Article

Network analysis identifies DAPK3 as a potential biomarker for lymphatic invasion and colon adenocarcinoma prognosis

WGCNA with Colon Cancer Gene Expression

Lymphatic Invasion

Module

GO Term Enrichment

Differential Expression

CAC

Normal

Actin cytoskeleton org.
Cell projection assembly
ECM structural constituent

DAPK3

Intramodular Connectivity

Survival Analysis

DAPK3 low

DAPK3 high

Normal   CAC

Gene

Gene

Differential Correlation

Huey-Miin Chen, Justin A. MacDonald

jmacdo@ucalgary.ca

Highlights

WCGNA reveals a gene module linked to lymphatic invasion in colon adenocarcinoma

DAPK3 is a pseudohub gene with differential expression in colon cancer

Gene ontology identified relationships to cytoskeletal organization and apoptosis

DAPK3 was correlated with lymphatic invasion and poor overall survival

Article

Network analysis identifies DAPK3 as a potential biomarker for lymphatic invasion and colon adenocarcinoma prognosis

Huey-Miin Chen[1] and Justin A. MacDonald[1,2,*]

SUMMARY

Colon adenocarcinoma is a prevalent malignancy with significant mortality. Hence, the identification of molecular biomarkers with prognostic significance is important for improved treatment and patient outcomes. Clinical traits and RNA-Seq of 551 patient samples in the UCSC Toil Recompute Compendium of The Cancer Genome Atlas TARGET and Genotype Tissue Expression project datasets (primary_site = colon) were used for weighted gene co-expression network analysis to reveal the association between gene networks and cancer cell invasion. One module, containing 151 genes, was significantly correlated with lymphatic invasion, a histopathological feature of higher risk colon cancer. *DAPK3* (death-associated protein kinase 3) was identified as the pseudohub of the module. Gene ontology identified gene enrichment related to cytoskeletal organization and apoptotic signaling processes, suggesting modular involvement in tumor cell survival, migration, and epithelial-mesenchymal transformation. Although *DAPK3* expression was reduced in patients with colon cancer, high expression of *DAPK3* was significantly correlated with greater lymphatic invasion and poor overall survival.

INTRODUCTION

Adenocarcinomas of the colon and rectum, collectively referred to as colorectal cancer (CRC), are highly aggressive and common contributors to global cancer morbidity and mortality (Malki et al., 2020; Dekker et al., 2019; Marmol et al., 2017). The likelihood of CRC diagnosis is about 5%, and risk factors include age and sex as well as diet and other lifestyle habits (Marmol et al., 2017). The distinction between colon and rectal cancer is largely anatomical, but the tumor location does impact both surgical and therapeutic management strategies as well as prognosis. In 2018, adenocarcinoma of the colon was the fourth most common malignancy and was documented to account for >10% of all diagnosed cancers globally (Ferlay et al., 2019). Colorectal adenocarcinomas are caused by genomic instability and mutations that occur in tumor suppressor genes, oncogenes, and genes related to DNA repair. Genetic and epigenetic alterations contribute to molecular events that lead to the neoplastic transformation of healthy colorectal epithelium which then progresses to malignancy (Fearon and Vogelstein, 1990). Several canonical signaling pathways are known to be affected by genomic aberrations in CRC (i.e., mitogen-activated protein kinases (MAPKs), phosphoinositide 3-kinase (PI3K), transforming growth factor (TGF)-β, tumor protein P53 (TP53), and wingless and Int-1 (Wnt/β-catenin) and ultimately lead to changes in fundamental cellular processes that drive tumor development, including differentiation, apoptosis, cell cycle, cell proliferation, and survival (Malki et al., 2020; Marmol et al., 2017).

In addition, cancer cell migration is particularly important to the process of metastasis, by which cancer cells escape from the primary tumor and travel to distant sites to invade a new niche (Chow et al., 2012). Two important pathologic processes for cancer cell migration include lymphovascular and perineural invasion, whereby tumor cells invade nervous tissue (perinuclear invasion, PNI) or small lymphatic and vascular tissues (lymphovascular invasion, LVI), respectively. Patients with locally advanced rectal cancer who demonstrated PNI and/or LVI had significantly higher risk of recurrence and poorer survival outcomes (Sun et al., 2019). Moreover, LVI and PNI had detrimental effects on survival after diagnosis of stage II or III adenocarcinoma of the colon (Mutabdzic et al., 2019; Skancke et al., 2019). In patients with stage III colon cancer, increased lymph node involvement was also associated with a need for adjuvant therapy with

[1]Department of Biochemistry & Molecular Biology, Cumming School of Medicine, University of Calgary, 3280 Hospital Drive NW, Calgary, AB T2N 4Z6, Canada

[2]Lead contact

*Correspondence: jmacdo@ucalgary.ca
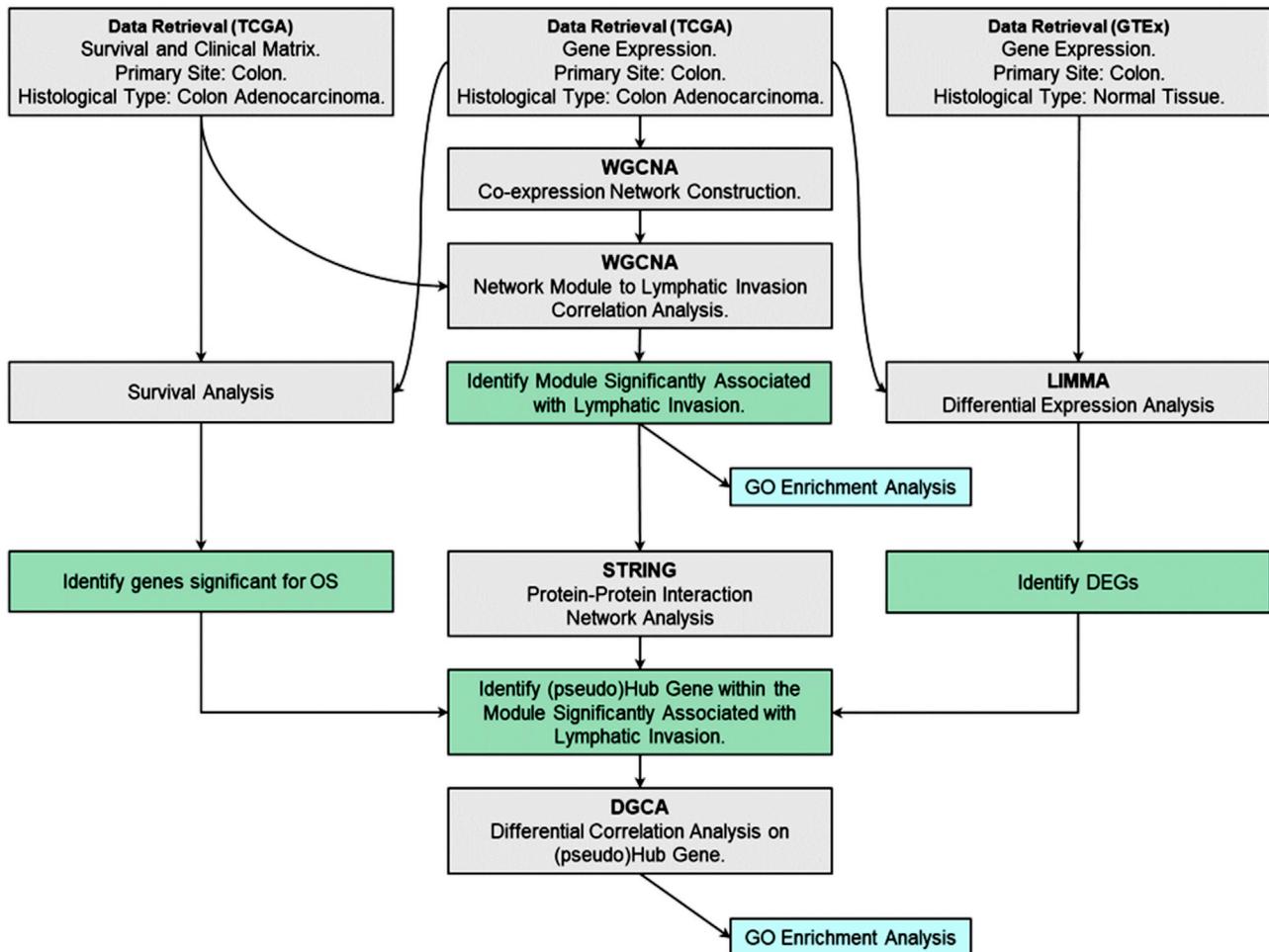
Figure 1. Workflow scheme. The workflow for data acquisition, preparation, processing, and analyses

worsening overall survival and shorter time to recurrence (Sjo et al., 2012). Thus, an improved understanding of the genes associated with PNI and/or LVI of colon adenocarcinoma may reveal candidate biomarkers and novel therapeutic opportunities by identifying predictive and/or prognosis markers for metastatic invasion events.

In this study, we assess the ubiquity between tumor and LVI molecular markers that are associated with colon cancer to improve prediction of cancer cell migration and invasion with links to overall survival. Weighted gene co-expression network analysis (WGCNA) is commonly used to explore next-generation RNA-Seq datasets for correlations among genes and clinical phenotypes (Li et al., 2018; Langfelder and Horvath, 2008; Zhang and Horvath, 2005). WGCNA can provide insights into the molecular networks that connect clinical features of tumor progression and has been previously used to identify candidate prognostic biomarkers or therapeutic targets (Chen et al., 2019; Zou and Jing, 2019; Qin et al., 2019; Zhai et al., 2017; Yang et al., 2014). Co-expression modules were constructed using normalized RNA-Seq data and then related to clinicopathological outcomes for patients with colon adenocarcinoma. Search tool for the retrieval of interacting genes/proteins (STRING) network analysis and gene ontology (GO) enrichment analysis were performed to identify hub genes and the primary biological function of selected modules that were aligned with LVI and PNI characteristics of colon adenocarcinoma.

RESULTS

The primary objective of our study was to resolve tumor-associated molecular markers by their linkage to lymphatic invasion and overall survival. To achieve this, we used a schema that is outlined in Figure 1.
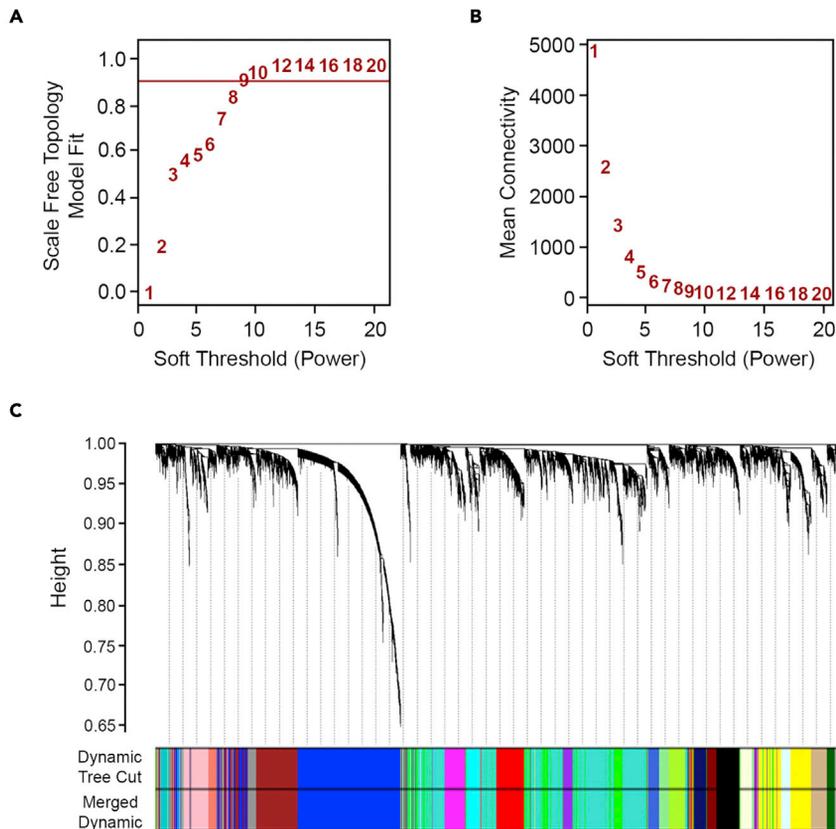
Figure 2. Weighted gene co-expression network analysis (WGCNA) of primary tumors from colon adenocarcinoma

A range of soft-thresholding power (β) was screened within the WGCNA pipeline (A), and the plot shows the scale-free topology fit index for different β powers. The mean connectivity for various soft-thresholding powers is also provided (B). Hierarchical clustering and module assignment was completed based on a topological overlap matrix to calculate the corresponding dissimilarity (dissTOM). In (C), the connected genes were grouped into colored modules with the dynamic tree cut method. Highly correlated modules were merged based on a cut height of 0.25.

WGCNA identified a gene network associated with lymphovascular invasion in colon adenocarcinoma

The presence of tumor cells within the lymphatics or blood vessels is associated with increased risk of axillary lymph node and distant metastases. To identify candidate biomarkers for lymphatic invasion, we performed WGCNA on RNA-Seq data obtained from The Cancer Genome Atlas (TCGA) with the |primary_site| originating from colon, |histological_type| limited to colon adenocarcinoma, and non-protein-coding genes excluded. When the soft thresholding was set at 10, the scale-free topology fit index reached 0.95 (Figures 2A and 2B). Modules of co-expressed genes were determined with hierarchical clustering and the dynamic tree cut procedure, and each of the modules was marked by a color. In all, 24 different modules resulted from WGCNA with the clustering dendrograms of genes shown in Figure 2C. Merging of modules with similar expression profiles was completed with a cut height of 0.25, corresponding to a 0.75 correlation; however, this did not reduce the number of modules. The *darkgrey* module was the smallest, with 100 genes, while the *turquoise* and *blue* modules were the largest, with 2,005 and 1,742 genes, respectively (Table S1). All genes were assigned, and no genes were grouped in *gray* (i.e., the module reserved for genes with no assigned classification).

The clinical features of patients colon adenocarcinoma, including lymphatic invasion, venous invasion, and perineural invasion occurrence, as well as pathological stage and pathological T (tumor, T), pathological N (node, N), and pathological M (metastasis, M) staging categories were extracted from the TCGA dataset and then related to the module eigengenes (MEs). As shown in Figure 3, ME*darkred* was correlated with

Module-Trait Relationships
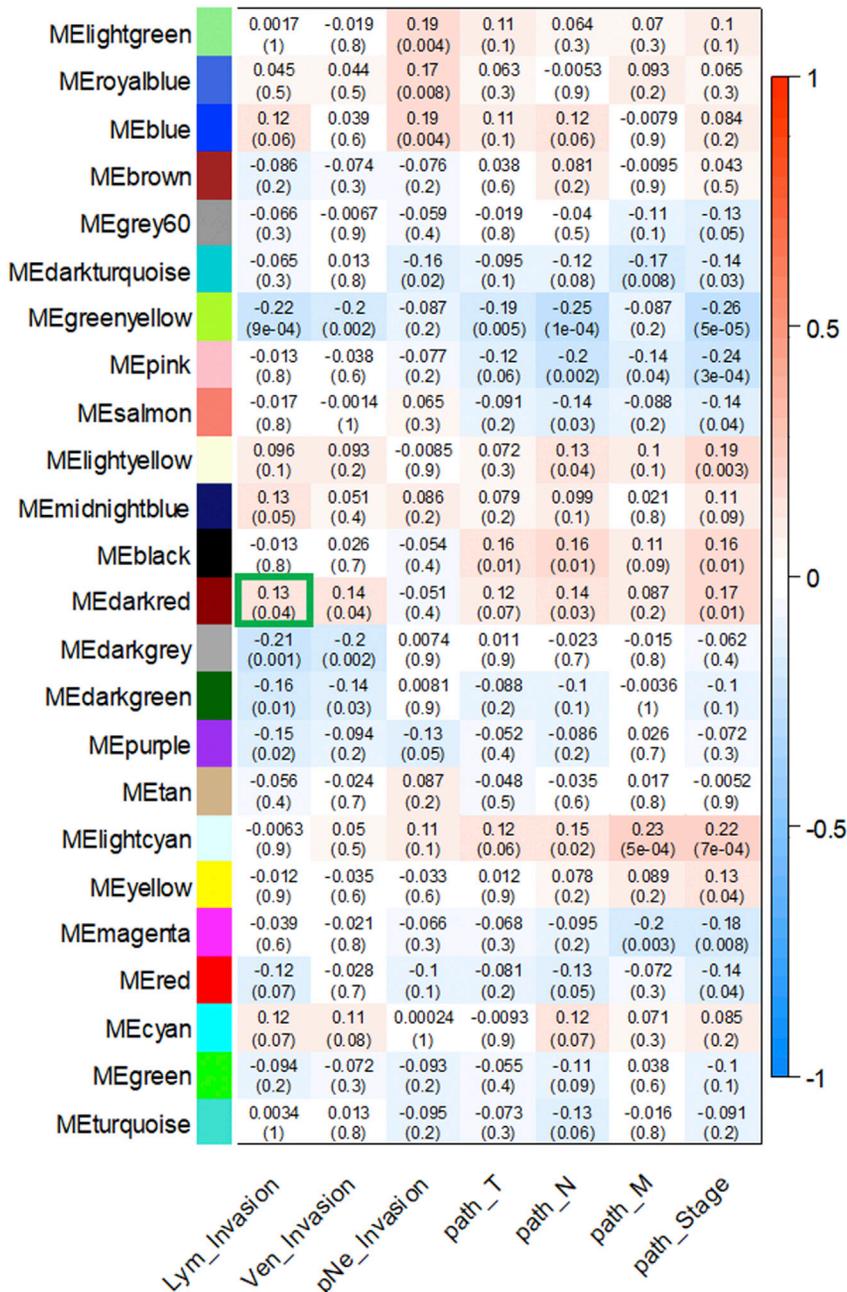
Figure 3. Module-trait relationship matrix correlating MEs with clinical traits

For the module-trait relationships, each row in the matrix corresponds to a module eigengene (ME) and each column to a clinical trait. Each cell contains the corresponding correlation with the p values of the correlations in parentheses. Each module-trait relationship is color-coded by correlation as indicated in the color legend on the right; blue indicates a negative correlation, while red represents a positive one. The ME*darkred* (outlined in green) had the highest correlation with lymphatic invasion and contains *DAPK3*.

lymphatic invasion (r = 0.13, p = 0.04) and venous invasion (r = 0.14, p = 0.04). Additional correlations were observed for ME*darkred* with pathology N (r = 0.14, p = 0.03) and pathological stage (r = 0.17, p = 0.01). Notably, the *darkred* module contained *DAPK3* (death-associated protein kinase 3) along with 150 other genes (Table S2). *DAPK3* expression is attenuated in several squamous cell carcinomas (Kake et al.,

2017; Das et al., 2016; Li et al., 2015; Kocher et al., 2015; Bi et al., 2009). Moreover, loss-of-function mutations or deletion of *DAPK3* promote increased cell survival, proliferation, cellular aggregation, and increased resistance to chemotherapy (Brognard et al., 2011). Considering this evidence, *DAPK3* is now implicated as a tumor-suppressing kinase.

To further validate the module-clinical trait relationships, we completed additional examination of the correlation between the gene significance (GS) and the module membership (MM) measures. Five modules possessed a greater proportion of gene members with positive significance toward lymphatic invasion, including the *darkred* module (r = 0.29; p = 0.0003), as well as *blue*, *cyan*, *pink*, and *lightyellow* modules (Figure 4). As the module of primary interest, *darkred* contained a majority of members (79%) with positive gene significance toward lymphatic invasion. MEs were also used as representative profiles to assess module similarity by eigengene correlation. Construction of the eigengene network provided the identification of groups of correlated eigengenes termed meta-modules. ME*darkred*, ME*black*, ME*midnightblue,* and ME*lightyellow* modules were highly related. The mutual correlation of these four MEs was much stronger than their correlation with lymphatic invasion; however, lymphatic invasion was present as a part of the meta-module (Figure S1).

The gene module associated with lymphovascular invasion is linked to apoptosis and actin cytoskeleton processes

Functional enrichment analysis was conducted with topGO R on the *darkred* module, and the enriched GO biological processes were found to be dominated by regulation of apoptosis and cellular architecture (Figure 5A). Importantly, the significantly enriched pathways were aligned with "Actin cytoskeleton organization; GO:0030036," the "Positive regulation of apoptotic signaling pathway; GO:2001235," the "Regulation of mitochondrion organization; GO:0010821," and the "Regulation of mitochondrial membrane permeability involved in apoptotic process; GO:1902108".

DAPK3 is revealed as the pseudohub gene in the module associated with lymphovascular invasion

Differential expression analysis is routinely used to interpret the biological importance of transcriptomics experiments (Crow et al., 2019). RNA-Seq data from TCGA and Genotype Tissue Expression project (GTEx) were analyzed to obtain a list of differentially expressed genes (DEGs). Based on the threshold of |log2(FC)| > 0.58 and an adjusted p value cut-off of 5%, 9,390 DEGs were identified. For a stricter definition of significance, the treat method was used to calculate the p values from empirical Bayes-moderated t-statistics with a minimum |log2(FC)| requirement (McCarthy and Smyth, 2009). The number of DEGs was reduced to a total of 8,479 when treat testing required a |log2(FC)| greater than 0.58. Among the 8,479 DEGs, 3,579 were significantly upregulated in CAC samples, whereas 4,900 were significantly downregulated in CAC samples (Figure 5B). The list of DEGs generated via the limma R pipeline was amalgamated with WGCNA-derived modules. An examination of the consolidated analyses revealed 57 of 151 (38%) genes in the *darkred* module to be differentially expressed. The R function *coxph* |survival| was used to determine the association between differential gene expression and overall survival. This analysis identified 18 of 151 (12%) genes in the *darkred* module that passed the |log_rank = 0.05| threshold. Of this subset, 6 of 151 genes were significantly differentially expressed and significantly associated with overall survival (Figure 5C and Table S2).

As shown in Figure 6, genes within the *darkred* module were used as input for protein-protein interaction network construction with the STRING database. Intramodular connectivity was used to identify putative intramodular hub genes with the top five candidates having greater than 25 intramodular connections (i.e., RHOT2, mitochondrial Rho GTPase 2; RIPK3, receptor-interacting serine/threonine-protein kinase 3; SSH3, slingshot homolog 3; PIDD1, p53-induced-death-domain-containing protein 1; and TELO2, telomer length regulation protein TEL2 homolog). Importantly, none of the five hub candidates were both differentially expressed and linked to overall survival probability. Of genes that were differentially expressed and were significant for overall survival, *DAPK3* was revealed as the pseudo-hub gene of the network (Figure 6). It also possessed a high degree of connectivity as revealed by degree distribution and closeness centrality (Table S3). In addition to *DAPK3, ANKRD13D* (ankyrin-repeat-domain-containing protein 13D), *ENKD1* (enkurin-domain-containing protein 1), *DEF8* (differentially expressed in FDCP 8 homolog), *NOL3* (nucleolar protein 3), and *ARRDC1* (α-arrestin-domain-containing protein 1) exhibited differential expression profiles that significantly impacted overall survival probability.

A

Module	Gene Count	Gene Count GS<0 (%)	Gene Count GS>0 (%)	Correlation Pearson r	Pearson P (two-tailed)	P value summary
blue	1742	27.10%	72.90%	0.3565	<0.0001	****
lightcyan	171	56.14%	43.86%	0.3232	<0.0001	****
grey	162	70.37%	29.63%	0.3189	<0.0001	****
darkred	**151**	**20.53%**	**79.47%**	**0.288**	**0.0003**	***
cyan	185	14.59%	85.41%	0.2599	0.0004	***
pink	322	48.45%	51.55%	0.2499	<0.0001	****
lightgreen	157	50.96%	49.04%	0.2072	0.0092	**
lightyellow	156	27.56%	72.44%	0.1883	0.0186	*
salmon	225	56.44%	43.56%	0.1874	0.0048	**
tan	234	72.65%	27.35%	0.1645	0.0118	*
turquoise	2005	50.72%	49.28%	0.08345	0.0002	***
darkgreen	114	85.09%	14.91%	-0.3128	0.0007	***

B
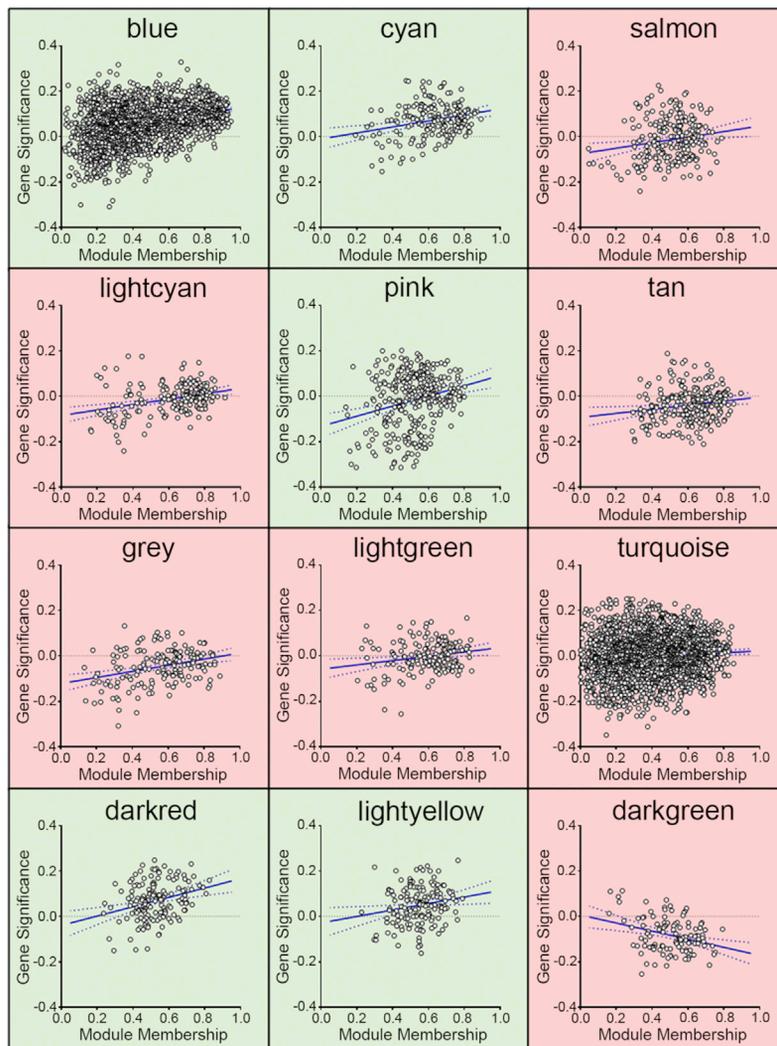
Figure 4. Intramodular analysis for modular membership versus gene significance

(A) The correlation between the gene significance (GS) and module membership (MM) measures was explored. Shown are the Pearson correlations for modules with p values (two-tailed) indicating significance.

(B) Scatterplots of GS versus MM for lymphatic invasion. GS is the relationship between the gene and the clinical trait, and the MM describes the correlation between the ME and the expression profile of a given gene with said module. Highlighted in green are modules with a majority of members having positive GS toward lymphatic invasion. Red highlights modules with most members having negative GS toward lymphatic invasion.

A
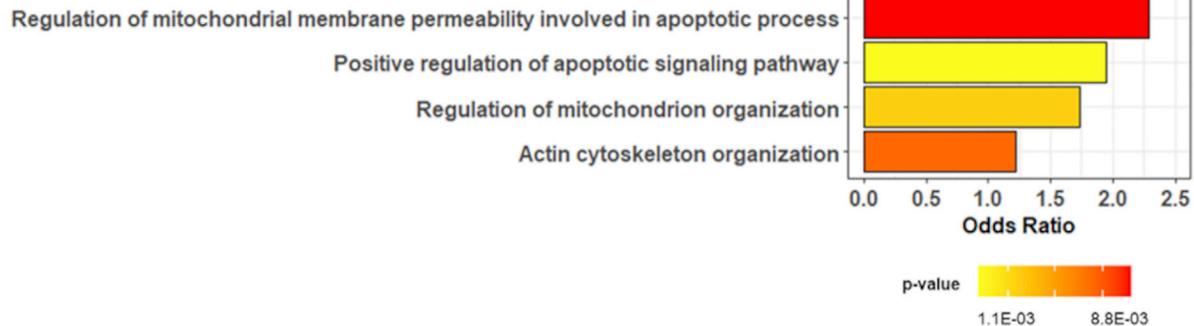
B
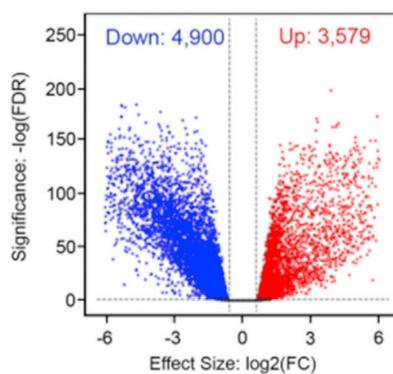
C
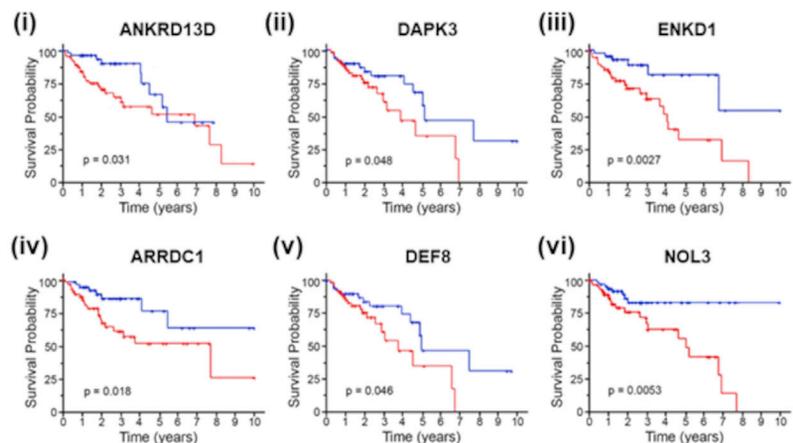
Figure 5. Biomolecular characteristics of the *darkred* module

(A) Gene ontology (GO) enrichment analysis was conducted for genes in the *darkred* module. Biological process (BP) GO terms are listed along the y axis. GO terms were deemed significant when |p value| $\leq$ 0.01 and |total number of annotated genes| $\geq$ 30.

(B) A volcano plot of differentially expressed genes (DEGs, n = 8,479) derived from the TCGA and GTEx datasets and plotted as the log2-fold change differences (log2FC) between the compared samples. The upregulated (red) and downregulated (blue) genes were assessed for statistical significance using the *treat* method with empirical Bayes-moderated t-statistics (-log(FDR)).

(C) Six genes were revealed to be differentially expressed and be significantly associated with overall survival. Kaplan-Meier survival analysis by log-rank test was used to examine survivability with respect to high or low expression of (i) *ANKRD13D*, ankyrin repeat domain-containing protein 13D; (ii) *DAPK3*, death-associated protein kinase 3; (iii) *ENKD1*, enkurin domain-containing protein 1; (iv) *ARRDC1*, arrestin domain-containing protein 1; (v) *DEF8*, differentially expressed in FDCP 8 homolog; and (vi) *NOL3*, nucleolar protein 3.

To investigate *DAPK3* expression in human colon cancer, the Toil recomputed expression of *DAPK3* in GTEx and TCGA (healthy colon vs. colon adenocarcinoma, primary tumor) was juxtaposed via unpaired t test with Welch's correction. As shown, the level of *DAPK3* expression is significantly downregulated (Figure 7A; −0.825 ± 0.089, p < 0.0001) in colon adenocarcinoma. A subsequent analysis of *DAPK3* expression (healthy colon vs. colon adenocarcinoma, primary tumor; unified by Wang et al., 2018) generated consistent results and verified reduced *DAPK3* expression in colon cancer samples (Figure 7B; −0.879 ± 0.048, p < 0.0001). Thus, the gene expression analysis further substantiates *DAPK3* as a candidate tumor suppressor gene. Survival data were also retrieved as part of the UCSC dataset downloaded from Xena and subjected to Kaplan-Meier analysis to determine the significance of *DAPK3* expression on survival probability in patients with colon adenocarcinoma. Figure 8A shows that high *DAPK3* expression (>10.92, upper quartile) strongly correlates with poor overall survival probability when compared with low DAPK3 expression (<10.39, lower quartile). However, *DAPK3* expression does not make a significant impact on disease-specific survival (Figure 8B). One feasible explanation for this discrepancy is an experimentally defined role for *DAPK3* in the actin filament and focal adhesion dynamics (Nehru et al., 2013), as well as cellular migration (Li et al., 2015; Komatsu and Ikebe, 2004), which may play unfavorably within the context of colon cancer metastasis. Invasion data were not available for the Toil Recomputed Compendium; however, an examination of *DAPK3* expression versus invasion for the 'TCGA Colon Cancer' dataset

Figure 6. Biomolecular interaction network constructed for DAPK3 and other gene products of the *darkred* module

Genes within the *darkred* module were used as input for protein-protein interaction network construction using the search tool for the retrieval of interacting genes/proteins (STRING). The network was visualized with Cytoscape. The node size reflects the degree of interactions (i.e., edges) linked to a given gene. The node color describes the differential expression: blue, downregulation; orange, upregulation; and white, not differentially expressed. The node border color is reflective of the association with overall survival: red, |log_rank <0.05| and blank, not significant. Genes that were differentially expressed and significantly associated with overall survival are noted in red text.

via Xena does reveal correlation between *DAPK3* expression and metastatic invasion occurrence. In this case, incidents of venous, lymphatic, and perineural invasions were paradoxically associated with higher *DAPK3* expression (Figure 8C).

Differential correlation of *DAPK3* in normal and tumor samples revealed loss of correlation with genes linked to cell migration and cell adhesion

Genes that are functionally related tend to have similar expression profile; therefore, differential gene correlation analysis (DGCA) comparing expression correlation in normal and disease samples can illuminate *DAPK3*-dependent biological processes and/or molecular pathways that may be distinctly involved between the two states. DGCA was used to compare gene expression correlation in normal and colon adenocarcinoma disease samples, and the outcomes were sorted into nine possible categories (Figure S2). Of the 8,528 protein-coding genes surveyed, 8,222 (93%) were found to be differentially correlated with *DAPK3*

Figure 7. DAPK3 expression in the colon of healthy normal and primary colonic adenocarcinoma tumors

The expression of *DAPK3* in the healthy colon (GTEx) and colon adenocarcinoma primary tumor (TCGA) was examined in UCSC Toil RNA-Seq Recompute Compendium (A) and MSKCC unified (B) datasets. Statistical significance was identified by Welch's t test for two independent samples of unequal variance. Data are presented as a violin plot showing the median (dashed line) along with the upper and lower limits of the interquartile range between the 25th and 75th percentiles (dotted lines). The n values are indicated on the figure. TCGA, The Cancer Genome Atlas.

between healthy controls and colon cancer conditions. When compared against the GTEx normal healthy population, *DAPK3* lost correlation with 46% of the genes surveyed and gained correlation with 47%. Finally, 6% of the surveyed genes had no significant correlation in either condition (healthy colon or colon adenocarcinoma) or had nonsignificant differential correlation. GO enrichment analysis was used to identify the biological processes and/or molecular pathways wherein differential correlation of *DAPK3* was implicated (Figure 9). *DAPK3* differential correlations changed most within the biological process of "Organic acid catabolic process; GO:0016054," "Cell projection assembly; GO:0030031, and "Muscle organ development; GO:0007517" and the molecular pathways of "Endoribonuclease activity; GO:0004521" and "Extracellular matrix structural constituent; GO:0005201."

DISCUSSION

Several patient and disease characteristics are known to affect survival of patients with colon cancer, including age, sex, primary tumor location, tumor grade, and lymph node involvement (Dienstmann et al., 2019; Weiser et al., 2011). In addition, LVI and PNI are histopathological features associated with higher risk colon cancer (Skancke et al., 2019; Mutabdzic et al., 2019; Cienfuegos et al., 2017; Liebig et al., 2009). However, the molecular biomarkers that distinguish between normal and invasive colon cancer are not well described. In this study, we used the WGCNA method to evaluate the system-level functionality of genes with clinical invasion features of colon adenocarcinoma.

The power of WGCNA lies in its ability to reveal gene modules (i.e., gene co-expression networks) and identify the central players (i.e., hub genes) within modules that are usually related to biological functions (Li et al., 2018; Langfelder and Horvath, 2008; Zhang and Horvath, 2005). RNA-Seq data and clinical information obtained from the TCGA database were included in the WGCNA to identify robust co-expression modules associated with tumor invasion characteristics. A total of 24 distinct gene co-expression modules that included 9,750 protein-coding genes were identified from primary tumors of patients with colon cancer. After examining the associations between the modules and clinical invasion traits, *darkred* was identified to be a clinically significant gene module that required further interrogation. Significantly the module was linked with positive gene correlation toward lymphatic and venous invasion as well as pathological stage. The 151 genes within this module were considered to possess functional associations with each other and, thus, could potentially identify important molecular pathways and hub genes that could advance understanding, detection, and/or treatment opportunities for colon cancer.

Functional annotation of the *darkred* module revealed a core set of genes that were involved in the "Actin cytoskeleton pathway." Moreover, the enrichment of GO terms "Cell projection assembly" and
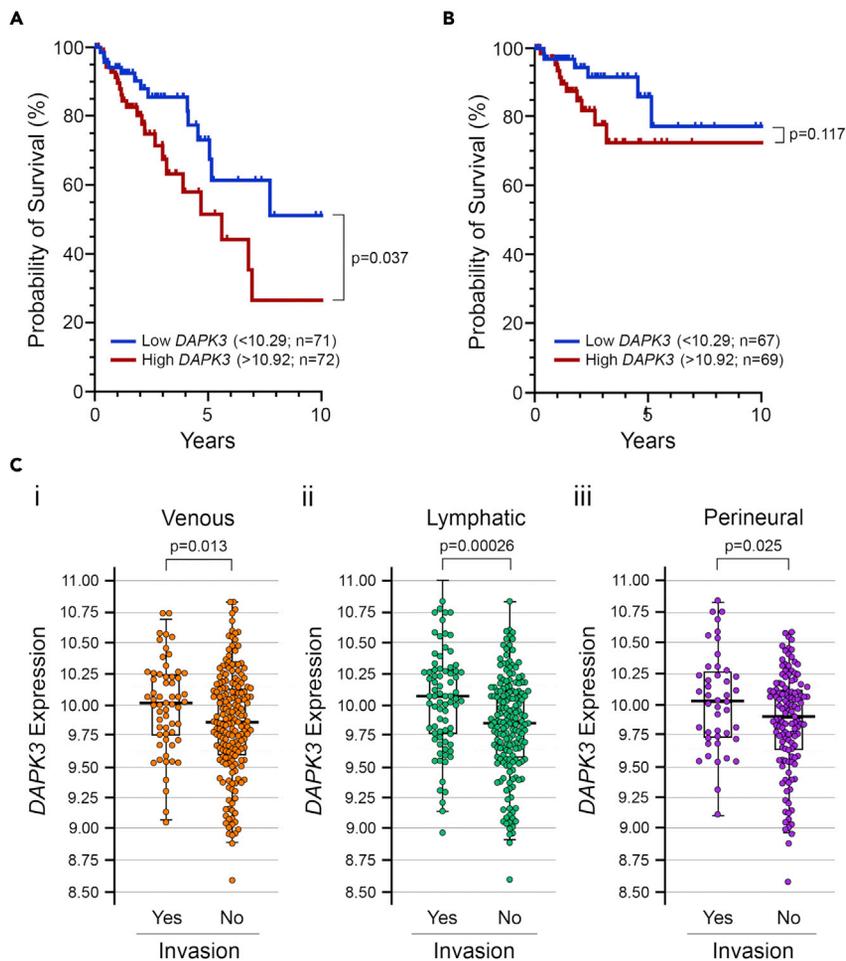
Figure 8. High *DAPK3* expression is associated with poor overall survival of patients with colon adenocarcinoma
Overall survival data (A) and disease-specific survival data (B) were retrieved as part of the UCSC Toil Recompute dataset and subjected to Kaplan-Meier analyses. In (C), a comparison of *DAPK3* expression in patients with colon adenocarcinoma with or without incidents of venous (i), lymphatic (ii), or perineural (iii) invasion was completed using data retrieved from the Xena TCGA data hub (https://tcga.xenahubs.net). Statistical significance was identified by Welch's t test for two independent samples of unequal variance (n = 245 (venous), 251 (lymphatic), and 171 (perineural)). Data are presented as boxplots showing the median (solid line) along with the upper and lower limits of the interquartile range between the 25th and 75th percentiles (box) and Tukey whiskers. TCGA, The Cancer Genome Atlas.

"Extracellular matrix structural constituent" in differential correlations identified for *DAPK3* are of particular interest. These terms are associated with the molecular dynamics of cell migration and adhesion, which are particularly important to the process of metastasis. The cytoskeleton, especially the contractile tension generated by actomyosin, is also inherent to the pathogenesis and metastasis of tumors. Indeed, there is growing appreciation that a wide range of cellular activities that are highly relevant to tumorigenesis are dependent on the composition and organization of the cytoskeletal architecture, including tumor cell migration and invasion (Fletcher and Mullins, 2010; Vasiliev, 2004), epithelial-mesenchymal transformation (Rana et al., 2018; Lamouille et al., 2014), nuclear dysmorphia and genome stability (Takaki et al., 2017; Irianto et al., 2016; Chow et al., 2012), and tumor cell survival under hemodynamic shear stress (Xin et al., 2019). Furthermore, genes in the *darkred* module were also enriched in processes related to "Positive regulation of apoptotic signaling pathway" and "Regulation of mitochondrial membrane permeability involved in apoptotic process." It is well known that the attenuation of apoptotic signaling is a hallmark of tumor biology (Zhang and Yu, 2013; Yang et al., 2009), and a plethora of alterations have been revealed in key apoptotic pathways that increase tumor cell survival and reduce the efficacy of chemotherapy. Enhanced invasiveness is also observed in cancer cells that experience failed apoptosis (Berthenet et al., 2020) or incomplete mitochondrial outer membrane permeabilization (Ichim and Tait, 2016).
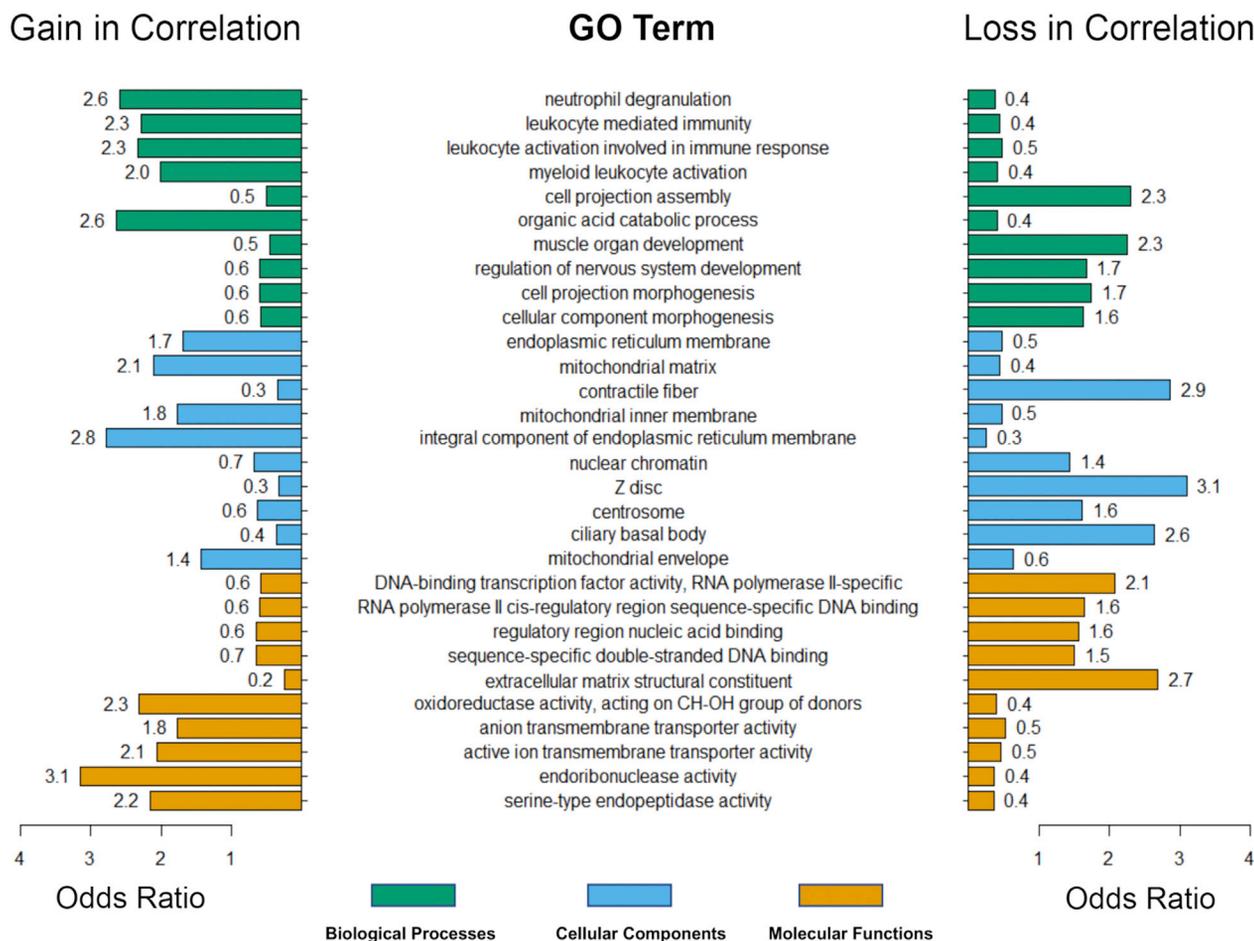
Figure 9. Differential gene correlation analysis of *DAPK3*-dependent biological processes
DGCA R was used to compare gene expression correlation in normal and colon adenocarcinoma disease samples. GO enrichment analysis was completed using |minSize| = 30, |p value cutoff| = 0.01, and |filterSigThresh| = 0.01. GO, gene ontology.

When the genes in the *darkred* module were mapped using STRING, *DAPK3* was identified as a pseudohub gene. However, *RHOT2* (mitochondrial Rho GTPase 2; also known as *MIRO2*) was the true core hub based on node degree when the (i) differential expression or (ii) Kaplan-Meier overall survival estimate was disregarded. RHOT2 is a mitochondrial GTPase involved in mitochondrial trafficking and serves as a docking site for parkin (PRKN)-mediated mitophagy. While the RhoA GTPase family is closely associated with cancer progression, there are few studies on the role of the RHOT2 protein in tumor cell movement. In one published study, however, mitochondrial trafficking to the cortical cytoskeleton and tumor cell invasion were suppressed when siRNA-mediated silencing of Miro2 in LN229 glioma cells was used under conditions of cellular stress (Caino et al., 2016).

The *RHOT2* network also included *ENKD1*, *ARRDC1*, *NOL3*, *DAPK3*, *ANKRD13D*, and *DEF8*. Each of these genes was differentially expressed in colon adenocarcinoma and was correlated with patient prognosis for overall survival analysis. The ANKRD13D protein possesses a ubiquitin-interacting motif (UIM) and forms a complex with other ANKRD13 family members to bind the Lys63-linked ubiquitin chains appended to epidermal growth factor receptor (EGFR) to regulate the internalization of ligand-activated EGFR (Mattioni et al., 2020; Tanno et al., 2012). *ANKRD13* has also been used in a panel of DNA methylation markers to identify colorectal cancer (CRC) in cell-free DNA samples obtained from patients (Cho et al., 2020). The

ARDDC1 protein controls the generation of endosomal microvesicles that bud from the plasma membrane (Nabhan et al., 2012), a type of protein cargo (Anand et al., 2018), and, consequently, signal transduction by receptors subjected to endosomal sorting mechanisms. Other studies suggest that ARRDC1 may act as a tumor suppressor by modulating the levels of Yes-associated protein (YAP) 1 and other membrane receptors (Xiao et al., 2018). NOL3 is an RNA-binding protein that has been implicated in tumorigenesis and metastasis. NOL3 was associated with worse prognosis in CRC and was overexpressed in patients with CRC with lymph node metastasis (Zhang et al., 2020; Toth et al., 2016). Finally, the gene products of *ENKD1* and *DEF8* are poorly described in the literature, so it is unclear how these gene products may act to impact tumor cells.

The emergence of *DAPK3* as a pseudohub gene in the *darkred* network with differential expression and significant linkage to survival outcome was particularly interesting. DAPK3 (also known as zipper-interacting protein kinase, ZIPK) is a serine/threonine protein kinase that acts as a molecular switch for a multitude of cellular processes, including the induction of apoptotic and autophagic cell death, cell proliferation, actomyosin contraction, and cellular migration (Shiloh et al., 2014). The proapoptotic influence of DAPK3 is aligned with its association with promyelocytic leukemia (PML) oncogenic domains, nuclear structures implicated in transcription of regulation of apoptotic factors (Kawai et al., 2003). A marked drop in basal apoptosis is observed at the polyp-to-adenoma stage of colon cancer, suggesting that resistance to apoptotic programming is acquired early in tumorigenesis (Termuhlen et al., 2002). The reduced expression of *DAPK3*, as a tumor-suppressing kinase, could provide a means for metastatic colon cancer cells to exhibit resistance to particular pathways of apoptosis. In addition, DAPK is associated with autophagic cell death; DAPK3 interaction with autophagy-related (ATG)-1 protein kinase allows for regulation of autophagosome formation via actomyosin activation (Tang et al., 2011). DAPK3 was also found to phosphorylate and increase the activity of Unc-51-like autophagy activating kinase (ULK1; the mammalian homolog of ATG1), thereby driving autophagy induction (Li et al., 2021). Further to this finding, the co-expression of *ULK1* and *DAPK3* was correlated with favorable survival outcomes in patients with gastric cancer (Li et al., 2021). DAPK3 promotes actin reorganization and actomyosin contraction by controlling the phosphorylation of myosin-regulatory light chain (Moffat et al., 2011) which has a broad range of cellular impacts, including cell migration, focal adhesion regulation, stress fiber bundling, and autophagic cell death. In addition, DAPK3 can attenuate myosin light chain phosphatase activity (Borman et al., 2002; MacDonald et al., 2001), and the loss of myosin phosphatase was further linked to cancer cell nuclear dysmorphia and genome instability (Takaki et al., 2017).

Although *DAPK*3 displays tumor suppressor properties (Li et al., 2015, 2021; Das et al., 2016; Kocher et al., 2015; Brognard et al., 2011; Bi et al., 2009), it also plays an oncogenic role with regulation of transcriptional and translational programs that are tightly connected to cell survival, proliferation, and growth. In this regard, DAPK3 promotes epithelial-mesenchymal transition (Kake et al., 2017; Li et al., 2015). DAPK3 also provides transcriptional regulation of canonical Wnt/β-catenin signaling through an interaction with Nemo-like kinase to drive enhanced proliferation of colon carcinoma cells (Togi et al., 2011). Additionally, DAPK3 can directly enhance the transcriptional activity of both STAT3 and the DNA-binding androgen receptor (Felten et al., 2013; Sato et al., 2006). As previously hypothesized by Kake et al (Kake et al., 2017), DAPK3 may change from a tumor suppressor into a tumor promoter during the course of tumor progression. In the present study, we show that *DAPK3* expression is reduced in samples obtained from patients with colon cancer; however, there is significant correlation for *DAPK3* association with LVI and PNI properties and worse survival probability. Further investigations will be required to solve the incongruity observed with respect to the proapoptotic role of DAPK3 in tumor suppression versus its proliferative influence on tumorigenesis.

TCGA database is frequently used to uncover molecular mechanisms and potential biomarkers associated with CRC progression and prognosis. For instance, Liu et al utilized TCGA to show that the prognostic power of the cancerous inhibitor of protein phosphatase 2A (*CIP2A*) was linked to activating transcription factor 6 (ATF6)-dependent ER stress signaling in CRC (Liu et al., 2018). By the same token, V-set and transmembrane-domain-containing 2A (*VSTM2A*) was identified as one of the most significantly downregulated secretory proteins in CRC, with low expression levels equating to worse overall survival in patients with CRC by way of Wnt signaling suppression (Dong et al., 2019). As well, the TCGA database was used by Taha-Mehlitz et al to support their discovery of adenylosuccinate lyase (*ADSL*) as a contributor to mitochondrial dysfunction within the context of CRC (Taha-Mehlitz et al., 2021). Finally, TCGA has facilitated the

construction of prognostic indices (Hou et al., 2018) and nomograms (Lee et al., 2019) for CRC which outperform conventional prediction tools that are based solely on clinicopathological risk factors such as age and Tumor, Node, Metastasis (TNM) staging.

WGCNA of TCGA data has been used in multiple studies to derive system-level properties, to identify genes and/or gene modules associated with specific clinicopathological variables, or to relate prespecified genes and/or gene sets to pathways or functions within the context of CRC. Application of WGCNA to TCGA data by Wang et al revealed patients with CRC with median immunity to have enhanced activation of TGF-β, vascular endothelial growth factor (VEGF), and MAPK signaling pathways and have better survival outcome; in contrast, a low-immunity subgroup displayed the worst prognosis (Wang et al., 2020). Similarly, Su et al found that higher neutrophil scores associate with poorer prognosis and that a *SOD2-CXCL8* neutrophil recruitment axis may play a role in CRC progression (Su et al., 2021). Recently, Jiang et al combined the power of TCGA and WGCNA to investigate the oncogenic functions of B-cell lymphoma 9 (*BCL9*). This investigation uncovered a β-catenin-independent function of BCL9 in a poor-prognosis subtype of CRC tumors that enhances tumor progression through its interaction with paraspeckle proteins (Jiang et al., 2020).

Some studies that utilize WGCNA to identify modules of highly correlated biomolecules restrict the data input to include only differentially expressed transcripts. A study by Zhou et al. constructed miRNA-gene interaction networks, and the authors subsequently identified two important gene modules with 11 of the 20 hub genes demonstrating prognostic value in survival analyses (Zhou et al., 2018). Although not the focus of the study, LVI was associated with both gene modules and subsequent KEGG analysis (i.e., Kyoto Encyclopedia of Genes and Genomes) identified extracellular matrix (ECM)-receptor interaction, focal adhesion, and vascular smooth muscle contraction as significantly enriched pathways. A study by Wu et al. cross-referenced DEGs with target genes for differentially expressed miRNAs from which *GCNT4*, *EDN2*, and miR-1295 were observed as network hubs (Wu et al., 2017). Given the theoretical distinction between differential expression and differential correlation and the process by which WGCNA determines intramodular hub genes, filtering genes by differential expression before WGCNA can result in loss of information. This information is pertinent to gene network construction, the loss of which may negatively impact the integrity of the co-expression network analysis (Langfelder and Horvath, 2008). Nevertheless, differential gene expression is a valuable component in network analyses that look to pinpoint candidate biomarkers, therapeutic targets, and/or gene signatures for diagnosis or prognosis.

One viable method of integrating differential gene expression into a co-expression network that is constructed with all genes (i.e., DEGs and non-DEGs) is to project the DEGs *post hoc* onto said network after the identification of gene modules. An example of this approach is apparent in the study conducted by Jiang et al, where candidate BCL9-interacting proteins and genes downstream of BCL9 were projected onto a WGCNA correlation matrix constructed with all genes (Jiang et al., 2020). The authors found that most of the genes downstream of BCL9, but not the BCL9-interacting proteins, mapped into specific gene modules involved in processes such as ECM remodeling, neuron differentiation, and wound healing. Similarly, we applied the WGCNA method to systematically identify co-expression network modules associated with lymphatic invasion in colon adenocarcinoma. Our co-expression network was constructed using all genes (minus those removed during quality control), and the DEGs were then projected onto the gene module with the highest significant correlation with LVI. Finally, for our differential expression analysis, we compared TCGA tumor samples against GTEx normal samples that were reanalyzed using the same RNA-Seq pipeline to eliminate batch effects (Vivian et al., 2017). This eliminates potential noises of tumor microenvironments stemming from the proximity of "TCGA solid tissue normal" to "TCGA primary tumor" samples. To our knowledge, no previous published studies have investigated LVI in colon adenocarcinoma utilizing the methods described in the present study.

Limitations of the study

The presented analyses were completed using the TCGA and GTEx RNA-Seq datasets. Although these datasets excel at having sufficient observations for statistically sound correlation studies, similar analyses need to be repeated on additional datasets to confirm our findings. The described networks characterized a core gene set associated with invasion and survival of tumor bulk tissues. Given that genes may demonstrate diverse functions across different cell types, gene sets identified from averaged dataset need to be reexamined in a cell-specific manner for the identification of susceptible cell types and converged pathways among different cells. The described networks were based on the analysis of gene-to-gene correlations, which did not indicate causal relationships.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Data preprocessing
  - Identification of DEGs
  - Weighted gene co-expression network analysis
  - Generation of module-trait relationships and pathway enrichment analysis of genes
  - Survival analysis
  - Differential gene correlation analysis
- QUANTIFICATION AND STATISTICAL ANALYSES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102831.

AUTHOR CONTRIBUTIONS

H.-M.C. conceived the study, completed the data analysis, and prepared figures. J.A.M. coordinated the study, secured grant funding support, wrote the manuscript, supervised trainees, and provided intellectual contributions to the project. All authors reviewed the results and approved the final version of the manuscript.

DECLARATION OF INTERESTS

Justin A. MacDonald is a cofounder and a shareholder of Arch Biopartners Inc. Huey-Miin Chen has no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

REFERENCES

Alexa, A., Rahnenführer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics *13*, 1600–1607. https://doi.org/10.1093/bioinformatics/btl140.

Anand, S., Foot, N., Ang, C.S., Gembus, K.M., Keerthikumar, S., Adda, C.G., Mathivanan, S., and

Kumar, S. (2018). Arrestin-domain containing protein 1 (Arrdc1) regulates the protein cargo and release of extracellular vesicles. Proteomics *18*, e1800266. https://doi.org/10.1002/pmic. 201800266.

Berthenet, K., Castillo Ferrer, C., Fanfone, D., Popgeorgiev, N., Neves, D., Bertolino, P., Gibert,

B., Hernandez-Vargas, H., and Ichim, G. (2020). Failed apoptosis enhances melanoma cancer cell aggressiveness. Cell Rep. *31*, 107731. https://doi. org/10.1016/j.celrep.2020.107731.

Bi, J., Lau, S.H., Hu, L., Rao, H.L., Liu, H.B., Zhan, W.H., Chen, G., Wen, J.M., Wang, Q., Li, B., and Guan, X.Y. (2009). Downregulation of ZIP kinase is

associated with tumor invasion, metastasis and poor prognosis in gastric cancer. Int. J. Cancer *124*, 1587–1593. https://doi.org/10.1002/ijc.24164.

Borman, M.A., MacDonald, J.A., Murányi, A., Hartshorne, D.J., and Haystead, T.A. (2002). Smooth muscle myosin phosphatase-associated kinase induces Ca2+ sensitization via myosin phosphatase inhibition. J. Biol. Chem. *277*, 23441–23446. https://doi.org/10.1074/jbc.M201597200.

Brognard, J., Zhang, Y.W., Puto, L.A., and Hunter, T. (2011). Cancer-associated loss-of-function mutations implicate DAPK3 as a tumor-suppressing kinase. Cancer Res. *71*, 3152–3161. https://doi.org/10.1158/0008-5472.CAN-10-3543.

Caino, M.C., Seo, J.H., Aguinaldo, A., Wait, E., Bryant, K.G., Kossenkov, A.V., Hayden, J.E., Vaira, V., Morotti, A., Ferrero, S., et al. (2016). A neuronal network of mitochondrial dynamics regulates metastasis. Nat. Commun. *7*, 13730. https://doi.org/10.1038/ncomms13730.

Chen, W., Zhang, W., Wu, R., Cai, Y., Xue, X., and Cheng, J. (2019). Identification of biomarkers associated with histological grade and prognosis of gastric cancer by co-expression network analysis. Oncol. Lett. *18*, 5499–5507. https://doi.org/10.3892/ol.2019.10869.

Cho, N.Y., Park, J.W., Wen, X., Shin, Y.J., Kang, J.K., Song, S.H., Kim, H.P., Kim, T.Y., Bae, J.M., and Kang, G.H. (2020). Blood-based detection of colorectal cancer using cancer-specific DNA methylation markers. Diagnostics (Basel) *11*, 51. https://doi.org/10.3390/diagnostics11010051.

Chow, K.H., Factor, R.E., and Ullman, K.S. (2012). The nuclear envelope environment and its cancer connections. Nat. Rev. Cancer *12*, 196–209. https://doi.org/10.1038/nrc3219.

Cienfuegos, J.A., Martinez, P., Baixauli, J., Beorlegui, C., Rosenstone, S., Sola, J.J., Rodriguez, J., and Hernandez-Lizoain, J.L. (2017). Perineural invasion is a major prognostic and predictive factor of response to adjuvant chemotherapy in Stage I-II colon cancer. Ann. Surg. Oncol. *24*, 1077–1084. https://doi.org/10.1245/s10434-016-5561-0.

Crow, M., Lim, N., Ballouz, S., Pavlidis, P., and Gillis, J. (2019). Predictability of human differential gene expression. Proc. Natl. Acad. Sci. U S A *116*, 6491–6500. https://doi.org/10.1073/pnas.1802973116.

Das, T.P., Suman, S., Papu John, A.M., Pal, D., Edwards, A., Alatassi, H., Ankem, M.K., and Damodaran, C. (2016). Activation of AKT negatively regulates the pro-apoptotic function of death-associated protein kinase 3 (DAPK3) in prostate cancer. Cancer Lett. *377*, 134–139. https://doi.org/10.1016/j.canlet.2016.04.028.

Dekker, E., Tanis, P.J., Vleugels, J.L.A., Kasi, P.M., and Wallace, M.B. (2019). Colorectal cancer. Lancet *394*, 1467–1480. https://doi.org/10.1016/S0140-6736(19)32319-0.

Dienstmann, R., Villacampa, G., Sveen, A., Mason, M.J., Niedzwiecki, D., Nesbakken, A., Moreno, V., Warren, R.S., Lothe, R.A., and Guinney, J. (2019). Relative contribution of clinicopathological variables, genomic markers, transcriptomic

subtyping and microenvironment features for outcome prediction in stage II/III colorectal cancer. Ann. Oncol. *30*, 1622–1629. https://doi.org/10.1093/annonc/mdz287.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Doncheva, N.T., Morris, J.H., Gorodkin, J., and Jensen, L.J. (2019). Cytoscape StringApp: network analysis and visualization of proteomics data. J. Proteome. Res. *18*, 623–632. https://doi.org/10.1021/acs.jproteome.8b00702.

Dong, Y., Zhang, Y., Kang, W., Wang, G., Chen, H., Higashimori, A., Nakatsu, G., Go, M., Tong, J.H., Zheng, S., et al. (2019). VSTM2A suppresses colorectal cancer and antagonizes Wnt signaling receptor LRP6. Theranostics *9*, 6517–6531. https://doi.org/10.7150/thno.34989.

Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell *61*, 759–767. https://doi.org/10.1016/0092-8674(90)90186-i.

Felten, A., Brinckmann, D., Landsberg, G., and Scheidtmann, K.H. (2013). Zipper-interacting protein kinase is involved in regulation of ubiquitination of the androgen receptor, thereby contributing to dynamic transcription complex assembly. Oncogene *32*, 4981–4988. https://doi.org/10.1038/onc.2012.503.

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D.M., Pineros, M., Znaor, A., and Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int. J. Cancer *144*, 1941–1953. https://doi.org/10.1002/ijc.31937.

Fletcher, D.A., and Mullins, R.D. (2010). Cell mechanics and the cytoskeleton. Nature *463*, 485–492. https://doi.org/10.1038/nature08908.

Goldman, M.J., Craft, B., Hastie, M., Repecka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat. Biotechnol. *38*, 675–678. https://doi.org/10.1038/s41587-020-0546-8.

Hou, X., He, X., Wang, K., Hou, N., Fu, J., Jia, G., Zuo, X., Xiong, H., and Pang, M. (2018). Genome-wide network-based analysis of colorectal cancer identifies novel prognostic factors and an integrative prognostic index. Cell. Physiol. Biochem. *49*, 1703–1716. https://doi.org/10.1159/000493614.

Ichim, G., and Tait, S.W. (2016). A fate worse than death: apoptosis as an oncogenic process. Nat. Rev. Cancer *16*, 539–548. https://doi.org/10.1038/nrc.2016.58.

Irianto, J., Pfeifer, C.R., Ivanovska, I.L., Swift, J., and Discher, D.E. (2016). Nuclear lamins in cancer. Cell. Mol. Bioeng. *9*, 258–267. https://doi.org/10.1007/s12195-016-0437-8.

Jiang, M., Kang, Y., Sewastianik, T., Wang, J., Tanton, H., Alder, K., Dennis, P., Xin, Y., Wang, Z., Liu, R., et al. (2020). BCL9 provides multi-cellular communication properties in colorectal cancer by interacting with paraspeckle proteins. Nat.

Commun. *11*, 19. https://doi.org/10.1038/s41467-019-13842-7.

Kake, S., Usui, T., Ohama, T., Yamawaki, H., and Sato, K. (2017). Death-associated protein kinase 3 controls the tumor progression of A549 cells through ERK MAPK/c-Myc signaling. Oncol. Rep. *37*, 1100–1106. https://doi.org/10.3892/or.2017.5359.

Kawai, T., Akira, S., and Reed, J.C. (2003). ZIP kinase triggers apoptosis from nuclear PML oncogenic domains. Mol. Cell Biol. *23*, 6174–6186. https://doi.org/10.1128/MCB.23.17.6174-6186.2003.

Kocher, B.A., White, L.S., and Piwnica-Worms, D. (2015). DAPK3 suppresses acini morphogenesis and is required for mouse development. Mol. Cancer Res. *13*, 358–367. https://doi.org/10.1158/1541-7786.MCR-14-0333.

Komatsu, S., and Ikebe, M. (2004). ZIP kinase is responsible for the phosphorylation of myosin II and necessary for cell motility in mammalian fibroblasts. J. Cell Biol. *165*, 243–254. https://doi.org/10.1083/jcb.200309056.

Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. Nat. Rev. Mol. Cell Biol. *15*, 178–196. https://doi.org/10.1038/nrm3758.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559. https://doi.org/10.1186/1471-2105-9-559.

Law, C.W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G.K., and Ritchie, M.E. (2016). RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Res. *5*. ISCB Comm J-1408. https://doi.org/10.12688/f1000research.9005.3.

Lee, J.H., Jung, S., Park, W.S., Choe, E.K., Kim, E., Shin, R., Heo, S.C., Lee, J.H., Kim, K., and Chai, Y.J. (2019). Prognostic nomogram of hypoxia-related genes predicting overall survival of colorectal cancer-analysis of TCGA database. Sci. Rep. *9*, 1803. https://doi.org/10.1038/s41598-018-38116-y.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323. https://doi.org/10.1186/1471-2105-12-323.

Li, G.M., Li, L., Li, M.Q., Chen, X., Su, Q., Deng, Z.J., Liu, H.B., Li, B., Zhang, W.H., Jia, Y.X., et al. (2021). DAPK3 inhibits gastric cancer progression via activation of ULK1-dependent autophagy. Cell Death Differ. *28*, 952–967. https://doi.org/10.1038/s41418-020-00627-5.

Li, J., Deng, Z., Wang, Z., Wang, D., Zhang, L., Su, Q., Lai, Y., Li, B., Luo, Z., Chen, X., et al. (2015). Zipper-interacting protein kinase promotes epithelial-mesenchymal transition, invasion and metastasis through AKT and NF-kB signaling and is associated with metastasis and poor prognosis in gastric cancer patients. Oncotarget *6*, 8323–8338. https://doi.org/10.18632/oncotarget.3200.

Li, J., Zhou, D., Qiu, W., Shi, Y., Yang, J.J., Chen, S., Wang, Q., and Pan, H. (2018). Application of weighted gene co-expression network analysis

for data from paired design. Sci. Rep. *8*, 622. https://doi.org/10.1038/s41598-017-18705-z.

Liu, C.Y., Hsu, C.C., Huang, T.T., Lee, C.H., Chen, J.L., Yang, S.H., Jiang, J.K., Chen, W.S., Lee, K.D., and Teng, H.W. (2018). ER stress-related ATF6 upregulates CIP2A and contributes to poor prognosis of colon cancer. Mol. Oncol. *12*, 1706–1717. https://doi.org/10.1002/1878-0261.12365.

Liebig, C., Ayala, G., Wilks, J., Verstovsek, G., Liu, H., Agarwal, N., Berger, D.H., and Albo, D. (2009). Perineural invasion is an independent predictor of outcome in colorectal cancer. J. Clin. Oncol. *27*, 5131–5137. https://doi.org/10.1200/JCO.2009.22.4949.

MacDonald, J.A., Borman, M.A., Murányi, A., Somlyo, A.V., Hartshorne, D.J., and Haystead, T.A. (2001). Identification of the endogenous smooth muscle myosin phosphatase-associated kinase. Proc. Natl. Acad. Sci. U S A *98*, 2419–2424. https://doi.org/10.1073/pnas.041331498.

Malki, A., Elruz, R.A., Gupta, I., Allouch, A., Vranic, S., and Al Moustafa, A.E. (2020). Molecular mechanisms of colon cancer progression and metastasis: recent insights and advancements. Int. J. Mol. Sci. *22*, 130. https://doi.org/10.3390/ijms22010130.

Marmol, I., Sanchez-De-Diego, C., Pradilla Dieste, A., Cerrada, E., and Rodriguez Yoldi, M.J. (2017). Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. Int. J. Mol. Sci. *18*, 197. https://doi.org/10.3390/ijms18010197.

Mattioni, A., Boldt, K., Auciello, G., Komada, M., Rappoport, J.Z., Ueffing, M., Castagnoli, L., Cesareni, G., and Santonico, E. (2020). Ring finger protein 11 acts on ligand-activated EGFR via the direct interaction with the UIM region of ANKRD13 protein family. FEBS J. *287*, 3526–3550. https://doi.org/10.1111/febs.15226.

McCarthy, D.J., and Smyth, G.K. (2009). Testing significance relative to a fold-change threshold is a TREAT. Bioinformatics *25*, 765–771. https://doi.org/10.1093/bioinformatics/btp053.

McKenzie, A.T., Katsyv, I., Song, W.M., Wang, M., and Zhang, B. (2016). DGCA: a comprehensive R package for differential gene correlation analysis. BMC Syst. Biol. *10*, 106. https://doi.org/10.1186/s12918-016-0349-1.

Moffat, L.D., Brown, S.B., Grassie, M.E., Ulke-Lemée, A., Williamson, L.M., Walsh, M.P., and MacDonald, J.A. (2011). Chemical genetics of zipper-interacting protein kinase reveal myosin light chain as a bona fide substrate in permeabilized arterial smooth muscle. J. Biol. Chem. *286*, 36978–36991. https://doi.org/10.1074/jbc.M111.257949.

Mutabdzic, D., O'Brien, S.B.L., Handorf, E.A., Devarajan, K., Reddy, S.R., Sigurdson, E.R., Denlinger, C.S., Meyer, J.E., and Farma, J.M. (2019). Evaluating the prognostic significance of lymphovascular invasion in stage II and III colon cancer. J. Clin. Oncol. *37*, 685.

Nabhan, J.F., Hu, R., Oh, R.S., Cohen, S.N., and Lu, Q. (2012). Formation and release of arrestin domain-containing protein 1-mediated microvesicles (ARMMs) at plasma membrane by recruitment of TSG101 protein. Proc. Natl. Acad. Sci. U S A *109*, 4146–4151. https://doi.org/10.1073/pnas.1200448109.

Nehru, V., Almeida, F.N., and Aspenstrom, P. (2013). Interaction of RhoD and ZIP kinase modulates actin filament assembly and focal adhesion dynamics. Biochem. Biophys. Res. Commun. *433*, 163–169. https://doi.org/10.1016/j.bbrc.2013.02.046.

Qin, D., Wei, R., Liu, S., Zhu, S., Zhang, S., and Min, L. (2019). A circulating miRNA-based scoring system established by WGCNA to predict colon cancer. Anal. Cell. Pathol. (Amst). *2019*, 1571045. https://doi.org/10.1155/2019/1571045.

Rana, M.K., Aloisio, F.M., Choi, C., and Barber, D.L. (2018). Formin-dependent TGF-beta signaling for epithelial to mesenchymal transition. Mol. Biol. Cell *29*, 1465–1475. https://doi.org/10.1091/mbc.E17-05-0325.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

Sato, N., Kamada, N., Muromoto, R., Kawai, T., Sugiyama, K., Watanabe, T., Imoto, S., Sekine, Y., Ohbayashi, N., Ishida, M., et al. (2006). Phosphorylation of threonine-265 in zipper-interacting protein kinase plays an important role in its activity and is induced by IL-6 family cytokines. Immunol. Lett. *103*, 127–134. https://doi.org/10.1016/j.imlet.2005.10.015.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504. https://doi.org/10.1101/gr.1239303.

Shiloh, R., Bialik, S., and Kimchi, A. (2014). The DAPK family: a structure-function analysis. Apoptosis *19*, 286–297. https://doi.org/10.1007/s10495-013-0924-5.

Sjo, O.H., Merok, M.A., Svindland, A., and Nesbakken, A. (2012). Prognostic impact of lymph node harvest and lymph node ratio in patients with colon cancer. Dis. Colon Rectum *55*, 307–315. https://doi.org/10.1097/DCR.0b013e3182423f62.

Skancke, M., Arnott, S.M., Amdur, R.L., Siegel, R.S., Obias, V.J., and Umapathi, B.A. (2019). Lymphovascular invasion and perineural invasion negatively impact overall survival for Stage II adenocarcinoma of the colon. Dis. Colon Rectum *62*, 181–188. https://doi.org/10.1097/DCR.0000000000001258.

Su, H., Cai, T., Zhang, S., Yan, X., Zhou, L., He, Z., Xue, P., Li, J., Zheng, M., Yang, X., and Feng, B. (2021). Identification of hub genes associated with neutrophils infiltration in colorectal cancer. J. Cell. Mol. Med. *25*, 3371–3380. https://doi.org/10.1111/jcmm.16414.

Sun, Q., Liu, T., Liu, P., Luo, J., Zhang, N., Lu, K., Ju, H., Zhu, Y., Wu, W., Zhang, L., et al. (2019). Perineural and lymphovascular invasion predicts for poor prognosis in locally advanced rectal cancer after neoadjuvant chemoradiotherapy and surgery. J. Cancer *10*, 2243–2249. https://doi.org/10.7150/jca.31473.

Taha-Mehlitz, S., Bianco, G., Coto-Llerena, M., Kancherla, V., Bantug, G.R., Gallon, J., Ercan, C., Panebianco, F., Eppenberger-Castori, S., von Strauss, M., et al. (2021). Adenylosuccinate lyase is oncogenic in colorectal cancer by causing mitochondrial dysfunction and independent activation of NRF2 and mTOR-MYC-axis. Theranostics *11*, 4011–4029. https://doi.org/10.7150/thno.50051.

Takaki, T., Montagner, M., Serres, M.P., Le Berre, M., Russell, M., Collinson, L., Szuhai, K., Howell, M., Boulton, S.J., Sahai, E., and Petronczki, M. (2017). Actomyosin drives cancer cell nuclear dysmorphia and threatens genome stability. Nat. Commun. *8*, 16013. https://doi.org/10.1038/ncomms16013.

Tang, H.W., Wang, Y.B., Wang, S.L., Wu, M.H., Lin, S.Y., and Chen, G.C. (2011). Atg1-mediated myosin II activation regulates autophagosome formation during starvation-induced autophagy. EMBO J. *30*, 636–651. https://doi.org/10.1038/emboj.2010.338.

Tanno, H., Yamaguchi, T., Goto, E., Ishido, S., and Komada, M. (2012). The Ankrd 13 family of UIM-bearing proteins regulates EGF receptor endocytosis from the plasma membrane. Mol. Biol. Cell *23*, 1343–1353. https://doi.org/10.1091/mbc.E11-09-0817.

Termuhlen, P.M., Sweeney-Gotsch, B.M., Berman, R.S., Ellis, L.M., Bucana, C., Shen, Y., Cleary, K.R., and McConkey, D.J. (2002). Increased apoptosis in metastatic human colonic adenocarcinomas. Cancer Biol. Ther. *1*, 58–63. https://doi.org/10.4161/cbt.1.1.43.

Togi, S., Ikeda, O., Kamitani, S., Nakasuji, M., Sekine, Y., Muromoto, R., Nanbo, A., Oritani, K., Kawai, T., Akira, S., and Matsuda, T. (2011). Zipper-interacting protein kinase (ZIPK) modulates canonical Wnt/beta-catenin signaling through interaction with Nemo-like kinase and T-cell factor 4 (NLK/TCF4). J. Biol. Chem. *286*, 19170–19177. https://doi.org/10.1074/jbc.M110.189829.

Toth, C., Meinrath, J., Herpel, E., Derix, J., Fries, J., Buettner, R., Schirmacher, P., and Heikaus, S. (2016). Expression of the apoptosis repressor with caspase recruitment domain (ARC) in liver metastasis of colorectal cancer and its correlation with DNA mismatch repair proteins and p53. J. Cancer Res. Clin. Oncol. *142*, 927–935. https://doi.org/10.1007/s00432-015-2102-3.

Vasiliev, J.M. (2004). Cytoskeletal mechanisms responsible for invasive migration of neoplastic cells. Int. J. Dev. Biol. *48*, 425–439. https://doi.org/10.1387/ijdb.041806jv.

Vivian, J., Rao, A.A., Nothaft, F.A., Ketchum, C., Armstrong, J., Novak, A., Pfeil, J., Narkizian, J., Deran, A.D., Musselman-Brown, A., et al. (2017). Toil enables reproducible, open source, big biomedical data analyses. Nat. Biotechnol. *35*, 314–316. https://doi.org/10.1038/nbt.3772.

Wang, Q., Armenia, J., Zhang, C., Penson, A.V., Reznik, E., Zhang, L., Minet, T., Ochoa, A., Gross, B.E., Iacobuzio-Donahue, C.A., et al. (2018). Unifying cancer and normal RNA sequencing data from different sources. Sci. Data 5, 180061. https://doi.org/10.1038/sdata.2018.61.

Wang, X., Duanmu, J., Fu, X., Li, T., and Qiang, Q. (2020). Analyzing and validating the prognostic value and mechanism of colon cancer immune microenvironment. J. Transl. Med. 18, 324. https://doi.org/10.1186/s12967-020-02491-w.

Weiser, M.R., Gonen, M., Chou, J.F., Kattan, M.W., and Schrag, D. (2011). Predicting survival after curative colectomy for cancer: individualizing colon cancer staging. J. Clin. Oncol. 29, 4796–4802. https://doi.org/10.1200/JCO.2011.36.5080.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (New York: Springer-Verlag).

Wu, F., Yuan, G., Chen, J., and Wang, C. (2017). Network analysis based on TCGA reveals hub genes in colon cancer. Contemp. Oncol. (Pozn) 21, 136–144. https://doi.org/10.5114/wo.2017.68622.

Xiao, J., Shi, Q., Li, W., Mu, X., Peng, J., Li, M., Chen, M., Huang, H., Wang, C., Gao, K., and Fan, J. (2018). ARRDC1 and ARRDC3 act as tumor suppressors in renal cell carcinoma by facilitating YAP1 degradation. Am. J. Cancer Res. 8, 132–143.

Xin, Y., Chen, X., Tang, X., Li, K., Yang, M., Tai, W.C., Liu, Y., and Tan, Y. (2019). Mechanics and actomyosin-dependent survival/chemoresistance of suspended tumor cells in shear flow. Biophys. J. 116, 1803–1814. https://doi.org/10.1016/j.bpj.2019.04.011.

Yang, S.Y., Sales, K.M., Fuller, B., Seifalian, A.M., and Winslet, M.C. (2009). Apoptosis and colorectal cancer: implications for therapy. Trends Mol. Med. 15, 225–233. https://doi.org/10.1016/j.molmed.2009.03.003.

Yang, Y., Han, L., Yuan, Y., Li, J., Hei, N., and Liang, H. (2014). Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. Nat. Commun. 5, 3231. https://doi.org/10.1038/ncomms4231.

Zhai, X., Xue, Q., Liu, Q., Guo, Y., and Chen, Z. (2017). Colon cancer recurrence-associated genes revealed by WGCNA coexpression network analysis. Mol. Med. Rep. 16, 6499–6505. https://doi.org/10.3892/mmr.2017.7412.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 4. Article17. https://doi.org/10.2202/1544-6115.1128.

Zhang, L., and Yu, J. (2013). Role of apoptosis in colon cancer biology, therapy, and prevention. Curr. Colorectal Cancer Rep. 9. https://doi.org/10.1007/s11888-013-0188-z.

Zhang, Z., Wang, L., Wang, Q., Zhang, M., Wang, B., Jiang, K., Ye, Y., Wang, S., and Shen, Z. (2020). Molecular characterization and clinical relevance of RNA binding proteins in colorectal cancer. Front. Genet. 11, 580149. https://doi.org/10.3389/fgene.2020.580149.

Zhou, X.G., Huang, X.L., Liang, S.Y., Tang, S.M., Wu, S.K., Huang, T.T., Mo, Z.N., and Wang, Q.Y. (2018). Identifying miRNA and gene modules of colon cancer associated with pathological stage by weighted gene co-expression network analysis. Onco. Targets Ther. 11, 2815–2830. https://doi.org/10.2147/OTT.S163891.

Zou, Y., and Jing, L. (2019). Identification of key modules and prognostic markers in adrenocortical carcinoma by weighted gene co-expression network analysis. Oncol. Lett. 18, 3673–3681. https://doi.org/10.3892/ol.2019.10725.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
TCGA TARGET GTEx study	Xena Toil RNA-Seq Recompute Compendium	https://toil.xenahubs.net
TCGA survival data	Xena TCGA data hub	https://tcga.xenahubs.net
COAD clinicalMatrix data	Xena TCGA data hub	https://tcga.xenahubs.net
Software and algorithms		
R (v4.0.2)	The R Project	https://www.r-project.org
RStudio (1.3.1093-1)	RStudio Team	http://www.rstudio.com
Cytoscape (v3.8.2)	Shannon et al. (2003)	https://cytoscape.org
Limma (v3.48.0)	Ritchie et al. (2015)	https://bioconductor.org
stringApp	Doncheva et al. (2019)	http://apps.cytoscape.org/apps/stringapp
topGO (release 3.13)	Alexa et al. (2006)	https://bioconductor.org
Xena browser	Computational Genomics Lab of the University of California Santa Cruz	http://xena.ucsc.edu
ggplot2 (v3.3.3)	Wickham (2016)	http://ggplot2.tidyverse.org
DGCA (v1.0.2)	McKenzie et al. (2016)	https://cran.rproject.org/web/packages/DGCA/index.html

RESOURCE AVAILABILITY

Lead contact

Requests for further information should be directed to and will be fulfilled by the lead contact, Justin A. MacDonald (jmacdo@ucalgary.ca).

Materials availability

This study did not generate new unique reagents.

Data and code availability

1. Data

This study analyzes existing, publicly available RNA-Seq data. The sources for the datasets are listed in the key resources table.

2. Code

This study does not report original code. All codes were used in this study in alignment with recommendations made by authors of R packages in their respective user's guide, which can be accessed at https://bioconductor.org.

3. Additional information requests

Any additional information required to reanalyze the data used in this study is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

RNA sequencing data (RNA-Seq) for the GTEx (https://gtexportal.org/home/) and TCGA (https://portal.gdc.cancer.gov) were retrieved from the Xena Toil RNA-Seq Recompute Compendium (https://toil.

xenahubs.net) on 30 November 2020. Although TCGA includes normal samples, these "solid tissue normal" samples are derived from normal tissues located proximal to the tumor. Consequently, they may also possess tumor transcriptomic profiles. In contrast, samples from the GTEx project provide expression data from the normal tissue of healthy, cancer-free individuals. The datasets from the two projects cannot be compared directly, so the "TCGA TARGET GTEx study" with samples restricted to |Primary_-site| = "colon" was retrieved from the Xena Toil data hub where TCGA, TARGET, and GTEx samples were reanalyzed (i.e., realigned to hg38 genome with expressions defined using RSEM and Kallisto methods) by the same RNA-Seq pipeline. The Toil Compendium provides recomputed TCGA and GTEx expression raw data based on a uniform bioinformatic/RNA-Seq pipeline to eliminate batch effect due to different computational processing (Vivian et al., 2017). The Xena Toil pipeline utilized STAR for uniform realignment (Dobin et al., 2013); RSEM and Kallisto were used to produce quantification data (Li and Dewey, 2011). The realignment also enabled the removal of degraded samples before quantification and batch effect correction after quantification. In addition, TCGA survival data and COAD clinicalMatrix information of patients with colon cancer were retrieved from the Xena TCGA data hub (https://tcga.xenahubs.net). The Xena browser (http://xena.ucsc.edu), hosted by the Computational Genomics Lab of the University of California Santa Cruz (UCSC; https://cglgenomics.ucsc.edu), was utilized to compare gene expressions (Goldman et al., 2020). Data retrieval was completed via the UCSCXenaTools R package.

METHOD DETAILS

Data preprocessing

The log-transformed expected count retrieved from Xena Toil was back-transformed via the |round(((2^x)-1),0)| function. The data were further refined to include only "Normal Tissue" from the GTEx gene sets and only "Primary Tumor" with |histological_type| = "Colon Adenocarcinoma" from the TCGA gene sets. TCGA and GTEx datasets were filtered to exclude non-protein-coding genes. The TARGET dataset was excluded to permit focus on adult samples. Data from metastatic and recurrent tumors were also excluded owing to small sample sizes. After these preprocessing steps, the dataset included 551 patient samples and 18,205 genes. Data processing was completed using the R (v4.0.2) programming language. All codes used in this study align with recommendations made by authors of R packages in their respective user's guide, which can be accessed at https://bioconductor.org.

Identification of DEGs

The limma workflow was used to detect DEGs between cancer and normal samples (Robinson et al., 2010; Ritchie et al., 2015; Law et al., 2016). The input data (18,205 genes, 551 samples) were screened with function |filterByExpr.default| to eliminate outliers and lowly expressed genes. Scaling factors were computed with function |calcNormFactors.upperquartile|. The final DGEList object contained 15,164 genes and 547 samples (303 GTEx, 244 TCGA). Heteroscedasticity was removed via function |voom|. Expression values of TCGA were then compared with that of GTEx through linear modeling; namely, the limma functions |lmFit|, |contrasts.fit|, and |eBayes|. All functions were operated with default settings. The transformed expression values, generated by |voom|, were exported for subsequent analyses.

Weighted gene co-expression network analysis

To identify candidate biomarkers for lymphatic invasion within the colon adenocarcinoma subset, we applied WGCNA to voom-transformed expression values (Langfelder and Horvath, 2008). Genes with negative normalized expression levels were removed, and then, the WGCNA function |goodSamples-Genes| was used to identify genes with zero variance. Sample outliers were detected via function |hclust| and then removed with the WGCNA function |cutreeStatic.minSize=10|. The data input for WGCNA consisted of 231 TCGA samples and 9,750 genes. The network topology was calculated for a range of soft-thresholding power (β). Then, the scale-free topology fit index was plotted as a function of the soft-thresholding power, and the mean connectivity as a function of the soft-thresholding power was determined. Although β = 9 was the lowest power for which the scale-free topology fit index reached an R^2 value of 0.90, β = 10 was ultimately used as it was recommended as the minimum soft-threshold power to be used for a signed network (Langfelder and Horvath, 2008). The co-expression similarity was raised to achieve scale-free topology with this soft-thresholding power. Co-expression similarity and adjacency were calculated for a signed network, transformed into a topological overlap matrix (TOM) to calculate the corresponding dissimilarity (dissTOM) which in turn was used to conduct hierarchical clustering. Module identification was completed with the function |cutreeDynamic| with method=hybrid,

minClusterSize=100, deepSplit=4, pamStage=TRUE, and pamRespectDendro=FALSE. Any highly similar modules with the height of clustering lower than 0.25 were merged. A clustering dendrogram was used to display the results.

Generation of module-trait relationships and pathway enrichment analysis of genes

Correlations between modules generated by WGCNA and clinical parameters were determined by module-trait relationship (MTR) analysis. MEs were calculated and related to clinical features via standard WGCNA pipeline. GO enrichment analysis was completed with topGO [algorithm=elim, statistic=fisher] to facilitate biological interpretation (Alexa et al., 2006). Finally, intramodular connectivity was analyzed using genes within the *darkred* module as search inputs. STRING (Doncheva et al., 2019) and the Cytoscape plugin NetworkAnalyzer (Shannon et al., 2003) were used for computation of topological parameters for an undirected network with protein query as the data source.

Survival analysis

Survival data were retrieved as part of the TCGA dataset downloaded from Xena and were subjected to Kaplan-Meier analysis in R. The Cox proportional-hazards model was used to investigate the association between survival time and candidate biomarkers for lymphatic invasion. Z-scale cut-offs were set at $\geq$ 0.647 for high expression and $\leq$ 0.647 for low expression. Survival time cut-off was set at 10 years. The coxph R function was used in combination with the |log-rank| test for comparison of the high/low survival curves. Survival curves were plotted to show the differences in patient mortality between high- and low-expression groups.

Differential gene correlation analysis

Using the R package DGCA (McKenzie et al., 2016), the differential correlation between *DAPK3* and a filtered list of human protein-coding genes (8,528 genes; exclusion criteria: bottom 25th percentile of median expression and/or dispersion index of expression) was computed with the |ddCorAll| function with [nPerms=100]. This pipeline provided the Pearson coefficient (*r*) and the corresponding *p* values for each pair of genes across samples (n=547). Significant changes in differential correlation between the two conditions (healthy colon vs. colon adenocarcinoma) were then identified using a Fisher's Z-test. GO term enrichment analysis of differential correlation-classified genes was completed with the DGCA function |ddcorGO| with [adjusted=TRUE, conditional=TRUE, calculateVariance=TRUE].

QUANTIFICATION AND STATISTICAL ANALYSES

Statistical analyses were performed within specific R packages. For DGCA, the pipeline provided Pearson coefficient (r) and the corresponding p values for each pair of genes across samples (n=547). Significant changes in differential correlation between the two conditions (healthy colon vs. colon adenocarcinoma) were then identified using a Fisher's Z-test. For survival analyses, the Kaplan-Meier method was used to estimate the survival cure, univariate Cox analyses was used to compute hazard ratios, and outputs from log-rank testing was used to describe overall significance of the model. For the differential expression of genes analysis, the *treat* method with a nonparametric empirical Bayes approach for the analysis of factorial data provide a paired t-test for every gene within the limma R environment (McCarthy and Smyth, 2009). For the comparison of DAPK3 expression levels, Welch's t-test for two independent samples of unequal variance was used. Data were plotted using GraphPad Prism 8.3.0 (GraphPad Software, La Jolla, CA). DAPK3 expression levels are presented graphically with median values along with the upper and lower limits of the interquartile range (IQR) between the 25th and 75th percentiles. The statistical tests used to assess experimental results are provided within the figure legends. Statistical test results (i.e., p values) are provided for all comparisons in figures, and p values < 0.05 were considered significant. The 'n' value represents the number of patient samples obtained from the RNA-Seq datasets and is marked within the figure or in the figure legend. As detailed earlier, RNA-Seq data were filtered to include only samples from adults and to exclude non-protein-coding genes; in addition, data for metastatic and recurrent tumors were excluded owing to small sample sizes.