IDENTIFICATION OF NOVEL REGULATORS OF COLORECTAL ADENOCARCINOMA AND HEAD AND NECK SQUAMOUS CELL CARCINOMA

By

Christi Lynn French

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

CANCER BIOLOGY

August, 2015

Nashville, TN

Approved:

Alissa M. Weaver, M.D., Ph.D.

Al Reynolds, Ph.D.

Andries Zijlstra, Ph.D.

Donna Webb, Ph.D.

To my wonderful husband, Will, whom I love to the ceiling

And to those in my family that have battled cancer

ACKNOWLEDGEMENTS

I must first thank my advisor, Alissa Weaver, for accepting me into her lab. Her critique and advice saw me through long years and several projects, and I appreciate her allowing me to take advantage of numerous opportunities to advance a nontraditional career. To the other members of the Weaver lab (past and present), I am exceedingly grateful to have had you as colleagues and friends. To Kellye Kirkbride, Aron Parack, Bong Hwan Sung, and Kevin Branch – thank you for your mentorship and patience when I first joined the lab. To Seema Sinha and Nan Hyung Hong, my graduate student partners-in-crime – thank you for the cupcakes and encouragement. Selma Maacha, merci beaucoup pour ton amitié et tous les cours de français. To Kaitlin Costello, Tyne Miller, Rachel Jerrell, Eric Wirtz, Andy McKenzie, and Mallory Miles – thank you for your friendship. Without the frequent coffee breaks, lab lunches, girl's nights, happy hours, and thoughtful scientific discussions I wouldn't be the scientist or person I am today.

Thank you to my committee members Al Reynolds, Andries Zijlstra, Donna Webb, and Dell Yarbrough for their mentorship and guidance. Their advice was hugely helpful to me, and I always left committee meetings with a renewed excitement about my project and dozens of new ideas.

Numerous people were invaluable to me for their help and patience as I learned about bioinformatics, statistical analysis, and computer programming, especially Fei Ye, Bing Zhang, and Darren Tyson. I am also grateful to Kay Washington, Natasha Deane, and Dan Beauchamp for their helpful suggestions to the project and manuscript. Thank you to Frank Revetta for doing the TMA IHC staining, and to Joe Roland in the Epithelial Biology Core for his help in the TMA image analysis.

The Yarbrough lab, specifically Brandee Brown, Jose Godoy, and Drew Sewell, were very generous with their time and reagents as they trained me to do mouse surgeries. A huge thank you to Brandee Brown, Kellye Kirkbride, and Seema Sinha for helping me set up multiple *in vivo* experiments. I am also grateful to the TPSR Core for performing the subsequent tissue processing and staining.

Thank you to the Cancer Biology Department for the training opportunities, and all the professors, post-docs, and graduate students that provided valuable scientific advice and feedback during symposia and Science Hour presentations. I would also like to thank Tracy Tveit and Toni Shepard for their kind guidance and support. Thank you also to the IGP program for my first year of training and funding, and to Vanderbilt University for providing a wonderful, collaborative atmosphere to conduct research. I am also grateful to the BRET office for their numerous career development events.

While earning a Ph.D. takes an indescribable amount of work, none of it would have been possible without some semblance of work/life balance. To the friends I made in graduate school, thank you for helping to keep me sane. In no particular order, thank you to Hank and Hayley Clay, Sudipta Chakraborty, Chris Muller, Emily Anderson-Baucum, Jamie and James Saxon, Scott and Sarah Collier, Shann Yu, Tish and Jonathan Warren, John and Nicki Burnam, Lara Jazmin, Laura Hover, Jason Ballengee, Margarita Prieto, Stijn Smeets, and many, many others. Thank you also to the Graduate Christian Fellowship and Trinity Church for providing me with a wonderful supportive community. I am especially thankful to the members of the various small groups I've been in over the years.

I am also thankful to my family for being constant cheerleaders in all of my life endeavors. To all of my family members, thank you: Karen, Coye, and McKenzie Mann; Carol Gambrell; Daniel and Jana Gambrell; Josh and Lisa Gambrell; Wendy and Elliot Schuler; and Bill and Sarah Faulkner. To my family by marriage, thank you for always

making me feel at home and teaching me how to be an honorary Texan: Lisa French, Randy and Cheryl French, Doug and Kara French, Phyllis and Aaron Babin, and Tyler and Emma Kelly. And particularly to my parents Fred and Joni Lynn – thank you for your unwavering encouragement and love, and for raising me into the person I am today. I am proud to be your daughter!

A very special thank you to goes to my husband Will. His love, support, and encouragement have been instrumental in motivating me for the past five years. His enthusiasm for molecular simulations and other computational work helped inspire my interest in bioinformatics. I couldn't imagine the past five years without him!

Last, but certainly not least: this work was made possible by a predoctoral F31 grant from the National Institute of Dental and Craniofacial Research. I am grateful to the National Institutes of Health for accepting my grant application and for the incredibly useful professional and training opportunities their funding provided me. I am also indebted to Alissa's grant funding from the National Institutes of Health and the GI SPORE.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Ab	Antibody
AJCC	American Joint Committee on Cancer
AKT	Protein kinase b
AML	Acute myeloid leukemia
Ank	Ankyrin
APC	Adenomatous polyposis coli
Arp2/3	Actin related proteins 2 and 3
bFGF	Basic fibroblast growth factor
BRAF	Rapidly accelerated fibrosarcoma B
BSA	Bovine serum albumin
CCLE	Cancer cell line encyclopedia
CCND1	Cyclin D1 (gene)
CortBP1	Cortactin binding protein 1
CRC	Colorectal cancer
CTTN	Cortactin (gene)
Cys	Cysteine-rich
DAB	Diaminobenzidine
Dlg1	Drosophila disc large tumor suppressor
DMEM	Dulbecco's modified eagle medium
DPD	Dihydropyrimidine dehydrogenase
Dx	Diagnosis
ECM	Extracellular matrix
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor

ER	Estrogen receptor
ERCC1	Excision repair cross-complementation group 1
Erk	Extracellular signal-regulated kinase
EtOH	Ethanol
F-actin	Filamentous actin
FITC	Fluorescein isothiocyanate
FL	Full-length
FN	Fibronectin
FOXO3a	Transcription factor forkhead box O3a
GSEA	Gene set enrichment analysis
HBD	Heparin binding domain
HDR	Hypoparathyroidism, Sensorineural Deafness and Renal Disease
HGF	Hepatocyte growth factor
HNSCC	Head and neck squamous cell carcinoma
HR	Hazard ratio
IF	Immunofluorescence
IGF	Insulin-like growth factor
IGFBP2	IGF binding protein 2
IGFR	Insulin growth factor receptor
IHC	Immunohistochemistry
KD	Knockdown
kDa	Kilodaltons
KRAS	Kirsten rat sarcoma viral oncogene homolog
Lasso	Least absolute shrinkage and selection operator
LE	Late endosome

LOH	Loss of heterozygosity
LZRS	Retroviral vector used to create cell lines
mAb	Monoclonal antibody
MAPK	Mitogen activated protein kinase
MDM2	Mouse double minute homolog 2
MEF	Mouse embryonic fibroblasts
MEK	Mitogen activated protein kinase kinase
MeOH	Methanol
MET	Mesenchymal to epithelial
mGluR	Metabotropic glutamate receptor
mRNA	Messenger ribonucleic acid
MSI	Microsatellite instability
N	Number of replicates in an experiment
n.s.	Not significant
Na-H	Sodium-Hydrogen
NaPiIIa	Sodium-phosphate cotransporter IIa
NLS	Nuclear localization sequence
NMDA	N-methyl-D-aspartate
NRAS	Neuroblastoma rat sarcoma protein
NTA	N-terminal acidic
O/N	Overnight
OE	Overexpressed
Pak1	p21 activated kinase 1
PBS	Phosphate-buffered saline
PDZ	PSD95/Dlg1/Zo-1

PFA	Paraformaldehyde
PI3K	Phosphoinositide 3 kinase
PIGF	Phosphatidylinositol-glycan biosynthesis class F protein
PLC- β3	Phospholipase C beta 3
PPARα	Peroxisome proliferator activated receptor alpha
PPI	Polyproline I
PR	Progesterone receptor
PRD	Proline-rich domain
ProSAP1a	Proline-rich synapse-associated protein 1a
PSD95	Post synaptic density protein 95
PTEN	Phophatase and tensin homolog
Ras	Rat sarcoma protein
Rb	Retinoblastoma protein
RGD	Integrin binding domain (Arg-Gly-Asp)
RNAseq	Ribonucleic acid sequencing
RPMI	Roswell Park Memorial Institute medium
RPPA	Reverse phase protein array
RT	Room temperature
RTK	Receptor tyrosine kinase
SAM	Sterile alpha motif
SH3	Src homology 3 domain
SHANK2	SH3 and multiple ankyrin repeat domains protein 2 (CAPS: gene; lower-case: protein)
SMAD	Portmanteau of SMA (small) and MAD (mothers against decapentaplegic) genes

TA	Transactivation
TCGA	The cancer genome atlas
TGF-β	Transforming growth factor beta
TGN	Trans-Golgi network
TOPO1	Type 1 topoisomerase
TP53	Tumor protein 53
TS	Thymidylate synthase
UGT1A1	Uridine diphosphate glucuronosyltransferase 1A1
UICC	International Union for Cancer Control
UTR	Untranslated region
VEGF	Vascular endothelial growth factor
WASP	Wiskott–Aldrich Syndrome protein
WT	Wild-type
ZF	Zinc finger
Zo-1	Zonula occludens-1 protein

CHAPTER I: INTRODUCTION

INTRODUCTION TO CANCER

Cancer is the second leading cause of death in the United States, surpassed only by heart disease. While overall cancer deaths have declined 20% in the past 22 years, it still causes one out of every four deaths in the US [1]. However, much work remains to be done to elucidate underlying mechanisms of cancer aggressiveness in order to improve treatment of this disease.

Cancer arises from normal tissue in a multi-step process called tumor progression. It has been hypothesized that this progression occurs in an evolutionary fashion: cancer cells acquire and accumulate multiple genetic alterations, and advantageous alterations result in improved cell growth or survival, causing a clonal expansion of that cell population [2]. This process of genetic alteration followed by subsequent clonal expansion repeats multiple times, eventually producing an aggressive tumor. This evolutionary progression results in tumors that are unique in every patient, which would benefit from specific therapy targeted towards its distinct genetic alterations [3].

Although this progression causes varied individual tumors, cancer cells do acquire similar biological capabilities termed the “hallmarks of cancer” that regulate their transformation from normal cells and progression to malignancy [4]. The first hallmarks of cancer that were identified were sustaining proliferative signaling, evading growth suppressors, activating invasion and metastasis, enabling replicative immortality, inducing angiogenesis, and resisting cell death [4]. More recently identified hallmarks of cancer are deregulating cellular energetics, avoiding immune destruction genome instability and mutation, and tumor-promoting inflammation [5]. The latter two are

considered enabling characteristics for tumor cells [5]. The hallmark of activating invasion and metastasis is a defining feature of aggressive cancer; non-aggressive tumors that do not invade adjacent tissue or metastasize are benign.

CANCER AGGRESSIVENESS

Metastasis is the process by which cancer cells gain the ability to leave the primary tumor and grow at a distant location in the body. This occurs in a multistep cascade of discrete events. First, the cells must gain the ability to degrade the surrounding matrix and invade into the local stromal tissue [6]. Cells secrete factors that recruit blood vessels in a process termed called angiogenesis, which provides both nutrients to the growing tumor as well as a means of escaping the primary tumor site [7]. A cell must enter a blood or lymphatic vessel (intravasation), survive transit, and exit the vessel at a distant site (extravasation) [5]. Finally, the cell(s) must survive and proliferate at this distant site, creating a metastatic lesion [5]. Importantly, metastasis is the leading cause of death in cancer patients [8]; it is therefore crucial that we develop new ways to target this process.

Clinically, tumor progression and aggressiveness is most commonly classified using the American Joint Committee on Cancer (AJCC) or International Union for Cancer Control (UICC) tumor staging system [9]. This system takes into account primary tumor size (T), the number and location of regional cancer-positive lymph nodes (N), and whether or not the cancer has metastasized (M). This TNM designation is widely used for treatment decisions and survival predictions. However, it has long been recognized that the TNM staging system is often inadequate to predict patient prognosis [10]. One limitation is that TNM staging does not take into account additional biologic

factors, such as tumor molecular subtypes, tumor heterogeneity, microenvironmental factors, or genetic instability. A system that is able to better predict patient prognosis or treatment response would aid patient treatment decisions and have the potential to increase survival rates.

GENETIC ALTERATIONS IN CANCER

Normal cells have a system of checks and balances that regulate growth and other functions [2]. The genetic alterations that drive tumor progression and metastasis often deregulate these systems. For example, oncogenes encode proteins that promote tumor progression, while tumor suppressors are proteins that inhibit tumor progression [2]. Oncogenes can be activated at the gene level by gene duplication, induction of gene expression, or chromosomal translocations resulting in fusion proteins with increased oncogenic activity; additionally, a mutation can occur that results in an abnormally active form of the protein [11]. Conversely, tumor suppressors can be repressed by gene deletion or inactivation, or a mutation causing a defective protein. Most tumor suppressors require both alleles to be inactivated, since only one copy of the protein is typically necessary for its proper function; however, some tumor suppressors are haploinsufficient and inactivation of one copy is sufficient [12]. Oncogenes or tumor suppressors can also be activated at the protein level by epigenetics, miRNA, or other mechanisms [11–13].

Identification of oncogene or tumor suppressors involved in the processes of tumor progression or metastasis is crucial to further our understanding of this disease in order to better diagnose and treat patients. For example, new biomarkers can be identified that can be used for assessing a person's risk of developing cancer (BRCA1/2

in breast cancer [14]), identification of high-risk patients needing more aggressive treatment (S-100B in melanoma [15]), or stratification of already diagnosed patients into distinct subtypes that might receive a different therapy (HER2 positive breast cancer [16]). Discovery of proteins that drive tumor behavior opens up the possibility of treatment with small molecule inhibitors. Finally, identification of new cancer regulators leads to better understanding of the intrinsic tumor biology. An important area with the potential to aid in identification of these novel cancer regulators is bioinformatics.

BIOINFORMATICS AND CANCER

The term "bioinformatics" was first coined in 1970 as a working concept for analyzing information processing of living systems [17], but the field began to expand rapidly in the 1990s, recognized by the creation of the National Center for Biotechnology Information (NCBI) [18]. Today, bioinformatics is defined as a field that uses advanced mathematical, statistical, and computing techniques to solve biological problems. With the onset of high-throughput technologies enabling rapid and cheap generation of information, the rate-limiting step has increasingly become data analysis and interpretation of these data. Bioinformatics methods are used widely across multiple fields that produce this "big data"; specifically in the field of cancer research, bioinformatics analysis of data at multiple levels (DNA, RNA, protein) has been used extensively for personalized medicine, drug discovery, metabolomics, immunology, immunotherapy, gene discovery, epigenetics, gene function prediction, protein structure prediction, and molecular profiling [19–26].

High throughput data from the Cancer Genome Atlas (TCGA) and other publically available datasets are becoming widely available and are a rich resource for

data mining and biological discovery. A challenge for the field is to identify both biological drivers and strong prognostic markers of cancer. Gene expression datasets have been commonly used to classify tumors, due to their wide availability. However, additional types of high throughput datasets are now available and may provide a different starting point for molecular analysis of tumors. Protein expression datasets generated by mass spectrometry or reverse phase protein array (RPPA) are becoming widely available for many TCGA tumors. Since gene expression frequently does not correlate well with protein levels [27], such datasets may give additional insight into molecular mechanisms that drive tumor behaviors. In addition, phospho-protein levels may identify activation of specific signaling pathways.

A common approach to the analysis of tumor data is to first classify patients by molecular characteristics, such as KRAS mutation status or gene expression clusters, and then determine prognosis or treatment differences [28–30]. Alternatively, one can directly identify molecular differences that are statistically associated with patient outcome characteristics. We previously used the latter approach with RPPA data from head and neck squamous cell carcinoma to identify a PI 3-kinase-high, PKCα-low signaling state that drives invasive behavior [31]. Although it is limited by the availability of patient follow-up data, this type of bioinformatics approach is potentially powerful for identifying novel molecular drivers of tumor aggressiveness.

Brief Overview of Key Proteins Identified Affecting Cancer Aggressiveness

The following molecules are described more comprehensively within their specific chapters.

IGFBP2

Insulin-like growth factor binding protein 2 (IGFBP2) can function in an insulin growth factor (IGF) dependent or independent manner. In the IGF pathway, IGFBP2 binds to growth factors IGF1 and IGF2, controlling their half-life and availability and thus downstream signaling effects [32]. In addition, IGFBP2 can modulate angiogenesis, cell growth, and metabolic activity independently of IGF signaling by binding to other molecules, such as extracellular matrix (ECM) components, proteoglycan receptor, integrin receptors, and nuclear transport complexes [33]. In breast, prostate, gliomas, leukemia, and lung cancer, high levels of IGFBP2 are associated with metastasis, invasive cancer, poor prognosis, and promote growth, invasion, and chemoresistance [34–52]. However, in CRC, the role of IGFBP2 is more controversial: some studies show correlation with or promotion of tumor aggressiveness and poor prognosis, similar to its role in other cancer types [53–57], while a few other studies show that IGFBP2 inhibits tumor aggressive phenotypes such as cell proliferation and tumor size [58–61]. In this study, we find that high levels of IGFBP2 are associated with CRC poor prognosis in an initial cohort, and this is validated in a secondary TMA dataset that shows IGFBP2-high expressing patients have significantly decreased survival and recurrence-free survival.

GATA3

GATA3 (GATA-binding protein 3) has been most highly characterized in T-cells. GATA3 functions as a transcription factor that induces production of various cytokines, thereby promoting differentiation of naïve T-cells into a TH2 phenotype [62–68]. GATA3

also plays a role in differentiation in epithelial tissues, such as skin cells, hair follicles, kidney, and mammary luminal cells [69–72]. Some recent papers also highlight the importance of GATA3 in breast cancer. GATA3 is frequently mutated breast cancer, and these mutations are a defining feature of a subtype with improved prognosis (luminal subtype A) [73–75]. Interestingly, low levels of GATA3 are correlated with poor prognosis in cell lines and patients, such as metastatic potential, tumor progression, estrogen receptor (ER) and progesterone receptor (PR) negative status, Her2/Neu overexpression, high grade, and poor differentiation [76,77]. Functionally, GATA3 expression in breast cancer induces cell differentiation, regulates cell proliferation, suppresses metastasis, induces a mesenchymal-to-epithelial transition (MET), decreases invasion, and negatively regulates TGF-β signaling and invadopodia formation [76,78–82]. In addition to these breast cancer studies, one paper has also shown that GATA3 loss promotes invasion and predicts recurrence in prostate cancer [83]. However, prior to the work outlined in this dissertation, a role for GATA3 in CRC has never been reported. We show that low levels of GATA3 are associated with poor prognosis in CRC, and overexpression of GATA3 in CRC cell lines reduces cellular invasion and 3D colony growth.

Shank2

Shank2 is highly expressed in the brain, and plays a crucial role in the post-synaptic density by scaffolding together large protein signaling complexes [84]. Shank2 is necessary for proper synaptic functioning; mutations in Shank2 are associated with autism spectrum disorder [85–88]. Shank2 is also expressed in epithelial cells of other tissue types, and regulates salt and water transport and calcium signaling [89–92].

Shank2 binds to cortactin [93], a known driver of HNSCC aggressiveness (reviewed in [94]), but the effect of this binding interaction is unknown. Interestingly, their genes SHANK2 and CTTN are both located on the 11q13 amplicon, which is amplified in up to 40% of HNSCC cases [95–97]. However, the role of Shank2 in cancer of any type has not been investigated prior to the research outlined in this dissertation. We see that the cortactin SH3 domain responsible for Shank2 binding is crucial for tumor-associated phenotypes. However, since cortactin binds to multiple proteins through this domain it is unclear whether this effect is due to Shank2 binding. We find that Shank2 is expressed in HSNCC patients and cell lines, and is upregulated in tumor tissue compared to normal tissue. We also show that a knockdown of Shank2 in HNSCC cell lines results in decreased invasion, and preliminarily decreased invadopodia formation and MMP activity.

Brief Overview of Results

This dissertation will discuss two types of cancer, Colorectal Adenocarcinoma (CRC) and Head and Neck Squamous Cell Carcinoma (HNSCC), and the discovery of novel regulators in each using different methods. A bioinformatics analysis was used to identify which proteins were most associated with poor prognosis in CRC. We characterized IGFBP2, a known but understudied molecule in CRC progression, and showed that patients with high expression of IGFBP2 had significantly decreased survival and recurrence-free survival. In addition, we identified the transcription factor GATA3 as a novel regulator of CRC through functional studies with GATA3-OE cell lines. We show that overexpression of GATA3 in CRC cell lines decreases invasiveness and 3D colony growth, but not intrinsic 2D proliferation rates.

We also investigated the role of the cortactin SH3 domain and the adaptor protein Shank2, a cortactin SH3-binding partner, in HNSCC. A cortactin SH3 domain mutant was unable to rescue the cortactin dependent phenotypes Golgi morphology or *in vivo* tumor growth. Knockdown of Shank2 decreased invasiveness, with preliminary results suggesting a defect in invadopodia activity and MMP secretion. This shows that the cortactin SH3 domain, and its binding partner Shank2, regulates HNSCC tumor aggressiveness. Identification of these novel regulators of cancer gives insight into the mechanisms of tumor progression in CRC and HNSCC.

CHAPTER II: METHODS

ANTIBODIES AND REAGENTS

We used three GATA3 antibodies: catalog number 558686 from BD Biosciences (GATA3 BD), catalog number sc-265 from Santa Cruz (GATA3 SC), and catalog number LS-B4163 from LifeSpan Biosciences (GATA3 LS). IGFBP2 antibody was catalog number LS-C138280 from LifeSpan Biosciences and β−actin antibody was catalog number A2228 from Sigma Aldrich. Transwell invasion chambers were from Corning.

Shank2 rabbit polyclonal (8034, 8035) and mouse monoclonal (3B6, 4B3, 1D4, 2F5, 3F8) antibodies were a generous gift from Tom Parsons. Commercial Shank2 polyclonal antibody was from Santa Cruz Biotechnologies. Cortactin C-90 antibody has been described previously [98]. GM130 was from BD Labs, CD63 was from Abcam and Phalloidin and all secondary antibodies were from AlexaFluor.

Matched HNSCC tumor and normal samples were obtained from the Vanderbilt-Barry Baker Head and Neck Biospecimen Repository.

TCGA DATA ANALYSIS

CRC RPPA level 3 and clinical information (including survival, recurrence, and follow-up time) were downloaded from the TCGA data portal. Summary spreadsheets of patient clinical information and RPPA expression were created in Excel. All primary data analyses were performed in R 1.3.1 [99] (details on statistical analyses are in the following section).

HNSCC RNAseq V2 and clinical data were downloaded from the TCGA data portal. The analysis was done in Excel and graphs were made using GraphPad Prism. Gene_normalized RNAseq data was used for the analysis. The cutoff for high expression was 14,000 RPKM for CTTN and 1,000 for SHANK2.

BIOINFORMATICS STATISTICAL ANALYSES

A univariate Cox's proportional hazard's model analysis was performed for each protein (survival package in R) [100,101]. Patients with <30 days of follow-up information were excluded. The Wilma algorithm works in a greedy forward strategy and optimizes a combination of the Wilcoxon and Margin statistics for finding clusters of predictor variables (supclust package in R) [102]. Regsubsets (Leaps package) [103] is a model selection method that carries out an exhaustive search for the best subsets of independent variables that predict the dependent variable in linear regression. Nvmax (the maximum number of proteins that could be included in the model) was set to 5 and nbest (the number of models of each size) was set to 10. The RPPA data were median-centered and scaled to one standard deviation before performing analyses in order to make comparisons between different proteins. For the Wilma and Regsubsets analyses, patients were divided into good prognosis (living patients or patients with recurrence-free survival were only included if they were alive or recurrence-free for at least 3 years, indicated subsequently by ≥ 3 years follow- up) or poor prognosis (all patients with a recurrence or death were included regardless of follow-up time).

HEATMAPS

Heatmaps were created with unsupervised clustering of patients and proteins, using the package “heatmap.plus” in R 1.3.1 based on Euclidian distance and complete linkage [104].

SURVIVAL PLOTS

For each protein, patients were divided into high-expressing (at or above median RPPA expression) and low-expressing (below median RPPA expression). Using SPSS, multivariable cox proportional hazard model was used to estimate overall survival and recurrence-free survival, adjusting for patient stage, and Kaplan-Meier curves were generated to compare survival and recurrence-free survival between high-expressing and low-expressing groups.

CELL CULTURE

CRC cells were grown in previously published optimal media for each cell line (for DLD1 and KM12c, DMEM + 10% FBS and non-essential amino acids). DMEM was purchased from Corning, FBS was purchased from Denville Scientific, and non-essential amino acids were purchased from Sigma. To create GATA3-OE cells, DLD1 or KM12c cells were transduced with retrovirus created by transfecting Phoenix packaging cells with pBabePuro-GATA3 (plasmid 1286 from Addgene). Pooled transduced cells were selected by puromycin treatment and used for experiments. Empty vector pBabePuro was used as a control.

HNSCC cells were cultured in optimal growth media for each cell type: Cal-27 in DMEM + 10% FBS and 1% Gln; Detroit-562 in EMEM + 10% FBS; FaDu and SCC61 in DMEM + 20% FBS and 0.4 ug/mL HC; Hep3-GFP in DMEM + 10% FBS; HSC-2 in DMEM + 10% FBS and 1.25 ug/mL amphotericin B; and SCC25 in DMEM/F12 + 20% FBS and 0.4 ug/mL of HC.

To create Shank2-KD cells, shRNA was designed to the 3' untranslated region (UTR) of Shank2 using the Invitrogen website and cloned into pLenti-dest vector. To create Shank2-KD cells, Cal27 or FaDu cells were transduced with retrovirus created by transfecting Phoenix packaging cells with shRNA. Pooled transduced cells were selected by puromycin treatment and used for experiments. An shRNA to LacZ was used as a control. KD was quantitated in ImageJ and graphed in Excel.

TRANSWELL INVASION ASSAY

For CRC cells, 50,000 cells/well were plated in triplicate on matrigel-coated Transwell inserts in serum-free DMEM. Normal growth media was used on the bottom as a chemoattractant. Cells were allowed to invade for 48 hours and then fixed with a three-step stain (Thermo Scientific). Five random fields from each Transwell insert at 10x magnification were taken on an EVOS microscope for quantitation.

For HNSCC experiments, cells were serum starved for 24 hours prior to plating on invasion chambers. 35,000 cells/well were plated in duplicate on matrigel-coated Transwell inserts (Corning) in serum-free DMEM. Normal growth media was used on the bottom as a chemoattractant. Cells were allowed to invade for 48 hours; cells were removed from the top of the filter by scraping, and invaded cells were then fixed with a

three-step stain (Thermo Scientific). The total number of invaded cells was counted manually.

3D MATRIGEL CULTURE

Embedded three-dimensional culture was carried out as previously published [105]. Briefly, 35 mm glass-bottomed Mat-tek dishes (Mat-tek Corporation) were coated with 60 μL Matrigel (Corning). 4,000 cells were plated in each dish in 200 μL 90% Matrigel, 10% growth medium. 2 mL of growth media was added to each dish after 30 minutes and replaced every four days. Cells were imaged at 10x magnification every two days starting at day 3; eight random fields at various depths from each dish were imaged and the diameter of each in-focus colony was quantitated.

PROLIFERATION ASSAY

1500 cells/well were plated in triplicate in the presence or absence of 10% serum in 96 well plates and grown for five days. Each day the plates were imaged on a Cellavista automated microscope after the addition of Calcein to identify live cells, Propidium iodide to identify dead cells, and Hoechst to identify nuclei (all from Invitrogen). Data were quantitated with the Cellavista imaging software to determine the number of live cells for each day.

TMA CONSTRUCTION AND IRB INFORMATION

All use of human tissue samples was conducted under IRB-approved protocols. The colorectal cancer tissue microarray (TMA) was constructed with 99 cases of colorectal cancer, using duplicate 1-mm cores of each colorectal cancer in the GI SPORE Tissue Core facility (IRB # 020338). All samples in the TMA are from formalin-fixed paraffin-embedded blocks in the pathology archives, and are from tissue removed during the course of routine clinical care. Associated outcome and demographic data are extracted from the Colorectal Carcinoma Data and Virtual Archival Specimen Repository (IRB# 101531), and are stripped of all identifiers when released to investigators. The array is enriched for special histologic subtypes of CRC such as mucinous, signet ring cell, and medullary carcinoma, and contains the full spectrum of histologic grades and tumor stages. Twelve control cases of histologically normal colorectal mucosa from surgical resections for non-neoplastic disease such as diverticulosis coli are included.

TMA STAINING

Antigen retrieval was performed in pH 6.0 citrate buffer, by using a pressure cooker at 104 ^{o}C for 20 minutes with a 10 minute bench cool down, followed by quenching with 0.04% H_20_2 w/sodium azide for 5 minutes. After blocking in a serum-free protein block for 20 min, primary antibody was incubated with the samples for an hour, followed by detection with Dako Envision + HRP Labeled Polymer for 20 minutes followed by incubation with chromogen DAB+ for 5 minutes.

TMA Analysis

To be included in the survival or recurrence-free curves, patients needed to have the following information: stage, days until event (if deceased or recurrent), and a follow-up time of at least 30 days (if living or nonrecurrent). Through the Vanderbilt University Digital Histology Shared Resource in the Epithelial Biology Center, immunostained TMA slides were imaged at 20X magnification to a resolution of 0.5 μm/pixel with the Leica SCN400 Slide Scanner (Leica Biosystems). Tissue cores were analyzed with Ariol® Review software SL-50. Upper and lower thresholds for brown DAB positive staining were set for color, saturation, and intensity. Tumor areas with staining that registered between these thresholds were determined to be DAB-positive in an automated analysis. Brown (DAB-positive) area of each tumor core was thus used to determine cytokeratin (tumor area), IGFBP2, and GATA3 stained area. The percent of the tumor area positive for IGFBP2 was calculated by dividing the IGFBP2-positive area by the cytokeratin-positive area and multiplying by 100.

TGF-β Reporter Assay

Cells were transfected with both CAGA-Luciferase to measure TGF-β activity and Renilla as an internal control, both generous gifts from Natasha Deane. After 24 hours, cells were lysed, samples were prepared using the Dual-Glo Liciferase Assay System (Promega) and fluorescence was measured on a luminometer.

Numbers and Statistics

For comparison of good and poor prognosis TCGA patients, a Fisher's exact test was used to analyze categories with two variables (gender, M). A Chi-squared test was used to analyze categories with more than two variables (Stage, T, N). Age and gender were analyzed using a Student *t*-test. All analyses were performed in GraphPad. For experimental data from CRC cell lines, data from the engineered cell lines were plotted and statistically analyzed in GraphPad using a Student *t*-test. Data plotted in bar graphs were represented as mean+/-standard error. For growth curves, error bars represent 95% confidence intervals.

Microscopy and Image Analysis

For immunofluorescent imaging, cells were seeded (50,000) on coverslips. After 24 h, the cells were fixed in 4% paraformaldehyde, permeabilized with TritonX100, incubated with primary and secondary antibodies for 1 h each at room temperature. Hoechst was added on the last wash for 5 min, and the coverslips were then mounted using Aqua Poly/Mount (Polysciences).

Images for quantification of Golgi area were acquired with an Eclipse TE2000-E widefield fluorescent microscope (Nikon) equipped with a Plan Apo VC 60×, 1.4NA, oil immersion lens, and a cooled charge-coupled device camera (HAMAMATSU, C4742-80-12AG) by the use of Metamorph software. Golgi area was quantitated by thresholding the images based on GM130 and a goat anti-mouse Alexa Fluor 488 staining. Cell area was determined manually by tracing the cell perimeter in rhodamine–phalloidin stained images, and the ratio of Golgi to cell area was calculated based on the thresholded area

using Metamorph software. Graphs were generated using Prism Graphpad version 5. Statistical analysis was performed using SPSS version 18. Data with a non-normal distribution (e.g., Golgi area/cell area) were analyzed using a Mann–Whitney U test with Bonferroni correction. Data are presented as box and whiskers plots with the box indicating the 25th and 75th percentiles, solid line indicating the median, and the whiskers indicating the 95% confidence intervals.

For colocalization experiments, images were acquired on a confocal laser scanning microscope (model LSM 510; Carl Zeiss), using a Plan-APOCHROMAT 63×/1.4 oil DIC objective and either a Argon/2, HeNe1, and/or HeNe2 laser. Z-axis sectioning was at 0.2 µm intervals.

For the invadopodia analysis, images were acquired with an Eclipse TE2000-E widefield fluorescent microscope (Nikon) equipped with a Plan Apo VC 40×, 1.4NA, oil immersion lens, and a cooled charge-coupled device camera (HAMAMATSU, C4742-80-12AG) by the use of Metamorph software. Cell area was determined manually by tracing the cell perimeter in actin or cortactin stained images. This cell area was then transferred over to the FN stained image, which was thresholded to determine degradation area. Data were tested for normality using SPSS, and non-normal data were analyzed using a Mann–Whitney U test with Bonferroni correction. Graphs were generated in GraphPad Prism.

SEMI-ORTHOTOPIC HNSCC MOUSE MODEL SYSTEM

The mouse model system was developed by the laboratory of Dell Yarbrough [106]. Rat tracheas were obtained from rats sacrificed in the Vanderbilt animal facility and denuded by three rounds of freezing at -80°C and thawing at room temperature.

Tracheas were placed into human keratinocyte media containing 40 ug/mL tobramycin, 100 ug/Ml ceftazidime, and 1 ug/mL amphotericin, and media was changed after the first thaw. After the second freeze, tracheas were stored at -80°C until ready for use.

For stability, denuded tracheas were first clipped using sterile wound clips (Vitalitec) to plastic tubing. Sutures were then used to secure one end of the trachea slightly above the wound clip. A small incision was made at the other end of the trachea, $1X10^6$ HNSCC cells in 40 uL were inserted, and the incision was then sutured shut below the wound clip. Would clips were then removed, leaving only the sutures securing the trachea to the plastic tubing.

Nude mice were anesthetized and prepared tracheas were inserted into subcutaneous pockets along the flanks. Two tracheas were inserted into each mouse (one per flank). Tumors were grown for 8 weeks; mice were sacrificed and tracheas were removed.

After removal, sutures were cut and plastic tubing was removed from trachea/tumor. For processing, the tracheas were cut in half in the middle of the trachea (or the middle of the tumor, if the tumor was concentrated at one end of the trachea). Half was snap-frozen in liquid nitrogen and the other half was embedded in paraffin. For staining, cross-sections of the trachea were cut to visualize the tumor inside the tracheal ring. The Vanderbilt Translational Pathology Shared Resource Core performed staining and provided the cytokeratin antibody.

INVADOPODIA ASSAY

Fibronectin (BD Biosciences) was labeled with Fluorescein isothiocyanate (FITC) in borate buffer (170 mM $Na_2B_4O_7$ pH 9.3, 40 mM NaCl), then dialyzed extensively

against phosphate-buffered saline (PBS), followed by dialysis with 50% glycerol. FITC-FN was stored at -20°C until ready for use. Immediately before use, FITC-FN was diluted to 50 ug/mL and ultracentrifuged at 70,000 rpm for 15 minutes at 4°C to remove protein aggregates.

Poly-d-lysine covered Mat-tek dishes were then coated with a solution of 1% gelatin (Polysciences) and 1% sucrose, heated to 37°C to dissolve gelatin, for 1 minute at room temperature (RT). Gelatin was aspirated until only a thin layer remained, and was allowed to dry for 1 hour at RT. Gelatin was cross-linked by the addition of 0.5% glutaraldehyde in PBS for 15 minutes on ice, followed by 30 minutes at RT. Glutaraldehyde was deactivated by the addition of 1mg/mL of sodium borohydride for 3 minutes at RT. Ultracentrifuged FITC-FN (see above paragraph) was added for 1 hour at RT. Coated plates and lids were then sterilized with 70% ethanol (EtOH) in a sterile cell culture hood. Plates were washed 2X with 1X PBS and equilibrated in invadopodia media (1:1 DMEM:RPMI, 20% FBS, 10% Nu-serum) for 30 minutes. Immediately before plating cells, new invadopodia media with EGF to stimulate invadopodia formation was added. 50,000 cells were plated onto each coated Mat-tek dish and cultured for 24 or 48 hours at normal conditions or in a hypoxia chamber.

The next day, cells were fixed in 4% paraformaldehyde (PFA) for 15 minutes, permabilized in 0.5% Triton X-100 for 5 minutes, and blocked for 1 hour in 3% bovine serum albumin (BSA). Primary and secondary antibody incubations were each performed for 1 hour in 3% BSA. Coverslips were mounted on Mat-tek dishes using Aqua Poly/Mount (Polysciences).

ZYMOGRAPHY

FaDu cells were plated at 90% confluency on 60 mm Mat-tek dishes. After cells adhered, growth media was replaced with serum-free media. After 24 hours, conditioned media was harvested and concentrated 10X. Samples were run on 10% SDS substrate gels with 1% gelatin, 0.75 mm gels thick. Gels were washed with 2.5% Triton X-100 to remove SDS and incubated overnight (O/N) in substrate buffer (50 mM Tris-HCl, 10 mL $CaCl_2$, pH 7.6). The next day, gels were stained with 0.5% Coomassie Blue in 50% methanol (MeOH) and 10% acetic acid. Gels were destained in 50% MeOH and 10% acetic acid, incubated in ddH_2O to reverse gel shrinking, and imaged. Analysis of bands was done in ImageJ.

CHAPTER III: LINKING PATIENT OUTCOME TO HIGH THROUGHPUT EXPRESSION DATA IDENTIFIES NOVEL REGULATORS OF COLORECTAL ADENOCARCINOMA AGGRESSIVENESS

Parts of this chapter are published in F1000Research.

INTRODUCTION

Colorectal Adenocarcinoma

Colorectal carcinoma (CRC) is the third leading cause of cancer deaths for both men and women in the United States; each year 142,000 people are diagnosed and 50,000 people die from this disease [1]. Like most cancers, CRC progression is associated with acquisition of a series of genetic alterations, including APC mutation/loss, KRAS mutation, Smad2/4 loss, and TP53 mutation/loss (**Figure 1**). Importantly, survival rates dramatically decrease when distant metastases are present: survival for localized stage II CRC is 90%, while survival for metastatic stage IV CRC is only 12% [1]. Therefore, it is crucial for patient survival and quality of life that cancer is diagnosed as early as possible and treated appropriately.

CRC staging is based on the TNM tumor staging system, including an assessment of the primary tumor, regional lymph nodes, and distant metastases. The primary tumor (T) is classified from T0-T4 based on the amount of invasion into the surrounding tissue. Regional lymph nodes are classified from N0-N2 based on the number of regional lymph nodes containing metastases. Distant metastases are a binary classification of M0 or M1, indicating whether distant metastases are absent or present, respectively. If any of these categories cannot be assessed, they are assigned

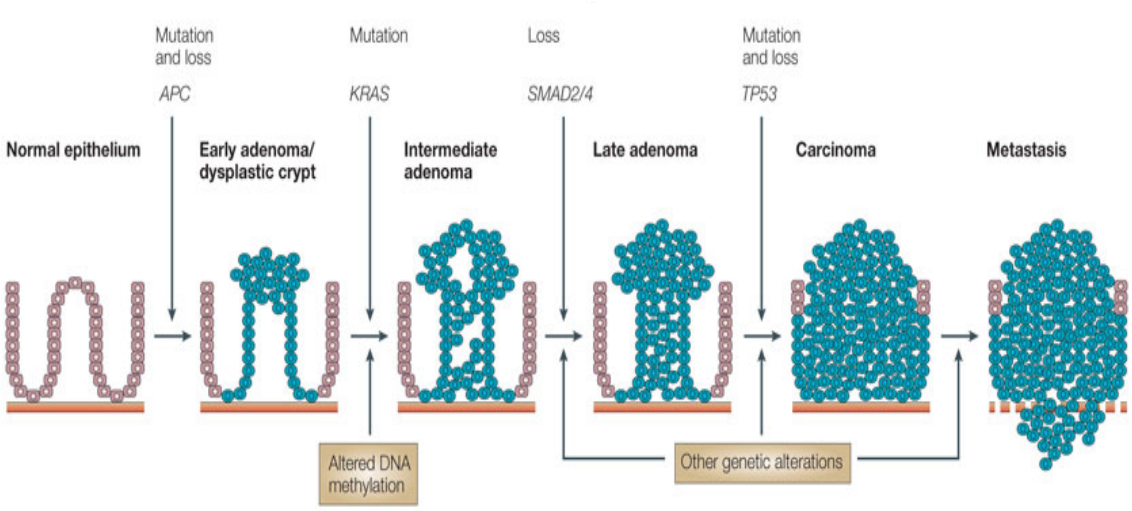


Figure 1: Genetic alterations associated with CRC progression
This depicts the progression of normal colonic tissue to a metastatic cancer. Common genetic alterations are shown in *italics.* Abbreviations: APC, adenomatous polyposis coli; KRAS, Kirsten rat sarcoma viral oncogene homolog; SMAD, small mothers against decapentaplegic; TP53, tumor protein 53. Figure is from [115].

a designation of TX, NX, or MX [107]. Once the TNM status of the tumor is assessed, the stage is determined (**Table 1**).

CRC stage is a central prognostic factor; therefore, current treatment decisions for patients depend on tumor stage to a large degree. While it is agreed that Stage III patients benefit from adjuvant chemotherapy [107], it remains controversial whether the less advanced Stage II patients should receive adjuvant treatment. It has been reported in several papers that Stage II CRC patients do not demonstrate a significant benefit from adjuvant treatment [108,109]. Conversely, other studies show either smaller benefit to Stage II patients [110,111] or a similar benefit in Stages II and III cancer regardless of other prognostic risk factors [112]. In general, adjuvant therapy is recommended for "high-risk" (clinically, the invasion of the cancer into the serosa) Stage II patients [107]. Additionally, there are both prognostic and predictive molecular/genetic factors identified in CRC: prognostic factors are thought to determine patient prognosis, while predictive factors are used to guide treatment decisions (**Table 2**). There are two established molecular poor prognosis factors: MSI status in early-stage CRC and BRAF mutation in advanced CRC [113]. In addition, there are several potential molecular poor prognosis factors identified. For early-stage CRC, these include KRAS mutation, thymidylate synthase (TS) positivity, 18q loss of heterozygosity (LOH), high p53 expression, and SMAD4 loss; for advanced CRC, MSI status, EGFR expression, and KRAS mutation [113]. While these prognostic genomic markers are promising, none are considered ready for use in the clinic. Treatment is instead determined based on clinical presentation and tumor biology (e.g. metastases, symptoms, and progression), patient-related factors, drug efficacy/toxicity, and drug availability, or the following predictive markers [113]. There are three molecular/genetic predictive factors that are recommended for use (outside of clinical trials) for advanced CRC: EGFR-mAb for treatment of patients with a KRAS mutation, irinotecan for treatment of patients with TA

Primary tumor (T)			
TX	Primary tumor cannot be assessed		
T0	No evidence of primary tumor		
Tis	Carcinoma in situ: intraepithelial or invasion of the lamina propria		
T1	Tumor invades submucosa		
T2	Tumor invades muscularis propria		
T3	Tumor invades through the muscularis propria into the subserosa, or into the nonperitonealized pericolic tissues		
T4	Tumor directly invades other organs or structures and/or perforates the visceral peritoneum		
Regional lymph nodes (N)			
NX	Regional nodes cannot be assessed		
N0	No regional lymph node metastases		
N1	Metastases in 1–3 regional lymph nodes		
N2	Metastases in ≥ 4 regional lymph nodes		
Distant metastases (M)			
MX	Presence of distant metastases cannot be assessed		
M0	No distant metastases		
M1	Distant metastases		
Stage grouping			
	T	***N***	***M***
Stage 0	Tis	N0	M0
Stage I	T1 or T2	N0	M0
Stage IIA	T3	N0	M0
Stage IIB	T4	N0	M0
Stage IIIA	T1 or T2	N1	M0
Stage IIIB	T3 or T4	N1	M0
Stage IIIC	Any T	N2	M0
Stage IV	Any T	Any N	M1

Table 1: TNM staging classification of CRC
The TNM staging system used by AJCC. The primary tumor, regional lymph nodes, and distant metastases are all given a number designation; higher numbers indicate more advanced disease. From these three designations, the cancer stage is determined. Adapted from [107]

		Potential	Established
Prognostic Factors	Early CRC	KRAS mutation TS positivity (>25% of cells) 18qLOH p53 (high) SMAD4 (any loss) Multi-gene signatures	MSI-H/dMMR
	Advanced CRC	MSI-H EGFR (IHC) KRAS mutation	BRAF mutation
Predictive Factors	Advanced CRC	Predictive for chemoradiation in localized rectal cancer • High TS • Low EGFR • TS polymorphism *3/*3 or *3/*4 (less benefit from CRT, than *2/*2, *2/*3, or *2/*4) For treatment with EGFR-mAB • BRAF mutation • Ligands: amphiregulin and epiregulin levels PI3K (exon 20 versus exon 9) mutation • PTEN mutation • NRAS mutation For treatment with bevacizumab • VEGF >98 pg/ml • bFGF, HGF, PlGF increase before progression under Bevacizumab (+chemotherapy) For chemotherapy toxicity or efficacy • High ERCC1 for oxaliplatin • High TOPO 1 for irinotecan ± oxaliplatin	For treatment with EGFR-mAB • KRAS mutation For chemotherapy toxicity • UGT1A1*28 genotype for irinotecan • DPD deficiency for fluoropyrimidines

Table 2: Molecular and genetic prognostic and predictive factors for CRC

Potential and established molecular markers are shown for early and advanced CRC. Abbreviations: KRAS, Kirsten rat sarcoma viral oncogene homolog; TS, Thymidylate synthase; LOH, Loss of heterozygosity; SMAD, small mothers against decapentaplegic; MSI-H, Microsatellite instability-high; EGFR, Epidermal growth factor receptor; IHC, Immunohistochemistry; BRAF, Rapidly accelerated fibrosarcoma B; mAb, Monoclonal antibody; PTEN, Phophatase and tensin homolog; NRAS, Neuroblastoma rat sarcoma protein; VEGF, Vascular endothelial growth factor; bFGF, Basic fibroblast growth factor; HGF, Hepatocyte growth factor; PlGF, Phosphatidylinositol-glycan biosynthesis class F protein; ERCC1, Excision repair cross-complementation group 1; TOPO1, Type 1 topoisomerase; UGT1A1, uridine diphosphate glucuronosyltransferase 1A1; DPD, Dihydropyrimidine dehydrogenase. Table from [113].

repeats in the TATA region of the uridine diphosphate glucuronosyltransferase 1A1 gene (UGT1A1*28 genotype), and fluoropyrimidines for treatment of patients with a dihydropyrimidine dehydrogenase (DPD) deficiency [113]. Additionally, there are several potential predictive markers that are not currently recommended for clinical use (**Table 2**).

Although overall survival rates are significantly higher for early-stage patients, those with undetectable micrometastases at the time of diagnosis are in danger of subsequent metastatic growth that worsens their outcome. In CRC, 30% of stage II patients have a recurrence at a distant site after removal of the primary tumor [114]. Identification of patients that have latent metastases would presumably improve patient prognosis by helping physicians to decide which patients would benefit from adjuvant chemotherapy. Appropriate treatment with chemotherapy of Stage II patients might prevent these latent metastases from later recurrence, thus improving survival rates.

IGFBP2

IGFBP2 is part of the insulin-like growth factor (IGF) signaling axis: ligands IGF1 and IGF2 activate insulin growth factor receptor (IGFR) by direct binding, which leads to downstream Ras/MAPK and phosphoinositide 3 kinase/protein kinase b (PI3K/AKT) signaling that result in decreased apoptosis, and increased proliferation, protein synthesis, and glucose metabolism (reviewed in [32]). IGFBP2 binds to IGF1 and IGF2 in the extracellular space, controlling their half-life and availability. Additionally, IGFBP2 can bind to molecules outside of the IGF pathway, such as extra-cellular matrix (ECM) components, proteoglycan receptor, integrin receptors, nuclear transportation complexes, and other factors, resulting in IGF-independent functions such as promotion of angiogenesis, cell growth, and metabolic activity [33]. IGFBP2 structural domains and binding sites are shown in **Figure 2**.

IGF signaling has long been recognized as important for driving cancer progression, and is deregulated in diverse cancer types, such as colon, prostate, osteosarcoma, melanoma, breast, liver, pancreatic, multiple myeloma, mesothelioma, glioblastoma, and childhood malignancies [116]; both expression level and functional studies have highlighted the importance of IGFBP2. In breast cancer, IGFBP2 promotes growth and invasion [34], and high expression levels are associated with lymph node metastases [35] and phosphatase and tensin homolog (PTEN) loss [36]. Overexpression of IGFBP2 is also associated with prostate cancer [37–40] and in high-grade or very invasive ovarian cancer [41,42], and promotes invasion of ovarian cancer cells [43]. In gliomas, IGFBP2 expression correlates with high-grade and poor prognosis patients [44] and stem cells [45]. Its overexpression in gliomas results in increased proliferation [45], higher grade tumors and increased AKT pathway activity [46],

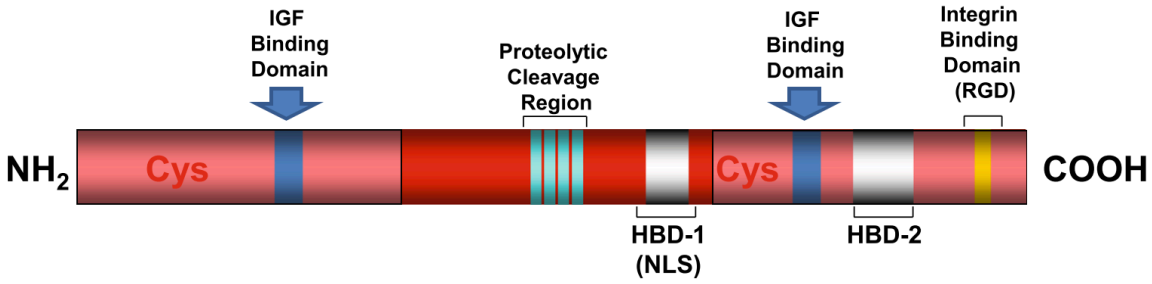


Figure 2: IGFBP2 domain structure
IGFBP2 has two cysteine-rich (Cys) regions at its N and C termini, separated by a linking domain containing a proteolytic cleavage region. Both the N and C terminal Cys regions contain IGF binding domains. There are two heparin-binding domains (HBD) located in the link domain and C-terminal Cys domain. The C-terminal Cys domain also contains an integrin-binding domain (RDG). Abbreviations: Cys, Cysteine; IGF, Insulin-like growth factor; HBD-1, heparin-binding domain; NLS, Nuclear localization sequence. Figure from [33].

increased invasion and MMP2 activity [47], while knockdown experiments show decreased invasiveness and tumorigenicity [48]. IGFBP2 is also associated with chemoresistance and therapy response in adult acute myeloid leukemia (AML) [49,50]. In lung cancer, high IGFBP2 autoantibody levels in the blood are used as a biomarker [51], and it inhibits caspase 3-mediated apoptosis [52].

While the studies described above indicate an oncogenic, tumor-promoting role of IGFBP2 in most cancer types, its role in CRC is more controversial. Some evidence points toward a similar tumor-promoting role: IGFBP2 is overexpressed in tumors [53–56] and elevated in blood serum levels [56,57] of human CRC patients. Overexpression of IGFBP2 resulted in increased proliferation, motility, tumorigenesis, and metastasis in a subcutaneous model, while IGFBP2 knockdown caused a reduction of these phenotypes [54]. Conversely, IGFBP2 has been shown to inhibit proliferation in vitro [58,59]. Additionally, in human tissue samples IGFBP2 mRNA was expressed in all samples but IGFBP2 protein was only expressed in a small subset [60]. This study also showed that extracts taken from cancer tissue (but not normal colonic tissue extracts) were able to degrade IGFBP2 protein, suggesting the presence of factors in cancer tissue has the ability to decrease IGFBP2 expression by protein degradation [60]. Furthermore, transgenic mice with an overexpression of IGFBP2 in the intestine show reduced tumor size and proliferating (Ki67-positive) cells [61]. IGFBP2 controls the availability of IGFs [33]; therefore, an extreme overabundance of IGFBP2 may sequester IGFs and prevent the tumor-promoting activities of its downstream signaling. It is also possible that IGFBP2 has opposing effects on tumor initiation and tumor progression; however, it is currently unknown what causes these contradictory effects and more research is needed to elucidate the mechanism. The work described in this dissertation supports the tumor-promoting role of IGFBP2 in CRC.

GATA3

GATA3 was originally identified as a transcription factor in T-cells [117,118]. GATA3 is expressed in naïve T-cells, and upon activation it is either upregulated or downregulated to produce at TH2 or TH1 phenotype, respectively [62]. GATA3 encourages a TH2 cell phenotype in two ways: by promoting differentiation of TH2 cells and suppressing differentiations of TH1 cells. GATA3 induces the differentiation of TH2 cells by production of the cytokines IL-4 and IL-5, and is both necessary and sufficient for TH2 cell differentiation [63,64]. Conversely, GATA3 suppresses TH1 cell differentiation by suppression of IFN-γ, Stat4, and RUNX3 production [65–68].

Although much of the literature focuses on the role of GATA3 in T-cells, it is also important for growth and development in non-hematopoietic tissues. GATA3-null mice are embryonically lethal and die due to hemorrhaging, anemia, central nervous system defects, and defects in liver hematopoiesis, suggesting that GATA3 is crucial for early development in multiple tissues [119]. In humans, haploinsufficiency of GATA3 causes Hypoparathyroidism, Sensorineural Deafness and Renal Disease (HDR) Syndrome, characterized by hypoparathyroidism, heart and renal defects, immune deficiency, and deafness, indicating that GATA3 plays an essential role in the development of the thyroid, heart, kidney, and immune system [120]. GATA3 is expressed in epithelial cells and regulates the differentiation of skin cells and hair follicles [69], kidney [70], and luminal cells in the mammary gland [71,72].

Recently, some work has also been published on the role of GATA3 in breast cancer. GATA3 levels inversely correlate with the metastatic potential of breast cancer cell lines, and GATA3 loss occurred with tumor progression in three different mouse models of breast cancer [76]. In human patients, low GATA3 expression is associated with a number of poor prognostic indicators, such as estrogen receptor (ER) and

progesterone receptor (PR) negative status, Her2/Neu overexpression, high grade, and poor differentiation (reviewed in [77]). Whether GATA3 has independent prognostic value is disputed; several studies report that it does [121–123], but this has been refuted by another study [124].

In addition to these GATA3 expression levels studies, GATA3 mutations are also commonly found in breast cancer [73–75]. Interestingly, the TCGA publication reports that GATA3 is the most frequently mutated protein in breast cancer [73]. These mutations are a defining feature of luminal subtype A of breast cancer that has an improved prognosis [75]. Most of these mutations are in the zinc finger domain of GATA3 that is necessary for its ability to bind DNA (**Figure 3**) [74].

Functional studies have shown that GATA3 expression induces cell differentiation and decreases metastasis [76,78]. GATA3 also regulates tumorigenesis and cell proliferation through its direct target INK4C [79]. Another downstream target of GATA3 is miR-29b, which promotes differentiation, alters the microenvironment, and suppresses metastasis [80]. It has also been reported that GATA3 induces a mesenchymal-to-epithelial (MET) transition that results in decreased invasion, smaller tumors, and decreased metastasis by binding to the E-cadherin promoter [81]. Finally, GATA3 has also been shown to negatively regulate TGF-β signaling and invadopodia formation through binding to Smad4 [82].

In addition to the breast cancer studies, one paper has reported GATA3 loss promotes cellular invasion and predicts recurrence in prostate cancer [83]; however, the role of GATA3 in CRC had not been reported prior to the work done in our laboratory.

In this study, we analyzed publicly available data from TCGA to identify proteins that are predictive of poor prognosis in colorectal adenocarcinoma (CRC) (7). We analyzed RPPA data, which includes protein and phospho-protein expression levels. Our analysis identified both known and novel candidate CRC drivers statistically

Figure 3: GATA3 domain structure

GATA3 consists of two N-terminal transactivation (TA) domains, and two zinc finger (ZF) domains that are each followed by conserved basic regions. Both ZF domains bind directly to DNA; ZF1 binds specifically to the canonical GATA motif (A/T)GATA(A/G), while ZF2 binding is broader. The basic regions mediate DNA binding. Abbreviations: TA, transactivation; ZF, zinc finger. Figure from [77].

associated with tumor recurrence or patient survival. Of these, we characterized two molecules in more detail. IGFBP2 was associated with both death and recurrence. Validation in an independent patient dataset by immunohistochemical (IHC) staining of a tissue microarray (TMA) demonstrated that high levels of IGFBP2 are associated with poor patient prognosis. Interestingly, low protein levels of the transcription factor GATA3 were highly associated with death of CRC patients in the TCGA data set. Experimental studies in colon cancer cell lines indicate that GATA3 expression acts to suppress invasive, aggressive CRC behavior. Since GATA3 protein and RNA levels are not correlated with each other, this association would not have been detected using RNA expression data.

RESULTS

Bioinformatics Analysis of TCGA Data

To identify molecular drivers of aggressive CRC behavior, we used statistical methods to link patient outcome data to protein and phospho-protein expression in the TCGA RPPA dataset. The RPPA dataset includes protein and phospho-protein levels from tumor biopsies taken at the time of diagnosis (for a complete list of proteins included in the analyses, see **Table 3**). The clinical information for these patients is also available, including recurrence and survival information, stage, and follow up time (**Table 4**, **Table 5**).

Therefore, we used a combination of univariate and multivariate approaches to identify proteins associated with recurrence or death. Univariate Cox proportional hazard regression analysis [100,101] relates the time to an event to a covariate (gene or protein expression) and is a common method to identify associations of protein expression with patient outcome. We also used Wilma and Regsubsets multivariate algorithms to select groups of proteins with predictive power [103,125]. Wilma is a greedy forward algorithm, meaning it selects the best solution at each step; conversely, Regsubsets is an exhaustive algorithm. The use of all 3 methods allowed us to identify whether certain proteins were chosen independent of the statistical method used. Patient characteristics are shown in **Table 5** for the Cox regression analysis and in **Table 4** for the Wilma/Regsubsets analyses. As expected, there were some significant differences in good vs. poor prognosis groups in tumor stage, metastasis status for patients included in the multivariate analysis (**Table 4**), and tumor stage, primary tumor size T, lymph node status N, and metastasis status M for patients in included in the Cox

14-3-3_epsilon-M-C	CD49b-M-V	IGFBP2-R-V	PKC-delta_pS664-R-V
4E-BP1_pS65-R-V	CDK1-R-V	INPP4B-G-C	PR-R-V
4E-BP1_pT37-R-V	Chk1_pS345-R-C	IRS1-R-V	PRAS40_pT246-R-V
4E-BP1_pT70-R-C	Chk1-R-C	JNK_pT183_Y185-R-V	PTCH-R-C
4E-BP1-R-V	Chk2_pT68-R-C	JNK2-R-C	PTEN-R-V
53BP1-R-C	Chk2-M-C	K-Ras-M-C	Rab11-R-V
ACC_pS79-R-V	cIAP-R-V	Ku80-R-C	Rab25-R-C
ACC1-R-C	Claudin-7-R-V	Lck-R-V	Rad50-M-C
AIB1-M-V	Collagen_VI-R-V	LKB1-M-NA	Rad51-M-C
Akt_pS473-R-V	COX-2-R-C	MAPK_pT202_Y204-R-V	Rb_pS807_S811-R-V
Akt_pT308-R-V	Cyclin_B1-R-V	MEK1_pS217_S221-R-V	Rb-M-V
Akt-R-V	Cyclin_D1-R-V	MEK1-R-V	S6_pS235_S236-R-V
alpha-Catenin-M-V	Cyclin_E1-M-V	MIG-6-M-V	S6_pS240_S244-R-V
AMPK_alpha-R-C	Cyclin_E2-R-C	Mre11-R-C	S6-R-NA
AMPK_pT172-R-V	DJ-1-R-C	MSH2-M-C	SETD2-R-NA
Annexin_I-R-V	Dvl3-R-V	MSH6-R-C	Shc_pY317-R-NA
AR-R-V	E-Cadherin-R-V	mTOR_pS2448-R-C	Smac-M-V
ARID1A-M-V	eEF2-R-V	mTOR-R-V	Smad1-R-V
ATM-R-C	eEF2K-R-V	N-Cadherin-R-V	Smad3-R-V
B-Raf-M-NA	EGFR_pY1068-R-V	NF-kB-p65_pS536-R-C	Smad4-M-V
Bak-R-C	EGFR_pY1173-R-C	NF2-R-C	Snail-M-C
Bax-R-V	EGFR_pY992-R-V	Notch1-R-V	Src_pY416-R-C
Bcl-2-R-NA	EGFR-R-C	Notch3-R-C	Src_pY527-R-V
Bcl-X-R-C	eIF4E-R-V	P-Cadherin-R-C	Src-M-V
Bcl-xL-R-V	ER-alpha_pS118-R-V	p21-R-C	STAT3_pY705-R-V
Beclin-G-V	ER-alpha-R-V	p27_pT157-R-C	STAT5-alpha-R-V
beta-Catenin-R-V	ERCC1-M-C	p27_pT198-R-V	Stathmin-R-V
Bid-R-C	ERK2-R-NA	p27-R-V	Syk-M-V
Bim-R-V	FAK-R-C	p38_MAPK-R-C	Tau-M-C
c-Jun_pS73-R-C	Fibronectin-R-C	p38_pT180_Y182-R-V	TAZ_pS89-R-C
c-Kit-R-V	FOXO3a_pS318_S321-R-C	p53-R-V	TAZ-R-C
c-Met_pY1235-R-C	FOXO3a-R-C	p70S6K_pT389-R-V	Transglutaminase-M-V
c-Met-M-C	GAB2-R-V	p70S6K-R-V	Tuberin-R-C
c-Myc-R-C	GATA3-M-V	p90RSK_pT359_S363-R-C	VASP-R-C
C-Raf_pS338-R-C	GSK3_pS9-R-V	PARP_cleaved-M-C	VEGFR2-R-C
C-Raf-R-V	GSK3-alpha-	Paxillin-R-V	XBP1-G-C
Caspase-3_active-	GSK3-alpha-beta-M-V	PCNA-M-V	XIAP-R-C
Caspase-	HER2_pY1248-R-V	PDK1_pS241-R-V	XRCC1-R-C
Caspase-8-M-C	HER2-M-V	PEA-15-R-V	YAP_pS127-R-C
Caspase-	HER3_pY1298-R-C	PI3K-p110-alpha-R-C	YAP-R-V
Caveolin-1-R-V	HER3-R-V	PI3K-p85-R-V	YB-1_pS102-R-V
CD20-R-C	HSP70-R-C	PKC-alpha_pS657-R-V	YB-1-R-V
CD31-M-V	IGF-1R-beta-R-C	PKC-alpha-M-V	

Table 3: List of all proteins included in RPPA analysis

A complete list of the 171 proteins that were included in the TCGA RPPA analysis, listed in alphabetical order.

Patients included in Wilma and Regsubsets analyses						
	Recurrence			***Death***		
	Recurrent	**Non-recurrent (3 yr. follow-up)**	**p-value**	**Deceased**	**Living (3 yr. follow-up)**	**p-value**
Total number	22	12		23	20	
Average age	66.86	59.83	0.2339	73.96	63.65	0.0274*
Average weight	77.55	79.75	0.6605	67.21	77.79	0.1732
Male	13	7	1	14	10	0.5472
Female	9	5		9	10	
Stage I	0	2	0.0341*	3	3	0.2379
Stage II	6	2		6	5	
Stage III	6	7		4	8	
Stage IV	9	1		9	3	
T0	0	0	0.0834	1	0	0.1315
T1	0	1		0	1	
T2	0	2		4	3	
T3	17	9		11	15	
T4	5	0		7	1	
N0	7	4	0.3392	10	8	0.9904
N1	8	7		8	7	
N2	6	1		5	4	
M0	7	10	0.0161*	10	13	0.1516
M1	10	1		9	3	

Table 4: Characteristics and comparison of good vs. poor prognosis of TCGA patients with RPPA data included in Wilma and Regsubsets analyses for death and recurrence

The features of patients included in the TCGA Wilma and Regsubsets analyses for both definitions of poor prognosis (recurrence and death). For good prognosis, a minimum of 3 years of follow-up data was required to be reasonably sure an event would not occur in the future. Age and weight are averages, while other categories have the number of patients present in each (gender, stage, T, N, and M). To see if there were any differences between good vs. poor prognosis for any of these categories, we used a student's T-test (age, weight), a Fisher's exact test (gender, M), and a Chi-square test (stage, T, N). P-values are reported significance is denoted by asterisks.

Patients included in Cox regression analysis						
	Recurrence			***Death***		
	Recurrent	**Non-recurrent (3 yr. follow-up)**	**p-value**	**Deceased**	**Living (3 yr. follow-up)**	**p-value**
Total number	22	125		23	168	
Average age	66.86	63.53	0.3418	73.96	65.21	0.003*
Average weight	77.55	82.84	0.2136	67.21	82.62	0.0459*
Male	13	63	0.4951	14	86	0.5052
Female	9	62		9	82	
Stage I	0	22	0.0009*	3	27	0.0016*
Stage II	6	46		6	67	
Stage III	6	41		4	53	
Stage IV	9	13		9	17	
T0	0	0	0.1259	1	0	0.0037*
T1	0	3		0	3	
T2	0	22		4	27	
T3	17	87		11	120	
T4	5	10		7	15	
Tis	0	1		0	1	
N0	7	75	0.0137*	10	102	0.2263
N1	8	39		8	45	
N2	6	11		5	20	
M0	7	99	< 0.0001*	10	130	0.0006*
M1	10	13		9	18	

Table 5: Characteristics and comparison of good vs. poor prognosis of TCGA patients with RPPA data included in Cox regression analysis for death and recurrence

The features of patients included in the TCGA cox regression analysis for both definitions of poor prognosis (recurrence and death). For good prognosis, patients with less than 30 days of follow-up time were discarded. Age and weight are averages, while other categories have the number of patients present in each (gender, stage, T, N, and M). To see if there were any differences between good vs. poor prognosis for any of these categories, we used a student's T-test (age, weight), a Fisher's exact test (gender, M), and a Chi-square test (stage, T, N). P-values are reported significance is denoted by asterisks.

regression analysis (**Table 5**). Age was also significantly different in deceased vs. living patients (but not recurrent vs. non-recurrent patients) for both analysis groups, which could be a possible confounding factor.

The Wilma and Regsubsets algorithms compare groups (clusters) of patients, which we predefined by patient prognosis (recurrence or death), and find proteins that are able to predict these clusters. For these multivariate methods, patients were divided into “good” or “poor” prognosis groups according to survival or recurrence data. “Good prognosis” patients were classified either as living or as having no recurrence with a minimum of 3 years follow-up time. We chose 3 years as a reasonable cut-off time since the great majority of colon cancer cases (91%) have a recurrence within this time frame [114]. Although this did reduce our sample size for patients included in the multivariate analyses compared to the univariate Cox regression (**Table 4** vs. **Table 5**), we felt it was necessary to ensure that our “good prognosis” group was accurate. For the “poor prognosis” patient group, recurrence or death could occur at any time point. To determine whether any proteins had stage-specific statistical associations, we performed the analyses using patient groups of stages I-II, stages I-III, or stages I-IV ("all stages"). However, we did not use stage, node or metastasis status as traits for identification of molecular correlates for several reasons. First, we reasoned that identifying molecular correlates of stage would not add prognostic information for clinical decision making, since stage is already gathered on every patient. Second, an initial test using the Wilma algorithm suggested that RPPA protein expression changes selected to be associated with node and metastasis negativity (e.g. N0M0 vs. N+M+) did not segregate patients well into groups (**Figure 4**). Thus, the two-dimensional projections indicate that proteins selected by both recurrence and death had the ability to separate patients into distinct groups, indicating good predictive power, while N/M status at the time of diagnosis did

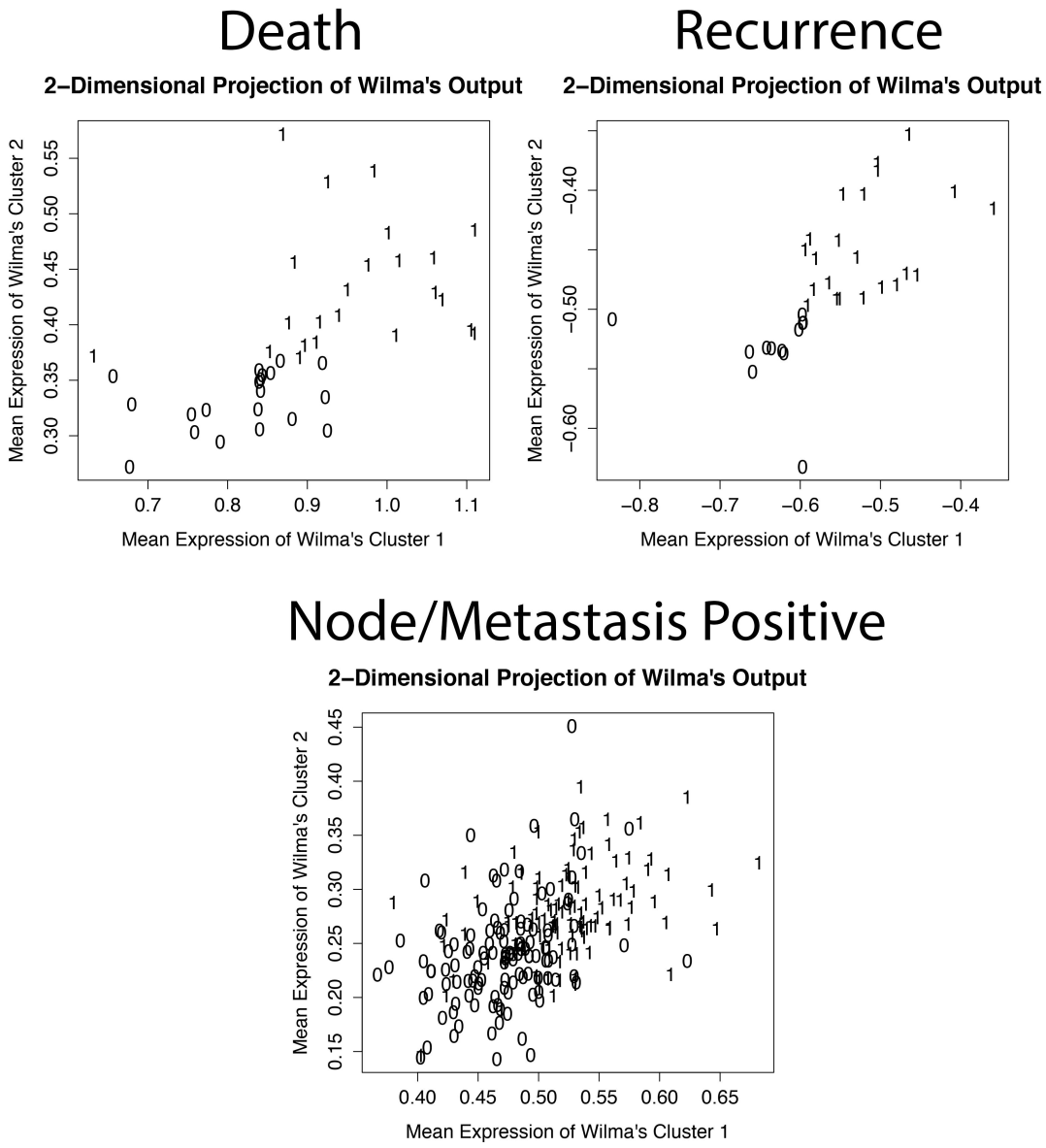


Figure 4: Comparison of Wilma 2D projections for death, recurrence, and node/metastasis status

These plots show the mean RPPA expression for all proteins (plotted in two dimensions) identified by the Wilma algorithm for patients of all stages. Each number shows the position of an individual patient in the two dimensional projection; a 0 indicates good prognosis (living with 3 years of follow-up data, non-recurrent with 3 years of follow-up data, or N0M0) while a 1 indicates poor prognosis (deceased, recurrent, or N+M+). The good separation between good and poor prognosis patients in the death and recurrence plots indicates these definitions are superior to node/metastasis positivity in identifying proteins associated with poor prognosis.

not (**Figure 4**). After this initial test, we did not perform any more analyses with node or metastasis status.

Proteins (indicated by "Antibody") significantly ($p<0.05$) associated with death using the univariate Cox proportional hazard regression are shown in **Table 6**, while proteins significantly ($p<0.05$) associated with recurrence are shown in **Table 7**. Hazard ratios indicate whether increased (HR > 1) or decreased (HR < 1) protein expression is associated with an event (death or recurrence). For example, if the hazard ratio is 3, the patient is 3 times more likely to have an event when expression of that protein is increased. All proteins that had a significant p-value were included, regardless of their hazard ratio.

Two-dimensional plots showing the results of the Wilma algorithm for all stages, stages I-II, and stages I-III are shown in **Figure 5** (death-associated proteins) and **Figure 6** (recurrence-associated proteins). Each number represents an individual patient; a 1 indicates a poor prognosis patient and a 0 indicates a good prognosis patient. The distinct separation between groups of patients indicates the algorithm is clustering poor prognosis patients separately from good prognosis patients. Proteins selected for both death and recurrence had the ability to segregate patients based on prognosis in stages I-II, stages I-III, and all patients, indicating that these are both good definitions of prognosis (**Figure 5**, **Figure 6**).

Plots showing the results of the Regsubsets algorithm for stages I-II, stages I-III, and all stages are in **Figure 7** (death) and **Figure 8** (recurrence). We set nvmax, the maximum number of proteins that could be included in a model, to 5. Nbest, the number of models of each size to determine, was set at 10. The Regsubsets algorithm returns nbest number of models of all sizes up to nvmax. Thus, each Regsubsets analysis returned 50 models (10 models times 5 model size groups). The adjusted r-squared value for each model is the plotted on the y-axis, while the proteins included in each

Death			
	Antibody	**Hazard_Ratio**	**P_value**
All stages	XBP1.G.C	7.12	0.003
	AMPK_alpha.R.C	4.82	0.011
	Tau.M.C	0.19	0.015
	Smad3.R.V	3.48	0.019
	Bid.R.C	0.06	0.028
	INPP4B.G.C	1.96	0.029
	IGFBP2.R.V	1.44	0.030
	GATA3.M.V	0.24	0.038
	Bim.R.V	2.67	0.044
	Fibronectin.R.C	0.58	0.046
Stages I-II	GATA3.M.V	0.04	0.0003
	K.Ras.M.C	0.14	0.002
	cIAP.R.V	15.84	0.005
	IGFBP2.R.V	2.25	0.006
	Rb.M.V	0.02	0.008
	Cyclin_D1.R.V	0.00	0.010
	AMPK_alpha.R.C	24.32	0.012
	X14.3.3_epsilon.M.C	0.00	0.017
	Tau.M.C	0.05	0.025
	Notch3.R.C	3.33	0.036
	PR.R.V	0.02	0.047
Stages I-III	GATA3.M.V	0.07	0.001
	Rb.M.V	0.06	0.004
	Bid.R.C	0.01	0.006
	IGFBP2.R.V	1.83	0.007
	K.Ras.M.C	0.23	0.008
	Tau.M.C	0.07	0.009
	AMPK_alpha.R.C	5.91	0.015
	c.Kit.R.V	2.25	0.027
	TAZ_pS89.R.C	0.00	0.031
	PR.R.V	0.06	0.046
	XBP1.G.C	6.28	0.046

Table 6: Hazard Ratios and p-values for Cox Regression Analysis Identifying Proteins Associated with Patient Death

A hazard ratio above 1 indicates that increased expression of the protein is associated with death; a hazard ratio above 1 indicates that decreased expression of the protein is associated with death. Only proteins with a significant p-value (<0.05) were included

Recurrence			
	Antibody	**Hazard_Ratio**	**P_value**
All stages	COX.2.R.C	2.02	0.001
	YAP_pS127.R.C	2.31	0.021
	Caspase.7_cleavedD198.R.C	0.63	0.028
	c.Jun_pS73.R.C	4.18	0.042
	MEK1_pS217_S221.R.V	2.42	0.049
Stages I-II	XIAP.R.C	41.60	0.012
	HSP70.R.C	2.44	0.031
	YAP_pS127.R.C	8.12	0.035
	IGFBP2.R.V	2.01	0.040
	COX.2.R.C	2.68	0.045
	PEA.15.R.V	52.19	0.046
	CDK1.R.V	0.00	0.050
Stages I-III	IGFBP2.R.V	1.96	0.004
	COX.2.R.C	2.30	0.007
	YB.1_pS102.R.V	7.69	0.019
	XRCC1.R.C	16.08	0.028
	XIAP.R.C	12.56	0.030
	DJ.1.R.C	0.07	0.043

Table 7: Hazard Ratios and p-values for Cox Regression Analysis Identifying Proteins Associated with Tumor Recurrence

A hazard ratio above 1 indicates that increased expression of the protein is associated with recurrence; a hazard ratio above 1 indicates that decreased expression of the protein is associated with recurrence. Only proteins with a significant p-value (<0.05) were included.

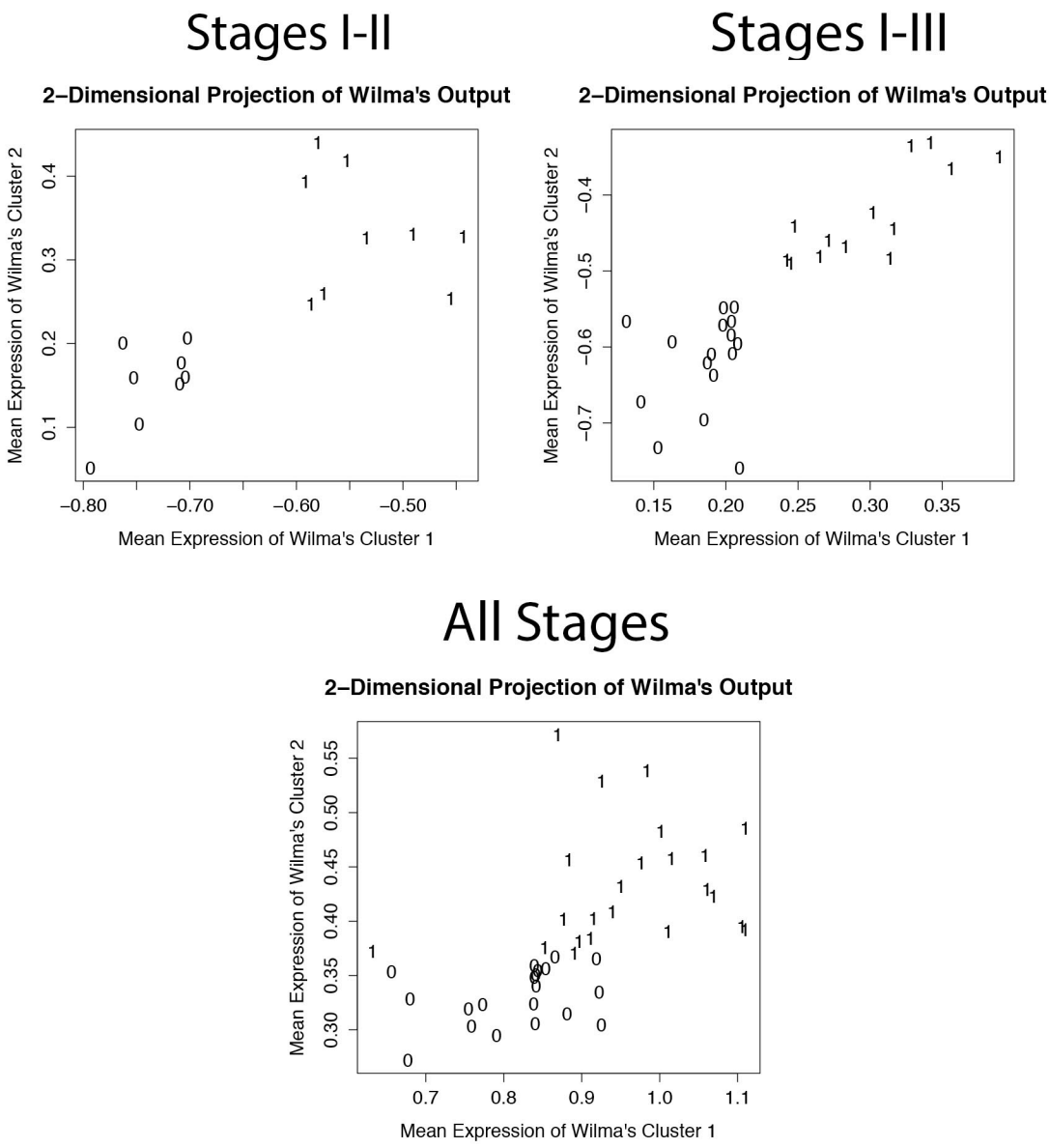


Figure 5: Two-Dimensional Wilma Algorithm Plots for Death
These plots show the mean RPPA expression for all proteins (plotted in two dimensions) identified by the Wilma algorithm for patients of all stages. Each number shows the position of an individual patient in the two dimensional projection; a 0 indicates good prognosis (living with 3 years of follow-up data) while a 1 indicates poor prognosis (deceased).

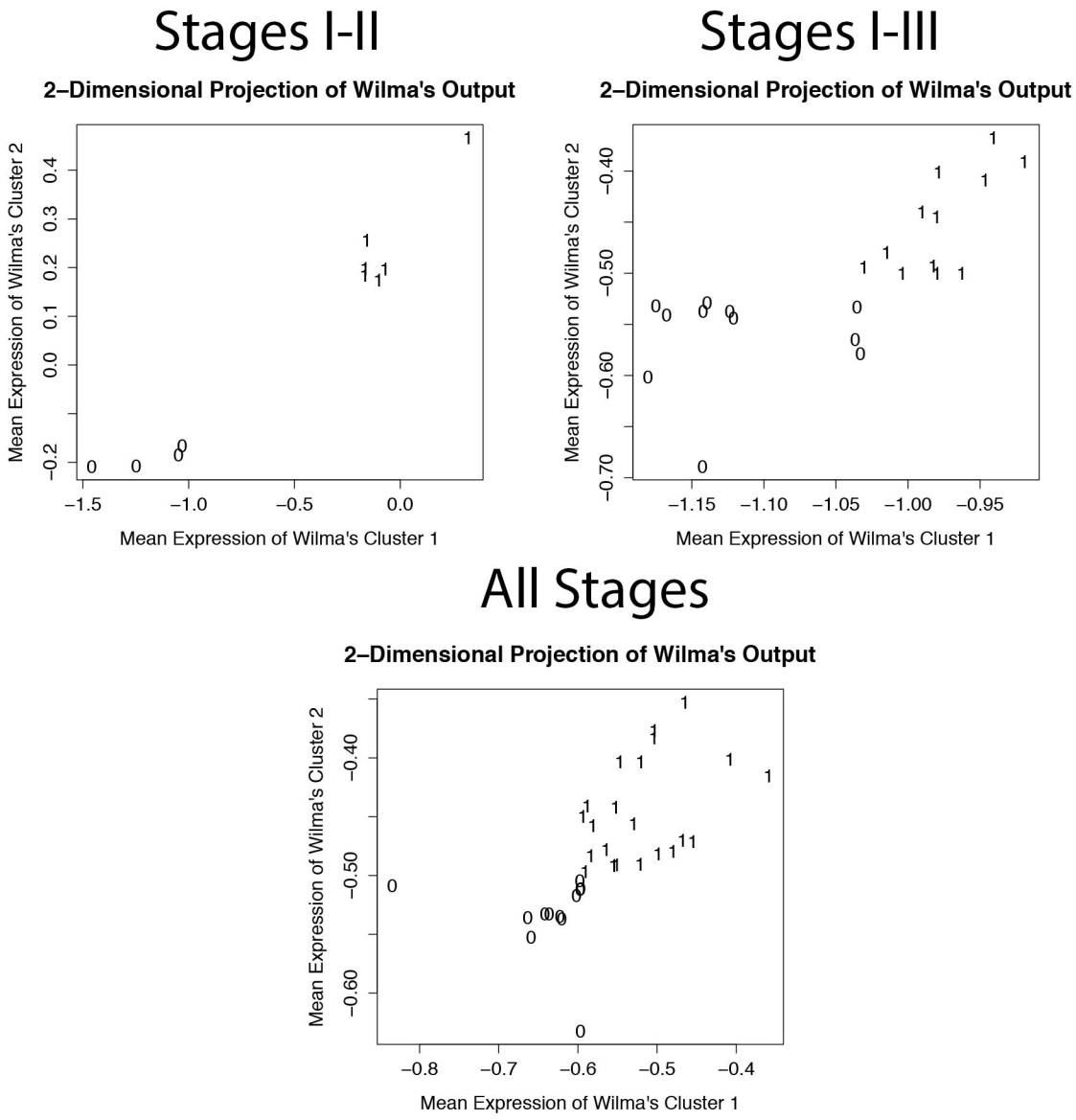


Figure 6: Two-Dimensional Wilma Algorithm Plots for Recurrence
These plots show the mean RPPA expression for all proteins (plotted in two dimensions) identified by the Wilma algorithm for patients of all stages. Each number shows the position of an individual patient in the two dimensional projection; a 0 indicates good prognosis (non-recurrent with 3 years of follow-up data) while a 1 indicates poor prognosis (recurrent).

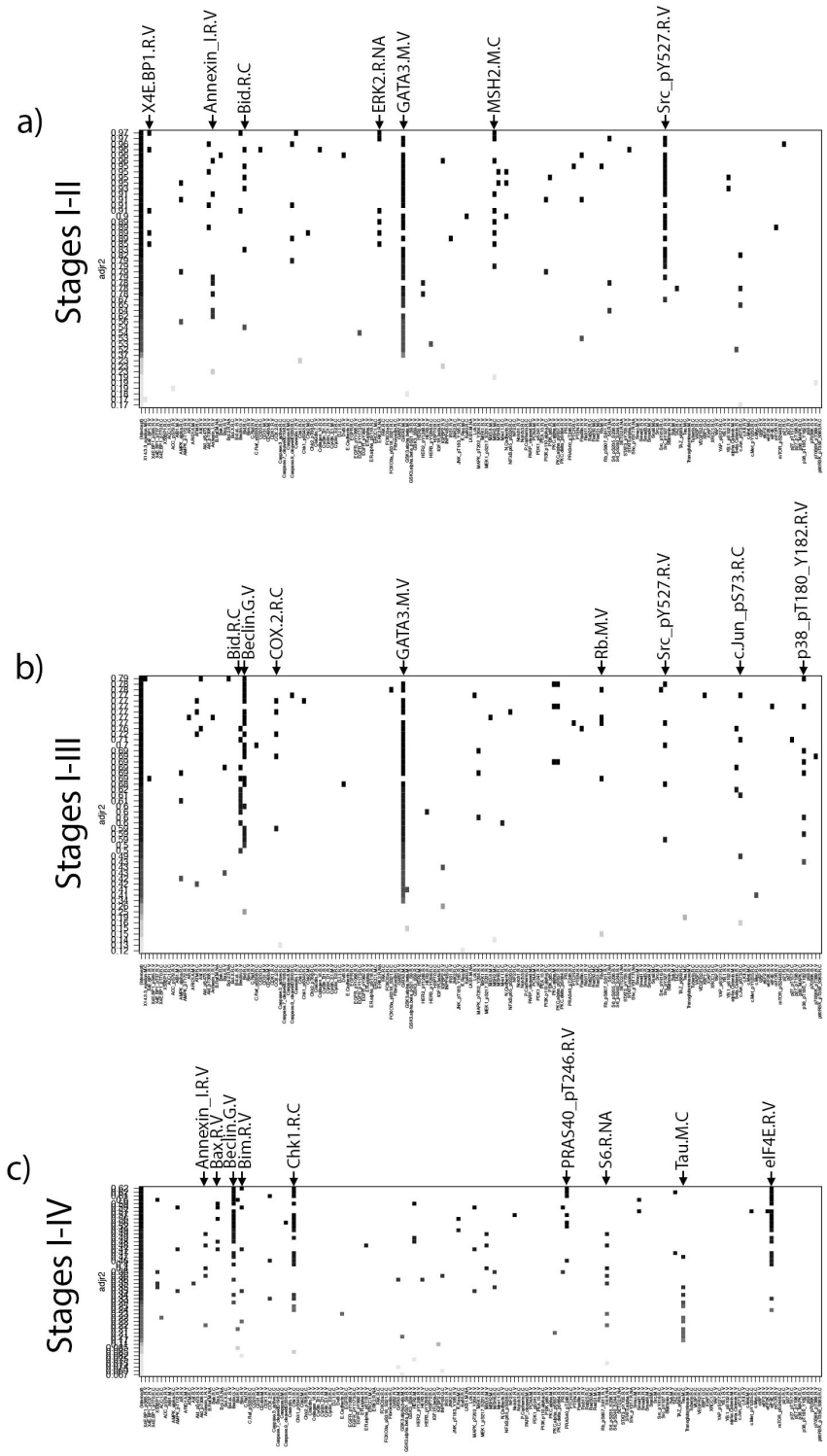


Figure 7: Regsubsets Algorithm Plots for Death

These plots show the output of the Regsubsets algorithm for **a)** Stages I-II, **b)** Stages I-III, and **c)** All stages. Proteins are on the x-axis, and adjusted r-squared value is on the y-axis. Each horizontal line indicates a separate model, and the black boxes show which proteins are included in each model. Proteins identified in more than 5 models are marked with an arrow and enlarged protein name above the plot.

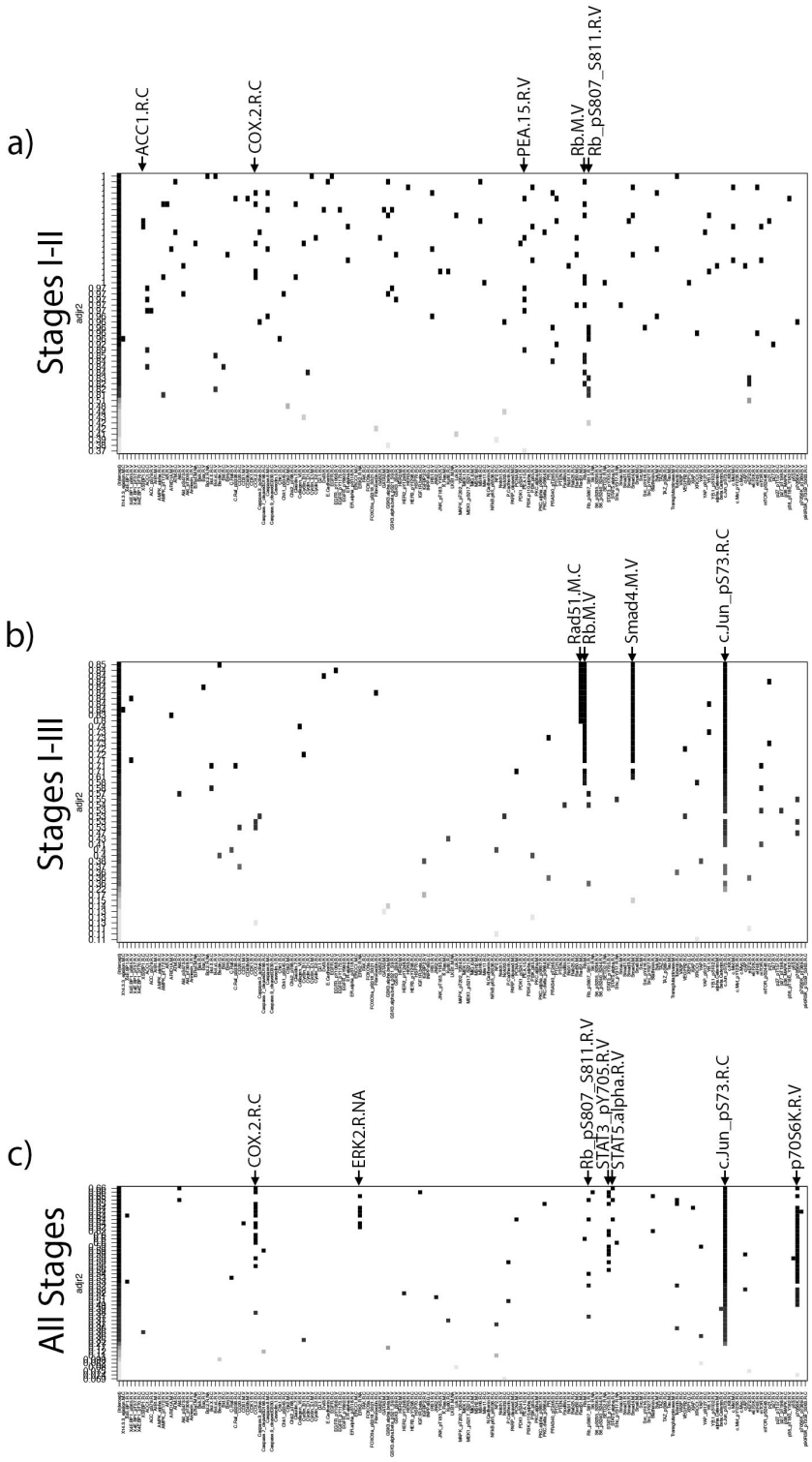


Figure 8: Regsubsets Algorithm Plots for Recurrence

These plots show the output of the Regsubsets algorithm for **a)** Stages I-II, **b)** Stages I-III, and **c)** All stages. Proteins are on the x-axis, and adjusted r-squared value is on the y-axis. Each horizontal line indicates a separate model, and the black boxes show which proteins are included in each model. Proteins identified in more than 5 models are marked with an arrow and enlarged protein name above the plot.

model are indicated on the x-axis. As the number of proteins in the model increases, the adjusted r-squared value also increases, so the models with nvmax proteins (in our case, 5) are the top nbest (in our case, 10) rows of the Regsubsets plots. For our analysis, we set a threshold that a protein must be identified five times by Regsubsets to be included in the summary charts (indicated by enlarged protein names inserted above arrows in **Figure 7** and **Figure 8**).

The full results of all three bioinformatics analyses for molecules statistically associated with death or recurrence are summarized in **Table 8** and **Table 9**, with bold proteins indicating the proteins that were identified by more than one of the three bioinformatics methods. To more easily visualize this, we created modified volcano plots of these proteins showing the number of times a protein was identified vs. the difference in RPPA expression for either death or recurrence (**Figure 9a**). In these charts, proteins with negative values are downregulated in patients with poor outcome (such as the well-known tumor suppressor, Rb) and proteins with positive values are upregulated (such as the oncogene c-Jun). Proteins that were identified by more than one method are indicated in red in the volcano plots (**Figure 9a**).

Proteins associated with death included known CRC drivers, including SMAD3, SMAD4, and MSH2, which respectively regulate TGF-β signaling [126] and microsatellite instability [127] (**Table 8**). In addition, a number of apoptosis and cell cycle proteins were associated with death, including Bid, Bim, Rb, and Chk1. Interestingly, the transcription factor GATA3 was our top hit associated with patient death and was identified 8 times out of a potential maximum of 9 times (3 stage groups analyzed by 3 statistical methods). GATA3 is frequently mutated in breast cancer and is known to promote luminal cell differentiation in the mammary gland [71,73,74,76], but has not been previously studied in colon cancer. IGFBP2, which was linked with both patient death and tumor recurrence in our analysis, was another interesting hit, as it has been

Death											
	Method	**Cox**			**Wilma**			**Regsubsets**			
	Stages	*All*	*I-II*	*I-III*	*All*	*I-II*	*I-III*	*All*	*I-II*	*I-III*	**Total**
Antibody	**GATA3.M.V**	✓	✓	✓	✓	✓	✓		✓	✓	8
	Bid.R.C	✓		✓		✓	✓		✓	✓	6
	Rb.M.V		✓	✓		✓	✓			✓	5
	AMPK_alpha.R.C	✓	✓	✓	✓	✓					5
	Tau.M.C	✓	✓	✓	✓			✓			5
	IGFBP2.R.V	✓	✓	✓			✓				4
	Beclin.G.V				✓		✓	✓		✓	4
	Src_pY527.R.V				✓	✓			✓	✓	4
	c.Jun_pS73.R.C					✓	✓			✓	3
	X4E.BP1.R.V					✓	✓		✓		3
	Bim.R.V	✓			✓			✓			3
	COX.2.R.C				✓					✓	2
	Smad4.M.V				✓		✓				2
	ERK2.R.NA				✓				✓		2
	PR.R.V		✓	✓							2
	Annexin_I.R.V							✓	✓		2
	Chk1.R.C					✓		✓			2
	K.Ras.M.C		✓	✓							2
	MSH2.M.C				✓				✓		2
	p27_pT157.R.C				✓		✓				2
	p70S6K_pT389.R.V					✓	✓				2
	Smad3.R.V	✓				✓					2
	X4E.BP1_pT37.R.V				✓	✓					2
	XBP1.G.C	✓		✓							2
	ACC1.R.C				✓						1
	Akt.R.V				✓						1
	AMPK_pT172.R.V				✓						1
	Caspase.3_active.R.C						✓				1
	CDK1.R.V				✓						1
	Fibronectin.R.C	✓									1
	STAT5.alpha.R.V					✓					1
	TAZ.R.C					✓					1
	YB.1_pS102.R.V						✓				1
	B.Raf.M.NA				✓						1
	Bax.R.V							✓			1
	c.Kit.R.V			✓							1
	cIAP.R.V		✓								1
	Cyclin_D1.R.V		✓								1
	Cyclin_E1.M.V				✓						1
	eIF4E.R.V							✓			1
	INPP4B.G.C	✓									1
	Notch3.R.C		✓								1
	p27.R.V					✓					1
	p38_pT180_Y182.R.V									✓	1
	PRAS40_pT246.R.V							✓			1
	S6.R.NA							✓			1
	TAZ_pS89.R.C			✓							1
	X14.3.3_epsilon.M.C		✓								1

Table 8: All Proteins Associated with Patient Death, Sorted in Descending Order by the Number of Times Identified

Table on previous page. Proteins (indicatd by Antibody) identified by all three bioinformatics methods (Cox regression, Wilma algorithm, and Regsubsets algorithm) to be associated with death for Stages I-II, I-III, and All stages. Proteins identified by more than one bioinformatics method are indicated in **bold**.

Recurrence											
	Method	**Cox**			**Wilma**			**Regsubsets**			
	Stages	*All*	*I-II*	*I-III*	*All*	*I-II*	*I-III*	*All*	*I-II*	*I-III*	**Total**
Antibody	**COX.2.R.C**	✓	✓	✓	✓		✓	✓	✓		7
	c.Jun_pS73.R.C	✓			✓		✓	✓		✓	5
	Rb.M.V					✓	✓		✓	✓	4
	IGFBP2.R.V		✓	✓			✓				3
	Rb_pS807_S811.R.V					✓		✓	✓		3
	Beclin.G.V				✓		✓				2
	Smad4.M.V						✓			✓	2
	GSK3.alpha.beta.M.V				✓		✓				2
	HSP70.R.C		✓		✓						2
	p70S6K.R.V				✓			✓			2
	PEA.15.R.V		✓						✓		2
	PI3K.p85.R.V					✓	✓				2
	XIAP.R.C		✓	✓							2
	XRCC1.R.C			✓			✓				2
	YAP_pS127.R.C	✓	✓								2
	GATA3.M.V				✓						1
	ERK2.R.NA							✓			1
	PR.R.V						✓				1
	ACC1.R.C								✓		1
	AMPK_pT172.R.V					✓					1
	Caspase.3_active.R.C				✓						1
	CDK1.R.V		✓								1
	Fibronectin.R.C					✓					1
	STAT5.alpha.R.V							✓			1
	TAZ.R.C				✓						1
	YB.1_pS102.R.V			✓							1
	ACC_pS79.R.V					✓					1
	Akt_pS473.R.V				✓						1
	ARID1A.M.V				✓						1
	Caspase.7_cleavedD198.R.C	✓									1
	DJ.1.R.C			✓							1
	GSK3_pS9.R.V						✓				1
	Lck.R.V					✓					1
	MEK1_pS217_S221.R.V	✓									1
	NF2.R.C					✓					1
	Rad50.M.C					✓					1
	Rad51.M.C									✓	1
	Shc_pY317.R.NA						✓				1
	STAT3_pY705.R.V							✓			1

Table 9: All Proteins Associated with Tumor Recurrence, Sorted in Descending Order by the Number of Times Identified

Proteins (indicatd by Antibody) identified by all three bioinformatics methods (Cox regression, Wilma algorithm, and Regsubsets algorithm) to be associated with recurrence for Stages I-II, I-III, and All stages. Proteins identified by more than one bioinformatics method are indicated in **bold**.

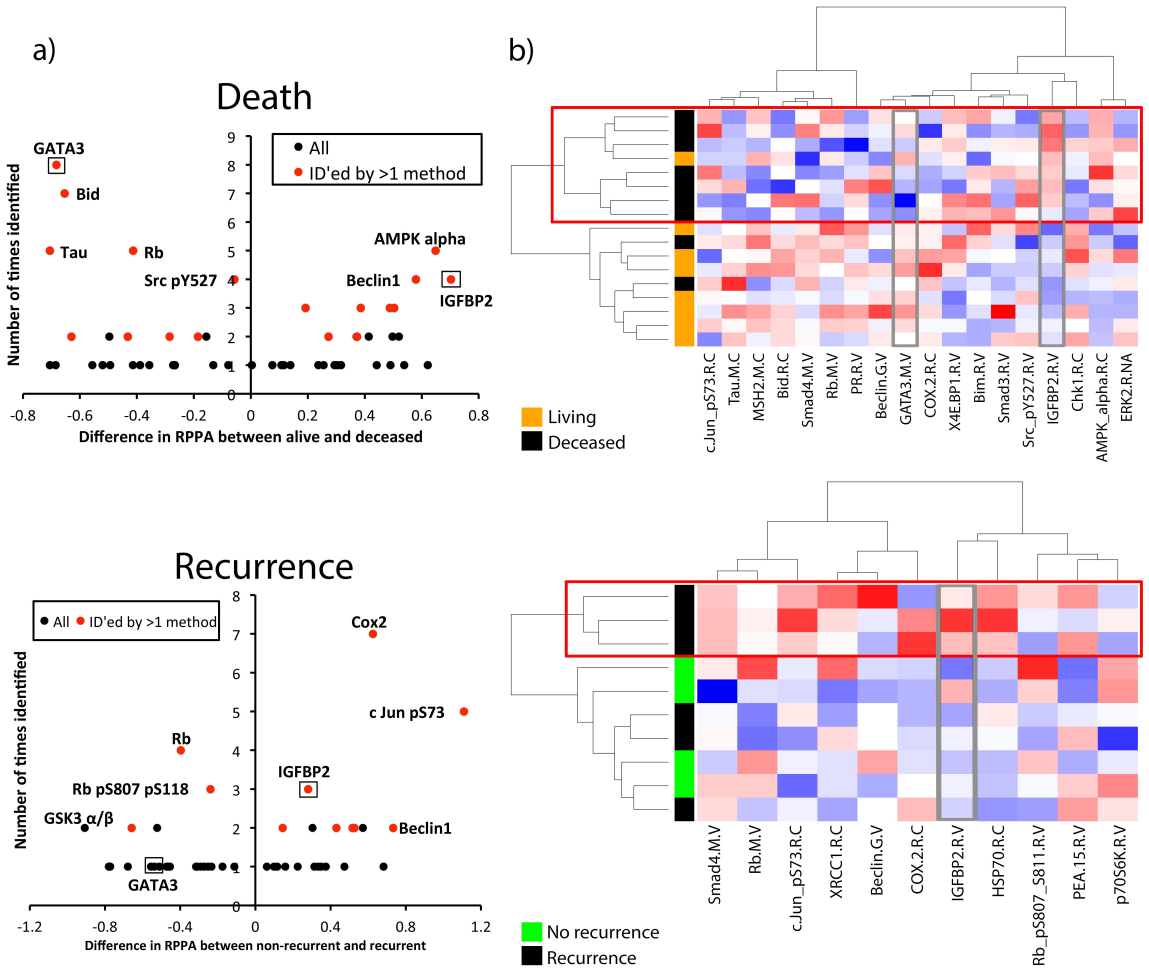


Figure 9: Visualization of Proteins Identified by Bioinformatics Analysis
a) Volcano plots were created by plotting the difference in the scaled RPPA expression for each protein vs. the number of times that protein was identified in the bioinformatics analysis. A positive value on the y-axis means that protein is upregulated in poor prognosis (recurrent or deceased) patients, while negative value on the y-axis means that protein is downregulated in poor prognosis (recurrent or deceased) patients. Proteins identified by more than one bioinformatics method (**Table 8** and **Table 9, bold**) are shown in red, and proteins selected for further analysis are boxed and labeled.
b) Heatmaps were created using unsupervised clustering of all top hits (**Table 8** and **Table 9, bold**) in stage I-II patients. Each row is a patient; each column is a protein. Red boxes outline poor prognosis (recurrence or death) clusters. Proteins selected for further analysis (GATA3 and IGFBP2) are outlined in grey boxes. RPPA, Reverse phase protein array.

associated with a number of cancer types but few studies have addressed its role in CRC [53–55].

Proteins associated with recurrence (**Table 9**) also included known CRC regulators, including the pro-inflammatory enzyme COX2 [128,129], phospho-c-Jun [130] and SMAD4 (reviewed in [131]). Some proteins were identified to be statistically associated with both death and recurrence, including the cell cycle regulator Rb, the autophagy regulator Beclin1, and IGFBP2. To visualize the expression of top hits identified by more than one bioinformatics method (**Table 8** and **Table 9**, indicated in bold; **Figure 9a**, red) in individual patient tumor samples, we created heatmaps using unsupervised clustering. Interestingly, clustering of data from Stage I-II patient tumors gave superior segregation of prognosis groups by the proteins than using data from Stages I-III or I-IV patient tumors. For both recurrence and survival, there was a "poor prognosis" cluster that segregated away from the remaining patients (**Figure 9b**, red boxes). Notably, the ability of the chosen proteins to cluster patients according to poor prognosis was also superior when using death as the outcome, perhaps due to the larger number of significant proteins or the larger sample size of Stage I-II patients with that follow-up metric (**Figure 9b**, compare death and recurrence heat maps).

Of the proteins identified in our analyses, GATA3 and IGFBP2 were the most novel as regulators of CRC. Visualization by heatmaps shows a decreased expression in GATA3 and increased IGFBP2 expression in tumors within the poor prognosis clusters (**Figure 9b**, grey boxes). Stage-adjusted survival plots revealed that TCGA patients with low GATA3 expression levels had a significantly increased risk of death, compared with patients whose tumors had high GATA3 levels. Patients whose tumors had high IGFBP2 expression had a trend towards decreased survival, but this did not reach statistical significance (**Figure 10a**). Importantly, both GATA3 and IGFBP2 had significantly altered RPPA expression in deceased patients for all stages, stages I-II, and

stages I-III (**Figure 10b, c**). Similar trends were seen in recurrent vs. non-recurrent patients, but the data did not reach statistical significance, potentially due to the smaller number of patients with recurrence follow-up data (**Figure 11**).

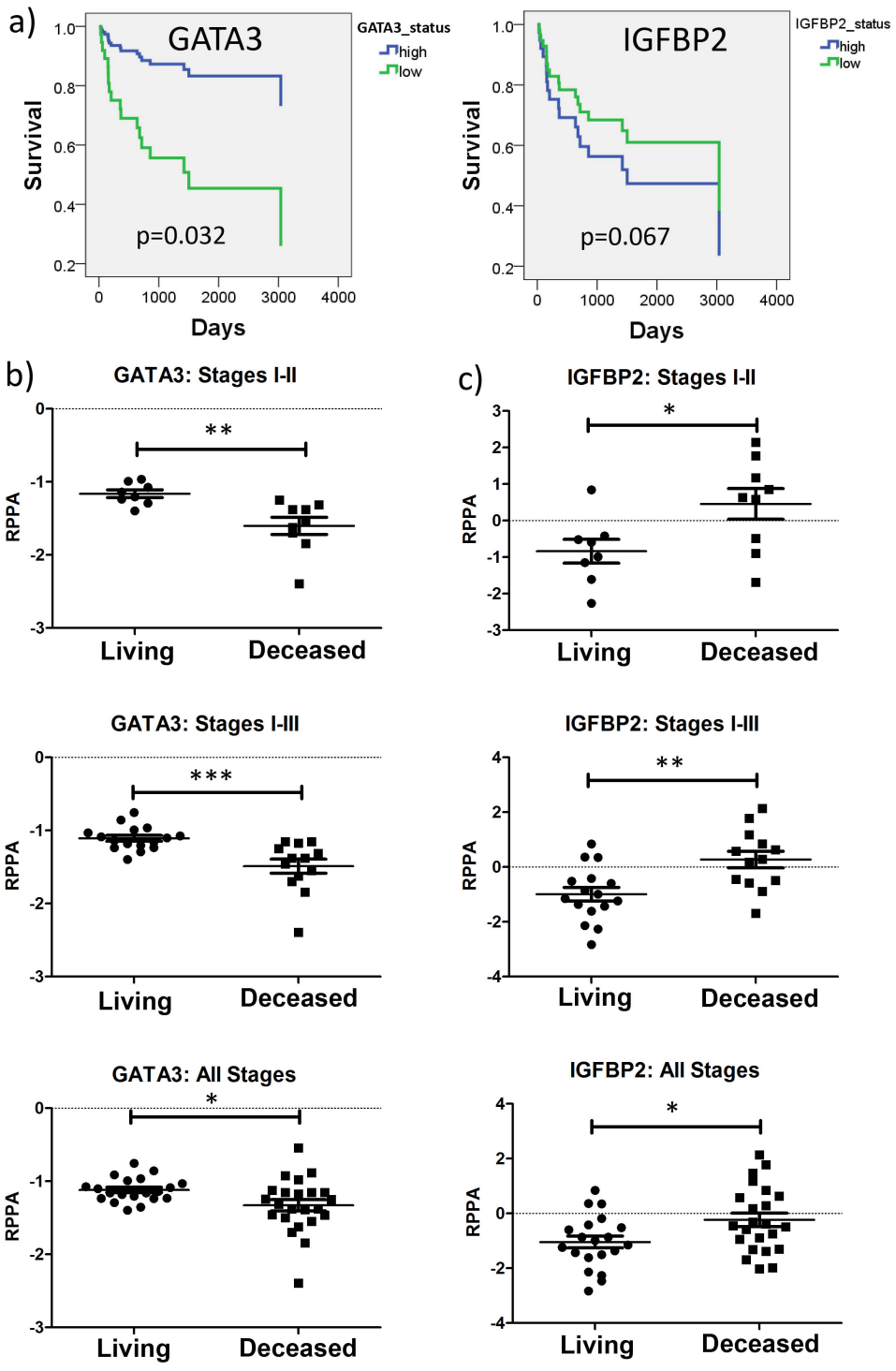


Figure 10: Survival Analysis of Selected Proteins from Bioinformatics Analysis
(**a**) Stage-adjusted survival plots for GATA3 and IGFBP2. (**b**) and (**c**) Comparison of RPPA-determined expression in living and deceased patients for GATA3 (**b**) and IGFBP2 (**c**). A cox proportional hazard analysis shows that IGFBP2 expression is significantly increased in deceased patients in Stages I-II, I-III, and I-IV, while GATA3 is significantly decreased in deceased patients in Stages I-II, I-III, and I-IV. * $p<0.05$, ** $p<0.01$, *** $p<0.001$. RPPA, Reverse phase protein array.

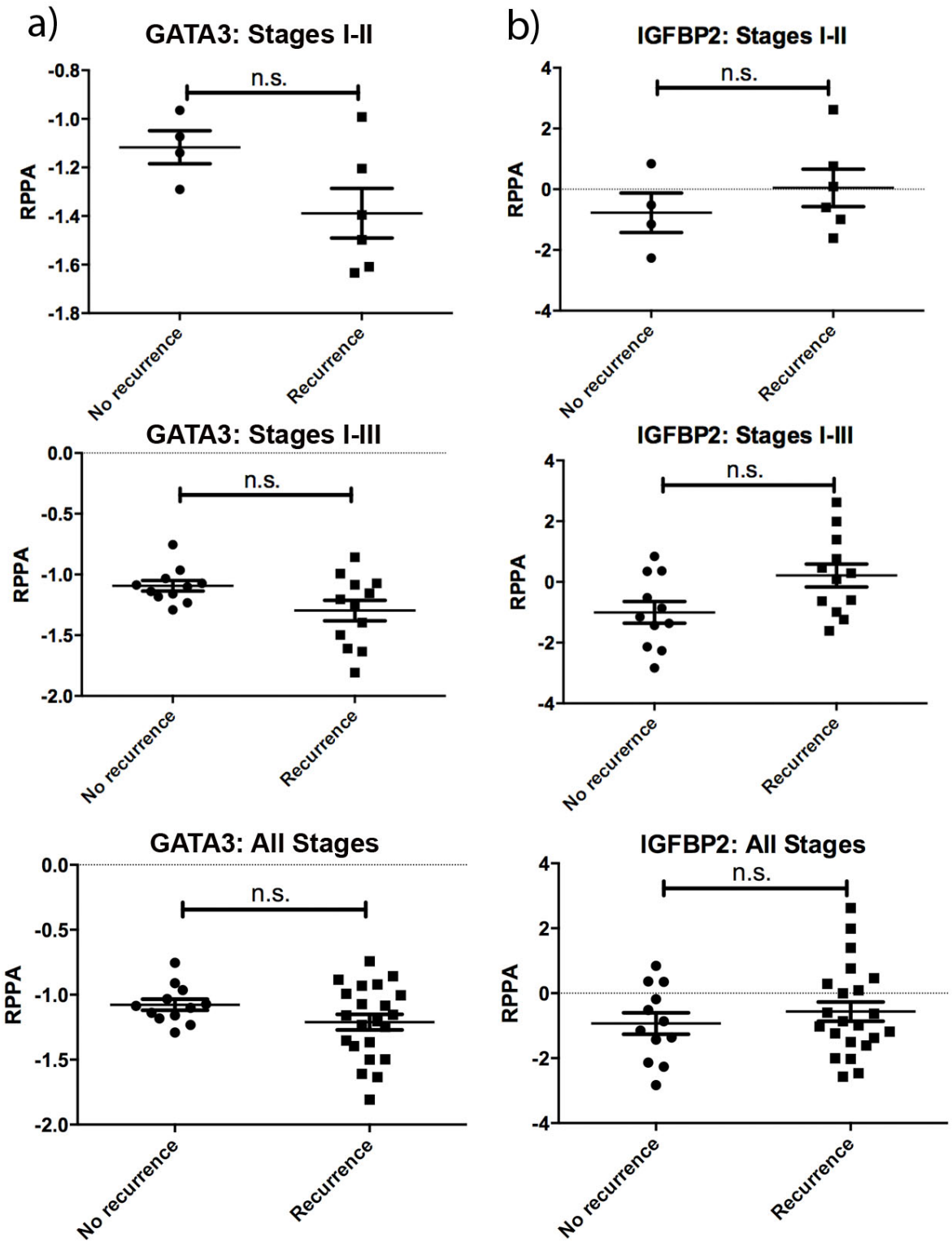


Figure 11: Recurrence Analysis of Selected Proteins from Bioinformatics Analysis
Decreased GATA3 **(a)** and increased IGFBP2 **(b)** expression are evident in recurrent patient tumors, but the data were not significant (n.s.). RPPA, Reverse phase protein array.

Validation of IGFBP2 in Human CRC TMA

To validate our findings in an independent tumor cohort, we obtained a tissue microarray (TMA) that contained 61 CRC samples and had available patient follow-up data. Patient characteristics are shown in Suppl. Table 5. Note that some clinical information, such as age or gender, was not available for all patients (**Table 10**). We stained the TMA slides with antibodies against IGFBP2 as well as with the epithelial marker cytokeratin in order to identify tumor cells (**Figure 12a, b**). We quantified the areas of both IGFBP2 staining and cytokeratin staining (representing total tumor area), and calculated the percent IGFBP2 positive area per tumor area in order to normalize to the amount of tumor present in each sample. This metric was used to divide patients into high or low IGFBP2 by median expression, and their survival or recurrence-free survival was compared. The results revealed that patients with IGFBP2 staining at or above the median had a significant reduction in both survival and recurrence-free survival time, independent of tumor stage (**Figure 12a, b, lower panels**). Staining of normal colon tissue also revealed strong staining in the bottom of the crypts (**Figure 12c**), consistent with a previous report [54].

TMA Patients		
Characteristics	Average age (years)	63.05
	Male	26
	Female	34
	Gender unknown	1
	Average follow up time (days)	2074.08
Recurrence	Average days to recurrence	674.95
	Recurrent	22
	Non-recurrent	39
Death	Average days to death	1180.05
	Deceased	21
	Living	40
Stage	Stage I	0
	Stage II	34
	Stage III	26
	Stage IV	0

Table 10: Characteristics of patients included in TMA analysis
Age, sex, follow-up time, recurrence status and time, vital status and time, and cancer stage are shown for TMA patients.

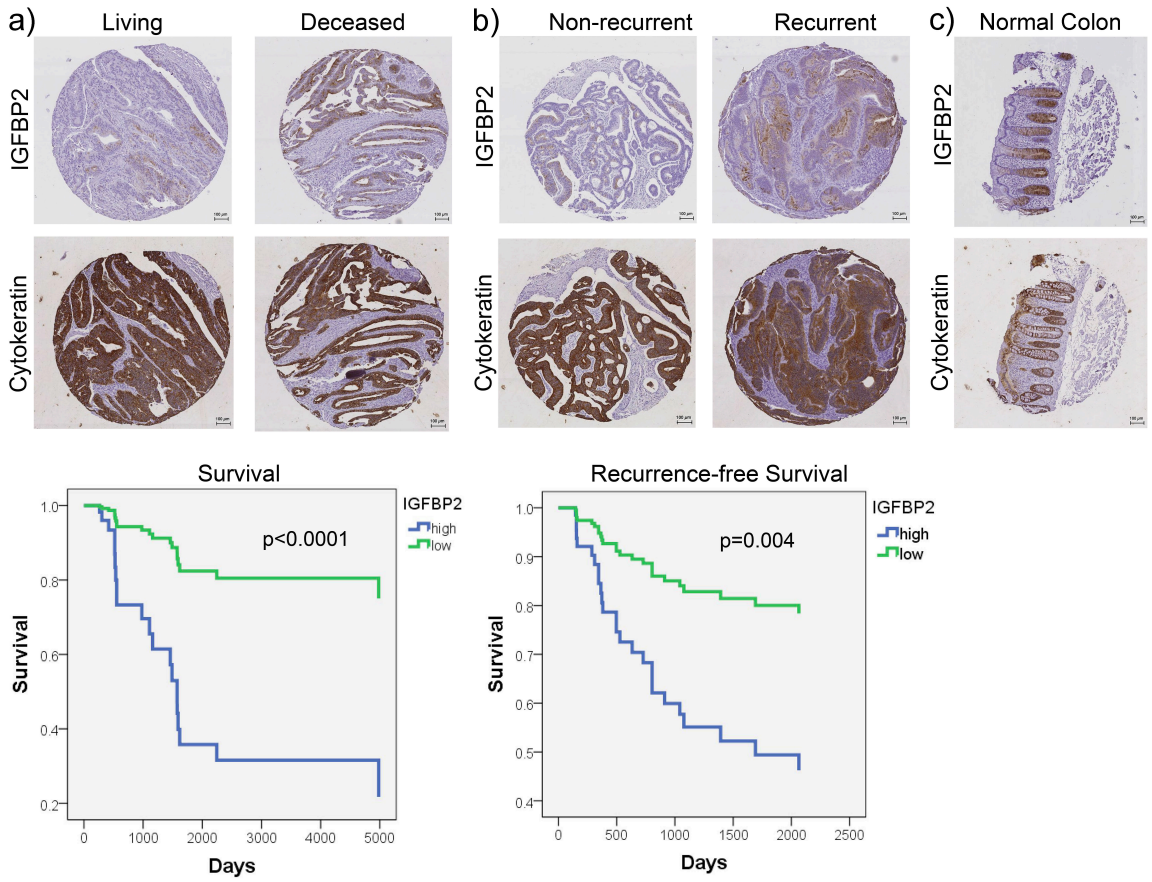


Figure 12: IGFBP2 Expression is Associated with Recurrence and Death in CRC in a Secondary Dataset

IHC immunostaining of a CRC tissue microarray for IGFBP2 and cytokeratin (epithelial marker) was performed. **a)** Representative IGFBP2 staining in living and deceased patients and Kaplan-Meier curve comparing survival of patients with low (below median) vs. high (at or above median) IGFBP2 staining. **b)** Representative IGFBP2 staining in non-recurrent and recurrent patients and Kaplan-Meier curve comparing recurrence-free survival of patients with low (below median) vs. high (at or above median) IGFBP2 staining. %IGFBP2-positive area of tumor was calculated using IGFBP2 area and cytokeratin area to identify tumor. Survival and recurrence-free survival plots are adjusted for stage. **c)** Representative IGFBP2 and cytokeratin staining in a representative normal colon sample. Scale bars indicate 100 μm.

Validation of GATA3 *in Vitro*

GATA3 is a transcription factor that was originally identified as a T-cell differentiation factor [117,118]. However, recent data indicates that GATA3 is also expressed in some epithelia (reviewed in [77]) . In breast cancer, GATA3 is frequently mutated [73,74]. In addition, low levels of GATA3 correlate with decreased breast cancer patient survival [121,124,132–134]. To determine whether GATA3 is expressed in CRC cells or only in T-cells, we stained CRC TMAs as well as matched normal and colon cancer tissue (**Figure 13**). Antibodies to cytokeratin (CK) and CD3 respectively marked the epithelial tumor cell and T-cell compartments. We found variable staining patterns with two different anti-GATA3 antibodies. Using the same antibody that was used to probe the TCGA RPPA samples (**Figure 13a**, GATA3 BD), there was weak cytoplasmic and occasional nuclear staining in the tumor cells and a small amount of nuclear staining in cells in the stromal compartment. It should be noted that this antibody had not been validated for IHC. Furthermore, we noticed variable staining of TMA sections from normal colon tissue, suggesting high sensitivity of this antibody to fixation conditions. We therefore tested two more antibodies that were validated for IHC. Using an antibody that has successfully been used for breast cancer stratification [121], we detected very light cytoplasmic staining of epithelial cells with some nuclear staining of stromal cells in normal colon samples, but no staining of epithelial or stromal cells in paired colon cancer samples (GATA3 SC, **Figure 13b**). Using a second validated IHC antibody (GATA3 LS), we found strong staining of the epithelial component of both normal colon tissue and colon cancer (**Figure 13b**). Interestingly, with both the SC and LS antibodies, it appeared that in normal colon tissue there was increased staining in epithelial cells at the mucosal surface with nuclear localization, compared to the deep

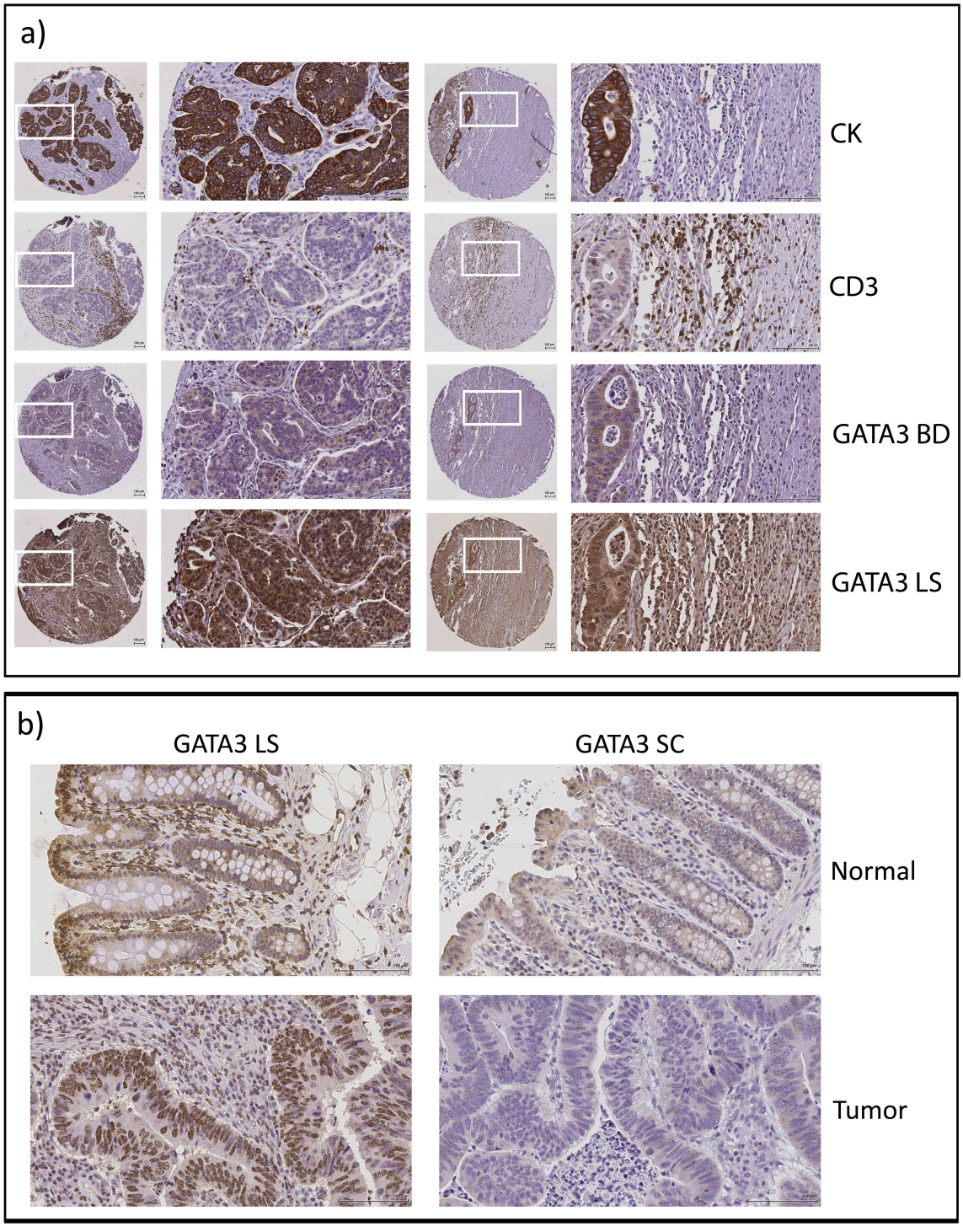


Figure 13: GATA3 is Expressed in Human CRC

a) Representative immunostained tissue sections from two patient tumors from the CRC TMA showing staining for epithelial tumor (cytokeratin, CK), T-cells (CD3), and two different GATA3 antibodies (BD and LS). **b)** Representative staining of matched normal colonic tissue and colon cancer samples for two different GATA3 antibodies (LS and SC). Note the variability of GATA3 staining with different antibodies.

crypts (**Figure 13b**). Staining of the TMA with GATA3 LS gave strong staining in both the nuclei and cytoplasm of tumor cells. However, there was a high background in many of the samples with apparently nonspecific staining throughout both the tumor and stromal compartment (**Figure 13a**), which made the samples unsuitable for quantitation. This high background may be due to overfixation of some of the TMA blocks, since it was not apparent on separate fixed tissues that were not part of the TMA (compare **Figure 13a** to **Figure 13b**, GATA3 LS staining).

We also checked the Human Protein Atlas (HPA) [135] for staining of colon tissues by GATA3 antibodies (**Figure 14**). The HPA also used three different antibodies. One of them, CAB016217, is the same as the antibody we tested that gave little to no staining of colon tissue (GATA3 SC). Likewise, they found little nuclear staining, and weak or negative cytoplasmic staining across both normal and colon cancer samples. The other two antibodies stained the epithelial component of both normal and colon cancer samples with primarily nuclear or nuclear + cytoplasmic staining patterns. Thus, with 4 out of the 5 antibodies tested by our laboratory and the HPA, nuclear GATA3 staining was seen in colon epithelial and cancer cells. However, due to the variability in intensity and pattern of staining, we were not able to perform quantitations to obtain information about prognostic significance.

To determine if we could use a gene expression dataset for validation, we tested whether GATA3 RNA expression by RNA sequencing correlated with GATA3 protein expression by RPPA in TCGA samples that had both types of data. There was no correlation between GATA3 RNA and protein expression (**Figure 15a**), so we were not able to use GATA3 RNA expression for correlative studies in a secondary tumor dataset. By contrast, IGFBP2 protein levels correlate well with IGFBP2 RNA levels (**Figure 15b**). There was no correlation between IGFBP2 protein and GATA3 protein levels, indicating there is likely no mechanistic link between these two proteins (**Figure 15c**).

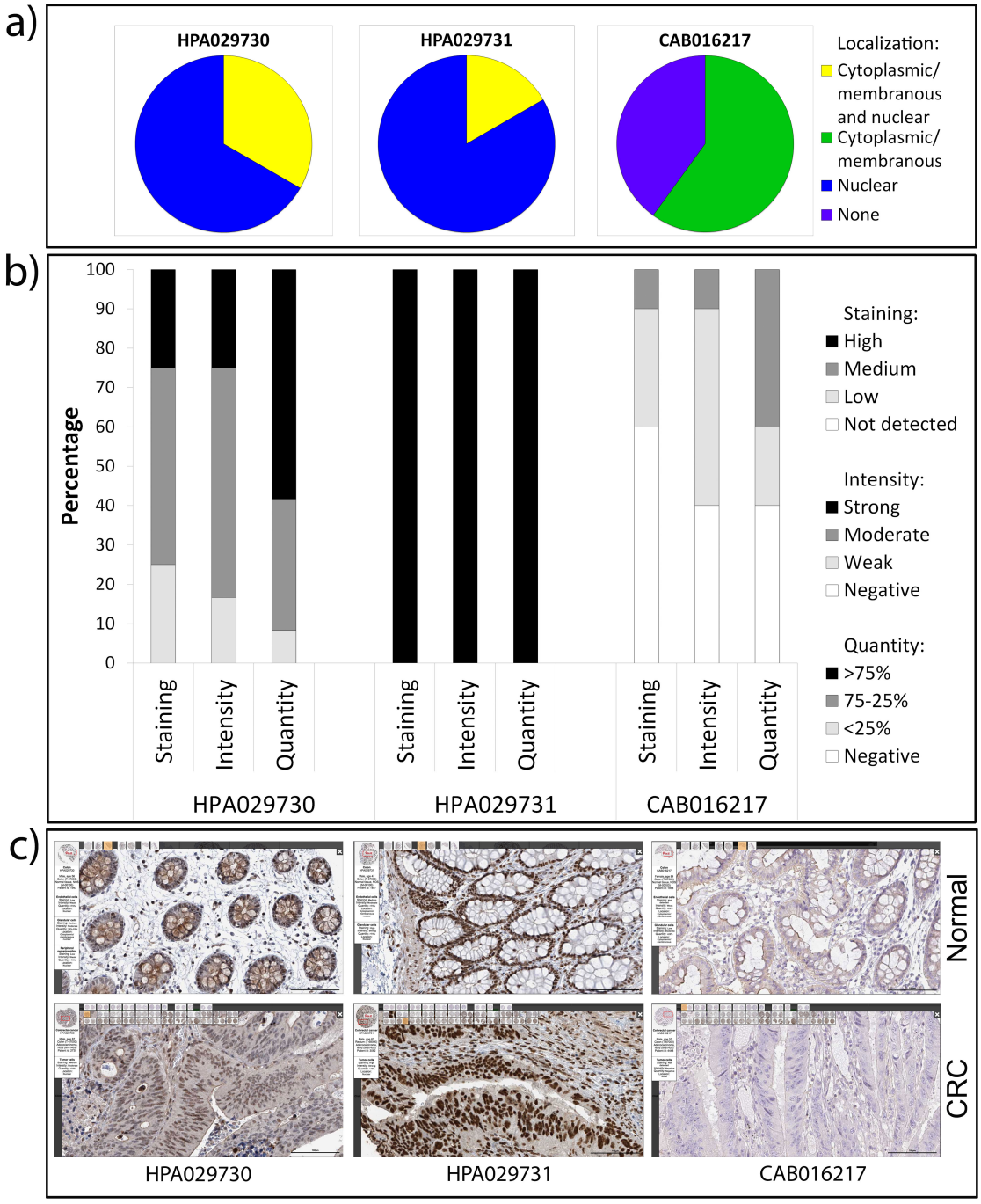


Figure 14: Comparison of GATA3 Staining Patterns in CRC Tumor Samples in the Human Protein Atlas (HPA) Using Three Different Antibodies

a) Subcellular localization; **b**) Staining, intensity, and quantity plots from the HPA. **c**) Representative images from matched normal colon tissue and colorectal cancer (CRC) samples with three different GATA3 antibodies, as indicated.

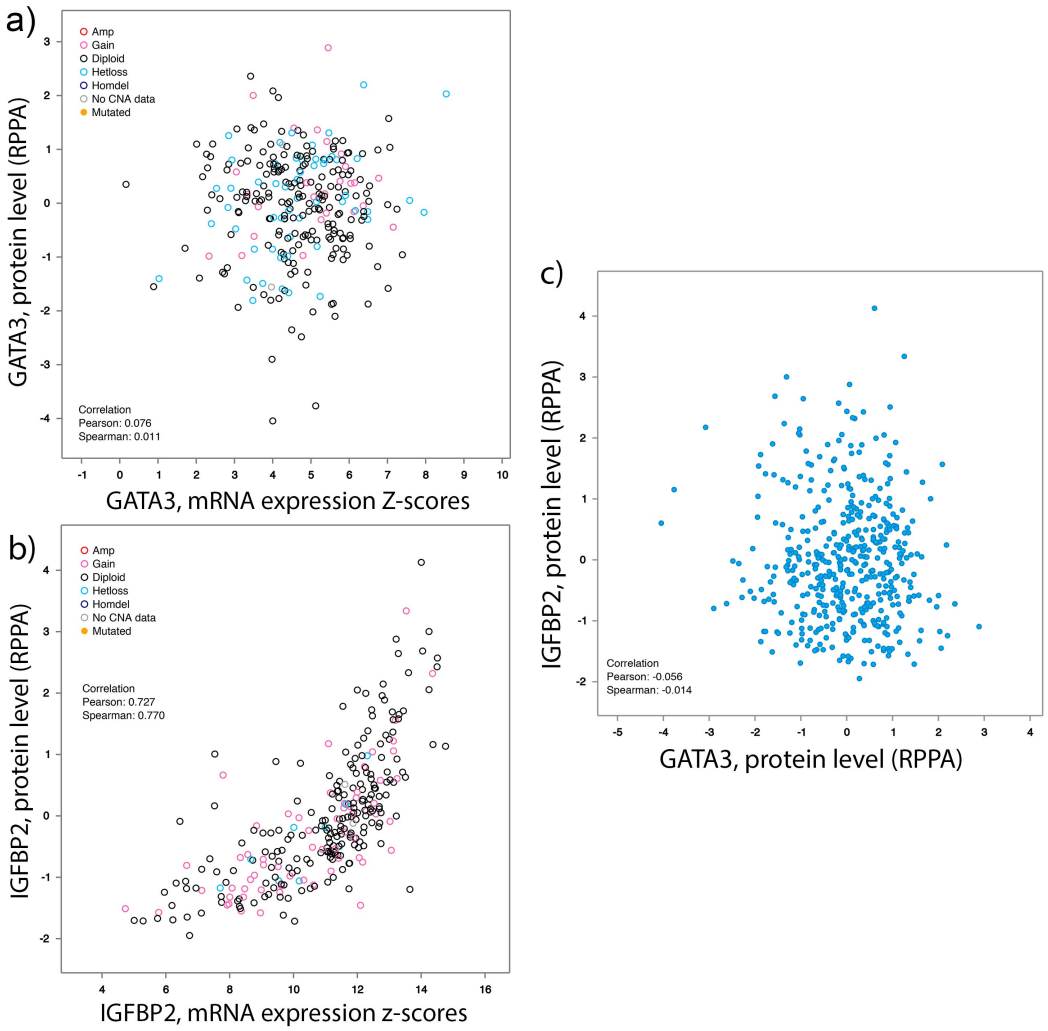


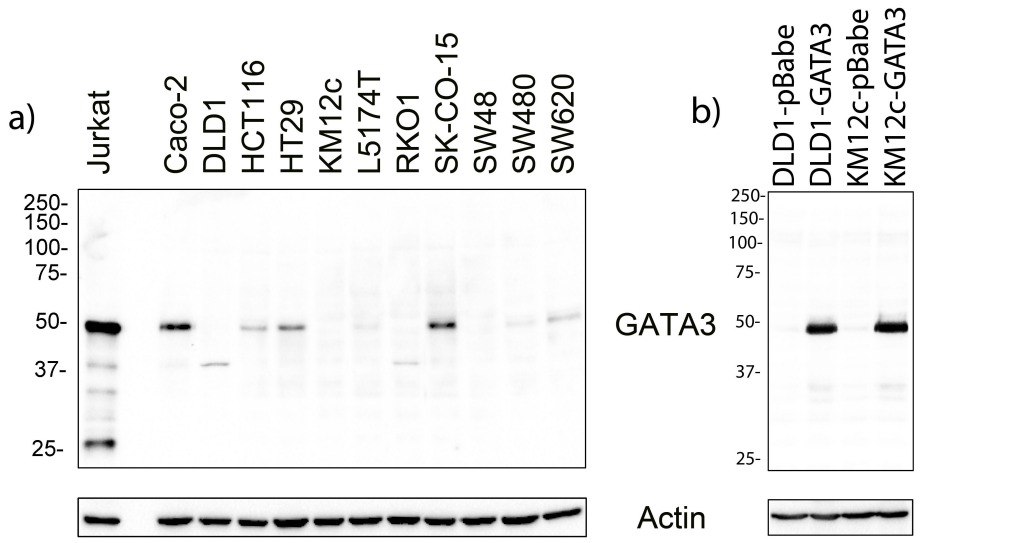
Figure 15: Correlations of RPPA Protein Expression with RNA Expression for selected proteins

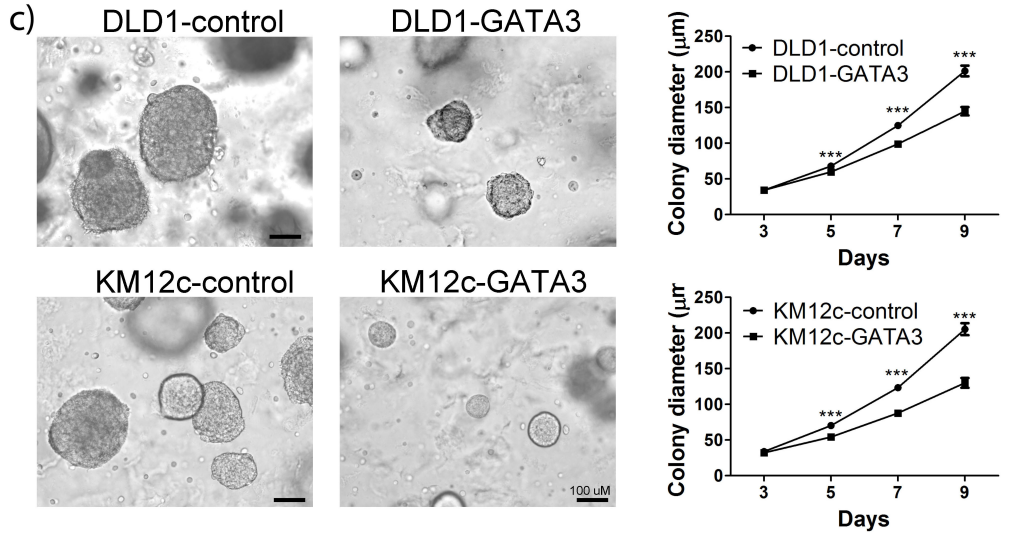
GATA3 and IGFBP2 RPPA and mRNA expression values from TCGA datasets were plotted and analyzed on an individual tumor basis. **(a)** GATA3 mRNA expression does not correlate with protein expression. **(b)** IGFBP2 mRNA expression does correlate with protein expression. **(c)** GATA3 and IGFBP2 RPPA expression levels do not correlate. All plots were created with cBioPortal using TCGA (2012) dataset [149]. RPPA, Reverse phase protein array.

Because we were unable to use GATA3 RNA expression, we tried gene set enrichment analysis (GSEA) in a secondary patient dataset as an alternative to gene expression validation sets. This method determines if a predefined set of genes (such as target genes from a transcription factor) is enriched in different biological phenotypes [136]. Using GSEA, we did not see significant enrichment of GATA3 target genes in poor prognosis patients in a secondary dataset. However, this is not entirely surprising: while GATA3 has been identified in the driver network of a differentiated CRC subtype with improved survival, but GATA3 target genes were not significantly enriched in this subtype [30].

As an alternative to validation with tissue samples or gene expression sets, we decided to investigate the biological role of GATA3 in colon cancer with *in vitro* experiments. We performed Western blot analysis of GATA3 levels in a panel of CRC cell lines with Jurkat T-cells as a positive control for GATA3 expression. Using the same antibody that was used in the TCGA RPPA analyses (GATA3 BD), we detected a band of the correct 48 kilodalton (kDa) size for GATA3 was detected by Western blot analysis. Compared with Jurkat cell expression, GATA3 was expressed at a much lower level in most CRC cell lines. GATA3 expression was undetectable in about half of the cell lines tested, including several with invasive characteristics, e.g. DLD1, SW480, and SW620 [137,138]. Consistent with the known role of GATA3 in cellular differentiation [64,69,77,118,139–142], the highest GATA3 expression was observed in the more differentiated cell lines, Caco-2, SK-CO-15 and HT-29 [143–145] (**Figure 16a**).

To investigate the role of GATA3 in CRC growth and invasion, we chose two of the invasive cell lines with undetectable GATA3 expression and stably expressed GATA3 in them using retroviral transduction (**Figure 16b**). We first tested the ability of the GATA3-expressing cells to form colonies after seeding as single cells in an embedded 3D Matrigel growth assay. Colony growth in this assay represents a


a)
Jurkat
Caco-2
DLD1
HCT116
HT29
KM12c
L5174T
RKO1
SK-CO-15
SW48
SW480
SW620
250-
150-
100-
75-
50-
37-
25-
b)
DLD1-pBabe
DLD1-GATA3
KM12c-pBabe
KM12c-GATA3
GATA3
Actin


c)
DLD1-control
DLD1-GATA3
KM12c-control
KM12c-GATA3
100 uM
Colony diameter (µm)
Days
***

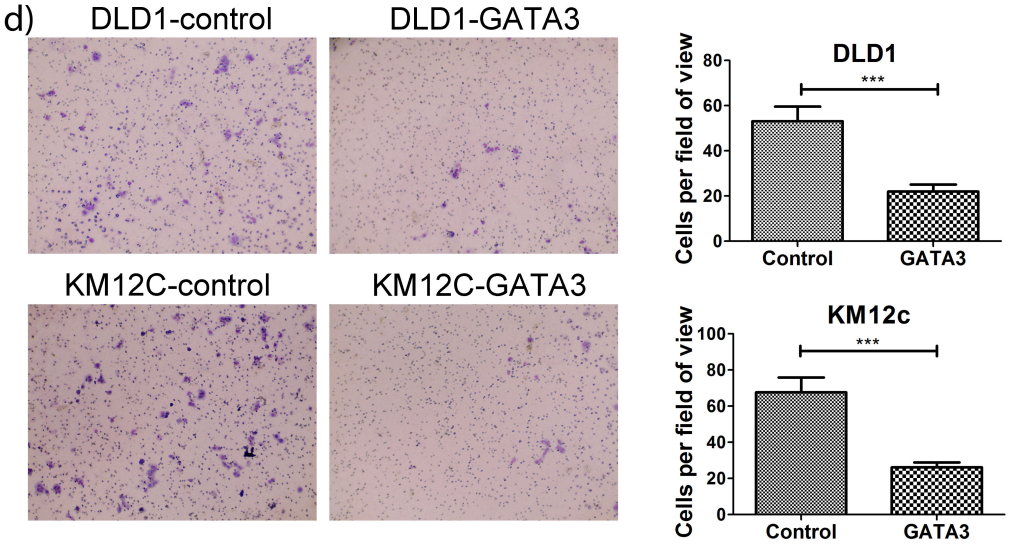

d)
DLD1-control
DLD1-GATA3
KM12C-control
KM12C-GATA3
DLD1
KM12c
Cells per field of view
Control
GATA3
***

Figure 16: GATA3 Expression Affects CRC Aggressiveness

Figure on previous page. a) Representative Western blot (of 2 blots) showing that GATA3 is expressed in a subset of CRC cell lines. Jurkat is a T-cell line and used as a positive control. Higher expression is seen in the more differentiated cell lines Caco-2, HT-29, and SK-CO-15. **b)** Western blot showing engineered expression of GATA3 in DLD1 and KM12c CRC cell lines. pBabe is an empty vector control. **c)** Colony growth of engineered CRC cell lines in 3D Matrigel. Left: Representative images from day 9. Right: Growth curves. Data were gathered from duplicate wells from 3 independent experiments. The mean is plotted and error bars represent 95% CI. **d)** Invasion of CRC cell lines across Transwell filters. Left: Representative images of the bottom of Transwell filters after 48 hours invasion. Right: Quantitation of invaded cells/field. Data from five random fields per filter x triplicate filters for each of 3 independent experiments. Error bars represent +/- SEM. *** $p<0.001$.

combination of growth and matrix remodeling activity, since the cells are fully embedded in 90% Matrigel [146–148]. Compared with control cells, GATA3-expressing cells formed smaller colonies in this 3D culture environment, an effect that was statistically significant beginning at day 5 (**Figure 16c**). To determine whether the smaller colony size of GATA3-expressing cells was due to an intrinsic decrease in proliferation rate, we cultured them in 2D in the presence or absence of serum and used automated microscopy to follow the number of cells over a period of 5 days. GATA3 expression had no effect on cell numbers in the presence or absence of serum (**Figure 17**). To determine if GATA3 specifically controls CRC invasiveness, control and GATA3-expressing cells were allowed to invade for 48 h across a bed of Matrigel in a Transwell invasion assay. For both of the tested CRC cell lines, GATA3-expressing cells exhibited significantly decreased invasion compared to control cells (**Figure 16d**). Taken together, these data indicate that GATA3 controls CRC invasiveness.

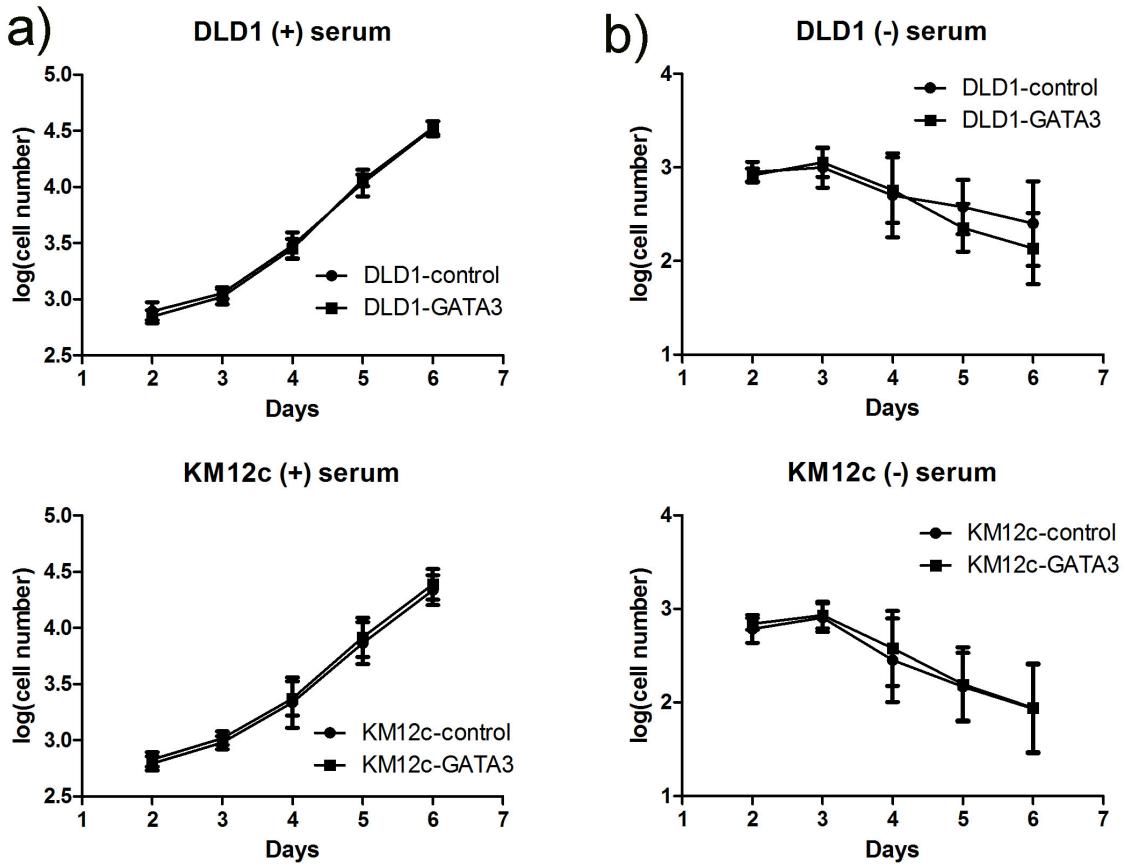


Figure 17: GATA3 Does Not Affect 2D CRC Proliferation

Growth curves (log base 10 of the cell number) from CRC cell lines grown in the presence of 10% serum (**a,** "(+) serum)") or the absence of serum (**b,** "(-) serum"). Cells were plated in triplicate and imaged on a Cellavista automated microscope in 3 independent experiments. Mean is plotted and error bars represent 95% confidence intervals. No significant differences were observed between control and GATA3-OE cells for either cell line.

DISCUSSION

In this study, we used high throughput protein and phospho-protein expression data from the TCGA to identify candidate drivers of CRC aggressiveness. By linking RPPA data to patient death or recurrence and using multiple statistical approaches, we identified both known and novel biomarkers of CRC aggressiveness. The top hit in our survival analysis was the transcription factor GATA3, for which low levels were correlated with death. Follow-up experiments indicated that GATA3 is expressed in CRC and suppresses the invasive behavior of CRC cells. We also validated the prognostic value of the known but understudied molecule IGFBP2 in a secondary CRC dataset. These data indicate that RPPA and other high throughput protein datasets are useful for identifying potential biomarkers and drivers of aggressive tumor behavior, especially for proteins whose RNA expression does not correlate to protein expression, such as GATA3.

Gene expression signature discovery has been dominated by transcript profiling technologies. Since we previously found that a small RPPA dataset from human tumors can be useful as a biological discovery tool [31], we tested its utility in a larger dataset from TCGA in this study. In addition to identifying proteins known to drive CRC progression, we identified several novel or understudied proteins associated with recurrence or death of CRC patients. These included IGFBP2 and GATA3, which were identified by multiple statistical methods, and a number of additional proteins that were detected by multiple (**bold**) or any method (**Table 8**,**Table 9**). Validation of IGFBP2 by TMA staining and GATA3 *in vitro* suggests that our bioinformatic approach has utility and biological validity. Moreover, our analysis showed that GATA3 mRNA levels were not predictive of GATA3 protein levels (**Figure 15**). Consistent with recent reports

showing that RNA and protein expression levels frequently do not correlate with each other [27,150], these data highlight the necessity of incorporating proteomics data into gene signature studies.

Our approach uses a comparison of tumor tissue between good and poor prognosis patients, which differs from previous proteomics studies that have either focused on differences between tumor and normal control tissues or on stage-specific differences [151–158]. These studies have given insight in to the pathophysiology of CRC progression. However, our goal was to identify markers that are independent of stage and could be potentially used in the future to predict prognosis in early stage patients. It is agreed that Stage III and IV patients universally benefit from chemotherapy [107], but the treatment decision for early Stage II patients is more complicated: there is disagreement over whether Stage II patients should [110–112] or should not [108,109] receive additional chemotherapy. While these findings are clearly a long way away from translation to the clinic, we posit that our general approach has the potential to identify biomarkers that can be used to identify early patients that could benefit from additional adjuvant therapy.

A limitation of our study was that the TCGA CRC patient sample set was smaller for RPPA than for more standard analyses such as RNA Seq or DNA mutations (196, compared to 244 and 224 patient samples) [149]. In addition, many samples either did not have clinical follow-up or had only short follow-up time, further reducing sample size. Additionally, there were no other published RPPA datasets in CRC that contained analysis of our proteins of interest. Therefore, validation of our findings required either staining of tissue microarrays or *in vitro* experiments. As RPPA datasets accumulate, we anticipate that there will be larger and multiple independent validation datasets with longer follow-up times. Finally, because RPPA is an antibody-based technique, it is usually typically limited in the number of proteins detected. Higher throughput proteomic

approaches may solve this problem, although they are often unsuitable for quantitation of posttranslational modifications such as phosphorylation.

We identified increased expression of insulin-like growth factor binding protein 2 (IGFBP2) to be associated with CRC recurrence and death. High levels of IGFBP2 have been associated with poor prognosis in several cancer types. In breast cancer, IGFBP2 has increased expression compared to normal samples [159]. IGFBP2 has also been shown to promote invasion of ovarian cancer cells [43]. In CRC, IGFBP2 has been reported to be upregulated compared to normal colon epithelia [55] with a trend towards higher expression in more advanced CRC [53]. Interestingly, IGFBP2 is expressed predominantly in the crypts of normal colon tissue (Fig 3a and [54]), opposite to the pattern we observed with GATA3 expression and suggesting a stem-cell-like expression pattern. Notably, IGFBP2 has been connected to both hematopoietic and glioma stem cell expansion and survival [149,160]. Stem cells are progenitor cells that have limitless capacity for self-renewal and can differentiate into a number of other cell types; in cancer, similar signaling pathways regulate cancer cell growth, and cancer stem cells can drive tumorigenesis [161]. In addition, IGFBP2 overexpression in CRC cell lines was recently found to promote CRC tumorigenesis and metastasis [54]. Those data are consistent with our finding that high IGFBP2 expression in CRC tumors is significantly associated with death and recurrence in two independent datasets of CRC patients (**Table 8**,**Table 9; Figure 12**).

The top hit in our survival analysis was GATA3, which has not previously been studied in CRC. GATA3 is a transcription factor that was originally identified in T-cells, and controls the differentiation of TH2 cells [64,118,141,142], skin cells [69], hair follicles [139] and luminal cells in the mammary gland [71,76]. The importance of GATA3 for mammary luminal cell proliferation and differentiation is suggested by the high expression of GATA3 in luminal breast cancers and recurrent mutations in the luminal

subtype that stabilize GATA3 protein expression levels [73,74]. Conversely, similar to our findings in CRC, low GATA3 levels are associated with poor patient prognosis in breast cancer [121,124,132–134]. At this point it is unclear whether that represents the overall poor outcome of non-luminal breast cancers or an active role for GATA3 in suppressing aggressive behavior. Support for the latter possibility is provided by data indicating that re-expression of GATA3 in non-luminal breast cancer cells is sufficient to induce differentiation and suppress lung metastases [76].

FUTURE DIRECTIONS

To improve our analysis, more stringent bioinformatics techniques could be employed to have more confidence in proteins selected for further analysis. One such method is called elastic net, which uses both L1 and L2 regularization; specifically, it improves upon L1-penalized least squares method used by the least absolute shrinkage and selection operator (lasso) [162] by combining it with an L2 penalty on the residual sum of squares (called ridge regression) to improve variable selection [163]. This results in the ability of elastic net to select groups of correlated predictors (genes or proteins, for example) even when the number of predictors is far greater than the number of observations; it was originally suggested to be used in analysis of microarray data [164]. We attempted to utilize the elastic net technique for analysis of TCGA data; unfortunately, the sample size was too small in most cases, so these data were not included in our overall summary. However, the algorithm was able to compute results for two groups: Recurrence/All Stages and Survival/Stages I-III (**Table 11**). Importantly, many of the proteins identified by Cox regression, Wilma, and Regsubsets algorithms were also identified by elastic net, including GATA3 and IGFBP2 (**Table 11**). This confirms that our method using these techniques has merit. It is unknown why the algorithm was successful for these two particular groups; however, increasing the samples size by adding more patients would greatly help this analysis.

The specific mechanistic role of IGFBP2 in CRC is still unclear. In most cancer types, high expression of IGFBP2 has been correlated with more aggressive cancers, and it been shown to promote proliferation, invasion, and other tumor-promoting phenotypes [34–52]. However, in CRC studies have shown contradictory roles of IGFBP2. Some studies have indicated that IGFBP2 has a tumor-promoting role [53–57],

	Recurrence / All Stages	Survival / Stages I-III
Proteins Identified	Beclin.G.V	AMPK_alpha.R.C
	COX.2.R.C	Annexin_I.R.V
	GSK3.alpha.beta.M.V	Beclin.G.V
	NF2.R.C	Bid.R.C
	c.Jun_pS73.R.C	FOXO3a_pS318_S321.R.C
	p70S6K.R.V	GATA3.M.V
		IGFBP2.R.V
		MIG.6.M.V
		NF2.R.C
		Rb.M.V
		XBP1.G.C
		c.Jun_pS73.R.C

Table 11: Proteins identified by elastic net analysis
For the elastic net analysis, the same sample set was used for Wilma and Regsubsets (**Table 4**). Proteins are listed in alphabetical order.

while other studies have shown a tumor-inhibitory effect [58–61]. The reason for these converse effects of IGFBP2 on CRC is unclear. Tumor-promoting effects of IGFBP2 are generally due to its binding to factors unrelated to the IGF signaling pathways, such as ECM components, proteoglycan receptor integrin receptors, and nuclear transportation complexes [33]. It is probable that the tumor inhibitory effects are due to IGF-dependent effects: when IGFBP2 binds IGF1 or IGF2, it sequesters them and prevents the tumor-promoting downstream Ras/MAPK and PI3K/AKT signaling pathways [32]. Additionally, it has been suggested that exceedingly high IGFBP2 levels could account for the inhibition of tumor growth in IGFBP2 transgeneic mice [33,61]. However, why this effect is seen in CRC and not in any other cancer type to date is unknown. The work presented in this thesis supports a tumor-promoting activity of IGFBP2, since high levels of IGFBP2 were found in recurrence and deceased patients as compared to non-recurrent and living patients with at least three years of follow-up data (**Figure 12**). Interesting future directions would include studies to determine the mechanism of how IGFBP2 is promoting CRC pathogenesis and in particular, through which binding partners it is acting. IGFBP2 represents a possible target for future chemotherapy; however, much more research must be done to ensure specific targeting of IGFBP2 tumor-promoting activities and not the tumor-inhibitory effects.

In CRC, the mechanistic role of GATA3 also still remains to be defined. One possibility is that GATA3 controls CRC differentiation, similar to its function in T-cells and luminal breast cells. Consistent with our prediction, IHC stains of normal colon tissue showed higher staining in the superficial mucosa, where the most differentiated cells should be (**Figure 13**). In addition, the most differentiated CRC cell lines in our panel had the highest GATA3 expression (**Figure 16**). Additionally, we previously identified three transcriptional subtypes of CRC and then identified subtype-specific driver networks by integrating mutation and copy number alteration data from each subtype

with a protein signaling network using a random walk approach [30]. GATA3 was included in the driver network for the "differentiated subtype" with relatively good survival outcome; GATA3 mRNA was not significantly upregulated in this subtype, although that is not surprising since we show that GATA3 mRNA and protein levels do not correlate (**Figure 15**). Consistent with our prediction, IHC stains of normal colon tissue showed higher staining in the superficial mucosa, where the most differentiated cells should be (**Figure 13b**, normal). In addition, the most differentiated CRC cell lines in our panel had the highest GATA3 expression (**Figure 16a**).

Another nonexclusive possibility is that GATA3 regulates TGF-β signaling, a key pathway regulating CRC aggressiveness, as reported in breast cancer [82]. TGF-β signaling is widely recognized for its effect on tumor initiation and progression of many cancers, including CRC [165–167]. The role of TGF-β in cancer is complex: early in tumor development, TGF-β has a tumor suppressive role, but in later stages it can act as a tumor promoter [167]. It has been reported that GATA3 negatively regulates TGF-β signaling in breast cancer by binding to Smad4, thus regulating cell migration, invasion, and invadopodia formation [82]. To investigate if a similar mechanism existed in CRC cell lines, we obtained a TGF-β luciferase reporter and transfected it into control and GATA3-OE cells along with an internal Renilla control. Interestingly, we see decreased activity of a TGF-β reporter in GATA3-OE cell lines (**Figure 18**), indicating that GATA3 overexpression causes decreased TGF-β signaling in these cell lines. Further work is required to determine if any of these or other mechanisms are responsible for the role of GATA3 in CRC.

In the future, it is feasible that GATA3 could be used as a prognostic biomarker in the clinic to guide treatment decisions for stage II CRC patients. To determine if a biomarker is an independent prognostic factor, a multivariate analysis must be done to

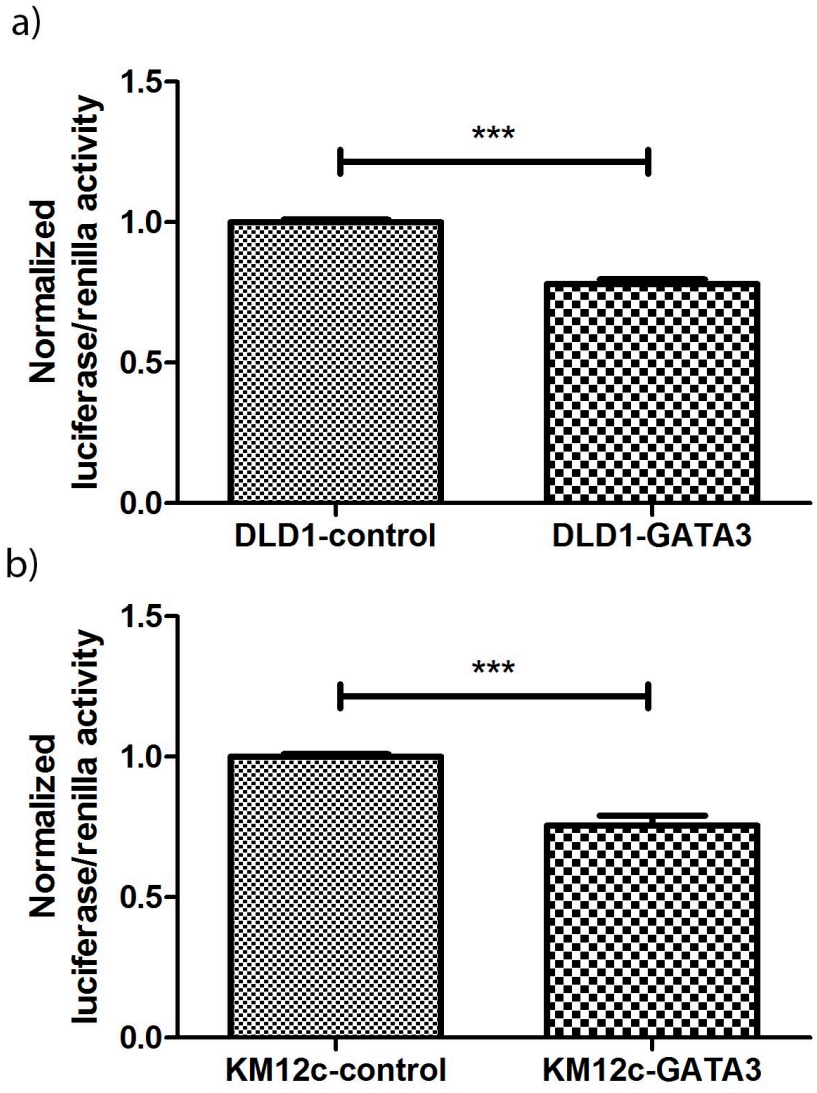


Figure 18: TGF-β Activity is Decreased in GATA3-OE CRC Cell Lines Compared to Control

Bar charts showing the average normalized TGF-β activity for two control and GATA3-OE cell lines. Cells were transfected with a TGF-β luciferase reporter and an internal renilla control, and fluorescence of cell lysates was measured. N=3.

compare the new biomarker with established ones [168]. Unfortunately, there was not enough data for us to perform an extensive multivariate analysis; in general, 10 events per variable are needed so overfitting does not occur [169]. We instead compared GATA3 RPPA expression in different categories of known prognostic determinants, such as tumor stage, tumor size/spread (T), lymph node spread (N), metastasis status (M), gender, age, venous invasion, lymphatic invasion, prior diagnosis, microsatellite instability, non-nodal tumor deposits, perineural invasion, synchronous colon cancer present, colon polyps present, history of colon polyps, expression loss of mismatch repair proteins, anatomic location, histological type, and race (**Figure 19**). The only significant differences were a slight increase in GATA3 RPPA expression from Stage II to III patients and an increase from N0 to N1 patients. Since this is the opposite effect of what we saw in our TCGA analysis and *in vitro* studies, it suggests that GATA3 may be independent of stage and lymph node status. All other differences were not significant, indicating that GATA3 RPPA expression is not linked to any of these prognosis factors. While this is suggestive that GATA3 may be a useful biomarker in CRC, more data needs to be collected to perform a multivariate analysis with these factors to determine if GATA3 is an independent prognostic biomarker.

It is also possible that future therapeutics could target GATA3 in CRC. The downregulation of GATA3 in more aggressive tumors and our work showing overexpression of GATA3 decreases tumor phenotypes points to a tumor suppressor-like role for GATA3 in CRC. However, tumor suppressors are more difficult to target than oncogenes since their activity must be restored, not inhibited [170]. Most cancer therapeutics are small molecule inhibitors that bind to and inhibit the activity of kinases [171]. It is possible to use small molecule inhibitors upstream of tumor suppressor genes, but this is challenging since a specific target upstream of the tumor suppressor gene that increases its expression must first be identified [170]. One example is

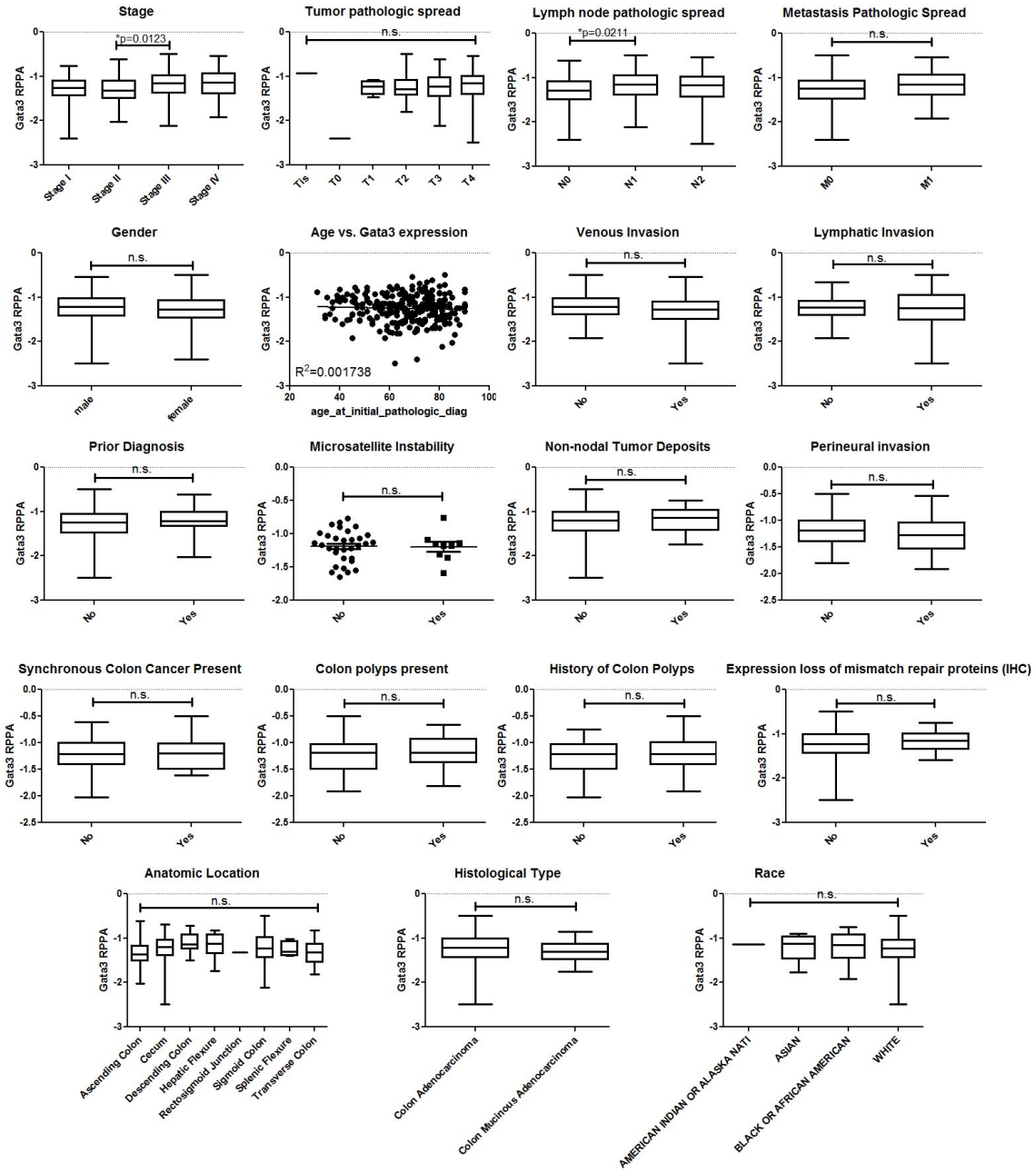


Figure 19: GATA3 RPPA expression is not significantly affected by clinical prognostic biomarkers.

Bar charts or scatter plots comparing GATA3 RPPA expression in different categories. A student's T-test was used to compare two variables, an ANOVA was used to compare more than two variables, and a Pearson's correlation was used for the age vs. RPPA correlation. Significant p-values are denoted by an asterisks. RPPA, Reverse phase protein array.

restoring retinoblastoma protein (Rb) function. Rb was the first identified tumor suppressor gene [172], and its expression can be indirectly restored using inhibitors against oncogenic kinases such as MEK (Trametinib) [173], PPARα (Fenofibrate) [174], and PI3K (LY294002) [175]; when used in combination with 5-fluorouracil (5-FU), a common treatment for CRC, these three Rb-restoring agents enhance its efficacy [176]. Another common tumor suppressor, p53, is negatively regulated by its interaction with mouse double minute homolog 2 (MDM2) [177]. Thus, the p53 pathway can be reactivated with small molecule inhibitors of MDM2 [178,179]. For this strategy to be successful with GATA3 in CRC, more work needs to be done to identify an appropriate upstream target of GATA3.

CHAPTER IV: IDENTIFICATION OF THE CORTACTIN BINDING PROTEIN SHANK2 AS A NOVEL REGULATOR OF HNSCC AGGRESSIVENESS

INTRODUCTION

11q13 Amplification in HNSCC

Genetic instability is one of the hallmarks of cancer [4], and HNSCC has a number of genetic alterations that occur during disease progression (**Figure 20**) [180]. One of these alterations is amplification of 11q13, which is a late-stage event thought to occur through a breakage-fusion-bridge cycle [181]. 11q13 amplification occurs in 30-40% of HNSCC patients [95–97], and is found less commonly (up to 24%) in breast [73,182], ovarian [183], small cell lung cancer [184], pancreatic [185], esophageal [186], and bladder [187] cancers. 11q13 has also been identified by the TCGA as an important alteration in a pan-cancer analysis: twelve human cancer types were divided into distinct classes characterized by either mutations or copy number changes, and 11q13 was a defining feature of one of these subclass [188]. Importantly, 11q13 amplification correlates with more aggressive disease in multiple cancer types: increased tumor grade, increased recurrence, increased lymph node metastases, and decreased survival in HNSCC [189–194]; survival rate and distant metastasis in esophageal cancer [186]; decreased recurrence-free survival, lymph node metastases, decreased survival, increased recurrence, and increased distant metastasis in breast cancer [195–197]; poor prognosis in pancreatic cancer [185]; and larger and poorly differentiated tumors in lung cancer [198]. 11q13 amplification is thought to be a late stage event, driving tumor progression instead of initiation [180,199].

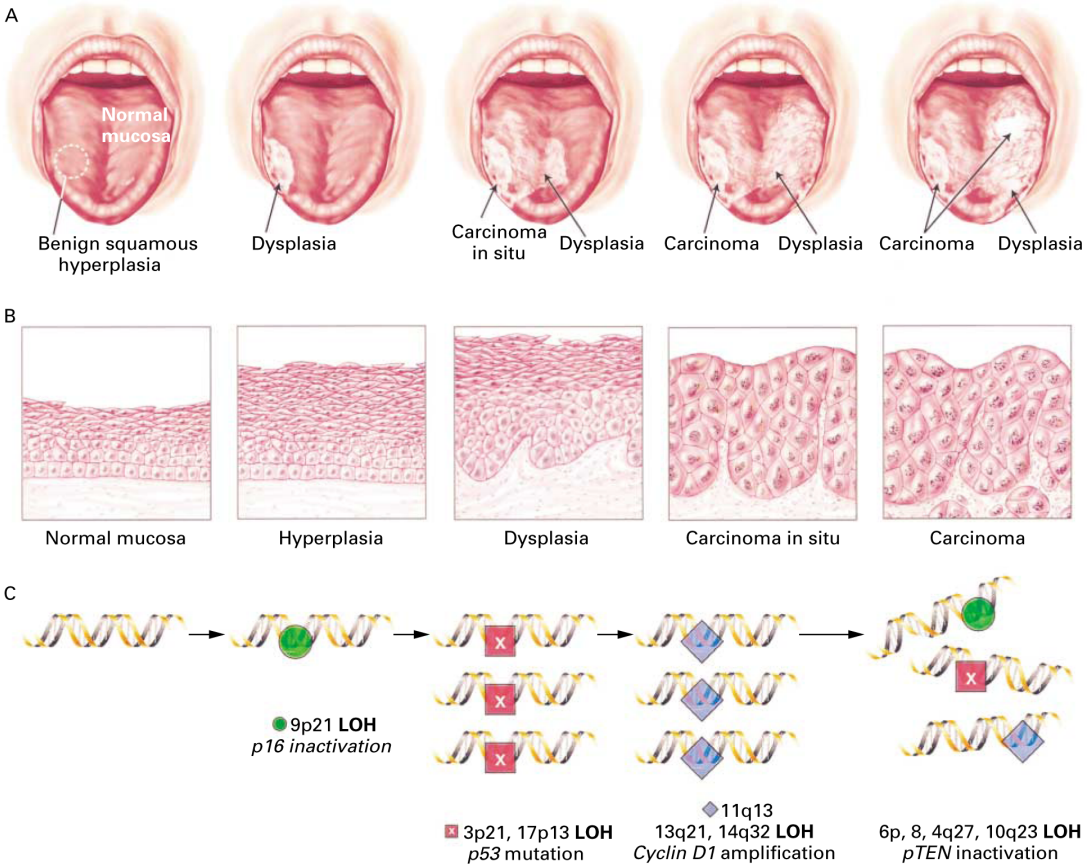


Figure 20: Genetic alterations associated with HNSCC progression
Common genetic alterations that take place in the progression of HNSCC from normal mucosa to carcinoma. Note that 11q13 amplification is a late stage event. Abbreviations: LOH, Loss of heterozygocity; pTEN, Phophatase and tensin homolog. Figure from [180].

The core 11q13 amplicon is 1.5 Mb and contains twelve genes (**Table 12**) [97]. Among these, CCND1 (cyclin D1) and CTTN/EMS1 (cortactin) have been the most studied and are thought to be likely drivers of this amplification due to their consistent correlation of protein overexpression with gene amplification [196,200,201]. One study examined the cases of independent amplifications of CTTN and CCND1 in HSNCC (these cases occur because CCND1 and CTTN are located at opposite ends of the 11q13 amplicon), and found that poor prognosis markers such as decreased survival were correlated with CTTN but not CCND1 amplification [202]. This provides evidence that cortactin is more likely than CCND1 to drive 11q13 amplicon-associated poor prognosis. Additionally, cortactin protein levels are correlated with poor prognosis in HNSCC [203,204]. Mechanistic experimental studies have also shown that cortactin drives tumor aggressive phenotypes in HNSCC (reviewed in [94,205]), including invasiveness. However, additional genes within the amplicon are also overexpressed (**Table 12**) and may contribute to 11q13-driven HNSCC aggressiveness [206,207].

One of these genes, SHANK2, is located on the edge of the 11q13 amplicon directly adjacent to CTTN and is usually reported amplified [206,208], although one study using cell lines reported only some SHANK2 overexpression as a result of amplification [209]; when SHANK2 is gene-amplified, this leads to its protein overexpression [206]. Interestingly, Shank2 is also a cortactin binding partner [93]. However, the significance of this binding interaction and CTTN and SHANK2 co-amplification in HNSCC is unclear. Furthermore, the role of Shank2 in cancer progression of any type has never been studied.

Gene	Description	Function	Protein overexpression?
TPCN2	Two pore segment channel 2	Ion channel	Yes [208]
MYEOV	Myeloma overexpressed gene	Proliferation and invasion	No [208]
CCND1	Cyclin D1	Cell cycle	Yes [97,208,209]
ORAOV1 (TAOS1)	Oral cancer overexpressed 1	Prevention of ROS damage	Yes [97,208,209]
FGF19	Fibroblast growth factor 19	Growth factor in FGF signaling	Yes [208]
FGF4	Fibroblast growth factor 4	Growth factor in FGF signaling	Yes [210] No [208,209]
FGF3	Fibroblast growth factor 3	Growth factor in FGF signaling	No [208,209]
ANO1	Anoctamin 1	Ca^{2+} activated Cl^- channel	Yes [211]
FADD	Fas-associated via death domain	Apoptosis	Yes [208,209]
PPFIA1	Protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF) interacting protein (liprin) a1	Tyrosine phosphatase	Yes [208,209]
CTTN	**Cortactin** (previously EMS1)	**Activates Arp2/3 complex and N-WASp; binds F-actin**	**Yes** [200,202,208,209]
SHANK2	**SH3 and multiple ankyrin repeat domains 2**	**Scaffolding protein**	**Yes** [206]

Table 12: Genes and their functions in the HNSCC 11q13 amplicon core
Genes in the 11q13 amplicon, listed in the order they are located in the amplicon. Protein overexpression as a result of gene amplification is also indicated, along with appropriate references.

Cortactin

Cortactin was originally identified as a cytoskeleton-associated Src kinase substrate [212] and was subsequently named for its localization to cortical actin structures [213]. Cortactin is ubiquitously expressed, with the exception of hematopoietic cells that instead express the cortactin homolog hematopoietic specific 1 (HS1) [214]. Structurally, cortactin consists of a NTA (N-terminal acidic) domain, six and a half tandem repeat domains, a proline-rich domain, and an SH3 domain (**Figure 21**). The N-terminal end of cortactin regulates branched actin assembly through direct binding interactions with the Arp2/3 complex and F-actin [215,216]. The C-terminus, consisting of proline-rich and SH3 domains, is thought to be the regulatory or signaling part of the molecule. Cortactin can be phosphorylated by various kinases in the proline-rich domain and has a number of different cytoskeletal, trafficking, and signaling SH3 binding partners (**Table 13**). The role of the C-terminus in cortactin function is poorly understood.

Cortactin controls many different actin-based cellular processes, such as cell motility, invasion, cell-cell adhesion, and vesicular trafficking [94]. Interestingly, our lab and others have identified a role for cortactin in HNSCC tumor aggressiveness phenotypes. Specifically, cortactin regulates tumor size, invasion, vascularization, invadopodia activity, and protein secretion [147,217–220].

Although interactions with Arp2/3 complex and actin filaments are critical for regulation of cortactin-mediated actin dynamics, the role of additional cortactin-binding partners in cortactin-dependent phenotypes is unclear. In particular, the role of cortactin SH3 interactions and how these contribute to its function in tumor progression is poorly understood. In general, SH3 domains bind to target proteins with proline-rich regions

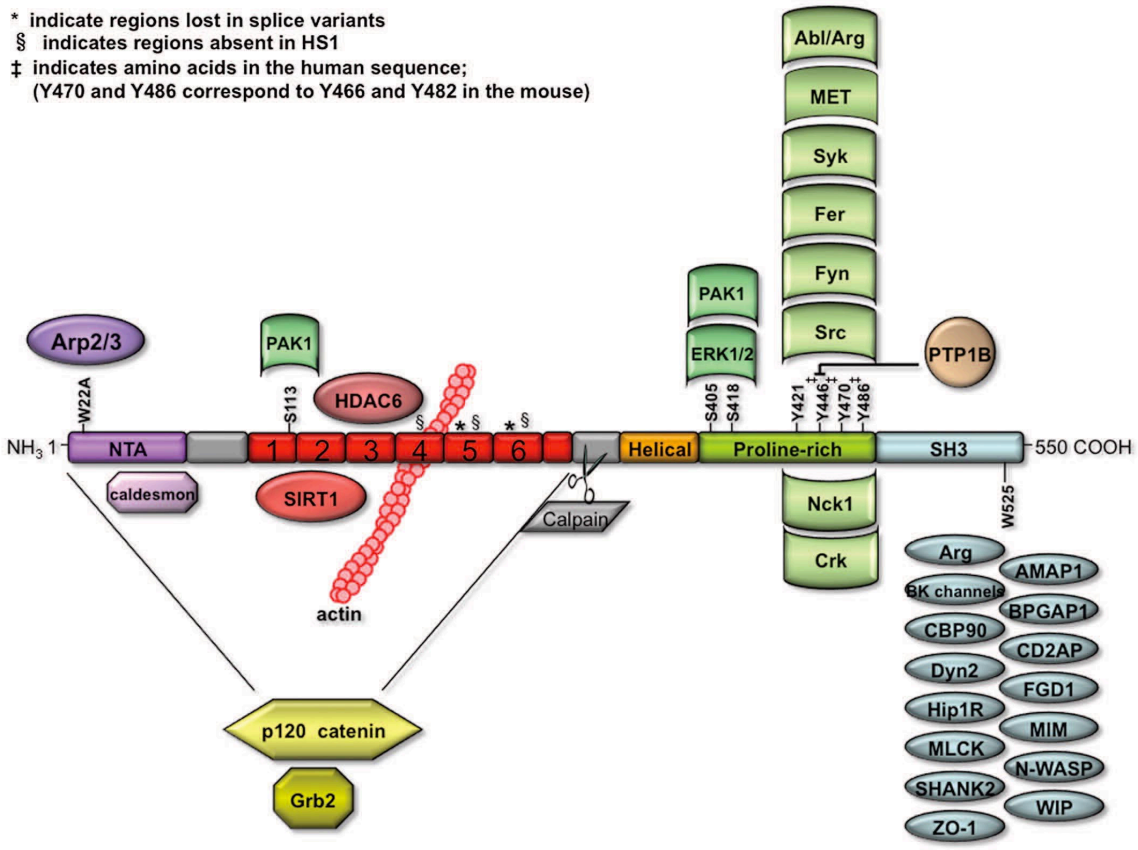


Figure 21: Cortactin domain structure
A schematic of cortactin domains and binding partners is shown. Cortactin consists of four main functional domains: NTA, tandem repeats (1-6), Abbreviations: NTA, N-terminal acidic, SH3, Src homology 3. Figure from [94]

Protein	Description	Function
MIM	Missing in metastasis	Actin assembly
N-WASp	Neural Wiskott–Aldrich Syndrome protein	Actin assembly
WIP	WASP-interacting protein	Actin assembly
CD2AP	CD2-associated protein	Adaptor protein
Shank2	SH3 and multiple ankyrin repeat domains protein 2	Adaptor protein
BK channels	Big potassium channels	Membrane excitability
Dyn2	Dynamin2	Membrane trafficking
Hip1R	Huntingtin interacting protein 1 related	Membrane trafficking
AMAP1/ASAP1	Arf-GTPase activating protein 1	Membrane trafficking; Signal transduction
BPGAP1	RhoA-GAP	Rho-GAP
FGD1	Faciogenital dysplasia protein 1	Rho-GEF
MLCK	Myosin light chain kinase	Signal transduction
ZO-1	Zonus occudens 1	Tight junction adaptor
Abl/Arg	Abl-related gene	Tyrosine kinase
CBP90	Cortactin binding protein 90	Unknown

Table 13: Cortactin SH3 Binding Partners and Their Functions
Proteins that have been shown to bind directly to the cortactin SH3 domain, and their cellular functions. Adapted from [94,205].

containing one or more PXXP motifs; cortactin binds to proteins with positively charged residues surrounding this PXXP motif [221]. A conserved tryptophan residue in SH3 domains [222], found at position 525 in cortactin [93], is crucial for SH3 binding.

Cortactin SH3 binding interactions have been implicated in a number of actin-related cellular phenotypes. *In vitro* actin polymerization assays show that W525K-cortactin has decreased actin bundles compared to WT-cortactin [223]. Overexpression studies show that W525K-cortactin has decreased actin dynamics and membrane protrusion compared to WT-cortactin [224,225]. Smaller growth cones and shorter filopodia are seen when cortactin-KD cells are rescued with either the W525K or a mutant where the entire SH3 domain is deleted [226]. However, Src-phosphorylated cortactin-W525K has been shown to enhance actin polymerization as well as Src-phosphorylated wild-type (WT) cortactin *in vitro* [227]. There are also some contradictory studies on cell migration. Cell migration defects caused by KD of cortactin can be fully rescued using the N-terminus of cortactin containing only the NTA domain and tandem repeats, indicating the SH3 domain is not important for cell migration [98]. However, it has also been shown that the W525K mutant does not rescue migration of mouse embryonic fibroblast (MEF) cells in scratch assays [228]. Our laboratory has recently reported that the W525K mutant does rescue fibrosarcoma random cell migration; however, the migration levels of W525K-cortactin-expressing cells were somewhat decreased (although not statistically significant from WT-cortactin) [218].

Additionally, phosphorylation of cortactin regulates the affinity of SH3 binding proteins. Erk1/2 phosphorylation has been proposed to “liberate” the SH3 domain for binding and increases the ability of cortactin to bind the SH3 binding partner N-WASp [229]. Similarly, p21-activated kinase 1 (Pak1) phosphorylation of cortactin at the same residues (S405 and S418) also increases cortactin binding to N-WASp [230]. Conversely, Src phosphorylation of cortactin inhibits the ability of cortactin to bind N-

WASp with purified proteins [229], although a later study showed the tyrosine phosphorylation-mediated binding of cortactin to N-WASp was dependent on the adaptor protein Nck [227], indicating opposing effects may be seen dependent on which proteins are present. However, Src phosphorylation of cortactin enhances its ability to bind to dynamin-2 [231], indicating this regulation may be specific to different SH3 binding proteins. Importantly, cortactin can be phosphorylated by both Erk1/2 and Src concurrently, and the serine phosphorylation is independent from tyrosine phosphorylation [232].

Taken together, these data suggest that the SH3 domain plays a role in cortactin-dependent cellular phenotypes, but this role may vary depending on which SH3 binding partner is interacting with cortactin and the phosphorylation status of cortactin. However, since only one SH3 binding partner can interact with cortactin at a time, distinguishing which protein is bound to the SH3 domain and its subsequent role in cortactin-associated phenotypes is impossible with experiments using cortactin SH3 domain mutants. Therefore, studying individual SH3 binding partners is important.

.

Protein Trafficking and Secretion

Protein trafficking and secretion are complex processes, and it is well known that they are altered during tumor progression. For example, improper trafficking of the receptor tyrosine kinase (RTK) EGFR through the endocytic system can result in deregulated signaling pathways that can contribute to tumor initiation, promotion, and metastasis [233–235]. Cancer cells secrete a milieu of proteins termed the "cancer secretome", which is being researched for discovery of potential cancer biomarkers [236–239] and control the tumor microenvironment. As one example, secreted vesicles known as exosomes are increasingly being recognized as important for tumor aggressiveness [240–242].

Sorting at the Golgi complex determines the destination of secretory vesicles coming from the biosynthetic pathway; thus, this organelle plays an important role in secretion [243]. Additionally, the Golgi complex has been reported to sense stress and trigger apoptosis [244], and has been proposed as a future target of cancer therapy [245]. Interestingly, it has been reported that morphology of the Golgi complex is collapsed in weakly metastatic cells and extended in highly metastatic cells [246]. Previous studies have shown that a dominant-negative cortactin mutant (without the SH3 domain) blocks vesicle exit from the trans-Golgi network (TGN) [247,248]. Cortactin has also been implicated in other trafficking pathways, such as endocytosis [231,249]. However, although multiple trafficking proteins bind to cortactin through its SH3 domain (**Table 13**), it is unknown what role the cortactin SH3 domain plays in these secretory phenotypes.

Shank2

Shank2 is a scaffolding protein that binds to neurotransmitter receptors, and has primarily been characterized for its role in the nervous system as a core component of the post-synaptic density [250]. Specifically, Shank2 links together the NMDA- and mGluR- receptor complexes through its interactions with GKAP and Homer [84]. This scaffolding role of Shank2 is crucial for proper synaptic functioning, and mutations in SHANK2 have recently been linked to autism spectrum disorder [85–88]. However, Shank2 is also expressed in tissues outside the brain, including kidney and liver [89,90], and was also reported to regulate salt and water transport in epithelial cells by binding to Na-H exchanger 3 [91]. Shank2 also regulates calcium signaling through an interaction with PLC-β3 [92].

Shank2 has three reported isoforms (**Figure 22**), the longest of which is known as Shank2e and consists of ankyrin repeats, an SH3 domain, PDZ domain, a region of proline-rich clusters, and a C-terminal SAM domain [90]. Two shorter Shank2 proteins, ProSAP1a and CortBP1, are produced by alternative start sites and start respectively with the SH3 and PDZ domains [93,251]. Cortactin binds to all three Shank2 isoforms to the PPI motif in the proline-rich region; in neurons, this interaction helps organize the post-synaptic density [93,252].

In addition to Shank2, there are two other Shank family proteins: Shank1 and Shank3. Their domain structure is similar to Shank2e, although like Shank2, there are multiple alternative start, stop, and splicing sites that produce shorter isoforms [89]. Both Shank2 and Shank3 bind to cortactin, but Shank1 lacks the PPI motif necessary for this interaction [93,253]. Shank1 is found exclusively in the brain, while Shank3 is expressed in the heart, brain, and spleen [89]. Additionally, the proline-rich domains of Shank1, 2, and 3 have much less sequence homology than the SH3, PDZ, and SAM

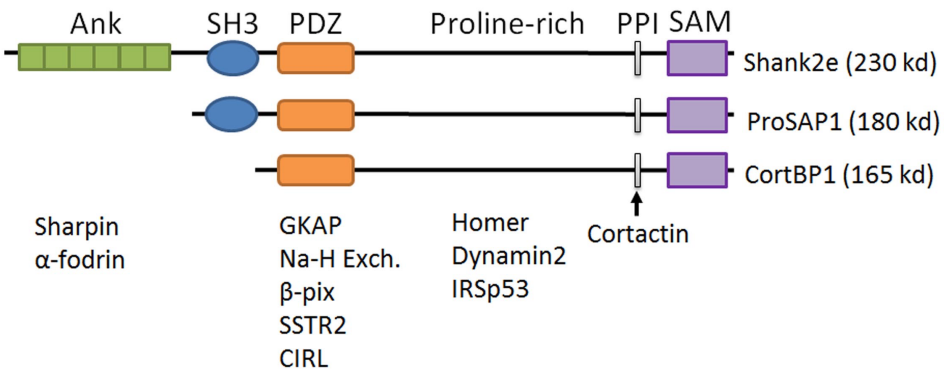


Figure 22: Shank2 Isoforms and Domain Structure
Schematic showing Shank2 domain structure, isoforms, and binding partners for each domain. Abbreviations: Ank, Ankyrin repeats; SH3, Src homolog 3, PDZ, PSD95/Dlg1/Zo-1; PPI, Polyproline I; SAM, Sterile alpha motif.

domains [89]. Although all three Shank family members have been implicated in autism spectrum disorder [254], the differential expression patterns and distinct proline-rich domains suggest unique functions of the Shank2 family protein members in tissues other than the brain.

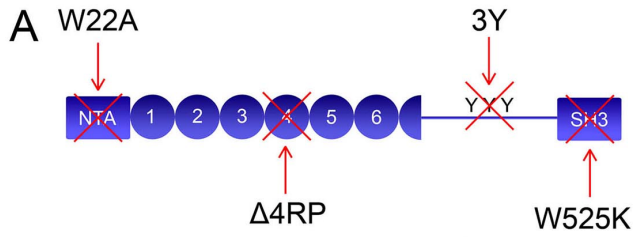
Although Shank2 is co-amplified with cortactin in the 11q13 amplicon, the functional significance of this is unclear and the role of Shank2 in cancer has not been investigated. It is possible that Shank2 is merely amplified as a result of its gene location, and has no role in driving this amplification. However, we previously identified Shank2 in a random walk of our invadopodia network [31], suggesting that Shank2 may play a functional role in HNSCC. Interestingly, Shank2 also binds to sharpin, a protein with a newly identified role in tumor progression and metastasis by inhibition of β1 integrin and activation of the nuclear factor κB pathway [255–258].

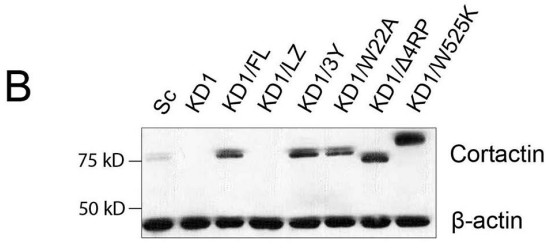
RESULTS

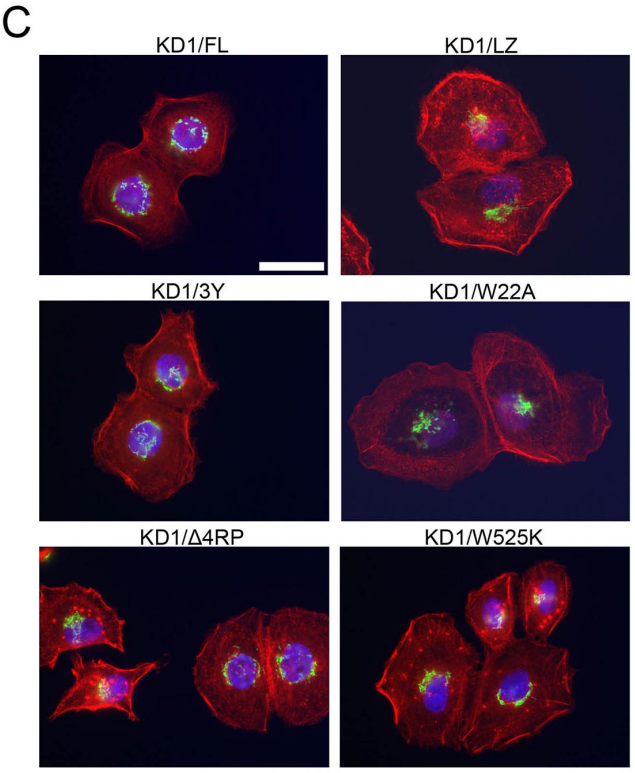
Cortactin SH3 Domain is Crucial for Cellular and Tumor Phenotypes

In this study, we investigated the role of cortactin in regulating Golgi size, morphology, and trafficking. Surprisingly, we found that cortactin regulates late endosomal/lysosomal trafficking, indirectly controlling Golgi morphology [259]. The knockdown of cortactin is already known to decrease tumor aggressiveness [147], and cortactin-KD cells exhibited dramatically decreased Golgi area similar to the Golgi phenotype reported previously in less aggressive cell lines [246]. Interestingly, rescue of the Golgi phenotype with cortactin W525K mutants showed a variable phenotype, with some cells displaying a collapsed Golgi complex while others were dispersed (**Figure 23**). We suggest this may due to a dominant-negative effect [259].

While the effect of the cortactin SH3 domain mutant has been characterized in several *in vitro* phenotypes, such as actin dynamics, cell migration, FN secretion, and Golgi morphology [218,223–226,228,259,260], its effect on *in vivo* tumor growth has never been reported. We used a semi-orthotropic HNSCC model developed by the Yarbrough lab where HNSCC cells are placed inside a denuded rat trachea and implanted into the flank of a nude mouse [106]. Our laboratory has previously used this model to report that HNSCC tumor size and aggressiveness correlates with expression of cortactin [147]. It is a good system to model tumor growth and invasion, by measuring both total tumor size and cells that have invaded through the tracheal ring [147]. To determine if cortactin's SH3 domain is important for tumor growth in this model, we expressed either full-length (FL)-cortactin or the cortactin SH3 mutant W525K in both WT and cortactin-KD tumors for overexpression and rescue studies, respectively.


A
W22A
3Y
NTA
1
2
3
4
5
6
Y Y Y
SH3
Δ4RP
W525K


B
Sc
KD1
KD1/FL
KD1/LZ
KD1/3Y
KD1/W22A
KD1/Δ4RP
KD1/W525K
75 kD
50 kD
Cortactin
β-actin


C
KD1/FL
KD1/LZ
KD1/3Y
KD1/W22A
KD1/Δ4RP
KD1/W525K

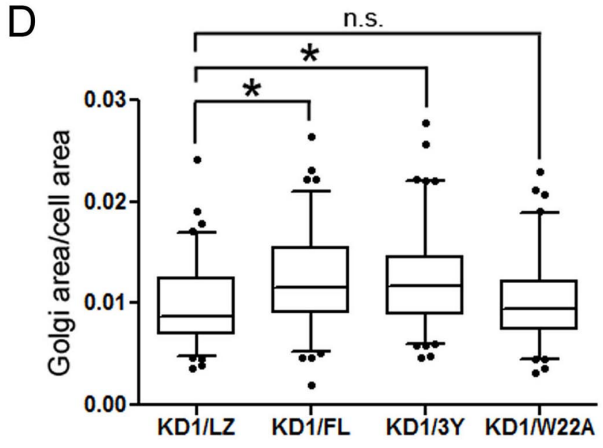

D
n.s.
*
*
Golgi area/cell area
0.03
0.02
0.01
0.00
KD1/LZ
KD1/FL
KD1/3Y
KD1/W22A

Figure 23: Cortactin SH3 Domain Mutant Partially Rescues Golgi Morphology
Figure on previous page. **A**) Schematic of cortactin with mutant sites indicated: W22A (Arp2/3 binding mutant); 3Y (Src phosphorylation mutant); Δ4RP (F-actin binding mutant); and W525K (SH3 binding mutant). NTA represents the N-terminal acidic domain, YYY represents Src phosphorylation sites, and circles represent cortactin repeat domains. For **B**–**D**: FL = wild-type cortactin, LZ = empty vector, Sc = scrambled oligo. Others as indicated in (**A**) and (**B**) Western blot analysis of cortactin mutant expression (top panel) and actin as a loading control (bottom panel) in stable SCC61 cells. (C) Representative widefield images of the Golgi (GM130, green), actin (rhodamine-phalloidin, red), and Hoechst (blue) in the cortactin mutants. (D) Quantitation of Golgi area to cell area ratio from $n = 3$; > 20 cells per independent experiment. Scale bar = 25 μm. Data are presented as box and whiskers plots with the box indicating the 25th and 75th percentiles, solid line indicating the median, and the whiskers indicating the 95% confidence intervals. *$P < 0.05$. From [259]

Interestingly, expression of cortactin FL in cortactin-KD cells resulted in large and invasive tumors, while tumors grown from the cortactin W525K mutant were a similar size as cortactin-KD cells (**Figure 24**). Similar results were seen in the overexpression studies, although without any invasion; however, these results were insignificant and more replicates need to be completed. These data suggest that the cortactin SH3 domain may be critical for *in vivo* tumor size, although it is not known which SH3 binding partners contribute to this phenotype.

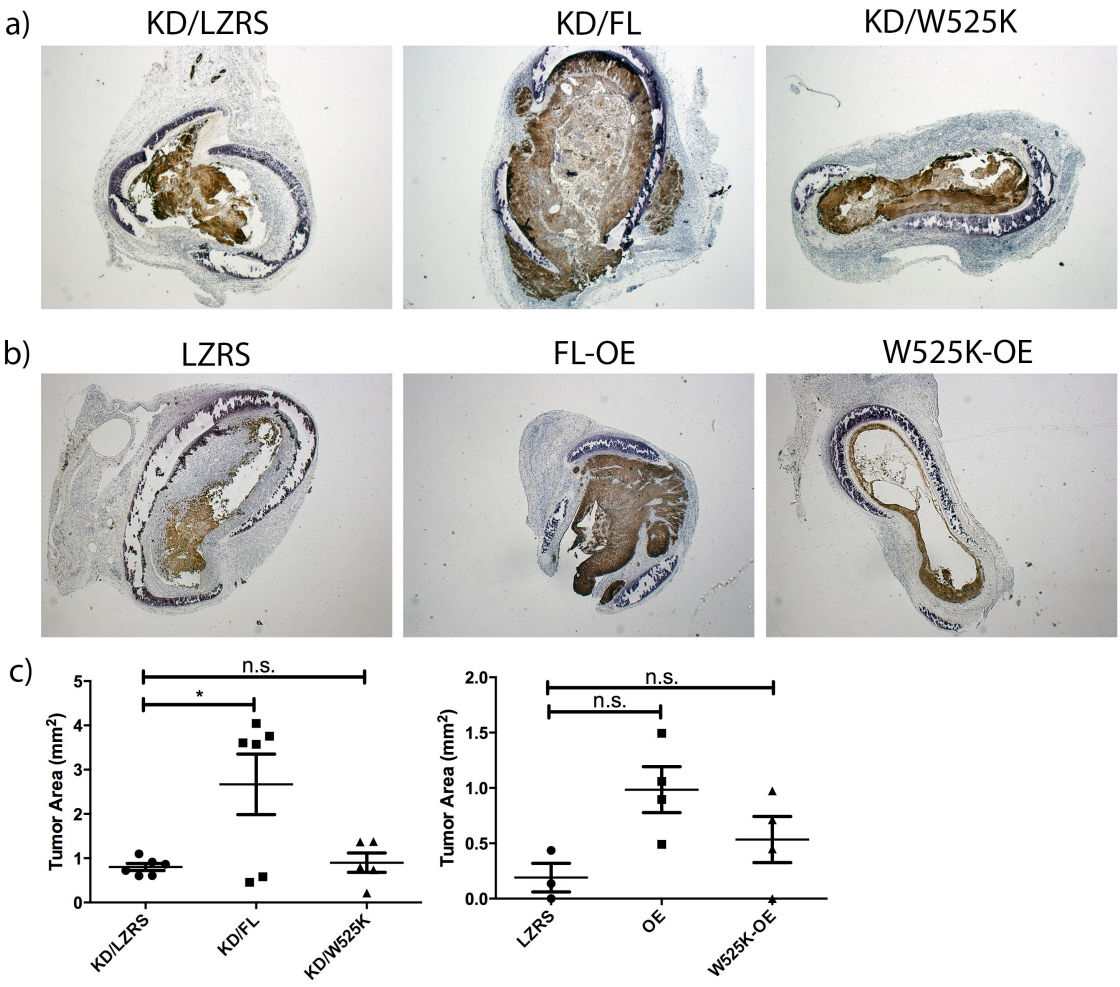


Figure 24: Cortactin SH3 Domain is Crucial for *In Vivo* Tumor Growth

a) Representative images from cytokeratin immunohistochemical staining of tumors grown with cortactin-rescued cells lines. Cortactin FL or W525K mutant were re-expressed in SCC61-cortactin-KD cells, and LZRS vector was used as a control. Note the invasion outside of the tracheal ring in the cortactin full-length rescue (KD/FL) tumor. Three independent replicates with one mouse each were completed (n=3 mice; 6 tracheas). **b)** Representative images from cytokeratin immunohistochemical staining of tumors grown with cortactin-overexpressed cells lines. Cortactin FL or W525K mutant were overexpressed in SCC61 cells, and LZRS vector was used as a control. One independent replicates of two mice each were completed (n=2 mice; 4 tracheas). **c)** Tumor area measured from cytokeratin staining to identify tumor epithelial cells. Cytokeratin-positive tumor area is brown. Abbreviations: KD, knockdown; LZRS, retroviral vector used to create cell lines; FL, full-length; OE, overexpressed; W525K, Cortactin SH3 binding mutant; n.s., not significant. * $p<0.05$

Characterization of Shank2 Expression in HNSCC Cell Lines and Patients

Because Shank2 has not been studied in the context of cancer, we wanted to characterize its expression in HNSCC. We found that Shank2e was expressed in five out of eight HNSCC cell lines (**Figure 25a**). Additionally, high Shank2e expression was correlated to 11q13 amplification; FaDu, Detroit 562, and SCC25 are known to have 11q13 amplification [147,261]. CortBP1 was also expressed, but to a lesser degree, and no significant ProSAP1a expressed was observed (**Figure 25a**). We also analyzed paired normal and tumor samples (**Table 14**) from two Stage IV HPV-negative HNSCC patients and saw that Shank2e (quantitated) and CortBP1 were increased in tumor tissue as compared to normal tissue (**Figure 25b**). This data indicates that Shank2, primarily the epithelial isoform Shank2e, is expressed in HNSCC and is upregulated in tumors.

Additionally, we checked the localization of Shank2 by immunofluorescence (IF) using a commercial polyclonal antibody. Shank2 localizes both to the cytoplasm and co-localizes with actin at membrane ruffles (**Figure 26a**); similarly, Shank2e has previously been reported to localize to apical membranes [90]. Shank 2 localization is unaltered in cortactin-manipulated cells (control, KD, and rescue), indicating that Shank2 is not recruited to membrane ruffles by cortactin (**Figure 26a**). Shank2 binds to dynamin2 [252], and is also endocytosed with sodium-phosphate cotransporter IIa (NaPiIIa) and regulates its apical retention and intracellular distribution [262,263], suggesting a possible role in membrane trafficking. Cortactin has recently been shown to play a role in the LE/lysosomal maturation and trafficking and cortactin-KD leads to enlargement of LE [218,259,264]. Based on these previous studies, we hypothesized that Shank2 may also participate in this process. We do see some possible Shank2 colocalization with

the late endosomal marker CD63 (**Figure 26b**), indicating that it is positioned to play a role in LE trafficking, although further studies need to be done to ensure this is not the result of cytoplasmic overlap. Shank2 cytoplasmic localization is unaltered in cortactin-manipulated cells (**Figure 26b**), indicating that cortactin expression does not play a role in Shank2 localization.

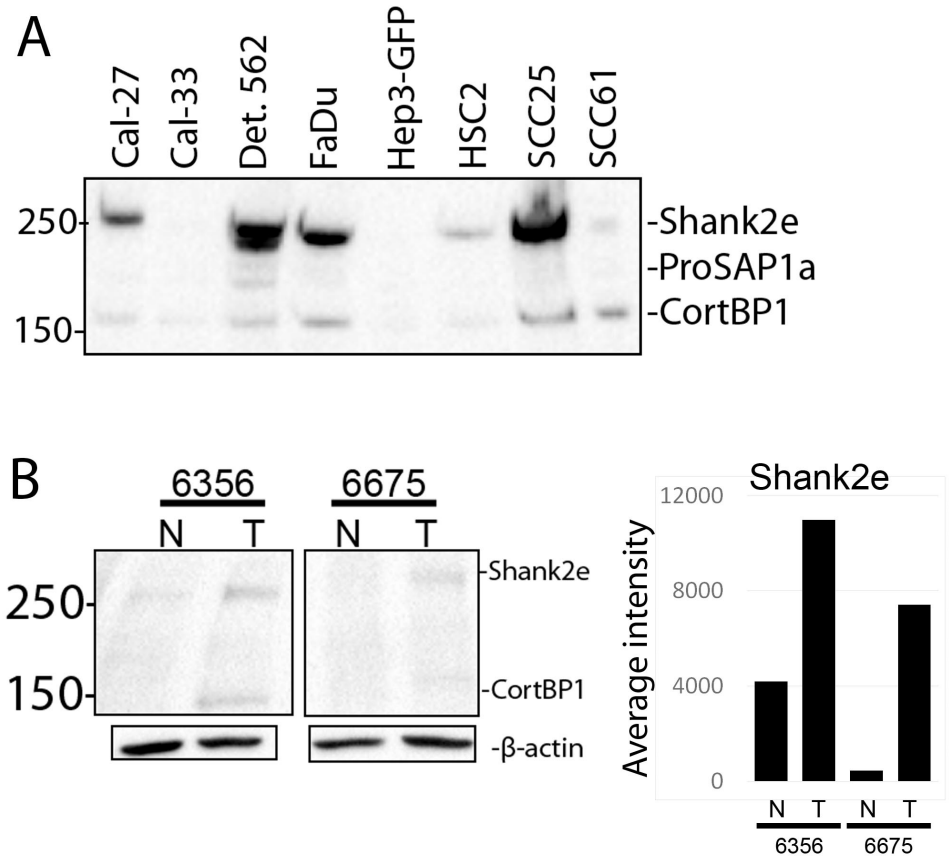


Figure 25: Shank2 is Expressed in HNSCC and Upregulated in Tumor Compared to Normal Tissue

a) A panel of HNSCC cell lines stained with a polyclonal Shank2 antibody. Three forms of Shank2 were identified in cell lines: Shank2e (230 kDa), ProSAP1a (180 kDa), and CortBP1 (165 kDa). Shank2e was the most highly expressed isoform, and expression was increased in 11q13 amplified cell lines (Detroit 562, FaDu, and SCC25). **b)** Paired normal (N) and tumor (T) tissue lysates from patients with Stage IV HPV-negative HNSCC stained with polyclonal Shank2 antibody. Faint expression of both Shank2e and CortBP1 is seen, and is upregulated in tumor tissue compared to normal. The graph shows the quantitation of Shank2e bands.

ID	*6356*	*6675*
Site	Larynx	Larynx
Sex	M	M
Age at dx	67	53
Race	W	W
Smoking history (Pack Years)	50	2 packs per day
Pathology staging	T4aN3	T3N2b
	T4N2bM0	T3N2b
Clinical staging	IVB	IVA
Treatment before collection	No treatment	No treatment
Treatment after collection	Surgery, concurrent chemo/radiation	Surgery/radiation
Months from dx to last follow-up	4	2
Months from dx to recurrence	N/A	N/A
Current status	LWD	LWD

Table 14: Clinical Attributes of Patients Tested for Shank2 Expression
Patient information is from the Vanderbilt-Barry Baker Head and Neck Repository. Paired tumor and normal tissue from these patients was used to assess Shank2 expression (**Figure 25**). Abbreviations: dx, diagnosis; M, male; W, white.

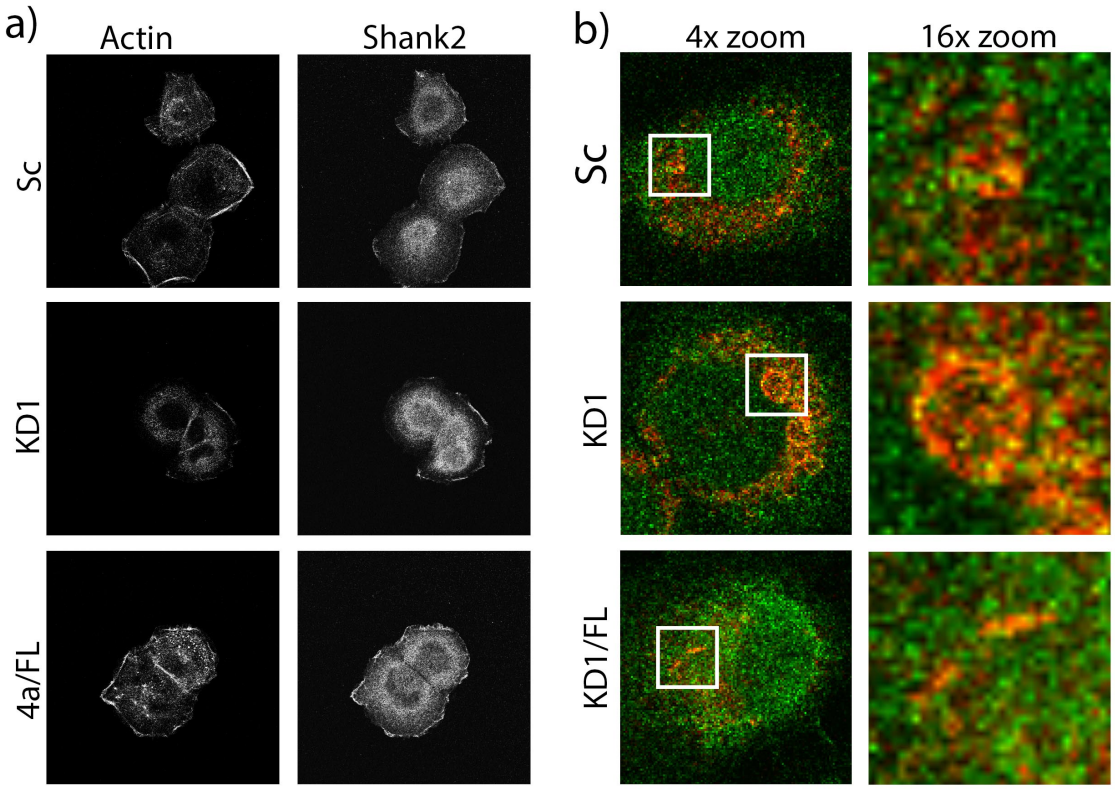


Figure 26: Shank2 localizes to cell ruffles and cytoplasm in HNSCC cells
Representative confocal microscopy images (.2 μM slices) are shown from stained SCC61 cells. **a)** Shank2 staining (right) localizes to the cytoplasm and membrane ruffles (marked by actin staining, left). **b)** Shank2 (green) partially colocalizes with late endosomes (red, CD63). Yellow indicates colocalization.

Analysis of TCGA Patients Suggests Shank2 is an Independent Predictor of Tumor Aggressiveness

One possibility is that the Shank2 protein is merely overexpressed as a result of 11q13 amplification of the SHANK2 gene, but has no effect on tumor growth. To determine if SHANK2 expression is an independent predictor of tumor aggressiveness, we analyzed TCGA RNAseq data. We first graphed SHANK2 vs. CTTN expression (**Figure 27a**) to determine if there was a correlation. We do see a slight correlation between these; however, there were some tumors with high SHANK2 and low CTTN or vice versa, allowing us to assess whether SHANK2 correlates with tumor phenotypes independent of CTTN expression. We observed a cluster of patients with both low SHANK2 and CTTN expression; thus, we set the cutoff for "high" and "low" expression based on this observation (**Figure 27a**, red lines indicate cutoff points). This divided our patients into four groups: CTTN and SHANK2 high, CTTN high and SHANK2 low, CTTN low and SHANK2 high, and CTTEN and SHANK2 high. We then looked at the stage, tumor size, and lymph node status of these four groups of patients. In CTTN-high expressing patients, SHANK-high expressing patients display decreased aggressive indicators (Stage IV, N2-3, T3-4) and increased less aggressive indicators (Stages I-III, N1-2, T1-2) when compared to SHANK2-low expressing patients. Similar results were seen in CTTN-low expressing patients, indicating that SHANK2 may correlate with tumor aggressiveness independent of CTTN expression (**Figure 27b-d**). Unfortunately, there were not enough patients available to do a statistical analysis, or to look at vital/recurrence status. Therefore, these data are suggestive that Shank2 may play a significant role in tumor progression independent of 11q13 amplification measured by CTTN expression, but we are unable to draw any definitive conclusions at this time.

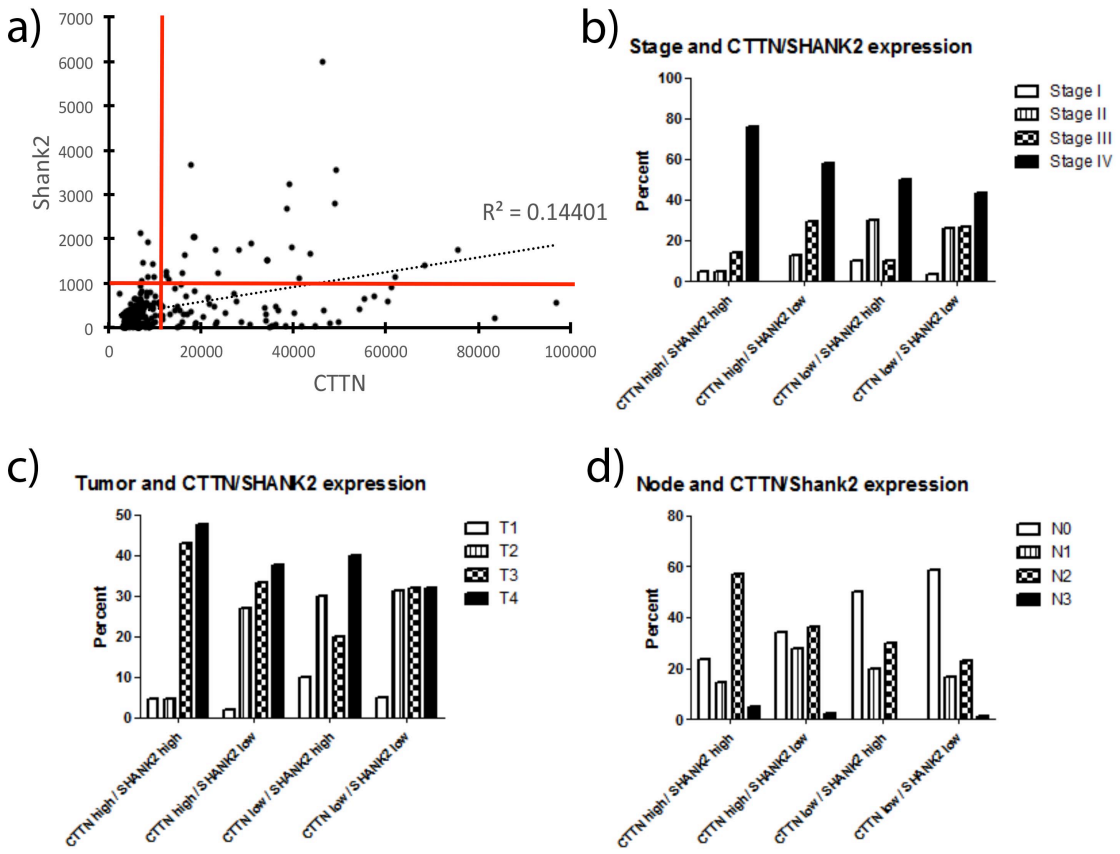


Figure 27: TCGA RNAseq Data Shows Increased Tumor Aggressiveness with High Shank2 Expression

Data from TCGA [149]. **a)** CTTN vs. SHANK2 RNAseq correlation for all tumor types except hematopoietic malignancies. Each data point indicates a patient. Red lines indicate cutoff points for "low" vs. "high" RNAseq expression. **b)** Analysis of tumor stage and CTTN/SHANK2 high/low groups. SHANK2 high expressing patients have a higher percentage of aggressive Stage III-IV tumors, in both CTTN-low and CTTN-high groups. c) Analysis of tumor size and CTTN/SHANK2 high/low groups. SHANK2 high expressing patients have a higher percentage of aggressive T4 tumors, in both CTTN-low and CTTN-high groups. d) Analysis of lymph node status and CTTN/SHANK2 high/low groups. SHANK2 high expressing patients have a higher percentage of more aggressive N2 tumors and a lower percentage of non-aggressive N0 tumors, in both CTTN-low and CTTN-high groups. Percentage of patients was used instead of total patient number due to the differing numbers of patients in each group.

Shank2 Knockdown Affects Tumor Phenotypes

To determine the role of Shank2 in tumor aggressiveness, we knocked down Shank2 in one 11q13-amplified cell line (FaDu) and one 11q13-non-amplified cell line (Cal-27) cells using shRNA to the 3'UTR [261,265] (**Figure 28a**). One important hallmark of cancer is invasion and metastasis [4]. To determine if Shank2 controls HNSCC invasiveness, we tested the ability of these cells to invade through Matrigel-coated Transwell filters. The number of cells able to invade was significantly reduced in FaDu Shank2-KD cells compared to control cells; a similar trend is seen in Cal27-KD cells, but more replicates are needed to reach statistical significance (**Figure 28b**). This indicates that Shank2 does play a specific role in HNSCC invasiveness, independent of 11q13 amplification status.

Invadopodia are subcellular structures that secrete proteases such as MMPs to degrade ECM, giving cancer cells the ability to move through tissues in the body; they are formed preferentially by aggressive cancer cells and invadopodia activity is used to identify aggressive cancer cells [266]. To determine if Shank2 affects HSNCC invasiveness by specifically controlling invadopodia activity, we tested the ability of Shank2-KD cells to degrade a fluorescent bed of ECM. Shank2-KD cells exhibit decreased invadopodia activity compared to control cells, seen by a decrease in degradation area of the FN matrix (dark holes) (**Figure 29**). This result is significant for KD3 (the more robust KD, see **Figure 28a**) but not KD2; more data needs to be collected to reach statistical significance for this less robust KD.

A decrease in invadopodia activity can be a result of either a decrease in the number of invadopodia formed, or a defect in secretion of MMPs at invadopodia. To test whether the effect of Shank2 manipulation on invadopodia activity was due to a

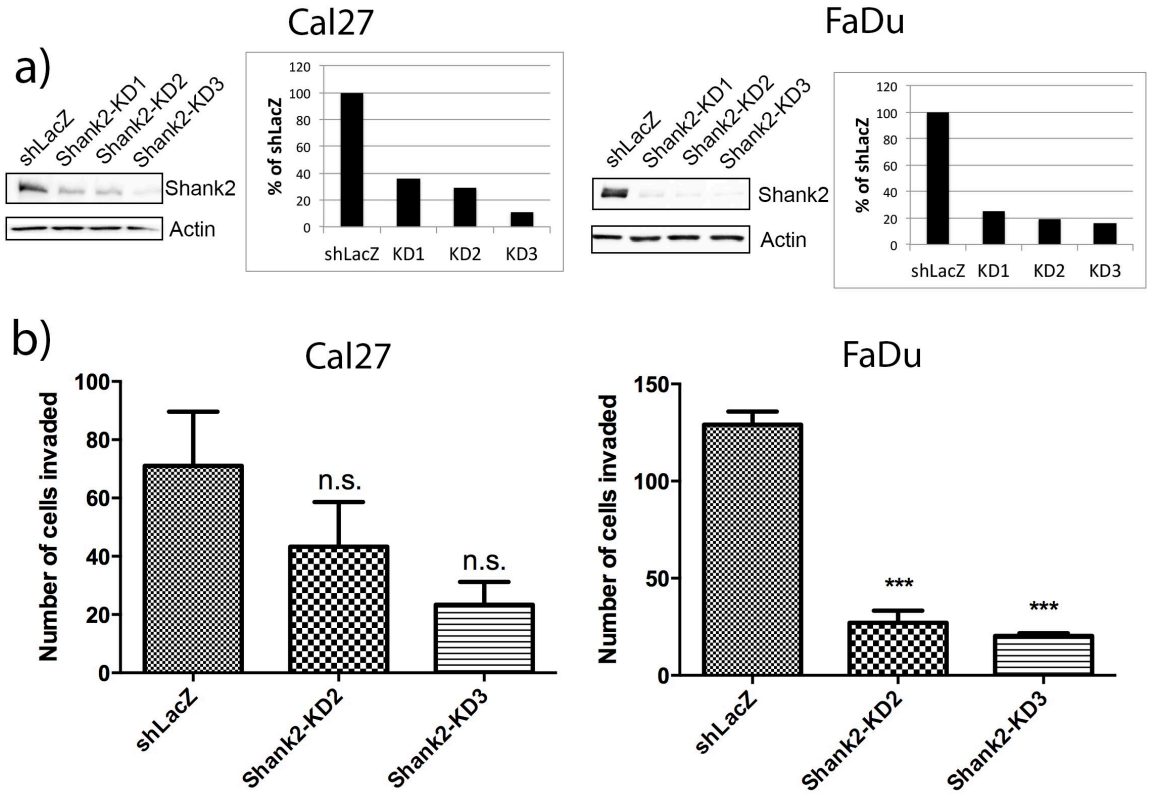


Figure 28: KD of Shank2 Decreases Transwell Invasion

a) Western blots of cell lysates from shLacZ control and three separate knockdowns (KD) constructs of Shank2. β-actin is shown as a loading control. Graphs show the quantitation of Shank2 KD as a percentage to the shLacZ control. N=1 is quantitated. **b)** A quantitation of the number of cells invaded across a Matrigel-coated Transwell filter after 48 hours. N=2 replicates; experiments were plated in duplicate.

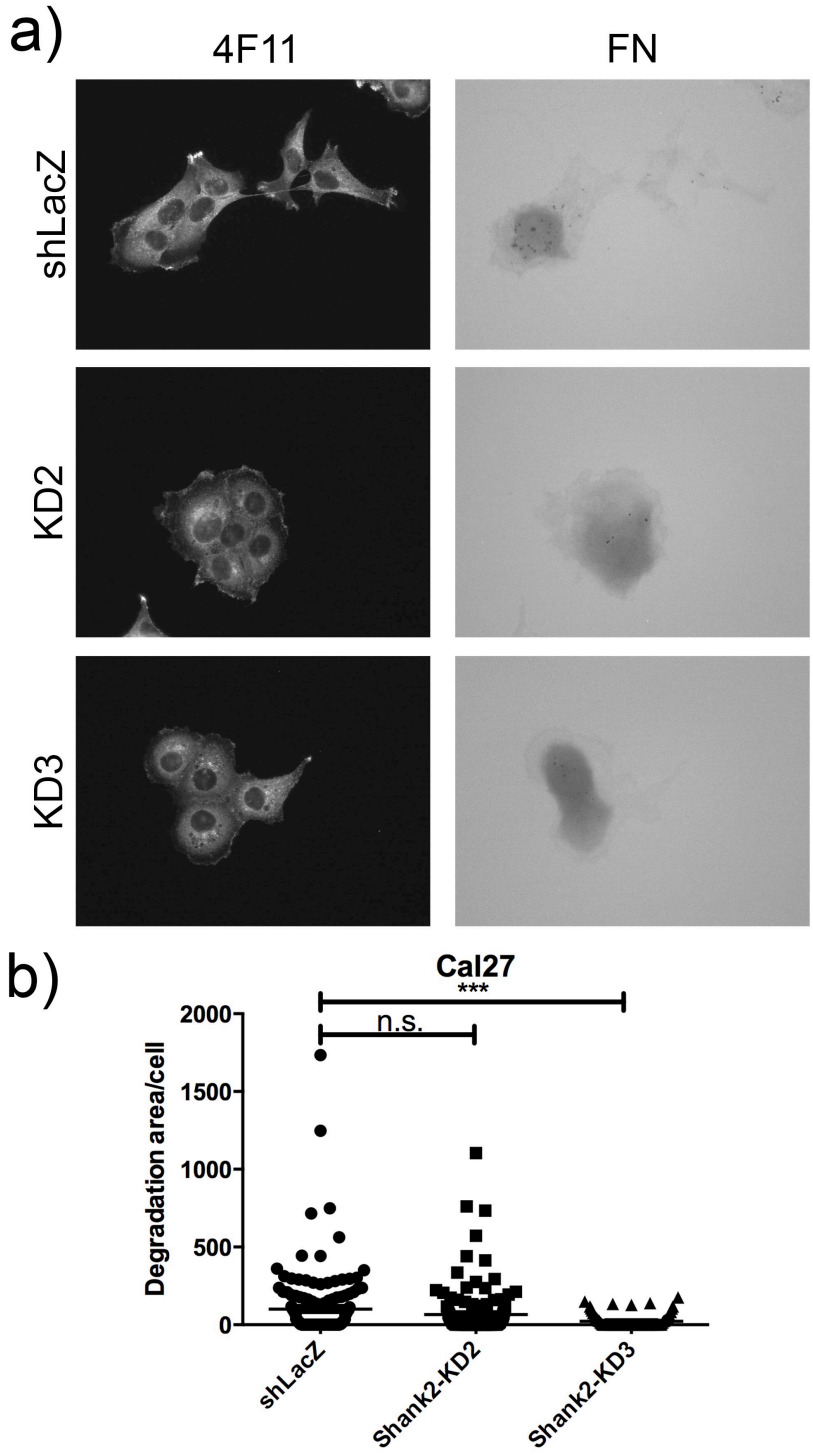


Figure 29: KD of Shank2 Decreases Invadopodia Activity

a) Representative images of control (shLacZ) and Shank2-knockdown (KD) cells stained with actin (4F11, left) and FITC-labeled fibronectin matrix (FN, right). Some shadows from cells settling on the FN matrix are seen in all groups. However, invadopodia activity can still be visualized as dark puncta, and their area was quantitated. b) The quantitation of degradation area normalized to cell area. ***p<0.001; n.s., not significant

decrease in MMP levels, we tested the ability of these cells to secrete MMPs. Conditioned media from Shank2-manipulated FaDu cells was run on a non-reducing gel containing an MMP substrate (gelatin), and incubated overnight to allow degradation of the gelatin by MMPs. Interestingly, we see a decrease in MMP9 and active MMP2, but similar levels of pro-MMP2 (**Figure 30**). Taken together, these data indicate that Shank2 controls HNSCC aggressiveness, specifically invasion through invadopodia activity and MMP secretion.

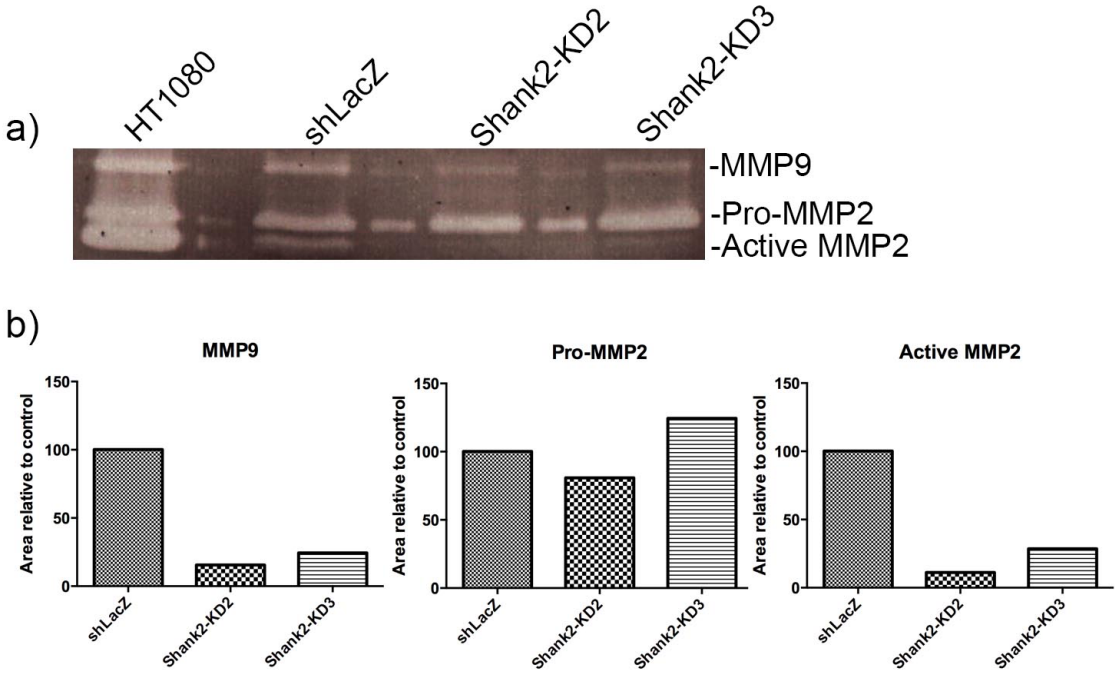


Figure 30: KD of Shank2 Decreases MMP Activity

a) Zymography using conditioned media from control (shLacZ) and Shank2-knockdown (KD) cells showing gelatin degradation of MMPs. Shank2-KD cells exhibit decreased MMP9 and active MMP2 activity. Conditioned media from HT1080 cell line was used as a positive control. b) Quantitation of band intensity for MMP9, Pro-MMP2, and active MMP2. N=1 replicate.

DISCUSSION

In this study, we set out to determine the roles of the cortactin SH3 domain and one of its binding partners Shank2 in tumor aggressiveness. We showed that the cortactin SH3 domain is crucial for the cortactin-dependent phenotypes *in vitro* Golgi morphology and *in vivo* HNSCC tumor size [147,259]. Cortactin with a point mutation abolishing its SH3 binding activity (W525K) did not rescue collapsed Golgi phenotype or *in vivo* tumor growth defects caused by cortactin KD (**Figure 23**, **Figure 24**). These data indicate that that cortactin SH3 domain is important for tumor aggressive phenotypes. However, since cortactin binds to numerous proteins through its SH3 domain (**Table 13**), it is unknown which of these interactions is contributes to these phenotypes. In addition, cortactin has been reported to affect various other tumor-related phenotypes, such as invasion, vascularization, invadopodia formation and activity, protein secretion, and exosomes secretion [147,217–220]. Although cortactin only has the ability to interact with one protein at once through its SH3 domain, it is possible that cortactin plays multiple roles in HNSCC through temporal regulation of SH3 binding. It is also likely that cortactin SH3 interactions may be important in different locations throughout the cell. Indeed, cortactin is found localized at many places throughout the cell, including cell membranes, cell ruffles, lamellipodia, invadopodia, and intercellular vesicles. Many cortactin SH3 binding partners have also been implicated in tumor progression or crucially related processes, such as actin assembly, membrane trafficking, and signaling (**Table 13**).

The cortactin SH3 binding partner Shank2 has never been studied in the context of cancer, although it is overexpressed along with cortactin as a result of 11q13 amplification, which occurs in multiple cancer types [95–97,182,185–187]. In this study,

we identified Shank2 as a novel regulator of HNSCC aggressiveness. Shank2 knockdown resulted in a decrease of cellular invasiveness, invadopodia activity, and MMP secretion (**Figure 28**, **Figure 29**, **Figure 30**). We used two HNSCC cell lines with differing 11q13 amplification status: FaDu cells are highly 11q13 amplified while Cal-27 cells are not amplified. Shank2 regulated these phenotypes regardless of 11q13 status, although higher Shank2 expression was seen in 11q13 amplified cells. Previously, it was unclear whether Shank2 overexpression is simply a result of 11q13 amplification, or whether Shank2 function may help drive this amplification. It is currently unknown which genes are responsible for 11q13 amplification, although it is assumed to be either CTTN or the cell cycle control protein CCND1 [196,200,201]. The breakage-fusion-bridge cycle that amplifies 11q13 can result in cases where only one end of the core amplicon is amplified. Interestingly, an analysis of this in HNSCC shows that poor prognosis was associated with CTTN amplification (located on one end of the amplicon) but not with CCND1 amplification (located on the other end). Since the SHANK2 gene is found immediately adjacent to the CTTN gene, this is consistent with our results that identify Shank2 as an active regulator of tumor aggressiveness. Additionally, our preliminary results suggest that tumors with high SHANK2 expression are more aggressive than those with low SHANK2 expression, regardless of CTTN levels (**Figure 27**).

The mechanism of how Shank2 contributes to HNSCC aggressiveness, and whether or not the interaction between cortactin and Shank2 plays a role, is currently unknown. One possibility is that similar to its role in the post-synaptic density, Shank2 serves as a scaffolding protein. This could occur at the cell edge and/or in the cytoplasm (potentially at late endosomal vesicles), two locations where we observed Shank2 localization (**Figure 26**). It is unknown which proteins Shank2 would link together in HNSCC cells. In the PSD, Shank2 scaffolds together two separate receptor complexes (NMDA- and mGluR-) through interactions with GKAP and Homer; deletion of

Shank2 in mice results in defective synaptic signaling leading to autistic behaviors [85–88]. Interestingly, cortactin, Shank2, and dynamin2 are all direct binding partners [93,252,267]. Additionally, Shank2 binds to α-fodrin, another cytoskeletal protein that interacts with f-actin [268]. These interactions suggest that Shank2 may link signaling molecules to the actin cytoskeleton. Another possible scaffolding role of Shank2 is linking the exocyst complex protein Sec8 to cortactin through its binding to PSD-95 [253,269]. The exocyst complex is important for numerous cellular functions, such as cytokinesis, exocytosis, endocytic recycling, autophagosome biogenesis, cell migration, and invadopodia [270]. It is feasible that Shank2 and cortactin could work together to promote cell migration, invadopodia formation and activity, or trafficking.

Additionally, we previously performed a random walk analysis on a manually curated list of core invadopodia machinery to identify new molecules, or hubs, associated with invadopodia. Interestingly, Shank2 was significantly associated with this invadopodia network [31]. This is consistent with our results that Shank2 expression affects invadopodia activity and MMP secretion (**Figure 29**, **Figure 30**). It is likely that Shank2 controls invadopodia activity by regulating upstream trafficking of MMPs, resulting in decreased MMP secretion. Whether or not this trafficking defect MMP-specific or is more general remains to be determined.

FUTURE DIRECTIONS

It is interesting to note that the genes CTTN and SHANK2 are located immediately adjacent on the 11q13 amplicon; it is possible that their overexpression may have a synergistic effect on HNSCC tumor progression. To test that hypothesis, cell lines manipulated with high levels of cortactin and Shank2, both separately and together, could be used for tumor growth or other phenotypic experiments. Cortactin is the only known binding partner of the PPI motif in Shank2; therefore, a Shank2 PPI binding mutant that is unable to bind cortactin could be used to determine whether any observed effects were due to the Shank2-cortactin binding interaction. Re-expression of this Shank2 PPI mutant in Shank2-KD cells would be a crucial rescue experiment for the phenotypes tested in this study (transwell invasion, invadopodia activity, and MMP secretion), as well as *in vivo* tumor growth and exosomes secretion. Testing of additional phenotypes would determine if the mechanistic role of Shank2 in HSNCC progression is trafficking-specific or more broad.

Another interesting future direction would be use of Shank2 KD cells to test its role on vesicular trafficking. One way Shank2 could regulate trafficking is through its interaction with dynamin2. Dynamin2 controls vesicle scission at the cell membrane and other organelles; however, its mechanism is poorly understood [271,272]. We showed that Shank2 regulates MMP secretion (**Figure 30**) and matrix degradation by invadopodia (**Figure 29**); dynamin2 is also necessary for focal adhesion matrix degradation [273]. It is feasible that Shank2 could act with dynamin2 at or upstream of invadopodia to regulate matrix degradation or MMP secretion at invadopodia. The interaction of Shank2 with cortactin could also affect trafficking and secretion, since cortactin has a reported role in LE/lysosomal trafficking and secretion of FN [218,259].

More work is needed to see if this trafficking defect is general, or specific to MMP secretion. Dynamin2 and cortactin bind to different residues of Shank2, meaning it is possible for Shank2 to bind to both cortactin and dynamin2 simultaneously, but whether or not this occurs and what effect it would have is unknown. Competitive binding assays could be performed to determine if Shank2 and dynamin2 compete for binding to the cortactin SH3 domain, or if cortactin and Shank2 compete for binding to the dynamin2 proline-rich domain (PRD). This competition could regulate dynamics on vesicles or at the cell surface.

It would also be interesting to test the effect of Shank2 on progression of tumors from other tissue types. 11q13 amplification is the most common in HNSCC, but is also found in other cancers, such as breast, esophageal, pancreatic, ovarian, lung, and bladder [95–97,182–188]. Although the overexpression of Shank2 is likely driven by 11q13 amplification in HNSCC and possibly other cancers, our results show that the Shank2 promotes tumor-associated phenotypes regardless of 11q13 amplification status. Therefore, Shank2 could also play a role in non-11q13 amplified cells.

More replicates of the following experiments are needed to achieve statistical significance or a greater level of confidence in our results: tumor growth with the cortactin W525K mutant (OE cell lines only), expression of Shank2e in paired tumor and normal samples, and invadopodia assay and zymography with Shank2-KD cell lines. Both of the tumors included in the paired tumor and normal tissue Western blots that showed increased expression of Shank2e (and CortBP1) are from Stage IV patients (**Figure 25**, **Table 14**). These patients were selected because 11q13 amplification is thought to be a late stage in HNSCC progression [180]. However, inclusion of patients in varying stages, as well as a comparison of cortactin and/or cyclin D1 overexpression, would give a clearer picture of when Shank2 overexpression occurs in HNSCC and whether or not this is concurrent and comparable with overexpression of other markers

in the 11q13 amplicon. For tumor growth, it is unusual that the OE tumors were smaller than the KD/rescue, which should have less expression of cortactin (**Figure 24**). Since only one independent experiment was performed for the OE tumors, this may simply be due to inherent variability in the experiment.

Additionally, MMP secretion should be tested in non-11q13 amplified Shank2-KD cells (Cal27), and invadopodia activity should be tested in 11q13-amplified Shank2-KD cells (FaDu). Interestingly, even though FaDu cells are highly amplified and very aggressive in vivo [147], they do not form invadopodia nor degrade matrix under our usual experimental conditions. Therefore, we must either use a different 11q13-amplified cell line, or determine a way to induce invadopodia formation in FaDu cells. We hypothesized that hypoxia, a known promoter of tumor progression [274] that has been reported to induce invadopodia formation in other cell types [275], may also stimulate invadopodia formation in FaDu cells. We saw no effect after 24 hours in hypoxic conditions, but after 48 hours we do see significantly increased invadopodia activity of FaDu cells (**Figure 31**). We see a similar effect in SCC61 cells, which are well characterized for their ability to form invadopodia [266,276,277], indicating that hypoxia induces invadopodia formation in multiple HNSCC cell types. Therefore, we can test the ability of FaDu Shank2-KD cells to form invadopodia and degrade matrix under hypoxic conditions.

Additionally, more patients are needed for our TCGA analysis in order to perform the proper statistical analysis (chi-square test) (**Figure 27**). It is worth noting that M was not a good method of analysis, since HNSCC doesn't usually have distant metastases (only 2 patients out of 312 total). We would also be able to perform a survival and recurrence-free survival analysis with additional patients. With the data used, nearly 50% of the deceased patients had no available follow-up information, making it difficult to do a Kaplan-Meier analysis.

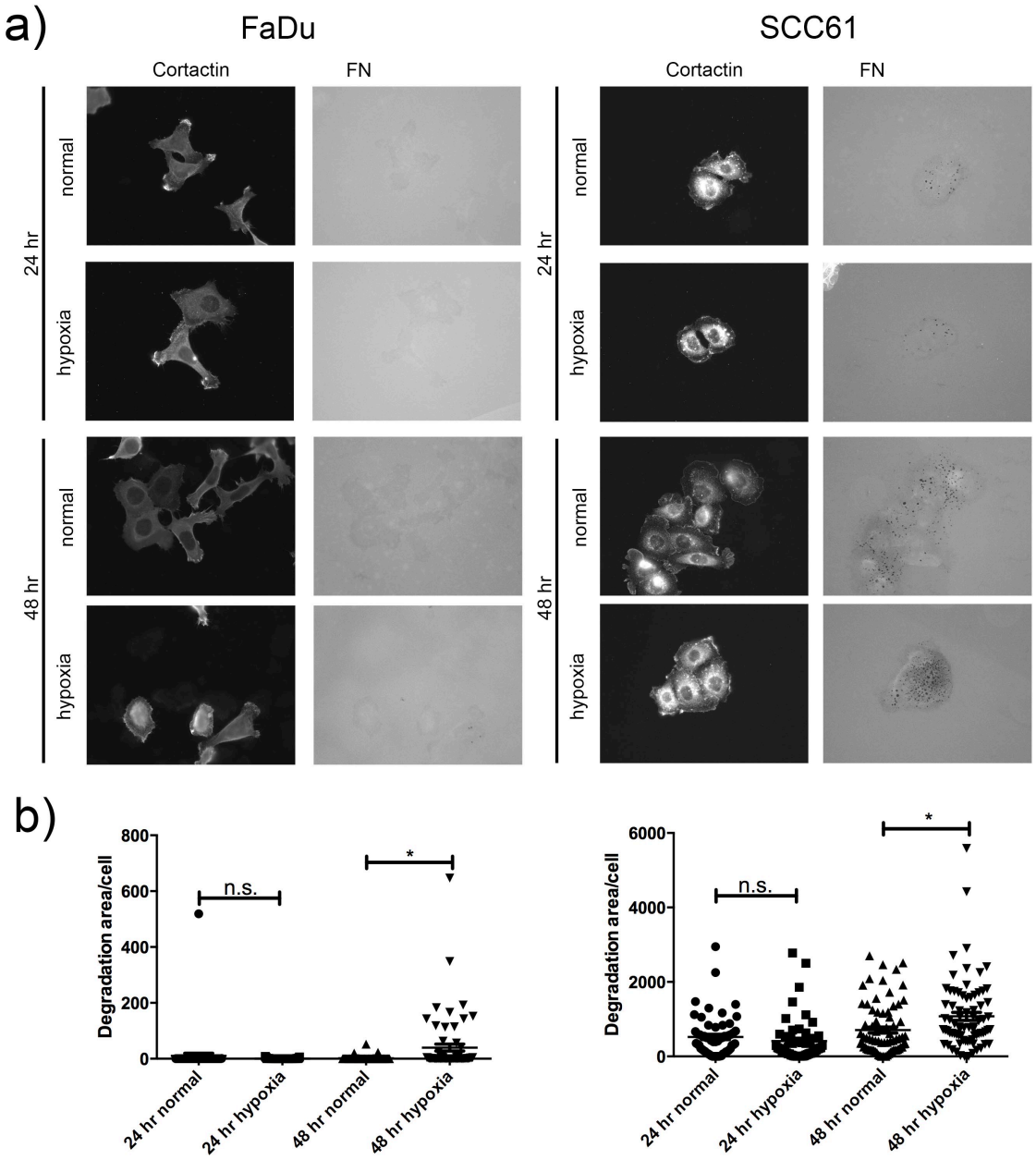


Figure 31: Hypoxia increases invadopodia degradation after 48 hours

a) Representative images of FaDu cells (left) and SCC61 cells (right) stained with cortactin and fibronectin (FN). Cells were plated on fluorescently labeled FN matrix, and fixed and stained after 24 or 48 hours in hypoxic or normal O_2 conditions. Invadopodia activity can be visualized by dark puncta in the FN matrix colocalized with bright cortactin puncta. b) The quantitation of degradation area normalized to cell area. n.s., not significant; * $p<0.05$; n.s., not significant.

CHAPTER V: SUMMARY AND CONCLUSIONS

SUMMARY

Part I

In this dissertation, I use different methods to identify novel regulators of two different cancer types. In CRC, we used a novel bioinformatics approach that identified proteins associated with poor prognosis (defined as patient death or tumor recurrence) in the publically available TCGA dataset. Many of these proteins are well known to be associated with CRC (such as KRAS, SMAD2, SMAD4, and MSH2), indicating that this approach has merit. We followed up on two of these proteins, IGFBP2 and GATA3, to validate our initial results. High levels of IGFBP2 were associated with poor prognosis in our initial TCGA dataset. Using a TMA stained with IGFBP2 for our secondary dataset, we confirmed that patients with high levels of IGFBP2 have significantly decreased survival and recurrence-free survival. Interestingly, the protein most highly associated with death was GATA3, a transcription factor important for T-cell differentiation. While low levels of GATA3 have been also linked with breast cancer, this protein has not previously been studied in the context of CRC. Using cell lines with stable overexpression of GATA3, we show that a higher level of GATA3 decreased invasiveness and 3D colony growth in Matrigel, but did not affect 2D proliferation rates. The mechanism is currently unknown, but our data suggests two nonexclusive possibilities: differentiation or TGF-β signaling. Our results show that consistent with the known role of GATA3 in T-cell differentiation, higher levels of GATA3 are found in more highly differentiated, less aggressive cell lines (**Figure 16**), and higher GATA3 staining is

seen in more differentiated mucosal cells (**Figure 13**). Additionally, since GATA3 has been shown to regulate TGF-β signaling in breast cancer [82], we used a TGF-β reporter and see significantly decreased TGF-β activity in our GATA3-OE cell lines as compared to control (**Figure 18**). Taken together, these data indicate that GATA3 plays a role in CRC invasion.

Part II

In HNSCC, I identified the cortactin SH3 binding partner Shank2 as a novel regulator of cancer aggressiveness. Cortactin is a protein that is overexpressed in up to 40% of HNSCC through the 11q13 amplicon. This overexpression correlates with poor prognosis, and functional studies have shown cortactin regulates tumor size, vascularization, invasiveness, and protein secretion. Cortactin binds to multiple proteins through its SH3 domain (**Table 13**), and it is unknown what mechanistic effect these interactions produce. Here, I showed that the SH3 domain of cortactin was crucial for both Golgi morphology (a phenotype previously connected to tumor aggressiveness [246]) and *in vivo* tumor growth. This suggests that the regulation of tumor aggressiveness phenotypes depends on SH3 binding interactions. Interestingly, one of cortactin's SH3 binding partners, Shank2, is in 11q13 immediately adjacent to CTTN and is overexpressed via its amplification. Therefore, we hypothesized that the interaction of cortactin with Shank2 may regulate tumor aggressiveness. Indeed, we found that knockdown of Shank2 resulted in decreased invasiveness of HNSCC cells, with some preliminary evidence also showing a decrease in invadopodia activity and MMP secretion. This identifies Shank2 as a novel regulator of HNSCC.

SIGNIFICANCE

Bioinformatics is an incredibly valuable tool to use in the analysis of biological data, and the development and characterization of new methods for these analyses is crucial. In particular, the field of cancer research has been tremendously helped by application of bioinformatics techniques to personalized medicine, drug discovery, metabolomics, immunology, immunotherapy, gene discovery, epigenetics, gene function prediction, protein structure prediction, and molecular profiling [19–26]. It is becoming increasingly easier to collect and store large amounts of various types of information. For example, at the date of writing this dissertation, TCGA has copy number alterations, clinical data, DNA methylation, exon expression, gene expression, protein expression, miRNA expression, somatic mutations, and miRNAseq for over 11,000 human tumors samples from 34 different cancer types [278]. Another publically available dataset, the Cancer Cell Line Encyclopedia (CCLE), has DNA copy number, mRNA expression, mutations, and drug treatment data for over 1,000 cell lines [279,280]. A number of other datasets produced by individual researchers or institutions can be obtained through articles published in various scientific journals. Computational and bioinformatics analysis saves time, money, and resources by using this collected data for multiple purposes. In this dissertation, I analyzed high throughput data from two different cancer types. Using RPPA protein expression data and patient clinical data, I used three different algorithms (Cox regression, Wilma, and Regsubsets) to identify proteins associated with poor prognosis (death or recurrence). While these algorithms have been previously developed and described, using them concurrently to gain more confidence in selected proteins was novel. Additionally, I used TCGA RNAseq data to provide evidence suggesting that SHANK2 contributes to cancer aggressiveness

independent of CTTN expression in HNSCC. These findings highlight the utility of these high throughput datasets in making biological discoveries.

Protein overexpression of oncogenes is crucial for driving tumor progression and metastasis; it is therefore important to identify these proteins and the mechanisms underlying their overexpression. Oncogene activation can occur through a mutation of the oncogene, resulting in a protein that is constitutively active. Additionally, oncogenes can be overexpressed at the gene or protein level by gene duplication, induction of gene expression, chromosomal translocation, or epigenetics [11]. When gene duplication occurs frequently at distinct regions, the unit of DNA amplification is called an amplicon [2]. Many oncogenes are overexpressed through amplicons, such as c-MYC (8q24), KRAS (6q12), erbB1 (7p12-13), akt-1 (14q32-33), akt2 (19q13), cdk4-mdm2-sas-gli (12q13), cyclin e (19q12), and others [2,188]. Interestingly, overexpression of the 11q13 amplicon was found to define one subclass of cancer based on a pan-analysis of 12 different TCGA cancer types [188]. In this dissertation, an oncogenic role for the protein IGFBP2 in CRC was suggested: IGFBP2 was associated with poor prognosis in a bioinformatics analysis, and patients in a secondary dataset that expressed high IGFBP2 levels had decreased survival and recurrence-free survival. Additionally, functional studies demonstrating decreased invasiveness and preliminary evidence for decreased invadopodia activity and MMP secretion provide evidence for an oncogenic role of Shank2 in HNSCC.

Tumor progression and metastasis is also driven by repression of tumor suppressor genes. This process occurs at the gene or protein level through gene deletion, gene inactivation, epigenetics, or mutations producing a defunct protein, and in most cases requires inactivation of both tumor suppressor alleles [12]. In this dissertation, low levels of GATA3 were associated with CRC tumor recurrence and patient death, indicating it may play a tumor suppressive role (**Figure 9**, **Figure 10**,

Figure 11). This is further supported by functional studies in CRC cells where overexpression of GATA3 decreases 3D colony growth and invasiveness (**Figure 16**). Interestingly, GATA3 protein levels do not correlate with mRNA levels (**Figure 15**), indicating that the regulation of GATA3 occurs at the protein level.

Clinically, cancer aggressiveness is measured using the TNM staging system, which takes into account primary tumor size and invasiveness (T), spread to regional lymph nodes (N), and metastasis to distant organs (M) [9]. This system is commonly used to make treatment decisions for patients, even though it has several limitations. TNM staging does not take into account additional biologic factors (such as subtypes, heterogeneity, microenvironment, or genetic instability), it is often inadequate to predict patient prognosis [10], and it limited by the current technology for detecting spread to lymph nodes or distant organs. This is underscored by the fact that we find N and M status to be a poor way to define poor prognosis. Using the Wilma algorithm, we see that proteins associated with recurrence and death have the ability to separate patients into distinct populations based on prognosis, while N/M status surprisingly does not (**Figure 4**). Methodologies such as the one presented in this dissertation are a crucial step in the direction of finding biomarkers that more accurately describe tumor aggressiveness to predict patient prognosis and improve patient treatment.

Treatment decisions are important for drugs that are effective only on a subset of patients (for example, targeted therapies). Additionally, treatment decisions are especially crucial for early stage patients. For example, the standard of care for Stage II is surgical resection of the primary tumor [107]. However, after this resection, 30% of CRC Stage II patients have a recurrence at a distant site [114]. This likely means that there were undetectable micrometastases present at the time of diagnosis that subsequently grew after removal of the primary tumor. This subset of patients would benefit from adjuvant therapy to treat these micrometastases. Therefore, biomarkers to

identify this subset of patients would be helpful to clinicians and would likely improve patient prognosis. In this dissertation, I identify GATA3 to be associated with tumor recurrence and patient death by three different bioinformatics analysis. GATA3 expression is significantly lower in deceased Stage I-II patients compared to Stage I-II living patients, and a stage-adjusted Cox proportional hazards shows that GATA-low expressing patients have significantly decreased survival compared to GATA3-high expressing patients. This indicates that GATA3 expression is associated with patient prognosis independent of stage, and has exciting potential as a novel biomarker in CRC. Because RPPA expression data was taken at the time of diagnosis, it is feasible that tumor biopsies in the future could be measured for GATA3 RPPA expression to guide treatment decisions.

Identification of novel oncogenes and tumor suppressors that regulate cancer progression and metastasis is important for developing new biomarkers to diagnose cancer, guide treatment decisions, or identify populations that are at a high risk of developing cancer so preventive measures can be taken. Once identified, targetable proteins can be treated with small molecule inhibitors. Additionally, novel cancer regulators provide mechanistic insight into how individual cancer cells are functioning. The work presented within this dissertation identifies two novel cancer regulators, GATA3 in CRC and Shank2 in HNSCC. While these proteins need to be further characterized with respect to their roles in those cancers and are clearly years away from clinical usefulness, these discoveries have the potential in the future to act as biomarkers and impact cancer treatment.

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA. Cancer J. Clin. [Internet]. 2013; 63:11–30. doi:10.3322/caac.21166.

2. Weinberg RA. The Biology of Cancer. Garland Science, Taylor & Francis Group, LLC; 2007.

3. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976; 194(4260):23–28. doi:10.1126/science.959840.

4. Hanahan D. The Hallmarks of Cancer. Cell. 2000; 100(1):57–70. doi:10.1016/S0092-8674(00)81683-9.

5. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011; 144(5):646–674. doi:10.1016/j.cell.2011.02.013.

6. Van Zijl F, Krupitza G, Mikulits W. Initial steps of metastasis: Cell invasion and endothelial transmigration. Mutat. Res. - Rev. Mutat. Res. 2011; 728(1-2):23–34. doi:10.1016/j.mrrev.2011.05.002.

7. Senger DR, Davis GE. Angiogenesis. Cold Spring Harb. Perspect. Biol. 2011; 3(8):1–19. doi:10.1101/cshperspect.a005090.

8. Sporn MB. The war on cancer. Lancet. 1996; 347(9012):1377–1381.

9. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann. Surg. Oncol. 2010; 17(6):1471–1474. doi:10.1245/s10434-010-0985-4.

10. Piccirillo JF, Feinstein AR. Clinical symptoms and comorbidity: Significance for the prognostic classification of cancer. Cancer. 1996; 77(5):834–842. doi:10.1002/(SICI)1097-0142(19960301)77:5<834::AID-CNCR5>3.0.CO;2-E.

11. Croce C. Oncogenes and cancer. N Engl J Med. 2008; 358(5):502–511. doi:10.1126/science.7878455.

12. Sherr CJ. Principles of Tumor Suppression. Cell. 2004; 116(2):235–246. doi:10.1016/S0092-8674(03)01075-4.

13. Negrini M, Ferracin M, Sabbioni S, Croce CM. MicroRNAs in human cancer: from research to therapy. J. Cell Sci. 2007; 120(Pt 11):1833–1840. doi:10.1242/jcs.03450.

14. Musolino A, Bella MA, Bortesi B, Michiara M, Naldi N, Zanelli P, et al. BRCA mutations, molecular markers, and clinical variables in early-onset breast cancer:

a population-based study. Breast [Internet]. 2007; 16(3):280–292. doi:S0960-9776(06)00224-4 [pii]\r10.1016/j.breast.2006.12.003.

15. Kruijff S, Hoekstra HJ. The current status of S-100B as a biomarker in melanoma. Eur. J. Surg. Oncol. 2012; 38(4):281–285. doi:10.1016/j.ejso.2011.12.005.

16. Orphanos G, Kountourakis P. Targeting the HER2 receptor in metastatic breast cancer. Hematol. Oncol. Stem Cell Ther. [Internet]. 2012; 5(3):127–37. doi:10.5144/1658-3876.2012.127.

17. Hesper B, Hogeweg P. Bioinformatica: een werkconcept. Kameleon. 1970; 1(6):28–29.

18. Benson D, Boguski M, Lipman DJ, Ostell J. The National Center for Biotechnology Information. Genomics. 1990; 6:389–391.

19. Blekherman G, Laubenbacher R, Cortes DF, Mendes P, Torti FM, Akman S, et al. Bioinformatics tools for cancer metabolomics. Metabolomics. 2011; 7(3):329–343. doi:10.1007/s11306-010-0270-3.

20. Desany B, Zhang Z. Bioinformatics and cancer target discovery. Drug Discov. Today. 2004; 9(18):795–802. doi:10.1016/S1359-6446(04)03224-6.

21. Charoentong P, Angelova M, Efremova M, Gallasch R, Hackl H, Galon J, et al. Bioinformatics for cancer immunology and immunotherapy. Cancer Immunol. Immunother. 2012; 61(11):1885–1903. doi:10.1007/s00262-012-1354-x.

22. Ram PT, Mendelsohn J, Mills GB. Bioinformatics and systems biology. Mol. Oncol. 2012; 6(2):147–154. doi:10.1016/j.molonc.2012.01.008.

23. Narayanan R. Bioinformatics approaches to cancer gene discovery. Methods Mol. Biol. 2007; 360:13–31. doi:10.1385/1-59745-165-7:13.

24. Yang HH, Lee MP. Application of bioinformatics in cancer epigenetics. In: Annals of the New York Academy of Sciences. 2004. p. 67–76. doi:10.1196/annals.1310.008.

25. Kihara D, Yang YD, Hawkins T. Bioinformatics resources for cancer research with an emphasis on gene function and structure prediction tools. Cancer Inform. 2006; 2:25–35.

26. Manning AT, Garvin JT, Shahbazi RI, Miller N, McNeill RE, Kerin MJ. Molecular profiling techniques and bioinformatics in cancer research. Eur. J. Surg. Oncol. 2007; 33(3):255–265. doi:10.1016/j.ejso.2006.09.002.

27. Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, et al. Proteogenomic characterization of human colon and rectal cancer. Nature [Internet]. 2014; doi:10.1038/nature13438.

28. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene Expression Classification of Colon Cancer into Molecular Subtypes: Characterization, Validation, and Prognostic Value. PLoS Med. 2013; 10. doi:10.1371/journal.pmed.1001453.

29. Phipps AI, Limburg PJ, Baron J a., Burnett-Hartman AN, Weisenberger DJ, Laird PW, et al. Association Between Molecular Subtypes of Colorectal Cancer and Patient Survival. Gastroenterology [Internet]. 2015; 148(1):77–87.e2. doi:10.1053/j.gastro.2014.09.038.

30. Zhu J, Wang J, Shi Z, Franklin JL, Deane NG, Coffey RJ, et al. Deciphering genomic alterations in colorectal cancer through transcriptional subtype-based network analysis. PLoS One [Internet]. 2013; 8(11):e79282. doi:10.1371/journal.pone.0079282.

31. Hoshino D, Jourquin J, Emmons SW, Miller T, Goldgof M, Costello K, et al. Network analysis of the focal adhesion to invadopodia transition identifies a PI3K-PKCalpha invasive signaling axis. Sci Signal [Internet]. 2012; 5(241):ra66. doi:10.1126/scisignal.2002964.

32. Denduluri SK, Idowu O, Wang Z, Liao Z, Yan Z, Mohammed MK, et al. Insulin-like growth factor (IGF) signaling in tumorigenesis and the development of cancer drug resistance. Genes Dis. [Internet]. 2015; 2(1):13–25. doi:10.1016/j.gendis.2014.10.004.

33. Yau SW, Azar WJ, Sabin M a., Werther G a., Russo VC. IGFBP-2 - taking the lead in growth, metabolism and cancer. J. Cell Commun. Signal. [Internet]. 2015; (Braulke 1999). doi:10.1007/s12079-015-0261-2.

34. Foulstone EJ, Zeng L, Perks CM, Holly JMP. Insulin-like growth factor binding protein 2 (IGFBP-2) promotes growth and survival of breast epithelial cells: Novel regulation of the estrogen receptor. Endocrinology. 2013; 154(5):1780–1793. doi:10.1210/en.2012-1970.

35. Wang H, Arun BK, Wang H, Fuller GN, Zhang W, Middleton LP, et al. IGFBP2 and IGFBP5 overexpression correlates with the lymph node metastasis in T1 breast carcinomas. Breast J. 2008; 14(3):261–267. doi:10.1111/j.1524-4741.2008.00572.x.

36. Dean SJR, Perks CM, Holly JMP, Bhoo-Pathy N, Looi LM, Mohammed NAT, et al. Loss of PTEN expression is associated with IGFBP2 expression, younger age, and late stage in triple-negative breast cancer. Am. J. Clin. Pathol. 2014; 141(3):323–333. doi:10.1309/AJCPR11DEAYPTUSL.

37. Richardsen E, Ukkonen T, Bjørnsen T, Mortensen E, Egevad L, Busch C. Overexpression of IGBFB2 is a marker for malignant transformation in prostate epithelium. Virchows Arch. 2003; 442(4):329–335. doi:10.1007/s00428-003-0786-2.

38. Cohen P, Peehl DM, Stamey TA, Wilson KF, Clemmons DR, Rosenfeld RG. Elevated levels of insulin-like growth factor-binding protein-2 in the serum of prostate cancer patients. J. Clin. Endocrinol. Metab. 1993; 76(4):1031–1035.

39. Ho PJ, Baxter RC. Insulin-like growth factor-binding protein-2 in patients with prostate carcinoma and benign prostatic hyperplasia. Clin. Endocrinol. (Oxf). 1997; 46(3):333–342.

40. Kanety H, Madjar Y, Dagan Y, Levi J, Papa MZ, Pariente C, et al. Serum insulin-like growth factor-binding protein-2 (IGFBP-2) is increased and IGFBP-3 is decreased in patients with prostate cancer: correlation with serum prostate-specific antigen. J. Clin. Endocrinol. Metab. [Internet]. 1993; 77(1):229–233. doi:10.1210/jc.77.1.229.

41. Karasik A, Menczer J, Pariente C, Kanety H. Insulin-like growth factor-I (IGF-I) and IGF-binding protein-2 are increased in cyst fluids of epithelial ovarian cancer. J. Clin. Endocrinol. Metab. 1994; 78(2):271–276. doi:10.1210/jc.78.2.271.

42. Wang H, Rosen DG, Wang H, Fuller GN, Zhang W, Liu J. Insulin-like growth factor-binding protein 2 and 5 are differentially regulated in ovarian cancer of different histologic types. Mod. Pathol. 2006; 19(9):1149–1156. doi:10.1038/modpathol.3800637.

43. Lee E-J, Mircean C, Shmulevich I, Wang H, Liu J, Niemistö A, et al. Insulin-like growth factor binding protein 2 promotes ovarian cancer cell invasion. Mol. Cancer. 2005; 4:7. doi:10.1186/1476-4598-4-7.

44. McDonald KL, O'Sullivan MG, Parkinson JF, Shaw JM, Payne CA, Brewer JM, et al. IQGAP1 and IGFBP2: valuable biomarkers for determining prognosis in glioma patients. J. Neuropathol. Exp. Neurol. 2007; 66(5):405–417. doi:10.1097/nen.0b013e31804567d7.

45. Hsieh D, Hsieh A, Stea B, Ellsworth R. IGFBP2 promotes glioma tumor stem cell expansion and survival. Biochem Biophys Res Commun [Internet]. 2010; 397(2):367–372. doi:10.1016/j.bbrc.2010.05.145.

46. Dunlap SM, Celestino J, Wang H, Jiang R, Holland EC, Fuller GN, et al. Insulin-like growth factor binding protein 2 promotes glioma development and progression. Proc. Natl. Acad. Sci. U. S. A. 2007; 104(28):11736–11741. doi:10.1073/pnas.0703145104.

47. Wang H, Wang H, Shen W, Huang H, Hu L, Ramdas L, et al. Insulin-like growth factor binding protein 2 enhances glioblastoma invasion by activating invasion-enhancing genes. Cancer Res. 2003; 63(15):4315–4321.

48. Fukushima T, Tezuka T, Shimomura T, Nakano S, Kataoka H. Silencing of insulin-like growth factor-binding protein-2 in human glioblastoma cells reduces both invasiveness and expression of progression-associated gene CD24. J. Biol. Chem. 2007; 282(25):18634–18644. doi:10.1074/jbc.M609567200.

49. Kühnl A, Kaiser M, Neumann M, Fransecky L, Heesch S, Radmacher M, et al. High expression of IGFBP2 is associated with chemoresistance in adult acute myeloid leukemia. Leuk. Res. 2011; 35(12):1585–1590. doi:10.1016/j.leukres.2011.08.006.

50. El-Naggar AA, Mahmoud BF, Abou Shamaa LA, Salama MA. Changes of serum growth factors (IGF-I & IGFBP-2) and prediction of response to chemotherapy in patients with acute myeloid leukemia. 2008.

51. Zhang Y, Ying X, Han S, Wang J, Zhou X, Bai E, et al. Autoantibodies against insulin-like growth factor-binding protein-2 as a serological biomarker in the diagnosis of lung cancer. Int. J. Oncol. 2013; 42(1):93–100. doi:10.3892/ijo.2012.1699.

52. Migita T, Narita T, Asaka R, Miyagi E, Nagano H, Nomura K, et al. Role of insulin-like growth factor binding protein 2 in lung adenocarcinoma: IGF-independent antiapoptotic effect via caspase-3. Am. J. Pathol. 2010; 176(4):1756–1766. doi:10.2353/ajpath.2010.090500.

53. Renehan AG, Jones J, Potten CS, Shalet SM, O'Dwyer ST. Elevated serum insulin-like growth factor (IGF)-II and IGF binding protein-2 in patients with colorectal cancer. Br J Cancer [Internet]. 2000; 83(10):1344–1350. doi:10.1054/bjoc.2000.1462.

54. Ben-Shmuel A, Shvab A, Gavert N, Brabletz T, Ben-Ze'ev A. Global analysis of L1-transcriptomes identified IGFBP-2 as a target of ezrin and NF-kappaB signaling that promotes colon cancer progression. Oncogene [Internet]. 2013; 32(27):3220–3230. doi:10.1038/onc.2012.340.

55. Mishra L, Bass B, Ooi BS, Sidawy A, Korman L. Role of insulin-like growth factor-I (IGF-I) receptor, IGF-I, and IGF binding protein-2 in human colorectal cancers. Growth Horm IGF Res [Internet]. 1998; 8(6):473–479. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10985759.

56. Miraki-Moud F, Jenkins PJ, Fairclough PD, Jordan S, Bustin SA, Jones AM, et al. Increased levels of insulin-like growth factor binding protein-2 in sera and tumours from patients with colonic neoplasia with and without acromegaly. Clin. Endocrinol. (Oxf). 2001; 54(4):499–508. doi:10.1046/j.1365-2265.2001.01221.x.

57. El Atiq F, Garrouste F, Remacle-Bonnet M, Sastre B, Pommier G. Alterations in serum levels of insulin-like growth factors and insulin-like growth-factor-binding proteins in patients with colorectal cancer. Int. J. Cancer. 1994; 57(4):491–497.

58. Corkins MR, Vanderhoof JA, Slentz DH, Macdonald RG, Park JHY. Growth-Stimulation By Transfection Of Intestinal Epithelial-Cells With an Antisense Insulin-Like Growth-Factor Binding Protein-2 Construct. Biochem. Biophys. Res. Commun. 1995; 211(3):707–713.

59. Höflich A, Lahm H, Blum W, Kolb H, Wolf E. Insulin-like growth factor-binding protein-2 inhibits proliferation of human embryonic kidney fibroblasts and of IGF-responsive colon carcinoma cell lines. FEBS Lett. 1998; 434(3):329–334.

60. Michell NP, Langman MJ, Eggo MC. Insulin-like growth factors and their binding proteins in human colonocytes: preferential degradation of insulin-like growth factor binding protein 2 in colonic cancers. Br J Cancer [Internet]. 1997; 76(1):60–66. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9218734.

61. Diehl D, Hessel E, Oesterle D, Renner-Müller I, Elmlinger M, Langhammer M, et al. IGFBP-2 overexpression reduces the appearance of dysplastic aberrant crypt foci and inhibits growth of adenomas in chemically induced colorectal carcinogenesis. Int. J. Cancer. 2009; 124(9):2220–2225. doi:10.1002/ijc.24193.

62. Yagi R, Zhu J, Paul WE. An updated view on transcription factor GATA3-mediated regulation of Th1 and Th2 cell differentiation. Int. Immunol. 2011; 23(7):415–420. doi:10.1093/intimm/dxr029.

63. Zhang DH, Cohn L, Ray P, Bottomly K, Ray A. Transcription factor GATA-3 is differentially expressed in murine Th1 and Th2 cells and controls Th2-specific expression of the interleukin-5 gene. J. Biol. Chem. 1997; 272(34):21597–21603. doi:10.1074/jbc.272.34.21597.

64. Zheng W, Flavell RA. The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. Cell. 1997; 89:587–596. doi:10.1016/S0092-8674(00)80240-8.

65. Ouyang W, Ranganath SH, Weindel K, Bhattacharya D, Murphy TL, Sha WC, et al. Inhibition of Th1 development mediated by GATA-3 through an IL-4-independent mechanism. Immunity. 1998; 9(5):745–755. doi:10.1016/S1074-7613(00)80671-8.

66. Ferber IA, Lee HJ, Zonin F, Heath V, Mui A, Arai N, et al. GATA-3 significantly downregulates IFN-gamma production from developing Th1 cells in addition to inducing IL-4 and IL-5 levels. Clin. Immunol. 1999; 91(2):134–144. doi:10.1006/clim.1999.4718.

67. Usui T, Nishikomori R, Kitani A, Strober W. GATA-3 suppresses Th1 development by downregulation of Stat4 and not through effects on IL-12R-beta2 chain or T-bet. Immunity. 2003; 18(3):415–428. doi:10.1016/S1074-7613(03)00057-8.

68. Yagi R, Junttila IS, Wei G, Urban JF, Zhao K, Paul WE, et al. The Transcription Factor GATA3 Actively Represses RUNX3 Protein-Regulated Production of Interferon-gamma. Immunity. 2010; 32(4):507–517. doi:10.1016/j.immuni.2010.04.004.

69. Kaufman CK, Zhou P, Pasolli HA, Rendl M, Bolotin D, Lim KC, et al. GATA-3: An unexpected regulator of cell lineage determination in skin. Genes Dev. 2003; 17:2108–2122. doi:10.1101/gad.1115203.

70. Grote D, Boualia SK, Souabni A, Merkel C, Chi X, Costantini F, et al. Gata3 acts downstream of b-catenin signaling to prevent ectopic metanephric kidney induction. PLoS Genet. 2008; 4(12). doi:10.1371/journal.pgen.1000316.

71. Asselin-Labat M-L, Sutherland KD, Barker H, Thomas R, Shackleton M, Forrest NC, et al. Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. Nat. Cell Biol. 2007; 9(2):201–209. doi:10.1038/ncb1530.

72. Kouros-Mehr H, Slorach EM, Sternlicht MD, Werb Z. GATA-3 Maintains the Differentiation of the Luminal Cell Fate in the Mammary Gland. Cell. 2006; 127(5):1041–1055. doi:10.1016/j.cell.2006.09.048.

73. Cancer Genome Atlas N, Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, et al. Comprehensive molecular portraits of human breast tumours. Nature [Internet]. 2012; 490(7418):61–70. doi:10.1038/nature11412.

74. Usary J, Llaca V, Karaca G, Presswala S, Karaca M, He X, et al. Mutation of GATA3 in human breast tumors. Oncogene. 2004; 23:7669–7678. doi:10.1038/sj.onc.1207966.

75. Jiang YZ, Yu K Da, Zuo WJ, Peng WT, Shao ZM. GATA3 mutations define a unique subtype of luminal-like breast cancer with improved survival. Cancer. 2014; 120:1329–1337. doi:10.1002/cncr.28566.

76. Kouros-Mehr H, Bechis SK, Slorach EM, Littlepage LE, Egeblad M, Ewald AJ, et al. GATA-3 Links Tumor Differentiation and Dissemination in a Luminal Breast Cancer Model. Cancer Cell. 2008; 13(February):141–152. doi:10.1016/j.ccr.2008.01.011.

77. Chou J, Provot S, Werb Z. GATA3 in development and cancer differentiation: Cells GATA have it! J. Cell. Physiol. 2010; 222(August):42–49. doi:10.1002/jcp.21943.

78. Dydensborg AB, Rose AAN, Wilson BJ, Grote D, Paquet M, Giguère V, et al. GATA3 inhibits breast cancer growth and pulmonary breast cancer metastasis. Oncogene. 2009; 28(29):2634–2642. doi:10.1038/onc.2009.126.

79. Pei XH, Bai F, Smith MD, Usary J, Fan C, Pai SY, et al. CDK Inhibitor p18INK4c Is a Downstream Target of GATA3 and Restrains Mammary Luminal Progenitor Cell Proliferation and Tumorigenesis. Cancer Cell. 2009; 15(5):389–401. doi:10.1016/j.ccr.2009.03.004.

80. Chou J, Lin JH, Brenot A, Kim J, Provot S, Werb Z. GATA3 suppresses metastasis and modulates the tumour microenvironment by regulating microRNA-

29b expression. Nat. Cell Biol. [Internet]. 2013; 15(2):201–13. doi:10.1038/ncb2672.

81. Yan W, Cao QJ, Arenas RB, Bentley B, Shao R. GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition. J. Biol. Chem. 2010; 285(18):14042–14051. doi:10.1074/jbc.M110.105262.

82. Sun J, He H, Pillai S, Xiong Y, Challa S, Xu L, et al. GATA3 transcription factor abrogates Smad4 transcription factor-mediated fascin overexpression, invadopodium formation, and breast cancer cell invasion. J. Biol. Chem. 2013; 288:36971–36982. doi:10.1074/jbc.M113.506535.

83. Nguyen AHT, Tremblay M, Haigh K, Hervékoumakpayi I, Paquet M, Pandolfi PP, et al. Gata3 antagonizes cancer progression in pten-deficient prostates. Hum. Mol. Genet. 2013; 22(12):2400–2410. doi:10.1093/hmg/ddt088.

84. Tu JC, Xiao B, Naisbitt S, Yuan JP, Petralia RS, Brakeman P, et al. Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of postsynaptic density proteins. Neuron. 1999; 23(3):583–592. doi:10.1016/S0896-6273(00)80810-7.

85. Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, et al. Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nat. Genet. 2010; 42(6):489–491. doi:10.1038/ng.589.

86. Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, Huguet G, et al. Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. PLoS Genet. 2012; 8(2). doi:10.1371/journal.pgen.1002521.

87. Won H, Lee H-R, Gee HY, Mah W, Kim J-I, Lee J, et al. Autistic-like social behaviour in Shank2-mutant mice improved by restoring NMDA receptor function. Nature. 2012; 486(7402):261–265. doi:10.1038/nature11208.

88. Schmeisser MJ, Ey E, Wegener S, Bockmann J, Stempel AV, Kuebler A, et al. Autistic-like behaviours and hyperactivity in mice lacking ProSAP1/Shank2. Nature. 2012; doi:10.1038/nature11015.

89. Lim S, Naisbitt S, Yoon J, Hwang JI, Suh PG, Sheng M, et al. Characterization of the Shank family of synaptic proteins. Multiple genes, alternative splicing, and differential expression in brain and development. J. Biol. Chem. 1999; 274(41):29510–29518. doi:10.1074/jbc.274.41.29510.

90. McWilliams RR, Gidey E, Fouassier L, Weed SA, Doctor RB. Characterization of an ankyrin repeat-containing Shank2 isoform (Shank2E) in liver epithelial cells. Biochem. J. 2004; 380(Pt 1):181–191. doi:10.1042/BJ20031577.

91. Han W, Kyung HK, Min JJ, Ji HL, Yang J, Doctor RB, et al. Shank2 associates with and regulates Na+/H+ exchanger. J. Biol. Chem. 2006; 281(3):1461–1469. doi:10.1074/jbc.M509786200.

92. Hwang JI, Hyeon SK, Jae RL, Kim E, Sung HR, Suh PG. The interaction of phospholipase C-??3 with Shank2 regulates mGluR-mediated calcium signal. J. Biol. Chem. 2005; 280(13):12467–12473. doi:10.1074/jbc.M410740200.

93. Du Y, Weed SA, Xiong WC, Marshall TD, Parsons JT. Identification of a novel cortactin SH3 domain-binding protein and its localization to growth cones of cultured neurons. Mol. Cell. Biol. 1998; 18(10):5838–5851.

94. Kirkbride KC, Sung BH, Sinha S, Weaver AM. Cortactin: A multifunctional regulator of cellular invasiveness. Cell Adhes. Migr. 2011; 5(2):187–198. doi:10.4161/cam.5.2.14773.

95. Gollin SM. Chromosomal alterations in squamous cell carcinomas of the head and neck: window to the biology of disease. Head Neck. 2001; 23(March):238–253.

96. Bockmühl U, Petersen I. DNA ploidy and chromosomal alterations in head and neck squamous cell carcinoma. Virchows Arch. 2002; 441:541–550. doi:10.1007/s00428-002-0729-3.

97. Huang X, Gollin SM, Raja S, Godfrey TE. High-resolution mapping of the 11q13 amplicon and identification of a gene, TAOS1, that is amplified and overexpressed in oral cancer cells. Proc. Natl. Acad. Sci. U. S. A. 2002; 99(17):11369–11374. doi:10.1073/pnas.172285799.

98. Bryce NS, Clark ES, Leysath JL, Currie JD, Webb DJ, Weaver AM. Cortactin promotes cell motility by enhancing lamellipodial persistence. Curr. Biol. 2005; 15(14):1276–1285. doi:10.1016/j.cub.2005.06.043.

99. R Core Team. R: A language and environment for statistical computing [Internet]. 2014; Available from: http://www.r-project.org/.

100. Grambsch TMT and PM. Modeling Survival Data: Extending the Cox Model. New York: Springer; 2000.

101. Therneau T. A Package for Survival Analysis in S [Internet]. 2014; Available from: URL: http://CRAN.R-project.org/package=survival.

102. Dettling M, Bühlmann P. Supervised clustering of genes. Genome Biol. 2002; 3:RESEARCH0069. doi:10.1186/gb-2002-3-12-research0069.

103. Lumley T. leaps: regression subset selection. Compr. R Arch. Netw. [Internet]. 2009; Available from: http://cran.r-project.org/web/packages/leaps/index.html.

104. Day A. heatmap.plus: Heatmap with more sensible behavior [Internet]. 2012; Available from: http://cran.r-project.org/package=heatmap.plus.

105. Debnath J, Muthuswamy SK, Brugge JS. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement

membrane cultures. Methods. 2003; 30:256–268. doi:10.1016/S1046-2023(03)00032-X.

106. Shores CG, Yarbrough WG. Three-dimensional xenograft model of dysplastic human laryngeal mucosa. Laryngoscope. 1998; 108(9):1358–1362. doi:10.1097/00005537-199809000-00019.

107. Labianca R, Nordlinger B, Beretta GD, Brouquet a., Cervantes a. Primary colon cancer: ESMO clinical practice guidelines for diagnosis, adjuvant treatment and follow-up. Ann. Oncol. 2010; 21(Supplement 5):70–77. doi:10.1093/annonc/mdq168.

108. Marsoni I, Analysis P, Trials CC. Efficacy of Adjuvant Fluorouracil and Leucovorin Stage B2 and C Colon Cancer. 28(I). doi:10.1053/sonc.2001.

109. Benson AB, Schrag D, Somerfield MR, Cohen AM, Figueredo AT, Flynn PJ, et al. American Society of Clinical Oncology recommendations on adjuvant chemotherapy for stage II colon cancer. J. Clin. Oncol. 2004; 22(16):3408–3419. doi:10.1200/JCO.2004.05.063.

110. Figueredo A, Charette ML, Maroun J, Brouwers MC, Zuraw L. Adjuvant therapy for stage II colon cancer: A systematic review from the cancer care Ontario program in evidence-based care's gastrointestinal cancer disease site group. J. Clin. Oncol. 2004; 22(16):3395–3407. doi:10.1200/JCO.2004.03.087.

111. Gill S, Loprinzi CL, Sargent DJ, Thomé SD, Alberts SR, Haller DG, et al. Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: Who benefits and by how much? J. Clin. Oncol. 2004; 22(10):1797–1806. doi:10.1200/JCO.2004.09.059.

112. Mamounas E, Wieand S, Wolmark N, Bear HD, Atkins JN, Song K, et al. Comparative Efficacy of Adjuvant Chemotherapy in Patients With Dukes' B Versus Dukes' C Colon Cancer: Results From Four National Surgical Adjuvant Breast and Bowel Project. 1999; 17(5):1349–1355.

113. Schmoll HJ, Van Cutsem E, Stein a, Valentini V, Glimelius B, Haustermans K, et al. ESMO Consensus Guidelines for management of patients with colon and rectal cancer. a personalized approach to clinical decision making. Ann. Oncol. [Internet]. 2012; 23(10):2479–516. doi:10.1093/annonc/mds236.

114. Sadahiro S, Suzuki T, Ishikawa K, Nakamura T, Tanaka Y, Masuda T, et al. Recurrence patterns after curative resection of colorectal cancer in patients followed for a minimum of ten years. Hepatogastroenterology. 2003; 50:1362–1366.

115. Davies RJ, Miller R, Coleman N. Colorectal cancer screening: prospects for molecular stool analysis. Nat. Rev. Cancer. 2005; 5(3):199–209. doi:10.1038/nrc1545.

116. Sachdev D, Yee D. Disrupting insulin-like growth factor signaling as a potential cancer therapy. Mol. Cancer Ther. 2007; 6(1):1–12. doi:10.1158/1535-7163.MCT-06-0080.

117. Joulin V, Bories D, Eléouet JF, Labastie MC, Chrétien S, Mattéi MG, et al. A T-cell specific TCR delta DNA binding protein is a member of the human GATA family. EMBO J [Internet]. 1991; 10(7):1809–1816. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2050118.

118. Ho IC, Vorhees P, Marin N, Oakley BK, Tsai SF, Orkin SH, et al. Human GATA-3: a lineage-restricted transcription factor that regulates the expression of the T cell receptor alpha gene. EMBO J [Internet]. 1991; 10(5):1187–1192. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1827068.

119. Pandolfi PP, Roth ME, Karis A, Leonard MW, Dzierzak E, Grosveld FG, et al. Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. Nat. Genet. 1995; 11(1):40–44. doi:10.1038/ng0995-40.

120. Van Esch H, Groenen P, Nesbit MA, Schuffenhauer S, Lichtner P, Vanderlinden G, et al. GATA3 haplo-insufficiency causes human HDR syndrome. Nature. 2000; 406(6794):419–422. doi:10.1038/35019088.

121. Mehra R, Varambally S, Ding L, Shen R, Sabel MS, Ghosh D, et al. Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. Cancer Res. 2005; 65:11259–11264. doi:10.1158/0008-5472.CAN-05-2495.

122. Dolled-Filhart M, Rydén L, Cregger M, Jirström K, Harigopal M, Camp RL, et al. Classification of breast cancer using genetic algorithms and tissue microarrays. Clin. Cancer Res. 2006; 12(21):6459–6468. doi:10.1158/1078-0432.CCR-06-1383.

123. Yoon NK, Maresh EL, Shen D, Elshimali Y, Apple S, Horvath S, et al. Higher levels of GATA3 predict better survival in women with breast cancer. Hum. Pathol. 2010; 41(12):1794–1801. doi:10.1016/j.humpath.2010.06.010.

124. Voduc D, Cheang M, Nielsen T. GATA-3 expression in breast cancer has a strong association with estrogen receptor but lacks independent prognostic value. Cancer Epidemiol. Biomarkers Prev. 2008; 17(February):365–373. doi:10.1158/1055-9965.EPI-06-1090.

125. Maechler MD and M. supclust: Supervised Clustering of Predictor Variables such as Genes [Internet]. 2011; Available from: http://cran.r-project.org/package=supclust.

126. Zhang Y, Feng XH, Derynck R. Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-beta-induced transcription. Nature [Internet]. 1998; 394(6696):909–913. doi:10.1038/29814.

127. Fishel R, Ewel A, Lee S, Lescoe MK, Griffith J. Binding of mismatched microsatellite DNA sequences by the human MSH2 protein. Science. 1994; 266:1403–1405. doi:10.1126/science.7973733.

128. Tsujii M, Kawano S, DuBois RN. Cyclooxygenase-2 expression in human colon cancer cells increases metastatic potential. Proc Natl Acad Sci U S A [Internet]. 1997; 94(7):3336–3340. doi:10.1073/pnas.94.7.3336.

129. Tsujii M, Kawano S, Tsuji S, Sawaoka H, Hori M, DuBois RN. Cyclooxygenase regulates angiogenesis induced by colon cancer cells. Cell [Internet]. 1998; 93(5):705–716. doi:10.1016/S0092-8674(00)81433-6.

130. Nateri AS, Spencer-Dene B, Behrens A. Interaction of phosphorylated c-Jun with TCF4 regulates intestinal cancer development. Nature. 2005; 437(September):281–285. doi:10.1038/nature03914.

131. Yang G, Yang X. Smad4-mediated TGF-beta signaling in tumorigenesis. Int J Biol Sci [Internet]. 2010; 6(1):1–8. doi:20087440.

132. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A [Internet]. 2001; 98(19):10869–10874. doi:10.1073/pnas.191367098.

133. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature [Internet]. 2000; 406(6797):747–752. doi:10.1038/35021093.

134. Jenssen T-K, Kuo WP, Stokke T, Hovig E. Associations between gene expressions in breast cancer and patient survival. Hum. Genet. 2002; 111:411–420. doi:10.1007/s00439-002-0804-5.

135. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based Human Protein Atlas. Nat. Biotechnol. 2010; 28(12):1248–1250. doi:10.1038/nbt1210-1248.

136. Subramanian A, Subramanian A, Tamayo P, Tamayo P, Mootha VK, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. [Internet]. 2005; 102(43):15545–50. doi:10.1073/pnas.0506580102.

137. Vermeulen SJ, Bruyneel EA, Bracke ME, De Bruyne GK, Vennekens KM, Vleminckx KL, et al. Transition from the noninvasive to the invasive phenotype and loss of alpha-catenin in human colon cancer cells. Cancer Res [Internet]. 1995; 55(20):4722–4728. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7553655.

138. Yoon WH, Lee SK, Song KS, Kim JS, Kim TD, Li G, et al. The tumorigenic, invasive and metastatic potential of epithelial and round subpopulations of the

SW480 human colon cancer cell line. Mol. Med. Rep. 2008; 1:763–768. doi:10.3892/mmr-00000026.

139. Kurek D, Garinis G a, van Doorninck JH, van der Wees J, Grosveld FG. Transcriptome and phenotypic analysis reveals Gata3-dependent signalling pathways in murine hair follicles. Development. 2007; 134:261–272. doi:10.1242/dev.02721.

140. Ho I-C, Tai T-S, Pai S-Y. GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. Nat. Rev. Immunol. 2009; 9(February):125–135. doi:10.1038/nri2476.

141. Zhu J, Yamane H, Cote-Sierra J, Guo L, Paul WE. GATA-3 promotes Th2 responses through three different mechanisms: induction of Th2 cytokine production, selective growth of Th2 cells and inhibition of Th1 cell-specific factors. Cell Res. 2006; 16:3–10. doi:10.1038/sj.cr.7310002.

142. Ting CN, Olson MC, Barton KP, Leiden JM. Transcription factor GATA-3 is required for development of the T-cell lineage. Nature. 1996; 384:474–478. doi:10.1038/384474a0.

143. Pinto M, Robine-Leon S, Appay M-D, Kedinger M, Triadou N, Dussaulx E, et al. Enterocyte-like differentiation and polarization of the human colon carcinoma cell line Caco-2 in culture. Biol. Cell. 1983; 47(323):323–330.

144. Rousset M. The human colon carcinoma cell lines HT-29 and Caco-2: two in vitro models for the study of intestinal differentiation. Biochimie [Internet]. 1986; 68(9):1035–1040. doi:10.1016/S0300-9084(86)80177-8.

145. Kolegraff K, Nava P, Helms MN, Parkos CA, Nusrat A. Loss of desmocollin-2 confers a tumorigenic phenotype to colonic epithelial cells through activation of Akt/beta-catenin signaling. Mol Biol Cell [Internet]. 2011; 22(8):1121–1134. doi:10.1091/mbc.E10-10-0845.

146. Glondu M, Liaudet-Coopman E, Derocq D, Platet N, Rochefort H, Garcia M. Down-regulation of cathepsin-D expression by antisense gene transfer inhibits tumor growth and experimental lung metastasis of human breast cancer cells. Oncogene [Internet]. 2002; 21(33):5127–5134. doi:10.1038/sj.onc.1205657.

147. Clark ES, Brown B, Whigham a S, Kochaishvili a, Yarbrough WG, Weaver a M. Aggressiveness of HNSCC tumors depends on expression levels of cortactin, a gene in the 11q13 amplicon. Oncogene. 2009; 28:431–444. doi:10.1038/onc.2008.389.

148. Sabeh F, Shimizu-Hirota R, Weiss SJ. Protease-dependent versus -independent cancer cell invasion programs: three-dimensional amoeboid movement revisited. J Cell Biol [Internet]. 2009; 185(1):11–19. doi:10.1083/jcb.200807195.

149. Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond J a., Fowler G, et al. Comprehensive molecular characterization of human colon and rectal cancer. Nature [Internet]. 2012; 487(7407):330–337. doi:10.1038/nature11252.

150. Chen G, Gharib TG, Huang C-C, Taylor JMG, Misek DE, Kardia SLR, et al. Discordant protein and mRNA expression in lung adenocarcinomas. Mol. Cell. Proteomics. 2002; 1:304–313. doi:10.1074/mcp.M200008-MCP200.

151. Friedman DB, Hill S, Keller JW, Merchant NB, Levy SE, Coffey RJ, et al. Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. Proteomics. 2004; 4(3):793–811. doi:10.1002/pmic.200300635.

152. Nibbe RK, Markowitz S, Myeroff L, Ewing R, Chance MR. Discovery and scoring of protein interaction subnetworks discriminative of late stage human colon cancer. Mol. Cell. Proteomics. 2009; 8(4):827–845. doi:10.1074/mcp.M800428-MCP200.

153. Besson D, Pavageau A-H, Valo I, Bourreau A, Belanger A, Eymerit-Morin C, et al. A Quantitative Proteomic Approach of the Different Stages of Colorectal Cancer Establishes OLFM4 as a New Nonmetastatic Tumor Marker. Mol. Cell. Proteomics. 2011; 10(12):M111.009712–M111.009712. doi:10.1074/mcp.M111.009712.

154. Han C-L, Chen J-S, Chan E-C, Wu C-P, Yu K-H, Chen K-T, et al. An informatics-assisted label-free approach for personalized tissue membrane proteomics: case study on colorectal cancer. Mol. Cell. Proteomics. 2011; 10(4):M110.003087. doi:10.1074/mcp.M110.003087.

155. Jankova L, Chan C, Fung CLS, Song X, Kwun SY, Cowley MJ, et al. Proteomic comparison of colorectal tumours and non-neoplastic mucosa from paired patient samples using iTRAQ mass spectrometry. Mol. Biosyst. 2011; 7(11):2997. doi:10.1039/c1mb05236e.

156. Kang UB, Yeom J, Kim HJ, Kim H, Lee C. Expression profiling of more than 3500 proteins of MSS-type colorectal cancer by stable isotope labeling and mass spectrometry. J. Proteomics. 2012; 75(10):3050–3062. doi:10.1016/j.jprot.2011.11.021.

157. O'Dwyer D, Ralton LD, O'Shea A, Murray GI. The proteomics of colorectal cancer: Identification of a protein signature associated with prognosis. PLoS One. 2011; 6(11). doi:10.1371/journal.pone.0027718.

158. Yang H-Y, Kwon J, Park H-R, Kwon S-O, Park Y-K, Kim H-S, et al. Comparative proteomic analysis for the insoluble fractions of colorectal cancer patients. J. Proteomics. 2012; 75(12):3639–3653. doi:10.1016/j.jprot.2012.04.018.

159. Busund L-T, Richardsen E, Busund R, Ukkonen T, Bjørnsen T, Busch C, et al. Significant expression of IGFBP2 in breast cancer compared with benign lesions. J. Clin. Pathol. 2005; 58:361–366. doi:10.1136/jcp.2004.020834.

160. Huynh H, Zheng J, Umikawa M, Zhang C, Silvany R, Iizuka S, et al. IGF binding protein 2 supports the survival and cycling of hematopoietic stem cells. Blood [Internet]. 2011; 118(12):3236–3243. doi:10.1182/blood-2011-01-331876.

161. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature. 2001; 414(6859):105–111. doi:10.1038/35102167.

162. Tibshirani R. Regression Selection and Shrinkage via the Lasso [Internet]. J. R. Stat. Soc. B. 1994; 58:267–288. doi:10.2307/2346178.

163. McDonald GC. Ridge regression. Wiley Interdiscip. Rev. Comput. Stat. 2009; 1(1):93–100. doi:10.1002/wics.14.

164. Zou H, Hastie T. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat. Methodol. 2005; 67(2):301–320. doi:10.1111/j.1467-9868.2005.00503.x.

165. Bellam N, Pasche B. Tgf-beta signaling alterations and colon cancer. Cancer Treat Res [Internet]. 2010; 155:85–103. doi:10.1007/978-1-4419-6033-7_5.

166. Lampropoulos P, Zizi-Sermpetzoglou A, Rizos S, Kostakis A, Nikiteas N, Papavassiliou AG. TGF-beta signalling in colon carcinogenesis. Cancer Lett. 2012; 314(1):1–7. doi:10.1016/j.canlet.2011.09.041.

167. Bierie B, Moses HL. TGF-beta and cancer. Cytokine Growth Factor Rev. [Internet]. 2006; 17(1-2):29–40. doi:10.1016/j.cytogfr.2005.09.006.

168. Schwab M. Encyclopedia of Cancer. 2012. doi:10.1007/978-3-642-16483-5.

169. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstem AR. A simulation study of the number of events per variable in logistic regression analysis. J. Clin. Epidemiol. 1996; 49(12):1373–1379. doi:10.1016/S0895-4356(96)00236-3.

170. Guo XE, Ngo B, Modrek AS, Lee W-H. Targeting tumor suppressor networks for cancer therapeutics. Curr. Drug Targets [Internet]. 2014; 15(1):2–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24387338.

171. Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. Nat. Rev. Cancer. 2009; 9(1):28–39. doi:10.1038/nrc2559.

172. Murphree AL, Benedict WF. Retinoblastoma: clues to human oncogenesis. Science. 1984; 223(4640):1028–1033. doi:10.1126/science.6320372.

173. Flaherty KT, Robert C, Hersey P, Nathan P, Garbe C, Milhem M, et al. Improved Survival with MEK Inhibition in BRAF-Mutated Melanoma. N. Engl. J. Med. 2012; 367(2):107–114. doi:10.1056/NEJMoa1203421.

174. Saidi SA, Holland CM, Charnock-Jones DS, Smith SK. In vitro and in vivo effects of the PPAR-alpha agonists fenofibrate and retinoic acid in endometrial cancer. Mol. Cancer. 2006; 5:13. doi:10.1186/1476-4598-5-13.

175. Yuan J, Mehta PP, Yin M-J, Sun S, Zou A, Chen J, et al. PF-04691502, a Potent and Selective Oral Inhibitor of PI3K and mTOR Kinases with Antitumor Activity. Mol. Cancer Ther. 2011; 10(11):2189–2199. doi:10.1158/1535-7163.MCT-11-0185.

176. Watanabe M, Sowa Y, Yogosawa M, Sakai T. Novel MEK inhibitor trametinib and other retinoblastoma gene (RB)-reactivating agents enhance efficacy of 5-fluorouracil on human colon cancer cells. Cancer Sci. 2013; 104(6):687–693. doi:10.1111/cas.12139.

177. Moll UM, Petrenko O. The MDM2-p53 interaction. Mol. Cancer Res. 2003; 1(14):1001–1008.

178. Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z, et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science. 2004; 303(5659):844–848. doi:10.1126/science.1092472.

179. Shangary S, Qin D, McEachern D, Liu M, Miller RS, Qiu S, et al. Temporal activation of p53 by a specific MDM2 inhibitor is selectively toxic to tumors and leads to complete tumor growth inhibition. Proc. Natl. Acad. Sci. U. S. A. 2008; 105(10):3933–3938. doi:10.1073/pnas.0708917105.

180. Forastiere A, Koch W, Trotti A, Sidransky D. Head and Neck Cancer. N Engl J Med. 2001; 345:1890–1900. doi:10.1056/NEJMra001375.

181. Shuster MI, Han L, Le Beau MM, Davis E, Sawicki M, Lese CM, et al. A consistent pattern of RINI rearrangements in oral squamous cell carcinoma cell lines supports a breakage-fusion-bridge cycle model for 11q13 amplification. Genes Chromosom. Cancer. 2000; 28(2):153–163. doi:10.1002/(SICI)1098-2264(200006)28:2<153::AID-GCC4>3.0.CO;2-9.

182. Dickson C, Fantl V, Gillett C, Brookes S, Bartek J, Smith R, et al. Amplification of chromosome band 11q13 and a role for cyclin D1 in human breast cancer. In: Cancer Letters. 1995. p. 43–50. doi:10.1016/0304-3835(94)03676-A.

183. Brown LA, Kalloger SE, Miller MA, Shih LM, McKinney SE, Santos JL, et al. Amplification of 11q13 in ovarian carcinoma. Genes Chromosom. Cancer. 2008; 47(6):481–489. doi:10.1002/gcc.20549.

184. Xu J, Tyan T, Cedrone E, Savaraj N, Wang N. Detection of 11q13 amplification as the origin of a homogeneously staining region in small cell lung cancer by chromosome microdissection. Genes Chromosom. Cancer. 1996; 17(3):172–178. doi:10.1002/(SICI)1098-2264(199611)17:3<172::AID-GCC5>3.0.CO;2-1.

185. Gansauge S, Gansauge F, Ramadani M, Stobbe H, Rau B, Harada N, et al. Overexpression of cyclin D1 in human pancreatic carcinoma is associated with

poor prognosis. Cancer Res [Internet]. 1997; 57(9):1634–1637. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9134998.

186. Kitagawa Y, Ueda M, Ando N, Shinozawa Y, Shimizu N, Abe O. Significance of int-2/hst-1 coamplification as a prognostic factor in patients with esophageal squamous carcinoma. Cancer Res. 1991; 51(5):1504–1508.

187. Proctor AJ, Coombs LM, Cairns JP, Knowles MA. Amplification at chromosome 11q13 in transitional cell tumours of the bladder. Oncogene. 1991; 6(5):789–795.

188. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat. Genet. [Internet]. 2013; 45(10):1127–1133. doi:10.1038/ng.2762.

189. Akervall JA, Jin Y, Wennerberg JP, Zätterström UK, Kjellén E, Mertens F, et al. Chromosomal abnormalities involving 11q13 are associated with poor prognosis in patients with squamous cell carcinoma of the head and neck. Cancer. 1995; 76(5):853–859. doi:10.1002/1097-0142(19950901)76:5<853::AID-CNCR2820760520>3.0.CO;2-6.

190. Callender T, El-Naggar AK, Lee MS, Frankenthaler R, Luna MA, Batsakis JG. PRAD-1 (CCND1)/cyclin D1 oncogene amplification in primary head and neck squamous cell carcinoma. Cancer. 1994; 74(1):152–158. doi:10.1002/1097-0142(19940701)74:1<152::AID-CNCR2820740124>3.0.CO;2-K.

191. Williams ME, Gaffey MJ, Weiss LM, Wilczynski SP, Schuuring E, Levine PA. Chromosome 11Q13 amplification in head and neck squamous cell carcinoma. Arch. Otolaryngol. Head. Neck Surg. 1993; 119(11):1238–1243.

192. Bockmühl U, Küchler I, Petersen I. Improved prognostic assessment of head-neck carcinomas by new genetic markers. HNO. 2000; 48(6):451–456.

193. Namazie A, Alavi S, Olopade OI, Pauletti G, Aghamohammadi N, Aghamohammadi M, et al. Cyclin D1 amplification and p16(MTS1/CDK4I) deletion correlate with poor prognosis in head and neck tumors. Laryngoscope. 2002; 112(3):472–481. doi:10.1097/00005537-200203000-00013.

194. Meredith SD, Levine PA, Burns JA, Gaffey MJ, Boyd JC, Weiss LM, et al. Chromosome 11q13 amplification in head and neck squamous cell carcinoma. Association with poor prognosis. Arch. Otolaryngol. Head. Neck Surg. 1995; 121(7):790–794. doi:10.1001/archotol.1995.01890070076016.

195. Borg A, Sigurdsson H, Clark GM, Fernö M, Fuqua SA, Olsson H, et al. Association of INT2/HST1 coamplification in primary breast cancer with hormone-dependent phenotype and poor prognosis. Br. J. Cancer. 1991; 63(1):136–142.

196. Schuuring E, Verhoeven E, van Tinteren H, Peterse JL, Nunnink B, Thunnissen FB, et al. Amplification of genes within the chromosome 11q13 region is indicative of poor prognosis in patients with operable breast cancer. Cancer Res. 1992; 52(19):5229–5234.

197. Kenny FS, Hui R, Musgrove EA, Gee JM, Blamey RW, Nicholson RI, et al. Overexpression of cyclin D1 messenger RNA predicts for poor prognosis in estrogen receptor-positive breast cancer. Clin. Cancer Res. [Internet]. 1999; 5(8):2069–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10473088.

198. Berenson JR, Koga H, Yang J, Pearl J, Holmes EC, Figlin R. Frequent amplification of the bcl-1 locus in poorly differentiated squamous cell carcinoma of the lung. The Lung Cancer Study Group. Oncogene. 1990; 5(9):1343–1348.

199. Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. Nat Rev Cancer [Internet]. 2011; 11(1):9–22. doi:10.1038/nrc2982\rnrc2982 [pii].

200. Schuuring E. The involvement of the chromosome 11q13 region in human malignancies: Cyclin D1 and EMS1 are two new candidate oncogenes - A review. In: Gene. 1995. p. 83–96. doi:10.1016/0378-1119(94)00562-7.

201. Ormandy CJ, Musgrove EA, Hui R, Daly RJ, Sutherland RL. Cyclin D1, EMS1 and 11q13 amplification in breast cancer. Breast Cancer Res. Treat. 2003; 78(3):323–335. doi:10.1023/A:1023033708204.

202. Rodrigo JP, García LA, Ramos S, Lazo PS, Suárez C. EMS1 gene amplification correlates with poor prognosis in squamous cell carcinomas of the head and neck. Clin. Cancer Res. 2000; 6:3177–3182.

203. Gibcus JH, Mastik MF, Menkema L, de Bock GH, Kluin PM, Schuuring E, et al. Cortactin expression predicts poor survival in laryngeal carcinoma. Br J Cancer. 2008; 98(5):950–955. doi:10.1038/sj.bjc.6604246.

204. Hofman P, Butori C, Havet K, Hofman V, Selva E, Guevara N, et al. Prognostic significance of cortactin levels in head and neck squamous cell carcinoma: comparison with epidermal growth factor receptor status. Br. J. Cancer. 2008; 98(5):956–964. doi:10.1038/sj.bjc.6604245.

205. Weaver AM. Cortactin in tumor invasiveness. Cancer Lett. 2008; 265(2):157–166. doi:10.1016/j.canlet.2008.02.066.

206. Freier K, Sticht C, Hofele C, Flechtenmacher C, Stange D, Puccio L, et al. Recurrent coamplification of cystoskeleton-associated genes EMS1 and SHANK2 with CCND1 in oral squamous cell carcinoma. Genes Chromosom. Cancer. 2006; 45(2):118–125. doi:10.1002/gcc.20270.

207. Gibcus JH, Kok K, Menkema L, Hermsen M a., Mastik M, Kluin PM, et al. High-resolution mapping identifies a commonly amplified 11q13.3 region containing multiple genes flanked by segmental duplications. Hum. Genet. 2007; 121:187–201. doi:10.1007/s00439-006-0299-6.

208. Gibcus JH, Menkema L, Mastik MF, Hermsen MA, De Bock GH, Van Velthuysen MLF, et al. Amplicon mapping and expression profiling identify the Fas-associated death domain gene as a new driver in the 11q13.3 amplicon in

laryngeal/pharyngeal cancer. Clin. Cancer Res. 2007; 13(21):6257–6266. doi:10.1158/1078-0432.CCR-07-1247.

209. Huang X, Godfrey TE, Gooding WE, McCarty KS, Gollin SM. Comprehensive genome and transcriptome analysis of the 11q13 amplicon in human oral cancer and synteny to the 7F5 amplicon in murine oral carcinoma. Genes Chromosom. Cancer. 2006; 45(11):1058–1069. doi:10.1002/gcc.20371.

210. Belbin TJ, Singh B, Smith R V, Socci ND, Wreesmann VB, Sanchez-Carbayo M, et al. Molecular profiling of tumor progression in head and neck cancer. Arch. Otolaryngol. Head. Neck Surg. 2005; 131(1):10–18. doi:10.1001/archotol.131.1.10.

211. Ruiz C, Martins JR, Rudin F, Schneider S, Dietsche T, Fischer CA, et al. Enhanced expression of ANO1 in head and neck squamous cell carcinoma causes cell migration and correlates with poor prognosis. PLoS One. 2012; 7(8). doi:10.1371/journal.pone.0043265.

212. Wu H, Reynolds AB, Kanner SB, Vines RR, Parsons JT. Identification and characterization of a novel cytoskeleton-associated pp60src substrate. Mol. Cell. Biol. 1991; 11(10):5113–5124.

213. Wu H, Parsons JT. Cortactin, an 80/85-kilodalton pp60src substrate, is a filamentous actin-Binding protein enriched in the cell cortex. J. Cell Biol. 1993; 120(6):1417–1426. doi:10.1083/jcb.120.6.1417.

214. Van Rossum AGSH, Schuuring-Scholtes E, van Buuren-van Seggelen V, Kluin PM, Schuuring E. Comparative genome analysis of cortactin and HS1: the significance of the F-actin binding repeat domain. BMC Genomics. 2005; 6:15. doi:10.1186/1471-2164-6-15.

215. Weed SA, Karginov A V., Schafer DA, Weaver AM, Kinley AW, Cooper JA, et al. Cortactin localization to sites of actin assembly in lamellipodia requires interactions with F-actin and the Arp2/3 complex. J. Cell Biol. 2000; 151(1):29–40. doi:10.1083/jcb.151.1.29.

216. Weaver AM, Heuser JE, Karginov A V., Lee W lih, Parsons JT, Cooper JA. Interaction of cortactin and N-WASp with Arp2/3 complex. Curr. Biol. 2002; 12(15):1270–1278. doi:10.1016/S0960-9822(02)01035-7.

217. Clark ES, Weaver AM. A new role for cortactin in invadopodia: Regulation of protease secretion. Eur. J. Cell Biol. 2008; 87(8-9):581–590. doi:10.1016/j.ejcb.2008.01.008.

218. Sung BH, Zhu X, Kaverina I, Weaver AM. Cortactin controls cell motility and lamellipodial dynamics by regulating ECM Secretion. Curr. Biol. 2011; 21(17):1460–1469. doi:10.1016/j.cub.2011.06.065.

219. Oser M, Yamaguchi H, Mader CC, Bravo-Cordero JJ, Arias M, Chen X, et al. Cortactin regulates cofilin and N-WASp activities to control the stages of

invadopodium assembly and maturation. J. Cell Biol. 2009; 186(4):571–587. doi:10.1083/jcb.200812176.

220. Ayala I, Baldassarre M, Giacchetti G, Caldieri G, Tetè S, Luini A, et al. Multiple regulatory inputs converge on cortactin to control invadopodia biogenesis and extracellular matrix degradation. J. Cell Sci. 2008; 121(Pt 3):369–378. doi:10.1242/jcs.008037.

221. Sparks a B, Rider JE, Hoffman NG, Fowlkes DM, Quillam L a, Kay BK. Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLCgamma, Crk, and Grb2. Proc. Natl. Acad. Sci. U. S. A. 1996; 93(4):1540–1544. doi:10.1073/pnas.93.4.1540.

222. Tanaka M, Gupta R, Mayer BJ. Differential inhibition of signaling pathways by dominant-negative SH2/SH3 adapter proteins. Mol. Cell. Biol. 1995; 15(12):6829–6837.

223. Mooren OL, Kotova TI, Moore AJ, Schafer DA. Dynamin2 GTPase and cortactin remodel actin filaments. J. Biol. Chem. 2009; 284(36):23995–24005. doi:10.1074/jbc.M109.024398.

224. Schafer DA, Weed SA, Binns D, Karginov A V., Parsons JT, Cooper JA. Dynamin2 and cortactin regulate actin assembly and filament organization. Curr. Biol. 2002; 12(21):1852–1857. doi:10.1016/S0960-9822(02)01228-9.

225. Perrin BJ, Amann KJ, Huttenlocher A. Proteolysis of cortactin by calpain regulates membrane protrusion during cell migration. Mol. Biol. Cell. 2006; 17(1):239–250. doi:10.1091/mbc.E05-06-0488.

226. Yamada H, Abe T, Satoh A, Okazaki N, Tago S, Kobayashi K, et al. Stabilization of actin bundles by a dynamin 1/cortactin ring complex is necessary for growth cone filopodia. J. Neurosci. [Internet]. 2013; 33(10):4514–26. doi:10.1523/JNEUROSCI.2762-12.2013.

227. Tehrani S, Tomasevic N, Weed S, Sakowicz R, Cooper JA. Src phosphorylation of cortactin enhances actin assembly. Proc. Natl. Acad. Sci. U. S. A. 2007; 104(29):11933–11938. doi:10.1073/pnas.0701077104.

228. Kowalski JR, Egile C, Gil S, Snapper SB, Li R, Thomas SM. Cortactin regulates cell migration through activation of N-WASP. J. Cell Sci. 2005; 118(Pt 1):79–87. doi:10.1242/jcs.01586.

229. Martinez-Quiles N, Ho H-YH, Kirschner MW, Ramesh N, Geha RS. Erk/Src phosphorylation of cortactin acts as a switch on-switch off mechanism that controls its ability to activate N-WASP. Mol. Cell. Biol. 2004; 24(12):5269–5280. doi:10.1128/MCB.24.12.5269-5280.2004.

230. Grassart A, Meas-Yedid V, Dufour A, Olivo-Marin JC, Dautry-Varsat A, Sauvonnet N. Pak1 phosphorylation enhances cortactin-N-WASP interaction in clathrin-

caveolin-independent endocytosis. Traffic. 2010; 11(8):1079–1091. doi:10.1111/j.1600-0854.2010.01075.x.

231. Zhu J, Yu D, Zeng XC, Zhou K, Zhan X. Receptor-mediated endocytosis involves tyrosine phosphorylation of cortactin. J. Biol. Chem. 2007; 282(22):16086–16094. doi:10.1074/jbc.M701997200.

232. Kelley LC, Hayes KE, Ammer AG, Martin KH, Weed SA. Cortactin phosphorylated by ERK1/2 localizes to sites of dynamic actin regulation and is required for carcinoma lamellipodia persistence. PLoS One. 2010; 5(11). doi:10.1371/journal.pone.0013847.

233. Lin SY, Makino K, Xia W, Matin A, Wen Y, Kwong KY, et al. Nuclear localization of EGF receptor and its potential new role as a transcription factor. Nat. Cell Biol. 2001; 3(9):802–808. doi:10.1038/ncb0901-802.

234. Hung M-C, Link W. Protein localization in disease and therapy. J. Cell Sci. [Internet]. 2011; 124(Pt 20):3381–92. doi:10.1242/jcs.089110.

235. Wang SC, Hung MC. Cytoplasmic/nuclear shuttling and tumor progression. In: Annals of the New York Academy of Sciences. 2005. p. 11–15. doi:10.1196/annals.1339.002.

236. Pavlou MP, Diamandis EP. The cancer cell secretome: A good source for discovering biomarkers? J. Proteomics. 2010; 73(10):1896–1906. doi:10.1016/j.jprot.2010.04.003.

237. Makridakis M, Vlahou A. Secretome proteomics for discovery of cancer biomarkers. J. Proteomics. 2010; 73(12):2291–2305. doi:10.1016/j.jprot.2010.07.001.

238. Xue H, Lu B, Lai M. The cancer secretome: a reservoir of biomarkers. J. Transl. Med. 2008; 6:52. doi:10.1186/1479-5876-6-52.

239. Karagiannis GS, Pavlou MP, Diamandis EP. Cancer secretomics reveal pathophysiological pathways in cancer molecular oncology. Mol. Oncol. 2010; 4(6):496–510. doi:10.1016/j.molonc.2010.09.001.

240. Andre F, Escudier B, Angevin E, Tursz T, Zitvogel L. Exosomes for cancer immunotherapy. Ann. Oncol. 2004; 15(SUPPL. 4). doi:10.1093/annonc/mdh918.

241. Hoshino D, Kirkbride K, Costello K, Clark E, Sinha S, Grega-Larson N, et al. Exosome secretion is enhanced by invadopodia and drives invasive behavior. Cell Rep. 2013; 5(5):1159–1168. doi:10.1016/j.celrep.2013.10.050.

242. Kahlert C, Kalluri R. Exosomes in tumor microenvironment influence cancer progression and metastasis. J. Mol. Med. 2013; 91(4):431–437. doi:10.1007/s00109-013-1020-6.

243. Griffiths G, Simons K. The trans Golgi network: sorting at the exit site of the Golgi complex. Science. 1986; 234(4775):438–443. doi:10.1126/science.2945253.

244. Hicks SW, Machamer CE. Golgi structure in stress sensing and apoptosis. Biochim. Biophys. Acta - Mol. Cell Res. 2005; 1744(3 SPEC. ISS.):406–414. doi:10.1016/j.bbamcr.2005.03.002.

245. Wlodkowic D, Skommer J, McGuinness D, Hillier C, Darzynkiewicz Z. ER-Golgi network-A future target for anti-cancer therapy. Leuk. Res. 2009; 33(11):1440–1447. doi:10.1016/j.leukres.2009.05.025.

246. Ayala I, Babià T, Baldassarre M, Pompeo A, Fabra A, Kok JW, et al. Morphological and biochemical analysis of the secretory pathway in melanoma cells with distinct metastatic potential. FEBS Lett. 1999; 451(3):315–320.

247. Cao H, Weller S, Orth JD, Chen J, Huang B, Chen J-L, et al. Actin and Arf1-dependent recruitment of a cortactin-dynamin complex to the Golgi regulates post-Golgi transport. Nat. Cell Biol. 2005; 7(5):483–492. doi:10.1038/ncb1246.

248. Salvarezza SB, Deborde S, Schreiner R, Campagne F, Kessels MM, Qualmann B, et al. LIM kinase 1 and cofilin regulate actin filament population required for dynamin-dependent apical carrier fission from the trans-Golgi network. Mol. Biol. Cell. 2009; 20(1):438–451. doi:10.1091/mbc.E08-08-0891.

249. Cao H, Orth JD, Chen J, Weller SG, Heuser JE, McNiven MA. Cortactin is a component of clathrin-coated pits and participates in receptor-mediated endocytosis. Mol. Cell. Biol. 2003; 23(6):2162–2170. doi:10.1128/MCB.23.6.2162-2170.2003.

250. Boeckers TM, Bockmann J, Kreutz MR, Gundelfinger ED. ProSAP/Shank proteins - A family of higher order organizing molecules of the postsynaptic density with an emerging role in human neurological disease. J. Neurochem. 2002; 81(5):903–910. doi:10.1046/j.1471-4159.2002.00931.x.

251. Boeckers TM, Kreutz MR, Winter C, Zuschratter W, Smalla KH, Sanmarti-Vila L, et al. Proline-rich synapse-associated protein-1/cortactin binding protein 1 (ProSAP1/CortBP1) is a PDZ-domain protein highly enriched in the postsynaptic density. Ann. Anat. 2001; 183(2):101. doi:10.1016/S0940-9602(01)80024-8.

252. Okamoto PM, Gamby C, Wells D, Fallon J, Vallee RB. Dynamin Isoform-specific Interaction with the Shank/ProSAP Scaffolding Proteins of the Postsynaptic Density and Actin Cytoskeleton. J. Biol. Chem. 2001; 276(51):48458–48465. doi:10.1074/jbc.M104927200.

253. Naisbitt S, Eunjoon K, Tu JC, Xiao B, Sala C, Valtschanoff J, et al. Shank, a novel family of postsynaptic density proteins that binds to the NMDA receptor/PSD-95/GKAP complex and cortactin. Neuron. 1999; 23(3):569–582. doi:10.1016/S0896-6273(00)80809-0.

254. Jiang Y hui, Ehlers MD. Modeling Autism by SHANK Gene Mutations in Mice. Neuron. 2013; 78(1):8–27. doi:10.1016/j.neuron.2013.03.016.

255. Jung J, Kim JM, Park B, Cheon Y, Lee B, Choo SH, et al. Newly identified tumor-associated role of human Sharpin. Mol. Cell. Biochem. 2010; 340(1-2):161–167. doi:10.1007/s11010-010-0413-x.

256. Rantala JK, Pouwels J, Pellinen T, Veltel S, Laasola P, Mattila E, et al. SHARPIN is an endogenous inhibitor of β1-integrin activation. Nat. Cell Biol. 2011; 13(11):1315–1324. doi:10.1038/ncb2340.

257. Lim S, Sala C, Yoon J, Park S, Kuroda S, Sheng M, et al. Sharpin, a novel postsynaptic density protein that directly interacts with the shank family of proteins. Mol. Cell. Neurosci. 2001; 17(2):385–397. doi:10.1006/mcne.2000.0940.

258. Zhang Y, Huang H, Zhou H, Du T, Zeng L, Cao Y, et al. Activation of nuclear factor κB pathway and downstream targets survivin and livin by SHARPIN contributes to the progression and metastasis of prostate cancer. Cancer. 2014; doi:10.1002/cncr.28796.

259. Kirkbride KC, Hong NH, French CL, Clark ES, Jerome WG, Weaver AM. Regulation of late endosomal/lysosomal maturation and trafficking by cortactin affects Golgi morphology. Cytoskeleton. 2012; 69(9):625–643. doi:10.1002/cm.21051.

260. Cantarelli V V., Takahashi A, Yanagihara I, Akeda Y, Imura K, Kodama T, et al. Cortactin is necessary for F-actin accumulation in pedestal structures induced by enteropathogenic Escherichia coli infection. Infect. Immun. 2002; 70(4):2206–2209. doi:10.1128/IAI.70.4.2206-2209.2002.

261. Kalish LH, Kwong R a, Cole IE, Gallagher RM, Sutherland RL, Musgrove E a. Deregulated cyclin D1 expression is associated with decreased efficacy of the selective epidermal growth factor receptor tyrosine kinase inhibitor gefitinib in head and neck squamous cell carcinoma cell lines. Clin. Cancer Res. 2004; 10(22):7764–7774. doi:10.1158/1078-0432.CCR-04-0012.

262. Dobrinskikh E, Giral H, Caldas YA, Levi M, Doctor RB. Shank2 redistributes with NaPilla during regulated endocytosis. Am. J. Physiol. Cell Physiol. 2010; 299(6):C1324–C1334. doi:10.1152/ajpcell.00183.2010.

263. Dobrinskikh E, Lanzano L, Rachelson J, Cranston D, Moldovan R, Lei T, et al. Shank2 contributes to the apical retention and intracellular redistribution of NaPiIIa in OK cells. Am. J. Physiol. Cell Physiol. [Internet]. 2013; 304(6):C561–73. doi:10.1152/ajpcell.00189.2012.

264. Puthenveedu MA, Lauffer B, Temkin P, Vistein R, Carlton P, Thorn K, et al. Sequence-dependent sorting of recycling proteins by actin-stabilized endosomal microdomains. Cell. 2010; 143(5):761–773. doi:10.1016/j.cell.2010.10.003.

265. Tan KD, Zhu Y, Hiang KT, Rajasegaran V, Aggarwal A, Wu J, et al. Amplification and overexpression of PPFIA1, a putative 11q13 invasion suppressor gene, in head and neck squamous cell carcinoma. Genes Chromosom. Cancer. 2008; 47(4):353–362. doi:10.1002/gcc.20539.

266. Weaver AM. Invadopodia: Specialized cell structures for cancer invasion. Clin. Exp. Metastasis. 2006; 23(2):97–105. doi:10.1007/s10585-006-9014-1.

267. McNiven MA, Kim L, Krueger EW, Orth JD, Cao H, Wong TW. Regulated interactions between dynamin and the actin-binding protein cortactin modulate cell shape. J. Cell Biol. 2000; 151(1):187–198. doi:10.1083/jcb.151.1.187.

268. Böckers TM, Mameza MG, Kreutz MR, Bockmann J, Weise C, Buck F, et al. Synaptic scaffolding proteins in rat brain: Ankyrin repeats of the multidomain Shank protein family interact with the cytoskeletal protein α-fodrin. J. Biol. Chem. 2001; 276(43):40104–40112. doi:10.1074/jbc.M102454200.

269. Riefler GM, Balasingam G, Lucas KG, Wang S, Hsu S-C, Firestein BL. Exocyst complex subunit sec8 binds to postsynaptic density protein-95 (PSD-95): a novel interaction regulated by cypin (cytosolic PSD-95 interactor). Biochem. J. 2003; 373(Pt 1):49–55. doi:10.1042/BJ20021838.

270. Heider MR, Munson M. Exorcising the Exocyst Complex. Traffic. 2012; 13(7):898–907. doi:10.1111/j.1600-0854.2012.01353.x.

271. Henley JR, Cao H, McNiven MA. Participation of dynamin in the biogenesis of cytoplasmic vesicles. FASEB J. 1999; 13 Suppl 2:S243–S247. doi:10.1111/j.1460-9568.2006.04654.x.

272. Ramachandran R. Vesicle scission: Dynamin. Semin. Cell Dev. Biol. 2011; 22(1):10–17. doi:10.1016/j.semcdb.2010.09.001.

273. Baldassarre M, Pompeo A, Beznoussenko G, Castaldi C, Cortellino S, McNiven MA, et al. Dynamin participates in focal extracellular matrix degradation by invasive cells. Mol. Biol. Cell. 2003; 14(3):1074–1084. doi:10.1091/mbc.E02-05-0308.

274. Brahimi-Horn MC, Chiche J, Pouysségur J. Hypoxia and cancer. J. Mol. Med. 2007; 85(12):1301–1307. doi:10.1007/s00109-007-0281-3.

275. Lucien F, Brochu-Gaudreau K, Arsenault D, Harper K, Dubois CM. Hypoxia-induced invadopodia formation involves activation of NHE-1 by the p90 ribosomal s6 kinase (p90RSK). PLoS One. 2011; 6(12). doi:10.1371/journal.pone.0028851.

276. Alexander NR, Branch KM, Parekh A, Clark ES, Iwueke IC, Guelcher SA, et al. Extracellular Matrix Rigidity Promotes Invadopodia Activity. Curr. Biol. 2008; 18(17):1295–1299. doi:10.1016/j.cub.2008.07.090.

277. Clark ES, Whigham AS, Yarbrough WG, Weaver AM. Cortactin is an essential regulator of matrix metalloproteinase secretion and extracellular matrix degradation in invadopodia. Cancer Res. 2007; 67(9):4227–4235. doi:10.1158/0008-5472.CAN-06-3928.

278. The Cancer Genome Atlas Data Portal [Internet]. Available from: https://tcga-data.nci.nih.gov/tcga/.

279. MacConaill LE, Campbell CD, Kehoe SM, Bass AJ, Hatton C, Niu L, et al. Profiling critical cancer gene mutations in clinical tumor samples. PLoS One. 2009; 4(11). doi:10.1371/journal.pone.0007887.

280. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012; 483(7391):603–307. doi:10.1038/nature11003.