

Supplementary Information

Accounting for diverse evolutionary forces reveals mosaic patterns of selection on human preterm birth loci

LaBella, Abraham et al.

Table of Contents:

Supplementary Table:

1. Number of sPTB regions with evolutionary signatures

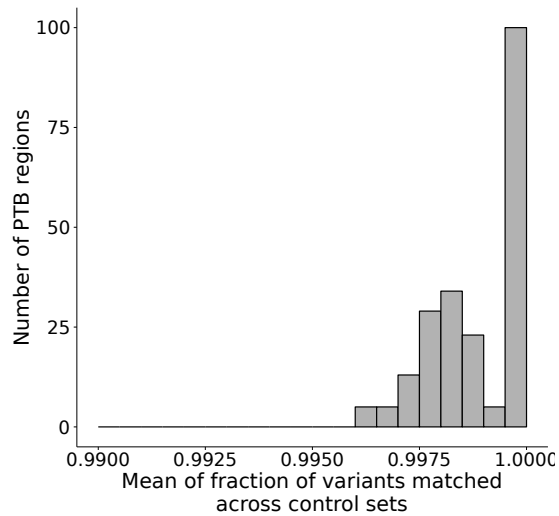
Supplementary Figures:

1. Match quality of sPTB regions across control regions
2. Matched and unmatched distributions for diverse evolutionary measures
3. Sensitivity analysis of diverse evolutionary forces acting on sPTB-associated genomic regions

Supplementary Table 1: Number of sPTB regions with evolutionary signatures. Number of sPTB regions out of 215 that have evidence for evolutionary signatures (top 5th percentile) when compared to 1,000 iteration of randomly selected regions (n=5,0000 per PTB region) for each evolutionary measure.

Evolutionary Measure	Mean	Standard Deviation
ARGweaver-TMRCA	39.5	1
Beta Score	23.9	0.2
F _{ST} Eas-Afr	30.2	0.5
F _{ST} Eur-Afr	36	0.3
F _{ST} Eas-Eas	32.5	0.7
GERP	23.4	0.8
iES Sabeti Eur	33	0.2
LINSIGHT	33.3	0.8
PhastCons	19	0.1
PhyloP	19.5	0.8
XP-EHH Afr-Eas	20	0.1
XP-EHH Afr-Eur	16.5	0.7
XP-EHH Eas-Eas	15.3	0.7

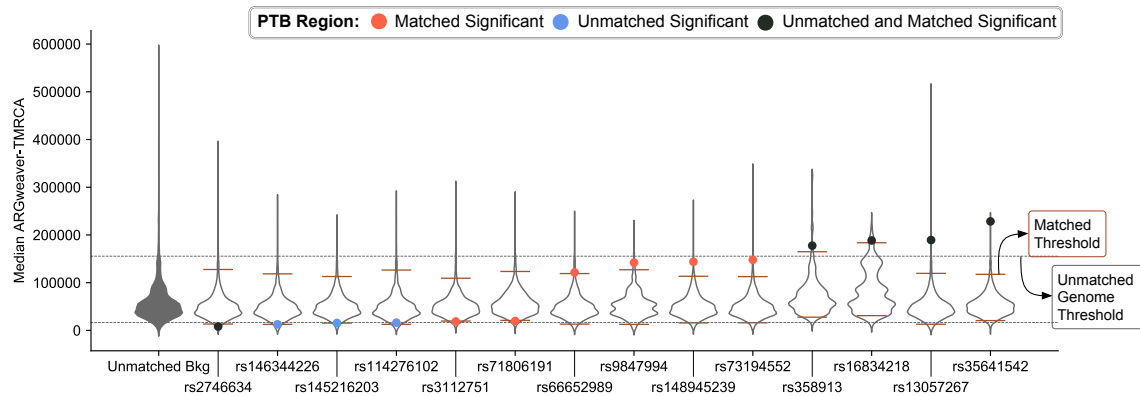
Supplementary Figure 1: Match quality of sPTB regions across control regions. For each 215 sPTB-associated genomic regions, we generated control regions matched on minor allele frequency and linkage disequilibrium structure. The match quality per genomic region, defined as the fraction of sPTB variants with a matched variant averaged across all control regions, is $\geq 99.6\%$ for all sPTB-associated genomic regions (Figure 1B).



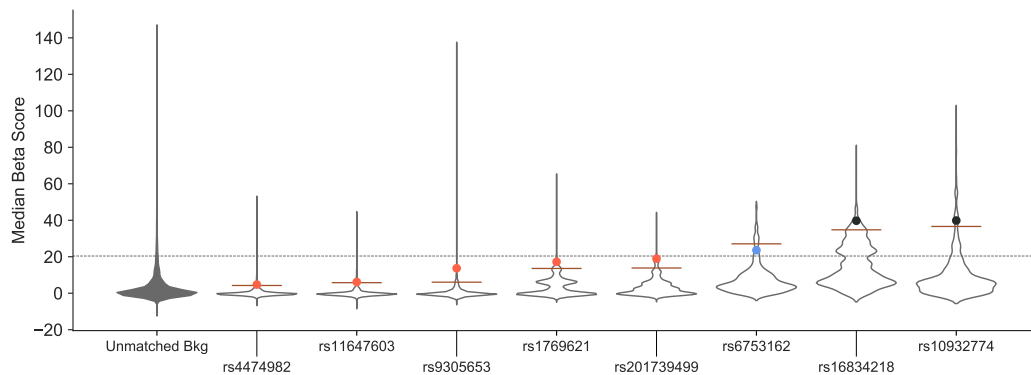
Supplementary Figure 2: Matched and unmatched distributions for diverse evolutionary measures.

Each figure below includes sPTB-associated genomic regions labeled by the lead variant (x-axis) that was statistically significant using either a genome wide background (blue dot), a matched background (red dot), or both (black dot) for the median value for a specific evolutionary measure (y-axis). Genome wide background distribution ('Unmatched Bkg, x-axis) is the filled violin plot; all unfilled violin plots are the MAF and LD matched distributions labeled by the lead variant for that sPTB genomic region. Dotted horizontal lines are genome wide unmatched thresholds for one or two-sided empirical p-value of 0.05 calculated from the genome-wide background distribution. Gold horizontal lines for each violin plot represents the threshold for empirical p-values of 0.05 derived from the matched distributions. Allele age (TMRCA from ARGweaver), PhyloP, and alignment block age are bi-directional measures, therefore, we calculated two-tailed p-values; all other evolutionary measures used one-tailed p-values. The evolutionary measures tested with sPTB regions include (a) ARGweaver-TMRCA, (b) Beta Score, (c) F_{ST} Eas-Afr, (d) F_{ST} Eur-Afr, (e) F_{ST} Eur-Eas (f) GERP, (g) iES, (h) LINSIGHT, (i) PhastCons, (j) PhyloP, (k) Syntentic age, (l) XP-EHH Afr-Eas, (m) XP-EHH Afr-EUR, (n) XP-EHH Eas-Eur.

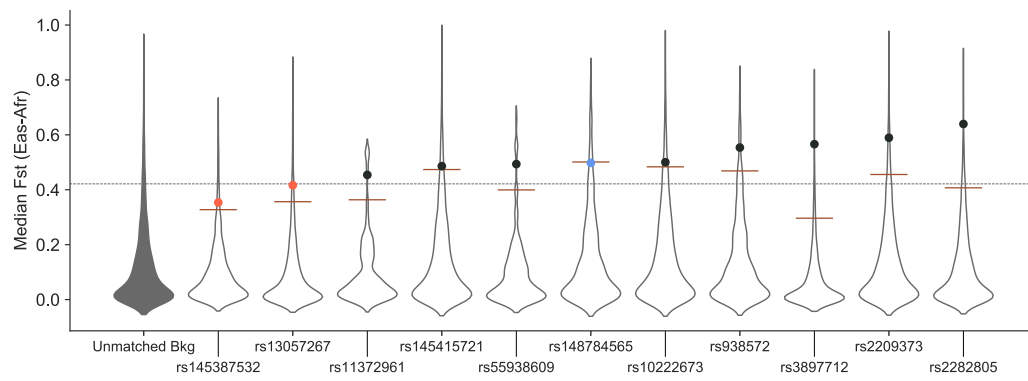
a)



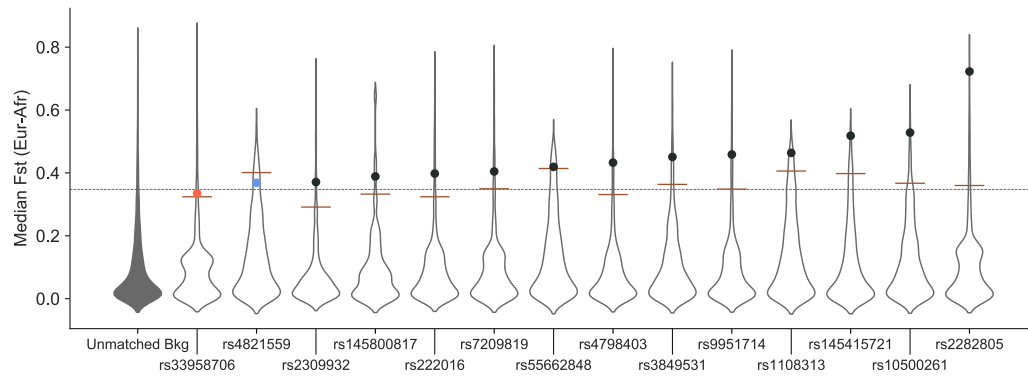
b)



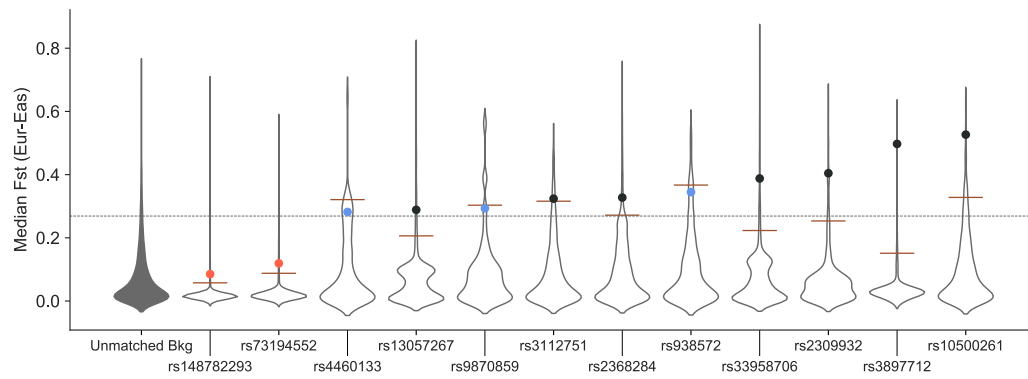
c)



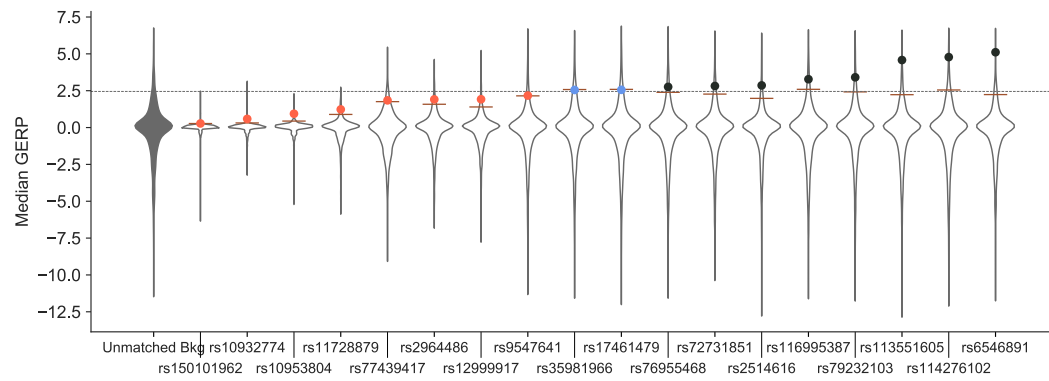
d)



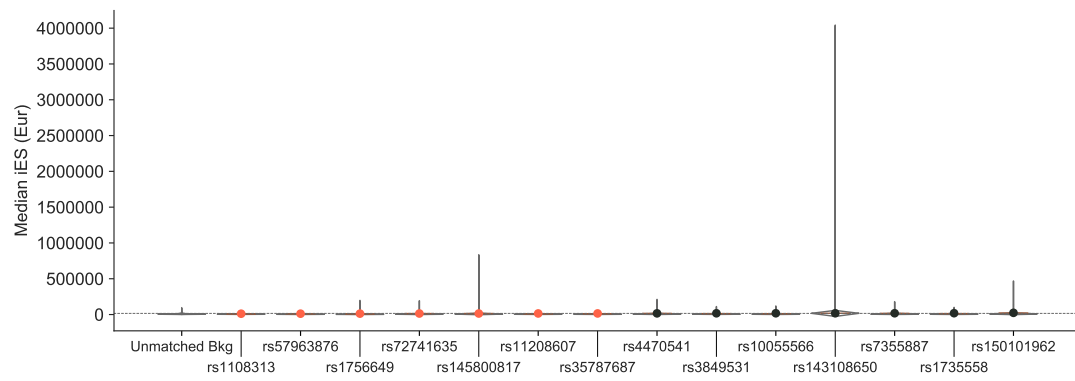
e)



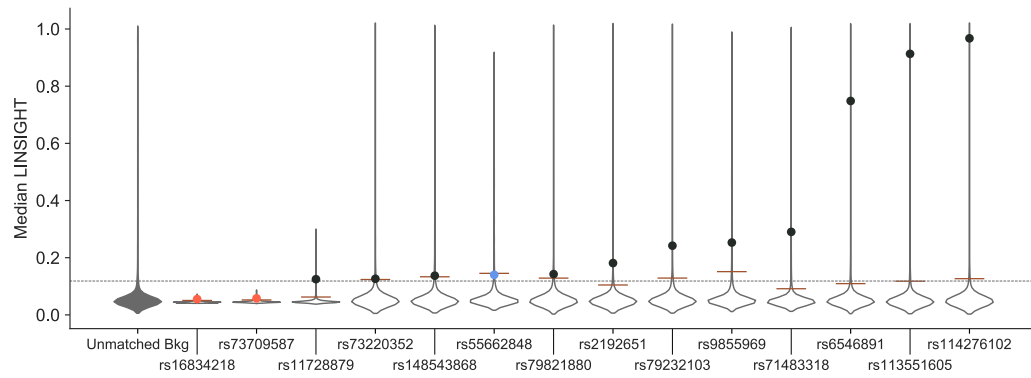
f)



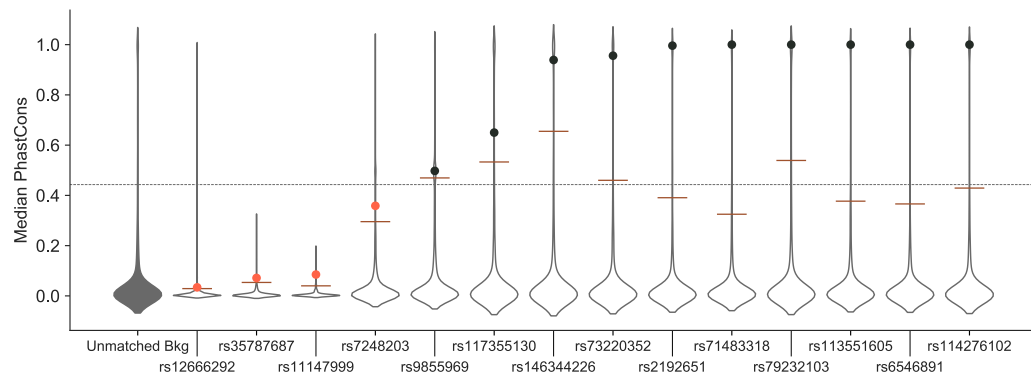
g)



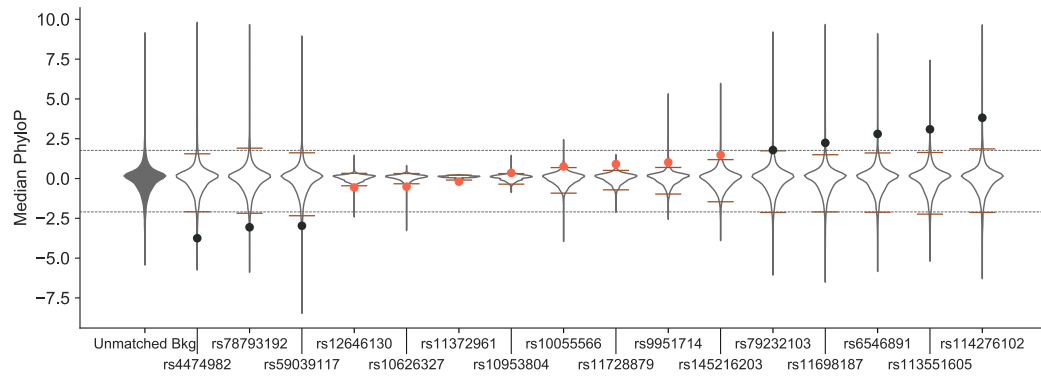
h)



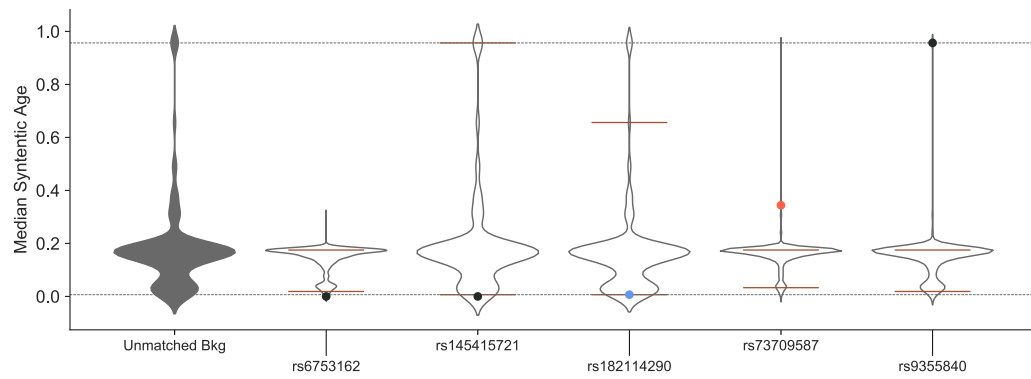
i)



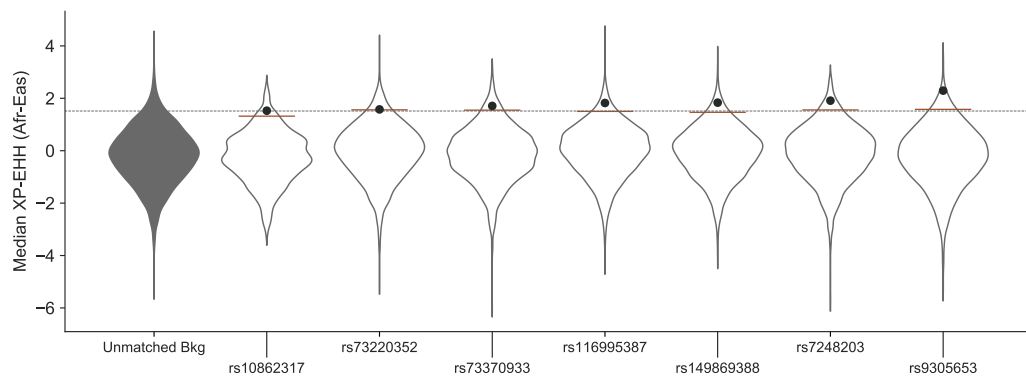
j)



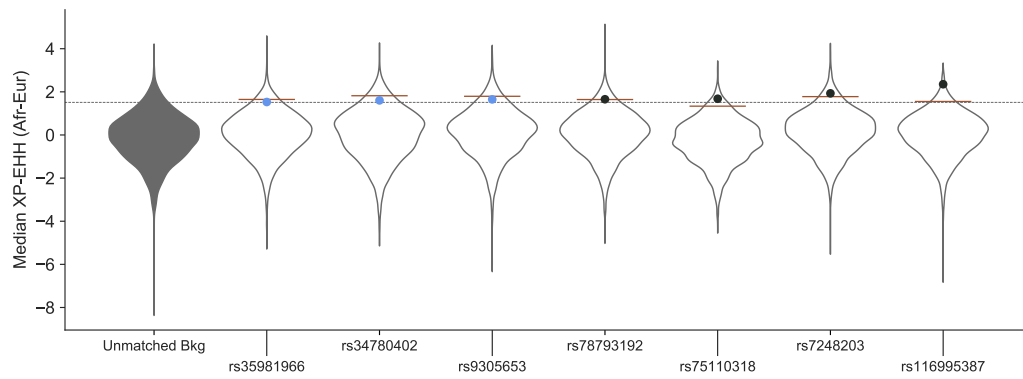
k)



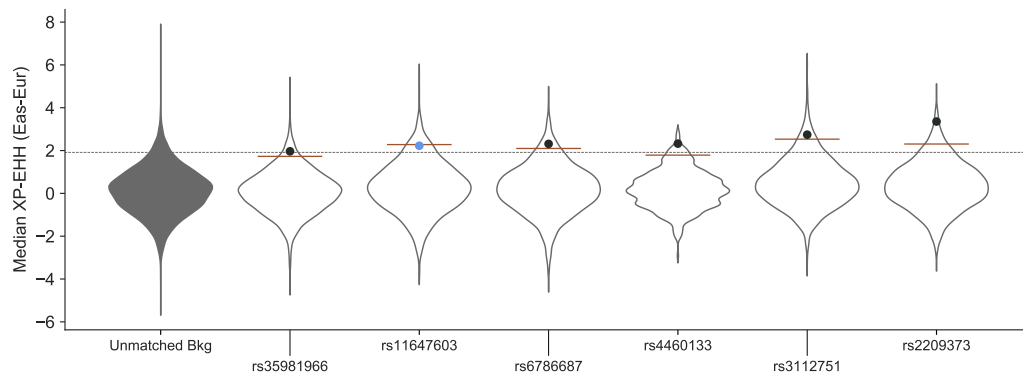
l)



m)



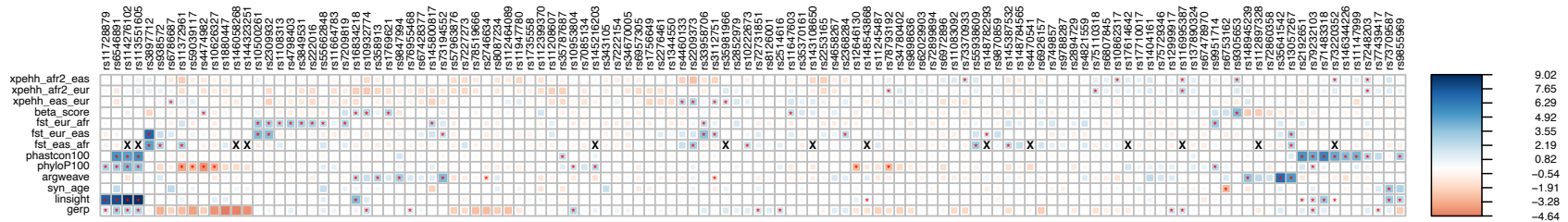
n)



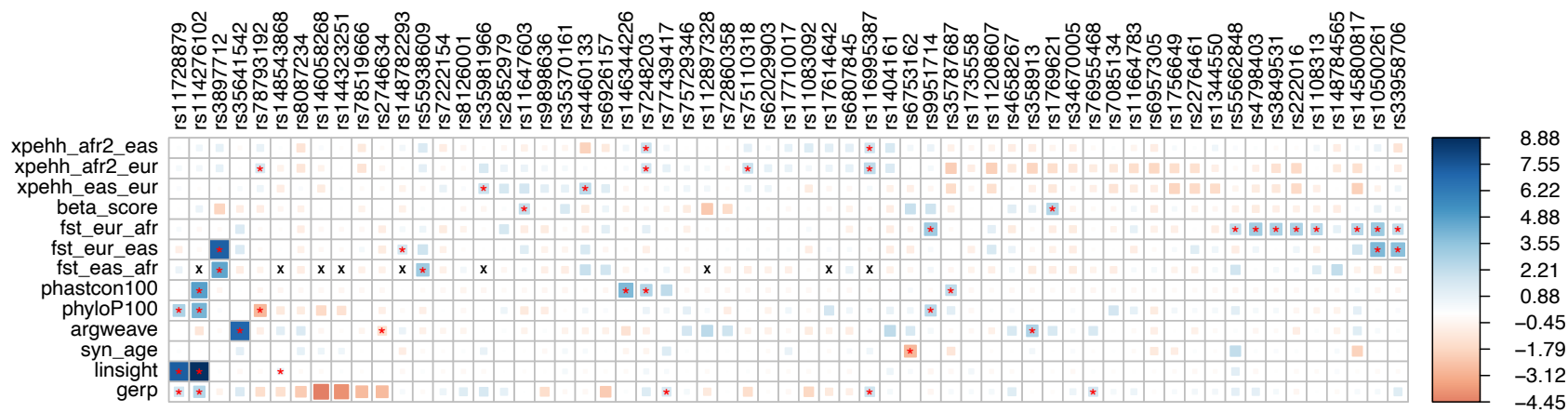
Supplementary Figure 3: Sensitivity analysis of diverse evolutionary forces acting on sPTB-associated genomic regions

We tested 215 sPTB-associated genomic regions tested for diverse evolutionary forces (as described in Figure 2 and Methods). Heatmaps show sPTB-associated genomic region (columns) referred to by the lead SNP rsID and evolutionary forces (y-axis). The relative strength (size of square) and direction (color) of each selection measure for each PTB region is presented as a z-score calculated from that region's background control distribution. Statistical significance was assessed by comparing the median value of the evolutionary measure to the background control distribution to derive an empirical one or two-sided p-value (* $p < 0.05$). Allele age (TMRCA from ARGweaver), PhyloP, and alignment block age are bi-directional measures, therefore, we calculated two-tailed p-values; all other evolutionary measures used one-tailed p-values. We evaluated the pattern of evolutionary forces observed under three different p-value thresholds for the lead variant sPTB GWAS p-value in each sPTB-associated genomic region: a) $p < 10E-4$, b) $5 * 10E-5$, c) $1 * 10E-5$. Across all these thresholds we observed strong signals of diverse evolutionary forces.

a)



b)



c)

