

Cis-meQTL mapping reveals genetic architecture of DNA methylation in a giga-genome species

Victor Chano^{1,2,*}, Konstantin V. Krutovsky^{1,2}, Kai Wang³, Matti Haapanen⁴, Fred O. Asiegbu³, Oliver Gailing^{1,2,*}

¹ Department of Forest Genetics and Forest Tree Breeding, University of Göttingen. Büsgenweg 2, 37077 Göttingen, Germany.

² Center for Integrated Breeding Research (CiBreed), University of Göttingen. Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany.

³ Department of Forest Science, University of Helsinki. FIN-00014 Helsinki, Finland.

⁴ Natural Resources Institute Finland (LUKE). Latokartanonkaari 9, Helsinki 00790, Finland.

* Corresponding authors

ORCID:

Victor Chano 0000-0002-3423-9090

Konstantin V. Krutovsky 0000-0002-8819-7084

Kai Wang 0000-0002-5205-7346

Matti Haapanen 0000-0003-3294-501X

Fred O. Asiegbu 0000-0003-0223-7194

Oliver Gailing 0000-0002-4572-2408

23 **Supplementary Information**

24 **Supplementary Notes:**

- 25 – Supplementary Note 1. Processing of raw sequencing data and nucleotide read mapping
- 26 – Supplementary Note 2. Cohort genetic structure, genomic relationships and identity-by-
27 descent
- 28 – Supplementary Note 3. Statistical analysis of methylation levels, and detection of
29 significantly variable methylated positions (SVMPs)
- 30 – Supplementary Note 4. Genetic-epigenetic concordance
- 31 – Supplementary Note 5. Cis-methylation quantitative trait loci (cis-meQTL) mapping and
32 genomic annotation of associated markers

33 **Supplementary Tables**

34 **Supplementary Figures:**

- 35 – Supplementary Figure S1. Procrustes biplots of genomic and epigenomic ordinations
- 36 – Supplementary Figure S2. Quantile–quantile (QQ) plots of MatrixEQTL cis-meQTL *p*-
37 values
- 38 – Supplementary Figure S3. Quantile–quantile (QQ) plots of GENESIS cis-meQTL *p*-values
- 39 – Supplementary Figure S4. Venn diagrams of SNPs detected by MatrixEQTL and
40 GENESIS

41

42

Supplementary Note 1

Processing of raw sequencing data and nucleotide read mapping

Raw nucleotide sequencing reads from both ECS and TMS libraries were quality-evaluated and trimmed, then mapped to the Norway spruce v2.0 reference genome (<https://doi.org/10.21203/rs.3.rs-6502828/v1>) using separate alignment pipelines suited to standard and converted reads. PCR duplicates were removed from all aligned files prior to downstream processing. We have successfully sequenced 602 Norway spruce seedlings (209 trees from the breeding cohort and 393 trees from the natural cohort) by both ECS and TMS approaches (Supplementary Tables S1 and S2). Quality control of raw reads confirmed high overall sequencing quality across all ECS and TMS libraries, with per-base quality scores predominantly above Q30 and low adapter contamination rates before trimming. After adapter removal and quality trimming, paired-end read yields were sufficient for downstream analysis. Across all retained samples, ECS produced an average of $3,030,916 \pm 602,711$ reads per sample (Supplementary Table S1), whereas TMS yielded $2,947,908 \pm 1,034,358$ reads per sample. As expected, alignment performance differed markedly between approaches, with ECS showing a very high overall alignment rate ($99.28 \pm 0.07\%$), while TMS exhibited a lower total alignment rate ($77.16 \pm 8.10\%$).

Supplementary Note 2

Cohort genetic structure, genomic relationships and identity-by-descent

Initially, SNPs were called jointly across all samples, filtered on minor allele frequency ($MAF \geq 0.05$), missingness ($\leq 20\%$), and minimum coverage threshold ($\geq 10\times$). We identified a set of 488,826 high-quality SNPs, which was then cohort-based split. After removing within-cohort monomorphic sites, we obtained 450,947 high-quality SNPs for the breeding cohort and 462,617 for the natural cohort (Supplementary Table S1). These results reflect the high genetic diversity expected for this outcrossing conifer and a broad genomic coverage of the exome capture design. Following BEAGLE imputation, phased and imputed dosage genotype datasets were available for both cohorts for all subsequent analyses.

Plotting PC1 against PC3 (8.9%) revealed additional family differentiation in the breeding cohort (Fig. 1B), and when PC2 was plotted against PC3, the family clusters remained tight and well-defined (Fig. 1C). In the natural cohort, the third component (PC3, 5.3%) introduced a further axis of variation, yet the overall pattern remained a diffuse cloud with no discrete population breaks (Fig. 1E), and examining PC2 against PC3 confirmed that the residual axes of variation were distributed uniformly around the origin, with natural stands broadly overlapping and only scattered outlier individuals projecting toward the extremes (Fig. 1F). Taken together, the breeding cohort showed a strong, pedigree-driven family structure whereas the genetic differentiation in natural spruce populations followed a continuous gradient consistent with widespread historical gene flow across the landscape.

The genomic relatedness matrix (GRM) of the breeding cohort (Extended Data Fig. 1A) revealed possible cross-family relatedness, notably between Families 38 and 47, between

Families 40 and 41, and Family 33 with Families 32 and 27. This possibly reflects shared ancestry in the breeding program. The identity-by-descent (IBD) violin plot for the breeding cohort (Extended Data Fig. 1B) showed IBD estimate values for between-family pairs ($n = 20,473$) concentrated between 0.05 and 0.10, while within-family pairs ($n = 1,263$) demonstrated a broader distribution centered between the values corresponding to half-sib ($\pi = 0.25$) and full-sib ($\pi = 0.50$) pairs according to their pedigree, with no anomalous low values suggesting pedigree misassignment. The natural cohort GRM (Extended Data Fig. 1C) lacked any discrete block structure, with uniformly low off-diagonal values on a compressed scale (-0.01 to 0.02). Accordingly, the natural cohort violin plots (Extended Data Fig. 1D) showed that both between-stand ($n = 72,505$) and within-stand ($n = 4,523$) IBD estimate value distributions were nearly indistinguishable and concentrated below the half-sib threshold, consistent with a design in which parent trees were from large, freely outbreeding natural populations and produced largely unrelated offspring in each stand group.

Supplementary Note 3

Statistical analysis of methylation levels, and detection of significantly variable methylated positions (SVMPs)

To facilitate and accelerate computational analysis, methylation calling was performed separately for the breeding and natural cohorts and independently for each methylation context (CpG, CHG, and CHH sequence sites). To avoid confounding sequence polymorphism with methylation variation, 158,584 sites previously identified as true C>T conversions from the ECS dataset were first removed prior to methylation site filtering and further analysis. After coverage filtering ($5-50\times$) and uniting across cohort samples, the final methylation datasets retained 34,595 CpG, 85,836 CHG, and 262,566 CHH sites for the breeding cohort, and 63,109 CpG, 150,543 CHG, and 479,363 CHH sites for the natural cohort (Supplementary Table S2). Application of minor epiallele frequency filtering ($MEF \geq 0.05$), retaining only epiloci with sufficient within-cohort methylation variability, reduced them to 855 CpG, 3,871 CHG, and 8,496 CHH sites (breeding cohort) and 829 CpG, 2,841 CHG, and 6,143 CHH sites (natural cohort) (Supplementary Table S2).

CpG methylation in the breeding cohort (Extended Data Fig. 2A) showed a wide range across the 17 families, with Family 16 forming a clearly distinct low-methylation group; one-way ANOVA revealed significant between-family differences ($p = 2.61 \times 10^{-3}$), with 7 of 136 pairwise comparisons reaching significance after Tukey's HSD correction. In the natural cohort (Extended Data Fig. 2B), CpG methylation showed substantial between-stand differentiation ($p = 2.80 \times 10^{-8}$; 13 of 300 pairs were significant). CHG methylation in the breeding cohort (Extended Data Fig. 2C) showed the most pronounced between-family differentiation of all six panels ($p = 3.93 \times 10^{-25}$; 30 of 136 pairs were significant), with Family 27 standing out as markedly hypermethylated. In the natural cohort (Extended Data Fig. 2D), CHG methylation occupied a compressed range, with 13 of 300 stand-pairs differing significantly ($p = 2.21 \times 10^{-7}$). CHH methylation in the breeding cohort (Extended Data Fig. 2E) showed no significant between-family variation ($p = 0.098$). In the natural cohort (Extended Data Fig. 2F), CHH methylation showed six of 300 stand-pairs differing significantly ($p = 1.58 \times 10^{-3}$) (Supplementary Table S2).

Kruskal-Wallis tests applied independently to each MEF-retained locus revealed a striking asymmetry between cohorts. In the breeding cohort, 5,229 SVMPs were identified at the adjusted p -value < 0.05 after the Benjamini-Hochberg (BH) correction: 616 CpG, 2,224 CHG, and 2,389 CHH loci, corresponding to 72%, 57%, and 28% of MEF-retained loci in each methylation context, respectively (Supplementary Tables S2 and S3). In the natural cohort, by contrast, only six SVMPs survived correction: three CpG and three CHH, with no CHG locus reaching significance. The contrast between the number of SVMPs in both cohorts likely reflected the fundamental difference in epigenetic architecture between heritable, family-specific epialleles producing strong discrete signals in the structured breeding pedigree, versus diffuse, environmentally modulated variation distributed across thousands of loci in the natural population. The hierarchically clustered heatmap of the breeding cohort (Extended Data Fig. 3A), displaying the top-150 highest-ranking loci (50 per methylation context), revealed a clear context-driven clustering, with CpG loci uniformly showing high methylation, while CHG and CHH loci were predominantly hypomethylated. This methylation site-specific polarity is most clearly illustrated in the natural cohort heatmap (Extended Data Fig. 3B), where the three CpG SVMPs clustered together showing hypermethylation in most of the samples (regardless of stand identity), while the three CHH SVMPs were predominantly hypomethylated.

Supplementary Note 4

Genetic-epigenetic concordance

Pairwise genetic and epigenomic distance matrices were computed independently for each cohort and cytosine context, and their concordance was assessed using three permutation-based approaches, the Mantel test based on Pearson correlation between distance matrices, Procrustes/PROTEST analysis based on alignment of ordinations after optimal rotation, and the RV coefficient based on multivariate generalization of squared Pearson correlation; all tests used 9,999 permutations.

In the breeding cohort, all three statistics consistently indicated significant concordance across all three cytosine contexts (Supplementary Table S4). Mantel r values ranged from 0.21 (CHH) to 0.38 (CpG), all with $p < 1 \times 10^{-4}$. Procrustes t_0 values were 0.75 (CpG), 0.60 (CHG), and 0.61 (CHH), all highly significant ($p < 1 \times 10^{-4}$), indicating that the PCoA ordinations derived from genetic and epigenetic distance matrices coalign substantially after optimal rotation. The RV coefficient confirmed the same trend (CpG: 0.57; CHG: 0.47; CHH: 0.38; all with $p < 1 \times 10^{-4}$), with the strongest concordance consistently observed for CpG methylation. These results indicate that genetically similar individuals also tend to have more similar methylation profiles, reflecting the shared family background in the structured pedigree. Visually, the Procrustes biplots for the breeding cohort (Supplementary Fig. S1A-C) showed similar shapes and relatively short residuals across both ordinations between the genetic (SNPs) and epigenetic (methylation sites) ordinations.

In the natural cohort, concordance was markedly weaker and largely non-significant across all three contexts. Mantel r values ranged from only 0.02 to 0.06 (all insignificant with $p > 0.05$), and Procrustes and RV statistics were similarly insignificant for CpG and CHG, with the only exception of the CHH context, where both the Procrustes statistic ($t_0 = 0.12$, $p = 0.008$) and the

RV coefficient ($RV = 0.016$, $p = 0.012$) reached significance. The Procrustes biplots for the natural cohort (Supplementary Fig. S1D-F) illustrate this decoupling, with the epigenetic ordination collapsing into a central cluster while the genetic ordination retained appreciable spread, indicating that individual methylation profiles were essentially unrelated to genetic distances at the population level.

Supplementary Note 5

Cis-methylation quantitative trait loci (cis-meQTL) mapping and genomic annotation of associated markers

At $FDR < 1 \times 10^{-10}$, MatrixEQTL identified 6,406, 6,311, and 4,294 significant SNP-methylation site pairs involving 2,406, 2,298, and 2,005 unique methylation sites and 4,533, 2,568, and 3,557 unique SNPs for the CpG, CHG, and CHH methylation sites in the breeding cohort, respectively, and 11,750, 15,097, and 14,223 pairs involving 4,322, 5,776, and 8,253 unique methylation sites and 7,412, 6,129, and 10,620 unique SNPs for the CpG, CHG, and CHH methylation sites in the natural populations (Table 1). The genome-wide distribution of these associations for the breeding and natural cohorts are shown in Extended Data Fig. 4A and 4B, respectively, with significant associations distributed across all 12 chromosomes. Quantile–quantile plots presented for the CpG, CHG, and CHH sites in the breeding (Supplementary Fig. S2A-C) and natural (Supplementary Fig. S2D-F) cohorts, and genomic inflation factors (λ ; Table 1) reveal notably higher genomic inflation for the CpG site in the breeding cohort ($\lambda = 1.34$), with values of 0.86 and 0.69 for the CHG and CHH sites in this cohort, and 0.92, 0.70, and 0.53 for the CpG, CHG, and CHH sites in the natural cohort. On the other hand, and also at $FDR < 1 \times 10^{-10}$, GENESIS identified consistently smaller numbers of significant associations, with 2,863, 2,645, and 1,976 pairs involving 1,257, 1,210, and 953 unique methylation sites and 2,054, 1,232, and 1,667 unique SNPs for the CpG, CHG, and CHH sites in the breeding cohort, respectively; and 7,054, 8,219, and 5,385 pairs involving 2,475, 2,887, and 2,477 unique methylation sites and 4,548, 3,135, and 4,239 unique SNPs for the CpG, CHG, and CHH sites in the natural cohort (Table 1). Genome-wide associations are shown in Fig. 3A for the breeding and in Fig. 3B for the natural cohorts. Quantile–quantile plots for the CpG, CHG, and CHH sites in the breeding (Supplementary Fig. S3A-C) and natural (Supplementary Fig. S3D-F) cohorts, and genomic inflation factors (λ ; Table 1) confirmed better-controlled p-value distributions with $\lambda = 1.21$, 0.84, and 0.72 for the CpG, CHG, and CHH sites in the breeding cohort, and 0.93, 0.87, and 0.73 for the CpG, CHG, and CHH sites in the natural cohort, respectively (Table 1). However, due to the structured relatedness in the breeding cohort, principal components alone do not fully absorb pedigree confounding, which is reflected in the higher λ for CpG sites under MatrixEQTL. The kinship-based random effect in GENESIS explicitly models this family structure, accounting for the better-calibrated λ values.

SNP-methylation site pairs (same pair at $FDR < 1 \times 10^{-10}$ in both methods) were selected as robust associated markers (pair-level criterion; Supplementary Fig. S4A-F). While most GENESIS-significant SNPs were also detected by MatrixEQTL, a substantial fraction of MatrixEQTL-only SNPs lacked support from the mixed model, consistent with those being inflated associations driven by unmodelled relatedness. In the breeding cohort, robust associations comprised 2,733, 2,491, and 1,730 SNP–methylation site pairs involving 1,204, 1,138, and 769

unique sites and 1,978, 1,166, and 1,477 unique SNPs in the CpG, CHG, and CHH sites, respectively, for a total of 6,954 robust pairs (3,111 unique sites; 3,952 unique SNPs). In the natural cohort, robust associations comprised 6,611, 7,346, and 4,523 pairs involving 2,387, 2,660, and 1,797 unique sites and 4,299, 2,840, and 3,630 unique SNPs in the CpG, CHG, and CHH sites, for a total of 18,480 robust pairs (6,844 unique sites; 8,498 unique SNPs) (Supplementary Table S5).

We compared robust SNP sets across methylation contexts within each cohort, revealing predominantly cohort-specific genetic control. In the breeding cohort, 1,501, 740, and 1,105 SNPs were exclusively associated with CpG, CHG, and CHH methylation sites, respectively, with smaller fractions spanning two contexts (CpG $\cap$ CHG: 234 SNPs; CpG $\cap$ CHH: 180; CHG $\cap$ CHH: 129), and 63 SNPs were significant across all three context-specific analyses (Fig. 4A). In the natural cohort, the pattern was analogous but at a larger scale, including 2,801, 1,456, and 2,307 context-exclusive SNPs for CpG, CHG, and CHH, two-context overlaps of 611 (CpG $\cap$ CHG), 550 (CpG $\cap$ CHH), and 436 (CHG $\cap$ CHH), and 337 SNPs common to all three methylation contexts (Fig. 4B). The methylation context-specific classification reflects which one of the three separate meQTL analyses, each testing the same SNP against three distinct sets of methylation sites, one per context (CpG, CHG, and CHH), identifies that variant as significant. A SNP in the three-way overlap therefore predicts methylation across all three methylation contexts simultaneously. Additionally, the comparison of robust cis-meQTL positions between cohorts within each methylation context revealed extensive cross-cohort conservation. Cross-cohort comparisons were performed using true genomic positions (chromosome name and base-pair coordinate from the SNP variant annotation file), as integer SNP identifiers are not transferable across the independently constructed cohort-specific genotype datasets. For CpG, 1,409 of 1,978 breeding robust SNP positions were also found in the natural cohort (71%), with 569 exclusive to the breeding cohort and 2,890 exclusive to the natural cohort (Fig. 4C). For CHG, 843 of 1,166 positions were shared (72%), with 323 breeding-exclusive and 1,997 natural-exclusive (Fig. 4D). For CHH, 1,026 of 1,477 positions were shared (70%), with 451 breeding-exclusive and 2,604 natural-exclusive (Fig. 4E). These fractions are recoveries of the breeding positions. The natural cohort carries 2.2-fold more robust positions (8,498 against 3,952), so the reciprocal fractions are 1,409 of 4,299 natural CpG positions (33%), 843 of 2,840 for CHG (30%) and 1,026 of 3,630 for CHH (28%), and the union-based overlap is 28.9%, 26.7% and 25.1%, respectively. The asymmetry is expected as the breeding cohort is smaller ($n = 209$ against 393) and its allele frequencies are constrained by 17 pedigrees. The consistent $\sim 70\%$ recovery across all three methylation contexts indicates that the genetic architecture of DNA methylation control detected within the pedigree is substantially conserved in natural populations, while cohort-exclusive positions likely reflect differences in local allele frequencies.

Cis-meQTL SNPs and methylation sites associated with them were annotated using the Norway spruce v2.0 gene models, classifying markers as genic (located within genes), proximal intergenic (located within less than 2 Kb apart from genes), or distal intergenic (located more than 2 Kb apart from genes). For each SNP, the full set of associated methylation sites across all significant contexts was recorded. Methylation sites were characterized by the mean beta value and classified as hypomethylated, intermediate, or hypermethylated using context-specific

thresholds (CpG: $< 0.30 / \geq 0.70$; CHG: $< 0.20 / \geq 0.50$; CHH: $< 0.10 / \geq 0.30$). Cis-meQTL SNPs mapped to 2,552 unique genes in total, including 1,068 in both cohorts, 247 exclusively in the breeding cohort, and 1,237 exclusively in the natural cohort. The majority were genic (87% in the breeding cohort, 83% in the natural cohort), consistent with the ECS design. Concerning TMS, meQTL methylation sites mapped to 2,520 unique genes, including 1,056 that were shared between cohorts, 244 exclusive to breeding, and 1,220 to the natural cohort. As expected, the majority were again genic (87% and 84%, respectively). The source gene set (2,552) and target gene set (2,520) were largely co-extensive, with 2,477 genes (97.1%) appearing in both, consistent with a distributed cis-regulatory architecture in which regulatory SNPs and their methylation targets largely reside within the same genomic regions rather than in a hierarchical hotspot structure. The 2,477 genes appearing in both source and target sets harbour both regulatory SNPs and regulated methylation sites at the same locus, consistent with pervasive local cis-regulatory relationships across the meQTL landscape.

Most cis-meQTL SNPs were associated within a single methylation context, consistent with the independent enzymatic maintenance of CpG, CHG, and CHH methylation. A smaller subset showed cross-context pleiotropic effects; the source genes whose cis-meQTL SNPs were each associated with the most methylation sites simultaneously across all three contexts in the breeding cohort were PA_chr02_G000527 (no annotation; 24 methylation sites), PA_chr05_G000628 (encoding purple acid phosphatase; 22 sites), and PA_chr12_G003172 (encoding a pumilio homolog; 21 sites); and in the natural cohort, PA_chr05_G003300 (no annotation; 68 sites), PA_chr01_G007153 (no annotation; 47 sites), and PA_chr01_G003672 (encoding BTB/POZ domain-containing protein; 39 sites). PA_chr05_G003300 is notable in that both the meQTL SNP and regulated methylation sites reside within the same gene body, consistent with a possible cis-regulatory association within this locus.

Table 2 reports the top two cis-meQTL source genes containing the largest number of unique SNPs in the CpG∩CHG∩CHH cytosine methylation site category (complete annotated list of robust SNPs is provided in Supplementary Table S6). The two loci associated with all three methylation contexts with the largest total numbers of unique cis-meQTL SNPs were 1) PA_chr11_G000565, encoding guanosine monophosphate (GMP) synthase (glutamine-hydrolyzing), with 29 and 45 SNPs in the breeding and natural cohorts targeting 8 and 11 distinct methylation sites, respectively, and 2) PA_chr07_G001491, encoding a cell number regulator protein, with 19 and 41 SNPs in the breeding and natural cohorts targeting 13 and 32 distinct methylation sites, respectively. No robust cis-meQTL SNPs were detected at loci encoding canonical DNA methyltransferases (MET1, CMT3, or DRM2). This is consistent with purifying selection constraining functional variation at these essential genes, although limited polymorphism or reduced capture coverage at these loci cannot be excluded as contributing factors. One locus encoding a methyl-CpG binding domain (MBD) reader protein, PA_chr11_G001754, contained 2 and 6 SNPs in the breeding and natural cohorts, respectively, associated with methylation across multiple contexts, suggesting that inter-individual genetic variation may also affect the reading of methylation signals.

The top cis-meQTL target genes harbouring the most unique methylation sites under significant cis-genetic control are presented in Table 3 (the complete annotated list of robust methylation sites is provided in Supplementary Table S7). Because each SNP and each methylation site are annotated to their own nearest gene independently, source genes (Table 2) and target genes (Table 3) were annotated independently, and each SNP and its associated methylation site were assigned to their nearest gene model. With 97.1% of source genes also appearing in the target set, the two lists are largely co-extensive rather than non-overlapping. Across all three methylation contexts, the top targets were two genes (PA_chr01_G003672, encoding BTB/POZ domain protein and containing 25 and 39 sites in breeding and natural cohorts, respectively, and PA_chr10_G004214, encoding structure-specific endonuclease catalytic subunit and containing 22 and 38 sites). Several target genes also encoded components of the methylation and chromatin regulatory machinery. Most notably, the PA_chr10_G001822 gene, encoding RNA-dependent RNA polymerase (RdRP), included 13 and 19 genetically regulated methylation sites restricted to the CHG and CHH contexts, which is a core enzyme of the RdDM pathway. Additional RdDM components with genetically regulated methylation sites included 1) an Argonaute protein encoded by the PA_chr04_G002131 gene with one and five CHH sites in the breeding and natural cohorts, respectively, 2) an RdDM factor encoded by the PA_chr10_G001694 gene with one CHH site in each cohort, and 3) a cytosine-5-DNA methyltransferase encoded by the PA_chr03_G006112 gene with one CpG and one CHH site in the natural cohort only. The recurrence of RdDM machinery components as cis-meQTL targets specifically in the CHH context raises the possibility that local genetic variants directly influence the methylation state of genes encoding the machinery responsible for the establishment of this context.

310 **Supplementary Tables**

311 Supplementary Table S1. Exome capture sequencing (ECS) data processing summary

312 Supplementary Table S2. Targeted enzymatic methylation sequencing (TMS) data summary by
313 cohort and cytosine context

314 Supplementary Table S3. Selected SVMPs with annotation

315 Supplementary Table S4. Genomic-epigenomic distance statistics

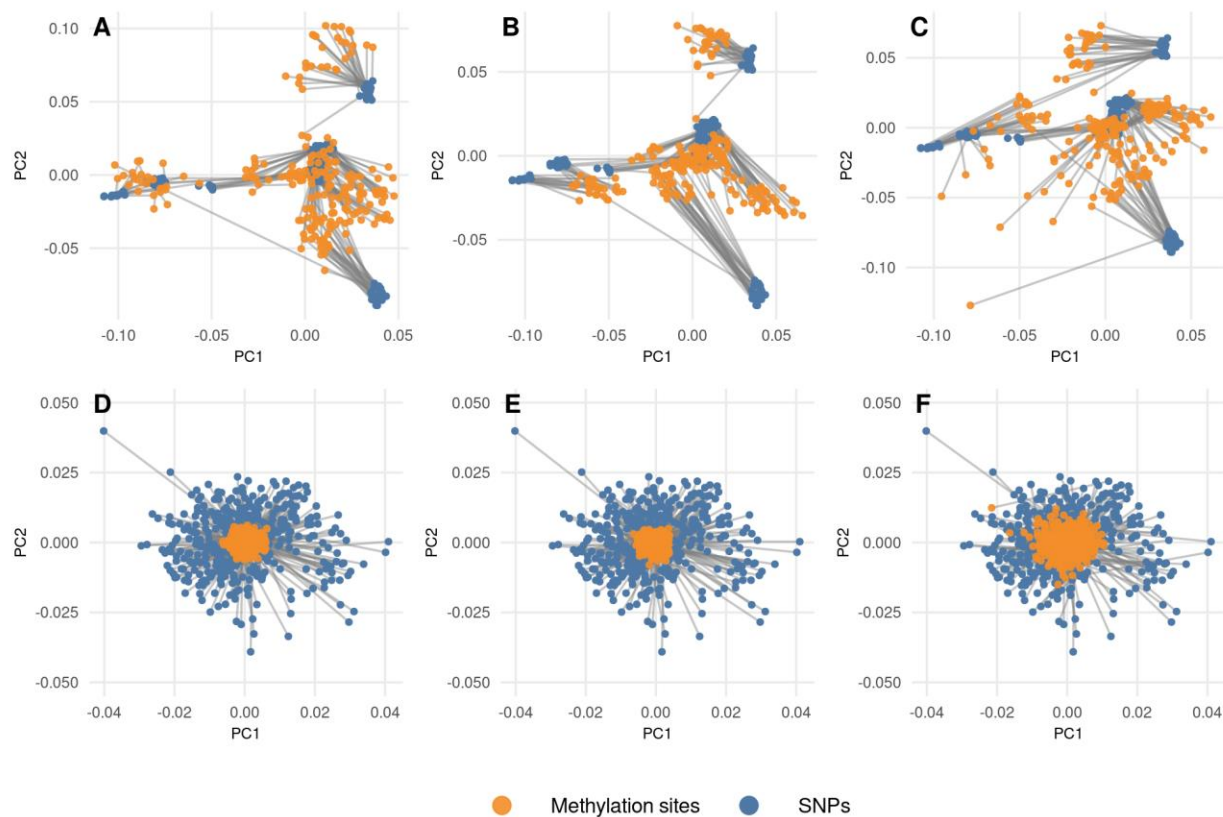
316 Supplementary Table S5. Robust meQTL pairs

317 Supplementary Table S6. Annotated robust SNPs

318 Supplementary Table S7. Annotated robust methylation sites

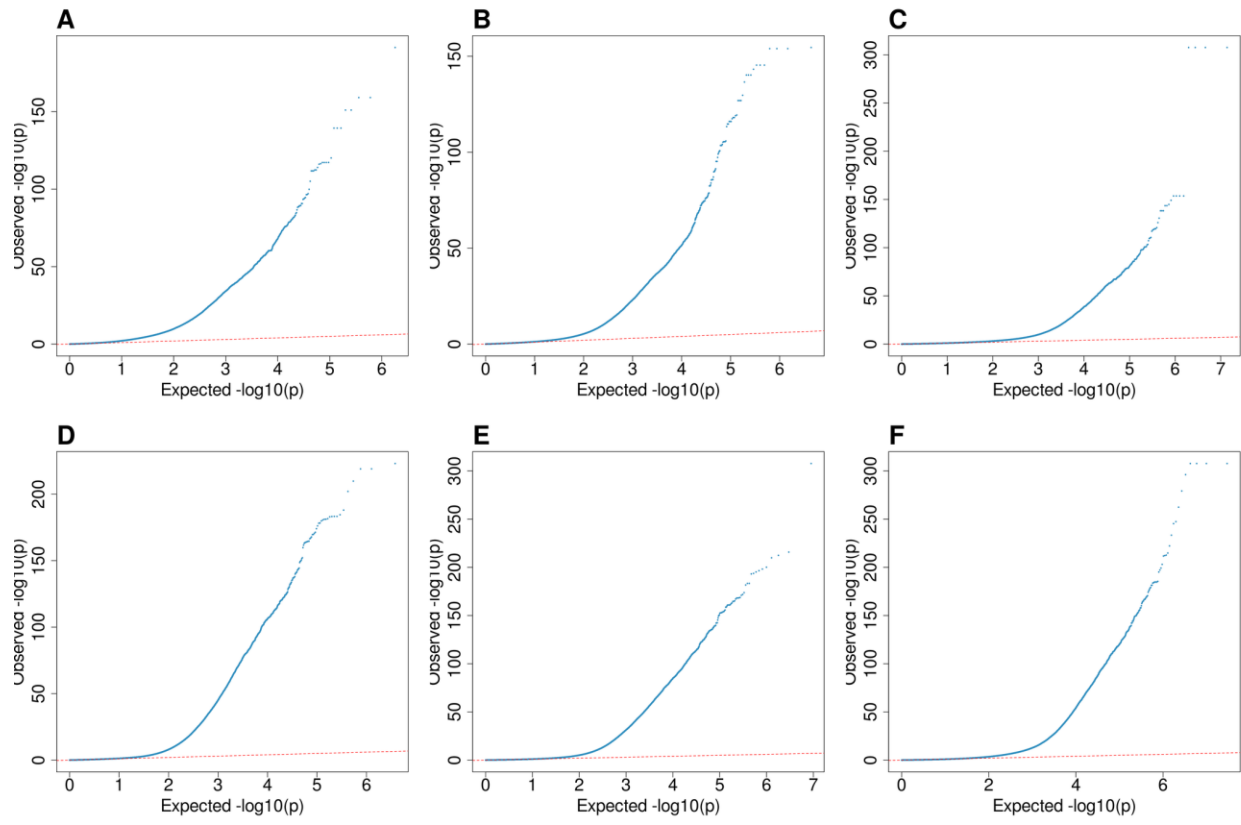
319

Supplementary Figure S1



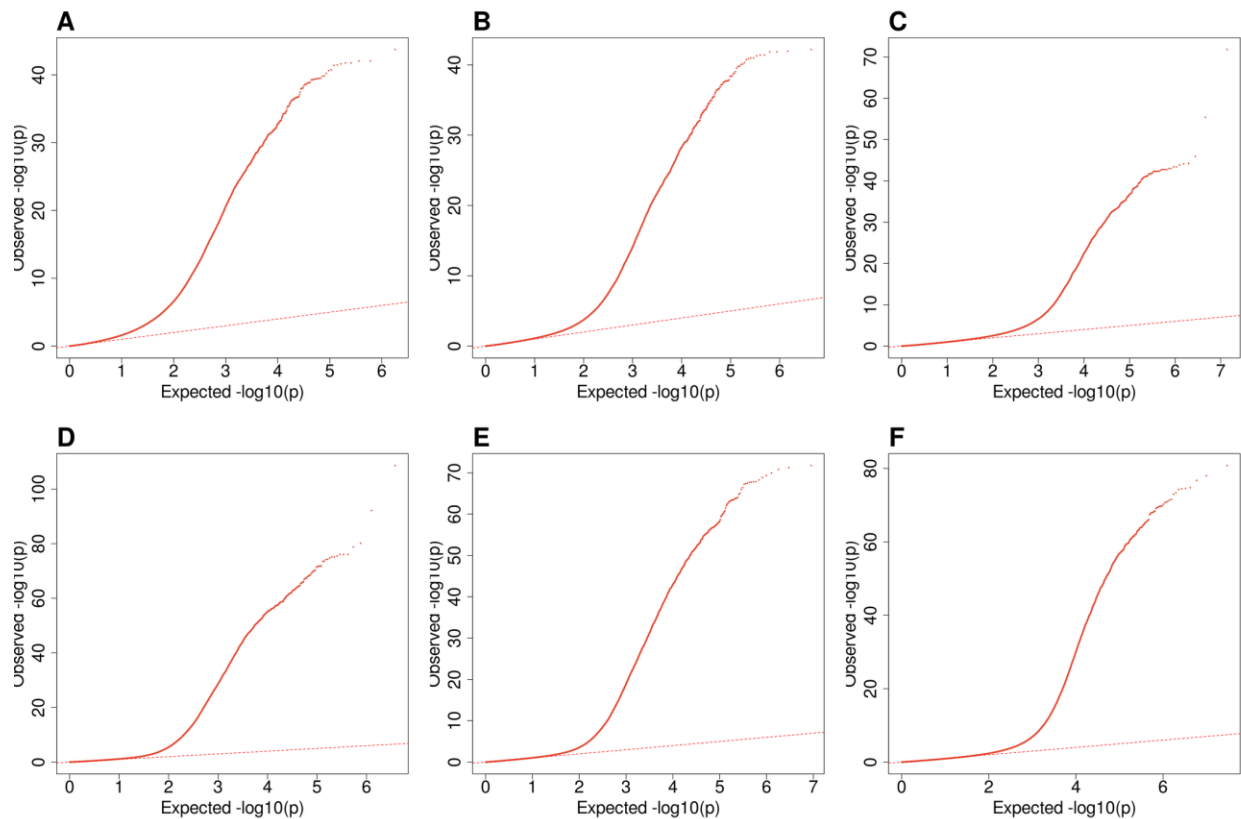
Supplementary Fig. S1 | Procrustes biplots of genomic and epigenomic ordinations. A–C Procrustes biplots for the breeding cohort (n = 209 individuals; 17 full-sib families) for CpG (A), CHG (B), and CHH (C) methylation contexts. D–F Procrustes biplots for the natural cohort (n = 393 individuals; 25 natural stands) for CpG (D), CHG (E), and CHH (F) methylation contexts. Each biplot shows the alignment between PCoA ordinations derived from the SNP-based genomic distance matrix and the methylation-based epigenomic distance matrix after optimal Procrustes rotation. Residual arrows connect paired sample positions in the two ordinations; short arrows indicate high concordance. Procrustes t_o -statistics and permutation p -values (9,999 permutations) are given in Supplementary Table S4. PCoA, principal coordinate analysis; SNP, single-nucleotide polymorphism; CpG, CHG, and CHH, cytosine methylation contexts (H = A, T or C).

332 **Supplementary Figure S2**



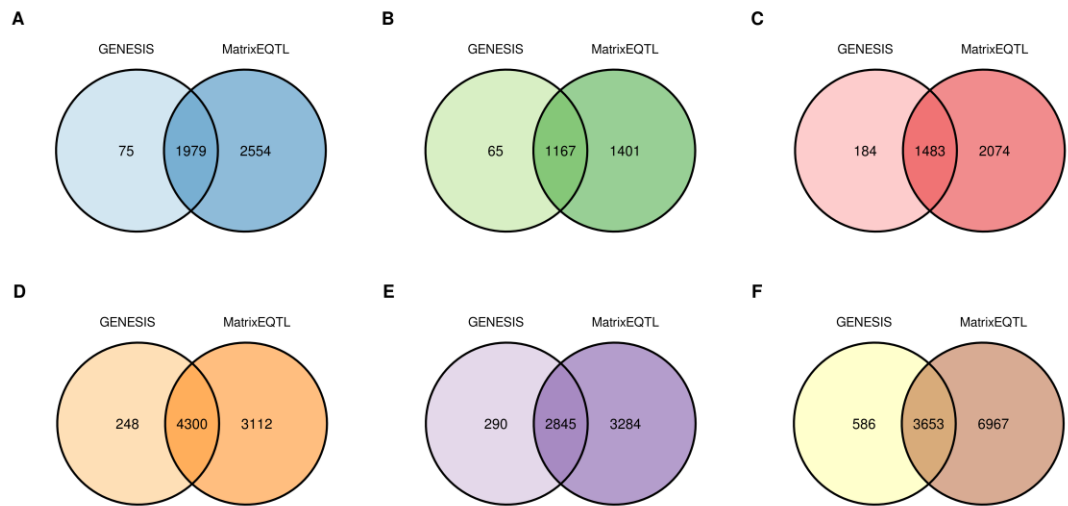
333 **Supplementary Fig. S2 | Quantile–quantile (QQ) plots of MatrixEQTL cis-meQTL p -values.** **A–C** QQ
334 plots for the breeding cohort ($n = 209$ individuals; 17 full-sib families) for CpG (**A**), CHG (**B**), and CHH (**C**)
335 methylation contexts. **D–F** QQ plots for the natural cohort ($n = 393$ individuals; 25 natural stands) for CpG
336 (**D**), CHG (**E**), and CHH (**F**) methylation contexts. Expected $-\log_{10}(p)$ under the null is plotted on the x-
337 axis; observed $-\log_{10}(p)$ on the y-axis. The diagonal line represents the null expectation. Genomic inflation
338 factors (λ) are given in Table 1. CpG, CHG, and CHH, cytosine methylation contexts (H = A, T or C).
339

341 **Supplementary Figure S3**



342 **Supplementary Fig. S3 | Quantile–quantile (QQ) plots of GENESIS cis-meQTL p -values.** A–C QQ
343 plots for the breeding cohort ($n = 209$ individuals; 17 full-sib families) for CpG (A), CHG (B), and CHH (C)
344 methylation contexts. D–F QQ plots for the natural cohort ($n = 393$ individuals; 25 natural stands) for CpG
345 (D), CHG (E), and CHH (F) methylation contexts. Expected $-\log_{10}(p)$ under the null is plotted on the x-
346 axis; observed $-\log_{10}(p)$ on the y-axis. The diagonal line represents the null expectation. The reduced
347 genomic inflation relative to MatrixEQTL (Extended Data Fig. 4; Supplementary Fig. S2) reflects absorption
348 of residual pedigree structure by the GRM random effect. Genomic inflation factors (λ) are annotated in
349 each panel. CpG, CHG, and CHH, cytosine methylation contexts (H = A, T or C).
350

Supplementary Figure S4



Supplementary Fig. S4 | Venn diagrams of SNPs detected by MatrixEQTL and GENESIS. A–C Number of unique SNPs significant at $FDR < 1 \times 10^{-10}$ found only by MatrixEQTL, only by GENESIS, or by both tools for CpG (A), CHG (B), and CHH (C) methylation contexts in the breeding cohort ($n = 209$ individuals; 17 full-sib families). **D–F** Corresponding diagrams for the natural cohort ($n = 393$ individuals; 25 natural stands) for CpG (D), CHG (E), and CHH (F) methylation contexts. The intersection defines the robust SNP set under the pair-level criterion: the same SNP-methylation site pair passed $FDR < 1 \times 10^{-10}$ in both MatrixEQTL and GENESIS independently. Used for all downstream annotation analyses. SNP, single-nucleotide polymorphism; CpG, CHG, and CHH, cytosine methylation contexts (H = A, T or C).