Published in final edited form as:

Ann Surg Oncol. 2017 December ; 24(13): 4051–4058. doi:10.1245/s10434-017-6054-5.

Proteomic features of colorectal cancer identify tumor subtypes independent of oncogenic mutations and independently predict relapse-free survival

Callisia N. Clarke, MD¹, Michael S. Lee, MD², Wei Wei, PhD³, Ganiraju Manyam, PhD³, Zhi-Qin Jiang, PhD⁴, Yiling Lu, MD⁵, Jeffrey Morris, PhD³, Bradley Broom, PhD³, David Menter, PhD⁴, Eduardo Vilar-Sanchez, MD, PhD⁶, Kanwal Raghav, MD⁴, Cathy Eng, MD⁴, George J. Chang, MD⁷, Iris Simon, PhD⁸, Rene Bernards, PhD⁹, Michael Overman, MD⁴, Gordon B. Mills, MD, PhD⁵, Dipen Maru, MD¹⁰, and Scott Kopetz, MD, PhD⁴

¹ Department of Surgery, Medical College of Wisconsin, 9200 W. Wisconsin Ave, Milwaukee WI 53226, USA ² Division of Hematology and Oncology, University of North Carolina, 101 Manning Drive, Chapel Hill, NC 27514, USA ³Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, 1400 Pressler St, Houston, TX 77030, USA ⁴Department of GI Medical Oncology, The University of Texas MD Anderson Cancer Center, 1400 Pressler St., Houston, TX 77030, USA ⁵Department of Systems Biology, The University of Texas MD Anderson Cancer Center, 1400 Pressler St., Houston, TX 77030, USA ⁶Department of Clinical Cancer Prevention, The University of Texas MD Anderson Cancer Center, 1400 Pressler St., Houston, TX 77030, USA ⁷Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, 1400 Pressler St., Houston, TX 77030, USA ⁸Agendia BV, Amsterdam, The Netherlands ⁹Division of Molecular Carcinogenesis, Cancer Systems Biology Centre and Cancer Genomics Centre, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands ¹⁰Department of Pathology, The University of Texas MD Anderson Cancer Center, 1400 Pressler St., Houston, TX 77030, USA

Abstract

Background: The directed study of the functional proteome in colorectal cancer (CRC) has identified critical protein markers and signaling pathways. However, the prognostic relevance of many of these proteins remains unclear.

Methods: We determined the prognostic implications of the functional proteome on 263 CRC tumor samples from patients treated at MD Anderson Cancer Center (MDACC) and 462 patients from The Cancer Genome Atlas (TCGA) to identify patterns of protein expression that drive tumorigenesis. 163 validated proteins were analyzed by RPPA. Unsupervised hierarchical clustering of the tumor proteins from MDACC was performed. Clustering was validated using RPPA data from TCGA CRC. Cox regression was used to identify predictors of tumor recurrence.

Results: Clustering revealed dichotomization, with Subtype A notable for a high EMT protein signature while Subtype B was notable for high Akt/TSC/mTOR pathway components. Survival data was only available for the MDACC cohort and was used to evaluate prognostic relevance of these protein signatures. Group B demonstrated worse relapse-free survival (HR 2.11, 95% CI 1.04–4.27, $p=0.039$), though there was no difference in known genomic drivers between the two proteomic groups. Proteomic grouping and stage were significant predictors of recurrence on multivariate analysis. 8 proteins were found to be significant predictors of tumor recurrence on multivariate analysis: Collagen VI, FOXO3a, INPP4B, LcK, phospho-PEA15, phospho-PRAS40, Rad51, phospho-S6.

Conclusion: CRC can be classified into distinct subtypes by proteomic features independent of common oncogenic driver mutations. Proteomic analysis has identified key biomarkers with prognostic importance, however these findings require further validation in an independent cohort.

INTRODUCTION

Colorectal cancer (CRC) remains a leading cause of cancer death in the United States. This is largely due to high tumor heterogeneity and inherent or acquired resistance to therapies. Efforts to classify CRC independent of TNM staging are largely gene-expression based, however, they have only provided limited differentiation in prognosis and treatment response in early stage CRC. Protein expression and phosphorylation accurately reflects cellular function and tumor biology, and may provide unique and meaningful differences as we strive to define CRC taxonomy.

The Cancer Genome Atlas (TCGA) project interrogated CRC tumor biology by performing comprehensive molecular characterization of human colon and rectal cancers. This increased our knowledge of the genetic and epigenetic events involved in CRC pathophysiology, and importantly demonstrated that colon and rectal tumors are molecularly indistinguishable [1]. Zhang et al. [2] further investigated these genomic alterations by assessing tumor phenotype through proteomic analysis of 95 CRC tumor samples. Utilizing liquid chromatography-tandem mass spectrometry (LC-MS/MS), they identified discrete proteomic CRC subtypes and driver pathways critical in CRC tumorigenesis. However, little is known about the prognostic relevance of these proteins and protein signatures. To that end, we analyzed the functional proteome using reverse phase protein array (RPPA) in a large cohort of patients with colorectal cancer in the TCGA and a second cohort treated at the University of Texas M.D. Anderson Cancer Center (MDACC). We sought to determine if CRC tumor subtypes could be identified based on unique protein signatures and investigate the prognostic implications of these functional proteomes.

MATERIALS AND METHODS

Collection of Tumor Protein Lysates and RPPA

Primary tumor tissue from 263 MDACC archived CRC specimen with 30% malignant cells were macrodissected and cellular proteins collected and denatured by 1% sodium dodecyl sulfate (SDS) with β -mercaptoethanol. Protein lysates were diluted in five 2-fold serial dilutions, and serial diluted lysates arrayed on nitrocellulose-coated slides (Grace Bio-

labs, Bend, OR) by Aushon 2470 Arrayer (Aushon BioSystems, Billerica, MA). A total of 5808 array spots were arranged on each slide with positive and negative controls, mixed cell lysates and dilution buffer respectively. Slides were probed with validated primary antibodies plus a biotin-conjugated secondary antibody. Only antibodies with a Pearson correlation coefficient between RPPA and Western blotting of greater than 0.7 were used in RPPA. Antibodies with a single or dominant band on Western blotting were further assessed by direct comparison to RPPA using cell lines with differential protein expression or modulated with ligands/inhibitors or siRNA for phospho structural proteins, respectively. The signal obtained was amplified using a DakoCytomationcatalyzed system (Dako, Carpinteria, CA) and visualized by DAB colorimetric reaction. Slides were scanned, analyzed, and quantified using a customized software, Array-Pro Analyzer software (MediaCybernetics) to generate spot intensity. Dilution curves were fitted with a logistic model (<http://bioinformatics.mdanderson.org/OOMPA>), which fits a single curve using all the dilution series on a slide, with signal intensity as the response variable and dilution steps as the independent variable. The fitted curve is plotted with the signal intensities, both observed and fitted, on the y-axis, and the $\log^2$ concentration of proteins on the x-axis. Protein concentrations were normalized by median polish, corrected across samples by linear expression values using the median expression of all antibody experiments to calculate a loading correction factor for each sample.

Independently, RPPA data from 462 primary tumor tissue in the TCGA archived colorectal tumor bank were also analyzed. Specimen were prepared using the technique described above and run on the same RPPA platform. This cohort was used for validation.

Unsupervised Hierarchical Clustering

RPPA was performed to determine protein levels of 163 phospho- and total proteins in the MD Anderson cohort (MDACC), and of 171 phospho- and total proteins in the TCGA cohort. There were 127 proteins common to both the MD Anderson and TCGA cohorts. The median centered levels of the common proteins were used to perform hierarchical clustering of genes and patients via centroid linkage using GeneCluster 3.0 (<http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm>). Clustering was performed independently in the MDACC and TCGA cohorts. LIMMA linear regression analysis was performed to determine the top discriminating proteins between the clusters in each cohort.

Statistical Analysis

Baseline characteristics were determined for each cluster in the MDACC and TCGA cohorts. Mutation testing for MDACC patients included analysis of *PIK3CA* (codons 542, 545, 1047); *BRAF*; and *KRAS* (codons 12, 13) (Sequenom, San Diego, CA). Fisher's exact test or Pearson chi-square test was used to determine two-sided p-value. Cox univariate analysis was performed for baseline variables. Survival data was only available in the MDACC cohort. Multivariate Cox regression analysis was performed on relapse-free survival and overall survival in 232 patients with Stage II/ III CRC in the MDACC cohort with outcome data. Variables analyzed in survival were age, T stage, N Stage, lymphatic invasion (LVI), BRAF mutation, and proteomic grouping.

To identify independent prognostic proteins, protein expression values were median centered. Cox regression was used for univariate analysis with bootstrap validation. A Beta-Uniform mixture (BUM) model [3] was used to calculate false discovery rates (FDR) to correct for multiple comparisons. Genes with p-value less than or equal to 0.028 were considered significant with a specified FDR of 0.3. Multivariate analysis was performed using significant proteins, clinical and pathologic variables.

RESULTS

Unsupervised hierarchical clustering

Unsupervised clustering resulted in dichotomization of patients into distinct groups, subtypes A and B, each with similar patterns of protein expression between both cohorts (Figure 1). The top 40 significant proteins independently identified for both MDACC and TCGA cohorts are listed in Supplemental Table 1. Of these, 24 proteins were common to both. Subtype A accounted for 57% (150/263) and 60.6% (280/462) of tumors in the MDACC and TCGA cohorts respectively. Subtype A was notable for increased levels of fibronectin, vascular marker CD31 and decreased levels of E-cadherin, indicative of epithelial-mesenchymal transition (EMT) or reactive stroma. Conversely, subtype B demonstrated increased levels of phosphorylated AMP kinase T172, phospho-MTOR, and phosphorylated acetyl-coA carboxylase (ACC) S79, indicating deranged energy sensing and metabolic regulation. Phosphorylated AMP kinase is associated with phosphorylation of the tuberous sclerosis complex and decreased MTOR activation, suggesting that uncoupling of phospho-AMPK and phospho-MTOR in these tumors may facilitate ongoing tumor growth despite AMPK phosphorylation. Subtype B also exhibited increased total and phospho-levels of receptor tyrosine kinases, including HER2, HER3, and BRAF, and demonstrated increased phospho-CRAF S338 and phospho-EGFR Y1173. These findings suggest that increased signaling through the receptor tyrosine kinase pathways may facilitate the upregulation of phospho-MTOR despite an increase in activated AMP kinase.

Baseline characteristics of patients in subtypes A and B are listed in Table 1. In the MDACC cohort, subtype A had more stage II and fewer stage IV patients. Tumors were significantly more right-sided (65.5% vs. 44.5%, $p=0.001$) and less likely to have LVI (58.1% vs. 71.8%, $p=0.023$). Subtype A tumors more frequently displayed mucinous histology (25.3% vs. 12.5%, $p=0.012$) and *BRAF* mutation (15.7% vs. 3.9%, $p=0.011$). There was no significant difference in *PIK3CA* or *KRAS* mutations between subtypes.

TCGA cohort subtype A similarly had more stage II disease (42.0% vs. 34.5%) and less stage IV (12.7% vs. 16.4%), though this did not reach significance. Subtype A trended toward higher proportion of right-sided tumors (46.5% vs. 39.0%, $p=0.11$). As seen in the MDACC cohort, subtype A also had less LVI (39.6% vs. 51.5%, $p=0.017$) and higher *BRAF* mutations (15.5% vs. 0.6%, $p=0.002$). Unlike the MDACC cohort, subtype A had a lower rate of *KRAS* mutations (38.8% vs. 49.7%, $p=0.029$). Notably, a higher rate of *KRAS* mutations was seen in the TCGA cohort overall (43.1% vs. 37.5%). The MDACC cohort only tested for mutations in codons 12 and 13, which may explain the lower mutation incidence. The proportion of patients with microsatellite instability tested in the TCGA was

only 18% (83/462) so the distribution of MSI tumors between subtypes could not be accurately validated.

Survival Analysis

After a median follow up of 71.3 months, 44 patients (18.9%) patients in the MDACC cohort with stage II/III disease experienced recurrence. On multivariate Cox regression analysis of relapse-free survival (RFS) in proteomic subtype B was associated with a significant decrease in RFS with a HR 2.11 (95% CI 1.04–4.28, $p=0.039$) (Figure 2). Stage at diagnosis was also a predictor of RFS with stage III disease conferring a HR of 3.12 (95% CI 1.38–7.09, $p=0.007$), while age, BRAF mutation, and LVI were not significant. On multivariate analysis for overall survival (OS) age (HR 1.08, 95% CI 1.05–1.12, $p<0.001$) and stage at diagnosis (HR 2.66, 95% CI 1.21–5.86, $p=0.015$) were significant predictors of survival, but proteomic grouping was not. (Table 2a).

Prognostic implications of individual proteins

14 proteins were significant ($p \leq 0.05$) independent predictors of recurrence on univariate analysis after bootstrap validation. These proteins were notable for components of the energy balance/MTOR signaling pathway including AMPK, mTOR, FoxO3a. Using a FDR of 0.3, nine proteins remained significant with a $p \leq 0.028$. On multivariate analysis, inclusive of known prognostic clinical and pathology variables, Node positive disease, Collagen VI, FoxO3a, INPP4B, LcK, phospho-PEA15 (Ser116), phospho-PRAS40, Rad51, and phospho-S6 (Ser240–244) remained significant prognostic factors for tumor recurrence (Table 2b). Outcome data in the TCGA cohort was limited so we were unable to validate the proteomic subtype or individual protein prognostic relevance seen in the MDACC cohort.

DISCUSSION

Tumorigenesis results when somatic and/or acquired genetic and epigenetic alterations result in dysregulation of oncogenic signaling pathways affecting subsequent transcriptional activity. Efforts to successfully characterize these aberrant signaling pathways in colorectal cancer through genomic or mRNA based approaches have largely been limited by their imprecision in predicting tumor protein expression [4–6]. For this reason, direct evaluation of the functional proteome provides us the opportunity to evaluate tumor phenotype and link it directly to the expanding knowledge of tumor genomics.

TCGA has successfully performed proteogenomic analysis across several cancer types and identified tumor subtypes that cluster based on protein expression [7]. In doing so, they identified key pancancer pathways and potentially actionable proteins. Similarly, the Clinical Proteomic Tumor Analysis Consortium (CPTAC) sought to investigate the role of proteomic tumor characterization in CRC. Using 95 tumors from the TCGA, they performed proteomic analysis of 124,823 peptides identified with LC-MS/MS and identified five proteomic colorectal subtypes with associated genomic features [2]. Here we present our study of proteomic analysis using RPPA in a larger cohort of 263 colorectal patients treated at MDACC with long term follow up and an emphasis on evaluating the prognostic implications of the CRC functional proteome.

Clustering analysis of 127 proteins from 263 CRC samples in the MDACC cohort demonstrated dichotomization into proteomic subtype A; characterized by high expression of proteins indicative of EMT/reactive stroma, and subtype B; characterized by high expression of RTK proteins. When RPPA data from 462 colorectal tumors in the TCGA study were subjected to the same unsupervised hierarchical clustering these tumors also dichotomized into subtypes A and B with similar protein signatures. BRAF mutation was the only genetic driver found to be associated with a proteomic subtype, with BRAF mutant tumors clustering to proteomic subtype A in both cohorts. These data demonstrate that CRC can reliably and reproducibly be subtyped by unique protein signatures that appear to be independent of known oncogenic drivers. The ability to validate the protein signatures and subtypes seen in our study population with the TCGA data further strengthens our findings. Although the two proteomic subtypes likely represent an oversimplification of the diversity in protein expression patterns in CRC, this dichotomous split allows for early comparisons to start to tease out biologically significant differences.

Our study also uniquely demonstrates the prognostic relevance of the functional proteome. Unfortunately, these findings could not be validated in the TCGA cohort owing to a lack of outcome data. In the MDACC cohort, our data revealed that in patients with Stage II/ III CRC, proteomic clustering predicts tumor recurrence with proteomic subtype A demonstrating superior RFS when compared to subtype B. This improved RFS is observed despite higher expression of an EMT-like signature. The TCGA also reported similar findings in their multiomic evaluation of CRC, where CRC Cluster V, found to be enriched for reactive stroma and EMT signatures, similarly demonstrated superior outcomes when compared to the remaining CRCs clusters [7]. They hypothesized that this improvement in outcomes in Cluster V was due to depleted TP53, APC, and KRAS mutations. We are not able to confirm or refute this hypothesis with our current hotspot mutation data.

Additionally, we identified individual proteins with prognostic implications. Our analysis revealed 8 protein independent predictors of RFS in Stage II/ III CRC, more than identified in the TCGA study [8]. High expression of 2 proteins, FOXO3 and INPP4B, was associated with poor RFS [HR= 2.27 (95% CI 1.15, 4.48, p= 0.018)] and [HR= 3.03 (95% CI 1.15, 4.48, p= 0.018)] respectively. FOXO3, a transcription factor that plays an important role in cell cycle modulation and apoptosis, may play a role in the TCGA reported CRC Cluster VI; a high EMT, cell cycle, apoptotic cluster. INPP4B is a potent suppressor of AKT signaling (Figure 3.) While it was not implicated in CRC in the study TCGA, it is associated with hyperactive PI3K signaling and HER2 overexpression in breast cancer possibly conferring more aggressive disease [9].

The remaining 6 proteins were positive prognostic markers for RFS. Collagen VI, a stromal component, has been widely studied however, its role in tumor biology is still controversial. Collagen VI has been found to promote tumor progression, mitigating its effect through direct activation of the Akt-GSK- β -catenin-TCF/LEF pathway in malignant cells, resulting in increased angiogenic and transcription factors [10]. In the TCGA study however, Collagen VI is well represented in Cluster_V notable for a reactive/EMT signature and better relapse-free survival when compared to other clusters [7]. Additionally, Dennison et al. recently identified a similarly reactive subtype of breast cancer by RPPA notable for high

stromal content similarly associated with favorable outcomes [11]. Our findings are concordant with improved outcomes in high Collagen VI CRC expression [HR= 0.31 (95% CI 0.15, 0.65; p= 0.002)].

Phospho- PEA15, phospho-PRAS40 and phospho-S6 are critical regulators of the energy/Akt/mTOR signaling pathway and are implicated in Clusters I and IV in the TCGA paper. In our study, all three were prognostic for longer RFS and may represent a potentially targetable pathway. Lck tyrosine kinase plays a major role in T-cell receptor signaling and expression correlates closely with tumor-infiltrating lymphocyte volume[12]. The role of the immune system in CRC tumorigenesis is increasingly appreciated and immunotherapy therapies have demonstrated promise in subsets of CRC. Studies have shown a protective role of tumor inflammatory infiltrates in CRC [13, 14]. In our study high levels of Lck expression was also a positive prognostic marker [HR =0.30 (95% CI 0.14, 0.61; p= 0.001)]

Rad51 is a key mediator of DNA repair and is critical for cellular response to DNA damage. Tennstedt et al. demonstrated a strong association between MSI-H CRC and RAD51 overexpression, consistent with what is known about mismatch repair deficiencies [15]. Despite this finding in the same study, there was a negative prognosis associated with RAD51 expression, discordant with the widely accepted good prognosis of MSI-H CRC. Our data confirms a weak association between MSI-H and RAD51 overexpression (r=0.21), and reiterates that both are biomarkers for good prognosis.

CONCLUSION:

Functional proteomic analysis can be used to reliably and reproducibly classify CRC into subtypes based on clustering proteomic features that reflect distinct differences in cellular signaling. These differences appear to be independent of the common oncogenic driver mutations that have been previously described. Additionally, we have identified key proteomic markers with prognostic importance independent of known clinical/pathological variables. These protein signatures may enhance our understanding of colorectal tumorigenesis and represent potential novel therapeutic targets. Further study is needed to validate these prognostic findings in an independent cohort.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institute of Health under award number T325T32CA009599–24, the MD Anderson Cancer Center Sister Institution Network Fund (SINF) Grant and CCSG grant CA016672. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Cancer Genome Atlas N Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012; 487: 330–337.22810696

2. Zhang B, Wang J, Wang X Proteogenomic characterization of human colon and rectal cancer. *Nature* 2014; 513: 382–387.25043054
3. Pounds S, Morris SW. Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *Bioinformatics* 2003; 19: 1236–1242.12835267
4. Segal E, Friedman N, Koller D, Regev A. A module map showing conditional activity of expression modules in cancer. *Nat Genet* 2004; 36: 1090–1098.15448693
5. Margolin AA, Nemenman I, Basso K ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 2006; 7 Suppl 1: S7.
6. Wang K, Saito M, Bisikirska BC Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat Biotechnol* 2009; 27: 829–839.19741643
7. Akbani R, Ng PK, Werner HM A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat Commun* 2014; 5: 3887.24871328
8. Li J, Lu Y, Akbani R TCPA: a resource for cancer functional proteomics data. *Nat Methods* 2013; 10: 1046–1047.
9. Miller TW, Rexer BN, Garrett JT, Arteaga CL. Mutations in the phosphatidylinositol 3-kinase pathway: role in tumor progression and therapeutic implications in breast cancer. *Breast Cancer Res* 2011; 13: 224.22114931
10. Chen P, Cescon M, Bonaldo P. Collagen VI in cancer and its biological mechanisms. *Trends Mol Med* 2013; 19: 410–417.23639582
11. Dennison JB, Shahmoradgoli M, Liu W High Intratumoral Stromal Content Defines Reactive Breast Cancer as a Low-risk Breast Cancer Subtype. *Clin Cancer Res* 2016; 22: 5068–5078.27172895
12. Dalerba P, Maccalli C, Casati C Immunology and immunotherapy of colorectal cancer. *Crit Rev Oncol Hematol* 2003; 46: 33–57.12672517
13. Jass JR. Lymphocytic infiltration and survival in rectal cancer. *J Clin Pathol* 1986; 39: 585–589.3722412
14. Ropponen KM, Eskelinen MJ, Lipponen PK Prognostic value of tumour-infiltrating lymphocytes (TILs) in colorectal cancer. *J Pathol* 1997; 182: 318–324.9349235
15. Tennstedt P, Fresow R, Simon R RAD51 overexpression is a negative prognostic marker for colorectal adenocarcinoma. *Int J Cancer* 2013; 132: 2118–2126.23065657

SYNOPSIS

Colorectal cancers can be classified into distinct subsets by proteomic features reflective of differences in cellular signaling that are independent of common oncogenic driver mutations. Additionally, we have identified protein biomarkers with prognostic importance that may identify pathways for molecular targeting.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1:

Heatmaps showing proteins expressed in rows and patient samples in columns. : a) Unsupervised clustering of tumors in the MDACC cohort resulted in dichotomization: the yellow bar corresponds to CRC subtype A, blue bar corresponds to subtype B. b) Subtypes were then validated in the TCGA cohort with similar dichotomization and clustering of proteins: again the yellow bar corresponds to Group A, and the blue bar corresponds to Group B.

Figure 2:
Proteomic subtype is prognostic for stage II-III CRC. Subtype B was associated with worse Relapse free survival (RFS) on multivariate cox regression analysis.

Figure 3:

Protein interaction network of prognostic proteins in CRC: INPP4B, FOXO3, PEA15, PRAS40 and pS6 are all involved in the Akt/mTOR signaling cascade. RAD51 is critical in DNA double strand break repair. Lck regulates T-cell signaling. Collagen type VI is an extracellular matrix protein that recruits macrophages and endothelial cells.

Table 1:

Clinicopathological characteristics and mutation incidence of patients in a) MDACC cohort and b) TCGA cohort. $p < 0.05$ was considered significant.

A.			
MDACC	Subtype A n= 150 (%)	Subtype B n= 150 (%)	p-value
Mean Age	63	61	0.13
Stage			
I	0 (0)	0 (0)	0.04
II	69 (46)	48 (43)	
III	71 (47)	45 (40)	
IV	10 (7)	18 (16)	
Site			
Proximal	97 (66)	49 (45)	0.001
Distal	51 (35)	61 (56)	
MSI - high	12 (18)	10 (13)	0.33
LVI present	86 (58)	79 (72)	0.02
PIK3CA mut	18 (16)	11(14)	0.8
BRAF mut	18 (16)	3 (4)	0.01
KRAS mut	44 (38)	28 (37)	0.79

B.			
TCGA	Subtype A n= 280(%)	Subtype B n= 182 (%)	p-value
Mean Age	66	69	0.013
Stage			
I	43 (16)	28 (16)	0.38
II	116 (42)	61 (35)	
III	82 (30)	59 (33)	
IV	35 (13)	29 (16)	
Site			
Proximal	128 (47)	71 (39)	0.11
Distal	147(54)	111 (61)	
MSI - high	10 (15)	0 (0)	0.11
LVI present	101 (40)	84 (52)	0.02
PIK3CA mut	57 (23)	30 (18)	0.24

B.			
TCGA	Subtype A n= 280(%)	Subtype B n= 182 (%)	p-value
BRAF mut	38 (16)	9 (1)	0.002
KRAS mut	95 (39)	81 (49)	0.03

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2a:

On multivariate analysis only Stage at diagnosis and Proteomic Subtype were significant predictors of relapse free survival. Proteomic grouping was not prognostic of Overall Survival.

Multivariate Analysis of Relapse Free Survival				
	HR	Lower Limit	Upper Limit	p-value
Node Positive (Stage)	3.12	1.38	7.09	0.01
Lymphatic Invasion	0.54	0.24	1.202	0.13
Proteomic Subtype A	0.48	0.23	0.96	0.04
Proteomic Subtype B	2.11	1.04	4.28	0.04
Age >50 years	1.023	0.99	1.05	0.13
Multivariate Analysis of Overall Survival				
Node Positive (Stage)	2.66	1.21	5.86	0.02
Lymphatic Invasion	0.64	0.3	1.37	0.24
Proteomic Grouping	0.77	0.39	1.54	0.46
Age >50 years	1.08	1.05	1.12	<0.001

Table 2b:

Eight Proteins were found to be independent predictors of recurrence on multivariate analysis.

Protein	Protein Function	Hazard Ratio	p- value	95% CI
FOXO3a	Apoptosis & cell cycle regulation	2.7	0.018	1.15, 4.48
INPP4B	Suppresses Akt signaling	3.03	0.004	1.43, 6.4
Collagen VI	Extra cellular matrix component	0.31	0.002	0.15, 0.65
LcK	T-cell signaling	0.3	0.001	0.14, 0.61
pPEA15	Akt driven anti-apoptotic	0.33	0.003	0.16, 0.67
pPRAS40	Akt/mTOR signaling	0.27	0.001	0.12, 0.57
Rad51	DNA repair	0.39	0.01	0.19, 0.80
pS6	Cell cycle progression	0.43	0.01	0.22, 0.83