

High-Throughput Architecture for Discovering Combination Cancer Therapeutics

Matt Gianni
 Yong Qin
 Geert Wenes
 Becca Bandstra
 Anthony P. Conley
 Vivek Subbiah
 Raya Leibowitz-Amit
 Suhendan Ekmekcioglu
 Elizabeth A. Grimm
 Jason Roszik

Author affiliations and support information (if applicable) appear at the end of this article.

Corresponding author: Jason Roszik, PhD, Departments of Melanoma Medical Oncology and Genomic Medicine, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX, 77030; e-mail: jason6219@gmail.com.

Creative Commons Attribution Non-Commercial No Derivatives 4.0 License.



abstract

Purpose The amount of available next-generation sequencing data of tumors, in combination with relevant molecular and clinical data, has significantly increased in the last decade and transformed translational cancer research. Even with the progress made through data-sharing initiatives, there is a clear unmet need for easily accessible analyses tools. These include capabilities to efficiently process large sequencing database projects to present them in a straightforward and accurate way. Another urgent challenge in cancer research is to identify more effective combination therapies.

Methods We have created a software architecture that allows the user to integrate and analyze large-scale sequencing, clinical, and other datasets for efficient prediction of potential combination drug targets. This architecture permits predictions for all genes pairs; however, Food and Drug Administration-approved agents are currently lacking for most of the identified gene targets.

Results By applying this approach, we performed a comprehensive study and analyzed all possible combination partners and identified potentially synergistic target pairs for 38 approved targets currently in clinical use. We further showed which genes could be synergistic prediction markers and potential targets with MAPK/ERK inhibitors for the treatment of melanoma. Moreover, we integrated a graph analytics technique in this architecture to identify pathways that could be targeted synergistically to enhance the efficacy of certain therapeutics in cancer.

Conclusion The architecture and the results presented provide a foundation for discovering effective combination therapeutics.

Clin Cancer Inform. © 2018 by American Society of Clinical Oncology

INTRODUCTION

Next-generation sequencing is becoming less expensive, which is promoting its application in routine genomic analysis of clinical samples. The enormous amount of data generated by gene sequencing in recent years, along with progress in related molecular network signaling and clinical information databases, is making it challenging to store, integrate, analyze, and interpret the relationships among such data. Big data initiatives, like the Genomic Data Commons,¹ ORIEN,² and GENIE,³ have been proposed and implemented to help store and analyze large datasets, and the analysis platforms of these databases are being developed. As of today, most of the available software is either web-based and limited by data transfer challenges, or is not easy to install and use, preventing scientists and clinicians from working directly with such databases and tools. Thus, there is a significant and unmet need for

applications that provide more sophisticated analyses of next-generation sequencing and related data in a high-throughput manner.

Because monotherapy seldom leads to cure in oncology, combination therapies are warranted. One of the major challenges in cancer drug discovery and development is to identify effective combination-therapy strategies. To predict drug response, multiple computational tools have been developed using genome-wide molecular data; however, predictions for drug combinations are not optimal or the number of cancers or drugs included is limited.⁴ Although several regimens including multiple approved clinical agents have been tested in clinical trials to identify synergistic activities of therapeutics, determining predictive biomarkers for establishing combination drugs remains a critically important obstacle.⁵ Current drug selection for combination therapies is often limited to drugs that have

recently achieved widespread clinical use. Furthermore, various noncancer drugs may have potential for the treatment of cancers, but those drugs are rarely tested in combination with cancer therapeutics. Current use of “big data” analytics methods to identify potentially synergistic targets for developing combinations is limited.

To address these issues, we leveraged a user-friendly distributed system architecture, Spark,⁶ that bundles a variety of tools and techniques originally developed for at-scale distributed processing, including meta-scheduling of multiple dependent applications for optimized queries and workflows, rapid in-memory processing for large data volumes, and graph processing. Spark also continues to provide support to several visualization and analysis tools and languages, such as the R programming language, that are frequently used in biomedical research. Spark is flexible with respect to data import/export and data management, and it also has tools for visualizing results, while making the work accessible between collaborators in an open and reproducible framework. To demonstrate the power of this approach, we also show how other databases, like Human Genome Organization (HUGO) and Reactome, can be included in our analyses.

In summary, we created a system that allows users to efficiently analyze large amounts of sequencing and clinical data from thousands of patients to identify combinational biomarkers to aid the discovery of potential targets for combination cancer therapeutics. We show how big databases can be integrated and visualized to allow quick selection of the most promising candidate targets. The source code of our application is publicly available on GitHub for use and further development.

METHODS

Spark

Initially developed by researchers at the University of California, Berkeley, Apache Spark⁶ is an open-source parallel processing framework for large-scale data analytics. Spark was originally designed to provide similar scalability and fault tolerance to Hadoop but with better memory use to deliver higher performance supporting iterative machine learning and interactive data analytics.⁷ Spark developers report that programs execute up to 100 times faster than Hadoop

MapReduce in memory or 10 times faster on disk. Spark’s key data abstraction, the resilience distributed dataset (RDD), insulates developers from complexity of distributed, in-memory analytics by providing an easy-to-use application programming interface delivered through a variety of language bindings, including Java, Scala, Python, and R.

Resource Description Framework /Triplestore

The Resource Description Framework (RDF) is a World Wide Web Consortium specification (<https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>) that was designed as a framework for representing information in the Web and it is commonly used in knowledge management applications. The fundamental unit of data in RDF is the triple (<https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/#dfn-rdf-triple>), consisting of three components: subject, predicate, and object. A triple can be conceptualized as an edge in a directed graph in which the subject and object identify the nodes, and the predicate represents a labeled edge connecting them. An RDF graph is a set of RDF triples. RDF’s graph abstraction provides a powerful framework for representing the relationships in highly connected datasets such as pathways databases.

An RDF store or triplestore is a database system specifically designed to store and explore RDF triples using semantic queries. Much like SQL (structured query language) for relational database systems, triplestore typically supports a standard query language called SPARQL. Like RDF itself, SPARQL was also standardized by the World Wide Web Consortium (<http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/>).

Cray Graph Engine

The Cray Graph Engine (CGE) is an RDF triplestore that leverages high-performance hardware and parallel software design expertise to accelerate graph analytics at scale. CGE provides semantic pattern matching and filtering coupled with custom graph algorithms and analysis tools that allow scale to graphs to billions of edges, while maintaining interactive query performance. We used CGE to identify and depict relevant pathway targets by organizing our results as a graph and labeling each of the

candidate gene pairs with their respective Reactome pathway categories. We were then able to query and visualize this network to illuminate how the candidate genes are connected in terms of their pathway categorizations.

We built our graph from our results and two publicly available datasets, HUGO and Reactome. The HUGO dataset allowed us to translate between The Cancer Genome Atlas (TCGA) gene symbols and Reactome's gene identifiers. We used European Bioinformatics Institute's Reactome for the pathway ontology, which allowed us to connect genes to their respective pathways.

We used CGE to build a single graph from all three data sources. Using CGE's SPARQL interface to query/update this integrated graph, we were able to label each of our candidate genes with its associated top-level Reactome pathway category. CGE also allowed us to export these networks and visualize them using standard graph visualization tools like Cytoscape.

Datasets

Gene expression and clinical data from 33 TCGA cancer projects were obtained from public TCGA repositories (<https://tcga-data.nci.nih.gov> and <http://gdac.broadinstitute.org/>). Cancer types, with TCGA abbreviations, and total input RNA-sequencing sample counts in parentheses, are as follows: adrenocortical carcinoma (ACC; n = 79); bladder urothelial carcinoma (BLCA; n = 427); breast invasive carcinoma (BRCA; n = 1212); cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC; n = 309); cholangiocarcinoma (CHOL; n = 45); colon adenocarcinoma (COAD; n = 191); lymphoid neoplasm diffuse large B-cell lymphoma (DLBC; n = 33); esophageal carcinoma (ESCA; n = 196); glioblastoma multiforme (GBM; n = 166); head and neck squamous cell carcinoma (HNSC; n = 566); kidney chromophobe (KICH; n = 91); kidney renal clear cell carcinoma (KIRC; n = 606); kidney renal papillary cell carcinoma (KIRP; n = 323); acute myeloid leukemia (LAML; n = 173); brain lower grade glioma (LGG; n = 530); liver hepatocellular carcinoma (LIHC; n = 423); lung adenocarcinoma (LUAD; n = 576); Lung squamous cell carcinoma (LUSC; n = 552); mesothelioma (MESO; n = 87); ovarian serous cystadenocarcinoma (OV; n = 307); pancreatic adenocarcinoma (PAAD; n = 183); pheochromocytoma

and paraganglioma (PCPG; n = 187); prostate adenocarcinoma (PRAD; n = 550); rectum adenocarcinoma (READ; n = 72); sarcoma (SARC; n = 265); skin cutaneous melanoma (SKCM; n = 473); stomach adenocarcinoma (STAD; n = 409); testicular germ cell tumors (TGCT; n = 156); thyroid carcinoma (THCA; n = 568); thymoma (THYM; n = 122); uterine corpus endometrial carcinoma (UCEC; n = 381); and uterine carcinosarcoma (UCS; n = 57); uveal melanoma (UVM; n = 80). TCGA normal tissues were not used in the survival analyses. To analyze normal-tissue gene expressions we used mRNA expression data from the Genotype-Tissue Expression project (GTEx, <http://www.gtexportal.org>).⁸ We used the transcripts per million⁹ unit to compare mRNA expressions from RNA sequencing. Cell-line expression and drug-sensitivity data were obtained from the website of the Cancer Cell Line Encyclopedia (CCLE) project.¹⁰

The pathway data from the Reactome dataset were obtained from European Bioinformatics Institute's RDF Platform (<https://www.ebi.ac.uk/rdf/services/reactome/>). An RDF version of HUGO was also used in the knowledge graph and was obtained from <https://biportal.bioontology.org/ontologies/HUGO>.

Survival Analyses

We used the "survival" R package (<https://cran.r-project.org/web/packages/survival/index.html>) to perform Kaplan-Meier analyses and calculate log-rank *P* values. To compare two groups, we used two-tailed Student *t* tests. Differences were considered significant when $P < .05$.

Description of Scala Classes

Algorithm 1: Calculate survival statistics. Input: survival times per patient, gene expression flags

1. Read survival data from distributed cache
2. Read incoming gene expression flags from standard input line by line (produced from algorithm 2)
 - a. Parse standard input to identify gene pairs, median expression values, and flags
 - b. Filter out gene pairs that do not have all four expression-flags groups represented
 - c. Calculate survival statistics using R survival package (<https://cran.r-project.org/web/packages/survival/citation.html>)

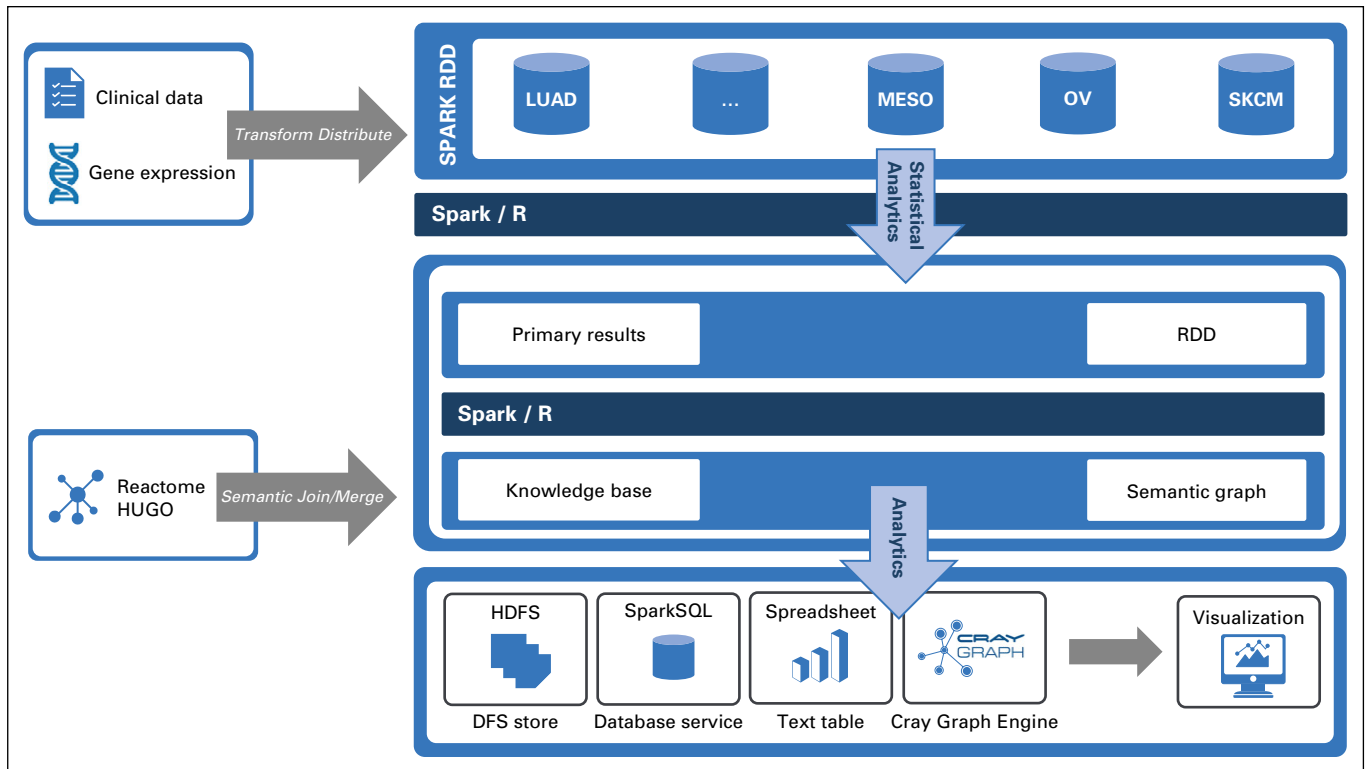


Fig 1. The high-throughput computing architecture. Clinical and gene expression data are distributed and permuted as RDDs and used to calculate the survival statistics (see algorithms 1 and 2 in the text) that make up the primary results. These primary survival results are then filtered and merged, using the CGE, into a knowledge graph built from publicly available databases representing known protein associations. This knowledge graph was used to explore and identify interesting relationships between genes and pathways that showed significant association with survivability. Tableau Desktop and Cytoscape were used to visualize these data and their relationships. CGE, Cray Graph Engine; HDFS, Hadoop Distributed File System; LUAD, lung adenocarcinoma; MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; RDD, resilient distributed dataset; SKCM, skin cutaneous melanoma.

d. Emit(Type, Gene₁, Gene₂, *p*-value, MedianExp₁, MedianExp₂, Quantile₁, Quantile₂, Quantile₃, Quantile₄)

Algorithm 2: Data preparation and permutation method. Input: patient survival and gene expression data file

Output: RDD_{flags}, survival data file

1. Extract *R*-Script from JAR file
2. Distribute *R*-Script to compute nodes
3. RDD_{raw} = Map(Expression Text File): Emit(Barcode, Status, Months, Expression₁, ..., Expression_g)
4. RDD_{surv} = Map(RDD_{raw}): Emit(Barcode, Status, Months)
5. Extract RDD_{surv} to the file system as survival data file
6. Distribute survival data file to all the compute nodes
7. RDD_{exp} = Map(RDD_{raw}): Emit(Expression₁, ..., Expression_g)
8. RDD_{exp} = Transpose(RDD_{exp})
9. RDD_{flags} = Map(RDD_{exp}):
 - a. For each Gene_g: Calculate MedianExp_g across all patients
 - b. Emit(Key: Gene_g, Value: list<character flag indicating whether the patient's expression value for Gene_g is greater than or equal to the MedianExp_g or less than this median>
10. RDD_{perm} = Cartesian(RDD_{flags}, RDD_{flags})
11. RDD_{cartesian} = RDD_{perm}.Filter(G₁ < G₂)
12. RDD_{prepared} = Map(RDD_{cartesian}): Emit(Gene₁, Gene₂, Median₁, Median₂, List<Flags>)
13. RDD_{result} = Map(RDD_{prepared}): Emit(Algorithm #1)

Building the Pair Hunter Program

The source code for the Pair Hunter program is available on GitHub at <https://github.com/jason1/pair-hunter>. Pair Hunter is built using sbt.¹¹ Executing the “sbt assembly” command builds the application. Testing whether the application build process completed successfully can be performed by executing the “run_test.sh” command. Before building the application, the following software needs to be installed: SBT (we used version 0.7.0), Apache Spark (we have tested 1.5.2 and 2.0.0), and R (we used version 3.1.0).

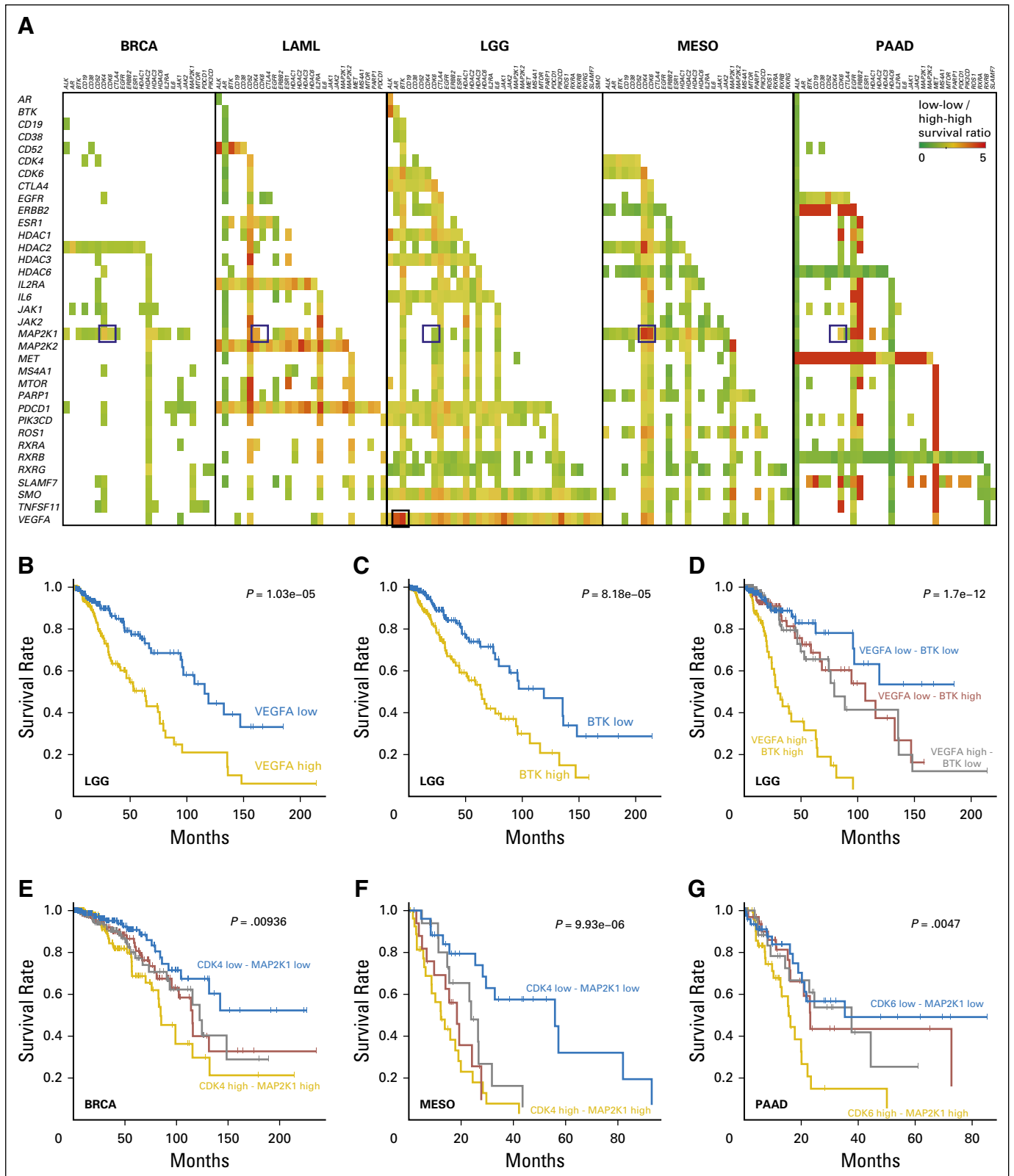


Fig 2. Combination predictions for US Food and Drug Administration-approved targeted therapeutics. (A) Drug-target gene and survival correlations are denoted by colored rectangles; the color represents the ratio of survival in the groups in which both targets are low (below median) or both are high (above median). The x- and y-axes show the same genes. Black rectangle in the LGG panel highlights the VEGFA–BTK combination, for which (B, C) individual and (D) combined survival analyses are shown. Blue rectangles denote the MAP2K1 (MEK1) and CDK4/6 combinations with Kaplan-Meier curves shown for (E) breast cancer, (F) mesothelioma, and (G) pancreatic cancer. BRCA, breast invasive carcinoma; LAML, acute myeloid leukemia; LGG, brain lower grade glioma; MESO, mesothelioma; PAAD, pancreatic adenocarcinoma.

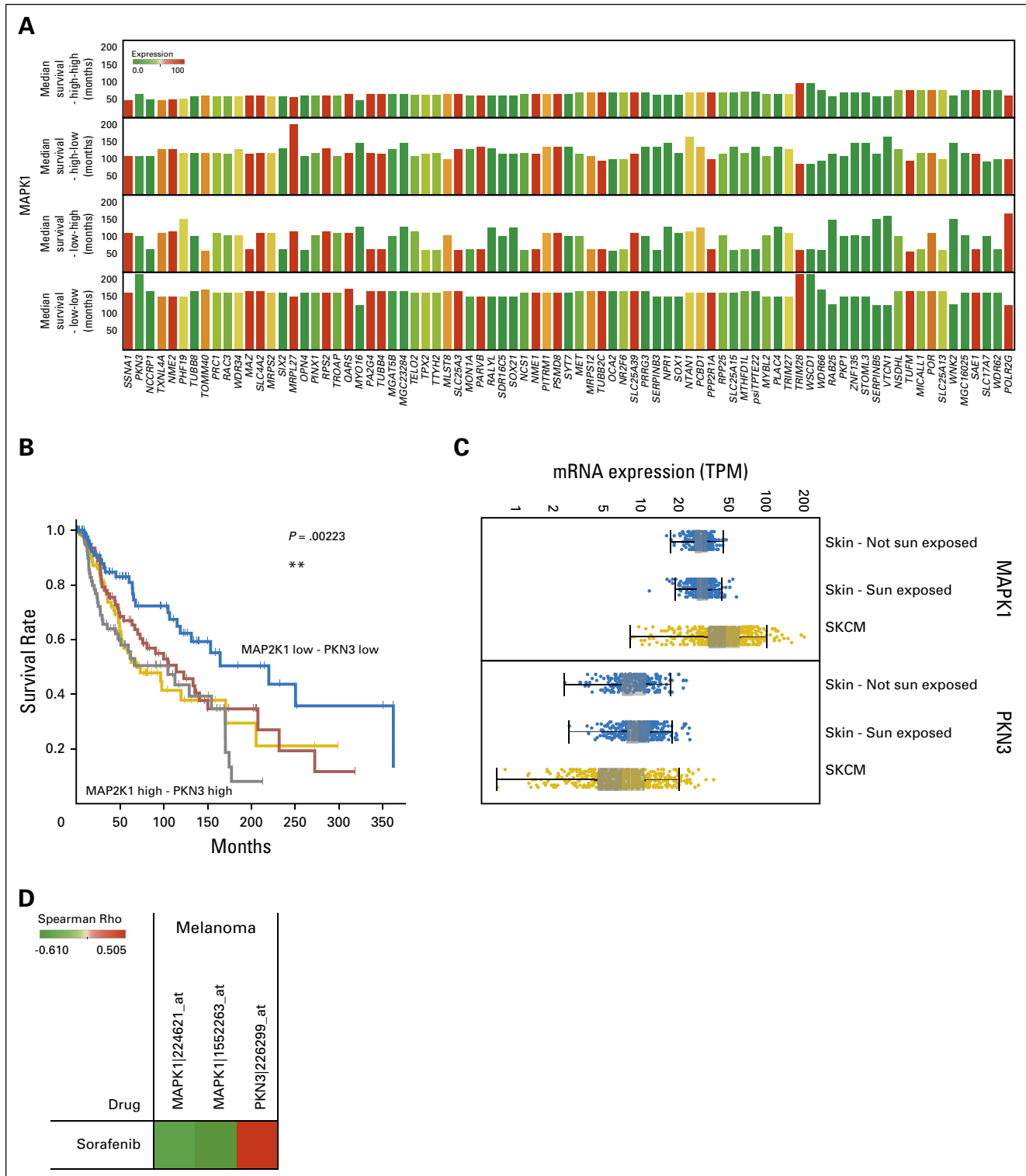


Fig 3. Predicted combinations to target the MAPK/ERK pathway in melanoma. Median survival in months is shown for the following four groups of patients with melanoma: expression of both genes are above the median expression levels (high-high group), the expression of one gene is above while the other is below median expression (high-low group), the expression of the first gene is below while the second is above median expression (low-high group), and the expression of both genes are below the median expression levels (low-low group). One of the combination partners is always MAPK1, and the other gene is shown on the x-axis. (A) Color represents gene expression. MAPK1–PKN3 survival analysis is shown as an example. (B) Blue: MAPK1 low, PKN3 low; red: MAPK1 low/PKN3 high; yellow: MAPK1 high/PKN3 low; gray: MAPK1 high/PKN3 high, together with (C) expression analysis of these genes, and (D) a correlation analysis of sorafenib concentration achieving half-maximal response and MAPK1 and PKN3 expression in the Cancer Cell Line Encyclopedia database. BRCA, breast invasive carcinoma; LGG, brain lower grade glioma; MESO, mesothelioma; PAAD, pancreatic adenocarcinoma; SKCM, skin cutaneous melanoma; TPM, transcripts per million.

RESULTS

Construction of the High-Throughput Computing Architecture

To analyze survival correlations of all possible combinations of all 20,531 genes with expression data in TCGA, we constructed a computational architecture (Fig 1) to efficiently process data from 10,395 samples of 33 cancer types in the TCGA. Importantly, this approach can be readily applied to any other datasets. Clinical and gene expression data were loaded into Spark RDD and overall survival (OS) and gene expression correlations were calculated for all possible gene pairs, using four patient groups: (1) the expression of both genes is above the median expression levels of each gene (high-high group), (2) the expression of the first gene is above whereas the second gene is below median expression (high-low group), (3) the expression of the first gene is below whereas the expression of the second is above median expression (low-high group), and (4) the expression of both genes is below the median expression levels (low-low group). This combination analysis shows that expression of two genes together shows a stronger association with prognosis and those genes may be good combinatory prediction makers for prognosis and may be tested as combination targets. Survival correlation results were integrated with the analyses of Reactome pathway and the HUGO databases. The results were visualized using the Tableau Desktop software¹² (Tableau Software, Seattle, WA) and the Cytoscape¹³ network analysis tool (<http://www.cytoscape.org/>).

Combination Predictions for Approved Targeted Therapeutics

To identify gene pairs that potentially synergistically affect patient survival, we used a list of 38 genes (*ALK, AR, BTK, CD19, CD38, CD52, CDK4, CDK6, CTLA4, EGFR, ERBB2, ESRI, HDAC1, HDAC2, HDAC3, HDAC6, IL2RA, IL6, JAK1, JAK2, MEK1[MAP2K1], MEK2[MAP2K2], MET, MS4A1, MTOR, PARP1, PDCD1, PIK3CD, ROS1, RXRA, RXRB, RXRG, SLAMF7, SMO, TNFSF11, VEGFA, VEGFR2, and VEGFR3*) that encode proteins targeted by therapeutics approved by the US Food and Drug Administration (FDA).¹⁴ Taking advantage of gene expression

(mRNA levels) and OS data from the TCGA, our analysis tool allowed us to systematically examine the correlation of expression levels of two specific genes with patient outcome. Figure 2A shows examples of predictions for all gene pairs of 36 of 38 FDA-approved targets that correlated significantly with patient survival for five selected cancers in the TCGA: BRCA, LAML, LGG, MESO, and PAAD. As expected, targets that correlated well with survival showed significant associations with multiple combination partners. Specifically, MAP2K1 in all five cancers; RXRB in PAAD; HDAC2 in BRCA, HDAC6 in MESO and PAAD; and VEGFA in LGG.

Kaplan-Meier analyses of a selected combination pair in LGG show significant survival correlation with VEGFA (Fig 2B) and BTK (Fig 2C) expression; however, the survival difference becomes very large when the expression of both these genes is low in comparison with when they are both high (Fig 2D), suggesting that these genes could be efficient targets in a combination therapy. The MAP2K1 (MEK1) and CDK4/6 combination, which has been shown to be synergistic¹⁵ and has been tested in the clinic in solid tumors (ClinicalTrials.gov identifier: NCT02065063), shows significant correlations in all five cancers (Fig 2A, blue boxes). As shown by survival curves for BRCA (Fig 2E), MESO (Fig 2F), and PAAD (Fig 2G), the OS of patients with high levels of MAP2K1 and CDK4/6 was significantly lower than the OS of patients with low levels of MAP2K1 and CDK4/6. These data further validate the rationale for combining drugs targeting MAP2K1 and CDK4/6 in various cancers.

New Targets to Combine With MAPK/ERK Pathway Inhibition to Enhance Survival of Patients With Melanoma

The components in the MAPK/ERK pathway are known important targets in various cancers, including cutaneous melanoma, and several therapies combining MAPK/ERK inhibition with other therapeutics are being tested in clinical trials for the treatment of advanced solid tumors.¹⁶ To identify novel combination therapies to enhance MAPK inhibition in cutaneous melanoma, we applied our tool to discover potential genes or targets that show strong correlation with patient survival when

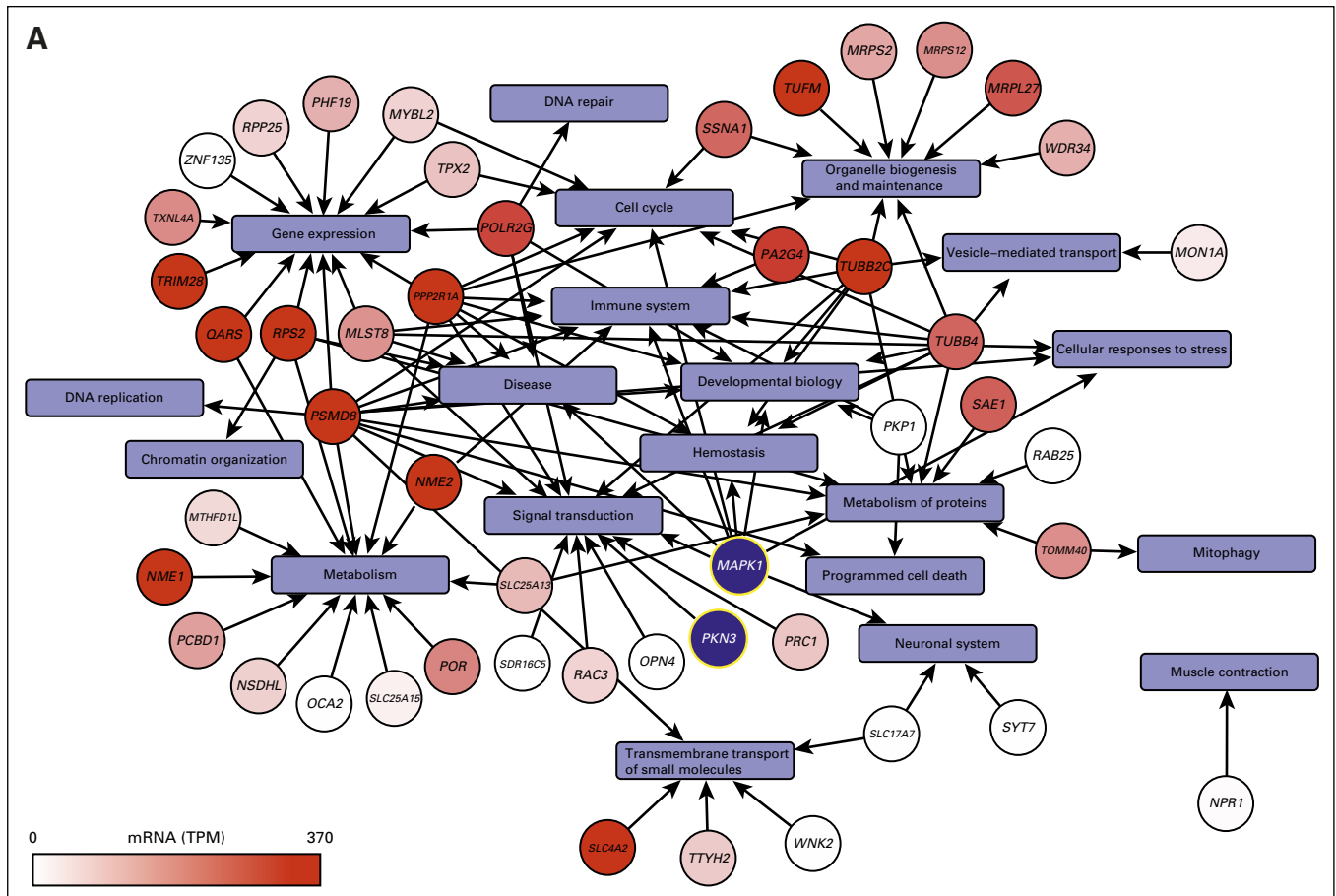


Fig 4. Pathway analysis of predicted potential targets to combine with MAPK/ERK inhibition, using graph analytics. Genes predicted to be potential MAPK1 combination partners are shown in circles; the red background color indicates median gene expression. MAPK1 and PKN3 are highlighted with blue. Pathways associated with these genes in the Reactome database are connected to the gene names and shown at (A) the top pathway level and also in the signal transduction pathway group at (B) the second level.

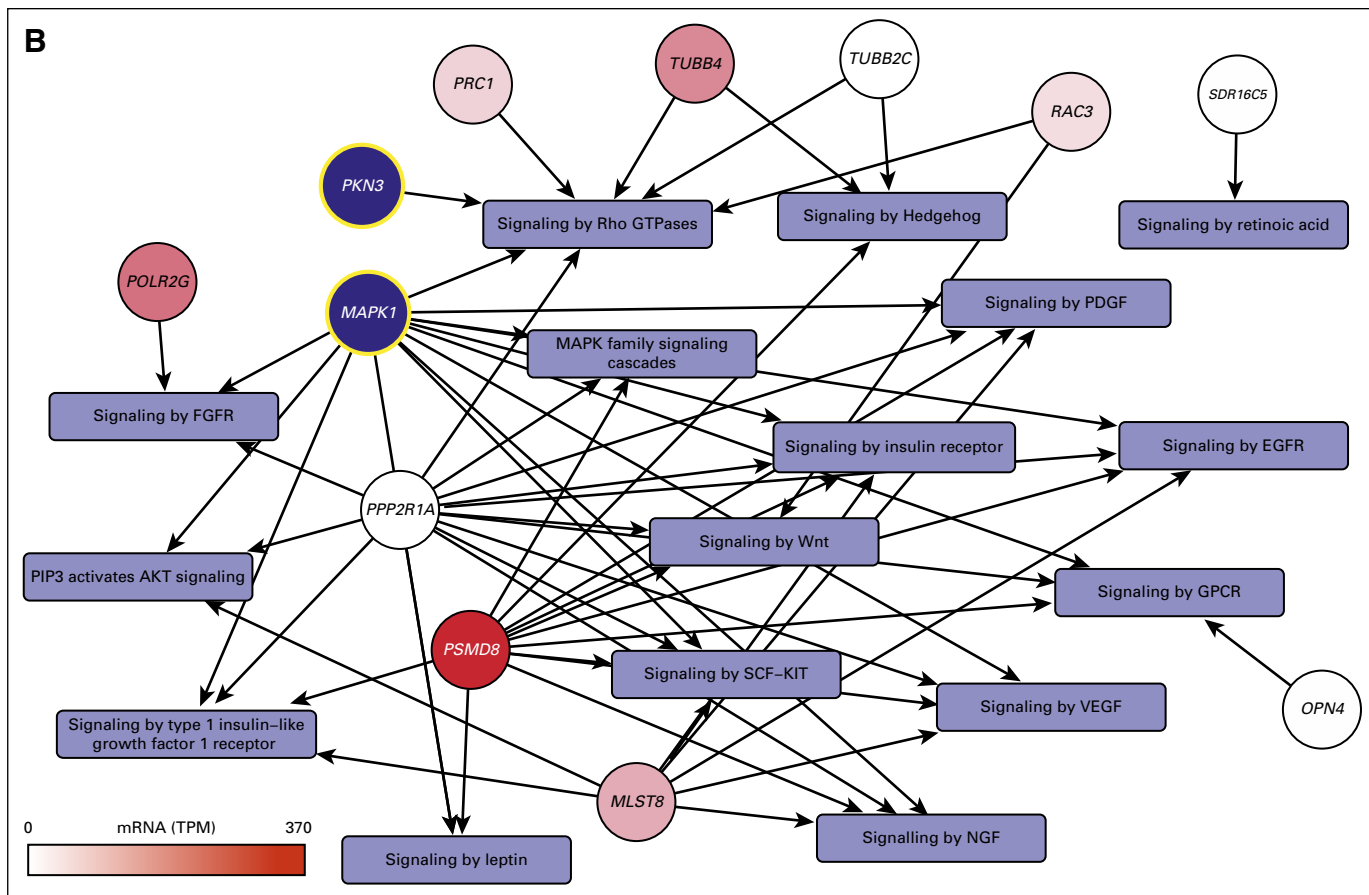
combining with MAPK1 inhibition. As shown in Figure 3, synergistic effect was predicted for several MAPK1 combinations evidenced by short OS in the high (ie, more than the median) MAPK1-high other gene-expression group and significantly better OS within the cohort in which expression of both of these genes is low (ie, below median).

As a detailed analysis example, we show that expression of protein kinase N3 (PKN3), which has been identified as a metastasis regulator in melanoma,¹⁷ is associated with OS when analyzed together with MAPK1 ($P < .01$; Fig 3B). MAPK1 is also overexpressed in melanoma ($P < .001$), and whereas PKN3 expression was lower ($P < .001$) in melanoma than in normal skin samples (which are not optimal controls for melanoma), it showed increased expression in sun-exposed skin compared with non-sun-exposed skin ($P < .05$; Fig 3C). To provide further validation of the MAPK1-PKN3 combination, we analyzed how expression of these genes correlates with drug response in the CCLE¹⁰ database, using the tool we recently developed.¹⁸ We found that high MAPK1 levels correlated with lower concentration

achieving half-maximal response of the sorafenib BRAF/MAPK inhibitor, as expected (Fig 3D). Interestingly, high PKN3 levels were associated with worse efficacy (ie, higher concentration achieving half-maximal response), suggesting that targeting or lowering of PKN3 may help increase efficacy of inhibitors of the MAPK pathway.

Pathway Analysis of Predicted MAPK/ERK Combination Partners Points to Targeting Metabolic Pathways

To show the power of our architecture in integration and visualization of results with pathway analysis tools, we converted our survival correlation results to RDF format and integrated them with the HUGO gene symbol and Reactome pathway databases. The predicted MAPK1 combination partners and associated Reactome pathways are shown in Figure 4. Importantly, we found that multiple potential combination targets are associated with the metabolism and metabolism of proteins pathways.



DISCUSSION

Integration and analysis of large amounts of information from next-generation genomics platforms present challenges. To be able to make a meaningful impact, we need computational tools that scientists and clinicians can use to analyze data in an efficient and straightforward way.¹⁹ Because the analytical load is very high, stand-alone analytical systems may provide advantages to the website-based approaches.¹⁹ The price of supercomputer systems is decreasing and they are becoming available to many laboratories and clinics; however, analyzing large sequencing datasets and integration of genomic data with other molecular and clinical information is challenging to researchers and clinicians with limited supercomputer and bioinformatics knowledge.

Recent advances in information technology help make precision medicine a reality through applying massive amounts of genetic, clinical, and other data.²⁰ These advances include the rapid development of next-generation sequencing technologies, including RNA sequencing,

which may provide information on expression of target genes and on associations with outcome and other clinical variables. Overexpression of the target on tumor cells is important for specificity and reduced toxicity in healthy tissues. Combining targeted and immune therapies is also a promising strategy in precision medicine and may help significantly improve outcomes for patients with cancer.²⁰ Multiple immunotargeted therapy combinations are being tested in the clinic; however, every possible combination of the currently available therapies is not feasible, and new approaches are needed to prioritize combination therapies for experimental testing.²¹ This study is intended to support prioritization and provide justification for further experimental and clinical testing. However, the results from this study should be interpreted with caution, because this study is not discovering combination cancer therapeutics. The study identifies associations of genes upregulated in a single tumor type with known combinations of FDA-approved agents shown to improve outcome. Furthermore, overexpression of the target gene is the basis for the association

with existing, FDA-approved agents that have been shown to increase outcome and other clinical variables. In solid tumors, many target genes associated with disease are mutated but not upregulated and, therefore, would not be detected through this method.

Furthermore, testing already approved noncancer drugs in combination with novel cancer treatments is also limited, although using these might provide the quickest way to the clinic. To address these challenges, we created a software architecture that allows the user to integrate and analyze large datasets, including sequencing, clinical, pathway, and other data to efficiently predict potential combination drug targets that can be tested in preclinical and clinical models. Our system provides efficient parallelization, quick analysis and visualization of large datasets, and inclusion opportunity for additional databases. To show the power of our architecture, we analyzed all possible combinations of FDA-approved cancer therapeutics and identified potentially synergistic target pairs for cancers in the TCGA project. Although in the current study we used data from samples in 33 TCGA cancer projects, cancer subtypes and more refined gene sets can similarly be studied using this approach.

To study novel targets, we have identified multiple targets that could potentially be synergistic with MAPK/ERK inhibitors for the treatment of skin cutaneous melanoma, for which sequencing data are available for many patients in the TCGA.²² Specifically, we have shown that targeting PKN3 may increase efficacy of MAPK/ERK inhibition. ERK inhibitors are being tested in the clinic for the treatment of tumors with aberrant MAPK pathway signaling¹⁶; however, combination therapies will probably be necessary to achieve durable tumor control. Furthermore, in the RAF-MEK-ERK path-

way, ERK inhibitors seem to be more effective than MEK inhibition at reducing MAPK activity and they are also superior at inhibiting the proliferation of BRAF inhibitor-resistant melanoma cells.²³ ERK inhibition may also be the best way to disrupt this pathway in other RAS-driven tumors.²⁴

Our results show that expression of multiple novel targets might synergistically affect patient survival, and inhibition of these genes may increase the efficacy of ERK inhibitors. We demonstrate that expression of MAPK1 and PKN3 both have to be low to predict better OS, and PKN3 inhibition may increase efficacy of the MAPK inhibitor sorafenib. Of note, it was recently shown that PKN3 is a key regulator of angiogenesis and metastasis in melanoma.¹⁷ This further supports that our predictions aid the discovery of potentially relevant combination targets. We have also shown that graph analytics and visualization can help identify relevant pathways to target. Specifically, we found that targeting metabolism pathways is of potential importance in skin cutaneous melanoma, supporting the current view that exploiting metabolic vulnerabilities may be helpful to overcome MAPK pathway inhibitor resistance.²⁵ Although these results are promising, we note that the potential beneficial effect of FDA-approved drug combinations identified by using this architecture remains theoretical. There is potential for serious, unforeseen adverse effects between previously untested drug combinations. In addition, drug combinations may lead to feedback effects on the pathways of interest or other pathways.

DOI: <https://doi.org/10.1200/CCI.17.00054>

Published online on ascopubs.org/journal/cci on January 11, 2018.

AUTHOR CONTRIBUTIONS

Conception and design: Matt Gianni, Becca Bandstra, Jason Roszik

Collection and assembly of data: Matt Gianni, Anthony P. Conley, Vivek Subbiah, Jason Roszik

Data analysis and interpretation: Matt Gianni, Yong Qin, Geert Wenes, Anthony P. Conley, Vivek Subbiah, Raya Leibowitz-Amit, Suhendan Ekmekcioglu, Elizabeth A. Grimm, Jason Roszik

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this

manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/site/ffc.

Matt Gianni

Employment: Sutter Health (I)

Yong Qin

No relationship to disclose

Geert Wenes

No relationship to disclose

Becca Bandstra

No relationship to disclose

Anthony P. Conley

Consulting or Advisory Role: Novartis, Nektar

Research Funding: MedImmune (Inst), Ignyta

Travel, Accommodations, Expenses: Nektar

Vivek Subbiah

Research Funding: Novartis, GlaxoSmithKline, Nanocarrier, NorthWest Biotherapeutics, Roche, BergPharma, Bayer, Incyte, FujiFilm, Pharmamar, D3, Pfizer Amgen, Abbvie, Multivir, Blueprint Medicines.

Travel, Accommodations, Expenses: Novartis, Pharmamar, Fujifilm

Raya Leibowitz-Amit

Honoraria: Bristol-Myers Squibb, Janssen Oncology, MSD

Consulting or Advisory Role: Sanofi

Travel, Accommodations, Expenses: Pfizer

Suhendan Ekmekcioglu

No relationship to disclose

Elizabeth A. Grimm

No relationship to disclose

Jason Roszik

No relationship to disclose

Affiliations

Matt Gianni, **Geert Wenes**, and **Becca Bandstra**, Cray Inc, Seattle, WA; **Yong Qin**, **Anthony P. Conley**, **Vivek Subbiah**, **Suhendan Ekmekcioglu**, **Elizabeth A. Grimm**, and **Jason Roszik**, The University of Texas MD Anderson Cancer Center, Houston, TX; and **Raya Leibowitz-Amit**, Tel Aviv University Sackler School of Medicine, Tel-Hashomer, Israel.

Support

Supported by philanthropic contributions to The University of Texas MD Anderson Moon Shots Program and by the Dr Miriam and Sheldon G. Adelson Medical Research Foundation. MD Anderson Cancer Center is supported by MD Anderson Cancer Center Support Grant P30 CA016672.

REFERENCES

1. National Cancer Institute: Genomic Data Commons. <https://gdc.cancer.gov>
2. Oncology Research Information Exchange Network: ORIEN homepage. <http://oriencancer.org/>
3. American Association for Cancer Research: Project GENIE. <http://www.aacr.org/Research/Research/Pages/aacr-project-genie.aspx>
4. Azuaje F: Computational models for predicting drug responses in cancer research. *Brief Bioinform* 18:820-829, 2017
5. Webster RM: Combination therapies in oncology. *Nat Rev Drug Discov* 15:81-82, 2016
6. Apache Software Foundation: Apache Spark. <http://spark.apache.org>
7. Zaharia M, Chowdhury M, Franklin MJ, et al: Spark: cluster computing with working sets. *Hot-Cloud* 10(10)http://static.usenix.org/legacy/events/hotcloud10/tech/full_papers/Zaharia.pdf
8. GTEx Consortium: The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45:580-585, 2013
9. Wagner GP, Kin K, Lynch VJ: Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* 131:281-285, 2012
10. Barretina J, Caponigro G, Stransky N, et al: The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483:603-607, 2012
11. Lightbend: The interactive build tool. <http://www.scala-sbt.org/index.html>
12. Tableau Software: Tableau Desktop. <http://www.tableau.com/products/desktop>
13. Shannon P, Markiel A, Ozier O, et al: Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498-2504, 2003
14. National Cancer Institute: Targeted cancer therapies. <https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies/targeted-therapies-fact-sheet>

15. Franco J, Witkiewicz AK, Knudsen ES: CDK4/6 inhibitors have potent activity in combination with pathway selective therapeutic agents in models of pancreatic cancer. *Oncotarget* 5:6512-6525, 2014
16. Infante JR, Janku F, Tolcher SW, et al: Dose escalation stage of a first-in-class phase I study of the novel oral ERK 1/2 kinase inhibitor BVD-523 (ulixertinib) in patients with advanced solid tumors. *J Clin Oncol* 33:2506, 2015 (15_suppl)
17. Mukai H, Muramatsu A, Mashud R, et al: PKN3 is the major regulator of angiogenesis and tumor metastasis in mice. *Sci Rep* 6:18979, 2016
18. Qin Y, Conley AP, Grimm EA, et al: A tool for discovering drug sensitivity and gene expression associations in cancer cells. *PLoS One* 12:e0176763, 2017
19. Hawkins RD, Hon GC, Ren B: Next-generation genomics: An integrative approach. *Nat Rev Genet* 11:476-486, 2010
20. Bertier G, Carrot-Zhang J, Ragoussis V, et al: Integrating precision cancer medicine into health-care-policy, practice, and research challenges. *Genome Med* 8:108, 2016
21. Morrissey KM, Yuraszeck TM, Li CC, et al: Immunotherapy and novel combinations in oncology: current landscape, challenges, and opportunities. *Clin Transl Sci* 9:89-104, 2016
22. Cancer Genome Atlas Network: Genomic classification of cutaneous melanoma. *Cell* 161:1681-1696, 2015
23. Carlino MS, Todd JR, Gowrishankar K, et al: Differential activity of MEK and ERK inhibitors in BRAF inhibitor resistant melanoma. *Mol Oncol* 8:544-554, 2014
24. Ryan MB, Der CJ, Wang-Gillam A, et al: Targeting RAS-mutant cancers: Is ERK the key? *Trends Cancer* 1:183-198, 2015
25. Smith LK, Rao AD, McArthur GA: Targeting metabolic reprogramming as a potential therapeutic strategy in melanoma. *Pharmacol Res* 107:42-47, 2016