

Calculations

Weighted average scores

dist_tss_ave Gene 1: $20 \text{ kb} * 0.6 + 50 \text{ kb} * 0.4 = 32 \text{ kb}$

dist_tss_ave Gene 2: $70 \text{ kb} * 0.6 + 50 \text{ kb} * 0.4 = 62 \text{ kb}$

dist_foot_ave Gene 1: $0 \text{ kb} * 0.6 + 10 \text{ kb} * 0.4 = 4 \text{ kb}$

dist_foot_ave Gene 2: $35 \text{ kb} * 0.6 + 15 \text{ kb} * 0.4 = 27 \text{ kb}$

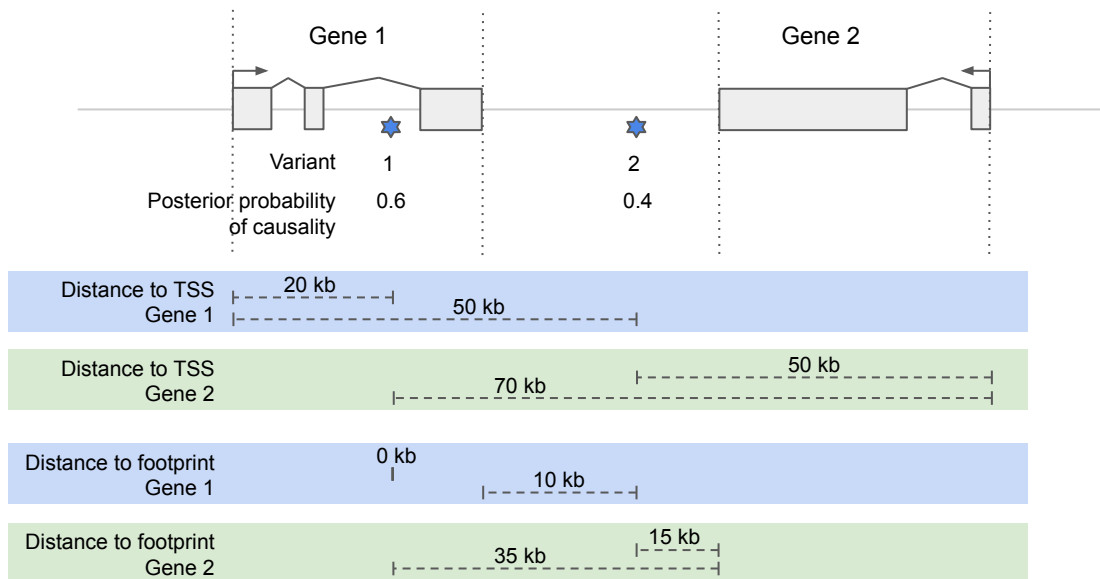
Neighborhood scores

dist_tss_ave_nbh Gene 1: $-\log(32 \text{ kb} / 32 \text{ kb}) = 0$

dist_tss_ave_nbh Gene 2: $-\log(62 \text{ kb} / 32 \text{ kb}) = -0.29$

dist_foot_ave_nbh Gene 1: $-\log(4 \text{ kb} / 4 \text{ kb}) = 0$

dist_foot_ave_nbh Gene 2: $-\log(27 \text{ kb} / 4 \text{ kb}) = -0.83$



Calculations

Weighted average scores

dist_tss_ave Gene 1: $20 \text{ kb} * 0.6 + 50 \text{ kb} * 0.4 = 32 \text{ kb}$

dist_tss_ave Gene 2: $70 \text{ kb} * 0.6 + 50 \text{ kb} * 0.4 = 62 \text{ kb}$

dist_foot_ave Gene 1: $0 \text{ kb} * 0.6 + 10 \text{ kb} * 0.4 = 4 \text{ kb}$

dist_foot_ave Gene 2: $35 \text{ kb} * 0.6 + 15 \text{ kb} * 0.4 = 27 \text{ kb}$

Neighborhood scores

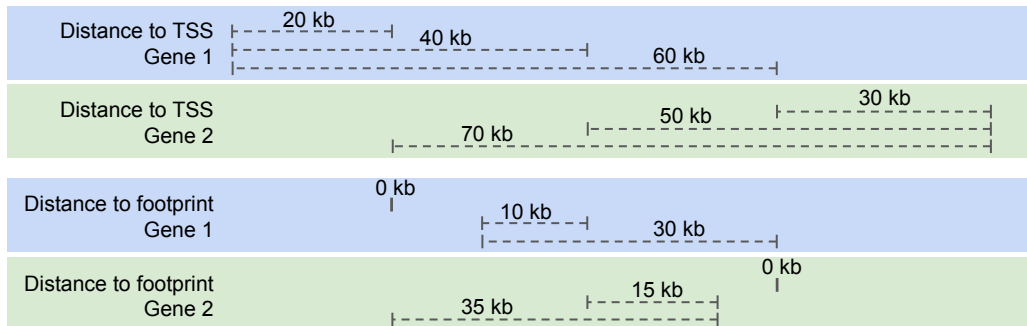
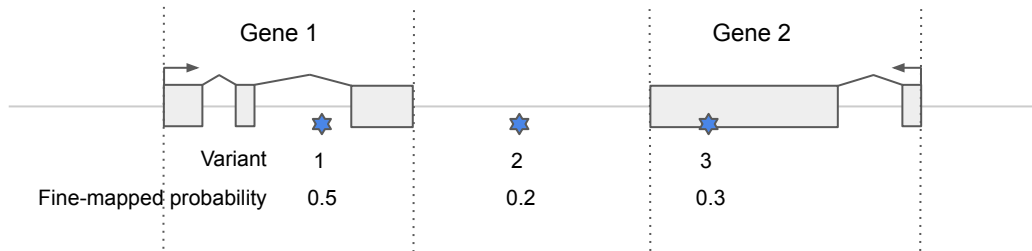
dist_tss_ave_nbh Gene 1: $-\log(32 \text{ kb} / 32 \text{ kb}) = 0$

dist_tss_ave_nbh Gene 2: $-\log(62 \text{ kb} / 32 \text{ kb}) = -0.29$

dist_foot_ave_nbh Gene 1: $-\log(4 \text{ kb} / 4 \text{ kb}) = 0$

dist_foot_ave_nbh Gene 2: $-\log(27 \text{ kb} / 4 \text{ kb}) = -0.83$

Schematic showing how variant to gene distance features are calculated. Distance to the transcription start site (TSS) is the number of bases from the variant to the TSS of the canonical transcript of the gene as defined by Ensembl. Distance to the gene footprint is the smallest number of base pair between the variant and any position between the TSS and transcription end site of the canonical transcript. (Calculations, left) Some L2G distance features are an average across variants, weighted by each variant's posterior probability from fine mapping. (Calculations, right) "Neighborhood" features are defined on a log scale relative to the gene with the best score in that category (here, smallest distance). The negative log is used so that genes with better feature values have higher neighborhood scores.



Calculations

Weighted TSS distance Gene 1: $20 \text{ kb} * 0.5 + 40 \text{ kb} * 0.2 + 60 \text{ kb} * 0.3 = 36 \text{ kb}$

Weighted TSS distance Gene 2: $70 \text{ kb} * 0.5 + 50 \text{ kb} * 0.2 + 30 \text{ kb} * 0.3 = 54 \text{ kb}$

Weighted footprint distance Gene 1: $0 \text{ kb} * 0.5 + 10 \text{ kb} * 0.2 + 30 \text{ kb} * 0.3 = 11 \text{ kb}$

Weighted footprint distance Gene 2: $35 \text{ kb} * 0.5 + 15 \text{ kb} * 0.2 + 0 \text{ kb} * 0.3 = 20.5 \text{ kb}$

Schematic showing how variant to gene distance features are calculated. Distance to the transcription start site (TSS) is the number of bases from the variant to the TSS of the canonical transcript of the gene as defined by Ensembl. Distance to the gene footprint is the smallest number of base pair between the variant and any position between the TSS and transcription end site of the canonical transcript. The distance feature values for each gene are an average across variants, weighted by each variant's fine-mapped probability.