Resource

Cell

Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities

Suhas Vasaikar,[1,2,14] Chen Huang,[1,2,14] Xiaojing Wang,[1,2,12,14] Vladislav A. Petyuk,[3,14] Sara R. Savage,[4,14] Bo Wen,[1,2] Yongchao Dou,[1,2] Yun Zhang,[1] Zhiao Shi,[1,2] Osama A. Arshad,[3] Marina A. Gritsenko,[3] Lisa J. Zimmerman,[5] Jason E. McDermott,[3] Therese R. Clauss,[3] Ronald J. Moore,[3] Rui Zhao,[3] Matthew E. Monroe,[3] Yi-Ting Wang,[3] Matthew C. Chambers,[5] Robbert J.C. Slebos,[5] Ken S. Lau,[6] Qianxing Mo,[7,13] Li Ding,[8] Matthew Ellis,[1,7] Mathangi Thiagarajan,[9] Christopher R. Kinsinger,[10] Henry Rodriguez,[10] Richard D. Smith,[3] Karin D. Rodland,[3,11,15,*] Daniel C. Liebler,[5,15,*] Tao Liu,[3,15,*] Bing Zhang,[1,2,7,15,16,*] and Clinical Proteomic Tumor Analysis Consortium

[1]Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX 77030, USA
[2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
[3]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA
[4]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37232, USA
[5]Department of Biochemistry, Vanderbilt University, Nashville, TN 37232, USA
[6]Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN 37232, USA
[7]Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA
[8]The McDonnell Genome Institute, Washington University in St. Louis, Forest Park Avenue, Campus Box 8501, St. Louis, MO 63108, USA
[9]Leidos Biomedical Research Inc., Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA
[10]Office of Cancer Clinical Proteomics Research, National Cancer Institute, Bethesda, MD 20892, USA
[11]Department of Cell, Developmental, and Cancer Biology, Oregon Health & Science University, Portland, OR 97221, USA
[12]Present address: Department of Epidemiology and Biostatistics, Greehey Children's Cancer Research Institute, The University of Texas Health Science Center at San Antonio, San Antonio, TX 78229, USA
[13]Present address: Department of Biostatistics & Bioinformatics, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL 33612, USA
[14]These authors contributed equally
[15]Senior author
[16]Lead Contact
*Correspondence: karin.rodland@pnnl.gov (K.D.R.), daniel.liebler@vanderbilt.edu (D.C.L.), tao.liu@pnnl.gov (T.L.), bing.zhang@bcm.edu (B.Z.)
https://doi.org/10.1016/j.cell.2019.03.030

SUMMARY

We performed the first proteogenomic study on a prospectively collected colon cancer cohort. Comparative proteomic and phosphoproteomic analysis of paired tumor and normal adjacent tissues produced a catalog of colon cancer-associated proteins and phosphosites, including known and putative new biomarkers, drug targets, and cancer/testis antigens. Proteogenomic integration not only prioritized genomically inferred targets, such as copy-number drivers and mutation-derived neoantigens, but also yielded novel findings. Phosphoproteomics data associated Rb phosphorylation with increased proliferation and decreased apoptosis in colon cancer, which explains why this classical tumor suppressor is amplified in colon tumors and suggests a rationale for targeting Rb phosphorylation in colon cancer. Proteomics identified an association between decreased CD8 T cell infiltration and increased glycolysis in microsatellite instability-high (MSI-H) tumors, suggesting glycolysis as a potential target to overcome the resistance of MSI-H tumors to immune checkpoint blockade. Proteogenomics presents new avenues for biological discoveries and therapeutic development.

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer worldwide and the fourth leading cause of cancer-related deaths (Arnold et al., 2017). Recent studies of the genomic, transcriptomic, and proteomic landscapes of human CRC have identified many genomic alterations and have revealed extensive molecular heterogeneity of the disease (Cancer Genome Atlas Network, 2012; Guinney et al., 2015; Zhang et al., 2014). However, the rapidly accumulating omics data have yet to bring novel biomarkers and drug targets to the clinic.

Global proteomic differences between tumor and normal tissues, which are critical for cancer biomarker discovery, have not been systematically characterized in large tumor cohorts. Signaling proteins and pathways are often attractive therapeutic targets for cancer treatment, yet global phosphoproteomic analyses on human CRC are lacking. Recent advances in cancer immunotherapy underscore the critical need for biomarkers to predict response to immune checkpoint inhibition and to select neoantigens for personalized vaccine development (Sharma et al., 2017). Proteogenomics can provide fresh approaches to these needs. Here, we describe a proteogenomic study from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) on

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Specimens and Clinical Data

Tumor, adjacent normal, and blood samples were collected by several tissue source sites in strict accordance to the CPTAC-2 colon procurement protocol (https://brd.nci.nih.gov/brd/sop/download-pdf/321) with an informed consent from the patients. The Washington University in St. Louis Institutional Review Board (IRB) reviewed the individual informed consent documents at each tissue source site and determined that the materials sent to the CPTAC biospecimen core resource met the requirements for Exemption 4. The cohort had an inclusion criterion of newly diagnosed, untreated patients undergoing primary surgery for colon adenocarcinoma. Because untreated rectal tumors are difficult to obtain, we only included colon cancers, which represent approximately 70% of all CRCs. Patients with prior history of other malignancies within 12 months, any systemic chemotherapy, endocrine or biological therapy as well as prior radiation therapy to the abdomen or pelvis for any cancer type were excluded from the study. Required clinical information regarding patient history and status of surgery along with relevant diagnostic information were collected using case reports forms. One year follow up information with updated history after completion of the initial treatment regimen were also collected through follow up forms. Deidentified pathology reports and representative diagnostic slide images were utilized to review and qualify cases for this study. The peripheral venous blood from the same patient were collected prior to administration of anesthesia. Segments from qualified tumor specimens were greater than 300mg in mass with at least 60% tumor cell nuclei and less than 20% necrosis. To ensure tissue suitability for phosphoprotein analysis, the tissue and the adjacent normal specimens were collected in less than 30 minutes total ischemic time and embedded in optimal cutting temperature (OCT) compound for processing at a common CPTAC-2 specimen core resource center. Pathologically qualified cases underwent further molecular qualification for extraction and co-isolation of nucleic acids. Tissue segments that were pathology and molecular qualified were shipped to the proteomic characterization centers. DNA and RNA from the same tumor segment and DNA from germline blood were further aliquoted and quantified per protocol. DNA quality was confirmed using gel electrophoresis and Nano drop methods. RNA quality was confirmed using Nano drop and Agilent bioanalyzer. Sufficient yield, a good gel score and passing value of 7 or greater RIN qualified the DNA and RNA, respectively, for sequencing. The analytes were then shipped to the sequencing center. The corresponding clinical data were formatted and distributed through the CPTAC data coordinating center (https://cptac-data-portal.georgetown.edu/cptac/s/S037). Table S1 summarizes the clinical and pathological characteristics of the tumors and the specific numbers of samples analyzed by each omics platform. Among the 110 patients in the cohort, there were 65 females (60%) and 45 males (40%), with an average age of 65 (range 40 to 93 years). We did not perform analyses on the two sexes separately because the sample size is too small after sex stratification. Moreover, we were interested in results common to both sexes, and the sex distribution is reasonably balanced.

METHODS DETAILS

PCR-based MSI Analysis

The MSI Analysis System (version 1.2, Promega), a fluorescent PCR-based assay, was used to detect microsatellite instability (MSI) in the colon tumors. The analysis compares allelic profiles of microsatellite markers generated by amplification of DNA from matching tumor and normal samples, and alleles that are present in the tumor sample but not in corresponding normal samples indicate MSI. The system uses seven markers including five mononucleotide repeat markers (BAT-25, BAT-26, NR-21, NR-24, and MONO-27) and two pentanucleotide repeat markers (Penta C and Penta D). The output data were analyzed with GeneMapper® software (Applied Biosystems) to determine MSI status of the colon tumor samples.

Genotyping Array Analysis

Genomic DNA samples were prepared according to Illumina's Infinium LCG Quad Assay manual protocol. Processed samples were loaded on the HumanOmni5-Quad BeadChips and run on the HiScan platform. SNP and CNP genotyping were performed with the Genome Studio Genotyping Module (Version 2.0, Illumina).

Whole Exome Sequencing

Genomic DNA samples were used to prepare indexed libraries using the Nextera Rapid Capture Exome kit from Illumina. Library preparation was performed using a semi-automated 96-well plate method, with washing and clean-up/concentration steps performed on the Beckman Coulter Biomek NXP platform and with ZR-96 DNA Clean & Concentrator-5 plates, respectively. Libraries were quantified using the Agilent 2100 Bioanalyzer. Pooled libraries were run on HiSeq4000 (2x150 paired end runs) to achieve a minimum of 150x on target coverage per each sample library. The raw Illumina sequence data were demultiplexed and converted to fastq files, adaptor and low-quality sequences were trimmed. Whole exome sequencing (WXS) data were used for somatic mutation detection, microsatellite instability prediction, and somatic copy number alteration (SCNA) analysis as described below.

Somatic Mutation Detection

We followed the Genome Analysis Toolkit (GATK, version 3.8.0) best practice guideline for somatic short variant discovery (https://software.broadinstitute.org/gatk/best-practices/workflow?id=11146). Briefly, we aligned paired-end WXS reads to the human reference genome (hg19) with BWA-mem (version 0.7.15-r1140). The bam files were further processed by adding read groups, marking