

PRECISION MEDICINE IN THE AGE OF "BIG DATA":
LEVERAGING MACHINE LEARNING AND GENOMICS FOR
DRUG DISCOVERIES

A Dissertation

Presented to the Faculty of the Weill Cornell Graduate School
of Medical Sciences

in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

By

Kaitlyn M. Gayvert

August 2017

© 2017 Gayvert

PRECISION MEDICINE IN THE AGE OF "BIG DATA":
LEVERAGING MACHINE LEARNING AND GENOMICS FOR
DRUG DISCOVERIES

Kaitlyn M. Gayvert, Ph.D.

Cornell University 2017

Targeted therapies designed to specifically target molecules involved in carcinogenesis have achieved remarkable antitumor efficacy. However resistance inevitably develops and many cancer patients are not candidates for these targeted therapies. Furthermore the clinical attrition rate continues to rise, which remains a barrier in the development of novel targeted therapies. Integration of extensive genomics datasets with large drug databases allows us to begin to tackle questions about target discovery and drug toxicity with the ultimate goal of accelerating personalized anticancer drug discovery. The purpose of this dissertation was to address these problems through the development of drug repurposing, toxicity prediction, and drug synergy prediction models.

First to target the role of transcription factors as drivers of oncogenic activity, we developed a computational drug repositioning approach (CRAFTT) that makes predictions about drugs that specifically disrupt transcription factor activity. To do this, CRAFTT integrates transcription factor binding site information with drug-induced expression profiling. We found that CRAFTT was able to recover a significant number of known drug-transcription factor interactions and identified a novel interaction that we subsequently validated. Our work in drug discovery led us to ask

questions about what makes a drug safe. We developed a data-driven approach (PrOCTOR) that integrates the properties of a compound's targets and its structure to directly predict the likelihood of toxicity in clinical trials and was able to accurately classify known safe and toxic drugs. Finally to address the problem of drug resistance, we developed a machine learning approach to identify synergistic and effective drug combinations based on single drug efficacy information and limited drug combination testing. When applied to mutant BRAF melanoma, this approach exhibited significant predictive power upon evaluation with cross-validation and further experimental testing of previously untested drug combinations in cell lines independent of the training set.

Altogether this work demonstrates how the integration of orthogonal datasets gives us power to address difficult questions that are critical for precision medicine and drug discovery. Approaches such as these have the potential to make a direct impact on how patients are treated, as well as to help prioritize and guide additional focused studies.

BIOGRAPHICAL SKETCH

Kaitlyn Gayvert graduated from the State University of New York at Geneseo in 2012 with a Bachelors of Arts in Mathematics. During this time, she researched population and tumor growth modeling problems under the mentorship of Drs. Christopher Leary, Caroline Haddad and Gregg Hartvigsen. She also participated in a Research Experience for Undergraduates program at North Carolina State University in 2011, working under mentorship of Dr. H.T. Banks on disease modeling and algorithm selection. She joined the Tri-Institutional Training Program in Computational Biology and Medicine Program in 2012 and became a member of the Elemento lab in 2013, where her work has focused on the development of predictive methods for drug repositioning, toxicity, and synergy.

ACKNOWLEDGEMENTS

I would like to thank the Elemento, Rickman, and Mezey lab members, the members of the Tri-Institutional Computational Biology and Medicine program, and my committee members John Chodera, Olga Boudker and Ekta Khurana for their feedback and discussions. This work was supported by the CAREER grant from National Science Foundation (DB1054964), NIH grant R01CA194547, the Starr Cancer Foundation, as well as by startup funds from the Institute for Computational Biomedicine. Support was also provided by the PhRMA Foundation Pre Doctoral Informatics Fellowship and by the Tri-Institutional Training Program in Computational Biology and Medicine (via NIH training grant T32GM083937).

TABLE OF CONTENTS

Biographical Sketch	iii
Acknowledgements	iv
Table of Contents	v
List of Figures	vi
List of Tables	vii
Chapter One. Introduction	1
Chapter Two. A computational drug repositioning approach for targeting oncogenic transcription factors	4
Chapter Three. A data-driven approach to predicting drug toxicity	27
Chapter Four. A computational approach for identifying synergistic drug combinations	44
Perspective	60
Appendix	62
References	70

LIST OF FIGURES

Figure 2.1. Analysis of COSMIC and TCGA reveals high prevalence of transcription factor mutations in cancer	7
Figure 2.2. CRAFTT Methodology Overview	10
Figure 2.3. Systematic analysis of CRAFTT Predictions	13
Figure 2.4. Identification of dexamethasone as a candidate drug for inhibition of ERG activity	18
Figure 2.5. Experimental support of dexamethasone for modulation of ERG activity	21
Figure 2.6. Extended Analyses of CRAFTT Predictions	24
Figure 3.1. PrOCTOR Method Schematic	30
Figure 3.2. Distributions of select features	31
Figure 3.3. Benchmarking PrOCTOR's performance	34
Figure 3.4. Correlation of Side Effects with PrOCTOR Score	38
Figure 3.5. PrOCTOR Feature Importance and Model Robustness	40
Figure 4.1. Drug Combination Feature construction schematic	48
Figure 4.2. Method schematic and Evaluation of individual model performance	50
Figure 4.3. Predictions for Previously Untested Combinations	52
Figure 4.4. Experimental Validation of Predicted BRAF Effective and Synergistic Combinations	55
Figure 4.5. Identification of Synergistic Combinations involving the BRAF Inhibitor PLX4720	56

LIST OF TABLES

Table 4.1. Drug Combination Model Performance	48
Table 4.2. Drug Combination Experimental Validation	53

CHAPTER ONE

INTRODUCTION

Over the past few decades, great strides have been made in the treatment of cancer through the adoption of precision medicine approaches. One major effort of precision medicine is the greater application of targeted therapies, which seek to selectively kill tumor cells. However there are many challenges associated with the development and application of these therapies, including identification of tractable targets(Gashaw et al., 2011), challenges with drug toxicity(Ledford, 2011), and drug resistance(Fitzgerald et al., 2006; Komarova et al., 2013). Furthermore the rising clinical attrition rate due to biological activity and safety issues is a major hurdle for the development of new compounds(Ledford, 2011). Therefore approaches that rescue compounds that lack efficacy or identify toxic compounds before expensive preclinical and clinical studies have the potential to be highly impactful.

Drug repositioning (or repurposing) is the process of finding new uses for existing drugs(Li et al., 2012). Drug repositioning is especially advantageous in terms of cost and time efficiency when applied to drugs that have passed human-safety and toxicity conditions(Hurle et al., 2013; Li et al., 2012). However, historically, drug repositioning has typically been done through a target-based discovery method based on a priori mechanistic data(Hurle et al., 2013). Computational repositioning methods have begun to emerge largely due to an expansion of available data resources(Hurle et al., 2013), e.g. Connectivity Map (CMaP)(Lamb et al., 2006) and ENCODE(Encode Project Consortium, 2011). Common computational drug repositioning methods include transcriptomic methods, which identify existing drugs whose transcriptional profile is inversely and unexpectedly

correlated to disease expression signatures(Campbell et al., 2012; Dudley et al., 2011; Hurle et al., 2013) and methods based on drug side-effects, which seek to identify drugs that unexpectedly share side-effects with drugs used against a given disease and therefore likely share activity against the disease(Hurle et al., 2013). There are also many efforts underway that aim to better elucidate the mechanisms, targets and effects of drugs. This requires the knowledge of protein-protein interactions and pathways, as a drug's effect on one protein will often impact other interacting proteins. This information is then used to direct drug repositioning approaches.

The identification of toxic drugs before they reach the clinic is another critical unmet need. Drug likeness measures are commonly used in early stages of drug development to weed out compounds with features that are likely to be associated with safety issues, such as poor bioavailability(Leeson et al., 2007). Additionally before drugs enter human trials, toxicology and efficacy are evaluated in animal models. Yet the majority of drugs that have good drug-likeness characteristics and are safe in animal models still fail in human trials(Shanks et al., 2009).

Another challenge that targeted therapies face is the seemingly inevitable development of drug resistance. However it has been proposed that combination therapies have the potential to prevent and overcome drug resistance(Fitzgerald et al., 2006; Komarova et al., 2013). Indeed there have been a number of instances in which drug combinations have been successfully approved and utilized to prevent resistance, most notably in the treatment of hypertension, asthma, and HIV(Foucquier et al., 2015). There is also great interest in utilizing combination therapies for the treatment of cancer and they are FDA approved for use in various cancer types(Foucquier et al., 2015). However combination

therapies are typically identified in highly focused studies based on detailed mechanistic knowledge about each drug. As a result, the large drug combinatorial space remains largely unexplored.

The goal of this thesis is to address these diverse challenges that drug development pipelines currently face. Three methods are described which address drug repositioning, toxicity prediction, and synergy prediction. The first method is a drug repositioning approach that can be used to target drivers of oncogenic activity. The second method is machine learning method that directly predicts whether a compound is likely to have manageable toxicity in clinical trials. The third method is a generalizable machine learning method that can predict drug synergy based on limited combination testing.

CHAPTER TWO

A COMPUTATIONAL DRUG REPOSITIONING APPROACH FOR TARGETING ONCOGENIC TRANSCRIPTION FACTORS*

PREAMBLE

This chapter consists of a paper that was published in *Cell Reports* in June 2016. The method (CRAFTT) was conceived in partnership with Dr. Olivier Elemento. I implemented the method and subsequent computational analyses. The experimental follow-up was done by the Rickman lab (D.R., C.C., E.D.) and the electronic medical record analysis by the Tatonetti lab (N.T., T.L., M.R.B.).

INTRODUCTION

Transcription factors (TFs) are frequently mutated in cancer. These include factors that function in a variety of ways, including nuclear hormone receptors, resident nuclear proteins, and latent cytoplasmic factors (Darnell, 2002). Classic examples of recurrently altered TFs include the tumor suppressor TF gene p53, which is mutated in up to 40% of human tumors (Libermann et al., 2006) and yet has remained a highly elusive target for reactivation (Mees et al., 2009).

Examples also include c-Myc, which is also among the most commonly altered genes in cancer (Ablain et al., 2011), and ERG and other ETS-family factors, which are fused to the androgen-controlled promoters in over 50% of prostate cancer patients (Rickman et al., 2012).

Inhibition of oncogenes and reactivation of tumor-suppressors have become well-established goals in anticancer drug development (Darnell, 2002). Yet TFs are

* Gayvert KM, Dardenne E, Cheung C, et al. A Computational Drug Repositioning Approach for Targeting Oncogenic Transcription Factors. *Cell Rep.* 2016;15(11):2348-56.

generally considered difficult to drug (Mees et al., 2009). If a strategy could be developed for safely and effectively modulating the activity of specific TFs, it would have a broad impact on the treatment of tumor types and subtypes driven by oncogenic TFs. In theory a similar strategy could be applied to reactivate the lost activity of tumor suppressive factors. Potential mechanisms for pharmacological activation or inhibition include disruption of direct DNA binding, perturbation or prevention of the interaction with cofactors and other interacting proteins(Libermann et al., 2006), as well as disruption or activation of upstream signaling mechanisms(Mees et al., 2009). Disrupting interactions with co-factors and other regulatory proteins is broadly viewed as one of the most promising approaches to altering the activity and function of TFs implicated in disease.

One of the first and best-understood successes in disrupting TFs was the identification of the combination of retinoic acid and arsenic trioxide for inhibition of the PML/RARA fusion oncogene in acute promyelocytic leukemia (APL). The PML/RARA fusion results in the repression of many genes, which in turn blocks the differentiation phenotype that is characteristic of APL(Ablain et al., 2011). The retinoic acid-arsenic combination induces PML/RARA degradation which reactivates the silenced genes(Ablain et al., 2011). A small-molecule, JQ1, was recently discovered to inhibit c-Myc and n-Myc, both key regulators of cell proliferation, by inhibiting BET bromodomain proteins which function as regulatory factors for c-Myc and n-Myc(Delmore et al., 2011; Puissant et al., 2013). While important, these studies are based on extremely detailed knowledge of the mechanisms and structures of the co-factors required for TF activity. Such knowledge is not always available and as a result there is no systematic way to identify small molecules that can specifically disrupt TF activity.

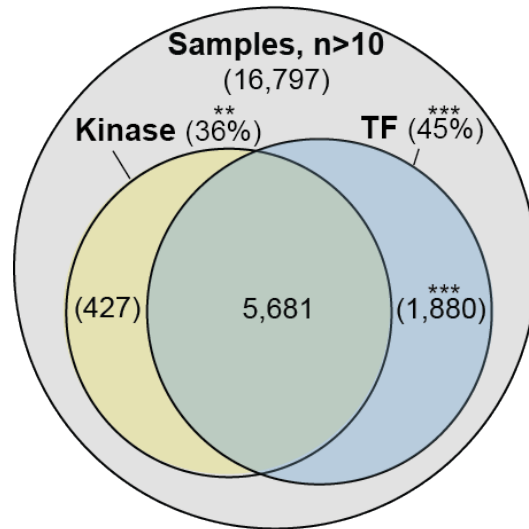
To address this unmet need, we developed CRAFTT, a broadly applicable Computational drug-Repositioning Approach For Targeting Transcription factors. Altogether, our method provides a broadly applicable strategy to identify drugs and small molecules that specifically target the activity of individual TFs. Since a significant number of tumors are driven by oncogenic TFs or have lost tumor suppressive TFs, our approach could potentially have an important impact on the development of new therapeutic strategies. For example, our method may be applicable to other therapeutically elusive factors with oncogenic activity, such as FOXA1 or for reactivating the expression program of tumor suppressive TFs such as p53.

RESULTS

Computational drug repositioning approach rediscovers JQ1 for MYC inhibition

We first set out to quantify the prevalence of somatic mutations in TF genes. We found that 45.1% ($p < 0.001$, Permutation test) of cancer samples in COSMIC reported a mutation in a TF. Furthermore TFs constitute a significant proportion (18.1%) of the genes in the Sanger cancer gene census (**Figure 2.1**). This confirmed that the prevalence of genomic alterations in TF genes in cancer is indeed substantial and further indicates that TFs should constitute a major class of anticancer drug targets.

Transcription Factor Prevalence in COSMIC



n=number of genes in sample

- Samples with Mutations in Kinase Genes
- Samples with Mutations in Transcription Factor Genes

Figure 2.1 - Analysis of COSMIC and TCGA reveals high prevalence of transcription factor mutations in cancer, Related to Figure 1. Frequency of alterations in COSMIC for transcription factor and kinase genes. Statistical significance was assessed for each category for using a permutation test with 1000 random gene sets of the same size. Statistical significance for the comparison of transcription factor to kinase alteration frequency was assessed using the chi-squared test. (** $p < 0.01$, *** $p < 0.001$).

To address this need, we reasoned that if drugs could be identified that specifically disrupt the expression of the direct target genes of a given TF, then these drugs would represent good candidates for perturbing the driving role of that particular TF in cancer. We propose CRAFTT, a Computational drug-Relocation Approach For Targeting Transcription factor activity. CRAFTT consists of two major steps: (1) prediction and (2) prioritization using network analysis.

For the prediction step, we compute a score that represents how the direct targets of a TF are modulated by a particular drug. Direct transcriptional target genes are identified using CHIP-seq binding data. The drug treatment-induced modulation profiles are obtained by analyzing expression profiles from drug perturbation experiments, such as those in the Broad Institute's Connectivity Map (CMap)(Lamb et al., 2006), and generating ranked gene lists by sorting the genes from most down-regulated to most up-regulated upon treatment. For a given TF and drug pair, we implement the Broad Institute's gene set enrichment analysis (GSEA)(Subramanian et al., 2005) approach using the drug-induced ranked gene list and the TF's direct target gene set. Each GSEA analysis yields a normalized enrichment score (*NES*) and corresponding *p*-value indicating whether the TF target gene set is mobilized as a whole by the drug, either towards down-regulation ($NES > 0$) or up-regulation ($NES < 0$). *p*-values are corrected for multiple testing using family-wise error rate (*FWER*) controlling procedures. This multiple testing procedure is applied to each drug perturbation profile individually, correcting across all TF gene sets that we are testing. We consider a drug to be predicted to affect TF activity if the *FWER* adjusted *p*-value for the pair was less than 10% ($FWER < 0.1$).

Next we use network analysis to prioritize the predictions made in the first step of CRAFTT. We reasoned that if many of our predictions are indeed true drug-TF modulatory interactions, the network path between drug and their predicted target TF should be relatively short. This is due to the presumed mechanisms underlying the interaction, which would involve signaling molecules immediately upstream of TFs in signaling pathways and transcriptional co-factors. More broadly, we expected that drug and target TFs would be functionally related and therefore be located in vicinity of each other in a global drug-protein network. We curated a biological network that contains 22,399 protein-coding genes, 6,679 drugs and 170 TFs. The protein-protein interactions represent established interactions (Aksoy et al., 2013; Das et al., 2012; Khurana et al., 2013), which include both physical (protein-protein interactions) and non-physical (phosphorylation, metabolic, signaling, regulatory) interactions. The drug-protein interactions were curated from several drug target databases (Aksoy et al., 2013; Knox et al., 2011).

For each drug-TF pair, we calculated the network path length (shortest path) between the TF and the drug. To account for the biases associated with TFs or drugs with large numbers of targets we calculated a normalized path length, which we defined to be the probability that the path length would be observed given randomized networks that conserved TF and drug degrees (Gobbi et al., 2014). We then generate a final prediction score, which we term the modulation index (MI). The MI is a weighted score that scales the NES score for the drug-TF pair ($NES_{d,TF}$) by the normalized network path length ($NPL_{d,TF}$). We note that the proposed approach does not make any assumptions about the mechanisms by which a drug can disrupt the expression program of TFs (**Figure 2.2A**). Such

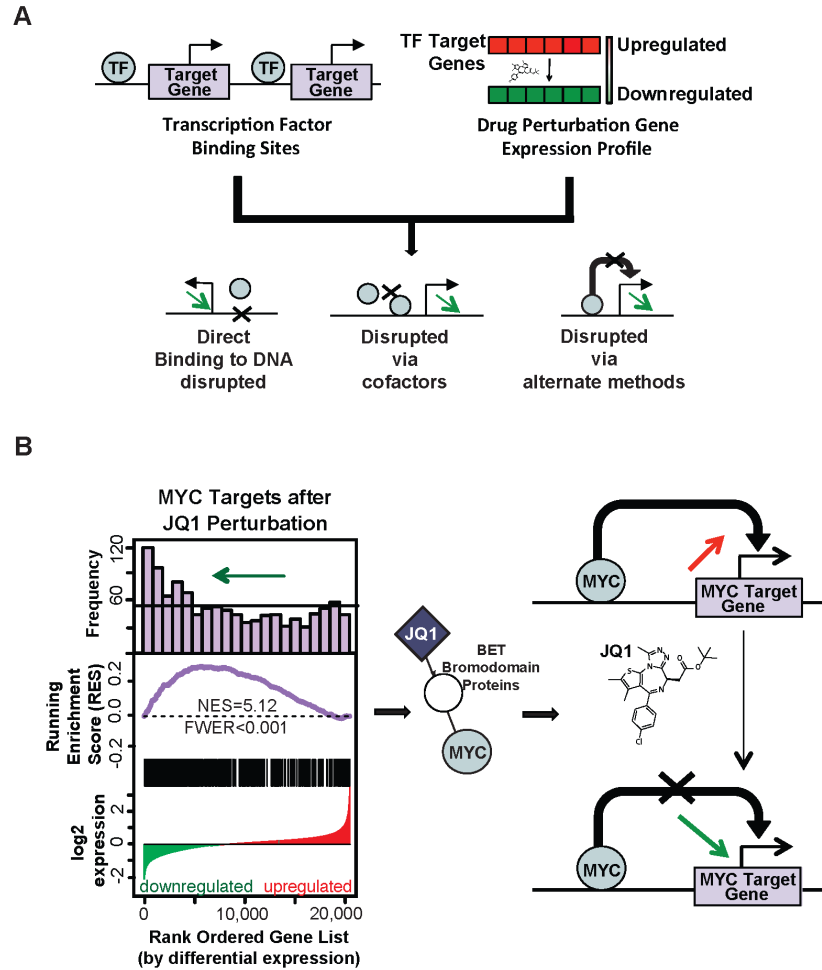


Figure 2.2 - Methodology overview.

- A. Alterations in transcription factors are frequently observed in tumors, leading to aberrant activity. Our method integrates transcriptional binding data and drug-induced gene expression profiles to make predictions about drugs that may affect transcriptional activity. This disruption can occur through a variety of mechanisms, including the inhibition or reactivation of direct binding to DNA or disruption via cofactors.
- B. Application of our method to JQ1 expression profiles and MYC ChIP-seq. The **(left)** panel illustrates the results for the GSEA involving JQ1 and MYC. The lowest plot in the left panel shows the log₂ differential expression profile for JQ1, with the locations of the MYC target genes marked directly above. Directly above that are the running enrichment score and a histogram of the MYC target gene frequency across the drug-induced ranked list, which illustrate whether the MYC target gene set is enriched in the under- or over-expression regions. In the **(middle)** panel, the shortest path between JQ1 and MYC is shown, with BET Bromodomain proteins lying between the two. On the **(right)**, we illustrate that the application of JQ1 results in the downregulation of MYC target genes.

disruption can occur in a variety of ways, e.g., disruption of interaction with co-factors and DNA binding disruption.

As a first proof-of-principle, we applied this approach to JQ1-induced gene expression profiles derived from a recent study (Puissant et al., 2013), all CMap drug-induced expression profiles (Lamb et al., 2006), and to MYC direct target genes, which were derived from ENCODE ChIP-seq data (Encode Project Consortium, 2011). We found that JQ1 significantly down-regulated a substantial fraction (47%) of the 1,250 MYC direct target genes identified by ChIP-seq ($FWER < 0.001$). Furthermore we found that JQ1 had the lowest $FWER$ adjusted p -value, highest enrichment score ($NES = 5.12$) and the shortest possible network path length of 2 given the underlying mechanisms of the true interaction. This indicated that JQ1 is the best candidate ($MI_{JQ1,MYC} = 5120$) out of the 1,310 drugs that we investigated. Thus, as predicted, our method correctly identified the inhibitory effect of JQ1 on MYC-induced transcription (**Figure 2.2B**).

Systematic drug-TF analysis predicts that candidate small molecules can disrupt TFs

We next applied our drug repositioning approach to 166 ChIP-seq experiments from ENCODE (Encode Project Consortium, 2011) and to the 1,309 drug perturbation experiments in CMap (Lamb et al., 2006). This approach identified 37,638 candidate drug-TF pairs (out of 218,603 possible combinations) (**Figure 2.3A**). These candidates included 21,495 predicted activating interactions (a drug induces activation of many direct TF targets) and 16,143 inhibiting interactions (a drug induces repression of many direct TF targets). In particular, there were 1,673 selective predictions involving 49 TFs and 1308 drugs that we have greater confidence in due to the selectivity of the prediction.

Several predicted drug-TF interactions are consistent with the known activity of the drugs involved. For example, all four known HSP90 inhibitors that were both included in our biological network and in CMap were predicted to repress HSF1 activity, which was expected given HSP90's chaperone effect on HSF1 (Conde et al., 2009). These four HSP90 inhibitors were radicicol ($FWER=0.054$), 17-AAG ($FWER=0.031$), 17-DMAG ($FWER=0.085$), and geldanamycin ($FWER<0.001$). Additionally novobiocin, whose antagonism of HSP90 is reported in literature but was not annotated in our network, was also recovered by CRAFTT for disruption of HSF1 ($FWER=0.031$). Novobiocin and geldanamycin had been previously identified to disrupt HSF1 activity through inhibition of HSP90 chaperone activity, operating through the inhibition of HSP90 autophosphorylation for novobiocin and the binding to the HSP90 site in geldanamycin (Conde et al., 2009). We found experimental evidence for numerous other predicted drug-TF interactions for both inhibition and reactivation.

Since experimental validations are not available for the majority of all drug-TF pairs, we turned to network analysis to further evaluate the prediction step of our approach. Within our curated biological network, there were 35 known drug-TF interactions that were also present in both the ENCODE and CMap datasets. The majority of these combinations involved a GR agonist (26 combinations) or a

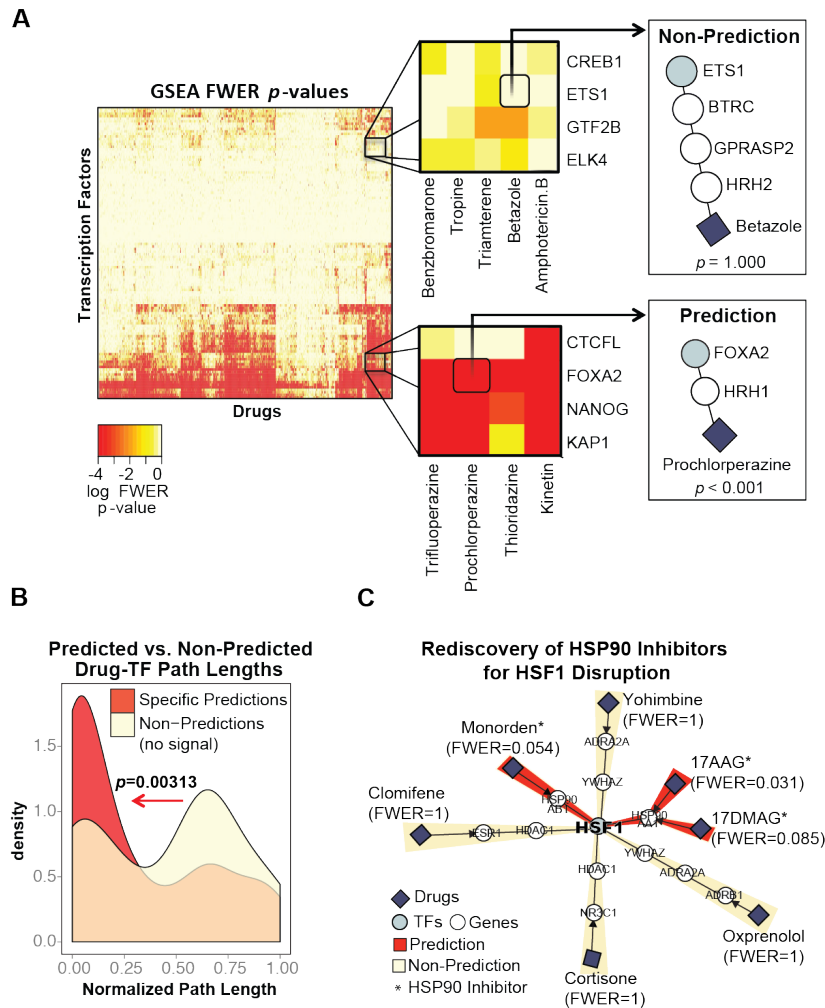


Figure 2.3 - Systematic analysis of 166 TFs and 1309 drug perturbation experiments identifies approximately 38,000 candidate TF-drug pairs.

- Heatmap of the FWER p -values for all TF-drug pairs involving 168 TFs from ENCODE and the 1309 drugs from CMap. In the middle panels, we highlight a subset of non-predictions with high GSEA FWER scores (top) and predictions with low GSEA FWER scores (bottom). On the right, we illustrate that we would expect the candidate TF-drug pairs to have shorter network path lengths than non-predictions. For example, the non-predicted pair ETS1-betazole ($p=1$, GSEA nominal p -value) has a path length of 4 while the predicted pair FOXA2-prochlorperazine ($p<0.001$, GSEA nominal p -value) has a path length of 2.
- Normalized network path lengths for the specific predictions (FWER<0.1) and non-predictions (FWER=1). Statistical significance was evaluated using the Mann-Whitney Test.
- Network visualization of HSF1, all three HSP90 inhibitors covered in CMap and our network (monorden, 17AAG, 17DMAG), and four other drugs not predicted to disrupt HSF1 disruption (clomifene, yohimbine, oxprenolol, cortisone)

HDAC inhibitor (7 combinations). Out of the 35 known drug-TF combinations, CRAFTT was able to correctly predicted more than expected ($n=21$, $p= 1.708e-08$, Binomial Test). In particular, CRAFTT predicted both the GR agonists ($p=6.524e-08$, Binomial Test) and HDAC inhibitors ($p=0.01978$, Binomial Test) well. Furthermore we observed that the drug perturbation profiles within these classes were quite distinct, thus this is not likely due to recovery of the same signal. Additionally, about 85% of these combinations were nominally significant ($p=3.42e-08$), which indicates that our approach was able to identify evidence of the targeting event. The drug-TF pairs that were not rediscovered generally involved drugs or TFs that targeted many genes or were predicted to interact with most other drugs or TFs (non-specific). In general, we found that CRAFTT had limited predictive ability for drugs with more than 25 targets and TFs with more than 2300 target genes.

To further assess CRAFTT's predictive ability, we performed a global network analysis by computing the network path lengths for all drug-TF pairs that were found to be significant ($FWER<0.1$) in the predictive GSEA step of our approach. As described above, we reasoned that true drug-TF interactions should be short given the underlying mechanisms of the interactions (**Figure 2.3A**). Network analysis indeed revealed that the network path lengths (normalized shortest path) of our predicted specific drug-TF pairs were significantly shorter than the path lengths of non-predictions ($FWER=1.0$) ($p=0.00313$, Mann-Whitney test) (**Figure 2.3B**). This is illustrated in **Figure 2.3C** where we show a subnetwork centered on HSF1 that includes drugs connected to HSF1 via one or more intervening proteins. Predicted HSF1 inhibitors by our transcriptomic approach are indeed closer to HSF1 in this subnetwork (red paths) compared to non-predicted molecules (yellow paths). Altogether, this analysis indicates that our

predictions are not random and confirms that many drugs might disrupt TFs by targeting regulatory or interacting co-factors. The network analysis provided increased confidence in our approach's predictive capacity. Moving forward, we used shorter drug-TF paths to further prioritize drug-TF predictions using our combined score (*MI*).

Identification and validation of small molecules that inhibit the ERG TF

We hypothesized that CRAFTT could be used to identify molecules that inhibit the activity of the pro-invasive, oncogenic TF ERG. This is of an interest due to ERG's overexpression resultant of a tissue specific gene fusion event that occurs in as many as 50% of prostate cancer patients. This overexpression results in a pro-invasive phenotype in prostate cancer (Elemento et al., 2012; Rickman et al., 2010; Tomlins et al., 2008). We had previously identified ERG target genes using ChIP-seq in RWPE1 benign prostate cells (Rickman et al., 2012). We therefore applied our approach to all Connectivity Map drug profiles to identify candidate drugs for inhibition of ERG.

From the prediction step of CRAFTT, we identified eight candidate drugs that down-regulate ERG target genes: dexamethasone (*FWER*=0.086), naproxen (*FWER*=0.048), acetaminophen (*FWER*=0.087), ondansetron (*FWER*=0.061), epitiostanol (*FWER*=0.069), diloxanide (*FWER*=0.003), methanthelium bromide (*FWER*=0.046) and isoflupredone (*FWER*=0.088). Five of these candidate drugs were contained in our biological network: dexamethasone (*MI*=1015.85), naproxen (*MI*=530.90), acetaminophen (*MI*= 2167.88), ondansetron (*MI*= 3.35), and epitiostanol (*MI*= 520.99) (**Figure 2.4A**). An initial network analysis suggested

that dexamethasone, naproxen, acetaminophen and epitiostanol were the best candidates due to their large modulation indices.

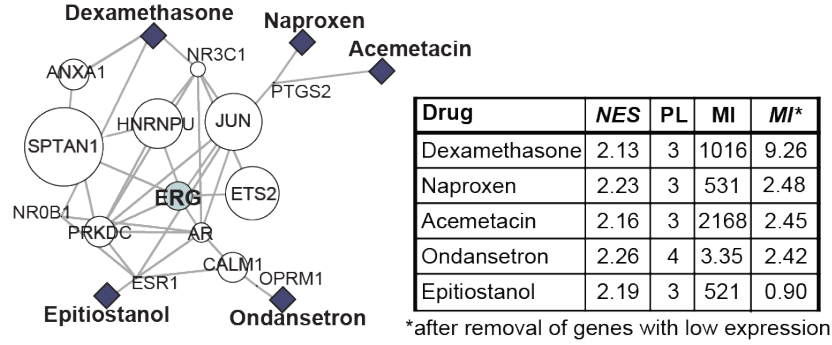
Next we performed an additional analysis to use with our CRAFTT methodology to further prioritize our drug candidate list. We used gene expression (RNAseq) from RWPE1 prostate cells to filter out genes that have low expression in the network, which we defined as RPKM<4. This analysis resulted in dexamethasone being identified as the drug with the shortest path length and highest modified modulation index (MI=9.26, **Figure 2.4B**).

Since dexamethasone has not previously been linked to ERG, we next sought to experimentally test our hypothesis that dexamethasone would be able to reverse ERG-induced oncogenic phenotypes through disruption of ERG in ERG-expressing prostate cancer cells. One of the top target genes that was reversed by dexamethasone in the CMap profile was the urokinase plasminogen activator (*PLAU*), which is a known ERG target gene that has been previously implicated in ERG-mediated cell invasion in multiple cancers and models (Tomlins et al., 2008). We found experimentally that dexamethasone abrogated expression of the ERG target gene *PLAU* in both DU145 cells expressing ERG and in VCaP cells with high endogenous levels of ERG (**Figure 2.4C**). In comparison, dexamethasone was weakly active in the control GFP cells (**Figure 2.4C**).

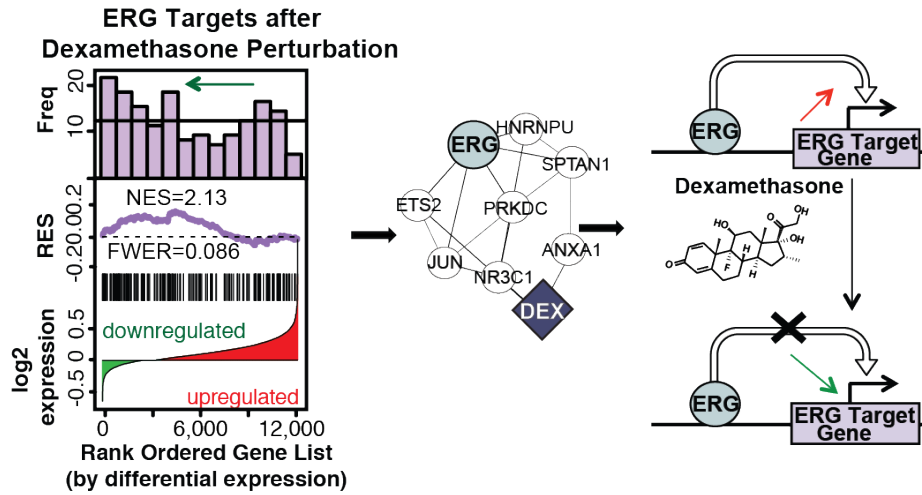
Figure 2.4 - Identification of dexamethasone as a candidate drug for inhibition of ERG activity.

- A. Network visualization illustrating path lengths from ERG to five candidate drugs for ERG inhibition (dexamethasone, naproxen, acetaminophen, ondansetron, epitiostanol). The node sizes correspond to the gene expression levels, with the larger size representing a higher expression level. If low expression genes are removed (RPKM<4), the path lengths for naproxen and acetaminophen are increased while the paths from ondansetron and epitiostanol are completely disrupted. The corresponding table shows metrics that describe each of these drugs in relation to ERG: NES is the Normalized Enrichment Score obtained from GSEA, PL is the shortest network path length required to connect ERG to the drug, MI is the modulation index and MI * is modulation index respectively after low expression genes were removed (RPKM<4).
- B. Application of our method to dexamethasone expression profiles and ERG target genes. The **(left)** panel illustrates the results of the GSEA for ERG and dexamethasone. The lowest plot of the left panel shows the log₂ differential expression profile for dexamethasone, with the ERG target genes marked directly above. Above are the running enrichment score and a histogram of the ERG target gene frequency, which illustrates whether the gene set is enriched in the under- or over-expression regions. The **(middle)** panel shows a subnetwork including all genes that were members of any shortest path between ERG and dexamethasone. The **(right)** panel illustrates our prediction that the application of dexamethasone would result in the downregulation of activity of ERG target genes.
- C. ERG target gene *PLAU* expression by RT-PCR in cell lines expressing ERG (DU145-ERG, VCaP) and controls (DU145-GFP) after treatment with vehicle or dexamethasone. Data are shown as mean ± SEM. Asterisks indicate statistically significant differences by paired t test and n = 3 for each condition (*p < 0.05, **p < 0.01, ***p < 0.001, ns - not significant).
- D. Cell invasion and migration in cell lines expressing ERG (DU145-ERG) and controls (DU145-GFP). The data are shown as mean ± SEM and at n=4 representation 10x field of view. Asterisks indicate statistically significant differences by paired t test and n = 3 for each condition (*p < 0.05, **p < 0.01, ***p < 0.001, ns - not significant).
- E. The binding of ERG and a control (IgG) by ChIP-PCR at the promoter of its target gene *PLAU* and at a negative control (ARHGEF) in cell lines expressing ERG (DU145-ERG, VCaP). Data are shown as mean ± SEM. Asterisks indicate statistically significant differences by paired t test and n = 3 for each condition (*p < 0.05, **p < 0.01, ***p < 0.001, ns - not significant).

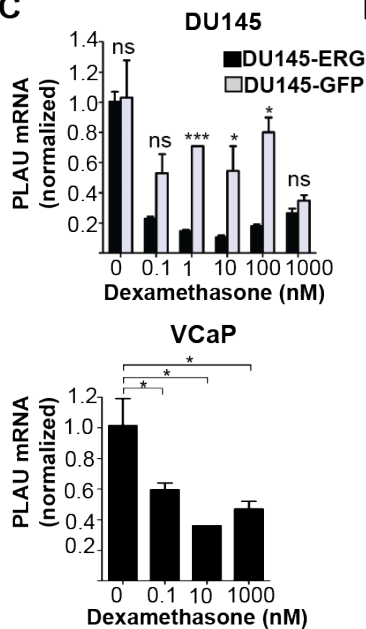
A



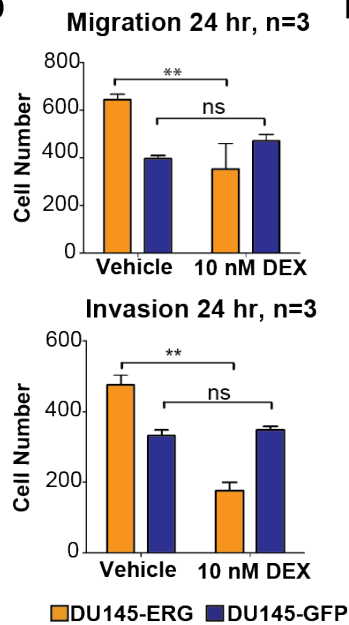
B



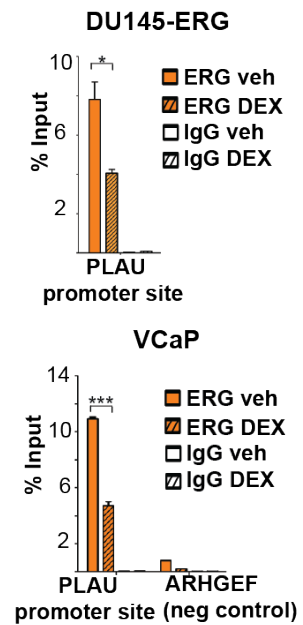
C



D



E



To further test the inhibitory effect of dexamethasone on ERG activity, we treated a newly derived ERG over-expressing cell line derived from PTEN^{-/-}/ERG^{Rosa26} prostate tumors in transgenic mice (Chen et al., 2013). Consistent with the commercially available human prostate cancer cells, dexamethasone treatment resulted in a dose-dependent decrease in mouse *PLAU* mRNA expression (**Figure 2.5A**).

Using cell invasion and migration assays, we then found that dexamethasone significantly decreased cell invasion and migration in DU145 prostate cancer cells over-expressing ERG, but not in isogenic control cells (**Figure 2.4D, Figure 2.5B**). High-resolution microscopic images revealed that dexamethasone helps the cells partially regain polarity, which may be a potential mechanism for reduced cell invasion (**Figure 2.5C**). As expected from published literature on the mostly invasive oncogenic role of ERG, we found that ERG inhibition via dexamethasone treatment had no effect on cell viability *in vitro* (**Figure 2.5D**). Finally, we found using ChIP-PCR that dexamethasone substantially decreased binding of ERG at the *PLAU* promoter in both DU145-ERG and VCaP cells (**Figure 2.4E**). Altogether, these experimental results support CRAFTT's computationally derived prediction that dexamethasone inhibits ERG activity.

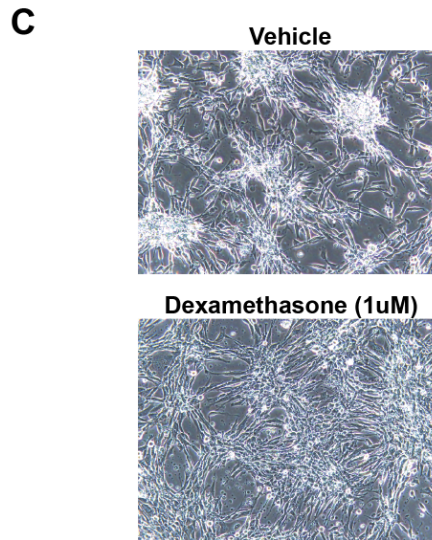
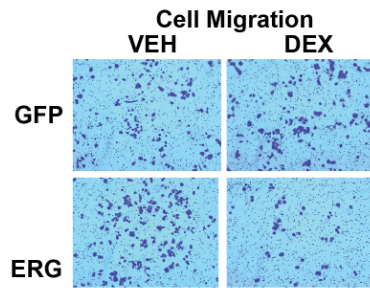
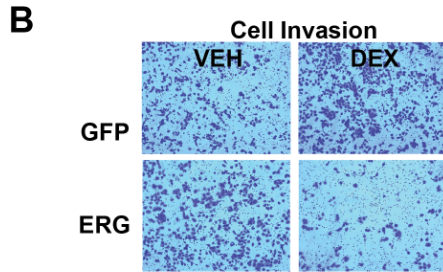
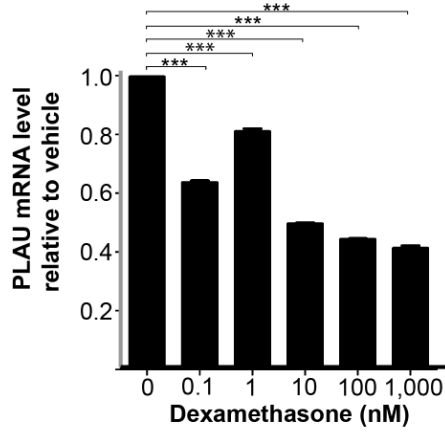
CRAFTT's predicted Dexamethasone-ERG interaction is independent of AR and GR

Dexamethasone is a glucocorticoid receptor (GR) agonist, which suggests that GR, encoded by *NR3C1*, may play a role in ERG-mediated gene expression. We found that siRNAs targeting *NR3C1* mRNA lowered GR levels by 80% in the DU145-ERG cells (**Figure 2.5E**). Although GR seems to play a role in *PLAU* regulation in the absence of ERG, lowering GR levels did not significantly alter

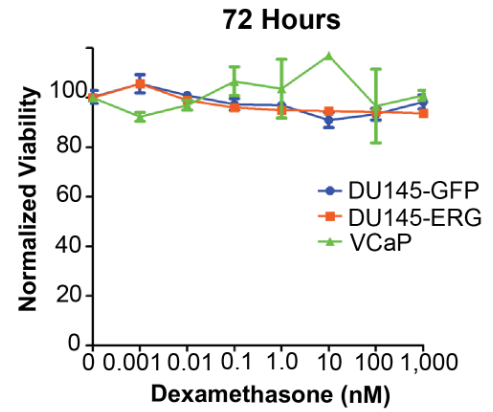
Figure 2.5 - Experimental support of dexamethasone for modulation of ERG activity.

- A. PLAU expression by RT-PCR after treatment with dexamethasone in mouse ERG over-expressing cell lines derived from prostate cancer tumors in transgenic mice. Data are shown as mean \pm SEM. Asterisks indicate statistically significant differences by paired t test for each dosage compared to 0nM (**p < 0.01, ***p < 0.001).
- B. Cell invasion and in DU145 ERG-expressing cell lines and isogenic controls after treatment with vehicle or dexamethasone.
- C. Imaging in ERG/PTEN cells after treatment with vehicle or dexamethasone.
- D. Dose response curves following a 72 hr incubation at the indicated dose of dexamethasone for DU145 clones stably over-expressing ERG (orange) or GFP (blue) or VCaP (green) cells.
- E. DU145 cells were treated with siRNAs targeting NR3C1 mRNA that lowered GR levels by 80% in the DU145-ERG cells. Data are shown as mean \pm SEM. Asterisks indicate statistically significant differences by two-tailed paired Student's t test (*p < 0.05, **p < 0.01, ***p < 0.001, ns – not significant).
- F. DU145 cells were treated with siRNAs targeting NR3C1 mRNA or controls. PLAU expression in DU145 cell lines expressing ERG and controls (GFP) was quantified using RT-PCR after treatment with vehicle or dexamethasone. Data are shown as mean \pm SEM. Asterisks indicate statistically significant differences by paired t test (*p < 0.05, **p < 0.01, ***p < 0.001, ns - not significant).
- G. Quantification of AR, PSA, and TMPRSS2 expression in VCaP after treatment with dexamethasone. Data are shown as mean \pm SEM. Asterisks indicate statistically significant differences by paired t test for each dosage compared to 0 nM (*p < 0.05, **p < 0.01, ***p < 0.001, ns - not significant).

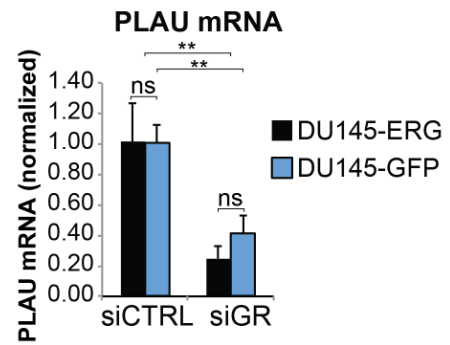
A ERG-overexpressing Mouse Cell Line



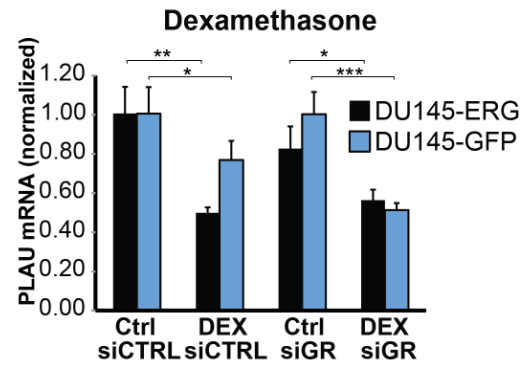
D



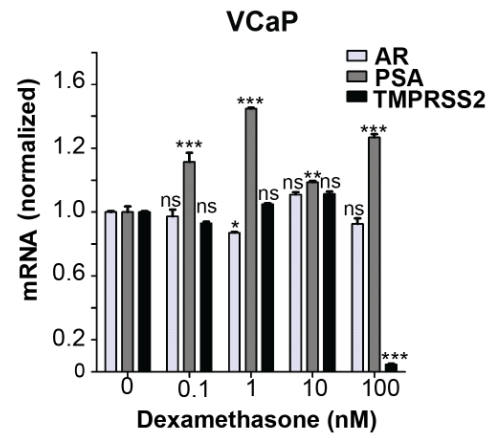
E



F



G



dexamethasone's impact on *PLAU* expression in ERG positive cells (**Figure 2.5F**). Additionally, we found that AR target genes were not substantially mobilized by dexamethasone and screening of VCaP cells showed that dexamethasone had little effect on AR signaling (**Figure 2.5G**). Altogether, these results indicate that dexamethasone-mediated ERG inhibition occurs independently of GR and AR signaling.

We next looked to see what CRAFTT would predict for another glucocorticosteroid that is used in the treatment of prostate cancer, prednisone. We found that CRAFTT predicted that prednisone would not inhibit ERG activity and subsequent experiments involving the active form of prednisone, prednisolone, supported this finding. Recent clinical trials for castration refractory prostate cancer (CRPC), in the absence of ERG fusion status, have suggested that there is an advantage to using dexamethasone over prednisolone, the active form of prednisone, due to improved patient PSA response rates (37% on dexamethasone compared to 17% on prednisolone)(Venkitaraman et al., 2013).

Electronic Health Record analyses support CRAFTT's predictions

To further investigate the correlation between dexamethasone treatment and prostate cancer, we performed a retrospective analysis of electronic health records (EHRs) at Columbia University Medical Center (CUMC). Kaplan-Meier survival analysis was performed using the time from first-prescription of drug to prostate cancer diagnosis (censor point) on an age-adjusted cohort of male patients. Significance was assessed using the Cox proportional hazards test. Dexamethasone patients had a statistically significant greater likelihood of not getting diagnosed with prostate cancer than patients on prednisone ($p < 0.001$), patients on simvastatin ($p < 0.001$), and patients on any of the top 100 prescribed

drugs ($p < 0.001$) (**Figure 2.6A**). We next constructed a logistic regression model to assess the relationship of the dexamethasone and other control treatments and prostate cancer diagnosis independent of known prostate cancer confounders. The results of our regression model showed a protective effect for dexamethasone administration versus other control treatment groups that was independent of other known risk factors. Thus dexamethasone appears to both be protective against prostate cancer (perhaps through its inhibitory effect on ERG-rearranged tumors as predicted in this study) and more active than prednisolone both in its protective effect and in the treatment of CRPC. We note that these results are still largely correlative in the absence of ERG molecular status for EMR patients, which we could not obtain for this study.

CRAFTT predicts candidate drugs for reactivating TF activity

CRAFTT also made predictions about drugs for transcriptional reactivation. We found that there was an enrichment of histone deacetylase inhibitors ($p < 0.0001$, Permutation test) amongst our reactivation predictions, indicating that CRAFTT is successful in identifying true drug-TF interactions. Thus we hypothesized that we could identify a drug that reactivates the tumor suppressor TF p53. The application of CRAFTT to p53 ChIP-seq (Kittler et al., 2013) and subsequent network analysis identified promethazine ($FWER < 0.001$) as a therapeutic option for reactivation of p53 activity. Analysis of DTP-NCI60 drug sensitivity data (Reinhold et al., 2012) further supported this prediction, as we found that the mutant p53 cell lines were significantly more sensitive to promethazine than the wild-type p53 cell lines ($p = 0.0376$, Mann-Whitney test, **Figure 2.6B**). We next looked to see whether any predicted drugs for p53 activity reactivation targeted genes had been previously identified as necessary for growth in *TP53* deficient cells (Xie et al., 2012). We found that seven of the drugs predicted by CRAFTT to

reactivate p53 activity target genes from that list: pentetrazol, naftopidil, oxedrine, capsaicin, ifenprodil, flumetasone, and dexpropranolol. Altogether this suggests that that our approach can be used to identify candidates for reactivation of TFs frequently lost in cancer.

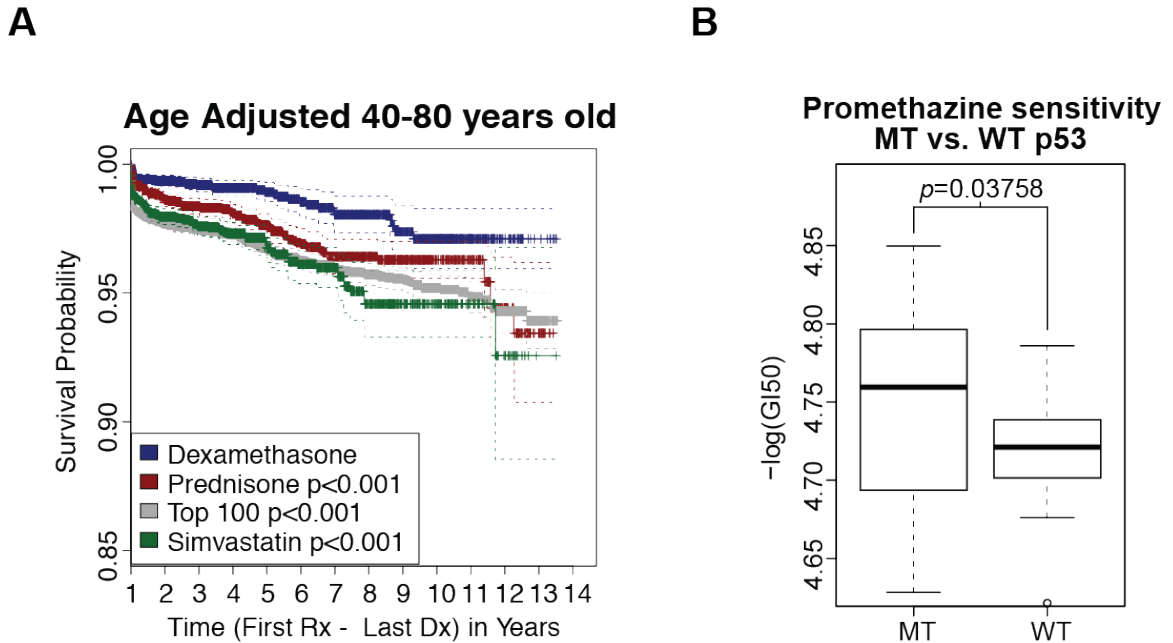


Figure 2.6 – Extended Analyses of CRAFTT Predictions.

- Kaplan-Meier survival analysis for time from first-prescription of drug to prostate cancer diagnosis (censor point) using an age-adjusted cohort of male patients was performed for patients treated with dexamethasone, prednisone, simvastatin and top-100 prescribed drugs. Statistical significance was assessed using cox proportional hazards test for the comparison of dexamethasone to each other drug.
- The drug concentration required to inhibit 50% growth (GI_{50}) in mutant p53 and wild-type p53 cell lines in the NCI DTP. Statistical significance was assessed using the Mann-Whitney test.

DISCUSSION

Traditionally, TFs have been considered difficult to drug and attempts at identifying drugs that affect TFs unfruitful. While recent breakthroughs have begun to experimentally identify molecules that indirectly modulate transcriptional activity, we propose a method (called CRAFTT) to do so computationally and systematically. Since cancer subtypes are frequently associated with aberrant TF activity often due to somatic mutations, our approach has the potential to broadly impact the development of new therapeutic strategies in these subtypes.

We first looked to see if CRAFTT could rediscover known cases of drugs that affect TF activity. We found that when we applied our method to transcriptional binding site data and drug profiles from known cases, we could indeed rediscover these connections. We then used CRAFTT to identify dexamethasone as a candidate for inhibition of ERG activity and follow-up experiments supported this prediction. We also found that dexamethasone had a similar effect in recently isolated mouse cell lines as it did in the human cell lines. This suggests that mouse models could be used to further follow-up on the therapeutic use of dexamethasone in treatment of the ERG-overexpression cancer subtypes.

While CRAFTT was successful in the identification of drugs for affecting transcriptional activity, there are areas that could further improve its predictive capacities. While the shortest path analysis provides support for our predictions and is only used in prediction prioritization, we cannot rule out that individual predictions may be affected by bad edges, especially in our protein-protein interaction network. However a network sensitivity analysis does suggest that our network is robust to missing network edges. This is likely due to the high interconnectivity of the network, which has an average path length of 3.6. This

high interconnectivity also explains the bimodality in the normalized path lengths, with the first and second peaks corresponding to shorter and longer observed path lengths than average for a drug-TF pair with the same network degree respectively.

Additionally, the ChIP-seq that we used to derive binding site data was obtained from wild-type TFs. However we note that our approach was able to capture true drug-TF interactions, at least in part due to these variants often causing constitutive expression and binding of the TF instead of dramatic disruption and changes to binding sites. However as more mutant TF binding data becomes available, we will be able to adapt and apply our approach in a more targeted and physiologically relevant manner. ChIP-seq peak calling procedures are also known to be error-prone. While we have taken steps to control for binding hotspots, our method will also benefit as improved peak calling methods become available.

Finally, the Connectivity Map data that we analyzed was in a collapsed format, which limits the robustness of the predictions. The Broad Institute has recently released an updated version of the Connectivity Map, which includes a 1000-fold scale up and will better allow us to utilize the variability in replicates. We also intend to apply CRAFTT for the identification of candidate drugs for modulating the activity of other TFs that are historically elusive but desirable for targeting, such as FOXA1 and XBP1.

CHAPTER THREE

A DATA DRIVEN APPROACH TO PREDICTING DRUG TOXICITY*

PREAMBLE

This chapter consists of a paper that was published in *Cell Chemical Biology* in October 2016. The method (PrOCTOR) was conceived in partnership with Dr. Olivier Elemento. I implemented the method and subsequent analyses. Neel Madhukar contributed to model interpretation and follow-up analyses.

INTRODUCTION

Failures in all phases of clinical trials have skyrocketed over the past three decades, with a substantial portion occurring for safety reasons (Hay et al., 2014; Ledford, 2011). This is occurring despite improvements in all stages of the drug development pipeline (Scannell et al., 2012). One of the key areas of improvement has been the screening for drugs likely to fail clinical trials.

Drug-likeness measures have been widely accepted as a useful guide for filtering out toxic molecules in the early stages of drug discovery. Lipinski first proposed this concept over a decade ago with his Rule of 5 (Ro5), a set of four physicochemical features associated with orally active drugs that were derived from analyzing clinical drugs that reached Phase II trials or beyond (Lipinski et al., 1997). This concept enhanced the drug discovery process by providing a set of practical filters that became widely adopted in drug development pipelines. However Lipinski noted that the Ro5 is a very conservative predictor and passing the rule does not guarantee drug-likeness (Lipinski, 2004). Modified rule sets

* Gayvert KM, Madhukar NS, Elemento O. A Data-Driven Approach to Predicting Successes and Failures of Clinical Trials. *Cell Chem Biol.* 2016;23(10):1294-1301.

have since been proposed, such as Veber's Rule (Veber et al., 2002) and Ghose's Rule (Ghose et al., 1999), to include more properties associated with bioavailability, such as Polar Surface Area, and to improve upon the concept proposed by Lipinski. More recently, the Quantitative Estimate for Drug-likeness (QED) was proposed as an alternative to rule-based methods (Bickerton et al., 2012).

The adoption of drug-likeness concepts early in the drug discovery process has been shown to reduce attrition rates (Leeson et al., 2007). However despite these advances in identifying potentially toxic drugs, clinical trial attrition rates have continued to rise (Hay et al., 2014). While oral bioavailability is highly relevant to drug toxicity, there are other factors that also contribute to clinical trial toxicity events. To address this problem, we propose a new approach for predicting odds of clinical trial outcomes (PrOCTOR).

RESULTS

Analysis of clinical trials data reveals limitations of structural-based approaches

Drug-likeness approaches have been important and informative in guiding the drug development process. However they cannot distinguish drugs with unmanageable toxicity profiles from safe ones (Bickerton et al., 2012; Leeson et al., 2007). We verified this quantitatively by comparing drugs that have failed clinical trials with FDA approved drugs. To this end, we downloaded data from The Database for Aggregate Analysis of ClinicalTrials.gov (AACT) ^{41 39 38 38 38 34 34} _{34 34 34 34 34} at [ClinicalTrials.gov](https://clinicaltrials.gov) and extracted the names of the drugs associated with 108 clinical trials of any phase that were annotated as having failed for

toxicity reasons. The comparative list was developed from the 1013 FDA approved drugs that were annotated as FDA approved in the DrugBank database(Law et al., 2014).

For the drugs in these lists, we tested existing methods for their ability to distinguish approved drugs from those that failed for toxicity in trials (FTT drugs). Most FDA approved drugs pass Lipinski's Rule of Five(Lipinski et al., 1997) (80.6%) and Ghose's(Ghose et al., 1999) (64.9%) rules, but so do most of the FTT drugs (73% Lipinski, 54% Ghose). In contrast, Veber's rule(Veber et al., 2002) appears to be a far too conservative measure, with 75.2% of approved and 92% of FTT drugs being predicted to fail. Finally the QED approach, which calculates a continuous score(Bickerton et al., 2012), is also unable to significantly distinguish the two classes ($p=0.1069$, $D=0.10703$, Kolmogorov-Smirnov Test). This analysis further highlights the unmet need to develop strategies for predicting the likelihood of toxicity in clinical trials.

Computational approach accurately predicts likelihood of clinical trial failure

Because all of the drug-likeness methods consider only the chemical properties of a molecule, we reasoned that a new approach that includes overlooked features related to the results of a drugs performance could prove to be highly impactful, similar to the effect that adopting sabermetrics had on the baseball scouting process as described in Michael Lewis's *Moneyball*(Lewis, 2003). A specific example is the consideration of target-related properties, such as tissue selectivity (an ideal target would be found only in diseased tissue and sparsely anywhere else). We suggest that such considerations could be useful in determining potential toxic effects.

The inferences gained from the analysis of the various methods and the consideration of additional characteristics in the prediction of tolerable toxicity in clinical trials led to the development of our new approach for predicting odds of clinical trial outcomes using random-forest (PrOCTOR). PrOCTOR integrates established informative chemical features of the drugs with target-based features to produce a classifier that is able to distinguish FDA approved drugs from FTT drugs. Random forest (Breiman, 2001b), a decision tree based machine learning model, is used to address the classification problem of clinical trial drug toxicity (**Figure 3.1**). The random forest model builds a set of 50 decision trees with a subset of features (see below) within each tree and assigns the predicted outcome to be the consensus of the trees.

The set of 48 features describing each drug contains 10 molecular properties, 34 target-based properties and 4 drug-likeness rule features. Given their established validity, we chose to include the molecular properties considered by the Lipinski, Veber and Ghose rules. We found that, individually, some of these properties had slight but significant power to discriminate between FDA approved drugs and FTT drugs when applied to our lists of drugs in the two categories (**Figure 3.2A**). Additional features represent the compatibility of the compounds with the drug-likeness approaches. Each drug's known targets were annotated from the DrugBank dataset (Law et al., 2014) and used to derive an additional set of target-based properties. We considered the median expression of the gene targets in 30 different tissues, such as the liver and the brain, calculated from the Genotype-Tissue Expression (GTEx) project (Consortium, 2015). Other target-

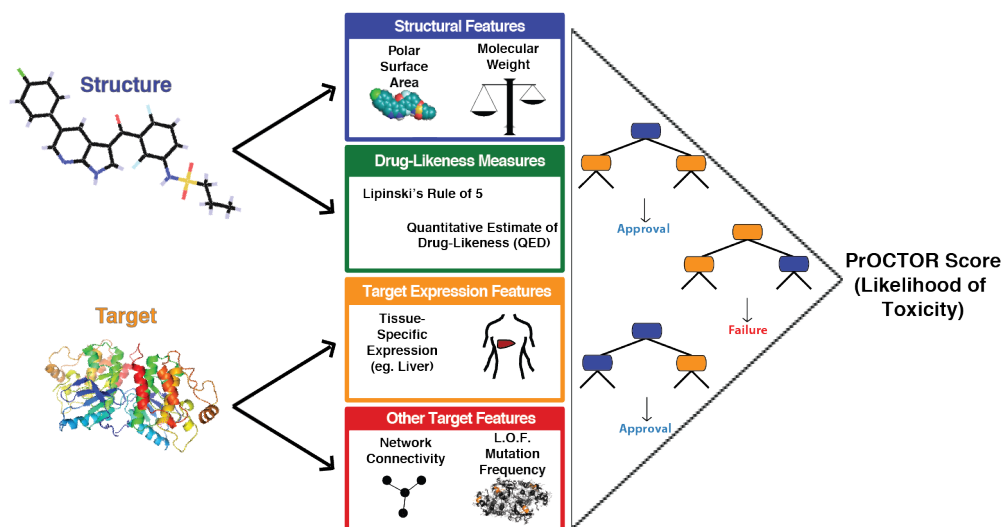


Figure 3.1. Method Schematic. Our approach integrates chemical properties, drug-likeness measures and target-based properties of a molecule into a random forest model to predict whether the drug is likely to be a member to fail clinical trials for toxicity reasons.

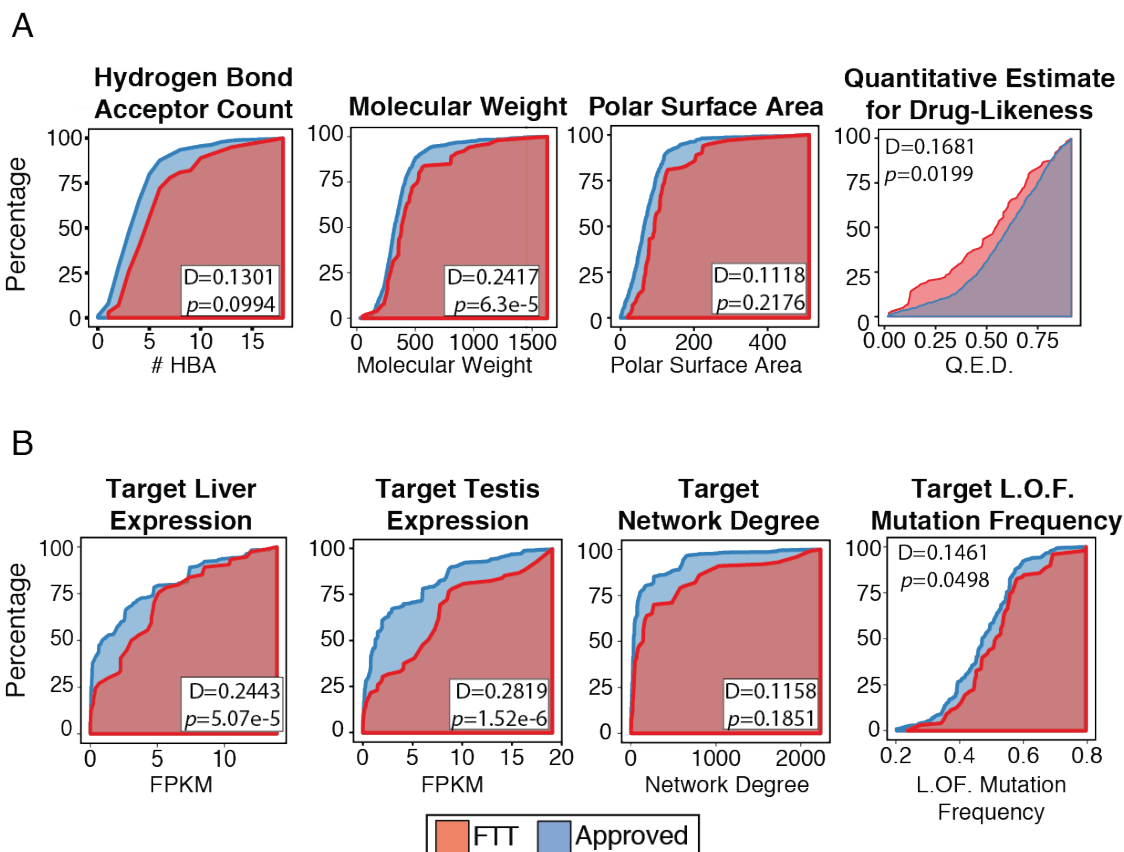


Figure 3.2. Distributions of select (a) chemical features, and (b) target-based model features. The Kolmogorov-Smirnov D statistic and p-value are shown for the comparison of failed toxic clinical trial (FTT) drugs (red) and FDA approved drugs (blue).

based features represent the network connectivity of the target, with gene degree and betweenness features, computed using an aggregated gene-gene interaction network (Aksoy et al., 2013; Das et al., 2012; Khurana et al., 2013), and a feature that represents the loss of function mutation frequency in the target gene, extracted from the Exome Aggregation Consortium (ExAC) database (Exome Aggregation Consortium (ExAC)). Like the chemical properties, we found that some of these target-based features also were able to weakly but significantly discriminate between FDA approved drugs and FTT drugs (**Figure 3.2B**). Not surprisingly, many of the features within the target-based or the chemical category were highly correlated with each other. Since we found the target expression values to be highly correlated, principle component analysis was applied to all target expression values in order to reduce the feature dimensionality. In place of the raw expression values, the first three principle components were instead used. However there was little correlation between the two classes of features (maximum Pearson correlation of $r=0.1942$). Thus the target-based features add information independent of the chemical features into the model.

The approach was tested by performing 10-fold cross validation on a set of 784 FDA approved drugs with known targets and the drugs associated with 100 FTT that had at least one annotated target and known chemical structure. We found that ProCTOR had significant predictive performance, with an area under the receiver operator curve (AUC) of 0.8263 (**Figure 3.3A**). At the optimal point of the curve the method achieved an accuracy (ACC) of 0.7529, with both high sensitivity (true positive rate (TPR) of 0.7544), and high specificity (true negative rate (TNR) of 0.7410). By comparison, on this same dataset the Ro5 and Ghose rules had a TPR of 0.8030 and 0.6468, respectively, and a TNR of 0.27 and 0.46

respectively. Application of the Veber method achieved a TPR of 0.2465, and a TNR of 0.92. (**Figure 3.3A**). The ROC curve of both the unweighted and weighted versions of the QED method fell significantly below that of ProOCTOR's ROC curve (AUC=0.581, $p < 2.2e-16$, Wilcoxon signed rank test), indicating that ProOCTOR is able to better distinguish the FTT and approved drug classes. Furthermore, ProOCTOR's approval probability allows for the separation of the drugs of the FTT and FDA approved classes ($D=0.5343$, $p < 2.2e-16$, Kolmogorov-Smirnov test) (**Figure 3.3B**) on a continuous scale.

We further assessed the approach by applying ProOCTOR to drugs that are approved in Europe (EMA-Approved) or in Japan (JP17) but not annotated as being FDA approved in our dataset. When compared to the FTT drugs in our training set, we found that EMA-Approved ($p < 2.2e-16$, Mann-Whitney U Test) and JP17 drugs ($p = 9.84e-14$, Mann-Whitney U Test) were predicted to be significantly safer and had a similar distribution of ProOCTOR scores to the class of FDA Approved Drugs (**Figure 3.3C**).

Next, we applied ProOCTOR to 3,236 drugs that were in DrugBank and not in our training set. We found that the predicted toxic drugs had significantly more frequent reports of serious adverse events, such as death and renal failure, than predicted safe drugs in the openFDA resource of drug adverse events (<https://open.fda.gov>) (**Figure 3.3D**). Furthermore, we found that safe predictions were enriched for classes of drugs that are known to be relatively safe, such as antidepressants, stimulants, and serotonin-related drugs. In comparison, toxic predictions were enriched for known toxic classes of drugs, such as immunosuppressive agents and antineoplastic agents.

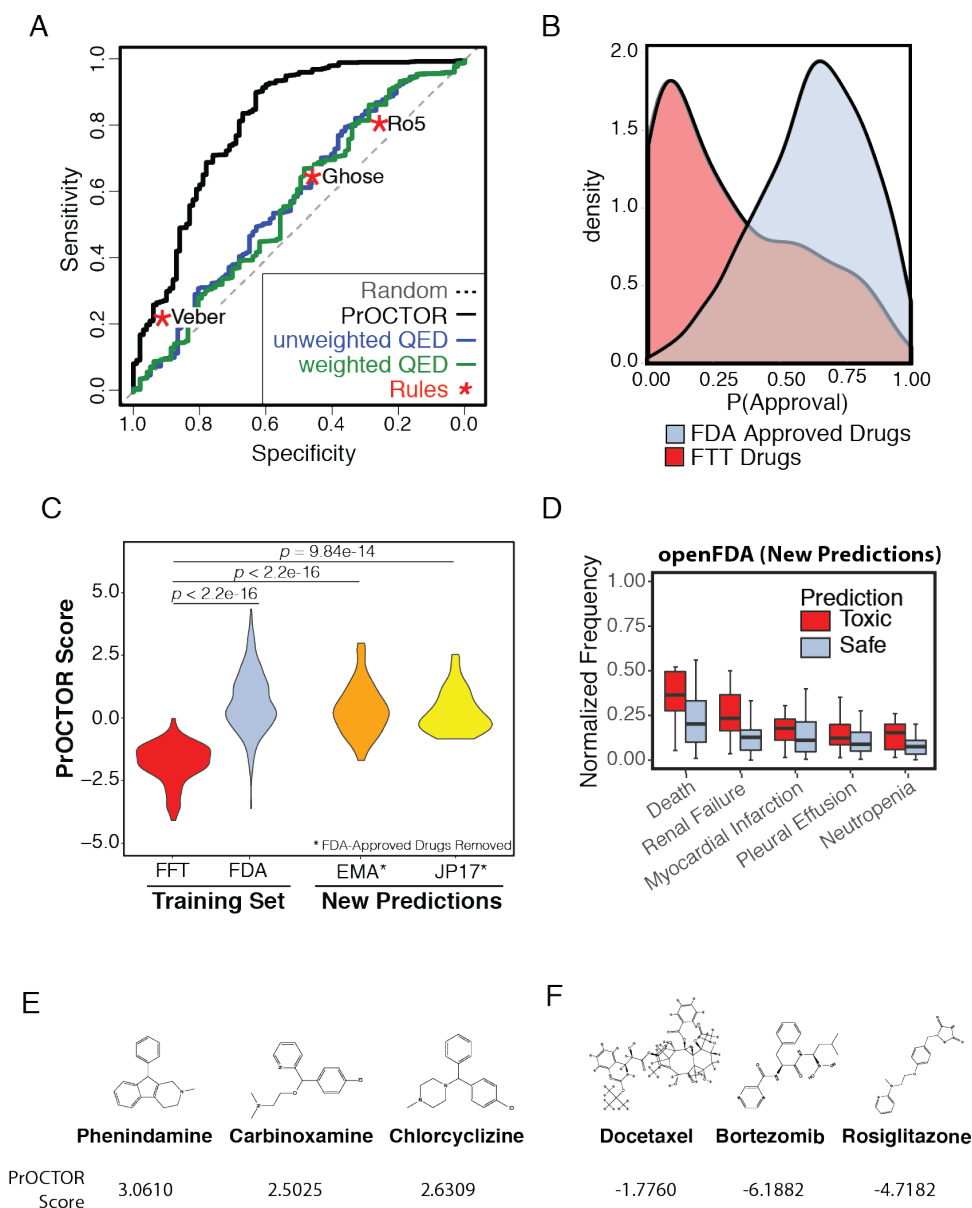


Figure 3. Benchmarking Model performance. (a) Receiver operating characteristic (ROC) curves for ProCTOR, three drug-likeness rules (Ro5, Veber, Ghose) and both the weighted and unweighted QED metrics. (b) ProCTOR scores and the Q.E.D. metric for approved and failed toxic clinical trial (FTT) drugs. (c) ProCTOR scores for the FDA approved and FTT drugs in the training set, as well as EMA-Approved and Japanese-Approved (JP17) drugs after removal of FDA approved drugs. Statistical significance was assessed for FDA, EMA, and JP17 vs FTT drugs using the Mann-Whitney U Test. (d) Reported frequencies, normalized to the most frequently reported adverse event, in the openFDA database for predicted toxic (red, score < -1) and predicted safe drugs from the DrugBank dataset. (e) The top three molecules predicted by ProCTOR as most likely to be FDA approved are phenindamine, carbinoxamine, and chlorcyclizine. (f) The three molecules predicted by ProCTOR as most likely to fail clinical trials for toxicity reasons are docetaxel, bortezomib, and rosiglitazone.

We also applied our approach to 137 drugs annotated as most-DILI-concern and 65 drugs of no-DILI-concern by the FDA. We found that the most-DILI-concern drugs had 1.5-fold higher odds of being classified as toxic by ProCTOR than the no-DILI-concern drugs. More generally, the most-DILI-concern drugs had higher ProCTOR scores than the no-DILI-concern drugs ($p= 0.0005$, Mann–Whitney U Test). This suggests that our model is able to generalize beyond the training set.

Identification of FDA drugs with increased likelihood of toxicity events

Next we looked to evaluate the predictions of our approach by analyzing ProCTOR's predictions for FDA approved drugs. A ProCTOR score expressing the $\log_2(\text{odds of approval})$ was calculated taking the \log_2 of the ratio of the ProCTOR-predicted probability of approval to the probability of failure.

The three molecules identified by ProCTOR as most likely to receive FDA approval were phenindamine, carbinoxamine, and chlorcyclizine (**Figure 3.3E**). All three of these drugs are FDA approved antihistamines with highly tolerable side effects. Interestingly, all three of these drugs pass the Ro5 but have relatively low QED values (0.311, 0.242, and 0.499 respectively).

The three molecules with the worst ProCTOR score and thus predicted as most likely to fail clinical trials for toxicity reasons were docetaxel, bortezomib, and rosiglitazone (**Figure 3.3F**). Of note, all are FDA approved drugs that have been associated with serious toxicity events. Docetaxel is a chemotherapy agent used to treat a number of cancers (Massacesi et al., 2004; Puisset et al., 2007). The most frequent adverse event associated with docetaxel is neutropenia, a potentially life threatening event that often results in delay of treatment (Puisset et al., 2007). It also fails the Ro5 and has an extremely low QED of 0.147, suggesting that this prediction is consistent with other drug screening methods.

Bortezomib is a proteasome inhibitor used for treatment of relapse multiple myeloma that has a moderate QED value of 0.476 and passes the Ro5. While it was FDA approved due to its significant antitumor activity, it has been associated with frequent adverse events, such as peripheral neuropathy, that are thought to in part be due to nonproteasomal targets (Arastu-Kapur et al., 2011).

Rosiglitazone is an antidiabetic drug that also passes the Ro5 and has a high QED value of 0.825. However it has been linked with an elevated risk of heart attack (Nissen et al., 2007) and consequently was withdrawn from the market in Europe in 2010 (Blind et al., 2011). This suggests that existing methods were not necessarily able to foresee the adverse events associated with these latter two compounds.

These compounds bring to attention the importance of context when considering toxicity events. In general, more frequent and serious side effects will be acceptable for drugs that are used to treat severe and otherwise untreatable conditions, such as cancer. This is an important consideration to keep in mind when determining acceptable score ranges in drug development. Additionally, it highlights the shortcomings of rule-based methods, which are unable to quantify the extent to which a drug may have undesirable characteristics since a molecule that just barely fails one requirement is equivalent to one that substantially fails all requirements.

We further assessed what insights the predictions from ProCTOR can offer regarding toxic effects using the SIDER side effect resource database (Kuhn et al., 2010). We hypothesized that drugs with better ProCTOR scores would have less frequent severe side effects reported due to their more tolerable toxicity profiles. We first compared all drugs predicted to be approved by ProCTOR (via cross-validation), including those misclassified, to those predicted to be of the

FTT class. We found that the predicted FTT drugs had significantly more frequent severe side effects, such as neutropenia (37.3% vs 14.3%, $p=1.78 \times 10^{-7}$, Fisher-Exact test) (**Figure 3.4A**). When comparing the drugs with the top 10% best PROCTOR scores to those within the bottom 10%, this distinction was even greater with severe toxic events, such as neutropenia (54.8% vs 13.4%, $p=1.72 \times 10^{-6}$, Fisher-Exact test) and pleural effusion (47.6% vs 5.2%, $p=2.59 \times 10^{-7}$, Fisher-Exact test), occurring far more frequently in the predicted FTT class.

Furthermore, we found that these severe side effects were significantly negatively correlated with the PROCTOR score. For example, the spearman's correlation coefficient of the binned pleural effusion frequency against the PROCTOR score was $\rho=-0.9792$ (**Figure 3.4B**) and for neutropenia was $\rho=-0.9613$ (**Figure 3.4C**). In comparison, the frequent side effect of dizziness still occurred more frequently in the predicted toxic drugs but had a much weaker correlation of $\rho=-0.5070$. Thus the predictions of PROCTOR are consistent with reported adverse events, with the PROCTOR score negatively correlating with the reported severe side effects that would ultimately contribute to a drug's success in clinical trials.

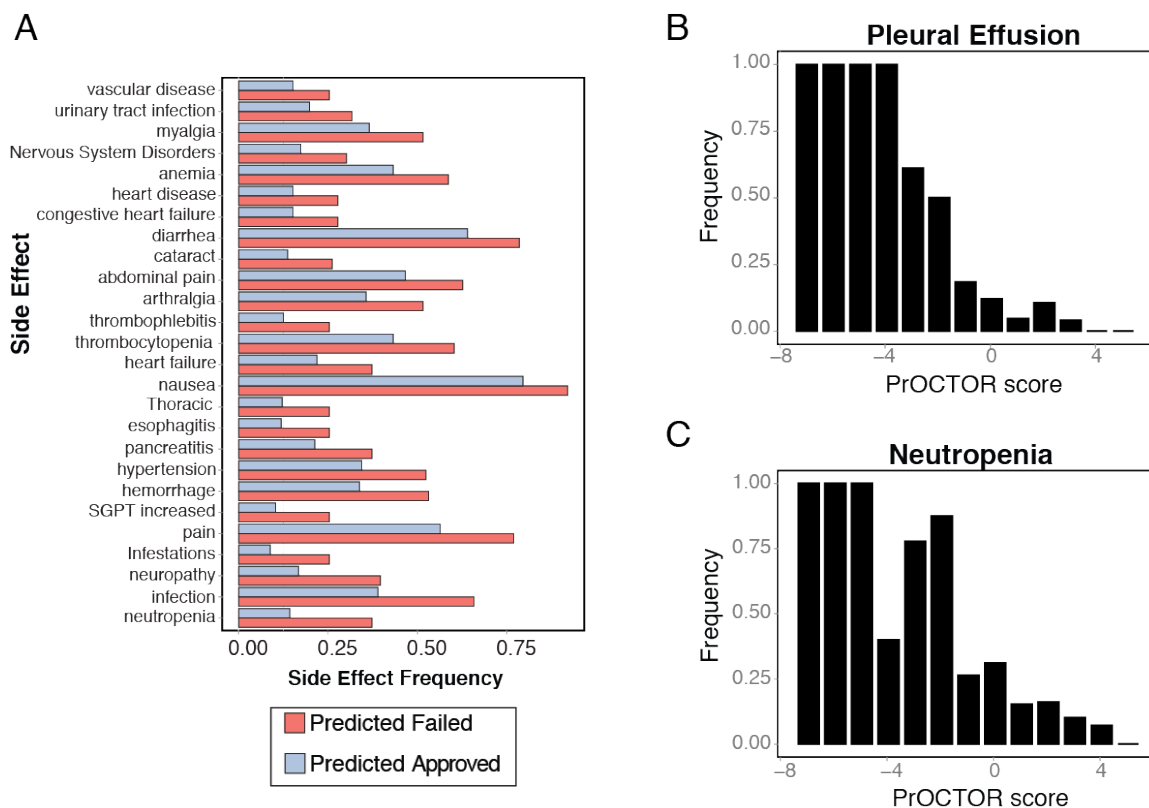


Figure 3.4. Side Effects. (a) Adverse events that occur more frequently in predicted failed toxic clinical trial (FTT) drugs compared to predicted approved drugs. (b) Binned frequency of pleural effusion across ProCTOR score bins. (c) Binned frequency of neutropenia across ProCTOR score bins.

Model reveals insights about how various properties can contribute to or help avert toxicity

We evaluated what insights ProCTOR can offer about successful drugs. A feature importance analysis showed that both the chemical and target-based features contribute significantly to the performance of the ProCTOR algorithm. The first expression principle component, QED metric, polar surface area, and the drug target’s network connectivity emerged as the four most important features (**Figure 3.5A-B**), thus target-based features were identified as highly

important features for predicting toxicity. Using target-based features alone, PrOCTOR achieved a significant predictive performance (ACC=0.7115). Our approach relies on existent annotation of drug targets to calculate these features. However this information is often not available during the drug development stage. We found that our method is robust to removal of targets (**Figure 3.5C**) and additionally maintains a significant predictive performance (ACC= 0.6708) in absence of known target information. However PrOCTOR's performance remains strongest when including both the chemical and target-based features (ACC=0.7529).

We next investigated the relationships between the features in the model. We found that certain combinations of uncorrelated features provided greater discriminative power. For example, Bickerton *et al.* (Bickerton et al., 2012) reported that the QED approach outperformed other drug-likeness methods when the threshold was set at 0.35. We found that 75% of drugs with QED<0.35 were approved. However when high testis expression (FPKM>10) was added into consideration, 88.5% of FTT drugs were accurately be classified. Additionally, tissue selectivity is a useful consideration in determining potential toxic effects. We hypothesize that this may be due to some tissue-specific toxicity events being associated with the drug target's expression in normal tissue. We found that 84% (38/45) of drugs with high molecular weight (MW>500) but low general tissue expression (PC1< -2) were FDA approved. Thus if a gene appears to be a promising target for mechanistic reasons while appearing ill-suited due to high global expression profiles, it still may remain a viable candidate given that certain molecular properties are satisfied.

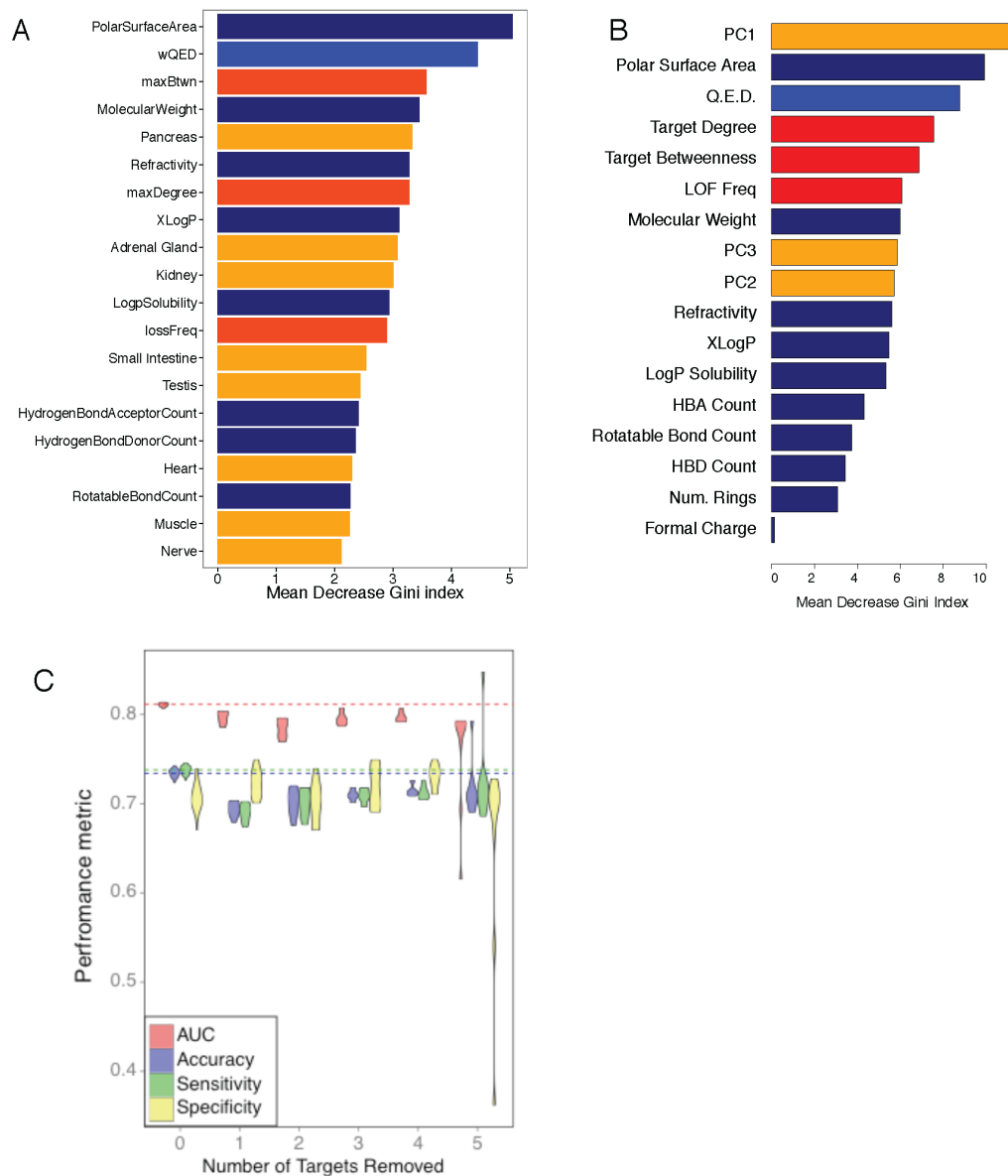


Figure 3.5. Feature Importance and Model Robustness. Mean decrease Gini coefficient observed upon feature removal for the top 20 features **(a)** with all individual expression features and **(b)** with top 3 expression principle components instead of individual expression features. **(c)** Violin plots showing the range of AUC, accuracy, sensitivity and specificity for 0-5 targets removed.

DISCUSSION

Drug-likeness approaches, as first proposed by Lipinski almost two decades ago, have become a key tool for the pre-selection of compounds that are likely to have manageable toxicity in clinical studies. However all these methods consider only the molecular properties of the drug itself. We have proposed a data-driven approach (PrOCTOR) for predicting likelihood of toxic events in clinical trials that moves beyond existing drug likeness rules and measures by not only considering the chemical properties of a molecule, but also the properties of the drug's target. When trained on failed clinical trials and FDA approved drugs, the PrOCTOR score performs at high accuracy, specificity and sensitivity. Furthermore, the PrOCTOR score strongly correlates with reported severe adverse events.

While phase I trials are designed to investigate safety, drugs can fail at any stage for toxicity reasons and additionally can fail phase I trials for non-safety reasons. Lipinski's Ro5 was derived using the set drugs that had succeeded to phase II trials, under the assumption that undesirable drugs would have been eliminated in Phase I (Lipinski et al., 1997). However it has been observed that a substantial number of drugs fail in Phase II trials and beyond for safety reasons (Ledford, 2011). Additionally many of the drug-likeness measures were developed using larger representative datasets in place of clinical trial data (Bickerton et al., 2012). While these methods are important, they are focused on subtly different problems such as bioavailability. We have shown above that these approaches are not able to sufficiently capture clinical trial safety. There have been a number of other methods that have been developed to predict toxicity events as well. A recent DREAM Challenge focused on predicting cytotoxicity in lymphoblastoid cell lines, however primarily focused on environmental toxins (Eduati et al., 2015). Similarly, the EPA's extensive ToxCast dataset is covered predominantly by non-

therapeutic chemicals(USEPA, 2016). Other toxicity prediction methods, such as those in AMBIT, have been developed to address other toxicity-based questions, including model organism and tissue-specific toxicities(Jeliazkova et al., 2011). QSAR models are also frequently used for toxicity prediction. However they have generally been applied to the prediction of specific toxicity endpoints, such as drug LD₅₀ values, tissue-specific toxicity events or for the estimation of maximum tolerated dose levels (Patlewicz et al., 2016). Finally, PK/PD models are highly valuable tools for identifying toxicological properties of drugs preclinically, but must be independently constructed for every drug and thus would benefit from more high-throughput methods for toxicity prediction(Sahota et al., 2016). Consequently, we selected the set of drugs that failed any phase of clinical trials for toxicity reasons to develop our approach.

We have also only addressed the issue of general clinical trial toxicity. However some indications, such as cancers, have more critical needs and consequently allow for higher toxicity levels. As a result, our model may predict some promising anti-cancer drugs to have unmanageable toxicity levels. Since PrOCTOR outputs a score, instead of just a prediction, a different threshold for allowable toxicity may be considered for different indications. A preliminary testing of this idea on cancer-only drugs with cancer type added as a feature demonstrated improved predictive power on this subset of drugs (ACC=0.74, AUC=0.80). However given the small sample size of this training set (n=89), this cancer-specific model is not optimal at this time. Additionally many new therapies are currently being developed to target specific isoforms and mutations. While our model is not currently accounting for these specific targets, it can straightforwardly be adapted using publicly available or user-provided target-based information. There are also areas in which PrOCTOR could be further

improved such that leads to better predictive capacities. The use of 3D fingerprinting methods may allow for the structural features to be better represented. Co-expression networks from the GTEx data may also be useful features, as they may provide a stronger biological signal. Biological interaction networks are generally incomplete and also vary between cellular contexts and populations, which may limit the power of the network metrics. Finally, our method is largely dependent on existing target annotation for drugs, which is generally incomplete. Thus we will likely benefit from advancements in drug target identification.

Furthermore over two-thirds of clinical trials fail for other reasons, including efficacy, strategic and financial reasons (Ledford, 2011). The problem of efficacy is a highly complex issue, since each drug must demonstrate improvement over existing drugs in addition proving a context-specific efficacy. Thus while this problem remains important, it is not likely to be tractable using this style of approach.

Our approach has the potential to impact the preclinical drug development pipeline by quantifying how likely a given compound is to have manageable toxicity in clinical trials. In order to facilitate interaction with and application of our model, we have developed an interactive tool that we have made available on github (<https://github.com/kgayvert/PrOCTOR>). PrOCTOR may also help flag drugs for increased post-approval surveillance of adverse effects and toxicity. Perhaps even more importantly, the model will help design better drugs by providing insights about how various chemical and target-based properties can contribute to or help avert toxicity.

CHAPTER FOUR

A COMPUTATIONAL APPROACH FOR IDENTIFYING SYNERGISTIC DRUG COMBINATIONS*

PREAMBLE

This chapter consists of a paper that was published in *PLoS Computational Biology* in January 2017. The method was conceived in partnership with Dr. Olivier Elemento. I implemented the method and subsequent computational analyses. The experimental validation was performed by Omar Aly. The Stern and Bosenberg labs generated the data and contributed the expertise to results interpretation (D.F.S., M.W.B., J.P.).

INTRODUCTION

Targeted therapies designed to specifically target molecules involved in carcinogenesis have achieved remarkable antitumor efficacy. In melanoma, over half of patients are reported to harbor activating mutations in the BRAF oncogene (Curtin et al., 2005; Held et al., 2013). BRAF inhibitors, such as vemurafenib, have been developed to selectively kill mutant BRAF positive cells (Chapman et al., 2011). Patients initially exhibited significant responses to these drugs, with 48% responding to vemurafenib in phase 1 and 2 clinical trials, however resistance developed within months (Chapman et al., 2011).

Combination therapy has been proposed for preventing and overcoming resistance. This is thought to be a promising option because resistance to the

* Gayvert KM, Aly O, Platt J, Bosenberg MW, Stern DF, Elemento O. A Computational Approach for Identifying Synergistic Drug Combinations. *PLoS Comput Biol.* 2017;13(1):e1005308.

combinatorial therapy would require either acquisition of multiple mutations rapidly (Fitzgerald et al., 2006) or an individual mutation that is able to bypass both drugs (Komarova et al., 2013), both of which are low probability events. Additional goals of combination therapy are to lower drug dosage levels in order to reduce the frequency and severity of adverse events and to achieve enhanced effectiveness through either drug additivity or synergy (Fitzgerald et al., 2006).

Drug synergy can occur through a variety of mechanisms. These include enhancement of bioavailability, through inhibition of parallel pathways (Cokol et al., 2011), and chemosensitization, in which the first compound primes the cells to be sensitive to the second drug (Bansal et al., 2014). Synergy is generally quantified through either effect based or dose-effect based methodologies. Effect-based methods compare the independent effects of drugs, while dose-effect based methods assume nonlinear individual dose-effect curves (Foucquier et al., 2015). The most popular effect based method is the Bliss Independence model, which assumes that drugs act independently and the expected additive effect is based on the common probabilistic independence formula (Foucquier et al., 2015). However limitations of this approach include that it does not account for nonlinearity in dose response curves (Yeh et al., 2009) and its independence assumption. Since the mechanism of action for many drugs remain unknown, the validity of the independence assumption is often not met (Foucquier et al., 2015). A popular dose-effect based method is the Chou-Talalay Combination Index (CI), which is a median-effect equation based on the “mass-action law” (Chou, 2010). A major limitation of the Chou-Talalay method is its dependence on accurate and well-defined dose-effect curves, which are not always available (Foucquier et al., 2015).

Given the poor prognosis of BRAF melanoma and the rapid rate at which resistance develops and tumors progress, there is an urgent need to identify suitable combinations. The first drug combination for treatment of advanced melanoma was approved in January 2014 and involved the BRAF inhibitor, dabrafenib, and the MEK inhibitor trametinib. This combination was pursued due to the great response rates of the individual drugs with the goal of preventing drug resistance. Indeed, it has been observed that the combination delays the development of resistance and prolongs progression free and overall survival (Robert et al., 2015). However it is certain that many more combinations exist and are not yet known. Additionally, a subset of patients that were treated with the combination of dabrafenib and trametinib have developed resistance to this combination therapy (Robert et al., 2015).

Existing methods that have been developed to predict synergistic combinations have generally relied on mechanistic insights. However they have been applied only in limited specific contexts, such as in B cells (Bansal et al., 2014).

Furthermore numerous studies have shown that synergy is very dependent on context (Bansal et al., 2014; Cokol et al., 2011; Held et al., 2013). This makes it difficult to utilize any prior knowledge about synergistic drug combinations from other different cancers or genotypes. A systematic method for identifying optimal combinations would therefore be highly impactful. Here we propose a computational approach utilizing existing high-throughput drug screen data to help identify other combinations that are both synergistic and effective in the context of mutant BRAF melanomas.

RESULTS

Single Dose Response Predictive of Combinatorial Synergy and Effectiveness

We set out to determine whether combination efficacy and synergy could be predicted from single agent efficacies. Previous computational approaches to drug combinations have shown that the dose response curves of single agents exhibit predictive power for identifying synergistic combinations (Bansal et al., 2014). To further investigate this, we utilized a high-throughput drug screen that was performed by Held *et al.* (Held et al., 2013). In this study, the response of 150 single agents and a large combinatorial drug screen involving 40 drugs were experimentally tested across mutant BRAF, mutant RAS, and wild-type BRAF and RAS (WT) cell lines. For each drug pair, we derived a feature set that consisted of the mean and difference of the single agent dose response in each tested cell line. The single agent dose response was represented as the percent of concentration required to inhibit 50% of growth inhibition (GI_{50}). Altogether, this resulted in a total of 54 total features representing the similarity of a drug pair's efficacies in 27 melanoma cell lines (**Figure 4.1**).

The results of the combinatorial drug screen were used to identify genotype-selective and synergistic combinations. Genotype-selective combinations were defined to be those that yielded an average 15% or greater growth inhibition exclusively in the genotypic group and achieved at least 50% growth inhibition within the genotypic group. We further defined a general effective combination to be one that achieved at least 70% growth inhibition. Finally we computed synergy labels for each combination using the Chou-Talalay synergy combination index (CI) metric. We decided to use the Chou-Talalay approach because the Chou-

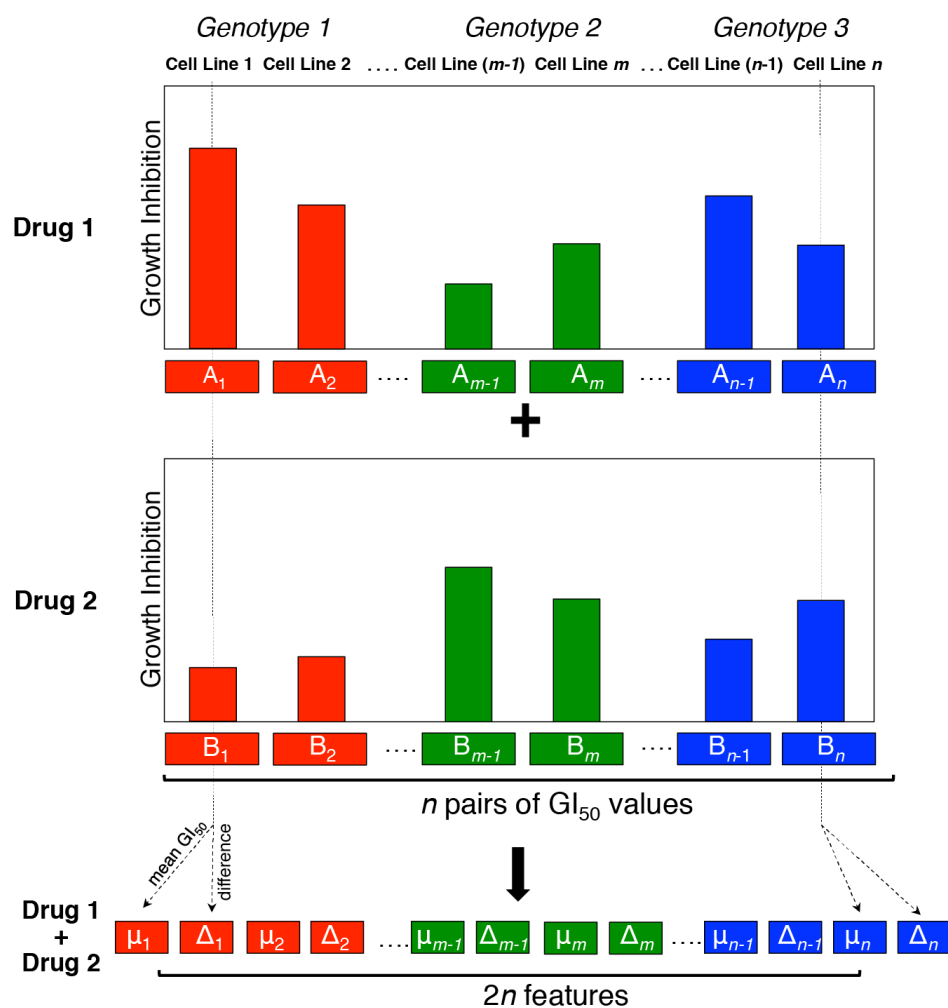


Figure 4.1 – Feature construction schematic. For each drug pair, we combined the drug pair’s efficacies by taking the mean (μ) and difference (Δ) in 27 mutant BRAF (red), mutant RAS (green) and WT (blue) melanoma cell lines.

Table 1 - Model Performance

	AUC	Accuracy	Sensitivity	Specificity
BRAF-specific effectiveness	0.8809	0.8230	0.6911	0.8894
General BRAF-effectiveness	0.8630	0.7800	0.6818	0.8418
BRAF synergy	0.8683	0.8213	0.4196	0.9494

Talalay CI was provided in the original dataset and furthermore the mechanism for many drugs in the training set is unknown. We defined a synergistic combination to be a pair of drugs that demonstrated $CI < -1$ at any concentration level. Overall, the combinatorial screen helped identify 248 BRAF-selective and 161 synergistic combinations.

We then trained random forest models (Breiman, 2001a) on 780 drug combinations for each of the outcomes described above in context of BRAF and RAS melanomas (**Figure 4.2A**). We evaluated our approach using 10-fold cross-validation and found that our model exhibits significant power (**Table 1**) for predicting both synergy (AUC=0.8663, Accuracy=0.8213) and genotype-selective efficacy (AUC=0.8809, Accuracy=0.8230) in context of BRAF melanomas (**Figure 4.2B**). Importantly both models maintained high specificity rates (0.9494 and 0.8894 for synergy and effectiveness respectively) which suggests that there would be few false leads. As a control to identify limitations of our approach, we evaluated our model by making predictions for “sham combinations”, which were cases in which a drug is combined with itself. We found that 98% (147/150) of the sham combinations were predicted to not be synergistic. We also found that both the effectiveness and synergy models exhibit significant robustness (**Figure 4.2C**). At 25% of the original number of combinations used in the training set, the BRAF-specific effectiveness approach maintained 77.56% accuracy, 89.27% specificity, and 54.91% sensitivity. This suggests that fewer combination testing could be performed while maintaining strong confidence in the positive predictions, given that the specificity remains high. Since high-throughput screens require significant resources and time, this type of approach could prove to be valuable in screening the larger space of drug combinatorial pairs given that a suitable representative set is chosen for initial testing.

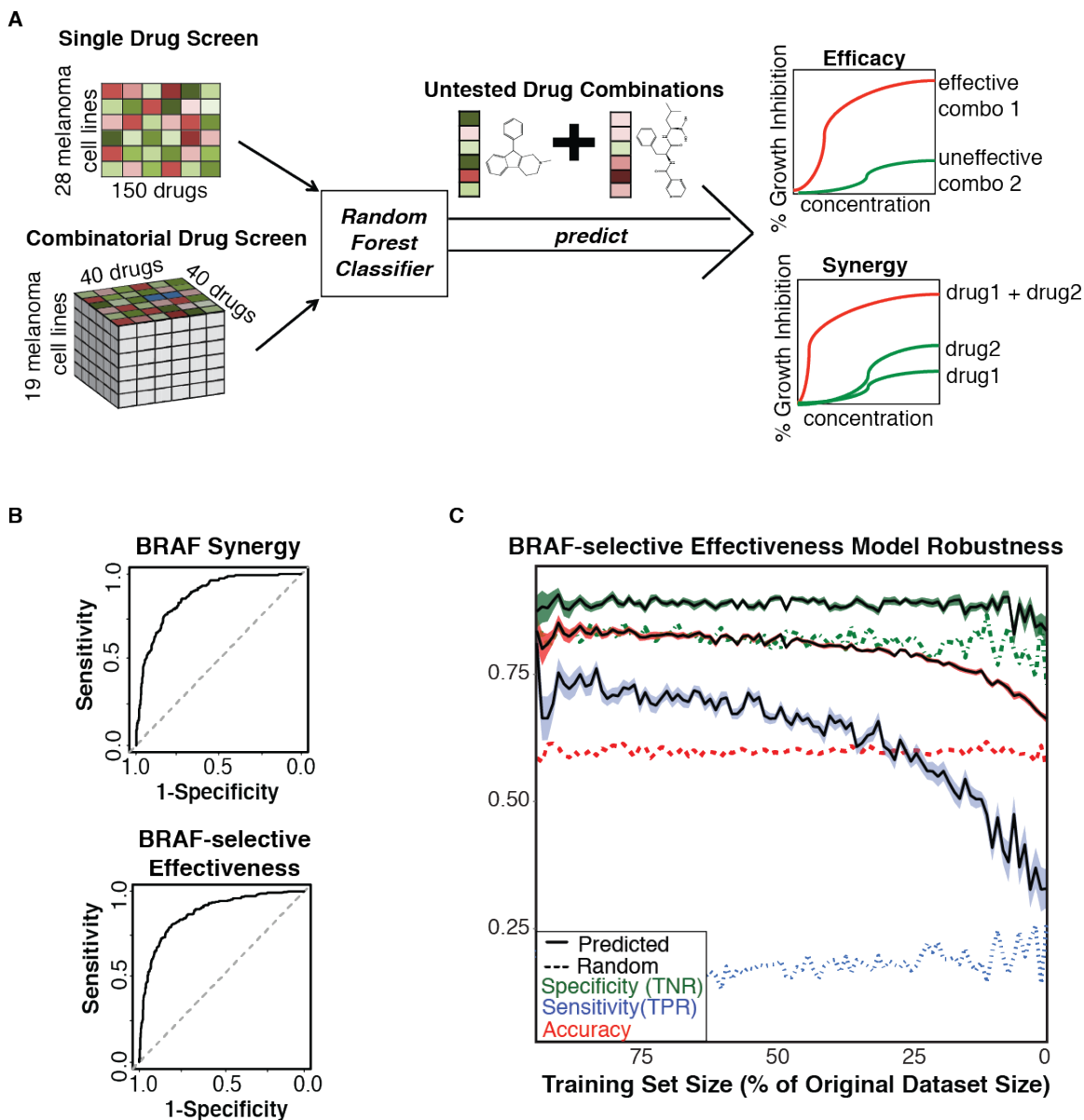


Figure 4.2 – Method schematic and Evaluation of individual model performance.

- A Our approach integrates a large single drug screen of 150 drugs with a combinatorial drug screen, which tested 780 combinations of 40 unique drugs. This was used to train a random forest model that predicts synergy and genotype-selective efficacy for untested drug combinations.
- B Receiver operating characteristic (ROC) curves for 10-fold cross-validation of the BRAF-specific effectiveness (top) and synergy (bottom) models.
- C The effect of randomly removing samples on model accuracy, sensitivity and specificity. At 25% of the original number of combinations was used to train the model, the approach maintained 77.56% accuracy, 89.27% specificity, and 54.91% sensitivity.

The single agent screens performed by Held *et al* (Held et al., 2013) included 110 drugs that were not tested in the combinatorial screen, so we applied our approach to the 10,395 additional untested combinations. We predicted 842 combinations to be synergistic, 890 to be effective, and 304 to be both effective and synergistic in context of mutant BRAF melanoma. We found that our predictions had noticeable patterns of synergy and effectiveness (**Figure 4.3A**). Predicted synergistic combinations involved drugs that had varying levels of efficacy across the different mutant BRAF cell lines, with synergistic combinations demonstrating a trend towards lower correlation of GI50 values across the mutant BRAF cell lines ($p=0.07929$, Kolmogorov-Smirnov Test). In contrast, combinations involving drugs with similar efficacy profiles across the different cell lines were generally predicted to be non-synergistic. Thus it appears that our approach drew its strength from the large number of tested cell lines.

To further evaluate our approach, we compared these predictions to an independent high-throughput screen that tested 5,778 combinations involving 108 drugs at two concentration levels, high and low (Friedman et al., 2015). Our prediction dataset contained 274 combinations that overlapped with this independent dataset. We found that our predicted effective combinations had a significantly higher growth inhibition levels than our predicted non-effective combinations ($p=0.002602$, Student's t Test) (**Figure 4.3B**).

Experimental Validation of Novel Synergistic and Effective Combinations for BRAF Melanoma

We next identified a subset of 7 drugs with a diverse set of predictions (**Figure 4.3C**). These drugs included a BRAF inhibitor (PLX4720), a statin (Simvastatin), two chemotherapies (Doxorubicin, Paclitaxel), and three drugs of other various mechanisms (Fak Inhibitor 14, Gefitinib, 17AAG). We tested each of these drugs

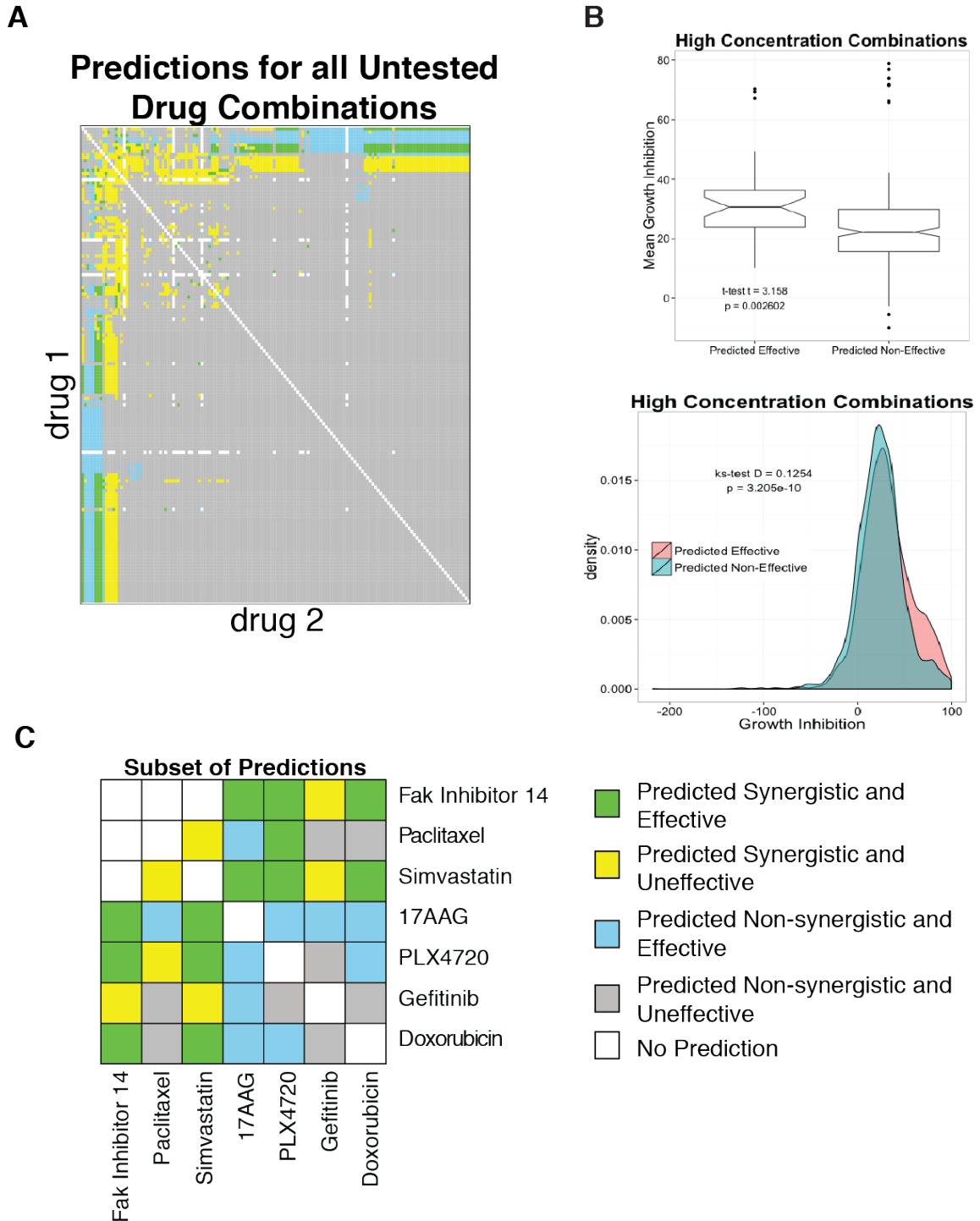


Figure 4.3 – Predictions for Previously Untested Combinations. (a) We applied our trained model to make predictions for 10,395 additional untested combinations. (b) Comparison of growth inhibition levels, as reported by Friedman *et al.*, for 274 predicted effective or ineffective drug combinations. (c) We focused on a subset of 7 drugs to experimentally follow-up on our predictions for previously untested combinations.

Table 2 - Experimental Validation

	Accuracy	Sensitivity (TPR)	Specificity (TNR)	FDR
BRAF effectiveness	0.73	0.71	0.75	0.125
BRAF synergy	0.64	0.67	0.5	0.14

both alone and in combination in the mutant BRAF melanoma cell line MALME-3M at low, medium, and high concentrations, estimated from their GI_{10} , GI_{25} , and GI_{50} values respectively.

We found that our method continued to demonstrate significant predictive power when tested on cell lines that were independent of the original training set (**Figure 4.4A, Table 2**). We validated 82% and 64% of the effectiveness (**Figure 4.4B**) and synergy predictions (**Figure 4.4C**) respectively. Importantly, we also found that the false discovery rates (FDR) for both synergy and effectiveness predictions remained relatively low (14.3% for synergy, 12.5% for effectiveness) despite being tested in a different setting.

BRAF inhibitors are of high interest for treating BRAF-mutant melanomas due to their selectivity and effectiveness. To further investigate the efficacy and synergy of combinations involving PLX4720, we performed more extensive experiments for predicted synergistic and non-synergistic drug partners. PLX4720 was predicted to be synergistic with FAK inhibitor 14 and non-synergistic with 17AAG, a Hsp90 inhibitor. Instead we found both combinations to be synergistic when tested across the 9 combinations of varying concentrations. Interestingly, we observed that while 17AAG appeared to be synergistic when combined with PLX4720 held at a constant rate, PLX4720 itself was not synergistic when 17AAG was held constant (**Figure 4.5A**). However PLX4720 was very

synergistic when combined with a constant dosage of FAK Inhibitor 14 (**Figure 4.5B**), consistent with the synergy prediction. Additionally we observed that PLX4720 was highly synergistic with paclitaxel. This combination represents a potentially impactful combination since paclitaxel and vemurafenib are both used in clinical trials for the treatment of melanoma (Chapman et al., 2011; Pflugfelder et al., 2011). Furthermore previous reports have suggested that the combination of BRAF inhibitors with paclitaxel represents a promising therapeutic approach for overcoming resistance in BRAF melanomas (Thang et al., 2015).

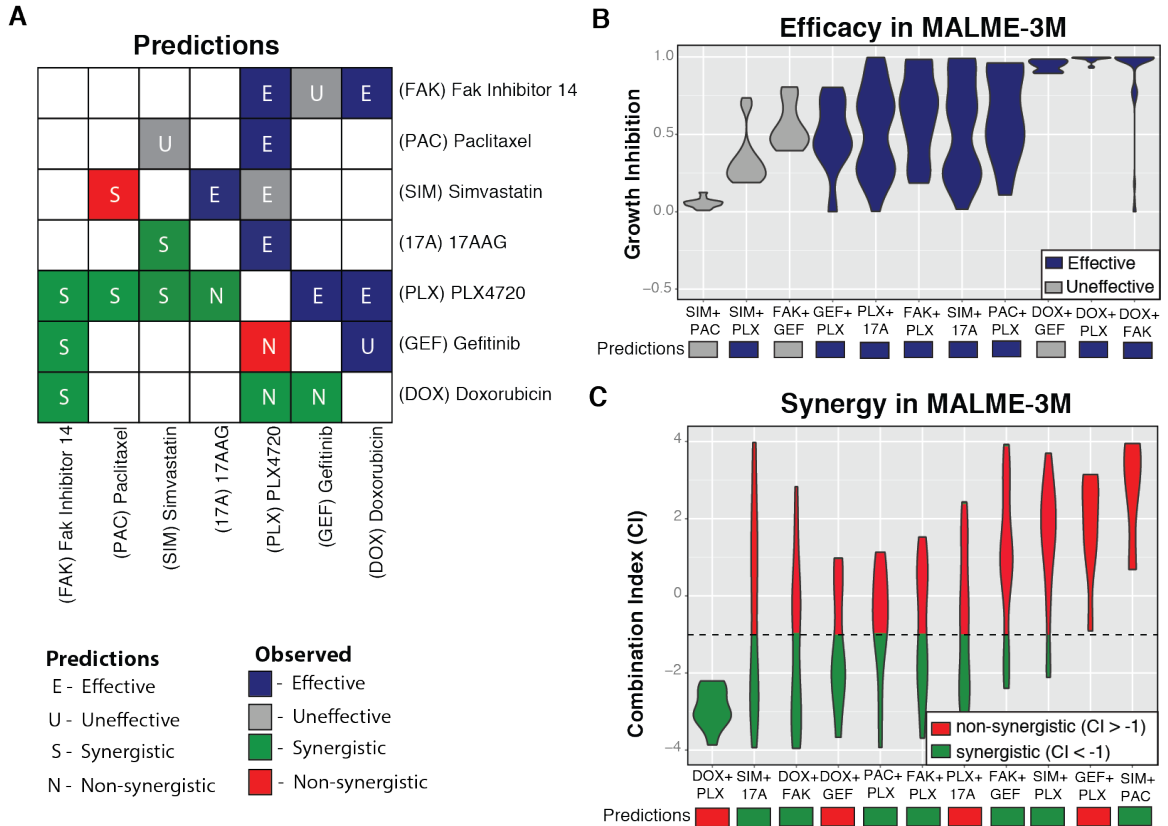


Figure 4.4 – Experimental Validation of Predicted BRAF Effective and Synergistic Combinations.

A We selected a set of 11 combinations involving 7 drugs with a diverse set of predictions for experimental validation. These included traditional chemotherapeutic agents (doxorubicin, paclitaxel), targeted agents (PLX4720, gefitinib, FAK Inhibitor 14), a statin (simvastatin), and an antitumor antibiotic (17AAG).

B Each drug was tested in combination at medium, and high concentrations, estimated from their GI_{10} , GI_{25} , and GI_{50} values respectively. The observed growth inhibition levels for all dosage level combinations involving each tested drug combination are shown in violin plots. Violin plots that are colored navy blue are those whose third quantile values where 70% or greater. The predictions for each combination are shown below the plot, with dark blue representing an effective prediction and grey representing an ineffective prediction.

C For each tested drug combination, the Chou-Talalay synergy scores were calculated. The observed synergy scores for all dosage level combinations involving each tested drug combination are shown in violin plots. The predictions for each combination are shown below the plot, with green representing a synergy prediction and red representing a non-synergy prediction.

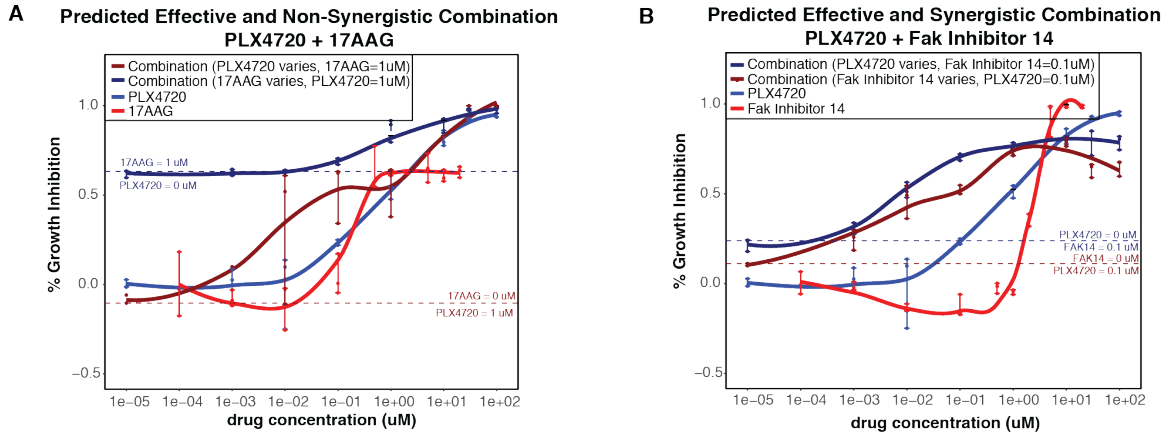


Figure 4.5 – Identification of Synergistic Combinations involving the BRAF Inhibitor PLX4720.

- A The observed growth inhibition levels for PLX4720 alone (blue) 17AAG alone (red), PLX4720 varying while 17AAG held constant at 1uM (navy blue), and 17AAG varying while PLX4720 held constant at 1 uM (dark red).
- B The observed growth inhibition levels for PLX4720 alone (blue) FAK Inhibitor 14 alone (red), PLX4720 varying while FAK Inhibitor 14 held constant at 0.1uM (navy blue), and FAK Inhibitor 14 varying while PLX4720 held constant at 0.1 uM (dark red).

DISCUSSION

We found that drug synergy and combinatorial effectiveness can be predicted from a relatively small subset of combinations based only upon single drug efficacies. We experimentally validated novel predictions involving 7 drugs in a BRAF mutant cell line with $FDR < 0.15$. This analysis included compounds that span a variety of drug classes, including targeted therapies and chemotherapies. Additionally, an analysis of the model robustness suggested that it is possible to confidently make these predictions with even smaller subsets, while maintaining significant confidence in the positive predictions. We note that while we propose that a smaller training set can be used to infer these combinations, it is critical to retain a distinctive and representative set of drug combinations in the training set.

The classification errors, particularly those involving synergy predictions, may be due in part to varying genetic conditions since the combinations were tested in a different cell line than the original training set. This is supported by one combination that was tested both in our experiments and in the larger combinatorial drug screen that had inconsistent synergy and effectiveness levels. The combination of Fak Inhibitor 14 and gefitinib was found to be synergistic in the combinatorial screen, however we found it to be non-synergistic in our experiments. Consequently we believe that our false discovery rate would have been even lower if we had tested the combinations in the same setting that was used to generate the training set. However the validation of the majority of our predictions across slightly variable contexts is highly relevant for the treatment of patient cancers, in which the treated patient populations involve different individuals and thus have slightly different genomic profiles than the population under which the therapy was conceived. Additionally we would like to emphasize that any effectiveness (or synergy) predictions that our model makes are in the context of mutant BRAF cell lines. While we hope that some of these findings may be translatable to human patients, many other factors must first be considered.

It is important to note that we did not consider maximum tolerated doses (MTD) in our analysis. A retrospective analysis revealed that there were two drugs included in our analysis that were tested at levels above MTD for humans: Obatoclax and Tamoxifen. Additionally there are many drugs included in the study do not have known MTD. This is because these drugs are experimental and thus have not yet had this evaluated in clinical trials(Marshall, 2012). While we do not believe that this biases the model, it does highlight the importance of clear model interpretation. In particular, we observed in our experiments that

synergy generally did not occur at high dosage levels (**Figure 4.5**), which further suggests that the model would not be biased by including drugs that were tested above MTD. Importantly our approach allows subsequent experimental studies to be prioritized on promising combinations, which can be focused on more clinically relevant information such as MTD.

The use of synergistic drug combinations has the potential to help prevent and overcome drug resistance. It is hypothesized that the application of drug combinations with initial treatment lower the odds of resistance occurring because it requires multiple mutations to bypass both drugs, which is a lower probability event than each drug individually(Fitzgerald et al., 2006; Komarova et al., 2013). This could be particularly impactful for combinations involving BRAF inhibitors, which individually have demonstrated remarkable responses in patients but suffer from the rapid development of resistance(Chapman et al., 2011). To further explore how models could be used to identify drug combinations that overcome drug resistance, we used our training set to derive a set of 24 effective and 12 non-effective combinations involving vemurafenib in resistant cell lines. We applied the framework of our approach to this small dataset and found that there is an underlying signal for predicting combinations that overcome resistance (AUC=0.677, Accuracy=0.697). Thus approaches such as ours may be applicable both to directly predicting combinations that overcome resistance, as well as predicting combinations that may help prevent resistance from developing.

Existing methods have previously found that the inclusion of features representing the biological mechanisms of the drug have been most successful(Bansal et al., 2014). However this information is often not available. Indeed only 50% (20/40) of the drugs in our training set have information about

the drug's target. Thus our method would likely improve as this type of information becomes more widely available enough to include in the model. However our model was able to exhibit significant predictive power despite this information not being available. We hypothesize that this was due to the large number of cell lines that each single agent was tested in. We found that generally the predicted synergistic combinations involved drugs that had varied single agent efficacies across the different mutant BRAF cell lines.

While we have trained and tested our approach in the context of BRAF mutant melanoma, the approach itself is applicable to other types of cancers. As additional large combinatorial screens become available, this methodology could prove to be impactful for the identification of drug synergy within the larger universe of possible drug combinations.

PERSPECTIVE

In recent years, enormous amounts of biomedical data has been generated and collected. These span from genomics to drug to other healthcare-related datasets. In genomics, large consortiums have focused on establishing large collections of patient sequencing data (eg. TCGA, 1000 Genomes, GTEx) and have identified and analyzed patterns that emerge (eg. EXaC, COSMIC). Similar efforts have been put forward in collecting information about drugs, such as characterizing their targets (DrugBank), general structure and properties (PubChem, ChEMBL), and their biological effects (NCI-DTP, CMAP, LINCS). Additional relevant information exists in other healthcare records, such as Electronic Medical Records (EMR/EHR), adverse event reports (FAERS), and clinical trial results.

The integration of these disparate datasets through machine learning and other quantitative methods remains largely unexplored, which creates an opportunity to answer interesting and important questions. Addressing inefficiencies in the drug development pipeline is one critical application with large potential. From drug repositioning to understanding drug mechanisms, there are many areas of the drug development process in which computational and data-driven methods could provide powerful insights. There is also potential in using these data to better understand disease and patient outcomes.

However there also remain many important challenges to working with and integrating these different data types. Pipelines for generating and analyzing genomics datasets are constantly evolving and vary across the field, which limits the extent to which different genomics datasets can be combined to make more powerful conclusions. In the drug space, inconsistent annotation and naming

nomenclature presents a similar problem. Furthermore there is a lack of clear and consistent definitions of clinical outcomes across healthcare datasets. The system of medical ontology used in health records varies between hospitals. Reports of clinical trial outcomes are often incomplete, with only 13.4% and 38.3% of trial results reported within 1 and 5 years respectively after trial completion (Anderson et al., 2015). Finally patient privacy concerns add an additional barrier to acquiring and leveraging datasets.

Addressing these challenges require large-scale efforts to be put into place. Fortunately many such efforts are already underway. In New York State and City, there is a public initiative to develop a comprehensive health information infrastructure with the goals of empowering better data-driven population health surveillance. In genomics, it is often mandated that data generated using grants or presented in publications be deposited in public repositories, including raw files, which will allow for better utilization of these datasets. However the problem of poor clinical trial reporting persists despite already existent federal mandates. This area of work, as well as many other federally funded projects, would benefit from both the mandate of data deposition and the follow-up of penalties for non-compliance. Additionally, most compounds fail before they reach clinical trials and most data related to these compounds is not released outside the company in which it was developed. As a result, information that could be relevant for the purposes of repositioning and novel drug design is not utilized, further contributing to inefficiencies in the drug development process.

Altogether there exists great promise for the integration of large, distinct biomedical datasets into machine learning and other quantitative models to answer important biological questions.

APPENDIX

MATERIALS AND METHDOS - CRAFTT

The CRAFTT approach

CRAFTT requires two inputs for its predictions, ChIP-seq for a TF that is used to derive its target gene set and drug-induced expression profiles. The CRAFTT procedure (1) uses Gene Set Enrichment Analysis with the target gene set for a TF and drug-induced expression profiles as inputs to make predictions about what drugs modulate TF's activity, and then (2) prioritizes predictions using network analysis. For the network analysis we compute a normalized path length score (*NPL*), in which we calculate the probability of observing the path length between the drug *d* and TF *X* ($P(PL|d,X)$) using 500 degree-preserving randomized networks(Gobbi et al., 2014). These steps are combined to generate a prediction score, which we term the modulation index (MI):

$$MI_{d,X} = \frac{NES_{d,X}}{P(PL|d,X)}$$

where $NPL = P(PL|d,X) = \frac{\sum_{i=1}^{500} PL_{d,X_{si}} < PL_{d_g,X_g}}{500}$

Statistical Analysis

The statistical significance for each of our predictions was estimated according to the GSEA procedure (Subramanian et al., 2005). For the analysis of Electronic Medical Records (EMR), Kaplan-Meier Survival Analysis was performed on an age-adjusted cohort, using time to first diagnosis of prostate cancer as the end-point in our study and excluding all patients with any prior diagnosis of cancer and cox proportional hazards test was used to assess significance. Statistical

analysis of RT-PCR, ChIP-PCR, cell invasion, and cell migration experiments was done in Prism using paired t test and $n = 3$ for each condition.

All other statistical significance values were calculated in R. The permutation test (using 1000 random permutations) was used to assess significance of the enrichment of TF alterations in COSMIC and the enrichment of drug categories (e.g. HDAC inhibitors) within our predictions. The chi-squared test was used to compare the transcription factor enrichment to that of kinases. The significance for the enrichment of known interactions was calculated using the exact binomial test, comparing the enrichment of known pairs to the total percentage of drug-pairs that were predicted. The Mann-Whitney test was used to assign significance to the network analysis and to the difference between GI50 (concentration required to inhibit 50% of growth) values in WT and MT p53 cell lines.

Experimental validation

RWPE1, VCaP and DU145 were obtained from ATCC and maintained according to manufacture's protocol. Isgogenic DU145 or RWPE1 +/-ERG cell lines were generated to over-express either truncated ERG as previously described(Rickman et al., 2010; Rickman et al., 2012). The PTEN^{-/-}/ERG^{Rosa26} prostate cancer cells were derived from PTEN^{-/-}/ERG^{Rosa26} prostate tumors as previously described(Chen et al., 2013). For treatments the cells PBS and incubated with the appropriate media at the indicated drug or vehicle dose for 24 or 48 hours.

Cells were then analyzed using ChIP-PCR, quantitative RT-PCR (QRT-PCR) or invasion/migration assay. Quantitative RT-PCR, ChIP-PCR, cell invasion and migration assays were performed as previously described(Rickman et al., 2010;

Rickman et al., 2012). Each sample was run in triplicate. The amounts of target genes were calculated relative to the reference gene HMBS.

MATERIALS AND METHDOS - PrOCTOR

Clinical Trials Training Set

We downloaded data from [ClinicalTrials.gov](https://clinicaltrials.gov) from The Database for Aggregate Analysis of ClinicalTrials.gov (AACT) ^{41 39 38 38 38 34 34 34 34 34 34 10 10}. To extract the names of the drugs associated with clinical trials that failed toxicity reasons, we identified any clinical trials that were annotated as “Terminated”, “Suspended” or “Withdrawn” and described as failing for toxicity reasons. The list of FDA approved drugs was obtained from the drug annotations within the DrugBank 4.0 database (Law et al., 2014).

Model Feature Derivation

Chemical Features

The structures (sdf format) were downloaded for all of the drugs in DrugBank. The molecular weight, polar surface area, hydrogen bond donor and acceptor counts, formal charge and number of rotatable bounds were extracted from the sdf file for each of these compounds. When that information was missing, it was filled in by querying PubChem or by computationally estimating these values using ChemmineR in R. The rule outcomes were then derived from these features. The QED values were computed using the author-released script.

Network features

We constructed the aggregated biological network by taking the union across multiple databases of gene-gene interactions. (Aksoy et al., 2013; Das et al.,

2012; Khurana et al., 2013). The network degree of a gene was calculated as the number of gene neighbors that a particular gene has. For drug's with multiple genes, the maximum value was take. The network betweenness for a particular gene (i.e. vertex) is defined as the number of shortest paths that travel through the vertex. This was calculated using the betweenness function in R's igraph package(Csardi et al., 2006).

Tissue features

The Gene RPKM RNA-Seq data from the Genotype-Tissue Expression (GTEx) project(Consortium, 2015) was downloaded from <http://www.gtexportal.org/home/>. This dataset has 2921 samples spanning 30 tissues. For each tissue, the median RPKM was calculated for each gene. For drugs with more than one target gene, the maximum RPKM was used.

Target Loss Frequency

The Exome Aggregation Consortium (ExAC) database (Exome Aggregation Consortium (ExAC)) was downloaded from www.exac.broadinstitute.org. For each gene, we counted the deleterious and total number of mutations that was reported. We calculated the loss frequency to be percentage of mutations that were reported in the gene that were deleterious.

The ProCTOR Model

We trained the ProCTOR approach on the clinical trials dataset using the features described above. It was trained using the random forest model, an ensembl decision tree based approach, which constructs 50 bootstrapped decision trees. A sub-sampling approach was used to account for the imbalanced ratio of approved drugs to FTT drugs, by randomly sampling the FDA approved class of samples to the size of the FTT drugs. To reduce the odds of poor

representatives being sampled, this was repeated 30 times. The labels were assigned by taking the consensus across the set of bootstrapped trees and replicates. This approach also yields a probability for each test sample. This probability was used to calculate an odds score = $\frac{P(\text{approval})}{P(\text{failure})}$. To better visualize the distribution of this score, the \log_2 of the odds score was used.

Independent Datasets

To further assess our approach, we applied ProCTOR to European (EMA) and Japanese (JP17) approved drugs, as well as 3236 drugs in DrugBank (version 4.2) (Law et al., 2014). The list of EMA-approved drugs were downloaded from the EMA website (<http://www.ema.europa.eu/ema>) and the JP17 list was downloaded from KEGG(Anders et al., 2015). Drugs that were already annotated as FDA approved were removed from these lists and the trained ProCTOR model was used to make predictions for the remaining drugs. The openFDA resource (<https://open.fda.gov>) was used to query adverse events of drugs in the DrugBank dataset but not in our training set. FDA annotated drug-induced liver toxicity (DILI). The DILI dataset was downloaded from the FDA website at <http://www.fda.gov/ScienceResearch/BioinformaticsTools/LiverToxicityKnowledgeBase/ucm226811.htm>. The SIDER side effect resource database (Kuhn et al., 2010) was used to annotate side effects of each drug in the clinical trials dataset. The *meddra_adverse_effects.txt* table was used to extract reported adverse events, using the *MedDRA Preferred Term* descriptor to group similar side effects.

Statistical Analyses

We used area under the receiver operating characteristic (ROC) curve and 10-fold cross validation to evaluate the predictive power of our approach. For the independent analysis of predictions in the DrugBank dataset, we tested for enrichment of drug classes using the binomial test. We tested for differences of serious adverse event frequency between predicted toxic (score<-1) and predicted safe (score>1) drugs in the DrugBank dataset and not in the training set using the unpaired Student's t-test. For the EMA, JP17, and DILI datasets, we tested for differences in ProCTOR scores between predictions using the Mann-Whitney *U* Test. For the side effects of drugs in the training set, we used the Fisher's Exact Test to identify the side effects that occurred more frequently in predicted toxic drugs using a *p*-value cutoff of 0.01.

MATERIALS AND METHODS – Drug Combination Predictions

The Model

150 Single agent and 780 combinatorial efficacies were obtained from the Held *et al* (Held et al., 2013) study. The single agent efficacies were collapsed to their GI_{50} values, which is the concentration of the drug required to inhibit 50% of cell growth. Features representing a drug pair were constructed by taking the mean and difference (**Figure 4.1**) of the GI_{50} values for each of 27 tested cell lines (15 mutant BRAF, 6 mutant RAS, 6 wtBRAF/wtRAS).

The combinatorial results in 19 cell lines (8 mutant BRAF, 6 mutant RAS, and 5 WT) were then used to construct labels for each of the 780 drug pairs. Genotype-selective combinations were defined to be those that yielded an average 15% or greater growth inhibition exclusively in the genotypic group and achieved at least

50% growth inhibition within the genotypic group and a general effective combination was defined to be one that achieved at least 70% growth inhibition. Synergy labels for each combination were determined using the Chou-Talalay synergy combination index (CI) metric. We defined a synergistic combination to be a pair of drugs that yielded a CI less than -1 at any concentration level. A random forest model was then trained on the above-described data and evaluated using 10-fold cross-validation. Predictions were made using the trained model for the 10,395 untested combinations that had single agent efficacy information in the dataset.

Experimental validation

Melanoma Cell Culture

Malme-3m, SK-Mel-28, and SK-MEL-2 cell lines were generously donated by the Houvras lab at Weill Cornell Medicine. Cell lines were cultured in basal medium [DMEM (Gibco) supplemented with 10% FBS and 1% penicillin/streptomycin (P/S)] and maintained in a 37° incubator at 5% CO₂.

Agent Screening

Cells were pipetted into 384-well plates at 750 cells per well using a multi-channel pipette (Eppendorf) in 20 ul basal medium, and placed in an 37° incubator at 5% CO₂ overnight prior to exposure to the agent. Drug stock plates for single therapy agents were created by serial dilution, generally 1:10 from 10 mmol/L stock. Using a multi-channel pipette, 2.5 ul drug volume from drug stock plates were added to the 384-well cell plates. A total of 0.1% and 10% DMSO was used as negative and positive controls, respectively.

For combinatorial, dual-agent screens, a 2.5 ul volume of Drug X was plated at GI-50 concentration (determined from single-agent screen), and 2.5 ul volume

from drug stock plate of Drug Y was added in a range of concentrations. Initial dual agent screening was carried out in 384-well plates at 750 cells per well in 20 ul basal medium. Further exploration screening were carried out in 96-well plates, in 100ul basal medium with 3500 cells per well.

All experiments were carried out in triplicates at the Weill Cornell Medicine's Meyer Cancer Center (WCMC MCC, New York, NY). Cells were exposed to drugs for 72 hours, followed by GI measurement by CellTiter-Glo ATP detection assay (Promega) following the manufacturer's recommended. A 20 ul volume of basal medium served as a background for luminescence.

Prediction Evaluation

A combination was considered to be effective if the third quantile value of observed growth inhibition values was at least 70%. Synergistic combinations were those that achieved a Chou-Talalay synergy score of -1 in at least one dosage-level pair.

References

- AACT database. Clinical Trials Transformation Initiative website.
- Ablain, J., Nasr, R., Bazarbachi, A., and de The, H. (2011). The drug-induced degradation of oncoproteins: an unexpected Achilles' heel of cancer cells? *Cancer Discov* *1*, 117-127.
- Aksoy, B.A., Gao, J., Dresdner, G., Wang, W., Root, A., Jing, X., Cerami, E., and Sander, C. (2013). PiHelper: an open source framework for drug-target and antibody-target data. *Bioinformatics* *29*, 2071-2072.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* *31*, 166-169.
- Anderson, M.L., Chiswell, K., Peterson, E.D., Tasneem, A., Topping, J., and Califf, R.M. (2015). Compliance with Results Reporting at ClinicalTrials.gov. *New England Journal of Medicine* *372*, 1031-1039.
- Arastu-Kapur, S., Anderl, J.L., Kraus, M., Parlati, F., Shenk, K.D., Lee, S.J., Muchamuel, T., Bennett, M.K., Driessen, C., Ball, A.J., *et al.* (2011). Nonproteasomal targets of the proteasome inhibitors bortezomib and carfilzomib: a link to clinical adverse events. *Clin Cancer Res* *17*, 2734-2743.
- Bansal, M., Yang, J., Karan, C., Menden, M.P., Costello, J.C., Tang, H., Xiao, G., Li, Y., Allen, J., Zhong, R., *et al.* (2014). A community computational challenge to predict the activity of pairs of compounds. *Nature biotechnology* *32*, 1213-1222.
- Bickerton, G.R., Paolini, G.V., Besnard, J., Muresan, S., and Hopkins, A.L. (2012). Quantifying the chemical beauty of drugs. *Nature chemistry* *4*, 90-98.
- Blind, E., Dunder, K., de Graeff, P.A., and Abadie, E. (2011). Rosiglitazone: a European regulatory perspective. *Diabetologia* *54*, 213-218.
- Breiman, L. (2001a). Random forests. *Machine Learning* *45*, 5– 32.
- Breiman, L. (2001b). Random forests. *Machine learning* *45*, 5-32.
- Campbell, J.D., McDonough, J.E., Zeskind, J.E., Hackett, T.L., Pechkovsky, D.V., Brandsma, C.A., Suzuki, M., Gosselink, J.V., Liu, G., Alekseyev, Y.O., *et al.* (2012). A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK. *Genome medicine* *4*, 67.
- Chapman, P.B., Hauschild, A., Robert, C., Haanen, J.B., Ascierto, P., Larkin, J., Dummer, R., Garbe, C., Testori, A., Maio, M., *et al.* (2011). Improved survival with

vemurafenib in melanoma with BRAF V600E mutation. *The New England journal of medicine* 364, 2507-2516.

Chen, D.W., Saha, V., Liu, J.Z., Schwartz, J.M., and Krstic-Demonacos, M. (2013). Erg and AP-1 as determinants of glucocorticoid response in acute lymphoblastic leukemia. *Oncogene* 32, 3039-3048.

Chou, T.C. (2010). Drug combination studies and their synergy quantification using the Chou-Talalay method. *Cancer Res* 70, 440-446.

Cokol, M., Chua, H.N., Tasan, M., Mutlu, B., Weinstein, Z.B., Suzuki, Y., Nergiz, M.E., Costanzo, M., Baryshnikova, A., Giaever, G., *et al.* (2011). Systematic exploration of synergistic drug pairs. *Mol Syst Biol* 7, 544.

Conde, R., Belak, Z.R., Nair, M., O'Carroll, R.F., and Ovsenek, N. (2009). Modulation of Hsf1 activity by novobiocin and geldanamycin. *Biochemistry and cell biology = Biochimie et biologie cellulaire* 87, 845-851.

Consortium, G.T. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348, 648-660.

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695.

Curtin, J.A., Fridlyand, J., Kageshita, T., Patel, H.N., Busam, K.J., Kutzner, H., Cho, K.H., Aiba, S., Brocker, E.B., LeBoit, P.E., *et al.* (2005). Distinct sets of genetic alterations in melanoma. *The New England journal of medicine* 353, 2135-2147.

Darnell, J.E., Jr. (2002). Transcription factors as targets for cancer therapy. *Nat Rev Cancer* 2, 740-749.

Das, J., and Yu, H. (2012). HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC systems biology* 6, 92.

Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastiris, E., Gilpatrick, T., Paranal, R.M., Qi, J., *et al.* (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* 146, 904-917.

Dudley, J.T., Sirota, M., Shenoy, M., Pai, R.K., Roedder, S., Chiang, A.P., Morgan, A.A., Sarwal, M.M., Pasricha, P.J., and Butte, A.J. (2011). Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Science translational medicine* 3, 96ra76.

Eduati, F., Mangravite, L.M., Wang, T., Tang, H., Bare, J.C., Huang, R., Norman, T., Kellen, M., Menden, M.P., Yang, J., *et al.* (2015). Prediction of human population

responses to toxic compounds by a collaborative competition. *Nat Biotechnol* 33, 933-940.

Elemento, O., Rubin, M.A., and Rickman, D.S. (2012). Oncogenic transcription factors as master regulators of chromatin topology: a new role for ERG in prostate cancer. *Cell cycle* 11, 3380-3383.

Encode Project Consortium (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS biology* 9, e1001046.

Exome Aggregation Consortium (ExAC) (Cambridge, MA).

Fitzgerald, J.B., Schoeberl, B., Nielsen, U.B., and Sorger, P.K. (2006). Systems biology and combination therapy in the quest for clinical efficacy. *Nat Chem Biol* 2, 458-466.

Foucquier, J., and Guedj, M. (2015). Analysis of drug combinations: current methodological landscape. *Pharmacol Res Perspect* 3, e00149.

Friedman, A.A., Amzallag, A., Pruteanu-Malinici, I., Baniya, S., Cooper, Z.A., Piris, A., Hargreaves, L., Igras, V., Frederick, D.T., Lawrence, D.P., *et al.* (2015). Landscape of Targeted Anti-Cancer Drug Synergies in Melanoma Identifies a Novel BRAF-VEGFR/PDGFR Combination Treatment. *PLoS one* 10, e0140310.

Gashaw, I., Ellinghaus, P., Sommer, A., and Asadullah, K. (2011). What makes a good drug target? *Drug Discov Today* 16, 1037-1043.

Ghose, A.K., Viswanadhan, V.N., and Wendoloski, J.J. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *Journal of combinatorial chemistry* 1, 55-68.

Gobbi, A., Iorio, F., Dawson, K.J., Wedge, D.C., Tamborero, D., Alexandrov, L.B., Lopez-Bigas, N., Garnett, M.J., Jurman, G., and Saez-Rodriguez, J. (2014). Fast randomization of large genomic datasets while preserving alteration counts. *Bioinformatics* 30, i617-623.

Hay, M., Thomas, D.W., Craighead, J.L., Economides, C., and Rosenthal, J. (2014). Clinical development success rates for investigational drugs. *Nat Biotechnol* 32, 40-51.

Held, M.A., Langdon, C.G., Platt, J.T., Graham-Steed, T., Liu, Z., Chakraborty, A., Bacchicchi, A., Koo, A., Haskins, J.W., Bosenberg, M.W., *et al.* (2013). Genotype-selective combination therapies for melanoma identified by high-throughput drug screening. *Cancer discovery* 3, 52-67.

- Hurle, M.R., Yang, L., Xie, Q., Rajpal, D.K., Sanseau, P., and Agarwal, P. (2013). Computational drug repositioning: from data to therapeutics. *Clinical pharmacology and therapeutics* 93, 335-341.
- Jeliazkova, N., and Jeliazkov, V. (2011). AMBIT RESTful web services: an implementation of the OpenTox application programming interface. *J Cheminform* 3, 18.
- Khurana, E., Fu, Y., Chen, J., and Gerstein, M. (2013). Interpretation of genomic variants using a unified biological network approach. *PLoS computational biology* 9, e1002886.
- Kittler, R., Zhou, J., Hua, S., Ma, L., Liu, Y., Pendleton, E., Cheng, C., Gerstein, M., and White, K.P. (2013). A comprehensive nuclear receptor network for breast cancer cells. *Cell reports* 3, 538-551.
- Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V., *et al.* (2011). DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic acids research* 39, D1035-1041.
- Komarova, N.L., and Boland, C.R. (2013). Cancer: calculated treatment. *Nature* 499, 291-292.
- Kuhn, M., Campillos, M., Letunic, I., Jensen, L.J., and Bork, P. (2010). A side effect resource to capture phenotypic effects of drugs. *Molecular systems biology* 6, 343.
- Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., *et al.* (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929-1935.
- Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V., *et al.* (2014). DrugBank 4.0: shedding new light on drug metabolism. *Nucleic acids research* 42, D1091-1097.
- Ledford, H. (2011). Translational research: 4 ways to fix the clinical trial. *Nature* 477, 526-528.
- Leeson, P.D., and Springthorpe, B. (2007). The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov* 6, 881-890.
- Lewis, M. (2003). *Moneyball : the art of winning an unfair game*, 1st edn (New York: W.W. Norton).
- Li, Y.Y., and Jones, S.J. (2012). Drug repositioning for personalized medicine. *Genome medicine* 4, 27.

- Libermann, T.A., and Zerbini, L.F. (2006). Targeting transcription factors for cancer gene therapy. *Curr Gene Ther* 6, 17-33.
- Lipinski, C.A. (2004). Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol* 1, 337-341.
- Lipinski, C.A., Lombardo, F., Dominy, B.W., and Feeney, P.J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews* 46, 3-26.
- Marshall, J. (2012). Maximum-tolerated dose, optimum biologic dose, or optimum clinical value: dosing determination of cancer therapies. *J Clin Oncol* 30, 2815-2816.
- Massacesi, C., Marcucci, F., Rocchi, M.B., Mazzanti, P., Pilone, A., and Bonsignori, M. (2004). Factors predicting docetaxel-related toxicity: experience at a single institution. *J Chemother* 16, 86-93.
- Mees, C., Nemunaitis, J., and Senzer, N. (2009). Transcription factors: their potential as targets for an individualized therapeutic approach to cancer. *Cancer gene therapy* 16, 103-112.
- Nissen, S.E., and Wolski, K. (2007). Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 356, 2457-2471.
- Patlewicz, G., and Fitzpatrick, J.M. (2016). Current and Future Perspectives on the Development, Evaluation, and Application of in Silico Approaches for Predicting Toxicity. *Chemical research in toxicology* 29, 438-451.
- Pflugfelder, A., Eigentler, T.K., Keim, U., Weide, B., Leiter, U., Ikenberg, K., Berneburg, M., and Garbe, C. (2011). Effectiveness of carboplatin and paclitaxel as first- and second-line treatment in 61 patients with metastatic melanoma. *PloS one* 6, e16882.
- Puissant, A., Frumm, S.M., Alexe, G., Bassil, C.F., Qi, J., Chanthery, Y.H., Nekritz, E.A., Zeid, R., Gustafson, W.C., Greninger, P., *et al.* (2013). Targeting MYCN in neuroblastoma by BET bromodomain inhibition. *Cancer Discov* 3, 308-323.
- Puisset, F., Alexandre, J., Treluyer, J.M., Raoul, V., Roche, H., Goldwasser, F., and Chatelut, E. (2007). Clinical pharmacodynamic factors in docetaxel toxicity. *Br J Cancer* 97, 290-296.
- Reinhold, W.C., Sunshine, M., Liu, H., Varma, S., Kohn, K.W., Morris, J., Doroshow, J., and Pommier, Y. (2012). CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer research* 72, 3499-3511.

Rickman, D.S., Chen, Y.B., Banerjee, S., Pan, Y., Yu, J., Vuong, T., Perner, S., Lafargue, C.J., Mertz, K.D., Setlur, S.R., *et al.* (2010). ERG cooperates with androgen receptor in regulating trefoil factor 3 in prostate cancer disease progression. *Neoplasia* *12*, 1031-1040.

Rickman, D.S., Soong, T.D., Moss, B., Mosquera, J.M., Dlabal, J., Terry, S., MacDonald, T.Y., Tripodi, J., Bunting, K., Najfeld, V., *et al.* (2012). Oncogene-mediated alterations in chromatin conformation. *Proceedings of the National Academy of Sciences of the United States of America* *109*, 9083-9088.

Robert, C., Karaszewska, B., Schachter, J., Rutkowski, P., Mackiewicz, A., Stroiakovski, D., Lichinitser, M., Dummer, R., Grange, F., Mortier, L., *et al.* (2015). Improved overall survival in melanoma with combined dabrafenib and trametinib. *The New England journal of medicine* *372*, 30-39.

Sahota, T., Danhof, M., and Della Pasqua, O. (2016). Pharmacology-based toxicity assessment: towards quantitative risk prediction in humans. *Mutagenesis* *31*, 359-374.

Scannell, J.W., Blanckley, A., Boldon, H., and Warrington, B. (2012). Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discov* *11*, 191-200.

Shanks, N., Greek, R., and Greek, J. (2009). Are animal models predictive for humans? *Philos Ethics Humanit Med* *4*, 2.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* *102*, 15545-15550.

Thang, N.D., Nghia, P.T., Kumasaka, M.Y., Yajima, I., and Kato, M. (2015). Treatment of vemurafenib-resistant SKMEL-28 melanoma cells with paclitaxel. *Asian Pac J Cancer Prev* *16*, 699-705.

Tomlins, S.A., Laxman, B., Varambally, S., Cao, X., Yu, J., Helgeson, B.E., Cao, Q., Prensner, J.R., Rubin, M.A., Shah, R.B., *et al.* (2008). Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia* *10*, 177-188.

USEPA (2016). ToxCast & Tox21 Chemicals Distributed Structure-Searchable Toxicity Database from DSSTox_20151019.

Veber, D.F., Johnson, S.R., Cheng, H.Y., Smith, B.R., Ward, K.W., and Kopple, K.D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry* *45*, 2615-2623.

Venkitaraman, R., Thomas, K., Murthy, V., Woode-Amisshah, R., Dearnaley, D.P., Horwich, A., Huddart, R.A., De Bono, J.S., and Parker, C. (2013). A randomized phase II trial of dexamethasone versus prednisolone as secondary hormonal therapy in CRPC. Paper presented at: Genitourinary Cancers Symposium (Orlando, FL: Journal of Clinical Oncology).

Xie, L., Gazin, C., Park, S.M., Zhu, L.J., Debily, M.A., Kittler, E.L., Zapp, M.L., Lapointe, D., Gobeil, S., Virbasius, C.M., *et al.* (2012). A synthetic interaction screen identifies factors selectively required for proliferation and TERT transcription in p53-deficient human cancer cells. *PLoS genetics* 8, e1003151.

Yeh, P.J., Hegreness, M.J., Aiden, A.P., and Kishony, R. (2009). Drug interactions and the evolution of antibiotic resistance. *Nat Rev Microbiol* 7, 460-466.