MEMÒRIA DEL TREBALL DE FI DE GRAU DEL GRAU (ESCI-UPF)

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

AUTOR/A: Taras Yuziv Duda	**NIA:** 107748
GRAU: Bachelor's Degree in Bioinformatics (BDBI)	
CURS ACADÈMIC: 2024-2025	
DATA: 18/06/2025	
TUTOR/S: Lea Lemler and Jose A. Seaone	

FULL DE RESUM DEL TREBALL DE FI DE GRAU DEL BDBI (ESCI-UPF)

TÍTOL DEL PROJECTE: Transcriptional and Regulatory Divergence Between BRAF-V600E and non-V600E Colon Cancers	
AUTOR/A: Taras Yuziv Duda	**NIA:** 107748
CURS ACADÈMIC: 2024-2025	
DATA: 18/06/2025	
TUTOR/S: Lea Lemler and Jose A. Seoane	

PARAULES CLAU (mínim 3)

- Català: Càncer colorectal, Mutació de BRAF, Transcriptòmica, Perfil d'expressió gènica, Interferència de xarxes reguladores, Subtipus tumorals

- Castellà: Cáncer colorrectal, Mutación en BRAF, Transcriptómica, Perfil de expresión génica, Inferencia de redes regulatorias, Subtipos tumorales

- Anglès: Colorectal cancer, BRAF mutation, Transcriptomics, Gene expression profiling, Regulatory network inference, Tumor subtypes

RESUM DEL PROJECTE (extensió màxima: 100 paraules per llengua)

- Català: Aquest projecte investiga les diferències transcriptòmiques i reguladores entre tumors de còlon amb mutacions BRAF-V600E, BRAF no-V600E i BRAF WT. Utilitzant dades d'RNA-seq i mutacions de les cohorts TCGA-COAD i E-MTAB-12862, vam dur a terme un anàlisi d'expressió diferencial i vam inferir l'activitat dels reguladors amb l'algoritme VIPER. Els tumors amb BRAF-V600E mostraren una forta activació immune i desregulació dels programes epitelials, mentre que els tumors amb mutacions no-V600E eren més heterogenis. L'anàlisi de xarxes reguladores va identificar conjunts diferenciats de factors de transcripció i proteïnes de senyalització segons el subtipus tumoral

- Castellà: Este proyecto investiga las diferencias transcriptómicas y regulatorias entre tumores de colon con mutacions BRAF-V600E, BRAF no-V600E y BRAF WT. Utilizando datos de RNA-seq y mutaciones de las cohortes TCGA-COAD y E-MTAB-12862, realizamos un anàlisis de expresión diferencial e inferimos la actividad de reguladores mediante el algoritmo VIPER. Los tumores con BRAF-V600E mostraron señales inmunes fuertes y programas epiteliales desregulados, mientras que los tumores con mutaciones no-V600E fueron más heterogéneos. El anàlisis de redes regulatorias identifico conjuntos distintos de factores de transcripcion y proteinas de señalizacion según el subtipo tumoral.

- Anglès: This project investigates transcriptional and regulatory differences between BRAF-V600E, BRAF non-V600E, and BRAF WT colon tumors. Using RNA-seq and mutation data from TCGA-COAD and E-MTAB-12862, we performed differential expression

analysis and inferred regulator activity via the VIPER algorithm. BRAF-V600E tumors showed strong immune signals and deregulated epithelial programs, while non-V600E tumors were more heterogeneous. Regulatory network analysis identified distinct sets of active transcription factors and signaling proteins across the diferent tumor subtypes.

ESCI upf.
School of International Studies

Bachelor's Degree in Bioinformatics (UPF-UPC-UB)

Final Grade Project

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

Taras Yuziv Duda

Scientific directors: Lea Lemler[1], Jose A. Seoane[1]

[1]Cancer Computational Biology Group, Vall d'Hebron Institute of Oncology, Barcelona, Spain

Abstract

Colorectal cancer (CRC) is a genetically diverse disease, and mutations in the *BRAF* gene (particularly the *BRAF-V600E* variant) are linked to aggressive behavior and poor outcomes. While *BRAF-V600E* tumors have been studied extensively, less is known about *non-V600E BRAF* mutations and how they differ at the molecular level. In this study, we analyzed gene expression and regulatory programs across *BRAF-V600E*, *non-V600E*, and *BRAF wild-type* tumors using two independent colon cohorts. We found that *BRAF-V600E* tumors show strong immune activation and loss of epithelial regulators, consistent with dedifferentiation and transcriptional reprogramming. In contrast, *non-V600E* tumors displayed more heterogeneous profiles but retained more features of epithelial identity and WNT signaling. *BRAF wild-type* tumors showed the most stable regulatory landscape. Our results distinguish biological trajectories within *BRAF-mutant* colon cancer and support the need for subtype-specific therapeutic strategies.

1 Introduction

Colorectal cancer (CRC) is a diverse disease resulting from the buildup of genetic, epigenetic, and transcriptomic changes that disturb cellular balance and promote malignant transformation. Worldwide, CRC is the third most diagnosed cancer and the second foremost cause of cancer-related mortality, with approximately 1.9 million new cases and more than 934,000 fatalities noted in 2020 (Sung et al., 2021). This load is projected to rise significantly in the forthcoming years, highlighting the necessity of comprehending the molecular mechanisms that influence tumor progression and treatment response (Arnold et al., 2017).

The development of CRC is traditionally outlined by the multistep adenoma-carcinoma sequence suggested by Vogelstein and his team, identified by the initial inactivation of *APC*, activating mutations in *KRAS*, and eventual loss of *TP53* (Vogelstein et al., 1988). However, Vogelstein model does not capture the full complexity of CRC biology. Tumors originating in the proximal (right) colon, in particular, frequently diverge from this sequence (Baran et al., 2018). These right-sided tumors are often defined by a distinct molecular profile, including high rates of microsatellite instability (MSI), the CpG island methylator phenotype (CIMP), and, most notably, activating mutations in the *BRAF* oncogene (Jass, 2007) (Toyota et al., 1999). Additionally, CRC is stratified into four consensus molecular subtypes (CMS1-4) with distinct clinical and biological features, with CMS1, characterized by immune activation, often associated with right-sided, MSI-H, and BRAF-mutated tumors (Guinney et al., 2015). MSI, resulting from defects in the DNA mismatch repair system, is observed in ~15% of CRCs and is particularly common among older patients, females, and those with right-sided tumors (Lachit & Vinita, 2022). Among these characteristics, *BRAF* mutations (specifically *BRAF-V600E*) are frequently associated with more aggressive cancers and a worse prognosis for patients (Barras et al., 2017).

The *BRAF-V600E* variant, found in around 8-12% of metastatic CRC (mCRC) cases, is clinically significant as it correlates with aggressive disease, poor prognosis and resistance to standard first-line treatments, including chemotherapy paired with *EGFR* inhibitors (Guerrero et al., 2022). The replacement of valine with glutamic acid at codon 600 results in the continuous activation of the *MAPK* signaling pathway, enhancing tumor cell growth and resistance to apoptosis (Yao et al., 2017). *BRAF-V600E*-mutant CRCs are typically found on the right side, exhibit CIMP-high levels and show high MSI (MSI-H) (Margonis et al., 2018) (Baran et al., 2018).

To address this treatment resistance, several targeted combination strategies have been explored in clinical trials. The biggest of these was the BEACON CRC trial which evaluated a triplet therapy using a *BRAF* inhibitor (encorafenib), a *MEK* inhibitor (binimetinib), and an *EGFR* inhibitor (cetuximab) (Kopetz et al., 2024a). While this trial showed improvements in survival and response rates compared to standard chemotherapy, most patients eventually developed resistance, and the underlying mechanisms remain incompletely understood (Kopetz et al., 2024a).

Importantly, not all *BRAF* mutations have the same biological behavior or clinical trajectory. In contrast to *BRAF-V600E*, *non-V600E BRAF* mutations (approximately 2-3% of CRCs) represent a functionally diverse subgroup (Yaeger et al., 2019). These alterations divide into two classes: class II mutations, which exhibit constitutive dimerization and are *RAS*-independent, and class III mutations, which are kinase-impaired but *RAS*-dependent (Yaeger et al., 2019). Clinically, patients with *non-V600E BRAF* mutations tend to be younger, more often present with left-sided tumors, and exhibit significantly better prognoses than those with *BRAF-V600E* mutations (Jones et al., 2017). Despite these observations, the molecular and transcriptional characteristics for these clinical outcomes remain poorly understood. (Kopetz et al., 2024a).

While gene expression profiling has shown that cancer-driving mutations like *BRAF-V600E* can reshape how genes are regulated, we still know very little about what actually controls these changes (Bradner et al., 2017). For example, *BRAF-V600E* tumors can be grouped into two subtypes–*BRAF* mutant 1 (BM1) and *BRAF* mutant 2 (BM2)–that have different signaling and immune profiles, but the reason behind these differences remain poorly understood (Barras et al., 2017). That's why the activity of regulator proteins like transcription factors (TF), signaling proteins and chromatin regulatory genes (CRG) could potentially explain how these tumors develop. However, their functional activity isn't always reflected in RNA levels alone. To do this, we can apply tools that go beyond traditional gene expression analysis and instead focus on the regulatory networks driving these changes (Margolin et al., 2006).

1.1 Objectives

We hypothesized that *BRAF-V600E* and *non-V600E* mutations engage in different regulatory programs that contribute to their different clinical trajectories and therapeutic responses. To investigate this, we aimed to look into the transcriptional and regulatory landscape of *BRAF-V600E* and *non-V600E* colon cancers. By integrating gene expression data from multiple cohorts, we analyzed differential expression profiles and pathway enrichments between *BRAF-V600E, BRAF non-V600E* and *BRAF WT*. Primary analyses were conducted using The Cancer Genome Atlas Colorectal Adenocarcinomas (TCGA-COAD) (Muzny et al., 2012a) dataset as a discovery cohort, with replication in the independent E-MTAB-12862 cohort (Nunes et al., 2024). In addition, we applied the VIPER algorithm to infer the activity of transcriptional regulators, signaling proteins and chromatin-modifying enzymes driving these phenotypes.

2 Methods

This study used RNA-Sequencing (RNA-seq) and mutational data from publicly available CRC datasets to investigate transcriptional and regulatory differences between *BRAF-V600E*, *BRAF non-V600E* and *BRAFWT* tumors. Primary analyses were performed on TCGA-COAD dataset, with validation in the E-MTAB-12862 cohort (Muzny et al., 2012b) (Nunes et al., 2024). Samples were categorized into three groups on *BRAFmt* status: *BRAF-V600E, BRAF non-V600E BRAF* mutations, and *BRAF WT*.

Data Acquisition and Processing

From the TCGA cohort, RNA-seq raw count and TPM counts data were retrieved using the TCGAbiolinks R package (v2.35.3), along with the somatic mutation annotation files (MAF format) (*TCGAbiolinks*, n.d.). In contrast, the E-MTAB-12862 dataset provided raw Variant Call Format (VCF) files, without centralized mutation annotations. To identify *BRAFmt* samples in this cohort, we first extracted all relevant mutational lines, targeting entries containing the *BRAF* gene. These filtered lines were then imported into R and converted into a structured data frame. RNA-seq raw counts from the E-MTAB-12862 cohort were extracted from ArrayExpress (BioStudies, n.d.). Only primary tumor samples were included in both cohorts.

Gene expression data were processed using the DESeq2 package (v1.36.0) (Love et al., 2014). Raw counts were normalized using the median-of-ration method (DESeq2 built-in normalization) to account for differences in sequencing depth across samples. To minimize noise and improve computational efficiency, genes with fewer than 10 counts in the smallest sample group were filtered out, following DESeq2 documentation (*DESeq2*, n.d.). Additionally, samples in the E-MTAB-12862 cohort that derived from rectal tumors were excluded to focus the analysis on primary adenocarcinomas. TCGA only included colon samples.

Cohort description

After stratifying for *BRAFmt* status and filtering for gene expression availability and rectal tumors, 405 samples from TCGA and 782 samples from the E-

MTAB-12862 cohort were retained for analysis. The TCGA cohort included 45 *BRAFV600E*, 11 *non-V600E BRAF*, and 349 *BRAF WT* tumors. In the E-MTAB-12862 dataset, 196 samples were classified as *BRAF-V600E*, 28 as *non-V600E BRAF*, and 558 as *BRAF WT* (Supplementary Table 1).

Sex distribution showed a female predominance in the *BRAFV600E* group in both cohorts (71.1% in TCGA, 68.9% in E-MTAB-12862), while *BRAF WT* tumors were more balanced. *Non-V600E BRAF* cases showed greater heterogeneity but leaned toward a female predominance in E-MTAB-12862 (approximately 71%).

An enrichment of MSI-H was observed among *BRAFV600E* tumors: 80.0% in TCGA and 77.0% in E-MTAB-12862. *Non-V600E BRAF* mutants also showed a high frequency of MSI-H (36.4% in TCGA; 71.4% in E-MTAB-12862), much greater than in *BRAF WT* tumors (10.6% in TCGA; 8.8% in E-MTAB-12862).

Tumor location showed more samples toward right-sided origin in *BRAFV600E* cases: 93.3% of TCGA and 90.3% of E-MTAB-12862 *BRAFV600E* tumors arose in the right colon. A similar pattern was observed in *non-V600E BRAF* tumors, with 81% of E-MTAB-12862 cases located in the right colon. *BRAF WT* tumors had a more even distribution between left and right.

A complete breakdown of clinical and molecular characteristics stratified by *BRAF* mutation subtype is presented in Supplementary Table S1.

Microsatellite Instability Inference

MSI status was missing for a subset of TCGA samples (n=12). To overcome this, I applied the PreMSI classifier (v1.0.0), which predicts MSI status from transcriptomic data based on a validated 15-gene expression panel (Li et al., 2020). This method was selected due to its robustness and demonstrated accuracy across TCGA colon, gastric, and endometrial cancers (Li et al., 2020).

Tumor Location Prediction

Tumor location (right-sided vs. left-sided) was not fully annotated in the TCGA dataset (n=14). To infer this inconsistency between samples, we first performed an exploratory analysis including Principal Component Analysis (PCA) using the prcomp() function in R in both cohorts. Then, developed a machine learning predictive classifier using a shrunken centroid approach using the pamr package (v1.57) (Hastie et al., 2024). The model was trained on the E-MTAB-12862 cohort, where all of the samples were annotated for location. The raw counts were normalized using Variance Stabilizing Transformation (VST) before training.

The classifier was trained via pamr.train(), and the optimal number of genes and classification threshold were determined using cross-validation error plots (pamr.cv()). To evaluate the generalizability of the model, we tested it on the subset of TCGA samples with known tumor location annotations. Performance was assessed using the confusionMatrix() function from the caret package (v6.0-93) (Kuhn, n.d.), reporting accuracy, sensitivity, and specificity. Once validated, the final model was applied to infer tumor location for the remaining unannotated TCGA samples.

Differential Gene Expression Analysis

To investigate transcriptional alterations associated with *BRAF* mutational subtypes, three differential gene expression (DGE) analyses were performed, including 1) *BRAF-V600E* vs. *BRAF WT*, 2) *BRAF non-V600E* vs. *BRAF WT*, and 3) *BRAF-V600E* vs. *BRAF non-V600E* using the DESeq2 package (v1.36.0) (Love et al., 2014). All models were adjusted for gender, MSI status, tumor purity, age and tumor site. Additionally, we implemented a "one-vs-rest" strategy to better capture subtype-specific expression patterns. In this framework, each *BRAF* category was compared against the union of the other two groups. Including 1) *BRAF-V600E* vs. *non-V600E* + *BRAF WT*, 2) *non-V600E* vs. *BRAF-V600E* + *BRAF WT*, and *BRAF WT* vs. *BRAF-V600E* + *BRAF nonV600E*. Genes were considered significantly differentially expressed if they had an adjusted p-value (p-adj) < 0.05 and an absolute log2 fold change greater than 1 (|log2FC| > 1). Results were adjusted for multiple testing using the Benjamini Hochberg procedure for false discovery rate (FDR) (Benjamini & Hochberg, 1995).

Gene Set Enrichment Analysis

To interpret the biological relevance of differentially expressed genes, Gene Set Enrichment Analysis (GSEA) was conducted using the fgsea package (v1.20.0) (Korotkevich et al., 2021). Genes were ranked by the Wald statistic from DESeq2, preserving the direction of effect size to distinguish between upregulated and downregulated pathways. Enrichment was tested in three curated gene set collections from the Molecular Signature Database (MSigDB), including Hallmark gene sets, KEGG pathways, and Reactome pathways (Liberzon et al., 2015) (Kanehisa & Goto, 2000) (Croft et al., 2011). Significantly enriched pathways were defined based on adjusted p-value <0.05 and absolute normalized enrichment score (NES) > 1.

Regulatory Network Analysis: ARACNe, VIPER

To identify upstream regulatory programs driving the transcriptional differences across *BRAFmt* subtypes, we inferred the activity of TFs , CRGs and signaling proteins using the VIPER algorithm (Alvarez et al., 2016). This method estimates the activity of a regulatory protein by analyzing the coordinated expression of its downstream targets, collectively known as a regulon.

We first constructed custom regulatory networks (regulons) using the ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks) algorithm (Margolin et al., 2006). The input expression matrix for ARACNe was constructed from TPM values for TCGA and E-MTAB-12862 separately, which were log2-transformed. While not explicitly required by the ARACNe or VIPER algorithms, this transformation is commonly applied in network inference pipelines for RNA-seq (Alvarez et al., 2016).

To define candidate regulators, we first extracted a list of TFs/signaling proteins from a Colon Adenocarcinoma context-specific regulatory networks from the arcane.networks library using *names(reguloncoad)* (*Aracne.Networks*, n.d.) and a curated custom list of CRGs based on in-house annotations. The ARACNe pipeline (*Califano-Lab/ARACNe-AP*, 2017/2025) was run using default parameters with 1000 bootstrap iterations, using a mutual information threshold of 1×10^{-8}, calculated empirically via the -–calculateThreshold mode. The data processing inequality (DPI) step was kept active to remove indirect interactions and retain only high confidence direct regulator-target relationships. The final network was generated using the --consolidate mode, which combines bootstrap networks and assigns empirical p-values to each iteration based on their recurrence across bootstraps (*Califano-Lab/ARACNe-AP*, 2017/2025).

To infer TF/CRG/signaling protein activity from the constructed regulons, we applied the msviper() function in the VIPER R package, using the E-MTAB-12862 cohort. This function estimated the enrichment of each regulon (set of target genes) within a gene expression signature, using aREA (analytic rank-based enrichment analysis) (Alvarez et al., 2016).

Expression signatures were generated using the limma package with voom transformation (Law et al., 2014). Although we had previously used DESeq2 for gene-level analysis, limma-voom was selected for VIPER due to its computational efficiency and compatibility with rank-based enrichment methods (Law et al., 2014). For each *BRAF* subgroup comparison, we computed moderated t-statistics using limma:eBayes() and used these ranked statistics as input for enrichment analysis in VIPER.

To assess statistical significance, we generated a custom null model by permuting phenotype labels (*BRAF* subtypes) 1,000 times, as recommended in the VIPER vignette. This null model was used to calculate empirical p-values for enrichment scores, which were then adjusted for multiple testing. Regulators with p-value < 0.05 were considered significantly activated or repressed in the phenotype of interest.

Shadow Analysis for Specificity Filtering

To distinguish primary regulators from those enriched due to shared or overlapping target genes, a shadow analysis using VIPER's built-in tools was performed. This method evaluated the dependency of a regulator's enrichment on more dominantly enriched regulators within the same network. Regulators identified as shadows were prioritized to improve the specificity of biological interpretations (Alvarez et al., 2016).

3 Results and Discussion

3.1 Tumor Location Prediction Using PAMR Classifier and exploratory analysis.

Accurate annotation for tumor location (left vs. right colon) is significant in CRC due to its strong association with distinct molecular subtypes and clinical behavior (Baran et al., 2018). However, the TCGA-COAD dataset lacks complete location metadata. To address this, we developed a predictive model using a shrunken centroid approach trained on the fully annotated E-MTAB-12862 cohort.

Exploratory analysis on the E-MTAB-12862 cohort using PCA did not reveal a clear separation of tumors by anatomical site along PC1 and PC2 (Supplementary Figure S1). However, separation by site became more apparent along PC3 and PC4 (Supplementary Figure S2).

As shown in Supplementary Figure S3 (top panel), the overall misclassification error gradually decreased with increasing threshold values and reached a minimum at a threshold of 6.14, where the model retained 51 genes.

The bottom panel of Supplementary Figure S3 displays class-specific error curves, revealing how performance varied between left- and right-sided tumors across thresholds. Notably, misclassification errors were consistently lower for right-sided tumors (green line), indicating slightly higher sensitivity in this group. In contrast, error increased more sharply for left-sided tumors (red line) at higher thresholds, suggesting reduced robustness when fewer genes were used.

The final model was validated on the subset of TCGA-COAD samples with known tumor locations. It achieved 85.6% accuracy, with balanced accuracy of 85.9%. In addition, it reached a sensitivity of 87.5% and a specificity of 84.2% (Supplementary Table S2 and S3).

3.2 Transcriptional Profiling of BRAF Subtypes

To study the transcriptional landscape of *BRAFmt* colon cancers, we performed DGE analysis across both pairwise (*BRAF-V600E* vs. *BRAF WT*, *BRAF non-V600E* vs. *BRAF WT*, and *BRAF-V600E* vs. *non-V600E*) and one-vs-rest (each subtype vs. the combined other two) comparisons on the discovery and validation cohorts.

3.2.1 BRAF-V600E vs Wild-Type Tumors Display an Immune-Enriched and WNT-Suppressed Transcriptional Program

Several genes emerged significantly differentially expressed in both cohorts (TCGA *BRAF-V600E* n=45, *BRAFWT* n=349; E-MTAB-12862 *BRAF-V600E* n=196, B*RAFWT* n=558) (Figure 1, Supplementary figures S4 and S10). *ANXA10*, a gene implicated in gastric and serrated epithelial differentiation, was among the most strongly upregulated (log2FC > 5, p-adj < 8.20e-13) reinforcing its known association with the CIMP-H and MSI status frequently observed in *BRAF-V600E* tumors (Bae et al., 2015). On the other hand, classical WNT pathway components and tumor suppressors such as *AXIN2*, *PTPRD*/*PTPRO*, *SOX1* and *PRDM13* were consistently downregulated, suggesting a departure from the canonical *APC-KRAS-TP53* adenoma-carcinoma sequence (Jass, 2007) described by Vogelstein (Vogelstein et al., 1988). Additional downregulated genes of interest included *RNF43*, a negative regulator of WNT signaling recurrently mutated in serrated CRCs (significantly decreased in both cohorts) (Elez et al., 2022), and *MLH1*, a mismatch repair gene (observed significantly decreased only in TCGA) (Vilkin et al., 2009). These patterns were further reinforced by the one-vs-rest analysis (*BRAF-V600E* vs. *BRAF WT* + *non-V600E*) in the discovery cohort, which independently recapitulated many of the same top-ranked genes (Supplementary figure S12).

GSEA revealed biologically coherent transcriptional programs in *BRAF-V600E* tumors across both pairwise comparisons and the one-vs-rest analysis. Multiple immune-related gene sets in Hallmarks were significantly enriched in *BRAF-V600E* samples (discovery and validation cohorts), including interferon gamma & alpha response, inflammatory response, complement and allograft rejection (Figure 2, Supplementary figures S5, S9).

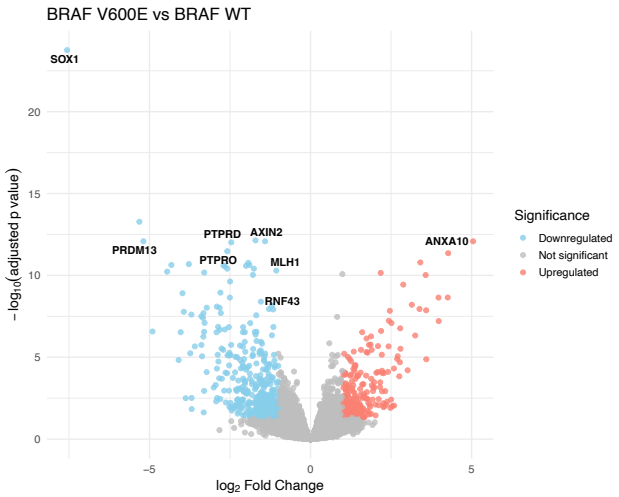

Figure 1: Volcano plot showing differentially expressed genes between *BRAF-V600E* and *BRAF WT* in the TCGA cohort. Each point represents a gene; genes with |log2 fold change| > 1 and adjusted p-value < 0.05. Selected genes of biological relevance to the *BRAF-V600E* phenotype are annotated.

These signatures align with the CMS1, which is defined by MSI-H status, strong immune infiltration, and frequent right-sided tumor origin (Guinney et al., 2015). The upregulation of interferon pathways reflects the hypermutated, antigen-rich environment of *BRAF-V600E* tumors (Baran et al., 2018). Interestingly, we also observed upregulation of Epithelial-Mesenchymal Transition (EMT) in E-MTAB-12862, (p-adj < 0.05) (Supplementary figure S9) and of genes downregulated by *KRAS* in TCGA (p-val < 0.05) (Supplementary figure 7). EMT activation has been linked to tumor invasiveness, immune evasion, and resistance to *EGFR* inhibitors, particularly in metastatic *BRAF-V600E* cases (Kopetz et al., 2024b). The enrichment of *KRAS* signaling down in *BRAF-V600E* tumors reflects the known mutual exclusivity between *KRAS* and *BRAF* mutations, suggesting that these tumors activate the MAPK pathway downstream of *KRAS*, via constitutive *BRAF* activation (Yao et al., 2017).

In contrast, multiple proliferation- and differentiation-associated pathways in Hallmarks were significantly downregulated in *BRAF-V600E* tumors across both cohorts, including MYC targets V1 & V2, WNT/β-catenin signaling, oxidative phosphorylation and peroxisome (Supplementary figures S7 and S9).

These repressed gene sets underscore the non-canonical oncogenic trajectory of *BRAFV600E* tumors (Jass, 2007). Rather than progressing through the classic *APC-KRAS-TP53* axis, these tumors likely arise via the serrated neoplasia pathway, bypassing WNT-driven proliferation (Jass, 2007).

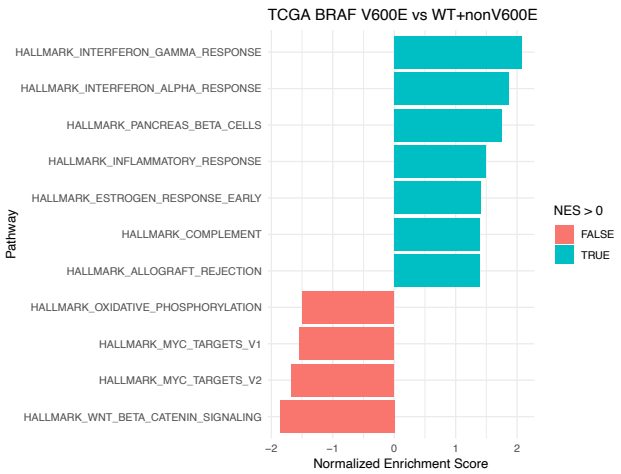

Figure 2: GSEA of Hallmark pathways comparing *BRAF-V600E* tumors to the combined WT and *non-V600E BRAF* tumors in the TCGA cohort. Positively enriched pathways are in blue and negatively enriched pathways in red.

These trends were further supported by KEGG and Reactome pathway enrichment analyses (Supplementary figures S6, SS8). Upregulated pathways reinforced immune activation (interferon signaling, intestinal IgA production) and epithelial remodeling (tight junctions, glycerophospholipid biosynthesis), while downregulated signatures were dominated by mitochondrial respiration, translation, and RNA metabolism, suggesting relatively lower engagement of oxidative and biosynthetic programs in *BRAF-V600E* tumors compared to *BRAF WT*.

Together, these findings confirm that *BRAF-V600E* tumors constitute a molecularly distinct and immune-enriched colon cancer subtype, consistent with CMS1. The transcriptional programs observed reinforce their serrated origin, MSI tendency, and potential therapeutic vulnerabilities, particularly to MAPK-targeted strategies (Kopetz et al., 2024b).

3.2.2 Non-V600E BRAF Tumors Exhibit Modest and Heterogeneous Transcriptional Programs

Compared to the gene expression changes seen in *BRAF-V600E* tumors, *non-V600E BRAFmt* colon cancers showed more subtle and variable differences when compared to *BRAF WT* cases. Differential gene expression analysis across both the TCGA and E-MTAB-12862 cohorts identified relatively few significantly altered genes, and there were no strongly overlapping DEGs between the two datasets using standard thresholds (padj < 0.05, |log2FC| > 1) (Supplementary figures S14, S18). This limited overlap likely reflects the small number of *non-V600E* samples available for analysis (only 11 in TCGA and 28 in E-MTAB-12862) and the biological diversity within this group. Non-V600E *BRAF* mutations include both class II (*RAS*-independent) and class III (*RAS*-dependent) variants, which differ in how they signal and how they behave clinically (Yaeger et al., 2019).

GSEA results from the E-MTAB-12862 cohort showed that these tumors have increased expression of genes linked to pancreas beta cells and the interferon alpha response, along with decreased expression of genes involved in EMT (padj < 0.05). The beta cell signature likely reflects changes in tumor metabolism and it partially overlaps with patterns seen in *BRAF-V600E* tumors. The drop in EMT signaling might point to a less invasive tumor type, which aligns with clinical data showing that *non-V600E BRAF* colon cancers generally have better outcomes and are less likely to spread (Jones et al., 2017) (Supplementary figure S19).

In contrast, the TCGA dataset did not show any significantly enriched pathways at the adjusted p-value threshold (padj < 0.05). However, several trends emerged (p-val < 0.05), including increased activity of WNT/β-catenin signaling, the G2/M checkpoint, E2F targets, and the mitotic spindle (Supplementary figure S15). There was also decreased expression of genes typically downregulated in *KRAS*-driven tumors, as well as those involved in coagulation. Together, these results point toward a more proliferative tumor phenotype mediated by the activation of cell cycle and WNT signaling pathways, which are hallmarks of the traditional adenoma carcinoma sequence (Vogelstein et al., 1988). At least in some cases, non-V600E tumors may engage IN tumorigenic programs more similar to the canonical CRC development, unlike the distinct biology seen in *BRAF-V600E* tumors.

A one-vs-rest analysis (*non-V600E* vs. *BRAF-V600E* + *BRAF WT*) in the TCGA cohort revealed a small set of differentially expressed genes and one significantly enriched pathway (mitotic spindle) suggesting increased cell division in non-V600E tumors (Supplementary figures S20, S21). Notably, genes like *MAGEA1* and *SERPINB4*, which play roles in immune regulation, were downregulated, suggesting

at a less immune-active tumor environment (Koning et al., 2011; Zajac et al., 2017).

Additionally, KEGG and Reactome pathway enrichment in the discovery cohort revealed significant upregulation of melanoma-related and actin cytoskeleton regulation pathways (Supplementary figure S16), as well as extensive activation of fibroblast growth factor receptor (FGFR) signaling cascades, particularly involving PI3K/AKT and PLCγ branches (Supplementary figure S17). These findings may reflect alternative growth and survival pathways activated in *non-V600E BRAF* tumors, distinct from the programs observed in *BRAF-V600E* CRCs (Lu et al., 2023).

These results suggest that *non-V600E BRAF* tumors are highly diverse and lack a clear, consistent transcriptional signature. Some show signs of mild immune or metabolic activity, while others resemble more traditional, proliferation-driven colon cancer. The small number of differentially expressed genes and varying pathway enrichments across cohorts likely reflect both this biological heterogeneity and the limited sample sizes. These findings underscore the need to consider BRAF mutation class (II vs. III) when studying or treating these tumors.

3.2.3 Transcriptional Divergence Between BRAF-V600E and Non-V600E Tumors Reflects Distinct Oncogenic Programs

The comparison of *BRAF-V600E* and *non-V600E* tumors revealed only modest differences in gene expression, likely due to the small number of *non-V600E* samples (Supplementary figures S22, S25). However, one gene (*PTPRO*) was consistently and significantly downregulated in *BRAF-V600E* tumors when doing the gene-overlap between both datasets (p-adj < 0.05). *PTPRO* plays a role in tumor suppression and cell differentiation, so its loss may signal reduced regulatory control in these tumors (Dai et al., 2022). *PTPRD* a related gene, was also significantly downregulated in the TCGA cohort (p-adj < 0.05) and E-MTAB-12862 (p-val < 0.05), implying a possible disruption of this phosphatase pathway in *V600E*-driven CRC (FUNATO et al., 2011).

GSEA revealed contrasts between the subtypes. In the TCGA cohort (Supplementary figure S23), *BRAF-V600E* tumors showed upregulation of oxidative phosphorylation and genes downregulated by *KRAS* signaling, suggesting a shift toward mitochondrial energy metabolism and possible *RAS* feedback suppression. In contrast, *non-V600E* tumors demonstrated upregulation of apical junction, WNT/β-catenin signaling, myogenesis, UV response, EMT and genes upregulated by *KRAS* signaling which aligns with the biological behavior of class III *BRAF* mutations that are RAS-dependent and rely on upstream RAS or RTK signaling for MAPK pathway activation (Yaeger et al., 2019).

In contrast, the GSEA on the E-MTAB-12862 cohort presented a divergent pattern (Supplementary figure S26). Here, *BRAF-V600E* tumors showed enrichment for EMT, angiogenesis, coagulation, inflammatory response and genes upregulated by *KRAS* signaling.

Meanwhile, *non-V600E* tumors in E-MTAB-12862 showed upregulation of bile acid metabolism, WNT/β-catenin signaling, oxidative phosphorylation, and peroxisome activity, suggesting a WNT-driven tumor biology.

As highlighted, several pathways showed opposing trends between cohorts. For example, oxidative phosphorylation was enriched in *BRAF-V600E* tumors in TCGA, while in E-MTAB-12862 it was enriched in *non-V600E* tumors. A similar pattern was observed for EMT, which was downregulated in *BRAF-V600E* tumors in TCGA but strongly upregulated in the same group in E-MTAB-12862. Additionally, *KRAS* signaling up was enriched in *non-V600E* tumors in TCGA (consistent with their *RAS*-dependence) but inversely enriched in *BRAF-V600E* tumors in E-MTAB-12862.

These contradictions likely reflect cohort-specific variability combined with strong imbalances in sample size. In TCGA, the comparison involved 45 *BRAF-V600E* vs. only 11 *non-V600E* tumors, while in E-MTAB-12862, the contrast was between 196 *BRAF-V600E* and just 28 *non-V600E* cases. These differences limit statistical power and may lead to unstable or non-reproducible enrichment results. Furthermore, as previously mentioned, *non-V600E BRAF* mutations have multiple functional classes (class II and III), adding biological heterogeneity that cannot be resolved with limited sample sizes.

KEGG enrichment analyses further underscored cohort-specific transcriptional differences seen in Hallmark pathways, with several pathways showing opposite enrichment directions across datasets (Supplementary figures S24, S27). In TCGA, *BRAF-V600E* tumors showed enrichment of ribosome, oxidative phosphorylation, and chromatin modification pathways, while Extracellular Matrix (ECM) interaction and focal adhesion were suppressed. In contrast, E-MTAB-12862 *BRAF-V600E* tumors showed the

inverse pattern with ECM receptor interaction, cytokine signaling, and coagulation cascades enriched and with fatty acid metabolism, ribosome, and peroxisome activity downregulated.

In summary, *BRAF-V600E* and *non-V600E* tumors appear to follow different biological paths with *V600E* tumors showing immune and epithelial changes, while non-V600E tumors have increased WNT signaling and metabolic stability. However, many differences were not consistent across datasets, likely due to small and uneven sample sizes. These findings highlight the diversity within BRAFmt colon cancers and the need for larger, more balanced studies to draw clearer conclusions.

3.3 Regulatory Program Inference

To identify upstream regulatory factors driving the transcriptomic differences observed across *BRAFmt* subtypes, we applied the VIPER algorithm using ARACNe-inferred regulatory networks in the E-MTAB-12862 cohort. This approach enables the estimation of TF, CRG, and signaling proteins activity based on coordinated expression of their downstream targets (regulons), rather than mRNA abundance alone (Alvarez et al., 2016). We performed msVIPER on the one-vs-rest comparison for each subtype and filtered the results through shadow analysis to prioritize the non-shadowed regulatory elements.

3.3.1 Regulatory Programs Underlying the BRAFV600E Subtype

BRAF-V600E tumors showed activation of a distinct set of transcriptional regulators, including several homeobox and developmental transcription factors (p-val < 0.05) such as *HOXC6*, *HOXC9*, *HOXC11*, *EMX1*, *ONECUT2*, *FOXD1*, and *PRDM16* (Supplementary figure S28). Their activity suggests that these tumors may adopt a more flexible, less differentiated state (Yadav et al., 2024). This pattern fits with what is known about the serrated pathway of CRC, which does not follow the typical *APC-KRAS-TP53* sequence and is consistent with the CMS1 subtype which is often marked by MSI-H, strong immune signals, and a preference for right-sided tumor location (Jass, 2007).

In addition, *BRAF-V600E* tumors showed repression of regulators of epithelial differentiation and WNT pathway control, including *AXIN2*, *RNF43*, *PTPRO*, *PTPRD*, *SOX1*, *PRDM13*, and *NEUROG2*. These factors are necessary for maintaining epithelial identity and regulating WNT signaling which supports a WNT-suppressed, MSI-high phenotype, further distinguishing *BRAF-V600E* tumors from following conventional adenoma-carcinoma progression (Elez et al., 2022; Tsao et al., 2012) (FUNATO et al., 2011).

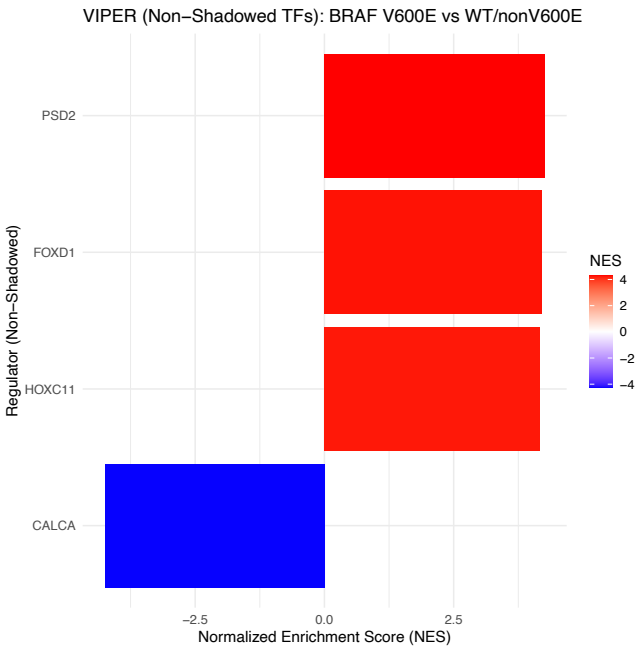

Figure 3: Bar plot showing the top regulators that remained significantly differentially active after shadow analysis between *BRAF-V600E* tumors and the combined group of *BRAF non-V600E* and *BRAF WT*. Three transcriptional regulators were significantly upregulated, while *CALCA* was significantly downregulated in *BRAF-V600E* tumors.

Following the shadow analysis, *FOXD1*, *HOXC11*, and *PSD2* were the only transcriptional regulators that remained significantly active in *BRAF-V600E* tumors (Figure 3). These regulators are not typically expressed in colon epithelium but have been linked to oncogenic processes in other contexts. For example, *FOXD1* has been shown to promote stem-like properties and resistance to chemotherapy in CRC by enhancing β-catenin signaling (Feng et al., 2023). *HOXC11* has been shown to drive progression and poor prognosis in colon adenocarcinoma (L. Liu et al., 2021). *PSD2* remains less studied in CRC, and CALCA, the only regulator that was repressed in *BRAF-V600E* tumors, has not yet been specifically studied in the context of CRC.

3.3.2 Regulatory Programs Underlying the BRAF Non-V600E Subtype

Compared to *BRAF-V600E* tumors, *non-V600E BRAF* tumors show a different pattern of regulator activity. Several regulators that were repressed in

BRAF-V600E tumors, like *AXIN2*, *RNF43*, *PTPRO*, *PTPRD*, *SOX1*, and *PRDM13*, were more highly expressed in the *non-V600E* group (Supplementary figure S29). These regulators, as mentioned, are involved in preserving epithelial traits and many (*AXIN2*, *RNF43*) act as negative regulators of the WNT pathway, restricting excessive cell growth (Elez et al., 2022).

In contrast, several regulators active in *BRAF-V600E* tumors, such as *HOXC11*, *FOXD1*, *PRDM16*, and *EMX1*, were significantly repressed in the non-V600E group. These include developmental and mesenchymal regulators linked to dedifferentiation and plasticity (Yadav et al., 2024), reinforcing the notion that non-V600E tumors follow a different oncogenic trajectory.

After shadow analysis, six regulators remained significantly upregulated in *non-V600E* tumors: *NFIA*, *NFIB*, *HOXB9*, *RGMB*, *MAP2K6*, and *NES* (Supplementary figure S30). *NFIA* and *NFIB* are transcription factors defined to promote CRC cell proliferation and EMT (Z. Liu et al., 2019), while *HOXB9* was found to be significantly associated with worse outcomes and promotion of metastasis, indicating its role in driving developmental and EMT programs (Martinou et al., 2021). *MAP2K6*, also known as *MKK6*, is an upstream activator of the *p38 MAPK* stress pathway, which is triggered by inflammation and environmental stress of the *p38 MAPK* cascade (Loesch & Chen, 2008).

In contrast, three regulators were significantly repressed: *ZNF530*, *ZNF543*, and *EFNB3*. *EFNB3* has been found to contribute to colitis-associated CRC tumor growth and poorer patients' survival (Qiao et al., 2024)

These findings suggest that *non-V600E* tumors engage with other tumor-promoting programs that could also influence disease behavior and treatment response.

3.3.3 Regulatory Programs Underlying the BRAF WT Subtype

When comparing *BRAF WT* tumors to *BRAFmt* (*V600E* and *nonV600E*), we found that the overall pattern of regulator activity was similar to what was observed in the *non-V600E* vs. *WT + V600E* comparison (Supplementary figure S31). However, the shadow analysis revealed a different set of regulators (Supplementary figure S32).

Only *RAB32* and *PTPRD* remained significantly more active in *BRAF WT* tumors. *RAB32* was previously found to be frequently hypermethylated in MSI-high CRC tumors (Shibata et al., 2006), while *PTPRD* as mentioned, suppresses cancer cell migration by maintaining cell-cell adhesion through β-catenin/TCF signaling and is frequently lost or silenced in CRC tumors (FUNATO et al., 2011). Based on these results, we can conclude that *BRAF WT* tumors maintain stronger epithelial structure compared to *BRAFmt* tumors, which often show loss of this characteristic.

In contrast, *BRAF WT* tumors showed repression of *PSD2* and *NTSR2*, two regulators consistently activated in *BRAF-V600E* tumors. While *PSD2* remains understudied in CRC, *NTSR2* belongs to the neurotensin receptor family, which has been linked to tumor progression and *MAPK* pathway activation (Qiu et al., 2017). However, *NTSR2* itself has not been well studied in CRC, and its specific role in this context remains unclear (Qiu et al., 2017)

In summary, these results suggest that *BRAF WT* tumors keep more of their original epithelial identity and regulatory balance, unlike *BRAFmt* tumors which show signs of transcriptional reprogramming and deregulated signaling.

4 Conclusions

In this study, we explored the transcriptional and regulatory differences between *BRAF-V600E*, *non-V600E BRAF*, and *BRAF WT* colon tumors. By analyzing gene expression data from two independent cohorts and applying regulatory network tools, we found that these three groups follow different biological paths.

BRAF-V600E tumors showed strong immune signals, reduced expression of WNT-related genes, and activation of developmental regulators linked to dedifferentiation and plasticity. This implies that *BRAF-V600E* tumors follow the serrated pathway of CRC, which is different from the classic stepwise progression seen in most CRCs. In addition, these tumors also showed signs of transcriptional reprogramming, suggesting they rely on different biological mechanisms that may explain their aggressive behavior and resistance to therapy.

In contrast, *non-V600E* BRAF tumors were more variable. They did not share a single dominant gene expression pattern and showed modest, often

inconsistent changes across datasets. Still, they tended to maintain higher activity of genes involved in cell organization and WNT signaling, which are typical of more stable and structured tumor biology. This could be the reason why *BRAF non-V600E* tumors have better clinical outcomes than *BRAF V600E* cases.

BRAF WT tumors showed expression of regulators involved in keeping epithelial traits and feedback control systems preserved. These findings are consistent with a more stable and less reprogrammed tumor state, aligning with their known better prognosis and different treatment responses.

Overall, this work highlights that clear differences between *BRAF-V600E* and *BRAF non-V600E* colon cancers exist and that understanding these differences is essential for improving how we classify and treat *BRAFmt* colon cancer.

5 Future Directions

This project gave us a better understanding of how *BRAF-V600E* and *non-V600E* colon cancers behave differently at the transcriptional and regulatory level. Nonetheless, we would like to further explore the following aspects.

One interesting next step could be to repeat this analysis in the Hartwig Medical Foundation cohort, which is the largest database of metastatic tumor samples with whole-genome sequencing in the world (n=465 mCRC). Since the cohort is well annotated (including MSI, *BRAF* mutation status and gene expression), it would allow us to validate previous findings in metastatic tissue.

Another direction could be to perform a survival analysis, to check whether certain genes or regulators are linked to better or worse outcomes. This could help us identify markers that might guide treatment decisions and management strategy.

Finally, while our study focused on data-driven analysis, it would be valuable to test some of these findings in the lab. For example, regulators like *FOXD1* or *HOXC11*, which seem especially active in *BRAF-V600E* tumors, could be interesting targets for further research in cell lines or animal models.

6 Code and data availability

You can access all the scripts and data used in this project through the following GitHub repository. Click here.

7 Supplementary material

Please download/view the supplementary material for this project through this GitHub link: Supplementary Material

8 Acknowledgements

Getting to work on this project has been an incredibly rewarding journey. I met amazing people and learned a lot along the way. I want to give special thanks to Lea Lemler for mentoring me during these six wonderful months, to all members of the Cancer Computational Biology group, to Jose A. Seoane for making this possible and to the Vall d'Hebron Institute of Oncology.

References

Alvarez, M. J., Shen, Y., Giorgi, F. M., Lachmann, A., Ding, B. B., Ye, B. H., & Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nature Genetics*, *48*(8), 838–847. https://doi.org/10.1038/ng.3593

Aracne.networks. (n.d.). Bioconductor. Retrieved June 3, 2025, from http://bioconductor.org/packages/aracne.networks/

Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2017). Global patterns and trends in colorectal cancer incidence and mortality. *Gut*, *66*(4), 683–691. https://doi.org/10.1136/gutjnl-2015-310912

Bae, J. M., Kim, J. H., Rhee, Y.-Y., Cho, N.-Y., Kim, T.-Y., & Kang, G. H. (2015). Annexin A10 expression in colorectal cancers with emphasis on the serrated neoplasia pathway. *World Journal of Gastroenterology : WJG*, *21*(33), 9749–9757. https://doi.org/10.3748/wjg.v21.i33.9749

Baran, B., Mert Ozupek, N., Yerli Tetik, N., Acar, E., Bekcioglu, O., & Baskin, Y. (2018). Difference Between Left-Sided and Right-Sided Colorectal Cancer: A Focused Review of Literature. *Gastroenterology Research*, *11*(4), 264–273. https://doi.org/10.14740/gr1062w

Barras, D., Missiaglia, E., Wirapati, P., Sieber, O. M., Jorissen, R. N., Love, C., Molloy, P. L., Jones, I. T., McLaughlin, S., Gibbs, P., Guinney, J., Simon, I. M., Roth, A. D., Bosman, F. T., Tejpar, S., & Delorenzi, M. (2017). BRAF V600E Mutant Colorectal Cancer Subtypes Based on Gene Expression. *Clinical Cancer Research*, *23*(1), 104–115. https://doi.org/10.1158/1078-0432.CCR-16-0140

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

BioStudies. (n.d.). *BioStudies < The European Bioinformatics Institute < EMBL-EBI*. Retrieved June 10, 2025, from https://www.ebi.ac.uk/biostudies/arrayexpress

Bradner, J. E., Hnisz, D., & Young, R. A. (2017). Transcriptional Addiction in Cancer. *Cell*, *168*(4), 629–643. https://doi.org/10.1016/j.cell.2016.12.013

Califano-lab/ARACNe-AP. (2025). [Java]. Califano Lab. https://github.com/califano-lab/ARACNe-AP (Original work published 2017)

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., … Stein, L. (2011). Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Research*, *39*(suppl_1), D691–D697. https://doi.org/10.1093/nar/gkq1018

Dai, W., Xiang, W., Han, L., Yuan, Z., Wang, R., Ma, Y., Yang, Y., Cai, S., Xu, Y., Mo, S., Li, Q., & Cai, G. (2022). PTPRO represses colorectal cancer tumorigenesis and progression by reprogramming fatty acid metabolism. *Cancer Communications (London, England)*, *42*(9), 848–867. https://doi.org/10.1002/cac2.12341

DESeq2. (n.d.). Bioconductor. Retrieved June 3, 2025, from http://bioconductor.org/packages/DESeq2/

Elez, E., Ros, J., Fernández, J., Villacampa, G., Moreno-Cárdenas, A. B., Arenillas, C., Bernatowicz, K., Comas, R., Li, S., Kodack, D. P., Fasani, R., Garcia, A., Gonzalo-Ruiz, J., Piris-Gimenez, A., Nuciforo, P., Kerr, G., Intini, R., Montagna, A., Germani, M. M., … Toledo, R. A. (2022). RNF43 mutations predict response to anti-BRAF/EGFR combinatory therapies in BRAFV600E metastatic colorectal cancer. *Nature Medicine*, *28*(10), 2162–2170. https://doi.org/10.1038/s41591-022-01976-z

Feng, W.-Q., Zhang, Y.-C., Gao, H., Li, W.-C., Miao, Y.-M., Xu, Z.-F., Xu, Z.-Q., Zhao, J.-K., Zheng, M.-H., Zong, Y.-P., & Lu, A.-G. (2023). FOXD1 promotes chemotherapy resistance by enhancing cell stemness in colorectal cancer through β-catenin nuclear localization. *Oncology Reports*, *50*(1), 134. https://doi.org/10.3892/or.2023.8571

FUNATO, K., YAMAZUMI, Y., ODA, T., & AKIYAMA, T. (2011). Tyrosine phosphatase PTPRD suppresses colon cancer cell migration in coordination with CD44. *Experimental and Therapeutic Medicine*, *2*(3), 457–463. https://doi.org/10.3892/etm.2011.231

Guerrero, R. M., Labajos, V. A., Ballena, S. L., Macha, C. A., Lezama, M. S., Roman, C. P., Beltran, P. M., & Torrejon, A. F. (2022, December 15). *Targeting BRAF V600E in metastatic colorectal cancer: Where are we today?* https://doi.org/10.3332/ecancer.2022.1489

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., Bot, B. M., Morris, J. S., Simon, I. M., Gerster, S., Fessler, E., De Sousa E Melo, F., Missiaglia, E., Ramay, H., Barras, D., … Tejpar, S. (2015). The consensus molecular subtypes of colorectal cancer. *Nature Medicine*, *21*(11), 1350–1356. https://doi.org/10.1038/nm.3967

Hastie, T., Tibshirani, R., Narasimhan, B., & Chu, G. (2024). *pamr: Pam: Prediction Analysis for Microarrays* (Version 1.57) [Computer software]. https://cran.r-project.org/web/packages/pamr/index.html

Jass, J. R. (2007). Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology*, *50*(1), 113–130. https://doi.org/10.1111/j.1365-2559.2006.02549.x

Jones, J. C., Renfro, L. A., Al-Shamsi, H. O., Schrock, A. B., Rankin, A., Zhang, B. Y., Kasi, P. M., Voss, J. S., Leal, A. D., Sun, J., Ross, J., Ali, S. M., Hubbard, J. M., Kipp, B. R., McWilliams, R. R., Kopetz, S., Wolff, R. A., & Grothey, A. (2017). Non-V600BRAF Mutations Define a Clinically Distinct Molecular Subtype of Metastatic Colorectal Cancer. *Journal of Clinical Oncology*, *35*(23), 2624–2630. https://doi.org/10.1200/JCO.2016.71.4394

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Koning, P. J. A. de, Kummer, J. A., Poot, S. A. H. de, Quadir, R., Broekhuizen, R., McGettrick, A. F., Higgins, W. J., Devreese, B., Worrall, D. M., & Bovenschen, N. (2011). Intracellular Serine Protease Inhibitor SERPINB4 Inhibits Granzyme M-Induced Cell Death. *PLOS ONE*, *6*(8), e22645. https://doi.org/10.1371/journal.pone.0022645

Kopetz, S., Murphy, D. A., Pu, J., Ciardiello, F., Desai, J., Van Cutsem, E., Wasan, H. S., Yoshino, T., Saffari, H., Zhang, X., Hamilton, P., Xie, T., Yaeger, R., & Tabernero, J. (2024a). Molecular profiling of BRAF-V600E-mutant metastatic colorectal cancer in the phase 3 BEACON CRC trial. *Nature Medicine*, *30*(11), 3261–3271. https://doi.org/10.1038/s41591-024-03235-9

Kopetz, S., Murphy, D. A., Pu, J., Ciardiello, F., Desai, J., Van Cutsem, E., Wasan, H. S., Yoshino, T., Saffari, H., Zhang, X., Hamilton, P., Xie, T., Yaeger, R., & Tabernero, J. (2024b). Molecular profiling of BRAF-V600E-mutant metastatic colorectal cancer in the phase 3 BEACON CRC trial. *Nature Medicine*, *30*(11), 3261–3271. https://doi.org/10.1038/s41591-024-03235-9

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N., & Sergushichev, A. (2021). *Fast gene set enrichment analysis* (p. 060012). bioRxiv. https://doi.org/10.1101/060012

Kuhn, M. (n.d.). *The caret Package*. Retrieved June 4, 2025, from https://topepo.github.io/caret/

Lachit, K., & Vinita, P. (2022). Study of Mismatch Repair Protein Expression by Using Immunohistochemistry in Various Carcinomas with Special Reference to Colorectal Adenocarcinomas—At a Tertiary Referral Laboratory in India. *Asian Pacific Journal of Cancer Biology*, *7*(4), Article 4. https://doi.org/10.31557/apjcb.2022.7.4.341-347

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*(2), R29. https://doi.org/10.1186/gb-2014-15-2-r29

Li, L., Feng, Q., & Wang, X. (2020). PreMSIm: An R package for predicting microsatellite instability from the expression profiling of a gene panel in cancer. *Computational and Structural Biotechnology Journal*, *18*, 668–675. https://doi.org/10.1016/j.csbj.2020.03.007

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*, *1*(6), 417–425. https://doi.org/10.1016/j.cels.2015.12.004

Liu, L., Jia, S., Jin, X., Zhu, S., & Zhang, S. (2021). HOXC11 Expression Is Associated with the Progression of Colon Adenocarcinoma and Is a Prognostic Biomarker. *DNA and Cell Biology*, *40*(9), 1158–1166. https://doi.org/10.1089/dna.2021.0368

Liu, Z., Chen, J., Yuan, W., Ruan, H., Shu, Y., Ji, J., Wu, L., Tang, Q., Zhou, Z., Zhang, X., Cheng, Y., He, S., & Shu, X. (2019). Nuclear factor I/B promotes colorectal cancer cell proliferation, epithelial-mesenchymal transition and 5-fluorouracil resistance. *Cancer Science*, *110*(1), 86–98. https://doi.org/10.1111/cas.13833

Loesch, M., & Chen, G. (2008). The p38 MAPK stress pathway as a tumor suppressor or more? *Frontiers in Bioscience : A Journal and Virtual Library*, *13*, 3581–3593. https://doi.org/10.2741/2951

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, Y., Gu, D., Zhao, C., Sun, Y., Li, W., He, L., Wang, X., Kou, Z., Su, J., & Guo, F. (2023). Genomic landscape and expression profile of consensus molecular subtype four of colorectal cancer. *Frontiers in Immunology*, *14*, 1160052. https://doi.org/10.3389/fimmu.2023.1160052

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R. D., & Califano, A. (2006). ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics*, *7*(1), S7. https://doi.org/10.1186/1471-2105-7-S1-S7

Margonis, G. A., Buettner, S., Andreatos, N., Kim, Y., Wagner, D., Sasaki, K., Beer, A., Schwarz, C., Løes, I. M., Smolle, M., Kamphues, C., He, J., Pawlik, T. M., Kaczirek, K., Poultsides, G., Lønning, P. E., Cameron, J. L., Burkhart, R. A., Gerger, A., … Weiss, M. J. (2018). Association of BRAF Mutations With Survival and Recurrence in Surgically Treated Patients With Metastatic Colorectal Liver Cancer. *JAMA Surgery*,

153(7), e180996. https://doi.org/10.1001/jamasurg.2018.0996

Martinou, E., Falgari, G., Bagwan, I., & Angelidi, A. M. (2021). A Systematic Review on HOX Genes as Potential Biomarkers in Colorectal Cancer: An Emerging Role of HOXB9. *International Journal of Molecular Sciences*, *22*(24), Article 24. https://doi.org/10.3390/ijms222413429

Muzny, D. M., Bainbridge, M. N., Chang, K., Dinh, H. H., Drummond, J. A., Fowler, G., Kovar, C. L., Lewis, L. R., Morgan, M. B., Newsham, I. F., Reid, J. G., Santibanez, J., Shinbrot, E., Trevino, L. R., Wu, Y.-Q., Wang, M., Gunaratne, P., Donehower, L. A., Creighton, C. J., … Tissue source sites and disease working group. (2012a). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, *487*(7407), 330–337. https://doi.org/10.1038/nature11252

Muzny, D. M., Bainbridge, M. N., Chang, K., Dinh, H. H., Drummond, J. A., Fowler, G., Kovar, C. L., Lewis, L. R., Morgan, M. B., Newsham, I. F., Reid, J. G., Santibanez, J., Shinbrot, E., Trevino, L. R., Wu, Y.-Q., Wang, M., Gunaratne, P., Donehower, L. A., Creighton, C. J., … Tissue source sites and disease working group. (2012b). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, *487*(7407), 330–337. https://doi.org/10.1038/nature11252

Nunes, L., Li, F., Wu, M., Luo, T., Hammarström, K., Torell, E., Ljuslinder, I., Mezheyeuski, A., Edqvist, P.-H., Löfgren-Burström, A., Zingmark, C., Edin, S., Larsson, C., Mathot, L., Osterman, E., Osterlund, E., Ljungström, V., Neves, I., Yacoub, N., … Sjöblom, T. (2024). Prognostic genome and transcriptome signatures in colorectal cancers. *Nature*, *633*(8028), 137–146. https://doi.org/10.1038/s41586-024-07769-3

Qiu, S., Pellino, G., Fiorentino, F., Rasheed, S., Darzi, A., Tekkis, P., & Kontovounisios, C. (2017). A Review of the Role of Neurotensin and Its Receptors in Colorectal Cancer. *Gastroenterology Research and Practice*, *2017*, 6456257. https://doi.org/10.1155/2017/6456257

Shibata, D., Mori, Y., Cai, K., Zhang, L., Yin, J., Elahi, A., Hamelin, R., Wong, Y. F., Lo, W. K., Chung, T. K. H., Sato, F., Karpeh, M. S., & Meltzer, S. J. (2006). RAB32 hypermethylation and microsatellite instability in gastric and endometrial adenocarcinomas. *International Journal of Cancer*, *119*(4), 801–806. https://doi.org/10.1002/ijc.21912

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

TCGAbiolinks. (n.d.). Bioconductor. Retrieved June 4, 2025, from http://bioconductor.org/packages/TCGAbiolinks/

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J. G., Baylin, S. B., & Issa, J. P. (1999). CpG island methylator phenotype in colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(15), 8681–8686. https://doi.org/10.1073/pnas.96.15.8681

Tsao, C.-M., Yan, M.-D., Shih, Y.-L., Yu, P.-N., Kuo, C.-C., Lin, W.-C., Li, H.-J., & Lin, Y.-W. (2012). SOX1 functions as a tumor suppressor by antagonizing the WNT/β-catenin signaling pathway in hepatocellular carcinoma. *Hepatology (Baltimore, Md.)*, *56*(6), 2277–2287. https://doi.org/10.1002/hep.25933

Vilkin, A., Niv, Y., Nagasaka, T., Morgenstern, S., Levi, Z., Fireman, Z., Fuerst, F., Goel, A., & Boland, C. R. (2009). Microsatellite Instability, MLH1 Promoter Methylation, and BRAF Mutation Analysis in Sporadic Colorectal Cancers of Different Ethnic Groups in Israel. *Cancer*, *115*(4), 760–769. https://doi.org/10.1002/cncr.24019

Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Smits, A. M. M., & Bos, J. L. (1988). Genetic Alterations during Colorectal-Tumor Development. *New England Journal of Medicine*, *319*(9), 525–532. https://doi.org/10.1056/NEJM198809013190901

Yadav, C., Yadav, R., Nanda, S., Ranga, S., Ahuja, P., & Tanwar, M. (2024). Role of HOX genes in cancer progression and their therapeutical aspects. *Gene*, *919*, 148501. https://doi.org/10.1016/j.gene.2024.148501

Yaeger, R., Kotani, D., Mondaca, S., Parikh, A. R., Bando, H., Van Seventer, E. E., Taniguchi, H., Zhao, H., Thant, C. N., de Stanchina, E., Rosen, N., Corcoran, R. B., Yoshino, T., Yao, Z., & Ebi,

H. (2019). Response to Anti-EGFR Therapy in Patients with BRAF non-V600-Mutant Metastatic Colorectal Cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *25*(23), 7089–7097. https://doi.org/10.1158/1078-0432.CCR-19-2004

Yao, Z., Yaeger, R., Rodrik-Outmezguine, V. S., Tao, A., Torres, N. M., Chang, M. T., Drosten, M., Zhao, H., Cecchi, F., Hembrough, T., Michels, J., Baumert, H., Miles, L., Campbell, N. M., de Stanchina, E., Solit, D. B., Barbacid, M., Taylor, B. S., & Rosen, N. (2017). Tumours with class 3 BRAF mutants are sensitive to the inhibition of activated RAS. *Nature*, *548*(7666), 234–238. https://doi.org/10.1038/nature23291

Zajac, P., Schultz-Thater, E., Tornillo, L., Sadowski, C., Trella, E., Mengus, C., Iezzi, G., & Spagnoli, G. C. (2017). MAGE-A Antigens and Cancer Immunotherapy. *Frontiers in Medicine*, *4*. https://doi.org/10.3389/fmed.2017.00018