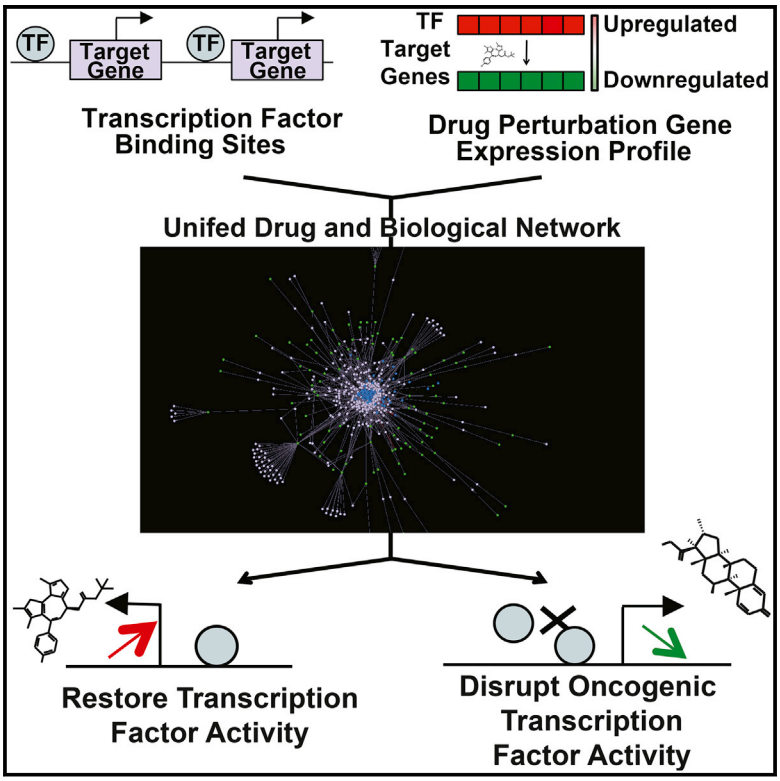
Report

Cell Reports

A Computational Drug Repositioning Approach for Targeting Oncogenic Transcription Factors

Graphical Abstract



Highlights

- A computational approach predicts drugs that modulate transcription factor activity
- Known drug-transcription factor interactions are recovered
- Dexamethasone is identified as a modulator of ERG activity
- Experimental data functionally validate dexamethasone-ERG interaction

Authors

Kaitlyn M. Gayvert, Etienne Dardenne, Cynthia Cheung, ..., Nicholas P. Tatonetti, David S. Rickman, Olivier Elemento

Correspondence

dsr2005@med.cornell.edu (D.S.R.), ole2001@med.cornell.edu (O.E.)

In Brief

Gayvert et al. present a broadly applicable systems biology method for identifying small molecules and drugs that modulate transcription factor activity. They identified dexamethasone as a candidate for the inhibition of the oncogenic transcription factor ERG and validate this prediction experimentally in several systems.

Gayvert et al., 2016, Cell Reports *15*, 2348–2356
June 14, 2016
http://dx.doi.org/10.1016/j.celrep.2016.05.037

OPEN ACCESS
CellPress

Cell Reports
Report

A Computational Drug Repositioning Approach for Targeting Oncogenic Transcription Factors

Kaitlyn M. Gayvert,[1,2,3] Etienne Dardenne,[4] Cynthia Cheung,[4] Mary Regina Boland,[5] Tal Lorberbaum,[5,6] Jackline Wanjala,[7] Yu Chen,[7] Mark A. Rubin,[2,4] Nicholas P. Tatonetti,[5] David S. Rickman,[2,4,*] and Olivier Elemento[1,2,*]

[1]Institute for Computational Biomedicine, Department of Physiology and Biophysics, Weill Cornell Medical College, New York, NY 10021, USA
[2]Institute for Precision Medicine, Weill Cornell Medical College, New York, NY 10021, USA
[3]Tri-Institutional Training Program in Computational Biology and Medicine, New York, NY 10065, USA
[4]Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, NY 10021, USA
[5]Department of Biomedical Informatics, Columbia University, New York, NY 10027, USA
[6]Department of Physiology and Cellular Biophysics, Columbia University, New York, NY 10032, USA
[7]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA
*Correspondence: dsr2005@med.cornell.edu (D.S.R.), ole2001@med.cornell.edu (O.E.)
http://dx.doi.org/10.1016/j.celrep.2016.05.037

SUMMARY

Mutations in transcription factor (TF) genes are frequently observed in tumors, often leading to aberrant transcriptional activity. Unfortunately, TFs are often considered undruggable due to the absence of targetable enzymatic activity. To address this problem, we developed CRAFTT, a computational drug-repositioning approach for targeting TF activity. CRAFTT combines ChIP-seq with drug-induced expression profiling to identify small molecules that can specifically perturb TF activity. Application to ENCODE ChIP-seq datasets revealed known drug-TF interactions, and a global drug-protein network analysis supported these predictions. Application of CRAFTT to ERG, a pro-invasive, frequently overexpressed oncogenic TF, predicted that dexamethasone would inhibit ERG activity. Dexamethasone significantly decreased cell invasion and migration in an ERG-dependent manner. Furthermore, analysis of electronic medical record data indicates a protective role for dexamethasone against prostate cancer. Altogether, our method provides a broadly applicable strategy for identifying drugs that specifically modulate TF activity.

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). One classic example of a recurrently altered TF is the tumor suppressor TF gene p53, which is mutated in up to 40% of human tumors (Libermann and Zerbini, 2006) yet has remained a highly elusive target for reactivation (Mees et al., 2009). Other examples include c-Myc, which is also among the most commonly altered genes in cancer (Ablain et al., 2011), ERG, and other ETS-family factors, which are fused to the androgen-controlled promoters in more than 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 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 affect 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 and Zerbini, 2006), and disruption or activation of upstream signaling mechanisms (Mees et al., 2009). Disrupting interactions with cofactors 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, has been 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 cofactors 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.

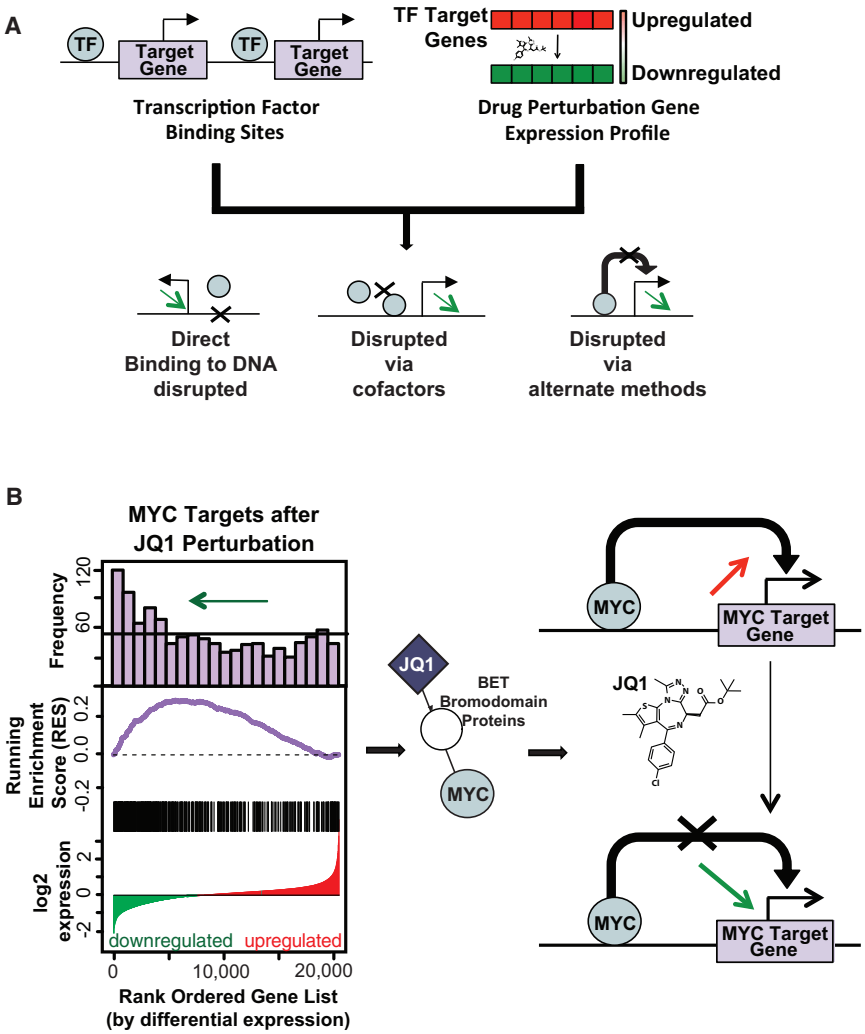


Figure 1. Methodology Overview

(A) Alterations in TFs 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 overexpression 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. See also Figure S1.

TFs constitute a significant proportion (18.1%) of the genes in the Sanger Institute's Cancer Gene Census. This confirmed that the prevalence of genomic alterations in TF genes in cancer is substantial and further indicates that TFs should constitute a major class of anticancer drug targets.

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 (Figure 1A). We propose CRAFTT, which 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 chromatin immunoprecipitation sequencing (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) (Figure 1A) (Lamb et al., 2006), and generating ranked gene lists by sorting the genes from most downregulated to most upregulated 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 (see Experimental Procedures). Each GSEA yields a normalized enrichment score (NES) and a corresponding p value

To address this unmet need, we developed CRAFTT, a computational drug-repositioning approach for targeting TFs. Altogether, our method provides a broadly applicable strategy to identify drugs and small molecules that specifically target the activity of individual TFs. Because a significant number of tumors are driven by oncogenic TFs or have lost tumor-suppressive TFs, our approach could potentially affect 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 the Catalogue of Somatic Mutations in Cancer (COSMIC) reported a mutation in a TF (Figure S1). Furthermore,

indicating whether the TF target gene set is mobilized as a whole by the drug, either toward downregulation (NES > 0) or toward upregulation (NES < 0). The 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 true drug-TF modulatory interactions, the network path between a drug and its 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 cofactors. More broadly, we expected that the drug and target TFs would be functionally related and therefore be located near 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 and Yu, 2012; Khurana et al., 2013), which include both physical (protein-protein) and non-physical (phosphorylation, metabolic, signaling, and 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 for the drug-TF pair ($NES_{d,TF}$) by the normalized network path length ($NPL_{d,TF}$) (see Experimental Procedures). The proposed approach does not make any assumptions about the mechanisms by which a drug can disrupt the expression program of TFs (Figure 1A). Such disruption can occur in a variety of ways, e.g., disruption of interaction with cofactors and DNA binding disruption.

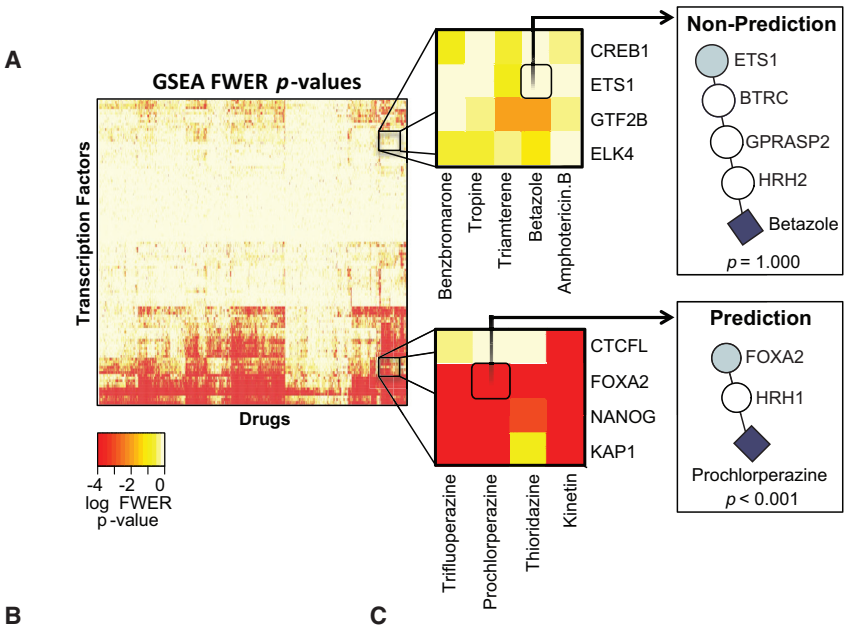
As a first proof of principle, we applied this approach to JQ1-induced gene expression profiles derived from another study (Puissant et al., 2013), all CMap drug-induced expression profiles (Lamb et al., 2006), and MYC direct target genes, which were derived from Encyclopedia of DNA Elements (ENCODE) ChIP-seq data (ENCODE Project Consortium, 2011). We found that JQ1 significantly downregulated 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 shortest possible network path length of 2, given the underlying mechanisms of the true interaction. This indicated that JQ1 is the best candidate ($MI_{JQ1,MYC}$ = 5,120) 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 1B).

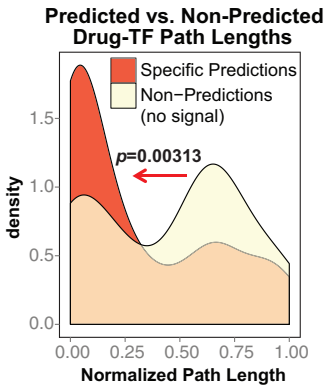
Systematic Drug-TF Analysis Predicts that Candidate Small Molecules Can Disrupt TFs

Next, we 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 (Figure S2) (Lamb et al., 2006). This approach identified 37,638 candidate drug-TF pairs (out of 218,603 possible combinations) (Figure 2A). 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 1,308 drugs that we have greater confidence in due to the selectivity of the prediction (see Supplemental Experimental Procedures for more details). Top specific predictions for both drug-induced inhibition and drug-induced activation of each TF are shown in Table S1.

Several predicted drug-TF interactions are consistent with the known activity of the drugs involved. For example, all four known heat shock protein 90 (HSP90) inhibitors that were included both 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 monorden (FWER = 0.054), 17-AAG (FWER = 0.031), 17-DMAG (FWER = 0.085), and geldanamycin (FWER < 0.001). In addition, novobiocin, whose antagonism of HSP90 is reported in literature but was not annotated in our network, was 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, which can be found in Table S2.

Because experimental validations are not available for most 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 the CMap datasets. Most of these combinations involved a glucocorticoid receptor (GR) agonist (26 combinations) or a HDAC inhibitor (7 combinations). Of the 35 known drug-TF combinations, CRAFTT was able to correctly predict more than expected (n = 21, p = 1.708e-8, binomial test). In particular, CRAFTT predicted well both the GR agonists (p = 6.524e-8, binomial test) and the HDAC inhibitors (p = 0.01978, binomial test). 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. In addition, about 85% of these combinations were nominally significant (p = 3.42e-8), 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 2,300 target genes (see Supplemental Experimental Procedures for more details).





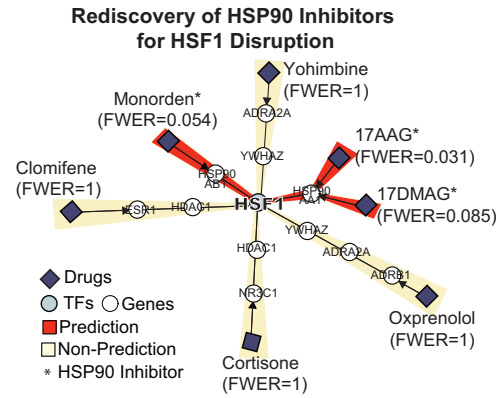


Figure 2. Systematic Analysis of 166 TFs and 1,309 Drug Perturbation Experiments Identifies Approximately 38,000 Candidate TF-Drug Pairs

(A) Heatmap of the FWER p values for all TF-drug pairs involving 166 ChIP-seq experiments from ENCODE and the 1,309 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 those of 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.

(B) 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.

(C) Network visualization of HSF1, three HSP90 inhibitors covered in CMap and our network (monorden, 17-AAG, and 17-DMAG), and four other drugs not predicted to disrupt HSF1 disruption (clomifene, yohimbine, oxprenolol, and cortisone).

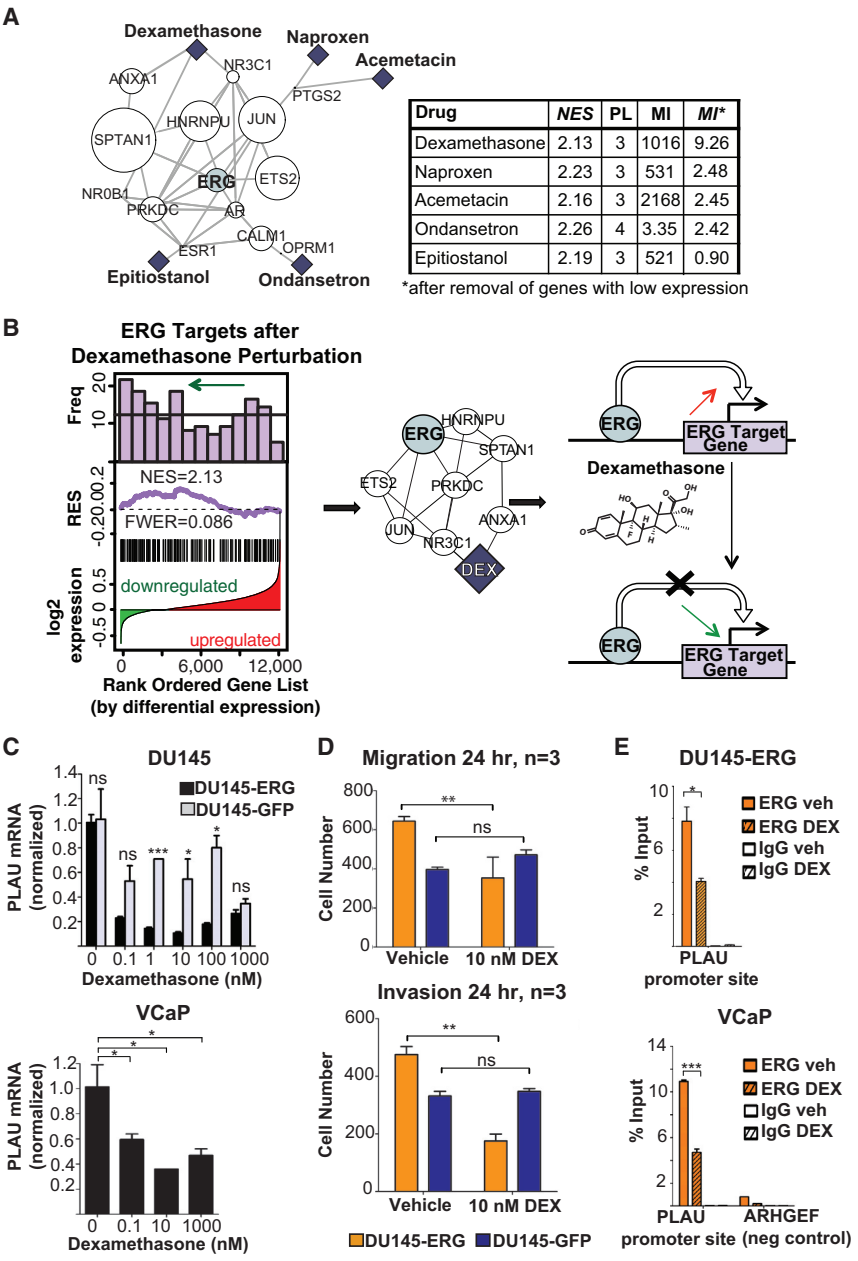
See also Tables S1 and S2.

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 earlier, we reasoned that true drug-TF interactions should be short, given the underlying mechanisms of the interactions (Figure 2A). Network analysis 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 2B). This is illustrated in Figure 2C, 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 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 cofactors. 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) (see Experimental Procedures).

Identification and Validation of Small Molecules that Inhibit the TF ERG

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 overexpression resulting from 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 applied our approach to all CMap drug profiles to identify candidate drugs for inhibition of ERG.

From the prediction step of CRAFTT, we identified eight candidate drugs that downregulate ERG target genes: dexamethasone (FWER = 0.086), naproxen (FWER = 0.048), acemetacin (FWER = 0.087), ondansetron (FWER = 0.061), epitiostanol (FWER = 0.069), diloxanide (FWER = 0.003), methanthelinium bromide (FWER = 0.046), and isoflupredone (FWER = 0.088). Five of these candidate drugs were contained in our biological network: dexamethasone (MI = 1,015.85), naproxen (MI = 530.90), acemetacin (MI = 2,167.88), ondansetron (MI = 3.35), and epitiostanol (MI = 520.99) (Figure 3A). An initial network analysis suggested that dexamethasone, naproxen, acemetacin,



Drug	*NES*	PL	MI	*MI**
Dexamethasone	2.13	3	1016	9.26
Naproxen	2.23	3	531	2.48
Acemetacin	2.16	3	2168	2.45
Ondansetron	2.26	4	3.35	2.42
Epitiostanol	2.19	3	521	0.90

*after removal of genes with low expression

Figure 3. 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, acemetacin, ondansetron, and 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 acemetacin are increased while the paths from ondansetron and epitiostanol are disrupted. The corresponding table shows metrics that describe each of these drugs in relation to ERG: NES is obtained from GSEA, PL is the shortest network path length required to connect ERG to the drug, and MI* is the MI 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 overexpression regions. The middle panel shows a subnetwork that includes 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 and VCaP) and controls (DU145-GFP) after treatment with vehicle or dexamethasone.

(D) Cell invasion and migration in cell lines expressing ERG (DU145-ERG) and controls (DU145-GFP). The data are shown at n = 4 representation 10× field of view.

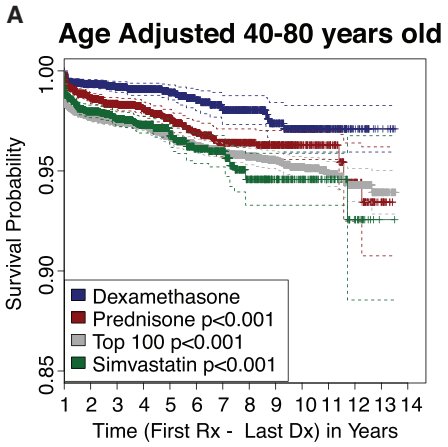
(E) The binding of ERG and a control (immunoglobulin G) by ChIP-PCR at the promoter of its target gene *PLAU* and at a negative control (ARHGEF) in cell lines expressing ERG (DU145-ERG and VCaP).

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. See also Figure S2.

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 (RNA sequencing) from RWPE1 prostate cells to filter out genes that have low expression in the network, which we defined as reads per kb of transcript per million mapped reads (RPKM) < 4. This analysis resulted in dexamethasone being identified as the drug with the shortest path length and highest-modified MI (9.26) (Figure 3B).

Because dexamethasone has not been previously 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*), 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 VCaP cells with high endogenous levels of ERG (Figure 3C). In comparison, dexamethasone was weakly active in the control GFP cells (Figure 3C).To further test the inhibitory



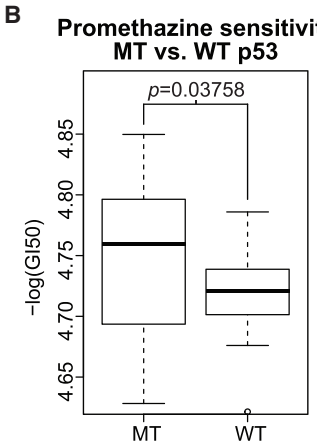


Figure 4. Extended Analyses of CRAFTT Predictions

(A) 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 the top-100 prescribed drugs. Statistical significance was assessed using a Cox proportional hazards test for comparison of dexamethasone to each other drug.

(B) The drug GI_{50} in mutant p53 and wild-type p53 cell lines in the National Cancer Institute (NCI) DTP. Statistical significance was assessed using the Mann-Whitney test.

See also Figures S3 and S4.

effect of dexamethasone on ERG activity, we treated a ERG-overexpressing cell line newly 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 S2A).

Using cell invasion and migration assays, we then found that dexamethasone significantly decreased cell invasion and migration in DU145 prostate cancer cells overexpressing ERG but not in isogenic control cells (Figure 3D; Figure S2B). High-resolution microscopic images revealed that dexamethasone helps the cells partially regain polarity, which may be a potential mechanism for reduced cell invasion (Figure S2C). 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 S2D). 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 3E). 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 GR agonist, which suggests that GR, encoded by *NR3C1*, may play a role in ERG-mediated gene expression. We found that small interfering RNAs targeting *NR3C1* mRNA lowered GR levels by 80% in the DU145-ERG cells (Figure S2E). Although GR seems to play a role in *PLAU* regulation in the absence of ERG, lowering GR levels did not significantly alter dexamethasone's impact on *PLAU* expression in ERG-positive cells (Figure S2F). In addition, 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 S2G). Altogether, these results indicate that dexamethasone-mediated ERG inhibition occurs independent of GR and AR signaling.

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

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 (Figure S3). 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 did 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 4A). We next constructed a logistic regression model to assess the relationship of the dexamethasone and other control treatments and the 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 be both protective against prostate cancer (perhaps through its inhibitory effect on ERG-rearranged tumors, as predicted in this study) and more active than prednisolone in its protective effect and in the treatment of CRPC. These results are still largely correlative in the absence of ERG molecular status for electronic medical record (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) among our reactivation predictions, indicating that CRAFTT is successful in identifying true drug-TF interactions (Figure S4). 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 Developmental Therapeutics Program (DTP)-NCI60 drug sensitivity data (Reinhold et al., 2012) supported this prediction, because we found that the mutant p53 cell lines were significantly more sensitive to promethazine than were the wild-type p53 cell lines ($p = 0.0376$, Mann-Whitney test) (Figure 4B). We next looked to see whether any predicted drugs for p53 activity reactivation targeted genes that had been previously identified as necessary for growth in *TP53*-deficient cells (Xie et al., 2012). From that list, we found seven of the drugs predicted by CRAFTT to reactivate p53 activity target genes: pentetrazol, naftopidil, oxedrine, capsaicin, ifenprodil, flumetasone, and dexpropranolol. Altogether, this suggests that our approach can be used to identify candidates for reactivation of TFs frequently lost in cancer.

DISCUSSION

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

We first looked to see whether 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 rediscover these connections. We then used CRAFTT to identify dexamethasone as a candidate for inhibition of ERG activity; follow-up experiments supported this prediction. We also found that dexamethasone had a similar effect in isolated mouse cell lines and in human cell lines. This suggests that mouse models could be used for further follow-up on the therapeutic use of dexamethasone in treatment of the ERG-overexpression cancer subtypes.

CRAFTT was successful in the identification of drugs for affecting transcriptional activity, but some areas 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 suggests that our network is robust to missing network edges (see Supplemental Experimental Procedures). 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, respectively, than average for a drug-TF pair with the same network degree.

In addition, the ChIP-seq that we used to derive binding site data was obtained from wild-type TFs. However, our approach was able to capture true drug-TF interactions, at least partly because these variants often cause constitutive expression and binding of the TF instead of dramatic disruption and changes to binding sites. As more mutant TF binding data become 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 prone to error. While we have taken steps to control for binding hotspots, our method will benefit as improved peak calling methods become available.

Finally, the CMap data that we analyzed was in a collapsed format, which limits the robustness of the predictions. The Broad Institute has released an updated version of CMap, which includes a 1,000-fold scale-up and will better allow us to use the variability in replicates. We also intend to apply CRAFTT to identification of candidate drugs for modulating the activity of other TFs that are historically elusive but desirable for targeting, such as FOXA1 and XBP1.

EXPERIMENTAL PROCEDURES

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 GSEA, with the target gene set for a TF and drug-induced expression profiles as inputs, to make predictions about which drugs modulate the 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, the *MI*:

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

where

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

See the Supplemental Experimental Procedures for a more detailed description of the approach.

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 EMR, Kaplan-Meier survival analysis was performed on an age-adjusted cohort, using time to first diagnosis of prostate cancer as the endpoint in our study and excluding all patients with prior diagnosis of cancer. The Cox proportional hazards test was used to assess significance. See the Supplemental Experimental Procedures for a more detailed description of the EMR analysis. 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 statistically significance values were calculated in R. The permutation test (using 1,000 random permutations) was used to assess the significance of the enrichment of TF alterations in COSMIC and the enrichment of drug categories (e.g., HDAC inhibitors) within our predictions. The chi-square test was used to compare the TF 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-TF pairs that were predicted. The Mann-Whitney test was used to assign significance to the network analysis and to the difference between the concentration required to inhibit 50% of growth (GI_{50}) values in wild-type and the GI_{50} values in mutant p53 cell lines.

Experimental Validation

RWPE1, VCaP, and DU145 were obtained from ATCC and maintained according to the manufacturer's protocol. Isogenic DU145 or RWPE1 ± ERG cell lines were generated to overexpress truncated ERG, as previously described (Rickman et al., 2010, 2012). The $PTEN^{-/-}/ERG^{Rosa26}$ prostate cancer cells were derived from $PTEN^{-/-}/ERG^{Rosa26}$ prostate tumors, as previously described (Chen et al., 2013). The cells were treated with PBS and incubated with the appropriate media at the indicated drug or vehicle dose for 24 or 48 hr. Cells were then analyzed using ChIP-PCR, qRT-PCR, or invasion or migration assay. ChIP-PCR, qRT-PCR, cell invasion assay, and migration assay were performed as previously described (Rickman et al., 2010, 2012).

ChIP-PCR primers for all sites are listed in Table S3. Each sample was run in triplicate. The amounts of target genes were calculated relative to the reference gene HMBS. See the Supplemental Experimental Procedures for more details.

Supplemental Website

A supplemental website for CRAFTT has been made available at http://physiology.med.cornell.edu/faculty/elemento/lab/data/CRAFTT/. This includes an expanded overview of the methodology, with examples and a tool for querying predictions. We also have released our code on the website so that users can test our approach on their own drug perturbation profiles and/or TF target gene set of interest.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.05.037.

AUTHOR CONTRIBUTIONS

Conceptualization, K.G., D.R., and O.E.; Methodology, K.G., D.R., and O.E.; Investigation, E.D., C.C., J.W., Y.C., and D.R.; Formal Analysis – General, K.G., M.A.R., D.R., and O.E.; Formal Analysis – EMR, M.R.B., T.L., and N.P.T; Writing – Original Draft, K.G., M.A.R., D.R., and O.E.; Writing – Review & Editing, K.G. and O.E.; Funding Acquisition, D.R. and O.E.; Resources, D.R.; Supervision, D.R. and O.E.

ACKNOWLEDGMENTS

The authors would like to thank Lew Cantley and the O.E., D.R., N.P.T., and M.A.R. lab members for their feedback and discussions. This work was supported by a CAREER grant from National Science Foundation (DB1054964), the NIH under awards R01CA194547 and R01CA179100, and the Starr Cancer Foundation, as well as by startup funds from the Institute for Computational Biomedicine and the Department of Pathology, Weill Cornell Medical College. K.G. was also supported in part by the Tri-Institutional Training Program in Computational Biology and Medicine (via NIH training grant T32GM083937) and by the PhRMA Foundation Pre Doctoral Informatics Fellowship. T.L. and N.P.T. were supported by the National Institute of General Medical Sciences (NIGMS) (R01GM107145); T.L. was also supported by supported by the NIGMS (T32GM082797). M.R.B. was supported by the NLM (T15LM00707). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the NIH. O.E. is a consultant for Novartis, Roche, and Prolias Technologies.

Received: August 7, 2015
Revised: March 18, 2016
Accepted: May 7, 2016
Published: June 2, 2016

REFERENCES

Ablain, J., Nasr, R., Bazarbachi, A., and de Thé, 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.

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.

Conde, R., Belak, Z.R., Nair, M., O'Carroll, R.F., and Ovsenek, N. (2009). Modulation of Hsf1 activity by novobiocin and geldanamycin. Biochem. Cell Biol. *87*, 845–851.

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 Syst. Biol. *6*, 92.

Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastritis, 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.

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 Biol. *9*, e1001046.

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–i623.

Khurana, E., Fu, Y., Chen, J., and Gerstein, M. (2013). Interpretation of genomic variants using a unified biological network approach. PLoS Comput. Biol. *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 Rep. *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 Res. *39*, D1035–D1041.

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.

Libermann, T.A., and Zerbini, L.F. (2006). Targeting transcription factors for cancer gene therapy. Curr. Gene Ther. *6*, 17–33.

Mees, C., Nemunaitis, J., and Senzer, N. (2009). Transcription factors: their potential as targets for an individualized therapeutic approach to cancer. Cancer Gene Ther. *16*, 103–112.

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.

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 Res. *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. Proc. Natl. Acad. Sci. USA *109*, 9083–9088.

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., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

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.

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., and Green, M.R. (2012). A synthetic interaction screen identifies factors selectively required for proliferation and TERT transcription in p53-deficient human cancer cells. PLoS Genet. *8*, e1003151.