EXTRACTING BIOLOGICAL KNOWLEDGE FROM HIGH THROUGHPUT PHENOTYPIC SCREENS

A Dissertation

By

REID TRENTON POWELL

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of committee,	Peter Davies
Committee members,	Mike Mancini
	Arvind Rao
	Yubin Zhou
Head of Program,	Warren Zimmer

August 2018

Major Subject: Medical Sciences

ABSTRACT

High throughput drug screening has greatly progressed drug discovery for diseases such as cancer. Specifically, the introduction of the automated microscope in high throughput drug screening has facilitated the acquisition of very large amounts of phenotypic data of how biological systems respond to drugs. Indeed, phenotypic analysis has become preferential over traditional biochemical methods due to the ability to measure multiple factors simultaneously, making it a highly efficient method to establish drug sensitivity data. However, these methods heavily rely on the ability to reliably extract quantitative metrics describing alterations in phenotypes in order to begin to approach this data format and make decisions about how populations of cells are responding to perturbagens. A major contribution of this research has been the establishment of a paradigm aimed at translating raw HTS data from phenotypic screens into functional knowledge about biological systems. Importantly, this paradigm starts at the evaluation of the biological question, which dictates the model system, influences the method of data acquisition and processing, and level of statistical analysis. Furthermore, the newly generated information must then be integrated with pre-existing knowledge to facilitate extraction of the new knowledge about the biological process being studied. I have applied this paradigm to multiple biological studies including: the identification a novel class of drug that targets SKP2 mediated degradation of p27 via interference of a protein-protein interaction, the discovery of bexarotene as a novel regulator of cellular primary cilia, and profiling the pharmacologic susceptibilities of panels of cells lines. Each of these projects have contributed new

knowledge to their respective fields, however, they also serve as examples of generalizable screening pipelines and methods that can be applied to additional studies. While these applications have had specific emphasis on understanding the interactions of drugs on cancer cells, the image processing tools and methodological approaches developed here are generalizable to application in other fields including histology, toxicology, and microbiology.

I would like to dedicate this dissertation to my
mother and father for their continual love and support
throughout this transformative process

ACKNOWLEDGEMENTS

There are many people who deserve recognition for making this dissertation possible for me. First and foremost, I would like to thank my parents, Loren and Charles Powell, for raising me in a loving environment, even through my more rebellious years, and providing me the tools and opportunities for limitless success. Likewise, I would like to thank the rest of my family, for which there are too many to name here, for providing a support network throughout this degree and in my life. Finally, I would like to thank my loving fiancé, Kelsey Andrade, for supporting me and providing grammatical edits for this dissertation.

I would also like to thank my early life mentor, Father James Vorwoldt, who challenged me to learn the art of photography through self-exploration and creativity. I feel that the skills that I learned through him have helped me identify a field of science which integrates multiple life passions, gave me the skill to think “outside the box” and to understand things on a conceptual basis. I would also like the thank my graduate mentors/committee members: Peter Davies, Mike Mancini, Arvind Rao, and Yubin Zhou. Each has provided me with key insights into the fields of cancer biology, technology, drug discovery, and bioinformatics in addition to providing an ear when my graduate studies and research projects seemed insurmountable.

Finally, I would like to thank the members of my lab families, both from the Walker and Davies laboratories, especially: Cheryl Walker, Ruhee Dere, In Young Park, Durga Tripathi, Pratim Chowdry, Tia Berry, Lindsey Trevino, Cliff Stephan, Mary Sobiesky, Nghi (Ivy) Nguyen, Yong Sung Park, Marijke Shrock, and Denelle Orellana.

The support and knowledge that each of these people have provided was crucial to my graduate career.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

The following work was supervised by a dissertation committee consisting of Peter Davies of the Center for Translational Cancer Research at Texas A&M University and Mike Mancini of the Department of Cell and Molecular Biology at Baylor College of Medicine, Arvind Rao of the Department of Bioinformatics and Statistics at UT MD Anderson, and Yubin Zhou of the Center for Translational Cancer research at Texas A&M University.

The data presented in Chapter III is done in collaboration with Cheryl Walker of BCM who is the lead PI on the supporting grant, Scott Gilbertson of UTMB who performed medicinal chemistry, and Shuxing Zhang and John Morrow of MDACC who performed the *in silico* modeling. The data presented in Chapter IV is done in collaboration with Cheryl Walker and Ruhee Dere of BCM who provided interest in screening small molecules to pharmacologically rescue the primary cilia and the work presented has been submitted to the journal Cilia. The data in Chapter V is done in collaboration with Rajat Bhattacharya and Lee Ellis who provided the cells and funding to perform the experiments. The example sarcoma screening data presented in Chapter VI is from screens done in collaboration with Anthony Conley and Neeta Somaiah lab at MDACC.

All other work conducted for this dissertation was completed by the student independently.

Funding Sources

This work was made possible in part by the following grant support: CPRIT (RP110532 and RP150578) awarded to Dr. Peter Davies, CPRIT (DR150086) and Welch endowed Chair in Chemistry (BE-0023) awarded to Dr. Cheryl Walker, the Advanced Technology Core at Baylor College of Medicine (NCICA125213 and NIDDK-56338-13/15) and DoD (CA140515) awarded to Dr. Lee Ellis. Its contents are solely the responsibility of the author and do not necessarily represent the official views of the above funding agencies.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

NOMENCLATURE

HT – High throughput

HTS – High throughput screen

HC - High content

HCA – High content analysis

MCTA – multicellular tumor aggregate

PDC – patient-derived cells

ADME – absorption, distribution, metabolism and excretion

ECM – extracellular matrix

CTG – CellTiter Glo

CV – Coefficient of variance

ULA – Ultra low attachment

RZ – Robust Z

MoA – Mechanism of Action

VHL – Von Hippel-Lindau

ccRCC – clear cell reval cell carcinoma

TKI – Tyrosine kinase inhibitors

CHAPTER I

INTRODUCTION

The overarching goal of this dissertation has been to increase the biological and translational relevance of high content (HC) and high throughput (HT) cell based assays used in the field of drug discovery. This has largely been facilitated through the development of HC phenotypic screening pipelines and application of advanced cellular models. Therefore, the focus of Chapter II will be to review the state of knowledge in cancer biology, the implications that the modern understanding of cancer biology has had in drug screening, as well as the roles of phenotypic screening in these applications. Furthermore, a detailed description of the best statistical practices that are important in the design, implementation, and evaluation of modern HT assays will be discussed.

In Chapter III, I will describe a direct application of a phenotypic screening which was used to perform validation experiments from an *in silico* screen aimed at identifying a chemical scaffold that inhibits $SCF^{SKP2\text{-}CKS1}$ mediated degradation of p27, which regulates the cell cycle. This study resulted in the identification of a chemical scaffold that could increase nuclear concentrations of p27, which result in either cell cycle arrest or death. Importantly, this chemical scaffold can now serve as a starting point for additional medicinal chemistry to be further advanced into a biological probe or therapeutically useful drug. Methodologically, this application represents a relatively simple fluorescent analysis of cells grown in monolayers, which is a heavily used paradigm in HTS. Therefore, the methods and procedures outlined in the research can generally be applied to other studies where fluorescent analysis of cells grown as monolayers is the primary endpoint.

In Chapter IV we present a different type of phenotypic screening method that quantifies the morphometric properties of a specific organelle known as the primary cilium. The primary cilium, is a specialized cellular structure that serves as a signaling antenna and when lost is associated with disease, such as cancer. One genetic event that is heavily associated with loss of the primary cilium is the deletion of *VHL*, which is part of an ubiquitination complex and is lost in the early development of certain cancers. Importantly, there are limited therapeutic options that specifically target *VHL*-null cancers. Ultimately, this research resulted in the identification of a retinoid, bexarotene, which pharmacologically rescued the primary cilia in *VHL*-null cells. This led to ongoing studies of the anti-tumor activity of bexarotene in *VHL*-null xenograft models. This application demonstrates how the development of a phenotypic assay facilitated in the discovery of a novel mode of action for an established drug that cannot otherwise be detected through traditional biochemical methods. Finally, the methods generated in this chapter can also serve as a generalizable research tool for those studying ciliopathies *in vitro*.

To this point, all the described applications have been based on cells grown as monolayers; however, there is an overwhelming shift in the field of HTS towards the utilization of advanced cell culture models, such as tumor spheroids, which more faithfully recapitulate *in vivo* tumor physiology. Therefore, the focus of Chapter V is the development and validation of HT phenotypic screening methods which are used to gain an understanding of how the mode of growth can alter pharmacological susceptibility of biological systems. Importantly, this research provides empiric evidence that the trends predicted from spheroid cultures do indeed more faithfully recapitulate clinical observations when compared to monolayer cultures. As a method, this research describes a practical, yet highly robust, approach to implementing multiplate spheroid assays to HTS drug discovery. Therefore, we anticipate that these methods will greatly contribute to the future of HTS which are more heavily relying on these types of models to establish drug sensitivity data.

It is clear that HT drug screening is a useful tool used to generate information about how a biological system responds to a perturbagen. However, the data that it produces is also high dimensional and often times requires extensive data mining in order to obtain a basic understanding of what the collective data means. Historically, one of the most powerful tools to convey large amounts of information is through the generation of informative charts and graphics. The most prevalent methods of displaying HTS data are in the form of waterfall plots, Z-score plots, and clustered heatmaps. However, we find these graphics to be generally uninformative outside of identifying the top performer in a library. Therefore, the focus of Chapter VI is the generation of informative graphics used to provide context to HTS results by embedding them in a network which integrates target profiles and molecular signaling information that the drugs work through. Importantly, this method is designed to be self-compiling and is therefore generalizable to any library that contains combinations of drugs held in the in house annotation database.

CHAPTER II

LITERATURE REVIEW OF THE STATE OF HTS IN CANCER BIOLOGY

Introduction

High throughput screening is now an established scientific discipline that leverages technology to rapidly test many thousands of experimental conditions in a well-defined and statistically validated model system. Therefore, when interoperating HTS data it is important to understand the domain of applicability, which deals with the statistical, biological, and translational relevance of the results. The term “biological significance” is a loosely defined term and is often presented in contrast to statistical significance. To further distinguish these two concepts, statistical relevance deals with the confidence and reproducibility of a measurement, while biological relevance relates to the functional relevance of the test results to the biological system being tested.[1,2] However, the distinction between these two concepts does not limit the importance of both in testing and interoperating HTS results. The related concept of translational relevance expands upon the biological relevance by specifically contextualizing the results in their potential to affect clinical diagnostics or management of a particular disease. In the following chapter, I will discuss the evolution of HTS under the selective pressure of providing biologically and translationally relevant small molecules. Likewise, the discussion will explain how a modern understanding in cancer biology has altered the way we model the disease with the specific aim at increasing the translational relevance of pre-clinical pharmacology. Finally, statistical integration and retrieval of significant results, with a specific interest in bolstering the rigor and reproducibility of HTS systems, will be discussed.

Historical prospective of HTS in targeted drug discovery

HTS is an unbiased discovery based platform that became popularized in the early to mid 90's, when multiple pharmaceutical companies adapted this technology for drug discovery applications.[3] At this point, enzyme kinetics assays where miniaturized, automated, and used to identify potent enzyme inhibitors. Likewise, many of the early discoveries were focused around targets with enzymatic or binding activity, such as kinases and GPCRs; which to the credit of this system, represent the largest classes of molecular targets to date.[4,5] However, one apparent limitation of this paradigm of HTS is a failure to translate activity established in cell-free in *vitro* assays into appreciable cell-based assays and clinical activity. Thus, the introduction of cell-based screening for drug discovery represents a critical stride towards providing more biologically relevant results. In this context, cell-based models can be engineered to provide direct readouts of molecular activity in the presence of physiologic relevant features, such as membrane barriers, metabolic processes, complex signaling interactions, compartmentalization, and the presence of serums in the culture media, which are unaccounted for in many homogenous cell-free assays.[6] A specialized form of a cell-based screen, known as a phenotypic screen, further capitalizes on cell-based screening, but allows one to identify bioactive molecules on the basis of being able to perturb a given phenotype. This method of analysis has resulted in the identification of targets that alter cell morphology, chromosomal segregation, migration, invasion, and survival.[7-10]

The underlying principle behind performing phenotypic screening as a method of discovery is that the phenotype of biological systems is the sum aggregate of many different biological processes. Likewise, if one modulates the activity of an integral component feeding into the phenotype, measurable changes may be observed. This makes research tools, such as the microscope, incredibly powerful in that a researcher can directly visualize alterations in phenotypes. When the microscope is further combined with digitized cameras and integrated with automated systems, a HTS-ready platform is produced with the ability to produce vast amounts of data at the single-cell level and below. Indeed, many HTS facilities have integrated image-based phenotypic screening capabilities due to the efficiency of these methods to provide multiple simultaneous readouts of a biological response in a single assay, making high content analysis (HCA) possible. Despite these positive aspects of phenotypic screening, it is potentially limited in that it does not necessarily confirm the direct mode of action of a bioactive molecule, it comes with significant overhead and instrumentation expenses, and often requires tailored automated methods of analysis. Thus, there is a continual demand to develop new tools to address the needs of new biological questions.

Review of the state of the art in phenotypic image analysis

Given the importance and prevalence of phenotypic screening in drug discovery applications, there has been a fair amount of research devoted to developing image analysis tools. Interestingly, many of the strides in this field have come from generalizing methods developed for non-biological image analysis applications, such as food and industrial quality control assement.[9] Due to the broad scope of applications for image analysis, a few different nomenclatures describing similar underlying concepts have emerged in the field. To simplify this, we subdivide image analysis into two major paradigms, referred to here as high-level and low-level.[11] Importantly, these terms are not to be confused with superiority of either paradigm over the other; but rather how meticulous the paradigm is at describing the content of the image. In a high-level image analysis approach, the goal is to provide a detailed description of individual components of an image, often termed as regions of interest (ROI). Indeed, many of the image analysis pipelines used in HTS fall into this category due the ability to perform single-cell analysis.[12] Limitations in high-level image analysis are that the quality of the data is related to the ability to accurately segment an image, which may not always be possible and is difficult to quantitatively address. In contrast, a low-level image analysis approach attempts to describe the content of an image by generalizing patterns of pixels within a field. Thus, low-level image analysis circumvents the need to define ROI making it highly generalizable. However, one apparent limitation of the low-level image analysis approach is that the features produced are not directly human-interpretable, and it often requires advanced data mining to be used as a quantitative tool. However, simple evaluation of whether a phenotype is significantly different between different experimental conditions can readily be achieved with low-level analysis, making this a valid tool for primary screening. Likewise, open source and commercial tools, which capitalize on both of these paradigms, have been developed specifically for bio-image analysis.[13-15]

The next most critical aspect of image processing is the selection of the feature space. This is analogous to a pathologist or cellular biologist selecting the right collection of words to describe a phenotype. Likewise, there are various types of descriptors that capture morphometric, intensity, and textural differences that may be used independently or collectively to describe a particular phenotype. Morphometric analysis is generally the most robust and human-interpretable method used to quantify a change in phenotype. Morphometric features rely on a high-level image analysis, which provides discrete objects to be measured. Ideally, the major visual change can be described by a key morphometric factor, such as an object becoming smaller or rounder, however, some phenotypes require a more detailed description in order to successfully differentiate them. Intensity-based features are also commonly used for biomedical image analysis applications. Similar to morphometric features, intensity features are most often derived from ROI but may also be calculated from a field. Applications that capitalize on intensity-based features include the quantification of protein levels, localization, and cell cycle stage; which use changes in the mass, concentration, or variance of signals from dyes or stains.[16,17] The most apparent limitation of intensity-based features is reproducibility of the signal across assay batches. Therefore, intensity features require further normalization and systematic evaluation of batch effects in order to be integrated across a screening campaign, which are discussed in detail in a subsequent section. An emerging feature space that is now being explored in biological image analysis quantifies textural information. Multiple texture feature extraction methods have been proposed, most of which quantify the relational attributes between neighboring pixels. Likewise, texture is most commonly used in low-level image analysis methods.[18,19] In addition to primary features, there are a number of calculable features. Applications of these include the quantification of the penetrance of a phenotype and long range organization of cells.[20] Finally, the integration of machine learning techniques has provided powerful new tools to quantify a phenotype using mixed feature spaces.

Implications of cancer biology in *in vitro* model systems

There is a continuously evolving view of the features that define a successful cancer; likewise, the lack of representation of these features in *in vitro* models has been proposed as a major cause for the low rate of translation of drugs identified using these systems into clinical activity. To summarize these features, a set of 6 hallmarks of cancer were initially proposed by Hanahan and Weinberg. The hallmarks of a successful cancer proposed here are: resistance to cell death, sustaining proliferative growth, evading growth suppressors, activating invasion and metastasis, enabling replicative immortality, and inducing angiogenesis.[21] Later, the initial 6 hallmarks were supplemented with dysregulation of cellular metabolism, ability to avoid immune destruction, enabling tumor promoting inflammation, and genomic instability.[22] Interestingly, many of these hallmarks rely on the ability of a tumor to promote a heterogeneous environment, to the extent that tumor heterogeneity can be considered a distinct hallmark.[22,23] Indeed, tumor heterogeneity can be observed at multiple levels, which include heterogeneity in the cellular composition of a tumor, heterogeneity between tumors within a patient, and genetic heterogeneity within a given tumor. Through the recognition of tumor heterogeneity, multiple novel therapeutic targets have been identified and are now being explored. Though, it also resulted in the recognition of the extent of limitations of the current paradigm of *in vitro* modeling and has led to the development of model systems that better recapitulate this important aspect of tumor biology, which are discussed in detail in subsequent sections.

The tumor microenvironment consists of a heterogeneous mixture of multiple cell lineages which range from normal supportive cells, infiltrating immune cells, vascular endothelium, and transformed tumor cells.[22,23] Through understanding the cellular composition of a tumor, multiple prognostic factors and drug targets have been identified.[22] For example, in breast cancer tumor-stroma interactions can be used to predict the course of the disease by analyzing the gene expression profile of stromal cells, opposed to traditional prognostication based on genetic markers of transformed cancer cells, demonstrating the importance of paracrine signals being provided by the stroma.[24] In other diseases, such as cervical cancer, targeting of platelet-derived growth factors supplied by cancer-associated fibroblasts and pericytes presents a new and promising molecular target based on targeting the cellular heterogeneity of the tumor stroma.[25] In yet another example, the fraction of infiltrative $CD4^+$ T helper cells or $CD8^+$ T killer cells serves as a prognostic marker for disease survival in gliomas and has the potential to predict the efficacy of utilizing immune-based therapies.[26] Thus, it is clear that understanding the heterogeneous milieu of cells that compose a tumor has provided novel prognostic factors and targets used in the treatment of cancer.

Tumor heterogeneity may also refer to heterogeneity in the genetic and epigenetic composition of the tumor. The classical model of cancer development assumed a tumor consisted of relatively homogeneous clonally derived cells that had undergone linear evolution.[22,27] However, in light of the observance of high levels of genomic instability and elevated levels of heterogeneity at the genomic level, this notion has shifted towards a branched model where different clonal lineages co-evolve together and those lineages that have accrued mutations that confer an increase in fitness will grow out to form the bulk of the tumor.[28] In terms of utilizing targeted therapies, this presents additional challenges because the targetable profile of the various clonal lineages that compose the tumor may be different and can alter over time, especially in the presence of selective pressures. Indeed, treating cancer using a single target agent has been generally unsuccessful because the provision of a single selective pressure can result in the outgrowth of a resistant sub-population.[27,29] Thus, it has become increasingly more common to use combinations of agents to avoid the development of secondary resistance and minimize toxicity. In addition to genetic heterogeneity, tumor cells also demonstrate epigenetic heterogeneity. This level of heterogeneity can result from paracrine signaling between the tumor cells and the stroma, or as a result of hypoxic conditions, to name a few potential contributors. Epigenetic heterogeneity can similarly provide alterations in the evolutionary fitness of cancers, result in secondary resistance through regulation of the gene expression profile, and produce novel stem-like phenotypes.[30] The observance of epigenetic heterogeneity has led to the exploration of combination therapies, which inhibit epigenetic processes to homogenize the population and increase the susceptibility to a secondary agent.[31,32]

Review of cell-based model systems used for HTS

It is clear that tumor heterogeneity is a repeating motif in tumor physiology and that it is important in the development, growth, and spread of multiple cancers. Unfortunately, many of the models that are used in biological research, such as immortalized cell lines propagated for many years as monolayers, fail to fully recapitulate this aspect of tumor physiology. Despite this limitation, immortalized cell lines grown as monolayers do have value in that they have been instrumental in the establishment of many of the driver mutations in cancer in addition to uncovering the mechanism of action that drugs work through. Indeed, drug sensitivity profiles generated across panels of immortalized cancer cell lines, representing different tumor lineages, are able to cluster drugs by mode of action and cell lines by linage specific sensetivities.[33] From a logistical perspective, immortalized cell lines are available to the broader research community, are isogenic, and well-characterized. These features facilitate bench marking and peer validation of results, which are important for scientific advancement and credibility. To address the limitation of immortalized cell lines, researchers have suggested testing the effects of potential therapeutics in panels of representative cell lines, which provides genomic heterogeneity and allows researchers to estimate the generalizability of the observed response. Others have moved away from immortalized cell lines towards patient-derived models, which recapitulate the genetic and epigenetic heterogeneity found in the clinical setting. The gold standard for establishing drug sensitivity data for patient-derived cell lines has been xenographt mouse models.[34] However, these are limited in the number of drugs that can be tested due to the amount of time and cost associated with performing these types of studies. For these reasons, adapting patent-derived cell cultures towards HTS is a major trend in the field. In this context, others have established biobanks composed of patient-derived cell lines that form organoids *in vitro*; a system which further benefits from its ability to recapitulate physiological features which are discussed later in this section.[35]

Another family of methods being explored to increase the translational relevance of HTS rely on the integration of physiologic relevant features into cell culture models.

Importantly, these systems model non-genomic characteristics of a tumor which include: the formation of 3D cell-cell interactions, secretion of extracellular matrix (ECM) and formation of cell-ECM interactions, chronic hypoxic cores, proliferative margins, metabolic heterogeneity, drug permeability gradients, and spontaneous development of secondary chemoresistance and radioresistance.[36-38] To integrate these features, traditional *in vitro* cell culturing methods which form monolayers that adhere to a solid interface are giving way to non-adherent multicellular aggregates grown in 3D suspensions.[39] At present, multiple methodologies have been proposed to facilitate the formation of cells into multicellular aggregates, the benefits and limitations for which are discussed in the sections below.

The first 3D culturing methods discussed here are those that rely on using a supportive matrix to facilitate the growth of clonally derived multicellular aggregates. Typically, these systems either use animal-derived basement membrane hydrogels or defined synthetic extracellular matrix scaffolds, which were initially characterized in the mid-1980's, to support the growth of cells by providing physiological cell-ECM interactions.[40-42] Importantly, matrix-based systems have been demonstrated to support the growth of multicellular aggregates in multiple established cell lines and in primary patient-derived cell lines (PDC). Indeed, prolonged culturing of primary PDC has been developed using scaffold-based technologies, which are now being used for personalized pharmacologic profiling applications.[35,40,43] Therefore, scaffold-based technologies achieve increased biological and translational relevance through supporting the growth of cell lines that are otherwise difficult to culture, such as PDCs, and by providing a supportive growth matrix which facilitates non-adherent 3D multicellular aggregate formation. In the context of HTS, scaffold-based model systems are able to be adapted to multiwell plates using automation, and can be optimized to generate a few hundred to thousands of clonally derived organoids per multiwell plate. While this feature helps improve population-based statistical analysis, it also presents certain limitations; including heterogeneity in the size and morphology of spheroids, which can result in novel cellular behaviors and may confound the results of certain analyses.[44] For HTS campaigns using cells grown in a matrix, biochemical endpoint analysis is often performed over image-based methods. In part, this is due to the resulting organoids growing in different focal planes and issues with automated focusing through the supportive matrix.[43] Interestingly, it has been my observation that imaging is often done as a method of quality control for these screens and has potential to be highly informative, pending the development of further methods. Other established limitations in scaffold-based methods deal with the reproducibility in the composition of animal-derived hydrogels, which often display a high degree of inter-batch variation.[45] For this reason, the use of synthetic matrix sources, such as the commercially available product Matrigel (Corning), are commonly used. While this synthetic matrix decreases variability associated with the composition of the matrix; it is limited in that the full repertoire of ligands that are present in animal-derived extracellular matrices are not well represented, resulting in the inability of synthetic matrices to facilitate as diverse of a range of cell lines. Finally, these systems also often require prolonged culture in order to generate detectable organoids from a single cell.

The next method discussed produce a single uniform spheroid using specialized multiwell plates or plating accessories. These include hanging drop, ultra-low attachment (ULA) round (U) bottom plates, and magnetic bioprinting, with each method presenting certain benefits and limitations.[36,46-49] Each of these systems facilitate in the rapid formation of large multicellular aggregates that grow in a non-adherent manner. Furthermore, the characteristics of spheroids with a diameter greater than 300-500 microns recapitulate the physiological features of small avascular tumor nodules and micrometastisis as demonstrated by the presence of chronically hypoxic cores, formation of 3D cell-cell contacts, resistance to chemotherapies and radiotherapies that mimic clinical observations, and establishment of drug permeability gradients.[37,50] To further increase the physiological relevance of these systems, cells may also be co-cultured using combinations of established cell lines, PDC, as well as stromal and immune cells in order to mimic the cellular and genomic heterogeneity encountered in a tumor.[51,52] Thus, these methods are able to provide highly biological and translational relevant models amenable to HTS, but are also associated with certain limitations which are briefly described here. Hanging drop culturing methods were adapted from embryonic stem cell biology and represent the oldest non-scaffold-based method used for spheroid production.[53,54] These systems can robustly generate uniform spheroids, but are limited in HTS applications by their difficulty to be manipulated using automation and the requirement of the spheroids to be transferred into a separate apparatus in order to be analyzed. ULA U-bottom plates have a similar performance in the ability to robustly generate uniform spheroids, but are more readily manipulated with automation tools and can be directly used for both image-based and biochemical assays. However, ULA U-bottom plates are limited by the inability of the microscope to focus on a non-uniform surface and the inability of the manufacturer to generate a high-standard optical bottom. Finally, magnetic bioprinting systems supplement the growth media with magnetic nano-particles, resulting in cells that can be manipulated by a magnetic field.[48] When combined with a specialized magnetic plate rack, cells are drawn together in the center or corner of a well, stabilizing cell-cell interactions and allowing a single aggregate to form. This method is limited in that it requires supplementation of exogenous materials into the cells, which can introduce artifacts and requires highly specialized equipment.

The final model systems aimed at increasing the physiological relevance of *in vitro* systems are organ-on-chip and microfluidic cell culturing methods. These models use combinations of supportive matrices, advanced co-culturing techniques, and specialized apparatuses that facilitate in the maintenance and formation of mimetic tissue. To date, multiple organ-on-chips have been developed as models for liver, skin, intestine, and kidneys.[55] Additionally, different organ systems can be linked together to mimic the physiology of entire organisms.[55] While many of these systems have yet to be fully developed for the use in HTS, they do have the potential to model novel physiological features and test conditions that are not readily achieved by other 3D systems. Thus, these models can achieve high levels of biological significance through generating mimetic multicellular tissue, but require highly specialized equipment, optimization, and expertise to generate the chip and perform subsequent experiments, yet, may represent a novel technology with the potential to impact HTS.

With the development of these novel cell culture systems, there is an increased pressure to develop new methods to quantify the effects of drugs using these systems. Likewise, the development of novel approaches or adaption of pre-existing biochemical and phenotypic endpoints must be explored. At present, both biochemical and image-based phenotypic assays are in development and applied in HTS. The most common biochemical approach uses ATP as a surrogate of viability using reagents such as CTG. However, it has been observed that CTG is partially confounded by the presence of different morphologies rather than alterations in cellular viability.[56] In light of this, Zanoni *et al.* have proposed to prescreen spheroids and only select morphologically similar spheroids to be included in analysis. This highlights the importance of performing an image-based analysis used to phenotype spheroids, and argues that an image-based analysis may be more informative. Indeed, multiple phenotypic assays have been previously reported in the literature and consist of both label-free and fluorescent methods. Label-free methods do not interfere with cellular behavior and can be performed readily across many different cell lines. These methods typically measure alterations in the equatorial area, diameter, perimeter, and circularity as surrogate measurements of viability. Other high content image assays have been performed, which combine morphometric analysis with fluorescent dyes that measure viability using caspase activity (CellEvent), membrane permeability (Propidium Iodide), and metabolic potential (Calcein AM).[46] However, caution should be applied when quantifying fluorescent intensity in larger spheroids, which may be confounded by the inability of light to penetrate into the core of the spheroid and a high amount of optical scattering occurs. In turn, this results in the artificial depletion of signal in the central region of the spheroid which may represent cell death or optical proclusion.[57] Other low throughput image-based method analyses have been applied to the characterization and validation of spheroids. These methods provide a very high content snapshot of the composition, structures, and status of cells within a spheroid by applying antibody labels to sections from formalin-fixed paraffin-embedded spheroids.[58]

Statistical methods used for assay quality control and data normalization

HTS relies on the miniaturization and automation of a biological assay, which facilitates the ability to collect hundreds to thousands of experimental conditions simultaneously. Often times, the results of a screening campaign are obtained from aggregating the results of many experimental batches together. Likewise, there is a potential that hits may be falsely identified through random and/or systematic variation associated with a particular batch of experiments, which is termed as “batch effects”. Indeed, in the early days of HTS this was considered a major limitation of the technology that needed to be reconciled in order to transform HTS into a *bona fide* drug discovery method.[59,60] Therefore, the integration of specialized statistical analysis aimed at bolstering confidence in the reliability of HTS data has been a major development in the field.[61-64] Likewise, it is important to understand the potential sources of error that are encountered in HTS and the appropriate use of statistics to detect and correct for these errors.

Multiple factors contribute to inter-batch variation including the exact environmental conditions, alterations in cellular behavior, and the quality of different experimental procedures that have been performed. Each of these factors can result in an over or under representation of data obtained from a given experimental batch and are known as “batch effects”, as previously mentioned. This type of error is distinguishable from random error, which is the observance of alterations in the data that occur from random chance, and can be identified and handled using replication. In contrast, the correction of batch effects often requires additional post-processing and statistical normalization used to generate comparable metrics across all the batches in a screening campaign. In order to understand how to correct for batch effects, it is important to first identify the potential sources of error so that relevant statistical analysis may be performed. To discuss the potential sources of batch-related error we subdivide it into two classes: systematic and spatial. In the context of high throughput image-based assays, systematic batch effects uniformly alter the magnitude or range of the data. In contrast, spatial batch effects introduce biases in the data as a function of spatial organization in a plate. Spatial batch effects are harder to account for and include radial vignettes, gradients, or striping patterns across a multiwell plate. In the following sections, we discuss the various types of batch effects in addition to the underlying issues that cause them and present appropriate statistical manipulations that can be used to correct for these issues.

When addressing high throughput data it is important to systematically evaluate the quality of the data across plates and use this information as inclusion criteria for the larger dataset. In a well-designed study, each plate will contain clearly defined positive and negative controls which are used both in assay robustness assessment and in normalization, which is discussed in the subsequent section. The most common method used to quantify the results of a high throughput run is through the calculation of a metric known as the Z prime factor (Z'), shown below.

$$Z' = 1 - \frac{3(\sigma_p + \sigma_n)}{|\mu_p - \mu_n|}$$

Where σ_p and σ_n represent the standard deviations and μ_p and μ_n are the means of the positive and negative controls, respectively. However, this method relies on how well the mean and standard deviation represent the data and may be subject to outliers. To increase the robustness of this metric, a robust variant of Z' may be used, which substitutes the mean and standard deviation with the median and mean average distance (MAD) of both reference populations.[62,65] Other methods used to qualify the validity of an experimental batch include strictly standardized mean difference (SSDM) and taking the area under the curve (AUC) of a receiver operator characteristic curve (ROC).[66] The advantage of using SSDM is that it is linked to a strictly probabilistic model of the data and does not depend on sample size. However, SSDM is limited in that the thresholds used to determine the quality of the assay are based on subjective classification of the strength of the controls. Utilization of the AUC of a ROC curve method is useful for data that is not normally distributed and can be used to optimize a threshold used to determine a positive result. This method is limited due to its requirement of larger sample sizes. Another commonly utilized method to determine the consistency of the results of batch is the minimum significance ratio (MSR) which quantify the error between calculated EC50 values. This is useful in that it provides a method to evaluate the consistence of dose response data. Functionally this is calculated using the formula shown below where S is the standard deviation of EC50 values and a MSR value less then 3 is generally considered to be good:

$$MSR = 10^{2\sqrt{2}S}$$

Other important quality control metrics that should continuously be monitored through out a screening campagn are baseline signal and signal to noise ratio. Alterations in base line single can arise from differences in factors such as alterations in cell seeding density which intern has the potential to alter the rate of growth, shape and location of fitted dose response curves based on these metrics.[64] Therefore, the ability to reliably seed the same number of cells between batches is important in the generation of comparable data across batches.

After assay validation and assurance that the data are of an acceptable quality, correction for systematic variation in the data can be evaluated and managed. The source of systematic errors can include variation in the way an assay was performed, alterations in the settings of the instrument, and degradation of the reagents used to perform the assay. The most common method to account for systematic batch effects rely on data normalization, which are applied to scale the data to a standardized range using on-plate controls. The most fundamental method of scaling data is calculation of a fold change, which is formulaically shown below:

$$FC = \frac{X_{ij}}{\mu}$$

Where Xij is the individual observation being normalized and μ is a representative value for a control, typically the negative. Likewise, a representative value can be calculated by taking the mean, 50%-trimmed mean, or median of the control. The latter two methods are typically used to provide a more robust representation of the data and limit the potential effects of outliers from random variation. Fold-change normalization is applicable in situations where the controls exhibit minimal variance and are approximately Gaussian. In datasets where the confidence interval of the controls is known or calculable, a z-transformation may be more appropriate and is formulaically shown below:

$$Z = \frac{X_{ij} - \mu_n}{\sigma_n}$$

Z-normalization scales and centers the data to the mean and standard deviation of a control, such as the on plate negative control. For similar reasons mentioned previously, the median and MAD may also be used as representative values of the control and when used are referred to as a robust z-score. Similar to fold-change, z-scoring based methods assume the data is approximately Gaussian. Fraction affected normalization scales raw data to the positive and negative control, and is shown bellow.

$$FA = \frac{X_{ij} - \mu_{Neg}}{\mu_{Pos} - \mu_{Neg}}$$

This scales the data from 0 to 1, making AC50 values easy to determine and does not assume the underlying data is Gaussian. Examples of where the above-mentioned statistics are appropriately applied in the normalization of phenotypic screening data include the standardization of fluorescent and morphometric features to on-plate controls. Interestingly the above mentioned statistical normalizations can also be used to standardize pixel values in the image matrix.

More recently, additional statistical methods have been suggested for assays where cell counts are used as the endpoint, with the rational that this metric is potentially confounded by multiple factors, including the cell growth rate and time point from which the data were collect. To correct for these issues, it has been proposed to correct the cell number to the initial cell count at the time of treatment and growth rate of the negative control.[64,67,68] Functionally this can be achieved using the following formula:

$$GR = \left(2^{\frac{\log_2(FC_{ij})}{\log_2(FC_n)}}\right) - 1$$

The use of normalized growth rates as a surrogate of viability simultaneously removes the effects of the rate of growth on cell count data, allows one to infer the biological mode of action (i.e. whether a drug is non-effective, cytostatic, or cytotoxic), and minimize inter-batch variability.

Compared to systematic batch effects, spatial batch effects represent a more challenging artifact encountered in high throughput screening campaigns. Spatial artifacts can occur on an isolated plate or systematically across an entire batch. In general, multiple methods of statistical normalization have been proposed to adjust for spatially related batch effects; however, many researchers question the validity of the data because these types of artifacts are usually due to a systematic failure of various automated processes used to perform the experiment. Therefore, in the assessment and correction of spatial batch effects, we will discuss procedural modification in assay design, the best lab practices to avoid these artifacts, methods to detect and differentiate the various types of spatial artifacts, and the use of appropriate statistical methods for correction. The first spatial batch effect that we will discuss is radial gradient, also known as “vignettes”. The source of this type of pattern is usually due to evaporation of the media on the outer edges of plates, which can increase salt and drug concentrations in wells located in the outer region of a plate. Likewise, this artifact is generally more apparent in assays that require prolonged incubation. Detection of this artifact is most readily apparent when viewing data in the form of a plate heatmap, but can also be detected by a poor Z’ value and by evaluating the averaged values of the rows and columns. Procedural modifications to account for this include randomizing the well positions, performing subsequent replicate analysis on different plates, and averaging values together. While this method is a statistically simple way to remove the effects of arbitrarily positioned drugs on a plate it is procedurally more difficult to randomize treatment positions in the context of a large HTS. Therefore, if a static plate layout is used one can simply apply gas permeant stickers or lids to the plate which have been shown to improve the uniformity of a plate.[43] Other considerations relevant to handling this artifact is to increase the humidity in the incubator where plates are stored, proper maintenance of incubators, and increasing the volume of media. Despite utilizing the best procedural practices, radial gradient patterns can still occur. In this event, statistical correction may be applied, which include utilizing global parametric smoothing, local fitting, B-scores, and model based approaches.[61,66] Each of these methods attempts to estimate the contribution of the observed artifact to the measured signal using different advanced statistical models which are used to correct spatial bias. However, these methods should be used sparingly because the inappropriate application to plates not exhibiting a vignette can result in secondary artifacts.[61,69] Other spatial batch effects, such as row and column biases, can be detected via visual inspection of plate heatmaps, or by calculating the mean and standard deviation of rows and columns and performing statistical tests, such as t-tests, Wilcoxon test, or ANOVA. Using these statistics allows the evaluation of the null hypothesis, which is that no row or column effects are present and the mean and variance of the rows or columns are the same. However, this method assumes that the majority of the data are inactive and may result in significant differences without the presence of a true artifact. A practical example of when this occurs is when dose curves are present on the plate that results in a gradual change in signal across the plate. Other more relevant sources of row and column batch effects are failures in cell seeding robots, drug transfer failures, loss of focus, and procedural failures during plate production or processing. Thus, if this type of artifact is detected it is best to exclude the data from those plates and redo the experiment rather than attempt to correct the data with statistical normalization. In conclusion, the recognition of batch effects is critical in the integration and ultimately the interperatation of HTS results. There are multiple types of batch effects that can be detected and classified into different categories using visual and statistical methods. Likewise, the appropriate classification of the nature of the batch effect can inform on the proper use of statistical methods used for correction. When reviewing pre-existing data, it is important to have all the information regarding treatment conditions and batches to determine if systematic errors occurred and how to best correct for them. Likewise, archiving HTS data in a format that allows for retrospective analysis needs to become standardized.

CHAPTER III

BIOLOGICAL VALIDATION OF *IN SILICO* SCREENING

Synopsis

In this chapter, I perform a validation experiment for an ultra-high throughput *in silico* screen using a relatively simple cell-based assay. Importantly, through the development of a cell-based phenotypic screening system, we increase the biological relevance of the primary *in silico* screen by providing empiric data directly quantifying the endogenous levels of nuclear localized p27 when treated with small molecules aimed at rescuing this target. Additionally, these data also indirectly confirm that the lead molecules can cross the cell membrane, provide functional data, and help determine on- vs off-target cytotoxicity. Finally, when combined with structural information of the molecules, a quantitative structure activity relationship (QSAR) is observed, which can guide future modifications that may be made to the core structure with the aim of improving the core scaffold towards a translationally relevant drug.

Introduction

Targeted therapies consist of biological and chemical probes designed to modulate the activity of specific targets. Likewise, the utilization of targeted therapies presents an opportunity to specifically interfere with proteins or other targets that are associated with the development or progression of human disease.[70,71] As the scientific body of knowledge continues to expand, new molecular targets and mechanisms of action (MoA) have been identified. However, in order to successfully capitalize on newly identified targets and MoA, one must identify molecular probes that can modulate them. Traditionally, discovery of molecular probes was achieved by performing large HTS campaigns on *in vitro* models treated with diversity libraries consisting of tens of thousands to millions of small molecule scaffolds. Indeed, this strategy has successfully identified a large number of GPCR, kinase, and ion transport modulators which have been translated into clinically utilized drugs.[5] However, the massive number of chemical scaffolds that are required to identify lead molecules using “shotgun” HTS approaches make these methods highly inefficient and have failed to meet the demand of generating new drugs against recently characterized targets.[60] This has altered the traditional HTS paradigm towards strategies which employ focused and fragment-based libraries in addition to utilizing *in silico* approaches to predict drug activity to minimize the time and cost associated with drug development.[59,72] Of these strategies, *in silico* modeling has a high potential to rapidly identify lead molecules; however, it is limited in that they do not fully capture the complexity of a biological system and therefore require further biological validation. Thus, utilization of a comprehensive screening strategy that combines *in silico* lead identification with empiric testing in biological systems balances the need for high throughput and biological relevance.

To exemplify the role of phenotypic screening in the biological validation of *in silico* screening, I present a team-based project where the goal was to identify a small molecule inhibitor that selectively restores nuclear p27 through interfering with the protein-protein interaction between the specificity domain of an E3-ligase complex and p27. The p27 protein is a cyclin-dependent kinase inhibitor (CKI), for which the primary role is to negatively regulate the cell cycle by inactivating cyclin A- and E-CDK2 complexes in the nucleus. While aberrant cell growth is a hallmark of tumorigenesis and defects in cell cycle checkpoint proteins are common in cancer cells[73], mutations of the *p27* gene in cancers are rare, at approximately 3% across all cancers.[74] The ability of p27 to inhibit cell cycle progression is predominantly regulated by post-translational modifications, which determine the stability and subcellular localization of p27. Key regulatory marks include phosphorylation of p27 at serine 10 (S10), which stabilizes and sequesters this CKI in the cytoplasm; and at threonine 187 (T187), which targets it for degradation in the nucleus by the SCF^{SKP2} complex.[75,76] When localized to the cytoplasm, p27 represses RhoA, leading to reorganization of the actin cytoskeleton, diminished focal adhesion formation, and promotion of cell motility, in addition to inhibition of apoptosis and promotion of autophagy.[77] Furthermore, increased CDK2 activity caused by loss of nuclear p27 results in phosphorylation of the estrogen receptor (ER), thereby increasing ER activity and promoting growth of hormone-dependent tumors. Thus, p27 is a tumor suppressor in the nucleus, but in the cytoplasm acts as an oncogene. Notably, p27 is targeted for degradation differently in the cytosol vs. nucleus. In the nucleus, p27 is specifically targeted by the S-phase kinase-associated protein 1 and 2 (Skp1/2)-Cullin-1 (Cul-1)-F-Box (SCF) ubiquitin E3-ligase; while in the cytoplasm, p27 is targeted for degradation by the Cul1-KPC1/2 E3-ligase.[78,79] Specificity for p27 binding to SCF is conferred via the formation of a pocket between Skp2 and the cyclin-dependent kinase subunit 1 (Cks1). Indeed, pharmacologic rescue of p27 through targeting different interactions in the SCF^{SKP2} complex has been previously explored and published.[80,81] However, the small molecules identified in these papers are limited in that they require super-physiological doses (high micromolar range) to achieve appreciable activity, lack selectivity, and do not necessarily inhibit E3-Ligase activity of the complex which has been discussed elsewhere.[82] Thus, there remains an unfulfilled need to identify potent and selective inhibitors that target this mechanism in order to advance this class of molecules into a useful translational agent. To address this need, we hypothesized that targeting the CKS1:SKP2 specificity pocket, over targeting SCF formation or SKP1:SKP2 interactions, represents a novel targetable region with the potential to effectively rescue nuclear p27.

We developed a comprehensive team-based screening strategy, schematized in Fig 3.1A. The first step leverages *in silico* screening and was done in collaboration with the Zhang lab at MDACC. In brief, this method used the publicly available crystal structure of the CKS1:SKP2 binding interface acquired from PDB[83] and screened against a diverse library of small molecules using HiPCDock and GOLD virtual screening methods.[84,85] This resulted in the identification of 164 high-priority unique chemical agents, a visualization of the molecular docking, and stick structure for the top two scaffolds, which are shown in Fig 3.1 B,C. Of the high-priority molecules, 78 were commercially available and 16 were synthesized according to the suggestion of a medicinal chemist. Therefore a total of 94 compounds were tested in the cell-based HTS assay which was developed and validated in the subsequent sections.

The goal of developing a HTS cell-based validation assay is to confirm the findings of the *in silico* screen, thereby providing increased biological relevance of the results. The first consideration made during assay development is the selection of a relevant model system. In cancers such as endometrial carcinoma (EndoCa), which is the fourth most common cancer in women and has no targeted therapy currently approved, loss or mis-localization of p27 to the cytoplasm occurs frequently (~56%) yet remains unmutated; providing a compelling translational rationale for targeting p27 as a therapeutic strategy for this cancer. Therefore, we validated this screening method in multiple prototypic cell line models for endometrial carcinoma; however, due to cost associated with performing an antibody-based primary screen, only the most well-behaved cell line model, HEC1b, was tested in the primary screen. The top confirmed hits are now being subjected to a panel of secondary biochemical assays in multiple cell lines to further confirm the mechanism of action. The next major consideration in the construction of the phenotypic assay is the selection and validation of quantitative endpoints. The method we developed evaluates the ability to rescue nuclear p27 using an antibody conjugation system that measures endogenous p27 levels in response to small molecules. As p27 is a potent negative regulator of the cell cycle, the anti-proliferative effects of the drug are also quantified as an orthogonal functional read out. Thus, this method directly quantifies the functional rescue of p27; however, it is also limited as it does not directly confirm that the small molecules are functioning through the proposed mechanism of action. Finally, quantitative chemical features and empiric activity are used to perform QSAR to further understand the chemical attributes associated with biological activity.

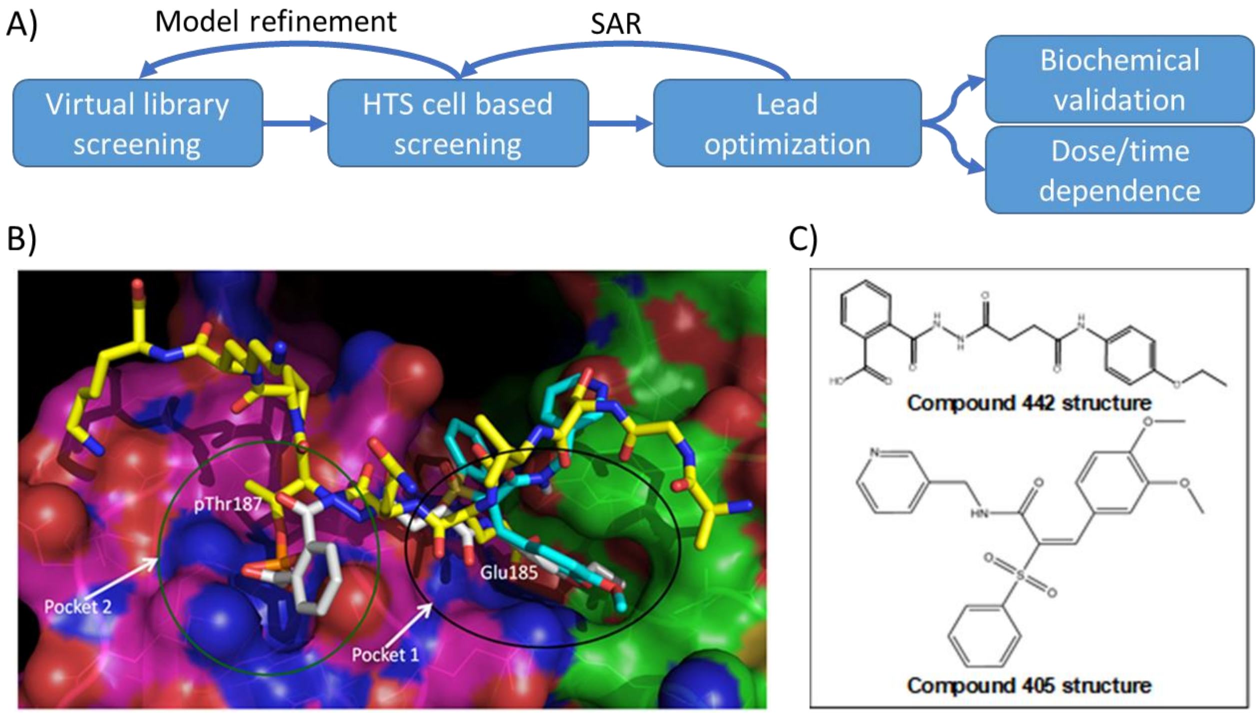


Fig 3.1) Comprehensive screening strategy and preliminary *in silico* results. A) Schematic summary of the screening strategy used to identify and optimize lead molecules with the potential to inhibit the SKP2:p27 interaction. B) Surface plot of SKP2 (green) and CKS1 (magenta) binding interface showing co-crystallized p27 (yellow sticks) overlayed with two lead molecules CMPD405 (gray) and CMPD442 (cyan). Pocket 1 is on the interface of Skp2/Cks1 where the Glu185 of p27 lies. Pocket 2 interacts with phosphorylated Thr187 of p27 and is primarily on Cks1. C) Structures of two lead scaffolds which we refer to as CMPD 442 and STM 209405.

Materials and Methods

Cell culture conditions

HEC1b (ATCC: HTB-113) were acquired from ATCC and cultured in DMEM (ThermoFisher, 11965) supplemented with 10% Fetal Bovine Serum (FBS) and incubated in a humidified 37℃ incubator at 5% CO_2. Cell stocks are made by suspending cells in 10% DMSO-FBS and flash frozen in liquid nitrogen. Cultured cells are passaged for up to 5 times before a fresh stock is brought up. Cells plated for a HTS/HCA are seeded onto replicate 384-well plates (Griener Cat: 781091) at a density of 6000 cells per well and allowed to recover in growth media overnight.

Drug addition

After cells have sufficiently recovered and formed a monolayer, cells are treated in 4-point dose response ranging from 1.25 µM to 20 µM with replicate wells on different plates. Compounds are arrayed on a low dead volume plate (Labcyte) and are subsequently transferred into assay plates using an Echo Acoustic dispenser (Labcyte). At the time of treatment, an additional plate is fixed and DAPI labeled which is used to establish a cell count at the time of treatment, which is used during statistical normalization. Each plate contains on plate positive (20 µM ST029405) and negative controls (DMSO) to assess the quality of measurements across plates.

Immunocytochemistry

Following a 24-hour incubation in the presence of compound, the plates are fixed and processed using an automated Biomek FX (Beckman) liquid handling platform. In brief, plates are fixed using a 4% paraformaldehyde-PBS solution for 30 minutes at room temperature (RT). Cells are then permeablized with a 0.5% Triton/PBS solution for 20 minutes at RT. Following 2 rounds of washing, a 5% milk/PBS blocking buffer is added for 1 hour at room temperature. Next, cells are labeled with a commercially available p27 antibody (Cell Signaling: D69C12) which is diluted 1:1000 in 5% milk/TBST blocking buffer and incubated overnight at 4^{o}C. The plate is then vigorously washed followed by adding a secondary antibody, Alexa Goat anti-Rabbit 546 (Invitrogen), diluted in blocking buffer and allowed to incubate for 1 hour at 37^{o}C. The cells are then washed and fixed again in order to stabilize the signal. Lastly, the cells are counter stained with DAPI in order to visualize the nucleus.

Image acquisition and analysis

Plates are imaged using an imageXpress micro confocal (Molecular Devices) using widefield settings. A total of four fields per well are sampled with a 10x NA=0.45 Plan Apo lens (Nikon), which covers the entire surface area of the well. Image analysis is performed using a custom script developed using the Imaging package in Pipeline Pilot (Dassault Systemes BIOVIA, Version 2018, San Diego) . In brief, images are first background subtracted in order to minimize variance due to background effects. Next, the nuclei are segmented using thresholding to initially identify putative regions, which are subsequently smoothed with morphological operators, and finally neighboring objects are separated using a marker directed watershed. Debris are filtered using nuclear size and intensity filters. The mean intensity value for the p27 label is then measured for each cell and summarized at the well level by taking the mean value. Additionally, cell counts from a plate fixed on the day of treatment and from wells treated with drug are collected.

Statistical normalization

Intensity (I) data are normalized using a robust z-score (RZ) transformation from the median (MED) and MAD of the DMSO treated wells according to the formula shown below:

$$RZ = \frac{I - MED_{DMSO}}{MAD_{DMSO}}$$

Normalized rate of growth are calculated using the method described in the NCI60 screens which is formulaically represented below where C_i is the cell count at time of harvest, C_0 is the cell count at time treatment and C_{DMSO} is the cell count of the negative control:[67,68]

$$Growth = if(C_i \geq \bar{C}_0)\left\{\left(\frac{C_i - \bar{C}_0}{C_{DMSO} - \bar{C}_0}\right) * 100\right\} Else \left\{\left(\frac{C_i - \bar{C}_0}{\bar{C}_0}\right) * 100\right\}$$

In brief, this method scales the response from one to negative one where values between one and zero represent varying degrees of cytosuppression and values between zero and negative one represent the extent of cytotoxicity. One deviation from this method from what is used is we offset the values by 100 to ensure all values are positive, making the values 200, 100, and 0 landmarks for inactive, cytosuppressive, and cytotoxic, respectively. RZ-intensity values and normalized growth rate values are then plotted by concentration and fitted to a 4-parameter logistic curve using the “robust” package in R/Pipeline Pilot. Here, the Normalized MAD and Tukey bi-weight function are used to minimize the influence of potential outliers. AUC values are calculated via numerically integrating fitted curves. Statistics, data integration, and visualization are automated using the Analysis and Statistics package in Pipeline Pilot.

Quantitative structure activity relationship

To establish a structure activity relationship, chemical attributes are classified against the empiric drug activity from the cell-based screen. Chemical attributes are calculated from the structure of the molecule using components in the chemistry library of Pipeline Pilot. The chemical features used in the model describe attributes related to formal charge, number of chiral centers, average bond length, total number of atoms, bonds, hydrogens, positive and negative charged atoms, rotatable bonds, aromatic rings, chain assembles, hydrogen bond acceptors/donors, and 3D volume and surface descriptors. Compounds are determined to be active when the AUC of the p27 vs concentration curve is larger than 100 (activity cut off). The robustness of the model is evaluated by calculating the AUC of the receiver operator curve using the categories mentioned above. Further statistical characterization of the model is performed on the optimal confusion matrix using the following formulas; where TP = true positive, TN = true negative, FP = false positive, and TN = true negative.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Missclassification\,rate = \frac{FP+FN}{TP+FP+TN+FN}\backslash$$

$$True\ Positive\ Rate = \frac{TP}{TP + FN}$$

$$False\ Positive\ Rate = \frac{FP}{FP + TN}$$

$$Specificity = \frac{TN}{TN + FP}$$

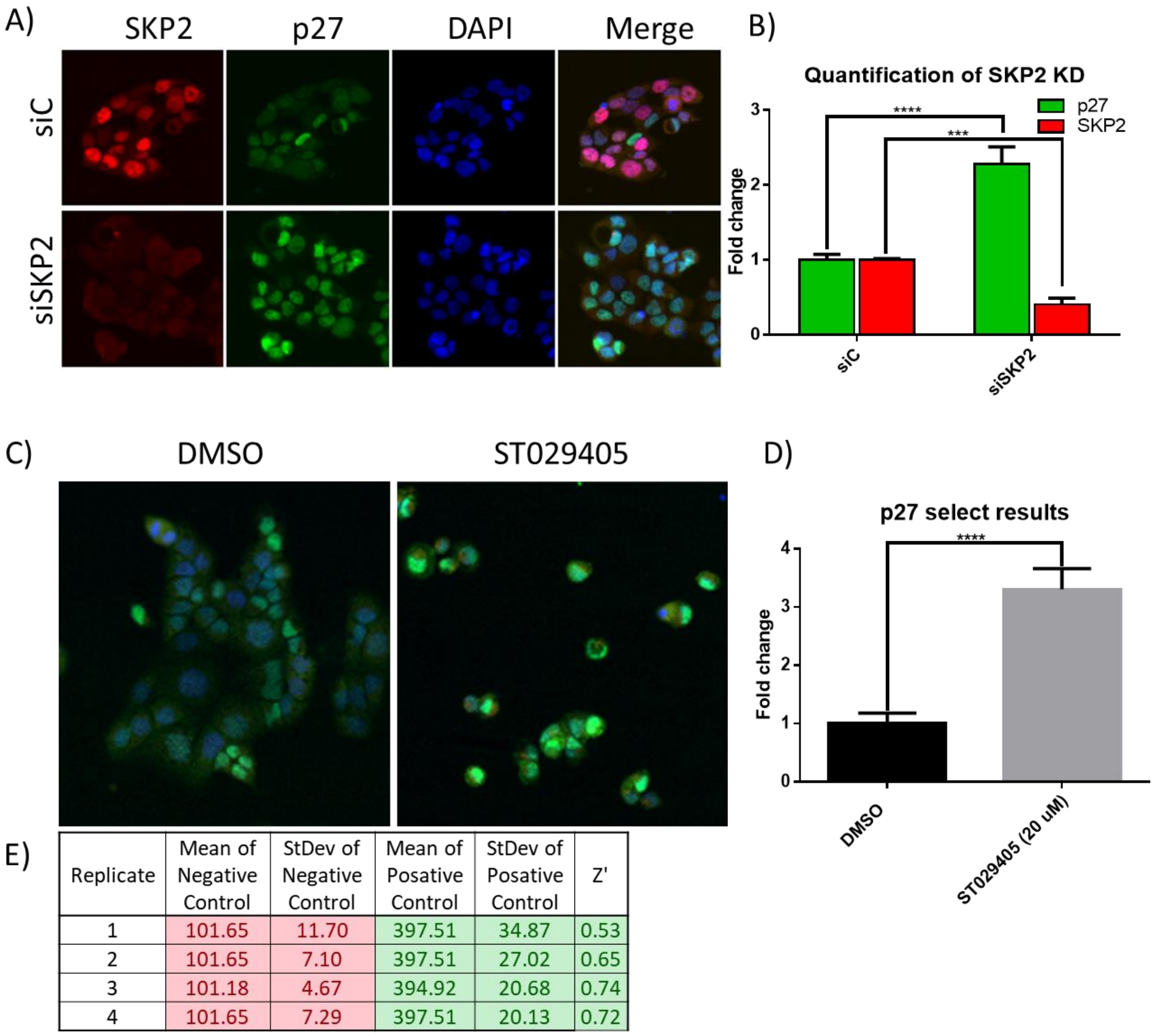
$$Precision = \frac{TP}{TP + FN}$$

$$F - measure = \frac{2 * precision * true\ positive\ rate}{precision + true\ positive\ rate}$$

Results

HTS assay validation

In order to confirm the biological activity of lead molecules identified by the *in silico* screening method we first had to develop a HT cell-based assay. Here, HEC1b cells were used as a model of endometrial cancer and were plated in 384-well plates. Next, the cells were immunofluorescently labeled with p27 in order to directly measure endogenous concentrations of nuclear p27. In rapidly growing cells p27 levels are generally low; therefore, in order to validate the responsivity of the p27 antibody, we knocked down SKP2. These data show a significant depletion in SKP2 protein levels and concurrent elevation in nuclear p27 levels in the cells where SKP2 had been knocked down (siSKP2) when compared to the control (siC), Fig 3.2A,B. This demonstrates both that the antibody is responsive to changes in p27 levels and that targeting SKP2 is a valid method to selectively upregulate nuclear p27. Next, we tested this method on a focused library of small molecules initially identified from the *in silico* screen which had preliminary data demonstrating activity. These data were used to provide validation of the previous findings and the image-based screening method (data not shown). From this analysis, we confirmed that ST029405 significantly elevates levels of p27 in the nucleus and arrests cell growth (Fig 3.2 C,D). To further validate the image-based assay, we performed statistical analysis using ST029405 as a positive control and DMSO as a negative control. Using these controls, we calculated Z’ values as a function of cell density, the number of field acquired per well, and number of replicate wells tested. From these data we found that a seeding density of 6000 cells per well, sampling 4 fields per well with four replicates resulted in Z’ values of approximately 0.7 (Fig 3.2 E), indicating a highly robust sufficient to perform screens on.



Replicate	Mean of Negative Control	StDev of Negative Control	Mean of Posative Control	StDev of Posative Control	Z'
1	101.65	11.70	397.51	34.87	0.53
2	101.65	7.10	397.51	27.02	0.65
3	101.18	4.67	394.92	20.68	0.74
4	101.65	7.29	397.51	20.13	0.72

Fig 3.2) Evaluation of reagents and statistical validation of an image-based p27 localization assay. In order to validate the materials used in the development of this assay, HEC1b cells are either treated with non-targeting siRNA or siSKP2. The cells are subsequently labeled using commercially available antibodies against SKP2 and p27. A) A panel of representative images shows successful knockdown of SKP2 with concurrent elevation in p27. B) The nuclear intensity of both antibody labels are quantified which show a significant reduction in SKP2 ($p<0.0009$) and significant elevation of p27 ($p<0.0001$). Bar plot shows the mean and standard deviation of 3 technical replicates. C-D) The pharmacological control is tested and quantified using the same method. These data show a significant elevation in p27 when treated with STM209405 when compared to vehicle control ($p<0.0001$).

HTS validates the biological activity of a new class of drugs

A total of 98 small molecules were advanced into the imaged-based screening system, schematized in Fig 3.3A. We observed similar Z’ scores for on plate controls from what had been established previously (ranging between 0.5-0.7) for most assay plates, however, we did note that on some assay plates ST029405 lost potency, resulting in decreased activity and lower Z’ values, despite showing clear dose response curves for screening molecules. We later found this to be related to the stability of ST029405, which is intolerant to repetitive freeze thaw cycles. Thus, both manual and statistical criteria are used in the assessment and integration of individual batches into the collective screening results. Of the compounds tested in the cell-based screen, twenty-nine upregulated p27 in the nucleus and displayed dose sensitivity, Fig 3.3B and Table 3.1. Of these, 10 resulted in cells becoming cytostatic and 19 resulted in various degrees of toxicity. Representative hits that elevate p27 and result in either cytostatic or cytotoxicity are shown in Fig 3.3C, where the dose response curve that plateaus near the green line represents a cytostatic interaction and the one that dips below the green line is cytotoxic.

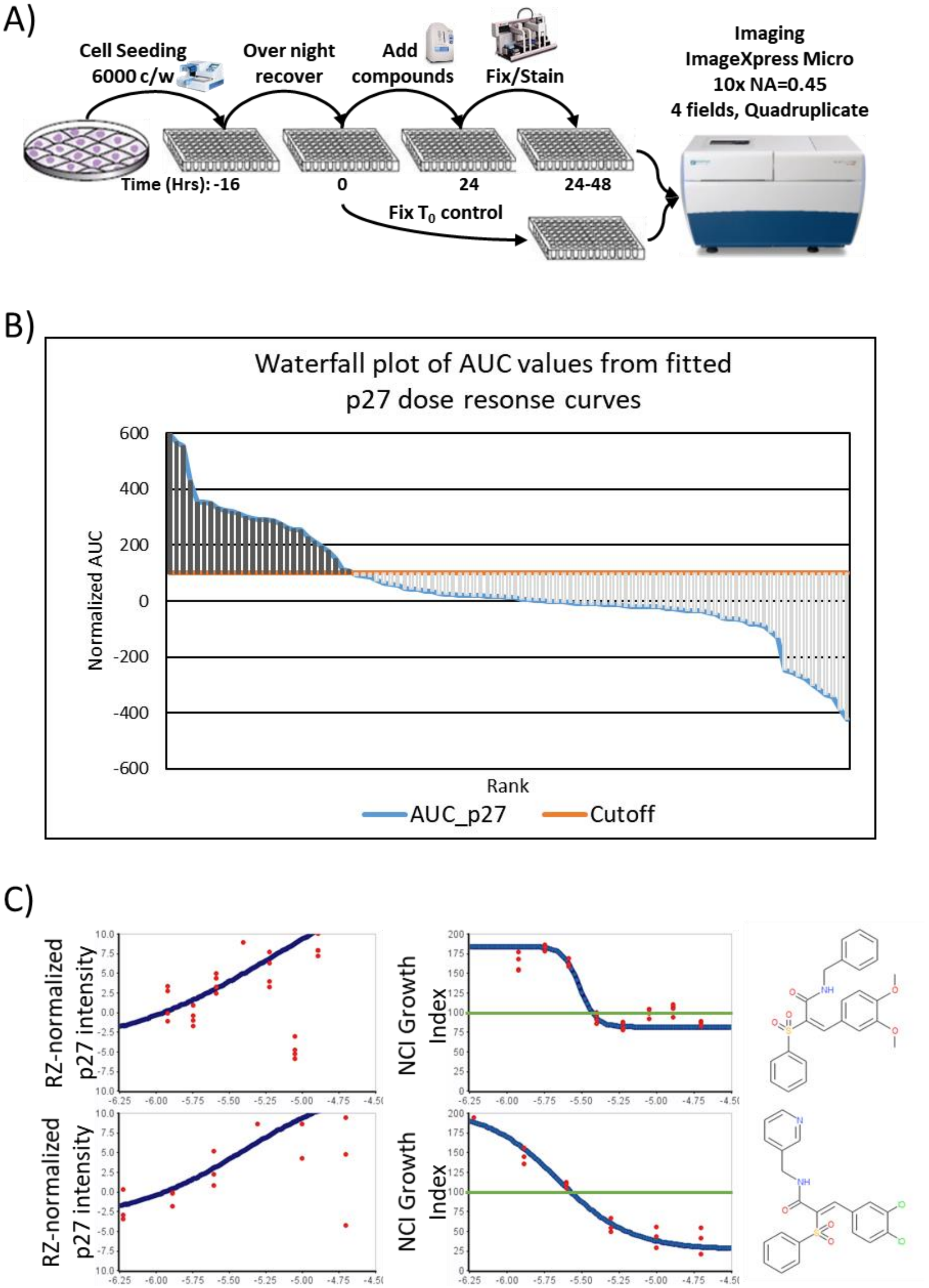
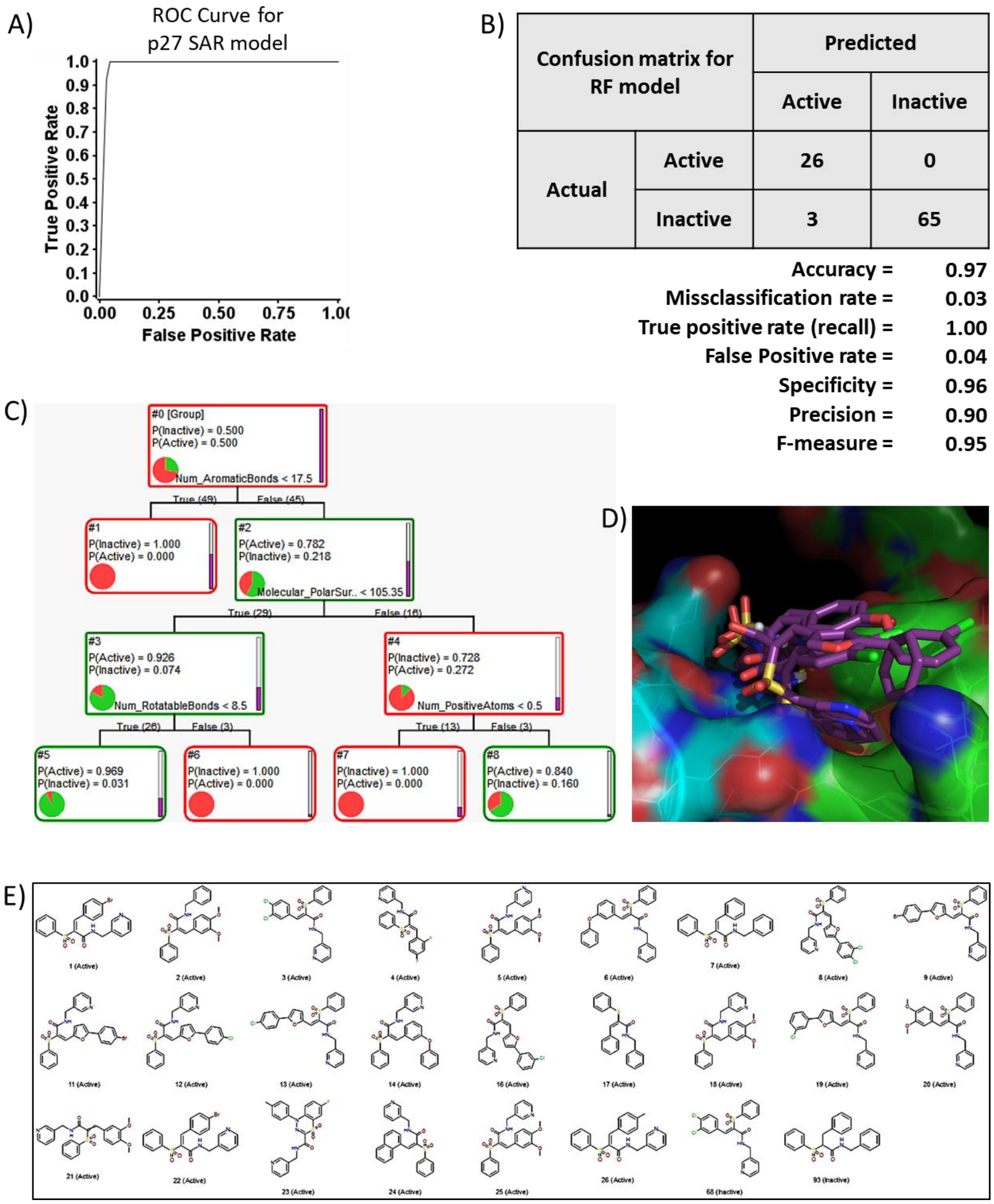


Fig 3.3) Cell based screening results. A) A schematic summary of the cell based screening method. B) Waterfall plot of AUC values from fitted dose response curves using Robust Z-normalized nuclear p27 intensity values of 4 technical replicates treated in 6-point dose response (2 log). Active compounds are represented as dark gray bars. C) RZ-normalized p27 intensity vs concentration (Left), NCI growth index vs concentration (middle) of Hec1b cells treated with the accompanying small molecule (right)

Establishment of a quantitative structure activity (QSAR) relationship using regressive modeling

Multiple methodologies have been proposed to perform QSAR. One such method is through the usage of machine learning models such as random forest (RF), which are uniquely well suited to QSAR due to the ability to automatically identify and weight relevant features and produce interpretable trees that guided decision making.[86,87] Importantly, the chemical space used as an input for this model is fairly narrow, thus the domain of the model is limited to the description of molecular motifs around the core scaffold we identified and not as a generalizable model used in the discovery of new scaffolds. Likewise, a relatively small set of human-interpretable molecular descriptors focusing on quantifying the prevalence of molecular motifs and geometric attributes of the volumetric surface are used versus higher dimensional atomic features; which may result in overfitting, are computationally complex to produce, and difficult to interoperate on the back end. Another important consideration made during model construction is to balance decision weights, which we did by activity class. This is important because the representation of truly active molecules is approximately half of inactive and may bias decisions if not accounted for. In constructing forest-based models it is also important to consider the method used to split nodes. Here, we used Gini index to calculate node purity, however, Shannon entropy was also explored but did not have a significant effect on the model performance.[88] The final consideration made in the construction of the model was to perform a 10-fold cross validation in order to estimate the robustness of the model system. Evaluation of the QSAR model using the mention chemical feature versus activity categories (active vs. inactive) showed robust performance, demonstrated by a high area under the curve of the receiver operator curve (AUC=0.98, 10-fold cross validation AUC = 0.90), Fig 3.4A. Further analysis of the confusion matrix constructed from the optimal model showed that this is a highly accurate model (accuracy=0.97) with a high recall of 1.0, and a low false-positive rate of 0.04, Fig 3.4B. Finally, from interoperating the results of the optimal decision tree, a clear association between molecular activity and the number of aromatic bonds, polar surface area, and number of rotatable bonds was discovered, Fig 3.4C. Rationally, these features make sense when considering the 3D structure of the CKS1:SPK2 pocket. Here, the CKS1:SKP2 interface forms two hydrophobic pockets that are rigidly spaced. Therefore, the chemical selectivity towards systems with more aromatic bonds, which are primarily contained in 3-4 aromatic rings, makes sense as they fit into the clefts, Fig 3.4 D,E. Likewise, the exclusion of highly polar surfaces from these hydrophobic pockets also makes rational sense. Due to the rigidity of the pocket, it also is well rationed that a more structurally rigid molecule facilitates the orientation and strengthens binding of the molecule.



Confusion matrix for RF model		Predicted	
		Active	Inactive
Actual	Active	26	0
	Inactive	3	65

Accuracy =	0.97
Missclassification rate =	0.03
True positive rate (recall) =	1.00
False Positive rate =	0.04
Specificity =	0.96
Precision =	0.90
F-measure =	0.95

Fig 3.4) Structure activity relationship decision tree and model evaluation. A) Receiver operator characteristic (ROC) curve for a categorical random forest model evaluating molecular features associated with activity. B) A confusion matrix of the actual vs predicted drug activity from the optimal forest model with accompanying statistical parameters used in the evaluation of categorical models. C) Tree representation of molecular features associated with drug activity. D) Molecular docking of top 4 lead molecules in the CKS1:SKP2 pocket. E) Chemical structures molecules from bin #3, which is the most highly enriched group representing active drugs. Calculation of chemical features, model construction and evaluation are performed using components in the Chemistry and Statistics libraries in Pipeline Pilot.

Discussion

The *in silico* screening method used here provided a set of molecular scaffolds that putatively interferes with CKS1:SKP2 binding to p27. Importantly, we validated that a subset (~30%) of the scaffolds could functionally restore nuclear p27 in a cell-based system, providing biological relevance to these results. These data also highlight the logistical benefit of performing *in silico* screens to provide a focused library used in lead identification, which is both more time and cost efficient when compared to “shotgun” HTS approaches which identify lead molecules at significantly lower rates. However, the data also argues that a strictly *in silico* approach is not sufficient for lead identification when done in isolation. In this context, we observed that a significant proportion (~70%) of the drugs tested failed to be active in a cell-based screen. While not explicitly proven here, this could be attributed to a failure in the molecular docking model to account for various biological and chemical aspects that are hard to quantify and test *in silico*. It does, however, open the possibility of performing QSAR studies, which require both molecular diversity and varying degrees of activity. Therefore, QSAR increases the biological and translational relevance of this research by providing a chemical basis of activity. In turn, this has the potential to guide additional modifications to the core scaffold to further improve activity or drug-like properties and advance this class of molecules into a translationally relevant drug for diseases where loss or mis-localization of p27 occurs through SKP2-mediated degradation.

CHAPTER IV

ROLES OF PHENOTYPIC SCREENING IN DRUG REPURPOSING

Synopsis

The full spectrum of molecular targets and cellular processes that drugs work through are not necessarily well known. Thus, the identification of novel targets and mechanisms of action for established drugs is an attractive method to supplement the drug development pipeline. An attractive benefit of taking a drug repurposing approach is that pharmacokinetic, pharmacodynamics, and tolerability data are often known for the drugs being tested; which further facilitates rapid translation into the clinic. Phenotypic screening has emerged as an important tool for drug repurposing because it can be tailored to identify drugs which can modulate specific phenotypes, thereby implicating drugs in potential new molecular pathways or cellular processes. Here, we develop a phenotypic screening method aimed at identifying novel modulators of the primary cilia in a disease relevant VHL-null model system. Using this system, we identified bexarotene as a small molecule that protects cells from ciliary resorption in response to acute depletion of *von Hippel Lindau* (VHL) protein. In turn, this research has the potential to increase the translational impact of bexarotene towards the management of ciliary diseases associated with loss of VHL.

Introduction

The phenotype of a biological system is the product of many different cellular processes. Likewise, phenotypic screens have the potential to identify drug activity independently from the established mechanism of action by using an unbiased approach that measures the change in a phenotype. Indeed, this has facilitated an entire field of research where existing or historic drugs find a renewed purpose towards novel disease or activity against a previously unknown target. Likewise, there has been a devoted effort in the field of drug repurposing to identify the profile of molecular targets, signaling pathways, and cellular processes that drugs function through, which is achieved through a number of different methods. At present these include methods that leverage *in silico* screening methods, similar to the one described in the previous chapter, but applied to a library of FDA-approved drugs which are virtually screened against non-canonical binding partners.[89] Other more empiric efforts aimed at describing the profile of molecular targets that drugs work through have also been performed, especially for drugs targeting kinases and GPCR, which show that most of these drugs can modulate the activity of multiple targets.[90,91] Interestingly, clinicians have been non-systematically performing drug repositioning through off-label usage of drugs which have been observed to have a clinically significant impact without explicit scientific testing or knowledge regarding the mechanism.[92] Importantly, multiple databases cataloging the profile of molecular targets that a drug functions through are now available, which further facilitates drug repositioning efforts.[93,94]

We developed an unbiased phenotypic screening method aimed at identifying novel positive regulators of the primary cilium in a deficient model. The primary cilium is a microtubule-based structure that grows and resorbs in response to cues arising from phases of the cell cycle and environmental factors.[95] It is also an important structure that integrates multiple signaling cascades such as Hedgehog, Wnt, PDGF, Calcium signaling, and TGF-β which are reviewed elsewhere.[96] Several critical regulators of cilium disassembly have been identified and include cancer-associated genes such as VHL and AURKA.[97-101] Importantly, the primary cilium is composed of a generic set of proteins making biochemical analysis impossible; likewise, much of the discovery of ciliary proteins has relied on genetic studies where observable changes in the primary cilium are visualized in response to genetic perturbagen. Indeed, many of the early studies which identified critical regulators of the primary cilium relied on high-resolution electron microscopy to characterize these phenotypic defects.[96] However, these methods are cumbersome and difficult to perform on larger chemical screens. Thus, methods which fluorescently label the primary cilium greatly increased throughput by minimizing the resolution required to detect the structure. The most common fluorescently-labeled target for the primary cilium is acetylated tubulin.[102,103] However, acetylated tubulin is also present in a number of other cellular structures, including long-lived microtubules and the midbody, amongst others;[104] thus is not specific and can potentially contribute to the false-positive rate for image analysis methods solely based on labeling acetylated tubulin. Therefore, the addition of a second label against the basal body greatly improves the accuracy of the analysis; however, image analysis pipelines that incorporate this feature have yet to be developed.

Loss or mutations in the *VHL* gene are most commonly associated with VHL disease and clear cell renal cell carcinoma (ccRCC). [105] VHL disease is part of a diverse group of human disorders referred to as ‘ciliopathies’, characterized by the structural loss or functional abnormality of the primary cilium that is directly linked to the cystic kidney phenotype commonly associated with these diseases.[106] Recent advances in the understanding of the role of VHL in multiple molecular pathways, have resulted in the development of targeted therapies, including tyrosine kinase inhibitors (TKIs), monoclonal antibodies, mammalian target of rapamycin (mTOR) inhibitors, and immune checkpoint therapy for the treatment of ccRCC.[107] Although successful, these strategies have plateaued since their inception in large part due to the development of resistance. The molecular mechanisms underlying RCC pathogenesis are only beginning to emerge with the identification of non-canonical targets of VHL,[108-117] linking it to a range of cellular functions including maintenance of mitotic spindle orientation [118,119] and the primary cilium. [120-122] Thus, we developed a ciliopathy model using the well-established retinal pigmented epithelial (RPE1) cell line that was deficient in the ability to properly form the primary cilium by acutely knocking down VHL protein levels using small interfering RNAs. We then used this model to screen a set of FDA approved drugs for the ability to pharmacologically restore the primary cilium.

Methods

Cell culture and drug treatments

Immortalized retinal pigmented epithelial (hTERT RPE1) cells (gift from Dr. Gregory Pazour, University of Massachusetts Medical School, Worcester, MA, USA) were maintained in Dulbecco’s modified Eagle’s medium/F-12 media (Life Technologies, Carlsbad, CA, USA). Human 786-0 *VHL*-deficient RCC cell line was maintained in RPMI-1640 media (Life Technologies, Carlsbad, CA, USA). Cells were maintained in media supplemented with 10% fetal bovine serum (Sigma-Aldrich, St Louis, MO, USA). All human cell lines were short tandem repeats fingerprinted and validated using the

Characterized Cell Line Core Facility (U.T. M.D. Anderson Cancer Center). In addition, all cells used in these studies were routinely tested and confirmed negative for mycoplasma.

Compounds used for the primary screen are solubilized in DMSO and used at a final concentration of 10 µM. Bexarotene,used in secondary validation assays was obtained from Sigma Aldrich (St. Louis, MO, USA) and Selleck Chem (Houston, TX, USA).

Primary screening conditions

hTERT RPE1 cells (7000 cells) were plated in each well of 384 well plates and transfected with siC or siVHL. Next, cells were simultaneously treated with the compound library (Custom Clinical library with 256 compounds) and serum starved for 48 hours to induce primary cilia. Cells were subsequently fixed and stained for the primary cilia using antibodies against pericentrin and acetylated tubulin which mark the basal body and axoneme respectively. Images were obtained at 20X magnification, 4 fields per well, 13 z-sections (0.5 µm steps) using an InCell6000 confocal microscope (GE Healthcare Life Sciences, Pittsburg, PA, USA).

Immunocytochemistry

Immunofluorescence staining to observe primary cilia was performed as published previously [109,120]. Briefly, hTERT RPE-1 cells plated on 384 well plates (primary screen) or glass coverslips (secondary validation) were transfected, starved, and treated for 48 hours to induce cilia formation. Cells were fixed using 4% paraformaldehyde (15 min), permeabilized with 0.5% Triton-X (10 min), followed by blocking in 3.75% bovine serum albumin solution (1 hour). Primary antibodies for acetylated α-tubulin (clone 6-11B-1, 1:5000; Sigma-Aldrich, St. Louis, MO, USA) and pericentrin (1:5000; Abcam, Cambridge, MA, USA) were applied in blocking buffer for 1 hour. AlexaFluor 488 and 546 goat anti-mouse or anti-rabbit secondary antibodies (Life Technologies, Carlsbad, CA, USA) were subsequently applied for another hour. Cells were counterstained using DAPI (1:4000 of 1 mg/ml stock, Thermo Fisher Scientific, Waltham, MA, USA) and visualized using a InCell6000 (GE) at 20X (primary screen) or a Deltavision deconvolution microscope (Applied Precision, Pittsburgh, PA, USA) at 60X magnification (secondary validation assays).

RT-PCR analysis

Reverse transcriptase–PCR analysis was performed as described previously [109,120]. Briefly, mRNA was isolated from cells transfected with siControl/siVHL, and complementary DNA prepared by reverse transcription (Superscript III, Life Technologies, Carlsbad, CA, USA). Gene expression was assessed by real-time quantitative PCR using specific TaqMan probes (Thermo Fisher Scientific, Waltham, MA, USA), and a TaqMan Fast Universal master mix on a Viia7 system (Thermo Fisher Scientific, Waltham, MA, USA). mRNA expression was evaluated for VHL, AURKA, GLI1, PTCH, and PPIA (endogenous control). The following set of conditions were used for each real-time reaction: 95°C for 20 min followed by 40 cycles of 1 sec at 95°C and 20 sec at 60°C. The real-time PCR reactions were all performed in triplicate and were quantified using the $-\Delta\Delta$ cycle threshold (CT) method.

Statistics

To further validate the robustness of the HTS method, we calculated a Z’ value for the percentage of ciliated cells using siC (scrambled control), DMSO treated wells as a positive control, and siVHL treated with sulforaphane (sulphoraphane causes loss of primary cilia, unpublished data) as a negative control. These results showed an average Z’ factor of approximately 0.6, indicating that the assay was robust and suitable for high content screening. Hits were identified using a Z-scoring method which used the mean and standard deviation of the negative control (siC DMSO). Drugs with an average response outside of 3 standard deviations of the negative control were considered as significant. An additional toxicity filter, which was defined by a 20% decrease in cell number from the negative control, was also used.

Results

Development of an image analysis pipeline to detect the primary cilia

As previously mentioned, the primary cilium is composed of the basal body the axomeme. In the development of this image analysis algorithm we take advantage of the structural composition of the primary cilium by labeling both the basal body (pericentrin) and the axoneme (Acetylated-tubulin). Importantly, we found that by using both markers, the false discovery rate of other acetylated-tubulin based structures was minimized. In the preliminary development of this assay, we also explored the utilization of other primary cilia markers, which rely on recruitment and or activation, and are often cell line specific (data not shown). Thus, the method developed here benefits from being able to identify the primary cilium in a cell line and signaling agnostic manner.

Methodologically, we developed a custom image analysis pipeline that reliably identifies the primary cilium at low magnification (20x) in order to facilitate phenotypic screening. The method was developed using the Pipeline Pilot (Biovia, San Diego) developer’s environment. Procedurally, images were first read into the environment and the background was removed by subtracting the mode pixel value shown in blue in the background subtraction panel of Fig 4.1A. Next, nuclei were identified using a watershed segmentation-based method, which identifies and separates neighboring nuclei. The cellular area was then defined using a tessellation of the nuclear mask shown by the green lines in the cellular tessellation box of Fig 4.1 and incorrectly identified objects were filtered out using an expected size range. A cropped example of what ciliated cells look like at 20x magnification is shown in Fig 4.1B. These show the pericentrin, in red, and acetylated tubulin in green, and DAPI in blue. The basal body was identified by performing a regional white top hat on the basal body, labeled by pericentrin, and filtered by co-localization to the cell mask. Further, filtering was performed using a combination of size and circularity filters in order to remove other image artifacts. Next, the axoneme was segmented using a similar regional white top hat for the acetylated α-tubulin (Ac-Tubulin) channel. In order to reduce the false-positive rate for the primary cilia over other microtubule structures, a combination of regional proximity filters, axoneme contrast, and other morphometric features were used and are schematized in Fig 4.1C.

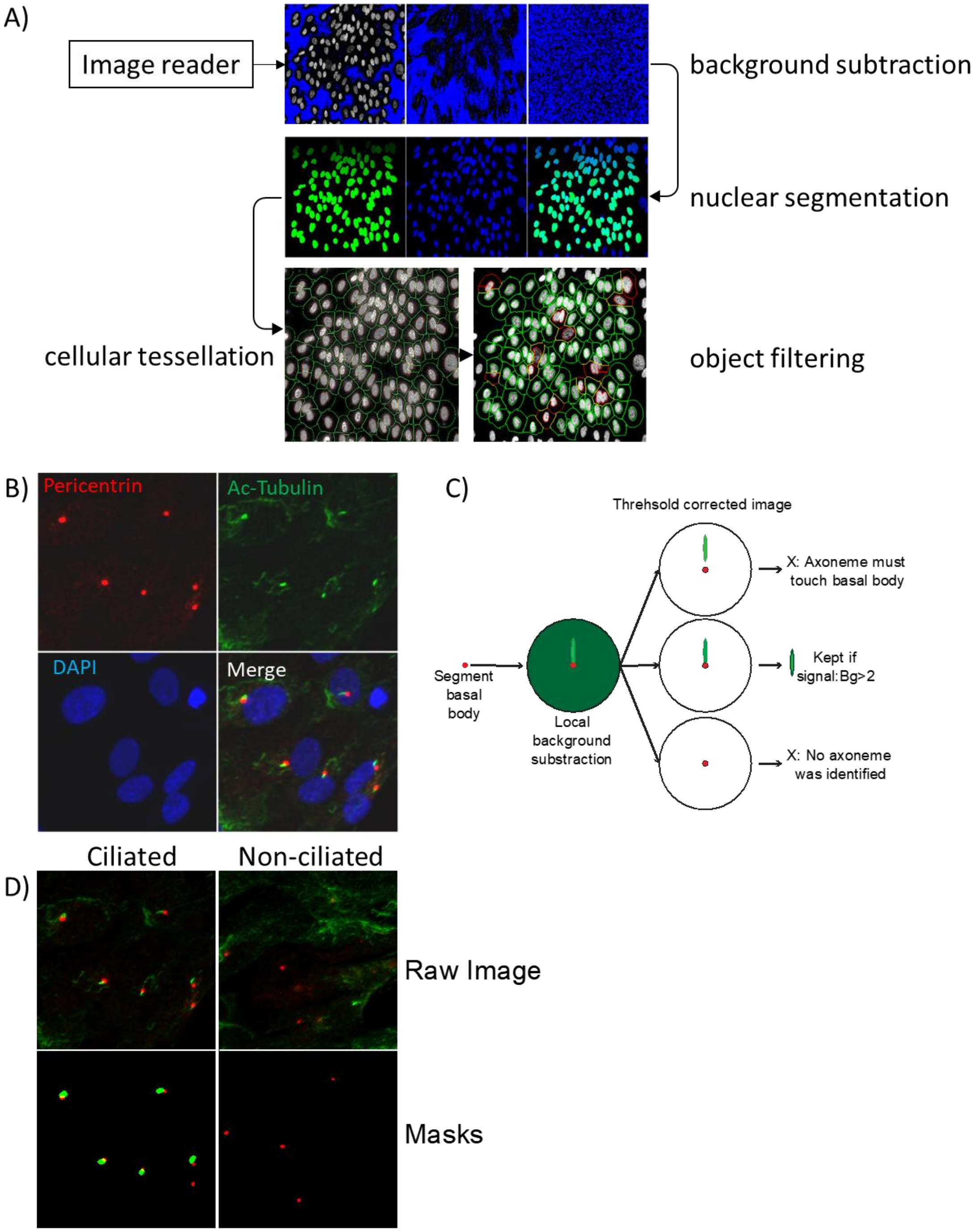


Fig 4.1) Development of an image analysis method for the automated detection of the primary cilium. A) Schematic of the core image analysis pipeline showing, mode background subtraction, nuclear segmentation, cell body estimation, and object filter steps. B) Cropped image of cells acquired at 20x of; pericentrin, red; acetylated(Ac)-tubulin, green; DAPI, blue; and merged. C) schemeatic decision tree of filters used for qualifying the primary cilia using two markers. D) Qualitative image for retrieval of correctly identified primary cilia.

Example raw images of ciliated verse non-ciliated cells and masks showing the correct identification of the primary cilia were used to assess the quality of segmentation, shown in Fig 4.1D. Finally, the percentage of cells containing a primary cilium and the length of the cilium were then used as the primary read-outs of the assay.

Primary HTS identifies small molecules that rescue primary cilia in VHL-deficient cells. We developed a HTS assay to identify small molecules that could restore primary cilia in *VHL*-deficient cells, with the dual goal of identifying novel therapeutic targets, and signaling pathways involved in aberrant ciliogenesis associated with loss of *VHL*. We used an *in vitro* ciliogenesis model, wherein immortalized human RPE1 cells transfected with VHL siRNA (to induce an acute loss of VHL) were stimulated to ciliate by serum withdrawal for 48 hours. We previously showed that this acute loss of VHL (siVHL) resulted in a significant decrease in the ability of RPE1 cells to ciliate compared to control siRNA transfected cells.[109,120] The primary screen was established using 384 well plates with automated processing, thus, this method is amenable to HTS and is schematized in Fig 4.2A. In brief, RPE1 cells were transfected with siC or siVHL, 24 hours after seeding and allowed to recover for an additional 24 hours. At this point, ciliagenesis was induced by the simultaneous withdrawal of serum and treatment with either vehicle (DMSO) or compound (10 µM) for 48 hours. At the end of the incubation period (48 hours), cells were immuno-stained using acetylated α-tubulin (cilia marker) and pericentrin (basal body marker), and imaged at 20x magnification (4 fields/well) using an InCell6000. The efficiency of VHL knockdown was assessed using RT-PCR, which showed a 70-80% decrease in VHL transcript levels, shown in Fig 4.2B, which corresponded to ~41% and ~27% ciliated cells for siC DMSO and siVHL DMSO, respectively, shown in Fig 4.2D, with a coefficient of variance of approximately 10-11% for both controls.[109,120] One limitation of the method developed was that it did not take into account heterogeneity introduced by knockdown efficiency, which potentially under-represents the true response. Using siC as a positive control and sulphoraphane (which we had preliminary evidence demonstrating complete destruction of the primary cilia, unpublished) we obtained a Z' averaging ~0.5, when comparing siC to siVHL we obtained a Z' of ~0.3. Collectively, these indicate the data produced during screening are acceptable quality to categorize drugs into binary classes. We next defined "hit" criteria as: a drug that can significantly restore the percentage of ciliated cells, defined as having a response greater than or equal to two standard deviations from siVHL DMSO, and is non-toxic, which is defined by containing at least 80% of the cell number of the siC DMSO treated wells. A more formal growth analysis was not performed as cells are confluent and serum starved in order to promote ciliation. Of the 256 drugs tested, 100 met the toxicity criteria while only 4 drugs met both. The top hits identified using this method include bexarotene (rentinoid X receptor agonist), BEZ235 (a dual PI3K and mTOR inhibitor), PI-103 (PI3K inhibitor), and PIK90 (PI3K inhibitor),Fig 4.2 C,D and Table 4.1. Importantly, others have shown that the PI3K/mTOR pathway can positively regulate the primary cilium in other disease models, which further confirms the findings of this system; however, bexarotene had not been previously indicated to restore the primary cilium.[102] We therefore followed up on bexarotene in a separate experiment with new compounds and siRNA, and performed on coverslips to validate the findings and obtain representative images, Fig 4.2E.

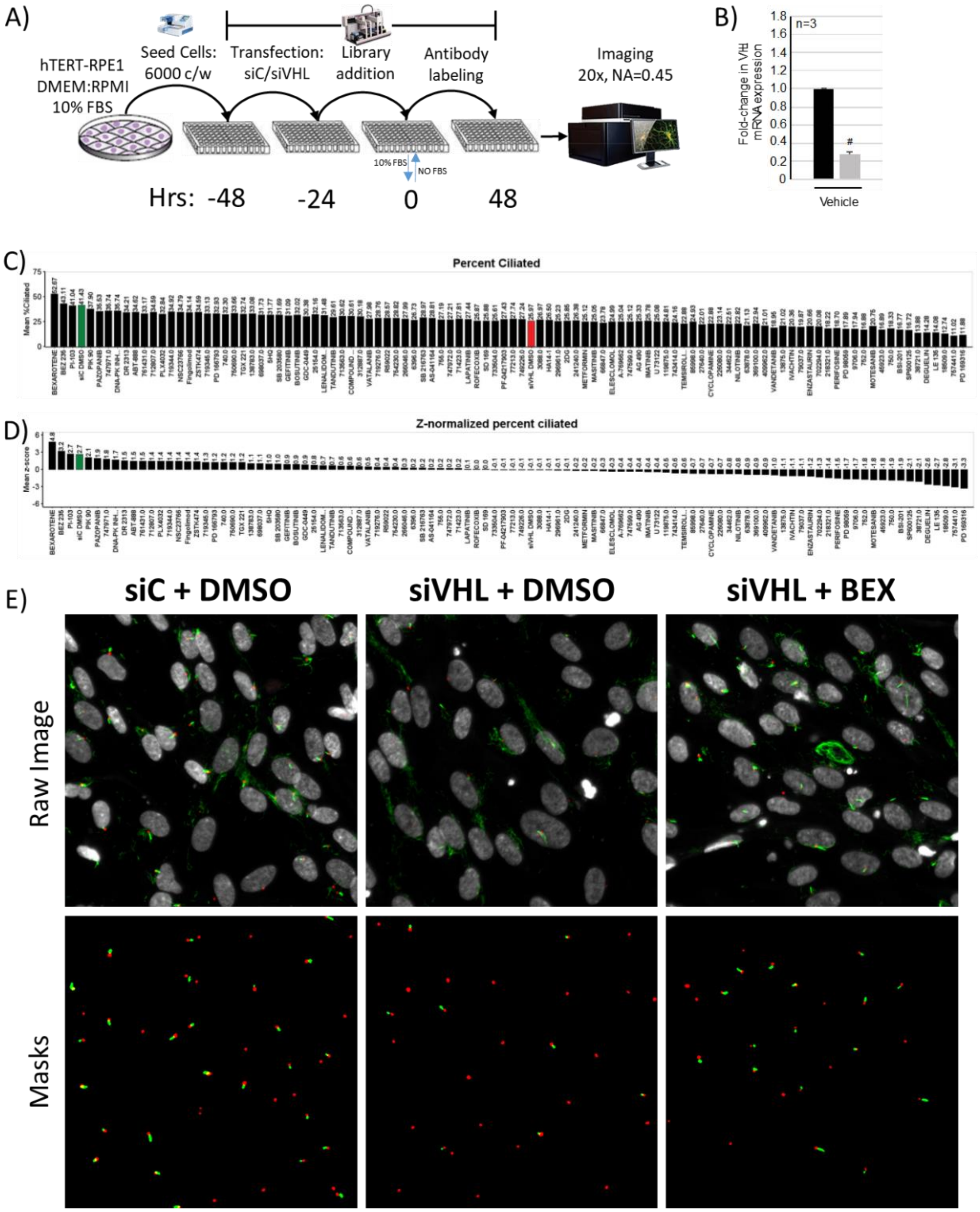


Fig 4.2) Screening schematic and results for ciliary rescue screen A) schematic representation of the screening protocol. B) mRNA knock down efficiency of VHL. C) Bar chart of the percent of ciliated cells by drug treatment. The genetic positive (siC) and negative (siVHL) controls are shown in green and red, respectively. D) Z-transform using the mean and standard deviation of siVHL control, results that are greater than 2 standard deviations are considered significant. E) representative images (cropped from a larger field to visualize cilia better) of siC+DMSO, siVHL+DMSO, and siVHL+Bex (3µM)

Table 4.1) Results from HTS for small molecules that promote ciliation

COMPOUND Name	Cell Count	%Ciliated	%Ciliated (Z-Score)	Result
BEXAROTENE	644	52.67	4.84	HIT
BEZ 235	574.5	43.11	3.21	HIT
PI-103	582.5	41.04	2.72	HIT
PIK 90	602.5	37.90	2.13	HIT
PAZOPANIB	720	35.53	1.95	Inactive
747971	610.5	35.74	1.78	Inactive
DNA-PK INHIBITOR V	531	35.74	1.68	Inactive
DR 2313	580	34.21	1.51	Inactive
ABT-888	526	34.62	1.49	Inactive
761431	587.5	33.17	1.45	Inactive
712807	579	34.59	1.45	Inactive
PLX4032	649.5	32.84	1.44	Inactive
719344	529	34.92	1.44	Inactive
NSC23766	536.5	34.79	1.44	Inactive
Fingolimod	634	34.14	1.41	Inactive
ZSTK474	577.5	34.59	1.39	Inactive
719345	545.5	33.13	1.29	Inactive
PD 166793	567	32.93	1.24	Inactive
740	523	32.30	1.24	Inactive
750690	593	33.66	1.23	Inactive
TGX 221	598.5	32.74	1.23	Inactive
138783	543.5	33.08	1.07	Inactive
698037	632	31.73	1.05	Inactive
5HQ	579	31.77	1.01	Inactive
SB 203580	524.5	31.69	1.01	Inactive
GEFITINIB	635	31.09	0.92	Inactive
BOSUTINIB	869	32.02	0.89	Inactive
GDC-0449	564	30.38	0.86	Inactive
25154	551	32.16	0.80	Inactive
LENALIDOMIDE	542	31.48	0.71	Inactive
TANDUTINIB	762	29.61	0.68	Inactive
713563	531	30.62	0.64	Inactive
COMPOUND 401	633.5	30.62	0.62	Inactive
312887	688	30.18	0.55	Inactive
VATALANIB	655.5	27.98	0.51	Inactive
719276	629	28.76	0.44	Inactive
R59022	734.5	28.57	0.44	Inactive
754230	602.5	28.82	0.38	Inactive

Table 4.1 continued				
COMPOUND Name	Cell Count	%Ciliated	%Ciliated (Z-Score)	Result
266046	694.5	27.99	0.29	Inactive
6396	690	26.73	0.25	Inactive
SB 216763	654	28.97	0.24	Inactive
AS-041164	543.5	28.82	0.24	Inactive
755	644.5	27.19	0.23	Inactive
747972	587.5	27.21	0.23	Inactive
71423	626.5	27.81	0.23	Inactive
LAPATINIB	618	27.44	0.08	Inactive
ROFECOXIB	618.5	25.87	0.05	Inactive
SD 169	762	25.88	0.05	Inactive
733504	757	25.61	-0.07	Inactive
PF-04217903	716.5	27.43	-0.07	Inactive
77213	757	27.74	-0.08	Inactive
749226	652	27.24	-0.08	Inactive
3088	713.5	26.97	-0.09	Inactive
HA14-1	609.5	26.50	-0.10	Inactive
296961	617.5	25.23	-0.11	Inactive
2DG	804	25.85	-0.15	Inactive
241240	618.5	26.39	-0.18	Inactive
METFORMIN	705.5	25.12	-0.24	Inactive
MASITINIB	759	25.06	-0.24	Inactive
66847	714	23.78	-0.31	Inactive
ELESCLOMOL	735	24.99	-0.34	Inactive
A-769662	602.5	25.05	-0.38	Inactive
747599	835	25.12	-0.38	Inactive
AG 490	608.5	25.33	-0.39	Inactive
IMATINIB	699.5	25.78	-0.40	Inactive
U 73122	692.5	25.08	-0.43	Inactive
119875	646	24.81	-0.47	Inactive
743414	661	24.16	-0.61	Inactive
TEMSIROLIMUS	749.5	22.88	-0.65	Inactive
85998	740.5	24.93	-0.67	Inactive
27640	728.5	22.01	-0.70	Inactive
CYCLOPAMINE	718.5	22.88	-0.74	Inactive
226080	524	23.14	-0.74	Inactive
34462	795	22.51	-0.80	Inactive
NILOTINIB	697	22.82	-0.80	Inactive
63878	523	21.13	-0.89	Inactive

Table 4.1 continued

COMPOUND Name	Cell Count	%Ciliated	%Ciliated (Z-Score)	Result
369100	617	22.94	-0.90	Inactive
409962	559	21.01	-0.91	Inactive
VANDETANIB	580	18.95	-0.99	Inactive
13875	829	21.02	-1.10	Inactive
IVACHTIN	616	20.36	-1.10	Inactive
79037	692	19.87	-1.15	Inactive
ENZASTAURIN	552.5	20.66	-1.16	Inactive
702294	865	20.08	-1.29	Inactive
218321	760.5	18.22	-1.44	Inactive
PERIFOSINE	553.5	18.70	-1.53	Inactive
PD 98059	871	17.89	-1.67	Inactive
9706	578	17.94	-1.69	Inactive
752	741.5	16.88	-1.77	Inactive
MOTESANIB	554.5	20.75	-1.81	Inactive
45923	836	16.89	-1.81	Inactive
750	653	18.33	-1.88	Inactive
BSI-201	817	16.77	-1.90	Inactive
SP600125	607	16.72	-2.06	Inactive
38721	644	13.88	-2.13	Inactive
DEGUELIN	598	14.28	-2.62	Inactive
LE 135	608	14.08	-2.70	Inactive
18509	848	12.74	-2.85	Inactive
757441	673.5	11.02	-3.07	Inactive
PD 169316	846	11.88	-3.31	Inactive
ABT-263	81	8.87	-3.25	Toxic
AXITINIB	510.5	38.61	2.14	Toxic
ABT-869	52	27.57	0.14	Toxic
17-AAG		33.81	2.33	Toxic
AZD 0530	77.5	19.15	-1.26	Toxic
BORTEZOMIB		0.76	-5.08	Toxic
BIBF-1120	410.5	45.11	3.24	Toxic
BMS-536924		25.29	-1.20	Toxic
CEDIRANIB	439	20.43	-1.62	Toxic
CANERTINIB		0.47	-5.14	Toxic
CI 1040	349	11.15	-3.02	Toxic
DASATINIB	115.5	49.39	4.46	Toxic
PD 0325901	9	29.17	0.77	Toxic
17-DMAG		1.46	-4.95	Toxic

Table 4.1 continued				
COMPOUND Name	Cell Count	%Ciliated	%Ciliated (Z-Score)	Result
CHIR 258	310	34.72	1.48	Toxic
SFN	210	8.12	-3.55	Toxic
246131	195.5	37.12	1.67	Toxic
82151		5.88	-4.24	Toxic
715055	472	30.65	0.78	Toxic
26980	23.5	5.54	-3.99	Toxic
613327	195.5	38.57	2.24	Toxic
3053	223	2.50	-4.77	Toxic
SFN	419	3.75	-4.46	Toxic
STF-62247	69.5	27.49	-0.13	Toxic
GELDENAMYCIN		8.10	-3.47	Toxic
TRICHOSTATIN A		1.35	-4.96	Toxic
H-89	79.5	36.74	1.82	Toxic
Y-27632	432	46.02	3.29	Toxic
LESTAURTINIB		0.40	-5.16	Toxic
OBATOCLAX	4	0.00		Toxic
AZD 2281	479.5	33.30	1.40	Toxic
SB 431542	503	53.47	5.20	Toxic
PF-2341066		0.42	-5.14	Toxic
NVP AUY922		1.00	-5.04	Toxic
PKC412	8	0.94	-5.05	Toxic
SFN	422	2.71	-4.69	Toxic
122819		13.08	-3.12	Toxic
754143		51.17	2.96	Toxic
737754	385	49.70	4.38	Toxic
718781	435	38.86	2.26	Toxic
SFN	381	2.09	-4.83	Toxic
ROSCOVITINE	272.5	13.32	-2.61	Toxic
SB 202190	357	31.61	1.17	Toxic
MK-2206	62.5	27.54	-0.06	Toxic
PD 153035	298.5	26.21	0.24	Toxic
SORAFENIB		0.60	-5.12	Toxic
SU 11274	518.5	31.61	0.96	Toxic
SUNITINIB	217	8.80	-3.27	Toxic
Brivanib	511	23.51	-0.54	Toxic
STAUROSPORINE		1.10	-5.01	Toxic
PXD101		0.00		Toxic
PCI-24781	156	0.00		Toxic

Table 4.1 continued

COMPOUND Name	Cell Count	%Ciliated	%Ciliated (Z-Score)	Result
KU-55933	344	26.06	0.03	Toxic
SFN	269	2.58	-4.68	Toxic
747973		0.00		Toxic
732517	33	2.99	-4.73	Toxic
279836		13.64	-1.16	Toxic
169780	516	21.44	-1.17	Toxic
762	413.5	24.53	-0.48	Toxic
SFN	322.5	2.45	-4.72	Toxic
GSK1904529A	392.5	27.38	0.09	Toxic
NVP LAQ824		12.06	-3.26	Toxic
JNJ-26481585		27.63	0.91	Toxic
AG014699	423	34.83	1.39	Toxic
AZD1152-HQPA	427	33.75	1.37	Toxic
PLX4720	266.5	26.92	-0.16	Toxic
DMSO	503.5	30.73	0.87	Toxic
SFN	374.5	2.72	-4.63	Toxic
616348	436	33.04	1.08	Toxic
609699	53.5	32.47	0.78	Toxic
122758	413.5	53.36	4.93	Toxic
67574		0.56	-3.23	Toxic
SFN	388.5	3.19	-4.57	Toxic
BMS 204352	67.5	31.19	0.60	Toxic
FK-506	358.5	35.26	1.58	Toxic
CHIR 98014	435.5	26.41	-0.01	Toxic
RAPAMYCIN	396	35.61	1.77	Toxic
BX 912	2	100.00	13.93	Toxic
SARACATINIB	103	34.59	1.70	Toxic
GW 843682X	398	29.77	0.41	Toxic
AZD 6244	290	16.42	-1.93	Toxic
SU 6656	508	32.89	0.94	Toxic
ABT-737		0.00		Toxic
PD 180970	183	44.23	3.23	Toxic
GSK 269962A	69	9.24	-3.17	Toxic
BI 2536		1.04	-5.04	Toxic
BMS-599626	347.5	34.29	1.27	Toxic
BIBW2992	46	5.94	-4.16	Toxic
ERLOTINIB	437.5	35.58	1.53	Toxic
SFN	259	5.90	-4.02	Toxic

Table 4.1 continued				
COMPOUND Name	Cell Count	%Ciliated	%Ciliated (Z-Score)	Result
745750	484	30.19	0.63	Toxic
760766	475	40.54	2.70	Toxic
14229	442	29.01	0.55	Toxic
26271	423.5	16.91	-2.19	Toxic
758252		1.00	-5.03	Toxic
49842		3.56	-4.65	Toxic
SFN	148.5	4.37	-4.39	Toxic
DMSO	429	53.05	5.05	Toxic
NSC 625987	449.5	37.53	1.70	Toxic
GW 441756	512	24.61	-0.55	Toxic
HYPOTHEMYCIN	38.5	21.37	-1.21	Toxic
AEG 3482	514	12.76	-2.71	Toxic
PIK 75	13	2.90	-4.61	Toxic
LBH-589	2	9.36	-3.16	Toxic
MGCD0103		7.52	-4.01	Toxic
MS-275		0.42	-5.15	Toxic
LY 294002	495	36.32	1.53	Toxic
MOTESANIB	464.5	35.86	1.56	Toxic
SFN	175	9.40	-3.49	Toxic

Discussion

We took an unbiased phenotypic screening approach to identify FDA approved drugs that can restore the primary cilia in a cell line with engineered deficiencies in ciliogenesis. The strategy employed in our study differs from previous screens in its use of an unbiased approach, agnostic to the genes involved in ciliogenesis, to identify small molecule compounds that can restore cilia in VHL-deficient cells. Use of a dual-labelling scheme increased the accuracy of image analysis and minimized false positives arising from the inclusion of cytoplasmic acetylated tubulin. Likewise, this method does not rely on using other protein markers which require recruitment, do not label the entirety of the primary cilia, and are often cell line specific. Thus, this method is not confounded by inhibition of recruitment factors and is more generalizable to other cell line models. Using this approach, we identified that multiple drugs targeting the PI3K/mTOR pathway could significantly restore the primary cilia in VHL-null cells, which has been previously observed by others.[102] In addition, we demonstrated the pharmacologic ability of bexarotene to restore the primary cilium in VHL-deficient cells. In the context of drug repositioning, we implicated bexarotene as a potential chemoprotective drug for VHL-related ciliopathies. This has led to molecular biology studies which are currently ongoing in the Walker and Dere labs aimed at clarifying the mechanism of action bexarotene is functioning through. From the literature, bexarotene is a FDA approved synthetic rexinoid with known function in activating RXRs.[123] Ligand binding and activation of the cognate RXR receptors modulates gene expression arising from the homo- or hetero-dimerization of these receptors (with RAR, PPAR, VDR, and others).[124] More recently, Src was identified as a novel off-target protein modulated by bexarotene; where Src kinase activity was inhibited by bexarotene, which was speculated to bind to the ATP binding pocket using *in silico* modeling studies.[125] In addition, bexarotene was also reported to activate p53 by post-translational modifications (*i.e*. phosphorylation on serine 15 (S15)), which regulates p53 downstream targets, such as p21. Thus, there are multiple mechanisms of action that bexarotene can be working through in order to restore the primary cilium. More broadly, retinoids promote differentiation and apoptosis in cells, and this attribute has led to their successful therapeutic application in several cancers, including acute promyelocytic leukemia (APL), Kaposi's sarcoma, squamous cell skin cancer, cutaneous T-cell lymphoma, neuroblastoma, lung cancer, and breast cancer.[124] In contrast to retinoids, /* 8th Generation Intel® Core™ i7-8550U Processortreatment with rexinoids show reduced teratogenicity and toxicity with manageable side-effects, such as hyperlipidemia [126]. The effects of retinoids in RCC treatment were deemed more efficacious when used as part of a combination approach compared to the anti-tumor activity of single agents.[127-130] These data showing limited efficacy of retinoids in RCC, led to recent efforts refocusing on identifying receptor subtypes associated with RCC and their prognostic significance [131-133]. Bexarotene as a prevention strategy (either singly or in combination) has found efficacy in cancers, such lung cancer [134] and oral cavity carcinogenesis [135]. In the case of VHL disease and RCC, it would be valuable to assess the synergistic effects of bexarotene with anti-angiogenic drugs, and with AURKA inhibitors that can rescue ciliation defects associated with VHL-deficiency.[109,120] The identification of bexarotene as a modulator of the cilium now opens new avenues to investigate the merit of rexinoids in ciliopathies.

CHAPTER V

IMPLICATIONS OF TUMOR SPHEROIDS ON HTS PROFILING

Synopsis

Three dimensional multicellular aggregates represent a more physiologically relevant model system that is increasingly being used to acquire drug sensitivity data. Likewise, the applications where 3D models are being utilized range from drug discovery, toxicology, and personalized medicine. Here, we present a high throughput screening method to screen a specialized multicellular aggregate known as spheroids. To validate this method, we empirically compared other fixed endpoint analysis including image-based and biochemical methods. From these data we found that the rate of growth is a highly sensitive method to establish pharmacologic response. This method is further improved by the ability to calculate from label-free images, and is therefore fairly economical and can be run on most standard instruments. Using this method, we generated pharmacological responses in two prototypic models of BRAF and KRAS driven cancers. In parallel, we performed an analogous screen using the same cells grown as monolayers to demonstrate how the 3D spheroid system better recapitulates clinical observation.

Introduction

Historically, 2D monolayers have been the primary model system used for biological validation in drug development and repositioning pipelines. However, many drugs and investigation agents that show positive results *in vitro* using these models fail to translate into *in vivo* efficacy. One explanation for this observation is that 2D monolayers do not recapitulate the physiological microenvironment that is present in a tumor, in turn limiting the predictive value of these models. In contrast, 3D-MCTA represent an *in vitro* model that better mimics the *in vivo* physiological environment of small avascular tumor nodules and micrometastasis. It has been demonstrated that larger spheroids, defined as those with a diameter greater than or equal to 500 µm, contain physiologically similar features of tumor lesions which include; metabolic and proliferative gradients, development of a hypoxic core, and formation of more biologically relevant cell-cell and cell-matrix contacts.[50,136,137] Furthermore, others have demonstrated that these features have the capability to confer secondary chemoresistance and radioresistance, which is also seen in the clinic.[38,136,138] With these observations, it is clear that 3D-MCTA models present a unique *in vitro* system that bridges the gap between *in vitro* 2D monolayers and *in vivo* efficacy. However, it is still technically challenging to screen large amounts of bioactive molecules using these model systems. Therefore, the development of robust HTS assays to implement 3D-MCTA models is paramount.

In order to perform HTS, it is important to first reproducibly generate spheroids in an automation friendly manor. To achieve this, different spheroid generation techniques have been described and are now commercially available which include: hanging drop, magnetic levitation/printing, and ULA U-bottom plates.[47-49,139] Each of these methods has certain benefits and limitations when used with automated platforms and are reviewed elsewhere.[49] For the purpose of this manuscript, we utilize the ULA U-bottom plate method to generate spheroids. This method facilitates the formation of a uniform single spheroid per well, has a transparent bottom with opaque walls, and does not require plate transfers or pre-treatment of cells with exogenous materials, making this the most automation-friendly method to generate large spheroids. Multiple biochemical and phenotypic analysis methods have been used to provide low throughput analysis of spheroids using U-bottom plates. Here, detection of surrogate endpoints of cellular viability such as measuring ATP levels, metabolic potential, and membrane integrity are the most commonly used biochemical endpoints to establish drug sensitivity.[50,56,140,141] Likewise, a number of phenotypic assays have also been proposed and include label-free and fluorescent methods. Label-free methods segment the spheroid being imaged using transmitted light, and often quantify the diameter, which is used to calculate the area or volume.[56,139,142-144] These methods benefit from being cheap to perform and easy to run; however, fixed endpoint analysis of the diameter is potentially confounded by the rate of expansion and the calculation of area or volume relies on the assumption that the spheroid is perfectly round.[64,67,68] Multiplexed image-based fluorescent methods have also been utilized to correct for artifacts in fixed endpoint analysis. These methods rely on adding fluorescent dyes such as Calcein AM (metabolic activity), CellEvent (caspase 3/7 activation), Mitotracker (mitochondrial potential), and Draq7 or Ethidium (membrane permeability), which are often used in combination with a nuclear dye such as Hoechst. However, light scattering and optical penetration through the core of larger spheroids can confound these methods.[46,57] Thus, we proposed to explore methods to increase the robustness of label-free approaches.

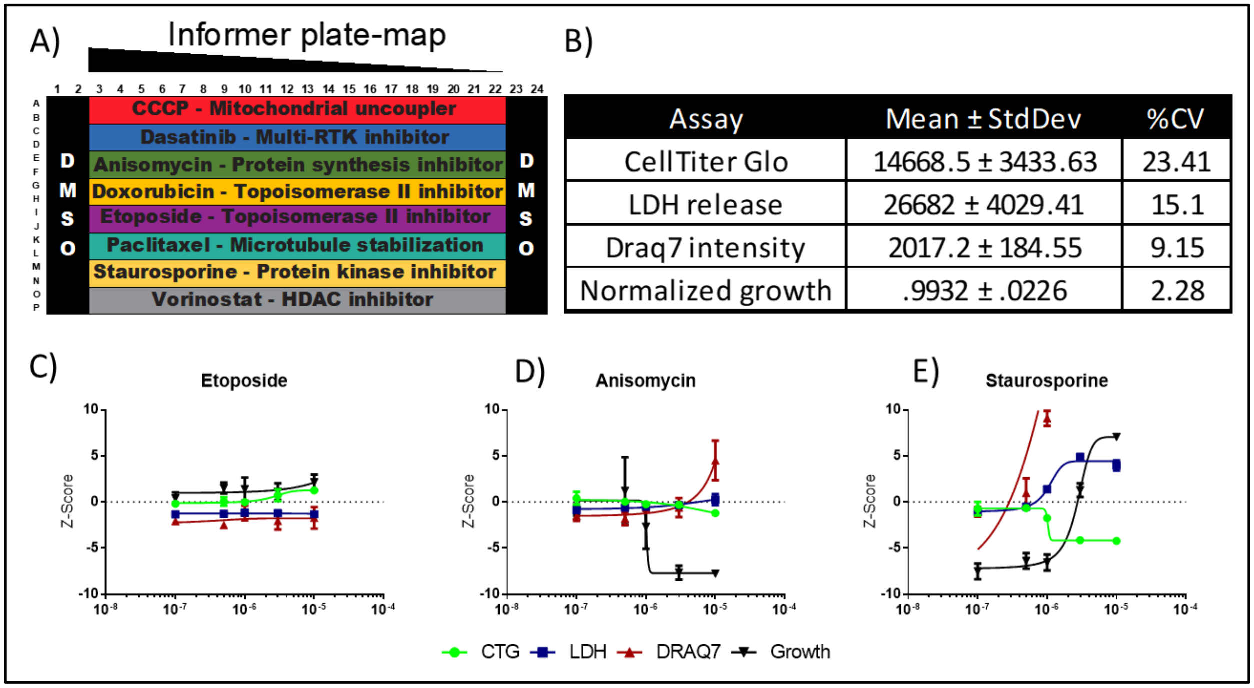
Results

Comparison of HTS methods used to establish viability in spheroids

To establish the utility of using rate of growth as the primary endpoint we first benchmarked it against other routine HTS endpoints. In order to obtain a fair comparison between endpoints a multiplexed analysis of HT29 and HCT116 spheroids was performed. Here, cells were plated into U-bottom plates and treated with a focused library of 8 mechanistically diverse drugs in quadruplicate at 10 doses, Fig 5.1A. The plates were imaged over a seven day period in order to track growth. At the end of the live cell assay, half the media was transferred into a blank 384-well plate and an LDH release assay was performed. CellTiter-Glo (Promega) was then dispensed at a 1:1 ratio into the remaining media of the original plate. Next, we evaluated the coefficient of variance of the negative controls and found that the normalized rate of growth had the lowest variance, Fig 5.1B. Finally, we highlighted three different response patterns to demonstrate the strengths and weaknesses of using only growth as the primary endpoint, Fig 5.1C-E. These data show that etoposide, a topoisomerase II inhibitor, was inactive by all the tested methods, indicating that this is a true negative and that a growth based metric is not over ambitious. Anisomycin, a protein synthesis inhibitor, showed no activity using CTG and LDH release endpoints, while cell growth was significantly inhibited at micromolar ranges and Draq7 accumulation was significantly increased at 10 micromolar. Thus, CTG and LDH release resulted in a false negative for this drug. Staurosporine, a non-specific protein kinase inhibitor, showed a significant decrease in the CTG signal with a concurrent increase in the LDH released into the media and significant accumulation of Draq7, indicating that staurosporine is a highly cytotoxic drug. However, the growth analysis shows a different trend where spheroids are seemingly growing with increasing concentrations of drugs, therefore, resulting in a false negative and suggesting a limitation to this approach. In this case, the potency of staurosporine killed the spheroid very rapidly, leaving behind a carcass which “grows” as the spheroid disaggregates. However, this anomaly can be corrected for by using earlier timepoints and by multiplexing using dyes such as Draq7.

High level comparison of HT29 and HCT116 grown as spheroids and monolayers

A total of 242 unique small molecules were tested in three point dose response (0.1, 1.0, 10 µM) under two and three dimensional growth conditions. As a primary endpoint, the normalized rate of growth as proposed by Hafner et al. was used to normalize cell counts for 2D screens and area was used for the 3D screens.[64] The adjusted rate of growth was then fitted to a 4-parameter logistic regression curve and the GR_{50}, LD_{50}, and normalized AUC (AUC of 1 equals DMSO control) were calculated, Table 5.1. These data show a range of activity across the various cell line and growth formats, Fig 5.2. Next, we defined a low stringency criterion of activity for drugs that can inhibit growth by at least 50 percent within or below the tested ranges. When this criterion was applied to the data, we observed a relatively high hit rate of 48% and 39% for HT29 and HCT116 grown as two-dimensional monolayers, respectively. Interestingly, despite a high hit ratio using the GR_{50}, a low percentage of the drugs tested resulted in significant cell loss as defined by 50 percent loss in cell number, i.e. the LD_{50} was determined to be within or below the tested ranges. When applying this more stringent criterion only 1% of the tested drugs remained for HT29 and none for HCT116. Likewise, using the same low stringency criteria for the 3D screen resulted in 37% and 45% activity in HT29 and HCT116, respectively. However, a significantly larger proportion of the drugs were able to reduce spheroid size by at least 50 percent, which we observed to be 13% for both HT29 and HCT116. In order to compare the results from both cell lines and growth conditions a Venn diagram was constructed using the list of drugs from the less stringent cutoff, Fig 5.3. These data show that the vast majority of compounds that were active remain active across the cell lines and growth conditions, however, certain classes of drugs displayed selectivity towards a cell type and/or growth condition.



Assay	Mean ± StdDev	%CV
CellTiter Glo	14668.5 ± 3433.63	23.41
LDH release	26682 ± 4029.41	15.1
Draq7 intensity	2017.2 ± 184.55	9.15
Normalized growth	.9932 ± .0226	2.28

Fig5.1) Experimental comparison between HTS methods in 3D-spheroids. A) Plate layout of a focused library of 8 mechanistically diverse set of compounds treated in quadruplicate in 10 point dose response. B) Table representing the variance of the negative control for each assay. C-D) Dose response curves for three representative classes of response observed for multiplexed analysis of HT29 grown as spheroids.

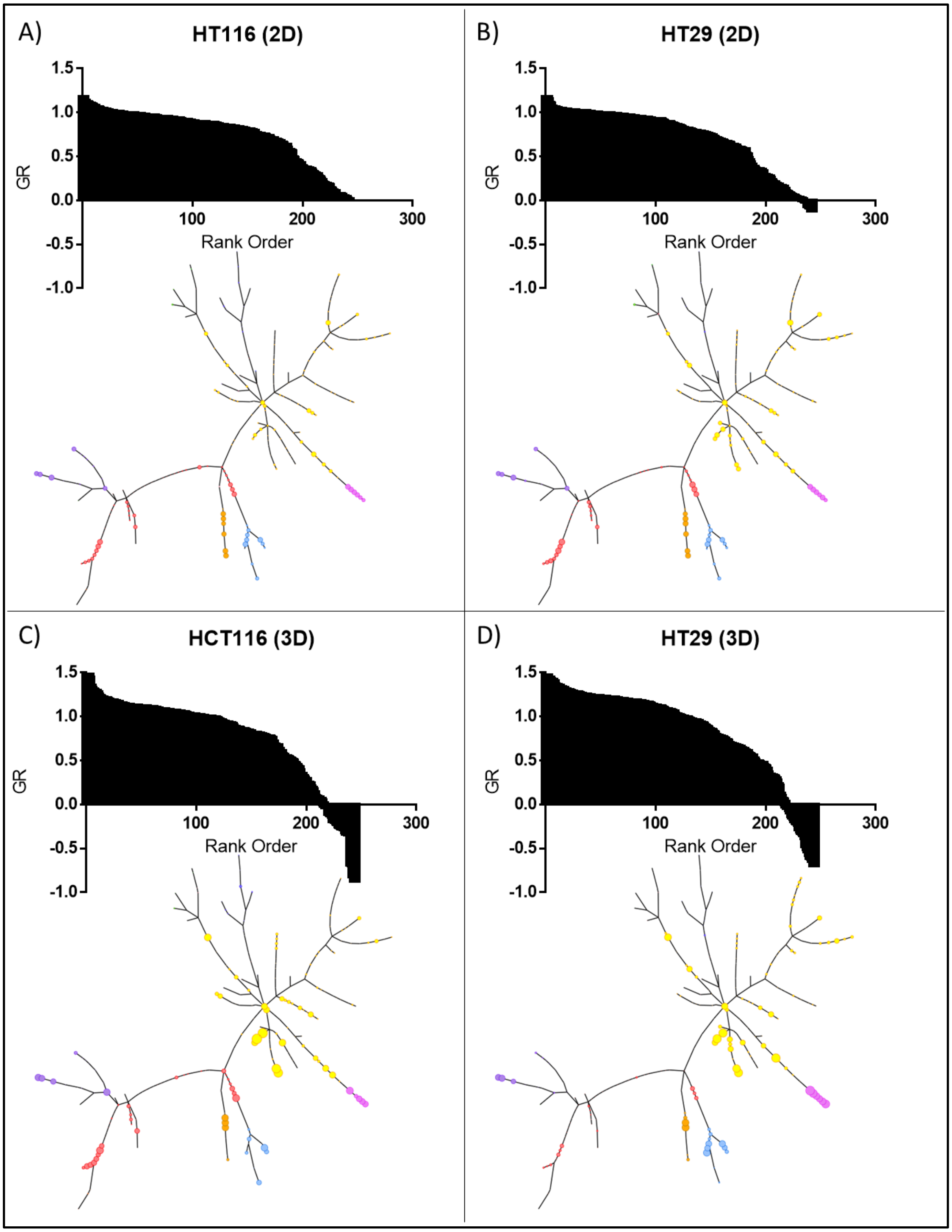


Fig5.2) Screening results and comparison between HT29 and HCT116 grown as monolayers and spheroids. A-D) High throughput screening results from treatment of HCT116 and HT29 cells grown as either a monolayer or spheroid. Data is summarized as a waterfall plot and as an in house data visualization of pharmacologic relatedness

Effects of growth format on the identification of drug-genome interactions

Two of the most common driver mutations in colorectal cancer are BRAF and KRAS, which are considered to be mutually exclusive and occur between 5-10% and 40-50% in the patient population, respectively. Thus, we selected two prototypic cell lines, HT29 and HCT116, which harbor BRAF and KRAS mutations, respectively, to model colorectal cancer. In order to explore the effects of how cells are grown on the identification of drug-genome interactions, we further mined HTS data from HCT116 and HT29 grown either as a spheroid or monolayer. These data show that all three targeted agents against BRAF resulted in inhibition of growth in HT29 but had minimal activity in HCT116, with the exception of PLX4032, which had activity only at the highest dose tested, Fig 5.3A and B. These data are consistent with the results of the 2D screen and provide increased confidence that the model system is behaving properly and the assay is sufficiently sensitive to identify relevant drug-genome interactions. Further comparison of selectively active drugs between HCT116 and HT29 spheroids showed HT29 selectivity to EGFR, p38 MAPK, and HDAC inhibitors while HCT116 were more susceptible to mTOR inhibition and genotoxic stress, demonstrated by an enrichment of drugs that interfere with DNA synthesis, damage DNA, inhibit topoisomerase, and inhibit DNA repair proteins such as ATM and PARP, Fig 5.3D. Importantly, it has been observed that the presence of an activating KRAS mutation is a strong predictor of a poor response to anti-EGFR drugs and is often treated with a cocktail of genotoxic agents.[145] Thus, comparative analysis between HCT116 and HT29 spheroids revealed clinically relevant trends. However, when comparing HCT116 and HT29 grown as monolayers these patterns were not apparent, suggesting a limitation of utilizing a 2D screening format in clinically oriented research. A less consolidated group of non-specific drugs that provide proteotoxic or replicative stress were found to be active in all the tested cell lines and growth formats. However, certain targeted agents such as those that target MEK and general RTK activity with MET inhibitors were being over represented as well.

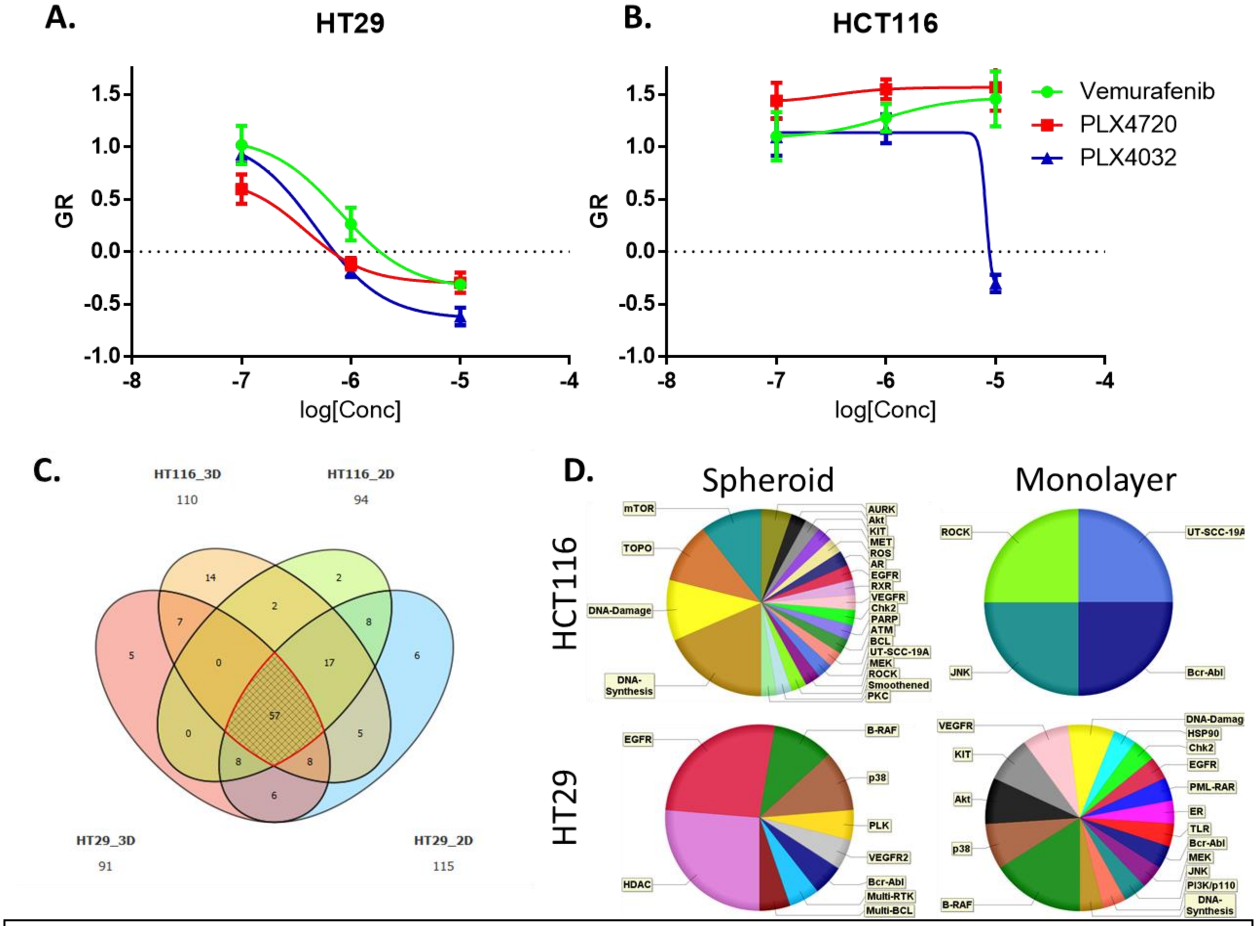


Fig 5.3) Effects of growth format on the identification of drug-genome interactions. A-B) Hafner growth index by concentration curves for three targeted agents against BRAF. C) Venn diagram of active drugs, using the low stringency criteria, for HCT116 and HT29 grown as a spheroid (3D) or monolayer (2D). D) Pie charts of the frequency of drug targets that selectively effect either HCT116 (top row) or HT29 (bottom row) grown as spheroids (left) or monolayers (right).

Materials and methods

Cell line models

HT29 and HCT116 cell lines were acquired from Dr. Bhattacharya, expanded, and subsequently suspended in 10% DMSO/FBS (v/v) and cryopreserved in liquid nitrogen. Cell stocks were then brought up, washed via centrifugation, and allowed to recover before passaging for experiments. Maintained cells were grown in Minimal Essential Medium Eagle (MEME, Thermo) supplemented with 10% Fetal Bovine Serum (FBS, Thermo) and passaged no more than 10 times before being discarded. When plated for spheroid assays the cells were suspended in MEME plus 5% FBS.

Spheroid generation

Single cell suspensions were made by adding trypsin-EDTA (Thermo) to the stock cell culture flask, incubating at 37°C for a sufficient amount of time for most cells to detach, followed by thorough resuspension by pipette. The cell number and viability was then determined by aliquoting 10 uL of media containing cells to 10 uL trypan blue in a disposable counting slide which was analyzed using a TC10 automated cell counter (Bio-Rab). 2500 cells/well were then transferred into barcoded ULA U-bottom plates (Corning, Cat No. 3830). The cells were then incubated for 3 days in a temperature (37°C), CO_2 (5%), and humidity (>95%) controlled cell culture incubator. This allowed for the spontaneous formation of spheroids.

Drug treatment

Experiments used for comparative analysis of different endpoints were treated with a mechanistically diverse array of 8 small molecule probes. Libraries used for primary screening included: 242 FDA approved and phase III investigational drugs acquired from the National Cancer Institute. All drugs used in this manuscript were diluted in

DMSO and arrayed on Echo certified low dead volume plates (LDV, Labcyte). Drugs were transferred from the LDV source plate into assay plates using an Echo liquid handling machine (Labcyte). Wells were treated such that the final concentration of DMSO in media did not exceed 1% (v/v). Wells containing DMSO and media only served as negative controls. For screening, each drug was tested in four replicates and at three concentrations with each replicate and concentration on separate plates.

Image acquisition

Plates were serial imaged using a robotically enabled ImageXpress Micro Confocal (Molecular Devices) equipped with Nikon Pillar Diascopic Illuminator with TE-ELWD Condensor and a 10x/0.45NA Plan Apo lens (Nikon). Additional automation included a Spinnaker Microplate Robot (Thermo) and Cytomat6000 (Thermo). An automated workflow to reimage multiple plates stored in the Cytomate6000 every day for 7 days was developed in Momentum (Thermo). At each time point, one centrally located field was collected per well. To obtain higher quality images, a z-stack covering a 100 µm section taken at 20 µm steps was collected. The resulting stack are projected using the “best focus” method in MetaXpress (Molecular Devices) and used for subsequent image analysis. In order to keep track of successively imaged plates, a systematic file structure was generated which incorporates the date, time, and assay plate barcode.

Image analyses

Batch image analysis was done using a custom script developed using the Imaging package in Pipeline Pilot 9.2 or 2018 (PLP, Biovia). Best focus projected images were first read into PLP and the background was corrected using a z-transformation. Images are then blurred using a Gaussian filter and multiple rounds of morphological elements applied in order to smooth the image. A global threshold was then applied to the smoothed image and used to define the background for a marker directed watershed segmentation. Markers were defined using a peak transformation on the smoothed image. Debris and other artifacts were removed using a combination of morphological and regional filters. Morphological features for remaining objects are then saved in CSV format.

Viability using CellTiter Glo

CellTiter Glo Luminescent Cell Viability Kit was purchased from Promega (Cat No. G7572). The assay was performed as described in the product manual. Luminescence was read with a Tecan M1000 Pro plate reader.

LDH release assay

A commercial fluorescent-based LDH release assay was purchased from Abcam (AB197004). Half of the media on the drug treated plate was transferred to a second LDH assay plate. The LDH regent was then added according to the instruction manual. Fluorescent intensity was measured using a Tecan M1000 Pro plate reader.

Statistical normalization

We developed an automated statistical pipeline which coordinates the plate and time point metadata, raw data from image analysis, performs appropriate statistical normalization, generates fitted curves, and summarizes the results using the Analysis and Statistics package in PLP. Metadata was populated using the file structure produced during image acquisition and merged with the results produced form our custom image analysis. The fold change in spheroid is calculated for each well position over time. The fixed endpoint growth index proposed by Hafner et al. is then applied using the fold change of drug treated cells and the 50 percent trimmed mean of the fold change in DMSO treated wells for each time point. Curves are fitted to a sigmodal curve using a robust least squares fitting method which uses the Normalized MAD and iterative reweighting using the Tukey Biweight method to provide outlier detection. Fitted growth rate curves were then numerically integrated for each drug across a 3 log concentration curve.

Discussion

We have developed a robust screening platform for spheroid models in order to facilitate translational relevant research. The primary endpoint used for this assay is the rate of growth, which we found to be highly sensitive, minimally noisy, and robust across multiple time points. Furthermore, this method is amenable to additional multiplexing which can further increase the power of this method. For multiplexed endpoints, we recommend dyes such as Draq7 or PI to label necrotic cores. Importantly, we also found that addition of these dyes prior to spheroid formation makes for a more consistent labeling. We have also explored the utilization of other fluorescent labels such as retroviral addition of Nuclight (IncuCyte) and soluble-GFP which label the nuclei or cytoplasmic space respectively and facilitate segmentation spheroids and provide evidence of cellular viability.

Using the 3D screening system we developed in addition to routine 2D cell growth assays, we screened two colorectal cancer models from both a KRAS and BRAF driven cancer. From these data, we saw a number of pharmacologic classes of drugs that can effectively inhibit growth or kill in a model and cell line independent manner. Demonstrating that many of the core susceptibilities identified using 2D screening formats are still relevant. These included selective activity of targeted agents against established drivers such as BRAF, which is identified in both screening formats. However, we also found that cells grown as spheroids are more sensitive to MEK inhibitors, which is a clinically relevant target for the management of both KRAS and BRAF driven cancers. We also observed a lack of activity of EGFR inhibitors in HCT116, a KRAS model, and activity in HT29, a BRAF driven model. Importantly, KRAS activation is a strong negative prognostic factor of EGFR therapy and selectivity was not observed in 2D screening format. Thus, the results obtained from screening cancer spheroids closely mimics clinically observed trends which indicates the strength of this method over traditional monolayer systems. Other potentially interesting classes of drugs that were active are HDAC and HSP90 inhibitors, which have been shown to effectively synergize and chemosensitize cells to other classes of drugs. Logically, this presents an opportunity to test these agents in combination with other pharmacologic agents. Indeed, HDAC therapies used in combination with genotoxic agents such as 5-flourouracil are being explored in colorectal cancers by others.[146] Additionally, HDAC inhibitors have also been shown to overcome acute resistance to MEK inhibitors.[147] Thus, a detailed high throughput combinatory screen of HDAC inhibitors with targeted and non-targeted agents may reveal novel combination therapies for the treatment of colorectal cancers.

Table 5.1) Results of 2D and 3D screens for HT29 and HCT116

Compound Name	HCT116 2D	HCT116 3D	HT29 2D	HT29 3D
PCI-24781	0.95	0.45	0.80	-0.31
CABOZANTINIB	0.91	0.78	0.72	0.89
MOTESANIB	0.89	1.32	0.89	1.24
LY 333531	0.86	1.19	0.75	1.16
SUNITINIB	0.84	0.99	0.77	0.61
ZM447439	0.84	0.80	0.84	1.16
MELPHALAN HYDROCHLORIDE	0.84	0.86	0.78	1.28
SORAFENIB	0.81	0.72	0.72	0.62
BIBW2992	0.80	0.85	0.59	0.49
THIOGUANINE	0.79	0.47	0.63	0.70
AXITINIB	0.77	0.68	0.78	0.93
ABT-869	0.77	0.72	0.64	0.82
AFATINIB	0.76	0.82	0.61	0.61
LY 294002	0.76	0.81	0.67	0.85
GW 843682X	0.76	0.51	0.69	0.70
TRAMETINIB	0.76	-0.69	0.14	-0.58
IRINOTECAN HYDROCHLORIDE	0.75	0.21	0.83	1.12
TRIETHYLENEMELAMINE	0.75	0.43	0.64	1.06
PD 180970	0.75	1.23	0.96	1.40
MS-275	0.74	0.18	0.60	0.33
BOSUTINIB	0.73	1.09	0.59	0.58
STF-62247	0.73	0.82	0.63	1.37
VORINOSTAT	0.72	0.61	0.61	0.49
BMS-536924	0.71	0.31	0.36	0.07
PONATINIB	0.71	0.82	0.76	0.45
PF-2341066	0.71	0.80	0.71	0.98
VALRUBICIN	0.69	0.79	0.53	1.30
AZD 6244	0.69	-0.70	0.17	-0.48
MECHLORETHAMINE HYDROCHLORIDE	0.69	0.55	0.60	0.88
JNK INHIBITOR X	0.68	0.68	0.97	1.23
GSK 269962A	0.68	0.55	1.07	1.27
BX 795	0.66	0.29	0.78	0.90
CRIZOTINIB	0.66	0.78	0.66	0.93
CANERTINIB	0.65	0.88	0.68	0.64
IDARUBICIN HYDROCHLORIDE	0.64	0.64	0.67	0.97
ETOPOSIDE	0.63	0.57	0.65	1.18
AZD 0530	0.63	0.95	0.43	0.65
HYPOTHEMYCIN	0.61	-0.23	0.22	-0.17

Table 5.1 continued

Compound Name	HCT116 2D	HCT116 3D	HT29 2D	HT29 3D
BX 912	0.58	0.09	0.69	0.80
AZACITIDINE	0.58	0.95	0.58	0.88
PP121	0.57	0.67	0.35	0.43
MITOMYCIN	0.56	0.82	0.37	0.86
CHLORAMBUCIL	0.49	0.35	0.36	0.64
RDEA119	0.49	-0.82	0.10	-0.61
NVP TAE684	0.46	0.33	0.17	0.13
AZD 7762	0.46	0.47	0.21	0.06
PRALATREXATE	0.45	-0.04	0.29	0.73
SARACATINIB	0.45	0.68	0.10	-0.07
FLOXURIDINE	0.43	0.00	0.16	0.58
BLEOMYCIN SULFATE	0.43	1.21	0.93	1.22
CLADRIBINE	0.40	-0.04	0.39	1.12
AZD1152-HQPA	0.38	-0.22	0.19	0.75
PKC412	0.38	-0.19	0.59	0.27
PI-103	0.38	0.88	0.36	0.97
CYTARABINE HYDROCHLORIDE	0.38	0.02	0.28	0.92
DECITABINE	0.37	0.08	0.38	0.77
MGCD0103	0.37	-0.28	0.41	0.03
CABAZITAXEL	0.36	-0.12	0.35	-0.51
TRICHOSTATIN A	0.36	0.81	0.26	0.37
PD 0325901	0.34	-0.87	0.04	-0.62
TOZASERTIB	0.33	-0.19	0.29	0.41
DR 2313	0.32	0.84	0.66	1.18
LESTAURTINIB	0.31	-0.34	0.34	-0.35
PACLITAXEL	0.30	-0.13	0.06	-0.34
TENIPOSIDE	0.30	0.07	0.28	0.74
PXD101	0.27	0.25	0.46	0.51
EPIRUBICIN HYDROCHLORIDE	0.27	0.15	-0.07	0.33
CLOFARABINE	0.25	-0.26	0.21	1.36
DOXORUBICIN HYDROCHLORIDE	0.23	0.07	-0.09	0.22
MITOXANTRONE	0.21	0.19	-0.02	0.45
OBATOCLAX	0.20	0.39	0.29	0.66
17-AAG	0.20	-0.27	0.00	-0.39
PIK 75	0.20	-0.02	0.19	-0.05
DACTINOMYCIN	0.19	-0.25	0.17	-0.15
NVP LAQ824	0.18	1.28	0.08	-0.13
STAUROSPORINE	0.17	-0.20	-0.02	-0.24

Table 5.1 continued

Compound Name	HCT116 2D	HCT116 3D	HT29 2D	HT29 3D
PLICAMYCIN	0.14	-0.15	0.19	-0.20
VINORELBINE TARTRATE	0.13	-0.11	0.24	-0.22
ELESCLOMOL	0.12	0.98	0.03	1.02
DASATINIB	0.11	0.94	-0.08	0.49
TOPOTECAN HYDROCHLORIDE	0.11	-0.34	0.06	0.32
17-DMAG	0.08	-0.29	-0.02	-0.33
VINBLASTINE SULFATE	0.08	0.30	0.07	-0.28
GEMCITABINE HYDROCHLORIDE	0.08	-0.34	0.11	0.56
BI 2536	0.07	0.05	0.03	-0.65
JNJ-26481585	0.07	1.44	0.08	-0.21
NVP AUY922	0.07	-0.28	0.16	0.34
DOCETAXEL	0.05	1.07	0.03	-0.53
LBH-589	0.05	1.58	0.05	-0.09
VINCRISTINE SULFATE	0.04	-0.33	0.04	-0.70
BORTEZOMIB	0.04	0.54	0.00	0.48
CARFILZOMIB	0.03	1.07	0.03	0.78
DAUNORUBICIN HYDROCHLORIDE	0.02	0.08	-0.12	0.30
OMACETAXINE MEPESUCCINATE	0.02	-0.04	0.02	0.10
LE 135	1.18	0.91	1.15	1.34
AS 252424	1.14	0.86	1.02	1.08
RALOXIFENE	1.12	0.93	0.98	1.21
AMINOLEVULINIC ACID HYDROCHLORIDE	1.12	1.00	1.00	1.19
URACIL MUSTARD	1.11	1.03	1.04	1.10
METHOTREXATE	1.10	1.23	1.04	1.42
PLERIXAFOR	1.10	1.12	1.07	1.15
ESTRAMUSTINE PHOSPHATE SODIUM	1.10	1.15	1.02	1.35
ENZALUTAMIDE	1.09	0.82	1.05	1.13
NELARABINE	1.08	1.12	0.99	1.22
PEMETREXED	1.08	0.99	0.93	1.02
CISPLATIN	1.07	1.13	0.99	1.17
CELECOXIB	1.06	1.16	0.88	1.24
ALTRETAMINE	1.06	1.12	1.03	1.22
LETROZOLE	1.05	1.06	1.03	1.20
AMIFOSTINE	1.05	1.12	0.97	1.17
IFOSFAMIDE	1.05	1.10	1.02	1.23
CYCLOPHOSPHAMIDE	1.04	1.12	1.01	1.26
ZOLEDRONIC ACID	1.04	1.18	1.03	1.29
REGORAFENIB	1.04	0.78	0.94	0.86

Table 5.1 continued				
Compound Name	HCT116 2D	HCT116 3D	HT29 2D	HT29 3D
MK-2206	1.04	0.61	0.70	0.79
ERLOTINIB	1.03	1.08	0.93	0.47
U 73122	1.03	1.14	0.96	1.26
SIROLIMUS	1.02	0.54	0.79	0.67
BENDAMUSTINE HYDROCHLORIDE	1.02	1.07	0.97	1.24
MERCAPTOPURINE	1.02	0.77	0.81	0.97
LOMUSTINE	1.02	1.13	0.99	1.25
METFORMIN	1.02	1.01	1.03	1.19
LENALIDOMIDE	1.01	0.93	1.05	0.96
CARBOPLATIN	1.01	1.09	0.98	1.23
PD 153035	1.01	1.00	0.97	0.57
BMS-599626	1.01	0.99	0.89	0.56
PROCARBAZINE HYDROCHLORIDE	1.01	1.12	0.99	1.31
AICAR	1.01	1.20	0.95	1.22
PD 169316	1.01	1.08	1.02	1.08
CAPECITABINE	1.00	1.13	1.02	1.33
PENTOSTATIN	1.00	1.22	1.01	1.45
PIK 90	1.00	0.97	0.99	1.02
PLX4032	1.00	0.85	0.53	-0.05
VX-702	1.00	1.12	1.02	1.00
ABIRATERONE	1.00	1.15	1.05	1.24
SB 431542	1.00	1.13	0.96	1.24
FINGOLIMOD	1.00	1.09	1.00	1.37
PF-04217903	0.99	0.92	0.98	0.99
THALIDOMIDE	0.99	1.16	1.06	1.25
PAZOPANIB	0.99	1.18	1.11	1.43
VISMODEGIB	0.99	1.12	1.03	1.25
PERIFOSINE	0.99	0.35	0.82	0.76
BSI-201	0.99	1.04	1.02	1.08
FLUDARABINE PHOSPHATE	0.98	0.94	0.86	1.24
PIPOBROMAN	0.98	1.02	1.01	1.23
METHOXSALEN	0.98	1.22	1.00	1.20
PD 98059	0.98	0.88	0.98	1.18
Y-27632	0.98	1.13	1.18	1.30
GSK1904529A	0.98	1.10	0.94	1.31
NU 1025	0.98	1.07	0.95	1.19
KU-55933	0.97	0.78	0.98	1.04
VATALANIB	0.97	1.21	0.94	1.22

Table 5.1 continued				
Compound Name	HCT116 2D	HCT116 3D	HT29 2D	HT29 3D
AG 490	0.97	1.07	1.03	1.08
EVEROLIMUS	0.97	0.58	0.80	0.68
BEZ 235	0.97	1.00	0.85	1.03
IMIQUIMOD	0.97	1.03	0.83	1.34
SD 208	0.97	1.19	0.99	1.18
A-769662	0.97	1.10	0.93	1.08
GEFITINIB	0.97	1.10	0.88	0.78
AZD 2281	0.96	0.88	1.00	1.20
MEGESTROL ACETATE	0.96	1.14	1.06	1.26
EXEMESTANE	0.96	1.02	0.90	1.22
FULVESTRANT	0.96	1.01	0.98	1.24
2DG	0.96	1.05	1.02	1.10
VEMURAFENIB	0.95	1.30	0.71	0.34
PAZOPANIB HYDROCHLORIDE	0.95	1.20	0.98	1.20
TEMOZOLOMIDE	0.95	1.19	1.00	1.15
ARSENIC TRIOXIDE	0.95	1.22	0.68	1.23
MITOTANE	0.95	0.96	0.97	0.79
BUSULFAN	0.95	1.11	0.97	1.28
NSC 625987	0.95	1.17	0.93	1.28
TEMSIROLIMUS	0.95	0.53	0.79	0.67
TANDUTINIB	0.95	1.15	0.99	1.21
SD 169	0.95	1.06	1.04	1.16
ANASTROZOLE	0.94	1.10	0.96	1.23
SP600125	0.94	1.14	0.97	1.25
ERLOTINIB HYDROCHLORIDE	0.94	1.01	0.85	0.52
STREPTOZOCIN	0.94	1.13	1.02	1.24
BMS 204352	0.94	1.04	1.00	1.13
COMPOUND C	0.94	1.17	0.98	1.42
IMATINIB	0.94	1.12	0.76	1.24
PD 166793	0.93	1.14	1.03	1.28
CEDIRANIB	0.92	0.84	0.89	0.96
FLUOROURACIL	0.92	0.80	0.86	0.99
FK-506	0.92	1.07	1.04	1.19
SB 203580	0.92	1.35	0.79	0.83
NILOTINIB	0.92	1.33	1.03	1.30
DACARBAZINE	0.92	0.43	0.90	0.68
ABT-263	0.91	1.00	0.95	0.85
CYCLOPAMINE	0.91	0.88	1.00	0.96

Table 5.1 continued				
Compound Name	HCT116 2D	HCT116 3D	HT29 2D	HT29 3D
PLX4720	0.91	1.54	0.36	-0.03
ROSCOVITINE	0.90	1.11	0.95	1.20
LAPATINIB	0.90	0.97	0.79	0.48
TGX 221	0.90	0.81	0.90	1.03
ZSTK474	0.90	1.04	0.86	1.06
RAPAMYCIN	0.90	0.52	0.71	0.64
CARMUSTINE	0.90	1.02	0.82	0.99
ALLOPURINOL	0.90	1.14	1.00	1.22
GDC-0449	0.89	1.19	1.01	1.22
8-DODECYLSULFANYL-1,3-DIMETHYL-7H-PURINE-2,6-DIONE	0.89	1.29	0.95	1.36
H-89	0.89	1.11	1.01	1.19
TRETINOIN	0.89	1.02	0.99	1.18
DEXRAZOXANE	0.89	1.08	0.93	1.01
AS-604850	0.89	1.06	1.03	1.30
ROFECOXIB	0.89	1.09	0.95	1.20
AS-041164	0.89	1.07	0.94	1.13
ENZASTAURIN	0.89	0.93	0.94	0.99
COMPOUND 401	0.89	1.06	0.96	1.19
R59022	0.89	1.05	0.93	1.28
SB 218078	0.89	0.94	0.93	0.94
SB 202190	0.88	1.29	0.77	0.77
CI 1040	0.88	-0.32	0.49	0.00
CHIR 258	0.87	0.77	0.77	0.70
TRICIRIBINE	0.87	0.50	0.59	1.44
HA14-1	0.87	1.08	0.99	1.20
ABT-737	0.86	1.03	0.91	0.93
NSC23766	0.86	1.13	0.96	1.12
MASITINIB	0.86	0.80	0.75	1.23
BIBF-1120	0.86	1.03	0.92	0.95
CHK2 INHIBITOR II	0.86	0.83	0.79	1.07
BEXAROTENE	0.85	0.55	1.11	0.97
IVACHTIN	0.85	1.03	1.01	1.22
TAMOXIFEN CITRATE	0.85	1.02	0.78	0.73
DABRAFENIB MESYLATE	0.85	1.25	0.28	0.33
VANDETANIB	0.85	1.09	0.75	0.69
HYDROXYUREA	0.85	1.02	0.90	1.12
THIOTEPA	0.84	0.85	0.82	1.18
DNA-PK INHIBITOR V	0.84	1.01	1.02	1.14

Table 5.1 continued

Compound Name	HCT116 2D	HCT116 3D	HT29 2D	HT29 3D
CHIR 99021	0.84	1.17	0.82	1.29
SB 216763	0.83	1.00	1.04	1.16
GDC 0941	0.83	0.24	0.63	0.24
OXALIPLATIN	0.83	0.87	0.82	1.09
SU 6656	0.82	0.26	1.00	0.49
GELDENAMYCIN	0.82	0.94	0.81	0.68
GW 441756	0.81	1.12	1.01	1.48
BRIVANIB	0.81	1.09	1.01	1.13
AG014699	0.81	0.86	0.93	1.19
SU 11274	0.81	0.87	1.05	1.21
ABT-888	0.81	1.07	0.94	1.17
CHIR 98014	0.79	1.01	0.87	1.12
DEGUELIN	0.75	0.10	0.85	0.66
AEG 3482	0.69	1.03	0.67	1.33

CHAPTER VI

CURRENT AND FUTURE WORK

Development of visualization tools for HTS/HCA data

Data visualization is an important final step in the HTS/HCA analytical workflow that summarizes the collective results of a screen into a discrete graphic. At present, the most prevalent ways to visualize HTS data are in the form of tabular reports or waterfall plots. However, these methods are limited in that they often times fail to contextualize the data outside of identifying the most active drug in a dataset. Others have aimed to provide such context by clustering drug sensitivity data across multiple cell lines.[33,148] These type of analysis can provide a two-way inference of the molecular mechanism of action (MoA) of the drug, and provide a low resolution genetic profile based on drug-driver susceptibilities through the identification of regional hotspots.[33] Therefore, two-way clustering is a good method for identifying novel MoA, crude characterization of the genomic characteristics of a cell line, and clustering pharmacologic patterns across an unannotated array of cells. However, this method is limited in that it relies on testing panels of drugs across an array of cell lines with different underlying molecular susceptibility to properly bin the mode of action, which is not always feasible. Fortunately, public databases which archive pharmacologic profiles are available and can be used to provide such context if the same endpoints and methods of analysis are used.[64] Yet another approach relies on compiling the target and MoA of established drugs into organized networks.[149,150] While this method does generate an aesthetically pleasing and informative visualization, there are limitations in the underlying method including that it is proprietary, remains unpublished, and does not appear to have the ability to dynamically self-compile when new libraries are tested. Thus, we aimed to generate a robust visualization method using open source tools that provides context to screening data. To achieve this, we first had to generate in-house databases of molecular targets, drug classifications, and modes of action. This was done under two frameworks, one database that is manually curated and the other which is automatically populated. The benefits of utilizing the manually curated database are that it provides better congression of related drugs; however, this database is tedious to maintain and potentially subject to bias. Conversely, the database that automatically compiles vectors from multiple online sources is less subject to bias, easier to maintain, and includes information about drugs that active against multiple targets; however, there is more noise in the descriptive vector. With the database structure developed, we then generated a method to convert descriptive features archived in the database into a self-compiling network, which can subsequently be used to extract biological information from screening campaigns regardless of library size or content.

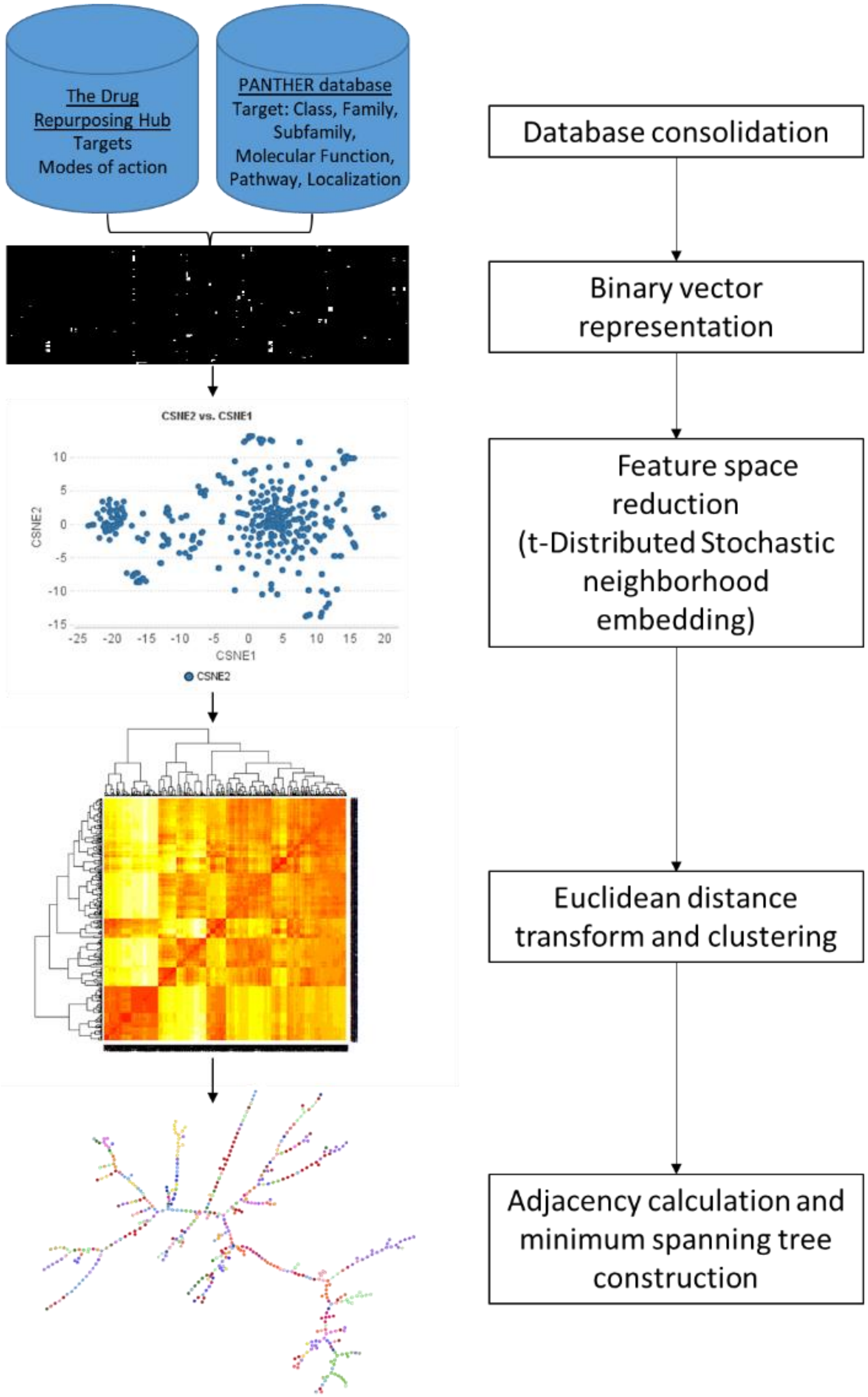


Fig 6.1) Comprehensive workflow for generating pharmacologic trees. Data base sources are used to build drug annotation vectors which are summarized in a binary vector format. T-SNE is then used to reduce the dimensionality of the data while preserving the global and local structures within the data. A pairwise distance map is then constructed to identify nearest neighbors and compiled into a minimum spanning tree.

Materials and methods

Constructing the drug annotation database

The initial drug annotation database was manually curated and cataloged the class of the drug, canonical target, molecular signaling pathway, and the functional implication. Raw data was obtained either from vendor sources or databases including the Drug Repurposing Hub, Selleck, ChemSpider, PubChem, and PubMed. Summarized terms for the molecular signaling cascade and functional implications were derived from inferring information about these drugs through primary literature. Thus, this initial implementation, which we will refer to as method 1, is limited in that it relies on manual inference of drug MoA and functional implication which is potentially subjective and highly labor intensive. However, we believe that the ability of this method to self-compile HTS data into a meaningful visualization makes this a valid tool despite this limitation.

Automated methods for compiling drug annotations

In order to remove the biases associated with manually inferring data stored in discordant data silos, an automated method to build a binary description vector of drugs was generated and referred to as method 2. This method pulls the target information from the Drug Bank database, providing target information, which is then merged with the PANTHER database, providing signaling pathways, molecular processes, and localization of the target proteins.[5,151-153] These data are then stored in a tabular format where each row represents a single drug and each column represents a term describing targets, molecular modes of action, signaling pathways, molecular processes, and localization.

Generating a minimum spanning tree of pharmacologic MoA

In order to convert the descriptive vectors into an adjacency matrix an intermediate distance matrix must first be obtained. For method 1, this is done simply by calculating the gower distance using the daisy function in R. However, for method 2 direct calculation of gower distances fails to recapitulate local and global structures of the data due to the high dimensionality of the drug description vector. Therefore, method 2 requires that the dimensionality of first is reduced before identifying nearest neighbors. While many different methods of data reduction are available we found that non-linear data reduction methods, such as t-distributed stochastic neighborhood (t-sne) embedding, were better than linear data reduction techniques, such as principle component analysis (PCA) or multidimensional scaling (MDS), due to the ability of t-sne to maintain both global and regional interactions on large datasets. A simple Euclidean distance matrix can then be used to identify nearest neighbors and construct an adjacency matrix using the graph_from_adjacency_matrix function in the R igrpah library. Finally, a network of pharmacologic relations was constructed by linking nearest neighbors into a minimum spanning tree using the mst function in the R igraph library, Fig 6.1. In order to make this visualization tool interactive, the igraph network object was then converted into an interactive visNetwork object using the visNetwork package. The code and tabular drug annotation databases used to generate trees are provided in supplemental material 6.1.

Preliminary results

Generalizability of VisPharma-Network visualization tool

In order to increase the generalizability of this application we aimed to make a flexible algorithm used to visualize pharmacologic relations of experimental drugs tested in a HTS campaign, regardless of the library size and profile. In order to achieve this, targeting and MoA information must be stored in a format that allows the algorithm to self assemble a network of related drugs. To demonstrate this capability, we simulated libraries of various size ranging from 250 to 2000 drugs. When annotated, these data clearly show that despite being different sizes the algorithm was able to spatially isolate and connect related drugs, Fig 6.2. Importantly, a reproducible pharmacologic network can be obtained given the same input of drugs. However, we found certain limitations in this method including the ability to connect drugs that are distantly related which may lead one to over interpret the importance of a given region. We also found that larger networks, loosely defined as those with more than 1500 nodes, become visually very compact and difficult to interpret, which is a general limitation of network-based analysis.

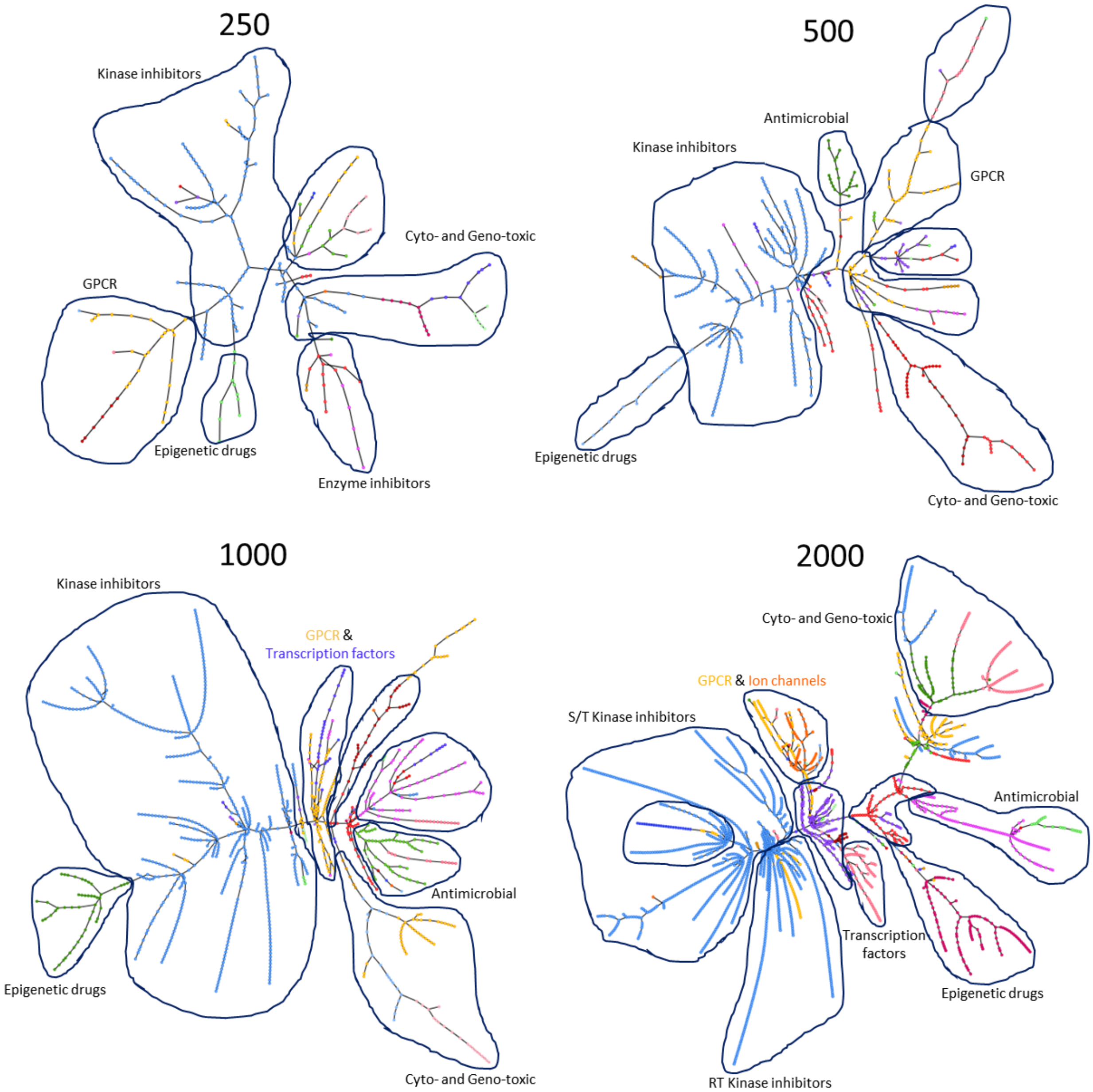


Fig 6.2) Pharmacologic trees from libraries of different size and content. In order to demonstrate the dynamic scalability and generalizability of this method to visually congregate similar drugs, VisPharmaNetwork graphics are constructed using a simulated library containing 250, 500, 1000, and 2000 drugs. Regions displayed in the above graphic are manually annotated for their content after using the interactive reviewing tools. These data show the trade-off in molecular resolution and visual interpretability. All graphics are made using R Studios, igraph, and visNetwork libraries.

Comparing pharmacologic profiles of sarcoma cell lines using various resolutions of data visualization

Multiple levels of information can be extracted from a HTS campaign. To demonstrate how various methods of data visualization can influence the interpretation of data, we present a case study of a HTS campaign performed on a panel of 25 sarcoma cell lines screened against 1163 drugs. The cell lines used in this study represent a diverse panel of sarcomas that were isolated from soft tissue and bone tumors. Likewise, one would expect the molecular susceptibilities of these tumors to be unique and result in a signature drug susceptibility profile. The rate of growth for each condition was calculated using the fixed-endpoint method proposed by Hafner et al. and fitted against the log of three concentrations (0.1, 1.0, 10.0 uM). Finally, the AUC of the fitted curve was used to make comparisons across the different cell lines. Principle component analysis was then used to provide an abstract preview of the overall pattern of susceptibility between cell line models. Here, two clusters can be seen which are enriched for either bone or soft-tissue linages, Fig 6.3A. These data therefore provide

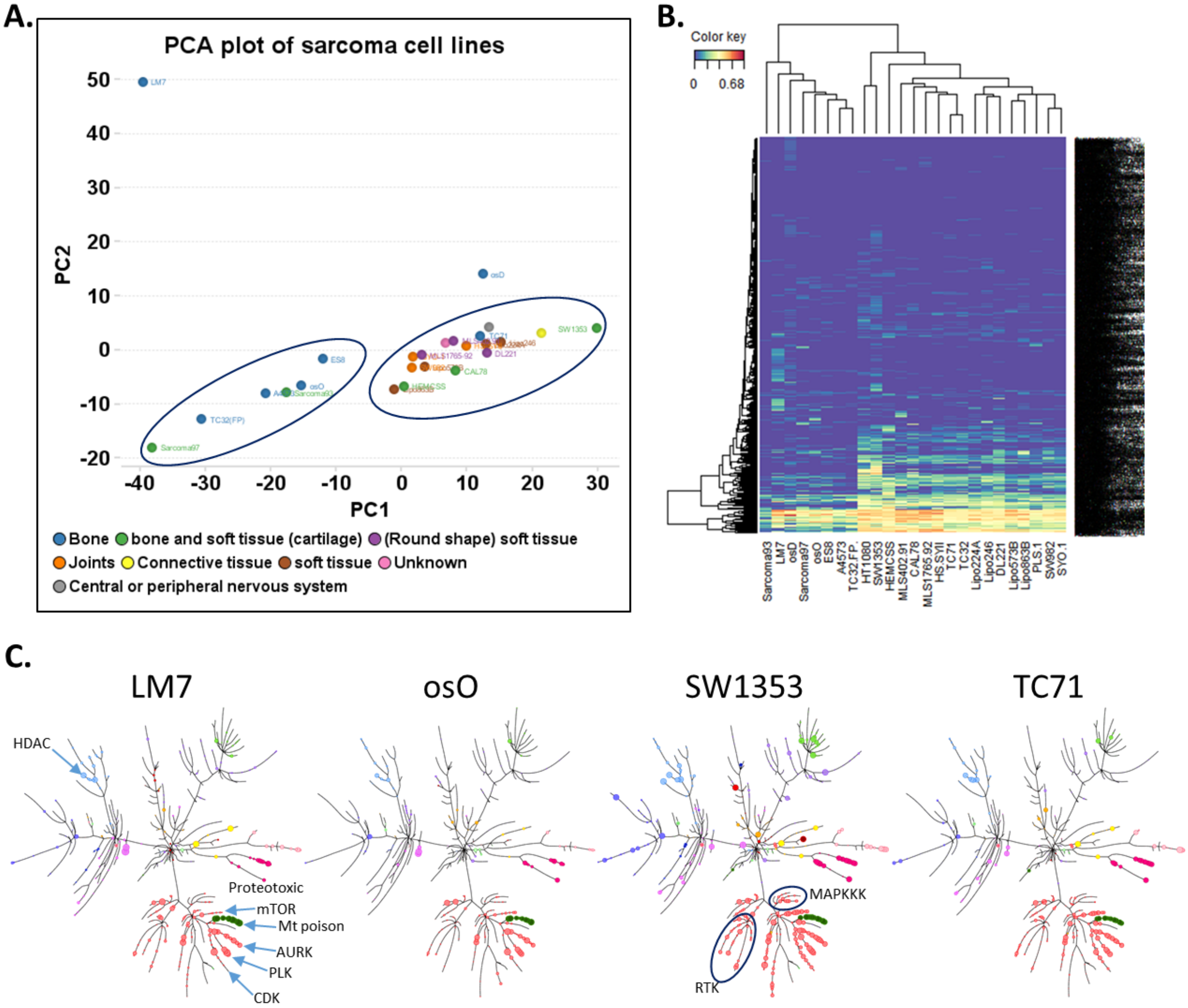


Fig 6.3) Exploratory analysis of sarcoma cell lines. A) Principle component analysis (PCA) of pharmacologic vectors for 25 sarcoma cell lines. Cell lines are color coded according to the tissue of origin. PCA analysis is performed using Pipeline Pilot and R. B) Two-way clustering heatmap of pharmacologic vectors and cell lines. Color code represents the AOC of the Hafner Growth index by concentration curve. A value of 0 (blue) represents no activity and 1 (red) represents complete killing of cells. Clustering is performed using R and cim function in the mixedOmics library. C) Pharmacologic networks of representative cell lines identified by the clustering analysis presented in B. Networks are constructed using R, igraph, and visNetwork libraries.

evidence that there was an underlying difference in the models, which resulted in a signature profile associated with the origin of the tumor. However, this embedding only accounted for 40% of the variance of the total dataset and has limited resolution of the pharmacologic profile. In order to ascertain that level of information a more granular data representation is more suitable. Here, we embedded the screening data in an analogous method to the ACME analysis performed by Seashore et al. This data representation similarly confirmed the presence of two larger clusters of response patterns that were associated with disease lineage but also shows that there was a gradient of activity with blocks of pan-active, selective, and inactive drugs. In general, less activity was observed in bone derived sarcomas when compared to soft tissue derived; however, this method does not easily result in a mechanistic understanding of what those drugs are targeting. To further enrich these data, visPharma-networks were constructed using method one and merged with the drug susceptibility data so that the size represents the strength of response. When comparing these cell lines, it is clear that much of the differences were in drugs that target kinases, shown in Fig 6.3C. Importantly, each circle is mechanistically annotated and allows for interactive exploration of the data. When applied here, it is apparent that drugs that target mTOR, CDK inhibitors displayed soft tissue selectivity while PLK, AURK, in addition to microtubule poisons were pan-active. Additional pan-active classes included proteotoxic stress induction via HSP or proteasome inhibitors, and HDAC inhibitors. Interestingly, these data also show that subset (3/7) of chondrosarcoma cell lines (HT1080, HMCSS, and SW1353) showed selectivity towards downstream MAPK inhibitors, farnesyl transferase inhibitors, a broad spectrum of RTK inhibitors, and JAK inhibitors. Indicating that this subset has a strong RAS signature.

Discussion

Data visualization is a useful tool used to contextualize the results of large data sets. At present, there is no single method of data visualization that allows one to easily understand the full complexity of HTS data sets. Here, we presented a generalizable method which embeds HTS data into a network formed from the target and MoA of the drug. This method is tunable towards libraries of different size and content, and results in visually congressing similar drugs into regions. Importantly, this method can provide context for screens performed on a single cell line or across multiple lineages. However, comparison of large numbers of cell lines using this method can become unmanageable. To overcome this, we performed a multi-tiered analysis aimed at understanding different levels of information in a given dataset. Here, we proposed using methods such as PCA and two-way clustering to identify prototype cell lines, representing which are further explored using the visPharma-Network method.

To exemplify this workflow, we applied it to a dataset containing drug sensitivity data from a panel of sarcoma cell lines. These data show how data visualization can be used to extract a basic understanding of the pharmacologic susceptibility profile between cell lines starting with a highly abstract method and honing into highly mechanistic views. From these data we identified clear differences between sarcomas of different lineages and identified a subset of chondrosarcomas with a strong RAS-like profile. Indeed, one of the chondrosarcomas, HT1080, is well established to have a driver mutation in N-

RAS. However, empirically testing the presence and activity levels of these signaling cascades across all the tested cell lines remains to be tested.

CHAPTER VII

CONCLUSIONS AND FINAL REMARKS

The research conducted throughout this dissertation are multifaceted and far ranging in their scientific and technological merit. At the most fundamental level, we developed and validated tools used for phenotypic screening. The application of these tools has supported a number of drug discovery and repurposing projects, each of which have contributed to the scientific body of knowledge in their respective fields

The first application discussed in Chapter III demonstrated the role of phenotypic screening in early drug development and optimization. The primary endpoint of this assay was the pharmacologic restoration of p27, which was directly measured using the florescent intensity of immunolabeled p27 and functionally confirmed using growth inhibition. The ability to simultaneously measure a direct and indirect readout of activity without the need to perform additional assays highlights the efficiency of image-based screening. Additionally, the image analysis method presented here has already served as a platform for additional ongoing screening projects with drastically different endpoints than the ones presented. Indeed, the same image analysis pipeline is currently being used to quantify the morphometric attributes of cells grown as monolayers and labeled with an actin marker. The combined morphometric feature vector was then interpreted using a random forest model which classifies cells into phenotypic bins with very high precision. Thus, showing the generalizability of this image analysis method towards applications outside of the one for which it was initially proposed.

The second application discussed in Chapter IV is the most specialized image analysis pipeline developed during this dissertation. Here, multiple fluorescent markers were used in combination to identify a specific organelle. Despite being highly specialized, this method has the potential to be further adapted towards assays where proximal labels need to be identified. Similar to the method presented in Chapter III, the screen performed in Chapter IV also benefited from having multiple image-based endpoints including the penetrance of the primary cilia and an orthangonal toxicity assay; further demonstrating the efficiency of image-based screening.

The third application, presented in Chapter V, represents a practical approach to preforming HTS on spheroid models. The results in this section strongly argue the necessity of using models such as 3D spheroids in drug discovery because it alters the pharmacologic susceptibility profile in a manner that is more consistent with clinical observation. Likewise, the development and implementation of these models in HTS is important for the next generation of *in vitro* drug testing. Indeed, the screening pipeline developed in Chapter V is currently being used to profile multiple cancers including breast, colon, and pancreatic. For some of these projects, primary patient-derived cells are being used. Scientifically, these represent interesting models due to their ability to recapitulate the genomic and epigenetic attributes of the clinical population. Thus, when combined with a physiologically relevant growth format these serve as a powerful predictive model of how a patient will respond to potential treatments. Traditionally, these type of data are acquired by performing experiments in xenograft mouse models, which are subsequently treated with potential therapeutic drugs. However, the time required expanding the mice, establishing a tumor, and the cost associated with testing multiple drug treatment arms makes this both cost and time prohibitive. However, application of *ex vivo* organoids grown in multi-well format provides a balance of speed, efficiency, and cost with biological and translational relevance.

Multiple HTS screens aimed at profiling the pharmacologic susceptibility of cell lines have been performed directly or by using tools developed for this dissertation, resulting in a large amount data which requires interpretation. Thus, one of the most powerful tools in data science to perform such a task is data visualization. Here, we compiled a knowledge base of pharmacologic modes of action which was used to construct a network of pharmacologic relations. Importantly, this method of visualization provides molecular context to the results of a HTS. It also allows one to make a rapid visual comparison between model systems and gain a high-level understanding of the mechanistic drivers on which each model relies. Finally, it shifts the notion of HTS as being performed in order to pick the best candidate drug towards using HTS as a method to profile and understand the molecular biology of a cell, which in turn can be used to prioritize potential therapies or combinatorial studies.

REFERENCES

1. Martínez-Abraín A. Statistical significance and biological relevance: A call for a more cautious interpretation of results in ecology. Acta Oecologica 2008; 34:9-11.

2. Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S*, et al.* Guidance on the assessment of the biological relevance of data in scientific assessments. EFSA Journal 2017; 15:e04970.

3. Pereira DA, Williams JA. Origin and evolution of high throughput screening. British Journal of Pharmacology 2007; 152:53-61.

4. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D*, et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 2008; 36:D901-6.

5. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P*, et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 2006; 34:D668-72.

6. Finkbeiner S, Frumkin M, Kassner Paul D. Cell-Based Screening: Extracting Meaning from Complex Data. Neuron 86:160-74.

7. Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK*, et al.* Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. Proceedings of the National Academy of Sciences 2009; 106:1826-31.

8. Neumann B, Held M, Liebel U, Erfle H, Rogers P, Pepperkok R*, et al.* High-throughput RNAi screening by time-lapse imaging of live human cells. Nature methods 2006; 3:385.

9. Walter T, Held M, Neumann B, Hériché J-K, Conrad C, Pepperkok R*, et al.* Automatic identification and clustering of chromosome phenotypes in a genome wide RNAi screen by time-lapse imaging. Journal of structural biology 2010; 170:1-9.

10. Neumann B, Walter T, Hériché J-K, Bulkescher J, Erfle H, Conrad C*, et al.* Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature 2010; 464:721.

11. Cavallaro A, Ebrahimi T. Interaction between High-Level and Low-Level Image Analysis for Semantic Video Object Extraction. EURASIP Journal on Advances in Signal Processing 2004; 2004:783262.

12. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O*, et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome biology 2006; 7:R100.

13. Uhlmann V, Singh S, Carpenter AE. CP-CHARM: segmentation-free image classification made accessible. BMC bioinformatics 2016; 17:1.

14. Orlov N, Shamir L, Macura T, Johnston J, Eckley DM, Goldberg IG. WND-CHARM: Multi-purpose image classification using compound image transforms. Pattern recognition letters 2008; 29:1684-93.

15. Rajaram S, Pavie B, Wu LF, Altschuler SJ. PhenoRipper: software for rapidly profiling microscopy images. Nature methods 2012; 9:635-7.

16. Roukos V, Pegoraro G, Voss TC, Misteli T. Cell cycle staging of individual cells by fluorescence microscopy. Nature Protocols 2015; 10:334.

17. Di Z, Herpers B, Fredriksson L, Yan K, van de Water B, Verbeek FJ*, et al.* Automated Analysis of NF-κB Nuclear Translocation Kinetics in High-Throughput Screening. PLOS ONE 2012; 7:e52337.

18. Ershad SF. Texture classification approach based on combination of edge & co-occurrence and local binary pattern. arXiv preprint arXiv:12034855 2012.

19. Chen Z, Karim MA, Hayat MM. Displacement co-occurrence statistics for binary digital images. ELECTIM 2002; 11:127-35.

20. Jiao Y, Berman H, Kiehl T-R, Torquato S. Spatial Organization and Correlations of Cell Nuclei in Brain Tumors. PLoS ONE 2011; 6:e27323.

21. Hanahan D, Weinberg RA. The Hallmarks of Cancer. Cell 100:57-70.

22. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144:646-74.

23. Pietras K, Östman A. Hallmarks of cancer: interactions with the tumor stroma. Experimental cell research 2010; 316:1324-31.

24. Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H*, et al.* Stromal gene expression predicts clinical outcome in breast cancer. Nat Med 2008; 14:518-27.

25. Pietras K, Pahler J, Bergers G, Hanahan D. Functions of paracrine PDGF signaling in the proangiogenic tumor stroma revealed by pharmacological targeting. PLoS medicine 2008; 5:e19.

26. Domingues P, González-Tablas M, Otero Á, Pascual D, Miranda D, Ruiz L*, et al.* Tumor infiltrating immune cells in gliomas and meningiomas. Brain, behavior, and immunity 2016; 53:1-15.

27. Turner NC, Reis-Filho JS. Genetic heterogeneity and cancer drug resistance. The lancet oncology 2012; 13:e178-e85.

28. Calderwood SK. Tumor Heterogeneity, Clonal Evolution, and Therapy Resistance: An Opportunity for Multitargeting Therapy. Discovery medicine 2013; 15:188-94.

29. Huang M, Shen A, Ding J, Geng M. Molecularly targeted cancer therapy: some lessons from the past decade. Trends in Pharmacological Sciences 2014; 35:41-50.

30. Prasetyanti PR, Medema JP. Intra-tumor heterogeneity from a cancer stem cell perspective. Molecular Cancer 2017; 16:41.

31. Juergens RA, Wrangle J, Vendetti FP, Murphy SC, Zhao M, Coleman B*, et al.* Combination Epigenetic Therapy Has Efficacy in Patients with Refractory Advanced Non–Small Cell Lung Cancer. Cancer Discovery 2011; 1:598-607.

32. Carew JS, Giles FJ, Nawrocki ST. Histone deacetylase inhibitors: Mechanisms of cell death and promise in combination cancer therapy. Cancer Letters 2008; 269:7-17.

33. Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price EV, Coletti ME*, et al.* Harnessing connectivity in a large-scale small-molecule sensitivity dataset. Cancer discovery 2015; 5:1210-23.

34. Teicher BA. Anticancer drug development guide: preclinical screening, clinical trials, and approval. ed). Springer Science & Business Media, 2013.

35. Pauli C, Hopkins BD, Prandi D, Shaw R, Fedrizzi T, Sboner A*, et al.* Personalized In Vitro and In Vivo Cancer Models to Guide Precision Medicine. Cancer Discovery 2017.

36. Kimlin LC, Casagrande G, Virador VM. In vitro three-dimensional (3D) models in cancer research: an update. Mol Carcinog 2013; 52:167-82.

37. Hirschhaeuser F, Menne H, Dittfeld C, West J, Mueller-Klieser W, Kunz-Schughart LA. Multicellular tumor spheroids: an underestimated tool is catching up again. Journal of biotechnology 2010; 148:3-15.

38. Durand RE, Olive PL. Resistance of tumor cells to chemo- and radiotherapy modulated by the three-dimensional architecture of solid tumors and spheroids. Methods in Cell Biology 2001; 64:211-33.

39. Horvath P, Aulner N, Bickle M, Davies AM, Nery ED, Ebner D*, et al.* Screening out irrelevant cell-based models of disease. Nat Rev Drug Discov 2016; advance online publication.

40. Sato T, Stange DE, Ferrante M, Vries RG, Van Es JH, Van Den Brink S*, et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology 2011; 141:1762-72.

41. Caliari SR, Burdick JA. A practical guide to hydrogels for cell culture. Nat Meth 2016; 13:405-14.

42. Kleinman HK, McGarvey ML, Hassell JR, Star VL, Cannon FB, Laurie GW*, et al.* Basement membrane complexes with biological activity. Biochemistry 1986; 25:312-8.

43. Boehnke K, Iversen PW, Schumacher D, Lallena MJ, Haro R, Amat J*, et al.* Assay establishment and validation of a high-throughput screening platform for three-dimensional patient-derived colon cancer organoid cultures. Journal of biomolecular screening 2016; 21:931-41.

44. Eke I, Hehlgans S, Sandfort V, Cordes N. 3D matrix-based cell cultures: Automated analysis of tumor cell survival and proliferation. International Journal of Oncology 2016; 48:313-21.

45. Vukicevic S, Kleinman HK, Luyten FP, Roberts AB, Roche NS, Reddi A. Identification of multiple active growth factors in basement membrane Matrigel suggests caution in interpretation of cellular activity related to extracellular matrix components. Experimental cell research 1992; 202:1-8.

46. Sirenko O, Mitlo T, Hesley J, Luke S, Owens W, Cromwell EF. High-content assays for characterizing the viability and morphology of 3D cancer spheroid cultures. Assay and drug development technologies 2015; 13:402-14.

47. Tung Y-C, Hsiao AY, Allen SG, Torisawa Y-s, Ho M, Takayama S. High-throughput 3D spheroid culture and drug testing using a 384 hanging drop array. Analyst 2011; 136:473-8.

48. Tseng H, Gage JA, Shen T, Haisler WL, Neeley SK, Shiao S*, et al.* A spheroid toxicity assay using magnetic 3D bioprinting and real-time mobile device-based imaging. Scientific reports 2015; 5:13987.

49. Raghavan S, Mehta P, Horst EN, Ward MR, Rowley KR, Mehta G. Comparative analysis of tumor spheroid generation techniques for differential in vitro drug toxicity. Oncotarget 2016; 7:16948-61.

50. Friedrich J, Seidel C, Ebner R, Kunz-Schughart LA. Spheroid-based drug screen: considerations and practical approach. Nature protocols 2009; 4:309-24.

51. Kim S-A, Lee EK, Kuh H-J. Co-culture of 3D tumor spheroids with fibroblasts as a model for epithelial–mesenchymal transition in vitro. Experimental Cell Research 2015; 335:187-96.

52. Herter S, Morra L, Schlenker R, Sulcova J, Fahrni L, Waldhauer I*, et al.* A novel three-dimensional heterotypic spheroid model for the assessment of the activity of cancer immunotherapy agents. Cancer Immunology, Immunotherapy 2017; 66:129-40.

53. Keller GM. In vitro differentiation of embryonic stem cells. Current Opinion in Cell Biology 1995; 7:862-9.

54. Kelm JM, Timmins NE, Brown CJ, Fussenegger M, Nielsen LK. Method for generation of homogeneous multicellular tumor spheroids applicable to a wide variety of cell types. Biotechnology and bioengineering 2003; 83:173-80.

55. Maschmeyer I, Lorenz AK, Schimek K, Hasenberg T, Ramme AP, Hübner J*, et al.* A four-organ-chip for interconnected long-term co-culture of human intestine, liver, skin and kidney equivalents. Lab on a Chip 2015; 15:2688-99.

56. Zanoni M, Piccinini F, Arienti C, Zamagni A, Santi S, Polico R*, et al.* 3D tumor spheroid models for in vitro therapeutic screening: a systematic approach to enhance the biological relevance of data obtained. Scientific reports 2016; 6.

57. Boutin ME, Hoffman-Kim D. Application and assessment of optical clearing methods for imaging of tissue-engineered neural stem cell spheres. Tissue Engineering Part C: Methods 2014; 21:292-302.

58. Cheng V, Esteves F, Chakrabarty A, Cockle J, Short S, Bruning-Richardson A. High-content analysis of tumour cell invasion in three-dimensional spheroid assays. Oncoscience 2015; 2:596-606.

59. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. British Journal of Pharmacology 2011; 162:1239-49.

60. Fox S, Farr-Jones S, Sopchak L, Boggs A, Nicely HW, Khoury R*, et al.* High-throughput screening: update on practices and success. J Biomol Screen 2006; 11:864-9.

61. Caraus I, Alsuwailem AA, Nadon R, Makarenkov V. Detecting and overcoming systematic bias in high-throughput screening technologies: a comprehensive review of practical issues and methodological solutions. Briefings in bioinformatics 2015; 16:974-86.

62. Zhang J-H, Chung TD, Oldenburg KR. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. Journal of biomolecular screening 1999; 4:67-73.

63. Malo N, Hanley JA, Cerquozzi S, Pelletier J, Nadon R. Statistical practice in high-throughput screening data analysis. Nat Biotech 2006; 24:167-75.

64. Hafner M, Niepel M, Chung M, Sorger PK. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nature methods 2016; 13:521-7.

65. Iversen P, Beck B, Chen Y, Dere W, Devanarayan V, Eastwood B*, et al.* Assay guidance manual. In. Bethesda, MD: Eli Lilly & Co and the National Center for Advancing Translational Sciences, 2004.

66. Bray MA, Carpenter A. Advanced Assay Development Guidelines for Image-Based High Content Screening and Analysis. In: Assay Guidance Manual. Sittampalam GS, Coussens NP, Brimacombe K*, et al.*: Bethesda (MD), 2004.

67. O'Connor PM, Jackman J, Bae I, Myers TG, Fan S, Mutoh M*, et al.* Characterization of the p53 tumor suppressor pathway in cell lines of the National Cancer Institute anticancer drug screen and correlations with the growth-inhibitory potency of 123 anticancer agents. Cancer Res 1997; 57:4285-300.

68. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nature Reviews Cancer 2006; 6:813-23.

69. Makarenkov V, Zentilli P, Kevorkov D, Gagarin A, Malo N, Nadon R. An efficient method for the detection and elimination of systematic error in high-throughput screening. Bioinformatics 2007; 23:1648-57.

70. Druker BJ, Tamura S, Buchdunger E, Ohno S, Segal GM, Fanning S*, et al.* Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr–Abl positive cells. Nature medicine 1996; 2:561.

71. Chang L, Graham PH, Ni J, Hao J, Bucci J, Cozzi PJ*, et al.* Targeting PI3K/Akt/mTOR signaling pathway in the treatment of prostate cancer radioresistance. Critical Reviews in Oncology/Hematology 2015; 96:507-17.

72. Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. British Journal of Pharmacology 2007; 152:9-20.

73. Sherr CJ. Cancer cell cycles. Science 1996; 274:1672-7.

74. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR*, et al.* Mutational Landscape of Metastatic Cancer Revealed from Prospective Clinical Sequencing of 10,000 Patients. Nature medicine 2017; 23:703-13.

75. Rodier G, Montagnoli A, Di Marcotullio L, Coulombe P, Draetta GF, Pagano M*, et al.* p27 cytoplasmic localization is regulated by phosphorylation on Ser10 and is not a prerequisite for its proteolysis. The EMBO Journal 2001; 20:6672-82.

76. Hara T, Kamura T, Nakayama K, Oshikawa K, Hatakeyama S, Nakayama K-I. Degradation of p27 Kip1 at the G0-G1 Transition Mediated by a Skp2-independent Ubiquitination Pathway. Journal of Biological Chemistry 2001; 276:48937-43.

77. Denicourt C, Saenz CC, Datnow B, Cui X-S, Dowdy SF. Relocalized p27^{Kip1} Tumor Suppressor Functions as a Cytoplasmic Metastatic Oncogene in Melanoma. Cancer Research 2007; 67:9238-43.

78. Schulman BA, Carrano AC, Jeffrey PD, Bowen Z, Kinnucan ER, Finnin MS*, et al.* Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. Nature 2000; 408:381-6.

79. Zheng N, Schulman BA, Song L, Miller JJ, Jeffrey PD, Wang P*, et al.* Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. Nature 2002; 416:703-9.

80. Chen Q, Xie W, Kuhn DJ, Voorhees PM, Lopez-Girona A, Mendy D*, et al.* Targeting the p27 E3 ligase SCF^{Skp2} results in p27- and Skp2-mediated cell-cycle arrest and activation of autophagy. Blood 2008; 111:4690-9.

81. Wu L, Grigoryan AV, Li Y, Hao B, Pagano M, Cardozo T. Specific Small Molecule Inhibitors of Skp2-Mediated p27 Degradation. Chemistry & biology 2012; 19:1515-24.

82. Chan C-H, Morrow John K, Li C-F, Gao Y, Jin G, Moten A*, et al.* Pharmacological Inactivation of Skp2 SCF Ubiquitin Ligase Restricts Cancer Stem Cell Traits and Cancer Progression. Cell 2013; 154:556-68.

83. Hao B, Zheng N, Schulman BA, Wu G, Miller JJ, Pagano M*, et al.* Structural basis of the Cks1-dependent recognition of p27(Kip1) by the SCF(Skp2) ubiquitin ligase. Mol Cell 2005; 20:9-19.

84. Zhang S, Du-Cuny L. Development and evaluation of a new statistical model for structure-based high-throughput virtual screening. Int J Bioinform Res Appl 2009; 5:269-79.

85. Zhang S, Kumar K, Jiang X, Wallqvist A, Reifman J. DOVIS: an implementation for high-throughput virtual screening using AutoDock. BMC Bioinformatics 2008; 9:126.

86. Svetnik V, Liaw A, Tong C, Wang T. Application of Breiman's Random Forest to Modeling Structure-Activity Relationships of Pharmaceutical Molecules. 2004; Berlin, Heidelberg. Springer Berlin Heidelberg

87. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. Journal of Chemical Information and Computer Sciences 2003; 43:1947-58.

88. Ceriani L, Verme P. The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. The Journal of Economic Inequality 2012; 10:421-43.

89. Sohraby F, Bagheri M, Aliyar M, Aryapour H. In silico drug repurposing of FDA-approved drugs to predict new inhibitors for drug resistant T315I mutant and wild-type BCR-ABL1: A virtual screening and molecular dynamics study. Journal of Molecular Graphics and Modelling 2017; 74:234-40.

90. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G*, et al.* Comprehensive analysis of kinase inhibitor selectivity. Nat Biotechnol 2011; 29:1046-51.

91. Zhang R, Xie X. Tools for GPCR drug discovery. Acta Pharmacologica Sinica 2012; 33:372-84.

92. Radley DC, Finkelstein SN, Stafford RS. Off-label prescribing among office-based physicians. Archives of Internal Medicine 2006; 166:1021-6.

93. Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE*, et al.* The Drug Repurposing Hub: a next-generation drug library and information resource. Nat Med 2017; 23:405-8.

94. Eid S, Turk S, Volkamer A, Rippmann F, Fulle S. KinMap: a web-based tool for interactive navigation through human kinome data. BMC Bioinformatics 2017; 18:16.

95. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. Nat Rev Mol Cell Biol 2017.

96. Seeley ES, Nachury MV. The perennial organelle: assembly and disassembly of the primary cilium. Journal of Cell Science 2010; 123:511-8.

97. Wang W, Wu T, Kirschner MW. The master cell cycle regulator APC-Cdc20 regulates ciliary length and disassembly of the primary cilium. Elife 2014; 3:e03083.

98. Kinzel D, Boldt K, Davis EE, Burtscher I, Trumbach D, Diplas B*, et al.* Pitchfork regulates primary cilia disassembly and left-right asymmetry. Dev Cell 2010; 19:66-77.

99. Inoko A, Matsuyama M, Goto H, Ohmuro-Matsuyama Y, Hayashi Y, Enomoto M*, et al.* Trichoplein and Aurora A block aberrant primary cilia assembly in proliferating cells. J Cell Biol 2012; 197:391-405.

100. Inaba H, Goto H, Kasahara K, Kumamoto K, Yonemura S, Inoko A*, et al.* Ndel1 suppresses ciliogenesis in proliferating cells by regulating the trichoplein-Aurora A pathway. J Cell Biol 2016; 212:409-23.

101. Kim S, Lee K, Choi JH, Ringstad N, Dynlacht BD. Nek2 activation of Kif24 ensures cilium disassembly during the cell cycle. Nat Commun 2015; 6:8087.

102. Khan N, Willemarck N, Marchand A, Binda M, Rueda N, Rueda-Rincon N*, et al.* Identification of Drugs that Restore Primary Cilium Expression in Cancer Cells. ed). 2016.

103. Dummer A, Poelma C, DeRuiter MC, Goumans M-JTH, Hierck BP. Measuring the primary cilium length: improved method for unbiased high-throughput analysis. Cilia 2016; 5:7.

104. Li L, Yang XJ. Tubulin acetylation: responsible enzymes, biological functions and human diseases. Cell Mol Life Sci 2015; 72:4237-55.

105. Kaelin WG. Von Hippel-Lindau disease. Annual review of pathology 2007; 2:145-73.

106. Seeger-Nukpezah T, Little JL, Serzhanova V, Golemis EA. Cilia and cilia-associated proteins in cancer. Drug discovery today Disease mechanisms 2013; 10:e135-e42.

107. Hsieh JJ, Purdue MP, Signoretti S, Swanton C, Albiges L, Schmidinger M*, et al.* Renal cell carcinoma. Nat Rev Dis Primers 2017; 3:17009.

108. Anderson K, Nordquist KA, Gao X, Hicks KC, Zhai B, Gygi SP*, et al.* Regulation of cellular levels of Sprouty2 protein by prolyl hydroxylase domain and von Hippel-Lindau proteins. J Biol Chem 2011; 286:42027-36.

109. Hasanov E, Chen G, Chowdhury P, Weldon J, Ding Z, Jonasch E*, et al.* Ubiquitination and regulation of AURKA identifies a hypoxia-independent E3 ligase activity of VHL. Oncogene 2017.

110. Ivan M, Kondo K, Yang H, Kim W, Valiando J, Ohh M*, et al.* HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. Science 2001; 292:464-8.

111. Kuznetsova AV, Meller J, Schnell PO, Nash JA, Ignacak ML, Sanchez Y*, et al.* von Hippel-Lindau protein binds hyperphosphorylated large subunit of RNA polymerase II through a proline hydroxylation motif and targets it for ubiquitination. Proc Natl Acad Sci U S A 2003; 100:2706-11.

112. Lai Y, Qiao M, Song M, Weintraub ST, Shiio Y. Quantitative proteomics identifies the Myb-binding protein p160 as a novel target of the von Hippel-Lindau tumor suppressor. PLoS One 2011; 6:e16975.

113. Mikhaylova O, Ignacak ML, Barankiewicz TJ, Harbaugh SV, Yi Y, Maxwell PH*, et al.* The von Hippel-Lindau tumor suppressor protein and Egl-9-Type proline hydroxylases regulate the large subunit of RNA polymerase II in response to oxidative stress. Mol Cell Biol 2008; 28:2701-17.

114. Na X, Duan HO, Messing EM, Schoen SR, Ryan CK, di Sant'Agnese PA*, et al.* Identification of the RNA polymerase II subunit hsRPB7 as a novel target of the von Hippel-Lindau protein. EMBO J 2003; 22:4249-59.

115. Okuda H, Saitoh K, Hirai S, Iwai K, Takaki Y, Baba M*, et al.* The von Hippel-Lindau tumor suppressor protein mediates ubiquitination of activated atypical protein kinase C. J Biol Chem 2001; 276:43611-7.

116. Xie L, Xiao K, Whalen EJ, Forrester MT, Freeman RS, Fong G*, et al.* Oxygen-regulated beta(2)-adrenergic receptor hydroxylation by EGLN3 and ubiquitylation by pVHL. Science signaling 2009; 2:ra33.

117. Xue J, Lv DD, Jiao S, Zhao W, Li X, Sun H*, et al.* pVHL mediates K63-linked ubiquitination of nCLU. PLoS One 2012; 7:e35848.

118. Thoma CR, Toso A, Gutbrodt KL, Reggi SP, Frew IJ, Schraml P*, et al.* VHL loss causes spindle misorientation and chromosome instability. Nature cell biology 2009; 11:994-1001.

119. Hell MP, Duda M, Weber TC, Moch H, Krek W. Tumor suppressor VHL functions in the control of mitotic fidelity. Cancer Res 2014; 74:2422-31.

120. Dere R, Perkins AL, Bawa-Khalfe T, Jonasch D, Walker CL. beta-catenin links von Hippel-Lindau to aurora kinase A and loss of primary cilia in renal cell carcinoma. Journal of the American Society of Nephrology : JASN 2015; 26:553-64.

121. Lolkema MP, Mans DA, Ulfman LH, Volpi S, Voest EE, Giles RH. Allele-specific regulation of primary cilia function by the von Hippel-Lindau tumor suppressor. European journal of human genetics : EJHG 2008; 16:73-8.

122. Thoma CR, Frew IJ, Hoerner CR, Montani M, Moch H, Krek W. pVHL and GSK3beta are components of a primary cilium-maintenance signalling network. Nature cell biology 2007; 9:588-95.

123. Henney JE. From the Food and Drug Administration. JAMA 2000; 283:1131.

124. Uray IP, Dmitrovsky E, Brown PH. Retinoids and rexinoids in cancer prevention: from laboratory to clinic. Semin Oncol 2016; 43:49-64.

125. Kim MS, Lim DY, Kim JE, Chen H, Lubet RA, Dong Z*, et al.* Src is a novel potential off-target of RXR agonists, 9-cis-UAB30 and Targretin, in human breast cancer cells. Mol Carcinog 2015; 54:1596-604.

126. Assaf C, Bagot M, Dummer R, Duvic M, Gniadecki R, Knobler R*, et al.* Minimizing adverse side-effects of oral bexarotene in cutaneous T-cell lymphoma: an expert opinion. Br J Dermatol 2006; 155:261-6.

127. Motzer RJ, Mazumdar M, Bacik J, Berg W, Amsterdam A, Ferrara J. Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma. J Clin Oncol 1999; 17:2530-40.

128. Motzer RJ, Murphy BA, Bacik J, Schwartz LH, Nanus DM, Mariani T*, et al.* Phase III trial of interferon alfa-2a with or without 13-cis-retinoic acid for patients with advanced renal cell carcinoma. J Clin Oncol 2000; 18:2972-80.

129. Aass N, De Mulder PH, Mickisch GH, Mulders P, van Oosterom AT, van Poppel H*, et al.* Randomized phase II/III trial of interferon Alfa-2a with and without 13-cis-retinoic acid in patients with progressive metastatic renal cell Carcinoma: the European Organisation for Research and Treatment of Cancer Genito-Urinary Tract Cancer Group (EORTC 30951). J Clin Oncol 2005; 23:4172-8.

130. Boorjian SA, Milowsky MI, Kaplan J, Albert M, Cobham MV, Coll DM*, et al.* Phase 1/2 clinical trial of interferon alpha2b and weekly liposome-encapsulated

all-trans retinoic acid in patients with advanced renal cell carcinoma. J Immunother 2007; 30:655-62.

131. Buentig N, Stoerkel S, Richter E, Dallmann I, Reitz M, Atzpodien J. Predictive impact of retinoid X receptor-alpha-expression in renal-cell carcinoma. Cancer Biother Radiopharm 2004; 19:331-42.

132. Goelden U, Ukena SN, Pfoertner S, Hofmann R, Buer J, Schrader AJ. RAR-beta(1) overexpression in chromophobe renal cell carcinoma: a novel target for therapeutic intervention? Exp Oncol 2005; 27:220-4.

133. Obara W, Konda R, Akasaka S, Nakamura S, Sugawara A, Fujioka T. Prognostic significance of vitamin D receptor and retinoid X receptor expression in renal cell carcinoma. J Urol 2007; 178:1497-503.

134. Wang Y, Zhang Z, Yao R, Jia D, Wang D, Lubet RA*, et al.* Prevention of lung cancer progression by bexarotene in mouse models. Oncogene 2006; 25:1320-9.

135. Tang XH, Osei-Sarfo K, Urvalek AM, Zhang T, Scognamiglio T, Gudas LJ. Combination of bexarotene and the retinoid CD1530 reduces murine oral-cavity carcinogenesis induced by the carcinogen 4-nitroquinoline 1-oxide. Proc Natl Acad Sci U S A 2014; 111:8907-12.

136. Santini MT, Rainaldi G. Three-Dimensional Spheroid Model in Tumor Biology. Pathobiology 1999; 67:148-57.

137. Kunz-Schughart LA, Freyer JP, Hofstaedter F, Ebner R. The use of 3-D cultures for high-throughput screening: the multicellular spheroid model. Journal of biomolecular screening 2004; 9:273-85.

138. Imamura Y, Mukohara T, Shimono Y, Funakoshi Y, Chayahara N, Toyoda M*, et al.* Comparison of 2D-and 3D-culture models as drug-testing platforms in breast cancer. Oncology reports 2015; 33:1837-43.

139. Kessel S, Cribbes S, Déry O, Kuksin D, Sincoff E, Qiu J*, et al.* High-Throughput 3D Tumor Spheroid Screening Method for Cancer Drug Discovery Using Celigo Image Cytometry. Journal of laboratory automation 2016:2211068216652846.

140. Perche F, Torchilin VP. Cancer cell spheroids as a model to evaluate chemotherapy protocols. Cancer Biology & Therapy 2012; 13:1205-13.

141. Ivanov DP, Parker TL, Walker DA, Alexander C, Ashford MB, Gellert PR*, et al.* Multiplexing spheroid volume, resazurin and acid phosphatase viability assays for high-throughput screening of tumour spheroids and stem cell neurospheres. PloS one 2014; 9:e103817.

142. Piccinini F. AnaSP: a software suite for automatic image analysis of multicellular spheroids. Computer methods and programs in biomedicine 2015; 119:43-52.

143. Chen W, Wong C, Vosburgh E, Levine AJ, Foran DJ, Xu EY. High-throughput image analysis of tumor spheroids: a user-friendly software application to measure the size of spheroids automatically and accurately. JoVE (Journal of Visualized Experiments) 2014:e51639-e.

144. Monazzam A, Razifar P, Lindhe Ö, Josephsson R, Långström B, Bergström M. A new, fast and semi-automated size determination method (SASDM) for studying multicellular tumor spheroids. Cancer cell international 2005; 5:1.

145. Markman B, Javier Ramos F, Capdevila J, Tabernero J. EGFR and KRAS in colorectal cancer. Advances in clinical chemistry 2010; 51:72.

146. Hamam R, Ali D, Vishnubalaji R, Alsaaran ZF, Chalisserry EP, Alfayez M*, et al.* Enhanced Efficacy of 5-Fluorouracil in Combination with a Dual Histone Deacetylase and Phosphatidylinositide 3-Kinase Inhibitor (CUDC-907) in Colorectal Cancer Cells. Saudi Journal of Gastroenterology : Official Journal of the Saudi Gastroenterology Association 2017; 23:34-8.

147. Carson R, Celtikci B, Fenning C, Javadi A, Crawford N, Carbonell LP*, et al.* HDAC Inhibition Overcomes Acute Resistance to MEK Inhibition in BRAF-Mutant Colorectal Cancer by Downregulation of c-FLIPL. Clin Cancer Res 2015; 21:3230-40.

148. Gray JW, Mills GB. Large-scale drug screens support precision medicine. Cancer discovery 2015; 5:1130-2.

149. Atkinson JM, Shelat AA, Carcaboso AM, Kranenburg TA, Arnold A, Boulos N*, et al.* An integrated in vitro and in vivo high throughput screen identifies treatment leads for ependymoma. Cancer cell 2011; 20:384-99.

150. Morfouace M, Shelat A, Jacus M, Freeman BB, Turner D, Robinson S*, et al.* Pemetrexed and gemcitabine as combination therapy for the treatment of Group3 medulloblastoma. Cancer cell 2014; 25:516-29.

151. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R*, et al.* PANTHER: a library of protein families and subfamilies indexed by function. Genome Res 2003; 13:2129-41.

152. Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Research 2010; 38:D204-D10.

153. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR*, et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 2017.

154. Park IY, Powell RT, Tripathi DN, Dere R, Ho TH, Blasius TL*, et al.* Dual Chromatin and Cytoskeletal Remodeling by SETD2. Cell 2016; 166:950-62.

155. Park IY, Chowdhury P, Tripathi DN, Powell RT, Dere R, Terzo EA*, et al.* Methylated α-tubulin antibodies recognize a new microtubule modification on mitotic microtubules. mAbs 2016; 8:1590-7.

156. Cooper LA, Kong J, Gutman DA, Dunn WD, Nalisnik M, Brat DJ. Novel genotype-phenotype associations in human cancers enabled by advanced molecular platforms and computational analysis of whole slide images. Laboratory investigation 2015; 95:366-76.

157. Lehrer M, Powell RT, Barua S, Kim D, Narang S, Rao A. Radiogenomics and Histomics in Glioblastoma: The Promise of Linking Image-Derived Phenotype with Genomic Information. In: Advances in Biology and Treatment of Glioblastoma. Somasundaram K. Springer International Publishing: Cham, 2017;143-59.

158. Powell RT, Olar A, Narang S, Rao G, Sulman E, Fuller GN*, et al.* Identification of Histological Correlates of Overall Survival in Lower Grade Gliomas Using a Bag-of-words Paradigm: A Preliminary Analysis Based on Hematoxylin & Eosin Stained Slides from the Lower Grade Glioma Cohort of The Cancer Genome Atlas. Journal of Pathology Informatics 2017; 8:9.

APPENDIX

Summary of SETD2 roles in genomic stability

Gene expression is a finely regulated process governed in part by post-translational modification of histones. These modifications include acetylation, methylation, and ubiquitination, which are catalyzed by histone acetyltransferases (HAT)/deacetylases (HDAC), histone methyltransferases (HMT)/demethylases (KDM), and ubiquitin ligases, respectively. Interestingly, some of the proteins that regulate the addition and removal of these post-translational modifications do not exclusively localize to the nucleus and can be found in the cytoplasm or at specific cellular structures including components of the cytoskeleton. We showed that SetD2, the histone methyltransferase which exclusively tri-methylates histone H3 at lysine residue 36 (H3K36me3), also directly interacts with microtubules and has the catalytic potential to tri-methylate lysine 40 of alpha tubulin.[154] We further hypothesized that the addition of methyl residues to microtubules is not a passive mark; but rather, directly participates in microtubule stability and function. In order to determine when and where methylation occurs on microtubules, we performed immunofluorescence labeling of H3k36me3 and alpha tubulin which showed co-localization of tri-methyl antibodies to a specific microtubule structure known as the midbody, which is important in symmetric cellular division.[155] In order to quantify these phenotypic behaviors, we developed a number of HC imaging pipelines to identify midbodies from a field of cells and methods to quantify multiple aspects of nuclear content and integrity. These included methods to determine the rate of polyploidy nuclei and presence of micronuclei. We also showed that acute homozygous depletion of Setd2, a common event in a broad spectrum of cancers including clear cell renal cell carcinoma, resulted in increased rates of genomic instability, consistent with aberrant cytokinesis.

Bio-medical image analysis beyond HTS/HCA applications

The ability to extract meaningful information form biological images is not limited to HTS. Other biomedical fields such as histology and radiology can also greatly benefit from automated image analysis. Here, automated image analysis can serve as a tool to quantitatively evaluate the cellular structures in tissue and describe perturbation in diseased tissue. In doing so, it pushes these fields towards a more standardized description of tissue and overcomes issues with qualitative analysis, which can vary between day, institution, and medical examiner. Image analysis can further be combined with machine learning to generate predictive models of disease. In part, these types of studies are made possible through large clinical archives, such as the TCGA, which store histological data with matched clinical and molecular data. Indeed, this has resulted in the development of fields such as radio-genomics and histo-genomics where phenotypic features are used to predict genomic characteristics of the disease.[156,157] Other applications include the development of predictive models which integrate image derived features which are used to predict the course and severity of disease.[158] Interestingly, unlike HTS pipelines, analysis of tissues and radiological images rely more heavily on field-based methods as the images are often difficult to segment. In tissues where segmentation is possible, direct morphometric measurements can be used to sub-classify disease in addition to more advanced statistical features which quantify the relational spatial distributions of cells within the field. Conceptually, development of image processing tools, identification of genotype-phenotype relations, and machine learning methods that are being developed in this field can also be applied to HTS and vice-versa. Therefore, it is important to carefully watch both fields to identify novel applications of these technology.