

## MISSION STATEMENT and EXECUTIVE SUMMARY

**DarwinHealth™: Precision Therapeutics for Cancer Medicine** is leveraging breakthrough discoveries and novel systems biology methodologies to implement a radically new approach for precision cancer medicine, making patient- and tumor-specific targeted therapy a practical reality for virtually every cancer patient.

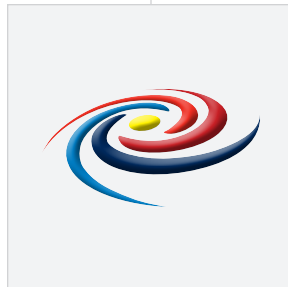
While individual health is increasingly managed on a predictive and preventative basis, the true promise of personalized, genomic medicine is yet to be fulfilled. This is especially true in the context of human malignancies, where current DNA sequencing-based approaches—widely employed across the continuum of cancer care, despite well known limitations—are not likely to significantly impact the life expectancy of most cancer patients. Indeed, despite a few spectacular successes, only a minority of patients (<25%) presents genetic alterations that can be targeted pharmacologically, and even fewer (<10%) are poised to derive direct clinical benefit from such knowledge.

As a result, radically novel approaches are needed to make precision cancer medicine a clinically relevant and broadly applicable paradigm for all cancer patients. In response to this interventional shortfall, **DarwinHealth™** has pioneered and is deploying proprietary technologies to tailor targeted therapy to the critical tumor dependencies that exist in all individual patients, thus paving the road for a more universal application of personalized medicine.

It is now broadly accepted that cancer and other diseases are driven by the aberrant activity of multiple proteins acting in concert to dysregulate normal cellular function. Genetic alterations in their corresponding genes are only one way among many to induce such pathologic behavior. Indeed, paradigm-changing, systems biology research at Columbia University has shown that malignant transformation, metastatic progression, and drug resistance are all driven by the abnormal activity of pivotal

master regulator (MR) proteins that are rarely if ever directly mutated. Critically, these master regulators work in concert to implement highly conserved *tumor checkpoints*, whose activity is critical for cancer cell survival and can be exploited by targeted therapy. Not surprisingly, any pattern of mutations that can activate such checkpoints, no matter how complex, can also induce cancer initiation or progression. Much as in the way a vase may break in a million different pieces yet never in the same exact way, the number of mutational patterns that can spark tumor checkpoint activation is virtually infinite. In fact, simple math indicates that the number of distinct tumorigenic patterns far exceeds the number of cells on our planet; hence, the difficulty in understanding how cancer begins, how it subsequently evolves, and how it may be managed therapeutically, using only DNA mutation data.

As a result, much of genetics-driven cancer therapy currently finds itself at an impasse, with only a handful of statistically interpretable mutations (<5%) that may inform clinical decision-making. In contrast, the role of other cancer genome mutations remains obscure and has been compared to the dark matter of the universe. In summary, the disconnect between our ability to fully characterize and leverage the mutational signature of tumors into actionable therapy can be explained by: (a) the regulatory, mechanical, and functional separation between the cancer mutations we can detect by sequencing and the aberrant activity of the full complement of potentially targetable oncoproteins that comprise critical tumor checkpoint; and (b) the lack of methodologies to directly assess functional cancer drivers on an individual patient basis—typically, aberrantly activated proteins *rather* than mutated genes. Understanding of the former combined with availability of the latter is critically required to identify the most vulnerable action points or “treatment turnstiles” at which cancer agents should be aimed to be maximally effective.



## MISSION STATEMENT and EXECUTIVE SUMMARY

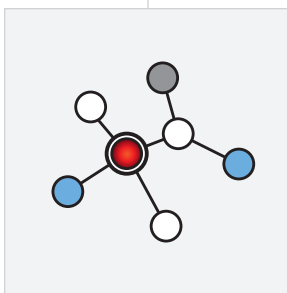
Not surprisingly, perhaps, the daunting challenge of making precision medicine — and especially precision-focused cancer medicine—more “precise” has presented a major hurdle for advocates of genomics-based cancer therapeutics that relies almost exclusively on mutation-based analysis at the DNA level. Indeed, genetic alterations alone, while providing useful information to inform therapeutic strategy for a minority of patients, represent woefully, statistically underpowered markers for extending the precision cancer medicine paradigm to all patients or even to a majority of them. As a result, alternative, more integrative approaches are needed that account for the complexity of cancer cell regulation when analyzing the wealth of information that genomics technologies have put at our fingertips.

**DarwinHealth™** is precisely dedicated to negotiating the current barriers between a purely mutation-driven view of cancer therapeutics and its actual functional implementation at the protein level, using the regulatory logic of the cancer cell as an interpretative and integrative blueprint. Indeed, the functionally critical proteins that are necessary for survival of a specific cancer can now be systematically and accurately elucidated by proprietary **DarwinHealth™** computational methods that make deeper probes into cellular dynamics possible. These methodologies are supported by a wealth of high impact publications and by work spanning from the basic science to preclinical and clinical applications at Columbia University (see our comprehensive bibliography).

Leveraging IP exclusively licensed from Columbia University, **DarwinHealth™** has developed the most innovative, consistently validated, reproducible, and practical approach for the systematic identification of aberrantly activated master regulator (MR) proteins within critical tumor checkpoints that drive cancer initiation, progression, and drug resistance (*TumorCheckPoint™*). This has proven instrumental in bridging the clinically and therapeutically problematic disconnect between DNA-based mutational analysis and the true functional implementers of cancer phenotypes—and, by extension, aberrant cell behaviors characteristic of the malignant state. Critically, rather than requiring costly DNA mutational profiles, this technology allows identification of MR proteins that have been extensively validated as the gate-keepers of the “cancer cell state” using only mRNA abundance data from RNA sequencing (RNA-Seq) profiles—a methodology that easily works with either fresh frozen or formalin fixed paraffin embedded (FFPE) tumor samples harvested from individual patients.

Developing and refining the tumor checkpoint models responsible for driving tumorigenesis and the analytical technology that permits their systematic elucidation has been the life’s work of **DarwinHealth™** Co-Founder, Professor Andrea Califano, Dr., the Wu Professor of Chemical Systems Biology, Chair of the Department of Systems Biology, Director, JP Sulzberger Columbia Genome Center, Associate Director of the Herbert Irving Comprehensive Cancer Center at Columbia University. Professor Califano has pioneered computational algorithms based on information theory to precisely reverse engineer the regulatory logic of the cancer cell, as implemented by the physical interaction between regulatory proteins and their targets and substrates, across a broad range of tumor types. This has allowed systematic elucidation of the precise tumor checkpoints and MR proteins that drive tumor initiation, progression, and drug resistance. These checkpoints and associated MR proteins have been systematically validated in animal studies; and more recently, in human

clinical studies. Importantly, he has shown that tumor checkpoints act as functional bottlenecks to integrate the effect of genetic alterations in upstream pathways to induce aberrant activity of their MR proteins, thus activating the genetic programs necessary for cancer initiation and maintenance—a concept best characterized as the “*cancer bottleneck hypothesis*.”



In the process, he and his team of co-investigators have made a whole new world of molecular-level regulatory interactions both visible and computationally queryable. This knowledge has not only supported the systematic elucidation of regulatory oncoproteins and master regulator (MR) proteins but has allowed prediction of novel tumor therapies that have been confirmed experimentally and are now being tested in the cancer clinic setting with promising results. Employing **DarwinHealth™**-based technologies, he and his collaborators have identified an expanding universe of plausible and actionable targets for cancer treatment, which, until recently, had remained opaque, impenetrable, and poorly understood.

Together with Dr. Califano, **DarwinHealth™** co-founder, Gideon Bosker, MD and **DarwinHealth™** Chief Scientific, Technology, and Innovation Officer, Mariano Alvarez, PhD, who has worked alongside Califano at Columbia for seven years, believe that evolving technologies based in systems biology, are now poised to change the world of cancer therapeutics by establishing a rigorous analytical framework for pharmacological hypothesis generation and testing. Importantly, these

## MISSION STATEMENT and EXECUTIVE SUMMARY

methodologies are ideally positioned and implementation-ready to provide a new generation of therapeutic advances and pharmacological options that would allow every cancer patient to immediately benefit from targeted therapies.

It is the position of **DarwinHealth™** leadership that their proprietary computational methods and algorithms, which mechanistically leapfrog the previous world of statistical associations, are now ready to be systematically applied to prioritizing tumor therapeutics using inexpensive molecular profiles (i.e., RNA-Seq based). These approaches facilitate the “unveiling” of critical tumor checkpoints and will allow clinicians to “act upon” these pivotal targets that, for nearly all others using alternative technologies, remain largely invisible and/or indecipherable.

How does **DarwinHealth™** technology accomplish its lofty objective of bringing precision therapeutics to the world of cancer medicine? What are the nuts-and-bolts mechanics of its analytical methods based on mRNA profiling? *Gene expression profiles (GEPs) of mRNA abundance (RNA-Seq) of specific tumors in specific patients are easily and routinely measured.* Yet, when interrogated at the individual gene level, they are even less reliable than DNA mutations in predicting activity of drugs. For instance, expression of individual genes is poorly reproducible across distinct samples and is not indicative of the activity of their encoded proteins.

**DarwinHealth™** has adopted and improved key technology developed by the Califano lab to assess activity of proteins from the expression of their tumor-specific targets. While this requires extremely accurate maps of protein-target regulation, which vary from cancer to cancer, the resulting measurements are extremely reproducible and truly characterize the tumor relevant activity of these functional units. Assessment of differential protein activity (e.g., in tumor vs. normal tissue) has been shown to be critical in determining the key players that underpin the cancer initiation, propagation and invasion processes, from unregulated proliferation and immunoevasion to metastatic invasion.

The ability to precisely pinpoint those master regulators that are primarily responsible for these “executive” breakdowns of cellular function is the arduous, proprietary level of analysis that **DarwinHealth™** has mastered. At the simplest level, it may be compared to reconstructing an image from the noisy bits in its holographic representation. The

**DarwinHealth™** analytical platform performs complex, information theory-based computational “crunching” of the mRNA abundance measurements using the Virtual Proteomics by Enriched Regulon algorithm (VIPER), which reconciles the measured expression values against a pre-validated matrix of known regulatory protein interactions (an interactome). *The result of these calculations is the accurate assessment of protein activity, all the way from aberrantly inactivated (e.g., tumor suppressors) to aberrantly activated (e.g., oncogenes), for all regulatory proteins in the cell.*

Leveraging these discoveries and fine-tuning them to address unmet market needs, **DarwinHealth™** has developed three novel, pillar technologies and platforms, **DarwinOncoTarget™**, **DarwinOncoTreat™**, and **DarwinOncoDiscovery™**, each of which provides a distinct, but complementary, methodology to advance precision therapeutics in the field of cancer medicine. The first two products address cancer institution and patient-centric needs for optimizing alignment of therapeutic agents with specific tumors (at all their stages of progression and invasiveness) in individual patients. In the case of **DarwinOncoTarget™**, we report the master regulator oncoproteins that represent potential druggable targets of currently approved or investigational agents; and, in the case of **DarwinOncoTreat™**, the full, expansive corral of master regulator proteins and a matched suite of therapeutic agents capable of reversing their entire abnormal signatures, either individually or in combination, is generated.

The third platform, **DarwinOncoDiscovery™**, is a fully integrated biodiscovery and pharmanalytics program designed to ascertain, predict, and characterize proprietary pipelines, from lead compounds to FDA-approved drugs. Individual, tightly integrated components within this platform assess compound potential within novel therapeutic areas (*repurposing*), aid in the design of clinical studies through deployment of robust biomarker to identify responders in the population (*stratification*), assess genome wide compound mechanism of action (*MoA*), and support first-principle based development of combination therapy (*drug synergy*). This is accomplished through the interlinked, seamless integration of computational and experimental methodologies, supporting full pre-clinical characterization of compounds and compound combinations (see below).

**DarwinOncoTarget™** and **DarwinOncoTreat™** support, for the first time, the systematic identification of



## MISSION STATEMENT and EXECUTIVE SUMMARY

druggable tumor checkpoints, on an *individual patient basis*. Furthermore, they provide a repertoire of therapeutic agents assessed as the optimal tumor-specific inhibitors for these dependencies, including individual drugs and synergistic drug combinations. Critically, these technologies require no costly tumor DNA sequencing and can be applied to minimal amounts of fresh-frozen or FFPE tissue down to single cells harvested from individual cancer patients. This allows systematic matching of cancer patients to a growing repertoire of targeted therapeutic agents, even in the absence of actionable mutations and supports a new paradigm for drug development and repositioning. As shown by recent N-of-1 clinical studies at Columbia University, these tumor checkpoint-based approaches implemented by **DarwinHealth™** can bring targeted therapy to *every* cancer patient.

More specifically, **DarwinOncoTarget™** assesses whether any known targets of FDA-approved and experimental inhibitors are abnormally activated in a specific tumor, thus establishing the rationale for delivering targeted therapy to a broad universe of patients and across the cancer continuum, regardless of the stage or degree of tumor invasiveness. Specifically, this analytical approach is applicable across the full natural history of cancer progression, from cancer predisposition to diagnosis, remission, relapse to drug resistant disease, metastatic progression, and/or cure in an individual patient. Delivery of this technology at major partnering cancer centers will complement and extend current genomic profiling approaches, quadrupling the number of patients that may benefit from targeted therapy, at a fraction of DNA profiling costs.

In contrast, **DarwinOncoTreat™** is designed to identify FDA-approved and experimental compounds that target *the full repertoire of tumor dependencies identified by the DarwinCheckPoint™ methodology*, either individually or in combination. We have shown that compounds and combinations that abrogate the full signature of MR protein activity in a specific cancer induce substantial tumor regression *in vivo*, leading to complete response or stable disease.

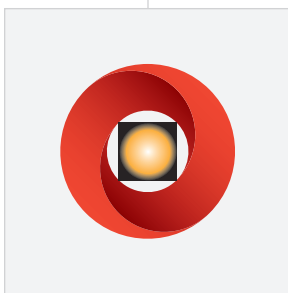
It is well documented that cancer patients, in particular, are increasingly taking control of their disease and want to be directly informed—and take advantage—of potential therapeutic strategies that are at the forefront of cancer medicine. Their search for individualized care at the extremes of scientific, technological, and clinical capabilities has led a growing number of patients to Cancer

Centers of Excellence, where specialized teams and research centers offer the option for enrolling in therapeutic trials that may produce significant benefits where other standardized approaches have failed or produced suboptimal outcomes.

**DarwinHealth™** has responded to this burgeoning movement by offering a service called, **DarwinOncoMe™**, which addresses directly the need to more precisely delineate tumor characteristics—including their pathological, behavioral, and molecular sensitivity profiles. This level of specificity and customization of cancer care is acutely relevant and takes on unique importance in patients with rare, rapidly progressive, or untreatable malignancies in which initial analysis of tumor tissues reveals the absence of any actionable mutations that might establish a presumptive roadmap for the selection of appropriate therapeutic options, either established or investigational. These challenges, and the accompanying need for personalized and customized strategies, address an identified unmet need for a broad range of malignancies on the cancer continuum, from metastatic gastroenteropancreatic neuroendocrine tumors (GEP-NETs) to glioblastoma, to most tumors with RAS pathway mutations, which fail to achieve long-term remission following routinely available therapeutic strategies.

To address this critical deficiency on the therapeutic landscape of cancer, **DarwinHealth™** principals have designed individualized studies for patients who can afford a personal—and, in some cases, sizable—out-of-pocket investment in the assessment and treatment of their malignancy. The key expectation of this “concierge service” for cancer management is that successful treatment of these landmark patients will open the door to inexpensive, broadly available therapies for a large community of individuals with poorly studied or otherwise lethal tumors. This population of cancer patients will be able take advantage of the discoveries made for, and underwritten by the landmark patients, including access to existing and novel therapeutic agents.

**DarwinOncoMe™** is articulated around a complex, proven process that starts with the harvesting of patient-derived tumor tissue. The tissue is subsequently prepared and processed to create appropriate models for testing candidate therapeutic options, including patient-derived cell (PDCL) lines, patient-derived xenografts (PDX), and organotypic cultures (explants). The same tissue is then profiled by RNA Sequencing for analysis using the **DarwinOncoTreat™** methodology. This helps prioritize a





## MISSION STATEMENT and EXECUTIVE SUMMARY

list of FDA-approved and experimental drugs that, either individually or in combination, are most likely to reverse activity of the specific tumor checkpoints identified by the **DarwinCheckPoint™** analysis. The most promising drugs and drug combinations emerging from these analyses are then tested in PDX models to assess their ability to abrogate tumorigenesis **in vivo**.

Of special note is the observation that the evolution of a tumor—especially aggressive, high-grade malignancies—in an individual patient is characterized by unpredictable trajectories over time. This is manifested by variable degrees of invasiveness, drug-resistance, inconsistent relapse profiles, and fluctuating lengths of cancer remission. In actionable, precision cancer medicine terms, this means that therapies directed against an evolving malignancy must uniquely and precisely target those progressive, sequential cellular derangements—in particular, the time-linked shifts in oncogenic **TumorCheckPoint** Signatures—that are triggering dynamic changes in tumor behavior, resistance, invasiveness, and metastatic spread. Thus, cancer-directed intervention will need to unfold in precise, but flexible and nimble sequences that respond dynamically to both incremental and significant shifts in protein-driven programs that perpetuate malignant behavior and tumor evolution in a patient.

This longitudinal, probe-and-pinpoint- and protein pick-off capability, which has been incorporated into the **DarwinOncoMet™** and **DarwinOncoRelapse™** treatment analysis, emphasizes just this kind of dynamic assessment and re-assessment of oncoproteomic signatures in tumor samples over time. This is the most reliable way to ensure that with each new shift in the suite of tumor drivers orchestrating malignant evolution in a specific host, a precise and optimal alignment between these newly differentiated targets and effective therapies can be formulated.

It should be stressed that only the **DarwinOncoTarget™** and **DarwinOncoTreat™** methodologies offer the multi-temporal responsiveness to tumor evolution and aggressiveness, since it is rare that novel druggable mutations may emerge over the lifespan of a tumor. Indeed, while DNA-based tumor evolution analysis frequently provides the rationale for therapeutic failure (e.g. emerging KRAS or p53 mutations), it only infrequently provides additional rationale for therapy refinement. Therefore, mutational analysis provides scant, if any, additional actionable information about how to address progressive changes in tumor invasiveness, the expansion of aggressive

metastatic sub-clones, or treatment failure and clinical relapses due to drug resistance. The **DarwinOncoMet™** and **DarwinOncoRelapse™** treatment analysis developed by **DarwinHealth™** will address the pressing need to provide precise assessment and treatment-optimizing clinical interventions for patients with advanced cancers shown resistant to standard therapies and protocols, as shown in several publications.

The **DarwinOncoDiscovery™** platform, based on the **DarwinOncoMatch™** technology that aligns compounds with aberrantly activated master regulators, is aimed at the urgent and unmet need among biopharmaceutical companies to identify, target, and ultimately disable the complex oncogenic programs driving cellular transformation and tumorigenesis using the repertoire of proprietary compounds at their disposal. When underwritten by pharmaceutical companies, deployment of proprietary and foundational **DarwinOncoMatch™** algorithms will not only optimize therapeutic options for individual patients but will provide an unprecedented roadmap of actionable insights for drug development and repositioning, from genome-wide elucidation of mechanism of action, to toxicity profile analysis, to development of biomarkers for stratification of responder populations, to the *in vivo* validation of predicted tumor-target-therapy alignments.

When required, these methods also can be adapted to focus on the targeting profiles and anti-cancer activity of a single approved cancer agent. By providing a precise and comprehensive analysis of where, why, and in what tumor types a singular agent will likely be active, the **DarwinOncoDiscovery™** platform will define a productive label extension trajectory and clinical development roadmap for biopharmaceutical customers seeking to expand indications for, and maximize monetization potential of their in house anti-cancer agents already marketed for narrow indications.

By facilitating the prioritization and repurposing of proprietary compound pipelines for clinical development—either by improved characterization of tumor-specific drug mechanism(s) of action, identifying novel therapeutic opportunities for existing drugs, including molecular biomarkers for patient stratification, and/or prioritizing sensitizer compounds to rescue tumor sensitivity and augment clinical responsiveness following patient relapse—the **DarwinOncoDiscovery™** platform offers an accelerated, precision- and proteomic-focused approach for discovering new cancer therapies with



## MISSION STATEMENT and EXECUTIVE SUMMARY

uncanny accuracy. It does so by predicting and validating the full and robust range of therapeutic opportunities that are possible—but which, without this discovery platform, would remain invisible—within the existing and near-future armamentarium of anti-cancer therapies.

Representing hundreds of man-years of complex and breakthrough research and development, as well as hundreds of successful scientific collaborations between **DarwinHealth™** co-founders and top laboratories across the world, these *in silico* methodologies, combined with *in vitro* cellular perturbation studies and *in vivo* validation assays, precisely pinpoint novel master regulator proteins that govern cancer cell dysregulation as a result of mutations in other genes and represent ideal therapeutic targets. Using this foundational technology, **DarwinOncoDiscovery™** will illuminate previously undiscovered, even unanticipated and novel match-ups among a vast inventory of FDA-approved and investigational compounds and specific tumor cell lines, against which these agents, either as monotherapy, or in synergistic combination regimens predicted by **DarwinOncoMatch™**, will deliver anti-cancer activity at a level of precision, personalization, and effectiveness not currently achievable with conventional methods.

Researchers, oncologists, and leading cancer centers across the world have been waiting for the day when the “bad” macromolecules, wreaking havoc in cancer’s cellular black box, could be reliably and systematically illuminated, pinpointed, and predictably disabled. Myriad trajectories have spun out to achieve this elusive goal, but none of these paths has proven as direct and fertile as the one carefully hypothesized and thoroughly validated by **DarwinHealth™** co-founders, which now informs the technological infrastructure of **DarwinHealth™**. These statements are factually supported by a wealth of highly cited publications in prestigious, high-impact scientific and medical journals—from *Nature*, *Nature Genetics*, *Cell*, and *Cancer Cell*, to *Nature Biotechnology*, *Genes & Development*, *Molecular Systems Biology*, and *Genome Research*, among others—and by their presentation as keynote lectures to scientific and clinical audiences at major international meetings focusing on cancer research and precision medicine.

The integrated suite of **DarwinHealth™** technologies—focused on a broad base of distinct customer groups, among them patients, oncologists, biopharmaceutical companies, hospitals, national health systems, cancer centers—will dramatically and immediately extend the potential applicability, value and precision potential of our current therapeutic arsenal for cancer to the front lines of patient care. Taken together, these opportunities represent

a unique business proposition whose methodologies are expected to emerge as a dominant paradigm for both health and disease management over the next decade and beyond.

With 1.7 million new cases of cancer, and almost 600,000 cancer-related deaths in the U.S. estimated by the American Cancer Society for the year 2015 alone—and with more than 14 million *new cases* of cancer reported worldwide by the Cancer Research Institute in the U.K.—the need for innovative and validated strategies to treat cancer has never been greater; and, this need will only continue to grow. **DarwinHealth™** has been launched to address these global health needs and to respond to national and international mandates that envision, that are being aggressively supported, and increasingly, that will demand and require more precise patient-and-pathsospecific interventions to optimize patient survival and response rates to cancer.

Merging the technology and IP foundations developed at Columbia University with the translational, product formulation, partnering, and commercialization capabilities of **DarwinHealth™** represents a singular, leapfrogging step forward in bringing targeted, precision-focused therapeutics to a global universe consisting of millions of individuals afflicted with cancer.

With these scientific, technological, public health and clinical issues in clear focus, it is the mission of **DarwinHealth™** to use novel, highly sophisticated, and experimentally validated computational methodologies and clinical tools to translate our evolved understanding of cancer’s intra- and inter-cellular trelliswork to real world, day-to-day decision-making and therapy for patients with cancer. Working within the rigorous, tested framework of a proprietary, cancer-positioning system (CPS), **DarwinHealth™** is poised to apply the proven, essential capabilities of *in silico*-based analysis and wisdom to optimize clinical outcomes in patients worldwide suffering from this grand-challenge disease.

**DarwinHealth™: Precision Therapeutics for Cancer Medicine** is committed to delivering a molecularly precise, analytically rigorous approach to cancer treatment, where other methodologies have faltered; to find actionable targets and precise treatments for millions of individual cancer patients who, at the present time, do not have access to the tumor-target-therapy alignments that **DarwinHealth™** technology makes possible.

Ours is a meticulously plotted journey from disruption and innovation to information and precision-driven application at the front lines of cancer care. Like so many other disciplines—among them economics, chemistry, meteorology,

## MISSION STATEMENT and EXECUTIVE SUMMARY

computer science, and physics—that have applied rigorous quantitative methods to produce stunning products with enormous commercial impact and value, **DarwinHealth™** will employ numbers, edge advances in bioscience, and analytics to turn the dark matter of cancer genomes into light, so that novel therapeutic strategies can be illuminated and applied with confidence and precision where it matters most—in the individual patient with cancer. The trajectory

can be stated simply: There will be paradigms lost, paradigms gained and—most importantly—patients cured.

**Andrea Califano, PhD**

**Gideon Bosker, MD**

**Co-Founders, DarwinHealth™: Precision Therapeutics for Cancer Medicine**

Selected Publications by **DarwinHealth™** Co-Founder Andrea Califano, PhD and **DarwinHealth™** Chief Scientific, Technology and Innovations Officer, Mariano Alvarez, PhD

**—Foundational Methodologies, Computational Tools, and Scientific Discoveries Informing the DarwinHealth™ Product Line and Analytical Platforms for Precision-Based Cancer Medicine**

### REVERSE ENGINEERING:

1. Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nat Genet* 37, 382-390.
2. Wang, K., Saito, M., Bisikirska, B.C., Alvarez, M.J., Lim, W.K., Rajbhandari, P., Shen, Q., Nemenman, I., Basso, K., Margolin, A.A., et al. (2009). Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat Biotechnol* 27, 829-839.
3. Zhang, Q.C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C.A., Bisikirska, B., Lefebvre, C., Accili, D., Hunter, T., et al. (2012). Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* 490, 556-560.

### MASTER REGULATORS OF TUMOR INITIATION AND PROGRESSION:

4. Compagno, M., Lim, W.K., Grunn, A., Nandula, S.V., Brahmachary, M., Shen, Q., Bertoni, F., Ponzoni, M., Scandurra, M., Califano, A., et al. (2009). Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* 459, 717-721.
5. Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H., et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 463, 318-325.
6. Della Gatta, G., Palomero, T., Perez-Garcia, A., Ambesi-Impiombato, A., Bansal, M., Carpenter, Z.W., De Keersmaecker, K., Sole, X., Xu, L., Paietta, E., et al. (2012). Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL. *Nat Med* 18, 436-440.
7. Aytes, A., Mitrofanova, A., Lefebvre, C., Alvarez, M.J., Castillo-Martin, M., Zheng, T., Eastham, J.A., Gopalan, A., Pienta, K.J., Shen, M.M., et al. (2014). Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* 25, 638-651.
8. Alvarez, M.J., Shen, Y., Giorgi, F., Lachmann, A., Ding, B.B., Ye, B.H., and Califano, A. (2015). Network-based inference of protein activity disentangles the mutational and drug-response landscape of cancer. *Nat Genet* (in review).

### MASTER REGULATORS OF DRUG RESISTANCE:

9. Piovan, E., Yu, J., Tosello, V., Herranz, D., Ambesi-Impiombato, A., Da Silva, A.C., Sanchez-Martin, M., Perez-Garcia, A., Rigo, I., Castillo, M., et al. (2013). Direct reversal of glucocorticoid resistance by AKT inhibition in acute lymphoblastic leukemia. *Cancer Cell* 24, 766-776.
10. Rodriguez-Barrueco, R., Yu, J., Llobet-Navas, D., Paula-Saucedo, L., Castillo-Martin, M., Cordon-Cardo, C., Maurer, M., Kalinsky, K., Califano, A., and Silva, J.M. (2015). Inhibition of the autocrine loop IL6-JAK2-STAT3-Calprotectin as targeted therapy for HR-/HER2+ breast cancers. *Genes Dev*, in press

## MISSION STATEMENT and EXECUTIVE SUMMARY

### DRUG MECHANISM OF ACTION AND SYNERGY:

11. Bansal, M., Yang, J., Karan, C., Menden, M.P., Costello, J.C., Tang, H., Xiao, G., Li, Y., Allen, J., Zhong, R., et al. (2014). A community computational challenge to predict the activity of pairs of compounds. *Nat Biotechnol* 32, 1213-1222.
12. Woo, J.H., Shimoni, Y., Yang, W.S., Subramaniam, P., Iyer, A., Nicoletti, P., Rodriguez Martinez, M., Lopez, G., Mattioli, M., Realubit, R., et al. (2015). Elucidating Compound Mechanism of Action by Network Dysregulation Analysis in Perturbed Cells. *Cell* 162, 441-451.

### COMPLEXITY OF CANCER GENETICS:

13. Califano, A., Butte, A.J., Friend, S., Ideker, T., and Schadt, E. (2012). Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nat Genet* 44, 841-847.
14. Chen, J.C., Alvarez, M.J., Talos, F., Dhruv, H., Rieckhof, G.E., Iyer, A., Diefes, K.L., Aldape, K., Berens, M., Shen, M.M., et al. (2014). Identification of Causal Genetic Drivers of Human Disease through Systems-Level Analysis of Regulatory Networks. *Cell* 159, 402-414.
15. Sumazin, P., Yang, X., Chiu, H.S., Chung, W.J., Iyer, A., Llobet-Navas, D., Rajbhandari, P., Bansal, M., Guarnieri, P., Silva, J., et al. (2011). An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell* 147, 370-381.









## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

### INTRODUCTION TO DARWINHEALTH™ TECHNOLOGY AND PRODUCT PORTFOLIO

This document introduces the key elements of the **DarwinOncoSphere™** ecosystem of interconnected products, technologies and deliverables for precision cancer medicine. The three product lines include:

- (1) The service line (**DarwinOncoTarget™** and **DarwinOncoTreat™**) dedicated to supporting therapeutic decision-making by oncologists and related specialists within large, influential, and magnet organizations—cancer referral centers, Cancer Centers of Excellence (CCoE), regional cancer centers, large healthcare systems, cancer research organizations, and large cancer clinics—dedicated to cancer prevention, consultation, and treatment;
- (2) A personalized “oncology concierge service” (**DarwinOncoMe™**) providing precision-based tumor analysis, customized cancer treatment recommendations, and targeted management strategies for patients with rare, poorly responsive, or untreatable tumors, and;
- (3) **DarwinHealth™**’s suite of drug-focused discovery platforms (**DarwinOncoDiscovery™**) and pharmanalytic products supporting corporate clinical development for the biopharmaceutical and biotechnology industry; with a focus on identifying unique and novel therapeutic and commercial trajectories for established and investigational proprietary compounds, including their repurposing and/or label extensions across the continuum of cancer care;

Each of these products—as well as the universe of unique and interlinked proprietary databases (**DarwinOncoBase™**) that support their implementation—are based on breakthrough technologies, processes, and algorithms (**DarwinCheckPoint™**, **DarwinOncoMatch™**, and **DarwinOncoSynergy™**) that, in aggregate, are internally consistent, share common interfaces, and supply the foundational rafters and girders that underpin all **DarwinHeath™** products.

Finally, we describe a suite of cancer- and precision medicine-focused databases (**DarwinOncoBase™**) that consolidates patient tumor data and the results of drug perturbation assays conducted with hundreds of cancer-approved and investigational agents, thereby comprising *the largest and most comprehensive tumor target/targeted therapy alignment database in the cancer space.*

Because data emanating from these in-house discovery and analytical methods represent an integral, continuously evolving—and living treasury—of cancer-focused, tumor therapy alignments across the malignancy continuum, **DarwinOncoBase™** represents a long-term strategic asset that can be deployed across multiple product lines; and, as important, it can be used to enhance the precision and actionability of **DarwinHeath™** deliverables for a broad range of customer segments.



DARWIN|OncoTarget™



DARWIN|OncoTreat™



DARWIN|OncoMe™



DARWIN|OncoDiscovery™

## DARWINHEALTH™ PRODUCT LINE



### Assessing Aberrant Activity of Established Oncogenes and Master Regulator Proteins

A key advance in personalized treatment of the individual patient with cancer has been the discovery that some oncogene mutations (e.g., EGFR amplifications in lung adenocarcinoma) induce dependency of the cancer cell on the aberrant activity of the corresponding protein. This paradigm, best known by the term *oncogene addiction*<sup>1</sup>, has rapidly become the foundation of modern approaches to precision cancer medicine and has led to development of such highly successful oncogene-targeting drugs as imatinib (BCR/ABL) and trastuzumab (HER2).

Most oncologists are thoroughly familiar with this paradigm and routinely rely on tumor DNA mutational profiles to prioritize clinical use of targeted inhibitors. This approach presents two significant advantages: First, mutational analysis has become economically feasible and ubiquitously available, at a cost of approximately \$3000 to \$5000 per patient; and second, interpretation of mutational data for established oncogene mutations is straightforward and highly reproducible.

Although DNA-based mutational profiles have wedged open a small, but significant clinical channel for cancer precision medicine, a plethora of scientific and medical studies point to important limitations of this approach. Not surprisingly, questions about the widespread applicability of this strategy to the front lines of cancer care—and the inherent limits of its precision—have started to emerge. First and foremost, only about 25% of adult malignancies present with oncogene mutations that are directly actionable (i.e., can be targeted pharmacologically). Indeed, direct mutation of an *oncogene* is only one, out of a staggeringly large number of ways by which the corresponding *oncoprotein*—i.e., the functional and operational interface responsible for cellular dysregulation, resulting in tumorigenesis, progression, and drug sensitivity—may become aberrantly activated.

Put another way, upon evaluation of a cancer patient, it may be determined that a tumor-specific DNA mutation exists and that it *may* be therapeutically actionable. Unfortunately, however, this is the case only in a minority of patients with cancer. This is because the presence of an

oncogenic mutation, in and of itself, is *neither necessary nor sufficient* to induce tumor initiation and progression.

Indeed, many tumors that depend on the aberrant activity of a specific oncogene present with no mutations in the corresponding gene. The relationship, perhaps, is best described as partially causal and, therefore, is problematic in yielding definitive clinical action points on a predictable basis across a broad swath of cancer patients. *In more quantitative and predictive terms, the resolutive capacity of DNA mutational analysis, as articulated by the oncogene addiction model, is simply not sufficient to direct therapeutic decision-making in up to 75% of all cancers.* In addition, a significant fraction of patients harboring oncogene-activating mutations fail to respond to targeted inhibitors, while most of those who initially do respond eventually relapse with drug-resistant tumors, suggesting other factors and regulating checkpoints may be operative.

In stark contrast, **DarwinHealth™** has developed and validated methodologies to directly identify and pharmacologically inhibit the oncoproteins that are aberrantly activated in cancer (i.e., over- or under-activated compared to normal tissue)—both individually and especially when working in concert as an oncogenic program—independent of their mutational state. This provides a much more direct and precision therapy-focused roadmap for treating the vast majority of tumors.

Critically, such an oncoprotein-directed approach to targeted, therapeutic intervention is highly effective both in the initial treatment of a tumor at the time of its presentation, as well as in the therapeutic management of tumors undergoing metastatic progression or relapse.

It should be noted that, almost without exception, such behavioral manifestations signaling tumor resistance and aggressiveness *will* be accompanied by a modified repertoire of *aberrantly activated proteins*; however, these advanced tumors undergoing “malignant evolution” generally *will lack additional targetable (DNA) mutations* to guide progressive treatment decisions. The lack of new actionable mutated targets that accompany tumor evolution in vivo undermines any reliable, prescriptive approach that might otherwise be possible using DNA mutational analysis. Indeed, mutational studies may generally help us understand why a cancer evolved to a drug-resistant state but it is only infrequently that they may provide us with novel actionable targeted for therapeutic intervention.

## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

In contrast, the oncoprotein-focused methodology developed by **DarwinHealth™** is especially sensitive to a cancer's natural history, in that it allows clinicians to make dynamic, “analyze-and-target-and-treat” interventions across the entire temporal and morbidity continuum of tumor behavior in the human host.

To achieve its goal of making precision therapeutics possible in every patient and to address the limitations of strictly mutation-dependent cancer therapeutics, **DarwinHealth™** is utilizing a seminal and groundbreaking proprietary methodology (**DarwinOncoTarget™**) to systematically detect and assess the full repertoire of aberrantly active and pharmacologically actionable proteins in a tumor, *independent of the tumor's DNA mutational state*. This critical advance in cancer therapeutics is achieved by a complex systems biology analysis of the mRNA profile data of either fresh frozen (FF) or formalin-fixed paraffin-embedded (FFPE) tumor samples, as obtained by RNA sequencing methodologies (RNA-Seq). This yields an approach that produces unprecedented reproducibility and accuracy in the assessment of aberrantly activated oncoproteins. Moreover, use of RNA profiles provides major savings, compared to exome or genome-based sequencing approaches; and, importantly, requires much less tissue for analysis. (See Figure 1.)

In summary, mutations in an oncogene are only one of a virtually infinite number of ways to abnormally activate their associated oncoproteins, whose aberrant activity and direct regulatory control of the tumorigenic state, more often than not, depends on a confluence of factors independent of their direct mutational status. *As a result, a new paradigm introduced by DarwinHealth™ for pinpointing and pharmacologically targeting aberrantly activated oncoproteins in virtually all cancer patients represents the most disruptive, innovational, and productive frontier on the current landscape of personalized, tumor-specific cancer medicine.*

### Technology Underpinnings of DarwinOncoTarget™

**DarwinHealth™** intends to establish, improve upon, and deploy these disruptive, post- “oncogene addiction” paradigms to revolutionize precision-based therapeutics at the front lines of patient care. Using proprietary technology based on the VIPER algorithm<sup>2, 3, 4</sup>, a radical extension of the Master Regulator Inference algorithm (MARINA), **DarwinOncoTarget™** *generates a complete report of the aberrant activity of established oncoproteins, including those for which targeted inhibitors are available, resulting in the identification of multiple candidate (for drug therapy) targets and associated FDA-approved or experimental inhibitors, for virtually every cancer patient.*

As shown by an extensive body of published work, VIPER assesses activity of targetable oncoproteins based on the expression of their tumor-specific transcriptional targets, which dramatically increases assessment reproducibility and sensitivity. Specifically, ***VIPER-based inference—a computational model—of protein activity dramatically outperforms mutational analysis, gene expression analysis, and protein and phosphoprotein abundance analysis based on Reverse Phase Protein Arrays (RPPA) technology in predicting targeted inhibitor response*** (e.g., erlotinib for aberrant EGFR activity in lung adenocarcinoma cells, as assessed by these methods) (See Figure 2.)

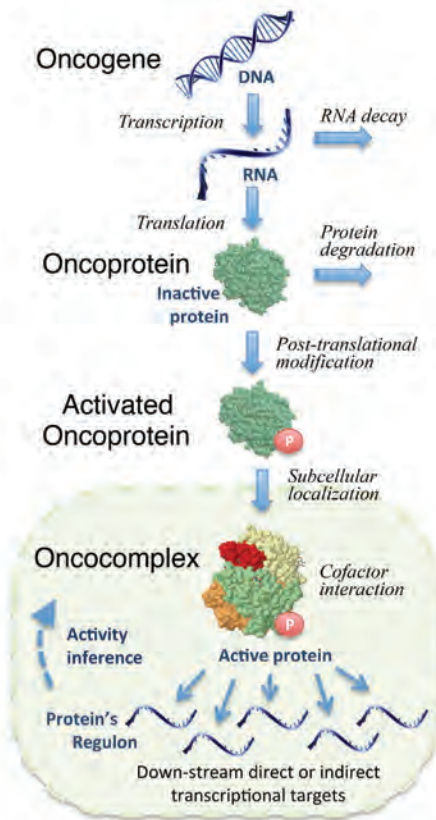
**DarwinOncoTarget™** analyses can also be combined with mutational data, if available, thereby further increasing prediction confidence. For instance, the **DarwinOncoTargetReport™** that accompanies each tumor- and patient-specific analysis will indicate whether a previously unreported mutation in an established oncoprotein likely represents a neutral (no activity change), gain of function (aberrant increase in oncoprotein activity), or loss of function (aberrant decrease in oncoprotein activity) mutation, thus allowing optimal patient data management.

**Competitive Advantages:** **DarwinOncoTarget™** provides additional, therapeutically actionable information on >6,000 proteins, including >500 established oncoproteins, to help oncologists design optimal therapeutic approaches on an individual patient basis; (See Table 1.)

Critical readouts that introduce a level of precision currently unmatched by and distinct from any other methodologies include:

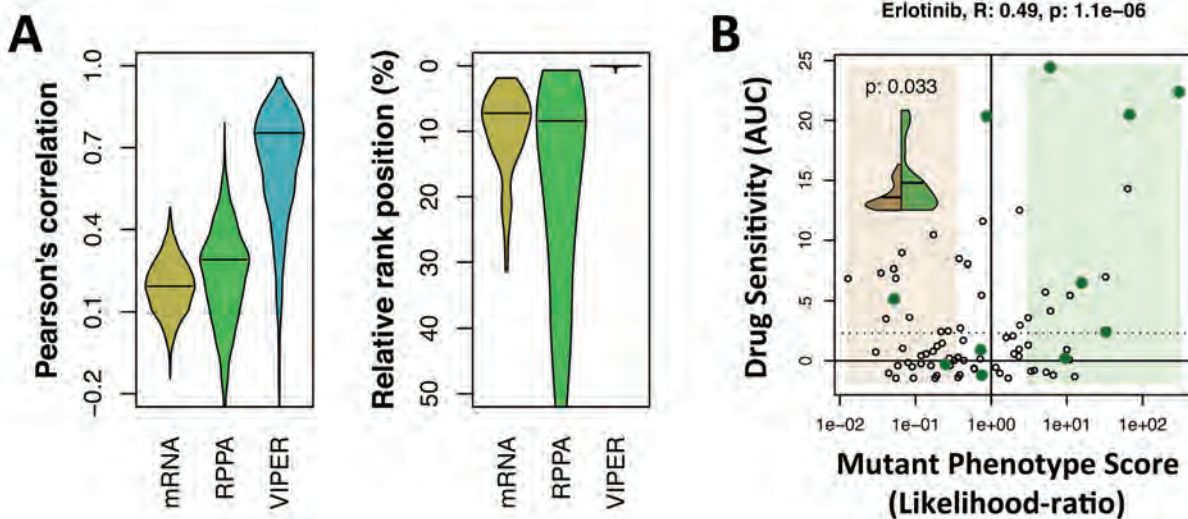
- (1) Generating oncoprotein activity profiles and associated tumor-specific inhibitors across multiple tumor sites (e.g., multiple metastatic sites);
- (2) Assessing aberrant activity of proteins that may induce relapse or resistance to first or second line treatment and associated tumor-specific inhibitors;
- (3) Identifying candidate tumor-specific inhibitors of aberrantly activated master regulators *de novo*, even though these may not be known or reported in the literature and;
- (4) Allowing precise and highly reproducible assessment of aberrant protein activity from FFPE samples. (See Figure 3.)





**Figure 1.**

Measuring protein activity is critical for understanding dysregulation of cancer cells, so that master regulators, which are necessary and/or sufficient for tumor initiation/progression, can be identified.



**Figure 2.**

(A) Reproducibility of VIPER results compared to mRNA differential expression and Protein abundance as assessed by Reverse Phase Protein Array (RPPA) technology. The left graph shows the correlation between triple negative breast cancer samples, compared to luminal samples, based on the differential expression, protein abundance, and VIPER activity. On the right, we show the distribution of the top 10 most differentially expressed genes, abundant protein, and VIPER-activity proteins in one triple negative sample when assessed in all other samples in the TCGA cohort. It is clearly visible that the protein assessed to be the most aberrantly activated in one sample are conserved across all the patients in the subtype. In contrast, differential expression and different protein abundance are very poorly conserved.

(B) VIPER activity (x-axis) is a much better predictor of erlotinib activity (y-axis) than mutational status of 80 lung cancer cell lines from the cancer cell line encyclopedia ( $p = 1.1 \times 10^{-6}$  vs.  $p = 3 \times 10^{-2}$ ).

## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

Unlike any available methodology—or, to the knowledge of **DarwinHealth™**, technology that is even under development—the systematic tumor-site specific assessment of aberrantly activated proteins, over the time-dependent evolutionary trajectory of a patient-specific tumor, represents a powerful new framework for optimizing cancer treatment based on precise alignment of tumor-dependencies with available targeted inhibitors.

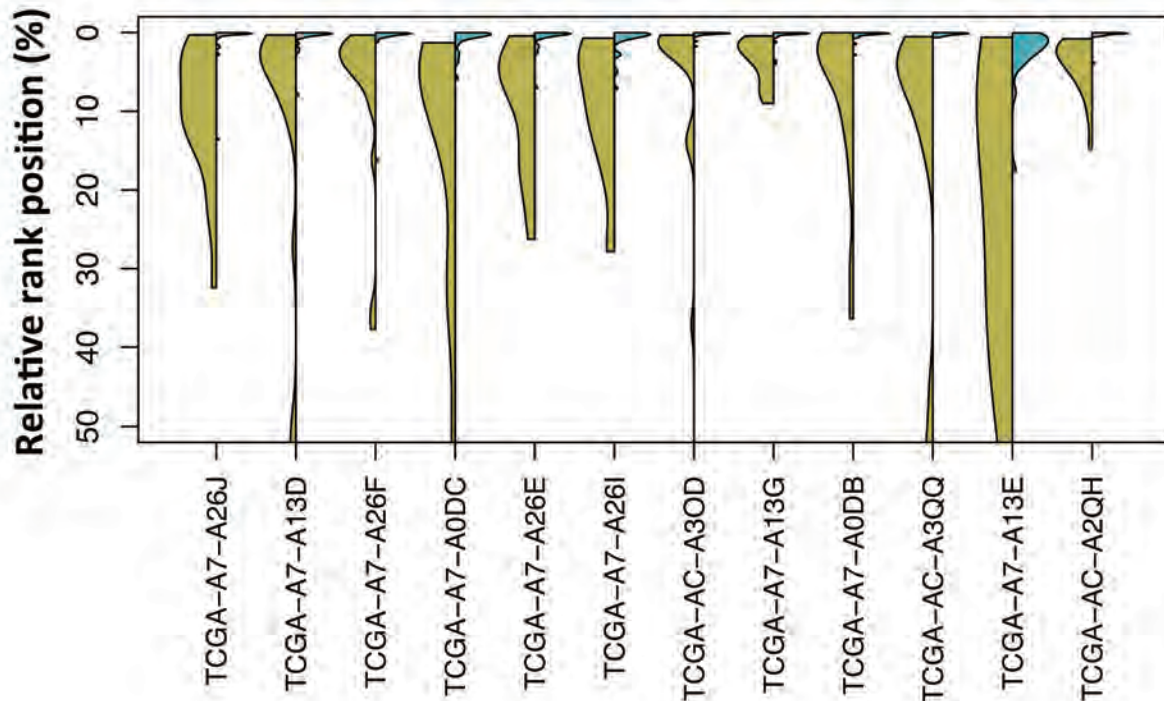
As indicated in points (1)–(4) above, the analytical capabilities of **DarwinOncoTarget™** will extend from the time a tumor is initially discovered over the course of its natural history, which may include recurrence, long periods of remission, declining responsiveness, treatment failure, and metastatic progression.

For the first analysis (1) above, **DarwinOncoTarget™** generates an independent report for each patient-specific tumor biopsy or resection. This initial, baseline report can be used to guide initial therapeutic choices and, equally important, it can be used over time to modify treatment approaches and fine tune the drug selection calculus according to changes in the tumor's oncoprotein activity profile over time.

For the second assessment (2) above, we identify treatment vulnerabilities linked to drugs associated with relapse and resistance. In this regard, as just one example, we have shown that resistance emerging from the use of targeted agents, including trastuzumab in HER2+ breast cancer and glucocorticoids in T-ALL, leads to identification of key actionable targets that restore sensitivity via combination therapy<sup>5,6</sup>.

For the third stage of precision-driven guidance, (3) above, we leverage extensive proprietary databases generated by molecular perturbation studies that identify tumor-specific protein inhibitors, using the **DarwinOnco-Match™** algorithm (See Section Below, **DarwinHealth™ Proprietary Technologies**).

Finally, the ability to perform these analyses with unaffected accuracy from FFPE sections, (4) above, represents a dramatic improvement in the ability to identify therapeutic options for the majority of patients whose samples are not preserved as fresh frozen specimens, including those obtained from retrospective clinical studies.



**Figure 3.**

Reproducibility of VIPER results based on 12 TCGA samples for which both FFPE and fresh frozen (FF) tissue was available. The yellow plot represents the distribution of the rank of the top 10 most differentially expressed proteins, as computed from the FF sample, in the in the corresponding FFPE sample. The cyan plot represents the distribution of the top 10 most differentially active proteins, as assessed by VIPER from the FF sample data, in the corresponding FFPE sample. This shows that, while differential expression is not reproducible, VIPER activity is highly reproducible.



**Table 1. COMPARATIVE FEATURES AMONG TESTS USED TO IDENTIFY PRECISE THERAPEUTIC TARGETS FOR CANCER**

Features and Properties	 DARWIN OncoTarget™	Foundation One	Differential Expression	RPPA
Reproducibility	<b>High</b>	High	Low	Medium
Genes or Proteins Assessed	<b>6,000 + All Rearrangements</b>	315 + 28 Arrangements	25,000	~300
Fully Captures Aberrant Oncogene Activity, Independent of Oncogene Mutations	<b>YES</b>	No	No	No
Generates Site Specific Reports of Aberrantly Active Oncogenes and Matched Inhibitors	<b>YES</b>	No	No	No
Assesses Proteins Responsible for Relapse and Matched Inhibitors	<b>YES</b>	No	No	No
Assesses Candidate Inhibitors Even if Currently Unknown or Unknowable	<b>YES</b>	No	No	No
Applicable to Minimal Tissue	<b>YES</b>	No	Yes	No
Applicable to Single Cells	<b>YES</b>	No	Yes	No
Applicable to Fresh Frozen Biopsies	<b>YES</b>	Yes	Yes	Yes
Applicable to FFPE	<b>YES</b>	Partially	Poor	No
Cost	<b>LOW</b>	High	Low	Medium

**Table 1: Key Differentiations of Competitor Technologies for DarwinOncoTarget™**

The **DarwinOncoTarget™ Report** provides clinicians and cancer centers with specific recommendations addressing the most actionable targets, based on the cancer literature and completed clinical studies; as well as a listing of all potentially relevant clinical studies that are open for enrollment and key literature references.



DARWINOncoTreat™

### Targeting the Full Repertoire of Tumor Dependencies

As emphasized in the preceding sections, the **DarwinOncoTarget™** analysis-to-treatment platform for the individual patient represents a dramatic step forward in precision cancer medicine because it can identify aberrantly active oncogenes **independent of their mutational state**. *This extends the reach of targeted therapeutics to virtually all cancer patients, instead of the current ~25% of patients who harbor potentially actionable mutations.*

Despite the significant advance and level of precision that **DarwinOncoTarget™** will introduce into the personalized medicine-focused treatment landscape for cancer, this approach also has limitations because it requires oncologists to think about therapy **one oncogene—or, more precisely, one aberrantly active oncoprotein—at a time**. This is suboptimal because **tumorigenesis requires coordinated dysregulation of multiple proteins overseeing key tumorigenic programs** (i.e., **the hallmarks of cancer**<sup>7</sup>), ranging from programmed cell death avoidance and immunoevasion to increased proliferation and dysregulated cell adhesion. Given the multiplicity of protein-induced derangements that underpin the cancer process, it is not surprising that most patients treated with inhibitors targeting a single oncogene relapse to drug-resistant tumors.

Although the **DarwinOncoTreat™** platform for tumor assessment and treatment (see below for a detailed description of this product line) has clear advantages, there are critical factors and entrenched usage patterns among clinicians that explain why **DarwinOncoTarget™** is poised to play an important and immediate role in the marketplace; and, why this **DarwinHealth™** product will be sustainable as a go-to analytical and theranostics platform for precision therapeutics in cancer for years to come. Specifically: (a) **DarwinOncoTreat™** directly extends the “actionable mutation” paradigm, which focuses on a single oncogene as a clinical trigger for drug selection. This represents an already established, widespread, familiar, and well-accepted rationale for therapeutic decision-making in the cancer space with a wealth of existing targeted inhibitors. Thus, even

though the technology is radically different, conceptual barriers among oncologists and related specialists are likely to be very low for its prompt adoption, especially because **DarwinOncoTarget™** recapitulates the results they are expected to see from mutational analysis; (b) given the ever increasing availability of targeted inhibitors, a technology that expands the single target actionable oncogene model from the current 25% of patients identified based on mutational analysis to the 100% of patients identified by **DarwinOncoTarget™** will be transformational and game-changing; and (c) **DarwinOncoTreat™** requires availability of cancer specific perturbational databases (**DarwinOncoDrugBases™**). These are not yet available for all tumor subtypes, including rare tumors. As a result, while **DarwinHealth™** develops these resources, **DarwinOncoTarget™** will continue to represent the premier technology for the identification of druggable targets on an individual patient basis.

Addressing the challenge to improve upon the actionable mutation paradigm for drug selection in cancer patients, **DarwinOncoTreat™ is the first proven methodology for the identification of individual drugs and drug combinations that target the complete repertoire of regulatory proteins necessary for tumor survival, on a patient-by-patient basis, rather than just individual oncogenes**. In support of this cancer treatment strategy breakthrough, **DarwinHealth™** co-founders have shown that master regulator (MR) proteins necessary for tumor survival are organized in tightly coupled modules (tumor checkpoints), whose aberrant activity is strictly required for tumor cell survival. These tumor checkpoints are much more conserved across patients with similar cancer subtypes—and even across distinct cancer types—than the specific mutations that induce tumorigenesis. (See Figure 4).

Critically, they have also shown that MR proteins can be systematically inferred and validated, based on analysis of tumor specific regulatory networks (**interactomes**). This foundational work underpinning the **DarwinHealth™** enterprise—and, more specifically, its proprietary portfolio of technologies focused on precision therapeutics for cancer medicine—has been published in a series of high impact, disruptive contributions to the scientific, cancer, and medical literature addressing multiple tumor types<sup>4</sup>, including lymphoma<sup>8,9</sup>, leukemia<sup>5,10,11</sup>, glioblastoma<sup>2,12,13</sup>, breast cancer<sup>2,6,14</sup>, and prostate cancer<sup>3,15</sup>, among others.

Of unique scientific and commercial significance is the following: Principals of the **DarwinHealth™** science, technology, and innovation group have shown that genetic or pharmacological inhibition of some MR proteins, either individually (essential MRs) or in combination (synthetic lethal MRs), leads to **predictable collapse of the entire**



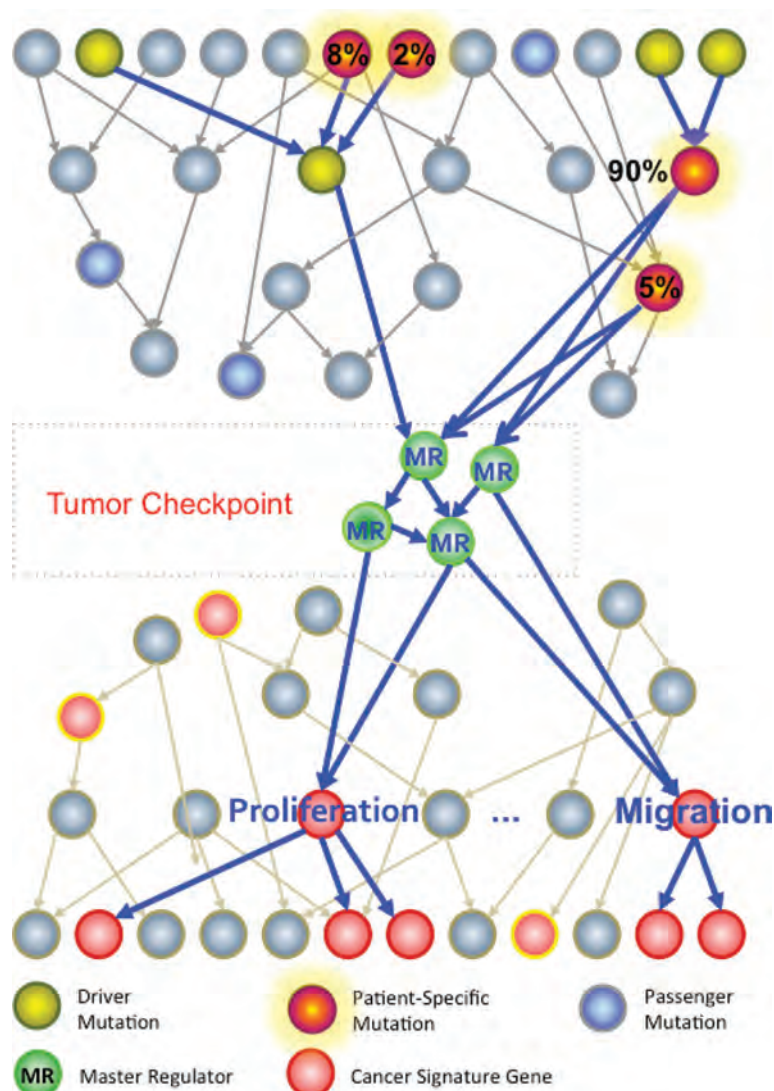
## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

**tumor-checkpoint activity and loss of tumor cell viability both *in vitro* and *in vivo*.** This is producing promising preliminary results in a clinical context, for instance, in HER2+/ER- breast cancer patients who had progressed to trastuzumab-resistant disease. In preliminary results, as predicted by the algorithm, co-inhibition of HER2 and JAK2, using a combination of trastuzumab and ruxolitinib, resulted in profound response and stable disease for both of the first two HER2+/ER- patients enrolled in the clinical trial and produced no response (also as predicted) in an HER2+/ER+ patient.

From both an analytical and treatment perspective, *DarwinOncoTreat™* effectively supplants the concept of *oncogene*

addiction with the more actionable and precise paradigm of *tumor checkpoint* addition, where multiple master regulator (MR) proteins that comprise a *tumor checkpoint* (*DarwinCheckPoint™ signature*) are simultaneously identified and targeted. (See **technology** section for a precise definition of master regulators and tumor checkpoints.)

An important corollary is that **tumor checkpoint activity (i.e., the activity of a tumor's full MR protein repertoire) represents an ideal reporter assay for identifying drugs and drug combinations capable of inducing tumor remission *in vivo***<sup>15,16</sup>. This addresses one of the most critical limitations of current drug screening approaches. Specifically, *in vitro* cell line viability is a very poor predictor of drug



**Figure 4.**

DarwinHelath™'s Tumor Checkpoint activity (i.e., the activity of a tumor's full Master Regulator Protein repertoire) represents an ideal reporter assay for identifying drugs and drug combinations capable of inducing tumor remission *in vivo*.



## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

activity *in vivo*. To overcome this challenge, **DarwinOncoTreat™** replaces viability assays with tumor checkpoint activity assays, whose results are extremely well correlated *in vitro* and *in vivo*. Indeed, a drug's ability to abrogate a target protein's activity is almost always assessed *in vitro* and the outcome is highly predictive of *in vivo* activity. For instance, an EGFR inhibitor will inhibit the receptor both in cell lines and in human tumors.

Leveraging and improving upon breakthrough algorithms developed at Columbia University, **DarwinHealth™** introduces proprietary technology to elucidate MR proteins in tumor checkpoints (**DarwinCheckPoint™**), as well as the individual drugs (**DarwinOncoMatch™**) and synergistic drug combinations (**DarwinOncoSynergy™**) that may abrogate their activity, thus inducing tumor regression *in vivo*.

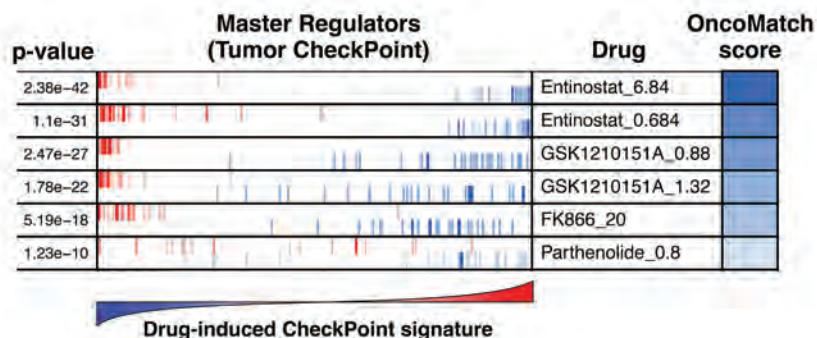
Ultimately, widespread deployment and application of the **DarwinOncoTreat™** platform depends on availability of large-scale, proprietary drug perturbation databases, representing each specific tumor subtype of interest. Specifically, **DarwinHealth™** has already created extensive databases (**DarwinOncoDrugBase™**), representing the response of tumor subtype specific cell lines to FDA-approved and experimental compound perturbations, as measured by genome-wide RNA Seq profiling of perturbed cells at multiple drug concentrations and multiple time points following perturbation (See Technology Section for details and currently available perturbational databases).

**Competitive Advantages of DarwinOncoTreat™.** An unique differentiating element of **DarwinOncoTreat™** is that MR proteins representing either individual or synergistic tumor dependencies are identified strictly—and automatically—from the analysis of patient-derived tissue. As a result, these dependencies are not contaminated by the idiosyncratic biology of a cell line or a mouse model,

such as those that may be identified by functional screens. Indeed, cell lines are used exclusively to characterize the genome-wide mechanism of action (MoA) of FDA-approved and experimental compounds, which is known to be predictive of their *in vivo* MoA.

Similarly, the ability to match pre-screened compounds to the full MR program of a specific patient's tumor, within minutes of receiving the tumor RNA Seq data, represents a remarkable and desperately needed improvement in precision cancer medicine. The MR target-focused approach permits drug-mediated effects to disrupt, shut down, and/or undermine tumor viability mechanistically, across the full range of their molecular and operational dependencies at the oncoproteomic level, rather than by simply selecting a specific onco-gene to target. Systematic, targeting of the entire suite of critical tumor checkpoints—the **DarwinCheckPoint™** Signature—significantly decreases the likelihood of tumor relapse for several reasons, including the ability to counter the effect of alternative mutations before they may start feeding the same tumor checkpoint, leading to sub-clonal selection (See Figure 4).

Much like **DarwinOncoTarget™** discussed above, **DarwinOncoTreat™** (See Table 2) provides critical, additional information to help oncologists tailor their therapeutic approach to the cancer patient, including: (a) the ability to predict distinct drugs and drug combinations across multiple tumor sites (e.g., multiple metastatic sites); and (b) the ability to rescue sensitivity to first or second line treatment by targeting the full complement of proteins that induce resistance. Finally, much like **DarwinOncoTarget™**, **DarwinOncoTreat™** provides oncologists with specific recommendations with respect to drugs and drug combinations that may be available through open clinical studies, as well as any relevant literature references.



**Figure 5:** Results of the **DarwinOncoMatch™** compound prioritization in a **DarwinOncoTreat™** report. Each Box represents the full repertoire of VIPER analyzed proteins, from the one that is mostly inactivated by the drug (on the left) to the one that is most activated by the drug (on the right). Red and blue bars identify proteins that were identified as aberrantly activated and inactivated in a tumor checkpoint of a specific neuroendocrine tumor subtype, respectively. This shows that some drugs are able to implement a complete inversion of tumor checkpoint activity. For instance, the compound entinostat and GSK1210151 dramatically reduced activity of activated master regulator proteins and increased activity of inactivated ones, thus inducing tumor checkpoint collapse. This was validated *in vivo*, where both compounds induced tumor viability reduction in mouse xenografts, with entinostat producing complete tumor regression at 25 days.



**Table 2. COMPARATIVE FEATURES AMONG TESTS USED TO IDENTIFY PRECISE AND COMPREHENSIVE SET OF TUMOR DEPENDENCIES**

Features and Properties	DARWIN OncoTreat™	Foundation One	Differential Expression	RPPA
Reproducibility	<b>High</b>	High	Low	Medium
Tumor Relevance of Inferred Genes/Proteins	<b>Full Set of Tumor Dependencies</b>	Activated Oncogenes	Unclear	Unclear
Identifies Single Drugs and Combinations That Inhibit the Entire Set of Tumor Dependencies	<b>YES</b>	No	No	No
Can Identify Drivers of Tumor Relapses	<b>YES</b>	Unknown	Unknown	Unknown
Applicable to Minimal Tissue	<b>YES</b>	No	Yes	No
Applicable to Single Cells	<b>YES</b>	No	Yes	No
Applicable to Fresh Frozen Biopsies	<b>YES</b>	Yes	Yes	Yes
Applicable to FFPE	<b>YES</b>	Partially	Poor	No
Cost	<b>Low</b>	High	Low	Medium

**Table 1: Key Differentiations of Competitor Technologies for DarwinOncoTreat™**

## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING



### Bringing Precision and Personalized Cancer Care to a Single Patient

It is well documented that cancer patients, in particular, are increasingly taking control of their disease and want to be directly informed—and take advantage—of potential therapeutic strategies that are at the forefront of cancer medicine. Their search for individualized care at the extremes of scientific, technological, and clinical capabilities has led a growing number of patients to Cancer Centers of Excellence, where specialized teams and research centers offer the option for enrolling in therapeutic trials that may produce significant benefits where other standardized approaches have failed or produced suboptimal outcomes.

The need to more precisely delineate tumor characteristics—including their pathological, behavioral, and molecular sensitivity profiles—is acutely relevant and takes on unique importance in rare, rapidly progressive, or untreatable malignancies in which initial analysis of tumor tissues reveals the absence of any kind of actionable mutations that might establish a presumptive roadmap for the selection of appropriate therapeutic options, either established or investigational. These challenges, and the accompanying need for personalized and customized strategies, address an identified unmet need for a broad range of malignancies on the cancer continuum, from metastatic gastroenteropancreatic neuroendocrine tumors (GEP-NETs) to glioblastoma, to most tumors with RAS pathway mutations, which fail to achieve long-term remission following routinely available therapeutic strategies.

To address this critical deficiency in the therapeutic landscape of cancer, **DarwinHealth™** principals have designed individualized studies for patients who can afford a personal—and, in some cases, sizable—out-of-pocket investment in the assessment and treatment of their malignancy. The key expectation is that successful treatment of these landmark patients will open the door to inexpensive, broadly available therapies for a large community of individuals with poorly studied or otherwise lethal tumors. These will be able take advantage of the discoveries made for, and underwritten by the landmark patients, including access to existing and novel therapeutic agents.

**DarwinOncoMe™** (See Table 3) is articulated around a complex, proven process that starts with the harvesting of patient-derived tumor tissue. The tissue is subsequently prepared and processed to create appropriate models for testing candidate therapeutic options, including patient-derived cell (PDCL) lines, patient-derived xenografts (PDX), and organotypic cultures (explants). The same tissue is then profiled by

RNA Sequencing for analysis using the **DarwinOncoTreat™** methodology. This helps prioritize a list of FDA-approved and investigational drugs that, either individually or in combination, are most likely to reverse activity of the specific tumor checkpoints identified by the **DarwinCheckPoint™** analysis.


The most promising drugs and drug combinations emerging from these analyses are then tested in PDX models to assess their ability to abrogate tumorigenesis *in vivo*. Whenever the patient tumor fails to engraft into a PDX, **DarwinHealth™** will either attempt to match the specific patient checkpoint to those of existing PDX models or will use patient explants to test the drugs prioritized by **DarwinOncoTreat™**.

As discussed, this process relies on the availability of comprehensive *drug perturbation databases* (**DarwinOncoDrugBase™**) for each patient-specific tumor subtype. Although **DarwinHealth™** is increasing the number of profiled malignancies—both common and rare—it is likely that in some cases these resources may not have yet been developed. In such cases, **DarwinHealth™** will produce patient-matched perturbational databases for a cost ranging from \$50,000 to \$250,000, depending on the number of cell lines and tested compounds. The resulting perturbational database becomes a universal resource that can then be used to prioritize drugs and drug combinations for all future **DarwinOncoMe™** patients with the same rare or untreatable tumor subtype. As novel inhibitors are developed by pharmaceutical and biotech companies, they can be easily added to the existing databases for a minimum additional investment.

**Competitive Advantages.** A few companies currently exist that provide patient-specific drug assessment using either patient-derived xenografts or patient-derived explants. **However, the critical limitation of their approaches can be summarized as follows:**

- (1) Only a limited number of drugs (typically  $N < 5$ ) can be tested in a mouse model or explant, due to cost and tissue availability; and
- (2) **DarwinHealth™**'s competitors have no proven methodology to prioritize drugs from the full repertoire of FDA-approved and experimental compounds. As a result, they typically start from a small, predefined number of agents that have been used successfully in related tumors. In contrast, **DarwinHealth™ can prioritize hundreds of drugs, and drug combinations, based on mechanistic delineation of tumor checkpoints** and, therefore, provides a dramatically more robust and precise assessment of tumor-target-therapeutic alignments than all other current technologies on the market.

**Table 3.** COMPARATIVE FEATURES AMONG TESTS USED TO IDENTIFY TUMOR-SPECIFIC TARGETS AND THERAPIES IN INDIVIDUAL PATIENTS

Features and Properties	 <b>DARWIN</b> OncoMe™	Champions
Patient Derived Xenograft	YES	Yes
Patient Derived Cell Lines	YES	No
Tumor Growth Assessed by MRI	YES	No
Drugs and Drug Combinations Prioritized by Checkpoint Analysis	YES	No
Concierge Service	YES	Yes
PDX Testing of Drug Combinations	YES	No
Patient Matched to Existing PDX Models if Tumor Transplant Fails	YES	No
Use of Patient Explants if PDX Cannot Be Established or Matched	YES	No

**Table 3:** Key Differentiations of Competitor Technologies for *DarwinOncoMe*™

## DIFFERENTIATING DARWINHEALTH™ TECHNOLOGY AND THERANOSTIC PLATFORMS FROM THE CURRENT ACTIONABLE MUTATION PARADIGM

Advancing and Applying Therapeutic Precision to the Real World, Front Lines of Cancer Management

*What Are the Differentiating Aspects of the **DarwinHealth™** Approach to Precision Therapeutics for Cancer Medicine? How, Why, and in Which Patients is DarwinHealth™ Technology Superior to the Established Oncogenic Paradigm for Cancer Treatment?*

Although **DarwinHealth™** is introducing an entirely novel approach for identifying actionable therapeutic targets—and, in the process, corresponding treatments that are optimally aligned with tumor-specific vulnerabilities—it is important to understand that its methodologies and conceptual approaches are not antithetic to the current oncogene addiction paradigm used in precision cancer medicine. Rather, they largely complement and extend it to make it applicable and relevant to every cancer patient.

Indeed, its baseline technology (**DarwinOncoTarget™**) is driven by the basic tenets that underpin genetics-based cancer medicine, albeit dramatically expanded to identify druggable targets in virtually any tumor. Moreover, when deployed at an even deeper analytical level and informed by systematic cell perturbation studies (**DrawinOncoTreat™** and **DarwinHealthDrugBase™**), these technologies make it possible to identify and target critical tumor checkpoint dependencies that completely elude the established approaches based on tumor mutation analysis. As a result, they consistently yield prescriptive information that far exceeds the precision capacity of all other methods currently available for assessing which therapeutic strategies are most likely to optimize clinical outcomes in cancer patients.

These improvements and advances in precision therapeutics are possible because **DarwinHealth™** technologies directly, systematically, and quantitatively address a critical and yet highly elusive question in cancer biology: that is, *what are the oncoproteins whose concerted aberrant activity is responsible for maintaining the disease state of tumor cells in the human host*. In contrast, current precision medicine strategies approach this problem only indirectly, for instance by studying tumor cells in conditions that do not represent the tumor pathophysiology or by focusing only on the relatively small subset of oncoproteins that harbor activating mutations. Yet, *the mere presence of an activating mutation in the DNA is neither necessary nor sufficient to induce aberrant activity of an oncoprotein* and it is the latter (i.e., the pattern of dysregulated oncoprotein

activities) that is ultimately responsible for inducing tumorigenesis and progression. *Similarly, most targets identified from cell lines or mouse model studies have failed in follow up clinical studies.*

While *all* the technologies that comprise the **DarwinHealth™** portfolio represent clinically actionable advances—and extend the potential benefits of precision medicine to a dramatically expanded segment of the cancer population—it is critically important to differentiate: (a) the capabilities, methodology, and prescriptive outcomes afforded by its two primary products, **DarwinOncoTarget™** and **DarwinOncoTreat™**, and how they compare to each other; and, (b) how the theranostic information and treatment roadmaps emanating from these products, each in their own ways, are distinct from the oncogene addiction model (DNA mutations) that presently informs a significant fraction of clinical decision-making in cancer care.

From a broad and defining perspective, the most critical difference between the current *actionable mutation* paradigm and **DarwinHealth™**'s analytical framework, which identifies tumor checkpoints that represent and are manifestations of critical disease dependencies, is that the **DarwinHealth™** tumor-specific analyses *are performed independent of tumor mutational data*; although, as discussed below, they can benefit from mutational profile data, if available.

This mutation-independent approach addresses two critical constraints that limit the methodologies currently used for precision-focused cancer treatment: (a) **DarwinHealth™** analyses are associated with a five- to ten-fold cost reduction, and; (b) **DarwinHealth™** analyses afford a profound extension of the number of patients who can benefit from genetics-based cancer treatment, from the current benchmark of about 25% of cancer patients who harbor actionable mutations, to 100% of cancer patients who harbor aberrantly activated, druggable oncoproteins—thereby expanding the universe of cancer patients who can benefit from genetics-based, precision medicine by almost four-fold.



## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

### Commonalities and Differences Between Existing and Novel Approaches to Tumor Analytics

How Does DarwinHealth™ Technology Improve Precision of the Drug Selection Equation for Cancer Therapy?

Another approach to clarifying these important distinctions—especially those demonstrating the advantages of the **DarwinHealth™** technology over more traditional *actionable mutation* approaches—is to ask when the pharmacological targets and therapeutic agents identified by these approaches (i.e., **DarwinHealth™** products versus traditional actionable mutation analysis) overlap, and when are they different; as well as whether these analyses are independent of each another or inform each other in some way. To specifically address these points, it is important that we distinguish between the two **DarwinHealth™** products, i.e., **DarwinOncoTarget™** and **DarwinOncoTreat™**.

### Important Advances in and Extensions to the Current Standard (Actionable Mutation Analysis) of Genetics-Based Cancer Treatment



**DarwinOncoTarget™**, generates results that are *most directly comparable to those generated by actionable mutation analyses*, the current standard for tumor-target-therapy alignment. Both methodologies focus on proteins whose aberrant activity is necessary for tumor survival and for which targeted inhibitors may exist in the clinical setting. The difference is that the actionable mutation model identifies these proteins based on the presence of activating mutations in their corresponding genes. In contrast, **DarwinOncoTarget™** *identifies actionable oncoproteins by directly measuring their aberrant activity using a gene reporter assay represented by the expression of their transcriptional targets*.

These differences have important, clinically and therapeutically relevant consequences. Because analyses such as those conducted by **DarwinHealth™** are based on measuring aberrant activity among proteins, two important advantages are observed: (a) the method can assess the aberrant activity of any regulatory protein in the cell (~6,000 in total) independently of whether it is mutated or not; and (b), this method generally identifies aberrantly activated proteins *for which an inhibitor is available in virtually 100% of tumor patients*.

Therefore, for this specific technology the answers to the questions raised above can be summarized as follows:

#### 1. In what ways and in which situations do the pharmacological targets identified by these approaches (i.e., **DarwinOncoTarget™** versus actionable mutation analysis) overlap; and when are the pharmacologic targets identified by these analytical approaches different?

—*The pharmacologic targets identified by these two methods will overlap* when an oncoprotein is aberrantly activated because of one or more activating mutations in the corresponding oncogene.

—*The pharmacologic targets identified by these two methods will be different when either:*

- An oncoprotein is aberrantly activated despite the fact that it does not harbor any activating mutations in the corresponding oncogene; or
- An oncoprotein is not aberrantly activated despite the presence of one or more activating mutations in the corresponding oncogene.

#### 2. Are these analyses independent of each another or do they inform each other in some way?

—The methods and technology employed for these analyses *are completely independent of each other*. However, **DarwinOncoTarget™** can inform the analysis addressing actionable mutations by assessing whether a specific mutation induces aberrant activity of the oncoprotein or not.

#### 3. Will the therapeutic agents prioritized by the **DarwinOncoTarget™** approach be different from those agents dictated by the actionable mutation model?

—*The therapeutic agents prioritized by these methodologies will be the same when:* a druggable oncoprotein is aberrantly activated because of one or more activating mutations in the corresponding oncogene.

—*The therapeutic agents prioritized by these methodologies will be different when:*

- An oncoprotein is aberrantly activated despite the fact that it does not harbor any activating mutations in the corresponding oncogene, or;
- An oncoprotein is not aberrantly activated despite the presence of one or more activating mutations in the corresponding oncogene.

## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING



**DARWIN**OncoTreat™

### A New Paradigm, Methodological Approach, and Technology for Precision Therapeutics in Cancer Medicine

Novel Tumor Checkpoint-Driven Strategies for Optimizing  
Alignment of Specific Malignancies in Specific Patients with  
Corresponding Treatment Options

Vastly superior and more specific than either actionable mutation analysis, or *DarwinOncoTarget™*, the *DarwinOncoTreat™* product uses a radically different approach to identify druggable tumor dependencies. Rather than focusing on aberrantly activated oncoproteins, *it focuses on master regulator (MR) proteins, organized into small, compact modules (tumor checkpoints), whose overall activity is both necessary and sufficient for tumor survival, independent of the specific tumor mutation pattern.* As a result, and in very general terms, this technology opens the door to identifying precise and dramatically more effective therapeutic alignments between drugs and patient-specific tumors that simply are not possible with mutation-dependent analyses.

These MR proteins are generally downstream of canonical oncoproteins (i.e., those that are recurrently mutated in cancer) and are themselves rarely, if ever mutated. Accordingly, *therapeutic agents are not prioritized based on their ability to inhibit the activity of a specific oncoprotein (i.e., a classic tumor target) but rather based on the ability of therapeutic agents to inhibit the activity of the entire MR protein signature representing a tumor checkpoint.*

Therefore, for this technology the answers to the questions raised above are as follows:

#### **1. In what ways and in which situations do the pharmacological targets identified by these approaches (i.e., *DarwinOncoTreat™* versus actionable mutation) overlap and when are the pharmacologic targets different?**

*The pharmacologic targets will overlap only when a single oncoprotein harboring one or more activating mutations is both necessary and sufficient to induce aberrant activation of all the MR proteins in the tumor checkpoint. This is obviously a relatively rare event. In all other cases, the predictions generated by actionable mutation analysis and DarwinOncoTreat™ will be different.*

#### **2. Are these analyses independent of each another or do they inform each other in some way?**

These analyses are completely independent of each other. *However, DarwinOncoTreat™ may inform and*

*improve upon the analysis of actionable mutations by assessing whether a specific mutation may be responsible for inducing aberrant activity of the tumor checkpoint.* Note that this implies that the mutation is sufficient but not *necessary* to induce activity.

#### **3. Will the therapeutic agents prioritized by *DarwinOncoTreat™* approaches be different from those dictated by the actionable mutation model?**

*Yes, and this is a critical point of departure: In most cases the therapeutic agents prioritized for a specific patient and tumor using DarwinOncoTreat™ technology will be different.* This is because rather than targeting a specific aberrantly activated oncoprotein, *drugs will be prioritized based on their ability to inactivate the entire repertoire of MR proteins in a tumor checkpoint.* There are rare cases, when targeting an oncoprotein may be both necessary and sufficient to induce collapse of the entire tumor checkpoint. In these *rare* cases the agents prioritized by *DarwinOncoTreat™* and those prioritized by analysis of actionable mutations will be the same.

In general, while targeting individual, aberrantly activated oncoproteins can be extremely useful (e.g., erlotinib for EGFR mutated lung cancer or trastuzumab for HER2 amplified breast cancer), this approach is less than optimal and often leads to relapse with drug-resistant tumors because **tumorigenesis requires the coordinated dysregulation of multiple proteins overseeing key tumorigenic programs—or, put differently, the mutational overload that underpins the tumorigenic state finds its operational success (and, therefore, its vulnerabilities) embedded in a constellation of dysregulated proteins.**

Mutational heterogeneity across multiple sub-clones in the same tumor provides exactly the evolutionary recipes to select bypass and alternative mutations that can independently activate the relevant tumor checkpoints, even when the activity of the primary mutated oncoprotein is targeted with a pharmacological inhibitor, thus resulting in tumor relapse to a drug resistant state.

Consequently, the most definitive and successful therapies are those that are **capable of disabling multiple MR proteins whose concerted aberrant activity is both necessary and sufficient to maintain the tumor state of the cell.** This is similar to a building with a few load-bearing columns and many that are not load-bearing. Destruction of a single load-bearing column or of any number of non-load-bearing ones is unlikely to affect building stability. Yet, there are specific, minimal patterns of load-bearing columns which, if significantly disrupted, will bring down the entire edifice.

## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

### Superiority of the DarwinHealth™ Technology Platform

Extending the Universe of Cancer Patients and Improving  
Predictive Capacity for Identifying Responders

The combination of **DarwinOncoTarget™** and **DarwinOncoTreat™** analyses offered by **DarwinHealth™** provides a unique opportunity to prioritize therapeutic strategies on a tumor-specific and individual patient basis. On one hand, **DarwinOncoTarget™** extends the ability to target established oncogenes, using an ever-increasing repertoire of targeted inhibitors. On the other, **DarwinOncoTreat™** identifies compounds and compound combinations that target the **full repertoire** of proteins whose activity is required for tumor maintenance, independent of the specific mutations. (See Table 4.)

Having made this point, the following is a legitimate question that needs to be explicitly addressed.

**Specifically, if both the *DarwinOncoTarget™* and actionable mutation models are pegged to a “single oncogene” strategy for prioritizing therapeutic agents, then why can DarwinHealth™ claim that the *DarwinOncoTarget™* technology is vastly superior?**

**DarwinOncoTarget™** greatly improves on an approach that oncologists are already largely accustomed to and with which they are familiar in the current clinical setting when they prioritize targeted inhibitors for their cancer patients. Specifically, we already have a large repertoire of FDA-approved compounds targeting oncogene dependencies in a variety of tumors, among them, CML, lung cancer, breast cancer, and many others. In addition, the repertoire of investigational inhibitors is continually expanding in the hope that they may be used to treat specific malignancies on a molecular biomarker basis. Currently, these molecular biomarkers are mostly represented by the presence of activating somatic mutations (e.g., ESR1 amplification) or germ line variants (e.g., BRCA1 inactivating variants). **However, these biomarkers provide are relevant and currently offer actionable treatment roadmaps in less than 25% of adult cancer patients.**

**What about the other 75% of patients with cancer who may have aberrantly activated oncogenes not detectable by current technologies?** Clearly, any technological breakthrough that will bring precision-focused cancer care to this vast universe of patients is desperately needed. By focusing on the aberrant activity of these same oncoproteins, **rather than just on their mutations, DarwinOncoTarget™** significantly improves and expands upon how clinicians may deploy targeted

therapeutic strategies based on the information already afforded by the oncogene addiction paradigm. Specifically, ***it dramatically empowers both physicians and patients by reproducibly identifying multiple aberrantly activated oncogenes, for which targeted inhibitors are available in the clinics, for virtually 100% of cancer patients.*** In addition, it can identify the specific changes in the repertoire of aberrantly activated oncogenes following relapse or progression, ***thus allowing therapy to be constantly adapted to target the most stage-relevant oncogene addiction.***

Hence, in this very important and decisive way, **DarwinOncoTarget™** technology represents an extraordinary advance in front line capabilities for the cancer specialist, as compared to the current oncogene addiction paradigm, which deprives millions of patients with cancer, and their oncology specialists, the opportunity to deploy precision-focused treatments for their care. The oncogene addiction paradigm, as emphasized, only identifies one out of a very large number of ways by which an oncogene may become aberrantly activated, i.e., the presence of one or more activating mutations. Thus, while **DarwinOncoTarget™** may in some cases prioritize candidate therapeutic agents—i.e., FDA-approved and investigational compounds targeting a single oncogene dependency—that are consistent and similar to those inferred from mutational model, ***it extends these capabilities to the full universe of patients with cancer.***

While **DarwinOncoTarget™** offers several, immediately applicable advantages over the current standard of DNA mutational analysis, its effectiveness is nevertheless restricted by its confinement to an analytical framework that is “**pre-keyed**” to those reasonably well-known proteins for which clinically tested inhibitors may be available, including those that were originally discovered as classic mutation-harboring oncogenes.

This brings us to **DarwinOncoTreat™**, the next level of **DarwinHealth™** technology (See sections below for detailed discussion of this product), which not only extends tumor-specific analysis to the full universe of cancer patients, but it does so with a technology that establishes an entirely new level of **precisional, prescriptive alignment—between tumors, their full suites of oncoprotein dependencies, and available therapeutic agents.** This technology and its implications for revolutionizing cancer treatment are unprecedented and exclusive to the **DarwinHealth™** portfolio.

How does it do this? In contrast to prior models based on oncogene addiction, the **DarwinOncoTreat™** platform utilizes proprietary drug databases, generated

**Table 4. COMPETITIVE MARKET ANALYSIS OF DARWINHEALTH™ PRECISION ANALYTICS FOR TUMOR-TARGET-THERAPY ALIGNMENT**

Products	 <b>DARWIN</b> OncoTarget™	 <b>DARWIN</b> OncoTreat™	 <b>FOUNDATIONONE™</b>
Database Against Which Patient DNA/RNA is Tested	1. -800 Cosmic + TCGA Oncogenes 2. -6,000 Regulatory Proteins 3. DarwinHealth Proprietary Drug Perturbation DB	1. -6,000 Regulatory Proteins 2. DarwinHealth Proprietary Drug Perturbation DB	1. 315 Oncogenes 2. 30 Translocations
Resolution Level for Tumor Dependencies	Individual Tumor Dependencies, Including Druggable Oncogenes and Other Regulatory Proteins	Full Tumor Dependency Signature. I.e., All Aberrantly Activated MR Proteins	Oncogenes Containing Somatic or Germline Functional Variants
Percent of Patient for Which Analysis Produces Actionable Results	100%	100%	<25%
Quality and Characteristics of Drug/Dependency Alignment	<b>2 to 3 Order of Magnitude Better P-Value Than Using Mutations or GEP</b>  <b>E.g. <math>p = 1E-6</math> for EGFR</b>	<b>2 to 3 Order of Magnitude Better P-Value Than Using Mutations or GEP</b>  <b>E.g. <math>p = 1E-6</math> for EGFR</b>	<b>Systematically Worst Than Based on MR Activity.</b>  <b>E.g. <math>p = 1E-3</math> for EGFR</b>
Key Deliverables and Oncologist Clinical Action Points Provided	1. Aberrantly Activated Proteins 2. Aberrantly Inactivated Proteins 3. Targeted Inhibitors and Combinations 4. Clinical Trials 5. Literature	1. Aberrantly Active MR Signature 2. Targeted Inhibitors and Combinations 3. Clinical Trials 4. Literature	1. Mutated Genes 2. Targeted Inhibitors 3. Clinical Trials 4. Literature
Estimated Cost	RNA-Seq: \$500 DH Report: \$100	RNA-Seq: \$500 DH Report: \$200	DNA-Seq: \$3,000 FO Report: \$???

**Table 4:** Competitive Market Analysis of DarwinHealth™ Precision Analytics for Tumor-Target-Therapy Alignment



by multi-dimensional cell perturbation studies conducted in multiple tumor lines. These data allow the systematic identification of drugs and drug combinations that can target the activity of the full repertoire of MR proteins in a tumor-specific checkpoint, whose activity is critical for tumor cell survival *in vivo*.

As a result **DarwinOncoTreat™** provides treatment recommendations that are entirely independent and significantly more effective than those targeting only individual oncogenes. Indeed, it would generally be a rare event when the drugs prioritized by **DarwinOncoTreat™** would overlap with those identified by single oncogene-mutation analysis. In general, **DarwinOncoTreat™** will: (a) identify therapeutic strategies that cannot be “envisaged or assessed” by any other methods or technology; and (b) identify targets and drugs that are significantly more predictive of and likely to effect, respectively, tumor response than those prioritized by other approaches.

#### Specific Examples, Based on Scientific Discovery, Demonstrating Superiority and Application of DarwinHealth™ Technology in Cancer Patients

To understand why **DarwinHealth™** technology represents a truly disruptive advance on the landscape of cancer treatment, a number of critical distinctions and comparisons should be emphasized.

First, drugs prioritized by **DarwinHealth™** technology and by actionable mutation analysis *may or may not* overlap. This is a critical point of distinction. For instance, in Figure 6, we show that most lung adenocarcinoma (LUAD) patients with a mutation in the EGFR receptor are identified by **DarwinOncoTarget™** as aberrantly activated. Hence, **DarwinOncoTarget™** and mutational analysis would prioritize the same drugs (erlotinib or some other EGFR inhibitor) for treating this group of patients.

Second (and this represents a critical point of departure), there also are: (a) several patients with EGFR mutations *but no aberrant activity of EGFR* (we have shown that these tumors are highly unlikely to benefit from the inhibitor); and, more importantly, (b) *many patients without any EGFR mutations that are identified by DarwinOncoTarget™ technology as having highly aberrant activity of the receptor*. These tumors, as we have shown, have a high probability of responding to EGFR inhibitors ( $p = 10\text{--}6$ ).

And third, a decision to treat with a specific agent based only on mutational data is much less predictive ( $p = 10\text{--}3$ ) than decisions made using **DarwinHealth™** technology. In general, **DarwinOncoTarget™**’s ability to predict responders based on aberrant activity of known oncogenes **outperforms**

**mutational analysis by 2 to 4 orders of magnitude in terms of statistical significance of the prediction.**

As shown in Figure 6, limited by space to only a few oncogenes and tumor types, the vast majority of patients with functionally relevant mutation in key oncogenes have aberrant activity of the protein, as assessed by **DarwinHealth™** analysis.

#### Differences Between Activated Oncoproteins and Tumor Checkpoint™ Signatures

What Advantages Does Tumor Checkpoint Analysis Bring to the Landscape of Precision Therapeutics for Cancer Medicine?

Aberrantly activated oncoproteins, whether or not they may be targeted pharmacologically, do not necessarily represent a “tumor checkpoint.” (See **DarwinOncoTreat™: A Closer Look below for discussion of Tumor CheckPoint™ signatures**.) Rather, they generally represent upstream proteins that are responsible for activating the tumor checkpoints. The distinction is that the proteins in a tumor checkpoint are both necessary *and* sufficient to maintain the cell’s tumor state. In contrast, most oncogenes are strictly sufficient and not necessary. For instance, while 35% of breast cancers have mutations and aberrant activity of the PI3K kinase, 65% of virtually identical tumors can maintain their status without such mutations, which is likely due to mutations in other bypass or alternative genes. These capacity—and focus—on such distinctions highlights the key difference between **DarwinOncoTarget™** and **DarwinOncoTreat™**.

Thus, based on **DarwinOncoTarget™** analysis, EGFR aberrant activity would be detected and targeted. However, based on **DarwinOncoTreat™** analysis (which relies on its proprietary **DarwinOncoDrugBase™** database) the drug prioritized for that tumor would abrogate the activity of a number of key MR proteins downstream of EGFR. This represents the actual tumor checkpoint that needs to be targeted pharmacologically to induce loss of lung cancer cell viability, *independent of EGFR mutations*. We have shown that **DarwinHealth™** technology can effectively identify MR proteins representing the universal tumor checkpoint, while also assessing aberrant activity of upstream classic oncogenes.

We know, for example, that about 20% of lung cancer patients with EGFR mutations do not respond to erlotinib. Consistently, about 20% of patients with mutations are detected as not having any aberrant activity of EGFR. And, about 50% of patients who have aberrant EGFR activity detected by **DarwinOncoTarget™** technology have **no EGFR mutation whatsoever**. So the detection technology offered by **DarwinOncoTarget™** may increase by about two-fold the number of patients with NSCLC who may benefit from the inhibitor, erlotinib.

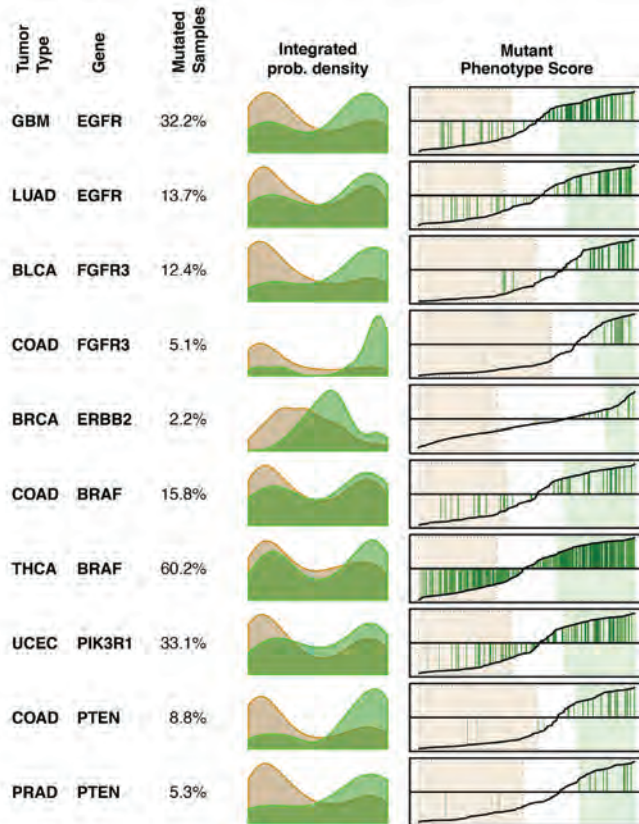


## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

This increase in coverage is automatically applied to account for all targetable oncogenes; and, characteristically, each patient with a tumor will present with several oncogenes that are aberrantly activated. This is why **DarwinOncoTarget™** increases coverage and actionable therapeutic insights for patients with targetable oncogenes from <25% of the cancer population, as is currently detected by actionable mutation analysis, to 100% of cancer patients, if their tumors undergo analysis with **DarwinOncoTarget™**.

Figure 6 below summarizes these important points using results obtained from patients in The Cancer Genome Atlas repository. Patients are ranked left to right from the one with the lowest to the highest predicted activity of a specific oncoprotein in a specific tumor type, e.g., EGFR in glioblastoma (GBM), in the first graph. Patients in the green shaded area are those identified by **DarwinOncoTarget™** as having aberrant activity of the oncoprotein. Those marked by a thin, green vertical line are those with an activating mutation in the oncogene.

From a clinically actionable perspective, clearly the vast majority of patients with a mutation in the oncogene are identified as having aberrant activity of the oncoprotein (green shaded area). **However**, depending on the specific oncogene, **there are many patients who may harbor mutations and yet show no aberrant activity of the protein**. This may be because the mutation is not expressed or because of a host of other reasons. These patients would be unlikely to benefit from the drug identified by the oncogene addiction mutational model, but currently, they would be likely to be receiving such an agent. Similarly, patients in the green shaded area that lack the green vertical bar are identified as having aberrantly high activity of the oncoprotein, **but they have no associated mutation**. These patients would likely benefit from the drug but, using the current oncogene model, they would likely not be offered this therapy because the oncogene model fails to identify them as appropriate candidates.



**Figure 6.**

**Figure 6.**

Comparison of **DarwinOncoTarget™** predictions of aberrantly activated oncoproteins to those based on the actionable mutation paradigm. Only ten tumor specific oncoproteins are shown for space reasons. Each horizontal stripe includes five elements: (a) the specific TCGA tumor cohort in which the analysis was performed, including glioblastoma (GBM), lung adenocarcinoma (LUAD), colon adenocarcinoma (COAD), bladder adenocarcinoma (BLCA), breast adenocarcinoma (BRCA), and renal adenocarcinoma (READ); (b) the oncogene name; (c) the fraction of samples in the cohort harboring oncogene mutations that alter the oncoprotein sequence; (d) graph showing the probability of samples to harbor the mutated oncogene (y-axis) as a function of the **DarwinOncoTarget™** predicted oncoprotein activity (x-axis). The latter is sorted left to right from the lowest to the highest predicted activity; (e) oncoprotein activity on a sample by sample basis (y-axis) ranked from the lowest (left) to highest (right) predicted activity. Samples in the green shaded area are predicted to have aberrant oncoprotein activity by **DarwinOncoTarget™** analysis. Samples marked with a vertical green bar harbor protein altering mutations in the corresponding oncogene.

As shown, **DarwinOncoTarget™** correctly predicts aberrant oncoprotein activity for the majority of mutated samples. However, it both identifies mutated samples with low oncoprotein activity that have low probability of responding to the inhibitor as well as non-mutated samples with highly aberrant oncoprotein activity that are likely to respond to the inhibitor.



DARWIN|OncoTreat™

**DarwinOncoTreat™: A Science-to-Strategy Framework  
for Definitive Advancement of Cancer Care Beyond  
the Mutational Analysis Model**

A Closer Look at the Methods, Rationale, and Databases that  
Underpin the Disruptive, DarwinCheckpoint™ Paradigm for  
Precision-Focused Cancer Therapy

Addressing what is arguably the most urgent and pervasive unmet need in cancer therapeutics, **DarwinOncoTreat™** is the first proven methodology for the identification of individual drugs and drug combinations that target the complete repertoire of regulatory proteins necessary for tumor survival on a patient-by-patient basis, rather than just individual oncoproteins based on their mutational state.

In support of this cancer treatment breakthrough, **DarwinHealth™** co-founders have shown that master regulator (MR) proteins necessary for tumor survival are organized in tightly coupled modules (tumor checkpoints), whose concerted aberrant activity is necessary for tumor cell survival in a variety of cancers, including cancers with very different genetic alteration patterns. Indeed, **tumor checkpoints** are much more conserved across patients with similar cancer subtypes—and even across distinct cancer types—than the specific mutations that induce tumorigenesis.

**But why is it important to target multiple proteins in cancer?** It is well known that *even the most potent oncogene is not sufficient to induce tumorigenesis in isolation*. For instance, many of our nevi have BRAF or KRAS mutations. This is because our cells contain very sophisticated failsafe mechanisms to prevent transformation. For instance, presence of mutated KRAS in the cell triggers senescence, a mechanism by which cells go “dormant” and eventually die.

In a similar way, triggering oncogenesis (making normal cells transform into malignant cells) requires the concerted activity of many different proteins that, together, modulate multiple genetic programs (also known as the hallmarks of cancer) necessary for cancer survival and proliferation. Let’s return, for a moment to our analogy in which multiple, load-bearing columns are required to stabilize a building. Much like targeting a single oncogene, destroying a single load-bearing column is unlikely to bring down the entire building, although it may damage part of the building. However, destruction of even a small

number of load-bearing columns can lead to collapse of the entire structure, including all remaining load-bearing columns via a domino effect.

In this sense, **DarwinHealth™** technology has been developed specifically to identify drugs that are best able to target the entire “load-bearing” structure of the cancer cell, i.e., the key MR proteins that maintain a tumor checkpoint in its aberrantly active state. This precisional capacity allows successful drug prioritization to be employed at the level of the individual cancer patient; or, more importantly, across a significant fraction of cancer patients whose tumors depend on the same tumor checkpoint mechanism.

Put slightly differently, a tumor checkpoint—sometimes also called a “tumor bottleneck”—comprises the *minimal set of MR proteins* that are dysregulated by the tumor’s mutational load **and** that are necessary and sufficient to control the genetic programs necessary for tumor cell state maintenance.

Reliance on the tumor checkpoint paradigm is what distinguishes the **DarwinOncoTreat™** methodology from both the actionable mutation and the **DarwinOncoTarget™** strategies, both of which are pegged to single oncogenic mutations, albeit with very different application potential and accuracy.

*This explains why the tumor checkpoint paradigm, which seeks to identify a set of proteins working together to implement the tumor phenotype, represents a critical point of scientific, conceptual, and analytical departure from all previous strategies of therapeutic attack focused on anti-tumor therapy.*

By analogy, one can think of tumor checkpoints as those necessary and sufficient columns that ensure a skyscraper’s stability. At the cancer cell level, **DarwinOncoTreat™** may identify 20 proteins in a tumor checkpoint, such that *precise inactivation of a specific protein or protein-pair in the set will result in the systematic inversion of all checkpoint protein the activities (i.e., tumor checkpoint collapse) and, in turn, induce irreversible loss of tumor viability in vivo*.

Critically, *we have shown that there are multiple individual proteins and protein pairs that can induce tumor checkpoint collapse*. For instance, in follicular lymphoma progression, we identified at least 9 protein pairs in a tumor checkpoint that induce checkpoint collapse and dramatic loss of viability. While there likely are many other pharmacological targets that may not be identified by this approach, *the unprecedented ability to systematically*

*identify relevant targets that are much more universal than individual mutated oncogenes, using quantitative analysis of cancer regulatory models, represents an extraordinarily precise approach that meets an unmet need to develop more precise, tumor- and patient-specific treatments at the front lines of cancer care.*

Naturally, there are multiple ways to “bring down” those cancer load-bearing proteins. The optimal approach, of course, would be to deploy drugs that directly target the tumor checkpoint proteins. However, if such drugs are not available, *one can also utilize drugs that target the pathway just upstream of the checkpoints.* For instance, ibrutinib is a bruton kinase (BTK) inhibitor. BTK is never mutated in diffuse large B-cell lymphoma. However, it is a critical upstream modulator of the aberrant activity of Nf-kB, which is the key tumor checkpoint in the ABC subtype of the disease (Compagno et al. Nature 2009). As such, it is now used clinically, with great success, to treat patients with this type of cancer.

It follows that the farther away from the target tumor checkpoint you go, the more likely it is that the cell may “rewire” itself into a state that blocks the effect of the drug, an adaptation the frequently induces resistance. It is for this reason that patients managed within the framework of **DarwinHealth™** technology will have to be monitored over time using a dynamic, multi-temporal approach. This longitudinal approach to patient care will help discover new entry points to address relapse as the disease progresses. However, once drugs are developed that target tumor checkpoints directly, relapses will be much less likely to occur.

The drugs selected by **DarwinHealth™** technology are those that *can optimally reverse the activity of all, or at least, the majority of proteins in the bottleneck.* If a drug hits a relevant mutation, one may see a partial reversion effect but this is unlikely to cause a global reversion of all bottleneck proteins (i.e., MR proteins). As a result, the full tumor checkpoint will not collapse.

#### Pharmacological Targeting of Oncogenes and Tumor Checkpoints

To address the urgent and unmet need for more precision-oriented cancer treatment, it is highly desirable to have a real world, operational framework that would facilitate immediate and automatic identification and deployment of the available repertoire of FDA-approved drugs and investigational compounds, or their combination, for the purpose of either: (a) inactivating proteins identified by **DarwinOncoTarget™**; or (b) inactivating

tumor checkpoints identified by **DarwinOncoTreat™**.

Obviously, having direct knowledge of a targeted inhibitor for a specific protein or proteins identified by these analyses is one possible way to make these alignments, especially for activated oncoproteins identified by **DarwinOncoTarget™**. For instance, if EGFR is identified by the analysis as an aberrantly activated oncogene, a plethora of clinically approved inhibitors producing such inhibition already have been identified. However, we have also demonstrated that, within specific contexts, **the current repertoire of FDA-approved and investigational drugs can be repurposed to target almost any protein of interest by taking an unbiased, data-driven (i.e. automatic) approach to screening for compound activity.** This methodology, unique to the **DarwinHealth™** technology portfolio, permits a dramatic expansion in our understanding of when approved agents may be effective against tumors against which their anti-tumor activity is not yet known, and has not been well characterized.

It should be noted that the manner in which MR proteins—including MR proteins in tumor checkpoints—are identified by **DarwinHealth™** technologies, is by assessing the expression of their transcriptional targets. Precisely the same method can be used to assess whether any drug is available that represents a viable inhibitor of that MR protein in the specific cancer context in which it is aberrantly activated.

Indeed, by performing RNA Seq profiling of representative tumor cells before and after “treatment” (i.e., titrated exposure to) with a large number of compounds, *we can determine whether expression of the targets of a MR protein has been globally inverted. I.e., whether all the targets that were overexpressed are now underexpressed and vice versa as a result of exposure to a specific compound.* Such a finding would identify the compound as a specific inhibitor of the MR protein activity. Similarly, for tumor checkpoints identified by **DarwinOncoTreat™**, *one can simply repeat this assessment across all the MR proteins in the tumor checkpoint to assess inversion of their activity, thereby identifying compounds capable of inducing tumor checkpoint collapse.*

To achieve this goal, **DarwinHealth™** co-founders have developed and will continue to develop large scale drug databases representing the RNA Seq profile of FDA-approved and investigational compounds in a large set of tumor subtypes, from triple-negative breast cancer (TNBC) and lung cancer to mesenchymal glioblastoma and neuroendocrine tumors. This resource, the **DarwinOncoDrugBase™**, represents both a valuable commercial asset and it offers an



## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

unique mechanism through which **DarwinHealth™** *can directly match the specific vulnerabilities of individual tumors to the optimal profile of either individual drugs or drug combinations.*

These critical alignments between specific tumors and agents is made possible using **DarwinHealth™** technology, because assessment of the proteins whose activity is modulated by a compound (i.e., the compound mechanism of action) can be performed very effectively in cell lines. These alignments can be made very accurately with **DarwinHealth™** technology because, although the ability of a drug to kill a tumor in a patient is rarely recapitulated in cell lines, **the mechanism of action (MoA) of a drug is very effectively studied in cell lines and its effects are well conserved in vivo.**

For instance, an inhibitor of the EGFR receptor will inhibit it both in cell lines and in a patient's tumor, assuming that it already has acceptable pharmacokinetic and pharmacodynamic (PK/PD) properties in the human host. The acceptability of the latter (i.e. its PK/PD profile, as well as its toxicity profile) is a given in **DarwinHealth™** analyses because these analyses, and corresponding drug recommendations, are limited to FDA-approved drugs and investigational compounds that are already being used in the cancer clinical setting.

To eliminate the possibility that the desired effect on the target protein may be idiosyncratic to the presence of specific mutations, **the database is generated using multiple cell lines representing great mutational variability.** Similarly, it's important to ensure that compound selection is not driven by specific concentrations of the drug or evidence of its effectiveness that is observed only at specific time points. To avoid this, data used to construct the **DarwinOncoDrugBase™** are assembled using perturbations of cancer cells at multiple concentrations and the effects also are profiled at multiple time points following perturbation, using a proprietary methodology. **This approach has produced remarkably reproducible results, with highly statistically significant correlation between drug profiles representing related drugs or the same drug.**

In summary, above and beyond the actionable information provided by **DarwinOncoTarget™**, the ability of **DarwinOncoTreat™** technology to match pre-screened compounds in the **DarwinOncoDrugBase™** to the full MR program of a specific patient's tumor, within minutes of receiving the tumor RNA Seq data, **represents a remarkable and desperately needed improvement in precision cancer medicine.**

The MR target-focused approach permits rapid and automatic identification of cancer drugs—and delineation of their precise mechanisms of action—that can disrupt, shut down, and/or undermine tumor viability. Unlike any other analytical approach currently used on the landscape of cancer medicine, the drugs identified by **DarwinOncoTreat™** technology are those that have been proven to operate **mechanistically** across the full range of tumor-specific, molecular and operational dependencies at the oncoproteomic level, rather than by simply selecting a specific oncogene to target. Systematic targeting of the entire suite of critical tumor checkpoints—the **DarwinCheckPoint™** Signature—significantly decreases the likelihood of tumor relapse for several reasons, including the ability to counter the effects of alternative mutations before they start feeding and reconfiguring vulnerabilities of the same tumor checkpoint, leading to sub-clonal selection.



DARWINOncoDiscovery™

### An Innovative and Disruptive Preclinical Platform for Drug Discovery and Compound Characterization

Based on the products discussed and described above, it is evident that the technology, IP, and products in the **DarwinHealth™** portfolio represent a significant and unprecedented discovery platform for biopharmaceutical and biotechnology companies that are interested in prioritizing or repositioning their approved drugs and developmental compounds.

Implementing new paradigms based on MR-driven, tumor checkpoint signatures, **DarwinHealth™** provides cutting-edge research technology and expertise to decrease the uncertainty associated with product development in the biopharmaceutical and life sciences space.

The **DarwinOncoDiscovery™** represents the full gamut of **DarwinHealth™** proprietary technologies, methodologies, and databases that are designed to address key challenges in drug development and compound repositioning.

The unique services that can be executed with the proprietary **DarwinOncoDiscovery™** platform include: (a) compound characterization in terms of assessing its full therapeutic potential across an extensive range of tumor types; (b) elucidation of compound mechanism of action in diverse tissues; (c) identification of biomarkers for stratification of responder or drug-resistant populations; (d) elucidation of mechanisms of resistance and associated small molecule inhibitors that will rescue sensitivity;

## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

(e) characterization of compound synergy with other FDA approved and investigational compounds; and (e) characterization of potential compound toxicity.

**DarwinHealth™** already maintains databases with characterization of activity of most FDA-approved and investigational compounds, from several companies. This leads to systematic identification of novel applications and therapeutic areas for compounds that currently have a limited application range. To establish itself—within the framework of its unique systems-based, data-driven approach to cellular dysregulation across a broad spectrum of disease states—as the leading provider of actionable discoveries and analytics focused on drug therapy, **DarwinHealth™** is in the process of establishing collaborative agreements and contracts with several pharmaceutical and biotech companies to perform a thorough analysis of their developmental pipeline. Based on the results, **DarwinHealth™** will provide specific information, suggestions, and developmental roadmaps for compounds that show specific and outsized potential—and unique competitive advantages—for targeting key, unaddressed tumor checkpoints in specific tumor subtypes.

In general, such an assessment is based on matching compound-specific perturbational assays to master regulators that comprise subtype-specific tumor checkpoints, as inferred by **DarwinHealth™**'s proprietary methodologies. Based on these preliminary “proprietary” discoveries and previously unrecognized tumor-target/targeted-drug alignments, **DarwinHealth™** will engage the company which owns the IP rights to the compound in a collaborative effort to generate a comprehensive, detailed, and clinically relevant compound characterization program implemented within the **DarwinOncoDiscovery™** service line. In addition, we welcome requests initiated by pharmaceutical and biotechnology companies to perform full or partial characterization of their proprietary compounds, from a single compound of specific relevance to their full developmental pipeline.

**DarwinOncoDiscovery™** is designed to provide a comprehensive, 360° evaluation of the pharmacological properties of any compound of interest. It does so by employing a proprietary **DarwinHealth™** combination of *in vitro*, *in vivo*, and *in silico* assays to assess the compound's likely potential as a novel and highly therapeutic agent for a tumor subtype-specific indication. This evaluation includes a complete array of validation assays, identification of specific molecular biomarkers to stratify sensitive and resistant patients, full elucidation of tumor context specific compound mechanisms of action, and elucidation of additional sensitizers compounds that

rescue sensitivity in resistant cells or that synergize with the compound of interest. This is accomplished through a set of interdependent and coordinated sub-projects, as discussed below.

### DARWIN!OncoValidate™

#### Validation of the Potential of a Proprietary Compound within a Specific Tumor Subtype

Once a proprietary compound has been identified as a candidate inhibitor for a subtype-specific tumor checkpoint, the first step in the **DarwinOncoDiscovery™** drug delineation platform is the broader validation of a compound's therapeutic potential across an extremely robust set of representative models. Specifically, this step entails a concerted series of proprietary perturbational assays, at the cellular level, covering a comprehensive repertoire of subtype specific cell lines and patient-derived xenograft models. This comprehensive approach to analysis of a compound(s) ensures that the initial assessment was not due to cell line-specific idiosyncrasies and that it properly accounts for inter and intra-tumor heterogeneity.

To achieve these goals, we first identify an appropriate number of well-characterized cell lines that recapitulate the master regulators and overall tumor checkpoint addiction of the tumor subtype of interest. Depending on tumor subtype and availability of relevant cell line models, this may range from 3 to 20 cell lines. Then, rather than assessing compound activity based on cell viability assays, **DarwinHealth™** assesses its ability to reverse the tumor-specific MR signature, based on molecular profiles obtained from systematic perturbational assays.

The rationale for these analysis is that, while it is widely known that *in vitro* viability is rarely if ever predictive of compound activity *in vivo*, we have shown that **MR activity signature reversal by a compound represents an excellent predictor of activity in vivo, both for individual compounds and for compound combinations**<sup>19</sup>. (See Figure 7.)

The value of this approach is predicated on the following experimentally validated assumptions: (a) that tumor checkpoints identified from analysis of primary tumors are highly enriched in both essential and synthetic lethal master regulator proteins; (b) that MR signature reversal, as assessed by compound perturbation analysis of cell lines that recapitulate the primary tumor's MR signature, is highly conserved *in vivo*; and, (c) that complete reversal of

## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

MR activity signature induces tumor checkpoint collapse, resulting in loss of tumor viability *in vivo*. (See Figures 8 and 9.)

Following successful confirmation of compound-mediated MR-signature reversal *in vitro*, we will assess the compound's activity in an appropriate number subtype specific PDX models or patient-derived explants. The specific number of models is we will evaluate is clearly dependent on the number that are available for a specific tumor sub-type. Some tumors, such as triple negative breast cancer, have great availability of PDX models, whereas others, such as gastroenteropancreatic neuroendocrine tumors (GEP-NETs), have no available models.

In general, we will attempt to validate findings *in vivo* in at least five (5) distinct PDX models but a smaller or larger set can be determined in agreement with the partnering company, depending on the specific tumor context. For tumors with no PDX model availability, we will use recent advances in organotypic culture models, based on treatment of freshly harvested tumor explants. Since our assays have a molecular endpoint, assessed by RNA sequencing at 6h – 24h, drug perturbation assays can be easily performed using appropriately sectioned explants, whose cells can be maintained in a viable state for time periods exceeding 72h. **DarwinHealth™** will have broad availability of PDX models via strategic partnerships, either for pre-existing models, such as those established by Champions Oncology, or based on models generated *de novo* in collaboration with the mouse hospital core resource at Columbia University.

### DARWIN|OncoBiomarker™

#### Identification of Biomarkers for Patient Stratification

Clinical testing of investigational compounds is increasingly dependent on availability of high quality molecular biomarkers to identify responder populations as well as to identify specific patients who may relapse with drug resistant tumors. For the former, population stratification to identify candidate responders will allow FDA approval of successful compounds using dramatically smaller cohorts, thus reducing both timelines and cost. For the latter, effective identification of patients at risk of relapse will help elucidate the specific mechanism of relapse and identify candidate sensitizer compounds to rescue drug sensitivity, as we have shown both for

glucocorticoid resistance in T-ALL and for trastuzumab resistance in HER2 amplified breast cancer<sup>5,6</sup>.

**DarwinHealth™** founders have demonstrated that **master regulator proteins not only represent optimal therapeutic targets but also constitute exceptional biomarkers that significantly exceed mutation-based biomarkers in terms of their predictive ability**. This is because these genes are part of tumor checkpoints whose activity is both necessary and sufficient for tumor phenotype maintenance. **As a result, their activity is directly and causally associated with the specific tumor phenotype of interest.**

For instance, we have shown that the two MR proteins FOXM1 and CENPF are not only synergistic master regulators of aggressive prostate cancer but, as a pair, represent optimal predictors of prostate cancer progression (See Figure 10). Indeed less than 1% of the patients that presented with negative immunohistochemistry for both proteins at diagnosis died of prostate cancer, while the patients with co-staining for both protein represented 90% of the tumor mortality in a cohort of over 900 patients collected at Memorial Sloan Kettering and followed for more than 20 years. Critically, neither of the two proteins in isolation was predictive and neither was significantly differentially expressed in patients with aggressive tumors<sup>8</sup>. As a result, they could only be identified with the advanced methodologies that are at the core of **DarwinHealth™**'s product line.

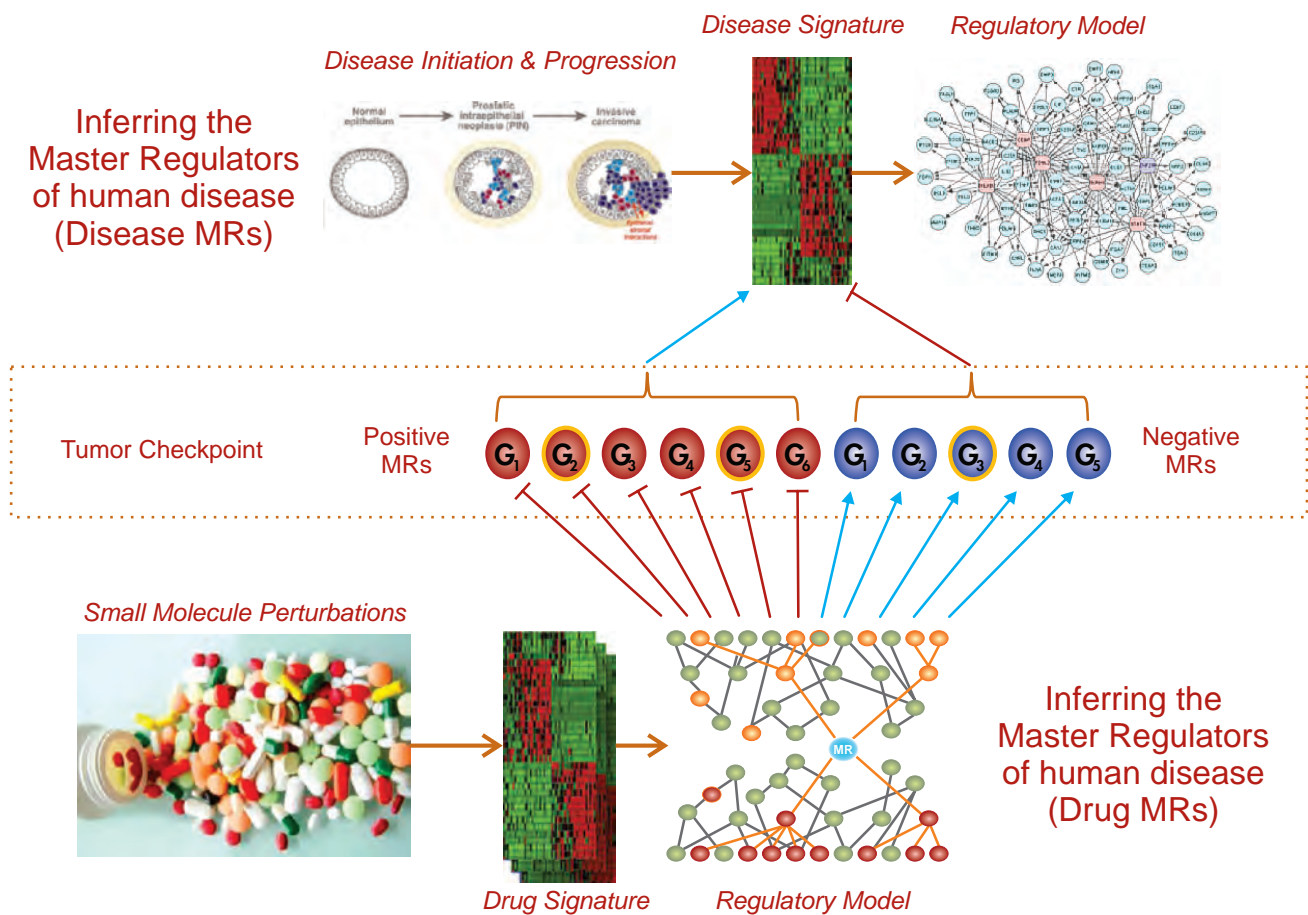
In terms of characterizing and addressing drug resistance, **DarwinHealth™** founders identified AKT1 as the master regulator of glucocorticoid resistance in T-ALL and phospho-STAT3 as the master regulator of HER2+/ER- resistance to trastuzumab. These findings have led to clinical trial development resulting in promising preliminary data. In both cases, identification of the MR protein representing the biomarker of drug resistance led to identification of re-sensitizer compounds that fully rescued sensitivity to the original therapy *in vivo* and, for the latter, in the clinic, i.e., MK2206 and ruxolitinib, respectively.

**These analyses can be easily extended to any compound of interest.** Additionally, given the ability of this methodology to perform MR analysis from FFPE blocks, this approach can be easily applied retroactively to patients enrolled in clinical trials that have been completed, thus leading to elucidation of sensitivity and resistance MRs, as well as to identification of sensitizer compounds that can rescue sensitivity when used in combination therapy.





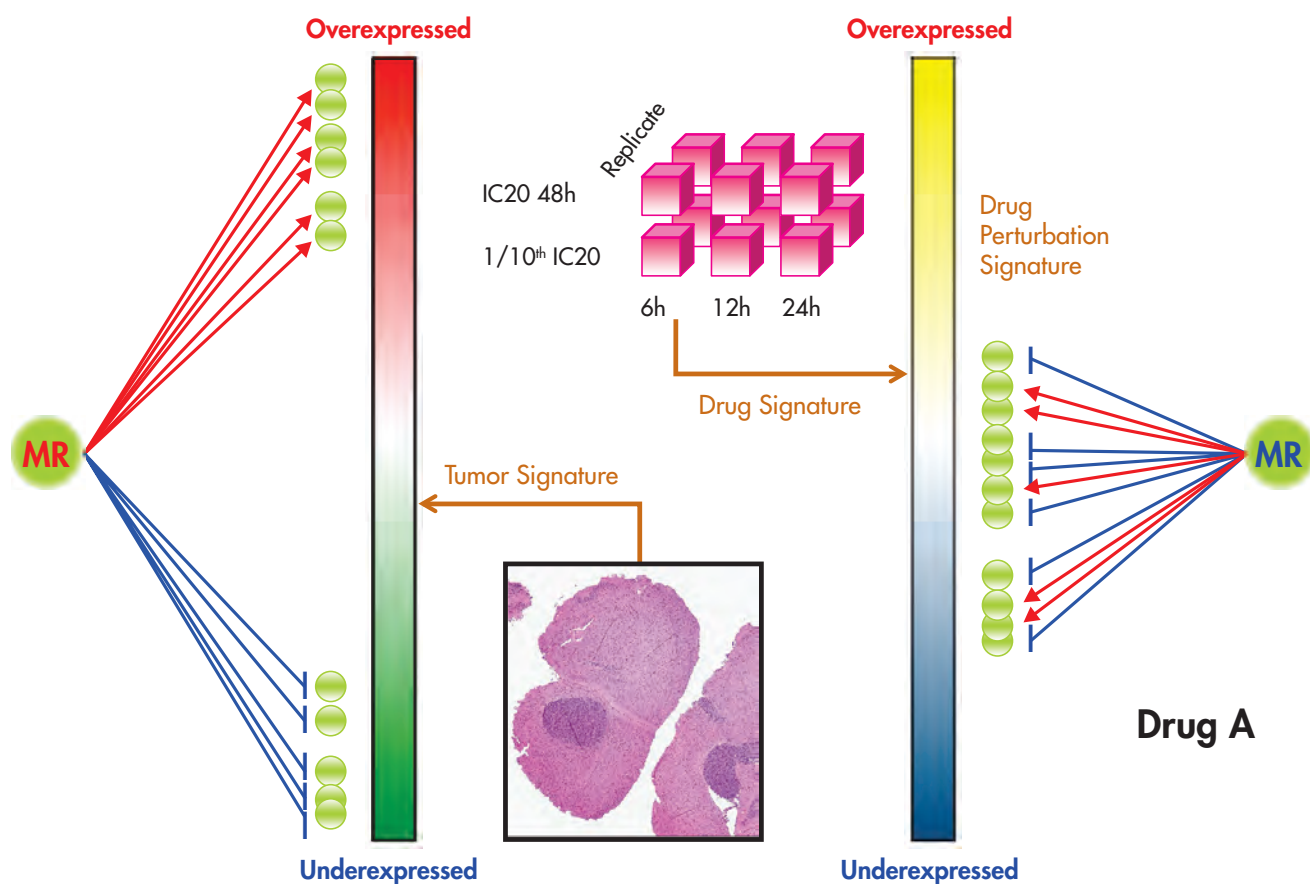
## Matching Diseases to Drugs



**Figure 7.**  
**MATCHING DISEASES TO DRUGS**



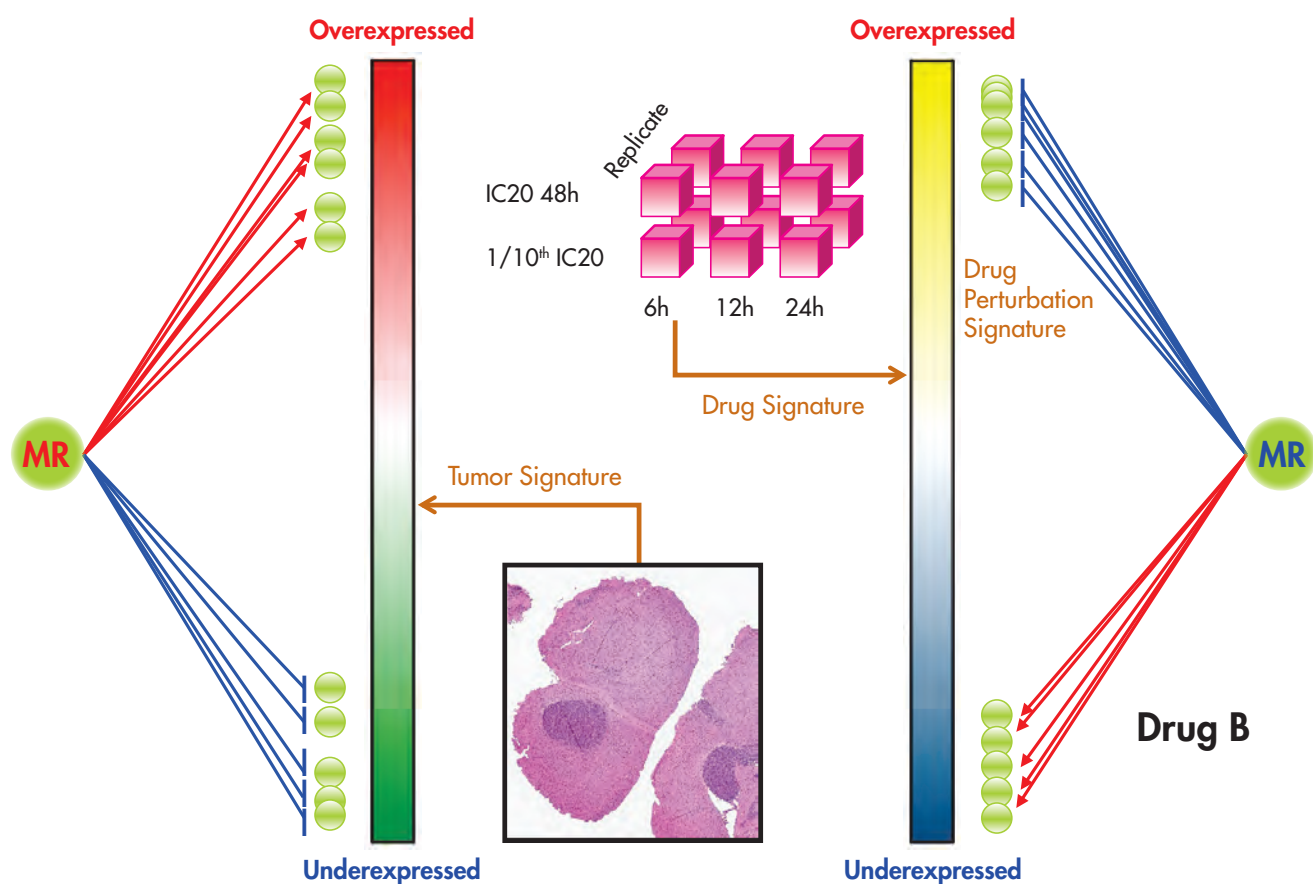
## VIPER: VIRTUAL Proteomics Algorithm



**Figure 8.**  
**VIPER PROTEOMICS ALGORITHM**  
Drug A does **not** promote reversal of tumor signature.



## VIPER: VIRTUAL Proteomics Algorithm



*Figure 9.*  
VIPER PROTEOMICS ALGORITHM

Drug B **promotes reversal** of tumor signature, thereby making it a viable therapeutic option.

## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

### DARWIN<sup>®</sup>OncoMoA™

#### Precise, Genome-wide Assessment of Compound Mechanism of Action

A critical issue in evaluating the suitability of a compound for clinical studies is the characterization of its mechanism of action (MoA), i.e., the full repertoire of proteins responsible for its pharmacological activity and potential toxicity. These analyses can identify critical effector proteins that may be harnessed to modulate activity of tumor checkpoints as well as of off-target proteins that may be critically relevant in determining compound's toxicity.

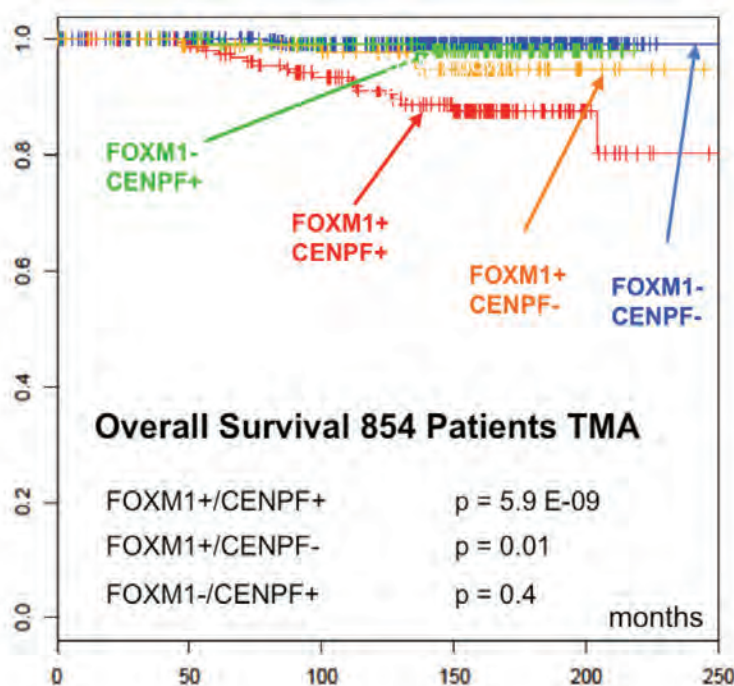
We assess compound MoA by determining: (a) the proteins whose activity is affected following compound perturbation, using the VIPER algorithm<sup>4</sup> and; (b) the proteins whose molecular interactions are most significantly dysregulated by the compound, using the DeMAND algorithm<sup>17</sup>. The integrated analysis of these contributions provides a highly accurate and comprehensive repertoire of proteins whose activity is either critical in mediating or in modulating the pharmacological effects of a compound.

In addition, MoA analysis can be used to performed molecular screening for compounds that are specifically suited to inhibit a desired and pharmacologically relevant

target, including targets that are considered undruggable, such as transcription factors. Using PLATE-Seq, a novel low-cost, high-throughput technology for RNA-Seq profiling of cells following compound perturbation, **DarwinHealth™** can build universal gene reporter assays that can report on the compound's ability to modulate the activity on any protein of interest.

Since the cost per well for this technology is ~\$30, this compares favorably with more specialized assays that report on the activity of a single target and cannot be tailored to report on additional targets of interest *a posteriori*. In combination with the VIPER algorithm, this allows the unprecedented ability to screen for inhibitors of arbitrary targets of interest for a fraction of the cost typically associated with the design, miniaturization, and implementation of specialized high-throughput screening assays.

The technology requires generation of RNA Seq profiles following compound perturbation in multiple tumor specific cell lines, at multiple concentrations, and at multiple time points. A minimum of  $N = 6$  profiles per cell line is required by DeMAND (optimal  $N = 10$ ) while VIPER can work even with individual compound perturbation signatures. This allows full characterization of any proprietary compound library of interest both in terms of compound MoA and in terms of their prioritization as agonists or antagonists of any target protein of interest.



**Figure 10.**

Kaplan-Meier analysis of prostate cancer patient survival based on immunohistochemistry of FOXM1 and CENPF MR proteins at diagnosis. Double negative (blue curve) had <1% mortality, while double positives (red curve) had >90% mortality. FOXM1 positives/CENPF negatives were borderline significant and FOXM1 negative/CENPF positives were not statistically significant.

## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

### DARWIN|OncoTox™

#### Assessing Potential Compound Toxicity.

Using **DarwinHealth's** proprietary technology, we can match a compound's MoA, as inferred by VIPER and DeMAND, to the MoA of compounds known to present significant toxicity in the liver, kidney, erythropoietic, nervous, and cardiac systems. This analysis will alert of any potential toxicity mechanisms that may require monitoring during follow up clinical studies.

Today's current methodologies for the study of compound toxicity largely rely on differential expression analysis and phenotypic effect in mouse models, which are often not predictive of human compound toxicity. **In contrast, our methodologies are based on a systematic, comparative analysis of a compound's mechanism(s) of action. This is accomplished using a full genome-wide repertoire of MoA proteins, thereby permitting the compound of interest to be compared to compounds known to be associated with and produce relevant toxicity in the human host.** Critically, this analysis can be performed in cell lines or *in vivo*. In the latter case, by harvesting distinct tissues from a treated animal **DarwinHealth™** can generate a comprehensive toxicity profile for any compound of interest across all relevant tissues. Even if the compound toxicity is not manifested by an overt phenotype, an MoA footprint comparison to compounds with established human toxicity will be recapitulated in model organisms.

### DARWIN|OncoExtend™

#### Repurposing Additional Compounds in Biopharma's Developmental Repertoire as Candidate Therapeutic Agents in Additional Diseases.

**DarwinHealth™** will work with partner companies to perform systematic characterization of any proprietary compounds in their developmental and pre-developmental pipeline as candidate therapeutic agents in a full range of tumor subtypes. **An unique advantage of our proprietary methodology is that it does not require the presence of specific mutations in cancer patients.** Rather compound prioritization is based exclusively on the effects that the agent has on the aberrant activity of the master regulator proteins that comprise a subtype specific tumor checkpoint. These proteins have been shown to either elicit tumor-specific essentiality, when individually

inhibited, or tumor specific synthetic lethality, when co-inhibited in pairs.



### DARWIN|OncoSynergy™

#### Elucidating Synergistic Compound Activity

**DarwinHealth™** will deploy its proprietary **DarwinOncoSynergy™** technology<sup>19</sup> to identify additional compounds among non-proprietary FDA-approved compounds as well as from the partner's proprietary developmental pipeline that synergize with any proprietary compound of interest. Identification of such synergies will dramatically increase the activity of the compound(s) in patients with specific tumor checkpoint additions. These data are already available for all compounds that were previously tested in our perturbational assays and that are represented in our proprietary **DarwinOncoBase™** repository. These analyses can also be easily extended to any additional compound of interest, Compound synergy predictions will be followed by *in vitro* and *in vivo* validation assays using established methodologies for synergy assessment<sup>16</sup>.





## DARWINHEALTH™ PROPRIETARY TECHNOLOGIES

### DARWIN|CheckPoint™

Over the past decade, **DarwinHealth™** founders have shown that *cell states representing both normal human physiology and human disease—from cancer to neurodegenerative disease—are governed/determined by the concerted activity of a relatively small number of proteins, organized within tightly interconnected regulatory modules*<sup>2, 3, 5, 8-10, 20-26</sup>. Such “master regulator” (MR) proteins are only infrequently affected by genetic alterations. Rather, their abnormal activity (i.e., *dysregulation*) results from the cooperative effect of multiple germ line variants, somatic mutations, and environmental signals in their upstream pathways<sup>7, 15</sup>.

These MR proteins represent the operational infrastructure of critical disease-checkpoint-modules (i.e., *disease checkpoints*, for short), that are responsible for: (a) integrating the effect of complex upstream genetic patterns (i.e., the variants and mutations that contribute to disease risk); and, (b) activating the specific transcriptional programs necessary for the maintenance of the disease cell state. **DarwinHealth™** has implemented proprietary technology (**DarwinCheckPoint™**) to systematically identify and target disease checkpoint proteins, on an individual patient basis. (See Figure 11.)

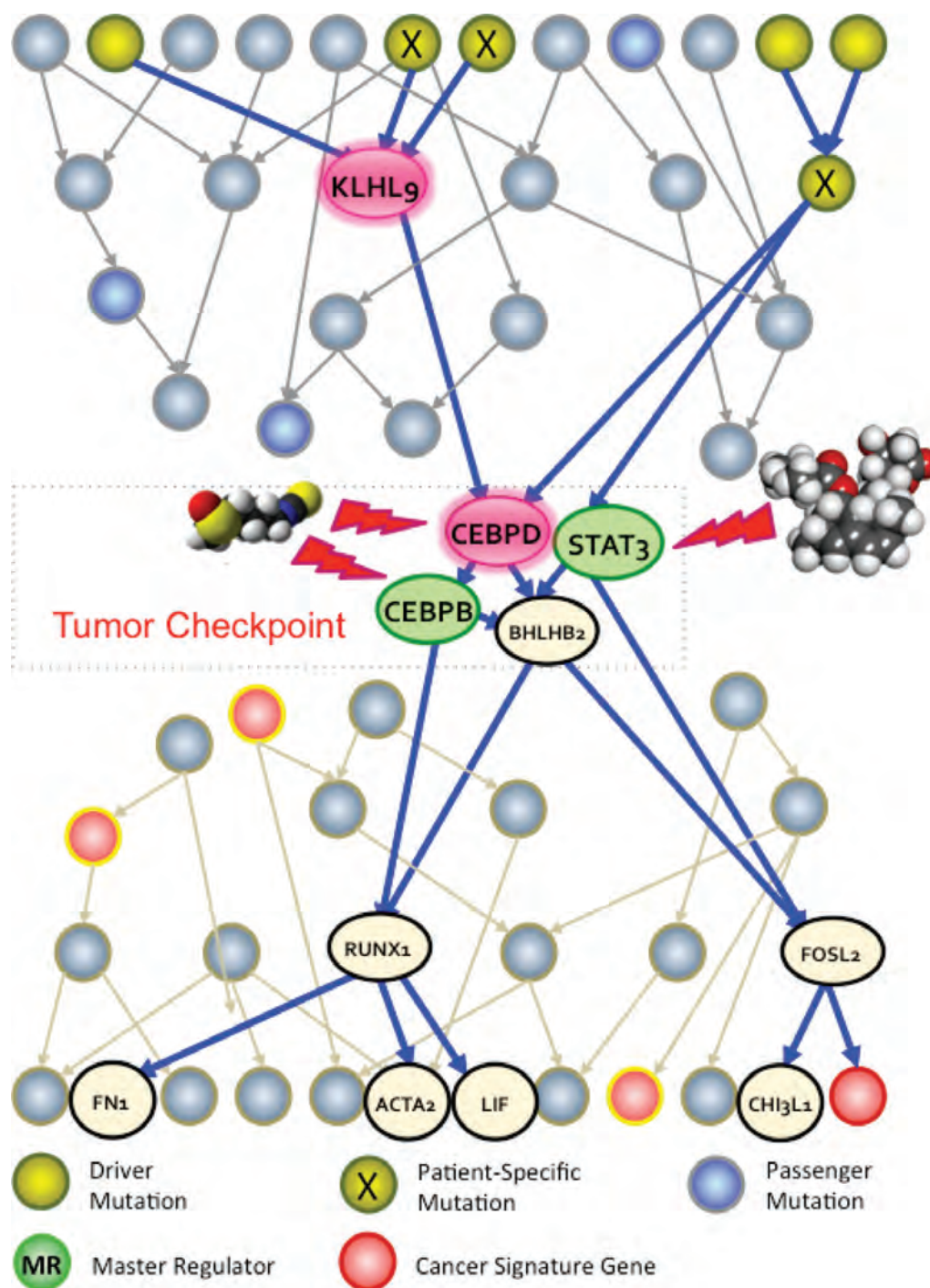
There are two important consequences stemming from the discovery of disease checkpoints:

First, *disease checkpoints provide a much more compact and manageable representation of disease states than genetic alterations*. Indeed, there are innumerable combinations of genetic alterations and variants that may induce abnormal activity of MR proteins in a disease checkpoint, including any combination of genetic and

epigenetic alterations in their upstream regulators<sup>2</sup>, in their cognate binding partners, and even in other molecular species including mRNAs and microRNAs<sup>19</sup>. As an example, a recent study by **DarwinHealth™** co-founders has shown that just two proteins (FOXMI and CENPF) provide an almost ideal determinant of aggressive prostate cancer, capturing at least 5 distinct genetic alterations patterns associated with poor prognosis and far outperforming any genomic classifier<sup>3</sup>.

Second, *disease checkpoints represent much more direct and universal disease dependencies, compared to genes harboring genetic alterations*. Indeed, MR proteins in disease checkpoints represent direct mechanistic regulators of the genetic programs necessary for disease implementation. As such, abnormally activated disease checkpoints not only represent better prognostic biomarkers but also more universal therapeutic targets, especially compared to the individual genetic alterations that contribute to their abnormal activity. For instance, since tumor checkpoints are downstream of bypass or alternative sub-clonal mutations that may induce drug resistance, their pharmacological targeting is less likely to induce relapse. Consistently, despite their mutational heterogeneity, disease cohorts only present a handful of distinct disease checkpoint mechanisms, thus supporting far simpler clinical study design.

Methods for the systematic elucidation of MR proteins<sup>8, 12</sup> have led to discovery, validation, and translation of many critical tumor checkpoints including, among others: (a) FOXMI and CENPF as the checkpoint of aggressive prostate cancer<sup>3, 15</sup>, with direct clinical application both in terms of prognostic biomarkers and therapeutic strategies, (b) CEBPB, CEBPD, and STAT3 as the checkpoint for the mesenchymal subtype of glioblastoma (GBM)<sup>2, 12</sup>, (c) AKT1 as the checkpoint for glucocorticoid



**Figure 11.**  
Master regulators and tumor checkpoints for mesenchymal glioma.

## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

resistance in T cell acute lymphoblastic leukemia (T-ALL)<sup>5</sup>, and (d) JAK2 and STAT3 as the checkpoint for trastuzumab resistance in HER2+/ER- breast cancer<sup>6</sup>.

Pharmacological targeting of the latter, using trastuzumab + ruxolitinib combination therapy, has led to striking response in a clinical trial for trastuzumab refractory patients (NCT02066532). Similarly, pharmacological targeting of tumor checkpoints discovered on an individual patient basis has become the foundation for a number of innovative N of 1 clinical studies at Columbia University, including a 260 patient landmark study for nine rare or untreatable malignancies.

Supporting the generality of this concept, **disease checkpoints have also been discovered in neurodegenerative diseases, including amyotrophic lateral sclerosis (ALS)<sup>3</sup>, Parkinson's (PD)<sup>4</sup>, and Alzheimer's Disease (AD)<sup>2, 20</sup>**, as well as in neurobehavioral syndromes, such as alcohol addiction<sup>22</sup>, and in control of stem cell pluripotency<sup>21</sup>. Thus, while initially targeting the oncology space, **DarwinHealth™**'s technology is equally relevant to a variety of additional human disease states.

Systematic discovery, characterization, and pharmacological targeting of disease checkpoints, using proprietary algorithms, represent the unifying strategy behind every **DarwinHealth™** product.

**Key Differentiating Features of DarwinHealth™ Technology Platform.** All established technologies to identify candidate targets in oncology analyze candidate genes one at the time, based on the presence of mutations, copy number alterations, differential methylation, or differential gene or protein expression in tumors. **DarwinCheckPoint™** is the only experimentally validated technology to directly identify the full repertoire of proteins (See Table 5), whose differential activity is causally responsible for tumor survival. Multiple, high-impact publications show that targeting these proteins either individually or as synergistic pairs leads to tumor checkpoint collapse and tumor regression in vivo or restores normal cellular function for non-tumor-related diseases.



DARWIN|OncoMatch™

and



DARWIN|OncoSynergy™

One of the most critical limitations of current high-throughput screening methodologies is that cell viability *in vitro* (i.e., outside of the human organism) is an extremely poor predictor of drug activity in a patient-relevant clinical context. In contrast, the mechanism of action (MoA) of drugs (i.e., the full complement of proteins that mediate and modulate their pharmacological activity and represent their targets and effectors) is relatively well conserved *in vitro* and *in vivo*. For instance, even though a compound may kill tumor cells *in vitro* with high specificity, the probability that it will yield a therapeutically relevant drug for the same tumor type in patients is relatively remote (<<10%). *Yet, if a compound is characterized as a potent inhibitor of a specific protein, it is very likely that its inhibitory activity will be preserved in a human tumor context.*

Conservation of drug mechanism of action represents the foundation for the **DarwinHealth™** approach to targeting MR proteins in tumor checkpoints. *As shown in several publications, genetic or pharmacological inhibition of either individual essential MRs or of synthetic lethal MR pairs within a tumor checkpoint leads to its complete collapse and to inversion of the aberrant activity of the individual MRs. That is, MR proteins that were aberrantly activated, such as oncogenes, will become inactivated while MR proteins that were aberrantly inactivated, such as tumor suppressors, will be activated.*

***As a result, the overall activity of MR proteins that comprise a specific tumor-checkpoint represents an ideal reporter assay to identify drugs and drug combinations capable of inducing tumor remission in vivo<sup>15, 16</sup>.***

By precisely assessing activity changes of all proteins in response to drug perturbations, **DarwinOncoMatch™** (See Table 6), allows identification of compounds that optimally invert the activity of MR proteins in any tumor checkpoint, leading to tumor regression in vivo. A complementary technology, **DarwinOncoSynergy™**, allows identification of drug pairs that invert the activity of complementary subsets of tumor checkpoint MRs, thus also leading to checkpoint collapse when used in combination. Both technologies represent the only published and thoroughly experimentally validated methodologies of this kind<sup>15, 16, 22</sup>.


The ability to match pre-screened compounds to patient specific MR programs, within minutes of receiving the tumor RNA Seq data, represents a remarkable achievement in precision cancer medicine. It allows attacking tumor viability mechanistically, across the full range of their molecular dependencies rather than by selecting a specific oncogene. The latter is far more likely to induce tumor relapse, for instance due to sub-clonal selection.

**Table 5. KEY DIFFERENTIATORS OF COMPETITIVE TECHNOLOGIES FOR CHARACTERIZATION OF CANDIDATE TUMOR DEPENDENCIES**

Features and Applicability	 <b>DARWIN</b> CheckPoint™	Mutations	mRNA expression	RPPA
Reproducibility	<b>High</b>	High	Low	Medium
Captures majority of tumor dependencies	<b>YES</b>			
Captures synergistic dependencies	<b>YES</b>			
Directly captures aberrant oncogene activity	<b>YES</b>			
Applicable to Minimal Tissue	<b>YES</b>		Yes	
Applicable to Single Cells	<b>YES</b>		Yes	
Applicable to Fresh Frozen Biopsies	<b>FULL</b>	Full	Full	Full
Applicable to FFPE	<b>FULL</b>	Partially	Poor	No
Independent of mutational analysis	<b>YES</b>		Yes	Yes
Cost	<b>LOW</b>	High	Low	Medium

**Table 5:** Key differentiators of competitor technologies for characterizing candidate tumor dependencies

**Table 6.** ADVANCED, “POST-ONCOGENE ADDICTION” PARADIGMS FOR IDENTIFICATION OF CANDIDATE TUMOR TARGETS

Features and Distinguishing Characteristics	 <b>DARWIN</b> OncoMatch™	Mutations
Captures aberrant activities of mutated oncogenes	<b>YES</b>	Yes
Predictive of compound activity	<b>HIGH</b>	Medium
Validated in pre-clinical contexts	<b>YES</b>	Yes
Validated in a clinical context	<b>PARTIAL/ ONGOING</b>	Yes
Identifies responders to targeted inhibitors that lack mutations	<b>YES</b>	No
Percent of applicable tumors	<b>100%</b>	25%
Applicable outside of oncology	<b>YES</b>	No
Cost	<b>Low</b>	High
Requires minimal tissue, down to single cells	<b>YES</b>	No
Predictive of synergistic inhibitor activity	<b>YES</b>	No
<i>Table 2: Key differentiators of competitor technologies for matching therapeutic agents to individual tumors</i>		

**Table 6:** Advanced, “Post-Oncogene Addiction” Paradigms for Identification of Candidate Tumor Targets



## DESCRIPTION of PROPRIETARY TECHNOLOGIES, CLINICAL APPLICATIONS, and MARKET POSITIONING

**Key Differentiating Points.** *There are currently no other established approaches to systematically and accurately match FDA-approved and investigational drugs to specific tumor dependencies, unless the compound is specifically designed to target the relevant oncogene. There are also no other validated predictive technologies to predict compound synergy based on tumor dependency analysis.*



A Foundational Element—Scientific,  
Clinical, and Commercial—of the  
DarwinOncoUniverse™ Technology Portfolio

The methods used to execute and generate **DarwinOncoMatch™** and **DarwinOncoSynergy™** analyses and recommendations are critically dependent on the availability of large-scale, proprietary drug perturbation databases, representing each specific tumor subtype of interest (**DarwinOncoDrugBase™**). For each tumor subtype of interest, this growing database represents the response of subtype-specific cell lines to FDA-approved and experimental compound perturbations, as measured by genome-wide RNA Seq profiling of these cells following compound perturbation at multiple concentrations and multiple timepoints, as well as the associated drug MAP, as inferred by **DarwinOncoDrugMoA** technology.


The current version of the **DarwinOncoDrugBase™** (See Table 7) includes drug response profiles for all subtypes of breast cancer (BRCA), the mesenchymal and proneural subtypes of glioblastoma (GBM), gastroenteropancreatic neuroendocrine tumors (GEP-NET), meningioma (MEN), GIST sarcoma (GSARC), pediatric neuroblastoma (NBL), and diffuse large B cell lymphoma (DLBCL). Additional tumor subtypes perturbations will be generated over the next six

months, including metastatic/aggressive colon adenocarcinoma (COAD), lung adenocarcinoma (LUAD), prostate adenocarcinoma (PRAD), and pancreatic ductal adenocarcinoma (PDAC).

**Key Differentiating Points: *DarwinOncoDrugBase™*** represents a radical departure from previous attempts to build large scale compound signatures, such as the Connectivity Map (CMap) and LINCS perturbational databases, for the following reasons:

1. **DarwinOncoDrugBase™** uses an advanced high-throughput, low-cost technology (PLATESeq), developed at Columbia University, to perform genome wide RNA Seq profiling of perturbed cells. This technology permits and achieves systematic coverage of large compound libraries across a large number of cell lines at multiple compound concentrations and time points. In contrast, the genome-wide profiling used by the CMap is too expensive to allow assembly of such databases and the reduced-coverage technology adopted by LINCS does not provide sufficiently descriptive signatures.
2. Rather than using fixed, isomolar compound concentrations, as done in CMap and LINCS, **DarwinOncoDrugBase™** starts from systematic dose-response curves to precisely assess the highest tolerated sub-lethal compound concentrations that are most informative of compound mechanism of action.
3. **DarwinOncoDrugBase™** relies on multiple time points, compound concentrations, and assay replicates to generate systematic and accurate elucidation of compound MoA, including direct target proteins and compound activity effectors and modulators<sup>1,16,17</sup>, thereby producing accurate predictions of compound activity *in vivo*.

**Table 7. KEY DIFFERENTIATING FEATURES AND CAPABILITIES OFFERED BY DARWINONCODRUGBASE™—NEW PARADIGMS AND DRUG PERTURBATION PROFILING**

Features and Distinguishing Characteristics	 DARWIN OncoDrugBase™	Connectivity Map	LINCS
Large-scale Compound Libraries	YES	YES	YES
Multiple Tumor Subtypes/Lines	>20	3	>10
Multiple Concentrations	YES		
Sub-lethal compound conc.	YES		
Genome-wide coverage	YES	YES	
Full Microfluidic Automation	YES		
Demonstrated MoA elucidation	YES		
Demonstrated in vivo activity prediction	YES		
Selected FDA approved and experimental compounds	YES		
Cost	LOW	HIGH	LOW

**Table 7:** Key Differentiating Features and Capabilities Offered by DarwinOncoDrugBase™—New Paradigms and Drug Perturbation Profiling



## REFERENCES

- Weinstein, I.B. Cancer. Addiction to oncogenes--the Achilles heel of cancer. *Science* **297**, 63-64 (2002).
- Chen, J.C. et al. Identification of Causal Genetic Drivers of Human Disease through Systems-Level Analysis of Regulatory Networks. *Cell* **159**, 402-414 (2014).
- Aytes, A. et al. Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* **25**, 638-651 (2014).
- Alvarez, M.J. et al. Network-based inference of protein activity disentangles the mutational and drug-response landscape of cancer. *Nat Genet (in review)* (2015).
- Piovan, E. et al. Direct reversal of glucocorticoid resistance by AKT inhibition in acute lymphoblastic leukemia. *Cancer Cell* **24**, 766-776 (2013).
- Rodriguez-Barrueco, R. et al. Inhibition of the autocrine loop IL6-JAK2-STAT3-Calprotectin as targeted therapy for HR-/HER2+ breast cancers *Genes & development* (2015).
- Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
- Lefebvre, C. et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol Syst Biol* **6**, 377 (2010).
- Compagno, M. et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* **459**, 717-721 (2009).
- Della Gatta, G. et al. Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL. *Nature medicine* **18**, 436-440 (2012).
- Palomero, T. et al. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc Natl Acad Sci U S A* **103**, 18261-18266 (2006).
- Carro, M.S. et al. The transcriptional network for mesenchymal transformation of brain tumours. *Nature* **463**, 318-325 (2010).
- Sonabend, A.M. et al. The transcriptional regulatory network of proneural glioma determines the genetic alterations selected during tumor progression. *Cancer research* **74**, 1440-1451 (2014).
- Lim, W., Lyashenko, E. & Califano, A. Master Regulators Used As Breast Cancer Metastasis Classifier. *Pac Symp Biocomp* **14**, 492-503, PMID: 19209726 (2009).
- Mitrofanova, A., Aytes, A., Shen, C., Abate-Shen, C. & Califano, A. A systems biology approach to predict drug response for human prostate cancer based on in vivo preclinical analyses of mouse models. *Cell reports* (2015 in review).
- Bansal, M. et al. A community computational challenge to predict the activity of pairs of compounds. *Nature Biotechnology* **32**, 1213-1222 (2014).
- Woo, J.H. et al. Elucidating Compound Mechanism of Action by Network Dysregulation Analysis in Perturbed Cells. *Cell*, (in press) (2015).
- Califano, A., Butte, A.J., Friend, S., Ideker, T. & Schadt, E. Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nat Genet* **44**, 841-847 (2012).
- Sumazin, P. et al. An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell* **147**, 370-381 (2011).
- Aubry, S. et al. Assembly and Interrogation of Alzheimer's disease genetic networks reveal novel regulators of progression. *PLoS One, in press* **10** (2015).
- Kushwaha, R. et al. Interrogation of a context-specific transcription factor network identifies novel regulators of pluripotency. *Stem Cells* **33**, 367-377 (2015).
- Repunte-Canonigo, V. et al. Identifying candidate drivers of alcohol dependence-induced excessive drinking by assembly and interrogation of brain-specific regulatory networks. *Genome biology* **16**, 68 (2015).
- Ikiz, B. et al. Dissecting the regulatory machinery of neurodegeneration in an in vitro model of amyotrophic lateral sclerosis. *Cell reports* (2015).
- Brichta, L. et al. Satb1 orchestrates an intrinsic network against dopaminergic neuron degeneration. *Nature neuroscience*, (in press) (2015).
- Chudnovsky, Y. et al. ZFH4 interacts with the NuRD core member CHD4 and regulates the glioblastoma tumor-initiating cell state. *Cell reports* **6**, 313-324 (2014).
- Zhao, X. et al. The N-Myc-DLL3 cascade is suppressed by the ubiquitin ligase Huwe1 to inhibit proliferation and promote neurogenesis in the developing brain. *Dev Cell* **17**, 210-221 (2009).

# The transcriptional network for mesenchymal transformation of brain tumours

Maria Stella Carro<sup>1\*†</sup>, Wei Keat Lim<sup>2,3\*†</sup>, Mariano Javier Alvarez<sup>3,4\*</sup>, Robert J. Bollo<sup>8</sup>, Xudong Zhao<sup>1</sup>, Evan Y. Snyder<sup>9</sup>, Erik P. Sulman<sup>10</sup>, Sandrine L. Anne<sup>1†</sup>, Fiona Doetsch<sup>5</sup>, Howard Colman<sup>11</sup>, Anna Lasorella<sup>1,5,6</sup>, Ken Aldape<sup>12</sup>, Andrea Califano<sup>1,2,3,4</sup> & Antonio Iavarone<sup>1,5,7</sup>

**The inference of transcriptional networks that regulate transitions into physiological or pathological cellular states remains a central challenge in systems biology. A mesenchymal phenotype is the hallmark of tumour aggressiveness in human malignant glioma, but the regulatory programs responsible for implementing the associated molecular signature are largely unknown. Here we show that reverse-engineering and an unbiased interrogation of a glioma-specific regulatory network reveal the transcriptional module that activates expression of mesenchymal genes in malignant glioma. Two transcription factors (C/EBP $\beta$  and STAT3) emerge as synergistic initiators and master regulators of mesenchymal transformation. Ectopic co-expression of C/EBP $\beta$  and STAT3 reprograms neural stem cells along the aberrant mesenchymal lineage, whereas elimination of the two factors in glioma cells leads to collapse of the mesenchymal signature and reduces tumour aggressiveness. In human glioma, expression of C/EBP $\beta$  and STAT3 correlates with mesenchymal differentiation and predicts poor clinical outcome. These results show that the activation of a small regulatory module is necessary and sufficient to initiate and maintain an aberrant phenotypic state in cancer cells.**

High-grade gliomas (HGGs) are the most common brain tumours in humans and are essentially incurable<sup>1</sup>. The defining hallmarks of the aggressiveness of glioblastoma multiforme (GBM) are local invasion and neo-angiogenesis<sup>2,3</sup>. A recently established notion postulates that neoplastic transformation in the central nervous system (CNS) converts neural cells into cell types manifesting a mesenchymal phenotype—a state associated with an uncontrolled ability to invade and stimulate angiogenesis<sup>4,5</sup>. Gene expression studies have established that overexpression of a ‘mesenchymal’ gene expression signature (MGES) and loss of a proneural signature (PNGES) co-segregate with the poor prognosis group of glioma patients<sup>4</sup>. Yet, differentiation along the mesenchymal lineage is virtually undetectable in normal neural tissue during development. Thus, it is unclear whether drift towards the mesenchymal lineage is an aberrant event that occurs during brain tumour progression or whether glioma cells recapitulate the rare mesenchymal plasticity of neural stem cells (NSCs)<sup>4–7</sup>. The molecular events that activate the MGES while suppressing the PNGES signature, thus imparting a highly aggressive phenotype to glioma cells, remain unknown.

Efforts to identify transcription factors (TFs) that are master regulators (MRs) of specific cancer signatures, on the basis of cellular-network models, have yet to produce experimentally validated discoveries, probably because these networks are still poorly mapped, especially within specific mammalian cellular contexts<sup>8</sup>. Notwithstanding, recent developments in genome-wide reverse engineering were successful in identifying causal, rather than associative

interactions<sup>9–12</sup>, and showed promise in the identification of dysregulated genes within developmental and tumour-related pathways<sup>13–17</sup>. Thus, we reasoned that context-specific regulatory networks, inferred by unbiased reverse engineering algorithms, may provide sufficient accuracy to allow estimation of (1) the activity of TFs from that of their transcriptional targets or regulons, and (2) the identity of TFs that are MRs of specific eukaryotic signatures<sup>18,19</sup> from the overlap between their regulons and the signatures. We applied the above mechanisms to unravel the MRs causally linked to activation of the MGES in malignant glioma (Supplementary Fig. 1).

## A transcriptional module is linked to the MGES of HGGs

We first addressed whether copy number variation may account for the aberrant expression of MGES genes in HGGs. Integrated analysis of gene expression profiles and array comparative genomic hybridization (aCGH) of 76 HGGs showed no correlation between mean expression and DNA copy number of MGES genes in proneural, mesenchymal and proliferative tumours (Supplementary Fig. 2).

We thus used the ARACNe algorithm<sup>9</sup> to assemble a genome-wide repertoire of HGG-specific transcriptional interactions (the HGG-interactome) from 176 gene expression profiles of grade III (anaplastic astrocytoma) and grade IV (GBM) samples previously classified into three molecular signature groups: proneural, proliferative and mesenchymal (Supplementary Table 1a–c)<sup>4,20,21</sup>. ARACNe is an information theoretical approach for the inference of TF-target interactions from large sets of gene expression profiles<sup>9,16</sup>, further refined to

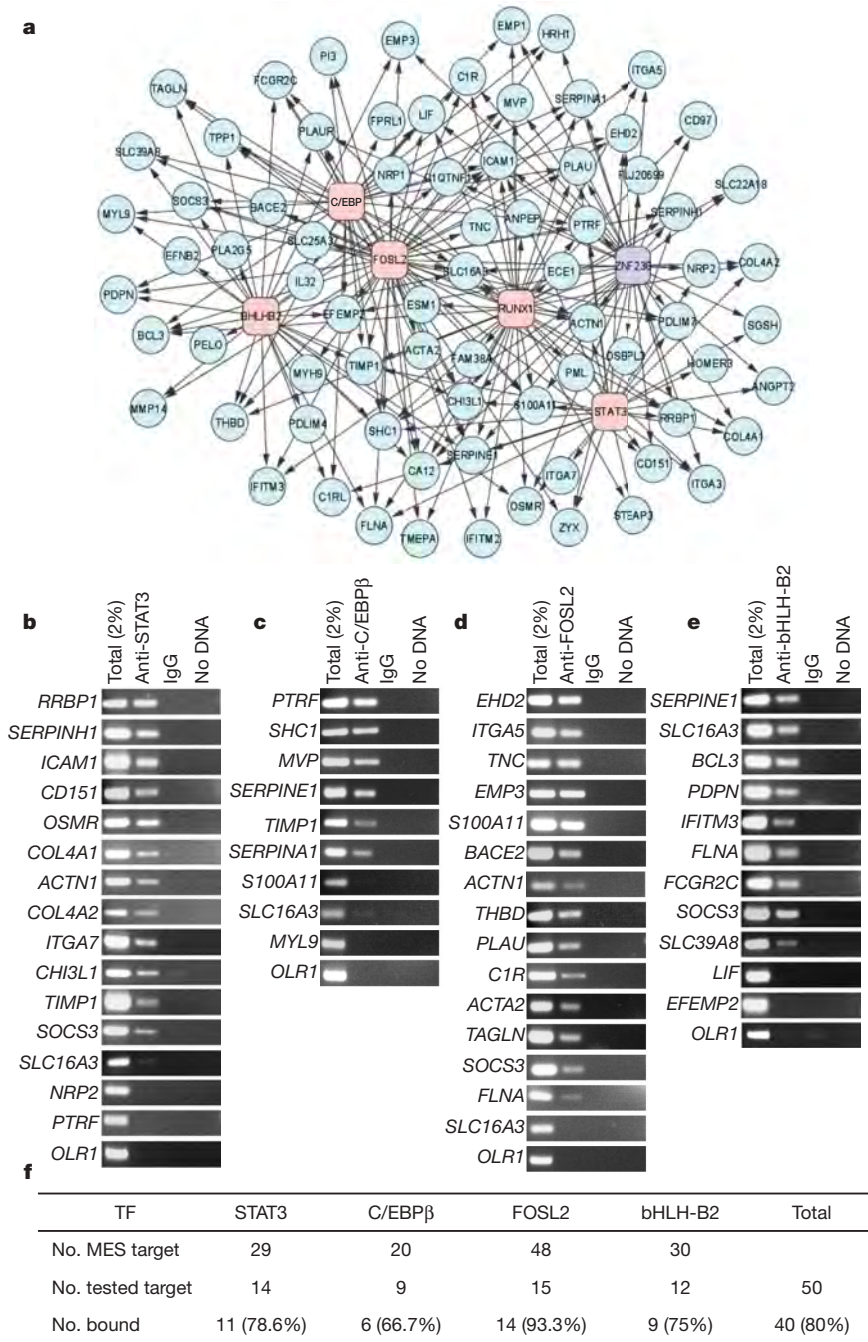
<sup>1</sup>Institute for Cancer Genetics, <sup>2</sup>Department of Biomedical Informatics, <sup>3</sup>Center for Computational Biology and Bioinformatics, <sup>4</sup>Joint Centers for Systems Biology, <sup>5</sup>Department of Pathology, <sup>6</sup>Department of Pediatrics, <sup>7</sup>Department of Neurology, Columbia University Medical Center, New York, New York 10032, USA. <sup>8</sup>Department of Neurosurgery, New York University School of Medicine & NYU Langone Medical Center, New York, New York 10016, USA. <sup>9</sup>Burnham Institute for Medical Research, La Jolla, California 92037, USA. <sup>10</sup>Division of Radiation Oncology, <sup>11</sup>Department of Neuro-Oncology, <sup>12</sup>Department of Pathology, M.D. Anderson Cancer Center, Houston, Texas 77030, USA. †Present addresses: Department of General Neurosurgery, Neurocenter and Comprehensive Cancer Research Center, University of Freiburg, Breisacher Str. 64, D-79106 Freiburg, Germany (M.S.C.); Therasis, Inc., 462 First Avenue, Suite 908, New York, New York 10016, USA (W.K.L.); Rockefeller University, RRB 750, 1230 York Avenue, New York, New York 10065, USA (S.L.A.).

\*These authors contributed equally to this work.

determine directed (that is, causal) interactions<sup>12,22</sup> (see Methods). ARACNe predicted 92,660 transcriptional interactions, 1,217 of which were between TFs and 102 out of 149 MGES genes<sup>4</sup>, represented across all the gene expression profile data.

Next, we applied a new master regulator analysis (MRA) algorithm to the HGG-interactome. The algorithm computes the statistical significance of the overlap between the regulon of each TF (that is, its ARACNe-inferred targets) and the MGES genes (*P* values computed by Fisher's exact test, FET). From a list of 928 TFs (Supplementary Table 2), MRA inferred 53 MGES-specific TFs, at a false discovery rate (FDR) < 5% (Supplementary Table 3a). These were ranked on the

basis of the total number of MGES targets they regulated. The top six TFs (STAT3, C/EBP, bHLH-B2 (also known as bHLHE40), RUNX1, FOSL2 and ZNF238) collectively controlled >74% of the MGES genes (Fig. 1a). C/EBP $\beta$  and C/EBP $\delta$  were grouped as they form stoichiometric homo- and heterodimers with identical DNA-binding specificity and redundant transcriptional activity<sup>23</sup>. We thus use the term C/EBP to indicate the TF complex with the union of their targets as the corresponding regulon. Consistent with their previously reported activity<sup>24,25</sup>, Spearman's correlation analysis showed that five of these TFs are probably activators (STAT3, C/EBP, bHLH-B2, RUNX1 and FOSL2) and one is probably a repressor (ZNF238).



**Figure 1 | The mesenchymal signature of HGGs is controlled by six TFs.** **a**, TFs involved in activation of MGES targets are shown in pink, those involved in repression are in purple. MGES targets controlled by these TFs are in cyan. Overall, the six TFs control 74% of the genes in the mesenchymal signature of high-grade glioma. A region between 2 kb upstream and downstream the transcription start site of the target genes identified by ARACNe was analysed for the presence of putative binding sites.

**b–e**, Genomic regions of genes containing putative binding sites for specific TFs were immunoprecipitated in SNB75 cells by antibodies specific for STAT3 (**b**), C/EBP $\beta$  (**c**), FOSL2 (**d**), and bHLH-B2 (**e**). SOCS3 was included as positive control of STAT3 binding. Total chromatin before immunoprecipitation was used as positive control for PCR. The *OLR1* gene was used as negative control. **f**, Summary of binding results of the tested TFs to mesenchymal targets.



Overlap between the regulons of these TFs was highly significant (Supplementary Table 4). MRA of the PNGES and Proliferative (PROGES) signatures of HGGs detected virtually no overlap among candidate MRs of the three signatures, with the notable exception of few TFs inversely associated with MGES and PNGES activation (Supplementary Table 5).

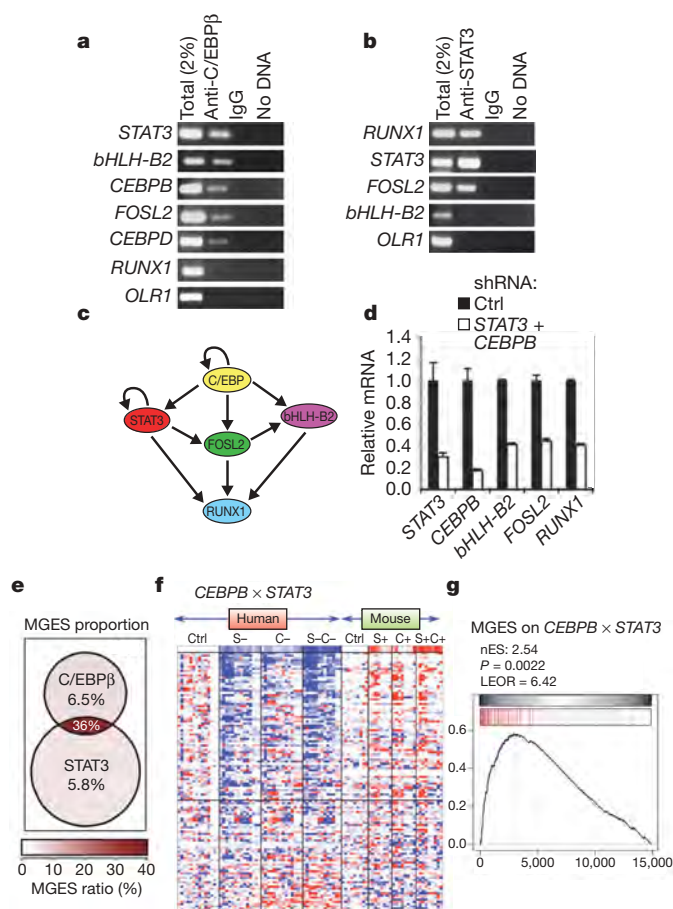
Next, we used stepwise linear regression to determine simple, quantitative regulation models for each MGES gene. Specifically, the log-expression of each MGES gene was fitted by the linear combination of the log-expression of a small number of genes (1–5; see Methods), selected among 53 ARACNe-inferred and 52 additional TFs, whose DNA-binding signatures were enriched in the promoters of MGES genes. Six TFs were in both lists, for a total of 99 TFs (Supplementary Table 3b). TFs were then ranked on the basis of the fraction of MGES genes they regulated. Surprisingly, the top six MRA-inferred TFs were also among the eight controlling the largest number of MGES targets, based on stepwise linear regression analysis (Supplementary Table 6). Indeed, the three with the highest linear-regression coefficient values were C/EBP ( $\alpha = 0.40$ ), bHLH-B2 ( $\alpha = 0.41$ ) and STAT3 ( $\alpha = 0.40$ ), further establishing them as probable MGES-MR candidates. The next strongest TF, ZNF238, had a negative coefficient ( $\alpha = -0.34$ ) confirming its role as a candidate MGES repressor.

### Validation of the mesenchymal regulatory module

To determine whether these TFs bind the promoter region of their predicted MGES targets, we performed chromatin immunoprecipitation (ChIP) in a human glioma cell line. On average, TF-specific antibodies but not control antibodies immunoprecipitated 80% of the tested genomic regions (Fig. 1b–f). Lentivirus-mediated short hairpin RNA (shRNA) silencing of *CEBPB*, *STAT3*, *bHLH-B2*, *FOSL2* and *RUNX1* in glioma cells, followed by gene expression profiling and Gene Set Enrichment Analysis (GSEA), showed that after silencing of each TF, differentially expressed genes were highly enriched in their ARACNe-inferred targets but not in those of control TFs with equivalent regulon size (Supplementary Table 7a). Furthermore, differentially expressed genes were also enriched in MGES genes (Supplementary Table 7b).

Promoter occupancy analysis revealed a hierarchical and highly modular topology, with eight out of ten possible intra-module interactions implemented (modularity  $P = 1.0 \times 10^{-8}$  by FET, Fig. 2c). Specifically, CEBP $\beta$  and STAT3 occupy their own promoter (Fig. 2a, b); C/EBP $\beta$  occupies the *STAT3*, *FOSL2*, *bHLH-B2*, *CEBPB* and *CEBPD* promoters (Fig. 2a); STAT3 occupies those of *FOSL2* and *RUNX1* (Fig. 2b); *FOSL2* occupies those of *RUNX1* and *bHLH-B2* (Supplementary Fig. 3a), and bHLH-B2 only occupies the *RUNX1* promoter (Supplementary Fig. 3b). C/EBP and STAT3 are at the top of this hierarchical regulatory module. They have autoregulatory loops and form feed-forward loops with a larger fraction of MGES genes (43%) than any of the other TF pairs. shRNA-mediated co-silencing of C/EBP $\beta$  and STAT3 in glioma cells produced >2-fold reduction of the messenger RNAs coding for the second layer TFs in the feed-forward loops (*bHLH-B2*, *FOSL2* and *RUNX1*; Fig. 2d), supporting their role as MRs. C/EBP $\beta$  and STAT3 also bound the promoters of their MGES targets in primary human GBM (Supplementary Fig. 3c, d).

To validate functionally the role of C/EBP $\beta$  and STAT3 as MRs of the MGES, we conducted gain- and loss-of-function experiments. We transduced *v-myc* immortalized mouse NSCs (mNSCs) known as C17.2 (refs 26–28), as well as primary mNSCs derived from the mouse telencephalon at embryonic day (E)13.5, with retroviruses expressing C/EBP $\beta$  and a constitutively active form of STAT3 (STAT3C)<sup>29</sup>. shRNA-mediated silencing targeted C/EBP $\beta$  and STAT3 in the human glioma cell line SNB19, and in serum-free cultures of tumour cells derived from primary GBM that propel the formation of GBM-like tumours after intracranial transplantation in immunodeficient mice<sup>30</sup> (GBM-derived brain tumour initiating cells, GBM-BTICs; see later).



**Figure 2 | A hierarchical transcriptional module regulates the MGES.**

**a, b**, ChIP for C/EBP $\beta$  (**a**) and STAT3 (**b**). **c**, Transcriptional network emerging from promoter occupancy analysis. **d**, qRT-PCR of mesenchymal TFs in glioma cells infected with STAT3 and CEBPB shRNA or control (ctrl) lentiviruses. Error bars are s.d. **e**, Venn-diagram depicts the proportion of mesenchymal genes identified by ARACNe as targets of only C/EBP $\beta$ , STAT3 or both TFs. **f**, Heatmap of MGES gene expression analysis of mouse and human cells carrying perturbations of C/EBP $\beta$  plus STAT3. Samples (columns) were grouped according to species and treatment. Control, control shRNA or empty vector; S–, STAT3 knockdown; S+, STAT3 overexpression; C–, CEBPB knockdown; C+, CEBPB overexpression; S–/C–, STAT3 and CEBPB knockdown; S+/C+, STAT3 and CEBPB overexpression. **g**, GSEA of the MGES on the gene expression profile rank-sorted according to the correlation with the CEBPB  $\times$  STAT3 metagene. The bar-code plot indicates the position of MGES genes, red and blue colours represent positive and negative correlation, respectively. The grey scale bar indicates the Spearman's rho coefficient used as weighting score for GSEA. LEOR, leading-edge odds ratio; nES, normalized enrichment score; *P*, sample-permutation-based *P* value.

We generated a global data set of 89 individual samples, including 55 knockdown experiments in human glioma cells and 34 ectopic expression experiments in mouse NSCs. Of the 149 genes in the MGES, 118 could be mapped to murine genes represented on the array (Supplementary Table 8). Quantitative PCR with reverse transcription (qRT-PCR) analysis showed that, after CEBPB or STAT3 shRNA silencing in GBM-BTICs and SNB19 cells, the corresponding mRNA levels were significantly reduced compared to non-target control transduced cells (CEBPB fold ratio = 0.26,  $P \leq 0.00108$ , STAT3 fold ratio = 0.205,  $P \leq 0.00109$  by U-test). Reciprocal changes followed ectopic expression of the two TFs in C17.2 cells and NSCs (Supplementary Table 9). qRT-PCR values and microarray-based measurements were highly correlated for STAT3 but not for CEBPB mRNA (Supplementary Fig. 4). Thus, we used the qRT-PCR values for CEBPB and STAT3 as more accurate read-outs for their mRNA expression. GSEA confirmed that genes co-expressed with the two TFs

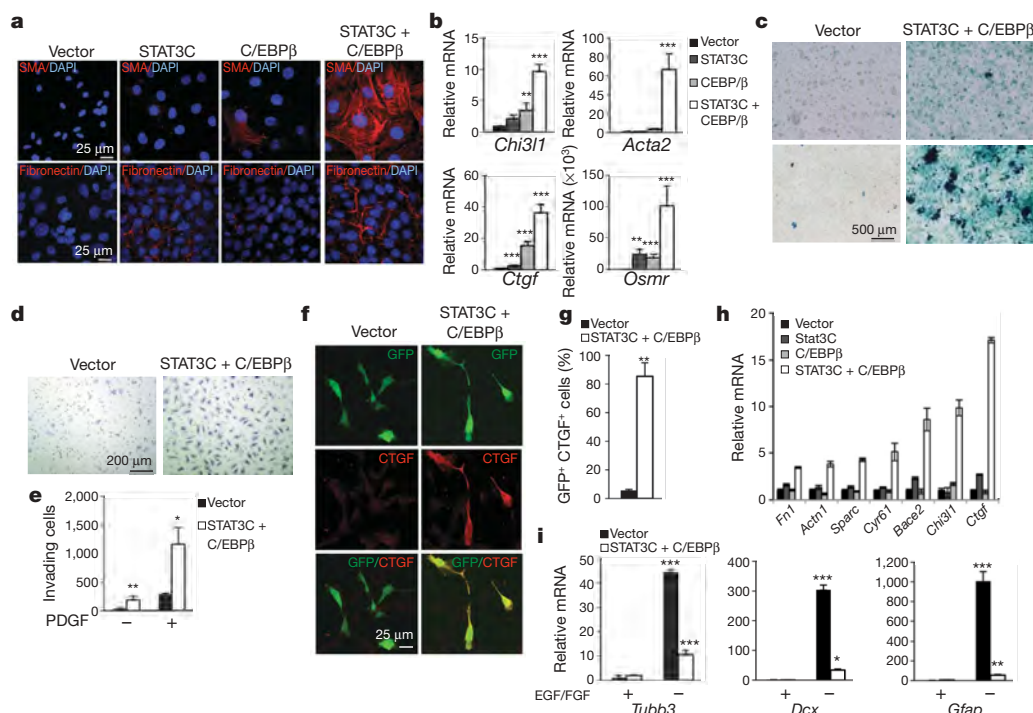
were enriched in their respective ARACNe-inferred regulon genes but not in those of control TFs (Supplementary Table 10). Perturbation of C/EBP $\beta$  (Supplementary Fig. 5a, c) or STAT3 (Supplementary Fig. 5b, d) specifically affected the MGES signature ( $P = 2.69 \times 10^{-2}$  and  $P = 2.0 \times 10^{-4}$ , respectively, by GSEA). Common targets of C/EBP and STAT3 were eightfold more enriched in MGES genes than targets controlled individually by each TF (Fig. 2e,  $P = 2.25 \times 10^{-5}$ ). To test whether the two TFs may be involved in synergistic MGES control, we computed a metagene (CEBPB  $\times$  STAT3) with expression proportional to the product of their mRNAs, such that the metagene should be highly correlated with the expression profile of any target synergistically regulated by the two TFs, under a multiplicative model (Fig. 2f). GSEA confirmed that genes ranked by Spearman's correlation to the CEBPB  $\times$  STAT3 metagene were significantly enriched in MGES genes (Fig. 2g), suggesting that at least a subset of the MGES is synergistically regulated by the CEBPB  $\times$  STAT3 pair.

We sought to establish (1) whether MRs inferred by our procedure would also be inferred when using an independent glioma sample data set, and (2) whether MRs identified on the basis of clinical outcome would overlap significantly with those inferred from MGES analysis. The Atlas-TCGA data set<sup>31</sup> includes 77 and 21 samples associated with worst- and best-prognosis, respectively (92 samples with intermediate prognosis were not considered). Differential expression analysis identified a TCGA worst-prognosis signature (TWPS), comprising 884 genes differentially expressed in the worst-prognosis versus best-prognosis samples (at  $P \leq 0.05$  by Student's *t*-test, Supplementary Table 11). GSEA confirmed that MGES genes were markedly enriched in the TWPS signature ( $P \leq 1.0 \times 10^{-4}$ , Supplementary Fig. 6) indicating that the poor-prognosis group in the Atlas-TCGA data set displays marked mesenchymal features. Despite partial overlap between

MGES and TWPS genes (22.8%), five of the six MRs identified by MRA from the original data set were also found among the ten most significant TFs identified by MRA of the Atlas-TCGA data set using the TWPS signature. C/EBP was the most significant TF, and STAT3 was in seventh position. C/EBP $\beta$  and C/EBP $\delta$  had the first and second best linear-regression coefficient by stepwise linear regression analysis, respectively (Supplementary Table 12). These results indicate significant robustness of the approach both to data set and signature selection.

### Mesenchymal reprogramming of NSCs by C/EBP $\beta$ and STAT3

We tested whether combined and/or individual expression of STAT3C and C/EBP $\beta$  in NSCs is sufficient to trigger the mesenchymal phenotype that characterizes HGGs. The introduction of C/EBP $\beta$  and STAT3C into C17.2 NSCs caused loss of neuronal differentiation and manifestation of a fibroblast-like morphology (Supplementary Fig. 7a, b). The morphological changes were associated with gain of expression of the mesenchymal marker proteins smooth muscle alpha actin (SMA, encoded by *Acta2*) and fibronectin (Fig. 3a and Supplementary Fig. 7c, d), and with induced expression of the mesenchymal genes *Chi3l1* (also known as *Ykl40*), *Acta2*, *Ctgf* and *Osmr* (Fig. 3b). The individual expression of STAT3C or C/EBP $\beta$  was generally insufficient to induce expression of either mesenchymal marker proteins or mesenchymal genes (Fig. 3a, b). Removal of mitogens to STAT3C plus C/EBP $\beta$ -expressing C17.2 cells resulted in a further increase in the expression of mesenchymal genes, and complete acquisition of mesenchymal properties such as positive alcian blue staining, a specific assay for chondrocyte differentiation (Fig. 3c and Supplementary Fig. 7e, f). The expression of STAT3C and C/EBP $\beta$  promoted migration in a wound assay, and triggered invasion through



**Figure 3 | Ectopic expression of C/EBP $\beta$  and STAT3C in NSCs induces mesenchymal transformation and inhibits neural differentiation.**

**a**, Immunofluorescence analysis for SMA and fibronectin in C17.2 cells expressing the indicated TFs. DAPI, 4',6-diamidino-2-phenylindole. **b**, qRT-PCR of mesenchymal targets in C17.2 cells expressing the indicated TFs.  $n = 3$ ; error bars are mean  $\pm$  s.d. **c**, Alcian blue staining of C17.2 cells expressing STAT3C and C/EBP $\beta$  or the empty vector cultured in growth medium (top), or chondrocyte differentiation medium (bottom). **d**, Microphotographs of invading C17.2 cells expressing STAT3C plus C/EBP $\beta$  or empty vector. **e**, Quantification of invading cell in the absence or

presence of PDGF.  $n = 3$ ; error bars denote mean  $\pm$  s.e.m.

**f**, Immunofluorescence analysis for CTGF in NSCs expressing STAT3C and C/EBP $\beta$  or the empty vector. Green fluorescent protein (GFP) identifies infected cells. **g**, Quantification of GFP $^{+}$  CTGF $^{+}$  cells. Error bars denote mean  $\pm$  s.d. of three independent experiments. **h**, qRT-PCR of representative mesenchymal genes in primary NSCs expressing the indicated TFs.  $n = 3$ ; error bars indicate mean  $\pm$  s.d. **i**, qRT-PCR of *Tubb3*, *Dcx* and *Gfap* in NSCs expressing STAT3C plus C/EBP $\beta$  or the empty vector.  $n = 3$ ; error bars indicate mean  $\pm$  s.d. qRT-PCR data are normalized to fold changes in 18S ribosomal RNA. \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ .



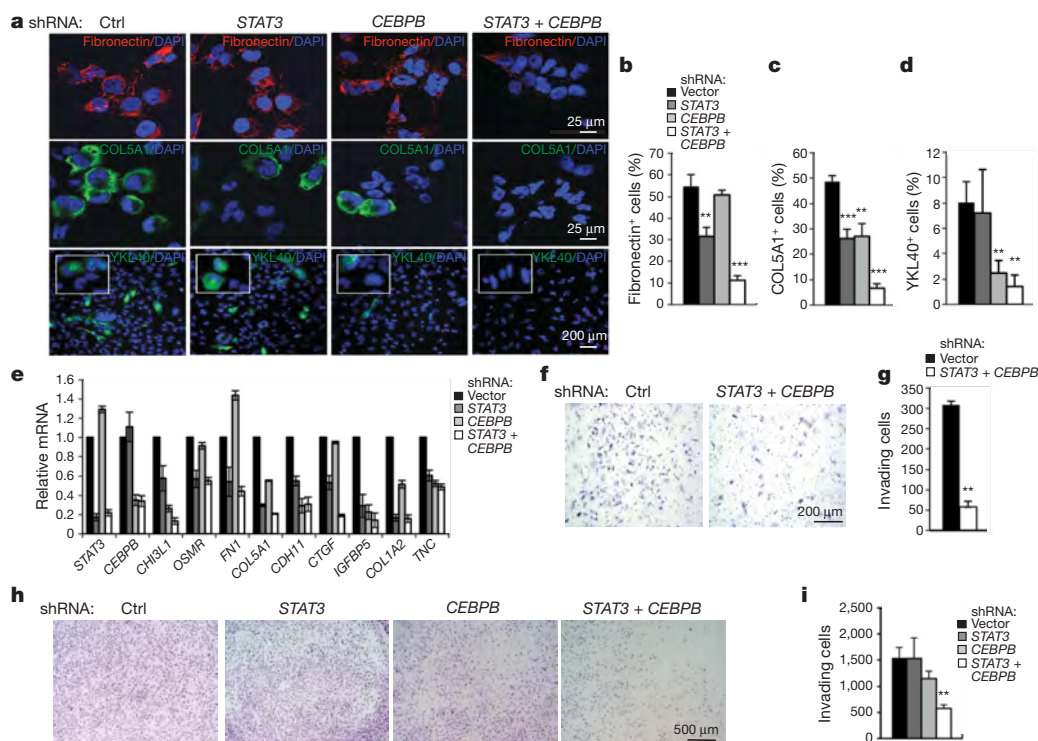
the extracellular matrix in a Matrigel invasion assay in the absence or presence of PDGF (Fig. 3d, e and Supplementary Fig. 7g). The combined but not individual expression of STAT3C and C/EBP $\beta$  efficiently induced mesenchymal marker proteins and mesenchymal gene expression in primary NSCs (Fig. 3f–h). Conversely, STAT3C and C/EBP $\beta$  abolished differentiation along the neuronal and glial lineages (Fig. 3i and Supplementary Fig. 7h). The C/EBP $\beta$ - and STAT3C-induced mesenchymal transformation of primary NSCs was associated with withdrawal from the cell cycle (data not shown). Thus, the combined introduction of C/EBP $\beta$  and STAT3C in NSCs prevents neural differentiation and triggers reprogramming towards an aberrant mesenchymal lineage.

### Mesenchymal aggressiveness of glioma by C/EBP $\beta$ and STAT3

Transduction of GBM-BTIC cultures derived from two GBM patients (BTSC-20 and BTSC-3408) with specific shRNA-carrying lentiviruses silenced endogenous C/EBP $\beta$  and STAT3, eliminated expression of mesenchymal genes, and depleted the tumour cells of the mesenchymal marker proteins fibronectin, collagen-5 $\alpha$ 1 and YKL40. Individual silencing of C/EBP $\beta$  or STAT3 produced variable inhibitory effects, with the silencing of C/EBP $\beta$  typically carrying the most severe consequences (Fig. 4a–e and Supplementary Fig. 8a). Combined or individual silencing of C/EBP $\beta$  and STAT3 in the human glioma cell line SNB19 produced similar effects (Supplementary Fig. 8b–e). Silencing of the two TFs in SNB19 cells and GBM-BTICs reduced their ability to invade through Matrigel by >70% (Fig. 4f–i). Next, we determined the effect of CEBPB and STAT3 knockdown on brain tumorigenesis after intracranial injection of SNB19 cells in immunocompromised mice. We observed efficient tumour formation in all mice injected with control or STAT3 shRNA cells. However, only one of four mice from the

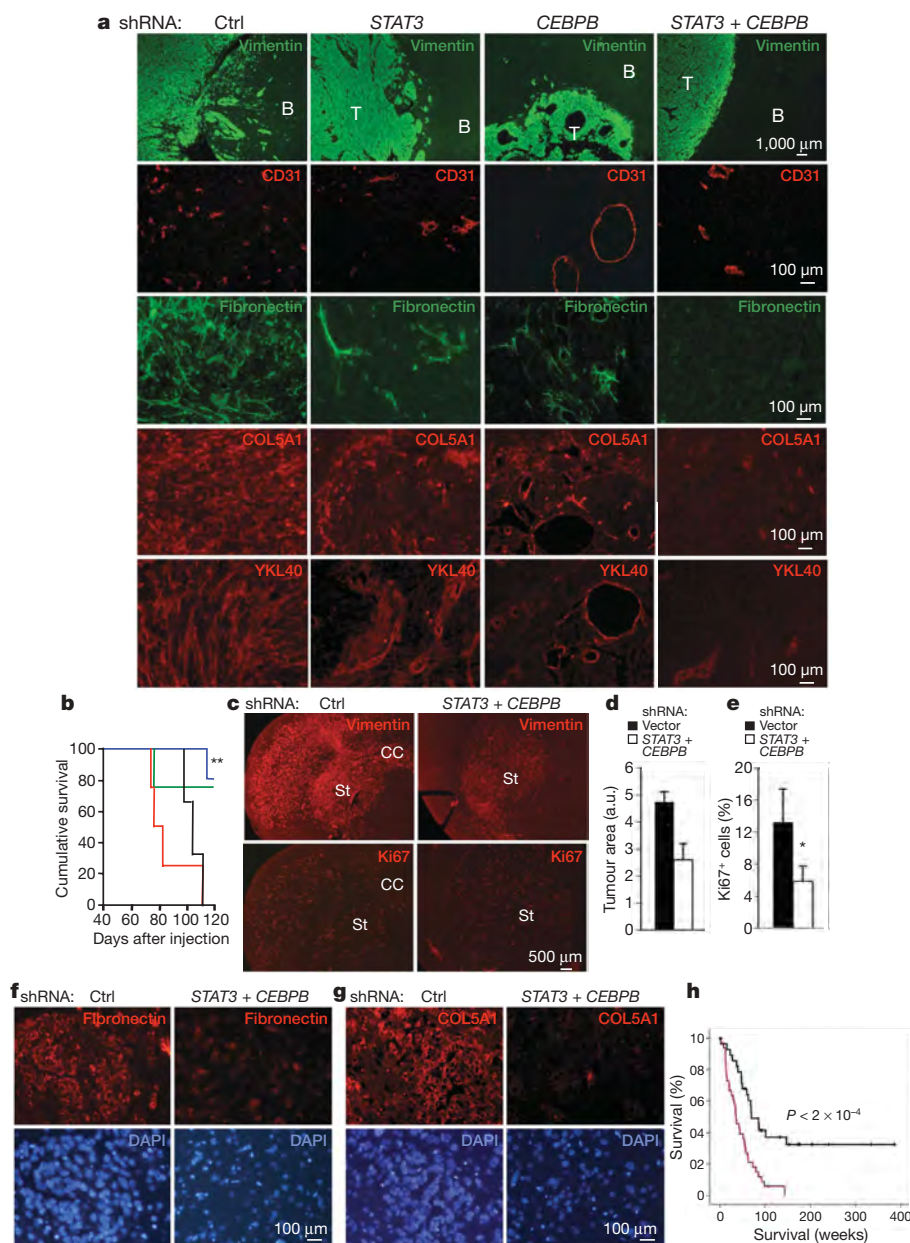
CEBPB shRNA and one of five mice from the combined CEBPB and STAT3 shRNA groups developed tumours 120 days after the injection (Fig. 5b). The histological analysis demonstrated high-grade tumours, which displayed peripheral invasion of the surrounding brain as single cells and cell clusters in the shRNA control group, as shown by anti-human vimentin staining (Fig. 5a). Staining for the endothelial marker CD31 revealed marked vascularization in the control shRNA tumour group. Conversely, the single tumour in the CEBPB plus STAT3 shRNA group grew well circumscribed and was less angiogenic. Tumours in the STAT3 shRNA group and the single tumour in the CEBPB shRNA group had an intermediate growth pattern and limited angiogenesis. Staining for fibronectin, collagen-5 $\alpha$ 1 and YKL40 was readily detected in the tumours from the control group but absent or barely detectable in the single tumours from the CEBPB shRNA and CEBPB plus STAT3 shRNA groups (Fig. 5a). Tumours derived from STAT3 shRNA cells showed an intermediate phenotype, with reduced expression of mesenchymal markers compared with tumours in the control shRNA group, but higher than that observed in the tumours in the CEBPB shRNA and CEBPB plus STAT3 shRNA groups (control shRNA > STAT3 shRNA > CEBPB shRNA > CEBPB+STAT3 shRNA). Intracranial transplantation of GBM-BTICs transduced with control shRNA lentivirus produced extremely invasive tumour cell masses extending through the corpus callosum to the contralateral brain. Combined knockdown of CEBPB and STAT3 led to a significant decrease in the tumour area and tumour cell density (as evaluated by human vimentin staining), markedly reduced the proliferation index (Fig. 5c–e), and abolished expression of the mesenchymal markers fibronectin and collagen-5 $\alpha$ 1 (Fig. 5f, g).

Finally, we conducted immunohistochemical analysis for C/EBP $\beta$  and active, phospho-STAT3 in human tumour specimens, and



**Figure 4 | C/EBP $\beta$  and STAT3 maintain the mesenchymal phenotype of human glioma cells.** **a**, Immunofluorescence for fibronectin, collagen-5 $\alpha$ 1 (COL5A1) and YKL40 in BTSC-3408 infected with lentiviruses expressing STAT3, CEBPB or STAT3 plus CEBPB shRNA. **b–d**, Quantification of cells positive for fibronectin (**b**), COL5A1 (**c**) and YKL40 (**d**). **e**, qRT-PCR of mesenchymal genes in BTSC-20 infected with lentiviruses expressing STAT3, CEBPB or STAT3 plus CEBPB shRNA. Gene expression was normalized to 18S ribosomal RNA expression.  $n = 3$ ; error bars indicate mean  $\pm$  s.d.

**f**, Microphotographs of invading SNB19 cells infected with lentiviral vectors expressing control or STAT3 plus CEBPB shRNA. **g**, Quantification of SNB19-invading cells. Error bars indicate mean  $\pm$  s.d.;  $n = 6$  (two independent experiments, each performed in triplicate). **h**, Invading BTSC-3408 cells infected with control, STAT3, CEBPB or STAT3 plus CEBPB shRNA lentiviruses. **i**, Quantification of invading BTSC-3408 cells. Error bars indicate mean  $\pm$  s.d.;  $n = 6$  (two independent experiments, each performed in triplicate). \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ .



**Figure 5 | C/EBP $\beta$  and STAT3 are essential for glioma tumour aggressiveness in mice and humans.** **a**, Immunofluorescence staining for human vimentin, CD31, fibronectin, COL5A1 and YKL40 in tumours derived from SNB19 cells infected with lentiviruses expressing shRNA targeting STAT3, CEBPB, or STAT3 plus CEBPB. B, normal brain; T, tumour. **b**, Kaplan–Meier survival curve of NOD/SCID mice transplanted intracranially with SNB19 glioma cells transduced with control shRNA (red), STAT3 shRNA (black), CEBPB shRNA (green) or STAT3 plus CEBPB shRNA (blue) lentiviruses. **c**, Immunostaining for human vimentin and Ki67 on

representative brain sections from mice injected with BTSC-3408 after silencing of C/EBP $\beta$  and STAT3. CC, corpus callosum; St, striatum. **d**, Quantification of human vimentin-positive area. a.u., arbitrary units. Error bars are s.d. **e**, Quantification of Ki67-positive cells.  $n = 5$  for each group; error bars are mean  $\pm$  s.d. **f**, **g**, Immunostaining for fibronectin (**f**) and COL5A1 (**g**) on representative brain sections from mice injected with BTSC-3408 transduced as indicated. **h**, Kaplan–Meier analysis comparing survival of patients carrying tumours double positive for C/EBP $\beta$  and STAT3 (red) and single- or double-negative tumours (black). \* $P \leq 0.05$ , \*\* $P \leq 0.01$ .

compared their expression with that of YKL40 (a well-established mesenchymal protein expressed in primary human GBM)<sup>21,32</sup> and with patient outcome in a collection of 62 GBM (Supplementary Fig. 9). Expression of either C/EBP $\beta$  or STAT3 was significantly associated with YKL40 expression (C/EBP $\beta$ ,  $P = 4.9 \times 10^{-5}$ ; STAT3,  $P = 2.2 \times 10^{-4}$ ), with higher association in double-positive tumours (C/EBP $\beta^+$  STAT3 $^+$ ,  $P = 2.7 \times 10^{-6}$ ) versus double-negative ones (C/EBP $\beta^-$  STAT3 $^-$ , Supplementary Table 13). Double-positive tumours were associated with worse clinical outcome than either single- or double-negative tumours (log-rank test,  $P = 0.0002$ , Fig. 5h). Positivity for either of the two TFs remained predictive of negative outcome but with lower statistical strength than double positivity (C/EBP $\beta$ ,  $P = 0.0022$ ; STAT3,  $P = 0.0017$ ).

## Discussion

We have shown that inference of context-specific regulatory network identifies the transcriptional module that controls expression of the mesenchymal signature associated with poor prognosis in HGGs. In this approach, the traditional model of gene-expression-profile-based cancer research, yielding long lists of differentially expressed genes (that is, cancer signatures), becomes only a starting point for a cellular-network analysis, where a causal regulatory model identifies the TFs controlling the signatures and related phenotypes.

Recently, there have been several unsuccessful attempts to identify common expression signatures predictive of the same cellular phenotype<sup>33</sup>. Our approach produced virtually identical regulatory MR modules when applied to two completely distinct data sets and



signatures associated with poor prognosis in HGGs, thus indicating that MRs of mammalian phenotype signatures may be significantly more conserved than the complement of differentially expressed genes. Other methods, including differential expression analysis, DNA-binding-site enrichment analysis<sup>8</sup> and relevance network analysis<sup>34</sup> could not identify C/EBP $\beta$  and STAT3 as MRs (see Supplementary Note 2). This suggests that enrichment analysis of ARACNe-inferred TF regulons is specifically useful for the identification of MRs of cellular phenotypes. Our results do not exclude that other graph-theoretical methods such as Bayesian networks might provide further fine-grain regulatory insight once the number of candidate MRs is reduced to a handful by methods such as those proposed here. Yet, once a relatively small number of TFs is identified, direct experimental validation is feasible.

The experimental follow-up established that C/EBP $\beta$  and STAT3 are MRs sufficient in NSCs and necessary in human glioma cells for mesenchymal transformation. Interestingly, C/EBP $\beta$  and STAT3 are expressed in the developing nervous system<sup>35–38</sup>. However, although STAT3 induces astrocyte differentiation and inhibits neuronal differentiation of neural stem/progenitor cells, C/EBP $\beta$  promotes neurogenesis and opposes gliogenesis<sup>39–41</sup>. A question remains as to how the combined activity of C/EBP $\beta$  and STAT3 can reprogram NSCs towards an aberrant lineage (mesenchymal) and oppose the genesis of the normal neural lineages (neuronal and glial). We propose that mesenchymal transformation results from the concurrent activation of two conflicting transcriptional regulators normally operating to funnel opposing signals (neurogenesis versus gliogenesis). This condition is intolerable by normal neural stem/progenitor cells, whereas it operates to permanently drive the aberrant mesenchymal phenotype in the context of the genetic and epigenetic changes that accompany high-grade gliomagenesis (for example, EGFR amplification, PTEN loss, Akt activation)<sup>4</sup>. Because expression of C/EBP $\beta$  and STAT3 in human glioma is essential to maintain the tumour initiating capacity and the ability to invade the normal brain, the two TFs provide important clues for diagnostic and pharmacological intervention. Consistent with this, the combined expression of C/EBP $\beta$  and STAT3 is linked to the mesenchymal state of primary GBM, and provides an excellent prognostic biomarker for tumour aggressiveness.

Thus, systems biology methods can be effectively used to infer MRs that choreograph malignant transformation. This model will be applicable to the dissection of other phenotypic states.

## METHODS SUMMARY

**Cell culture.** Primary NSCs were isolated from E13.5 mouse telencephalon, and cultured in the presence of FGF2 and EGF as described<sup>42</sup>. Differentiation was induced by culturing NSCs in NSC medium without EGF and FGF2. GBM-derived BTICs were grown in Neurobasal media supplemented with FGF2 and EGF.

**Generation of transcriptional network, microarrays and qRT-PCR.** GBM transcriptional network was generated by ARACNe<sup>12</sup>. Total RNA was reverse transcribed to complementary DNA and amplified using primers specific for human and murine transcripts. Expression values were calculated relative to the 18S ribosomal RNA. RNA was used for analysis on Illumina HumanHT-12v3 or MouseWG-6 expression BeadChip. Sample information is in Methods.

**Master regulator analysis.** For each TF, the statistical significance of the intersection between the TF-regulon and the gene expression signature was computed by FET. Significant genes were ranked on the basis of the number of overlapping genes.

**GSEA.** The statistical significance of the enrichment of a ranked list of genes in a smaller set of genes was determined as described<sup>43</sup>.

**Stepwise linear regression.** The regulatory model of each gene was determined by identifying the smallest number of TFs that were informative for the expression of that gene across the data set. TFs were added to the model one at the time, until the error reduction produced by adding another TF was no longer statistically significant. Models had on average 1–5 TFs.

**Intracranial injection of glioma cells.** SNB19 glioma cell line and GBM-derived BTICs were injected into the brain of 6–8 weeks NOD/SCID mice 48 h after infection with lentiviruses carrying shRNAs using a stereotaxic frame. Animals were monitored and euthanized when they presented with signs of tumour.

Mouse research was approved by the Committee for Animal Care, and conducted in compliance with the Animal Welfare Act Regulations.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 2 September; accepted 30 November 2009.

Published online 23 December 2009.

- Ohgaki, H. & Kleihues, P. Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. *J. Neuropathol. Exp. Neurol.* **64**, 479–489 (2005).
- Demuth, T. & Berens, M. E. Molecular mechanisms of glioma cell migration and invasion. *J. Neurooncol.* **70**, 217–228 (2004).
- Kargiotis, O., Rao, J. S. & Kyritsis, A. P. Mechanisms of angiogenesis in gliomas. *J. Neurooncol.* **78**, 281–293 (2006).
- Phillips, H. S. et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* **9**, 157–173 (2006).
- Tso, C. L. et al. Primary glioblastomas express mesenchymal stem-like properties. *Mol. Cancer Res.* **4**, 607–619 (2006).
- Takashima, Y. et al. Neuroepithelial cells supply an initial transient wave of MSC differentiation. *Cell* **129**, 1377–1388 (2007).
- Wurmser, A. E. et al. Cell fusion-independent differentiation of neural stem cells to the endothelial lineage. *Nature* **430**, 350–356 (2004).
- Rhodes, D. R. & Chinnaiyan, A. M. Integrative analysis of the cancer transcriptome. *Nature Genet.* **37** (suppl.), S31–S37 (2005).
- Basso, K. et al. Reverse engineering of regulatory networks in human B cells. *Nature Genet.* **37**, 382–390 (2005).
- Chen, Y. et al. Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
- Margolin, A. A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7** (suppl. 1), S7 (2006).
- Margolin, A. A. et al. Reverse engineering cellular networks. *Nature Protocols* **1**, 662–671 (2006).
- Zhao, X. et al. The N-Myc-DLL3 cascade is suppressed by the ubiquitin ligase Huwe1 to inhibit proliferation and promote neurogenesis in the developing brain. *Dev. Cell* **17**, 210–221 (2009).
- Lim, W. K., Lyashenko, E. & Califano, A. Master regulators used as breast cancer metastasis classifier. *Pac. Symp. Biocomput.* **14**, 504–519 (2009).
- Mani, K. M. et al. A systems biology approach to prediction of oncogenes and perturbation targets in B cell lymphomas. *Mol. Syst. Biol.* **4**, 169–178 (2008).
- Palomero, T. et al. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc. Natl Acad. Sci. USA* **103**, 18261–18266 (2006).
- Taylor, R. C., Acquah-Mensah, G., Singhal, M., Malhotra, D. & Biswal, S. Network inference algorithms elucidate Nrf2 regulation of mouse lung oxidative stress. *PLoS Comput. Biol.* **4**, e1000166 (2008).
- Hanauer, D. A., Rhodes, D. R., Sinha-Kumar, C. & Chinnaiyan, A. M. Bioinformatics approaches in the study of cancer. *Curr. Mol. Med.* **7**, 133–141 (2007).
- Lander, A. D. A calculus of purpose. *PLoS Biol.* **2**, e164 (2004).
- Freije, W. A. et al. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res.* **64**, 6503–6510 (2004).
- Nigro, J. M. et al. Integrated array-comparative genomic hybridization and expression array profiles identify clinically relevant molecular subtypes of glioblastoma. *Cancer Res.* **65**, 1678–1686 (2005).
- The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
- Ramji, D. P. & Foka, P. CCAAT/enhancer-binding proteins: structure, function and regulation. *Biochem. J.* **365**, 561–575 (2002).
- Aoki, K. et al. RP58 associates with condensed chromatin and mediates a sequence-specific transcriptional repression. *J. Biol. Chem.* **273**, 26698–26704 (1998).
- Fuks, F., Burgers, W. A., Godin, N., Kasai, M. & Kouzarides, T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J.* **20**, 2536–2544 (2001).
- Lee, J. P. et al. Stem cells act through multiple mechanisms to benefit mice with neurodegenerative metabolic disease. *Nature Med.* **13**, 439–447 (2007).
- Park, K. I. et al. Acute injury directs the migration, proliferation, and differentiation of solid organ stem cells: evidence from the effect of hypoxia-ischemia in the CNS on clonal “reporter” neural stem cells. *Exp. Neurol.* **199**, 156–178 (2006).
- Parker, M. A. et al. Expression profile of an operationally-defined neural stem cell clone. *Exp. Neurol.* **194**, 320–332 (2005).
- Bromberg, J. F. et al. Stat3 as an oncogene. *Cell* **98**, 295–303 (1999).
- Lee, J. et al. Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell* **9**, 391–403 (2006).
- The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
- Pelloski, C. E. et al. YKL-40 expression is associated with poorer response to radiation and shorter overall survival in glioblastoma. *Clin. Cancer Res.* **11**, 3326–3334 (2005).



33. Ein-Dor, L., Kela, I., Getz, G., Givol, D. & Domany, E. Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* **21**, 171–178 (2005).
34. Butte, A. J. & Kohane, I. S. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac. Symp. Biocomput.* **5**, 418–429 (2000).
35. Barnabé-Heider, F. *et al.* Evidence that embryonic neurons regulate the onset of cortical gliogenesis via cardiotrophin-1. *Neuron* **48**, 253–265 (2005).
36. Bonni, A. *et al.* Regulation of gliogenesis in the central nervous system by the JAK-STAT signaling pathway. *Science* **278**, 477–483 (1997).
37. Sterneck, E. & Johnson, P. F. CCAAT/enhancer binding protein  $\beta$  is a neuronal transcriptional regulator activated by nerve growth factor receptor signaling. *J. Neurochem.* **70**, 2424–2433 (1998).
38. Nadeau, S., Hein, P., Fernandes, K. J., Peterson, A. C. & Miller, F. D. A transcriptional role for C/EBP  $\beta$  in the neuronal response to axonal injury. *Mol. Cell. Neurosci.* **29**, 525–535 (2005).
39. Ménard, C. *et al.* An essential role for a MEK-C/EBP pathway during growth factor-regulated cortical neurogenesis. *Neuron* **36**, 597–610 (2002).
40. Nakashima, K. *et al.* Synergistic signaling in fetal brain by STAT3-Smad1 complex bridged by p300. *Science* **284**, 479–482 (1999).
41. Paquin, A., Barnabe-Heider, F., Kageyama, R. & Miller, F. D. CCAAT/enhancer-binding protein phosphorylation biases cortical precursors to generate neurons rather than astrocytes *in vivo*. *J. Neurosci.* **25**, 10747–10758 (2005).
42. Bachoo, R. M. *et al.* Epidermal growth factor receptor and Ink4a/Arf: convergent mechanisms governing terminal differentiation and transformation along the neural stem cell to astrocyte axis. *Cancer Cell* **1**, 269–277 (2002).
43. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by National Institute of Health grants R01CA109755 (A.C.), R01CA101644 (A.L.), R01CA085628 and R01NS061776 (A.I.), NCI Grand Opportunities TDDN Network 1RC2CA148308-01 (A.C.), In Silico Research Centre of Excellence NCI-caBIG 29XS192 (A.C.), National Centers for Biomedical Computing NIH Roadmap Initiative U54CA121852 (A.C.) and National Institute of General Medical Sciences grant P20GM075059 (E.Y.S.). M.S.C. is supported by a fellowship from the Italian Ministry of Welfare/Provincia di Benevento and S.L.A. by a fellowship from Fondation de Recherche Medicale. We thank N. Ramirez-Martinez for technical assistance with mouse husbandry and *in vivo* procedures.

**Author Contributions** A.C. and A.I. conceived the ideas for this study. A.C. designed the computational systems biology approach and A.I. the experimental platform. M.S.C. prepared constructs, performed the biochemical experiments and the microarrays, conducted biological experiments and analyses, assisted in mouse intracranial injections and performed tumour xenograft immunohistochemistry and tumour analysis. W.K.L. performed reverse engineering, master regulator, and statistical analyses. M.J.A. conducted gene expression, bioinformatics and statistical analyses. R.J.B. and E.Y.S. provided experimental material. X.Z. and F.D. assisted in mouse intracranial injections. E.P.S., H.C. and K.A. provided reagents, performed the arrayCGH/expression analysis and primary human tumour immunohistochemistry. S.L.A. performed cell culture immunofluorescence microscopy and analysis. A.L. assisted in primary NSC experiments, performed intracranial injections and assisted in the analysis of mouse xenografts. A.I. and A.C. wrote the manuscript with contributions from all other authors. M.S.C., W.K.L. and M.J.A. contributed equally to this work.

**Author Information** Gene expression data have been deposited in Gene Expression Omnibus (GEO) with the following accession numbers: GSE19113 for mouse and GSE19114 for human data. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.C. ([califano@c2b2.columbia.edu](mailto:califano@c2b2.columbia.edu)) or A.I. ([ai2102@columbia.edu](mailto:ai2102@columbia.edu)).

## METHODS

**Array comparative genomic hybridization expression correlation.** The correlation between gene expression and DNA copy number for the MGES genes was determined using data from 76 high-grade gliomas for which both gene expression array and aCGH profiling were performed<sup>4</sup>. Tumours were grouped on the basis of molecular subtype (mesenchymal, proneural or proliferative) and the mean expression for MGES genes were determined in each group. The normalized copy number of each gene was interpolated based on the copy number of the nearest genomic clone on the CGH array as determined by comparison of the sequence annotation of both array platforms, as previously described<sup>21</sup>.

**ARACNe network reconstruction.** ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks), an information-theoretic algorithm for inferring transcriptional interactions, was used to identify a repertoire of candidate transcriptional regulators of the MGES genes. Expression profiles used in the analysis were previously characterized using Affymetrix HU-133A microarrays and preprocessed by MAS 5.0 normalization procedure<sup>4</sup>. First, candidate interactions between a TF ( $x$ ) and its potential target ( $y$ ) are identified by computing pairwise mutual information,  $MI[x, y]$ , using a Gaussian kernel estimator<sup>12</sup> and by thresholding the mutual information based on the null-hypothesis of statistical independence ( $P < 0.05$ , Bonferroni corrected for the number of tested pairs). Then, indirect interactions are removed using the data processing inequality, a well-known property of the mutual information. For each TF–target pair ( $x, y$ ) we considered a path through any other TF ( $z$ ) and remove any interaction such that  $MI[x, y] < \min(MI[x, z], MI[y, z])$ .

**TF classification.** To identify human TFs, we selected the human genes annotated as ‘transcription factor activity’ in Gene Ontology and the list of TFs from TRANSFAC. From this list, we removed general TFs (for example, stable complexes such as polymerases or TATA-box-binding proteins), and added some TFs not annotated by Gene Ontology, producing a final list of 928 TFs that were represented on the HU-133A microarray gene set.

**Master regulator analysis.** The MRA has two steps. First, for each TF its signature-enrichment is computed as the  $P$  value of the overlap between the TF-regulon and the signature genes (that is, the MGES genes in this case), assessed by FET. Because FET depends on regulon size, it can be used to assess signature-enriched TFs but not to rank them. TFs are thus ranked based on the total number of signature genes included in their regulon, under the assumption that TFs controlling a larger fraction of the signature will be more likely to determine its activity.

**Stepwise linear regression analysis.** A regulatory program for each MGES gene was computed as follows: the  $\log_2$  expression of the  $i$ -th MGES gene was considered as the response variable and the  $\log_2$ -expression of the TFs as the explanatory variables in the linear model  $\log x_i = \sum \alpha_{ij} \log f_j + \beta_{ij}$  (ref. 44). Here,  $f_j$  represents the expression of the  $j$ -th TF in the model and the  $(\alpha_{ij}, \beta_{ij})$  are linear coupling coefficients computed by standard regression analysis. TFs are iteratively added to the model, by choosing each time the one producing the smallest relative error  $E = \sum |x_i - x_{i0}|/x_{i0}$  between predicted and observed target expression. This is repeated until the decrease in relative error is no longer statistically significant, based on permutation testing. To avoid excessive multiple hypothesis testing correction, TFs were chosen only among the following: (1) the 53 inferred by ARACNe at FDR  $< 0.05$ , and (2) TFs whose DNA-binding signature was significantly enriched in the proximal promoter of the MGES genes and that are expressed in the data set, based on the coefficient of variation ( $CV \geq 0.5$ ). TFs were then ranked based on the number of MGES target they regulated, with the average linear-regression coefficient providing further insight. The log-transformation allows convenient linear representation of multiplicative interactions between TF activities<sup>44,45</sup>. TFs were individually added to the model, each time selecting the one contributing the most significant reduction in relative expression error (predicted versus observed), until error-reduction was no longer significant.

**Enrichment analysis.** The FDRs are computed using procedures described by Benjamini and Hochberg<sup>46</sup>, where the adjusted  $P$  values,  $q = p^*/n$  ( $p = P$  value,  $n$  = total number of tests,  $i$  = sorted rank of  $P$  value). It is a less conservative procedure to correct for multiple comparisons than family-wise error rate (FWER), especially when the number of tests is large.

**Cell lines and cell culture conditions.** SNB75, SNB19, 293T and Phoenix cell lines were grown in DMEM plus 10% FBS (Gibco/BRL). GBM-derived BTICs were grown as neurospheres in Neurobasal media (Invitrogen) containing N2 and B27 supplements (Invitrogen), and human recombinant FGF2 and EGF (50 ng ml<sup>-1</sup> each; Peprotech). mNSCs (from an early passage of clone C17.2) (26–28) were cultured in DMEM plus 10% heat-inactivated FBS (Gibco/BRL), 5% horse serum (Gibco/BRL) and 1% L-glutamine (Gibco/BRL). Neuronal differentiation of mNSCs was induced by growing cells in DMEM supplemented

with 0.5% horse serum. For chondrocyte differentiation, cells were treated with STEMPro chondrogenesis differentiation kit (Gibco/BRL) for 20 days. Primary mNSCs were isolated from E13.5 mouse telencephalon and cultured in the presence of FGF2 and EGF (20 ng ml<sup>-1</sup> each) as described<sup>42</sup>. Differentiation of NSCs was induced by culturing neurospheres on laminin-coated dishes in NSC medium in the absence of growth factors. mNSCs expressing STAT3C and C/EBP $\beta$  were generated by retroviral infections using supernatant from Phoenix ecotropic packaging cells transfected with pBabe-STAT3C-Flag and/or pLZRS-T7-His-C/EBP $\beta$ -2-IRES-GFP.

**Promoter analysis and ChIP.** Promoter analysis was performed using the MatInspector software (<http://www.genomatix.de>). A sequence 2 kb upstream and 2 kb downstream from the transcription start site was analysed for the presence of putative binding sites for each TF. Primers used to amplify sequences surrounding the predicted binding sites were designed using the Primer3 software ([http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)) and are listed in Supplementary Table 14.

ChIP was performed as described<sup>47</sup>. SNB75 cell lysates were precleared with Protein A/G beads (Santa Cruz) and incubated at 4 °C overnight with 1  $\mu$ g of polyclonal antibody specific for C/EBP $\beta$  (sc-150, Santa Cruz), STAT3 (sc-482, Santa Cruz), FOSL2 (Fra2, sc-604, Santa Cruz), bHLH-B2 (A300-649A, BETHYL Laboratories), or normal rabbit immunoglobulins (Santa Cruz). DNA was eluted in 200  $\mu$ l of water and 1  $\mu$ l was analysed by PCR with Platinum Taq (Invitrogen). For primary GBM samples, 30 mg of frozen tissue was transferred in a tube with 1 ml of culture medium, fixed with 1% formaldehyde for 15 min and stopped with 0.125 M glycine for 5 min. Samples were centrifuged at 1,500g for 2 min, washed twice and diluted in PBS. Tissues were homogenized using a pestle and suspended in 3 ml of ice-cold immunoprecipitation buffer with protease inhibitors and sonicated. ChIP was then performed as described above.

**qRT-PCR and microarray analysis.** RNA was prepared with RiboPure kit (Ambion), and used for first-strand cDNA synthesis using random primers and SuperScriptII Reverse Transcriptase (Invitrogen). qRT-PCR was performed using Power SYBR Green PCR Master Mix (Applied Biosystems). Primers are listed in Supplementary Table 15. qRT-PCR results were analysed by the  $\Delta\Delta C_T$  method<sup>48</sup> using 18S as a housekeeping gene.

RNA amplification for array analysis was performed with Illumina TotalPrep RNA Amplification Kit (Ambion). One-and-a-half micrograms of amplified RNA was hybridized on Illumina HumanHT-12v3 (including 24,385 human genes) or MouseWG-6 (including 20,311 mouse genes) expression BeadChip according to the manufacturer's instructions. Hybridization data was obtained with an iScan BeadArray scanner (Illumina) and pre-processed by variance stabilization and robust spline normalization implemented in the lumi package under the R-system<sup>49</sup>.

**Immunofluorescence and immunohistochemistry.** Immunofluorescence staining was performed as previously described<sup>50</sup>. Primary antibodies and dilutions were: SMA (mouse monoclonal, Sigma, 1:200), fibronectin (mouse monoclonal, BD Biosciences, 1:200), Tau (rabbit polyclonal, Dako, 1:400),  $\beta$ III-tubulin (mouse monoclonal, Promega, 1:1,000), CTGF (rabbit polyclonal, Santa Cruz, 1:200), YKL40 (rabbit polyclonal, Quidel, 1:200) and COL5A1 (rabbit polyclonal, Santa Cruz, 1:200). Confocal images acquired with a Zeiss Axioscop2 FS MOT microscope were used to score positive cells. At least 500 cells were scored for each sample. Quantification of the fibronectin intensity staining in mNSCs was performed using NIH Image J software (<http://rsb.info.nih.gov/ij/>). The histogram of the intensity of fluorescence of each point of a representative field for each condition was generated. The fluorescence intensity of three fields from three independent experiments was scored, standardized to the number of cells in the field and divided by the intensity of the vector. For immunostaining of xenograft tumours, mice were perfused transcardially with 4% paraformaldehyde (PFA), brains were dissected and post-fixed for 48 h in 4% PFA. Immunostaining was performed as previously described<sup>51</sup>. Primary antibodies and dilutions: fibronectin (mouse monoclonal, BD Biosciences, 1:100), COL5A1 (rabbit polyclonal, Santa Cruz, 1:100), YKL40 (rabbit polyclonal, Quidel, 1:100), human vimentin (mouse monoclonal, Sigma, 1:50), and Ki67 (rabbit polyclonal, Novocastra laboratories, 1:1,000). Quantification of the tumour area was obtained by measuring the human vimentin-positive area in the section using the NIH Image J software (<http://rsb.info.nih.gov/ij/>). Five tumours for each group were analysed. For quantification of Ki67, the percentage of positive cells was scored in five tumours per each group. In histograms the values represents the mean values; error bars are standard deviations. Statistical significance was determined by  $t$ -test (with Welch's correction) using GraphPad Prism 4.0 software (GraphPad Inc.). Immunohistochemistry of primary human GBM was performed as previously described<sup>52</sup>. The primary antibodies and dilutions were: anti-YKL40 (rabbit polyclonal, Quidel, 1:750), anti-C/EBP $\beta$ , (rabbit polyclonal, Santa Cruz, 1:250) and anti-p-STAT3 (rabbit monoclonal, Cell Signaling, 1:25). Scoring for YKL40 was

based on a three-tiered system, where 0 was <5% of tumour cells positive, 1 was 5–30% positivity, and 2 was >30% of tumour cells positive. Scores of 1 and 2 were later collapsed into a single value for display purposes on Kaplan–Meier curves. Associations between C/EBP $\beta$ /STAT3 and YKL40 were assessed using the FET. Associations between C/EBP $\beta$ /STAT3 and patients survival were assessed using the log-rank (Mantel–Cox) test of equality of survival distributions.

**Migration and invasion assays.** For the wound assay testing migration, mNSCs were plated in 60-mm dishes and grown until 95% confluence. A scratch of approximately 1,000  $\mu$ m was made with a P1000 pipette tip and images were taken every 24 h with an inverted microscope. For the Matrigel invasion assay, mNSCs and SNB19 cells ( $1 \times 10^4$ ) were added to the upper compartment of a 24-well BioCoat Matrigel Invasion Chamber (BD Biosciences) in serum-free DMEM. The lower compartment of the chamber was filled with DMEM containing either 0.5% horse serum or 20  $\mu$ g ml $^{-1}$  PDGF-BB (R&D systems) as a chemoattractant. After 24 h, invading cells were fixed, stained according to the manufacturer's instructions and counted. For GBM-derived BTICs,  $5 \times 10^4$  cells were plated on the upper chamber in the absence of growth factors. In the lower compartment Neurobasal medium containing B27 and N2 supplements plus 20  $\mu$ g ml $^{-1}$  PDGF-BB (R&D systems) was used as chemoattractant.

**Lentivirus infection.** Lentiviral expression vectors carrying shRNAs were purchased from Sigma. The sequences are listed in Supplementary Table 16. To generate lentiviral particles, each shRNA expression plasmid was co-transfected with pCMV-dR8.91 and pCMV-MD2.G vectors into human embryonic kidney 293T cells using Fugene 6 (Roche). Lentiviral infections were performed as described<sup>51</sup>.

**Intracranial injection.** Intracranial injection of the SNB19 glioma cell line and GBM-derived BTICs was performed in 6–8-week NOD/SCID mice (Charles River laboratories) in accordance with guidelines of the International Agency for Research on Cancer's Animal Care and Use Committee. In brief, 48 h after

lentiviral infection,  $2 \times 10^5$  SNB19 cells or  $3 \times 10^5$  BTICs were injected 2 mm lateral and 0.5 mm anterior to the bregma, 3 mm below the skull. Mice were monitored daily and euthanized when neurological symptoms appeared. A Kaplan–Meier survival curve of the mice injected with SNB19 glioma cells was generated using the DNA Statview software package (AbacusConcepts).

44. Tegner, J., Yeung, M. K., Hasty, J. & Collins, J. J. Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling. *Proc. Natl Acad. Sci. USA* **100**, 5944–5949 (2003).
45. Bussemaker, H. J., Li, H. & Siggia, E. D. Regulatory element detection using correlation with expression. *Nature Genet.* **27**, 167–174 (2001).
46. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
47. Frank, S. R., Schroeder, M., Fernandez, P., Taubert, S. & Amati, B. Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. *Genes Dev.* **15**, 2069–2082 (2001).
48. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. *Methods* **25**, 402–408 (2001).
49. Du, P., Kibbe, W. A. & Lin, S. M. *lumi*: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).
50. Rothschild, G., Zhao, X., Iavarone, A. & Lasorella, A. E proteins and Id2 converge on p57<sup>Kip2</sup> to regulate cell cycle in neural cells. *Mol. Cell. Biol.* **26**, 4351–4361 (2006).
51. Zhao, X. *et al.* The HECT-domain ubiquitin ligase Huwe1 controls neural differentiation and proliferation by destabilizing the N-Myc oncoprotein. *Nature Cell Biol.* **10**, 643–653 (2008).
52. Simmons, M. L. *et al.* Analysis of complex relationships between age, p53, epidermal growth factor receptor, and survival in glioblastoma patients. *Cancer Res.* **61**, 1122–1128 (2001).

# Cross-Species Regulatory Network Analysis Identifies a Synergistic Interaction between *FOXM1* and *CENPF* that Drives Prostate Cancer Malignancy

Alvaro Aytes,<sup>1,8,15</sup> Antonina Mitrofanova,<sup>2,15</sup> Celine Lefebvre,<sup>2,15,16</sup> Mariano J. Alvarez,<sup>2</sup> Mireia Castillo-Martin,<sup>9</sup> Tian Zheng,<sup>7,10</sup> James A. Eastham,<sup>11</sup> Anuradha Gopalan,<sup>12</sup> Kenneth J. Pienta,<sup>13,14</sup> Michael M. Shen,<sup>1,2,3,4,7</sup> Andrea Califano,<sup>2,5,7,\*</sup> and Cory Abate-Shen<sup>1,2,6,7,\*</sup>

<sup>1</sup>Department of Urology, Columbia University Medical Center, New York, NY 10032, USA

<sup>2</sup>Department of Systems Biology, Columbia University Medical Center, New York, NY 10032, USA

<sup>3</sup>Department of Medicine, Columbia University Medical Center, New York, NY 10032, USA

<sup>4</sup>Department of Genetics & Development, Columbia University Medical Center, New York, NY 10032, USA

<sup>5</sup>Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY 10032, USA

<sup>6</sup>Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY 10032, USA

<sup>7</sup>Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY 10032, USA

<sup>8</sup>Translational Research Laboratory, Catalan Institute of Oncology, Bellvitge Institute for Biomedical Research, L'Hospitalet de Llobregat, Barcelona 08907, Spain

<sup>9</sup>Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>10</sup>Department of Statistics, Columbia University, New York, NY 10027, USA

<sup>11</sup>Department of Urology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

<sup>12</sup>Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

<sup>13</sup>The University of Michigan, Ann Arbor, MI 48109, USA

<sup>14</sup>Brady Urological Institute at the Johns Hopkins School of Medicine, Baltimore, MD 21231, USA

<sup>15</sup>Co-first author

<sup>16</sup>Present address: Inserm U981, Gustave Roussy Cancer Institute, 94805 Villejuif Cedex, France

\*Correspondence: [califano@c2b2.columbia.edu](mailto:califano@c2b2.columbia.edu) (A.C.), [cabateshen@columbia.edu](mailto:cabateshen@columbia.edu) (C.A.-S.)

<http://dx.doi.org/10.1016/j.ccr.2014.03.017>

## SUMMARY

To identify regulatory drivers of prostate cancer malignancy, we have assembled genome-wide regulatory networks (interactomes) for human and mouse prostate cancer from expression profiles of human tumors and of genetically engineered mouse models, respectively. Cross-species computational analysis of these interactomes has identified *FOXM1* and *CENPF* as synergistic master regulators of prostate cancer malignancy. Experimental validation shows that *FOXM1* and *CENPF* function synergistically to promote tumor growth by coordinated regulation of target gene expression and activation of key signaling pathways associated with prostate cancer malignancy. Furthermore, co-expression of *FOXM1* and *CENPF* is a robust prognostic indicator of poor survival and metastasis. Thus, genome-wide cross-species interrogation of regulatory networks represents a valuable strategy to identify causal mechanisms of human cancer.

## INTRODUCTION

It is widely appreciated that cancer is not a single entity but rather a highly individualized spectrum of diseases characterized by a

large number of molecular alterations (Hanahan and Weinberg, 2011). Distinguishing those that constitute true drivers of cancer phenotypes from the multitude that are simply deregulated has proven to be a daunting task, which is further exacerbated by

### Significance

Genetically engineered mouse models have been widely used for *in vivo* analyses of cancer phenotypes as well as preclinical investigations. However, inherent species differences often hinder the appropriate extrapolation of studies performed in mice to human cancer. Here we introduce a strategy using cross-species computational analysis of context-specific regulatory networks for the effective integration of experimental findings from mouse models and human cancer. This approach enables the identification of conserved master regulators of malignant prostate cancer, as well as elucidation of their synergistic interactions. This computational paradigm should be broadly applicable for elucidating causal mechanisms of cancer, as well as integrating preclinical analyses from mouse to man.



the complexity of elucidating how such drivers interact synergistically to elicit cancer phenotypes. In this regard, prostate cancer is particularly challenging because its notorious heterogeneity, combined with a relative paucity of recurrent gene mutations, has made it especially difficult to identify molecularly distinct subtypes with known clinical outcomes (Baca et al., 2013; Schoenborn et al., 2013; Shen and Abate-Shen, 2010). Additionally, whereas most early-stage prostate tumors are readily treatable (Cooperberg et al., 2007), advanced prostate cancer frequently progresses to castration resistance, which is often metastatic and nearly always fatal (Ryan and Tindall, 2011; Scher and Sawyers, 2005). Thus, there is a pressing need to identify bona fide determinants of aggressive prostate cancer as well as prognostic biomarkers of disease outcome.

Analysis of genetically engineered mouse models (GEMMs) can circumvent inherent challenges associated with the intrinsic complexity of more heterogeneous human cancer phenotypes. Indeed, investigations of mouse models of prostate cancer have contributed to characterization of disease-specific pathways, led to the identification of biomarkers of disease progression, and provided useful preclinical models for prevention and therapy (Irshad and Abate-Shen, 2013; Ittmann et al., 2013). Following the description of an initial transgenic model nearly 20 years ago, there are now numerous GEMMs that collectively model key molecular pathways deregulated in human prostate cancer, and recapitulate the various stages of disease progression including preinvasive lesions (prostatic intraepithelial neoplasia; PIN), adenocarcinoma, castration resistance, and metastasis (Irshad and Abate-Shen, 2013; Ittmann et al., 2013).

However, inherent species differences often hinder direct comparative analyses of mouse models and human cancer. Indeed, such analyses would greatly benefit from computational approaches that enable accurate cross-species integration of regulatory information from mouse to man. Recent advances in systems biology have led to the reverse engineering of regulatory networks (interactomes) that integrate large-scale data sets encompassing expression profiles, protein-protein interactions, genomic alterations, and epigenetic changes associated with cancer and other diseases (Lefebvre et al., 2012). However, whereas individual analyses of human and murine interactomes have led to relevant biological discoveries, their cross-species interrogation has not been systematically implemented.

Here, we introduce an approach for accurate cross-species analysis of conserved cancer pathways based on reverse engineering of genome-wide regulatory networks (i.e., interactomes) representing both human and mouse prostate cancer. To accomplish this, we have produced a regulatory network based on *in vivo* perturbation of a repertoire of mouse cancer models and implemented comparative analysis with a complementary regulatory network generated from human prostate cancer data sets. Cross-species computational interrogation of these paired interactomes, followed by experimental and clinical validation, has elucidated the synergistic interaction of *FOXM1* and *CENPF* as a driver of prostate cancer malignancy. We propose that analyses of genome-wide, cross-species regulatory networks will provide an effective paradigm for elucidating causal mechanisms of human cancer and other complex diseases.

## RESULTS

We developed a strategy for genome-wide interrogation of cancer phenotypes based on accurate integration of experimental data from model organisms and human cancer (Figure 1). First, we generated regulatory networks (interactomes) for human and mouse prostate cancer using the Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNe; Basso et al., 2005; Margolin et al., 2006b). We next evaluated the suitability of these mouse and human interactomes for cross-species interrogation using a computational approach to assess the global conservation of their transcriptional programs. We then used the Master Regulator Inference algorithm (MARINA; Carro et al., 2010; Lefebvre et al., 2010) to infer candidate master regulators that act individually or synergistically to drive malignant prostate cancer. Finally, we performed experimental studies to validate synergistic interactions of master regulators, to elucidate underlying mechanisms, and to evaluate their clinical relevance.

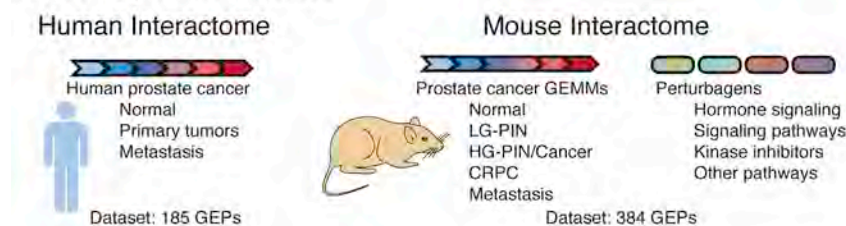
### Assembly of Interactomes for Human and Mouse Prostate Cancer

ARACNe is an unbiased algorithm that infers direct transcriptional interactions based on the mutual information between each transcriptional regulator and its potential targets. For optimal analyses, ARACNe requires large data sets of gene expression profiles ( $\geq 100$ ) having significant endogenous (i.e., genetic) and/or exogenous (i.e., perturbation-induced) heterogeneity. To assemble a human prostate cancer interactome, we analyzed the expression profile data set reported elsewhere (Taylor et al., 2010), which is ideally suited for ARACNe because: (1) it is relatively large ( $n = 185$ ) and diverse, including primary tumors, adjacent normal tissue, metastases, and cell lines; (2) its primary tumors encompass the full range of pathological Gleason scores and have well-annotated clinical outcome data; and (3) it displays extensive genetic diversity and tumor heterogeneity, as shown by t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis (Figure 2; Table S1 available online). Notably, interactomes assembled from three alternative human prostate cancer data sets (Table S1) were neither as complete nor as extensive (data not shown).

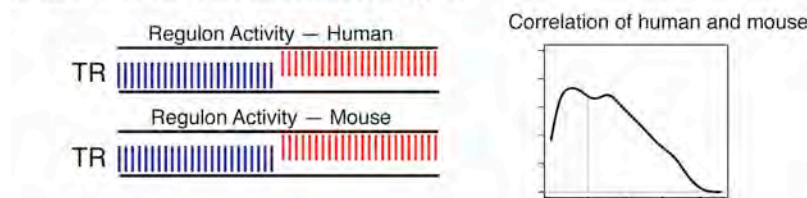
To assemble a corresponding mouse prostate cancer interactome, it was first necessary to generate an expression profile data set of appropriate size and representing sufficient expression variability. We selected 13 distinct GEMMs, which together represent the full spectrum of prostate cancer phenotypes, including normal epithelium (wild-type), low-grade PIN (*Nkx3.1* and *APT*), high-grade PIN, and adenocarcinoma (*APT-P*, *APC*, *Myc*, *NP*, *Erg-P*, and *NP53*), castration resistance (*NP-AI*), and metastatic prostate cancer (*NPB*, *NPK*, and *TRAMP*; Figure S1A; Table S2). To further increase the variability of the expression profiles, we introduced a controlled set of exogenous perturbations by *in vivo* administration of 13 small-molecule perturbagens to each GEMM. Perturbagens were selected for their clinical relevance and/or ability to modulate key prostate cancer pathways, including hormone signaling (testosterone, calcitriol, and enzalutamide); PI3 kinase activity (MK2206, LY294002, and rapamycin); MAP kinase activity (PD035901); tyrosine kinase activity (imatinib, dasatinib, and sorafenib); NF $\kappa$ B signaling (BAY



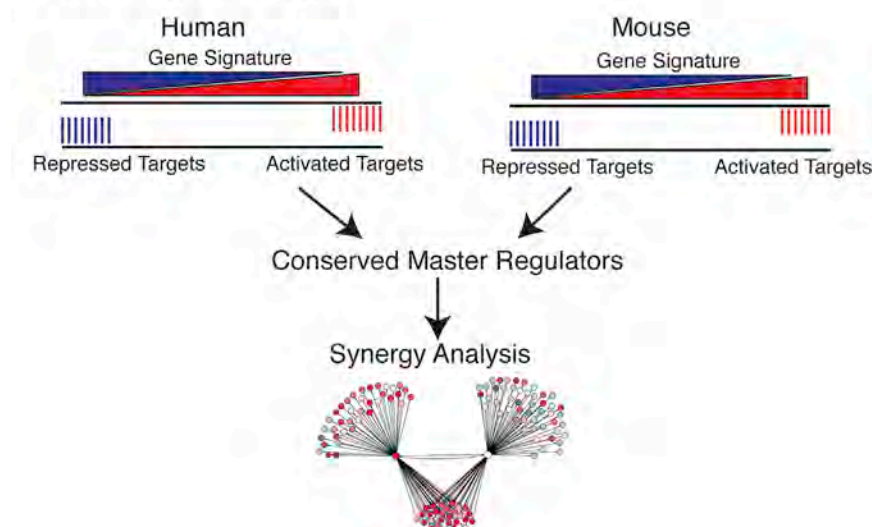
## I. Assembly of interactomes



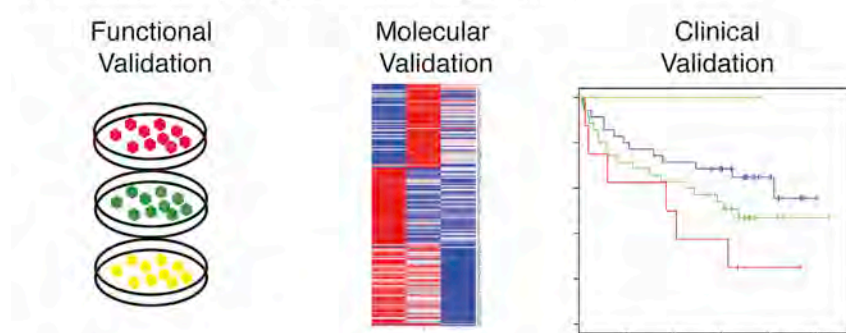
## II. Genome-wide conservation analysis



## III. Master regulator analysis

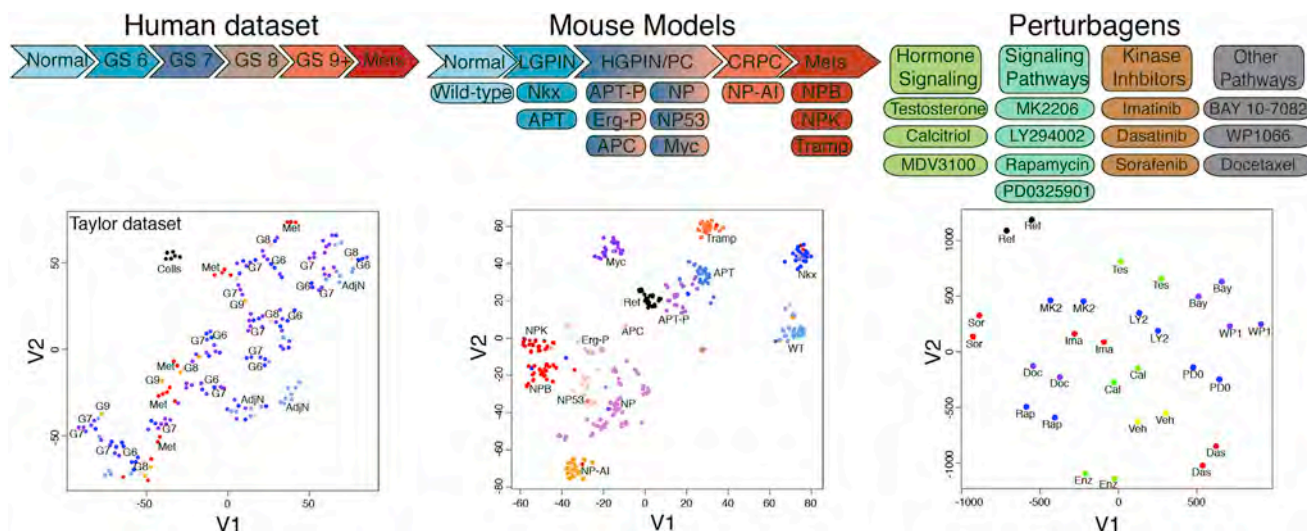


## IV. Validation of Synergistic Master Regulators



### Figure 1. Strategy for Genome-wide Cross-Species Analyses of Prostate Cancer

Schematic representation of the overall strategy. Step I: assembly of human and mouse prostate cancer interactomes. Step II: genome-wide computational analysis of conservation of transcriptional regulon activity in the mouse and human prostate cancer interactomes. Step III: master regulator analysis for identification of conserved master regulators and prediction of synergy. Step IV: validation of candidate master regulators using functional, molecular, and clinical analyses.



**Figure 2. Heterogeneity of Human and Mouse Data Sets Used for Interactome Assembly**

t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis of human and mouse data sets used to assemble the prostate cancer interactomes. (Left) t-SNE analysis of the Taylor data set relative to Gleason score (GS). (Middle) Schematic representation of GEMMs used to assemble the mouse prostate cancer interactome. t-SNE analysis showing relative distribution of the GEMMs. (Right) Schematic diagram depicting perturbagens used to treat the GEMMs. t-SNE analysis showing the relative distribution of perturbagens for a representative GEMM (i.e., the NP mice).

See also [Figure S1](#) and [Tables S1](#) and [S2](#).

11-7082); JAK/STAT activity (WP1066); and chemotherapy (docetaxel; [Supplemental Experimental Procedures](#)). Following pilot studies to define appropriate doses and schedule ([Figures S1B–S1D](#)), we adopted a universal treatment schedule of one treatment per day for 5 days with dosage determined independently for each perturbagen ([Supplemental Experimental Procedures](#)).

The resulting data set comprises 384 gene expression profiles, corresponding to the 13 GEMMs each treated with the 13 perturbagens or vehicles. t-SNE analysis revealed that the resulting mouse data set represented an extensive range of expression variability, as required for ARACNe ([Figure 2](#)). Specifically, whereas expression profiles from the same GEMMs and perturbagens clustered together, the diverse GEMMs and perturbagens provided independent and highly effective axes to modulate gene expression variability.

ARACNe was run independently on the human and mouse data sets using a conservative mutual information threshold ( $p \leq 1.0 \times 10^{-9}$ , i.e.,  $p \leq 0.05$  Bonferroni corrected for all candidate interactions). This resulted in highly robust regulatory networks—in particular, the “human interactome” represented 249,896 interactions between 2,681 transcriptional regulators and their inferred target genes ([Figure 3A](#); [Table S3](#)), whereas the “mouse interactome” represented 222,787 interactions for 2,072 transcriptional regulators ([Figure 3A](#); [Table S4](#)).

### Analysis of Genome-wide Conservation of Transcriptional Regulatory Pathways

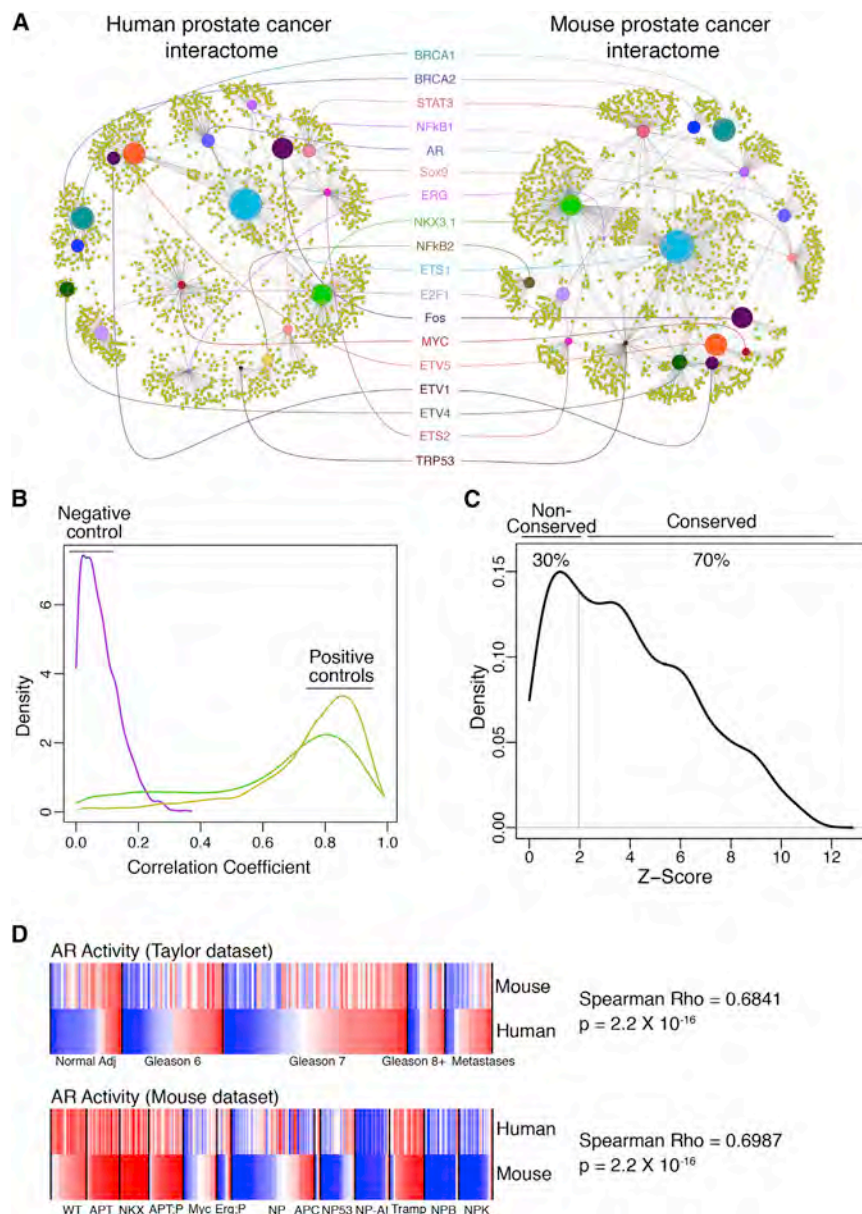
Because it has been previously established that target-by-target analysis may not be optimal to evaluate cross-species interactome conservation (e.g., [Zhang et al., 2012](#)), we developed a quantitative metric to compare conservation of the human and mouse interactomes. In particular, we developed a modification

of the MARINA algorithm that allows for single-sample analysis to infer the differential activity of all 2,028 transcriptional regulators represented in both interactomes. Analysis was performed on 1,009 expression profiles across the four human data sets ([Table S1](#)) and the mouse data set (described herein) to determine whether the inferred activities of each regulator were significantly correlated ( $p \leq 0.05$ ), indicating that the murine and human regulatory programs were conserved ([Supplemental Experimental Procedures](#)). The accuracy of this metric was evident by comparing two equivalent same-species interactomes from the human and mouse data sets (i.e., positive controls), in which virtually all transcriptional regulators were conserved (>90%), in contrast to randomized interactomes (i.e., negative controls), which had virtually no conservation ([Figure 3B](#)).

Using this metric, we found that 70% of the transcriptional regulators in the human and mouse prostate cancer interactomes regulate statistically conserved programs ( $p \leq 0.05$ ; [Figure 3C](#); [Table S5](#)). Notably, among the conserved transcriptional regulators are many genes important in prostate cancer, such as *AR*, *ETS1*, *ETV4*, *ETV5*, *STAT3*, *MYC*, *BRCA1*, and *NKX3.1* ([Shen and Abate-Shen, 2010](#); [Figure 3A](#); [Table S5](#)). In particular, *AR* displayed extensive correlation of its transcriptional activity between the human and mouse interactomes ([Figure 3D](#)), consistent with its known role as a key regulator of prostate development and tumorigenesis ([Ryan and Tindall, 2011](#); [Shen and Abate-Shen, 2010](#)).

### Cross-Species Computational Analysis Identifies Synergistic Master Regulators of Malignant Prostate Cancer

To identify master regulators (MRs) of malignant prostate cancer ([Figure 4](#)), we used the MARINA algorithm, which identifies candidate MRs based on the concerted differential expression



**Figure 3. Genome-wide Conservation Analyses of the Human and Mouse Prostate Cancer Interactomes**

(A) ARACNe subnetworks from the human and mouse prostate cancer interactomes highlighting selected conserved transcriptional regulators. The scaled size of the transcriptional regulator nodes (colored circles) indicates the degree of conservation while the relative distance between them approximates the strength of their association.

(B and C) Histograms (density plots) showing conservation of transcriptional regulator activity between the human and mouse prostate cancer interactomes. (B) Distribution of correlation coefficients of activity profiles of transcriptional regulators for randomized interactomes (negative control; purple line) and the positive control interactomes for human (yellow) and mouse (green; Supplemental Experimental Procedures). (C) Distribution of Z scores for conservation of activity profiles between the human and mouse interactomes at  $p \leq 0.05$ .

(D) Comparison of the androgen receptor (AR) activity levels in each sample from Taylor et al. (top) and the mouse data set (bottom) showing the Spearman correlation coefficient.

See also Tables S3, S4, and S5.

of their ARACNe-inferred targets (i.e., their inferred differential activity, DA). Specifically, “activated” MRs have positively regulated and repressed targets significantly enriched among over- and underexpressed genes, respectively, while “repressed” MRs have the converse. We interrogated the human prostate cancer interactome using a gene signature representing prostate cancer malignancy derived from the Taylor data set as described elsewhere (Aytes et al., 2013), which compares aggressive prostate tumors (Gleason score  $\geq 8$  with rapid biochemical recurrence;  $n = 10$ ) to indolent ones (Gleason score 6 tumors with no biochemical recurrence;  $n = 39$ ). This analysis identified 175 candidate MRs, including 49 activated and 126 repressed ( $p \leq 0.05$ ; Figure 4A; Table S6).

To investigate the robustness of these MRs, we performed MARINA using an independent malignancy signature derived from the Balk data set (Table S1; Stanbrough et al., 2006), which

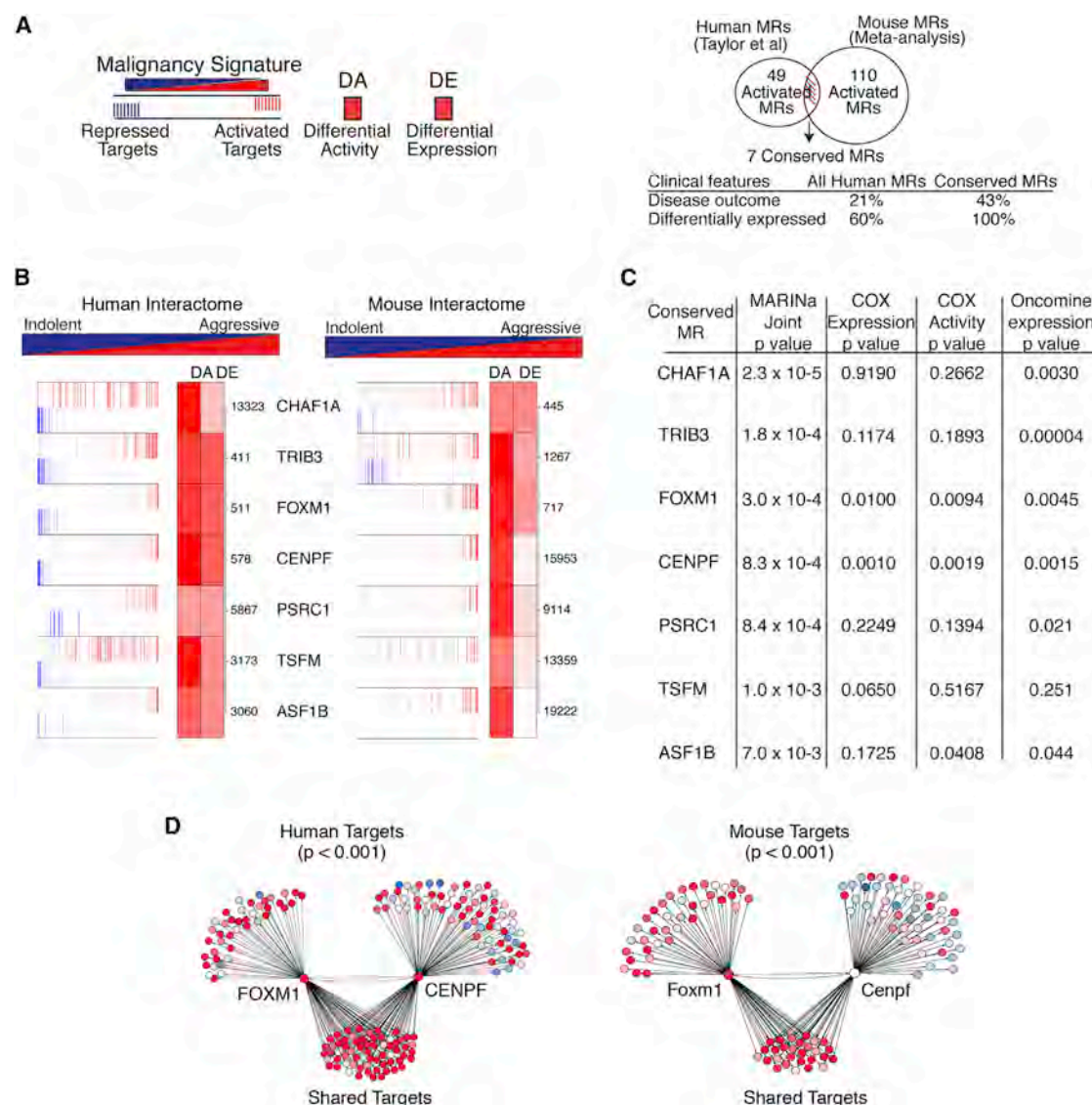
compares castrate-resistant ( $n = 29$ ) with hormone-naïve disease ( $n = 22$ ). The MRs identified from the Balk malignancy signature (Table S6) significantly overlapped with those identified from the Taylor malignancy signature (36 in common; Fisher exact test  $p < 0.0001$ ; Table S6). Furthermore, MARINA analyses of 15 independent interactomes revealed that the MRs inferred from two independent prostate cancer interactomes significantly overlapped with those inferred from the Taylor prostate cancer interactome ( $p < 7 \times 10^{-9}$  and  $p < 8 \times 10^{-20}$ , Fisher exact test), whereas the overlap of MRs inferred from 13 nonprostate

cancer-specific interactomes were orders of magnitude less significant (Supplemental Experimental Procedures). Thus, MRs of human prostate cancer malignancy are consistent across independent prostate cancer malignancy signatures, but dependent on a context-specific prostate cancer interactome.

To identify corresponding mouse MRs of malignancy, we performed MARINA on four independent GEMM signatures, which embody the diverse range of prostate cancer phenotypes represented among the GEMMs (Figure S2A; Table S2; Supplemental Experimental Procedures). Meta-analysis of the resulting MRs from these independent GEMM signatures led to identification of 229 candidate mouse MRs, including 110 activated and 119 repressed MRs ( $p \leq 0.001$ ; Figure 4A; Table S7).

The independent list of human and mouse MRs were then integrated to produce a ranked list of 20 conserved MRs, including seven activated and 13 repressed (joint  $p$  value:  $p \leq 0.0074$  by





**Figure 4. Conserved Master Regulators of Malignant Prostate Cancer**

(A) (left) Master regulators (MRs) were identified using human or mouse interactomes malignancy signatures; differential activity (DA) is based on enrichment of activated (red) and repressed (blue) targets. DE, differential expression. (Right) Venn diagram showing integration of independent lists of activated MRs from human (49) and mouse (110) with an overlap of seven conserved MRs. Clinical features of all human MRs versus the conserved MRs showing the percentage associated with disease outcome (using a COX proportional hazard model) and the percentage that are differentially expressed in advanced prostate cancer (from Oncomine).

(B) Conserved activated MRs are shown for the human (left) and mouse (right) malignancy signatures, depicting their positive (activated; red bars) and negative (repressed; blue bars) targets. The ranks of differential activity (DA) and differential expression (DE) are shown by the shaded boxes; the numbers indicate the rank of the DE in the malignancy signature.

(C) Summary of conserved MRs showing joint p value from human and mouse MARINA analysis, calculated using Stouffer's method; p value for COX proportional hazard regression model applied to mRNA expression levels and predicted MR activity; and average p values for differential expression of MRs in metastatic versus nonmetastatic primary tumors.

(D) Computational synergy analysis depicting FOXM1 and CENPF regulons from the human (left) and mouse (right) interactomes showing shared and nonshared targets. Red corresponds to overexpressed targets and blue to underexpressed targets; the p value for the enrichment of shared targets is shown.

See also Figure S2 and Tables S6 and S7.

Stouffer's method; Figures 4A and 4B; Figure S2B). Notably, these conserved MRs were more likely to be associated with disease outcome than the nonconserved ones, as assessed by a univariate COX proportional hazard regression model ( $p \leq 0.05$ ), and were also more likely to be differentially expressed

in aggressive prostate tumors (Figures 4A and 4C; Supplemental Experimental Procedures).

We focused our subsequent analysis on the activated conserved MRs, each of which has been associated with cancer-related biological processes: *CHAF1A* (chromatin activity);

*TRIB3* (regulation of cell signaling in transcriptional control); *FOXM1* (cell cycle progression); *CENPF* (mitosis); *PSRC1* (growth control); *TSFM* (translational elongation); and *ASF1B* (regulation of nucleosome assembly; Figure 4B). We further prioritized these MRs by computationally evaluating their potential synergistic interactions. In particular, any pair of MRs was considered “synergistic” if their coregulated ARACNe-inferred targets were significantly enriched in the malignancy signature relative to their individual targets ( $p \leq 0.001$ ; Carro et al., 2010; Lefebvre et al., 2010). Among all possible pairs of conserved activated MRs, the only pair that was statistically significant was *FOXM1* and *CENPF* (Figure 4D). Notably, both *FOXM1* and *CENPF* are expressed in aggressive prostate tumors and predicted to be associated with disease outcome (Figure 4C; Supplemental Experimental Procedures). Strikingly, among all activated (rather than conserved) human MRs identified, only *FOXM1* and *CENPF* were predicted to be both synergistic and of potential clinical relevance (Supplemental Experimental Procedures). Thus, cross-species analyses of conserved MRs identified a single MR synergy pair of potential clinical relevance.

### Cosilencing *FOXM1* and *CENPF* Synergistically Abrogates Prostate Tumor Growth

To evaluate their individual and potential synergistic functions in prostate cancer, we silenced *FOXM1* and/or *CENPF* individually or together in four human prostate cell lines, DU145, PC3, LNCaP, and 22Rv1, which have differing tumorigenic properties and responses to androgen signaling (Figure 5A; Figure S3A). Notably, each of these cell lines express high levels of *FOXM1* and *CENPF* mRNA; however, LNCaP does not have detectable *CENPF* protein (Figure 5B; Figures S3B–S3E), and therefore provides an excellent negative control for synergy analysis. To silence *FOXM1* and/or *CENPF*, we engineered doxycycline-inducible lentiviral vectors expressing shRNAs for *FOXM1* or *CENPF* or a control shRNA, as well as an RFP or GFP reporter (Figure 5A; Supplemental Experimental Procedures); analyses were done using two independent shRNA to minimize concerns about off-target effects (Figure S3). We distinguish “synergistic” versus “additive” effects of *FOXM1* and *CENPF* by first extrapolating their “predicted additivity” based on their individual silencing using a log-linear model, and then comparing this predicted value to the “actual” (observed) effect following their cosilencing using a one-sample *t* test; if the “actual” is statistically greater than the “predicted additive,” we conclude that *FOXM1* and *CENPF* are synergistic rather than additive (Supplemental Experimental Procedures).

Whereas individual silencing of *FOXM1*, and, to a lesser extent, *CENPF*, resulted in reduced cellular proliferation, the actual reduction following their cosilencing was statistically greater ( $p < 0.01$ ; one-sample *t* test) than the predicted additive, and is therefore synergistic for each cell line that expresses both *FOXM1* and *CENPF* proteins (Figure S3F). Similarly, with respect to colony formation, whereas individual silencing of *FOXM1* or *CENPF* reduced the number of colonies, their cosilencing resulted in nearly complete abrogation of colony formation in each cell line expressing both *FOXM1* and *CENPF* proteins ( $p < 0.001$ ; one-sample *t* test; Figures 5C and 5D; Figures S3G and S3H). Importantly, cosilencing of *FOXM1* and *CENPF* was

not associated with reduced viability, apoptosis, or further cell cycle arrest relative to their individual silencing (Figures S3I–S3K), suggesting that their observed synergy was not simply due to induction of cell death or was secondary to cell cycle arrest.

To investigate their consequences for tumor growth in vivo, we engrafted DU145 cells expressing silencing vectors for *FOXM1* and/or *CENPF* (or controls) into immunodeficient mice and monitored tumor growth (Figures 5E–5H). Consistent with the cell culture studies, individual silencing of *FOXM1* or *CENPF* resulted in a modest but statistically significant reduction in tumor growth (2-fold,  $p \leq 0.002$  and 1.5-fold,  $p \leq 0.002$ , respectively), as well as tumor weight (2.3-fold,  $p \leq 0.007$ , and 1.6 fold,  $p \leq 0.01$ , respectively) (Figures 5F and 5G). However, cosilencing of *FOXM1* and *CENPF* resulted in a complete abrogation of tumor growth (10.2-fold reduced,  $p \leq 0.000013$ ) and a profound reduction in tumor weight (12.9-fold,  $p \leq 0.000011$ ; Figures 5F and 5G). Notably, the actual inhibition of tumor growth following cosilencing of *FOXM1* and *CENPF* was significantly greater than the predicted additive inhibition (3.3-fold difference,  $p \leq 0.00026$ ; one sample *t* test; Figure 5F), supporting the conclusion that *FOXM1* and *CENPF* synergistically regulate tumor growth in vivo.

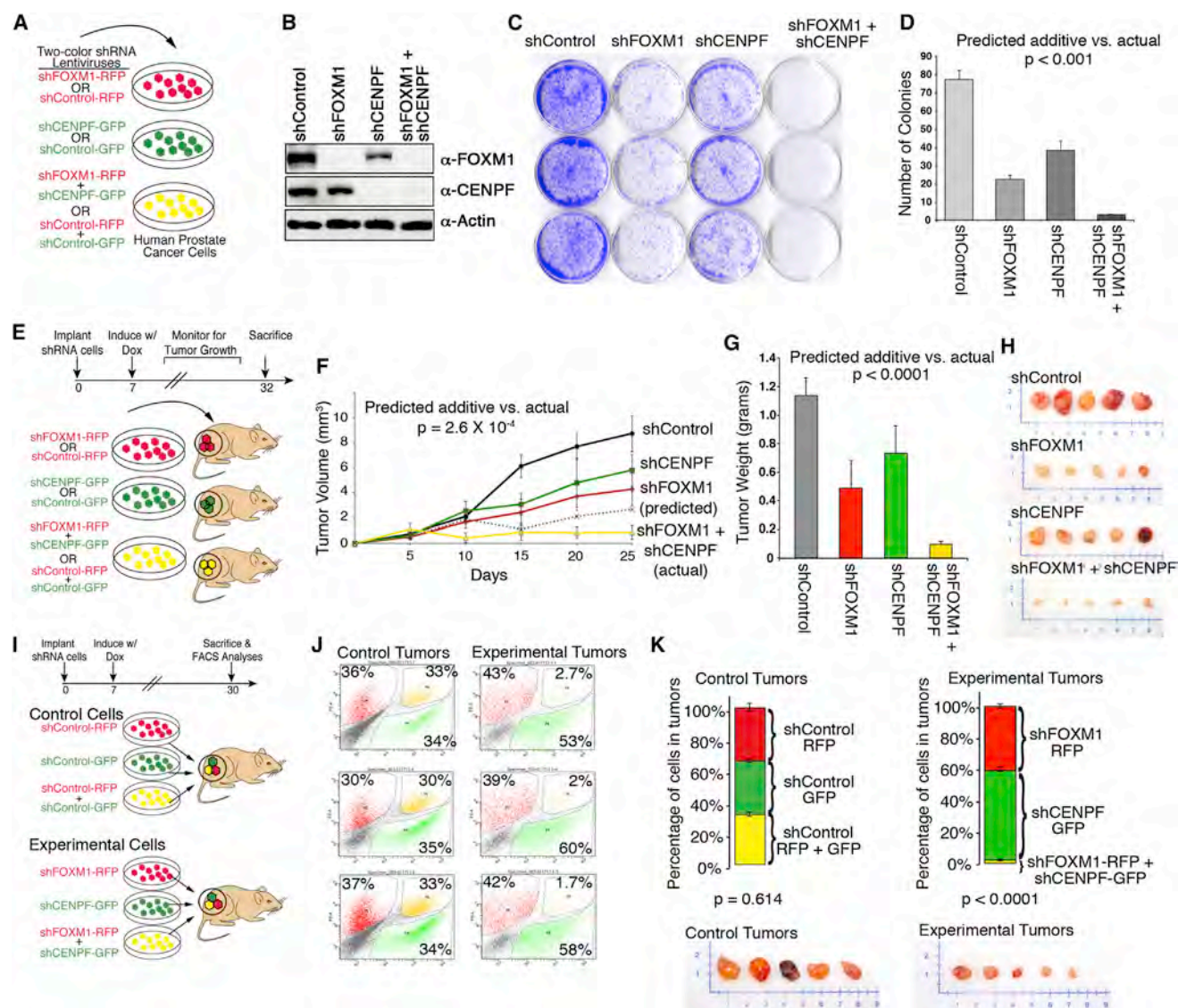
To further evaluate the synergistic activity of *FOXM1* and *CENPF* for tumor growth, we developed an in vivo competition assay (Figures 5I–5K). Specifically, we infected DU145 cells with silencing vectors expressing an *FOXM1* shRNA and an RFP reporter (red) or a *CENPF* shRNA and a GFP reporter (green), or both lentiviruses (yellow; Figure 5I). As negative controls, we infected DU145 cells with control vectors lacking the *FOXM1* or *CENPF* shRNA but expressing the fluorescent reporters. We then implanted equal numbers of viable red, green, or yellow cells from the experimental or control groups into immunodeficient mice. Following 1 month of growth in vivo, the resulting tumors were isolated and the percentage of red, green, and yellow cells were quantified by fluorescence-activated cell sorting.

Tumors derived from control cells ( $n = 4$ ) were comprised of equivalent numbers of red ( $34\% \pm 0.6\%$ ), green ( $34\% \pm 2.7\%$ ), and yellow ( $33\% \pm 1.2\%$ ) cells, indicating that the respective lentiviral vectors offer no selective growth advantage ( $p \leq 0.614$ ; Hotelling’s one-sample T-squared test; Figures 5J and 5K). In striking contrast, tumors derived from the experimental cells ( $n = 7$ ) were comprised primarily of green *CENPF*-silenced cells ( $57\% \pm 3.5\%$ ) and red *FOXM1*-silenced cells ( $41\% \pm 2.6\%$ ), whereas there were virtually no yellow cosilenced cells ( $2.0\% \pm 0.3\%$ ; Figures 5J and 5K). This profound selection against cells cosilenced for *FOXM1* and *CENPF* was highly significant ( $p \leq 0.0001$ ; Hotelling’s one-sample T-squared test; Figure 5K), which further supports the conclusion that *FOXM1* and *CENPF* synergistically regulate tumor growth in vivo.

### *FOXM1* and *CENPF* Coregulate Gene Expression and Control Tumorigenic Signaling Pathways in Prostate Cancer

To investigate the mechanism(s) underlying the observed activities of *FOXM1* and *CENPF*, we assessed the consequences of their individual versus cosilencing for expression of their ARACNe-inferred common (shared) target genes (Table S3).





**Figure 5. Functional Validation of *FOXM1* and *CENPF***

(A) Human prostate cancer cells were infected with lentiviral silencing vectors expressing shRNA for *FOXM1* and/or *CENPF* (or control) and either an RFP (red) or GFP (green) reporter. Unless otherwise indicated, analyses were done using two independent shRNAs for each gene and in four independent prostate cancer cell lines (DU145, PC3, LNCaP, 22Rv1); in most cases data using shRNA1 are shown.

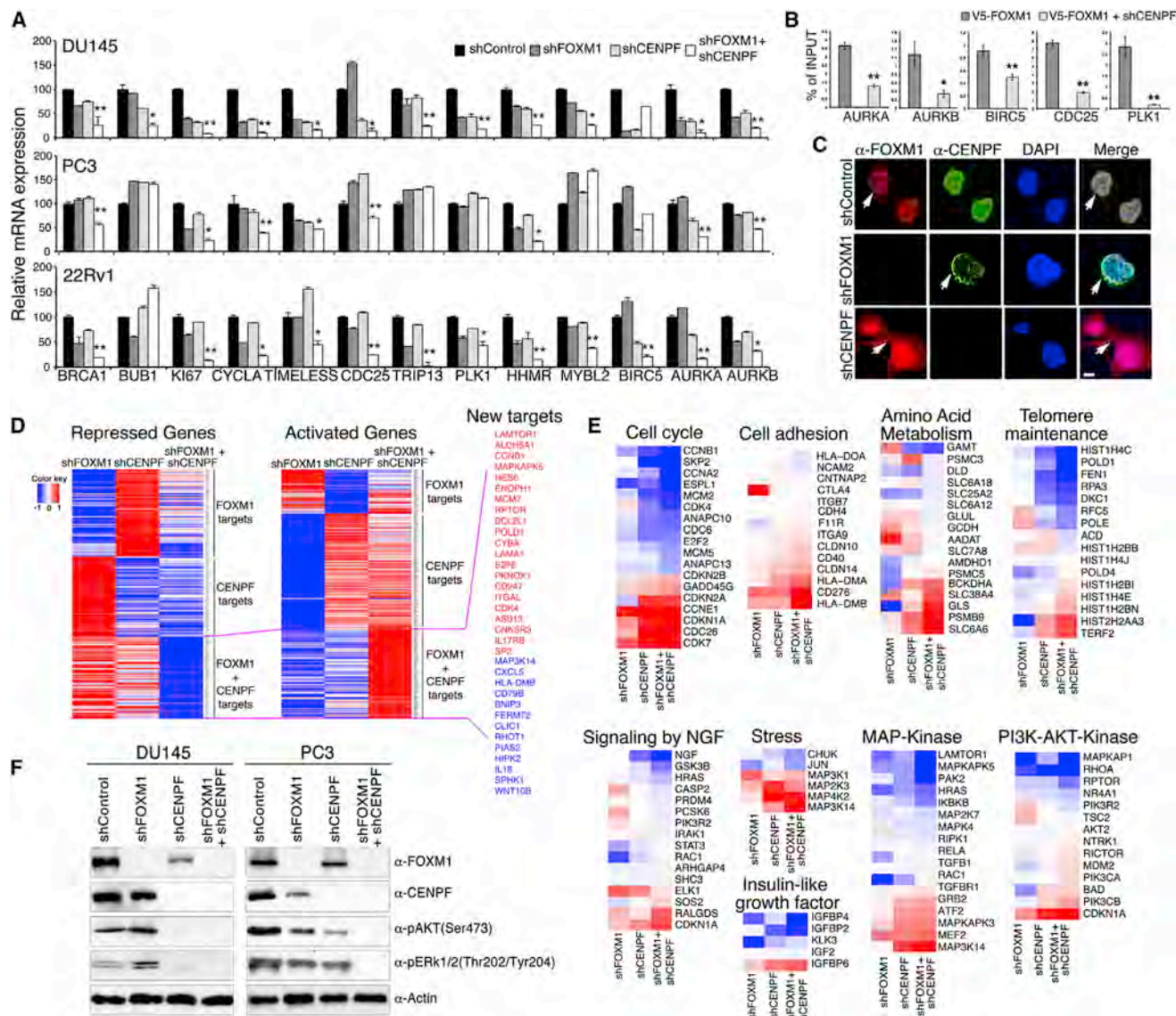
(B) Western blot analysis showing expression of FOXM1 or CENPF proteins in DU145 cells with the indicated shRNAs.

(C and D) Colony formation assay. (C) Representative analyses of DU145 cells with an shRNA for *FOXM1* and/or *CENPF* (or the control) with colonies visualized using crystal violet. (D) Quantification of colonies using ImageJ.

(E–H) Analysis of tumor growth in vivo. (E) DU145 cells expressing an shRNA for *FOXM1* and/or *CENPF*, or the control, were implanted subcutaneously into mouse hosts. Beginning on day 7, mice were administered doxycycline to induce shRNA expression and tumor growth was monitored for 1 month. (F) Tumor growth curves for the indicated shRNA. The dashed line shows the predicted additive effect of cosilencing *FOXM1* and/or *CENPF*. (G) Tumor weights at the time of sacrifice. (H) Representative tumors. In (D), (F), and (G) the predicted additive was estimated based on the consequences of individual silencing of *FOXM1* and *CENPF* using a log-linear model; the  $p$  value, calculated using a one-sample  $t$  test, indicates the significance between the predicted additive versus the actual (observed) consequences of cosilencing *FOXM1* and *CENPF*.

(I–K) In vivo competition assay. (I) Equal numbers of DU145 cells expressing the control shRNA (control cells), or the experimental shRNA for *FOXM1* and/or *CENPF* (experimental cells) as well as RFP or GFP were implanted into mouse hosts. Beginning on day 7, mice were administered doxycycline to induce shRNA expression and tumor growth was monitored for 1 month, then tumors were collected and fluorescence-activated cell sorting was performed to quantify the total number of red, green, or yellow cells in individual tumors for control and experimental groups. (J) Representative fluorescence-activated cell sorting plots showing the percentage of red, green, or yellow cells relative to the total number of fluorescent cells. (K) (top) Graphs show the average percent of red, green, and yellow cells in the control tumors ( $n = 4$ ) or experimental tumors ( $n = 7$ );  $p$  values correspond to a Hotelling's one-sample  $t$  test. (Bottom) Representative tumors. Error bars represent  $\pm$  SD.

See also Figure S3.



**Figure 6. *FOXM1* and *CENPF* Synergistically Regulate Gene Expression and Control Tumorigenic Signaling Pathways in Prostate Cancer**

(A) Validation of ARACNe-inferred shared targets of *FOXM1* and *CENPF*. The graphs show relative mRNA expression levels, normalized to *GADPH*, for the indicated genes in the cell lines shown following individual or co-silencing of *FOXM1* and *CENPF*. The p values (indicated by \*) show the significance of the predicted additive effect versus actual effect on gene expression calculated using a one-sample t test (\* $p < 0.01$ ; \*\* $p < 0.001$ ).

(B) Chromatin immunoprecipitation followed by qPCR of genomic binding sites of *FOXM1*. Cells were infected with a lentivirus expressing V5-tagged *FOXM1* as well as an shRNA for *CENPF* (or a control) and ChIP was done using an anti-V5 antibody. Data are expressed as fold enrichment of *FOXM1* binding normalized to input.

(C) Subcellular localization of *FOXM1* and *CENPF* in prostate cancer cells after silencing. Shown are microphotographs of immunofluorescence staining for *FOXM1* or *CENPF* in the control or silenced cells as indicated. Arrows indicate subcellular localization. Scale bars represent 1  $\mu$ m.

(D and E) Consequences of silencing *FOXM1* and/or *CENPF* for gene expression profiling in DU145 cells. (D) Heatmaps of differentially expressed genes. Colors correspond to levels of differential expression; red corresponds to overexpression and blue to under-expression. Selected genes differentially expressed following cosilencing are indicated. (E) Heatmaps showing leading edge genes of biological pathways enriched by cosilencing of *FOXM1* and *CENPF* as assessed by gene set enrichment analysis.

(F) Western blot analyses showing expression of the indicated markers of the PI3-kinase and MAP kinase signaling pathways in DU145 and PC3 prostate cancer cells silenced for *FOXM1* and/or *CENPF*, as indicated. Error bars represent  $\pm$  SD.

See also Figure S4 and Tables S8 and S9.

Although target gene expression was somewhat reduced by their individual silencing, cosilencing of *FOXM1* and *CENPF* produced a significantly greater reduction for the majority of targets,

consistent with coregulation of target gene expression by *FOXM1* and *CENPF* (Figure 6A; Figure S4). Furthermore, using chromatin immunoprecipitation (ChIP) followed by quantitative



PCR, we found that binding by FOXM1 to its known genomic sites on shared target genes was reduced following silencing of *CENPF* (Figure 6B), suggesting that *CENPF* is required for appropriate genomic binding by FOXM1. Interestingly, although we did not observe a direct protein-protein interaction of FOXM1 and *CENPF* in co-immunoprecipitation assays (data not shown), we observed that FOXM1 and *CENPF* were colocalized in the nucleus of prostate cancer cells and that their subcellular colocalization was mutually dependent (Figure 6C). In particular, silencing of *CENPF* resulted in the redistribution of FOXM1 to the cytoplasm as well the nucleus, and conversely silencing of *FOXM1* resulted in the accumulation of *CENPF* at the nuclear periphery (Figure 6C). Notably, subcellular colocalization of FOXM1 and *CENPF* was also observed in human prostate tumors and was associated with disease outcome (see below). Taken together, our findings suggest that FOXM1 and *CENPF* coregulate expression of shared target genes in prostate cancer cells, at least in part, through their subcellular colocalization.

To elucidate molecular pathways underlying the synergistic interaction of *FOXM1* and *CENPF* for tumor growth, we analyzed expression profiles from prostate cancer cells in which they were individually silenced or cosilenced (Figure 6D; Table S8). The differentially expressed genes following individual silencing of *FOXM1* or *CENPF* included a majority of their ARACNe-inferred targets ( $p = 0.0028$  for enrichment of FOXM1 targets;  $p \leq 0.001$  for *CENPF* targets), further confirming the accuracy of the ARACNe analysis (Table S8). Inspection of these differentially expressed genes, as well as gene set enrichment analysis of enriched biological pathways confirmed the known individual functions of *FOXM1* and *CENPF* as regulators of cellular proliferation and/or mitosis (Tables S8 and S9).

However, cosilencing of *FOXM1* and *CENPF* revealed an additional repertoire of significantly differentially expressed genes and enriched biological pathways (Figures 6D and 6E; Tables S8 and S9), including several pathways associated with tumorigenesis: “cell cycle” (normalized enrichment score [NES] 1.32;  $p \leq 0.001$ ), “stress pathway” (NES 1.58;  $p \leq 0.01$ ), “regulation of insulin-like growth factor” (NES 1.89;  $p \leq 0.001$ ), “signaling by NGF” (NES 1.25;  $p \leq 0.001$ ), “metabolism of amino acids” (NES 1.25;  $p \leq 0.01$ ), “PI3-Akt signaling” (NES 1.89;  $p \leq 0.001$ ), “MAP kinase pathway” (NES 1.34;  $p \leq 0.008$ ), “telomere maintenance” (NES 1.35;  $p \leq 0.01$ ), and “cell adhesion molecules” (NES 1.32;  $p \leq 0.001$ ).

Notable was the enrichment of PI3-kinase and MAP kinase signaling pathways following cosilencing of *FOXM1* and *CENPF* (Figure 6E; Table S9) because these constitute established hallmarks of aggressive prostate cancer (Aytes et al., 2013; Taylor et al., 2010). As evident by western blot analysis, both pathways are completely abrogated following cosilencing of *FOXM1* and *CENPF* (Figure 6F), suggesting that therapeutic targeting of FOXM1 and *CENPF* in prostate cancer cells may be effective for inactivation of these signaling pathways.

### Co-Expression of FOXM1 and CENPF Is a Prognostic Indicator for Human Prostate Cancer

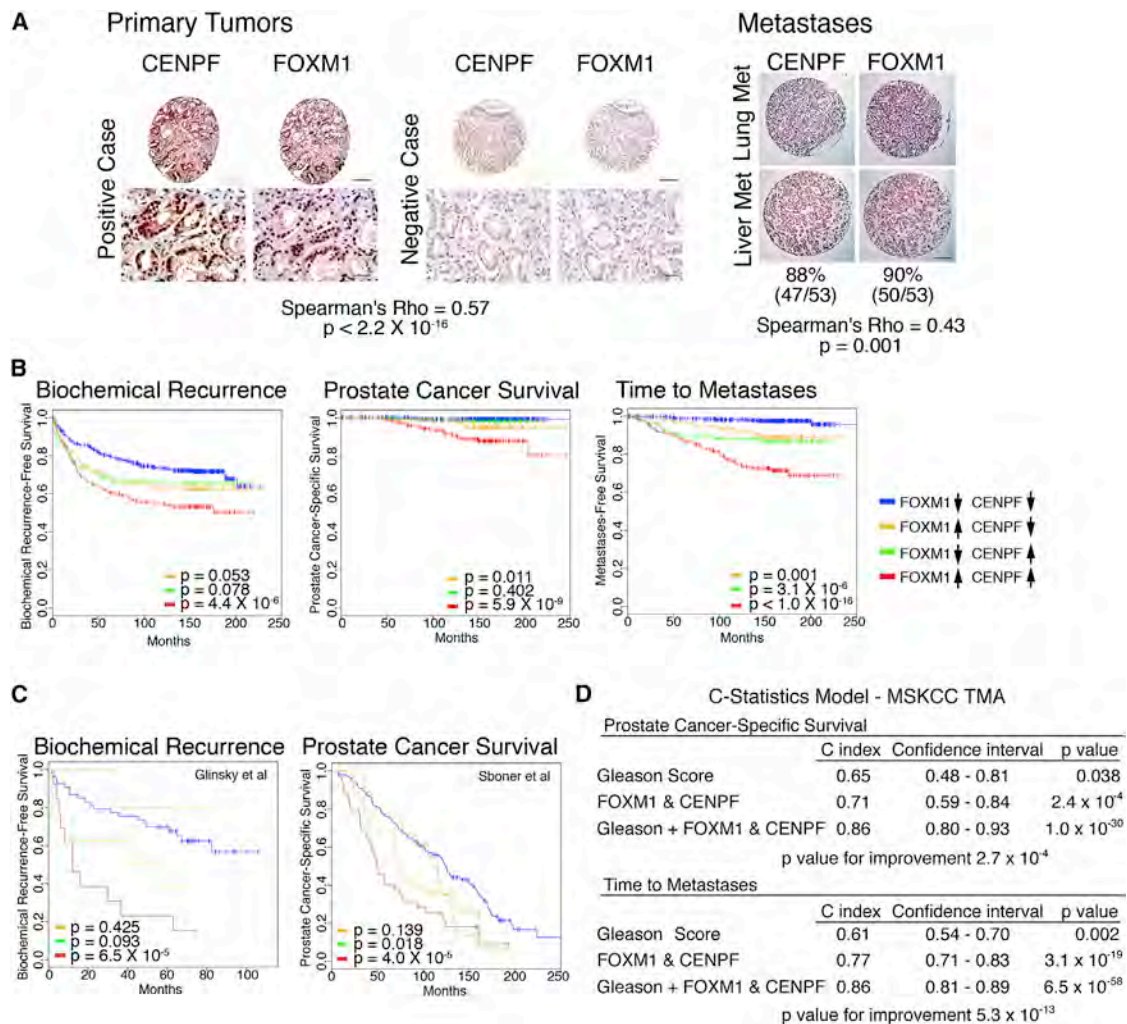
We next asked whether expression of FOXM1 and/or *CENPF* is associated with cancer progression and/or outcome by analysis of tissue microarrays (TMAs; Figure 7A; Table S1). In particular, we analyzed a high-density TMA containing primary tumors

from a large cohort of patients ( $n = 916$ ) that had undergone prostatectomy at Memorial Sloan-Kettering Cancer Center (MSKCC) from 1985 to 2003 (Donovan et al., 2008). These cases have extensive clinical follow-up data for up to 20 years, including time to biochemical recurrence, prostate cancer-specific survival, and time to metastasis (Table S1). We also evaluated a second TMA from the rapid autopsy program at the University of Michigan, which contains prostate cancer metastases ( $n = 60$ ), including 6 lung, 11 liver, 22 lymph node, and 14 other sites (Shah et al., 2004).

Analysis of the MSKCC prostatectomy TMA revealed that FOXM1 and *CENPF* were overexpressed in 33% and 37% of all cases, respectively, ( $n = 821$  informative cases) with a trend toward increased expression in tumors with higher Gleason scores (Figure 7A; Figure S5A). Furthermore, analysis of the Michigan metastasis TMA revealed that FOXM1 and *CENPF* were expressed in most of the prostate cancer metastases (88% and 90%, respectively;  $n = 53$  informative cases; Figure 7A). Moreover, FOXM1 and *CENPF* were frequently co-expressed and colocalized in the nucleus in both the MSKCC prostatectomy TMA (Spearman's  $Rho = 0.57$ ,  $p \leq 2 \times 10^{-16}$ ) and the Michigan metastasis TMA (Spearman's  $Rho = 0.43$ ,  $p \leq 1 \times 10^{-3}$ ; Figure 7A). Additionally, at the mRNA level, overexpression of *FOXM1* and *CENPF* was well correlated in advanced prostate cancer and metastases from independent cohorts of human prostate cancer (Figure S5B).

To determine whether expression of FOXM1 and/or *CENPF* is associated with disease outcome, we first defined four groups of patients from the MSKCC TMA based on their expression levels: (1) low/normal expression of both FOXM1 and *CENPF* ( $n = 418$ ); (2) high expression of FOXM1 and low/normal expression of *CENPF* ( $n = 97$ ); (3) high expression of *CENPF* and low/normal expression of FOXM1 ( $n = 133$ ); and (4) high expression of both FOXM1 and *CENPF* ( $n = 173$ ). Kaplan-Meier survival analysis revealed that patients having elevated expression of both FOXM1 and *CENPF* had the worst outcome for three independent clinical endpoints, namely, time to biochemical-free recurrence ( $p \leq 4.4 \times 10^{-6}$ ), death due to prostate cancer ( $p \leq 5.9 \times 10^{-9}$ ), and time to metastasis ( $p \leq 1.0 \times 10^{-16}$ ; Figure 7B). Notably, subcellular colocalization of FOXM1 and *CENPF* in prostate tumors was also associated with the worst outcome for all three independent clinical endpoints (Figure S5A). In contrast, elevated expression of only FOXM1 or *CENPF* was either not significant or marginally significant for biochemical recurrence and prostate-specific survival ( $p \leq 0.053$  and  $p \leq 0.011$  for FOXM1, respectively;  $p \leq 0.078$  and  $p \leq 0.402$  for *CENPF*, respectively), and was 10 to 13 orders of magnitude less significant for time to metastasis ( $p \leq 0.001$  for FOXM1 and  $p \leq 3.1 \times 10^{-6}$  for *CENPF*; Figure 7B).

We independently corroborated the association of *FOXM1* and *CENPF* with disease outcome in two independent human prostate cancer data sets that had not been used for training purposes elsewhere in this study; namely, the Glinsky data set, in which biochemical recurrence is the clinical endpoint (Glinsky et al., 2004), and the Sboner data set, in which the clinical endpoint is prostate cancer-specific overall survival (Sboner et al., 2010; Table S1). Using these independent cohorts, we evaluated the mRNA expression levels of *FOXM1* and *CENPF* as well as their MARINA-inferred transcriptional activity



**Figure 7. Clinical Validation of FOXM1 and CENPF in Human Prostate Cancer**

(A) Analysis of tissue microarrays (TMAs). Representative images from the MSKCC prostatectomy TMA and the Michigan metastasis TMA showing FOXM1 and CENPF protein expression; Spearman correlation of their co-expression with p value is shown. Scale bars represent 10  $\mu$ m (lower primary tumor) or 100  $\mu$ m (all others).

(B) Kaplan-Meier survival analysis based on protein expression levels of FOXM1 and CENPF in MSKCC prostatectomy TMA with respect to time to biochemical recurrence, time to prostate cancer-specific death, or time to metastatic progression.

(C) Kaplan-Meier survival analysis based on the ARACNe-inferred activity levels of FOXM1 and CENPF (Supplemental Experimental Procedures) in two independent human prostate cancer data sets using biochemical recurrence-free survival (Glinsky et al., 2004) or prostate cancer-specific survival (Sboner et al., 2010) as disease endpoints. In (B) and (C), the p values correspond to a log-rank test and indicate the statistical significance of the association with outcome for each indicated branch compared to control (i.e., patients with low activity levels of both FOXM1 and CENPF, blue line curve).

(D) C-statistics analysis, based on the protein levels of FOXM1 and CENPF from the MSKCC TMA, using death due to prostate cancer and time to metastasis as evaluation endpoints.

See also Figure S5.

(Supplemental Experimental Procedures; Figures 7C and S5C). We then performed Kaplan-Meier survival analysis comparing four patient groups: (1) those with low inferred activity or expression for FOXM1 and CENPF; (2) those with high inferred activity or expression only for FOXM1; (3) those with high inferred activity or expression only for CENPF; and (4) those with high inferred activity or expression for both FOXM1 and CENPF. Similar to our analysis of the TMA, patients with high inferred activity or mRNA expression for both CENPF and FOXM1 were associated with the worst outcome in both cohorts, as measured by

biochemical recurrence ( $p \leq 0.000065$ ) and prostate cancer-specific survival ( $p \leq 0.000040$ ; (Figures 7C and S5C). Notably, these findings reveal that their MARINA-inferred activities are well correlated with the actual protein expression of FOXM1 and CENPF, and further demonstrate the striking association of their co-expression/co-activity with poor disease outcome.

Finally, association of FOXM1 and CENPF protein expression with disease outcome using C-statistics revealed its robust prognostic value for disease-specific survival ( $C = 0.71$ ; CI 0.59–0.84,  $p \leq 0.00024$ ), as well as time to metastasis ( $C = 0.77$ ;

CI 0.71–0.83,  $p \leq 3.0 \times 10^{-19}$ ; Figure 7D). Notably, co-expression of FOXM1 and CENPF proteins dramatically improves prognosis over Gleason score alone for both disease-specific survival ( $C = 0.86$ ; CI 0.80–0.93,  $p \leq 1.0 \times 10^{-30}$ ;  $p$  value for improvement  $p \leq 0.00020$ ) and time to metastasis ( $C = 0.86$ ; CI 0.81–0.89,  $p \leq 6.5 \times 10^{-58}$ ;  $p$  value for improvement,  $p \leq 5.3 \times 10^{-13}$ ; Figure 7D). Taken together, these analyses of independent clinical cohorts and using distinct statistical models demonstrate that co-expression of FOXM1 and CENPF is a robust prognostic indicator of poor disease outcome and metastasis.

## DISCUSSION

Recent advances in whole-genome analyses are providing an increasingly high-resolution view of the multitude of genetic, genomic, and epigenetic alterations associated with cancer phenotypes. Given the staggering number of potential interactions, identification of the true causal drivers and essential synergistic interactions represents a considerable challenge. In the current study, we have demonstrated that cross-species interrogation of genome-wide context-specific regulatory networks can address this challenge by dramatically winnowing the candidate gene interactions that implement the regulatory programs underlying cancer phenotypes. In particular, we have introduced a comprehensive systems approach to interrogate complementary regulatory networks for human and mouse prostate cancer to identify conserved causal regulators and to elucidate synergistic interactions among them. These studies have led to the identification of FOXM1 and CENPF as synergistic master regulators of prostate cancer malignancy and robust prognostic biomarkers of aggressive prostate cancer. We propose that this overall approach for genome-wide cross-species analysis will be generally applicable for identifying synergistic interactions that drive physiologic and pathologic phenotypes in cancer and other diseases.

The genome-wide assembly and cross-species interrogation of human and mouse prostate cancer regulatory networks represents a major conceptual advance. A critical requirement was the generation of a mouse prostate cancer interactome from a data set of appropriate size and expression heterogeneity. We incorporated the diversity afforded by genetically and phenotypically distinct mouse models of prostate cancer obtained through a community effort (see Supplemental Experimental Procedures), in combination with exogenous perturbations administered to each mouse model. This strategy has led to the successful construction of a genome-wide, context-specific mouse interactome for the study of prostate cancer, and is generalizable for the generation of interactomes for a wide range of physiologic and pathologic phenotypes.

A second critical requirement was the development of an informative algorithm to determine whether the human and mouse prostate cancer interactomes represented conserved regulatory programs, thus enabling accurate and robust cross-species integrative analysis. Toward this end, we introduced a metric for quantitative assessment of conservation of regulatory networks, which revealed that the large majority of regulatory programs represented by these networks are highly conserved (>70%). Although the current study is focused on

prostate cancer interactomes assembled using ARACNe, this general approach for evaluating conservation can be used for cross-species analyses of regulatory networks for other cancers or other diseases and can be readily adapted for analyses of networks inferred using alternative algorithms, such as those based on the Context Likelihood of Relatedness and Bayesian-networks algorithms (Akavia et al., 2010; Faith et al., 2007). Indeed, we envision that the ability to quantitatively evaluate conservation of cross-species regulatory programs will be broadly applicable for other physiological and pathological comparisons, and particularly beneficial for accurate integration of preclinical findings from genetically engineered mice to human clinical trials.

A third critical requirement for the success of our approach was our ability to effectively mine these cross-species regulatory networks to identify conserved master regulators of cancer malignancy and to identify their synergistic interactions. The MARINA algorithm used for these analyses defines “master regulators” as genes that most significantly regulate the transcriptional program associated with a particular phenotype (in this case, prostate cancer malignancy), and hence are rate-limiting drivers of the phenotype (Carro et al., 2010; Lefebvre et al., 2010). Notably, MARINA also provides an effective computational tool for analyses of synergistic interactions among master regulators (Carro et al., 2010); indeed, our unbiased interrogation of ~2,000 transcriptional regulators represented in the interactomes led to identification of a single synergistic pair, namely FOXM1 and CENPF. The power of this approach suggests that it may be of general value in dissecting polygenic dependencies in cancer and other diseases.

Although both FOXM1 and CENPF have been implicated in various cancers, our study has uncovered their unexpected synergistic interaction. FOXM1 encodes a forkhead domain transcription factor that is frequently overexpressed in many different types of cancer, including prostate (Alvarez-Fernandez and Medema, 2013; Halasi and Gartel, 2013a; Kalin et al., 2011; Koo et al., 2012). Many previous studies have established a role for FOXM1 in regulation of cellular proliferation, DNA damage, genomic stability, drug resistance, and metastasis, and have shown that it interacts with other key regulators such as  $\beta$ -catenin and MYB (Lefebvre et al., 2010; Zhang et al., 2011). In prostate cancer, gain- or loss-of-function of FOXM1 in vivo have been shown to elicit modest effects on tumor growth (Cai et al., 2013; Kalin et al., 2006).

CENPF (also known as mitotin or LEK1 in mouse), a known target of FOXM1, has also been implicated in various cancers, although not previously in prostate, and in some cases has been shown to undergo gene amplification and to be associated with disease outcome (Ma et al., 2006; Varis et al., 2006). However, the actual functional role of CENPF has been more elusive and difficult to reconcile. In particular, whereas CENPF is named for its association with the centromere-kinetochore protein complex, such association is transient and, in fact, CENPF has other functions, including regulation of mitosis and cellular proliferation (Bomont et al., 2005; Feng et al., 2006; Holt et al., 2005), which are mediated in part by protein interactions (Ma et al., 2006; Varis et al., 2006).

Thus, although the individual functions of FOXM1 and CENPF in cancer had been well studied, their synergistic interaction



could not have been anticipated from previous analyses. Cumulatively, our findings suggest that co-expression of *FOXM1* and *CENPF* in prostate cancer leads to coregulation of transcriptional programs, which ultimately result in activation of the key signaling pathways associated with prostate cancer malignancy, including the PI3K and MAPK signaling pathways. Because *FOXM1* and *CENPF* can each be targeted pharmacologically (Halasi and Gartel, 2013b; Pan and Yeung, 2005; Radhakrishnan et al., 2006), their inhibition may provide an effective means of treating advanced prostate cancer; indeed, therapeutic targeting of *FOXM1* and *CENPF* may help overcome the complex feedback mechanisms that have hindered therapeutic targeting of PI3K and MAPK signaling pathways.

Furthermore, we envision that by using alternative gene signatures that represent other prostate cancer phenotypes, genome-wide cross-species analysis of master regulators and their potential synergistic interactions may help to define molecular subtypes of prostate cancer, which have thus far been elusive. More broadly, our general approach to elucidate conserved and functionally relevant gene interactions can be applied to many tumor contexts as well as other human diseases for which appropriate model systems are available.

## EXPERIMENTAL PROCEDURES

### Assembly of Interactomes and Master Regulator Analyses

Expression profile data sets for human prostate cancer are described in Table S1. GEMMs are described in Table S2 and their representative histopathology shown in Figure S1A. A description of perturbation treatments is provided in the Supplemental Experimental Procedures. Human and mouse interactomes were assembled using the ARACNe algorithm (Margolin et al., 2006a). Details of the resulting human and mouse networks are provided in Tables S3 and S4, respectively. Analysis of cross-species network conservation was done using a modification of the MARINA algorithm described in the Supplemental Experimental Procedures. Master regulator analysis and computational synergy analysis were performed using MARINA (Carro et al., 2010; Lefebvre et al., 2010). Master regulators for the human and mouse interactomes are provided in Tables S6 and S7, respectively.

### Functional Validation

Gene silencing of *FOXM1* and *CENPF* as well as forced expression of *FOXM1* were done using lentiviral shRNAs or expression vectors (Open Biosystems and CCSB Human ORFeome Library, respectively). Human cancer cell lines used for functional studies were obtained from ATCC. All experiments using animals were performed according to protocols approved by the Institutional Animal Care and Use Committee at Columbia University Medical Center.

### Tissue Microarray Analyses

All studies involving human subjects were approved by the Institutional Review Board of MSKCC or University of Michigan. TMAs were constructed with approval from the Human Biospecimen Utilization Committee and Institutional Review Board; consent was obtained from all patients, as required. Analysis of protein expression of *FOXM1* and *CENPF* was performed using a high-density TMA (Donovan et al., 2008) and a metastasis TMA (Shah et al., 2004). Available clinicopathological features of these TMAs are summarized in Table S1.

### Statistical Methods

Statistical analysis was performed with survcomp package using R v2.14.0. Cox proportional hazard model was estimated with the surv and coxph functions. Kaplan-Meier survival analysis was performed using surv, survfit, and survdiff functions. Concordance indexes (c-index) were estimated and compared using coxph and concordance.index (counting ties) and cindex.comp functions. Details of all statistical analyses and all computational procedures are provided in the Supplemental Experimental Procedures. An

executable SWEAVE document and R data objects are deposited in Figshare at <http://dx.doi.org/10.6084/m9.figshare.928353>.

## ACCESSION NUMBERS

The Gene Expression Omnibus accession number for the mouse expression profiles reported in this paper is GSE53202.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and nine tables and can be found with this article online at <http://dx.doi.org/10.1016/j.ccr.2014.03.017>.

## AUTHOR CONTRIBUTIONS

A.A., A.M., M.J.A., M.M.S., A.C., and C.A.-S. designed experiments, analyzed results, and wrote the manuscript; C.L. contributed to initial interactome assembly and analyses; M.C.-M. provided pathological expertise; T.Z. provided statistical expertise; and J.A.E., A.G., and K.L.P. provided critical reagents.

## ACKNOWLEDGMENTS

We are indebted to our colleagues who generously provided genetically engineered mouse models for assembly of the mouse prostate cancer interactome, namely Drs. Bart Williams (Van Andel), Terry Van Dyke (NCI-Frederick), Barbara Foster (Roswell Park Cancer Center), Yu Chen (MSKCC), and Charles Sawyers (MSKCC). We are also indebted to Dr. Victor Reuter (MSKCC) for generously sharing the prostatectomy tissue microarray. We thank Jose Silva and members of his laboratory for invaluable help with establishment of the shRNA validation studies and Drs. Adolfo Ferrando and Riccardo Dalla Favera for critical reading of the manuscript.

This work was supported by grants CA084294 (to C.A.S., M.M.S., and A.C.), U54 CA121852 (to A.C., C.A.S., and M.M.S.), CA154293 (to M.M.S. and C.A.S.), DK076602 (to M.M.S.), Cancer Target Discovery and Development Centers NCI U01 CA168426 (to A.C.), the Michigan Center for Translational Pathology, SPOR grant P50 CA69568 (to K.J.P.), and an award from the V-Foundation for Cancer Research (to C.A.S.). A.A. was a recipient of a Marie Curie International Outgoing Fellowship (PIOF-GA-2009-253290), cosponsored with the Catalan Institute of Oncology-Bellvitge Institute for Biomedical Research, Barcelona, Spain. A.M. is a recipient of a Prostate Cancer Foundation Young Investigator Award. C.A.S. is an American Cancer Society Research Professor supported in part by a generous gift from the F.M. Kirby Foundation.

Received: August 31, 2013

Revised: January 9, 2014

Accepted: March 14, 2014

Published: May 12, 2014

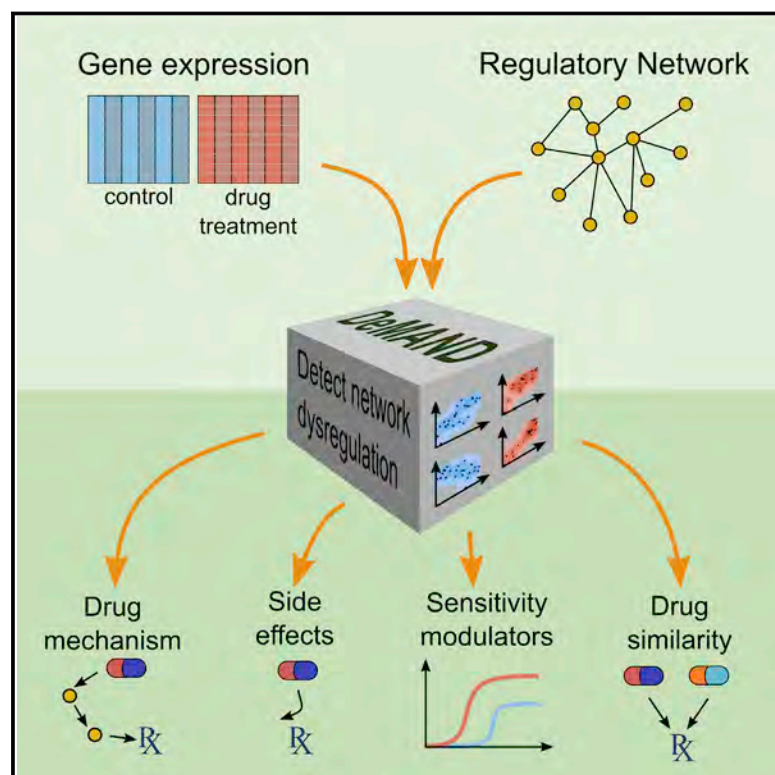
## REFERENCES

- Akavia, U.D., Litvin, O., Kim, J., Sanchez-Garcia, F., Kotliar, D., Causton, H.C., Pochanard, P., Mozes, E., Garraway, L.A., and Pe'er, D. (2010). An integrated approach to uncover drivers of cancer. *Cell* 143, 1005–1017.
- Alvarez-Fernandez, M., and Medema, R.H. (2013). Novel functions of FoxM1: from molecular mechanisms to cancer therapy. *Front. Oncol.* 3, 30.
- Aytes, A., Mitrofanova, A., Kinkade, C.W., Lefebvre, C., Lei, M., Phelan, V., LeKaye, H.C., Koutcher, J.A., Cardiff, R.D., Califano, A., et al. (2013). ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. *Proc. Natl. Acad. Sci. USA* 110, E3506–E3515.
- Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., Ghandi, M., et al. (2013). Punctuated evolution of prostate cancer genomes. *Cell* 153, 666–677.
- Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* 37, 382–390.

- Bomont, P., Maddox, P., Shah, J.V., Desai, A.B., and Cleveland, D.W. (2005). Unstable microtubule capture at kinetochores depleted of the centromere-associated protein CENP-F. *EMBO J.* 24, 3927–3939.
- Cai, Y., Balli, D., Ustiyani, V., Fulford, L.A., Hiller, A., Misetic, V., Zhang, Y., Paluch, A.M., Waltz, S.E., Kasper, S., and Kalin, T.V. (2013). Foxm1 expression in prostate epithelial cells is essential for prostate carcinogenesis. *J. Biol. Chem.* 288, 22527–22541.
- Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H., et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 463, 318–325.
- Cooperberg, M.R., Broering, J.M., Kantoff, P.W., and Carroll, P.R. (2007). Contemporary trends in low risk prostate cancer: risk assessment and treatment. *J. Urol.* 178, S14–S19.
- Donovan, M.J., Hamann, S., Clayton, M., Khan, F.M., Sapir, M., Bayer-Zubek, V., Fernandez, G., Mesa-Tejada, R., Teverovskiy, M., Reuter, V.E., et al. (2008). Systems pathology approach for the prediction of prostate cancer progression after radical prostatectomy. *J. Clin. Oncol.* 26, 3923–3929.
- Faith, J.J., Hayete, B., Thaden, J.T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J.J., and Gardner, T.S. (2007). Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* 5, e8.
- Feng, J., Huang, H., and Yen, T.J. (2006). CENP-F is a novel microtubule-binding protein that is essential for kinetochore attachments and affects the duration of the mitotic checkpoint delay. *Chromosoma* 115, 320–329.
- Glinsky, G.V., Glinskii, A.B., Stephenson, A.J., Hoffman, R.M., and Gerald, W.L. (2004). Gene expression profiling predicts clinical outcome of prostate cancer. *J. Clin. Invest.* 113, 913–923.
- Halasi, M., and Gartel, A.L. (2013a). FOX(M1) news—it is cancer. *Mol. Cancer Ther.* 12, 245–254.
- Halasi, M., and Gartel, A.L. (2013b). Targeting FOXM1 in cancer. *Biochem. Pharmacol.* 85, 644–652.
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674.
- Holt, S.V., Vergnolle, M.A., Hussein, D., Wozniak, M.J., Allan, V.J., and Taylor, S.S. (2005). Silencing Cenp-F weakens centromeric cohesion, prevents chromosome alignment and activates the spindle checkpoint. *J. Cell Sci.* 118, 4889–4900.
- Irshad, S., and Abate-Shen, C. (2013). Modeling prostate cancer in mice: something old, something new, something premalignant, something metastatic. *Cancer Metastasis Rev.* 32, 109–122.
- Ittmann, M., Huang, J., Radaelli, E., Martin, P., Signoretti, S., Sullivan, R., Simons, B.W., Ward, J.M., Robinson, B.D., Chu, G.C., et al. (2013). Animal models of human prostate cancer: the consensus report of the New York meeting of the Mouse Models of Human Cancers Consortium Prostate Pathology Committee. *Cancer Res.* 73, 2718–2736.
- Kalin, T.V., Wang, I.C., Ackerson, T.J., Major, M.L., Detrisac, C.J., Kalinichenko, V.V., Lyubimov, A., and Costa, R.H. (2006). Increased levels of the FoxM1 transcription factor accelerate development and progression of prostate carcinomas in both TRAMP and LADY transgenic mice. *Cancer Res.* 66, 1712–1720.
- Kalin, T.V., Ustiyani, V., and Kalinichenko, V.V. (2011). Multiple faces of FoxM1 transcription factor: lessons from transgenic mouse models. *Cell Cycle* 10, 396–405.
- Koo, C.Y., Muir, K.W., and Lam, E.W. (2012). FOXM1: From cancer initiation to progression and treatment. *Biochim. Biophys. Acta* 1819, 28–37.
- Lefebvre, C., Rajbhandari, P., Alvarez, M.J., Bandaru, P., Lim, W.K., Sato, M., Wang, K., Sumazin, P., Kustagi, M., Bisikirska, B.C., et al. (2010). A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* 6, 377.
- Lefebvre, C., Rieckhof, G., and Califano, A. (2012). Reverse-engineering human regulatory networks. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 4, 311–325.
- Ma, L., Zhao, X., and Zhu, X. (2006). Mitosin/CENP-F in mitosis, transcriptional control, and differentiation. *J. Biomed. Sci.* 13, 205–213.
- Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006a). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7 (Suppl 1), S7.
- Margolin, A.A., Wang, K., Lim, W.K., Kustagi, M., Nemenman, I., and Califano, A. (2006b). Reverse engineering cellular networks. *Nat. Protoc.* 1, 662–671.
- Pan, J., and Yeung, S.C. (2005). Recent advances in understanding the anti-neoplastic mechanisms of farnesyltransferase inhibitors. *Cancer Res.* 65, 9109–9112.
- Radhakrishnan, S.K., Bhat, U.G., Hughes, D.E., Wang, I.C., Costa, R.H., and Gartel, A.L. (2006). Identification of a chemical inhibitor of the oncogenic transcription factor forkhead box M1. *Cancer Res.* 66, 9731–9735.
- Ryan, C.J., and Tindall, D.J. (2011). Androgen receptor rediscovered: the new biology and targeting the androgen receptor therapeutically. *J. Clin. Oncol.* 29, 3651–3658.
- Sboner, A., Demichelis, F., Calza, S., Pawitan, Y., Setlur, S.R., Hoshida, Y., Perner, S., Adami, H.O., Fall, K., Mucci, L.A., et al. (2010). Molecular sampling of prostate cancer: a dilemma for predicting disease progression. *BMC Med. Genomics* 3, 8.
- Scher, H.I., and Sawyers, C.L. (2005). Biology of progressive, castration-resistant prostate cancer: directed therapies targeting the androgen-receptor signaling axis. *J. Clin. Oncol.* 23, 8253–8261.
- Schoenborn, J.R., Nelson, P., and Fang, M. (2013). Genomic profiling defines subtypes of prostate cancer with the potential for therapeutic stratification. *Clin. Cancer Res.* 19, 4058–4066.
- Shah, R.B., Mehra, R., Chinnaiyan, A.M., Shen, R., Ghosh, D., Zhou, M., Macvicar, G.R., Varambally, S., Harwood, J., Bismar, T.A., et al. (2004). Androgen-independent prostate cancer is a heterogeneous group of diseases: lessons from a rapid autopsy program. *Cancer Res.* 64, 9209–9216.
- Shen, M.M., and Abate-Shen, C. (2010). Molecular genetics of prostate cancer: new prospects for old challenges. *Genes Dev.* 24, 1967–2000.
- Stanbrough, M., Bubley, G.J., Ross, K., Golub, T.R., Rubin, M.A., Penning, T.M., Febbo, P.G., and Balk, S.P. (2006). Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. *Cancer Res.* 66, 2815–2825.
- Taylor, B.S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B.S., Arora, V.K., Kaushik, P., Cerami, E., Reva, B., et al. (2010). Integrative genomic profiling of human prostate cancer. *Cancer Cell* 18, 11–22.
- Varis, A., Salmela, A.L., and Kallio, M.J. (2006). Cenp-F (mitosin) is more than a mitotic marker. *Chromosoma* 115, 288–295.
- Zhang, N., Wei, P., Gong, A., Chiu, W.T., Lee, H.T., Colman, H., Huang, H., Xue, J., Liu, M., Wang, Y., et al. (2011). FoxM1 promotes  $\beta$ -catenin nuclear localization and controls Wnt target-gene expression and glioma tumorigenesis. *Cancer Cell* 20, 427–442.
- Zhang, Q.C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C.A., Bisikirska, B., Lefebvre, C., Accili, D., Hunter, T., et al. (2012). Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* 490, 556–560.

# Elucidating Compound Mechanism of Action by Network Perturbation Analysis

## Graphical Abstract



## Authors

Jung Hoon Woo, Yishai Shimoni, Wan Seok Yang, ..., Brent R. Stockwell, Mukesh Bansal, Andrea Califano

## Correspondence

mb3113@cumc.columbia.edu (M.B.), califano@cumc.columbia.edu (A.C.)

## In Brief

The mechanism of action (MoA) of small-molecule compounds is elucidated by analyzing regulatory networks to identify proteins whose interactions are affected following compound perturbation. Experimental validation of novel MoA predictions revealed that the anticancer drug altretamine acts as an inhibitor of *GPX4* lipid repair activity, revealing unexpected similarity to the activity of sulfasalazine.

## Highlights

- DeMAND—a method to predict genes involved in mechanism of action of a compound
- DeMAND predictions can be used to identify compound similarity
- Known MoA genes are identified with high precision, sensitivity, and specificity
- Novel predictions of both MoA and similarity were experimentally validated

## Accession Numbers

GSE60408



Woo et al., 2015, Cell 162, 441–451  
 July 16, 2015 ©2015 Elsevier Inc.  
<http://dx.doi.org/10.1016/j.cell.2015.05.056>

# Elucidating Compound Mechanism of Action by Network Perturbation Analysis

Jung Hoon Woo,<sup>1,13</sup> Yishai Shimoni,<sup>2,3,13</sup> Wan Seok Yang,<sup>4,13</sup> Prem Subramaniam,<sup>2,3</sup> Archana Iyer,<sup>2,3</sup> Paola Nicoletti,<sup>2,3</sup> María Rodríguez Martínez,<sup>2,3,12</sup> Gonzalo López,<sup>2,3</sup> Michela Mattioli,<sup>5</sup> Ronald Realubit,<sup>6</sup> Charles Karan,<sup>6</sup> Brent R. Stockwell,<sup>2,4,7,8</sup> Mukesh Bansal,<sup>2,3,14,\*</sup> and Andrea Califano<sup>1,2,3,9,10,11,14,\*</sup>

<sup>1</sup>Department of Biomedical Informatics (DBMI), Columbia University, New York, NY 10032, USA

<sup>2</sup>Department of Systems Biology, Columbia University, New York, NY 10032, USA

<sup>3</sup>Center for Computational Biology and Bioinformatics (C2B2), Columbia University, New York, NY 10032, USA

<sup>4</sup>Department of Biological Sciences, Columbia University, New York, NY 10027, USA

<sup>5</sup>Center for Genomic Science of IIT@SEMM, Fondazione Istituto Italiano di Tecnologia (IIT), 20139 Milano, Italy

<sup>6</sup>Columbia Genome Center, High Throughput Screening Facility, Columbia University, New York, NY 10032, USA

<sup>7</sup>Department of Chemistry, Columbia University, New York, NY 10027, USA

<sup>8</sup>Howard Hughes Medical Institute, Columbia University, New York, NY 10027, USA

<sup>9</sup>Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA

<sup>10</sup>Institute for Cancer Genetics, Columbia University, New York, NY 10032, USA

<sup>11</sup>Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY 10032, USA

<sup>12</sup>Present address: IBM Research-Zurich, 8803 Rüschlikon, Switzerland

<sup>13</sup>Co-first author

<sup>14</sup>Co-senior author

\*Correspondence: [mb3113@cumc.columbia.edu](mailto:mb3113@cumc.columbia.edu) (M.B.), [califano@cumc.columbia.edu](mailto:califano@cumc.columbia.edu) (A.C.)

<http://dx.doi.org/10.1016/j.cell.2015.05.056>

## SUMMARY

**Genome-wide identification of the mechanism of action (MoA) of small-molecule compounds characterizing their targets, effectors, and activity modulators represents a highly relevant yet elusive goal, with critical implications for assessment of compound efficacy and toxicity. Current approaches are labor intensive and mostly limited to elucidating high-affinity binding target proteins. We introduce a regulatory network-based approach that elucidates genome-wide MoA proteins based on the assessment of the global dysregulation of their molecular interactions following compound perturbation. Analysis of cellular perturbation profiles identified established MoA proteins for 70% of the tested compounds and elucidated novel proteins that were experimentally validated. Finally, unknown-MoA compound analysis revealed altretamine, an anticancer drug, as an inhibitor of glutathione peroxidase 4 lipid repair activity, which was experimentally confirmed, thus revealing unexpected similarity to the activity of sulfasalazine. This suggests that regulatory network analysis can provide valuable mechanistic insight into the elucidation of small-molecule MoA and compound similarity.**

## INTRODUCTION

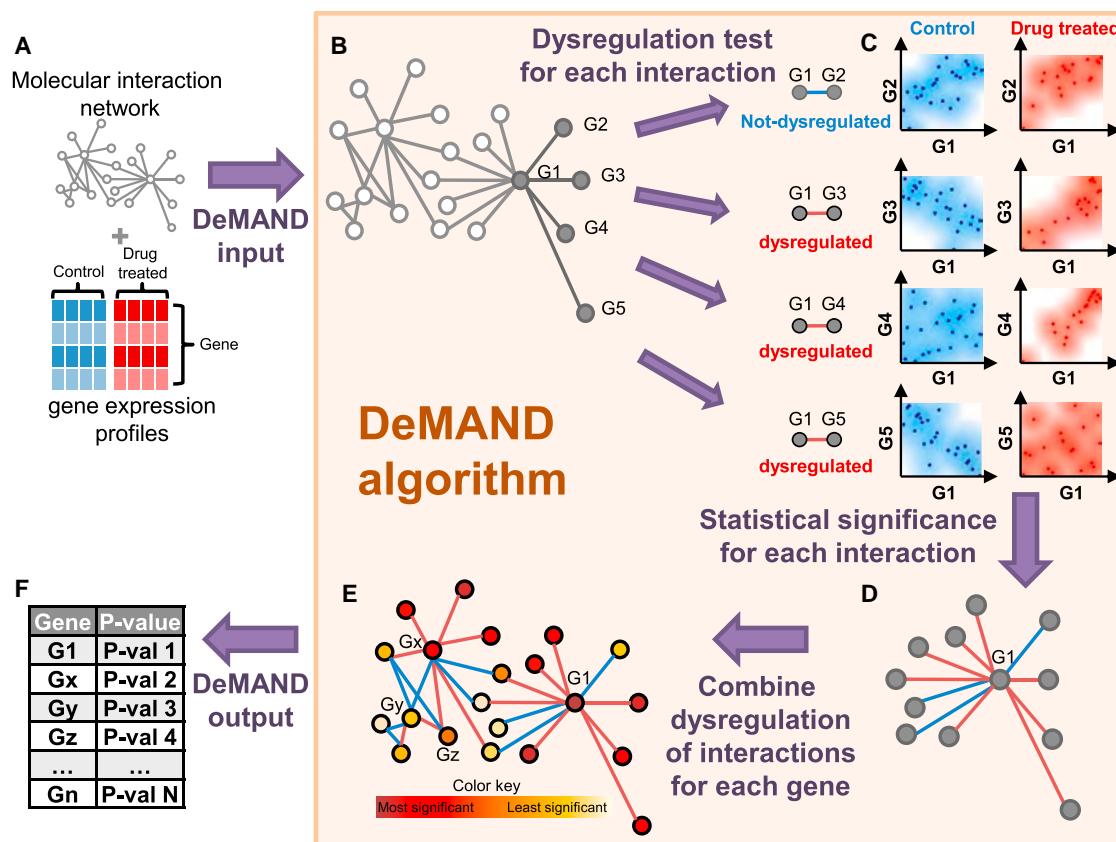
The mechanism of action of a compound (MoA) is defined as the set of target and effector proteins necessary to produce its phar-

macological effect in a specific cellular context. Its elucidation is critical in assessing both on-target compound activity as well as off-target effects associated with potential toxicity, thus providing critical insight into the two major challenges of drug development (Scannell et al., 2012). Since most compounds in clinical trials fail due to toxicity or lack of efficacy (Wehling, 2009), any improvements in systematic MoA characterization may increase the yield of pharmacological discovery pipelines.

MoA characterization remains a major challenge that is only partially addressed by experimental and computational strategies. Most experimental approaches rely on direct binding assays, such as affinity purification (Hirota et al., 2012; Ito et al., 2010) or affinity chromatography (Aebersold and Mann, 2003). These methods are labor-intensive and generally limited to the identification of high-affinity binding targets, rather than of all proteins responsible for compound activity. They may thus miss important indirect effectors, as well as lower-affinity targets responsible for both desirable and undesirable pharmacological properties. For instance, compounds can be effectively screened against all protein kinases, while missing equally relevant targets, as shown by the recent reclassification of the MET inhibitor tivantinib as a microtubule inhibitor (Basilico et al., 2013). In addition, these assays work in vitro and may miss effects from tissue-specific interactions and signals.

Chemo-informatics methods have also been developed. Yet, these are mostly designed to assess compound MoA similarity or specific compound/target interactions (Keiser et al., 2009; Lomenick et al., 2009; Miller, 2002), by leveraging the integration of structural and genomic information (Yamanishi et al., 2008), text-mining algorithms (Li et al., 2009), or machine learning methods for data-mining (Hansen et al., 2009). As such, they rely on detailed three-dimensional structures of both compound and target proteins or on prior literature or database knowledge of





**Figure 1. Schematics of the DeMAND Algorithm**

(A) DeMAND requires both a regulatory network and a set of gene expression profiles from compound perturbed and control samples, as an input.

(B) DeMAND evaluates the dysregulation of each interaction in the regulatory network.

(C) To evaluate interaction dysregulation co-expression scatter plots for the two interacting genes are smoothed using a Gaussian Kernel method to generate an interaction probability density. The probability density difference before and after compound perturbation is evaluated using the KL-divergence. The top example illustrates no change in probability density (i.e., no dysregulation). The other three examples illustrate various examples of compound dysregulation, including correlation inversion, gain, and loss (top to bottom, respectively).

(D) The statistical significance of the KL-divergence is assessed by gene pair shuffling.

(E) The global dysregulation of each gene is determined by integrating the p values of all its network interactions, while accounting for their dependencies.

(F) DeMAND produces a list of all network genes and the statistical significance of their dysregulation.

See also [Figure S1](#), [Table S7](#), and [Supplemental Experimental Procedures](#).

related MoA compounds. More recently, assembly of large reference compendia by systematic gene expression profiles (GEP) analysis of cells following compound perturbations has spurred development of MoA analysis methods ([Ganter et al., 2005](#); [Lamb et al., 2006](#); [Wolpaw et al., 2011](#)). In general, however, these methods are mostly comparative in nature and thus poorly suited to de novo MoA elucidation or to recognize subtle MoA differences that may induce unexpected toxicity. Network-based methods have also been recently proposed ([Bansal et al., 2006](#); [di Bernardo et al., 2005](#); [Gardner et al., 2003](#); [Mani et al., 2008](#)). Rather than focusing on individual genes, these methods perform integrative analyses over interacting gene subsets or pathways. Yet, these methods either rely on prior knowledge of the pathways that mediate compound activity, making them unsuitable for genome-wide analyses, or require very large samples sizes ( $n > 100$ ), thus making them impractical even for small compound libraries. As a result, there is still a pressing

need for experimentally validated methodologies for the de novo prediction of genome-wide compound targets and effectors or to mechanistically elucidate MoA proteins associated with differential activity or toxicity.

To address this challenge, we introduce detecting mechanism of action by network dysregulation (DeMAND), a hybrid computational and experimental approach for MoA analysis. DeMAND elucidates compound MoA by interrogating tissue-specific regulatory networks using small-size GEP datasets ( $n \geq 6$  samples) representing in vitro or in vivo, compound perturbations ([Figure 1](#)). Using GEPs from human lymphoma cells perturbed with libraries of 14 and 92 compounds, respectively, we systematically assessed the algorithm's ability to infer known compound targets (from public databases) and then experimentally validated novel compound activity effector and modulator predictions (hereafter *MoA-proteins*). DeMAND identified established MoA proteins for >70% of these compounds, as well as novel

proteins that were experimentally validated, such as RPS3A, VHL, and CCNB1 for the mitotic spindle inhibitor vincristine and JAK2 for mitomycin C. We also tested the algorithm's ability to assess compound MoA similarity. More than 50% of top predicted compound pairs were confirmed by literature and database analysis or by experimental validation. For instance DeMAND identified altretamine, an unknown MoA compound, as a novel GPX4 inhibitor based on predicted MoA similar to sulfasalazine, a system  $x_c^-$  cystine-glutamate antiporter-mediated GPX4 inhibitor (Yang et al., 2014).

DeMAND is freely available to the research community, both as a Bioconductor package (Gentleman et al., 2004) and as a web-based geWorkbench module (Floratos et al., 2010).

## RESULTS

### Overview of DeMAND Algorithm

Consider the regulon of a gene  $G$ , i.e., all its interactions ( $G \leftrightarrow G_i$ ) with other genes  $G_i$ , including transcriptional, signaling, and protein-complex interactions. If  $G$  belongs to a compound's MoA, then it is reasonable to assume that its regulon gene interactions will be dysregulated by the compound. This can be optimally assessed by measuring changes in the joint gene expression probability density  $p(G, G_i)$ , for each of its regulon genes. Such analysis can capture direct effects on gene expression and more importantly modulation of the interacting partner's expression via either direct or indirect regulatory mechanisms (e.g., feedback loops). Consider for instance a transcription factor regulating a set of targets. A targeted inhibitor will significantly alter the joint expression probabilities  $p(G, G_i)$ , as the expression of the targets will be dysregulated even though the expression of  $G$  is not generally affected (Figures 1 and S1; Experimental Procedures).

The Kullback-Leibler divergence (KLD) (Kullback and Leibler, 1951) provides an ideal metric to quantitatively assess probability density changes in one or more variables. From information theory, the KLD is easily interpreted as the loss of information resulting from using a probability density as a surrogate for another. For each regulon interaction ( $G \leftrightarrow G_i$ ), we estimate the KLD of each probability density  $p(G, G_i)$ , before and after compound perturbation. Their statistical significance is then integrated, thus producing a global statistical assessment of the compound-mediated dysregulation of  $G$ . To avoid overestimating such integrative significance, due to interaction dependencies, we use a modification of Brown's method that compensates for the integration of correlated evidence (Brown, 1975). All genes are then ranked based on their global KLD statistics.

To identify the regulon of each gene-product of interest, we used a set of established network reverse engineering algorithms (see Experimental Procedures). However, DeMAND is agnostic to the specific approach and can use networks generated by any alternative means, both computational and experimental.

### DeMAND Predictions Are Enriched in Established High-Affinity Binding Targets

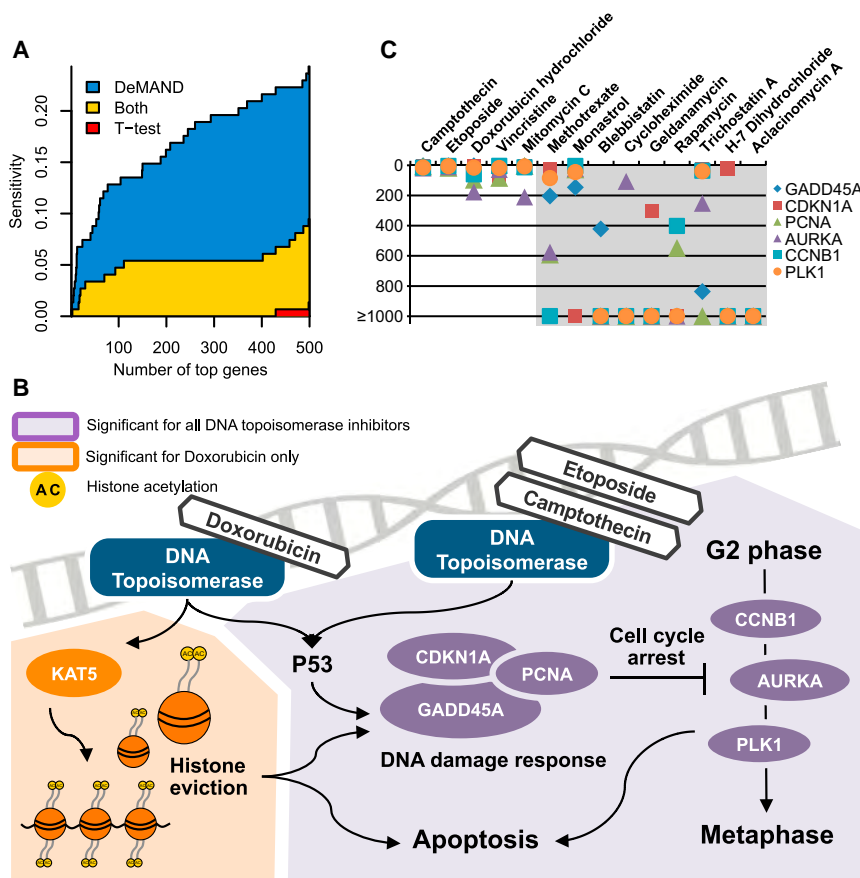
We first evaluated the accuracy of DeMAND-inferred MoA genes for 14 selected compounds, using the perturbation dataset

(DP14) from the DREAM/NCI compound synergy challenge (Bansal et al., 2014). This includes 276 GEPs of diffuse large B cell lymphoma cells (OCI-LY3), following perturbation with 14 distinct compounds, of which 11 have established primary targets (Supplemental Experimental Procedures; Table S1), and DMSO as control media, at two concentrations and three time points, in triplicate. The network for these analyses was produced as described in Lefebvre et al. (2010), using a published dataset of 226 U133p2 GEPs representing both normal and tumor-related human B cells (Basso et al., 2010) (Supplemental Experimental Procedures). Although DeMAND is designed to predict both compound targets (i.e., high-affinity binding proteins) and effectors/modulators, its performance can only be systematically benchmarked against the former, because gold-standard datasets to systematically assess the latter are not yet available.

DeMAND identified the known primary targets of 7 of the 11 tested compounds as statistically significant, at a 10% false discovery rate (FDR) (Figure S2A; Table S2; Experimental Procedures). Since the GEPs used in this analysis were obtained at multiple time points (6 hr, 12 hr, and 24 hr), we further assessed whether individual time points may be more informative. Intriguingly, several targets were best predicted at specific time points (Figure S2B), consistent with expectations that compound activity may be mediated over different timescales. Yet, integration over all time points performed as well or better than the optimal time point for all but two compounds (monastrol and doxorubicin). For these, the direct target was significant only when specific time point GEPs were used. In total, targets for 9 of the 11 compounds could be elucidated either from multi-point or single time point analysis. Replacing interaction dysregulation with the differential expression of neighbors reduces the performance (see Supplemental Experimental Procedures).

Differential expression analysis has been proposed to elucidate compound substrates (Ganter et al., 2005; Lamb et al., 2006; Wolpaw et al., 2011). We thus compared DeMAND's performance with differential expression analysis, by t test statistics. DeMAND systematically outperformed t test analysis, except for blebbistatin for which neither method identified myosin II as statistically significant (Figure S2A). Indeed, DeMAND had an almost 5-fold better sensitivity in the top 100 predictions, compared to t test analysis (15% versus 3%), which was highly statistically significant ( $p = 5 \times 10^{-4}$  and  $p = 0.06$  by  $\chi^2$  test, respectively) (see Supplemental Experimental Procedures and Figure 2A). Furthermore, any targets that were significant by t test analysis were also significant by DeMAND analysis, but not the opposite. Considering the full area under the receiver operator characteristic (ROC) curve (AUC), DeMAND also consistently outperformed the t test, AUC = 0.70 ( $p = 2 \times 10^{-16}$  by Fisher integration of individual Mann-Whitney p values for each compound) versus AUC = 0.60 ( $p = 3.5 \times 10^{-7}$ ), respectively, reflecting higher overall sensitivity and specificity (Figure S2C).

To assess DeMAND's performance on MoA proteins other than high-affinity targets, we focused on two of the four compounds, whose direct targets were missed, including camptothecin (a TOP1 inhibitor) and doxorubicin (a TOP2A inhibitor), which severely disrupt DNA repair and mitosis. DeMAND identified growth arrest and DNA damage-inducible gene 45A



**Figure 2. DP14 Dataset Analysis**

(A) The average sensitivity (true-positive rate) for identifying known direct targets in all DP14 compounds, as a function of the number of top selected predictions, using either DeMAND (blue + yellow areas) or t test analysis (red + yellow areas). DeMAND consistently outperforms t test. For instance, DeMAND achieves ~15% sensitivity across the top 100 predictions, compared to only 3% for t test. Furthermore, virtually all targets that are significant by t test analysis are also significant by DeMAND analysis (no red area for up to 400 genes). In contrast, DeMAND identifies many targets that are missed by t test (large blue area).

(B) Comparative schematics of established MoA genes for camptothecin, doxorubicin, and etoposide. Doxorubicin-specific DeMAND inferred MoA genes are shown with an orange background, while common inferred MoA genes for all compounds are shown with a purple background. The common genes include the core DNA-damage repair machinery (GADD45A, PCNA, and CDKN1A) and cell-cycle arrest genes (CCNB1, AURKA, PLK1). Doxorubicin's specific MoA includes KAT5, a mediator of histone eviction.

(C) Rank of DNA damage response genes across all DP14 compounds. DeMAND predicts GADD45A, the canonical DNA-damage-inducible gene and its well-known partners CDKN1A, PCNA, CCNB1, AURKA, and PLK1 among the most significant genes only for the five DNA damaging agents (i.e., camptothecin, doxorubicin, etoposide, mitomycin C, and vincristine).

See also Figure S2 and Tables S1 and S2.

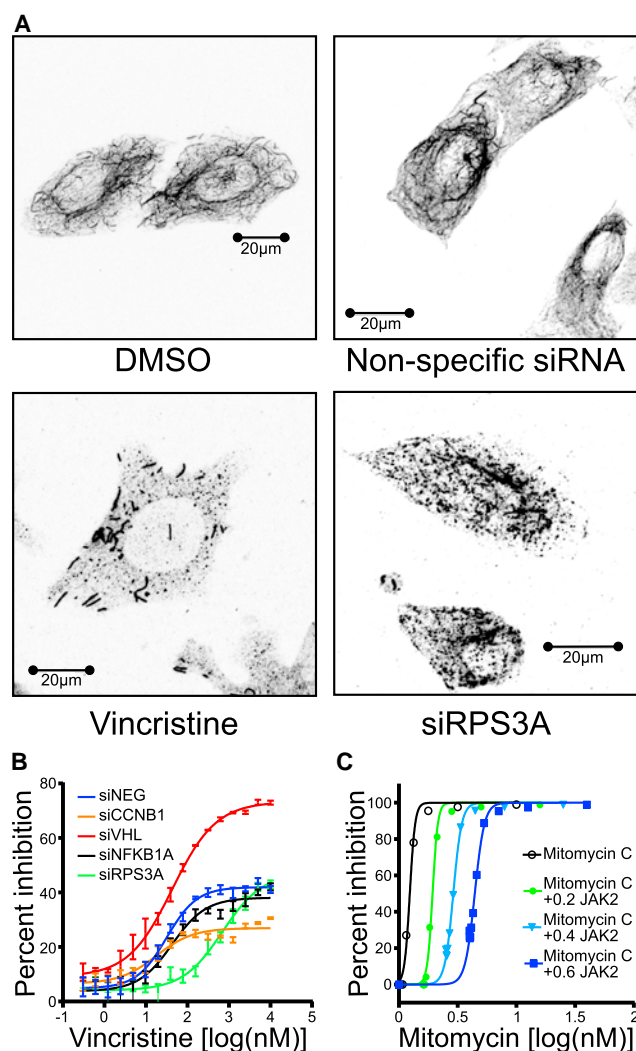
(GADD45A), cyclin-dependent kinase inhibitor 1A (CDKN1A), proliferating cell nuclear antigen (PCNA), Aurora Kinase A (AURKA), polo-like kinase 1 (PLK1), and cyclin B1 (CCNB1) among the most statistically significant genes for both compounds (mostly in the top 20), which are known key downstream effectors of TOP1 and TOP2A inhibition (Figure 2B). DeMAND therefore identifies key MoA proteins for both these compounds. More specifically, GADD45A, an established DNA damage response effector (Goldwasser et al., 1996), acts by forming protein complexes with CDKN1A and PCNA, a processivity factor of DNA polymerase delta required for high-fidelity DNA replication and excision repair (Smith et al., 1994). In turn, if DNA damage is detected, CDKN1A, PCNA, and GADD45A regulate the activity of CCNB1 (a critical effector of the G2/M cell-cycle checkpoint) (Zhan et al., 1999), PLK1, and AURKA (a mitosis regulator) either at the RNA or protein level (Shao et al., 2006). Of these six genes, only GADD45A and CDKN1A were differentially expressed, albeit at a much lower rank.

#### DeMAND Identifies Specific Differences in Compounds with Similar MoA

Detailed assessment highlighted key differences and commonalities in DeMAND-inferred MoA of compounds with similar targets, which were undetectable by t test analysis. For instance, camptothecin (TOP1), doxorubicin (TOP2A), and etoposide (TOP2A) are all topoisomerase (TOP) inhibitors, which induce

single or double strand breaks following covalent trapping of the TOP-DNA cleavable complex (Gilbert and Hemann, 2010). Consistently, DeMAND identified a significant common footprint in their inferred MoA, as shown in the previous section. However, it also identified highly specific effectors, such as KAT5/TIP60 for doxorubicin (ranked fourth), suggesting potentially relevant MoA differences (Figure 2B). Indeed, contrary to etoposide and camptothecin, doxorubicin is also a strong DNA intercalator, inducing KAT5-dependent histone acetylation and release from open chromatin (histone eviction) (Choi et al., 2009; Ikura et al., 2000), leading to cell-cycle arrest (Pang et al., 2013). Similarly, DeMAND identified *SIK1* as a doxorubicin-specific effector (ranked 36<sup>th</sup>), which is required for cardiac progenitor cell maintenance (CPCs) (Romito et al., 2010), thus pinpointing the compound's key adverse event, i.e., cardiomyopathy followed by congestive heart failure (Zhang et al., 2012b). Both KAT5 and SIK1 were completely missed by t test analysis.

Finally, DeMAND successfully stratified compounds based on MoA gene overlap, further emphasizing its specificity. For instance, for all DNA damaging agents, including camptothecin, doxorubicin, etoposide, mitomycin C, and vincristine, DeMAND predicted GADD45A, the canonical DNA-damage-inducible gene, and its well-known interactors (CDKN1, CCNB1, PCNA, and AURKA) among the most significant genes (Figure 2C). Yet, these genes were not significant for other compounds (Figure 2C), confirming the algorithm's specificity.



**Figure 3. Validation of Novel Effectors of Vincristine and Mitomycin C**

(A) Immunohistochemistry-based imaging of microtubule networks in cells treated with DMSO, vincristine, non-target siRNA, and siRNA-targeting RPS3A. Non-target siRNA is indistinguishable from DMSO controls. Both vincristine and siRPS3A significantly alter the microtubule network in U-2-OS cells (4 nM of vincristine for 24 hr). Images represent red channel intensity.

(B) Vincristine dose response curves in U-2-OS following transfection with non-target siRNA (blue) or siRNA-targeting CCNB1 (orange), VHL (red), NFKB1A (black), and RPS3A (green). RPS3A and CCNB1 silencing reduces cell sensitivity to vincristine, while VHL silencing increases sensitivity by 2-fold. The error bars indicate the SD from the mean using three replicates. See also Supplemental Experimental Procedures and Table S3.

(C) Mitomycin C dose-response curves in OCI-LY3 normalized to DMSO treatment (black) or following treatment with TG101348 (a JAK2 inhibitor), at 0.2  $\mu$ M (green), 0.4  $\mu$ M (cyan), and 0.6  $\mu$ M (blue). JAK2 inhibition induces loss of sensitivity to mitomycin C.

### Validation of Novel Effectors and Modulators of Compound Activity

To assess whether DeMAND can identify novel compound effectors and modulators, we validated novel predictions for vincristine and mitomycin C, an inhibitor of microtubule formation in mitotic

spindle and an antineoplastic antibiotic, respectively. DeMAND successfully identified the known high-affinity target of vincristine (TUBB), as well as CCNB1, VHL, RPS3A, and NFKB1A, in the top five predictions. While RPS3A and VHL are known to affect mitotic spindle assembly (Jang et al., 2012; Thoma et al., 2009), and CCNB1 is a microtubule activity marker, their function in mediating/modulating vincristine's activity is unknown.

Probing the microtubule network with an anti-tubulin antibody, following small interfering RNA (siRNA)-mediated silencing of these genes, confirmed that loss of RPS3A (but not of VHL, CCNB1, or NFKB1A) disrupts microtubules in adherent U-2-OS cells (Figure 3A). To further validate the role of these genes in mediating vincristine's activity, we performed dose-response curve assays in U-2-OS cells, following silencing of each gene (Supplemental Experimental Procedures; Table S3). These assays confirmed that all of these genes, except for NFKB1A, are key vincristine activity effectors and mediators. Specifically, VHL silencing increased vincristine sensitivity by more than 2-fold (Figure 3B), while RPS3A and CCNB1 silencing had the opposite effect. Thus, four out of five of the top DeMAND-inferred genes were confirmed vincristine activity modulators, including its primary target (TUBB), suggesting that, for some compounds, false positive rates may be as low as 20%. None of these genes were significant by t test analysis.

DeMAND also inferred the JAK2 kinase as an exclusive mitomycin C MoA protein (i.e., JAK2 was not significant by DeMAND analysis for any other compound). This is of potential importance since constitutive activity of JAK2 causes chemo-resistance in lymphocytes (Gupta et al., 2012), while constitutive JAK2 activity may also affect DNA damage, repair, and recombination outcome (Hoser et al., 2003). Confirming the prediction, dose-response curves for mitomycin C, following treatment with varying amounts of TG101348 (a JAK2 inhibitor), revealed highly significant, dose-dependent antagonism between JAK2 inhibition and mitomycin C activity (Figure 3C; Experimental Procedures).

Finally, we analyzed DeMAND-inferred results for rapamycin. While DeMAND could not predict the highest-affinity targets, MTOR and FKBP1A, many genes downstream of MTOR pathways (Hsieh et al., 2012) were highly enriched in the top DeMAND-inferred genes (Figure S2E), including many ribosomal genes. The only other compound with significant ribosomal gene enrichment was cycloheximide, a known ribosomal activity inhibitor, thus further highlighting the algorithm's specificity.

### Algorithm Robustness and Requirements

We then benchmarked DeMAND's performance as a function of network accuracy and size, as well as of the number of samples in the perturbation dataset. First, we compared the results obtained using an independent B cell gene regulatory network, reconstructed from a distinct dataset of 254 Affymetrix U95av2 GEPs (see Experimental Procedures). We tested the enrichment of statistically significant DeMAND-inferred genes ( $FDR \leq 0.1$ ), using the U95av2 network, against those inferred using the U133p2 network, by Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005). The analysis confirmed that DeMAND predictions were almost identical, independent of network model ( $p < 1 \times 10^{-9}$  by GSEA; Figure S3A). Furthermore, predictions were virtually unaffected when up to 60% of the network



**Table 1. Thirteen Compound Perturbation Datasets from the GEO Database**

Compound	Cellular Context	GEO ID
Zoledronate	metastatic breast cancer cell lines (MDA-MB-231)	GSE33552
Valproic acid	chronic lymphocytic leukemia (patient-derived B cells)	GSE14973
Genistein	breast cancer cell lines (MCF-7)	GSE9936
S-Equol	breast cancer cell lines (MCF-7)	GSE9936
Estradiol	breast cancer cell lines (MCF-7)	GSE9936
Rituximab	B cell non-Hodgkin's lymphoma cell lines (K422)	GSE7292
Thapsigargin	lytic-permissive lymphoblastoid cell lines	GSE31447
Fluvastatin	metastatic breast cancer cell lines (MDA-MB-231)	GSE33552
MALT1 inhibitor	diffuse large B cell lymphoma (patient-derived B cells)	GSE40003
Docetaxel	breast cancer cell lines (MCF-7)	GSE5149
$\gamma$ -Secretase inhibitor	MCL cell lines	GSE34602
Triptolide	breast cancer cell lines (MCF-7)	GSE28662
Actinomycin D	breast cancer cell lines (MCF-7)	GSE28662

See also [Figure S4](#) and [Table S4](#).

interactions were randomly removed ([Figure S3B](#); [Experimental Procedures](#)). Similarly, predictions were virtually identical, as long as six or more GEPs representative of compound perturbation were used ([Figure S3C](#); [Supplemental Experimental Procedures](#)). Taken together, these data suggest that DeMAND is highly robust to network noise and especially to false negative interactions and that it can be applied to datasets with as few as six treatment and six untreated controls GEPs.

We then selected 13 datasets representing compound perturbations (GEO13) from the GEO database ([Table 1](#); [Table S4](#)). Only compounds with established targets with at least six treatment/control GEPs were selected, including seven human breast cancer and six human B cell lymphoma datasets. Confirming results on the DP14 dataset, DeMAND inferred known direct targets for 62% of these compound perturbations ( $\text{FDR} \leq 0.1$ ; [Figure S4A](#)), while still significantly outperforming t test-based methods ( $\text{AUC} = 0.82$  versus  $0.74$ , respectively,  $p$  value =  $2.2 \times 10^{-16}$  versus  $p$  value =  $5.9 \times 10^{-8}$ , respectively, by Fisher integration of individual Mann-Whitney  $p$  values for each compound) ([Figure S4B](#)). Among top-predicted MoA proteins, DeMAND again achieved  $\sim 5$ -fold better performance than t test ([Figure S4C](#)).

### DeMAND-Inferred MoA Stratifies Pharmacological Effect

We then assessed whether DeMAND-inferred MoA overlap was predictive of pharmacological compound similarity. We first computed the significance of MoA overlap for each DP14 compound pair ( $\text{FDR} \leq 0.1$  by Fisher's exact test [FET]) ([Figure 4A](#); [Table S5](#); [Experimental Procedures](#)). Among all 91 possible

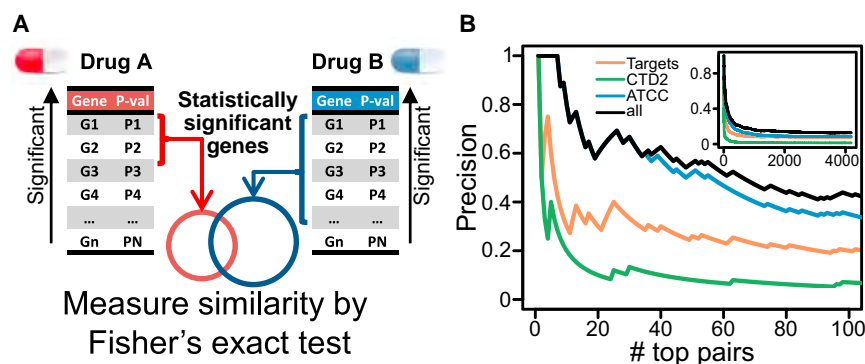
compound pairs, the six most similar ones included only topoisomerase inhibitors and other DNA-damaging agents (etoposide, doxorubicin, camptothecin, and mitomycin C). Thus, DeMAND successfully assessed high compound MoA similarity between topoisomerase inhibitors and other DNA-damaging agents even though it could not identify TOP1 or TOP2A among the inferred MoA genes, suggesting that key effector proteins may be as informative as direct targets in terms of compound similarity.

To further evaluate this hypothesis, we applied the method to a much larger compound perturbation dataset (DP92), representing GEPs from three B cell lymphoma cell lines (OCI-LY3, OCI-LY7, and U-2932), following perturbation with 92 unique FDA-approved, late-stage experimental, and tool compounds ([Table S6](#); [Supplemental Experimental Procedures](#)). Since only three GEPs per compound and cell line are available in this dataset, we used it only for compound-pair similarity assessment (see [Experimental Procedures](#)).

DeMAND performance was objectively evaluated by comparison with three independent data sources: (1) compounds sharing established targets, (2) compounds sharing therapeutic and chemical characteristics, according to the Anatomical Therapeutic Chemical (ATC) classification system, and (3) compounds with correlated drug-response profiles, as assessed by the Cancer Target Discovery and Development (CTD2) consortium ([Basu et al., 2013](#)) (see [Supplemental Experimental Procedures](#)). The latter dataset recapitulates dose-response curve vectors representing 338 unique compounds profiled against 257 distinct cancer lines. We evaluated the fraction of validated similar pairs (precision), based on each of the three evidence datasets, as a function of the number of significant pairs (precision curves, [Figure 4B](#)). DeMAND-inferred pairs were highly enriched in pairs from three evidence datasets, as assessed by each of the evidences individually (i.e.,  $p$  value =  $2 \times 10^{-8}$ ,  $1.4 \times 10^{-5}$ , and  $9 \times 10^{-4}$ , by GSEA, for pairs sharing the same ATC class, common established targets, and high dose-response vector correlation in the CTD2 dataset, respectively; [Figure S5A](#)), and also when taken together (GSEA  $p$  value =  $7.6 \times 10^{-7}$ ). For instance, 8 of the top 10 and 43 of the top 100 DeMAND-inferred pairs were validated by at least one of the three datasets ( $p = 2.2 \times 10^{-16}$  by FET).

DeMAND outperformed predictions using similarity obtained by overlapping statistically significant differential expressed genes (e.g., by t test statistics) by consistently achieving higher sensitivity at any precision value ([Figure S5B](#)). DeMAND also outperformed another state of the art method, (MANTRA) ([Iorio et al., 2010](#)), which uses mutual gene set enrichment analysis ([Subramanian et al., 2005](#)) to compute similarity, again by achieving higher sensitivity at almost any desired precision value ([Figure S5B](#)).

Finally, we evaluated the correlation between compound-pair similarity as predicted by each method and their CTD2-based similarity. DeMAND prediction achieved significant Spearman correlation ( $\rho = 0.59$ ,  $p$  value =  $7.8 \times 10^{-5}$ ; [Figure S5C](#)), while both the t test and MANTRA methods did not achieve statistically significant correlation ([Figures S5D](#) and [S5E](#)). Thus, DeMAND could predict compounds with similar pharmacological effect and activity profile using only the GEP following their treatment in a single cell line.



**Figure 4. Compound Similarity Inference**

(A) Compound similarity is assessed based on the statistical significance (by FET) of the overlap of their DeMAND-inferred MoA proteins.

(B) DeMAND-inferred compound similarity in the DP92 dataset is assessed by (a) the overlap of known direct targets between two compounds (orange), (b) compound sensitivity profile similarity based on CTD2 data (green), (c) overlap in compound classification, according to the Anatomical Therapeutic Chemical (ATC) Classification (blue), or (d) any of the above evidences (black). See also Figure S5 and Tables S5 and S6.

### DeMAND Identifies GPX4 as a Novel MoA Effector for Altretramine

We identified altretamine and sulfasalazine as the compound pair with the highest DeMAND-inferred MoA similarity ( $p$  value =  $9.91 \times 10^{-81}$ ), among all pairs where the MoA of at least one compound was unknown. Altretramine is an FDA-approved anti-neoplastic drug with no established targets or effectors. Instead, sulfasalazine is an inhibitor of system  $x_c^-$ , the cystine-glutamate antiporter (Dixon et al., 2014), required for the biosynthesis of glutathione (GSH). Thus sulfasalazine inactivates enzymes that rely on reduced glutathione (GSH) as a cofactor, including glutathione peroxidase 4 (GPX4) (Dixon et al., 2012; Yang et al., 2014), leading to toxic accumulation of lipid reactive oxygen species (ROS).

We thus tested whether altretamine may also modulate the system  $x_c^-$ -GPX4 pathway. U-2932 cells were treated with altretamine and their GSH levels were assessed using Ellman's reagent (Figure 5A; Supplemental Experimental Procedures). Sulfasalazine was used as a positive control for GSH depletion in U-2932 cells, confirming depletion of GSH levels following compound treatment. In contrast, altretamine did not deplete GSH levels, even after doubling its  $IC_{50}$  at 24 hr concentration, suggesting that the compound may target mechanisms downstream of GSH in this pathway. We thus treated U-2932 cells with altretamine and prepared cell lysates for a liquid chromatography-mass spectrometry (LC-MS)-based GPX4 assay. Phosphatidylcholine hydroperoxide (PC-OOH), a specific substrate for GPX4 (Brigelius-Flohé and Maiorino, 2013), was added to cell lysates and PC-OOH to PC-OH reduction was assessed by the mass chromatogram of the  $[PC-OOH + H^+]$  ion ( $m/z = 790.5$ ). As shown in Figure 5B, lysates of untreated cells reduced PC-OOH levels completely, leaving no residual signal for the  $[PC-OOH + H^+]$  ion ( $m/z = 790.5$ ). In sharp contrast, lysates from altretamine-treated cells displayed a significant  $[PC-OOH + H^+]$  signal, indicating that abrogation of PC-OOH reduction was mediated by GPX4 inhibition (Experimental Procedures). Indeed, since GPX4 is the only enzyme capable of reducing lipid hydroperoxides (Yang et al., 2014), GPX4 inhibition is necessary to increase lipid-ROS levels (Thomas et al., 1990). As expected, both sulfasalazine and altretamine were confirmed to induce lipid-ROS accumulation in U-2932 cells, as assessed by BODIPY-C11 staining and flow cytometry (Figure 5C; Experimental Procedures). Thus, DeMAND correctly

predicted the unexpected mechanistic similarity between the MoA of two previously unrelated drugs (Figure 5D), showing altretamine as a new GPX4 inhibitor and suggesting a potential mechanism for its antineoplastic activity.

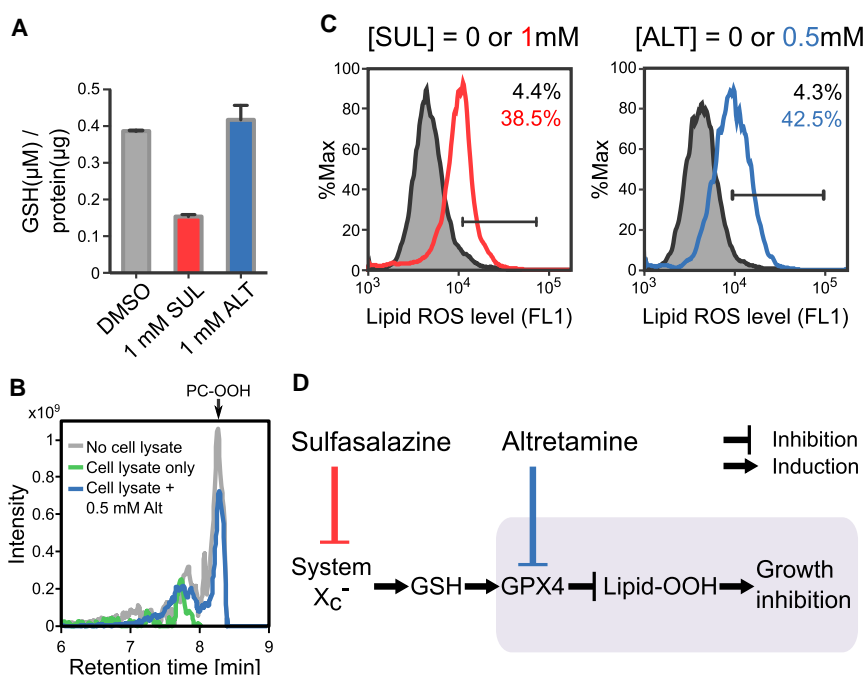
### DISCUSSION

DeMAND elucidates compound MoA by assessing compound-mediated dysregulation of gene-gene interactions on a genome-wide basis, from gene expression profiles of compound perturbations. DeMAND reliably identifies compound targets, effectors, and activity modulators, allowing effective assessment of compound MoA and MoA similarity. Indeed, DeMAND identified known and novel MoA genes for vincristine, mitomycin C, and altretamine that were experimentally validated. DeMAND also elucidated a novel MoA for altretamine, confirming its predicted similarity to sulfasalazine.

DeMAND was shown to be highly robust to network and sample variability. More importantly, unlike previous methods (di Bernardo et al., 2005; Mani et al., 2008), DeMAND can reliably predict compound MoA using as few as six control and six perturbation samples. This allows unprecedented applicability of the methods to elucidate MoA for novel developmental compounds within specific cellular contexts of interest, including in vivo.

DeMAND leverages integration of GEPs obtained at multiple time points and at multiple compound concentrations, thus simplifying experimental design when the precise concentration or time points at which the MoA may be revealed is unknown. Indeed, absent prior knowledge, compound MoA was optimally revealed by integrating multi-time-point compound perturbations for all but two of the tested compounds (Figure S2B).

DeMAND predictions are highly specific, allowing classification of compounds into groups of similar function and identification of pathways that are relevant to compound MoA. For instance, for DNA-damaging compounds (camptothecin, doxorubicin, etoposide, vincristine, and mitomycin C), DeMAND correctly predicted several of the hallmark genes involved in DNA-damage-induced response. The specificity was evidenced by the fact that relevant MoA proteins were inferred only for DNA-damage inducing compounds and not for any other compound (including compounds exhibiting significant polypharmacology like H-7 dihydrochloride or cycloheximide). In other examples,



**Figure 5. DeMAND Identifies the MoA of Al-tretamine**

(A) GSH concentration following treatment of cells by negative control (DMSO, gray), sulfasalazine as a positive control (red), and al-tretamine (blue) show that sulfasalazine reduces active GSH levels compared to control, while al-tretamine results in active GSH levels indistinguishable from the control. (B) The level of a GPX4-specific substrate (PC-OOH) is measured by mass spectrometry (a) without cell lysate (gray), (b) with untreated cell lysate (green), and (c) with cell lysate from al-tretamine-treated cells (blue). PC-OOH levels in al-tretamine-treated cells are similar to no-lysate and markedly different from untreated lysate, indicating that al-tretamine inhibits GPX4 activity. (C) Lipid reactive oxidative species (ROS) levels were measured by flow cytometry using DMSO-treated cells (black curve, as control) and compound-treated cells (red curve). Both al-tretamine and sulfasalazine significantly increases lipid-ROS levels, confirming the predicted similarity in their functional effect. (D) Sulfasalazine is a known inhibitor of the System X<sub>c</sub><sup>-</sup> cystine/glutamate antiporter. Its downstream effect on Glutathione (GSH) and GPX4 leads to accumulation of lipid ROS. DeMAND predicted significant similarity between sulfasalazine and al-tretamine and GPX4 but not GSH as al-tretamine-specific MoA proteins, as experimentally confirmed panels (A–C).

high MoA specificity was shown for doxorubicin, where DeMAND identified KAT5, consistent with recent findings of KAT5-mediated histone eviction, as well as SIK1, a gene required for cardiac progenitor cells maintenance, providing a potential mechanistic link between doxorubicin and its known cardiac toxicity. Critically, SIK1 was also detected in the MoA of other DNA damaging agents, albeit at much lower rank/significance, suggesting that these compounds should also be monitored for cardiac toxicity. Taken together, these findings suggest that the algorithm is equally effective in predicting both direct targets and indirect compound effectors, thus helping elucidate both on-target pharmacology and off-target toxicity. Overall, DeMAND identified known MoA proteins for >70% of tested compound, while experimental validation suggests that false discovery rates (FDR) may be as low as 20%, although more extensive FDR estimate is impossible at this time because compound MoA in databases is largely incomplete, producing significant FDR overestimate. For instance, following experimental validation, FDR for vincristine went from 80%, as only TUBB was an established compound target/effector, to 20%.

DeMAND relies on the existence of high quality context-specific gene regulatory networks, which may represent a limitation for specific cellular contexts. However, given the abundance of data generated by large-scale projects such as the Cancer Genome Atlas (TCGA) and other related consortia, as well as the availability of increasingly accurate and comprehensive methods for context-specific network reverse engineering (Califano et al., 2012; Zhang et al., 2012a), this limitation is at best temporary. However, network availability does not guarantee identification of MoA proteins that are poorly represented. For instance, for blebbistatin (a myosin II inhibitor), using the

U95av2 network, DeMAND identified *PTK2B*, *GRB2*, and *FYN*, all of which are both direct regulators of myosin II phosphorylation and responders to myosin II perturbation (Sieg et al., 1998) (Figure S2D). Yet, due to lack of *GRB2* representation in the U133p2 network, this gene could not be inferred. It is also important to highlight that DeMAND analysis of the DP14 and DP92 datasets, using a high quality context-free network from the STRING database (Franceschini et al., 2013), was still able to identify ubiquitous targets and effectors (e.g., those involved in cell-cycle and DNA damage repair mechanisms) with high precision and sensitivity, but exhibited lower performance both in compound similarity analysis and in the identification of genes with context-specific function/expression. This suggests that non-context-specific networks can still be used for DeMAND analyses, albeit with an increase in false positive and negative predictions.

An important, albeit not critical, limitation of the current methodology is the lack of prediction of compound activity sign, i.e., whether a compound will induce increase or decrease in an inferred MoA protein activity. Conversely, the method cannot predict whether inhibiting an inferred MoA protein will likely either increase or decrease drug activity. Presently, the only way to resolve this question is by follow-up experimental assays. In addition, the need for at least six GEPs at multiple concentrations and time points is a potential limitation when assessing MoA for large compound libraries. Despite these limitations, however, DeMAND has proven highly effective in the de novo identification of context-specific targets and effectors for arbitrary compounds of interest, providing important insight into the prioritization of novel compounds for development, or into the repositioning of previously approved compounds.

## EXPERIMENTAL PROCEDURES

### Networks Used in the Analysis

We generated context-specific gene-regulatory networks with both protein-DNA and protein-protein interactions (Table S7). The analysis used both context-specific GEPs and context-independent information from multiple experimental and computational databases, which was integrated into a final interactome using Naive Bayes Classifiers (see Lefebvre et al. [2010] and Supplemental Experimental Procedures for detailed information). B cell- and breast cancer-specific networks as well as the STRING database can be downloaded from <http://wiki.c2b2.columbia.edu/califanolab/index.php/Software/DeMAND>.

### Evaluating Interaction Dysregulation

For each pair of interacting genes in the network, we compute a two-dimensional probability density from their discrete rank-transformed expression in a given condition (treatment or control), by Gaussian kernel smoothing, using Silverman's approach (Silverman, 1986). The sum of the Gaussian probabilities densities from treatment samples, computed at each point of the discrete rank space, provides the perturbation probability distribution  $P$ , while that from control samples provides the control probability distribution  $Q$ . The distance between the two discrete probability distributions is evaluated using a symmetric form of the Kullback-Leibler divergence (KLD), obtained by averaging  $KLD(P|Q)$  and  $KLD(Q|P)$ .

KLD statistical significance is determined using a null distribution generated by  $10^5$  KLD values generated from random gene pairs (regardless of whether they share a network edge), providing individual edge dysregulation  $p$  values. These are integrated across all the interactions in a specific gene regulon, using the Fisher's method, and corrected using a modification of Brown's method for correction of  $p$  value dependence (Brown, 1975), using the covariance between the residuals from a linear fit to the common gene,  $\alpha$  (Figure S1A). A more detailed description of this method is available in the Supplemental Experimental Procedures.

### Determining Known Direct Targets of Compounds

Established targets for tested compounds were obtained from DrugBank (Wishart et al., 2008), MATADOR (Günther et al., 2008), and literature searches. For MATADOR, only genes annotated as "direct" or "direct-indirect" were considered as compound targets, while genes labeled as "indirect" were discarded. For a list of compound targets see Table S1.

### Assessing Drug Similarity

To evaluate compound similarity, we first selected statistically significant MoA genes ( $FDR \leq 0.1$ ) for each compound. We then computed the significance of their overlap by FET analysis. Many genes were not significant for any compounds, thus biasing this analysis. To reduce this effect, we removed these genes from the analysis. Notably this correction did not affect compound pair ranking but only their absolute similarity  $p$  values, by avoiding  $p$  value underestimation.

To compute similarity  $p$  values using the DP92 dataset we calculated  $p$  values for each of the three cell lines independently and used Fisher's method to combine them.

### Robustness Analysis

To evaluate the effect of network accuracy on DeMANDs' performance, we gradually removed interactions at random, in 10% increments and compared the overlap of significant perturbed and unperturbed MoA protein predictions by FET analysis. To evaluate the effect of sample size, we subsampled  $i$  samples ( $i = 3..18$ ) from the compound-treated and from the control samples and compared these results with the result obtained using all samples. Both analyses were performed independently on each of the 14 compounds in the DP14 dataset. See the Supplemental Experimental Procedures for additional information and Figure S3 for the results of the analysis.

### Cell Culture

Diffuse large B cell lymphoma (DLBCL) OCI-LY3 and OCI-LY7 cells were obtained from University Health Network (Toronto, Canada); the U-2932 DLBCL

cell line was purchased from the Leibniz-Institute DSMZ German Collection of Microorganisms and Cell Cultures; the U-2-OS osteosarcoma cell line was obtained from ATCC (ATCC HTB-96). OCI-LY3, OCI-LY7, U-2932 cells were cultured in Iscove's modified Dulbecco Medium (IMDM) supplemented with 10% fetal bovine serum at 37°C in a 5% CO<sub>2</sub> atmosphere. U-2-OS cells were cultured in McCoy's 5A medium, supplemented with 10% fetal bovine serum.

### Dose-Response Curves

The 92 compounds were selected based on primary activity screen of FDA-approved, late-stage experimental and tool compounds. OCI-LY3, OCI-LY7, and U-2932 cells were seeded in white tissue culture-treated 96-well plates, at a density of  $5 \times 10^4$  cells per well in 100  $\mu$ l total volume using the Janus automated liquid handling system (Perkin Elmer). After 12 hr of incubation at 37°C, plates were allowed to cool to room temperature, prior to compound addition via the Janus. Compounds were diluted in DMSO as a 7-point dilution curves in a stock plate, 1  $\mu$ l of these stock solutions were transferred into assay plates, in triplicate. These were subsequently placed on an orbital shaker for 5 min and then back in the incubator. At 24 hr, plates were removed from the incubator and equilibrated to room temperature before addition of 50  $\mu$ l of CellTiter-Glo Luminescent Cell Viability Assay (Promega) per well. Plates were shaken 5 min on an orbital shaker before data acquisition in an Envision (PerkinElmer) (0.5 s read time, enhanced luminescence). IC<sub>20</sub> values were assessed using a four parameter fit model (IDBS Activity Base).

### Compound Treatment for Gene Expression Profiling

Cells were seeded in tissue culture-treated 96-well plates at a density of  $5 \times 10^4$  cells per well using the Janus automated liquid handling system (Perkin Elmer). They were then treated with the 24 hr IC<sub>20</sub> of each compound (by DMSO dilution) for 6 hr, 12 hr, and 24 hr at 37°C, 5% CO<sub>2</sub> under humidified conditions. For each compound/condition combination one single data point was analyzed and 0.2% DMSO vehicle-treated samples were used as controls. Viability assay was run in parallel to monitor the compound effectiveness.

### Generation of Gene Expression Profiles

Total RNA was isolated with the RNeasy-96 Automated Kit (Ambion) on the Janus automated liquid handling system (Perkin Elmer), quantified by NanoDrop 6000 spectrophotometer and quality checked by Agilent Bioanalyzer. A total of 300 ng of each of the samples with an RNA integrity (RIN) value  $>7$  were converted to biotinylated cRNA with the Illumina TotalPre-96 RNA Amplification Kit (Ambion) using a standard T7-based amplification protocol and hybridized on the Human Genome U219 96-Array Plate (Affymetrix). Hybridization, washing, staining, and scanning of the array plates were performed on the GeneTitan Instrument (Affymetrix) according to manufacturer's protocols.

### GPX4 Enzymatic Activity Assay

GPX4 enzymatic activity assay was performed as described in Yang et al. (2014). Briefly,  $1 \times 10^6$  cells were re-suspended in the cell lysis buffer. Sonication was used to make cell lysates followed by centrifugation at 14,000 rpm for 10 min. Protein concentration of the cleared cell lysates was determined using a Bradford protein assay (Bio-Rad). Two hundred micrograms of cellular proteins was mixed with phosphatidyl choline hydroperoxide (PC-OOH), the GPX4-specific substrate, and reduced glutathione, a GPX4 cofactor. The mixture was incubated at 37°C for 30 min followed by lipid extraction using a chloroform:methanol (2:1) solution. The lipid extract was evaporated using a rotary evaporator and re-dissolved in 100% ethanol before injecting into LC-MS instrument for PC-OOH quantitation.

### Analysis of Lipid ROS Generation

U-2932 cells ( $2 \times 10^5$ ) were seeded in 6-well plates and incubated at 37°C for 16 hr. Cells were treated with test compounds for the indicated time, then harvested, pelleted, and washed once with PBS. For lipid ROS detections, cells were re-suspended with Hanks balanced salt solution (HBSS, Life Technologies) containing C11-BODIPY (581/591) (2  $\mu$ M) (Life Technologies) and incubated for 10 min at 37°C. Cells were then pelleted, re-suspended in 500  $\mu$ l HBSS, strained through 40- $\mu$ M cell strainer (BD Falcon), and analyzed using



BD Accuri C6 flow cytometer (BD Biosciences). C11-BODIPY signal was measured using FL1 channel. Experiments were done in biological triplicates, and a representative result was shown.

### Co-treatment with Mitomycin C and a JAK2 Inhibitor

The JAK2-selective inhibitor TG101348 (Wernig et al., 2008) and Mitomycin C were purchased from Selleckchem and Tocris Bioscience, respectively, and were dissolved in DMSO. OCI-LY3 cells were treated with the indicated compounds in 96-well plates and their growth was determined using the CellTiter-Glo Luminescent Cell Viability Assay (Promega). Typically, 30,000 OCI-LY3 cells per well in 200  $\mu$ l of growth medium were grown for 48 hr in the presence or absence (DMSO alone) of the desired compounds and then assayed with CellTiter Glo according to manufacturer's instructions.

### ACCESSION NUMBERS

The accession number for the DP92 expression data reported in this paper is GEO: GSE60408.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and seven tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.05.056>.

### AUTHOR CONTRIBUTIONS

M.B. and A.C. conceived the idea. J.H.W., Y.S., M.B., and A.C. developed the method. J.H.W., Y.S., W.S.Y., P.S., B.R.S., M.B., and A.C. wrote the manuscript. M.M., R.R., and C.K. generated the data. W.S.Y., P.S., and B.R.S. validated the predictions. G.L. generated interactomes. M.R.M. performed statistical analysis leading to method development. P.N., A.I., and P.S. performed literature-based analysis to establish the connection between DeMAND's predictions and the MoA of the compounds.

### ACKNOWLEDGMENTS

We thank Katia Basso for providing the U-2932 cell line for experimental validation, Wei Keat Lim for GEP normalization, and Beatrice Salvatori for helpful feedback on the manuscript. This work is supported in part by the CTD2 (5U01CA168426), LINC (1U01CA164184-02 and 3U01HL11566-02), and MAGNet (5U54CA121852-08) grants to A.C. B.R.S. is supported by the NIH (5R01CA097061, R01CA161061) and New York Stem Cell Science (C026715) and is an Early Career Scientist of the Howard Hughes Medical Institute.

Received: November 24, 2014

Revised: February 17, 2015

Accepted: May 28, 2015

Published: July 16, 2015

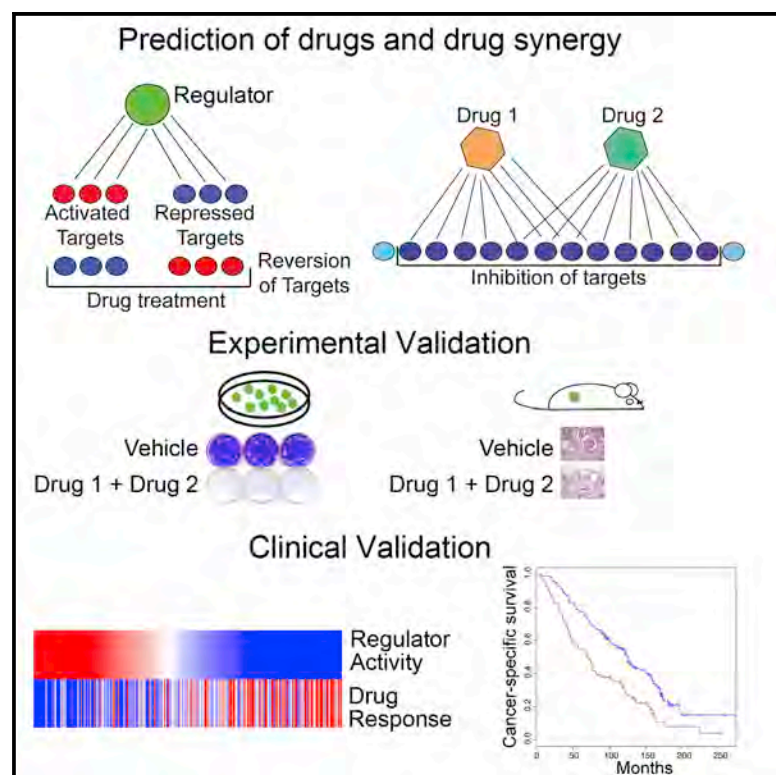
### REFERENCES

- Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* 422, 198–207.
- Bansal, M., Della Gatta, G., and di Bernardo, D. (2006). Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics* 22, 815–822.
- Bansal, M., Yang, J., Karan, C., Menden, M.P., Costello, J.C., Tang, H., Xiao, G., Li, Y., Allen, J., Zhong, R., et al.; NCI-DREAM Community; NCI-DREAM Community (2014). A community computational challenge to predict the activity of pairs of compounds. *Nat. Biotechnol.* 32, 1213–1222.
- Basilico, C., Pennacchietti, S., Vigna, E., Chiriaco, C., Arena, S., Bardelli, A., Valdembrì, D., Serini, G., and Michieli, P. (2013). Tivantinib (ARQ197) displays cytotoxic activity that is independent of its ability to bind MET. *Clin. Cancer Res.* 19, 2381–2392.
- Basso, K., Saito, M., Sumazin, P., Margolin, A.A., Wang, K., Lim, W.K., Kitagawa, Y., Schneider, C., Alvarez, M.J., Califano, A., and Dalla-Favera, R. (2010). Integrated biochemical and computational approach identifies BCL6 direct target genes controlling multiple pathways in normal germinal center B cells. *Blood* 115, 975–984.
- Basu, A., Bodycombe, N.E., Cheah, J.H., Price, E.V., Liu, K., Schaefer, G.I., Ebright, R.Y., Stewart, M.L., Ito, D., Wang, S., et al. (2013). An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* 154, 1151–1161.
- Brigelius-Flohé, R., and Maiorino, M. (2013). Glutathione peroxidases. *Biochim. Biophys. Acta* 1830, 3289–3303.
- Brown, M.B. (1975). Method for combining non-independent, one-sided tests of significance. *Biometrics* 31, 987–992.
- Califano, A., Butte, A.J., Friend, S., Ideker, T., and Schadt, E. (2012). Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nat. Genet.* 44, 841–847.
- Choi, J., Heo, K., and An, W. (2009). Cooperative action of TIP48 and TIP49 in H2A.Z exchange catalyzed by acetylation of nucleosomal H2A. *Nucleic Acids Res.* 37, 5993–6007.
- di Bernardo, D., Thompson, M.J., Gardner, T.S., Chobot, S.E., Eastwood, E.L., Wojtovich, A.P., Elliott, S.J., Schaus, S.E., and Collins, J.J. (2005). Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nat. Biotechnol.* 23, 377–383.
- Dixon, S.J., Lemberg, K.M., Lamprecht, M.R., Skouta, R., Zaitsev, E.M., Gleason, C.E., Patel, D.N., Bauer, A.J., Cantley, A.M., Yang, W.S., et al. (2012). Ferroptosis: an iron-dependent form of nonapoptotic cell death. *Cell* 149, 1060–1072.
- Dixon, S.J., Patel, D.N., Welsch, M., Skouta, R., Lee, E.D., Hayano, M., Thomas, A.G., Gleason, C.E., Tatonetti, N.P., Slusher, B.S., and Stockwell, B.R. (2014). Pharmacological inhibition of cystine-glutamate exchange induces endoplasmic reticulum stress and ferroptosis. *eLife* 3, e02523.
- Floratos, A., Smith, K., Ji, Z., Watkinson, J., and Califano, A. (2010). geWorkbench: an open source platform for integrative genomics. *Bioinformatics* 26, 1779–1780.
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., and Jensen, L.J. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808–D815.
- Ganter, B., Tugendreich, S., Pearson, C.I., Ayanoglu, E., Baumhueter, S., Bostian, K.A., Brady, L., Browne, L.J., Calvin, J.T., Day, G.J., et al. (2005). Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol.* 119, 219–244.
- Gardner, T.S., di Bernardo, D., Lorenz, D., and Collins, J.J. (2003). Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 301, 102–105.
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5, R80.
- Gilbert, L.A., and Hemann, M.T. (2010). DNA damage-mediated induction of a chemoresistant niche. *Cell* 143, 355–366.
- Goldwasser, F., Bae, I., Fornace, A.J., Jr., and Pommier, Y. (1996). Differential GADD45, p21CIP1/WAF1, MCL-1 and topoisomerase II gene induction and secondary DNA fragmentation after camptothecin-induced DNA damage in two mutant p53 human colon cancer cell lines. *Oncol. Res.* 8, 317–323.
- Günther, S., Kuhn, M., Dunkel, M., Campillos, M., Senger, C., Petsalaki, E., Ahmed, U., Urdiales, E.G., Gewiss, A., Jensen, L.J., et al. (2008). SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.* 36, D919–D922.
- Gupta, M., Han, J.J., Stenson, M., Maurer, M., Wellik, L., Hu, G., Ziesmer, S., Dogan, A., and Witzig, T.E. (2012). Elevated serum IL-10 levels in diffuse large

- B-cell lymphoma: a mechanism of aberrant JAK2 activation. *Blood* 119, 2844–2853.
- Hansen, N.T., Brunak, S., and Altman, R.B. (2009). Generating genome-scale candidate gene lists for pharmacogenomics. *Clin. Pharmacol. Ther.* 86, 183–189.
- Hirota, T., Lee, J.W., St John, P.C., Sawa, M., Iwaisako, K., Noguchi, T., Pongsawakul, P.Y., Sonntag, T., Welsh, D.K., Brenner, D.A., et al. (2012). Identification of small molecule activators of cryptochrome. *Science* 337, 1094–1097.
- Hoser, G., Majsterek, I., Romana, D.L., Slupianek, A., Blasiak, J., and Skorski, T. (2003). Fusion oncogenic tyrosine kinases alter DNA damage and repair after genotoxic treatment: role in drug resistance? *Leuk. Res.* 27, 267–273.
- Hsieh, A.C., Liu, Y., Edlind, M.P., Ingolia, N.T., Janes, M.R., Sher, A., Shi, E.Y., Stumpf, C.R., Christensen, C., Bonham, M.J., et al. (2012). The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* 485, 55–61.
- Ikura, T., Ogryzko, V.V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., and Nakatani, Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell* 102, 463–473.
- Iorio, F., Bosotti, R., Scacheri, E., Belcastro, V., Mithbaokar, P., Ferriero, R., Murino, L., Tagliaferri, R., Brunetti-Pierri, N., Isacchi, A., and di Bernardo, D. (2010). Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl. Acad. Sci. USA* 107, 14621–14626.
- Ito, T., Ando, H., Suzuki, T., Ogura, T., Hotta, K., Imamura, Y., Yamaguchi, Y., and Handa, H. (2010). Identification of a primary target of thalidomide teratogenicity. *Science* 327, 1345–1350.
- Jang, C.Y., Kim, H.D., Zhang, X., Chang, J.S., and Kim, J. (2012). Ribosomal protein S3 localizes on the mitotic spindle and functions as a microtubule associated protein in mitosis. *Biochem. Biophys. Res. Commun.* 429, 57–62.
- Keiser, M.J., Setola, V., Irwin, J.J., Laggner, C., Abbas, A.I., Hufeisen, S.J., Jensen, N.H., Kuijter, M.B., Matos, R.C., Tran, T.B., et al. (2009). Predicting new molecular targets for known drugs. *Nature* 462, 175–181.
- Kullback, S., and Leibler, R.A. (1951). On information and sufficiency. *Ann. Math. Stat.* 22, 79–86.
- 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.
- Lefebvre, C., Rajbhandari, P., Alvarez, M.J., Bandaru, P., Lim, W.K., Sato, M., Wang, K., Sumazin, P., Kustagi, M., Bisikirska, B.C., et al. (2010). A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* 6, 377.
- Li, J., Zhu, X., and Chen, J.Y. (2009). Building disease-specific drug-protein connectivity maps from molecular interaction networks and PubMed abstracts. *PLoS Comput. Biol.* 5, e1000450.
- Lomenick, B., Hao, R., Jonai, N., Chin, R.M., Aghajani, M., Warburton, S., Wang, J., Wu, R.P., Gomez, F., Loo, J.A., et al. (2009). Target identification using drug affinity responsive target stability (DARTS). *Proc. Natl. Acad. Sci. USA* 106, 21984–21989.
- Mani, K.M., Lefebvre, C., Wang, K., Lim, W.K., Basso, K., Dalla-Favera, R., and Califano, A. (2008). A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *Mol. Syst. Biol.* 4, 169.
- Miller, M.A. (2002). Chemical database techniques in drug discovery. *Nat. Rev. Drug Discov.* 1, 220–227.
- Pang, B., Qiao, X., Janssen, L., Velds, A., Groothuis, T., Kerkhoven, R., Nieuwland, M., Ovaa, H., Rottenberg, S., van Tellingen, O., et al. (2013). Drug-induced histone eviction from open chromatin contributes to the chemotherapeutic effects of doxorubicin. *Nat. Commun.* 4, 1908.
- Romito, A., Lonardo, E., Roma, G., Minchiotti, G., Ballabio, A., and Cobellis, G. (2010). Lack of sik1 in mouse embryonic stem cells impairs cardiomyogenesis by down-regulating the cyclin-dependent kinase inhibitor p57kip2. *PLoS ONE* 5, e9029.
- Scannell, J.W., Blanckley, A., Boldon, H., and Warrington, B. (2012). Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* 11, 191–200.
- Shao, S., Wang, Y., Jin, S., Song, Y., Wang, X., Fan, W., Zhao, Z., Fu, M., Tong, T., Dong, L., et al. (2006). Gadd45a interacts with aurora-A and inhibits its kinase activity. *J. Biol. Chem.* 281, 28943–28950.
- Sieg, D.J., Ilić, D., Jones, K.C., Damsky, C.H., Hunter, T., and Schlaepfer, D.D. (1998). Pyk2 and Src-family protein-tyrosine kinases compensate for the loss of FAK in fibronectin-stimulated signaling events but Pyk2 does not fully function to enhance FAK- cell migration. *EMBO J.* 17, 5933–5947.
- Silverman, B.W. (1986). Density estimation for statistics and data analysis (London, New York: Chapman and Hall).
- Smith, M.L., Chen, I.T., Zhan, Q., Bae, I., Chen, C.Y., Gilmer, T.M., Kastan, M.B., O'Connor, P.M., and Fornace, A.J., Jr. (1994). Interaction of the p53-regulated protein Gadd45 with proliferating cell nuclear antigen. *Science* 266, 1376–1380.
- 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.
- Thoma, C.R., Toso, A., Gutbrodt, K.L., Reggi, S.P., Frew, I.J., Schraml, P., Hergovich, A., Moch, H., Meraldi, P., and Krek, W. (2009). VHL loss causes spindle misorientation and chromosome instability. *Nat. Cell Biol.* 11, 994–1001.
- Thomas, J.P., Geiger, P.G., Maiorino, M., Ursini, F., and Girotti, A.W. (1990). Enzymatic reduction of phospholipid and cholesterol hydroperoxides in artificial bilayers and lipoproteins. *Biochim. Biophys. Acta* 1045, 252–260.
- Wehling, M. (2009). Assessing the translatability of drug projects: what needs to be scored to predict success? *Nat. Rev. Drug Discov.* 8, 541–546.
- Wernig, G., Kharas, M.G., Okabe, R., Moore, S.A., Leeman, D.S., Cullen, D.E., Gozo, M., McDowell, E.P., Levine, R.L., Doukas, J., et al. (2008). Efficacy of TG101348, a selective JAK2 inhibitor, in treatment of a murine model of JAK2V617F-induced polycythemia vera. *Cancer Cell* 13, 311–320.
- Wishart, D.S., Knox, C., Guo, A.C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., and Hassanali, M. (2008). DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 36, D901–D906.
- Wolpaw, A.J., Shimada, K., Skouta, R., Welsch, M.E., Akavia, U.D., Pe'er, D., Shaik, F., Bulinski, J.C., and Stockwell, B.R. (2011). Modulatory profiling identifies mechanisms of small molecule-induced cell death. *Proc. Natl. Acad. Sci. USA* 108, E771–E780.
- Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., and Kanehisa, M. (2008). Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 24, i232–i240.
- Yang, W.S., SriRamaratnam, R., Welsch, M.E., Shimada, K., Skouta, R., Viswanathan, V.S., Cheah, J.H., Clemons, P.A., Shamji, A.F., Clish, C.B., et al. (2014). Regulation of ferroptotic cancer cell death by GPX4. *Cell* 156, 317–331.
- Zhan, Q., Antinore, M.J., Wang, X.W., Carrier, F., Smith, M.L., Harris, C.C., and Fornace, A.J., Jr. (1999). Association with Cdc2 and inhibition of Cdc2/Cyclin B1 kinase activity by the p53-regulated protein Gadd45. *Oncogene* 18, 2892–2900.
- Zhang, Q.C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C.A., Bisikirska, B., Lefebvre, C., Accili, D., Hunter, T., et al. (2012a). Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* 490, 556–560.
- Zhang, S., Liu, X., Bawa-Khalife, T., Lu, L.S., Lyu, Y.L., Liu, L.F., and Yeh, E.T. (2012b). Identification of the molecular basis of doxorubicin-induced cardiotoxicity. *Nat. Med.* 18, 1639–1642.

## Predicting Drug Response in Human Prostate Cancer from Preclinical Analysis of In Vivo Mouse Models

### Graphical Abstract



### Authors

Antonina Mitrofanova, Alvaro Aytes, Min Zou, Michael M. Shen, Cory Abate-Shen, Andrea Califano

### Correspondence

cabateshen@columbia.edu (C.A.-S.),  
califano@c2b2.columbia.edu (A.C.)

### In Brief

Mitrofanova et al. describe a computational method using in vivo data from mouse models to identify drug combinations for human cancer treatment. They identify a drug combination that inhibits a key cancer dependency and show that genes responsive to treatment in mice identify human patients likely to benefit from treatment.

### Highlights

- Computational method uses in vivo data from mice to predict drugs for human cancer
- Method predicts optimal drug combinations using data from single agents
- Demonstrates efficacy of co-targeting PI3 and MAP kinases for prostate cancer
- Generalizable method to use mouse preclinical data to inform human cancer treatment

### Accession Numbers

GSE69211  
GSE69213

# Predicting Drug Response in Human Prostate Cancer from Preclinical Analysis of In Vivo Mouse Models

Antonina Mitrofanova,<sup>1,8</sup> Alvaro Aytes,<sup>2,8</sup> Min Zou,<sup>2</sup> Michael M. Shen,<sup>1,2,3,4,7</sup> Cory Abate-Shen,<sup>1,2,5,7,\*</sup> and Andrea Califano<sup>1,6,7,\*</sup>

<sup>1</sup>Department of Systems Biology

<sup>2</sup>Department of Urology

<sup>3</sup>Department of Medicine

<sup>4</sup>Department of Genetics and Development

<sup>5</sup>Department of Pathology and Cell Biology

<sup>6</sup>Department of Biochemistry and Molecular Biophysics

<sup>7</sup>Institute of Cancer Genetics

Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY 10032, USA

<sup>8</sup>Co-first author

\*Correspondence: [cabateshen@columbia.edu](mailto:cabateshen@columbia.edu) (C.A.-S.), [califano@c2b2.columbia.edu](mailto:califano@c2b2.columbia.edu) (A.C.)

<http://dx.doi.org/10.1016/j.celrep.2015.08.051>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## SUMMARY

Although genetically engineered mouse (GEM) models are often used to evaluate cancer therapies, extrapolation of such preclinical data to human cancer can be challenging. Here, we introduce an approach that uses drug perturbation data from GEM models to predict drug efficacy in human cancer. Network-based analysis of expression profiles from in vivo treatment of GEM models identified drugs and drug combinations that inhibit the activity of FOXM1 and CENPF, which are master regulators of prostate cancer malignancy. Validation of mouse and human prostate cancer models confirmed the specificity and synergy of a predicted drug combination to abrogate FOXM1/CENPF activity and inhibit tumorigenicity. Network-based analysis of treatment signatures from GEM models identified treatment-responsive genes in human prostate cancer that are potential biomarkers of patient response. More generally, this approach allows systematic identification of drugs that inhibit tumor dependencies, thereby improving the utility of GEM models for prioritizing drugs for clinical evaluation.

## INTRODUCTION

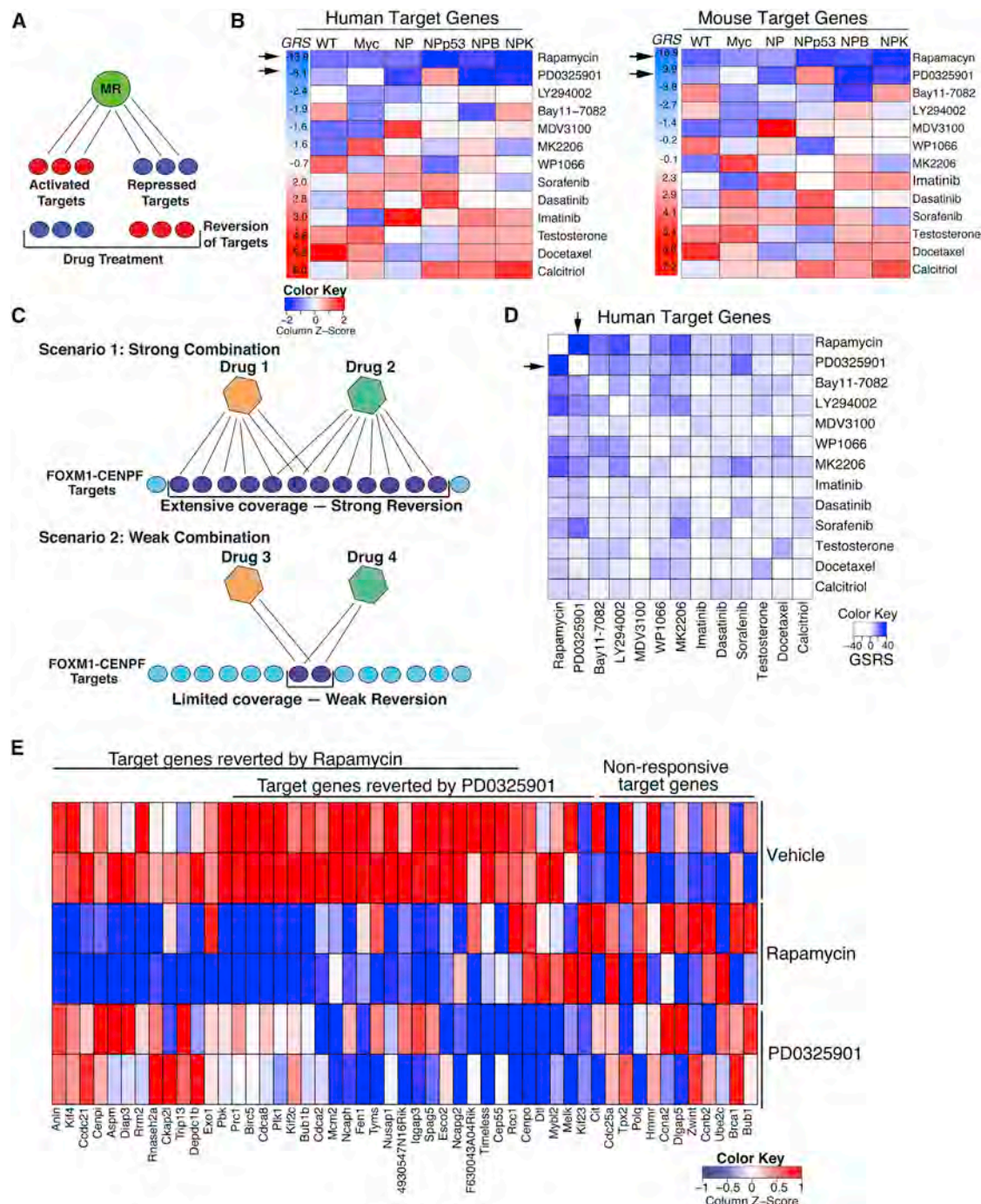
Recent large-scale genomic analyses have led to the identification of “actionable” driver genes of specific cancers that are therapeutically accessible, including oncogene and non-oncogene dependencies (Al-Lazikani et al., 2012; Garraway and Lander, 2013; Luo et al., 2009; Rubio-Perez et al., 2015). However, the accurate and efficient identification of drugs and drug combinations that inhibit such drivers within specific tumor contexts represents a major challenge, particularly for transcrip-

tional regulators that, in general, are pharmacologically inaccessible. Genetically engineered mouse (GEM) models are well suited to empower investigations of targeted inhibitors in the context of the native tumor microenvironment in vivo (Abate-Shen and Pandolfi, 2013; Politi and Pao, 2011; Sharpless and Depinho, 2006). However, species differences with respect to tumor histology, physiology, pharmacology, and metabolism often preclude direct extrapolation of preclinical findings from mouse models to human cancer.

In the current study, we introduce an innovative regulatory-network-based method that uses expression profiles from drug-treated GEM models to predict drugs and drug combinations that specifically inhibit the activity of established human cancer dependencies. We focus this proof-of-concept study on prostate cancer, a disease characterized by heterogeneity of its causal mechanisms and range of disease outcomes (Chang et al., 2014; Cooperberg et al., 2005; Roychowdhury and Chinnaiyan, 2013; Shen and Abate-Shen, 2010). In particular, while most locally invasive prostate tumors are curable, recurrent or aggressive tumors initially respond to androgen deprivation therapy but ultimately relapse to castration-resistant metastatic disease, which is nearly always fatal (Ryan and Tindall, 2011; Scher and Sawyers, 2005). While treatment options for castration-resistant metastatic prostate cancer have significantly improved in recent years (Mukherji et al., 2014; Rathkopf and Scher, 2013; Wong et al., 2014), none of the available treatments are as yet curative.

We have recently generated genome-wide reverse-engineered regulatory networks (henceforth “interactomes”) for both mouse and human prostate cancer (Aytes et al., 2014). Interrogation of these interactomes identified FOXM1 and CENPF as master regulators (i.e., key driver genes), which function synergistically to elicit synthetic lethality and are robust predictors of poor patient outcome (Aytes et al., 2014). Here, we show that interrogation of in vivo drug perturbation signatures from GEM models represents an effective strategy for systematic identification of specific drugs and drug combinations that





**Figure 1. Computational Prediction of Drugs that Inhibit FOXM1/CENPF Activity In Vivo**

(A) Shown is the strategy for prediction of single drugs. Drug reversion scores were calculated based on the degree to which target genes that are activated (red) by a master regulator (MR) are inhibited (blue) following drug treatment, and conversely, the degree to which target genes that are repressed (blue) by the MR are activated (red) following drug treatment (see [Supplemental Experimental Procedures](#)).

(B) Heatmap representations of GSEA used to calculate drug reversion scores across a series of GEM models with a series of drugs, as indicated (see [Supplemental Experimental Procedures](#)). GSEA were done using the mouse in vivo drug perturbation signatures as the reference and human or mouse FOXM1/CENPF target genes inferred from their respective prostate cancer interactomes, as indicated, as the query gene set. Global reversion scores (GRSs) were calculated for each drug by combining the individual NES for each GEM model using a metric based on the Stouffer integration formulation (see [Supplemental Experimental Procedures](#)). Arrows point to the two drugs with the highest GRSs.

(C) Shown is the strategy for prediction of drug synergy. Pairwise combinations of data from individual drug treatments (as in A) were assessed to predict drugs that effectively revert FOXM1/CENPF target genes when used in combination. Scenario 1 illustrates two drugs that inhibit (i.e., revert) many target

(legend continued on next page)

inhibit the transcriptional activity of FOXM1/CENPF. Strikingly, drug combinations that revert transcriptional activity of these proteins are highly effective in abrogating tumorigenesis in vivo and well correlated with patient outcome. We propose that this computational method can be generalized for more effective utilization of preclinical data from GEM models to predict optimal drugs and drug combinations and thereby dramatically improve the utilization of GEM models to prioritize compounds for clinical investigation.

## RESULTS

### Systematic Inference of FOXM1/CENPF Inhibitors In Vivo

The current methodology is predicated on our previous analyses showing that expression of the target genes of a given master regulator (MR) (its regulon) represents an effective reporter to predict the activity of the MR for a given cancer phenotype (Aytes et al., 2014; Carro et al., 2010; Chen et al., 2014). Here, we have extended this concept to evaluate whether such regulon can be used as a reporter to quantitatively measure the ability of a drug or drug combination to inhibit the activity of the corresponding MR. In general, reversion of MR activity would correspond to the ability of a given drug to *downregulate* its *activated* target genes and *upregulate* its *repressed* targets (Figure 1A). As a proof of concept for this approach, we evaluated drugs for their ability to inhibit the master regulator pair, FOXM1/CENPF, which we have previously established to be a key synthetic lethal dependency of prostate tumor malignancy (Aytes et al., 2014). In particular, we tested whether candidate therapeutic agents could be prioritized based on in vivo perturbation by assessing their ability to “reverse” the FOXM1/CENPF regulon. We focused on the activated shared targets of FOXM1/CENPF, since the number of repressed targets is too few for analysis. However, both activated and repressed targets may be used in general.

To assess this strategy, we used a drug perturbation dataset that includes drugs with known prostate cancer relevance, such as those that inhibit the androgen receptor, or key signaling pathways such as phosphatidylinositol 3-kinase (PI3K)/mTOR or MAP kinase or standard chemotherapy (Aytes et al., 2014; see Supplemental Experimental Procedures). In vivo drug perturbation studies were performed using multiple GEM models representative of advanced prostate cancer (Aytes et al., 2014; see Supplemental Experimental Procedures) to avoid potential bias introduced by any individual model. The in vivo drug perturbation data were analyzed by gene set enrichment analysis (GSEA) (Subramanian et al., 2005) to assess the inhibition (i.e., reversion) of FOXM1/CENPF shared target genes; analyses were per-

formed separately for the mouse and human targets (Figure 1B). Using GSEA, we obtained a normalized enrichment score (NES) for each drug signature and each GEM model, which we define as the reversion score ( $RS_{\text{FOXM1/CENPF}}$ ), to assess the compound's ability to inhibit FOXM1/CENPF activity in a specific GEM model (Table S1). From these analyses, a global reversion score ( $GRS_{\text{FOXM1/CENPF}}$ ) was assigned for each drug by integrating each of the GEM-specific  $RS_{\text{FOXM1/CENPF}}$  scores, using a metric based on Stouffer's integration formulation (Whitlock, 2005) (Figure 1B; Figure S1; Supplemental Experimental Procedures). Thus, drugs that most effectively inhibit FOXM1/CENPF activity are those having the most negative  $GRS_{\text{FOXM1/CENPF}}$ . Notably, FOXM1/CENPF target genes from either mouse or human yielded equivalent  $GRS_{\text{FOXM1/CENPF}}$  (Figure 1B; Table S1), indicating conservation of the predicted drug response between mouse and human prostate cancer.

Among the individual drugs tested in the GEM models, the two with the most significant negative  $GRS_{\text{FOXM1/CENPF}}$  were rapamycin and PD0325901. These drugs inhibit the PI3K/mTOR and MAP kinase signaling pathways, respectively, which are frequently dysregulated in advanced prostate cancer (Aytes et al., 2013; Kinkade et al., 2008; Taylor et al., 2010). Specifically, the GRSs for rapamycin were  $GRS_H = -13.9$  (human targets) and  $GRS_M = -16.9$  (mouse targets) and for PD0325901 were  $GRS_H = -8.1$  and  $GRS_M = -9.9$  (Figure 1B; Table S1). In contrast to rapamycin and PD0325901, other drugs including docetaxel, a standard-of-care chemotherapy for advanced prostate cancer (Pienta and Smith, 2005), were not predicted to be effective for inhibiting the FOXM1/CENPF regulon ( $GRS_H = 5.8$  and  $GRS_M = 5.6$ ; Figure 1B; Figure S1).

### Systematic Inference of Drug Synergy

Next, we tested whether this computational approach could be extended to infer drug combinations that cooperate to inhibit MR activity, again using FOXM1 and CENPF as a proof of concept. These analyses are based on the hypothesis that effective drug combinations should induce a more significant reversal of MR-specific regulon expression, compared to the individual drugs (Figure 1C; Supplemental Experimental Procedures). Notably, such logic can be implemented based on individual drug signatures, without requiring in vivo signatures from drug combinations, which vastly increases the experimental efficiency for prioritizing drug combinations based on in vivo preclinical data.

To estimate a global synergistic reversion score (GSRS) for each drug pair, we assessed the predicted reversion score for all possible combinations of two drug treatments across each of the GEM models. First, the synergistic reversion score (SRS) was calculated for each GEM model as an harmonic mean

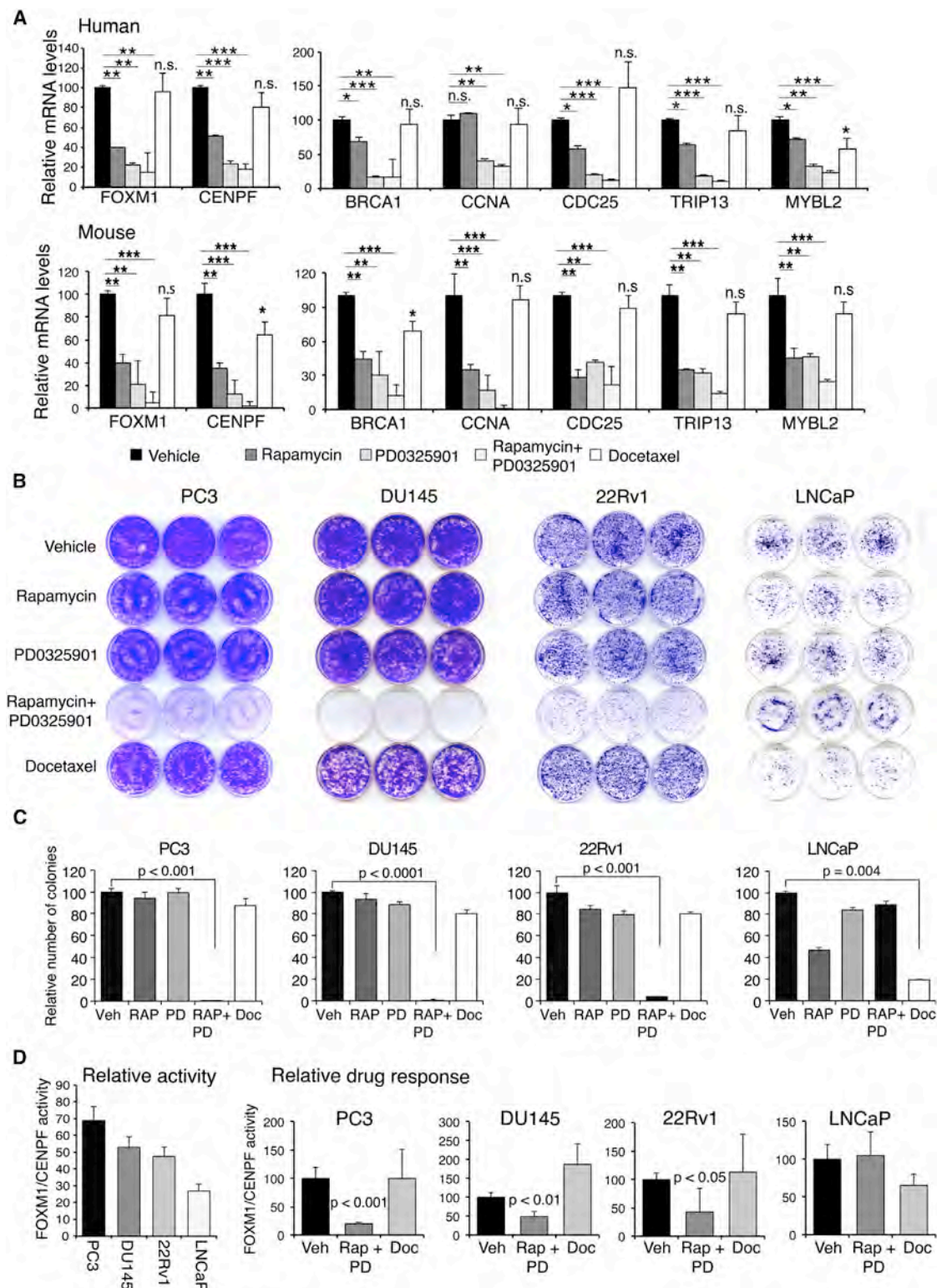
genes, thereby resulting in strong reversion. Scenario 2 illustrates two other drugs that inhibit (i.e., revert) relatively few target genes, thereby resulting in weak reversion.

(D) Heatmap representation depicting global synergistic reversion scores (GSRSs) for each possible pair of drugs across the series of GEM models based on FOXM1/CENPF human target genes. GSRSs were calculated by combining the synergistic reversion scores for targets affected by the drug combinations (see Supplemental Experimental Procedures). Heatmap intensity (blue) represents the predicted degree of reversion (GSRS); arrows indicate drug pairs with the highest combined GSRS.

(E) Heatmap showing the relative expression levels of FOXM1/CENPF target genes reverted by treatment with rapamycin or PD0325901 versus vehicle; also shown are genes that are not reverted (i.e., non-responsive) to these drugs.

Figure S1 is related to Figure 1; computational predictions of GRSs and GSRSs are provided in Tables S1 and S2, respectively.





**Figure 2. Validation of Drug Efficacy, Synergy, and Specificity in Prostate Cancer Cells**

(A) Real-time PCR of mRNA expression levels of *FOXM1* and *CENPF* and their shared target genes following treatment with rapamycin and/or PD0325901 or docetaxel in DU145 human prostate cancer cells (top) or NPK mouse prostate tumors (bottom).

(B and C) Colony formation assays in the indicated human prostate cancer cells, PC3, DU145, 22Rv1, and LNCaP, following treatment with rapamycin (Rap) and/or PD0325901 (PD), or docetaxel (Doc). (B) Representative colony formation assays. (C) Quantification of independent assays performed in triplicate.

(legend continued on next page)

(F-score) that first maximizes the number of unique targets affected by each drug, and then the total number of targets affected by both drugs (Figure 1C; Supplemental Experimental Procedures). These analyses identified several combinations, most of which included rapamycin or PD0325901, which were predicted to be more effective than the individual compounds based on their GSRSs (Figure 1D; Table S2). In particular, the rapamycin + PD0325901 combination was predicted to have the strongest global inhibition of the FOXM1/CENPF regulon, both with respect to total number of targets affected by both drugs and the number of unique targets affected by each drug, resulting in the most significant negative global synergistic reversion score (GSRS<sub>H</sub> = -40.4; p value < 0.001 compared to a random model; see Supplemental Experimental Procedures). This theoretical prediction was validated by assessment of FOXM1/CENPF target genes that were reverted by rapamycin or PD0325901 following drug treatment in vivo (Figure 1E).

### Experimental Validation of Drug Specificity and Synergy in Cell Culture

Based on these computational predictions, we performed experimental validation to assess whether rapamycin and/or PD0325901 specifically inhibit the FOXM1/CENPF regulon in relevant mouse and human prostate cancer cell culture models, and if so, whether these drugs affect cell growth and tumorigenicity in a FOXM1/CENPF-dependent manner. First, we validated the underlying computational prediction that treatment with rapamycin and PD0325901 reverts the expression of shared target genes of FOXM1/CENPF. Using real-time PCR, we found that treatment with rapamycin and PD0325901, but not docetaxel, inhibited expression of both FOXM1 and CENPF as well as their shared target genes in several human and mouse prostate cancer models (Figure 2A; Figure S2A). This inhibition of target genes was coincident with inhibition of the corresponding signaling pathways, namely PI3K/mTOR and MAP kinase in the mouse and human cells (Figures S2B and S2C). Notably, inhibition of colony formation was significantly greater when the drugs were combined than when used individually (Figure 2B, C), which supports the computational prediction of rapamycin + PD0325901 synergy.

To address the specificity of the rapamycin + PD0325901 drug combination for inhibition of FOXM1/CENPF activity, we assessed whether this combination was preferentially more potent in contexts having high levels of FOXM1/CENPF activity. First, we surveyed the expression and activity of FOXM1/CENPF in a series of human and mouse cell lines; “activity” was determined experimentally by analyses of the expression of FOXM1/CENPF shared target genes (Figure 2D; Figures S2D–S2F). These studies revealed that PC3 cells have the highest levels of FOXM1/CENPF activity, whereas LNCaP cells have lower levels (Figure 2D). Correspondingly, human prostate cancer cells with higher levels of FOXM1/CENPF activity had greater response

to rapamycin + PD0325901 treatment, as evident from the strong inhibition of activity and colony formation, whereas LNCaP cells, which have low levels of FOXM1/CENPF activity, had a modest response to rapamycin + PD0325901 (Figures 2B–2D). In contrast, this relationship to FOXM1/CENPF activity was not observed following docetaxel treatment of these cells (Figures 2B–2D). Similar findings were observed in mouse prostate cancer cells wherein response to rapamycin + PD0325901 treatment was correlated with the relative levels of FoxM1/Cenpf activity (Figures S2D and S2E).

Moreover, the dependence on FOXM1/CENPF in the human prostate cancer cells was evident by the reduction in the half maximal inhibitory concentration (IC<sub>50</sub>) for rapamycin + PD0325901, but not docetaxel, following the silencing of both FOXM1 and CENPF in human prostate cancer cell lines with high levels of activity (Figure S3). Conversely, overexpression of FOXM1 and CENPF in a non-prostate cancer cell line, HEK293, resulted in an increase in the IC<sub>50</sub> for rapamycin + PD0325901, but not docetaxel (Figure S3). Taken together, these findings validate the computational prediction that FOXM1/CENPF activity is specifically inhibited by rapamycin + PD0325901.

### Experimental Validation of Drug Efficacy and Specificity In Vivo

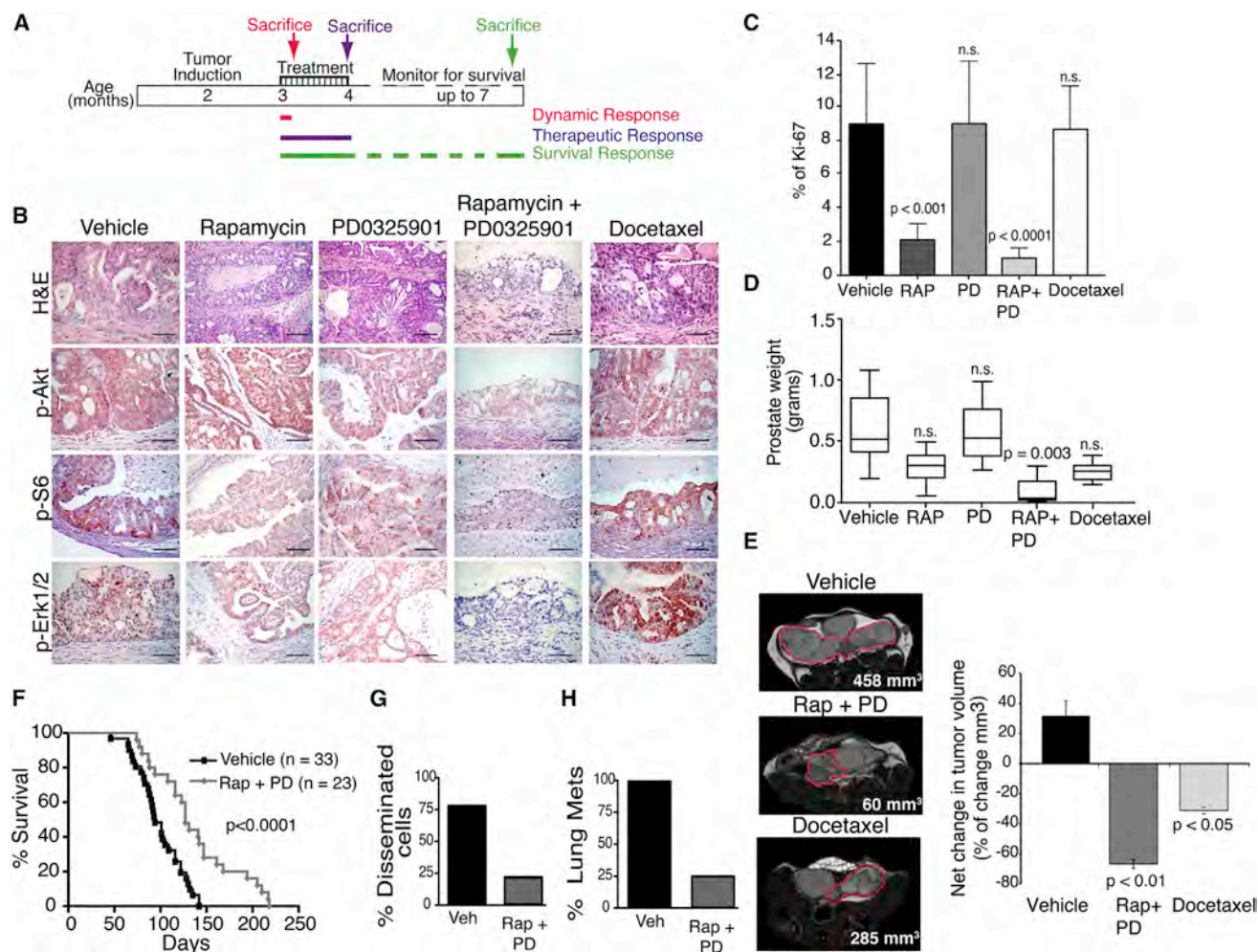
The synergistic effects of combined treatment with rapamycin + PD0325901 were even more dramatic in vivo. In particular, we performed preclinical studies using *NPK* mice (*Nkx3.1<sup>CreERT2</sup>; Pten<sup>flx/flx</sup>; Kras<sup>LSL-G12D/+</sup>*), which model aggressive, metastatic prostate cancer that is dependent on FOXM1/CENPF activity (Aytes et al., 2013, 2014). Tumor-bearing *NPK* mice were treated with rapamycin and/or PD0325901, or docetaxel, for 5 days (i.e., the dynamic response cohort) or 1 month (i.e., the therapeutic response cohort) (Figure 3A; Table S3). Mice were then either sacrificed for analysis or monitored for the effects of drug treatment on survival and metastasis (i.e., the survival response cohort) (Figure 3A; Table S3).

Whereas treatment with either drug individually had a modest therapeutic benefit at the various endpoints, the combination of rapamycin + PD0325901 had a profound effect at all tumor endpoints in the therapeutic response cohort (Figures 3B–3E). In particular, treatment with rapamycin + PD0325901, but not docetaxel, resulted in profound abrogation of the histological phenotype, coincident with inhibition of relevant signaling pathways, as evident by immunohistochemistry (Figure 3B). Moreover, tumors treated with rapamycin + PD0325901, but not docetaxel, displayed a significant decrease in cellular proliferation (p < 0.0001) (Figure 3C), as well as significant reduction in tumor burden, as measured by tumor weight (p = 0.003) and tumor volume using MRI (p < 0.01) (Figures 3D and 3E). Furthermore, these effects on phenotype and tumor burden were accompanied by a significant improvement in survival (p < 0.0001) (Figure 3F), as well as a 3-fold reduction in the incidence of

(D) (Left) Relative activity of FOXM1/CENPF in human prostate cancer cell lines. Activity levels were calculated based on expression levels of 10 FOXM1/CENPF target genes (see Figure S2F). (Right) Relative drug response assessed for FOXM1/CENPF activity levels in the human prostate cancer cell lines following treatment with rapamycin + PD0325901 (Rap + PD) or docetaxel (Doc). Differences between treatment groups were assessed using Student's t test. When indicated, p values are represented as \*p < 0.01, \*\*p < 0.001, and \*\*\*p < 0.0001. Bars represent mean ± SD.

Figures S2 and S3 are related to Figure 2.





**Figure 3. Validation of Drug Synergy and Specificity In Vivo**

(A) Shown is the design of preclinical studies. *NPK* mice were induced to form tumors by delivery of tamoxifen at 2 months of age as in (Aytes et al., 2013). Mice were treated with rapamycin and/or PD0325901, or docetaxel for 5 days (dynamic response cohort) or 1 month, following which mice were sacrificed for analyses (therapeutic response cohort) or monitored for survival (survival response cohort).

(B–E) Analysis of the therapeutic response cohort following treatment with rapamycin (Rap) and/or PD0325901 (PD) or docetaxel (Doc) as indicated ( $n = 5$  mice/treatment group). (B) Representative sections of H&E staining or immunohistochemical staining for the indicated markers of the PI3K/mTOR or MAP kinase signaling pathways; scale bars represent 100  $\mu$ m. (C) Relative cellular proliferation following drug treatment as determined by the percent of Ki67-positive cells relative to total epithelial cells. (D) Prostate weight (in grams) following drug treatment. (E) Longitudinal MRI imaging showing representative images following drug treatment with tumor volumes indicated. The panel to the right represents the net change in tumor volume following 1 month of drug treatment.

(F–H) Shown is analysis of the survival response cohort. (F) Survival analysis showing the improvement in survival following treatment with rapamycin and PD0325901 (Rap + PD) compared with the vehicle-treated mice. (G) Percentage of mice with disseminated cells in the bone marrow and (H) percentage of mice with lung metastases following treatment with vehicle (Veh) or rapamycin + PD0325901 (Rap + PD) ( $n = 10$  mice/treatment group). Differences between groups were assessed using Student's *t* test; bars represent mean  $\pm$  SD. In (F), *p* value corresponds to a log-rank test.

disseminated tumor cells in the bone marrow and a 4-fold reduction in the incidence of lung metastases (Figures 3G and 3H). Together, these findings validate the concept that treatment with rapamycin + PD0325901 inhibits growth of FOXM1/CENPF-dependent tumors.

### Relationship of Mouse Drug-Treatment Signatures to Human Cancer

Given the striking reduction in tumor and metastatic burden following treatment with rapamycin + PD0325901, we evaluated

whether this combination might be sufficient to broadly inhibit molecular processes associated with advanced, FOXM1/CENPF-dependent prostate cancer. We addressed this question by analyzing signatures obtained by differential gene expression analysis of *NPK* prostate tumors treated with vehicle or rapamycin + PD0325901 for 1 month (i.e., the therapeutic response cohort; Table S3), which resulted in extensive abrogation of the tumor phenotype (see Figure 3). We compared this “therapeutic response” signature to a reference mouse “tumor” signature, corresponding to differential gene expression

between phenotypically wild-type prostates and *NPK* prostate tumors, which captures the transition from normal prostate to fully malignant prostate cancer (Table S3). Strikingly, genes that were differentially expressed in the therapeutic response signature were strongly negatively enriched in the mouse tumor signature ( $NES = -8.58$ ;  $p < 0.001$ ) (Figure S4A). Further evidence that rapamycin + PD0325901 treatment results in broad inhibition beyond their respective target signaling pathways was provided by biological pathway analysis. In particular, pathways that were significantly inhibited (i.e., reverted) following treatment of the *NPK* tumors with rapamycin + PD0325901, but not docetaxel, include several that are important for tumor progression and are not directly related to mTOR/PI3K/MAP kinase signaling (Figure 4A; Table S4).

To evaluate molecular processes that are inhibited immediately following drug treatment, we analyzed a “dynamic response” signature, representing a time point wherein the drugs are active but the tumor phenotype has not yet been abrogated (Figure 3A; Table S3; and data not shown). In particular, this short-term treatment with rapamycin + PD0325901 resulted in reversion of FOXM1/CENPF targets, as predicted by our computational approach (Figure S4B; see Figure 1E). Comparison of this “dynamic response” signature to a reference mouse “malignancy signature,” based on comparison of non-malignant prostate tumors from *NP* mice to fully malignant *NPK* tumors (Aytes et al., 2013), revealed a striking negative enrichment (i.e., strong reversion) ( $NES = -8.34$ ;  $p < 0.001$ ) (Figure 4B), suggesting that the rapamycin + PD0325901 combination inhibits molecular processes associated with *NPK* tumor malignancy even prior to their overt effects on the tumor phenotype.

To assess conservation of these molecular changes with human prostate cancer, we performed GSEA to compare a humanized version of the mouse dynamic response signature with human prostate cancer signatures (see Supplemental Experimental Procedures). We used three independent human prostate cancer signatures, each of which is based on distinct clinical endpoints (Table S3): (1) a malignancy signature based on the Taylor dataset (Taylor et al., 2010), which compares patients having low Gleason score and no biochemical recurrence ( $n = 39$ ) to those with high Gleason score and a short time to biochemical recurrence ( $n = 10$ ) (Aytes et al., 2013); (2) a metastasis signature based on the Balk dataset (Stanbrough et al., 2006), which compares hormone-naïve prostate tumors ( $n = 22$ ) to bone metastases from castration-resistant prostate cancer ( $n = 29$ ) (Aytes et al., 2014); and (3) a survival signature based on the Sboner dataset (Sboner et al., 2010), which compares transurethral resections from patients who survived for nearly 200 months ( $n = 12$ ) to those who died of prostate cancer within 12 months ( $n = 6$ ) (Wang et al., 2013). Strikingly, the mouse dynamic response signature was strongly negatively enriched when compared with each of these human signatures, indicating that genes that are consistently overexpressed in aggressive prostate cancer are inhibited following drug treatment (Taylor signature  $NES = -5.48$ ,  $p < 0.001$ ; Balk signature  $NES = -5.26$ ,  $p < 0.001$ ; and Sboner signature  $NES = -6.40$ ,  $p < 0.001$ ) (Figure 4C). In contrast, the docetaxel treatment response signature was either minimally or not negatively enriched in these human signatures (Figure S4C).

We then asked whether the mouse dynamic response signature could reverse a “FOXM1/CENPF activity” signature in human prostate cancer. This FOXM1/CENPF activity signature, defined using the Sboner dataset (Sboner et al., 2010), corresponds to differential gene expression between patient samples having low versus high levels of FOXM1/CENPF activity, which was measured by enrichment of the FOXM1/CENPF regulon in each patient using single-sample master regulator inference algorithm (ssMARINA) as in Aytes et al. (2014) (see Supplemental Experimental Procedures). GSEA comparing the “FOXM1/CENPF” activity signature with the mouse “dynamic response” signature showed strong negative enrichment ( $NES = -6.43$ ,  $p < 0.001$ ) (Figure 4D), which supports the concept that patients with high levels of FOXM1/CENPF activity should respond more effectively to rapamycin + PD0325901 treatment. Notably, similar comparison with a docetaxel treatment response signature did not indicate such relationship ( $NES = 0.37$ ,  $p = 0.77$ ) (Figure S4D).

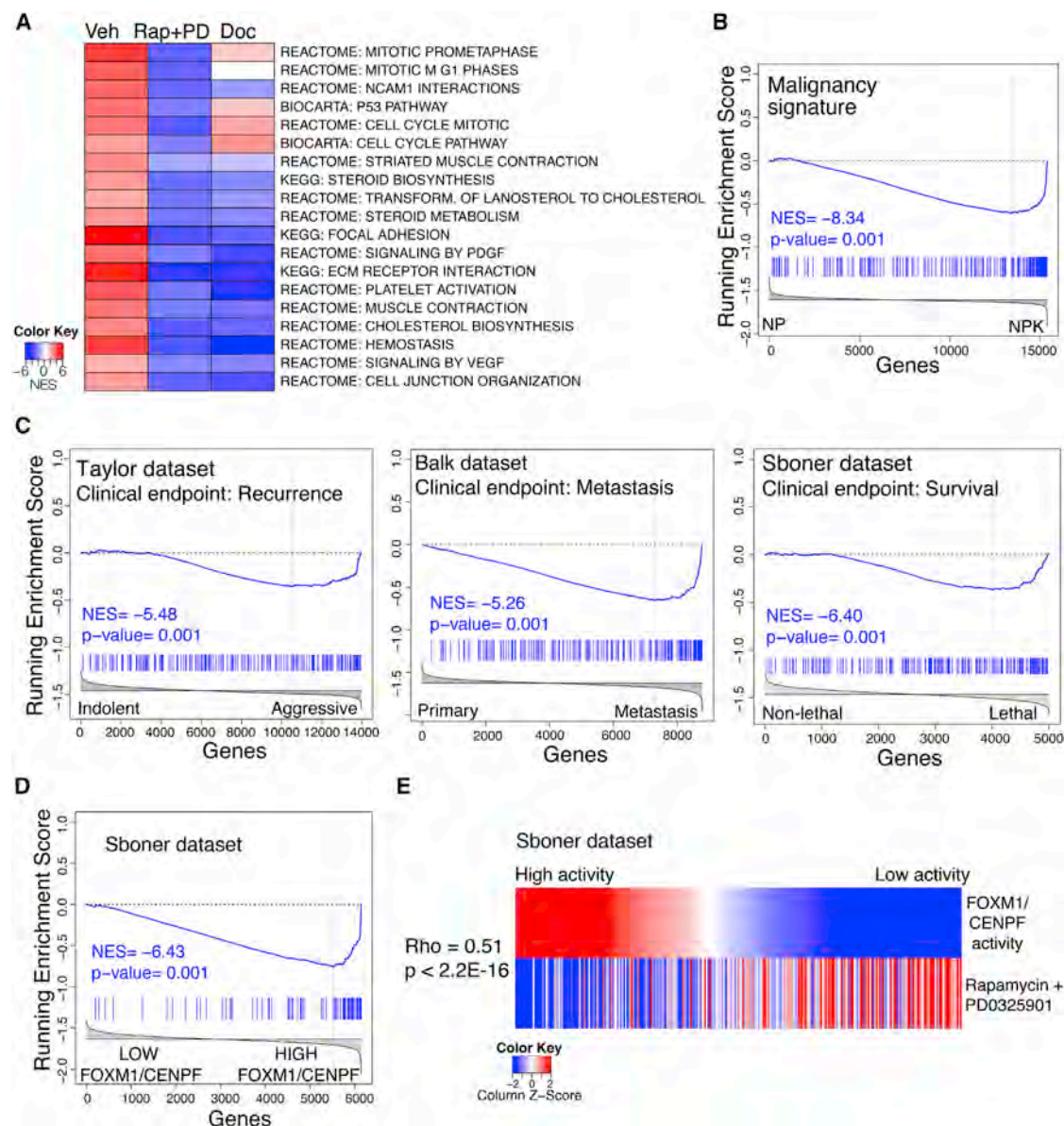
We further evaluated the correlation between FOXM1/CENPF activity levels and predicted treatment response in each patient in the Sboner dataset estimated using ssMARINA and GSEA, respectively. We found that inferred FOXM1/CENPF activity levels and predicted treatment response were strongly correlated (Spearman’s  $\rho = 0.51$ ,  $p < 2.2 \times 10^{-16}$ ) (Figure 4E), which was not the case for the docetaxel treatment response (Figure S4E). Taken together, these computational analyses suggest that the molecular programs (i.e., genes and pathways) specifically inhibited (reverted) by rapamycin + PD0325901 in the mouse model are conserved with those that drive aggressive human prostate cancer, and in particular in patients having high levels of FOXM1/CENPF activity.

### Conservation of Treatment Response in Mouse and Human Prostate Cancer

Given the conservation in the molecular programs affected by drug treatment in the GEM models and human prostate cancer, we next asked whether we could use the mouse treatment response signature to identify genes predicted to be associated with treatment response in humans. First, we identified candidate rapamycin + PD0325901-responsive genes by interrogating the mouse prostate cancer interactome (Aytes et al., 2014) with the dynamic response signature using the standard MARINA algorithm to identify MRs of treatment response in the mouse (Lefebvre et al., 2010). We then compared these MRs with the orthologous human genes to identify those predicted both to be regulated by FOXM1/CENPF in human prostate cancer and to be downregulated by drug treatment; we refer to these as “predicted treatment-responsive genes” and distinguish them from other FOXM1/CENPF target genes that are not predicted to be responsive to the treatment (Figure 5A). Notably, real-time PCR analyses confirmed that the expression levels of these predicted treatment-responsive genes were indeed inhibited by rapamycin + PD0325901 in human prostate cancer cell lines, whereas the expression levels of the predicted non-responsive genes were not inhibited by such treatment (Figure 5B).

These treatment-responsive genes (including *TOP2A*, *UHRF1*, *ASF1B*, *MCM4*, *WHSC1*, *MCM2*, *SUV39H1*, *BLM*, *BRCA1*,





**Figure 4. Cross-Species Analyses of Drug-Treatment Response**

(A) Heatmap depiction showing representative pathways that are significantly changed following treatment with rapamycin + PD0325901 (Rap + PD) or docetaxel (Doc) relative to vehicle treatment (Veh). Pathway analysis was done by GSEA using a “humanized” version of the dynamic response allograft tumor signature (see Table S3 and Supplemental Experimental Procedures). A complete list of pathways is provided in Table S4.

(B–D) GSEA using as the query gene set the mouse rapamycin + PD0325901 dynamic treatment response signature (B) or a “humanized” version of this signature (C and D); normalized enrichment score (NES) and associated p values are shown. In (B), the reference is mouse “malignancy” signature, which represents differentially expressed genes from NP versus NPK mouse tumors as reported previously (Aytes et al., 2013). In (C), the references are three independent human tumor signatures (i.e., Taylor, Balk, or Sboner), each of which compare differentially expressed genes representing less aggressive versus more aggressive prostate cancer specimens (Table S3). In (D), the reference signature represents differentially expressed genes in patients from the Sboner dataset having low versus high levels of FOXM1/CENPF activity, which was inferred using single sample MARiNa (ssMARiNa) (see Supplemental Experimental Procedures).

(E) Heatmap showing the correlation in human patients from the Sboner dataset of FOXM1/CENPF activity levels (top) with the corresponding predicted drug-treatment response for rapamycin + PD0325901 (bottom). As above, FOXM1/CENPF activity was estimated for each patient using ssMARiNa. The treatment response for each patient was inferred using a “humanized” version of the mouse dynamic treatment signature using GSEA (see Supplemental Experimental Procedures). Correlation coefficient ( $\rho$ ) and associated p value were estimated using Spearman’s correlation.

Figure S4 and Tables S3 and S4 are related to Figure 4.



(G) Kaplan-Meier survival analysis based on the master regulator activity levels of predicted treatment-responsive genes in the Sboner dataset using prostate cancer-specific survival as the endpoint. The p value was estimated using a log-rank test of the difference in outcome between patients with higher activity levels (red) and those with lower/no activity (blue).



*CCNA2*, *E2F1*, and *MYBL2*) have known functions in DNA repair, epigenetic modifications, cell cycle, proliferation, and/or survival, which are all associated with cancer malignancy. Notably, each of these is overexpressed in advanced human prostate cancer, and their activity levels are associated with disease outcome, as shown by univariate analyses using a COX proportional hazard model on the Sboner dataset (Figures 5C and 5D). Moreover, analyses based on the Balk dataset revealed robust activity levels of the treatment-responsive genes in metastatic samples compared to primary tumors (Figure 5E).

We further demonstrated the association of the activity levels of the treatment-responsive genes with drug response on a patient-by-patient basis, estimated using ssMARINA and GSEA, respectively, on the Sboner dataset (Figure 5F). In particular, the average activity levels of treatment-responsive genes were strikingly correlated with the rapamycin + PD0325901 drug response (Spearman  $\rho = 0.57$ ,  $p < 2.2 \times 10^{-16}$ ) (Figure 5F), similar to that observed for the FOXM1/CENPF activity (see Figure 4D). Most notably, multivariate Kaplan-Meier survival analysis on the Sboner dataset to evaluate disease-specific survival revealed that patients with higher activity levels of the treatment-responsive genes have a shorter time to prostate cancer-specific death compared to patients with lower activity levels (log-rank  $p$  value =  $1.7 \times 10^{-5}$ ) (Figure 5G). Importantly, the activity of the treatment-responsive subnetwork of the FOXM1/CENPF regulon was more significant than the FOXM1/CENPF regulon (log-rank  $p$  value =  $1.3 \times 10^{-4}$ ) and also outperformed a random comparable set of genes with respect to the COX proportional hazard model ( $p$  value for improvement  $< 0.001$ ) and Kaplan-Meier survival analysis ( $p$  value for improvement  $< 0.015$ ) (see Supplemental Experimental Procedures). Taken together, these findings suggest that computationally predicted treatment-responsive genes can be used to identify patients that are likely to benefit from treatment with drugs that co-target the PI3K/mTOR and MAP kinase signaling pathways and provide a proof of concept for the overall approach.

## DISCUSSION

In this study, we introduce a generalizable computational approach to extrapolate in vivo preclinical treatment data from GEM models to inform on human cancer treatment. Our method infers drug efficacy based on the ability of a given drug to revert the transcriptional regulon of key dependencies that drive the tumor phenotype. Importantly, we show that this method can be used to prioritize drug combinations based on analysis of individual compounds, which greatly enhances the value of in vivo preclinical analyses of compounds in mice. We demonstrate this approach with a proof-of-concept study based on identification of drugs and drug combinations that inhibit the activity of FOXM1/CENPF, which were chosen for their established relevance for lethal prostate cancer (Aytes et al., 2014). However, this approach should be applicable to identify candidate drugs and drug combinations for many other driver gene(s) of interest and not limited to prostate cancer. Notably, the molecular programs affected by drug treatment in the GEM model are well conserved with human prostate cancer, which supports the concept that analyses of drug-treatment data from mouse

models can be used to identify treatment-responsive genes for human prostate cancer. Thus, we have described a method to identify drugs and drug combinations that specifically inhibit cancer driver genes, as well as to identify potential biomarkers to predict the efficacy of drug treatments for individual patients.

Several features of our approach distinguish it from other strategies previously used to screen for drug response in human cancer. First, most other approaches have been based on analyses of cancer cell lines in culture (e.g., Barretina et al., 2012; Garnett et al., 2012), whereas our study is based on drug perturbation of GEM models in vivo. Thus, we evaluate drug efficacy in the context of the native tumor microenvironment and intact immune system, which are now widely recognized as being essential for drug response in vivo, particularly given recent advances in immunotherapy. Although the tumor context of any individual GEM model is unlikely to fully recapitulate that of human cancer, we address this limitation by analyzing multiple distinct GEM models to avoid idiosyncratic GEM-specific biology. Indeed, we have observed a remarkable concordance of the molecular consequences of drug treatment between our “consensus” analyses of mouse models and human prostate cancer.

A second distinguishing feature of our approach is its ability to identify synergistic drug combinations based on single-agent treatment data. From a practical standpoint, the number of drugs that can be feasibly evaluated using in vivo perturbations in a series of GEM models is limited. Therefore, the ability to evaluate the efficacy of drug combinations by profiling a relatively small number of single drugs (e.g., the 100 most relevant compounds) would allow assessment of a very large potential combination therapy space (e.g., 4,950 combinations), thus affording significant economy of scale.

A third important feature is that our computational method identifies drugs based on their ability to inhibit specific drivers of the tumor phenotype, rather than on overall toxicity or inhibition of more general tumor-related parameters. In particular, the method evaluates the efficacy of drug response based on inhibition of the transcriptional regulon of specific master regulators of interest. Furthermore, our computational analysis of treatment response in the GEM models in vivo has also identified treatment-responsive genes that are conserved in human prostate cancer. We propose that such treatment-responsive genes may serve as surrogate biomarkers to infer the potential efficacy of drug treatments in patients. In particular, our current findings suggest that previous analyses may have underestimated the value of molecular inference of preclinical data from GEM models for not only predicting optimal drug combinations but also identifying molecular markers for predicting treatment response to such drugs.

The PI3K/mTOR and MAP kinase signaling pathways are known to be dysregulated in many advanced prostate cancers (Aytes et al., 2013; Kinkade et al., 2008; Taylor et al., 2010). Currently, drugs that target these pathways (albeit not rapamycin and PD0325901) are being evaluated in numerous clinical trials for prostate cancer and many other solid tumors, including combination-therapy regimes. Results from the current study as well as previous work (Aytes et al., 2014) suggest that aberrant levels of FOXM1 and CENPF, as assessed by immunostaining of tumor samples, may identify patients who would likely benefit from

treatment with agents that target the PI3K/mTOR and/or MAP kinase signaling pathways. In addition, our study suggests that the treatment-responsive genes we have identified could provide intermediate biomarkers to assess short-term efficacy of combination therapy in patients, a strategy that can be readily generalized to other targets and therapies. Thus, our studies may inform or modulate the design of clinical trials or help provide a mechanistic basis for clinical findings.

Beyond prostate cancer, our computational methodology may be beneficial to identify drugs that target key actionable targets in vivo for a wide range of tumor types, oncogene and non-oncogene dependencies, and therapeutic agents, including both US Food and Drug Administration-approved and experimental compounds. Since many cancer types now have relevant GEM models that are being used in many preclinical studies, it would be advantageous to use our approach to apply these preclinical data from GEMs to analyze treatment response in human cancer.

## EXPERIMENTAL PROCEDURES

### Computational Prediction of Drug Synergy

Computational inference of drugs that inhibit FOXM1 and CENPF activity was done using their shared target genes predicted from the mouse or human prostate cancer interactomes and using in vivo drug perturbation data, which were described previously (Aytes et al., 2014). Target gene reversion (i.e., inhibition) was assessed using GSEA for each drug across each GEM model. GRSs for each drug were then inferred by integrating the reversion scores across each GEM model using a metric based on the Stouffer integration formulation (Whitlock, 2005). Optimal drug combinations were predicted from the single-agent in vivo drug perturbation data by determining the SRSs for each drug using an harmonic mean (F1 statistical measure), which maximizes the number of unique targets affected by each drug as well as the total number of targets affected by two drugs. GSRs were then estimated as an average SRS weighted by the number of mouse models in which a drug pair was estimated to be effective (i.e., to share a non-zero SRS). Details of the computational methods used to compute GRSs and GSRs are provided in Supplemental Experimental Procedures, and the data are provided in Tables S1 and S2.

### Efficacy of Drug Treatment

All experiments using animals were performed according to protocols approved by the Institutional Animal Care and Use Committee at Columbia University Medical Center. Cell culture studies were done as described previously (Aytes et al., 2014) using human prostate cancer cell lines obtained from ATCC and mouse cell lines derived from the GEM models used herein (Aytes et al., 2013; unpublished date). Rapamycin and docetaxel were purchased from LC Laboratories, and PD0325901 was provided by Pfizer. Cell culture assays were performed in a minimum of two independent experiments each done in triplicate; data are presented by the mean  $\pm$  SD. For in vivo studies, tumor-bearing NPK mice (Aytes et al., 2013) or allografted NPK tumors were treated with vehicle or rapamycin (10 mg/kg) and/or PD0325901 (10 mg/kg) or docetaxel (10 mg/kg) as described previously (Kinkade et al., 2008). At the time of sacrifice, tissues were collected for histopathological and molecular analysis as described elsewhere (Aytes et al., 2013; Kinkade et al., 2008). GraphPad Prism software (version 5.0) was used for statistical analyses and to generate data plots. A complete list of primers used in this study is provided in Table S5.

### Cross-Species Computational Analysis of Drug-Treatment Signatures

Gene expression profiles based on Illumina expression arrays as in Aytes et al. (2014) were used to generate drug-treatment signatures for the mouse tumors or allografts, as detailed in Table S3. For comparison of mouse treatment sig-

natures with human signatures, the mouse genes were mapped to their corresponding human orthologs. Single-sample computation of FOXM1/CENPF activity levels or drug treatment across human patients was inferred for each patient sample using single-sample MaRNA (ssMaRNA) and GSEA, respectively (Aytes et al., 2014) (see Supplemental Experimental Procedures). COX proportional hazard model and Kaplan-Meier analysis were done using the "surv" and "coxph" functions from the survcomp package in R v2.14.0.

## ACCESSION NUMBERS

The accession numbers for the gene expression profiling data reported in this study are GEO: GSE69211 and GEO: GSE69213.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and five tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2015.08.051>.

## AUTHOR CONTRIBUTIONS

A.M. performed all of the computational analyses, and A.A. performed all of the experimental analyses. M.Z. provided the new mouse cell lines described herein. A.M., A.A., M.M.S., C.A.-S., and A.C. designed experiments, analyzed the data, and wrote the paper.

## CONFLICTS OF INTEREST

A.C. is a founder and stockholder of Darwin Health and Therasis and a consultant for Dow AgroSciences, Thermo Fisher Scientific, and Cancer Genetics Inc.

## ACKNOWLEDGMENTS

This work was supported by grants CA173481 (to C.A.-S.), U01 CA084294 (to C.A.-S., M.M.S., and A.C.), U54 CA121852 (to A.C., C.A.-S., M.M.S.), P01 CA154293 (to M.M.S. and C.A.-S.), U01HL11566-02S2 (to A.C.) and U01CA168426 (to A.C.). A.A. was a recipient of a Marie Curie International Outgoing Fellowship (PIOF-GA-2009-253290), co-sponsored with the Catalan Institute of Oncology-Bellvitge Institute for Biomedical Research, Barcelona, Spain, and a recipient of a pilot award from the Irving Institute for Clinical and Translational Research at Columbia University supported by the National Center for Advancing Translational Sciences, NIH (UL1 TR000040). A.M. is a recipient of a Prostate Cancer Foundation Young Investigator Award. C.A.-S. is an American Cancer Society Research Professor supported in part by a generous gift from the F.M. Kirby Foundation.

Received: December 16, 2014

Revised: June 2, 2015

Accepted: August 17, 2015

Published: September 17, 2015

## REFERENCES

- Abate-Shen, C., and Pandolfi, P.P. (2013). Effective utilization and appropriate selection of genetically engineered mouse models for translational integration of mouse and human trials. *Cold Spring Harb. Protoc.* 2013, 2013.
- Al-Lazikani, B., Banerji, U., and Workman, P. (2012). Combinatorial drug therapy for cancer in the post-genomic era. *Nat. Biotechnol.* 30, 679–692.
- Aytes, A., Mitrofanova, A., Kinkade, C.W., Lefebvre, C., Lei, M., Phelan, V., Le-Kaye, H.C., Koutcher, J.A., Cardiff, R.D., Califano, A., et al. (2013). ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. *Proc. Natl. Acad. Sci. USA* 110, E3506–E3515.
- Aytes, A., Mitrofanova, A., Lefebvre, C., Alvarez, M.J., Castillo-Martin, M., Zheng, T., Eastham, J.A., Gopalan, A., Pienta, K.J., Shen, M.M., et al.

- (2014). Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* 25, 638–651.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607.
- Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H., et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 463, 318–325.
- Chang, A.J., Autio, K.A., Roach, M., 3rd, and Scher, H.I. (2014). High-risk prostate cancer-classification and therapy. *Nat. Rev. Clin. Oncol.* 11, 308–323.
- Chen, J.C., Alvarez, M.J., Talos, F., Dhruv, H., Rieckhof, G.E., Iyer, A., Diefes, K.L., Aldape, K., Berens, M., Shen, M.M., and Califano, A. (2014). Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. *Cell* 159, 402–414.
- Cooperberg, M.R., Moul, J.W., and Carroll, P.R. (2005). The changing face of prostate cancer. *J. Clin. Oncol.* 23, 8146–8151.
- Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483, 570–575.
- Garraway, L.A., and Lander, E.S. (2013). Lessons from the cancer genome. *Cell* 153, 17–37.
- Kinkade, C.W., Castillo-Martin, M., Puzio-Kuter, A., Yan, J., Foster, T.H., Gao, H., Sun, Y., Ouyang, X., Gerald, W.L., Cordon-Cardo, C., and Abate-Shen, C. (2008). Targeting AKT/mTOR and ERK MAPK signaling inhibits hormone-refractory prostate cancer in a preclinical mouse model. *J. Clin. Invest.* 118, 3051–3064.
- Lefebvre, C., Rajbhandari, P., Alvarez, M.J., Bandaru, P., Lim, W.K., Sato, M., Wang, K., Sumazin, P., Kustagi, M., Bisikirska, B.C., et al. (2010). A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* 6, 377.
- Luo, J., Solimini, N.L., and Elledge, S.J. (2009). Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell* 136, 823–837.
- Mukherji, D., Omlin, A., Pezaro, C., Shamseddine, A., and de Bono, J. (2014). Metastatic castration-resistant prostate cancer (CRPC): preclinical and clinical evidence for the sequential use of novel therapeutics. *Cancer Metastasis Rev.* 33, 555–566.
- Pienta, K.J., and Smith, D.C. (2005). Advances in prostate cancer chemotherapy: a new era begins. *CA Cancer J Clin* 55, 300–318, quiz 323–305.
- Politi, K., and Pao, W. (2011). How genetically engineered mouse tumor models provide insights into human cancers. *J. Clin. Oncol.* 29, 2273–2281.
- Rathkopf, D., and Scher, H.I. (2013). Androgen receptor antagonists in castration-resistant prostate cancer. *Cancer J.* 19, 43–49.
- Roychowdhury, S., and Chinnaiyan, A.M. (2013). Advancing precision medicine for prostate cancer through genomics. *J. Clin. Oncol.* 31, 1866–1873.
- Rubio-Perez, C., Tamborero, D., Schroeder, M.P., Antolin, A.A., Deu-Pons, J., Perez-Llamas, C., Mestres, J., Gonzalez-Perez, A., and Lopez-Bigas, N. (2015). In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. *Cancer Cell* 27, 382–396.
- Ryan, C.J., and Tindall, D.J. (2011). Androgen receptor rediscovered: the new biology and targeting the androgen receptor therapeutically. *J. Clin. Oncol.* 29, 3651–3658.
- Sboner, A., Demicheli, F., Calza, S., Pawitan, Y., Setlur, S.R., Hoshida, Y., Perner, S., Adami, H.O., Fall, K., Mucci, L.A., et al. (2010). Molecular sampling of prostate cancer: a dilemma for predicting disease progression. *BMC Med. Genomics* 3, 8.
- Scher, H.I., and Sawyers, C.L. (2005). Biology of progressive, castration-resistant prostate cancer: directed therapies targeting the androgen-receptor signaling axis. *J. Clin. Oncol.* 23, 8253–8261.
- Sharpless, N.E., and Depinho, R.A. (2006). The mighty mouse: genetically engineered mouse models in cancer drug development. *Nat. Rev. Drug Discov.* 5, 741–754.
- Shen, M.M., and Abate-Shen, C. (2010). Molecular genetics of prostate cancer: new prospects for old challenges. *Genes Dev.* 24, 1967–2000.
- Stanbrough, M., Bubley, G.J., Ross, K., Golub, T.R., Rubin, M.A., Penning, T.M., Febbo, P.G., and Balk, S.P. (2006). Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. *Cancer Res.* 66, 2815–2825.
- 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.
- Taylor, B.S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B.S., Arora, V.K., Kaushik, P., Cerami, E., Reva, B., et al. (2010). Integrative genomic profiling of human prostate cancer. *Cancer Cell* 18, 11–22.
- Wang, Z.A., Mitrofanova, A., Bergren, S.K., Abate-Shen, C., Cardiff, R.D., Califano, A., and Shen, M.M. (2013). Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity. *Nat. Cell Biol.* 15, 274–283.
- Whitlock, M.C. (2005). Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J. Evol. Biol.* 18, 1368–1373.
- Wong, Y.N., Ferraldeschi, R., Attard, G., and de Bono, J. (2014). Evolution of androgen receptor targeted therapy for advanced prostate cancer. *Nat. Rev. Clin. Oncol.* 11, 365–376.