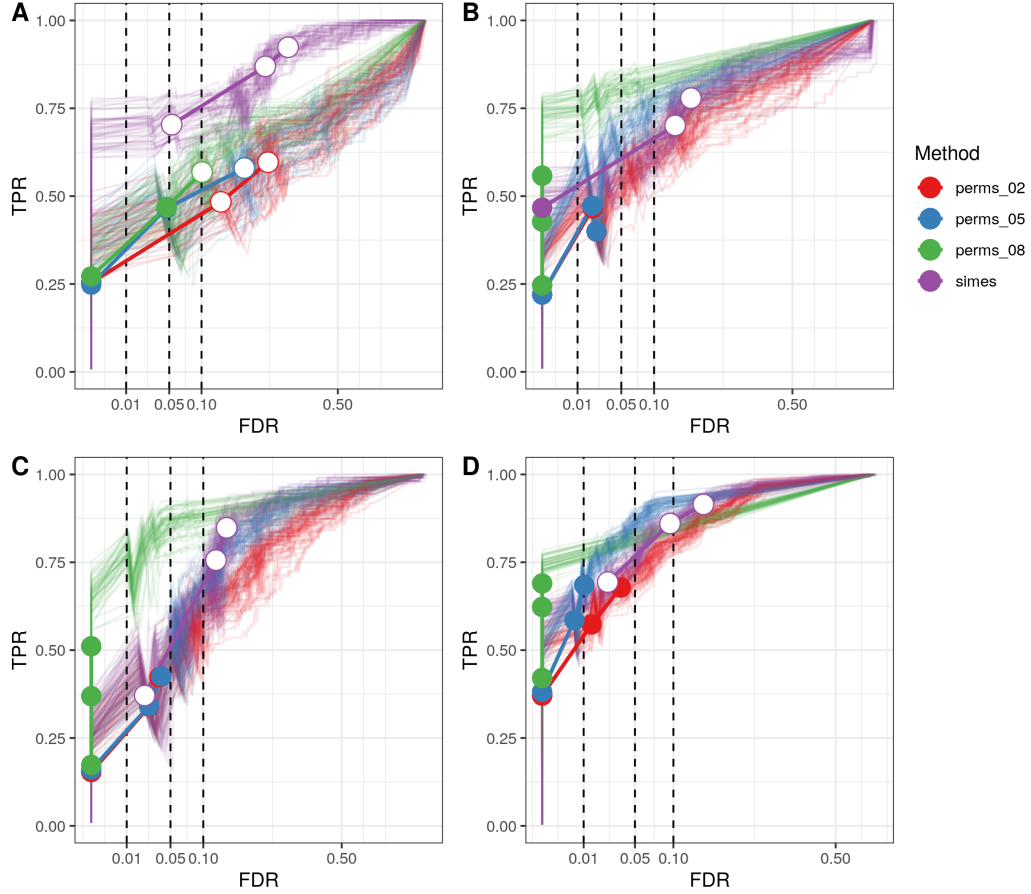


Additional File 1

DAMEfinder: A Method to Detect Differential Allele-Specific
Methylation

Stephany Orjuela, Dania Machlab, Mirco Menigatti,
Giancarlo Marra, Mark D Robinson

May 28, 2020



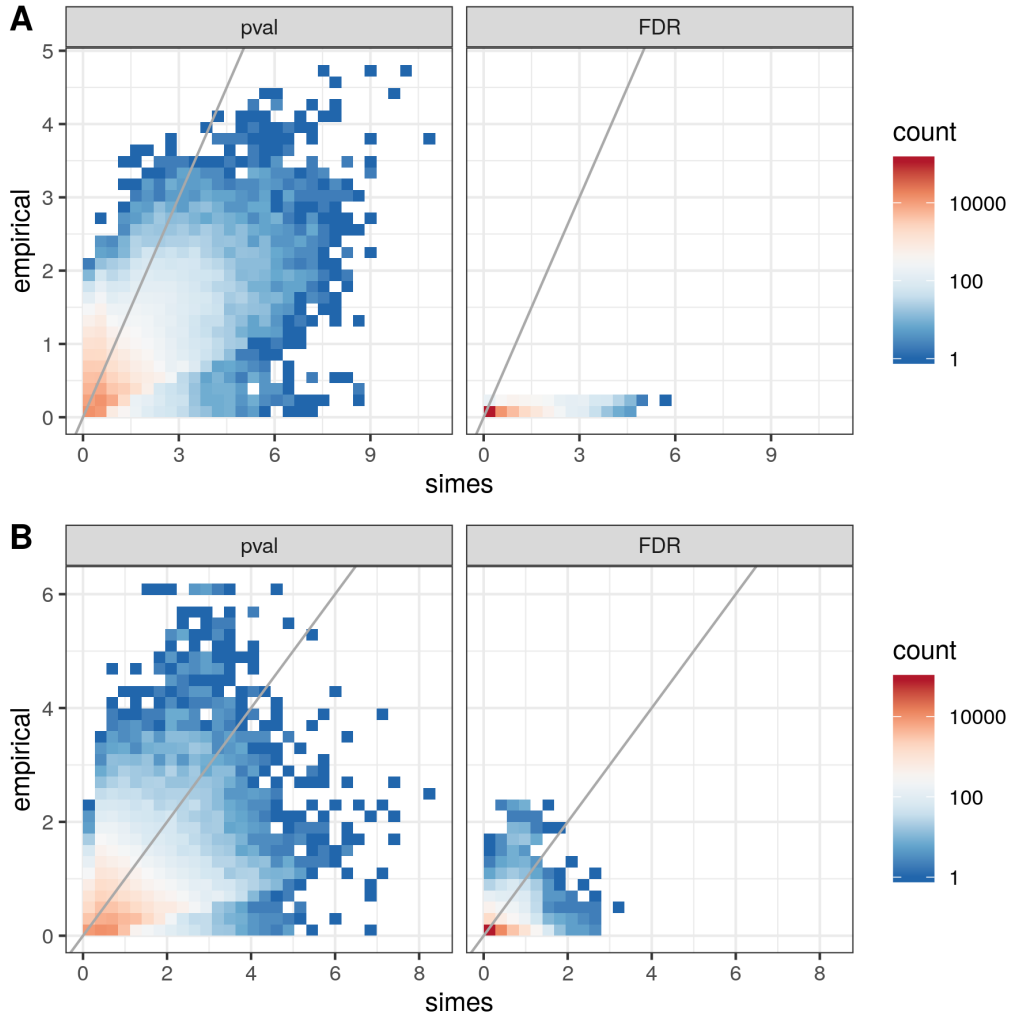


Figure S2: Permutation and simes p-values assigned to CIMP (A) and non-CIMP (B) DAMEs. Left panels: p-values assigned to DAMEs using the permutation (empirical) method in the y-axis, and the cluster-wise correction (simes) in the x-axis. Right panels: Adjusted p-values using the Benjamini and Hochberg method. All axes are negative-log transformed.

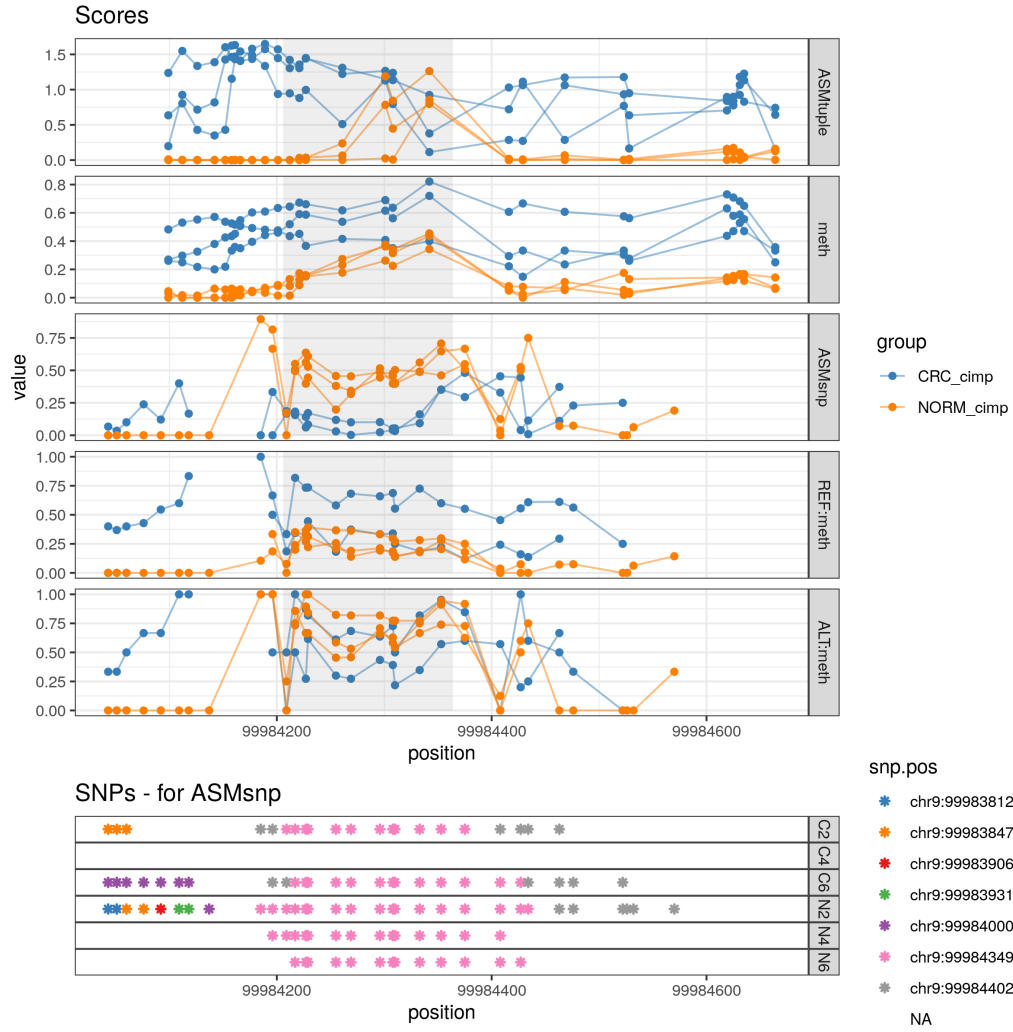


Figure S3: Increase in ASM_{tuple} score, decrease in ASM_{snp} score. DAME located in chr9:99,984,206-99,984,364. ASM_{tuple} shows gain of ASM in CRC samples, whereas ASM_{snp} track shows loss of ASM in CRC samples.

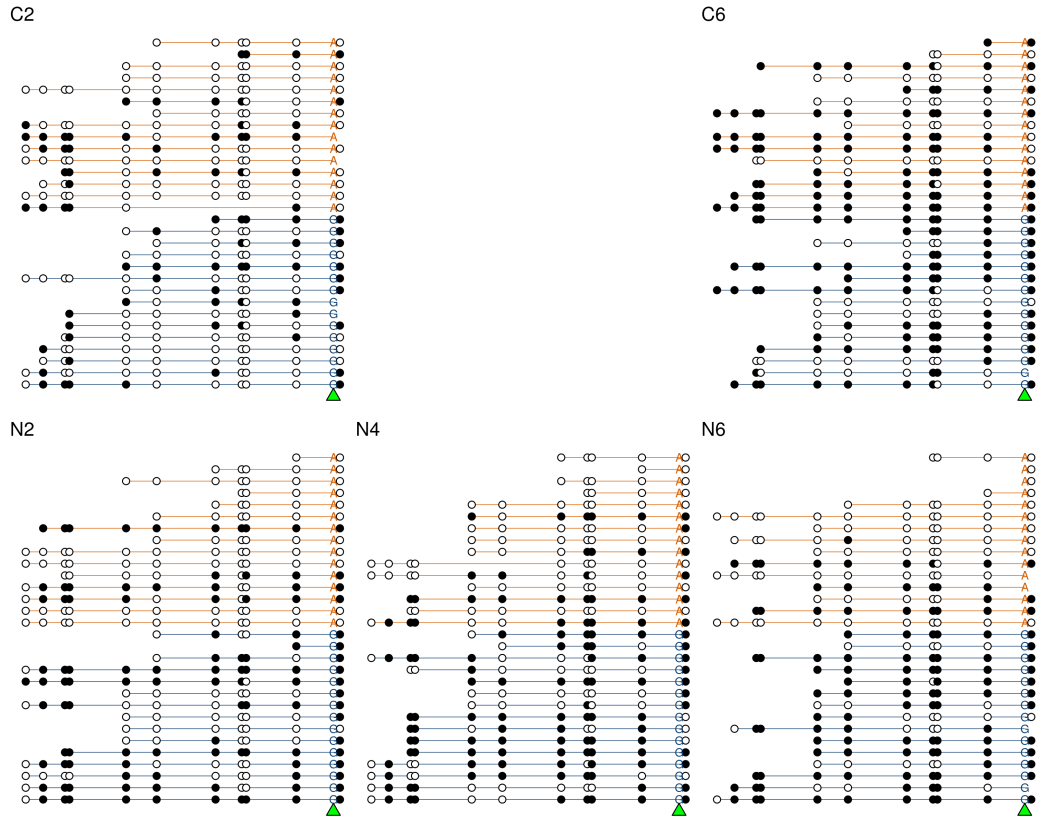


Figure S4: Reads of DAME located in chr9:99,984,206-99,984,364 (from figure above 3) Reads sorted by allele show a different pattern than looking at unsorted reads (ignoring the allele). Figures labeled from C2 to N6 show the reads (horizontal lines colored by allele) overlapping the SNP (green triangle pointing to alleles) used to calculate this DAME from the ASM_{snp} score. Samples with C are CRCs, samples with N are normal tissue. White circles are unmethylated CpG sites, black circles are methylated CpG sites. A random sample of 15 reads is shown for each allele.

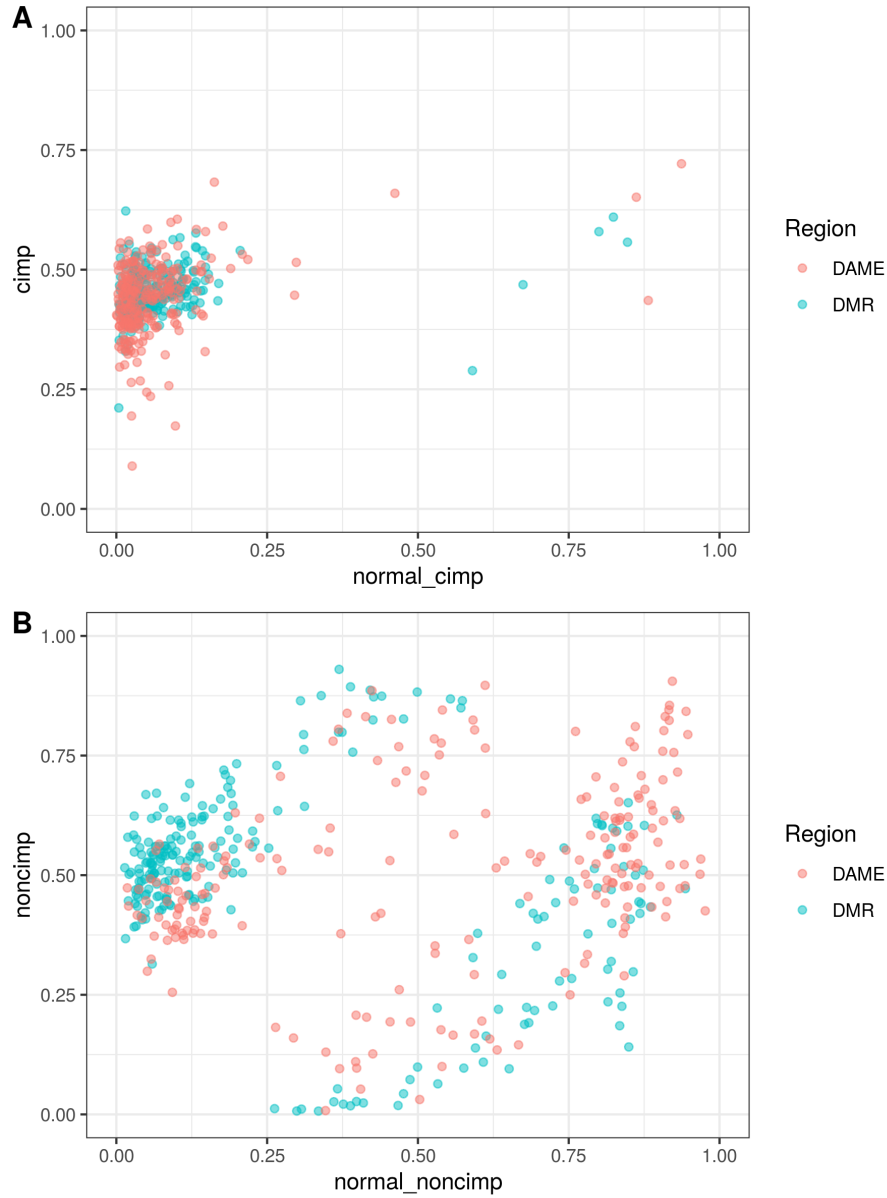


Figure S5: Average methylation in top 250 DAMEs and DMRs. x-axis: Average methylation in normal tissue. y-axis: Average methylation in CRC tissue (average is across samples per group and then per region). **A.** CIMP CRCs **B.** non-CIMP CRCs.

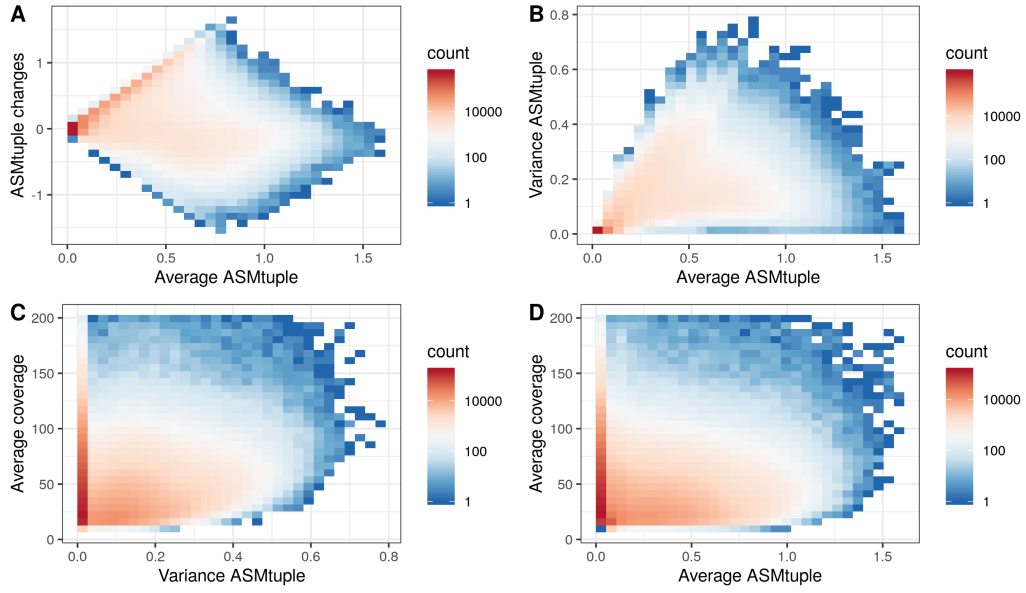


Figure S6: ASM_{tuple} score diagnostics in the CRC dataset. **A.** mean-difference plot, differences between all normals and all CRCs. **B.** mean-variance plot. **C.** Variance (x-axis) and mean coverage (y-axis). **D.** Mean of score (x-axis) and mean coverage (y-axis). Variance and mean do not depend on coverage. Each point corresponds to a CpG tuple.

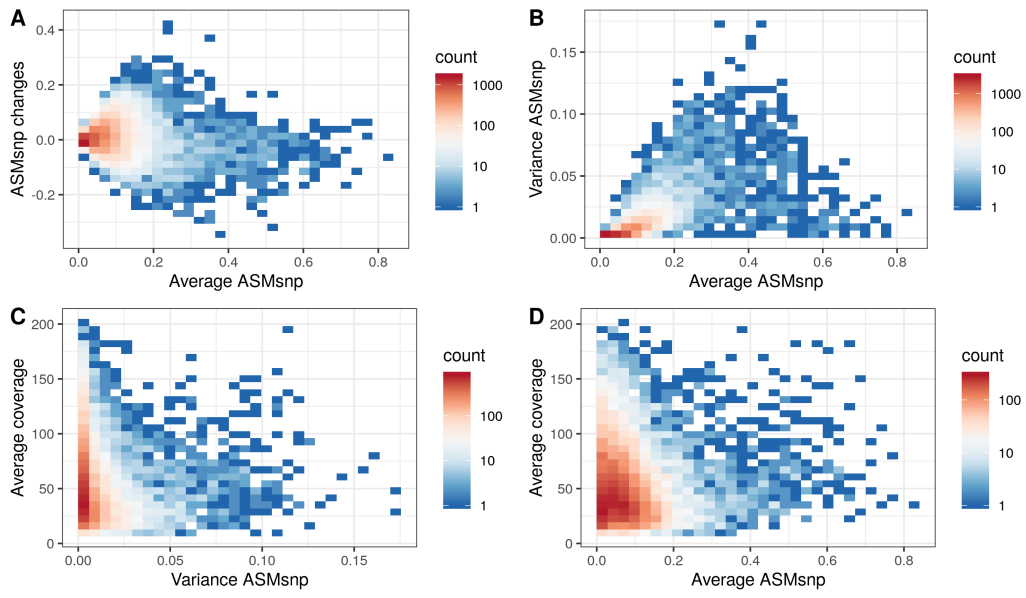


Figure S7: ASM_{snp} score diagnostics in the CRC dataset. **A.** Mean-difference plot, differences between all normals and all CRCs. **B.** Mean-variance plot. **C.** Variance (x-axis) and mean coverage (y-axis). **D.** Mean (x-axis) and mean coverage (y-axis). Each point corresponds to a CpG site.

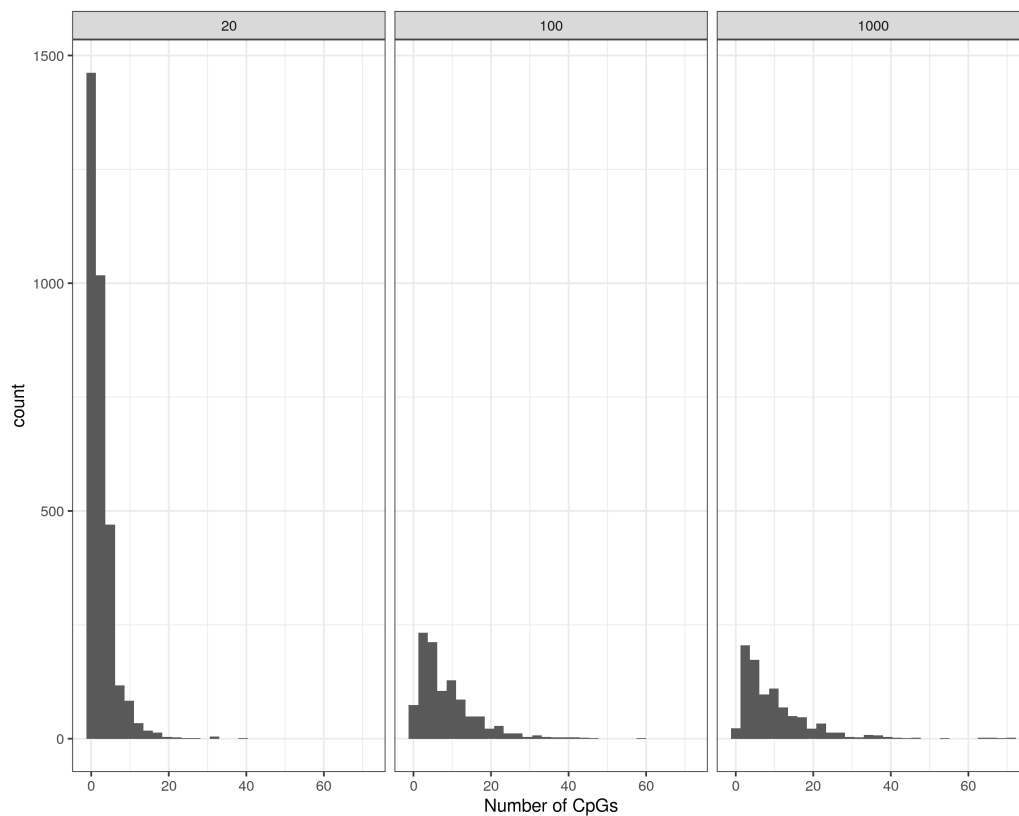


Figure S8: Sizes of generated clusters (regions) for simulation. x-axis is the number of CpG sites per generated cluster. Simulation in main text was based on $\text{maxGap} = 100$ (middle facet). Other values for maxGap were used for Supp. Fig. 1

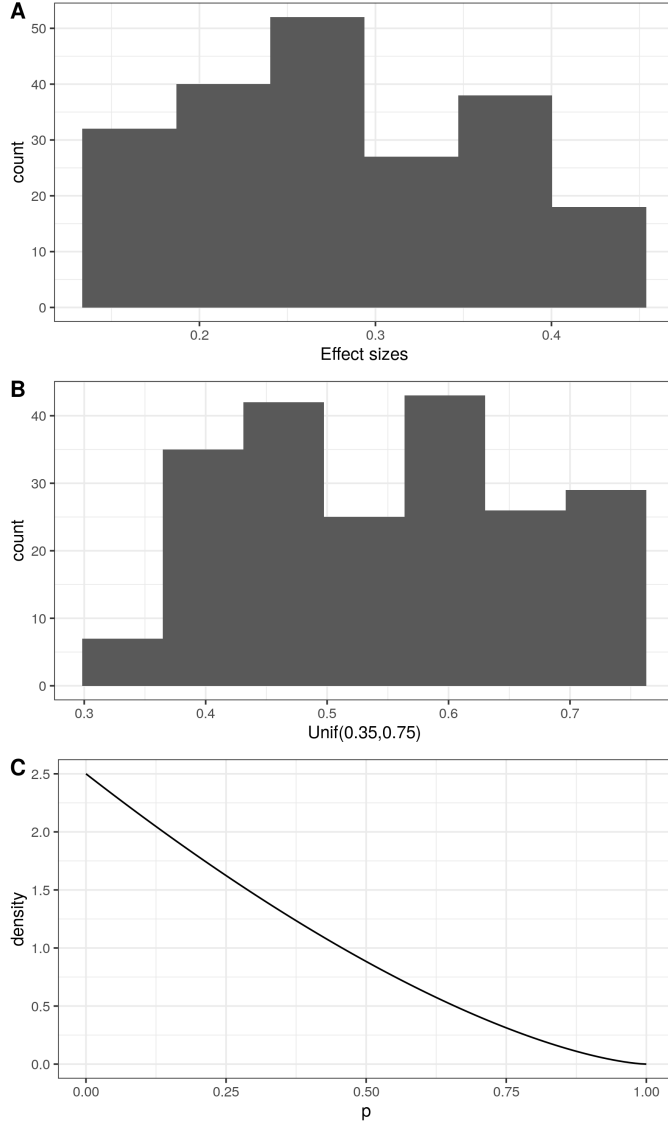


Figure S9: Simulation parameters for effect size. **A.** Real effect sizes generated for 1 of the 50 simulations with inverse transform sampling of the form $F_X^{-1}(u) = x$, where $u \sim Unif(0.35, 0.75)$, and $F_X(x)$ is the CDF of $Beta(1, 2.5)$. **B.** 207 random uniform values from $Unif(0.35, 0.75)$. **C.** Beta distribution PDF with $\alpha = 1$ and $\beta = 2.5$.

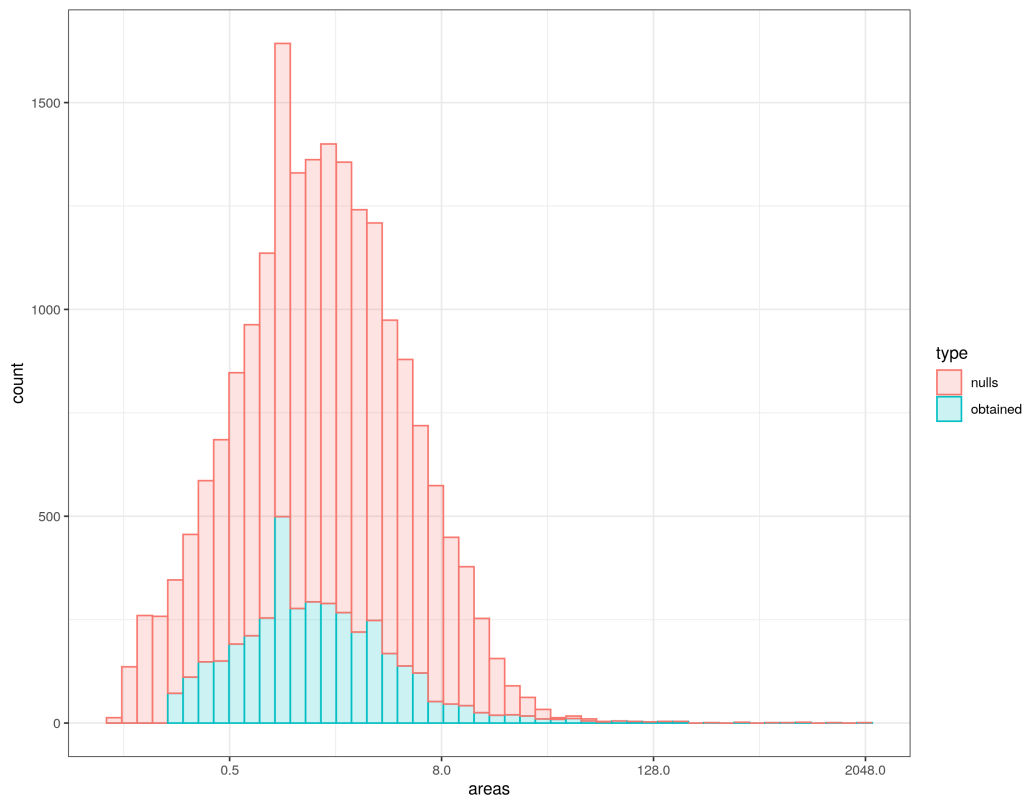


Figure S10: Distribution of areas obtained from one of the 50 simulations by running the permutation method with $K = 0.2$. X-axis log2 transformed.

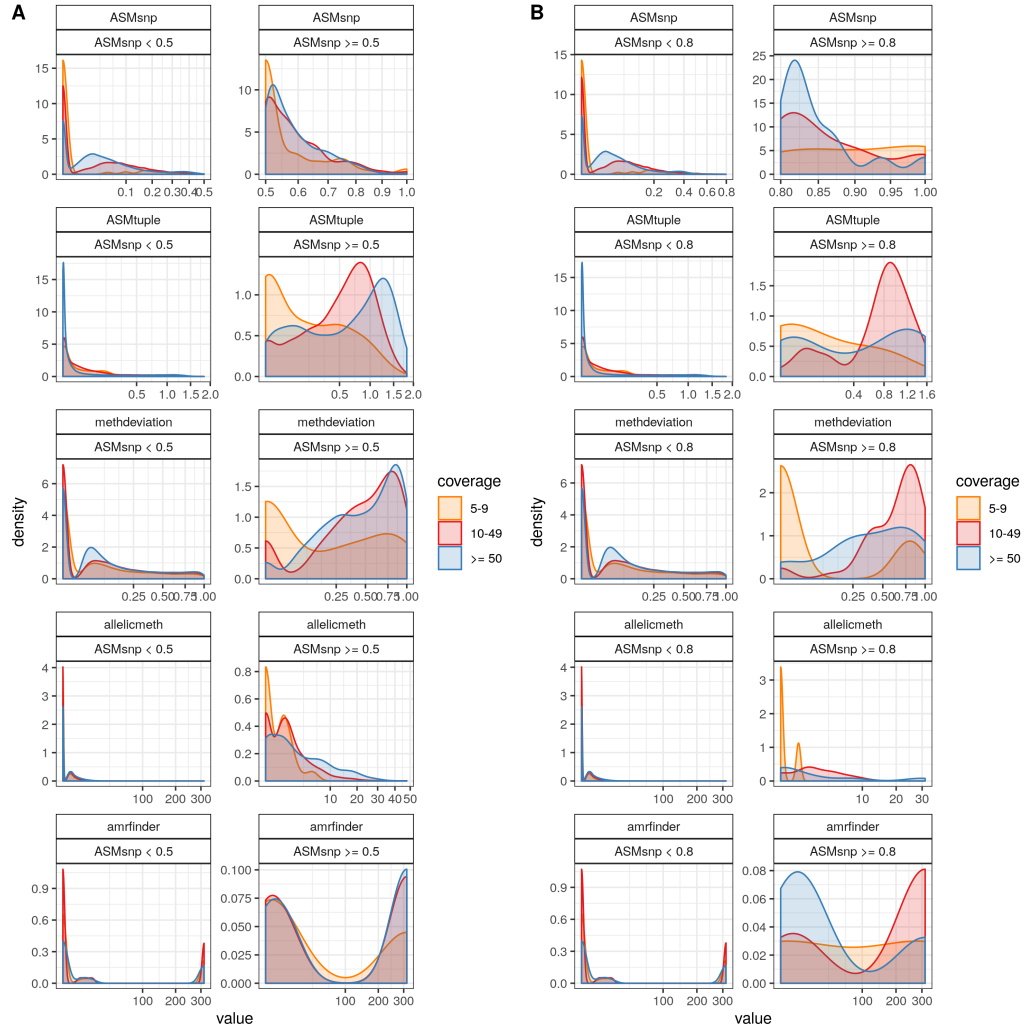


Figure S11: Distributions of evaluated scores included in main Figure 2 ($ASMTuple$, $methdeviation$, $allelicmeth$, $amrfinder$), and the ASM_{snp} score, considered as the truth. Colors correspond to coverage groups (5 to 9, 10 to 49, ≥ 50). **A.** Values when $ASM_{snp} < 0.5$ (sizes 5-9: 5909, 10-49: 108618, ≥ 50 : 166286), and ≥ 0.5 (5-9: 67, 10-49: 591, ≥ 50 : 626). **B.** Values when $ASM_{snp} < 0.8$ (5-9: 5972, 10-49: 109169, ≥ 50 : 166887), and ≥ 0.8 (5-9: 4, 10-49: 40, ≥ 50 : 25). X-axis is square-root transformed.

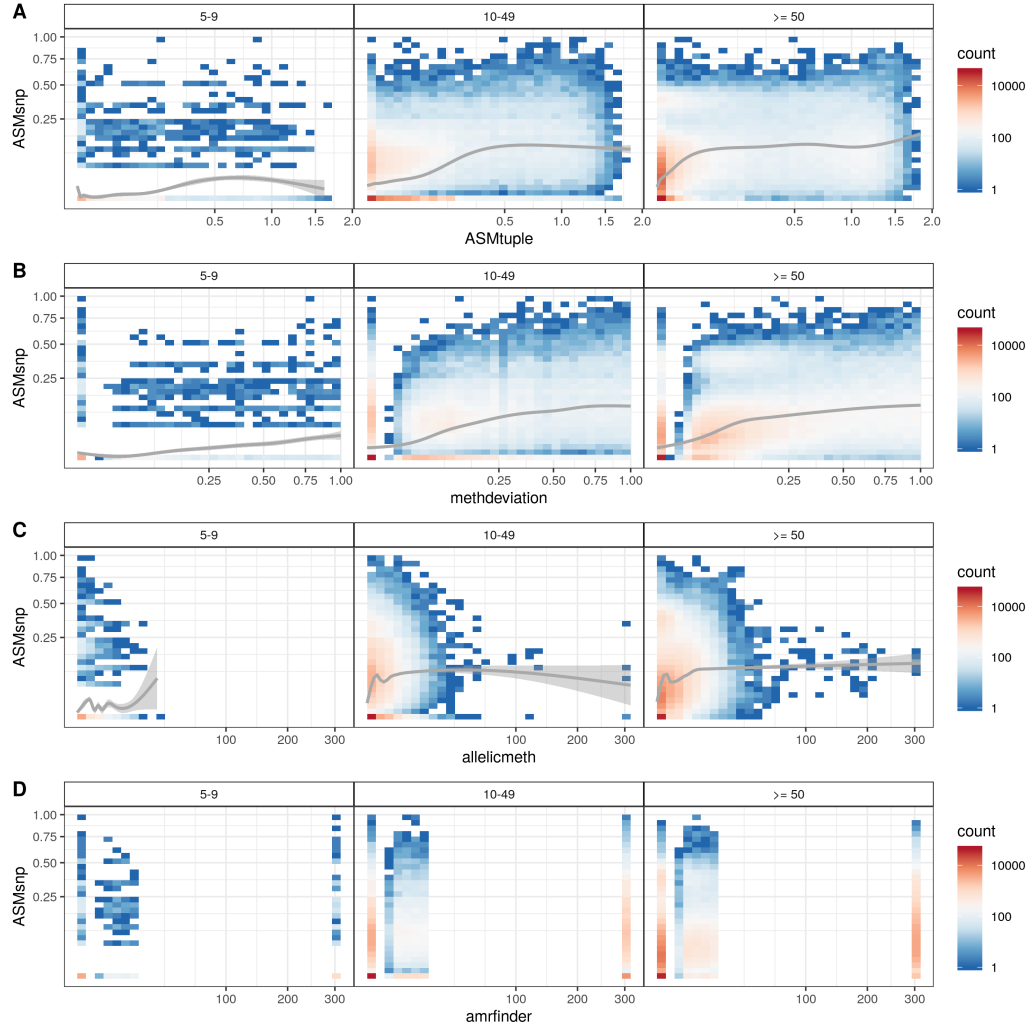


Figure S12: Scatter plots of evaluated scores included in main Figure 2. X-axis: **A.** ASM_{tuple} , **B.** $methdeviation$, **C.** $allelicmeth$, **D.** $amrfinder$. Y-axis: ASM_{snp} score in all plots. Facets correspond to coverage groups (5 to 9, 10 to 49, ≥ 50). X and Y-axis are square-root transformed.

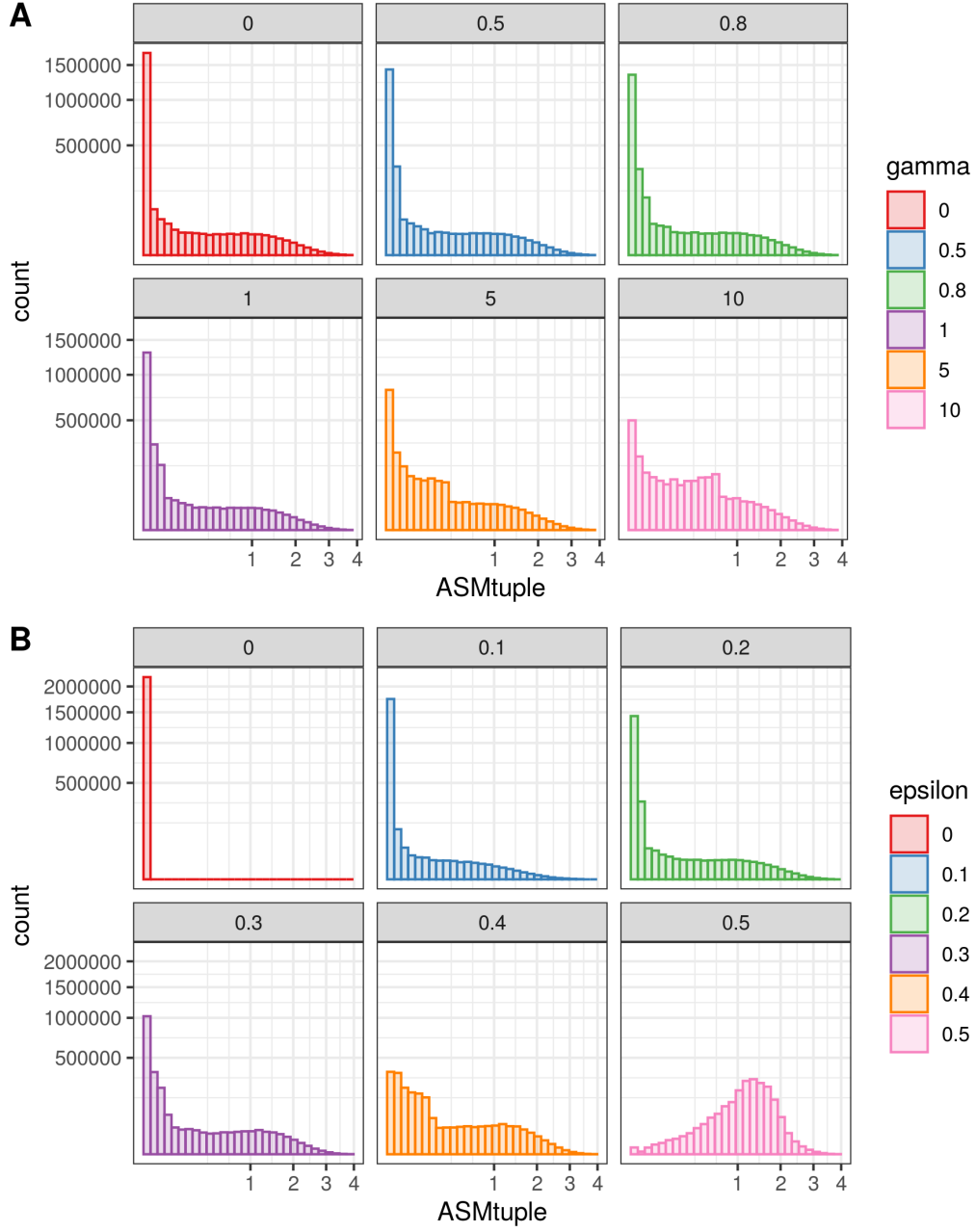


Figure S13: Histograms of the ASM_{tuple} score with different γ (A) and ϵ (B) values, from Equation 3. X and Y-axis are square-root transformed.