Gastroenterology 2017;153:1082–1095

Colorectal Cancer Cell Line Proteomes Are Representative of Primary Tumors and Predict Drug Sensitivity

Jing Wang,[1,2,*] **Dmitri Mouradov**,[3,4,*] **Xiaojing Wang**,[1,2,*] Robert N. Jorissen,[3,4] Matthew C. Chambers,[5] Lisa J. Zimmerman,[5] Suhas Vasaikar,[1,2] Christopher G. Love,[3,4] Shan Li,[3] Kym Lowes,[3] Karl-Johan Leuchowius,[3] Helene Jousset,[3] Janet Weinstock,[6] Christopher Yau,[7,8] John Mariadason,[9,10] Zhiao Shi,[1] Yuguang Ban,[11] Xi Chen,[11,12] Robert J. C. Coffey,[13,14] Robbert J. C. Slebos,[15] Antony W. Burgess,[4,6,16] Daniel C. Liebler,[5] Bing Zhang,[1,2] and Oliver M. Sieber[3,4,16,17]

[1]Lester and Sue Smith Breast Center; [2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; [3]Systems Biology and Personalised Medicine Division, The Walter and Eliza Hall Institute of Medial Research; [4]Department of Medical Biology, The University of Melbourne, Parkville, VIC, Australia; [5]Department of Biochemistry, Vanderbilt University, Nashville, Tennessee; [6]Structural Biology Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia; [7]Wellcome Trust Centre for Human Genetics; [8]Department of Statistics, University of Oxford, Oxford, United Kingdom; [9]Olivia Newton-John Cancer Research Institute, Heidelberg; [10]La Trobe University School of Cancer Medicine, Melbourne, VIC, Australia; [11]Sylvester Comprehensive Cancer Center, [12]Department of Public Health Sciences, University of Miami Miller School of Medicine, Miami, Florida; [13]Department of Medicine, Vanderbilt University School of Medicine, Nashville; [14]Veterans Affairs Medical Center, Nashville, Tennessee; [15]Clinical Science Lab, Moffitt Cancer Center, Tampa, Florida; [16]Department of Surgery, The University of Melbourne, Parkville; and [17]School of Biomedical Sciences, Monash University, Clayton, VIC, Australia

BACKGROUND AND AIMS: Proteomics holds promise for individualizing cancer treatment. We analyzed to what extent the proteomic landscape of human colorectal cancer (CRC) is maintained in established CRC cell lines and the utility of proteomics for predicting therapeutic responses. **METHODS:** Proteomic and transcriptomic analyses were performed on 44 CRC cell lines, compared against primary CRCs (n=95) and normal tissues (n=60), and integrated with genomic and drug sensitivity data. **RESULTS:** Cell lines mirrored the proteomic aberrations of primary tumors, in particular for intrinsic programs. Tumor relationships of protein expression with DNA copy number aberrations and signatures of post-transcriptional regulation were recapitulated in cell lines. The 5 proteomic subtypes previously identified in tumors were represented among cell lines. Nonetheless, systematic differences between cell line and tumor proteomes were apparent, attributable to stroma, extrinsic signaling, and growth conditions. Contribution of tumor stroma obscured signatures of DNA mismatch repair identified in cell lines with a hypermutation phenotype. Global proteomic data showed improved utility for predicting both known drug-target relationships and overall drug sensitivity as compared with genomic or transcriptomic measurements. Inhibition of targetable proteins associated with drug responses further identified corresponding synergistic or antagonistic drug combinations. Our data provide evidence for CRC proteomic subtype-specific drug responses. **CONCLUSIONS:** Proteomes of established CRC cell line are representative of primary tumors. Proteomic data tend to exhibit improved prediction of drug sensitivity as compared with genomic and transcriptomic profiles. Our integrative proteogenomic analysis highlights the potential of proteome profiling to inform personalized cancer medicine.

Keywords: Colorectal Cancer; Cell Lines; Proteomics; Drug Sensitivity.

BASIC AND TRANSLATIONAL AT

Studies of the genomic and transcriptomic landscapes of human colorectal cancer (CRC) have been instrumental in advancing our understanding of disease biology and the identification of clinically actionable aberrations.[1–3] While the major genomic and transcriptomic hallmarks and subtypes of CRC have been defined,[4,5] these explain only part of tumor clinical heterogeneity. The next challenge is to gain a detailed understanding of the dynamic protein pathways involved in normal and disease states, and we have recently characterized the proteome of primary CRCs from patients participating in The Cancer Genome Atlas (TCGA) project, identifying 5 major proteomic subtypes (Clinical Proteomic Tumor Analysis Consortium [CPTAC]).[6] From a therapeutic perspective, most drug targets are proteins rather than nucleic acids, and we have shown that DNA- or mRNA-level measurements are poor predictors of protein abundance.[6]

Cancer cell lines are the most commonly utilized model systems in tumor biology and therapy development. Large cancer cell line-based projects, such as NCI-60,[7] Cancer Cell Line Encyclopedia (CCLE),[2] and Genomics of Drug Sensitivity in Cancer (GDSC),[3] have used molecularly heterogeneous

*Authors share co-first authorship.

Abbreviations used in this paper: **5-FU, 5-fluorouracil; CMS, consensus molecular subtype; CPTAC, Clinical Proteomic Tumor Analysis Consortium; CRC, colorectal cancer; FDR, false discovery rate; GDSC, Genomics of Drug Sensitivity in Cancer; GSEA, gene set enrichment analysis; IHC, immunohistochemistry; PAM, Prediction Analysis of Microarrays; SNP, single nucleotide polymorphism; SNV, single nucleotide variant; TCGA, The Cancer Genome Atlas.**

Most current article

0016-5085/$36.00
http://dx.doi.org/10.1053/j.gastro.2017.06.008

EDITOR'S NOTES

BACKGROUND AND CONTEXT

Proteomics holds promise for cancer medicine.

NEW FINDINGS

CRC cell line proteomes maintain cell-intrinsic programs, proteogenomic relationships and proteomic subtypes observed in primary tumors. Proteomics improves identification of known drug-target relationships as compared to genomic or transcriptomic measurements.

LIMITATIONS

Proteome-drug response relationships observed in CRC cell lines could not be investigated in patient tumors due to insufficient cases with single-agent treatment and outcome data.

IMPACT

An integrative proteomic analysis demonstrates both the value and limitation of CRC cell lines as models for primary disease, and highlights the potential of proteomic data to inform cancer treatment.

cancer cell lines to identify stratification biomarkers and signatures for precision medicine. Nonetheless, controversy remains whether cell lines provide an appropriate representation of primary tumors, given the lack of organismal context, different growth conditions, and selection or acquisition of additional aberrations in vitro. Genomic analyses indicate that established cancer cell lines are suitable molecular proxies for primary tumors in many cancer types,[2] yet findings at the transcriptomic level have been variable, with data for hepatocellular carcinoma[8] and CRC[9] indicating similarity between cell lines and primary tumors, whilst data for breast cancer suggest pronounced differences.[10] Although some global proteomics data sets for cancer cell lines are available,[11,12] no large-scale proteomic study exists comparing cell lines with primary tumors. It remains unknown whether cancer cell lines are representative of primary tumors at the proteome level, and to what extent molecular programs and proteogenomic relationships are maintained in vitro. The relative utility of proteomic data as a predictor of anti-cancer drug responses in comparison with genomic and transcriptomic modalities has not been systematically investigated.

Herein, we generated proteomic and transcriptomic data for a panel of 44 human CRC cell lines previously characterized at the genomic level.[13] We performed a comprehensive integrative proteogenomic analysis across these 44 cell lines and 95 CRCs and 60 normal tissue biopsies analyzed in our CPTAC project[6] to systematically evaluate cell lines as tumor models. We further integrated cell line proteogenomic data with drug sensitivity measurements to assess the utility of different types of omics data for predicting therapeutic responses and to connect tumor proteomic subtypes to drug sensitivity.

Materials and Methods

CRC Cell Lines and Primary Tumors

A total of 44 CRC cell lines were studied (Supplementary Table 1, Supplementary Methods). In addition, we retrieved previously published genomic, transcriptomic, and proteomic data on 95 primary tumor specimens from 90 CRC patients and proteomics data from 60 normal colon biopsies from 30 patients from our original CPTAC study,[6] as well as RNA-Seq data for 48 normal colon and rectum samples deposited by the TCGA (Supplementary Table 2 and 3).

Liquid Chromatography-Tandem Mass Spectrometry

The protein extraction and tryptic digestion of the frozen cell line pellets were performed as previously described for TCGA CRC specimens[6] (Supplementary Methods). Raw data for the cell lines, database search results, and the 2 versions of assemblies can be found at the Mass spectrometry Interactive Virtual Environment (MassIVE, ftp://massive.ucsd.edu/MSV000080374).

Transcriptome Sequencing

RNA samples from CRC cell lines were extracted from pellets collected at the same time as the samples for proteomics analysis and sequenced to a depth of at least 28 million reads. Reads were subsequently aligned to human genome build Hg19 using Tophat (Supplementary Methods).

Single Nucleotide Polymorphism Microarray Analysis

Single nucleotide polymorphism (SNP) array data on 38 cell lines from our cohort have been published previously.[13] SNP array assays on the additional DiFi, GEO, IS1, IS2, IS3, and V9P cells were performed at the Australian Genome Research Facility using CytoSNP-850K v1.1 and processed using OncoSNP v2.18 suite (https://sites.google.com/site/oncosnp/; Supplementary Methods).

Exome-Capture Sequencing

Whole exome mutation data on 35 CRC cell lines from our cohort have been published previously.[13] Additionally, DIFI, GEO, IS1, IS2, IS3, LIM1863, LIM2537, V9P, and VAC05 cells were sequenced using the Nextera Rapid Capture Expanded Exome Enrichment Kit (Illumina, San Diego, CA) on an Illumina HiSeq 2000 System at the Australian Genome Research Facility. Sequence alignment and calling of single nucleotide variants (SNVs) and INDEL in the absence of matched normal tissue were performed using a hybrid of the GATK Germline and Somatic Best Practice Variant Detection Protocols (Supplementary Methods).

Variant Peptide Identification and Analysis

To identify variant peptides, we derived customized protein sequence databases from matched whole exome sequencing and RNA-Seq data and then performed database searches using these databases for individual samples (Supplementary Methods).

Voom/Limma Analysis

Voom/limma analyses were performed using Limma and edgeR R packages, and method sensitivity and specificity for spectral count data were validated using the spike-in data set generated by the 2015 study of the Proteome Informatics Research Group (iPRG) of the Association of Biomolecular