Cancer Therapy: Preclinical

Gene Expression Patterns in Mismatch Repair-Deficient Colorectal Cancers Highlight the Potential Therapeutic Role of Inhibitors of the Phosphatidylinositol 3-Kinase-AKT-Mammalian Target of Rapamycin Pathway

Eduardo Vilar,1 Bhramar Mukherjee,7 Rork Kuick,6 Leon Raskin,1 David E. Misek,5 Jeremy M.G. Taylor,7 Thomas J. Giordano,1,4 Samir M. Hanash,8 Eric R. Fearon,1,3,4 Gad Rennert,9 and Stephen B. Gruber1,2,3

Abstract

Purpose: High-frequency microsatellite-instable (MSI-H) tumors account for ~15% of colorectal cancers. Therapeutic decisions for colorectal cancer are empirically based and currently do not emphasize molecular subclassification despite an increasing collection of gene expression information. Our objective was to identify low molecular weight compounds with preferential activity against MSI colorectal cancers using combined gene expression data sets.

Experimental Design: Three expression/query signatures (discovery data set) characterizing MSI-H colorectal cancer were matched with information derived from changes induced in cell lines by 164 compounds using the systems biology tool “Connectivity Map.” A series of sequential filtering and ranking algorithms were used to select the candidate compounds. Compounds were validated using two additional expression/query signatures (validation data set). Cytotoxic, cell cycle, and apoptosis effects of validated compounds were evaluated in a panel of cell lines.

Results: Fourteen of the 164 compounds were validated as targeting MSI-H cell lines using the bioinformatics approach; rapamycin, LY-294002, 17-(allylamino)-17-demethoxygeldanamycin, and trichostatin A were the most robust candidate compounds. In vitro results showed that MSI-H cell lines due to hypermethylation of MLH1 are preferentially targeted by rapamycin (18.3 versus 4.4 μM; P = 0.0824) and LY-294002 (15.02 versus 10.37 μM; P = 0.0385) when compared with microsatellite-stable cells. Preferential activity was also observed in MSH2 and MSH6 mutant cells.

Conclusion: Our study shows that the phosphatidylinositol 3-kinase-AKT-mammalian target of rapamycin pathway is of special relevance in mismatch repair-deficient colorectal cancer. In addition, we show that amalgamation of gene expression information across studies provides a robust approach for selection of potential therapies corresponding to specific groups of patients.

Colorectal cancer is the third most common cancer in both men and women in western countries (1). Although the management of patients with colorectal cancer has changed dramatically over the past decade, with relevant improvement in survival (2), a significant proportion of patients relapse after a combination of surgery and adjuvant therapies and do not respond to treatment in the metastatic setting. Therefore, it is crucial to find additional therapies for this disease. Moreover, although our knowledge of the molecular biology of colorectal cancer has advanced, the molecular data have not yet offered insights into clinically significant tumor subgroups with distinct therapeutic responses, other than the recently described evidence that K-ras mutant colorectal cancers are less responsive to treatments targeted against the epidermal growth factor receptor (3–5). Colorectal cancer can also be classified based on the relative proportion of microsatellite loci that are found to show instability in cancer cells relative to the patient’s germline configuration for the locus, and consensus criteria distinguish colorectal cancer showing high-frequency microsatellite instability (MSI-H), low-frequency MSI (MSI-L), or microsatellite stability (MSS) (6). About 15% of apparently sporadic colorectal cancer are MSI-H (7, 8) mainly due to
Translational Relevance

Therapeutic decisions for colorectal cancer currently do not emphasize molecular subclassification. In addition, drugs selected for further development rest heavily on the activities observed in preclinical models albeit in the context of an incomplete understanding of the biology of the disease. Our aim was to select small molecules with preferential activity against microsatellite-instable (MSI) colorectal cancer using a systems biology tool to combine gene expression data coming from different sources. From 164 candidate compounds, rapamycin and LY-294002 showed preferential selectivity in vitro to cell lines displaying high-frequency MSI. Statistical methods developed for this analysis show that an amalgamation of gene expression information across studies provides a robust approach for selection of potential therapies. In addition, these results point toward the therapeutic importance of the phosphatidylinositol 3-kinase-AKT-mammalian target of rapamycin pathway in MSI and open the door for the eventual evaluation of anti-mammalian target of rapamycin and anti-phosphatidylinositol 3-kinase therapies in this recognized subset of colorectal cancers.

epigenetic silencing of the promoter of MLH1 gene (9). A small fraction of the MSI-H cases, perhaps 2% to 3%, are due to germ-line mutations in one allele of one of the mismatch repair genes (e.g., MLH1, MSH2, MSH6, and PMS2; refs. 7, 8), which are the cause of hereditary nonpolyposis colorectal cancer (Lynch syndrome; ref. 10). MSI-H tumors are characterized by right-sided location, diagnosis at younger age, lower pathologic stage, high histologic tumor grade, mucinous phenotype with prominent tumor-infiltrating lymphocytes, and better prognosis than MSI-L/MSS cases (11–13). Several studies have shown an apparent lack of responsiveness and, in some cases, a resistance to chemotherapy regimens containing 5-fluorouracil in MSI-H colorectal cancers (14, 15). On the other hand, some preliminary data suggest that MSI-H colorectal cancers may be more sensitive to irinotecan (16–18). Therefore, the MSI-H phenotype has the potential to guide therapeutic decisions based on drugs specifically selected to take advantage of the molecular characteristics of the tumor.

In the last decade, a large collection of gene expression data describing tumor subtypes as well as the response of cells to various small molecules has been obtained. Analytic tools that interrogate the vast collection of gene expression data offer an opportunity to identify possible agents targeting specific tumor subtypes based of this biological information. The "Connectivity Map" is a novel resource that has been developed by researchers at the Broad Institute (19). It has the potential to reveal statistical "connections" between diseases and drug treatments using gene expression data. To use the "Connectivity Map," researchers select a list of differentially expressed genes, named as a query signature, often distinguishing a biological state of interest. In parallel, the "Connectivity Map" stores a reference database of gene expression profiles from a diverse set of established cell lines that have been exposed to 164 small-molecule "perturbagens." Similarities between each of the reference expression profiles stored in this repository and the initial query signature are assessed using a nonparametric, rank-based pattern-matching strategy akin to the well-known approach of Gene Set Enrichment Analysis (20). The query signature is then compared with each rank-ordered reference list to determine whether up-regulated (down-regulated) query genes tend to appear near the top (bottom) of the reference list (positive connectivity) or vice versa (negative connectivity), and a connectivity score between -1 and +1 is assigned (19). Based on this analytic method, "perturbagens" are classified as positively or negatively correlated with the query signature and ranked between them according to the strength of connectivity scores. Those agents receiving negative connectivity scores are inversely associated with a particular biological state (19, 20).

Here, we describe our approach for combining results obtained through the "Connectivity Map" using five independent gene expression profiles from primary colorectal cancers divided into discovery and validation data sets. Our analysis has identified that MSI-H colorectal cancer is specifically sensitive to compounds inhibiting the phosphatidylinositol 3-kinase (PI3K)-AKT-mammalian target of rapamycin (mTOR) pathway. In our in vitro experiments, these compounds have shown their preferential antiproliferative and cytotoxic activities in MSI-H cell lines when compared with MSS cells.

Materials and Methods

Gene expression data sources. Five gene expression data sets were used to identify the gene expression signatures distinguishing MSI-H from MSS cancers. Four of them were publicly available and one was generated from our own data of the Molecular Epidemiology of Colorectal Cancer (MECC) study. The MECC study is a population-based, matched case-control study, which has been described previously (21). The study was approved by the institutional review boards at the University of Michigan and Carmel Medical Center in Haifa. Written informed consent was required for eligibility.

We characterized gene expression in MSI-H using a subset of 51 frozen colorectal cancer specimens originally studied by our group for molecular classification of epithelial cancers. The Affymetrix HuGeneFL array platform was used (22). Methods regarding RNA isolation and microarray procedures are detailed in Supplementary Methods. Normal and tumor DNA were extracted from microdissected DNA and analyzed for the consensus panel of five markers (6) to assess the MSI status as described previously (13). Of these tumors, 38 were identified as MSS and 13 were identified as MSI-H (Supplementary Table S1). As expected, the MSI-H group showed a statistically significant association with the presence of BRAF V600E mutation (P = 0.004, χ² test) and right side of the colon (P = 0.006, χ² test). Although tumor location may be related with gene expression (23), we have reported previously in a larger cohort of the MECC study that MSI status is more strongly associated with gene expression than location (24). Using a two-sample t test, we identified a list of genes (71 probe sets) that were differentially expressed at significant levels (P < 0.00001) between tumors grouped by microsatellite status (Fig. 1; Supplementary Table S2), whereas in 1,000 data sets in which the sample labels were randomly permuted we found only 0.084 probe sets giving this small P value on average, so that we estimated the false discovery rate for these 71 probe sets to be ~0.1%. Forty-one probe sets showed higher and 30 lower expressions in MSI-H compared with MSS cancers. Finally, enrichment testing of functional lists of genes was done using this data set. A detailed explanation is provided in Supplementary Information.

We then identified a total of four previously published studies characterizing the gene expression of the MSI subtype in primary tumors through searches in PubMed (23, 25–27). Results in these studies were generated using the HG-U133A Affymetrix platform, the
same platform used in “Connectivity Map.” The consistency of the platform used in these four studies minimized potential differences in annotation and facilitated relatively homogeneous data sets. Each column corresponds to a separate sample and grouped by microsatellite status. Each row represents a probe set with \( P < 0.00001 \) (false discovery rate < 0.1%) whose expression is color-coded according to the indicated scale and displayed in alphabetical order. Gene symbols of HuGeneFL probe sets and corresponding \( P \) values are shown (right). Note that general transcription factor II A (GTF2A2), eukaryotic translation initiation factor 5a (EIF5A), splicing factor arginine/serine-rich 6 (SFRS6), thymidylate synthetase (TYMS), a member of ras oncogene family Rab27b (RAB27B), and protein kinase cAMP-dependent type II (pka2b) showed upregulation in MSI-H cancers. On the other hand, protein kinase C (PRKCI), murine homologue 1 (MLH1), and transforming growth factor-\( \beta \) receptor II (TGFBR2) showed lower expression in MSI-H cancers. A complete list with HuGeneFL probe set ID, gene symbol, and gene title is presented in Supplementary Table S2.

**Cell culture.** HCT-116, LoVo, HCT-15, HCT-8, SW-48, RKO, SW-480, and HT-29 cell lines were selected for experiments based on mismatch repair gene mutations and microsatellite status obtained from the Wellcome Trust Sanger Institute Cancer Genome Project Web site and from previous publications (refs. 28, 29; Supplementary Table S9). Cell lines were obtained from the American Type Culture Collection. HCT-116 and LoVo harbor mutations in MLH1 and MSH2. HCT-15 and HCT-8 have mutations in MSH6. These two groups of cells display MSI-H and are consistent with the mechanism of mismatch repair deficiency in patients affected by hereditary nonpolyposis colorectal cancer. RKO and SW-48 are MSI-H lines due to hypermethylation in the promoter of hMLH1 gene, similar to sporadic colorectal cancer displaying MSI-H. HT-29 and SW-480 are wild-type for mismatch repair genes and are MSS. Cells were grown in DMEM/F-12 supplemented with 10% fetal bovine serum. All tissue culture reagents were obtained from Life Technologies.

**Treatment of cell lines with small molecules.** 17-(Allylamino)-17-demethoxygeldanamycin (17-AAG), LY-294002, rapamycin, and trichostatin A were obtained from Sigma-Aldrich. All the drugs were dissolved in DMSO to give the stock solutions (Supplementary Table S14). Stock concentrations were dissolved in supplemented medium to obtain seven serial 3-fold dilutions to perform cytotoxicity experiments.

**Cytotoxicity experiments.** Cytotoxicity assays were done in 96-well plates and cell proliferation was assessed using the WST-1 (Roche
Diagnostics) reagent according to the manufacturer’s protocol. The protocol followed is detailed in Supplementary Methods.

Analyses of BRAF and KRAS mutations. Tumor DNA extracted from paraffin-embedded slides and from cell lines was amplified by PCR. Primers used are described in Supplementary Methods.

MSI analysis of cell lines. Genomic DNA was prepared from cells exposed to 3 μmol/L rapamycin for 5 days and control cells. The sequences of the primers for PCR and details of the technique are described in Supplementary Methods.

Cell cycle analysis and Annexin V assays. We followed standard procedures as detailed in Supplementary Methods.

Results

Gene expression data defining MSI colorectal tumors show an apparent lack of overlap. As shown in Fig. 2, we separated the signatures in two groups: (a) a discovery set including two previously defined signatures (Watanabe et al. and Koinuma et al.) and the MECC signature we report above, which we used to search for the candidate compounds (23, 25), and (b) a validation set, which included the remaining two signatures (26, 27). This second set was used to test the reproducibility of the ranking and selection mechanism and assess concordance between the two generated lists of compounds. In spite of using the same platform, gene expression data defining MSI-H colorectal tumors showed an apparent lack of overlap most likely related to significance thresholds, already described in other settings (30). To summarize the information across the studies, for each set, we decided to build two artificial signatures, named “intersection” and “union,” using the component gene lists. The “intersection” signature integrates those probe sets highly differentially expressed between MSI and MSS tumors reported in original publications that appear in common in at least two signatures (Fig. 2). As shown in Supplementary Tables S3 and S4, this signature was composed of a total of six up-regulated and four down-regulated genes in the discovery set and three up-regulated and four down-regulated genes in the validation set. Therefore, the “union” signature was built grouping together the probe sets from all the highly differentially expressed genes from every signature in each set. Ultimately, we had three original and two artificial signatures leading to five lists in the discovery set and two original and two artificial signatures leading to four lists in the validation set.

Fig. 2. Gene expression data sets used and the flowchart for generation of a final compounds list. Gene expression data coming from five different studies were divided in two sets, discovery and validation sets. In addition, two artificial signatures were created for each set, the “intersection” and the “union” signature. Information on the total number of MSI and MSS tumors, number of significantly expressed up-regulated and down-regulated probes, and the array platform used to generate expression data are indicated in the rectangles representing every study. Number of compounds that passed the two filters per study is indicated in parentheses. Note that filters applied to "Connectivity Map" results rule out those compounds with only one experimental instance, thus limiting the total number of compounds to 95.
Identification of small molecules using “Connectivity Map” and the process of validation using an independent data set. After applying every individual query signature (including the artificially synthesized “union” and “intersection” signature) to the “Connectivity Map,” a total of five and four lists of compounds were retrieved from the discovery and validation sets, respectively (Fig. 2). To select those drugs whose reference profiles consistently yield negative connectivity scores across the various input query lists, sequential filter and ranking methods were developed. Our filter integrated two sequential criteria, excluding those compounds that failed to pass to the next step. The first filter selects those candidate compounds that have more than one negative connection, and the second filter sorts those having at least 50% of the observed non-null connections negative for a given signature in an attempt to eliminate compounds, which exhibit both positive and negative connections in different experimental instances. The two filters were applied to each individual gene list. It is relevant that the first sequential filter eliminates any compound that has been tested only once in the “Connectivity Map” and then reduces the final number of compounds introduced in our analysis to 95 small molecules. Due to this fact, we considered a separate strategy for examining the one-time tested compounds \((n = 69)\) where we restricted our attention to those compounds obtaining a strong negative connectivity score (less than -0.6) in at least two signatures (Supplementary Tables S5 and S6).

Those compounds fulfilling the above two criteria in at least two signatures of every set were selected to be eligible for being ranked. Ranking compounds was based on four different criteria: \((a)\) total number of negative connectivity scores, \((b)\) percentage of the difference between negative and positive connectivity scores \([\text{(no. negative connectivity scores - no. positive connectivity scores)} / \text{(no. negative connectivity scores + no. positive connectivity scores)]}\), \((c)\) mean connectivity score across signatures, and \((d)\) aggregated measure of strength of statistical evidence obtained by combining the permutation-based \(P\) values, which were retrieved as direct outputs from the “Connectivity Map” \([-\Sigma \log(P)\text{, the sum extending over all signatures included in a particular set under consideration}]\).

### Assessing reproducibility of the selection procedure by comparing the compound lists generated from discovery and validation sets

In the final lists, the compounds were ordered based on the average of the four different ranking criteria. Of the 164, the ranked list contains a total of 27 in the discovery set and 20 in the validation set (Supplementary Tables S7 and S8). As shown in Table 1, a direct comparison of the final lists of the discovery and validation sets revealed that 14 of 27 compounds were replicated. Also, 6 compounds appeared consistently within the 15 highest ranked in the two sets: LY-294002, 17-AAG, monorden, rapamycin, geldanamycin, and SC-58125. A close examination of the entire list revealed the presence of compounds that are recognized demethylators or that act through the inhibition of related molecules as histone deacetylases (trichostatin A and valproic acid). Others interact with heat shock proteins (17-AAG, 17-dimethylaminoethylamino-demethoxy-geldanamycin, monorden and geldanamycin) or inhibit the PI3K-mTOR pathway (rapamycin, LY-294002, and wortmannin).

Based on the results from the discovery set, we selected 17-AAG, rapamycin, LY-294002, and trichostatin A as the most promising candidate drugs to carry forward in vitro using different MSI and MSS colorectal cancer cell lines. Note that 17-AAG, rapamycin, and LY-294002 appeared to be highly rated in both the discovery set and the validation set. Trichostatin A, although highly ranked in the discovery set, was not in the validation list. Hence, the validation set was constructed not for selecting the compounds but for assessing the reproducibility of our ranking and selection mechanism.

Our hypothesis about the biological value of this tool is that connections established by the “Connectivity Map” between disease states and drugs highlight molecular events that are essential for a specific tumor subtype. Therefore, the effects of these compounds on the in vitro models should be predicted based on the consequences that the inhibition of these pathways have on cancer cells. Inhibition of heat shock proteins \((31)\), PI3K \((32)\) and mTOR pathways \((33)\), as well as demethylating agents \((34)\) have proven to affect survival, proliferation, and cell growth on different models of cancer. Therefore, in a first step, we decided to assess the cytotoxicity of these compounds. In a second step and based on the rationale that compounds selected using the “Connectivity Map” could reverse a disease state, we explored whether changes on microsatellite markers are present on cells after treatment with the most active compound.

### Testing final candidate compounds in cell lines sorted by microsatellite status

As shown in Fig. 3A and B, cytotoxicity experiments revealed that rapamycin and LY-294002 preferentially target cell lines with MSI due to hypermethylation in the MLH1 promoter. MSS cell lines showed 4-fold higher

---

**Table 1. List of 27 ranked compounds from discovery set**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Compound</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LY-294002</td>
</tr>
<tr>
<td>2</td>
<td>17-AAG</td>
</tr>
<tr>
<td>3</td>
<td>Trichostatin A</td>
</tr>
<tr>
<td>4</td>
<td>17-Dimethylaminoethylamino-demethoxy-geldanamycin</td>
</tr>
<tr>
<td>5</td>
<td>5224421</td>
</tr>
<tr>
<td>6</td>
<td>Geldanamycin</td>
</tr>
<tr>
<td>7</td>
<td>Wortmannin</td>
</tr>
<tr>
<td>8</td>
<td>Valproic acid</td>
</tr>
<tr>
<td>9</td>
<td>Monorden</td>
</tr>
<tr>
<td>10</td>
<td>Resveratrol</td>
</tr>
<tr>
<td>11</td>
<td>Prazosin</td>
</tr>
<tr>
<td>12</td>
<td>Rapamycin</td>
</tr>
<tr>
<td>13</td>
<td>SC-58125</td>
</tr>
<tr>
<td>14</td>
<td>Calmidazolium</td>
</tr>
<tr>
<td>15</td>
<td>15-β-Prostaglandin J2</td>
</tr>
<tr>
<td>16</td>
<td>Fluphenazine</td>
</tr>
<tr>
<td>17</td>
<td>Raloxifene</td>
</tr>
<tr>
<td>18</td>
<td>Monastrol</td>
</tr>
<tr>
<td>19</td>
<td>Trifluoperazine</td>
</tr>
<tr>
<td>20</td>
<td>Fulvestrant</td>
</tr>
<tr>
<td>21</td>
<td>Rofecoxib</td>
</tr>
<tr>
<td>22</td>
<td>Wortmannin</td>
</tr>
<tr>
<td>23</td>
<td>estradiol</td>
</tr>
<tr>
<td>24</td>
<td>celecoxib</td>
</tr>
<tr>
<td>25</td>
<td>Celecoxib</td>
</tr>
<tr>
<td>26</td>
<td>Tetraethylpentamine</td>
</tr>
<tr>
<td>27</td>
<td>Genistein</td>
</tr>
</tbody>
</table>

**NOTE:** Compounds have been filtered and then ranked based on sequential criteria as detailed in the text. Those compounds that appear in the validation set are indicated in bold.
IC50 compared with those harboring hypermethylation of MLH1 promoter when treated with rapamycin (18.3 versus 4.4 μmol/L; \( P = 0.0824 \)) and 1.5-fold difference (15.02 versus 10.37 μmol/L; \( P = 0.0385 \)) when treated with LY-294002 (Fig. 3C and D). Comparisons of the activity of these compounds between MSI-H and MSS showed no differences due to the outlier effect of HCT-116. After excluding this cell line from the analysis, significant differences between MSS and MSI-H were detected for both compounds (Supplementary Tables S10 and S11). However, note that cells harboring mutations in MSH2 (LoVo) and MSH6 (HCT-15 and HCT-8) were the most sensitive to both compounds (Supplementary Table S10). This fact underlines the possible application of this therapeutic strategy to Lynch syndrome cases.

Proliferation data at specific drug concentrations confirm the above results. Cell lines with hypermethylation of MLH1 promoter showed higher inhibition with rapamycin (\( P = 0.0534 \), two-sided t test) and LY-294002 (\( P = 0.015 \), two-sided t test) than stable ones (Supplementary Fig. S1). In addition, LY-294002-treated cells showed highly significant differences in growth between MSI-H and MSS cell lines (\( P = 0.0021 \), two-sided t test). No significant differences between groups were found for 17-AAG and trichostatin A (Supplementary Table S10). Finally, we have performed mutational studies on KRAS and BRAF to analyze the effect that both mutations could have on the sensitivity of a compound (Supplementary Table S9). Classification of cell lines based on KRAS mutational status showed no differences between groups.

**Fig. 3.** A and B, differential sensitivity of colorectal cancer cell lines to treatment with rapamycin and LY-294002 for 5 d. MLH1 promoter hypermethylated cell lines (SW-48 and RKO) and MSS cell lines (HT-29 and SW-480). Points, mean (\( n = 3 \) experiments); bars, SD. C and D, MSI-H confers sensitivity to rapamycin and low levels of sensitivity to LY-294002. Column 1, mean IC50 as a function of mismatch repair genes status (\( n = 3 \) experiments); bars, SE. IC50 for growth inhibition by rapamycin differed markedly in a manner that correlated with microsatellite status. MSS cells exhibited resistance to rapamycin at concentrations >15 μmol/L. Column 2, IC50 for growth inhibition by LY-294002 differed in a manner that correlated with microsatellite status. MSS cells exhibited resistance to LY-294002 at concentrations >15 μmol/L. Mean IC50 of MSS was statistically significantly different with respect to Meth-MSI cell lines (*, \( P = 0.0358 \)). Meth-MSI, MSI due to hypermethylation of MLH1, MSI-H, MSI-high; MSI-H*, MSI-high excluding HCT-116.
The cytotoxic effects of rapamycin and LY-294002 highlight the relevance of the pathway inhibited by both drugs in MSI-H tumors. Then, we examined the gene expression profile generated using the MECC samples by enrichment testing. We found indirect evidences suggesting the role of PI3K-AKT-mTOR pathway in MSI-H tumors represented by relevant genes integrated in the VEGF (PIK3R3 and MAPK1), phosphatidylinositol signaling system (PIP4K2A, INPP1, ITPK1, and ITPKA), and inositol phosphate metabolism (PLCB3). Moreover, mTOR, mitogen-activated protein kinase, type II diabetes mellitus, and insulin signaling system pathways yielded 3, 13, 4, and 6 genes up in MSI-H respectively, although these did not reach statistical significance. A complete list for enrichment testing is provided in Supplementary Tables S12 and S13.

Assessing phenotype reversibility under treatment with rapamycin using MSI markers. One of the advantages of using MSI colorectal cancer cell lines is that the genetic marker of the phenotype is easily evaluated to test whether these drugs are

**Fig. 4.** MSI in mismatch repair-deficient cell lines following treatment with rapamycin. For this specific purpose, we relied on four mononucleotide microsatellite markers (BAT25, BAT26, β-catenin, and transforming growth factor-β) because there is no corresponding normal tissue to measure MSI using dinucleotide markers in cell lines. Markers are presented as follows: column 1, BAT26 and β-catenin; column 2, transforming growth factor-β and BAT25. Gray columns, position of major allele after 5 d of treatment. No dose-dependent stabilization of microsatellite markers was observed in cells treated with 3 μmol/L rapamycin for 5 consecutive days.
inducing changes in expression of genes related to mismatch repair function or whether they are inducing a selective decrease of the targeted cell population. Therefore, we assessed the status of microsatellite markers in HCT-116, RKO, and SW-48 after treatment. As expected, treatment with rapamycin at concentrations around IC_{50} values for 5 days induced no changes in mononucleotide markers (Fig. 4), suggesting that the accumulated genetic changes of MSI cell lines (and cancers) with defective mismatch repair are not reversed when therapeutically targeted.

**Discussion**

The results of our study show that mismatch repair-deficient cell lines arising based on either somatic inactivation or germ-line mutations, and potentially their corresponding human...
tumors, respond better to therapies that preferentially target the PI3K-AKT-mTOR pathway. To derive this potentially therapeutic information, we have successfully combined gene expression data generated from different sources. Therefore, systems biology strategies such as the “Connectivity Map” can be used to identify optimal indications for testing drugs in vitro and potentially to accelerate the introduction of compounds into the clinic.

Deregulation of the PI3K-AKT signaling pathway has been previously implicated in the pathogenesis of colorectal cancer (33, 35). Somatic mutations in the gene encoding the p110α catalytic subunit (PI3KCA) have been described in 32% of colorectal tumors (36–38); also, mutations in the p85α regulatory subunit of PI3K has been reported (39). Interestingly, somatic mutations in PI3KCA tend to occur more frequently in mismatch repair-deficient tumors (37, 38, 40). Moreover, loss of PTEN tumor suppressor activity by hypermethylation of the promoter occurs almost exclusively in the MSI-H tumor subgroup (40). In addition, functional studies published previously have shown that mismatch repair-deficient colorectal cell lines harboring mutations in PI3KCA gene are more sensitive to PI3K inhibitors than the wild-type clones coming from the same cell lines (32).

We subsequently found another strong line of evidence supporting the efficacy of blockade of PI3K-AKT-mTOR in patients with deficient mismatch repair mechanism from the literature. A clinical case report of a patient with the Muir-Torre variant of Lynch syndrome harboring a germ-line mutation in MSH2 presented with a profound eruption of multiple MSI sebaceous adenomas after starting immunosuppressive therapy due to kidney transplantation (41, 42). The patient's sebaceous neoplasms improved dramatically when the immunosuppressive therapy was changed to rapamycin and recurred when treatment was stopped secondary to patient intolerance. Restoration of rapamycin treatment stopped the progression of cutaneous lesions. Although anecdotal, the obvious clinical activity of rapamycin in the Muir-Torre variant of Lynch syndrome provides a compelling clinical correlation with our bioinformatic and experimental conclusions. In fact, the cell line harboring a mutation in MSH2 (LoVo) and the two MSH6-mutant cells showed high responsiveness to rapamycin and LY-294002 in our experiments. This fact underlines the potential applicability of these therapies to hereditary nonpolyposis colorectal cancer patients. Clearly, more attention should be focused toward this group of patients as a model of study of therapeutic interventions in MSI-H cancers. Finally, we believe that our results showing the selective inhibition of the pathway by rapamycin and LY-294002 provide further evidence implicating the PI3K-AKT-mTOR pathway in MSI-H colorectal cancer. However, these compounds should be interpreted as models, so other rapalogs as well as other PI3K inhibitors with more favorable pharmacogenomic and pharmacokinetic characteristics should be explored in this context (35, 43).

Our in vitro studies failed to validate the activity of two of our candidate compounds (trichostatin A and 17-AAG). This observation could be explained by the presence of mutations in their target molecules or others related within the same pathway. As an example to illustrate this point, mutations in HDAC2 has been observed more frequently between MSI-H tumors and have been pointed as the cause for resistance to certain types of histone deacetylase inhibitors as trichostatin A but not to others as valproic acid and butyric acid (44). We found valproic acid and trichostatin A as potential compounds of interest in our discovery set, although these were not validated. Although we elected to pursue trichostatin A in our experimental models, we do not anticipate that cell lines would be strongly differentially sensitive because this compound was not confirmed in our validation set. The potential for agents interacting through epigenetic-related pathways is still attractive for sporadic MSI tumors. However, these limitations highlight the lack of perfect precision to find the most suitable agent for a subset of cancers based on a priori knowledge.

Microarray experiments provide an enormous amount of gene expression information that at first blush shows an apparent lack of overlap. In this study, we have shown that data sets from different sources characterizing the same biological state can be effectively combined to reveal a limited number of compounds that appear consistently across signatures. In our approach, we created artificial union/intersection gene lists using the original signatures to enhance the power of our classifying and ranking techniques in presence of limited data sources and to aggregate the gene-signature information. Although union-intersection of probe lists have been previously used for meta-analytic purposes with successful results (45), an obvious disadvantage of using these artificially created signatures is that they may introduce redundancy, sometimes amplifying the relevance of certain compounds and thus introducing bias. Moreover, these artificial signatures are not independent of the constituent original studies. In case one has at least three independent constituent studies in discovery and validation sets, we recommend applying our algorithm to only the original studies and omit the union/intersection signatures from the analytic process. However, in our current study, when we applied the algorithm after excluding synthetic signatures, the results were consistent in the discovery set but not in the validation set due to the fact that the validation set had only two gene expression data sets available (data not shown). With only two independent probe lists, our filtering algorithm requires the compounds to be consistently behaving in both the lists and thus making it harder to qualify. Another critical issue is the number of probes used to generate the compound lists. “Connectivity Map” will provide more accurate enrichment scores as more probes are introduced, making optimal cutoff values to include genes a very relevant question. We established a cutoff for probes coming from our own gene expression information based on P values and we used the probes included in the final classifiers as have been published originally by others. It is clear to us that this choice may decrease the homogeneity of probe sets selection across studies and that selection of probes may have used expression fold changes. However, our strategy was designed to ensure minimal data manipulation and to encourage researchers and clinical investigators to adopt this tool in drug development initiatives that take advantage of the information contained in large expression data sets.

Recently, several groups have developed computational strategies to discover new targeted therapies (46, 47). “Connectivity Map” presents several advantages over the existing approaches (31, 48). First, gene expression information is matched agnostically without focusing on a specific pathway. Second, off-target effects induced by drugs at specific concentrations can be discovered by matching all the information
provided by the signature in terms of expression. Third, a total number of 164 small molecules can currently be scanned at the same time. However, future editions of “Connectivity Map” are likely to include a broader selection of drugs (49), making a compelling argument for development of automated ranking and selection procedure of compounds. Fourth, the biological and biomedical value of the concept of reversing a gene expression signature through the treatment with a small molecule is not fully understood, we have proven that these connections can contribute to the identification of compounds than effectively target pathways that are essential for growth and survival in cancer cell models.

Our selection approach has been designed to minimize the intrinsic limitations of the “Connectivity Map” such as the fact that not all small molecules have been tested in every cell model, and not all were tested across the same spectrum of concentrations, thus raising the possibility that compounds might be prioritized not only based on connectivity score but also as a function of differing number of experimental instances. Therefore, to address this fact, we initially filtered out those compounds that have been tested less than two times based on the idea that, in the context of a relatively low number of experimental instances, it is easier to obtain consistent results having higher connectivity scores. Secondly, the use of several independent data sets combined with our two-step validation strategy has added robustness, consistency, and balance to our candidate compound list, diminishing the number of false positives. We are aware that direct construction of a weighted Kolmogorov-Smirnov scan statistic, with weights depending on positives. We are aware that direct construction of a weighted Kolmogorov-Smirnov scan statistic, with weights depending on positives. We are aware that direct construction of a weighted Kolmogorov-Smirnov scan statistic, with weights depending on positives.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Drs. María D. Iniesta and Gabriel Capella for critical reading of the article and Michelle Vinco for the excellent technical support.

Accession numbers. The raw array data as well as the data set with the statistical tests have been deposited in National Center for Biotechnology Information’s GEO and are accessible through GEO Series accession no. GSE15434.

References

8. Hampel H, Frankel WL, Martin E, et al. Screening for the Lynch syndrome (hereditary nonpolyposis colo-
9. Herman JG, Umar A, Polyak K, et al. Incidence and functional consequences of hMLH1 promoter hyper-
10. de la Chapelle A. Genetic predisposition to colorec-
12. Popat S, Hubner R, Houston RS. Systematic review of microsatellite instability and colorectal cancer pro-
13. Greenisen JG, Bonner JD, Ben-Yzhak O, et al. Pheno-
type of microsatellite unstable colorectal carcinomas: well-differentiated and focally mucinous tumors and the absence of the dinky neoscorre correlate with microsatel-
14. Carethers JM, Chauhan DP, Fink D, et al. Mismatch repair proficiency and in vitro response to 5-fluoroura-
15. Ribic CM, Sargent DJ, Moore M, et al. Tumormicro-
16. Fallik E, Bonini F, Boige V, et al. Microsatellite insta-
libility is a predictive factor of the tumor response to iri-
18. Vilar E, Scadutti M, Balmana J, et al. Microsatellite instability due to HMLH2 deficiency is associated with increased cytotoxicity to irinotecan in human colorec-
19. Lamb J, Crawford ED, Peck D, et al. The Connectiv-
22. Giordano TJ, Shedden KA, Schwartz DR, et al. Or-
gan-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene ex-
24. Rozek LS, Lipkin SM, Fearon ER, et al. CDX2 poly-
sion signatures for colorectal cancer microsatellite sta-
Gene Expression Patterns in Mismatch Repair-Deficient Colorectal Cancers Highlight the Potential Therapeutic Role of Inhibitors of the Phosphatidylinositol 3-Kinase-AKT-Mammalian Target of Rapamycin Pathway

Eduardo Vilar, Bhramar Mukherjee, Rork Kuick, et al.


Updated version
Access the most recent version of this article at:
http://clincancerres.aacrjournals.org/content/15/8/2829

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2011/07/19/1078-0432.CCR-08-2432.DC1

Cited articles
This article cites 48 articles, 18 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/15/8/2829.full.html#ref-list-1

Citing articles
This article has been cited by 8 HighWire-hosted articles. Access the articles at:
/content/15/8/2829.full.html#related-urls

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.