# Developing and applying systems-based approaches to kinase-centered biology

Thomas Bello

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Taranjit S. Gujral, Chair

Nathan Price

Richard Gardner

Program authorized to offer degree:

Molecular and Cellular Biology

University of Washington

## Abstract

Developing and applying systems-based approaches to kinase-centered biology

Thomas Bello

Chair of supervisory committee:

Taranjit S. Gujral

Division of Human Biology, Fred Hutchinson Cancer Research Center

Kinases play diverse and critical roles that span nearly the entire field of molecular biology. As the available methods and knowledge regarding kinase biology continues to grow in the high-throughput era of the biosciences, there is a great need for new tools to interpret the resulting data and connect it to known information. Here, I develop and apply three such tools for the study of kinase biology. The first tool, KInhibition, addresses the polypharmacology of small molecule kinase inhibitors (KIs) by connecting researchers to available data from large-scale drug screens. By leveraging these data in combination with a newly-developed metric for quantifying selectivity, I constructed a portal that enables data-driven decisions on the most selective kinase inhibitor to use for a specific application. I further exploit this polypharmacology in my application of Kinome Regularization (KiR) to the study of late stage prostate cancer. I performed a series of small-scale KI screens in a panel of prostate cancer cell

lines and applied an updated and improved version of KiR to build predictive models of kinase signaling. These models predicted two compounds, PP121 and SC-1, that effectively inhibited growth of late-stage prostate cancer tumors in multiple *in vitro* and *in vivo* model systems. Finally, I extend the scope of KiR to a network-based modeling tool named KiRNet. KiRNet integrates functional hypotheses and additional large-scale, molecular data types with known protein-protein interactions to predict differentially regulated subnetworks within a model system. I demonstrate the potential of KiRNet by identifying known and new regulators of a mesenchymal cell state in hepatocellular carcinoma. All three of these tools bridge important gaps in the field of kinase biology and will accelerate future research progress in this area.

# Acknowledgements

*"If I have seen further, it is by standing on the shoulders of Giants"*
— Sir Isaac Newton

I thank my PI, mentor, and boss, Taran Gujral, for his guidance and support over the past four years. I could not have done any of this without his patience and expertise.

I thank my lab mate, fellow graduate student, and bay buddy, Andrew Bondesson, for always being willing to engage in a conversation, short or long, about anything from Src homology domains to the different tiers of memes.

I thank all current and former members of the Gujral lab for their endless assistance and encouragement throughout my time in the lab.

To my parents, Larry and Kate Bello: thank you. You have blessed me in more ways than I ever thought possible.

To my family and friends that have continued to walk by my side throughout this journey: thank you. Graduate school has taken me through many uncharted waters, and I am grateful to have emerged with you all still by my side.

To Anna Russell: thank you. You are amazing. I never imagined I would have someone as wonderful as you to walk through life with. You have kept me going when I would have given up. You have inspired me, challenged me, and helped me to become the person that I am today.

**Funding sources:**

Fred Hutchinson Cancer Research Center Interdisciplinary Dual Mentor Fellowship in Cancer Research; NIH/NCI K22CA201229; Pacific Northwest Prostate Cancer SPORE (P50CA097186-14); Sage Bionetworks; and Fred Hutchinson Cancer Research Center new investigator startup fund.

**Chapter 2:**

I thank Dirk Peterson, Robert McDermott, and the Fred Hutch Scientific Computing Core for their help and support in Shiny, Git, and in deploying the application to the web.

**Chapter 3:**

The *in vivo* experiments in this chapter were performed by Eleonora and Claudia Paindelli at The University of Texas MD Anderson Cancer Center.

**Chapter 4:**

The cell culture experiments in this section were performed by Marina Chan and Taran Gujral. The TCGA analysis was performed by Nithisha Khasnavis, using methods previously developed by Andrew Xue.

Some of the results shown in this chapter here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

# Table of Contents

# Lists of Figures, Tables, and Equation

**Figures:**



**Tables:**



**Equations:**

# Chapter 1: Introduction

## *1.1 KINASES ARE CRITICAL MEDIATORS OF CELLULAR ACTIVITY*

There are approximately 20,000 protein-coding genes in the human genome.[1] Each of these protein products performs a specific molecular and cellular function that, together, constitute the set of activities we recognize as human life. Of these 20,000 genes, approximately 538 encode for a particular class of enzyme called kinases.[2–4] Kinases catalyze the transfer of the γ-phosphate group of adenosine triphosphate (ATP) to serine, threonine, or tyrosine residues on their substrates, inducing conformational and functional changes on the target protein.[5] This phosphorylation is reversible through the action of another class of enzymes called phosphatases.[2,5] Kinases themselves are often regulated via phosphorylation by a different kinase, creating a chain of phosphorylation and activation events referred to as a "kinase cascade".[5,6] As one of the most common forms of post-translational modification, this molecular "switch" regulates nearly all cellular processes, including metabolism, transcription, cell cycle progression, migration, differentiation, and apoptosis.[2,4] Consequently, aberrant or misregulated kinase activity has been implicated in a plethora of human diseases, including 448 kinases that are directly linked to specific hallmarks of cancer.[2–5,7,8] Indeed, the first proto-oncogene identified was the tyrosine kinase Src.[7] Studying kinase biology therefore offers a lucrative opportunity to expand our understanding of cellular regulation and develop novel (or repurposed) agents for treating disease and improving human life.

Kinases are categorized into distinct groups based on their target substrates and evolutionary phylogeny (Figure 1.1).[5,9,10] The majority of kinases phosphorylate other proteins, while a small number of kinases target lipids. The most notable lipid kinases are the phosphotidyl-inositol-3 kinases (PI3Ks), which are extremely prominent upstream activators of

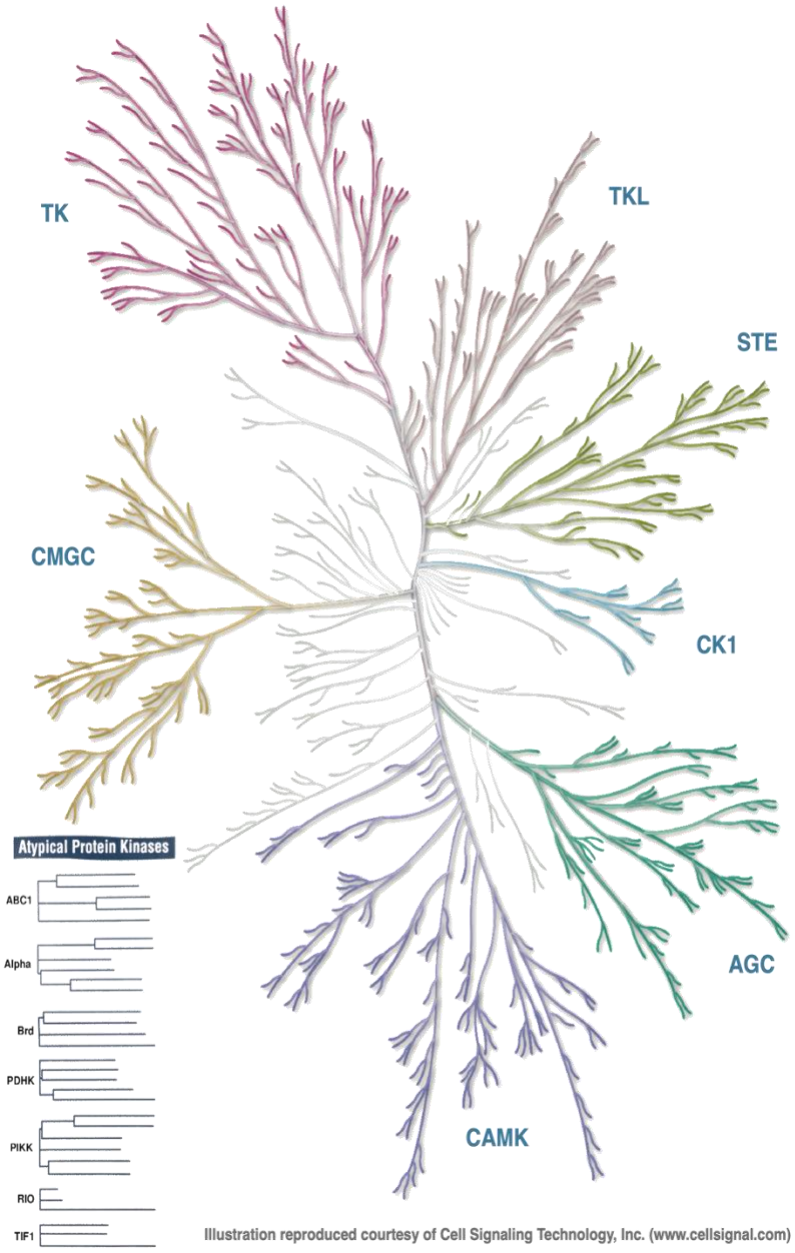


**Figure 1.1. Phylogenetic tree of the human protein kinome.** Image generated using KinMap (http://www.kinhub.org/kinmap/).

many cellular activities and have been heavily implicated in numerous cancers.[9] Protein kinases are first divided based on their target amino acid residue: serine/threonine kinases (STKs), tyrosine kinases (TKs), and dual-specificity kinases, which can phosphorylate all three residues. STKs constitute the largest group and can be further subdivided into the following families: **AGC** (protein kinase A, protein kinase G, and protein kinase C); **CaMK** (calcium/calmodulin

dependent kinases); **CK1** (cell kinase 1, previously casein kinase 1); **CMGC** (Cyclin-dependent kinases (CDKs), mitogen-activated protein kinases (MAPKs), glycogen synthase kinase 3 (GSK3), and CLK kinases); **STE** (sterile kinases); and **TKL** (tyrosine kinase like).[5] Tyrosine kinases are frequently divided into receptor tyrosine kinases (RTKs, e.g. EGFR) and non-receptor tyrosine kinases (e.g. Src). While tyrosine is by far the least commonly phosphorylated residue of the three (between 100 and 1000 times less common than serine), TKs are nevertheless regarded as some of the most crucial regulators of cellular activity and are heavily implicated in numerous pathologies.[3,7,11–19] Despite these divisions, all active kinases contain a catalytic domain that includes an ATP binding pocket.[5] As a result, nearly all kinases share a high degree of similarly in their 3D structures, particularly near this domain.[20,21] Kinases thus represent a fascinating paradox, wherein their molecular shapes and functions are relatively homogenous, but the consequences of their activities are incredibly diverse and span nearly the entire field of cellular biology.

### *1.2 INTERROGATING THE HUMAN KINOME THROUGH PERTURBATION BIOLOGY*

Since the discovery of reversible phosphorylation in 1955 by Edmond Fischer and Edwin Krebs, there have been enormous efforts to describe, perturb, and exploit the human kinome.[2,4,22,23] Initial studies used predominantly genetic approaches to uncovering the roles of newly-identified kinases. In 2002, Manning et al. published the first comprehensive catalogue of 518 human protein kinases.[4] Modern genetic approaches include kinome-wide RNA-interference (RNAi) or CRISPR/Cas9 screens to reduce or remove expression of individual kinases and observe the effects of this perturbation on the phenotype.[24–27] These approaches have numerous advantages, including the ability to target a single gene or transcript, making them appropriate

for answering many biological questions. However, these approaches can be extremely costly and time consuming to perform. Furthermore, genetic manipulations are seldom suitable for clinical applications, creating a need for a more translational approach to studying kinase biology.

In the past three decades, there has been an explosion in the number of available small molecule drugs that can inhibit kinase activity, collectively called kinase inhibitors (KIs).[3,16,20] The first kinase inhibitor to be approved for clinical use was the ROCK inhibitor fasudil, approved in Japan in 1994 for the treatment of cerebral vasospasm in aneurysm patients.[3,28] The first FDA approved KI was rapamycin (sirolimus) in 1999 to prevent graft rejection, followed by Imatinib (Gleevec®) 2001 for the treatment of BCR-Abl mutated chronic myeloid leukemia.[3,16,20,29–31] By 2019, there were 48 FDA-approved small molecule kinase inhibitors, 41 of which are for oncology-related uses.[6] Outside of clinical uses, there are thousands of KIs available for research purposes. The ability to specifically target and block a protein-protein interaction has proven to be invaluable in biological research, and KIs provide a powerful tool to accomplish this purpose.[31]

Due to the aforementioned structural similarity among kinases, nearly all widely available KIs bind to and inhibit more than one target. This polypharmacology can be challenging to work with, but it confers many benefits as well. For example, in clinical use, off-target effects of a drug can cause toxic side effects for the patient, potentially making it unsuitable. On the other hand, an inhibitor that targets multiple kinases can combat the development of therapeutic resistance by simultaneously targeting many pathways that would otherwise act as compensatory escape mechanisms.[32–35] In fact, it has been shown for multiple FDA-approved KIs that a genetic knockout of their purported targets is not sufficient to mimic

the phenotypic results, suggesting this polypharmacology is intrinsically critical to the clinical efficacy of KIs.[36] This polypharmacology can also be exploited for research purposes. In particular, there have been multiple large-scale studies that systematically quantify the precise effects of a given KI on a particular human kinase.[21,37–42] When iterated over a panel containing hundreds of KIs, this creates not only a wealth of data, but new opportunities for perturbation biology.

### *1.3 COMPUTATIONAL MODELS OF KINASE SIGNALING*

The scale and complexity of kinase biology necessitates the use of mathematical and computational models to help describe and understand observed phenomena. For example, one can perform a drug screen wherein a panel of KIs are applied to a cellular system, and the effect of each treatment on a phenotype (e.g. cell proliferation) is observed. Such an experiment will reveal which KIs can affect the phenotype, but, due to the polypharmacology of KIs, may not identify the underling kinases and signaling paths that specifically regulate that phenotype. One approach to tackle this challenge is through a previously-developed method called Kinome Regularization (KiR).[43] KiR is a powerful approach that integrates a small-scale functional KI screen with biochemical profiling via regression-based mathematical models. It began with one of the aforementioned large-scale KI studies, performed by Anastassiadis et al., that profiled the effect of 178 KIs on approximately 298 human kinases.[42] Briefly, KIs were incubated in a cell-free system containing a recombinant human protein kinase and an appropriate substrate. The catalytic activity of the kinase was quantified using radioisotope labeled phosphate and a filtration-binding assay, wherein only the phosphorylated products will bind to the filter and be detected.[42,44] This activity was expressed as a percentage of the uninhibited kinase's activity (i.e.

"percent of control" data, as opposed to binding affinities such as $K_i$ values). Multiplexing this approach resulted in a large, pairwise matrix of kinases and KIs, where each entry is the residual activity of that particular kinase in the presence of that KI (at the measured concentration). These >400 inhibition profiles already represent a powerful dataset, but this is only the starting place for KiR.

From these profiles, Gujral et al. computationally selected a panel of ~32 KIs that best captured the variation in the inhibition profiles and provided the best coverage of the profiled kinome.[43] These 32 KIs were then used as the basis for functional drug screens to study virtually any quantifiable phenotype, from cell migration to malarial liver stage infection.[43,45] For example, to study cell migration *in vitro*, cells were plated in 96-well plates and a scratch wound was made through the middle of each well. Cells were allowed to grow and close the wound in the presence of the 32 chosen KIs. The endpoint measurement was the relative wound density (area of the initial wound occupied by cells at the final timepoint). This gave a phenotypic response value for each of the 32 tested KIs. From the prior screen, each KI also corresponds to a set of residual activities for 298 human protein kinases. Thus, these two were combined to regress the phenotypic responses (response variables) onto the 298 kinase activities (explanatory variables). They applied elastic net regularization, a modified form of multiple linear regression that imposes an added penalty for each non-zero coefficient included in the final model.[46,47] Specifically, elastic net regularization seeks to solve the following minimization problem:

$$\min_{\beta,\beta_0}\left(\frac{1}{2N}\sum_{i=1}^{N}(y_i-\beta_0-x_i^T\beta)^2+\lambda\sum_{j=1}^{p}\left(\frac{(1-\alpha)}{2}\beta_j^2+\alpha\left|\beta_j\right|\right)\right) \qquad \textit{Eq. 1.1}$$

where the variables are as defined in table 1.1. This regularization serves the dual purpose of allowing a solution to be reached even in the case of only ~32 observations and 298 possible

| Variable | Meaning |
|---|---|
| $N$ | Number of observations (KIs used) |
| $p$ | Number of explanatory variables (kinases) |
| $\beta_j$ | Coefficient of kinase $j$ |
| $\beta_0$ | Constant coefficient (intercept) |
| $x_i$ | Inhibition profile of drug $i$ |
| $y_i$ | Measured phenotype for drug $i$ |
| $\lambda$ | Regularization penalty weight coefficient |
| $\alpha$ | Relative weighting between LASSO and ridge regularization |

**Table 1.1. Definition of variables used in elastic net formula (eq. 1.1).**

explanatory variables (the $p \gg N$ problem), while also eliminating those kinases that are poorly correlated with the phenotype by driving their coefficient to zero. The result is a set of coefficients, one for each kinase, that represents that kinase's predicted contribution to the phenotype. A higher coefficient suggests the activity of that kinase more heavily contributes to the phenotype (and thus inhibiting that kinase would abrogate the phenotype), while a coefficient of zero means the corresponding kinase is not predicted to contribute to the phenotype (and thus inhibiting that kinase it would have no effect on the phenotype). A typical KiR model contains between 5 and 30 nonzero coefficients, representing a semi-ordered list of "key kinases" that are predicted to affect the measured phenotype. Finally, these coefficients, combined with the previous inhibition profiles, allow the KiR model to predict the phenotypic response to all 178 KIs in the dataset, including over 100 untested KIs. In short, KiR produces a vast trove of data from a relatively small set of experiments through the integration of chemical biology and mathematical modeling.

Systems-based approaches, such as KiR, seek to describe and understand how all of the components as a system operate together, rather than studying each component in isolation. A common theme in these approaches is network-based modeling.[48–54] For example, the binary

status of phosphorylation conferred by kinases is often modeled using Boolean networks, where each node assumes one of only two possible states.[50,52] A kinase cascade can be represented as a series of such nodes, each one representing a kinase or substrate, and a set of edges connecting the nodes, each one representing a specific phosphorylation event. Molecular signals, in the form of phosphorylation-mediated state changes, are represented by successive state changes of these nodes as the information flows through the network. Such networks are a particular type of protein-protein interaction (PPI) networks, wherein each node represents a specific protein, and each edge represents a known interaction between two proteins (e.g. binding, activation, phosphorylation, methylation).[55] These networks serve to visually and computationally represent observations, place results into a broader context, and guide future experiments in meaningful directions. The sheer complexity of kinase signaling gives systems-based approaches such as these a huge advantage in developing an accurate and comprehensive understanding of its biological implications.

### *1.4 Addressing current challenges in kinase biology*

Studying kinase biology offers fascinating insights into the molecular underpinnings of cellular activity and provides ample opportunities for translational research. However, there are still many challenges to interrogating the human kinome using current approaches. One challenge that has already been mentioned is the widespread promiscuity of KIs. While this polypharmacology creates numerous unique opportunities, it nevertheless remains an obstacle when one desires to inhibit only a specific kinase. In chapter 2, I present a new tool named KInhibition designed to directly address this issue. I defined a new selectivity metric that allows for a user-defined set of targets and applied this metric to four large datasets containing

quantitative drug-target profiles of KIs. The results are publicly accessible through an easy-to-use web portal that allows researchers to make data-driven decisions regarding KI use.

As if the polypharmacology of KIs wasn't difficult enough, kinases themselves can exhibit a diversity of functions, with the same kinases having vastly different phenotypic consequences in different contexts. Deconvolving both the underlying kinase drivers of a phenotype, and then identifying the optimal inhibition strategy to abrogate that phenotype, remains an ever-present challenge for early stage drug discovery. In chapter 3, I take a systems based approach to identifying kinase targeted therapies for late-stage prostate cancer (PCa). While much of PCa research focuses on hormonal signaling, there is growing evidence that abberant kinase signaling may drive late-stage tumor activity and underly therapeutic resistance.[17,19,56–64] Using KiR, I identified two small molecule KIs that effectively inhibited PCa growth in a panel of *in vitro* model systems. In a collaborative effort, we were able to demonstrate these compounds were also effective *in vivo*, demonstrating a promising therapeutic potential of kinase-targeted therapies for late-stage PCa. This also demonstrates the power of a systems-based approach to translational biology.

Finally, there is an inherent challenge in placing the results of a functional screen like KiR into the context of prior knowledge. KiR is a powerful tool for predicting critical kinases that mediate a phenotype, and it's predictions do not exist in a vacuum. In chapter 4, through a joint effort between myself and other Gujral lab members, we apply KiR to study the cellular properties of a particular hepatocellular carcinoma (HCC) cell line. By integrating the results of KiR with a PPI database and additional large-scale datasets, I identified a kinase-centered subnetwork of proteins that regulates a mesenchymal-like cell state in HCC. By validating the

importance of this network in across model systems, I demonstrate the power of network-based models in analyzing the human kinome.

# Chapter 2: KInhibition: a kinase inhibitor selectivity metric and selection portal

A version of the following work is published in[65]:

Bello T, Gujral TS. KInhibition: A Kinase Inhibitor Selection Portal. iScience. 2018;8:49-53. doi:10.1016/j.isci.2018.09.009

## *2.1 INTRODUCTION*

Despite the enormity of research that has been done on kinase signaling, there remains a confounding challenge when attempting to selectively inhibit a desired molecule. In large part due to high structural similarity among kinases, nearly all available small molecule kinase inhibitors exhibit some promiscuity, causing undesired "off-target" effects.[37,40,42,66] Even many compounds described as being "specific" or "selective" have confounding off-target effects, making the selection, use, and analysis of the appropriate kinase inhibitor difficult. A number of large-scale kinase inhibitor screens have been undertaken in an attempt to quantify these effects, and many of these results are publicly available.[21,38,39,41,42,67] However, these data are decentralized and difficult to query, with results often being spread across multiple data files that must be downloaded and opened individually. Furthermore, there are multiple methods employed for representing or quantifying "selectivity", which may yield conflicting results.[37,40–42,68–70] Thus, the ultimate challenge of identifying the right kinase inhibitor for a biological task remains unresolved. Here I present KInhibition, a powerful platform tool that allows researchers search through multiple kinase inhibitor screens, visualize the relevant data, and choose the most selective and appropriate kinase inhibitor for the task at hand. We anticipate that KInhibition will

be adopted by the broader research community equivalent to RNAi or CRISPR guide RNA selection tools.

### *2.2 THE THEORY BEHIND KINHIBITION*

KInhibition is a platform tool designed to answer the question "Which compound should be used to inhibit a chosen kinase or pathway?" The first, but often most critical, step in answering this question is to locate and format the relevant data from the publicly available kinase inhibitor screens. These datasets may be initially found in somewhat intractable formats, but nearly all of them can be summarized by a matrix of drug-target interactions, with rows representing the compounds tested, columns representing the kinases screened, and entries being that compound's effect on that kinase. The main requirement for a dataset to be included in KInhibition is for it to be formatted in this manner, making the addition of future datasets to this platform a relatively trivial task.

A unified data format allows us to address the more nuanced issue of how to quantify "selectivity" of a given inhibitor. While there have been many proposed metrics for selectivity, very few are computationally robust enough to apply across datasets.[41,42,68–70] I developed a "KInhibition Selectivity Score", which quantifies the selectivity of a compound based on its on-target inhibition ("inhibition score") and its off-target effects ("inhibition penalty"). The inhibition score is simply the inhibition of the selected kinase, or a geometric mean if multiple kinases are chosen. The inhibition penalty is further divided into two sub-penalties. The first quantifies the broad inhibition activity of a compound, and will therefore best account for the extreme case where a compound inhibits nearly every kinase tested, but to a small degree (e.g. 10% of control) compared to the intended target. The second sub-penalty specifically quantifies

off-target effects that are close in magnitude to the inhibition of the intended target. This accounts for the other extreme case, where a compound inhibits only a few (e.g. 2–10) off-target kinases, but with a magnitude comparable to or greater than the intended target. Both of these extremes represent distinct types of off-target effects that must be considered when choosing the appropriate kinase inhibitor for an experiment.

The KInhibition Selectivity Score has numerous advantages that give the user the most relevant information and the most control over the decision of which compound to use. First, it is designed to work with percent-of-control data, rather than binding coefficients ($K_i$) or $IC_{50}$ values like previously reported partition indices or entropy-based scores.[70] This allows it to be used for screens performed even at a single dose, treating additional doses as separate compounds or entries in the matrix. Secondly, this score quantifies selectivity for a user-defined set of up to 10 "on-target" kinases, rather than simply basing all calculations on the most inhibited kinase for each compound. Thirdly, as previously mentioned, this score accounts for both the number and the magnitude of off-target effects better than previous scores (such as the Gini coefficient), allowing researchers to select the inhibitor most suited to their needs. Finally, this scoring metric does not rely on any hard-coded values, arbitrary thresholds, or data binning, giving it an advantage over S(x) scores or Ambit scores.[68]

### *2.2 Calculating the KInhibition Selectivity Score*

The KInhibition Selectivity Score, $KISS$, quantifies the selectivity of a given compound screened at a specific dose against a large panel of kinases. The data from such screens is treated as an $m \times n$ drug-target interaction matrix, where $m$ is the number of compounds (with different doses of the same compound being treated as separate compounds for this purpose) and

$n$ is the number of kinases screened (Figure 2.2). All entries in this matrix fall in the range $[0, 100]$ and represent the inhibition of the kinase by that compound. This is "percent of control" data, with 0 being no inhibition (activity of the kinase equal to or greater than the control), and 100 being complete inhibition (no detected kinase activity in the presence of that compound).

Before computing the score, each row in the drug-target interaction matrix is scaled so that the maximum of every row is exactly 100. This decouples compounds' selectivity from their efficacy to avoid unnecessarily penalizing compounds that are highly selective, but display low absolute inhibition at the dose tested. Efficacy at the tested dose can still be determined in the final results table, which displays the unscaled inhibition values.

The following calculations are computed row-wise on the scaled drug-target matrix, so that each compound has an assigned selectivity score based solely on the properties of that compound, independent of the other compounds (rows) in the matrix. The non-missing elements of each row, $z_1 \dots z_n$, can be partitioned into on-target effects, $x_1 \dots x_k$, and off-target effects, $y_1 \dots y_{n-k}$. First, the Inhibition Score, $IS$, is computed from the $k$ chosen on-target inhibitions $x_1 \dots x_k$ as a geometric mean:

$$IS = \sqrt[k]{\prod_{i=1}^{k} x_i} \qquad \textit{Eq. 2.1}$$

I used a geometric mean to best represent compounds that inhibit only some of the chosen on-target kinases, as a geometric mean is always less than or equal to an arithmetic mean. The value from equation 2.1 thus lies in the interval of the scaled data:

$$0 \leq IS \leq 100 \qquad \textit{Eq. 2.2}$$

The Inhibition Penalty, $IP$, is then computed in two parts. The first part, $IP_1$, is computed as an arithmetic mean of the off-target effects $y_1 \dots y_{n-k}$:

$$IP_1 = \left(\frac{1}{n-k}\right)\sum_{i=1}^{n-k} y_i \qquad \textit{Eq. 2.3}$$

This portion of the penalty represents the baseline inhibition across all of the off-target kinases, which is useful for penalizing compounds with extremely broad activity. However, it poorly accounts for compounds that have a small number (i.e. <10) of very large (near 100% inhibition) off-target effects compared to the on-target effects. To better capture this aspect, $IP_2$ is computed as a ratio of normalized variances between the off-target effects $y_1 \dots y_{n-k}$ and the total effects $z_1 \dots z_n$ as follows. First, the variances are computed as

$$s_y^2 = \left(\frac{1}{n-k-1}\right)\sum_{i=1}^{n-k}(y_i - \bar{y}) \qquad \textit{Eq. 2.4}$$

$$s_z^2 = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(z_i - \bar{z}) \qquad \textit{Eq. 2.5}$$

where $\bar{y} = \left(\frac{1}{n-k}\right)\sum_{i=1}^{n-k} y_i$ and $\bar{z} = \left(\frac{1}{n}\right)\sum_{i=1}^{n} z_i$ are the arithmetic means of the off-target and total inhibitions, respectively. When $k \ll n$, we expect these variances to be approximately equal for all but the most selective compounds, whose total variances will instead be explained largely by the difference between the on- and off-target effects. An approximate lower bound for the off-target variance can be computed using a modified version of the von Szokefalvi Nagy Inequality:

$$s_{lo}^2 = \frac{max(y_i)^2}{2n} \qquad \textit{Eq. 2.6}$$

The variances can then be centered using this lower bound in order to shift $IP_2$ to have a minimum closer to $0$, and $IP_2$ is calculated as the ratio between the two shifted variances:

$$IP_2 = \frac{s_y^2 - s_{lo}^2}{s_z^2 - s_{lo}^2} \qquad \textit{Eq. 2.7}$$

While the modifications to the lower bound (namely, using $n$ rather than $n - k$ in the denominator) mean that $IP_2$ can fall outside the interval of $[0, 1]$, in practice this only occurs for the most selective compounds, and then only by a very small margin. I then combine the two penalties in an empirically determined manner:

$$IP = \frac{IP_1 + 100(IP_2)^5}{2} \qquad \textit{Eq. 2.8}$$

With the rare exception mentioned above, both terms in the numerator of this equation lie in the interval $[0, 100]$, and therefore their arithmetic mean $IP$ does as well. Finally, $KISS$ is calculated as a different between $IS$ and $IP$

$$KISS = IS - IP \qquad \textit{Eq. 2.9}$$

Based on the bounds mentioned above, $KISS$ generally lies in the interval $[-100, 100]$, with $100$ representing near perfect selectivity (i.e. only on-target inhibition), and $-100$ representing the opposite (i.e. only off-target inhibition).

### *2.3 USING KINHIBITION*

The KInhibition app is run entirely in-browser, and does not require the user to upload or download any data or files. It can be found at https://kinhibition.fredhutch.org/. Upon loading the app webpage, users select a kinase or group of kinases they wish to inhibit (Figure 2.1). Kinase names are standardized across all datasets, as described towards the bottom of the "Datasets" tab. After selecting the kinase(s), the "Table of Results" tab will automatically update to list the inhibitors in the first available dataset, sorted based on their KInhibition Selectivity Score for the chosen kinase(s). A set of radio buttons will also appear to allow the user to choose between all


**Figure 2.1. Using KInhibition.** Users first select a kinase or group of kinases that they wish to inhibit, and then select a dataset in which to search for compounds. The table of compounds updates and sorts compounds based on the calculated KInhibition Selectivity Score. Clicking on a row opens a second table below, which displays off-target effects of that compound. The Heatmap tab displays the full inhibition profiles across all kinases for the compounds displayed in the first page of the Table of Results. The table, and consequently the heatmap, can be reordered by clicking on any of the column headers**.**

of the datasets that include their chosen kinase(s). The data from each dataset are kept separate and must be searched one at a time, as each screen is done using different experimental conditions, kinase panels, and compound doses, and thus results may not be comparable enough to simply merge the datasets. Details about each dataset can be found in the "Datasets" tab (Table 2.1, Figure 2.2). The Table of Results can be searched using the search box in the top right, can display more or fewer compounds per page using the drop-down menu in the top left, and can be sorted based on the values in any of the columns by clicking on the column header. Furthermore, clicking on any row generates a new table below this one which lists the significant off-targets of that compound. "Significant off-targets" are defined as kinases inhibited at least half the amount of the chosen kinase (or half of the geometric mean for multiple chosen kinases).

| Dataset | Compounds Tested | Compound+Dose Combinations | Kinases Screened | Pairwise Coverage | Reference |
|---|---|---|---|---|---|
| Reaction Biology* | 178 | 178 | 300 | 98.9% | Anastassiadis et al., 2011 |
| HMS LINCS | 121 | 134 | 471 | 88.5% | Koleti et al., 2017 |
| GSK PKIS | 367 | 734 | 224 | 99.9% | Dranchak et al., 2013 |
| EMD Millipore | 128 | 255 | 234 | 100% | Gao et al., 2013 |

**Table 2.1. Informational summary of datasets currently included in KInhibition.**

Finally, after exploring the Table of Results, users can click on the "Heatmap" tab to load a heatmap of the inhibition profiles for the compounds displayed in the Table of Results. The compounds included in the heatmap will mirror those in the first page of the Table of Results, including any changes to the number or order of compounds in this table. Inhibition is represented as a color spectrum from black (no inhibition) to yellow (moderate inhibition) to red (maximal inhibition) (Figure 2.1). Users can mouse over the heatmap to view the details of any individual point, or click and drag to selectively zoom in on a portion (double clicking zooms out to the full heatmap). This heatmap, along with all of the previously mentioned tables, can be downloaded using buttons below each element.

### *2.4 Vignette: Finding a Selective Inhibitor of EGFR and ErbB Family Kinases*

To begin, navigate to https://kinhibition.fredhutch.org in a modern desktop browser (while mobile browsers may load the page, some functionality may be lost). Chrome, Firefox, and Safari have all been tested for compatibility. On loading, the sidebar on the left will have two entry fields, "Kinase(s) of Interest" and "Dataset", as well as brief instructions for use. The

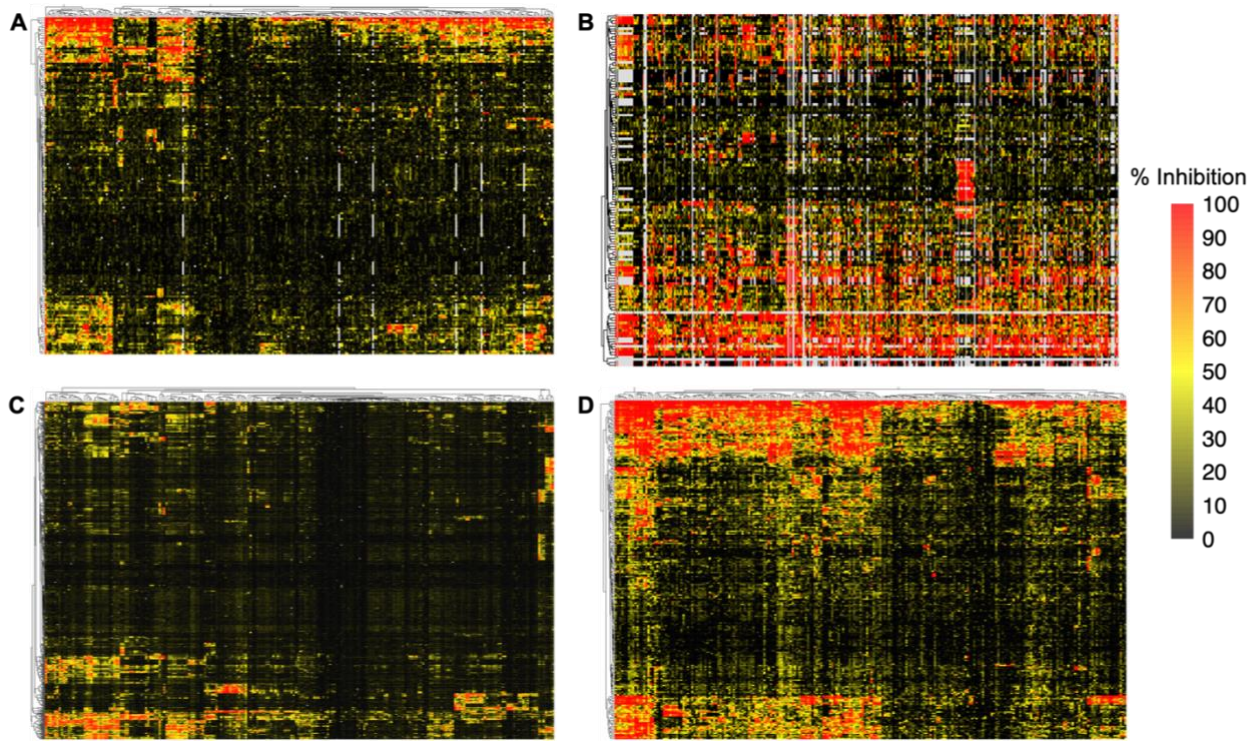


**Figure 2.2. Heatmaps showing the inhibition profiles of all the compounds in each dataset.** Rows represent the compounds tested, columns represent the kinases screened. White boxes denote missing data. **(A)** Reaction Biology dataset **(B)** HMS LINCS dataset **(C)** GSK PKIS dataset **(D)** EMD Millipore dataset.

first step is to select a kinase or group of kinases to inhibit. In this vignette, we will search for inhibitors of EGFR and other ErbB family members. Click on the kinase selection field (labeled with the placeholder text "Select Kinase(s) of interest…"). This will open a drop-down menu with all of the kinases profiled in the chosen dataset. Select "EGFR" by either scrolling down to find it, or by typing the first few letters into the field to filter the results, then clicking on "EGFR" in the drop-down menu. Note that you must click the option even if you have typed the entire name into the field in order for the tool to register your selection. Once EGFR is selected, the Table of Results to the right will automatically populate with the top ten selective inhibitors from the first available dataset (in this case, "Reaction Biology"), initially sorted by KInhibition Selectivity Score (KISS). Additionally, a set of radio buttons will appear to allow you to select any dataset that includes the selected kinase(s) (descriptions of each dataset can be found by

navigating to the “Datasets” tab). For now, let us keep “Reaction Biology” selected. If it has changed, click on the “Reaction Biology” button.

In the Table of Results, we see that the top result is “EGFR Inhibitor”, with a KISS of 98.72, followed by PD 174265 with a KISS of 88.37. To the right side of the table, we notice that PD 174265 has one off-target effect that is greater than half of its inhibition of EGFR, while EGFR inhibitor has none. Note that you may need to scroll to the right to see the whole table, depending on the size of your browser window. To check this off target effect, click on the second row of the Table of Results. This will open a second table below the “Table of Results” that displays the off-target results of the selected compound (you may need to scroll down to see this table). Here, we see that PD 174265 also has significant inhibition on ErbB2 that is considered “off-target” and is lowering its KISS. To get a better idea of the full inhibition profiles of these compounds, click on the “Heatmap” tab towards the top of the screen. Note that the heatmap may take a few seconds to load, and may require scrolling to the right to see the entire image. From this heatmap, we can see that EGFR Inhibitor is, indeed, highly selective for EGFR at 0.5 μM. The white tiles represent areas of missing data, so we can see that EGFR Inhibitor was not profiled for 6 kinases in this dataset. If we hover over these tiles, a pop-up window reveals that these kinases are (from left to right) EPHA6, MYO3B, PRKACG, STK38, TLK2, and ULK3.

From the heatmap, we can see that Lapatinib (row 3) also has high inhibition of EGFR, but also inhibits two other kinases. Hovering the mouse over these bright red tiles reveals that these are ErbB2 and ErbB4. To modify our search to include these as on-target inhibitions, first navigate back to the “Table of Results” tab. Now, click on the “Kinase(s) of Interest” field on the left again and, without removing EGFR, add “ERBB2” and “ERBB4” the same way as before

(either by scrolling to find them, or typing the first few characters, then clicking on the desired kinase). With these selected, we see that Lapatinib is now the top result with a KISS of 96.31. Use the drop-down menu to the top-left of the table to change “Show 10 entries” to “Show 20 entries”. Towards the bottom of this table, we see that “EGFR Inhibitor” now has a KISS of 10.58 due to its low inhibition of ErbB2 and ErbB4. If we desire an inhibitor that has higher inhibition of ErbB4 specifically, we can click on the column name “ERBB4 % Inhibition” twice to reorder the compounds based on their inhibition of ErbB4. The top result, Bosutinib, does not appear very selective (KISS = 27.42). However, the second result is a compound named “EGFR/ErbB2/ErbB4 Inhibitor” (KISS = 87.50). Once again, we can click on this row to generate a table of the 7 notable off-target effects.

To get a holistic view of which inhibitor to use, it is a good idea to search multiple datasets. With the same group of kinases selected (EGFR, ERBB2, ERBB4), click on the “HMS LINCS” button under the “Datasets” header on the left. When the table loads, we see that, in this dataset, Lapatinib (10μM) has a KISS of 88.37. Clicking on the second row to display the off-target effects of Lapatinib (10μM) shows that this dataset also includes many mutant forms of EGFR that are significantly inhibited by Lapatinib. Since these are counted as off-target kinases, they will lower the KISS. If we wish to get a better idea of Lapatinib’s score, we can add each of these mutant forms to our “Kinase(s) of Interest). If we wish to search for a particular subset of compounds, we can type “AZ” in the search field to the top right of the table to restrict the compounds displayed to only those with “AZ” in their names. After doing this, clicking on the “Heatmap” tab will display a heatmap of only these compounds. Finally, all of these outputs (Table of Results, off-target effect table, and heatmap) can be downloaded by clicking on the buttons below each object.

## *2.5 DISCUSSION AND LIMITATIONS*

Data on kinase inhibitors is a lucrative asset in both academia and life science industries. With the above functionality, KInhibition fills a much-needed role in modern cell biology by allowing researchers to make data-driven decisions regarding kinase inhibitors. By following the above steps, researchers can easily find and choose the most selective and appropriate compound for their particular target or pathway. Due to the robust and careful design of the method, this platform can be easily updated and expanded to include additional datasets and information as they become available. I therefore expect this portal to see broad use and adoption akin to other selection tools.

The KInhibition platform, and the associated KInhibition Selectivity Score (KISS) were designed to best leverage the presently available data. However, it should be noted that the KISS and all other metrics listed in this tool are based only on a single dose of the compound used. The efficacy, selectivity, and off-target effects of a given compound depend heavily on the concentration used in the actual experiment, as well as the biological context in which the compound is applied. Thus, the information presented in the Table of Results (i.e. Percent Inhibition) may not directly translate to a cellular or *in vivo* context. Therefore, the goal of this portal, and the datasets included in it, is to obtain a qualitative assessment of selectivity, using the quantitative metrics presented as data-driven guidelines for decisions making in the context of past experience and other pharmacologic properties of the compounds in question (i.e. solubility, bioavailability, metabolism, etc.).

# Chapter 3: Computational modeling of kinase signaling predicts effective therapies for metastatic, castration-resistant prostate cancer

A version of the following work has been submitted to *Nature Communications*:

Bello T, Paindelli C, Diaz-Gomez LA, Melchiorri A, Mikos AG, Nelson PS, Dondossola E, Gujral TS. Computational modeling of kinase signaling predicts effective therapies for metastatic, castration-resistant prostate cancer. (With editor).

## *3.1 INTRODUCTION*

Multiple studies have demonstrated that most KIs, even ones used clinically, display some level of promiscuity in their targets.[39,41,42] While this can increase the potential for toxicity and side effects, this polypharmacology may also reduce the chance of *de novo* resistance by simultaneous inhibition of multiple parallel signaling pathways that could otherwise act as compensatory growth signals, as seen in the multitude of trials combining multiple KIs to target the same or parallel pathways.[34] It also allows for the repurposing of already-approved molecules in different disease contexts where genetic drivers may differ, but the key signaling nodes remain the same.[35,71–73] Finally, the recent approval of Midostaurin, a broad-acting derivative of staurosporine, for treatment of *FLT3*-mutated acute myeloid leukemia demonstrates that selectivity is not a prerequisite for clinical utility.[74]

Gujral and colleagues previously established an unbiased approach called Kinome Regularization (KiR) that exploits the polypharmacology of KIs to identify the kinases whose activities are most likely contributing to a measured phenotype, as described in chapter 1.[43] In this project, I made modifications to refine KiR and best leverage newly available data. The initial 178 KI profiling has since been expanded to 427 KIs, some at additional doses. Using this

new data, I identified a new optimal set of KIs to test, as well as changed many of the calculations performed in the method to make it more refined and robust.

I sought to test the potential of KiR to predict KIs that target castration resistant prostate cancer (CRPC), a leading cause of cancer-related deaths in US males and a challenge for treatment.[75] Localized, early-stage prostate cancer (PCa) can be successfully treated, with nearly 100% survival at 5 years from diagnosis. The five-year survival rate drops to ~30% in metastatic, late-stage CRPC, where bone represents the most common site for distant growth (84% of metastatic patients).[75,76] Late-stage CRPC exhibits poor responses to current therapeutic strategies and results in progressive disease relapse, pain, and other morbidities, imposing a significant negative social and economic impact.[75] Besides amplification of the androgen receptor (AR), expression of AR splice variants, and stromal-mediated survival, progression may be marked by upregulation of compensatory kinase-mediated signaling pathways.[19,56,57,60,77–84] Here, I pursued an unbiased, systems-based approach to investigate the potential of targeting kinase signaling within CRPC as a means of improving therapeutic outcomes. I identified two multi-target small-molecule inhibitors, PP121 and SC-1, that effectively inhibited CRPC growth across many model CRPC systems. *In vitro* assays demonstrated that both PP121 and SC-1 significantly reduced cellular confluence across 4 CRPC cell lines, as well as abrogating phosphorylation on multiple growth signaling proteins. Both compounds also showed significant growth suppression of *in vivo* subcutaneous xenograft CRPC tumors. Finally, orthotopic xenografts in mouse tibiae showed that, while neither KI inhibited growth of the tumors on their own, both PP121 and SC-1 can dramatically improve survival and decrease tumor sizes when administered in combination with docetaxel, a standard of care chemotherapeutic for CRPC. These results demonstrates the power of our KiR models to make translational predictions for

cancer therapies. The marked ability of these KIs to improve on docetaxel in both tumor response and survival represents a promising new avenue of combining multi-target acting KIs with chemotherapy and revitalizes research into KI combination therapy for late stage CRPC.

### *3.2 UPDATING AND REFINING THE KIR METHOD*

Before applying KiR to the study of CRPC, I made modifications to the pipeline to include more recent data and make the models more robust. The previously published dataset of 178 KIs has since been expanded to over 400 KIs, with some KIs tested at additional doses. For my purposes, different doses of KIs are treated as separate drugs, bringing the total number of entries (KI/dose combinations) to 427. Since this is a much larger set than the previous one, I chose a new optimal set of drugs with which to perform the initial screen. Unsupervised hierarchical clustering of the full inhibition profiles divided these 427 KIs into distinct groups, from which I formed a subset of ~32 KIs by picking one from each group so as to maximize the overall coverage of the profiled human kinome.[42] These ~32 computationally chosen KIs formed the initial screening set from which the KiR models were built. As before, the residual activities of the 298 human kinases in the presence of these ~32 selected inhibitors were used as the explanatory variables for elastic net regularized multivariate linear regression, where the response variable is a quantitative cellular phenotype (in this study, cellular confluence).[42,43,46,47,85]

To make models more consistent across runs, I began every model with series of 300 regularization penalties ($\lambda$), beginning at 0.005 and geometrically distributed. The optimal regularization penalty was selected via leave-one-out cross validation (Figure 3.2C).[43,46,47] To get a more comprehensive view of the potential models, I generated 11 total models for each run,

one for each alpha value between 0 and 1.0 (evenly spaced; 0.1, 0.2, etc.). I also allowed the input of a "control" observation, where the residual activity of each kinase is set to 100%. This control was used to scale the phenotypic responses so that the control has a response of 1.0, making the coefficients more standard across models.

Finally, I added an optional setting that allowed for the selection of a regularization penalty ($\lambda$) that is not the absolute lowest MSE in the cross-validation. This is helpful in cases where the lowest MSE produces a model with little to no kinases that have non-zero coefficients. The method worked by identifying two new types of "points of interest": 1) local minimums, identified using step-wise derivatives of a window-averaged MSE curve; and 2) points where the slope of the MSE curve matches a user-specific fraction (e.g. 75%) of the average slope of the entire curve. The former is useful to force the model to select points that are still locally optimal from a cross-validation standpoint, while being more useful by having a more desirable number of non-zero coefficients. The latter is useful specifically in the case of a monotonically increase MSE, where the global minimum is the maximum penalty, resulting in a null model. This setting allows the user to define a particular point where the MSE is still larger than the minimum, but has not yet begun a typical exponential increase and is thus still at acceptable levels. Users can force models to choose any of these sub-optimal penalties by specifying a minimum and/or maximum number of non-zero coefficients (strangely referred to as "degrees of freedom" by glmnet). The program will then identify all possible points of interest (including the global minimum that is normally the default), and select the point with the lowest MSE that satisfies these restrictions. While I caution overruse of this functionality, it nevertheless provides much more flexibility and can lead to biological insights that would otherwise be missed.

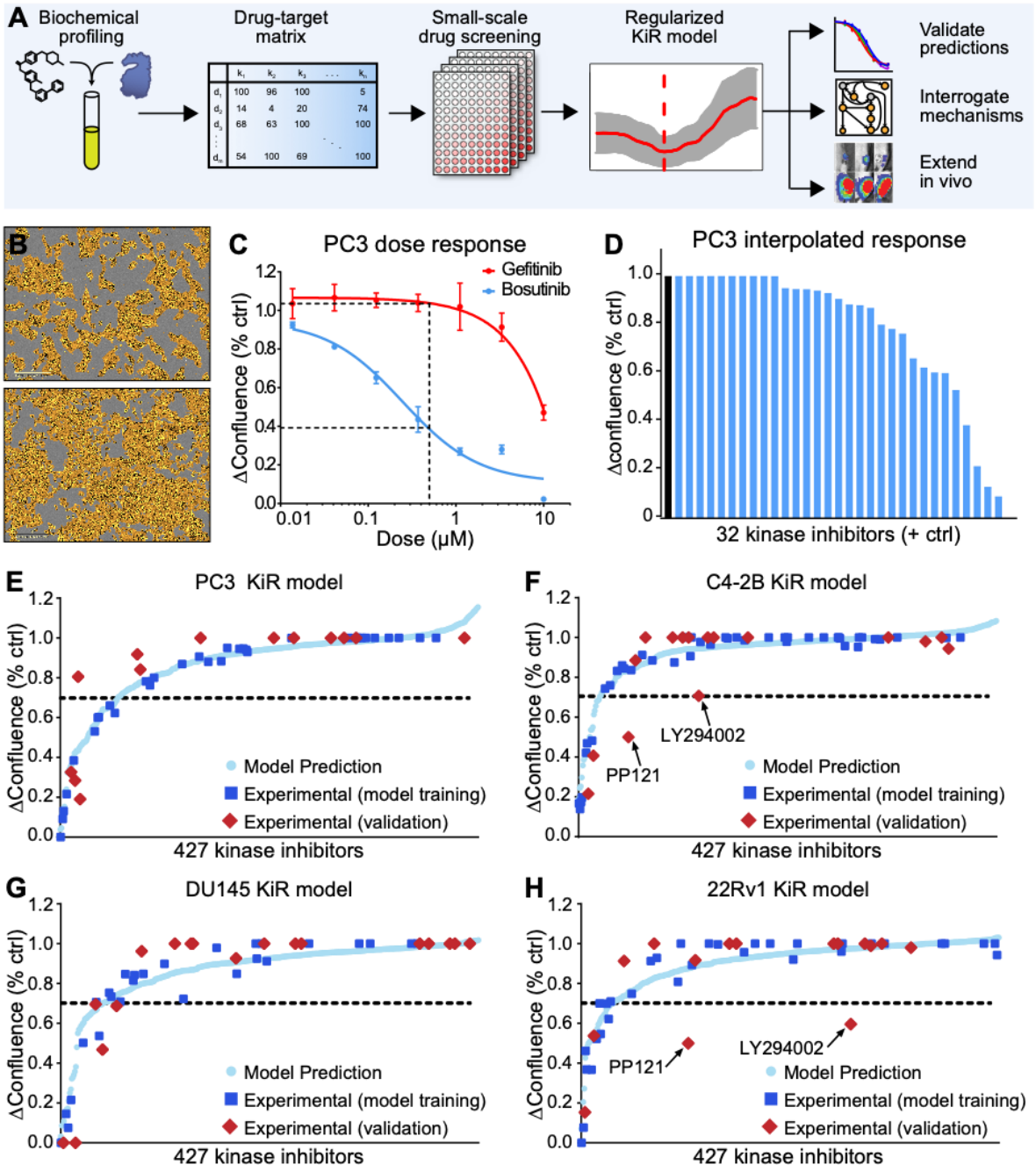


**Figure 3.1.** *Figure legend on following page.*

### *3.3 KINASE INHIBITOR FUNCTIONAL SCREEN LEADS TO PREDICTIVE MODELS OF CRPC GROWTH*

To begin predicting suitable KI candidates for CRPC treatment, I performed a functional KI screen in four CRPC cell lines: PC3, C4-2B, DU145, and 22Rv1 (Figure 3.1A). These lines have been used extensively in prostate cancer research and account for a variety of biological phenotypes, including AR expression and tissue of origin (C4-2B and 22Rv1 exhibit AR activity,

**Figure 3.1. Constructing predictive KiR models from an in vitro kinase inhibitor screen. (A)** Schematic of KiR workflow. Previous biochemical profiling of kinase inhibitors resulted in a quantitative drug–target matrix. From this, an optimal set of ~32 KIs are used for a small-scale drug screen. The results enable the generation of a regularized, cross-validated model that predicts the responses to over 400 KIs. **(B)** Representative images of PC3 cells after 52 hours of treatment with Bosutinib at 3.33μM (top) or 0.01μM (bottom). Cell confluence mask (shown in orange) was drawn and quantified via IncuCyte ZOOMTM. Top image was quantified at 45.6% confluent, bottom image at 72.6% confluent. **(C)** Representative dose-response curves of PC3 treated with Bosutinib or Gefitinib. Curves are fitted with a 3-parameter logistic equation, and responses were interpolated at 500 nM. Mean ± SEM, n = 3 wells/dose. **(D)** Interpolated responses of PC3 cells to the tested inhibitors. **(E–H)** KiR model plots for four CRPC cell lines showing initial measured values (dark blue), the resulting KiR model predictions (light blue), and measured validations not included in the model training set (red). Poorly-predicted inhibitors that target PI3K (which is not profiled as a variable in the linear model) are labeled. Dotted line indicates the threshold for an inhibitor to be considered effective.

while PC3 and DU145 do not; PC3 and C4-2B are derived from bone metastases, DU145 from a brain metastasis, and 22Rv1 from a primary tumor).[86,87] In our assay, each line was seeded in 96-well cell culture plates and treated with the aforementioned ~32 KIs at 7 doses (plus untreated control) ranging from 10μM to 0.01μM, in triplicate for each dose. This dose range captures the full response range for nearly all inhibitors, with the lowest dose rarely differing from the controls for any inhibitor, and the highest dose displaying near maximal potency for most drugs without exhibiting significant non-specific effects (such as DMSO vehicle toxicity). Images of each well were taken every 2 hours using the IncuCyte ZOOM$^{TM}$ imaging system for 48-96 hours.[88] Cellular confluence was quantified using the IncuCyte ZOOM$^{TM}$ software to identify a cell mask and divide the area covered occupied by cells by the total image area, as described previously (Figure 3.1B).[89] To account for varying growth rates between cells, I calculated the endpoint response as the net change in confluence (Δconfluence) between the first image post treatment and the time point when the control samples within the same plate reached 75% confluence (Figure 3.2A). I generated dose-response data for each cell line–inhibitor combination, fitted the data with a three-parameter logistic equation, and interpolated the resulting curve at the doses for which the compounds were profiled (generally 500 nM) (Figure

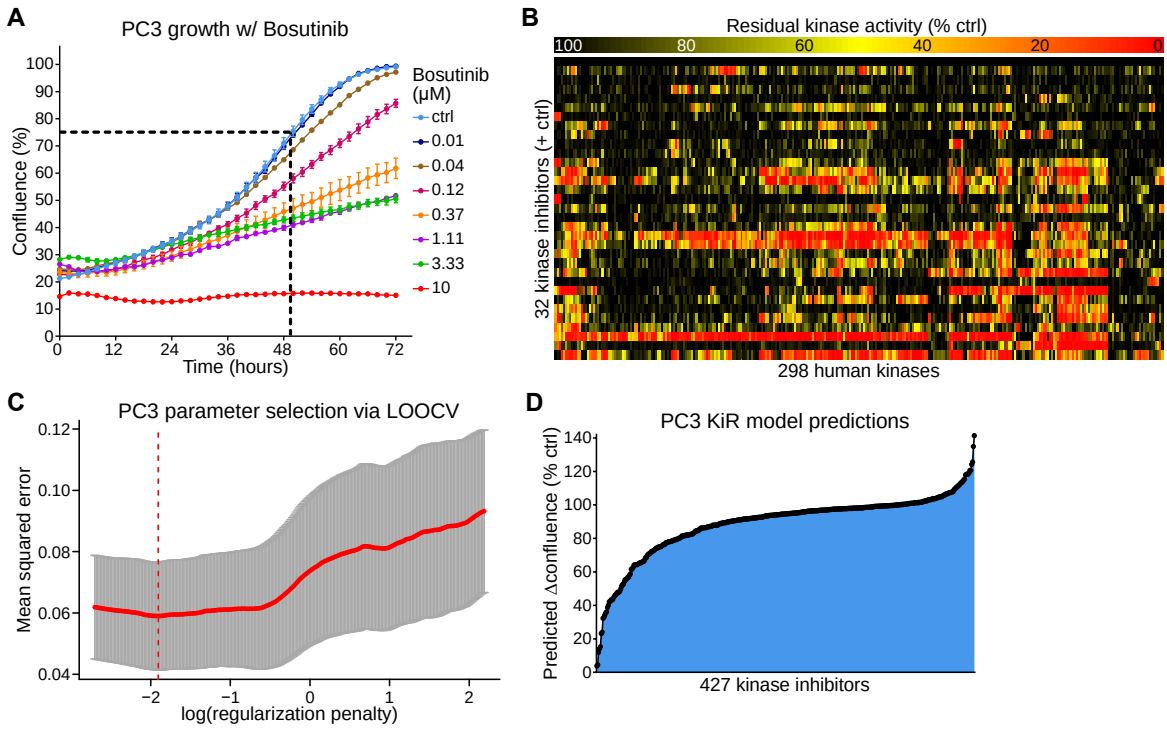


**Figure 3.2. Generation of final PC3 KiR model. (A)** Time course growth of PC3 cells treated with various doses of Bosutinib. Confluence (percentage of total area covered by cells) is quantified by the Incucyte ZOOMTM software. Dotted line shows time point where untreated control cells reach 75% confluence. Data presented as mean ± SEM, n = 3 wells per condition. **(B)** Heatmap of inhibition profiles for the 32 inhibitors tested on PC3 cells. Data represents residual activity of the kinase in the presence of the inhibitor as a percent of uninhibited activity, where 100 represents no inhibition (100% kinase activity). **(C)** Leave-one-out cross validation (LOOCV) plot for the PC3 elastic net model. Horizontal axis displays the log of the penalty hyperparameter λ, vertical axis shows the mean squared error from the LOOCV (mean ± SD, n = 32 folds). Dotted red line indicated the lowest error and the selected value of λ for the final model. **(D)** KiR model predictions for the response (Δconfluence, % ctrl) of PC3 cells to all 427 profiled kinase inhibitors.

3.1C–D).[42] At these doses, roughly 30% of molecules tested in each cell line displayed significant confluence reduction (at least 30% reduction in Δconfluence compared to control), and nearly 30% of all tested compounds displayed no significant effect. C4-2B was the most resilient line, with only 8 out of 43 initially tested compounds reducing Δconfluence by 30% or more. Four inhibitors in the initially tested sets were effective in all four cell lines: Cdk1/2 Inhibitor III, Staurosporine, PDK1/AKT/FLT Dual Pathway Inhibitor, and Dasatinib. The first three of these are very broad-acting kinase inhibitors that tend to inhibit most model systems,

while Dasatinib is a broad-acting SRC family kinase inhibitor that has been previously noted to inhibit PCa growth in many model systems .[17,90,91]

From the ~32 inhibitors that we tested in each cell line, I generated KiR models using elastic net regularized linear regression as described above. The interpolated Δconfluence for each KI was used as the response variable. The resulting linear models had normalized root mean squared errors (RMSE) between the predicted responses and the measured data between 0.0291 for PC3 and 0.09191 for DU145, giving me confidence that the models fit the training set well even with the additional regularization and cross validation (Table 3.1). I used the KiR model's coefficients calculated from the training set to predict the Δconfluence response for all >400 profiled KIs (Figure 3.2D). To experimentally validate these predictions, we tested the responses of each cell line to an additional ~20 KIs not included in the training set, using the same experimental and analysis protocol as before. The measured responses to most inhibitors followed the model predictions, with validation set RMSEs ranging between 0.127 for DU145 and 0.163 for 22Rv1 (Figure 3.1E–H, Table 3.1). As notable exceptions, PP121 and LY294002 inhibited growth of C4-2B by >50% and ~30%, respectively, and of 22Rv1 cells by >50% and ~40%, respectively, despite a lack of predicted efficacy in these cases. Both of these compounds are reported to be inhibitors of phosphatidylinositol 3-kinases (PI3Ks), a lipid kinase that is not profiled in the pairwise catalytic assay and not included as an explanatory variable in the KiR model, which likely acts as a common confounding factor in these models.[32,92,93] Such issues were resolved via careful parameter selection and interpretation of the models. The final, validated KiR models resulted in measured or predicted cellular responses to over 400 KIs across the four CRPC cell lines (Figure 3.1E–H).

| Cell Line | Alpha Value | Cross Validation RMSE (normalized) | Training Set RMSE (normalized) | Validation Set RMSE |
|---|---|---|---|---|
| C4-2B | 0.6 | 0.23869318 | 0.0447 | 0.142332465 |
| 22Rv1 | 0.6 | 0.32513731 | 0.09048 | 0.16263531 |
| DU145 | 0.8 | 0.32770855 | 0.09191 | 0.126522616 |
| PC3 | 0.7 | 0.31523296 | 0.0291 | 0.150294567 |

**Table 3.1. Properties of KiR models for CRPC cell lines.** Cross validation is performed as LOOCV. Normalized RMSEs are calculated by dividing the RMSE by the average of the training set responses, which are scaled to have a pseudo-maximum of 1.00 for untreated samples. Alpha is user defined when generating the model.

### *3.4 Model-predicted compounds PP121 and SC-1 inhibit CRPC growth and signaling*

While I identified several cell-type specific inhibitors, I chose to focus on compounds that were effective in all four cell lines. I defined an "effective" compound as one that had a measured or predicted (if measured was unavailable) Δconfluence at least 30% less than control at the interpolated dose. This cutoff was chosen to maximize the proportion of tested compounds correctly assigned to either above or below this threshold while still maintaining a meaningful partition. A total of eight compounds met this criterion (Figure 3.3A). Pearson correlations of these compounds' target profiles showed they generally fell into four clusters (Figure 3.3B). The first cluster included three broad-spectrum inhibitors (CDK 1/2 inhibitor III, Staurosporine, and PDK/AKT/FLT dual pathway inhibitor) that we know to inhibit cellular processes in most model systems (data not shown). Using my previously-published KI selection portal, KInhibition, I quantified the selectivity of these compounds towards various receptor tyrosine kinases (RTKS) and signaling nodes (growth factor receptors, Src family kinases, Abl, etc.) via the KInhibition Selectivity Score (KISS). The KISS is a value between -100 and 100 that summarizes the on-target vs off-target effects of a KI.[65] A KISS of 100 represents full on-target inhibition and no

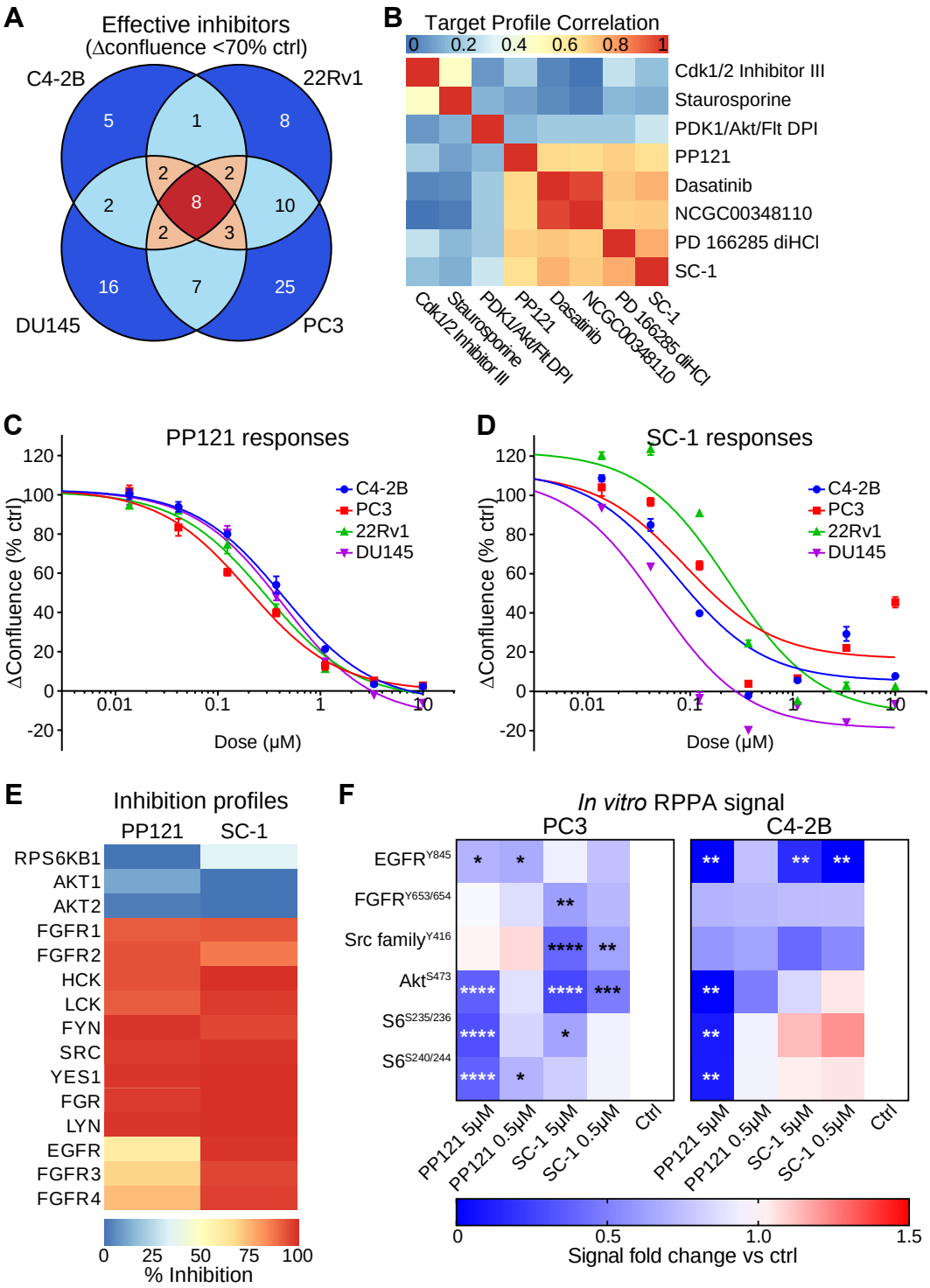


**Figure 3.3.** *Figure legend on following page.*

off-target effects (among the kinases profiled in the underlying dataset), while a KISS of -100 represents the opposite. CDK 1/2 inhibitor III, Staurosporine, and PDK/AKT/FLT dual pathway inhibitor had KISS values of -24.49, -12.65, and -22.47, respectively, reinforcing their

**Figure 3.3. PP121 and SC-1 effectively target CRPC cells *in vitro*. (A)** Euler diagram of the effective inhibitors in all four cell lines. Eight inhibitors reduced the change in confluence to <70% of untreated control in all four cell lines. **(B)** Pearson correlation heatmap of the inhibition profiles of the eight effective inhibitors from panel (A). **(C–D)** Cell line dose responses to two of the most effective inhibitors, PP121 and SC-1. Mean ± SEM, n = 3 wells/dose. **(E)** Inhibition profiles of PP121 and SC-1 towards receptors EGFR and FGFR, and downstream signaling molecules Src family kinases, Akt, and S6 kinase. **(F)** Heatmaps of RPPA phosphorylation signals when PC3 and C4-2B cells are treated with PP121 or SC-1 in vitro. Signal is normalized to β-actin and presented relative to DMSO control. *$p$<0.05, **$p$<0.01, ***$p$<0.001, ****$p$<0.0001, two-way ANOVA with Holm-Sidak multiple comparisons test.

promiscuity, and were thus excluded from further analysis. Dasatinib (KISS of -11.07) has already been extensively studied for its noted efficacy in PCa, so we chose not to pursue this compound and a very similar compound (NCGC00348110) despite their notable *in vitro* results.[17,78,90,91,94,95] Of the remaining two clusters, PP121 (KISS of 15.49) seemed the most unique, and SC-1 (KISS of 23.45) showed slightly greater growth suppression than the molecule it clustered with (PD 166285 dihydrochloride, KISS of 4.10). Both PP121 and SC-1 showed potent, dose-dependent growth suppression across all cell lines (Figure 3.3C–D), so I chose to proceed with these compounds for validation in additional *ex vivo* and *in vivo* models.

PP121 is a pyrazolopyrimidine small molecule developed to act as a dual inhibitor of tyrosine kinases and PI(3)Ks.[32] It is reported to have a relatively broad effect on PI(3)Ks, many RTKs, and Src-family kinases, but much less inhibition of serine-threonine kinases. PP121 was reported to block proliferation of cancer cells *in vitro*, as well as overcome common resistance mutations via inhibition of multiple kinases. SC-1, also known as Pluripotin, is a 3,4-dihydropyrimido[4,5-d]pyrimidine developed in a screen to identify molecules that promote embryonic stem cell self-renewal.[96] It is reported to inhibit Erk1 (MAPK3) and RasGAP (RASA1) and block differentiation pathways in embryonic stem cells, although its kinase inhibition profile demonstrates broader activity against many tyrosine kinases.[42,96] To further investigate the molecular effects of PP121 and SC-1 on prostate cancer cells *in vitro*, I performed reverse-phase protein array (RPPA). RPPA is a miniaturized dot blot technique that allows screening of

multiple signaling proteins and pathways across many samples.[97] Members of the Gujral lab have previously applied RPPA to uncover the state of signaling pathway in tumor lysates and lysates from cell culture models.[89,98] PC3 and C4-2B cells were treated *in vitro* with 5μM or 0.5μM of PP121 or SC-1 (plus DMSO control). I assayed a small subset of proteins involved in kinase-mediated growth signaling, including EGFR, FGFR, Src family kinases, Akt, and small ribosomal protein S6. All of these proteins have been reported to play key roles in the growth and progression of late stage prostate cancer.[17,19,58–60,62–64,77,99–101] The biochemical inhibition profiles of PP121 and SC-1 suggest these compounds effectively inhibit EGFR, FGFR, and Src family kinase signaling (Figure 3.3E). Neither appear to exhibit direct inhibition of Akt or S6 kinase, but they may still abrogate signaling along this axis via indirect inhibition (such as PP121's reported effects on PI(3)K). RPPA results were quantified as the fluorescent signal intensity from each spot, normalized to the β-actin signal intensity from the same spot and presented relative to untreated controls. The results indicate a strong trend towards decreasing growth factor signaling in both cell lines, with PC3 tumors exhibiting more significant responses (Figure 3.3F). EGFR, Akt, and S6 phosphorylation significantly decreased in both cell lines treated with the higher dose of PP121. SC-1 abrogated Src family kinase and Akt phosphorylation at both doses in PC3 cells, while only significantly affecting EGFR phosphorylation in C4-2B cells. Interestingly, Akt and S6 phosphorylation slightly increased under SC-1 treatment in C4-2B cells, potentially representing the previously-observed compensatory growth signaling through the PI(3)K-Akt-mTOR axis.[62,81,99,101] Overall, the trend of decreasing phosphorylation signals in both PC3 and C4-2B under PP121 and SC-1 treatment provides a preliminary molecular basis for these compounds' inhibition of CRPC growth.

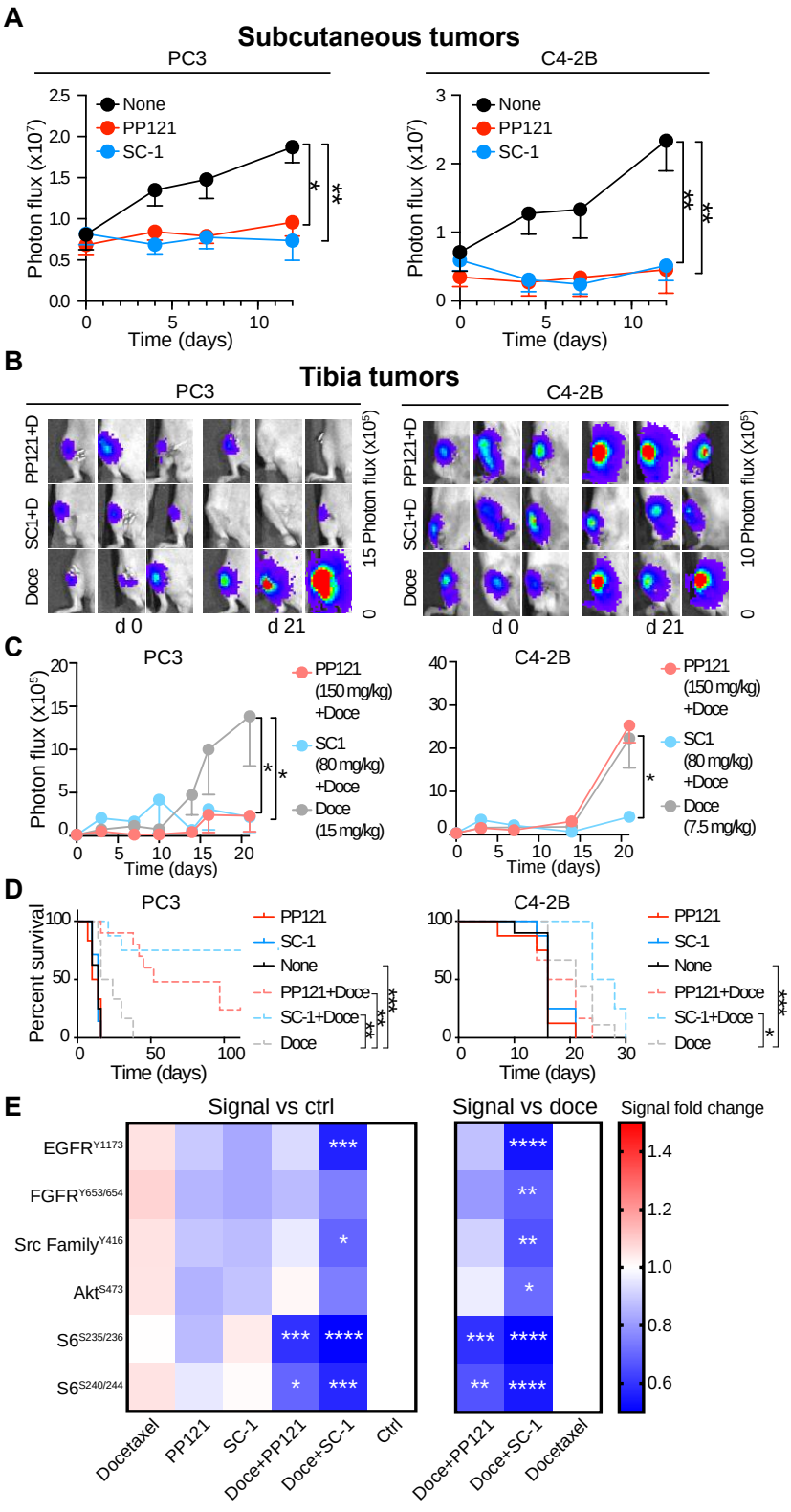


**Figure 3.4.** *Figure legend on following page.*

**Figure 3.4. PP121 and SC-1 inhibit tumor growth *in vivo*. (A)** Growth of PC3 or C4-2B tumors implanted subcutaneous treated with either PP121 or SC-1. Response to KI treatment was assessed longitudinally by monitoring the bioluminescence signal. **p*<0.05, ***p*<0.01, one-way ANOVA followed by Tukey's HSD post-hoc test. **(B–C)** Growth of PC3 or C4-2B tumors implanted via intra-tibial injection and treated with combination KI+docetaxel therapy. Shown as mean - SD, n = 8–10 tibiae per group. **p*<0.05, one-way ANOVA followed by Tukey's HSD post-hoc test. **(D)** Overall survival was estimated over time. **p*<0.05, ***p*<0.01, ****p*<0.001. **(E)** Heatmap of RPPA signal from short-term in vivo PC3 tibia xenografts treated with KI monotherapy or combination therapy with Docetaxel. Signal is normalized to β-actin and presented relative to untreated control (left heatmap) or docetaxel monotherapy (right heatmap). **p*<0.05, ***p*<0.01, ****p*<0.001, *****p*<0.0001, two-way ANOVA with Holm-Sidak multiple comparisons test. Note: data and figure in panels (A–D) were generated by Eleonora Dondossola and Claudia Paindelli at the University of Texas MD Anderson Cancer Center.

### *3.5 PP121 AND SC-1 INHIBIT CRPC GROWTH IN VIVO*

After the success of PP121 and SC-1 at inhibition CRPC growth *in vitro*, I sought to extend these findings to further preclinical models. Through a collaboration with Eleonora Dondossola and Claudia Paindelli at the University of Texas MD Anderson Cancer Center, we were able to test these KIs in two *in vivo* models. First, mice were implanted with fluorescently-labeled PC3 or C4-2B tumors subcutaneously. The mice were randomly assigned to receive PP121, SC-1, or vehicle control. Both KIs showed significant growth suppression in both tumor types, as assessed by photon flux (Figure 3.4A). As bone represents one of the most common sites of metastasis, my collaborators next tested KI therapies on CRPC cells implanted via intria-tibial injection. Neither PP121 nor SC-1 displayed any growth suppression of tibia tumors when administered as single agents (data now shown). However, when mice were co-treated with these KIs and docetaxel, a standard-of-care therapeutic for late-stage CRPC, they exhibited vastly decreased tumor growth compared to even docetaxel as a single agent (Figure 3.4 B–C). Both PP121 and SC-1 significantly decreased PC3 tumor growth, and SC-1 significantly inhibited C4-2B tumor growth as well. Furthermore, mice implanted with PC3 tumors and treated with the KI + docetaxel combination therapy displayed significantly longer survival times than either untreated controls or mice treated only with docetaxel, with tumor-free mice up to 120 days

(Figure 3.4D). Mice implanted with C4-2B tumors did not respond to PP121 as a single agent or in combination, but they did exhibit significantly increased survival when treated with SC-1 + docetaxel, even compared to docetaxel alone. Taken together, these results demonstrate the ability of KI therapy to successfully improve therapeutic outcomes of CRPC in an *in vivo* system.

To better understand the molecular basis for these observations, I again performed RPPA on treated samples. Tumors were implanted into tibiae of mice, allowed to grow for 3 days, and treated for 3 additional days with KI single or combination therapies. We chose to focus only on PC3 tumors as they exhibited the strongest phenotypic response across treatments. The tumors were lysed, homogenized, and subjected to RPPA, as described previously. Interestingly, none of the tumors treated with KIs or docetaxel as single agents exhibited any decreases in the phosphorylation signals assayed (Figure 3.4E). However, PC3 tumors treated with SC-1 in combination with docetaxel displayed significantly decreased levels of phosphorylated $EGFR^{Y1173}$, $FGFR^{Y653/654}$, $Src^{Y416}$, $Akt^{S473}$, $S6^{S235/236}$ and $S6^{S240/244}$ (fold changes of 0.54, 0.69, 0.66, 0.71, 0.50, and 0.55 compared to docetaxel-only treated tumors, respectively). Tumors treated with PP121 showed significantly decreased levels of phosphorylated $S6^{S235/236}$ and $S6^{S240/244}$ (fold changes of 0.60 and 0.66 compared to docetaxel-only treated tumors, respectively). These molecular results align with the observed phenotypes, wherein the combination therapy of SC-1 and docetaxel elicits an extremely strong response, with sharply decreased tumor size, markedly increased lifespans, and lower levels of canonical kinase-mediated growth signals. Taken together, these results provide promising evidence for the ability of KIs to improve therapeutic outcome in late-stage CRPC.

### *3.6 DISCUSSION AND LIMITATIONS*

In this chapter, I explored the role of kinase signaling in PCa by taking a combined chemical and systems biology approach. I began with a functional screen employing multiple computationally-selected, multi-targeting KIs. I mathematically combined my observations with data from others' previous work to generate regularized linear models that predicted the effectiveness of hundreds of untested KIs against CRPC. Importantly, neither PP121 nor SC-1 were included in the training set for the final KiR models, yet they still emerged as promising hits and proved effective even in more complex model systems. Thus, by computationally integrating complementary data types, I was able to significantly improve the physiological relevance and predictive power of an *in vitro* model system. Of note, the KiR models predicted two inhibitors (PP121 and LY294002) to be ineffective in both C4-2B and 22Rv1 cell lines, yet PP121 inhibited growth of both cell lines by more than 50%, and LY294002 inhibited the growth of C4-2B by ~30% and 22Rv1 by ~40%. Both of these compounds are known inhibitors of phosphatidylinositol 3-kinases (PI3K), a lipid kinase that is not profiled in the pairwise catalytic assay and therefore not included as an explanatory variable in the KiR model.[92,93] Given the known role of PI3K in late-stage prostate cancer, it is likely that the efficacy of these compounds is mediated, in part, by PI3K inhibition, which would explain the disparity between the predicted and observed effects.[62,64,81] I was able to circumvent this limitation of the KiR approach via careful variable selection and additional validation. It is worth noting that this trend only existed in the AR-positive cell lines, which is consistent with prior findings of significant cross-talk between the PI3K and AR signaling pathways in CRPC; however, further studies are needed to

completely elucidate the role AR signaling may have in modulating therapeutic response to KI treatment.[81,99,102]

The RPPA results from both *in vitro* and *in vivo* studies confirm our prior biochemical knowledge that both PP121 and SC-1 are relatively broad acting kinase inhibitors (Figure 3.3E–F, Figure 3.4E). This indicates that their effect may not be mediated through a single kinase or pathway, but through simultaneous multipronged effects, lending further credence to the benefits of KI polypharmacology. The fact that these molecular changes only occurred *in vivo* when KI therapy was combined with docetaxel implies there is an additional confounding factor, particularly in the bone microenvironment, that complicates KI efficacy. Further studies are needed to optimize the pharmacokinetic and biochemical properties KI therapy *in vivo*, particularly in cases of treatment resistant and/or metastatic disease. Nevertheless, these results demonstrate a possible synergy between KIs and chemotherapeutics for tumors in the bone environment that warrants further investigation and consideration during experimental drug development.

There are several remaining questions raised by our study, including why C4-2B tumors implanted in mice did not respond to PP121 + docetaxel and demonstrated far less detectable molecular responses to SC-1 + docetaxel than PC3 tumors. C4-2B cells exhibit AR activity, the canonical PCa driver, while PC3 cells do not, which may account for the differential sensitivity.[87] Future preclinical investigation could assess a response in combination with an AR targeted therapy, such as Enzalutamide or Abiraterone; or via biomarker-guided therapy, wherein KI combination therapies are reserved for patients with AR-negative tumors. As Bluemn and colleagues recently noted, in the time since second-generation AR-targeted therapies (enzalutamide and abiraterone) have been approved, the proportion of AR-null / neuroendocrine-

null patients rose from 5.4% to over 20% of PCa patients, resulting in a new therapeutic niche with few available options.[60] Here, my collaborators and I demonstrated a promising proof of concept for employing multitargeted KI therapy in combination with chemotherapy to address this emerging patient population.

## *3.7 Materials and Methods*

### Cell culture

PC3, 22Rv1, and DU145 cells were purchased from the American Type Culture Collection (Manassas, VA). C4-2B cells were purchased from ViroMed laboratories (Minnetonka, MN). PC3, C4-2B, and 22Rv1 cells were cultured in RPMI media supplemented with 10% fetal bovine serum, 1% penicillin / streptomycin, and 1% sodium pyruvate. DU145 cells were cultured in DMEM media supplemented with 10% fetal bovine serum and 1% penicillin / streptomycin. PC3 DC cells expressing nuclear H2B-GFP and cytoplasmic DsRed2 were from Anticancer; PC3 luciferase expressing cells were a gift of Dr. Gary Gallick from UT MD Anderson Cancer Center. Cells were cultivated in DMEM (Corning) supplemented with 10% fetal bovine serum (FBS, Sigma) and Penicillin-Streptomycin (100µg/ml each, Sigma). C4-2B cells (a gift of Dr. Timothy Thompson from UT MD Anderson Cancer Center) expressing nuclear H2B-mCherry and cytoplasmic GFP-LifeAct or luciferase were cultivated in RPMI (Corning) supplemented with 10% fetal bovine serum (FBS, Sigma), Penicillin-Streptomycin (100µg/ml each, Sigma) and 1% HEPES. The identity of cancer cell lines was verified by the "Characterized Cell Line Core Facility", M.D. Anderson Cancer Center, through Short Tandem Repeat DNA profiling. Adipose tissue-derived mesenchymal stem cells (hMSCs, ASC52telo telomerase reverse transcriptase immortalized from ATCC) were cultivated in *MEM 1X* (Corning) supplemented

with 17% fetal bovine serum (FBS, Sigma), Penicillin-Streptomycin (100µg/ml each, Sigma), vitamins (Sigma), non-essential amino acids (Sigma), and sodium pyruvate (Gibco). To induce differentiation towards the osteoblastic lineage, hMSCs were kept in osteogenic induction medium (DMEM 1X, supplemented with 10% fetal bovine serum, penicillin and streptomycin, 50µg/mL ascorbic acid,10 mM β-glycerophosphate, 0.1 µM dexamethasone from Sigma).

**Drug response and quantification**

Unless otherwise specified, all drug response data was collected by plating cells in 96-well tissue culture treated plates, allowing cells to adhere overnight, and treating with small molecules suspended in DMSO. Drug dilutions were made using threefold serial dilutions in the appropriate media for each cell type, and each dose was administered to the cells in triplicate. Microscopy images were taken with an Incucyte using the 10X objective. Confluence as percentage of total area covered by cells was quantified using the Incucyte ZOOM 2016B software.

**Computational analysis and modeling**

Unless otherwise specified, all dose-response plots, fitted curves, and interpolations were performed using Graphpad Prism version 7.03. Elastic net KiR models, predictions, and cross-validation plots were generated in R Studio version 1.0.153, running R version 3.4.1 "Single Candle", using custom scripts that employ the "glmnet" package (Supplementary File S5).[46] The input parameters to the *cv.glmnet* command were as follows: the residual kinase activities of the KIs test as x-values, the Δconfluence of the corresponding KIs (as fraction of untreated control) as y-values, no standardization of x-values, uniform observation weights and variable penalties, system-generated lambda sequence of 300 lambdas and a minimum ratio of 0.005, gaussian (quantitative) family response type with naive algorithm, number of cross-validation folds equal to the number of observations (the length of the response vector; LOOCV), and alpha values

ranging from 0.0 to 1.0, incrementing by 0.1 (the command is run iteratively for each value of alpha). Additional calculations and analyses were performed in Microsoft Excel 365. Heatmaps were generated either in R Studio using the "pheatmap" package[103], or in Graphpad Prism.

**Reverse-phase protein array**

Protein microarrays were printed and processed as described previously.[89,98] Briefly, PC3 or C4-2B cells were plated in 6-well plates at 500,000 cells/well and allowed to adhere overnight. The following morning, the cells were treated with PP121 or SC-1 at 2 doses (5μM and 0.5μM, plus DMSO control) for 1 hour. Cells were then washed twice in PBS and cell lysates were prepared in 2% SDS lysis buffers as described previously.[89,98] Whole cell lysates were printed onto 16-pad nitrocellulose coated slides (Grace Biolabs, GBL505116) using Aushon 2470 microarrayer (Aushon BioSystems). Each sample was printed in triplicate and slides were stored at - 20°C until processing. RPPA slides were washed with 1 M Tris-HCl (pH 9.0) for 2–4 days to remove SDS. Slides were then washed 2–3 times with PBS for 5 min each and blocked with Odyssey Blocking Buffer (OBB) (Licor, 927-40000) for one hour at room temperature. After blocking, arrays were incubated with primary antibodies in OBB at 4°C overnight. Next day, arrays were washed thrice with PBS and incubated with IRDye labeled secondary antibodies in OBB for 1 hr at room temperature. Following incubation, slides were scanned using Licor Odyssey CLX Scanner (LiCOR). Total signal intensity from each spot was quantified using Array-Pro analyzer software package (Media Cybernetics). The measurement of a specific protein from an individual sample was normalized to total β-actin (Sigma, A1978).

In vivo samples: tumors were surgically extracted from mice and kept cryopreserved until use. Small pieces of the tumors were homogenized and lysed in 2% SDS lysis buffer using a BioGen

PRO200 homogenizer (PRO scientific 01-01200). Samples were then processed as described above.

**Kinase inhibitor treatment *in vivo***

All *in vivo* studies and treatments in this chapter were performed by Eleonora Dondossola and Claudia Paindelli at The University of Texas MD Anderson Cancer Center.

Animal studies were approved by the Institutional Animal Care and Use Committee of The University of Texas MD Anderson Cancer Center and performed according to the institutional guidelines for animal care and handling. Athymic nude and NOD-SCID mice were injected with luciferase-expressing PC3 and C4-2B cells, respectively.

Subcutaneous tumors: Mice were injected subcutaneously with $5 \times 10^6$ PC3 cells/mouse or $6 \times 10^6$ C4-2B subcutaneously, in 30% Matrigel in PBS (n=5-8 mice/group). After 10 days, mice were randomized and treatment started. The growth of C4-2B tumors was monitored through macroscopic bioluminescence imaging using an IVIS 200 imaging system (Perkin Elmer, Waltham, MA). The growth of PC3 tumors was monitored measuring tumor volume using a caliper.

Bone tumors: Mice were injected with $2.5 \times 10^5$ PC3 cells/mouse or $1 \times 10^6$ C4-2B cells intra-tibia (8–10 tibiae/group), randomized after 3 days and treatment started. The growth of tumors was monitored through macroscopic bioluminescence imaging using an IVIS 200 imaging system. In survival studies, photon flux values of $1.2 \times 10^6$ or $2 \times 10^6$ (total counts) were considered as the endpoint for PC3- or C4-2B-implanted mice, respectively.

In vivo treatments: Mice received kinase inhibitor treatment 5 days/week for the following 10 days. Mice were administered a total daily dose of 150 mg/kg of PP121 (Tocris) in 10% DMSO (Sigma), 40% PEG-300 (Sigma), 5% TWEEN-80 (Sigma) and 45% PBS through oral gavage

(100 µl volume/administration, 75 mg/kg twice a day) or 80 mg/kg of SC-1 (Selleckchem) in 10% 2methyl2pirrolidinone (Sigma) and 90% PEG-300 (Sigma) through oral gavage (100µl volume/administration, 40 mg/kg twice a day). For combination therapy, athymic nude mice were administered 15 mg/kg of docetaxel (Sanofi) by intra-peritoneal injection while NOD-SCID mice were given 7.5mg/kg of docetaxel at day 5 after cell injection.

**Statistical analysis**

Statistical analysis was performed using GraphPad Prism 7.0 (GraphPad Software, San Diego, CA). To address the difference among different groups, one-way ANOVA was performed followed by Tukey's honestly significant difference (HSD) post-hoc test. All statistical tests were two-sided, and a *p*-value of less than 0.05 was considered statistically significant. Statistical analysis of RPPA data was done in Graphpad Prism version 7.03 using a 2-way ANOVA with a Holm-Sidak multiple comparisons test.

# Chapter 4: KiRNet: Integrated, kinase-centered network model for investigating kinase-driven phenotypes

A version of the following work has been submitted to *Cell Systems*:

Bello T, Chan M, Golkowski M, Khasnavis N, Ceribelli M, Ong SE, Thomas C, and Gujral TS. KiRNet: Integrated, kinase-centered network model for investigating kinase-driven phenotypes. (With editor).

## *4.1 INTRODUCTION*

In the past few decades, the biosciences have seen an incredible rise in high-throughput technologies and large-scale databases.[104–106] This explosion in the size and availability of prior knowledge necessitates the development of new tools to integrate dispersed and distinct data into actionable, functionally relevant output. In particular, human protein kinases are critical signaling enzymes with numerous available high-throughput molecular approaches.[2,3,5,6,33] Nevertheless, it remains challenging to construct a cohesive understanding of the kinome signaling landscape driving a particular phenotype from these vast and disparate sources of data. To address this need, I developed KiRNet, a flexible, systems-based method for integrating responses from a kinase-centered functional screen with prior knowledge.

KiRNet builds on our previously established, computationally-enhanced functional screen Kinome Regularization (KiR) that leverages the polypharmacology of kinase inhibitors to identify essential kinases that mediate a measured phenotype.[43] The result of KiR is a list of "hits" (in this case, kinases) that are predicted to act as key mediators of the given cellular function. However, the "hits" from KiR and other functional screens are often disjoint and disordered, making it difficult to identify the best direction for further studies. To transform these

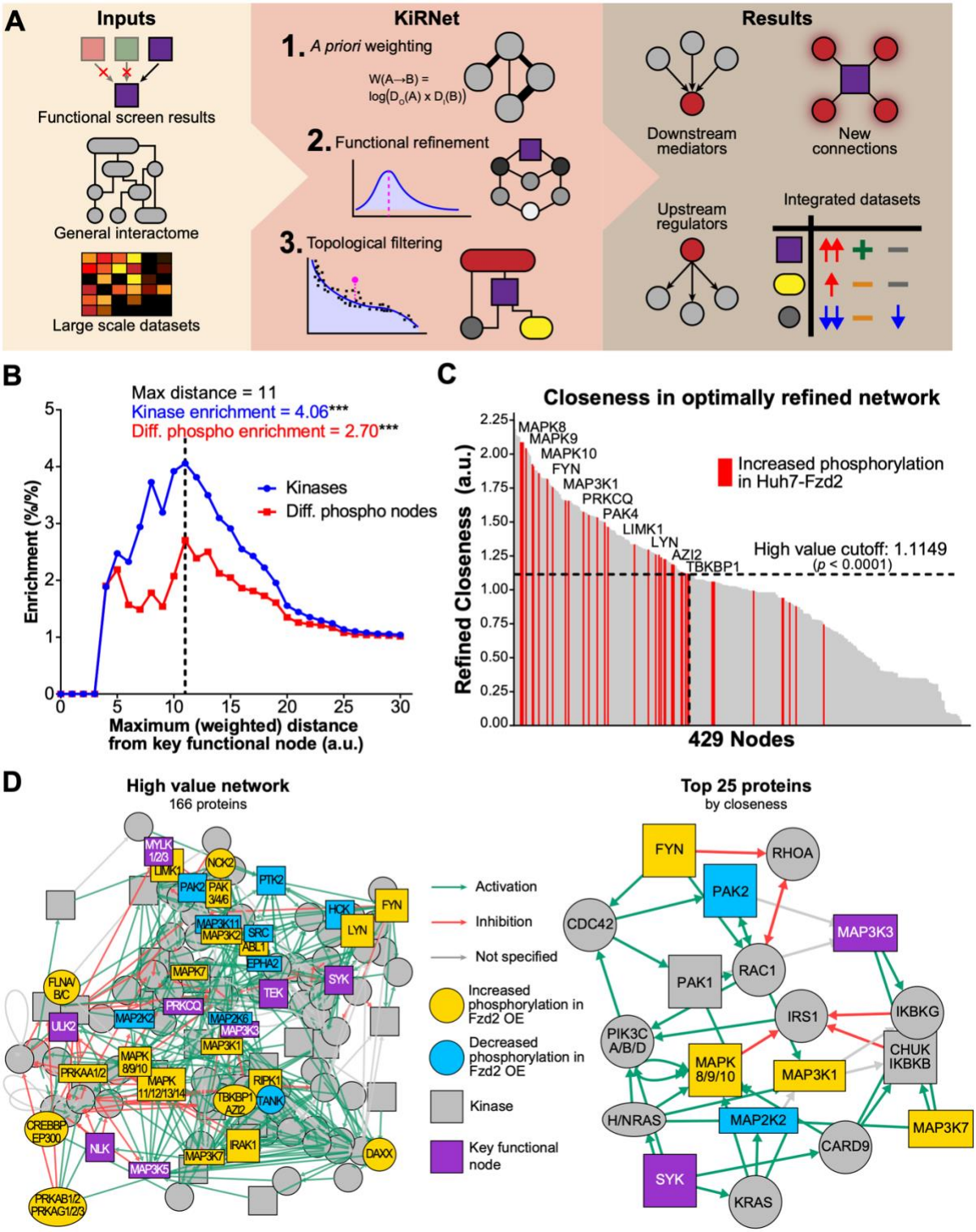


**Figure 4.1.** *Figure legend on following page.*

results into a more actionable, contextualized output, KiRNet maps KiR "hits" onto known protein-protein interaction (PPI) networks and amalgamates multiple large-scale datasets to

**Figure 4.1. KiRNet expands lists of hits from functional screens into network-level hypotheses. (A)** Schematic overview of the KiRNet method. **(B)** Kinase enrichment and differential phosphorylation enrichment at different cutoffs of distance from key functional nodes in the network. Enrichment of both reaches a peak at a distance of 11 (arbitrary units), establishing the cutoff for the refined subnetwork. ***$p<0.001$, Fisher's exact test. **(C)** Closeness for the 429 nodes in the optimally refined subnetwork, normalized to a minimum of 0 and mean of 1. Red bars indicate that node exhibits significantly higher phosphorylation in Huh7-Fzd2 cells compared to Huh7 WT cells. Selected differentially phosphorylated nodes are labeled. An empirically determined closeness cutoff of 1.1149274 was enforced to identify the high value subnetwork of 166 proteins ($p<0.0001$, Fisher's exact test). **(D)** Subnetworks produced by KiRNet for Huh7-Fzd2. High value subnetwork (left) contains 166 proteins chosen via an empirical cutoff for closeness (as calculated in the refined subnetwork). Top 25 proteins (right) shows the subnetwork of the 25 proteins with the highest closeness in the refined subnetwork. In both networks, sibling nodes (nodes with identical edges) have been collapsed into single nodes, thus each node may represent more than one protein.

create unified, kinase-centered functional network models. Generating these models requires only a list of functional hits and has the option of integrating multiple additional molecular data types (mRNA expression, genetic mutations, phosphopeptide quantification, etc.). This flexibility makes KiRNet a simple yet powerful tool for systems-level analyses.

Other members of the Gujral lab and I validated KiRNet performance using an epithelial hepatocellular carcinoma (HCC) cell line, Huh7, engineered to overexpress FZD2 (Huh7-Fzd2), a receptor for WNT5A/B that drives epithelial-mesenchymal transition (EMT) and cancer metastasis in HCC.[107,108] KiRNet produced a kinase-centered network model consisting of 166 proteins predicted to mediate this FZD2-driven mesenchymal cancer cell state. Mass spectrometry data showed that proteins in this network displayed strong differential phosphorylation in a FZD2-dependent manner, and perturbing this network significantly affected multiple EMT-associated phenotypes, including cell migration and E-Cadherin expression. Further, the expression of the network members was increased considerably in mesenchymal tumors across nine cancer types in TCGA, including HCC. These results demonstrate the power of KiRNet to generate predictive, functional network models of complex disease phenotypes, such as Fzd2-driven HCC.

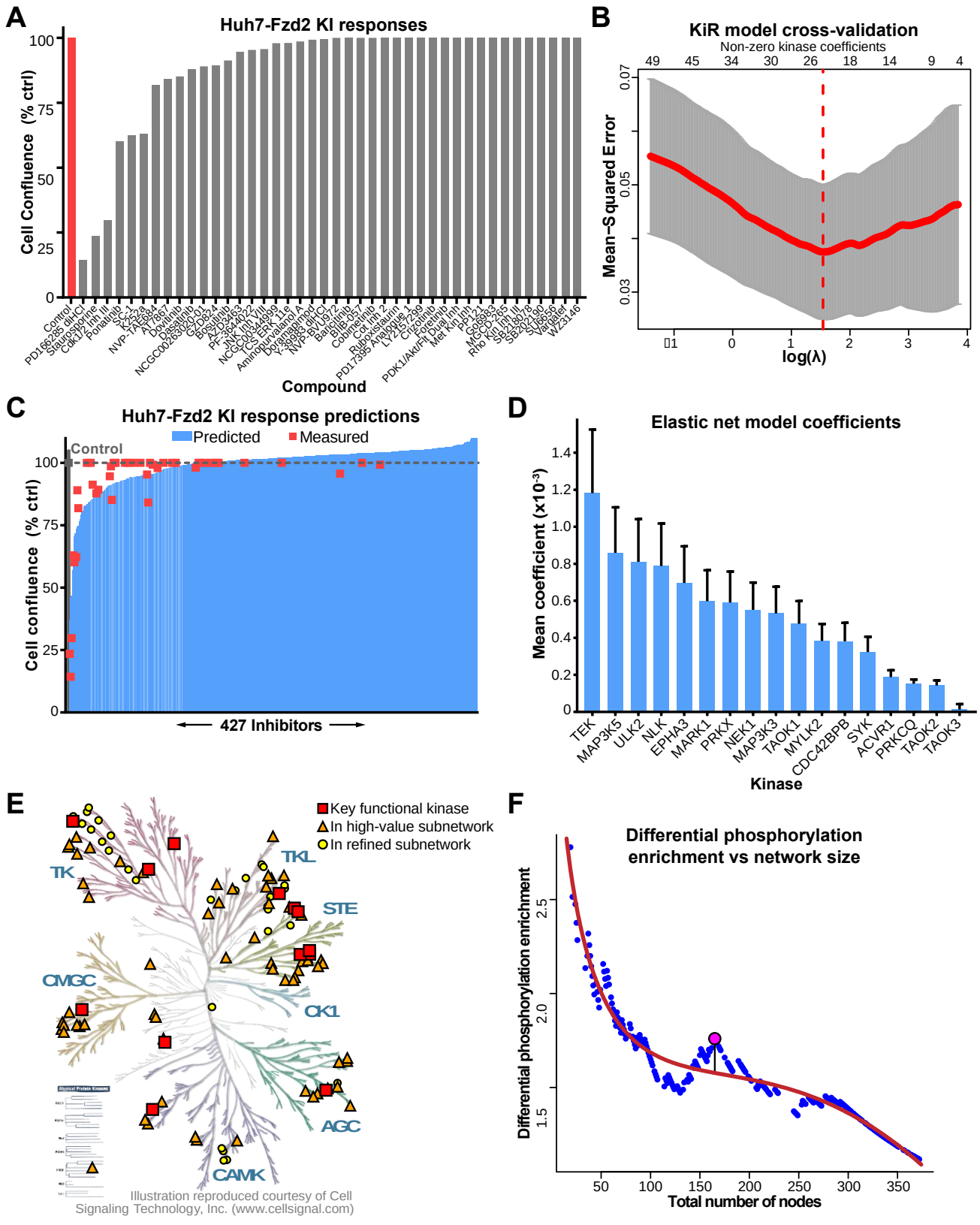


**Figure 4.2.** *Figure legend on following page.*

**Figure 4.2. Generation of Huh7-Fzd2 KiR model and refinement of KiRNet model. (A)** Interpolated cellular confluence responses of Huh7-Fzd2 cells to 42 KIs (plus control). Cells were treated with 7 doses of each KI, and their endpoint cellular confluence was quantified as the response. These responses were fit with a sigmoidal Hill equation and interpolated at the dose at which the compound was profiled (usually 500 nM). Data presented as % of untreated control. **(B)** Cross validation curve for the elastic net KiR model demonstrating the selection of the optimal penalty scalar λ. Models were computed for 300 geometrically distributed values of λ and evaluated via leave-one-out cross validation (LOOCV). Horizontal axis displays the logarithm of λ, vertical axis displays the LOOCV mean-square error ± standard deviation for that value. The value of λ for which this error reaches a minimum, indicated by the red dotted line, is chosen for the final model. **(C)** KiR model predictions for Huh7-Fzd2 response to all KIs profiled. Horizontal axis displays 427 KIs, vertical axis displays the predicted response to that KI based on the elastic net KiR model. Red dots represent experimentally measured values used to build the model, the same values displayed in panel (A). Blue bars represent the KiR model predicted response to each KI. **(D)** Elastic net coefficients for the kinases selected as key functional nodes for Huh7-Fzd2 cell proliferation. Data presented as mean ± standard deviation for the coefficients across 10 values of the elastic net hyperparameter α. **(E)** KinMap of kinases predicted by KiR and KiRNet to be important for Huh7-Fzd2 cell proliferation. Red squares are KiR predicted kinases, orange triangles are kinases present in the high value subset, and yellow circles are kinases present in the optimally refined subnetwork. **(F)** Plot of the enrichment of differentially phosphorylated nodes versus the size (total number of nodes) of the high value network, for each empirically tested cutoff for refined closeness. Data was fit with a third-degree polynomial model of the square root of the size ($R^2 = 0.947$). Pink point represents the point with the maximum residual that was chosen for the final model, to maximize enrichment while minimizing the size (relative to the fitted relationship).

### *4.2 GENERATING A CONTEXT-SPECIFIC PROTEIN INTERACTION NETWORK*

The starting point for KiRNet is a list of genes or proteins that are predicted to be functionally important for the phenotype of interest (Figure 4.1A). This is a typical output of many functional screens such as large-scale RNAi or CRISPR-Cas9 screens. In our case, we applied KiR to study epithelial (Huh7 WT) and mesenchymal (Huh7-Fzd2) HCC cells *in vitro*, as described previously.[43] Briefly, in experiments performed by Marina Chan, Huh7-Fzd2 cells were treated with a panel of kinase inhibitors whose inhibitions against hundreds of human protein kinases have been previously quantified.[42] These quantitative inhibition profiles and the cells' responses to each drug were used as the explanatory and response variables in the generation of the KiR elastic net model, respectively (Figure 4.2A–C). This KiR model yields a semi-ordered list of kinases that were found to be the most explanatory in predicting the

responses of the model system (Figure 4.2D–E). These kinases are heretofore referred to as the "key functional nodes" inputted into KiRNet.

The second essential input to KiRNet is a generalized PPI network; that is, a set of nodes and edges where nodes represent proteins (or genes), and edges represent interactions between those proteins. This can be sourced from publicly accessible PPI databases such as STRING, KEGG, HPRD, OmniPath, BioGrid, or others.[55,106,109,110] For this purpose, I chose to use the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database,[109] which, although smaller than many others, is high-quality and well annotated, potentially making our models more robust. The interactions in KEGG were collated and simplified, resulting in a network of 6,021 nodes (genes/proteins) and 59,142 interactions.

Finally, the third (optional) input for KiRNet is any additional information describing the state of the network. Here, I collected both primary and previously-published data on genetic mutations, RNA expression levels, protein levels, and phosphopeptide levels for Huh7 and Huh7-Fzd2 cell lines.[108,111–113] The mass spectrometry experiments quantifying protein and phosphopeptide levels were performed by Martin Golkowski. While none of these datasets are required to run KiRNet, incorporating them gives the user better metrics with which to optimize and assess the models, such as by filtering unexpressed genes or scoring subnetworks based on differential phosphorylation. When working with multiple large-scale datasets such as these, mapping datum to the correct gene or protein becomes a non-trivial task, in part due to the large number of possible identifiers for the same node. To aid in these conversions, I generated a curated, non-comprehensive table of identifiers, including Entrez IDs, HGNC gene symbols, and Ensembl gene IDs. Once the identifiers in each dataset were unified, these data were all

incorporated as annotations to the nodes. All of these inputs are loaded into the KiRNet pipeline to generate an initial network model.

To develop a context-specific network for mesenchymal HCC cells, several optimization steps were undertaken. First, I removed any nodes annotated as having loss-of-function mutations or RNA levels below an empirical cutoff. Second, I removed a further 6% of the nodes that were isolated from the rest of the network in small (>33 node) clusters. Finally, I added *a priori* weights to all of the edges in the network to represent distances, or "costs", of each interaction, with higher weights signifying a less likely or more difficult interaction. Importantly, these weights are derived entirely *a priori* from the network topology based on the degree (number of edges) of each node, a strategy that penalizes large hub nodes and prevents them from dominating the network results. By annotating the nodes with data specific to the Huh7-Fzd2 context, optimizing the network, and defining these edge weights, I created an initial, context-specific network ready to be further refined and interrogated to produce systems-level, functional hypotheses.

### *4.3 INITIAL PROTEIN-PROTEIN INTERACTION NETWORK GENERATION AND EDGE WEIGHTING*

The initial protein-protein interaction network used in this paper was based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database.[109] Custom R scripts were used to download the KEGG Markup Language (KGML) file for each pathway. All pathways that had human (hsa) identifiers were collated to form a single master list of interactions. Because many interactions are present in multiple pathways, this master list was further simplified by aggregating interactions with the same starting node, ending node, and interaction type, and concatenating their subtype annotations. This simplified list was used as the input to igraph using

custom R scripts to generate a single network consisting of all KEGG Pathway human interactions.[114] The final result contains 6,021 unique proteins and 59,142 unique interactions. This network was further refined to remove indirect interactions (often present in KEGG pathways to visually condense long signaling cascades into a single step). Genes annotated as having loss-of-function mutations had their corresponding nodes removed from the network. Nodes with RNA expression levels, expressed as log2(counts per million + 1), less than a value of 1 were similarly removed. The network was then examined to determine how many weakly connected "components" it contains. A "component" is defined as a subnetwork wherein every node is reachable by every other node in the subnetwork; a *weakly* connected component does not consider the direction of interactions. The largest weakly connected component in the initial Huh7-Fzd2 network contained 94% of all nodes (5,301 nodes out of 5,649 total), while the second largest component contained only 32 nodes. We therefore chose to consider only the largest weakly connected component.

Once the initial network was generated, *a priori* weights (or "distances") were assigned to each edge based on the degree of each node, so as to penalize edges connected to large hub nodes with many interactions. One strategy for assigning edge weights to represent distances (i.e. higher numbers mean less likely interactions) is to take the negative logarithm of the probability of the interaction[115]:

$$W_{A\to B} = -\log(P(A \to B)) = \log\left(\frac{1}{P(A \to B)}\right) \qquad \textit{Eq. 4.10}$$

In many cases, the probabilities of interactions are derived from the underlying evidence for that interaction. Because KEGG Pathway consists mainly of very well-studied pathways, there are no probabilities or confidence values associated with interactions, as they are largely presumed to be correct in at least some contexts. Therefore, I chose an *a priori* strategy that

bases the probability of any single interaction occurring dependent only on the degree of the starting and ending nodes. This strategy assumes all interactions to or from a node are equally likely, but each starting node must select a single edge to which to transmit information or signals, and each ending node must similarly select a single edge from which to receive information or signals. Assuming that these selections are independent, the resulting probability of a specific interaction $A \rightarrow B$, starting at node $A$ and ending at node $B$, is the product of the inverses of the degrees:

$$P(A \rightarrow B) = \frac{1}{D_{out}(A) \times D_{in}(B)} \qquad \textit{Eq. 4.11}$$

where $D_{out}(A)$ is the out-degree of node A (the number of edges coming from node A) and $D_{in}(B)$ is the in-degree of node B (the number of edges ending at node B). Substituting into equation 1 yields:

$$W_{A \rightarrow B} = -\log(P(A \rightarrow B)) = \log(D_{out}(A) \times D_{in}(B)) \qquad \textit{Eq. 4.12}$$

All edges in the initial network were weighted using this strategy. These original edge weights were preserved throughout the modeling process, even when analyzing smaller subnetworks where some interactions have been removed.

### *4.4 REFINING INITIAL NETWORK TO AN OPTIMAL SUBNETWORK AND IDENTIFYING HIGH VALUE NODES*

The initial, context-specific network I generated still contained the majority of nodes and edges from the entire source interactome (5,301 nodes). My goal was to identify a smaller and more testable subnetwork that best captures the differential regulation and signaling that occurs in epithelial versus mesenchymal HCC cells by capitalizing on the functional insight provided by the prior screen. I assigned every node a "functional distance" equal to the length of the shortest undirected path between that node and the nearest key functional node. The aforementioned edge

weighting strategy ensured that these distances are not dominated by connections through large hub nodes, but did not completely exclude these hubs from the models. To continue the focus on kinase-centered subnetworks, I created series of hypothetical subnetworks by enforcing a functional distance cutoff at integer values between 1 and 30 (arbitrary units) and calculated the fold enrichment of the number of kinases included in each subnetwork. This enrichment peaked at a cutoff of 11 ($p<0.0001$; Figure 4.1B). For mesenchymal cancer cells (Huh7-Fzd2), I exploited quantitative phospho-peptide data that provided additional insight into which proteins are likely to be functionally important for the growth of these cells. The enrichment of differentially phosphorylated nodes (compared to Huh7 WT) also reached a maximum at a distance of 11, reinforcing this as a meaningful cutoff (Figure 4.1B). While these two metrics are not independent, as the phospho-peptide data are biased towards the kinome and kinase-interacting proteins, this validates the use of kinase enrichment as an *a priori* means of identifying a meaningful kinome-centered subnetwork and provides further support for the flexibility of KiRNet.

The resulting subnetwork, referred to as the "optimally refined" subnetwork, served as my hypothesis-generating platform for determining the functionally important nodes and edges in regulating and mediating the phenotype of interest. For our mesenchymal HCC model, this subnetwork contained 429 nodes. I wanted to predict which of these nodes are "high value" nodes, or critical nodes, that can be targeted to perturb the network and alter the cellular phenotype. To this end, I leveraged our phospho-peptide data and tested many established measures of a node's "importance" or "centrality", including closeness, betweenness, degree, influence, and combinations thereof.[115–120] I found that, out of all of the metrics we tested, the closeness (the inverse of the average distance between a node and all other nodes in the

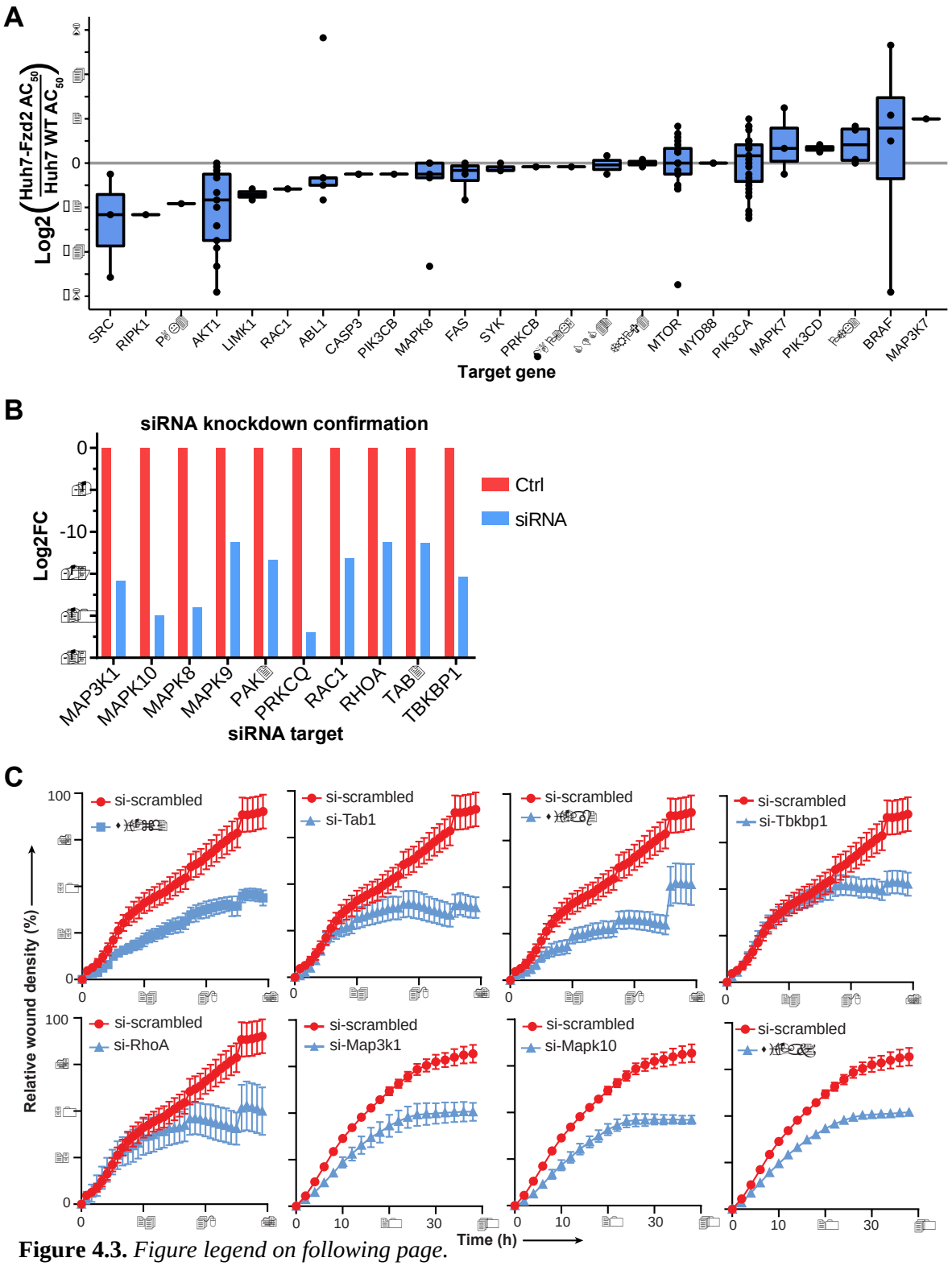


**Figure 4.3.** *Figure legend on following page.*

network[116]) in the refined subnetwork was the most effective at identifying subsets with further enrichment of differentially phosphorylated proteins (Figure 4.1C). Thus, in cases where there

**Figure 4.3. Additional validation of final KiRNet high value model. (A)** Responses to MIPE drugs annotated as targeting genes present in the high value subset of the Huh7-Fzd2 KiRNet model, selected from a large-scale drug screen. Horizontal axis displays the annotated drug targets, while the vertical axis displays the log2 ratio of the AC50 for Huh7-Fzd2 versus Huh7 WT. Negative values indicate increased sensitivity in Huh7-Fzd2 compared to Huh7 WT. **(B)** Confirmation of siRNA knockdown for select genes. Quantitative, real time PCR was used to quantify expression levels. Data was normalized using the $2^{-\Delta\Delta Ct}$ method, normalized to GAPDH. **(C)** Selected time course plots of wound-healing assay. Huh7-Fzd2 cells were transfected with transient siRNA knockdowns of the indicated high value gene

are no additional data present, this "refined closeness" of a node can be used as a topological predictor of that node's importance in the network and, consequently, for the phenotype. I identified an empirical cutoff that resulted in the largest enrichment of differentially phosphorylated nodes while also minimizing the size of the subnetwork (Figure 4.1C; Figure 4.2F), creating a high value subnetwork of 166 proteins (Figure 4.1D). This subnetwork serves as the final KiRNet model for our experimental system: a focused subnetwork, centered around the original functional hits, that predicts the proteins and relationships most critical for the phenotype of interest.

### *4.5 VALIDATING HYPOTHESES FOR THE MESENCHYMAL CANCER CELL MODEL*

This KiRNet model represents the predicted kinase-centered subnetwork of proteins and interactions that regulate the mesenchymal cancer cell state observed in FZD2-expressing Huh7 cells. As preliminary validation for this prediction, I compared the phosphorylation changes in Huh7-Fzd2 (versus Huh7 WT) to an independent HCC cell line. The FOCUS HCC cell line exhibits endogenously high levels of FZD2 and is commonly used in Wnt signaling studies.[107] Other members of the Gujral lab had previously generated a stable cell line with a short-hairpin knockdown of FZD2 that we refer to as FOCUS-shFZD2.[107] I hypothesized that the high value subnetwork from Huh7-Fzd2 would overlap with FZD2 signaling in FOCUS cells; thus, the shRNA knockdown should reverse the changes in phosphorylation we see in the Huh7-Fzd2

network. I observed 11 phosphosites that exhibited decreased phosphorylation in Huh7-Fzd2 (compared to Huh7 WT) and increased phosphorylation in FOCUS-shFZD2 (compared to FOCUS WT) (Figure 4.4A, left). A further set of 34 phosphosites exhibited increased phosphorylation in Huh7-Fzd2 and decreased in FOCUS-shFZD2. Together, these two sets account for approximately half of the 96 measured phosphosites in the high value network that correlate with FZD2 expression in both cell lines. This supports the hypothesis that these high value nodes, chosen by their topological closeness, are indeed functionally important for the FZD2-mediated phenotype in both cell lines.

In addition to phosphorylation data, we used Huh7 WT and Huh7-Fzd2 cell lines to carry out a drug screen that annotates each inhibitor with a specific gene target.[121] These experiments were performed by Michele Ceribelli and Craig Thomas. The sensitivities of Huh7 WT and Huh7-Fzd2 were quantified and compared using the log fold change of $AC_{50}$ (concentration for half-maximal activity) values. A lower $AC_{50}$ value indicates increased sensitivity to the drug, thus a negative log fold change in $AC_{50}$ indicates a higher sensitivity in Huh7-Fzd2 versus Huh7 WT. Drugs annotated as targeting one of the high value proteins tended to exhibit increased sensitivity in Huh7-Fzd2 cells (Figure 4.3A). Of note, drugs targeting SRC, PAK4, LIMK1, RAC1, and MAPK8 all exhibited increased efficacy in Huh7-Fzd2 cells (median log2 fold change of -2.3, -1.8, -1.4, -1.2, and -0.5, respectively). These data reinforce the hypothesis that these proteins are functionally critical in the FZD2-driven cells.

Next, I selected a panel of specific nodes from the mesenchymal cell KiRNet model to further validate their role in the mesenchymal phenotype. I chose nodes from the following categories: a kinase predicted by our original functional screen and KiR model (PRKCQ); kinases not predicted by our functional screen (EPHA2, MAP2K2/6, MAP3K11, MAPK8/9/10,

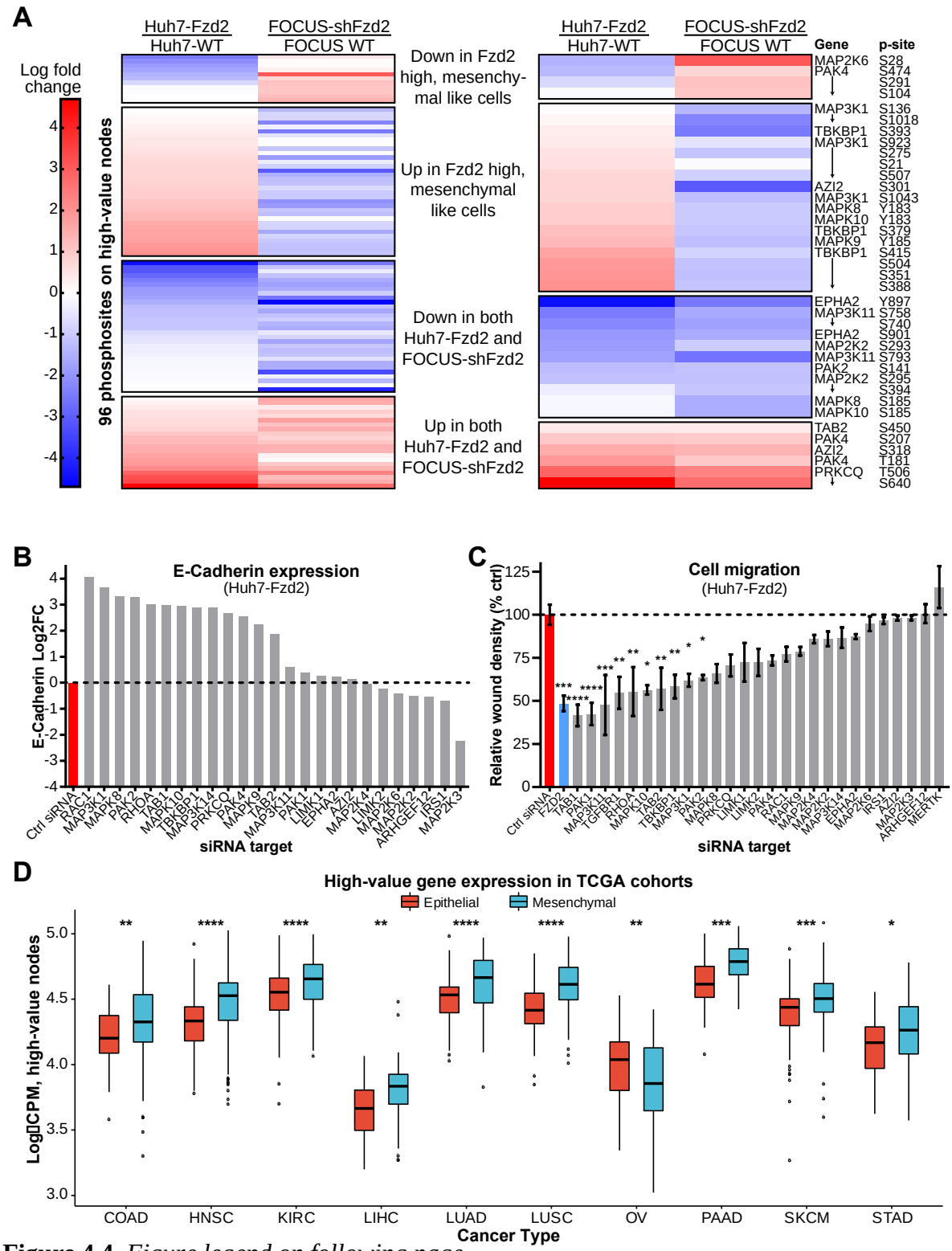


**Figure 4.4.** *Figure legend on following page.*

PAK1/2/4, LIMK1, MAP2K3/4, MAP3K1/14), and non-kinases (ARHGEF12, AZI2, IRS1, RAC1, RHOA, TAB1, TAB2, TBKBP1). The phosphosites on these chosen proteins followed

**Figure 4.4. KiRNet model identifies a critical subnetwork regulating a FZD2-driven, mesenchymal cancer cell state. (A)** Heatmap of phosphorylation changes in Huh7-Fzd2 and FOCUS cells. Left panel shows all 96 detected phosphosites on all of the proteins from the high value model, right panel shows selected phosphosites from proteins chosen for further validation. Data presented as the log fold change in phosphopeptide level between the modified cell line (Huh7-Fzd2 or FOCUS shFZD2) and the parental cell line (Huh7 WT or FOCUS WT, respectively). Top group contains sites decreased in Huh7-Fzd2 and increased in FOCUS shFZD2; second group contains sites increased in Huh7-Fzd2 and decreased in FOCUS shFZD2; third group contains sites decreased in both; bottom group contains sites increased in both. **(B)** Quantitative, real-time PCR results for E-Cadherin (*CDH1*) expression in Huh7-Fzd2 cells transfected with transient siRNA knockdowns targeting various high value genes. Presented as the log2 fold change compared to a non-targeting siRNA. **(C)** Quantitative changes in cell migration as assessed by a wound-healing assay in Huh7-Fzd2 transfected with transient siRNA targeting various high value genes. siFZD2 and siTGFBR1 are included as positive controls, and siMERTK is included as a negative control. Data are presented as mean±SEM of at least three biological replicates. *$p$<0.05, **$p$<0.01, ***$p$<0.001, ****$p$<0.0001, one-way ANOVA with two-tailed Holm-Sidak multiple comparisons test. **(D)** Expression of the high value nodes from the Huh7-Fzd2 KiRNet model in TCGA cohorts. Samples were stratified into epithelial or mesenchymal sample types based on gene expression signatures. LogCPM signature of the high value genes was significantly higher in mesenchymal samples in 9 out of 17 tested cancer types, including hepatocellular carcinoma. Signature expression was significantly decreased in mesenchymal ovarian cancer samples, and not significantly different in the other cancer types. *$p$<0.05, **$p$<0.01, ***$p$<0.001, ****$p$<0.0001, Wilcoxon rank-sum test. COAD = colon adenocarcinoma, HNSC = Head and neck squamous cell carcinoma, KIRC = kidney renal clear cell carcinoma, LIHC = liver hepatocellular carcinoma, LUAD = lung adenocarcinoma, LUSC = lung squamous cell carcinoma, OV = ovarian serous cystadenocarcinoma, PAAD = pancreatic adenocarcinoma, SKCM = skin cutaneous melanoma, STAD = stomach adenocarcinoma.

the same trend observed for the high value nodes, with approximately half of the sites changing in ways that correspond to the FZD2 status of both Huh7 and FOCUS cells (Figure 4.4A, right). To quantify the effects of these proteins on the mesenchymal-like state driven by FZD2, Taran Gujral depleted their expression levels in Huh7-Fzd2 cells via RNAi (Figure 4.3B). Expression levels of E-cadherin, a common epithelial cell state marker, exhibited over a 3-fold increase for approximately half of the genes tested, including RAC1, MAPK8/9/10, PAK2/4, RHOA, TAB1/2, TBKBP1 and MAP3K1, reinforcing the role these proteins play in maintaining the mesenchymal state of these cells (Figure 4.4B). Taran Gujral also performed a wound-healing assay as an aggregate measure of the migratory and proliferative phenotype exhibited by FZD2-driven cells. We observed appreciable decreases in wound closure for the majority of tested siRNA targets (Figure 4.4C). Depletions of MAPK10, PAK1/2, RHOA, TAB1/2, TBKBP1 and

MAP3K1 each resulted in more than a 35% decrease in wound closure compared to scrambled siRNA control ($p<0.05$; Figure 4.3C, Figure 4.4C). Together, these data identify kinases and non-kinases that play a role in the mesenchymal cancer cell phenotype.

Finally, I asked whether the KiRNet identified nodes in the mesenchymal HCC cells are also expressed in a broader set of mesenchymal-like tumors. Using methods developed by Andrew Xue, Nithisha Khasnavis analyzed mRNA expression of the high value genes across 17 different cancer types in The Cancer Genome Atlas (TCGA) database.[122] Samples in TCGA were stratified into epithelial- or mesenchymal-like cell states based on the expression of key marker genes, as defined previously.[113] A high value signature was computed for each sample from expression levels of the 166 high value genes. Nine of the tested cancer types, including HCC, showed significantly higher expression of this high value signature in mesenchymal-like cancers, while only ovarian cancer showed significantly decreased expression (Figure 4.4D). This provides further support that the high value network predicted using KiRNet mediates a mesenchymal-like phenotype, found in both FZD2-driven cells and mesenchymal-like cancers.

### *4.6 Discussion*

As the cost and time required for many functional screens decreases, there will be an ever increasing need to follow up the resulting leads. Traditional approaches of manual literature curation and one-by-one experiments can take years to interpret and contextualize these results. KiRNet integrates functional predictions with PPI networks to place these results in the broader context of cell biology. Furthermore, KiRNet enables the discovery of additional nodes that may have been missed in screens due to either biases within the methods or redundancies in the underlying biology that are still physiologically important. Generating context-specific networks

provides robust intermediate step for testable and accurate hypotheses that best leverage existing knowledge.

I demonstrated the power of KiRNet by identifying a "high value" network of proteins that are functionally important for mesenchymal cancer cells, specifically HCC cells driven by *FZD2*. Other members of the Gujral lab and I validated that these proteins are indeed critical for maintaining the migration and proliferation that accompanies this phenotype. Some proteins, such as RAC1, RHOA, and SRC, have already been identified as key players in this area.[123–125] Others, such as TBKBP1, PAK1/2/4, and MAPK8/9/10, expand on the initial functional screen hits and allow us to consider new proteins and interactions that otherwise would have been missed. KiRNet is especially powerful when handling limited amounts of data, as high value predictions can be made based on nothing more than an original list of functional hits. Moreover, as additional chemical biology tools become available, KiRNet may serve as a blueprint for investigating network models for gene families beyond kinases, such as epigenetic regulators, phosphatases, or GTPases.[121,126–131] I expect that KiRNet will pave the way for a broader, contextual understanding of biological systems.

## *4.7 Materials and Methods*

### Cell culture and media

All cell culture experiments in this chapter were performed by Marina Chan and Taran Gujral. Hepatocellular Huh7 cells were obtained from American Type Culture Collection. FOCUS cells were obtained from J. Wands (Brown University). Both cell lines were grown at 37°C under 5% CO2, 95% ambient atmosphere and maintained in Dulbecco's minimum essential medium (DMEM) supplemented with 10% FBS (Sigma) and 1% Penn Strep. Stable cell lines, FOCUS

cells with depleted levels of FZD2 and Huh7 cells expressing *FZD2*, were generated and cultured as described previously.[107,113]

**Kinase inhibitor screening**

The KI dosing experiments and the original drug screen were performed by Marina Chan. Kinase inhibitor screening was performed as described previously.[43] Huh7-Fzd2 cells were plated in 96 well plates and treated with a panel of 42 kinase inhibitors at 6–8 different doses each. Images of the cells were taken every 2 hours, and cell confluence (the proportion of total area occupied by cells) was quantified via Incucyte ZOOM™ software. Endpoint confluence at each dose was plotted and fitted with a dose-response curve using GraphPad Prism. These curves were interpolated at 500 nM (the dose used in the drug-target matrix), and the interpolated response was used as the response variable in building the KiR models.

**Elastic Net regularized models**

KiR models and the list of predicted "key kinases" (or "key functional nodes") were generated as previously described.[43] Briefly, a panel of 427 kinase inhibitors previously had their pairwise effects on 298 human kinases profiled. The result is a quantitative drug-target matrix, where each entry is a percentage between 0 and 100 that represents that kinases residual activity (as a percent of control, uninhibited activity) in the presence of that inhibitor.[42] A small panel of these inhibitors were tested on both Huh7 WT and Huh7-Fzd2 cells as described above, with the end result being a single interpolated response for each drug that represents the cell line's growth (as % control) at the profiled dose of the inhibitor (usually 500 nM). The kinase inhibition profile of each inhibitor, and the cell line's quantitative response to that inhibitor, were used as the explanatory and response variables, respectively, for elastic net regularized multiple linear regression models.[47] Custom R scripts employing the "glmnet" package were used to generate

the final models.[46] Leave-one-out cross validation (LOOCV) was used to select the optimal value for the penalty scaling factor λ (Figure 4.2B). Models were computed for 11 evenly-spaced values of α (the relative weighting between LASSO and Ridge regularization) ranging from 0 to 1.0 inclusive. Kinases with non-zero coefficients in at least one of these models (with the exception of $\alpha = 0$, which always has non-zero coefficients for every kinase) were considered "hits" (Figure 4.2D–E). Model accuracy was assessed via the LOOCV error as well as the root-mean-squared error of the predictions for the tested inhibitors (Figure 4.2B–C).

**KinMap**

The KinMap in Figure 4.2E was generated using the online KinMap tool available at http://www.kinhub.org/kinmap.[10]

**Generating identifier conversion table**

I combined multiple databases and tools with manual curation to generate a table containing multiple gene/protein identifiers to aid in conversion. STRING and KEGG IDs were sourced from their respective databases and used as the starting point, as these are the genes/proteins in which we are interested. Ensembl's online BioMart tool was used to find ENSG, ENSP, and Entrez IDs.[132] HGNC's online conversion tools and symbol checkers were also employed for additional conversions and verifications.[133] Missing entries or entries that did not match across databases were subject to manual curation. The final table contained 11,466 entries each with Ensembl Gene IDs, Ensembl protein IDs, Entrez IDs, HGNC gene symbols, HGNC gene descriptions, identifiers (where available). Since genes, proteins, and identifiers do not always correspond one-to-one, for both biological and technical reasons, preference was given to the most abundant isoform or the most widely-used identifier.

**Identifying refined closeness cutoff for high-value subnetwork**

The “refined closeness” of each node was computed using the “closeness” function in the igraph package. These values were shifted to a minimum of 0 and scaled to a mean of 1. In order to choose the best cutoff for refined closeness, I plotted the enrichment of differentially phosphorylated nodes that would be included in the high value set versus the total number of nodes included for each value of the refined closeness. Because the enrichment was a (mostly) decreasing function of subnetwork size, we fit this relationship with a polynomial model and selected the point with the maximum residual; that is, the subnetwork that, relative to the fitted relationship, had the largest enrichment for its size (Figure 4.2F). The empirically derived cutoff value for (normalized) closeness was 1.1149, resulting in a final network size of 166 nodes and an enrichment of 1.76 ($p$<0.0001, Fisher’s exact test).

**Small interfering RNA transfection**

All small interfering RNA (siRNA) experiments in this chapter were performed by Taran Gujral. All small interfering RNA (siRNA) were obtained from Dharmacon (Thermo). siRNA’s transfections in both 12-well plate for expression profiling and in 96-well plates for cell migration were carried out using Lipofectamine RNAiMax (Invitrogen) according to manufacturer instructions.

**RNA Extraction and Quantitative Real-Time PCR**

All RNA extractions and first-strand reactions in this chapter were performed by Taran Gujral. mRNA expression changes in *CDH1* was determined using quantitative PCR (qPCR). Total cellular RNA was isolated using an RNeasy Mini Kit (QIAGEN). Briefly, 1 μg of total RNA was reverse transcribed into first-strand cDNA using an RT2 First Strand Kit (QIAGEN). The resultant cDNA was subjected to qPCR using human *CDH1*-specific primer (Realtimeprimers.com) and *GAPDH* (housekeeping control). The qPCR reaction was performed

with an initial denaturation step of 10 min at 95 °C, followed by 15 s at 95 °C and 60 s at 58 °C for 40 cycles using Biorad CFX384 thermocycler (Biorad). The mRNA levels of *CDH1* were normalized relative to the mean levels of the housekeeping gene and compared using the $2^{-\Delta\Delta Ct}$ method as described previously.[43,107]

**Migration analysis**

All wound healing assays in this chapter were performed by Taran Gujral. To study the effect of KiRNet predicted nodes on migration of mesenchymal cancer cells, a wound-healing assay was employed as described previously.[43,107] Briefly, siRNAs targeting various proteins and scrambled control were transfected in Huh7-Fzd2 cells using Lipofectamine RNAiMax (Invitrogen) according to manufacturer instructions. Cells were plated on 96-well plates (Essen Image Lock, Essen Instruments) and 48 hours post transfections, a wound was scratched with wound scratcher (Essen Instruments). Wound confluence was monitored with Incucyte Live-Cell Imaging System and software (Essen Instruments). Wound closure was observed every 2 hours for 24-72 hours by comparing the mean relative wound density of at least three biological replicates.

**TCGA Analysis**

The TCGA analysis in this chapter was performed by Nithisha Khasnavis, using methods developing by Andrew Xue.[113] Patient data and clinical manifests were downloaded from selected TCGA (The Cancer Genome Atlas) projects through the National Cancer Institute's Genomic Data Commons data portal (https://portal.gdc.cancer.gov/) using the GenomicDataCommons Bioconductor package in R. 17 TCGA patient cohorts, containing 7881 patients in total, were selected, representing both high incidence and highly aggressive cancer subtypes. Data was processed as described previously.[113] To assess co-expression between KiRNet predicted nodes and EMT, we used previously reported epithelial and mesenchymal

marker genes to rank patients in each TCGA cohort by calculating the mean-rank of their epithelial marker expression and mesenchymal marker expression, giving an E-score and M-score respectively. Patients with above median E-score and below median M-score were labelled as the epithelial-like group, while patients with above median M-score and below median E-score were labelled as mesenchymal-like. LogCPM of the high value signature was calculated by taking the average of the log(CPM+1) values for all available genes from the signature within each sample. *P*-values were calculated using a Wilcoxon rank-sum test. The comprehensive list of cancer types analyzed is as follows: breast invasive carcinoma, cervical squamous cell carcinoma and endocervical adenocarcinoma, colon adenocarcinoma, glioblastoma multiforme, head and neck squamous cell carcinoma, kidney renal clear cell carcinoma, liver hepatocellular carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, mesothelioma, ovarian serous cystadenocarcinoma, pancreatic adenocarcinoma, prostate adenocarcinoma, rectum adenocarcinoma, sarcoma, skin cutaneous melanoma, and stomach adenocarcinoma.

**Kinase affinity enrichment and on-bead digestion**

All mass spectrometry experiments and analyses in this section were performed by Martin Golkowski and Shao-En Ong. Kinase affinity enrichment and on-bead digestion was performed as previously described.[108] Briefly, three micro tubes containing 35 µl of a 50% slurry of the in-house-made, optimized kinobead mixture in 20% aq. ethanol were prepared for each pulldown experiment. The beads were washed twice with 300 µl modified RIPA buffer (50 mM Tris, 150 mM NaCl, 0.25% Na-deoxycholate, 1% NP-40, 1 mM EDTA and 10 mM NaF, pH 7.8). 1 mg of protein extract in mod. RIPA buffer containing HALT protease inhibitor cocktail (100x, Thermo Fisher Scientific, Waltham, MA) and phosphatase inhibitor cocktail II and III (100x, Sigma-Aldrich, St Louis, MO) were added to the first tube. The mixture was incubated on a tube rotator

for 1h at 4°C and then the beads were spun down rapidly at 2000 rpm on a benchtop centrifuge (5s). The supernatant was pipetted into the next tube with kinobeads for the second round of affinity enrichment. The procedure was repeated once more for a total of three rounds of affinity enrichment. After removal of the supernatant, the beads were rapidly washed twice with 300 µl of ice-cold mod. RIPA buffer and three times with 300 µl ice-cold tris-buffered saline (TBS, 50 mM tris, 150 mM NaCl, pH 7.8) to remove detergents. 100 µl of the denaturing buffer (20% trifluoroethanol (TFE)[134], 25 mM Tris containing 5 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP*HCl) and 10 mM chloroacetamide (CAM), pH 7.8), were added and the slurry vortexed at low speed briefly. At this stage, kinobeads from the three tubes are combined and heated at 95°C for 5 min. The mixture was diluted 2-fold with 25 mM triethylamine bicarbonate (TEAB), the pH adjusted to 8-9 by addition 1 N aq. NaOH; 5 µg LysC were added and the mixture agitated on a thermomixer at 700 rpm at 37°C for 2 h. Then 5 µg MS-grade trypsin (Thermo Fisher Scientific, Waltham, MA) were added, and the mixture agitated on a thermomixer at 700 rpm at 37°C overnight. 600 µl of 1% formic acid was added and the mixture acidified by addition of another 6 µl of formic acid to yield 1.2 ml peptide solution in total. An aliquot of 120 µl (10%) of the peptide solution was desalted using StageTips and analyzed in single nanoLC-MS/MS runs for protein quantification. The remaining peptide solution (90%) was dried under vacuum at RT on a SpeedVac. 300 µl of 70% aq. ACN + 0.1 % TFA was added to each tube, the mixture vortexed, and sonicated in a bath sonicator until dried peptide residue dissolved. In case the dried residue could not be fully resuspended, additional 0.1% aq. TFA can be added in 10 µl increments until dissolved. The solution was subjected to IMAC phosphopeptide enrichment protocol and desalted using StageTips (see ‘IMAC phosphopeptide enrichment’ and ‘Peptide and phosphopeptide desalting with StageTips’ above).

**nanoLC-MS/MS analyses**

The LC-MS/MS analyses were performed as described previously with the following minor modifications.[108,135] Peptide samples were separated on a Thermo-Dionex RSLCNano UHPLC instrument (Sunnyvale, CA) using 20 cm long fused silica capillary columns (100 µm ID) packed with 3 µM 120 Å reversed phase C18 beads (Dr. Maisch, Ammerbuch, DE). For whole peptide samples the LC gradient was 120 min long with 10−35% B at 300 nL/min. For phosphopeptide samples the LC gradient was 120 min long with 3−30% B at 300 nL/min. LC solvent A was 0.1% aq. acetic acid and LC solvent B was 0.1% acetic acid, 99.9% acetonitrile. MS data was collected with a Thermo Fisher Scientific Orbitrap Elite instrument. Data-dependent analysis was applied using Top15 selection with CID fragmentation.

**Computation of MS raw files**

Data .raw files were analyzed by MaxQuant/Andromeda[136] version 1.5.2.8 using protein, peptide and site FDRs of 0.01 and a score minimum of 40 for modified peptides, 0 for unmodified peptides; delta score minimum of 17 for modified peptides, 0 for unmodified peptides. MS/MS spectra were searched against the UniProt human database (updated July 22nd, 2015). MaxQuant search parameters: Variable modifications included Oxidation (M) and Phospho (S/T/Y). Carbamidomethyl (C) was a fixed modification. Max. missed cleavages was 2, enzyme was Trypsin/P and max. charge was 7. The MaxQuant "match between runs" feature was enabled. The initial search tolerance for FTMS scans was 20 ppm and 0.5 Da for ITMS MS/MS scans.

**MaxQuant output data processing**

MaxQuant output files were processed and statistically analyzed using the Perseus software package v1.5.6.0.[137] Human gene ontology (GO) terms (GOBP, GOCC and GOMF) were loaded from the 'Perseus Annotations' file downloaded on 01.08.2017. Expression columns (protein and

phosphopeptide intensities) were log2 transformed and normalized by subtracting the median log2 expression value from each expression value of the corresponding data column. Potential contaminants, reverse hits and proteins only identified by site were removed. Reproducibility between LC-MS/MS experiments were analyzed by column correlation (Pearson's r) and replicates with a variation of $r > 0.25$ compared to the mean r-values of all replicates of the same experiment (cell line or knockdown experiment) were considered outliers and excluded from the analyses. Data imputation was performed using a modeled distribution of MS intensity values downshifted by 1.8 and having a width of 0.2. For statistical testing of significant differences in expression, a two-sample Student's T-test with Benjamini-Hochberg correction for multiple hypothesis testing was applied (FDR = 0.05).

**Quantification and statistical analysis**

Enrichment tests for kinases and differentially phosphorylated nodes were performed using custom R scripts employing the fisher.test function to implement a Fisher's Exact test for overrepresentation.

Significant differences in expression of peptides and phosphosites were assessed using a two-sample Student's T-test with Benjamini-Hochberg correction for multiple hypothesis testing (FDR = 0.05).

Cell migration in the wound healing assay was quantified using the Incucyte ZOOM$^{TM}$ software, which computes the area percentage of the of the initial wound that is occupied by cells. Significance was assessed in GraphPad Prism using a one-way ANOVA with the Holm-Sidak multiple comparisons test.

TCGA significance was calculated using custom R scripts implementing a Wilcoxon rank sum test.

# Chapter 5: Conclusions and future directions

Kinase biology has been a hotspot for innovation and discovery for decades. As catalysts of a reversible reaction, they represent one of the most functionally important molecular signaling paradigms to be understood and exploited. In chapter 1, I introduced the challenges and opportunities presented by the polypharmacology of KIs. KIs provide a powerful tool for both chemical biology and clinical therapies, but their lack of specificity creates obstacles in both realms. I developed KInhibition, an accessible web portal that employs quantitative analyses to make data-driven decisions regarding KI selectivity. At its core, KInhibition seeks to connect users to the wealth of available information about KIs, placing a specific scientific query into the larger context of existing knowledge. This enables a streamlined approach to chemical biology.

The current functions of KInhibition are relatively simple, but they provide a platform upon which more in-depth and complex analyses may be built. As it stands, each KI present in the datasets contains only a few necessary pieces of identifying information (dose, identifier, etc.). However, more detailed information can be added later, such as literature references, chemical structures, and even more specific chemical descriptions such as functional groups, polarity, and solubility. These later data points, in particular, could open even more doors for correlative analyses. For example, since KInhibition dynamically computes selectivity scores, these scores could be compared to the chemical properties of the profiled KIs to identify specific functional groups or other moieties that are closely associated with selectivity. This would require both the generation of full chemical descriptions of the KIs and the development of a robust classifier that can identify the most relevant properties. However, in the era of high-throughput analyses and machine learning, both of these are entirely feasible.

Given that kinases mediate nearly every cellular phenotype, it is unsurprising that they are one of the most common classes proteins implicated in many cancers. Despite its traditional association with androgen signaling, prostate cancer is no exception. In chapter 2, I applied a systems-based approach to modeling kinase signaling that predicted two KIs that effectively inhibited the growth of CRPC across multiple model systems. First, I updated and refined the KiR pipeline to leverage newly available data on KI inhibition profiles. By integrating this data with the results of an *in vitro* functional screen, I generated KiR models for four CRPC cell lines. These models predicted that both PP121 and SC-1 would effectively abrogate growth across all 4 cell lines, which I validated experimentally. With the help of my amazing collaborators at MD Anderson, we further demonstrated the efficacy of these KIs in subcutaneously-implanted CRPC tumors *in vivo*. Finally, we also showed that these KIs can synergize with the chemotherapeutic docetaxel to inhibit the growth of CRPC tumors implanted into the tibia, providing powerful preclinical evidence. Future studies can perform a similar test for KI synergy with androgen receptor targeting therapies such as enzalutamide or abiraterone to investigate the interplay between androgen signaling and kinase signaling in late stage CRPC. My predictive models led to powerful preclinical evidence for KI efficacy in CRPC. I am hopeful that future studies will continue this work and move closer to clinical applications.

The predictions made by KiR have proven invaluable resources across many areas of research. In every application, these predictions must be understood within the context of the larger system at hand. My attempt to streamline and standardize this process led to the creation of KiRNet in chapter 4. KiRNet is based on the idea that none of the predictions of methods such as KiR exist in a vacuum. A process as simple as examining the known PPIs that surround a protein of interest reveal new properties of the cellular system that would have otherwise been

missed. Integrating additional information as it is available provides added context and gives a more concrete form to these properties. KiRNet provides a tool that connects researchers with the right prior knowledge to interpret their predictions at a systems level.

As it is still new, there is a lot of potential for KiRNet to expand into a more diverse and powerful tool. As it stands, all PPIs are considered equally likely independent of any available data. This is intended to make KiRNet as flexible and available as possible. However, when additional molecular data are available, such as mRNA expression levels or quantitative phosphorylation levels, these could be taken into account to reassess the likelihood of a give interaction. For example, Bayes' formula could be used to update the probabilities of each interaction based on new observations. This could even evolve into a Bayesian subnetwork classifier that employs a more complex integration of the available data. On a simpler note, the underlying PPI could be expanded by amalgamating multiple sources, similar to OmniPath, rather than relying on a single interactome to serve as the base of the network. The flexibility of KiRNet in its current form gives ample room for future growth in any of these directions.

I have presented three examples of tools that aid researchers in interrogating and exploiting the human kinome to increase our understanding of human biology. Certain themes resonate throughout my work. For one, at the heart of each of my projects, there is a fundamental mathematical tool or concept upon which the rest of the functionality is built. For KInhibition, the derivation of the KISS is the crucial innovation that enables KInhibition to quantify and present selectivity. In applying KiR to prostate cancer, knowing how to apply elastic net regularization and how to interpret the results is fundamental to being able to make meaningful models and predictions. In KiRNet, the specific edge weighting formula employed is critical in narrowing down the entire interactome to only the relevant portions. In each of these cases, the

underlying mathematics are not immensely complicated, but applying them in just the proper context is what gives power to the methods.

Along similar lines, there is a core concept that each of these tools exhibit: they all enable researchers to connect their newly gathered data to a general wealth of prior knowledge. KInhibition is the most clear example of this, with the only user-supplied information being the target kinase(s). The rest is simply making previous drug-target screens more accessible. KiR relies heavily on such particular screen to connect new phenotypes to known kinase signaling. Finally, KiRNet does not generate a novel PPI network with each use, but simply connects each new set of predictions to their locations in an existing interactome. This pattern is consistent with the goals of systems biology to understand each piece of cellular information not just in isolation, but in the broader context of what is known. Doing so reveals incredible connections between previously unrelated fields and drives the interdisciplinary collaborative efforts that define the biological sciences. It has been an honor and a pleasure to play a part, however small, in unraveling the molecular mysteries of human life.

*"Yet such is oft the course of deeds that move the wheels of the world:*
*small hands do them because they must, while the eyes of the great are elsewhere."*

— J.R.R. Tolkien

## References


1. Hatje K, Mühlhausen S, Simm D, Kollmar M. The Protein-Coding Human Genome: Annotating High-Hanging Fruits. *BioEssays*. 2019;41(11):1900066. doi:10.1002/bies.201900066
2. Baharani A, Trost B, Kusalik A, Napper S. Technological advances for interrogating the human kinome. *Biochem Soc Trans*. 2017;45(1):65-77. doi:10.1042/BST20160163
3. Wilson LJ, Linley A, Hammond DE, et al. New Perspectives, Opportunities, and Challenges in Exploring the Human Protein Kinome. *Cancer Res*. 2018;78(1):15-29. doi:10.1158/0008-5472.CAN-17-2291
4. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science*. 2002;298(5600):1912-1934. doi:10.1126/science.1075762
5. Ardito F, Giuliani M, Perrone D, Troiano G, Muzio L Lo. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int J Mol Med*. 2017;40(2):271-280. doi:10.3892/ijmm.2017.3036
6. Roskoski R. Properties of FDA-approved small molecule protein kinase inhibitors. *Pharmacol Res*. 2019;144:19-50. doi:10.1016/j.phrs.2019.03.006
7. Fleuren EDG, Zhang L, Wu J, Daly RJ. The kinome "at large" in cancer. *Nat Rev Cancer*. 2016;16(2):83-98. doi:10.1038/nrc.2015.18
8. Lun X-K, Szklarczyk D, Gábor A, et al. Analysis of the Human Kinome and Phosphatome by Mass Cytometry Reveals Overexpression-Induced Effects on Cancer-Related Signaling. *Mol Cell*. 2019;74:1-17. doi:https://doi.org/10.1016/j.molcel.2019.04.021
9. Fabbro D, Cowan-Jacob SW, Moebitz H. Ten things you should know about protein

kinases: IUPHAR Review 14. *Br J Pharmacol*. 2015;172(11):2675-2700. doi:10.1111/bph.13096

10. 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(1):16. doi:10.1186/s12859-016-1433-7
11. Gallick GE, Corn PG, Zurita AJ, Lin S-H. Small-molecule protein tyrosine kinase inhibitors for the treatment of metastatic prostate cancer. *Future Med Chem*. 2012;4(1):107-119. doi:10.4155/fmc.11.161
12. Gordus A, Krall JA, Beyer EM, et al. Linear combinations of docking affinities explain quantitative differences in RTK signaling. *Mol Syst Biol*. 2009;5(1):235. doi:10.1038/msb.2008.72
13. Lisle JE, Mertens-Walker I, Rutkowski R, Herington AC, Stephenson SA. Eph receptors and their ligands: Promising molecular biomarkers and therapeutic targets in prostate cancer. *Biochim Biophys Acta - Rev Cancer*. 2013;1835(2):243-257. doi:10.1016/j.bbcan.2013.01.003
14. Wang B. Cancer Cells Exploit the Eph-Ephrin System to Promote Invasion and Metastasis: Tales of Unwitting Partners. *Sci Signal*. 2012;4(175):1-6. doi:10.1126/scisignal.2002153.Cancer
15. Nazarian R, Shi H, Wang Q, et al. Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature*. 2010;468(7326):973-977. doi:10.1038/nature09626
16. Bhullar KS, Lagarón NO, McGowan EM, et al. Kinase-targeted cancer therapies: progress, challenges and future directions. *Mol Cancer*. 2018;17(1):48.

doi:10.1186/s12943-018-0804-2

17. Park SI, Zhang J, Phillips KA, et al. Targeting Src family kinases inhibits growth and lymph node metastases of prostate cancer in an orthotopic nude mouse model. *Cancer Res*. 2008;68(9):3323-3333. doi:10.1158/0008-5472.CAN-07-2997

18. Walker-Daniels J, Coffman K, Azimi M, et al. Overexpression of the EphA2 tyrosine kinase in prostate cancer. *Prostate*. 1999;41(4):275-280. doi:10.1002/(SICI)1097-0045(19991201)41:4<275::AID-PROS8>3.0.CO;2-T

19. Drake JM, Graham NA, Stoyanova T, et al. Oncogene-specific activation of tyrosine kinase networks during prostate cancer progression. *Proc Natl Acad Sci*. 2012;109(5):1643-1648. doi:10.1073/pnas.1120985109

20. Wu P, Nielsen TE, Clausen MH. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol Sci*. 2015;36(7):422-439. doi:10.1016/j.tips.2015.04.005

21. Duong-Ly KC, Devarajan K, Liang S, et al. Kinase Inhibitor Profiling Reveals Unexpected Opportunities to Inhibit Disease-Associated Mutant Kinases. *Cell Rep*. 2016;14(4):772-781. doi:10.1016/j.celrep.2015.12.080

22. FISCHER EH, KREBS EG. *Conversion of Phosphorylase b to Phosphorylase a in Muscle Extracts*. Vol 216.; 1955:121-132. Accessed February 21, 2019. http://www.ncbi.nlm.nih.gov/pubmed/13252012

23. KREBS EG, FISCHER EH. Phosphorylase activity of skeletal muscle extracts. *J Biol Chem*. 1955;216(1):113-120. Accessed February 21, 2019. http://www.ncbi.nlm.nih.gov/pubmed/13252011

24. Silva-Evangelista C, Barret E, Ménez V, et al. A kinome-wide shRNA screen uncovers vaccinia-related kinase 3 (VRK3) as an essential gene for diffuse intrinsic pontine glioma

survival. *Oncogene*. 2019;38(38):6479-6490. doi:10.1038/s41388-019-0884-5

25. Mohseni M, Sun J, Lau A, et al. A genetic screen identifies an LKB1-MARK signalling axis controlling the Hippo-YAP pathway. *Nat Cell Biol*. 2014;16(1):108-117. doi:10.1038/ncb2884

26. Shen J, Najafi S, Stäble S, et al. A kinome-wide RNAi screen identifies ALK as a target to sensitize neuroblastoma cells for HDAC8-inhibitor treatment. *Cell Death Differ*. 2018;25(12):2053-2070. doi:10.1038/s41418-018-0080-0

27. de Wilde AH, Wannee KF, Scholte FEM, et al. A Kinome-Wide Small Interfering RNA Screen Identifies Proviral and Antiviral Host Factors in Severe Acute Respiratory Syndrome Coronavirus Replication, Including Double-Stranded RNA-Activated Protein Kinase and Early Secretory Pathway Proteins. *J Virol*. 2015;89(16):8318-8333. doi:10.1128/jvi.01029-15

28. Shibuya M, Suzuki Y, Sugita K, et al. Effect of AT877 on cerebral vasospasm after aneurysmal subarachnoid hemorrhage: Results of a prospective placebo-controlled double-blind trial. *J Neurosurg*. 1992;76(4):571-577. doi:10.3171/jns.1992.76.4.0571

29. Li J, Kim SG, Blenis J. Rapamycin: One drug, many effects. *Cell Metab*. 2014;19(3):373-379. doi:10.1016/j.cmet.2014.01.001

30. Wu P, Nielsen TE, Clausen MH. Small-molecule kinase inhibitors: An analysis of FDA-approved drugs. *Drug Discov Today*. 2016;21(1):5-10. doi:10.1016/j.drudis.2015.07.008

31. Jin L, Wang W, Fang G. Targeting Protein-Protein Interaction by Small Molecules. *Annu Rev Pharmacol Toxicol*. 2014;54(1):435-456. doi:10.1146/annurev-pharmtox-011613-140028

32. Apsel B, Blair JA, Gonzalez B, et al. Targeted polypharmacology: Discovery of dual

inhibitors of tyrosine and phosphoinositide kinases. *Nat Chem Biol*. 2008;4(11):691-699. doi:10.1038/nchembio.117

33. Gossage L, Eisen T. Targeting multiple kinase pathways: A change in paradigm. *Clin Cancer Res*. 2010;16(7):1973-1978. doi:10.1158/1078-0432.CCR-09-3182

34. Knight ZA, Lin H, Shokat KM. Targeting the cancer kinome through polypharmacology. *Nat Rev Cancer*. 2010;10(2):130-137. doi:10.1038/nrc2787

35. Ma X, Lv X, Zhang J. Exploiting polypharmacology for improving therapeutic outcome of kinase inhibitors (KIs): An update of recent medicinal chemistry efforts. *Eur J Med Chem*. 2018;143:449-463. doi:10.1016/j.ejmech.2017.11.049

36. Lin A, Giuliano CJ, Palladino A, et al. Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci Transl Med*. 2019;11(509):eaaw8412. doi:10.1126/scitranslmed.aaw8412

37. Davis MI, Hunt JP, Herrgard S, et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol*. 2011;29(11):1046-1051. doi:10.1038/nbt.1990

38. Dranchak P, MacArthur R, Guha R, et al. Profile of the GSK Published Protein Kinase Inhibitor Set Across ATP-Dependent and-Independent Luciferases: Implications for Reporter-Gene Assays. Jung M, ed. *PLoS One*. 2013;8(3):e57888. doi:10.1371/journal.pone.0057888

39. Gao Y, Davies SP, Augustin M, et al. A broad activity screen in support of a chemogenomic map for kinase signalling research and drug discovery. *Biochem J*. 2013;451(2):313-328. doi:10.1042/BJ20121418

40. Karaman MW, Herrgard S, Treiber DK, et al. A quantitative analysis of kinase inhibitor selectivity. *Nat Biotechnol*. 2008;26(1):127-132. doi:10.1038/nbt1358

41. Klaeger S, Heinzlmeir S, Wilhelm M, et al. The target landscape of clinical kinase drugs. *Science (80- )*. 2017;358(6367):eaan4368. doi:10.1126/science.aan4368

42. Anastassiadis T, Deacon SW, Devarajan K, Ma H, Peterson JR. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol*. 2011;29(11):1039-1045. doi:10.1038/nbt.2017

43. Gujral TS, Peshkin L, Kirschner MW. Exploiting polypharmacology for drug target deconvolution. *Proc Natl Acad Sci*. 2014;111(13):5048-5053. doi:10.1073/pnas.1403080111

44. Ma H, Deacon S, Horiuchi K. The challenge of selecting protein kinase assays for lead discovery optimization. *Expert Opin Drug Discov*. 2008;3(6):607-621. doi:10.1517/17460441.3.6.607

45. Arang N, Kain HS, Glennon EK, et al. Identifying host regulators and inhibitors of liver stage malaria infection using kinase activity profiles. *Nat Commun*. 2017;8(1):1232. doi:10.1038/s41467-017-01345-2

46. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22. doi:10.18637/jss.v033.i01

47. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol*. 2005;67(2):301-320. doi:10.1111/j.1467-9868.2005.00503.x

48. Alvarez MJ, Shen Y, Giorgi FM, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet*. 2016;48(8):838-847. doi:10.1038/ng.3593

49. Savino A, Avalle L, Monteleone E, et al. Network analysis allows to unravel breast cancer molecular features and to identify novel targets. *bioarXiv*. Published online 2019:1-20.

doi:10.1101/570051

50. Saez-Rodriguez J, Alexopoulos LG, Zhang MS, Morris MK, Lauffenburger DA, Sorger PK. Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. *Cancer Res*. 2011;71(16):5400-5411. doi:10.1158/0008-5472.CAN-10-4453

51. Tang J, Gautam P, Gupta A, et al. Network pharmacology modeling identifies synergistic Aurora B and ZAK interaction in triple-negative breast cancer. *npj Syst Biol Appl*. 2019;5(1):1-11. doi:10.1038/s41540-019-0098-z

52. Saez-Rodriguez J, Alexopoulos LG, Epperlein J, et al. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol Syst Biol*. 2009;5. doi:10.1038/msb.2009.87

53. Echeverría PC, Bernthaler A, Dupuis P, Mayer B, Picard D. An interaction network predicted from public data as a discovery tool: Application to the Hsp90 molecular chaperone machine. Cotterill S, ed. *PLoS One*. 2011;6(10):e26044. doi:10.1371/journal.pone.0026044

54. Tang J, Karhinen L, Xu T, et al. Target Inhibition Networks: Predicting Selective Combinations of Druggable Targets to Block Cancer Survival Pathways. *PLoS Comput Biol*. 2013;9(9). doi:10.1371/journal.pcbi.1003226

55. Lehne B, Schlitt T. Protein-protein interaction databases: keeping up with growing interactomes. *Hum Genomics*. 2009;3(3):291-297. doi:10.1186/1479-7364-3-3-291

56. Drake JM, Paull EO, Graham NA, et al. Phosphoproteome Integration Reveals Patient-Specific Networks in Prostate Cancer. *Cell*. 2016;166(4):1041-1054. doi:10.1016/j.cell.2016.07.007

57. Park S, Kim YS, Kim DY, So I, Jeon JH. PI3K pathway in prostate cancer: All resistant roads lead to PI3K. *Biochim Biophys Acta - Rev Cancer*. 2018;1870(2):198-206. doi:10.1016/j.bbcan.2018.09.001

58. Gururajan M, Cavassani KA, Sievert M, et al. SRC family kinase FYN promotes the neuroendocrine phenotype and visceral metastasis in advanced prostate cancer. *Oncotarget*. 2015;6(42):44072-44083. doi:10.18632/oncotarget.6398

59. Chang YM, Bai L, Liu S, Yang JC, Kung HJ, Evans CP. Src family kinase oncogenic potential and pathways in prostate cancer as revealed by AZD0530. *Oncogene*. 2008;27(49):6365-6375. doi:10.1038/onc.2008.250

60. Bluemn EG, Coleman IM, Lucas JM, et al. Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. *Cancer Cell*. 2017;32(4):474-489.e6. doi:10.1016/j.ccell.2017.09.003

61. Lin JZ, Wang ZJ, De W, et al. Targeting AXL overcomes resistance to docetaxel therapy in advanced prostate cancer. *Oncotarget*. 2017;8(25):41064-41077. doi:10.18632/oncotarget.17026

62. Kaarbø M, Mikkelsen ØL, Malerød L, et al. PI3K-AKT-mTOR pathway is dominant over androgen receptor signaling in prostate cancer cells. *Cell Oncol*. 2010;32(1-2):11-27. doi:10.3233/CLO-2009-0487

63. Feng S, Shao L, Castro P, et al. Combination treatment of prostate cancer with FGF receptor and AKT kinase inhibitors. *Oncotarget*. 2017;8(4):6179-6192. doi:10.18632/oncotarget.14049

64. Goc A, Al-Husein B, Kochuparambil ST, Liu J, Heston WWD, Somanath PR. PI3 kinase integrates Akt and MAP kinase signaling pathways in the regulation of prostate cancer. *Int*

*J Oncol*. 2011;38(1):267-277. Accessed May 14, 2018. https://www.spandidos-publications.com/ijo/38/1/267

65. Bello T, Gujral TS. KInhibition: A Kinase Inhibitor Selection Portal. *iScience*. 2018;8:49-53. doi:10.1016/j.isci.2018.09.009

66. Bain J, Plater L, Elliott M, et al. The selectivity of protein kinase inhibitors: a further update. *Biochem J*. 2007;408(3):297-315. doi:10.1042/BJ20070797

67. Koleti A, Terryn R, Stathias V, et al. Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Res*. Published online November 13, 2017. doi:10.1093/nar/gkx1063

68. Cheng AC, Eksterowicz J, Geuns-Meyer S, Sun Y. Analysis of kinase inhibitor selectivity using a thermodynamics-based partition index. *J Med Chem*. 2010;53(11):4502-4510. doi:10.1021/jm100301x

69. Graczyk PP. Gini coefficient: A new way to express selectivity of kinase inhibitors against a family of kinases. *J Med Chem*. 2007;50(23):5773-5779. doi:10.1021/jm070562u

70. Uitdehaag JC, Zaman GJ. A theoretical entropy score as a single value to express inhibitor selectivity. *BMC Bioinformatics*. 2011;12(1):94. doi:10.1186/1471-2105-12-94

71. Rao S, Du G, Hafner M, Subramanian K, Sorger PK, Gray NS. A multi-targeted probe-based strategy to identify signaling vulnerabilities in cancers. *J Biol Chem*. Published online 2019:jbc.RA118.006805. doi:10.1074/jbc.RA118.006805

72. Knapp S. New opportunities for kinase drug repurposing and target discovery. *Br J Cancer*. 2018;118(7):936-937. doi:10.1038/s41416-018-0045-6

73. Zhou AY, Johnson DB. Combinatorial Therapies in Melanoma: MAPK Inhibitors and

Beyond. *Am J Clin Dermatol*. 2018;19(2):181-193. doi:10.1007/s40257-017-0320-y

74. Levis M. Midostaurin approved for FLT3-mutated AML. *Blood*. 2017;129(26):3403-3406. doi:10.1182/blood-2017-05-782292

75. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer Statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7-34. doi:10.3322/canjclin.57.1.43

76. Gandaglia G, Abdollah F, Schiffmann J, et al. Distribution of metastatic sites in patients with prostate cancer: A population-based analysis. *Prostate*. 2014;74(2):210-216. doi:10.1002/pros.22742

77. Gao H, Ouyang X, Banach-Petrosky WA, Gerald WL, Shen MM, Abate-Shen C. Combinatorial activities of Akt and B-Raf/Erk signaling in a mouse model of androgen-independent prostate cancer. *Proc Natl Acad Sci U S A*. 2006;103(39):14477-14482. doi:10.1073/pnas.0606836103

78. Lara PN, Evans CP. Dasatinib and docetaxel in advanced prostate cancer. *Lancet Oncol*. 2013;14(13):1248-1249. doi:10.1016/S1470-2045(13)70500-X

79. Wise HM, Hermida MA, Leslie NR. Prostate cancer, PI3K, PTEN and prognosis. *Clin Sci*. 2017;131(3):197-210. doi:10.1042/CS20160026

80. Nevedomskaya E, Baumgart SJ, Haendler B. Recent advances in prostate cancer treatment and drug discovery. *Int J Mol Sci*. 2018;19(5):1359. doi:10.3390/ijms19051359

81. Crumbaker M, Khoja L, Joshua AM. AR signaling and the PI3K pathway in prostate cancer. *Cancers (Basel)*. 2017;9(4). doi:10.3390/cancers9040034

82. Ciccarese C, Santoni M, Brunelli M, et al. AR-V7 and prostate cancer: The watershed for treatment selection? *Cancer Treat Rev*. 2016;43:27-35. doi:10.1016/j.ctrv.2015.12.003

83. Chandrasekar T, Yang JC, Gao AC, Evans CP. *Mechanisms of Resistance in Castration-*

*Resistant Prostate Cancer (CRPC)*. Vol 6. AME Publications; 2015. doi:10.3978/j.issn.2223-4683.2015.05.02

84. Drake JM, Graham NA, Lee JK, et al. Metastatic castration-resistant prostate cancer reveals intrapatient similarity and interpatient heterogeneity of therapeutic kinase targets. *Proc Natl Acad Sci*. 2013;110(49):E4762-E4769. doi:10.1073/pnas.1319948110

85. Krzywinski M, Altman N. Points of Significance: Multiple linear regression. *Nat Methods*. 2015;12(12):1103-1104. doi:10.1038/nmeth.3665

86. Russell P, Kingsley E. Human Prostate Cancer Cell Lines. *Methods Mol Med Prostate Cancer Methods Protoc*. 2003;81(12):21-39. doi:10.1385/1-59259-372-0:21

87. Cunningham D, You Z. In vitro and in vivo model systems used in prostate cancer research. *J Biol Methods*. 2015;2(1):17. doi:10.14440/jbm.2015.63

88. Sartorius. IncuCyte ZOOM. Accessed September 4, 2019. https://www.essenbioscience.com/en/products/incucyte/

89. Gujral TS, Kirschner MW. Hippo pathway mediates resistance to cytotoxic drugs. *Proc Natl Acad Sci*. Published online 2017:E3729-E3738. doi:10.1073/pnas.1703096114

90. Nam S, Kim D, Cheng JQ, et al. Action of the Src Family Kinase Inhibitor, Dasatinib (BMS-354825), on Human Prostate Cancer Cells. *Cancer Res*. 2005;65(20):9185-9189. doi:10.1158/0008-5472.CAN-05-1731

91. Koreckij T, Nguyen H, Brown LG, Yu EY, Vessella RL, Corey E. Dasatinib inhibits the growth of prostate cancer in bone and provides additional protection from osteolysis. *Br J Cancer*. 2009;101(2):263-268. doi:10.1038/sj.bjc.6605178

92. Che H yong, Guo H yuan, Si X wei, You Q ying, Lou W ying. PP121, a dual inhibitor of tyrosine and phosphoinositide kinases, inhibits anaplastic thyroid carcinoma cell

proliferation and migration. *Tumor Biol*. 2014;35(9):8659-8664. doi:10.1007/s13277-014-2118-3

93. Vlahos CJ, Matter WF, Hui KY, Brown RF. A specific inhibitor of phosphatidylinositol 3-kinase, 2-(4-morpholinyl)-8-phenyl-4H-1-benzopyran-4-one (LY294002). *J Biol Chem*. 1994;269(7):5241-5248. Accessed September 5, 2019. http://www.ncbi.nlm.nih.gov/pubmed/8106507

94. Araujo JC, Mathew P, Armstrong AJ, et al. Dasatinib combined with docetaxel for castration-resistant prostate cancer: Results from a phase 1-2 study. *Cancer*. 2012;118(1):63-71. doi:10.1002/cncr.26204

95. Araujo JC, Trudel GC, Saad F, et al. Docetaxel and dasatinib or placebo in men with metastatic castration-resistant prostate cancer (READY): a randomised, double-blind phase 3 trial. *Lancet Oncol*. 2013;14(13):1307-1316. doi:10.1016/S1470-2045(13)70479-0

96. Chen S, Do JT, Zhang Q, et al. Self-renewal of embryonic stem cells by a small molecule. *Proc Natl Acad Sci U S A*. 2006;103(46):17266-17271. doi:10.1073/pnas.0608156103

97. Espina V, Wulfkuhle J, Calvert VS, Liotta LA, Petricoin EF. Reverse Phase Protein Microarrays for Theranostics and Patient-Tailored Therapy. In: *Methods in Molecular Biology (Clifton, N.J.)*. Vol 441. ; 2008:113-128. doi:10.1007/978-1-60327-047-2_8

98. Gujral TS, Karp RL, Finski A, et al. Profiling phospho-signaling networks in breast cancer using reverse-phase protein arrays. *Oncogene*. 2013;32(29):3470-3476. doi:10.1038/onc.2012.378

99. Carver BS, Chapinski C, Wongvipat J, et al. Reciprocal Feedback Regulation of PI3K and Androgen Receptor Signaling in PTEN-Deficient Prostate Cancer. *Cancer Cell*. 2011;19(5):575-586. doi:10.1016/j.ccr.2011.04.008

100. Varkaris A, Katsiampoura AD, Araujo JC, Gallick GE, Corn PG. Src signaling pathways in prostate cancer. *Cancer Metastasis Rev*. 2014;33(2-3):595-606. doi:10.1007/s10555-013-9481-1

101. Kosaka T, Miyajima A, Shirotake S, Suzuki E, Kikuchi E, Oya M. Long-term androgen ablation and docetaxel up-regulate phosphorylated Akt in castration resistant prostate cancer. *J Urol*. 2011;185(6):2376-2381. doi:10.1016/j.juro.2011.02.016

102. Mulholland DJ, Tran LM, Li Y, et al. Cell autonomous role of PTEN in regulating castration-resistant prostate cancer growth. *Cancer Cell*. 2011;19(6):792-804. doi:10.1016/j.ccr.2011.05.006

103. Kolde R. pheatmap: Pretty Heatmaps. Published online 2015. https://cran.r-project.org/package=pheatmap

104. Moutsatsos IK, Parker CN. Recent advances in quantitative high throughput and high content data analysis. *Expert Opin Drug Discov*. 2016;11(4):415-423. doi:10.1517/17460441.2016.1154036

105. D'Argenio V. The high-throughput analyses era: Are we ready for the data struggle? *High-Throughput*. 2018;7(1):8. doi:10.3390/ht7010008

106. Türei D, Korcsmáros T, Saez-Rodriguez J. OmniPath: Guidelines and gateway for literature-curated signaling pathway resources. *Nat Methods*. 2016;13(12):966-967. doi:10.1038/nmeth.4077

107. Gujral TS, Chan M, Peshkin L, Sorger PK, Kirschner MW, Macbeath G. A noncanonical frizzled2 pathway regulates epithelial-mesenchymal transition and metastasis. *Cell*. 2014;159(4):844-856. doi:10.1016/j.cell.2014.10.032

108. Golkowski M, Lau H-T, Chan M, et al. Pharmacoproteomics Identifies Kinase Pathways

that Drive the Epithelial-Mesenchymal Transition and Drug Resistance in Hepatocellular Carcinoma. *Cell Syst*. 2020;11(2):196-207.e7. doi:10.1016/j.cels.2020.07.006

109. Kanehisa M. (KEGG) Toward pathway engineering: A new database of genetic and molecular pathways. *Sci Technol Japan*. 1996;59:34-38. Accessed February 24, 2019. http://www.genome.ad.jp/

110. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45(D1):D362-D368. doi:10.1093/nar/gkw937

111. Cerami E, Gao J, Dogrusoz U, et al. The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-404. doi:10.1158/2159-8290.CD-12-0095

112. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):1-1. doi:10.1126/scisignal.2004088

113. Xue AG, Chan M, Gujral TS. Pan-cancer analysis of the developmental pathways reveals non-canonical wnt signaling as a driver of mesenchymal-type tumors. *Transl Res*. Published online June 6, 2020. doi:10.1016/j.trsl.2020.06.003

114. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal*. 2006;Complex Sy:1695. Accessed July 30, 2020. http://igraph.org

115. Costa A, Petrenko AA, Guizien K, Doglioli AM. On the calculation of betweenness centrality in marine connectivity studies using transfer probabilities. *PLoS One*. 2017;12(12). doi:10.1371/journal.pone.0189021

116. Freeman LC. Centrality in social networks conceptual clarification. *Soc Networks*.

1978;1(3):215-239. doi:10.1016/0378-8733(78)90021-7

117. Pržulj N, Wigle DA, Jurisica I. Functional topology in a network of protein interactions. *Bioinformatics*. 2004;20(3):340-348. doi:10.1093/bioinformatics/btg415

118. Gibbs DL, Shmulevich I. Solving the influence maximization problem reveals regulatory organization of the yeast cell cycle. Maini PK, ed. *PLoS Comput Biol*. 2017;13(6):e1005591. doi:10.1371/journal.pcbi.1005591

119. Jalili M, Salehzadeh-Yazdi A, Gupta S, et al. Evolution of centrality measurements for the detection of essential proteins in biological networks. *Front Physiol*. 2016;7(AUG):375. doi:10.3389/fphys.2016.00375

120. Santolini M, Barabási AL. Predicting perturbation patterns from the topology of biological networks. *Proc Natl Acad Sci U S A*. 2018;115(27):E6375-E6383. doi:10.1073/pnas.1720589115

121. Mathews Griner LA, Guha R, Shinn P, et al. High-throughput combinatorial screening identifies drugs that cooperate with ibrutinib to kill activated B-cell-like diffuse large B-cell lymphoma cells. *Proc Natl Acad Sci U S A*. 2014;111(6):2349-2354. doi:10.1073/pnas.1311846111

122. The Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell*. 2017;169(7):1327-1341.e23. doi:10.1016/j.cell.2017.05.046

123. Jansen S, Gosens R, Wieland T, Schmidt M. Paving the Rho in cancer metastasis: Rho GTPases and beyond. *Pharmacol Ther*. 2018;183:1-21. doi:10.1016/j.pharmthera.2017.09.002

124. Xia L, Lin J, Su J, et al. Diallyl disulfide inhibits colon cancer metastasis by suppressing

Rac1-mediated epithelial-mesenchymal transition. *Onco Targets Ther*. 2019;12:5713-5728. doi:10.2147/OTT.S208738

125. Patel A, Sabbineni H, Clarke A, Somanath PR. Novel roles of Src in cancer cell epithelial-to-mesenchymal transition, vascular permeability, microinvasion and metastasis. *Life Sci*. 2016;157:52-61. doi:10.1016/j.lfs.2016.05.036

126. Steger M, Diez F, Dhekne HS, et al. Systematic proteomic analysis of LRRK2-mediated rab GTPase phosphorylation establishes a connection to ciliogenesis. *Elife*. 2017;6. doi:10.7554/eLife.31012

127. Steger M, Tonelli F, Ito G, et al. Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *Elife*. 2016;5(JANUARY2016). doi:10.7554/eLife.12813.001

128. Sangodkar J, Farrington CC, McClinch K, Galsky MD, Kastrinsky DB, Narla G. All roads lead to PP2A: Exploiting the therapeutic potential of this phosphatase. *FEBS J*. 2016;283(6):1004-1024. doi:10.1111/febs.13573

129. Fuhrmann J, Subramanian V, Kojetin DJ, Thompson PR. Activity-Based Profiling Reveals a Regulatory Link between Oxidative Stress and Protein Arginine Phosphorylation. *Cell Chem Biol*. 2016;23(8):967-977. doi:10.1016/j.chembiol.2016.07.008

130. Wu Q, Heidenreich D, Zhou S, et al. A chemical toolbox for the study of bromodomains and epigenetic signaling. *Nat Commun*. 2019;10(1):1-14. doi:10.1038/s41467-019-09672-2

131. Fuhrmann J, Clancy KW, Thompson PR. Chemical Biology of Protein Arginine Modifications in Epigenetic Regulation. *Chem Rev*. 2015;115(11):5413-5461.

doi:10.1021/acs.chemrev.5b00003

132. Yates AD, Achuthan P, Akanni W, et al. Ensembl 2020. *Nucleic Acids Res*. 2020;48(D1):D682-D688. doi:10.1093/nar/gkz966

133. Braschi B, Denny P, Gray K, et al. Genenames.org: The HGNC and VGNC resources in 2019. *Nucleic Acids Res*. 2019;47(D1):D786-D792. doi:10.1093/nar/gky930

134. Wang H, Qian WJ, Mottaz HM, et al. Development and evaluation of a micro- and nanoscale proteomic sample preparation method. *J Proteome Res*. 2005;4(6):2397-2403. doi:10.1021/pr050160f

135. Golkowski M, Vidadala RSR, Lombard CK, Suh HW, Maly DJ, Ong SE. Kinobead and Single-Shot LC-MS Profiling Identifies Selective PKD Inhibitors. *J Proteome Res*. 2017;16(3):1216-1227. doi:10.1021/acs.jproteome.6b00817

136. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen J V, Mann M. Andromeda: A peptide search engine integrated into the MaxQuant environment. *J Proteome Res*. 2011;10(4):1794-1805. doi:10.1021/pr101065j

137. Tyanova S, Temu T, Sinitcyn P, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*. 2016;13(9):731-740. doi:10.1038/nmeth.3901