Published in final edited form as:

Clin Cancer Res. 2009 April 15; 15(8): 2829–2839. doi:10.1158/1078-0432.CCR-08-2432.

Gene Expression Patterns in Mismatch Repair-Deficient Colorectal Cancers Highlight the Potential Therapeutic Role of Inhibitors of the PI3K-AKT-mTOR pathway

Eduardo Vilar¹, Bhramar Mukherjee⁷, Rork Kuick⁶, Leon Raskin¹, David E. Misek⁵, Jeremy MG. Taylor⁷, Thomas J. Giordano^{1,4}, Samir M. Hanash⁸, Eric R. Fearon^{1,3,4}, Gad Rennert⁹, and Stephen B. Gruber^{1,2,3}

¹Department of Internal Medicine, The University of Michigan Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI 48109

²Department of Epidemiology, The University of Michigan Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI 48109

³Department of Human Genetics, The University of Michigan Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI 48109

⁴Department of Pathology, The University of Michigan Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI 48109

⁵Department of Surgery, The University of Michigan Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI 48109

⁶Department of Biostatistics Core, The University of Michigan Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI 48109

⁷Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109

⁸Molecular Diagnostics, Fred Hutchinson Cancer Research Center, Seattle, WA 98109

⁹Clalit Health Services, National Cancer Control Center and Department of Community Medicine and Epidemiology, Carmel Medical Center and B. Rappaport Faculty of Medicine, Technion, Haifa 31096, Israel

Abstract

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

Experimental Design—Three expression/query signatures (discovery data set) characterizing MSI-H CRC 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

Request for reprints: Stephen B. Gruber, Departments of Internal Medicine, Human Genetics and Epidemiology, University of Michigan, 1524 BSRB 109 Zina Pitcher, Ann Arbor, MI 48109. Phone: (734) 615-9712; Fax: (734) 647-7950; sgruber@umich.edu.

Accession Numbers

The raw array data as well as the data set with the statistical tests have been deposited in NCBI's GEO and are accessible through GEO Series accession number GSE11543.

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 cells lines using the bioinformatics approach; Rapamycin, LY-294002, 17-AAG 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 vs 4.4 μ M, $P=0.0824$) and LY-294002 (15.02 vs 10.37 μ M, $P=0.0385$) when compared to MSS cells. Preferential activity was also observed in *MSH2* and *MSH6*-mutant cells.

Conclusion—Our study demonstrates that the PI3K-AKT-mTOR pathway is of special relevance in mismatch repair-deficient CRC. 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.

Keywords

Microsatellite instability; colorectal cancer; gene expression patterns; rapamycin; mTOR pathway; PI3K inhibitors

Introduction

Colorectal cancer (CRC) is the third most common cancer in both men and women in western countries (1). While the management of patients with CRC 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 CRC 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 CRCs are less responsive to treatments targeted against the epidermal growth factor receptor (3–5). CRC can also be classified on the basis of the relative proportion of microsatellite loci that are found to demonstrate instability in cancer cells relative to the patient's germline configuration for the locus, and consensus criteria distinguish CRC showing high-frequency microsatellite instability (MSI-H), low frequency MSI (MSI-L), or microsatellite stability (MSS) (6). About 15% of apparently sporadic CRC are MSI-H (7, 8), mainly due to epigenetic silencing of the promoter of *MLH1* gene (9). A small fraction of the MSI-H cases, perhaps 2–3%, are due to germline mutations in one allele of one of the mismatch repair genes (e.g., *MLH1*, *MSH2*, *MSH6* and *PMS2*) (7, 8), which are the cause of Hereditary Non-Polyposis CRC (HNPCC or Lynch Syndrome) (10). MSI-H tumors are characterized by right-sided location, diagnosis at younger age, lower pathological stage, high histological 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 CRCs (14, 15). On the other hand, some preliminary data suggest that MSI-H CRCs 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 biologic 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 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 (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 analytical 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 biologic state (19, 20).

Here we describe our approach for combining results obtained through the “*Connectivity Map*” using five independent gene expression profiles from primary CRCs divided into discovery and validation data sets. Our analysis has identified that MSI-H CRC is specifically sensitive to compounds inhibiting the PI3K-AKT-mTOR pathway. In our in vitro experiments these compounds have demonstrated their preferential anti-proliferative and cytotoxic activities in MSI-H cell lines when compared to 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 previously been described (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 CRC 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) in order to assess the MSI status as previously described (13). Of these tumors, 38 were identified as MSS and 13 as MSI-H (Supplementary Table S1). As expected, the MSI-H group showed a statistically significant association with the presence of *BRAF* V600E mutation ($P=0.004$, χ^2 test) and right side of the colon ($P=0.006$, χ^2 test). Although tumor location may be related with gene expression (23), we have previously reported in a larger cohort of the MECC study that MSI status is more strongly associated with gene expression than location (24). Using 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 (Figure 1 and Supplementary Table S2), whereas in 1000 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 approximately 0.1%. Forty-one probe sets showed higher and thirty lower expressions in MSI-H compared with MSS cancers. Finally, enrichment testing of functional lists of genes was performed 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 homogenous data sets. Complete information corresponding to each data set was retrieved from the original publications (23, 25, 26) and are detailed in Supplementary Tables S16–19. Probe sets from each data set were then submitted to “*Connectivity Map*” in two independent lists, one for upregulated and one for the downregulated probes, containing the highly significantly expressed genes at levels reported in the original publications. Therefore, no statistical re-calculations on the original data have been performed.

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 (28, 29) (Supplementary Table S9). Cell lines were obtained from the American Type Culture Collection (Manassas, VA). 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 HNPCC. RKO and SW-48 are MSI-H lines due to hypermethylation in the promoter of *hMLH1* gene, similar to sporadic CRC displaying MSI-H. HT-29 and SW-480 are wild type for mismatch repair genes and are MSS. Cells were grown in DMEM/F12 medium supplemented with 10% of fetal bovine serum. All tissue culture reagents were obtained from GIBCO.

Treatment of cell lines with small-molecules—17-(Allylamino)-17-demethoxygeldanamycin (17-AAG), LY-294002, Rapamycin, and Trichostatin-A were obtained from Sigma-Aldrich (Saint Louis, MO). 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 in order to perform cytotoxicity experiments.

Cytotoxicity Experiments—Cytotoxicity assays were performed in 96-well plates and cell proliferation was assessed using the WST-1 (Roche Diagnostics GmbH, Germany) 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 μ M of Rapamycin during five 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 the Supplementary Methods.

Results

Gene expression data defining MSI colorectal tumors show an apparent lack of overlap

As shown in Figure 2, we separated the signatures in two groups: (i) a discovery set including two previously defined signatures (Watanabe et al and Koinuna et al) and the MECC signature we report above, which we used to search for the candidate compounds (23, 25); and (ii) a validation set, that 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 list 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). In order 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 (Figure 2). As is shown in Supplementary Tables S3 and S4, this signature was composed of a total of six up regulated and four downregulated genes in the discovery set, and three upregulated and four downregulated 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.

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 set, respectively (Figure 2). In order 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 which 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 are exhibiting 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: (i) the total number of negative connectivity scores, (ii) percentage of the difference between negative and positive connectivity scores $[(\text{No. negative connectivity scores} - \text{No. positive connectivity scores}) / (\text{No. negative connectivity scores} + \text{No. positive connectivity scores})]$, (iii) the mean connectivity score across signatures; and finally (iv) an 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*” (namely, $-\Sigma \log(P\text{-value})$, 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. Out 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 is shown in Table 1, a direct comparison of the final lists of the discovery and validation set revealed that 14 out of 27 compounds were replicated. Also, six compounds appeared consistently within the fifteen 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 which are recognized demethylators or that act through the inhibition of related molecules as HDACs (Trichostatin-A and Valproic Acid). Others interact with heat shock proteins (17-AAG, 17-dimethylaminoethylamino-demethoxygeldanamycin (17-DAG), 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 CRC cell lines. Note that 17-AAG, Rapamycin and LY-294002 appeared to be highly rated both in the discovery and in 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. So, 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 explore if 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 Figures 3A and 3B 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 IC50 compared to those harboring hypermethylation of *MLH1* promoter when treated with Rapamycin (18.3 vs 4.4 μM , $P=0.0824$) and 1.5-fold difference (15.02 vs 10.37 μM , $P=0.0385$) when treated with LY-294002 (Figures 3C and 3D). 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 Table 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 (two-sided *t*-test, $P=0.0534$) and LY-294002 (two-sided *t*-test, $P=0.015$) than stable ones (Supplementary Figure S1). In addition, LY-294002 treated-cells showed highly significant differences in growth between MSI-H and MSS cell lines (two-sided *t*-test, $P=0.0021$). 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 impact that both mutations could have on a compound's sensitivity (Supplementary Table S9). Classification of cell lines based on *KRAS* mutational status showed no differences between groups.

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*, *MAPK1*), Phosphatidylinositol signaling system (*PIP4K2A*, *INPP1*, *ITPK1*, *ITPKA*) and Inositol phosphate metabolism (*PLCB3*). Moreover mTOR, MAPK, 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 Information Table S12 and S13.

Assessing phenotype reversibility under treatment with Rapamycin using microsatellite instability markers

One of the advantages of using MSI CRC cell lines is that the genetic marker of the phenotype is easily evaluated to test whether these drugs are 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 IC50 values for five days induced no changes in mononucleotide markers (Figure 4), suggesting that the accumulated genetic changes of MSI cell lines (and cancers) with defective mismatch repair are not reversed when therapeutically targeted.

CRC cells lines under treatment with Rapamycin are arrested in G1 phase

We then performed apoptosis in cells exposed to Rapamycin. Staining revealed no changes with Rapamycin in any of the cell lines assessed (Figure 5A). Likewise quantification of apoptosis through sub-G1 did not render significant differences. Therefore, this fact is consistent with the reported cytostatic effect of Rapamycin. When we compared the percentage of cells in G1 phase after treatment with the control, we observed an increased in the G1 peak in MSI-H cell lines that was consistent with our expectations and previous studies (Figures 5B and Supplementary Figure S2).

Discussion

The results of our study show that mismatch repair-deficient cell lines arising on the basis of either somatic inactivation or germline mutations, and potentially their corresponding human tumors, respond better to therapies that preferentially target the PI3K-AKT-mTOR pathway. In order 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 PI3K-AKT signaling pathway has been previously implicated in the pathogenesis of CRC (33, 35). Somatic mutations in the gene encoding the p110 α catalytic subunit (*PI3KCA*) have been described in 32% of colorectal tumors (36–38) and also mutations in the p85 α 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 CRC 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 germline 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 HNPCC patients. Clearly, more attention should be focused towards 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 CRC. 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 HDAC 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 since 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 apriori 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 demonstrated that data sets from different sources characterizing the same biologic 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 in order to aggregate the gene-signature information. Though 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 set, we recommend applying our algorithm to only the original studies and omit the union/intersection signatures from the analytical 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 cut-off values to include genes a very relevant question. We established a cut-off 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, although 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 proved 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 on the basis of connectivity score but also as a function of differing number of experimental instances. Therefore, in order 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 the sample size, signature platform and strength of the different studies may be an attractive

alternative for the combination of signatures (50), but it is also true that a weighted analysis may introduce other types of noise and bias.

MSI is the molecular fingerprint that characterizes a well-known subset of colorectal tumors that show specific molecular and clinico-pathological characteristics (23, 25–27). MSI can be easily detected through comparison of DNA from the tumor and normal mucosa allowing its use in the routine clinical setting. MSI analysis of tumor biopsies is used for the selection of those patients candidate for germline analysis in MMR genes. In addition, there is compelling evidence showing that colorectal MSI tumors associate with a better outcome. Although it is under debate at the present time, MSI may play a role as a predictive factor of response to 5-Fluorouracil and Irinotecan-based chemotherapy. Our results point towards the therapeutic importance of the PI3K-AKT-mTOR pathway in MSI-H and open the door for the evaluation of anti-mTOR and anti-PI3K therapies in this recognized colorectal tumor subgroup.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant support: E. Vilar was supported by a fellowship from “la Caixa”, Barcelona, Spain. This work was supported in part by a NCI grant 1R01CA81488; by a grant from the NCI Director’s Challenge Program at the University of Michigan (NIH CA84952); by a grant from the University of Michigan Comprehensive Cancer Center Bioinformatics Core (NIH 5P30CA46592); by gifts from the Irving Weinstein Foundation and the Ravitz Foundation.

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

References

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2008. *CA Cancer J Clin.* 2008; 58:71–96. [PubMed: 18287387]
2. SEER. US Mortality Age-Adjusted Rates, Total US 1990–1999: US National Center for Health Statistics. 2002.
3. Amado RG, Wolf M, Peeters M, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol.* 2008; 26:1626–34. [PubMed: 18316791]
4. Khambata-Ford S, Garrett CR, Meropol NJ, et al. Expression of epiregulin and amphiregulin and K-ras mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab. *J Clin Oncol.* 2007; 25:3230–7. [PubMed: 17664471]
5. Lievre A, Bachet JB, Boige V, et al. KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. *J Clin Oncol.* 2008; 26:374–9. [PubMed: 18202412]
6. Boland CR, Thibodeau SN, Hamilton SR, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* 1998; 58:5248–57. [PubMed: 9823339]
7. Aaltonen LA, Salovaara R, Kristo P, et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N Engl J Med.* 1998; 338:1481–7. [PubMed: 9593786]
8. Hampel H, Frankel WL, Martin E, et al. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med.* 2005; 352:1851–60. [PubMed: 15872200]

9. Herman JG, Umar A, Polyak K, et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci U S A*. 1998; 95:6870–5. [PubMed: 9618505]
10. de la Chapelle A. Genetic predisposition to colorectal cancer. *Nat Rev Cancer*. 2004; 4:769–80. [PubMed: 15510158]
11. Gryfe R, Kim H, Hsieh ET, et al. Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. *N Engl J Med*. 2000; 342:69–77. [PubMed: 10631274]
12. Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol*. 2005; 23:609–18. [PubMed: 15659508]
13. Greenson JK, Bonner JD, Ben-Yzhak O, et al. Phenotype of microsatellite unstable colorectal carcinomas: Well-differentiated and focally mucinous tumors and the absence of dirty necrosis correlate with microsatellite instability. *Am J Surg Pathol*. 2003; 27:563–70. [PubMed: 12717242]
14. Carethers JM, Chauhan DP, Fink D, et al. Mismatch repair proficiency and in vitro response to 5-fluorouracil. *Gastroenterology*. 1999; 117:123–31. [PubMed: 10381918]
15. Ribic CM, Sargent DJ, Moore M, et al. Tumor Microsatellite-Instability Status as a Predictor of Benefit from Fluorouracil-Based Adjuvant Chemotherapy for Colon Cancer. *N Engl J Med*. 2003; 349:247–57. [PubMed: 12867608]
16. Fallik D, Borrini F, Boige V, et al. Microsatellite instability is a predictive factor of the tumor response to irinotecan in patients with advanced colorectal cancer. *Cancer Res*. 2003; 63:5738–44. [PubMed: 14522894]
17. Jacob S, Aguado M, Fallik D, Praz F. The role of the DNA mismatch repair system in the cytotoxicity of the topoisomerase inhibitors camptothecin and etoposide to human colorectal cancer cells. *Cancer Res*. 2001; 61:6555–62. [PubMed: 11522654]
18. Vilar E, Scaltriti M, Balmana J, et al. Microsatellite instability due to hMLH1 deficiency is associated with increased cytotoxicity to irinotecan in human colorectal cancer cell lines. *Br J Cancer*. 2008; 99:1607–12. [PubMed: 18941461]
19. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006; 313:1929–35. [PubMed: 17008526]
20. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102:15545–50. [PubMed: 16199517]
21. Poynter JN, Gruber SB, Higgins PD, et al. Statins and the risk of colorectal cancer. *N Engl J Med*. 2005; 352:2184–92. [PubMed: 15917383]
22. Giordano TJ, Shedden KA, Schwartz DR, et al. Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. *Am J Pathol*. 2001; 159:1231–8. [PubMed: 11583950]
23. Watanabe T, Kobunai T, Toda E, et al. Distal colorectal cancers with microsatellite instability (MSI) display distinct gene expression profiles that are different from proximal MSI cancers. *Cancer Res*. 2006; 66:9804–8. [PubMed: 17047040]
24. Rozek LS, Lipkin SM, Fearon ER, et al. CDX2 polymorphisms, RNA expression, and risk of colorectal cancer. *Cancer Res*. 2005; 65:5488–92. [PubMed: 15994917]
25. Koinuma K, Yamashita Y, Liu W, et al. Epigenetic silencing of AXIN2 in colorectal carcinoma with microsatellite instability. *Oncogene*. 2006; 25:139–46. [PubMed: 16247484]
26. Banerjee A, Ahmed S, Hands RE, et al. Colorectal cancers with microsatellite instability display mRNA expression signatures characteristic of increased immunogenicity. *Mol Cancer*. 2004; 3:21. [PubMed: 15298707]
27. Kruhoffer M, Jensen JL, Laiho P, et al. Gene expression signatures for colorectal cancer microsatellite status and HNPCC. *Br J Cancer*. 2005; 92:2240–8. [PubMed: 15956967]
28. Suter CM, Norrie M, Ku SL, Cheong KF, Tomlinson I, Ward RL. CpG island methylation is a common finding in colorectal cancer cell lines. *Br J Cancer*. 2003; 88:413–9. [PubMed: 12569385]
29. Li HR, Shagisultanova EI, Yamashita K, Piao Z, Perucho M, Malkhosyan SR. Hypersensitivity of tumor cell lines with microsatellite instability to DNA double strand break producing chemotherapeutic agent bleomycin. *Cancer Res*. 2004; 64:4760–7. [PubMed: 15256444]

30. Fan C, Oh DS, Wessels L, et al. Concordance among Gene-Expression-Based Predictors for Breast Cancer. *N Engl J Med*. 2006; 355:560–9. [PubMed: 16899776]
31. Hieronymus H, Lamb J, Ross KN, et al. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell*. 2006; 10:321–30. [PubMed: 17010675]
32. Samuels Y, Diaz LA Jr, Schmidt-Kittler O, et al. Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell*. 2005; 7:561–73. [PubMed: 15950905]
33. Vivanco I, Sawyers CL. The phosphatidylinositol 3-Kinase AKT pathway in human cancer. *Nat Rev Cancer*. 2002; 2:489–501. [PubMed: 12094235]
34. Minucci S, Pelicci PG. Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nat Rev Cancer*. 2006; 6:38–51. [PubMed: 16397526]
35. Hennessy BT, Smith DL, Ram PT, Lu Y, Mills GB. Exploiting the PI3K/AKT pathway for cancer drug discovery. *Nat Rev Drug Discov*. 2005; 4:988–1004. [PubMed: 16341064]
36. Velho S, Oliveira C, Ferreira A, et al. The prevalence of PIK3CA mutations in gastric and colon cancer. *Eur J Cancer*. 2005; 41:1649–54. [PubMed: 15994075]
37. Samuels Y, Wang Z, Bardelli A, et al. High frequency of mutations of the PIK3CA gene in human cancers. *Science*. 2004; 304:554. [PubMed: 15016963]
38. Abubaker J, Bavi P, Al-Harbi S, et al. Clinicopathological analysis of colorectal cancers with PIK3CA mutations in Middle Eastern population. *Oncogene*. 2008; 27:3539–45. [PubMed: 18193083]
39. Philp AJ, Campbell IG, Leet C, et al. The phosphatidylinositol 3'-kinase p85alpha gene is an oncogene in human ovarian and colon tumors. *Cancer Res*. 2001; 61:7426–9. [PubMed: 11606375]
40. Goel A, Arnold CN, Niedzwiecki D, et al. Frequent inactivation of PTEN by promoter hypermethylation in microsatellite instability-high sporadic colorectal cancers. *Cancer Res*. 2004; 64:3014–21. [PubMed: 15126336]
41. Levi Z, Hazazi R, Kedar-Barnes I, et al. Switching from tacrolimus to sirolimus halts the appearance of new sebaceous neoplasms in Muir-Torre syndrome. *Am J Transplant*. 2007; 7:476–9. [PubMed: 17229076]
42. Harwood CA, Swale VJ, Bataille VA, et al. An association between sebaceous carcinoma and microsatellite instability in immunosuppressed organ transplant recipients. *J Invest Dermatol*. 2001; 116:246–53. [PubMed: 11180000]
43. Easton JB, Houghton PJ. mTOR and cancer therapy. *Oncogene*. 2006; 25:6436–46. [PubMed: 17041628]
44. Ropero S, Fraga MF, Ballestar E, et al. A truncating mutation of HDAC2 in human cancers confers resistance to histone deacetylase inhibition. *Nat Genet*. 2006; 38:566–9. [PubMed: 16642021]
45. Varambally S, Yu J, Laxman B, et al. Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell*. 2005; 8:393–406. [PubMed: 16286247]
46. Dressman HK, Berchuck A, Chan G, et al. An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. *J Clin Oncol*. 2007; 25:517–25. [PubMed: 17290060]
47. Wong DJ, Nuyten DS, Regev A, et al. Revealing targeted therapy for human cancer by gene module maps. *Cancer Res*. 2008; 68:369–78. [PubMed: 18199530]
48. Stegmaier K, Wong JS, Ross KN, et al. Signature-based small molecule screening identifies cytosine arabinoside as an EWS/FLI modulator in Ewing sarcoma. *PLoS Med*. 2007; 4:e122. [PubMed: 17425403]
49. Lamb J. The Connectivity Map: a new tool for biomedical research. *Nat Rev Cancer*. 2007; 7:54–60. [PubMed: 17186018]
50. Folks, JL. Combination of Independent Tests. In: Krishnaiah, PR.; Sen, PK., editors. *Handbook of statistics*. New York, N.Y., U.S.A: Elsevier Science Pub. Co; 1984. p. 113-21.

Statement of Translational Relevance

Therapeutic decisions for colorectal cancer (CRC) currently do not emphasize molecular subclassification. In addition, drugs selected for further development rest heavily on the activities observed in pre-clinical 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) CRC using a systems biology tool to combine gene expression data coming from different sources. From 164 candidate compounds, Rapamycin and LY-294002 demonstrated preferential selectivity *in vitro* to cell lines displaying high-frequency MSI. Statistical methods developed for this analysis demonstrate that an amalgamation of gene expression information across studies provides a robust approach for selection of potential therapies. In addition, these results point towards the therapeutic importance of the PI3K-AKT-mTOR pathway in MSI and open the door for the eventual evaluation of anti-mTOR and anti-PI3K therapies in this recognized subset of colorectal cancers.

Figure 1. Differentially expressed genes in MSI-H and MSS tumors from the MECC study. Tumors exhibiting the MSI-L phenotype were excluded. Each column corresponds to a separate sample and they are grouped by microsatellite status. Each row represents a probe set with P -value < 0.00001 ($FDR < 0.1\%$) whose expression is color-coded according to the indicated scale and are displayed by alphabetical order. Gene symbols of HuFL probe sets and its corresponding P -values are shown at the 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 (*PRKAR2B*) showed overexpression in MSI-H cancers. On the other hand, Protein Kinase C (*PRKCI*), mutL Homolog 1 (*MLH1*) and Transforming Growth Factor β Receptor II (*TGFBR2*) showed lower expression in MSI-H cancers. A complete list with HuFL probe set ID, gene symbol and gene title is presented in Supplementary Table S2.

Figure 2.

Gene expression data sets used and the flowchart for generation of a final compounds list. Gene expression data coming from 5 different studies were divided in two sets, discovery and validation set. 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 upregulated and downregulated 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.

Figure 3.

(A, B) Differential sensitivity of CRC cell lines to treatment with Rapamycin and LY-294002 for 5 days. Shown are the *MLH1* promoter hypermethylated cell lines (SW-48 and RKO) and MSS cell lines (HT-29 and SW-480). Points, mean ($n=3$ experiments); bars, $\pm$ standard deviation (SD) (C, D) MSI-H confers sensitivity to Rapamycin and low levels of sensitivity to LY-294002. *First column*, mean IC50 values as a function of mismatch repair genes status ($n=3$ experiments); bars, $\pm$ standard error (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 higher than 15 μ M. *Second column*, 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 higher than 15 μ M. Mean IC50 of MSS were statistically significantly different with respect to Meth-MSI cell lines (* $P=0.0358$) Meth-MSI, microsatellite-instability due to

hypermethylation of *MLH1*; MSI-H, microsatellite instability-high; MSI-H*, microsatellite instability-high excluding HCT-116.

HCT-116 CONTROL **HCT-116 RAPAMYCIN [3μM]** **RKO CONTROL** **RKO RAPAMYCIN [3μM]** **SW-48 CONTROL** **SW-48 RAPAMYCIN [3μM]** **Figure 4.**

Microsatellite instability 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 TGF- β) since there is no corresponding normal tissue to measure MSI using dinucleotide markers in cell lines. Markers are presented as follow: first column, BAT26 and β -Catenin; second column, TGF- β and BAT25. Grey bars represent the position of mayor allele after five days of treatment. No dose-dependent stabilization of microsatellite markers was observed in cells treated with Rapamycin 3 μ M during five consecutive days.

Figure 5.

(A) Effects of Rapamycin on apoptosis in SW-48 (first row), RKO (second row) and HT-29 (third row). AnnexinV/PI cytograms of cell lines following treatment with Rapamycin 3 μ M during five consecutive days. Control cells were grown in free-drug medium. X axis, AnnexinV; y axis, PI staining. Annexin V–positive, PI-negative cells reflect cells in the early stages of apoptosis (lower right quadrant), whereas Annexin V–positive, PI-positive cells reflect dead cells or cells at the late stages of apoptosis (upper right quadrant) in at least three independent experiments. Mean of cell percentages $\pm$ SD at late apoptosis are quantified in upper right quadrants. Untreated cells showed 81–88% viability (first column); Rapamycin did not induce apoptosis, consistent with its previously described cytostatic effect (second column).

(B) Effects of Rapamycin on cell cycle using PI staining. DNA histograms of SW-48 (first row), RKO (second row) and HT-29 (third row) CRC cell lines following treatment with Rapamycin 3 μ M for five days. Control cells were grown in free-drug medium. No changes in apoptosis were observed by sub-G1 results. G1 arrest under treatment with Rapamycin was observed in MSI-H cells being statistically significant in RKO ($P=0.006$) and no effect on the MSS HT-29 cell line (Figures 5B and Supplementary Figure S2). Results are reported as the mean $\pm$ SD done in at least two replicate samples. Vilar et al - Table 1

List of 27 ranked compounds from discovery set. Compounds have been filtered and then ranked based on sequential criteria as detailed in the text. In bold are those compounds that appear in the validation set.

Table 1

Rank	Compound
1	LY-294002
2	17AAG
3	Trichostatin A
4	17-DAG
5	5224221
6	Geldanamycin
7	Trifluoperazine
8	Monorden
9	Resveratrol
10	Prazosin
11	Rapamycin
12	Haloperidol
13	SC-58125
14	Calmidazolium
15	Valproic Acid
16	15-delta-Prostaglandin J2
17	Fluphenazine
18	Raloxifene
19	Monastrol
20	Tretinoin
21	Fulvestrant
22	Rofecoxib
23	Wortmannin
24	alpha-Estradiol
25	Celecoxib
26	Tetraethylenepentamine
27	Genistein