs density (y-axis). Chromosome 19 is indicated by orange dot, and other chromosomes are indicated by blue dots. **c**, Relation among the number of trans-mQTL hotspot, proportion of mSNPs near TF and the threshold of defining hotspots. With threshold increases, the number of trans-mSNP loci above the threshold goes down (red line, right y-axis), but the proportion of mSNPs near-TFs goes up (blue line, left y-axis). **d**, The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr19. P-values and correlation coefficients were from Spearman's rank correlation test. **e**, The distribution of the associated unlinked trans-mCpGs count (y-axis) for each trans-mSNP identified in NSPT. The 16 trans-mQTL hotspots are labeled by "H1" to "H16". "*" indicates the hotspots are also reported in the FHS cohort². **f**, Annotation of cis-eQTLs and super enhancer according to each TF/DBP in 16 trans-mQTL hotspot regions; trans-mSNPs in each hotspot are located in the same TAD regions, enriched with HiChIP interactions and significant OpenCausal signals.

A *FOSL2* mediated trans-mQTL hotspot is causally associated with eosinophil counts

We next set out to explore the potential significance of our hotspots in disease (**Extended Fig. 9, Fig. S29**). One hotspot localized on chr2 (H2, **Fig. 5a, Fig. S30**) harbors *FOSL2* gene, a leucine zipper type TF that forms the AP-1 (activator protein 1) complex involved in a variety of biological processes such as cell differentiation and immune system development⁵⁵⁻⁵⁸. The index SNP (rs4666078) was trans-associated with 232 unlinked mCpGs, all of which displayed hypomethylation with the G allele. The 232 trans-mCpGs were significantly enriched for *FOSL2* motifs (P-value= 1.31×10^{-136} , **Table 1, Fig. S31a**) and *FOSL2* ChIP-seq binding sites (data from ChIP-Atlas database⁵⁹) (enriched proportion = 78%, enrichment fold = 5.2, P-value = 1.25×10^{-101} , **Fig. 5a**). The majority (141, 60.8%) of the 232 mQTLs were detected exclusively in the myeloid lineage (**Fig. 5b&c, Fig. S31b**). Using the EWAS Atlas³¹, we found the 232 mCpGs to be mainly enriched in immune system disorders (P.bfadjust =

313 1.01×10^{-8}), neurological disorders ($P_{\text{badjust}} = 1.30 \times 10^{-8}$) and other diseases like asthma (P_{badjust}
 314 $= 1.03 \times 10^{-3}$) (Hypergeometric test, **Fig. 5d**). Importantly, we also observed significant enrichment
 315 (Fold = 5.3, FDR = 1.2×10^{-5}) for genes associated with eosinophil count (using the R package
 316 *disgenet2r* and DisGeNET knowledgebase⁶⁰) (**Fig. 5e**). In line with this and according to a recent meta-
 317 analysis based on EAS populations⁶¹, rs4666078 has been associated with eosinophil counts in blood
 318 with the G allele associating negatively with counts ($N = 86,890$, Beta = -0.015, P-value = 4.3×10^{-3}).
 319 Based on a recent epigenome-wide association study of tissue eosinophilia in 147 Korean individuals⁶²,
 320 we found that 20% ($n = 46$) of the 232 *FOSL2* mediated mCpGs were significantly (FDR < 0.05)
 321 associated with the risk of tissue eosinophilia, a highly significant enrichment of eosinophilia-
 322 associated CpGs compared to background CpGs from the genome (Hypergeometric test, P-value =
 323 2.19×10^{-24} , **Fig. 5f**). The distribution of P-values of the 232 mCpG also deviated significantly from
 324 those of background CpGs (Kolmogorov-Smirnov test, P-value = 7.31×10^{-7} , **Fig. 5f**). For the majority
 325 of the 46 sites ($N = 39$, 85%), methylation levels were positively correlated with the risk of tissue
 326 eosinophilia, and negatively correlated with the number of the risk allele (G allele), in accord with
 327 GWAS meta-analysis results⁶¹ (**Extended Fig. 10**). To shed light on potential causal mediation, we
 328 performed two-sample Mendelian Randomization (MR) analysis between the 232 mCpGs and blood
 329 eosinophil counts, which identified 24 CpGs (FDR < 0.05) as putative causal factors of blood
 330 eosinophil counts (reverse was not significant, **Fig. 5g, Table S7**). In addition, the methylation levels
 331 of all 24 CpGs were positively correlated with the phenotype. Of note, five of these 24 CpGs also
 332 showed significant positive associations with tissue eosinophilia in the study of Kim et al (2021)⁶² and
 333 one of them (cg22652934) mapped to *RUNX1*, a TF well-known to be functionally involved in the
 334 specification of myeloid and lymphoid cell-lineages from hematopoietic stem cells⁶³. These results
 335 suggest that the *FOSL2*-mediated hotspot may modulate eosinophil counts via modulation of DNAm
 336 at multiple trans-mCpGs.

Fig. 5 The *FOSL2* trans-mQTL hotspot (H2) associated with eosinophil counts.

a, Circos plot displaying the enrichment of the 232 rs4666078 trans-mCpGs in *FOSL2* ChIP-seq binding sites (data from ChIP-Atlas⁵⁹). The inner link lines show the 232 trans-mCpGs, the ones overlapped with *FOSL2* binding sites denoted by orange. The inner barplots show $-\log_{10}(P)$ of trans-mQTL associations. The outer blue dots represent hypomethylation with the G allele. The outer green barplots show $-\log_{10}(Q)$ of peak-caller MACS2 indicating ChIP-seq signals score. **b**, Barplot showing the proportion of the 232 mQTL pairs that existing in myeloid or lymphoid lineage (cell-specific mQTL results filtered by $P\text{-value} < 0.05/232 = 2.16 \times 10^{-4}$). **c**, Circos plot displaying the enrichment of the 141 rs4666078 myeloid-lineage trans-mCpGs in *FOSL2* ChIP-seq binding sites. **d**, Enrichment of the 232 mCpGs for EFO parent categories based on the curated EWAS results³¹ compared with all rs4666078 trans-wide CpGs. P-values of hypergeometric test are adjusted by Bonferroni method. **e**, Enrichment of rs4666078 trans-mCpGs annotated genes for human diseases and traits based on DisGeNET database⁶⁰ compared with all genes in human genome. **f**, Upper panel showing the enrichment of these trans-mCpGs in tissue eosinophilia associated CpGs reported in 147 Korean individuals⁶² (Hypergeometric test). Lower panel showing the association significance with tissue eosinophilia (y-axis) of the 232 mCpGs (red) compared with all tested CpGs (blue). Kolmogorov-Smirnov test is used to evaluate the significance. **g**, Two-Sample MR results showing 24 CpGs are causal for eosinophil counts in whole blood ($FDR < 5\%$). The effect (x-axis), the significance (y-axis) of MR, and CpG-annotated genes are shown.

A *NFKB1* mediated trans-mQTL hotspot is causally associated with BMI

Another important trans-mQTL hotspot was driven by a GWAS SNP associated with ulcerative colitis (UC) and linked *in-cis* with the transcription factor *NFKB1*, consistent with the findings of Bonder et al (2017)¹, and thus validating this *NFKB1* trans-mQTL network in an EAS population (**Fig. 6, Fig. S32**). Here we verified that trans-mQTLs linked to this SNP were not only strongly enriched for *NFKB1* binding motifs, but also for *NFKB1* binding sites as assessed using ChIP-Seq data from the ChIP-Seq atlas⁵⁹ (**Methods, Fig. 6a&b**). Given that a recent EWAS for BMI has revealed strong enrichment for the *NFKB*-pathway⁶⁴, we hypothesized that some of the *NFKB1* trans-mQTLs may be associated with BMI. In order to test this in the context of our Han Chinese cohort, we first established

that a list of 364 CpGs known to be associated with BMI (as derived from Wahl et al (2019)⁶⁴ and other studies), did so also in our Asian cohort and with the same directionality of DNAm change (**Fig. 6c**). We observed that 15 of the 364 BMI CpGs were *NFKB1* trans-mQTLs (Fisher-test, P-value = 2.7×10^{-19}), and that *NFKB1* trans-mQTLs displayed a very strong association with BMI in our Chinese cohort, as well as in an independent white Caucasian cohort (**Fig. 6d&e**). Next, we asked if *NFKB1* trans-mQTLs enriched for NFKB1 binding motifs/sites and which are also associated with BMI in our cohort, displayed any evidence of being causal mediators of BMI. Using a 2-step MR strategy⁶⁴⁻⁶⁶ (**Methods**), we observed stronger evidence for DNAm levels at *NFKB1* trans-mQTLs being causal mediators for BMI, as opposed to being a consequence of BMI (**Fig. 6f, Table S8**). In contrast, when considering the BMI-associated CpGs (excluding the subset of *NFKB1* trans-mQTLs) we observed stronger evidence for DNAm at these sites being a consequence of BMI (**Fig. 6g, Table S9**), consistent with previous observations⁶⁴. Thus, these data suggest that DNAm at a subset of loci of NFKB1 binding targets may be causal for high BMI, with a component of variation at these loci being under genetic influence. We next reasoned that if DNAm at these loci mediates risk of obesity, that these loci may also display a significant interaction between genotype and BMI that better models their DNAm variation. We were able to confirm a number of *NFKB1* trans-mQTLs that displayed a significant interaction between genotype and BMI (**Fig. 6h**). Of note, among the 3 CpGs exhibiting both causal and interaction effects, one mapped to *PTPN3*, a protein-tyrosine-phosphatase gene that has been linked causally to obesity⁶⁷, and another to *NOD2*, an intracellular innate immunity protein gene that has been shown to be protective of diet-induced obesity and colitis⁶⁸⁻⁷⁰ and which also has polymorphisms associated with inflammatory bowel disease (IBD)⁷¹. Other genes highlighted by our interaction analysis have also been previously implicated in obesity, IBD or type-2 diabetes, including *FOXP1*⁷², *CDKAL1*⁷³⁻⁷⁵, *HSD17B12*⁷⁶, *ALAS1*⁷⁷, *COL15A1*⁷⁸ and *DNAH10*⁷⁹.

Fig. 6 The *NFKB1* trans-mQTL hotspot associated with BMI.

a-b, Circos plots displaying the enrichment of *NFKB1* binding motifs (left) and *NFKB1* ChIP-Seq binding sites (right) among trans-mQTLs associated with the SNP linked in-cis with *NFKB1*. The fraction of trans-mQTLs enriched for the binding motif/site is given below circos plot. The inner track labels the enriched trans-mQTLs (red). Red barplots display the corresponding $-\log_{10}(P)$ of enrichment. Blue barplots display $-\log_{10}(P)$ of trans-mQTLs.

Outer track represents the directionality of DNAm change of the trans-mQTLs: cyan (hypermethylation), magenta (hypomethylation). **c**, scatterplots displaying the signed statistical significance of 364 BMI-associated CpGs reported in the literature (y-axis) against their corresponding values in NSPT (x-axis). **d**, Association of *NFKB1* trans-mQTLs with BMI in 2 independent cohorts (NSPT and Buccal-IC). Plots display the statistical significance (y-axis) vs. the CpG index position (x-axis) for the *NFKB1* trans-mQTLs (red line) vs that of randomly selected CpGs (blue). P-value is from a Binomial test. **e**, Scatterplots displaying the t-statistics of association of DNAm of *NFKB1* trans-mQTLs with risk allele (y-axis) vs. their corresponding t-statistics of association of DNAm with BMI in the 2 independent cohorts (x-axis). **f**, MR for the 35 *NFKB1* trans-mQTLs associated with BMI. x-axis labels statistical significance (adjusted by FDR). Major color indicates direction of effect, with left panel displaying results for model where DNAm mediates the effect of SNP on obesity, right panel is for model where obesity affects DNAm. **g**, As f, but for BMI-associated CpGs reported by BMI EWAS studies. **h**, 3d scatterplot displaying the interaction between BMI and genotype on BMI-associated *NFKB1* trans-mQTLs. x-axis labels the significance level of the univariate association between DNAm and genotype, y-axis labels the significance level of the univariate association between DNAm and BMI, z-axis labels the significance level of the interaction term between genotype and BMI.

Discussion

This work advances our understanding of mQTLs in the following ways. First of all, our results clearly indicate that at least 80% of mQTLs are pan-ethnic, and therefore that the mechanisms underlying the formation of mQTLs unlikely differs between ethnic groups. While ethnic-specific mQTLs likely exist, as indeed a small number of mQTL hotspots identified in our EAS population were not reported in previous studies, we are not able to discard other factors (e.g., differences in power or Illumina beadarray version) driving these differences. The implications of these findings are profound: it suggests that variations in genetic disease predisposition between ethnic groups should not be studied in isolation, pointing towards the need to incorporate the ethnic and individual specific exposome, to better understand how DNA methylation may mediate disease risk through additional modulation associated with environmental factors. Second, our results indicate that cell-type specific mQTLs are

420 relatively uncommon. A cautionary note here is that we have only assessed cell-type specificity
421 between blood cell subtypes, still leaving open the possibility that mQTLs could vary substantially
422 when compared across more distinct cell-types (e.g., immune cells vs epithelial or fibroblasts). In line
423 with the non-specificity in blood, TFs implicated in mQTLs (e.g., *NFKB1* or *CTCF*) are generally not
424 blood-cell subtype specific, displaying variable but consistently non-zero expression between such cell
425 subtypes. To ascertain in which specific cell-types a given mQTL is truly not present in will however
426 require more adequately powered studies¹⁵. Third, we have been able to demonstrate that mQTLs are
427 strongly enriched for open-causal variants and that a much higher proportion of cis- and lcis-mQTLs
428 arise due to modulation of a pioneer TF that affects chromatin accessibility and subsequent propagation
429 of distal DNAm changes via chromatin looping within TADs. Fourth, we could provide several lines
430 of evidence, although observational, to further strengthen the key role of TFs bridging genetic variants
431 and methylation levels in the formation of trans-mQTLs and hotspots. Undoubtedly, other factors, e.g.,
432 SNPs affecting the activity of methylation binding domain (MBD), DNA methyltransferases (DNMTs)
433 or ten-eleven translocations (TETs)⁴⁴ could give rise to mQTLs (**Extended Fig. 11**), and future studies
434 should explore this. Fifth, our trans-mQTL hotspot analysis has shown that many key hotspots (e.g.,
435 *NFKB1*) are independent of ethnicity, while also revealing a novel hotspot associated with *FOSL2* for
436 which MR-analysis supports a potentially causal role of specific mQTLs modulating eosinophil counts,
437 possibly involving the *RUNX1* TF. To establish causality will require experimental validation⁸⁰.
438 Besides the potential role of the *NFKB1* hotspot in mediating the risk of ulcerative colitis¹, our analyses
439 has revealed a separate role in obesity. By performing MR on the subset of trans-mQTLs that co-
440 localized with *NFKB1* binding and that are also associated with BMI, we have discovered that this
441 subset could be causally implicated in mediating the risk of obesity. This is a novel and important
442 finding since MR performed on BMI EWAS associated differentially methylated cytosines
443 overwhelmingly point to DNAm variation at these loci being a consequence of obesity⁶⁴, a result which
444 we were able to replicate here. To our knowledge, this is the first study to explore if specific trans-

445 mQTLs co-localizing with TF-binding are causally implicated in relation to a disease or outcome, and
446 the positive findings we report could represent a novel paradigm for identifying potentially causal
447 mQTLs. In support of this, we note that several of the identified mCpGs map to genes (e.g., *NOD2*,
448 *PTPN3*) that in mouse models have already been causally implicated in obesity⁶⁷⁻⁷⁰. We note that these
449 findings do not contradict the dearth of casual associations reported recently by Min et al (2021)⁸, as
450 this latter study did not specifically focus on trans-mCpGs co-localizing with TF-binding.

451 In summary, this work advances our understanding of the mQTL landscape across ethnicities, cell-
452 types and in terms of the underlying mechanistic models that shape this complex landscape. In so doing,
453 we contribute the first mQTL database in an Asian population, an invaluable resource for
454 understanding the genetic and epigenetic variations in disease predisposition between ethnic groups.

455

456   **Table 1 nearby TFs in 16 trans-mQTL hotspots and the enrichment of corresponding trans-mCpG regions for the motifs**

HotSpot	Index SNP	Index SNP (CHR:BP)	EA	OA	Number of trans- mCpGs	POS/NE G	TF (<1Mbp)	Putative TF (Motif)	TFmotifView			PWMErich		Mye/Lym ^a
									MotifEnrich	Enrich Fold	P _{Enrich}	MotifEnrich	P _{Enrich}	
H1	rs78823853	1:61606620	A	G	140	0/140	<i>NFIA</i>	<i>NFIA</i> (MA0670.1)	49 (35.00%)	3.75	2.07×10 ⁻¹⁶	46 (32.86%)	1.20×10 ⁻¹¹	78/0
H2	rs4666078	2:28691245	G	A	232	0/232	<i>ZNF512, FOSL2</i>	<i>FOSL2</i> (MA0478.1)	178 (76.72%)	17.48	2.50×10 ⁻¹⁷⁷	183 (78.88%)	1.31×10 ⁻¹³⁶	141/0
H4	rs58408429	4:57769824	C	T	455	385/70	<i>REST</i>	<i>REST</i> (MA0138.2)	90 (19.78%)	180.00	3.48×10 ⁻¹¹³	110 (24.18%)	7.77×10 ⁻²⁵³	219/0
H5	rs3774937	4:103434253	C	T	448	3/445	<i>NFKB1</i>	<i>NFKB1</i> (MA0105.4)	21 (4.69%)	241.50	1.54×10 ⁻²⁷	255 (56.92%)	2.76×10 ⁻²⁸¹	152/2
H8	rs143396005	7:115645268	T	C	682	10/672	<i>TFEC</i>	<i>TFEC</i> (MA0871.2)	79 (11.58%)	5.59	2.10×10 ⁻³⁰	276 (40.47%)	4.26×10 ⁻¹⁶³	197/3
H10	rs10883359	10:101274033	G	A	592	155/437	<i>NKX2-3</i>	<i>NKX2_3</i> (MA0672.1)	149 (25.17%)	1.25	2.42×10 ⁻³	52 (8.78%)	5.81×10 ⁻⁵	134/5
H11	rs3809627	16:30103160	C	A	195	39/156	<i>TBX6, MAZ, ZNF48, ZNF771, ZNF768, ZNF747, ZNF764, ZNF688, ZNF785, ZNF689, SRCAP, ZNF629, FBXL19, ZNF668, ZNF646</i>	<i>TBX6</i> (MA1567.1)	173 (88.72%)	1.40	1.30×10 ⁻¹⁵	74 (37.95%)	5.06×10 ⁻⁶⁴	148/4
							<i>HLF</i>	<i>HLF</i> (MA0043.3)	18 (9.63%)	2.88	8.33×10 ⁻⁵	20 (10.70%)	2.67×10 ⁻³	92/1
							<i>RELB, ZNF296, FOSB, SIX5, FOXA3, MYPOP, HIF3A, ARHGAP35</i>	<i>FOXA3</i> (MA1683.1)	86 (68.25%)	2.50	2.79×10 ⁻²¹	43 (34.13%)	2.07×10 ⁻³⁷	41/1
H15	rs28789846	20:31275545	A	G	402	40/20	<i>FOXSI, PLAGL2, E2F1</i>	<i>PLAGL2</i> (MA1548.1)	101 (25.12%)	1.47	4.62×10 ⁻⁵	35 (8.71%)	6.34×10 ⁻⁶	91/1
H3	rs62175733	2:174231184	T	C	1056	105/60	<i>SP3, SP9</i>	<i>SP9</i> (MA1564.1)	162 (15.34%)	0.46	1.00	8 (0.76%)	1.00	8/8
H6	rs651297	6:44260324	T	C	180	1/179	<i>ZNF318, CDC5L</i>							130/0
H7	rs4473920	7:21974610	G	A	192	0/192	<i>SP4</i>							107/0
H9	rs1038353	8:61670476	G	A	242	47/195								146/0
H13	rs11082385	18:42040114	T	C	218	94/124	<i>SETBP1</i>							26/1
H16	rs7275212	21:39852551	T	A	1149	919/230	<i>ERG, ETS2</i>	<i>ERG</i> (MA0474.2)	143 (12.45%)	1.14	0.07	57 (4.96%)	0.71	341/2

457   ^a Significant myeloid/lymphoid lineage mCpGs that passed Bonferroni corrected thresholds among all of the trans-mCpGs.

458 **Acknowledgement:** This work is supported by the National Science Foundation of China (NSFC)
459 (grant numbers 31571359, 31771464, 31970632, 81930056, 32170657), the National Key Research
460 and Development Project (Grant No. 2018YFC0910403, 2018YFE0201603), Ministry of Science
461 and Technology (2015FY111700), Shanghai Municipal Science and Technology Major Project
462 (Grant No. 2017SHZDZX01), the Strategic Priority Research Program of Chinese Academy of
463 Sciences (Grant No. XDB38010400, XDC01000000), Science and Technology Service Network
464 Initiative of Chinese Academy of Sciences (KFJ-STQ-QYZD-2021-08-001, KFJ-STQ-ZDTP-079),
465 CAS Interdisciplinary Innovation Team, CAS Youth Innovation Promotion Association (Grant No.
466 2020276), China Postdoctoral Science Foundation (Grant No. 2021M693274 and BX2021336). We
467 are grateful to Prof. Chaolong Wang from Huazhong University of Science and Technology for
468 helpful discussion.

470 **Author contributions:** SW, FL and AT designed and drafted the work. QP, XL, HJ, and WL
471 performed the statistical analyses and contributed to data interpretation and writing of the
472 manuscript. XG, QL, CB, GL and SP contributed to data analysis. JL, NY, JQ, LY and GZ generated
473 the mQTL database. CY and SD contributed to preprocessing of methylation and SNP chip data in
474 the discovery cohort. YZ, XL and LS contributed to the design and acquisition of data in the
475 validation panel. LJ, JW, JT and ZY contributed to the design and acquisition of data in the discovery
476 cohort. SG, YL, TN and BW contributed to design of this work. All the authors revised this work,
477 approved the submitted version, agreed with personal contributions and are responsible for the
478 integrity of the data and the accuracy of the data analysis.

480 **Data Availability:** Our mQTL database (NSPT, N=3,523) is available at
481 <https://www.biosino.org/meqtl/home>. The statistics of mQTLs in NSPT cohort can also be viewed at
482 NODE <https://www.biosino.org/node> under accession number OEP002902, or directly at

483 <https://www.biosino.org/node/project/detail/OEP002902>, and accessed by submitting a request for
484 data-access. The Illumina EPIC DNAm data and SNP chip data in NSPT and CGZ are accessed by
485 submitting a request for data-access. Data usage shall be in full compliance with the Regulations on
486 Management of Human Genetic Resources in China. The buccal swab DNAm data are only available
487 by submitting data requests to mrc1ha.swiftinfo@ucl.ac.uk; see full policy at
488 <http://www.nshd.mrc.ac.uk/data.aspx>. all other relevant data supporting the key findings of this study
489 are available within the article and its Supplementary files or from the corresponding author upon
490 reasonable request. The mQTL results of the EUR cohort (FHS, N=4170) were downloaded from
491 https://ftp.ncbi.nlm.nih.gov/eqtl/original_submissions/FHS_mQTLs/. The annotation of CpG probes
492 was downloaded from <https://zwdzwd.github.io/InfiniumAnnotation>. Significant GWAS results were
493 downloaded from GWAS Catalog (<https://www.ebi.ac.uk/gwas/docs/file-downloads>) and significant
494 EWAS results was downloaded from EWAS Atlas (<https://ngdc.cncb.ac.cn/ewas/downloads>). The
495 cis-eQTL results in whole blood were downloaded from GTEx V8 database
496 (<https://www.gtexportal.org/home/datasets>). Human gene information was downloaded from
497 GENCODE (https://www.encodegenes.org/human/release_37lift37.html), and the data of human
498 transcription factors were from <http://humantfs.cabr.utoronto.ca/download.php>. Motifs information
499 of TFs was obtained from JASPAR 2020 database (<http://jaspar.genereg.net/>). ChIP-seq signals of
500 TFs were downloaded from ChIP-Atlas database (<http://chip-atlas.org/>). Other data sources used in
501 this study include: BLUEPRINT mQTLs summary statistics ([https://ega-](https://ega-archive.org/datasets/EGAD00001005200)
502 [archive.org/datasets/EGAD00001005200](https://ega-archive.org/datasets/EGAD00001005200)); Phenoscanner GWAS summary statistics
503 (<http://www.phenoscanner.medschl.cam.ac.uk/>); Functional genomic regions from the Functional
504 Annotation of ANimal Genomes (FAANG) Project (<https://www.faang.org>); PCHi-C data
505 (<https://osf.io/u8tzp>); H3K27ac HiChIP data (<https://www.ncbi.nlm.nih.gov/geo/>, GSE101498); The
506 DNase-seq data for B cells and T cells and the H3K27ac ChIP-seq data of neutrophil cells
507 (<https://www.encodeproject.org>).

508 **Code Availability:** Most operations were carried out by R (<https://cran.r-project.org/>), and the plots
509 were mainly made by ggplot2 R package ([https://cran.r-](https://cran.r-project.org/web/packages/ggplot2/index.html)
510 [project.org/web/packages/ggplot2/index.html](https://cran.r-project.org/web/packages/ggplot2/index.html)). FastQTLmapping is available at
511 <https://github.com/TianTTL/xQTLmapping>. CellDMC is available as part of the EpiDISH
512 Bioconductor R-package, which is freely available from
513 <http://bioconductor.org/packages/devel/EpiDISH>. eFORGE was run with the webserver at
514 <https://eforge.altiusinstitute.org/>. The GO and KEGG pathway enrichment analyses of mCpGs were
515 conducted using R package missMethyl
516 (<https://bioconductor.org/packages/3.13/bioc/html/missMethyl.html>). Genes enrichment for
517 diseases/traits analysis was performed by the R package disgenet2r
518 (<https://www.disgenet.org/disgenet2r>). The Two-Sample MR analysis was conducted using the
519 TwoSampleMR R package (<https://mrcieu.github.io/TwoSampleMR/>). The HiChIP loops were
520 processed by HiCCUPS implemented in the Juicer Tools (version 0.7.5) with default parameter
521 settings. The influence of SNPs on regulatory elements was calculated using the tool OpenCausal
522 (<https://github.com/liwenran/OpenCausal>).

523

524 **Ethics:**

525 This research is a sub project of The National Science & Technology Basic Research Project which
526 was approved by the Ethics Committee of Human Genetic Resources of School of Life Sciences,
527 Fudan University, Shanghai (14117). The Declaration of Helsinki Principles was followed and all
528 participants provided written informed consent.

529

530 **Competing interests:**

531 The authors declare that they have no competing interests.

532

533

Reference

1. Bonder, M.J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* **49**, 131-138 (2017).
2. Huan, T. *et al.* Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. *Nat Commun* **10**, 4267 (2019).
3. McRae, A.F. *et al.* Identification of 55,000 Replicated DNA Methylation QTL. *Sci Rep* **8**, 17605 (2018).
4. van Dongen, J. *et al.* Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun* **7**, 11115 (2016).
5. Gaunt, T.R. *et al.* Systematic identification of genetic influences on methylation across the human life course. *Genome Biol* **17**, 61 (2016).
6. Hannon, E. *et al.* Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the Relationship between Methylomic Variation, Gene Expression, and Complex Traits. *Am J Hum Genet* **103**, 654-665 (2018).
7. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* **167**, 1398-1414.e24 (2016).
8. Min, J.L. *et al.* Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet* **53**, 1311-1321 (2021).
9. McClay, J.L. *et al.* High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction. *Genome Biol* **16**, 291 (2015).
10. Lemire, M. *et al.* Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. *Nat Commun* **6**, 6326 (2015).
11. Banovich, N.E. *et al.* Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLoS Genet* **10**, e1004663 (2014).
12. Bell, C.G. *et al.* Obligatory and facilitative allelic variation in the DNA methylome within common disease-associated loci. *Nat Commun* **9**, 8 (2018).
13. Liu, Y. *et al.* GeMes, clusters of DNA methylation under genetic control, can inform genetic and epigenetic analysis of disease. *Am J Hum Genet* **94**, 485-95 (2014).
14. Zheng, S.C., Breeze, C.E., Beck, S. & Teschendorff, A.E. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods* **15**, 1059-1066 (2018).
15. You, C. *et al.* A cell-type deconvolution meta-analysis of whole blood EWAS reveals lineage-specific smoking-associated DNA methylation changes. *Nat Commun* **11**, 4779 (2020).
16. Teschendorff, A.E., Zhu, T., Breeze, C.E. & Beck, S. EPISCORE: cell type deconvolution of bulk tissue DNA methylomes from single-cell RNA-Seq data. *Genome Biol* **21**, 221 (2020).
17. Tsai, P.C. *et al.* Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health. *Clin Epigenetics* **10**, 126 (2018).
18. Kumasaka, N., Knights, A.J. & Gaffney, D.J. High-resolution genetic mapping of putative causal interactions between regions of open chromatin. *Nat Genet* **51**, 128-137 (2019).
19. Li, W., Duren, Z., Jiang, R. & Wong, W.H. A method for scoring the cell type-specific impacts of noncoding variants in personal genomes. *Proc Natl Acad Sci U S A* **117**, 21364-21372 (2020).
20. Liang, D. *et al.* Cell-type-specific effects of genetic variation on chromatin accessibility during human neuronal differentiation. *Nature Neuroscience* **24**, 941-953 (2021).
21. Moran, S., Arribas, C. & Esteller, M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* **8**, 389-99 (2016).

- 577 22. Zaimi, I. *et al.* Variation in DNA methylation of human blood over a 1-year period using the Illumina
578 MethylationEPIC array. *Epigenetics* **13**, 1056-1071 (2018).
- 579 23. Gao, X. *et al.* FastQTLmapping: an ultra-fast package for mQTL-like analysis. (bioRxiv, 2021).
- 580 24. Shabalin, A.A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-
581 8 (2012).
- 582 25. Shi, J. *et al.* Characterizing the genetic basis of methylome diversity in histologically normal human lung
583 tissue. *Nat Commun* **5**, 3365 (2014).
- 584 26. DeFronzo, R.A. Chiglitazar: a novel pan-PPAR agonist. *Science Bulletin* **66**, 1497-1498 (2021).
- 585 27. Ji, L. *et al.* Efficacy and safety of chiglitazar, a novel peroxisome proliferator-activated receptor pan-agonist,
586 in patients with type 2 diabetes: a randomized, double-blind, placebo-controlled, phase 3 trial (CMAP).
587 *Science Bulletin* **66**, 1571-1580 (2021).
- 588 28. Jia, W. *et al.* Chiglitazar monotherapy with sitagliptin as an active comparator in patients with type 2 diabetes:
589 a randomized, double-blind, phase 3 trial (CMAS). *Science Bulletin* **66**, 1581-1590 (2021).
- 590 29. Lin, D. *et al.* Characterization of cross-tissue genetic-epigenetic effects and their patterns in schizophrenia.
591 *Genome medicine* **10**, 13 (2018).
- 592 30. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted
593 arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-d1012 (2019).
- 594 31. Li, M. *et al.* EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids*
595 *Res* **47**, D983-d988 (2019).
- 596 32. Teschendorff, A.E., Jing, H., Paul, D.S., Virta, J. & Nordhausen, K. Tensorial blind source separation for
597 improved analysis of multi-omic data. *Genome Biol* **19**, 76 (2018).
- 598 33. Breeze, C.E. *et al.* eFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data.
599 *Bioinformatics* **35**, 4767-4769 (2019).
- 600 34. Breeze, C.E. *et al.* eFORGE: A Tool for Identifying Cell Type-Specific Signal in Epigenomic Data. *Cell Rep*
601 **17**, 2137-2150 (2016).
- 602 35. Thurman, R.E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82 (2012).
- 603 36. Adams, D. *et al.* BLUEPRINT to decode the epigenetic signature written in blood. *Nat Biotechnol* **30**, 224-
604 6 (2012).
- 605 37. Bernstein, B.E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-8
606 (2010).
- 607 38. Javierre, B.M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease
608 Variants to Target Gene Promoters. *Cell* **167**, 1369-1384.e19 (2016).
- 609 39. Mumbach, M.R. *et al.* Enhancer connectome in primary human cells identifies target genes of disease-
610 associated DNA elements. *Nat Genet* **49**, 1602-1612 (2017).
- 611 40. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*
612 **489**, 57-74 (2012).
- 613 41. Currin, K.W. *et al.* Genetic effects on liver chromatin accessibility identify disease regulatory variants. *Am*
614 *J Hum Genet* **108**, 1169-1189 (2021).
- 615 42. Duren, Z., Chen, X., Jiang, R., Wang, Y. & Wong, W.H. Modeling gene regulation from paired expression
616 and chromatin accessibility data. *Proc Natl Acad Sci U S A* **114**, E4914-e4923 (2017).
- 617 43. Choi, J. *et al.* Haemopedia RNA-seq: a database of gene expression during haematopoiesis in mice and
618 humans. *Nucleic Acids Res* **47**, D780-d785 (2019).
- 619 44. Héberlé, É. & Bardet, A.F. Sensitivity of transcription factors to DNA methylation. *Essays Biochem* **63**, 727-

620 741 (2019).

621 45. Domcke, S. *et al.* Competition between DNA methylation and transcription factors determines binding of
622 NRF1. *Nature* **528**, 575-9 (2015).

623 46. Feldmann, A. *et al.* Transcription factor occupancy can mediate active turnover of DNA methylation at
624 regulatory regions. *PLoS Genet* **9**, e1003994 (2013).

625 47. Lienert, F. *et al.* Identification of genetic elements that autonomously determine DNA methylation states.
626 *Nat Genet* **43**, 1091-7 (2011).

627 48. Stadler, M.B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*
628 **480**, 490-5 (2011).

629 49. Albert, F.W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet*
630 **16**, 197-212 (2015).

631 50. Yao, C. *et al.* Dynamic Role of trans Regulation of Gene Expression in Relation to Complex Traits. *Am J*
632 *Hum Genet* **100**, 571-580 (2017).

633 51. Leporeq, C. *et al.* TFmotifView: a webserver for the visualization of transcription factor motifs in genomic
634 regions. *Nucleic acids research* **48**, W208-W217 (2020).

635 52. Stojnic, R. & Diez, D. PWMEnrich: PWM enrichment analysis. (R package version 4.28.1, 2021).

636 53. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).

637 54. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity
638 genes. *Cell* **153**, 307-19 (2013).

639 55. Davies, J.S., Klein, D.C. & Carter, D.A. Selective genomic targeting by FRA-2/FOSL2 transcription factor:
640 regulation of the Rgs4 gene is mediated by a variant activator protein 1 (AP-1) promoter sequence/CREB-
641 binding protein (CBP) mechanism. *The Journal of biological chemistry* **286**, 15227-15239 (2011).

642 56. Ubieta, K. *et al.* Fra-2 regulates B cell development by enhancing IRF4 and Foxo1 transcription. *The Journal*
643 *of experimental medicine* **214**, 2059-2071 (2017).

644 57. Li, K. *et al.* Landscape and Dynamics of the Transcriptional Regulatory Network During Natural Killer Cell
645 Differentiation. *Genomics, proteomics & bioinformatics* **18**, 501-515 (2020).

646 58. Butler, T.L. & Pennypacker, K.R. The transcriptional response to hypoxic insult controlled by FRA-2. *Gene*
647 *expression* **12**, 61-67 (2005).

648 59. Oki, S. *et al.* ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO*
649 *Rep* **19**(2018).

650 60. Piñero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic acids*
651 *research* **48**, D845-D855 (2020).

652 61. Chen, M.H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5
653 Global Populations. *Cell* **182**, 1198-1213.e14 (2020).

654 62. Kim, K.W. *et al.* Integrated genetic and epigenetic analyses uncover MSI2 association with allergic
655 inflammation. *The Journal of allergy and clinical immunology* **147**, 1453-1463 (2021).

656 63. Goyama, S., Huang, G., Kurokawa, M. & Mulloy, J.C. Posttranslational modifications of RUNX1 as
657 potential anticancer targets. *Oncogene* **34**, 3483-3492 (2015).

658 64. Wahl, S. *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity.
659 *Nature* **541**, 81-86 (2017).

660 65. Relton, C.L. & Davey Smith, G. Two-step epigenetic Mendelian randomization: a strategy for establishing
661 the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol* **41**, 161-76 (2012).

662 66. Dekkers, K.F. *et al.* Blood lipids influence DNA methylation in circulating cells. *Genome Biol* **17**, 138 (2016).

- 663 67. Gurzov, E.N., Stanley, W.J., Brodnicki, T.C. & Thomas, H.E. Protein tyrosine phosphatases: molecular
664 switches in metabolism and diabetes. *Trends Endocrinol Metab* **26**, 30-9 (2015).
- 665 68. Rodriguez-Nunez, I. *et al.* Nod2 and Nod2-regulated microbiota protect BALB/c mice from diet-induced
666 obesity and metabolic dysfunction. *Sci Rep* **7**, 548 (2017).
- 667 69. Gurses, S.A. *et al.* Nod2 protects mice from inflammation and obesity-dependent liver cancer. *Sci Rep* **10**,
668 20519 (2020).
- 669 70. Kreuter, R., Wankell, M., Ahlenstiel, G. & Hebbard, L. The role of obesity in inflammatory bowel disease.
670 *Biochim Biophys Acta Mol Basis Dis* **1865**, 63-72 (2019).
- 671 71. Hugot, J.P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease.
672 *Nature* **411**, 599-603 (2001).
- 673 72. Liu, P. *et al.* Foxp1 controls brown/beige adipocyte differentiation and thermogenesis through regulating
674 beta3-AR desensitization. *Nat Commun* **10**, 5070 (2019).
- 675 73. Palmer, C.J. *et al.* Cdkal1, a type 2 diabetes susceptibility gene, regulates mitochondrial function in adipose
676 tissue. *Mol Metab* **6**, 1212-1225 (2017).
- 677 74. Consortium, U.I.G. *et al.* Genome-wide association study of ulcerative colitis identifies three new
678 susceptibility loci, including the HNF4A region. *Nat Genet* **41**, 1330-4 (2009).
- 679 75. Anderson, C.A. *et al.* Investigation of Crohn's disease risk loci in ulcerative colitis further defines their
680 molecular relationship. *Gastroenterology* **136**, 523-9 e3 (2009).
- 681 76. Hachim, M.Y. *et al.* An Integrative Phenotype-Genotype Approach Using Phenotypic Characteristics from
682 the UAE National Diabetes Study Identifies HSD17B12 as a Candidate Gene for Obesity and Type 2
683 Diabetes. *Genes (Basel)* **11**(2020).
- 684 77. Moreno-Navarrete, J.M. *et al.* Heme Biosynthetic Pathway is Functionally Linked to Adipogenesis via
685 Mitochondrial Respiratory Activity. *Obesity (Silver Spring)* **25**, 1723-1733 (2017).
- 686 78. Cox, B. *et al.* A Co-expression Analysis of the Placental Transcriptome in Association With Maternal Pre-
687 pregnancy BMI and Newborn Birth Weight. *Front Genet* **10**, 354 (2019).
- 688 79. Huang, L.O. *et al.* Genome-wide discovery of genetic loci that uncouple excess adiposity from its
689 comorbidities. *Nat Metab* **3**, 228-243 (2021).
- 690 80. Birney, E., Smith, G.D. & Greally, J.A.-O. Epigenome-wide Association Studies and the Interpretation of
691 Disease -Omics.

692

693 **Extended figures**

694

695 **Extended Fig. 1 Overlap between cis-, lcis- and trans-mQTL SNPs and CpGs.**

696 **a-b**, Two venn plots demonstrating the overlap between cis-, lcis- and trans-mQTL SNPs (plot a) and CpGs (plot b).

697

Extended Fig. 2 mQTL effects along with distance between SNPs and CpGs.

a-c, Three scatter plots showing the relationship between the mQTLs distance (x-axis) and the significance (absolute T value, plot a), the explained mCpG variance (R^2 , plot b) and the effect size (absolute value, plot c) within a local 10Mbp region. These plots are based on the loose threshold results ($P < 1 \times 10^{-10}$).

Extended Fig. 3 EAF difference is larger for FHS-only significant SNPs.

Density plot showing the effect allele frequency (EAF) difference between EAS (NSPT) and EUR (FHS²). The effect alleles in EUR are harmonized with EAS. The x-axis shows the difference of the EAF (EUR-EAS), and different colors represent two SNP sets (red, SNPs in both significant mQTLs; blue, SNPs in FHS-only significant mQTLs).

P-value is from one-tailed T-test.

Extended Fig. 4 mQTLs enrichment for different functional elements.

a, Comparison of distance distributions of mQTLs and that of 3D loops; **b**, proportion of SNP-CpG pairs of mQTL within the same TAD; **c&d**, Enrichment of mQTL SNPs (**c**) and CpGs (**d**) in six functional elements: CTCF-enriched elements (CTCF), enhancers (E), promoters (P), promoter flanking regions (PF), regulatory elements (RE), and TF binding sites (TFBS). **e**, Enrichment of cis-mQTL pairs (left), lciis-mQTL pairs (middle), and trans-mQTL pairs (right) in all combinations of the six functional categories (i.e., CTCF-E, P-E, RE-E, and etc). The fold changes are labeled within each box. The y-axis indicates the fold changes (see Methods) and the significance of

719 enrichment is denoted by different symbols on each bar, i.e., *, P-value < 0.05; **, P-value < 0.01; ***, P-value <
 720 0.001.
 721

722
 723 **Extended Fig. 5 cis-mQTL SNPs are enriched in cis-eQTLs (whole blood, GTEx v8).**
 724 **a**, Bars show the ratio of cis-eQTL SNPs in all tested SNPs (blue) and cis-mQTL SNPs (red). The significance of the
 725 cis-mQTL SNPs enrichment for cis-eQTLs (Hypergeometric test) is labeled on the top of the bars (P.Hypergeometric).
 726 **b**, Enrichment of cis-mQTL SNPs that are also cis-eQTLs in different genomic regions (left) and functional elements
 727 (right). The y-axis indicates the fold enrichment compared with all cis-mQTL SNPs. Significance of one-tailed
 728 hypergeometric test is denoted by different symbols on each bar, i.e., *, P.bfadjust (the P-value is adjusted by
 729 Bonferroni method) < 0.05; **, P.bfadjust < 0.01; ***, P.bfadjust < 0.001. **c-e**, Enrichment of the overlaps between
 730 cis-mQTL SNPs and cis-eQTLs in GO terms (c), KEGG pathways (d), and Reactome pathways (e).

732 **Extended Fig. 6 Mye and Lymph mCpGs enriched in different immune cell types.**
733 **a**, Comparison of DHS numbers per sample suggests that sample quality, rather than immune cell type,
734 is a likely main driver of DHS number; **b**, comparison of FORGE2 chromatin state enrichments for different
735 top Mye and Lymph mQTL CpGs indicates that Lymph results are localized in promoters and enhancers shared across
736 different immune cell types (Mye and Lymph, top), while Mye results localize to enhancers present in a few Mye cell
737 types including hematopoietic progenitors and monocytes (bottom).

738

Extended Fig. 7 Density of protein-coding genes, TFs, trans-mSNPs and all SNPs on each chromosome.

The density (count/Mbp) of all protein-coding genes (PCGs, grey bar), TFs (orange bar), TFs after adjusting for PCGs density (red bar), the trans-mSNPs (blue bar) and all SNPs (black bar) on each chromosome (x-axis).

Extended Fig. 8 10 trans-mQTL hotspots are enriched with TF motifs.

The plot showing the TF motif enrichment results of 10 trans-mQTL hotspots using TFmotifView ⁵¹, and the ten points represent the ten hotspots in which the trans-mCpGs related with the index SNPs significantly ($P < 0.05/16 = 3.13 \times 10^{-3}$) enriched for the binding motifs of corresponding nearby-TFs. The text next to each point describes the index SNP of the hotspot and the TF with its motif matrix ID. The significance of the enrichment is denoted by the size of the points. The x-axis shows the percentage of the index SNP associated trans-mCpGs overlapped with the TF motifs, and the y-axis shows the fold enrichment (log2 transform) for the TF motifs compared with the background CpGs (see Methods). We also use different red color depths to indicate the mQTL effect consistency of the index SNPs, i.e., the proportion of the trans-mCpGs with the same effect direction.

Extended Fig. 9 Enrichment of trans-mQTLs in TFBSs and disease/traits.

a-b, Enrichment of trans-mQTLs in TFs and diseases.

Extended Fig. 10 rs4666078 G allele negatively associated with all of the mCpGs while most mCpGs positively associated with tissue eosinophilia.

This plot showing the 46 trans-mCpGs that are also the CpGs (DMPs) associated with tissue eosinophilia reported in 147 Korean individuals⁶². The x-axis shows the $-\log_{10}(FDR) \times \text{effect sign of DMPs}$ associated with tissue eosinophilia which indicates the significance and effect direction of these CpGs to tissue eosinophilia. And the y-axis shows the $-\log_{10}(P) \times \text{effect sign of rs4666078 G allele}$ associated with these CpGs, indicating the significance and effect direction of rs4666078 to these CpGs. We mark the points with different colors to show if the effect of the two associations are in the same directions (red) or opposite (blue).

Extended Fig. 11 Speculated mechanisms of cis- and trans-mQTLs.

a, speculated mechanisms of cis-mSNPs affect cis-mCpGs: cis-mSNPs may affect binding of transcriptional factor (TF) or DNA binding protein (DBP) in local region and the corresponding TF/DBP would inhibit or recruit DNA methyltransferase (DNMT) or Ten-Eleven Translocation proteins (TET) (①②③), or cis-mSNPs may directly affect binding of DNMTs (④), methyl-binding protein domain (MBD, ⑤) or TETs (⑥); **b**, speculated mechanisms of trans-mSNPs affect multiple trans-mCpGs: trans-mSNPs may affect the translation of TF/DBP in the local region through chromatin changes, TF and mediator cooperation, and as the TF/DBP plays crucial functions in defining cell identity, they would affect methylation at multiple remote CpG sites through recruiting or competing with DNMTs or TETs (⑦⑧⑨).

Figure legends

Fig. 1 Identification and validation of mQTLs in East Asians.

a, Flowchart of mQTL study in Han Chinese. **b**, Comparing the efficiency of fastQTLmapping with MatrixEQTL²⁴. Every test is repeated five times to take an average. **c**, Atlas of the significant mQTLs in Han Chinese (NSPT), red, cis-; orange, lcis-; blue, trans-. Left, the distribution of the genome-wide significant mQTLs (x-axis: SNPs, y-axis: CpGs), point size indicates significance; Middle, zoom in chromosome 6; Right, an example of a single trans-mQTL, y-axis shows the methylation levels (beta values) of cg19268652 (body of *HIVEP1*) and x-axis shows the three genotypes of rs2233437 (intronic of *NFKBIE*). **d**, Comparing the scaled effect sizes of 2.14M mQTL pairs with an independent Han Chinese dataset (CGZ). **e**, Comparing the scaled effect sizes between Han Chinese (NSPT) and Europeans (FHS)² for 13.03M mQTL associations. **f**, Enrichment of mSNPs and mCpGs in different genomic regions. The y-axis indicates the fold enrichment compared with all tested SNPs or CpGs. The P-value of hypergeometric test is adjusted by Bonferroni method (P.bfadjust). **g**, Enrichment of mQTL pairs in functional elements (CTCF: CTCF-enriched elements, E: enhancers, P: promoters, PF: promoter flanking regions, RE: regulatory elements, TFBS: TF binding sites). Heatmap shows the fold changes (see **Methods**) of SNP-CpG pairs in all combinations of functional categories. **h**, Enrichment of mSNPs and mCpGs in phenotype-related sites from GWAS Catalog³⁰ and EWAS Atlas³¹ compared with permuted sites (1×10^4). The y-axis shows the percentage of GWAS or EWAS significant signals overlapped with mQTLs or permuted sites.

Fig. 2 Cell-type and cell-lineage specific mQTLs and validation.

a, Number of mQTLs found in each cell-type (top-panel: red=lymphoid cell-types, blue=myeloid cell-types) and cell-lineage (lower panel), by application of CellDMC. **b**, The estimated proportion of shared and specific mQTLs in the two main cell lineages; **c&d**, Validation of cell-type and cell-lineage specific mQTLs in the independent Chinese Han cohort (CGZ). Scatterplots display the t-statistics of association between genotype and DNAm in the discovery cohort (x-axis), vs the corresponding statistics in the validation cohort (y-axis). Odds Ratio and one-tailed Fisher's test P-value are given. **e**, Validation of cell-lineage specific mQTLs in BLUEPRINT (BP) data. Scatterplots display the signed statistical significance of myeloid or lymphoid mQTLs (x-axis) vs the corresponding values in the purified samples from BP (y-axis). P-value of a linear regression is given. Barplots compare the overlap of myeloid/lymphoid mQTLs with those identified in the BP data (BP summary statistics only included index-SNPs and not other SNPs in LD), against the expected number under the null. Binomial P-value is given. Neut = neutrophil, Mono = monocytes,

809 CD4T = CD4⁺ T-cells. **f**, Violin plots of effect-size for cell-lineage specific mQTLs. P-value is from a Wilcoxon rank
810 sum test. **g**, eFORGE analysis of myeloid and lymphoid mCpGs. Significance levels of enrichment (y-axis) of
811 myeloid and lymphoid mCpGs for DHSs as derived in lymphoid and myeloid cell-types (x-axis) from BLUEPRINT,
812 ENCODE, Consolidated Roadmap and original Roadmap. Color of cell-type indicates its source.

813 **Fig. 3 Mechanistic interpretation of cis- and lcis-mQTLs.**

814 **a**, The schema of mQTL pairs located within/across/outside a TAD (above); Comparison of the number of SNP-CpG
815 pairs within/across/outside TADs for mQTLs and two control groups (below). **b**, The schema of mQTL pairs validated
816 by a chromatin loop (above); Comparison of the number of SNP-CpG pairs validated by PChIP or HiChIP loops
817 (below). **c**, The schema of mQTL pairs being co-opening or co-active; Comparison of the number of co-opening/co-
818 active SNP-CpG pairs (below). **d**, Schematic overviews of the M1 regulatory process. **e**, Schematic overviews of the
819 M2 regulatory process. **f**, Enrichment of mQTLs in OpenCausal variants of blood tissue derived from an external
820 database (Fisher's test: P-value = 1.6×10^{-16} , 2.02-fold); **g**, Proportion of mQTLs interpreted by the two mechanisms
821 M1 and M2; **h**, The proportion of potential opening-causal mQTLs overlapped/specific in the two main cell lineages,
822 myeloid and lymphoid.

823 **Fig. 4 trans-mQTL hotspots mediated by TFs and super enhancers.**

824 **a**, Relation between trans-mQTL SNPs and near-TFs (± 1 Mbp) SNPs. y-axis shows the proportion of trans-mSNPs
825 in the SNPs sets. P-value and odds ratio (OR) were from Fisher's exact test. **b**, The relation between density
826 (count/Mbp) of trans-mSNPs (x-axis) and TFs after correcting PCGs density (y-axis). Chromosome 19 is indicated
827 by orange dot, and other chromosomes are indicated by blue dots. **c**, Relation among the number of trans-mQTL
828 hotspot, proportion of mSNPs near TF and the threshold of defining hotspots. With threshold increases, the number
829 of trans-mSNP loci above the threshold goes down (red line, right y-axis), but the proportion of mSNPs near-TFs
830 goes up (blue line, left y-axis). **d**, The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right
831 y-axis) for each trans-mSNP on chr19. P-values and correlation coefficients were from Spearman's rank correlation
832 test. **e**, The distribution of the associated unlinked trans-mCpGs count (y-axis) for each trans-mSNP identified in
833 NSPT. The 16 trans-mQTL hotspots are labeled by "H1" to "H16". "*" indicates the hotspots are also reported in the
834 FHS cohort². **f**, Annotation of cis-eQTLs and super enhancer according to each TF/DBP in 16 trans-mQTL hotspot
835 regions; trans-mSNPs in each hotspot are located in the same TAD regions, enriched with HiChIP interactions and
836 significant OpenCausal signals.

Fig. 5 The *FOSL2* trans-mQTL hotspot (H2) associated with eosinophil counts.

a, Circos plot displaying the enrichment of the 232 rs4666078 trans-mCpGs in *FOSL2* ChIP-seq binding sites (data from ChIP-Atlas⁵⁹). The inner link lines show the 232 trans-mCpGs, the ones overlapped with *FOSL2* binding sites denoted by orange. The inner barplots show $-\log_{10}(P)$ of trans-mQTL associations. The outer blue dots represent hypomethylation with the G allele. The outer green barplots show $-\log_{10}(Q)$ of peak-caller MACS2 indicating ChIP-seq signals score. **b**, Barplot showing the proportion of the 232 mQTL pairs that existing in myeloid or lymphoid lineage (cell-specific mQTL results filtered by $P\text{-value} < 0.05/232 = 2.16 \times 10^{-4}$). **c**, Circos plot displaying the enrichment of the 141 rs4666078 myeloid-lineage trans-mCpGs in *FOSL2* ChIP-seq binding sites. **d**, Enrichment of the 232 mCpGs for EFO parent categories based on the curated EWAS results³¹ compared with all rs4666078 trans-wide CpGs. P-values of hypergeometric test are adjusted by Bonferroni method. **e**, Enrichment of rs4666078 trans-mCpGs annotated genes for human diseases and traits based on DisGeNET database⁶⁰ compared with all genes in human genome. **f**, Upper panel showing the enrichment of these trans-mCpGs in tissue eosinophilia associated CpGs reported in 147 Korean individuals⁶² (Hypergeometric test). Lower panel showing the association significance with tissue eosinophilia (y-axis) of the 232 mCpGs (red) compared with all tested CpGs (blue). Kolmogorov-Smirnov test is used to evaluate the significance. **g**, Two-Sample MR results showing 24 CpGs are causal for eosinophil counts in whole blood ($FDR < 5\%$). The effect (x-axis), the significance (y-axis) of MR, and CpG-annotated genes are shown.

Fig. 6 The *NFKB1* trans-mQTL hotspot associated with BMI.

a-b, Circos plots displaying the enrichment of *NFKB1* binding motifs (left) and *NFKB1* ChIP-Seq binding sites (right) among trans-mQTLs associated with the SNP linked in-cis with *NFKB1*. The fraction of trans-mQTLs enriched for the binding motif/site is given below circos plot. The inner track labels the enriched trans-mQTLs (red). Red barplots display the corresponding $-\log_{10}(P)$ of enrichment. Blue barplots display $-\log_{10}(P)$ of trans-mQTLs. Outer track represents the directionality of DNAm change of the trans-mQTLs: cyan (hypermethylation), magenta (hypomethylation). **c**, scatterplots displaying the signed statistical significance of 364 BMI-associated CpGs reported in the literature (y-axis) against their corresponding values in NSPT (x-axis). **d**, Association of *NFKB1* trans-mQTLs with BMI in 2 independent cohorts (NSPT and Buccal-IC). Plots display the statistical significance (y-axis) vs. the CpG index position (x-axis) for the *NFKB1* trans-mQTLs (red line) vs that of randomly selected CpGs (blue). P-value is from a Binomial test. **e**, Scatterplots displaying the t-statistics of association of DNAm of *NFKB1* trans-

864 mQTLs with risk allele (y-axis) vs. their corresponding t-statistics of association of DNAm with BMI in the 2
865 independent cohorts (x-axis). **f**, MR for the 35 *NFKB1* trans-mQTLs associated with BMI. x-axis labels statistical
866 significance (adjusted by FDR). Major color indicates direction of effect, with left panel displaying results for model
867 where DNAm mediates the effect of SNP on obesity, right panel is for model where obesity affects DNAm. **g**, As f,
868 but for BMI-associated CpGs reported by BMI EWAS studies. **h**, 3d scatterplot displaying the interaction between
869 BMI and genotype on BMI-associated *NFKB1* trans-mQTLs. x-axis labels the significance level of the univariate
870 association between DNAm and genotype, y-axis labels the significance level of the univariate association between
871 DNAm and BMI, z-axis labels the significance level of the interaction term between genotype and BMI.

Supplementary note

I. Methods	3
1. Sample and recruitment details	3
2. DNAm assessment of the discovery panel (NSPT)	3
3. SNP genotyping and imputation of the discovery panel (NSPT)	4
4. DNAm assessment of the validation panel (CGZ)	5
5. SNP genotyping of the validation panel (CGZ)	5
6. mQTL mapping analysis	6
i) fastQTLmapping	6
ii) MatrixEQTL	6
iii) mQTL replication in another Han Chinese sample (CGZ)	7
iv) Explore mQTL consistence with those reported in EUR	8
v) Estimating the heritability of DNAm	8
7. Functional annotation and enrichment of mQTL SNPs and CpGs	9
i) Genomic enrichment	9
ii) GWAS and EWAS signal enrichment	10
iii) GO and KEGG pathway enrichment	10
iv) cis-mQTL enrichment in cis-eQTL (GTEx)	11
v) Functional analysis for the overlaps between mQTLs and eQTLs	11
8. Cell-lineage specific mQTL	11
i) Cell-lineage specific mQTL mapping using CellDMC	11
ii) Validation of cell-lineage specific mQTL in BLUEPRINT data	13
iii) eFORGE analysis of myeloid and lymphoid mQTLs	13
9. Functional enrichment of mQTLs	14
i) Enrichment of mQTLs in OpenCausal	14
ii) Enrichment of mQTLs in 3D chromatin contacts	14
iii) Comparison of distance distributions of mQTLs and that of 3D loops	15
iv) Calculation of co-opening and co-activity of mQTL pairs	15
v) Evaluating the sensitivity of mQTLs to chromatin accessibility	16
vi) trans-mQTL SNPs linked with SNPs near TFs	16
vii) Protein-coding genes and TFs density	16

32	viii) motif and ChIP-seq enrichment of trans-mCpGs.....	17
33	10. Functional enrichment of trans-mQTL hotspots	18
34	i) Identifying whole-genome trans-mQTL hotspots	18
35	ii) TF binding site enrichment for trans-mCpGs in each trans-mQTL hotspot	18
36	iii) Super enhancer enrichment in trans-mQTL hotspots	19
37	iv) OpenCausal enrichment in trans-mQTL hotspots	19
38	v) TAD, HChIP enrichment in trans-mQTL hotspots	20
39	11. Downstream effect of trans-mQTLs on diseases/traits	20
40	i) trans-mCpG annotated genes enrichment for diseases/traits	20
41	ii) trans-mCpGs enrichment in nasal tissue eosinophilia	20
42	iii) Two-Sample Mendelian randomization test to identify potential causality between CpGs and	
43	eosinophil counts	21
44	iv) BMI EWAS	21
45	v) Mendelian randomization test to identify potential causality between CpGs and obesity	22
46	vi) BMI and SNP interaction on CpGs	23
47	II. Supplementary Figures	24
48	III. Supplementary Tables	57
49	IV. Supplementary References	65
50		
51		

I. Methods

1. Sample and recruitment details

Samples in the discovery dataset included 3,523 Han Chinese individuals who were recruited in three regional districts of China: Zhengzhou, Taizhou and Nanning (NSPT, 1,310 males and 2,213 females, aged from 18 to 83 years old, mean $\pm$ SD = 50.21 $\pm$ 12.75). This research is a sub project of The National Science & Technology Basic Research Project which was approved by the Ethics Committee of Human Genetic Resources of School of Life Sciences, Fudan University, Shanghai (14117). All the participants had provided written informed consents.

Samples in the validation dataset included 798 Han Chinese individuals who were recruited in the two clinical trials of chiglitazar (CGZ), a novel pan-PPAR agonist developed by Shenzhen Chipscreen Biosciences Co., Ltd. (Shenzhen, China) for treating type 2 diabetes¹⁻³. The CGZ cohort included 492 males and 306 females, aged from 24 to 70 years old, mean $\pm$ SD = 51.0 $\pm$ 9.7). Written informed consent was obtained from each patient. The ethics committee at each participating clinical center approved the study, which was conducted in accordance with the principles of Good Clinical Practice, the Declaration of Helsinki, and the relevant Chinese regulations.

2. DNAm assessment of the discovery panel (NSPT)

Blood samples were kept in Fudan University Taizhou Institute of Health Sciences for storage at -80C until DNA extraction. DNA extraction was performed using a TGuide M48 Automated nucleic acid extractor. Genome-wide DNA methylation was profiled using the Infinium MethylationEPIC BeadChips (Illumina). Five hundred nanogram of genomic DNA from each whole blood sample was bisulfite converted using the EZ DNA Methylation Kit (Zymo Research). BeadChips were processed following the manufacturer guide and protocol for Infinium MethylationEPIC array. DNA was hybridized to BeadChips and single base extension were performed using a Freedom EVO robot (Tecan). BeadChips were subsequently imaged using the iScan Microarray Scanner (Illumina). Illumina .idat files were then processed with the *minfi* Bioconductor package⁴ without background

77 correction (although background correction reduces bias it does so at the expense of increased variance,
78 which is generally something to be avoided, unless the DNAm data are used for copy-number
79 estimation). Probes with SNPs were removed using the *dropLociWithSnps* function from *minfi*.
80 This function uses the SNP information provided by Illumina and UCSC Common SNP tables
81 (including version 132, 135, 137, 138, 141, 142, 144, 146, and 147) with preset MAF (0 is the default
82 value and was used here) to filter SNP CpGs. We further removed probes on chromosomes X and Y.
83 We further used the Illumina definition of beta-values and derived P-values of detection for the rest of
84 probes by comparing the total intensity U+M to that of the background distribution (given by negative
85 control probes), as implemented in *minfi*. Beta-values with P-values of detection greater than 0.01 were
86 set to NA. Of note, the threshold of detection ($P < 0.01$) is more stringent than the $P < 0.05$ threshold
87 used in the other cohorts, partly because sample coverages were very high, allowing for a more
88 stringent threshold whilst also retaining a high coverage over probes. Only probes with less than 5%
89 missing values were retained. The missing beta values were then imputed with the *impute.knn* function
90 (using $k = 5$) in R. Type-2 probe bias was corrected using BMIQ⁵. All this resulted in a 811902 probe
91 times 3,523 sample data matrix. Based on principal component analyses, we found a significant
92 slide/beadchip effect. Therefore we used ComBat⁶ on M-values (logit of beta-values) to correct for the
93 slide effect and then M-values were used in mQTL mapping.

94 **3. SNP genotyping and imputation of the discovery panel (NSPT)**

95 Genome-wide SNP genotyping was profiled using Illumina Infinium Global Screening Array which
96 analyzes over 710,000 SNPs. It is a fully custom array designed by WeGene
97 (<https://www.wegene.com/>). Samples with call rate $< 98\%$, ambiguous sex and duplicates were
98 excluded. No sample failed heterozygosity and inbreeding test. 3,523 samples were kept in discovery
99 NSPT dataset. SNPs were excluded if they had call rate $< 98\%$, MAF $< 1\%$ and P-value of violations
100 from Hardy-Weinberg equilibrium (P_{HWE}) < 0.001 . After quality control (QC), 433,485 SNPs of 3,523
101 samples is left for further analyses. Imputation was done by SHAPEIT2 and IMPUTE2 using 1000

102 genome phase 3 as reference. Imputed variants were filtered with MAF > 0.01, imputation quality
103 score > 0.8 and violations from Hardy-Weinberg (HW) equilibrium ($p < 1 \times 10^{-5}$), leaving 8,615,463
104 variants. Then, we excluded SNPs that had multiple alleles, belonged to insertion or deletion. After
105 that, there were 7,576,990 SNPs (tested SNPs) left for mQTL mapping.

106 **4. DNAm assessment of the validation panel (CGZ)**

107 Blood samples of the CGZ cohort were kept in Shenzhen Chipscreen Biosciences Co. Ltd. until DNA
108 extraction. DNA extraction was performed using QIAamp DNA Blood Mini Kit (Qiagen). Genome-
109 wide DNA methylation was profiled using the Infinium MethylationEPIC BeadChips (Illumina). Five
110 hundred nanogram of genomic DNA from each whole blood sample was bisulfite converted using the
111 EZ DNA Methylation Kit (Zymo Research). BeadChips were processed following the manufacturer
112 guide and protocol for Infinium MethylationEPIC array. DNA was hybridized to BeadChips and single
113 base extension were performed using a Freedom EVO robot (Tecan). BeadChips were subsequently
114 imaged using the iScan Microarray Scanner (Illumina). Illumina .idat files were then processed with
115 the *CHAMP* Bioconductor package⁷, which resulted in a 717,100 probe times 798 sample data matrix.
116 Beta values were transformed into M values and M values were used in mQTL mapping.

117 **5. SNP genotyping of the validation panel (CGZ)**

118 CGZ samples and QC samples were genotyped with Affymetrix Axiom PMR Arrays under instructions
119 from the manufacturer. The raw genotyping data (CEL files) were generated with the GeneTitan
120 workflow, and subsequently went through the Best Practices Workflow embedded in Axiom Analysis
121 Suite for genotype calling. Each sample was genotyped with 902,560 probes. All samples passed a
122 default Dish QC threshold of 0.82 and a QC call rate threshold of 0.97, and had a call rate over 0.98.
123 After the successful completion of the Best Practices Workflow, genotyping calls in VCF and numeric
124 call codes were exported, as well as the QC tables for samples and probes. After that, 874,438 SNPs
125 were left for mQTL mapping.

6. mQTL mapping analysis

We adapted a two-step analysis strategy for mQTL mapping. Step 1, we adjusted methylation M values for bisulfite slide number, batch, age, sex, predicted blood cell fractions⁸, top ten genomic principal components (PC) and top two DNAm PCs by linear model regression. Then we tested associations between methylation residuals and SNP dosages using fastQTLmapping⁹ and retained significant SNP-CpG pairs at a loose threshold ($P\text{-value} < 1 \times 10^{-10}$). Step 2, we excluded the outliers of methylation M values (out the range of $\text{mean} \pm 3\text{SD}$) and the genotypes which presented in less than 5 samples, then we repeated analyses in step 1. We applied Bonferroni correction to maintain an experiment-wide type I error rate of 0.05 for cis-, lcis- and trans-pairs respectively, i.e., $P_{\text{cis}} < 0.05/4.7 \times 10^9 = 1.06 \times 10^{-11}$, $P_{\text{lcis}} < 0.05/1.75 \times 10^{10} = 2.86 \times 10^{-12}$ and $P_{\text{trans}} = 0.05/6.13 \times 10^{12} = 8.16 \times 10^{-15}$.

i) fastQTLmapping

FastQTLmapping is a linear regression solver in C++ for exhaustive regression analysis between two extraordinary large matrices allowing for covariates, which is particularly useful for mQTL-like studies⁹. The input contains three text files of floating-point matrices for independent variables X , dependent variables Y , and covariates C . Note that rows represent variables and columns represent observations for faster loading. FastQTLmapping also supports Plink binary file format. The output file records the test statistics of all regressions, optionable under a preset significance threshold. FastQTLmapping is developed on Linux using MKL (<https://software.intel.com/tools/onemkl>) and GSL (<http://www.gnu.org/software/gsl/>) library, and is run from the command line. C++ source code, an example run, and documentation of fastQTLmapping are freely available at <https://github.com/TianTTL/xQTLmapping>.

ii) MatrixEQTL

We also applied MatrixEQTL¹⁰ (MatrixEQTL R package v2.3) in the mQTL mapping study. MatrixEQTL is widely-used QTL mapping tool which processed large data with large matrix operations with no loss of precision. Linear additive models were calculated, adjusted for age, sex,

151 batch, bisulfite slide number, array position, top ten genomic PCs and predicted blood cell fractions
152 (B cells, CD4⁺ and CD8⁺ T cells, NK cells, monocytes and neutrophils). SNPs within 1 Mb of the
153 CpGs were considered as cis-mQTLs. SNPs located further than 1Mb of the CpGs or in other
154 chromosomes were considered as trans-mQTLs. The P-value threshold for defining mQTLs was
155 1.00×10^{-8} .

156 Both fastQTLmapping and MatrixEQTL produced the exact results under all investigated settings. The
157 operating time of both fastQTLmapping and MatrixEQTL was largely linear to the computation size.
158 In the single-thread setting, fastQTLmapping completed the analyses of sizes 10^9 , 10^{10} , and 10^{11} in 24,
159 210, and 2,058 seconds, respectively. MatrixEQTL completed these analyses in 103, 714, and 6,733
160 seconds, respectively, meaning that fastQTLmapping was 3 - 4 times faster than MatrixEQTL. In the
161 32 CPU threads setting, fastQTLmapping completed these analyses of sizes 10^9 , 10^{10} , and 10^{11} in 3,
162 14, and 100 seconds, respectively. MatrixEQTL completed these analyses in 33, 162, and 682 seconds,
163 respectively, meaning that fastQTLmapping was 6-11 times faster than MatrixEQTL (Fig. 1). In
164 addition, the peak memory usage of fastQTLmapping under the computation size of 10^{11} and 32-thread
165 (17.6 GB) was much small than that of MatrixEQTL (105.4GB). Applying fastQTLmapping in the
166 analysis of our data (8.6M SNPs $\times$ 811K CpGs $\times$ 3,523 samples) consumed 6.79 hours and 321GB
167 memory under 32-threads parallel.

168 **iii) mQTL replication in another Han Chinese sample (CGZ)**

169 To validate our results, we conducted mQTL analysis using another independent Han Chinese cohort
170 (CGZ). First, we matched our mQTL SNPs and CpGs in CGZ according to genomic positions. Then,
171 we excluded SNPs with palindromic alleles, i.e., A/T, C/G. For the left SNPs, we harmonized (switch
172 and flip) the alleles if needed. After preprocessing, there were 2,463,146 (3.91%) mQTLs available in
173 CGZ. Then, we performed mQTL mapping by MatrixEQTL in a linear model adjusted for adjusted for
174 age, sex, batch, bisulfite slide number, top ten genomic PCs and predicted blood cell fractions (B cells,
175 CD4⁺ and CD8⁺ T cells, NK cells, monocytes and neutrophils). The P-values were adjusted for

multiple-testing using the Benjamini-Hochberg FDR method and 2.14M (87.05%) were successfully replicated (FDR < 0.05). For these 2.14M mQTLs, we compare the mQTL effect sizes between the two cohorts using β_{ijc}^s which defined as the effect sizes divided by the maximum of absolute effect values in each cohort.

$$\beta_{ijc}^s = \beta_{ijc} / \max_{i,j}(|\beta_{ijc}|) \quad (1)$$

where, β_{ijc} is the effect size of mQTL SNP i on CpG j in cohort c .

Based on β_{ijc}^s , we assessed the similarity of effect patterns and directions of mQTLs among the two cohorts. Pattern's similarity was estimated by Spearman's rank correlation.

iv) Explore mQTL consistence with those reported in EUR

To explore the consistence of mQTLs between populations, we downloaded the summary statistics of genome-wide significant mQTLs reported in the Framingham Heart Study (FHS, N=4170)¹¹. First, we matched the FHS mQTL SNPs and CpGs in our discovery dataset (NSPT) according to genomic positions. Then, we excluded SNPs with palindromic alleles, i.e., A/T, C/G. For the left SNPs, we harmonized (switch and flip) the alleles if needed. After preprocessing, there were 16.07M FHS mQTLs that available in NSPT. Then, we extracted these mQTLs in our discovery results and the P-values were adjusted for multiple-testing using the Benjamini-Hochberg FDR method. Only the mQTLs passed FDR < 0.05 were kept. We calculated the scaled effect sizes as shown in equation (1) in each cohort, and assessed the similarity of effect patterns and directions of mQTLs among two cohorts. Pattern's similarity was estimated by Spearman's rank correlation.

v) Estimating the heritability of DNAm

Narrow-sense heritability of each CpG (h_{CpG}^2) was estimated by a self-designed pipeline. We first adjust DNAm M values for age, sex, batch, bisulfite slide number, top ten genomic PCs and predicted

198 blood cell fractions (B cells, CD4⁺ and CD8⁺ T cells, NK cells, monocytes and neutrophils). In our
199 algorithms, we used the same hybrid estimation scheme in GCTA¹² and utilize the factored spectral
200 transformation (FaST)¹³ to accelerate the calculation. By parameterization, we also transformed the
201 constrained optimization problem into an unconstrained. The run time and memory footprint were
202 linear in the cohort size, making it far efficient than GCTA.

203

204 **7. Functional annotation and enrichment of mQTL SNPs and CpGs**

205 **i) Genomic enrichment**

206 The tested SNPs were annotated using ANNOVAR¹⁴ with its pre-built RefSeq Gene annotation files
207 on hg19 build. The annotations included: upstream (within 1500bp region upstream of transcription
208 start site, TSS), UTR5 (within a 5' untranslated region), exonic (within a coding region), splicing
209 (within 2-bp of a splicing junction), intronic (overlapping an intron), ncRNA (non-coding RNA items)
210 and UTR3 (within a 3' untranslated region).

211 The CpGs were annotated according to manufacturer's manifest files¹⁵ (hg19 build). The genomic
212 annotations included: Enhancer (in FANTOM5¹⁶ project defined enhancer regions), TSS1500 (200–
213 1500 bases upstream of the TSS), TSS200 (0–200 bases upstream of the transcriptional start site),
214 UTR5 (within the 5' untranslated regions), 1stExon (the first exon), ExonBnd (within 20 bases of an
215 exon boundary, i.e. the start or end of an exon), Body (gene body) and UTR3 (within a 3' untranslated
216 region). The CpG island annotations included: N_Shelf (upstream 2–4 kb from CpG islands), N_Shore
217 (upstream 0–2 kb from CpG islands), S_Shore (downstream 0–2 kb from CpG islands) and S_Shelf
218 (downstream 2–4 kb from CpG islands). Hypergeometric tests were used to evaluate if the identified
219 cis-, lcis- and trans-mQTL SNPs and CpGs showed enrichment for these genomic functional features.
220 And the P-values were adjusted by Bonferroni correction method.

221 We also collected functional genomic regions from the Functional Annotation of ANimal Genomes
222 (FAANG) Project for neutrophil cells¹⁷, totally yielding 598,398 functional regions of CTCF binding

223 sites, enhancers, open chromatin regions, promoters, promoter flanking regions, and TF binding sites.
224 We examined the functional classes to which the SNPs and CpGs of mQTL pairs belonged and counted
225 the number of annotated mQTL pairs. Hypergeometric tests were used to evaluate if the identified cis-,
226 lcis- and trans-mQTL SNPs and CpGs showed enrichment for these genomic functional classes. And
227 the P-values were adjusted by Bonferroni correction method. Then, we generated random SNP-CpG
228 pairs at the same sample size as mQTLs and annotated those random pairs to functional regions. For
229 each combination of two functional classes, the fold change is calculated as the ratio between the
230 number of annotated mQTL pairs and that of annotated random SNP-CpG pairs.

231 **ii) GWAS and EWAS signal enrichment**

232 We downloaded the curated significant GWAS results from GWAS Catalog¹⁸ ($P < 1 \times 10^{-5}$, date:
233 2020.12.25) including 227,262 records and 137,450 sites, and significant EWAS results from EWAS
234 Atlas¹⁹ ($P < 10^{-4}$ or $Q < 0.05$, date: 2020.12.25) containing 416,331 records and 226,959 sites. We
235 applied permutation test to examine if mQTL SNPs were enriched for the known GWAS signals. First,
236 we randomly sampled sites from all tested SNPs to keep the same number and similar MAF distribution
237 (bin = 0.05) with the mQTL SNPs, which was replicated 10^4 times. Then based on the 10^4 sampled
238 sets, we calculated the coverage ratio of GWAS signals for each set and got the null distribution. Finally,
239 we observed the percentile of the actual ratio on the null distribution and took it as the permutation P-
240 value. The same test was also applied in the enrichment of EWAS signals on mCpGs.

241 **iii) GO and KEGG pathway enrichment**

242 We tested the mQTLs and mCpGs enrichment for different Experimental Factor Ontology (EFO)
243 parent categories using one-tailed hypergeometric test with all tested sites as background. The Gene
244 Ontology (GO) and KEGG pathway enrichment of mCpGs were conducted by R package
245 missMethyl²⁰ which could deal with two different bias: 1) the different number of probes per gene on
246 the array, and 2) CpGs that annotated to multiple genes.

iv) cis-mQTL enrichment in cis-eQTL (GTEx)

We downloaded the significant cis-eQTL data in whole blood ($N = 670$) from GTEx V8[ref²¹], then we harmonized the alleles (switch and flip if necessary) to match SNPs between the cis-eQTLs and our tested SNPs. We performed one-tailed hypergeometric test to evaluate the enrichment of cis-mQTL SNPs for cis-eQTL SNPs.

v) Functional analysis for the overlaps between mQTLs and eQTLs

We collected 3804 house-keeping (HK) genes from Eisenberg et al (2013)²² and checked the proportion of those HK genes in the overlapped SNPs between mQTLs and eQTLs, denoted as A. Then, we calculated the proportion of HK genes in all genes derived from Ensembl release v104[ref²³], denoted as B. The fold change is then defined as the ratio of A and B. In addition, we downloaded GO terms, KEGG pathways, and Reactome pathways from the Molecular Signatures Database²⁴. We checked the enrichment of the overlapped mQTLs and eQTLs in the functional terms/pathways using hypergeometric tests.

8. Cell-lineage specific mQTL

i) Cell-lineage specific mQTL mapping using CellDMC

Due to the heavy computational burden, we restricted this analysis to mQTLs passing a significance threshold of 10^{-8} . The inference of cell-lineage specific mQTLs was implemented via the CellDMC⁸ algorithm (EpiDISH R package v2.8). CellDMC identifies interactions between phenotype and cell-lineage fraction, thus allowing for the detection of cell-lineage specific differentially methylated cytosines (DMCTs). In this application, the phenotype is the genotype of the mQTL, and the DMCTs inferred by CellDMC would be a measurement of the cell-lineage specific nature of the mQTL. On the NSPT cohort, we ran CellDMC with confounders as additional covariates, which included age, sex, batch and array position. Blood cell fractions (B cells, $CD4^{+}$ and $CD8^{+}$ T cells, NK cells, monocytes and neutrophils) were estimated using EpiDISH^{25,26}. Prior to running CellDMC, DNAm data were

272 normalized for slide number using ComBat. Reason for not including slide number within CellDMC
273 itself is that there are only 8 samples per slide, adjustment of which thus benefits from a Bayesian
274 shrinkage approach as implemented in Combat. CellDMC was run at 2 resolutions: 6 cell-type
275 resolution (B cells, CD4⁺ and CD8⁺ T cells, NK cells, monocytes, neutrophils) and 2 cell-lineage
276 resolution (lymphoid and myeloid lineage). When running CellDMC at the resolution of 2 lineages,
277 we summed the estimated cell fractions of B cells, CD4⁺ T cells, CD8⁺ T cells and NK cells to give a
278 total lymphocyte fraction, whereas for the myeloid lineage we summed the fractions of monocytes,
279 neutrophils and eosinophils. The P-values inferred from CellDMC were adjusted for multiple-testing
280 using the Benjamini-Hochberg FDR method. Using FDR < 0.05, there were 6,865,414 B cells specific
281 mQTLs, 10,378,286 CD4⁺ T cells specific mQTLs, 12,374,232 CD8⁺ T cells specific mQTLs,
282 14,279,365 NK cells specific mQTLs, 7,178,916 monocytes specific mQTLs, 26,390,493 neutrophils
283 specific mQTLs, 22,744,847 lymphocytes specific mQTLs and 63,964,415 myeloid cells specific
284 mQTLs.

285 To estimate the fraction of shared mQTLs between lymphoid and myeloid lineages, we first calculated
286 the overlap between lymphoid and myeloid mQTLs, expressing this as a fraction of the total number
287 of mQTLs for which CellDMC was run. This overlap fraction was 0.16 (16%). For this analysis, and
288 because CellDMC was only run for significant mQTLs, we relaxed the significance threshold for
289 calling lineage-specific mQTLs, using an unadjusted P-value threshold of 0.05. This is justified since
290 by definition an mQTL must be an mQTL in at least one cell-lineage. Next, the observed overlap
291 fraction is the product of the true fraction of shared mQTLs times the sensitivity to detect such a non-
292 specific mQTL in the lymphoid lineage, times the sensitivity to detect such a non-specific mQTL in
293 the myeloid lineage. To estimate these sensitivities, we used the simulation framework from You C et
294 al (2020)²⁷ with an average mQTL effect size of 1.1, where effect size is defined by the difference in
295 average DNAm between the two extreme genotypes divided by the pooled standard deviation. The
296 average effect size was estimated from the list of mQTLs (P-value < 10⁻⁸). This resulted in a 90%

297 sensitivity to detect a non-specific mQTL in the myeloid lineage, but only a 20% sensitivity to detect
298 it in the lymphoid lineage. The lower sensitivity in the lymphoid lineage is because the lymphocyte
299 fraction is much lower and less variable than the myeloid one in whole blood. From this, we then
300 estimated the true fraction of shared mQTLs to be $0.16/(0.2*0.9) = 0.89$, i.e. approximately 90% of
301 mQTLs are shared between myeloid and lymphoid lineages.

302

303 **ii) Validation of cell-lineage specific mQTL in BLUEPRINT data**

304 We validated the cell-lineage specific mQTLs inferred in previous step against the BLUEPRINT data²⁸,
305 which contains matched genotype and DNAm profiles for purified neutrophils, monocytes and CD4⁺
306 T cells from approximately 200 samples. Specifically, we first downloaded the BLUEPRINT mQTL
307 summary statistics for the three primary blood cell subtypes (CD4⁺ T cells, monocytes and neutrophils),
308 which we note only includes the summary statistics of significant mQTLs defined at a threshold of
309 6.3×10^{-5} and only for the corresponding index SNPs.

310 For the validation analysis, we focused on the subset of mQTLs derived at 2 cell-lineage resolution
311 and which displayed clear specificity (lymphoid lineage and myeloid lineage), defined as those with
312 P-values less than 10^{-8} in one cell lineage but larger than 0.9 in the other cell lineage. With this stringent
313 definition, there were 58,105 lymphocyte-specific mQTLs and 407,069 myeloid-specific mQTLs.
314 Then, we matched these mQTLs with the index mQTLs in CD4⁺ T cells, monocytes and neutrophils
315 as reported in BLUEPRINT, obtaining an overlap of 22 index mQTLs for CD4⁺ T cells, 221 for
316 monocytes and 278 for neutrophils. Statistical significance of the validation was assessed at the level
317 of (i) overlap number using a binomial test against a null distribution derived from the expected overlap
318 as obtained from the opposite lineage, as well as (ii) in terms of the consistency of directionality, as
319 determined from a linear regression of the corresponding signed $-\log_{10}(P\text{-values})$.

320 **iii) eFORGE analysis of myeloid and lymphoid mQTLs**

321 To test whether the cell-lineage specific mQTLs capture the tissue specific regulatory component at

CpG level, we conducted a functional enrichment analysis of the cell-lineage specific mQTLs derived at 2 cell-lineage resolution using eFORGE2^{29,30}. Given a set of CpGs, eFORGE2 assesses if these CpGs map to cell-type-specific DHS regions, which are naturally enriched for functional regulatory elements such as enhancers. eFORGE2 makes use of a large database of DHS signals as derived from the Epigenome Roadmap, BLUEPRINT and ENCODE. eFORGE2 analysis was performed for the highly specific mQTLs defined earlier and due to eFORGE2 server restrictions, we used as input the top 1000 lymphocyte and myeloid specific mQTLs.

9. Functional enrichment of mQTLs

i) Enrichment of mQTLs in OpenCausal

We downloaded the list of opening-causal SNPs in blood tissue from Li et al (2020)³¹, where the SNPs are predicted to be sensitive to the chromatin accessibility of the regulatory elements. We considered the total number of SNPs as N , the number of opening-causal SNPs as M , the number of mQTLs as R , and the overlap between mQTLs and those opening-causal SNPs as k . Then, the enrichment significance P-value is then calculated as the summation of the hypergeometric distribution, as

$$\text{P-value} = 1 - \sum_{i=0}^k \frac{\binom{M}{i} \binom{N-M}{R-i}}{\binom{N}{R}} \quad (2)$$

ii) Enrichment of mQTLs in 3D chromatin contacts

PCHi-C data of primary blood cells including total B cells (tB), naive B cells (nB), monocytes (Mon), neutrophil (Neut), total CD4⁺ T cells (tCD4), naive CD4⁺ T cells (nCD4), total CD8⁺ T cells (tCD8), naive CD8⁺ T cells (nCD8) are collected from Javierre et al (2016)³², where PCHi-C interactions were filtered with a threshold of CHiCAGO scores > 5 [ref³³]. The loops and TADs were directly downloaded from the website given by this study (<https://osf.io/u8tzip>). The H3K27ac HiChIP data for naive T cells were obtained from Mumbach et al (2017)³⁴. The JuiceBox output was processed by HiCCUPS implemented in the Juicer Tools (version 0.7.5) with default parameter settings to obtain the HiChIP

346 loops³⁵.

347 We counted a mQTL pair as validated if its SNP and CpG can respectively locate within the two
348 anchors of a 3D interacting loop derived from PCHi-C or HiChIP data. We built two groups of
349 randomly selected SNP-CpG pairs as controls to evaluate the enrichment of mQTLs in chromatin
350 contacts. The first control group was generated by randomly sampling SNP-CpG pairs from all SNP-
351 CpG combinations, regarded as genomic background. The other control group was sampled SNP-CpG
352 pairs with the same distance distribution as mQTL pairs, named distance-matched SNP-CpG group.
353 We compared the number of validated mQTL pairs with that of genomic background and distance-
354 matched SNP-CpG pairs. The enrichment significance and fold changes are calculated with the
355 hypergeometric test.

356 **iii) Comparison of distance distributions of mQTLs and that of 3D loops**

357 We checked the distance distribution of mQTL pairs and that of 3D chromatin loops and found that
358 our mQTLs tend to capture interactions whose SNP and CpG are quite close to each other: 39.26% of
359 mQTL interactions occur within sub-50-kb distances, typically a region of low sensitivity for 3C-based
360 techniques (2.9% Hi-C loops, 13.2% PCHi-C loops, 2.7% HiChIP loops). This suggests that many
361 functional interactions may be below the resolution of conventional 3C-based techniques, which also
362 partly explained why a proportion of mQTLs are not colocalized with chromatin loops.

363 **iv) Calculation of co-opening and co-activity of mQTL pairs**

364 The peaks of DNase-seq data for B cells and T cells and the H3K27ac ChIP-seq data of neutrophil
365 cells were all downloaded from ENCODE³⁶. We derived the chromatin opening degree of SNPs and
366 CpGs by mapping them to the DNase-seq peaks and assigned the opening score of the overlapped peak
367 to SNPs or CpGs. Similarly, the chromatin activity of SNPs and CpGs is derived from ChIP-seq data.
368 We define a SNP-CpG pair as co-opening/co-active if both the SNP and the CpG are overlapped with
369 the peaks. Then, we generated two control groups by sampling SNP-CpG pairs from genomic
370 background or with the same distance distribution as mQTL pairs. We compared the number of co-

opening/co-active mQTL pairs with that of genomic background and distance-matched SNP-CpG pairs.
The enrichment significance and fold changes are calculated with the hypergeometric test.

v) Evaluating the sensitivity of mQTLs to chromatin accessibility

We quantified the influence of mQTL SNPs to the regulatory elements where they locate, using a tool named OpenCausal³¹. OpenCausal first predicts the chromatin accessibility score of a given region using the TF expression and genomic sequence information as input, where the sequence information can be derived from the reference genome, or alternatively, it can be formed based on the SNP information. Then, OpenCausal uses the change of chromatin accessibility scores before and after SNP mutation to measure the influence of a variant on the regulatory element. In this study, we collected the processed fragments per kilobase million (FPKMs) of RNA-seq data for blood tissue from GTEx project. Using the TF expression and mQTL information as input, OpenCausal calculated an opening-causal score for each mQTL. We consider the mQTLs with non-zero scores as the ones that are sensitive to the chromatin accessibility.

vi) trans-mQTL SNPs linked with SNPs near TFs

The list of human transcription factors (TFs) was downloaded from Lambert et al (2018)³⁷, which included 1,639 manually examined TF genes. Based on these TFs and gene map information from GENCODE³⁸, we defined near-TF SNPs if the SNPs in our tested SNPs located in ± 1 Mbp regions of any TFs. Then, the SNPs were divided into two sets, near-TF SNPs and not near-TF SNPs. We calculated the proportion of trans-mQTLs in these two sets and applied one-tailed fisher's exact test to examine whether trans-mQTL SNPs were more likely to be SNPs near TFs.

vii) Protein-coding genes and TFs density

Gene list of human protein-coding genes (PCGs) was downloaded from GENCODE³⁸, which was based on hg19 (GRCh37). We calculated the number of PCGs and TFs per Mbp in each chromosome, defined as density of PCGs and TFs. The TFs density controlled by PCGs density (TFs/PCGs) was obtained through dividing the number of TFs by the number of PCGs. Trans-mQTLs on each

396 chromosome were also summarized. Then, the PCGs, TFs, TFs/PCGs and trans-mQTL densities on
397 each chromosome were generally compared.

398 **viii) motif and ChIP-seq enrichment of trans-mCpGs**

399 To assess if trans-mQTLs can be explained by differential transcription factor binding, we first filtered
400 trans-mQTLs where the SNP is associated in-trans with 100 or more CpGs. This was done to ensure
401 adequate power to detect an enrichment of TF-binding motifs or ChIP-Seq peaks. Next, SNPs driving
402 such trans-mQTLs were filtered further by the requirement that the SNP be a GWAS hit for a
403 disease/trait, as recorded by the GWAS catalogue database¹⁸ and PhenoScanner^{39,40}. We then filtered
404 SNPs further by the requirement that a transcription factor (TF) be in-cis (<1Mbp) with the SNP. The
405 TF list was downloaded from The Human Transcription Factors database³⁷ which contains 1,639
406 human TFs (and their motifs). Next, we examined if the trans-mQTL CpGs associated with a given
407 SNP were enriched for either the corresponding TF-binding motif or TF ChIP-Seq signal. ChIP-seq
408 data were downloaded from the ChIP-Seq atlas⁴¹. For each relevant TF, the ChIP-Seq atlas provides
409 the binding intensity score to potential target genes defined at 3 different distances from the TSS (± 1 kb,
410 ± 5 kb and ± 10 kb). Trans-mQTL CpGs were assigned corresponding binding intensity scores if they co-
411 localized to within ± 1 kb, ± 5 kb and ± 10 kb of the target gene's TSS. To obtain a P-value of enrichment
412 we performed a one-tailed Wilcoxon rank sum test comparing these binding intensity values to those
413 of 5,000 randomly selected CpGs. Separate to this, we also performed enrichment for TF-binding
414 motifs using Position Weight Matrices (PWMs) from MotifDB using the package PWMEnrich. In
415 more detail, we scanned 200bp sequences (± 100 bp from the trans-mQTL CpGs) and statistical
416 enrichment was assessed against background sequences randomly selected from a pre-compiled
417 threshold-free lognormal background sequence distribution of human promoters. From the TF ChIP-
418 Seq and binding-motif analysis we thus obtained a union list of trans-mQTLs with the GWAS trait
419 SNP linked in-cis to a TF and with the corresponding trans-mQTL CpGs enriched for TF-binding sites
420 (as determined either by ChIP-Seq or motif-enrichment) (**Fig. S33**). For traits from the GWAS catalog,

we obtained a total of 2,182 unique SNP-Trait-TF validated trans-mQTL networks. For traits from Phenoscanner, we got 125,133 unique SNP-Trait-TF validated trans-mQTL networks.

10. Functional enrichment of trans-mQTL hotspots

i) Identifying whole-genome trans-mQTL hotspots

The definition of trans-mQTLs hotspots was parallel to that of trans-eQTLs and trans-pQTLs^{42,43}, which were identified by an index mQTL and nearby SNPs in high LD ($R^2 > 0.8$) associated with a number of trans-mCpGs. We first clumped trans-mCpGs with 500Kbp window to exclude potential linkage among adjacent CpGs. Then we calculated the number of unlinked trans-mCpGs associated with each trans-mQTL. To detect trans-mQTL hotspots regions, we divided genome-wide trans-mQTL SNPs into 1Mbp chunks, and then we selected an index mQTL which was associated with the largest number of trans-mCpGs in each chunk. The region comprised of an index SNP and trans-mQTLs linked with this index SNP in each chunk was taken as potential trans-mQTL hotspot. Then trans-mQTL hotspots could be filtered according to the number of associated trans-mCpGs. In this study, to better learn the properties of trans-mQTL hotspots, we defined trans-mQTL hotspots as the region which comprised of an index SNP associated more than 97 unlinked trans-mCpGs (top 1% percentile of the number of associated trans-mCpGs on each trans-mQTL, **Fig. S6**) and trans-mQTLs linked with this index SNP in 2Mbp region.

ii) TF binding site enrichment for trans-mCpGs in each trans-mQTL hotspot

We used two motif enrichment tools (TFmotifView⁴⁴ and PWMEnrich⁴⁵) to evaluate if the trans-mCpGs associated with the index SNP in each hotspot were enriched for TF binding (TF nearby the index SNP within ± 1 Mbp, **Table 1**). The motifs of each TF were downloaded from JASPAR 2020 database⁴⁶ and we tested if the motif was enriched in regions (± 100 bp) of trans-mCpGs associated with the index SNP for each hotspot. The background CpGs for TFmotifView were randomly selected more than 10K sites from all tested CpGs to ensure a consistent distribution of methylation variation (SD)

446 with the interested CpGs. As for PWMEnrich, the background were all trans-wide CpGs according to
447 the index SNP (SNP-CpG distance > 5Mbp).

448 For hotspot H2, we also downloaded *FOSL2* ChIP-seq signals data from ChIP-Atlas database⁴¹ (Using
449 the signals with peak-caller MACS2 q-value < 1×10^{-5} . We didn't find the data from blood, so used all
450 the cell types instead), and performed one-tailed hypergeometric test on the trans-mCpGs nearby-
451 regions ($\pm 100\text{bp}$) compared with all trans-wide CpGs according to the index SNP rs4666078
452 (rs4666078-CpG distance > 5Mbp). The same strategy was used in the analysis of myeloid-specific
453 mQTLs.

454 **iii) Super enhancer enrichment in trans-mQTL hotspots**

455 We checked if the interested TFs or DNA binding proteins (DBP) located in trans-mQTL hotspots were
456 enriched with super enhancers as over two thirds of the interested TFs were enriched with binding sites
457 in remote mCpG regions. Super enhancers have crucial functions in defining cell identity⁴⁷⁻⁴⁹. The link
458 between trans-mQTL hotspots and super enhancer would disclose the biological importance of these
459 hotspots. Three super enhancer databases⁵⁰⁻⁵² were used in this study. We checked if there was super
460 enhancer for each interested TF/DBP in trans-mQTL hotspot in different tissues or cell lineages. We
461 applied hypergeometric distribution tests to evaluate the enrichment of super enhancers in trans-mQTL
462 hotspots.

463 **iv) OpenCausal enrichment in trans-mQTL hotspots**

464 For each trans-mQTL hotspots, there were trans-mSNPs with significant OpenCausal scores. We
465 examined if the trans-mSNPs in each trans-mQTL hotspot were enriched with significant OpenCausal
466 scores compared with randomly selected genomic regions in the same length. For each trans-mQTL
467 hotspot, we first generated 1,000 random numbers according to the chromosome length. Then we
468 generate a region of 2Mbp with each random number as the center. Next, we calculated the proportion
469 of SNPs with significant OpenCausal scores in these regions, and take their distribution as a null
470 distribution. At last, we checked the posterior probability of the real value on this distribution.

v) TAD, HChIP enrichment in trans-mQTL hotspots

The signatures of trans-mQTL hotspots include trans-mSNPs located in TAD, HChIP promoter-promoter and promoter-other interaction regions. To test if these signatures were enriched in trans-mQTL hotspots, we compared these signatures with randomly selected SNPs in the whole genome. We applied hypergeometric distribution tests to evaluate the enrichment of these signatures in trans-mQTL hotspots.

11. Downstream effect of trans-mQTLs on diseases/traits

i) trans-mCpG annotated genes enrichment for diseases/traits

The trans-mCpGs of the index SNP rs4666078 in H2 annotated to 151 genes. We tested if these genes were enriched in human diseases/traits or biological processes by using R package disgenet2r. disgenet2r is based on the DisGeNET knowledgebase⁵³, i.e., one of the largest publicly available collections of genes and variants associated to human diseases. We used all genes in the genome as the background, and applied Benjamini-Hochberg FDR test to correct P-values from Fisher's tests.

ii) trans-mCpGs enrichment in nasal tissue eosinophilia

We downloaded the summary statistics about different methylation probes (DMPs) associated with nasal tissue eosinophilia which was closely related to the blood eosinophil numbers⁵⁴ (as there was almost no research about the blood eosinophil) in a published study of Korean population (Kim, N=147)⁵⁵. The DMPs were defined as: 1) mean methylation level difference between case and control > 0.01 (1%), 2) ANOVA FDR < 0.05 and Tukey honest significant difference P-value < 0.05. Based on these results, we evaluate the enrichment of 232 trans-mCpGs associated with rs4666078 in the DMPs. We performed one-tailed hypergeometric test whilst taking all tested CpGs in our study as background. We also checked if these trans-mCpGs were significant enriched in tissue eosinophilia compared with background CpGs using one-tailed Kolmogorov-Smirnov test.

iii) Two-Sample Mendelian randomization test to identify potential causality between CpGs and eosinophil counts

We downloaded the GWAS-meta summary statistics of blood eosinophil counts from a published study in EAS (N=86,890)⁵⁶. We conducted Two-Sample Mendelian Randomization analysis (TwoSampleMR) with taking CpGs as exposures and blood eosinophil counts as outcome. Instrument SNPs are selected from the whole-genome mQTL (including cis-, lcis- and trans-) at a loose threshold (P-value $< 1 \times 10^{-10}$) for each trans-mCpG associated with rs4666078. The analysis was performed by using R package TwoSampleMR and referring to previous studies⁵⁷⁻⁵⁹. In detail, we compared the 232 CpGs whole-genome mQTL results and eosinophil counts GWAS summary data, and kept the results of intersect SNPs (harmonized alleles). Then, we clumped the CpGs results to remove the SNPs which were in high LD with others (LD was calculated from 1000 genomes EAS data. Clumping amongst those SNPs that had LD R^2 above the specified threshold 0.2 within 1000kb window, only the SNP with the lowest P-value were retained), to get independent instrumental variables. We excluded the CpGs which had less than three instruments (as the minimum number of variables required for multi-instrument MR analysis was three), and used the inverse variance weighted (IVW) method on the left CpGs (136 CpGs with 659 instruments) to perform MR. The MR P-values were then corrected by Benjamini-Hochberg method to control false discovery rate. Meanwhile, we did sensitivity analyses to test heterogeneity and horizontal pleiotropy. Consequently, we got 28 significant causal CpGs according to 3 criteria: 1) MR FDR < 0.05 ; 2) heterogeneity P-value > 0.05 ; 3) horizontal pleiotropy P-value > 0.05 . We also estimated the reverse MR effect of the 28 CpGs using the same strategy and finally, we obtained the 24 causal CpGs of eosinophil counts in blood whose reverse MR were not significant.

iv) BMI EWAS

We obtained a list of 364 BMI-associated validated CpGs from 4 EWAS BMI studies⁶⁰⁻⁶³. To test if these high-confidence BMI-associated CpGs were correlating with BMI in our NSPT cohort, we

520 performed linear regressions of DNAm against BMI with age, sex, batch, array position and array slide
521 as covariates. We derived the signed $\log_{10}$ (P-values) for a total of 302 BMI-associated CpGs with
522 representation in our EPIC array, and compared them with the corresponding values reported in the
523 BMI EWAS studies.

524 **v) Mendelian randomization test to identify potential causality between CpGs and obesity**

525 To identify causality of *NFKB1* trans-mQTLs and BMI, we conducted Mendelian Randomization
526 analysis using R package TwoSampleMR. We focused on a list containing 554 BMI-associated CpGs
527 in *NFKB1* trans-mQTL list and 364 BMI-associated validated CpGs from 4 EWAS BMI studies as
528 candidate CpGs in MR model. Firstly, to estimate the causal effects, we run causal test in Two-sample
529 MR model with candidate CpGs as exposures and BMI as outcome. We filtered the NSPT mQTLs list
530 with candidate CpGs at the threshold P-value $< 1 \times 10^{-8}$ as exposure input and took the BMI GWAS
531 summary statistics of instrument SNPs as outcome input. SNPs with LD score larger than 0.2 within
532 1000 kb window were removed by clumping, retaining only SNP with the lowest P-value in the
533 window. We used the inverse variance weighted (IVW) method to obtain an estimate of the causal
534 effect when performing MR analysis on the candidate CpGs. The MR P-values were adjusted by
535 Benjamini-Hochberg method. We then did sensitivity analyses including testing heterogeneity and
536 horizontal pleiotropy. Secondly, we conducted the inverse test with BMI as exposures and CpGs as
537 outcome. Here, we used the GWAS summary statistics of 32 BMI associated SNPs from a large GWAS
538 study⁶⁴ as exposure input and a list of mQTL statistics of candidate CpGs at the threshold P-value $<$
539 1×10^{-8} as outcome input. Similar strategy as casual test was conducted in the inverse consequential
540 test including LD clumping, IVW MR analysis and sensitivity test. In causal test and inverse test, CpGs
541 that passed 3 following criteria would be defined as casual CpGs and consequential CpGs respectively:
542 1) MR FDR < 0.05 ; 2) heterogeneity P-value > 0.05 ; 3) horizontal pleiotropy P-value > 0.05 . After
543 excluding CpGs that did not pass the sensitivity test, we obtained 170 causal CpGs of BMI in *NFKB1*
544 trans-mQTL list, 58 causal CpGs and 3 consequential CpGs in BMI EWAS CpG list.

545 **vi) BMI and SNP interaction on CpGs**

546 To interpret the relation between BMI and *NFKB1*-mediated trans-mQTLs, we carried out interaction
547 analysis of BMI and trans-mSNPs on trans-mCpGs via linear regression in R.

548
$$mCpG = \alpha + \beta_{BMI}BMI + \beta_{SNP}G + \beta_{interaction}G * BMI \quad (3)$$

549

551

552 **Fig. S1 unique features of mQTL SNPs and CpGs.**

553 **a-d**, these four histograms showing the differences between cis-, lcis- and trans-mQTLs. **a**, the minor allele frequency
554 (MAF) of SNPs; **b**, the average CpGs methylation beta values among samples; **c**, the methylation variation (standard
555 deviation, SD) among samples; **d**, the genetic explanation proportion (heritability) of CpGs. Grey, non-mQTL; red,
556 cis-mQTL; orange, lcis-mQTL; blue, trans-mQTL.

557

558

559 **Fig. S2 GO and KEGG pathway enrichment of mQTL CpGs.**

560 **a-c**, these barplots showing the Gene Ontology (GO, left plots) and KEGG pathway (right plots) enrichment of mQTL
561 CpGs (**a**, cis-; **b**, lcis-; and **c**, trans-) compared with all background CpGs conducted by R package missMethyl²⁰.
562 Only the top 10 enrichment results are displayed.

563

Fig. S3 enrichment of mQTL SNPs and CpGs for different EFO parent categories.

a-c, these scatter plots showing the enrichment of mQTL (**a**, cis-; **b**, lcis-; and **c**, trans-) SNPs (left plots) and CpGs (right plots) for different Experimental Factor Ontology (EFO) parent categories using one-tailed hypergeometric test compared with background based on significant GWAS signals from GWAS Catalog¹⁸ and EWAS signals from EWAS Atlas¹⁹. The P-values of enrichment are corrected by Bonferroni method (P.bfadjust) for 17 EFO categories in each test set.

Fig. S4 Estimated blood cell fraction in our DNAm data.

Fig. S5 The relation between density (count/Mbp) of trans-mQTL SNPs (x-axis) and TFs (y-axis).

Chromosome 19 is indicated by orange dot, and other chromosomes are indicated by blue dots.

Fig. S6 the count of unlinked trans-mCpGs for each trans-mQTL Loci and SNP.

a-b, the two lines showing the count of unlinked unlinked trans-mCpGs for each trans-mQTL SNP Locus (left, see Methods) and SNP (right).

Fig. S7 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr1. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S8 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr2. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S9 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr3. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S10 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr4. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S11 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr5. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S12 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr6. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S13 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr7. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S14 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr8. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S15 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr9. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S16 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr10. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S17 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr11. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S18 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr12. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S19 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr13. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S20 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr14. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S21 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr15. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S22 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr16. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S23 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr17. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S24 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr18. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S25 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr20. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S26 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr21. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S27 The number of unlinked trans-mCpGs (red, left y-axis) and nearby TFs (blue, right y-axis) for each trans-mSNP on chr22. x-axis shows the location of trans-mSNPs. P-value and correlation coefficient are from Spearman's rank correlation test.

Fig. S28 functional enrichment of TF/DBP genes and SNPs in trans-mQTL hotspots.

TF/DBP enrichment in super enhancer, SNPs enrichment in TAD, HiChIP promoter-promoter interaction and promoter-other interaction.

Fig. S30 SNPs in H2 affect expression of *FOSL2* through chromatin accessibility changes or chromatin interaction.

Fig. S31 Enrichment of rs4666078 trans-mCpGs in *FOSL2* binding motif.

a, Circos plot displaying the enrichment of the 232 rs4666078 trans-mCpGs for *FOSL2* binding motif (MA0478.1) (data from JASPAR 2020 database⁴⁶ and the motif enrichment test was using PWMEnrich⁴⁵, see Methods). The fraction of trans-mCpGs overlapped with the binding sites and the significance of the enrichment are given below. In the circos plots, the inner link lines show the 232 trans-mQTLs, the ones overlapped with the *FOSL2* binding motif are denoted by orange and others denoted by grey. Shown in the inner barplots are the $-\log_{10}(P)$ of trans-mQTL associations. The dots in outer tracks represent the directionality of DNA methylation change of the trans-mQTLs: red and blue denote hypermethylation and hypomethylation with the G allele, respectively (although all of these are hypomethylation). The green barplots in outer most tracks are the $-\log_{10}(P)$ of *FOSL2* binding motif enrichment.

b, Circos plot displaying the enrichment of the 141 rs4666078 myeloid-lineage trans-mCpGs in *FOSL2* binding motif (MA0478.1).

Fig. S32 SNPs in H4 affect expression of *NFKB1* through chromatin openness changes and chromatin interaction and cis-eQTL effects.

Fig. S33 Schematic overview for elucidating the downstream effects on trait/disease of trans-mQTLs.

726 **III. Supplementary Tables**

727 Table S1 The density of protein-coding genes (PCGs), TFs, TFs/PCGs (TFs after adjusting for PCGs density), trans-mQTL SNPs and all SNPs on each chromosome

CHR	Length (Mbp)	N.PCGs	N.PCGs/Mbp	N.TFs	N.TFs/Mbp	N.trans-mSNPs	N.trans-mSNPs/Mbp	TFs/PCGs	N.SNPs	N.SNPs/Mbp
chr1	249.25	2082	8.35	128	0.51	85253	342.04	0.06	597503	2397.20
chr2	243.20	1251	5.14	97	0.40	77193	317.41	0.08	624170	2566.50
chr3	198.02	1076	5.43	67	0.34	65915	332.87	0.06	544452	2749.45
chr4	191.15	756	3.95	47	0.25	82522	431.70	0.06	555649	2906.81
chr5	180.92	884	4.89	50	0.28	62043	342.94	0.06	472859	2613.70
chr6	171.12	1042	6.09	78	0.46	105745	617.98	0.07	518078	3027.66
chr7	159.14	921	5.79	97	0.61	79639	500.44	0.11	449615	2825.30
chr8	146.36	689	4.71	68	0.46	56175	383.80	0.10	409269	2796.24
chr9	141.21	802	5.68	67	0.47	44806	317.29	0.08	334920	2371.73
chr10	135.53	739	5.45	50	0.37	64153	473.33	0.07	387054	2855.75
chr11	135.01	1311	9.71	63	0.47	66945	495.86	0.05	368587	2730.14
chr12	133.85	1035	7.73	85	0.64	52781	394.32	0.08	365597	2731.35
chr13	115.17	319	2.77	23	0.20	35001	303.91	0.07	273446	2374.28
chr14	107.35	610	5.68	38	0.35	29146	271.51	0.06	248465	2314.54
chr15	102.53	603	5.88	32	0.31	27755	270.70	0.05	224038	2185.07
chr16	90.35	855	9.46	68	0.75	35355	391.29	0.08	233958	2589.33
chr17	81.20	1185	14.59	70	0.86	36766	452.81	0.06	202258	2491.01
chr18	78.08	269	3.45	28	0.36	22888	293.15	0.10	215488	2759.93
chr19	59.13	1473	24.91	331	5.60	54842	927.50	0.22	177571	3003.11
chr20	63.03	546	8.66	54	0.86	17653	280.09	0.10	161187	2557.49
chr21	48.13	243	5.05	13	0.27	15450	321.01	0.05	106539	2213.57
chr22	51.30	447	8.71	19	0.37	19998	389.79	0.04	106287	2071.69

N.PCGs the number of PCGs; N.TFs the number of TFs; N.trans-mSNPs the number of trans-mSNPs; N.SNPs the number of SNPs.

731 Table S2 the number of trans-mQTL loci and the proportion of SNPs that near-TFs as the hotspot threshold increasing

Hotness- index	N.trans-mQTL loci	Prop.trans-mQTL loci	N.trans-mQTL SNPs	Prop.trans-mQTL SNPs	N.nearTF.trans-mQTL SNPs	Prop.nearTF.trans-mQTL SNPs
=1	768	0.44	688695	0.61	369585	0.54
>=1	1727	1	1138024	1	660699	0.58
>1	959	0.56	449329	0.39	291114	0.65
>10	183	0.11	57743	0.05	44997	0.78
>20	85	0.05	14016	0.01	13081	0.93
>40	33	0.02	3925	0.003	3763	0.96
>60	23	0.01	1625	0.0018	1565	0.96
>97	16	0.01	806	0.001	776	0.96

732 Hotness-index the number of unlinked trans-mCpGs associated with the mSNP; N.trans-mQTL loci the number of trans-mQTL loci; Prop.trans-mQTL loci proportion of trans-
733 mQTL loci in hotspots compared with all trans-mQTL loci.

734

735 Table S3 overlap of nearby TF/DBPs related cis-eQTL and trans-mQTLs in each trans-mQTL hotspots

hotspot	TF_2mb	GTEx v8		CAGE	
		No. of cis-eQTL	overlap of trans-mQTL and cis-eQTL	No. of cis-eQTL	overlap of trans-mQTL and cis-eQTL
H1	<i>NF1A</i>	0	0	0	0
H2	<i>FOSL2</i>	34	0	0	0
H4	<i>REST</i>	0	0	0	0
H5	<i>NFKB1</i>	136	136	194	153
H8	<i>TFEC</i>	0	0	0	0
H10	<i>NKX2-3</i>	0	0	0	0
H11	<i>TBX6</i>	280	224	0	0
H12	<i>HLF</i>	0	0	0	0
H14	<i>FOXA3</i>	0	0	0	0
H16	<i>ERG</i>	0	0	0	0
H6	<i>CDC5L/NFKBIE</i>	0	0	103(<i>NFKBIE</i>)	62(<i>NFKBIE</i>)
H13	<i>SETBP1</i>	0	0	0	0
H3	<i>SP3/SP9</i>	18 (<i>SP3</i>)	0	0	0
H7	<i>SP4</i>	4	0	258	39
H9	<i>CHD7</i> (DBP) ^a	180	38	171	12
H15	<i>E2F1/FOXS1/PLAGL2</i>	8 (<i>E2F1</i>)	0	0	0

736 cis-eQTL P-value threshold <10⁻⁵

737 ^a *CHD7* is a DNA binding protein.

738

739 Table S4 overlap of nearby TF/DBPs related cis-eQTL in GTEx V7 (P-value<0.05) and trans-mQTLs in each trans-mQTL hotspots

hotspot	TF_2mb	GTEx v7	
		No. of cis-eQTL	overlap of trans-mQTL and cis-eQTL
H1	<i>NFIA</i>	189	2
H2	<i>FOSL2</i>	209	0
H4	<i>REST</i>	634	49
H5	<i>NFKB1</i>	237	194
H8	<i>TFEC</i>	297	102
H10	<i>NKX2-3</i>	0	0
H11	<i>TBX6</i>	446	315
H12	<i>HLF</i>	144	11
H14	<i>FOXA3</i>	0	0
H16	<i>ERG</i>	265	2
H6	<i>CDC5L/NFKBIE</i>	60 (<i>NFKBIE</i>)	0
H13	<i>SETBP1</i>	405	79
H3	<i>SP3/SP9</i>	0	0
H7	<i>SP4</i>	262	30
H9	<i>CHD7</i> (DBP) ^a	322	53
H15	<i>E2F1/FOXS1/PLAGL2</i>	0	0

740 cis-eQTL P-value threshold <0.05

741 ^a *CHD7* is a DNA binding protein.

748 Table S5 enrichment of TF/DBP with super enhancers in trans-mQTL hotspots

hotspot	TF/DBP	Annotation from Super enhancer database ^a		
		SEdb ⁵⁰	dbSUPER ⁵¹	SEanalysis ⁵²
H1	<i>NFIA</i>	YES(M) ^b	YES(M)	YES(M)
H2	<i>FOSL2</i>	YES(M)	YES(M)	YES(M)
H4	<i>REST</i>	YES(Embryo)	NO	YES(Embryo)
H5	<i>NFKB1</i>	YES (M)	YES (blood)	YES (M)
H8	<i>TFEC</i>	YES (M)	YES (NCI-H69)	YES (M)
H10	<i>NKX2-3</i>	YES (M)	YES(Spleen/Sigmoid Colon/Small Intestine)	YES (M)
H11	<i>TBX6</i>	YES (Spleen)	YES (Spleen)	YES (M)
H12	<i>HLF</i>	YES (M)	YES (Esophagus)	YES (M)
H14	<i>FOXA3</i>	YES (M not blood) ^c	YES (Pancreatic islets/Colon Crypt)	YES (M not blood)
H16	<i>ERG</i>	YES (M)	YES (M)	YES (M)
H6	<i>CDC5L</i> <i>NFKBIE</i>	CDC5L(M not blood) NFKBIE(blood/Neuroblastoma)	CDC5L (HSMM/NHEK/u87) NFKBIE (Ly3)	CDC5L(M)/NFKBIE(M)
H13	<i>SETBP1</i>	YES (M)	YES (M)	YES (M)
H3	<i>SP3</i> <i>SP9</i>	SP3(M) SP9(Skeletal muscle)	YES (M)	SP3(M)/SP9(Skeletal muscle)
H7	<i>SP4</i>	YES(Embryo/peripheral blood)	YES (M)	YES(Embryo/peripheral blood)
H9	<i>CHD7</i>	YES (M)	YES (M)	YES (M)
H15	<i>E2F1</i> <i>FOXS1</i> <i>PLAGL2</i>	FOXS1(M not blood) PLAGL2(M not blood)	PLAGL2(Fetal Intestine /CD4p cancer / HMEC)	E2F1(M not blood) /FOXS1(M not blood) /PLAGL2(M)

749 ^a three super enhancer database applied here.
750 ^b M indicates multiple tissues or cell types.
751 ^c M not blood indicates multiple tissues except blood.
752

Table S7 24 putative causal CpGs (rs4666078 trans-associated) of eosinophil counts in blood (reverse was not significant) identified by Two-Sample MR analysis

DNA _m	Gene	Instrument SNPs	MR beta	MR se	MR P-value	MR FDR	Heterogeneity P-value	Horizontal pleiotropy P-value	Tissue eosinophilia	DisGe NET
cg22652934	<i>RUNX1</i> (Body)	3	0.18	0.04	1.64×10^{-6}	1.71×10^{-5}	0.90	0.76	POS	Yes
cg22425699	5q14.2(intergenic)	5	0.20	0.03	9.32×10^{-13}	1.27×10^{-10}	0.84	0.83	POS	
cg02734358	<i>GPRIN3</i> (5'UTR)	4	0.27	0.05	1.15×10^{-7}	1.57×10^{-6}	0.22	0.97	POS	
cg14784010	<i>DOCK2</i> (Body)	5	0.15	0.04	3.09×10^{-4}	1.83×10^{-3}	0.42	0.92	POS	
cg04331199	<i>TCTEX1D1</i> (Body)	16	0.08	0.03	3.88×10^{-3}	0.02	0.11	0.81	POS	
cg00947809	10q22.1(intergenic)	4	0.33	0.05	8.69×10^{-12}	5.91×10^{-10}	0.56	0.80		
cg04915079	<i>KIAA1683</i> (TSS1500)	5	0.51	0.08	8.99×10^{-11}	4.08×10^{-9}	0.30	0.70		
cg02343439	<i>SAG</i> (1stExon,5'UTR)	3	0.31	0.05	1.42×10^{-10}	4.76×10^{-9}	0.54	0.98		
cg16902078	<i>KCCAT333</i> (Body)	3	0.32	0.05	1.81×10^{-10}	4.76×10^{-9}	0.62	0.79		
cg23599196	<i>SAG</i> (1stExon,5'UTR)	3	0.28	0.04	2.10×10^{-10}	4.76×10^{-9}	0.53	0.95		
cg15346406	5p13.1(intergenic)	3	0.34	0.06	1.38×10^{-9}	2.69×10^{-8}	0.34	0.93		
cg21320221	<i>DYSFIP1</i> (TSS200)	5	0.24	0.04	1.68×10^{-9}	2.85×10^{-8}	0.39	0.96		
cg02862897	<i>DYSFIP1</i> (TSS200)	6	0.25	0.05	5.04×10^{-8}	7.62×10^{-7}	0.20	0.61		
cg04308793	20q11.21(intergenic)	4	0.23	0.05	4.32×10^{-7}	5.28×10^{-6}	0.20	0.53		
cg24380489	<i>DYSFIP1</i> (TSS200)	5	0.21	0.04	4.66×10^{-7}	5.28×10^{-6}	0.16	0.94		
cg06938558	<i>PEX7</i> (Body)	3	0.20	0.05	4.03×10^{-5}	3.43×10^{-4}	0.65	0.52		
cg09117950	<i>NCOR1</i> (Body)	6	0.08	0.02	1.15×10^{-4}	8.69×10^{-4}	0.38	0.37		
cg05041552	<i>RFT1</i> (Body)	3	0.20	0.05	1.60×10^{-4}	1.09×10^{-3}	0.86	0.68		
cg16286413	<i>IQGAP2</i> (Body)	6	0.11	0.03	3.09×10^{-4}	1.83×10^{-3}	0.57	0.36		
cg18213700	19q13.11(intergenic)	4	0.18	0.05	4.81×10^{-4}	2.73×10^{-3}	0.11	0.28		
cg15568069	<i>LINC01607</i> (Body), <i>LOC101927040</i> (Body)	5	0.11	0.03	6.06×10^{-4}	3.30×10^{-3}	0.49	0.41		
cg04674809	7q36.1(intergenic)	6	0.11	0.04	1.78×10^{-3}	0.01	0.64	0.61		
cg21554670	<i>GAS7</i> (Body)	3	0.18	0.06	0.01	0.03	0.32	0.38		
cg18346531	<i>GAS7</i> (Body)	3	0.19	0.08	0.01	0.04	0.28	0.36		

759 Table S8&S9 see Table S8-S9.xls

760

761

IV. Supplementary References

1. DeFronzo, R.A. Chiglitazar: a novel pan-PPAR agonist. *Science Bulletin* **66**, 1497-1498 (2021).
2. Ji, L. *et al.* Efficacy and safety of chiglitazar, a novel peroxisome proliferator-activated receptor pan-agonist, in patients with type 2 diabetes: a randomized, double-blind, placebo-controlled, phase 3 trial (CMAP). *Science Bulletin* **66**, 1571-1580 (2021).
3. Jia, W. *et al.* Chiglitazar monotherapy with sitagliptin as an active comparator in patients with type 2 diabetes: a randomized, double-blind, phase 3 trial (CMAS). *Science Bulletin* **66**, 1581-1590 (2021).
4. Aryee, M.J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-9 (2014).
5. Teschendorff, A.E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189-96 (2013).
6. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-27 (2007).
7. Tian, Y. *et al.* ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* **33**, 3982-3984 (2017).
8. Zheng, S.C., Breeze, C.E., Beck, S. & Teschendorff, A.E. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods* **15**, 1059-1066 (2018).
9. Gao, X. *et al.* FastQTLmapping: an ultra-fast package for mQTL-like analysis. (bioRxiv, 2021).
10. Shabalin, A.A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-8 (2012).
11. Huan, T. *et al.* Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. *Nat Commun* **10**, 4267 (2019).
12. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
13. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nat Methods* **8**, 833-5 (2011).
14. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
15. Zhou, W., Laird, P.W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic acids research* **45**, e22 (2017).
16. Lizio, M. *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome biology* **16**, 22 (2015).
17. Bernstein, B.E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-8 (2010).
18. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-d1012 (2019).
19. Li, M. *et al.* EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res* **47**, D983-d988 (2019).
20. Phipson, B., Maksimovic, J. & Oshlack, A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics (Oxford, England)* **32**, 286-288 (2016).
21. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
22. Eisenberg, E. & Levanon, E.Y. Human housekeeping genes, revisited. *Trends in Genetics* **29**, 569-574 (2013).

805 23. Howe, K.L. *et al.* Ensembl 2021. *Nucleic Acids Res* **49**, D884-d891 (2021).

806 24. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-
807 wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).

808 25. Teschendorff, A.E., Breeze, C.E., Zheng, S.C. & Beck, S. A comparison of reference-based algorithms for
809 correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinformatics* **18**, 105
810 (2017).

811 26. Zheng, S.C. *et al.* EpiDISH web server: Epigenetic Dissection of Intra-Sample-Heterogeneity with online
812 GUI. *Bioinformatics* **36**, 1950-1 (2019).

813 27. You, C. *et al.* A cell-type deconvolution meta-analysis of whole blood EWAS reveals lineage-specific
814 smoking-associated DNA methylation changes. *Nat Commun* **11**, 4779 (2020).

815 28. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell*
816 **167**, 1398-1414.e24 (2016).

817 29. Breeze, C.E. *et al.* eFORGE: A Tool for Identifying Cell Type-Specific Signal in Epigenomic Data. *Cell Rep*
818 **17**, 2137-2150 (2016).

819 30. Breeze, C.E. *et al.* eFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data.
820 *Bioinformatics* **35**, 4767-4769 (2019).

821 31. Li, W., Duren, Z., Jiang, R. & Wong, W.H. A method for scoring the cell type-specific impacts of noncoding
822 variants in personal genomes. *Proc Natl Acad Sci U S A* **117**, 21364-21372 (2020).

823 32. Javierre, B.M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease
824 Variants to Target Gene Promoters. *Cell* **167**, 1369-1384.e19 (2016).

825 33. Cairns, J. *et al.* CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome*
826 *Biol* **17**, 127 (2016).

827 34. Mumbach, M.R. *et al.* Enhancer connectome in primary human cells identifies target genes of disease-
828 associated DNA elements. *Nat Genet* **49**, 1602-1612 (2017).

829 35. Durand, N.C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments.
830 *Cell Syst* **3**, 95-8 (2016).

831 36. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*
832 **489**, 57-74 (2012).

833 37. Lambert, S.A. *et al.* The Human Transcription Factors. *Cell* **172**, 650-665 (2018).

834 38. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic acids*
835 *research* **47**, D766-D773 (2019).

836 39. Kamat, M.A. *et al.* PhenoScanner V2: an expanded tool for searching human genotype-phenotype
837 associations. *Bioinformatics* **35**, 4851-4853 (2019).

838 40. Staley, J.R. *et al.* PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* **32**,
839 3207-3209 (2016).

840 41. Oki, S. *et al.* ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO*
841 *Rep* **19**(2018).

842 42. Albert, F.W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet*
843 **16**, 197-212 (2015).

844 43. Yao, C. *et al.* Dynamic Role of trans Regulation of Gene Expression in Relation to Complex Traits. *Am J*
845 *Hum Genet* **100**, 571-580 (2017).

846 44. Leporcq, C. *et al.* TFmotifView: a webserver for the visualization of transcription factor motifs in genomic
847 regions. *Nucleic acids research* **48**, W208-W217 (2020).

848 45. Stojnic, R. & Diez, D. PWMEnrich: PWM enrichment analysis. (R package version 4.28.1, 2021).

849 46. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles.
850 *Nucleic acids research* **48**, D87-D92 (2020).

851 47. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).

852 48. Stower, H. Gene expression: Super enhancers. *Nat Rev Genet* **14**, 367 (2013).

853 49. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity
854 genes. *Cell* **153**, 307-19 (2013).

855 50. Jiang, Y. *et al.* SEdb: a comprehensive human super-enhancer database. *Nucleic Acids Res* **47**, D235-d243
856 (2019).

857 51. Khan, A. & Zhang, X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids*
858 *Res* **44**, D164-71 (2016).

859 52. Qian, F.C. *et al.* SEanalysis: a web tool for super-enhancer associated regulatory analysis. *Nucleic Acids Res*
860 **47**, W248-w255 (2019).

861 53. Piñero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic acids*
862 *research* **48**, D845-D855 (2020).

863 54. Wang, K. *et al.* Concordant systemic and local eosinophilia relates to poorer disease control in patients with
864 nasal polyps. *The World Allergy Organization journal* **12**, 100052 (2019).

865 55. Kim, K.W. *et al.* Integrated genetic and epigenetic analyses uncover MSI2 association with allergic
866 inflammation. *The Journal of allergy and clinical immunology* **147**, 1453-1463 (2021).

867 56. Chen, M.H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5
868 Global Populations. *Cell* **182**, 1198-1213.e14 (2020).

869 57. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome.
870 *Elife* **7**(2018).

871 58. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured
872 traits using GWAS summary data. *PLoS genetics* **13**, e1007081 (2017).

873 59. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on
874 complex diseases. *Nature genetics* **52**, 1122-1131 (2020).

875 60. Dick, K.J. *et al.* DNA methylation and body-mass index: a genome-wide analysis. *Lancet* **383**, 1990-8 (2014).

876 61. Mendelson, M.M. *et al.* Association of Body Mass Index with DNA Methylation and Gene Expression in
877 Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach. *PLoS Med*
878 **14**, e1002215 (2017).

879 62. Wahl, S. *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity.
880 *Nature* **541**, 81-86 (2017).

881 63. Demerath, E.W. *et al.* Epigenome-wide association study (EWAS) of BMI, BMI change and waist
882 circumference in African American adults identifies multiple replicated loci. *Hum Mol Genet* **24**, 4464-79
883 (2015).

884 64. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body
885 mass index. *Nat Genet* **42**, 937-48 (2010).

886

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS8S9.xls](#)