Molecular pathways associated with Kallikrein 6 overexpression in colorectal cancer

Ritu Pandey^{1,2*}, Muhan Zhou³, Yuliang Chen³, Dalila Darmoul^{4,5}, Conner C. Kisiel²,

Valentine N. Nfonsam⁶, Natalia A. Ignatenko^{1,2*}

¹Department of Cellular and Molecular Medicine, University of Arizona, Tucson, AZ,

²University of Arizona Cancer Center, Tucson, AZ, 85724

³Bioinformatics Shared Resource, University of Arizona Cancer Center

⁴Institut National de la Santé et de la Recherche Médicale (INSERM), Hôpital Lariboisière,
F-75010 Paris, France.

⁵Université de Paris, UMR1275, F-75010 Paris, France.

⁶Department of Surgery, Section of Surgical Oncology, University of Arizona, Tucson, AZ

* Correspondence: ritu@email.arizona.edu; nai@email.arizona.edu

Key words: colorectal cancer, kallikrein 6, TCGA, gene set enrichment analysis, K-RAS-
oncogene, regulatory pathways, organoid culture

List of abbreviations: KLK6: Kallikrein 6; CRC: colorectal cancer; GDC: Genomics Data
Commons; GO: Gene Ontology; TCGA: The Cancer Genome Atlas; GEO: Gene
Expression Omnibus; KEGG: Kyoto Encyclopedia of Genes and Genomes; DEGs:
differentially expressed genes; VST: Variance Stabilizing Transformation; OS: Overall
Survival.

28 Abstract

29 Colorectal cancer (CRC) remains one of leading causes of cancer-related death worldwide.
 30 The high mortality of CRC is related to its ability to metastasize to distant organs. Current
 31 management of the CRC disease is limited due to the need of more extensive molecular
 32 characterization of tumor biomarkers. Kallikrein 6 (KLK6) is a member of the fifteen-gene
 33 family of kallikrein-related peptidases. KLK6 is overexpressed in CRC and contributes to
 34 cancer cell invasion through its proteolytic degradation of the extracellular matrix and
 35 recently discovered regulation of metastasis signaling pathways.

36 The goal of this study was to evaluate the clinical features, CRC molecular subtypes,
 37 mutational and gene expression patterns of the CRC tumors with overexpressed KLK6 in
 38 order to identify KLK6-associated markers for the CRC prognosis and treatment. RNA-Seq
 39 data from the CRC patients with a significantly elevated KLK6 transcript levels (KLK6-high
 40 group) (Z-score higher than -1.96) were identified in the Cancer Genome Atlas (TCGA)
 41 database and processed for clinical, molecular evaluation and bioinformatic analysis, using
 42 Gene Ontology (GO), Phenotype and Reactome enrichment and protein interaction methods.

43 KLK6-high cases had a distinct spectrum of mutations in titin (TTN), APC, K-RAS and MUC
 44 16 genes. Differentially expressed genes (DEGs) unique to KLK6-high group were clustered
 45 to regulatory pathways controlling the cell signaling, extracellular matrix organization, and
 46 cell communication regulation. The members of kallikrein family (KLK7, KLK8, KLK10),
 47 keratins (KRT6A, 6B, 15, 16, 19, 80), extracellular matrix proteins (integrin 4B), small
 48 proline rich repeat proteins (SPRRs), S100A families, protein trafficking genes (SYL1) and
 49 signaling genes within TGF- β , FOS and Ser/Thr protein kinase pathways were recognized as
 50 the top KLK6-interaction partners. Expression of selected KLK6-associated genes was
 51 validated in a subset of paired normal and tumor CRC patient-derived organoid cultures.

52 The performed analyses identified KLK6 itself and a set of genes, which are co-expressed

53 with KLK6, as potential clinical biomarkers for the management of the CRC disease.

54

55 **Introduction**

56 Colorectal cancer (CRC) manifests from the accumulated defects in oncogenes and tumor
57 suppressor genes, which alter downstream molecular pathways regulating cell proliferation,
58 differentiation, and apoptosis (Wong *et al.*, 2007; Fearon, 2011). One of the key hallmarks of
59 cancer is an ability of transformed cells to invade surrounding tissues, which occurs through
60 induction of cell- and tissue-specific signaling pathways [1]. Proteases, which catalyses the
61 hydrolysis of peptide bonds, have long been associated with colon cancer progression because
62 of their ability to degrade extracellular matrices. Tissue kallikrein-related peptidases (KLK)
63 are a subgroup of serine proteases, known to serve a variety of functions, ranging from the
64 normal physiological processes to different disease conditions, including cancer [2]. Several
65 members of kallikrein family are overexpressed in colon cancer patients, i.e. KLK4, KLK6,
66 KLK7, KLK10 and KLK14 [3-6]. Because of association of kallikreins with malignancy they
67 have been proposed as potential diagnostic/prognostic and therapeutic markers [7].
68 Particularly, KLK6 mRNA levels correlated with serosal invasion, liver metastasis, advanced
69 Duke's stage, and overall poor patients' prognosis [8]. Presence of KLK6 transcripts in lymph
70 nodes of CRC patients was associated with the shorter average survival time after surgery and
71 was more sensitive indicator of poor prognosis than carcinoembryonic antigen (CEA) [9]. As
72 we previously reported, KLK6 expression in colon cancer can be induced by oncogenic *K*-
73 *RAS*, which is the one of the major colon cancer driver genes [9-12].
74 The aim of our study was to analyze Tissue Cancer Genome Atlas (TCGA) RNA-Seq data
75 from the colorectal cancer samples with high KLK6 expression with the goal to define
76 KLK6-specific gene signature in the CRC for identification of potential markers of tumor

77 progression in colorectal metastatic disease. Our analysis revealed a set of genes, which were
78 differentially expressed in the CRC tumors, which overexpressed KLK6. The set includes
79 genes involved in the regulation of cell signaling, cell-cell communication and proteolysis.
80 The results were confirmed in another GEO dataset (GSE39582). We also analyzed the
81 expression of KLK6 and selected co-expressed genes in the subset of colonic organoids
82 developed from the patients' surgical tissue (paired normal and tumor cultures).

83

84 **Materials and Methods**

85 **RNA-Seq Data Processing**

86 Colorectal Cancer samples from The Cancer Genome atlas (TCGA) were downloaded from
87 Genomic Data Common Website (GDC at <https://gdc.cancer.gov/>). A total of 480 tumor
88 samples were assayed for transcription. We used the GDC data transfer tool client and GDC
89 API to download all the RNA-seq raw counts data, metadata and available clinical data. Data
90 was analyzed using R (v 3.) and open source methods implemented in R. Raw HTSEQ counts
91 data was normalized using Variance Stabilizing Transformation (VST) method [13] and data
92 was processed further. Z scores were calculated for KLK6 expression across samples and
93 high (N=16) and low (N=7) groups were defined by ≥ 1.96 and < -1.96 cutoff values, they
94 represent statistically significant outliers. Differential analysis was done using the DESEQ2
95 [14]. Pearson correlation coefficient was used to identify genes that correlate with KLK6
96 amongst high and low samples. GEO dataset GSE39582 was downloaded and the array data
97 for 566 colon tumor samples was analyzed for high KLK6 expression. 30 samples were found
98 to be outliers with high KLK6 expression (z score > 1.96). Genes results from TCGA high
99 KLK6 tumors were compared to high KLK6 tumors in GSE39582.

100 **Survival analysis of samples with overexpressed KLK6**

101 The normalized data was combined with clinical data for high KLK6 samples (16), low
102 KLK6 (7) vs. other groups (N=457) for TCGA data. The overall survival (OS) in KLK6 high
103 samples compared to other samples was done by utilizing Kaplan-Meier survival plot. GEO
104 dataset was divided into top and bottom quartiles based on KLK6 expression and for high
105 KLK6 samples (30) vs others (536). Logrank test was applied to assess the statistical
106 significance between the survival curves. The survdiff function under rms library in R was
107 used to perform the log-rank test.

108 **Enrichment analysis**

109 Differentially expressed genes (DEG) were analyzed for Gene Ontology using enrichGO
110 from Cluster Profiler and Reactome pathways gene enrichment analysis using
111 enrichPathway from ReactomePA in Bioconductor. Protein interaction data was analyzed
112 using String data resource [15]. The list of differential protein coding genes between high and
113 low KLK6 were searched against Stringdb. The network was filtered for first interacting
114 partners of kallikreins and their associations. Plots: All plots and heatmaps were generated
115 using gplot2 library and other data visualization packages in R.

116 **CRC patient-derived organoid cultures**

117 The 3D cultures were generated from the fresh surgical material collected as part of the
118 University of Arizona Cancer Center (UACC) Tissue Acquisition and Molecular Analysis
119 Shared Resource (TACMASR) Biorepository efforts. The CRC patients were consented for
120 the tissue collection and all individual patient-related reports and data were de-identified
121 before processing for the tissue biobanking and developing of organoid cultures. Because all
122 information related to the used surgical material was de-identified and did not include any
123 protected health information (PHI), it is not considered a clinical research according on the
124 Human Subjects Protection Program (HSPP) determination. Tissue collection was approved
125 by the Biorepository Oversight Committee at the UACC. The paired normal and tumor

specimens, retrieved from the resected surgical material were verified by a certified pathologist. The collected material was placed in the cold Advanced DMEM media and transported to the lab on ice, where they were processed for establishment of organoid cultures as described elsewhere [16]. The organoid cultures generated from paired normal and tumor tissue samples of eight CRC patients were analysed in this study.

Quantitative Reverse-Transcription Polymerase Chain Reaction

Organoid cultures were washed with AdDMEM/F12 media (Thermo Fisher Scientific. Onc. Carlsbad, CA) once and processed for total RNA isolation using RNeasy® Mini Kit (Qiagen GmbH, Germany, Cat. 74104) according to the manufacturer instructions. Reverse transcription to produce cDNA template was completed using the Applied Biosystems High Capacity cDNA Reverse Transcription Kit (Part #4368814). Quantitative Reverse-Transcription Polymerase Chain reaction (qRT-PCR) was performed using TaqMan® probes (Applied Biosystems, Thermo Fisher Scientific, Inc.) specific for the mRNAs of interest: KLK6 (Hs00160519_m1), KRT6A (Hs00749101_s1), KRT6B (Hs04194231_s1), FOSL1 (Hs04187685_m1), MYC (Hs00153408_m1), HMGA2 (Hs04397751_m1). 0.2 µg of total RNA was reversed transcribed into cDNA in a 20 µL reaction with random hexamers under thermal condition recommended by the protocol. Real-time PCR amplification was performed with the ABI PRISM 7700 SDS instrument (Applied Biosystems, Life Technologies, Inc.), under the universal thermal cycling conditions recommended by the Assay-on-Demand products protocol. Negative controls without template were included in each plate to monitor potential PCR contamination. The expression of genes was tested in triplicate and each reaction was run in duplicate. To determine the relative expression level of each target gene, the comparative *CT* method was used. The *CT* value of the target gene was normalized by the endogenous reference

150 β 2-microglobulin (β 2M, FAM (Hs99999907_m1)). The relative expression of each target
151 gene was calculated via the equation $2^{-\Delta C_T}$ where $\Delta C_T = C_{T(\text{target})} - C_{T(\text{endogenous control})}$.

152 **Enzyme-Linked Immunosorbent Assay (ELISA) for KLK6**

153 An ELISA kit for the detection of human KLK6 was obtained from Boster Immunoleader
154 (Pleasanton, CA). The assay was performed according to manufacturer's protocol.
155 Conditioned media was collected from the cultured organoids on days 2,4,7, and 10 after
156 subculture. KLK6 levels were expressed as picograms per milliliter of media. The plate
157 was read at 490 nm within 30 min of assay ending on a Synergy 2 Multi-Detection
158 Microplate Reader (Bio-Tek Instruments, Inc., Winooski, VT).

159

160 **Results**

161 **Pattern of Kallikrein related peptidases family expression in CRC** 162 **patients**

163 In this study we initially analyzed the relative expression pattern of the members of
164 kallikrein-related peptidase genes family in 480 CRC tumor samples from TCGA. We noted
165 that the transcript levels of a majority of kallikrein-related peptidases vary significantly in the
166 normal and tumor samples amongst the CRC patient's. (**Fig. 1A**). The normalized transcript
167 levels of KLK6,7,8, 10, 11 and 12 were elevated in tumor samples, compared to normal ones,
168 the levels of KLK2,5,9, and 14 were not altered, and levels of KLK2,3,4, and 13 were found
169 to be lower in tumors, compared to normal tissues. Focusing on KLK6, we further selected
170 the samples that have unusually high or low *KLK6* expression. We normalized row data using
171 Variance Stabilizing Transformation (VST) method [13] and identified outlier samples with
172 $> \pm 1.96$ z score as samples with high and low *KLK6* expression, respectively. Using these
173 criteria, sixteen samples were identified as high KLK6 expressers and seven samples were

174 selected as low KLK6 expressers. These groups were named as KLK6-high and KLK6-low
 175 groups, respectively. Relative expression of other kallikreins family members was plotted for
 176 these two groups (**Fig. 1B** upper panel - KLK6 -high group, and **Fig. 1B** lower panel -
 177 KLK6-low group). The above analysis revealed the distinct pattern of expression of KLK
 178 family members in KLK6-high group, with particularly high KLK10 co-expression, followed
 179 by KLK7, and KLK8. We also noted that KLK10 transcript levels were significantly higher
 180 than KLK6, KLK7 and KLK8 in normal tissues (**Fig. 1A**), and KLK10 was elevated in tumor
 181 samples in KLK6-low group but to the lesser degree than in KLK6-high samples (**Fig. 1B**
 182 lower panel).

Fig. 1. Expression of all members of Kallikrein related peptidase family in colon adenocarcinomas TCGA patient samples. (A) Box plot shows expression of KLK gene family normalized by Variance Stabilizing Transformation (VST) method in colon tumors and their relative expression in normal tissues. (B) Box plots of KLK family from the CRC patient samples that have been identified as KLK6-high group and KLK6-low group.

Upper panel: Box plot of expression of all Kallikrein related peptidases in samples with high KLK6 expression (KLK6-high group) and **Lower panel:** Box plot of expression of all Kallikrein related peptidases in samples with low KLK6 expression (KLK6-low group)

183 **Clinical and molecular characterization of KLK6-high expressing** 184 **samples**

185 The clinical, gender, histopathological and molecular characteristics for TCGA patients with
 186 KLK6-high and KLK6-low (expression are presented in **Table 1**). The tumors in both groups
 187 were identified as colon adenocarcinomas. In KLK6-high group, the number of patients with
 188 the advanced stages of the disease constituted more than third of cases (stage III and stage IV:
 189 18.8% and 25%, respectively). In contract, more than half patients in KLK6-low group had
 190 the early stages of the disease (stage I and stage II: 28.8% and 43%, respectively), with no

191 stage IV patients.

Table 1. Clinical and molecular analysis of the TCGA colon adenocarcinoma KLK6-high and KLK6-low groups.

TCGA dataset		Percentage	
Cases		KLK6-high group (n=16)	KLK6-low group (n=7)
Gender	female	44	71
	male	56	29
Tumor stage	Stage I	19	28.6
	Stage II A	37	28.6
	Stage II B	0	14.29
	Stage III	6.25	0
	Stage III B	6.25	14.29
	Stage III C	6.25	14.29
	Stage IV	19	0
	Stage IV A	6.25	0
	Metastasis ≥M1	25	0
Lymph node positive		44	28.6
Molecular subtype	MSS	43.75	42.85
	MSI-L	31.25	14.2
	MSI-H	25	42.85
Mutations in >50%	APC	75	100
	Titin (TTN)	75	nd
	K-RAS	68.75	Nd
	MUC16	56.25	Nd
	P53	50	66.67

MSS-microsatellite stable; MSI-microsatellite instable; MSH-H- MSI-high; MSI-L - MSI-low; nd-not detected

192

193 **Molecular subtypes in KLK6-high and KLK6-low CRC TCGA samples**

194 The CRC tumors with high and low KLK6 expression were stratified into the microsatellite
195 stable (MSS) and microsatellite instable (MSI) molecular subtypes as defined in other TCGA
196 studies [17]. Both groups had comparable numbers of MSS tumors. Stratification of MSI
197 tumors with high or low number of mutations (MSI-H) and MSI-L, respectively) did not
198 identify any correlation with KLK6 expression (**Table 1**). Similarly, no association or

correlation with any particular CMS (Consensus Molecular Subtype) [18] (**S1 Fig.**) was found for high KLK6 expression in tumors.

Mutation frequencies in KLK6-high and KLK6-low CRC samples

We further determined the spectrum of somatic gene mutations in the KLK6-high and KLK6-low groups by analyzing the available TCGA DNA sequencing results. As demonstrated in **Table 1** and **Fig. 2A**, (based on **S1 Table A** data), the most frequent cancer-related mutations observed in KLK6-high group were mutations in the titin (TTN) and APC genes (found in 75% of KLK6-high tumors), followed by mutations in K-RAS oncogene (68.75%) and MUC16 (mutated in 56.25 %). The mutations in p53 tumor suppressor gene were detected in half samples with high KLK6 expression (**Table 1**).

In contrast, 100% of tumors with low KLK6 expression had APC mutations 66.67% of tumor had p53 mutations. No mutations in TTN or K-RAS were detected in KLK6-low group (**Table 1, Fig. 2B** and **S1 Table B**).

Fig. 2. Mutation frequencies in the colon adenocarcinoma TCGA samples with the differential KLK6 transcript level. (A) KLK6-high group, (B) KLK6-low group.

We also assessed the status of KLK6 gene in KLK6-high and KLK6-low samples and did not find any changes in KLK6 gene copy number in these groups. Overall, twenty-five tumors within the remaining 469 cases of the TCGA cohort had alteration in KLK6 gene, with 80% of the cases carrying the missense mutations in KLK6 coding region and 20% of the cases having deletions in the 3'UTR or splice region of KLK6 (**S1 Table C**). This analysis indicates that overexpression of KLK6 in the CRC samples from the TCGA can't be attributed to alterations in the KLK6 gene copy number or gene's structural aberrations.

Survival analysis

Kaplan-Meier survival curves of patients with high KLK6 from TCGA demonstrated that they had overall poor survival compared to other patient samples, although the *p* value did not

reach significance (**Fig. 3A**). Patients with a significantly low KLK6 expression (KLK-low group) had no deaths reported. Because of the limited number of samples with the significantly elevated levels of KLK6 in TCGA, we evaluated the OS rates of the CRC cases from the GEO dataset, GSE39582 of 566 samples. The clinical, histopathological and molecular characteristics of the samples in this dataset are summarized in **S2 Table**. Analysis of GEO samples with the high KLK6 expression levels (30 samples with z score > 1.96) showed a statistically significant decrease in OS rates for these patients (**Fig. 3B**) compared to all other patients (536 samples) (P=0.0066). Dividing the GEO samples into top and bottom quartiles based on KLK6 expression also showed that top quartile has poor survival compared to other groups (P = 0.01) (**Fig. 3C**).

Fig. 3. Overall Survival analysis of patients based on KLK6 gene expression from TCGA and GEO datasets. (A) Kaplan-Meier Plot of colon adenocarcinoma samples from TCGA. KLK6 High (16 samples), KLK6 low (7 samples) and others were plotted for Overall Survival. (B) Kaplan-Meier plot of GEO dataset for KLK6 high samples vs others. (C) Kaplan-Meier plot of GEO dataset: based on KLK6 gene expression, data was plotted for upper quartile, bottom quartile and intermediate expression for overall survival.

KLK6 expression in left-and right sided CRC cases

The right-sided and left-sided colon cancer differ in their clinical and molecular features, and the right-sided colon cases have the worse prognosis [19]. We stratified TCGA CRC cases by the KLK6 expression in the ascending (right) and descending (left) portions of the colon. The KLK6 expression was highly variable within the right, left and transverse colons with no significant difference between tumor locations. The expression of KLK6 was significantly higher in all tumor areas compared to the normal tissues (p=0.001, normal vs descending tumors, p=1.6x10⁻⁶, normal vs ascending tumors) (**Fig. 4**).

Fig. 4. Analysis of KLK6 expression in the right side, left side and transverse CRC cases from TCGA. Box plot shows distribution of KLK6 expression across site of colon adenocarcinoma tumor and normal samples.

240 **Differential analysis of samples with overexpressed KLK6**

241 A differential expression analysis was done to assess the differences in gene expression
 242 between the KLK6-high and KLK6-low patient samples. DESEQ2 analysis revealed 620
 243 differentially expressed genes (DEGs) between these groups (**S3 Table A**). The DEGs set
 244 contained 475 protein coding genes and 145 long non-coding RNAs as well as antisense and
 245 uncharacterized transcripts. Cluster analysis of altered gene expression between the KLK6-
 246 high and KLK6-low groups showed the KLK6-dependend pattern of gene expression across
 247 the groups (**Fig. 5A**). Further differential expression analysis was done between KLK6-high
 248 group and the rest of the CRC cases to determine whether any specific genes were altered in
 249 KLK6-high group compared to all other samples in the tumor dataset. In the list of all DEGs
 250 we identified 236 protein-coding genes which were altered in KLK6-high group (**S3 Table**
 251 **B**). A heatmap of these 236 genes, shown in **Fig. 5B**, demonstrates the pattern of KLK6-
 252 dependent gene expression in protein coding genes. Depicted in **Fig. 5C** Venn diagram
 253 demonstrates no overlapping in gene expression between KLK6 -high cases and KLK6-low
 254 cases, with 171 genes upregulated genes and 54 downregulated genes in KLK6 high group
 255 compared to KLK6-low samples and other samples. Therefore, this set of genes is uniquely
 256 attributed to the high KLK6 expression in CRC samples. Correlation analysis of KLK6 across
 257 high and low KLK6 groups identified a set of transcripts that strongly correlate with KLK6
 258 expression and showed same expression trend (correlation of 0.6 and p value <0.05 were used
 259 as significance of association). The correlation plot of top genes, which co-express positively
 260 or negatively with KLK6 overexpression is shown in **Fig. 5D**. The highly associated genes
 261 include the members of kallikrein family, KLK7, KLK8, KLK10, and KLK11, keratins 80,

262 6A, 6B, 19, 15, 19, extracellular matrix proteins, vesicle-mediated protein trafficking and
 263 with signaling genes, e.g. BMP (TGF- β signaling pathway), GJB3 (Gap Junction protein beta
 264 3), and signaling genes and transcription factors (**S3 Table C**).

Figure 5. Clustering of the colon adenocarcinoma TCGA samples with differential expression of KLK6. **(A)** Heatmap showing differentially expressed protein-coding genes in KLK6-high vs KLK6-low group samples. **(B)** Heatmap showing a group of 236 differentially expressed protein-coding genes, which are common in high KLK6 vs low KLK6 group and high KLK6 vs other samples. **(C)** Venn diagram showing number of genes which were found to be up and down when KLK6-high group was compared with all other CRC cases. **(D)** Correlation plot of genes co-expressed with KLK6 in KLK6-high group.

265 Differential expression analysis of high KLK6 samples (30 samples) and other samples (536)
 266 in GEO dataset (total of 566 samples) identified 691 significantly altered genes. When we
 267 compared the set of 236 KLK6 high specific genes from TCGA dataset with 691 genes from
 268 GEO dataset, we found 121 common and the functionally significant genes between the two
 269 comparisons (**S3 Table D**).

270 **Pathways and significant cellular functions that associated with**

271 **KLK6 overexpression**

272 We next applied Gene Ontology (GO) and pathway enrichment analyses to find which GO
 273 terms are over-represented and under-represented in TCGA KLK6-high DEGs set. In the
 274 KLK6-high group we found the most significant association (P value <0.05) of GO terms for
 275 biological process (**Fig. 6A**) with the genes involved in response to stimulus, regulation of
 276 signaling and regulation of cell communication. **Fig. 6B** demonstrates significantly enriched
 277 GO cellular component genes in KLK6-high group involving the plasma membrane, both
 278 basolateral and basement, gap junctions and connexon complex as well as membrane-bound
 279 vesicle and extracellular exosome. Protein binding genes and peptidase activity (**Fig. 6C**)

280 were significantly enriched for GO molecular function terms (**S4 Table A**).

281 Reactome pathways analysis [20] was applied to search for pathways and several gene sets in

282 Type I hemidesmosome assembly, Laminin interactors, extracellular matrix organization,

283 degradation of the extracellular matrix, assembly of collagen fibrils, Anchoring fibril

284 formation and Activation of Matrix metalloproteinases were found to be differentially

285 enriched in KLK6 high expression samples (**Fig. 6D, S4 Table B**).

Figure 6. Analysis of Gene Ontology (GO) and signaling pathways associated with KLK6 overexpression in TCGA colon adenocarcinoma samples. Analysis of DEGs for GO shows (A) Biological processes. (B) Cellular Components and C) Molecular Functional gene sets that are significantly enriched in high expressing KLK6 group. (D) Reactome pathways that are enriched in DEGs.

286 Finally, the network of kallikreins interacting genes (KLK6,7,8,10,13) and their first

287 interacting partners was identified using Stringdb (**Fig. 7 and S4 Table C**). We report here the

288 top 34 extracellular matrix proteins with protein binding function. The list includes

289 metalloproteases, serine proteases, keratins and proteins involved in keratin differentiation

290 (**S4 Table D**).

Fig. 7. Protein Interaction network of kallikriens and differentially expressed genes found in high KLK6 samples using Stringdb interaction dataset.

291 **Analysis of expression and secretion of KLK6 in the CRC patient-**

292 **derived organoid cultures**

293 We analyzed RNA levels of *KLK6* and co-expressers *KRT6A*, *KRT6B* and *FOSL1* in the

294 paired normal and tumor organoid cultures established from the surgical material of the CRC

295 patients, which were previously untreated. Additionally, because of the importance of *c-MYC*

296 oncogene in colon tumorigenesis [21] and the association of *HMGA2* with *KLK6* expression

297 and poor disease outcome in the CRC [22, 23], *c-MYC* and *HMGA2* transcripts levels were

measured in these culture. Clinicopathological characteristics of the tumors processed for organoid cultures are presented in **S5 Table**. We found the variable levels of *KLK6* transcript in the screened cases with no association between *KLK6* expression and the molecular subtype of the analyzed tumors (**Fig. 8A**). The *HMGA2* and *FOSL1* transcript levels were detected in 4 out of 7 cases and the *FOSL1* level was found to be low compared to others genes. The *KRT6A* and *KRT6B* gene transcripts were significantly elevated in the adenoma (one case) compared to malignant tumors (**Fig. 8B**). The *c-MYC* transcript level was significantly higher in P#2T and P#13T organoid cultures (both MSS subtype and stage T3N0) along with overexpression of *KLK6*, *HMGA2*, and *FOSL1*, which suggests the aggressive phenotype of these tumors. We also measured the levels of secreted *KLK6* in the conditioned media of organoids derived from the normal and tumor tissue samples using *KLK6* ELISA assay (**Fig. 8C**). The analysis of the secreted *KLK6* levels in three cases, representing the colon cancer samples (P#3 and P#5) and a colon adenoma (P#7) showed that the elevated levels of secreted *KLK6* were detected in the tumor cultures compared to normal ones and they increased with the time in culture. The secreted *KLK6* was below the level of detection in the normal cultures of P#3 and P#7. The detected secreted *KLK6* in the normal culture of P#5 suggests that the tissue has acquired some invasive properties. Overall, this pilot analysis confirms the variability of *KLK6* expression in the CRC and demonstrates the experimental platform for more detailed analysis of key genes in the colon cancer.

Figure 8. Evaluation of the transcript levels of selected co-expressed genes and *KLK6* secretion in the CRC patient-derived organoid cultures. (A) of *KLK6*, *FOSL 1* and *HMGA2* transcript levels by quantitative RT PCR. (B) Expression of keratin 6A and 6B and *c-Myc*. (C) Levels of secreted *KLK6* in the conditioned media of the tumor organoid cultures by ELISA.

318 Discussion

319 The ongoing challenge in cancer treatment is to identify the most appropriate targeted
 320 therapy for managing the disease based on the individual properties of the tumor. This
 321 precision medicine approach requires identification of patients individual genetic, epigenetic
 322 and other alterations for further stratification into subgroups with matching drug treatment
 323 options. Kallikreins have become a subject of intense investigation as prognostic markers and
 324 candidates for drug targeting because of their important functions in normal physiology and
 325 pathological state, which have been discovered in the past years (reviewed in [24]). Recent
 326 analysis of kallikrein gene family across 15 different cancers was done with a goal to identify
 327 the suitable biomarkers within kallikrein gene family and highlighted KLK6, KLK7, KLK8
 328 and KLK10 as the excellent diagnostic biomarkers candidates for colon adenocarcinoma [25].
 329 Another analysis of TCGA dataset, which focused on search of prognostic biomarkers in
 330 colon adenocarcinomas, identified KLK6 gene as one of the two overexpressed genes (MYC
 331 was the second one), which are significantly associated with overall survival in
 332 adenocarcinoma patients and therefore have high accuracy in predicting the disease outcome
 333 [26].

334 In our study, we selected the extremely high KLK6 expressing cases (KLK6-high
 335 group, 16 cases) from TCGA database and analyzed the clinical parameters and molecular
 336 pathways and genes, which are associated with KLK6 overexpression. KLK6-high CRC cases
 337 in TCGA revealed that 25% of patients in this group had an advanced clinical stage (Stage IV
 338 and IVA), while none patients in KLK6-low group were diagnosed with Stage IV of the
 339 disease. No difference in the molecular subtypes of the adenocarcinomas were noted between
 340 KLK6-high and KLK6-low cases. At the same time, the analysis of the mutations revealed
 341 that the KLK6-high cases had the high frequencies of mutations in titin (TTN), APC, K-RAS
 342 and MUC 16 genes, while in KLK6-low group APC mutations were found at 100% frequency

343 and TP53 mutation was identified in 66% of cases. We have previously reported the mutant
344 K-RAS-dependent expression of KLK6 gene in the colon cancer cells [11, 27], so the current
345 results from TCGA patient data confirm the validity of the previous findings.

346 The mutational landscape in the KLK6-high group suggests the possibility of utilizing
347 the tumor mutational burden (TMB) biomarker for estimating the therapy outcomes in the
348 CRC patients with high KLK6 expression. The TMB is a novel and a promising biomarker
349 used for predicting responsiveness to Immune Checkpoint Blockage (ICB) [28, 29]. Tumors
350 with high TMB, independent from MSI-H status, such as KLK6 high expression tumors, are
351 not well studied. Titin (TTN), a structural protein in striated muscles [30] and MUC16, which
352 encodes the cancer antigen CA-125, are the two longest genes in the genome. The correlation
353 between the high TMB and the mutated status of the genes having the larger size has been
354 previously suggested [31]. Recent analysis of a discovery cohort of 34 solid tumors and 7
355 cohorts from clinical trials treated with programmed cell-death ligand 1 (PD-L1) inhibitor
356 demonstrated that TTN mutations in solid tumors can be used to stratify into groups with
357 different clinical responses to ICB monotherapy [31]. The unique presence of both of these
358 mutations along with high number of K-RAS and APC mutations in the KLK6-high cases
359 from TCGA dataset warrants further validation of the TBM applicability in the CRC patients
360 with KLK6-high cases for estimating their eligibility for ICB therapy in prospective clinical
361 trials. PIK3CA, another one of the most commonly mutated genes in CRC is also found in
362 ~43% of high KLK6 tumors. In the course of DEGs analysis, a set of the 236 protein-
363 coding DEGs were identified, which is unique to KLK6-high group. The correlation analysis
364 identified the members of kallikrein family, KLK7, KLK8, KLK10, as the most highly co-
365 expressed with KLK6. This finding is in agreement with limited analysis of kallikrein family
366 in the CRC reported by Tailor and co-workers [25]. Co-expression of KLK6 and KLK10 has
367 also been reported in pancreatic adenocarcinomas, where it was defined as an unfavorable

368 factor associated with a poor OS [32].

369 The other overexpressed genes, which correlated with KLK6, were. keratins 6A, 6B,
370 15, 16, 19, 80. This set of upregulated keratins is different from the one, identified in KLK6
371 overexpressing serous ovarian cancer [33]. This finding suggests that KLK6 regulates
372 expression of keratins in cancer in a tissue-specific manner. The list of altered genes included
373 extracellular matrix proteins ING4B (integrin 4B), LAMB3 (laminin subunit beta3), vesicle-
374 mediated transport protein coding genes SYL1 (Synaptotagmin Like 1) and GJB3 (Gap
375 Junction protein beta 3), as well as signaling genes, such as BMP4 (Bone Morphologic
376 Protein 4, in TGF- β signaling pathway) and transcription factor FOSL1 (FOS-like
377 transcription factor subunit). Analysis of the functional roles of the most altered genes
378 showed that they are important structural components of the cell cytoskeleton, gap junction
379 and connexon complexes, the basolateral membrane and actin cytoskeleton, as well as
380 extracellular exosomes, which indicates that KLK6 overexpression contributes to aggressive
381 growth and invasion in tumor cells.

382 Protein interaction network for KLK6-overexpressing cases allowed us to identify
383 proteins which may interact with KLK6, either through formation of protein complexes or via
384 modifications of other proteins. Particularly, the presence of the kallikrein family members,
385 KLK5, KLK7, KLK8, KLK10, KLK13, and the metalloproteinase family members MMP7
386 and MMP11 in this network suggests that active KLK6 enzyme may induce activation and
387 cross activation of other proteases. The constructed network predicts the direct interactions
388 between KLK7, KLK8 and KLK10, but not KLK6. This observation highlights complex
389 relations among kallikreins in the CRC. Through protein interaction network we found
390 keratins 6A, 6B, 15 and 16, as well as Small Proline Rich Repeat (SPRR) genes (SPRR1A,
391 SPRR1B, SPRR2A, SPRR2D and SPRR3), which are involved in the keratinization pathway,
392 as interacting partners of KLK6. This suggest the mechanism of KLK6-mediated

upregulation of a subgroup of keratin genes in KLK6-high group. Keratins are established epithelial cell markers and can regulate multiple signaling pathways in cancer. The keratins altered in KLK6-high colorectal tumors may be further evaluated as biomarkers in the CRC progression. SPRRs proteins are the keratinocytes differentiation markers, involved in formation of the cornified envelope in skin cells, but their function in cancer is not well defined. The SPRR3 has been reported as tumor suppressor in esophageal squamous carcinoma cells [34]. The S100A6 gene, which was found to be an interacting partner of SPRR3 and SPRR1A in KLK6 high group, has been associated with cells motility and tumor metastasis in cervical cancer cells and lung carcinoma [35, 36]. SERPIN F2, a serine protease inhibitor, which was identified in connection with KLK6 overexpression, is known inhibitor of plasmin, trypsin and chemotrypsin.

Although in this study we focused on protein-coding genes, we noted top two highly up regulated non-coding transcripts in KLK6 high group samples, i.e. CTB-147C22.8 and CTB147C22, which were reported to be co-expressed with kallikrein genes KLK5, KLK6 and KLK7 in HPV16⁺ head and neck tumors [37]. Using a set of 7 patient derived tumor organoid cultures we confirmed the variability in expression and secretion of KLK6, and necessity of individualized approach in designing the CRC chemotherapy.

The findings presented here suggest that KLK6-high tumors may be sensitive to the specific inhibitors of kallikreins. The reversible inhibitors of the members of kallikrein family have been previously reported [38-40]. Applicable to our targets, the promising inhibitors of KLK6 [41, 42] and KLK7 [43] have been recently characterized. The pre-clinical testing of selected kallikrein inhibitors should be done to validate their use as investigational drugs in clinical trials especially because of demonstrated here high complexity of interactions between kallikreins.

417 The analysis of KLK6 -overexpressing cases from TCGA was verified using a GEO
418 CRC sample set and the overall findings were in agreement with the results of TCGA
419 analysis. There were higher number of KLK6-overexpression cases in this set (30 samples
420 vs 16 in TCGA) and the GEO patients had more advance stages of the disease (no patients
421 with Stage I and the higher percent of patients had Stage II and Stage III of the disease).
422 Perhaps because of these differences, the analysis of OS rates in the GEO dataset showed
423 that patient's survival rates significantly correlated with KLK6 overexpression in tumors,
424 while in TCGA samples the Kaplan-Meier curves demonstrated a similar trend.

425

426 **Conclusions**

427 Overall, applying different types of analysis we discovered the distinct sets of genes,
428 such as the members of kallikreins, keratins, extracellular matrix proteins, SPRRs, S100A
429 families, protein trafficking and signaling genes within TGF- β , FOS and Ser/Thr protein
430 kinase families. These genes are involved in regulation and coordination of colon cancer cell
431 invasion and metastasis in KLK6-overexpressed CRCs. Our analysis suggests that identifying
432 colon cancer patients with KLK6 overexpression and targeting it using specific inhibitors can
433 be beneficial for the suppression of metastasis. In a future we hope to match the altered
434 pathways and genes in the individual patient cases with available or investigational drug
435 treatment options.

436

437 **Acknowledgements**

438 We would like to thank Dr David B. Stewart, MD, FACS, FASCRS, Professor of Surgery at
439 University of Arizona for facilitating the clinical samples acquisition and Ms. Carole
440 Kepler, B.S., MT (ASCP), Biospecimen Operations Manager of the University of Arizona

441 Cancer for her efforts in the tissue collection program. We also thank the patients who were
442 consent for tissue collection for organoid cultures.

443

444 **Funding**

445 The presented work was supported by the National Cancer Institute of the National
446 Institutes of Health (www.nih.gov) under award numbers R01CA157595 (to N.A.
447 Ignatenko), Research in this manuscript was directly supported by the Tissue Acquisition
448 and Molecular Analysis Shared Resource funded by the National Cancer Institute Award
449 P30CA023074 (to J. Sweasy, Cancer Center Support Grant). The funders had no role in
450 study design, data collection and analysis, decision to publish, or preparation of the
451 manuscript.

452

453 **Authors' contributions**

454 RP and NI designed the study, interpreted data and wrote the manuscript. ZM and CY
455 performed the bioinformatics data acquisition and analysis and assisted with manuscript
456 preparation. DD was involved in the analysis and interpretation of the data. CK gave
457 technical assistance with *in vitro* experiments. VN supervised the clinical samples
458 acquisition and evaluation. RP and NI are the corresponding authors. All authors read and
459 approved the final manuscript.

460

461 **References**

1. Cavallo F, De Giovanni C, Nanni P, Forni G, Lollini PL. 2011: the immune hallmarks of cancer. *Cancer Immunol Immunother*. 2011; 60: 319-26. doi: 10.1007/s00262-010-0968-0.
2. Sotiropoulou G, Pampalakis G, Diamandis EP. Functional roles of human kallikrein-related peptidases. *J Biol Chem*. 2009; 284: 32989-94. doi: 10.1074/jbc.R109.027946.
3. Devetzi M, Trangas T, Scorilas A, Xynopoulos D, Talieri M. Parallel overexpression and clinical significance of kallikrein-related peptidases 7 and 14 (KLK7KLK14) in colon cancer. *Thromb Haemost*. 2013; 109: 716-25. doi: 10.1160/TH12-07-0518.
4. Vakrakou A, Devetzi M, Papachristopoulou G, Malachias A, Scorilas A, Xynopoulos D, Talieri M. Kallikrein-related peptidase 6 (KLK6) expression in the progression of colon adenoma to carcinoma. *Biol Chem*. 2014; 395: 1105-17. doi: 10.1515/hsz-2014-0166.
5. Alexopoulou DK, Papadopoulos IN, Scorilas A. Clinical significance of kallikrein-related peptidase (KLK10) mRNA expression in colorectal cancer. *Clin Biochem*. 2013; 46: 1453-61. doi: 10.1016/j.clinbiochem.2013.03.002.
6. Alexopoulou DK, Kontos CK, Christodoulou S, Papadopoulos IN, Scorilas A. KLK11 mRNA expression predicts poor disease-free and overall survival in colorectal adenocarcinoma patients. *Biomark Med*. 2014; 8: 671-85. doi: 10.2217/bmm.13.151.
7. Borgono CA, Diamandis EP. The emerging roles of human tissue kallikreins in cancer. *Nat Rev Cancer*. 2004; 4: 876-90. doi:
8. Ogawa K, Utsunomiya T, Mimori K, Tanaka F, Inoue H, Nagahara H, Murayama S, Mori M. Clinical significance of human kallikrein gene 6 messenger RNA expression in colorectal cancer. *Clinical Cancer Research*. 2005; 11: 2889-93. doi:

9. Ohlsson L, Lindmark G, Israelsson A, Palmqvist R, Oberg A, Hammarstrom ML, Hammarstrom S. Lymph node tissue kallikrein-related peptidase 6 mRNA: a progression marker for colorectal cancer. *Br J Cancer*. 2012; 107: 150-7. doi: 10.1038/bjc.2012.220.
10. Fearon ER. Molecular genetics of colorectal cancer. *Annu Rev Pathol*. 2011; 6: 479-507. doi: 10.1146/annurev-pathol-011110-130235.
11. Henkhaus RS, Gerner EW, Ignatenko NA. Kallikrein 6 is a mediator of K-RAS-dependent migration of colon carcinoma cells. *Biol Chem*. 2008; 389: 757-64. doi: 10.1515/BC.2008.087.
12. Henkhaus RS, Roy UK, Cavallo-Medved D, Sloane BF, Gerner EW, Ignatenko NA. Caveolin-1-mediated expression and secretion of kallikrein 6 in colon cancer cells. *Neoplasia*. 2008; 10: 140-8. doi:
13. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010; 11: R106. doi: 10.1186/gb-2010-11-10-r106.
14. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15: 550. doi: 10.1186/s13059-014-0550-8.
15. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019; 47: D607-D13. doi: 10.1093/nar/gky1131.
16. Sato T, Stange DE, Ferrante M, Vries RG, Van Es JH, Van den Brink S, Van Houdt WJ, Pronk A, Van Gorp J, Siersema PD, Clevers H. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology*. 2011; 141: 1762-72. doi: 10.1053/j.gastro.2011.07.050.

- 510 17. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon
511 and rectal cancer. *Nature*. 2012; 487: 330-7. doi: 10.1038/nature11252.
- 512 18. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, Marisa
513 L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, et al. The
514 consensus molecular subtypes of colorectal cancer. *Nat Med*. 2015; 21: 1350-6. doi:
515 10.1038/nm.3967.
- 516 19. Lee MS, Menter DG, Kopetz S. Right Versus Left Colon Cancer Biology:
517 Integrating the Consensus Molecular Subtypes. *J Natl Compr Canc Netw*. 2017; 15: 411-9.
518 doi: 10.6004/jnccn.2017.0038.
- 519 20. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal
520 B, Jupe S, Korkner F, McKay S, Matthews L, May B, Milacic M, et al. The Reactome
521 pathway Knowledgebase. *Nucleic Acids Res*. 2016; 44: D481-7. doi: 10.1093/nar/gkv1351.
- 522 21. Gerner EW. Role of c-Myc in intestinal tumorigenesis of the ApcMin/+ mouse.
523 *Technol Cancer Res Treat*. 2006; 5: 1658-64. doi:
- 524 22. Morishita A, Zaidi MR, Mito A, Sankarasharma D, Szabolcs M, Okada Y,
525 D'Armiento J, Chada K. HMGA2 is a driver of tumor metastasis. *Cancer Res*. 2013; 73:
526 4289-99. doi: 10.1158/0008-5472.CAN-12-3848.
- 527 23. Chen H, Sells E, Pandey R, Abril ER, Hsu CH, Krouse RS, Nagle RB, Pampalakis
528 G, Sotiropoulou G, Ignatenko NA. Kallikrein 6 protease advances colon tumorigenesis via
529 induction of the high mobility group A2 protein. *Oncotarget*. 2019; 10: 6062-78. doi:
530 10.18632/oncotarget.27153.
- 531 24. Prassas I, Eissa A, Poda G, Diamandis EP. Unleashing the therapeutic potential of
532 human kallikrein-related serine proteases. *Nat Rev Drug Discov*. 2015; 14: 183-202. doi:
533 10.1038/nrd4534.

- 534 25. Tailor PD, Kodeboyina SK, Bai S, Patel N, Sharma S, Ratnani A, Copland JA, She
535 JX, Sharma A. Diagnostic and prognostic biomarker potential of kallikrein family genes in
536 different cancer types. *Oncotarget*. 2018; 9: 17876-88. doi: 10.18632/oncotarget.24947.
- 537 26. Dong S, Ding Z, Zhang H, Chen Q. Identification of Prognostic Biomarkers and
538 Drugs Targeting Them in Colon Adenocarcinoma: A Bioinformatic Analysis. *Integr Cancer*
539 *Ther*. 2019; 18: 1534735419864434. doi: 10.1177/1534735419864434.
- 540 27. Ignatenko NA, Zhang H, Watts GS, Skovan BA, Stringer DE, Gerner EW. The
541 chemopreventive agent alpha-difluoromethylornithine blocks Ki-ras-dependent tumor
542 formation and specific gene expression in Caco-2 cells. *Mol Carcinog*. 2004; 39: 221-33.
543 doi:
- 544 28. Yarchoan M, Hopkins A, Jaffee EM. Tumor Mutational Burden and Response Rate
545 to PD-1 Inhibition. *N Engl J Med*. 2017; 377: 2500-1. doi: 10.1056/NEJMc1713444.
- 546 29. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY,
547 Barron DA, Zehir A, Jordan EJ, Omuro A, Kaley TJ, Kendall SM, Motzer RJ, et al. Tumor
548 mutational load predicts survival after immunotherapy across multiple cancer types. *Nat*
549 *Genet*. 2019; 51: 202-6. doi: 10.1038/s41588-018-0312-8.
- 550 30. Ciferri A, Crumbliss AL. The Assembling and Contraction Mechanisms of Striated
551 Muscles. *Front Chem*. 2018; 6: 570. doi: 10.3389/fchem.2018.00570.
- 552 31. Jia Q, Wang J, He N, He J, Zhu B. Titin mutation associated with responsiveness to
553 checkpoint blockades in solid tumors. *JCI Insight*. 2019; 4. doi: 10.1172/jci.insight.127901.
- 554 32. Ruckert F, Hennig M, Petraki CD, Wehrum D, Distler M, Denz A, Schroder M,
555 Dawelbait G, Kalthoff H, Saeger HD, Diamandis EP, Pilarsky C, Grutzmann R. Co-
556 expression of KLK6 and KLK10 as prognostic factors for survival in pancreatic ductal
557 adenocarcinoma. *Br J Cancer*. 2008; 99: 1484-92. doi: 10.1038/sj.bjc.6604717.

- 558 33. Wang P, Magdolen V, Seidl C, Dorn J, Drecolle E, Kotzsch M, Yang F, Schmitt M,
559 Schilling O, Rockstroh A, Clements JA, Loessner D. Kallikrein-related peptidases 4, 5, 6
560 and 7 regulate tumour-associated factors in serous ovarian cancer. *Br J Cancer*. 2018; 119:
561 1-9. doi: 10.1038/s41416-018-0260-1.
- 562 34. Zhang Y, Feng YB, Shen XM, Chen BS, Du XL, Luo ML, Cai Y, Han YL, Xu X,
563 Zhan QM, Wang MR. Exogenous expression of Esophagin/SPRR3 attenuates the
564 tumorigenicity of esophageal squamous cell carcinoma cells via promoting apoptosis. *Int J*
565 *Cancer*. 2008; 122: 260-6. doi: 10.1002/ijc.23104.
- 566 35. Li A, Gu Y, Li X, Sun H, Zha H, Xie J, Zhao J, Huang M, Chen L, Peng Q, Zhang
567 Y, Weng Y, Zhou L. S100A6 promotes the proliferation and migration of cervical cancer
568 cells via the PI3K/Akt signaling pathway. *Oncol Lett*. 2018; 15: 5685-93. doi:
569 10.3892/ol.2018.8018.
- 570 36. Li P, Lv X, Zhang Z, Xie S. S100A6/miR193a regulates the proliferation, invasion,
571 migration and angiogenesis of lung cancer cells through the P53 acetylation. *Am J Transl*
572 *Res*. 2019; 11: 4634-49. doi:
- 573 37. Salyakina D, Tsinoremas NF. Non-coding RNAs profiling in head and neck cancers.
574 *NPJ Genom Med*. 2016; 1: 15004. doi: 10.1038/npjgenmed.2015.4.
- 575 38. Goettig P, Magdolen V, Brandstetter H. Natural and synthetic inhibitors of
576 kallikrein-related peptidases (KLKs). *Biochimie*. 2010; 92: 1546-67. doi:
577 10.1016/j.biochi.2010.06.022.
- 578 39. Liang G, Chen X, Aldous S, Pu SF, Mehdi S, Powers E, Giovanni A, Kongsamut S,
579 Xia T, Zhang Y, Wang R, Gao Z, Merriman G, et al. Virtual Screening and X-ray
580 Crystallography for Human Kallikrein 6 Inhibitors with an Amidinothiophene P1 Group.
581 *ACS Med Chem Lett*. 2012; 3: 159-64. doi: 10.1021/ml200291e.

- 582 40. Sotiropoulou G, Pampalakis G. Targeting the kallikrein-related peptidases for drug
583 development. Trends Pharmacol Sci. 2012; 33: 623-34. doi: 10.1016/j.tips.2012.09.005.
- 584 41. Sananes A, Cohen I, Shahar A, Hockla A, De Vita E, Miller AK, Radisky ES, Papo
585 N. A potent, proteolysis-resistant inhibitor of kallikrein-related peptidase 6 (KLK6) for
586 cancer therapy, developed by combinatorial engineering. J Biol Chem. 2018; 293: 12663-
587 80. doi: 10.1074/jbc.RA117.000871.
- 588 42. De Vita E, Schuler P, Lovell S, Lohbeck J, Kullmann S, Rabinovich E, Sananes A,
589 Hessling B, Hamon V, Papo N, Hess J, Tate EW, Gunkel N, et al. Depsipeptides Featuring
590 a Neutral P1 Are Potent Inhibitors of Kallikrein-Related Peptidase 6 with On-Target
591 Cellular Activity. J Med Chem. 2018; 61: 8859-74. doi: 10.1021/acs.jmedchem.8b01106.
- 592 43. de Veer SJ, Furio L, Swedberg JE, Munro CA, Brattsand M, Clements JA,
593 Hovnanian A, Harris JM. Selective Substrates and Inhibitors for Kallikrein-Related
594 Peptidase 7 (KLK7) Shed Light on KLK Proteolytic Activity in the Stratum Corneum. J
595 Invest Dermatol. 2017; 137: 430-9. doi: 10.1016/j.jid.2016.09.017.

596

597

Supporting information

- 598 **S1 Figure. KLK6 expression in normal colon and tumors from TCGA stratified**
599 **according to Consensus Molecular subtypes (CMS).** CMS1-4 are four subtypes defined
600 as in [18]. NoLBL are TCGA tumors with no labels or could not be subtype classified.
- 601 **S1 Table. Supplemental data for molecular characterization of TCGA samples. Data**
602 **are from Genomics Data Commons (GDC).** **A)** Top 100 frequently mutated genes in
603 Colon adenocarcinoma TCGA KLK6 high samples. **B)** Top 100 mutated genes in Colon
604 adenocarcinoma TCGA low KLK6 samples. **C)** Analysis of alterations in KLK6 gene in
605 TCGA cohort.

606 **S2 Table. Patient Characteristics of high KLK6 samples from GEO dataset**

607 **GSE39582.**

608 **S3 Table. Differential expression analysis of genes in the KLK6-high and KLK6-low**

609 **patient samples. A)** Differential Expressed Genes between high KLK6 expressed Group

610 and low KLK6 expressed groups in TCGA Colon Adenocarcinoma samples; **B)** 236

611 Differential Expressed Genes between high KLK6 expressed Group and low KLK6

612 expressed groups and high KLK6 expressed groups and rest all the samples in TCGA colon

613 tumor; **C)** Genes correlated with KLK6 in Colon adenocarcinoma TCGA samples. **D)**

614 Genes found in common between TCGA DEG list of 236 genes and GEO dataset

615 GSE39582 of KLK6 high samples compared with rest of the samples.

616 **S4 Table. Supplemental data for pathways associated with KLK6 overexpression in**

617 **colorectal cancer. A)** Gene Ontology Terms enriched in KLK6 high samples in TCGA for

618 **Fig. 6C. B)** Reactome Pathways enriched in KLK6 high samples in TCGA for **Fig. 6D. C)**

619 Protein Interaction of Kallikreins and differential expression genes in KLK6 High Samples

620 in TCGA from Stringdb for **Fig.7. D)** Annotation-Protein Interaction partners with

621 Kallikreins in **Fig. 7.**

622 **S5 Table. Clinicopathological characteristics of surgical cases established as organoid**

623 **cultures.**

Figure 1A.

Figure 1B.

Figure 2.

A.

B.

Figure 3

A.

B.

Figure 3

C.

Figure 4.

Figure 5.

Figure 6

Figure 7.

Figure 8.