35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Identification of signaling pathways related to drug efficacy in hepatocellular carcinoma via integration of phosphoproteomic, genomic and clinical data

Ioannis N. Melas, Douglas A. Lauffenburger, Leonidas G. Alexopoulos

Abstract— **Hepatocellular Carcinoma (HCC) is one of the leading causes of death worldwide, with only a handful of treatments effective in unresectable HCC. Most of the clinical trials for HCC using new generation interventions (drug-targeted therapies) have poor efficacy whereas just a few of them show some promising clinical outcomes [1]. This is amongst the first studies where the mode of action of some of the compounds extensively used in clinical trials is interrogated on the phosphoproteomic level, in an attempt to build predictive models for clinical efficacy. Signaling data are combined with previously published gene expression and clinical data within a consistent framework that identifies drug effects on the phosphoproteomic level and translates them to the gene expression level. The interrogated drugs are then correlated with genes differentially expressed in normal versus tumor tissue, and genes predictive of patient survival. Although the number of clinical trial results considered is small, our approach shows potential for discerning signaling activities that may help predict drug efficacy for HCC.**

I. Introduction

HCC is one of the leading causes of death worldwide [2]. Traditionally, the etiology of the disease is attributed to genetic alterations that accumulate during chronic inflammation of the liver. Mutations are found in several important genes including p73, p53, Rb, APC, DLC-1 (deleted in liver cancer), p16, PTEN, IGF-2, BRCA2, SOCS-1, Smad2 and Smad4, β-catenin, c-myc, and cyclin D1 [3]. Moreover, as in other cancers, HCC is characterized by an imbalance in growth promoting signals and the MAPK cascade [3]. Approved treatments so far for unresectable HCC include sorafenib and erlotinib [4,5] that target the VEGFR, PDGFR and RAF kinase, and the EGFR respectively. However, with the average survival benefit of these treatments at about 3 months, it is evident that identification of new targets for HCC is of the utmost importance.

On this front, fields like systems biology attempt to take advantage of the data generated by the new -omic technologies to identify suitable genes/proteins whose biological activity can be directly linked to pathological processes. E.g. an increasing number of studies tackle the complete characterization of tumors' gene expression profiles and protein content [6-8]. These approaches have succeeded in identifying several hundreds of genes and proteins that are differentially expressed in tumor vs normal tissue on the same patient, or genes that are differentially expressed across different patients and are predictive of cancer metastasis, or patient survival. However, applying this knowledge in drug discovery is not a straightforward procedure. Data must also be incorporated that capture the way cells function and respond to factors of its microenvironment (i.e. signaling data). Signaling data can provide the causality/directionality much needed in gene expression networks and uncover the genes that truly regulate the disease phenotype.

The importance of intracellular signaling in HCC has been well established and interrogated [9], while a number of new drugs target kinases or receptors differentially expressed in disease. However, with most of these drugs (especially the approved ones) being highly promiscuous and their effects on the cell's signaling pathways not yet studied in a systematic manner [10], we have yet to discover the key features that are predictive of drugs efficacy, reflecting also the fact that key aspects of this disease elude us.

Herein, we propose a consistent framework for the integration of signaling, gene expression and clinical data, aiming at the identification of signaling pathways related to drug efficacy in HCC. We have put together a signaling dataset consisting of the phosphoproteomic response of 3 HCC cell lines, presence of 8 drugs for unresectable HCC, most of which of known clinical efficacy, and attempted using recursive feature extraction to identify the phosphoproteomic signatures that are most predictive of drug efficacy. We, subsequently, translated our findings to the gene expression level, where we inferred regulatory networks between the identified phosphoprotein features and gene sets known to be implicated in HCC (either differentially expressed between tumor and normal tissue on the same patient, or differentially expressed across different patients and predictive of metastasis, or survival), leading to the identification of a subset of genes that could possibly govern patient survival and/or drug efficacy. The analysis presented herein could serve both for the identification of drug targets, as well as a new framework for the integration of signaling, gene expression and clinical data, aiming towards the holistic study of mechanisms implicated in drug efficacy.

II. Methods

A. Data collection and normalization

3 HCC cell lines were interrogated (huh7, hep3b, hepg2), by measuring the activation level of 16 key phosphoproteins (P90RSK, AKT, SRC, CREB, IRβ, MEK1,

IN Melas and LG Alexopoulos is with the Department of Mechanical Engineering, National Technical University of Athens, 15780 Zografou, and with ProtATonce Ltd, 15343 Athens, Greece

DA Lauffenburger is with the Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and the Center for Cell Decision Processes, Massachusetts Institute of Technology and Harvard Medical School, Boston, Massachusetts, USA

IKβ, HSP27, P70S6, GSK3, HISTH3, JNK, STAT3, ERK12, P38, IRS1), under 7 stimuli (IL1β, TGFα, Heregulin (HER), Insulin (INS), HGF, IL6 and TNFα) and presence of the following 8 drugs for unresectable HCC [10]: Lapatinib, Gefitinib, Erlotinib (EGFR inhibitors), Sorafenib (inhibitor of VEGFR, PDGFR and of Raf kinases C-Raf and B-Raf), Vandetanib (VEGFR and EGFR antagonist), Sunitinib (PDGFR and VEGFR kinase inhibitor), Dasatinib (multi-BCR/ABL and Src family kinase inhibitor), and Bortezomib (proteasome inhibitor). Regarding clinical trial results for these drgus, Lapatinib, Gefitinib, Vandetanib, Sunitinib shown poor outcomes [11-14], Sorafenib and Erlotinib had good clinical outcomes [4,5], while Dasatinib and Bortezomib were still under investigation [15].

The 16 signals were chosen based on assay availability and quality controls performed at early stages of the experimental setup. The 6 stimuli were chosen to perturb most of the pathways observed by the 16 signals. The data was averaged in an "average cancer cell type" and normalized using a linear regression model, that modeled the measured value of each signal as a linear function of the stimuli and inhibitor introduced, and analyte measured [16].

B. *Identification of signaling pathways predictive of drug efficacy*

Recursive feature extraction (RFE) on the phospho-proteomic data was implemented to identify the signaling "features" that are predictive of drug efficacy. RFE was implemented using the Matlab "***classify()"*** function within a custom Matlab script. Data was formatted as a 2-D matrix, with rows corresponding to the different drugs and columns corresponding to observations or features (i.e. signals measured under the various stimuli). 5 of the 8 interrogated drugs were used: Lapatinib, Gefitinib, Vandetanib, Sorafenib and Erlotinib, while Dasatinib, Bortezomib and Sunitinib were omitted due to limited clinical trial outcomes, or limited effects on the measured phosphoproteins. The differential response of the drugs was used, computed as follows. For every feature (column), the mean response over all drugs was evaluated and the differential response of each drug was obtained by subtracting the response of that drug from the mean.

Ultimately, every one of the remaining observations was used separately and a classifier was trained to distinguish the drugs between PASS and FAIL according to the clinical trial data available. This process was repeated as many times as the available drugs, every time leaving one of the drugs out as a test dataset to evaluate the classifiers performance. The average (over all drugs) performance of the classifier for each feature was evaluated. The features that best classify the drugs into PASS or FAIL are predictive of drug efficacy.

D. *Linking extracted features to significant and survival genes in HCC*

The extracted features, that are predictive of drug efficacy, are linked to significant genes in HCC. The gene expression data published in [7] was obtained from GEO (Series GSE3500) and was used to infer a network, connecting the measured phosphoproteins to genes differentially expressed in normal versus cancer tissue (hereafter referred to as significant genes). To this effect the signaling pathway published in [10] was used to obtain the transcription factors (TFs) downstream the measured phosphoproteins (P53, CREB, FOS, JUN, ATF2, ELK1, STAT1, STAT3 and NFKB); subsequently, TRED (Transcriptional Regulatory Element Database) was used to obtain the target genes of the TFs and ARACNE [17] was used to infer a mutual information network connecting the target genes to the significant genes.

Every one of the target genes was scored based on the number of connections to significant genes. The 10 most highly scored target genes are obtained and their connectivity to random subsets of GSE3500 (of equal size to the significant genes) is examined to identify which of these 10 target genes are most strongly correlated to the significant genes than to any other gene set (see Fig. 3A). In this manner a regulatory network is constructed from the phosphoprotein level, where the interrogated drugs act, through the affected TFs, to the gene expression level, where genes differentially expressed in normal versus cancer tissue have been identified.

The same procedure was also used to link the extracted features to survival genes in HCC. Lee et al in [6] identified a set of genes to be highly correlated to hazard ratios in HCC (hereafter refered to as survival genes). ARACNE was used to infer a mutual information network connecting the TFs target genes to the survival genes (see Fig. 3B).

III. RESULTS

A. *Phosphoproteomic data*

3 HCC cell lines were interrogated by measuring the activation level of 16 key phosphoproteins under 6 stimuli and presence of 8 drugs for unresectable HCC. The phosphoproteomic data for the average cancer cell type is shown in Fig. 1 (figure is placed at the end of the manuscript due to size). Regarding the effects of the interrogated drugs: Lapatinib inhibited AKT activation and partly CREB, MEK1 and ERK12 under TGFα. Also, inhibited CREB activation, MEK1, ERK12 and partly AKT under HER, while had no significant effects on any other pathway. Gefitinib had the same effects as Lapatinib under HER stimulation and also inhibited most signals under TGFα, such as AKT, CREB, MEK1, ERK12, P38, and IRS1. Gefitinib also partly inhibited HSP27 under IL1β. Sorafenib had no clear effects on the TGFα or HER pathway, but did inhibit MEK1, P70S6, and ERK12 under HGF and HSP27 in the IL1β pathway. Erlotinib and Vandetanib had very similar effects to Gefitinib under TGFα and on most of the signals under HER, apart from AKT (Gefitinib and Lapatinib both inhibited AKT under HER, while Erlotinib and Vandetanib left AKT unaffected). Sunitinib had no clear effects on any of the signals, apart from IRB under INS. Dasatinib also had no clear effects on the measured signals, indicating that the drug's mode of action is outside the observable part of the pathway. Bortezomib, being a proteasome inhibitor, increased activation of IKβ and IRS1 under all stimuli treatments.

B. *Phosphoproteomic features predictive of drug efficacy*

Recursive feature extraction (RFE) on the phospho-proteomic data was implemented to identify the "features" that are predictive of drug efficacy. Results are shown in Fig. 2. The most predictive phosphoprotein features are the measurement of (i) AKT under TGFα, (ii) AKT under HER, and (iii) ERK12 under HGF. In more detail, the feature extraction dictates that inhibition of (i) and (ii) is indicative of a drug that failed in clinical trials, while inhibition of (iii), of a drug that succeeded in clinical trials. (accuracy 80%). To ensure the significance of these results, the same analysis was performed after scrambling the classes (not shown here).

C. *Correlations between extracted features and significant/survival genes in HCC*

In this step, an independent approach is used to correlate gene expression results of HCC to the phosphoproteomic signatures. The extracted phosphoprotein features, that are predictive of drug efficacy, are linked via their downstream TFs to significant and survival genes in HCC. The target genes of the TFs are obtained from TRED and ARACNE is used to infer a mutual information network connecting the target genes to significant and survival genes in HCC. In this manner the biological relevance of the extracted phospho-protein features is validated and the mechanism by which they govern drug efficacy is identified. In Fig. 3 the top scoring target genes are shown.

SIAH1 and NME1, are the most highly scored genes (genes that are found to correlate more to the significant genes than to any other gene set), are both target genes of P53. P53 is inhibited by AKT via MDM2 [18]. This supports our previous finding that inhibition of AKT under TGFα and HER is predictive of drug efficacy in HCC and identifies the respective mechanism: a drug that inhibits AKT will lead to an increase in the activity of P53, that will affect SIAH1 and NME1 that correlate strongly to genes differentially expressed in cancer versus normal tissue. Apart from P53, JUN is also correlated to significant genes through its target genes LOX and IL7R, and is too affected by AKT, supporting our speculation. Results shown in Fig. 3B further validate our findings, as the most highly scored gene (KLK3) is a P53 target gene.

IV. Conclusion

Herein, we identified signaling pathways implicated in drug efficacy in HCC by combining signaling, gene expression and clinical data. A key finding of this work is that inhibition of AKT under TGFa (the most popular target for many types of cancer) may have a negative effect on drug efficacy, apart from the positive effect of blocking HCC cell proliferation [19]. Inhibition of AKT interferes with the p53 branch (via MDM2 [18]) that according to our TF analysis is a key regulator of genes differentially expressed in HCC. SIAH1, NME1 and KLK3 all target genes of p53 (and differentially expressed in HCC) are highly correlated to patient survival and have been identified before to be implicated in cancer [20, 21, 22]. The importance of p53 in HCC is also underlined in [20], where reduced p53 activity was found to induce HCC progression and thus have negative effects on clinical efficacy.

Even though our methodology is limited by the small number of drugs, this is amongst the first attempts to build a framework for integrating signaling, gene expression and clinical data in order to build strong hypotheses for drug targets in HCC and other diseases.

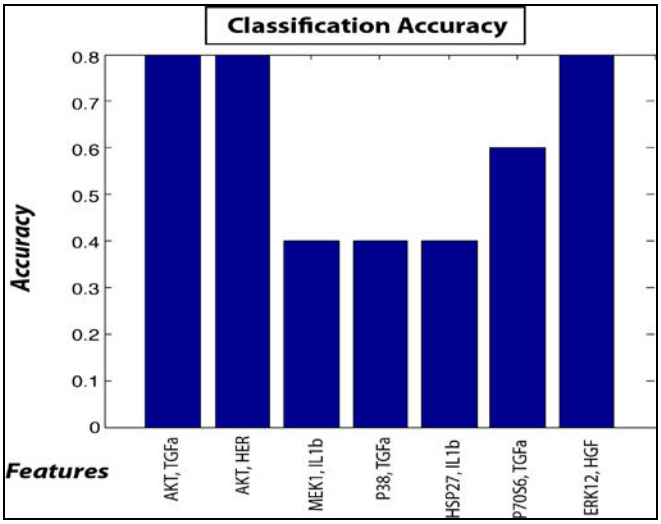

Figure 2. Classification accuracy for different phosphoprotein features. The y-axis corresponds to classification accuracy (out of 1.0), the x-axis corresponds to the different phosphoprotein features.

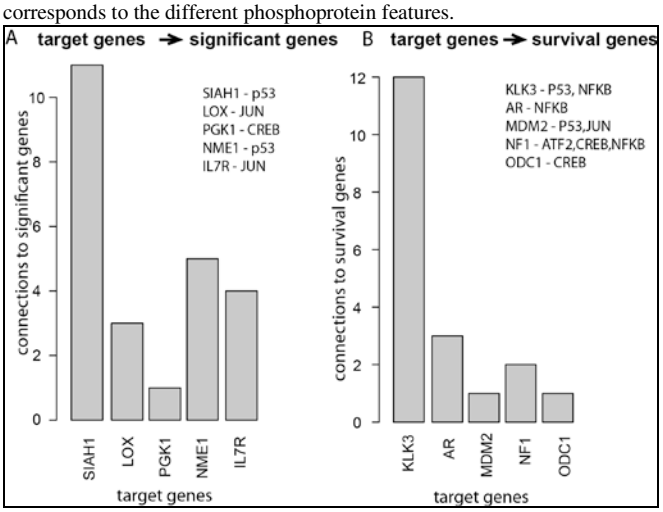

Figure 3. Correlation of the TFs' target genes to (A) significant in HCC and (B) survival genes. The y-axis corresponds to the number of connections to (A) significant in HCC or (B) survival genes, the x-axis corresponds to the most highly connected target genes.

V. Acknowledgements

We thank Katerina Chairakaki for help with the experimental procedure. Work by LGA and INM was supported by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program Education and Lifelong Learning of the National Strategic Reference Framework (NSRF) - Research Funding Program: ERC Grant Schemes. Work by DAL was supported by NIH grants U54-CA119267 and R01-CA96504.

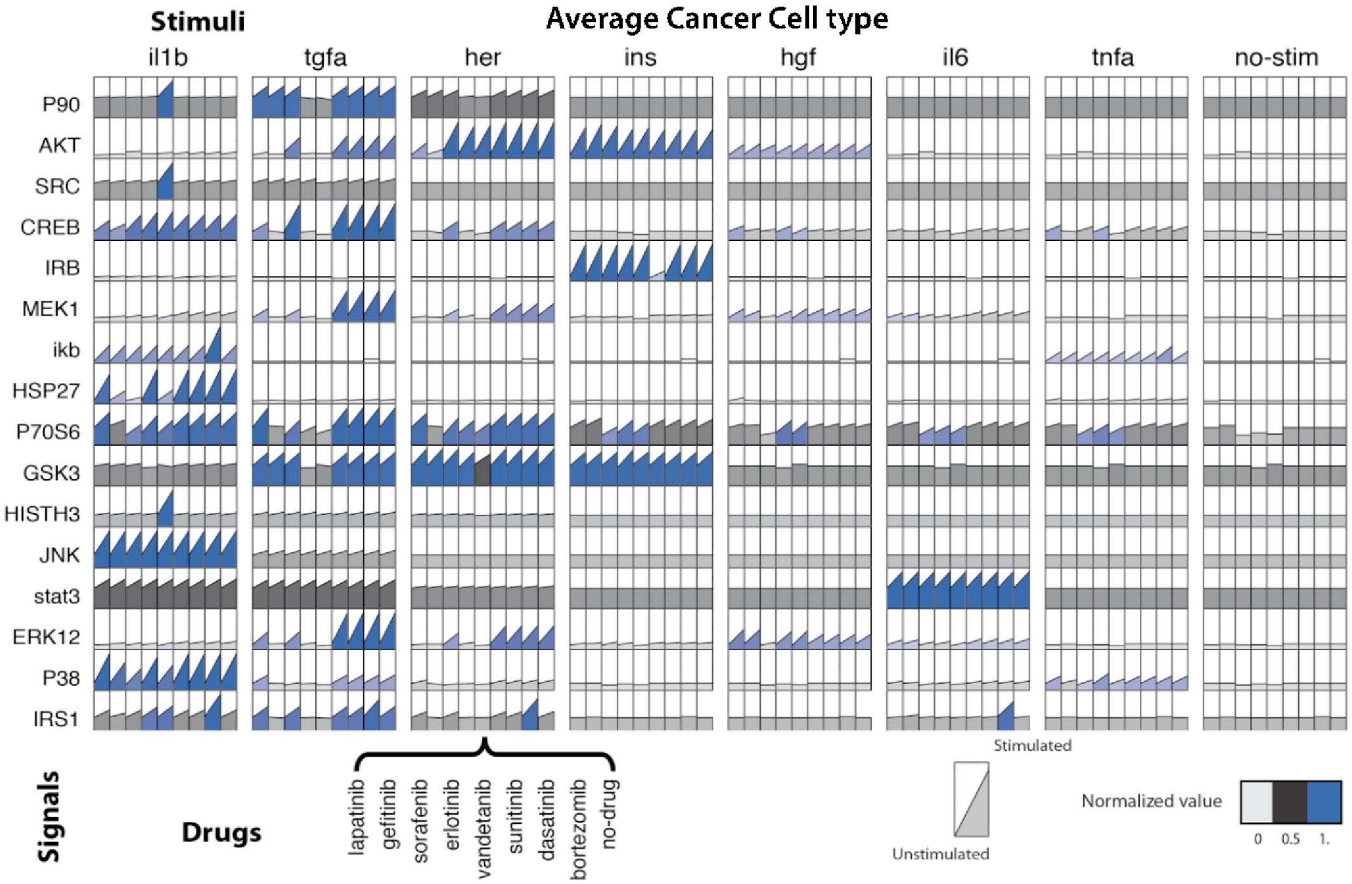

Figure 1. Phosphoproteomic data of the average cancer cell type under 8 stimuli (including the no-stimuli treatment) and 9 drugs for unresectable HCC (including the no-drug treatment). The time course of the 16 phosphoprotein signals from the unstimulated state to the average early response is illustrated. The rows correspond to the 16 signals, the main columns to the 8 stimuli treatments and the 9 subcolumns to the drugs. In each subplot, the first point shows the unstimulated activity of the respective signal (zero time point) and the second point shows the normalized value of the signal 5+25 minutes after stimulation.

VI. References

[1] http://clinicaltrials.gov, 2012.

[2] A. Jemal, "Global cancer statistics." CA: A Cancer Journal for Clinicians, 61(2):69–90, 2011.

[3] RC Alves, "Advanced hepatocellular carcinoma. review of targeted molecular drugs." Annals of Hepatology, 10(1):21–27, 2011.

[4] J. M. Llovet, "Sorafenib in advanced hepatocellular carcinoma." New England Journal of Medicine, 359(4):378–390, 2008.

[5] M. B. Thomas, "Phase 2 study of erlotinib in patients with unresectable hepatocellular carcinoma." Cancer, 110(5):1059–1067, 2007.

[6] J. Lee, "Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling." Hepatology, 40(3):667–676, 2004.

[7] X. Chen, "Gene expression patterns in human liver cancers." Molecular Biology of the Cell, 13(6):1929–1939, 2002.

[8] W. Sun, "Proteome analysis of hepatocellular carcinoma by two-dimensional difference gel electrophoresis." Molecular & Cellular Proteomics, 6(10):1798–1808, October 2007.

[9] S Whittaker, "The role of signaling pathways in the development and treatment of hepatocellular carcinoma." Oncogene, 29:4989–5005, 2010.

[10] A Mitsos, "Identifying drug effects via pathway alterations using an integer linear programming optimization formulation on phosphoproteomic data." PLoS Comput Biol, 5(12):e1000591, 12 2009.

[11] R Ramanathan, "A phase ii study of lapatinib in patients with advanced biliary tree and hepatocellular cancer." Cancer Chemotherapy and Pharmacology, 64:777–783, 2009. 10.1007/s00280-009-0927-7.

[12] P.J O'Dwyer, "Gefitinib in advanced unresectable hepatocellular carcinoma: Results from the eastern cooperative oncology group's study e1203." Journal of Clinical Oncology, 2006 ASCO Annual Meeting Proceedings Part I, 24, 2006.

[13] C Hsu, "Vandetanib in patients with inoperable hepatocellular carcinoma: A phase ii, randomized, double-blind, placebo-controlled study." Journal of Hepatology, 56(5):1097 – 1103, 2012.

[14] A. Cheng, "Phase iii trial of sunitinib (su) versus sorafenib (so) in advanced hepatocellular carcinoma (hcc)." Journal of Clinical Oncology, 29, 2011.

[15] D. Baiz, "Bortezomib arrests the proliferation of hepatocellular carcinoma cells hepg2 and jhh6 by differentially affecting e2f1, p21 and p27 levels." Biochimie, 91(3):373 – 382, 2009.

[16] D.C. Clarke, "Normalization and Statistical Analysis of Multiplexed Bead-based Immunoassay Data Using Mixed-effects Modeling", Mol Cell Proteomics. 2013 January; 12(1): 245–262

[17] A. Margolin, "Aracne: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context." BMC Bioinformatics, 7(Suppl 1):S7, 2006.

[18] Y. Ogawara, "Akt enhances mdm2-mediated ubiqui- tination and degradation of p53." Journal of Biological Chemistry, 277(24):21843–21850, 2002.

[19] R. Gedaly, "Pi-103 and sorafenib inhibit hepatocellular carcinoma cell proliferation by blocking ras/raf/mapk and pi3k/akt/mtor pathways." Anticancer Research, 30(12):4951–4958, December 2010.

[20] S. Seton-Rogers. "Hepatocellular carcinoma: Putting p53 in context." Nat Rev Cancer, 6:423, 2006.

[21] K. Matsuo, "Siah1 inactivation correlates with tumor progression in hepatocellular carcinomas." Genes, Chromosomes and Cancer, 36(3):283–291, 2003.

[22] S. Qu, "Genetic polymorphisms of metastasis suppressor gene nme1 and breast cancer survival." Clinical Cancer Research, 14(15):4787–4793, 2008.