

A Patient-to-Model-to-Patient (PMP) Cancer Drug Discovery Protocol for Identifying and Validating Therapeutic Agents Targeting Tumor Regulatory Architecture

Pasquale Laise,^{1,2} Gideon Bosker,¹ Andrea Califano,^{2,3,4,5,6,7}
and Mariano J. Alvarez^{1,2,7}

¹DarwinHealth, New York, New York

²Department of Systems Biology, Columbia University Irving Medical Center, New York, New York

³Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, New York

⁴Department of Medicine, Columbia University Irving Medical Center, New York, New York

⁵Department of Biochemistry & Molecular Biophysics, Columbia University Irving Medical Center, New York, New York

⁶Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, New York

⁷Corresponding authors: ac2248@cumc.columbia.edu; malvarez@darwinhealth.com

Published in the Pharmacology section

The current Achilles heel of cancer drug discovery is the inability to forge precise and predictive connections among mechanistic drivers of the cancer cell state, therapeutically significant molecular targets, effective drugs, and responsive patient subgroups. Although advances in molecular biology have helped identify molecular markers and stratify patients into molecular subtypes, these associational strategies typically fail to provide a mechanistic rationale to identify cancer vulnerabilities. Recently, integrative systems biology methodologies have been used to reverse engineer cellular networks and identify master regulators (MRs), proteins whose activity is both necessary and sufficient to implement phenotypic states under physiological and pathological conditions, which are organized into highly interconnected regulatory modules called tumor checkpoints. Because of their functional relevance, MRs represent ideal pharmacological targets and biomarkers. Here, we present a six-step patient-to-model-to-patient protocol that employs computational and experimental methodologies to reconstruct and interrogate the regulatory logic of human cancer cells for identifying and therapeutically targeting the tumor checkpoint with novel as well as existing pharmacological agents. This protocol systematically identifies, from specific patient tumor samples, the MRs that comprise the tumor checkpoint. Then, it identifies *in vitro* and *in vivo* models that, by recapitulating the patient's tumor checkpoint, constitute the appropriate cell lines and xenografts to further elucidate the tissue context-specific drug mechanism of action (MOA) and permit precise, biomarker-based preclinical validations of drug efficacy. The combination of determination of a drug's context-specific MOA and precise identification of patients' tumor checkpoints provides a personalized, mechanism-based biomarker to enrich prospective clinical trials with patients likely to respond. © 2022 The Authors. Current Protocols published by Wiley Periodicals LLC.

Keywords: biomarkers • cancer • drug discovery • regulatory networks • systems pharmacology

Laise et al.

1 of 19

How to cite this article:

Laise, P., Bosker, G., Califano, A., & Alvarez, M. J. (2022). A patient-to-model-to-patient (PMP) cancer drug discovery protocol for identifying and validating therapeutic agents targeting tumor regulatory architecture. *Current Protocols*, 2, e544. doi: 10.1002/cpz1.544

INTRODUCTION

According to the oncotecture and omnigenic hypotheses, the phenotypic/transcriptional state of every cell is controlled homeostatically by a set of transcriptional regulatory proteins, termed master regulators (MRs), whose concerted activity is necessary and sufficient to maintain the transcriptional identity of a specific cell (for recent perspectives, please see Boyle, Li, & Pritchard, 2017; Califano & Alvarez, 2017). In this article, we adhere to the definitions of transcriptional state and transcriptional identity as introduced by Paull et al. (2021) (Table 1). In this biological framework, MRs are organized in highly inter-regulated modules, or transcriptional regulatory checkpoints, which operate as molecular switches, controlling the transcriptional identity of normal/physiologic (Arumugam et al., 2020; Dutta et al., 2016; Kushwaha et al., 2015; Lefebvre et al., 2010; Talos, Mitrofanova, Bergren, Califano, & Shen, 2017) as well as aberrant/pathologic (Aytes et al., 2013; Aytes et al., 2014; Bisikirska et al., 2016; Brichta

et al., 2015; Carro et al., 2010; Compagno et al., 2009; Della Gatta et al., 2012; Ikiz et al., 2015; Laise et al., 2020; Pan et al., 2020; Paull et al., 2021; Putcha et al., 2015; Rajbhandari et al., 2018; Rodriguez-Barrueco et al., 2015; Son et al., 2021; Sonabend et al., 2014; Walsh et al., 2017) cell states. In the case of neoplastic malignancies, the aberrant protein activity levels and composition of such MR protein modules, or tumor checkpoints, are responsible for the establishment and maintenance of the tumor cell transcriptional identity (Califano & Alvarez, 2017; Paull et al., 2021).

MR proteins are typically transcriptional regulators, such as transcription factors (TFs) or transcriptional co-factors (coTFs), that exert their regulatory activity by mechanistically regulating the expression of their target genes. Consequently, the regulatory activity of a protein is mirrored by changes in the expression of its positive and negative target genes, collectively referred to as the “regulon.” In the case of MRs, the target genes are enriched in

Table 1 Glossary

Term	Definition
Interactome	Tissue lineage-specific collection of regulons indicating the set of transcriptional targets for each transcriptional regulator, as well as whether the transcription of each target is activated or repressed by each regulator.
Master regulator (MR)	Transcriptional regulatory protein whose activity is necessary and sufficient to maintain the transcriptional identity of a specific cell.
Regulatory checkpoint	Highly inter-regulated modules of MRs that operate as a molecular switch, controlling the transcriptional identity of a specific cell.
Transcriptional state	Gene expression vector describing the position of a cell in the N-dimensional space representing all the possible implementations of the cell transcriptome. Transcriptional states can be highly transient.
Transcriptional identity	High-probability-density regions in the transcriptional state space where cells persist over a long period of time and are characterized by similar phenotypic properties. Tumor subtypes, in particular, represent relevant tumor cell transcriptional identity implementations in the context of cancer.
Tumor Checkpoint Reporter Set (TCRS)	Set of 50 transcriptional regulators most differentially active—25 most activated and 25 most inactivated—in a particular tumor when compared against the “universal” tumor reference. The TCRS is used as a reporter for tumor checkpoint activity.

Table 2 Software and Software Packages

Software name	Use	Availability
ARACNe	Reverse engineering of transcriptional regulatory networks	Califano Lab ^a
aracne-networks	Collection of 25 lineage context-specific regulatory networks (interactomes)	Bioconductor ^b
DESeq2	Normalization of RNA-Seq data	Bioconductor ^c
VIPER	Inference of protein activity, aREA and viperSimilarity (reciprocal enrichment) algorithms	Bioconductor ^d

^a <https://califano.c2b2.columbia.edu/aracne>

^b <https://www.bioconductor.org/packages/release/data/experiment/html/aracne.networks.html>

^c <https://bioconductor.org/packages/release/bioc/html/DESeq2.html>

^d <https://www.bioconductor.org/packages/release/bioc/html/viper.html>

the gene expression signature characterizing the transcriptional identity of the cell. These mechanistic relationships provide a rationale for *inferring* the regulatory activity of a protein (gene product), hereafter referred to as “protein activity” for simplicity, from gene expression data through a systematic assessment of its regulon. Importantly, activity of a protein’s regulon can be quantitatively assessed by gene set enrichment analysis (GSEA) (Subramanian et al., 2005) or its derivatives, designed specifically to account for positive (induction) and negative (repression) transcriptional regulation, such as the analytical rank-based enrichment analysis (aREA) (Alvarez et al., 2016). The aREA algorithm is available from the VIPER R-system package (Table 2).

Coupled with the ability to infer context-specific target genes of a given protein by reverse engineering transcriptional interactions from genome-wide expression data (Califano & Alvarez, 2017; Carro et al., 2010; Margolin et al., 2006), this analytical framework represents the foundation of the Master Regulator Inference algorithm (MARINa) and its extension at the single-sample level, the algorithm for Virtual Inference of Protein activity by Enriched Regulons (VIPER) (Alvarez et al., 2016; Carro et al., 2010). Assessing MR protein activity level, using either the MARINa or the VIPER algorithm, requires two elements: a context-specific model of transcriptional regulation and a differential gene expression signature.

The context-specific model of transcriptional regulation is a network representing the candidate MRs, their regulons, and their regulatory relations. These networks can be inferred by reverse engineering transcriptional interactions from high-throughput gene expression data (such as microarray, RNA-Seq,

or scRNA-Seq) using dedicated algorithms, such as the Algorithm for the Reconstruction of Accurate Cellular Network (ARACNe) (Basso et al., 2005; Lachmann, Giorgi, Lopez, & Califano, 2016; Margolin et al., 2006). The context specificity of proteins’ regulons is critical for an accurate inference of protein activity, as the same protein can potentially regulate a different set of genes in different cell types (Califano, Butte, Friend, Ideker, & Schadt, 2012). The regulatory relation between a candidate MR protein and its target genes is called the mode of regulation (MoR) and is explicit whether a target gene is positively or negatively regulated by the MR protein (Alvarez et al., 2016). The gene expression signature is the full list of genes, as assessed by high-throughput technologies (e.g., microarray, RNA-Seq, scRNA-Seq), ranked according to their differential expression computed between two or more biological/experimental conditions.

Once the model of transcriptional regulation is computed, this can be used to infer the activity of each candidate MR from any lineage context-matched gene expression signature. Indeed, the regulatory activity of a given protein can be efficiently estimated by performing a two-tailed enrichment analysis of its regulon in a gene expression signature, ensuring preservation of the MoR and thereby allowing an estimation of whether there is an over-activation (positive enrichment) or an inactivation (negative enrichment) of the protein (Alvarez et al., 2016). The stronger the enrichment of the regulon of a given protein in a gene expression signature, the higher its activity is. The top-ranked proteins based on their inferred activity are typically enriched for MRs (Califano & Alvarez, 2017).

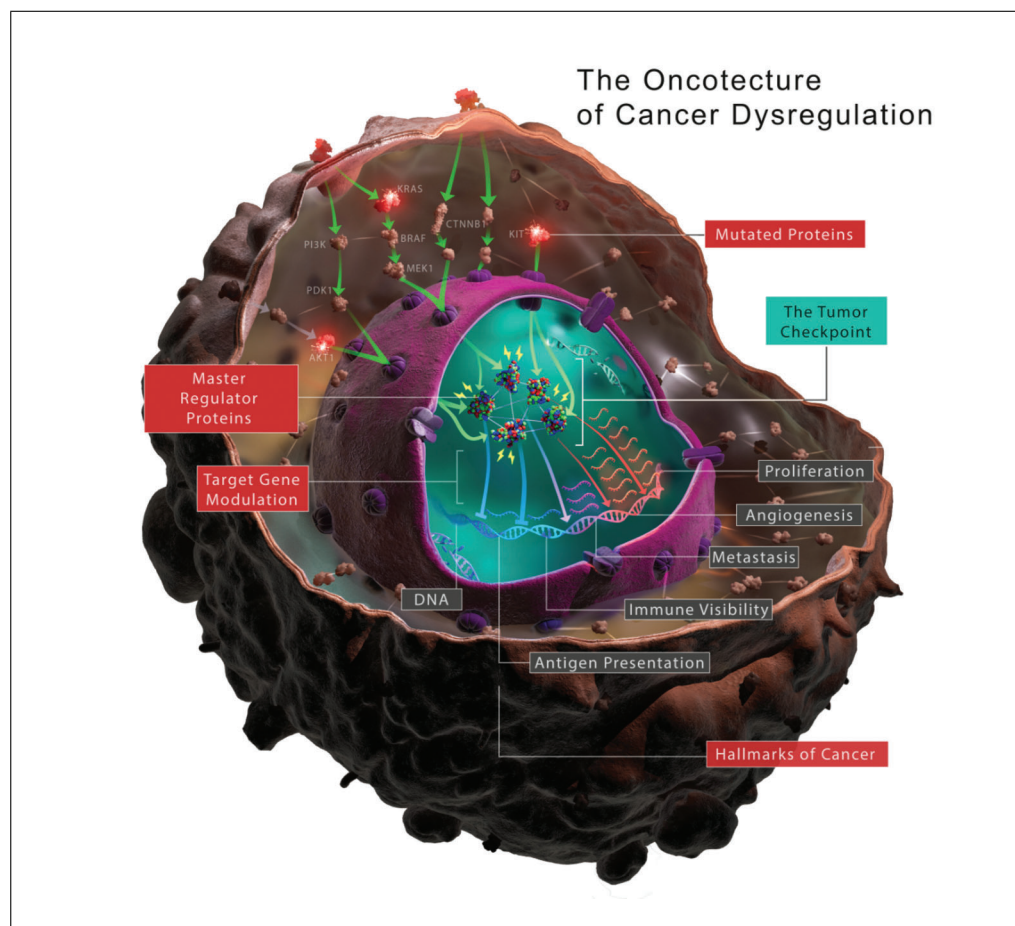


Figure 1 Oncotecture model. Diagram showing a highly inter-regulated module of five MR proteins, or a tumor checkpoint, which canalizes the effect of extracellular and intracellular signals on the cell transcriptome, directing the coordinated expression of different hallmarks of cancer genes and maintaining the transcriptional identity of the cancer cell.

The biological relevance of MR proteins in regulating cellular transcriptional identity, in both physiological and pathologic contexts, has been confirmed in multiple studies (see Arumugam et al., 2020; Dutta et al., 2016; Kushwaha et al., 2015; Lefebvre et al., 2010; and Talos et al., 2017 and Aytes et al., 2013; Aytes et al., 2014; Bisikirska et al., 2016; Brichta et al., 2015; Carro et al., 2010; Compagno et al., 2009; Della Gatta et al., 2012; Ikiz et al., 2015; Laise et al., 2020; Pan et al., 2020; Paull et al., 2021; Putcha et al., 2015; Rajbhandari et al., 2018; Rodriguez-Barrueco et al., 2015; Son et al., 2021; Sonabend et al., 2014; Walsh et al., 2017, respectively). We have recently shown that MR proteins are responsible for canalizing the genetics of individual tumor samples, effectively bridging the gap between the large space of potential genetic alterations and the reduced set of potential tumor cell transcriptional identities (Paull et al., 2021). Moreover, the organization of MR proteins into auto-regulated modules (tumor checkpoints or MR blocks) has

been shown to be mechanistically associated with the implementation of such cancer hallmarks as immune evasion, metastases, cell migration, and angiogenesis (Paull et al., 2021) (Fig. 1). Given the limited number of stable, aberrant tumor cell transcriptional identities, it has been postulated that MR proteins constitute a new class of non-oncogene-based tumor cell dependencies (Califano & Alvarez, 2017), and therefore, precision-based targeting of the tumor checkpoint, evaluated in a tissue context-specific manner, may represent a more effective and specific approach for identifying pharmacologic treatments for cancer directed at the aberrant regulatory architecture of cancer cells (Mundi et al., 2021).

Tumor-associated MR proteins have been systematically identified as fundamental dependencies for a diverse set of tumor types, among them hematopoietic malignancies (Bisikirska et al., 2016; Compagno et al., 2009; Della Gatta et al., 2012), glioblastoma (GBM) (Carro et al., 2010; Sonabend et al., 2014), breast carcinoma (Putcha et al., 2015;

Rodriguez-Barrueco et al., 2015; Walsh et al., 2017), gastroenteropancreatic neuroendocrine tumors (Alvarez et al., 2018), neuroblastoma (Rajbhandari et al., 2018), pancreas (Laise et al., 2020), and prostate carcinoma (Aytes et al., 2013; Aytes et al., 2014), among others (Paull et al., 2021). Moreover, by mechanistically regulating the aberrant transcriptional state of tumors, MR proteins operate downstream, canalizing the effect of most genetic alterations. As such, they can be viewed as a regulatory bottleneck whose targeting cannot be bypassed by additional alterations in upstream pathways (Califano & Alvarez, 2017) (Fig. 1). In contrast to classic approaches for biomarker discovery, based on empirical/statistical associations between features and outcomes, the analysis of regulatory networks provides a mechanistic rationale for the role of MR proteins in implementing and maintaining the tumor cell state and as such represents a method for identifying ideal candidate therapeutic targets and biomarkers.

THE PATIENT-TO-MODEL-TO-PATIENT PROTOCOL

In this article, we describe the patient-to-model-to-patient (PMP) protocol for tumor checkpoint-based drug discovery developed at DarwinHealth, based on algorithms and computational tools developed in the Califano Lab at Columbia University and licensed exclusively for commercial use and cancer drug discovery to DarwinHealth. The first step of the PMP protocol relies on the systematic analysis of gene expression profiles—RNA sequencing—of tumor samples from cancer **Patients** and the identification of MRs constituting the tumor checkpoint. Based on the conservation of the patient's tumor checkpoint—which consists of a signature characterized by MR protein activity levels—the PMP protocol identifies the most appropriate cell line and patient-derived xenograft (PDX) **Models**. The selected cell line models are then used for drugs' context-specific mechanism of action (MOA) elucidation and drugs' prioritization based on their ability to target the tumor checkpoint, whereas the PDX models are used for preclinical validation of drug efficacy and *in vivo* MOA. Specifically, for the preclinical validation, the PMP protocol leverages the tumor checkpoint as a mechanism-based biomarker, which, in concert with a drug's context-specific MOA, is used to predict drug response—including tumor growth (efficacy) and pharmacodynamic studies—in the context of genetically

engineered or PDX models. Finally, the PMP protocol returns to cancer **Patients**, leveraging the context-specific drug MOA/tumor checkpoint-based biomarkers to enrich clinical trial cohorts with patients most likely to show clinical response (tumor suppression) to the predicted pharmacologic intervention (see Fig. 2 for a schematic representation).

Part 1 of the PMP Protocol: Systematic Identification of Patients' Master Regulator Proteins and Tumor Checkpoints

A key aspect of the PMP protocol is the identification of non-oncogene-based tumor dependencies directly from *patients'* tumor samples, thereby avoiding the potential biased and idiosyncratic results commonly associated with the use of animal or cell culture-based tumor models (Fig. 2A).

MR elucidation in specific, patient-derived tumors requires gene expression profiling of fresh-frozen (recommended) or formalin-fixed tissue from tumor samples obtained from biopsies or tumor tissue following tumor resection/excision. Typically, such gene expression profiles require a minimum sequencing depth of 5 million reads aligned to the reference transcriptome for samples with a tumor cell content >60% (Alvarez et al., 2018). Satisfying these conditions is required because potential contamination of the expression profile with normal cells embedded in the tumor stroma dilutes the signal, thereby limiting our ability to characterize the transcriptional identity of the tumor cells specifically and therefore its MRs. We have previously shown that a sequencing depth of 5 million reads is adequate for protein activity quantification using the VIPER algorithm (see below). Moreover, it has been shown that VIPER-based results quantifying protein activity levels are extremely robust despite limited sequencing depth, producing virtually identical results once the sequence depth is higher than 2 million aligned reads, with Pearson's correlation ($R > 0.98$) when comparing protein activity signatures inferred at depths of 2 and 30 million reads (Alvarez et al., 2016).

The gene expression data (raw counts) are then normalized using a variance stabilization transformation—as implemented in the DESeq2 package, available from Bioconductor (Love, Huber, & Anders, 2014) (Table 2)—or by log-transforming the counts per million (CPM) or fragments per kilobase per million (FPKM); we typically use $\log_2(\text{CPM}+1)$. To account for differences in the

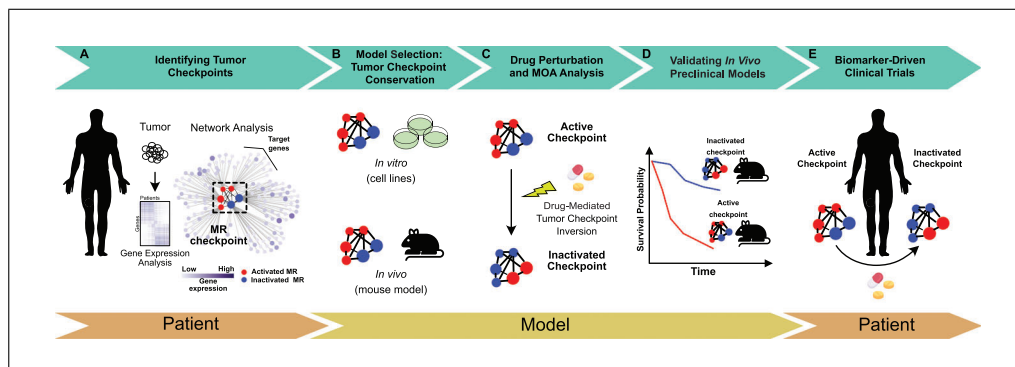


Figure 2 Diagram summarizing the PMP protocol. **(A)** Tumor checkpoints are identified from a patient's cohort archival tumor gene expression profiles using regulatory network analysis (part 1). **(B)** Cognate, tissue lineage-matched *in vitro* and *in vivo* models are identified based on the conservation of the patient's cohort tumor checkpoints (part 2). **(C)** Drug context-specific MOA is elucidated by network-based analysis of genome-wide expression profiles from drug-perturbed *in vitro* models. Drugs are prioritized based on their ability to target the tumor checkpoint, i.e., to invert the tumor checkpoint activity pattern (part 3). **(D)** Cognate *in vivo* models are used for preclinical validation to test the drug's ability to invert the tumor checkpoint activity pattern *in vivo* and its efficacy in reducing tumor growth (parts 4 and 5). **(E)** The tumor checkpoint elucidated for each individual patient's tumor is used, in conjunction with the drug context-specific MOA as a predictive biomarker of response to enrich prospective clinical trials with responder patients (part 6).

distribution of transcript abundance among genes, expression signatures are computed using a large dataset as a reference. The reference dataset we employ comprises a compendium of more than 10,000 tumor samples that have been profiled by The Cancer Genome Atlas (TCGA) (Mundi et al., 2021). Although the simplest approach is to subtract the mean and to scale by the standard deviation of each gene as estimated from the reference set, the use of a regularized reference set subsampling based on tumor subtypes or the results of fitting a mixture of Gaussian models to the transcript abundance distribution of each gene may help account for the unbalanced representation of different tumor subtypes in the reference set.

Because the critical regulatory architecture of tumor cells that we focus on is responsible for their transcriptional identity, the activity of such transcriptional regulatory proteins—typically TFs and their co-factors (coTFs), including chromatin-remodeling enzymes and proteins with transcriptional regulatory functions that do not directly bind DNA in a sequence-dependent fashion—is quantified by the VIPER algorithm (Alvarez et al., 2016). As aforementioned, VIPER analysis requires both (a) a gene expression signature and (b) a model of transcriptional regulation (regulon), concretizing the effect(s) of a given regulatory protein, either an inducer or a transcriptional repressor, over a set of its transcriptional target genes. Using each regulon as a probabilistic model of transcriptional regulation,

VIPER quantifies the coordinated enrichment of the transcriptionally induced and repressed targets of each regulatory protein according to the regulon model—i.e., akin to a highly multiplexed gene reporter assay—on the transcripts that are upregulated and downregulated, respectively, in the gene expression signature being analyzed (Alvarez et al., 2016). Given that chromatin accessibility, as well as TF and coTF availability is exquisitely tissue lineage context specific (Califano & Alvarez, 2017), regulons must be inferred in a tissue context-specific way. For this, we rely on the widely adopted ARACNe algorithm (Basso et al., 2005; Khatamian, Paull, Califano, & Yu, 2019; Lachmann et al., 2016; Margolin et al., 2006) developed in the Califano Lab at Columbia University and licensed for commercial applications to DarwinHealth.

ARACNe requires a large cohort ($n > 100$) of independent samples that, although belonging to the same lineage context, maximize the dynamic range of expression for the transcriptional regulators (Basso et al., 2005). Such a requirement is usually satisfied by the genetic and epigenetic heterogeneity represented in large tumor type-specific cohorts (Paull et al., 2021) but can be also obtained for non-transformed cells by using short-term perturbations with biological agents and small molecules (Aytes et al., 2014). ARACNe is freely available for academic research use from Columbia University (Table 2), and a collection of 25 tissue lineage context-specific interactomes is

Table 3 Interactomes Available from *aracne.networks* Package for R

Type ^a	Regulators			Targets	Interactions
	TF	coTF	Signaling		
Acute myeloid leukemia	1769	650	3588	19,269	531,535
Bladder urothelial carcinoma	1793	656	3605	19,785	489,101
Breast carcinoma	1793	656	3605	19,359	331,919
Cervical squamous cell carcinoma	1793	656	3607	19,839	583,961
Colon adenocarcinoma	1793	656	3607	19,820	413,789
Esophageal carcinoma	1760	640	3551	18,679	529,286
Glioblastoma	1793	656	3607	19,858	563,850
Head and neck squamous cell carcinoma	1793	656	3606	19,772	423,104
Kidney renal clear cell carcinoma	1793	656	3605	19,843	350,478
Kidney renal papillary cell carcinoma	1793	656	3606	19,858	452,653
Liver hepatocellular carcinoma	1793	656	3607	19,829	469,922
Lung adenocarcinoma	1793	656	3606	19,742	399,513
Lung squamous cell carcinoma	1793	656	3605	19,741	455,032
Neuroendocrine tumor	1818	665	3646	22,932	666,241
Ovarian carcinoma	1769	650	3588	19,140	647,358
Pancreatic adenocarcinoma	1793	656	3607	19,858	520,756
Pheochromocytoma and paraganglioma	1793	656	3607	19,861	603,617
Prostate adenocarcinoma	1793	656	3604	19,820	330,922
Rectum adenocarcinoma	1793	656	3607	19,856	557,911
Sarcoma	1812	665	3635	20,479	526,591
Stomach adenocarcinoma	1799	655	3602	21,663	561,858
Testicular germ cell tumor	1793	656	3607	19,860	432,621
Thymoma	1793	656	3607	19,862	387,923
Thyroid carcinoma	1792	656	3605	19,861	317,582
Uterine corpus endometrial carcinoma	1793	656	3606	19,716	469,845

^a Indicated are the tumor type, number of regulatory proteins (TFs, coTFs, and signaling proteins), number of target genes, and total number of regulatory interactions represented in each interactome.

publicly available as an R-system package (Tables 2 and 3).

Importantly, both ARACNe and VIPER have been extensively validated, demonstrating high accuracy, with validation rates exceeding 70% (Alvarez et al., 2016; Basso et al., 2005; Carro et al., 2010; Lefebvre et al., 2010). However, other approaches can be used instead, including DoRothEA (Garcia-Alonso et al., 2018) and ChEA3 (Keenan et al., 2019). An alternative approach, based on ARACNe regulons and VIPER-inferred protein activity but using a set of 488 actionable proteins—proteins that can be targeted by Food and Drug Administration (FDA)-approved drugs and

pharmacologic inhibitors in late-stage (phase 2 or 3) clinical development—also has been developed to predict drugs and investigational compounds targeting proteins aberrantly activated by tumors. This method, called DarwinOncoTarget, has shown value at predicting response in the preclinical (Mundi et al., 2021; Obradovic, Tomassoni, et al., 2022) and clinical (Marks et al., 2021; Zeleke et al., 2020) settings. For instance, DarwinOncoTarget quantification of HDAC6 activity has been found to be highly predictive of the response of breast carcinoma patients to the HDAC6 inhibitor ricolinostat, identifying the ricolinostat responder patients with perfect accuracy and

Laise et al.

7 of 19

sensitivity (Zelege et al., 2020). Moreover, DarwinOncoTarget identified XPO-1 as an aberrantly activated protein and vulnerability for rhabdoid tumors (Marks et al., 2021), and it has been successful at identifying targeted inhibitors for cholangiocarcinoma (Obradovic, Tomassoni, et al., 2022) and at predicting the response for 16 approved drugs and investigational compounds identified by analyzing clinical samples and tested in patient-derived xenografts for triple-negative breast carcinoma, pancreatic ductal carcinoma, colon adenocarcinoma, gastrointestinal stromal tumors, and meningioma (Mundi et al., 2021). DarwinOncoTarget has been certified by the New York State Clinical Laboratory Improvement Amendments (CLIA) and is available for clinical applications from the Personalized Genomic Medicine Laboratory, Department of Pathology, Columbia University (<https://www.pathology.columbia.edu/diagnostic-specialties/personalized-genomic-medicine/oncology-testing/darwin-oncotarget-tm-oncotreat>).

Finally, the set of the 50 most differentially active proteins—the 25 most activated and the 25 most inactivated—identified in part 1 of the PMP protocol is used as a reporter for tumor checkpoint activity. This is based on the observation that the top 20 to 30 most differentially active transcriptional regulators are able to canalize/account for virtually all functional genetic alterations in any of 9700 tumor samples from 20 TCGA cohorts analyzed (Paull et al., 2021). We will refer to this set of 50 proteins as the Tumor Checkpoint Reporter Set (TCRS).

Part 2 of the PMP Protocol: Selection of Cell Line Models Appropriate for Drug Context-Specific Mechanism-of-Action Elucidation

The selection of appropriate cell line models to assess drug MOA is strictly based on whether the *in vitro* models constitute a good reporter for the activity of pharmacologic perturbagens on the activity of MR proteins identified as the TCRS in part 1 of the protocol (Fig. 2B). Whether the cell line model recapitulates biological features of the modeled disease, including particular genetic alterations or *in vitro* sensitivity to pharmacologic perturbations, is, in principle, irrelevant, with the exception of mutations that could potentially affect, directly or indirectly, the response to the drug. This last criterion, however, is very difficult to assess because the same cell line models are used to profile a library of

drugs and compounds for which only limited knowledge about their primary targets is available. To compensate for potentially cell line–idiosyncratic responses, it is recommended to use several different (3 to 6) cell line models per tissue lineage/tumor type of interest to generate a consensus MOA for a specific drug in a tumor subtype–specific context.

Because the readout for PMP protocol part 4 (DarwinOncoTreat analysis, see below) is based on *inverting* the tumor checkpoint activity pattern—i.e., inactivating the aberrantly active MR proteins and activating the aberrantly inactive MRs for the purpose of abrogating the regulatory dependencies of tumor cells—conservation of such an activity pattern in the cell line maximizes its sensitivity for reporting inactivation of the active MR proteins and activation of the inactive MR proteins. Thus, prioritization of the cell lines is performed based on confirming enrichment of the TCRS on the cell line protein activity signature (Alvarez et al., 2019). Such enrichment can be quantified by the aREA (Alvarez et al., 2016) or GSEA (Subramanian et al., 2005) algorithm. Moreover, to maximize the chances of having a conserved signaling and transcriptional regulatory architecture between patient-derived tumors and cell line models, only cell lines of lineage-related origin should be considered.

The protein activity signatures for the cell line models are computed in a way that is analogous to the tumor protein activity signatures: the normalized gene expression is compared against a large collection of gene expression profiles that are used as a reference to account for differences between gene expression distributions, and VIPER is then used to quantify protein activity based on tissue lineage context–specific regulons. The reference used for the cell line models, however, is not a large compendium of tumors, given that systematic bias in gene expression due to *in vitro* growth and selection pressures would strongly affect such gene expression signatures. A more appropriate reference set for the cell line models is a large collection of cell line gene expression profiles, such as the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012). Figure 3 shows the conservation of the tumor checkpoint for 1045 prostate carcinoma tumors by lineage-matched cell line models.

We have shown that cell line models selected in this way are faithful reporters of drug MOA, showing the highest conservation of drug-induced protein activity signatures when compared to patient-derived tumor cell

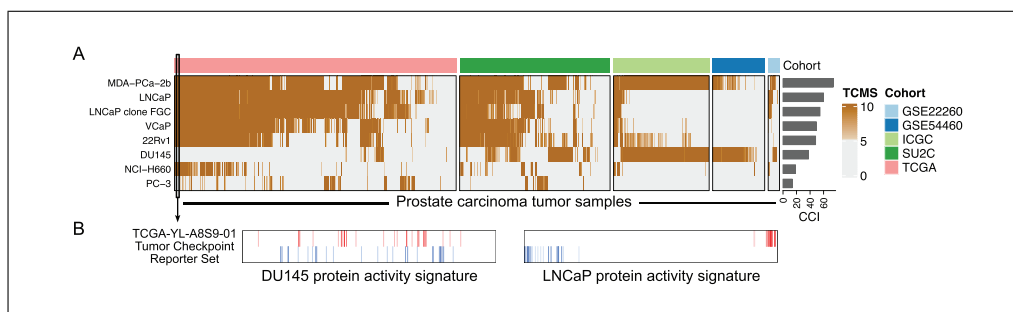


Figure 3 DarwinOncoMatch-based alignment of lineage-matched cell lines to prostate carcinoma tumors. **(A)** Heatmap showing the conservation of the Tumor Checkpoint Reporter Set (TCRS) for each of 1045 tumors, from five different prostate carcinoma cohorts, in eight prostate carcinoma cell lines. The conservation was quantified by the enrichment of the TCRS on the cell lines' protein activity signature and is expressed by the Tumor Checkpoint Matching Score (TCMS) as $-\log_{10}(p\text{-value})$, Bonferroni's corrected. The bar graph to the right of the heatmap shows the Checkpoint Coverage Index (CCI), indicating the proportion (%) of the analyzed tumors whose checkpoint is significantly conserved in the cell lines ($p < 10^{-5}$, $\text{TCMS} > 5$). **(B)** Example enrichment plots for the TCRS of one tumor sample (TCGA-YL-A8S9-01) on the protein activity signatures of the DU145 and LNCaP cell lines. The x-axis represents 2448 transcriptional regulatory proteins, sorted from the one showing the lowest relative activity (left) to the one showing the highest relative activity (right) for the specified cell line. The vertical lines indicate the top 25 (red) and bottom 25 (blue) proteins most activated and inactivated, respectively, that are part of the TCRS for sample TCGA-YL-A8S9-01. The plots show strong enrichment of the TCRS in the LNCaP, but not the DU145, protein activity signatures.

primary cultures and tumor explants (Alvarez et al., 2019).

Part 3 of the PMP Protocol: Elucidation of Context-Specific Drug Mechanism of Action

Context-specific drug MOA is elucidated by perturbing appropriate cell line models (see PMP protocol part 2, above) with a library of drugs and/or compounds in order to characterize the impact of the drug perturbations on the activity of MRs consisting of, as previously noted, transcriptional regulatory proteins (Fig. 2C). This analysis requires full transcriptome expression profiling followed by VIPER analysis to provide a precise MR protein activity signature (Shen et al., 2017).

To ensure that drug-mediated effects on protein activity levels are optimally characterized while also maximizing the value of the transcriptome profiles in cell lines as a faithful reporter for the specific effect of the drug on the activity of transcriptional regulatory proteins, drug perturbations are performed at their highest possible sublethal concentrations. These "equipotent" concentrations are defined by dose-response analysis using cell viability as a readout. Typically, 10- to 12-point titration curves are performed while quantifying viable cells at 48 to 96 hr and estimating the effective dose 20% (ED_{20}), i.e., the concentration leading to a loss of 20% viability. Then, a set of serial dilutions is defined

for the perturbational assays, with concentrations ranging from the estimated ED_{20} to 1 to 3 additional dilutions using a factor of 10 or 3, respectively, e.g., ED_{20} and $1/10 \text{ED}_{20}$ or ED_{20} and $1/3$, $1/9$, and $1/27 \text{ED}_{20}$.

Gene expression profiles after drug perturbation are obtained at 24 hr to evaluate the short-term effect of the perturbation on the activity of the MR proteins and at longer time points (48 to 96 hr) to corroborate a reproducible and sustained MOA not affected by cell adaptive/compensatory responses. Such corroboration of an enduring MOA at longer time points is essential for identifying drugs/compounds of developmental interest and prioritizing agents with potential therapeutic application in pre-clinical and clinical research. Although different protocols can be used to obtain the transcriptome-wide profile of perturbed cells, we typically use the low-cost, highly multiplexed Pooled Library Amplification for Transcriptome Expression (PLATE-Seq) protocol (Bush et al., 2017). The PLATE-Seq sequencing depth of 500,000 to 2 million mapped reads per well is sufficient for the accurate estimation of protein activity by VIPER, as we have previously discussed (Alvarez et al., 2016).

The Pan-cancer Analysis of Chemical Entity Activity (PanACEA) resource database provides genome-wide RNA-Seq profiles for 25 cell lines at 24 hr after perturbation with

more than 400 FDA-approved drugs and investigational compounds, which have been profiled at equipotent concentrations (ED₂₀) (Douglass et al., 2022). Expression profiles for part of the PanACEA database, including 32 kinase inhibitors profiled in 11 cell lines, are available from the Gene Expression Omnibus (GSE186341).

After data normalization, which is typically log-transformed CPM, gene expression signatures are computed by comparing each perturbed condition against its corresponding vehicle control, usually dimethyl sulfoxide (DMSO) or ethanol for small-molecule compounds and saline for antibodies and most biologicals. VIPER is then used, together with tissue lineage-matched regulons, to transform the drug perturbation-induced gene expression signatures into the drug perturbation-induced protein activity signatures we refer to as the context-specific drug MOA. As such, the context-specific MOA is a quantitative description of the tissue lineage context-specific effect the drug perturbation has on the activity of transcriptional regulatory proteins.

Part 4 of the PMP Protocol: Prioritization of Drugs Predicted to Invert the Activity of the Tumor Checkpoint

As described previously, the regulatory checkpoints, consisting of a set of MR proteins elucidated from gene expression profiles, operate as molecular switches, controlling the transcriptional identity of every cell. Due to their auto-regulation, tumor checkpoints can exist in two different, functionally opposed activation states. In its tumor-initiating and/or tumor-maintaining state, the checkpoint exerts transcriptional governance over a cell state characterized by expression of biological hallmarks that promote malignant behaviors, among them immune evasion, angiogenesis, metastases, and apoptosis (Paull et al., 2021). In the context of cancer drug discovery, if pharmacologically mediated perturbations—which may include a spectrum of both on- and off-target effects—invert the protein activity pattern that underpins the tumor cells' regulatory architecture, specifically by inactivating the over-active MRs and activating the under-active MRs, the drug-induced inversion of the regulatory checkpoint activity pattern will drive the cell into a regulatory state incompatible with the original transcriptional identity/regulatory program dependency re-

quired for maintenance of tumor cell state, thereby inducing tumor cell death (Califano & Alvarez, 2017). We have used this concept to develop the New York State CLIA-certified DarwinOncoTreat algorithm, which prioritizes drugs based on their predicted capacity to invert the activity pattern of an individual patient's tumor checkpoints (Alvarez et al., 2018). In addition, this approach has been extended to target checkpoints associated with specific biological hallmarks, such as cell migration (Paull et al., 2021), and even to block the phenotypic transition permissive of viral replication induced in human host cells by viral infections (Laise et al., 2022).

As part 4 of the PMP protocol, prioritization of drugs predicted to invert the pattern of activity of the tumor checkpoint is accomplished by deploying the DarwinOncoTreat algorithm, which prioritizes the anti-neoplastic, therapeutic potential of drugs by quantifying the inverse enrichment of a particular patient's MR proteins—using the set of the 50 most differentially active proteins in the patient as a reporter for the tumor checkpoint activity—on the drug-induced protein activity signatures (their context-specific MOA) obtained in appropriate cell line models, as described in part 3 of the PMP protocol (Fig. 2C). This is performed by quantifying the enrichment of the TCRS, i.e., the top 25 most activated and the top 25 most inactivated regulatory proteins in the patient's tumor, among the proteins inactivated and activated, respectively, in response to the drug perturbation in the cell line model. Enrichment of the TCRS on each drug-induced protein activity signature (drug context-specific MOA) can be quantified by the aREA (Alvarez et al., 2016) or GSEA (Subramanian et al., 2005) algorithm.

Importantly, the DarwinOncoTreat analysis can be performed retrospectively on a large cohort of clinically relevant tumors to estimate the proportion of patients predicted to respond to a panel of pharmacologic compounds whose context-specific MOA has been elucidated in cell line models. The proportion of patients predicted to respond is expressed as the Cohort Coverage Index (CCI), a metric that is used to prioritize those drugs with a higher likelihood of benefitting the greatest number of patients, based on the comparative propensity of drugs to invert tumor checkpoints in the patient cohort reference set (Mundi et al., 2021) (Fig. 4).

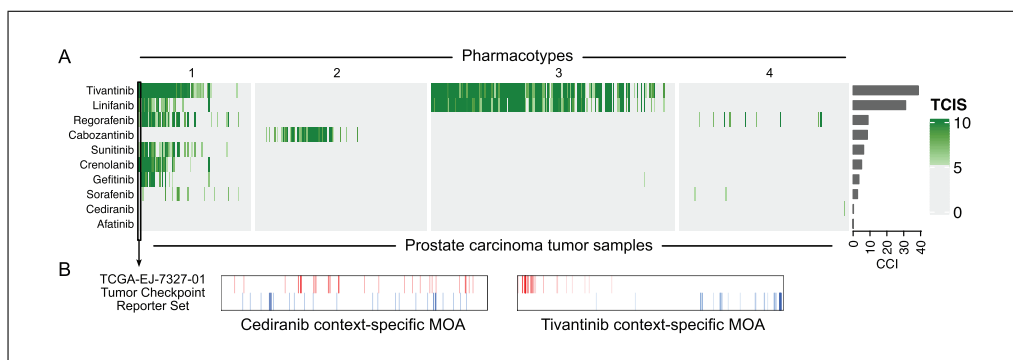


Figure 4 DarwinOncoTreat-based alignment of a set of kinase inhibitors to prostate carcinoma tumors. **(A)** Heatmap showing the inversion strength for the Tumor Checkpoint Reporter Set (TCRS), for each of 498 tumor samples represented in the TCGA prostate carcinoma cohort, induced by 10 kinase inhibitors when used to perturb LNCaP cells. The TCRS inversion strength was quantified by the negative enrichment of the TCRS on the drug-induced protein activity signature (drug context-specific MOA), and it is expressed by the Tumor Checkpoint Inversion Score (TCIS) as $-\log_{10}(p\text{-value})$, Bonferroni's corrected. The bar graph to the right of the heatmap shows the Checkpoint Coverage Index (CCI), indicating the proportion (%) of the analyzed tumors whose checkpoint is predicted by DarwinOncoTreat to be significantly inverted ($p < 10^{-5}$, TCIS > 5). **(B)** Example enrichment plots for the TCRS of one tumor sample (TCGA-EJ-7327-01) on the protein activity signatures induced when perturbing LNCaP cells with cediranib and tivantinib. The x-axis represents 2448 transcriptional regulatory proteins, sorted from the one most inactivated (left) to the one most activated (right) in response to the specified drug perturbation. The vertical lines indicate the top 25 (red) and bottom 25 (blue) proteins most activated and inactivated, respectively, that are part of the TCRS for sample TCGA-EJ-7327-01. The plots show strong inversion of the TCRS induced by tivantinib, but not by cediranib.

Part 5 of the PMP Protocol: Therapeutic and Mechanism-of-Action Validation of Pharmacologically Mediated Tumor Checkpoint Inversion and Assessment of Drug Efficacy *In Vivo*

Given the potential differences in drug pharmacodynamics between *in vitro* models and *in vivo* conditions and the potential drug pharmacokinetic issues inherent to *in vivo* conditions, a drug's context-specific MOA inferred from cell line models might not always be conserved *in vivo*. To address these potential discrepancies and to ensure that the experimental results and analyses generated by the PMP protocol are optimally applicable to the needs of preclinical and clinical cancer drug discovery and validation, drug pharmacodynamics, pharmacokinetics, and anti-tumor effects must be assessed within the context of *in vivo* models, most commonly by using genetically engineered mouse models (GEMMs) (Vasciaveo et al., 2022) or PDX models (Mundi et al., 2021; Vasciaveo et al., 2022). We will focus the discussion here on pharmacodynamics studies aimed at (a) elucidating the drug's context-specific MOA and (b) validating the inversion of the tumor checkpoint activity pattern *in vivo*, two criteria that must be evaluated to predict drug efficacy

and translational relevance in the clinical setting.

High-fidelity (cognate) *in vivo* tumor models for the study of drug MOA and drug response can be identified by assessing the conservation of tumor checkpoints from patient tumor samples (see part 1 of the PMP protocol, above) in the corresponding animal models (Alvarez et al., 2019; Vasciaveo et al., 2022) (Fig. 2B). Transcriptome profiles for *in vivo* tumor models are typically obtained by isolating RNA from a fresh-frozen biopsy or resected tumor sample and then performing RNA-Seq at a depth of at least 10 million reads aligning to the mouse genome (for GEMMs) or to the human genome (for PDX models). For PDX models, it is recommended to ignore reads corresponding to non-tumor murine cells, which can be identified by their alignment to the mouse reference genome. Gene expression signatures are then computed, and, to account for different distributions of transcript abundance between genes, *in vivo* model profiles are compared against a large compendium of tumor transcriptomes used as a reference. This reference must be the same as the one used to infer the protein activity signatures for the patient's tumors. For this, we typically use a compendium of more than 10,000 tumor samples that

have been profiled by TCGA (Mundi et al., 2021).

Importantly, we perform a tumor checkpoint “fidelity check” for all PDX models and GEMMs. This confirmation is designed to detect potential drifts in the transcriptional identity of the implanted cells while establishing and growing the tumor model and/or to identify cognate models from an *in vivo* tumor model collection. The aim is to document conservation of the tumor checkpoint activity pattern, as compared to the reference signature dissected from the cell line models used for drug context-specific MOA elucidation (see part 2 of the PMP protocol, above), in the PDX models/GEMMs employed for confirmation of drug MOA and assessment of their anti-tumor effects. To confirm this concordance, the activity of the tumor checkpoint reporter—i.e., the 25 most activated and 25 most inactivated transcriptional regulatory proteins in the patient’s tumor—is assessed in the *in vivo* model by quantifying the enrichment of the TCRS on the model’s protein activity signature (OncoMatch algorithm, see Alvarez et al., 2019; Mundi et al., 2021; Vasciaveo et al., 2022).

A drug’s context-specific MOA can then be dissected from *in vivo* models by profiling the transcriptome of the tumor under vehicle control and comparing the results to the drug’s “on-treatment” conditions. The optimal time point after the initiation of drug treatment *in vivo* to evaluate drug MOA will vary among drugs depending on their particular pharmacokinetic characteristics. However, as a general rule, we profile the animal model’s “on-treatment” arm by resecting the tumor or obtaining a biopsy of the tumor at 3 hr after the third drug administration. Gene expression signatures derived from tumor tissue for the drug-perturbed conditions can then be estimated by comparing the normalized gene expression data against vehicle-control conditions. *In vivo* drug context-specific MOA, using the VIPER algorithm and tissue lineage-matched regulons, can then be inferred by quantifying the differential activity of transcriptional regulatory proteins between the drug-perturbed samples and vehicle-control conditions (Mundi et al., 2021).

Conservation of context-specific MOA between *in vitro* and *in vivo* models can be assessed by correlation analysis or by reciprocal enrichment of the most differentially activated and inactivated proteins in response to the drug perturbation (Alvarez et al., 2019). The

reciprocal enrichment analysis is computed as the average of (a) the enrichment of the top k most differentially active proteins in response to the drug perturbation *in vitro* on the protein activity signature in response to the drug perturbation *in vivo* and (b) the enrichment of the top k most differentially active proteins in response to the drug perturbation *in vivo* on the protein activity signature in response to the drug perturbation *in vitro*. Reciprocal enrichment analysis is implemented by the function “viperSimilarity()” available from the VIPER R-system package (Table 2). Most importantly, inversion of the tumor checkpoint *in vivo* can be quantified by the inverse enrichment of the patient’s TCRS on the *in vivo*, drug-induced protein activity signature (Mundi et al., 2021). This is indeed DarwinOncoTreat analysis testing the inversion of the tumor checkpoint based on the drug context-specific MOA elucidated from *in vivo* drug perturbational data.

Finally, drug efficacy, for drugs with an MOA predicted to invert the activity levels of MRs comprising the tumor checkpoint, should be evaluated in the cognate models by assessing drug effect on tumor growth *in vivo* (Fig. 2D). Drug effect (tumor response) can be described and assessed based on commonly used parameters, including complete response (CR; >95% reduction in tumor volume from baseline), partial response (PR; >50% reduction in tumor volume from baseline), stable disease (SD; $0.5\times < \text{tumor volume} < 2\times$ from baseline tumor volume), and progressive disease (PD; >100% increase in tumor volume from baseline), and can be quantified by summary descriptors, like disease control rate (DCR = SD + PR + CR) and objective response rate (ORR = CR + PR). Tumor response can be evaluated at the time of treatment failure, usually defined as the day on which $\geq 75\%$ of the vehicle control-treated tumors met the criteria for PD, or at the end of the study (Mundi et al., 2021).

Validation of drug-mediated anti-tumor effects as outlined above has been demonstrated in preclinical studies. In a recently reported tumor type-agnostic trial across 14 distinct aggressive malignancies (IRB-AAAA7562), evaluating tumors that had progressed in patients after standard-of-care (SOC) systemic therapy, drugs predicted by DarwinOncoTreat to have tumor-suppressing properties showed a DCR of 91%, in contrast to a vehicle-control DCR of 12%, in the corresponding patient-derived xenograft models, as reported at the

end of the study. Notably, the DCR was only 6% for a set of negative-control anti-neoplastic drugs—including tyrosine kinase inhibitors—selected by a panel of physicians but that were not predicted to be effective by DarwinOncoTreat (Mundi et al., 2021).

Pharmacodynamic studies *in vivo* models can also be performed at the end of an *in vivo* drug efficacy study to obtain and assess transcriptional profiles of tumor cells representing residual disease. These data may be relevant and actionable for the purpose of characterizing the transcriptional identity of drug-resistant tumor niches, which may already be present in bulk tumor samples prior to treatment initiation, even while constituting only a small niche of the phenotypically heterogeneous tumor mass. This information may shed light on mechanisms of drug resistance. Accordingly, characterization of the regulatory checkpoints responsible for drug-resistant cell states can point to combination approaches using additional pharmacological agents, which by destabilizing (i.e., inverting) the transcriptional identity checkpoints in residual drug-resistant niches might (a) specifically induce cell death for tumor cells in such drug-resistant cell states or (b) induce differentiation away from or block differentiation toward such drug-resistant states, effectively depleting the drug-resistant tumor niche while favoring the more actionable, drug-sensitive state.

Part 6 of the PMP Protocol: Refinement of Drug Context-Specific Mechanism of Action from *In Vivo* Data for Use as a Biomarker to Enrich Clinical Trials with Responding Patients

From the perspective of optimizing clinical trial design, the DarwinOncoTreat algorithm employs the drug context-specific MOA elucidated from the cell line and/or *in vivo* models (parts 3 and 5 of the PMP protocol, above) as a biomarker of drug response (Alvarez et al., 2018; Mundi et al., 2021) (Fig. 2E). By quantitatively evaluating the enrichment of each tumor checkpoint—using the patient's tumor-specific TCRS—on each drug's context-specific MOA, the DarwinOncoTreat algorithm predicts whether drugs, based on their MOA elucidated in part 3 (*in vitro*) and/or part 5 (*in vivo*) of the PMP protocol, are acting as *inverters of the tumor checkpoint activity pattern*—so-called “inverting drugs,” and prioritizes these drugs. Importantly, DarwinOncoTreat-based predic-

tions of drug response are strictly mechanism based, both because the prediction is based on the regulatory architecture responsible for maintaining the transcriptional identity of the tumor cells (tumor checkpoint) and because it is based on the matched, context-specific MOA of the drug. Given the strong conservation of the tumor checkpoint among tumors of the same subtype, as identified by protein activity-based unsupervised cluster analysis (Paull et al., 2021), one would expect to see a similar stratification of drugs based on DarwinOncoTreat alignments. In fact, such clusters of patients based on DarwinOncoTreat drug alignments have been referred to as “pharmacotypes” (Fig. 4) (Mundi et al., 2021). Pharmacotypes help identify drugs consistently predicted by DarwinOncoTreat to benefit a responsive fraction of patients in a clinically relevant cohort and, by extension, provide a critical mechanism-based biomarker for identifying patients who should be recruited for drug evaluation in the context of biomarker-enriched clinical trials.

DarwinOncoTreat has been used in pre-clinical research to predict the response of individual tumors to FDA-approved drugs and investigational compounds. For instance, DarwinOncoTreat has been shown to be effective at identifying drugs for individual gastro-entero-pancreatic neuroendocrine tumors (Alvarez et al., 2019), cholangiocarcinomas (Obradovic, Tomassoni, et al., 2022), castration-resistant and metastatic prostate carcinoma (Vasciaveo et al., 2022), triple-negative breast carcinoma, pancreatic ductal carcinoma, colon adenocarcinoma, and gastrointestinal stromal tumors and meningioma (Mundi et al., 2021), and even to reprogram tumor-infiltrating regulatory T cells (Obradovic, Ager, et al., 2022). DarwinOncoTreat-based biomarkers of drug response are currently used in clinical research to predict the response of gastroenteropancreatic neuroendocrine tumors to entinostat (NCT03211988) and to predict the response of pancreas adenocarcinoma patients to several drugs in the umbrella HIPPOCRATES trial (NCT04476537).

SUMMARY AND CONCLUDING REMARKS

The PMP protocol we report is a patient-centered, systems biology-based pipeline for cancer drug discovery with broad applications in the precision oncology space. Each step of this compound-to-clinic drug development and validation pipeline is leveraged by the

integration of computational algorithms and experimental data in order to ensure that novel cancer targets—i.e., the MR proteins controlling the tumor cell transcriptional identity—and the postulated MOA of drugs being evaluated are consistent as the protocol progresses from cell lines to animal models and, finally, into patients for clinical translation.

The protocol begins with molecular profiling of the patient's tumor transcriptome by RNA-Seq, to which a regulatory network-based analysis is applied to identify the distinct set of dysregulated proteins responsible for maintaining the transcriptional identity of the cancer cell (part 1). These proteins, also known as tumor MRs, represent non-oncogene tumor vulnerabilities and are collectively known as the tumor checkpoint (Califano & Alvarez, 2017). Identification of the tumor checkpoint from patient databases (TCGA and others) using the VIPER algorithm represents the foundational element that underpins all subsequent steps of the PMP protocol. Critically, the identification and selection of cognate *in vitro* and *in vivo* model systems (part 2) for experimental study requires conservation of the patient's tumor checkpoint, thereby ensuring that the models selected for subsequent drug perturbation will recapitulate the activity of the patient's MRs, and the *in vivo* models selected for preclinical validation will faithfully recapitulate the patient's tumor vulnerabilities.

Parts 3 and 4, in conjunction, are aimed at identifying those “inverting drugs”—i.e., those that are most effective at inhibiting the activity of the most aberrantly over-activated MR proteins while increasing the activity level of the most under-activated MRs—that will collapse the tumor checkpoint and, by extension, abrogate the transcriptional identity state that is responsible for cancer-related bio-functional hallmarks that sustain tumor survival. The MOA is elucidated by first perturbing the *in vitro* cognate models (cell lines) identified above using a library of drugs and compounds of interest. Effects of the drug perturbations on the activity of MR proteins comprising the tumor checkpoint are quantified by RNA-Seq-based transcriptome profiling coupled with VIPER algorithm analysis. Upon completion of this step, drugs that were demonstrated to invert the tumor checkpoint are selected for preclinical validation in cognate *in vivo* models (part 5). The effects of the selected drugs on tumor growth kinetics, progression-free survival, and/or overall survival are compared against vehicle control. Fi-

nally, drugs showing significant improvement in the survival rate of the animals tested and/or significant reduction of tumor growth in the preclinical validation are selected for clinical evaluation. Importantly, the biomarker-driven trial (part 6) is enriched with patients with the highest chances of responding by recruiting patients whose tumor checkpoint signature (the presumptive biomarker) is predicted to be inverted as a result of the drug/drug MOA elucidated earlier (part 3) in the PMP protocol.

When deployed in a systematic fashion, as it is at DarwinHealth, where the commercial applications of the PMP protocol are under continual refinement and expansion, this novel cancer target and drug discovery pipeline provides a roadmap for precision- and MOA-based identification and clinical validation of novel, investigational, and FDA-approved therapeutic agents targeting the tumor regulatory architecture. Its universality in the context of cancer biology permits the PMP model to be applied across a broad spectrum of hematological and solid tumors, for which numerous clinical trials based on this drug discovery technology are currently in progress.

Current Limitations and Future Directions

One of the main limitations of the PMP protocol is the inability to address the challenges and confounding factors imposed by tumor tissue heterogeneity from bulk RNA-Seq data. Typically, analysis of tumor tissue and identification of tumor checkpoints using bulk RNA-Seq data can only capture the average signal from a complex, variable mixture of different cell types, a stromal infiltration, and tumor transcriptional states that co-exist in a heterogeneous tumor tissue sample. Analysis of such a convoluted signature will only allow a limited dissection of the tumor multi-niche/cell state while missing minor tumor and tumor microenvironment niches.

In recent years, however, the advent of single-cell profiling technologies has significantly increased our ability to characterize, resolve, and understand tumor heterogeneity (Suva & Tirosh, 2019), with the aim of (a) accounting for diverse transcriptional cell states and differentiated regulatory architecture co-existing within a single tumor and (b) identifying single drugs or combinations that can target multiple cell niches within a tumor. For instance, one of the earliest single-cell studies in cancer showed that the molecular subtypes of GBM defined through bulk tumor analysis, despite being predictive of patient prognosis,

did not recapitulate the molecular diversity of each of the individual GBM tumors (Patel et al., 2014). Indeed, multiple cell populations characterized by distinct phenotypic states and biological pathways were found to be co-existing in each individual tumor. These findings were later refined for GBM (Ding et al., 2019; Neftel et al., 2019) and extended to other tumor types and the tumor microenvironment (Puram et al., 2017; Wu et al., 2021).

Accordingly, the limitations imposed by the bulk RNA-Seq-based transcriptome analysis may obscure the potential impact that cell populations of relatively low representation are making on tumor progression and drug resistance, thereby limiting the identification of potential pharmacological targets and molecular biomarkers. Although scRNA-Seq has shown an unprecedented ability to characterize cell populations from tissue samples (Abdelaal et al., 2019; Tabula Muris Consortium et al., 2018), the transition from bulk tumor RNA-Seq to scRNA-Seq, in particular for clinical application, is not trivial. The low sequencing depth per cell characteristic of scRNA-Seq, resulting in part from small quantities of mRNA in individual cells and inefficient mRNA capture (Kharchenko, Silberstein, & Scadden, 2014), produces an elevated rate of drop-out events, i.e., undetected transcripts from the transcriptome snapshot of a particular cell while the gene is being actively expressed. In scRNA-Seq, drop-out events typically occur in up to 80% to 90% of genes, severely affecting the transcriptional readout of each single cell and, consequently, our ability to identify novel pharmacological targets and molecular biomarkers based on individual gene products.

Because VIPER estimates the activity of each regulatory protein based on a large set of the protein's transcriptional targets, it is inherently well suited to address the technical challenges presented by low sequencing depth and drop-out effects. In fact, it has been shown that although the number of detected transcripts in a particular regulon is higher than 25, VIPER is able to generate accurate estimations of protein activity (Alvarez et al., 2016). This has also been illustrated by *in silico* sequencing dilution experiments, showing a minimal drop in accuracy when the sequencing depth is reduced from 30 million to 100,000 reads, a sequencing depth that is compatible with most scRNA-Seq applications, even though gene expression signatures are strongly degraded (Alvarez et al., 2016). To account for multiple

cell types present in the tumor tissue, VIPER has been extended to use multiple tissue lineage context-specific regulon models, and this meta-lineage VIPER implementation, or metaVIPER, has shown comparably high accuracy for quantifying protein activity among individual cells from complex tissue type mixtures (Ding et al., 2018). metaVIPER can transform highly sparse scRNA-Seq expression matrices into complete and robust protein activity matrices. Indeed, metaVIPER has been shown to outperform scRNA-Seq-based transcript abundance and even antibody-based flow cytometry when dissecting the microenvironment of renal carcinoma at the single-cell level (Obradovic et al., 2021).

The inherent ability of VIPER to generate accurate and robust protein activity measurements from the sparse, low-depth scRNA-Seq data allows for the straightforward expansion of the PMP protocol to single-cell applications, including the elucidation and targeting of the tumor checkpoint for distinct tumor niches/transcriptional identities co-existing in the same tumor tissue (Laise et al., 2020), as well as specific components of the tumor stroma (Obradovic, Ager, et al., 2022). In fact, dissection of the tumor checkpoint at the single-cell level has facilitated identification of a novel transcriptional identity in pancreatic ductal adenocarcinoma, referred to as “oncogenic precursors,” which are present in all tumors that have been profiled at the single-cell level but undetectable by bulk RNA-Seq-coupled VIPER analysis (Laise et al., 2020). Moreover, metaVIPER analysis has been effective at identifying and characterizing the transcriptional identity and the corresponding regulatory checkpoint for specific tumor microenvironment components. For instance, metaVIPER has been used to dissect the regulatory checkpoint for recurrence-associated, tumor-infiltrating macrophages in renal carcinoma (Obradovic et al., 2021) and antigen-presenting, tumor-associated fibroblasts in pancreatic ductal adenocarcinoma (Elyada et al., 2019). Furthermore, application of DarwinOncoTreat to single-cell-level, tumor, and tumor microenvironment scRNA-Seq data has been effective at identifying drugs with potential to target the tumor checkpoints of individual cholangiocarcinoma tumor cell niches (Obradovic, Tomassoni, et al., 2022) and the regulatory checkpoint of tumor-infiltrating regulatory T cells in prostate adenocarcinoma, bladder cancer, clear cell renal

carcinoma, and GBM (Obradovic, Ager, et al., 2022).

ACKNOWLEDGMENTS

We would like to thank Drs. Xiaoyun Sun and Stuart Andrews for their assistance in collecting, organizing, and assembling the database of RNA-Seq profiles for tumor clinical cohorts used by the PMP protocol. We would like to extend our thanks to Drs. Charles Karan, Ronald Realubit, and Sergey Pampou and the Columbia Sulzberger High-Throughput Screening and Genome Center for their help generating the drug perturbational data used by the PMP protocol to infer drug context-specific MOA.

AUTHOR CONTRIBUTIONS

Pasquale Laise: Formal analysis, Methodology, Writing—original draft, Writing—review and editing; **Gideon Bosker:** Conceptualization, Writing—original draft, Writing—review and editing; **Andrea Califano:** Conceptualization, Writing—original draft, Writing—review and editing; **Mariano J. Alvarez:** Conceptualization, Formal analysis, Methodology, Writing—original draft, Writing—review and editing.

CONFLICT OF INTEREST

P.L. is Director of Single-Cell Systems Pharmacology at DarwinHealth, Inc., a company that has licensed some of the algorithms used in this article from Columbia University. G.B. is a founder, the CEO, and an equity holder of DarwinHealth, Inc. A.C. is a founder, an equity holder, and a consultant of DarwinHealth, Inc. M.J.A. is the CSO and an equity holder of DarwinHealth, Inc. Columbia University is also an equity holder of DarwinHealth, Inc.

DATA AVAILABILITY STATEMENT

All data used in this manuscript are publicly available as indicated.

LITERATURE CITED

Abdelaal, T., Michielsen, L., Cats, D., Hoogduin, D., Mei, H., Reinders, M. J. T., & Mahfouz, A. (2019). A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biology*, *20*, 194. doi: 10.1186/s13059-019-1795-z

Alvarez, M. J., Shen, Y., Giorgi, F. M., Lachmann, A., Ding, B. B., Ye, B. H., & Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nature Genetics*, *48*, 838–847. doi: 10.1038/ng.3593

Alvarez, M. J., Subramaniam, P. S., Tang, L. H., Grunn, A., Aburi, M., Rieckhof, G., ... Califano, A. (2018). A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. *Nature Genetics*, *50*, 979–989. doi: 10.1038/s41588-018-0138-4

Alvarez, M. J., Yan, P., Alpaugh, M. L., Bowden, M., Sicinska, E., Zhou, C. W., ... Califano, A. (2019). Unbiased assessment of H-STS cells as high-fidelity models for gastro-enteropancreatic neuroendocrine tumor drug mechanism of action analysis. *bioRxiv*, 677435. doi: 10.1101/677435

Arumugam, K., Shin, W., Schiavone, V., Vlahos, L., Tu, X., Carnevali, D., ... Cosma, M. P. (2020). The master regulator protein BAZ2B can reprogram human hematopoietic lineage-committed progenitors into a multipotent state. *Cell Reports*, *33*, 108474. doi: 10.1016/j.celrep.2020.108474

Aytes, A., Mitrofanova, A., Kinkade, C. W., Lefebvre, C., Lei, M., Phelan, V., ... Abate-Shen, C. (2013). ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, E3506–3515. doi: 10.1073/pnas.1303558110

Aytes, A., Mitrofanova, A., Lefebvre, C., Alvarez, M. J., Castillo-Martin, M., Zheng, T., ... Abate-Shen, C. (2014). Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell*, *25*, 638–651. doi: 10.1016/j.ccr.2014.03.017

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., ... Garraway, L. A. (2012). The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, *483*, 603–607. doi: 10.1038/nature11003

Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., & Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nature Genetics*, *37*, 382–390. doi: 10.1038/ng1532

Bisikirska, B., Bansal, M., Shen, Y., Teruya-Feldstein, J., Chaganti, R., & Califano, A. (2016). Elucidation and pharmacological targeting of novel molecular drivers of follicular lymphoma progression. *Cancer Research*, *76*, 664–674. doi: 10.1158/0008-5472.CAN-15-0828

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, *169*, 1177–1186. doi: 10.1016/j.cell.2017.05.038

Brichta, L., Shin, W., Jackson-Lewis, V., Blesa, J., Yap, E. L., Walker, Z., ... Greengard, P. (2015). Identification of neurodegenerative factors using translational-regulatory network analysis. *Nature Neuroscience*, *18*, 1325–1333. doi: 10.1038/nn.4070

Bush, E. C., Ray, F., Alvarez, M. J., Realubit, R., Li, H., Karan, C., ... Sims, P. A. (2017).

- PLATE-Seq for genome-wide regulatory network analysis of high-throughput screens. *Nature Communication*, 8, 105. doi: 10.1038/s41467-017-00136-z
- Califano, A., & Alvarez, M. J. (2017). The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nature Reviews Cancer*, 17, 116–130. doi: 10.1038/nrc.2016.124
- Califano, A., Butte, A. J., Friend, S., Ideker, T., & Schadt, E. (2012). Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nature Genetics*, 44, 841–847. doi: 10.1038/ng.2355
- Carro, M. S., Lim, W. K., Alvarez, M. J., Bollo, R. J., Zhao, X., Snyder, E. Y., ... Iavarone, A. (2010). The transcriptional network for mesenchymal transformation of brain tumours. *Nature*, 463, 318–325. doi: 10.1038/nature08712
- Compagno, M., Lim, W. K., Grunn, A., Nandula, S. V., Brahmachary, M., Shen, Q., ... Pasqualucci, L. (2009). Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature*, 459, 717–721. doi: 10.1038/nature07968
- Della Gatta, G., Palomero, T., Perez-Garcia, A., Ambesi-Impiombato, A., Bansal, M., Carpenter, Z. W., ... Ferrando, A. A. (2012). Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL. *Nature Medicine*, 18, 436–440. doi: 10.1038/nm.2610
- Ding, H., Burgenske, D. M., Zhao, W., Subramaniam, P. S., Bakken, K. K., He, L., ... Califano, A. (2019). Single-cell based elucidation of molecularly-distinct glioblastoma states and drug sensitivity. *bioRxiv*, 675439. doi: 10.1101/675439
- Ding, H., Douglass, E. F. Jr., Sonabend, A. M., Mela, A., Bose, S., Gonzalez, C., ... Califano, A. (2018). Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nature Communication*, 9, 1471. doi: 10.1038/s41467-018-03843-3
- Douglass, E. F. Jr., Allaway, R. J., Szalai, B., Wang, W., Tian, T., Fernandez-Torras, A., ... Califano, A. (2022). A community challenge for a pancancer drug mechanism of action inference from perturbational profile data. *Cell Reports Medicine*, 3, 100492. doi: 10.1016/j.xcrm.2021.100492
- Dutta, A., Le Magnen, C., Mitrofanova, A., Ouyang, X., Califano, A., & Abate-Shen, C. (2016). Identification of an NKX3.1-G9a-UTY transcriptional regulatory network that controls prostate differentiation. *Science*, 352, 1576–1580. doi: 10.1126/science.aad9512
- Elyada, E., Bolisetty, M., Laise, P., Flynn, W. F., Courtois, E. T., Burkhart, R. A., ... Tuveson, D. A. (2019). Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discovery*, 9, 1102–1123. doi: 10.1158/2159-8290.CD-19-0094
- Garcia-Alonso, L., Iorio, F., Matchan, A., Fonseca, N., Jaaks, P., Peat, G., ... Saez-Rodriguez, J. (2018). Transcription factor activities enhance markers of drug sensitivity in cancer. *Cancer Research*, 78, 769–780. doi: 10.1158/0008-5472.CAN-17-1679
- Ikiz, B., Alvarez, M. J., Re, D. B., Le Verche, V., Politi, K., Lotti, F., ... Przedborski, S. (2015). The regulatory machinery of neurodegeneration in in vitro models of amyotrophic lateral sclerosis. *Cell Reports*, 12, 335–345. doi: 10.1016/j.celrep.2015.06.019
- Keenan, A. B., Torre, D., Lachmann, A., Leong, A. K., Wojciechowicz, M. L., Utti, V., ... Ma'ayan, A. (2019). ChEA3: Transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Research*, 47, W212–W224. doi: 10.1093/nar/gkz446
- Kharchenko, P. V., Silberstein, L., & Scadden, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nature Methods*, 11, 740–742. doi: 10.1038/nmeth.2967
- Khatamian, A., Paull, E. O., Califano, A., & Yu, J. (2019). SJARACNe: A scalable software tool for gene network reverse engineering from big data. *Bioinformatics*, 35, 2165–2166. doi: 10.1093/bioinformatics/bty907
- Kushwaha, R., Jagadish, N., Kustagi, M., Tomishima, M. J., Mendiratta, G., Bansal, M., ... Chaganti, R. S. K. (2015). Interrogation of a context-specific transcription factor network identifies novel regulators of pluripotency. *Stem Cells*, 33, 367–377. doi: 10.1002/stem.1870
- Lachmann, A., Giorgi, F. M., Lopez, G., & Califano, A. (2016). ARACNe-AP: Gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics*, 32, 2233–2235. doi: 10.1093/bioinformatics/btw216
- Laise, P., Stanifer, M., Bosker, G., Sun, X., Triana, S., Doldan, P., ... Alvarez, M. (2022). A model for network-based identification and pharmacological targeting of aberrant, replication-permissive transcriptional programs induced by viral infection. *Communications Biology*, 5(1), 714. doi: 10.1038/s42003-022-03663-8
- Laise, P., Turunen, M., Maurer, H. C., Curiel, A. G., Elyada, E., Schmierer, B., ... Califano, A. (2020). Pancreatic ductal adenocarcinoma comprises coexisting regulatory states with both common and distinct dependencies. *bioRxiv*, 2020.2010.2027.357269. doi: 10.1101/2020.10.27.357269
- Lefebvre, C., Rajbhandari, P., Alvarez, M. J., Bandaru, P., Lim, W. K., Sato, M., ... Califano, A. (2010). A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Molecular Systems Biology*, 6, 377. doi: 10.1038/msb.2010.31
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion

- for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550. doi: 10.1186/s13059-014-0550-8
- Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., & Califano, A. (2006). ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, *7*(Suppl 1), S7. doi: 10.1186/1471-2105-7-S1-S7
- Marks, L. J., Diolaiti, D., Mundi, P., Gaviria, E. S., Rainey, A. R., Yamashiro, D. J., ... Kung, A. L. (2021). Identification and validation of a non-genetically encoded vulnerability to XPO1 inhibition in malignant rhabdoid tumors – expanding patient-driven discovery beyond the N-of-1. *bioRxiv*, 2021.2010.2002.462793. doi: 10.1101/2021.10.02.462793
- Mundi, P. S., Dela Cruz, F. S., Grunn, A., Diolaiti, D., Manguen, A., Rainey, A. R., ... Califano, A. (2021). Pre-clinical validation of an RNA-based precision oncology platform for patient-therapy alignment in a diverse set of human malignancies resistant to standard treatments. *bioRxiv*, 2021.2010.2003.462951. doi: 10.1101/2021.10.03.462951
- Neftel, C., Laffy, J., Filbin, M. G., Hara, T., Shore, M. E., Rahme, G. J., ... Suvà, M. L. (2019). An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell*, *178*, 835–849.e821. doi: 10.1016/j.cell.2019.06.024
- Obradovic, A., Ager, C., Turunen, M., Nirschl, T., Khosravi-Maharlooeei, M., Jackson, C., ... Califano, A. (2022). Systematic elucidation and pharmacological targeting of tumor-infiltrating regulatory T cell master regulators. *bioRxiv*, 2022.2002.2022.481404. doi: 10.1101/2022.02.22.481404
- Obradovic, A., Chowdhury, N., Haake, S. M., Ager, C., Wang, V., Vlahos, L., ... Drake, C. G. (2021). Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages. *Cell*, *184*, 2988–3005.e2916. doi: 10.1016/j.cell.2021.04.038
- Obradovic, A., Tomassoni, L., Yu, D., Guilan, K., Souto, K., Fraser, E., ... Califano, A. (2022). Case study of single-cell protein activity based drug prediction for precision treatment of cholangiocarcinoma. *bioRxiv*, 2022.2002.2028.482410. doi: 10.1101/2022.02.28.482410
- Pan, H., Xue, C., Auerbach, B. J., Fan, J., Bashore, A. C., Cui, J., ... Reilly, M. P. (2020). Single-cell genomics reveals a novel cell state during smooth muscle cell phenotypic switching and potential therapeutic targets for atherosclerosis in mouse and human. *Circulation*, *142*, 2060–2075. doi: 10.1161/CIRCULATIONAHA.120.048378
- Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., ... Bernstein, B. E. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, *344*, 1396–1401. doi: 10.1126/science.1254257
- Paull, E. O., Aytes, A., Jones, S. J., Subramaniam, P. S., Giorgi, F. M., Douglass, E. F., ... Califano, A. (2021). A modular master regulator landscape controls cancer transcriptional identity. *Cell*, *184*, 334–351.e320. doi: 10.1016/j.cell.2020.11.045
- Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., ... Bernstein, B. E. (2017). Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell*, *171*, 1611–1624.e1624. doi: 10.1016/j.cell.2017.10.044
- Putcha, P., Yu, J., Rodriguez-Barrueco, R., Saucedo-Cuevas, L., Villagrasa, P., Murga-Penas, E., ... Silva, J. (2015). HDAC6 activity is a non-oncogene addiction hub for inflammatory breast cancers. *Breast Cancer Research*, *17*, 149. doi: 10.1186/s13058-015-0658-0
- Rajbhandari, P., Lopez, G., Capdevila, C., Salvatore, B., Yu, J., Rodriguez-Barrueco, R., ... Califano, A. (2018). Cross-cohort analysis identifies a TEAD4-MYCIN positive feedback loop as the core regulatory element of high-risk neuroblastoma. *Cancer Discovery*, *8*, 582–599. doi: 10.1158/2159-8290.CD-16-0861
- Rodriguez-Barrueco, R., Yu, J., Saucedo-Cuevas, L. P., Olivan, M., Llobet-Navas, D., Putcha, P., ... Silva, J. M. (2015). Inhibition of the autocrine IL-6-JAK2-STAT3-calprotectin axis as targeted therapy for HR-/HER2+ breast cancers. *Genes & Development*, *29*, 1631–1648.
- Shen, Y., Alvarez, M. J., Bisikirska, B., Lachmann, A., Realubit, R., Pampou, S., ... Califano, A. (2017). Systematic, network-based characterization of therapeutic target inhibitors. *PLoS Computational Biology*, *13*, e1005599. doi: 10.1371/journal.pcbi.1005599
- Son, J., Ding, H., Farb, T. B., Efanov, A. M., Sun, J., Gore, J. L., ... Califano, A. (2021). BACH2 inhibition reverses beta cell failure in type 2 diabetes models. *Journal of Clinical Investigation*, *131*(24), e153876. doi: 10.1172/JCI153876
- Sonabend, A. M., Bansal, M., Guarnieri, P., Lei, L., Amendolara, B., Soderquist, C., ... Canoll, P. (2014). The transcriptional regulatory network of proneural glioma determines the genetic alterations selected during tumor progression. *Cancer Research*, *74*, 1440–1451. doi: 10.1158/0008-5472.CAN-13-2150
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... Mesirov, J. P. (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. doi: 10.1073/pnas.0506580102
- Suva, M. L., & Tirosh, I. (2019). Single-cell RNA sequencing in cancer: Lessons learned and emerging challenges. *Molecular Cell*, *75*, 7–12. doi: 10.1016/j.molcel.2019.05.003
- Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, ... Principal

- investigators. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, *562*, 367–372. doi: 10.1038/s41586-018-0590-4
- Talos, F., Mitrofanova, A., Bergren, S. K., Califano, A., & Shen, M. M. (2017). A computational systems approach identifies synergistic specification genes that facilitate lineage conversion to prostate tissue. *Nature Communication*, *8*, 14662. doi: 10.1038/ncomms14662
- Vasciaveo, A., Zou, M., Arriaga, J. M., de Almeida, F. N., Douglass, E. F., Shibata, M., ... Abate-Shen, C. (2022). OncoLoop: A network-based precision cancer medicine framework. *bioRxiv*, 2022.2002.2011.479456. doi: 10.1101/2022.02.11.479456
- Walsh, L. A., Alvarez, M. J., Sabio, E. Y., Reynolds, M., Makarov, V., Mukherjee, S., ... Chan, T. A. (2017). An integrated systems biology approach identifies TRIM25 as a key determinant of breast cancer metastasis. *Cell Reports*, *20*, 1623–1640. doi: 10.1016/j.celrep.2017.07.052
- Wu, S. Z., Al-Eryani, G., Roden, D. L., Junankar, S., Harvey, K., Andersson, A., ... Swarbrick, A. (2021). A single-cell and spatially resolved atlas of human breast cancers. *Nature Genetics*, *53*, 1334–1347. doi: 10.1038/s41588-021-00911-1
- Zelege, T. Z., Pan, Q., Chiuzan, C., Onishi, M., Alvarez, M. J., Honan, E., ... Silva, J. (2020). Network-based assessment of HDAC6 activity is highly predictive of pre-clinical and clinical responses to the HDAC6 inhibitor ricolinostat. *medRxiv*, 2020.2004.2023.20066928. doi: 10.1101/2020.04.23.20066928