

UNRAVELLING DRUG RESISTANCE MECHANISMS IN BREAST CANCER

Kumulative Dissertation

zur Erlangung
des mathematisch-naturwissenschaftlichen Doktorgrades
„Doctor rerum naturalium“
der Georg-August-Universität Göttingen

im Promotionsprogramm
„PhD Programme in Environmental Informatics“ (PEI)
der Georg-August University School of Science (GAUSS)

vorgelegt von
Silvia von der Heyde
aus Oldenburg in Holstein

Göttingen 2015

Betreuungsausschuss

Prof. Dr. Tim Beißbarth
Prof. Dr. Stephan Waack

Mitglieder der Prüfungskommission

Referent: Prof. Dr. Tim Beißbarth
Koreferent: Prof. Dr. Stephan Waack

Weitere Mitglieder der Prüfungskommission

Prof. Dr. Edgar Wingender
Prof. Dr. Burkhard Morgenstern
Prof. Dr. Ramin Yahyapour
Prof. Dr. Carsten Damm

Tag der mündlichen Prüfung: 04.06.2015

Dedicated to
My family

Acknowledgements

I would like to thank all of the people who supported my thesis project. They include my supervisors, collaborators, colleagues, friends and family.

I owe special thanks to my direct supervisor Prof. Tim Beißbarth who is leading the group *Statistical Bioinformatics* at the Department of Medical Statistics at the University Medical Center Göttingen. He also headed the *BreastSys* project within the *Medical Systems Biology (MedSys)* initiative funded by the German Federal Ministry of Education and Research (BMBF). It started in February 2009 and ended in January 2013. Its aim was to identify novel therapeutic approaches and efficient drug combinations for breast cancer treatment by establishing data-driven models of tumour progression. All of the thesis related studies were funded via this project.

Furthermore, Prof. Tim Beißbarth supported and guided me by constructive ideas and conceptual approaches. His helpful advice and fruitful discussions, especially on statistical Bioinformatics but also beyond that, formed the heart of the matter within my research.

I also thank my second supervisor and referee Prof. Stephan Waack for his engagement and advice. I also thank Prof. Winfried Kurth, the leader of the GAUSS PhD Programme in Environmental Informatics (PEI) and speaker of the PEI Management Board.

Many thanks to PD Dr. Klaus Jung heading the Core Facility *Medical Biometry and Statistical Bioinformatics* at the Department of Medical Statistics at the University Medical Center Göttingen. After the ending of the *BreastSys* project I got employed there and benefit from Klaus Jung's impressive statistical knowledge.

Furthermore, I thank all of the members of the Department of Medical Statistics, especially the head of the department, Prof. Tim Friede, the secretary, namely Dorit Meyer and Christina Lange, as well as my former colleagues in the group *Statistical Bioinformatics*, namely Michaela Bayerlová, Astrid Wachter, Dr. med. Annalen Bleckmann, Andreas Leha, Dr. Frank Kramer, Dr. Manuel Nietert and Alexander Wolff.

Among the collaborators within the *BreastSys* project I owe special thanks to former and current members of the German Cancer Research Center (DKFZ),

namely Dr. Ulrike Korf, Dr. Johanna Sonntag, Dr. Frauke Henjes and Dr. Christian Bender.

I further thank my friends for being patient whenever spare time was spent on studies.

Finally, I thank my beloved family for always being there.

Table of Contents

List of Figures	xi
List of Tables	xiii
List of Abbreviations	xiv
1 Introduction	1
1.1 Motivation	1
1.2 Theoretical background and prior research	2
1.2.1 The HER2-positive breast cancer subtype	2
1.2.2 Drug action and resistance mechanisms in targeted therapy	4
1.2.3 Genomic and (phospho-)proteomic measurement techniques	7
1.2.4 Boolean models of protein interactions	12
1.2.4.1 Boolean models	13
1.2.4.2 Network reconstruction	16
1.3 Aim and concept of this work	23
2 Summary in German (Zusammenfassung)	25
3 Summary	28
4 Cumulative part of the dissertation	31
4.1 Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines	31
4.1.1 Summary and discussion	31
4.1.2 Declaration of my contribution	36
4.2 RPPanalyzer Toolbox: An improved R package for analysis of reverse phase protein array data	37
4.2.1 Summary and discussion	37

4.2.2	Declaration of my contribution	44
4.3	mRNA Profiling Reveals Determinants of Trastuzumab Efficiency in HER2-Positive Breast Cancer	45
4.3.1	Summary and discussion	45
4.3.2	Declaration of my contribution	53
5	Outlook	54
	References	57
A	Publication “Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines”	73
B	Publication “RPPanalyzer Toolbox: An improved R package for analysis of reverse phase protein array data”	96
C	Publication “mRNA Profiling Reveals Determinants of Trastuzumab Efficiency in HER2-Positive Breast Cancer”	106

List of Figures

1.1	Scheme of ErbB receptor induced cell signalling (reproduced according to Henjes (2010); von der Heyde and Beissbarth (2012); http://openclipart.org/people/hs/hs_DNA_helix.svg).	4
1.2	Scheme of drug targets and ErbB pathway preferences (von der Heyde et al. (2014a), reproduced according to Henjes (2010)).	6
1.3	Example of a typical RPPA layout (von der Heyde et al., 2014b).	11
1.4	Example of a simple Boolean network model (Wang et al. (2012), ©IOP Publishing. Reproduced by permission of IOP Publishing. All rights reserved.).	15
1.5	Workflow of the MCMC-based network reconstruction via <i>ddepn</i> (von der Heyde et al. (2014a), reproduced according to Bender et al. (2010) and Bender et al. (2011)).	21
4.1	Scheme of the ErbB network modelling approach (von der Heyde et al., 2014a).	33
4.2	Reconstructed ErbB network models based on short-term measurements (von der Heyde et al., 2014a).	35
4.3	Workflow applying the <i>RPPAnalyzer Toolbox</i> (von der Heyde et al., 2014b).	38
4.4	Smoothing spline fit through dilution series data of the PRAS40 antibody (supplier code 2691) in the A549 cell line for extrapolation to zero (according to von der Heyde et al. (2014b)).	40
4.5	Exemplary graphics created with the <i>RPPAnalyzer</i> (according to von der Heyde et al. (2014b)).	42
4.6	Fold Changes (log ₂) of differentially expressed genes (BT474 versus HCC1954) with positive values indicating an upregulation in BT474 (von der Heyde et al., 2015).	47
4.7	Fold Changes (log ₂) of differentially expressed genes (BT474 plus trastuzumab versus BT474) with positive values indicating an upregulation upon drug treatment (von der Heyde et al., 2015).	48

- 4.8 Principle Component Analysis (PCA) plot of all measured samples, in which same colours denote that samples belong to the same conducted statistical test (von der Heyde et al., 2015). . . . 49

List of Tables

4.1	Table of detected SNPs in the candidate genes.	50
4.2	Table of detected SNPs in AKT1, HER2, MAPK1 and PIK3CA.	51

List of Abbreviations

ADCC	antibody-dependent cellular cytotoxicity
BIC	Bayesian information criterion
BN	Bayesian network
bp	base-pairs
cDNA	complementary DNA
CRAN	The Comprehensive R Archive Network
cSNP	single nucleotide polymorphism in cDNA
DAG	directed acyclic graph
DBN	dynamic Bayesian network
DDEPN	dynamic deterministic effects propagation networks
DEPN	deterministic effects propagation networks
DNA	deoxyribonucleic acid
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
ER	estrogen receptor
ER+	ER-positive
ER-	ER-negative
FDR	false discovery rate
GA	genetic algorithm
GEO	Gene Expression Omnibus

List of Abbreviations

GRN	gene regulatory network
HER2	human epidermal growth factor receptor 2
HER2+	HER2-positive
HGF	hepatocyte growth factor
HMM	hidden Markov model
HRG	heregulin
HTS	high-throughput sequencing
IGF-IR	insulin-like growth factor 1 receptor
LC	liquid-chromatography
MALDI	matrix-assisted laser desorption/ionization
MAPK	mitogen-activated protein kinase
MCMC	Markov chain Monte Carlo
mRNA	messenger RNA
MS	mass spectrometry
NGS	next generation sequencing
NIR	near infrared
PBN	probabilistic Boolean network
PCA	principle component analysis
PCR	polymerase chain reaction
PI3K	phosphoinositide 3-kinase
PR	progesterone receptor
PR+	PR-positive
RNA	ribonucleic acid
RNA-Seq	deep sequencing of RNA
RPPA	reverse phase protein array
RSS	residual sum of squares

List of Abbreviations

RT-qPCR	real-time quantitative PCR
RTK	receptor tyrosine kinase
SNP	single nucleotide polymorphism
TGF	transforming growth factor
VEGF	vascular endothelial growth factor

Chapter 1

Introduction

1.1 Motivation

Cancer and breast cancer in particular has been studied in-depth during the past decades. The disease has been mainly related to aberrant activity and corresponding miscommunication of cellular components leading to abnormal cell growth, proliferation and migration (Wieduwilt and Moasser, 2008).

Supported by the development of new technologies, individual molecular patterns have been derived for the diverse types not only by organ (Giordano et al., 2001) but even within the organ-specific types. According to gene expression profiles, breast cancer for example can be divided into five molecular subtypes, namely luminal A, luminal B, basal-like, HER2-positive, and normal breast-like phenotypes (Sorlie et al., 2003). Such molecular classifications paved the way for targeted therapies, specifically tailored to the individual genomic characteristics. Within such approaches particularly selected key factors of the cell, mainly genes or proteins, are targeted by therapeutic agents (Hynes and Lane, 2005). The monoclonal antibody drug trastuzumab is a role model for targeted therapy. It is in use against HER2-positive breast cancer, a subtype overexpressing the human epidermal growth factor receptor 2 (HER2) which is targeted by trastuzumab (Tinoco et al., 2013).

Moreover, the success story of fighting cancer has not stopped there. Recent research has revealed even more individual facets of the disease. Hence, we nowadays do not just distinguish cancer types by organ or molecular subtype

but also by the affected persons themselves. Within personalized medicine, each patient should be treated with an individual therapy providing efficient cure of dysfunctions while avoiding unnecessary overtreatment and side effects (Tessari et al., 2013).

The days of treating breast cancer in a shotgun approach like chemotherapy seem to be numbered. However, drug resistance casts a shadow on the achievements in so-called *precision oncology* (Workman et al., 2013). A major problem occurs when cancer cells bypass blocked communication pathways via alternative molecular factors activating tumour progression (Valabrega et al., 2007).

This work aims to reveal drug resistance mechanisms in HER2-positive breast cancer at the genomic and proteomic level in the context of personalized medicine. This comprises the detection of genes and variations as well as protein interactions contributing to drug resistance in targeted therapy. Supporting the gain of these novel biological insights, this work further includes the methodological improvement of a software tool for proteomic data processing and analysis.

1.2 Theoretical background and prior research

1.2.1 *The HER2-positive breast cancer subtype*

Breast cancer represents the most common type of cancer among women (Ferlay et al., 2010). The corresponding subgroups have different cellular properties and prognoses. The *HER2-positive* subtype represents 10-20% of breast tumours. It is associated with increased recurrence and mortality rates (Heil et al., 2012; Jelovac and Wolff, 2012) and characterized by an overexpression of the protein ‘human epidermal growth factor receptor 2’ (HER2, also termed ErbB-2) due to a gene mutation.

This receptor tyrosine kinase (RTK) is a member of the epidermal growth factor receptor (EGFR) family, besides HER1 (EGFR, ErbB-1), HER3 (ErbB-3) and HER4 (ErbB-4) (Roskoski, 2014). These receptors are located at the cell surface, spanning the cellular membrane with an extracellular and an intracellular

part (see Figure 1.1). Tyrosine kinases transfer phosphate groups to the amino acid tyrosine on cell proteins. Via such phosphorylation events proteins can communicate in the cell. In most cases the signal is an activating one. The activated protein then forwards the signal by transferring the phosphate group itself to another protein. This signal is further propagated until the actual effector protein is reached. These effector targets are generally transcription factors which regulate gene expression (Cooper, 2000; Chapter 13). Hence, genes and proteins act together in a *signalling network*.

The signalling pathways in Figure 1.1 are adapted from Galvez-Contreras et al. (2013), Scaltriti and Baselga (2006) and Kong and Yamori (2010). Additional protein interactions were derived from Klos et al. (2006) and Kolch et al. (1993). The scheme represents just an excerpt of dimer signalling. Sometimes intermediate factors may be acting in between two displayed connected proteins in the cascade. So the phosphorylation/activation does not necessarily have to be interpreted as a direct one.

Apart from HER2, the family members have to be activated via growth factors like EGF for EGFR or heregulin (HRG) for HER3 (Henjes et al., 2012). This binding takes place at the extracellular part of the receptors. The activated receptors then act in concert, forming homo- or heterodimers which get phosphorylated at their intracellular part and forward the activation signal into the inner part of the cell. In a pathological situation, as it is the case for HER2 overexpression or other permanently (in-)activated signalling molecules, the phosphorylation cascades become oncogenic, finally promoting uncontrolled cell growth and hence tumourigenesis as well as tumour progression.

Different receptor dimers induce different signal transduction pathways, like the major ones including phosphoinositide 3-kinase (PI3K) and mitogen-activated protein kinase (MAPK), also to a different extent (Olayioye et al., 2000). Overexpression of HER2 leads to a dysregulated protein signalling network and subsequently to uncontrolled tumour growth, proliferation and migration. The oncogenic potential of HER2 does not only lie in its quantity but beyond in its outstanding role as an orphan receptor. This means that it does not rely on growth factor activation but is constitutively active and the preferred

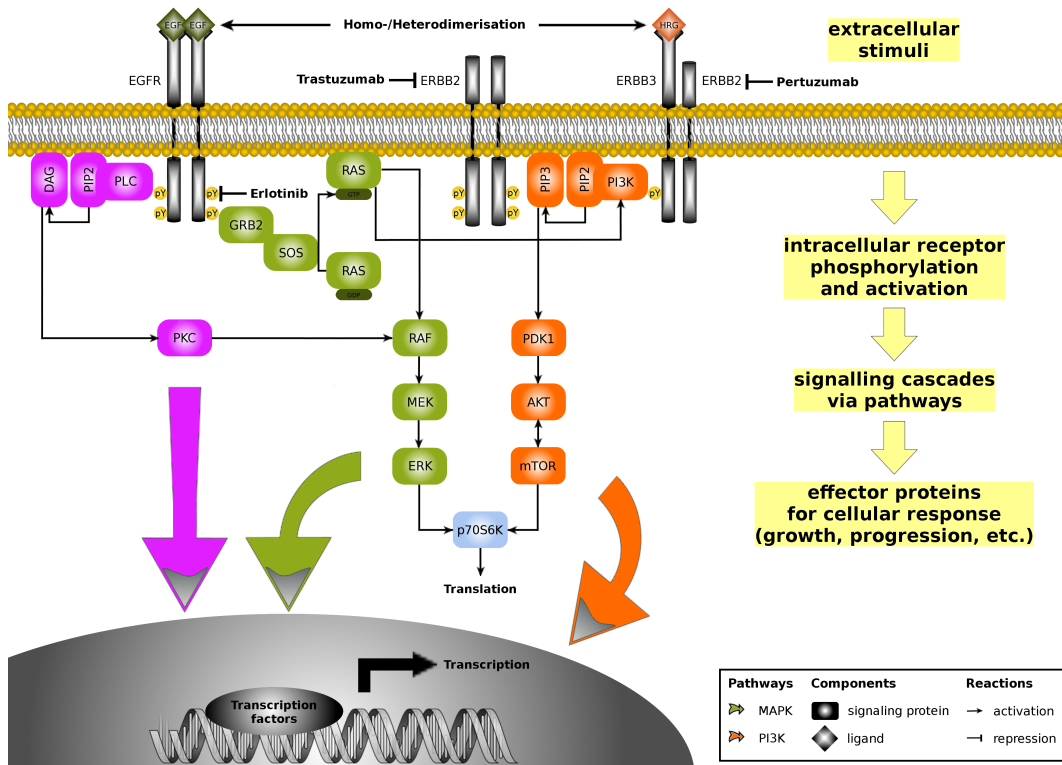


FIGURE 1.1. Scheme of ErbB receptor induced cell signalling (reproduced according to Henjes (2010); von der Heyde and Beissbarth (2012); http://openclipart.org/people/hs/hs_DNA_helix.svg).

dimerisation partner for the remaining EGFR family members (Olayioye et al., 2000; Park et al., 2008; Heil et al., 2012; Jelovac and Wolff, 2012).

1.2.2 Drug action and resistance mechanisms in targeted therapy

The heterogeneity of breast cancer requires therapies tailored towards each patient's molecular profile. Such individual approaches are currently applied within *precision medicine*. As molecular profiles change with tumour development and under drug treatment, latest therapy approaches imply sequential application of targeted therapies guided by biomarker changes (Sahin et al., 2014). HER2 is such a biomarker including its downstream targets such as mTOR.

Specific therapeutics were designed to prevent HER2 induced deregulated protein signalling, contributing to tumour progression. The monoclonal antibody drugs trastuzumab and pertuzumab have especially been designed to target

HER2 (Tinoco et al., 2013). Both bind to different extracellular domains of HER2.

The approved humanized monoclonal antibody trastuzumab is part of the adjuvant treatment of patients with HER2-positive (HER2+) early breast cancer (Valabrega et al., 2007). So it is in general combined with chemotherapy. The improved outcome due to the addition of trastuzumab is not completely understood. Besides inhibition of HER2, its dimerisation and cleavage, it has been associated with different mechanisms of action. These include inhibition of downstream signal transduction pathways like PI3K, antibody-dependent cellular cytotoxicity (ADCC), induction of cell cycle arrest and apoptosis or inhibition of tumour angiogenesis (Valabrega et al., 2007; Arteaga et al., 2012). Pertuzumab prevents the formation of HER2 dimers, especially the most potent ones including HER3. The combination with trastuzumab-based chemotherapy is synergistically associated with improved clinical outcomes and was approved as neoadjuvant therapy for HER2-positive breast cancer in 2013 (Reynolds et al., 2014).

The small-molecule inhibitor erlotinib targets the intracellular tyrosine kinase domain of EGFR and is already in use against non-small cell lung cancer (Pallis and Syrigos, 2013) and pancreatic cancer (Moore et al., 2007).

Figure 1.2 depicts the different targeting mechanisms of the therapeutics trastuzumab (T), pertuzumab (P) and erlotinib (E). Furthermore, it shows schematically the MAPK and PI3K pathways which are predominantly stimulated by the different ErbB receptor dimers upon ligand binding to ErbB-1 and ErbB-3 (Olayioye et al., 2000; Esteva and Pusztai, 2005; Mikalsen et al., 2006; Chen et al., 2009; Dienstmann et al., 2012). These pathways are intended to get inhibited by the mentioned drugs.

However, targeted therapeutics are limited in their success to inhibit the oncogenic signalling of overexpressed or mutated ErbB receptors. Frequently, therapy resistance occurs (Motoyama et al., 2002; Wilson et al., 2012; von der Heyde and Beissbarth, 2012), often due to deregulated pathway activity (Wang et al., 2011; Gallardo et al., 2012) or bypasses of pathway blockades via other

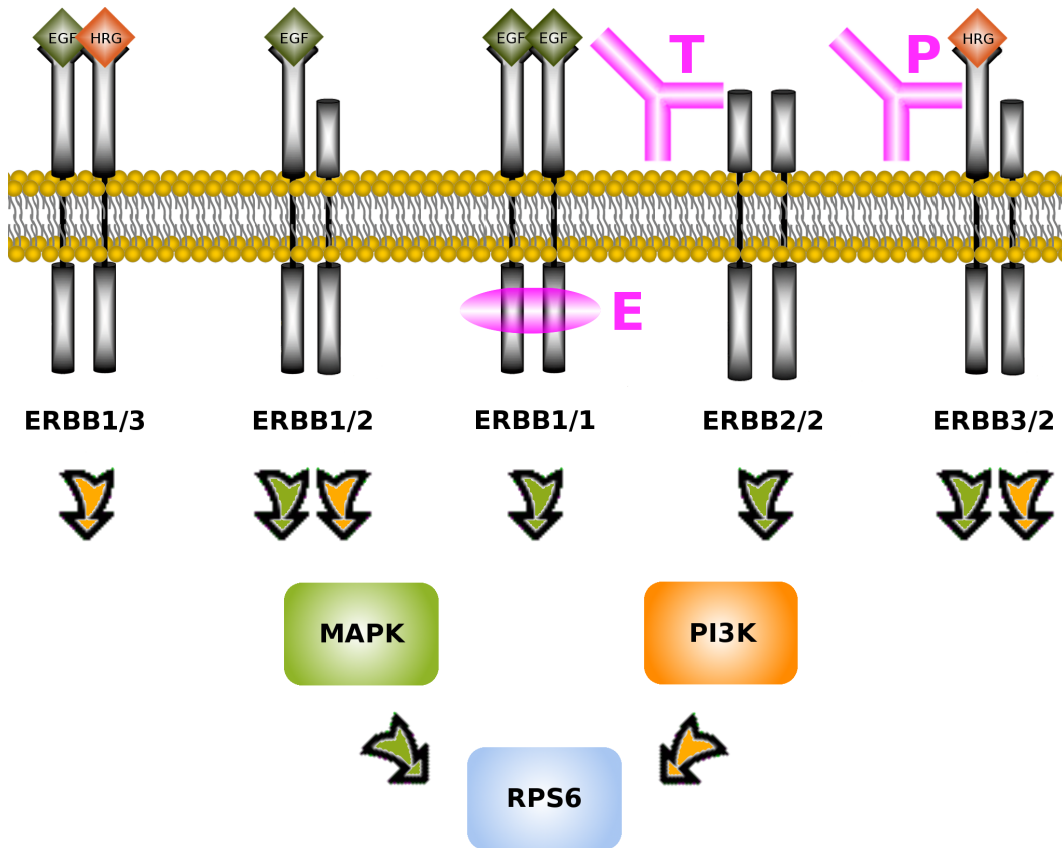


FIGURE 1.2. Scheme of drug targets and ErbB pathway preferences (von der Heyde et al. (2014a), reproduced according to Henjes (2010)).

RTKs, especially ErbB family dimers (Diermeier et al., 2005). EGFR in particular plays a major role in overcoming HER2 targeting. Resistance to trastuzumab is a major problem in treating HER2-positive breast cancer. So-called *intrinsic* (also termed *primary* or *innate*) resistance is pre-existent to drug treatment, e.g. due to mutations like PIK3CA (Kataoka et al., 2010). *Acquired* resistance on the contrary is developed over time induced by diverse mechanisms, e.g. due to molecular changes, despite initial drug response (O'Brien et al., 2010). Possible resistance mechanisms involve overexpression of EGFR, HER2 or HER3, which is accompanied by alternative cell signalling via different ErbB dimers. Alternative signalling pathways can be further induced by MET receptor or insulin-like growth factor 1 receptor (IGF-IR). Other mechanisms include constitutive PI3K pathway activation due to mutations in the PIK3CA gene or PTEN loss, steric hindrance of HER2-antibody interaction or overexpression of transforming growth factor (TGF)- α , HRG or vascular

endothelial growth factor (VEGF) (Valabrega et al., 2007; Arteaga et al., 2012).

To overcome resistance in the treatment of HER2-positive breast cancer, diverse novel drugs are in development. The small molecule tyrosine kinase inhibitor lapatinib and the HER2/3 antibody pertuzumab for example paved the way for improved therapeutic strategies (Figueroa-Magalhães et al., 2013). However, a need for a deeper understanding of the mechanisms of trastuzumab action and resistance persists to predict prior to treatment which patients will likely develop resistance and to develop improved agents or combinations.

1.2.3 Genomic and (phospho-)proteomic measurement techniques

Within the basic information flow in a cell, DNA is transcribed into messenger RNA (mRNA) which is translated into proteins (Crick, 1970). The corresponding cellular contents are distinguished by specific *-omic* terms (see Greenbaum et al. (2001)) of which important ones are explained in the following. The *genome* is represented by the DNA which consist of nucleotide sequences, representing genes as well as non-coding regions. The *exome* is represented by the DNA which is transcribed into mRNA, i.e. the protein-coding regions which constitute approximately just one percent of the human genome (Ng et al., 2009). The *transcriptome* defines the exome of a specified cell population. The *proteome* is represented by all expressed proteins under defined conditions. Both genomic/transcriptomic as well as proteomic information is important to understand the networking in a cell and how it becomes deregulated, pathologically causing cancer or drug resistance.

After the completion of the Human Genome project (2003), the Human Proteome Organization focusses on the 20,300 human protein-coding genes. The aim is the characterization of at least one protein product and post-translational modifications, single amino acid polymorphisms and splice variant isoforms from the protein-coding genes within the next 10 years (Wang et al., 2014).

Popular measurement techniques for genomic/transcriptomic information are DNA microarrays and more recently emerging Next Generation Sequencing (NGS), also termed high-throughput sequencing (HTS) (Malone and Oliver, 2011).

Microarrays generally consist of genomic DNA snippets complementary to transcripts of interest. Sample transcripts under investigation are fluorescently labelled and hybridize to their complementary snippets. After a washing step just hybridized transcripts remain. Via laser scanning the gene expression can then be measured based on light intensity, leading to continuous values.

Deep sequencing of RNA (RNA-Seq) is not limited to immobilized oligonucleotides which map to transcripts of interest. This NGS technology directly detects the transcripts in the sample of interest and maps them to a reference genome. The nucleotide sequences, so-called *reads*, which map to the reference sequence, are counted, leading to non-negative integer expression values.

The *HiSeq 2000* technology of Illumina Inc. (<http://www.illumina.com>) for example works as follows to identify transcripts (resummarized according to http://www.dkfz.de/gpcf/hiseq_technology.html). If one is interested in mRNA as protein draft, just mRNA is extracted from the sample under investigation. The mRNA is then translated into complementary DNA (cDNA) (Wang et al., 2009) of which 100-300 base-pairs (bp) long DNA fragments are created. The ends of these fragments are repaired and A-overhangs are added at the 3'-strand-ends. Afterwards, adaptors are ligated to both ends and fragments are selected according to size and purified. Illumina's device *cBot* is used then to attach the fragments to a flow cell with adaptor-complementary oligonucleotides. Within so-called *bridge amplification* hundreds of millions of unique clusters are generated. The last step before the actual sequencing is to remove the reverse strands and to add the sequencing primer to the DNA templates. In a base-wise manner, the templates are copied during the sequencing procedure applying the four fluorescently labelled nucleotides. Each synthesized base is then detected by a camera which captures its laser-induced fluorescent signal.

In contrast to microarrays, RNA-Seq is not limited to the hybridized probes but allows to measure genome-wide expression levels, independent of annotated regions. Another advantage is the ability to detect polymorphisms and isoforms (Malone and Oliver, 2011).

Single nucleotide polymorphisms (SNPs) represent sequence alternatives (alleles) at single base pair positions in genomic DNA in non-diseased individuals in population(s), wherein the least frequent allele has an abundance of at least

1% (Brookes, 1999). According to Brookes (1999), this definition does not hold for single base insertion/deletion variants (indels), but usually the term *SNP* is used very loosely. Thus, single base variants in cDNAs (cSNPs) are called SNPs as well, since they usually imply underlying genomic DNA variants. Another problem of the SNP definition is related to the incidence frequency, as in practice it is often not possible to attain representative global population samples. To sum up, the term *SNP* is widely used for diverse types of sequence variation.

SNPs are of special interest, as they are associated with heritable phenotypes, multifactorial diseases as cancer and drug response (Gray et al., 2000). The latter is especially interesting for approaches in personalized medicine. Due to their importance, SNP data are deposited in public databases (Sherry et al., 1999). The favoured method to identify SNPs is direct DNA sequencing (Gray et al., 2000; Kwok and Chen, 2003) but RNA-Seq as well has been proven to be a valuable method to identify SNPs (Quinn et al., 2013).

Popular gel-free proteomic measurement techniques are mass spectrometry (MS)-based proteomics and antibody microarrays (Alvarez-Chaver et al., 2014). The MS technique requires that the protein is digested into fragments (peptides). Within liquid-chromatography coupled to tandem mass spectrometry (LC-MS/MS) the peptides are at first separated via LC. Subsequently, they get ionized, e.g. via matrix-assisted laser desorption/ionization (MALDI). Finally, a dual stage MS is applied to identify/quantify the proteomic components by their mass and charge.

Another established technique are protein microarrays for simultaneous analysis of different proteins. Two types are distinguished, namely forward phase protein arrays and reverse phase protein arrays (RPPAs). Forward phase protein arrays resemble the DNA microarrays mentioned before, as they use immobilized capture molecules like antibodies which bind specifically a target protein of interest in a fluorescently labelled probe.

The RPPA approach works the other way around, i.e. here the samples are immobilized and a primary antibody binds the protein of interest. A fluorescently labelled secondary antibody then binds to the primary antibody, so that the continuous intensity values allow to quantify protein abundance or posttranslational modifications such as phosphorylation.

RPPAs were introduced by Paweletz et al. (2001) as a robust and reproducible technology, enabling efficient proteomic quantitation across large sample sets. According to Alvarez-Chaver et al. (2014), antibody microarrays are currently seen as a valuable method for cancer biomarker discovery due to their low-volume requirements, multiplexed detection capability, rapidness and automation. Indeed, the RPPA community is expanding with applications in the field of biomarker discovery and systems biology (Paweletz et al., 2001; Hennessy et al., 2010; Gonzalez-Angulo et al., 2011; Uhlmann et al., 2012; Federici et al., 2013; Sonntag et al., 2014; Ummanni et al., 2014).

In more detail, the RPPA workflow starts with spotting just minimal amounts (1 nl volume) of cell lysate on nitrocellulose-coated glass slides via a printing robot. Usually a serial dilution of control samples with defined protein concentrations is printed in addition for normalization. One slide consists of several subarrays which are then specifically incubated with the primary antibody for a protein of interest. In the next step, the secondary antibody is applied which is labelled with a near infrared (NIR) fluorescent dye. As it binds the primary antibody, this allows to detect for each spot intensity values for protein abundance or phosphorylation. Therefore, the slides are scanned and a microarray image analysis software is applied. In such a workflow 20-200 slides can be processed simultaneously in a parallelized high throughput fashion.

To account for technical spotting variance and corresponding deviant total protein concentrations, the signal intensities can be normalized spot-wise by using the dye Fast Green FCF (Loebke et al., 2007). A normalization slide is stained with the dye and the total protein content of each lysate spot is determined on which the correction factors are based. The signal intensities are then divided spot-wise by these factors and multiplied by the median value of the corresponding normalizer subarray to rescale the data to the native range. Figure 1.3 shows the RPPA layout used in von der Heyde et al. (2014b). Here, MET receptor signalling was analysed. A human alveolar adenocarcinoma cell line (A549) was serum-starved for 24 hours and stimulated with six different hepatocyte growth factor (HGF) concentrations ranging from 0 - 100 ng/ml. The activity of (phospho-)proteins was measured at six time points between 0 and 120 minutes.

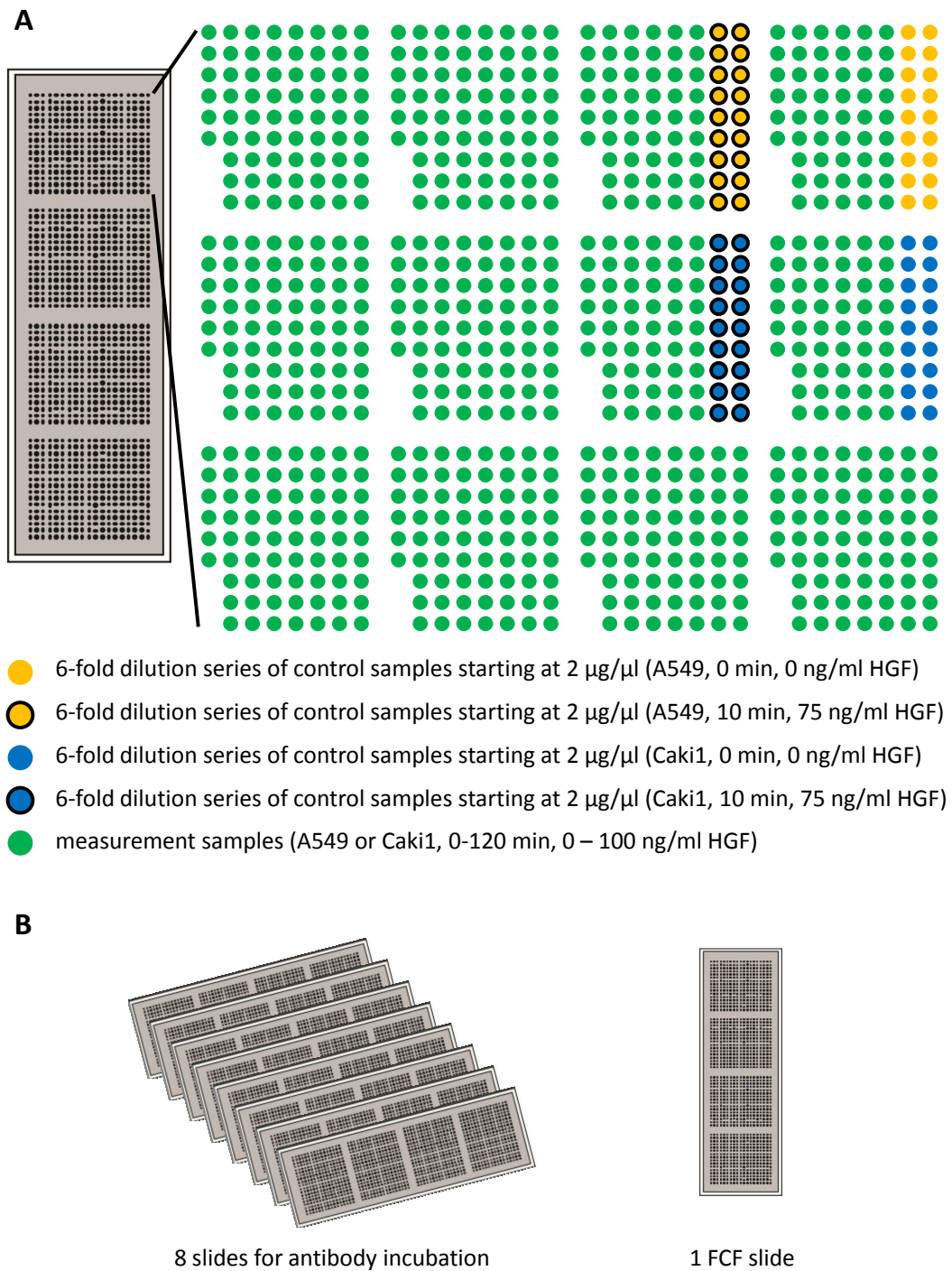


FIGURE 1.3. Example of a typical RPPA layout (von der Heyde et al., 2014b).

The samples were spotted as technical and biological triplicates on four sub-arrays of nine slides each. Control samples were spotted as 6-fold dilution series covering 3.125% to 100% concentration. Eight slides were incubated with target-specific antibodies per subarray. One slide was used for total protein staining via Fast Green FCF. The layout in Figure 1.3 also includes measurements for human clear cell renal cell carcinoma (Caki1). The bold marked yellow control spots of the dilution series were the ones that were used for normalization of the A549 measurements. The corresponding correction method will be described in more detail in Chapter 4 in Section 4.2, as it was part of the technical report about an extension of the R (R Core Team, 2013) package *RPPanalyzer* (von der Heyde et al., 2014b), an open-source software to process and analyse RPPA data, originally published by Mannsperger et al. (2010). In short, smoothing splines are fitted through the dilution series to estimate the intercept at concentration zero, representing background noise which should be corrected for. Subsequently, a linear model is identified which optimally predicts the intercept of the dilution series. This optimal model is then applied to predict the intercept of the measurement sample of interest. This intercept is finally subtracted from the sample signal intensity.

1.2.4 *Boolean models of protein interactions*

A prerequisite to counter deregulated protein signalling is to understand how the molecular factors interact. This knowledge can be deduced from genomic and proteomic measurements.

Diverse biochemical or biophysical methods exist and are emerging to infer interactions between genomic and proteomic components, i.e. gene interactions (Demuth and Wade, 2006), protein-DNA interactions (Rodriguez and Huang, 2005) or protein-protein interactions (Rao et al., 2014). Based on this knowledge interaction pathways have been inferred. The known interactions are collected in diverse databases which make this knowledge publicly available (Xenarios et al., 2002; Schaefer et al., 2009; Hornbeck et al., 2012; Chatr-Aryamontri et al., 2013; Croft et al., 2014).

Within a forward engineering approach, such derived biological signalling networks can then be translated into computational models which allow to simulate the qualitative temporal system behaviour as well as cellular responses to per-

turbances of the network (Oda et al., 2005; Feiglin et al., 2012; Wang et al., 2012). Hence, we can predict how knock-outs, drugs or stimuli influence protein interactions without the necessity for additional wet-lab experiments.

Within reverse engineering approaches, the regulatory relationships and signalling pathways are inferred from observed activity data (Bansal et al., 2007; Roukos, 2011; Wang et al., 2012). This way, network models can be established directly from measurement data through computational analysis.

1.2.4.1 Boolean models

If the system of interest has already been studied and chemical or kinetic parameters are determined, a common approach is to make use of these details within *quantitative* differential equation models with continuous protein concentration values (Schoeberl et al., 2002; Hatakeyama et al., 2003; Jones et al., 2006; Chen et al., 2009). Lacking such detailed information, less complex but more abstract *qualitative* Boolean models can be applied. Here, the activity states of the network components, i.e. genes or proteins, are modelled in a discrete way as active (1) or inactive (0) (Sahin et al., 2009; Samaga et al., 2009). This simplification is especially attractive for computationally extensive large-scale systems (Wang et al., 2012).

Wang et al. (2012) provide a methodology overview of Boolean modelling in systems biology. An excerpt, supplemented by information from the vignette of the *BoolNet* (Müssel et al., 2010) R package which provides analysis tools for Boolean networks, is described in the following. A Boolean network consists of a set of nodes with binary states which are determined by other network nodes through Boolean functions. For example, protein P_1 is activated by phosphorylation by protein P_2 or P_3 and inhibited by protein P_4 . Such dependencies between n variables are expressed via Boolean functions which map the set of all n -tuples over $\{0, 1\}$ to a binary output, i.e. $\{0, 1\}^n \rightarrow \{0, 1\}$. In the mentioned example n equals four. Boolean functions are based on logical operations via AND, OR and NOT. In the example the corresponding function to determine the activity state of protein P_1 would be $P_1 = (P_1 \text{ OR } P_2 \text{ OR } P_3) \text{ AND NOT } P_4$. In general, a Boolean network model consists of n Boolean variables $\{\sigma_1, \sigma_2, \dots, \sigma_n\}$ each assigned with a Boolean function $B = \{B_1, B_2, \dots, B_n\}$ which determines

the value of the variable. Hence, the state of the network at time t can be represented by the state vector $\sigma(t) = (\sigma_1(t), \sigma_2(t), \dots, \sigma_n(t))$. The so-called state space contains all possible 2^n states.

Within the *synchronous* scheme, the subsequent state of the network $\sigma(t+1)$ is computed by applying all transition functions $B_i(\sigma(t))$, $i = 1, 2, \dots, n$, simultaneously. Biologically this means that all genes are updated at the same time. Within the *asynchronous* scheme, it is assumed that gene expression levels are likely to change at different time points. A common stochastic asynchronous scheme looks like the following. At each time point t only one transition function $B_i \in B$ is chosen randomly and the corresponding variable is updated. In probabilistic Boolean networks (PBNs) more than one transition function can be specified per variable. Each function is chosen with a specific probability, and the probabilities of all functions for one variable sum up to one. A state transition is performed by choosing one function for each variable according to the probabilities and by applying the chosen functions synchronously.

State transition graphs represent all possible trajectories in the state space. The corresponding nodes represent the system states and the edges represent the state transitions which are allowed according to the updating scheme. The subsequent state updates at each time step form a trajectory of states which either leads to a steady state, which remains unchanged under further system updates, or a set of recurring states. Such states are called *attractors* and represent the stable states of cell function in which the network mostly resides. Consequently, they reflect system phenotypes under specific perturbations. All states that lead to a specific attractor form its so-called basin of attraction. Apart from steady-state attractors, simple and complex attractors exist. Simple attractors occur in synchronous Boolean networks and consist of a set of states among which the system oscillates in a cycle. Complex or loose attractors occur in asynchronous networks and consist of a set of states among which the system oscillates irregularly. A state of this set can be reached from all other states in the set. Wang et al. (2012) refer to several approaches for attractor calculations. As the R package *BoolNet* was applied in the modelling study of this dissertation, the following description is based on the methods of this software. Within an exhaustive synchronous search, the software starts from all possible network states and conducts synchronous state transitions until a simple or steady-state attractor is reached. This approach is rather

time and memory consuming with increasing network size. *BoolNet* supports this method just for networks up to 29 genes due to memory restrictions in R. However, *BoolNet* offers synchronous and asynchronous heuristic search algorithms as well, which start from a subset of predefined or random states. The asynchronous heuristic performs random transitions leading to an attractor with high probability which is afterwards tested for being complex. Finally, Markov chain simulations can be used to identify relevant states by calculating the probability to reach a state after a predefined number of iterations. If the number is large enough, it is very likely to reach an attractor.

A Boolean network can be expressed as a directed graph $G(V, E)$ in which the nodes $V = \{v_1, v_2, \dots, v_n\}$ represent Boolean variables and the edge set E is defined by Boolean functions. Edges can be activating or inhibiting according to the node effects in the Boolean functions. Figure 1.4 shows an exemplary Boolean network (a), related Boolean functions (b), truth tables mapping input to output values (c) and the synchronous state transition graph (d).

Figure 1 from Rui-Sheng Wang et al 2012 Phys. Biol. 9 055001

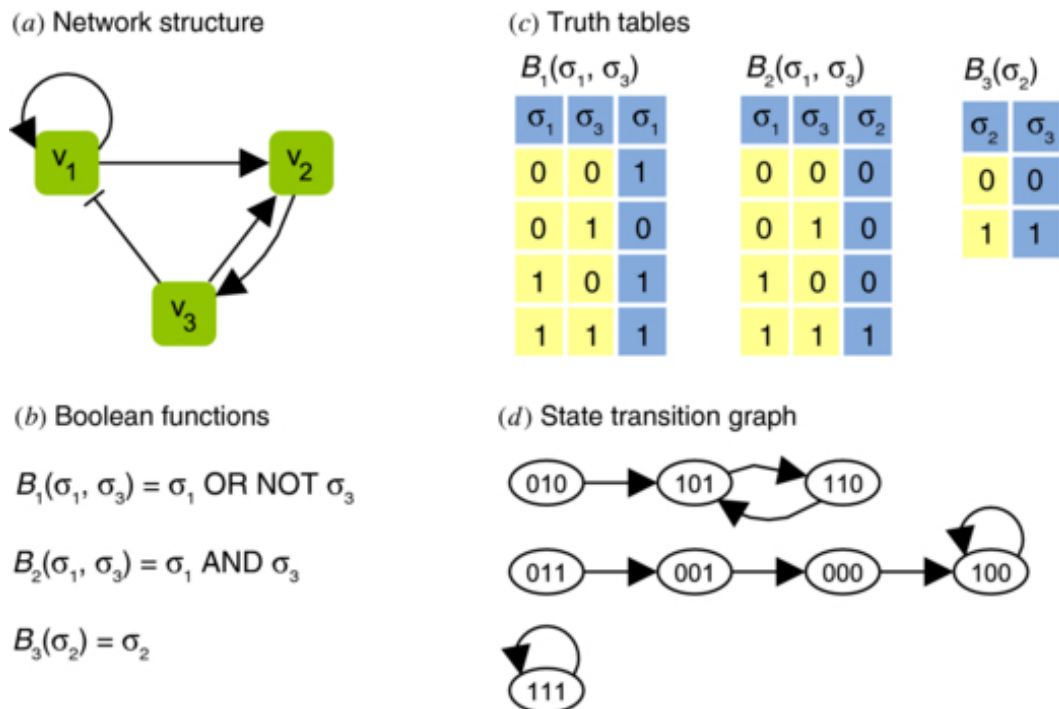


FIGURE 1.4. Example of a simple Boolean network model (Wang et al. (2012), ©IOP Publishing. Reproduced by permission of IOP Publishing. All rights reserved.).

The R package *BoolNet* provides tools for assembly, analysis and visualization of Boolean networks. Apart from synchronous, asynchronous and probabilistic Boolean networks, so-called *temporal* Boolean networks are supported. Here, the subsequent state $\sigma(t + 1)$ may not only depend on the direct preceding state $\sigma(t)$ but on any predecessor state or the time step itself. The software offers different methods for network assembly. Networks can be predefined and loaded from files, reconstructed from time course data or generated randomly. Furthermore, attractors can be calculated and perturbation simulations can be conducted including overexpressed or knock-out genes, for example, to test the robustness of networks to noise and mismeasurements. Above that, visualization tools are offered. Boolean networks can be graphically represented as graphs in which genes correspond to the vertices and the inputs defined by the transition functions determine the edges. Transition states can be visualized in table format with row-wise representation of the genes and column-wise representation of the consecutive states until the attractor is reached. One can also generate a transition graph of the network nodes and state transitions in which attractors are highlighted in bold and different colours allow to distinguish the basins of attraction. Simple attractors and the corresponding set of states can be visualized as well in table format. Also a graph representation of complex attractors is offered with nodes representing the related states and arrows representing the state transitions.

Several import and export interfaces exist, but as the R package *ddepn* (Bender et al., 2010, 2011) was chosen within this dissertation for network reconstruction, it was especially extended to convert the resulting adjacency matrices to logical rules which can be used as input for *BoolNet*.

1.2.4.2 Network reconstruction

Lee and Tzou (2009) reviewed different computational methods to infer gene regulatory networks (GRNs) which is summarized in the following. In principle, gene expression data are used to describe the phenotypic behaviour of a system of interest. First an initial model is built which represents the system behaviour under a specific condition. Then the model predictions for new conditions are compared with corresponding experimental data. This allows to validate the accuracy of the hypothetical model and to revise it accordingly.

Within network reconstruction experimental data, such as gene expression data from microarrays, are combined with prior knowledge to computationally infer a suitable network structure, which is also termed *reverse engineering*. Computational methods are applied to derive the network model parameters, to build the model, to simulate the system behaviour and to evaluate it by comparing it with the experimental data. Prior knowledge helps to narrow down the search space in the network and hence to save time and computational effort.

Lee and Tzou (2009) describe discrete models as well as models with continuous variables based on differential equations. The discrete models include Boolean network models, PBN models and Bayesian network models. As discrete network models have been applied in this thesis, the focus is put on those.

For Boolean models literature- and data-based approaches are distinguished. Within literature-based approaches the models are built from known gene interactions which are translated into logical rules. Within approaches based on experimental data two main classes of methods exist. One applies correlation measurements such as the mutual information between genes. The other one applies machine-learning approaches like the genetic algorithm (GA) which is also available in the *ddepn* package. Such global methods can further be combined with local search techniques to determine promising directions in the search space.

Bayesian network models are directed graphs in which the edges describe the conditional dependencies between the nodes. The modelling involves model structure selection and parameter learning. A Bayesian scoring metric is applied to evaluate the model. The aim is to infer a model which describes the data with high probability. Heuristic strategies can be applied to enhance computational performance.

Bansal et al. (2007) compared different software applications for reverse engineering. In this context they introduced diverse algorithms for network inference as well. They describe Bayesian networks more detailed as a graphical model for probabilistic relationships among n random variables $X_i, i = 1, \dots, n$. These variables represent the nodes of a directed acyclic graph (DAG) G in

static Bayesian networks (BNs). The relationships between the nodes are expressed via a joint probability distribution

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i = x_i | X_j = x_j \text{ for each } X_j \text{ which is a parent of } X_i)$$

with genes j (*parents*) as regulators of gene i on which the probability is conditioned. Hence, the joint probability density can be described by a product of conditional probabilities, assuming independence of probabilities and given that the variables X_i are just depending on their direct parents according to the *Markov assumption*.

Within network reconstruction, the DAG G is inferred whose network topology describes the steady-state gene expression data D best according to a scoring function. The score to be maximized can be expressed by the Bayes rule $P(G|D) = \frac{P(D|G)P(G)}{P(D)}$ with $P(G)$ as a constant or incorporating prior knowledge of the network structure. One of the most popular scores is the Bayesian Information Criterion (BIC) which also penalizes graph complexity to avoid overfitting.

The evaluation of all possible network topologies and choosing the one with maximum score is an NP-hard problem. Therefore, heuristic search methods like the Markov Chain Monte Carlo (MCMC) approach are applied. A further problem is that usually several high-scoring networks are inferred. Therefore, bootstrapping can be applied to get confidence estimates for the network edges and to select the most probable topology. Furthermore, prior knowledge can be incorporated into network reconstruction to choose the most reliable model. Anyhow, one has to be cautious interpreting reconstructed BNs, as they do not necessarily represent direct gene interactions but probabilistic dependencies. The parents of a node are not necessarily also the direct causes of its behaviour. The main drawback of BNs is their acyclicity, as feedback loops cannot be modelled but represent important biological network elements.

Dynamic Bayesian networks (DBNs) are not limited in this regard. They extend BNs and can be specifically applied to temporal processes and hence to time series data. An introduction is given by Friedman et al. (1998) and Murphy and Mian (1999). In DBNs the values of the random variables change over time which is denoted by $X_i(t)$ with $t = 0, 1, \dots, T$. The set

of all variables $X_i(t)$ is represented by $\mathbf{X}(t)$. Under the assumption that the temporal process is Markovian, the transition probability can be expressed as $P(\mathbf{X}(t+1)|\mathbf{X}(0), \dots, \mathbf{X}(t)) = P(\mathbf{X}(t+1)|\mathbf{X}(t))$. This is further assumed to be independent of t . A prior network B_0 specifies the distribution over the initial states $\mathbf{X}(0)$. A transition network B_{\rightarrow} over the variables $\mathbf{X}(0) \cup \mathbf{X}(1)$ specifies the transition probability $P(\mathbf{X}(t+1)|\mathbf{X}(t))$ for all t . The parents of $X_i(0)$ are those specified in B_0 . The parents of $X_i(t+1)$ are the nodes in step t and $t+1$ corresponding to the parents of $X_i(1)$ in B_{\rightarrow} . Friedman et al. (1998) define the transition probability as $P_{B_{\rightarrow}}(\mathbf{x}(1)|\mathbf{x}(0)) = \prod_{i=1}^n P_{B_{\rightarrow}}(x_i(1)|\mathbf{pa}(X_i(1)))$ with \mathbf{pa} denoting the parental nodes. The joint distribution over $\mathbf{X}(0), \dots, \mathbf{X}(T)$ is $P(\mathbf{x}(0), \dots, \mathbf{x}(T)) = P_{B_0}(\mathbf{x}(0)) \prod_{t=0}^{T-1} P_{B_{\rightarrow}}(\mathbf{x}(t+1)|\mathbf{x}(t))$.

To capture cell signalling dynamics over time, the input data for network modelling ideally covers several time points at which the gene or protein activities are measured under defined conditions (Hill et al., 2012; Park and Bader, 2012). Such time series can be further combined with knock-out or stimuli perturbation experiments to infer the interplay of the network components (Bender et al., 2011; Penfold et al., 2012; Wagner et al., 2013).

In addition, it is advantageous to integrate prior literature knowledge about protein interactions in the system of interest into the network model (Bender et al., 2011; Eduati et al., 2012; McDermott et al., 2013). Such constraints reduce the computational complexity and ensure that validated interactions are contained in the model. Reverse engineering approaches as implemented in R packages like *ddepn* or *CellNOptR* (Terfve et al., 2012) combine perturbed time course input data and prior knowledge to reconstruct Boolean networks.

The R package *ddepn* offers a Boolean network reconstruction method which is especially tailored to perturbed time course data and allows to integrate prior knowledge. The abbreviation *DDEPN* stands for *Dynamic Deterministic Effects Propagation Networks*. It is an extension of the *DEPN* (*Deterministic Effects Propagation Networks*) approach (Fröhlich et al., 2009). The main difference is that the version of Bender et al. (2010) is designed for longer time series and does not require as many perturbations as the original version. The approach models the protein networks as directed, transitively closed graphs

in which proteins and external perturbations, like drugs or growth factors, are represented by nodes. Their interactions form the edges. The method is deterministic with regard to the perturbation effect propagation from parent to child nodes.

Figure 1.5 shows a scheme of the *inhibMCMC* procedure of the *ddepn* package. According to Bender et al. (2011), this reconstruction workflow which is based on a stochastic MCMC approach looks like the following. The network is represented by an adjacency matrix Φ in which the rows and columns are defined by the network components, i.e. the nodes. The entries are assigned according to the edge types. If the component of row i activates the component in column j , the entry ϕ_{ij} is set to 1. In case of an inhibition it is set to -1. In case of no interaction between the components, the entry is zero. The measurement data for the diverse time points and replicates are stored in a separate matrix D . Within the MCMC approach a network structure is inferred which optimally fits to the measurement data. Therefore, the space of possible networks is sampled based on posterior probabilities. The posterior distribution of a network Φ given the data D is

$$P(\Phi|D) = \frac{P(D|\Phi)P(\Phi)}{P(D)} \propto P(D|\Phi)P(\Phi).$$

The prior probability distribution is represented by $P(\Phi)$. The likelihood of the data given the network is represented by $P(D|\Phi)$. To calculate $P(D|\Phi)$, the optimized system state matrix Γ^* has to be estimated which contains column-wise the measurement time points and row-wise the nodes. Its entries are either one or zero, representing an active (1) or inactive (0) node state.

The estimation of Γ^* starts with the computation of the matrix Γ which contains all reachable system states. It is similar to Γ^* but does not store the activity states of the nodes per time point but per transition step of Boolean signal propagation. The signalling starts when the network is perturbed by stimuli or drugs. These perturbations are assumed to be permanently active and hence set to an active state (1). The remaining nodes are initialised with inactive states (0). The perturbation signal is propagated then from parent to child nodes according to the transition rule that a child gets activated if at least one activating parent node is active and all inhibiting parents are inactive.

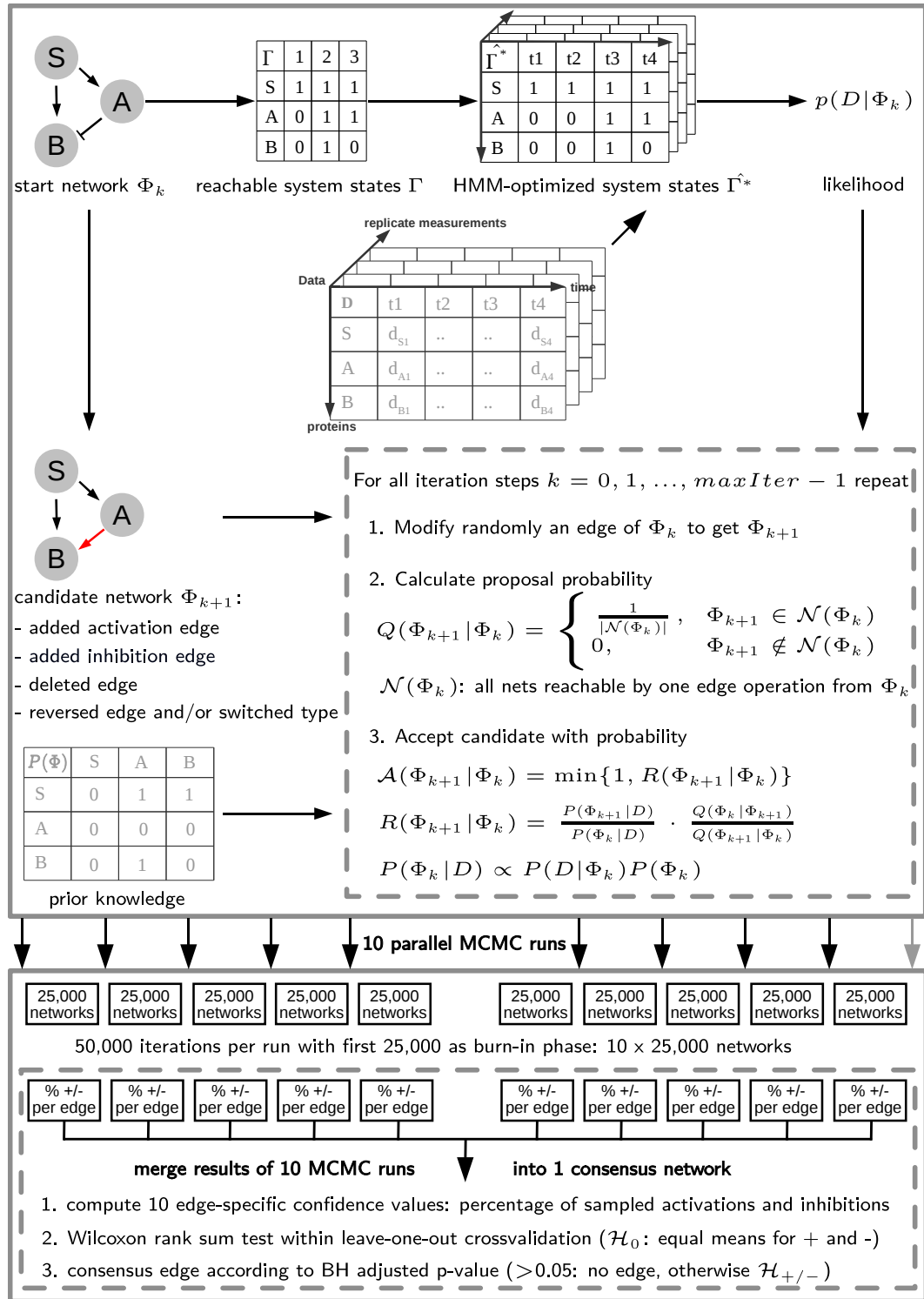


FIGURE 1.5. Workflow of the MCMC-based network reconstruction via ddepn (von der Heyde et al. (2014a), reproduced according to Bender et al. (2010) and Bender et al. (2011)).

To relate Γ^* to the measured time points and estimate the optimized Γ^* with the true unknown state sequence per node over time, a hidden Markov model (HMM) is applied. This leads to $\hat{\Gamma}^*$ which indicates whether a data point d_{itr} for a node i at time point t for replicate r has an underlying active (1) or passive (0) normal distribution

$$d_{itr} \sim \begin{cases} \mathcal{N}(\mu_{i0}, \sigma_{i0}), & \text{if } \hat{\gamma}_{itr}^* = 0 \\ \mathcal{N}(\mu_{i1}, \sigma_{i1}), & \text{if } \hat{\gamma}_{itr}^* = 1. \end{cases}$$

The distribution parameters for each node i are estimated as empirical mean and standard deviation of all measurements for the nodes in the same class. They are stored in the parameter matrix $\hat{\Theta} = \{\hat{\theta}_{i0}, \hat{\theta}_{i1}\} = \{(\hat{\mu}_{i0}, \hat{\sigma}_{i0}), (\hat{\mu}_{i1}, \hat{\sigma}_{i1})\}$. This way the first part of $P(\Phi|D) \propto P(D|\Phi)P(\Phi)$ is obtained, namely

$$p(D|\Phi) = p(D|\hat{\Gamma}^*, \hat{\Theta}) = \prod_{t=1}^T \prod_{i=1}^N \prod_{r=1}^R p(d_{itr}|\hat{\theta}_{i\hat{\gamma}_{itr}^*}),$$

assuming T time points, N nodes and R replicates.

Still the prior probability distribution $P(\Phi)$ has to be determined. For an edge ϕ_{ij} between nodes i and j it is defined as

$$P(\phi_{ij}|b_{ij}, \lambda, \gamma) = \frac{1}{2\lambda} e^{-\frac{\Delta_{ij}}{\lambda}}$$

when choosing the Laplace prior model (*laplaceinhib*) of *ddepn*. The weighted difference term $\Delta_{ij} = |\phi_{ij} - b_{ij}|^\gamma$ with $\gamma \in \mathbb{R}^+$ penalizes deviation of the network structure Φ from a user-defined prior belief matrix B . Its continuous entries can range from -1 to 1. The absolute values correlate with the confidence in an edge. As independence of the edge probabilities is assumed, the prior belief for a network structure Φ is simply the product of those over all nodes, i.e.

$$P(\Phi|B, \lambda, \gamma) = \prod_{i,j} P(\phi_{ij}|b_{ij}, \lambda, \gamma).$$

In Figure 1.5 the *inhibMCMC* procedure spans 50,000 iterations in 10 parallel runs with a burn-in phase of 25,000 iterative steps each. All runs are initialized with a starting network. The 10 x 25,000 networks resulting after the burn-in phase are merged into one consensus network. Therefore, significantly occurring

edges are identified by a Wilcoxon rank sum testing procedure in which the amount of sampled activations and inhibitions per edge is counted for each run and divided by the total number of sampled edges. The null-hypothesis is tested, whether the means of the ten edge-specific confidence values are identical for activation and inhibition. If the null-hypothesis is not rejected (adjusted p-value > 0.05), no edge is assumed. Otherwise, the alternative determines the interaction type. The procedure is embedded in a leave-one-out cross-validation in which each of the ten MCMC chains is left out once. An edge is included in the final consensus network if it occurred in all cross-validation runs.

After the data- and literature-based inference of the network model, it can be used for perturbation simulations to predict the system response towards stimuli, drugs or gene/protein knock-outs. The corresponding stable activity states of the network components (*attractors*) under a given condition reflect the long-term behaviour (homoeostasis) of the system. As the R package *BoolNet* accepts networks in form of Boolean formulas as input, it can be conveniently applied after network reconstruction.

1.3 Aim and concept of this work

The objective of this work is to shed light into the mechanisms of drug action and resistance mechanisms in individual types of HER2-positive breast cancer. The cumulative dissertation consists of three manuscripts, of which two are data-based research articles (von der Heyde et al., 2014a, 2015) and one is a software extension for proteomic data analysis (von der Heyde et al., 2014b).

The research article *Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines* (von der Heyde et al., 2014a) is based on (phospho-)proteomic RPPA data of three HER2-amplified human breast cancer cell lines with different resistance phenotypes, namely BT474, HCC1954 and SKBR3. While BT474 and SKBR3 are sensitive to trastuzumab, HCC1954 harbours a mutation and is intrinsically resistant. The activity of several (phospho-)proteins was measured after short- and long-term treatment with the drugs trastuzumab, erlotinib and pertuzumab

alone or in combination at 10 time points up to 60 minutes and 30 hours each. Based on these data protein signalling networks were reconstructed to detect protein interactions driving drug resistance and to reveal differences between the cell lines in the context of personalized medicine. In this regard the models were further used to simulate drug perturbations to predict optimal drug combinations for each cell line tailored to the individual phenotypes.

A more detailed description follows in Chapter 4 in Section 4.1.

The technical report *RPPanalyzer Toolbox: An improved R package for analysis of reverse phase protein array data* (von der Heyde et al., 2014b) is related to the analysis of RPPA data, as it describes a software extension of the R package *RPPanalyzer*.

Important data preprocessing steps were merged into a single function to provide a more comfortable usage of the software. The analysis workflow was streamlined, recommending at which step which function of the *RPPanalyzer* should be applied. These functions also include improvements or novel developments. For example, a new method for background noise correction based on dilution series was introduced as well as a new time course plotting function for data transformed by new methods for noise estimation and averaging of replicates.

The manuscript is presented in Chapter 4 in Section 4.2.

The research article *mRNA Profiling Reveals Determinants of Trastuzumab Efficiency in HER2-Positive Breast Cancer* (von der Heyde et al., 2015) is based on RNA-Seq data. Analogously to the research article described before, measurement data of cell lines were used to reveal drug resistance mechanisms. Two of the cell lines were identical, namely the trastuzumab sensitive BT474 cell line and the resistant HCC1954 cell line. While the previously mentioned article further includes the trastuzumab sensitive cell line SKBR3, this article includes a resistant version of BT474, i.e. BTR50, which represents acquired resistance. Here, the focus was put on gene expression and mutations (SNPs), potentially determining trastuzumab efficiency.

The manuscript is described more detailed in Chapter 4 in Section 4.3.

Chapter 2

Summary in German (Zusammenfassung)

Gegenstand dieser Dissertation ist es, auf genomischer und proteomischer Ebene Einsicht in die Wirkungsweise von Medikamenten zu gewinnen, die auf HER2-positive Brustkrebs-Zelllinien mit unterschiedlichen Resistenz-Phänotypen angewendet wurden. In der Ära der personalisierten Krebstherapie sind derartige Einsichten von essentieller Bedeutung für die Weiter- und Neuentwicklung oder Kombination von zielgerichteten Medikamenten und deren Anwendung in individuellen Therapieansätzen.

Auf proteomischer Ebene wurden RPPA-Zeitreihen-Messungen als Grundlage für Zelllinien-spezifische Boolesche Netzwerkrekonstruktionen genutzt, um Modelle der Signalübertragung zwischen den gemessenen Proteinen unter Medikamenteneinfluss zu entwickeln. Untersucht wurden die Wirkstoffe Trastuzumab und Pertuzumab, welche gegen HER2 bzw. dessen Dimerisierung gerichtet sind, sowie Erlotinib mit EGFR als Angriffspunkt. Die Zeitreihen umfassten einerseits eine Zeitspanne bis zu 60 Minuten und andererseits eine Zeitspanne bis zu 30 Stunden.

Mit Hilfe der gewonnenen Modelle wurde simuliert, wie unterschiedliche Medikamentenkombinationen auf die unterschiedlichen Zelllinien wirken. Hierzu wurde der Aktivitätszustand gewählter (Phospho-)Proteine der PI3K- und MAPK-Hauptsignalwege berechnet. Die Simulationsergebnisse deckten sich weitestgehend mit den tatsächlichen Messungen, sind allerdings differenziert

zu interpretieren, da die homöostatischen Zustände nicht zwingend nach den gemessenen maximalen Zeitpunkten erreicht worden sein müssen. Die Simulationen zeigten auch, dass in den Kurzzeitmessungen unterschiedliche Medikamente oder deren Kombinationen eine optimale Wirkung in den Zelllinien hatten, wohingegen in den Langzeitmessungen größtenteils ähnliche Inhibitoren einen optimalen Effekt erzielten. Dies betont die Notwendigkeit, bereits früh onkogene Proteine im Signalweg zu hemmen. Im Hinblick auf die Trastuzumab resistente Zelllinie HCC1954 wurde durch gesonderte Simulationen ermittelt, welche Proteininteraktionen verstärkten Einfluss auf den ohnehin schon hyperaktiven PI3K-Signalweg haben.

Zudem wurden die Modellstrukturen zwischen den Zelllinien verglichen, um potentielle Resistenzmechanismen aufzudecken. Dies ergab Zelllinien-spezifische Präferenzen für unterschiedliche Signalwege in Abhängigkeit des Brustkrebs-Phänotyps, was individuelle Therapieansätze erfordert. Außerdem deuteten die Modelle auf Rückkopplung, Signalweg-Interaktionen sowie hyperaktive Heterodimere als Resistenzmechanismen hin.

Um die Analyse von RPPA-Daten zu erleichtern, wurde weiterhin eine bestehende Software konzeptionell und methodisch weiterentwickelt.

Die Mehrheit der verfügbaren Methoden für RPPA-Daten umfasst kommerzielle oder nicht-standardisierte eigengefertigte Software-Lösungen. Zudem sind diese in der Regel auf Präprozessierung und Normalisierung limitiert ohne weitere Möglichkeiten zur grafischen und statistischen Analyse. Daher ist das Angebot entsprechend erweiterter, Plattform-übergreifender und zudem frei verfügbarer Alternativen für die Anwender der RPPA-Technologie von großer Wichtigkeit. Die *RPPAnalyzer*-Software stellt eine solche Alternative dar. Sie wurde um neue Methoden erweitert, welche in einen standardisierten Programmablaufplan integriert wurden. Dieser ermöglicht es dem Anwender, gewisse Routineschritte der Präprozessierung komfortabel automatisiert durchzuführen. Die neuen Funktionen dienen der Varianzschätzung, der Normalisierung und Visualisierung von Zeitreihenmessungen. Gleichzeitig wurde der modulare Charakter der Software bewahrt, welcher es je nach Bedarf des Anwenders gestattet, flexibel Funktionen anzupassen oder hinzuzufügen.

Auf genomischer Ebene wurden RNA-Seq-Daten verwendet, um Gene zu ermitteln, deren Expression entweder unter der Gabe von Trastuzumab variierte oder aber Unterschiede zwischen den analysierten Zelllinien zeigte. Beide Arten der differentiellen Expression deuten auf eine entscheidende Bedeutung dieser Gene für die Wirkungsweise von Trastuzumab hin. Die Phänotypen der Zelllinien umfassten Sensitivität, intrinsische und erworbene Resistenz. Die ermittelten validierten Gene waren differentiell exprimiert zwischen der sensitiven Zelllinie mit und ohne Trastuzumab-Behandlung sowie zwischen der sensitiven und der intrinsisch resistenten Zelllinie. Bei letzterem Vergleich wurden überwiegend Gene detektiert, denen bereits eine onkogene Wirkung sowie eine Behinderung der Trastuzumab-Wirkung zugeschrieben wurde und von denen bekannt ist, dass ihre Expression stark von Steroidrezeptoren beeinflusst wird. Dies stimmte mit den hier gemachten Beobachtungen überein. Auch hier spielte der PI3K-Signalweg eine bedeutende Rolle im Hinblick auf Medikamentenresistenz.

Des Weiteren ergaben die Untersuchungen, dass die intrinsisch resistente Zelllinie stärker von der sensitiven Zelllinie abwich als diejenige mit erworbener Resistenz. Zudem wurde sie weniger stark durch Trastuzumab beeinflusst als diejenige mit erworbener Resistenz.

Darüber hinaus wurden Mutationen in unbehandelten Zelllinien detektiert, welche ebenfalls einen Einfluss auf die Wirksamkeit des Medikaments haben könnten. In der intrinsisch resistenten Zelllinie wurden mehr Mutationen ermittelt als in derjenigen mit erworbener Resistenz, was die zuvor genannten Ergebnisse bestätigt. Die Schnittmenge der Mutationen in den resistenten Zelllinien beeinflusst sehr wahrscheinlich die Wirksamkeit von Trastuzumab.

Zusammenfassend konnten Einblicke in Resistenzmechanismen durch Modelle und direkte Messungen gewonnen werden. Die Netzwerkmodelle lieferten Proteininteraktionen als potentielle Resistenzmechanismen und erlaubten Simulationen, um optimale Medikamentenkombinationen zu ermitteln. Die genomischen Messungen lieferten Gene und Mutationen als potentielle Resistenzmechanismen. Innerhalb beider Ansätze wurden zudem Vergleiche zwischen unterschiedlichen Resistenz-Phänotypen angestellt, um Anwendungsmöglichkeiten in der personalisierten Medizin aufzuzeigen.

Chapter 3

Summary

The intention of this dissertation is to gain deeper insight into mechanisms of drug action on the genomic and proteomic level in HER2-positive breast cancer cell lines with different resistance phenotypes. In the era of personalized cancer therapy such insights are essential for the development and improvement of (combined) targeted therapeutics and their application in individual therapy approaches.

On the proteomic level, cell line specific Boolean networks were reconstructed based on RPPA time course data to model signal transduction between the measured proteins under drug treatment. The drugs of interest were trastuzumab and pertuzumab which target HER2 and its dimerisation, respectively. In addition, erlotinib was analysed which targets EGFR. The time series covered a range up to 60 minutes and 30 hours, respectively.

These models were used to simulate the effect of different drug combinations on the different cell lines. Therefore, the activity states of selected (phospho-)proteins in the PI3K and MAPK signalling pathways were computed. The simulation results were mostly confirmed by the actual measurement data, but still they have to be distinguished, as homeostasis does not necessarily have to be reached after the measured maximum time points. The simulations further revealed a more diverse drug response in the short-term measurements than in the long-term measurements. This underlines the importance of early drug intervention at the top level layer of the signalling network. For the trastuzumab resistant cell line HCC1954 additional simulations were performed.

These revealed specific protein interactions reinforcing the hyperactive PI3K signalling pathway in this cell line.

Furthermore, the model structures were compared between the cell lines to detect potential resistance mechanisms. Indeed, cell lines with different breast cancer and resistance phenotypes seem to prefer different signalling pathways, requiring individual therapeutic strategies. In addition, the models hint at feedback loops, pathway crosstalk and hyperactive heterodimers as main resistance mechanisms.

To facilitate the analysis of RPPA data, a corresponding software was conceptually refined and methodologically enhanced.

The majority of available tools for RPPA data comprises commercial or non standardized in-house solutions. Above that, these software solutions are generally limited to data preprocessing and normalization, lacking additional functions for graphical and statistical analysis. Hence, the users of the RPPA technology would benefit from improved and freely available alternatives which also allow data comparison across different RPPA platforms. The *RPPAnalyzer* software represents such an alternative. It was extended by new functions which were further integrated in a standardized workflow. This way, users can conveniently conduct automated standard preprocessing steps. The novel methods imply variance estimation, normalization and visualization of time course data. At the same time, the modular character of the software was preserved which allows users to flexibly integrate add-on functions and to choose or adapt existing functions of the toolbox according to their specific needs.

On the genomic level, RNA-Seq data were used to detect genes whose expression differed under trastuzumab treatment or between the analysed cell lines. Both kinds of differential expression point to an essential role of these genes in the action and for the efficiency of trastuzumab. The resistance phenotypes of the cell lines covered sensitivity, intrinsic and acquired resistance. The detected and validated genes were differentially expressed between the sensitive cell line and its trastuzumab treated version as well as between the sensitive cell line and the intrinsically resistant one. The latter comparison mainly revealed genes which are already known in an oncogenic context or suspected to hinder trastuzumab

efficiency. Their expression is strongly depending on steroid receptors which is in line with the results presented here. Analogously to the proteomic analysis mentioned before, also in this analysis the PI3K signalling pathway was ascribed an important role in drug resistance.

Furthermore, the analyses revealed that the intrinsically resistant cell line differs more from the sensitive one than it is the case for the one with acquired resistance. The intrinsically resistant cell line was further influenced less by trastuzumab treatment than the one with acquired resistance.

Additionally, mutations were detected in the untreated cell lines which also potentially impact trastuzumab action and efficiency. In the intrinsically resistant cell line more variations were detected than in the one with acquired resistance. This highlights the afore mentioned results. The intersect of mutations in the resistant cell lines very likely affects the efficiency of trastuzumab.

In conclusion, genomic and proteomic measurement data and corresponding models provided an insight into resistance mechanisms. The signalling network models unveiled protein interactions as potential resistance mechanisms and allowed simulations to infer optimal drug combinations. Based on the genomic measurements, genes and mutations were detected as potential resistance mechanisms. Additionally, within both approaches the different resistance phenotypes were compared to indicate prospective applications in personalized medicine.

Chapter 4

Cumulative part of the dissertation

4.1 Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines

The manuscript on which this chapter is based has already been published.

Silvia von der Heyde, Christian Bender, Frauke Henjes, Johanna Sonntag, Ulrike Korf and Tim Beißbarth*, Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines, *BMC Systems Biology*, **2014**, 8:75.

*Corresponding author

4.1.1 Summary and discussion

The intention of this work was to model the ErbB signalling network in different drug treated ErbB-1 expressing, ErbB-2 amplified breast cancer cell lines in order to reveal individual drug response, synergistic benefits of drug combinations and resistance mechanisms.

The network reconstructions via *ddepn* were based on longitudinal RPPA data

(Henjes et al. (2012), GEO accession number GSE50109) of ErbB receptors and downstream targets in the BT474, SKBR3 and HCC1954 cell lines with differing resistance phenotypes. HCC1954 is trastuzumab resistant due to a PIK3CA mutation, while BT474 is trastuzumab sensitive and exhibits wild type behaviour (Kataoka et al., 2010). The SKBR3 cell line is supposed to be pertuzumab resistant (Nahta et al., 2005; Henjes et al., 2012).

The cell lines were treated with erlotinib, trastuzumab and pertuzumab, alone or combined, up to 60 minutes and 30 hours, respectively. Both, short- and long-term data comprised ten time points each. In the short-term case, cells were stimulated with the growth factors EGF and HRG, and 11 phosphoproteins were measured. In the long-term case, cells were incubated in full growth medium, and 21 (phospho-)proteins were measured.

Besides the (phospho-)proteomic data, prior literature knowledge about protein interactions was integrated as a further source of information into the reconstruction algorithm. Two prior reference networks, i.e. one per time course, were determined as initial joint hypotheses for all of the three breast cancer cell lines. The impact strength of the prior networks was chosen in such a way, that only strongly deviating data influence the network structure, since ErbB signalling and especially the MAPK and PI3K pathways are well studied.

The Boolean models were inferred in a cell line and time course specific manner. These were then used to simulate network response to drug combinations which was represented by the attractor states of selected network components. Therefore, *BoolNet* was applied. The perturbation simulations started with predefined initial states of the network nodes. According to the experiment to be simulated, the activity states of the drug nodes were fixed to zero or one. The perturbations included all possible combinations of the three drugs, leading to eight fixed input combinations. The stimuli nodes were permanently fixed to one. The remaining protein activity start states were initialized with zero but were flexible towards updates. The attractor calculations allowed to predict optimal drug treatment customised to the topology of the different cell line phenotypes.

A scheme of the modelling workflow is depicted in Figure 4.1. The network reconstruction workflow is shown in Figure 1.5.

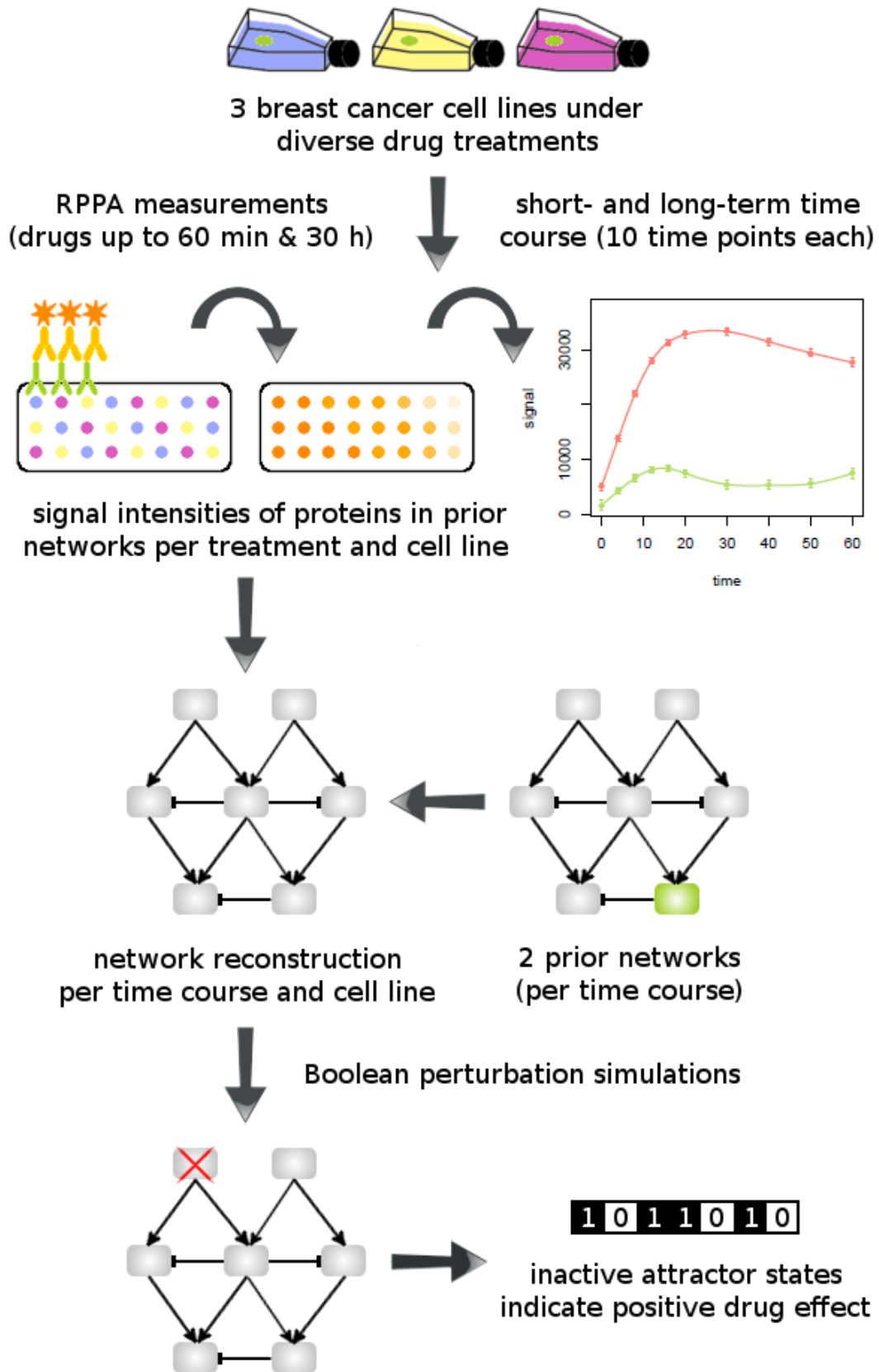


FIGURE 4.1. Scheme of the ErbB network modelling approach (von der Heyde et al., 2014a).

The reconstructed short-term networks are displayed in Figure 4.2. The reference network is marked in black, while the newly inferred edges are coloured according to the cell line they belong to. The individual network structures were compared to the reference network and between each other, respectively. The comparisons in short-term signalling pointed to cell line specific preferences for the MAPK and PI3K pathway, respectively. While BT474 showed a preference for the PI3K pathway, the novel edges in SKBR3 contributed mainly to signalling via the MAPK pathway. In HCC1954, the inferred edges contributed to both pathways to a similar extent. The different pathway preferences among the cell lines were also reflected in pathway specific drug response, driven by the type of growth factor stimulus.

In the perturbation simulations on the short-term networks, the PI3K and MAPK key regulators AKT, ERK1/2 and p70S6K (see Figure 1.1) were analysed. Main findings were that inhibition of PI3K signalling requires drug combinations, whereas inhibition of MAPK signalling mainly relies on erlotinib alone. Furthermore, the target p70S6K is influenced by both, PI3K and MAPK, pathways.

Special focus was put on drug resistance mechanisms in HCC1954 regarding the hyperactive PI3K pathway due to the known mutation. Indeed, feedback loops, hyperactive ErbB-1/2 heterodimers and pathway crosstalk were uncovered, amplifying PI3K signalling in this cell line. In addition to node perturbations, also edge perturbations were analysed in HCC1954 by removing each of the AKT stimulating edges outgoing from p70S6K, PDK1, mTOR and ErbB-3, alone or in combination. The resulting attractor states for the modified networks hinted at a less strong impact of mTOR on AKT, but indicated synergistic drug resistance potential of p70S6K, ErbB-3 and PDK1. Hence, one could regard the PIK3CA mutation as an edgetic mutation leading to AKT gain-of-function and contributing to trastuzumab inefficacy in HCC1954.

In the perturbation simulations on the long-term networks, RB and RPS6 as indicators for proliferation, cell cycle or tumour progression (Henjes et al., 2012) were analysed besides AKT and ERK1/2. The optimal long-term drug response for AKT and ERK1/2 mainly confirmed the short-term observations. Furthermore, a quick drug response was observed for RPS6 while RB showed a delayed response, which could be partially explained by newly inferred network edges. Also the long-term networks revealed feedback mechanisms as well as

novel edges or deletions which could contribute to drug resistance.

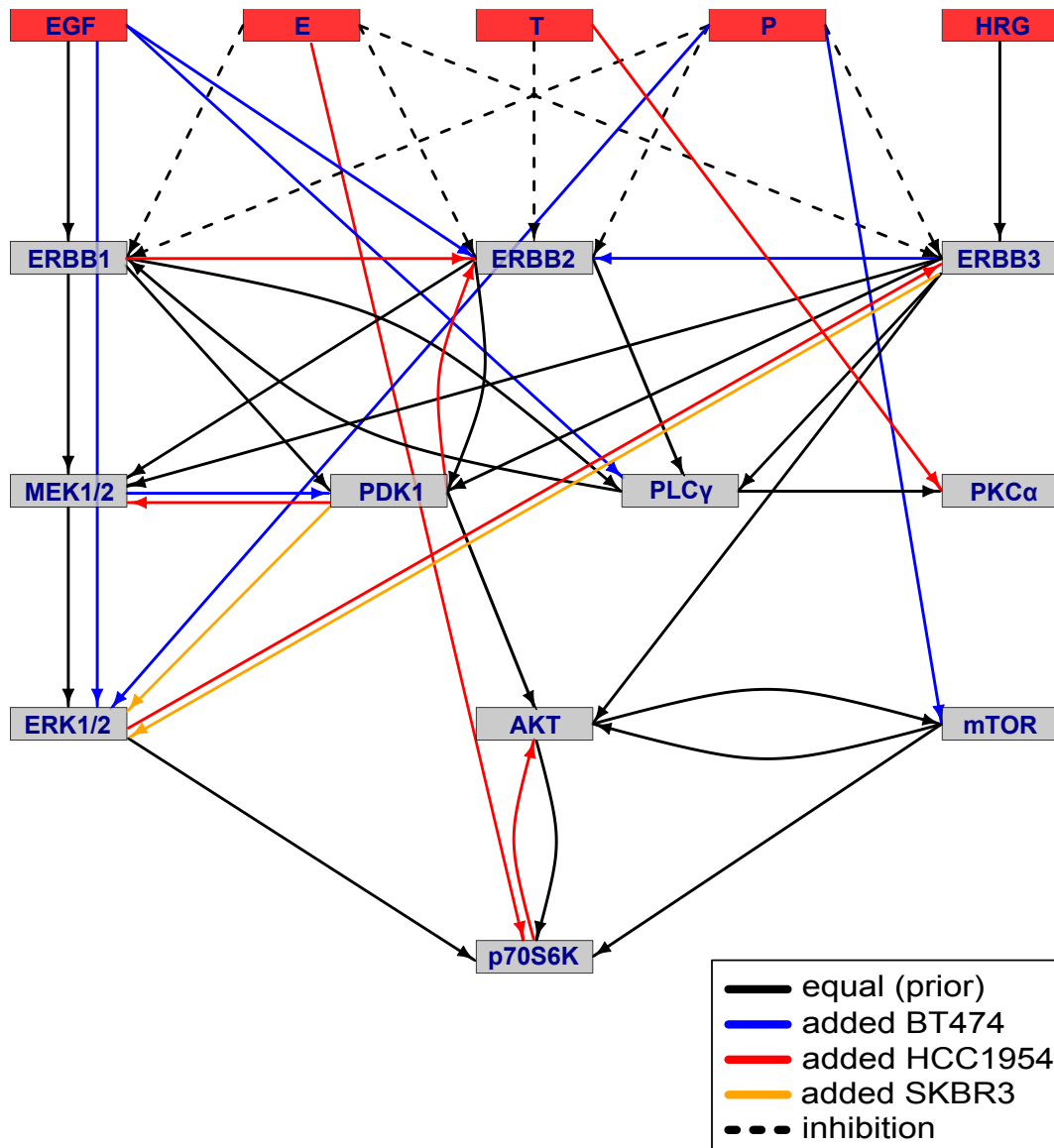


FIGURE 4.2. Reconstructed ErbB network models based on short-term measurements (von der Heyde et al., 2014a).

To sum up, protein interaction models for three breast cancer cell lines as representatives of different HER2-positive breast cancer phenotypes under short- and long-term drug treatment were reconstructed from data and literature. The conducted perturbation simulations to reveal optimal drug combinations were mostly consistent with the experimental data. This confirms combined

reverse and forward engineering techniques as valuable for drug discovery and personalized medicine. The network structures unveiled deregulated protein interactions in the ErbB signalling network, potentially contributing to drug resistance.

As proteomic signalling directly effects tumour development, it represents a promising target in targeted cancer therapy. Network models compose a first step to gain insight into individual mechanisms of drug response or resistance in breast cancer. Understanding such mechanisms in more detail will lead to optimized (combined) therapy approaches.

The detailed work is published in *BMC Systems Biology*, **2014**, 8:75, and attached in Appendix A.

4.1.2 Declaration of my contribution

I initiated the simulation study concepts together with Tim Beißbarth. I further carried out the literature research, the graphical and statistical analyses, the network reconstructions as well as the perturbation simulations followed by corresponding analyses.

Furthermore, I extended the *ddepn* software by an interface function called *adjacencyMatrix.to.logicalRules*, to directly combine reverse engineering via *ddepn* and forward engineering via *BoolNet*. The background was that *BoolNet* expects networks represented in form of logical interaction rules as input, whereas *ddepn* delivers the reconstructed networks in form of adjacency matrices. The new function translates the adjacency matrices to logical interaction rules of the network nodes.

Finally, I drafted the associated manuscript and made the data publicly available in GEO.

Frauke Henjes performed the RPPA measurements under supervision of Ulrike Korf and was mainly involved in target selection for the modelling approach. Johanna Sonntag was involved in discussions about the conducted RPPA experiments. Christian Bender and Tim Beißbarth developed the applied *ddepn* network reconstruction algorithm and participated in planning the modelling procedure. All authors edited, read and approved the final manuscript.

4.2 RPPanalyzer Toolbox: An improved R package for analysis of reverse phase protein array data

The manuscript on which this chapter is based has already been published.

Silvia von der Heyde, Johanna Sonntag, Daniel Kaschek, Christian Bender, Johannes Bues, Astrid Wachter, Jens Timmer, Ulrike Korf and Tim Beißbarth*, *RPPanalyzer* Toolbox: An improved R package for analysis of reverse phase protein array data, *BioTechniques*, **2014**, 57(3):125-135.

*Corresponding author

4.2.1 Summary and discussion

The intention of this work was to upgrade the *RPPanalyzer* open-source software package, originally developed by Mannsperger et al. (2010), since the analysis of proteomic large-scale RPPA data requires specialized software tools which can be tailored towards individual requirements. More precisely, the software extension arose out of practical user requirements at the German Cancer Research Center during collaborations.

The RPPA technology has already proven its potential for clinical applications, biomarker discovery and in systems biology (Hennessy et al., 2010; Gonzalez-Angulo et al., 2011; Uhlmann et al., 2012; Sonntag et al., 2014; Ummanni et al., 2014).

The R package *RPPanalyzer* provides functions for RPPA data preprocessing followed by basic statistical analyses and data visualization. It was especially designed for experimentalists to easily assess data quality and results. Its successful application has already been documented (Jozefczuk et al., 2012; Nelson et al., 2012; Sonntag et al., 2014; Ummanni et al., 2014). The enhanced version of the *RPPanalyzer* preserves its flexibility, but offers improved and streamlined standard preprocessing functions as well as new methods for variance estimation, normalization and visualization of time course data. The recommended workflow how to apply the *RPPanalyzer* is depicted in Figure 4.3.

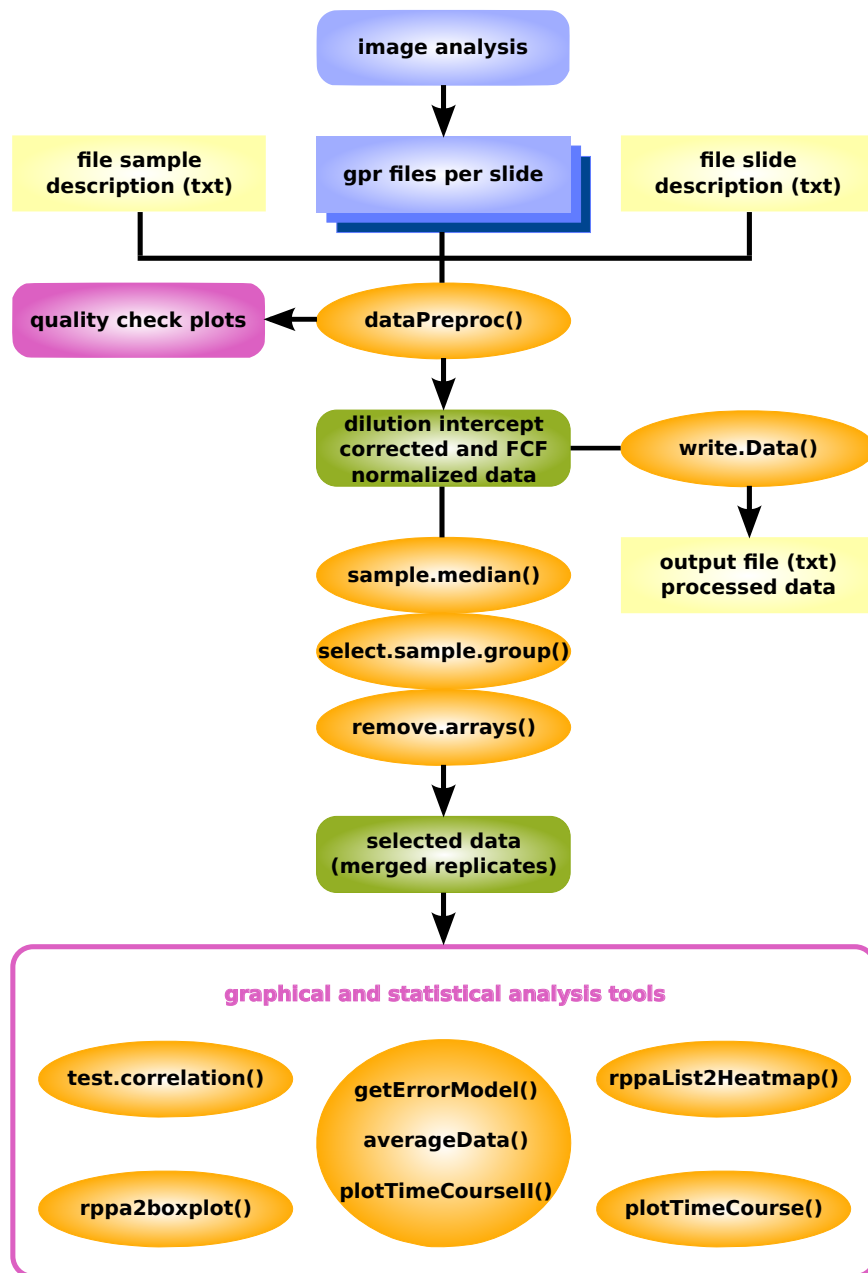


FIGURE 4.3. Workflow applying the RPPAnalyzer Toolbox (von der Heyde et al., 2014b).

Initial data preprocessing steps were merged into the new wrapper function `dataPreproc` leading to a standardized output.

In a first step, the raw data are imported from RPPA slide and sample description text files as well as image analysis result files in *gpr* format as obtained by

the commercial software GenePix[®] Pro (<http://www.moleculardevices.com>). The latter contain the spot-specific signal intensities as well as information about the array layout. The slide and sample description text files have to be provided by the user. They have to list the analysed target proteins and to contain sample-specific information, respectively. More detailed information about their set-up can be found in the corresponding user manual, i.e. the vignette of the *RPPanalyzer* R package.

The next step of the data preprocessing is to correct the data for background noise via the new method `correctDilinterc`. In detail, signal intensities are corrected by subtracting an intercept which is estimated for a total protein concentration of zero (0 $\mu\text{g}/\mu\text{l}$). This requires dilution series data. The main advantage of the new method is the independence of sample-wise dilution series, as correction factors are estimated based on just one representative. In the exemplary data set described in Chapter 1 in Section 1.2 the representative dilution series to normalize the A549 cell line includes triplicates of six concentrations starting at 2 $\mu\text{g}/\mu\text{l}$ (100%, 50%, 25%, 12.5%, 6.25%, 3.125%) for samples after 10 minutes of HGF (75 ng/ml) stimulation (see Figure 1.3 A). The actual samples cover measurements at six different time points ranging from 0 to 120 minutes, whereas the corresponding dilution series represents time point 10 minutes only. The actual samples are defined as *measurement* in the *sampledescription* text file while the serially diluted samples are defined as *control*. Additionally, the dilution series which should be used for background correction of certain samples of interest has to be flagged accordingly in this file, as usually several dilution series are measured, representing different cell lines or conditions. The selection of an appropriate dilution series strongly depends on the experimental setup.

In a first step, smoothing splines are applied to extrapolate the dilution series data to zero concentration. The uncertainty of this extrapolated intercept is estimated by non-parametric bootstrapping. The function provides plots of the dilution series and corresponding intercept estimations, as exemplarily shown in Figure 4.4.

A549 PRAS40 2691

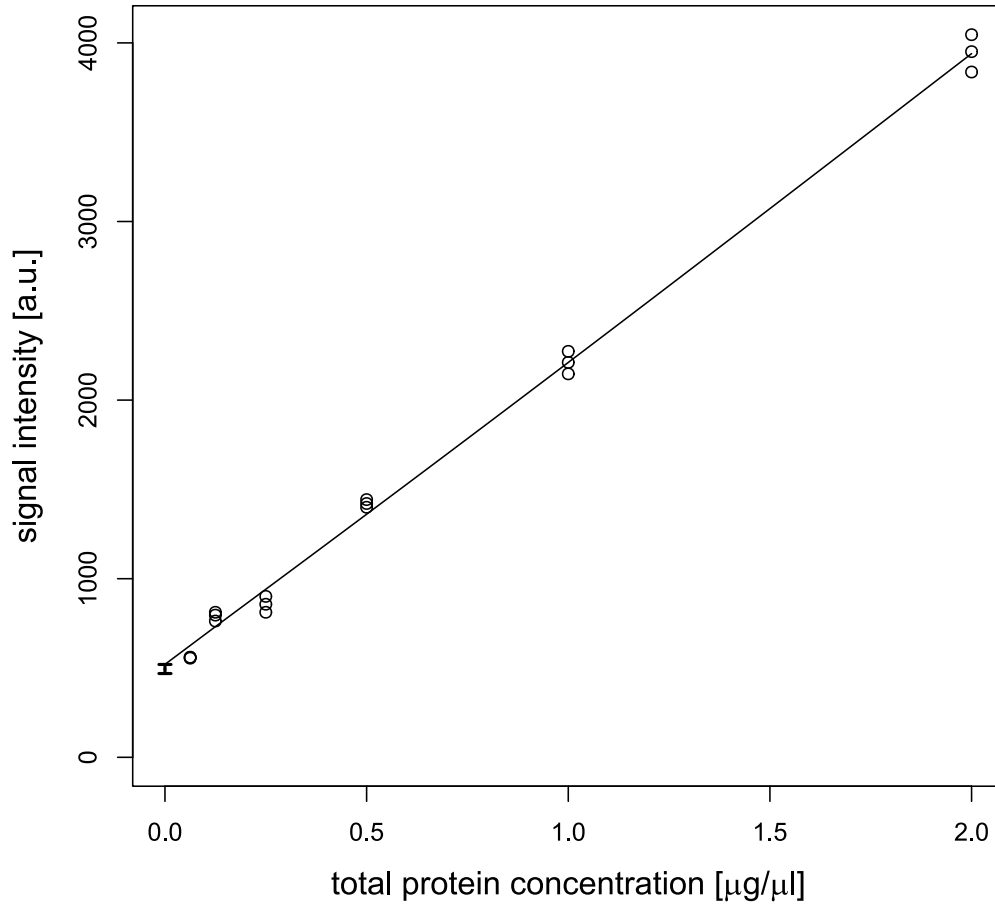


FIGURE 4.4. *Smoothing spline fit through dilution series data of the PRAS40 antibody (supplier code 2691) in the A549 cell line for extrapolation to zero (according to von der Heyde et al. (2014b)).*

The estimated intercepts \mathbf{I} are conditional on the detection antibody, the slide and the user-defined grouping factor, e.g. the cell line, of the samples which should be corrected by the same dilution series. Linear models are set up for the following hypotheses.

$$\mathbf{I} \sim \text{const.} \quad (1.1)$$

$$\mathbf{I} \sim \text{const.} + \text{antibody} \quad (1.2)$$

$$\mathbf{I} \sim \text{const.} + \text{antibody} + \text{slide} \quad (1.3)$$

$$\mathbf{I} \sim \text{const.} + \text{antibody} + \text{slide} + \text{group} \quad (1.4)$$

The models are weighted by the estimated uncertainties of the intercepts. A provided bar plot of the residual sum of squares (RSS) should be used to choose the model with the smallest RSS and least complexity. For example, if model 1.3 and 1.4 have the same smallest RSS, model 1.3 should be preferred, as the grouping factor in 1.4 does not provide additional information. Based on the chosen model, the noise of the actual *measurement* data at concentration zero is then predicted and subtracted from the raw intensities.

Subsequently, the background corrected signal intensities are normalized spot-wise by total protein concentration via the Fast Green FCF approach (Loebke et al., 2007) as explained in Chapter 1 in Section 1.2.

Apart from data-preprocessing the `dataPreproc` function generates plots to assess data quality. Visualization of the raw signal intensities of serially diluted control samples allows to check linearity of target protein detection. Furthermore, background noise corrected and FCF normalized target signals are plotted against the signal arising from the secondary antibody to detect bias caused by the secondary antibody. Finally, a quantile-quantile plot is generated to check whether the data are normally distributed.

The function returns matrices of the raw data, i.e. signal intensities as well as information about the samples and the arrays, matrices of the corrected data via `correctDilinterc` as well as matrices of the data additionally normalized by total protein content via Fast Green FCF. The raw data are further exported to a text file in table format. In general, data resulting from any of the processing steps can be exported manually in such a way via the `write.Data` function. After the data-preprocessing the next steps in the recommended workflow are to merge technical replicates via the `sample.median` function, to select data subsets of interest via the `select.sample.group` function and to remove data of target proteins which are not considered for further downstream analysis or data which did not pass the quality check via the `remove.arrays` function.

Data processed this way can then be analysed by applying diverse graphical and statistical methods of the *RPPanalyzer* which is not restricted to the suggested functions in Figure 4.3. The R package can be easily expanded and tailored to the specific user requirements by new or modified functions. The available and recommended methods are explained in the following.

The `test.correlation` function offers correlation plots to compare protein

expression and a numeric sample attribute of interest combined with statistical testing according to Pearson, Spearman or Kendall.

The function `rppa2boxplot` allows to visualize data as boxplots for sample groups of interest. Furthermore, statistical rank sum tests are applied. In case of testing against a reference group, two-sample Wilcoxon tests are performed with p-value adjustment for multiple testing according to Benjamini and Hochberg (1995). Otherwise, a Kruskal-Wallis rank sum test is applied to test for a general difference between all samples. An example for the data of MET receptor signalling described in Chapter 1 in Section 1.2 is shown in Figure 4.5. Figure 4.5 A displays a boxplot of MET receptor phosphorylation at site Y1349 at time point 15 minutes, depending on HGF concentration. The p-value resulting from a Kruskal-Wallis rank sum test is marked in green and indicates a significant difference between the signal intensities of the six sample groups. Figure 4.5 B shows a time course plot of MET receptor phosphorylation at site Y1349 over 120 minutes after HGF stimulation at different concentration levels.

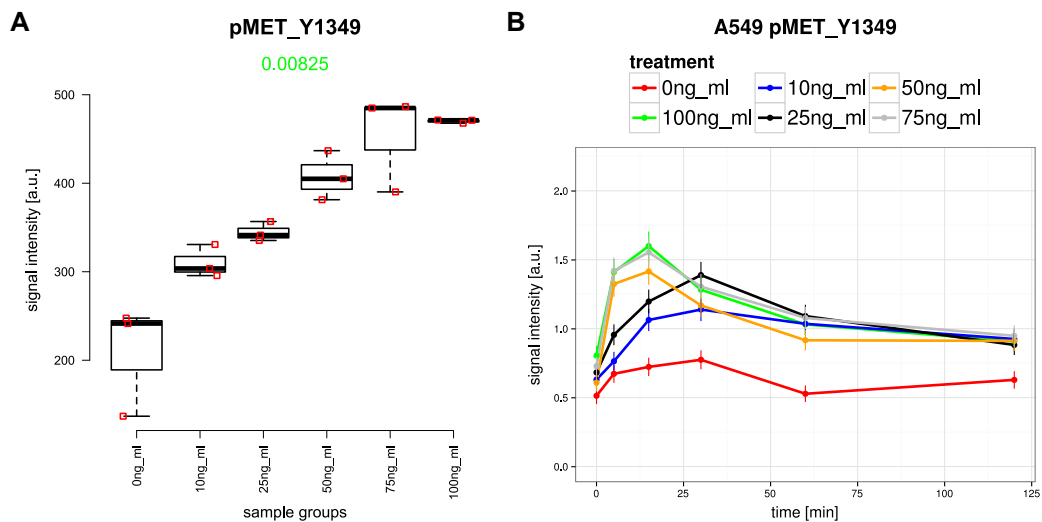


FIGURE 4.5. Exemplary graphics created with the *RPPAnalyzer* (according to von der Heyde et al. (2014b)).

While the boxplot function was just optimized regarding visual appearance and extended by the Wilcoxon testing, the time course plot was generated with the completely new function `plotTimeCourseII`. It was added as an alternative

to the already implemented function `plotTimeCourse` which offers different options for plotting time courses such as smoothed spline fits through the data. The new `plotTimeCourseII` function is applied after transformation of the preprocessed data by the new `getErrorModel` and `averageData` functions. This includes enhanced variance estimation which is appropriate for typical RPPA triplicate measurements.

The `getErrorModel` function estimates signal variance by combining information from triplicate measurements. Therefore, the parameters σ_0^2 and σ_{rel}^2 of the error model $\sigma^2 = \sigma_0^2 + S^2\sigma_{rel}^2$ are estimated in a maximum likelihood approach. The signal strength is represented by S . The variance at zero signal is represented by σ_0^2 . The relative error is represented by σ_{rel}^2 . The error model reflects the variance dependency on the signal strength. This further implies dependency on the subarrays and detection antibodies, as slides probed with different antibodies are scanned with different scanner settings to yield an optimal image for data analysis. Furthermore, the signal normalization via Fast Green FCF is related to the subarrays. Therefore, the parameter estimation is done for data subsets per slide and antibody. The model covers the situation when signal variances increase with signal strength as well as the case of constant or (almost) zero variance at zero signal.

The `averageData` function averages biological replicates which implies that lysate spots result from equal treatment of the samples. In contrast, technical replicates denote that the identical lysate is spotted several times on the same slide. The method assumes that a true signal dynamic $y_s(t)$ for a given stimulation s exists. The observed signal $S_{js}(t)$ for each biological replicate j on a specific slide under a specific stimulation s differs from $y_s(t)$ by a scaling factor s_j and biological variability. To estimate the true signal and the scaling factors per detection antibody, the objective function

$$\sum_{i,j,s} \frac{(s_j S_{ijs} - y_{is})^2}{s_j^2 \left[\sigma_{ijs,0}^2 + \left(\frac{y_{is}}{s_j} \right)^2 \sigma_{ijs,r}^2 \right]}$$

is minimized for the measurement time points t_i .

An already established method in the *RPPanalyzer* package to visualize clusters of samples in a heatmap is the `rppaList2Heatmap` function. In advance, the data can be logarithmized via the new `logList` function. Thus the data

look more normally distributed as required by statistical tests and the computation of distance measures. Furthermore, this leads to variance stabilization.

In conclusion, the upgrade of the *RPPanalyzer* package simplifies, enhances and standardizes the analysis of RPPA data. It is more user-friendly than the original version, as important functionalities for preprocessing are merged now. Furthermore, it is recommended how to integrate relevant functions in a standardized analysis workflow. This is concretely illustrated by an exemplary RPPA data set of longitudinal measurements of targets involved in MET receptor signalling upon stimulation with different HGF concentrations. Above that, the software has been extended by new functions and already established ones have been improved.

The detailed work is published in *BioTechniques*, **2014**, 57(3):125-135, and attached in Appendix B.

4.2.2 Declaration of my contribution

I developed the new wrapper function for data preprocessing and implemented all of the upgrades into the *RPPanalyzer* R package under supervision of Tim Beißbarth. Therefore, the functions had to be adapted to the *RPPanalyzer* format and the vignette had to be updated. Especially the new `correctDilinterc` function required a new default parameter in the sample description file to match samples with a specific dilution series. This parameter also had to be embedded in the function. I updated the boxplot function by a sorting option as well as a Wilcoxon test. Furthermore, I added the new `logList` function to logarithmize *RPPanalyzer* list elements. Above that, as one of the software package maintainers at *R-Forge* (<https://r-forge.r-project.org>), I corrected erroneous functions, solved compatibility problems and refined graphical representations in plots regarding labels, axes and margins.

Johannes Bues developed the new time course plotting function under supervision of Ulrike Korf. Daniel Kaschek developed the new functions for background correction as well as variance estimation under supervision of Jens Timmer. Johanna Sonntag, under supervision of Ulrike Korf, provided the RPPA example data and drafted together with me and Tim Beißbarth

the conceptual design of the software update. Christian Bender is the current maintainer of the R package at CRAN (<http://cran.r-project.org>). Astrid Wachter was involved in discussions and is currently one of the package maintainers at *R-Forge*. All authors contributed to conception, writing and editing of the manuscript.

4.3 mRNA Profiling Reveals Determinants of Trastuzumab Efficiency in HER2-Positive Breast Cancer

The manuscript on which this chapter is based has already been published.

Silvia von der Heyde, Steve Wagner, Alexander Czerny, Manuel Nietert, Fabian Ludewig, Gabriela Salinas-Riester, Dorit Arlt and Tim Beißbarth*, mRNA Profiling Reveals Determinants of Trastuzumab Efficiency in HER2-Positive Breast Cancer, *PLoS ONE*, **2015**, 10(2):e0117818.

*Corresponding author

4.3.1 Summary and discussion

The intention of this work was to infer genes and genetic variations (SNPs) affecting trastuzumab action and resistance in HER2-positive breast cancer cell lines. Therefore, three cell lines with different resistance phenotypes were analysed, namely BT474 as model of trastuzumab sensitivity, HCC1954 as model of intrinsic resistance, and BTR50, derived from BT474, as model of acquired resistance. Resistant BTR50 cells were developed by culturing parental BT474 cells in the presence of trastuzumab for around six months.

Differential expression analyses were performed on RNA-Seq data of these cell lines with and without trastuzumab treatment (GEO accession number GSE55005). In detail, five separate two-sample tests were performed on six samples, namely on the breast cancer cell lines with and without trastuzumab treatment. Two tests compared the gene expression in the resistant and the wild type cells, i.e. HCC1954 and BTR50 were compared to BT474. These

revealed 46 genes which are expected to contribute to trastuzumab resistance. The remaining three tests compared the gene expression in the untreated and trastuzumab treated cells, i.e. each of the three cell lines was compared to its trastuzumab treated version. The test for BT474 revealed 18 genes which are expected to contribute to trastuzumab efficiency. In general, significant ($\text{fdr} < 0.05$) differential expression was defined by a two-fold difference in gene expression. To exclude false positives from the combined set of 64 candidate genes, ten genes were excluded which were further differentially expressed in the test for BTR50 under trastuzumab treatment, as no drug effect is expected in the resistant cell line. The same would have held for HCC1954, but the related test revealed no genes overlapping with the candidate set.

This led to 54 genes which are very likely to determine trastuzumab efficiency in HER2-positive breast cancer. Many of these genes have already been discussed in the context of breast cancer and about ten percent especially in the context of trastuzumab action. These genes could contribute to develop novel strategies preventing trastuzumab resistance. 90% of 40 selected candidates were validated by real-time quantitative PCR (RT-qPCR). More than 85% of this validated selection of candidate genes were differentially expressed between the sensitive cell line and the intrinsically resistant cell line HCC1954. The remaining ones were differentially expressed between BT474 and its trastuzumab treated version.

The candidates of comparing BT474 and HCC1954 were mainly related to the different steroid receptor status of the cell lines. While BT474 is known to be positive for estrogen receptor (ER+) as well as for progesterone receptor (PR+) (Knutson et al., 2012), HCC1954 is negative for both receptors. These candidates included mainly tumour enhancers which were upregulated in HCC1954 and have already been ascribed to ER-negative (ER-) breast cancer subtypes. GDF15, IL8, LCN2, PTGS2 and 20 other genes were significantly higher expressed in HCC1954 than in BT474, while NCAM2, COLEC12, AFF3, TFF3, NRCAM, GREB1 and TFF1 were significantly lower expressed. Interestingly, the ER+ cell line BT474 resembles a more favourable luminal subtype of breast cancer, while the ER- cell line HCC1954 is similar to the more aggressive basal-like one (Gonçalves et al., 2008; Di Cello et al., 2013). This was confirmed by specific candidate genes, which have been reported in this context, such as S100A9 (Gonçalves et al., 2008) and CLDN1 (Di Cello et al., 2013) which were

overexpressed in HCC1954 and associated with basal subtypes.

ALPP, CALCOCO1, CAV1, CYP1A2 and IGFBP3 were significantly higher expressed in the trastuzumab treated than in the untreated BT474 cell line. Figure 4.6 and 4.7 show the expression fold changes of the validated genes.

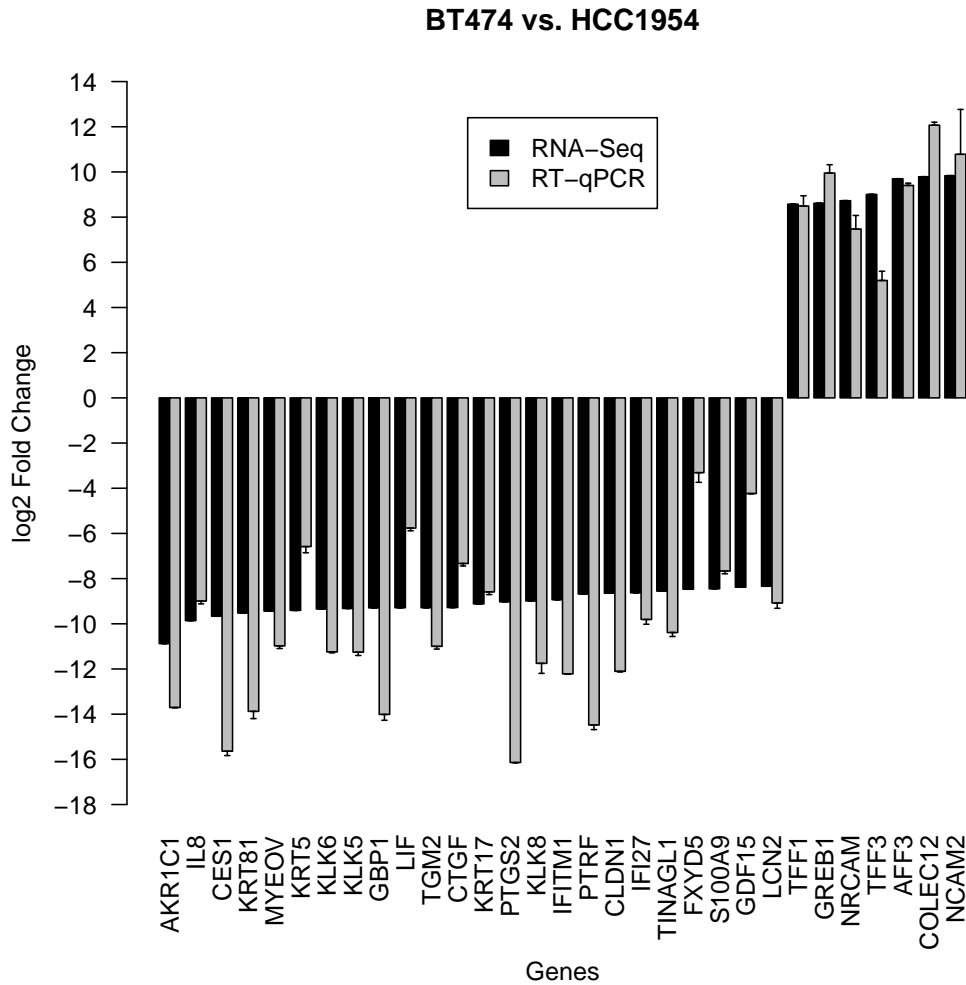


FIGURE 4.6. Fold Changes (\log_2) of differentially expressed genes (BT474 versus HCC1954) with positive values indicating an upregulation in BT474 (von der Heyde et al., 2015).

Intensive literature research was performed to elucidate the role of the candidate genes in affecting trastuzumab action. About 20% of the validated subset have already been discussed in the context of breast cancer and trastuzumab action. For example, IL8, PTGS2, GDF15 and LCN2 have already been reported to

hinder trastuzumab action. Interestingly, the PI3K pathway was considered to play a role in this context (Korkaya et al., 2012; Kumandan et al., 2013) which is in line with the PI3K gain-of-function mutation in the resistant cell line HCC1954. Furthermore, IL8 and LCN2 seem to cooperatively stabilize the resistant phenotype (Lin et al., 2011). Also PTGS2 and HER2 seem to cooperate within a positive feedback loop (Benoit et al., 2004; Wang et al., 2004).

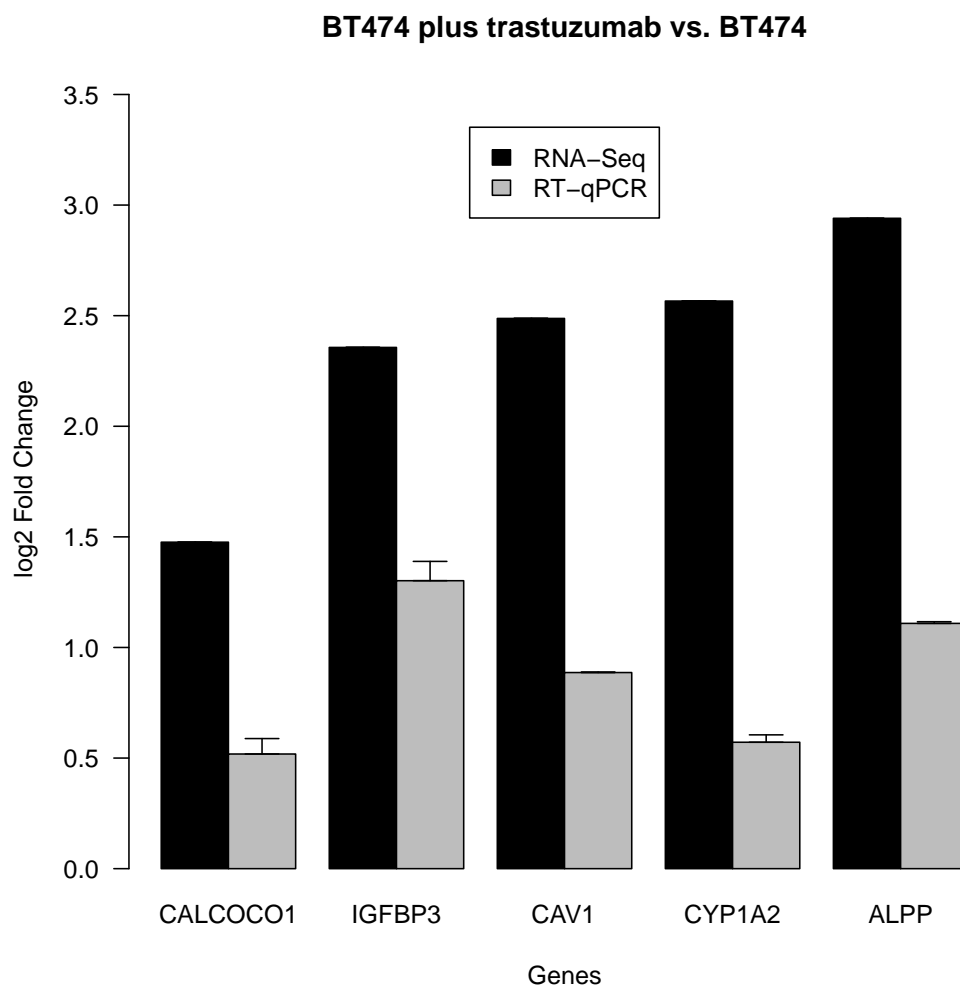


FIGURE 4.7. Fold Changes (\log_2) of differentially expressed genes (BT474 plus trastuzumab versus BT474) with positive values indicating an upregulation upon drug treatment (von der Heyde et al., 2015).

The two outstanding candidates revealed by comparing BT474 and its drug treated version were CAV1 and IGFBP3. Both genes have already been re-

ported to influence trastuzumab efficacy. Trastuzumab action seems to be supported by low CAV1 expression (Sekhar et al., 2013) but high IGFBP3 levels (Lu et al., 2001; Jerome et al., 2006; Dokmanovic et al., 2011).

A more general finding was a stronger discrepancy between HCC1954 and BT474 than between BTR50 and BT474. Furthermore, HCC1954 was influenced less by trastuzumab than BTR50. Hence, intrinsic resistance might be harder to overcome than acquired one. This has been deduced from the Principle Component Analysis (PCA) depicted in Figure 4.8.

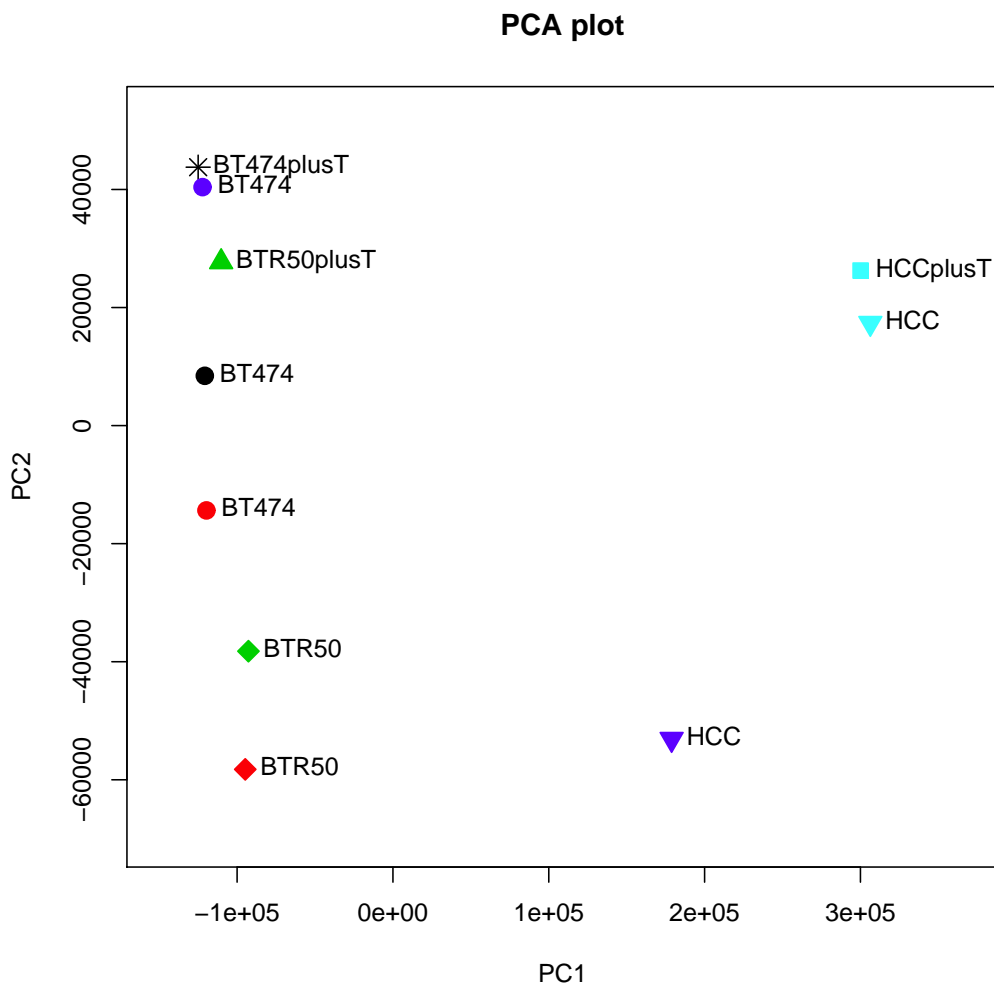


FIGURE 4.8. Principle Component Analysis (PCA) plot of all measured samples, in which same colours denote that samples belong to the same conducted statistical test (von der Heyde et al., 2015).

Interestingly, also more SNPs were detected in the intrinsically resistant cell line, which could contribute to the resistant phenotype. SNP calling and analysis within literature research was performed in the untreated cell lines BT474, HCC1954 and BTR50. Being aware of the definition mentioned in Chapter 1 in Section 1.2, the detected sequence variations which are potentially related to trastuzumab efficacy are called ‘SNPs’ in the following.

The SNP analysis focused on the validated gene candidates of the comparison between BT474 and its trastuzumab treated version (ALPP, CALCOCO1, CAV1, CYP1A2, IGFBP3) as well as the candidates of the comparison between BT474 and HCC1954 which were already reported in the context of trastuzumab efficacy, namely GDF15, IL8, LCN2 and PTGS2. Table 4.1 displays information about chromosomal location of the detected SNPs in the candidate genes of interest as well as corresponding SNP ID numbers (Sherry et al., 1999).

Gene	Chromosome	Position	Variation	Cell line
CAV1	7	116200554	rs374946197	HCC1954
IGFBP3	7	45952254	rs6670	BTR50
IGFBP3	7	45957678	rs3793345	HCC1954
IGFBP3	7	45960645	rs2854746	HCC1954
IL8	4	74609045	rs1126647	HCC1954
PTGS2	1	186641626	rs2853805	HCC1954
PTGS2	1	186642429	rs2206593	HCC1954
PTGS2	1	186643058	rs5275	HCC1954

TABLE 4.1. *Table of detected SNPs in the candidate genes.*

Additionally analysed genes were HER2 due to the HER2+ status of the cell lines, PIK3CA due to the known mutation in HCC1954, as well as MAPK1 and AKT1 as main players in the MAPK and PI3K/AKT signalling pathway, respectively. Table 4.2 displays information about the detected SNPs in the additional genes of interest. In case of missing SNP ID numbers (‘-’), a novel mutation could have been inferred.

Gene	Chromosome	Position	Variation	Cell line
AKT1	14	105235824	rs58565216	HCC1954
AKT1	14	105236287	rs35416681	HCC1954
AKT1	14	105241304	rs2230506	HCC1954
HER2	17	37855834	rs4252596	HCC1954
HER2	17	37859083	rs34284966	BTR50
HER2	17	37869895	-	HCC1954
HER2	17	37870378	rs2934967	HCC1954
HER2	17	37871081	-	HCC1954
HER2	17	37876179	rs4252639	BT474, BTR50
HER2	17	37876835	rs2952156	BTR50, HCC1954
HER2	17	37877221	-	BTR50
HER2	17	37877412	rs2952157	BT474, BTR50, HCC1954
HER2	17	37877447	rs11653998	HCC1954
HER2	17	37878113	rs115334808	BT474, BTR50
HER2	17	37878311	-	BTR50
HER2	17	37878371	-	BTR50, HCC1954
HER2	17	37878574	-	BTR50, HCC1954
HER2	17	37878635	-	BTR50, HCC1954
HER2	17	37878696	-	BTR50
HER2	17	37879030	rs2088126	BTR50, HCC1954
HER2	17	37884037	rs1058808, rs370420724	HCC1954
HER2	17	37885332	rs2952158	BTR50, HCC1954
MAPK1	22	22115004	rs6928	BT474, BTR50, HCC1954
MAPK1	22	22115353	rs9340	BTR50, HCC1954
MAPK1	22	22115498	rs3810610	BT474, BTR50, HCC1954
MAPK1	22	22115886	rs13515	BT474, BTR50
MAPK1	22	22116202	rs13943	BT474, BTR50, HCC1954
MAPK1	22	22116467	rs1063311	BT474, BTR50, HCC1954
MAPK1	22	22162072	-	BTR50
PIK3CA	3	178952085	rs121913279	HCC1954

TABLE 4.2. Table of detected SNPs in *AKT1*, *HER2*, *MAPK1* and *PIK3CA*.

For eight of the 13 genes of interest SNPs were detected according to pre-defined filter criteria in HCC1954. About 80% of these variations have already been listed in the *dbSNP* database (Sherry et al., 1999) but just about 40% of them have been analysed more deeply in literature, even less in the context of breast cancer. For HER2, 13 variations were inferred. Five variations were detected for MAPK1. Three variations were found for PTGS2 and AKT1, respectively. Two SNPs were called for IGFBP3, and one SNP was found for PIK3CA, CAV1 and IL8 each. The SNP rs2854746 of IGFBP3 in HCC1954 seems to influence IGFBP3 expression. It has already been reported that IGFBP3 is involved in breast carcinogenesis and associated with the SNP rs2854746 (Cheng et al., 2007; Patel et al., 2008; D'Aloisio et al., 2009; Su et al., 2010). Interestingly, IGFBP3 also had a variation in BTR50, namely rs6670, but it was not matching the ones of HCC1954, which points to a role in acquired resistance. For PIK3CA, the SNP rs121913279 was inferred which is referring to the known H1047R mutation in HCC1954 (Kataoka et al., 2010; Tong et al., 2012). This result supports usage of RNA-Seq data for SNP detection.

The detected mutations in BT474 formed a subset of the ones in BTR50, pointing at resistance related novel mutations in the cell line with acquired resistance. The detected mutations in BT474 included three variations for HER2 and five variations for MAPK1.

For three of the 13 genes of interest SNPs were detected in BTR50 which passed the filter criteria. For HER2, 13 variations were detected. Seven variations were inferred for MAPK1, and one variation was detected for IGFBP3. Two third of the SNPs have already been listed in the *dbSNP* database but just about 30% of those have been analysed more deeply in literature, even less in the context of breast cancer. Interestingly, four annotated SNPs in BTR50, which were not detected in BT474, overlapped with the SNPs in HCC1954. Three of these SNPs belonged to HER2, namely rs2952156, rs2088126 and rs2952158, and one belonged to MAPK1, namely rs9340. These could very likely contribute to trastuzumab resistance.

In principal, all of the variations listed in Table 4.1 and 4.2 could contribute to trastuzumab resistance and explain the different resistance cell phenotypes. Some inferred SNPs are already in the focus of breast cancer research, but mostly no influence on survival or risk has been detected for single ones. However, some seem to have an impact on the phenotype in combination with

further variations.

In summary, these results provide a basis for a deeper analysis of candidate genes and mutations within further experiments. More detailed information about mechanisms of trastuzumab action and resistance could help to predict the efficacy of trastuzumab therapy and hence improve stratification within personalized medicine approaches for HER2-positive breast cancer patients.

The detailed work is published in *PLoS ONE*, **2015**, 10(2):e0117818, and attached in Appendix C.

4.3.2 Declaration of my contribution

Tim Beißbarth, Dorit Arlt and me did the conceptual design of the experiments and data analysis. I performed the differential expression analysis on the RNA-Seq data. Above that, I interpreted the corresponding results of this analysis as well as of the SNP analysis, which was conducted by Alexander Czerny. The interpretation required intensive literature research and comparisons with external public data as well. Finally, I drafted the manuscript and made the data publicly available in GEO.

Manuel Nietert managed the data acquisition and was involved in discussions on differential expression analysis. Gabriela Salinas-Riester and Fabian Ludewig generated the RNA-Seq data. Steve Wagner cultivated and treated the cell lines and performed RT-qPCR validation.

Chapter 5

Outlook

Despite promising progress in precision oncology, intrinsic or acquired resistance remains problematic in personalized medicine. Modern technology allows to measure gene and protein expression in a way that becomes more and more efficient, i.e. faster and cheaper. This provides deeper insights into the molecular level of drug action and resistance mechanisms. The aim of the three studies which are summarized in this work was to contribute to a better understanding of such mechanisms in HER2-positive breast cancer by using existing technologies as well as by upgrading established software solutions.

The first study presented in this dissertation was based on proteomic time course measurements. Three HER2-positive breast cancer cell lines with different resistance phenotypes were treated with the drugs erlotinib, trastuzumab and pertuzumab alone or in combination, and the effect on several proteins of interest was analysed.

The reconstructed protein signalling network models per cell line revealed novel interactions which contributed to resistance mechanisms and partly explained the different cell line phenotypes by different pathway influences. As Boolean models represent a strong simplification of the actual underlying biology, the hypothetical results should be evaluated by additional experiments like protein knock-outs.

The network models were further used to simulate drug effects which were mainly confirmed by the input data. Also these results could be investigated further by additional experiments. For example, one could test whether other

techniques than RPPA would lead to the same results, especially not measuring drug influence over time but the homeostatic state of the system which was analysed in the simulations.

The second study presented in this thesis comprised a software extension for RPPA data processing and analysis. The R package *RPPanalyzer* was upgraded and streamlined leading to a more user-friendly handling.

Although a standardized workflow was suggested and key steps were applied on an example data set, R packages can be extended by any functions tailored to the specific needs of the users. Hence, the presented *RPPanalyzer Toolbox* won't be the final version of the software. The main functions for plotting and basic statistical analysis are provided, but higher sophisticated functions can be integrated at any time.

The third study presented in this dissertation was based on genomic measurements, i.e. RNA-Seq data providing genome-wide information. Three cell lines with different resistance phenotypes towards trastuzumab were analysed. Apart from a sensitive cell line, two resistant cell lines were analysed, of which one was intrinsically resistant due to a mutation, whereas the other one was cultured to resistance and hence represented acquired resistance.

Differential gene expression was revealed in the cell lines with and without trastuzumab treatment, leading to a set of candidate genes which are likely involved in trastuzumab action and resistance. Although the differentially expressed genes were mostly validated by RT-qPCR, it would be interesting to confirm the results within further experiments. Especially, several hypotheses were proposed regarding the genes' role in trastuzumab efficiency, which were mainly based on literature research and should be investigated in more detail. Furthermore, SNPs were detected in the untreated cell lines which could also play an important role for trastuzumab efficiency. As RNA-Seq data are usually not used for SNP calling, it would be of special interest to validate the presented results further.

Another interesting aspect of the study was the discrepancy between intrinsic and acquired resistance, which should be investigated in more detail. Revealing differences in drug susceptibility between specimens with intrinsic and acquired resistance would lead to better stratification in individualized therapeutic ap-

proaches to overcome resistance.

In conclusion, the presented studies in this thesis contribute to nowadays red-hot topic of personalized medicine with special focus on HER2-positive breast cancer and the corresponding therapeutic trastuzumab.

The related genomic and proteomic data analyses of different cancer phenotypes would benefit from additional experiments, either on the same data basis or in an analogous design, considering additional drugs or phenotypes, for example. Also the established protein signalling network models could be further validated and used for ongoing simulations. Finally, the provided software extension should be continuously upgraded according to user requirements and technology development.

Ever-expanding knowledge about molecular mechanisms of individual cancer development and drug resistance is indispensable for patient stratification in precision oncology and successful sequential application of biomarker-based targeted therapies.

References

- Alvarez-Chaver, P., Otero-Estevez, O., Paez de la Cadena, M., Rodriguez-Berrocal, F. J., and Martinez-Zorzano, V. S. (2014). Proteomics for discovery of candidate colorectal cancer biomarkers. *World Journal of Gastroenterology : WJG*, 20(14):3804–3824.
- Arteaga, C. L., Sliwkowski, M. X., Osborne, C. K., Perez, E. A., Puglisi, F., and Gianni, L. (2012). Treatment of HER2-positive breast cancer: current status and future perspectives. *Nature reviews. Clinical oncology*, 9(1):16–32.
- Bansal, M., Belcastro, V., Ambesi-Impiombato, A., and Bernardo, D. d. (2007). How to infer gene networks from expression profiles. *Molecular Systems Biology*, 3(1):78.
- Bender, C., Henjes, F., Fröhlich, H., Wiemann, S., Korf, U., and Beissbarth, T. (2010). Dynamic deterministic effects propagation networks: learning signalling pathways from longitudinal protein array data. *Bioinformatics (Oxford, England)*, 26(18):i596–602.
- Bender, C., Heyde, S. v., Henjes, F., Wiemann, S., Korf, U., and Beissbarth, T. (2011). Inferring signalling networks from longitudinal data using sampling based approaches in the R-package 'ddepn'. *BMC bioinformatics*, 12:291.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.
- Benoit, V., Relic, B., Leval Xd, X. d., Chariot, A., Merville, M.-P., and Bours, V. (2004). Regulation of HER-2 oncogene expression by cyclooxygenase-2 and prostaglandin E2. *Oncogene*, 23(8):1631–1635.

- Brookes, A. J. (1999). The essence of SNPs. *Gene*, 234(2):177–186.
- Chatr-Aryamontri, A., Breitkreutz, B.-J., Heinicke, S., Boucher, L., Winter, A., Stark, C., Nixon, J., Ramage, L., Kolas, N., O'Donnell, L., Reguly, T., Breitkreutz, A., Sellam, A., Chen, D., Chang, C., Rust, J., Livstone, M., Oughtred, R., Dolinski, K., and Tyers, M. (2013). The BioGRID interaction database: 2013 update. *Nucleic Acids Research*, 41(Database issue):D816–823.
- Chen, W. W., Schoeberl, B., Jasper, P. J., Niepel, M., Nielsen, U. B., Lauffenburger, D. A., and Sorger, P. K. (2009). Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Molecular systems biology*, 5:239.
- Cheng, I., DeLellis Henderson, K., Haiman, C. A., Kolonel, L. N., Henderson, B. E., Freedman, M. L., and Le Marchand, L. (2007). Genetic determinants of circulating insulin-like growth factor (IGF)-I, IGF binding protein (BP)-1, and IGFBP-3 levels in a multiethnic population. *The Journal of clinical endocrinology and metabolism*, 92(9):3660–3666.
- Cooper, G. M. (2000). *The Cell*. Sinauer Associates, 2nd edition.
- Crick, F. (1970). Central dogma of molecular biology. *Nature*, 227(5258):561–563.
- Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M. R., Jassal, B., Jupe, S., Matthews, L., May, B., Palatnik, S., Rothfels, K., Shamovsky, V., Song, H., Williams, M., Birney, E., Hermjakob, H., Stein, L., and D'Eustachio, P. (2014). The reactome pathway knowledgebase. *Nucleic Acids Research*, 42(Database issue):D472–477.
- D'Aloisio, A. A., Schroeder, J. C., North, K. E., Poole, C., West, S. L., Travlos, G. S., and Baird, D. D. (2009). IGF-I and IGFBP-3 polymorphisms in relation to circulating levels among African American and Caucasian women. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 18(3):954–966.

- Demuth, J. P. and Wade, M. J. (2006). Experimental methods for measuring gene interactions. *Annual Review of Ecology, Evolution, and Systematics*, 37(1):289–316.
- Di Cello, F., Cope, L., Li, H., Jeschke, J., Wang, W., Baylin, S. B., and Zahnow, C. A. (2013). Methylation of the claudin 1 promoter is associated with loss of expression in estrogen receptor positive breast cancer. *PLoS ONE*, 8(7):e68630.
- Dienstmann, R., De Dosso, S., Felip, E., and Tabernero, J. (2012). Drug development to overcome resistance to EGFR inhibitors in lung and colorectal cancer. *Molecular Oncology*, 6(1):15–26.
- Diermeier, S., Horváth, G., Knuechel-Clarke, R., Hofstaedter, F., Szöllosi, J., and Brockhoff, G. (2005). Epidermal growth factor receptor coexpression modulates susceptibility to herceptin in HER2/neu overexpressing breast cancer cells via specific erbB-receptor interaction and activation. *Experimental cell research*, 304(2):604–619.
- Dokmanovic, M., Shen, Y., Bonacci, T. M., Hirsch, D. S., and Wu, W. J. (2011). Trastuzumab regulates IGFBP-2 and IGFBP-3 to mediate growth inhibition: implications for the development of predictive biomarkers for trastuzumab resistance. *Molecular cancer therapeutics*, 10(6):917–928.
- Eduati, F., De Las Rivas, J., Di Camillo, B., Toffolo, G., and Saez-Rodriguez, J. (2012). Integrating literature-constrained and data-driven inference of signalling networks. *Bioinformatics (Oxford, England)*, 28(18):2311–2317.
- Esteva, F. J. and Pusztai, L. (2005). Optimizing outcomes in HER2-positive breast cancer: the molecular rationale. *Oncology (Williston Park, N.Y.)*, 19(13 Suppl 5):5–16.
- Federici, G., Gao, X., Slawek, J., Arodz, T., Shitaye, A., Wulfschlegel, J. D., De Maria, R., Liotta, L. A., and Petricoin, E. F. (2013). Systems analysis of the NCI-60 cancer cell lines by alignment of protein pathway activation modules with ”-OMIC” data fields and therapeutic response signatures. *Molecular cancer research: MCR*, 11(6):676–685.

- Feiglin, A., Hacothen, A., Sarusi, A., Fisher, J., Unger, R., and Ofran, Y. (2012). Static network structure can be used to model the phenotypic effects of perturbations in regulatory networks. *Bioinformatics (Oxford, England)*, 28(21):2811–2818.
- Ferlay, J., Shin, H.-R., Bray, F., Forman, D., Mathers, C., and Parkin, D. M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer*, 127(12):2893–2917.
- Figuroa-Magalhães, M. C., Jelovac, D., Connolly, R. M., and Wolff, A. C. (2013). Treatment of HER2-positive breast cancer. *The Breast*. PMID: 24360619.
- Fröhlich, H., Sahin, O., Arlt, D., Bender, C., and Beissbarth, T. (2009). Deterministic effects propagation networks for reconstructing protein signaling networks from multiple interventions. *BMC bioinformatics*, 10:322.
- Friedman, N., Murphy, K., and Russell, S. (1998). Learning the structure of dynamic probabilistic networks. In *Conf. on Uncertainty in Artificial Intelligence*, pages 139–147. Morgan Kaufmann.
- Gallardo, A., Lerma, E., Escuin, D., Tibau, A., Muñoz, J., Ojeda, B., Barnadas, A., Adrover, E., Sánchez-Tejada, L., Giner, D., Ortiz-Martínez, F., and Peiró, G. (2012). Increased signalling of EGFR and IGF1R, and deregulation of PTEN/PI3K/Akt pathway are related with trastuzumab resistance in HER2 breast carcinomas. *British journal of cancer*, 106(8):1367–1373.
- Galvez-Contreras, A. Y., Quinones-Hinojosa, A., and Gonzalez-Perez, O. (2013). The role of EGFR and ErbB family related proteins in the oligodendrocyte specification in germinal niches of the adult mammalian brain. *Frontiers in Cellular Neuroscience*, 7:258.
- Giordano, T. J., Shedden, K. A., Schwartz, D. R., Kuick, R., Taylor, J. M., Lee, N., Misek, D. E., Greenson, J. K., Kardia, S. L., Beer, D. G., Rennert, G., Cho, K. R., Gruber, S. B., Fearon, E. R., and Hanash, S. (2001). Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. *The American Journal of Pathology*, 159(4):1231–1238.

- Gonçalves, A., Charafe-Jauffret, E., Bertucci, F., Audebert, S., Toiron, Y., Esterni, B., Monville, F., Tarpin, C., Jacquemier, J., Houvenaeghel, G., Chabannon, C., Extra, J.-M., Viens, P., Borg, J.-P., and Birnbaum, D. (2008). Protein profiling of human breast tumor cells identifies novel biomarkers associated with molecular subtypes. *Molecular & cellular proteomics: MCP*, 7(8):1420–1433.
- Gonzalez-Angulo, A. M., Hennessy, B. T., Meric-Bernstam, F., Sahin, A., Liu, W., Ju, Z., Carey, M. S., Myhre, S., Speers, C., Deng, L., Broaddus, R., Lluch, A., Aparicio, S., Brown, P., Pusztai, L., Symmans, W. F., Alsner, J., Overgaard, J., Borresen-Dale, A.-L., Hortobagyi, G. N., Coombes, K. R., and Mills, G. B. (2011). Functional proteomics can define prognosis and predict pathologic complete response in patients with breast cancer. *Clinical Proteomics*, 8(1):11.
- Gray, I. C., Campbell, D. A., and Spurr, N. K. (2000). Single nucleotide polymorphisms as tools in human genetics. *Human molecular genetics*, 9(16):2403–2408.
- Greenbaum, D., Luscombe, N. M., Jansen, R., Qian, J., and Gerstein, M. (2001). Interrelating different types of genomic data, from proteome to secretome: 'oming in on function. *Genome Research*, 11(9):1463–1468.
- Hatakeyama, M., Kimura, S., Naka, T., Kawasaki, T., Yumoto, N., Ichikawa, M., Kim, J.-H., Saito, K., Saeki, M., Shirouzu, M., Yokoyama, S., and Konagaya, A. (2003). A computational model on the modulation of mitogen-activated protein kinase (MAPK) and Akt pathways in heregulin-induced ErbB signalling. *The Biochemical journal*, 373(Pt 2):451–463.
- Heil, J., Gondos, A., Rauch, G., Marmé, F., Rom, J., Golatta, M., Junkermann, H., Sinn, P., Aulmann, S., Debus, J., Hof, H., Schütz, F., Brenner, H., Sohn, C., and Schneeweiss, A. (2012). Outcome analysis of patients with primary breast cancer initially treated at a certified academic breast unit. *Breast (Edinburgh, Scotland)*, 21(3):303–308.
- Henjes, F., Bender, C., von der Heyde, S., Braun, L., Mannsperger, H. A., Schmidt, C., Wiemann, S., Hasmann, M., Aulmann, S., Beissbarth, T., and

- Korf, U. (2012). Strong EGFR signaling in cell line models of ERBB2-amplified breast cancer attenuates response towards ERBB2-targeting drugs. *Oncogenesis*, 1:e16.
- Henjes, F. M. (2010). *Analysis of ERBB signalling and the impact of targeted therapeutics using protein microarrays*. PhD thesis, Ruprecht-Karls-Universität Heidelberg, Heidelberg.
- Hennessy, B. T., Lu, Y., Gonzalez-Angulo, A. M., Carey, M. S., Myhre, S., Ju, Z., Davies, M. A., Liu, W., Coombes, K., Meric-Bernstam, F., Bedrosian, I., McGahren, M., Agarwal, R., Zhang, F., Overgaard, J., Alsner, J., Neve, R. M., Kuo, W.-L., Gray, J. W., Borresen-Dale, A.-L., and Mills, G. B. (2010). A technical assessment of the utility of reverse phase protein arrays for the study of the functional proteome in non-microdissected human breast cancers. *Clinical Proteomics*, 6(4):129–151.
- Hill, S. M., Lu, Y., Molina, J., Heiser, L. M., Spellman, P. T., Speed, T. P., Gray, J. W., Mills, G. B., and Mukherjee, S. (2012). Bayesian inference of signaling network topology in a cancer cell line. *Bioinformatics (Oxford, England)*, 28(21):2804–2810.
- Hornbeck, P. V., Kornhauser, J. M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Research*, 40(Database issue):D261–270.
- Hynes, N. E. and Lane, H. A. (2005). ERBB receptors and cancer: the complexity of targeted inhibitors. *Nature Reviews. Cancer*, 5(5):341–354.
- Jelovac, D. and Wolff, A. C. (2012). The adjuvant treatment of HER2-positive breast cancer. *Current treatment options in oncology*, 13(2):230–239.
- Jerome, L., Alami, N., Belanger, S., Page, V., Yu, Q., Paterson, J., Shiry, L., Pegram, M., and Leyland-Jones, B. (2006). Recombinant human insulin-like growth factor binding protein 3 inhibits growth of human epidermal growth factor receptor-2-overexpressing breast tumors and potentiates herceptin activity in vivo. *Cancer research*, 66(14):7245–7252.

- Jones, R. B., Gordus, A., Krall, J. A., and MacBeath, G. (2006). A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature*, 439(7073):168–174.
- Jozefczuk, J., Kashofer, K., Ummanni, R., Henjes, F., Rehman, S., Geenen, S., Wruck, W., Regenbrecht, C., Daskalaki, A., Wierling, C., Turano, P., Bertini, I., Korf, U., Zatloukal, K., Westerhoff, H. V., Lehrach, H., and Adjaye, J. (2012). A systems biology approach to deciphering the etiology of steatosis employing patient-derived dermal fibroblasts and iPS cells. *Frontiers in Physiology*, 3.
- Kataoka, Y., Mukohara, T., Shimada, H., Saijo, N., Hirai, M., and Minami, H. (2010). Association between gain-of-function mutations in PIK3CA and resistance to HER2-targeted agents in HER2-amplified breast cancer cell lines. *Annals of oncology: official journal of the European Society for Medical Oncology / ESMO*, 21(2):255–262.
- Klos, K. S., Wyszomierski, S. L., Sun, M., Tan, M., Zhou, X., Li, P., Yang, W., Yin, G., Hittelman, W. N., and Yu, D. (2006). ErbB2 increases vascular endothelial growth factor protein synthesis via activation of mammalian target of rapamycin/p70S6K leading to increased angiogenesis and spontaneous metastasis of human breast cancer cells. *Cancer research*, 66(4):2028–2037.
- Knutson, T. P., Daniel, A. R., Fan, D., Silverstein, K. A., Covington, K. R., Fuqua, S. A., and Lange, C. A. (2012). Phosphorylated and sumoylation-deficient progesterone receptors drive proliferative gene signatures during breast cancer progression. *Breast cancer research: BCR*, 14(3):R95.
- Kolch, W., Heidecker, G., Kochs, G., Hummel, R., Vahidi, H., Mischak, H., Finkenzeller, G., Marmé, D., and Rapp, U. R. (1993). Protein kinase C alpha activates RAF-1 by direct phosphorylation. *Nature*, 364(6434):249–252.
- Kong, D.-x. and Yamori, T. (2010). ZSTK474, a novel phosphatidylinositol 3-kinase inhibitor identified using the JFCR39 drug discovery system. *Acta Pharmacologica Sinica*, 31(9):1189–1197.
- Korkaya, H., Kim, G.-I., Davis, A., Malik, F., Henry, N. L., Ithimakin, S., Quraishi, A. A., Tawakkol, N., D’Angelo, R., Paulson, A. K., Chung, S.,

- Luther, T., Paholak, H. J., Liu, S., Hassan, K. A., Zen, Q., Clouthier, S. G., and Wicha, M. S. (2012). Activation of an IL6 inflammatory loop mediates trastuzumab resistance in HER2+ breast cancer by expanding the cancer stem cell population. *Molecular cell*, 47(4):570–584.
- Kumandan, S., Mahadevan, N. R., Chiu, K., DeLaney, A., and Zanetti, M. (2013). Activation of the unfolded protein response bypasses trastuzumab-mediated inhibition of the PI-3K pathway. *Cancer letters*, 329(2):236–242.
- Kwok, P.-Y. and Chen, X. (2003). Detection of single nucleotide polymorphisms. *Current issues in molecular biology*, 5(2):43–60.
- Lee, W.-P. and Tzou, W.-S. (2009). Computational methods for discovering gene networks from expression data. *Briefings in Bioinformatics*, 10(4):408–423.
- Lin, H.-H., Liao, C.-J., Lee, Y.-C., Hu, K.-H., Meng, H.-W., and Chu, S.-T. (2011). Lipocalin-2-induced cytokine production enhances endometrial carcinoma cell survival and migration. *International journal of biological sciences*, 7(1):74–86.
- Loebke, C., Sueltmann, H., Schmidt, C., Henjes, F., Wiemann, S., Poustka, A., and Korf, U. (2007). Infrared-based protein detection arrays for quantitative proteomics. *Proteomics*, 7(4):558–564.
- Lu, Y., Zi, X., Zhao, Y., Mascarenhas, D., and Pollak, M. (2001). Insulin-like growth factor-I receptor signaling and resistance to trastuzumab (Herceptin). *Journal of the National Cancer Institute*, 93(24):1852–1857.
- Malone, J. H. and Oliver, B. (2011). Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biology*, 9(1):34.
- Mannsperger, H. A., Gade, S., Henjes, F., Beissbarth, T., and Korf, U. (2010). RPPanalyzer: Analysis of reverse-phase protein array data. *Bioinformatics (Oxford, England)*, 26(17):2202–2203.
- McDermott, J. E., Wang, J., Mitchell, H., Webb-Robertson, B.-J., Hafen, R., Ramey, J., and Rodland, K. D. (2013). Challenges in biomarker discovery:

- Combining expert insights with statistical analysis of complex omics data. *Expert opinion on medical diagnostics*, 7(1):37–51.
- Mikalsen, T., Gerits, N., and Moens, U. (2006). Inhibitors of signal transduction protein kinases as targets for cancer therapy. In El-Gewely, M. R., editor, *Biotechnology Annual Review*, volume 12, page 153–223. Elsevier.
- Moore, M. J., Goldstein, D., Hamm, J., Figer, A., Hecht, J. R., Gallinger, S., Au, H. J., Murawa, P., Walde, D., Wolff, R. A., Campos, D., Lim, R., Ding, K., Clark, G., Voskoglou-Nomikos, T., Ptasynski, M., Parulekar, W., and Group, N. C. I. o. C. C. T. (2007). Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the national cancer institute of canada clinical trials group. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 25(15):1960–1966.
- Motoyama, A. B., Hynes, N. E., and Lane, H. A. (2002). The efficacy of ErbB receptor-targeted anticancer therapeutics is influenced by the availability of epidermal growth factor-related peptides. *Cancer Research*, 62(11):3151–3158.
- Murphy, K. and Mian, S. (1999). Modelling gene expression data using dynamic bayesian networks. Technical report, Computer Science Division, University of California, Berkeley, CA.
- Müssel, C., Hopfensitz, M., and Kestler, H. A. (2010). BoolNet—an R package for generation, reconstruction and analysis of Boolean networks. *Bioinformatics (Oxford, England)*, 26(10):1378–1380.
- Nahta, R., Yuan, L. X. H., Zhang, B., Kobayashi, R., and Esteva, F. J. (2005). Insulin-like growth factor-I receptor/human epidermal growth factor receptor 2 heterodimerization contributes to trastuzumab resistance of breast cancer cells. *Cancer research*, 65(23):11118–11128.
- Nelson, L. D., Bender, C., Mannsperger, H., Buergy, D., Kambakamba, P., Mudduluru, G., Korf, U., Hughes, D., Dyke, M. W. V., and Allgayer, H. (2012). Triplex DNA-binding proteins are associated with clinical outcomes revealed by proteomic measurements in patients with colorectal cancer. *Molecular Cancer*, 11(1):38.

- Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E., Bamshad, M., Nickerson, D. A., and Shendure, J. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 461(7261):272–276.
- O’Brien, N. A., Browne, B. C., Chow, L., Wang, Y., Ginther, C., Arboleda, J., Duffy, M. J., Crown, J., O’Donovan, N., and Slamon, D. J. (2010). Activated phosphoinositide 3-kinase/AKT signaling confers resistance to trastuzumab but not lapatinib. *Molecular cancer therapeutics*, 9(6):1489–1502.
- Oda, K., Matsuoka, Y., Funahashi, A., and Kitano, H. (2005). A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular systems biology*, 1:2005.0010.
- Olayioye, M. A., Neve, R. M., Lane, H. A., and Hynes, N. E. (2000). The ErbB signaling network: receptor heterodimerization in development and cancer. *The EMBO journal*, 19(13):3159–3167.
- Pallis, A. G. and Syrigos, K. N. (2013). Epidermal growth factor receptor tyrosine kinase inhibitors in the treatment of NSCLC. *Lung cancer (Amsterdam, Netherlands)*. PMID: 23384674.
- Park, J. W., Neve, R. M., Szollosi, J., and Benz, C. C. (2008). Unraveling the biologic and clinical complexities of HER2. *Clinical breast cancer*, 8(5):392–401.
- Park, Y. and Bader, J. S. (2012). How networks change with time. *Bioinformatics (Oxford, England)*, 28(12):i40–48.
- Patel, A. V., Cheng, I., Canzian, F., Le Marchand, L., Thun, M. J., Berg, C. D., Buring, J., Calle, E. E., Chanock, S., Clavel-Chapelon, F., Cox, D. G., Dorronsoro, M., Dossus, L., Haiman, C. A., Hankinson, S. E., Henderson, B. E., Hoover, R., Hunter, D. J., Kaaks, R., Kolonel, L. N., Kraft, P., Linseisen, J., Lund, E., Manjer, J., McCarty, C., Peeters, P. H. M., Pike, M. C., Pollak, M., Riboli, E., Stram, D. O., Tjonneland, A., Travis, R. C., Trichopoulos, D., Tumino, R., Yeager, M., Ziegler, R. G., Feigelson, H. S., and Breast and Prostate Cancer Cohort Consortium (2008). IGF-1, IGFBP-1, and IGFBP-3 polymorphisms predict circulating IGF levels but not breast

- cancer risk: findings from the breast and prostate cancer cohort consortium (BPC3). *PLoS ONE*, 3(7):e2578.
- Paweletz, C. P., Charboneau, L., Bichsel, V. E., Simone, N. L., Chen, T., Gillespie, J. W., Emmert-Buck, M. R., Roth, M. J., Petricoin III, E. F., and Liotta, L. A. (2001). Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene*, 20(16):1981–1989.
- Penfold, C. A., Buchanan-Wollaston, V., Denby, K. J., and Wild, D. L. (2012). Nonparametric Bayesian inference for perturbed and orthologous gene regulatory networks. *Bioinformatics (Oxford, England)*, 28(12):i233–241.
- Quinn, E. M., Cormican, P., Kenny, E. M., Hill, M., Anney, R., Gill, M., Corvin, A. P., and Morris, D. W. (2013). Development of strategies for SNP detection in RNA-Seq data: Application to lymphoblastoid cell lines and evaluation using 1000 genomes data. *PLoS ONE*, 8(3).
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rao, V. S., Srinivas, K., Sujini, G. N., and Kumar, G. N. S. (2014). Protein-protein interaction detection: methods and analysis. *International Journal of Proteomics*, 2014:147648.
- Reynolds, K., Sarangi, S., Bardia, A., and Dizon, D. S. (2014). Precision medicine and personalized breast cancer: combination pertuzumab therapy. *Pharmacogenomics and Personalized Medicine*, 7:95–105.
- Rodriguez, B. A. T. and Huang, T. H.-M. (2005). Tilling the chromatin landscape: emerging methods for the discovery and profiling of protein-DNA interactions. *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire*, 83(4):525–534.
- Roskoski, Jr, R. (2014). The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacological research: the official journal of the Italian Pharmacological Society*, 79:34–74.

- Roukos, D. H. (2011). Trastuzumab and beyond: sequencing cancer genomes and predicting molecular networks. *The pharmacogenomics journal*, 11(2):81–92.
- Sahin, O., Fröhlich, H., Löbke, C., Korf, U., Burmester, S., Majety, M., Mattern, J., Schupp, I., Chaouiya, C., Thieffry, D., Poustka, A., Wiemann, S., Beissbarth, T., and Arlt, D. (2009). Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. *BMC systems biology*, 3:1.
- Sahin, O., Wang, Q., Brady, S. W., Ellis, K., Wang, H., Chang, C.-C., Zhang, Q., Priya, P., Zhu, R., Wong, S. T., Landis, M. D., Muller, W. J., Esteva, F. J., Chang, J., and Yu, D. (2014). Biomarker-guided sequential targeted therapies to overcome therapy resistance in rapidly evolving highly aggressive mammary tumors. *Cell Research*, 24(5):542–559.
- Samaga, R., Saez-Rodriguez, J., Alexopoulos, L. G., Sorger, P. K., and Klamt, S. (2009). The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS computational biology*, 5(8):e1000438.
- Scaltriti, M. and Baselga, J. (2006). The epidermal growth factor receptor pathway: A model for targeted therapy. *Clinical Cancer Research*, 12(18):5268–5272.
- Schaefer, C. F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., and Buetow, K. H. (2009). PID: the pathway interaction database. *Nucleic Acids Research*, 37(Database issue):D674–679.
- Schoeberl, B., Eichler-Jonsson, C., Gilles, E. D., and Müller, G. (2002). Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nature biotechnology*, 20(4):370–375.
- Sekhar, S. C., Kasai, T., Satoh, A., Shigehiro, T., Mizutani, A., Murakami, H., El-Aarag, B. Y., Salomon, D. S., Massaguer, A., de Llorens, R., and Seno, M. (2013). Identification of caveolin-1 as a potential causative factor in the generation of trastuzumab resistance in breast cancer cells. *Journal of Cancer*, 4(5):391–401.

- Sherry, S. T., Ward, M., and Sirotkin, K. (1999). dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome research*, 9(8):677–679.
- Sonntag, J., Bender, C., Soons, Z., von der Heyde, S., König, R., Wiemann, S., Sinn, H.-P., Schneeweiss, A., Beißbarth, T., and Korf, U. (2014). Reverse phase protein array based tumor profiling identifies a biomarker signature for risk classification of hormone receptor-positive breast cancer. *Translational Proteomics*, 2:52–59.
- Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lønning, P. E., Brown, P. O., Børresen-Dale, A.-L., and Botstein, D. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences of the United States of America*, 100(14):8418–8423.
- Su, X., Colditz, G. A., Willett, W. C., Collins, L. C., Schnitt, S. J., Connolly, J. L., Pollak, M. N., Rosner, B., and Tamimi, R. M. (2010). Genetic variation and circulating levels of IGF-I and IGFBP-3 in relation to risk of proliferative benign breast disease. *International journal of cancer. Journal international du cancer*, 126(1):180–190.
- Terfve, C., Cokelaer, T., Henriques, D., MacNamara, A., Goncalves, E., Morris, M. K., van Iersel, M., Lauffenburger, D. A., and Saez-Rodriguez, J. (2012). CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. *BMC systems biology*, 6:133.
- Tessari, A., Palmieri, D., and Di Cosimo, S. (2013). Overview of diagnostic/targeted treatment combinations in personalized medicine for breast cancer patients. *Pharmacogenomics and Personalized Medicine*, 7:1–19.
- Tinoco, G., Warsch, S., Glück, S., Avancha, K., and Montero, A. J. (2013). Treating breast cancer in the 21st century: emerging biological therapies. *Journal of Cancer*, 4(2):117–132.

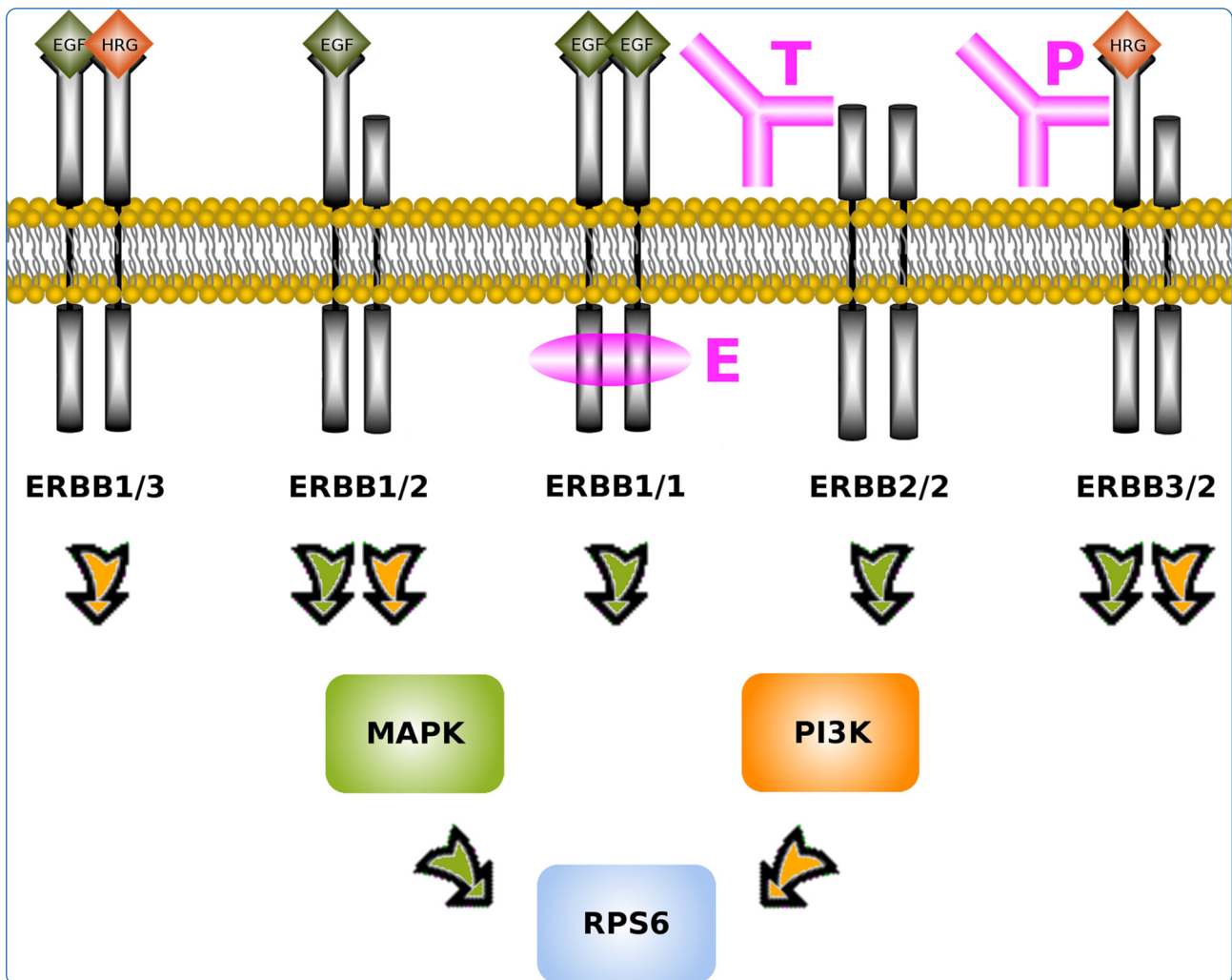
- Tong, L., Yang, X.-X., Liu, M.-F., Yao, G.-Y., Dong, J.-Y., Ye, C.-S., and Li, M. (2012). Mutational analysis of key EGFR pathway genes in Chinese breast cancer patients. *Asian Pacific journal of cancer prevention: APJCP*, 13(11):5599–5603.
- Uhlmann, S., Mannsperger, H., Zhang, J. D., Horvat, E.-g., Schmidt, C., Küblbeck, M., Henjes, F., Ward, A., Tschulena, U., Zweig, K., Korf, U., Wiemann, S., and Sahin, O. (2012). Global microRNA level regulation of EGFR-driven cell-cycle protein network in breast cancer. *Molecular Systems Biology*, 8:570.
- Ummanni, R., Mannsperger, H. A., Sonntag, J., Oswald, M., Sharma, A. K., König, R., and Korf, U. (2014). Evaluation of reverse phase protein array (RPPA)-based pathway-activation profiling in 84 non-small cell lung cancer (NSCLC) cell lines as platform for cancer proteomics and biomarker discovery. *Biochimica Et Biophysica Acta*, 1844(5):950–959.
- Valabrega, G., Montemurro, F., and Aglietta, M. (2007). Trastuzumab: mechanism of action, resistance and future perspectives in HER2-overexpressing breast cancer. *Annals of oncology: official journal of the European Society for Medical Oncology / ESMO*, 18(6):977–984.
- von der Heyde, S. and Beissbarth, T. (2012). A new analysis approach of epidermal growth factor receptor pathway activation patterns provides insights into cetuximab resistance mechanisms in head and neck cancer. *BMC Medicine*, 10(1):43.
- von der Heyde, S., Bender, C., Henjes, F., Sonntag, J., Korf, U., and Beißbarth, T. (2014a). Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines. *BMC Systems Biology*, 8(1):75.
- von der Heyde, S., Sonntag, J., Kaschek, D., Bender, C., Bues, J., Wachter, A., Timmer, J., Korf, U., and Beißbarth, T. (2014b). RPPAnalyzer Toolbox: An improved R package for analysis of reverse phase protein array data. *BioTechniques*, 57(3):125–135.

- von der Heyde, S., Wagner, S., Czerny, A., Nietert, M., Ludewig, F., Salinas-Riester, G., Arlt, D., and Beißbarth, T. (2015). mRNA Profiling Reveals Determinants of Trastuzumab Efficiency in HER2-Positive Breast Cancer. *PLoS ONE*, 10(2):e0117818.
- Wagner, J. P., Wolf-Yadlin, A., Sevecka, M., Grenier, J. K., Root, D. E., Lauffenburger, D. A., and MacBeath, G. (2013). Receptor tyrosine kinases fall into distinct classes based on their inferred signaling networks. *Science Signaling*, 6(284):ra58.
- Wang, K., Huang, C., and Nice, E. (2014). Recent advances in proteomics: towards the human proteome. *Biomedical chromatography: BMC*, 28(6):848–857.
- Wang, L., Zhang, Q., Zhang, J., Sun, S., Guo, H., Jia, Z., Wang, B., Shao, Z., Wang, Z., and Hu, X. (2011). PI3K pathway activation results in low efficacy of both trastuzumab and lapatinib. *BMC cancer*, 11:248.
- Wang, R.-S., Saadatpour, A., and Albert, R. (2012). Boolean modeling in systems biology: an overview of methodology and applications. *Physical Biology*, 9(5):055001.
- Wang, S.-C., Lien, H.-C., Xia, W., Chen, I.-F., Lo, H.-W., Wang, Z., Ali-Seyed, M., Lee, D.-F., Bartholomeusz, G., Ou-Yang, F., Giri, D. K., and Hung, M.-C. (2004). Binding at and transactivation of the COX-2 promoter by nuclear tyrosine kinase receptor ErbB-2. *Cancer cell*, 6(3):251–261.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63.
- Wieduwilt, M. J. and Moasser, M. M. (2008). The epidermal growth factor receptor family: biology driving targeted therapeutics. *Cellular and molecular life sciences: CMLS*, 65(10):1566–1584.
- Wilson, T. R., Fridlyand, J., Yan, Y., Penuel, E., Burton, L., Chan, E., Peng, J., Lin, E., Wang, Y., Sosman, J., Ribas, A., Li, J., Moffat, J., Sutherlin, D. P., Koeppen, H., Merchant, M., Neve, R., and Settleman, J. (2012). Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors. *Nature*, 487(7408):505–509.

- Workman, P., Al-Lazikani, B., and Clarke, P. A. (2013). Genome-based cancer therapeutics: targets, kinase drug resistance and future strategies for precision oncology. *Current Opinion in Pharmacology*, 13(4):486–496.
- Xenarios, I., Salwinski, L., Duan, X. J., Higney, P., Kim, S.-M., and Eisenberg, D. (2002). DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Research*, 30(1):303–305.

Appendix A

Publication “Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines”



Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines

von der Heyde *et al.*

RESEARCH ARTICLE

Open Access

Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines

Silvia von der Heyde¹, Christian Bender², Frauke Henjes³, Johanna Sonntag⁴, Ulrike Korf⁴
and Tim Beißbarth^{1*}

Abstract

Background: Despite promising progress in targeted breast cancer therapy, drug resistance remains challenging. The monoclonal antibody drugs trastuzumab and pertuzumab as well as the small molecule inhibitor erlotinib were designed to prevent ErbB-2 and ErbB-1 receptor induced deregulated protein signalling, contributing to tumour progression. The oncogenic potential of ErbB receptors unfolds in case of overexpression or mutations. Dimerisation with other receptors allows to bypass pathway blockades. Our intention is to reconstruct the ErbB network to reveal resistance mechanisms. We used longitudinal proteomic data of ErbB receptors and downstream targets in the ErbB-2 amplified breast cancer cell lines BT474, SKBR3 and HCC1954 treated with erlotinib, trastuzumab or pertuzumab, alone or combined, up to 60 minutes and 30 hours, respectively. In a Boolean modelling approach, signalling networks were reconstructed based on these data in a cell line and time course specific manner, including prior literature knowledge. Finally, we simulated network response to inhibitor combinations to detect signalling nodes reflecting growth inhibition.

Results: The networks pointed to cell line specific activation patterns of the MAPK and PI3K pathway. In BT474, the PI3K signal route was favoured, while in SKBR3, novel edges highlighted MAPK signalling. In HCC1954, the inferred edges stimulated both pathways. For example, we uncovered feedback loops amplifying PI3K signalling, in line with the known trastuzumab resistance of this cell line. In the perturbation simulations on the short-term networks, we analysed ERK1/2, AKT and p70S6K. The results indicated a pathway specific drug response, driven by the type of growth factor stimulus. HCC1954 revealed an edgetic type of PIK3CA-mutation, contributing to trastuzumab inefficacy. Drug impact on the AKT and ERK1/2 signalling axes is mirrored by effects on RB and RPS6, relating to phenotypic events like cell growth or proliferation. Therefore, we additionally analysed RB and RPS6 in the long-term networks.

Conclusions: We derived protein interaction models for three breast cancer cell lines. Changes compared to the common reference network hint towards individual characteristics and potential drug resistance mechanisms. Simulation of perturbations were consistent with the experimental data, confirming our combined reverse and forward engineering approach as valuable for drug discovery and personalised medicine.

Keywords: ErbB, RPPA, Network reconstruction, Boolean model, Breast cancer cell line, Drug resistance

*Correspondence: tim.beissbarth@ams.med.uni-goettingen.de

¹Statistical Bioinformatics, Department of Medical Statistics, University Medical Center Göttingen, Humboldtallee 32, 37073 Göttingen, Germany
Full list of author information is available at the end of the article

Background

Longitudinal time course data are the basis for modelling signalling networks in a holistic systems biology approach in order to uncover mechanisms of signal transduction dynamics [1,2]. Network models provide novel insight [3,4] and allow us to perform efficiently simulations to predict systems behaviour or evaluate certain hypotheses [5]. Furthermore, combining perturbation experiments with the measurements of system dynamics seems to be even more efficient than time series data on their own [6-8]. Knock-outs or stimuli as directed perturbations support the systematic identification of regulatory relationships.

Quantitative models, based on differential equations, require explicit knowledge on the kinetics of the system of interest [9-12]. In contrast, the qualitative Boolean abstraction considers the components' states as binary variables, being either active (1) or passive (0), but nevertheless encompasses the essential functionality [13,14]. Wang et al. stressed, that Boolean models have already been successfully applied in reverse engineering of proteomic signalling networks, and their reduced complexity is considered to be especially advantageous for large-scale systems [15]. To avoid the drawbacks of purely data- or literature-driven algorithms regarding completeness, generalisation or interpretability, combined approaches become more and more prominent in the area of network reconstruction [6,16,17]. Some reverse engineering approaches, like *ddepn* [6] or *CellNOptR* [18], ideally join perturbed time course input data and literature prior knowledge in network reconstruction, while preserving the simplicity of Boolean logic at the same time. Forward engineering methods allow subsequent analysis of the stable states of the reconstructed system. Hence, this may allow to deduce possible long-term behaviour of components activity under perturbations. Such approaches are integrated and freely available in the open source Python software package *BooleanNet* [19] or in the *R* [20] package *BoolNet* [21], for example. As reviewed by Samaga and Klampt [22], several software tools can be applied for the dynamic modeling of logical signal transduction networks. Among others, they exemplarily mentioned *GINsim* [23], *SQUAD* [24], *BooleanNet* [19], *ChemChains* [25], *Odefy* [26], and *BoolNet* [21].

Here we focus on protein signalling networks in breast cancer, representing the most common cancer type among women [27]. Breast cancer, as a heterogeneous disease, can be divided into subgroups, which differ in cellular properties as well as in prognosis. This requires individual therapy approaches, which are in the focus of current research and have partially already been realised.

Here we are interested in the 'HER2-positive' subtype of breast cancer, overexpressing the human epidermal growth factor receptor 2 (HER2, also termed ErbB-2).

ErbB-2 is a receptor tyrosine kinase (RTK) and member of the epidermal growth factor (EGF) receptor family, consisting of three further RTKs, namely ErbB-1, ErbB-3 and ErbB-4. These receptors cooperatively function as homo- or heterodimers after activation via growth factors like EGF for ErbB-1 or heregulin (HRG) for ErbB-3 [28]. This initialises signalling cascades, pathologically contributing to tumourigenesis and tumour progression. Interestingly, different dimer formations induce different signalling pathways, like PI3K and MAPK, also with differing signalling strengths [29]. The role of the orphan receptor ErbB-2 in dysregulation of the ErbB network is of major interest, due to its overexpression in 10-20% of breast tumours, diagnosed as HER2-positive. Furthermore, its role as favoured dimerisation partner independent on ligand-activation implies oncogenic potential [30-32]. The therapeutic antibodies trastuzumab and pertuzumab have especially been designed to target ErbB-2 [33].

However, frequently occurring therapy resistance reduces the efficiency of targeted therapeutics [34-36]. This resistance is often associated with deregulated pathway activity [37,38] or bypasses via other RTKs, especially ErbB family members [39]. Mainly ErbB-1 expression has been anticipated as molecular cause to overcome impact of ErbB-2 targeting drugs. Small-molecule inhibitors such as erlotinib are already in use against non-small cell lung cancer [40] and pancreatic cancer [41].

Here we aim as a first step at the identification of individual drug response patterns and insights into drug resistance in HER2-positive breast cancer. ErbB-2 amplified cell lines were therefore subjected to short- and long-term drug treatment with erlotinib, pertuzumab and trastuzumab, alone or in combinations. Samples were analysed by reverse-phase protein arrays (RPPA) [42]. We were interested in synergistic benefits of combining erlotinib, pertuzumab or trastuzumab in ErbB-1 expressing, ErbB-2 amplified tumours with differing resistance phenotypes. Therefore three representative breast cancer cell lines were selected as model systems, namely BT474, SKBR3 and HCC1954, of which the latter is known to be trastuzumab resistant due to a PIK3CA mutation, while BT474 exhibits wild type behaviour [43]. The SKBR3 cell line is supposed to be pertuzumab resistant [44].

ErbB dimers predominantly activate the MAPK and PI3K pathway [29]. Therefore, we concentrated on the involved key regulators in fast downstream signalling. Among those were ERK1/2 and AKT, and also p70S6K, which is upstream influenced by both of the signalling axes. Phosphorylation of RPS6 and RB was used as long-term indicator for proliferation, cell cycle or tumour progression [28]. Prior literature knowledge on ErbB signalling was used as input for protein network reconstruction per cell line via *ddepn*. Beyond that, we inferred

combined therapies that target ErbB family members, customised to the topology of the different subtypes. *BoolNet* was applied to compute stable cycles of protein activity states, so-called attractors, incorporating all possible treatment combinations. This way, optimal drug treatment to deactivate oncogenic proteins was identified.

Methods

Data

Protein abundance and phosphorylation measurements in BT474, SKBR3 and HCC1954 cells were carried out as described by Henjes et al. [28]. In principle, the RPPA protein array technology works as follows. Minimal amounts (1 nl volume) of cell lysate are spotted along with a serial dilution of control samples on nitrocellulose-coated glass slides using a printing robot (Aushon 2470 arrayer). Samples are organised as ordered subarrays so that they are addressable during the data analysis procedure, and a single slide can accommodate one or more subarrays. Each subarray is analysed using a highly specific detection antibody to measure the abundance of a certain protein or its phosphorylation rate. For each spot, the ratio of bound detection antibody is visualised using secondary antibodies labelled with near infrared (NIR) fluorescent dyes. Slides are scanned using the Odyssey scanner (LiCor Biosciences). Spot intensities are determined using a microarray image analysis software (GenePix).

Apart from the quantitative character, another advantage of the technology is the handling of large sample sets which protein abundance can be detected simultaneously in a high throughput fashion. 20-200 identical slides can be produced in parallel in a single print run.

In order to normalise the data spot-wise for deviant total protein concentrations due to spotting variance, staining with Fast Green FCF dye was employed [42]. Therefore, one slide was stained with the dye to determine the total protein content of each lysate spot and corresponding signal intensity correction factors. The spots on the remaining slides were divided by these correction factors and afterwards multiplied by the median value to scale the data back to the native range.

The RPPA data used here include data presented in Henjes et al. [28]. Additionally, further targets have been measured and were used for network reconstruction. The complete data set has been submitted to the Gene Expression Omnibus (GEO) with accession number GSE50109.

Short-term measurements

In the short-term measurements, trastuzumab, pertuzumab and erlotinib were added to the cells in starvation medium one hour before stimulation with the growth factors EGF and HRG. All possible 24 combinations of drugs and stimuli were measured. Application of the stimuli was defined as time point zero in the measurements.

The growth factors were chosen to activate explicitly the MAPK and PI3K pathway. Lysate preparation was performed at ten time points, namely after 0, 4, 8, 12, 16, 20, 30, 40, 50 and 60 minutes. The drug treatment experiments comprised three biological replicates, whereas the inhibitor-free experiments incorporated five biological replicates. The experiments for the SKBR3 cell line comprised only two biological replicates of HRG stimulated cells under the triple drug combination. Each biological replicate was spotted in triplicate on the RPPA slides. To obtain short-term signal intensities, eleven antibodies for specific phosphorylation sites were selected according to quality checks, including inspection of corresponding dilution series and comparison to signals arising from secondary antibodies only. The chosen target proteins and respective antibodies are listed in Additional file 1.

Long-term measurements

For long-term measurements, no explicit ligand stimulation was performed. Instead, cells were incubated in full growth medium for 24 hours prior to adding the three mentioned therapeutics in double combinations or as triplet. Single drug treatment was just conducted with erlotinib. Full growth medium was used to avoid confounding effects of nutrient deficiency. Protein abundance was also quantified without any drug application. The measuring points included 0, 1, 2, 4, 6, 8, 12, 18, 24 and 30 hours with three biological and technical replicates each. At time point 18, only two biological replicates were available. Additional file 1 displays the 21 targets of interest for long-term signalling.

Statistical inference of drug effects

To determine, whether a specific drug treatment revealed an inhibiting effect on the signal intensities of the proteins, we applied the following method. Firstly, for each protein and (combinatorial) drug treatment we linearly modelled the signal intensities as depending on the factors *time* and *group*, i.e. no drug treatment versus drug treatment. If the interaction of both factors showed a significant (p -value < 0.05) influence on the signal intensity, we further applied a Wilcoxon rank sum test for the measurements at time point 60 minutes for the short-term data, or at time point 30 hours for the long-term data. Thereby, we tested for significantly (p -value < 0.05) smaller intensity values in the drug treated group. The drug treatments with a significant test result were considered as efficient inhibitors. The therapeutic (combination) with the smallest p -value was defined as the optimal one.

Literature prior knowledge

We manually determined two reference networks, i.e. one for each time course, as initial joint hypotheses for all of the three breast cancer cell lines. Because emphasis

was put on phosphoproteomic signalling, this was mainly based on PhosphoSitePlus® [45]. Several publications confirm these assumptions, as depicted in Additional file 2.

Network reconstruction

For Boolean network reconstruction, we chose the method of dynamic deterministic effects propagation networks (DDEPN) [6]. This method was particularly tailored to perturbed longitudinal protein phosphorylation data. It is based on the DEPN approach [46], which stands for deterministic effects propagation networks. The determinism is related to the way of perturbation effect propagation in the networks from parent to child nodes, implying transitively closed graphs. The dynamic version of Bender et al. [6,47] differs with respect to the integration of perturbed time course measurements. While the DEPN approach requires many perturbations, like knock-downs, but only few time points, which are regarded as independent measurements, *ddepn* is designed for longer time series without the necessity of many or all network nodes being perturbed. The latter situation, i.e. few perturbations by drug interventions, reflected the design of the RPPA experiments under consideration here, hence leading to the application of *ddepn*. Most network reconstruction algorithms have been designed for gene expression data from microarray measurements [7], which differ from (phospho-)protein data regarding the amount of involved network nodes. Many current methods are tailored to the inference of gene regulatory networks based on static measurements at one time point, reflecting the steady state of the system under consideration [48]. The longitudinal time course data used here require a suitable method, as provided by *ddepn*. The method of Bender et al. was shown to outperform two dynamical Bayesian network approaches, and to be capable of inferring known signalling cascades in the ErbB pathway [47]. A further advantage was the public availability of *ddepn* as an R [20] package.

The reconstruction procedure is depicted in Additional file 3, and the core elements are described according to [6,47] in the following. The protein interaction networks are modelled as directed, possibly cyclic, graphs, with nodes $V = \{v_i : i \in 1, \dots, N\}$ representing proteins and edges representing interactions. Also the external perturbations, i.e. the drugs and growth factors in our case, are modelled as nodes. The edge types can be either activating or inhibiting, denoted by 1 and -1, respectively, in the adjacency matrix $\Phi = V \times V \rightarrow \{0, 1, -1\}$ of the network. An entry of zero indicates no edge between two nodes. So each edge incorporates a pair of nodes $\{\phi_{ij} : i, j \in 1, \dots, N\}$. The measurement data, which form the basis for the reconstruction, are stored in a matrix $D = \{d_{itr} : i \in 1, \dots, N, t \in 1, \dots, T, r \in 1, \dots, R\}$, considering T time points and R replicates.

For the inference of a network structure, optimally fitting to the data, we applied the stochastic Markov Chain Monte Carlo (MCMC) approach of *ddepn*, called *inhibMCMC*, in which the space of possible networks is sampled, based on posterior probabilities. It extends a Metropolis-Hastings type of MCMC sampler by the capability of sampling two edge types directly, i.e. activation and inhibition. The posterior distribution of a network Φ given the data D , is defined as $P(\Phi|D) = \frac{P(D|\Phi)P(\Phi)}{P(D)} \propto P(D|\Phi)P(\Phi)$, with $P(\Phi)$ as the prior probability distribution and $P(D|\Phi)$ as the likelihood of the data given the network. The latter is defined in [47] as $p(D|\Phi) = p(D|\hat{\Gamma}^*, \hat{\Theta}) = \prod_{t=1}^T \prod_{i=1}^N \prod_{r=1}^R p(d_{itr}|\hat{\theta}_{i\gamma_{itr}^*})$,

where $\Gamma^* = \{\gamma_{itr}^* : i \in 1, \dots, N, t \in 1, \dots, T, r \in 1, \dots, R\}$ denotes the optimized system state matrix, containing active and passive states per protein and time point. It is estimated in the following way. Assuming that the proteins can be either active (1) or inactive (0), signalling dynamics are modelled by Boolean signal propagation for a given network. All nodes, except the permanently active perturbations, are therefore initialised with inactive states. The transition rule is that children nodes get activated if at least one activating parent node is active and all inhibiting ones are inactive. In this way, all reachable system states are computed and stored in a matrix $\Gamma = \{\gamma_{ik} \in \{0, 1\} : i \in 1, \dots, N, k \in 1, \dots, M\}$, holding columnwise the activation states of all proteins at transition step k . The amount of transitions is limited by $0 < M \leq 2^N$. This state matrix has to be optimized, as it is not related to the measured time points yet. The true unknown state sequence over time is represented by Γ^* , which is estimated by a hidden Markov model (HMM). The resulting $\hat{\Gamma}^*$ indicates whether a data point d_{itr} has an underlying active (1) or passive (0) normal distribution

$$d_{itr} \sim \begin{cases} \mathcal{N}(\mu_{i0}, \sigma_{i0}), & \text{if } \hat{\gamma}_{itr}^* = 0 \\ \mathcal{N}(\mu_{i1}, \sigma_{i1}), & \text{if } \hat{\gamma}_{itr}^* = 1. \end{cases}$$

The distribution parameters are for each protein estimated as empirical mean and standard deviation of all measurements for the considered protein in the corresponding class, yielding the parameter matrix $\hat{\Theta} = \{\hat{\theta}_{i0}, \hat{\theta}_{i1}\} = \{(\hat{\mu}_{i0}, \hat{\sigma}_{i0}), (\hat{\mu}_{i1}, \hat{\sigma}_{i1})\} \forall i \in 1, \dots, N$.

The prior probability distribution $P(\Phi)$ includes penalisation of differences between the network structure Φ and a user-defined prior belief $B = V \times V \rightarrow [-1, 1]$, where the absolute value correlates with the confidence in an edge. Here we chose $B = V \times V \rightarrow \{0, 1, -1\}$, assuming in advance specific activating, inhibiting or missing edges with maximum confidence. We made use of the Laplace prior model (*laplaceinhib*), accounting for both edge types, i.e. activation and inhibition. The prior belief

for an edge is defined as $P(\phi_{ij}|b_{ij}, \lambda, \gamma) = \frac{1}{2\lambda} e^{-\frac{\Delta_{ij}}{\lambda}}$, including a weighted difference term $\Delta_{ij} = |\phi_{ij} - b_{ij}|^\gamma$ with a weight exponent $\gamma \in \mathbb{R}^+$. As the edge probabilities are assumed to be independent, the prior belief for a network structure Φ is derived as the product of those, i.e. $P(\Phi|B, \lambda, \gamma) = \prod_{i,j} P(\phi_{ij}|b_{ij}, \lambda, \gamma)$, $i, j \in \{1, \dots, N\}$. The

individual edge probabilities lie between 0 and $\frac{1}{2\lambda} \forall \lambda, \gamma \in \mathbb{R}^+$. The protein interactions corresponding to our chosen prior are displayed in Additional file 2. The prior's impact strength was emphasised in such a way, that only strongly deviating data influence the network structure, because the ErbB wiring as well as the MAPK and PI3K pathways are well examined in literature. This prioritisation is reflected in the hyperparameter λ set to 0.0001. For the parameter γ we chose one, neglecting extra penalisation of deviation from the prior. These settings should preserve robustness, but at the same time allow enough impact strength of strongly differing data values.

The network inference via *inhibMCMC* spanned 50,000 iterations with the first 25,000 iterative steps as burn-in phase. To ensure convergence, ten parallel MCMC chains were run, each initialised with a starting network. Convergence was validated via Gelman diagnostic [49]. Nine of the initial ten networks were randomly generated, i.e. for the defined nodes activating, inhibiting or no edges were sampled. The remaining network assumed no connections between the nodes. These initial networks were pruned to the following constraints. Firstly, the nodes related to the growth factors and drugs must not have any ingoing edges. Above that, the indegree of all nodes was limited to four. Finally, no self-loops were allowed. To find significantly occurring edges among the independent runs, merging into a consensus network, a Wilcoxon rank sum testing procedure was used. In detail, in each run the amount of sampled activations and inhibitions per edge was counted and divided by the total number of sampled edges. Subsequently the null-hypothesis was tested, whether the means of these ten edge-specific confidence values equal the same for activation and inhibition. In case of not rejecting the null-hypothesis, coming along with an adjusted p-value exceeding the significance level $\alpha = 0.05$, no edge was assumed. Otherwise, the respective alternative determined the type of interaction. Adjustment for multiple testing followed the method of Benjamini and Hochberg, controlling the false discovery rate [50]. The whole procedure was embedded into a leave-one-out cross-validation approach. So each of the ten MCMC chains was left out once, and the testing algorithm was applied to the remaining runs. An edge was included in the final consensus network if it occurred in all of the cross validation runs. Finally, to prevent excessive spurious or obsolete connections ascribable to transitivity, as argued by Bo Na Ki et al. [51], newly reconstructed edges

were successively added to the prior network according to *ddepn* significance and fit of resulting attractor states to the observations of Henjes et al. [28].

Perturbation simulations

To figure out which input of drug combination leads to a certain attractor state of the reconstructed network system, the *R* package *BoolNet* [21] was applied. The motivation was based on the assumption that attractors, representing cycles of states, comprise the stable states of cell function. In those states networks mostly reside. Hence, they mirror system phenotypes, dependent on the perturbation context. To the best of our knowledge, apart from *BoolNet*, there are hardly any *R* packages offering attractor calculations for Boolean networks. This package supports import of networks in form of files containing Boolean formulas. So it could be easily integrated in our workflow as subsequent analysis step after network reconstruction.

We used its functionality to identify attractors in a synchronous and an asynchronous way. The resulting attractors were steady-state attractors. These consist of only one state, in which all transitions from this state result. These attractors are identical for synchronous and asynchronous updates. We focused on the steady-states, as these should reflect the homeostatic system state of the cell lines. Intermediate transition states would be interesting as well, but due to the large amount of the involved targets, it would have been too complex to analyse those here in detail.

The search started from predefined initial states of the network nodes. The drug and growth factor nodes were fixed to specific values, reflecting the conducted experiment to be simulated. For short-term signalling, perturbations included all possible combinations of the therapeutics under the combined stimulus of EGF and HRG. Although the data of separate stimulation with EGF and HRG was used for network reconstructions, here we focused on the combined treatment, representing a more natural tumour environment than a single growth factor alone. Two possible binary states, i.e. active (1) or passive (0), to the power of three different drugs led to eight possible combinations. These were used as fixed input conditions, as the effect was assumed to be continuously valid. Analogously, the growth factors were permanently fixed to one. The remaining protein activity start states were initialised with zero. These components were flexible towards updates. In the long-term measurements, no growth factors were involved but full growth medium. This was defined as one stimulating input *S*, initially activating the ErbB receptors. This also led to eight fixed input combinations.

BoolNet expects network representation in form of logical interaction rules as input. In contrast, *ddepn* delivers network reconstruction output in terms of adjacency

matrices. Therefore, we incorporated an interface function into the *ddepn* package, called *adjacencyMatrix_to_logicalRules*. In detail, the *loadNetwork* function of *BoolNet* requires a file containing row-wise logical activation rules of each network node. Each row looks like 'target node, (activator_1 | activator_2) & !(inhibitor_1 | inhibitor_2)', here exemplary for a node with two incoming activating and inhibiting edges each. The logical OR operator is encoded by '|', the logical AND is encoded by '&', and logical negation is represented by '!'. Accordingly, all of the A inferred activating nodes $V_+ = \{v_a : a \in 1, \dots, A, A < N\}$ of a target node v_j , represented by an adjacency matrix entry $\phi_{aj} = 1$, and v_j itself were connected via OR operators. This ensured that at least one of the activators or the target protein itself had to be active to activate the target node. Analogously, the I inhibiting nodes $V_- = \{v_i : i \in 1, \dots, I, I < N\}$ with $\phi_{ij} = -1$ were connected via OR operators. A logical negation operator was attached to ensure that the activity of one of the nodes v_i would result in an inactive node v_j . Both sets of activators and negated inhibitors were then connected via a logical AND operator. After conversion of the adjacency matrices to logical rules, those were implemented in *BoolNet* into a computational model, to perform perturbation simulations per cell line and time course as well as subsequent analyses of the resulting attractor states.

Results and discussion

The complete workflow, holding for both, short- and long-term analysis, is depicted in Figure 1. For a better understanding of the discussion on MAPK and PI3K signalling, Figure 2 displays the interactions between the main MAPK and PI3K targets of the ErbB prior networks. It shows the preferred pathway activations by all possible homo- and heterodimers formed upon ligand binding to the ErbB-1 and ErbB-3 receptors [9,29,52-54]. The confidence values, representing the likeliness of the reconstructed network edges, are shown in Additional file 4.

Short-term signalling network reconstruction

The short-term signalling networks, reconstructed by the *ddepn* algorithm, are depicted in Figure 3. The equivalent Boolean logical interaction rules are listed in Additional file 5. In comparison to the prior network, newly inferred edges were specific for each cell line, and all of them were activating. For HCC1954 and BT474, seven additional edges were reconstructed, while in SKBR3 only two new edges were reconstructed. No prior edge deletion or type reversal took place.

HCC1954 is driven by the PI3K as well as the MAPK pathway

In HCC1954, the new edges contributed to both, PI3K and MAPK, signalling. The interaction ErbB-1→ErbB-2

reflected a dominant role of heterodimerisation of both receptors, as described by Henjes et al. [28]. The fact that it was specifically inferred for HCC1954, pointed to hyperactive ErbB-1/2 heterodimers here. These are known to trigger the MAPK but also, to a lesser extent, the PI3K pathway. The link PDK1→MEK1/2, supported by Sato et al. [55], stressed crosstalk between these pathways, placing PDK1 into a key position in the PI3K pathway, and MEK1/2 in the MAPK pathway, respectively. Two of the new edges in HCC1954, PDK1→ErbB-2 and p70S6K→AKT, contributed to feedback loops, which were not present in the other two cell lines. Such a topological network element could stabilise the known trastuzumab resistance by boosting the oncogenic effect of ErbB-2 and the mutant hyperactive PI3K pathway. Evidence for the feedback mechanism involving PDK1 was provided by Maurer et al. [56] and Tseng et al. [57]. Vega et al. noted an indirect activation of AKT by p70S6K via mTOR [58].

BT474 is driven by the PI3K pathway, while SKBR3 is driven by the MAPK pathway

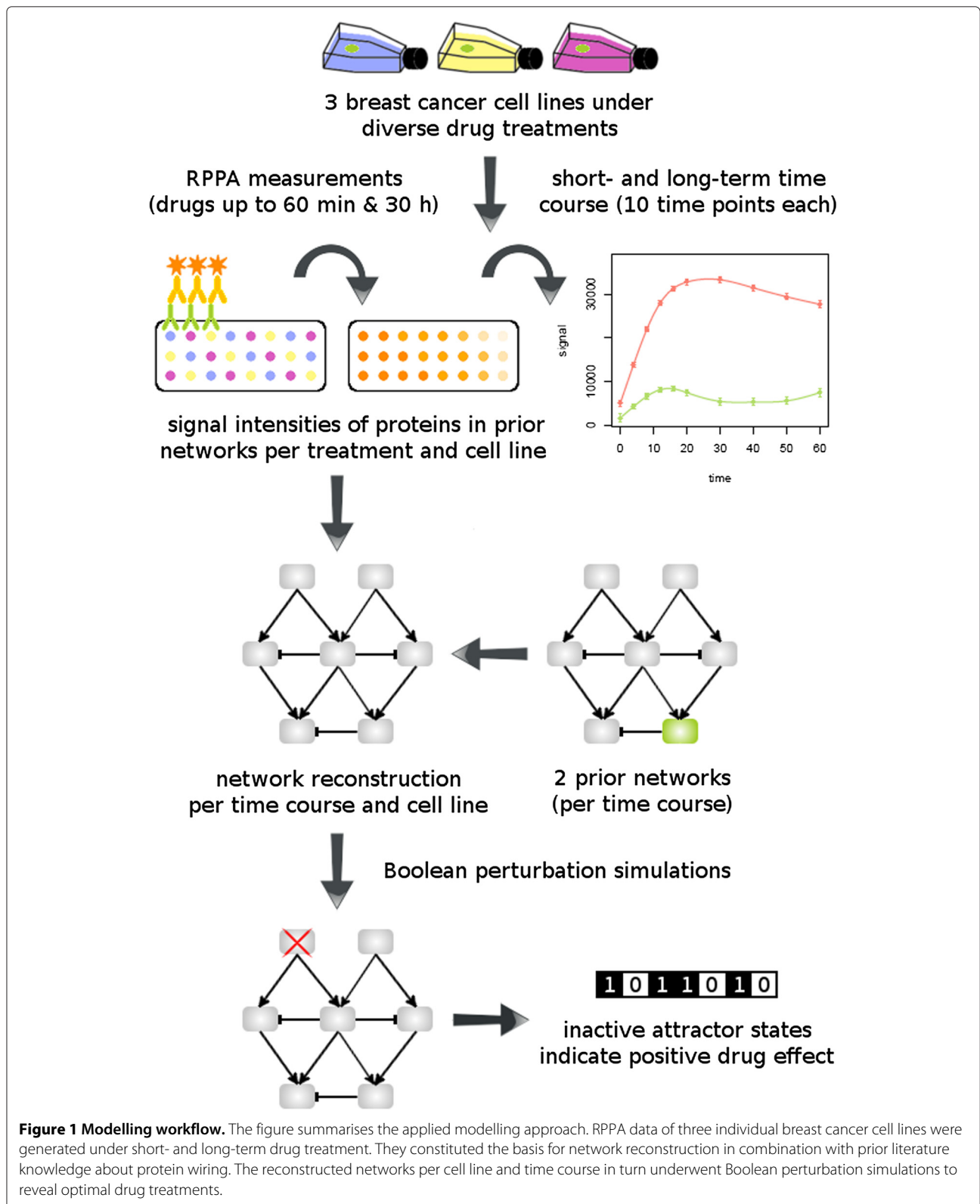
Comparably to HCC1954, in BT474 an edge indicating hyperactive heterodimers was found, namely ErbB-3→ErbB-2, here interestingly with a strong impact on AKT [28]. BT474 is known to contain a rare type of PIK3CA mutation [43]. Pathway crosstalk was also observed in BT474, but here MEK1/2 activated PDK1, and not vice versa like in HCC1954. This edge was supported by Frödin et al. [59], underlining dominant PI3K signalling in this cell line.

The newly detected interactions in SKBR3 started from ErbB-3 and PDK1, and both activated ERK1/2. This reflected a dominant MAPK pathway, in which ErbB-3→ERK1/2 was interpretable as indirect stimulation of ERK1/2 via MEK1/2, activated by ErbB-2/3 dimers [55].

Perturbation simulations on short-term networks

Perturbations included all possible combinations of the therapeutics erlotinib, pertuzumab and trastuzumab under combined stimulation of EGF and HRG. All inferred attractors were simple and consisted of one steady-state. This means that all transitions from this state result in the state itself. Table 1 summarises all simulation outcomes for the attractors of the AKT and ERK1/2 proteins, as those are key players in the PI3K (AKT) and MAPK (ERK1/2) pathways. Additionally, the results for p70S6K are listed there, as both pathways regulate this protein [60].

Stimulation with EGF and HRG should result in activation of ErbB-1 and ErbB-3, followed by dimerisation amongst ErbB members. This should initialise signalling cascades in the MAPK and PI3K pathways (Figure 2). Indeed, AKT, ERK1/2 and p70S6K got activated in



all cell lines, which was revealed by simulations as well as observations in graphical analyses (Table 1, Figures 4, 5, 6).

As we were interested in identifying optimal drug treatments, Table 2 summarises the corresponding statistical results. Most of them were supported by the perturbation

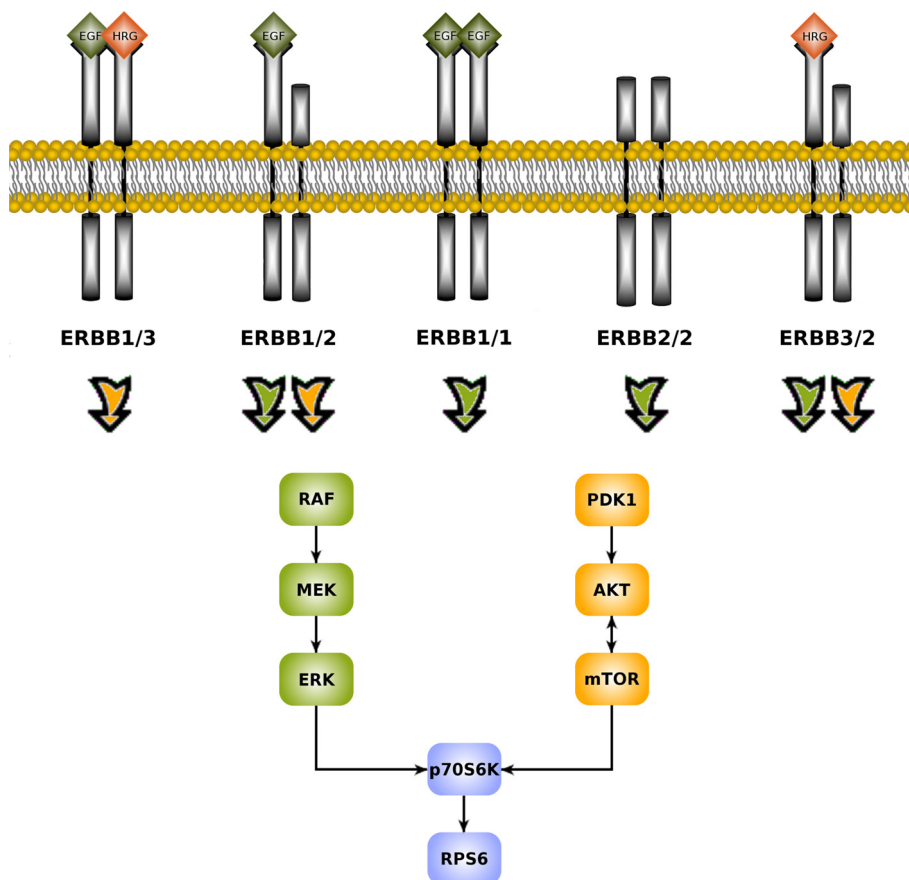


Figure 2 Scheme of ErbB dimers related MAPK and PI3K pathway activation. The figure depicts the different homo- and heterodimers of ErbB receptors, induced upon activation via the ligands EGF or HRG. The active dimers then initialise the MAPK and PI3K signalling cascades. The orange (PI3K) and green (MAPK) arrows denote, which dimer activates which pathway.

simulation results, corresponding to attractor states of AKT, ERK1/2 and p70S6K being zero. Four main conclusions were drawn from these results, which will be discussed in detail in the following subsections. Firstly, inhibition of PI3K signalling, reflected by downregulated AKT, required the combined treatment with erlotinib, pertuzumab and trastuzumab. Secondly, inhibition of the MAPK pathway, represented by ERK1/2, was reached with erlotinib alone in SKBR3 and HCC1954. BT474 additionally needed pertuzumab. Thirdly, the protein activity of p70S6K was influenced by both, PI3K and MAPK, pathways. The drug response differed between cell lines, indicating both pathways contribute to a different extent. Finally, the drug effect on PI3K signalling was much better in SKBR3 than in HCC1954, pointing to resistance in the latter cell line.

Inhibition of PI3K signalling requires drug combinations

In SKBR3, the triple drug combination was most effective in inhibiting AKT (Figure 4, Table 2). In BT474,

pertuzumab combined with erlotinib was most efficient, but AKT signalling was not fully suppressed as in SKBR3 (Figure 5). Statistically, we did not infer any significant positive drug effect in this cell line. Obviously, erlotinib in synergistic combination with at least pertuzumab was needed to block the ErbB-2 receptor and its heterodimerisation, mainly with ErbB-1, but also ErbB-3. The HRG activated ErbB-2/3 heterodimers and PI3K pathway in BT474, as revealed by the network reconstructions, might have prevented a potent drug efficacy.

Interestingly, BT474 and SKBR3 required pertuzumab. This drug was especially designed to prevent heterodimerisation with ErbB-2. The stimuli EGF and HRG together activate PI3K signalling by ErbB-2/3, ErbB-1/2 and ErbB-1/3 dimers (Figure 2). The need for pertuzumab combined with erlotinib indicated an important role of ErbB-1/2 dimers. This was supported by the fact, that in HCC1954 with dominant heterodimers of this type, as revealed by network reconstructions, none

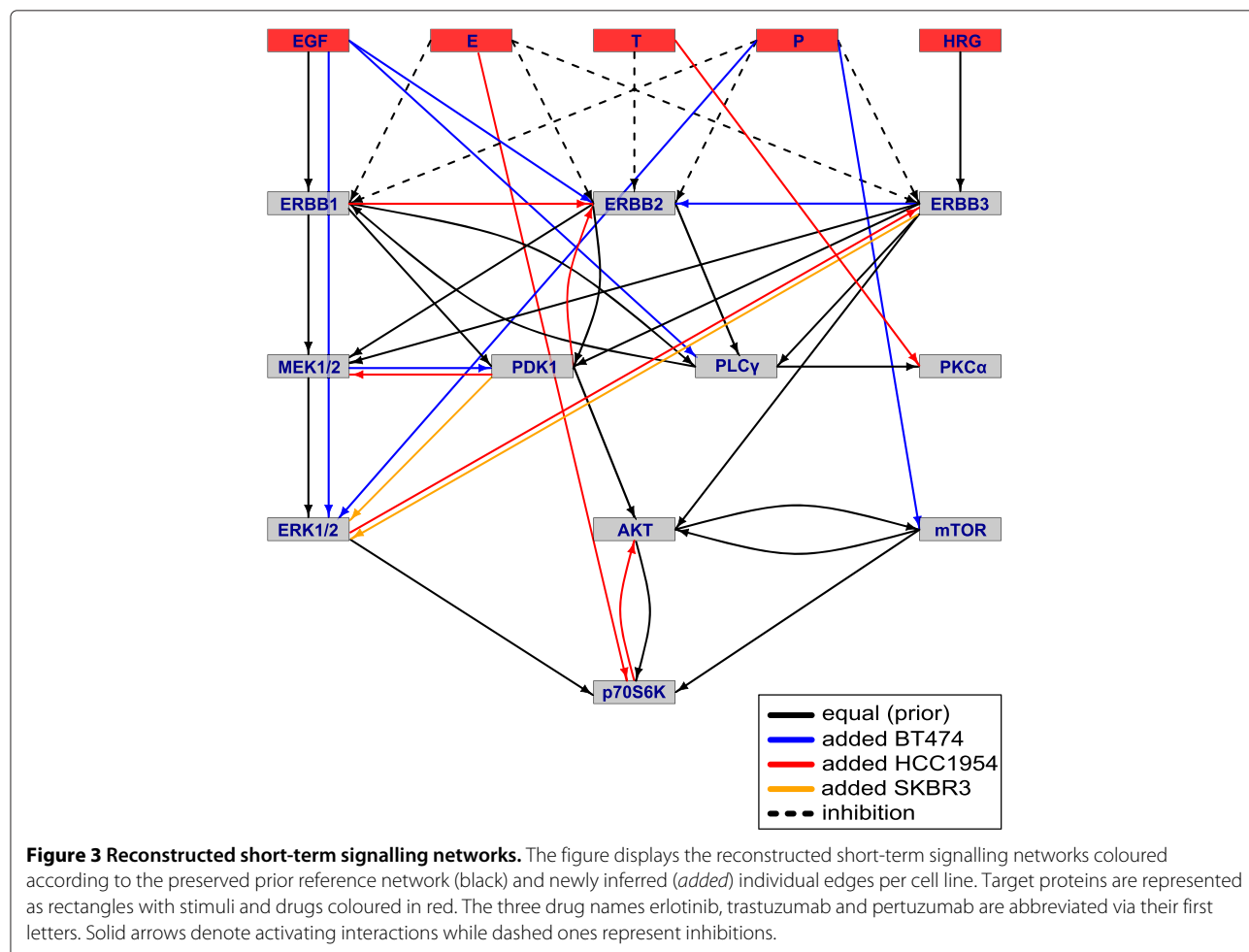


Table 1 Attractor states of short-term perturbation simulations

Simulation	BT474						HCC1954						SKBR3						
	AKT		ERK1/2		p70S6K		AKT		ERK1/2		p70S6K		AKT		ERK1/2		p70S6K		
	A	E	A	E	A	E	A	E	A	E	A	E	A	E	A	E	A	E	
X	1		1		1		1		1		1		1		1		1		1
E	0	1	1	0	1	0	1		0		1	0	1	0	0	1			1
P	1		1		1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
T	1		1		1		1		1		1		1		1		1		1
E,P	1	1	0	1	0	1	0	0		1	0	0		0	0	1			1
E,T	0	1	1	0	1		1		0	1	1	0	1	0	0	1			1
P,T	1		1		1	0	1	0		0	1	0		0	0	1			1
E,P,T	1	1	0	1	0	1	0	0		1		0		0	0	1			0

The therapeutics erlotinib, trastuzumab and pertuzumab, abbreviated by first letters, that were permanently active besides EGF and HRG in the simulated perturbation conditions are stored in the column *Simulation*. No simulated drug treatment is denoted by 'X'. The A columns hold the attractor states of the proteins AKT, ERK1/2 and p70S6K, associated with the perturbations. The E columns contain the protein activity status, statistically deduced from the experimental data. In case of a significant (p-value < 0.05) combined influence of both, drug treatment and time, on the protein signal intensity, a Wilcoxon rank sum test was conducted for the measurements at time point 60 minutes. The drug treatments leading to significantly (p-value < 0.05) smaller intensity values compared to the control measurement 'X' were considered as efficient inhibitors, resulting in a table entry of zero. Consistency between simulations and experimental observations is printed in bold.

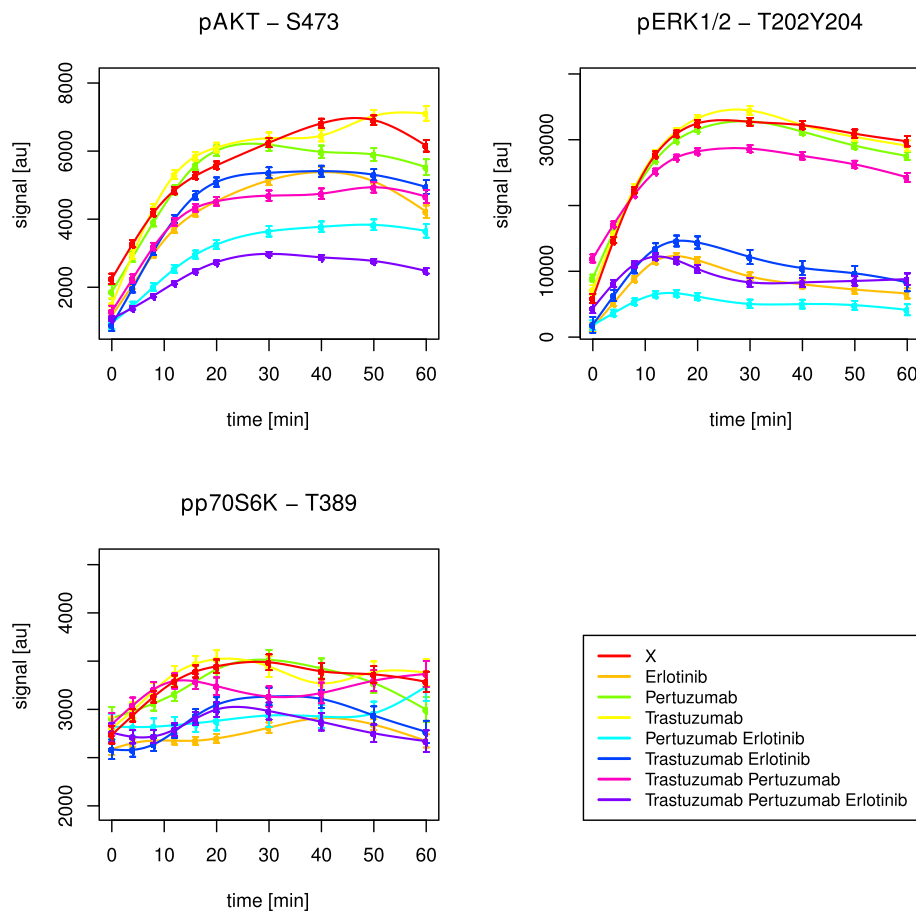


Figure 4 SKBR3 short-term time courses of AKT, ERK1/2 and p70S6K. The figure shows splines and related standard error bars of the measured RPPA data for AKT, ERK1/2 and p70S6K after combined EGF and HRG stimulation in the SKBR3 cell line. The measurements included ten time points up to 60 minutes. The different drug treatments are marked by different colours with 'X' denoting no drug treatment.

of the drugs was likewise efficient in inhibiting AKT (Figure 6). However, the optimal effect was revealed for the triple drug combination (Table 2). The simulations suggested pertuzumab alone or a combination of both monoclonal antibodies (Table 1). It has to be kept in mind, that the attractor states resembled a long-term steady state, which can differ from observations up to 60 minutes.

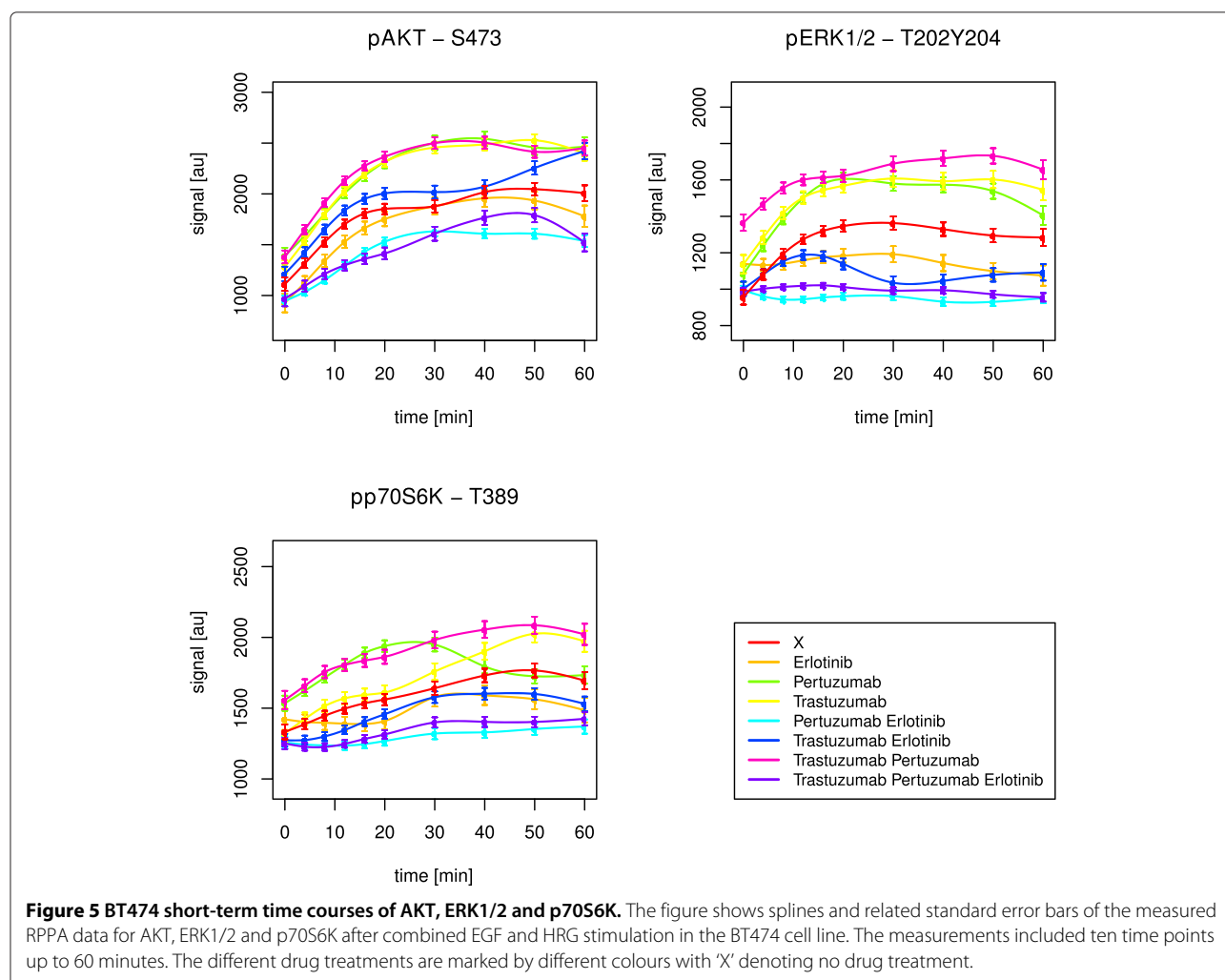
The perturbation simulations in BT474 did not lead to inactive AKT upon combined pertuzumab and erlotinib treatment. Instead, erlotinib alone or combined with trastuzumab was efficient (Table 1). Nevertheless, this supported the need for the small molecule inhibitor and a monoclonal antibody to suppress ErbB-2 induced PI3K signalling. In SKBR3, the attractor states confirmed the described optimal drug treatment to deactivate AKT. Trastuzumab, when applied alone, was the only treatment without a positive effect in the simulations (Table 1).

Inhibition of MAPK signalling requires erlotinib

Signalling through the MAPK pathway, represented by ERK1/2 activation, was efficiently inhibited by erlotinib alone in both, HCC1954 (Figure 6, Table 2) and SKBR3 (Figure 4, Table 2), cell lines. EGF activates the MAPK pathway via ErbB-1 homodimers and ErbB-1/2 heterodimers (Figure 2). Both are prevented by ErbB-1 inhibition via erlotinib, which was especially designed to target this receptor.

In BT474, pertuzumab plus erlotinib was required (Figure 5, Table 2). This was analogous to the situation in PI3K signalling.

HRG activates the MAPK pathway via ErbB-2/3 heterodimers (Figure 2). Obviously, BT474 needed the addition of the monoclonal antibody due to dominant ErbB-2/3 formation and activity. On the contrary, the other two cell lines just needed erlotinib alone. Here, in addition to the ErbB-1 dimers, the ligand-independent ErbB-2 homodimers might have driven ERK1/2 activation



and could be inhibited by the small molecule inhibitor. Efficacy of erlotinib towards ErbB-2 dimers was previously mentioned by Schaefer et al. [61].

In BT474, the simulations resulted in active ERK1/2 states, resisting drug treatment (Table 1). In HCC1954 and SKBR3, the positive effect of erlotinib was supported by the simulations. The attractor states were additionally inactive for all other (combinatorial) drug treatments, but not trastuzumab alone.

p70S6K is influenced by both, PI3K and MAPK, pathways

The target p70S6K is upstream influenced by the PI3K as well as the MAPK pathway (Figure 2). Hence, p70S6K merges both pathways, leading to activation of RPS6 [60].

The three cell lines showed different pathway preferences. BT474 required the combination of pertuzumab and erlotinib to suppress p70S6K (Figure 5). On the contrary, in SKBR3 the triple drug combination was shown to be optimal (Table 2). Obviously, the effect was driven by erlotinib (Figure 4), which was supported by the attractor

states of p70S6K (Table 1). This resembled the drug response of ERK1/2 and reflected a stronger influence by the MAPK pathway. In HCC1954, deactivation of p70S6K was reached via application of erlotinib combined with pertuzumab (Table 2). The treatment with erlotinib alone had a similar effect (Figure 6), while the simulations just confirmed a positive effect of pertuzumab (Table 1). Thus, this cell line seemed to be influenced by both, PI3K and MAPK, pathways.

These results were in line with the newly inferred edges in the network reconstructions. They pointed to a strong influence of PI3K in BT474 in contrast to a dominant MAPK pathway in SKBR3. HCC1954 was influenced by both pathways to a similar extent.

To follow up on the hypothesis that different pathways contribute to a different extent in individual cell lines, we tested correlation between the p70S6K time course and the ones of AKT and ERK1/2, respectively. In BT474, p70S6K correlated positively with AKT (p-value 0.01, Kendall's τ estimate 0.64). In HCC1954,

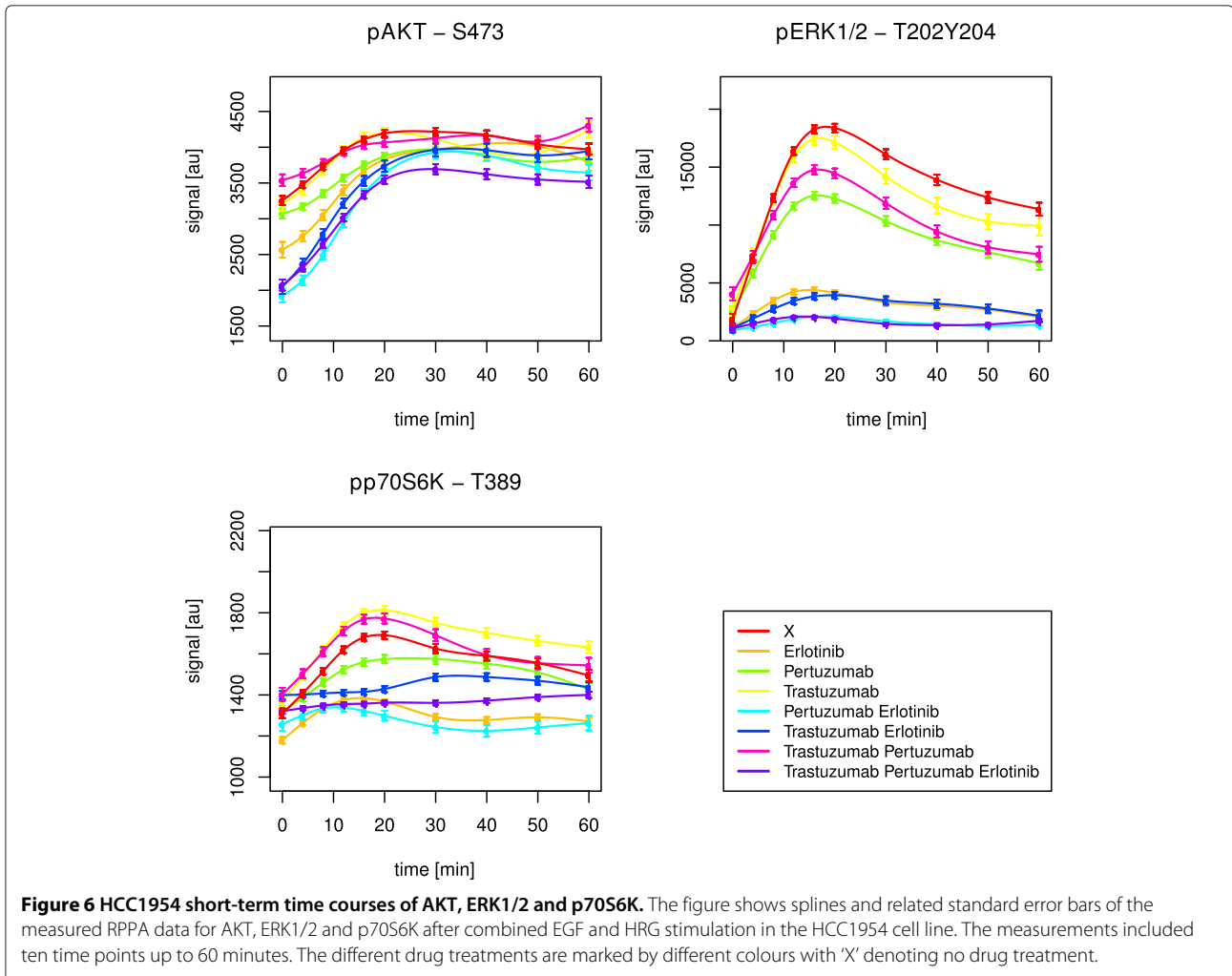


Figure 6 HCC1954 short-term time courses of AKT, ERK1/2 and p70S6K. The figure shows splines and related standard error bars of the measured RPPA data for AKT, ERK1/2 and p70S6K after combined EGF and HRG stimulation in the HCC1954 cell line. The measurements included ten time points up to 60 minutes. The different drug treatments are marked by different colours with 'X' denoting no drug treatment.

p70S6K correlated positively with both, AKT (p-value < $2.22 \cdot 10^{-16}$, Kendall's τ estimate 0.69) and ERK1/2 (p-value < $2.22 \cdot 10^{-16}$, Kendall's τ estimate 0.87). In SKBR3, p70S6K also correlated positively with both, AKT (p-value 0.05, Kendall's τ estimate 0.51) and ERK1/2 (p-value 0.02, Kendall's τ estimate 0.6), with a stronger

tendency towards MAPK signalling. The correlation was not as convincing as in the other two cell lines. One could speculate, that the dominance of the MAPK pathway in SKBR3 cells was not as strong as the dominance of the PI3K pathway in BT474. This was supported by the reconstructed networks. They revealed down-

Table 2 Optimal drug treatment in short-term signalling

Cell line	AKT	ERK1/2	p70S6K
BT474	-	PE	PE
HCC1954	PTE	E	PE
SKBR3	PTE	E	PTE

The table summarises the optimal drug treatments for the short-term data, leading to inactive AKT, ERK1/2 and p70S6K, respectively. In case of a significant (p-value < 0.05) combined influence of both, drug treatment and time, on the protein signal intensity, a Wilcoxon rank sum test was conducted for the measurements at time point 60 minutes, testing for significantly (p-value < 0.05) smaller intensity values under the drug treatment compared to the control measurement. The drug treatment with the smallest p-value was considered as the optimal inhibitor. No inferred significant positive drug effect is denoted by '-'. The therapeutics erlotinib, trastuzumab and pertuzumab are abbreviated by their first letters. More than one letter denotes drug combinations. The growth factors EGF and HRG were added in combination to the cell lines and permanently active in the simulated perturbation conditions. The column *Cell line* holds the cell lines under consideration. The columns *AKT*, *ERK1/2* and *p70S6K* hold the optimal drug combinations for each target. If those were confirmed by the attractor states (0) of perturbation simulations, they are printed in bold.

stream effects of MAPK signalling in SKBR3, while they revealed hyperactive ErbB-2/3 dimers in BT474. The dimers drive PI3K already at the receptor layer, and especially ErbB-2/3 dimers are regarded as the most potent heterodimer [29].

Drug resistance in HCC1954 regarding the PI3K pathway

In HCC1954, the inferred optimal treatment against AKT signalling with the triple drug combination was not convincing (Figure 6). Analogously, Henjes et al. did not monitor any positive drug effect on AKT under EGF application alone [28]. However, the simulations suggested pertuzumab alone or a combination of both monoclonal antibodies to inhibit AKT phosphorylation. In principle, divergence of simulations from experimental observations can be expected, as the simulated steady state of the system does not necessarily have to be reached after the measured period of time. Anyhow, the apparent resistance here pointed to a hyperactive PI3K pathway which was explainable by the newly inferred HCC1954 edges described in the previous subsection. They represented feedback loops, hyperactive ErbB-1/2 heterodimers and pathway crosstalk. On the contrary, in SKBR3, the triple drug combination worked well, as described before. The simulations even predicted efficacy of every other drug (combination) apart from trastuzumab alone. The drug efficacy towards AKT in this cell line could be explained by the fact that the two reconstructed interactions in SKBR3 mainly promoted the MAPK instead of the PI3K pathway.

The regulation of AKT activity under drug influence, highly diverging in HCC1954 and SKBR3, attracted our attention. Therefore we intended testing for edgetic mutations, as discussed by Zhong et al. [62], leading to AKT gain-of-function in HCC1954. Such mutations, perturbing not a node but an edge of a network, are speculated to have deeper impact on phenotypic manifestation of a disease. In detail, we removed each of the AKT stimulating edges outgoing from p70S6K, PDK1, mTOR and ErbB-3, alone or in all possible eleven combinations. We then computed the attractor states for the modified networks in HCC1954.

Removal of the connections of mTOR, PDK1 and ErbB-3 alone or combined had no influence on improving drug effects, i.e. AKT just got inactive under pertuzumab treatment. Involvement of p70S6K→AKT in the withdrawal process led to much better results. Removed alone or in double combinations with the aforementioned edges, as well as in the two triple combinations containing mTOR, AKT was deactivated under all drug treatments, but not yet trastuzumab alone. Finally, simultaneous removal of the outgoing connections from p70S6K, ErbB-3 and PDK1 with or without mTOR, turned out as the only combination enabling potency of all possible drug combinations,

including trastuzumab alone. This hinted at a less strong impact of mTOR on AKT here, but indicated synergistic drug resistance potential of p70S6K, ErbB-3 and PDK1, also due to the newly inferred edges.

Long-term signalling network reconstruction

The reconstructed long-term signalling networks per cell line are displayed in the Additional file 6. Additional file 5 lists the equivalent Boolean logical interaction rules. Compared to the prior network, most of the newly inferred edges were individual for each cell line, but HCC1954 shared ErbB-1→ERK1/2 with SKBR3, for example. This seemed to be an indirect edge via cRAF, as represented in the prior network. Besides activating connections, also inhibiting ones and edge deletions occurred. For HCC1954, ten new interactions were reconstructed, while two were deleted. In BT474, nine new links were added, and one edge was deleted. In SKBR3, we inferred 20 new connections and one deletion, namely the removal of p53 activation via p38, bearing oncogenic risk [63,64].

In contrast to the short-term networks, new feedback loops were reconstructed in every cell line, not exclusively in HCC1954. In HCC1954, the mutual activation between p53 and RB established such a feedback mechanism. For SKBR3 we even inferred two edges, each forming feedback loops. Contrary to HCC1954, p53 inhibited RB. The second loop connection was inhibition of ErbB-3 by AKT, pointing to a negative feedback against PI3K signalling [65-67].

In HCC1954, the newly inferred edges Cyclin B1→AKT and ErbB-3→ErbB-1 contributed to PI3K signalling, of which the latter was explainable as heterodimers. The newly inferred edge cJUN→ErbB-1 in HCC1954 also indicated raised activity of ErbB-1. Interestingly, in SKBR3 we conducted an inhibiting edge from Cyclin B1 to AKT but instead an activating one to ERK1/2, contributing to MAPK signalling, which was also stated by Abrieu et al. [68]. Another new edge in HCC1954 involved a cell cycle player, i.e. activation of Cyclin D1 by p70S6K [69]. Accordingly, we inferred RPS6→Cyclin D1 in BT474, with RPS6 as downstream target of p70S6K. In SKBR3, the edge p70S6K→Cyclin B1 was reconstructed. A further interesting new activating edge in HCC1954 led from RB to TSC2, while we inferred a reversed inhibition in SKBR3. Searle et al. discussed targeting RB deficient cancers by deactivating TSC2 [70].

Two novel interactions in BT474 activated Cyclin B1, arising from ErbB-1 and ErbB-3, respectively, which meant that mitosis was driven by ErbB-1/3 dimers in this cell line. This indicated a hyperactive PI3K pathway, as revealed in the short-term case.

In SKBR3, we reconstructed an outgoing edge from the artificial network stimulus S, representing full growth

medium, activating AKT. This could be explained as strong activation of AKT, driving PI3K signalling in this cell line. The new edges ErbB-2→TSC2 and ErbB-3→PRAS had to be interpreted as indirect effects, too. They pointed to activity of ErbB-2/3 dimers, feeding into both, MAPK and PI3K, pathways. The edge ErbB-2→TSC2 could imply an oncogenic role of TSC2. Liu et al. discussed a context dependent functionality of TSC2 [71].

Perturbation simulations on long-term networks

Similarly to the perturbation simulations for the short-term networks, we performed those for the long-term networks under all eight initial state combinations of the therapeutics erlotinib, pertuzumab and trastuzumab. Also here, all inferred attractors were simple and consisted of one steady-state. Table 3 contains the simulation results for the attractors of the RPS6 and RB proteins, as those are key players in cell growth and proliferation and mainly comparable to the experimental results of Henjes et al. [28] for HCC1954 and SKBR3. We analysed the attractor states of AKT and ERK1/2, too, but the results are not explicitly listed, since they mostly resembled the ones of RPS6.

The control measurements without any drug treatment should result in activation of ErbB members and dimerisation events, promoting cell growth and proliferation. In fact, this was expressed as reasonable activation of AKT, ERK1/2 and RPS6 in all cell lines, which held for simulations as well as experimental observations. In contrast, the attractor states of RB were inactive in all cell lines (Table 3). Actually, a continuously rising stimulation effect over 30 hours was not observed for HCC1954 and SKBR3 by Henjes et al. [28] either.

The attractor states of RPS6 and RB were identical in all cell lines (Table 3). All drugs, except trastuzumab under stimulation alone, led to inactive attractor states of RPS6. This was also the case for ERK1/2 in all cell lines, as well as AKT in BT474 and HCC1954. In SKBR3, the attractor states of AKT were just inactive without the stimulus. All therapeutics, including trastuzumab, resulted in deactivated attractor states of RB. The statistically inferred drug effects for AKT, ERK1/2, RB and RPS6 were slightly different. Table 4 summarises the optimal drug combinations, confirming and extending the observations of Henjes et al. [28]. Most of them were supported by the perturbation simulation results, corresponding to attractor states of AKT, ERK1/2, RB and RPS6 being zero.

The optimal long-term drug response for AKT and ERK1/2 confirms short-term observations

As shown in Figure 7, the best drug response in BT474 and HCC1954 regarding AKT was yielded for a combination of trastuzumab and erlotinib. Statistically, we inferred no positive effect in BT474 at all, which is explainable by the fact that we just considered a combined effect of drug treatment and time. Although the time courses of AKT signalling with and without the drug treatment were differing in the intensity strength, the signalling profiles were similar. This parallel shift indicated no time effect. Instead, the group effect was significant (p -value $< 2 \cdot 10^{-16}$). This was also the explanation, why we detected erlotinib, but not the combination with trastuzumab, as the optimal treatment in HCC1954 (Table 4). In SKBR3, we inferred the triple drug combination as the optimal one, but the combination of both monoclonal antibodies alone also had a significant effect over time (Figure 7).

Table 3 Attractor states of long-term perturbation simulations

Simulation	BT474				HCC1954				SKBR3			
	RPS6		RB		RPS6		RB		RPS6		RB	
	A	E	A	E	A	E	A	E	A	E	A	E
X	1		0	1	1		0	1	1		0	1
E	0	1	0		0	1	0		0	1	0	
P	0	-	0	-	0	-	0	-	0	-	0	-
T	1	-	0	-	1	-	0	-	1	-	0	-
E,P	0	1	0		0	1	0		0	1	0	
E,T	0	1	0		0	1	0		0		0	
P,T	0		0		0	1	0	1	0		0	
E,P,T	0	1	0		0	1	0	1	0		0	

The therapeutics erlotinib, trastuzumab and pertuzumab, abbreviated by first letters, that were permanently active in the simulated perturbation conditions besides the stimulus S, standing for the full growth medium, are stored in the column *Simulation*. No simulated drug treatment is denoted by 'X'. The A columns hold the attractor states of the proteins RPS6 and RB associated with the perturbations. The E columns contain the protein activity status, statistically deduced from the experimental data. In case of a significant (p -value < 0.05) combined influence of both, drug treatment and time, on the protein signal intensity, a Wilcoxon rank sum test was conducted for the measurements at time point 30 hours. The drug treatments leading to significantly (p -value < 0.05) smaller intensity values compared to the control measurement 'X' were considered as efficient inhibitors, resulting in a table entry of zero. Lacking comparable experiments is labelled as '-', while consistency between simulations and experimental observations is printed in bold.

Table 4 Optimal drug treatment in long-term signalling

Cell line	AKT	ERK1/2	RB	RPS6
BT474	-	TE	E	TP
HCC1954	E	TE	E	-
SKBR3	PTE	TE	TE	TE

The table summarises the optimal drug treatments for the long-term data, leading to inactive AKT, ERK1/2, RB and RPS6, respectively. In case of a significant (p-value < 0.05) combined influence of both, drug treatment and time, on the protein signal intensity, a Wilcoxon rank sum test was conducted for the measurements at time point 30 hours, testing for significantly (p-value < 0.05) smaller intensity values under the drug treatment compared to the control measurement. The drug treatment with the smallest p-value was considered as the optimal inhibitor. No inferred significant positive drug effect is denoted by '-'. The therapeutics erlotinib, trastuzumab and pertuzumab are abbreviated by their first letters. More than one letter denotes drug combinations. The column *Cell line* holds the cell lines under consideration. The columns *AKT*, *ERK1/2*, *RB* and *RPS6* hold the optimal drug combinations for each target. If those were confirmed by the attractor states (0) of perturbation simulations, they are printed in bold.

Hence, like in the short-term results, a drug combination was required to suppress PI3K signalling, here with an obvious need for trastuzumab. For BT474 and HCC1954, this was supported by the simulation results, in which trastuzumab alone had no effect, but was efficient within drug combinations. In HCC1954, even the best drug

response was not convincing (Figure 7), pointing to a dominant PI3K pathway, as revealed in the short-term analysis.

Interestingly, SKBR3 showed a strong activation peak of AKT phosphorylation between 8 and 18 hours (Figure 7), which was just suppressed under combined application of trastuzumab and pertuzumab. We revealed a positive correlation with ERK1/2 (p-value 0.02, Kendall's τ estimate 0.6) and RPS6 (p-value 0.01, Kendall's τ estimate 0.64). The reconstructed edges $S \rightarrow AKT$ and $ErbB-1 \rightarrow ERK1/2$ in SKBR3 indicated strong activation of AKT and ERK1/2. In addition to the prior network, in which AKT and ERK1/2 fed into RPS6 phosphorylation via p70S6K, some of the novel edges pointed to a positive feedback from p70S6K or RPS6 to ERK1/2. The feedback from p70S6K via Cyclin B1, for example, was expressed by the edges $p70S6K \rightarrow Cyclin\ B1$ and $Cyclin\ B1 \rightarrow ERK1/2$. Compared to the short-term results, indicating a dominant MAPK pathway, this long-term observation indicated strong signalling via both, PI3K and MAPK, pathways in SKBR3.

As displayed in Figure 8, erlotinib alone or in combination with trastuzumab showed the optimal effect against ERK1/2 in all of the three cell lines. This was in line with the short-term observations, and confirmed by the

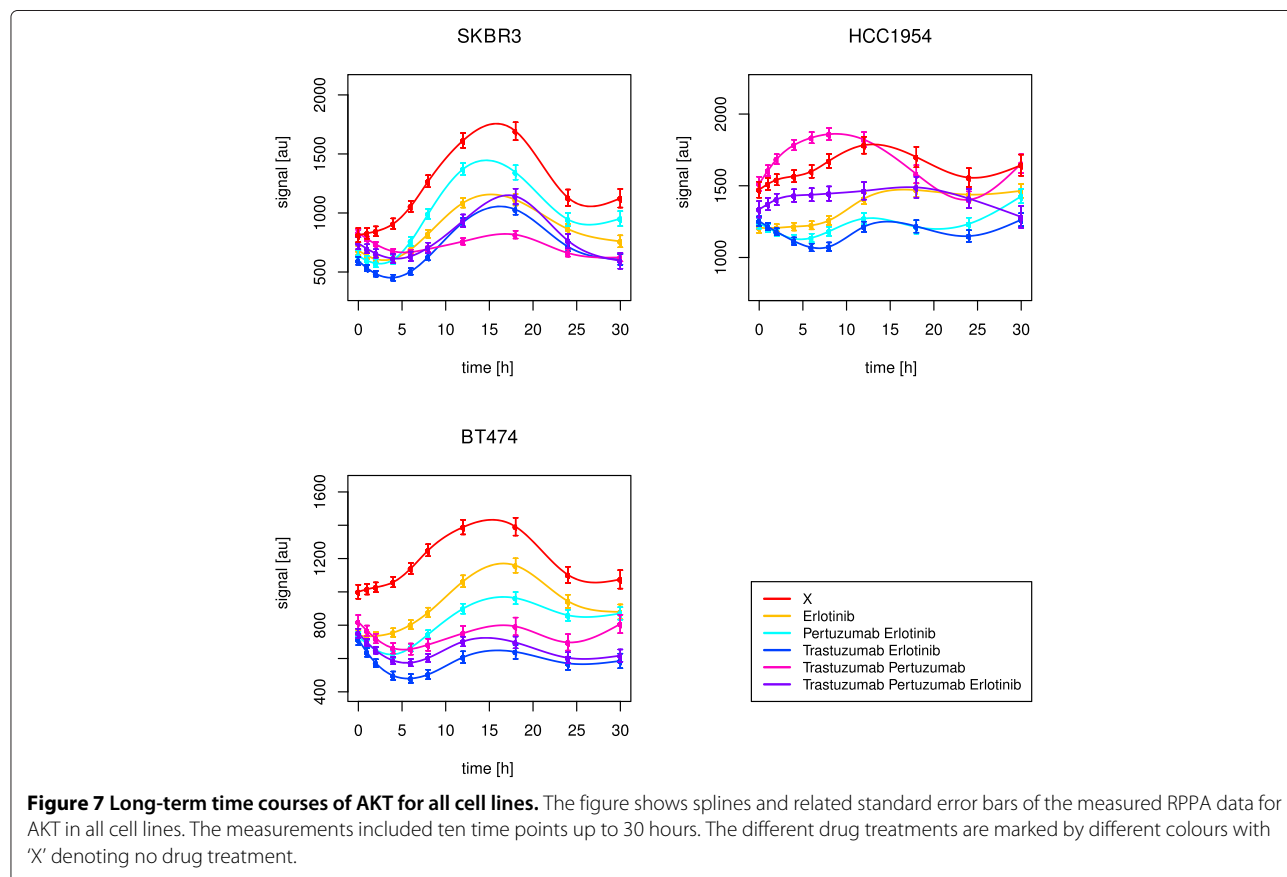


Figure 7 Long-term time courses of AKT for all cell lines. The figure shows splines and related standard error bars of the measured RPPA data for AKT in all cell lines. The measurements included ten time points up to 30 hours. The different drug treatments are marked by different colours with 'X' denoting no drug treatment.

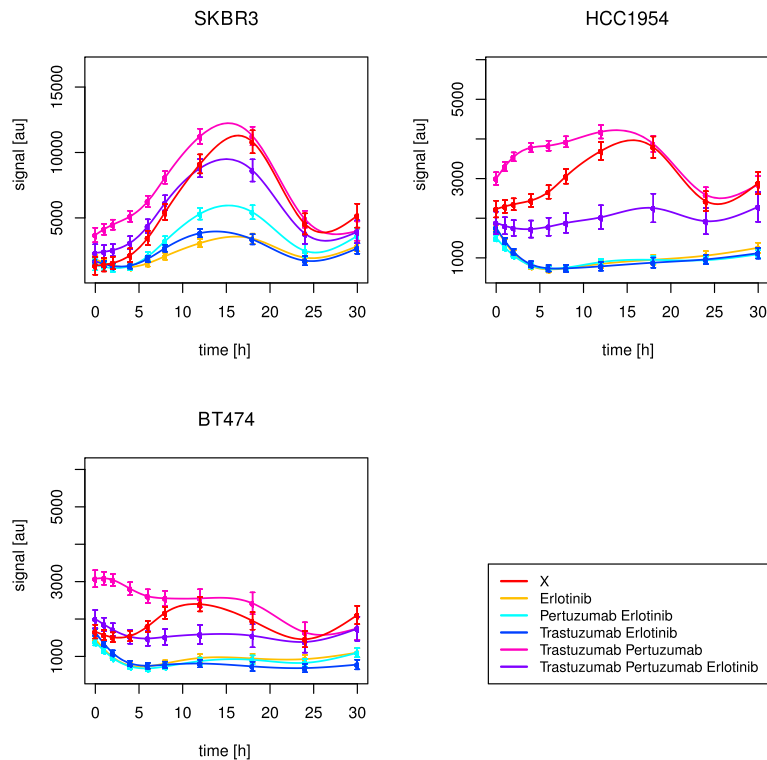


Figure 8 Long-term time courses of ERK1/2 for all cell lines. The figure shows splines and related standard error bars of the measured RPPA data for ERK1/2 in all cell lines. The measurements included ten time points up to 30 hours. The different drug treatments are marked by different colours with 'X' denoting no drug treatment.

perturbation simulations. Statistically, the most potent drug effect was yielded with the combination of erlotinib and trastuzumab (Table 4).

Quick drug response for RPS6 and delayed response for RB

As shown in Figure 9, in BT474, the simulation based predicted efficacy of erlotinib alone to counteract RPS6 (Table 3) was not as convincing as in case of drug combinations. A combination of pertuzumab and trastuzumab worked best (Table 4). For RB, the simulated drug effects in BT474 resembled the observed ones (Table 3, Figure 10), with a positive effect of all measured drug treatments. Erlotinib was inferred as the optimal treatment (Table 4). Though, the drug impact unfolded not before 18 hours.

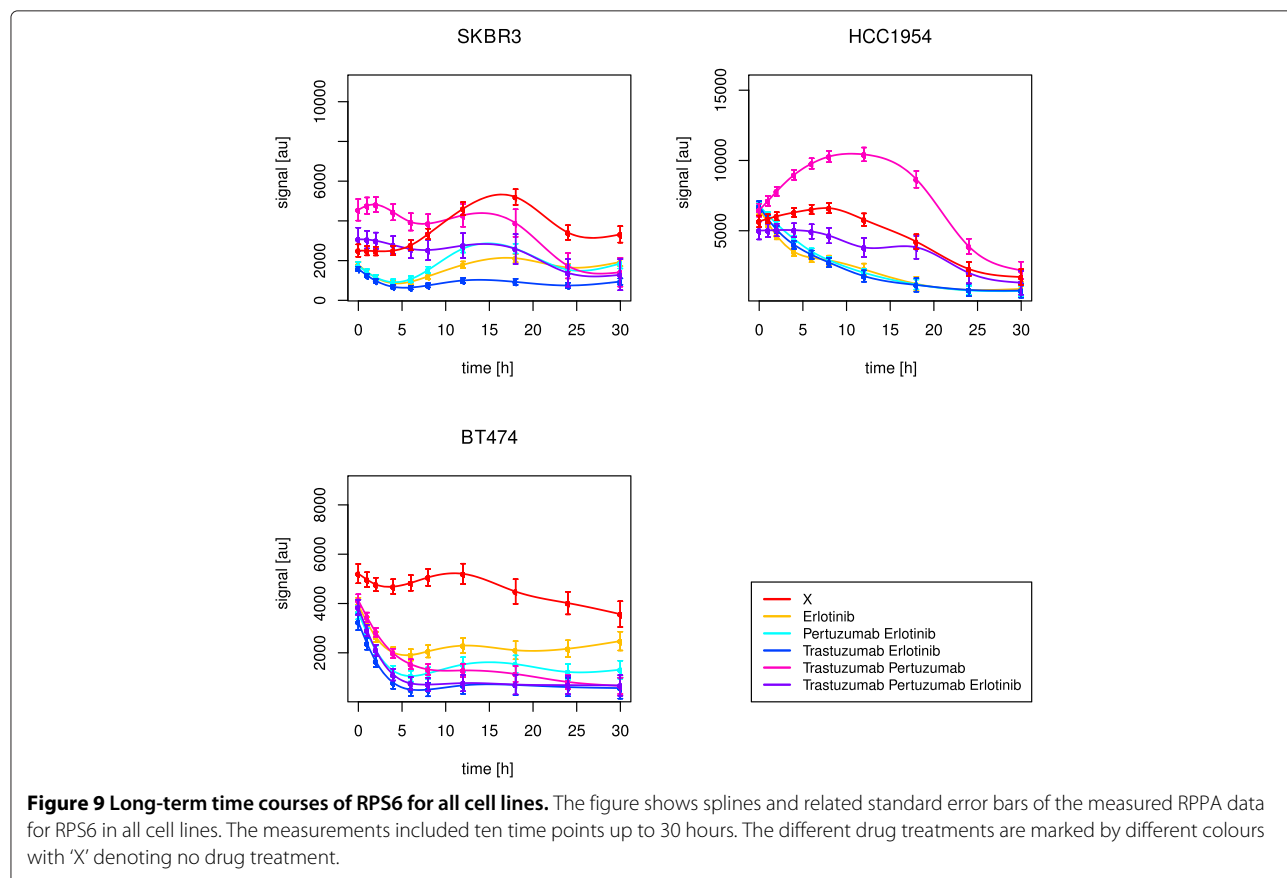
In HCC1954, it was the combination of both monoclonal antibodies, that failed in deactivating RPS6 (Figure 9), while the simulations predicted trastuzumab alone to fail (Table 3). The graphical observations were similar for RB (Figure 10). The newly inferred edges ErbB-3→ErbB-1 and cJUN→ErbB-1 in HCC1954 explained the necessity for erlotinib against ErbB-1 dimers. The positive impact of erlotinib, the optimal treatment against RB signalling (Table 4), was supported by simulations. However, it did not unfold before 12-18 hours, in case of RB as

well as RPS6. Regarding RPS6, no significant effect was detected for HCC1954 (Table 4).

According to Henjes et al. [28], in SKBR3 erlotinib and all therapeutic combinations helped to suppress RPS6, which was supported by the simulations (Table 3). As shown in Figure 9, the combination of trastuzumab and erlotinib was the only one, that revealed its continuous inhibiting effect already after one hour. This combined treatment was also statistically inferred as the optimal one (Table 4). The same combination was optimal with respect to RB activity, which was also in line with the simulations. Here, analogously to BT474 and HCC1954, the drug effect did not appear before 18 hours (Figure 10).

As the combination of trastuzumab and erlotinib was efficient in all of the three cell lines against RPS6 as well as RB phosphorylation, we further analysed target correlations under this drug combination to explain the different rapidness of drug responses.

In BT474, RB positively correlated with Cyclin B1 (p-value 0.02, Kendall's τ estimate 0.6), while RPS6 positively correlated with ERK1/2 (p-value 0.02, Kendall's τ estimate 0.6). Obviously, RPS6 was mainly stimulated by the MAPK pathway, which was efficiently inhibited by the combination of trastuzumab and erlotinib in a fast manner. On the contrary, RB seemed to be influenced



by Cyclin B1. The newly reconstructed edges ErbB-1→Cyclin B1 and ErbB-3→Cyclin B1 supported hyperactivity of Cyclin B1, driven by ErbB-1/3 heterodimers. In SKBR3, RB negatively correlated with PRAS (p-value 0.05, Kendall's τ estimate -0.51) and TSC2 (p-value 0.03, Kendall's τ estimate -0.56), while RPS6 positively correlated with AKT (p-value 0.03, Kendall's τ estimate 0.56) and ERK1/2 (p-value 0.02, Kendall's τ estimate 0.6). Obviously, like in BT474, RPS6 was mainly activated through the MAPK pathway. Interestingly, RB seemed to require inhibition via PRAS or TSC2. The latter was confirmed via one of the novel edges in SKBR3, namely inhibition of RB by TSC2. In addition, PRAS as well as TSC2 seemed to be especially active in this cell line with regard to the new edges ErbB-3→PRAS and ErbB-2→TSC2.

In HCC1954, the drug response was not only delayed for RB, but also for RPS6, which was in line with the positive correlation with RB (p-value $< 2.22 \cdot 10^{-16}$, Kendall's τ estimate 0.73). Like in BT474, Cyclin B1 seemed to be a driving force, since both, RPS6 (p-value 0.03, Kendall's τ estimate 0.56) and RB (p-value $< 2.22 \cdot 10^{-16}$, Kendall's τ estimate 0.82) positively correlated with this target. The new edge Cyclin B1→AKT supported special activation of RPS6 via PI3K signalling, leading to a delayed drug response. Interestingly, we revealed negative correlations,

as observed for SKBR3. In HCC1954, RPS6 and RB correlated with BAX (p-value 0.03, Kendall's τ estimate -0.56) and FoxO1/3a (p-value 0.05, Kendall's τ estimate -0.51), pointing to a delayed inhibition of RPS6 and RB via BAX or FoxO1/3a.

Conclusions

Using a combination of reverse and forward engineering techniques, we focused on deregulated protein interactions in the ErbB network in a Boolean modelling framework. The reconstructed hypothetical networks revealed individual protein interactions contributing to signalling pathway preferences as well as drug resistance via feedback loops, pathway crosstalk or hyperactive heterodimers. While this reverse engineering focused on the network edges, we concentrated in the subsequent forward engineering step on the network nodes. The perturbation simulations for AKT, ERK1/2, RB and RPS6 mainly confirmed our graphical and statistical analyses as well as the observations of Henjes et al. [28] regarding (combinatorial) drug efficacy. However they have to be interpreted as an independent, more prospective investigation, because stable system states do not necessarily have to be reached in temporally limited observations.

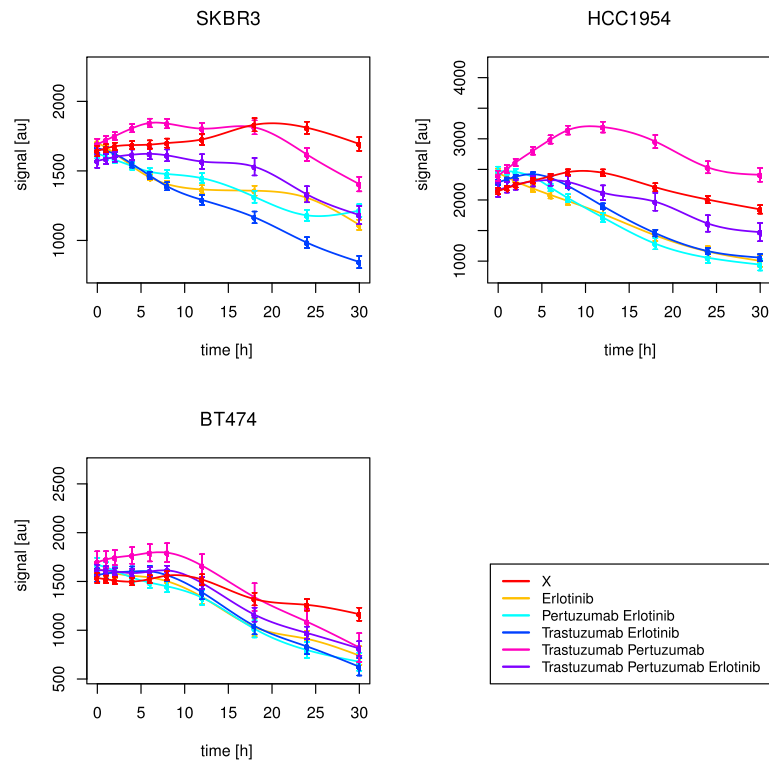


Figure 10 Long-term time courses of RB for all cell lines. The figure shows splines and related standard error bars of the measured RPPA data for RB in all cell lines. The measurements included ten time points up to 30 hours. The different drug treatments are marked by different colours with 'X' denoting no drug treatment.

In the first step, the combined Boolean modelling approach revealed the mechanisms underlying individual drug response. In the second step, it predicted the network propagation effects on protein activity, and hence the drug response itself.

One major finding is, that different breast cancer phenotypes seem to be driven by specific pathway preferences in the ErbB network. This leads to individual drug response, requiring different therapeutic treatments. The perturbation simulations revealed a more diverse drug response in short-term than in long-term signalling, which stresses the importance of early intervention at the top level layer of the signalling network.

Another interesting aspect is to combine edge and node perturbations in Boolean network models to reveal edgetic mutations, as we did in the HCC1954 cell line for AKT.

Basic molecular research, embedded in a Boolean modelling framework here, composes a first step to gain insight into individual mechanisms of drug response or resistance mechanisms in breast cancer. Especially, the proteomic signalling interplay directly effects tumour development and represents a promising target in cancer therapy, which has to be understood in more detail in the future.

Additional files

Additional file 1: Proteins and phosphorylation sites involved in RPPA measurements. The tables show the proteins and phosphorylation sites involved in RPPA short- and long-term measurements. The antibody catalogue numbers and providing companies are mentioned in brackets. For BT474, no experimental short-term data under EGF or HRG stimulation were available for PDK1. In case of total protein measurements, the column *Phosphosite* remains empty ('-') apart from the antibody number and supplier name.

Additional file 2: Literature references for the prior networks of short- and long-term signalling. The interactions between proteins are listed line by line in the tables. The column *Protein* denotes the source of the connection with the sink called *Target*. The interaction (*Type*) is encoded numerically, i.e. activation is marked by 1, while inhibition is labelled with 2, e.g. AKT activates mTOR. The column *Reference* specifies the supportive publication.

Additional file 3: Workflow of MCMC-based network structure inference. The *inhibMCMC* procedure of the *ddepn* package was run with *maxIter* = 50,000 in 10 parallel runs. The results of the 25,000 iterations after the burn-in phase were merged into one consensus network. It was applied for short- and long-term data separately per cell line, leading to six consensus networks. The figure is based on [6,47].

Additional file 4: Edge confidences for the reconstructed networks of the three cell lines per time course. In each of the ten MCMC runs, activation and inhibition edges were sampled. The percentage, i.e. the confidence, of sampled activation (red) and inhibition (blue) edges in the 25,000 iterations after the burn-in phase are depicted in the boxes. The sink nodes are displayed in each panel, while the activating, inhibiting or

missing influence of the source nodes is shown column-wise in the red, blue or missing boxes. The source node names are displayed at the x-axis with additional indicators, where ‘-’ refers to an inhibiting influence and ‘+’ is related to activation. The x-axis of the short-term plots is labelled as ‘AKT, E, EGF, ERBB1, ERBB2, ERBB3, ERK1/2, HRG, MEK1/2, mTOR, P, p70S6K, PDK1, PKC α , PLC γ , T’. The x-axis of the long-term plots is labelled as ‘AKT, BAX, cJUN, cRAF, CyclinB1, CyclinD1, E, ERBB1, ERBB2, ERBB3, ERK1/2, FOXO1/3a, GSK3 α/β , NF- κ B, P, p38, p53, p70S6K, PRAS, PTEN, RB, S, RPS6, T, TSC2’. An activating edge in the consensus network, as described in Additional file 3, means that the sampled activating edges have a significantly higher confidence value than the inhibiting ones. As self-loops and ingoing edges to the drug or growth factor nodes were not allowed during inference, the respective confidences are zero.

Additional file 5: Boolean interaction rules for the components of the short- and long-term signalling networks. The tables contain the rules that arose from network reconstructions based on short- and long-term RPPA data of BT474, HCC1954 and SKBR3. The three drug names erlotinib, trastuzumab and pertuzumab are abbreviated via their first letters. For the long-term networks, the stimulus is denoted by S. Symbols are interpretable in the following way: $\&$ \equiv AND, \vee \equiv OR and $!$ \equiv NOT.

Additional file 6: Reconstructed long-term signalling networks. The figure displays the reconstructed long-term signalling networks for BT474, HCC1954 and SKBR3. Target proteins are represented as rectangles with stimulus and drugs coloured in red. The three drug names erlotinib, trastuzumab and pertuzumab are abbreviated via their first letters. Stimulation via full growth medium is denoted by S. Solid arrows denote activating interactions while dashed ones represent inhibitions.

Abbreviations

DDEPN: Dynamic deterministic effects propagation networks; EGF: Epidermal growth factor; EGFR: Epidermal growth factor receptor; GEO: Gene expression omnibus; HER2: Human epidermal growth factor receptor 2; HMM: Hidden Markov model; HRG: Heregulin; MAPK: Mitogen-activated protein kinase; MCMC: Markov chain Monte Carlo; NIR: Near infrared; PI3K: Phosphoinositide 3-kinase; RPPA: Reverse phase protein array; RTK: Receptor tyrosine kinase.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

FH performed the RPPA measurements under supervision of UK and was mainly involved in target selection for the modelling approach. JS was involved in discussing the conducted RPPA experiments. CB and TB developed the applied network reconstruction algorithm and participated in planning the modelling procedure. TB and SvdH initiated the simulation study concepts. SvdH carried out the literature research, network reconstructions, perturbation simulations followed by associated analyses, and finally drafting the manuscript. All authors edited, read and approved the final manuscript.

Acknowledgements

Pertuzumab, trastuzumab and erlotinib were provided by Roche Diagnostics GmbH, Penzberg, Germany. This work was supported by a grant from the German Federal Ministry of Education and Research (BMBF) within the Medical Systems Biology programme BreastSys. We also acknowledge support by the Open Access Publication Funds of the Göttingen University.

Author details

¹Statistical Bioinformatics, Department of Medical Statistics, University Medical Center Göttingen, Humboldtallee 32, 37073 Göttingen, Germany. ²TRON - Translational Oncology at the University Medical Center Mainz, Langenbeckstraße 1, 55131 Mainz, Germany. ³Science for Life Laboratory, School of Biotechnology, KTH - Royal Institute of Technology, Box 1031, 17121 Solna, Sweden. ⁴Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120, Heidelberg, Germany.

Received: 20 December 2013 Accepted: 10 June 2014
Published: 25 June 2014

References

- Hill SM, Lu Y, Molina J, Heiser LM, Spellman PT, Speed TP, Gray JW, Mills GB, Mukherjee S: **Bayesian inference of signaling network topology in a cancer cell line.** *Bioinformatics (Oxford, England)* 2012, **28**(21):2804–2810. PMID: 22923301.
- Park Y, Bader JS: **How networks change with time.** *Bioinformatics (Oxford, England)* 2012, **28**(12):40–48. PMID: 22689777.
- Roukos DH: **Trastuzumab and beyond: sequencing cancer genomes and predicting molecular networks.** *Pharmacogenom J* 2011, **11**(2):81–92. PMID: 20975737.
- Oda K, Matsuoka Y, Funahashi A, Kitano H: **A comprehensive pathway map of epidermal growth factor receptor signaling.** *Mol Syst Biol* 2005, **1**:2005.0010. PMID: 16729045.
- Feiglin A, Hacohen A, Sarusi A, Fisher J, Unger R, Ofra Y: **Static network structure can be used to model the phenotypic effects of perturbations in regulatory networks.** *Bioinformatics (Oxford, England)* 2012, **28**(21):2811–2818. PMID: 22923292.
- Bender C, Heyde SV, Henjes F, Wiemann S, Korf U, Beissbarth T: **Inferring signalling networks from longitudinal data using sampling based approaches in the r-package ‘ddepn’.** *BMC Bioinformatics* 2011, **12**:291. PMID: 21771315.
- Penfold CA, Buchanan-Wollaston V, Denby KJ, Wild DL: **Nonparametric bayesian inference for perturbed and orthologous gene regulatory networks.** *Bioinformatics* 2012, **28**(12):233–241. PMID: 22689766.
- Wagner JP, Wolf-Yadlin A, Sevecka M, Grenier JK, Root DE, Lauffenburger DA, MacBeath G: **Receptor tyrosine kinases fall into distinct classes based on their inferred signaling networks.** *Sci Signaling* 2013, **6**(284):58.
- Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, Lauffenburger DA, Sorger PK: **Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data.** *Mol Syst Biol* 2009, **5**:PMID: 19156131.
- Hatakeyama M, Kimura S, Naka T, Kawasaki T, Yumoto N, Ichikawa M, Kim J-H, Saito K, Saeki M, Shirouzu M, Yokoyama S, Konagaya A: **A computational model on the modulation of mitogen-activated protein kinase (MAPK) and akt pathways in heregulin-induced ErbB signalling.** *Biochem J* 2003, **373**(Pt 2):451–463. PMID: 12691603.
- Jones RB, Gordus A, Krall JA, MacBeath G: **A quantitative protein interaction network for the ErbB receptors using protein microarrays.** *Nature* 2006, **439**(7073):168–174. PMID: 16273093.
- Schoeberl B, Eichler-Jonsson C, Gilles ED, Müller G: **Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors.** *Nat Biotechnol* 2002, **20**(4):370–375. PMID: 11923843.
- Sahin O, Fröhlich H, Läubke C, Korf U, Burmester S, Majety M, Mattern J, Schupp I, Chaoiyya C, Thieffry D, Poustka A, Wiemann S, Beissbarth T, Arlt D: **Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance.** *BMC Syst Biol* 2009, **3**:1. PMID: 19118495.
- Samaga R, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, Klamt S: **The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data.** *PLoS Comput Biol* 2009, **5**(8):1000438. PMID: 19662154.
- Wang R-S, Saadatpour A, Albert R: **Boolean modeling in systems biology: an overview of methodology and applications.** *Phys Biol* 2012, **9**(5):055001.
- McDermott JE, Wang J, Mitchell H, Webb-Robertson B-J, Hafen R, Ramey J, Rodland KD: **Challenges in biomarker discovery: Combining expert insights with statistical analysis of complex omics data.** *Expert Opin Med Diag* 2013, **7**(1):37–51. PMID: 23335946.
- Eduati F, De Las Rivas J, Di Camillo B, Toffolo G, Saez-Rodriguez J: **Integrating literature-constrained and data-driven inference of signalling networks.** *Bioinformatics (Oxford, England)* 2012, **28**(18):2311–2317. PMID: 22734019.
- Terfve C, Cokelaer T, Henriques D, MacNamara A, Goncalves E, Morris MK, van Iersel M, Lauffenburger DA, Saez-Rodriguez J: **CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms.** *BMC Syst Biol* 2012, **6**:133. PMID: 23079107.

19. Albert I, Thakar J, Li S, Zhang R, Albert R: **Boolean network simulations for life scientists**. *Source Code Biol Med* 2008, **3**:16. PMID: 19014577.
20. R Core Team: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2012. R Foundation for Statistical Computing. ISBN 3-900051-07-0. <http://www.r-project.org>.
21. Müssel C, Hopfensitz M, Kestler HA: **BoolNet—an r package for generation, reconstruction and analysis of boolean networks**. *Bioinformatics (Oxford, England)* 2010, **26**(10):PMID: 20378558.
22. Samaga R, Klamt S: **Modeling approaches for qualitative and semi-quantitative analysis of cellular signaling networks**. *Cell Commun Signaling: CCS* 2013, **11**(1):43. PMID: 23803171 PMCID: PMC3698152.
23. Gonzalez AG, Naldi A, Sánchez L, Thieffry D, Chaouiya C: **GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks**. *Bio Syst* 2006, **84**(2):91–100. PMID: 16434137.
24. Di Cara A, Garg A, De Micheli G, Xenarios I, Mendoza L: **Dynamic simulation of regulatory networks using SQUAD**. *BMC Bioinformatics* 2007, **8**:462. PMID: 18039375 PMCID: PMC2238325.
25. Helikar T, Rogers JA: **ChemChains: a platform for simulation and analysis of biochemical networks aimed to laboratory scientists**. *BMC Syst Biol* 2009, **3**:58. PMID: 19500393 PMCID: PMC2705353.
26. Krumsiek J, Pösterl S, Wittmann DM, Theis FJ: **Odefy—from discrete to continuous models**. *BMC Bioinformatics* 2010, **11**:233. PMID: 20459647 PMCID: PMC2873544.
27. Ferlay J, Shin H-R, Bray F, Forman D, Mathers C, Parkin DM: **Estimates of worldwide burden of cancer in 2008: Globocan 2008**. *Int J Cancer* 2010, **127**(12):2893–2917.
28. Henjes F, Bender C, Heyde SVD, Braun L, Mannsperger HA, Schmidt C, Wiemann S, Hasmann M, Aulmann S, Beissbarth T, Korf U: **Strong EGFR signaling in cell line models of ERBB2-amplified breast cancer attenuates response towards ERBB2-targeting drugs**. *Oncogenesis* 2012, **1**(7):16.
29. Olayioye MA, Neve RM, Lane HA, Hynes NE: **The ErbB signaling network: receptor heterodimerization in development and cancer**. *EMBO J* 2000, **19**(13):PMID: 10880430.
30. Heil J, Gondos A, Rauch G, Marmé F, Rom J, Golatta M, Junkermann H, Sinn P, Aulmann S, Debus J, Hof H, Schütz F, Brenner H, Sohn C, Schneeweiss A: **Outcome analysis of patients with primary breast cancer initially treated at a certified academic breast unit**. *Breast (Edinburgh, Scotland)* 2012, **21**(3):303–308. PMID: 22310244.
31. Jelovac D, Wolff AC: **The adjuvant treatment of HER2-positive breast cancer**. *Curr Treat Options Oncol* 2012, **13**(2):230–239. PMID: 22410709.
32. Park JW, Neve RM, Szollosi J, Benz CC: **Unraveling the biologic and clinical complexities of HER2**. *Clin Breast Cancer* 2008, **8**(5):392–401. PMID: 18952552.
33. Tinoco G, Warsch S, Glüsk S, Avancha K, Montero AJ: **Treating breast cancer in the 21st century: emerging biological therapies**. *J Cancer* 2013, **4**(2):117–132. PMID: 23386910.
34. Heyde Svd, Beissbarth T: **A new analysis approach of epidermal growth factor receptor pathway activation patterns provides insights into cetuximab resistance mechanisms in head and neck cancer**. *BMC Medicine* 2012, **10**(1):43. PMID: 22548923.
35. Motoyama AB, Hynes NE, Lane HA: **The efficacy of ErbB receptor-targeted anticancer therapeutics is influenced by the availability of epidermal growth factor-related peptides**. *Cancer Res* 2002, **62**(11):3151–3158.
36. Wilson TR, Fridlyand J, Yan Y, Penuel E, Burton L, Chan E, Peng J, Lin E, Wang Y, Sosman J, Ribas A, Li J, Moffat J, Sutherland DP, Koeppen H, Merchant M, Neve R, Settleman J: **Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors**. *Nature* 2012, **487**(7408):505–509. PMID: 22763448.
37. Gallardo A, Lerma E, Escuin D, Tibau A, Muñoz J, Ojeda B, Barnadas A, Adrover E, Sánchez-Tejada L, Giner D, Ortiz-Martínez F, Peiró G: **Increased signalling of EGFR and IGF1R, and deregulation of PTEN/PI3K/Akt pathway are related with trastuzumab resistance in HER2 breast carcinomas**. *Br J Cancer* 2012, **106**(8):1367–1373. PMID: 22454081.
38. Wang L, Zhang Q, Zhang J, Sun S, Guo H, Jia Z, Wang B, Shao Z, Wang Z, Hu X: **PI3K pathway activation results in low efficacy of both trastuzumab and lapatinib**. *BMC Cancer* 2011, **11**:248. PMID: 21676217.
39. Diermeier S, Horváth G, Knuedel-Clarke R, Hofstaedter F, Söllros J, Brockhoff G: **Epidermal growth factor receptor coexpression modulates susceptibility to herceptin in HER2/neu overexpressing breast cancer cells via specific erbB-receptor interaction and activation**. *Exp Cell Res* 2005, **304**(2):604–619. PMID: 15748904.
40. Pallis AG, Syrigos KN: **Epidermal growth factor receptor tyrosine kinase inhibitors in the treatment of NSCLC**. *Lung cancer (Amsterdam, Netherlands)* (2013). PMID: 23384674.
41. Moore MJ, Goldstein D, Hamm J, Figer A, Hecht JR, Gallinger S, Au HJ, Murawa P, Walde D, Wolff RA, Campos D, Lim R, Ding K, Clark G, Voskoglou-Nomikos T, Ptasynski M, Parulekar W, National Cancer, Institute of Canada Clinical Trials Group: **Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the national cancer institute of canada clinical trials group**. *J Clin Oncol* 2007, **25**(15):1960–1966. PMID: 17452677.
42. Loebke C, Suelmann H, Schmidt C, Henjes F, Wiemann S, Poustka A, Korf U: **Infrared-based protein detection arrays for quantitative proteomics**. *PROTEOMICS* 2007, **7**(4):558–564.
43. Kataoka Y, Mukohara T, Shimada H, Saijo N, Hirai M, Minami H: **Association between gain-of-function mutations in PIK3CA and resistance to HER2-targeted agents in HER2-amplified breast cancer cell lines**. *Ann Oncol* 2010, **21**(2):255–262. PMID: 19633047.
44. Nahta R, Yuan LXH, Zhang B, Kobayashi R, Esteva FJ: **Insulin-like growth factor-1 receptor/human epidermal growth factor receptor 2 heterodimerization contributes to trastuzumab resistance of breast cancer cells**. *Cancer Res* 2005, **65**(23):11118–11128. PMID: 16322262.
45. Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, Latham V, Sullivan M: **PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse**. *Nucleic Acids Res* 2011, **40**(D1):261–270.
46. Fröhlich H, Sahin O, Arlt D, Bender C, Beissbarth T: **Deterministic effects propagation networks for reconstructing protein signaling networks from multiple interventions**. *BMC Bioinformatics* 2009, **10**:322. PMID: 19814779.
47. Bender C, Henjes F, Fröhlich H, Wiemann S, Korf U, Beissbarth T: **Dynamic deterministic effects propagation networks: learning signalling pathways from longitudinal protein array data**. *Bioinformatics (Oxford, England)* 2010, **26**(18):596–602. PMID: 20823327.
48. Almudevar A, McCall MN, McMurray H, Land H: **Fitting boolean networks from steady state perturbation data**. *Stat Appl Genet Mol Biol* 2011, **10**(1):1–40.
49. Brooks SP, Gelman A: **General methods for monitoring convergence of iterative simulations**. *J Comput Graph Stat* 1998, **7**(4):434–455.
50. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: A practical and powerful approach to multiple testing**. *J R Stat Soc Series B (Methodological)* 1995, **57**(1):289–300. ArticleType: research-article/Full publication date: 1995/Copyright ©1995 Royal Statistical Society.
51. Bo Na Ki, D, Odenbrett MR, Wijs A, Ligtenberg W, Hilbers P: **Efficient reconstruction of biological networks via transitive reduction on general purpose graphics processors**. *BMC Bioinformatics* 2012, **13**(1):PMID: 23110660.
52. Mikalsen T, Gerits N, Moens U: **Inhibitors of signal transduction protein kinases as targets for cancer therapy**. In M Raafat El-Gewely (ed.) *Biotechnology Annual Review. Volume 12*: Elsevier:153–223. ISBN: 1387-2656 2006. <http://www.sciencedirect.com/science/article/pii/S1387265606120062>.
53. Dienstmann R, De Dosso S, Felip E, Tabernero J: **Drug development to overcome resistance to EGFR inhibitors in lung and colorectal cancer**. *Mol Oncol* 2012, **6**(1):15–26.
54. Esteva FJ, Pusztai L: **Optimizing outcomes in HER2-positive breast cancer: the molecular rationale**. *Oncology (Williston Park, N.Y.)* 2005, **19**(13 Suppl 5):5–16. PMID: 19364051.
55. Sato S, Fujita N, Tsuruo T: **Involvement of 3-phosphoinositide-dependent protein kinase-1 in the MEK/MAPK signal transduction pathway**. *J Biol Chem* 2004, **279**(32):33759–33767. PMID: 15175348.
56. Maurer M, Su T, Saal LH, Koujak S, Hopkins BD, Barkley CR, Wu J, Nandula S, Dutta B, Xie Y, Chin YR, Kim D-I, Ferris JS, Gruberger-Saal SK, Laakso M, Wang X, Memeo L, Rojzman A, Matos T, Yu JS, Cordon-Cardo C, Isola J, Terry MB, Toker A, Mills GB, Zhao JJ, Murty VV, Hibshoosh H, Parsons R: **3-phosphoinositide-dependent kinase 1 potentiates upstream lesions on the phosphatidylinositol 3-kinase pathway in breast carcinoma**. *Cancer Res* 2009, **69**(15):6299–6306. PMID: 19602588.

57. Tseng P-H, Wang Y-C, Weng S-C, Weng J-R, Chen C-S, Brueggemeier RW, Shapiro CL, Chen C-Y, Dunn SE, Pollak M, Chen C-S: **Overcoming trastuzumab resistance in HER2-overexpressing breast cancer cells by using a novel celecoxib-derived phosphoinositide-dependent kinase-1 inhibitor.** *Mol Pharmacol* 2006, **70**(5):PMID: 16887935.
58. Vega F, Medeiros LJ, Leventaki V, Atwell C, Cho-Vega JH, Tian L, Claret F-X, Rassidakis GZ: **Activation of mammalian target of rapamycin signaling pathway contributes to tumor cell survival in anaplastic lymphoma kinase-positive anaplastic large cell lymphoma.** *Cancer Res* 2006, **66**(13):6589–6597. PMID: 16818631.
59. Frödin M, Jensen CJ, Merienne K, Gammeltoft S: **A phosphoserine-regulated docking site in the protein kinase RSK2 that recruits and activates PDK1.** *EMBO J* 2000, **19**(12):2924–2934. PMID: 10856237.
60. Klos KS, Wyszomierski SL, Sun M, Tan M, Zhou X, Li P, Yang W, Yin G, Hittelman WN, Yu D: **ErbB2 increases vascular endothelial growth factor protein synthesis via activation of mammalian target of rapamycin/p70S6K leading to increased angiogenesis and spontaneous metastasis of human breast cancer cells.** *Cancer Res* 2006, **66**(4):2028–2037. PMID: 16489002.
61. Schaefer G, Shao L, Totpal K, Akita RW: **Erlotinib directly inhibits HER2 kinase activation and downstream signaling events in intact cells lacking epidermal growth factor receptor expression.** *Cancer Res* 2007, **67**(3):1228–1238. PMID: 17283159.
62. Zhong Q, Simonis N, Li Q-R, Charlotiaux B, Heuze F, Klitgord N, Tam S, Yu H, Venkatesan K, Mou D, Swearingen V, Yildirim MA, Yan H, Dricot A, Szeto D, Lin C, Hao T, Fan C, Milstein S, Dupuy D, Brasseur R, Hill DE, Cusick ME, Vidal M: **Edgetic perturbation models of human inherited disorders.** *Mol Syst Biol* 2009, **5**:321. PMID: 19888216.
63. McCurrach ME, Connor TM, Knudson CM, Korsmeyer SJ, Lowe SW: **bax-deficiency promotes drug resistance and oncogenic transformation by attenuating p53-dependent apoptosis.** *Proc Nat Acad Sci USA* 1997, **94**(6):2345–2349. PMID: 9122197.
64. Sherr CJ, McCormick F: **The RB and p53 pathways in cancer.** *Cancer Cell* 2002, **2**(2):103–112. PMID: 12204530.
65. Sithanandam G, Anderson LM: **The ERBB3 receptor in cancer and cancer gene therapy.** *Cancer Gene Therapy* 2008, **15**(7):413–448. PMID: 18404164.
66. Lynch DK, Daly RJ: **PKB-mediated negative feedback tightly regulates mitogenic signalling via gab2.** *EMBO J* 2002, **21**(1-2):72–82. PMID: 11782427.
67. Chakrabarty A, Sánchez V, Kuba MG, Rinehart C, Arteaga CL: **Feedback upregulation of HER3 (ErbB3) expression and activity attenuates antitumor effect of PI3K inhibitors.** *Proc Nat Acad Sci USA* 2012, **109**(8):2718–2723. PMID: 21368164.
68. Abrieu A, Dorée M, Fisher D: **The interplay between cyclin-b-cdc2 kinase (MPF) and MAP kinase during maturation of oocytes.** *J Cell Sci* 2001, **114**(Pt 2):257–267. PMID: 11148128.
69. Jirmanova L, Afanassieff M, Gobert-Gosse S, Markossian S, Savatier P: **Differential contributions of ERK and PI3-kinase to the regulation of cyclin d1 expression and to the control of the G1/S transition in mouse embryonic stem cells.** *Oncogene* 2002, **21**(36):5515–5528. PMID: 12165850.
70. Searle JS, Li B, Du W: **Targeting rb mutant cancers by inactivating TSC2.** *Oncotarget* 2010, **1**(3):228–232. PMID: 20706560.
71. Liu H, Radisky DC, Nelson CM, Zhang H, Fata JE, Roth RA, Bissell MJ: **Mechanism of akt1 inhibition of breast cancer cell invasion reveals a protumorigenic role for TSC2.** *Proc Nat Acad Sci USA* 2006, **103**(11):4134–4139. PMID: 16537497.

doi:10.1186/1752-0509-8-75

Cite this article as: von der Heyde et al.: Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines. *BMC Systems Biology* 2014 **8**:75.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



Appendix B

Publication “RPPanalyzer Toolbox: An improved R package for analysis of reverse phase protein array data”

RPPanalyzer Toolbox: An improved R package for analysis of reverse phase protein array data

Silvia von der Heyde¹, Johanna Sonntag², Daniel Kaschek³, Christian Bender⁴, Johannes Bues², Astrid Wachter¹, Jens Timmer^{3,5,6}, Ulrike Korf², and Tim Beißbarth¹

¹Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany, ²Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany, ³Institute of Physics, Freiburg University, Freiburg, Germany, ⁴TRON–Translational Oncology at the Johannes Gutenberg University of Mainz Medicine, Mainz, Germany, ⁵BIOSS Centre for Biological Signalling Studies, Freiburg University, Freiburg, Germany, and ⁶Freiburg Center for Systems Biology (ZBSA), Freiburg University, Freiburg, Germany

BioTechniques 57:125–135 (September 2014) doi 10.2144/000114205

Keywords: data analysis toolbox; open-source software; proteomics; RPPA; RPPanalyzer

Supplementary material for this article is available at www.BioTechniques.com/article/114205.

Analysis of large-scale proteomic data sets requires specialized software tools, tailored toward the requirements of individual approaches. Here we introduce an extension of an open-source software solution for analyzing reverse phase protein array (RPPA) data. The R package *RPPanalyzer* was designed for data preprocessing followed by basic statistical analyses and proteomic data visualization. In this update, we merged relevant data preprocessing steps into a single user-friendly function and included a new method for background noise correction as well as new methods for noise estimation and averaging of replicates to transform data in such a way that they can be used as input for a new time course plotting function. We demonstrate the robustness of our enhanced *RPPanalyzer* platform by analyzing longitudinal RPPA data of MET receptor signaling upon stimulation with different hepatocyte growth factor concentrations.

In principle, RPPAs are miniaturized dot blot immunoassays, enabling quantitative analysis of target protein abundance, including posttranslational modifications, across large sample sets. Originally introduced by Pawletz et al. as a robust and reproducible proteomics resource (1), over the past decade the RPPA approach has become a valuable method for biomarker discovery and systems biology research (2,3). The RPPA user base is also increasing, as demonstrated by recent publications in a number of fields (1,3–11). Thus, RPPA has realized its ground-breaking potential for clinical applications as well as for the elucidation of signaling

networks. Hence, RPPA technology would benefit from a robust, freely available software tool allowing data comparison across different RPPA platforms.

To date, the available analysis tools for RPPA have been mainly limited to non standardized in-house solutions or commercial platforms such as Microvigene (Vigene Tech, Carlisle, MA). A major drawback of open-source RPPA tools such *NormaCurve* (12), *Reno* (13), and others (14) is their limitation in data preprocessing because they are restricted to data normalization steps without offering additional functions to plot the resulting data or to carry out a statistical analysis. In contrast, the

R (<http://www.R-project.org>) package *RPPanalyzer* (15) offers data preprocessing in combination with graphical and statistical analysis options. Utilization of *RPPanalyzer* is on the rise and has been documented already (4,5,16–20).

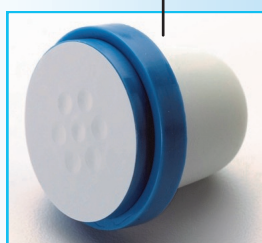
Here we describe an expanded and improved toolbox for *RPPanalyzer* providing additional useful features including a standardized workflow of upgraded data preprocessing combined with further improved graphical and statistical analyses. A novel background correction method has now been incorporated into the program. The main advantage of our background correction method is that

METHOD SUMMARY

We describe an improved version of *RPPanalyzer* that introduces a standardized workflow for RPPA data preprocessing, a novel functionality for background noise correction, and optional graphical and statistical analysis methods for robust analysis of reverse phase protein array (RPPA) data sets.

TAKE ACTION AGAINST CONTAMINATION WITH QMI® SAFE-SEPTUM

Sample, Inoculate Or Add Nutrients
To Your Bioreactor Aseptically !



The QMI Safe-Septum is:

- Aseptic
- Pressure & Temperature Safe
- Pre-Sterilized
- Easy To Retrofit
- Validated



651-501-2337

Email: info@qmisystems.com

www.qmisystems.com

it does not require the printing of all samples as serial dilutions; instead, it estimates correction factors based on a single representative dilution series. A new time course plotting function is also incorporated that embeds enhanced variance estimations, which is useful for triplicate measurements as usually obtained from cell-based perturbation experiments.

Materials and methods

The application of the *RPPanalyzer* (workflow depicted in Figure 1) is demonstrated using a typical systems biology RPPA data set (Figure 2A and 2B). The experimental setup was as follows: A549 cells, a human alveolar adenocarcinoma cell line, were serum-starved for 24 hours and subsequently stimulated with 6 different hepatocyte growth factor (HGF) concentrations ranging from 0 to 100 ng/mL. Samples were obtained at 6 different time points between 0 and 120 min. All experiments were done as biological triplicates, and the samples were analyzed by RPPA using antibodies directed against proteins and phosphoproteins involved in MET receptor signaling. This example RPPA data set and the corresponding R script are provided as Supplementary Material. The R package is available at the Comprehensive R Archive Network (CRAN, <http://cran.r-project.org/web/packages/RPPanalyzer>) and at R-Forge (<https://r-forge.r-project.org/projects/rppanalyzer>).

Results and discussion

The general *RPPanalyzer* workflow is depicted in Figure 1. For this updated version, the initial data preprocessing steps were bundled in the `dataPreproc` wrapper function, leading to standardized data output. Function tasks for `dataPreproc` include:

- *Import of raw data from slide, sample description text files, and corresponding gpr files.* In the text files, users provide sample-specific information and list the target proteins analyzed by RPPA. Instructions on the setup of slide and sample description files are provided in the *RPPana-*

lyzer vignette. The information in these files impacts all ensuing data processing steps. Hence, in case of any detected mistakes, files should be manually corrected and the workflow rerun. Spot-specific signal intensities obtained by image analysis using the commercial software GenePixPro are stored as gpr files that also contain information on array layout. *RPPanalyzer* is tailored to this format. In cases in which a user does not work with gpr files, a workaround is to save result files from other scanning systems as text files and then adjust the format to the syntax of gpr files. After adding the suffix `.gpr` to the text file generated this way, the data analysis can proceed as when processing gpr files.

- *Background noise correction using the new `correctDilinterc` function, as described next.*
- *Generation of plots to assess data quality.* These include: (i) *QC_dilutioncurve_raw.pdf* (raw signal intensities of serially diluted control samples to check linearity of target protein detection), (ii) *QC_target-VSblank_normed.pdf* (scatter plot of background noise corrected and normalized target signals versus “secondary antibody only” signal to assess bias caused by the secondary antibody), and (iii) *QC_qqPlot_normed.pdf* (quantile-quantile plot to control whether data are normally distributed).
- *Spot-specific normalization of signal intensities based on the total protein concentration determined by the FCF method (21).* Replicate slides, usually 1 slide out of 8 to 10, are stained with the dye Fast Green FCF to determine the total protein concentration of each individual lysate spot (Figure 2B). The signal intensity of FCF readout (f_i) is determined for each individual spot. The target protein specific signal intensities (x_i) obtained from slides probed with different detection antibodies are corrected for technical variance by dividing antibody signals by the factors f_i . Afterward, the corrected spot intensities are multiplied by the median f_m of the corresponding

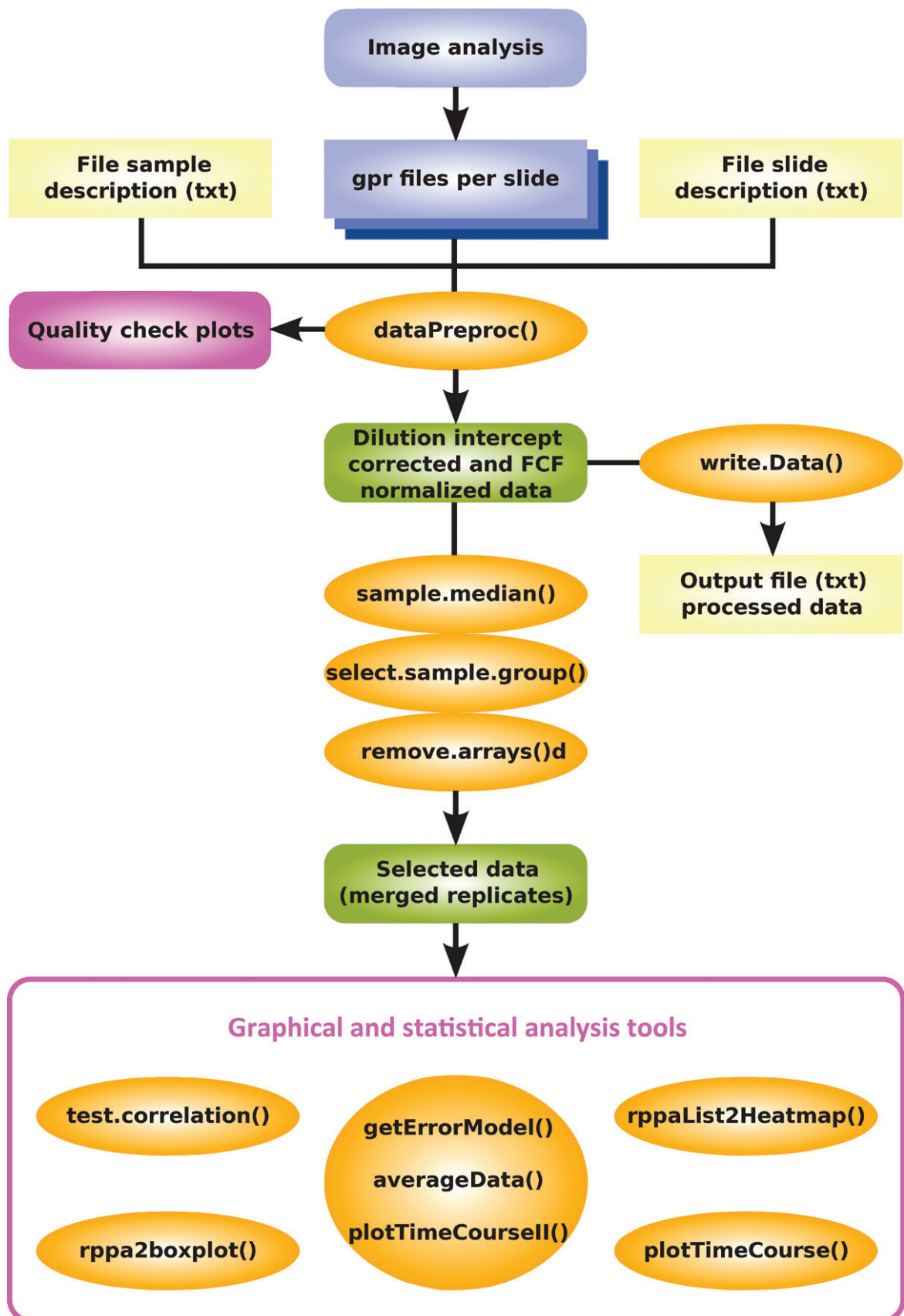


Figure 1. RPPanalyzer Toolbox workflow. The recommended workflow for reverse phase protein array (RPPA) data analysis starts with the import of the image analysis output files as well as the slide and sample description files. The data preprocessing returns dilution intercept corrected data, which were normalized by total protein concentration using the Fast Green FCF approach (21). Data preprocessed this way can then be analyzed using diverse graphical and statistical methods implemented in the RPPanalyzer toolbox.

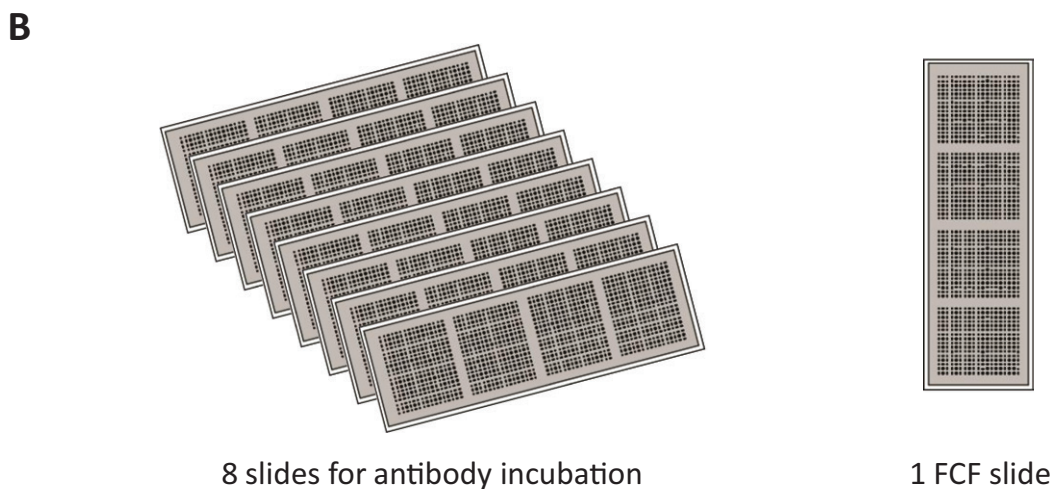
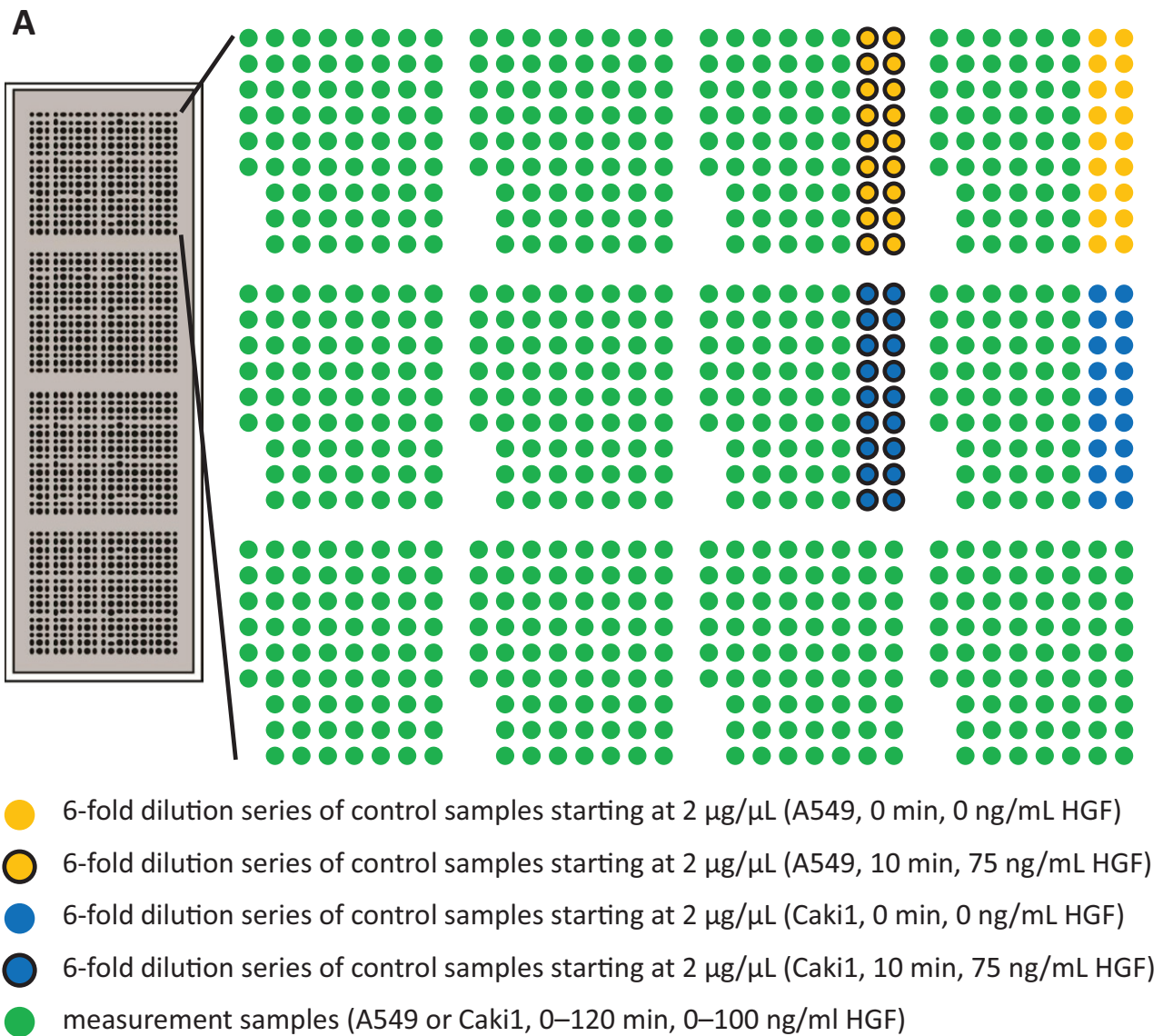


Figure 2. Exemplary reverse phase protein array (RPPA) layout. (A) Samples (control and measurement) are spotted as four identical subarrays on a series of nitrocellulose coated glass slides. All samples are spotted as technical triplicates. Only control samples are spotted as 6-fold dilution series. (B) Slide numbers 1 to 8 are used for incubation with target-specific antibodies; 4 subarrays allow for using 4 different antibodies per slide. Slide number 9 is used for total protein staining using the Fast Green FCF method (21).

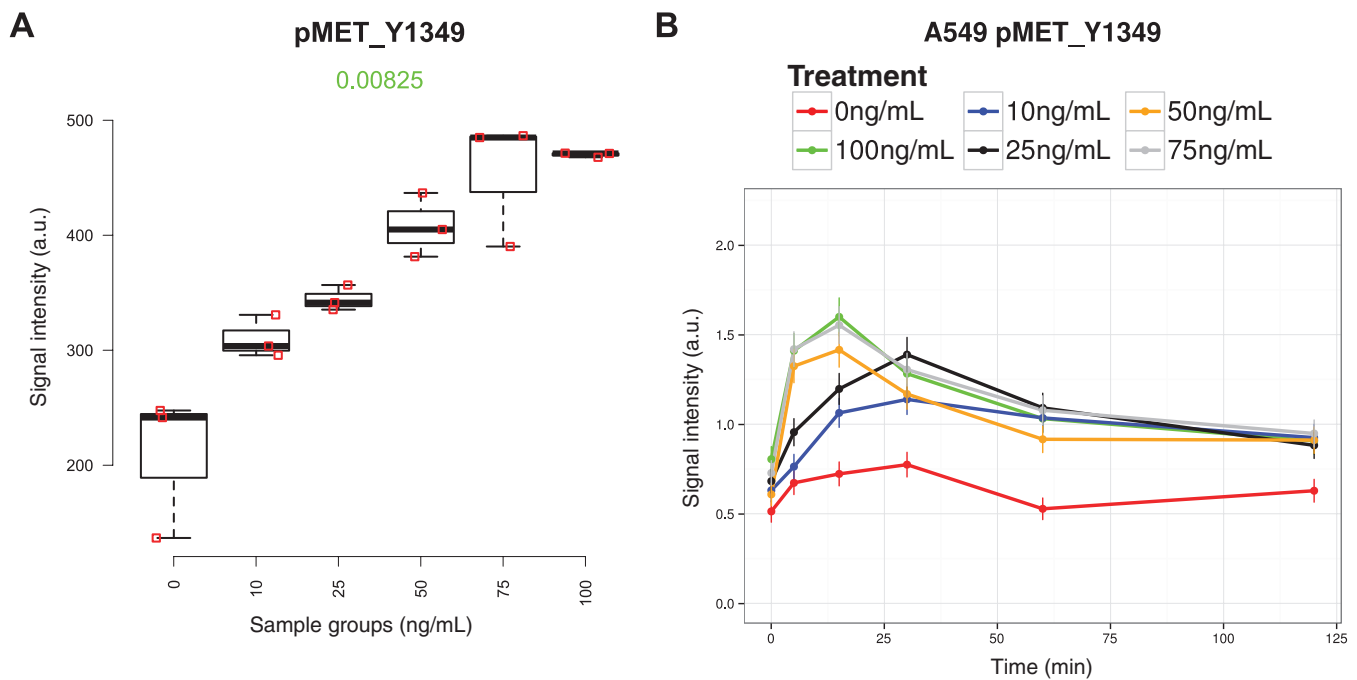


Figure 3. Examples of graphical output. (A) Boxplot of MET receptor phosphorylation (Y1349) signal intensity at the 15 min time point, revealing dependence on hepatocyte growth factor (HGF) concentration. The green number represents the *P* value result from a Kruskal-Wallis rank sum test, indicating significant difference between expression values of the six sample groups. (B) Time course plot of MET receptor phosphorylation (Y1349) over 120 min after HGF stimulation at different concentration levels.

normalizer subarray to scale data back to the original signal range. This can be summarized as normalized spot intensity $x_{FCF} = x_i / f_i * f_m$.

- *Returning a list of four items to users.* The first item lists raw data, that is, matrices of foreground and background signal intensities as well as information describing sample localization and providing keywords to identify individual samples as part of the downstream analysis. The second item is analogously built up but lists corrected foreground signal intensities generated using

the `correctDilinterc` function. If negative values are obtained after background subtraction applying the `correctDilinterc` function, the absolute minimum plus one is added, as for most applications (e.g., comparison of multiple conditions), a small positive value is required even for probes which are within the background levels. The third item is also structured as the first two but holds dilution intercept corrected and FCF normalized foreground data. The final item defines the directory for output storage.

- *Raw data export to a text file in table format.* Any other processed data (e.g., normalized data) can be exported manually using the `write.Data` function.

All output files are stored in a folder labeled with the date of analysis at the input files location.

After preprocessing, the next step is to merge technical replicates (`sample.median` function), select data subsets of interest (`select.sample.group` function), and then remove arrays or target proteins that are not required during downstream analysis, or data that did not pass the quality check (`remove.arrays` function).

Finally, different R functions can be applied for statistical analysis and graphical representation of the preprocessed RPPA data. According to the toolbox character, this output layer is not limited to the five implemented methods but can be easily expanded and tailored to a user's specific needs by new or modified R functions. The methods currently available in the *RPPAnalyzer* package are:

- `test.correlation`: A correlation plot to compare protein expression

Plasmid & Minicircle DNA Service

- Customized plasmids and *minicircles*
- *InStock* reporter genes
- AAV helper & packaging plasmids

PlasmidFactory.com

PlasmidFactory

The Minicircle Company

PlasmidFactory GmbH & Co. KG | Meisenstraße 96 | D-33607 Bielefeld | Germany | Fon ++49 521 2997350

and a numeric sample attribute, including statistical testing according to Pearson, Spearman or Kendall.

- `rppa2boxplot`: Allows for visualization of the data as boxplots for defined sample groups. In addition, statistical rank sum tests are applied. In case of a reference group to be tested against, two-sample Wilcoxon tests are performed, and the P values are adjusted for multiple testing according to Benjamini and Hochberg. To test for a general difference among all samples, a Kruskal-Wallis rank sum test is applied. The P values are displayed in the boxplot graph, as shown in the example data in Figure 3A.
- `rppaList2Heatmap`: Allows for the visualization of hierarchical clustering as a heatmap, adding specific column side colors to mark groups of selected phenodata. Before applying this function, the data should be logarithmized using the `logList` function. Scaling and viewing the data on a logarithmic scale have the effect of variance stabilization and are necessary for statistical tests and for the computation of distance measures that have the requirement that the data should look normally distributed.
- `plotTimeCourse`: Allows for the visualization of time course data. Different plotting options can be specified, such as smoothed spline fits through the data.

In addition to the `dataPreproc` function, four other functions have been added to this latest version of *RPPAnalyzer*.

- `correctDilinterc`: *Correction of background noise.* This function corrects signal intensities obtained for a protein of interest by subtracting an intercept estimated for a total protein concentration of 0 $\mu\text{g}/\mu\text{L}$. This function does require that the array contains a dilution series that passed the quality check mentioned before. The function is integrated in the data preprocessing function `dataPreproc`. The y-intercepts of the dilution

A549 PRAS40 2691

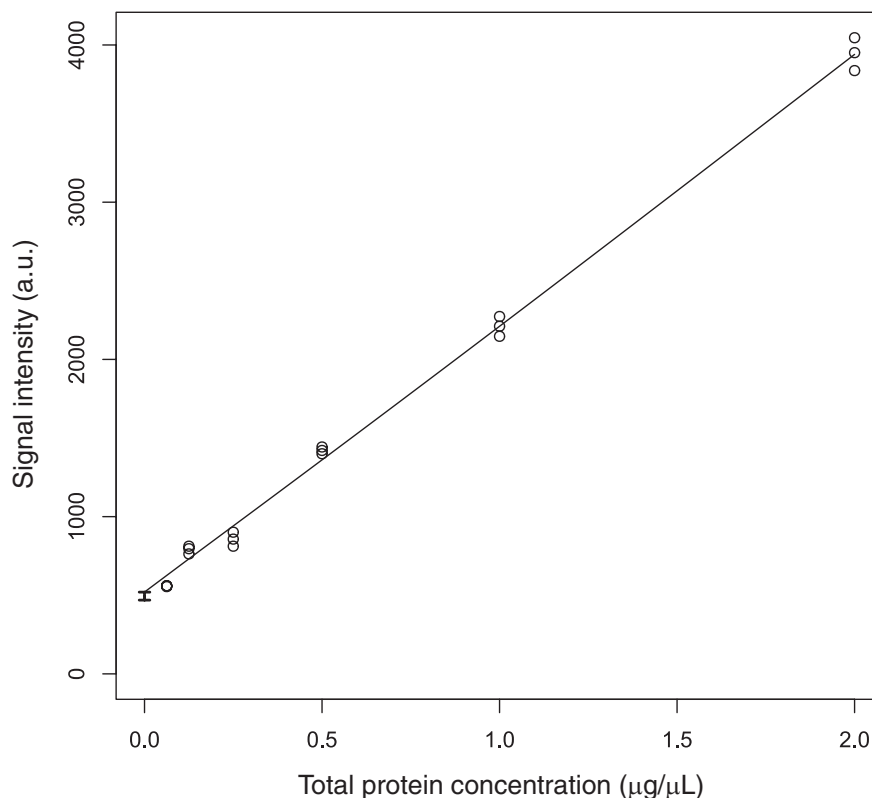


Figure 4. Dilution series plot visualizing the signal intercept estimation. In this example, a y-intercept of 453 a.u. was estimated as background noise for an antibody directed against PRAS40. Sample signal intensities will be background corrected by subtraction of this value.

series for corresponding samples are derived using `dilSeriesID`, a parameter defined in the sample description file, subarray-specific information, detection antibody and spotting run as summarized in the slide description file. Actual samples are defined as *measurement* in the *sampledescription.txt*, and serially diluted samples are specified as *control*.

A major advantage of this method is that it is independent of dilution series prepared from all samples as described by the original RPPA approach (1). Correction factors are obtained from representative samples chosen ahead of array printing during design. In our example data set, actual samples represent 6 different time points ranging from 0 to 120 min, whereas the corresponding dilution represents a 10 min time point only, as highlighted in Figure 2A. To link a sample

of interest with a certain control dilution series from a panel that mostly includes several dilution series (e.g., representing different cell lines or conditions), the corresponding identifier has to be entered in the column `dilSeriesID` of the *sampledescription.txt*. The selection of the sample to be used as dilution series strongly depends on the particular experimental setup.

For near-infrared (NIR) detection, signal and protein concentration are detectable over a linear range for the majority of detection antibodies. Hence, intercept subtraction will produce a parallel shift. For non-NIR fluorescence-based detection approaches, signals frequently saturate at higher protein concentrations. In this case, smoothing splines as nonparametric fits allow us to deal with several different RPPA detection methods and are

applied for extrapolation to zero concentration. The uncertainty of the extrapolated intercept is estimated by nonparametric bootstrapping.

The estimated intercepts I are conditional on the factors antibody, slide, and sample, of which the latter is defined by *dilSeriesID*. Linear models are established for the following hypotheses:

$$I \sim \text{const.} \quad [\text{Eq. 1.1}]$$

$$I \sim \text{const.} + \text{antibody} \quad [\text{Eq. 1.2}]$$

$$I \sim \text{const.} + \text{antibody} + \text{slide} \quad [\text{Eq. 1.3}]$$

$$I \sim \text{const.} + \text{antibody} + \text{slide} + \text{sample} \quad [\text{Eq. 1.4}]$$

The estimated uncertainties of the intercepts are used as weights. The provided bar plot of the residual sum of squares (RSS) (*anovaIntercepts_Output.pdf*) should be used to choose the model with the smallest RSS favoring less complexity. For example, if the bars of model 1.3 and 1.4 are the smallest and are equally high, model 1.3 should be preferred because the sample in 1.4 does not provide additional information. Based on the chosen model, the *measurement* intercepts are predicted and are then subtracted from the raw intensities. The function additionally generates plots of the dilution series and related intercept estimations (*getIntercepts_Output.pdf*), as shown in Figure 4.

- *getErrorModel: Estimation of signal variance.* Signal variance can be estimated from technical replicates. The variance estimator for triplicates is generally very poor because it is χ^2 distributed with $n-1$ degrees of freedom. In the case of $n = 3$, this is an exponential distribution. Therefore, information from several different triplicates is combined to estimate parameters of an error model.

A variance versus signal plot characteristic for RPPA data reveals the following:

- The signal dependency of the variance depends on the factors subarray, detection antibody, and median FCF normalizer value. Slides probed with different detection antibodies are scanned with different scanner settings to yield an optimal image for data analysis. Signals are scaled by the median FCF value.
- Signal variances can increase with the signal strength, are constant, or have (almost) zero variance at zero signal.

The error model used to account for these observations is $\sigma^2 = \sigma_0^2 + S^2\sigma_{rel}^2$, where S is the signal strength, σ_0^2 denotes the variance at zero signal, and σ_{rel}^2 reflects the relative error. To estimate σ_0^2 and σ_{rel}^2 from the variance versus signal plot, a maximum likelihood approach, exploiting the χ^2 distribution of the variances, is applied. This is due to the strong nonnormality of the triplicate variance estimator.

Safe & Secure Weighing

Secura®

Reduce Your risk with integrated and intelligent protection systems.



www.sartorius.com/secura

Unlimited BioImaging

Colorimetric
Bioluminescence
Fluorescence
Chemiluminescence
Multiplex



The **BioSpectrum Imaging System** with its advanced capabilities enables a wide range of imaging applications.

Unlimited in its ability to grow with your lab's developing life science research needs *today and into the future.*

UVP

An Analytik Jena Company

UVP, LLC: USA (909) 946-3197
UK +44 (0)1223 420022

Learn more at
uvp.com/biospectrum.html

- `averageData`: *Averaging biological replicates.*

Frequently, experiments provide two kinds of replicates: technical and biological. In this context, *biological replicate* means that spotted lysates result from equal treatment, and *technical replicate* means that the identical lysate is spotted several times on the same slide.

Averaging is done for each detection antibody separately. It is based on the assumption that a true dynamic behavior $y_s(t)$ for a given stimulation s exists. The observed dynamics $S_{j,s}(t)$ for each biological replicate j , that is, factor (biological replicate + slide), and stimulation s differ systematically from $y_s(t)$ by a scaling factor s_j . In addition, $s_j * S_{j,s}(t)$ differs statistically from $y_s(t)$ because of biological variability. To estimate the true dynamics and the scaling factors, the objective function (see Equation 2 below) for discrete measurement time points t_i is minimized. The response is weighted by the variance from the error model evaluated at the "true" value y_{is}/S_j . Strictly speaking, this is an approximation because each biological replicate has its own true dynamic behavior. A direct estimate of these signaling dynamics would be the mean of the technical replicates for each biological replicate, that is, $S_{i,j,s}$ itself. Depending on the value of σ^2_r , erroneous estimates can have a huge impact because smaller values are automatically favored. In contrast, the approximation by y_{is}/S_j is more robust.

- `plotTimeCourseII`: *Visualization of time course data.*

This function is applied after transforming the preprocessed data by the `getErrorModel` and `averageData` functions, and an example of this function is shown in Figure 3B.

In conclusion, we have simplified, enhanced and standardized RPPA data analysis by extending the functionality of the existing *RPPAnalyzer* package through the introduction of new approaches for variance estimation, background noise correction, and time course data visualization. In addition to this new functionality, the package was also streamlined by easing and improving several already existing functions. Although we recommend the use of the standard preprocessing steps, the newly introduced toolbox approach described here does allow integration of add-on functions that can be tailored to the specific needs of users.

Author Contributions

J.B. developed the new time course plotting function under the supervision of U.K.; D.K. developed the new functions for background correction as well as variance estimation under the supervision of J.T.; J.S. provided the RPPA example data and was involved in the conception of the software update under the supervision of U.K.; S.v.d.H. implemented the corresponding upgrades into the R package under the supervision of T.B.; C.B. is the current maintainer of the R package at CRAN; A.W. was involved in discussions and is currently one of the package maintainers at R-Forge. All authors contributed to conception, writing, and editing of the manuscript.

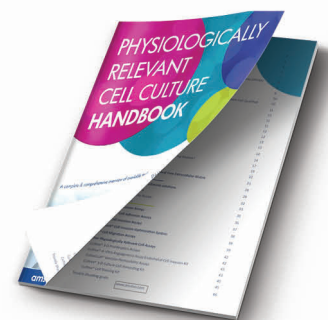
Acknowledgments

We thank Heiko Mannsperger for his initiative to develop the first version of the R package *RPPAnalyzer*. In addition, we thank all users of the *RPPAnalyzer* for valuable input to continuously improve its functions. We are grateful to Corinna Becki and Sabrina Schumacher for excellent technical assis-

$$x^2(\{s_j\}, \{y_{is}\}) = \sum_{i,j,s} \frac{(s_j S_{ijs} - y_{is})^2}{s_j^2 \left[\sigma_{ijs,0}^2 + \left(\frac{y_{is}}{s_j} \right)^2 \sigma_{ijs,r}^2 \right]} = \sum_{i,j,s} \frac{\left(S_{ijs} - \frac{y_{is}}{s_j} \right)^2}{\sigma_{ijs,0}^2 + \left(\frac{y_{is}}{s_j} \right)^2 \sigma_{ijs,r}^2} \quad [\text{Eq. 2}]$$

PHYSIOLOGICALLY RELEVANT CELL CULTURE HANDBOOK

- Popular 3D techniques
- Solutions to common problems
- How to obtain optimal 3D cell culture matrix for different applications



FREE ONLINE
DOWNLOAD

www.amsbio.com/PRCC.aspx

tance. This work was supported by the Medical Systems Biology program (grant 0315396 of the BreastSys project, and grant 0316173 of the eBio MetastaSys project) of the German Federal Ministry of Education and Research.

Competing interests

The authors declare no competing interests.

References

- Paweletz, C.P., L. Charboneau, V.E. Bichsel, N.L. Simone, T. Chen, J.W. Gillespie, M.R. Emmert-Buck, M.J. Roth, et al. 2001. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* 20:1981-1989.
- Kornblau, S.M., N. Singh, Y. Qiu, W. Chen, N. Zhang, and K.R. Coombes. 2010. Highly phosphorylated FOXO3A is an adverse prognostic factor in acute myeloid leukemia. *Clin Cancer Res.* 16:1865-1874.
- Uhlmann, S., H. Mannsperger, J.D. Zhang, E.-Á. Horvat, C. Schmidt, M. Küblbeck, F. Henjes, A. Ward, et al. 2012. Global microRNA level regulation of EGFR-driven cell-cycle protein network in breast cancer. *Mol. Syst. Biol.* 8:570.
- Sonntag, J., C. Bender, S. Soons, S. von der Heyde, R. König, S. Wiemann, H.-P. Sinn, A. Schneeweiss, et al. 2014. Reverse phase protein array based tumor profiling identifies a biomarker signature for risk classification of hormone receptor-positive breast cancer. *Transl. Proteomics* 2:52-59.
- Ummanni, R., H.A. Mannsperger, J. Sonntag, M. Oswald, A.K. Sharma, R. König, and U. Korf. 2013. Evaluation of reverse phase protein array (RPPA)-based pathway-activation profiling in 84 non-small cell lung cancer (NSCLC) cell lines as platform for cancer proteomics and biomarker discovery. *Biochim Biophys Acta.* 1844:950-959.
- Li, J., Y. Lu, R. Akbani, Z. Ju, P.L. Roebuck, W. Liu, J.-Y. Yang, B.M. Broom, et al. 2013. TCPA: a resource for cancer functional proteomics data. *Nat. Methods* 10:1046-1047.
- Chavez-Macgregor, M., S. Liu, D. De Melo-Gagliato, H. Chen, K.-A. Do, L. Pusztai, W. Fraser Symmans, L. Nair, et al. 2014. Differences in gene and protein expression and the effects of race/ethnicity on breast cancer subtypes. *Cancer Epidemiol. Biomarkers Prev.* 23:316-323.
- Hennessy, B.T., Y. Lu, A.M. Gonzalez-Angulo, M.S. Carey, S. Myhre, Z. Ju, M.A. Davies, W. Liu, et al. 2010. A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers. *Clin. Proteomics* 6:129-151.
- Federici, G., X. Gao, J. Slawek, T. Arodz, A. Shitaye, J.D. Wulfschlegel, R. De Maria, L.A. Liotta, and E.F. Petricoin 3rd. 2013. Systems analysis of the NCI-60 cancer cell lines by alignment of protein pathway activation modules with "-OMIC" data fields and therapeutic response signatures. *Mol. Cancer Res.* 11:676-685.
- Gonzalez-Angulo, A.M., B.T. Hennessy, F. Meric-Bernstam, A. Sahin, W. Liu, Z. Ju, M.S. Carey, S. Myhre, et al. 2011. Functional proteomics can define prognosis and predict pathologic complete response in patients with breast cancer. *Clin. Proteomics* 8:11.
- Shankavaram, U.T., W.C. Reinhold, S. Nishizuka, S. Major, D. Morita, K.K. Chary, M.A. Reimers, U. Scherf, et al. 2007. Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integrative microarray study. *Mol. Cancer Ther.* 6:820-832.
- Troncale, S., A. Barbet, L. Coulibaly, E. Henry, B. He, E. Barillot, T. Dubois, P. Hupé, and L. de Koning. 2012. NormaCurve: a SuperCurve-based method that simultaneously quantifies and normalizes reverse phase protein array data. *PLoS ONE* 7:e38686.
- Li, B., F. Liang, J. Hu, and X. He. 2012. Reno: regularized non-parametric analysis of protein lysate array data. *Bioinformatics* 28:1223-1229.
- Chiechi, A., C. Mueller, K.M. Boehm, A. Romani, M.S. Benassi, P. Picci, L.A. Liotta, and V. Espina. 2012. Improved data normalization methods for reverse phase protein microarray analysis of complex biological samples. *Biotechniques* 0:1-7.
- Mannsperger, H.A., S. Gade, F. Henjes, T. Beissbarth, and U. Korf. 2010. RPPAnalyzer: analysis of reverse-phase protein array data. *Bioinformatics* 26:2202-2203.
- Aleskandarany, M.A., O.H. Negm, E.A. Rakha, M.A.H. Ahmed, C.C. Nolan, G.R. Ball, C. Caldas, A.R. Green, et al. 2012. TOMM34 expression in early invasive breast cancer: a biomarker associated with poor outcome. *Breast Cancer Res. Treat.* 136:419-427.
- Alshareeda, A.T., O.H. Negm, N. Albarakati, A.R. Green, C. Nolan, R. Sultana, S. Madhusudan, A. Benhasouna, et al. 2013. Clinicopathological significance of KU70/KU80, a key DNA damage repair protein in breast cancer. *Breast Cancer Res. Treat.* 139:301-310.
- Jozefczuk, J., K. Kashofer, R. Ummanni, F. Henjes, S. Rehman, S. Geenen, W. Wruck, C. Regenbrecht, et al. 2012. A systems biology approach to deciphering the etiology of steatosis employing patient-derived dermal fibroblasts and iPS cells. *Front Physiol.* 3:339.
- Mannsperger, H.A., S. Uhlmann, C. Schmidt, S. Wiemann, O. Sahin, and U. Korf. 2010. RNAi-based validation of antibodies for reverse phase protein arrays. *Proteome Sci.* 8:69.
- Nelson, L.D., C. Bender, H. Mannsperger, D. Buegy, P. Kambakamba, G. Mudduluru, U. Korf, D. Hughes, et al. 2012. Triplex DNA-binding proteins are associated with clinical outcomes revealed by proteomic measurements in patients with colorectal cancer. *Mol. Cancer* 11:38.
- Loebke, C., H. Suelmann, C. Schmidt, F. Henjes, S. Wiemann, A. Poustka, and U. Korf. 2007. Infrared-based protein detection arrays for quantitative proteomics. *Proteomics* 7:558-564.

Received 21 February 2014; accepted 20 July 2014.

Address correspondence to Tim Beißbarth, Statistical Bioinformatics, Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany. Tim.Beissbarth@med.uni-goettingen.de

To purchase reprints of this article, contact: biotechniques@fosterprinting.com

Appendix C

Publication “mRNA Profiling Reveals Determinants of Trastuzumab Efficiency in HER2-Positive Breast Cancer”

RESEARCH ARTICLE

mRNA Profiling Reveals Determinants of Trastuzumab Efficiency in HER2-Positive Breast Cancer

Silvia von der Heyde^{1,2}, Steve Wagner^{3,4}, Alexander Czerny¹, Manuel Nietert¹, Fabian Ludewig⁵, Gabriela Salinas-Riester⁵, Dorit Arlt^{3,4}, Tim Beißbarth^{1*}

1 Statistical Bioinformatics, Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany, **2** IndiviuTest GmbH, Hamburg, Germany, **3** Division of Stem Cells and Cancer, German Cancer Research Center, Heidelberg, Germany, **4** Division of Molecular Genome Analysis, German Cancer Research Center, Heidelberg, Germany, **5** DNA Microarray and Deep-Sequencing Facility Göttingen, Department of Developmental Biochemistry, University of Göttingen, Göttingen, Germany

* tim.beissbarth@ams.med.uni-goettingen.de



 OPEN ACCESS

Citation: von der Heyde S, Wagner S, Czerny A, Nietert M, Ludewig F, Salinas-Riester G, et al. (2015) mRNA Profiling Reveals Determinants of Trastuzumab Efficiency in HER2-Positive Breast Cancer. PLoS ONE 10(2): e0117818. doi:10.1371/journal.pone.0117818

Academic Editor: Ming Tan, University of South Alabama, UNITED STATES

Received: July 29, 2014

Accepted: December 30, 2014

Published: February 24, 2015

Copyright: © 2015 von der Heyde et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The RNA-Seq data are available from the Gene Expression Omnibus (GEO) database (accession number GSE55005). All other relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the German Federal Ministry of Education and Research in the platform MedSys via the project BreastSys (grant 0315396A, URL <http://www.ams.med.uni-goettingen.de/breastsys/>), in the platform e:Bio via the project MetastaSys (grant 0316173A, URL <http://www.med.uni-goettingen.de/metastasy/>) and in the

Abstract

Intrinsic and acquired resistance to the monoclonal antibody drug trastuzumab is a major problem in the treatment of HER2-positive breast cancer. A deeper understanding of the underlying mechanisms could help to develop new agents. Our intention was to detect genes and single nucleotide polymorphisms (SNPs) affecting trastuzumab efficiency in cell culture. Three HER2-positive breast cancer cell lines with different resistance phenotypes were analyzed. We chose BT474 as model of trastuzumab sensitivity, HCC1954 as model of intrinsic resistance, and BTR50, derived from BT474, as model of acquired resistance. Based on RNA-Seq data, we performed differential expression analyses on these cell lines with and without trastuzumab treatment. Differentially expressed genes between the resistant cell lines and BT474 are expected to contribute to resistance. Differentially expressed genes between untreated and trastuzumab treated BT474 are expected to contribute to drug efficacy. To exclude false positives from the candidate gene set, we removed genes that were also differentially expressed between untreated and trastuzumab treated BTR50. We further searched for SNPs in the untreated cell lines which could contribute to trastuzumab resistance. The analysis resulted in 54 differentially expressed candidate genes that might be connected to trastuzumab efficiency. 90% of 40 selected candidates were validated by RT-qPCR. ALPP, CALCOCO1, CAV1, CYP1A2 and IGFBP3 were significantly higher expressed in the trastuzumab treated than in the untreated BT474 cell line. GDF15, IL8, LCN2, PTGS2 and 20 other genes were significantly higher expressed in HCC1954 than in BT474, while NCAM2, COLEC12, AFF3, TFF3, NRCAM, GREB1 and TFF1 were significantly lower expressed. Additionally, we inferred SNPs in HCC1954 for CAV1, PTGS2, IL8 and IGFBP3. The latter also had a variation in BTR50. 20% of the validated subset have already been mentioned in literature. For half of them we called and analyzed SNPs. These results contribute to a better understanding of trastuzumab action and resistance mechanisms.

platform e:Med via the project Her2Low (grant 031A429C, URL <http://www.ams.med.uni-goettingen.de/her2low/>). TB received the mentioned funding. Above that SvdH, MN, SW and DA were funded by BreastSys. AC was funded by MetastaSys. The authors also acknowledge support by the Open Access Publication Funds of the Göttingen University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

The ‘HER2-positive’ subtype of breast cancer overexpresses the human epidermal growth factor receptor 2 (HER2). This receptor tyrosine kinase is part of the epidermal growth factor receptor (EGFR) family, further including HER1 (EGFR), HER3 and HER4 [1]. It is overexpressed in 10–20% of breast tumors, and the related subtype is associated with increased recurrence and mortality rates [2, 3].

The humanized monoclonal antibody trastuzumab targets specifically the extracellular domain of HER2 and is part of the adjuvant treatment of patients with HER2-positive (HER2+) early breast cancer [4]. The improved outcome by adding trastuzumab to chemotherapy for example is not completely understood. So far it has been associated with different mechanisms of action apart from inhibiting HER2, its dimerization and cleavage [4, 5]. These mechanisms include inhibition of downstream signal transduction pathways like the PI3K pathway, antigen-dependent cellular cytotoxicity (ADCC), induction of cell cycle arrest and apoptosis or inhibition of tumor angiogenesis.

Although trastuzumab provides clinical benefit to women with HER2+ breast cancer, not all patients respond [6]. Primary or acquired resistance limits the success of trastuzumab. Diverse possible mechanisms have been discussed [4, 5]. Among others, these include increased HER2, HER1 or HER3 expression, steric hindrance of HER2-antibody interaction, constitutive activation of the PI3K pathway due to mutations in the PIK3CA gene or loss of PTEN, alternative cell signaling induced by EGFR family members, MET receptor or insulin-like growth factor 1 receptor (IGF-IR), and overexpression of transforming growth factor (TGF)- α , neuregulin or vascular endothelial growth factor (VEGF). Corresponding therapeutic strategies to overcome or avoid resistance to trastuzumab have been developed, and several new agents are in clinical development. Studies in metastatic disease led to the approval of new HER2-targeted therapies using small molecule tyrosine kinase inhibitors such as lapatinib and HER2/HER3 antibodies such as pertuzumab [7]. However, so far it is not possible to predict prior to trastuzumab treatment which patients will develop resistance. A need for a better understanding of the mechanisms of trastuzumab action and resistance persists.

This study aims at detecting genes and single nucleotide polymorphisms (SNPs) affecting trastuzumab efficiency in three cell lines with different resistance phenotypes. These include trastuzumab sensitivity, intrinsic resistance due to a mutated PIK3CA gene, and acquired resistance. Ten percent of the candidate genes inferred via mRNA profiling have already been supported by literature. The remaining ones, partially known to be involved in breast cancer, could also contribute to novel strategies preventing trastuzumab resistance.

SNPs are the most common genetic variations and can be associated with heritable phenotypes. Related data is deposited in public databases [8]. SNPs are defined as single base pair positions in genomic DNA at which different sequence alternatives (alleles) exist in normal (non-diseased) individuals in some population(s), wherein the least frequent allele has an abundance of at least 1% [9]. According to this definition, Brookes concluded that single base insertion/deletion variants (indels) would not formally be considered as SNPs. However, he stated that in practice the term SNP is used rather loosely. For example, single base variants in cDNAs (cSNPs) are usually called SNPs, since most of them reflect underlying genomic DNA variants. This could be misleading in case of disease predisposing single base variants, which occur in some non-diseased individuals. Also Brookes warned that the ‘some population’ component of the definition is limited by practical challenges of attaining representative global population samples. He summarized that the term ‘SNP’ is being widely used as a label for many different types of subtle sequence variation. Being aware of the definition mentioned before, we decided

to call our detected sequence variations, which are potentially related to trastuzumab efficacy, 'SNPs' in the following.

SNPs may unravel multifactorial diseases such as most cancers or drug response, bearing pharmacogenetic potential in the context of personalized medicine. Direct DNA sequencing is the favored high-throughput method for SNP identification [10, 11]. However, Quinn et al. evaluated the performance of different SNP calling methods by applying them to RNA-Seq data and comparing results with sequence variation data from *1000 Genomes* [12]. They regard RNA-Seq SNP data as a useful by-product of sequence-based transcriptome analysis. According to their results, one can detect a high proportion of mutations of expressed genes via RNA-Seq. We intended to reveal SNPs in our candidate genes which might determine trastuzumab efficiency in the untreated cell lines. Above that we analyzed SNPs that affect cell signaling in the MAPK and PI3K pathway. These variations could be responsible for the different resistance cell phenotypes. Hence, the combination of differential gene expression and SNP analysis could help to predict the efficacy of trastuzumab therapy. Consequently, the detection of corresponding genes and their variations could contribute to an improved patient stratification.

Materials and Methods

Cell lines

Three human HER2-amplified breast cancer cell lines (BT474, HCC1954 and BTR50) were chosen as model systems of trastuzumab sensitivity (BT474 [6]), intrinsic (HCC1954 [6, 13]) and acquired (BTR50) resistance. While HCC1954 is known to be trastuzumab resistant due to a hotspot PIK3CA mutation (H1047R, PI3K gain-of-function), BT474 exhibits PIK3CA wild type behavior [14]. The cell line BTR50 is a trastuzumab-conditioned version of BT474, as explained in the corresponding subsection.

The cells were grown in a monolayer and collected as a cell pellet after trypsin treatment. RNA was harvested from cell pellet using the miRNeasy kit (Qiagen).

BT474

The human breast cancer cell line BT474 was directly obtained from the American Type Culture Collection (ATCC), catalogue no. HTB-20. It was cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum, 0.01 mg/ml of insulin and 1% penicillin/streptomycin. The cells were cultured at 37°C in an atmosphere containing 5% CO₂. Cells were harvested with trypsin-ethylenediamine tetraacetic acid (EDTA) (0.5 g/L trypsin; 0.2 g/L EDTA; Sigma). The cells were split three times per week.

HCC1954

The human breast cancer cell line HCC1954 was directly obtained from ATCC, catalogue no. CRL-2338, and cultured in RPMI media (Gibco) supplemented with 10% fetal bovine serum (Gibco). The medium was supplemented with 1% penicillin/streptomycin (Gibco). The cells were cultured at 37°C in an atmosphere containing 5% CO₂. Cells were harvested with EDTA (0.5 g/L trypsin; 0.2 g/L EDTA; Sigma). The cells were split three times per week.

BTR50

Resistant cells (BTR) were developed by culturing the epithelial BT474 breast cancer cells (wild type, wt) in the presence of 50 µg trastuzumab (Roche) for around six months. Parental cells (wt) were cultured in parallel to resistant ones without the addition of trastuzumab (Fig. 1). Resistance of the cells to trastuzumab was verified by cell viability assays (Fig. 2). Trastuzumab

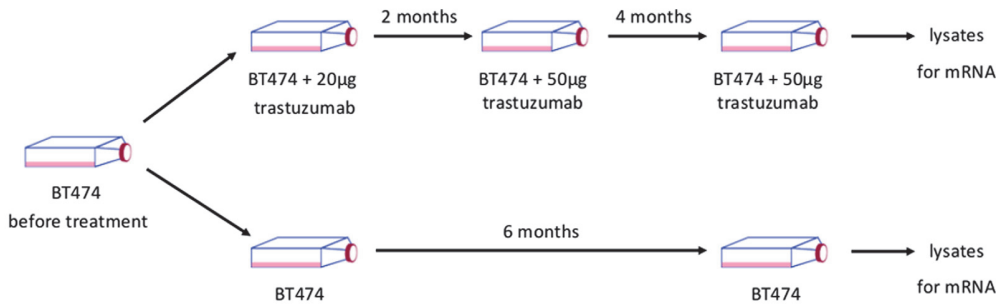


Fig 1. Development of BTR50 cells. Trastuzumab resistant cells were developed by culturing parental BT474 cells in the presence of 20/50 µg trastuzumab for around 6 months.

doi:10.1371/journal.pone.0117818.g001

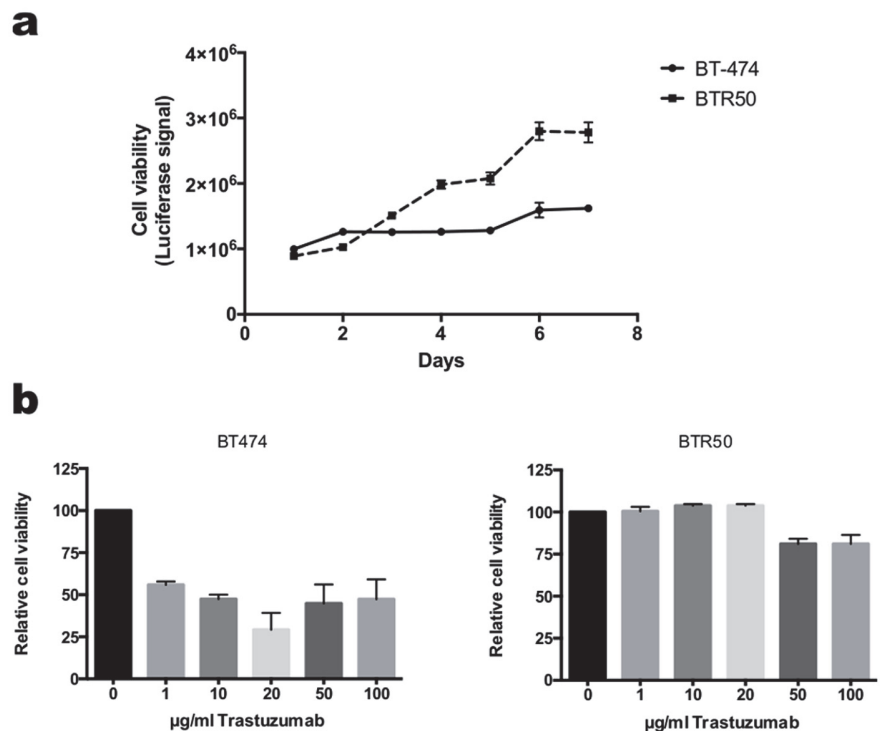


Fig 2. BT474 cells have acquired resistance to trastuzumab. (a) Proliferation rate of BT474 (parental) and trastuzumab resistant BT474 (BTR50) cells treated with 20 µg/ml trastuzumab. Proliferation rates were measured daily over 7 days by a luciferase-based viability assay. (b) Sensitivity of BT474 (parental) and trastuzumab resistant BTR50 cells towards increasing concentrations of trastuzumab. Cell viability was determined by a luciferase-based viability assay after 7 days of treatment.

doi:10.1371/journal.pone.0117818.g002

markedly reduced the growth of BT474 (wt) cells compared to trastuzumab resistant BT474 (BTR50) cells.

Cell viability assay and determination of trastuzumab sensitivity

For the measurement of cell viability, both parental BT474 and trastuzumab resistant BTR50 cells were seeded in 96 well plates. Cell viability was determined every day for in total 7 days using Cell Titer Glo Luminescent Cell Viability Assay (Promega) following the manufacturer’s

instructions (Fig. 2a). The sensitivity towards trastuzumab was defined by treatment of both cell lines with increasing concentrations of trastuzumab and a subsequent measurement of cell viability after 7 days (Fig. 2b).

Trastuzumab (Herceptin) treatment

For treatment experiments, 2×10^5 cells were seeded in T25 flasks and cultivated as described before. Cells were treated with 20 $\mu\text{g/ml}$ trastuzumab (Roche Diagnostics GmbH, Penzberg, Germany) or grown in full growth media without inhibitors. Cell pellets were harvested after 72 h. The cells were incubated in the standard media 24 h before addition of trastuzumab or fresh full growth media.

RNA sequencing

Total RNA was isolated from the cell lines BT474, HCC1954 and BTR50 using the Trizol (Invitrogen) method according to the manufacturer's recommendations. Afterwards, the samples were DNase I (Sigma) treated in order to remove DNA contamination. RNA quality was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) microfluidic electrophoresis. Only samples with comparable RNA integrity numbers were selected for deep sequencing.

Library preparation for RNA-Seq was performed using the TruSeq RNA Sample Preparation Kit (Illumina, catalog ID RS-122-2002) starting from 800 ng of total RNA. Accurate quantitation of cDNA libraries was performed by using the QuantiFluor[™] dsDNA System (Promega). The size range of final cDNA libraries was determined applying the DNA 1000 chip on the Bioanalyzer 2100 from Agilent (280 bp). cDNA libraries were amplified and sequenced by using the cBot and HiSeq 2000 from Illumina (SR, 1×51 bp, 6 GB ca. 30–35 million reads per sample).

Sequence images were transformed with Illumina software BaseCaller to bcl files, which were demultiplexed to fastq files with CASAVA (version 1.8.2). Quality check was done via FastQC (version 0.10.1, Babraham Bioinformatics).

The sequenced reads were hard (first five bases) and soft (last bases) trimmed as well as trimmed for adapter sequences via Flexbar [15] (version 2.32). Afterwards they were mapped to the human reference genome (GRCh37, Gencode [16] release 14) using STAR [17] (version 2.3.0), allowing maximal three mismatches. Conversion of SAM to BAM files and corresponding sorting was done via SAMtools [18] (version 0.1.18). Counting the reads to each gene to the UCSC gtf gene annotation file (March 2012) was done via HTSeq [19] (0.5.3p9, htseq-count).

SNP calling

The reads were aligned against the Ensembl [20] reference genome release 71 (GRCh37) with STAR (version 2.3.0), allowing for 5% mismatches of total read length. STAR was used with additional splice junction annotation. Read group definition as well as removal of duplicates after the alignment step was done via Picard command line tools (release 1.99, <http://picard.sourceforge.net>). For calling variants from RNA-Seq data, the Genome Analysis Toolkit (GATK) [21] (version 2.7.2) standard SNP best-practice protocol was used with the additional option -U ALLOW_N_CIGAR_READS. Standard hard filtering was applied. Read quality was reassigned from 255 to 60. Low quality reads were neglected. Just SNPs with a read depth ≥ 10 were selected.

For the analysis of the derived SNP candidates by applying GATK we used the Variant Effect Predictor (VEP) [22].

Differential expression analysis

Normalization of read counts to the library size, estimation of dispersions (method = 'blind', sharingMode = 'fit-only') and testing for differentially expressed (DE) genes based on a statistical test assuming negative binomial data distribution was computed via the DESeq [23] (version 1.12.1) R [24] package. Just genes exceeding 20 counts for at least one sample were kept for further analysis. The numerator and denominator of fold changes (FC) were increased by one to account for zero values. Significant genes were filtered to a minimum of 2xFC and $\text{fdr} < 0.05$ with multiple testing correction according to Benjamini and Hochberg [25].

Based on RNA-Seq data, we performed DE analyses on six samples, i.e. the breast cancer cell lines BT474, HCC1954 and BTR50 with and without trastuzumab treatment. In detail, five separate two-sample tests were performed and normalization was done per sample pair of consideration. First we tested for DE between resistant and wild type cells, i.e. HCC1954 and BTR50 vs. BT474, respectively. This revealed 46 significant genes which might contribute to resistance. Next we tested for DE between untreated and trastuzumab treated cells, i.e. each of the three cell lines vs. its trastuzumab treated version. The test for BT474 revealed 18 significant genes which might contribute to trastuzumab efficiency. To exclude false positives from the combined set of 64 genes, we removed ten genes that were also significant in the test for BTR50. No trastuzumab effect was expected for the resistant cell line. The same would have held for HCC1954, but the related test revealed no significant genes overlapping with our candidate set. This way we discovered 54 genes that might determine trastuzumab efficiency in HER2+ breast cancer cell lines. Annotation and functional association of the candidate genes to biological processes (BP, Gene Ontology annotation) was added via biomaRt [26] (version 2.16.0). The raw and normalized data have been submitted to the Gene Expression Omnibus (GEO) with accession number GSE55005.

Real-time quantitative PCR

Out of the 54 candidate genes detected by mRNA profiling we chose 40 for validation via real-time quantitative PCR (RT-qPCR). mRNA purification was performed at 4°C using the miR-Neasy Kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendation. cDNA was generated using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Darmstadt, Germany) from total RNA isolated from breast cancer cell lines. Primer combinations for the respective genes were designed according to the Harvard Primer Bank (<http://pga.mgh.harvard.edu/primerbank>) and are listed in S1 Table. 10 ng of each cDNA were used per approach. The final concentrations were 1x Applied Biosystems Power SYBR Green Master Mix (Applied Biosystems, Darmstadt, Germany) and 10 μM of each primer. RT-qPCR was performed in a total volume of 12 μl . Glyceraldehyde-3-phosphat-Dehydrogenase (GAPDH) was used as reference gene. The ViiA 7 Real-Time PCR System (Applied Biosystems, Darmstadt, Germany) was used for RT-qPCR analysis. Data analysis was performed as described in Livak and Schmittgen [27] with two biological replicates. The corresponding data are listed in S2 Table.

Results and Discussion

mRNA profiling reveals genes associated with trastuzumab efficiency

The differential expression (DE) revealed by RNA-Seq analysis was confirmed via RT-qPCR for 36 out of 40 candidate genes. The whole set consisted of 54 genes (S3 Table). The selection criterion was mainly based on literature support for association of genes with trastuzumab efficiency or breast cancer.

Six of the candidate genes were selected from the gene set inferred by testing DE between BT474 and its trastuzumab treated version. The whole subset consisted of eight genes. For the selected ones (ALPP, CALCOCO1, CAV1, CYP1A2, IGFBP3, L1CAM) a significantly higher expression after trastuzumab treatment of BT474 was validated apart from L1CAM.

Of the gene set inferred by testing DE between BT474 and HCC1954, 33 candidate genes were selected. The whole subset consisted of 45 genes. For all of the selected ones (AFF3, AKR1C1, CES1, CLDN1, COLEC12, CTGF, FXD5, GBP1, GDF15, GREB1, IFI16, IFI27, IFITM1, IL8, KLK5, KLK6, KLK8, KRT17, KRT5, KRT81, LCN2, LIF, MYEOV, NCAM2, NRCAM, PGR, PTGS2, PTRF, S100A9, TFF1, TFF3, TGM2 and TINAGL1) the expected DE was validated apart from PGR and IFI16. AFF3, COLEC12, GREB1, NCAM2, NRCAM, TFF1

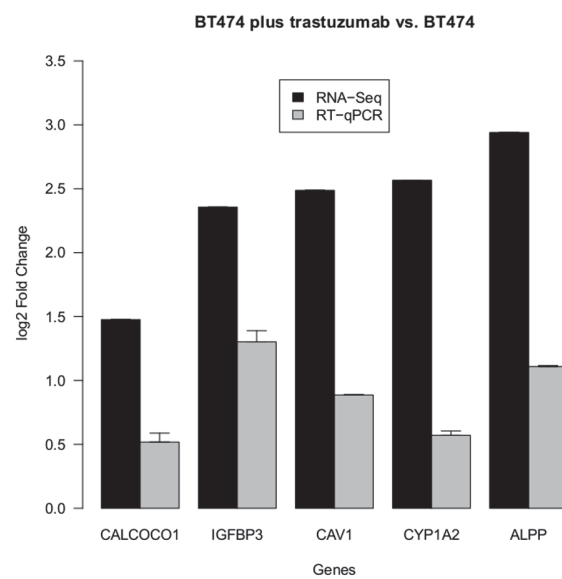


Fig 3. Fold Changes of DE genes (BT474 plus trastuzumab vs. BT474). The barchart displays the log2 fold changes of validated candidate genes, which significantly changed their expression in BT474 after trastuzumab treatment. The positive values indicate an upregulation upon drug treatment. Black bars denote values resulting from RNA-Seq analysis. Gray bars denote values resulting from RT-qPCR analysis.

doi:10.1371/journal.pone.0117818.g003

Table 1. Upregulated genes in BT474 upon trastuzumab treatment.

HGNC Symbol	Description	Log2 FC	FDR
ALPP	alkaline phosphatase placental	2.94	1.7e-04
CALCOCO1	calcium binding and coiled-coil domain 1	1.48	3.0e-02
CAV1	caveolin 1 caveolae protein 22kDa	2.49	2.7e-02
CYP1A2	cytochrome P450 family 1 subfamily A polypeptide 2	2.57	8.1e-03
IGFBP3	insulin-like growth factor binding protein 3	2.36	5.9e-08

The table displays the results of the RNA-Seq data comparison between BT474 and its trastuzumab treated version. It holds the description, log2 fold change (FC) and fdr of the differentially expressed genes. The positive log2 fold changes indicate upregulated gene expression after trastuzumab treatment.

doi:10.1371/journal.pone.0117818.t001

and TFF3 were significantly lower expressed in HCC1954 than in BT474, while the remaining ones were significantly higher expressed.

One candidate gene (MALAT1) was inferred by testing DE between BT474 and BTR50. Its significantly higher expression in BT474 was not validated.

Upregulated genes in BT474 upon trastuzumab treatment—indicators for drug sensitivity

The validated candidate genes ALPP, CYP1A2, CAV1, IGFBP3 and CALCOCO1 had positive log₂ fold changes (FC) indicating an upregulation in BT474 when treated with trastuzumab. [Fig. 3](#) shows the log₂ FC resulting from RNA-Seq and RT-qPCR analysis, respectively. [Table 1](#) lists the gene descriptions as well as the log₂ FC and fdr of the RNA-Seq analysis. It is an excerpt of [S3 Table](#).

ALPP and CYP1A2 showed the highest FC, but no association with trastuzumab has been reported so far. ALPP encodes placental alkaline phosphatase (PLAP) which is known as a tumor marker in seminoma and ovarian cancer [28]. CYP1A2 (cytochrome P450, family 1, subfamily A, polypeptide 2), which activity is known to be modulated by specific polymorphisms, is already supposed to influence breast cancer, as it is involved in breast carcinogen activation on the one hand, but produces beneficial estrogen and anti-inflammatory acids on the other hand [29].

IGFBP3 and CAV1 have already been reported in the context of trastuzumab efficacy. Regarding CAV1 (caveolin 1), Sekhar et al. discovered that although its tumor suppressor efficacy may be related to a HER2 downregulation in breast cancer cells, CAV1 and caveolae deficiency might be preferable under trastuzumab treatment [30]. Based on observations in the HER2+ human breast cancer cell line SKBR3, in which CAV1 was stably transduced, they speculated that an attenuated ADCC effect might be contributing to trastuzumab resistance. They found trastuzumab to be internalized and co-localized with CAV1, mediating endocytosis of HER2 by CAV1. Interestingly, in our study CAV1 was overexpressed upon trastuzumab treatment in the sensitive cell line BT474. The RT-qPCR data further revealed a stronger upregulation of CAV1 in HCC1954 compared to BT474, which was even stronger than the one in the trastuzumab treated BT474 cell line ([S2 Table](#)), underlining the benefit of a reduced CAV1 expression for drug efficacy. To further examine the role of CAV1 in trastuzumab treatment, we analyzed its expression in public available data from the transNOAH breast cancer trial (GEO series GSE50948) [31]. The original NeOAdjuvant Herceptin (NOAH) trial revealed the benefit of trastuzumab addition to treatment with neoadjuvant chemotherapy in HER2+ breast cancer patients [32]. We selected 11 patient samples with similar receptor status as BT474, i.e. estrogen receptor (ER)+/progesterone receptor (PR)+/HER2+ [33], of which four received neoadjuvant doxorubicin/paclitaxel treatment followed by cyclophosphamide/methotrexate/fluorouracil. The remaining seven patients received trastuzumab in addition for one year. CAV1 was differentially expressed between the two groups ($p < 0.05$) with an approximately two-fold upregulation in the trastuzumab treated group ([Fig. 4](#)) supporting our observation. [S1 Fig](#) shows the corresponding boxplots for all of the 54 candidate genes.

For IGFBP3, an insulin-like growth factor binding protein (IGFBP), several publications already hint at a positive association between its expression levels and trastuzumab efficacy. The corresponding biological processes of the Gene Ontology (GO) annotation included regulation of cell growth, negative regulation of signal transduction, protein phosphorylation and cell proliferation as well as positive regulation of apoptotic process, IGF receptor (IGFR) signaling pathway and MAPK cascade. According to Dokmanovic et al., trastuzumab enhanced IGFBP3 expression which in turn contributed to its sensitivity by growth inhibition [34]. They observed

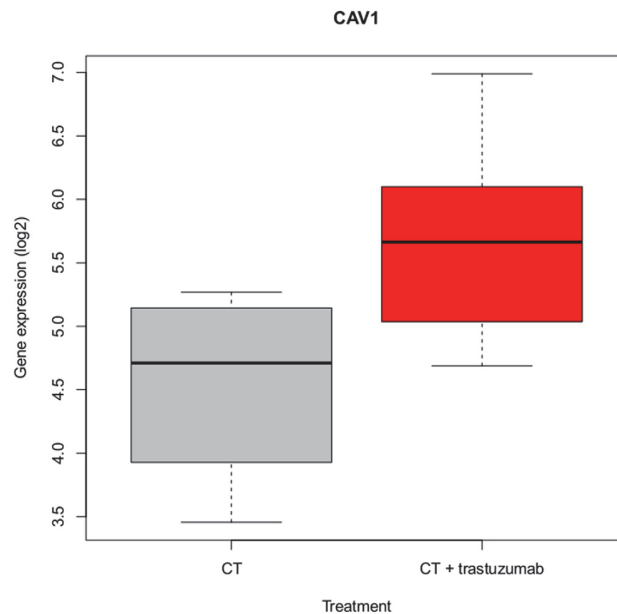


Fig 4. CAV1 expression is raised upon trastuzumab treatment. The boxplot displays CAV1 gene expression (log2) of seven patients treated with trastuzumab for one year in addition to neoadjuvant chemotherapy (red) and four patients treated with neoadjuvant chemotherapy only (gray). The patient data samples were selected from the transNOAH breast cancer trial (GEO series GSE50948).

doi:10.1371/journal.pone.0117818.g004

high IGFBP3 levels in the trastuzumab sensitive cell lines SKBR3 and BT474 and explained the way of growth inhibition by blocking crosstalk between the IGF-IR and HER2 signaling pathway. IGFBP2 and IGFBP3 were considered as potential predictive biomarkers for trastuzumab resistance by them, since IGFBP3 interfered with IGF-I-mediated mitogenic signaling and reduced IGFBP2-induced HER2 activation. Jerome et al. also observed that recombinant human IGFBP3 (rhIGFBP3) inhibited growth of HER2+ breast tumors and potentiated trastuzumab activity [35]. They observed that IGF-IR activation countered early effects of trastuzumab on HER2 signaling via AKT and ERK1/2. Downregulated phosphorylation of both proteins was associated with rhIGFBP3 inhibition of tumor growth. That is why Jerome et al. suggested simultaneous blockade of HER2 and IGF-IR pathways via rhIGFBP3 plus trastuzumab for HER2+ breast cancer patients. Lu et al. also reported a relation between IGF-IR signaling and trastuzumab resistance in HER2+ breast cancer cell models [36]. Their explanation for enhanced trastuzumab-induced growth inhibition by IGFBP3 was that IGFBP3 interferes with ligand-IGFR interactions.

Differentially expressed genes between HCC1954 and BT474—indicators for intrinsic resistance

Among the validated candidate genes, 24 had negative log2 fold changes (FC) and seven had positive ones. A negative log2 FC indicated significantly higher expression in HCC1954 compared to BT474. The genes with a positive log2 FC were significantly lower expressed in HCC1954 compared to BT474 and might contribute to trastuzumab efficiency. Fig. 5 shows the log2 FC resulting from RNA-Seq and RT-qPCR analysis, respectively. Table 2 lists the gene descriptions as well as the log2 FC and *fd*r of the RNA-Seq analysis. It is an excerpt of S3 Table.

Due to the large amount of candidates in this DE analysis, we focused on the ones which have already been mentioned in the context of trastuzumab efficacy, namely IL8, PTGS2,

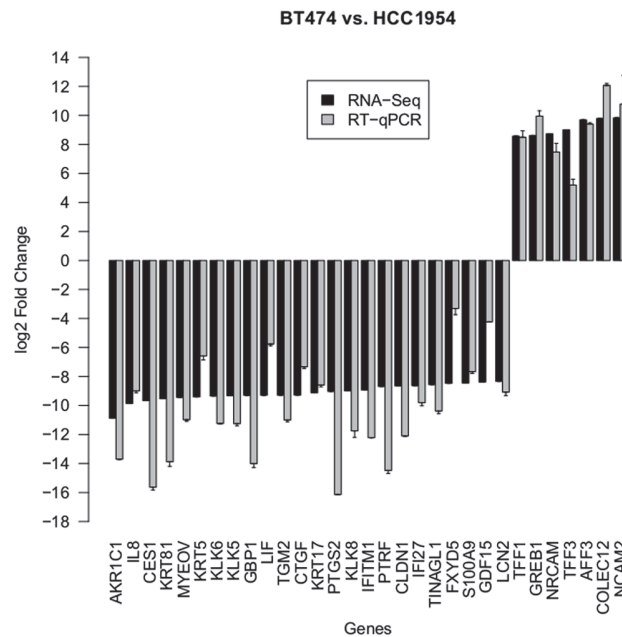


Fig 5. Fold Changes of DE genes (BT474 vs. HCC1954). The bar chart displays the log₂ fold changes of validated candidate genes, which showed significant differences in their expression in BT474 and HCC1954, respectively. Positive values indicate an upregulation in BT474. Negative values indicate an upregulation in HCC1954. Black bars denote values resulting from RNA-Seq analysis. Gray bars denote values resulting from RT-qPCR analysis.

doi:10.1371/journal.pone.0117818.g005

GDF15 and LCN2, all of which were upregulated in HCC1954. We analyzed the remaining genes of this group more generally in the context of breast cancer.

Upregulated genes in HCC1954 compared to BT474 IL8 (interleukin 8) is a chemotactic and inflammatory cytokine, which is produced upon inflammatory stimulation [37]. IL8 is associated with biological processes such as positive regulation of neutrophil chemotaxis, neutrophil activation, inflammatory and immune response. It was overexpressed in the resistant cell line HCC1954 compared to BT474. Korkaya et al. observed that PTEN downregulation and HER2 overexpression synergize to increase the expression of IL8 and IL6. Both of them are speculated to participate in an inflammatory loop, mediating trastuzumab resistance in HER2+ breast cancer by expanding the cancer stem cell population [38]. This loop implies AKT and subsequent NF- κ B activation. Interestingly, HCC1954 is already known for dysregulated PI3K/AKT signaling due to its PIK3CA mutation. Our results point to an important role of IL8 in driving trastuzumab resistance via a hyperactive PI3K pathway.

PTGS2 (cyclooxygenase prostaglandin-endoperoxide synthase 2), also known as COX2, was upregulated in HCC1954 as well and is connected to inflammation and carcinogenesis. Accordingly, it is usually not detectable in healthy tissues but can be induced in response to proinflammatory cytokines, growth factors and tumor promoters [39]. Ristimäki et al. stated that PTGS2 was detected in many breast tumors, correlated with poor patient prognosis, and its overexpression was shown to correlate with HER2 oncogene amplification in breast cancer [39]. Wang et al. also reported that nuclear HER2 bound at and transactivated the PTGS2 promoter, determining PTGS2 gene expression [40]. They further showed that trastuzumab triggered HER2 protein depletion and mitigated the association between nuclear HER2 and the PTGS2 promoter. Flowers and Thompson treated SKBR3 with t10c12 conjugated linoleic acid

Table 2. Differentially expressed genes between HCC1954 and BT474.

HGNC Symbol	Description	Log2 FC	FDR
AKR1C1	aldo-keto reductase family 1 member C1	-10.88	4.6e-02
CES1	carboxylesterase 1	-9.67	4.6e-02
CLDN1	claudin 1	-8.65	4.6e-02
CTGF	connective tissue growth factor	-9.28	4.6e-02
FXYD5	FXYD domain containing ion transport regulator 5	-8.47	4.6e-02
GBP1	guanylate binding protein 1 interferon-inducible	-9.30	4.9e-02
GDF15	growth differentiation factor 15	-8.38	4.6e-02
IFI27	interferon alpha-inducible protein 27	-8.63	4.6e-02
IFITM1	interferon induced transmembrane protein 1	-8.94	4.6e-02
IL8	interleukin 8	-9.86	4.6e-02
KLK5	kallikrein-related peptidase 5	-9.32	4.6e-02
KLK6	kallikrein-related peptidase 6	-9.35	4.6e-02
KLK8	kallikrein-related peptidase 8	-8.99	4.6e-02
KRT17	keratin 17	-9.12	4.6e-02
KRT5	keratin 5	-9.41	4.6e-02
KRT81	keratin 81	-9.53	4.6e-02
LCN2	lipocalin 2	-8.34	4.6e-02
LIF	leukemia inhibitory factor	-9.29	4.9e-02
MYEOV	myeloma overexpressed	-9.44	4.6e-02
PTGS2	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	-9.03	4.6e-02
PTRF	polymerase I and transcript release factor	-8.68	4.6e-02
S100A9	S100 calcium binding protein A9	-8.45	4.6e-02
TGM2	transglutaminase 2	-9.29	4.6e-02
TINAGL1	tubulointerstitial nephritis antigen-like 1	-8.55	4.6e-02
AFF3	AF4/FMR2 family member 3	9.70	4.6e-02
COLEC12	collectin sub-family member 12	9.79	4.6e-02
GREB1	growth regulation by estrogen in breast cancer 1	8.62	4.6e-02
NCAM2	neural cell adhesion molecule 2	9.83	4.6e-02
NRCAM	neuronal cell adhesion molecule	8.73	4.6e-02
TFF1	trefoil factor 1	8.57	4.6e-02
TFF3	trefoil factor 3 (intestinal)	9.01	4.6e-02

The table displays the results of the RNA-Seq data comparison between BT474 and HCC1954. It holds the description, log2 fold change (FC) and fdr of the differentially expressed genes. Positive log2 fold changes indicate higher gene expression in BT474, while negative log2 fold changes indicate higher gene expression in HCC1954.

doi:10.1371/journal.pone.0117818.t002

(CLA) and observed a HER2 suppression as well as enhanced apoptosis [41]. This confirmed the expected anti-tumor properties of t10c12 CLA via inhibition of NF-κB activity and PTGS2-derived prostaglandin E2 (PGE2). Both factors play a role in resistance against HER2-targeted therapy with trastuzumab, which usually improves disease-free survival (DFS) for women with HER2+ breast cancers. In this context we derived the GO terms ‘response to fatty acid’, ‘positive regulation of NF-κB import into nucleus’ and ‘positive regulation of prostaglandin biosynthetic process’. Regulation of HER2 oncogene expression by PTGS2 and PGE2 was analyzed by Benoit et al. [42]. According to their study, high levels of the proinflammatory and antiapoptotic enzyme PTGS2 have already been detected in HER2+ tumors, linked to HER2-mediated induction of PTGS2 gene transcription. In turn, PTGS2 expression and synthesis of its enzymatic product PGE2 lead to enhanced HER2 expression, which is known to correlate with adverse

prognosis in breast cancers. Thus, HER2 and PTGS2 transcriptionally regulate each other in a positive feedback loop. Indeed, PTGS2 inhibition reduced HER2 protein levels, increased cancer cells sensitivity to chemotherapeutic treatment and acted synergistically with trastuzumab. The association between PTGS2 and HER2 overexpression was also analyzed by Howe et al. [43]. They emphasize that PTGS2 or derived prostaglandins are known to promote angiogenesis and invasiveness of diverse cancer types. We inferred the GO term ‘positive regulation of cell migration involved in sprouting angiogenesis’ in this context. Especially the HER2/HER3 signaling pathway was shown to regulate expression of PTGS2 and its associated pathway [44]. As the HER2/HER3 heterodimer is known to stimulate the PI3K pathway [45], it confirms our observation that PTGS2 was overexpressed in the PIK3CA mutated HCC1954 cell line. Basu et al. detected that PTGS2 inhibition decreased cell growth and promoted apoptosis in metastatic breast cancer [46]. Indeed, PTGS2 is known to positively regulate apoptotic processes and growth factor (GF) production, e.g. TGF- β , VEGF, fibroblast GF (FGF) and platelet-derived GF (PDGF). As a link is assumed between overexpression of HER2 and PTGS2 activity in breast cancer, Gianni speculated that a combination of PTGS2 inhibitors with trastuzumab may be efficient in high-risk patients [47]. Within a corresponding phase II study, Dang et al. tested a combined treatment of metastatic breast cancer patients who have progressed after prior trastuzumab-based treatments with trastuzumab and celecoxib, a PTGS2 inhibitor [48], which already showed promising results in colorectal cancer [49]. Unfortunately, they revealed that the combination was not active in patients with HER2-overexpressing, trastuzumab-refractory disease. Cho et al. suggested to combine the HIV protease inhibitor nelfinavir and celecoxib or analogs to increase endoplasmic reticulum stress and toxicity in human breast cancer cell lines that were resistant to trastuzumab [50]. The background was that moderate activity of the stress response system exerts anti-apoptotic function, supporting tumor cell survival and chemoresistance, whereas more severe aggravation may turn on its pro-apoptotic module. In this context PTGS2 was also associated with the GO term ‘positive regulation of cell death’. Morrison et al. reported a marginal association of abnormal copy number of PTGS2 with objective trastuzumab response in metastatic breast cancer [51].

GDF15 (growth differentiation factor 15) was overexpressed in HCC1954, too. Hence, it could contribute to trastuzumab resistance as well. According to GO annotation, it is associated to cell-cell signal transduction, especially to the TGF- β signaling pathway. Indeed, Joshi et al. observed GDF15-mediated HER2 phosphorylation reducing trastuzumab sensitivity of HER2+ breast cancer cells [52]. They regarded GDF15-mediated activation of TGF- β receptor-Src-HER2 signaling crosstalk as a mechanism of trastuzumab resistance.

Another upregulated gene in HCC1954 was LCN2, encoding the chaperone protein lipocalin 2. Kumandan et al. reported that LCN2 is known to be highly upregulated during the unfolded protein response (UPR) in an NF- κ B-dependent manner [53]. As UPR activates the HER2/PI3K/AKT/NF- κ B signaling pathway, LCN2 is assumed to play an important role as a common downstream effector molecule. Kumandan et al. observed that trastuzumab inhibited the expression of LCN2 in SKBR3 cells. The downregulation of LCN2 was abrogated by activation of the UPR. An increase of LCN2 transcription and secretion was observed. It was speculated that UPR induction bypasses trastuzumab-mediated inhibition of the PI3K pathway. Lin et al. observed that exposure of human endometrial carcinoma cells to LCN2 for more than 24 hours reduced LCN2-induced cell apoptosis, changed proliferation and upregulated, amongst other cytokines, IL8 secretion [54]. Recombinant IL8 lead to decreased activity of the apoptotic enzyme caspase-3. The addition of IL8 antibodies resulted in significantly increased caspase-3 activity and decreased cell migration, indicating an important role of IL8 in the induction of cell migration. Interestingly, we associated the GO term ‘regulation of apoptotic process’ with LCN2 and UPR related GO terms with IL8, namely ‘activation of signaling

protein activity involved in unfolded protein response' and 'endoplasmic reticulum unfolded protein response'. Hence, IL8 does not only seem to be stimulated by UPR-induced LCN2, but also participates in the activation of UPR. This points to a synergistic feedback mechanism involving IL8 and LCN2, which could contribute to trastuzumab resistance.

Many of the remaining validated candidates which were differentially expressed between BT474 and HCC1954 with overexpression in HCC1954 have already been mentioned in the context of breast cancer, mainly as tumor promoters.

Regarding AKR1C1 (aldo-keto reductase family 1, member C1), a loss of expression and related decreased progesterone catabolism, leading to decreased growth in the presence of progesterone, has been observed in breast tumors by Ji et al. [55]. They speculated that loss of AKR1C1 in combination with increased PR expression may enhance progesterone-dependent PR activation, which plays an essential role in breast development and cancer formation. This fits to our observation that AKR1C1 was downregulated in BT474 compared to HCC1954. BT474 is known for its PR+ status [33], while HCC1954 is negative for expression of ER and PR according to the manufacturer's product information.

A relation to steroid receptors was also found for MYEOV (myeloma overexpressed), which DNA amplification correlated with ER+ and PR+ cancer in a study of Janssen et al. [56]. Although MYEOV is overexpressed in breast tumors, they did not detect its expression in BT474. They showed that abnormal expression levels of the candidate oncogene correlated only partially with DNA amplification and that overexpression could act independently or cooperatively with the cell cycle regulator Cyclin D1 to favor tumorigenesis. Hence, overexpression of MYEOV in HCC1954 underlined the outstanding role in tumor promotion of this cell line.

The keratin KRT5 (CK5) is related to progesterone as well, since CK5+ cells were shown to lose ER and PR expression [57], fitting to our observation of overexpression in HCC1954 compared to BT474. Furthermore Axlund et al. observed that cells overexpressing CK5 are more invasive, sphere-forming, and quiescent with an increased resistance to endocrine and chemotherapy [57]. CK5 is a basal marker in breast carcinoma, related to poor prognosis and to unfavorable overall survival (OS) as well as relapse-free interval [58]. Apart from KRT5, two further keratins were overexpressed in HCC1954, namely KRT17 (CK17) and KRT81 (KRTHB1). CK17 is a basal-type cytokeratin as well, and the expression of CK17 and CK5/6 appears to define a group of breast tumors with a poor clinical outcome, especially high mortality rates [59].

Another group of related genes was formed by the three kallikrein-related peptidases KLK5, KLK6 and KLK8, all of them overexpressed in HCC1954. High KLK5 expression was found more frequently in node-positive and ER-negative (ER-) patients [60], confirming our observation of upregulation in HCC1954. Yousef et al. showed that KLK5 overexpression was a predictor of reduced DFS and OS in breast cancer, and hence a marker of poor prognosis. They regarded KLK5 as a candidate that might stimulate cellular growth, angiogenesis or degradation of the extracellular matrix (ECM) [60]. Analogously higher KLK6 protein levels were found to be associated with tumors which are negative for ER and PR [61], as it is the case for HCC1954. Yousef et al. observed that the closely localized kallikreins KLK5, KLK6 and KLK8 are in general downregulated with a similar pattern of differential expression in breast cancer [62].

A promoting effect on tumorigenesis and metastasis of breast cancer via the AKT/mTOR signaling pathway, independent of ER status, was detected for the multi-functional cytokine protein LIF (leukemia inhibitory factor) by Li et al. [63]. We observed an upregulation of LIF in HCC1954. Li et al. reported that its overexpression has been observed in several types of cancers including breast cancer and is associated with a poorer relapse-free survival. They showed that LIF promotes cell proliferation and growth of breast cancer cells in vitro, and growth of

xenograft breast tumors *in vivo*. Furthermore it promoted invasion and migration of breast cancer cells *in vitro* and metastasis of breast cancer *in vivo*.

A further gene upregulated in HCC1954 was TGM2. As reviewed by Agnihotri et al., TGM2 is a stress-responsive gene, encoding the multifunctional ubiquitously expressed enzyme transglutaminase 2 (TG2) which seems to play a crucial role in promoting an aggressive phenotype in mammary epithelial cells [64]. Its expression is upregulated during inflammation and wounding, as it crosslinks ECM component proteins and stabilizes the matrix for increased cell attachment and motility. It has also been observed that anti-apoptotic TGM2 is upregulated in cancer, especially those resistant to chemo- and radiation therapy and those isolated from metastatic sites. In mammary epithelial cells, chronic TG2 expression initiates signaling contributing to drug resistance and an invasive phenotype, and high expression levels are associated with activation of indicators of aggressive tumors, such as AKT and NF- κ B in a feedback loop. Above that TG2 expression induces epithelial-to-mesenchymal transition and confers cancer stem cell trait, both of which have been implicated in metastasis and resistance to standard therapies. TG2 expression in tumor samples is associated with poor disease outcome, increased (chemotherapeutic) drug resistance, and increased incidence of metastasis [64].

Another gene highly upregulated in HCC1954 was CTGF (connective tissue growth factor). Although literature is ambivalent, assigning either a tumor suppressor or enhancer role to this gene, the oncogenic character of CTGF seems to dominate. Its expression is elevated in advanced stages of breast cancer, and Chen et al. observed enhanced cellular migratory ability in breast cancer cells overexpressing CTGF [65]. According to them, CTGF mediated ERK1/2 activation and hence cellular migration. Additionally, CTGF mediated upregulation of the prometastatic gene S100A4, dependent on ERK1/2. This points to an important role of CTGF in migration and invasion, and supports other investigators who linked overexpression to tumor size and lymph node metastasis or related CTGF to angiogenesis and bone metastasis in breast cancer. Interestingly, Chen et al. detected extremely low or no levels of CTGF mRNA in BT474, confirming our observation. Furthermore, the relation of CTGF and S100A4 is of special interest, as we also detected a downregulation of the S100 calcium binding protein family member S100A9 in BT474 compared to HCC1954. Gonçalves et al. found an association between S100A9, a protein expressed in invasive breast cancer, with basal subtypes as well as corresponding poor differentiation and prognosis value [66]. Most interestingly, they inferred BT474 and HCC1954 as different molecular subtypes, i.e. BT474 as luminal-like, and HCC1954 clustered together with basal-like cell lines, although no specific type was assigned. S100A9 expression was, amongst others, tightly associated with high grade, negative ER and PR status, and HER1 and HER2 expression. Furthermore, it was significantly associated with reduced metastasis-free and overall survival.

PTRF (polymerase I and transcript release factor) was upregulated in HCC1954 as well in our study. The RT-qPCR data of BT474 and HCC1954, respectively, revealed similar expression patterns for PTRF and CAV1 (S2 Table). Verma et al. detected that PTRF (cavin-1) and caveolin 1 are lost coordinately in cancer cells [67]. Membrane recruitment of the cavin complex requires caveolin 1, and PTRF is important for regulating caveolin 1 membrane dynamics.

Another upregulated gene in HCC1954 was CLDN1, encoding the tight junction protein claudin 1. Di Cello et al. pointed out that CLDN1 may alternatively function as a tumor suppressor or as an oncogene, but in breast cancer they assumed a tumor suppressor role, since CLDN1 is frequently downregulated in this type of cancer, associated with recurrence, metastasis, and reduced survival [68]. They explained its downregulation by epigenetic silencing via DNA methylation. Interestingly, they demonstrated this using amongst others the BT474 cell line, which showed low CLDN1 expression. According to their results, in breast cancer, CLDN1 expression varies according to the molecular subtype. While methylation of the

CLDN1 promoter CpG island was frequent in ER+ breast cancer, it was not the case in most of the ER- breast cancers samples, of which some overexpressed CLDN1. HCC1954 fell into a cluster composed mostly of ER-, basal-like cell lines, which featured loss of methylation at sites downstream of the CpG island. Lu et al. detected positive associations between claudin 1 and CK5/6 and HER1 [69], which is of special interest, as HER1 and HER2 function as heterodimers [45], and we detected CK5 as being overexpressed in HCC1954 as well. Above that they found a negative association between CLDN1 and ER, fitting to our results of upregulation in the ER- cell line HCC1954. Confirming the results of Di Cello et al., Lu et al. observed high levels of CLDN1 in the basal-like subtype.

Like CLDN1, FXYD5 (FXD domain containing ion transport regulator 5), also known as dysadherin, is involved in cell-cell adhesion, which was downregulated in BT474 compared to HCC1954 as well. Lee et al. detected no dysadherin expression in BT474, likely due to its ER+ status, as dysadherin is expressed in the more aggressive ER- breast cancer [70]. In their study the introduction of dysadherin cDNA into BT474 enhanced levels of AKT phosphorylation. Hence, they speculated that dysadherin, which overexpression is a predictor of metastasis, invasion and poor prognosis, might contribute to breast cancer progression through ER-dependent AKT activation.

Upregulated genes in BT474 compared to HCC1954 Many of the remaining validated candidates which were differentially expressed between BT474 and HCC1954 with overexpression in BT474 are induced by estrogen receptors (ER). This is explainable by the ER status of the cell lines. While BT474 is known to be an ER+/PR+ cell line [33], HCC1954 is an ER-/PR- one.

COLEC12 (collectin sub-family member 12) was highly overexpressed in BT474 and is likely to be an ER β primary target gene, indicating an antiproliferative role [71].

Analogously, TFF3 (trefoil factor 3) expression has been reported to correlate with ER α , hinting at induction by this receptor subtype. Additionally, a positive correlation was observed between PR and TFF3 protein expression by Ahmed et al. [72], confirming our observed TFF3 overexpression in the BT474 cell line. According to their results, TFF3 expression is induced by estrogens and hardly detected in breast tumors which do not express ER. The luminal subtype A, which is primarily composed of ER+ tumors, demonstrates a better prognosis than other breast cancer subtypes and is characterized by overexpression of estrogen-related genes such as TFF3 and ER α [73]. Ahmed et al. observed a higher TFF3 expression in well-differentiated tumors and an association with low histological grade. Nevertheless, they speculated that in invasive cancer the positive characteristics of TFF3 turn into oncogenic ones [72]. Obviously, the ER+/PR+ cell line BT474 benefits from the estrogen-related positive TFF3 effects. The ER α -responsive gene TFF1 is known to be coexpressed with TFF3 in ER+ malignant breast cancer cells and is likewise upregulated by estrogens. Although TFF1 is a disseminated tumor cells detection marker, TFF1 and TFF3 are components of a luminal epithelial signature defining a well-differentiated, low-grade breast cancer subtype [74].

The upregulation of NRCAM (neuronal cell adhesion molecule) in BT474 can be ascribed to the ER status, too. This gene was reported to be expressed in ER+ breast cancer cell lines but not in ER- ones [75].

Another estrogen-induced, ER-responsive gene upregulated in BT474 was GREB1 (growth regulation by estrogen in breast cancer 1). It is an important factor in ER induced proliferation in breast cancer cells [76]. Liu et al. reported that overexpression of the ER cofactor GREB1 also increased the clonogenic ability in breast cancer cells and that overexpression was observed in ER+ breast cancer patients compared to ER- ones [77]. They stressed that reduced levels of GREB1 are predictive of worse disease outcome for breast cancer patients in general, and for ER+ patients in particular.

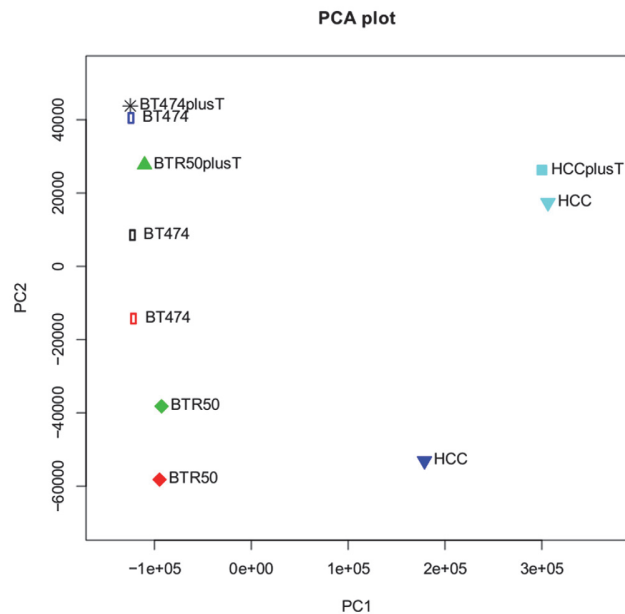


Fig 6. Principle Component Analysis (PCA) of all measured samples. The plot displays the result of a PCA on the pairwise normalized RNA-Seq expression values of all 23367 annotated genes in the samples BT474, BTR50 and HCC1954 with and without trastuzumab (T) treatment. Same colors denote that the samples belong to the same conducted statistical test and thus were normalized in pair. The five two sample tests included *BT474 vs. HCC1954*, *BT474 vs. BTR50*, *HCC1954+T vs. HCC1954*, *BTR50+T vs. BTR50*, and *BT474+T vs. BT474*.

doi:10.1371/journal.pone.0117818.g006

Interestingly, the RT-qPCR data of all genes overexpressed in BT474 compared to HCC1954 revealed a three-fold overexpression in the trastuzumab treated version of BT474 compared to its untreated version (S2 Table). Evans et al. already speculated that in HER2+/ER+ tumors increased TFF1 expression and estrogen signaling may occur as a result of inhibiting the EGFR/HER2 signaling pathway [78]. Collins et al. confirmed that trastuzumab treatment upregulated ER transcriptional activity and TFF1 expression [79]. However, they reported a mostly poor response to trastuzumab treatment in HER2+ cells with co-expression of the steroid pathway.

Furthermore, the RT-qPCR data also show that the mentioned genes are at least as strongly expressed in the BTR50 cell line compared to BT474, and not downregulated as in HCC1954 (S2 Table). This hints at a strong discrepancy between expression profiles in the cell line models of intrinsic and acquired drug resistance. This could be explained by the individual characteristics of the HCC1954 cell line, maybe also driven by specific SNPs, which are analyzed in the following section. Also, the Principal Component Analysis (PCA) plot of the pairwise normalized cell lines, considering all 23367 annotated genes, demonstrates that the intrinsically resistant cell line HCC1954 differs much more from BT474 than BTR50, and that trastuzumab treatment hardly influences expression in this cell line (Fig. 6).

SNP analysis of genes associated with trastuzumab efficiency

We searched for SNPs in the three untreated cell lines HCC1954, BTR50 and BT474. In the following, we describe the results for the validated gene candidates of the DE analysis on BT474 and its trastuzumab treated version (ALPP, CALCOCO1, CAV1, CYP1A2, IGFBP3) as well as

the candidates of the DE analysis on BT474 vs. HCC1954 which were already reported in the context of trastuzumab efficacy, namely GDF15, IL8, LCN2 and PTGS2. Additionally we focus on HER2 due to the HER2+ status of the cell lines, PIK3CA due to the known mutation in HCC1954, as well as MAPK1 and AKT1 as main players in the MAPK and PI3K/AKT pathway, respectively. [S1 Variant Call Format File](#), [S2 Variant Call Format File](#) and [S3 Variant Call Format File](#) store the SNPs called in the HCC1954 ([S1 Variant Call Format File](#)), BTR50 ([S2 Variant Call Format File](#)) and BT474 ([S3 Variant Call Format File](#)) sample.

SNPs in the HCC1954 cell line

[S1 Variant Call Format File](#) stores the SNPs called in the HCC1954 sample. For eight of the 13 genes of interest we detected SNPs in HCC1954 which passed our filter criteria.

We called 13 variations for HER2, five variations for MAPK1, three variations for PTGS2 and AKT1, respectively, two variations for IGFBP3 and one SNP for PIK3CA, CAV1 and IL8 each. The ALPP SNP at position 233243981 on chromosome 2 (rs13026692, COSM1162692)

Table 3. SNPs called in the HCC1954 cell line.

Gene	Chromosome	Position	Variation
AKT1	14	105235824	rs58565216
AKT1	14	105236287	rs35416681
AKT1	14	105241304	rs2230506
CAV1	7	116200554	rs374946197
HER2	17	37855834	rs4252596
HER2	17	37869895	-
HER2	17	37870378	rs2934967
HER2	17	37871081	-
HER2	17	37876835	rs2952156
HER2	17	37877412	rs2952157
HER2	17	37877447	rs11653998
HER2	17	37878371	-
HER2	17	37878574	-
HER2	17	37878635	-
HER2	17	37879030	rs2088126
HER2	17	37884037	rs1058808, rs370420724, CM087578
HER2	17	37885332	rs2952158
IGFBP3	7	45957678	rs3793345
IGFBP3	7	45960645	rs2854746, CM057457
IL8	4	74609045	rs1126647, CR025956
MAPK1	22	22115004	rs6928
MAPK1	22	22115353	rs9340
MAPK1	22	22115498	rs3810610
MAPK1	22	22116202	rs13943
MAPK1	22	22116467	rs1063311
PIK3CA	3	178952085	rs121913279, COSM776, COSM775, COSM249874, COSM94987, COSM94986
PTGS2	1	186641626	rs2853805
PTGS2	1	186642429	rs2206593
PTGS2	1	186643058	rs5275

The table displays information about chromosomal location of the detected SNPs in the genes of interest as well as corresponding SNP ID numbers [8]. In case of missing ID numbers ('-'), we likely inferred a novel mutation.

doi:10.1371/journal.pone.0117818.t003

had a read depth of nine, just slightly failing our filter criterion. The related locations and reference SNP ID numbers [8] are listed in Table 3. 80% of the SNPs have already been listed in the *dbSNP* database but just 40% of those have been analyzed more deeply in literature, even less in the context of breast cancer.

Regarding rs2952156 and rs4252596 of HER2, Einarsdóttir et al. did not observe an association with breast cancer survival [80]. Analogously, Benusiglio et al. analyzed whether common polymorphisms (frequency $\geq 5\%$) in HER2 were associated with breast cancer risk in a white British population [81]. Although they considered rs4252596 near the promoter as potentially regulatory functional and the nonsynonymous coding SNP rs1058808 of HER2 potentially affecting tyrosine kinase activity or protein structure, there was no evidence that these variations were associated with breast cancer. Breyer et al. analyzed the role of genetic variation of the HER2 gene in breast cancer risk in a study of invasive breast cancer cases from the Shanghai Breast Cancer Study [82]. According to their results, HER2 resides within a locus of high linkage disequilibrium, composed of three major ancestral haplotypes in the study population. These haplotypes are marked by simple tandem repeat and SNPs including the missense variants I655V and P1170A (rs1058808). However, they did not observe an association of the SNP haplotypes with breast cancer predisposition.

For PTGS2 we detected the polymorphism rs5275 in the 3' untranslated region of the PTGS2 gene. Festa-Vasconcellos et al. investigated the association between PTGS2 haplotypes and histopathological parameters with prognostic value on clinical outcome, i.e. risk of tumor recurrence, of breast cancer [83]. Their study involved women under treatment for non-metastatic breast cancer who were genotyped for rs689465, rs689466, rs20417 and rs5275. Eight haplotypes were inferred with a significant difference in their distribution as a function of tumor size, ER and HER2 status. No associations were found between rs5275 solely and histopathological features, which was in line with the observations of Gerger et al. [84] and Abraham et al. [85] who found no associations between rs5275 and DFS or OS of breast cancer patients. Above that two breast cancer meta-analyses indicated no strong risk association for rs5275 [86, 87]. Though, Festa-Vasconcellos et al. observed an rs5275 dependent positive association between higher tumor size and the haplotype formed by rs689465G, rs689466A, rs20417C and rs5275T. At least SNPs rs689465 and rs5275 were required to infer this haplotype, which was further positively associated with ER and PR negativity and HER2 positivity. Moore et al. supported an important influence of rs5275, reporting it disrupts micro-RNA-mediated regulation of PTGS2 mRNA degradation, leading to increased mRNA stability and PTGS2 protein levels [88]. Jung et al. analyzed polymorphisms of apoptosis-related genes and their impact on the survival of patients with early invasive ductal breast cancer [89]. The PTGS2 gene polymorphism rs5275 in a dominant model of the C allele was associated with a higher distant DFS. Cox et al. genotyped the five most common PTGS2 polymorphisms (rs20417, rs5277, rs20432, rs5275, and rs4648298) in the Nurses' Health Study to test for an association with breast cancer risk [90]. The SNP rs5275 was associated with a decrease in breast cancer risk and further genotyped in the Nurses' Health Study 2 and Harvard Women's Health Study. In pooled analyses, women homozygous for the T allele at rs5275 had a 20% lower risk of breast cancer than those homozygous for the C allele. Cox et al. concluded that this polymorphism may be associated with a decrease in breast cancer risk among Caucasian women, but is not associated with an increased risk of breast cancer.

We further identified the IGFBP3 SNP rs2854746. Qian et al. evaluated IGF-I and IGFBP3 genotypes in relation to their phenotypes in local breast tissues and in association with breast cancer risk for Chinese women [91]. No association was found between breast cancer risk and the IGFBP3 SNP rs2854746, but the genotype correlated with IGF-I phenotypes in tumor samples. Peptide levels of IGF-I were inversely correlated with age and menopause status. The

homozygous variant genotype of rs2854746 had lower IGF-I compared to the wild type. This suggested possible influences of the SNP on IGF-I activity in local tissues. IGF-I and its major binding protein IGFBP3 were also analyzed by Su et al. due to their implication in breast carcinogenesis [92]. They examined associations between genetic variants and circulating levels of IGF-I and IGFBP3 with proliferative benign breast disease (BBD), a marker of increased breast cancer risk. Higher circulating IGFBP3 levels were significantly associated with increased risk of proliferative BBD. The SNP rs2854746 was significantly associated with circulating IGFBP3 levels. D'Aloisio et al. reported that IGFBP3 plasma levels, associated with common diseases, are influenced by common IGFBP3 SNPs, especially rs2854746, among African American and Caucasian women [93]. Similar observations were made by Patel et al. [94] and Cheng et al. [95]. However, Tamimi et al. did not detect any significant association of the common haplotypes in three haplotype blocks, of which one included rs2854746, spanning the genes encoding IGFBP1/IGFBP3, with mammographic density, one of the strongest risk factors for breast cancer [96].

Regarding PIK3CA, we inferred the SNP rs121913279, which is referring to the known H1047R mutation in HCC1954. Tong et al. detected this SNP in Chinese breast cancer patients with invasive ductal carcinomas [97].

SNPs in the BTR50 cell line

[S2 Variant Call Format File](#) stores the SNPs called in the BTR50 sample. For three of the 13 genes of interest we detected SNPs in BTR50 which passed our filter criteria. We called 13

Table 4. SNPs called in the BTR50 cell line.

Gene	Chromosome	Position	Variation
HER2	17	37859083	rs34284966
HER2	17	37876179	rs4252639
HER2	17	37876835	rs2952156
HER2	17	37877221	-
HER2	17	37877412	rs2952157
HER2	17	37878113	rs115334808
HER2	17	37878311	-
HER2	17	37878371	-
HER2	17	37878574	-
HER2	17	37878635	-
HER2	17	37878696	-
HER2	17	37879030	rs2088126
HER2	17	37885332	rs2952158
IGFBP3	7	45952254	rs6670
MAPK1	22	22115004	rs6928
MAPK1	22	22115353	rs9340
MAPK1	22	22115498	rs3810610
MAPK1	22	22115886	rs13515
MAPK1	22	22116202	rs13943
MAPK1	22	22116467	rs1063311
MAPK1	22	22162072	-

The table displays information about chromosomal location of the detected SNPs in the genes of interest as well as corresponding SNP ID numbers [8]. In case of missing ID numbers ('-'), we likely inferred a novel mutation.

doi:10.1371/journal.pone.0117818.t004

Table 5. SNPs called in the BT474 cell line.

Gene	Chromosome	Position	Variation
HER2	17	37876179	rs4252639
HER2	17	37877412	rs2952157
HER2	17	37878113	rs115334808
MAPK1	22	22115004	rs6928
MAPK1	22	22115498	rs3810610
MAPK1	22	22115886	rs13515
MAPK1	22	22116202	rs13943
MAPK1	22	22116467	rs1063311

The table displays information about chromosomal location of the detected SNPs in the genes of interest as well as corresponding SNP ID numbers [8].

doi:10.1371/journal.pone.0117818.t005

variations for HER2, seven variations for MAPK1 and one variation for IGFBP3. The related locations and reference SNP ID numbers are listed in Table 4. Two third of the SNPs have already been listed in the dbSNP database but just 30% of those have been analyzed more deeply in literature, even less in the context of breast cancer. Approximately 60% of the SNPs were also detected in HCC1954 which would indicate that this intersection of mutations contributes to resistance. Interestingly, the SNP of IGFBP3 in BTR50 was not matching the ones of HCC1954.

SNPs in the BT474 cell line

S3 Variant Call Format File stores the SNPs called in the BT474 sample. For two of the 13 genes of interest we detected SNPs in BT474 which passed our filter criteria. We called three variations for HER2 and five variations for MAPK1. All mutations form a subset of the corresponding ones in BTR50. This makes sense, as BTR50 has been derived from BT474 and cultured to resistance, which obviously led to additional mutations in the cell line. The related locations and reference SNP ID numbers are listed in Table 5.

Conclusions

The aim of the study was to infer genes and corresponding variations, which influence trastuzumab action in HER2+ breast cancer. Comparing gene expression in three cell line models of different drug response characteristics with and without trastuzumab treatment, we inferred 54 candidate genes. Most of them have already been reported in the context of breast cancer but just a small proportion in the context of trastuzumab action.

The validated selection of the candidate genes was differentially expressed between the sensitive cell line BT474 and its trastuzumab treated version, as well as between BT474 and the resistant cell line HCC1954. The candidates of the latter comparison were mainly related to the different steroid receptor status of the cell lines, mostly including tumor enhancers with upregulation in HCC1954 which have already been ascribed to ER- breast cancer subtypes. Outstanding candidates were IL8, PTGS2, GDF15 and LCN2, as those have already been reported to hinder trastuzumab action. Interestingly, mechanisms involving the PI3K pathway have been reported, which fits to the PI3K gain-of-function mutation in the resistant cell line HCC1954. Furthermore, a cooperative functionality of IL8 and LCN2 would stabilize the resistant phenotype. The two outstanding candidates revealed in the comparison of BT474 and its drug treated version were CAV1 and IGFBP3, both of which have already been reported to

influence trastuzumab efficacy. While low *CAV1* expression seems to support trastuzumab action, high *IGFBP3* levels are associated with trastuzumab sensitivity.

We further detected that the intrinsically resistant cell line HCC1954 differs more from the sensitive cell line BT474 than the cell line BTR50, representing acquired resistance. HCC1954 was also influenced less by trastuzumab treatment than BTR50. Thus, intrinsically resistant specimens seem to respond less to trastuzumab treatment than those which acquired resistance. This was supported by the fact that we called more SNPs in the intrinsically resistant cell line.

All detected mutations in BT474 formed a subset of the ones detected in BTR50, pointing at resistance related novel mutations in the cell line which acquired resistance. Four annotated SNPs called in BTR50, which were not called in BT474, overlapped with the variations called in HCC1954, i.e. rs2952156, rs2088126 and rs2952158 of *HER2*, as well as rs9340 of *MAPK1*. Those SNPs could likely contribute to trastuzumab resistance. In general, some of the inferred SNPs are already in the focus of breast cancer research, but mostly no influence on survival or risk has been detected for single ones. However, some seem to have an impact on the phenotype in combination with further variations. Regarding the SNP rs5275 of *PTGS2* in HCC1954, we found hints at a relation to breast cancer risk. Additionally, the SNP rs2854746 of *IGFBP3* in HCC1954 seems to influence *IGF-I* and *IGFBP3* expression.

In conclusion, we applied successfully the RNA-Seq method to detect differentially expressed genes as well as SNPs in the context of trastuzumab resistance in different cell phenotypes. A high proportion of validated genes by RT-qPCR confirmed our results. Based on literature research, we elucidated the role of these candidate genes or mutations in influencing trastuzumab action. A deeper analysis of single candidates, supported by additional experiments, could reveal more detailed information about mechanisms leading to trastuzumab resistance. Hence, our results provide a basis to improve personalized medicine for HER2+ breast cancer patients.

Supporting Information

S1 Table. Primers used for RT-qPCR validation. The table lists the used forward and reverse primer sequences for the 40 candidate genes which were analyzed by RT-qPCR. Additionally, the sequences for the normalizer *GAPDH* are listed.
(XLS)

S2 Table. RT-qPCR validation data. The table lists the RT-qPCR data of 40 selected genes from the candidate set. The ViiA 7 Real-Time PCR System (Applied Biosystems, Darmstadt, Germany) was used for RT-qPCR analysis. Data analysis was performed as described in Livak and Schmittgen [27] with two biological replicates and *GAPDH* as reference gene.
(XLS)

S3 Table. DE genes revealed by RNA-Seq analysis. The table lists the genes which were significantly differentially expressed (DE), i.e. genes with a fold change (FC) of at least two and a false discovery rate (column *p_{adj}*) of at least 0.05. Multiple testing correction of p-values (column *p_{val}*) was done according to Benjamini and Hochberg. Genes are annotated with the HUGO Gene Nomenclature Committee (HGNC) [98] symbol and identifier (ID), Ensembl ID as well as Entrez ID [99]. Furthermore the table stores gene description as well as GO annotation for corresponding biological processes (BP). The conducted statistical test is given in the column *Test (B vs. A)*. The five two sample tests included *BT474 vs. HCC1954*, *BT474 vs. BTR50*, *HCC1954+T vs. HCC1954*, *BTR50+T vs. BTR50*, and *BT474+T vs. BT474*. The column *baseMean_A* holds the base mean, i.e. the mean of the counts divided by the size factors, for the counts of condition A. An analogous definition holds for *baseMean_B*. The column *baseMean*

holds the mean of `baseMeanA` and `baseMeanB`. The FC, i.e. the ratio $(\text{baseMeanB}+1)/(\text{baseMeanA}+1)$ is stored in the column `foldChange`. Analogously, the \log_2 FC is stored in the column `log2FoldChange`.

(XLS)

S1 Fig. Expression data of the candidate genes in the transNOAH breast cancer trial. The boxplot displays expression (\log_2) data of the 54 candidate genes from seven patients treated with trastuzumab for one year in addition to neoadjuvant chemotherapy (red) and four patients treated with neoadjuvant chemotherapy only (gray). The patient data samples were selected from the transNOAH breast cancer trial (GEO series GSE50948).

(PDF)

S1 Variant Call Format File. SNP calls in HCC1954. The VCF (Variant Call Format) file format is used by GATK for representation of variant calls. This file stores the SNPs called in the HCC1954 sample.

(VCF)

S2 Variant Call Format File. SNP calls in BTR50. The VCF (Variant Call Format) file format is used by GATK for representation of variant calls. This file stores the SNPs called in the BTR50 sample.

(VCF)

S3 Variant Call Format File. SNP calls in BT474. The VCF (Variant Call Format) file format is used by GATK for representation of variant calls. This file stores the SNPs called in the BT474 sample.

(VCF)

Acknowledgments

We thank Andreas Leha, Claudia Pommerenke, Annalen Bleckmann and Michaela Bayerlová for fruitful discussions.

Author Contributions

Conceived and designed the experiments: TB DA. Performed the experiments: SW GSR FL. Analyzed the data: SvdH TB AC. Contributed reagents/materials/analysis tools: SW DA GSR MN. Wrote the paper: SvdH SW AC GSR TB. Management of data acquisition and involvement in discussions: MN. Differential expression analysis: SvdH TB. SNP calling: AC. Generation of RNA-Seq data: GSR FL. Cultivation and treatment of the cell lines: SW. RT-qPCR validation: SW. Literature research and data interpretation: SvdH.

References

1. Roskoski R Jr (2014) The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacol Res* 79:34–74. doi: [10.1016/j.phrs.2013.11.002](https://doi.org/10.1016/j.phrs.2013.11.002) PMID: [24269963](https://pubmed.ncbi.nlm.nih.gov/24269963/)
2. Heil J, Gondos A, Rauch G, Marmé F, Rom J, et al. (2012) Outcome analysis of patients with primary breast cancer initially treated at a certified academic breast unit. *Breast* 21(3):303–308. doi: [10.1016/j.breast.2012.01.009](https://doi.org/10.1016/j.breast.2012.01.009) PMID: [22310244](https://pubmed.ncbi.nlm.nih.gov/22310244/)
3. Jelovac D, Wolff AC (2012) The adjuvant treatment of HER2-positive breast cancer. *Curr Treat Options Oncol* 13(2):230–239. doi: [10.1007/s11864-012-0186-4](https://doi.org/10.1007/s11864-012-0186-4) PMID: [22410709](https://pubmed.ncbi.nlm.nih.gov/22410709/)
4. Valabrega G, Montemurro F, Aglietta M (2007) Trastuzumab: mechanism of action, resistance and future perspectives in HER2-overexpressing breast cancer. *Ann Oncol* 18(6):977–984. doi: [10.1093/annonc/mdl475](https://doi.org/10.1093/annonc/mdl475) PMID: [17229773](https://pubmed.ncbi.nlm.nih.gov/17229773/)

5. Arteaga CL, Sliwkowski MX, Osborne CK, Perez EA, Puglisi F, et al. (2012) Treatment of HER2-positive breast cancer: current status and future perspectives. *Nat Rev Clin Oncol* 9:16–32. doi: [10.1038/nrclinonc.2011.177](https://doi.org/10.1038/nrclinonc.2011.177)
6. O'Brien NA, Browne BC, Chow L, Wang Y, Ginther C, et al. (2010) Activated phosphoinositide 3-kinase/AKT signaling confers resistance to trastuzumab but not lapatinib. *Mol Cancer Ther* 9(6):1489–1502. doi: [10.1158/1535-7163.MCT-09-1171](https://doi.org/10.1158/1535-7163.MCT-09-1171) PMID: [20501798](https://pubmed.ncbi.nlm.nih.gov/20501798/)
7. Figueroa-Magalhães MC, Jelovac D, Connolly RM, Wolff AC (2014) Treatment of HER2-positive breast cancer. *The Breast* 23(2):128–136. doi: [10.1016/j.breast.2013.11.011](https://doi.org/10.1016/j.breast.2013.11.011) PMID: [24360619](https://pubmed.ncbi.nlm.nih.gov/24360619/)
8. Sherry ST, Ward M, Sirotkin K (1999) dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* 9(8):677–679. PMID: [10447503](https://pubmed.ncbi.nlm.nih.gov/10447503/)
9. Brookes AJ (1999) The essence of SNPs. *Gene* 234(2):177–186. doi: [10.1016/S0378-1119\(99\)00219-X](https://doi.org/10.1016/S0378-1119(99)00219-X) PMID: [10395891](https://pubmed.ncbi.nlm.nih.gov/10395891/)
10. Gray IC, Campbell DA, Spurr NK (2000) Single nucleotide polymorphisms as tools in human genetics. *Hum Mol Genet* 9(16):2403–2408. doi: [10.1093/hmg/9.16.2403](https://doi.org/10.1093/hmg/9.16.2403) PMID: [11005795](https://pubmed.ncbi.nlm.nih.gov/11005795/)
11. Kwok PY, Chen X (2003) Detection of single nucleotide polymorphisms. *Curr Issues Mol Biol* 5(2):43–60. PMID: [12793528](https://pubmed.ncbi.nlm.nih.gov/12793528/)
12. Quinn EM, Cormican P, Kenny EM, Hill M, Anney R, et al. (2013) Development of strategies for SNP detection in RNA-seq data: Application to lymphoblastoid cell lines and evaluation using 1000 genomes data. *PLoS ONE* 8(3):e58815. doi: [10.1371/journal.pone.0058815](https://doi.org/10.1371/journal.pone.0058815) PMID: [23555596](https://pubmed.ncbi.nlm.nih.gov/23555596/)
13. Henjes F, Bender C, von der Heyde S, Braun L, Mannsperger HA, et al. (2012) Strong EGFR signaling in cell line models of ERBB2-amplified breast cancer attenuates response towards ERBB2-targeting drugs. *Oncogenesis* 1:e16. doi: [10.1038/oncsis.2012.16](https://doi.org/10.1038/oncsis.2012.16) PMID: [23552733](https://pubmed.ncbi.nlm.nih.gov/23552733/)
14. Kataoka Y, Mukohara T, Shimada H, Saijo N, Hirai M, et al. (2010) Association between gain-of-function mutations in PIK3CA and resistance to HER2-targeted agents in HER2-amplified breast cancer cell lines. *Ann Oncol* 21(2):255–262. doi: [10.1093/annonc/mdp304](https://doi.org/10.1093/annonc/mdp304) PMID: [19633047](https://pubmed.ncbi.nlm.nih.gov/19633047/)
15. Dodt M, Roehr JT, Ahmed R, Dieterich C (2012) FLEXBAR—Flexible barcode and adapter processing for next-generation sequencing platforms. *Biology* 1(3):895–905. doi: [10.3390/biology1030895](https://doi.org/10.3390/biology1030895) PMID: [24832523](https://pubmed.ncbi.nlm.nih.gov/24832523/)
16. Harrow J, Denoeud F, Frankish A, Reymond A, Chen CK, et al. (2006) GENCODE: producing a reference annotation for ENCODE. *Genome Biol* 7 Suppl 1:S4.1–9. doi: [10.1186/gb-2006-7-s1-s4](https://doi.org/10.1186/gb-2006-7-s1-s4) PMID: [16925838](https://pubmed.ncbi.nlm.nih.gov/16925838/)
17. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. doi: [10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635) PMID: [23104886](https://pubmed.ncbi.nlm.nih.gov/23104886/)
18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
19. Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169. doi: [10.1093/bioinformatics/btu638](https://doi.org/10.1093/bioinformatics/btu638) PMID: [25260700](https://pubmed.ncbi.nlm.nih.gov/25260700/)
20. Flicek P, Amode MR, Barrell D, Beal K, Billis K, et al. (2014) Ensembl 2014. *Nucleic Acids Res* 42(D1):D749–D755. doi: [10.1093/nar/gkt1196](https://doi.org/10.1093/nar/gkt1196) PMID: [24316576](https://pubmed.ncbi.nlm.nih.gov/24316576/)
21. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498. doi: [10.1038/ng.806](https://doi.org/10.1038/ng.806) PMID: [21478889](https://pubmed.ncbi.nlm.nih.gov/21478889/)
22. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, et al. (2010) Deriving the consequences of genomic variants with the ensembl API and SNP effect predictor. *Bioinformatics* 26(16):2069–2070. doi: [10.1093/bioinformatics/btq330](https://doi.org/10.1093/bioinformatics/btq330) PMID: [20562413](https://pubmed.ncbi.nlm.nih.gov/20562413/)
23. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106. doi: [10.1186/gb-2010-11-10-r106](https://doi.org/10.1186/gb-2010-11-10-r106) PMID: [20979621](https://pubmed.ncbi.nlm.nih.gov/20979621/)
24. R Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R-project website: <http://www.R-project.org/> Accessed 2015 Jan 9.
25. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 57:289–300.
26. Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* 4(8):1184–1191. doi: [10.1038/nprot.2009.97](https://doi.org/10.1038/nprot.2009.97) PMID: [19617889](https://pubmed.ncbi.nlm.nih.gov/19617889/)
27. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) method. *Methods* 25(4):402–408. doi: [10.1006/meth.2001.1262](https://doi.org/10.1006/meth.2001.1262) PMID: [11846609](https://pubmed.ncbi.nlm.nih.gov/11846609/)

28. Fishman WH (1987) Clinical and biological significance of an isozyme tumor marker—PLAP. *Clin Biochem* 20(6):387–392. doi: [10.1016/0009-9120\(87\)90003-8](https://doi.org/10.1016/0009-9120(87)90003-8) PMID: [3325192](https://pubmed.ncbi.nlm.nih.gov/3325192/)
29. Ayari I, Fedeli U, Saguem S, Hidar S, Khlifi S, et al. (2013) Role of CYP1A2 polymorphisms in breast cancer risk in women. *Mol Med Rep* 7:280–286. doi: [10.3892/mmr.2012](https://doi.org/10.3892/mmr.2012) PMID: [23128882](https://pubmed.ncbi.nlm.nih.gov/23128882/)
30. Sekhar SC, Kasai T, Satoh A, Shigehiro T, Mizutani A, et al. (2013) Identification of caveolin-1 as a potential causative factor in the generation of trastuzumab resistance in breast cancer cells. *J Cancer* 4(5):391–401. doi: [10.7150/jca.6470](https://doi.org/10.7150/jca.6470) PMID: [23833684](https://pubmed.ncbi.nlm.nih.gov/23833684/)
31. Prat A, Bianchini G, Thomas M, Belousov A, Cheang MCU, et al. (2014) Research-based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study. *Clin Cancer Res* 20(2):511–521. doi: [10.1158/1078-0432.CCR-13-0239](https://doi.org/10.1158/1078-0432.CCR-13-0239) PMID: [24443618](https://pubmed.ncbi.nlm.nih.gov/24443618/)
32. Gianni L, Eiermann W, Semiglazov V, Manikhas A, Lluch A, et al. (2010) Neoadjuvant chemotherapy with trastuzumab followed by adjuvant trastuzumab versus neoadjuvant chemotherapy alone, in patients with HER2-positive locally advanced breast cancer (the NOAH trial): a randomised controlled superiority trial with a parallel HER2-negative cohort. *The Lancet* 375(9712):377–384. doi: [10.1016/S0140-6736\(09\)61964-4](https://doi.org/10.1016/S0140-6736(09)61964-4) PMID: [20113825](https://pubmed.ncbi.nlm.nih.gov/20113825/)
33. Knutson TP, Daniel AR, Fan D, Silverstein KA, Covington KR, et al. (2012) Phosphorylated and sumoylation-deficient progesterone receptors drive proliferative gene signatures during breast cancer progression. *Breast Cancer Res* 14(3):R95. doi: [10.1186/bcr3211](https://doi.org/10.1186/bcr3211) PMID: [22697792](https://pubmed.ncbi.nlm.nih.gov/22697792/)
34. Dokmanovic M, Shen Y, Bonacci TM, Hirsch DS, Wu WJ (2011) Trastuzumab regulates IGFBP-2 and IGFBP-3 to mediate growth inhibition: implications for the development of predictive biomarkers for trastuzumab resistance. *Mol Cancer Ther* 10(6):917–928. doi: [10.1158/1535-7163.MCT-10-0980](https://doi.org/10.1158/1535-7163.MCT-10-0980) PMID: [21487052](https://pubmed.ncbi.nlm.nih.gov/21487052/)
35. Jerome L, Alami N, Belanger S, Page V, Yu Q, et al. (2006) Recombinant human insulin-like growth factor binding protein 3 inhibits growth of human epidermal growth factor receptor-2-overexpressing breast tumors and potentiates herceptin activity in vivo. *Cancer Res* 66(14):7245–7252. doi: [10.1158/0008-5472.CAN-05-3555](https://doi.org/10.1158/0008-5472.CAN-05-3555) PMID: [16849573](https://pubmed.ncbi.nlm.nih.gov/16849573/)
36. Lu Y, Zi X, Zhao Y, Mascarenhas D, Pollak M (2001) Insulin-like growth factor-I receptor signaling and resistance to trastuzumab (herceptin). *J Natl Cancer Inst* 93(24):1852–1857. doi: [10.1093/jnci/93.24.1852](https://doi.org/10.1093/jnci/93.24.1852) PMID: [11752009](https://pubmed.ncbi.nlm.nih.gov/11752009/)
37. Baggolini M, Clark-Lewis I (1992) Interleukin-8, a chemotactic and inflammatory cytokine. *FEBS Lett* 307:97–101. doi: [10.1016/0014-5793\(92\)80909-Z](https://doi.org/10.1016/0014-5793(92)80909-Z) PMID: [1639201](https://pubmed.ncbi.nlm.nih.gov/1639201/)
38. Korkaya H, Kim GI, Davis A, Malik F, Henry NL, et al. (2012) Activation of an IL6 inflammatory loop mediates trastuzumab resistance in HER2+ breast cancer by expanding the cancer stem cell population. *Mol Cell* 47(4):570–584. doi: [10.1016/j.molcel.2012.06.014](https://doi.org/10.1016/j.molcel.2012.06.014) PMID: [22819326](https://pubmed.ncbi.nlm.nih.gov/22819326/)
39. Ristimäki A, Sivula A, Lundin J, Lundin M, Salminen T, et al. (2002) Prognostic significance of elevated cyclooxygenase-2 expression in breast cancer. *Cancer Res* 62(3):632–635. PMID: [11830510](https://pubmed.ncbi.nlm.nih.gov/11830510/)
40. Wang SC, Lien HC, Xia W, Chen IF, Lo HW, et al. (2004) Binding at and transactivation of the COX-2 promoter by nuclear tyrosine kinase receptor ErbB-2. *Cancer Cell* 6(3):251–261. doi: [10.1016/j.ccr.2004.07.012](https://doi.org/10.1016/j.ccr.2004.07.012) PMID: [15380516](https://pubmed.ncbi.nlm.nih.gov/15380516/)
41. Flowers M, Thompson PA (2009) t10c12 conjugated linoleic acid suppresses HER2 protein and enhances apoptosis in SKBr3 breast cancer cells: possible role of COX2. *PLoS ONE* 4(4):e5342. doi: [10.1371/journal.pone.0005342](https://doi.org/10.1371/journal.pone.0005342) PMID: [19399184](https://pubmed.ncbi.nlm.nih.gov/19399184/)
42. Benoit V, Relic B, Leval Xd Xd, Chariot A, Merville MP, et al. (2004) Regulation of HER-2 oncogene expression by cyclooxygenase-2 and prostaglandin E2. *Oncogene* 23(8):1631–1635. doi: [10.1038/sj.onc.1207295](https://doi.org/10.1038/sj.onc.1207295) PMID: [14985703](https://pubmed.ncbi.nlm.nih.gov/14985703/)
43. Howe LR, Subbaramaiah K, Brown AM, Dannenberg AJ (2001) Cyclooxygenase-2: a target for the prevention and treatment of breast cancer. *Endocr Relat Cancer* 8(2):97–114. doi: [10.1677/erc.0.0080097](https://doi.org/10.1677/erc.0.0080097) PMID: [11397667](https://pubmed.ncbi.nlm.nih.gov/11397667/)
44. Vadlamudi R, Mandal M, Adam L, Steinbach G, Mendelsohn J, et al. (1999) Regulation of cyclooxygenase-2 pathway by HER2 receptor. *Oncogene* 18(2):305–314. doi: [10.1038/sj.onc.1202307](https://doi.org/10.1038/sj.onc.1202307) PMID: [9927187](https://pubmed.ncbi.nlm.nih.gov/9927187/)
45. Olayioye MA, Neve RM, Lane HA, Hynes NE (2000) The ErbB signaling network: receptor heterodimerization in development and cancer. *EMBO J* 19(13):3159–3167. doi: [10.1093/emboj/19.13.3159](https://doi.org/10.1093/emboj/19.13.3159) PMID: [10880430](https://pubmed.ncbi.nlm.nih.gov/10880430/)
46. Basu GD, Pathangey LB, Tinder TL, Lagioia M, Gendler SJ, et al. (2004) Cyclooxygenase-2 inhibitor induces apoptosis in breast cancer cells in an in vivo model of spontaneous metastatic breast cancer. *Mol Cancer Res* 2(11):632–642. PMID: [15561779](https://pubmed.ncbi.nlm.nih.gov/15561779/)

47. Gianni L (2002) The future of targeted therapy: combining novel agents. *Oncology* 63 Suppl 1:47–56. doi: [10.1159/000066197](https://doi.org/10.1159/000066197) PMID: [12422055](https://pubmed.ncbi.nlm.nih.gov/12422055/)
48. Dang CT, Dannenberg AJ, Subbaramaiah K, Dickler MN, Moasser MM, et al. (2004) Phase II study of celecoxib and trastuzumab in metastatic breast cancer patients who have progressed after prior trastuzumab-based treatments. *Clin Cancer Res* 10(12 Pt 1):4062–4067. doi: [10.1158/1078-0432.CCR-03-0463](https://doi.org/10.1158/1078-0432.CCR-03-0463) PMID: [15217939](https://pubmed.ncbi.nlm.nih.gov/15217939/)
49. Mann M, Sheng H, Shao J, Williams CS, Pisacane PI, et al. (2001) Targeting cyclooxygenase 2 and HER-2/neu pathways inhibits colorectal carcinoma growth. *Gastroenterology* 120(7):1713–1719. doi: [10.1053/gast.2001.24844](https://doi.org/10.1053/gast.2001.24844) PMID: [11375952](https://pubmed.ncbi.nlm.nih.gov/11375952/)
50. Cho HY, Thomas S, Golden EB, Gaffney KJ, Hofman FM, et al. (2009) Enhanced killing of chemoresistant breast cancer cells via controlled aggravation of ER stress. *Cancer Lett* 282:87–97. doi: [10.1016/j.canlet.2009.03.007](https://doi.org/10.1016/j.canlet.2009.03.007) PMID: [19345476](https://pubmed.ncbi.nlm.nih.gov/19345476/)
51. Morrison LE, Jewell SS, Usha L, Blondin BA, Rao RD, et al. (2007) Effects of ERBB2 amplicon size and genomic alterations of chromosomes 1, 3, and 10 on patient response to trastuzumab in metastatic breast cancer. *Genes Chromosomes Cancer* 46(4):397–405. doi: [10.1002/gcc.20419](https://doi.org/10.1002/gcc.20419) PMID: [17243161](https://pubmed.ncbi.nlm.nih.gov/17243161/)
52. Joshi JP, Brown NE, Griner SE, Nahta R (2011) Growth differentiation factor 15 (GDF15)-mediated HER2 phosphorylation reduces trastuzumab sensitivity of HER2-overexpressing breast cancer cells. *Biochem Pharmacol* 82(9):1090–1099. doi: [10.1016/j.bcp.2011.07.082](https://doi.org/10.1016/j.bcp.2011.07.082) PMID: [21803025](https://pubmed.ncbi.nlm.nih.gov/21803025/)
53. Kumandan S, Mahadevan NR, Chiu K, DeLaney A, Zanetti M (2013) Activation of the unfolded protein response bypasses trastuzumab-mediated inhibition of the PI-3K pathway. *Cancer Lett* 329(2):236–242. doi: [10.1016/j.canlet.2012.11.014](https://doi.org/10.1016/j.canlet.2012.11.014) PMID: [23200669](https://pubmed.ncbi.nlm.nih.gov/23200669/)
54. Lin HH, Liao CJ, Lee YC, Hu KH, Meng HW, et al. (2011) Lipocalin-2-induced cytokine production enhances endometrial carcinoma cell survival and migration. *Int J Biol Sci* 7:74–86. doi: [10.7150/ijbs.7.74](https://doi.org/10.7150/ijbs.7.74) PMID: [21278918](https://pubmed.ncbi.nlm.nih.gov/21278918/)
55. Ji Q, Aoyama C, Nien YD, Liu PI, Chen PK, et al. (2004) Selective loss of AKR1C1 and AKR1C2 in breast cancer and their potential effect on progesterone signaling. *Cancer Res* 64(20):7610–7617. doi: [10.1158/0008-5472.CAN-04-1608](https://doi.org/10.1158/0008-5472.CAN-04-1608) PMID: [15492289](https://pubmed.ncbi.nlm.nih.gov/15492289/)
56. Janssen JWG, Cuny M, Orsetti B, Rodriguez C, Vallés H, et al. (2002) MYEOV: a candidate gene for DNA amplification events occurring centromeric to CCND1 in breast cancer. *Int J Cancer* 102(6):608–614. doi: [10.1002/ijc.10765](https://doi.org/10.1002/ijc.10765) PMID: [12448002](https://pubmed.ncbi.nlm.nih.gov/12448002/)
57. Axlund SD, Yoo BH, Rosen RB, Schaack J, Kabos P, et al. (2013) Progesterone-inducible cytokeratin 5-positive cells in luminal breast cancer exhibit progenitor properties. *Horm Cancer* 4:36–49. doi: [10.1007/s12672-012-0127-5](https://doi.org/10.1007/s12672-012-0127-5) PMID: [23184698](https://pubmed.ncbi.nlm.nih.gov/23184698/)
58. Abd El-Rehim DM, Pinder SE, Paish CE, Bell J, Blamey RW, et al. (2004) Expression of luminal and basal cytokeratins in human breast carcinoma. *J Pathol* 203(2):661–671. doi: [10.1002/path.1559](https://doi.org/10.1002/path.1559) PMID: [15141381](https://pubmed.ncbi.nlm.nih.gov/15141381/)
59. van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, et al. (2002) Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. *Am J Pathol* 161(6):1991–1996. doi: [10.1016/S0002-9440\(10\)64476-8](https://doi.org/10.1016/S0002-9440(10)64476-8) PMID: [12466114](https://pubmed.ncbi.nlm.nih.gov/12466114/)
60. Yousef GM, Scorilas A, Kyriakopoulou LG, Rendl L, Diamandis M, et al. (2002) Human kallikrein gene 5 (KLK5) expression by quantitative PCR: an independent indicator of poor prognosis in breast cancer. *Clin Chem* 48(8):1241–1250. PMID: [12142380](https://pubmed.ncbi.nlm.nih.gov/12142380/)
61. Luo LY, Diamandis EP, Look MP, Soosaipillai AP, Foekens JA (2002) Higher expression of human kallikrein 10 in breast cancer tissue predicts tamoxifen resistance. *Br J Cancer* 86(11):1790–1796. doi: [10.1038/sj.bjc.6600323](https://doi.org/10.1038/sj.bjc.6600323) PMID: [12087468](https://pubmed.ncbi.nlm.nih.gov/12087468/)
62. Yousef GM, Yacoub GM, Polymeris ME, Popalis C, Soosaipillai A, et al. (2004) Kallikrein gene downregulation in breast cancer. *Br J Cancer* 90:167–172. doi: [10.1038/sj.bjc.6601451](https://doi.org/10.1038/sj.bjc.6601451) PMID: [14710225](https://pubmed.ncbi.nlm.nih.gov/14710225/)
63. Li X, Yang Q, Yu H, Wu L, Zhao Y, et al. (2014) LIF promotes tumorigenesis and metastasis of breast cancer through the AKT-mTOR pathway. *Oncotarget* 5(3):788–801. PMID: [24553191](https://pubmed.ncbi.nlm.nih.gov/24553191/)
64. Agnihotri N, Kumar S, Mehta K (2013) Tissue transglutaminase as a central mediator in inflammation-induced progression of breast cancer. *Breast Cancer Res* 15:202. doi: [10.1186/bcr3371](https://doi.org/10.1186/bcr3371) PMID: [23673317](https://pubmed.ncbi.nlm.nih.gov/23673317/)
65. Chen PS, Wang MY, Wu SN, Su JL, Hong CC, et al. (2007) CTGF enhances the motility of breast cancer cells via an integrin- α v β 3-ERK1/2-dependent S100A4-upregulated pathway. *J Cell Sci* 120 (Pt 12):2053–2065. doi: [10.1242/jcs.03460](https://doi.org/10.1242/jcs.03460) PMID: [17550972](https://pubmed.ncbi.nlm.nih.gov/17550972/)
66. Gonçalves A, Charafe-Jauffret E, Bertucci F, Audebert S, Toiron Y, et al. (2008) Protein profiling of human breast tumor cells identifies novel biomarkers associated with molecular subtypes. *Mol Cell Proteomics* 7(8):1420–1433. doi: [10.1074/mcp.M700487-MCP200](https://doi.org/10.1074/mcp.M700487-MCP200) PMID: [18426791](https://pubmed.ncbi.nlm.nih.gov/18426791/)

67. Verma P, Ostermeyer-Fay AG, Brown DA (2010) Caveolin-1 induces formation of membrane tubules that sense actomyosin tension and are inhibited by polymerase I and transcript release factor/cavin-1. *Mol Biol Cell* 21(13):2226–2240. doi: [10.1091/mbc.E09-05-0417](https://doi.org/10.1091/mbc.E09-05-0417) PMID: [20427576](https://pubmed.ncbi.nlm.nih.gov/20427576/)
68. Di Cello F, Cope L, Li H, Jeschke J, Wang W, et al. (2013) Methylation of the claudin 1 promoter is associated with loss of expression in estrogen receptor positive breast cancer. *PLoS ONE* 8(7):e68630. doi: [10.1371/journal.pone.0068630](https://doi.org/10.1371/journal.pone.0068630) PMID: [23844228](https://pubmed.ncbi.nlm.nih.gov/23844228/)
69. Lu S, Singh K, Mangray S, Tavares R, Noble L, et al. (2013) Claudin expression in high-grade invasive ductal carcinoma of the breast: correlation with the molecular subtype. *Mod Pathol* 26(4):485–495. doi: [10.1038/modpathol.2012.187](https://doi.org/10.1038/modpathol.2012.187) PMID: [23222490](https://pubmed.ncbi.nlm.nih.gov/23222490/)
70. Lee YK, Lee SY, Park JR, Kim RJ, Kim SR, et al. (2012) Dysadherin expression promotes the motility and survival of human breast cancer cells by AKT activation. *Cancer Sci* 103(7):1280–1289. doi: [10.1111/j.1349-7006.2012.02302.x](https://doi.org/10.1111/j.1349-7006.2012.02302.x) PMID: [22494103](https://pubmed.ncbi.nlm.nih.gov/22494103/)
71. Zhao C, Putnik M, Gustafsson JA, Dahlman-Wright K (2009) Microarray analysis of altered gene expression in ERbeta-overexpressing HEK293 cells. *Endocrine* 36(2):224–232. doi: [10.1007/s12020-009-9233-8](https://doi.org/10.1007/s12020-009-9233-8) PMID: [19680825](https://pubmed.ncbi.nlm.nih.gov/19680825/)
72. Ahmed ARH, Griffiths AB, Tilby MT, Westley BR, May FEB (2012) TFF3 is a normal breast epithelial protein and is associated with differentiated phenotype in early breast cancer but predisposes to invasion and metastasis in advanced disease. *Am J Pathol* 180(3):904–916. doi: [10.1016/j.ajpath.2011.11.022](https://doi.org/10.1016/j.ajpath.2011.11.022) PMID: [22341453](https://pubmed.ncbi.nlm.nih.gov/22341453/)
73. Doane AS, Danso M, Lal P, Donaton M, Zhang L, et al. (2006) An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. *Oncogene* 25(28):3994–4008. doi: [10.1038/sj.onc.1209415](https://doi.org/10.1038/sj.onc.1209415) PMID: [16491124](https://pubmed.ncbi.nlm.nih.gov/16491124/)
74. Lacroix M (2006) Significance, detection and markers of disseminated breast cancer cells. *Endocr Relat Cancer* 13(4):1033–1067. doi: [10.1677/ERC-06-0001](https://doi.org/10.1677/ERC-06-0001) PMID: [17158753](https://pubmed.ncbi.nlm.nih.gov/17158753/)
75. Yu JC, Hsiung CN, Hsu HM, Bao BY, Chen ST, et al. (2011) Genetic variation in the genome-wide predicted estrogen response element-related sequences is associated with breast cancer development. *Breast Cancer Res* 13:R13. doi: [10.1186/bcr2821](https://doi.org/10.1186/bcr2821) PMID: [21281495](https://pubmed.ncbi.nlm.nih.gov/21281495/)
76. Srinivasan S, Nwachukwu JC, Parent AA, Cavett V, Nowak J, et al. (2013) Ligand-binding dynamics rewire cellular signaling via estrogen receptor- α . *Nat Chem Biol* 9(5):326–332. doi: [10.1038/nchembio.1214](https://doi.org/10.1038/nchembio.1214) PMID: [23524984](https://pubmed.ncbi.nlm.nih.gov/23524984/)
77. Liu M, Wang G, Gomez-Fernandez CR, Guo S (2012) GREB1 functions as a growth promoter and is modulated by IL6/STAT3 in breast cancer. *PLoS ONE* 7(10):e46410. doi: [10.1371/journal.pone.0046410](https://doi.org/10.1371/journal.pone.0046410) PMID: [23056300](https://pubmed.ncbi.nlm.nih.gov/23056300/)
78. Evans AH, Pancholi S, Farmer I, Thornhill A, Evans DB, et al. (2010) EGFR/HER2 inhibitor AEE788 increases ER-mediated transcription in HER2/ER-positive breast cancer cells but functions synergistically with endocrine therapy. *Br J Cancer* 102(8):1235–1243. doi: [10.1038/sj.bjc.6605641](https://doi.org/10.1038/sj.bjc.6605641) PMID: [20386540](https://pubmed.ncbi.nlm.nih.gov/20386540/)
79. Collins DC, Cocchiglia S, Tibbitts P, Solon G, Bane FT, et al. (2014) Growth factor receptor/steroid receptor cross talk in trastuzumab-treated breast cancer. *Oncogene* [Epub ahead of print] doi: [10.1038/onc.2013.586](https://doi.org/10.1038/onc.2013.586)
80. Einarsdóttir K, Rosenberg LU, Humphreys K, Bonnard C, Palmgren J, et al. (2006) Comprehensive analysis of the ATM, CHEK2 and ERBB2 genes in relation to breast tumour characteristics and survival: a population-based case-control and follow-up study. *Breast Cancer Res* 8(6):R67. doi: [10.1186/bcr1623](https://doi.org/10.1186/bcr1623) PMID: [17132159](https://pubmed.ncbi.nlm.nih.gov/17132159/)
81. Benusiglio PR, Lesueur F, Luccarini C, Conroy DM, Shah M, et al. (2005) Common ERBB2 polymorphisms and risk of breast cancer in a white British population: a case-control study. *Breast Cancer Res* 7(2):R204–209. doi: [10.1186/bcr982](https://doi.org/10.1186/bcr982) PMID: [15743501](https://pubmed.ncbi.nlm.nih.gov/15743501/)
82. Breyer JP, Sanders ME, Airey DC, Cai Q, Yaspan BL, et al. (2009) Heritable variation of ERBB2 and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 18(4):1252–1258. doi: [10.1158/1055-9965.EPI-08-1202](https://doi.org/10.1158/1055-9965.EPI-08-1202) PMID: [19336545](https://pubmed.ncbi.nlm.nih.gov/19336545/)
83. Festa-Vasconcellos JS, Piranda DN, Amaral LM, Indio-do Brasil V, Koifman S, et al. (2012) Polymorphisms in cyclooxygenase-2 gene and breast cancer prognosis: association between PTGS2 haplotypes and histopathological features. *Breast Cancer Res Treat* 132:251–258. doi: [10.1007/s10549-011-1828-0](https://doi.org/10.1007/s10549-011-1828-0) PMID: [22037828](https://pubmed.ncbi.nlm.nih.gov/22037828/)
84. Gerger A, Renner W, Langsenlehner T, Hofmann G, Knechtel G, et al. (2010) Association of interleukin-10 gene variation with breast cancer prognosis. *Breast Cancer Res Treat* 119(3):701–705. doi: [10.1007/s10549-009-0417-y](https://doi.org/10.1007/s10549-009-0417-y) PMID: [19437115](https://pubmed.ncbi.nlm.nih.gov/19437115/)
85. Abraham JE, Harrington P, Driver KE, Tyrer J, Easton DF, et al. (2009) Common polymorphisms in the prostaglandin pathway genes and their association with breast cancer susceptibility and survival. *Clin Cancer Res* 15(6):2181–2191. doi: [10.1158/1078-0432.CCR-08-0716](https://doi.org/10.1158/1078-0432.CCR-08-0716) PMID: [19276290](https://pubmed.ncbi.nlm.nih.gov/19276290/)

86. Yu KD, Chen AX, Yang C, Qiu LX, Fan L, et al. (2010) Current evidence on the relationship between polymorphisms in the COX-2 gene and breast cancer risk: a meta-analysis. *Breast Cancer Res Treat* 122:251–257. doi: [10.1007/s10549-010-0738-x](https://doi.org/10.1007/s10549-010-0738-x) PMID: [20033767](https://pubmed.ncbi.nlm.nih.gov/20033767/)
87. Zhu W, Wei Bb, Shan X, Liu P (2010) -765G>C and 8473T>C polymorphisms of COX-2 and cancer risk: a meta-analysis based on 33 case-control studies. *Mol Biol Rep* 37:277–288. doi: [10.1007/s11033-009-9685-1](https://doi.org/10.1007/s11033-009-9685-1) PMID: [19669667](https://pubmed.ncbi.nlm.nih.gov/19669667/)
88. Moore AE, Young LE, Dixon DA (2012) A common single-nucleotide polymorphism in cyclooxygenase-2 disrupts microRNA-mediated regulation. *Oncogene* 31(12):1592–1598. doi: [10.1038/onc.2011.349](https://doi.org/10.1038/onc.2011.349) PMID: [21822307](https://pubmed.ncbi.nlm.nih.gov/21822307/)
89. Jung JH, Chae YS, Moon JH, Kang BW, Kim JG, et al. (2010) TNF superfamily gene polymorphism as prognostic factor in early breast cancer. *J Cancer Res Clin Oncol* 136(5):685–694. doi: [10.1007/s00432-009-0707-0](https://doi.org/10.1007/s00432-009-0707-0) PMID: [19890662](https://pubmed.ncbi.nlm.nih.gov/19890662/)
90. Cox DG, Buring J, Hankinson SE, Hunter DJ (2007) A polymorphism in the 3' untranslated region of the gene encoding prostaglandin endoperoxide synthase 2 is not associated with an increase in breast cancer risk: a nested case-control study. *Breast Cancer Res* 9:R3. doi: [10.1186/bcr1635](https://doi.org/10.1186/bcr1635) PMID: [17214885](https://pubmed.ncbi.nlm.nih.gov/17214885/)
91. Qian B, Zheng H, Yu H, Chen K (2011) Genotypes and phenotypes of IGF-I and IGFBP-3 in breast tumors among Chinese women. *Breast Cancer Res Treat* 130:217–226. doi: [10.1007/s10549-011-1552-9](https://doi.org/10.1007/s10549-011-1552-9) PMID: [21562710](https://pubmed.ncbi.nlm.nih.gov/21562710/)
92. Su X, Colditz GA, Willett WC, Collins LC, Schnitt SJ, et al. (2010) Genetic variation and circulating levels of IGF-I and IGFBP-3 in relation to risk of proliferative benign breast disease. *Int J Cancer* 126:180–190. doi: [10.1002/ijc.24674](https://doi.org/10.1002/ijc.24674) PMID: [19551864](https://pubmed.ncbi.nlm.nih.gov/19551864/)
93. D'Aloisio AA, Schroeder JC, North KE, Poole C, West SL, et al. (2009) IGF-I and IGFBP-3 polymorphisms in relation to circulating levels among African American and Caucasian women. *Cancer Epidemiol Biomarkers Prev* 18(3):954–966. doi: [10.1158/1055-9965.EPI-08-0856](https://doi.org/10.1158/1055-9965.EPI-08-0856) PMID: [19240240](https://pubmed.ncbi.nlm.nih.gov/19240240/)
94. Patel AV, Cheng I, Canzian F, Le Marchand L, Thun MJ, et al. (2008) IGF-1, IGFBP-1, and IGFBP-3 polymorphisms predict circulating IGF levels but not breast cancer risk: findings from the breast and prostate cancer cohort consortium (BPC3). *PLoS ONE* 3(7):e2578. doi: [10.1371/journal.pone.0002578](https://doi.org/10.1371/journal.pone.0002578) PMID: [18596909](https://pubmed.ncbi.nlm.nih.gov/18596909/)
95. Cheng I, DeLellis Henderson K, Haiman CA, Kolonel LN, Henderson BE, et al. (2007) Genetic determinants of circulating insulin-like growth factor (IGF)-I, IGF binding protein (BP)-1, and IGFBP-3 levels in a multiethnic population. *J Clin Endocrinol Metab* 92(9):3660–3666. doi: [10.1210/jc.2007-0790](https://doi.org/10.1210/jc.2007-0790) PMID: [17566087](https://pubmed.ncbi.nlm.nih.gov/17566087/)
96. Tamimi RM, Cox DG, Kraft P, Pollak MN, Haiman CA, et al. (2007) Common genetic variation in IGF1, IGFBP-1, and IGFBP-3 in relation to mammographic density: a cross-sectional study. *Breast Cancer Res* 9:R18. doi: [10.1186/bcr1655](https://doi.org/10.1186/bcr1655) PMID: [17300730](https://pubmed.ncbi.nlm.nih.gov/17300730/)
97. Tong L, Yang XX, Liu MF, Yao GY, Dong JY, et al. (2012) Mutational analysis of key EGFR pathway genes in Chinese breast cancer patients. *Asian Pac J Cancer Prev* 13(11):5599–5603. doi: [10.7314/APJCP.2012.13.11.5599](https://doi.org/10.7314/APJCP.2012.13.11.5599) PMID: [23317280](https://pubmed.ncbi.nlm.nih.gov/23317280/)
98. Gray KA, Daugherty LC, Gordon SM, Seal RL, Wright MW, et al. (2013) Genenames.org: the HGNC resources in 2013. *Nucleic Acids Res* 41(Database issue):D545–552. doi: [10.1093/nar/gks1066](https://doi.org/10.1093/nar/gks1066) PMID: [23161694](https://pubmed.ncbi.nlm.nih.gov/23161694/)
99. Maglott D, Ostell J, Pruitt KD, Tatusova T (2005) Entrez gene: gene-centered information at NCBI. *Nucleic Acids Res* 33(Database issue):D54–58. doi: [10.1093/nar/gki031](https://doi.org/10.1093/nar/gki031) PMID: [15608257](https://pubmed.ncbi.nlm.nih.gov/15608257/)