

**Phenocopy – A Strategy to Qualify Chemical
Compounds during Hit-to-Lead and/or
Lead Optimization**

Von der Fakultät Energie-, Verfahrens- und Biotechnik der
Universität Stuttgart zur Erlangung der Würde eines Doktors der
Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von
Patrick Baum
aus Mutlangen

Hauptberichter: Prof. Dr. Roland Kontermann

Mitberichter: Prof. Dr. Klaus Pfizenmaier

Tag der mündlichen Prüfung:

29.07.2010

Institut für Zellbiologie und Immunologie
Universität Stuttgart
2010

Table of contents

Table of contents	3
Abbreviations	6
Summary	8
Zusammenfassung	10
1. Introduction	12
1.1 Phenocopy strategy	13
1.2.1 RNA Interference (RNAi)	15
1.2.2 Kinase Inhibitors	18
1.3 Drug Development Process	21
1.4 Microarrays and Gene Expression Analysis	27
1.5 Transforming Growth Factor Beta (TGF-β)	32
1.6 Aim	36
2. Results	37
2.1 Phenocopy Platform	38
2.2 Modulator characterization	40
2.2.1 siRNA characterization.....	40
2.2.1.1 Transfection protocol.....	40
2.2.1.2 Control siRNA off-target profiling.....	42
2.2.1.3 Control siRNAs influence expression of different cytokines and MMP1.50	
2.2.1.4 Control siRNAs influence TNF α signaling	52
2.2.1.5 TGF- β R1 siRNA characterization	53
2.2.2 Kinase inhibitors.....	58
2.3 Phenocopy Experiment	66

2.3.1 Data normalization	66
2.3.2 TGF- β signature	71
2.3.3 Off-target signature	77
2.3.4 Molecular Function	84
2.3.5 Pathway Analysis	86
2.3.6 Wet Lab Validation.....	93
2.3.6.1 Cytotoxicity and Cell Death.....	93
2.3.6.2 Inflammation.....	99
2.3.7 Kinase Profiling	100
3. Discussion	105
3.1 NCE Ranking.....	106
3.2 Chemical genomic profiling	111
3.3 TGF-β Biology	113
3.4 siRNAs as modulators	117
3.4.1 Control siRNA characterization	117
3.4.2 TGF- β siRNAs.....	120
3.5 Conclusion	122
4. Methods	124
4.1 Wet laboratory experiments	125
4.1.1 Cell culture, NCE treatment and siRNA transfection	125
4.1.2 RNA extraction	126
4.1.3 Quantitative real time polymerase chain reaction (qRT-PCR).....	126
4.1.4 ELISA analysis of PAI-1, phospho Smad2/3, MMP1, IL8 and IL6.....	127
4.1.5 LDH release assay.....	129
4.1.6 Amplification, labeling and Beadchip hybridization of RNA samples.....	129
4.1.7 High content screen Cellomics.....	130
4.1.8 Caspase-3 Assay	130
4.1.9 <i>In vitro</i> kinase profiling	131
4.2 Data Analysis	131

4.2.1 Data processing.....	131
4.2.2 TGF- β signature (on-target signature).....	132
4.2.3 Off-target signature	132
4.2.4 Ingenuity Pathway Analysis and Gene Set Enrichment Analysis	135
Reference List	136
Danksagung	148
Erklärung	151
Lebenslauf	152

Abbreviations

°C	Degrees Celcius
ADME	Absorption, distribution, metabolism and excretion
AGC	Containing PKA, PKG , PKC families
ATP	Adenosine triphosphate
au	Arbitrary units
BI	Boehringer Ingelheim
CAMK	Calcium/calmodulin-dependent protein kinase
cDNA	complementary/copy DNA
CK1	Casein kinase 1
CMGC	Containing CDK, MAPK, GSK3, CLK families
conc	Concentration
COPD	Chronic obstructive pulmonary disease
Cpd	Compound
cRNA	copy RNA
CT	Cycle treshold
Ctrl	Control
d	Day(s)
DF	DharmaFECT
DMSO	Dimethylsulfoxid
DNA	Deoxyribonucleic acid
ELISA	Enzym-linked immunosorbent assay
EMA	European Medicines Agency
Ex	external
FCS	Fetal calf serum
FDA	Food and Drug Administration
FDR	False discovery rate
GEO	Gene Expression Omnibus
GSEA	Gene Set Enrichment Analysis
h	Hour(s)
H-t-L	Hit to lead
HTS	High-throughput screening
IC ₅₀	Half maximal inhibitory concentration
IND	Investigational New Drug
IPF	Idiopathic pulmonary fibrosis
I-Smad	Inhibitory Smad
KEGG	Encyclopedia of Genes and Genomes
l	Liter
LO	lead optimization
M	Molar
m	Meter
min	Minutes

Abbreviations

miRNA	micro RNA
mRNA	messenger RNA
MTD	Maximum tolerated dose
NCE	New chemical entity
NME	New molecular entity
PAI-1	Plasmingon activator inhibitor 1
PAMPS	Pathogen-associated molecular patterns
PK	Pharmacokinetic
PMSF	Phenylmethanesulfonylfluorid
POC	Percent of control
Prefix m	milli = 10^{-3}
Prefix μ	mikro = 10^{-6}
Prefix n	nano = 10^{-9}
Prefix p	piko = 10^{-12}
qRT-PCR	Quantitative real time polymerase chain reaction
RISC	RNA induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
R-Smad	Regulated Smad
SARA	Smad anchor for receptor activation
siRNA	small interfering RNA
Smad	Similar to MAD
Smurf	Smad ubiquitination regulatory factor
STE	Homologs of yeast Sterile 7, Sterile 11, Sterile 20 kinases
TGF- β	Transforming growth factor beta
TGF- β R1	Transforming growth factor beta receptor 1
TK	Tyrosine kinase
TKL	Tyrosine kinase-like
TLR	Toll-like receptor
TR	Transfection reagent
UT	Untreated
UTP	Uridine triphosphate
wotgf	Without TGF- β stimulation

Summary

A phenocopy is defined as an environmentally induced phenotype of one individual which is identical to the genotype-determined phenotype of another individual. In the present work, the phenocopy phenomenon has been translated to the drug discovery process as phenotypes produced by the treatment of cellular systems with small interfering RNAs (siRNAs) or new chemical entities (NCE) may resemble environmentally induced phenotypic modifications. Various new chemical entities exerting inhibition of the kinase activity of Transforming Growth Factor Beta Receptor I (TGF- β R1) were ranked by high-throughput RNA expression profiling. This chemical genomics approach was able to unravel both on-target effects (effects, caused by the inhibition of the drug target) and off-target effects (effects, caused by the interaction of the NCE with additional molecules). It resulted in a precise time-dependent insight into the TGF- β biology (referred to as on-target signature) and allowed furthermore a comprehensive analysis of each NCE's off-target effects (referred to as off-target signatures). Both signature types can support the drug discover process. The on-target signature helps to characterize the mode of action of the drug target (TGF- β R1) and thereby supports the target validation as well as the assay development process. Furthermore, the evaluation of off-target effects by the Phenocopy approach allows a more accurate and integrated view on the mode of action of the compounds, supplementing classical biological evaluation parameters such as potency and selectivity. The presented proof of concept study allowed the ranking of NCEs that were before indistinguishable solely based on potency and selectivity. According to the newly introduced criteria, several of the tested NCEs revealed liabili-

Summary

ties at e.g. the induction of off-target effects and of induction of gene regulation inverse to the desired TGF- β inhibition effect, at the induction of cell death, at acting as pro-inflammatory stimuli and as promoting cellular growth and at induction of cancer pathways. Ultimately, this approach has therefore the potential to become a novel method for ranking compounds during various drug discovery phases.

Zusammenfassung

Eine Phänokopie ist als ein Individuum definiert, dessen Phänotyp durch einen Umwelteinfluss mit dem eines anderen Individuums identisch ist, dessen Phänotyp durch seinen Genotyp bestimmt ist. In der vorliegenden Arbeit wurde das Phänokopiephänomen auf den Wirkstoffentwicklungsprozess übertragen. Hierbei wurden die speziellen Phänotypen in einem Zellsystem durch die Behandlung mit „new chemical entities“ (NCE), potentiellen neuen Wirkstoffkandidaten, oder mittels „small interfering RNAs“ (siRNAs) induziert. Somit können beide Molekülklassen als externe Umweltbedingung angesehen werden. Hierbei wurden diverse der potentiellen Wirkstoffkandidaten zur Inhibition der Kinaseaktivität von Transforming Growth Factor Beta Receptor I (TGF- β R1) mittels Hochdurchsatz-Expressionsprofilierung klassifiziert. Dieser „chemical genomics“ Ansatz war in der Lage, sowohl die On-target Effekte (Effekte, welche durch die Inhibition des Wirkstofftargets ausgelöst werden), als auch die Off-target Effekte (Effekte, welche durch die Interaktion des NCEs mit zusätzlichen Molekülen ausgelöst werden) zu identifizieren. Weiterhin ermöglichte er präzise, zeitlich aufgelöste Einblicke in die TGF- β Biologie (im Folgenden als On-target Signatur bezeichnet) und erlaubte eine umfassende Analyse der jeweiligen Off-target Effekte der Wirkstoffkandidaten (im Folgenden als Off-target Signatur bezeichnet). Beide Signaturtypen können zur Unterstützung des Wirkstoffentwicklungsprozesses herangezogen werden. Die On-target Signatur charakterisiert die Wirkungsweise des Wirkstoffziels (target) und unterstützt somit sowohl den „target validation“ Prozess, als auch den „assay development“ Prozess. Des Weiteren erlaubt das Ermitteln der Off-target Effekte durch den Phänokopieansatz einen prä-

zisen und ganzheitlichen Einblick in die Wirkungsweise der Wirkstoffkandidaten und ergänzt klassische, biologische Bewertungsparameter wie Wirksamkeit und Selektivität. Die hier präsentierte Machbarkeitsstudie ermöglicht die Klassifizierung von Wirkstoffkandidaten, die zuvor auf der Basis von Wirksamkeit und Selektivität nicht zu unterscheiden waren. Entsprechend der neu eingeführten Kriterien zeigen verschiedene der getesteten Wirkstoffkandidaten unterschiedliche Vorbelastungen, wie z.B. das Auslösen von Off-target Effekten und von Genregulation, welche gegenläufig zu dem gewünschten TGF- β Inhibitionseffekt verlaufen. Des Weiteren konnte gezeigt werden, dass manche NCEs Zelltod induzieren, als proinflammatorische Stimuli fungieren, das Zellwachstum steigern könnten und Krebsignalwege induzieren. Letztendlich hat dieser Ansatz somit das Potential eine neue Methode zur Klassifizierung von neuen Wirkstoffkandidaten, während verschiedenster Phasen des Wirkstoffentwicklungsprozesses, zu werden.

Ph.D. Thesis Patrick Baum

1. Introduction

Phenocopy – A Strategy to Qualify Chemical Compounds during Hit-to-Lead and Lead Optimization

1.1 Phenocopy strategy

A phenocopy is defined as an environmental induced, non-hereditary phenotype of one individual which is identical to the genotype-determined phenotype of another individual. In other words, the phenocopy induced by the environmental condition mimics the phenotype produced by a gene. For example, a phenocopy is observed in Himalayan rabbits which have a white colored coat along with a black tail, nose and ears when raised in moderate temperatures. However, when raised in colder climates, they develop phenotypically similar to genetically different black coated rabbits. The Himalayan rabbits exhibit black coloration of their coats, resembling the genetically encoded black rabbits. Hence, in colder climates the Himalayan rabbit is a phenocopy of the black rabbit¹. The phenocopy phenomenon can be translated and used for the drug discovery process through inhibiting a drug target with different functional modulation technologies and thereby mimicking a phenotype of interest. Inhibition can be achieved using RNA interference (RNAi), to knockdown a target, or by small molecule inhibitors (new chemical entities – NCEs), to block or inhibit the activity of the target. These modulators can be used as a particular environmental condition by treating *in vitro* cultured cells. Effects of the inhibition can be monitored by high-throughput RNA expression profiling and derived gene expression signatures represent either partial or exact phenocopies. Here, induced phenocopies consist of gene expression signatures caused by different pathway modulator treatments (siRNA and NCE). Subsequent analysis of the gene expression signatures will elucidate two critical issues for drug discovery. First, getting a deeper insight into a target's biology by identifying genes whose expression are transcriptionally altered af-

ter interfering with the target of interest, referred to as the on-target signature. Second, single observations for each modulator used can identify genes regulated independently of the target inhibition, referred to as the off-target signature. The on-target signature is independent of the used modulator and defines the biological mode of action of the target. In contrast, the off-target signature defines the mode of action for each modulator used, which has to be not necessarily limited to the inhibition of only the target. The obtained off-target signatures can subsequently be used to qualify the NCEs of interest. This strategy was applied to NCEs and siRNAs directed against the target TGF- β receptor 1 (Figure 1).

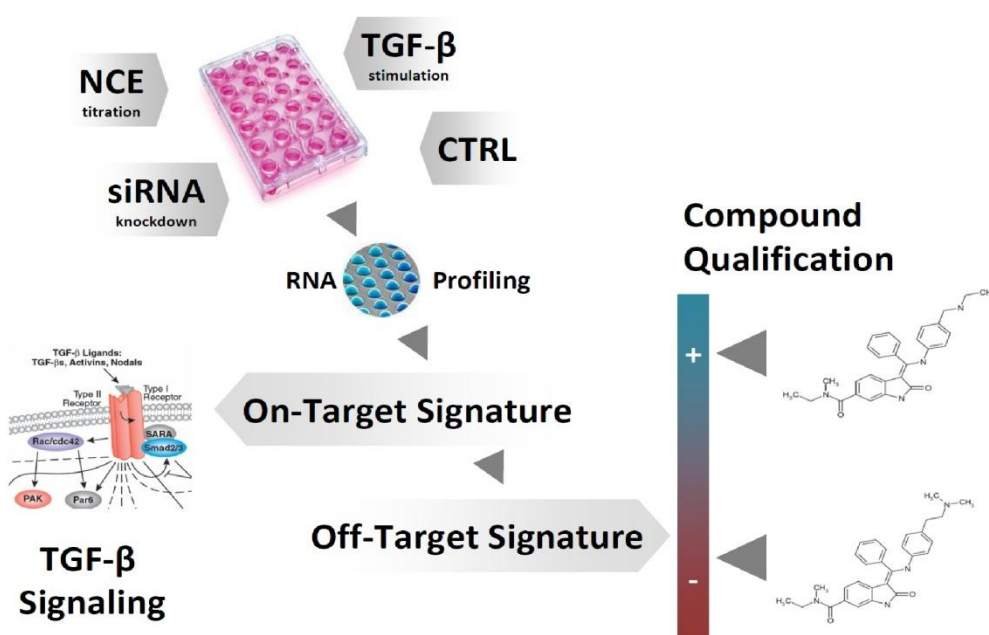


Figure 1 – Phenocopy workflow

In vitro cultured HaCaT cells stimulated with TGF- β were treated with NCEs inhibiting the kinase activity of TGF- β R1 or with a siRNA specific against TGF- β R1. After 2, 4 and 12 h total RNA was isolated for hybridization on Illumina Beadchips and expression profiles were generated. The concentration and time-dependent on-target (TGF- β signature) as well as the off-target signatures for every NCE were obtained by bioinformatic analysis. Compounds were qualified according to their off-target signatures by influencing other pathways.

1.2 Modulators – siRNAs & kinase inhibitors

There are different ways to modulate a signaling pathway. In the present study two functional modulation technologies were applied to either block the enzymatic activity of the receptor kinase by small molecular compounds (kinase inhibitors) or to perform a siRNA-mediated mRNA knockdown (RNAi) of the entire receptor. Pathway inhibition could also be achieved by the use of other external stimuli such as aptamers or blocking antibodies. Unfortunately, in the present case both molecule classes were not available for TGF- β receptor 1.

1.2.1 RNA Interference (RNAi)

The discovery of small interfering RNA molecules (siRNA) has been a mile stone in the field of molecular biology. RNA interference is not only used in basic research but is also efficiently applied in mammalian cells² for target identification and validation in pharmaceutical research and even finds its way into the clinic as third generation drug³. During the canonical RNAi pathway, siRNAs target complementary mRNAs for transcript cleavage and degradation in a process known as post-transcriptional gene silencing⁴. Endogenous siRNAs are generated from a long double-stranded RNA via digestion by the RNase Dicer. Resulting siRNAs are ~21-23 nucleotide duplexes with symmetric 2-3 nucleotides 3' overhangs and 5'-phosphate groups⁵. In contrast, synthetic siRNAs, initially unphosphorylated, are phosphorylated by hClp1 immediately after transfection into cells⁶. This phosphorylation is crucial for the subsequent incorporation of the siRNAs into a multi-component nuclease, the RNA-inducing silenc-

ing complex (RISC)⁷. During RISC assembly the double stranded siRNAs are unwinded and only one of the strands stably associates with the complex. Rules that govern selectivity of strand loading into RISC are based on differential thermodynamic stabilities of the ends of the siRNAs^{8,9}. The less thermodynamically stable end is favored for binding to RISC. The incorporated strand, called guide strand, directs target recognition by perfect or near-perfect Watson-Crick base pairing¹⁰ between the incorporated siRNA and the mRNA transcript¹¹ whereas the other strand, the passenger strand, of the RNA duplex is discarded. As part of RISC the endonuclease Ago-2 is responsible for the cleavage of the mRNA. The cleavage activity is very precise: the phosphodiester linkage between the nucleotides that are base-paired to siRNA residues 10 and 11 (relative to the 5' end of the siRNA) is cut to generate 5'-monophosphate and 3'-hydroxyl termini¹² leading to subsequent degradation through cellular exonucleases. Upon activation, RISC can undergo multiple rounds of cleavage resulting in a robust gene silencing.

However, the specificity of siRNA molecules has been a debate since the first description of RNAi as a technique for the down-regulation of a gene product. In the meantime, many studies on siRNA selectivity provided insight into the origin of siRNA regulations besides the intended reduction of the target sequence. These so called off-target effects are based on two distinct mechanisms: the siRNA sequence dependent and the double stranded RNA related off-target effects. The problem of the sequence dependent off-target effects is explained by either the binding of the passenger strand of the siRNA molecule to the inverted complementary sequence of an off-target transcript or by the low stringency binding of imperfect matches between the guide strand and the off-target transcript. Both events lead to undesired

transcript degradation after incorporation of the respective siRNA derived single stranded RNA into Ago proteins. The double stranded RNA related off-target effects can be summarized as an innate immune system response of the cell recognizing the presence of pathogen-associated molecular patterns (PAMPS) by host pattern recognition receptors¹³. The recognition is accomplished by cytoplasmatic, double stranded RNA binding proteins like PKR¹⁴ leading to the silencing of translation or by RIG-1 and MDA-5¹⁵, which leads to the activation of interferon regulator factor (IRF) or NFκB signaling. An additional group of membrane bound receptors also recognizes siRNA molecules either as double stranded RNA like TLR3 or by the presence of a ribose-backbone in close proximity to multiple uridine residues like TLR7¹⁶ or more specifically via GU-rich motifs like TLR7 and TLR8^{17, 18}. The recognition of PAMPS via these Toll-like receptors again activates the NFκB and interferon signaling cascades¹⁹ and markers of this activation have been described²⁰. Several strategies were developed to overcome these off-target effects¹⁸ by modifying the ribose backbone of the siRNA molecules. The chemical modifications avoid nuclease degradation, the incorporation of the passenger strand into RISC as well as the binding and activation of the PAMPS receptors. The strategies of chemical modifications include 2'-O-methyl modification of single²¹ or multiple sequences positions²², 2'-F-modifications²³ or even non-ribose backbones like locked nucleic acids²⁴ or arabinose²⁵ moieties.

1.2.2 Kinase Inhibitors

A new chemical entity (NCE), also referred to as new molecular entity (NME), is a novel pharmaceutical agent that does not contain an active chemical moiety previously approved by the United States Food and Drug Administration (FDA). This moiety is further defined as a molecule or ion, responsible for the physiological or pharmacological action of the drug substance. Thereby those appended portions of the molecule are excluded that cause the drug to be an ester, salt or other noncovalent derivative²⁶. The identification and characterization of these new molecular compounds represents one of the most important areas of pharmaceutical research in the pursuit of the next generation of therapeutic agents. Biological systems contain only four types of macromolecules that can be interfered using such a small chemical molecules: proteins, polysaccharides, lipids, and nucleic acids. However, the vast majority of successful drugs achieve their activity by binding to and modifying the activity of a protein. According to different estimations mainly based on the identification of druggable protein domains in the human genome there are between 2,000 and 3,000 gene products that can be treated by small chemical molecules^{27, 28}. The conservative estimation of Russ *et al.* assumes that with approximately 520, protein kinases represent with 22 % the biggest group of potential drug targets (Figure 2). Collectively, all these kinases, referred to as the human kinome, phosphorylate approximately 500,000 sites in the human proteome and are therefore involved in the regulation of virtually every cellular process²⁹.

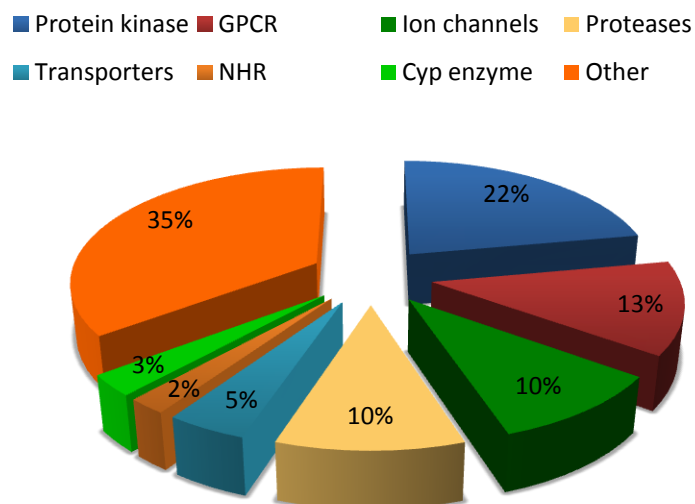


Figure 2 – Druggable genome

Conservative estimation of the druggable genome by Russ *et al.*²⁸ Abbreviation: GPCR: G-protein coupled receptor; NHR: nuclear hormone receptor; Cyp: cytochrome P450

Thus, kinases have become one of the most intensively investigated drug target classes with at least 30 targets being developed to a level of clinical phase I trial including 11 inhibitors approved for clinical use so far³⁰. Although kinase inhibition is predominantly discussed for the treatment of cancer, various other diseases, including immunological-, neurological-, vascular - and metabolic disorders, are linked to deregulation of kinase function^{30,31}. Kinase activity can be blocked by small chemical molecules in four different ways³⁰: i) Type I inhibitors. Kinases can adopt both active and inactive conformations. These molecules bind to both conformations in an ATP-mimetic manner to the ATP binding pocket of the kinase and thus prevent activity by blocking ATP binding (Figure 3). ii) Type II inhibitors. Kinase activity is blocked by preventing ATP binding and also by stabilizing the inactive conformation. This is achieved by the occupation of the ATP binding site and an adjacent hydrophobic pocket involved in the conformational change. iii) Covalent inhibitors. Comparable to

Type I inhibitors, they occupy the ATP binding pocket. However, once they are bound they form a covalent bond with an active site residue, usually a cysteine. iv) Allosteric inhibitors. These molecules are non-ATP-competitive and bind to a location in which binding modulates the catalytic activity of the kinase.

Collectively, due to the mode of action of these molecules and the partial highly preserved tertiary structure of kinases, it is often challenging to achieve a selective inhibition. Therefore, off-target effects of these inhibitors must be thoroughly investigated. Standard procedures to annotate kinase selectivity are *in vitro* enzymatic or binding assay counter screens against a panel of recombinant kinases and are available from a number of service providers. Discovering the full range of targets is also important for an alternative strategy for the use of kinase inhibitors that block the activity of more than one kinase to achieve a synergetic treatment effect. Such polypharmacology approaches seem to be a promising strategy for the treatment of disorders that tend to result from multiple molecular abnormalities such as cancer. So far, different of these promiscuous drugs have been approved. Among others, the most prominent examples are the anti-cancer drug Imatinib (Gleevec) and the schizophrenia drug Clozaril³².

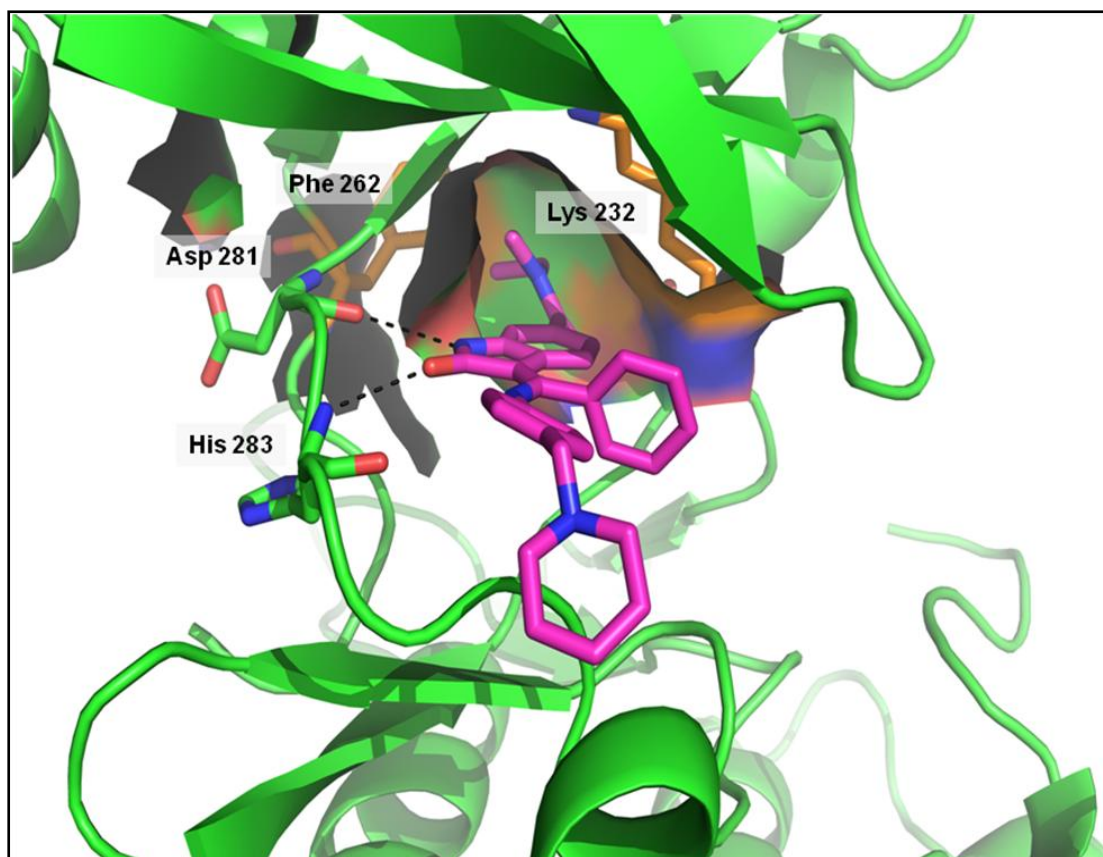


Figure 3 – Type I inhibitor

X-Ray structure of the Type I TGF β -R1 kinase inhibitor BI2 used in the present study. Typically for its chemotype, the indolinones, BI2 binds into the ATP pocket of the kinase and displays the canonical hydrogen bonds between the lactam moiety and Asp281 and His283 of the kinase hinge region. The 6-amido substituent on the Indolinone core points towards the TGF β -R1 specificity pocket flanked by Phe262 and Lys232³³.

1.3 Drug Development Process

The initial state of the drug development process is the identification of a potential therapeutic drug target whose modulation might inhibit or reverse disease progression. A wide variety of different approaches for target identification is used in both academic and pharmaceutical research, including high-throughput technologies like the profiling of changes on transcriptional³⁴ (genomics) or protein³⁵ (proteomics) levels or the use of (genome-wide) RNAi screens³⁶. Furthermore, reverse pharmacol-

ogy approaches are used as a more holistic approach. Here, phenotypes are screened after the treatment of an *in vitro* system with a compound library, desired phenotypes are selected before the hit drug target is identified³⁷. An alternative strategy that arrived in the “omics era” is data mining where mostly literature or microarray databases are screened for potential hits³⁸.

Subsequent to identification, target validation must demonstrate that the target is truly involved in the progress of the disease and that its modulation can result in a therapeutic effect. To unravel this therapeutic potential, extensive *in vitro* studies in a relevant cellular system as well as *in vivo* experiments using e.g. genetically modified organisms such as knock-out animal and/or disease related animal models need to be performed. Here, the biological contribution is characterized through activation, inhibition, overexpression or silencing of the target. A robust assay must be developed that can be used in high-throughput screening (HTS). Here, compound libraries are screened to identify substances that antagonize, agonize or modulate the drug target. These compounds are then classified as hits. The hits are further validated in counter screens before the best candidates will result in lead series (hit-to-lead phase) and finally enter the lead optimization (LO) process. Here, the structures of a lead class are chemically modified to first identify a suitable chemotype as lead class that is then further optimized in regards of potency, selectivity and pharmacokinetics of the molecules³⁹. However, very little information may be available that will give insights into the entire mode of action of the different hits that includes triggering of additional unwanted effects like toxicity. Selecting the best lead structures or compounds for optimization that are further promoted to pre-development candidates remains a challenging task.

For an optimal use of any drug it is pivotal to know its fate after administration to the body. Therefore, the pharmacokinetic (PK) and metabolism of the candidates must be determined. Four parameters (absorption, distribution, metabolism and excretion, referred to as ADME scheme) are mainly analyzed to identify what the organism does with the drug⁴⁰: i) Absorption, which is defined as the movement from the site of administration to the site of measurement (e.g. the bloodstream) in the body. ii) The distribution describes the process during which a drug moves from one site within the body to another. Drugs can either remain in the vascular system, get equally distributed in the body or are concentrated to specific organs or tissues mostly dependent on the molecular weight or solubility⁴¹. iii) Metabolism of the drug can be seen as a detoxification function of the body to xenobiotic molecules. Ideally, a drug is eliminated once its effect is no longer required. Although some drugs can be eliminated without structural changes, the vast majority must be metabolized to water soluble components to guarantee their excretion in bile or urine. A first group of enzymes (e.g. cytochrome P450 oxidases) carries out oxidation, reduction or hydrolytic reactions of the drug molecule (phase I reaction) before a second group (e.g. UDP-glucuronosyltransferases or glutathione S-transferases) introduces hydrophilic residues (phase II reaction)⁴². iv) Excretion is defined as the process whereby drugs or their metabolites are removed from the body. Hereby the kidney and the liver are major secretory organs. Excretion through saliva, sweat, tears, breast milk, hair and nails is also described⁴⁰ but only with minor contribution. Different experiments are performed to study ADME profiles, including absorption analysis on cells or artificial membranes⁴³ or protein binding of compounds⁴⁴, their stability in plasma or serum and their metabolism by e.g. screening their specific cytochrome P450 profile using

recombinant enzymes⁴⁵. Furthermore, LC-MS/MS based bio-analytical methods are applied to identify and quantify the compounds or their metabolites in blood, urine and other biological fluids or tissues⁴⁶.

Besides potency the drug candidates must also fulfill safety criteria. The therapeutic index compares both parameters: the dose necessary to produce a toxic effect with the dose sufficient to achieve efficacy. It is therefore a numerical measurement for both beneficial effect and relative safety of the drug candidate. The index is calculated upon preclinical *in vitro* and *in vivo* safety evaluation. While standard approaches to test *in vitro* cytotoxicity analyzes cell density, -integrity or -health parameters (e.g. ATP conc., lysosomal mass, mitochondrial membrane potential, nuclear fragmentation)^{47, 48}, *in vivo* toxicology involves studies in both rodent and non-rodent species and investigates toxic effects using well-proven markers such as histopathology, physiology- and blood-chemistry parameters⁴⁹. Additionally, toxicogenomics approaches are performed where RNA profiles of specific organs (mostly liver, kidney and heart) are analyzed after compound application, to mostly rodents, to screen for the induction of known toxicity related genes⁵⁰.

After positive non clinical safety assessment, the drug candidates can enter clinical trial as Investigational New Drug (IND) after acceptance of the regulatory authorities (FDA for US or EMA for Europe). During the clinical trials, the efficacy and safety/tolerability of a drug candidate is investigated in humans. The clinical trials are classified into up to five phases. In 2006, the FDA introduced the Exploratory Investigational New Drug guidance²⁶ that allows a so called phase 0 trial for exploratory first-in-man analyses. In this phase, also known as microdosing study, a subtherapeutic dose of the drug is administered to a small group of healthy subjects (10-15) to

investigate its pharmacodynamics and -kinetics. This allows to perform preliminary proof-of-concept studies at a very early stage and helps to reduce developmental costs⁵¹. In the next stage, phase I trials, the drug is tested in a small group of healthy subjects (20-80) to investigate its safety, and in depth pharmacodynamics and -kinetics. Phase I trials addresses dose escalation studies. Therefore a range of different drug concentration is administered to identify the maximum tolerated dose (MTD) and the suitable dose for future therapy. Subsequent phase II trials continue to test drug safety in a larger group of subjects (50-300) both healthy and diseased volunteers. Furthermore, the dose finding process must be completed and ultimately the effectiveness in disease treatment (proof-of-concept) is tested. After accomplishment of phase II goals the drug can be studied in phase III trial on a large group of patients (200-10,000). This so called pivotal trial needs to demonstrate the significant therapeutic effect of the drug compared to placebo treatment and ideally superior characteristics compared to current “gold standard” treatment of the disease⁵². Phase III trials are the most expensive and time consuming studies within the clinical trials and can last for up to several years, since they are often performed with large amounts of cohorts at different sites. After successful phase III trial, an application with all necessary information about the drug out of the data of preclinical and clinical studies can be submitted to the regulatory authorities (FDA and/or EMA) to achieve an approval to the market. Even after successful approval, the drug can be further investigate in a phase IV trial either to study rare adverse effects only detectable in a large collective of patients or for marketing reasons to find superior features of the treatment compared to competitor drug treatment.

However, the critical issue in drug development is the high attrition rate. The average success rate of a drug development program is approximately 11 %⁵³. However, the success rate differs strongly dependent on the therapeutic indication. While success rate for drugs treating cardiovascular disease is reasonably high with 20 %, only 5-15 % of all cancer programs succeed^{53, 54}. Additionally, the amount of FDA approvals decreased from 53 new drugs in 1996 to only 19 in 2009⁵⁵. Possible explanations are that recently diseases of great complexity are attacked but also that the entry bar for new drugs is higher than in the past because they are competing with enhanced standards of care. Furthermore, the demands of the regulatory authorities have increased over the past years⁵³. Responsible for most failures is the lack of efficacy of the treatment in man. Other drugs induce toxic adverse effects, fail due to chemistry and formulation issues, and have poor pharmacokinetic characteristics or their manufacturing process fails. Furthermore, a lack of clinical biomarkers prevents a distinct stratification of the patients or poor design of the clinical trials hampers identification of significant effects⁵⁶. Ultimately, there is a need for new strategies in all stages of drug development that optimizes the processes and supports the discovery of high quality drugs to cope with the challenges in a changing pharmaceutical industry. A summary of all stages of the drug development process is depicted in Figure 4.



Figure 4 – Drug development process

Drug development process can be divided in seven phases. It starts with the identification of the drug target whose modulation might inhibit or reverse disease progression. The target validation has to show that the target is causative of the disease symptoms and that modulation of the target ameliorates these symptoms. The assay development and high-throughput screening process have to find lead structures for the target modulation that are further improved during the lead optimization phase. Pharmacokinetic and metabolism investigate the fate of the drug after administration to an organism and animal safety assessment screens for toxic effects upon treatment. After successful assessment of all these phases the drug candidate can be investigated in clinical trials.

1.4 Microarrays and Gene Expression Analysis

Investigation and understanding of complex functional mechanism of a living organism requires a global and parallel analysis of different cellular processes. There are many possibilities to survey these cellular processes including mRNA levels, protein expression, epigenetic modifications or metabolite profiles⁵⁷. However, state-of-the-art method is the use of microarrays for the detection of transcriptional changes. The analysis of gene expression with microarrays is a well-established procedure⁵⁸ and various different technical approaches were developed. The most frequently used arrays were developed by Affymetrix⁵⁹ and Illumina⁶⁰. Affymetrix arrays are composed of 11 probes per gene that are 25 base pairs long. The probes are directly synthesized on the surface of the chip at a defined position by photolithography. In

contrast, Illumina arrays have 1-4 probes per gene that are 50 base pairs long. These probes are linked to beads that are then randomly arranged on the surface of the chip. Their location is subsequently decoded by a sequential hybridization procedure.

Over the past years microarray experiments have resulted in a big variety of gene expression compendia generated for different purposes. The first large-scale gene expression compendium was generated by Hughes et al.⁶¹ from 276 yeast deletion mutants, 11 conditional alleles and wild-type cells treated with 13 different drugs. Other efforts were made to generate a Global Cancer Map with samples of 218 human tumors out of 14 histological classes and 90 normal human tissues⁶². Furthermore, a large number of normal human and mouse tissue samples were combined to the Gene Expression Atlas⁶³. Additionally, several commercial databases such as Gene Logic's BioExpress have arisen, containing thousands of expression profiles from normal, diseased or drug-treated tissues or cell lines from different species. Ultimately, in a broader sense, ArrayExpress⁶⁴ and Gene Expression Omnibus (GEO)⁶⁵ can also be considered as huge compendia or repositories containing the stored data of more than 20,000 microarray experiments comprising more than 500,000 samples.

However, this almost vast quantity of data can be both boon and bane and the statistical analysis and most notably the interpretation of this data are great challenges. First, normalization of the data has to be performed to minimize systematic effects that are not constant between different samples of an experiment and that are not due to the factors under investigation (e.g. treatment, time). Optimal selection of a normalization method heavily depends on the nature of the experiment. In this re-

gard factors like comparability and quality of single runs play a major role. It has been shown that the normalization method used may influence further downstream analysis to a great extent⁶⁶, and thus, has to be carefully chosen based on the actual data.

Second, the selection criteria for differential expression of a gene have to be determined. Prior to selection an appropriate test statistic must be used to calculate p-values between the sample replicates. Commonly applied statistics are t-test, SAM-test or the U-Mann-Whitney-test⁶⁷. Afterwards additional statistical tests, using Bonferroni or Benjamini Hochberg procedures, are often applied to correct the p-values to reduce the amount of false positives⁶⁸. Then fold changes between the samples of interest must be determined. Finally, p-value and fold change cut-offs need to be determined, dependent on the conducted experiment and the biological system, to select differentially expressed genes. In order to avoid arbitrary cut-offs, other approaches use related samples to detect e.g. time or dose responses of a regulated gene^{69, 70}.

Ultimately, the critical and most challenging step in every expression analysis is to find a meaning in the orchestra of regulated genes. Each expression experiment regardless of which technology was used, normally results in a list of genes that are differentially expressed in one sample compared to another. Historically, each of the differentially expressed genes on this list was observed independently, relevant genes were assigned and a biological sense or meaning was deduced from this list, often without the use of prior knowledge. This time-consuming and partial low efficient approach can be bypassed through an incorporation of prior knowledge that helps to extract gene groups or gene expression signatures from the list of differen-

tially expressed genes. The use of gene expression signatures originally derived from early work on cancer classification from Golub *et al.*⁷¹. They profiled 38 bone-marrow samples, 27 from patients with acute lymphoblastic leukemia (ALL) and 11 from patients with acute myeloid leukemia (AML) and were able to predict the class of 34 other previously unclassified samples using a set of 50 genes known to be strongly correlated with either ALL or AML. Nowadays, a priori knowledge can be accessed through the use of various public or commercial databases (e.g. Biocarta⁷², Ingenuity Pathway Analysis⁷³, Kyoto Encyclopedia of Genes and Genomes⁷⁴, Wiki-Pathways⁷⁵, Gene Ontology, Reactome⁷⁶) developed by manual curation from literature, based on expert knowledge or assigned by computational annotation (Gene Ontology). Different algorithms can overlay the differentially expressed genes with biological pathways, networks or processes and are able to identify affected sets of genes with a sometimes small but coordinate change in expression. Additionally, the so called Gene Set Enrichment Analysis (GSEA)⁷⁷ introduces a rank based scoring metric that the magnitude of differential expression can also be included to the analysis.

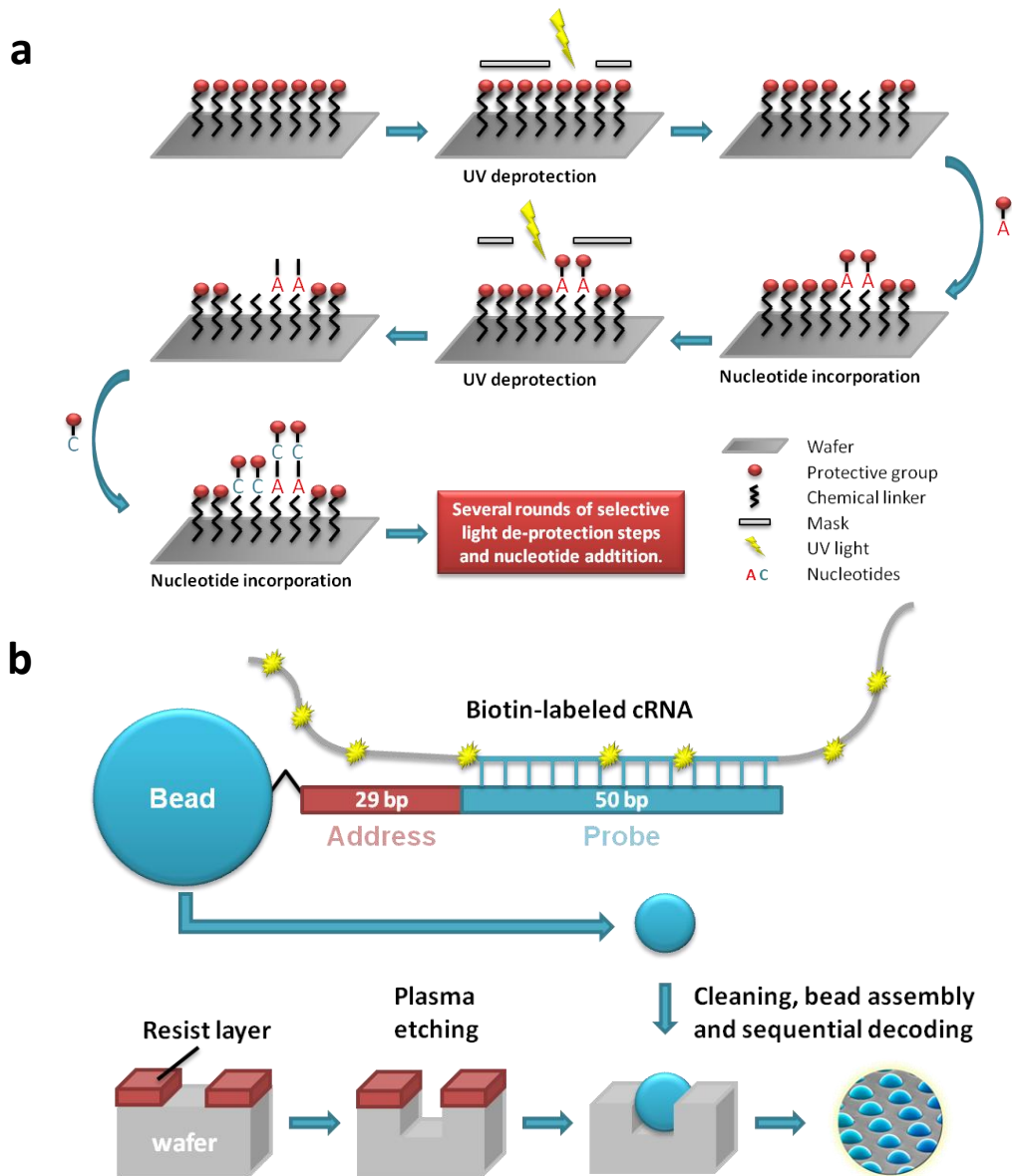


Figure 5 – Microarray manufacturing process

a) Affymetrix microarrays consist of 25mer probes that are chemically synthesized at specific locations on the surface of the silicon wafer. Synthesis is done in parallel, in a series of repetitive steps. Each step appends a particular nucleotide to selected regions of the chip. Selection occurs by exposure to UV light with the help of photolithographic masks. b) Illumina BeadChips are manufactured in four major steps. First, previously synthesized oligonucleotides, containing address and probe sequences, are linked to beads. Probe sequence is gene-specific for the binding of the transcribed and labeled cRNA. Always one species of oligonucleotide is covalently coupled to one bead, resulting in one bead type per surveyed gene. Second, the silicon wafer is coated with a micro-structured resist layer, then plasma etching is performed and subsequently the resist layer is removed. Third, the silicon wafer is flooded with the beads that self-assemble to a randomized microarray. Final, sequential decoding process⁷⁸, using decoder oligonucleotides complementary to the address sequence of the beads, is performed to map the probe position.

1.5 Transforming Growth Factor Beta (TGF- β)

TGF- β is a multifunctional cytokine with effects on cell growth, migration, adhesion, differentiation, apoptosis and epithel-to-mesenchymal transition⁷⁹. The canonical signaling pathway, as described by Shi and Massagué⁷⁹, consists of TGF- β receptor 1 and 2 (TGF- β R1 & R2), receptor serine/threonine protein kinases, and the family of Smad proteins. The ligands TGF- β 1, 2 and 3 but also Activin and Nodal initiate signaling through binding and subsequent formation of a receptor complex of TGF- β R1 and TGF- β R2. This allows TGF- β R2 phosphorylation of TGF- β R1, which then propagates the signal through phosphorylation of Smad2 and Smad3. Once phosphorylated, these regulated Smads (R-Smads) form heteromeric complexes with the Co-Smad, Smad4. This activated Smad complex enters the nucleus and in conjunction with other cofactors, such as CBP, TGIF and HDAC, interacts with the DNA and regulates gene expression. TGF- β signaling is regulated at several levels. First, the access of the R-Smads is controlled by the Smad anchor for receptor activation, SARA. Second, the E3 ubiquitin ligases, Smad ubiquitination regulatory factor-1 & 2 (Smurf-1 and 2), control the turnover of the TGF- β receptor and thereby the activation of Smads. Third, the inhibitory Smad (I-Smad), Smad7, acts antagonistically and inhibits receptor mediated activation of R-Smads. Finally, another level of control of the Smad pathway is via the regulation of nuclear accumulation of Smads, by the Ras-Extracellular signal kinase (ERK) pathway. There are also non-Smad mediated signaling events⁸⁰. For instance, TGF- β R1 can mediate JNK signaling by interacting with E3 ubiquitin ligase TRAF6 and subsequent activation of TAK1, MKK4 and 7 to trigger JNK-dependent apoptosis⁸¹.

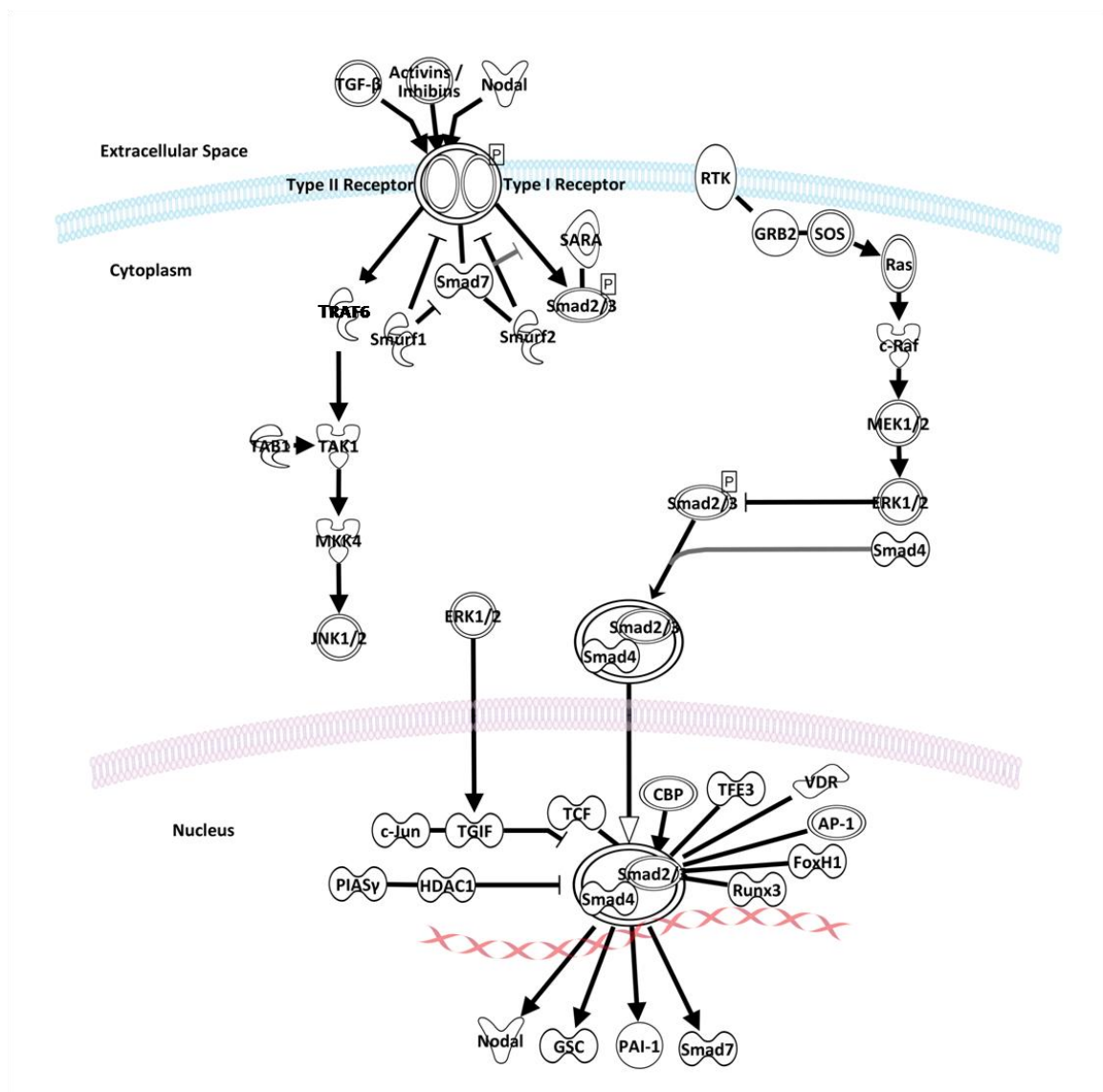


Figure 6 – Canonical TGF-β signaling pathway

Pathway illustration was taken from the Ingenuity Pathway Analysis (www.ingenuity.com) library of canonical signaling pathways.

Introduction

Malfunctions within the TGF- β signaling pathway may result in cancer, fibrosis and diverse hereditary disorders⁸²⁻⁸⁴. Therapeutic approaches to inhibit its signaling by targeting TGF- β R1 with antibodies, antisense molecules or kinase inhibitors⁸⁵ are widely discussed for the treatment of idiopathic pulmonary fibrosis (IPF) and solid tumor growth^{86, 87}. Lung diseases like asthma, chronic obstructive pulmonary disease (COPD) or idiopathic pulmonary fibrosis (IPF) are associated with an abnormal inflammatory response in combination with airway remodeling through fibrosis, goblet cell hyperplasia and smooth muscle thickening^{86, 88}. Among other growth factors and cytokines, TGF- β is highly expressed in fibrotic tissues and up-regulates the expression of adhesion molecules required for the recruitment of monocytes and neutrophils which both initiate inflammatory responses. Furthermore, TGF- β plays a pivotal role in the biosynthesis and turnover of extracellular matrix (ECM) proteins like collagens, fibronectin and proteoglycans and is thus contributing to fibrosis and stimulates smooth muscle cell proliferation⁸⁹. During cancerogenesis the function of TGF- β is Janus-faced. On the one hand, TGF- β acts as tumor suppressor by inhibiting the proliferation of normal epithelial, endothelial haematopoietic cells and early epithelial cancer cells. On the other hand, once tumorigenesis has been initiated tumor cells escape this growth control and produce high levels of TGF- β resulting in profound changes of the tumor's microenvironment in which TGF- β promotes tumor growth. The tumor-promoting effects include extracellular matrix degradation and epithelial mesenchymal transition by increased production of platelet-derived growth factor (PDGF), connective tissue growth factor (CTGF) and matrix metalloproteinases (MMP). Furthermore, angiogenesis is triggered by an up-regulation of vascular endothelial growth factor (VEGF)^{87, 90}. Clearly, inhibition of TGF- β R1 holds

promise as a new modality for the treatment of fibrotic diseases and cancer. Several small molecules targeting TGF- β R1 have been reported in literature⁸⁵ among them SB-431542, LY-2109761, SD-208, SM-16 and others. Recently, LY-2157299, a derivative of LY-2109761, was advanced into phase I clinical trials for the treatment of cancer⁹¹. Most known TGF- β R1 inhibitors are based on five-membered heterocyclic chemotypes (imidazoles, pyrazoles or thiazoles) occupying similar positions in the ATP pocket of the TGF- β R1 kinase domain.

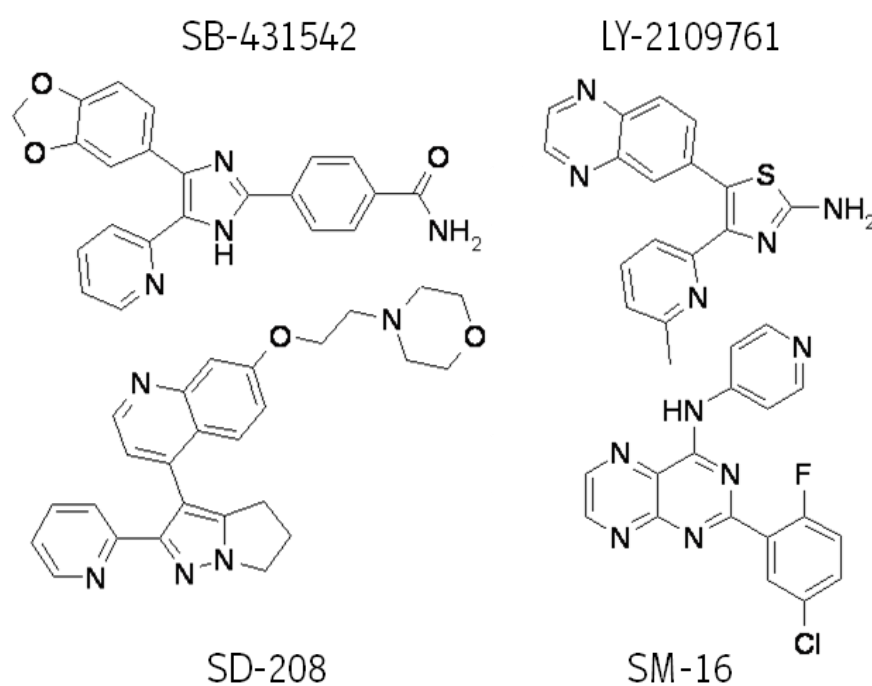


Figure 7 – TGF- β receptor 1 kinase inhibitors

1.6 Aim

The goal of the present work was to establish a procedure to support early phases in the drug development process. Therefore, the phenocopy phenomenon should be applied to an *in vitro* system by inhibiting TGF- β signaling with siRNAs and a novel structural lead class of TGF- β R1 kinase inhibitors and subsequent high-throughput expression profiling. This chemical genomics approach can then be used to elucidate the mode of action of TGF- β R1 as well as for the identification of the inhibitors' off-targets to allow a more accurate and integrated view on the optimized compounds, supplementing classical biological evaluation parameters such as potency and selectivity. The long term goal of such an early compound assessment is to reduce the attrition rate of advance clinical programs by an early discovery of liabilities of the analyzed NCEs.

Ph.D. Thesis Patrick Baum

2. Results

Phenocopy – A Strategy to Qualify Chemical Compounds during Hit-to-Lead and Lead Optimization

2.1 Phenocopy Platform

To perform the Phenocopy approach HaCaT cells (human keratinocytes) were cultured and treated to analyze TGF- β R1 modulators. First, it was important to demonstrate that the TGF- β signaling of the cells is functional and reproducible. Therefore, three readouts were carried out to cover the entire TGF- β signaling process that represent early, intermediate and late responses upon TGF- β stimulation. Direct downstream targets of the activated TGF- β R1 kinase are Smad2 and Smad3 proteins. Their phosphorylation initiates the intracellular signaling cascade. Smad phosphorylation was investigated by ELISA using an antibody specific for phosphorylated Smad2/3. This assay showed a significant increase of Smad2/3 phosphorylation already 15 minutes after stimulation with TGF- β . Phosphorylation was further enhanced after 30 and 60 minutes and remained stable for further 60 minutes (Figure 8a). A well characterized downstream target of TGF- β signaling is PAI-1 (SERPINE-1)⁹². The expression of PAI-1 at mRNA levels (qRT-PCR) and at protein levels (ELISA) was measured as surrogate marker for intermediate and late responses of TGF- β signal transduction. PAI-1 mRNA expression was TGF- β concentration- and time-dependent. An up to 70-fold up-regulation was detected 6 h after TGF- β stimulation. This expression decreased at later time points (Figure 8b). Subsequently, the supernatants were analyzed for PAI-1 protein expression. The expression of PAI-1 protein was delayed compared to the mRNA expression and can therefore be considered as a late response to TGF- β stimulation. The first significant increase was seen 6 h post stimulation and PAI-1 further accumulated at later time points (Figure 8c).

Results

After establishment and characterization of the cellular platform, it was possible to start the evaluation of the different pathway modulators.

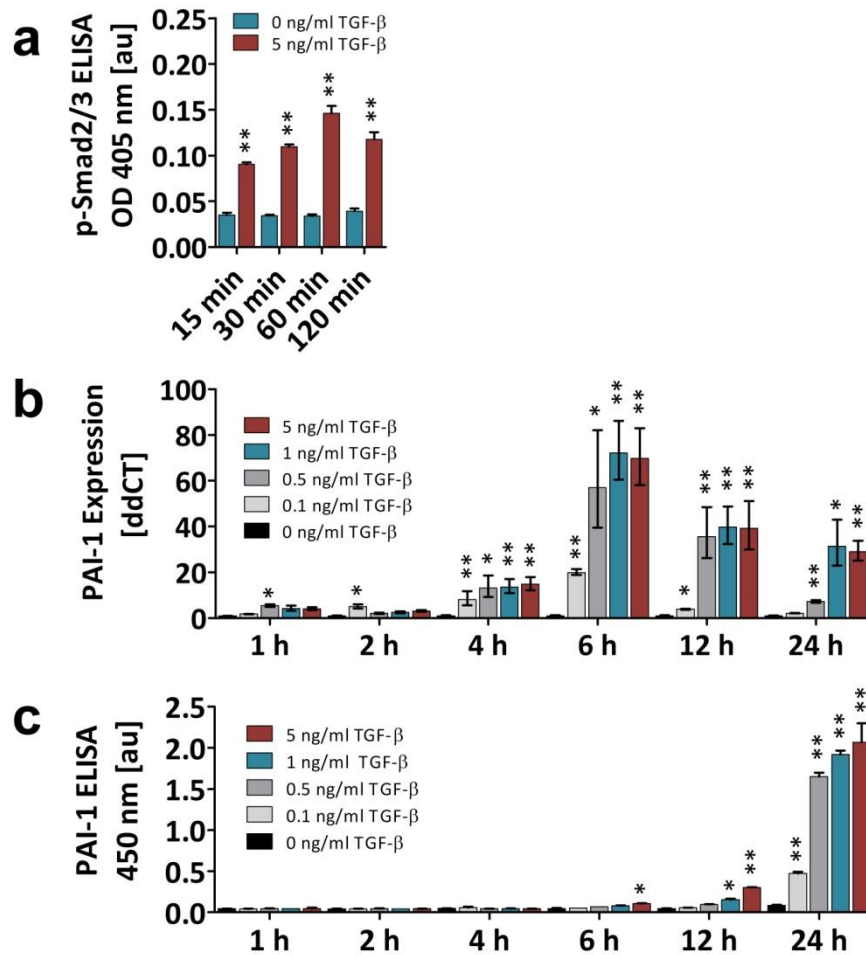


Figure 8 - Phenocopy platform read outs

Three readouts representing early (Smad2/3 phosphorylation), intermediate (PAI-1 mRNA) and late (PAI-1 protein) responses to TGF- β stimulation were performed. a: phospho-Smad2/3 ELISA. This assay showed a significant increase of Smad2/3 phosphorylation 15 minutes after stimulation with TGF- β . Phosphorylation was further enhanced after 30 and 60 minutes and remained stable for further 60 minutes. b: PAI-1 mRNA. Elevated PAI-1 expression was demonstrated by qRT-PCR after TGF- β stimulation in a concentration- and time-dependent manner. c: PAI-1 protein. The supernatants were analyzed with a PAI-1 ELISA for protein expression. The first significant increase was observed 6 h post stimulation. Subsequently, PAI-1 further accumulated in a concentration-dependent manner. All results are representative of three independent experiments. Student t-test was used to calculate the significance compared to unstimulated cells (* < 0.01 & ** < 0.001). All error bars indicate the standard deviation of $n=3$.

2.2 Modulator characterization

Two different functional modulation technologies were used to inhibit TGF- β signaling in HaCaT cells: RNAi and enzymatic kinase inhibition. siRNAs and seven NCEs (Table 8) were used to monitor and characterize mRNA transcriptional changes upon knockdown of TGF- β R1 mRNA or inhibition of TGF- β R1 kinase activity.

2.2.1 siRNA characterization

In the present study, siRNAs were used as alternative modulation technology and therefore as reference for the NCE treatment. However, the method as well as the treatment with siRNAs had to be adjusted to the cellular system of the Phenocopy platform. First, the protocol for the transfection of HaCaT cells was established, next an appropriate control molecule was identified and finally TGF- β R1 siRNAs were characterized and selected.

2.2.1.1 Transfection protocol

First, a protocol had to be chosen, suitable for an optimal transfection of HaCaT cells. Therefore, a well-characterized siRNA directed against GAPDH⁹³ was transfected using either lipofection or electroporation. Subsequently, knockdown efficacy was determined by qRT-PCR as surrogate for transfection efficacy. For lipofection four different transfection reagents from Dharmacon (DF1-DF4) were tested. Electroporation was carried out by Amaxa Nucleofector[®]. Lipofection using DF1 resulted in a knockdown of more than 95 %, but also the use of DF2 and DF4 led to knockdowns

Results

of approximately 90 % and comparable good results were also achieved by electroporation (Figure 9a). However, the lipofection protocol using DF1 was better tolerated by the HaCaT cells. LDH release to the supernatants of the cells was measured to analyze membrane integrity as measure for cell death after transfection. DF1 transfection resulted in less than 5 % cell death compared to the electroporation method with over 15 % dead cells (Figure 9b). Thus, HaCaT cells were transfected using DF1 in all further experiments due to slightly better transfection efficacy and less cell death compared to the other tested reagents or protocols.

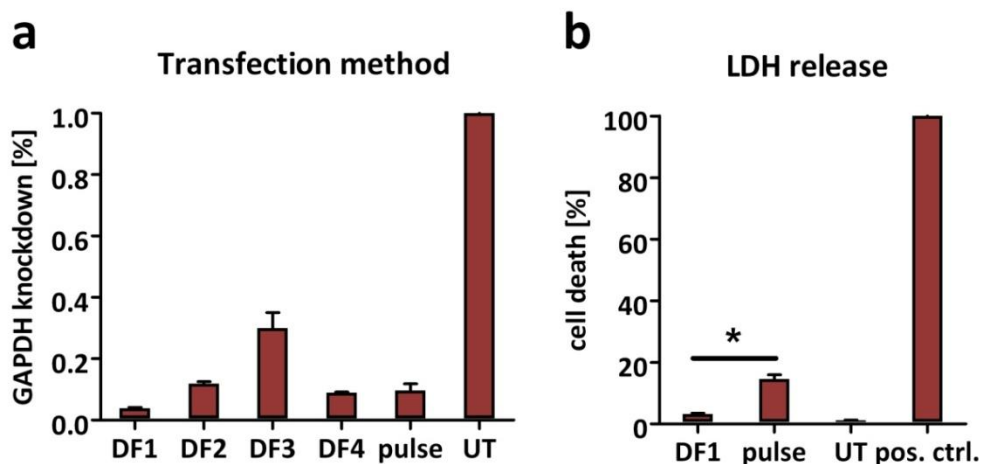


Figure 9 – Transfection method

a: HaCaT cells were transfected with a GAPDH-specific siRNA using either lipofection (DF1-DF4) or electroporation (pulse). Among the different lipofection reagents, highest knockdown was achieved by DF1 (> 95 %). Electroporation resulted in a knockdown of approximately 90 %. b: cell death caused by the transfection was analyzed through LDH activity in the supernatant of the cells. Fewer cells died after DF1 transfection (< 5 %) compared to the electroporation method (> 15 %). Student t-test was used to calculate the significance ($* < 0.01$). All error bars indicate the standard deviation of $n=2$.

2.2.1.2 Control siRNA off-target profiling

In general, control siRNA molecules, unable to mediate mRNA knockdown, are used for the normalization of specific siRNA effects. These molecules either consist of sequences that are not complimentary to any mRNA molecule (non-targeting siRNAs) or are not able to enter RISC (RISC-free siRNAs). In the most optimal situation, the use of these molecules should not introduce any gene regulation or phenotype. In order to find a control siRNA molecule suitable for a wide range of experiments the analysis was not only focused on the Phenocopy cell system. Therefore, the specificity of 13 control siRNA molecules from different vendors (Table 1) was compared by use of microarray expression profiling in two human cell lines HaCaT (keratinocyte phenotype) and HT1080 (fibroblast phenotype), as well as in the murine fibroblast line 3T3-L1. The control siRNA molecules were transfected using optimized conditions (high transfection efficacy without increased LDH release) in five independent experiments. All used control molecules are non-targeting siRNAs. The only exception is the siRNA D6ctrl representing the RISC-free control siRNA type. After 48 h the transcriptional changes introduced by these molecules were identified performing Illumina BeadChip based expression profiling. The profile of each treatment was compared with the respective untreated cell cultures. The differences in gene expression of HT1080 cultures are visualized as volcano plots in Figure 10. In the volcano plots, every point represents a single transcript. The x-axis shows the log₂ ratio (LR), representing the fold change, between transfected and untransfected cells. The y-axis is scaled as negative log₁₀ [p-value] as an indicator of significance. P-values were FDR-corrected according to Benjamini-Hochberg⁹⁴. Since the analysis is based

Results

on five replicates, it was possible to apply relative low fold change cut-offs and therefore regard all genes which revealed an expression difference of factor of 1.5 or higher with a FDR-corrected p-value of ≤ 0.01 as significantly regulated.

Table 1 – List of characterized control siRNA molecules

Vendor	Description	Cat. No.	Abbr.
Ambion	Silencer® Select Negative Control #1 siRNA	4390844	A1ctrl
Ambion	Silencer® Select Negative Control #2 siRNA	4390847	A2ctrl
Dharmacon	On-TARGET plus siControl Non-targeting Pool	D-001810-10-20	D1ctrl
Dharmacon	On-TARGET plus siControl Non-targeting siRNA	D-001810-01-20	D2ctrl
Dharmacon	On-TARGET plus siControl Non-targeting siRNA #2	D-001810-02-20	D3ctrl
Dharmacon	On-TARGET plus siControl Non-targeting siRNA #3	D-001810-03-20	D4ctrl
Dharmacon	On-TARGET plus siControl Non-targeting siRNA #4	D-001810-04-20	D5ctrl
Dharmacon	siControl RISC-free siRNA #1	D-001220-01-20	D6ctrl
Qiagen	Control siRNA_1248	customized	Q1ctrl
Qiagen	Random_2_siRNA	customized	Q2ctrl
Qiagen	Control siRNA_2904	customized	Q3ctrl
Qiagen	Random_1_siRNA	customized	Q4ctrl
Qiagen	Allstar Negative Control siRNA	1027281	Q5ctrl

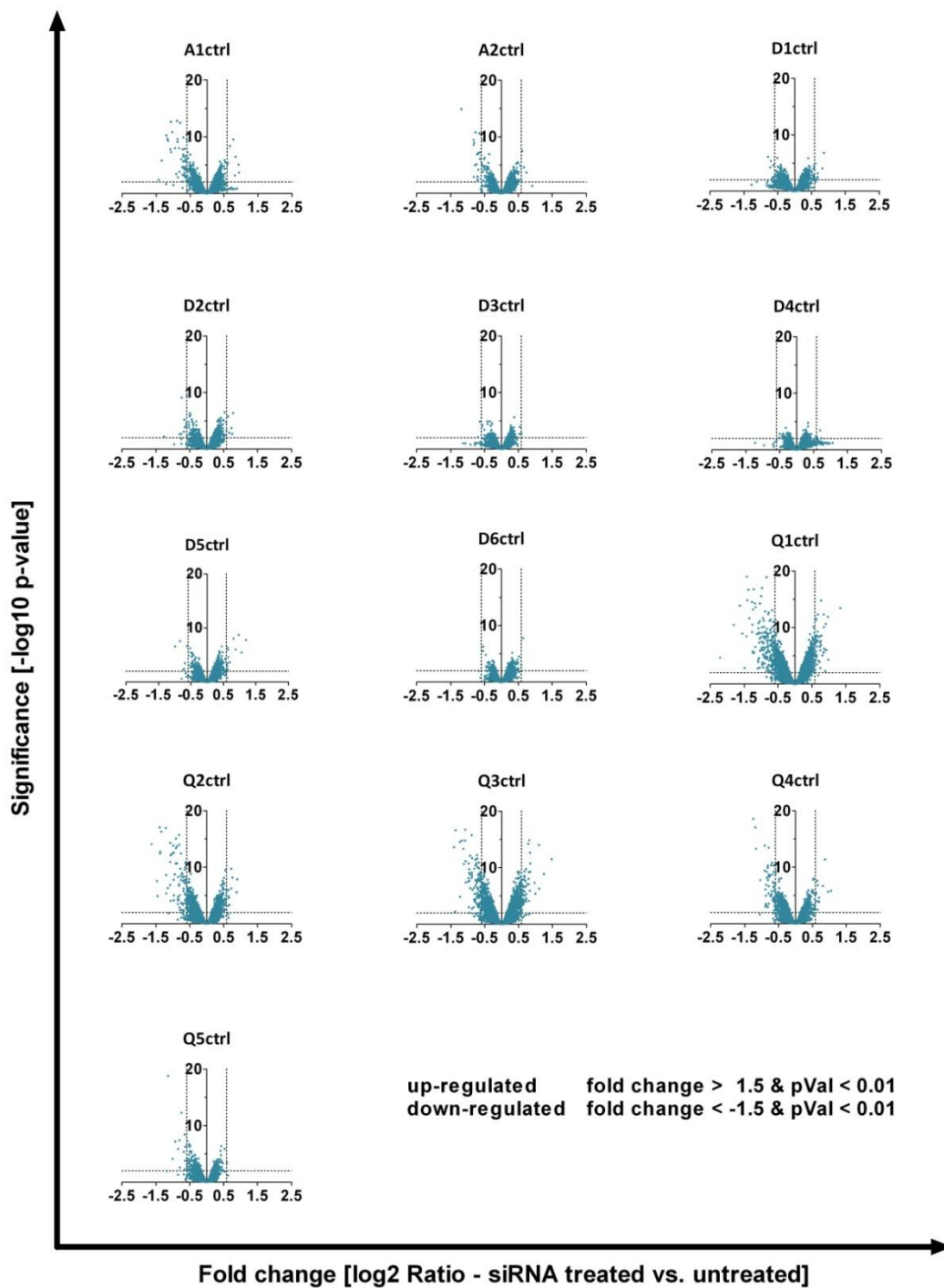


Figure 10 – Control siRNA off-targets in HT1080 cells

Shown are Log₂ Ratios (plotted on x-axes) and p-values (plotted on y-axes as negative Log₁₀ p-value) of control siRNA transfected versus untreated cells are shown. All genes with a fold change ≥ 1.5 and a p-value ≤ 0.01 were considered as off-target effects of the siRNA. Results represent five independent experiments. The p-values were FDR-corrected according to Benjamini-Hochberg.

Results

As quantitative specificity criteria for the control siRNA molecules the deregulated genes of each siRNA are summarized for all three cell lines (Table 2). Considering the number of differentially expressed genes the siRNA D6ctrl (RISC-free) shows the highest specificity among the 13 tested molecules with only four genes being significantly deregulated. Also only few differences were found for the treatment with the siRNAs D1ctrl (Non-targeting Pool), D3ctrl and D4ctrl. Here, less than 20 genes were altered compared to untransfected cells, whereas the control siRNAs A2ctrl, D5ctrl and Q5ctrl showed only moderate specificity with more than 20 differentially expressed genes. The lowest specificity exerted the siRNAs A1ctrl, Q2ctrl, Q4ctrl, Q1ctrl and Q3ctrl with the latter two showing several hundreds of differentially expressed genes compared to the controls. A summary of all genes that were significantly (p -value < 0.01) deregulated with a fold change larger than 1.5 within the three cell lines is listed in Table 2 and Supplement 1.

Table 2 - Number of control siRNA off-target transcripts in three different cell lines

<i>siRNA</i>	HT1080		HaCaT		3T3-L1		<i>sum</i>
	<i>down</i>	<i>up</i>	<i>down</i>	<i>up</i>	<i>down</i>	<i>up</i>	
D6ctrl	0	1	0	3	0	0	4
D3ctrl	2	1	0	4	0	0	7
D4ctrl	0	3	0	8	1	1	13
D1ctrl	8	7	0	2	0	0	17
D2ctrl	9	5	1	3	0	0	18
D5ctrl	6	13	0	3	0	0	22
Q5ctrl	12	1	10	2	0	0	25
A2ctrl	18	4	0	0	10	6	38
A1ctrl	45	8	5	1	17	2	78
Q4ctrl	54	13	15	17	6	2	107
Q2ctrl	81	10	47	8	9	0	155
Q3ctrl	112	55	60	54	50	26	357
Q1ctrl	194	58	60	40	110	9	471

Color code: blue indicates a small amount of control siRNA off-target effects, red indicates a high amount.

Results

Unspecific gene regulations induced by siRNA molecules can often be explained by partial sequence homologies with the respective mRNA off-targets. An important observation in this regard was that several of the gene regulations were recapitulated with other control siRNAs. Detailed analysis of the two human cell lines (HT1080 and HaCaT) identified 26 genes that are commonly deregulated by Q1ctrl and also 26 by Q2ctrl, 21 by Q3ctrl, 10 by Q4ctrl, 4 and 1 by A1ctrl and Q5ctrl (Supplement 1). However, only a small intersection is observed with the murine cell line 3T3-L1. Here, only the siRNA Q1ctrl down-regulated 3 genes (SPSB1, TOB1 and PPP1CC) in all three cell lines.

Most deregulated genes were found after transfection of HT1080 cells. Exemplary, for the effects siRNAs can have on cell lines, the analysis was focused on the effects of the control siRNAs in HT1080 cells. Hierarchical clustering of all 595 identified genes altered after treatment with any of the control molecules clearly illustrates different degrees of similarity among the used control molecules (Figure 11). All Dharmacon siRNAs have similar effects on gene expression and are all arranged in one sub cluster structure. The closest similarity of gene regulation is observed for the Non-targeting Pool D1ctrl and the siRNA D2ctrl. Treatment with one of the other control siRNAs resulted in more heterogeneous expression patterns. However, the control siRNA molecules A1ctrl and Q1ctrl are grouped in one subcluster due to their overlapping, multiple and strong effects on gene expression.

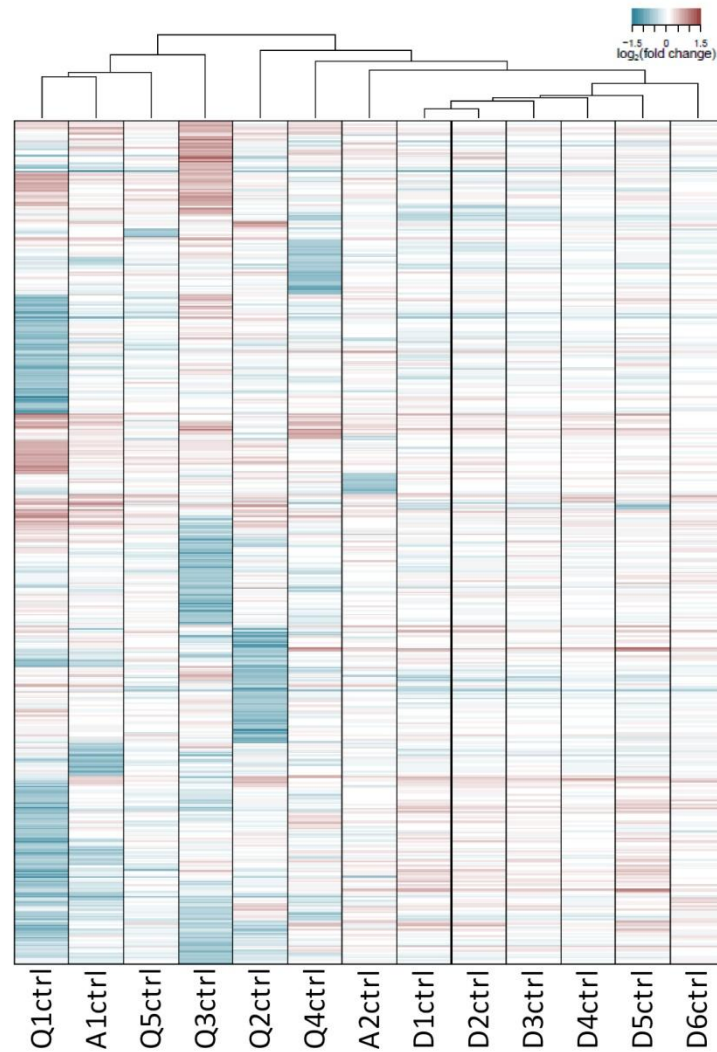


Figure 11 – Hierarchical clustering of control siRNA off-targets in HT1080 cells

Hierarchical clustering was performed based on all off-targets identified after expression profiling in HT1080 cells. The expression patterns of the different control siRNAs reveal several intersections in gene regulation. Blue indicates decreased expression relative to untreated cells, red indicates increased expression.

A detailed analysis on commonly deregulated genes revealed that 79 genes are differentially expressed ($FC > 1.5$, $p\text{-value} < 0.01$) after the treatment with two or more siRNAs. Interestingly, it was also possible to identify a group of genes that were commonly influenced by up to seven control siRNA molecules (Table 3).

Table 3 - List of Genes Altered by 2 or More siRNAs

Symbol	Ensembl Gene	Reg. by x siRNAs	siRNA	Regulation
IL24	ENSG00000181856, ENSG00000162892	7	A1, A2, Q1, Q2, Q3, Q4, Q5	down
TFRC	ENSG00000072274, ENSG00000163975	6	D2, D3, D5, Q2, Q4, Q5	down
FST	ENSG00000134363, ENSG00000125744	5	D1, D2, D5, Q1, Q4	down
IL1B	ENSG00000117480, ENSG00000125538	5	A1, A2, Q1, Q2, Q5	down
RGS4	ENSG00000115598, ENSG00000117152	5	A1, D1, D2, Q1, Q5	down
AMMECR1	ENSG00000101935, ENSG00000160957	3	D1, D3, Q2	down
CTGF	ENSG00000198898, ENSG00000118523	3	A1, Q1, Q2	down
ESM1	ENSG00000164283	3	A1, D2, Q1	down
NOX4	ENSG00000086991	3	A1, Q1, Q2	down
NUPR1_HUMAN	ENSG00000176046, ENSG00000180035	3	Q1, Q3, Q4	down
PNMA2	ENSG00000171362	3	Q1, Q3, Q4	down
RRM2	ENSG00000176076, ENSG00000171848	3	D1, D2, Q2	down
ABCA1	ENSG00000165029, ENSG00000129673	2	Q1, Q2	down
ADI1	ENSG00000182551	2	Q1, Q3	down
ATG2A	ENSG00000156802, ENSG00000110046	2	Q2, Q4	down
CBX6	ENSG00000183741, ENSG00000204149	2	D1, D5	down
CLIC3	ENSG00000169583, ENSG00000160199	2	Q1, Q3	down
DDAH1	ENSG00000153904, ENSG00000168172	2	A2, Q1	down
DPYSL2	ENSG00000131264, ENSG00000092964	2	Q1, Q4	down
DPYSL3	ENSG00000172352, ENSG00000113657	2	A1, Q3	down
DUSP4	ENSG00000120875	2	A1, Q1	down
E2F1	ENSG00000087237, ENSG00000101412	2	D2, Q4	down
ENPP1	ENSG00000197594	2	Q1, Q3	down
FLJ42986	ENSG00000196460	2	Q1, Q2	down
GALNT4	ENSG00000120322	2	A1, Q1	down
GDF15	ENSG00000130513	2	Q1, Q3	down
GFPT2	ENSG00000131459	2	Q1, Q3	down
H1FX	ENSG00000184897, ENSG00000011405	2	Q2, Q3	down
HAS2	ENSG00000134762, ENSG00000170961	2	A1, Q1	down
IL11	ENSG00000095752	2	Q1, Q2	down
IL6	ENSG00000151726, ENSG00000136244	2	Q1, Q3	down
INSIG1	ENSG00000137714, ENSG00000186480	2	Q2, Q4	down
JAG1	ENSG00000101384	2	A1, Q1	down
KLC1	ENSG00000126214	2	A1, Q2	down
MSRB3	ENSG00000174099	2	A1, Q3	down
NDRG1	ENSG00000104419, ENSG00000184779	2	Q1, Q3	down
PAPPA	ENSG00000182752	2	Q1, Q5	down
PDGFRB	ENSG00000113721	2	Q1, Q3	down
PRRT2	ENSG00000167371	2	Q1, Q3	down
PTPN1	ENSG00000196396	2	A1, Q1	down
PYGB	ENSG00000211598, ENSG00000100994	2	A1, Q1	down
RBPMS2	ENSG00000166831	2	Q1, Q3	down

Table 3 – continued

Symbol	Ensembl Gene	Reg. by x siRNAs	siRNA	Regulation
RDX	ENSG00000136634, ENSG00000137710	2	Q2, Q4	down
RNF144	ENSG00000151692, ENSG00000168118	2	A1, Q1	down
SBSN	ENSG00000189001	2	Q1, Q3	down
SCN2A	ENSG00000140859, ENSG00000136531	2	Q3, Q4	down
SLC2A3	ENSG00000059804, ENSG00000150457	2	Q2, Q4	down
SLC39A10	ENSG00000196950	2	Q2, Q3	down
SNAI2	ENSG00000019549	2	A1, Q1	down
SORBS2	ENSG00000154556	2	A1, Q1	down
SPIRE1	ENSG00000134278	2	A1, Q4	down
SPOCD1	ENSG00000134668	2	A1, Q1	down
SSFA2	ENSG00000138434	2	A1, Q1	down
STAG2	ENSG00000095539, ENSG00000101972	2	A1, Q3	down
THBS2	ENSG00000186340	2	Q1, Q3	down
UCN2	ENSG00000145040	2	Q1, Q2	down
UNG	ENSG00000076248	2	D2, Q4	down
VGF	ENSG00000128564	2	Q1, Q3	down
ZMAT3	ENSG00000172667	2	Q2, Q3	down
ZNF467	ENSG00000181444	2	Q1, Q3	down
DICER1	ENSG00000148655, ENSG00000100697	7	A1, A2, D1, D2, D5, Q1, Q4	up
PAEP	ENSG00000125944, ENSG00000122133	6	D1, D2, D3, D4, D5, Q4	up
COL22A1	ENSG00000169436	3	A2, D1, D5	up
HES4	ENSG00000188290	3	D1, D5, Q3	up
IL1A	ENSG00000115008	3	D2, Q3, Q4	up
KRT80	ENSG00000167767	3	A1, Q1, Q2	up
LIN28B	ENSG00000187772	3	A1, Q4, Q1	up
TGFBR3	ENSG00000152127, ENSG00000069702	3	D1, D5, Q2	up
ABCC3	ENSG00000108846, ENSG00000214570	2	A1, Q1	up
C2CD2	ENSG00000157617, ENSG00000104413	2	D2, D5	up
CALB2	ENSG00000172137, ENSG00000138363	2	Q1, Q4	up
CDC25A	ENSG00000137875, ENSG00000164045	2	Q1, Q3	up
EGR1	ENSG00000101204, ENSG00000120738	2	D5, Q4	up
GPR116	ENSG00000069122	2	D5, Q1	up
KIAA1324	ENSG00000116299	2	A1, Q1	up
MAFB	ENSG00000204103, ENSG00000117906	2	D1, D5	up
PHF17	ENSG00000077684	2	Q1, Q3	up
SKP2	ENSG00000168028, ENSG00000145604	2	Q1, Q3	up
STX3	ENSG00000099399, ENSG00000166900	2	Q1, Q3	up

2.2.1.3 Control siRNAs influence expression of different cytokines and MMP1

The gene expression analysis clearly indicated that the mRNA levels of several well-characterized genes are altered after the transfection of the control siRNAs. Especially siRNAs A1ctrl, Q1ctrl but also A2ctrl, Q2ctrl, Q3ctrl, Q4ctrl and Q5ctrl showed significant deregulations of cytokines like IL1 β or IL24. To further understand whether the observed expression differences correlate with differences in the respective signaling pathways the well-known influence of IL1 β on the gene expression of MMP1 as a marker for the status of the IL1 β signaling pathway⁹⁵ was observed. MMP1 levels in supernatants of control siRNA transfected cells were analyzed by ELISA and a strong MMP1 down-regulation was observed after A1ctrl (3.7-fold) and Q1ctrl (2.5-fold) transfection. Less significant effects on MMP1 levels were measured for the siRNAs A2ctrl, Q2ctrl, Q3ctrl and Q5ctrl. Again, the moderate effects on MMP1 release correlates with IL1 β and IL24 expression (Figure 12).

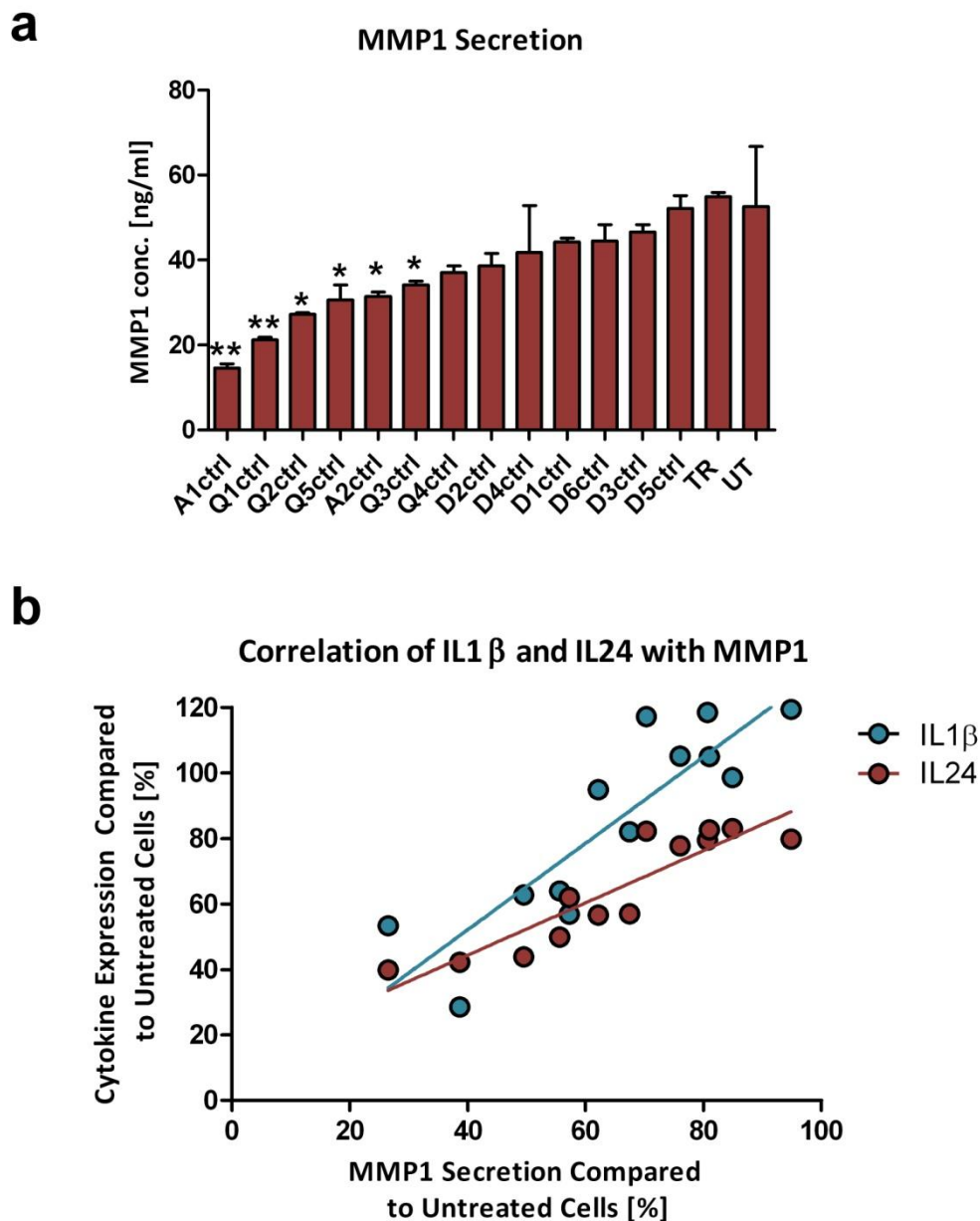


Figure 12 - Control siRNA-dependent MMP1 protein secretion

HT1080 cells were transfected with 13 different control siRNA molecules (A1-D5) a: 72 hours after the cells were transfected MMP1 levels in the supernatants were determined by ELISA. siRNAs A1ctrl, A2ctrl, Q1ctrl, Q2ctrl, Q3ctrl and Q5ctrl significantly decrease MMP1 levels. b: the protein secretion of MMP1 measured by ELISA correlates with the mRNA expression of IL1 β and IL24 measured by microarray. The results represent two independent MMP1 measurements performed in triplicates which are correlated to the expression profiling data of IL1 β and IL24. Student's t-test was used to calculate the significance compared to untreated cells (**< 0.0001, *< 0.01). All error bars indicate the standard deviation of n=3. (TR: transfection reagent, UT: untreated cells).

2.2.1.4 Control siRNAs influence TNF α signaling

Next, it was addressed whether the lack of specificity of the control siRNAs does not only reduce the basal cytokine expression levels or MMP1 secretion but also interferes with major cell signaling pathways. Since IL1 β and IL24 can both lead to NF κ B activation^{96, 97}, cells were stimulated with TNF α for activation of the IKK/NF κ B signaling cascade. IL8, a known target gene of NF κ B, showed a strong response upon TNF α stimulation (Fig. 4). Again ELISA analysis of supernatants derived from TNF α treated HT1080 cultures demonstrated a reduced sensitivity of IL8 release in the presence of the control siRNA molecules A1ctrl and Q1ctrl (Figure 13).

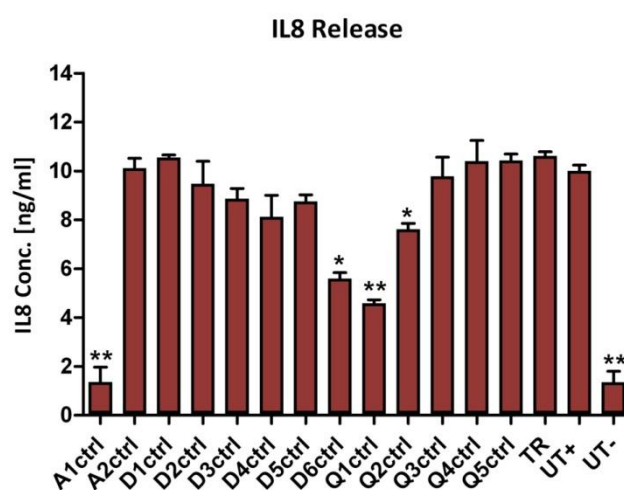


Figure 13 - Control siRNA-dependent IL8 release

HT1080 cells were transfected with 13 different control siRNA molecules (A1ctrl-D5ctrl). 72 hours after transfection cells were stimulated with 30 ng/ml TNF α for 8 hours. IL8 release was determined by ELISA of the cell supernatants. IL8 levels were 7.3 fold decreased after A1ctrl and 2.2 fold decreased after Q1ctrl treatment compared to untreated cells. The results are representative of two independent experiments performed in triplicates. Student's t-test was used to calculate the significance compared to untreated cells (**< 0.0001, *< 0.01). All error bars indicate the standard deviation of n=3. (TR: transfection reagent, UT+: untreated cells with TNF α stimulation, UT-: untreated cells without TNF α stimulation)

Results

In summary, the control siRNA D5ctrl was selected for further use in the Phenocopy experiment. D5ctrl was chosen due to its highest selectivity among all siRNAs in HaCaT cells. Hereby excluded were the siRNAs D1ctrl and D6ctrl although both siRNAs resulted in less off-target effects. However, both controls differ from the others. Within the Phenocopy experiment the goal was to use single siRNAs and not siRNA pools. Therefore, the use of siRNA D1ctrl, composed of four different siRNAs (SMARTpool), is not an appropriate control for this approach. Furthermore, a control molecule unable to enter RISC, such as D6ctrl, can also not be used to test the entire spectrum of a siRNA experiment since the RNAi machinery will not be affected through such a molecule.

2.2.1.5 TGF- β 1 siRNA characterization

Subsequent to the selection of a proper transfection protocol and the identification of a good control siRNA molecule, assessment of specific TGF- β 1 siRNAs was carried out. Criteria for a good candidate were a strong and stable mRNA knockdown, subsequent inhibition of the TGF- β signaling pathway and low off-target effects. To guarantee optimal siRNA-mediated TGF- β 1 knockdown, ten commercially available siRNAs were qualified (Table 4). First, knockdown efficacy was determined on mRNA level using qRT-PCR. Only five siRNAs (A1tgf, D1tgf, D2tgf, Q3tgf and Q4tgf) which led to a knockdown of more than 90 % were selected for off-target profiling (Figure 14a). Second, inhibition of downstream signaling of each selected siRNA was determined by phospho-Smad2/3 (Figure 14b) and PAI-1 ELISA (Figure 14c). Interestingly, although transfection of siRNA D1 resulted in the best mRNA knockdown (98 %), this

finding was not represented in the functional readouts. The strongest functional knockdown was observed for siRNA A1tgf.

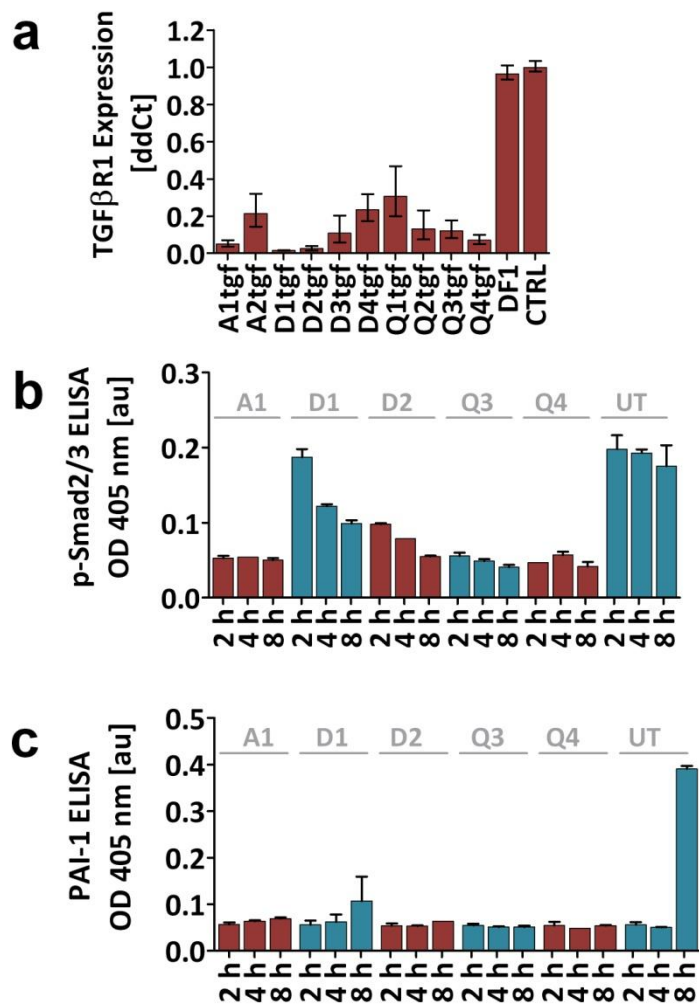


Figure 14 – TGF-β siRNA characterization

a: siRNA knockdown efficiency was measured by qRT-PCR 48 h post transfection. Ten different commercially available siRNAs (A - Ambion, D - Dharmacon & Q - Qiagen) were used. b and c: siRNAs with the best knockdown efficacy (A1tgf, D1tgf, D2tgf, Q3tgf & Q4tgf), as well as the untreated control (UT) were analyzed for functional blockade of TGF-β signaling determined by inhibition of p-Smad2/3 (b, p-Smad2/3 ELISA) or PAI-1 protein (c: PAI-1 ELISA). All error bars indicate the standard deviation of n=3.

Results

Finally, the off-target effects of all siRNAs were determined by microarray analysis using Illumina Beadchip technology. All deregulated genes (p-value < 0.01 and |LR| ≥ 1) were identified for the selected five siRNAs (Figure 15). To exclude genes from the off-target list that are relevant for the mechanism of the procedure or relevant for the TGF-β1 biology, only those genes were selected that were uniquely deregulated by the respective siRNA. The profiling was performed without TGF-β stimulation to focus the analysis on off-target effects. Due to its superior functional knockdown abilities (Figure 14b) and little off-target effects siRNA A1tgf (Figure 15b) was used in all further experiments.

Table 4 – List of TGF-β1 siRNAs

Vendor	Description	Cat. No.	Abbr.
Ambion	Silencer® Select TGF beta receptor 1	AM51331-556	A1tgf
Ambion	Silencer® Select TGF beta receptor 1	AM51331-557	A2tgf
Dharmacon	On-TARGET plus Duplex TGFBR1	J-003929-09-05	D1tgf
Dharmacon	On-TARGET plus Duplex TGFBR1	J-003929-10-05	D2tgf
Dharmacon	On-TARGET plus Duplex TGFBR1	J-003929-11-05	D3tgf
Dharmacon	On-TARGET plus Duplex TGFBR1	J-003929-12-05	D4tgf
Qiagen	Hs_TGFBR1_5_HP Validated siRNA	SI00301903	Q1tgf
Qiagen	Hs_TGFBR1_6_HP Validated siRNA	SI02223627	Q2tgf
Qiagen	Hs_TGFBR1_7_HP Validated siRNA	SI02223634	Q3tgf
Qiagen	Hs_TGFBR1_9_HP Validated siRNA	SI02664158	Q4tgf

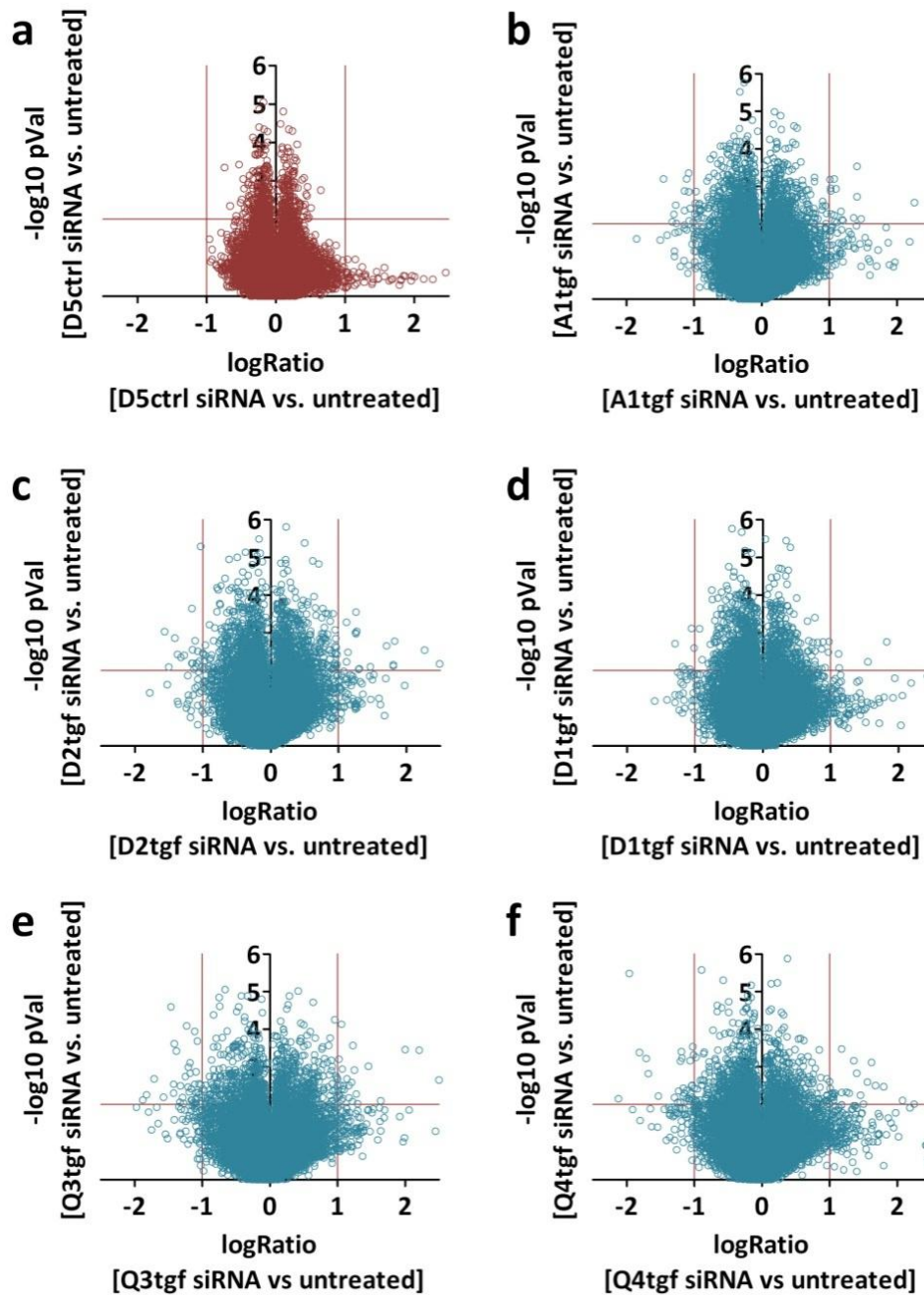


Figure 15 – TGF- β siRNA off-target effects

Volcano plots for siRNAs A1tgf, D1tgf, D2tgf, Q3tgf & Q4tgf. Total RNAs of biological triplicates were isolated post siRNA transfection and were hybridized to Illumina Beadchips. The off-target effects were analyzed by volcano plots. Each circle represents a single gene of the human genome. The x-axis depicts the log₂ ratio (LR) between each siRNA and untreated cells. The y-axis is scaled as $-\log_{10}$ [p-value] (Student t-test) as an indicator of significance. An off-target is defined to have a $|LR| \geq 1$ and a $-\log_{10}$ [p-value] > 2. D5ctrl siRNA vs. untreated D5ctrl revealed no off-target effects (a). The siRNAs A1tgf revealed 22 genes to be deregulated (b), the siRNA D1tgf - 8 genes (c), the siRNA D2tgf - 25 genes (d), the siRNA Q3tgf - 58 genes (e) & the siRNA Q4tgf - 42 genes (f).

Results

In addition, the stability of the receptor knockdown using siRNA A1tgf was determined. Therefore, HaCaT cells were transfected and TGF- β R1 mRNA was analyzed 1, 2, 3, 4 and 7 days after transfection. Transfection of siRNA A1tgf resulted in a stable knockdown of at least 70 % even 7 days after transfection (Figure 16).

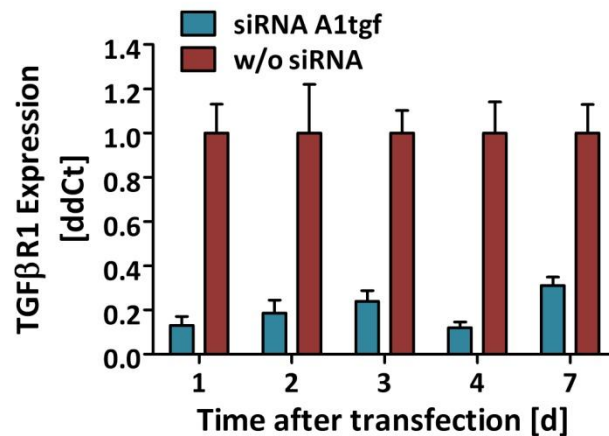


Figure 16 – Knockdown stability

HaCaT cells were transfected with siRNA A1tgf and knockdown efficacy was determined by qRT-PCR 1, 2, 3, 4 and 7 days post transfection. Even after 7 days knockdown was stable with approximately 70 %. All error bars indicate the standard deviation of n=2.

2.2.2 Kinase inhibitors

The kinase inhibitors used in the Phenocopy experiment derived from a Boehringer Ingelheim (BI) lead optimization project. Additionally, two previously described competitor compounds⁹⁸ were also incorporated in the analysis.

The BI inhibitors were identified in a high-throughput screening of the in-house compound collection using a biochemical TGF- β 1 substrate phosphorylation assay⁹⁹. Among the hits, compound 5 (Figure 17) and several other indolinone derivatives were identified, displaying down to double-digit nanomolar potency. Since it was known from earlier projects that the indolinone chemotype can show low cross-reactivity with the human kinase¹⁰⁰, it was chosen as starting point for further optimization. Structure-activity relationship studies were performed by substitution of the indolinone core in position R¹, R² and R³ of the scaffold (Figure 17) in order to increase the potency. Besides the inhibitory potency of the compounds in the primary biochemical kinase assay it needed to be shown that they also inhibit TGF- β signal transduction in a cellular setting. Therefore the pSmad2/3 ELISA was used to further characterize the inhibitors' potency. All compounds were analyzed in seven different concentrations: 0.0032, 0.016, 0.08, 0.4, 2, 10 and 50 μ M.

In the following the compounds were named with numbers adopted from Roth *et al.*³³ where their synthesis and design is described in more detail. The potency data derived from the TGF- β 1 substrate phosphorylation assay was also taken from this work.

Results

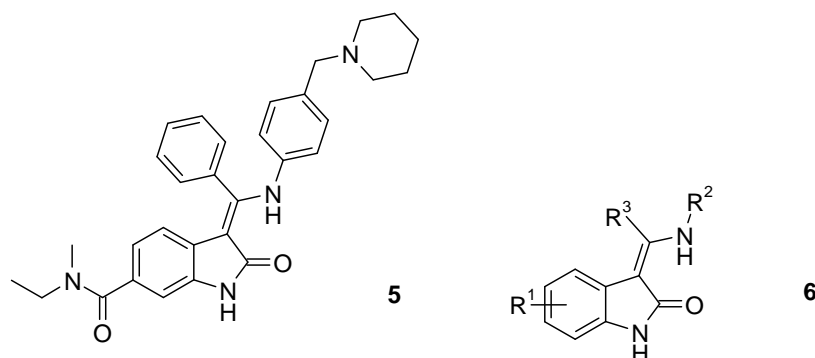


Figure 17 –TGF- β R1 hit compound 5 (BIBF0775) and general structure 6

The first subset of compounds contained indolinones substituted in position R¹ (Table 5). Hereby, the compounds substituted with smaller secondary and tertiary amides (**5**, **35-38**) in position 6 were favorable for a good TGF- β R1 potency. The unsubstituted indolinone **45** was significantly less active confirming that the 6-amido substituents contribute to the overall binding energy. In contrast, the same amides in position 5 resulted in less potency of the compound (compare compounds **46** and **37**). In general, the trend of inhibition in the cellular setting correlated with the data from the biochemical assay. However, the compounds were less potent when compared to the biochemical TGF- β R1 inhibition (factor 3-12 for potent compounds). Taken together, due to their good potency on a biochemical as well as on cellular level, these compounds looked promising as starting points for further optimization. All potencies of the modified indolinones are listed in Table 5.

Results

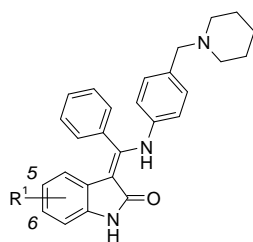


Table 5 – TGF- β R1/pSMAD inhibition of substituted indolinones

Cpd.	R ¹	TGF- β R1 IC ₅₀ (nM) ^a	pSMAD IC ₅₀ (nM) ^a
35	6-CONH(CH ₂ CH ₃)	24 ± 17	75 ± 53
36	6-CONHCH ₃	32 ± 26	NT ^b
5	6-CONEtMe	34 ± 30	105 ± 70
37	6-CONMe ₂	35 ± 22	246 ± 130
38	6-CONHnBu	91 ± 49	1065 ± 588
39	6-(pyrrolidine-1-carbonyl)	245 ± 170	696 ± 542
40	6-CONHiPr	318 ± 188	708 ± 453
41	6-CONH ₂	369 ± 213	>50000
42	6-CONH(CH ₂ CH ₂ OH)	430 ± 227	>50000
43	6-CONEt ₂	625 ± 396	NT ^b
44	6-CONHBn	1532 ± 908	NT ^b
45	H	3462 ± 2621	NT ^b
46	5-CONMe ₂	186 ± 104	198 ± 111

^a Values are averages ± SD of at least three independent determinations. Values “greater than” indicate that half-maximum inhibition was not achieved at the highest concentration tested.

^b Not tested.

To further explore structure-activity relationships, the R² side chain pointing towards the water phase was chosen for the next round of modifications (Table 6). Here, a large degree of freedom for structural variation was observed with various linkers between the aniline and the basic moiety were tolerated (**47a-u**). The distance between the basic moiety and the core seemed to play a minor role demonstrating that structure-activity relationships in this respect were shallow. Neutral compounds were clearly inferior (compare especially compound **47b** with **47q**). Anilines with smaller substituents (compounds **47r** and **47u**) or without any substitution (compound **47v**) were less active. Shifting the R⁶ substituent on the aniline (see Table 6) from position 4 into position 3 was also detrimental to activity (compound **47s**). Replacing the aniline by saturated cyclic systems such as **48** and **49** led to complete loss

Results

of potency. Comparable to hit compounds **5** and **35**, indolinones substituted in position 6 with either an ethylamido or an ethylmethylamido-substituent usually showed comparable activities with a tendency for higher potency for the ethylamido-substituted compounds (compare e.g. couples **47a/b**, **47c/d**, **47e/f** etc.). Although, structure-activity relationships were shallow, improved compounds with single-digit nanomolar TGF- β R1 inhibition could be identified by variation of position R², with compound **47a** showing an optimized IC₅₀ of 1 nM. Cellular efficacy of the compounds in Table 6 was again evaluated by pSmad2/3 ELISA assay. For many compounds, IC₅₀ values followed the same trend as mentioned before, i.e. cellular efficacy differed by a certain factor from biochemical IC₅₀. However, many compounds of this series displayed increased shifts between cellular and biochemical inhibition. Amongst them, potent compounds such as **47c**, **47e**, **47g**, and **47h** failed to show convincing cellular potency at all. This could be rationalized by the increased polarity of the compounds probably preventing them from permeating into the cells. However, despite their cellular shifts, several optimized compounds with attractive cellular activities such as **47a**, **47d**, **47i** or **47l** were identified by variation of R².

Results

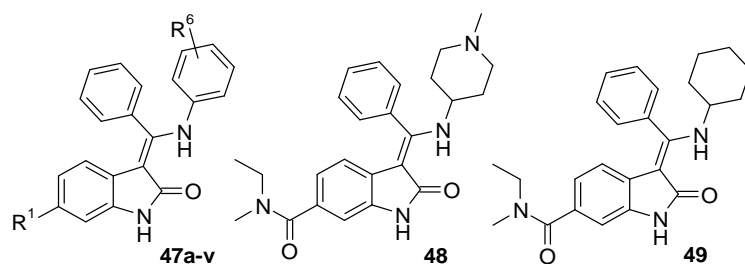


Table 6 – Inhibitory profile of 6-amido-substituted indolinones

Cpd.	R ¹	R ⁶	TGFβ-R1 IC ₅₀ [nM] ^a	pSmad IC ₅₀ [nM] ^a
47a	CONHEt	4-(NSO ₂ CH ₃)(CH ₂) ₂ NMe ₂	1 ± 1	108 ± 65
47b	CONEtMe	4-(NSO ₂ CH ₃)(CH ₂) ₂ NMe ₂	7 ± 5	209 ± 108
47c	CONHEt	4-(NCOCH ₃)(CH ₂) ₂ NMe ₂	3 ± 3	1066 ± 659
47d	CONEtMe	4-(NCOCH ₃)(CH ₂) ₂ NMe ₂	8 ± 5	102 ± 61
47e	CONHEt	4-(NCOCH ₃)(CH ₂) ₃ NMe ₂	3 ± 3	1249 ± 664
47f	CONEtMe	4-(NCOCH ₃)(CH ₂) ₃ NMe ₂	9 ± 6	411 ± 309
47g	CONHEt	4-(NCH ₃)COCH ₂ -(4-methyl-piperazin-1-yl)	9 ± 11	1370 ± 735
47h	CONEtMe	4-(NCH ₃)COCH ₂ -(4-methyl-piperazin-1-yl)	7 ± 6	833 ± 549
47i	CONHEt	4-CH ₂ NMe ₂	19 ± 13	185 ± 99
35	CONHEt	4-CH ₂ (piperidin-1-yl)	24 ± 17	75 ± 53
5	CONEtMe	4-CH ₂ (piperidin-1-yl)	34 ± 30	105 ± 70
47j	CONHEt	4-(NCH ₃)COCH ₂ NMe ₂	35 ± 22	3383 ± 3291
47k	CONEtMe	4-(NCH ₃)COCH ₂ NMe ₂	29 ± 16	382 ± 207
47l	CONEtMe	4-CH ₂ NHEt	32 ± 18	160 ± 83
47m	CONHEt	4-CONH(CH ₂) ₂ NEt ₂	33 ± 23	505 ± 264
47n	CONEtMe	4-CONH(CH ₂) ₂ NEt ₂	64 ± 53	542 ± 317
47o	CONHEt	4-(CH ₂ CH ₂)NMe ₂	47 ± 27	135 ± 70
47p	CONEtMe	4-(CH ₂ CH ₂)NMe ₂	64 ± 40	259 ± 190
47q	CONEtMe	4-(NSO ₂ CH ₃)CH ₂ CONMe ₂	86 ± 53	> 3000
47r	CONEtMe	4-(NCH ₃)SO ₂ Me	231 ± 129	> 10000
47s	CONEtMe	3-CH ₂ NEt ₂	616 ± 353	NT ^b
47t	CONEtMe	4-COOH	782 ± 435	NT ^b
47u	CONEtMe	4-COOCH ₃	1970 ± 1085	NT ^b
47v	CONEtMe	H	2586 ± 1844	NT ^b
48			> 50000	> 10000
49			> 50000	> 10000

^a Values are averages ± SD of at least three independent determinations. Values “greater than” indicate that half-maximum inhibition was not achieved at the highest concentration tested. ^b Not tested.

Results

Finally, the influence of substitution at R³ (Figure 17) was explored (see Table 7). When comparing compounds in which R³ = phenyl with the corresponding compounds with R³ = H (**47w-z**), a slight drop of potency could be observed for TGF-βR1 inhibition. Since compounds **47w-z** (R³ = H) are more polar than their phenyl counterparts, poor cell permeability is probably again the reason for the low activity of the compounds in the pSmad2/3 ELISA assay. Only the more lipophilic compound **47y** showed improved cellular activity in this series.

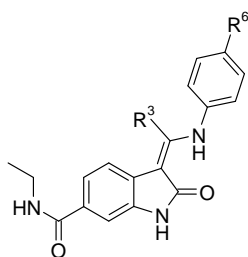


Table 7 – Inhibitory profile of various R⁴-substituted indolinones

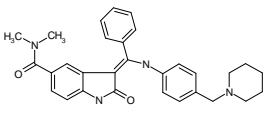
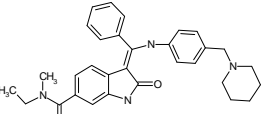
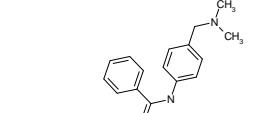
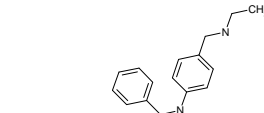
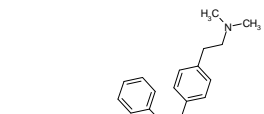
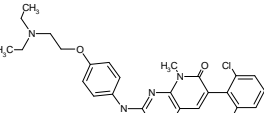
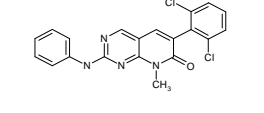
Cpd.	R ³	R ⁶	TGFβ-R1 IC ₅₀ [nM] ^a	pSmad IC ₅₀ [nM] ^a
47e	Ph	4-(NCOCH ₃)(CH ₂) ₃ NMe ₂	3 ± 3	1249 ± 664
47w	H	4-(NCOCH ₃)(CH ₂) ₃ NMe ₂	17 ± 10	597 ± 332
47g	Ph	4-(NCH ₃)COCH ₂ -(4-methyl-piperazin-1-yl)	9 ± 11	1370 ± 735
47x	H	4-(NCH ₃)COCH ₂ -(4-methyl-piperazin-1-yl)	15 ± 9	2175 ± 845
35	Ph	4-CH ₂ (piperidin-1-yl)	24 ± 17	75 ± 53
47y	H	4-CH ₂ (piperidin-1-yl)	69 ± 44	376 ± 246
47j	Ph	4-(NCH ₃)COCH ₂ NMe ₂	35 ± 22	3383 ± 3291
47z	H	4-(NCH ₃)COCH ₂ NMe ₂	52 ± 30	4355 ± 2298

^a Values are averages ± SD of at least three independent determinations.

Results

As mentioned above, five kinase inhibitors out of this lead optimization program were selected to further characterize them in the Phenocopy approach and to demonstrate that this approach is able to unravel new insights into the mode of action of these compounds beyond those that are currently investigated during a lead optimization phase. Therefore five BI compounds, in the following referred to as BI1-5 and two competitor substances (Ex1 and Ex2) with a wide range in inhibitory potency of TGF- β R1, from 19 nM (BI3) to 1537 nM (Ex2), were selected. A detailed list of the compounds' features, including biochemical and cellular IC₅₀ values, is provided in Table 8. The potencies (IC₅₀) for the inhibition of TGF- β R1 kinase, Smad2/3 phosphorylation (pSmad) and PAI-1 protein are indicated for compounds BI1 to BI5 (indolines [INDO]) and Ex1 and Ex2 (pyridopyrimidinones [PyPy]). The PubChem CIDs are indicated. According to the chemical synthesis of the compounds (Roth *et al.*³³), the corresponding compounds identification numbers are indicated in brackets.

Table 8 - List of profiled TGF- β R1 kinase inhibitors in the Phenocopy project

	B1	(Cpd #46)
	Formula	C₃₀H₃₂N₄O₂
	MW [g/mol]	480.609
	Structural Class	INDO
	IC50 TGFβR1 [nM]	186
	IC50 pSmad [nM]	198
	IC50 PAI-1 [nM]	438
PubChem CID	-	
	B2	(Cpd #5 / BIBF0775)
	Formula	C₃₁H₃₄N₄O₂
	MW [g/mol]	494.636
	Structural Class	INDO
	IC50 TGFβR1 [nM]	34
	IC50 pSmad [nM]	105
	IC50 PAI-1 [nM]	165
PubChem CID	-	
	B3	(Cpd #47i / B134659)
	Formula	C₂₇H₂₈N₄O₂
	MW [g/mol]	440.5442
	Structural Class	INDO
	IC50 TGFβR1 [nM]	19
	IC50 pSmad [nM]	185
	IC50 PAI-1 [nM]	227
PubChem CID	-	
	B4	(Cpd #47j)
	Formula	C₂₈H₃₀N₄O₂
	MW [g/mol]	454.571
	Structural Class	INDO
	IC50 TGFβR1 [nM]	32
	IC50 pSmad [nM]	160
	IC50 PAI-1 [nM]	223
PubChem CID	-	
	B5	(Cpd #47k)
	Formula	C₂₉H₃₂N₄O₂
	MW [g/mol]	468.598
	Structural Class	INDO
	IC50 TGFβR1 [nM]	64
	IC50 pSmad [nM]	259
	IC50 PAI-1 [nM]	1550
PubChem CID	-	
	Ex1	[PD166285]
	Formula	C₂₆H₂₇Cl₂N₅O₂
	MW [g/mol]	512.438
	Structural Class	PYPY
	IC50 TGFβR1 [nM]	25
	IC50 pSmad [nM]	211
	IC50 PAI-1 [nM]	220
PubChem CID	5311382	
	Ex2	[PD164199]
	Formula	C₂₀H₁₄Cl₂N₂O
	MW [g/mol]	397.264
	Structural Class	PYPY
	IC50 TGFβR1 [nM]	1537
	IC50 pSmad [nM]	400
	IC50 PAI-1 [nM]	855
PubChem CID	5327885	

INDO - Indolinone | PYPY - Pyridopyrimidinone

2.3 Phenocopy Experiment

The profiling of seven NCEs (BI1-5 & Ex1, Ex2) at seven different concentrations (0.0032, 0.016, 0.08, 0.4, 2, 10, 50 μ M) and three time points (2, 4, 12 h), including siRNA A1tgf, all appropriate controls and in biological triplicates for each condition resulted in an overall experimental setup of 651 samples to be submitted to array profiling. Subsequently, an optimal normalization method was selected for the expression data, the TGF- β signature as well as the NCE's off-target effects were determined and the obtained *in silico* results were validated in wet laboratory experiments.

2.3.1 Data normalization

Expression data was pre-processed in 24 different ways (Table 10). In this part, the analysis was focused on analyzing the TGF- β stimulated and control samples measured at three time points (2, 4, and 12 h). Generally speaking, first either background normalization from BeadStudio¹⁰¹ (bg_*) or no background modification (noBg_*) was applied. In a next step, the data was transformed using either log2-transformation (log) or vst¹⁰². Since background normalization can lead to negative values, the data had to be transformed to contain only positive values by using either the background correction of rma¹⁰³ or forcePos¹⁰⁴ to be able to apply log2-transformation. In a last step, the data was normalized using quantile, loess or rsn normalization. Alternatively, the transformation steps were skipped and vsn¹⁰⁵ or normalization methods supplied by BeadStudio (average, rankInvariant, cubicSpline) were used for normalization. Different pre-processing methods were evaluated by

Results

analyzing the resulting gene expression intensities via various statistical measures. These methods were scored from -2 to 2 based on how well they matched the required criteria for the different analyses (Table 10). Their impact on the quality for normalization of the presented expression study was analyzed in a separate study and was submitted for publication at *BMC Genomics*¹⁰⁶. In addition, to the investigation of the actual expression intensities, fold changes derived from resulting gene expression intensities were compared to fold changes based on quantitative measurements of RNA abundance as determined by qRT-PCR. Therefore, the expression of eight genes that are known to be regulated by TGF- β signaling (CDKN1A, CDKN2B, HAND1, JUNB, LINCR, RPTN, PAI-1 and TSC22D1) was analyzed. By this means, it is possible to compare the results of the normalization methods to values that reflect the real abundance of the respective mRNA in the cells. To guarantee that the comparisons of the normalization methods are not biased towards certain intensities, the mRNAs used in qRT-PCR experiments were chosen such that the respective signals on the chips cover a broad range of expression intensities (Table 9).

Table 9 – qRT-PCR fold changes of TGF- β dependent genes

Gene	Fold change of TGF- β stimulated vs. unstimulated HaCaT cells		
	2 h	4 h	12 h
PAI-1	70.07	98.32	124.85
CDKN1A	6.22	2.36	4.17
LINCR	5.06	0.74	0.05
CDKN2B	10.26	6.36	4.12
HAND1	3.65	2.68	0.54
TSC22D1	2.42	1.73	1.32
RPTN	9.03	4.63	0.03
JUNB	18.41	8.63	10.19

Results

Based on the different normalization procedures for the gene expression experiment and based on the qRT-PCR measurements, the Pearson correlation of the respective fold changes measured for TGF- β stimulated versus untreated cells at 2, 4 and 12 h were calculated.

Figure 18 displays the ranked correlation coefficients describing the relation between the different normalization methods and the qRT-PCR results. Quality values were assigned based on correlation cut-offs. A value of 2 is assigned to correlation coefficients ≥ 0.96 , a value of 1 to coefficients between 0.94 and 0.96, a value of 0 to coefficients between 0.92 and 0.94, a value of -1 to coefficients between 0.9 and 0.92 and a value of -2 to correlation coefficients ≤ 0.9 (Table 10). Values derived from most of the methods not utilizing background correction (noBg_*) show a lower correlation to the qRT-PCR results than expression intensities that are background corrected (bg_*). An exception in this regard are methods that are based on vst transformation (bg_vst_*). These three methods are amongst the six methods resulting in the lowest correlation coefficient values. Correlation coefficients exhibiting high values are delivered by methods introducing BeadStudio's background correction combined with either rma background correction and log2-transformation (bg_rma_log_*), cubic spline normalization (bg_cubicSpline) or variance stabilizing normalization (bg_vsn).

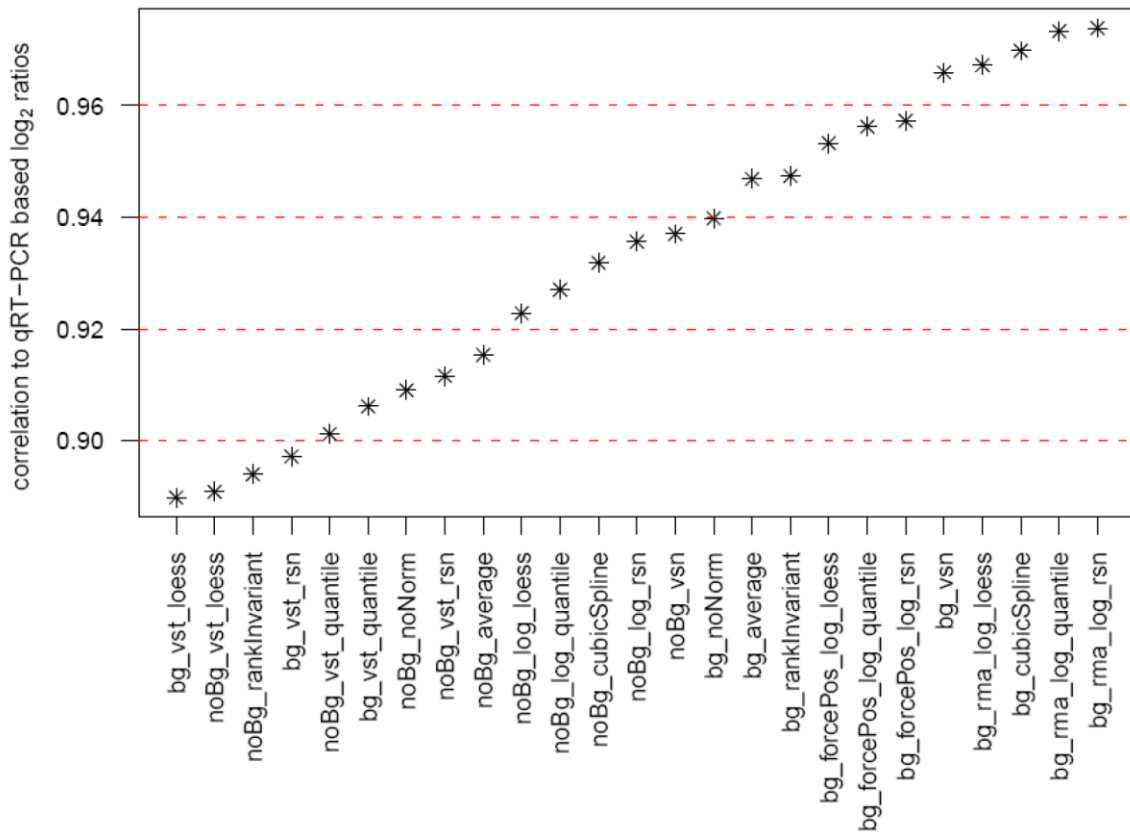


Figure 18 - Pearson correlation of log₂ ratios for different normalization methods

Correlations of log₂ ratios determined based on different normalization methods for gene expression data obtained from the Illumina array measurement to qRT-PCR results. On the x-axis, pre-processing methods are ranked according to their correlation. The dashed red lines indicate the cut-offs used for assigning quality score between -2 (< 0.9) and 2 (> 0.96).

As it is difficult to clearly categorize the methods based on the examined measures, the final decision of which score to assign to some extent stays subjective. However, it is unambiguously possible to separate better pre-processing methods from worse. A summary of all analyzed parameters, statistical measures and the fold change correlation is given in Table 10.

Table 10 –Quality scores given for the different pre-processing methods

	log10(p-value) versus MSQbetween	Boxplots MSQ	Density functions of MSQs	Volcano plots	Residual sd versus expression levels	Scatterplots	AUC	Slope of regression	Correlation to qRT-PCR	Sum
bg_average	0	-1	0	-1	-2	-1	0	0	1	-4
bg_cubicSpline	0	-1	1	-1	-2	-1	0	1	2	-1
bg_forcePos_log_loess	1	-1	1	0	-1	0	0	1	1	2
bg_forcePos_log_quantile	1	-1	1	0	-2	-1	0	1	1	0
bg_forcePos_log_rsn	1	-1	1	0	-2	-1	0	0	1	-1
bg_noNorm	1	-1	1	-1	-2	-1	0	0	0	-3
bg_rankInvariant	0	-2	-1	-1	-2	-1	1	0	1	-5
bg_rma_log_loess	-1	-1	-1	-2	-2	-1	-1	2	2	-5
bg_rma_log_quantile	-1	-1	0	-2	-2	-2	0	2	2	-4
bg_rma_log_rsn	-1	-1	0	-2	-2	-2	0	2	2	-4
bg_vsn	-1	-1	0	-1	-2	-2	0	1	2	-4
bg_vst_loess	2	0	-2	-1	2	1	-1	-1	-2	-2
bg_vst_quantile	2	2	1	1	2	2	0	-1	-1	8
bg_vst_rsn	2	2	1	1	2	2	0	-1	-1	8
noBg_average	1	-2	-2	1	1	1	-1	-1	-1	-3
noBg_cubicSpline	2	1	1	1	0	2	0	0	0	7
noBg_log_loess	2	0	-2	0	0	1	0	0	0	1
noBg_log_quantile	2	2	2	2	1	1	0	0	0	10
noBg_log_rsn	2	2	2	2	1	1	0	0	0	10
noBg_noNorm	1	2	1	-1	0	1	-1	-1	-1	1
noBg_rankInvariant	0	-2	-1	-1	0	2	1	-2	-1	-4
noBg_vsn	2	2	2	2	0	1	0	0	0	9
noBg_vst_loess	2	0	-2	0	2	2	-1	-1	-2	0
noBg_vst_quantile	2	2	1	1	2	2	0	-1	-1	8
noBg_vst_rsn	2	2	1	1	2	2	0	-1	-1	8

Displayed are the quality scores for the different pre-processing methods given for the analyses conducted. Quality scores range from -2 (bad/red) to 2 (good/blue). The last column displays the sum over the quality scores assigned. Based on this sum, the preprocessing method finally used to normalize the Phenocopy data has been chosen.

An integration of all scores for the analyzed parameters revealed that best results for the present expression data were obtained when no background modification is used in combination with log₂-transformation and either quantile or rsn normalization. It was therefore decided to use rsn normalization in the Phenocopy study.

2.3.2 TGF- β signature

To gain a deeper insight into the TGF- β biology first genes were identified that are regulated due to TGF- β stimulation (5 ng/ml). To unravel the time-dependent effects of TGF- β treatment, HaCaT cells were stimulated with TGF- β for 2, 4 and 12 h. While immediate early genes that are directly regulated by the TGF- β pathway are detected at 2 h post stimulation, more and more secondary effects linked to TGF- β signaling are found after 4 h and/or 12 h. To avoid arbitrary log ratio cut-offs a modulator-based approach was used to identify TGF- β dependent gene regulation. Therefore, two criteria were applied: only genes that were significantly deregulated (p -value < 0.01) in a basic comparison of TGF- β stimulation versus unstimulated cells were further analyzed. We found 1,046, 1,949 and 5,725 genes (6,525 non-redundant genes) to be regulated 2, 4 and 12 h after stimulation (Figure 19). Next, these genes were proven to be affected dose-dependently by kinase inhibitor treatment after TGF- β stimulation. The approach allowed separating these genes from potential compound related off-target effects. All transcripts identified for each NCE were merged to a common signature of TGF- β dependent genes. This strategy allowed the identification of a common on-target signature minimized for the amount of false positive and false negative genes. The Venn diagram (Figure 19b) depicts the number of genes that were identified after 2, 4, and 12 h of stimulation: 446 genes (2 h), 772 genes (4 h) and 1,932 genes (12 h). All gene identifier annotations and regulations are listed in Supplement 2.

Beyond the inhibition of the kinase activity by chemical compounds, the TGF- β pathway was also silenced by siRNA knockdown. All previously selected genes (TGF- β

Results

stim. vs. unstimulated cells, Figure 19a) were tested to be regulated by siRNA-mediated knockdown of TGF- β 1. To exclude mechanical effects, genes were only selected when they were regulated by siRNA A1tgf and not by a control siRNA (p-value < 0.01). By siRNA knockdown of TGF- β 1, 303 (2 h), 419 (4 h) and 1,112 (12 h) genes are identified as TGF- β dependent (Figure 19c). Although fewer genes were identified compared to the NCE approach, the vast majority of genes were identified by both approaches. According to the siRNA transfection procedure, a slightly different experimental setup was performed regarding cell seeding and culture conditions. This variation resulted in procedure specific changes in gene regulation, which had to be separated from the TGF- β signature. Addressing the level of TGF- β 1 activity upon siRNA transfection, the expression of PAI-1 as a surrogate marker for the TGF- β signaling activity was analyzed. Using NCEs it was possible to inhibit the PAI-1 expression by more than 95 % at all time points. In contrast, the use of siRNA A1tgf reduced PAI-1 levels only to 62 % after 2 h of TGF- β stimulation (Figure 19d). Despite the high efficiency of the siRNAs (A1 showed a mRNA knockdown efficiency of greater than 90 %; Figure 14a) treatment resulted only in a partial reduction of the TGF- β signaling.

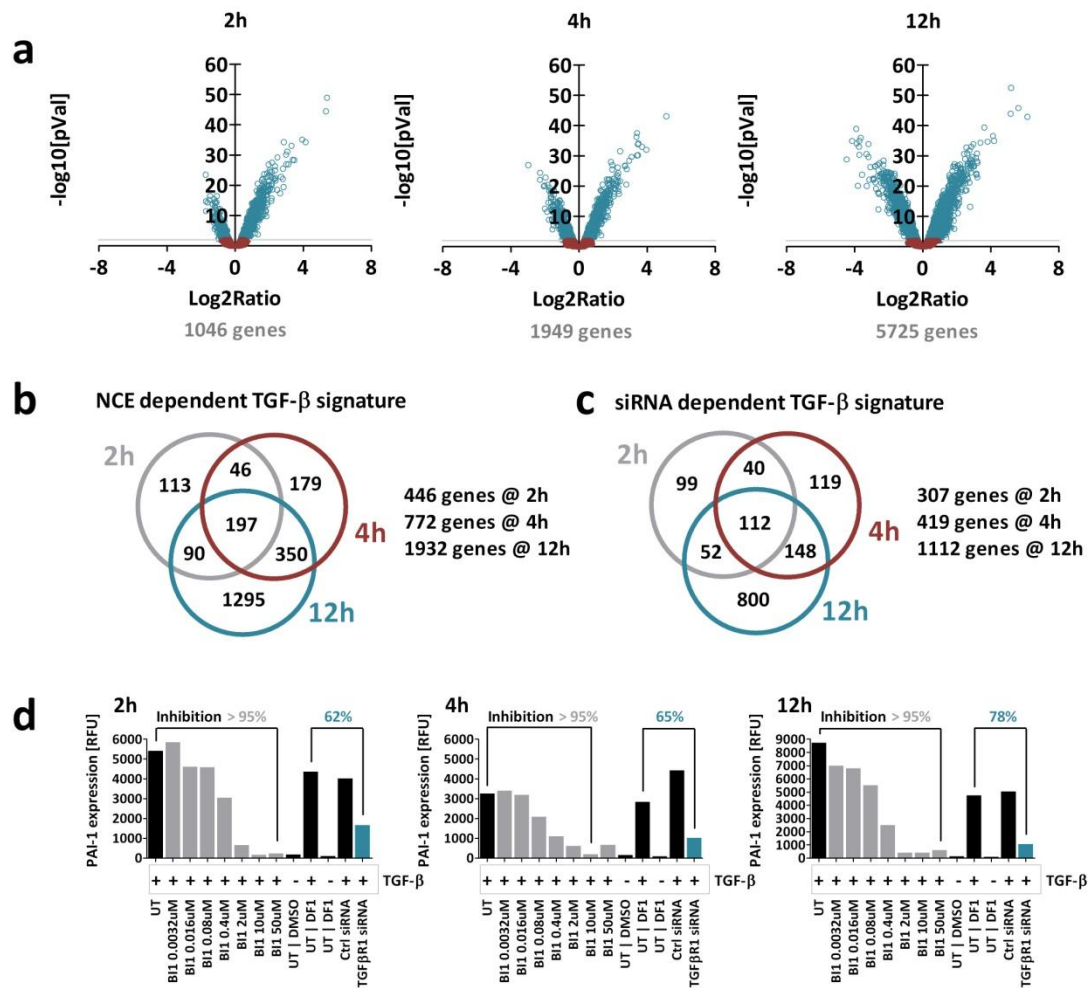


Figure 19 – TGF- β signature

The TGF- β signature was generated based on gene regulation upon treatment with TGF- β , TGF- β R1 kinase inhibitors (NCEs) or a siRNA. a: volcano plots of the comparison between TGF- β stimulated and non stimulated cells at 2, 4 and 12 h. Every circle represents a single transcript. The x-axis shows the log2 ratio (LR) between TGF- β stimulated vs. untreated HaCaT cells. The y-axis is scaled as negative log10 [p-value] as an indicator of significance. P-values were FDR-corrected according to Benjamini-Hochberg. Blue circled genes are significantly regulated by the stimulation with TGF- β (p-value < 0.01). b: the list of non-redundant genes was filtered for a concentration-dependent regulation upon NCE treatment and TGF- β stimulation.: 446, 772 and 1,932 genes were identified as the NCE-dependent on target TGF- β signature after 2, 4 and 12 h. c: the siRNA dependent TGF- β signature identified 307, 419 and 1,112 genes which were classified as siRNA-dependent on-target TGF- β signature genes after 2, 4 and 12 h. d: expression level of PAI-1 mRNA as a surrogate marker for TGF- β signaling pathway activity after treatment with NCE B1 or siRNA. Treatment with NCE B1 resulted in a complete knockdown of PAI-1 expression (>95 %) for all time points. In contrast the siRNA A1 mediated knockdown of the TGF- β signaling only reduced PAI-1 levels partially to 62 %, 65 % and 78 % after 2, 4 and 12 h of TGF- β stimulation.

Results

Subsequently, the genes of the TGF- β signature were used to perform gene set enrichment analysis (GSEA)^{107,108}. The annotation of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways delivered gene sets corresponding to 201 different pathways^{74, 109, 110}. The GSEA resulted in 16 different signaling pathways which were significantly influenced upon TGF- β stimulation of HaCaT cells. The signaling pathways were clustered in four groups (Figure 20). Not surprisingly, the TGF- β signaling pathway itself, as well as directly affected pathways like WNT and p53 signaling, were significantly regulated by the treatment of TGF- β (Cluster 1). In Cluster 2 MAPK, cytokine, ErbB, Hedgehog, as well as apoptosis signaling pathways are strongly affected immediately early upon TGF- β stimulation. Their modulation is reduced at later time points (4 h and 12 h), when more secondary effects, such as DNA polymerase, actin cytoskeleton, amino acid metabolism, gap junction and tight junction signaling become apparent (Cluster 3). The activation of these pathways in combination with the modulation of the cell cycle and cell communication activity (Cluster 4) seems to be the phenotypic consequences of TGF- β stimulation of HaCaT cells. To proof the findings obtained from the KEGG analysis, additionally Ingenuity Pathway Analysis (Ingenuity Systems®, www.ingenuity.com) was used to link and group genes from the TGF- β signature. In line with the KEGG results, the analysis identified the same connections and networks containing signaling but also WNT and ERK/MAPK signaling (Figure 21). In addition, diverse networks of genes were identified that play a role in embryonic development of different organs, but also in cellular proliferation and growth.

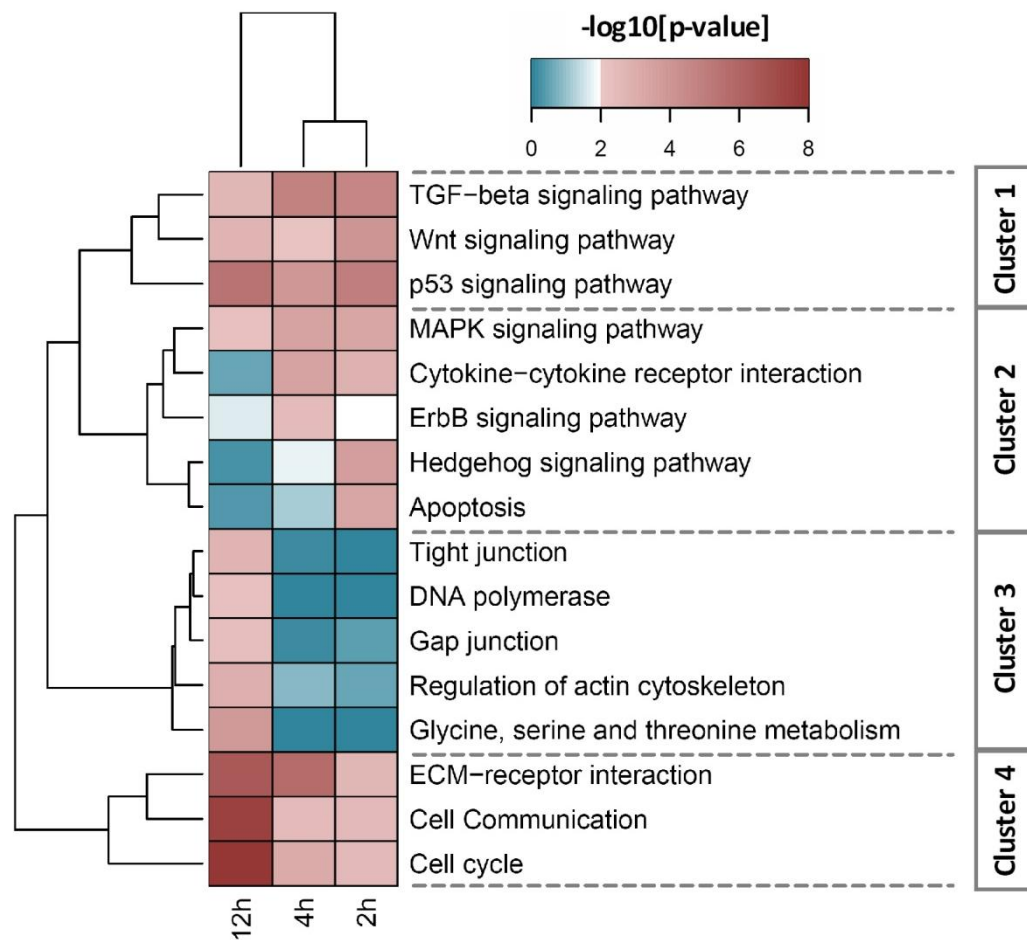


Figure 20 – TGF-β signature gene set enrichment analysis I

Gene set enrichment analysis (GSEA) using KEGG gene annotation resulted in 16 significantly affected genesets/signaling pathways. Clustering of $-\log_{10} [\text{p-values}]$ using complete linkage and manhattan distance resulted in four major clusters: immediate early affected pathways (cluster 2), permanently affected pathways with emphases at early (cluster 1) and late time points (cluster 4) or late established events (cluster 3). The color code defines the significance determined by Fisher’s exact test: blue < 2 – not significant; white = 2 – significant & red > 2 – highly significant).

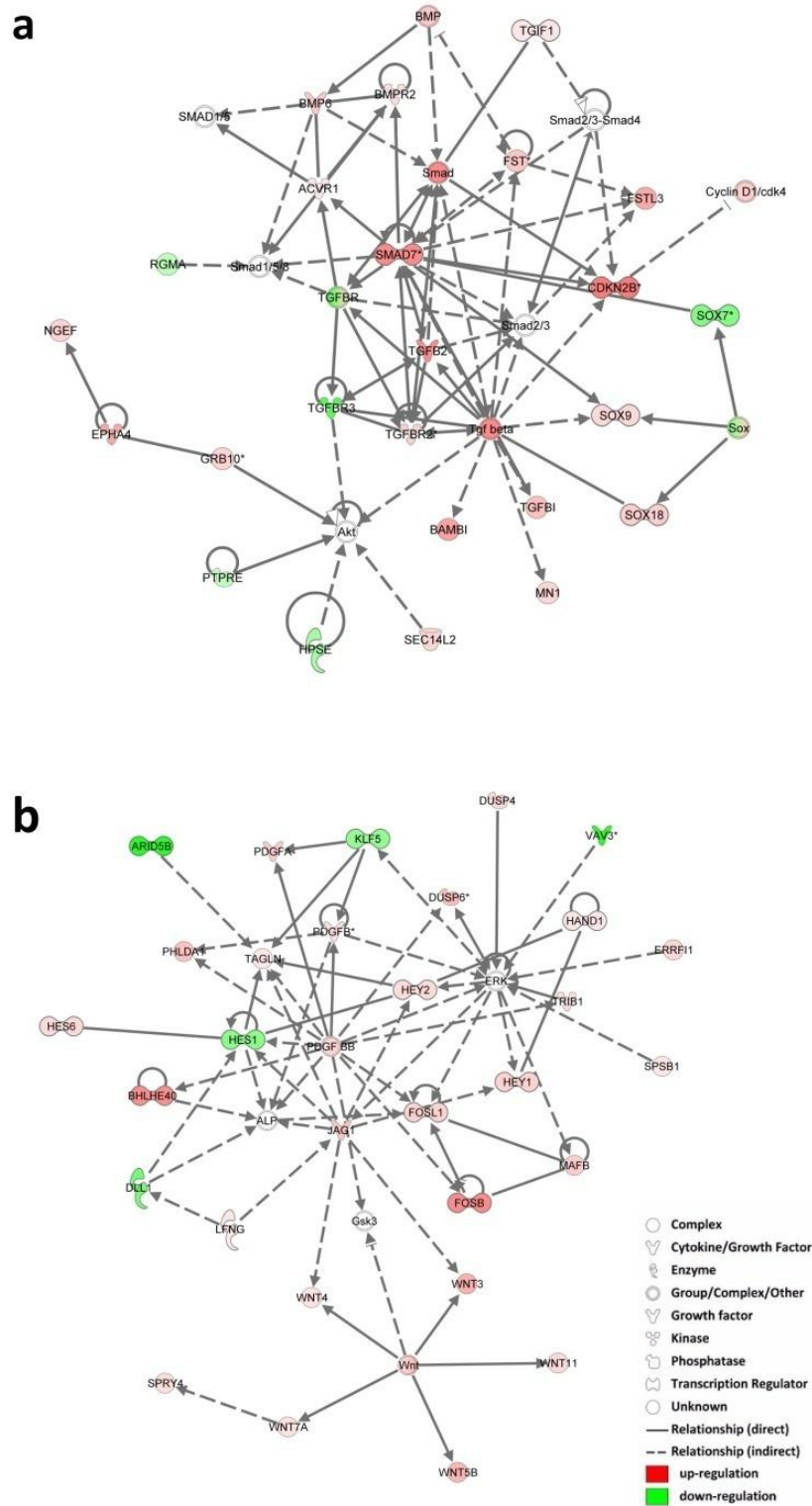


Figure 21 – TGF-β signature gene set enrichment analysis II

Networks of interacting and regulated molecules from the on-target signature generated by Ingenuity Pathway Analysis. All depicted molecules are represented as nodes and the biological relationship between two nodes is represented as an edge (line). All edges are supported by at least one literature reference. The intensity of the node color indicates the degree of up- (red) or down- (green) regulation. Nodes are displayed using various shapes that represent the functional class of the gene product. a: a network of molecules directly related to the canonical TGF-β signaling pathway containing genes involved in cell signaling, connective tissue development and function and in skeletal tissue development and function. b: a network of molecules of the WNT and the ERK/MAPK signaling pathways containing genes responsible for organ-, tissue and cellular development.

2.3.3 Off-target signature

After the identification of the TGF- β signature (on-target signature) as well as the affected pathways by GSEA using KEGG and IPA, the NCEs' off-target effects were identified and mapped to their molecular function and signaling pathways.

Each compound treatment resulted in a unique gene expression signature (phenocopy) of regulated genes. These signatures are composed of the cellular response to two different stimuli (TGF- β and NCE) and are integrated to the corresponding treatment signature. Thereby, elucidating the effects based on NCE treatment is more demanding since both TGF- β and off-target effects occur. Minor effects can also be observed for the interaction of the vehicle (DMSO) with the NCEs. The effects of the different stimuli overlap and also interfere with each other impeding a clear signature dissection. The profile of a given gene may therefore be dependent on which effect prevails and thus, dose-dependency might no longer be observed. In general, all regulated genes can be grouped into six classes, including single and integrated effects. Single effects derive either from the treatment with TGF- β (on-target effect) (Figure 22a) or from the treatment with NCEs (off-target effect). Single off-target effects can be further grouped in pure effects, where the genes are dose-dependently regulated (Figure 22b) and in so called common effects, where genes are dose-independently regulated by each of the seven compounds (Figure 22c).

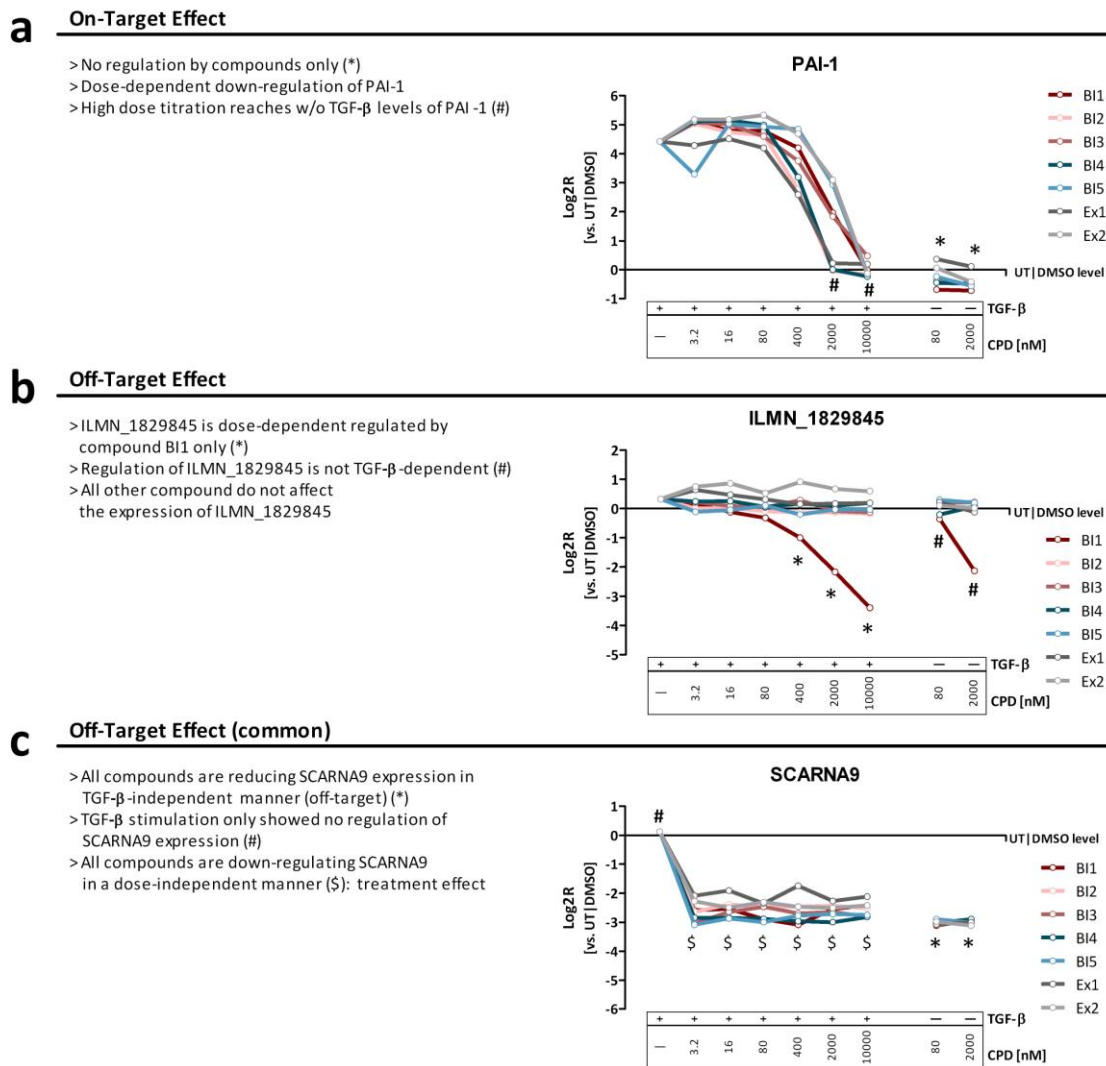


Figure 22 – Case profile definition (single effects)

NCE treatment and TGF- β stimulation resulted in different case profiles of gene regulation: representative examples for an on-target effect triggered by TGF- β (a), an off-target effect triggered by a NCE (b); common off-target effects induced by all seven NCEs in a dose-independent manner (c).

Additionally to the single-derived effects, an integration of both TGF- β and off-target effects can be detected: an NCE effect can be additive (Figure 23a) or inverse (Figure 23b) to the effect of TGF- β . Furthermore, opposed bipolar effects for high and low dosage of the NCE mostly linked with toxicity (Figure 23c), are observed.

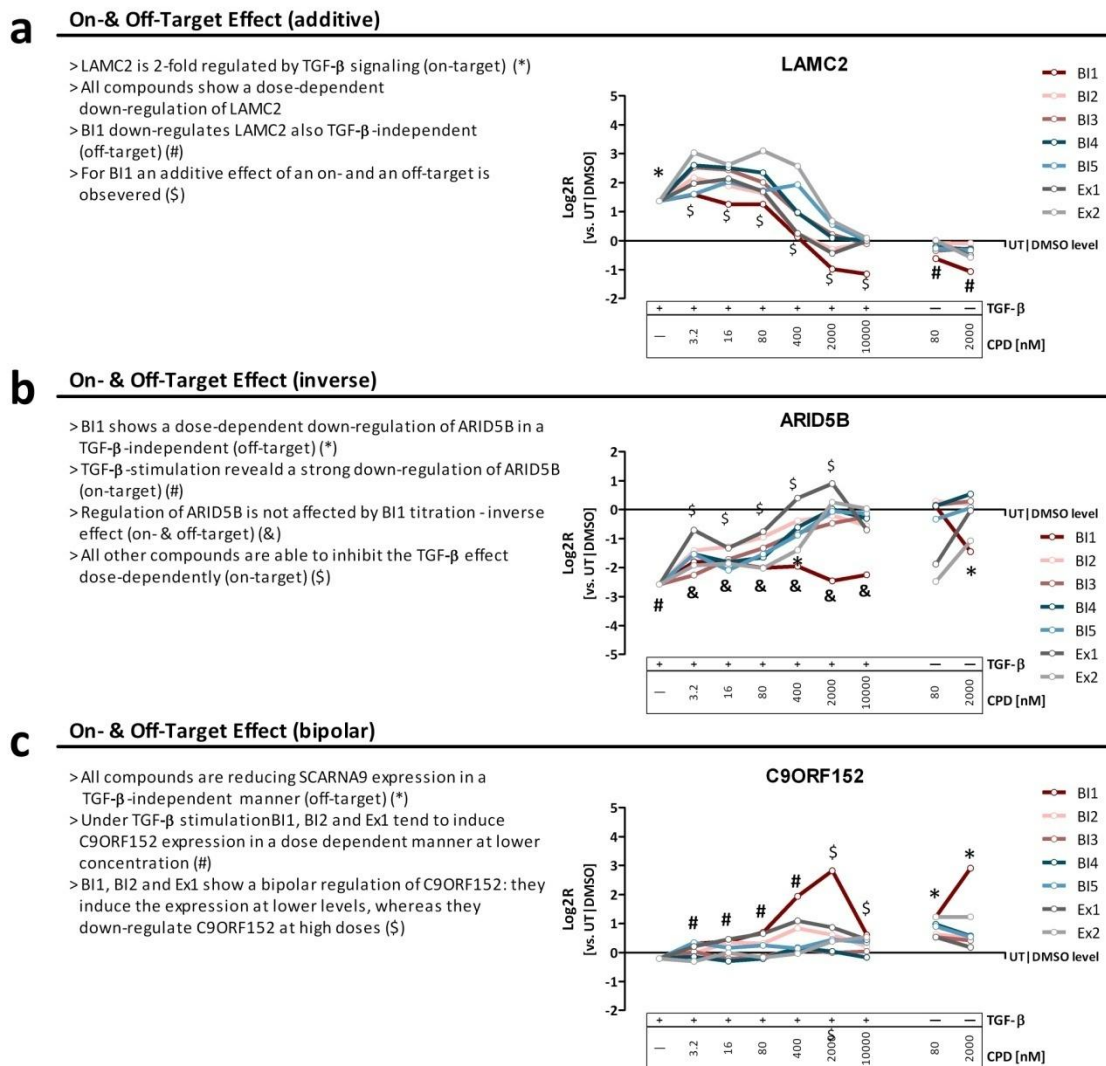


Figure 23 – Case profile definition (integrated effects)

NCE treatment and TGF- β stimulation resulted in different cases profiles of gene regulation: representative examples for integrated effects of on- and off-targets in an additive (c), inverse (d) and bipolar (c) manner.

Results

In a first approximation, the NCE treatment phenotypes (phenocopies) were determined as the total of all regulated genes (p -value < 0.01 and $|LR| \geq 1$) comparing NCE treated and TGF- β stimulated cells to DMSO control treated TGF- β stimulated cells. This analysis was done separately for each of the tested compounds at each concentration. Subsequently, the different phenotypes obtained after 2 h NCE treatment were clustered to unravel similarities between the different signatures (Figure 24). The early time point allowed focusing on primary affected genes that were altered as direct response to the treatment. Hierarchical clustering clearly revealed two major clusters separating the group of indolinones (BI1 to BI5) from the pyridopyrimidinones (Ex1 & Ex2). The fact that most obviously the specific chemotype has a major impact on differences in gene expression confirms that the classical notion of chemotypes determining biological profiles of NCEs holds true in this case. However, not only the scaffold itself, but also the specific decoration of each chemotype affected gene expression. The hierarchical cluster analysis demonstrates that treatment signatures can be used to differentiate even between analogs of the same chemotype. It was possible to clearly distinguish between signatures of BI1 treated cells from the other indolinones. These can be further subdivided into two clusters for either BI2 and BI3 and BI4 and BI5 that resulted in similar treatment effects. In general, treatment with each NCE resulted in clusters for high and low dose of the compound.

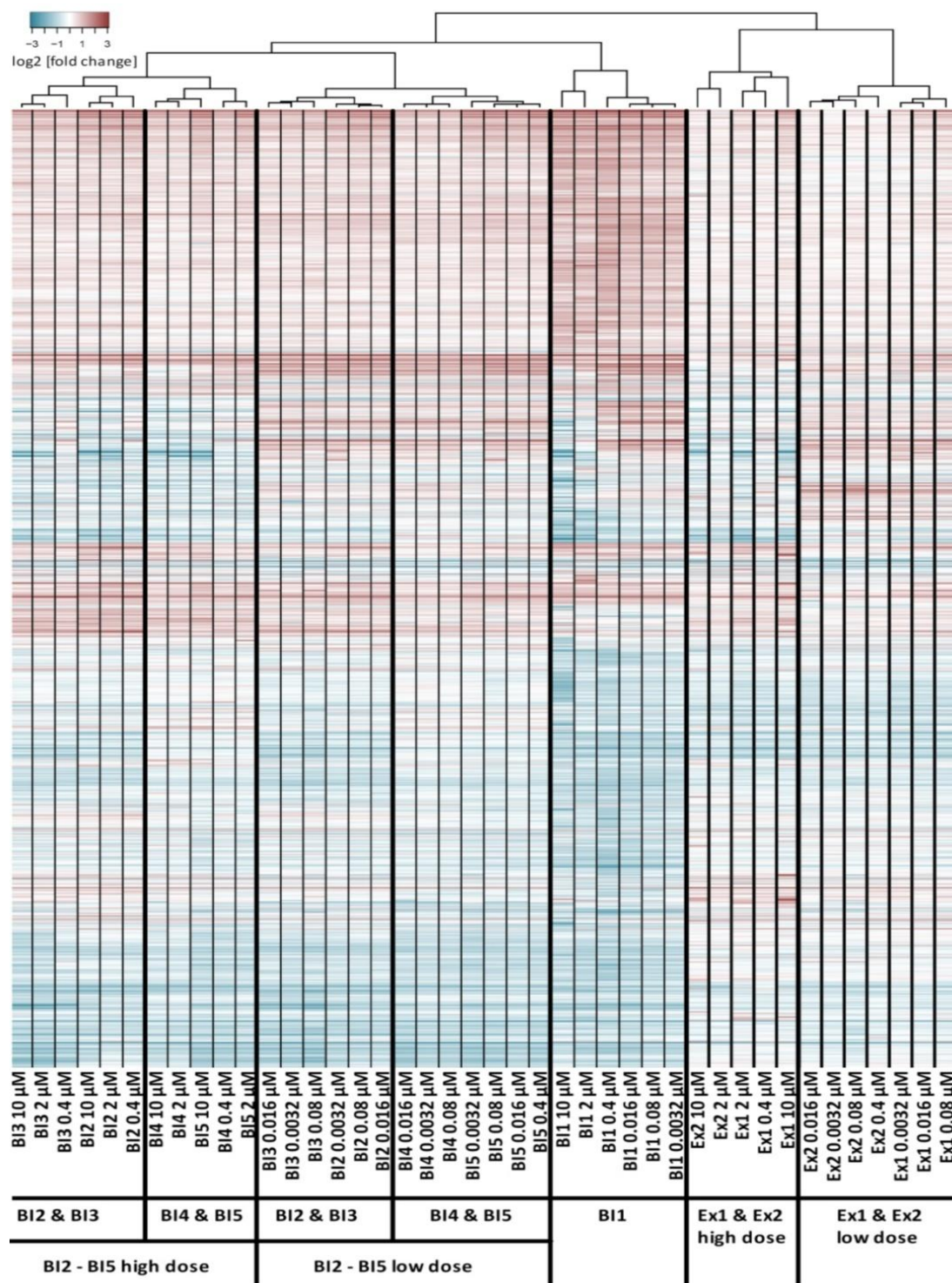


Figure 24 – Treatment signature

A hierarchical clustering of all 4,314 significant regulated genes ($|LR| \geq 1$ & $p\text{-value} < 0.01$) after NCE treatment and TGF- β stimulation for 2 h in HaCaT cells. The expression patterns of the different NCE-treated cells revealed several intersections in their effects on regulation: the five indolinones (BI1-BI5) are grouped and separated from the two pyridopyrimidinones (Ex1 & Ex2). Expression patterns are grouped in high vs. low dose fractions. The indolinone BI1 separates from the other class members, which can be further divided into two subgroups containing BI2 and BI3 and BI4 and BI5, respectively. Blue indicates decreased expression relative to untreated cells, red indicates increased expression.

Results

However, the identification of a particular off-target based on this approach is difficult. Further analyses were therefore performed to extract the compounds' off-target effects from the treatment signatures. As above mentioned not all off-target effects can be identified through dose dependence correlation due to overlapping, inverse and additive effects (Figure 23). Hence, off-targets can only be identified based on NCE treated samples in presence and absence of the TGF- β stimulus. Therefore, all regulated genes (p -value < 0.01 and $|LR| \geq 1$) comparing compound treated cells (either $0.08 \mu\text{M}$ or $2 \mu\text{M}$) to DMSO treated controls were selected. Genes were considered once the regulation was observed during compound treatment upon TGF- β stimulation as well as without TGF- β stimulation. Thus, it was ensured to select only drug target and TGF- β independent alterations. All genes that matched the criteria were allocated to the off-target signature of the NCE after 2, 4 and 12 h and are listed in Supplement 3. Based on this analysis, huge differences in the amount of off-target genes were observed. While treatment with BI1 deregulated 2,752 genes at all time points, BI3 deregulated "only" 973 genes. Slightly more off-target genes were identified for the indolinones BI2, BI4 and BI5 (1,050, 1,064 and 1,100). Both pyridopyrimidinones regulated 1,347 (Ex1) and 1,306 (Ex2) genes. The largest off-target increase over time was seen for Ex1 and Ex2 with almost four times more genes being regulated comparing the 12 h to the 2 h time point. In contrast, the amount of off-targets for the five indolinones was at a maximum doubled within this period (Figure 25). In summary, looking at the off-target signatures in general, the indolinones appeared more favorable compared to the pyridopyrimidinones at later points in time. Among the indolinones, BI2 to BI5 deregulated fewer genes than BI1 at all points in time which was confirmed by the different kinome

Results

specificities (see Chapter 'Kinase Profiling'). It also confirmed the structure-activity relationships described in Roth *et al.*³³ demonstrating that indolinones substituted in position 5 (such as BI1) showed a less favorable selectivity profile compared to indolinones substituted in position 6 (such as BI2-5). Among the indolinones, BI3 appeared to be the most attractive compound when merely looking at the off-target analysis.

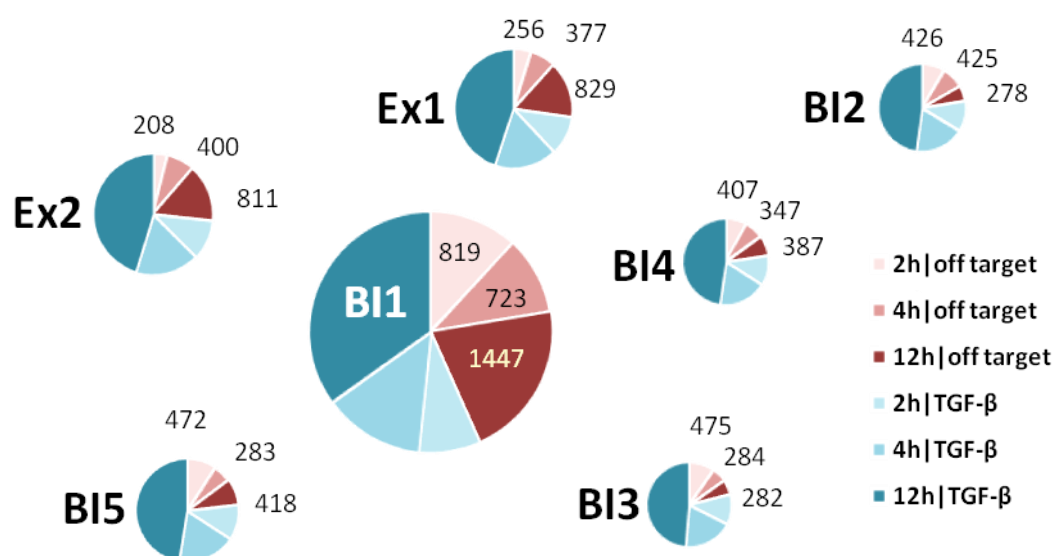


Figure 25 – Off-target effects (numbers)

Every circle represents one of the seven profiled compounds. The size of each circle corresponds to the number of off-target genes (in red). On-target genes numbers are shown in blue.

2.3.4 Molecular Function

Different *in silico* strategies can be applied to analyze the off-target signatures of the compounds in order to find a conclusion about their effects on the cells. Initially, the genes from the 12 h off-target signatures were assigned to their molecular function using Ingenuity Pathway Analysis (Figure 26). Hierarchical clustering resulted in one major cluster for both pyridopyrimidinones and in one for the indolinones (Figure 26). The genes regulated by the indolinones are distributed in more classes and genes involved in Vitamin and Mineral Metabolism, Cellular Compromise, Nucleic Acid- and Amino Acid Metabolism are exclusively regulated by the indolinones. In accordance with the structure-activity findings mentioned before, within the indolinone subcluster, BI1 stands apart from the four other indolinones and BI2 and BI3 as well as BI4 and BI5 are grouped in one cluster, respectively. The additionally by BI1 regulated genes are involved in RNA Post-Transcriptional Modification, Energy Production, Cellular Response to Therapeutics and in RNA Trafficking. Half of the molecular functional classes identified are regulated by all compounds. However, this does not necessarily mean that the same genes are regulated since the categories are rather spaciouly defined, such as Cell Cycle or Cellular Growth and Proliferation. Furthermore, different NCEs reach far higher significance scores for some categories than others caused by the higher amount of regulated genes in the respective biological process e.g. both pyridopyrimidinones regulate a huge amount of genes involved in Cell Death and Cellular Growth and Proliferation.

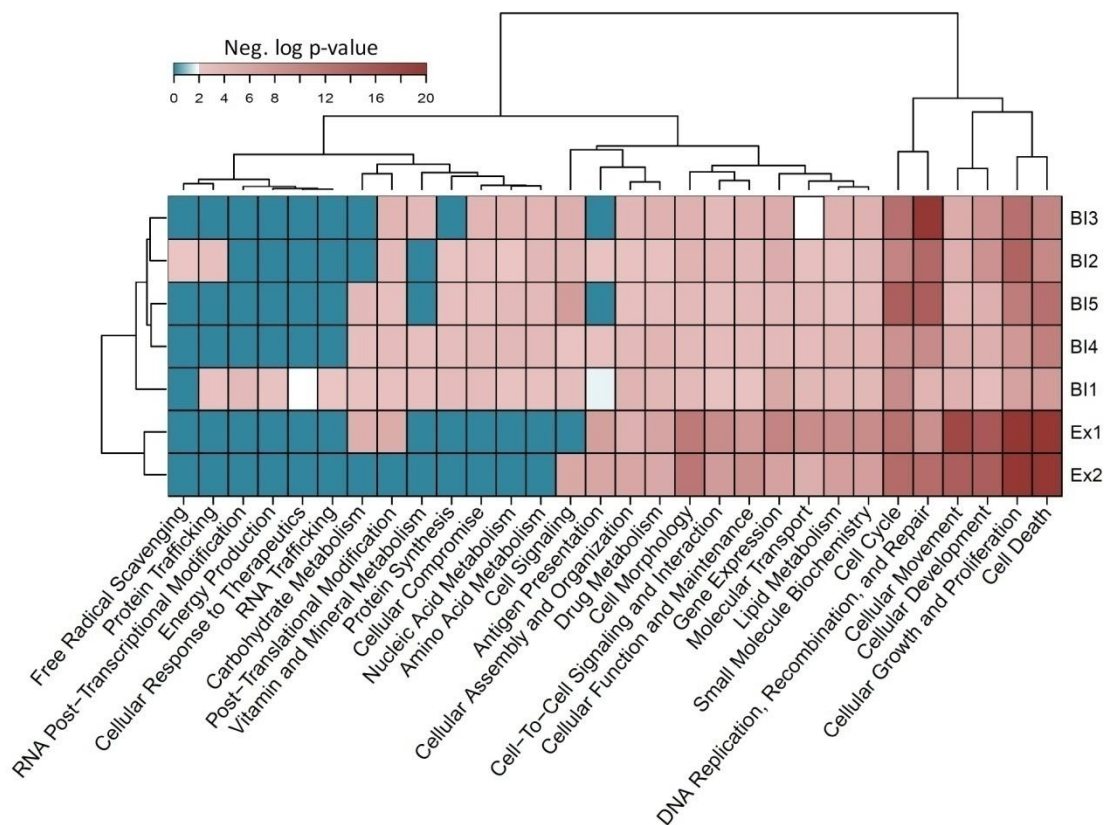


Figure 26 – Off-target effects (molecular function)

Molecular function of the off-target genes of all seven NCEs after 12 h treatment: clustering of the $-\log_{10}$ [p-values] using complete linkage and manhattan distance depicts the 30 significantly ranked categories. The color code defines the significance determined by Fisher’s exact test as $-\log_{10}$ [p-value]: blue < 2 – not significant; white = 2 – significant & red > 2 – highly significant.

2.3.5 Pathway Analysis

Subsequent to the annotation of the molecular function, Ingenuity Pathway Analysis was used to analyze the off-target signatures in order to enrich influenced signaling pathways to get a more precise idea about the compound effects. We found 39 (2 h), 38 (4 h) and 51 (12 h) canonical signaling pathways scored with a significant $-\log_{10}$ [p-value] > 2 (Fisher's exact test) for at least one of the NCEs (Figure 27 - Figure 29). While immediate early affected processes can be found 2 h after compound treatment, their effect, the treatment phenotype, however manifests during late phases. Caused NCE effects can thus be best observed 12 h after treatment. Hierarchical clustering of the pathway analysis results, again separated the indolinones from the pyridopyrimidinones, indicating that both series share not only a common mode of action like TGF- β inhibition, but also generate a distinct affection of other pathways by their specific off-target function. Again, BI1 stands apart from the four other indolinones with 5 significantly ranked pathways and the smallest overlap with the other indolinones. BI3 affects 15 signaling pathways and almost exclusively regulates genes involved in different cancer pathways. The indolinones BI2 and BI4 regulated genes that are significantly enriched in only 4 (BI2) and 2 (BI4) signaling pathways, respectively. However, pathways such as the Aryl Hydrocarbon Receptor Signaling and the LPS/IL-1 mediated inhibition of RXR function are also significantly ranked high for up to six compounds at all three time points, indicating a more general effect like a xenobiotic response to NCE treatment rather than a true compound specific effect. The highest numbers of significantly affected pathways are found for the two pyridopyrimidinones with 29 (Ex1) and 24 (Ex2). Additionally, genes involved in 30 out of the

Results

51 signaling pathways are exclusively regulated by Ex1 or Ex2 treatment. In line with the annotation of the molecular function of the off-targets (Figure 26), 13 out of the 51 identified pathways are known mediators of toxicity and cell death. These 13 pathways reach highest significance scores for either Ex1 or Ex2 with 8 being solely affected by the two pyridopyrimidinones indicating a cytotoxic mode of action for both of them. Interestingly, the amount of these pathways enrich over time. Pathways, such as p53 signaling and VDR/RXR Activation are significantly ranked at all time points and others, such as Death Receptor Signaling, are only found at the later time points (4 h and 12 h).

Besides cytotoxicity, these two NCEs deregulate genes involved in inflammatory processes like IL6 signaling, ERK/MAPK signaling and p38 MAPK signaling. In contrast to the findings for the pathways involved in cytotoxicity and cell death, all of these pathways can already be found at the 2 h and 4 h time point, although their significant scores slightly increase over time. These findings indicate a constant activation of pro-inflammatory processes right from the start of the treatment that are again increased upon later time points (Figure 27 -Figure 30).

Results

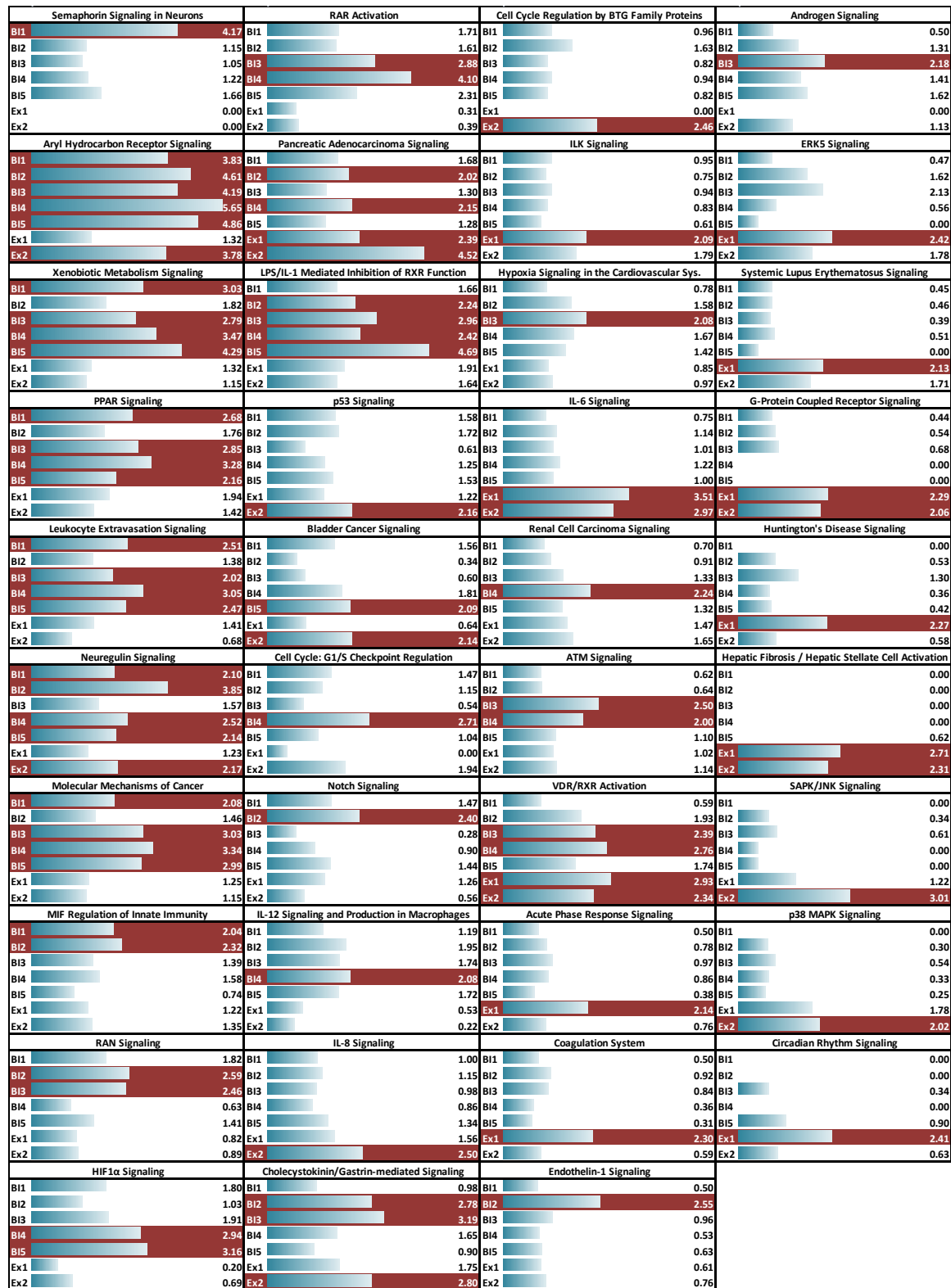


Figure 27 – Ingenuity signaling pathways (2h)

All 39 significantly ranked (Fisher's exact test: $-\log_{10} [p\text{-value}] > 2$; highlighted in red) signaling pathways from IPA analysis.

Results

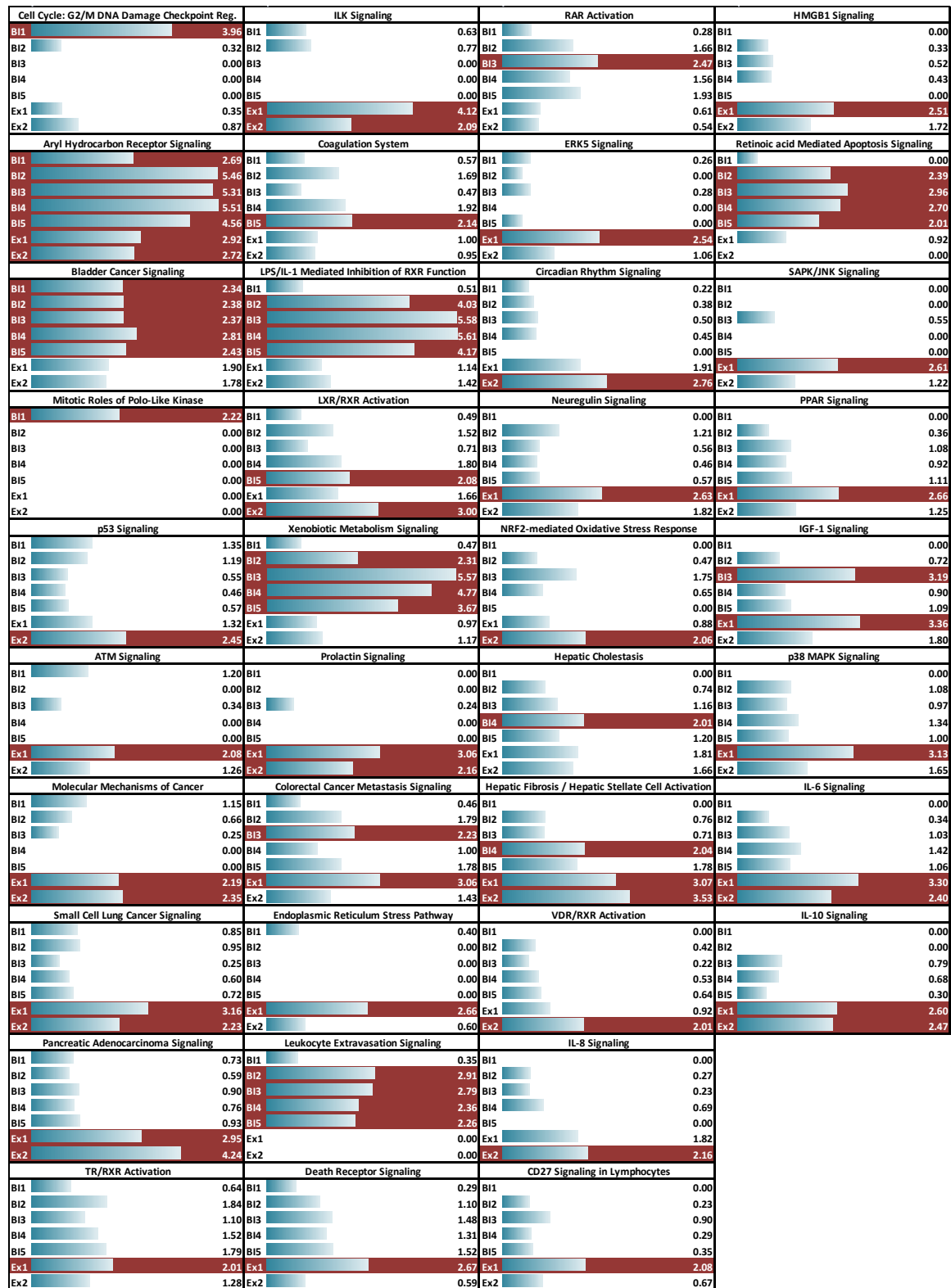


Figure 28 – Ingenuity signaling pathways (4h)

All 38 significantly ranked (Fisher's exact test: $-\log_{10} [p\text{-value}] > 2$; highlighted in red) signaling pathways from IPA analysis.

Results

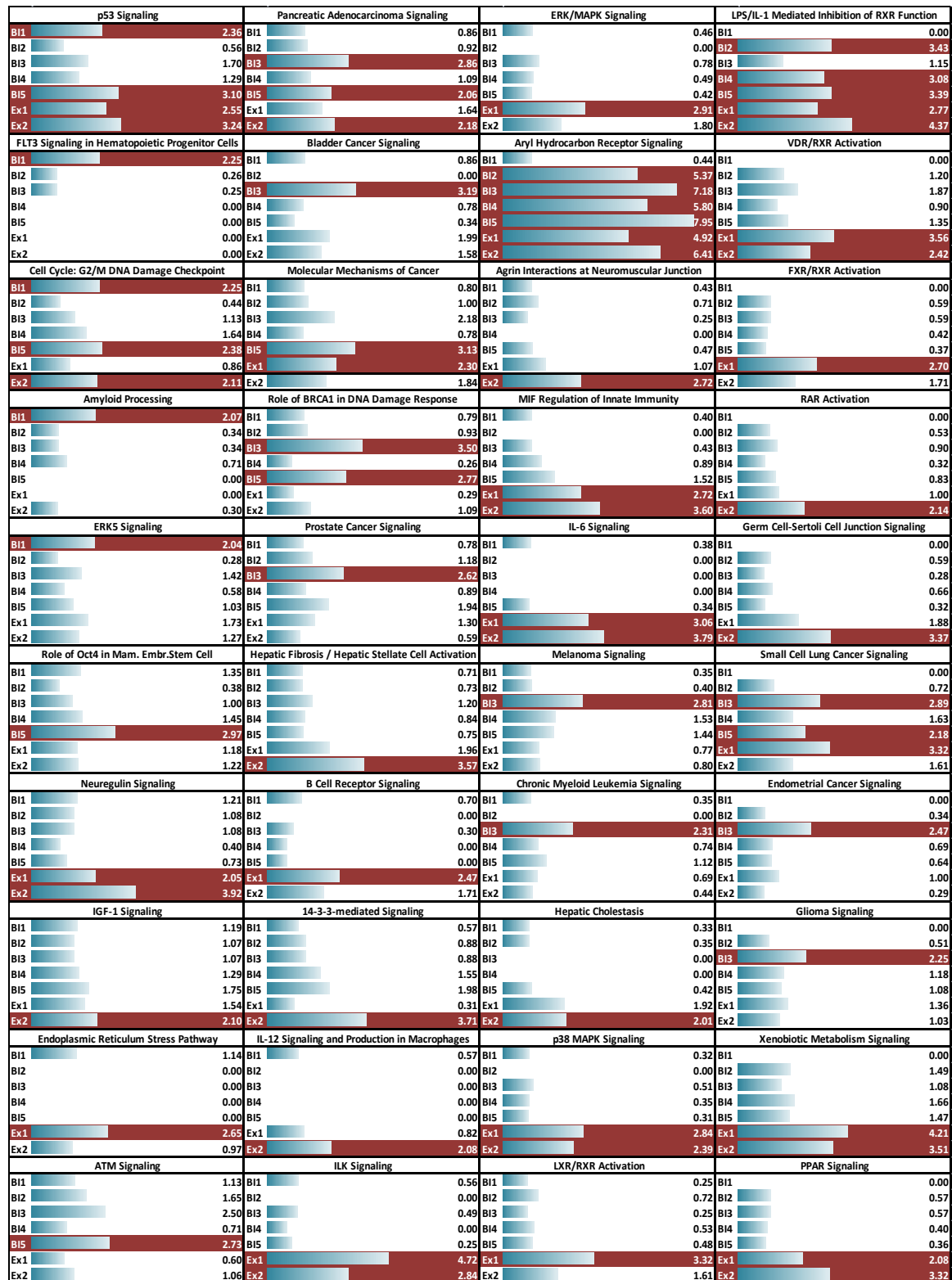


Figure 29 – Ingenuity signaling pathways (12h) – Part I

All 52 significantly ranked (Fisher's exact test: $-\log_{10} [p\text{-value}] > 2$; highlighted in red) signaling pathways from IPA analysis.

Results

Continued:

Cell Cycle: G1/S Checkpoint Regulation		Coagulation System		IL-10 Signaling		Ephrin Receptor Signaling	
B11	1.04	B11	0.50	B11	0.00	B11	0.89
B12	2.42	B12	0.48	B12	0.29	B12	0.47
B13	6.63	B13	1.21	B13	0.29	B13	0.47
B14	1.96	B14	0.37	B14	0.21	B14	0.28
B15	4.40	B15	1.68	B15	0.00	B15	0.00
Ex1	2.65	Ex1	2.26	Ex1	2.38	Ex1	1.81
Ex2	2.72	Ex2	1.61	Ex2	3.12	Ex2	2.74
Death Receptor Signaling		Docosahexaenoic Acid (DHA) Signaling		Leukocyte Extravasation Signaling		CD27 Signaling in Lymphocytes	
B11	0.93	B11	0.48	B11	0.00	B11	0.47
B12	0.30	B12	1.19	B12	1.68	B12	0.34
B13	0.00	B13	2.03	B13	1.18	B13	0.34
B14	1.18	B14	0.97	B14	0.25	B14	0.25
B15	1.10	B15	0.91	B15	0.00	B15	0.23
Ex1	3.13	Ex1	0.94	Ex1	2.44	Ex1	2.79
Ex2	1.38	Ex2	0.49	Ex2	2.15	Ex2	2.19
Role of CHK Proteins in Cell Cycle Checkpoint Ctrl		Cholecystokinin/Gastrin-mediated Signaling		Amyotrophic Lateral Sclerosis Signaling		NRF2-mediated Oxidative Stress Response	
B11	0.91	B11	0.48	B11	0.00	B11	0.00
B12	2.16	B12	0.00	B12	0.55	B12	0.00
B13	2.16	B13	0.00	B13	0.00	B13	0.00
B14	1.01	B14	0.00	B14	1.26	B14	0.00
B15	5.89	B15	0.62	B15	0.34	B15	0.00
Ex1	0.51	Ex1	1.32	Ex1	1.97	Ex1	1.49
Ex2	1.65	Ex2	2.33	Ex2	2.04	Ex2	2.05

Figure 29 – Ingenuity signaling pathways (12) – Part II

All 52 significantly ranked (Fisher's exact test: $-\log_{10} [p\text{-value}] > 2$; highlighted in red) signaling pathways from IPA analysis.

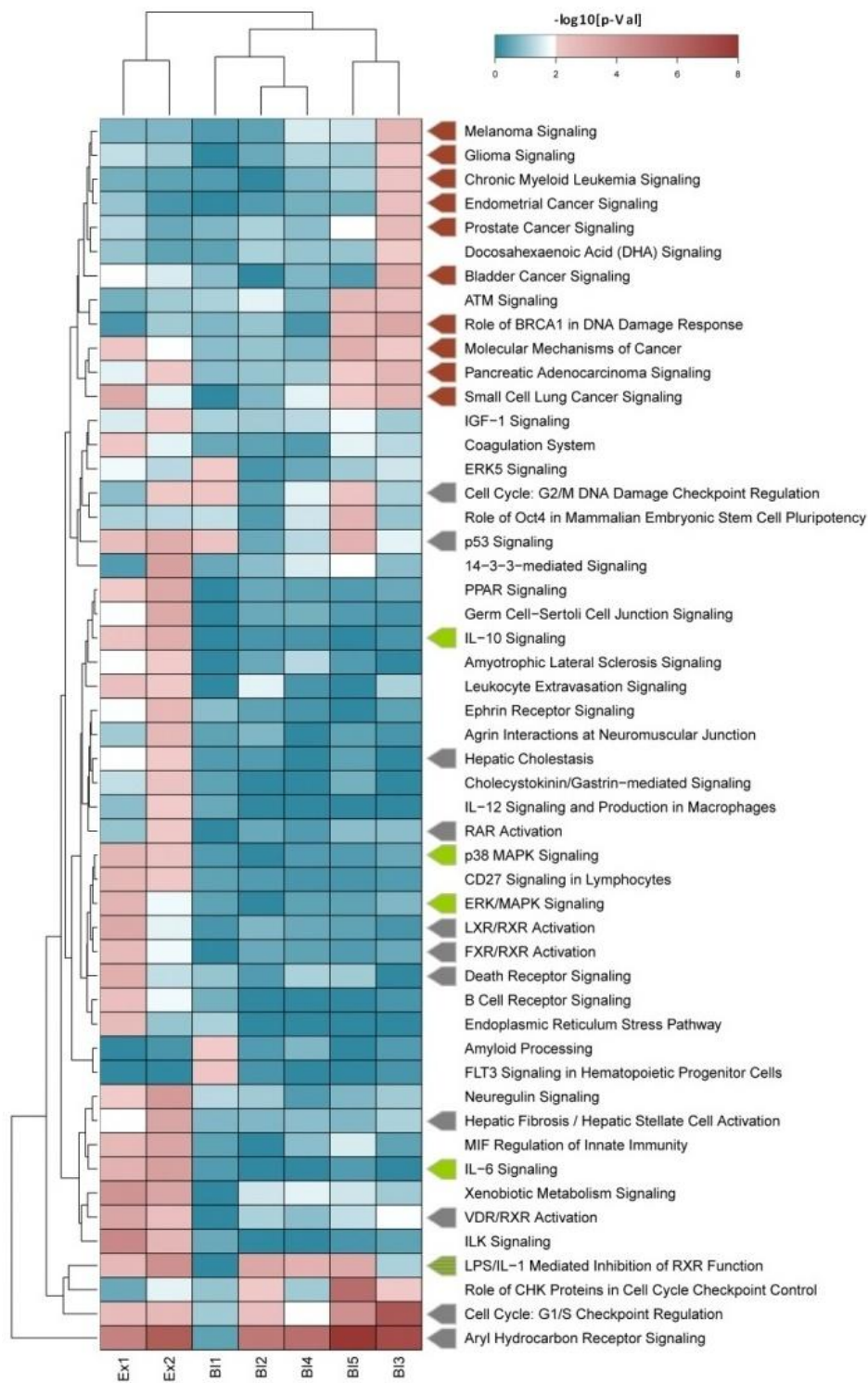


Figure 30 – Off-target effects (pathway analysis)

Ingenuity pathway analysis for the off-target genes of all seven NCEs after 12 hours: clustering of the $-\log_{10}$ [p-values] using complete linkage and manhattan distance depicts the 51 significantly ranked canonical signaling pathways. Off-target genes deregulated by BI3 treatment affect almost exclusively 10 cancer signaling pathways (red arrows). Ex1 & Ex2 off-target genes play a role in 12 pathways involved in cytotoxicity or cell death (grey arrows) and in 5 pathways involved in inflammation (green arrows). The color code defines the significance determined by Fisher's exact test as $-\log_{10}$ [p-value]: blue < 2 – not significant; white = 2 – significant & red > 2 – highly significant.

2.3.6 Wet Lab Validation

Results from the pathway analysis strongly implied different induced phenotypes after treatment with specific NCEs. However, pathway analysis tools only generate hypotheses and their proof of biological relevance must be verified. To address the accuracy of the pathway analysis, it was aimed to confirm the *in silico* generated hypotheses by experimental laboratory data.

2.3.6.1 Cytotoxicity and Cell Death

According to the expression data, both pyridopyrimidinones (Ex1 & Ex2) are involved in processes such as cell death and inflammation. To investigate several cytotoxicity parameters high content screen analysis using the high-capacity automated fluorescence imaging platform from Cellomics was performed. HaCaT cells were incubated with increasing compound concentrations (3.2 nM – 50 µM) for 24 h. Subsequently, the cells were stained with cytotoxicity cocktails and images were acquired and analyzed on the Cellomics ArrayScan II. Cells were stained using i) Hoechst DNA dye to count cell density and investigate nuclear condensation and fragmentation, ii) LysoTracker Red to analyze the amount of lysosomes per cell as an early marker for cytotoxicity, iii) Sytox Green as a membrane impermeable dye to detect loss of membrane integrity as late event for cytotoxicity. Example pictures of healthy and dying cells regarding nuclear fragmentation, membrane permeability and lysosomal mass per cell are given in Figure 31.

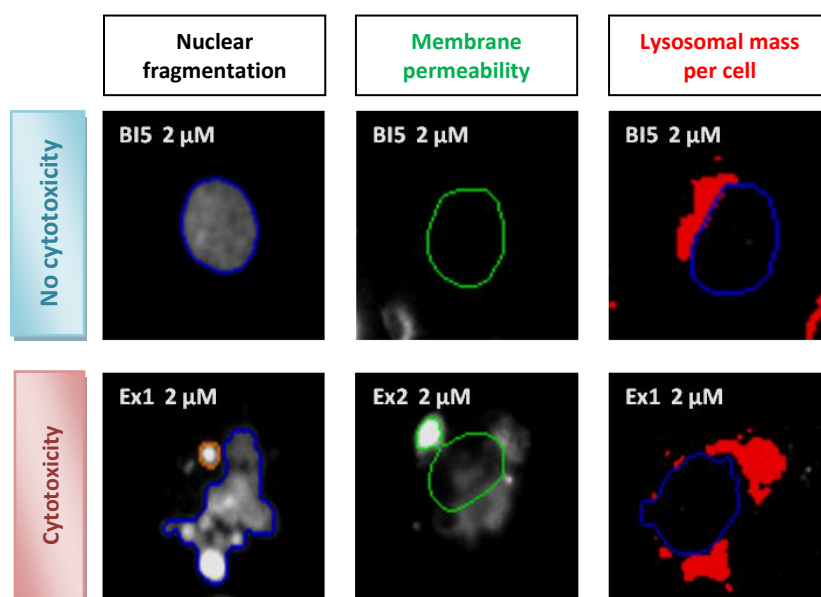


Figure 31 – Cytotoxicity parameters

Shown are examples for healthy in contrast to dying cells after NCE treatment regarding the measured parameters for nuclear fragmentation, membrane permeability and lysosomal mass per cell.

All four analyzed parameters were normalized to DMSO treated (vehicle) cells as negative control (0 %) and Valinomycin and Chloroquine treated cells as positive controls (100 %). Exemplary, the images of Ex1, BI5 and control treated cells are depicted in Figure 32. According to pathway analysis the highest toxicity is observed by treatment with Ex1 and Ex2. Cell density is strongly decreased to less than 10 % of control. Nuclear fragmentation, lysosomal mass per cell and membrane permeability are increased by 100 % for both pyridopyrimidinones and even lower concentrations of Ex2 were sufficient to raise the membrane permeability of more than 50 % of the control. Treatment with the indolinones resulted in mild toxicity effects for the treatment with BI1, BI2 and BI3 at high concentrations and almost no toxicity for BI4 and BI5 (Figure 33).

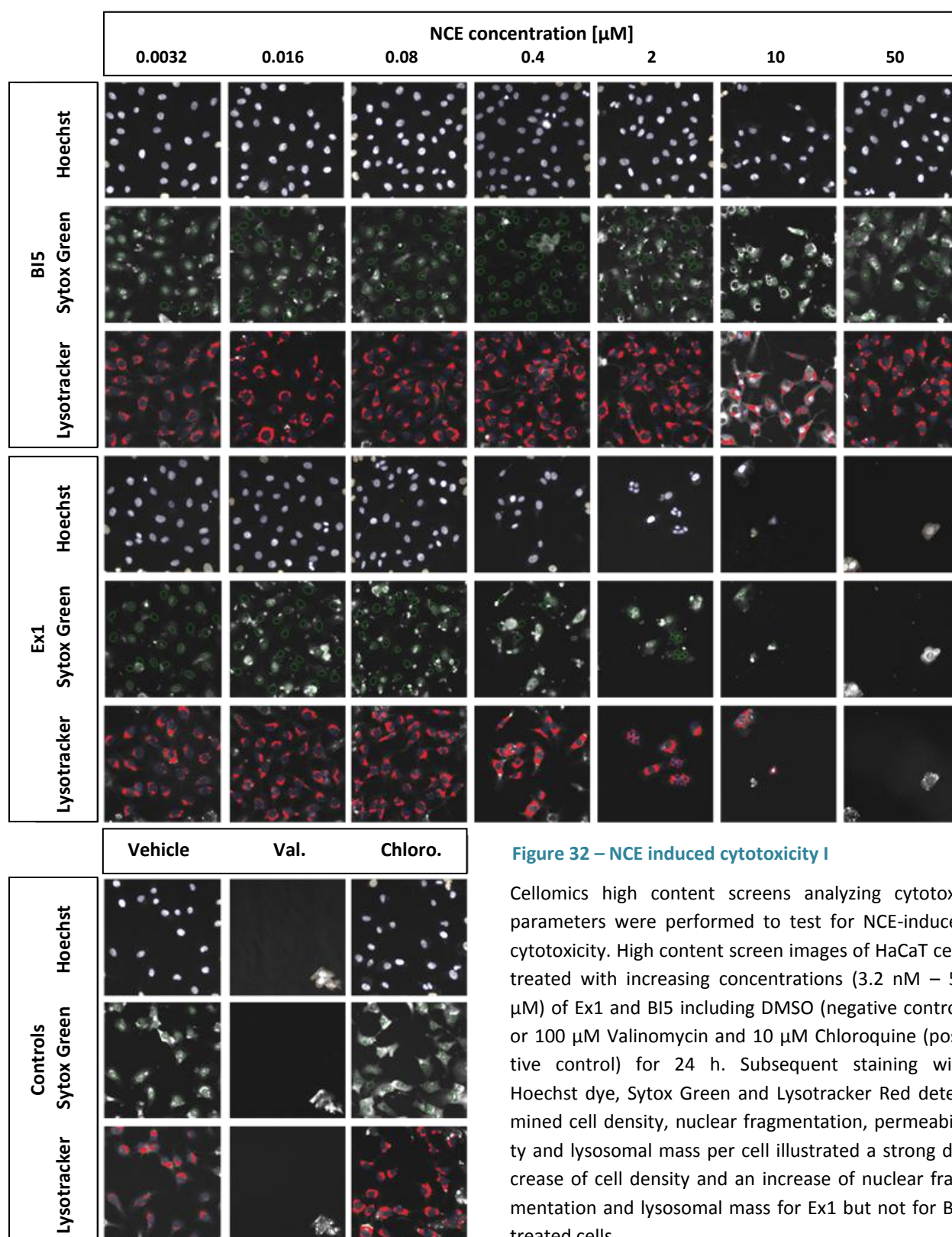


Figure 32 – NCE induced cytotoxicity I

Cellomics high content screens analyzing cytotoxic parameters were performed to test for NCE-induced cytotoxicity. High content screen images of HaCaT cells treated with increasing concentrations (3.2 nM – 50 μM) of Ex1 and BI5 including DMSO (negative control) or 100 μM Valinomycin and 10 μM Chloroquine (positive control) for 24 h. Subsequent staining with Hoechst dye, Sytox Green and Lysotracker Red determined cell density, nuclear fragmentation, permeability and lysosomal mass per cell illustrated a strong decrease of cell density and an increase of nuclear fragmentation and lysosomal mass for Ex1 but not for BI5 treated cells.

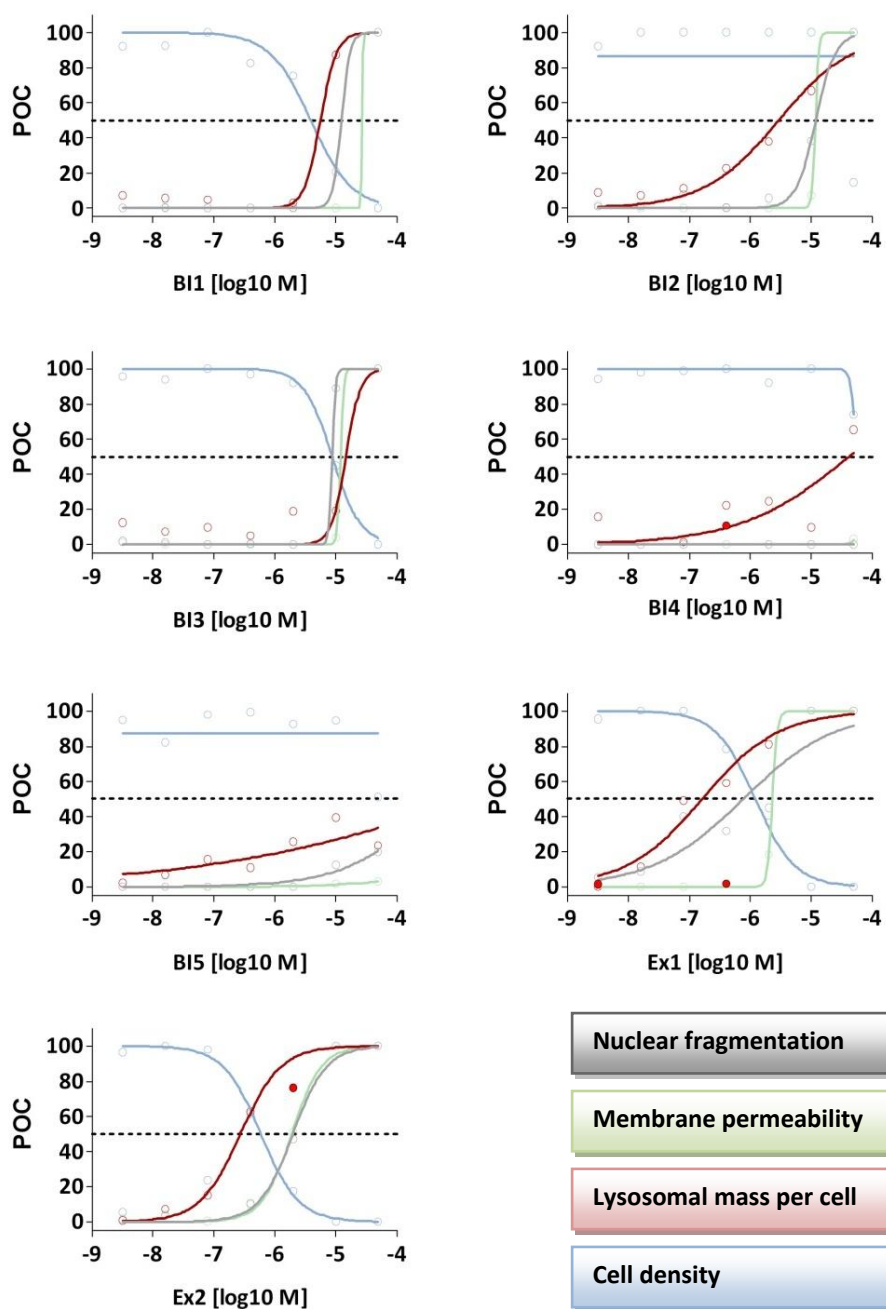


Figure 33 – NCE induced cytotoxicity II

Results of dose response experiments for all seven NCEs for cell density (blue), nuclear fragmentation (grey), permeability (green) and lysosomal mass (red) obtained from Cellomics high content screen analysis. Values of NCE treated cells are compared to DMSO treated cells and shown as percent of control (POC). Outlier data points are shown as filled red circles.

Results

To analyze the mode of cell death, Caspase-3 activation assays were performed to distinguish between apoptosis and necrosis, since activation of this executioner caspase is a clear marker for apoptotic cell death. Again, HaCaT cells were treated with 2 μ M of each compound for 24 h and subsequently Caspase-3 activation was detected using a Caspase-3 Detection Kit (Calbiochem) and quantified by flow cytometry. Among all NCEs tested, only treatment with Ex1 resulted in an activation of Caspase-3 with approximately 30 % positive cells (Figure 34b). This is in line with the results of the pathway analysis, in which only the off-target signature of Ex1 exceeded the significance threshold for Death Receptor Signaling after 4 h (Figure 28, Figure 34a) and was further increased after 12 h (Figure 29, Figure 34a).

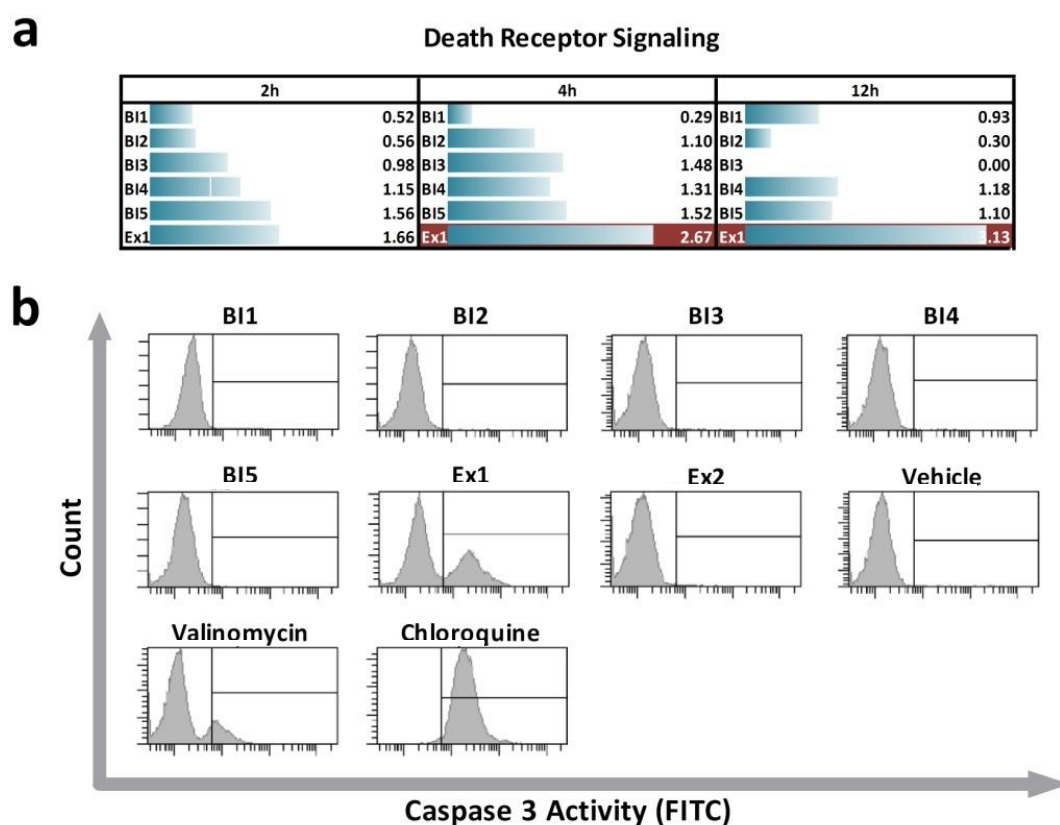


Figure 34 – Death Receptor Signaling I

a: Ingenuity pathway analysis results for Death Receptor Signaling obtained for the NCE off-targets. Only Ex1 treatment for 4 h and 12 h achieves the significance threshold of $-\log_{10} [\text{p-value}] > 2$ (Fisher's exact test; red). b: HaCaT cells were treated with 2 μM of each NCE and incubated for 24h. Caspase-3 activation was analyzed. A significant signal was identified only after treatment with Ex1 (30 %) or with Valinomycin (15 %) and Chloroquine (100 %).

An overlay of the off-targets of Ex1 with the canonical Death Receptor Signaling pathway revealed that six genes are deregulated after 12 h of treatment (Figure 35). NF κ B1a, a subunit of the I κ k-complex, is down-regulated, while five pro-apoptotic genes, such as Apaf1 as part of the apoptosome, the death receptors DR3 and 6, the death receptor ligand APO2L and the initiator Caspase-9, are up-regulated. Taken together, these findings clearly demonstrate that it was not only possible to predict the compound's cytotoxicity based on mRNA profiles but also its apoptotic mode of action.

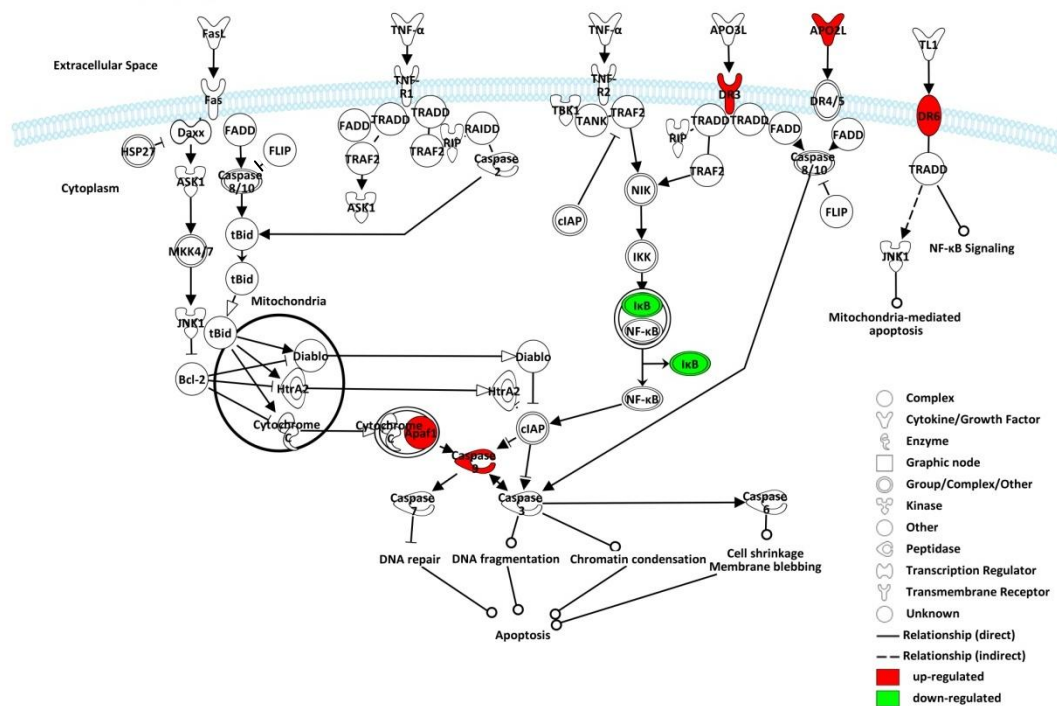


Figure 35 – Death Receptor Signaling II

Canonical Death Receptor Signaling pathway (Ingenuity Pathway Analysis) overlaid with the off-target of Ex1. Color code: red – up-regulated gene; green – down-regulated gene.

2.3.6.2 Inflammation

Pathway analysis not only labeled the pyridopyrimidinones for cell death induction, but also for affection of inflammatory mechanisms. To proof this prediction, HaCaT cells treated with each compound were analyzed for the induction of pro-inflammatory cytokines (IL1 β , TNF α , IL8 and IL6). While no significant alteration in release of IL1 β , TNF α and IL8 was observed, the IL6 levels were dose-dependently increased after treatment with the pyridopyrimidinones Ex1 and Ex2. Compared to DMSO treated control cells, 10 μ M of Ex2 increased IL6 secretion by factor 5 and Ex1 treatment at the same concentration even resulted in a 25-fold increase (Figure 36).

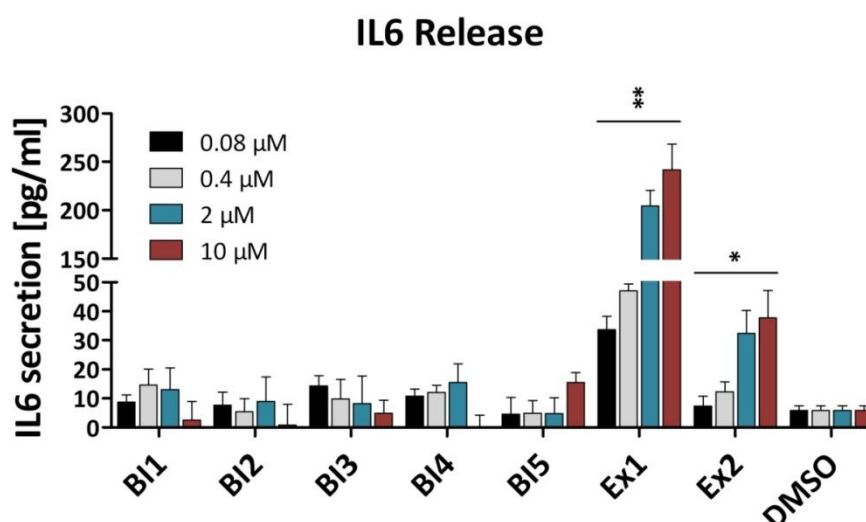


Figure 36 – IL6 release

HaCaT cells were treated with increasing NCE concentrations and DMSO for 12 h. IL6 levels were app. 25-fold increased after Ex1 treatment and 5-fold increased after Ex2 treatment compared to DMSO treated cells. Student t-test was used to calculate the significance compared to DMSO treated cells (* < 0.01 & ** < 0.001). All error bars indicate the standard deviation of $n=3$.

2.3.7 Kinase Profiling

One pivotal issue of kinase inhibitors is cross-reactivity with other kinases which may contribute one source of off-targets. *In vitro* kinase profiling is the state-of-the-art method to examine selectivity of kinase inhibitors. The Ingenuity Pathway Analysis database was used to extract the literature described downstream targets for 239 in HaCaT cells expressed kinases. For 147 kinases 807 non-redundant downstream targets had been described and annotated. The downstream targets were used as surrogate markers and overlaid with the NCE's off-target lists to assign off-target genes to off-target kinases (Figure 38). To proof the predictivity of our model all compounds were tested against 239 kinases available in the biochemical SelectScreen™ Kinase Profiling (Invitrogen) at concentrations of 2 μM and 200 nM. All kinases inhibited by at least 90 % at 2 μM conc. and additionally by at least 50 % at 200 nM were selected as off-target kinases (Table 11). Most off-target kinases were identified for

Results

BI1 (75) and Ex1 (60). All other NCEs showed a higher selectivity with only 21 (BI2), 17 (BI3), 12 (BI4), 15 (BI5) and 14 (Ex2) kinases inhibited by the respective compound (Figure 37b, Table 11). For BI1 84 of the 366 known surrogate marker genes were found to be regulated. A comparably good ratio was also identified for the two pyridopyrimidinones with 84 out of 361 (Ex1) and 55 out of 223 (Ex2). A summary of the data is shown in Figure 37. Integrating all criteria, the identification of kinase selectivities was limited, as shown for BI2 in Figure 38. Although, various annotated surrogate marker genes were identified as off-targets, fortifying the used identification procedure, a clear association to a specifically upstream-acting kinase was often not possible since too many surrogate markers were redundantly identified to act downstream of several receptor kinases. The cellular system might be optimized e.g. in regard to the addition of receptor ligands, but it will not replace testing NCEs in biochemical assays for kinase selectivity.

Results

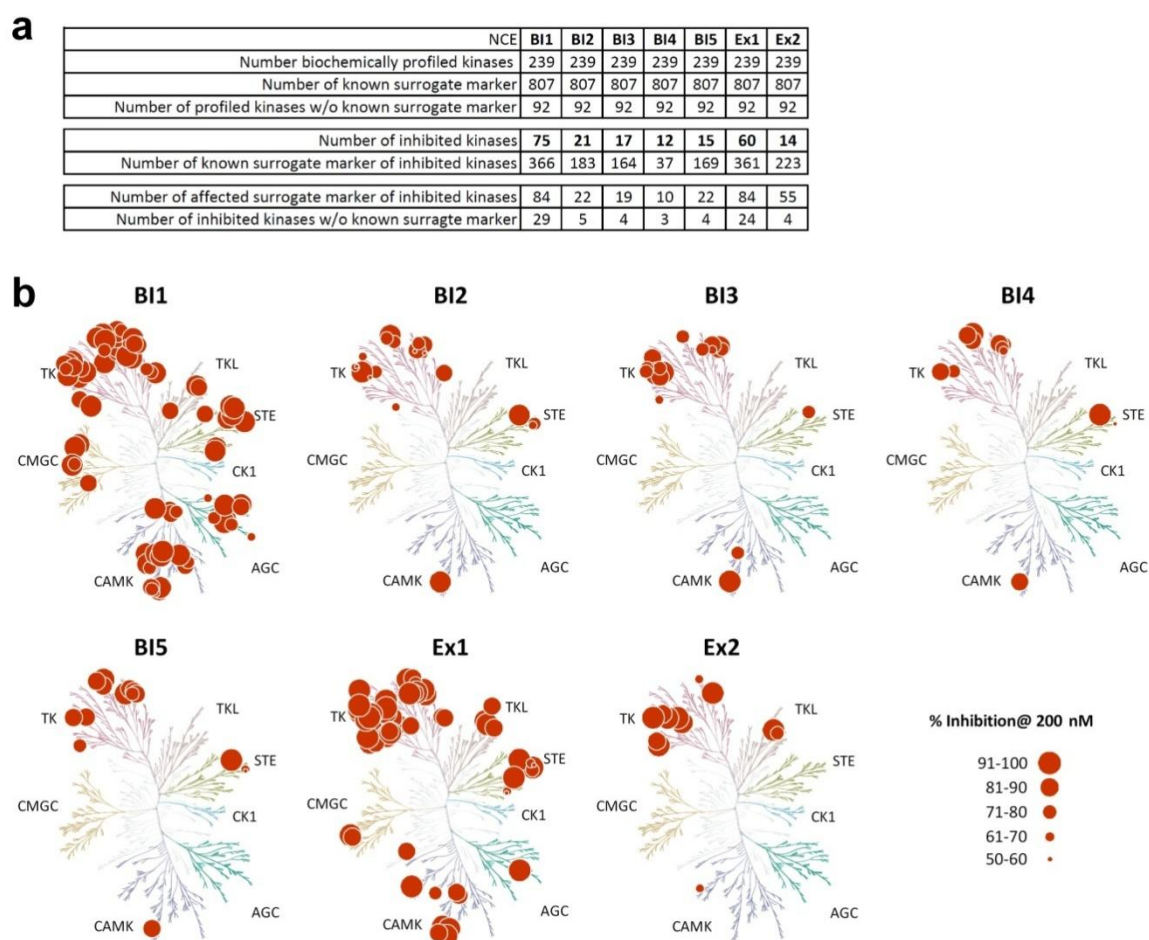


Figure 37 – Kinase Selectivity I

a: each compound was profiled against a panel of 239 protein kinases and the number of kinases inhibited by each compound is shown (Table 11). No surrogate marker (e.g. literature known downstream target of a given kinase) were identified for 92 enzymes, whereas 807 surrogate markers were identified for 147 of the enzymes. Based on the kinase inhibition profile of each compound, the expression of these kinases in HaCaT cells and the availability of surrogate markers the number of potentially effected surrogate marker genes was predicted. b: human kinome dendrograms showing the NCEs' kinase specificity profiles. Circle size corresponds to the precentage inhibition of the kinase at 200 nM concentration. AGC – Containing PKA, PKG, PKC families; CAMK – Calcium/calmodulin-dependent protein kinase; CK1 – Casein kinase 1; CMGC – Containing CDK, MAPK, GSK3, CLK families; STE – Homologs of yeast Sterile 7, Sterile 11, Sterile 20 kinases; TK – Tyrosine kinase; TKL – Tyrosine kinase-like. Kinome dendrograms used for visualization were shown with permission from Cell Signaling Technology, Inc. (<http://www.cellsignal.com>).

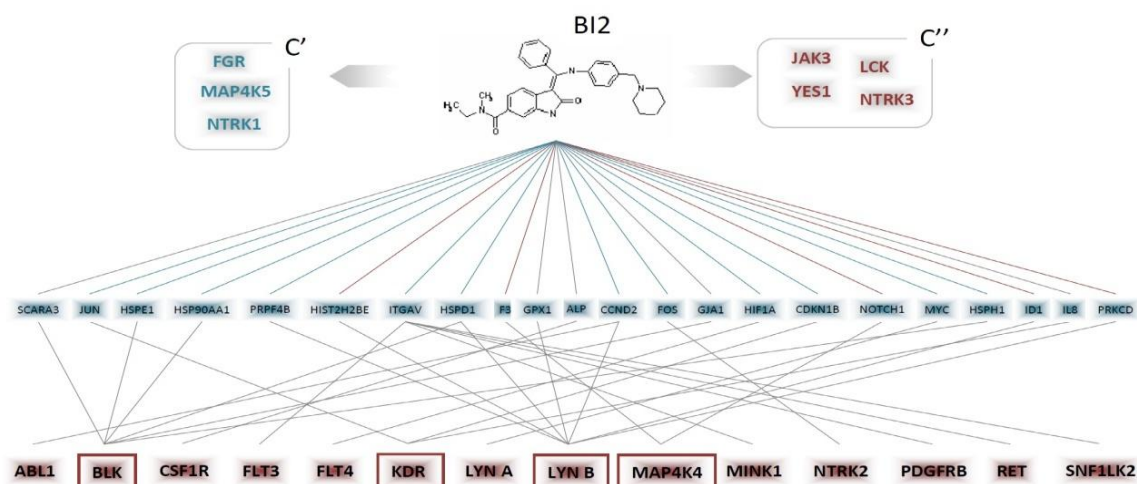


Figure 38 – Kinase Selectivity II

Projections of BI2-specifically regulated kinase surrogate marker genes. Indicated are all 21 inhibited kinases: 3 kinases with no affection of known surrogate marker genes (c'), 4 kinases of which no surrogate markers are described (c'') and 14 kinases with deregulated surrogate markers genes (blue line = transcriptional down-regulation, red line = transcriptional up-regulation, red box = *in silico* predicted BI2-specific kinase hits).

Table 11 – Kinase Selectivity

NCE	Kinase	% Inhibition [2 µM]	% Inhibition [200nM]	NCE	Kinase	% Inhibition [2 µM]	% Inhibition [200nM]	NCE	Kinase	% Inhibition [2 µM]	% Inhibition [200nM]
B11	ABL1	97	89	B11	STK4 (MST1)	103	102	Ex1	BMX	99	97
B11	ALK	97	99	B11	TYK2	95	80	Ex1	BRAF	102	100
B11	AMPK A1/B1/G1	98	92	B11	YES1	104	99	Ex1	BTX	98	100
B11	AURKA (Aurora A)	98	87	B12	ABL1	96	67	Ex1	CSF1R (FMS)	109	105
B11	AURKB (Aurora B)	101	80	B12	BLK	93	57	Ex1	EPHA1	91	99
B11	AURKC (Aurora C)	93	86	B12	CSF1R (FMS)	97	81	Ex1	EPHA2	98	100
B11	AXL	98	88	B12	FGFR	99	75	Ex1	EPHA5	98	101
B11	BLK	98	92	B12	FLT3	96	70	Ex1	EPHA8	97	99
B11	BRAF	98	87	B12	FLT4 (VEGFR3)	100	90	Ex1	EPHB1	95	101
B11	BRSK1 (SAD1)	93	76	B12	JAK3	94	90	Ex1	EPHB4	96	98
B11	CAMK2D (CaMKII delta)	100	81	B12	KDR (VEGFR2)	98	80	Ex1	FES (FPS)	102	86
B11	CDK7/cyclin H/MNAT1	97	82	B12	LCK	101	99	Ex1	FGFR1	100	97
B11	CHEK1 (CHK1)	94	92	B12	LYN A	96	58	Ex1	FGFR2	101	99
B11	CLK1	97	98	B12	LYN B	97	66	Ex1	FGFR3	98	86
B11	CLK2	98	99	B12	MAP4K4 (HGK)	102	78	Ex1	FGFR4	92	79
B11	FER	102	94	B12	MAP4K5 (KHS1)	100	104	Ex1	FGFR	96	105
B11	FGFR1	97	89	B12	MINK1	101	62	Ex1	FLT3	96	85
B11	FGFR2	103	94	B12	NTRK1 (TRKA)	100	86	Ex1	FLT4 (VEGFR3)	100	94
B11	FGFR3	97	76	B12	NTRK2 (TRKB)	106	85	Ex1	FRK (PTK5)	94	101
B11	FGFR	101	102	B12	NTRK3 (TRKC)	103	99	Ex1	FYN	98	93
B11	FLT1 (VEGFR1)	95	70	B12	PDGFRB (PDGFR beta)	96	58	Ex1	JAK2	98	94
B11	FLT3	100	101	B12	RET	95	57	Ex1	JAK3	92	83
B11	FLT4 (VEGFR3)	102	104	B12	SNF1LK2	103	93	Ex1	KDR (VEGFR2)	109	88
B11	FYN	97	90	B12	YES1	95	61	Ex1	LCK	101	96
B11	GRK7	92	61	B13	ABL1	93	67	Ex1	LYN A	100	101
B11	GSK3A (GSK3 alpha)	99	80	B13	BLK	97	85	Ex1	LYN B	90	102
B11	GSK3B (GSK3 beta)	99	93	B13	CSF1R (FMS)	94	82	Ex1	MAP2K1 (MEK1)	92	58
B11	HCK	100	80	B13	FGFR	99	91	Ex1	MAP2K2 (MEK2)	98	68
B11	JAK2	91	74	B13	FLT3	93	66	Ex1	MAP3K9 (MLK1)	106	89
B11	JAK3	101	97	B13	FYN	93	74	Ex1	MAP4K4 (HGK)	103	92
B11	KDR (VEGFR2)	100	98	B13	LCK	100	100	Ex1	MAP4K5 (KHS1)	103	100
B11	LCK	103	98	B13	LYN A	97	79	Ex1	MARK11 (p38 beta)	95	90
B11	LRRK2	98	88	B13	LYN B	97	78	Ex1	MAPK14 (p38 alpha)	101	99
B11	LYN A	99	87	B13	MAP4K5 (KHS1)	93	77	Ex1	MARK1 (MARK)	90	83
B11	LYN B	99	91	B13	MELK	93	80	Ex1	MARK3	96	92
B11	MAP2K1 (MEK1)	101	93	B13	NTRK3 (TRKC)	100	78	Ex1	MARK4	100	93
B11	MAP2K2 (MEK2)	100	98	B13	PDGFRA (PDGFR alpha)	99	91	Ex1	MINK1	104	77
B11	MAP3K8 (COT)	95	85	B13	PDGFRB (PDGFR beta)	94	76	Ex1	MST4	97	62
B11	MAP4K2 (GCK)	102	99	B13	RET	98	79	Ex1	MYLK2 (skMLCK)	107	95
B11	MAP4K4 (HGK)	104	102	B13	SNF1LK2	103	96	Ex1	NEK2	106	84
B11	MAP4K5 (KHS1)	101	105	B13	YES1	100	90	Ex1	PDGFRA (PDGFR alpha)	99	101
B11	MARK1 (MARK)	102	83	B14	CSF1R (FMS)	95	87	Ex1	PDGFRB (PDGFR beta)	100	91
B11	MARK2	92	63	B14	FGFR	93	80	Ex1	PRKCN (PKD3)	90	88
B11	MARK3	105	91	B14	FLT3	91	69	Ex1	PRKD1 (PKC mu)	96	92
B11	MARK4	103	95	B14	FLT4 (VEGFR3)	94	71	Ex1	PRKD2 (PKD2)	95	88
B11	MELK	93	99	B14	KDR (VEGFR2)	95	81	Ex1	PTK6 (Brk)	94	107
B11	MERTK (cMER)	102	98	B14	LCK	99	89	Ex1	RAF1 (cRAF) Y340D Y341D	103	98
B11	MINK1	103	93	B14	MAP4K4 (HGK)	99	59	Ex1	RET	99	99
B11	MUSK	90	75	B14	MAP4K5 (KHS1)	99	93	Ex1	SNF1LK2	100	104
B11	NTRK1 (TRKA)	101	98	B14	NTRK1 (TRKA)	95	83	Ex1	SRC	91	102
B11	NTRK2 (TRKB)	102	98	B14	NTRK2 (TRKB)	97	89	Ex1	SRMS (Srm)	94	90
B11	NTRK3 (TRKC)	103	96	B14	NTRK3 (TRKC)	103	94	Ex1	STK22D (TSSK1)	92	74
B11	PDGFRA (PDGFR alpha)	97	87	B14	SNF1LK2	100	82	Ex1	STK24 (MST3)	98	58
B11	PDGFRB (PDGFR beta)	98	87	B15	ABL1	90	75	Ex1	SYK	90	87
B11	PHKG1	91	72	B15	CSF1R (FMS)	94	84	Ex1	TAOK2 (TAO1)	103	96
B11	PKN1 (PRK1)	98	86	B15	FGFR	94	84	Ex1	YES1	96	102
B11	PRKCA (PKC alpha)	96	67	B15	FLT3	96	90	Ex2	ABL1	91	100
B11	PRKCN (PKD3)	95	104	B15	FLT4 (VEGFR3)	93	74	Ex2	ABL2 (Arg)	94	101
B11	PRKCO (PKC theta)	100	73	B15	KDR (VEGFR2)	95	81	Ex2	BRAF	95	74
B11	PRKD1 (PKC mu)	102	99	B15	LCK	98	90	Ex2	BTX	91	81
B11	PRKD2 (PKD2)	98	97	B15	MAP4K4 (HGK)	100	69	Ex2	CSF1R (FMS)	105	102
B11	PRKG1	99	79	B15	MAP4K5 (KHS1)	101	99	Ex2	EPHA2	94	95
B11	RAF1 (cRAF) Y340D Y341D	97	91	B15	MINK1	94	60	Ex2	EPHA8	93	93
B11	RET	102	102	B15	NTRK1 (TRKA)	95	82	Ex2	FGFR2	94	61
B11	RPS6KA1 (RSK1)	99	91	B15	NTRK2 (TRKB)	98	88	Ex2	FGFR	99	102
B11	RPS6KA2 (RSK3)	100	97	B15	NTRK3 (TRKC)	101	93	Ex2	FYN	97	84
B11	RPS6KA3 (RSK2)	94	85	B15	RET	100	93	Ex2	LCK	98	95
B11	SNF1LK2	103	88	B15	SNF1LK2	100	83	Ex2	LYN A	97	98
B11	SRC	97	87	Ex1	ABL1	94	100	Ex2	MYLK2 (skMLCK)	102	65
B11	STK22D (TSSK1)	106	93	Ex1	ABL2 (Arg)	99	105	Ex2	RAF1 (cRAF) Y340D Y341D	98	93
B11	STK25 (YSK1)	104	99	Ex1	ACVR1B (ALK4)	94	90				
B11	STK3 (MST2)	100	98	Ex1	BLK	96	100				

Ph.D. Thesis Patrick Baum

3. Discussion

Phenocopy – A Strategy to Qualify Chemical Compounds during Hit-to-Lead and Lead Optimization

Since the approval of Imatinib (Gleevec) in 2001, the first marketed kinase inhibitor, many additional kinase inhibitors have been advanced into clinical development. The most advanced kinase programs in research and development are aimed at the treatment of various cancers. However, additional therapeutic applications like immunological, metabolic- or infectious diseases and also the treatment of central nervous system disorders by kinase inhibitors are under investigation^{31, 111, 112}. During the optimization of kinase inhibitors one often has to cope with challenges like the improvement of kinase selectivity^{113, 114}. In combination with the overall high attrition rates of new drug candidates⁵³ there is a need for new strategies that support and optimize the drug discovery process. In the present study, an approach is introduced to support the drug development process by enabling a highly flexible and highly resolved ranking of the NCEs and by providing additional information of the drug target biology.

3.1 NCE Ranking

So far, the *in vitro* biological evaluation of NCEs has often been based on biochemical and cellular potencies as well as on the selectivity of the respective NCE. This limited view may result in wrong decisions for further time and cost consuming processes, such as *in vivo* experiments. In the present study, a workflow was established to alleviate the lead identification and optimization of NCEs in general and kinase inhibitors in particular by elucidating the mechanism of action of both the target and the NCE. Thereby, more knowledge about drug candidates is obtained at an early stage of drug discovery and several new categories for their qualification are available (Figure

39). Although the rating of each category remains subjective, one can clearly distinguish between good and bad results within a group of tested compounds. Especially the exclusion criteria (shown in red; Figure 39) are important breakpoints for the ranking since they are potential sources for failures of the respective NCE later on in the process.

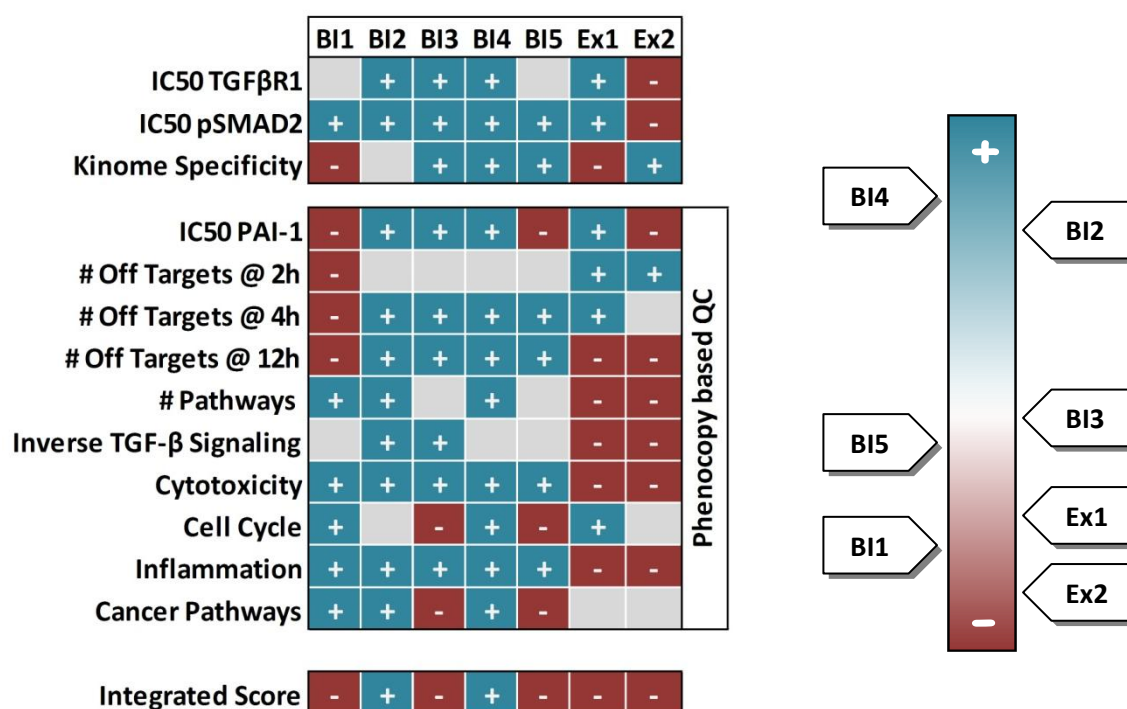


Figure 39 – NCE ranking

Quality parameters used to gauge the seven NCEs. The Phenocopy strategy introduces ten additional parameters dealing with potency, off-target numbers and affected pathways. NCEs are ranked from blue (good) to red (bad). Integration of all parameter scores identifies BI4 and BI2 as superior to all other NCEs.

In case of the tested inhibitors the decision based on classical parameters such as potency or kinase specificity can be misleading or fall far short. Besides Ex2 all other NCEs are potent inhibitors with comparable IC₅₀ values in cellular assays and only minor distinction in their biochemical activity. Also the kinome specificity profiling is not sufficient to arrive at a final decision, since only BI1 and Ex1 inhibit a wide panel

of other kinases. Furthermore, the Phenocopy data strongly indicates that off-target effects do not only derive from additionally inhibited kinases in line with the fact that within the human genome over 2,000 other nucleotide-dependent enzymes can be found¹¹⁵ which may be potentially affected by NCEs blocking an ATP binding site. In addition, identification of bioactive compounds revealed a high degree of promiscuity for kinases inhibitors with GPCRs and phosphodiesterases¹¹⁶. Hence, an approach using the Phenocopy strategy will deliver a wider view on the NCEs' selectivity. However, it can not replace a kinase selectivity screen, since mapping of the off-targets to published kinase surrogate markers were not sufficient to indentify all off-target kinases or did not result in unique predictions (Figure 38). This is potentially due to the experimental design and the current state of knowledge. The correlation of the *in vitro* predicted kinase inhibition with the off-target effects requires some criteria to be fulfilled: i) the kinase has to be expressed in the cellular system, ii) the signaling pathway must be functional, which iii) depends on the availability of appropriate ligands in the *in vitro* system. As the cells in our study had been starved for compound and/or TGF- β profiling, these criteria might have been only partially met. Finally, iv) surrogate markers have to be described. Nevertheless, the comparison of the compounds' off-target effects with the kinase profiling fortifies the Phenocopy findings and supports the chosen analysis procedure since a good overlap between surrogate markers and off-targets is found in many cases, where sufficient surrogate markers are published.

By evaluating of TGF- β R1 inhibitors, it was possible to clearly differentiate the indolinone chemotype from the pyridopyrimidinones in several parameters. Furthermore, even within the indolinone cluster differences between compounds with various

Discussion

decorations were identified. This can help at different steps during the drug development process. The differentiation between chemotypes helps to qualify the different hit classes after high-throughput screening that are promoted to the hit-to-lead phases. Later on, the differentiation between the compounds of a particular chemotype is suitable to support lead optimization programs.

As above mentioned, the presented approach enables a wider view on the compounds selectivity in a cellular context by the identification of the total amount of target-independent regulated genes and also sheds light on the chronological sequence when these regulations occurs. In earlier studies, compound-dependent gene regulation is either identified by treatment of the cells with only one concentration in comparison to untreated cells followed by the application of fold change and/or p-value cut-offs⁵⁷ or by the treatment with a series of concentrations, followed by the selection of dose-dependently regulated genes⁷⁰. The Phenocopy data however demonstrates that on the one hand, the use of only one concentration is not sufficient to identify all off-targets and might be too less stringent dependent on the used concentration. On the other hand, an off-target identification based on dose-dependency will only select that subgroup of regulated genes that are not subjected to integrated effects. Thus, the introduction of a selection procedure where different NCE concentrations are used in presents as well as in absence of TGF- β guarantees a sharp classification and enables a high resolution of detected genes.

The strategy can also help to interpret the identified off-target effects and offers the possibility to assign the regulated genes to relevant biological processes and networks. Good results that were reproduced by wet lab experiments were thereby achieved using Ingenuity Pathway Analysis (Ingenuity Systems®). However, the anno-

Discussion

tations of canonical pathways are superior to the analysis of the molecular function. Although the annotation of the molecular function of the regulated genes can help to understand the compound effects, e.g. in case of the predicted cytotoxicity of both pyridopyrimidinones, its informative value is limited since some of the categories are too general. Thus, it can not be consulted for the NCE ranking. In contrast, the pathway analysis can be used for the ranking in a flexible format by the respective scientist by defining context relevant processes or just by prioritizing the compounds in terms of the absolute number of affected processes or pathways. In terms of TGF- β inhibition for instance one goal is to reduce inflammation processes triggered by this cytokine⁸⁹ as one potential strategy to treat fibrotic diseases or cancer. Thus, the predicted and confirmed pro-inflammatory properties of both pyridopyrimidinones (Figure 30, Figure 36) are the opposite of the desired effect making both NCEs inferior to the indolinones. Furthermore, relevant and unwanted processes are regulation of growth and proliferation and obviously induction of cell death (Figure 30).

Finally, the combination of TGF- β and off-target signatures revealed that some compounds regulate genes inverse to the desired therapeutic effect (Figure 23). This can potentially affect the efficacy of the treatment with this NCE. This is of particular interest since three times more substances undergo attrition due to a lack of efficacy than due to clinical safety reasons⁵³. Especially both pyridopyrimidinones regulated 217 (Ex1) and 317 (Ex2) TGF- β dependent genes in the opposite direction to the desirable treatment effect. In contrast, the indolinones only affected a lower number (BI3: 42; BI2: 57, BI1: 70; BI5: 81; BI4: 101) of these genes.

According to the ten introduced Phenocopy criteria, a couple of compounds revealed liabilities through downstream inhibition of PAI-1 transcription (BI1, BI5 & Ex2), at the regulation of off-target genes per se (BI1, Ex1 & Ex2), at the affection of inverse TGF- β signaling (Ex1 & Ex2), at the induction of cell death (Ex1), at acting as pro-inflammatory stimuli (Ex1 & Ex2) and as promoting cellular growth and induction of cancer pathways (BI3 & BI5) (Fig.6). An integration of all obtained information recommends the use of BI4 and BI2 for further optimization due to superior overall performance of these two drug candidates (Figure 39).

3.2 Chemical genomic profiling

So far, microarray technology has been successfully applied during the drug development process for target discovery by profiling disease models, for target validation by profiling alterations caused by disease-related genes, for elucidating drug metabolism by measuring transcriptional changes of known drug metabolizing genes in rat livers or human hepatocytes and to address drug safety in toxicogenomics approaches. However, only few approaches have been tempted to fill the gap between target validation and drug metabolism and aimed to support the hit-to-lead or lead optimization processes. In fact, gene expression signatures have been used to functionally annotate and characterize small molecules in yeast and in mammalian cells^{57, 117, 118}. However, these approaches mainly focused on the identification of new NCEs directed against a given target, or to build novel connections to a disease, but not to obtain an in depth analysis of the off-target effects. In the present study, several optimized parameters are introduced to achieve a comprehensive qualification of

Discussion

the investigated compounds. First, the screening platform was chosen by the use of a relevant cellular system functionally expressing the drug target and its downstream signaling. Second, various time points and concentrations were monitored. Third, siRNAs against TGF- β R1 were used as an additional target modulation technology to confirm the results obtained with the NCEs. By combining those data it was demonstrated that the off-target signatures can be used to identify the most selective NCE among the tested compounds and to detect unwanted off-target effects such as induction of pro-inflammatory processes or of death receptor signaling. The data also allows identifying the target promiscuity of the NCE. These polypharmacological approaches, most notably discussed in fields of cancer treatment³², can not be faced with conventional single target-based assays but need approaches containing multi-parallel readouts for NCE characterization. Such a strategy will also help to accumulate an iterative knowledge about both the drug candidates itself and the structural classes. The drug candidates' off-target signatures can be overlaid with other databases containing drug-dependent gene signatures like the Connectivity Map⁵⁷. Integration of additional data sources will further characterize the NCEs by flagging them for potential side effects and the identification of desirable pharmacology profiles or even find a repositioning idea for other indications.

Nevertheless, there are some limitations inherently comprised in each chemical genomics approach. Most obviously, cells that grow on plastic in a laboratory are very different from tissues in a whole organism. Therefore, effects modulated by a specific microenvironment or based on the cooperation of multiple cell types can simply not be accessed by such approaches. Some of these limitations could be overcome by using more complex cell culture systems such as primary cells or mixed co-culture

systems, containing more than one cell type. The downside of such approaches however is the increased variability and the reduced throughput in combination with the rising costs of such cell systems. Another challenge of chemical genomic strategies derives from the determination of the appropriate incubation time or concentration for the treatment in order to obtain a strong signal response for the respective process. It was shown that being off by one log of concentration has turned a strong signal into a barely detectable one¹¹⁹. It is therefore mandatory to profile a wide range of concentrations and several time points to obtain an integrated view of the NCE effects. Furthermore, there is an uncertainty to which extent gene expression signatures of cells from one tissue differ in terms of NCE treatment from cells of another. Surprisingly, earlier chemical genomics studies analyzing histone modification, molecular chaperones and mTOR signaling have shown that gene expression signatures derived from cells of differing origin are remarkably similar⁵⁷. This might differ for other cellular processes and only empirical evidence can resolve this issue. However, the Phenocopy project does not claim to unravel the entire spectrum of NCE effects but to provide guidance along the decisions of early drug discovery projects, where such topics cannot be addressed.

3.3 TGF- β Biology

Gene set enrichment analysis resulted in 16 highly affected processes with known links to the TGF- β biology (Figure 20). TGF- β is involved in a plethora of biological processes and its cross-talk with other pathways can be detected in literally every stage of metazoan life from birth to death. During embryogenesis, complex interac-

Discussion

tions of TGF- β with BMP, Wnt Hedgehog, Notch and mitogen-activated protein kinase (MAPK) signaling are crucial for stem cell maintenance, body patterning, cell fate determination, organogenesis and further retain their role in regulation of cellular growth and functioning in adult tissues to control homeostasis. Often a deregulation of this pathway cross-talk is found in aged or diseased animals, e.g. in case of cancerogenesis^{84, 120-123}. Most of the above mentioned pathway interactions are also represented in the on-target signature of the Phenocopy data emphasizing their ubiquitous role in various cellular backgrounds. However, the respective cross-talks are complex and therefore hard to resolve. They result in the activation of various linked pathways and processes. MAPK signaling, for instance can either be regulated by TGF- β stimulation, which represents an important mechanism for non-Smad signaling⁸⁰, or vice versa modulate the function of TGF- β signaling, primarily triggered through regulation of Smad functions¹²⁴. Furthermore, TGF- β can potentially affect three different MAPK signaling pathways¹²⁵ and result in oppositional effects. While Erk1/2 phosphorylation of Smad3 sites is supposed to prevent the transcriptional activity¹²⁶ of TGF- β , p38 MAPK phosphorylation rather supports it¹²⁷. It is however demanding to dissect which of these processes are triggered in HaCaT cells simply by looking at the regulated genes, since the KEGG category contains genes from all MAPK signaling pathways.

Also not unexpected is the observed TGF- β effect on p53 signaling since the tumor suppressor protein physically interacts with Smad2/3 and jointly regulates the transcription of several TGF- β target genes¹²⁸. For instance, one of the best known TGF- β effects is triggered by this interaction: the induction of cell cycle arrest though the up-regulation of the CDK inhibitors CDKN1A and CDKN2B¹²⁹.

Discussion

Another example for complex cross-talks, also represented in the Phenocopy data, is the interaction of TGF- β with ErbB2 signaling. This finding is backed up by different studies showing that ErbB2 signaling, which activates both, MAPK and PI3K Akt pathways, strongly communicates with TGF- β signaling in the control of mammary epithelial cell biology and breast cancer development¹³⁰⁻¹³³. In fact, the synergy of TGF- β and ErbB2/Ras/MAPK signaling often leads to the secretion of different cytokines and growth factors, including TGF- β itself and thereby promote epithel-to-mesenchymal transition (EMT) and cell invasion¹³⁴⁻¹³⁶.

Also not surprisingly, TGF- β stimulation resulted in the regulation of genes involved in apoptosis. At this, TGF- β signaling can lead to opposing effects and provides signals for either cell survival or apoptosis. The outcome of TGF- β stimulation is thereby determined by the interaction and the balance of different stimuli as well as by the given cell type¹²⁵. Comparably to the findings of the MAPK signaling category, the KEGG apoptosis annotation is also ambiguously grouped regarding the different outcomes of apoptosis.

TGF- β is involved in the remodeling of the extracellular matrix resulting in EMT. An early event in EMT involves the dissolution of tight junctions triggered by an interaction of Par6, a regulator of cell polarity and tight junction assembly with TGF- β signaling components¹³⁷. Not only the tight junctions but also the gap junctions are affected by TGF- β through regulation of the gap junction subunit Cx43 expression which alters gap junction-mediated intercellular communication¹³⁸⁻¹⁴⁰. The outcome varies between positive and inverse relationships, depending on the cell type, the type of membrane receptors employed and the initial activation/phosphorylation state of the cells¹⁴¹. It has been further shown that this down-regulation at wound

Discussion

sites leads to a reduced inflammatory response, enhanced keratinocyte proliferation and wound fibroblast migration¹⁴².

Apart from the alteration of cell connection components, EMT is accompanied by a dramatic reorganization of the actin cytoskeleton also induced by TGF- β stimulation via Rho GTPase-dependent pathways^{143, 144}.

Interestingly, no influence or interaction of TGF- β signaling to the metabolisms of glycine, serine and threonine has been published so far. Therefore, the findings provided by the Phenocopy experiment are a good starting point for further experiments to confirm and unravel the influence of TGF- β on the metabolism of these amino acids.

The last three categories contained in Cluster 4 are wide-ranged and highly ambiguous. They most likely represent the integration of the above mentioned pathways and processes.

As mentioned for the off-target pathway prediction, *in silico* results based on pathway analysis tools merely draws attention to putatively affected processes and only wet laboratory validation can provide insight into their contribution to the targets biology. Furthermore, results obtained from only one cell line cannot cover the entire biology of the cytokine TGF- β and only provide initial insights and approaches for further experiments. Nevertheless, not only the off-target signatures but also an on-target signature can help to support the drug discovery process. On the one hand, these signatures can be overlaid with known disease signatures in order to annotate the targets contribution to the state of disease. On the other hand, it can be used to identify potent biomarkers for efficacy of the treatment and to support the clinical biomarker assay development process. This is especially important if the target's

biology is not as well characterized as for TGF- β R1. Despite the good characterization of the receptor biology, using the Phenocopy strategy it was possible to significantly increase the amount of known TGF- β regulated genes by several hundred compared to earlier studies¹⁴⁵⁻¹⁴⁹.

3.4 siRNAs as modulators

3.4.1 Control siRNA characterization

The aim of this experimental section was to shed light on how different negative control siRNA molecules can influence RNAi experiments. The obtained knowledge could then be directly incorporated into the siRNA part of the Phenocopy strategy. In addition, a broader approach was chosen to identify the influence of the control molecules in order to deduce information for siRNA experiments in general, beyond the use in the presented strategy. Dependent on the used cell line a wide range of differentially expressed genes was observed after the transfection with different siRNA molecules. Overall almost three times more genes were altered in HT1080 compared to 3T3-L1 cells indicating that the later are less sensitive to the treatment with siRNAs. The transfection of siRNA Q1ctrl resulted in a small overlap of three genes that are differentially expressed in all tested cell lines, arguing for a potential sequence homology of this siRNA to the respective off-target transcripts. Due to commercial reasons at the providers, the sequences of the control siRNA molecules cannot be provided but the alignment of the siRNA sequences with the respective off-targets showed a low stringency match to nearly all the identified transcripts.

Discussion

A larger overlap was found within the two human cell lines HT1080 and HaCaT. Here, the control siRNAs Q1ctrl, Q2ctrl, Q3ctrl and Q4ctrl share a common cell line independent off-target signature, suggesting that these genes are altered due to sequence similarities to the siRNA sequence. However, no perfect or seed matches to the control siRNA sequences were identified. Subsequently, the focus was set on the analysis of the specific off-targets observed in HT1080 cells. Since most of these off-targets do not fall into the classical group of siRNA induced interferon response genes¹⁵⁰, it can be anticipated that these regulations are sequence specific for the respective siRNA molecule^{151, 152}. However, among the 595 identified off-targets, 79 genes were altered by the treatment with more than one single siRNA. The hierarchical clustering revealed similar expression patterns for all Dharmacon siRNAs indicating that some off-target effects are sequence independent and one can speculate that these are related to chemical modification strategies of a specific manufacturer. The cluster analysis also demonstrated strong correlation between the cells treated with the molecules A1ctrl and Q1ctrl. In these experiments expression of the two cytokines IL1 β and IL24 is significantly altered compared to the untreated control samples. The low expression of both cytokines argued already for a reduced response of the siRNA treated cells towards an inflammation signaling through NF κ B¹⁵³. This hypothesis was further strengthened by the reduced level of MMP1 expression in the supernatants of these cultures⁹⁷, which correlates with the reduced expression of both cytokines. However, not only the basal NF κ B signaling cascade, but also the stimulated pathway, as shown by the activation of IL8 secretion in HT1080 cells after TNF α treatment, argues for a reduction of siRNA A1ctrl and Q1ctrl treated cells towards the inflammatory signaling.

Discussion

Furthermore, the unwanted side effects are the opposite of what is described as an inflammation stimulation effect via the double stranded RNA recognition system. The specificity in terms of number of identified off-targets did not correlate with the observed reduction of TNF α sensitivity, but was probably dependent on specific sequence motifs or chemical modifications. Variations in immune stimulation activity have already been described for a GFP control siRNA¹⁹ but the effects were restricted to the context of primary peripheral blood monocytes. Only very low expression levels of the toll-like-receptors 3, 7 and 8 were found in the HT1080 fibroblast cell line by the gene chip expression study. However, the intracellular, cytoplasmatic double stranded RNA binding domain (DRBD) proteins like RIG-1, PKR and MDA-5 were constitutively expressed with strong detection signals. It remains questionable whether the control siRNA-induced resistance of the NF κ B pathway implicates binding to DRBD-proteins since the interaction with these cytoplasmatic components would activate rather than silence this signaling cascade. The underlying mechanism for this observation remains obscure and will need a detailed understanding of the binding and signaling mechanisms of DRBD-proteins in general.

The findings have some major implications on the design of target identification and validation experiments with siRNA molecules. The reduced inflammatory response described is clearly cell type specific, since only the HT1080 cell line did show the IL1 β /IL24 effects. However, the data exemplifies that during the establishment of an assay protocol, the phenotypic analysis of the control siRNA treatment is essential. This is especially important since the identified off-target regulations are cell type specific and not a common feature of the molecules but related to the cellular background of the experiment. The cell type specific effects of siRNAs can derive from

different sources. First, double stranded RNA related effects can be different dependent on the expression of double stranded RNA binding domain (DRBD) proteins, discussed above, or in general by a different expression of the RNAi machinery components. Second, sequence specific effects can differ due to different expression profiles and the biology of the hit target. For instance, the knockdown of an essential key player in cell homeostasis or of an important transcription factor will later on influence the expression of far more genes than the knockdown of a more distal target. Finally, siRNAs can compete with endogenous miRNA pathways due to saturation effects that are specific for the respective cell line¹⁵⁴.

Although it is not mandatory to perform expression analysis upfront of each siRNA experiment, the impact of the control molecule on the assay system must be carefully analyzed. Therefore, only a close comparison of untreated to control siRNA treated cell cultures will help to identify the right control molecules, which show a controlled or even no impact on the assay system and will be appropriate for the normalization of the experimental observations.

3.4.2 TGF- β siRNAs

To guarantee an optimal siRNA-mediated knockdown it is mandatory to pre-select a siRNA, validated for the respective biological background and used cellular system. Additionally, different chemical modifications can lead to an altered activity, stability and tolerability of the respective siRNA¹⁵⁵. Since each siRNA provider follows its own modification strategy, it is beneficial to validate a panel of siRNAs from different providers. Therefore, a total of ten siRNAs from three providers was assessed by the

Discussion

analysis of three quality parameters: i) mRNA knockdown, ii) functional inhibition and iii) amount of off-target effects. Only the use of the siRNAs A2tgf, D4tgf and Q1tgf led to knockdowns of less than 80 %. In contrast, reasonable good knockdowns were achieved by more than 90 % for the siRNAs A1tgf, D3tgf, Q2tgf, Q3tgf and Q4tgf and superior knockdowns with more than 95 % for D1 and D2. Functional pathway inhibition, assayed by Smad phosphorylation and PAI-1 expression, led to good performances of the siRNAs A1tgf, D2tgf, Q3tgf and Q4tgf. Surprisingly, siRNA D1tgf that resulted in the best mRNA knockdown failed to inhibit TGF- β signaling. This finding remains obscure and can not be rationalized by any of its off-target effects. One can only speculate that the siRNA mimics a miRNA effect that can potentially promote TGF- β signaling. Moreover, the fact that a siRNA mediating a good or perfect knockdown does not result in the desired phenotype is not unique and can also be experienced in nearly each large scale siRNA screen. Vice versa, each hit in a siRNA screen is normally backed up by at least one counter experiment using a distinct siRNA to rule out single siRNA derived phenotypes. In parallel, the off-target analysis of these siRNAs favored both Dharmacon siRNAs D1tgf and D2tgf and the Ambion siRNA A1tgf and excludes the use of both Qiagen siRNAs with approximately twice as much off-target effects than the others. Collectively, siRNA A1tgf resulted in the best performance and was chosen for the use in the Phenocopy experiment.

Although treatment with the validated siRNA virtually erased TGF- β R1 mRNA, the use of siRNAs as pathway modulators came off badly in contrast to the kinase inhibitors and was not able to totally block TGF- β signaling according to PAI-1 mRNA levels (Figure 19d). One possible explanation is the internalization of the activated TGF- β receptor complex. TGF- β receptors are internalized through two distinct endocytic

pathways. On the one hand, TGF- β receptors can be internalized in caveolin positive vesicles for receptor degradation. On the other hand, Clathrin-dependent internalization into early endosomes is important for propagating signals. This was first elucidated by Di Guglielmo *et al.*¹⁵⁶ who showed that the internalized TGF- β receptor complex is first co-localized with EEA1 (Early Endosome Antigen 1) and then with Rab11 suggesting that the EEA1 compartment has subsequently entered Rab11-positive recycling endosomes^{156, 157}. Since a recycling of TGF- β receptors is possible, one can rational that due to the long half-life and turnover of the protein, even a complete blockage of its de novo synthesis using siRNAs will rather fade out TGF- β -mediated signaling then abruptly stop it.

3.5 Conclusion

Neither the complexity of a living organism nor a disease state can be entirely represented by profiling of a single cell line. Nevertheless, the Phenocopy strategy demonstrates one possibility to significantly alleviate the drug discovery process at an early stage. Comparing such an approach to classical toxicology testing or toxicogenomics studies, the Phenocopy strategy offers a couple of advantages: it addresses on- and off-target effects and is able to differentiate between target-related vs. compound-related events. This differentiation is only possible when a couple of compounds of different compound classes will be investigated. Due to costs and capacities, the analysis of a certain number of compounds can only be run *in vitro*. Although cellular systems cannot replace *in vivo* studies, they show less variability,

Discussion

guarantee the expression and signaling of the target protein and are less cost and time consuming.

The Phenocopy approach offers an opportunity to qualify and rank compound classes and single compounds early during hit-to-lead and lead optimization processes, which will subsequently reduce the attrition rates later on, e.g. during toxicological assessment of the development candidates. However, the addition of new technologies and checkpoints like Phenocopy is contributing to the ever-rising costs of getting innovative medicine to the market. But nevertheless, the assembly of workflows of successfully used tools during early lead generation processes will become crucial for the discovery of novel quality of entities in a changing pharmaceutical industry. One useful tool is the phenocopy principle, where external stimuli like the climate of the environment for the Himalayan rabbit or like NCEs, siRNAs, antibodies or aptameres for the inhibition of a cellular process are committing a certain phenotype. By investing in the qualification of NCEs during the early drug discovery process, later on the attrition rate during development phases will be reduced. Indirectly, this investment will reduce the overall cost for developing innovative medicine.

Ph.D. Thesis Patrick Baum

4. Methods

Phenocopy – A Strategy to Qualify Chemical Compounds during Hit-to-Lead and Lead Optimization

4.1 Wet laboratory experiments

4.1.1 Cell culture, NCE treatment and siRNA transfection

3T3-L1, HT1080 and HaCaT cells were maintained in DMEM (3T3-L1, HaCaT; Gibco; Cat. No. 31966) or RPMI (HT1080; RPMI; Cat. No. 61870) supplemented with 10 % (3T3-L1, HT1080) or 5 % (HaCaT) FCS, respectively. All cell lines were exponentially grown at 37 °C in a 5 % CO₂ atmosphere. Cells were seeded in 96-well (ELISA) or in 24-well (RNA expression profiling) plates and grown overnight to a confluence of approximately 70 %. For TGF- β expression experiments, cells were first starved for 3 h in DMEM containing no FCS. Cells were then pre-incubated with increasing NCE concentrations (0.0032, 0.016, 0.08, 0.4, 2, 10, 50 μ M) for 15 min and subsequently stimulated with 5 ng/ml of TGF- β (R&D Systems; Cat. No. 240-B) and incubated for the indicated time points. In parallel, control sample groups were analyzed that were either only stimulated with TGF- β , treated with NCEs but not TGF- β stimulated, treated with DMSO in presence and absence of TGF- β or left untreated. Finally, the medium was collected or discarded cells were washed with PBS and lysed using RLT buffer (Qiagen; Cat. No. 79216).

All siRNAs were purchased from Ambion, Dharmacon or Qiagen and prepared according to manufacturer's instructions (Table 1, Table 4). For transfection experiments, cells were seeded in 24-well plates and grown overnight to a confluency of 50-70 %. siRNAs were transfected at a final medium concentration of 20 nM. Cells were either transfected using Lipofectamine RNAiMAX (Invitrogen; Cat. No. 13778-150) for 3T3-L1, HT1080 or DharmaFECT1 reagent (Dharmacon; Cat. No. T-2001-03) for HaCaT cells. 24 h post transfection, the medium was replaced. 48 h after trans-

fection cells were then also treated starved and TGF- β stimulated according to the above mentioned conditions, only treated with the transfection reagent or left untreated dependent on the performed experiment. Subsequently, cells were washed with PBS and lysed using RLT buffer (Qiagen Cat. No. 79216). For electroporation siRNAs were transfected using an Amaxa Nucleofector[®] (Lonza) and the Amaxa Cell line Nucleofector[®] Kit L (Lonza, Cat. No. VCA-1005) according to the manufacturer's protocol.

4.1.2 RNA extraction

RNA isolation was carried out using a MagMAX[™] Express-96 Magnetic Particle Processor (Ambion) and the MagMAX[™]-96 Total RNA Isolation Kit (Ambion; Cat. No. AM1830) according to the manufacturer's protocol. Total RNA concentration was quantified by fluorescence measurement using SYBR Green II (Invitrogen; Cat. No. S-7568) and a Synergy HT reader (BioTek) as previously described¹⁵⁸. The RNA quality was characterized by the quotient of the 28S to 18S ribosomal RNA electropherogram peak using an Agilent 2100 bioanalyzer and the RNA Nano Chip (Agilent; Cat. No. 1511).

4.1.3 Quantitative real time polymerase chain reaction (qRT-PCR)

mRNA expression levels of TGF- β R1, PAI-1, GAPDH, CDKN1A, CDKN2B, LINCR, HAND1, RPTN and JUNB were determined by qRT-PCR analysis using a 7900HT Fast Real-Time PCR System (Applied Biosystems) and the Universal ProbeLibrary System (Roche; Cat. No. 04 683 633 001). Gene specific forward- and reverse primer se-

Methods

quences were designed using the Universal Probe Library Assay Design Center (Roche) and are shown in Table 12. Total RNA was transcribed into cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems; Cat. No. 4368814) according to the manufacturer's instructions. PCRs were performed according to the manufacturer's protocol using TaqMan reagents for 40 cycles. The threshold cycles (CT) for each cDNA were obtained from triplicate samples. The Δ CT was the calculated difference between the average CT for the target gene and the average CT for the house keeper gene POLR2A (RNA polymerase 2) as a control for total starting RNA quantity. $\Delta\Delta$ CT method was then used to relatively quantify mRNA levels of treated or stimulated samples compared to untreated, unstimulated or untransfected controls.

Table 12 – qRT-PCR primer

Gene	Forward primer	Reverse primer
TGF- β R1	aaattgctcgacgatgttc	cataataaggcagttggaatctca
PAI-1	aaggcacctctgagaactca	cccaggactaggcaggtg
GAPDH	gctctctgctcctcctgttc	acgaccaaaccggttgactc
CDKN1A	tcggttcacaggtgtttctg	agctgctcgctgtccact
CDKN2B	ttgtagaaaccaggctgcac	ttctcttttctgtggtttctcaat
LINCR	ccttcgagagcatttgccca	tgttggcagcgtgatagaag
HAND1	aactcaagaaggcggatgg	ggaggaaaaccttcgtgct
TSC22D1	tttccgttgaaggtgctac	ttgtcaatagctaccacacttg
RPTN	gctcttggtgagtttgag	aggttcaagatggtttccaca
JUNB	atacacagctacgggatacgg	gctcggtttcaggagttgt
POLR2A	ttgtgcaggacacactcaca	caggagttcatcacttcacc

4.1.4 ELISA analysis of PAI-1, phospho Smad2/3, MMP1, IL8 and IL6

To analyze Smad2/3 phosphorylation after TGF- β stimulation and/or NCE treatment cells were lysed in 100 μ l Cell Lysis Buffer (Cell Signaling; Cat. No. 9803) supplemented with 1 mM PMSF (Sigma; Cat. No. P7626). 96-well plates (Nunc MaxiSorp™;

Methods

Cat. No. 44-2404-21) were coated with anti-Smad2/3 monoclonal antibody (1 µg/ml; BD Bioscience; Cat. No. 610843) for 24 h at 4 °C. To reduce unspecific binding the wells were blocked with PBS (Gibco; Cat. No. 14190) + 2 % BSA (Sigma; Cat. No. A-7030) for 2 h at RT. After washing three times with PBS + 0.1 % Tween20 (Sigma; Cat. No. P-7949), the protein lysate was added and incubated for 2 h at RT. Wells were washed three times with wash buffer and incubated with an anti-phospho-Smad2/3 specific rabbit antisera (Eurogentec) diluted in PBS + 0.2 % BSA (Sigma; Cat. No. A-7030) + 0.02 % Tween20 (Sigma; Cat. No. P-7949) and incubated for 2 h at RT. An AP-conjugated monoclonal antibody mouse anti-rabbit IgG (Sigma; Cat. No. A2556) was added and incubated for 2 h at RT. pNPP Liquid Substrate System (Sigma; Cat. No. N7653) was added and developed in the dark at 37 °C for 2 h before the absorbance was measured at 405 nm in a Synergy HT plate reader (BioTek).

To analyze protein expression of PAI-1 supernatants of TGF-β stimulated cells were collected 1, 2, 4, 12 and 24 h after treatment and analyzed using the PAI-1 Antigen Kit (Haemochrom Diagnostica; Cat. No. HD44006), according to the manufacturer's protocol. To analyze protein expression of MMP1 and IL8 supernatants of transfected or transfection reagent treated cells were collected 72 h after control siRNA transfection. In case of IL8 cells were previously stimulated with 30 ng/ml TNFα (Alexis; Cat. No. ALX-520-002) for 8 h. Subsequently, protein expression was determined using the MMP1 ELISA kit (Calbiochem; Cat. No. QIA55) and the human IL8 ELISA kit (BD Bioscience; Cat. No. 555244) according to the manufacturer's protocol.

To analyze the expression of the four pro-inflammatory cytokines IL1β, TNFα, IL8 and IL6 cells were treated with NCEs at the indicated concentrations and incubated at 37

°C for 12 h. Supernatants were analyzed using a Mesoscale Discovery multiplex ELISA System (MSD) and the Hu ProInflammatory-4 II Tissue Culture Kit (MSD; Cat. No. K15025B-1) for detection according to the manufacturer's instruction.

4.1.5 LDH release assay

LDH release was measured to analyze cell toxicity after siRNA transfection using either lipofection or electroporation protocols. Therefore, the CytoTox-ONE™ Homogeneous Membrane Integrity Assay Kit (Promega; Cat. No. G7890) was used according to the manufacturer's instruction. As positive control, cells were treated with Triton X-100 (Sigma; Cat. No. T9284) and incubated for 3 h at 37 °C. The signal obtained for positive control treated cells was then set as 100 %.

4.1.6 Amplification, labeling and Beadchip hybridization of RNA samples

Illumina TotalPrep RNA Amplification Kit (Ambion; Cat. No. 4393543) was used to transcribe 350 ng of the isolated total RNA. Briefly, total RNA was first converted into single-stranded cDNA with reverse transcriptase using an oligo-dT primer containing the T7 RNA polymerase promoter site and then copied to produce double-stranded cDNA molecules. An overnight (14 h) in vitro transcription was performed using a T7 polymerase to generate single-stranded RNA molecules (cRNA). cRNA molecules were labeled by incorporation of biotin-UTP. A total of 700 ng of cRNA was hybridized at 58°C for 16 h to either the Illumina HumanRefseq-8 v2 Expression BeadChip (HaCaT, HT1080; Illumina; Cat. No. BD-102-0203) or the MouseRef-8 v1.1, (3T3-L1; Illumina; Cat. No. BD-202-0202), respectively for the identification of the control

Methods

siRNA off-target effects. All sample of the Phenocopy experiment were hybridized to the next generation of Illumina chips the HumanHT-12 Expression BeadChips (Illumina; Cat. No. BD-103-0203). After hybridization, the arrays were washed, blocked and the labeled cRNA was detected by staining with streptavidin-Cy3. BeadChips were scanned using an Illumina Bead Array Reader and the Bead Scan Software (Illumina).

4.1.7 High content screen Cellomics

The high-content cytotoxicity assay kit 2 was performed according to the manufacturer's instructions (ThermoFisher Cellomics; Cat. No. 8400102). Briefly, HaCaT cells were cultured overnight in black 96-well plates, incubated for 24 h with each NCE at the indicated concentrations and stained with cytotoxicity cocktail. Cells were fixed, washed and scanned on the Cellomics ArrayScan II platform. Images were analyzed with the Cell Health image analysis algorithm. Cytotoxicity indices were calculated for each of the four parameters to indicate the percentage of cells outside of the normal range which was defined using a vehicle-treated reference cell population.

4.1.8 Caspase-3 Assay

Cells were seeded in 6-well plates and grown overnight to a confluence of approximately 70 % before they were treated with 2 μ M of each NCE and incubated for 24 h. Caspase-3 activity was quantified using Facs Canto (BD Biosciences) and the Caspase-3 Detection Kit (Calbiochem; Cat. No. QIA91) according to the manufacturer's instruction.

4.1.9 *In vitro* kinase profiling

The SelectScreen™ kinase Profiling Service was performed (Invitrogen) to identify the compound selectivity against 239 kinases. Single-point kinase inhibitory activities of each compound at 2 μM and 0.2 μM were measured at 100 μM or K_m ATP concentrations. Downstream targets of the identified off-target kinases were manually extracted from Ingenuity's Knowledge Base and overlaid with the NCE off-targets for comparison.

4.2 Data Analysis

4.2.1 Data processing

Phenocopy data was processed with BeadStudio version 3.0 and the R Language and Environment for Statistical Computing (R) 2.7.0 in combination with Bioconductor 2.2¹⁵⁹. The Bioconductor lumi package¹⁰⁴ was used for quality control and normalization. The data was \log_2 transformed and normalized using robust spline normalization (rsn). Linear models (Bioconductor package limma)¹⁶⁰ were used to calculate \log_2 ratios, the resulting p-values were FDR-corrected⁶⁸. Other pre-processing methods tested prior to the Phenocopy experiment are summarized in the work of Schmid *et al.*¹⁰⁶.

Differential expression for all siRNA experiments were analyzed for siRNA treated samples versus transfection reagent treated samples. Identification of differential expression after TGF- β or NCE treatment is described below. The differentially expressed transcripts after control siRNA transfection in HT1080 cells (Figure 11) as well as the transcripts of the treatment signature (Figure 24) were clustered using

Spotfire® DescisionSite® 9.1.1. (Spotfire). Hierarchical clustering was performed using manhattan distance and complete linkage as similarity measure.

4.2.2 TGF- β signature (on-target signature)

To define genes deregulated by TGF- β signaling, three sequential filtering steps were applied to the \log_2 transformed expression values of each time point separately. i) The first filtering is based on the comparison of TGF- β stimulated cells against untreated cells by linear models¹⁶⁰ (FDR corrected⁶⁸ p-value < 0.01 and $|\log_2 \text{ratio}| \geq 0.5$). ii) A linear model was applied to the dose groups of each compound to extract all probes which are significantly deregulated (FDR-corrected p-value < 0.01) by at least one concentration compared to the respective control (cells treated with TGF- β and DMSO but no compound). iii) To detect probes with a dose dependent deregulation the likelihood ratio test statistic for monotonicity (R package IsoGene¹⁶¹) was used. IsoGene performs an isotonic regression based on the replicates for each concentration resulting in regression values $\mu_1, \mu_2, \dots, \mu_6$ for each probe and each compound treatment. Only probes that are significantly regulated by at least one compound with $|\mu_1 - \mu_6| \geq 1$ and an FDR corrected p-value < 0.01 are included in the further analysis. For each time point the probes that passed all three filters are pooled to the final TGF- β signature.

4.2.3 Off-target signature

To detect transcripts that are deregulated due to off-target effects of the compounds unstimulated cells (wotgf class) as well as TGF- β stimulated cells (tgf class)

Methods

were considered. For the wotgf class, the NCE samples 0.08 μM just as 2 μM were compared to untreated cells (d_{11} and d_{12} , respectively). Additionally, the 2 μM sample was compared to the 0.08 μM sample (δ_1) for each time point using linear models¹⁶⁰ (FDR-corrected⁶⁸ p-value < 0.01 and $|\log_2 \text{ratio}| \geq 1$). The same comparisons were made based on the tgf class (d_{21} , d_{22} and δ_2). Transcripts that are up- or down-regulated by either compound treatment (wotgf_{up} and wotgf_{down}, respectively) or by TGF- β stimulation together with compound treatment (tgf_{up} and tgf_{down}, respectively) were detected based on the described comparisons as follows: a transcript belongs to the class wotgf_{up} if either δ_1 is significantly up-regulated (δ_{1_up}) or if δ_1 is not significantly down-regulated ($-\delta_{1_down}$) but d_{11} , d_{12} , and δ_1 indicate an increasing course of expression intensity for higher compound concentrations. That is, if $-\delta_{1_down}$ holds true, five different trends render up-regulation: 1) d_{11} and d_{12} are both significantly up-regulated; 2) d_{11} but not d_{12} is significantly down-regulated and $\log_2\text{ratio}(\delta_1) \geq 1$, thereby showing an increasing trend of expression for increasing compound concentrations; 3) d_{11} but not d_{12} is significantly up-regulated and $\log_2\text{ratio}(\delta_1) > -1$, allowing for a small but not significant decreasing trend for increasing compound concentration; 4) d_{12} but not d_{11} is significantly down-regulated and $\log_2\text{ratio}(\delta_1) \geq 1$; 5) d_{12} but not d_{11} is significantly up-regulated and $\log_2\text{ratio}(\delta_1) > -1$. On the one hand, as soon as one of d_{11} or d_{12} is significantly up-regulated (cases 3 and 5), a small amount of noise was allowed by claiming $\log_2\text{ratio}(\delta_1) > -1$. On the other hand, as soon as one of d_{11} or d_{12} is significantly down-regulated (cases 2 and 4), one are more strict by claiming $\log_2\text{ratio}(\delta_1) \geq 1$ to call a transcript as being up-regulated.

Stated in a more mathematical fashion, transcripts up-regulated within the wotgf class are defined as follows:

$$\text{wotgf}_{\text{up}} = \left[\begin{array}{l} \delta_{1_{\text{up}}} \vee \\ [(\neg \delta_{1_{\text{down}}}) \wedge ((d_{11_{\text{up}}} \wedge d_{12_{\text{up}}}) \\ \vee (d_{11_{\text{down}}} \wedge \neg d_{12_{\text{down}}} \wedge \log_2 \text{ratio} \mathbb{C}_1 \gtrsim 1) \\ \vee (d_{11_{\text{up}}} \wedge \neg d_{12_{\text{up}}} \wedge \log_2 \text{ratio} \mathbb{C}_1 \gtrsim -1) \\ \vee (\neg d_{11_{\text{down}}} \wedge d_{12_{\text{down}}} \wedge \log_2 \text{ratio} \mathbb{C}_1 \gtrsim 1) \\ \vee (\neg d_{11_{\text{up}}} \wedge d_{12_{\text{up}}} \wedge \log_2 \text{ratio} \mathbb{C}_1 \gtrsim -1))] \end{array} \right]$$

The mirrored method was used to detect $\text{wotgf}_{\text{down}}$ and the analogous methods are used to detect tgf_{up} and tgf_{down} based on the cells stimulated with TGF- β .

$$\text{wotgf}_{\text{down}} = \left[\begin{array}{l} \delta_{1_{\text{down}}} \vee \\ [(\neg \delta_{1_{\text{up}}}) \wedge ((d_{11_{\text{down}}} \wedge d_{12_{\text{down}}}) \\ \vee (d_{11_{\text{down}}} \wedge \neg d_{12_{\text{down}}} \wedge \log_2 \text{ratio} \mathbb{C}_1 \lesssim 1) \\ \vee (d_{11_{\text{up}}} \wedge \neg d_{12_{\text{up}}} \wedge \log_2 \text{ratio} \mathbb{C}_1 \lesssim -1) \\ \vee (\neg d_{11_{\text{down}}} \wedge d_{12_{\text{down}}} \wedge \log_2 \text{ratio} \mathbb{C}_1 \lesssim 1) \\ \vee (\neg d_{11_{\text{up}}} \wedge d_{12_{\text{up}}} \wedge \log_2 \text{ratio} \mathbb{C}_1 \lesssim -1))] \end{array} \right]$$

$$\text{tgf}_{\text{up}} = \left[\begin{array}{l} \delta_{2_{\text{up}}} \vee \\ [(\neg \delta_{2_{\text{down}}}) \wedge ((d_{21_{\text{up}}} \wedge d_{22_{\text{up}}}) \\ \vee (d_{21_{\text{down}}} \wedge \neg d_{22_{\text{down}}} \wedge \log_2 \text{ratio} \mathbb{C}_2 \gtrsim 1) \\ \vee (d_{21_{\text{up}}} \wedge \neg d_{22_{\text{up}}} \wedge \log_2 \text{ratio} \mathbb{C}_2 \gtrsim -1) \\ \vee (\neg d_{21_{\text{down}}} \wedge d_{22_{\text{down}}} \wedge \log_2 \text{ratio} \mathbb{C}_2 \gtrsim 1) \\ \vee (\neg d_{21_{\text{up}}} \wedge d_{22_{\text{up}}} \wedge \log_2 \text{ratio} \mathbb{C}_2 \gtrsim -1))] \end{array} \right]$$

$$\text{tgf}_{\text{down}} = \left[\begin{array}{l} \delta_{2_{\text{down}}} \vee \\ [(\neg \delta_{2_{\text{up}}}) \wedge ((d_{21_{\text{down}}} \wedge d_{22_{\text{down}}}) \\ \vee (d_{21_{\text{down}}} \wedge \neg d_{22_{\text{down}}} \wedge \log_2 \text{ratio} \mathbb{C}_2 \lesssim 1) \\ \vee (d_{21_{\text{up}}} \wedge \neg d_{22_{\text{up}}} \wedge \log_2 \text{ratio} \mathbb{C}_2 \lesssim -1) \\ \vee (\neg d_{21_{\text{down}}} \wedge d_{22_{\text{down}}} \wedge \log_2 \text{ratio} \mathbb{C}_2 \lesssim 1) \\ \vee (\neg d_{21_{\text{up}}} \wedge d_{22_{\text{up}}} \wedge \log_2 \text{ratio} \mathbb{C}_2 \lesssim -1))] \end{array} \right]$$

Methods

The final off-target signature was defined based on the following transcripts:

(tgf_{up} wotgf_{up}) (tgf_{down} $\text{wotgf}_{\text{down}}$) (tgf_{up} $\text{wotgf}_{\text{down}}$) (tgf_{down} wotgf_{up})

The profiles of the respective transcripts can be assigned to different categories as described in Figure 22 and Figure 23.

tgf_{up} wotgf_{up} :

additive or bipolar on- and off-target effect or common off-target effect

tgf_{down} $\text{wotgf}_{\text{down}}$:

additive or bipolar on- and off-target effect or common off-target effect

tgf_{up} $\text{wotgf}_{\text{down}}$:

inverse of bipolar on- and off-target effect

tgf_{down} wotgf_{up} :

inverse or bipolar on- and off-target effect

4.2.4 Ingenuity Pathway Analysis and Gene Set Enrichment Analysis

Based on the on- and off-target signatures, standard IPAs were used to generate networks and perform GSEA using Fisher's exact test for molecular function and canonical pathways defined by the Ingenuity Knowledge Base⁷³.

Additionally, GSEA for the on-target signatures was conducted using Fisher's exact test based on gene sets defined by KEGG pathways as annotated by the Bioconductor package KEGG.db version 2.2.0⁷⁴. The p-values calculated based on Fisher's exact test were clustered using manhattan distance and complete linkage.

Reference List

1. Sawin, P.B. Hereditary Variation of the Chin-Chilla Rabbit: In Coat and Eye Color. *J. Hered.* **23**, 39-46 (1932).
2. Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P., & Sharp, P.A. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev.* **13**, 3191-3197 (1999).
3. Whitehead, K.A., Langer, R., & Anderson, D.G. Knocking down barriers: advances in siRNA delivery. *Nat. Rev. Drug Discov.* **8**, 129-138 (2009).
4. Zamore, P.D., Tuschl, T., Sharp, P.A., & Bartel, D.P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, 25-33 (2000).
5. Elbashir, S.M., Lendeckel, W., & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188-200 (2001).
6. Weitzer, S. & Martinez, J. hClp1: a novel kinase revitalizes RNA metabolism. *Cell Cycle* **6**, 2133-2137 (2007).
7. Nykanen, A., Haley, B., & Zamore, P.D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**, 309-321 (2001).
8. Khvorova, A., Reynolds, A., & Jayasena, S.D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209-216 (2003).
9. Schwarz, D.S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199-208 (2003).
10. Watson, J.D. & Crick, F.H. The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **18**, 123-131 (1953).
11. Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563-574 (2002).
12. Tomari, Y. & Zamore, P.D. Perspective: machines for RNAi. *Genes Dev.* **19**, 517-529 (2005).
13. Meylan, E., Tschopp, J., & Karin, M. Intracellular pattern recognition receptors in the host response. *Nature* **442**, 39-44 (2006).
14. Sledz, C.A. & Williams, B.R. RNA interference and double-stranded-RNA-activated pathways. *Biochem. Soc. Trans.* **32**, 952-956 (2004).

Reference List

15. Pippig,D.A. *et al.* The regulatory domain of the RIG-I family ATPase LGP2 senses double-stranded RNA. *Nucleic Acids Res.* **37**, 2014-2025 (2009).
16. Diebold,S.S. *et al.* Nucleic acid agonists for Toll-like receptor 7 are defined by the presence of uridine ribonucleotides. *Eur. J. Immunol.* **36**, 3256-3267 (2006).
17. Hornung,V. *et al.* Sequence-specific potent induction of IFN-[alpha] by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nat. Med* **11**, 263-270 (2005).
18. Judge,A. & MacLachlan,I. Overcoming the innate immune response to small interfering RNA. *Hum. Gene Ther.* **19**, 111-124 (2008).
19. Robbins,M. *et al.* Misinterpreting the therapeutic effects of small interfering RNA caused by immune stimulation. *Hum. Gene Ther.* **19**, 991-999 (2008).
20. Sledz,C.A., Holko,M., de Veer,M.J., Silverman,R.H., & Williams,B.R. Activation of the interferon system by short-interfering RNAs. *Nat. Cell Biol.* **5**, 834-839 (2003).
21. Jackson,A.L. *et al.* Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *RNA* **12**, 1197-1205 (2006).
22. Czauderna,F. *et al.* Structural variations and stabilising modifications of synthetic siRNAs in mammalian cells. *Nucleic Acids Res.* **31**, 2705-2716 (2003).
23. Cekaite,L., Furset,G., Hovig,E., & Sioud,M. Gene expression analysis in blood cells in response to unmodified and 2'-modified siRNAs reveals TLR-dependent and independent effects. *J. Mol. Biol.* **365**, 90-108 (2007).
24. Puri,N. *et al.* LNA incorporated siRNAs exhibit lower off-target effects compared to 2'-OMethoxy in cell phenotypic assays and microarray analysis. *Nucleic Acids Symp. Ser. (Oxf)*25-26 (2008).
25. Dowler,T. *et al.* Improvements in siRNA properties mediated by 2'-deoxy-2'-fluoro-beta-D-arabinonucleic acid (FANA). *Nucleic Acids Res.* **34**, 1669-1675 (2006).
26. U.S. Food and Drug Administration. www.fda.gov . 2010.
27. Hopkins,A.L. & Groom,C.R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727-730 (2002).
28. Russ,A.P. & Lampel,S. The druggable genome: an update. *Drug Discov. Today* **10**, 1607-1610 (2005).

Reference List

29. Kinexus. www.kinexus.ca . 2010.
30. Zhang,J., Yang,P.L., & Gray,N.S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9**, 28-39 (2009).
31. Chico,L.K., Van Eldik,L.J., & Watterson,D.M. Targeting protein kinases in central nervous system disorders. *Nat. Rev. Drug Discov.* **8**, 892-909 (2009).
32. Frantz,S. Drug discovery: playing dirty. *Nature* **437**, 942-943 (2005).
33. Roth,G.J. *et al.* Design, Synthesis and Evaluation of Indolinones as Inhibitors of the Transforming Growth Factor Beta Receptor I. (submitted).
34. Zou,J. *et al.* Microarray profile of differentially expressed genes in a monkey model of allergic asthma. *Genome Biol.* **3**, research0020 (2002).
35. Rajcevic,U., Niclou,S.P., & Jimenez,C.R. Proteomics strategies for target identification and biomarker discovery in cancer. *Front. Biosci.* **14**, 3292-3303 (2009).
36. Rodrigues,C.D. *et al.* Host scavenger receptor SR-BI plays a dual role in the establishment of malaria parasite liver infection. *Cell Host. Microbe* **4**, 271-282 (2008).
37. Terstappen,G.C., Schlupen,C., Raggiaschi,R., & Gaviraghi,G. Target deconvolution strategies in drug discovery. *Nat. Rev. Drug. Discov.* **6**, 891-903 (2007).
38. Yang,Y., Adelstein,S.J., & Kassis,A.I. Target discovery from data mining approaches. *Drug Discov. Today* **14**, 147-154 (2009).
39. Cheng,K.C., Korfmacher,W.A., White,R.E., & Njoroge,F.G. Lead Optimization in Discovery Drug Metabolism and Pharmacokinetics/Case study: The Hepatitis C Virus (HCV) Protease Inhibitor SCH 503034. *Perspect. Medicin. Chem.* **1**, 1-9 (2008).
40. Benedetti,M.S. *et al.* Drug metabolism and pharmacokinetics. *Drug Metab. Rev.* **41**, 344-390 (2009).
41. Tillement,J.P. A low distribution volume as a determinant of efficacy and safety for histamine (H1) antagonists. *Allergy* **50**, 12-16 (1995).
42. Zamek-Gliszczyński,M.J., Hoffmaster,K.A., Nezasa,K., Tallman,M.N., & Brouwer,K.L. Integration of hepatic drug transporters and phase II metabolizing enzymes: mechanisms of hepatic excretion of sulfate, glucuronide, and glutathione metabolites. *Eur. J. Pharm. Sci.* **27**, 447-486 (2006).

Reference List

43. Wexler,D.S. *et al.* Linking solubility and permeability assays for maximum throughput and reproducibility. *J. Biomol. Screen.* **10**, 383-390 (2005).
44. Cheng,Y., Ho,E., Subramanyam,B., & Tseng,J.L. Measurements of drug-protein binding by using immobilized human serum albumin liquid chromatography-mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **809**, 67-73 (2004).
45. Li,A.P. Screening for human ADME/Tox drug properties in drug discovery. *Drug Discov. Today* **6**, 357-366 (2001).
46. Ragu Ramanathan Mass spectrometry in drug metabolism and pharmacokinetics (2009).
47. Barile,F.A., Dierickx,P.J., & Kristen,U. In vitro cytotoxicity testing for prediction of acute human toxicity. *Cell Biol. Toxicol.* **10**, 155-162 (1994).
48. O'Brien,P.J. *et al.* High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening. *Arch. Toxicol.* **80**, 580-604 (2006).
49. Ulrich,R. & Friend,S.H. Toxicogenomics and drug discovery: will new technologies help us produce better drugs? *Nat. Rev. Drug Discov.* **1**, 84-88 (2002).
50. Khor,T.O., Ibrahim,S., & Kong,A.N. Toxicogenomics in drug discovery and drug development: potential applications and future challenges. *Pharm. Res.* **23**, 1659-1664 (2006).
51. Kinders,R. *et al.* Phase 0 clinical trials in cancer drug development: from FDA guidance to clinical practice. *Mol. Interv.* **7**, 325-334 (2007).
52. U.S. National Institute of Health - Clinical Trials. clinicaltrials.gov . (2010).
53. Kola,I. & Landis,J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug. Discov.* **3**, 711-715 (2004).
54. DiMasi,J.A., Hansen,R.W., & Grabowski,H.G. The price of innovation: new estimates of drug development costs. *J. Health Econ.* **22**, 151-185 (2003).
55. Hughes,B. 2009 FDA drug approvals. *Nat. Rev. Drug Discov.* **9**, 89-92 (2010).
56. Walker,I. & Newell,H. Do molecularly targeted agents in oncology have reduced attrition rates? *Nat. Rev. Drug Discov.* **8**, 15-16 (2009).

Reference List

57. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929-1935 (2006).
58. Schena, M., Shalon, D., Davis, R.W., & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470 (1995).
59. Lockhart, D.J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675-1680 (1996).
60. Kuhn, K. *et al.* A novel, high-performance random array platform for quantitative gene expression profiling. *Genome Res.* **14**, 2347-2356 (2004).
61. Hughes, T.R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109-126 (2000).
62. Ramaswamy, S. *et al.* Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. U. S. A* **98**, 15149-15154 (2001).
63. Su, A.I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U. S. A* **101**, 6062-6067 (2004).
64. ArrayExpress. www.ebi.ac.uk/microarray-as/ae . (2010).
65. Gene Expression Omnibus. www.ncbi.nlm.nih.gov/geo . (2010).
66. Lim, W.K., Wang, K., Lefebvre, C., & Califano, A. Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks. *Bioinformatics.* **23**, i282-i288 (2007).
67. Chen, J.J., Wang, S.J., Tsai, C.A., & Lin, C.J. Selection of differentially expressed genes in microarray data analysis. *Pharmacogenomics. J.* **7**, 212-220 (2007).
68. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57**, 289-300 (1995).
69. Lin, D. *et al.* Testing for trends in dose-response microarray experiments: a comparison of several testing procedures, multiplicity and resampling-based inference. *Stat. Appl. Genet. Mol. Biol.* **6**, Article26 (2007).
70. Ji, R.R. *et al.* Transcriptional profiling of the dose response: a more powerful approach for characterizing drug activities. *PLoS. Comput. Biol.* **5**, e1000512 (2009).

Reference List

71. Golub, T.R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531-537 (1999).
72. BioCarta. www.biocarta.com . (2010).
73. Ingenuity Pathway Analysis. www.ingenuity.com . (2010).
74. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27-30 (2000).
75. WikiPathways. www.wikipathways.org . (2010).
76. Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, D619-D622 (2009).
77. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A* **102**, 15545-15550 (2005).
78. Gunderson, K.L. *et al.* Decoding randomly ordered DNA arrays. *Genome Res.* **14**, 870-877 (2004).
79. Shi, Y. & Massague, J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685-700 (2003).
80. Zhang, Y.E. Non-Smad pathways in TGF-beta signaling. *Cell Res.* **19**, 128-139 (2009).
81. Sorrentino, A. *et al.* The type I TGF-beta receptor engages TRAF6 to activate TAK1 in a receptor kinase-independent manner. *Nat. Cell Biol.* **10**, 1199-1207 (2008).
82. Blobel, G.A., Schiemann, W.P., & Lodish, H.F. Role of transforming growth factor beta in human disease. *N. Engl. J. Med.* **342**, 1350-1358 (2000).
83. Massague, J., Blain, S.W., & Lo, R.S. TGFbeta signaling in growth control, cancer, and heritable disorders. *Cell* **103**, 295-309 (2000).
84. Massague, J. TGFbeta in Cancer. *Cell* **134**, 215-230 (2008).
85. Lahn, M., Kloeker, S., & Berry, B.S. TGF-beta inhibitors for the treatment of cancer. *Expert. Opin. Investig. Drugs* **14**, 629-643 (2005).
86. Kang, H.R., Cho, S.J., Lee, C.G., Homer, R.J., & Elias, J.A. Transforming growth factor (TGF)-beta1 stimulates pulmonary fibrosis and inflammation via a Bax-dependent, bid-activated pathway that involves matrix metalloproteinase-12. *J. Biol. Chem.* **282**, 7723-7732 (2007).

Reference List

87. Yingling, J.M., Blanchard, K.L., & Sawyer, J.S. Development of TGF-beta signalling inhibitors for cancer therapy. *Nat. Rev. Drug Discov.* **3**, 1011-1022 (2004).
88. Zanini, A., Chetta, A., & Olivieri, D. Therapeutic perspectives in bronchial vascular remodeling in COPD. *Ther. Adv. Respir. Dis.* **2**, 179-187 (2008).
89. Rosendahl, A. *et al.* Activation of the TGF-beta/activin-Smad2 pathway during allergic airway inflammation. *Am. J. Respir. Cell Mol. Biol.* **25**, 60-68 (2001).
90. Derynck, R., Akhurst, R.J., & Balmain, A. TGF-beta signaling in tumor suppression and cancer progression. *Nat. Genet.* **29**, 117-129 (2001).
91. Jachimczak, P. TGFbeta in cancer and other disease - AACR special conference in cancer research. *IDrugs.* **9**, 239-241 (2006).
92. Wilkins-Port, C.E. & Higgins, P.J. Regulation of extracellular matrix remodeling following transforming growth factor-beta1/epidermal growth factor-stimulated epithelial-mesenchymal transition in human premalignant keratinocytes. *Cells Tissues. Organs.* **185**, 116-122 (2007).
93. Ovcharenko, D., Jarvis, R., Hunicke-Smith, S., Kelnar, K., & Brown, D. High-throughput RNAi screening in vitro: from cell lines to primary cells. *RNA.* **11**, 985-993 (2005).
94. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., & Golani, I. Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* **125**, 279-284 (2001).
95. Vincenti, M.P. & Brinckerhoff, C.E. Transcriptional regulation of collagenase (MMP-1, MMP-13) genes in arthritis: integration of complex signaling pathways for the recruitment of gene-specific transcription factors. *Arthritis Res.* **4**, 157-164 (2002).
96. Borden, P. & Heller, R.A. Transcriptional control of matrix metalloproteinases and the tissue inhibitors of matrix metalloproteinases. *Crit Rev. Eukaryot. Gene Expr.* **7**, 159-178 (1997).
97. Klatt, A.R. *et al.* TAK1 downregulation reduces IL-1beta induced expression of MMP13, MMP1 and TNF-alpha. *Biomedecine & Pharmacotherapy* **60**, 55-61 (2006).
98. Klutchko, S.R. *et al.* 2-Substituted aminopyrido[2,3-d]pyrimidin-7(8H)-ones. structure-activity relationships against selected tyrosine kinases and in vitro and in vivo anticancer activity. *J. Med. Chem.* **41**, 3276-3292 (1998).
99. Chaudhary, N.I. *et al.* Inhibition of PDGF, VEGF and FGF signalling attenuates fibrosis. *Eur. Respir. J.* **29**, 976-985 (2007).

Reference List

100. Roth,G.J. *et al.* Design, synthesis, and evaluation of indolinones as triple angiokinase inhibitors and the discovery of a highly specific 6-methoxycarbonyl-substituted indolinone (BIBF 1120). *J. Med. Chem.* **52**, 4466-4480 (2009).
101. Illumina® BeadStudio. <http://www.illumina.com> . (2010).
102. Lin,S.M., Du,P., Huber,W., & Kibbe,W.A. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res.* **36**, e11 (2008).
103. Irizarry,R.A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* **4**, 249-264 (2003).
104. Du,P., Kibbe,W.A., & Lin,S.M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* **24**, 1547-1548 (2008).
105. Huber,W., von,H.A., Sultmann,H., Poustka,A., & Vingron,M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics.* **18 Suppl 1**, S96-104 (2002).
106. Schmid,R. *et al.* Comparison of Normalization Methods for Illumina BeadChip® HumanHT-12 v3. (submitted).
107. Subramanian,A., Kuehn,H., Gould,J., Tamayo,P., & Mesirov,J.P. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics.* **23**, 3251-3253 (2007).
108. Mootha,V.K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273 (2003).
109. Kanehisa,M., Goto,S., Furumichi,M., Tanabe,M., & Hirakawa,M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* (2009).
110. Kanehisa,M., Goto,S., Furumichi,M., Tanabe,M., & Hirakawa,M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* (2009).
111. Cohen,P. Protein kinases-the major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.* **1**, 309-315 (2002).
112. Zhang,J., Yang,P.L., & Gray,N.S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9**, 28-39 (2009).
113. Bain,J. *et al.* The selectivity of protein kinase inhibitors: a further update. *Biochem J.* **408**, 297-315 (2007).

Reference List

114. Cohen,P. Protein kinases--the major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.* **1**, 309-315 (2002).
115. Venter,J.C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351 (2001).
116. Paolini,G.V., Shapland,R.H., van Hoorn,W.P., Mason,J.S., & Hopkins,A.L. Global mapping of pharmacological space. *Nat. Biotechnol.* **24**, 805-815 (2006).
117. Hieronymus,H. *et al.* Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* **10**, 321-330 (2006).
118. Wei,G. *et al.* Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. *Cancer Cell* **10**, 331-342 (2006).
119. Cornwell,P.D., De Souza,A.T., & Ulrich,R.G. Profiling of hepatic gene expression in rats treated with fibric acid analogs. *Mutat. Res.* **549**, 131-145 (2004).
120. Attisano,L. & Labbe,E. TGFbeta and Wnt pathway cross-talk. *Cancer Metastasis Rev.* **23**, 53-61 (2004).
121. Mishra,L., Shetty,K., Tang,Y., Stuart,A., & Byers,S.W. The role of TGF-beta and Wnt signaling in gastrointestinal stem cells and cancer. *Oncogene* **24**, 5775-5789 (2005).
122. Pires-daSilva,A. & Sommer,R.J. The evolution of signalling pathways in animal development. *Nat. Rev. Genet.* **4**, 39-49 (2003).
123. Rao,M. Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. *Dev. Biol.* **275**, 269-286 (2004).
124. Guo,X. & Wang,X.F. Signaling cross-talk between TGF-beta/BMP and other pathways. *Cell Res.* **19**, 71-88 (2009).
125. Sanchez-Capelo,A. Dual role for TGF-beta1 in apoptosis. *Cytokine Growth Factor Rev.* **16**, 15-34 (2005).
126. Matsuura,I., Wang,G., He,D., & Liu,F. Identification and characterization of ERK MAP kinase phosphorylation sites in Smad3. *Biochemistry* **44**, 12546-12553 (2005).
127. Kamaraju,A.K. & Roberts,A.B. Role of Rho/ROCK and p38 MAP kinase pathways in transforming growth factor-beta-mediated Smad-dependent growth inhibition of human breast carcinoma cells in vivo. *J. Biol. Chem.* **280**, 1024-1036 (2005).

Reference List

128. Cordenonsi, M. *et al.* Links between tumor suppressors: p53 is required for TGF-beta gene responses by cooperating with Smads. *Cell* **113**, 301-314 (2003).
129. Cordenonsi, M. *et al.* Integration of TGF-beta and Ras/MAPK signaling through p53 phosphorylation. *Science* **315**, 840-843 (2007).
130. Janda, E. *et al.* Ras and TGF[beta] cooperatively regulate epithelial cell plasticity and metastasis: dissection of Ras signaling pathways. *J. Cell Biol.* **156**, 299-313 (2002).
131. Muraoka, R.S. *et al.* Increased malignancy of Neu-induced mammary tumors overexpressing active transforming growth factor beta1. *Mol. Cell Biol.* **23**, 8691-8703 (2003).
132. Seton-Rogers, S.E. & Brugge, J.S. ErbB2 and TGF-beta: a cooperative role in mammary tumor progression? *Cell Cycle* **3**, 597-600 (2004).
133. Ueda, Y. *et al.* Overexpression of HER2 (erbB2) in human breast epithelial cells unmasks transforming growth factor beta-induced cell motility. *J. Biol. Chem.* **279**, 24505-24513 (2004).
134. Lehmann, K. *et al.* Raf induces TGFbeta production while blocking its apoptotic but not invasive responses: a mechanism leading to increased malignancy in epithelial cells. *Genes Dev.* **14**, 2610-2622 (2000).
135. Oft, M. *et al.* TGF-beta1 and Ha-Ras collaborate in modulating the phenotypic plasticity and invasiveness of epithelial tumor cells. *Genes Dev.* **10**, 2462-2477 (1996).
136. Yue, J. & Mulder, K.M. Requirement of Ras/MAPK pathway activation by transforming growth factor beta for transforming growth factor beta 1 production in a smad-dependent pathway. *J. Biol. Chem.* **275**, 35656 (2000).
137. Ozdamar, B. *et al.* Regulation of the polarity protein Par6 by TGFbeta receptors controls epithelial cell plasticity. *Science* **307**, 1603-1609 (2005).
138. Hurst, V., IV, Goldberg, P.L., Minnear, F.L., Heimark, R.L., & Vincent, P.A. Rearrangement of adherens junctions by transforming growth factor-beta1: role of contraction. *Am. J. Physiol* **276**, L582-L595 (1999).
139. Rudkin, G.H. *et al.* Transforming growth factor-beta, osteogenin, and bone morphogenetic protein-2 inhibit intercellular communication and alter cell proliferation in MC3T3-E1 cells. *J. Cell Physiol* **168**, 433-441 (1996).
140. Dai, P., Nakagami, T., Tanaka, H., Hitomi, T., & Takamatsu, T. Cx43 mediates TGF-beta signaling through competitive Smads binding to microtubules. *Mol. Biol. Cell* **18**, 2264-2273 (2007).

Reference List

141. Chanson,M. *et al.* Gap junctional communication in tissue inflammation and repair. *Biochim. Biophys. Acta* **1711**, 197-207 (2005).
142. Mori,R., Power,K.T., Wang,C.M., Martin,P., & Becker,D.L. Acute down-regulation of connexin43 at wound sites leads to a reduced inflammatory response, enhanced keratinocyte proliferation and wound fibroblast migration. *J. Cell Sci.* **119**, 5193-5203 (2006).
143. Boland,S. *et al.* TGF beta 1 promotes actin cytoskeleton reorganization and migratory phenotype in epithelial tracheal cells in primary culture. *J. Cell Sci.* **109 (Pt 9)**, 2207-2219 (1996).
144. Lee,J., Ko,M., & Joo,C.K. Rho plays a key role in TGF-beta1-induced cytoskeletal rearrangement in human retinal pigment epithelium. *J. Cell Physiol.* **216**, 520-526 (2008).
145. Kretschmer,A. *et al.* Differential regulation of TGF-beta signaling through Smad2, Smad3 and Smad4. *Oncogene* **22**, 6748-6763 (2003).
146. Ranganathan,P. *et al.* Expression profiling of genes regulated by TGF-beta: differential regulation in normal and tumour cells. *BMC Genomics* **8**, 98 (2007).
147. Dawes,L.J., Elliott,R.M., Reddan,J.R., Wormstone,Y.M., & Wormstone,I.M. Oligonucleotide microarray analysis of human lens epithelial cells: TGFbeta regulated gene expression. *Mol. Vis.* **13**, 1181-1197 (2007).
148. Padua,D. *et al.* TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* **133**, 66-77 (2008).
149. Wu,X., Ma,J., Han,J.D., Wang,N., & Chen,Y.G. Distinct regulation of gene expression in human endothelial cells by TGF-beta and its receptors. *Microvasc. Res.* **71**, 12-19 (2006).
150. Smyth,G.K., Yang,Y.H., & Speed,T. Statistical issues in cDNA microarray data analysis. *Methods Mol. Biol.* **224**, 111-136 (2003).
151. Anderson,E.M. *et al.* Experimental validation of the importance of seed complement frequency to siRNA specificity. *RNA* (2008).
152. Birmingham,A. *et al.* 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods* **3**, 199-204 (2006).
153. Aggarwal,S. *et al.* Melanoma differentiation-associated gene-7/IL-24 gene enhances NF-kappa B activation and suppresses apoptosis induced by TNF. *J Immunol.* **173**, 4368-4376 (2004).
154. Svoboda,P. Off-targeting and other non-specific effects of RNAi experiments in mammalian cells. *Curr. Opin. Mol. Ther.* **9**, 248-257 (2007).

Reference List

155. Bramsen, J.B. *et al.* A large-scale chemical modification screen identifies design rules to generate siRNAs with high activity, high stability and low toxicity. *Nucleic Acids Res.* **37**, 2867-2881 (2009).
156. Di Guglielmo, G.M., Le, R.C., Goodfellow, A.F., & Wrana, J.L. Distinct endocytic pathways regulate TGF-beta receptor signalling and turnover. *Nat. Cell Biol.* **5**, 410-421 (2003).
157. De Renzis, S., Sonnichsen, B., & Zerial, M. Divalent Rab effectors regulate the sub-compartmental organization and sorting of early endosomes. *Nat. Cell Biol.* **4**, 124-133 (2002).
158. Schmidt, D.M. & Ernst, J.D. A fluorometric assay for the quantification of RNA in solution with nanogram sensitivity. *Anal. Biochem.* **232**, 144-146 (1995).
159. Gentleman, R.C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
160. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, (2004).
161. Lin, D. *et al.* Testing for trends in dose-response microarray experiments: a comparison of several testing procedures, multiplicity and resampling-based inference. *Stat. Appl. Genet. Mol. Biol.* **6**, Article26 (2007).

Danksagung

Knowledge is in the end based on acknowledgement. Ludwig Wittgenstein

Die vorliegende Arbeit wäre ohne die Unterstützung zahlreicher Personen nicht möglich gewesen.

Der größte Dank gebührt meinem Betreuer, Dr. Detlev Mennerich, nicht nur für die vorbildhafte, wissenschaftliche Anleitung und fachliche Diskussion, sondern auch dafür, immer ein offenes Ohr für meine Probleme gehabt zu haben und seinen unermüdlichen Einsatz diese dann auch zu lösen. Ferner, möchte ich mich bei ihm für die unkomplizierte und freundschaftliche Art der Betreuung bedanken, die mich auch in schwierigen Phasen motiviert hat und mich nie den Spaß an meiner Arbeit verlieren lies. Zu guter Letzt möchte ich ihm auf diesem Wege noch für die erteilten Lehrstunden in Doppelkopf danken. Durch seine Hilfe war es mir möglich, mich auch auf diesem Gebiet signifikant weiterzuentwickeln.

Des Weiteren bin ich meinen Laborkollegen Karoline Schwarz, Dagmar Knebel, Susanne Siewert, Dr. Tanja Fauti, Dr. Svenja Weikert, Martin Baur und Werner Rust zu großem Dank verpflichtet. Ich möchte mich bei Ihnen sowohl für die umfassende Einführung in die Nukleinsäureanalytik als auch für ihre tatkräftige Unterstützung bedanken. Darüber hinaus, war die gute Arbeitsatmosphäre in Labor mit Sicherheit auch einer der Eckpfeiler für das Gelingen dieser Arbeit.

Mein besonderer Dank gilt auch Dr. Tobias Hildebrandt, Dr. Jörg Rippmann und Dr. Sebastian Kreuz. Jedem einzelnen habe ich eine Erweiterung meines wissenschaftlichen Erfahrungsschatzes zu verdanken. Zusätzlich zu den zahlreichen aufschlussrei-

Danksagung

chen Diskussionen bekam ich hier auch immer gute Ratschläge und tatkräftige Unterstützung zur Bewältigung der jeweiligen Fragestellung.

Auch nicht vergessen will ich meine Bioinformatikerkollegen Dr. Carina Ittrich, Dr. Katrin Fundel-Clemens, Dr. Eric Simon, Dr. Fabian Birzele und Dr. Karsten Quast. Ich möchte mich zum einen für Rat und Tat bei den Analysen aber auch für so profane Dinge wie das Lösen von diversen Computerproblemen bedanken. Besonders hervorheben möchte ich hierbei meine Doktorandenkollegin und „Leidensgefährtin“ Ramona Schmid, da ihr Beitrag maßgeblich zum Gelingen des „Phenocopy Projects“ beigetragen hat. Durch ihren Einsatz war es möglich, die wirren Gedankengänge und Ideen eines Biologen in mathematische Bahnen zu lenken.

Ich möchte ebenfalls ausnahmslos allen Kollegen aus der Genomics Gruppe danken. Die äußerst angenehme Atmosphäre und der gute Zusammenhalt, auch über die Arbeit hinaus, schufen ideale Arbeitsbedingungen. Mein besonderer Dank gilt hier natürlich auch dem „Chefgenomiker“ Dr. Andreas Weith. Zum einen dafür mir die Möglichkeit gegeben zu haben, diese Arbeit in der Genomics Gruppe anzufertigen und zum anderen für Diskussionen, Ratschläge und die erhaltenen Freiräume bei der Realisierung des Projekts.

In diesem Zusammenhang möchte ich mich auch bei PD Dr. Florian Gantner bedanken, der dem Projekt immer wohlwollend gegenüber stand und ohne dessen Zustimmung eine derartige Studie nie möglich gewesen wäre.

Mein spezieller Dank gilt an dieser Stelle natürlich auch meinem Doktorvater Prof. Dr. Roland Kontermann, der ohne zu Zögern meine Betreuung seitens der Universität Stuttgart übernahm und die Arbeit immer sehr interessiert begleitet und unterstützt hat.

Danksagung

Schließlich, möchte ich mich noch bei all meinen Freunden und meiner Familie bedanken. Besonders hervorheben möchte ich hier die stetige Unterstützung und den großen Rückhalt durch meine Eltern Marlies und Manfred, meinen Bruder Philipp und meine Freundin Katja, die mir auch in schwierigen Zeiten immer die Kraft gegeben haben, mich allen Herausforderungen zu stellen.

Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Patrick Baum

Biberach, den 15.04.2010

Lebenslauf

Persönliche Angaben

Geburtsdatum/-ort: 17.06.1978, Mutlangen

Familienstand: Ledig

Staatsangehörigkeit: Deutsch

Schulische Ausbildung

09/89 bis 08/98 Allgemeine Hochschulreife am Limes-Gymnasium in Welzheim

Zivildienst

08/98 bis 09/99 Anstellung beim Bund für Umwelt und Naturschutz in Deutschland (BUND) in Welzheim

Studium

10/99 bis 08/00 Studium der Pharmazie an der Universität Würzburg

10/00 bis 10/06 Studium der technischen Biologie an der Universität Stuttgart in der Fakultät für Geo- und Biowissenschaften

02/05 bis 08/05 Studienarbeit am Tyndall National Institute, Cork, Irland mit dem Thema „The optimisation of the hybridisation of a PDITC modified silicon surface for use in biochip development“

09/05 bis 10/06 Diplomarbeit am Institut für Zellbiologie und Immunologie in der Arbeitsgruppe Biomedical Engineering mit dem Thema „Herstellung und Charakterisierung von single-chain Fv-Immunliposomen selektiv für das Fibroblast Activation Protein“

Doktorarbeit

02/07 bis 05/10 Doktorarbeit bei Boehringer Ingelheim Pharma GmbH & Co KG, Abteilung Atemwegsforschung, Group Genomics mit dem Thema „Phenocopy – A Strategy to Qualify Chemical Compounds during Hit-to-Lead and/or Lead Optimization“

Publikationen

Baum P., Müller D., Rüger R. & Kontermann R.E.

Single-chain Fv immunoliposomes for the targeting of fibroblast activation protein-expressing tumor stromal cells

(Journal of Drug Targeting, 2007, Vol. 15, No. 6, Pages 399-406)

Baum P., Fundel-Clemens K., Kreuz S., Kontermann R.E., Weith A., Mennerich D. & Rippmann J.F.

Off-Target Analysis of Control siRNA Molecules Reveals Important Differences in the Cytokine Profile and Inflammation Response of Human Fibroblasts

(Oligonucleotides, 2010, Vol. 20, No.1, Pages 17-26)

Birzele F., Schaub J., Rust W., Clemens C., **Baum P.**, Kaufmann H., Weith A. Schulz T.W. & Hildebrandt T.

Into the unknown: Expression profiling without genome sequence information in CHO by next generation sequencing

(Nucleic Acids Research, doi:10.1093/nar/gkq116)

Roth G.J., Heckel A., Brandl T., Grauert M., Hoerer S., Kley J., Schnapp G., **Baum P.**, Mennerich D., Schnapp A. & Park J.E.

Design, Synthesis and Evaluation of Indolinones as Inhibitors of the Transforming Growth Factor Beta Receptor I (TGF- β R1)

(submitted)

Schmid R., **Baum P.**, Ittrich C., Fundel-Clemens K., Huber W., Brors B., Eils R., Weith A., Mennerich D. & Quast K.

Comparison of Normalization Methods for Illumina BeadChip[®] HumanHT-12 v3

(submitted)

Baum P., Schmid R., Ittrich C., Rust W., Fundel-Clemens K., Siewert S., Baur M., Mara L., Gruenbaum L., Heckel A., Eils R., Kontermann R.E., Roth G.J., Ganter F., Schnapp A., Park J.E., Weith A., Quast K. & Mennerich D.

Phenocopy – A Strategy to Qualify Chemical Compounds during Hit-to-Lead and/or Lead Optimization

(submitted)

Schmid R., **Baum P.**, Ittrich C., Fundel-Clemens K., Lämmle B., Birzele F., Weith A., Brors B., Eils R., Mennerich D. & Quast, K.

Detecting Meaningful Protein Interactions by Accumulating Evidence

(submitted)