

RESEARCH HIGHLIGHT

# Molecular fingerprinting catches responders to therapeutic agents

Simeen Malik<sup>1\*</sup>, Patrick Tan<sup>1,2,3</sup> and Pauline C Ng<sup>4</sup>

See related Research, <http://genomebiology.com/2013/14/10/R110>

## Abstract

High-dimensional -omics profiling predicts responses to therapeutic agents in breast cancer cell lines that can be effectively applied to patient selection in clinical trials.

## Introduction

Breast cancer is the leading cancer among women and the second leading cause of cancer-related mortality in women. It is a heterogeneous disease with distinct histological and clinical outcomes and classifies into three basic therapeutic groups: (1) estrogen receptor (ER) positive, (2) human epidermal growth factor receptor 2 (HER2) amplified and (3) triple-negative breast cancers (TNBCs, which are negative for ER, progesterone receptor (PR) and HER2). This gave rise to the notion of tailoring patient treatment plans based on the patient's genomic characteristics, making breast cancer the poster child for precision medicine. In this issue of *Genome Biology*, Joe Gray and colleagues develop this concept of precision medicine for breast cancer using a machine learning approach to computer modeling [1].

Seminal work in techniques for gene expression profiling analysis has led to the classification of breast cancer into six different subtypes (luminal A, luminal B, HER2-enriched, basal-like, claudin-low and normal-like) [2,3]. Multi-gene supervised class-predictors based on the six subtypes were subsequently developed for prognostic classification [2,3]. The predictors are now clinically available (examples include PAM50<sup>®</sup>, Oncotype DX<sup>®</sup>, MammaPrint<sup>®</sup>, MapQuantDx<sup>®</sup>, Theros<sup>®</sup> and Endopredict<sup>®</sup>). These first-generation gene signatures can also be partially used as predictive gene signatures and have been instrumental in

sparing a subgroup of ER-positive breast cancer patients from adjuvant cytotoxic chemotherapy [4,5]. However, there has been less success in the development of gene signatures to predict response to specific therapeutic agents, and as such no commercially available tests are currently available.

Other genomic technologies, such as genome copy number, next-generation sequencing, DNA methylation arrays, RNA-seq and protein expression, add higher resolution and additional layers of information. The integration of these datasets has further refined breast cancer classification into additional distinct subtypes. Large-scale genomic efforts of The Cancer Genome Atlas (TCGA) [6] and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) [7] have comprehensively characterized key genomic changes in a large dataset of breast cancers but did not relate these omic characteristics to drug response.

## Identifying predictive biomarkers for targeted therapeutics

Application of integrated molecular datasets to develop clinically useful predictive gene signatures to guide precision medicine is currently under way [5]. However, there are three noteworthy studies that generated predictive multigene signatures. One study used a multifactorial approach to identify a multi-gene classifier as predictive for response to anthracyclines. The resulting anthracycline-based score (A-score) could serve as a valuable tool for sparing a subset of breast cancer patients from chemotherapy [4,5,8]. Similarly, the amplification of chromosome 8q22 and/or overexpression of YWHAZ/LAPTM4B can be used as predictive biomarkers to anthracycline response [5]. A 165-gene signature (termed the SET index) has also been developed to predict response to endocrine therapy [4,5].

Cancer cell lines that have molecular characteristics that closely mirror their tumor counterparts have proved useful as preclinical models for therapeutic response and

\* Correspondence: [simeen.malik@duke-nus.edu.sg](mailto:simeen.malik@duke-nus.edu.sg)

<sup>1</sup>Cancer and Stem Cell Biology Program, Duke-NUS Graduate Medical School, 8 College Road, Singapore 169857, Singapore

Full list of author information is available at the end of the article

predictive biomarker development for experimental drugs [9,10]. Sometimes, cancer cell lines are the only possible model to investigate experimental therapeutics. Two recent studies performed large-scale unbiased drug screens in collections of cancer cell lines across many types of cancer [9,10] and have substantially expanded the annotation of these cell line collections and our understanding of the predictive usefulness of gene signatures.

It is unclear whether the recently available molecular data types are essential and which combinations of these data types would provide the best predictors for breast cancer. In this issue of *Genome Biology*, Gray and colleagues address this question by developing disease-specific predictive signatures of response for a collection of breast cancer cell lines [1]. Their study expands the number of breast cancer-specific cell lines and therapeutic agents used as well as collecting comprehensive molecular data types: genomic data were obtained from 70 breast cancer cell lines in response to 90 experimental or approved therapeutic agents. The molecular profiling datasets included mutation status of selected genes of interest, copy number aberrations, gene expression (including splice variants), promoter DNA methylation and protein expression. In addition to identifying predictive markers of response and testing their performance in TCGA samples, the authors also examined the importance of specific and integrated datasets for response predictor development.

### No magic bullets

Gray and colleagues used two machine-learning approaches, the weighted Least Squares Support Vector Machine and Random Forests, to develop response signatures [1]. Cell lines were classified as sensitive or resistant based on the mean  $GI_{50}$  value for that compound (that is, the concentration at which growth is inhibited by 50%). Regardless of classification method, the signatures predicted response with high estimated accuracy (area under curve (AUC) >0.70) for 57% of the compounds tested. The two classification methods showed high correlation (Spearman correlation coefficient, 0.85;  $P$ -value, <0.001), especially for compounds that had strong biomarkers or no biomarkers of response. However, this did not hold for compounds that showed a weaker signal of drug response - for these compounds, selection of alternative response variables might improve performance in the future.

Cell line-derived response signatures were validated *in vivo* for tamoxifen and valproic acid, where the -omic signatures accurately predicted the chemotherapy response. For the compounds where *in vivo* data were not available, the presence of the signature was tested for in 536 breast TCGA tumors instead. Tumors and their

respective cell lines shared similar gene expression patterns, suggesting that the signatures might be effective in predicting response in patient samples. Many of these compounds are FDA approved or in clinical trials and hence could be validated in the near future.

Another important finding was that there was no single molecular dataset that always outperformed the rest. However, in general, RNA-seq performed better than the other data-types across all the compounds, and copy number array data performed worse. The basis for RNA-seq's superior performance in most cases was its improved sensitivity and dynamic range, and not its detection of splice sites. Splice-site awareness was only beneficial for prediction of response to ERBB2-targeting compounds.

The transcriptional dataset alone was sufficient to predict responses (AUC >0.70) in 25% of compounds. This is currently being done for ER<sup>+</sup> and HER2<sup>+</sup> breast cancers when guiding the selective use of chemotherapy, as described previously. Addition of other molecular datasets was able to significantly increase prediction for 65% of compounds. The mutation status of genes encoding TP53, PIK3CA, MLL3, CDH1, MAP2K4, PTEN and NCOR1 was primarily useful for predicting response to tamoxifen and the polyamine analog CGC-11144. This suggests that, for the majority of compounds, a combinatorial approach involving multiple molecular datasets (although not all) would prove beneficial. Validation studies in clinical trials will help clarify which combinations of data-types are most useful and whether there are similarities in response to certain classes of compounds. For example, the authors found that RNA-seq performed better for polyamine analogs and mitotic inhibitors, copy number array was better for inhibitors of ERBB2/EGFR, and DNA methylation profiling was best for inhibitors of CDK1.

There are a few caveats to this study. Although cell lines are good models for developing predictive drug signatures, they have certain drawbacks that limit their ability to recapitulate the primary tumors. Specifically, they disregard the molecular heterogeneity inherent to breast cancer and any associated influences of the microenvironment [4,5]. Most of the cell lines are epithelial in origin, and do not include stromal and immune components that are known to be important contributors to malignant progression [4,5]. Variations in oxygen content that are known to affect therapeutic responses are also not addressed by cell line models. Performing similar experiments in three-dimensional model systems or in mouse xenografts would help address some of these caveats and further refine the signatures. Furthermore, the clonal-evolutionary dynamics of the tumor are also largely unaddressed. Patient-specific molecular biomarker signatures could be developed by serial molecular monitoring

of disease progression in response to therapeutic agents. This could be especially useful in patients that fail to respond to the physician's first choice of drug.

### Future directions

The holy grail of precision medicine is matching the right drug to the patient. For new drugs in development, the inverse of finding patients most likely to respond to an experimental compound is also equally important for those clinical trials driven by -omics data. While both of these are momentous tasks, compounds can be ranked based on their predicted efficacies in individual patients and validated in prospective clinical trials. By creating a publicly available software tool that can predict drug response in individual tumors, the authors have taken us one step closer to the promise of precision medicine [1]. They applied this tool to 306 TCGA samples for which expression, copy number and DNA methylation data were available. Almost all patients (99.3%) had received at least one compound that they were predicted to respond to, and each patient was predicted to respond to an average of approximately six treatments. A future application of this tool could be to assign approved or experimental agents to individual patients in the clinic or biomarker-guided clinical trials, respectively.

Accurate predictors of response can be developed for compounds that have a strong associated molecular signature. For these compounds, combinatorial approaches involving multiple platforms are not necessary. These compounds are also the best candidates for transition to a single-platform lab diagnostic. For compounds with a weaker signal of drug response, adding additional platforms or identifying alternative response variables might improve efficacy.

As many of the compounds tested by Gray and colleagues are in clinical trials, further validation of their identified molecular signatures is likely to be imminent. A logical way forward for validating these signatures for approved therapeutic agents is a direct comparison of patient outcome when the drug selected is the physician's first treatment choice versus the top drug from the *in vitro* predictor tool. Validation of predictive biomarkers requires large multi-center randomized trials that are logistically challenging and expensive. This study is a good basis for such a multi-arm trial, where the simultaneous testing of a panel of compounds could accelerate the validation of the *in vitro* signatures. Optimizing the molecular features of the signature as well as the thresholds for tumor classification would also be possible in such a clinical trial setting. By developing the next generation of predictive biomarkers, Gray and colleagues' study, together with other research like it, gets us one step closer to precision medicine.

### Author details

<sup>1</sup>Cancer and Stem Cell Biology Program, Duke-NUS Graduate Medical School, 8 College Road, Singapore 169857, Singapore. <sup>2</sup>Infectious Diseases, Genome Institute of Singapore, 60 Biopolis Street, Genome, #02-01, Singapore 138672, Singapore. <sup>3</sup>Cancer Science Institute of Singapore, Yong Loo Lin School of Medicine, 14 Medical Drive, #12-01, Singapore 117599, Singapore. <sup>4</sup>Computational and Mathematical Biology, Genome Institute of Singapore, 60 Biopolis Street, Genome, #02-01, Singapore 138672, Singapore.

Published: 31 October 2013

### References

1. Daemen A, Griffith OL, Heiser LM, Wang NJ, Enach OM, Sanborn Z, Pepin F, Durinck S, Korkola JE, Griffith M, Hur JS, Huh N, Chung J, Cope L, Fackler MJ, Umbricht C, Sukumar S, Seth P, Sukhatme VP, Jakkula LR, Lu Y, Mills GB, Cho RJ, Collisson EA, Van't Veer LJ, Spellman PT, Gray JW: **Modeling precision treatment of breast cancer.** *Genome Biol* 2013, **14**:R110.
2. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lønning PE, Børresen-Dale AL: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci U S A* 2001, **98**:10869–10874.
3. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**:747–752.
4. Reis-Filho JS, Pusztai L: **Gene expression profiling in breast cancer: classification, prognostication, and prediction.** *Lancet* 2011, **378**:1812–1823.
5. Zardavas D, Pugliano L, Piccart M: **Personalized therapy for breast cancer: a dream or a reality?** *Future Oncol* 2013, **9**:1105–1119.
6. The Cancer Genome Atlas Network: **Comprehensive molecular portraits of human breast tumours.** *Nature* 2012, **490**:61–70.
7. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, METABRIC Group, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, et al: **The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.** *Nature* 2012, **486**:346–352.
8. Desmedt C, Di Leo A, de Azambuja E, Larsimont D, Haibe-Kains B, Selleslags J, Delaloge S, Duhem C, Kains JP, Carly B, Maervoet M, Vindevoghel A, Rouas G, Lallemand F, Durbecq V, Cardoso F, Salgado R, Rovere R, Bontempi G, Michiels S, Buyse M, Nogaret JM, Qi Y, Symmans F, Pusztai L, D'Hondt V, Piccart-Gebhart M, Sotiriou C: **Multifactorial approach to predicting resistance to anthracyclines.** *J Clin Oncol* 2011, **29**:1578–1586.
9. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jané-Valbuena J, Mapa FA, Thibault J, Bric-Furlong E, Raman P, Shipway A, Engels IH, Cheng J, Yu GK, Yu J, Aspesi P Jr, de Silva M, et al: **The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.** *Nature* 2012, **483**:603–607.
10. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J, Liu Q, Iorio F, Surdez D, Chen L, Milano RJ, Bignell GR, Tam AT, Davies H, Stevenson JA, Barthorpe S, Lutz SR, Kogera F, Lawrence K, McLaren-Douglas A, Mitropoulos X, Mironenko T, Thi H, Richardson L, Zhou W, Jewitt F, et al: **Systematic identification of genomic markers of drug sensitivity in cancer cells.** *Nature* 2012, **483**:570–575.

doi:10.1186/gb-2013-14-10-135

Cite this article as: Malik et al.: Molecular fingerprinting catches responders to therapeutic agents. *Genome Biology* 2013 **14**:135.