Proteogenomic Analysis Reveals Unanticipated Adaptations of Colorectal Tumor Cells to Deficiencies in DNA Mismatch Repair

Patrick J. Halvey1,6, Xiaojing Wang2, Jing Wang2, Ajaz A. Bhat3,4, Punita Dhawan3,4, Ming Li5, Bing Zhang2,4, Daniel C. Liebler1,6, and Robbert J.C. Slebos4,6

Abstract

A growing body of genomic data on human cancers poses the critical question of how genomic variations translate to cancer phenotypes. We used standardized shotgun proteomics and targeted protein quantitation platforms to analyze a panel of 10 colon cancer cell lines differing by mutations in DNA mismatch repair (MMR) genes. In addition, we performed transcriptome sequencing (RNA-seq) to enable detection of protein sequence variants from the proteomic data. Biologic replicate cultures yielded highly consistent proteomic inventories with a cumulative total of 6,513 protein groups with a protein false discovery rate of 3.17% across all cell lines. Networks of coexpressed proteins with differential expression based on MMR status revealed impact on protein folding, turnover and transport, on cellular metabolism and on DNA and RNA synthesis and repair. Analysis of variant amino acid sequences suggested higher stability of proteins affected by naturally occurring germline polymorphisms than of proteins affected by somatic protein sequence changes. The data provide evidence for multisystem adaptation to MMR deficiency with a stress response that targets misfolded proteins for degradation through the ubiquitin-dependent proteasome pathway. Enrichment analysis suggested epithelial-to-mesenchymal transition in RKO cells, as evidenced by increased mobility and invasion properties compared with SW480. The observed proteomic profiles demonstrate previously unknown consequences of altered DNA repair and provide an expanded basis for mechanistic interpretation of MMR phenotypes. Cancer Res; 74(1); 1–11. ©2013 AACR.

Introduction

Colon cancer development is characterized by a well-documented series of genetic changes that drive the progression from early adenomas to metastatic carcinomas (1). These include a chromosomal instability, microsatellite instability (MIN), and CpG island methylation (1–3). In addition to these global genetic and epigenetic characteristics, a relatively small number of oncogenes and tumor suppressor genes are frequently altered in colorectal carcinoma, including, APC (90%), p53 (~50%), and K-ras (~40%; refs. 1, 2). More recent global sequencing approaches have described somatic mutations in several human tumor types (4, 5) and larger scale network studies, such as The Cancer Genome Atlas initiative, have characterized mutations in hundreds of tumors, profiled tumor transcriptomes, and cataloged cancer-related gene amplifications and epigenetic silencing in colon and rectal carcinoma (6). The resulting wave of data poses the critical question of how genomic variations translate to cancer phenotypes. Genes and transcripts execute most of their functions through the proteins they encode. Systematic characterization of cancer proteomes thus provides a means to understand the translation of genomic variation to cancer phenotypes.

Here we address the largely unexplored problem of how specific cancer-related mutations translate to functional alterations through proteomes. A recent study demonstrated proteomic changes driven by gene copy number changes in cancer cells (7), but the proteomic consequences of gene mutations and gene silencing events remain unknown. We compared a panel of 10 colorectal carcinoma cell lines that display different mutations in DNA mismatch repair (MMR) genes, as well as other colon cancer-associated genes. We used shotgun proteomics by liquid chromatography/tandem mass spectrometry (LC/MS-MS), which enables global proteome surveys that can identify thousands of proteins from milligram quantities of cells or tissue (8, 9). Shotgun analyses provide an unbiased, global inventory of proteomes, together with quantitative estimates of protein abundances that translate to biologic phenotypes (10).
We previously described methods to enhance global proteomic analyses using mutational and gene expression data obtained by transcriptome sequencing (RNA-seq; refs. 11, 12). With these approaches, proteomic analysis yields higher numbers of identified proteins and detects specific sequence mutations and variants. In addition, RNA-seq data also provide transcript expression information, which can be combined with protein expression levels to identify regulatory changes in biologic systems (13). Here we applied a combined proteogenomic analysis to explore the impact of MMR deficiency due to several distinct mutations and epigenetic silencing events. The data broaden our understanding of phenotypes associated with MMR and provide a template for future studies of how genomic and proteomic changes generate important cell phenotypes in cancer.

Materials and Methods

**Cell lines and proteomic analysis by LC/MS-MS**

All cell lines were obtained from American Type Culture Collection and grown as described previously (13). A summary of genetic features of the cell lines is provided in Supplementary Table S1. Three separate replicate cultures for each cell line were analyzed by shotgun proteomics as described by Liu and colleagues (13). Spectral files were searched against the Human ENSEMBL protein database (version 36, release 52) using Myrimatch (version 1.5.6; ref. 14). IDpicker version 3.0 was used to assign protein identifications to the identified peptides. The resulting dataset consisted of 6,094 protein groups with a 4.63% protein false discovery rate (FDR; Supplementary Tables S2 and S3).

**Proteome analysis using RNA-seq data**

Knowledge on transcriptome data can greatly enhance protein identification and expression level analyses, including that of variant peptide sequences (12). We generated whole transcriptome analysis for 9 of the 10 cell lines as described by Wang and colleagues (12). Because DLD-1 and HCT-15 were derived from the same colon cancer (15), we only generated the HCT-15 RNA-seq data and used this dataset for both HCT-15 and DLD-1 analyses. FPKM (fragments per kilobase of exon per million fragments mapped) was extracted from cufflinks reports for genes for which expression levels could be determined, otherwise the gene FPKM values were set to 0. After removing genes with total FPKM less than 5 across 9 cell lines, we identified 14,846 expressed genes.

We used RNA-Seq data from 9 cell lines to generate customized protein databases that included all sequence variations found in the transcriptome. The HCT-15 customized protein database was used for DLD-1. Putative single nucleotide variants (SNV) were called one library at a time with SAMtools mpileup and varFilter scripts (16). Then an R package customProDB was used to annotate the SNVs and generate a database by keeping nonsynonymous protein coding SNVs for each cell line. Protein searches were performed using the customized FASTA databases for each cell line resulting in a dataset containing 6,513 protein groups with a protein FDR of 3.17%. (Supplementary Tables S4 and S5). Given the improvement of this dataset over the one generated using the standard database, all analyses were performed using the results based on the customized FASTA database.

**Proteome analysis using microsatellite mutation data from SelTarBase**

The SelTarBase database (www.seltarbase.org) provides a compilation of all published microsatellite mutations in expressed sequences. Single-nucleotide shifts in microsatellites can lead to truncated proteins that theoretically should contain novel sequences at the carboxy-terminal end, depending on where protein translation termination signals (stop-codons) are located within the alternative reading frame. If such shifted protein sequences are detectable, they would be unique to the mutant cell and could potentially be used as specific markers. We created a FASTA protein search database containing all possible frame shifts in the 167 genes listed in SelTarBase that would lead to novel predicted sequences of at least 5 amino acids and a total of 358 new database entries. A complete list of these genes is provided in Supplementary Table S6. To account for read-through events, the first 3 stop codons were replaced by tryptophan for a subset of proteins.

**Protein expression comparisons based on spectral count data**

Statistically significant differences in protein spectral counts between different groups were calculated using a quasi-likelihood Poisson distribution implemented in Quasitil (17). Complete results from Quasitil analyses are presented in Supplementary Tables S7 (MMR+ cell lines vs. MMR− cell lines) and S8 (RKO cells vs. all other cell lines).

**WebGestalt functional class enrichment analysis**

We used WebGestalt (18) at http://bioinfo.vanderbilt.edu/webgestalt for the functional interpretation of shotgun proteomics data. Our reference set consisted of all proteins identified across all cell lines. Differential proteins had at least a 2-fold difference in spectral counts and a FDR-corrected quasi-likelihood P value of less than 0.05.

**Multiple reaction monitoring analyses**

Cell samples for multiple reaction monitoring (MRM) were prepared as for LC/MS-MS proteomics, except that peptide extracts were not subjected to isoelectric focusing (IEF), and analyzed using our labeled reference peptide method (19). Four optimized transitions for each peptide of corresponding proteins were selected using Skyline (Supplementary Table S9; ref. 20). A stable isotope labeled β-actin peptide [U-13C6, U-15N4-Arg]-GYSFTTTAER was used as an internal standard (60 fmol/injection) for relative quantification of target proteins. An unpaired t test was used to test for significant differences between samples (n = 3).

**Western blotting**

Cell pellets were resuspended in ice-cold RIPA buffer (150 mmol/L NaCl, 50 mmol/L Tris, 0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate, and protease inhibitors), sonicated for 10 seconds, incubated on ice for 10 minutes and clarified by centrifugation at 13,000 × g. Protein concentrations were measured with the BCA assay. Lysates from each cell line
(40 μg protein) were combined with 4× SDS loading buffer (Invitrogen), incubated at 70°C for 10 minutes, and proteins were separated on 10% SDS-PAGE mini-gels. Proteins were transferred to polyvinylidene difluoride membranes, which were probed with primary antibodies for MLH1 or MSH2 (Santa Cruz Biotechnology) and β-actin (Abcam). Membranes were probed with fluorophore-conjugated secondary antibodies (Invitrogen) and proteins were visualized on a fluorescent scanner (LI-COR Odyssey; LIC-COR).

**Cell proliferation and invasion assays**

Cell mobility was measured in triplicate using a model wound healing assay in which scraped "wounds" were created in confluent monolayers of cells with a pipette tip. Cell migration was observed at 0, 24, and 48 hours by cell growth within the scrape line and representative scrape lines for each cell line were photographed. Cell invasion capacity was measured on Transwell filters in serum-free medium. Serum containing medium (0.5%) was utilized as the chemoattractant in the lower chamber. After 72 hours of incubation, cells that had invaded to the lower surface membrane were fixed, stained and counted under a light microscope. Significance was evaluated with Student t test.

**Hierarchical clustering**

For unsupervised cluster analysis of the 10 cell lines, the dataset was limited to the 300 proteins with the highest variation in expression in the dataset. Normalized spectral counts were log-transformed and the 300 proteins with the highest variance were chosen for hierarchical cluster analyses and visualization in heat maps. For supervised cluster analysis, proteins were selected based on differential expression between the MMR+ and MMR− groups (3-fold difference, adjusted quasi-likelihood P value ≤0.05) and processed as described for unsupervised clustering.

**Coexpression network construction and coexpression module identification**

Multiple group comparison based on quasi-likelihood modeling was used to identify proteins that are differentially expressed across groups but consistent within replicates. To reduce data redundancy and facilitate downstream functional analysis, expression data was summarized to the gene level. We used a modified version of our previously published Iterative Clique Enumeration algorithm (21) to identify 167 coexpression modules from the coexpression network. We identified 9 modules that showed differential expression using t test with a corrected P value of less than 0.1 between MMR+ and MMR−cell lines (Supplementary Table S10).

**Results**

**Proteomic analysis of 10 colorectal cancer cell lines**

We analyzed a collection of 10 colorectal carcinoma cell lines that differ in their DNA MMR status (Supplementary Table S1). Shotgun proteomic analyses performed on triplicate cultures of each of the cell lines yielded a total of 6,094 proteins at a protein FDR of 4.63% with the use of a generic ENSEMBL protein database. More than half of these proteins were observed across all cell lines (3,459 proteins total), whereas the average number of protein identifications per cell line was 4,373 (range: 3,183–4,953).

**Using RNA-seq transcriptome data to extend global proteomic analyses**

Protein identification in global proteomic surveys depends on prior knowledge of the protein sequences present in a given sample. Until recently, the typical shotgun proteomic identification step involved matching the observed MS/MS spectra sequences in standard protein databases, which generally do not contain variant sequences. We developed a novel strategy to use RNA-seq transcriptome data to create cell line-specific, custom databases (12) that reduce the search space of the database search and allow the identification of sequence variants. This analysis yielded a total of 6,513 protein groups (3.17% protein FDR) identified by 59,803 distinct peptides. Because of this improvement of 6.9% in protein identifications over standard database searches, the customized database results were used for all subsequent analyses (Supplementary Table S5).

**Identification of variant sequence peptides**

A total of 59,803 distinct peptides were identified in the dataset, but only 763 distinct peptides representing 564 unique variant sequences compared with the standard ENSEMBL protein database (Supplementary Table S11). Of these 571 variant peptide sequences, 329 (58%) had known genomic counterparts due to single-nucleotide polymorphisms (SNPs). Of the remaining 242 peptides, 235 peptides harbored a single amino acid change and 7 peptides were the result of larger sequence changes. RNA-seq analysis identified 19,613 nonsynonymous sequence variants in coding regions across all cell lines, which gives an overall ratio of (detected sequence variants)/(all possible variants from RNA-seq) = 2.9%. The reason that not all possible variants were detectable include (i) protein expression levels of many variants were too low for detection by our analysis platform; (ii) peptide sequences surrounding the variant were not favorable for ionization and detection by MS; and (iii) possible effects of variant sequence on protein stability. To further evaluate the last possibility, we separated variant peptides into those listed in the dbSNP database (22) and those that were newly detected in our study. On average, variant sequences representing germline polymorphisms were twice as likely of being detected in the proteome than non-dbSNP, and presumably somatic, variant sequences (4.0% vs. 2.2%, respectively; P < 0.0001, χ² test). This suggests that newly acquired sequence variants have a negative effect on protein stability.

Several known cancer-associated genetic mutations were observed at the proteome level, including the KRAS codon 12 G12V variant in SW480, a D140G mutation in the RAS homologue RAB1B in HCT-15, and a TP53 P309S variant in SW480. Of the remaining variant peptides, a subset was chosen based on high spectral counts and distribution for verification by manual inspection and comparison with spectra obtained with synthetic peptides of the same sequence. Variant peptides derived from GPATCH4, CCT8, ANXA11, SRF59, CEACAM1,
and EPB41L1 were confirmed using this strategy (Supplementary Figs. S1A–S1F). Interestingly, the EPB41L1 sequence alteration changes a known serine phosphorylation site at codon 75 to leucine. EPB41L1 is a competitive inhibitor of AGAP2 (also called PIKE or PI3 Kinase Enhancer), a protein that binds and activates PIK3CA (23). AGAP2 is a nuclear GTPase that is frequently overexpressed in cancer cells. Possible regulation of EPB41L1 activity by phosphorylation might be important for PI3-kinase activity in the cell.

Hierarchical clustering analysis and evidence for MMR-related protein signatures

Unsupervised clustering of the 300 most variable proteins in the dataset grouped all biologic replicates together (Fig. 1). Of note, cell lines DLD-1 and HCT-15 grouped together; these cell lines were originally cultured by 2 investigators from different tissues harvested from the same patient and were subsequently shown to be genetically identical (15). Cell line RKO was most dissimilar to all the other cell lines, in large part due to the absence of keratins and extracellular matrix proteins in this cell line. In addition to these specific features, each of the 10 cell lines displayed distinct protein expression patterns, reflecting their unique proteomic characteristics.

Proteomic characteristics of epithelial-to-mesenchymal transition in RKO cells

Protein expression data from RKO was characterized by dramatic decreases in expression of keratins, catenins, cadherins, and related proteins involved in cytoskeletal structures and cell adhesion (Supplementary Table S8). RKO cells differentially expressed 202 proteins as compared with the 9 other cell lines. The 105 proteins downregulated in RKO were significantly enriched in GO cellular component classes of cell–cell junction (GO:0005911, \( P < 0.0001 \)), cell periphery (GO:0071944, \( P < 0.0001 \)), and other classes related to extracellular matrix and cell–cell interactions. Similarly, KEGG categories for cell adhesion molecules (KEGG:04514, \( P = 0.007 \)), adherence junctions (KEGG:04520, \( P = 0.007 \)), and tight junctions (KEGG:04530, \( P = 0.007 \)) were downregulated in RKO. These differences suggested loss of epithelial characteristics in RKO cells, a hallmark characteristic of the epithelial-to-mesenchymal transition (EMT) phenotype. Transcription factor target
enrichment analysis of proteins with lower expression in RKO also were consistent with EMT. A large protein cluster was significantly enriched for targets of transcription factor TCF3 (also known as E2A; \(P = 0.0008\)). This analysis also identified significant enrichment for downregulated proteins associated with the transcription factor ZEB1 (also known as AREB6 and TCF8), which is implicated in EMT, suppresses CDH1, and is known to be highly expressed in RKO \((P = 0.008\); ref. 24).

The combined results provide additional evidence that RKO cells display an EMT phenotype, a characteristic of invasive and metastatic cancers characterized by loss of intercellular contacts, loss of baso-apical polarity, gain of mesenchymal markers and increased invasive properties (25). To determine whether the EMT-like proteomic features of RKO cells conferred a functional EMT phenotype, we compared RKO and SW480 (non-EMT) phenotypes using wound-healing and invasion assays. In the wound-healing assay, RKO cells completely covered a scratched surface within 24 hours, whereas open areas remained with SW480 cells (Fig. 2A). In a cell invasion assay (growth through TransWell filters), RKO cells were approximately 2.5-fold more invasive than SW480 (Fig. 2B), thus confirming growth characteristics consistent with an EMT phenotype.

Identification of proteomic changes associated with MMR deficiency

The genetic status of MMR genes in all of the cell lines in this study is known. This allowed us to study the proteomic correlates with known genetic and epigenetic alterations involving the MMR proteins MLH1, MSH2, and MSH6. Figure 3 summarizes the combined shotgun proteomic and MRM data for MLH1, MSH2, and MSH6, which indicate that the proteins corresponding to the mutated or silenced genes all are either absent or significantly downregulated. We also examined the proteins BAX, MRE11, LMAN1, and RAD50, which contain mononucleotide repeats within their coding sequences and are prone to frameshift mutations secondary to loss of MMR, thus leading to loss of protein expression. Accordingly, BAX expression was lost in LoVo and LS174T cells, LMAN1 expression was lost in LoVo while RAD50 levels were lowered, but not lost in all of the MMR–cell lines, as none have homozygous inactivation of the gene. We also observed low levels of MRE11A protein expression in LS174T, HCT–116, and LoVo, due to intronic microsatellite mutations, which lead to exon-skipping in these cell lines. MRM analyses confirmed all of the differences observed in shotgun MS analyses. Western blot analyses of MLH1 and MSH6 in several of the cell lines confirmed the protein and MRM data. MRM analyses of additional proteins found to be differentially expressed in MMR phenotypes verified the expression differences detected by shotgun proteomics (Supplementary Fig. S2A–S2T). These analyses demonstrate the validity of protein expression assessment through shotgun proteomic analyses and confirm expression changes in this group of MMR-associated proteins in the panel of colorectal carcinoma cell lines.

Proteomic consequences of MMR-associated slippage events in MMR–cell lines

Several hundred human genes harbor mononucleotide tracks in their coding sequences, tracks that are prone to DNA replication slippage in MMR–cells (Supplementary Table S6). Because such slippage events alter the reading frame, they lead to novel C-terminal amino acid sequences in the affected proteins. We generated a customized FASTA database containing 167 human genes with known mononucleotide elements from SelTarBase that predicted potentially detectable peptides and used this FASTA database to search the proteomic data for all 10 cell lines. The results from this search included several low-quality matches to shifted peptide sequences (approximately 1–2 matches per cell line), but none of these matches could be validated through manual inspection of the assigned MS/MS spectra (data not shown). The conclusion from this proteomic search is that cells either remove mRNAs containing premature stop codons or that the resulting altered proteins are unstable and do not reach levels that allow detection in survey-type shotgun proteomics experiments.

Coexpression network analysis reveals modular organization of proteomic adaptations to MMR deficiency

To detect coordinated proteomic adaptations to MMR deficiency, we performed coexpression network analysis and
Figure 3. Shotgun and MRM analyses provide consistent measurements of selective loss in expression of MMR proteins (MLH1, MSH2, and MSH6, A–D), selected protein products of MMR-sensitive target genes (BAX, MRE11, LMAN1, and RAD50, E–H). Shotgun proteomics data are plotted as spectral counts for triplicate analyses, whereas MRM data are plotted as summed signal intensity for measured transitions normalized to summed intensity for transitions measured for a reference peptide. Star symbols above the bars indicate genotype for the corresponding gene, as summarized in Supplementary Table S1 and in the cited literature. C, protein blot analysis for MLH1 and MSH2 confirming protein levels observed by mass spectrometry.
identified 167 coexpression modules. From these, we identified 9 modules that showed the largest differential expression between MMR\(^+\) and MMR\(^-\) cell lines that represent coordinated proteomic changes (Supplementary Table S10). We evaluated the 76 proteins comprising these modules to identify common functional classifications using Gene Ontology categories and identified 4 groups of related functions (Supplementary Table S12), including (i) protein turnover, transport, and folding (18 proteins), (ii) metabolic processes (16 proteins), (iii) DNA/RNA synthesis and repair (14 proteins), and (iv) transcription regulation (10 proteins; Fig. 4).

The largest group (group a) contains proteins involved in regulation of protein folding (DECR1, TRAP1, HSPH1, LMAN1, FKBP2, FKBP4, WFS1, and ERLIN2), protein turnover (UBXN1, DPP7, TFRC, SH3KBPI, EFTUD1, and C10orf118), and protein transport (DNM2, NAPA, TIMM23, and ABHD11). The functions represented in these coexpression modules suggest a coordinated program of adaptation to the translation of misfolded, variant polypeptides in MMR\(^-\) cells. Several of the proteins identified in this module are involved in HSP90-mediated cellular stress response (TRAP1, HSPH1, and LMAN1); other proteins involved in this response that are significantly upregulated in MMR\(^-\) cells, but were not identified through coexpression network analysis include STUB1, HSP90AA, HSP90AB1, CDC37, and ATXN2 (Supplementary Table S7). These data suggest enhanced chaperone expression and activity of the cellular response to proteins with variant sequences that need to be removed from the cell.

The second coexpression module consisted of 16 metabolic proteins with higher levels in MMR\(^-\) cell lines compared with MMR\(^+\) cell lines, most of which were located in the mitochondria. Several of these proteins, for example CISD1, NDUFAF3, ATP5B, MUT, OXCT1, and PDK1, play critical roles in glucose and fatty acid metabolism. The discovery of this network of proteins is unexpected and one could speculate that MMR deficiency places added metabolic demands on cells due to the need to maintain adaptive responses in protein turnover.

The third group, categorized as DNA/RNA synthesis and repair, included several DNA repair proteins from module 75 (RAD50 and MRE11A) and the related double-strand break repair protein NBN, which are constituents of the Mre11–Rad50–Nbs1 complex involved in nonhomologous end joining (NHEJ; ref. 26). In addition, module 75 contains a component of the origin recognition complex, ORC3, essential for the initiation of DNA replication in eukaryotic cells and a subunit of the CCR4-NOT complex, CNOT4 that functions as a general transcription regulation complex. All these proteins are coordinately downregulated in MMR\(^-\) cells, suggesting decreased DNA synthesis and repair in MMR\(^-\) cells (Fig. 4). Other proteins in this group are indicative of RNA surveillance (DIS3 and UTP20) and RNA splicing (DHX15, ESRP1, and SFPQ; Supplementary Table S12).

Taken together, the functions represented in these coexpression modules suggest a coordinated program of adaptation to the translation of misfolded, variant polypeptides in MMR\(^-\) cells. Although consistent with enhanced chaperone expression and protein translation, the coexpression of other proteins involved in metabolism, vesicle trafficking and transcription regulation suggest a previously unrecognized scope...
of adaptation to proteotoxic stress secondary to DNA MMR deficiency.

To study the proteomic consequences of MMR, we used quasi-likelihood modeling to identify proteins that were differentially expressed between the MMR+ and MMR− cell lines. A total of 245 protein entries were statistically significant at P < 0.05 and these were visualized in a heatmap in Fig. 5. This figure illustrates global common features that are affected in response to MMR status but also indicate proteomic characteristics that are unique for individual cell lines. For example, a cluster of proteins is highly expressed in Caco-2, including APOE, FN1, DCDC2, LAMA5, CD74, and MYL3; all proteins involved in cellular motility (GO:0048870). Of the proteins with higher expression in MMR− cell lines compared with MMR+ cell lines, HSP90AA1, HSP90AB1, CCT6B, CCT8, DNAJA2, and CDC37 bind to unfolded proteins (GO:0051082) and STUB1, HSP90AA1, and CCT6B are involved in chaperone-mediated protein complex assembly (GO:0051131). These results indicate a potential upregulation of proteins that manage degradation of aberrant proteins in the cell.

Discussion

Colorectal tumors with MIN comprise a major subset of colorectal cancers and are notable for distinct clinical characteristics (27) and high frequency of mutations (6). MIN tumors have been distinguished previously from other colon cancer types by gene expression profiles (28–30), which suggest broad adaptations to MMR deficiency. In MIN tumors, increased mutation frequency increases levels of frame-shifted sequences, which generate premature stop-codons in the mRNA. These abnormal mRNAs are selectively removed by nonsense-mediated decay (NMD), a process that is active in MIN tumors (31). This process is highly effective in removing mRNAs on which translation has stalled, but not all nonsense-containing mRNAs are sensitive to NMD and the process does not remove mRNAs that harbor single nucleotide sequence variants. Thus, increased production of sequence variant proteins in MIN tumors may demand adaptation of protein quality control mechanisms, which could contribute to the MIN phenotype.

Using gene expression analysis, Banerjea and colleagues (30) identified upregulation of immunomodulatory genes in MIN tumors, such as heat shock proteins, chaperone molecules and cytokines, and linked these findings to processing and presentation of antigenic peptides. Increased levels of somatically mutated peptides could result from increased mutational rates observed in MMR-defective tumors, and these peptides may be related to the strong immunogenic response triggered by MIN tumors (32, 33). However, the immunomodulatory response described by Banerjea and colleagues was observed in primary carcinomas and it is unclear if the gene expression changes occurred in the cancer cells or in the inflammatory component of the specimens. In addition, the study was based on mRNA expression profiling and did not study functional significance at the protein level. Our study provides additional information by studying proteomic changes in tumor cell lines and to our knowledge provides the most detailed documentation of the extent to which defects in MMR translate into the proteome.
Our proteomic findings are in agreement with the notion that MMR– cells have upregulation of cellular processes that handle aberrant proteins, either through protein folding, degradation and unfolded protein binding, which could enable MMR– cells to counteract the deleterious effects of abnormal protein load. Evidence for increased protein folding activity came from the presence of FKBP2 and FKBP4, LMAN1, TRAP1, and WFS1, whereas a separate coexpression module indicated increased ubiquitin-dependent protein degradation: UBXN1, DPP7, TFRC, and ERLIN2. The involvement of ubiquitin-dependent proteasomal degradation is in agreement with a recent paper by Kim and colleagues, demonstrating that aberrant mRNAs that escape NMD lead to mutant proteins that are degraded via the ubiquitin–proteasome system (34).

Apart from protein expression changes, our data provide a separate line of evidence for a cellular response to abnormal proteins. The acquisition of RNA-seq data for all cell lines and the application of our newly developed pipeline for the detection of variant peptide sequences from shotgun proteomic data (12) allowed us to quantify the levels of variant peptides resulting from germline polymorphisms (SNP) and from somatic mutations. This analysis clearly showed that SNP-encoded variant peptides were more likely to be detected in the shotgun proteomic datasets compared with somatically acquired variant peptides. These data suggest that proteins harboring new sequence variants are less stable in the cell than proteins with variant sequences encoded by germline SNPs. This observation is also consistent with our failure to detect novel peptides resulting from frame-shifted protein coding sequences in the limited number of genes that harbor repeated elements as targets for MMR. The unavailability of germline data precludes the analysis of individual sequence variants by comparing DNA from normal tissues to tumor DNA from the same patient. Nevertheless, for global comparisons, it seems reasonable to postulate that sequence variant peptides in the same patient. Nevertheless, for global comparisons, it seems reasonable to postulate that sequence variant peptides in the different SNP databases are more likely to be germline polymorphisms than new sequence variants.

Coexpression network analysis identified a large module of proteins that included proteins involved in DNA and RNA synthesis and repair. The driving proteins behind these modules are the MMR proteins responsible for the MMR phenotype (MLH1, MSH2, MSH6, etc.) and proteins involved in repair of double-strand breaks (RAD50, MRE11A, NBN, etc.). RAD50 and MRE11A have lower expression levels in MMR– cells because of mutations caused by defective MMR, but increased protein levels of SF3P, PFA3, ANXA3, and ORC3L suggested increased DNA synthesis and repair in MMR+ cells compared with MMR– cells. Our proteomic data did not indicate any increase in proteins involved in NMD, although several of the key NMD proteins were detectable in the dataset (UPF1, -2, and -3B). Coexpression network analysis also indicated differences in transcription regulation and in the process of metabolism between MMR+ and MMR– cells. To our knowledge, an association between MMR status and changes in cellular energy metabolism has not been reported. Our findings suggest that such an association might exist and provide a rationale for further study to elucidate the possible relationship between these biologic processes.

Our analyses also indicated an EMT phenotype of RKO cells based on proteomic analyses. The dramatic downregulation of CDH1 observed in RKO cells is a hallmark of EMT (25, 35). However, we observed 105 other downregulated proteins in RKO cells, which includes many components of cell adhesion and cytoskeleton and proteins that regulate these networks. Gene enrichment analysis of these proteins indicated activation of the EMT-associated transcription factor ZEB1. Although some EMT characteristics of RKO have been reported previously (24) and analyses of EMT-like models with other analysis platforms have been reported (36–38), our data provide the largest proteomic dataset to describe this phenomenon. Our detection of increased mobility and invasive properties of RKO compared with one of the other cell lines, SW480 provides preliminary confirmation of an EMT phenotype.

Here we demonstrate that standardized proteome analysis platforms can capture the protein expression characteristics of distinct phenotypes driven by cancer-related mutations. Our data provide the first insight into the proteomic consequences of DNA MMR deficiency and indicate that widespread mutational load leads to an adaptive stress response that allows cells to remove mutant proteins. Sequence variations may produce misfolded proteins that are subsequently degraded through the ubiquitin-dependent proteasomal pathway. In addition, global proteomic analyses proved sensitive enough to detect an EMT phenotype in one of the cell lines and features of this phenotype were detectable using biologic assays. This work demonstrates the potential of mass spectrometry-based global protein analyses and subsequent confirmation using targeted protein detection. Recent work by the NCI (National Cancer Institute) Clinical Proteomic Technology Assessment for Cancer network has demonstrated the feasibility of implementing standardized proteome analyses across multiple laboratories (39–41) and we anticipate that these tools will dramatically expand our ability to understand the association between cancer-related genomic variation and cancer phenotypes.

Disclosure of Potential Conflicts of Interest

P.J. Halvey is employed as a scientist at Momenta Pharmaceuticals. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions

Conception and design: P.J. Halvey, D.C. Liebler, R.J.C. Slebos
Development of methodology: P.J. Halvey, X. Wang, B. Zhang, D.C. Liebler, R.J.C. Slebos
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): P.J. Halvey, P. Dhawan, R.J.C. Slebos
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): P.J. Halvey, X. Wang, J. Wang, P. Dhawan, M. Li, B. Zhang, R.J.C. Slebos
Writing, review, and/or revision of the manuscript: P.J. Halvey, X. Wang, M. Li, B. Zhang, D.C. Liebler, R.J.C. Slebos
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): R.J.C. Slebos
Study supervision: B. Zhang, D.C. Liebler, R.J.C. Slebos

Acknowledgments

The authors thank Misti Martinez, Kristin Carpenter, and Sarah Stuart for technical assistance. They also thank Zhiao Shi for programming assistance.
Grant Support
This work was supported in part by NIH grants U54CA126479 and U24AI09988 (D.C. Liebler) and R01GM088822 (B. Zhang).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 29, 2013; revised October 30, 2013; accepted November 6, 2013; published OnlineFirst November 18, 2013.

References


Proteogenomic Analysis Reveals Unanticipated Adaptations of Colorectal Tumor Cells to Deficiencies in DNA Mismatch Repair

Patrick J. Halvey, Xiaojing Wang, Jing Wang, et al.

Cancer Res  Published OnlineFirst November 18, 2013.

Updated version  Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-13-2488

Supplementary Material  Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2013/11/18/0008-5472.CAN-13-2488.DC1

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.