

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

Supplementary material

Taras Yuziv Duda

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

Table S1. Clinical and molecular characteristics of colon cancer cohorts. This table contains the clinical and molecular characteristics of colon cancer samples from the TCGA (discovery) and E-MTAB-12862 (validation) cohorts.

Characteristics		TCGA n=405	E-MTAB-12862 n=782
Sex	Male	212	405
	Female	191	377
Age at Diagnosis	Median	68	74
	Range	34-90	30-94
BRAF Mutation Subtype	V600E	45	196
	Non-V600E	11	28
	WT	349	558
MSI Status	MSI-high	77	217
	BRAF-V600E	36	151
	BRAF-nonV600E	4	12
	BRAF WT	37	54
	MSI-Low/MSS	328	565
	BRAF-V600E	9	45
	BRAF-nonV600E	7	16
	BRAF WT	312	504
Tumor Stage	Stage I	74	90
	Stage II	158	326
	Stage III	110	287
	Stage VI	61	79
Tumor Location	Right-sided	239	284

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

	BRAF-V600E	42	177
	BRAF-nonV600E	9	19
	BRAF WT	188	302
	Left-sided	165	498
	BRAF-V600E	3	19
	BRAF-nonV600E	2	9
	BRAF WT	160	256

Table S2. Confusion matrix of E-MTAB-12862 trained classifier validated on TCGA.

This table presents the confusion matrix resulting from external validation of the classifier trained on the E-MTAB-12862 dataset, applied to annotated TCGA colon samples. The matrix compares predicted tumor locations (left vs. right) against the true labels.

	Reference	
Prediction	left	right
left	140	36
right	20	193

Table S3. External validation performance of the E-MTAB-12862-trained classifier on TCGA samples. This table presents the confusion matrix metrics for the classifier trained on the E-MTAB-12862 dataset and validated on annotated TCGA colon samples. The classifier achieved an overall accuracy of 85.6%. Sensitivity and specificity were 87.5% and 84.3%.

Accuracy	0.856
95% CI	(0.8172, 0.8894)
No Information Rate	0.5887
P-Value [ACcc > NIR]	< 2E-16
Kappa	0.7071
Mcnemar's Test P-Value	Mcnemar's Test P-Value
Sensitivity	0.875
Specificity	0.8428
Pos Pred Value	0.7955
Neg Pred Value	0.9061

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

Prevalence	0.4113
Detection Rate	0.3599
Detection Prevalence	0.4524
Balanced Accuracy	0.8589

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

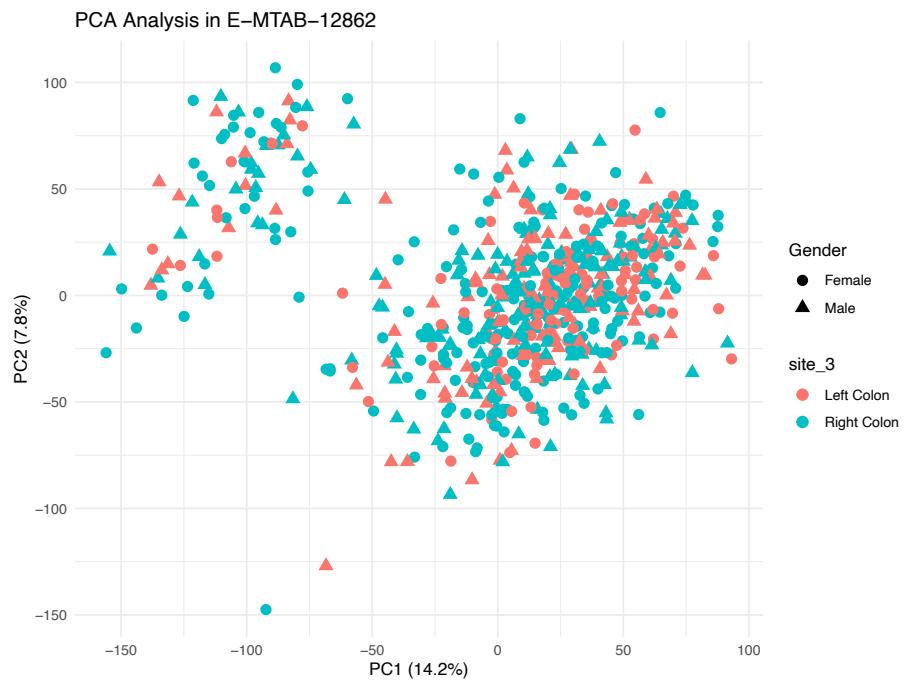


Figure S1. PCA of gene expression data before model training (PC1 vs. PC2). PCA was performed on the VST-normalized gene expression data (E-MTAB-12862) to explore clustering by tumor location. In this plot, the first two principal components, PC1 and PC2, which account for 14.2% and 7.8% of the total variance, show no clear separation between left-sided and right-sided colon tumors.

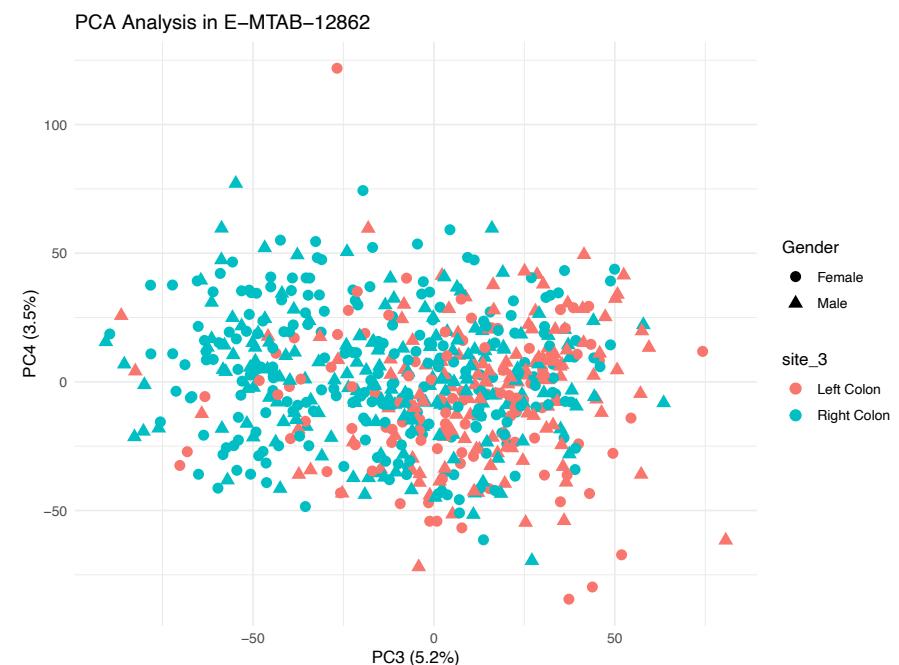


Figure S2. PCA of gene expression data before model training (PC3 vs. PC4). PCA revealed clearer separation between left-sided and right-sided colon tumor along the third and fourth principal components. PC3 and PC4 explain 5.2% and 3.5% of the total variance.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

Number of genes

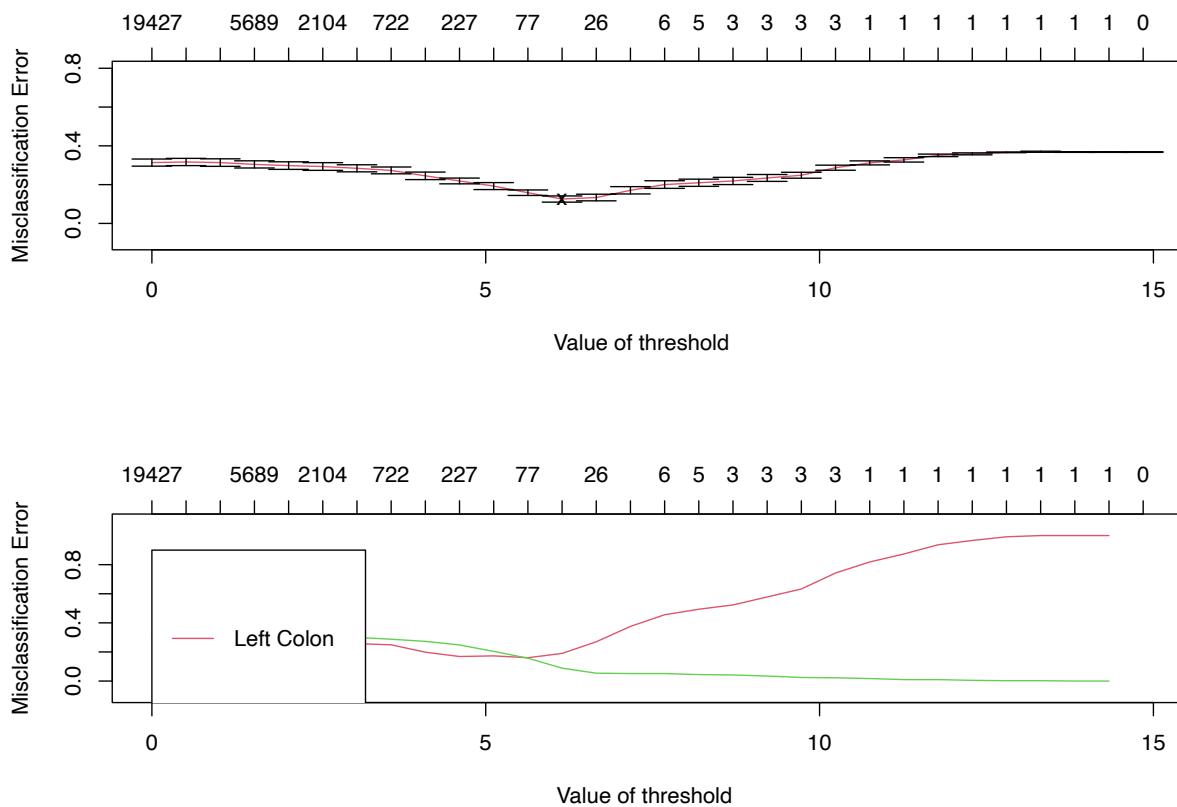


Figure S3. Cross-validation performance of the PAMr classifier. The top panel shows the overall cross-validated misclassification error of the model across a range of threshold values. Error decreases as the threshold increases and reaches a minimum at a threshold of 6.14, corresponding to a model that retained 51 genes.

The bottom panel displays class-specific misclassification error rates for left-sided (red line) and right-sided (green line) colon tumors. Left-sided tumors increased sharply at higher threshold, suggesting reduced classification robustness when fewer genes were used in the model. In contrast, right-sided tumors consistently had low error rates across thresholds, indicating slightly higher sensitivity.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

BRAF V600E vs BRAF WT – TCGA

EnhancedVolcano

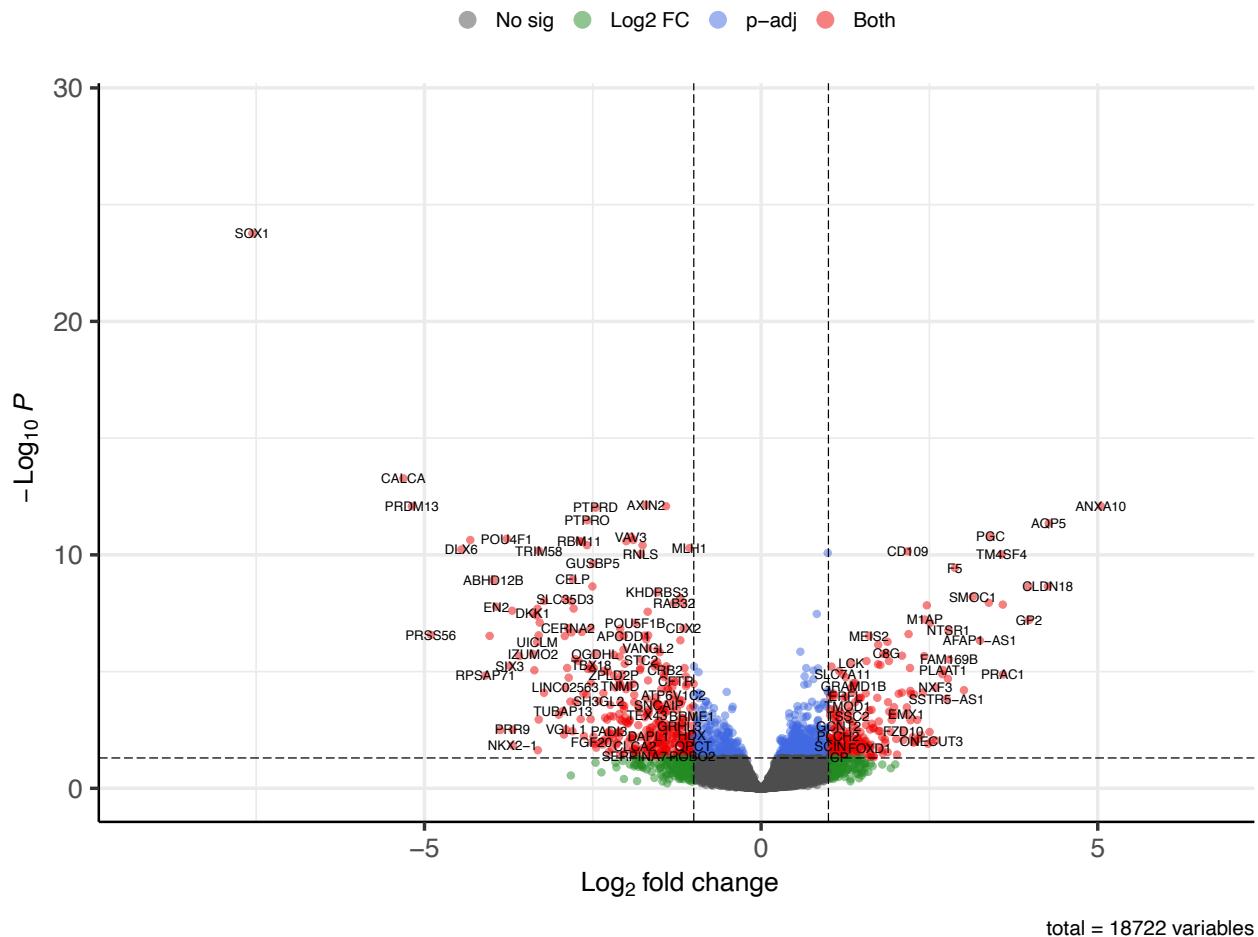


Figure S4. Differential gene expression between BRAF-V600E and BRAF WT tumors in TCGA. Volcano plot visualizing differential gene expression between BRAF-V600E and BRAF WT colon tumors in the TCGA cohort. Significantly differentially expressed genes were identified based on an absolute log2 fold change > 1 and an adjusted p-value <0.05 . Genes meeting both criteria are highlighted in red. Positive log2 fold change are genes upregulated in BRAF-V600E tumors and negative log2 fold change are genes downregulated in BRAF-V600E tumors.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

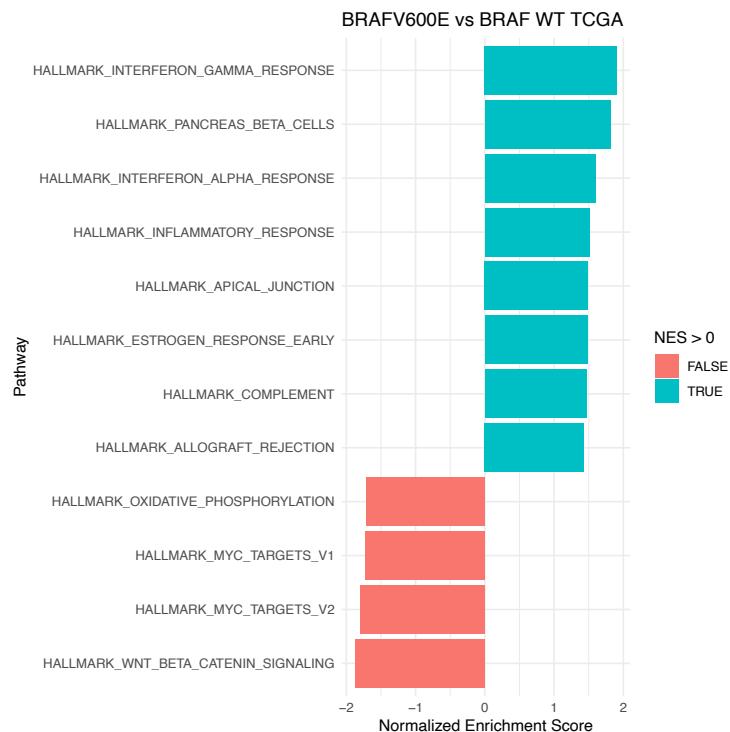


Figure S5. Gene set enrichment analysis (GSEA) of Hallmark pathways in BRAF-V600E vs. BRAF WT tumors (p-adj < 0.05, absolute NES > 1)(TCGA cohort). Bar plot showing the results of GSEA using the Hallmarks gene sets from MSigDB, comparing BRAF-V600E to BRAF WT tumors in the TCGA cohort

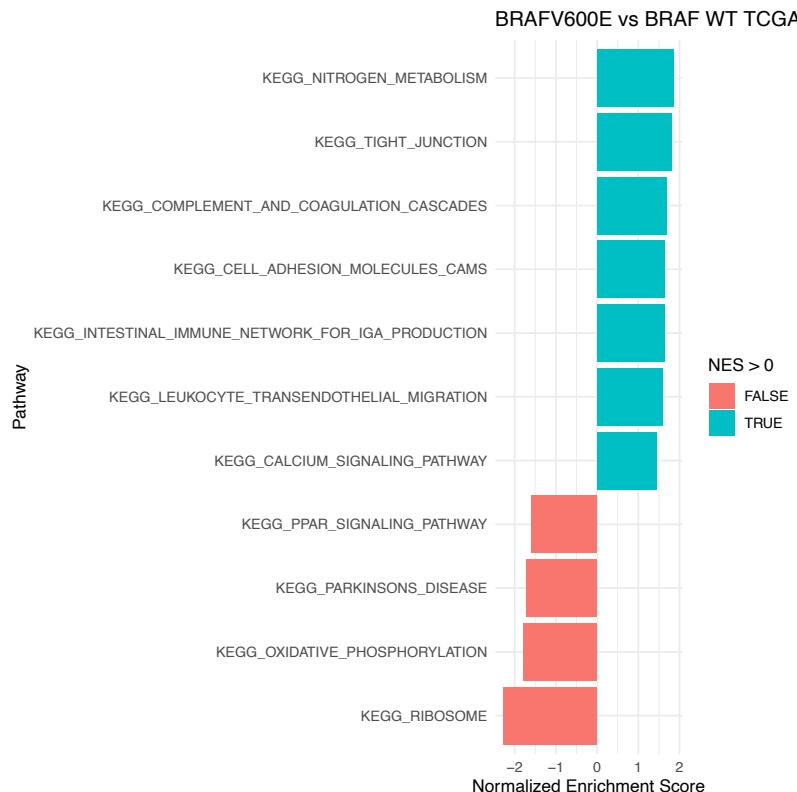


Figure S6. Gene set enrichment analysis (GSEA) of KEGG pathways in BRAF-V600E vs. BRAF WT tumors (p-adj < 0.05, absolute NES > 1) (TCGA cohort). Bar plot showing

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

the results of GSEA using the KEGG gene sets from MSigDB, comparing BRAF-V600E to BRAF WT tumors in the TCGA cohort.

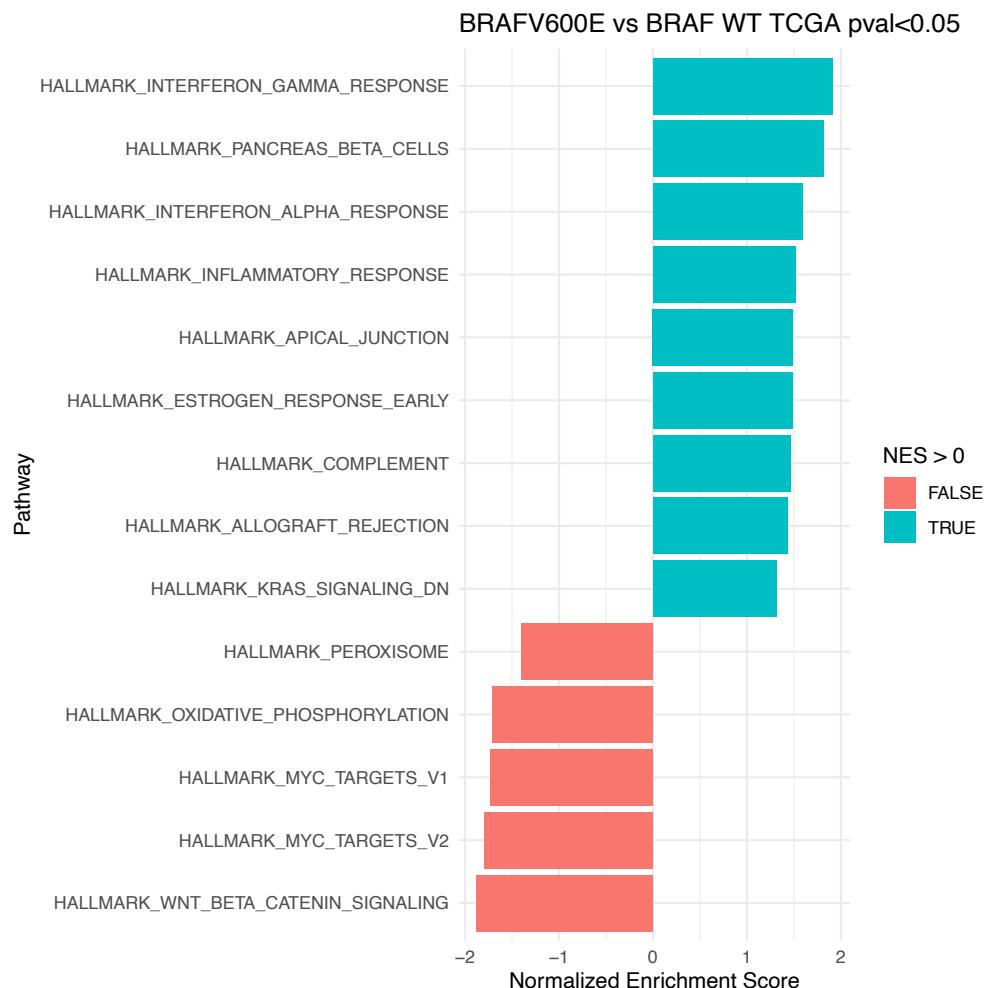


Figure S7. Gene set enrichment analysis (GSEA) of Hallmark pathways in BRAF-V600E vs. BRAF WT tumors (p-val< 0.05, absolute NES > 1) (TCGA cohort). Bar plot showing the results of GSEA using the Hallmarks gene sets from MSigDB, comparing BRAF-V600E to BRAF WT tumors in the TCGA cohort (p-val<0.05).

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

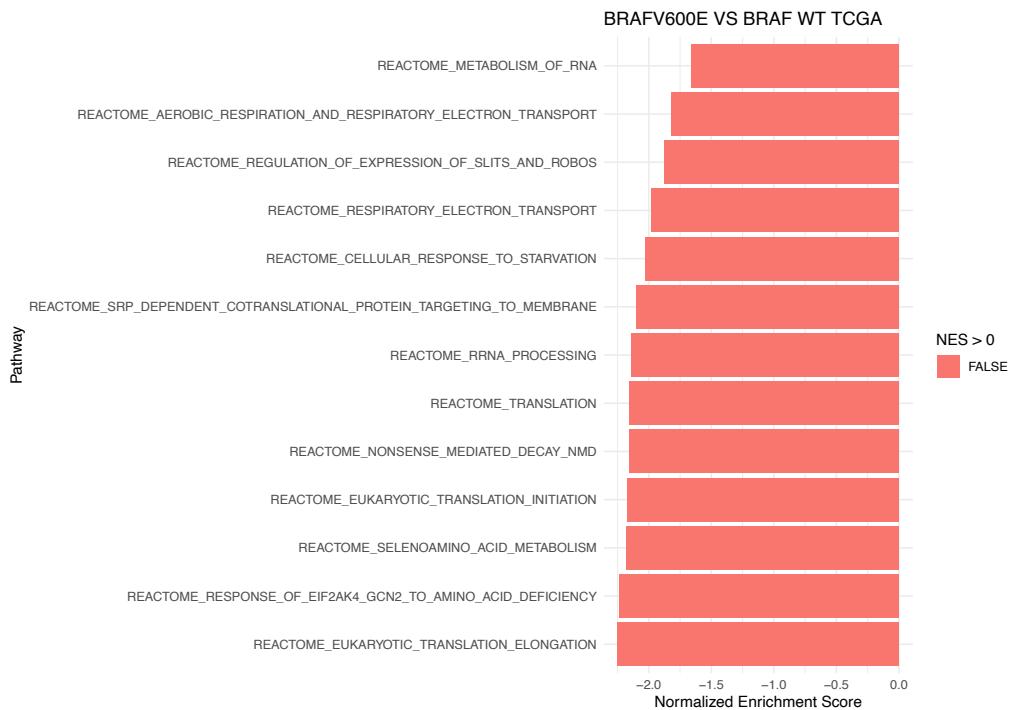


Figure S8. Gene set enrichment analysis (GSEA) of Reactome pathways in BRAF-V600E vs. BRAF WT tumors (p-adj < 0.05, absolute NES > 1) (TCGA cohort). Bar plot showing the results of GSEA using the Reactome gene sets from MSigDB, comparing BRAF-V600E to BRAF WT tumors in the TCGA cohort.

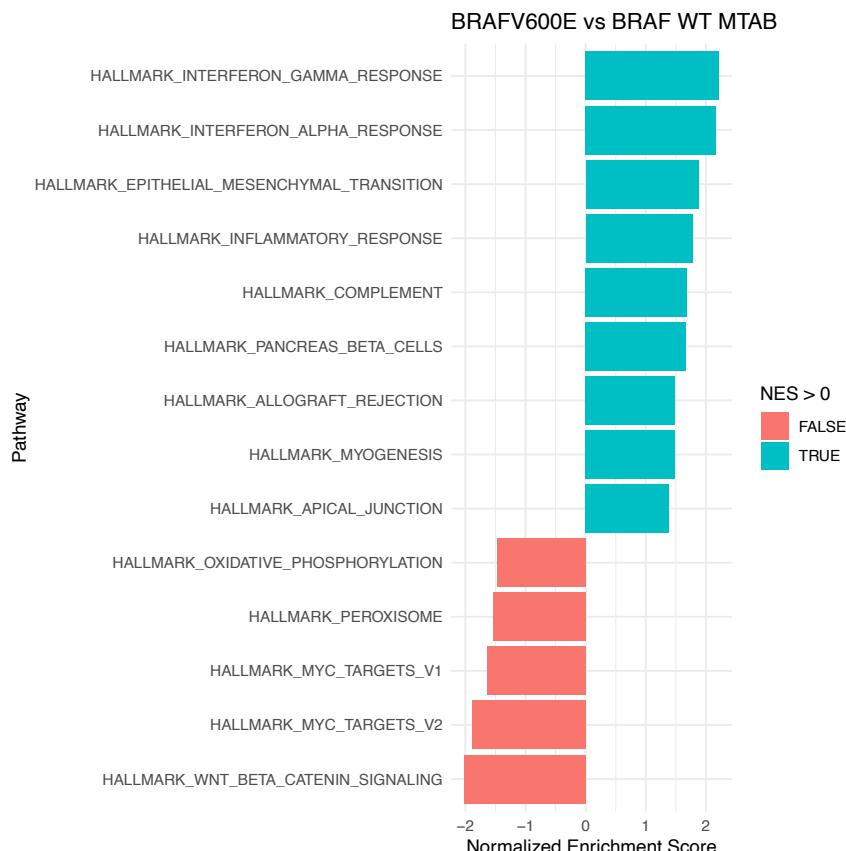


Figure S9. Gene set enrichment analysis (GSEA) of Hallmark pathways in BRAF-V600E vs. BRAF WT tumors (p-adj < 0.05, absolute NES > 1) (E-MTAB-12862 cohort).

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

Bar plot showing the results of GSEA using the Hallmark gene sets from MSigDB, comparing BRAF-V600E to BRAF WT tumors in the E-MTAB-12862 cohort.

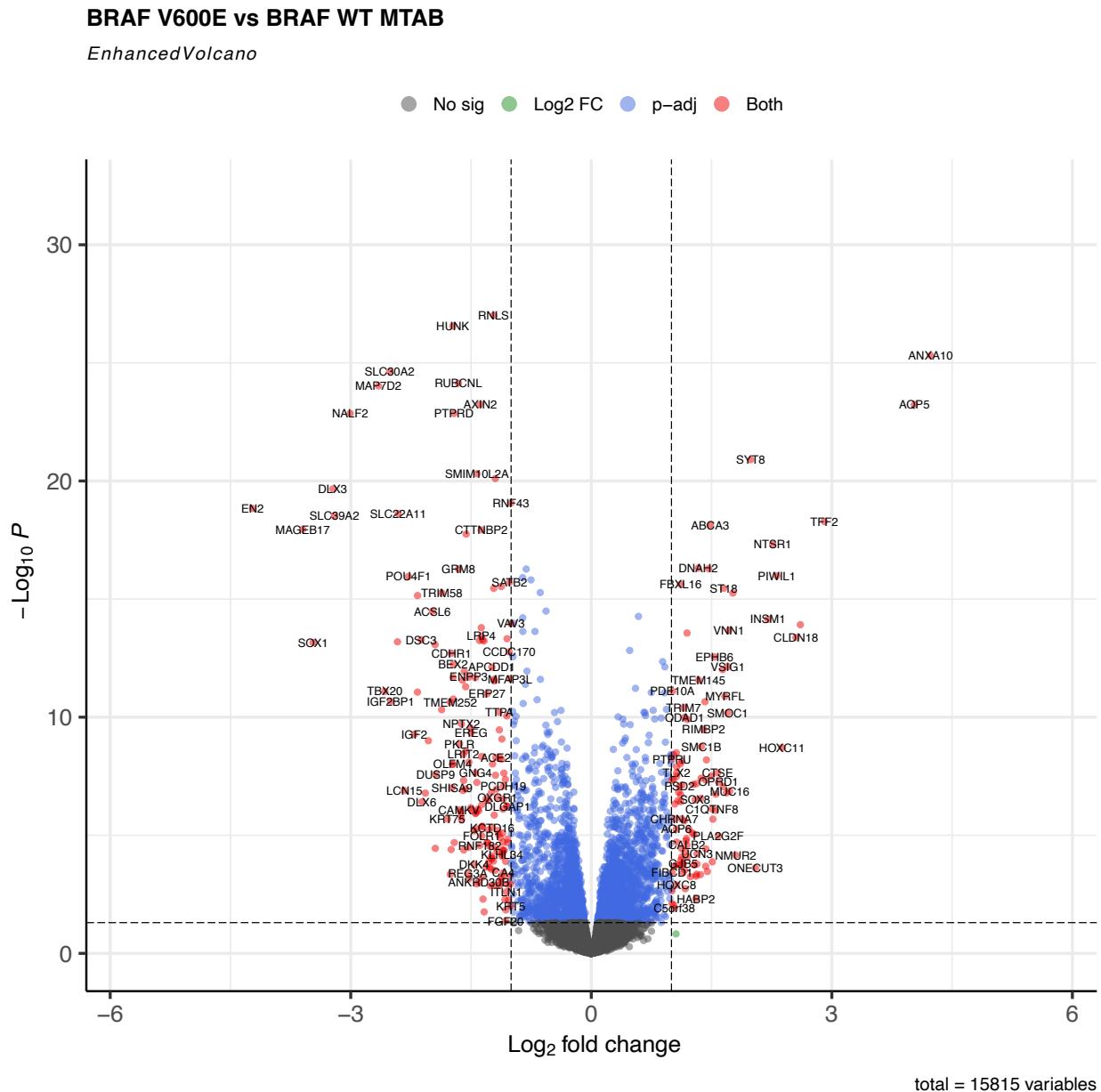


Figure S10. Differential gene expression between BRAF-V600E and BRAF WT tumors in E-MTAB-12862. Volcano plot visualizing differential gene expression between BRAF-V600E and BRAF WT colon cancer tumors in the E-MTAB-12862 cohort. Significantly differentially expressed genes were identified based on an absolute log2 fold change > 1 and an adjusted p-value < 0.05 . Genes meeting both criteria are highlighted in red. Positive log2 fold change are genes upregulated in BRAF-V600E tumors and negative log2 fold change are genes downregulated in BRAF-V600E tumors.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

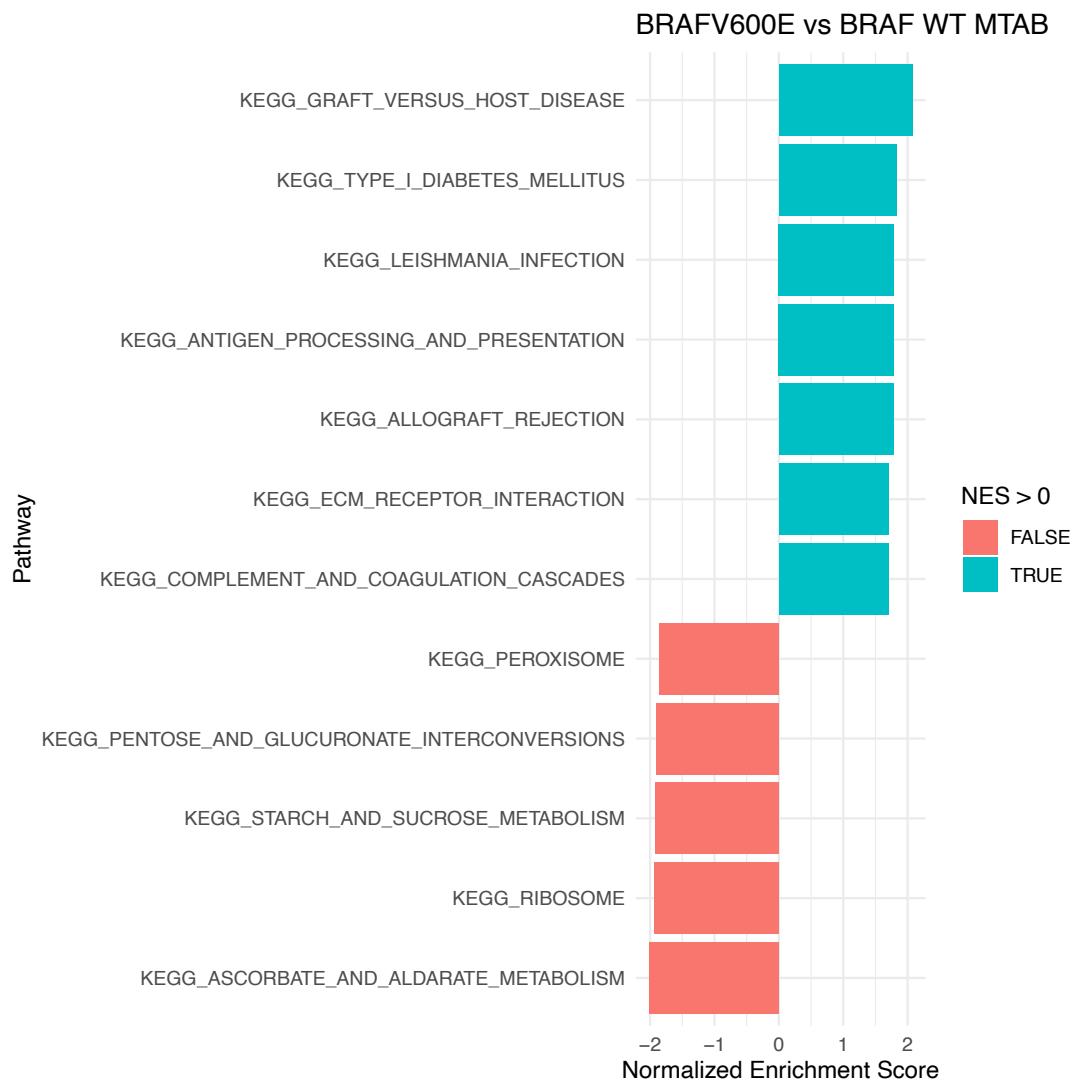


Figure S10. Gene set enrichment analysis (GSEA) of KEGG pathways in BRAF-V600E vs. BRAF WT tumors (p-adj < 0.05, absolute NES > 1) (E-MTAB-12862 cohort). Bar plot showing the results of GSEA using the KEGG gene sets from MSigDB, comparing BRAF-V600E to BRAF WT tumors in the E-MTAB-12862 cohort.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

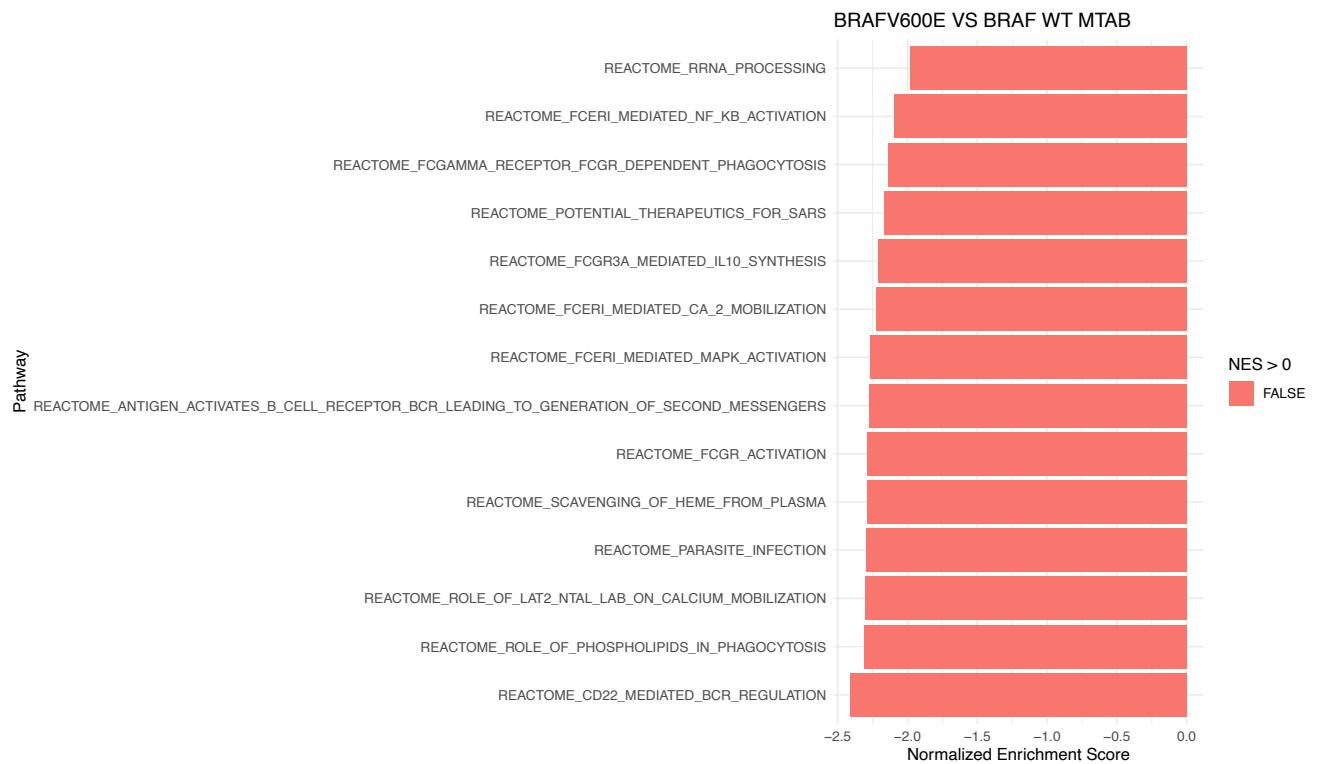


Figure S11. Gene set enrichment analysis (GSEA) of Reactome pathways in BRAF-V600E vs. BRAF WT tumors (p-adj < 0.05, absolute NES > 1) (E-MTAB-12862 cohort).
Bar plot showing the results of GSEA using the Reactome gene sets from MSigDB, comparing BRAF-V600E to BRAF WT tumors in the E-MTAB-12862 cohort.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

BRAF V600E vs BRAF WT+ nonV600E TCGA

EnhancedVolcano

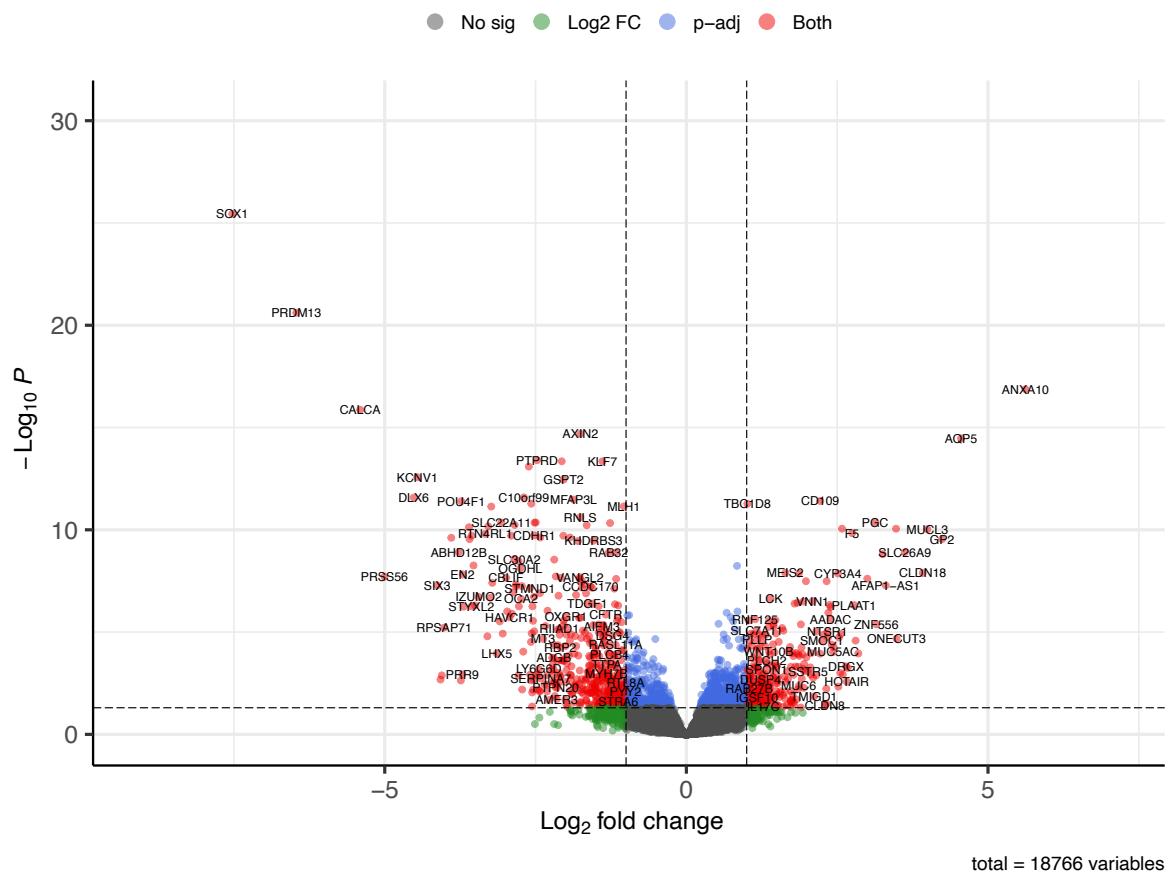


Figure S12. Differential gene expression between BRAF-V600E and BRAF WT + nonV600E tumors in TCGA. Volcano plot visualizing differential gene expression between BRAF-V600E and BRAF WT + nonV600E colon tumors in the TCGA cohort. Significantly differentially expressed genes were identified based on an absolute log2 fold change > 1 and an adjusted p-value < 0.05 . Genes meeting both criteria are highlighted in red. Positive log2 fold change are genes upregulated in BRAF-V600E tumors and negative log2 fold change are genes downregulated in BRAF-V600E tumors.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

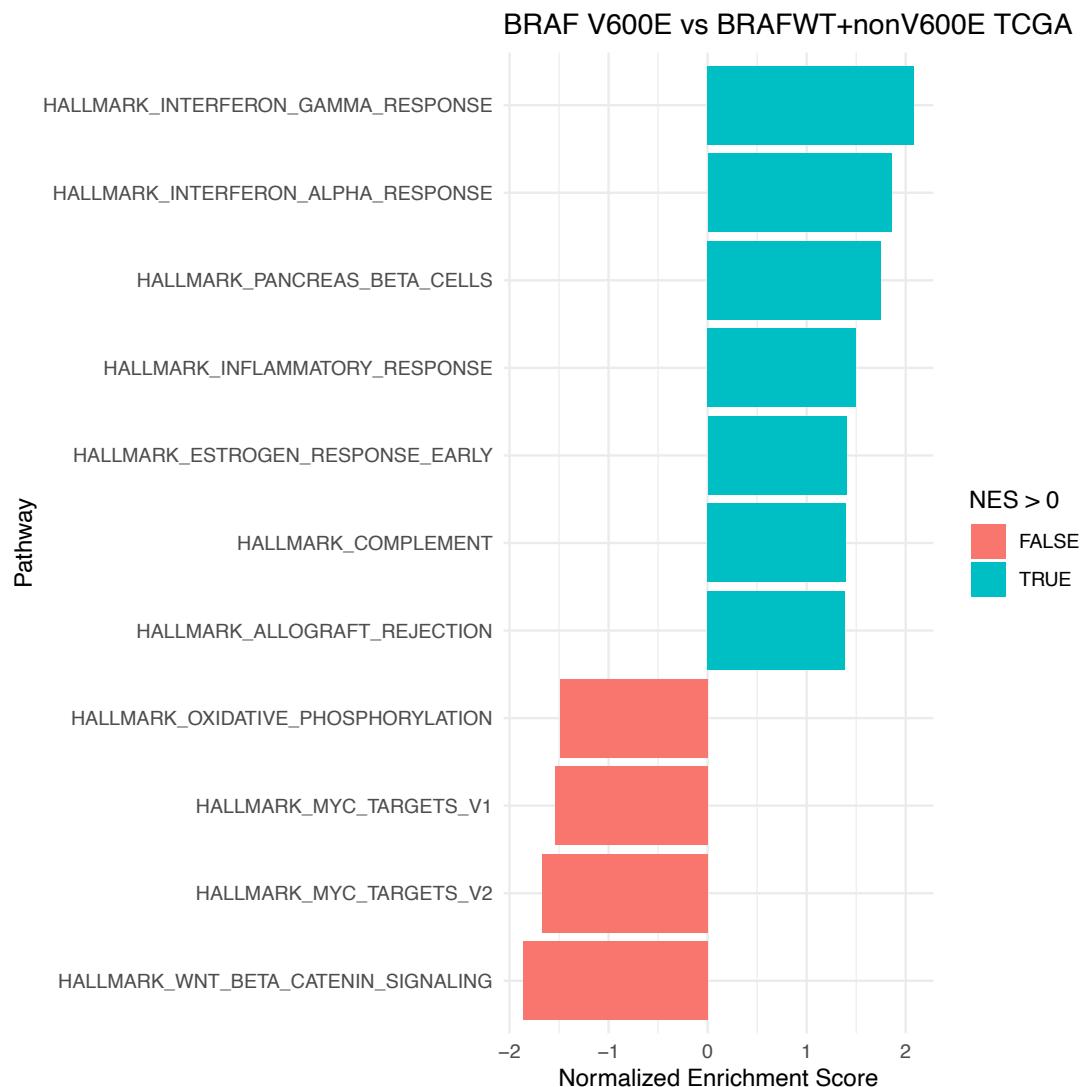


Figure S13. Gene set enrichment analysis (GSEA) of Hallmark pathways in BRAF-V600E vs. BRAF WT + nonV600E tumors (p-adj < 0.05, absolute NES > 1) (TCGA cohort). Bar plot showing the results of GSEA using the Hallmark gene sets from MSigDB, comparing BRAF-V600E to BRAF WT+nonV600E tumors in the TCGA cohort.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

BRAF nonV600E vs BRAF WT TCGA

EnhancedVolcano

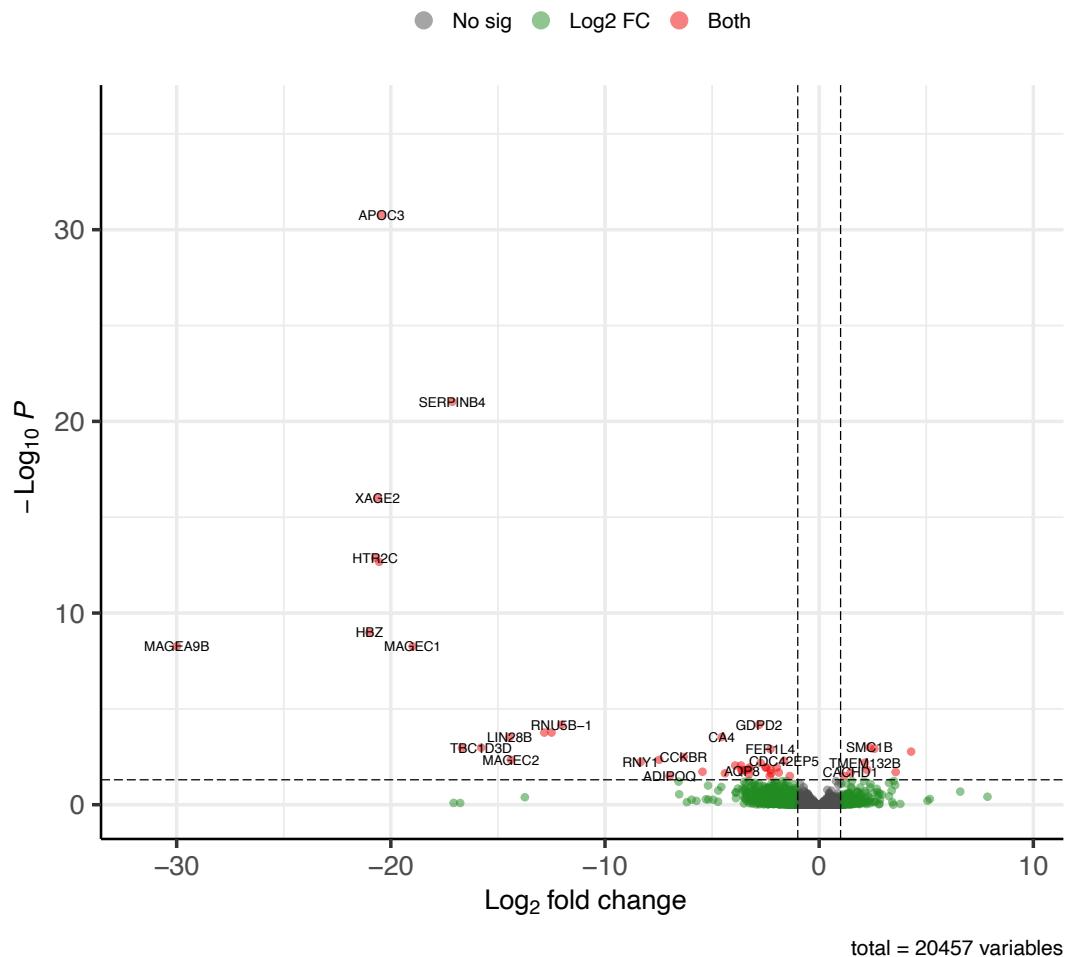


Figure S14. Differential gene expression between BRAF nonV600E and BRAF WT tumors in TCGA. Volcano plot visualizing differential gene expression between BRAF nonV600E and BRAF WT colon cancer tumors in the TCGA cohort. Significantly differentially expressed genes were identified based on an absolute log2 fold change > 1 and an adjusted p-value < 0.05 . Genes meeting both criteria are highlighted in red. Positive log2 fold change are genes upregulated in BRAF nonV600E tumors and negative log2 fold change are genes downregulated in BRAF nonV600E tumors.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

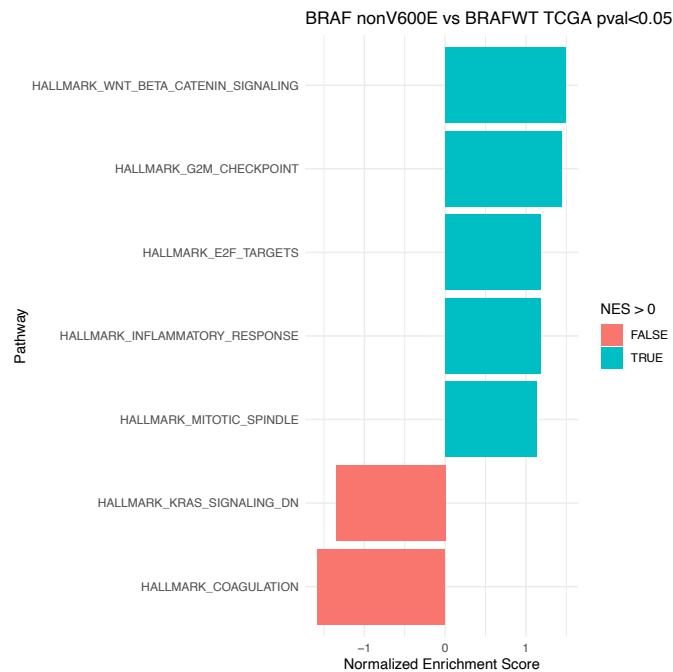


Figure S15. Gene set enrichment analysis (GSEA) of Hallmark pathways in BRAF nonV600E vs. BRAF WT tumors (p-adj < 0.05, absolute NES > 1) (TCGA cohort). Bar plot showing the results of GSEA using the Hallmark gene sets from MSigDB, comparing BRAF nonV600E to BRAF WT tumors in the TCGA cohort.

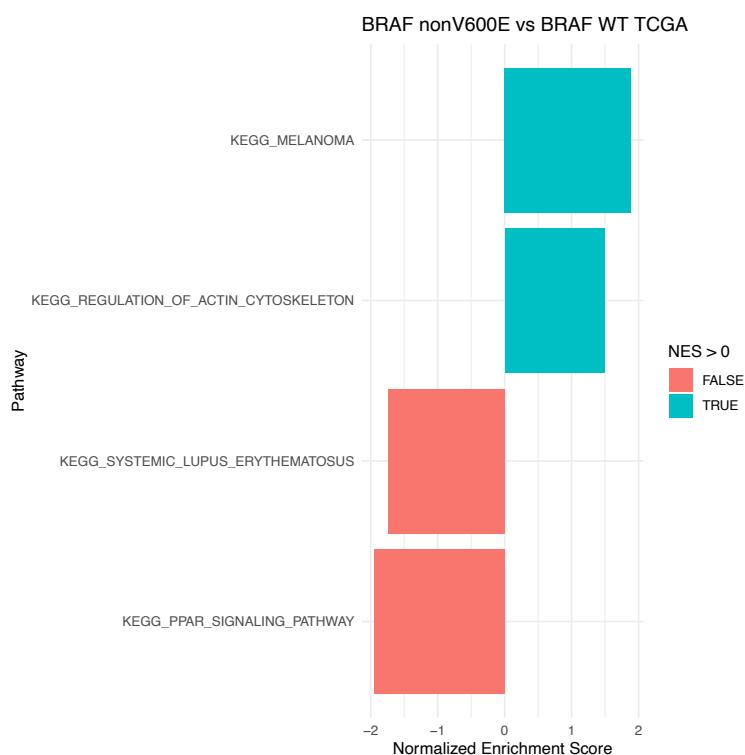


Figure S16. Gene set enrichment analysis (GSEA) of KEGG pathways in BRAF nonV600E vs. BRAF WT tumors (p-adj < 0.05, absolute NES > 1) (TCGA cohort). Bar plot showing the results of GSEA using the KEGG gene sets from MSigDB, comparing BRAF nonV600E to BRAF WT tumors in the TCGA cohort.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

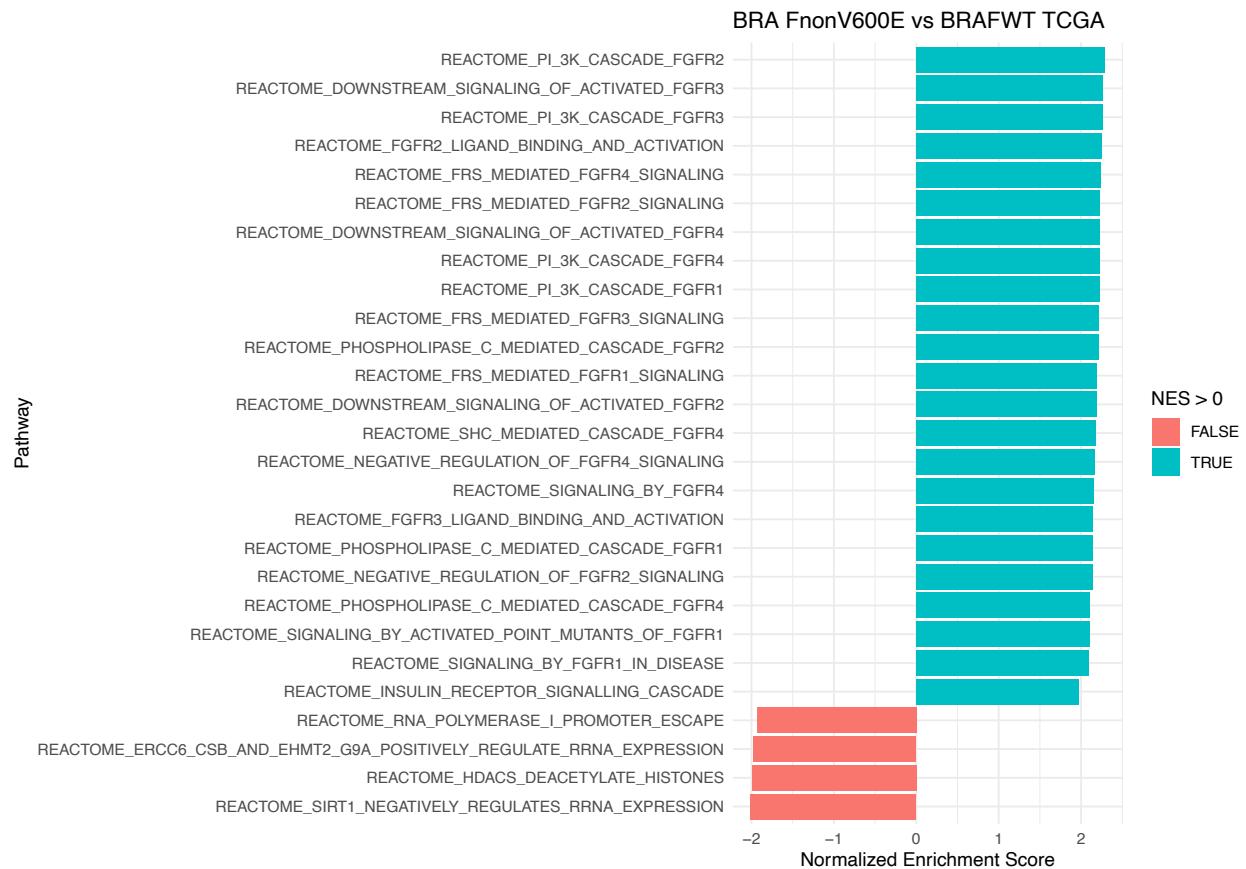


Figure S17. Gene set enrichment analysis (GSEA) of Reactome pathways in BRAF nonV600E vs. BRAF WT tumors (p-adj < 0.05, absolute NES > 1) (TCGA cohort). Bar plot showing the results of GSEA using the Reactome gene sets from MSigDB, comparing BRAF nonV600E to BRAF WT tumors in the TCGA cohort.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

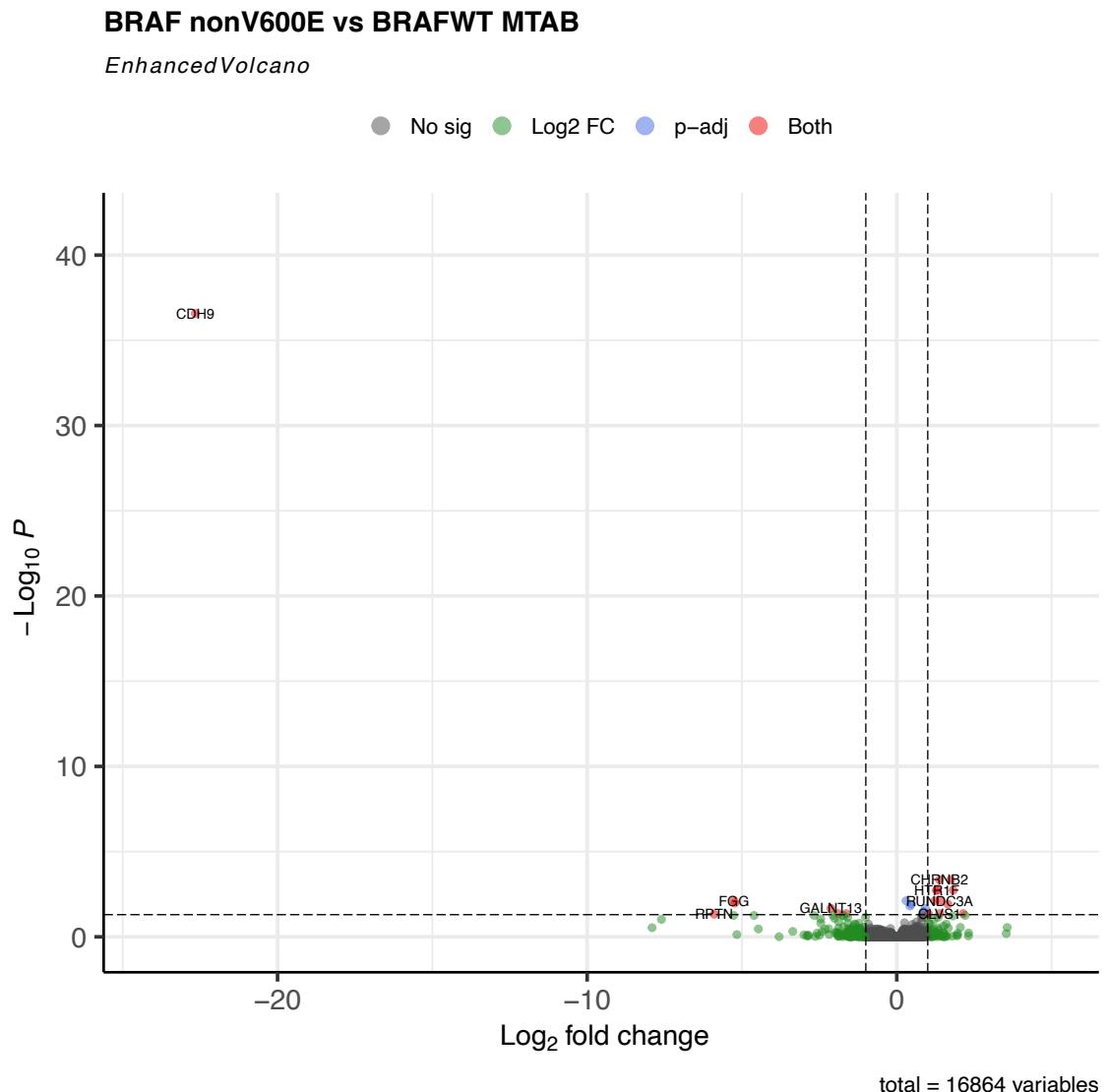


Figure S18. Differential gene expression between BRAF nonV600E and BRAF WT tumors in E-MTAB-12862. Volcano plot visualizing differential gene expression between BRAF nonV600E and BRAF WT colon cancer tumors in the E-MTAB-12862 cohort. Significantly differentially expressed genes were identified based on an absolute log2 fold change > 1 and an adjusted p-value < 0.05 . Genes meeting both criteria are highlighted in red. Positive log2 fold change are genes upregulated in BRAF nonV600E tumors and negative log2 fold change are genes downregulated in BRAF nonV600E tumors.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

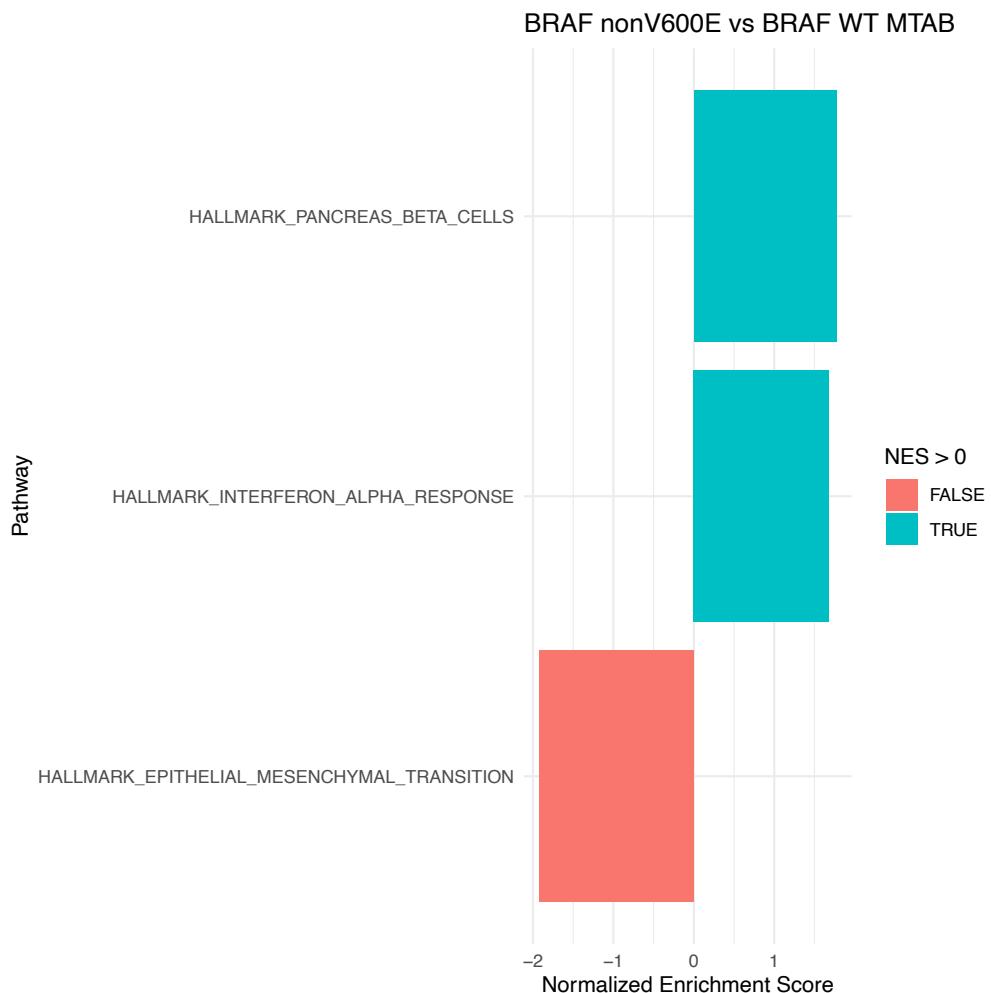


Figure S19. Gene set enrichment analysis (GSEA) of Hallmark pathways in BRAF nonV600E vs. BRAF WT tumors (p-adj < 0.05, absolute NES > 1) (E-MTAB-12862 cohort). Bar plot showing the results of GSEA using the Hallmark gene sets from MSigDB, comparing BRAF nonV600E to BRAF WT tumors in the E-MTAB-12862 cohort.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

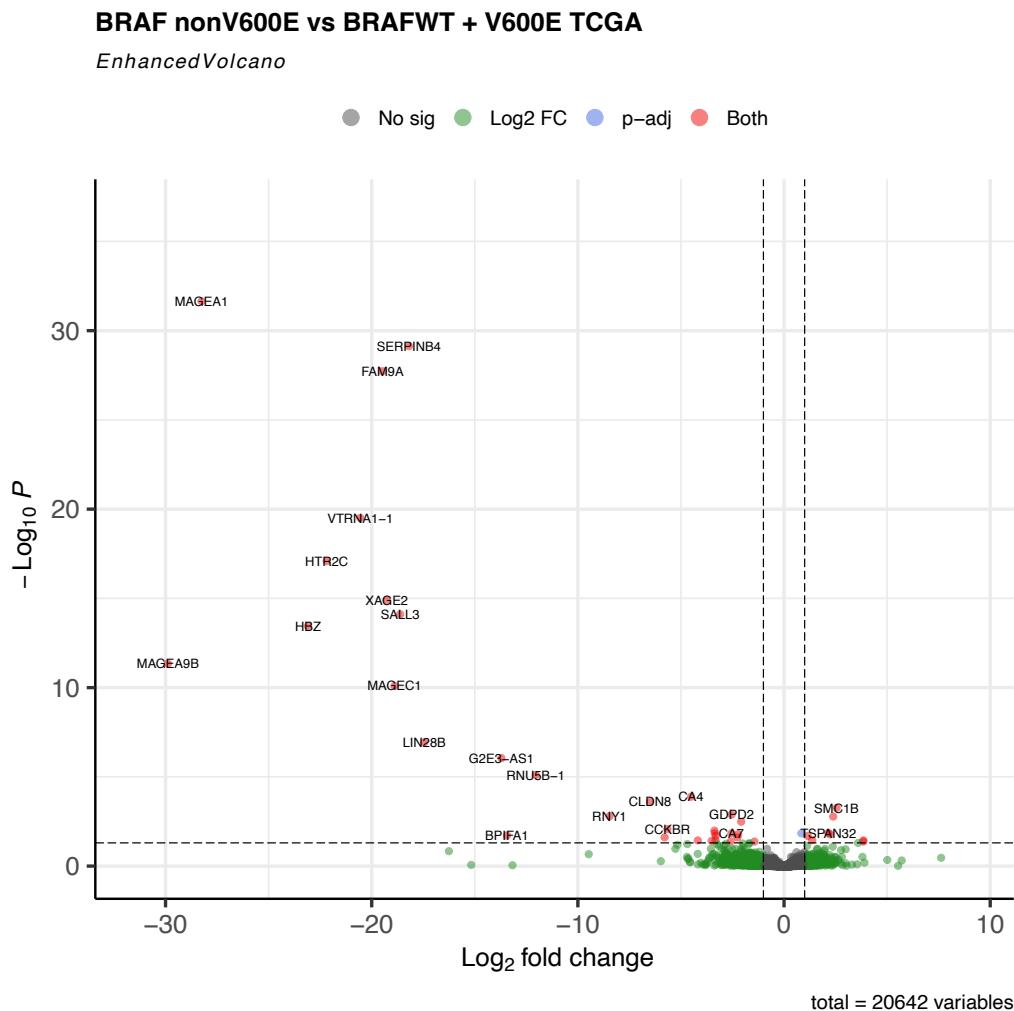


Figure S20. Differential gene expression between BRAF nonV600E and BRAF WT + V600E tumors in TCGA. Volcano plot visualizing differential gene expression between BRAF nonV600E and BRAF WT + V600E colon cancer tumors in the TCGA cohort. Significantly differentially expressed genes were identified based on an absolute log2 fold change > 1 and an adjusted p-value < 0.05 . Genes meeting both criteria are highlighted in red. Positive log2 fold change are genes upregulated in BRAF nonV600E tumors and negative log2 fold change are genes downregulated in BRAF nonV600E tumors.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

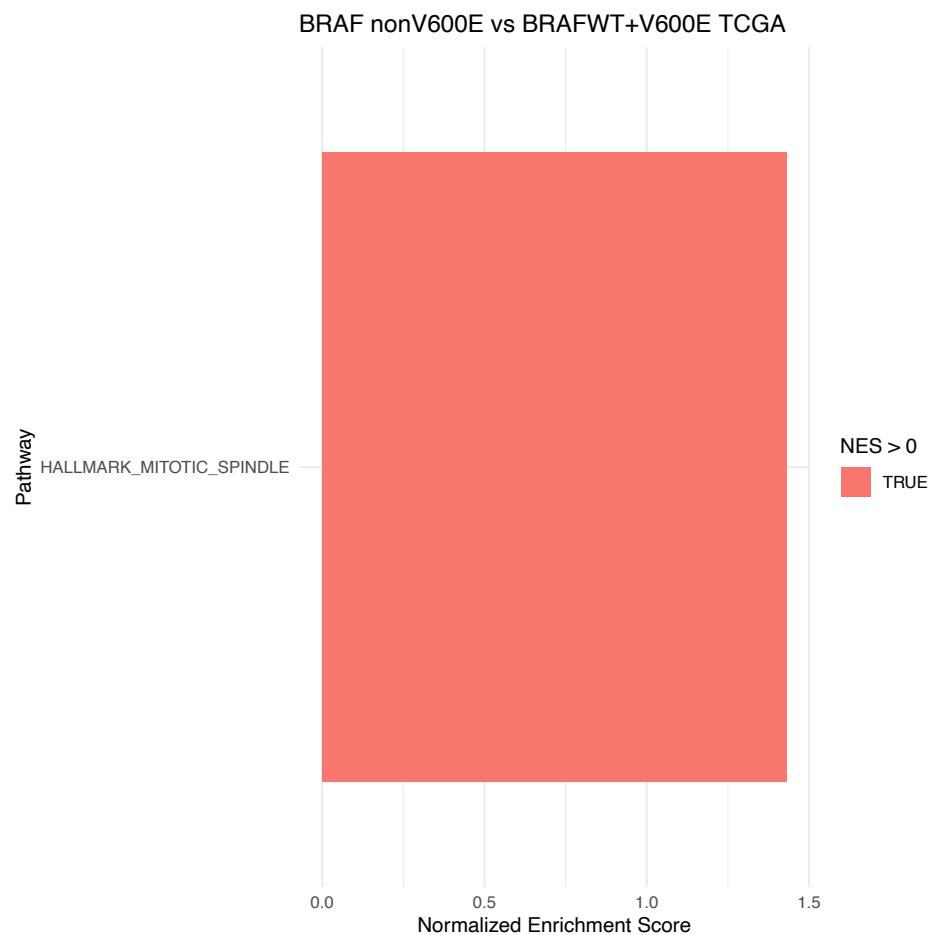


Figure S21. Gene set enrichment analysis (GSEA) of Hallmark pathways in BRAF nonV600E vs. BRAF WT + V600E tumors (p-adj < 0.05, absolute NES > 1) (TCGA cohort). Bar plot showing the results of GSEA using the Hallmark gene sets from MSigDB, comparing BRAF nonV600E to BRAF WT + V600E tumors in the TCGA cohort.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

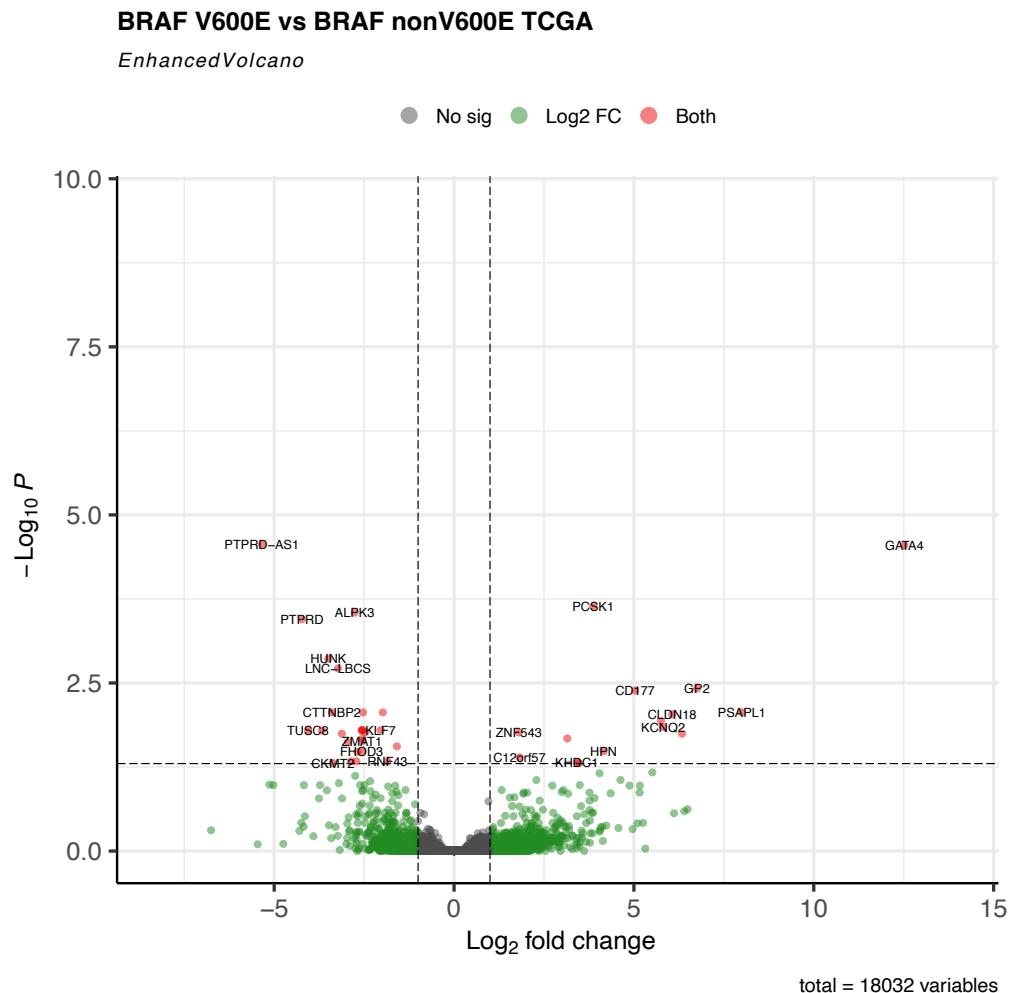


Figure S22. Differential gene expression between BRAF V600E and BRAF nonV600E tumors in TCGA. Volcano plot visualizing differential gene expression between BRAF V600E and BRAF nonV600E colon cancer tumors in the TCGA cohort. Significantly differentially expressed genes were identified based on an absolute log2 fold change > 1 and an adjusted p-value < 0.05 . Genes meeting both criteria are highlighted in red. Positive log2 fold change are genes upregulated in BRAF V600E tumors and negative log2 fold change are genes downregulated in BRAF V600E tumors.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

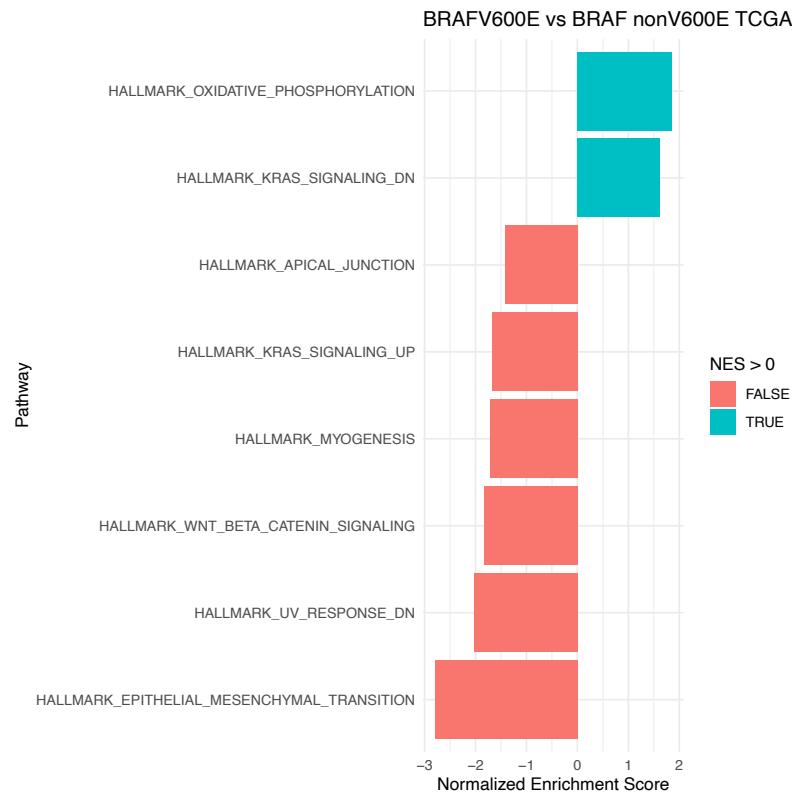


Figure S23. Gene set enrichment analysis (GSEA) of Hallmark pathways in BRAF-V600E vs. BRAF nonV600E tumors (p-adj < 0.05, absolute NES > 1) (TCGA cohort).
Bar plot showing the results of GSEA using the Hallmark gene sets from MSigDB, comparing BRAF-V600E to BRAF nonV600E tumors in the TCGA cohort.

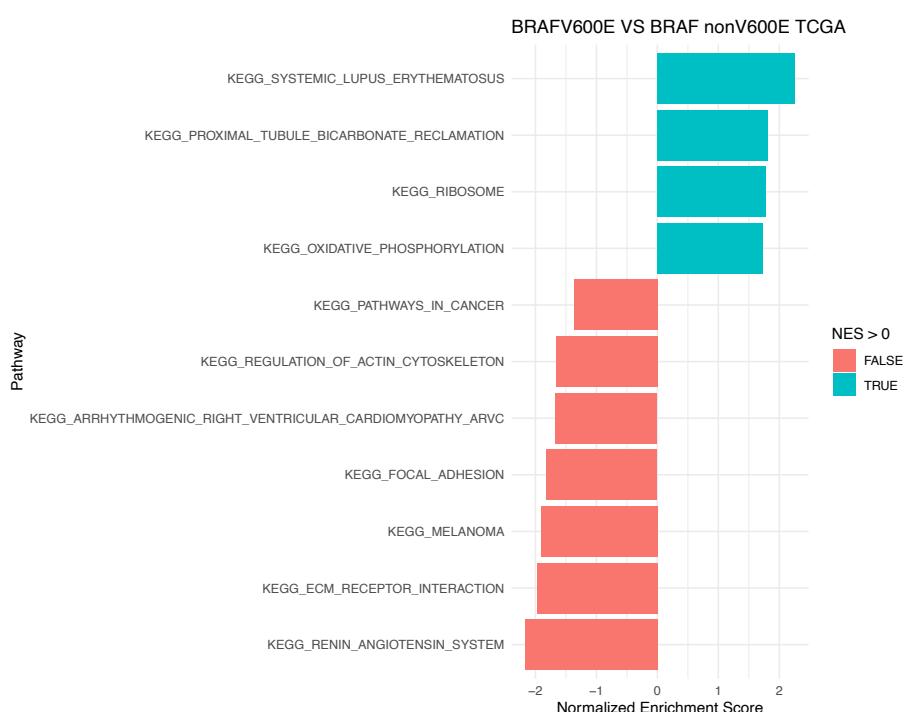


Figure S24. Gene set enrichment analysis (GSEA) of KEGG pathways in BRAF-V600E vs. BRAF nonV600E tumors (p-adj < 0.05, absolute NES > 1) (TCGA cohort). Bar plot

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

showing the results of GSEA using the KEGG gene sets from MSigDB, comparing BRAF-V600E to BRAF nonV600E tumors in the TCGA cohort.

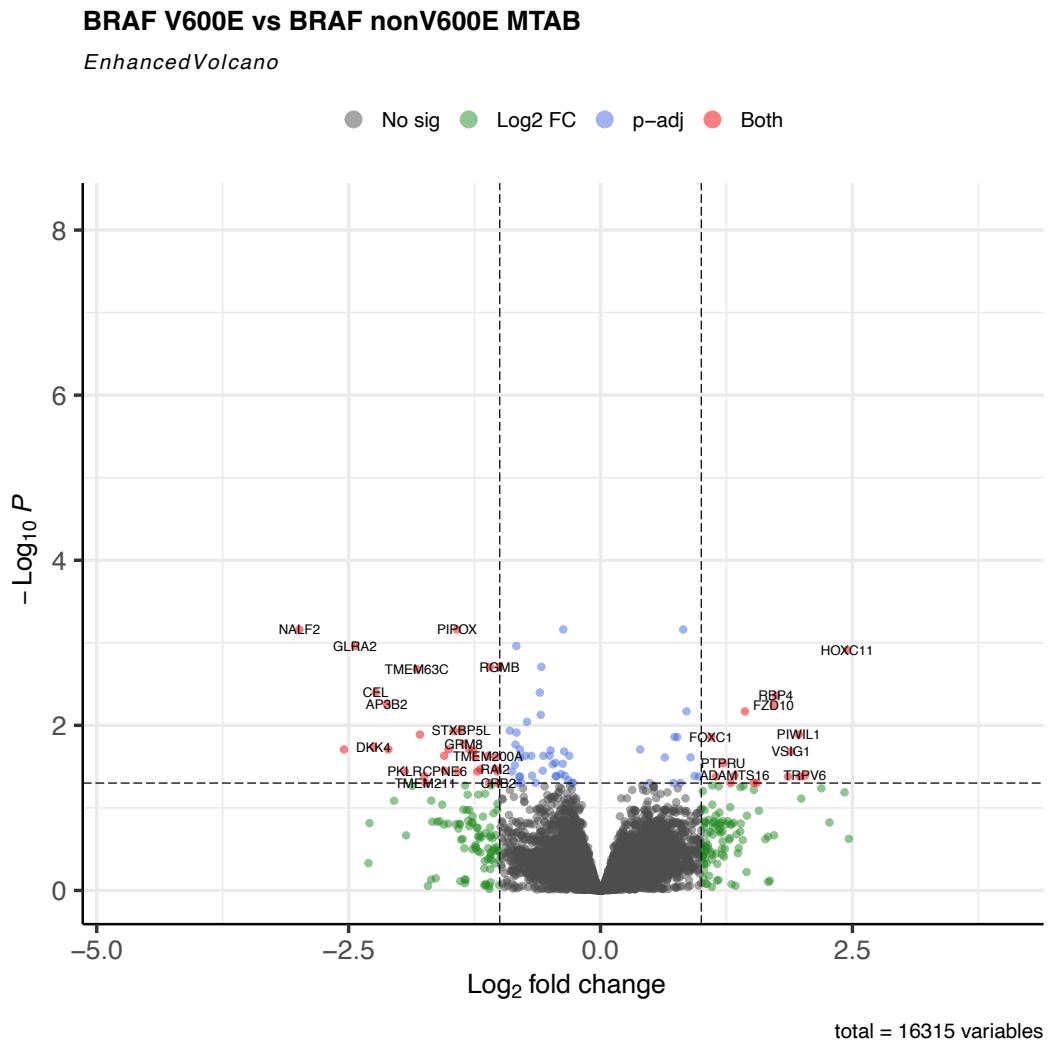


Figure S25. Differential gene expression between BRAF-V600E and BRAF nonV600E tumors in E-MTAB-12862. Volcano plot visualizing differential gene expression between BRAF V600E and BRAF nonV600E colon cancer tumors in the E-MTAB-12862 cohort. Significantly differentially expressed genes were identified based on an absolute log2 fold change > 1 and an adjusted p-value < 0.05 . Genes meeting both criteria are highlighted in red. Positive log2 fold change are genes upregulated in BRAF-V600E tumors and negative log2 fold change are genes downregulated in BRAF-V600E tumors.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

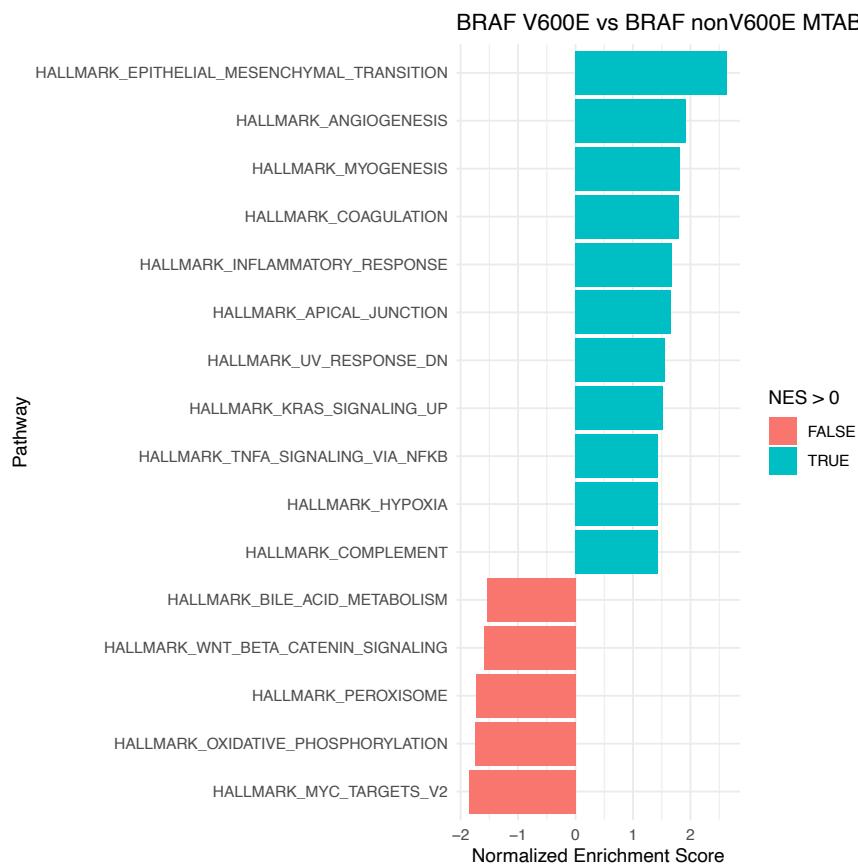


Figure S26. Gene set enrichment analysis (GSEA) of Hallmark pathways in BRAF-V600E vs. BRAF nonV600E tumors (p-adj < 0.05, absolute NES > 1) (E-MTAB-12862 cohort). Bar plot showing the results of GSEA using the Hallmark gene sets from MSigDB, comparing BRAF-V600E to BRAF nonV600E tumors in the E-MTAB-12862 cohort.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

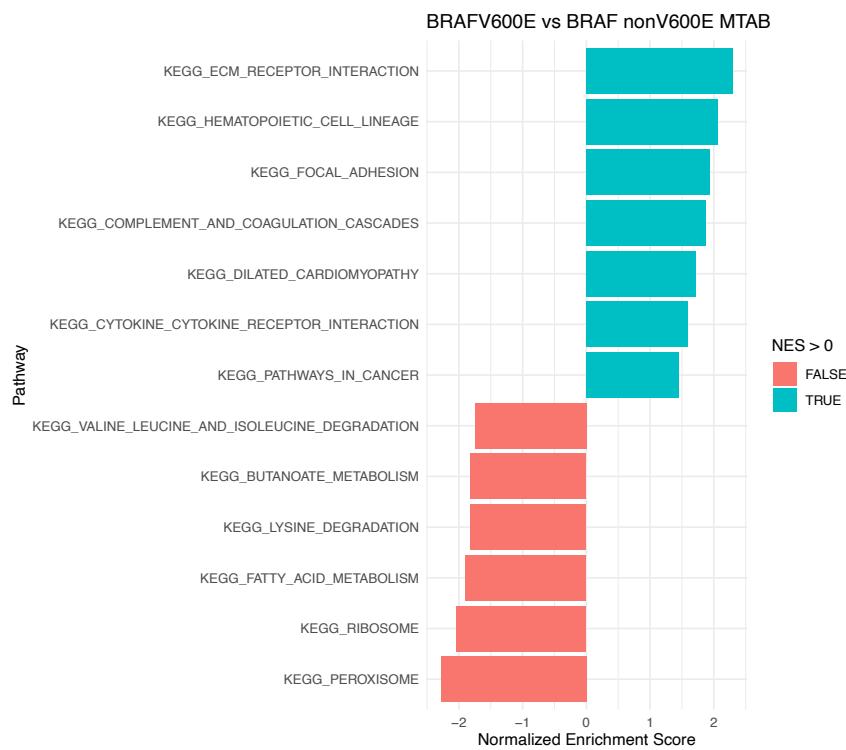


Figure S27. Gene set enrichment analysis (GSEA) of KEGG pathways in BRAF-V600E vs. BRAF nonV600E tumors (p-adj < 0.05, absolute NES > 1) (E-MTAB-12862 cohort).
Bar plot showing the results of GSEA using the KEGG gene sets from MSigDB, comparing BRAF-V600E to BRAF nonV600E tumors in the E-MTAB-12862 cohort.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

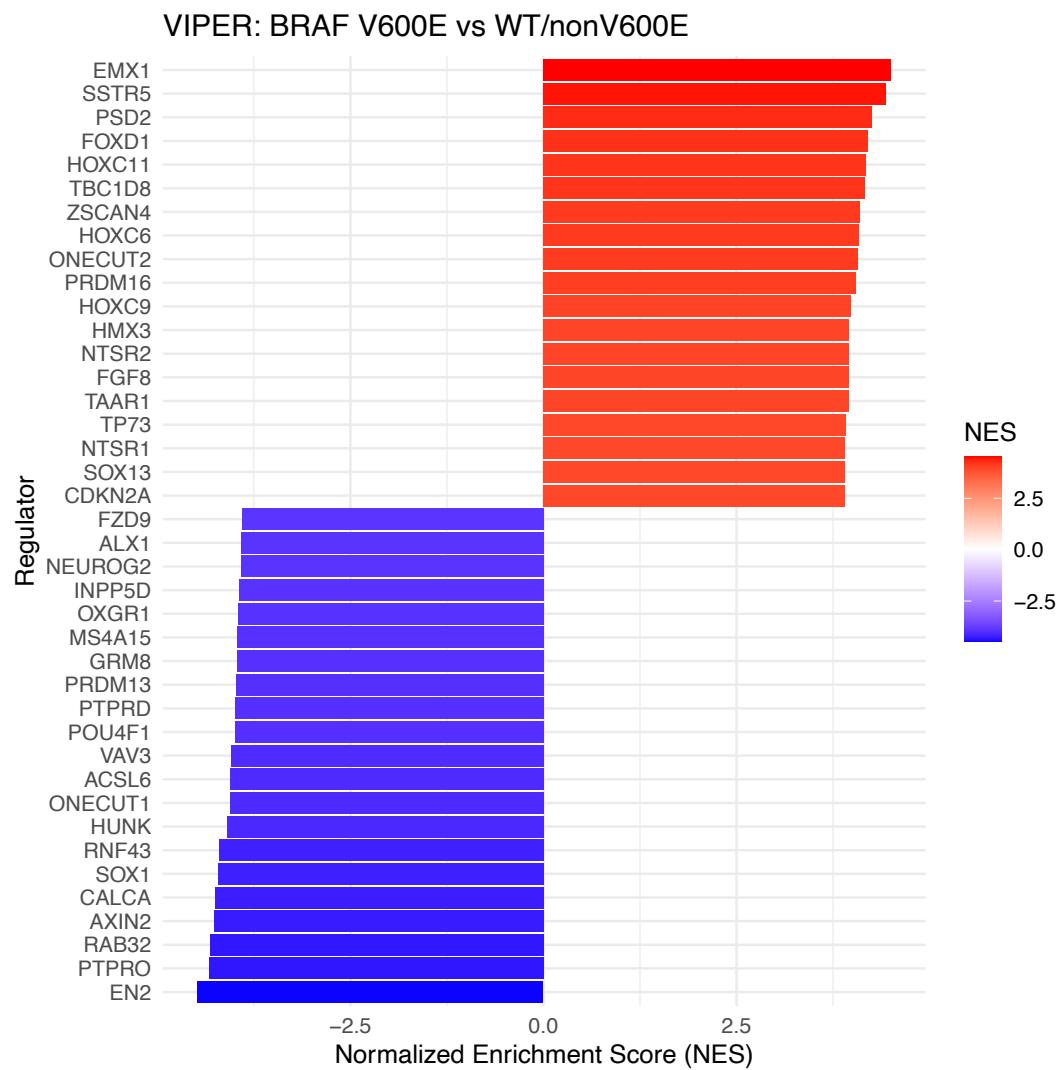


Figure S28. VIPER analysis of regulator activity in BRAF-V600E vs. BRAF WT + nonV600E tumors (E-MTAB-12862) ($p\text{-val} < 0.05$). Bar plot showing the normalized enrichment scores (NES) of inferred regulatory protein activity in BRAF-V600E compared to BRAF WT and nonV600E samples. Red bars indicate regulators increased activity in BRAF-V600E tumors, while blue bars represent regulators with decreased activity. VIPER analysis was based on ARACNe-inferred networks, and only significantly differentially active regulators are displayed ($p\text{-val} < 0.05$)

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

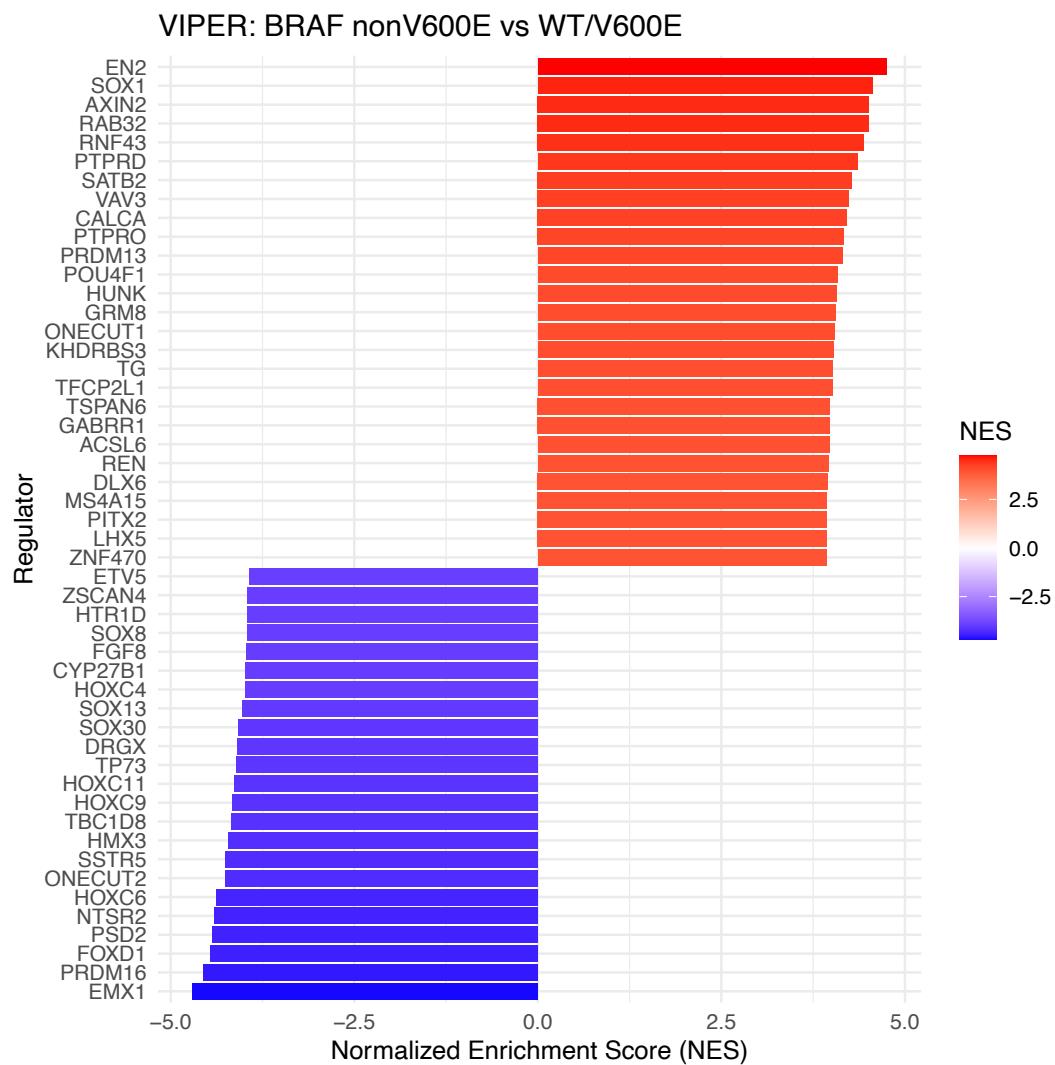


Figure S29. VIPER analysis of regulator activity in BRAF nonV600E vs. BRAF WT + V600E tumors (E-MTAB-12862) ($p\text{-val} < 0.05$). Bar plot showing the normalized enrichment scores (NES) of inferred regulatory protein activity in BRAF nonV600E compared to BRAF WT and V600E samples. Red bars indicate regulators increased activity in BRAF nonV600E tumors, while blue bars represent regulators with decreased activity. VIPER analysis was based on ARACNe-inferred networks, and only significantly differentially active regulators are displayed ($p\text{-val} < 0.05$)

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

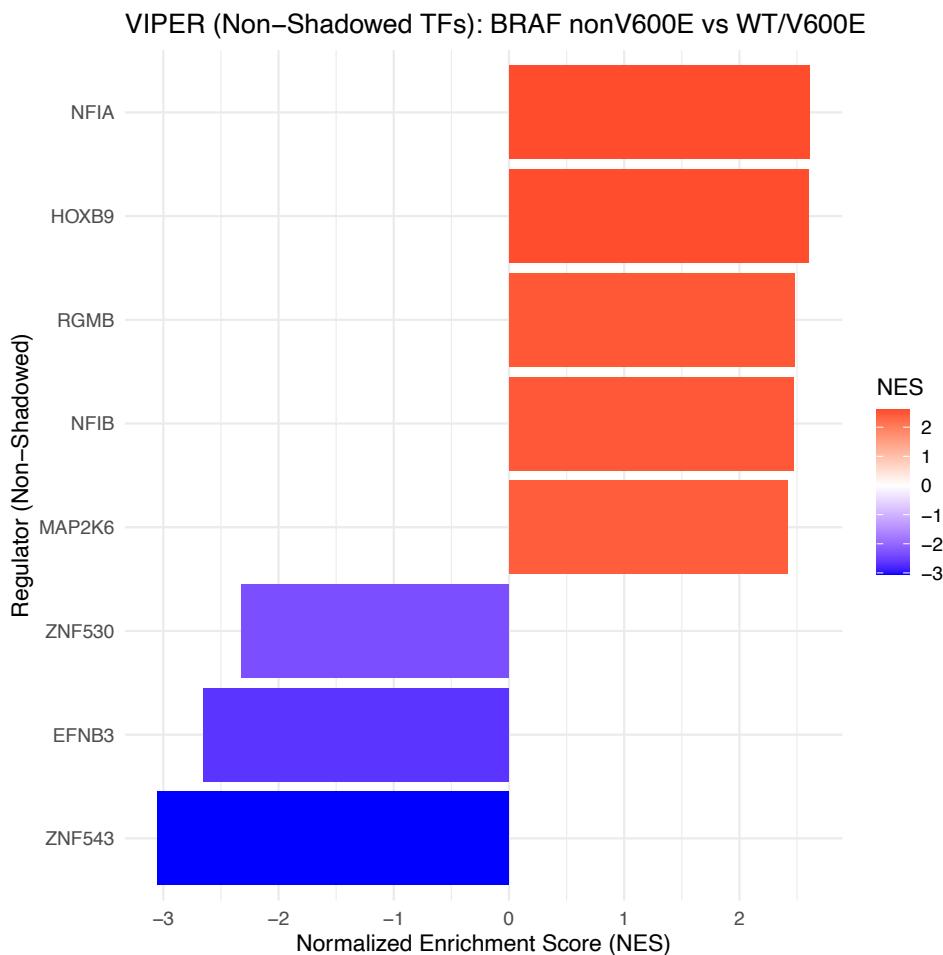


Figure S30. VIPER analysis of non-shadowed regulator activity in BRAF nonV600E vs. BRAF WT + V600E tumors (E-MTAB-12862). Bar plot displaying non-shadowed regulators with significantly altered activity in BRAF nonV600E vs BRAF WT + V600E tumors. Red bars represent regulators with increased activity in BRAF nonV600E tumors, while blue bars represent those with decreased activity. Only regulators that passed shadow analysis filtering are shown.

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

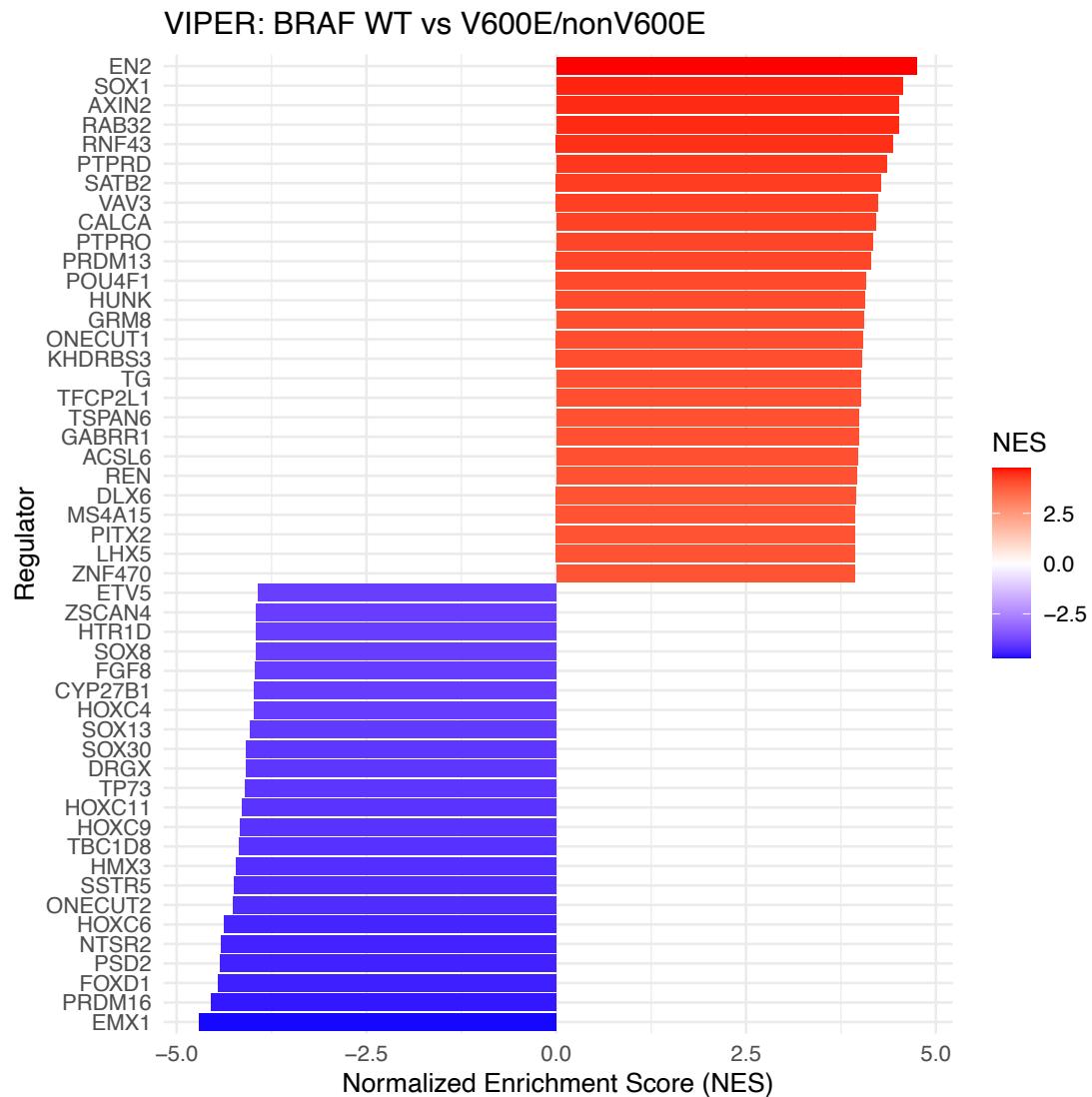


Figure S31. VIPER analysis of regulator activity in BRAF WT vs. BRAF V600E + nonV600E tumors (E-MTAB-12862) ($p\text{-val} < 0.05$). Bar plot showing the normalized enrichment scores (NES) of inferred regulatory protein activity in BRAF WT compared to BRAF V600E and nonV600E samples. Red bars indicate regulators increased activity in BRAF WT tumors, while blue bars represent regulators with decreased activity. VIPER analysis was based on ARACNe-inferred networks, and only significantly differentially active regulators are displayed ($p\text{-val} < 0.05$)

Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers

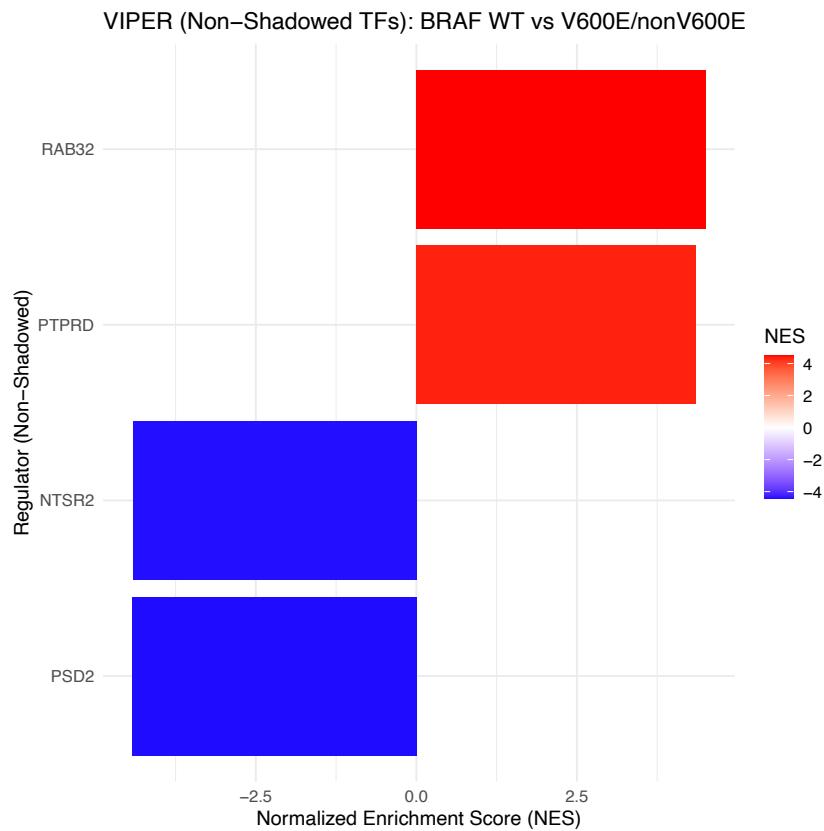


Figure S32. VIPER analysis of non-shadowed regulator activity in BRAF WT vs. BRAF V600E + nonV600E tumors (E-MTAB-12862). Bar plot displaying non-shadowed regulators with significantly altered activity in BRAF WT vs BRAF V600E+ nonV600E tumors. Red bars represent regulators with increased activity in BRAF WT tumors, while blue bars represent those with decreased activity. Only regulators that passed shadow analysis filtering are shown.