Predicting ligand-dependent tumors from multi-dimensional signaling features

Helge Hass[1,2], Kristina Masson[1], Sibylle Wohlgemuth[3], Violette Paragas[1], John E Allen[1], Mark Sevecka[1], Emily Pace[4], Jens Timmer[2,5], Joerg Stelling[3], Gavin MacBeath[1], Birgit Schoeberl[1], Andreas Raue[1,*]

[1]Merrimack Pharmaceuticals, Inc., Cambridge, MA 02139, USA; [2]Institute of Physics, University of Freiburg, Freiburg, Germany; [3]Department of Biosystems Science and Engineering and SIB Swiss Institute of Bioinformatics, ETH Zuerich, Zuerich, Switzerland; [4]Celgene, San Francisco, CA 94158, USA; [5]BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany

AACR #1312

Abstract

Receptor tyrosine kinases (RTKs) are high-affinity cell surface receptors for growth factors that are frequently deregulated in cancer. Signaling through these receptors has been associated with increased cancer cell proliferation and resistance to cytotoxic therapies. To block this detrimental signaling, many companies are developing inhibitory antibodies against various RTKs. A key challenge in clinical studies is the optimal stratification of patients who may benefit from these therapies. For an RTK targeted antibody, the detection of the respective growth factor in the tumor microenvironment may be an important bio-marker. Beyond the physical presence of the growth factor, the decision whether a cancer cell will respond to growth factor-induced signals is governed by complex intra-cellular signaling networks. We compared different approaches to predict cellular responses and will highlight a hybrid approach that combines mechanistic modeling based on ordinary differential equations with a machine learning algorithm. The models are trained on *in vitro* drug response screens and then applied to predict response in patient samples. The mechanistic models are trained on quantitative data from signal transduction studies as well as RNAseq data for cellular characterization. Using the hybrid approach, a correlation between growth factor expression in the tumor microenvironment and its predicted response was identified. This supports the hypothesis of addiction of tumors to growth factors abundant in the tumor microenvironment and might enable more robust patient stratification in the future.

In vitro proliferation screen

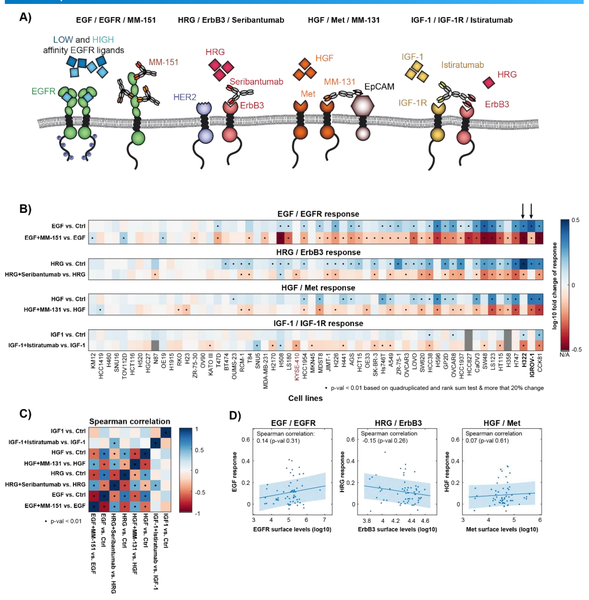

Figure 1. Proliferation screen across 58 cell lines. (a) Ligand/Receptor and antagonistic antibodies used in the *in vitro* proliferation screen. (b) Results of the proliferation screen across 58 cell lines. Dots mark a significant increase in ligand induced proliferation or decrease in the presence of ligand plus antibody. The ligand effect is normalized to the medium control, whereas the antibody plus ligand effect is relative to ligand alone. The two cell lines marked with an arrow, as well as five additional cell lines that were not included in the proliferation screen, were used to train the computational model to signaling data. (c) Correlation pattern of ligand and antibody effects across all cell lines. (d) Linear correlation of receptor expression to ligand induced proliferation. The proliferation in response to ligand (y-axis) is displayed as log10-fold change with respect to day 0. The receptor surface levels (x-axis) are absolute measurements of receptors/cell by qFACS on a log10-scale.

Strategies for predicting ligand induced growth

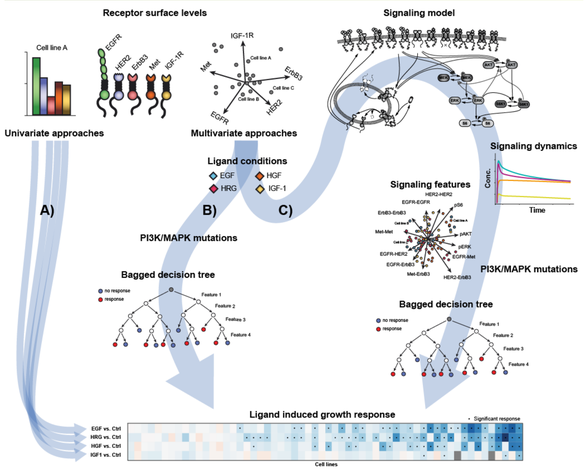

Figure 2. Based on the receptor expression of individual cancer cell lines, either a univariate or multivariate approach can be used to predict the phenotypic response to ligand stimulation. (a) Univariate approaches relate the respective receptor expression to the observed ligand induced proliferation for each of the four ligands separately. (b–c) Multivariate approaches such as bagged decision trees (BDTs) relate high-dimensional feature sets to the observed phenotype. (b) In this case the feature set consists of the five receptor surface levels as well as information about the respective ligand stimulation and mutation status. (c) The calibrated and validated signaling model allows to simulate the expected signaling dynamics for each individual cell line based on its receptor expression and ligands present. Based on the mechanistic knowledge that the signaling model incorporates, it can expand the initial five-dimensional feature set to a 12-dimensional feature set. This expanded feature set, together with information about mutation status is now connected to the observed growth responses by a bagged decision tree.

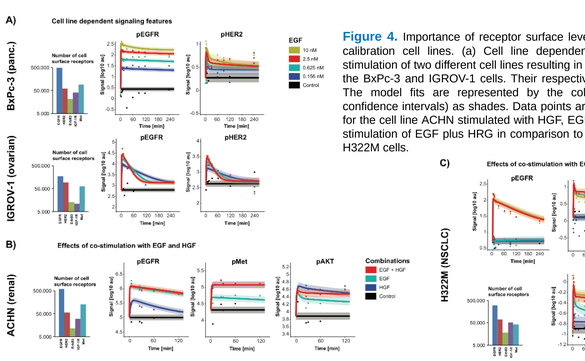

Figure 4. Importance of receptor surface levels for model response, shown for a selection of calibration cell lines. (a) Cell line dependent signaling features: Model response to EGF stimulation of two different cell lines resulting in sustained or transient receptor phosphorylation in the BxPc-3 and IGROV-1 cells. Their respective receptor surface levels are shown on the left. The model fits are represented by the colored lines with respective uncertainties (67% confidence intervals) as shades. Data points are shown as dots in the same color. (b) Model fits for the cell line ACHN stimulated with HGF, EGF and the combination. (c) Model response to co-stimulation of EGF plus HRG in comparison to the stimulation with EGF, HRG or IGF-1 alone in H322M cells.

Multi-pathway computational model

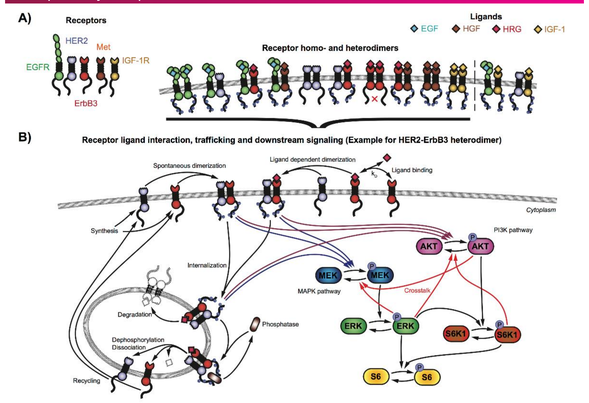

Figure 3. Structure of computational signaling model. (a) The receptors EGFR, HER2, ErbB3, Met, and IGF-1R can form several homo and heterodimers after ligand binding. (b) In the model, receptors are synthesized and either dimerize spontaneously or bind a ligand to form homo- and hetero-dimers, which results in trans-phosphorylation of the receptors. Activated receptors signal downstream and are prone for internalization, which leads to either degradation or dephosphorylation by a phosphatase followed by recycling to the cell surface. Downstream, the MAPK and PI3K cascade activate S6K1 and ultimately converge in the phosphorylation of S6. The MAPK and PI3K signaling pathways are interconnected via multiple crosstalk mechanisms.

Feature	BDT importance score
pS6 AUC	0.65
EGFR homodimerization	0.56
EGFR-HER2 heterodimerization	0.56
Met-ErbB3 heterodimerization	0.55
pAKT AUC	0.43
pERK AUC	0.37
PI3K mutation status	0.35
RAS mutation status	0.30

Table 1. Model features ranked by their BDT training efficiency

Application to patient data

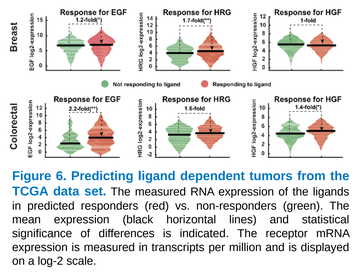

Figure 6. Predicting ligand dependent tumors from the TCGA data set. The measured RNA expression of the ligands in predicted responders (red) vs. non-responders (green). The mean expression (black horizontal lines) and statistical significance of differences is indicated. The receptor mRNA expression is measured in transcripts per million and is displayed on a log-2 scale.

Prediction accuracy and efficiency

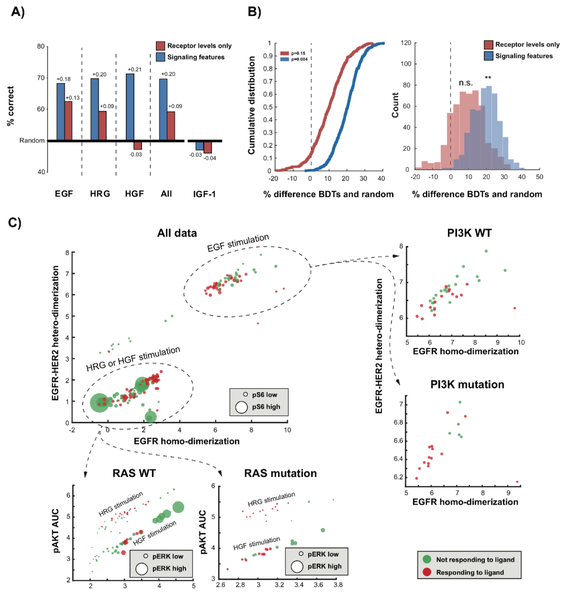

Figure 5. Prediction of ligand-induced proliferation using BDTs. (a) Ratio of true predictions after BDT training with simulated signaling features or receptor expression only, compared to random predictions in the presence of EGF, HRG, IGF or HGF. (b) For 500 random splits of training and testing cell lines, the BDT outcome is compared to random growth assessment as histogram and cumulative density function, showing the significant improvement due to mechanistic modeling. (c) Data of *in-vitro* cell viability screen showing proliferation response (green) or no significant response (red) in different 2D representations of the feature space.

Summary

- We developed a computational model describing multiple signaling pathways and show that a BDT algorithm using simulated signaling features can accurately predict ligand-dependent proliferation *in vitro*.
- Parameters of the model were estimated based on a variety of time-resolved data from seven different cell lines, including a wide range of ligand concentrations with comprehensive single ligand and co-stimulations.
- The computational model not only describes the data for the seven training cell lines but also predicts the signaling responses of two additional, independent validation cell lines.
- To predict proliferation using the BDT algorithm, simulated signaling features are advantageous over using receptor expression directly and result in significantly improved accuracy.
- To demonstrate the applicability of this novel approach to patient samples, we applied it to patient data and observed a significant correlation between the measured ligand expression and the predicted ligand-dependence for breast and colorectal cancer samples.
- The presented novel approach of using BDTs in conjunction with simulated signaling features is the beginning of how complex mechanistic models and large data sets can be combined to better understand cell-specific complexity and heterogeneous tumors.

Predicting ligand-dependent tumors from multi-dimensional signaling features. Hass et al. Nature Systems Biology and Applications, **3** (27), 2017. doi:10.1038/s41540-017-0030-3

A PASSION FOR OUTTHINKING CANCER

merrimack