Abstract # B27

OncoPredictor: High copy number gain of ERBB2 occurs in 1% of lung cancer patients and predicts sensitivity to lapatinib in lung cancer cell lines

Sean F. Eddy, Mark Tomilo, Seth Sadis, Paul D. Williams, Peter Wyngaard, Daniel R. Rhodes
Compendia Bioscience, Ann Arbor, MI, USA

Abstract:

Investigational new drugs for cancer must demonstrate convincing preclinical efficacy and a compelling strategy to translate preclinical observations to the clinical setting. Predictive biomarkers are gaining wider acceptance (i.e., HercepTest™, Oncotype DX® Breast Cancer Assay), however, there is no systematic approach to identify candidate predictive biomarkers for new therapies and to characterize their distribution and expression in cancer patient subsets. To meet this challenge, we developed an integrated platform to translate preclinical biomarker profiles, derived from screening compounds across large panels of cancer cell lines, to clinical tumor populations. This platform, known as OncoPredictor, consists of two parallel genomic databases that incorporate DNA copy number, mutation status, and gene expression biomarker annotations for hundreds of cell lines and the same biomarkers annotated across thousands of clinical tumor samples. To illustrate this approach, we used published drug response data for lapatinib (Greshock *et al.*, Cancer Res., 70: 3677-86, 2010) and investigated candidate biomarkers representing each of the genomic data types (mutation, DNA amplification and deletion, and gene over-expression) in an unbiased approach to identify biomarkers that associated with drug sensitivity in cell lines. Of all genomic biomarkers, high copy number gain of ERBB2 was the most significant biomarker that associated with sensitivity to lapatinib. As expected, several breast cell lines showed high level amplification of ERBB2 and were sensitive to lapatinib, consistent with approved indications for ERBB2-based therapies. Interestingly, two lung cancer cell lines (CALU3 and NCIH2170) were also exquisitely sensitive to lapatinib and both contained significant amplification of ERBB2. Using DNA copy data from >1,000 lung cancer tumor specimens, we estimated that approximately 0.5–1% of lung squamous cell carcinoma and lung adenocarcinoma contained high level copy number gain of ERBB2. Similar to observations in breast cancer, ERBB2 gains were typically focal (5–20 genes) and ERBB2 was the most commonly amplified gene in this region. Increased copy number was also significantly correlated with ERBB2 mRNA expression. Because KRAS mutations cause resistance to EGFR family targeting drugs, we investigated the clinical frequency of KRAS mutations in lung cancer. In >15,000 patient samples with available mutation data, approximately 17% of patient samples had KRAS mutations, including 20% of lung adenocarcinoma and 5% of squamous cell lung carcinoma samples. Infrequent high level amplification of ERBB2 in lung cancer coupled with the frequency of KRAS mutation, may account for apparent lack of efficacy in lung cancer patient populations. The cell line data support the use of high level ERBB2 amplification as a predictive biomarker for clinical response to lapatinib, which may lead to improved patient survival for a small fraction of patients with lung cancer.

Background and Rationale:

Cancer is a set of genetic diseases. Rational development of cancer therapeutics and patient treatment requires an understanding of the genetic biomarkers associated with drug response. New approaches that efficiently identify predictive biomarkers for targeted cancer therapies and place them into the correct clinical context are needed. Large cancer cell line panels are being increasingly used to discover predictive biomarkers of drug response. Large cell line panels generally preserve the genetic aberrations and molecular diversity unique to parent histology from which they were derived and can effectively recapitulate tumor associated genotypes that predict sensitivity to various kinase inhibitors. Here we demonstrate the use of parallel cell line and patient tumor databases for translating the findings that high copy number gain and focal amplification of ERBB2 sensitizes cell lines to lapatinib and identify appropriate and rare patient populations in lung cancer. Trastuzumab was previously considered ineffective in a lung cancer patient cohort selected for ERBB2 expression by IHC and FISH (Gatzemeier *et al.*, 2004), possibly due to over-estimation of ERBB2-driving aberrations in lung cancer. IHC-based expression classification estimates ERBB2 expression in non-small cell lung cancer in the range of 20–50% (Langer *et al.*, 2004). While IHC captures ERBB2 expression in many lung cancer patients, its use may not accurately identify cancer-driving aberrations. Use of a combined IHC and FISH testing that estimated 2% of non-small cell lung cancers have amplified ERBB2 (Heinmöller *et al.*, 2003), and subsets of patients with IHC3+ or FISH positive ERBB2 status, have been noted as showing a trend of increased progression free survival in response to trastuzumab (Gatzemeier *et al.*, 2004). Additionally, a report of a patient with an ERBB2 amplified tumor treated with lapatinib showed an unconfirmed decrease in tumor size (Ross *et al.*, 2010). Even within clinically selected ERBB2 patients, there likely remains a low frequency of ERBB2 driver amplifications in lung cancer as noted by the infrequent occurrence of high copy gain of ERBB2. This, coupled with cell response data to ERBB2 targeting therapies suggests a correctly selected subset of patients may be amenable to anti-ERBB2 therapy. Specifically, patients with high copy gain, focal amplification, and high over-expression of ERBB2 may be those that derive the most benefit from ERBB2 targeted therapies in lung cancer. While ERBB2-driven lung cancers are relatively rare, they represent a well-defined patient population that could be treated successfully with current approved therapies.

Methods:

Cell line profiling data. Drug response data for 228 cell lines from Greshock *et al.* (2010 Cancer Research 70:3677–86) was incorporated into Oncomine. Cell line sensitivity status was determined through visual inspection of the IC50 distributions.

Categorical biomarker analysis. To develop the cell line biomarker database, we collected genomic data from public sources that corresponded to the set of cell lines in the panel described by Greshock *et al.* (2010). Targeted mutation data was acquired from the Wellcome Trust Sanger Institute's Cancer Cell Line Project (Hinxton, U.K), which has sequenced the coding exons and immediate flanking intron sequences of 51 cancer-related genes in 700+ cell lines. DNA copy number data for 338 and gene expression profiling data for 318 cell lines was obtained from a dataset made public at caBIG (Bethesda, MD) by GlaxoSmithKline. Cell line names were standardized using terms provided by the source repositories, and a semi-automated text matching strategy was employed to match cell line names. Next, we developed a series of analysis pipelines to make binary (i.e., positive/negative) biomarker status calls for genes and pathways across the cell line panel. The mutation data was binary to begin with, as each cell line was called mutant or wild-type for each of the cancer-related genes tested. Cell lines were called positive for gene over-expression if a gene's expression level was greater than 64-fold (6 logs) above the gene's median expression level across all cell lines. Cell lines were called positive for a gene amplification or deletion if the gene's estimated copy number was at least two-fold above or below the gene's median copy number across all cell lines; this approximates greater than 4 copies for amplification calls and less than 1 copy for deletion calls, assuming a 2 copy baseline. In addition to these gene-level candidate biomarkers, we also derived biomarkers for cancer pathways. Cell lines were called positive for a pathway biomarker if they had an aberration in any gene in the pathway and negative if all genes in the pathway were negative. Pathway biomarkers combined different genes in the same pathway by considering mutations only, gene over-expression only, and a combined mutation/amplification/deletion annotation. To identify biomarkers predictive of drug response, we tested each biomarker from the cell line biomarker database for association with response to each compound by comparing the number of cell lines positive or negative for a biomarker that were also sensitive or resistant to the compound. Odds ratios and p-values were computed for each association and multiple hypothesis correction was performed using a false discovery rate approach.

Clinical Tumor Biomarker Database. To predict tumor populations likely to respond to a given agent based on predictive biomarkers observed in vitro, we required the ability to compute frequencies of the same biomarkers across all cancer types and subtypes (Fig. 1C). Thus, in the development of the clinical tumor biomarker database, we attempted to represent all of the candidate biomarkers from the cell line biomarker database in the largest collection of clinical tumor samples possible. Mutation data was downloaded from Sanger's COSMIC database and supplemented with data from 5 additional large-scale gene sequencing studies, including data from TCGA. Tumor samples were assigned controlled vocabulary terms describing general and specific cancer types. DNA copy number and gene expression data were collected from our Oncomine database, which contains genomic data from 500+ microarray studies, including several large scale initiatives such as the TCGA, the International Genome Consortium Expression Project for Oncology (expO), and Tumorscape. In order to compute biomarker frequencies, thresholds were applied to the quantitative DNA copy number and gene expression data such that each tumor sample was called positive or negative for each aberration. The DNA copy number data was processed to derive estimated copy number for each gene in each sample and samples were called positive for a gene amplification or deletion if the gene's copy number was two-fold above or below the gene's median copy number in a given dataset. Gene expression biomarkers were computed from a subset of the Oncomine database, comprising 25,647 clinical tumor specimens from 215 primary datasets analyzed on the Affymetrix HG-U133A or HG-U133 Plus 2.0 platforms. Samples were called positive for gene over-expression if the gene's expression level was 10-fold above the gene's median expression level across all samples. Finally, pathway biomarker statuses were computed similarly to the cell line biomarker database.

Results:

Figure 1. Analysis schematic.

(A) Cancer cell lines were tested for response to targeted agents including lapatinib and IC50 values were derived (Greshock *et al.*, 2010). Cutoffs were determined to establish sensitive and resistant sets of cell lines for each compound.

(B) Publicly available gene expression, copy number, and mutation data was collected for each cell line. KEGG was used to define genes involved in individual pathways, and a new pathway biomarker was created by rolling-up all the individual gene annotations into a single pathway-level annotation. Associations between cell lines with sensitive (or resistant) calls with each of the binary biomarkers were tested for using association analysis.

(C) Similarly, the same types of genomic data (gene expression, mutation, copy number, pathway) were collected and integrated for the clinical populations database. A controlled vocabulary for both general cancer type and detailed cancer type was implemented, allowing frequencies and detailed analysis (such as focal amplification) of genomic biomarkers to be associated with different disease terms. In a final step, biomarkers identified in the cell line drug response screen were assessed in the clinical populations database.

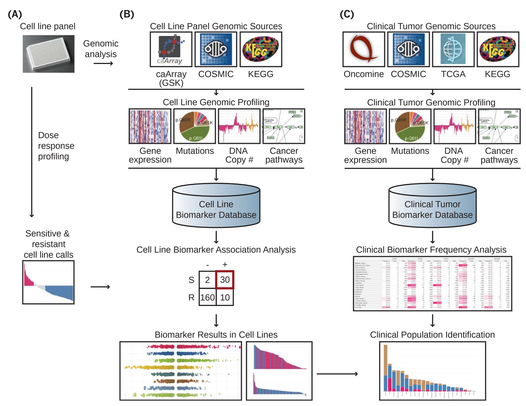

Figure 2. IC50 box plots for compounds analyzed in the Greshock *et al.* (2010) study.

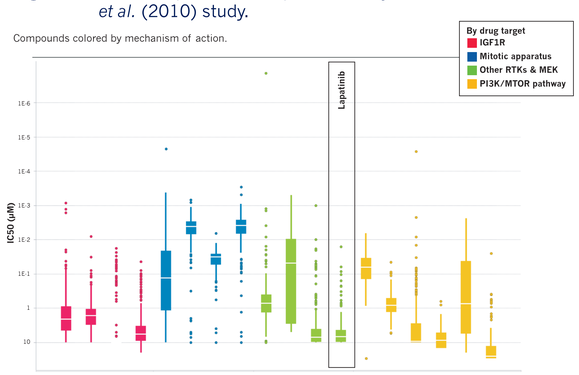

Figure 3. Volcano plot of biomarkers colored by associated drug.

Each spot on the plot represents a single biomarker and its corresponding p-value and Odds Ratio. Biomarkers Points were sized according to the –log(Q-value), the inverse log of false discovery correction. Resistance biomarkers were given a negative Odds Ratio value for plotting. The dashed red line indicates a p-value of 0.05. Amongst all biomarkers, ERBB2 amplification had the largest Odds Ratio and was highly significant by p-value and false discovery correction.

Figure 4. ERBB2 amplification and over-expression associates with sensitivity to lapatinib in breast cancer cell lines.

(A) Breast cancer cell line sensitivity plot. Breast cancer cell lines ranked by increasing lapatinib IC50 values and colored by sensitivity status. Sensitive cell lines were colored pink, resistant cell lines were colored green, and a moderately sensitive group of cell lines were colored gray.

(B) ERBB2 copy number in breast cancer cell lines. Breast cancer cell lines ranked by increasing lapatinib IC50 values and colored by ERBB2 DNA copy number. Cell lines colored in gradients of red indicate amplification of ERBB2 ranging from 8–12+ copies, cell lines colored in gradients of blue indicate copy number loss, while those colored light gray indicate 2 copies of ERBB2. Cell lines lacking copy number data were colored dark gray.

(C) ERBB2 expression in breast cancer cell lines. Breast cancer cell lines ranked by increasing lapatinib IC50 values and colored by Log2 ERBB2 expression. Cell lines colored red indicate highest expression of ERBB2, cell lines colored blue indicate lowest expression of ERBB2, while those colored light gray indicate median expression of ERBB2. Cell lines lacking transcriptional data were colored dark gray.

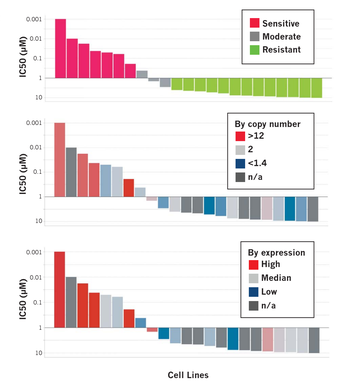

Figure 5. High copy number gain of ERBB2 was found in two lung cancer cell lines sensitive to lapatinib.

(A) Lung cancer cell line sensitivity plot. Lung cancer cell lines ranked by increasing lapatinib IC50 values and colored by sensitivity status. Sensitive cell lines were colored pink, resistant cell lines were colored green, and a moderately sensitive group of cell lines were colored gray.

(B) ERBB2 copy number in lung cancer cell lines. Lung cancer cell lines ranked by increasing lapatinib IC50 values and colored by ERBB2 DNA copy number. Cell lines colored in gradients of red indicate amplification of ERBB2 ranging from 11–13+ copies, cell lines colored in gradients of blue indicate copy number loss, while those colored light gray indicate 2 copies of ERBB2. Cell lines lacking copy number data were colored dark gray.

(C) ERBB2 expression in lung cancer cell lines. Lung cancer cell lines ranked by increasing lapatinib IC50 values and colored by Log2 ERBB2 expression. Cell lines colored red indicate highest expression of ERBB2, cell lines colored blue indicate lowest expression of ERBB2, while those colored light gray indicate median expression of ERBB2. Cell lines lacking transcriptional data were colored dark gray.

Figure 6. Copy number gain of ERBB2 in breast, gastric and lung cancers.

(A) A subset of lung cancer patients have estimated copy gains in ERBB2 comparable to estimated copy number of ERBB2 in breast and gastric cancer patients, cancer types for which ERBB2 targeted therapy has been approved.

(B) Expression levels of ERBB2 in a subset of lung cancer patients was comparable to ERBB2 expression in breast and gastric cancer patients, cancer types for which ERBB2 targeted therapy has been approved. Expression data for genes measured on Affymetrix U133 (U133A, U133 plus 2.0) arrays across 25,647 patient samples from all major cancer types was used to for normalization of gene expression. Fold change over the median expression of ERBB2 across all cancer types is reported.

(C) ERBB2 mRNA expression was correlated to ERBB2 DNA copy number. Utilizing the TCGA Lung datasets, 166 unique lung cancer patient samples were investigated for the correlation of mRNA expression with DNA copy number.

Table 1. Estimated copy number gain of ERBB2 in lung cancer subtypes.

1,051 lung cancer patient tumors were analyzed for copy number gains in ERBB2. Frequency of amplification was reported at increasing thresholds of estimated copy gain. Data was obtained from a meta-analysis of Oncomine datasets, Bass Lung, Beroukhim Multi-cancer, Jaiswal Mutli-cancer, Lu Lung, Ramos Lung, TCGA Lung 2, and Weir Lung.

	Estimated copies of ERBB2			
Cancer Type	>4	>6	>8	Patient Samples
Non-Small Cell Lung Carcinoma, NOS	1.4%	1.2%	0.5%	417
Lung Adenocarcinoma	1.2%	1.1%	0.5%	434
Squamous Cell Lung Carcimona	1.1%	1.1%	0.6%	174
Large Cell Lung Carcinoma	0.0%	0.0%	0.0%	9
Small Cell Lung Carcinoma	0.0%	0.0%	0.0%	17

Figure 7. ERBB2 amplification was focal in lung cancers.

The graph depicts ERBB2 amplification in lung cancer patients. The pink line indicates aggregate copy number gain of individual genes with at least 4 copies, while the blue line indicates number of patients with at least 4 copies of a subset of genes on chromosome 17q. An aggregate copy number peak indicates a high and focal copy number gain of ERBB2 relative to other genes on chromosome 17q.

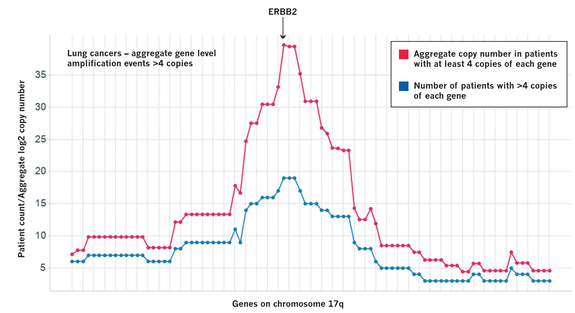

Table 2. Mutation frequency for KRAS across 15,245 lung cancer patient samples.

Data obtained from COSMIC v54 and Oncomine datasets, Ding Lung, Garraway Multi-cancer, Kan Multi-Cancer, and Kan Multi-cancer 2.

Cancer Type	KRAS Mutation Frequency	Count	Patient Samples
Lung Adenocarcinoma	20.9%	1,549	7,425
Large Cell Lung Carcinoma	17.5%	33	189
Lung Cancer, NOS	16.3%	251	1,537
Non-Small Cell Lung Cancer, NOS	15.7%	596	3,791
Adendosquamous Lung Carcinoma	10.6%	21	199
Squamous Cell Lung Carcinoma	5.8%	97	1,672
Lung Carcinoid Tumor	2.0%	3	148
Small Cell Lung Carcinoma	0.7%	2	294

Table 3. Incidence and mortality of lung adenocarcinoma and squamous cell lung carcinoma.

Data obtained from IARC GLOBOCAN 2008 (http://globocan.iarc.fr/) both sexes, all ages and rounded to the nearest hundred. The table assumes 40% of lung cancers are adenocarcinomas and 25% of lung cancers are squamous cell lung carcinomas (Travis WD, 2002).

	Incidence				Mortality			
Lung Cancer Subtype	World	European Union (EU-27)	USA	Japan	World	European Union (EU-27)	USA	Japan
Lung Adenocarcinoma	642,800	115,200	85,600	34,700	550,000	101,200	64,400	26,700
Squamous Cell Lung Carcinoma	401,700	72,000	53,500	21,700	343,700	63,200	40,200	16,700
Total Lung Cancers	1,607,000	288,000	214,000	86,800	1,375,000	253,000	161,000	67,000

Conclusions:

- ERBB2 amplification predicted sensitivity to lapatinib across a broad multi-cancer cell line panel.
- Two lung cancer cell lines with exquisite sensitivity to lapatinib both contained high copy number gain of ERBB2.
- Approximately 1% of patients with lung cancer have focal, high level copy number gain of ERBB2.
- The frequency of ERBB2 amplification was similar in lung adenocarcinoma and squamous cell lung carcinoma.
- Lung cancer patients with high copy number gain of ERBB2 may benefit from ERBB2 targeted therapy.

References:

Greshock J, Bachman KE, Degenhardt YY, Jing J, Wen YH, Eastman S, McNeil E, Moy C, Wegrzyn R, Auger K, Hardwicke MA, Wooster R. Molecular target class is predictive of in vitro response profile. Cancer Res. 2010 May 1;70(9):3677-86. Epub 2010 Apr 20.

Gatzemeier U, Groth G, Butts C, Van Zandwijk N, Shepherd F, Ardizzoni A, Barton C, Ghahramani P, Hirsh V. Randomized phase II trial of gemcitabine-cisplatin with or without trastuzumab in HER2-positive non-small-cell lung cancer. Ann Oncol. 2004 Jan;15(1):19-27.

Langer CJ, Stephenson P, Thor A, Vangel M, Johnson DH; Eastern Cooperative Oncology Group Study 2598. Trastuzumab in the treatment of advanced non-small-cell lung cancer: is there a role? Focus on Eastern Cooperative Oncology Group study 2598. J Clin Oncol. 2004 Apr 1;22(7):1180-7. Epub 2004 Feb 23.

Heinmöller P, Gross C, Beyser K, Schmidtgen C, Maass G, Pedrocchi M, Rüschoff J. HER2 status in non-small cell lung cancer: results from patient screening for enrollment to a phase II study of herceptin. Clin Cancer Res. 2003 Nov 1;9(14):5238-43.

Ross HJ, Blumenschein GR Jr, Aisner J, Damjanov N, Dowlati A, Garst J, Rigas JR, Smylie M, Hassani H, Allen KE, Leopold L, Zaks TZ, Shepherd FA. Randomized phase II multicenter trial of two schedules of lapatinib as first- or second-line monotherapy in patients with advanced or metastatic non-small cell lung cancer. Clin Cancer Res. 2010 Mar 15;16(6):1938-49. Epub 2010 Mar 9.

Travis WD. Pathology of lung cancer. Clin Chest Med. 2002 Mar;23(1):65-81, viii.