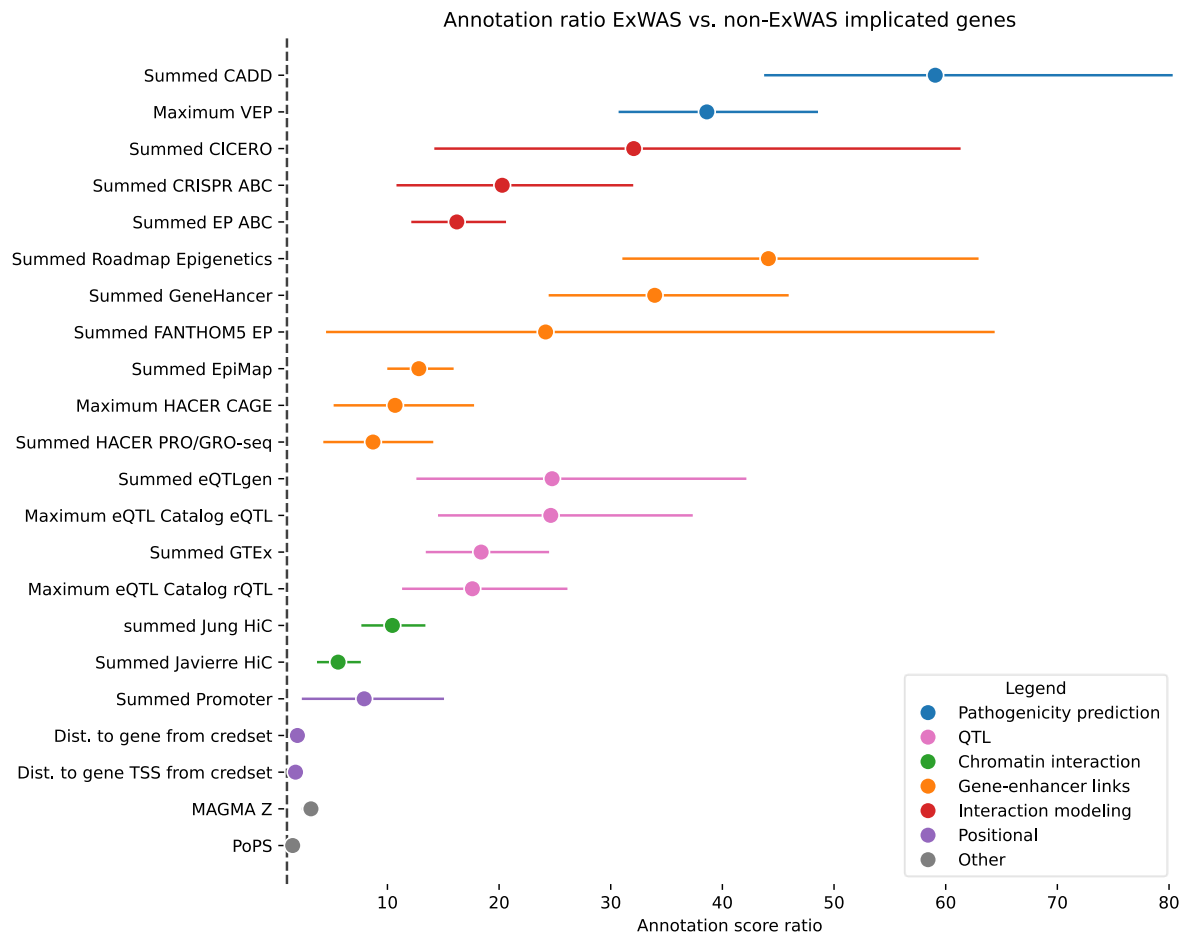**Supp. Figure 1 | Full feature impact by SHAP values.** Impact of features denoted by SHAP value. Blue denotes low feature values, red denote high feature values.

Supp. Figure 2 | Ratio of annotation scores in L2G expert-curated causal genes. Ratios of average annotation score per annotation in ExWAS implicated gene in GWAS locus vs. the rest of the genes in the locus. Error bars represent 95% confidence intervals were calculated by bootstrapping 1000 times.
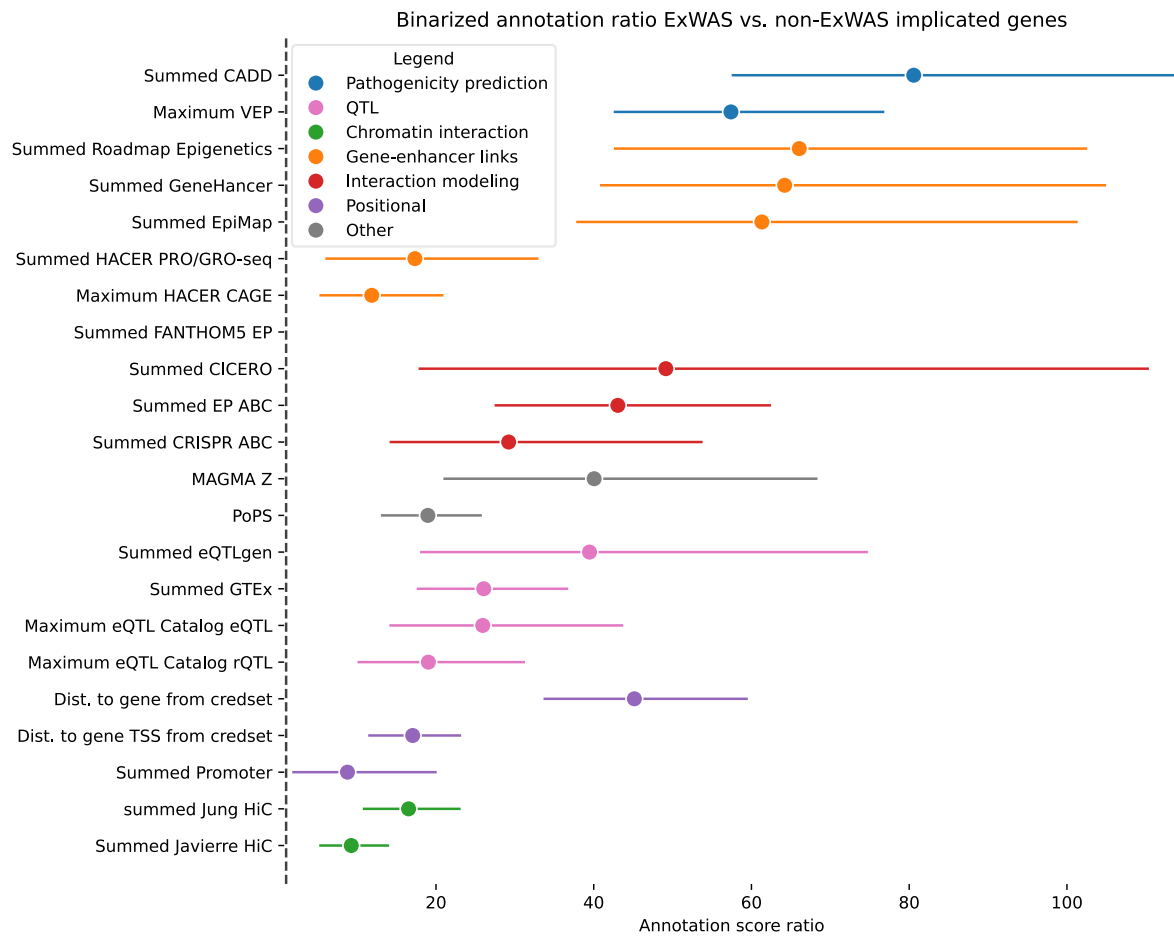
Supp. Figure 3 | Odds ratio of expert-curated causal gene having the highest annotation scores in the locus. Odds ratio of highest annotation score in the GWAS locus belonging to the ExWAS implicated gene in the locus. Error bars represent 95% confidence intervals were calculated by bootstrapping 1000 times.
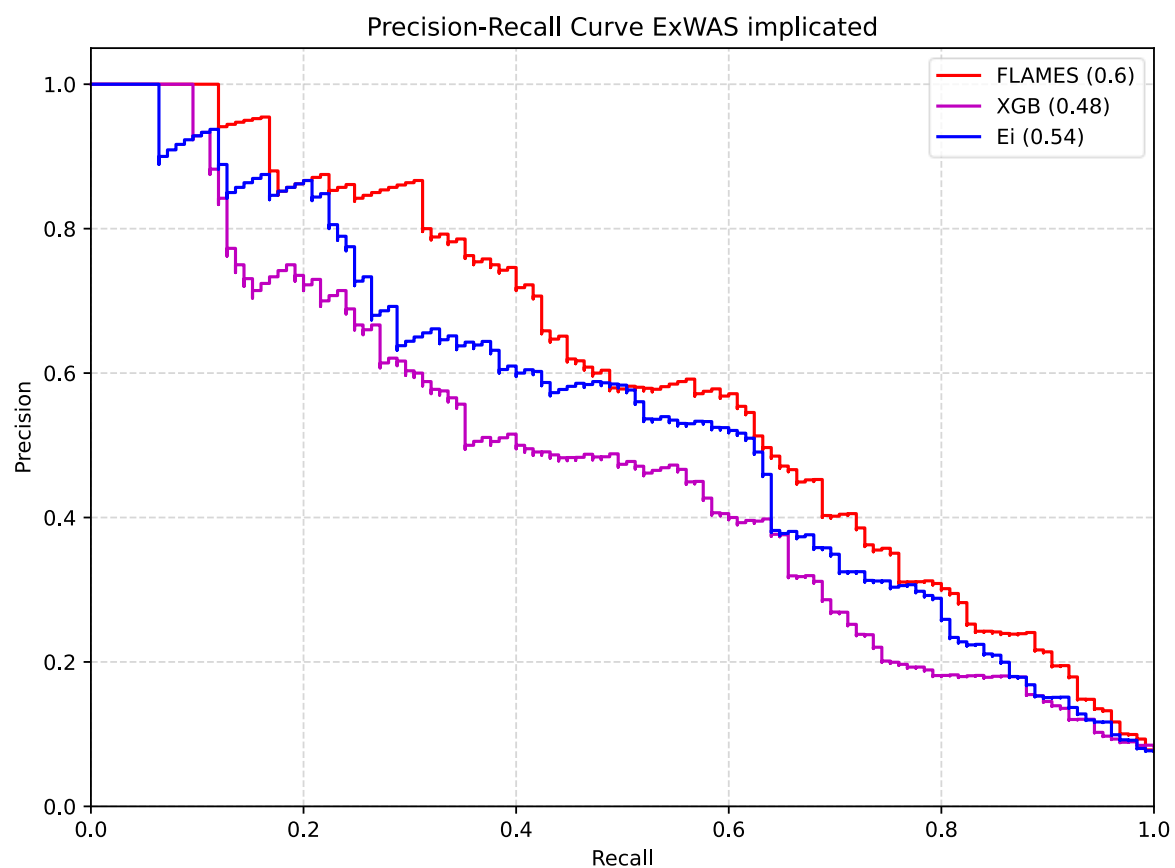
Supp. Figure 4 | Benchmark of FLAMES vs Ei. Benchmarking results of FLAMES versus Ei on ExWAS implicated benchmarking sets. For methods see Supp. Note.
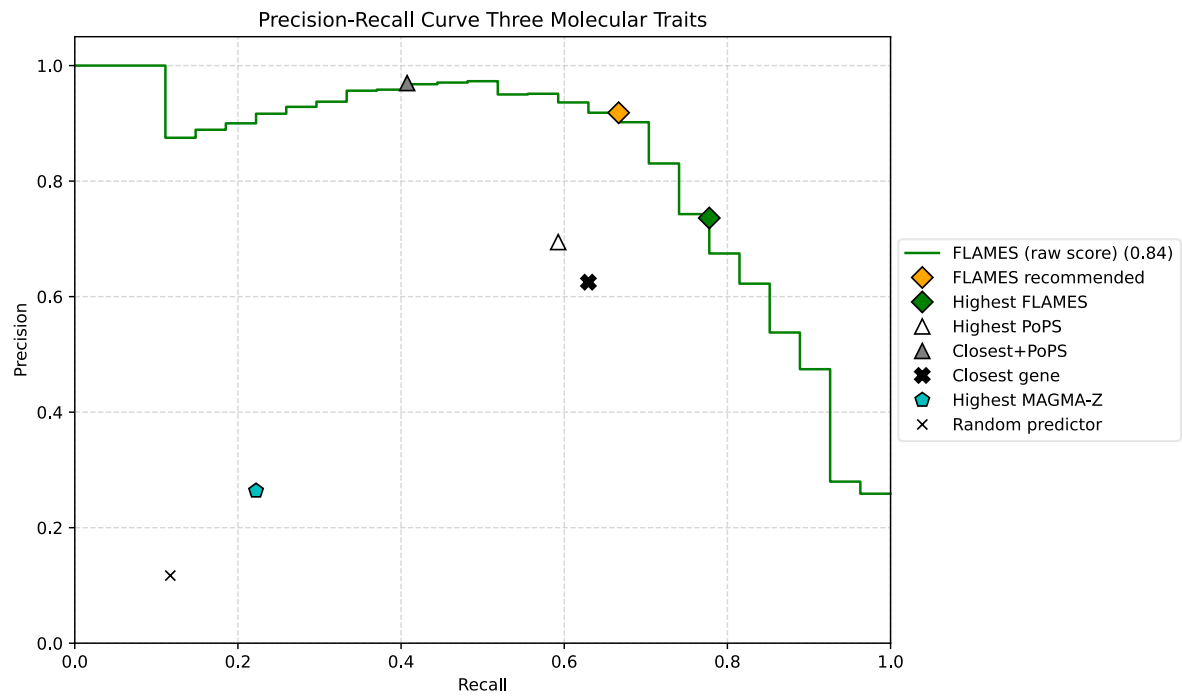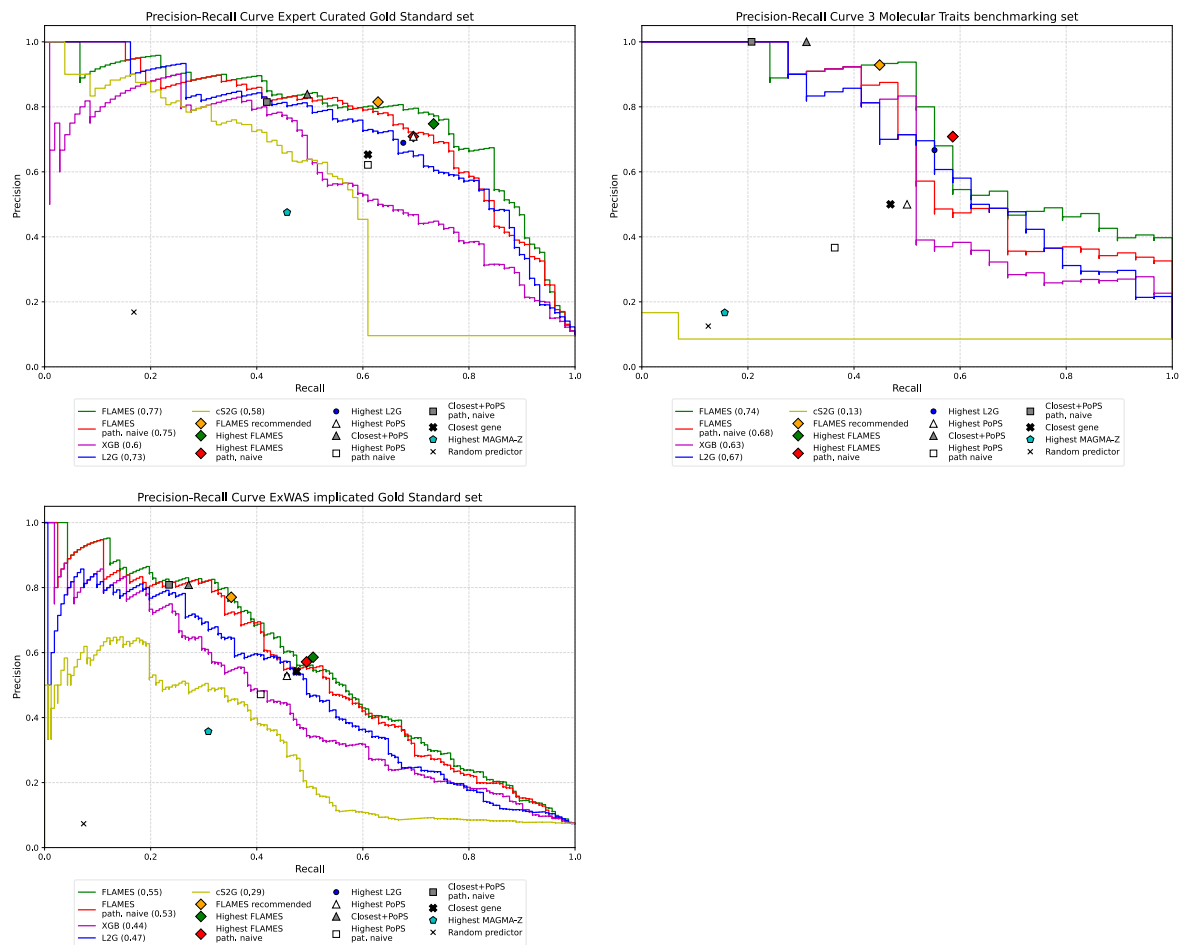
Supp. Figure 5 | Benchmark of FLAMES on three molecular traits with within-sample-LD fine-mapping. A more conservative raw score threshold was used, given that multiple signals in a single loci were possible (see Supp. Note).
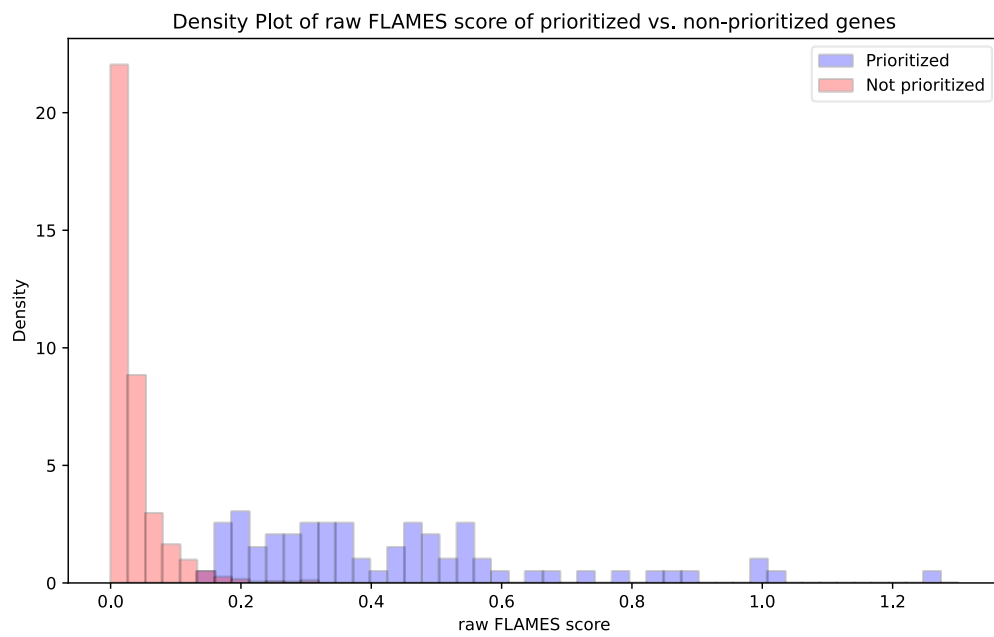
Supp. Figure 6 | Raw FLAMES score of prioritized vs non-prioritized genes. Scores were derived from ExWAS implicated benchmarking set. Prioritized genes in blue, not prioritized in red. To prioritize genes we used the recommended FLAMES threshold.
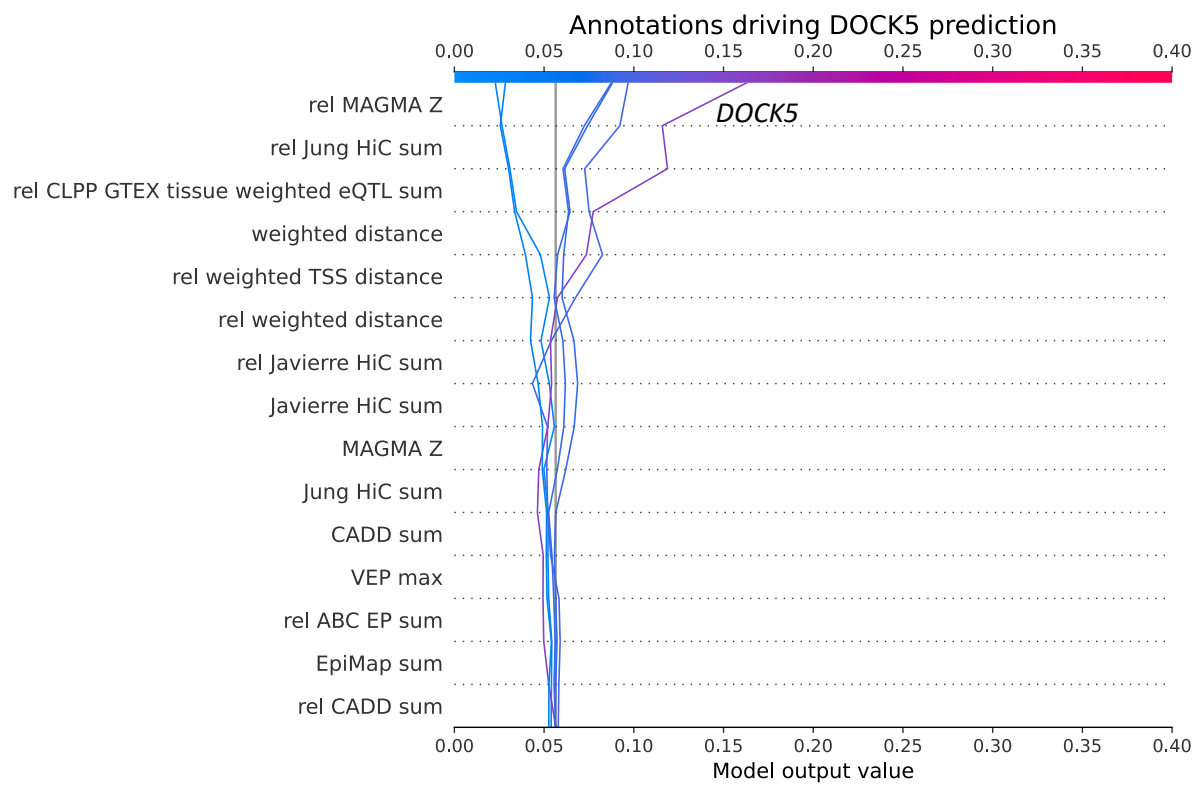
Supp. Figure 7 | Decision plot of *DOCK5* locus. Decision plot highlighting that *DOCK5* is prioritized mostly due to high MAGMA Z-scores, distance and eQTL evidence.

Supp. Figure 8 | Benchmarking results when including pathway naïve FLAMES. Benchmarking results in three distinct datasets. From top left to bottom left: Expert curated dataset, three molecular traits dataset, ExWAS implicated dataset. Similar to benchmarks in Fig. 4, however the red point and line represent FLAMES with pathway naïve FLAMES scores.
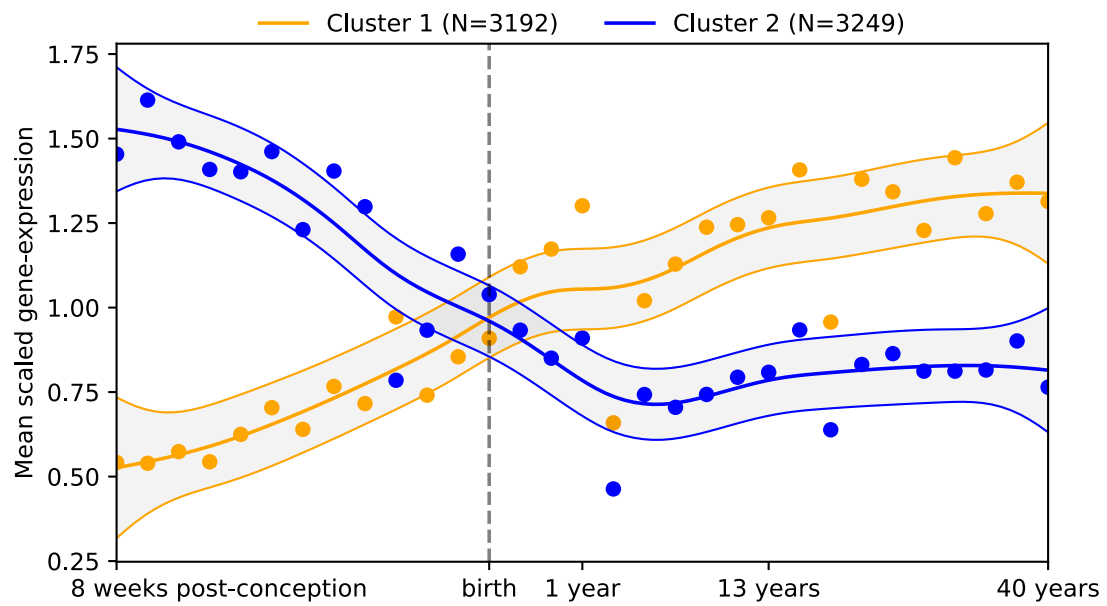
Supp. Figure 9 | Brain expressed genes expression profile in BRAINSPAN. Expression profile of BRAINSPAN brain expressed genes after k-means clustering. Gene expression is mean scaled per gene, and averaged across all genes in the cluster per timepoint. The expression of the two separate clusters is represented in orange and blue +/- 95% confidence intervals in grey.