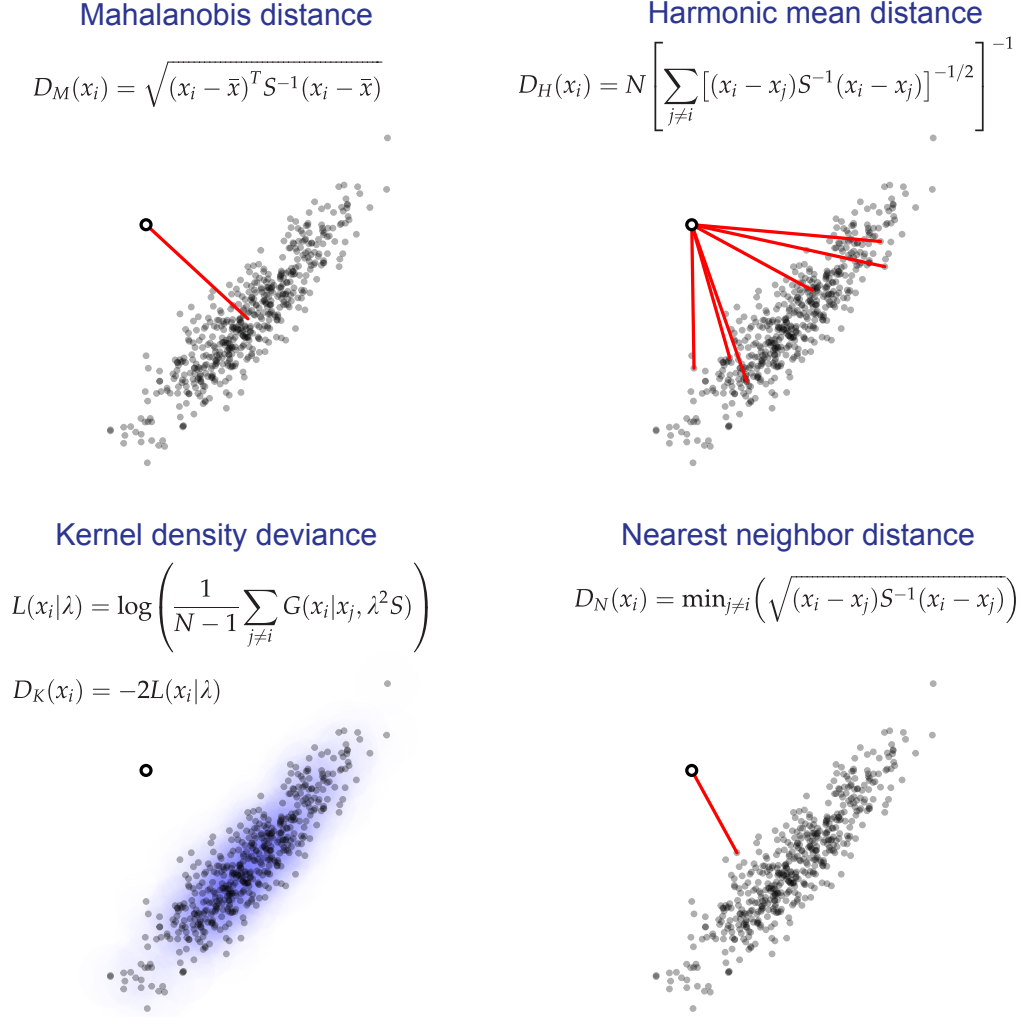# Supplementary Figures

Lotterhos et al. 2017. (Add Title Here). Methods in Ecology and Evolution.

Mahalanobis distance

$$D_M(x_i) = \sqrt{(x_i - \bar{x})^T S^{-1}(x_i - \bar{x})}$$

Harmonic mean distance

$$D_H(x_i) = N \left[\sum_{j \neq i} \left[(x_i - x_j)S^{-1}(x_i - x_j) \right]^{-1/2} \right]^{-1}$$

Kernel density deviance

$$L(x_i|\lambda) = \log \left(\frac{1}{N-1} \sum_{j \neq i} G(x_i|x_j, \lambda^2 S) \right)$$

$$D_K(x_i) = -2L(x_i|\lambda)$$

Nearest neighbor distance

$$D_N(x_i) = \min_{j \neq i} \left(\sqrt{(x_i - x_j)S^{-1}(x_i - x_j)} \right)$$

SUPPLEMENTARY FIGURE 1. Conceptual examples of the multivariate distance measures described in Verity et al. 2017 compared in this study (kernel density deviance was not included due to slow run times).
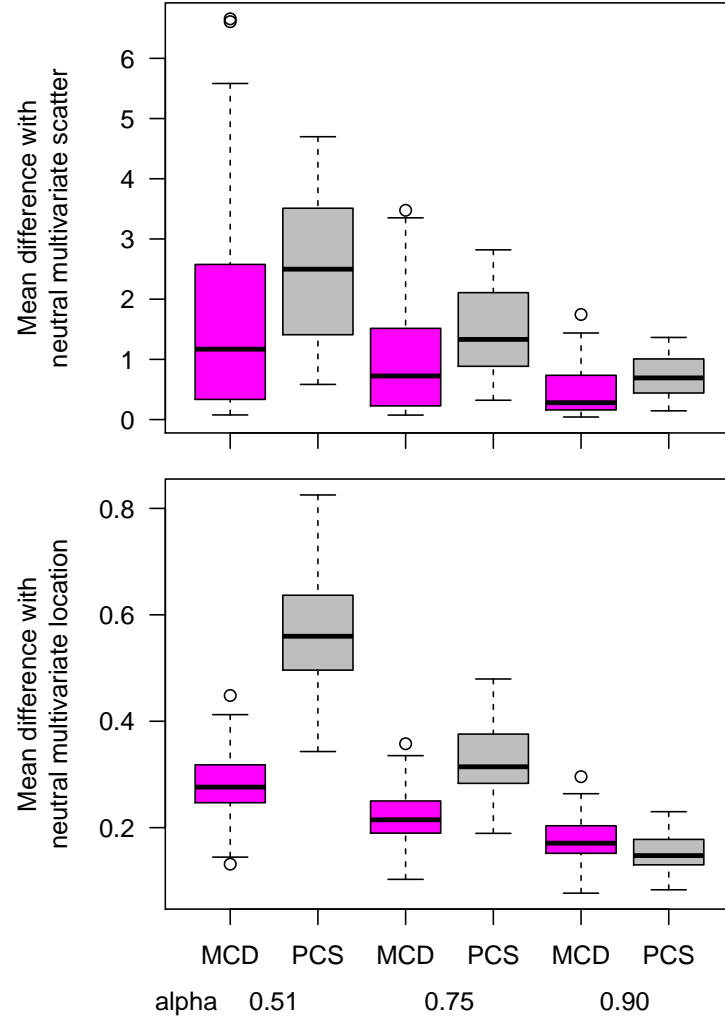
SUPPLEMENTARY FIGURE 2. Difference between the multivariate location and scatter estimated by the robust method and the true neutral estimate. MCD: minimum covariance determinant; PCS: projection congruent subset.
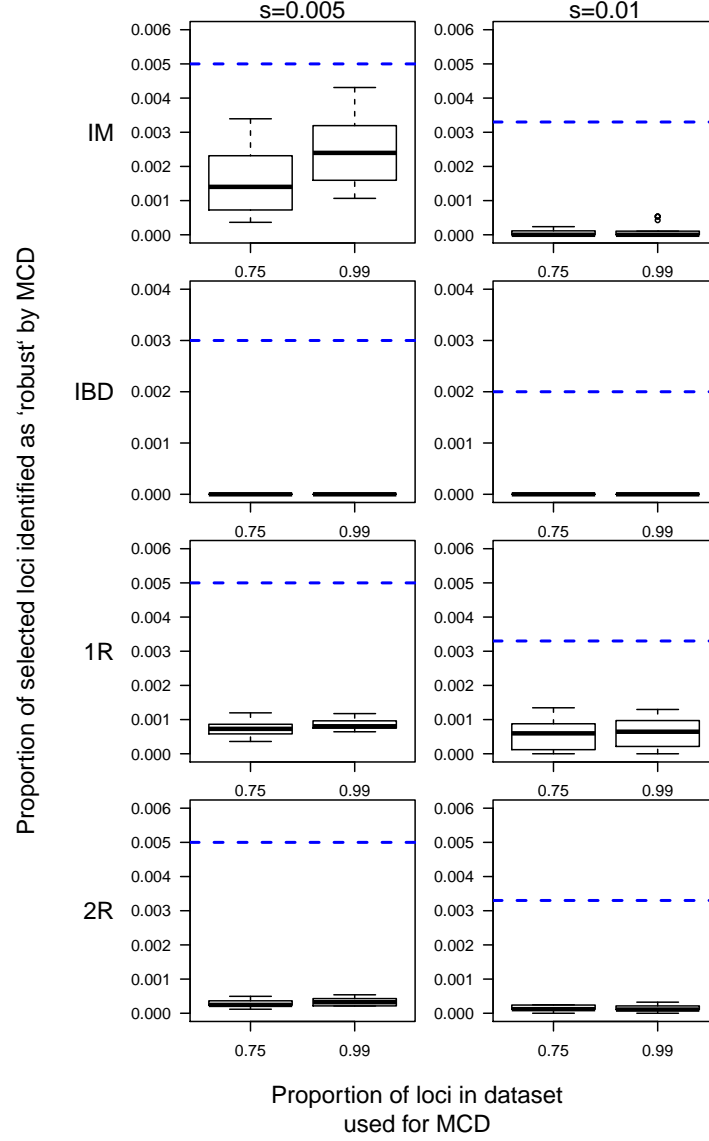
SUPPLEMENTARY FIGURE 3. Evaluation of whether loci affected by selection were identified as 'robust' points (e.g., not outliers) by the MCD algorithm. Loci under relatively weak selection (s=0.005) are in the left column, well loci under moderate selection are in the right column (s=0.01). The horizontal dashed line represents the null expectation, based on the proportion of loci under that strength of selection in the entire dataset. The MCD never identified loci simulated under strong selection (s=0.1) as robust (expected = 0.001 in IBD and 0.0017 in IM, 1R, and 2R; observed = 0 across all simulations).
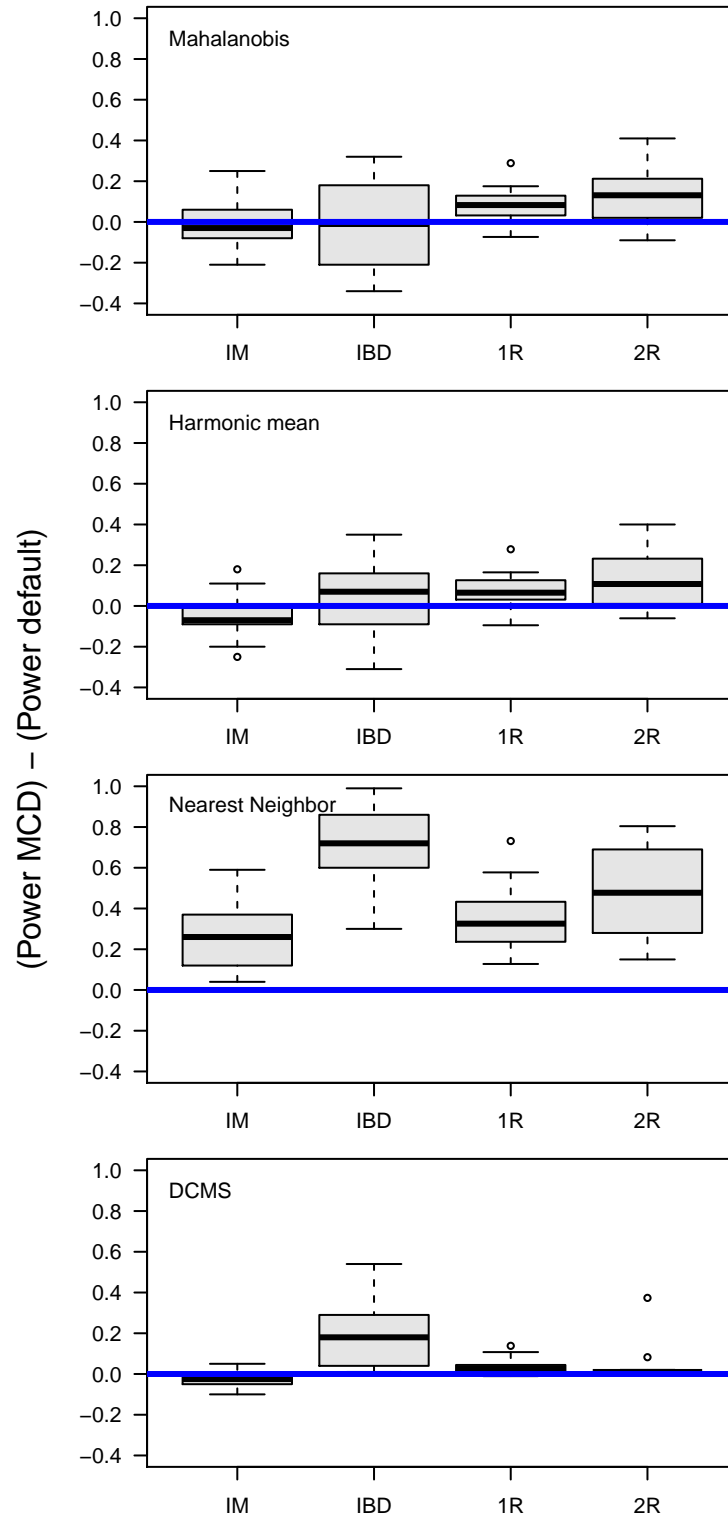
SUPPLEMENTARY FIGURE 4. The difference in empirical power of the statistic to detect selection with and without the robust points identified by minimum covariance determinant (MCD) used in the calculation for the four demographies: island model (IM), isolation by distance (IBD), range expansion from one refuge (1R), range expansion from two refugia (2R).
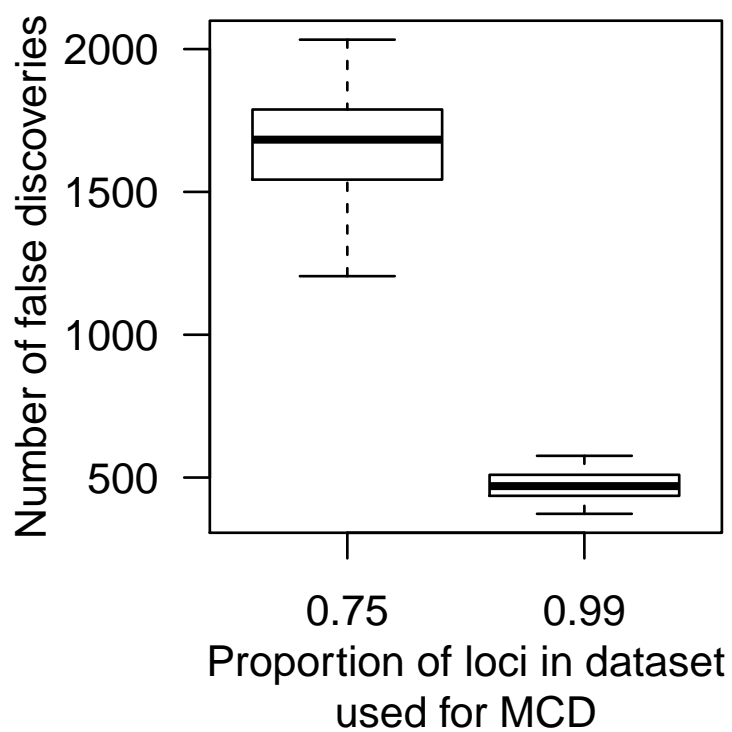
SUPPLEMENTARY FIGURE 5. The number of false discoveries (neutral loci inferred to be outliers, out of 9900 total neutral loci in the dataset) resulting if robust points identified by the MCD are used as an empirical null distribution.
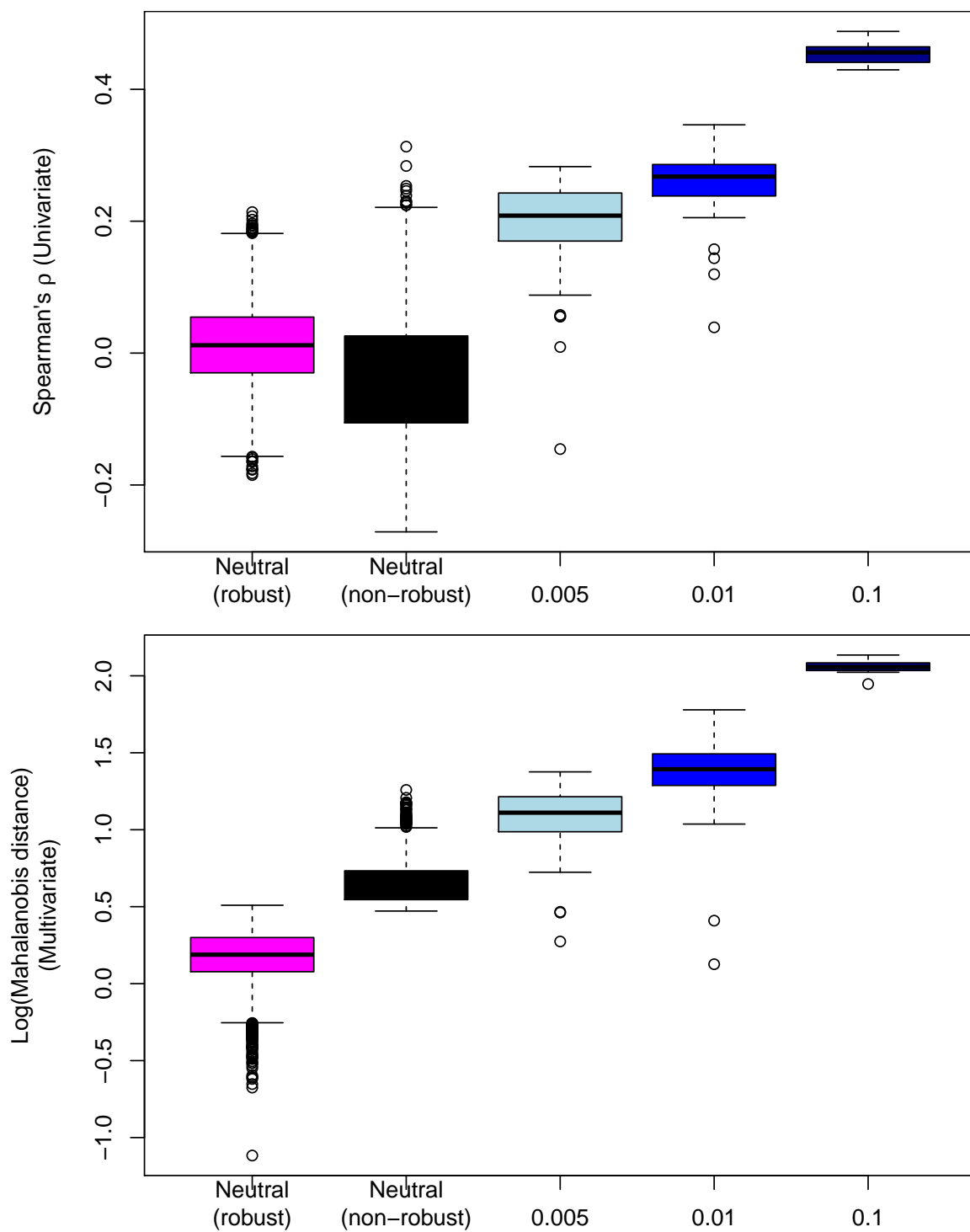
SUPPLEMENTARY FIGURE 6. Average signals as a function of type of locus for a univariate statistic (top) and a multivariate distance (bottom).
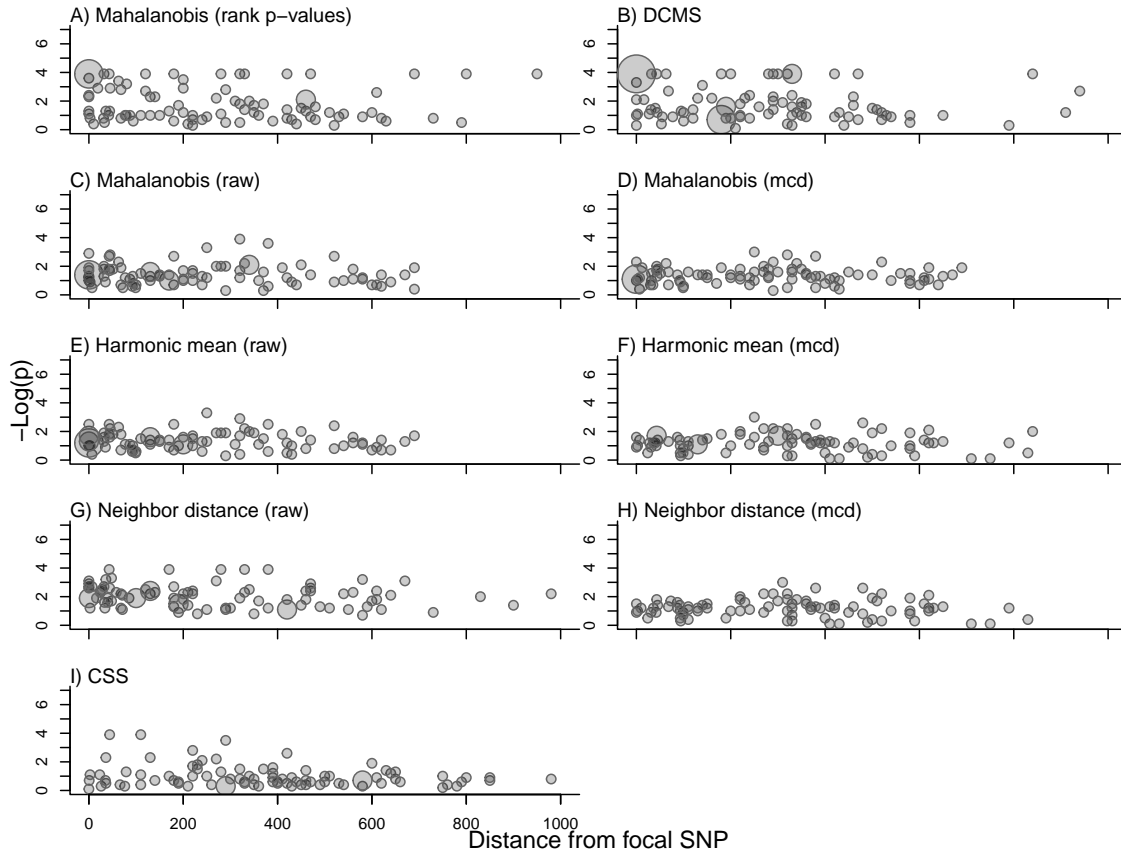
SUPPLEMENTARY FIGURE 7. Comparison among the compound measures for distance in the maximum of signal from the focal SNP versus the significance of the maximum signal, across all dog breeds and loci.. Upper left corner of the plot is higher performance.
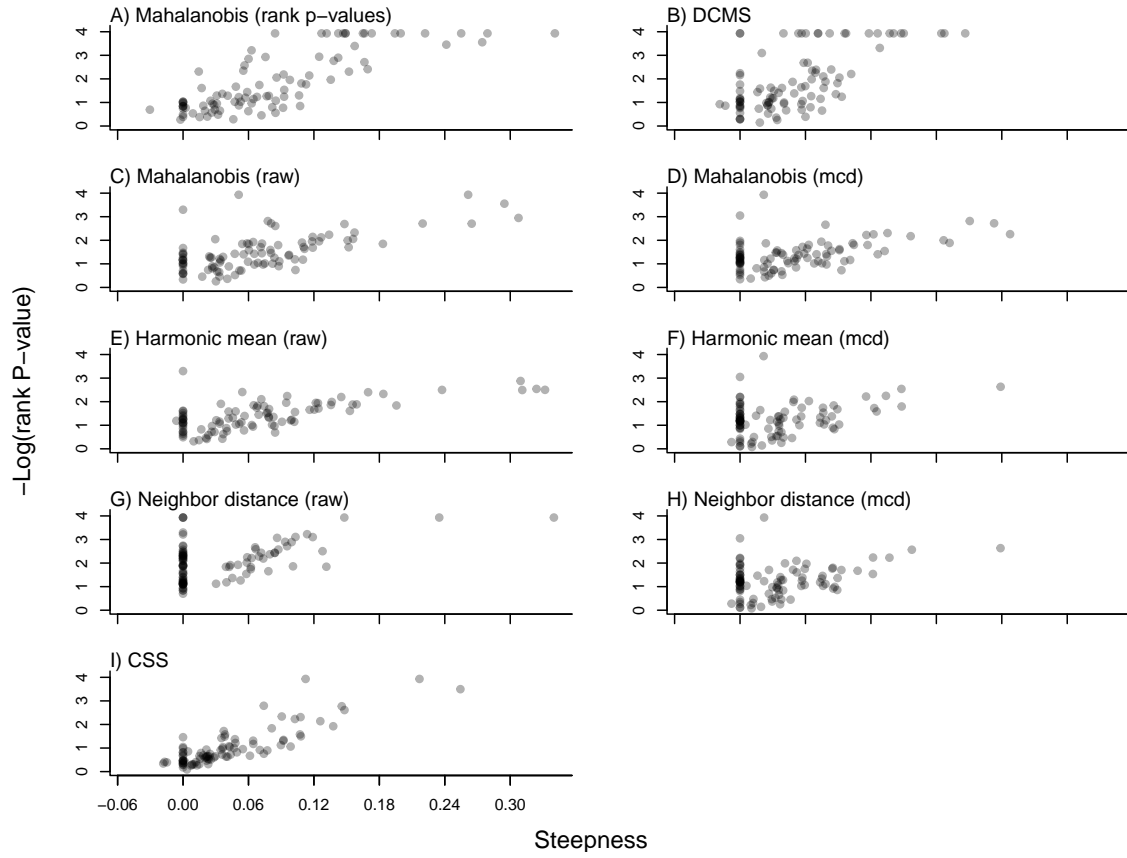
SUPPLEMENTARY FIGURE 8. Comparison among the compound measures for steepness in the maximum of signal versus the significance of the maximum signal, across all dog breeds and loci.. Upper right corner of the plot is higher performance.
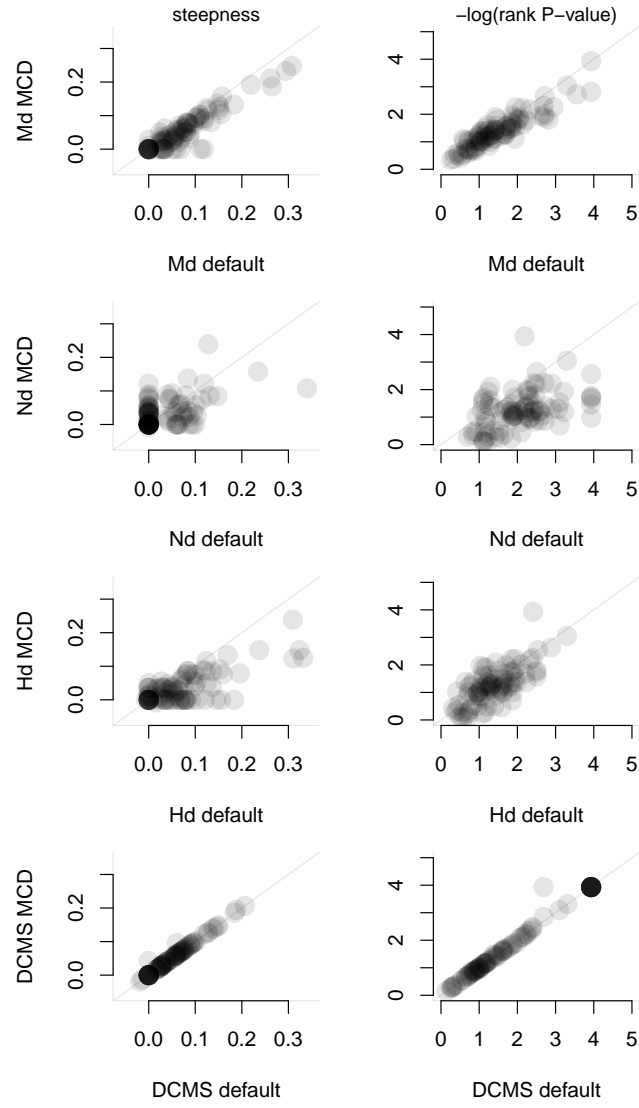
SUPPLEMENTARY FIGURE 9. Comparison between any compound measure with versus without the MCD, for steepness in the left column and minus-log ranked *P*-value in the right column, across all dog breeds and loci.
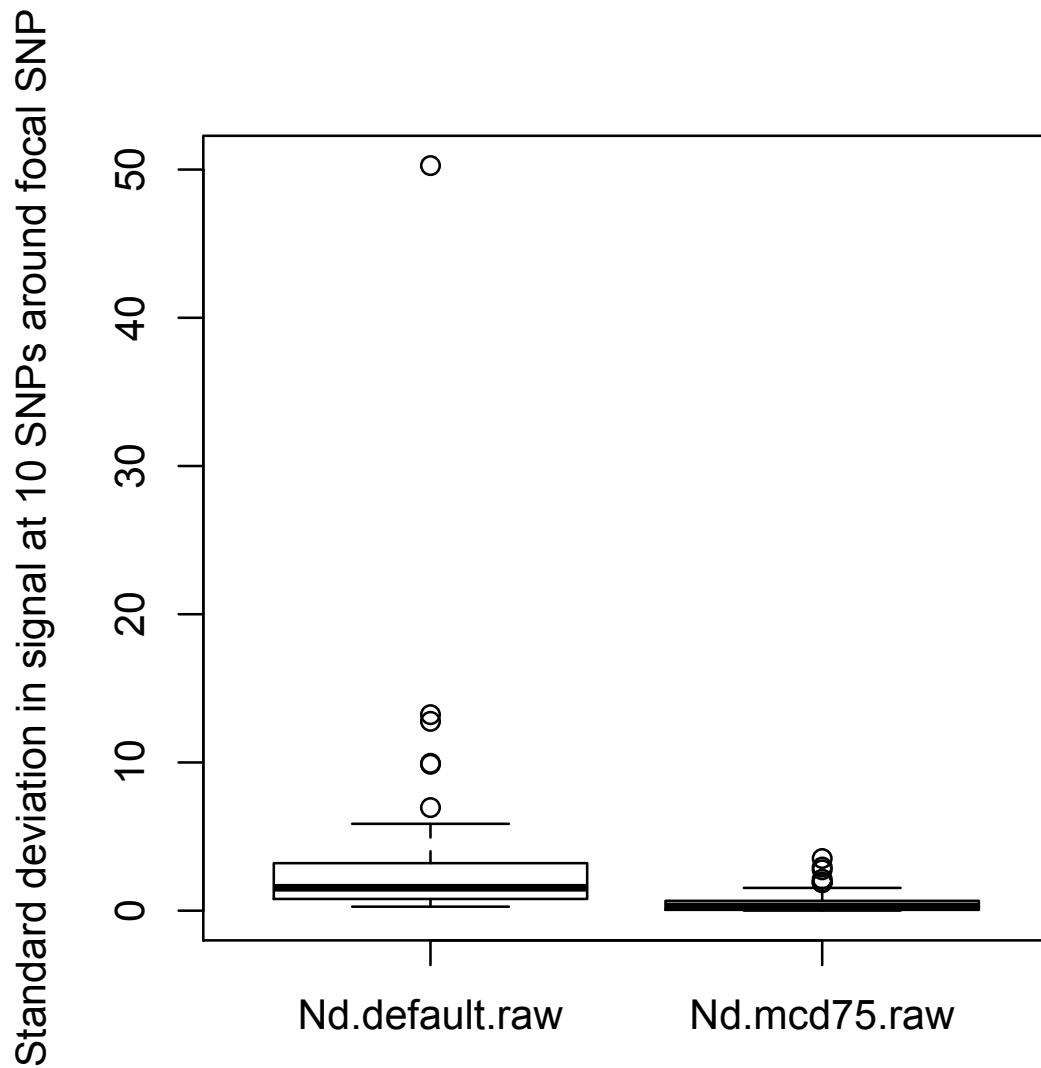
SUPPLEMENTARY FIGURE 10. Boxplot showing standard deviation in the signal at 10 SNPs around the max SNP, across all dog breeds and loci. The default nearest-neighbor distance (Nd.default.raw) is compared to the Nd with the MCD (Nd.mcd75.raw), showing higher variance in the former because of spurious signals.
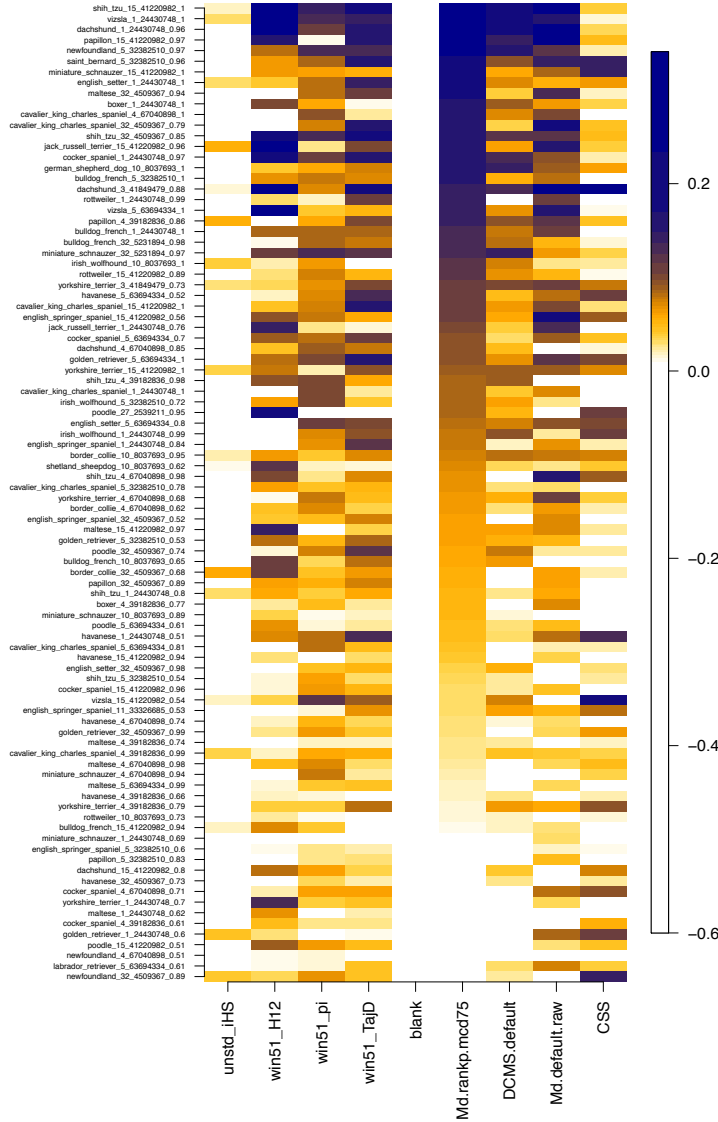
SUPPLEMENTARY FIGURE 11. Heatmap of steepness of the selection signal evaluated for each individual locus in the empirical dog dataset. Zero or negative values of steepness indicate the signal is absent or in the wrong direction. The univariate statistics include iHS, H12, nucleotide diversity π, and Tajima's D. The compound measures include: the Mahalanobis distance based on minus-log fractional P-values (Md-rank-P), the decorrelated composite of multiple signals (DCMS), the Mahalanobis distance based on the raw statistics, and the composite signal of selection (CSS). The row labels indicate the dog breed, chromosome, base pair of the focal SNP, and the allele frequency of the focal SNP.
TO DO: EDIT X-AXIS LABELS
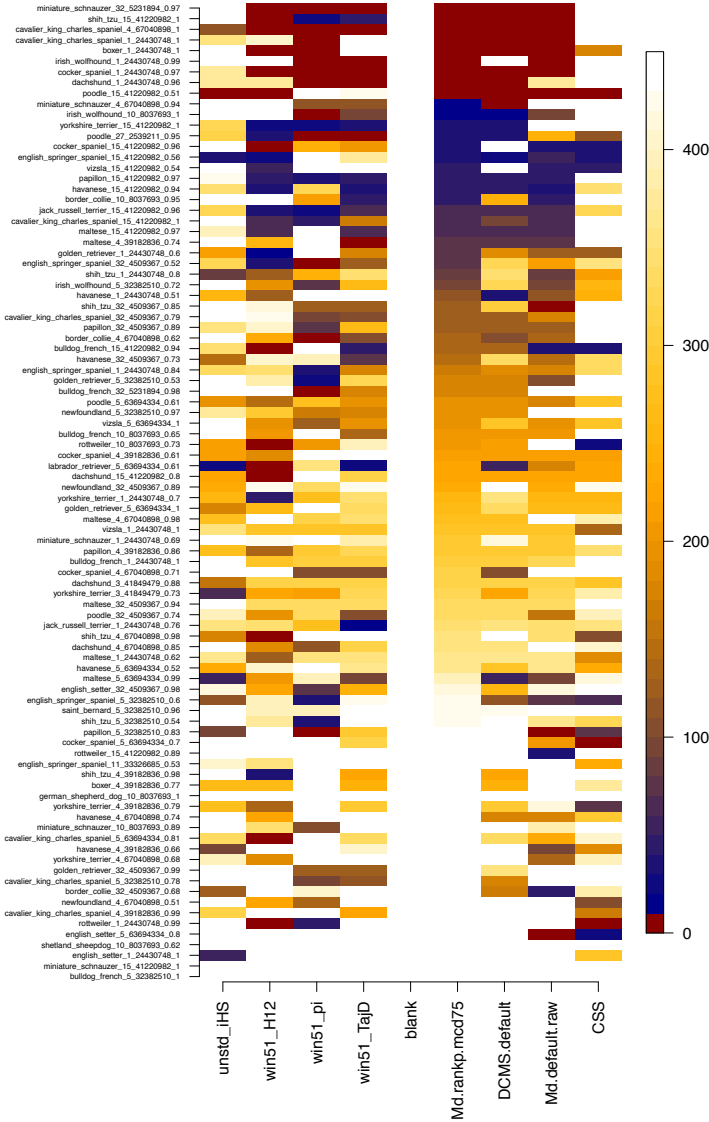
SUPPLEMENTARY FIGURE 12. Heatmap of distance in kilobases of the selection signal (max SNP) from the focal SNP, evaluated for each individual locus in the empirical dog dataset. Cases in which the focal SNP was genotyped and the maximum signal in the window was at this SNP (i.e., distance = 0) are colored dark red. The univariate statistics include iHS, H12, nucleotide diversity π, and Tajima's D. The compound measures include: the Mahalanobis distance based on minus-log fractional P-values (Md-rank-P), the decorrelated composite of multiple signals (DCMS), the Mahalanobis distance based on the raw statistics, and the composite signal of selection (CSS). The row labels indicate the dog breed, chromosome, base pair of the focal SNP, and the allele frequency of the focal SNP.

TO DO: EDIT X-AXIS LABELS
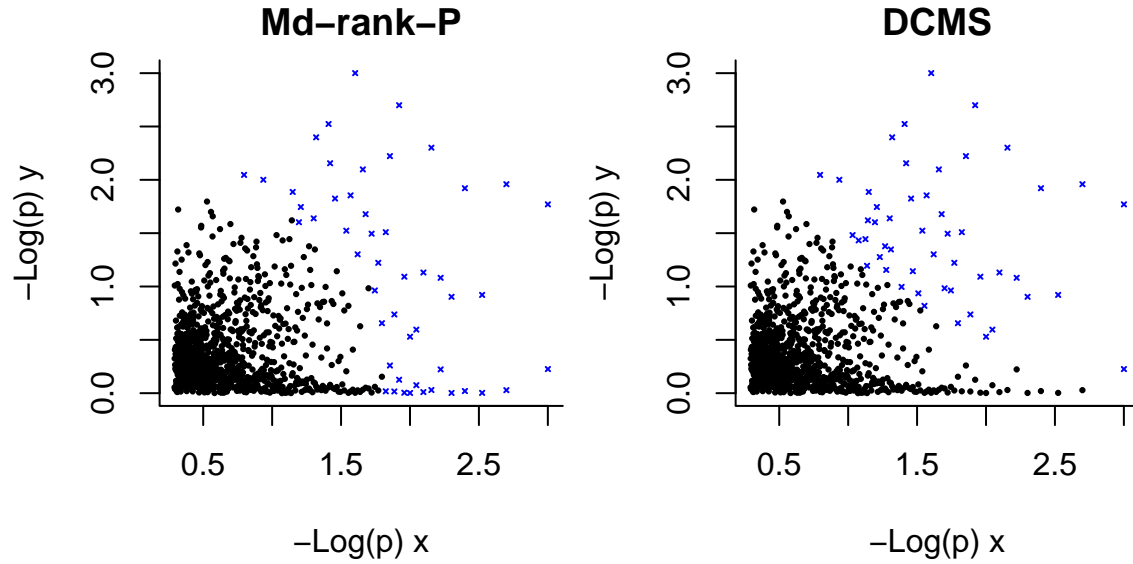
SUPPLEMENTARY FIGURE 13. Comparison of Md-rank-P and DCMS on P-values obtained from a bivariate normal distribution. Black circles indicate points less than the 0.95% of the compound measure and blue crosses indicate points above this threshold. In this toy example, some points are significant by both the x and y variables (points in upper right side of the graph), while other points are only significant by x variable (lower right part of the graph).