molecular systems biology

A pharmacogenomic method for individualized prediction of drug sensitivity

Adam Cohen, Raffaella Soldi, Haiyu Zhang, Adam Gustafson, Ryan Wilcox, Bryan Welm, Jeffrey Chang, Evan Johnson, Avrum Spira, Stefanie Jeffrey, Andrea Bild

Corresponding authors: Stefanie Jeffrey, Stanford University School of Medicine and Andrea Bild, University of Utah

Review timeline:

Submission date:	13 January 2011
Editorial Decision:	09 March 2011
Revision received:	20 April 2011
Editorial Decision:	24 May 2011
Revision received:	01 June 2011
Accepted	06 June 2011

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 09 March 2011

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the three referees who agreed to evaluate your manuscript. As you will see from the reports below, the referees find the topic of your study of interest. They raise, however, a series of substantial concerns on your work, which, I am afraid to say, preclude its publication in its present form.

In general, the authors recognized the potential value of this work and appreciated the effort made to experimentally validate your predictions. Nonetheless, they had a series of substantial concerns that they felt must be addressed before this work would be appropriate for publication. The reviewers have provided detailed and constructive suggestions for improvement; here, I highlight a few of the most important points:

1. The reviewers had clear concerns that some of the experimental results, particularly those in Fig. 5B and 6, remained rather inconclusive considering the small number of samples (especially samples predicted to be VPA resistant). All three reviewers indicated clearly that the correlation presented in Fig. 5B is currently unconvincing, and, as such, additional samples with low predicted VPA sensitivity should be tested.

2. Conceptually, reviewer #2 cautioned strongly against overstating the potential clinical relevance of these results, noting that MATCH predicted tissue sensitivity to VPA did not seem to coincide

with toxicity observed in clinical trials, and that the concentrations of VPA used here often exceeded clinically realistic levels.

3. The reviewers felt that the mechanistic connection between the VPA gene expression signature and tumor proliferation & survival remained largely unclear, and indicated that this issue should be explored and discussed in more detail. See in particular the "black box" comment by reviewer #3.

On a more subjective level, the editor feels that this work may benefit from a title that better describes the key findings of this work. Perhaps, something such as "Pharmacogenomic prediction of tumor-specific sensitivity to valproic acid." It should be sufficient to describe the MATCH acronym in the abstract.

If you feel you can satisfactorily deal with these points and those listed by the referees, you may wish to submit a revised version of your manuscript. Please attach a covering letter giving details of the way in which you have handled each of the points raised by the referees. A revised manuscript will be once again subject to review and you probably understand that we can give you no guarantee at this stage that the eventual outcome will be favorable.

PLEASE NOTE As part of the EMBO Publications transparent editorial process initiative (see http://www.nature.com/msb/journal/v6/n1/full/msb201072.html), Molecular Systems Biology now publishes online a Review Process File with each accepted manuscript. Please be aware that in the event of acceptance, your cover letter/point-by-point document will be included as part of this file, which will be available to the scientific community. More information about this initiative is available in our Instructions to Authors. If you have any questions about this initiative, please contact the editorial office (msb@embo.org).

Yours sincerely,

Editor

Molecular Systems Biology

Referee reports:

Reviewer #1 (Remarks to the Author):

I read the manuscript titled, "MATCH: Merging genomic and pharmacologic Analyses for Therapy CHoice" by Cohen et al. with great interest. The research describes a novel and important method of using publicly available molecular measurements of drug effect and cancer tumor expression to make inferences about therapeutic effect for individual tumors. The paper is generally well-written and the topic timely and relevant to the readership of Molecular Systems Biology. The authors should also be commended for performing the necessary experimental validations to support their computational predictions. That being said, I do feel that some aspects of the paper could be improved by clarification. My major concerns are the following:

1. It's not clear why the authors chose the binary regression based approach to generate the drug signature. Perhaps there is precedence from previous research by this group or others, but I can't

seem to find explanation of a principled choice of this method (e.g., performed better in cross-validation that SAMR, RANKPROD, XYZ). Just some clarification on why this method was best for generating the signature would be great.

2. The use of a second VPA treated experiment for validation of the signature is nice, but it's still hard to assess the quality signature model from LOOCV and the prediction on the second set. I'm still not convinced that it isn't overfitted. One thing that strikes me about this paper overall, is that the authors seem to ignore the power of the large data set of data found in connectivity map. It would be nice to see what you could predict by random chance against all the treated vs. untreated samples in connectivity map for example. Right now the drug side doesn't feel as systematic as the disease side. It would be great to see some kind of true positive vs. false positive tradeoff (ROC curve) across many samples to assess accuracy.

3. Following #2, is there a principled, or data-driven way to justify the choice of VPA? The provide a nice clinical justification for choosing VPA, but for the in silico part of the study, it would be interesting to see where VPA falls in relation to all the other anti-neoplastic drugs in connectivity map.

4. Figure 1 is a bit confusing. It's nice to see the steps, but the descriptions are a bit vague. It's hard to see how information is flowing through this decision process and what is flowing from one step to the next. I would suggest a rework of this figure. It's not immediately clear what is going on.

5. The 3-D cell culture is a really nice idea, but the regression line seems to be strongly influenced by one sample near the bottom right. It would be nice to see more samples in the low prediction range to have more belief in this trend. Also, a positive control like is done for the xenografts would be an added bonus.

6. One final thing that struck me about this paper was that there seemed to be no effort to functionally characterize the drug signature. The list of genes provided is nice, but I'm left a bit wanting to see which genes, pathways, functional groups, etc might be "explaining" some of the predicted and validated therapeutic disparities. Like the HER2 story the authors reference, it would be nice to get some insights about potential therapeutic development opportunities if any apparent functional group or pathway emerges. Even if it's not the main purpose of this paper, the results should be useful for this purpose.

Minor comments:

- The text under the "Generation of drug response signature" heading is a bit complex and jargony. It would be nice to walk the reader through this critical step of the research more clearly.

- Figure 1: I think the authors might mean to say "dysregulated" instead of "deregulated"

Reviewer #2 (Remarks to the Author):

Overall, I enjoyed reading this manuscript and it contained a lot of good ideas. I could not agree more strongly with the first two paragraphs of the Discussion. These clearly and succinctly state the problems with clinical trials, heterogeneity within patient populations, and the inability of current methods to address these issues. The results of the manuscript indicate a promising direction towards solving these problems, as well as providing promising preliminary results. I do, however, have a few problems with how the data is presented, interpreted, and discussed. I believe that this manuscript presents significant findings, but worry that some of the findings may be overstated.

The major concern is whether this genomic predictor really has promise to predict responses to VPA in the clinical arena and here I have major concerns.

First, the authors use clinically unachievable levels of VPA to validate their results. The free levels of VPA obtained in plasma range between 0.25 and 1.0 mM (Munster, JCO, 2007) which is far below what the authors are using in their validation experiments (where most of the EC50 values are in the 5 to 25 mM range). It is unclear if the authors hypothesize that VPA acts as an HDAC inhibitor for its primary mechanism and no data is shown indicating at which concentrations VPA inhibits HDAC activity. Furthermore, should other HDAC inhibitors yield similar predictions?

Second, MATCH predicts sensitivity of normal liver cells/tissues to VPA and the authors themselves acknowledge this as being predicted correctly based on the literature; however, the same Munster trial of VPA in cancer patients demonstrate no significant liver toxicity. The author's data would also suggest VPA causes esophageal toxicity which is not borne out in the clinic. This raises the concern that the cell based approach overstates predictive capacity in humans. It is also unclear why the authors excluded sarcomas and leukemia from their analysis; especially as leukemia is a potentially more sensitive tumor type than solid tumors.

Third, what is the author's definition of 'response', such as 'drug response'? Do they mean tumor regression (which matches the clinician's definition of response) or lump together tumor regression and stasis or mean tumors that have changes in gene expression from compounds irrespective of effects on tumor growth? Most of the clinical data to date suggests little tumor regressions (responses) to HDAC inhibitors in solid tumors including breast and lung cancers, two tumor types predicted to be enriched for sensitive tumors using MATCH.

Fourth, it seems this MATCH algorithm is best at defining genes that are expressed and modulated by VPA, but changes in these genes do not necessarily mechanistically explain, correlate with, or predict tumor cell death or stasis secondary to compounds. MATCH could similarly detect genes altered by other compounds and enrich for tumors that have differential expression of these genes. However, the mechanistic link between changes in gene expression and changes in tumor proliferation/survival may not exist. Thus, MATCH could not predict cytotoxicity or cytostasis of these compounds if the genes are not casually associated with cell death/proliferation.

An ovarian normal cell line is used for validation of VPA treatment (Figure 2C). But your other data suggests that ovarian cancer cell lines are resistant to VPA and there is no data on the normals to suggest either way. Although I'm sure that the use of ovarian normal was prompted by availability, this doesn't seem to be the ideal system for validation. Why are the ovarian cancer cell lines used for validation (ie. VPA treated) more sensitive to VPA? From the figure of predictive VPA it appears that ovarian cancer is ~0.10 or so for sensitivity. Yet the graph for GSE1615 shows 0.8 sensitivity. And when treated, the sensitivity goes toward the cancer range of 0.1. Perhaps this indicates acquired resistance to VPA but this isn't clear in the text. And this doesn't seem to validate much, however.

Why did the authors choose EC50 rather than IC50? For some cell lines, EC50 can be much lower than IC50 and thus falsely associate sensitivity of a cell line to treatment; for example, a tumor cell shows minimal growth arrest (magnitude of effect) but this change occurs quite low in the concentration curve and thus the EC50 is low but the IC50 is high.

The authors perform one experiment comparing VPA to doxorubicin and conclude that "VPA decreased tumor growth significantly more than doxorubicin". This is a dangerous overstatement as doxorubicin has been a highly effective therapy for breast cancer while the efficacy of VPA is still not clear. One mouse experiment is hardly strong data to support this concept.

For in vivo validation, 3 VPA-sensitive tumors (MATCH prediction) are examined while only 1 VPA-insensitive tumors (MATCH prediction) are examined. Ideally a few more resistant tumors would be examined.

R2 should be provided for all figures with best fit lines. Several of the fits are visually poor, with marginal p-values, and reporting R2 would make the quality of the fits more apparent. Similarly, the error bars on the predicted sensitivities are very large. In the case of breast tumor, the errors bars cover the entire range. This doesn't appear to be very specific.

Page 13, paragraph 1, sentence 3: interpretation of Fig 5B. This is a very poor fit, whose slope is defined largely by the right-most blue data point, at a clinically irrelevant concentration of VPA. If the right-most point were to be removed, the slope through the remaining five points would be near vertical in the opposite direction as the fit slope. There is not a significant correlation here, as the significance test makes distribution assumptions that do not hold in this case. From this figure, it can not be claimed that there is a significant correlation between VPA and EC50.

From the methods section, it appears that the binary regression is normalized to 0/1 scale across datasets. What if the proportion of responses is different across different datasets? If one is all resistant, then some of the resistant points will be scaled to sensitive?

Corrections:

Page 6, paragraph 2, sentence 2: "Therapy Choice" should be "Therapy CHoice" to match acronym capitalization elsewhere in the manuscript.

Page 13, pargraph 3, sentence 2: "No response to VPA was seen in the tumor predicted to be resistant (Figure 6B, p=0.86)". Fig 6B indicates p=0.96 for SUTI103. Either this sentence or Fig 6B needs to be altered to present the correct value.

Page 16, paragraph 2 (Methods), sentence 4: "Each datasets was" should be "Each dataset was"

Page 17, paragraph 2, sentence 1: "gse5364" should be capitalized to "GSE5364"

Suggestions:

Page 11, paragraph 2: mentions that Fig 3C is sorted by subtype. Looking at figure 3C, it can be difficult to see the general trend of sensitivity within each subtype, since there is no apparent order within each subtype and the colors appear randomly distributed. I would suggest sorting by sensitivity within each subtype. This may make the general trend of sensitivity within subtype more visually evident.

Minor suggestions:

Page 10, paragraph 1, sentence 4: "although other datasets were compared and similar results found". Either list the datasets, or indicate "(data not shown)".

Reviewer #3 (Remarks to the Author):

Here, Cohen et. al. propose a method for training a drug sensitivity predictor and apply the results to a case study of valproic acid (VPA) and its role as a cancer therapeutic. The topic of the study is relevant to the field of systems biology and is appropriate for this journal. Computational guidance for therapy choice is becoming an increasingly important facet of personalized medicine; thus, this study and its conclusions are potentially important to the field of cancer pharmacogenomics and ongoing clinical trials of VPA.

Major concerns:

ï The VPA signature is treated as a "black box" by the authors. This can be acceptable and effective, but the precise mechanism of VPA action remains an issue of fundamental importance. It would be useful for the authors to mention literature and mechanisms for VPA action that explain the variability in its effectiveness according to different patients and tissue types.

ï VPA is undergoing clinical trials for several cancers - do those tumor types agree with your assessment of sensitivity based on tumor type?

ï Is there a set of matched tumor and normal tissues from the same patients? Can you predict differential sensitivity in a single patient independent of tumor type? This would be more revealing than making assumptions about the underlying tumor histology.

ï The authors assess the sensitivity of the drugs across tumor types, and generally observe a wide range of sensitivities among the different tumors within each type. This suggests that every tumor type may harbor cases where VPA could be sensitive and useful. It is clearly important that the tumor vs. normal effects of VPA be studies on paired samples from a single patient in order to convincingly prove the tumor-specific action of VPA in sensitive patients.

ï Is absolute or relative sensitivity more important when deciding whether to treat with VPA? If there is substantial toxicity in normal tissue, how much does it matter whether the effects are greater in the tumor?

ï The authors show that fulvestrant sensitivity is breast cancer subtype specific, whereas VPA sensitivity is not as specific to a particular subtype. Along those lines, it would be prudent for the authors to perform additional analyses of other drugs that effect precise biochemical pathways as a further validation of the MATCH approach. For example, the authors could demonstrate the effectiveness of MATCH using data for the BRAF pathway in melanoma, the ERBB2 pathway (transtuzumab) in breast cancer, the EGFR pathway (erlontib), the ALK pathway (crizotinib) in lung cancer, and others. Since we already know how to target these drugs to specific mutations, we should expect MATCH to correctly assess the sensitivity of patients to these drugs using the same sensitivity predictions as were used for VPA. These validations would be strong support for the MATCH method.

ï The authors report 39% of all breast cancers are VPA sensitive based on a cutoff of 0.5. What does this cutoff represent theoretically? Is there a justification for using it?

ï How does VPA perform versus doxorubicin in control mice or mice where VPA is not predicted to be effective? This is a key question that is not addressed by the authors. Without knowing the answer to this it is difficult to judge whether VPA sensitivity is driving the increased response, or merely the fact that VPA is a better drug than doxorubicin overall. If these studies have been performed previously, please cite them or explain why they would not be appropriate here.

Minor Issues:

ï Please include the cell line information in a supplementary table (it is included in the methods section) and specify the predicted sensitivities of each cell line

ï Include the sample size and correlation coefficient in the main text and/or figure legend

ï The number of samples predicted to be insensitive to VPA is low and could allow isolated cell lines be driving the correlation coefficient towards significance.

ï In Figure 5B, just two tumors predicted to be insensitive to VPA are driving the correlation. The p-value is barely significant and the sample size is low. For these reasons, it is difficult to know how to interpret these results.

Overall, the study is relevant, encouraging, and potentially suitable for publication, but some key questions remain unanswered by the authors. We look forward to seeing a revised manuscript that addresses these questions. Further, we look forward to the application of the MATCH system to the study of other drugs and cancer types.

1st Revision - authors' response 20 April 2011

Thank you for considering our manuscript and for your thoughtful and thorough reviews. We are grateful for your time and advice. In response to your feedback, we have substantially revised our manuscript and believe these changes improve our work. Below we summarize comments from the editor and reviewers, and the changes we have made to address them.

1. Response to editorial comments:

The reviewers had clear concerns that some of the experimental results, particularly those in Fig. 5B and 6, remained rather inconclusive considering the small number of samples (especially samples predicted to be VPA resistant). All three reviewers indicated clearly that the correlation presented in Fig. 5B is currently unconvincing, and, as such, additional samples with low predicted VPA sensitivity should be tested.

We agree with the reviewers that the inclusion of additional tumor samples would strengthen the figures from the 3-D cultures of patient samples and from the xenografts. While these patient samples are difficult to obtain in large numbers, an additional five pleural effusion samples have been collected from the time of our initial submission and have been added to the studies depicted in Figure 5 (now 6) to increase confidence in the correlation between predicted and actual drug response of patient tumors grown in 3-dimensional culture. (Page 16 paragraph 2) Of these, four are predicted to be VPA resistant. Further, we have now included an additional mouse xenograft that had low predicted response. (Page 16 paragraph 3) Importantly, the correlation between predicted sensitivity and EC50 in Figure 5B (now 6B) remains statistically significant, with a p value of 0.006 (which is improved from 0.03). Moreover, we remain 100% accurate at predicting which tumor xenografts will have growth restriction following VPA treatment and which will have no effect on growth. It is important to note that the tumor xenograft experiments were performed in a blinded fashion, with actual and predicted response performed in independent labs and linked following the conclusion of the study.

Conceptually, reviewer #2 cautioned strongly against overstating the potential clinical relevance of these results, noting that MATCH predicted tissue sensitivity to VPA did not seem to coincide with toxicity observed in clinical trials, and that the concentrations of VPA used here often exceeded clinically realistic levels.

We thank Reviewer #2 for pointing out that additional detail and references were needed to clarify and support our statements on this section of the manuscript. In particular, we have added two references confirming the hepatotoxicity of valproic acid in rat models. (Page 11 paragraph 2) Although no grade III (Five times the upper limit of normal) changes in transaminases have been reported in clinical trials with VPA taken on a non-continuous basis, trials of continuous valproic acid (Candelaria et al, ; Rocca et al) have up to 12% incidence of lower grade transaminase elevations. Moreover, toxicology data suggests transaminase levels underestimate valproic acid hepatotoxicity (Lee et al). Time course experiments (Tong et al) suggest that hepatoxicity may not become clinically apparent until after several days of treatment, longer than the courses used in the Munster studies. Additionally, gastrointestinal toxicity is one of the most common reported side effects of valproic acid treatment, although the mechanism has not been well defined. Lastly, a case report of pill esophagitis due to valproic acid suggests that valproic acid can have some esophageal toxicity at high concentrations (Yamaoka et al).

In regards to the concentration of VPA used in this study, the doses needed for increased histone acetylation with VPA exceed the usual antiepileptic doses. In particular, the Rocca et al clinical trial showed that 1-2mM doses were needed for histone deacetylase inhibition. In two clinical trials, one with advanced melanoma patients (Rocca et al) and one with breast cancer patients (Munster et al), the valproic acid levels found in some patients were 1-2mM, which is very close to the EC50 dose of several of the sensitive tumor cells in our studies. Further, we have also achieved levels of VPA in the 2mM range in our ongoing clinical trial that uses valproic acid for breast cancer, thus providing an independent confirmation on the clinical relevance of this dose (unpublished data, trial is ongoing). The EC50 of many sensitive tumors in figures 5 and 6B is approximately 3-4mM, suggesting the doses used in this study are indeed relevant. Moreover, the xenograft experiments validate that tolerable doses can achieve tumor-inhibiting levels in an animal. Whether intermittent high dose therapy, as used by Munster, or continuous therapy, as has been used in other trials, is optimal remains an open question. Lastly, while the EC50 of responsive cells to VPA in the culture systems used here is ~1-5mM, we expect that a proportion of drug is protein bound and not free. In particular, our drug response assays using 2-dimensional cell culture conditions are carried out in 5% FBS (which contains proteins known to bind drug). Additionally, the matrigel used in the 3-dimensional culture system is derived from a crude tumor matrix extract, which contains large amounts of proteins that may affect protein binding and free VPA levels. Therefore, taking these factors into consideration, we estimate that we are within a relevant VPA dose range for our in vitro studies.

The reviewers felt that the mechanistic connection between the VPA gene expression signature and tumor proliferation & survival remained largely unclear, and indicated that this issue should be explored and discussed in more detail. See in particular the "black box" comment by reviewer #3.

A gene ontology analysis has been added to the results section to try to elucidate this mechanism further. (Page 10 paragraph 2 and figure 2G) This study, as well as an independent study in our lab, has shown that VPA alters the expression of genes related to cell cycle regulation. Whether cell cycle regulation remains its primary mechanism of anticancer activity is not known and is an active area of research in the Bild lab. However, it is important to note that one of the strengths of MATCH is that the mechanism of a drug need not be completely understood in order to identify individuals more or less likely to respond to that drug. While the use of gene expression profiles to uncover drug mechanism is an essential area of research, MATCH is not dependent on this information.

On a more subjective level, the editor feels that this work may benefit from a title that better describes the key findings of this work

The title has been changed to "A pharmacogenomic method for individualized prediction of drug sensitivity: Valproic acid as an example "

2. Response to Reviewer#1:

It's not clear why the authors chose the binary regression based approach to generate the drug signature.

The binary regression approach is an integrated approach for simultaneously generating the drug signature in the training set, reducing the dimension of the signature, and projecting the signature into the test set. We have used this method with great success in the past (Bild et al). The method was developed and compared with other competing methods by (West et al). To summarize the binary regression approach from West et al: the approach trains a binary prediction model on metagenes from training set in order to maximize the predictive ability on the model on the training set. The model is then used to predict treatment response in the test set. Part of the novelty of the algorithm is in the selection of metagenes, which utilizes a Bayesian regularized singular value decomposition that consists of placing informative prior weights on the individual gene contributions to the metagenes. The novelty of this method is that the regularization re-weights the genes in the metagenes based on their ability to improve the predictive power of the binary regression predictor in the training set. Therefore, the principle behind using this approach is that the signature developed from this approach is optimized for prediction and should outperform methods for selecting the signature that are independent of prediction, e.g. SAMR, RANKPROD, or XYZ as the reviewer suggests. In addition, in the past our group and West et al. have compared this binary regression approach with other prediction approaches such as logistic regression and support vector machines and have shown that the binary regression has consistently better performance.

The authors seem to ignore the power of the large data set of data found in connectivity map. It would be nice to see what you could predict by random chance against all the treated vs. untreated samples in connectivity map for example.

In order to capitalize further on data in the Connectivity Map, we projected the VPA signature into nine random batches from the Connectivity Map that contain at least one VPA treated sample and at least one sample treated with a different drug. (Page 9 paragraph 3) Figures showing the results have been added to the paper as figures 2D, 2E, and 2F. These results confirm that the signature consistently separates cells treated with VPA from untreated cells or cells treated with unrelated drugs, and the differences are statistically significant. There is some overlap between the predictions for cells treated with other HDAC inhibitors.

More generally, while there are many drugs tested in the Connectivity Map, currently the minority are cancer specific drugs, and many are without known effects against different cancer types. Because of its relatively low toxicity and unknown target group, VPA provided us with a unique opportunity to better understand the optimal target population for HDAC inhibitors, to test our findings using cell lines and patient tumors, and to then initiate a clinical trial based on our findings. As the number of drugs included in the connectivity map increase, we hope to include more drugs in MATCH for profiling and testing in cancer cells.

is there a principled, or data-driven way to justify the choice of VPA?

MATCH is a method of identifying a target population and validating this population for a targeted drug that is thought to have anticancer properties. That being said, not every drug will give a consistent signature. Drugs with fewer than 4 treated samples in the Connectivity Map, such as gefitinib, are unlikely to provide robust signatures for prediction of drug responsiveness. Similarly, drugs that require activation by enzymes not active in cancer cells may not give accurate results. In addition to the reasons listed above (unknown target population and clinical properties such as low

toxicity), VPA was chosen for further validation due to its ability to achieve consistently excellent LOOCV and external validation dataset predictions. Therefore, we select VPA as our proof of principle both because of both principle and data-driven factors. A sentence to this effect has been added to the introduction. (Page 6 paragraph 2)

Figure 1 is a bit confusing. It's nice to see the steps, but the descriptions are a bit vague. It's hard to see how information is flowing through this decision process and what is flowing from one step to the next.

We have modified Figure 1 to clarify information flow.

3. Response to Reviewer #2:

(Note: Please see response to editor above for replies to Reviewer #2's concern about VPA levels and toxicity, and request for the addition of extra samples for patient 3-D and xenograft work.)

Should other HDAC inhibitors yield similar predictions?

The Connectivity Map does include Trichostatin A (TSA), another putative HDAC inhibitor with similar mechanism of action. Our predictions with TSA do not correlate with the VPA predictions, likely due to the broader HDAC inhibition by TSA or due to dosing issues. In other experiments in our lab, signatures for VPA and vorinostat have a high correlation, thereby supporting the consistency and accuracy of the findings presented here.

It is also unclear why the authors excluded sarcomas and leukemia from their analysis; especially as leukemia is a potentially more sensitive tumor type than solid tumors

Because the Connectivity Map samples used to create the signature were from carcinomas, the signatures have not been validated except in carcinomas. Because of different gene expression backgrounds between hematopoietic cells and epithelial cells, we cannot assume that predictions in leukemias are accurate without a more detailed investigation. Therefore, we do not wish to present predictions in a class of tumors where we have not validated accuracy and therefore cannot be assured they are meaningful.

What is the author's definition of 'response', such as 'drug response'? Do they mean tumor regression (which matches the clinician's definition of response) or lump together tumor regression and stasis or mean tumors that have changes in gene expression from compounds irrespective of effects on tumor growth?

We have used the word response in two different contexts in the paper. The phrase "drug response signature" refers to changes in gene expression in cells in response to a drug, which is how the signatures are made. The phrase dose-response refers to decreased tumor growth or tumor shrinkage in response to drug treatment. We see decreased proliferation and increased cell death in our in vitro experiments. In some places we did use the word response to refer to decreased cell growth or increased cell death, but we have changed all of these to the word sensitivity to avoid confusion. (Page 6 paragraphs 1 and 2, page 10 paragraph 3, page 11 paragraph 2, page 12 paragraph 2, page 16 paragraph 2, and page 17 paragraphs 1 and 2) Our validation experiments show that drug response as measured by gene expression changes is indicative of sensitivity to the drug in assays for cell proliferation and apoptosis.

Fourth, it seems this MATCH algorithm is best at defining genes that are expressed and modulated by VPA, but changes in these genes do not necessarily mechanistically explain, correlate with, or

predict tumor cell death or stasis secondary to compounds.... Thus, MATCH could not predict cytotoxicity or cytostasis of these compounds if the genes are not casually associated with cell death/proliferation.

The signatures generated by our approach could be considered mechanism independent. Generally, the signatures represent the acute transcriptional changes in tumor cells that correlate to drug response. These gene expression changes may be a direct result of the drug or may be due to the cells' response to the effect of the drug. While genes modulated by drug could include immediate effectors of cell growth and apoptotic pathways, we do not specifically capture these changes, and instead focus on the genes that best describe a cell's response to a drug. Primary or secondary effects on gene transcription could include differentiation, apoptosis, or cell stasis specific genes, but our model is not dependent on their identification. For example, our signature for fulvestrant contains genes related to TGF signaling and to nucleotide metabolism. Although these are not in directly related to cell death or proliferation, it is their interaction with estrogen signaling that leads to cell death in estrogen addicted cancer cells.

An ovarian normal cell line is used for validation of VPA treatment (Figure 2C). But your other data suggests that ovarian cancer cell lines are resistant to VPA and there is no data on the normals to suggest either way.

The validation set (GSE1615) used ovarian theca cells, which are different than the cell of origin of ovarian cancer, which is probably either the epithelial lining of the ovary or the fallopian tube. (Dubeau) This has been clarified in the text. (Page 9 paragraph 2, page 25 paragraph 2) Valproic acid may cause apoptosis in theca cells. (Tauboll et al) The validation has been extended to other cell lines as well as described above.

Why did the authors choose EC50 rather than IC50? For some cell lines, EC50 can be much lower than IC50 and thus falsely associate sensitivity of a cell line to treatment; for example, a tumor cell shows minimal growth arrest (magnitude of effect) but this change occurs quite low in the concentration curve and thus the EC50 is low but the IC50 is high.

Because we were assessing effects on cell or organoid number and/or tumor size rather than on functional assays, such as of HDAC inhibition, EC50 is more appropriate than IC50. Although relative EC50 can be misleadingly low if the total effect is low; near complete cell death was achieved with high enough doses for all but the most resistant cell lines and tumors. Moreover, all of our dose-response curves were modeled from zero response to 100% response, so that the EC50 reported is the dose needed to achieve an absolute 50% decrease in cell or organoid number. For space considerations, dose-response curves were omitted from the manuscript, but can be provided in a supplement if requested.

The authors perform one experiment comparing VPA to doxorubicin and conclude that "VPA decreased tumor growth significantly more than doxorubicin". This is a dangerous overstatement as doxorubicin has been a highly effective therapy for breast cancer while the efficacy of VPA is still not clear. One mouse experiment is hardly strong data to support this concept

We agree that the implications of one experiment should not be overstated. We have moved this figure to supplemental data in order to avoid the suggestion that VPA is better than doxorubicin. Moreover, this sentence in the results has been qualified to indicate that it is a description of the results of this experiment, not a conclusion about the relative efficacy of these two drugs in general. (Page 17 paragraph 1) We note that neither the discussion nor the abstract reference this experiment or make claims about the relative efficacy of doxorubicin and VPA.

Most of the clinical data to date suggests little tumor regressions (responses) to HDAC inhibitors in solid tumors including breast and lung cancers, two tumor types predicted to be enriched for sensitive tumors using MATCH.

There is insufficient clinical trial experience to know what the response rate or the disease control rate, which includes stable disease, to valproic acid is in breast, and particularly lung cancer. Published clinical trials including valproic acid have included one non-small cell lung cancer patient. The responses of breast cancer patients are difficult to interpret because of differences in valproic acid schedule (three days vs continuous) and in accompanying chemotherapy. Moreover, the predicted sensitivity in these tumor types is variable. Therefore, trials are needed assessing response in those women predicted to be sensitive. Such trials are ongoing.

R2 should be provided for all figures with best fit lines…

R2 values have been added to figures 5 (previously 4) and 6B (previously 5B)

From the methods section, it appears that the binary regression is normalized to 0/1 scale across datasets. What if the proportion of responses is different across different datasets? If one is all resistant, then some of the resistant points will be scaled to sensitive?

The binary regression output is indeed normalized to a zero to one scale within each dataset. All datasets used in this paper are selected to be large and diverse enough to include both sensitive and resistant tumors. Leaving out the standardization has minimal effect on the results.

**Note: In addition to our response to the reviewer's main comments, all corrections and suggestions from Reviewer #2 have also been included in revised manuscript.

4. Response to Reviewer #3:

VPA signature is treated as a "black box" by the authors…

We have added detail concerning the genes in the signature, as discussed above. (Page 10 paragraph 2)

VPA is undergoing clinical trials for several cancers - do those tumor types agree with your assessment of sensitivity based on tumor type?

The results of many of these trials are not available, so we cannot know if they correspond to our predictions. There are ongoing trials of breast and lung cancer, cancer types we have highlighted in our manuscript. In addition, one advantage to our approach (in addition to identifying responsive tumor types) is to identify those tumors within each type that will be sensitive to VPA.

Is there a set of matched tumor and normal tissues from the same patients? Can you predict differential sensitivity in a single patient independent of tumor type? This would be more revealing than making assumptions about the underlying tumor histology.

The dataset from GSE 5364 contains tumor and matched adjacent normal tissue for most of the samples. We can, thus, predict differential sensitivity within one organ within a single patient. Therefore, the differential sensitivity is not simply a reflection of between person differences.

There is no set that we are aware of tumor and different normal tissues from a single patient that would allow within patient predictions of both tumor sensitivity and toxicity within one patient.

Is absolute or relative sensitivity more important when deciding whether to treat with VPA? If there is substantial toxicity in normal tissue, how much does it matter whether the effects are greater in the tumor?

We agree with the reviewer that both absolute and relative sensitivity are important when using a drug clinically. The importance of absolute toxicity, however, depends on the clinical context, which cannot be assessed by in silico or in vitro algorithms. For example, even drugs with very high absolute toxicity can be used if they have high relative therapeutic index and if a rescue for normal tissues is available, such as in high dose chemotherapy with autologous stem cell rescue or high dose methotrexate with leucovorin rescue. In addition, absolute toxicity may be dependent on supportive care, which also cannot be measured in vitro or in silico. For example, cisplatin became much more tolerable with the discovery of 5-HT3-antagonists and substance P-antagonists. Therefore, while the MATCH algorithm can identify a drug with a therapeutic index in specific patients, it does not preclude the need for clinical studies to determine the best way of exploiting this therapeutic index.

The authors show that fulvestrant sensitivity is breast cancer subtype specific, whereas VPA sensitivity is not as specific to a particular subtype. Along those lines, it would be prudent for the authors to perform additional analyses of other drugs that effect precise biochemical pathways as a further validation of the MATCH approach. For example, the authors could demonstrate the effectiveness of MATCH using data for the BRAF pathway in melanoma, the ERBB2 pathway (transtuzumab) in breast cancer, the EGFR pathway (erlontib), the ALK pathway (crizotinib) in lung cancer, and others.

Unfortunately, the Connectivity Map does not yet include antibodies targeted to growth factor receptor pathways. Further, the only tyrosine kinase inhibitors that are included in the Connectivity Map that have been used clinically (gefitinib and imatinib) have only a couple treated specimens, and therefore do not allow the generation of a robust drug response signature at this time. However, our approach is not absolutely dependent on the Connectivity Map, so to address this concern we leverage data from an additional group using melanoma cells treated with the BRAF inhibitor PLX4032 (GSE 20051). PLX4032 has a very high response rate in melanoma patients with a mutated BRAF. (Flaherty, 2010, NEJM) We generated a drug response signature for PLX4032 and applied the beginning steps of the MATCH algorithm to it. MATCH correctly predicted high sensitivity to PLX4032 in melanoma, thyroid cancer, and GI cancers, which is what is seen or expected clinically, and low sensitivity in neuroblastoma and small cell lung cancer, which do not depend on pathways involving RAF. The validation for this drug has been added to the paper. (Pages 13 and 14 and figure 4)

The authors report 39% of all breast cancers are VPA sensitive based on a cutoff of 0.5. What does this cutoff represent theoretically? Is there a justification for using it?

We chose 0.5 because it is intuitively appealing and because it is very close to the optimal point on the ROC curve for predicting whether a sample has been treated with VPA versus no treatment or treatment with a non-HDAC inhibitor. The ROC curve for VPA is now included as figure 2F, and the point corresponding to a cutoff of 0.5 is indicated on the curve. (Page 10 paragraph 1)

How does VPA perform versus doxorubicin in control mice or mice where VPA is not predicted to be effective?

This is a key question that is not addressed by the authors. Without knowing the answer to this it is difficult to judge whether VPA sensitivity is driving the increased response, or merely the fact that VPA is a better drug than doxorubicin overall. If these studies have been performed previously, please cite them or explain why they would not be appropriate here.

As discussed above, doxorubicin is very active in breast cancer, and it is very doubtful that VPA is a better drug than doxorubicin overall. The doxorubicin experiment is intended to show only that there exist tumors predicted to be sensitive to VPA for which the tumor may be more sensitive to VPA than doxorubicin. However, in response to reviewer comments, we have both clarified the conclusion from this experiment and placed the figure in the supplemental data section.

**Note: In addition to our response to the reviewer's main comments, all minor issues from Reviewer #3 have also been addressed in the text/data.

We appreciate the comments from all three reviewers and the editor. We believe we have addressed all of the concerns listed by the reviewers, which has significantly strengthened our manuscript. In particular, we have added additional patient samples to our assessment of predicting VPA sensitivity in patient tumors, and we have highlighted an additional drug to validate the generalizability and clinical utility of the MATCH approach. In closing, our results highlight the ability of MATCH to use genomic analysis with in vitro testing of patient tumors to select optimal drug regimens prior to clinical trial initiation, and believe this work is a good fit with the overarching focus of Molecular Systems Biology. We thank you for your consideration of our research for publication in Molecular Systems Biology

2nd Editorial Decision 24 May 2011

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from two of the three referees who agreed to evaluate this revised work, and have decided to render a decision now to avoid further delay. As you will see, the referees felt that the revisions had substantially improved this work, and they are generally supportive. The second reviewer, however, has a few of remaining concerns that may require some additional discussion or clarification, which we would ask you to carefully address in a final revision of the present work.

Reviewer #2 is still concerned that the correlation between predicted sensitivity to VPA and the experimental EC50 occurs over a range that extends well beyond the clinically plausible regime. The editor does not see that this compromises the usefulness of your method -- MATCH can still help classify the samples into tumors that could be treated with clinically achievable VPA concentrations and those that likely cannot. Nonetheless, it may be worthwhile to briefly to discuss what plausible treatment concentrations are likely to be, and acknowledge that MATCH would not be well-suited for predicting optimal VPA treatment dosage in a clinical setting (particularly given the high score variably at low EC50 values).

In addition, please consider the other points made this reviewer, including discussing the additional studies mentioned and incorporating data related to the TSA and vorinostat signature comparisons into the supplementary material.

The Editor also requests that you address the following format and content issues before submitting a revised manuscript:

1. Please combine all of the Supp. Figures and Supp. Table 2 into a single PDF file entitled Supplementary Information (Supp Table 1 is large enough that it should remain a separate file). This file should begin with a Table of Contents listing all supplementary material included with this work. Figure legends should be below or immediately following the Supp Figures.

2. In addition to our capacity to host datasets in our supplementary information section, we provide a new functionality that allows readers to directly download the 'source data' associated with selected figure panels (e.g. <http://tinyurl.com/365zpej>). We ask that the data now provided in Supp. Table 3 be provided in this format, and encourage you to submit a similar source file for Figure 7. A document with guidelines for the preparation of Figure Source Data has been attached.

3. Please supply high-resolution final images for each Figure in the main manuscript in TIFF, EPS, or PDF formats (the current resolution is rather low).

4. Several citations in the main text lack years (e.g. page 13, Joseph et al)

Thank you for submitting this paper to Molecular Systems Biology.

Yours sincerely,

Editor

Molecular Systems Biology

Referee reports

Reviewer #1 (Remarks to the Author):

I have read through the author's response to reviewer comments and read through the paper once again. I feel that the manuscript has much improved since the previous version. I am happy to see that the authors added additional patient samples which resulted in even stronger statistical results for their predictions. I am also happy that they compared the robustness of the VPA signature to that of drugs randomly selected from ConnectivityMap. This leads me to believe that there is a real effect from the VPA and that the predictions are not from some bias in the ConnectivityMap data, which is based on cancer cell lines. Also, the reasoning behind the choice of VPA, use of the binary regression method, and the rationale behind the entire analysis pipeline is much more clear. This reviewer is satisfied with the revisions made, and would suggest that the manuscript is now appropriate for publication in MSB.

Reviewer #2 (Remarks to the Author):

New data has been added to figure 6B; this adds additional tumors samples predicted to be resistant to VPA. Now the correlation between EC50 and predicted sensitivity remains statistically significant (actually, p-value improves). But my original concern about the clinical applicability of this

remains. The X-axis for the graph in Figure 6B is logEC50; evidence from the literature, cited by the authors, suggests it is doubtful that 10mM of VPA can be achieved in patients - even the authors acknowledge this in their rebuttal, as concentrations obtained in published studies are ~1-2 mM and the authors' own clinical trial obtains VPA levels of 2mM. If you now focus attention on the values less than 10mM, it is hard to see any correlation. The values >10mM really drive the regression line in Figure 6B. The authors should address this and modify the discussion. Otherwise readers can be misled by these types of curves, especially regarding the clinical correlations and predictions of this type of work. Thes science of MATCH is elegant, these types of approaches are important in drug development, but one has to be careful about overstating the capability of these pedictors to work in the clinical arena, as drug levels and in vivo tumor complexity enter into the equation.

Where are the results presented in the new version describing correlations between VPA and vorinostat and TSA?

Traynor and colleagues published a vorinostat (HDAC inhibitor) study in Journal of Thoracic Oncology 2009 finding no responses in advanced lung cancer; the study had low numbers, mainly because of deciding the therapy was not worth pursuing. You should mention this in your paper, as your study suggests lung cancer to be one responsive subtype. Similar negative results were reported by Vansteenkiste in Invest New Drugs in 2008. Maybe MATCH can identify subsets of patients who benefit for HDAC inhibitors, which would be welcome. But one should be cautious about the efficacy of these compounds in solid tumors, as the evidence to date is weak.

2nd Revision - authors' response 01 June 2011

Thank you again for considering our manuscript and for your thoughtful critiques. Below are the changes we have made in response to your letter on May 24, 2011.

1. Response to Reviewer #2:

The X-axis for the graph in Figure 6B is logEC50; evidence from the literature, cited by the authors, suggests it is doubtful that 10mM of VPA can be achieved in patients - even the authors acknowledge this in their rebuttal, as concentrations obtained in published studies are ~1-2 mM and the authors' own clinical trial obtains VPA levels of 2mM. If you now focus attention on the values less than 10mM, it is hard to see any correlation. The values >10mM really drive the regression line in Figure 6B. The authors should address this and modify the discussion. Otherwise readers can be misled by these types of curves, especially regarding the clinical correlations and predictions of this type of work.

In response to the reviewer’s request, we have added a new paragraph to the discussion (page 19, paragraph 3) to discuss the limitations of MATCH in determining the absolute risk-benefit of a drug and the inability of MATCH to predict the correct dose of a drug for clinical use as well as a sentence to the end of page 15 paragraph 1 focusing on how predicting sensitivity and predicting resistance are two sides of the same coin.

We would like to emphasize that the concentration -response relationship of drug in vitro is not generally the same as the plasma concentration-response relationship of that same drug in humans. Protein binding in media makes predicting the free VPA levels in a 4-10mM solution of VPA in cell culture media or matrigel-culture media difficult. Therefore, the equivalent drug concentrations to the in vitro concentrations in our experiments may be lower than expected. In fact, an average 2-fold

difference in EC50 between clinical studies and in vitro work is not unusual. Further, it is important to recognize that the dosing regimen of the in vitro experiments and our Valproic Acid Signature Trial (VAST) is not equivalent. In particular, for the in vitro work, cells are dosed only once on day 0 while embedded in matrigel, and then we measure proliferation of cells four days later (using a standard MTT based assay). In contrast, for our clinical trial, over the course of a week patients are given VPA twice a day at steadily increasing doses. The relationship of a tumor's response in an in vitro setting where cells are treated once to a tumor's response upon a relatively continuous treatment over the course of a week may not be comparable. Therefore, given the disparate systems and dosing regimens between the in vitro studies and clinical trial, it remains difficult to make a strict comparison of the doses and their respective effects.

Importantly, for the research presented in this paper, the absolute EC50s of the samples in the in vitro experiments are not nearly as important as relative EC50s. The ability of MATCH to identify tumors that are either sensitive or resistant is useful, even were it only able to identify those samples that are very resistant to a treatment. While in this particular setting MATCH (or our proliferation assay) may not distinguish between tumors that are 100% sensitive versus those that are 90% sensitive, being able to distinguish sensitive versus resistant tumors is of value.

Where are the results presented in the new version describing correlations between VPA and vorinostat and TSA?

We agree that comparing these different drugs and refining their use is important. We currently have a manuscript detailing the relationship between VPA, SAHA, and TSA in review at Pharmacogenomics journal. The manuscript under review at Pharmacogenomics focuses on what pathways are modulated by these drugs as well as synergistic combinations containing these drugs, and therefore are unable to add those results to the manuscript under review at Molecular Systems Biology at this time. We had initially focused on VPA in our basic science, computational, and clinical studies, and therefore had maintained this focus for the Molecular Systems Biology manuscript.

Traynor and colleagues published a vorinostat (HDAC inhibitor) study in Journal of Thoracic Oncology 2009 finding no responses in advanced lung cancer; the study had low numbers, mainly because of deciding the therapy was not worth pursuing. You should mention this in your paper, as your study suggests lung cancer to be one responsive subtype. Similar negative results were reported by Vansteenkiste in Invest New Drugs in 2008.

We agree with the reviewer for caution in the interpretation of our results, we agree that further studies are needed (such as our clinical trial) to assess the populations most sensitive to VPA. Therefore, we have added text to the end of the first paragraph of page 19 referencing these studies and discussing in more detail the limitations/advantages of a technique such as MATCH

Given our focus on VPA, we are hesitant to extrapolate our results to a different drug in the discussion, even one that may have a similar mechanism of action. Just as different anthracyclines or vinca alkaloids are used in leukemias and solid tumors, it is possible that different HDAC inhibitors will be useful in different sets of tumors. In fact, VPA, SAHA, and TSA have been found to block different HDACs, highlighting their potential to be effective in different populations. Moreover, non-small cell lung cancer was in the middle of sensitivity in our analysis of the GSK cell lines (Figure 3A). Therefore, we would expect that only a subset of lung cancer patients will be sensitive to VPA. Further, while the Traynor study did not find a sensitive population, the study had fewer than 20 people, included only relapsed NSCLC patients, contained a non-representative population of 81% women, and did use any method to select for a sensitive population. Additionally, in the Vansteenkste study, patients were treated only briefly. Despite the use of vorinostat as a single agent in recurrent NSCLC, we note that the Traynor study had a high rate of stable disease with several prolonged times to progression, leading them to conclude vorinostat may have a role in combination with other agents in lung cancer.

