Title

Development of a scalable in-vivo drug discovery platform allows for deep interrogation of mechanisms of KRAS.G12C inhibitors.

Author

Yu, John Xuzong

Publication Date

2022

Peer reviewed|Thesis/dissertation

A scalable drug discovery platform enables in-vivo dissection of drug mechanism of action and reveals novel insights into KRAS.G12C inhibitors.

by
John Yu

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Michael McManus
B47E685F33D1455...

Michael McManus
Chair

DocuSigned by:
LUKE GILBERT
DocuSigned by:
Hani Goodarzi
DocuSigned by:
A2E169ACDB154E2...

LUKE GILBERT

Hani Goodarzi

jonathan weissman

Committee Members

Acknowledgments

I want to thank Hani Goodarzi, Michael McManus, Luke Gilbert, Davide Ruggero, Jonathan Weissman, and Kevan Shokat for their generosity of time and willingness to help.

Thank you to Michael McManus for providing career advice at multiple stages during my PhD which helped me change course and drive towards an objective that was true to myself.

Thank you to Luke Gilbert for his hands-on generosity in helping with experiments and for his time and energy during collaborative projects.

Thank you to Davide Ruggero for welcoming me to UCSF and for teaching me how to write and think big about scientific problems.

Thank you to Jonathan Weissman for being a special mentor and more than generous host during Covid-19 at a time of significant personal challenges for myself and Katerina.

Thank you to Kevan Shokat for his interest, time, and faith in guiding my main project in graduate school from inception to completion; and for his pure scientific curiosity which consistently served as a personal source of inspiration.

Thank you especially to my thesis advisor, mentor, and friend Hani Goodarzi. His scientific guidance taught me how to conduct rigorous and risk-adjusted multi-disciplinary investigation. His personal support taught me how to go beyond that to live a life worthy of the scientific pursuit. During the darkest times of the PhD he was there in abundance and for that I will forever remember the depth of his dedication to his students.

Thank you especially to my unfailingly supportive wife Katerina who weathered the six years of PhD with me together through scientific, personal, national, and global turmoil. I look forward to many more years of growth together.

Thank you to the UCSF community and the Biomedical Sciences Program including Mark Ansel, Anita Sil, and Demian Sainz who created and nurtured myself and generations of young scientists to meet the challenges of the future. They are the invisible hand which enables opportunity and joy for all of us.

Contributions

This dissertation represents the scientific work I contributed to in inception and execution. **Chapter I** is an introduction into the current state of next-generation small molecule drug discovery methods. **Chapter II and III** of this dissertation comprises the development of my contribution to this field by invention of the GENEVA method and resulting discoveries within the space of KRAS.G12C inhibitors. **Chapter IV** describes a focused story on RBMS1 as a driver of metastatic gene networks.

Development of a scalable in-vivo drug discovery platform allows for deep interrogation of mechanisms of KRAS.G12C inhibitors.

Johnny Yu

Abstract

Drug discovery in small molecule oncology has traditionally been conducted using high-throughput, single-target screening approaches against a panel of arrayed test compounds. Since the inception of small-molecule screening around thirty years ago, the field has primarily progressed by automating this process using robotics or by moving to cell-based reporter assays. However, fundamental limitations exist namely: lack of in-vivo context of biochemical or reporter assays, lack of generalizability of screening discoveries across many different types of genetic backgrounds, and coarse data readouts that prevent simultaneous inference of deep biological interactions and mechanisms of compounds. The work herein describes in detail 1) the challenges and advances made in advancement of novel technologies aimed at addressing these problems as well as 2) the development of the GENEVA method as a method for simultaneously addressing these challenges in a multiplexed assay format scalable for in-vitro, organoid, and in-vivo model system 3) the discoveries enabled by GENEVA within the context of understanding KRAS.G12C inhibitor mechanisms.

In addition to a discussion of both methods and discoveries from GENEVA, I will also present a focused chapter on the dissection of RBMS1, an RNA binding protein that we go on to show rewires transcriptome networks to produce a pro-metastatic phenotype. I will describe the identification of this protein as a metastasis-relevant master regulator, validation of using in-vivo metastasis assays, and the downstream effects of shutting down RBMS1 on both specific transcripts and on gene networks broadly.

Table of Contents

List of Figures

Chapter 1: Advances in phenotype-based drug interrogation methods in the era of high-throughput genetic perturbation and sequencing methods

Drug discovery over the last few decades has been a challenging field with diminished clinical success rate, particularly in small molecule discovery. In oncology, only an estimated ~5% of compounds entering clinical trials ultimately reach approval (Wong et al. 2019). Failures can occur due to lack of understanding of mechanism of action (MOA) of the molecule, poor predictive power of preclinical models, or the lack of generalizability in broader patient populations (Fogel 2018, Nat Med 2010, Davis 2020, Pan et al. 2020, He et al. 2020). Furthermore, a persistent problem in oncology is the ability of cancers to evade the best targeted therapies over time by mechanisms involving adaptation, tumor heterogeneity, and novel mutations in escape genes (Vasan 2019). However, in recent years much progress has been made in development of more human-predictive model systems, targeting of drug-resistance tumor pathways, and biological dissection of molecular MOA.

Patient-derived Xenografts (PDXs) have seen an explosion of interest in recent years leading to now hundreds of available and established models in mice (Hidalgo et al 2014). PDX models offer the same histopathology in engrafted mice as in the original patient cancer representing a significant advantage over cell-line xenograft models (Dong et al 2016). In oncology, PDX models are viewed as an even higher level of 'gold-standard' validation compared to cell-line xenograft models. Furthermore, with increased availability PDX models have now been incorporated into large-scale screening panels to assess for generalizability of drug effect on a population-level scale (Ireson, 2019). Critically, PDX models remain inefficient to scale because of their more variable growth rates, requirements for in-vivo only propagation, and limited information content; predominantly the readout from PDX models is tumor size over time (Yoshida 2020, Letai 2021). Despite the promise of PDXs, ultimately cost and scalability hinder true discovery efforts using PDXs as the starting point.

Oncology also has the unique problem of drug resistance: despite the success of targeted therapies against mutational targets such as BRAF.V600E inhibitors, most targeted therapies inevitably develop resistance. Even among patients that respond to BRAF inhibitors ~50% of patients develop resistance within 6-8 months (Tanda et al 2020). As a response, combination therapies have been an active area of research but challenges remain; there are ultimately many routes to escape that exist and the large number of possible inhibitor combinations make testing these resource-intensive (Leary 2018).

The advent of high-throughput reverse genetics approaches and short-read sequencing methods have allowed for improved interrogation of mechanism of action (MOA). CRISPRi screens have shown that identification of key genetic dependencies to a molecule allow for the identification of pathways that inform MOA (Goncalves et al 2020). An illustration of this comes from a recent functional genomics study of Rigosertib (a molecule in Phase III trials) revealed the molecule had a vastly different MOA than previously known and, in fact, acted by targeting microtubule formation (Jost et al 2020). While it may seem surprising that a Phase III compound still had outstanding MOA questions, mechanism of action is often dirty and challenging to establish when the range of all possible biological perturbations are possible for any single moleculea and when polypharmacology comes into play. It may surprise the reader to learn that the commonly used home pain reliever Tylenol, or Paracetomol, has unclear and speculative MOA. For this reason, much of drug discovery is focused on interrogating MOAs and many compounds that proceed through phenotypic screening without a clear MOA can be discontinued as viable drug discovery programs. An alternative to perturbation-based interrogation, PRISM, which relies instead on the natural variation inherent in different cell lines has also been developed and been shown to inform genetic dependencies of small molecules as a way of understanding more deeply MOA (Yu 2016, Corsello 2020, McFarland 2020). The key difference here to genetic screens is that instead of perturbing genes and using these as

the source of experimental variation, endogenous variation assays rely on the inherent differences between many different cell lines as the source of controlled experimental variance. And while this method has been adapted to single-cell technologies as well (McFarland 2020), it is not clear just how powerful this type of assay informs MOA understanding beyond current methods and is also limited in its scope to short-term drug perturbations and strictly in-vitro based context.

In a landmark review from Stuart Shreiber, one of the pioneers of high-throughput screening and modern drug discovery, the authors illustrate different approaches including affinity-based, gene-expression, knockdown/overexpression screening, resistance screening, and molecule similarity clustering as novel methods that incorporate both phenotype and target identification in small molecule MOA discovery (Wagner & Shreiber 2016). Pros and cons exist for each type of method - for example, affinity-based capture is challenging and hard to scale for multiple molecules since direct capture assays require clever chemistry for target protein capture. Gene-expression based assays are often challenging to interpret because of the wide space of biological perturbations possible in a bulk cell population. Knockdown and overexpression genetic screens while valuable are labor intensive and suffer from scalability across many molecules. And finally, resistance screening is simply not scalable because of the length of time and many kinds of cell lines that can be profiled in a resistance assay.

Chapter I illustrates the advances that have been by incorporating novel methods developed in other spheres of biology and adapting them to phenotypic screening and MOA identification. However, even the most advanced methods illustrated here – genetic screens and natural variation cell pools – are limited in their scope. Genetic methods while useful, often provide an overabundance of information because of their ‘biological saturation’ approach to find all the perturbations that could go wrong. Scalability when it comes to compound libraries becomes a further challenge since screening is both time and scale intensive at a genome-wide level. Natural variation cell pools are a novel modality that allow for simultaneous profiling of

multiple genetic and phenotypic backgrounds. However, these methods have yet to demonstrate the power those genetic methods have in understanding first-order events in drug perturbations and lack the power of scaling to a different information space such as PDX, organoids, or in-vivo. Ultimately the information space targeted by these assays determine much of what can be discovered and it is possible that the intense focus of the last thirty years of drug discovery using in-vitro based systems may have led to the consistent reduction in success of clinical compounds which are ultimately tested in an in-vivo human context. Chapter II will discuss our approach to solving this problem and address specifically both model system limitations, scalability, and deep understanding of MOA and resistance mechanisms.

REFERENCES

Corsello, S. M., Nagari, R. T., Spangler, R. D., Rossen, J., Kocak, M., Bryan, J. G., Humeidi, R., Peck, D., Wu, X., Tang, A. A., Wang, V. M., Bender, S. A., Lemire, E., Narayan, R., Montgomery, P., Ben-David, U., Garvie, C. W., Chen, Y., Rees, M. G., … Golub, T. R. (2020). Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nature Cancer*, *1*(2), 235–248. https://doi.org/10.1038/s43018-019-0018-6

Davis, R. L. (2020). Mechanism of Action and Target Identification: A Matter of Timing in Drug Discovery. *IScience*, *23*(9), 101487. https://doi.org/10.1016/j.isci.2020.101487

Dong, R., Qiang, W., Guo, H., Xu, X., Kim, J. J., Mazar, A., Kong, B., & Wei, J.-J. (2016). Histologic and molecular analysis of patient derived xenografts of high-grade serous ovarian carcinoma. *Journal of Hematology & Oncology*, *9*, 92. https://doi.org/10.1186/s13045-016-0318-6

Fogel, D. B. (2018). Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemporary Clinical Trials Communications*, *11*, 156–164. https://doi.org/10.1016/j.conctc.2018.08.001

Goncalves. (2020). Drug mechanism-of-action discovery through the integration of pharmacological and CRISPR screens. *Molecular Systems Biology*, *16*(7), e9405. https://doi.org/10.15252/msb.20199405

He, Z., Tang, X., Yang, X., Guo, Y., George, T. J., Charness, N., Quan Hem, K. B., Hogan, W., & Bian, J. (2020). Clinical Trial Generalizability Assessment in the Big Data Era: A Review. *Clinical and Translational Science*, *13*(4), 675–684. https://doi.org/10.1111/cts.12764

Hidalgo, M., Amant, F., Biankin, A. V., Budinská, E., Byrne, A. T., Caldas, C., Clarke, R. B., de Jong, S., Jonkers, J., Mælandsmo, G. M., Roman-Roman, S., Seoane, J., Trusolino, L., & Villanueva, A. (2014). Patient Derived Xenograft Models: An Emerging Platform for Translational Cancer Research. *Cancer Discovery*, *4*(9), 998–1013. https://doi.org/10.1158/2159-8290.CD-14-0001

Ireson. (2019). *The role of mouse tumour models in the discovery and development of anticancer drugs | British Journal of Cancer*. https://www.nature.com/articles/s41416-019-0495-5

Jost, M., Chen, Y., Gilbert, L. A., Horlbeck, M. A., Krenning, L., Menchon, G., Rai, A., Cho, M. Y., Stern, J. J., Prota, A. E., Kampmann, M., Akhmanova, A., Steinmetz, M. O., Tanenbaum, M. E., & Weissman, J. S. (2020). Pharmaceutical-Grade Rigosertib Is a Microtubule-Destabilizing Agent. *Molecular Cell*, *79*(1), 191-198.e3. https://doi.org/10.1016/j.molcel.2020.06.008

Leary, M., Heerboth, S., Lapinska, K., & Sarkar, S. (2018). Sensitization of Drug Resistant Cancer Cells: A Matter of Combination Therapy. *Cancers*, *10*(12), 483. https://doi.org/10.3390/cancers10120483

Letai, A. (2021). *Functional precision oncology: Testing tumors with drugs to identify vulnerabilities and novel combinations: Cancer Cell*. https://www.cell.com/cancer-cell/fulltext/S1535-6108(21)00616-4

Mattar, M., McCarthy, C. R., Kulick, A. R., Qeriqi, B., Guzman, S., & de Stanchina, E. (2018). Establishing and Maintaining an Extensive Library of Patient-Derived Xenograft Models. *Frontiers in Oncology*, *8*, 19. https://doi.org/10.3389/fonc.2018.00019

McFarland, J. M., Paolella, B. R., Warren, A., Geiger-Schuller, K., Shibue, T., Rothberg, M., Kuksenko, O., Colgan, W. N., Jones, A., Chambers, E., Dionne, D., Bender, S., Wolpin, B. M., Ghandi, M., Tirosh, I., Rozenblatt-Rosen, O., Roth, J. A., Golub, T. R., Regev, A., … Tsherniak, A. (2020). Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. *Nature Communications*, *11*, 4296. https://doi.org/10.1038/s41467-020-17440-w

Mechanism matters. (2010). *Nature Medicine*, *16*(4), 347–347. https://doi.org/10.1038/nm0410-347

Pan, E., Bogumil, D., Cortessis, V., Yu, S., & Nieva, J. (2020). A Systematic Review of the Efficacy of Preclinical Models of Lung Cancer Drugs. *Frontiers in Oncology*, *10*, 591. https://doi.org/10.3389/fonc.2020.00591

Tanda, E. T., Vanni, I., Boutros, A., Andreotti, V., Bruno, W., Ghiorzo, P., & Spagnolo, F. (2020). Current State of Target Treatment in BRAF Mutated Melanoma. *Frontiers in Molecular Biosciences*, *7*, 154. https://doi.org/10.3389/fmolb.2020.00154

Vasan, N., Baselga, J., & Hyman, D. M. (2019). A view on drug resistance in cancer. *Nature*, *575*(7782), 299–309. https://doi.org/10.1038/s41586-019-1730-1

Wagner, B. K., & Schreiber, S. L. (2016). The Power of Sophisticated Phenotypic Screening and Modern Mechanism-of-Action Methods. Cell chemical biology, 23(1), 3–9. https://doi.org/10.1016/j.chembiol.2015.11.008

Wong, C. H., Siah, K. W., & Lo, A. W. (2019). Estimation of clinical trial success rates and related parameters. *Biostatistics*, *20*(2), 273–286. https://doi.org/10.1093/biostatistics/kxx069

Yoshida, G. J. (2020). Applications of patient-derived tumor xenograft models and tumor organoids. *Journal of Hematology & Oncology*, *13*(1), 4. https://doi.org/10.1186/s13045-019-0829-z

Yu, C., Mannan, A. M., Yvone, G. M., Ross, K. N., Zhang, Y.-L., Marton, M. A., Taylor, B. R., Crenshaw, A., Gould, J. Z., Tamayo, P., Weir, B. A., Tsherniak, A., Wong, B., Garraway, L. A., Shamji, A. F., Palmer, M. A., Foley, M. A., Winckler, W., Schreiber, S. L., … Golub, T. R. (2016). High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nature Biotechnology*, *34*(4), 419–423. https://doi.org/10.1038/nbt.3460

Chapter 2: Chapter 2: Development of GENEVA as a scalable in-vivo drug interrogation method and application in KRAS.G12C inhibitors

INTRODUCTION

Herein, we present a scalable method – GENEVA (Genetic and Natural Endogenous Variation Assay) – which allows for determination of MOA, resistance pathways, generalizability of drug effects across multiple genetic backgrounds, and the ability to gather this information from in-vivo and PDX cancer models in a single experiment. It builds upon the concept of using naturally occurring variation in pools of multiplexed cell lines with single-cell sequencing readouts, while significantly expanding the repertoire of available model systems (in-vivo and PDX) and utility (combination therapy and in-vivo specific insights). Using the KRAS.G12C covalent inhibitor ARS-1620 we show that GENEVA not only can discover genetic dependency of ARS-1620 on the KRAS.G12C mutation, but that the molecule acts to kill cells by inducing hyperactive mitochondria paired with increased oxygen consumption and increased mitochondrial reactive oxygen species production. We show that it combines the best features of both i) deep genetic perturbation approaches by discovery of highly complex mechanisms of action that are also orthogonally discovered in genetic systems and ii) natural variation assays using cell pools and elevate these mosaic technologies to highly challenging model systems such as PDX, PDO, and in-vivo. We further go to show that rapid combination therapy targets can be predicted, and that a specific in-vivo resistance mechanism of G12C inhibitor evasion can be enhanced by synergistic targeting in combination with TGF-beta inhibitor Galunisertib.

RESULTS

GENEVA as a platform for interrogation of genetic drivers of drug sensitivity

GENEVA aims to understand mechanism of action and tolerance of a drug compound in a single experiment. It relies on naturally existing genetic, transcriptomic, and phenotypic variation by pooling cell lines from different genetic backgrounds and then treating the pool with

the drug compound for an extended period of time. These pools of heterogeneous cell lines become the unit of experimentation and are implantable in different model systems: in vitro culture, 3D culture, and in vivo xenografts. Dissociation of cell pools post drug treatment is followed by single-cell RNA sequencing (scRNAseq), cell hashing and genetic demultiplexing. Using a multi-level computational demultiplexing pipeline, we categorize single cells by drug treatment, genetic background, and model system (Fig. 1a). Multiple lines of analysis can be performed from these single-experiment datasets to reveal how a compound behaves across multiple genetic backgrounds including reconstruction of relative sensitivities, determination of genetic drivers of sensitivity, and response mechanisms indicative of on-target mechanism of action and drug resistance mechanisms. By studying shared response across multiple genetic backgrounds cell-line specific artifacts are less likely to appear and generalizability of conclusions is increased.

We demonstrate that GENEVA can recapitulate discovery of genetic targets of targeted therapies. In a pool of cell lines containing both BRAF.V600E mutant and wild-type and KRAS.G12* mutant cell lines, we drugged cell pools with three compound conditions: DMSO (vehicle), Vemurafinib (BRAF.V600E targeting), and ARS-1620 (KRAS.G12C targeting) (Fig 1b). We calculated relative levels of sensitivity for different cell lines by counting cells assigned to each line in the vehicle and in the drug conditions. The most sensitive lines in Vemurafinib treated cells were BRAF.V600E mutant lines and most sensitive lines in ARS-1620 treated cells were KRAS.G12C mutant lines. Using a lasso variable selection regression model combining sensitivity output from GENEVA and whole exome derived mutation genotyping data, the model correctly predicted BRAF.V600E and KRAS.G12C as drivers of drug sensitivity to BRAF.V600E and ARS-1620 (Fig 1c). We further expanded GENEVA to conduct simultaneous IC50 estimation across multiple cell lines. Using ARS-1620 against a pool of G12C, G12D, and G12V lines we dosed pools at different concentrations and estimated relative survival using cell counts of each line at dose concentrations (Fig 1d). Reconstructed IC50 curves showed increased

sensitivity of G12C lines to ARS-1620, consistent with known mechanism of action of ARS-1620 targeting the KRAS.G12C mutation site (Fig 1e).

GENEVA allows for assessment *in vivo* cellular responses to drug and is applicable to PDX models

We developed GENEVA to also be a broadly flexible platform that allows for assaying chemical perturbations *in vivo* with the goal of understanding both in-vivo specific effects and patient-specific effects (Fig 2a). To demonstrate this, we multiplexed multiple KRAS G12* cell lines in a pooled flank xenograft to reconstruct drug response in an *in vivo* setting (Fig 2b). We proceeded to expand beyond cancer cell lines into patient-derived xenograft models (PDX) which are often considered high-fidelity models in oncology for their high-fidelity representation of patient tumors. We demonstrate that GENEVA can be conducted with pooled PDXs in flank xenograft *in vivo* systems (Fig 2d). Furthermore, PDX models grown as organoids in 3D culture can be pooled and deconvoluted (Fig 2e). Using cell cycle inference from the PDO GENEVA data, we estimated cell cycle suppression (G2+M/G1) for each genetic background and reconstructed a dose-response curve of cell cycle inhibition that implicates KRAS.G12C mutant PDX models as stronger responders to ARS-1620 than non-KRAS.G12C mutant models (Fig 2c). Importantly, GENEVA allows for reduction of the number of mice used per experiment to one or two per drug condition compared with hundreds in traditional mouse studies while also providing high-density phenotypic and transcriptomic readouts. Furthermore, GENEVA allows for the first time the ability to conduct these high-throughput natural variation experiments in an *in vivo* setting and with patient-relevant PDX models allowing for data collection that more accurately reflects the biology of tumors within the milieu of the human body.

ARS-1620 induces hyperactivity of mitochondrial electron transport chain flux and mitochondrial ROS production as a mechanism of cell death

GENEVA provides deep insights into mechanism of action of a drug by allowing for selection pressure on heterogeneous populations in response to long duration drug treatment.

GENEVA data from pools of KRAS.G12* mutant lines treated at multiple doses with ARS1620 in-vitro for a long duration (144h) (Fig 1d) revealed subpopulation structure across multiple cell lines indicating reduced mitochondrial-origin reads in cells surviving ARS-1620 treatment (Fig 3a). Deeper stratification of gene expression aggregated across all G12C lines indicated that specifically mitochondrial-origin, but not genomic-encoded mitochondrial targeted transcripts, were significantly downregulated after ARS-1620 treatment (Fig 3b). While genomic encoded mitochondrial transcripts were consistently unchanged across different lines, mitochondrial-encoded transcripts were specifically downregulated across G12C lines indicating a broadly generalizable mechanism of action (Fig 3c). We generated an ARS-1620 tolerant cell line derivative by treating KRAS.G12C mutant line H2030 with ARS-1620 (10 uM) for 30 days. Measurement of mitochondrial content using the mitochondrial dye Mitotracker Deep Red showed a decrease of mitochondrial content in H2030 Tolerant line compared to the drug-naïve cell line (Fig 3d). Taken together, GENEVA reveals mitochondrial-targeting action of ARS-1620 and experimental validation using a long-term drug tolerant line confirms that ARS1620 drug tolerant cells have reduced mitochondrial content.

We next investigated the mechanistic effects of KRAS.G12C inhibitors on mitochondrial dynamics. For these studies we turned to the more specific recently approved drug AMG510 because of its sub-micromolar IC50 and relevance to clinical outcomes. Using a Seahorse Respirometer, we measured oxygen consumption dynamics with and without AMG510 treatment at an acute two-hour timepoint. Strikingly, AMG510 induces a rapid and significant increase in basal respiration (Fig 4a). This effect is attributable to the increase in spare respiratory capacity at the site of the electron transport chain, with no significant change in glycolytic respiration or ATP production (Fig 4b,d). Interestingly, the increase in spare respiratory capacity attributable to AMG510 was absent in H2030 Persistor cells, possibly due to a compensation mechanism as part of a drug tolerance program (Fig 4c). We further characterized this mitochondrial hyperactivity phenotype by conducting a time-course study

measuring mitochondrial membrane potential, mitochondrial reactive oxygen species, and caspase cleavage using TMRE, mitoROS, Caspase 3/7 Detection Reagent respectively (Fig 4e). Initial response of cells to KRAS.G12C inhibition is an increase in membrane potential which is consistent with the observed acute upregulation in mitochondrial oxygen consumption. Following this at ~6 hours post-treatment mitochondrial reactive oxygen species increase followed by caspase 3/7 cleavage at ~24 hours post treatment. Membrane potential is tied directly to activity at the ETC and mitoROS is the direct output of ETC activity; caspase 3/7 cleavage is one of the last and irreversible steps in the apoptotic cascade. Therefore, our data suggests that AMG510 acts to kill cells via a cascade of hyperactive mitochondria triggering reactive oxygen species accumulation at the mitochondrial sites, followed by traditional executioner events such as caspase 3/7 cleavage.

To further validate that KRAS.G12C inhibitors act via creating hyperactive mitochondria, we used mitochondrial ETC inhibitor Antimycin to inhibit ECT activity in a co-dosing study with multiple KRAS.G12C inhibitors including ARS1620, AMG510, and MRTX849 (Fig 4g). The data shows a significant IC50 curve shift demonstrating that Antimycin rescues the lethality of KRAS.G12C inhibitors but only in KRAS.G12C lines (H1373, H2030) and not in non-KRAS.G12C lines (H441). This data confirms a novel mechanism of KRAS.G12C lethality via mitochondrial ETC that is specific to a KRAS.G12C context. We further analyzed a genome-wide CRISPRi screen designed to find synthetic lethal targets in a KRAS.G12C – ARS1620 context. We found that mitochondrial genes and specifically mitochondrial biogenesis genes (mitoribosomes) had a significant rescue effect on survival with mitochondrial biogenesis genes having the most pronounced rescue effect. This data confirms from an orthogonal reverse-genetics approach the mechanistic role that mitochondria play in ARS1620 lethality. Although ARS1620 is a highly specific covalent binder of the KRAS.G12C pocket, GENEVA reveals that its lethality mechanism relies on inducing hyper-respiratory mitochondria and that cells can evade lethality by downregulating their mitochondrial content.

ARS-1620 induces ferroptosis as a secondary mechanism of cell death

To determine mechanisms upregulated in response to KRAS.G12C inhibition we looked at upregulated genes across KRAS.G12C lines in the cell pool in cells surviving ARS1620 treatment. Strikingly, the most consistently upregulated genes across cell lines were involved in an anti-ferroptotic response mechanism (Fig 5a,b). Among this group of genes were FTH1 and FTL, the two components of the Ferritin Complex, which is responsible for sequestration of labile free iron. We hypothesized that because cells were upregulating an anti-ferroptotic response, ARS-1620 could be inducing ferroptosis as a secondary mechanism of cell death. Using a lipid peroxidation live cell probe we measured lipid peroxidation – one of the hallmark phenotypes of ferroptosis – in response to ARS-1620 treatment. A dose curve demonstrated that ARS-1620 induced lipid peroxidation in a dose dependent fashion, suggesting that ferroptosis was indeed an outcome of ARS-1620 treatment (Fig 5c). Furthermore, comparison of cell survival and lipid peroxidation dynamics with known inducers of ferroptosis (Erastin and Altretamine) showed that ARS-1620 behaves similarly in its ferroptotic/survival kinetics to these ferroptosis inducers (Fig 5d). Survival and normalized lipid peroxidation curves tended to crossover around the IC50 survival value in all three compounds demonstrating similar pharmacodynamics.

We then tested the effect of different KRAS.G12C inhibitors on ferroptosis in a KRAS.G12C (H2030) and non-KRAS.G12C (H441) cell line (Fig 5e). All G12C inhibitors significantly induced lipid peroxidation in G12C cell lines more than non-G12C cell lines indicating a specific and broadly consistent effect that is KRAS.G12C specific. We then attempted to determine if ferroptosis was a causal effect of cell lethality or a secondary mechanisms using the traditional ferroptosis rescue agent, ferrostatin-1 in co-dosing studies with multiple G12C inhibitors. We found that in ferrostatin-1 was unable to rescue cellular lethality with any of the G12C inhibitors, indicating that the ferroptosis phenotype is a secondary mechanism of action of the molecules. Lipid peroxidation has been known to lie downstream of

many cellular processes that produce reactive oxygen species and it appears that KRAS.G12C inhibitors by inducing the release of mitochondrial ROS, are able to trigger lipid peroxidation downstream as a secondary phenotype. GENEVA – in a single experiment – paints a complex mechanistic picture of the biological effects of a molecule and represents a focused and powerful method for rapidly assessing MOA.

GENEVA uncovers drug tolerance mechanisms to guide combination therapy targeting

We hypothesized that GENEVA could be further useful for understanding drug tolerance mechanisms as a way of identifying novel pathways to target for combination therapies and drug discovery because of its capacity for long drug exposure testing. From GENEVA data generated on ARS-1620, we aggregated gene expression changes in response to drug and compared differences between G12C and non-G12C cell lines. We identified several druggable targets that were upregulated in response to KRAS.G12C inhibition specifically in G12C lines but not in G12* lines representing specific on-target resistance mechanisms (Fig 6a,b). Notably, VEGF, JAK and mTOR pathways were implicated as conferring resistance to KRAS.G12C inhibition as previously reported. To test if these druggable resistance pathways would be effective combination therapy targets we co-dosed these molecules in survival assays with the KRAS.G12C inhibitors ARS-1620, AMG510, and MRTX849. Notably, several of these compounds showed strong Bliss drug synergy scores with INK128 – an mTOR inhibitor – as the strongest candidate which showed strong synergy with all three KRAS.G12C inhibitors (Fig 6c). A deeper analysis of the mTOR regulated geneset of ribosomal proteins showed that in fact it was strongly upregulated even in non-KRAS.G12C lines (Fig 6d). Therefore, we opted to conduct an *in vivo* combination therapy study with ARS1620 and INK128 to assay for in-vivo drug synergy using the KRAS.G12C cell line xenograft H1373. We found that INK128 was able to significantly reduce tumor growth *in vivo* (Fig 6f). Furthermore, calculation of the expected additive effect of the two compounds using a Bliss Independence model showed that the combination was able to reduce growth ten-fold compared to the expected growth rate if the

drugs were independent (Fig 6e). GENEVA was able to rapidly identify several targetable growth pathways expressed in drug-tolerant cells and treatment of these pathways demonstrated suppression of tumor growth both *in vivo* and *in vitro* models.

GENEVA *in vivo* reveals EMT as a potent adaptive mechanism targetable by combination therapy

We compared the ARS1620 datasets conducted *in vitro* against the data generated *in vivo* to look for differences and similarities attributable to the context of the model systems used (datasets from Fig 1d, 2b). We found that there were significant similarities across geneset responses to drug in both contexts including a downregulation of mitochondrially localized genes post-treatment (Fig 7b). However, one of the most upregulated genesets in response to drug was specific to the *in vivo* context and showed no difference *in vitro* (Fig 7a). This endothelial-mesenchymal transition hallmark gene expression signature was increased *in vivo* and represented a possible drug-adaptive mechanism to ARS1620 KRAS.G12C inhibition. Therefore, we designed a multi-arm combination therapy *in vivo* study to test the efficacy of an EMT/TGF-B inhibitor, Galunisertib, in combination with ARS1620. We found that the combination therapy was extremely effective in suppressing tumor growth (Fig 7c) and acted with ARS1620 to reduce growth synergistically according to a Bliss synergy model (Fig 7d). Further comparison of the effect size of *in vivo* synergy of Galunerstib+ARS1620 with INK128+ARS1620 showed similar growth rate reductions of ten-fold compared to the expected no additive model for each combination therapy (Fig 7e). Thus, this *in vivo* dataset was able to provide an insight into a resistance mechanism preferentially occurring *in vivo* and led to identification of a targetable pathway for combination therapy.

Analysis of the PDX GENEVA data showed that after drug treatment, cell populations changed to reflect an increase in EMT-high expressing cells in JAX877 (KRAS.G12C) and Ribosomal-high expressing cells in JAX233 (KRAS.G12C) (Fig 7f). This data suggests that one PDX is undergoing selective drug pressure to adapt by upregulating a specific EMT mechanism

of drug tolerance while the other PDX is adapting by upregulating ribosomal production as part of an mTOR signaling response, illustrating how GENEVA can be used to find patient-specific responses in a rapid fashion. In a different line of analysis, CLX data *in vivo* was compared to CLX data *in vitro* by calculating spearman correlation using drug-induced transcriptome responses. The data revealed that the *in vivo* dose (150 mg/kg ARS1620, daily gavage) was most correlated to a high *in vitro* dosage of 25 uM (Fig 7G). By comparing both model systems we were able to estimate relative pharmacodynamics of the drug compound across both systems, leading to faster estimation of bioavailability.

GENEVA as a method for combination therapy dissection *in vivo*

In addition to providing insights into resistance mechanisms to anti-cancer agents, we showed that GENEVA can also be used to conduct pooled *in vivo* combination therapy studies. We utilized a panel of G12C and non G12C lines pooled *in vivo* as xenografts to conduct combination therapy studies comparing ARS1620 with and without Galunisertib, INK128, and Antimycin for a total of eight different treatment conditions; these compounds were taken from our prior combination therapy discoveries and mechanistic understanding of mitochondrial action of ARS1620 (Fig 8a,b). Using cell cycle active/inactive proportions in each drug condition we calculated drug synergy across several cell lines for each drug. Ink128 and Galunisertib showed relative *in vivo* synergy consistent with the prior (Fig 8c) validation experiments demonstrating *in vivo* synergy (Fig 6e, Fig 7e), while antimycin showed an antagonistic effect demonstrating a relative rescue of ARS1620 consistent with prior rescue data (Fig 4e). A linear model built around gene expression and drug treatment + cell line of origin to discover genes that drove synergistic drug phenotype revealed that together Galunisertib and INK128 was able to further increase the synergistic decrease in mitochondrial reads consistent with ARS1620 general effect on mitochondria observed alone (Fig 8d). Subpopulation structure showed that indeed the drugs together were able to enhance the effect on reduction of mitochondrial gene expression, while antimycin was able to reverse the decrease of mitochondrial expression

attributable to ARS1620 (8e). Taken together, GENEVA is also useful as a scalable combination therapy tool that allows for assaying multiple cell lines together, conducting multiple drug treatment conditions quickly, deriving phenotypic synergy information from cell cycle state, while also allowing for gene-level inference of pathways that drive synergy or antagonism.

FIGURES

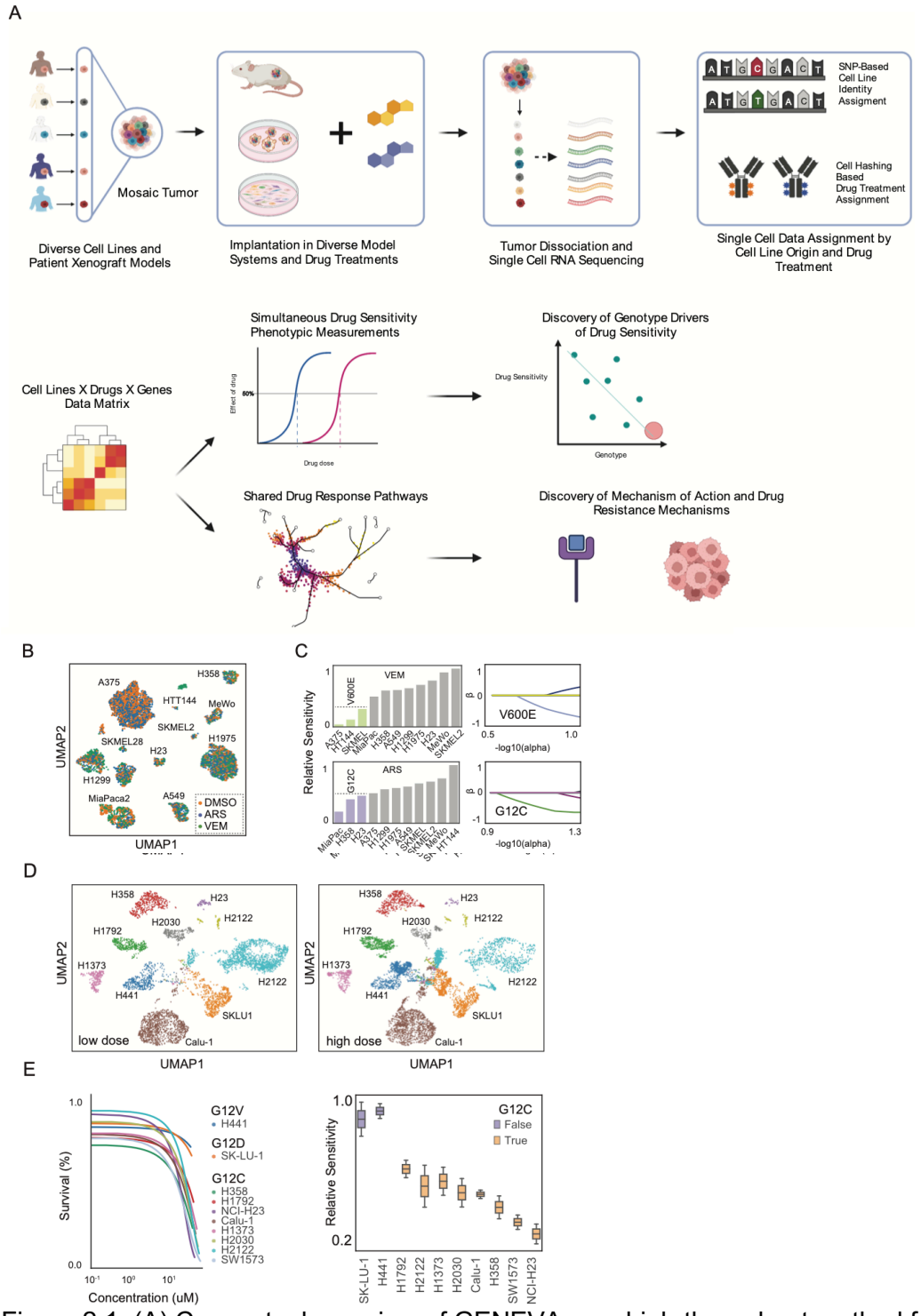


Figure 2.1. (A) Conceptual overview of GENEVA as a high-throughput method for understanding mechanisms of drug efficacy and escape using a mosaic tumor approach in diverse model systems with (B) an accurate pipeline for pooling and computational reassignment across drug and cell origin conditions in a 2D in vitro context with ARS1620 and Vemurafinib. (C) Analysis of drug sensitivity estimation using cell counts allows for identification of mutational targets of compounds using a feature selection regression model. (D) Multiple

dosing regimens of ARS1620 in a KRAS.G12C vs KRAS.G12* pool allows for (E) reconstruction of IC50 and relative drug sensitivity measure- ments recapitulating discovery of KRAS.G12C as the sensitizing mutational target for ARS1620.

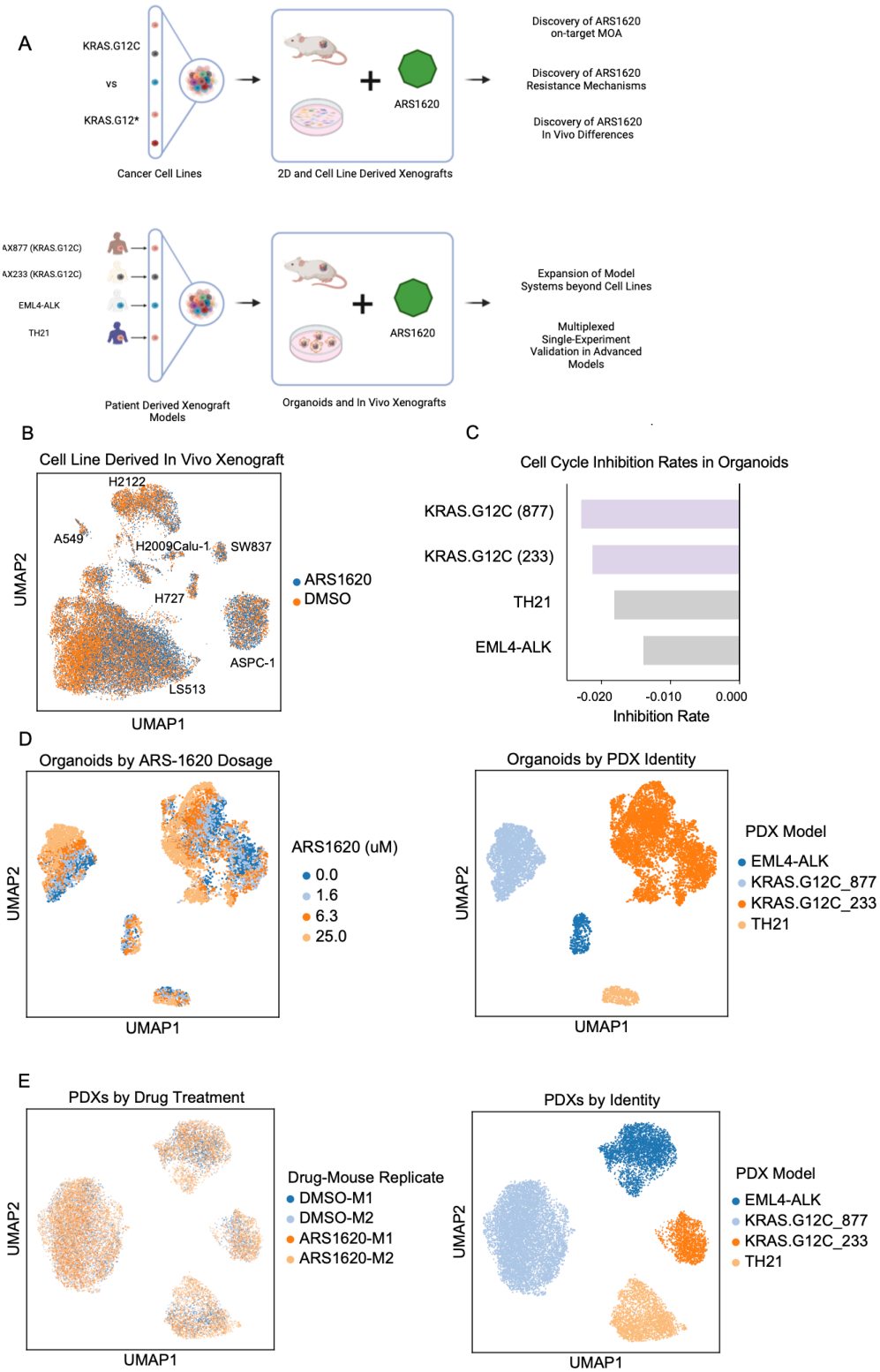


Fig 2.2 GENEVA can be conducted in-vivo and in organoids across PDX and CLX models to discover in-vivo specific drug response differences as well as expanding experimental drug testing in pools across multiple PDX/PDO models. (A) Applying GENEVA to In Vivo systems to investigate KRAS.G12C Inhibitor ARS1620 using both Cell line and Patient-Derived Xenograft Models (B) UMAP demonstrating successful treatment, implantation, harvest, and pooling of in

vivo mosaic tumors and computational deconvolution assignment for cell line in vivo xenografts in NSG mice (C) Cell cycle inhibition rate calculated from cell cycle active/inactive states in PDX organoid models showed specificity of KRAS.G12C inhibitor ARS1620 against G12C driven tumors (D) Mosaic organoids were treated across multiple doses of ARS1620 (E) Patient Derived Xenografts were successfully applied in an in vivo mosaic GENEVA experiment.

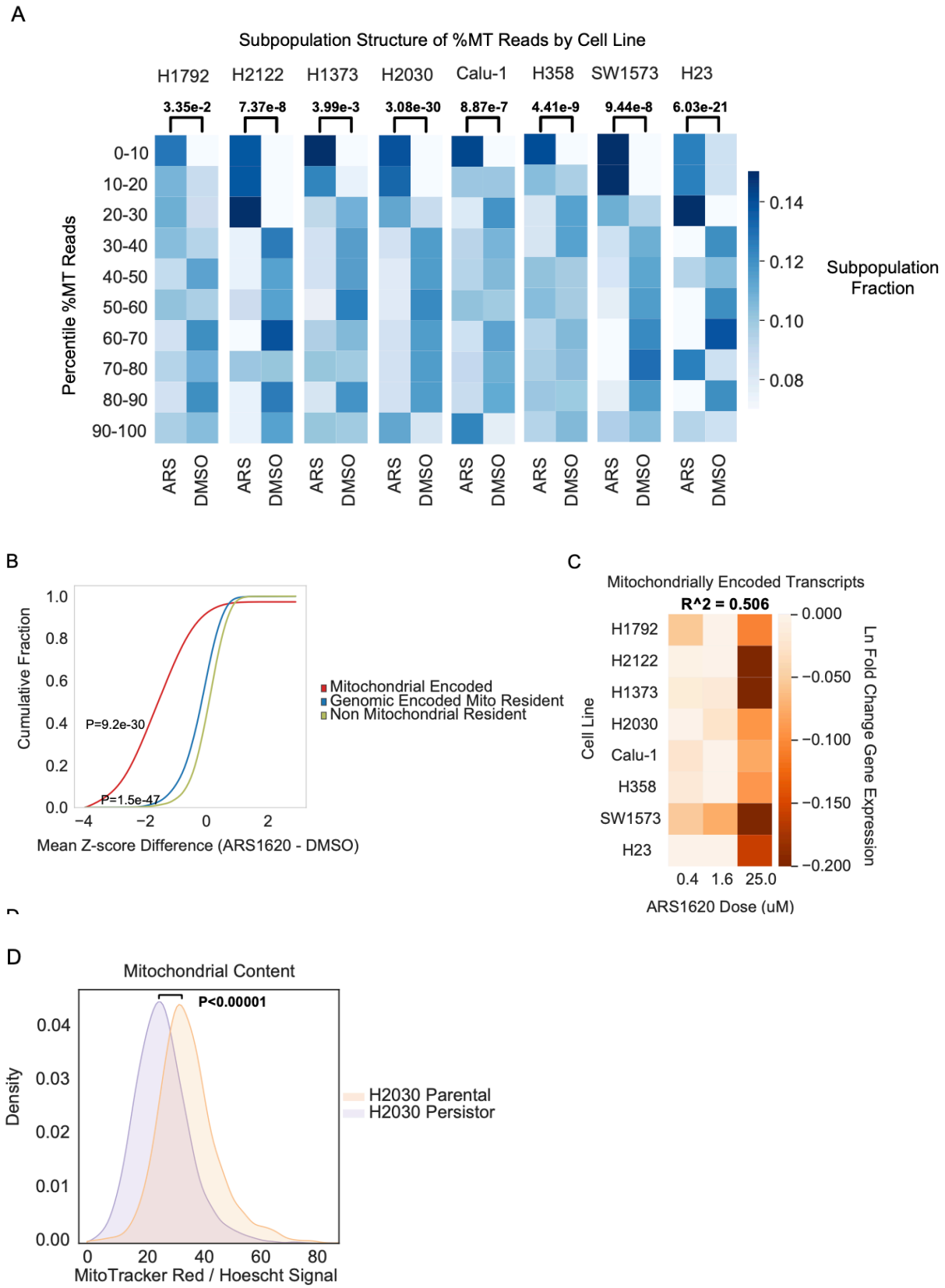
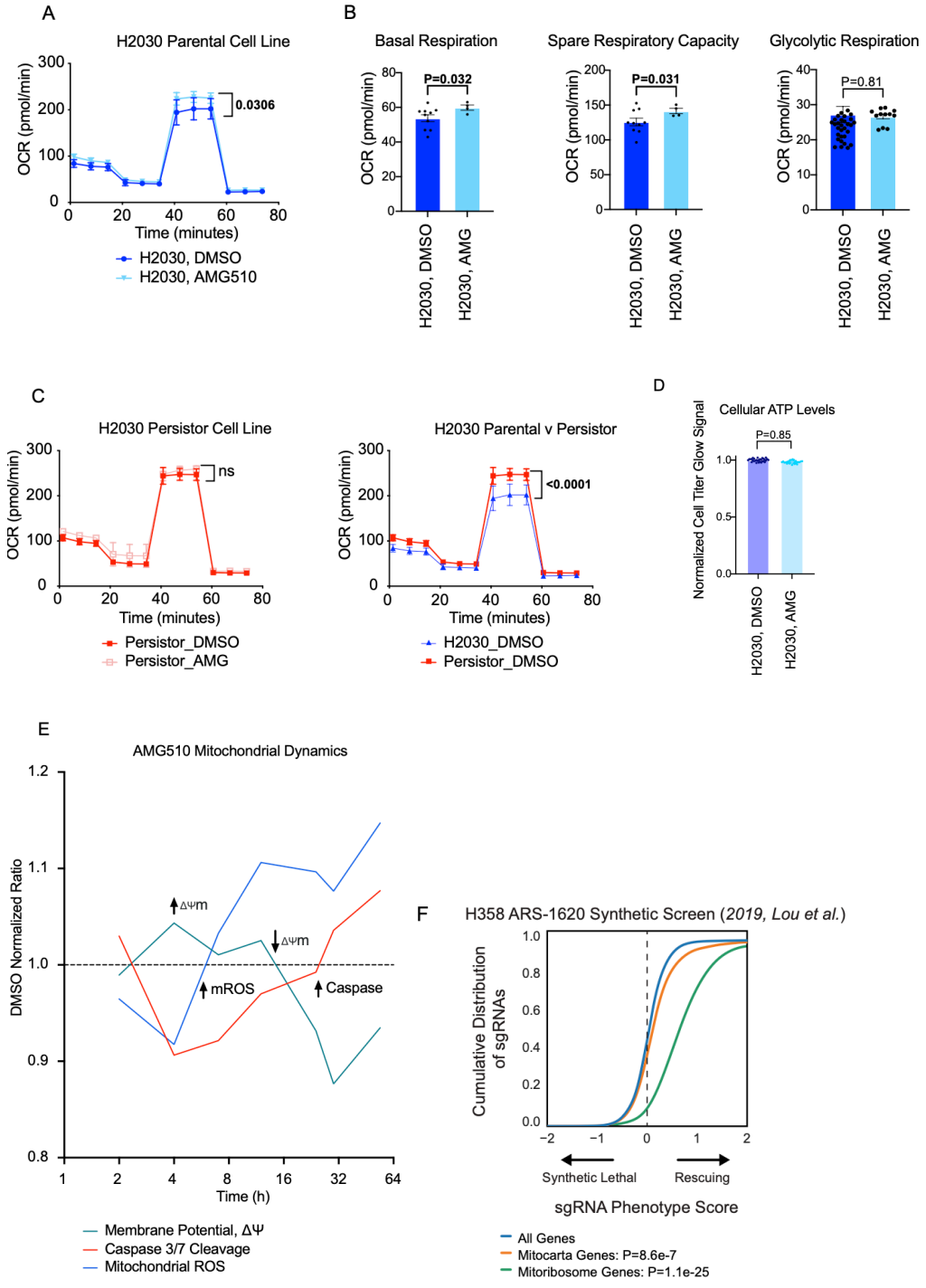

A
Subpopulation Structure of %MT Reads by Cell Line
H1792
H2122
H1373
H2030
Calu-1
H358
SW1573
H23
3.35e-2
7.37e-8
3.99e-3
3.08e-30
8.87e-7
4.41e-9
9.44e-8
6.03e-21
Percentile %MT Reads
0-10
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
ARS
DMSO
0.14
0.12
0.10
0.08
Subpopulation Fraction
B
Cumulative Fraction
1.0
0.8
0.6
0.4
0.2
0.0
P=9.2e-30
P=1.5e-47
−4
−2
0
2
Mean Z-score Difference (ARS1620 - DMSO)
Mitochondrial Encoded
Genomic Encoded Mito Resident
Non Mitochondrial Resident
C
Mitochondrially Encoded Transcripts
R^2 = 0.506
H1792
H2122
H1373
H2030
Calu-1
H358
SW1573
H23
Cell Line
0.000
−0.050
−0.100
−0.150
−0.200
Ln Fold Change Gene Expression
0.4
1.6
25.0
ARS1620 Dose (uM)
D
Mitochondrial Content
P<0.00001
Density
0.04
0.03
0.02
0.01
0.00
0
20
40
60
80
MitoTracker Red / Hoescht Signal
H2030 Parental
H2030 Persistor

Fig 2.3. (A) Single cells from GENEVA pools in-vitro treated with ARS1620 binned by percent mitochondrial reads demonstrate subpopulation structure associated with drug-tolerant cells. (B) Aggregated gene expression across KRAS.G12C lines in GENEVA pools of mitochondrially encoded and genomically encoded mitochondrially-targeted transcripts compared to gene expression of non-mitochondrial gene transcripts after ARS1620 treatment. (C) Gene expression of mito-encoded transcripts after ARS1620 treatment for each KRAS.G12C cell line. (D) Generation of a long term ARS1620 tolerant cell line (30 day treatment, 10uM) from H2030 (KRAS.G12C) profiled for mitochondrial content using fluorescent mitochondrial stain (Mitotracker Deep Red FM).


A
H2030 Parental Cell Line
OCR (pmol/min)
Time (minutes)
0.0306
H2030, DMSO
H2030, AMG510
B
Basal Respiration
Spare Respiratory Capacity
Glycolytic Respiration
P=0.032
P=0.031
P=0.81
H2030, DMSO
H2030, AMG
C
H2030 Persistor Cell Line
ns
Persistor_DMSO
Persistor_AMG
H2030 Parental v Persistor
<0.0001
H2030_DMSO
Persistor_DMSO
D
Cellular ATP Levels
Normalized Cell Titer Glow Signal
P=0.85
E
AMG510 Mitochondrial Dynamics
DMSO Normalized Ratio
ΔΨm
mROS
Caspase
Time (h)
Membrane Potential, ΔΨ
Caspase 3/7 Cleavage
Mitochondrial ROS
F
H358 ARS-1620 Synthetic Screen (2019, Lou et al.)
Cumulative Distribution of sgRNAs
Synthetic Lethal
Rescuing
sgRNA Phenotype Score
All Genes
Mitocarta Genes: P=8.6e-7
Mitoribosome Genes: P=1.1e-25

G

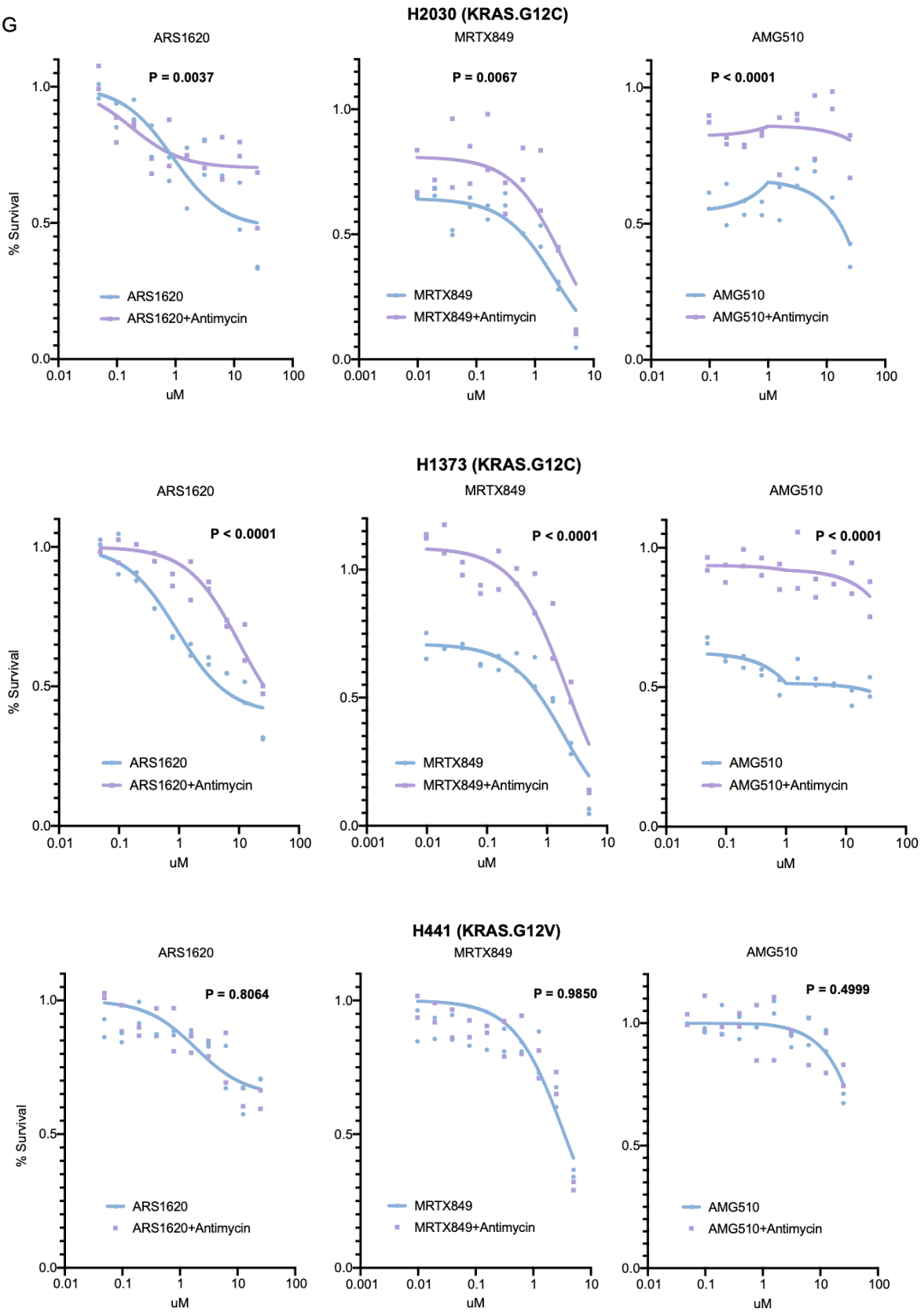
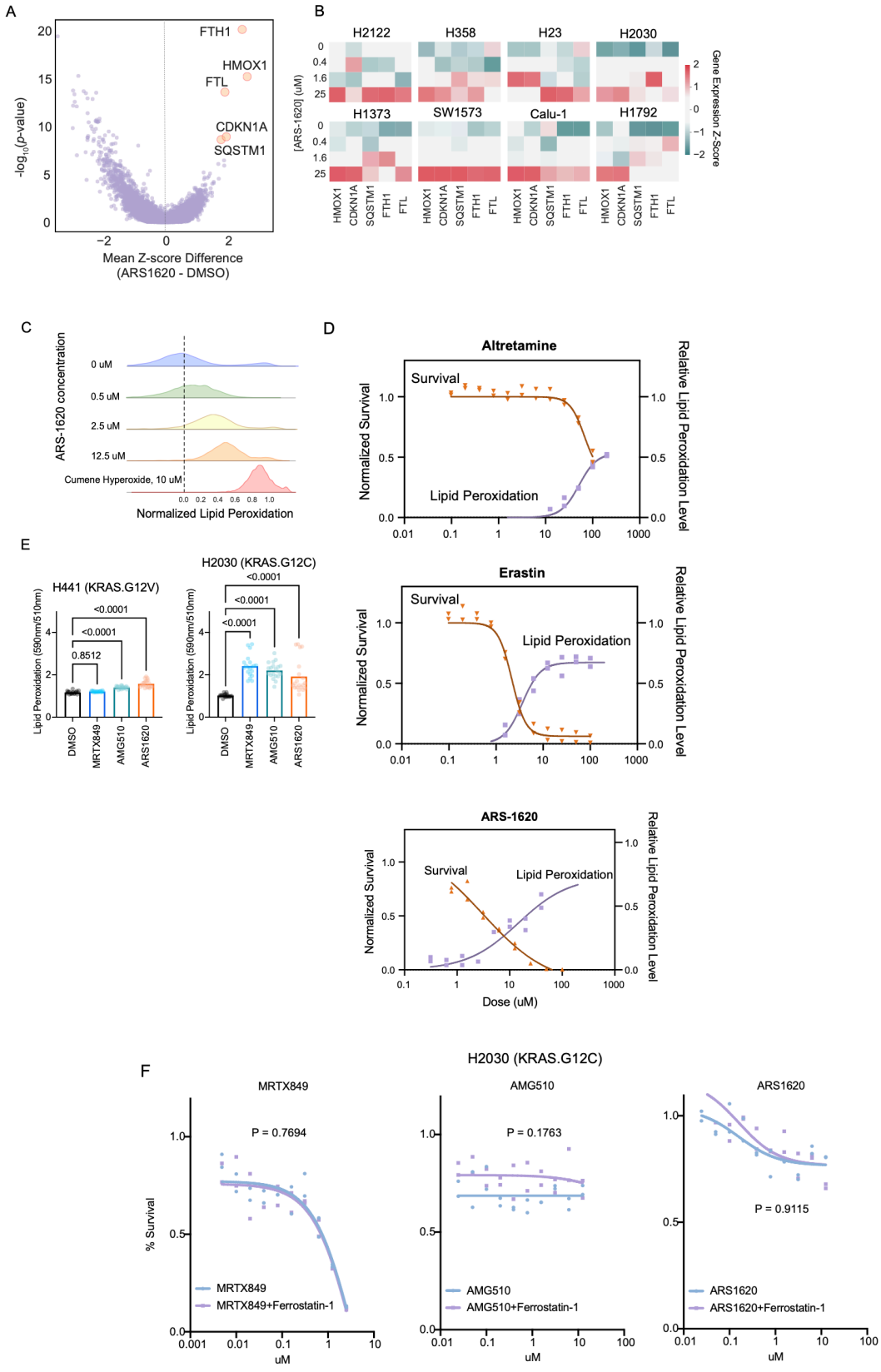

H2030 (KRAS.G12C)
ARS1620
MRTX849
AMG510
P = 0.0037
P = 0.0067
P < 0.0001
% Survival
ARS1620
ARS1620+Antimycin
MRTX849
MRTX849+Antimycin
AMG510
AMG510+Antimycin
uM
H1373 (KRAS.G12C)
ARS1620
MRTX849
AMG510
P < 0.0001
P < 0.0001
P < 0.0001
% Survival
ARS1620
ARS1620+Antimycin
MRTX849
MRTX849+Antimycin
AMG510
AMG510+Antimycin
uM
H441 (KRAS.G12V)
ARS1620
MRTX849
AMG510
P = 0.8064
P = 0.9850
P = 0.4999
% Survival
ARS1620
ARS1620+Antimycin
MRTX849
MRTX849+Antimycin
AMG510
AMG510+Antimycin
uM

Fig 2.4. AMG510 increases mitochondrial respiration and electron transport chain activity as lethality mechanims of KRAS.G12C inhibition (A) seahorse assay measuring oxygen consumption of H2030 (KRAS.G12C) cells at (2h) after AMG510 treatment (B) broken into components basal respiration, spare respiratory capacity, and glycolytic respiratory fractions. (C) Oxygen consumption measurements in H2030 ARS1620-persistor line after drug stimulation and comparison between untreated H2030 and H2030 ARS1620-persistor line. (D) Total cellular ATP measurements at timepoint concurrent with searhose assay dosing (500nM AMG510, 2h). (E) Investigation of AMG510 drug-induced mitochondrial dynamics in a time course assay with multiple measure- ments of mitochondrial phenotypes including mitochondrial membrane potential, mitochondrially localized reactive oxygen species generation, and caspase 3/7 cleavage. (F) Genome-wide synthetic lethal screen data analyzed for mito-resident genes and mito-ribosome genes compared to all-gene data in the context of effect on rescue of KRAS.G12C lethality. (G) Rescue of KRAS.G12C inhibitor lethality by Antimycin measured across inhibitors and cell lines with different KRAS.G12 mutational status.


A
FTH1
HMOX1
FTL
CDKN1A
SQSTM1
-log10(p-value)
Mean Z-score Difference
(ARS1620 - DMSO)
B
H2122
H358
H23
H2030
H1373
SW1573
Calu-1
H1792
[ARS-1620] (uM)
Gene Expression Z-Score
HMOX1
CDKN1A
SQSTM1
FTH1
FTL
C
ARS-1620 concentration
0 uM
0.5 uM
2.5 uM
12.5 uM
Cumene Hyperoxide, 10 uM
Normalized Lipid Peroxidation
D
Altretamine
Erastin
ARS-1620
Survival
Lipid Peroxidation
Normalized Survival
Relative Lipid Peroxidation Level
Dose (uM)
E
H441 (KRAS.G12V)
H2030 (KRAS.G12C)
Lipid Peroxidation (590nm/510nm)
<0.0001
0.8512
DMSO
MRTX849
AMG510
ARS1620
F
H2030 (KRAS.G12C)
MRTX849
AMG510
ARS1620
P = 0.7694
P = 0.1763
P = 0.9115
% Survival
MRTX849
MRTX849+Ferrostatin-1
AMG510
AMG510+Ferrostatin-1
ARS1620
ARS1620+Ferrostatin-1
uM

G

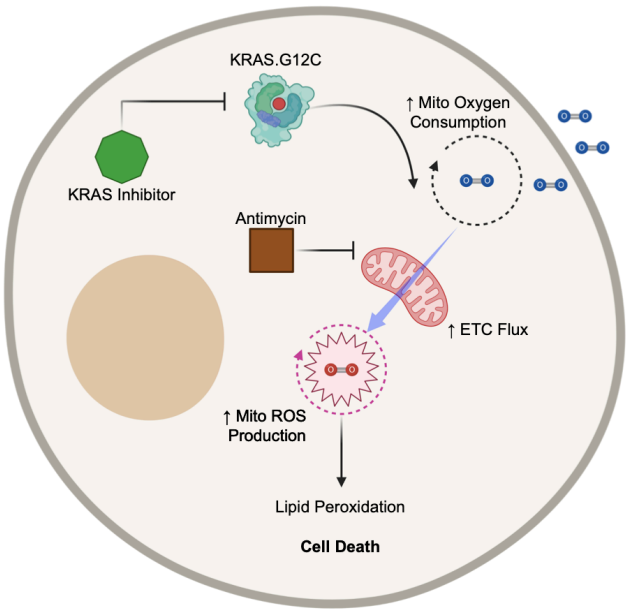


Figure 2.5. GENEVA uncovers secondary mechanism of cell death by lipid peroxidation specific to KRAS.G12C on-target engagement. (A) Z-score differences aggregated across multiple G12C lines from GENEVA pools drugged with ARS1620 in-vitro demonstrate a shared upregulation of the anti-ferroptotic genes and (B) gene expression of the anti-ferroptotic genes for each cell line respond to increasing ARS1620 dosage. (C) Experimental investigation of ferroptosis using a fluorescent lipid peroxidation sensor demonstrate dose-response of cells of lipid peroxidation to ARS1620 dosage (48h). (D) Survival and lipid peroxidation kinetics across known ferroptotic agents (Erastin, Altretamine) as compared to ARS1620 and (E) lipid peroxidation measurements in response to KRAS.G12C inhibitors in a KRAS.G12C cell line H2030 and KRAS.G12V cell line H441. (F) Rescue experiment of lethality of KRAS.G12C inhibitors with ferrostatin-1. (G) Mechanistic picture of KRAS.G12C inhibitors appears to be one of inducing hyperactive mitochondria causing increased oxygen consumption and ETC flux leading to elevated levels of mitochondrial ROS production which causes lipid peroxidation broadly in the cell and cell death.

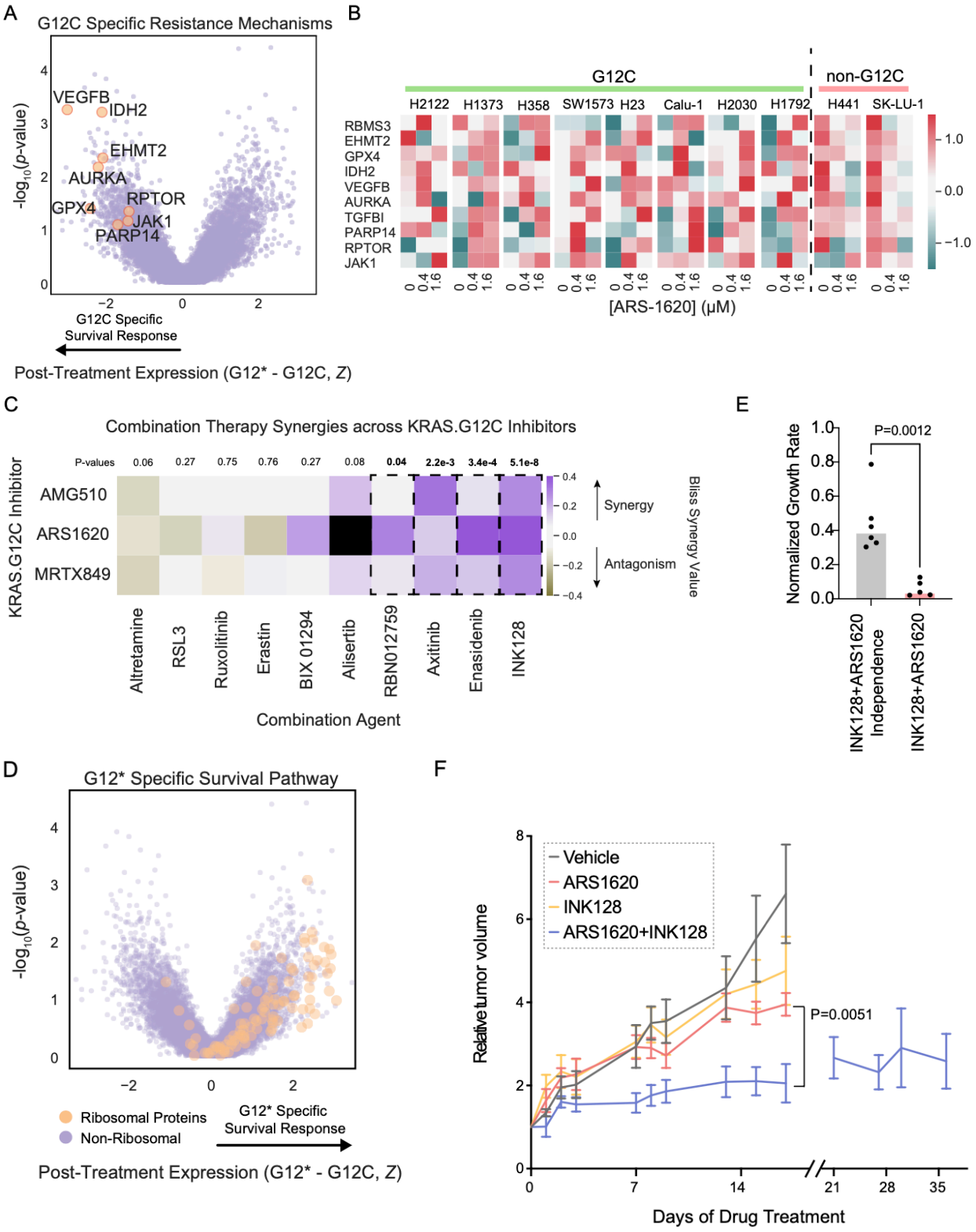


Figure 2.6. GENEVA predicts effective combination therapies by identification of upregulated pathways in treatment-tolerant cells. (A) Comparison of G12* against G12C KRAS mutant lines in GENEVA pools at moderate dosage (1.6 uM ARS1620) reveals expression upregulation of several druggable targets indicating possible survival mechanisms to G12C inhibition by

ARS1620. (B) Individual cell line expression of druggable targets across G12C and G12* lines across increasing doses of ARS1620. (C) Compounds inhbiting predicted combination therapy targets dosed in combination with three ARS1620 inhibitors assayed by cell survival for drug synergy. Highlighted columns indicate significant synergy and black boxes indicate missing data. (D) Further dissection of GENEVA data examining ribosomal proteins as a class are upregulated as a class non-G12C cell lines as part of survival response to ARS1620 treatment. (E) INK128 empirical in-vivo combination therapy Bliss drug synergy with ARS1620 compared to the expected no synergy expected growth rate. (F) Relative tumor volume of ARS1620 and INK128 in multi-arm combination therapy in-vivo study in H1373 KRAS.G12C mutant line (n=4-5 per condition)

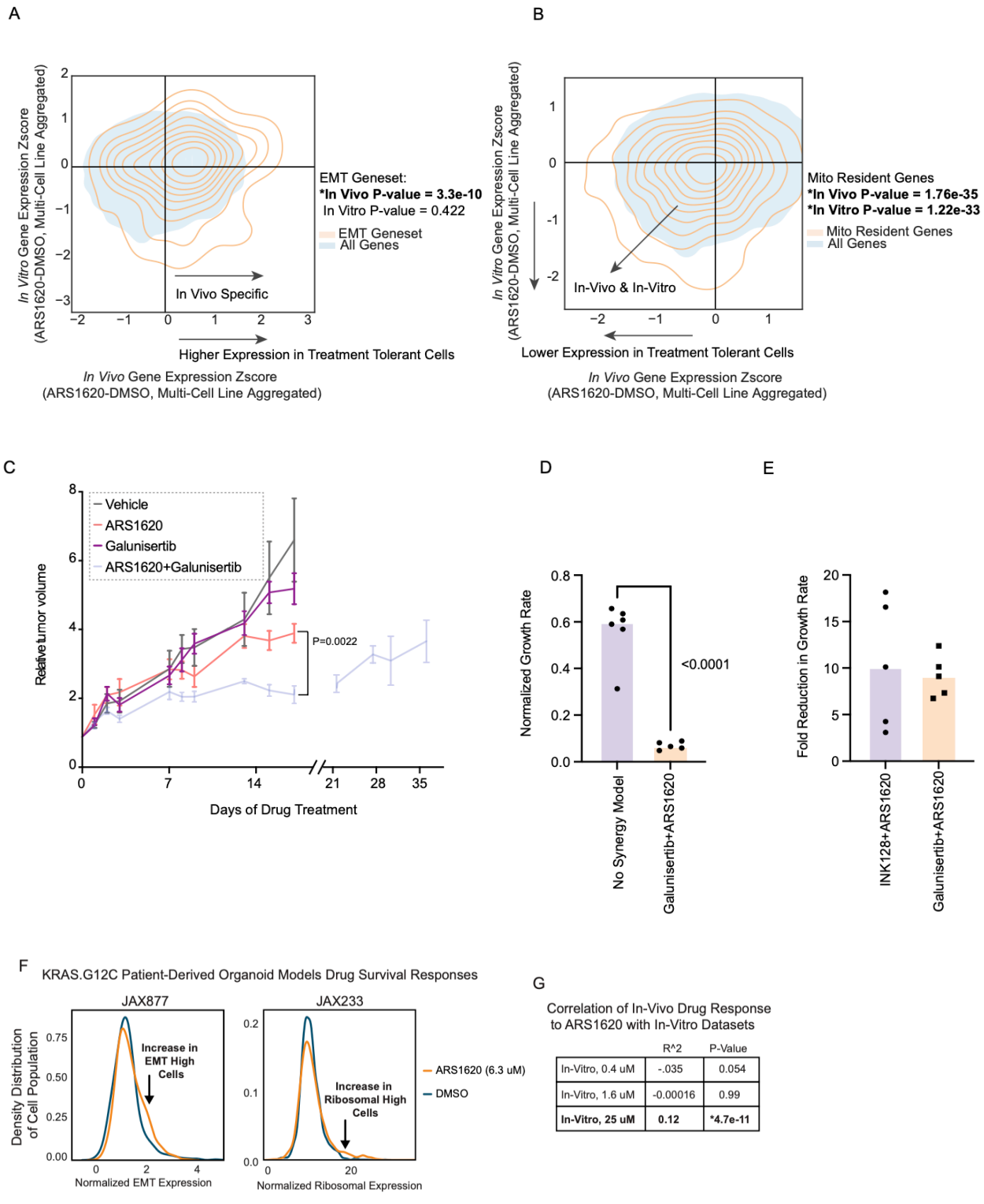


Figure 2.7. GENEVA reveals in-vivo specific upregulation of EMT as a response to KRAS.G12C inhibition tolerance as a targetable resistance mechanism (A) In vivo GENEVA vs in vitro GENEVA comparison of EMT geneset upregulation post-drug treatment with ARS1620 and (B) mitochondrial resident genes reveal in-vitro and in-vivo concordant and discordant effects. (C)

In-vivo multi-arm combination therapy (n=4-5) growth curves with ARS1620 and EMT inhibitor, Galunisertib; vehicle and ARS1620 single-dose curves reported from Fig 6F. (D) Expected growth rate assuming no synergistic effect by Bliss synergy model compared to measured growth rates of ARS1620+Galunisertib (E) Compari- son of G12C combination therapies predicted from GENEVA in vitro (Ink128) and GENEVA in vivo (Galunisertib) by fold reduction in growth rate compared to the respective expected null Bliss synergy models. (F) KRAS.G12C PDX models post-drug treatment with ARS1620 categorized for upregulation in EMT or Ribosomal transcript expression levels. (G) In vivo vs In vitro spearman correlation of GENEVA pools for estimation of relative cross-model system effective dose using full-transcriptome drug-specific expression response.

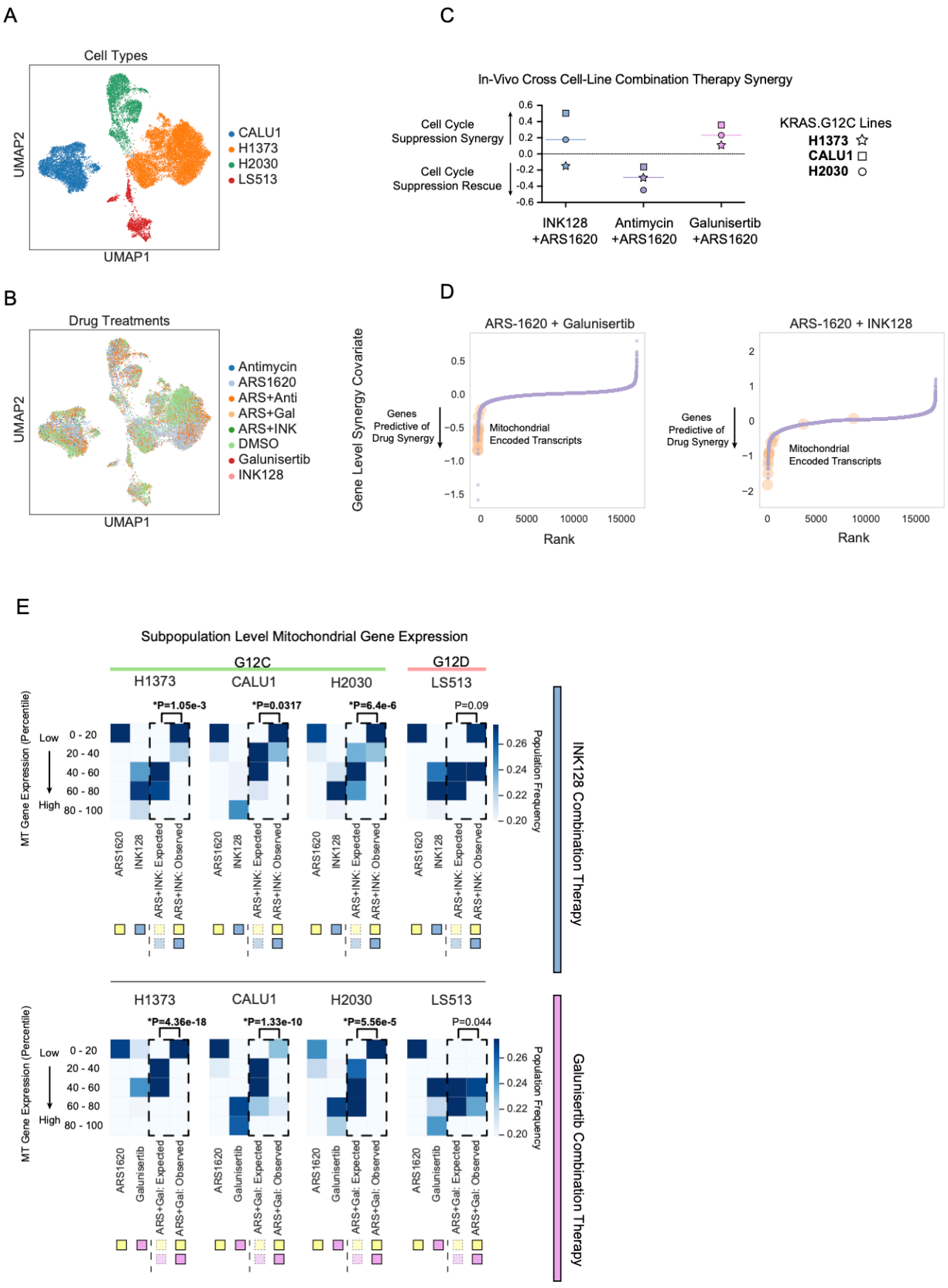

A
Cell Types
UMAP2
UMAP1
CALU1
H1373
H2030
LS513
B
Drug Treatments
UMAP2
UMAP1
Antimycin
ARS1620
ARS+Anti
ARS+Gal
ARS+INK
DMSO
Galunisertib
INK128
C
In-Vivo Cross Cell-Line Combination Therapy Synergy
Cell Cycle Suppression Synergy
Cell Cycle Suppression Rescue
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
INK128 +ARS1620
Antimycin +ARS1620
Galunisertib +ARS1620
KRAS.G12C Lines
H1373 ☆
CALU1 □
H2030 ○
D
Gene Level Synergy Covariate
ARS-1620 + Galunisertib
Genes Predictive of Drug Synergy
Mitochondrial Encoded Transcripts
0.5
0.0
−0.5
−1.0
−1.5
0
5000
10000
15000
Rank
ARS-1620 + INK128
Genes Predictive of Drug Synergy
Mitochondrial Encoded Transcripts
2
1
0
−1
−2
0
5000
10000
15000
Rank
E
Subpopulation Level Mitochondrial Gene Expression
G12C
G12D
H1373
CALU1
H2030
LS513
*P=1.05e-3
*P=0.0317
*P=6.4e-6
P=0.09
MT Gene Expression (Percentile)
Low
High
0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
Population Frequency
0.26
0.24
0.22
0.20
ARS1620
INK128
ARS+INK: Expected
ARS+INK: Observed
INK128 Combination Therapy
H1373
CALU1
H2030
LS513
*P=4.36e-18
*P=1.33e-10
*P=5.56e-5
P=0.044
MT Gene Expression (Percentile)
Low
High
0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
Population Frequency
0.26
0.24
0.22
0.20
ARS1620
Galunisertib
ARS+Gal: Expected
ARS+Gal: Observed
Galunisertib Combination Therapy

Fig 2.8. Scaling GENEVA in-vivo for combination therapy studies in KRAS.G12 mutant lines colored by (A)cell line of origin and (B) treatment condition. (C) Bliss synergy calculations performed on cell cycle states in each drug condition singly and in combination estimated for G12C lines across drug conditions. (D) Assessment of genes driving synergistic drug effect of Galunisertib and INK128 in combination with ARS1620 using a multifactorial linear model estimating gene-level synergy with mitochondrial transcripts highlighted. (E) Breakdown of mitochondrial transcript expression across G12C and G12D lines reveal significantly significant downregulation in G12C and G12D lines in a synergistic fashion when modeled against the expected Bliss null synergy effect.

Chapter 3: Conclusions, relevance and challenges for the GENEVA platform

My work has focused on bringing interdisciplinary experimental and computational research methods together to solve outstanding biological questions in the cancer space. It has lent insight into the highly studied field of KRAS.G12C – a field created by Kevan Shokat's group at UCSF by the development of the first KRAS.G12C inhibitor. It has also built methods for scaling *in vivo* drug MOA discovery methods for the first time, allowing for rapid in vivo assay of both phenotype and perturbation effects. I will discuss these and focus on summarizing the conclusions, discussing their relevance to the subfields they influence and then discussing the outstanding questions that remain in those spheres and how they may be solved.

Before the seminal 2012 paper indicating that KRAS could be drugged as a classic cancer target, the field of both academic and pharmaceutical science considered KRAS an important but untouchable drug target. KRAS is mutated as a driver mutation in over 46% of all cancer types, across almost all cancer indications. It is a powerful oncogene that when mutated also falls to specific sites: AA 12,13,61,121. Furthermore, the simple combination of KRAS with PTEN, or TP53 are the foundation for some of the most used syngeneic mouse tumor models reflecting the power of KRAS in rewiring cells to become tumorigenic. Despite this structural biologists and medicinal chemists have long been loath to chase KRAS as a tractable drug target. The reason lies in the protein surface and binding groove of KRAS. The surface of the protein is fairly smooth comparatively, without distinctive binding sites that would allow for small molecule specificity or optimization. Furthermore, the unique binding site takes a classic metabolite, GTP/GDP – hardly a site that allows for specificity to KRAS alone out of the many G-proteins that exist. For this reason, despite the fact that from an early stage RAS proteins and KRAS were clear candidates for small molecule inhibitor development, pharmaceutical companies have not funded development of internal RAS programs.

The advent of the 2012 paper from the Shokat lab showed how this classic boogeyman was tamed – by using both structure guided rational small molecule design and clever chemistry

inherent in covalent inhibitors. The authors used the insight that many of the oncogenic mutations in KRAS frequently mutated to cysteines, which allowed for a free reactive group that when bound formed irreversible covalent bonds. They developed a covalent small molecule inhibitor containing both i) structure optimized for the binding pocket as well as ii) a reactive group built into a site that would face the cysteine directly once bound in the groove. In this way they were able to significantly increase specificity to an otherwise hard to target protein. Since that time, pharmaceutical companies have rapidly taken notice and in only 8 years time Amgen has successfully produced and shepherded the first KRAS.G12C covalent small molecule inhibitor into approval for use in patients - a record-breaking time for any drug discovery program. Hot on their heels are companies focused on the RAS space broadly as the interest in G-proteins explodes due to both academic and pharmaceutical research successes.

It is within this context that GENEVA was developed to understand these inhibitors. Upon first inspection, this highly trafficked field with significant pharmaceutical potential would be quickly understood because of its enormous value. However GENEVA in a few experiments was able to uncover what was already well established in G12C space – namely that these were inhibitors designed to target and prove lethal to cells containing specific G12C mutations in KRAS. GENEVA however went significantly further and was able to show mitochondrial hyperoxidation, drug resistance pathways, and in vivo specific pathways that implicated EMT as an adaptive survival response.

Mitochondrial hyperoxidation

Mitochondrial hyperoxidation was shown in chapter two to be a clear GENEVA signal. After drug treatment rare subclones of cells were enriched and because of scRNAseq's ability to dissect cellular populations, these subclones were able to be detected. Furthermore, the consistency of the effect across different genetic backgrounds was able to boost the signal that remained consistent across many different cell types, and removed artifacts that would be specific to one or a few cell lines. KRAS.G12C inhibitors have not been shown to our knowledge to have

any effect on mitochondria and for us this signal was a mysterious clue. However in early literature, KRAS.G12 mutations have been shown to have an effect on mitochondria. In fact – one prominent publication even showed that in an inducible activation model of cancer genesis using a KRAS.G12 mutation, this protein even localized to the mitochondria and that this localization was critical to the oncogenic potential of KRAS! For this reason because of both present and past clues to a mitochondrial phenotype we decided to investigate this signal further. We showed there were hyperoxidation phenotypes associated with KRAS.G12C inhibitor dosing. While many compounds can inhibit respiration at the mitochondria (all ETC inhibitors), very few compounds are known to cause increased respiration. Furthermore, when we used an ETC inhibitor well known for its ability to inhibit mitochondrial respiration, it was able to prevent KRAS.G12C inhibitors' lethal effect. What this shows is that despite intense interest from the field, well-established MOA, and even the approval of a clinical compound for use in patients – something is missing in the way we as a field conduct biology. The mechanism of lethality at the site of the mitochondria remained unreported and undiscovered and only with a technology like GENEVA that could combine the ability to simultaneously interrogate multiple drugs, multiple genetic backgrounds, at an extended time point allowing for rare subclones to emerge and takeover, could a discovery that is both biologically deep and broadly generalizable appear.

Challenges remain however namely what is the signal cascade of events that take us from the molecule binding to the cysteine mutation, to the endpoint of cell death via hyper-respiration. A focused screen in which synthetic lethal perturbations using a readout that is not cell death but rather hyperoxidation, in a KRAS.G12C inhibitor context, could prove useful in defining the members of this pathway. Because of the classic deep nexus of cell death events at the site of the mitochondria (apoptosis is defined by activation of caspases on the outer mitochondrial membrane) it is possible there is more to previously understood death pathways that may in fact involve an alteration in respiration preceeding death. This future investigation would be critical to understanding G12C – but perhaps even more critical to understanding and reevaluating what we

know classically about how cells die. And of course, understanding these mechanisms will inevitably lead to methods of drug development that may take advantage of the renewed understanding of lethality.

Drug Resistance Evasion Mechanisms

The nature of GENEVA also allows for rapid identification of drug resistance mechanisms. Because the assay allows for long-term treatment of cell pools, achieved by apriori measurement and balancing of cellular growth rates, we are able to effectively extend the window of treatment duration to timepoints which allow sensitive cells to die, and tolerant cells with novel rewiring to survive. This analysis revealed VEGF, JAK, mTOR, and other programs as potential candidates for evasion pathways. All of these genes and some associated sets of genes such as ribosomal proteins, were upregulated post KRAS.G12C inhibitor treatment. While this does not establish causality, several of these prioritized candidates were tested in traditional drug synergy models and shown to be efficacious, where synergy is defined as the sum being greater than the parts. Combination therapy discovery in pharmaceutical research has been a notoriously tricky problem. It is akin to starting a new drug discovery target search: in a specific target-molecule pair, given enough time, what are the evasion pathways out of all possible targets that could appear? In some ways because of the adaptive nature of cancer cells, it is like beginning research into a space anew because ultimately there are many roads to the same resistance phenotype a tumor can travel. GENEVA allows for determination of this in a rapid assay format by finding both mechanisms that generalize across different genetic and phenotypic backgrounds by using many cell lines together, as well as by allowing for true drug resistance studies by extending the window on drug treatment assays past normal IC50 intervals (48-72h). Our data successfully predicted that mTOR inhibitors were the most likely candidate for advancement into in-vivo studies, and this corroborates well with both literature reporting the relevance of mTOR to G12C inhibitor survival, and to pharmaceutical research as several pharmaceutical companies have already advanced combination therapies of mTOR with G12C inhibitors. In summary, GENEVA allows for rapid

identification of pathways in a drug resistance setting recapitulating significant research investments.

Challenges remain however, specifically in constructing pools of mosaic cells for new target sets. Experimentally, banking and growing new pools of 20-30 cell lines for a GENEVA experiment and balancing the inputs ratiometrically to allow for even harvest of cells after a long timecourse can be time-intensive for different drug targets. Furthermore, in some cases of rare cancers it may be challenging to find enough cell lines to comprise a mosaic. However, we would emphasize that although the development of GENEVA here represents an important step it should be viewed as the beginning of a new way of conducting focused oncology research at scale in a different manner and that new improvements adding genetic manipulations or other measurements such as proteomics could prove useful. Technology development could significantly enhance the scale of these and require fewer cell lines to obtain similar information density.

In vivo specific insights

Lastly, I would like to address the important and novel advancement of a high-throughput MOA dissection method like GENEVA into *in vivo* space. In vivo is the classic gold-standard of validation in both the academic realm and the pharmaceutical realm. Hardly an oncology paper would be relevant without at least some in vivo validation of their most important claims. After all cancer is a disease that happens in the body, where cells grow unconstrained – and growth rate differences in vitro simply would not suffice as a model. However in vivo work has been constrained by the methods that exist. N=10 mice is common for a single arm in one study. Most studies require 2 arms at least, and in several cell lines. With experimental testing and mistakes, it is common to use several hundred mice for a single or few drug targets. Furthermore, with adding on different compound treatments and daily feeding of animals or injections with drug compound, the exercise of in vivo experimentation quickly becomes a massive operation for a very small drugsXcell types question.

GENEVA fundamentally scales this down to one mouse; one mouse contains all the patient cells from diverse populations and one mouse receives one drug. Because at the end of drug treatment the tumor is dissociated and scRNAseq is applied, the unit of information is in fact a single cell. Both miniaturization and clever experimental methods using cell hashing allow for reduction of information in both genetic and chemical/drug space to the single cell. GENEVA furthermore reads out different information than traditional in vivo studies. Most in vivo studies are focused on coarse readouts such as tumor volume reduction over time as a proxy for phenotype with and without drug. However, while daily tumor measurements may tell us something about if a drug is achieving its action, it does not tell us about the 'how' in which it acts. GENEVA captures phenotypes by relying on a combination of both cell counts pre/post drug and cell cycle state pre/post drug; GENEVA also simultaneously captures deep transcriptome shift information pre/post drug and is therefore able to go beyond and answer the 'how' question in the same experiment. GENEVA is not simply an experimental scaling method, rather it is a method that also delivers an enormous quantity of drug-informative data much greater than the coarse phenotype measurements commonplace in today's in vivo space.

An example of the relevance of GENEVA is the discovery and prioritization of KRAS.G12C resistance target EMT. Current interest in resistance targets in G12C space focus around SHP2/SOS which is a coactivator of KRAS, EGFR activation as a classic GPCR in a parallel pathway, or internal rewiring by mTOR upregulation. However, to date no significant clinical programs have been advanced looking at EMT or cell state shift in this space. Our in vitro data furthermore did not show EMT activation as a drug resistance mechanism, leading to speculation that perhaps the reason the field is not focusing on EMT is because of a model system limitation to in vitro discovery models. Our in vivo data was able to show us that EMT is a strong and multi-cell line generalizable phenomenon in G12C inhibitor evasion. An interesting data point supporting this also comes from early clinical trial histopathology data collected from cells that have become resistant inside patients to the G12C inhibitor AMG510. In a NEJM paper detailing

these patients, the pathologists noted a ‘non-squamous to squamous cell state change’ accompanying patient tumor resistance to drug. Upon inspection of the tumor images, it is clear that after tumor resistance the cells lose clearly defined cell boundaries, resulting in a highly dense and undefined packing of cells reminiscent of the classic endothelial to mesenchymal transition. For these reasons, we believe that GENEVA offers substantive information that could not be obtained at scale in traditional in vitro models and allows us to quickly dissect information that is more patient relevant.

Challenges and questions of course remain – are there effects in vivo that are cell-cell interaction specific? Are there immune modulatory effects in vivo that are confounding to patient results? And are there microenvironment effects that may skew results comparatively to patient data? Our response to these questions are two-fold: i) while these may and probably do occur to some extent, the fact that tumors grow overwhelmingly in vivo indicate clearly that the main driver of tumor growth are the cells themselves, and not the environmental interaction ii) the FDA consistently defines in vivo experiments in oncology to be the gold standard for a good reason, namely that out of all the experimental systems we have they are the best ones. Further research is of course expected as well and we hope to use the burden of validation to prove the validity of the hypothesis generation that comes from GENEVA.

GENEVA combines many different elements that are highly relevant to drug discovery in oncology in a single assay: MOA understanding, tumor resistance predictions, unprecedented access to in vivo information, and phenotype + transcriptome capture in a highly miniaturized assay. The power of the assay lies not necessarily in its relative strength in one direction or the other but rather in the ability to capture in multiple important information spaces simultaneously – work which would have taken many teams of scientists many years. The fact that this assay can then be scaled to do many compounds quickly promises to advance research and hopefully treatment timelines dramatically.

Chapter 4: RBMS1 suppresses colon cancer metastasis through targeted stabilization of its mRNA regulon

A version of this was published as:

Title: RBMS1 suppresses colon cancer metastasis through targeted stabilization of its mRNA regulon

Authors: Johnny Yu1,2,3,†, Albertas Navickas1,2,3,†, Hosseinali Asgharian1,2,3,†, Bruce Culbertson1,2,3, Lisa Fish1,2,3, Kristle Garcia1,2,3, John Paolo Olegario1,2,3, Maria Dermit7, Martin Dodel7, Benjamin Hänisch1,2,3, Yikai Luo1,2,3, Ethan M Weinberg4, Rodrigo Dienstmann5, Robert S Warren3,6, Faraz K. Mardakheh7, Hani Goodarzi1,2,3,*

Affiliations:

1Department of Biochemistry & Biophysics, University of California, San Francisco, San Francisco, California, USA.

2Department of Urology, University of California, San Francisco, San Francisco, California, USA.

3Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, California, USA.

4Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA.

5Medical Oncology Department, Vall d'Hebron University Hospital, Barcelona, Spain.

6Department of Surgery, University of California, San Francisco, San Francisco, California, USA.

7Centre for Cancer Cell & Molecular Biology, Barts Cancer Institute, Queen Mary University of London, Charterhouse square, London EC1M 6BQ, The United Kingdom.

*Correspondence to: hani.goodarzi@ucsf.edu

† These authors contributed equally.

ABSTRACT

Identifying master regulators that drive pathological gene expression is a key challenge in precision oncology. Here, we have developed an analytical framework, named PRADA, that identifies oncogenic RNA-binding proteins through the systematic detection of coordinated changes in their target regulons. Application of this approach to data collected from clinical samples, patient-derived xenografts, and cell line models of colon cancer metastasis revealed the RNA-binding protein RBMS1 as a suppressor of colon cancer progression. We observed that silencing RBMS1 results in increased metastatic capacity in xenograft mouse models, and that restoring its expression blunts metastatic liver colonization. We have found that RBMS1 functions as a post-transcriptional regulator of RNA stability by directly binding its target mRNAs. Together, our findings establish a role for RBMS1 as a previously unknown regulator of RNA stability and as a suppressor of colon cancer metastasis with clinical utility for risk stratification of patients.

INTRODUCTION

Metastatic progression in colorectal cancer (CRC) is accompanied by widespread gene expression reprogramming. Cancer cells often co-opt post-transcriptional regulatory mechanisms to achieve pathological expression of gene networks that drive metastasis (Goodarzi 2014, Goodarzi 2015, Goodarzi 2016). Colorectal cancer is the third most commonly diagnosed cancer (Siegel, 2018), therefore understanding the underlying regulatory programs that drive metastatic progression in this disease is a crucial step towards improving patient outcomes. Notably, there are not many predictive computational methods aimed at the discovery of unknown regulatory networks. By relying on annotated post-transcriptional regulatory pathways, e.g. those mediated by microRNAs, existing methods fail to capture previously unknown regulatory interactions. To tackle this problem, we have developed a computational approach called PRADA that identifies post-transcriptional master regulators responsible for aberrant mRNA stability and gene expression in cancer cells. By applying this

tool to a large compendium of gene expression profiling data from patient samples, patient-derived xenograft models, and established colon cancer cell lines, we identified a novel regulatory program involved in CRC metastasis. We find that this previously unknown regulatory pathway, which controls mRNA stability and is mediated by the RNA-binding protein RBMS1, is often silenced in highly metastatic CRC tumors. We demonstrate that RBMS1 stabilizes transcripts by binding to the last exons of target mRNAs in concert with the RNA-binding protein ELAVL1, and that in highly metastatic CRC cells and patient tumors, RBMS1 downregulation is associated with poor clinical outcome. We identify mRNA targets of RBMS1 that are functionally relevant to metastasis and reveal that RBMS1 silencing can be accomplished through epigenetic dysregulation. This study not only describes a disease-relevant post-transcriptional regulatory mechanism that governs the stability of a sizeable regulon, but also demonstrates the value of bottom-up discovery strategies like PRADA that do not solely rely on prior knowledge of annotated regulatory programs.

RESULTS

PRADA identifies RBMS1 as a novel regulator of CRC progression

Metastasis is a complex multistep process and requires modulation of many cellular pathways and functions. As such, increased metastatic capacity often involves broad reprogramming of the gene expression landscape in cancer cells. Thus, a mechanistic dissection of cancer progression relies on the systematic identification of the underlying regulatory programs that drive pathologic cellular states. To accomplish this, we developed PRADA (Prioritization of Regulatory Pathways based on Analysis of RNA Dynamic Alterations) that uses regulatory network predictions to identify the key RNA-binding proteins (RBPs) that may act as master regulators of mRNA stability and gene expression. Conceptually, PRADA solves a regression problem to predict changes in gene expression as a function of the expression of RBPs and their known or predicted targets across the transcriptome. In this customized regression analysis, the coefficient assigned to each RBP reflects its strength as a

regulator of gene expression and the direction of its effect (i.e. activator or repressor). Given our limited knowledge of post-transcriptional regulatory pathways and their role in human disease, PRADA provides an opportunity for the discovery of previously unknown disease pathways. To assess the contribution of post-transcriptional regulatory programs to colon cancer metastasis, we took advantage of a publicly available set of gene expression profiles from established colon cancer cell lines (GSE59857). We first categorized these cell lines as poorly metastatic or highly metastatic based on their liver colonization capacity in xenograft mouse models (Supplementary Fig. 1A; (Hamada, 2008)). We then performed differential gene expression analysis to compare gene expression changes across the transcriptomes of the two groups. Finally, we applied PRADA to this dataset to find RNA-binding proteins whose differential activity is most informative of metastasis-associated gene expression changes. PRADA identified the protein RBMS1 as the RNA-binding protein with the strongest regulatory potential in this dataset (Fig. 1A). The size and the direction of the regulatory coefficient assigned to RBMS1 implies that RBMS1 levels are strongly informative of changes in the expression of transcripts with sequence matches to RBMS1 binding sites. To confirm this, we performed motif enrichment analysis using FIRE, which uses mutual information to capture the association between the presence and absence of a given binding site and genome-wide transcriptomic measurements (Elemento, 2007). As shown in Fig. 1B, the RBMS1 consensus binding site is strongly informative of gene expression differences between poorly and highly metastatic cells, showing a highly significant enrichment among the transcripts that have decreased expression in highly metastatic cells. To ensure that this result is not sensitive to the choice of a specific RBMS1 binding motif, we tested two additional sequence models: (i) an independently derived representation of RBMS1 binding preferences (Ray, 2013), and (ii) predictions based on DeepBind models (Alipanahi, 2015) (Supplementary Table 1). Each of these three predicted RBMS1 binding motifs gave largely identical results, and we observed a similar decrease in the expression of the putative RBMS1 regulon derived from each motif (Supplementary Fig. 1B).

For consistency, we have used the DeepBind-derived RBMS1 regulon (referred to as the RBMS1 regulon hereafter) in our subsequent analyses as DeepBind is the state-of-the-art approach for predicting protein-nucleic acid interactions. Importantly, concordant with the results reported by PRADA and the lower expression of its regulon, the expression of RBMS1 was also significantly lower in highly metastatic cells (Fig. 1C). Moreover, consistent with RBMS1 acting as a post-transcriptional regulator of these genes, we observed a significant correlation between the expression of RBMS1 and the average expression of its regulon in multiple independent datasets (Fig. 1D, Supplementary Fig. 1C-D).

In order to further assess the association between RBMS1 and CRC metastasis in more directly disease-relevant models, we took advantage of patient-derived xenograft (PDX) models of colon cancer metastasis to liver. We used RNA-seq data from three parental CRC PDX models (CLR4, CLR27, and CLR32) and their highly liver metastatic derivatives (CLR-LvMs). The highly metastatic CLR-LvM models were derived from the parental PDXs by repeated splenic delivery and growth of cells in the liver of immunocompromised mice (NOD scid gamma), recapitulating the major site of CRC metastasis in humans (Yamaguchi, 2019). We observed that the increase in the metastatic capacity of the CLR-LvM PDX models compared to their parental PDXs was accompanied by a significant reduction in the expression of the RBMS1 regulon in the highly metastatic CLR-LvM models (Fig. 1E). More importantly, all three models also showed a significant and concomitant decrease in the expression of RBMS1 (Fig. 1F).

In addition to these cell line and PDX models of CRC liver metastasis, we also performed RNA-seq on matched primary tumors and liver metastases biopsied and frozen from two patients. As shown in Fig. 1G, in both cases, the RBMS1 regulon was enriched among transcripts that were downregulated in the metastases relative to their primary tumors. Interestingly, while the gene expression changes between the primary tumors and metastases were not generally correlated ($R = 0.02$, $P = 0.1$), the RBMS1 regulon was independently

downregulated in both metastatic tumors (Supplementary Fig. 1E). Consistently, RBMS1 was also strongly silenced in both metastases (Fig. 1H). Collectively, these findings in cell lines, PDXs, and clinical samples implicate RBMS1 silencing in both the downregulation of its putative regulon and in CRC metastatic progression.

RBMS1 acts as a post-transcriptional regulator of RNA stability

The correlated expression of RBMS1 and its putative regulon defined from its binding site preferences implies that RBMS1 modulates gene expression. To further assess the potential role of RBMS1 as a post-transcriptional regulator, we performed RNA sequencing of RBMS1 knockdown and control SW480 colon cancer cells. We chose the SW480 cell line for this experiment because: (i) RBMS1 expression in SW480 is among the highest in colon cancer lines tested by us and others, and (ii) it is an established xenograft model of liver metastasis and is considered to be poorly metastatic (relative to other CRC lines listed in Supplementary Fig. 1A). We used two independent shRNAs to silence RBMS1 in SW480 cells and confirmed its knockdown with both qPCR and Western blotting (Supplementary Fig. 2A). A 2.5-fold reduction in RBMS1 protein expression induced major changes in gene expression (Supplementary Fig. 2B) and resulted in a significant decrease in the expression of RBMS1 regulon (Fig. 2A). Since we are focused on the role of RBMS1 as an RNA-binding protein, we reasoned that its effect on gene expression is likely through regulation of the stability of its target RNAs. To test this hypothesis, we took advantage of REMBRANDTS, a computational framework we have developed to estimate RNA stability from RNA-seq data (Alkallas, 2017). We have previously established that REMBRANDTS accurately recapitulates experimental RNA stability measurements (Alkallas, 2017). Application of this method to RNA-seq data from RBMS1 knockdown and control cells found a significant enrichment of RBMS1 targets among transcripts destabilized upon RBMS1 silencing (Fig. 2B). REMBRANDTS relies on the comparison of exonic and intronic reads to measure changes in RNA stability. As shown in Supplementary Fig. 2C, RBMS1 silencing results in lower expression of its putative regulon as measured by exonic

reads; however, intronic reads, which reflect changes in pre-mRNA levels (and transcription rates) do not significantly change in response to RBMS1 depletion. To further strengthen our claim that RBMS1 functions as a post-transcriptional regulator of RNA stability rather than as a transcriptional activator, we also performed whole-genome RNA stability measurements in control and RBMS1 knockdown cells. For this, we used α-amanitin-mediated inhibition of RNA-polymerase II followed by RNA sequencing at two time points (0-hr and 9-hr). Consistent with our analyses with REMBRANDTS, we observed a similar reduction in the stability of the putative RBMS1 regulon upon RBMS1 silencing (Fig. 2C). We used RNA-seq data from the three matched PDX models (CLR panel) to compare RNA stability in highly liver metastatic models to that of their parental PDXs (using REMBRANDTS). As shown in Fig. 2D, lower RBMS1 expression in the LvM models (Fig. 1F) accompanies a reduction in the stability of its target regulon in all three independently derived models. Together, these findings establish RBMS1 as a post-transcriptional regulator of RNA stability with broad functional consequences for the transcriptome and clear implications for CRC progression.

RBMS1 CLIP-seq reveals 3' UTR binding of target mRNAs

To this point, our analyses were performed using in silico predicted binding targets of RBMS1. In order to create a transcriptome-wide snapshot of RBMS1 binding sites in colon cancer cells at nucleotide resolution, we performed UV crosslinking immunoprecipitation followed by sequencing (CLIP-seq). We carried out irCLIP (Zarnegar, 2016) for endogenous RBMS1 in SW480 cells, and identified hundreds of high-confidence RBMS1 binding sites across the transcriptome (Fig. 3A). Importantly, we observed that 90% of the CLIP-identified RBMS1 targets overlapped with our computationally-derived putative RBMS1 regulon (Supplementary Table 1). Consistently, these RBMS1-bound transcripts in our irCLIP data (RBMS1 targets hereafter) follow the same expression patterns as the computationally-derived RBMS1 regulon described above, exhibiting both decreased stability and expression upon RBMS1 silencing (Fig. 3B; Supplementary Fig. 3A).

In our analysis of RBMS1 binding sites, we noted a strong enrichment of RBMS1 binding to the last exon (on or close to 3' UTRs; Fig. 3A) of mRNAs. This is consistent with RBMS1 acting as a post-transcriptional regulator of RNA stability, and suggested that RBMS1 3' UTR binding could result in increased RNA stability. To assess this possibility, we built a bi-directional CMV promoter that drives the expression of both GFP and mCherry. We then cloned nine CLIP-seq-derived RBMS1 binding sites, plus ~150 nucleotides flanking the binding sites, downstream of GFP, and asked whether there was a reduction in GFP mRNA relative to mCherry mRNA upon RBMS1 depletion. As shown in Fig. 3C, there was a decrease in GFP reporter transcript levels in almost all instances. To further ensure that this effect was RBMS1-dependent, we also tested the two reporters with the strongest reduction in transcript levels (DAG1 and CD9) in LS174T cells, where RBMS1 is endogenously silenced. In this instance, as expected, there was no response in reporter mRNA levels upon transfection of RBMS1-targeting siRNAs (Supplementary Fig. 3B).

In order to gain insight into how RBMS1 regulates the stability of its targets, we carried out an unbiased search for its interacting protein partners. We performed immunoprecipitation of RBMS1, along with an IgG control, from SW480 cells and analyzed the resulting protein complexes by mass spectrometry (Fig. 3D). From this data, we found that the Gene Ontology terms RNA binding, RNA processing and negative regulation of mRNA metabolism were overrepresented among the RBMS1-interacting proteins (Supplementary Table 2). Of these proteins, we have confirmed an RNA-dependent interaction between RBMS1 and two well-characterized RBPs that predominantly bind 3' UTRs, PABPC1 and ELAVL1 (Fig. 3E). These findings suggest that RBMS1 stabilizes its targets at least partially in concert with ELAVL1 which preferentially binds A/U-rich elements (AREs) (Hinman, 2008). In line with this hypothesis, we found a significant enrichment of poly(U) motifs in the 3' UTRs of RBMS1 targets (Fig. 3F). As shown in Fig. 3G, 72% of RBMS1-bound 3' UTRs are also bound by ELAVL1 in vivo (as determined by PAR-CLIP (Lebedeva, 2016)). Furthermore, knockdown of ELAVL1 in

SW620 colon cancer cells (of the same genetic background as SW480 cells) resulted in a significant downregulation of RBMS1 targets, as determined by RNA-seq (Fischl, 2019) (Fig. 3H). Together, these results indicate that direct binding of RBMS1 to mRNA 3' UTRs, together with other stabilizing factors such as ELAVL1, results in increased mRNA stability.

RBMS1 is a suppressor of EMT and metastatic liver colonization in colon cancer cells

In order to carry out a functional study of RBMS1 and its downstream regulon in CRC metastasis, we picked an additional cell line model to complement the SW480 line. We chose the LS174T line because RBMS1 is almost completely silenced in this line as measured by qPCR and Western blotting (Supplementary Fig. 4A-B). LS174T is also an established xenograft model and is considered a highly liver metastatic cell line, ~100x more metastatic than the SW480 line as measured by liver colonization assays (Loo, 2015). We first performed RNA sequencing of both SW480 and LS174T lines and consistent with RBMS1 silencing in the LS174T cells, we observed a significant reduction in the expression of the RBMS1 regulon in this line (Supplementary Fig. 4C).

To establish a causal link between RBMS1 silencing and higher metastatic capacity, we performed liver colonization assays by splenically injecting RBMS1 knockdown and control SW480 cells and measuring metastatic burden in the livers of mice over time. While RBMS1 silencing did not have a strong effect on in vitro cell proliferation (Supplementary Fig. 4D), we observed a significant increase in liver colonization upon silencing RBMS1, based on in vivo bioluminescence measurements and gross liver mass at the experimental endpoint (Fig. 4A; Supplementary Fig. 4E). To control for possible off-target effects of the shRNAs, we repeated the experiment with an additional RBMS1-targeting hairpin and observed a similarly significant increase in metastatic liver colonization (Supplementary Fig. 4F). We also performed a gain-of-function experiment by expressing RBMS1 in the LS174T line, where RBMS1 is endogenously silenced. Consistent with our earlier findings, we observed that expressing RBMS1 in LS174T cells results in a marked reduction in their liver colonization capacity (Fig. 4B; Supplementary

Fig. 4G). It should be noted that as expected, expressing shRNAs targeting RBMS1 in highly metastatic cells in which RBMS1 is already silenced, namely LS174T, HCT116, and Colo320, did not have an impact on their liver colonization capacity in xenograft mouse assays, further supporting the on-target effect of the shRNAs used in these experiments (Supplementary Fig. 4H). In contrast, knockdown of RBMS1 in WiDr cells, in which RBMS1 is endogenously expressed, resulted in a significant increase in liver colonization in xenograft mouse models without an overt effect on in vitro cell proliferation (Supplementary Fig. 4I-J).

The metastatic cascade is a complex multistep process, involving cellular processes that not only impact cancer cell growth and survival, but also cell migration and invasion. As mentioned earlier, we did not observe a significant change in in vitro cell proliferation rates upon RBMS1 knockdown. Additionally, trans-well invasion assays of RBMS1 knockdown and control cells did not detect a significant role for RBMS1 in cancer cell invasion (Supplementary Fig. 4K). Therefore, to gain a better understanding of the key step(s) in the metastatic cascade that are regulated by RBMS1, we performed a systematic analysis of known and predicted gene-sets to identify those that may be modulated upon RBMS1 silencing. We performed this analysis using iPAGE, an information-theoretic framework we have developed for this type of analysis (Goodarzi, 2009) (Supplementary Fig. 4L). Downregulation of genes associated with negative regulation of epithelial-mesenchymal transition (EMT(-) signature gene set defined by various studies (Taube , 2010, Onder, 2008, Jechlinger, 2003) was the most significant pathway identified in RBMS1 knockdown cells (Fig. 4C; Supplementary Fig. 4L-M). For example, the canonical EMT marker E-Cadherin (CDH1) is significantly downregulated by ~2-fold in RBMS1 knockdown cells based on RNA-seq data (Supplementary Fig. 2B). We confirmed this by performing immunofluorescence staining for E-Cadherin in RBMS1 knockdown and control cells. We observed both a reduction in E-cadherin expression in RBMS1 knockdown cells, and also noted the appearance of spindle-like cell morphology that is associated with EMT (Fig. 4D). Moreover, the EMT(-) signature genes were also expressed at relatively lower levels in the

LS174T cell line, where RBMS1 is endogenously silenced (Supplementary Fig. 4M). Finally, analysis of the TCGA-COAD dataset revealed significantly lower E-cadherin expression in tumor samples with low RBMS1 expression (~20% reduction, P=0.006) and a significant general correlation between RBMS1 and E-Cadherin expression (Rho=0.1, P=0.03), which is indicative of the clinical relevance of this regulatory axis in CRC.

AKAP12 and SDCBP are functional downstream targets of RBMS1

In order to identify genes that are regulated by RBMS1 and act as suppressors of metastasis in CRC, we used an integrated analytical approach, which relies on mining relevant datasets to prioritize target genes based on their direct interaction with RBMS1 (3' UTR binding), their RBMS1-dependent magnitude of change in expression, the robustness of their response to RBMS1 silencing, as well as their lower expression in metastatic clinical samples. Using this approach, we prioritized two target genes, AKAP12 and SDCBP, that satisfy these criteria: (i) destabilized and downregulated in highly metastatic CRC cells and PDXs, (ii) destabilized and downregulated in RBMS1 knockdown cell lines, (iii) directly bound by RBMS1 based on our irCLIP data (Fig. 3A), and (iv) downregulated in liver metastases relative to primary colon cancers in a publicly available dataset (Sheffer, 2009). Consistently, silencing RBMS1 in SW480 cells resulted in the downregulation and destabilization of AKAP12 and SDCBP mRNAs (Fig. 5A-B). Conversely, overexpression of RBMS1 in LS174T cells resulted in upregulation and stabilization of these targets (Fig. 5A-B). Additionally, in line with the concerted action of RBMS1 and ELAVL1, both AKAP12 and SDCBP 3' UTRs are bound by ELAVL1 in vivo (Lebedeva, 2016). These observations establish AKAP12 and SDCBP as direct targets of RBMS1 that are post-transcriptionally regulated by this RBP.

To test the possible role of AKAP12 and SDCBP in driving CRC metastatic progression, we silenced them individually in SW480 cells using CRISPR interference (CRISPRi) (Horlbeck, 2016) and performed liver colonization assays. As shown in Supplementary Fig. 5A, silencing these genes did not have a significant impact on in vitro proliferation of SW480 cells. However,

reduced expression of these genes resulted in increased metastatic liver colonization in mice (Fig. 5C). To confirm that AKAP12, which, of the two targets tested, elicited a stronger phenotype when silenced, indeed functions downstream of RBMS1, we performed an in vivo epistasis experiment by generating cells with knockdown of both AKAP12 and RBMS1 in SW480 cells. As shown in Supplementary Fig. 5B, unlike in Fig. 5C, further silencing of AKAP12 failed to increase metastatic liver colonization in RBMS1 knockdown cells, in which AKAP12 is already downregulated. To evaluate the contribution of AKAP12 to the phenotype of RBMS1 silencing, we compared AKAP12 knockdown and the double knockdown cells. As shown in Supplementary Fig. 5C, we observed a slight additional increase (albeit not statistically significant) in metastatic liver colonization upon RBMS1 silencing in AKAP12 knockdown cells. This suggests that AKAP12 is a major effector downstream of RBMS1, and that other downstream targets, such as SDCBP, may also play minor roles in this process. Importantly, and consistent with AKAP12 acting downstream of RBMS1, gene expression profiling of AKAP12 knockdown and control cells revealed a significant induction of an EMT signature (Fig. 5D). Notably, this included the upregulation of canonical EMT transcriptional repressors, SNAI1 and SNAI2 (Supplementary Fig. 5D).

As an orthogonal approach, to assess if the effects of RBMS1 silencing on the transcriptome are reflected in the proteome, we used TMT labelling and tandem mass spectrometry to compare the global proteomes of SW480 cells with shRNA-mediated RBMS1 knockdown and control cells (Supplementary Fig. 5E, Supplementary Table 2). As shown in Supplementary Fig. 5F, we observed a positive and significant correlation between changes in mRNA abundance (as determined by RNA-seq) and changes in protein abundance. We found that RBMS1 targets, as determined by irCLIP, were enriched among the downregulated proteins in the RBMS1 knockdown cells (Supplementary Fig. 5G). Most notably, AKAP12, which we have identified as the downstream effector of RBMS1 in EMT regulation, was significantly downregulated at the protein level as well, mirroring our transcriptomic results (Supplementary

Fig. 5H). Taken together, our findings demonstrate that the RBMS1-AKAP12 regulatory axis acts as a suppressor of EMT and liver metastasis in models of CRC progression.

RBMS1 silencing and the downregulation of its targets is associated with CRC progression:

To further assess the clinical relevance of this previously unknown regulatory pathway, we performed a variety of measurements in clinical samples as well as analysis of clinical data to evaluate RBMS1 activity in CRC metastasis. First, we performed qRT-PCR for RBMS1 in two independent clinical cohorts, one cohort stratifying patients based on their tumor stage (n=96), and another comparing samples from normal mucosa, primary CRC tumors, and liver metastases (n=91). As shown in Fig. 6A and 6B, we observed a significant reduction in RBMS1 expression as the disease progresses, with stage IV clinical samples showing the lowest RBMS1 expression levels. We also carried out survival analysis in a large clinical cohort with publicly available expression data and clinical outcomes (Marisa, 2013). We observed a significant association between RBMS1 silencing and reduced relapse-free survival as well as overall survival in colon cancer patients (Fig. 6C). A multivariate Cox proportional-hazards model revealed that this association with RBMS1 silencing remains statistically significant even when controlled for other known clinical factors that may contribute to patient survival (Supplementary Fig. 6A). This observation emphasizes the relevance of RBMS1 silencing as an effective read-out for risk stratification of patients based on samples collected from their primary tumors.

Our results indicate that AKAP12 acts as a suppressor of metastasis downstream of RBMS1, and therefore is expected to show a similar association with metastatic disease. To test this hypothesis, we performed qRT-PCR measurements in the clinical cohorts mentioned above, and we observed a significant reduction in AKAP12 expression as a function of disease progression (Fig. 6D-E). Consistent with RBMS1 acting as a regulator of AKAP12, we observed a highly positive and significant correlation between the expression of these two genes (Supplementary Fig. 6B). Interestingly, our analyses indicate that the identified RBMS1 targets

provide a robust gene signature that, similar to RBMS1, is highly informative of clinical outcomes. For this analysis, we defined an RBMS1 target score as an aggregate measure of expression for the RBMS1 80-gene signature set (average normalized expression of target mRNAs that (i) interact in vivo with RBMS1 on or within 200 nucleotides of their 3' UTRs, and (ii) are downregulated upon RBMS1 knockdown; this is referred to as the RBMS1 80-gene signature hereafter (Supplementary Table 1)). We then stratified patients based on this score and we observed that lower expression of the RBMS1 80-gene signature is associated with lower relapse-free and overall survival in colon cancer patients (Fig. 6F). Consistently, the RBMS1 signature score remained a significant covariate in a multivariate Cox proportional-hazards model of relapse-free survival (Supplementary Fig. 6C). Moreover, we have not observed significant association between the RBMS1 signature score and validated clinico-pathological features or other molecular markers (such as microsatellite instability (MSI) status, left- or right-sidedness, or CRC consensus molecular subtypes (CMS)). Importantly, similar to our initial observation with the RBMS1 regulon (Fig. 1D), the RBMS1 signature score was significantly correlated with RBMS1 expression in a large TCGA pan-cancer dataset (Fig. 6G). We also observed a high correlation between the RBMS1 signature score and ELAVL1 expression (Fig. 6H), consistent with RBMS1 and ELAVL1 acting in concert to promote mRNA stabilization. The association between the RBMS1 regulon and CRC metastasis is not limited to a single dataset. Consistent with our initial observations in matched samples (Fig. 1G), comparison of gene expression changes in liver metastases relative to CRC primary tumors in a publicly available dataset (Del Rio, 2007), revealed a significant reduction in the expression of the RBMS1 regulon (Supplementary Fig. 6D-E). Furthermore, analysis of publicly available RNA-seq data using REMBRANDTS to infer changes in mRNA stability from a set of matched primary and metastatic samples (Kim, 2014), also confirmed (i) RBMS1 silencing in multiple liver metastases and (ii) a concordant reduction in the stability of the RBMS1 regulon (Supplementary Fig. 6F). Together, these results establish the clinical relevance of our findings

and the importance of RBMS1 silencing both as a prognostic factor and as a suppressor of liver metastasis in patients.

HDAC1-mediated promoter deacetylation leads to RBMS1 silencing in LS174T cells

Our findings described above establish a previously unknown regulatory pathway driven by the RNA-binding protein RBMS1 that plays a functional role in CRC metastasis to liver. Regulatory pathways can be expanded by uncovering the upstream pathways that influence their activity. We have developed computational and data analytical tools designed to integrate publicly available datasets from a variety of biological sources to identify such upstream regulatory mechanisms (Goodarzi, 2009). Using these tools, we sought to extend the RBMS1 regulatory pathway by exploring the mechanisms for RBMS1 silencing observed in the highly metastatic LS174T colon cancer cells. First, we found RBMS1 expression to be strongly associated with promoter acetylation, but did not identify any specific transcription factors associated with RBMS1 expression (Fig. 7A). Consistent with this observation, analysis of the connectivity map dataset (Lamb, 2006), which reports the impact of hundreds of small molecule treatments on gene expression, identified the histone deacetylase inhibitor Trichostatin A (TSA) as an activator of RBMS1 expression and of the RBMS1 regulon (Supplementary Fig. 7A-B). As RBMS1 is endogenously silenced in LS174T cells and expressed in the SW480 line, we tested the relative acetylation levels of the RBMS1 promoter in these two lines by performing H3K27Ac ChIP-qPCR. This data showed relatively low levels of H3K27 acetylation at the RBMS1 promoter in the LS174T line compared to the SW480 line, consistent with the differences in the level of RBMS1 in the two lines (Fig. 7B). TSA simultaneously inhibits multiple HDACs, however, we noted that among this group, HDAC1 is the most significantly upregulated in highly metastatic cells generally, and in LS174T cells compared to SW480 cells in particular (Fig. 7C-D; Supplementary Fig. 7C). Moreover, we observed that RBMS1 and HDAC1 expression are negatively correlated in colon cancer clinical samples (Supplementary Fig. 7D). Furthermore, we observed that silencing HDAC1 increases RBMS1 expression (Fig. 7E; Supplementary Fig.

7E-F). Increased expression of HDAC1, and other class I HDACs, has been reported as a strong predictor of survival in colon cancer patients (Weichert, 2008) and modulations in HDAC activity result in widespread gene expression reprogramming, including repression of tumor suppressor genes (Li, 2016). Our observations here indicate that HDAC1-mediated transcriptional repression may be one possible mechanism for RBMS1 silencing in colon cancer cells.

Discussion

Complex human diseases, including cancer, often accompany broad reprogramming of gene expression. In these cases, a comprehensive understanding of the disease state requires not only the identification of the differentially expressed genes, but also understanding the underlying regulatory pathways that explain the dysregulated expression patterns. A number of approaches have been developed by us (Goodarzi, 2009, Goodarzi, 2012) and others (Elemento, 2007, Lefebvre, 2010) to tackle this problem. These methods formalize the association between expression and/or activity of master regulators with those of their regulons. Since direct and indirect associations are challenging to disentangle, knowledge of binding preference (in vitro or in vivo) is also used to better define putative RNA-RBP interactions. Here, we have introduced PRADA to facilitate and formalize the discovery of post-transcriptional regulators that are involved in normal cell physiology and disease. RNA-binding proteins fall into families with highly similar binding preferences (Ray, 2013) and therefore have similar putative regulons. As a result, direct modeling of RBP-target interactions results in unstable predictions (i.e. the model fails to converge on similar outcomes for repeated runs) where the link between a given RBP and its putative regulon is not clear. PRADA solves this problem by integrating the knowledge of changes in the expression of RBPs (used as a proxy for their activities) directly into the analysis. This approach provides a one-step and often stable solution that effectively reveals the RNA-binding proteins whose differential expression is highly informative of changes in gene expression. As RBPs continue to emerge as key regulators of RNA dynamics with

crucial roles in human disease, systematic methods like PRADA provide a suitable approach for studying this class of regulators. In this study, we used this approach to discover a previously unknown regulatory network that counteracts colon cancer progression through RBMS1-mediated RNA stabilization.

We demonstrate that RBMS1 silencing accompanies CRC progression in cell line and PDX models as well as clinical samples, resulting in destabilization of RBMS1-bound transcripts (Fig. 7F). In line with this, we found several well-known RNA-stabilizing factors interacting with RBMS1. Importantly, we found that RBMS1-bound transcripts contain ELAVL1 motifs and binding sites in their 3' UTRs, and that ELAVL1 downregulation resulted in the decrease of RBMS1 target abundance, consistent with a model where RBMS1 and ELAVL1 have an RNA-dependent functional interaction that promotes mRNA stability. Furthermore, we observed that ELAVL1 expression is highly corelated with the expression of the RBMS1 80-gene signature set in the TCGA pan-cancer dataset. ELAVL1 has a well-documented role in stabilizing mRNA by a mechanism that is incompletely characterized. It is suggested that ELAVL1 competes for 3' UTR ARE binding with RNA destabilizing factors, such as TTP or BFR1. It is also proposed that ELAVL1 acts in counteracting microRNA-mediated decay and translational shutdown, thereby indirectly stabilizing mRNAs (Hinman, 2008). ELAVL1 upregulation has been broadly associated with oncogenesis and cancer progression (Schultz, 2020), suggesting a functional interplay between RBMS1 and ELAVL1.

Metastasis often hijacks developmental pathways to reprogram gene expression. EMT is such a pathway with a central role in cancer progression, including CRC (Vu, 2017). Our results show that RBMS1 downregulation results in the suppression of negative EMT regulators, thereby promoting the mesenchymal features of the tumor cells. One of the direct targets of RBMS1, AKAP12, participates in the protein kinase A and C signaling cascades. The association between AKAP12 silencing and colon cancer recurrence and prognosis has been previously described (Xu, 2017). Interestingly, AKAP12 has been described as a negative

regulator of SNAI1 (Cha, 2014), a transcription factor that represses epithelial markers during EMT. In line with this, we found that AKAP12 knockdown caused a global shift towards an EMT transcriptional program, including the upregulation of SNAI1/2. One of the main targets of SNAI1 is the epithelial marker E-cadherin, which is in agreement with our observation that E-cadherin is downregulated in RBMS1-deficient cells and clinical samples. This suggests that the RBMS1-AKAP12-SNAI1/2-E-cadherin axis is a potential route to EMT onset during CRC metastasis.

The Wnt/β-catenin signaling pathway has also been implicated in both EMT and CRC metastasis. β-catenin, a central factor in the canonical Wnt signaling cascade, has a dynamic role: it is known to interact with E-cadherin at the cell membrane, with TCF/LEF transcription factors in the nucleus to activate the downstream transcriptional program, and with APC, a member of the destruction complex, that targets β-catenin for proteasomal degradation in the absence of pathway activation (Schatoff, 2017). APC is a known mutational hotspot in CRC, and the SW480 cells used in this study harbor an inactivating APC mutation, indicating that the Wnt pathway is misregulated in this cell line (El-Bahrawy, 2004). We have shown that RBMS1 silencing results in the downregulation of E-cadherin and have detected upregulated β-catenin levels in RBMS1 knockdown cells (Supplementary Fig. 7G). Interestingly, one of the direct RBMS1 targets validated in this study is syndecan binding protein (SDCBP), which interacts both with Frizzled family proteins (canonical Wnt receptors) and syndecans (Wnt Co-receptors) (Wawrzak, 2009). As our results suggest multiple crossover points, future studies will be required to elucidate the role of RBMS1 in Wnt signaling cascades and their promotion of EMT.

Mirroring our observations from the CRC TCGA dataset, we found the expression of RBMS1 and its signature gene set were highly corelated in the TCGA pan-cancer dataset. However, when examined individually, RBMS1 and its signature gene set expression was informative for specific tumor types. RBMS1 silencing was associated with poor outcome in the CRC cohort, as discussed above (Fig. 6F); in contrast, different outcomes were observed in

other cancer types, ex. in the thyroid cancer cohort, patients with tumors with relatively low expression of RBMS1 had better outcomes (Supplementary Fig. 7H-I). We observed a similar trend in other pan-cancer cohorts, notably in a dataset of metastatic tumors (MET500) (Robinson, 2017). While RBMS1 and its signature expression remained highly corelated, we found reduced RBMS1 levels specifically in colorectal, prostate and gall bladder cancer metastases. Interestingly, when stratified by the metastasis site, RBMS1 expression was significantly downregulated in liver and bone marrow (independently of the primary tumor type), reflecting the principal organotropisms of metastatic CRC (Supplementary Fig. 7J-K). It is thus highly plausible that RBMS1 represents a regulatory node particularly suited for disease progression in the liver, potentially via activation of downstream signaling pathways as discussed above. This finding has implications for other cancer types that can metastasize to the liver.

Our work in the LS174T cell line model indicated that increased activity of histone deacetylase(s), such as HDAC1, could be responsible for RBMS1 silencing¬¬. HDAC inhibitors have been used in the clinic to treat metastatic CRC. A number of studies have previously noted increased expression of HDAC1/2 in colon cancer (Li, 2016) and in cases where RBMS1 silencing also occurs, HDAC inhibitors could be useful as they may also restore RBMS1 expression. While HDAC inhibitors are not specific in their effects, and their effects are not solely mediated through RBMS1, we speculate that in some cases their impact on RBMS1 may lead to better clinical outcomes. Identifying such cases, however, will require a deeper understanding of the mechanisms through which RBMS1 is silenced and the degree to which its expression can be effectively controlled.

Given our limited knowledge of the pathways and processes that regulate the RNA life-cycle in the cell, analytical tools that mine quantitative measurements of mRNA dynamics to identify key regulatory interactions can provide an effective avenue for identifying previously unknown molecular mechanisms with critical functions in health and disease. However, these

computational strategies must be paired with rigorous experimentation to functionally validate and characterize the putative regulons and their regulators. Using one such framework, we have established a novel functional association between RBMS1 silencing and increased liver metastatic capacity in colon cancer.

METHODS

Prioritization of Regulatory Pathways based on Analysis of RNA Dynamics Alterations (PRADA)

PRADA is a customized variation on lasso regression (least absolute shrinkage and selection operator). Therefore, PRADA can simultaneously perform feature selection and apply a custom penalty function as part of its regularization step. The goal of PRADA is to identify RNA-binding proteins whose differential expression explains changes in the expression of their targets that is observed in the data. We first generated an RBP-RNA interaction matrix based on binding preferences of RNA-binding proteins, as previously reported (Ray, 2013, Oikonomou, 2014). For this, we scanned mRNA sequences for matches to RBP-derived regular expression (as previously described (Elemento, 2007)). Given this interaction matrix, our goal is to identify RBPs whose change in expression is predictive of global changes in gene expression of their targets. In other words, we aim to solve the following: ΔExp(g)=∑_i〖α_i·t_(i,g)·ΔExp(〖RBP〗_i)〗+c_g, where ΔExp(g) stands for changes in the expression of gene g based on dataset of interest, t_(i,g) is a binary variable, 1 if RBPi is predicted to bind g and 0 otherwise, ΔExp(〖RBP〗_i) marks changes in the expression of RBPs themselves as a proxy for their differential activity. The resulting α_i represents the strength of regulatory interactions. To ensure that the most informative RBPs are selected, we introduced a customized penalty term that extends the lasso regression framework: min〖1/2n [‖αX-Exp‖+λ∑_i|α_i |/|ΔExp(〖RBP〗_i)|]〗. Here, the coefficients are penalized by |ΔExp(〖RBP〗_i)|^(-1), which is 1⁄|β̂| estimate of the linear model used to compare gene expression between the two groups. Other variations of this penalty can also be used, for example standard effect size can account for confidence in β̂as well (SE⁄|β̂|). This custom penalty term ensures that RBPs whose activity does not change are not selected by the model. This also stabilizes the resulting model, which would otherwise be a major issue as RBPs that belong to the same family often have very similar binding preferences resulting in correlated features in the interaction matrix. After running this regression analysis,

the RBPs with the largest assigned coefficients are prioritized for further study. In this study, we used PRADA to compare poorly and highly metastatic colon cancer lines, which revealed RBMS1 as the strongest candidate. The computational and experimental data flow used in this study is summarized in Supplementary Material 1. The PRADA benchmarking summary is available in Supplementary Material 2. All code and notebooks are available on Github at www.github.com/goodarzilab/PRADA.

Tissue Culture

SW480, LS174T, WiDr, HCT116, and COLO320 cell lines were obtained from ATCC. 293LTV cells were from Cell Biolabs. All cell lines were routinely screened for Mycoplasma infection by PCR (once a month in general and before every animal experiment in particular). SW480 and HCT116 cells were cultured in McCoy's 5A supplemented with 10% FBS, sodium pyruvate, and L-glutamine. LS174T and 293LTV cells were cultured in DMEM supplemented with 10% FBS, COLO320 cells were cultured in RPMI-1640 supplemented with 10% FBS, and WiDr cells were grown in EMEM supplemented with 10% FBS. All commonly used cell lines were routinely authenticated using STR profiling at UC Berkeley Sequencing Facility.

Animal studies

All animal studies were performed according to a protocol approved by UCSF Institutional Animal Care and Use Committee (AN179718). NOD/SCID gamma male mice (The Jackson Laboratory), aged 8 to 10 weeks, were used in all experiments. For splenic (portal circulation) injections, cells were injected directly into the spleen followed by splenectomy (250k for SW480 and LS174T lines and 100k for WiDr cells). In vivo bioluminescence was measured by retro-orbital injection of luciferin (Perkin Elmer) followed by imaging with an IVIS instrument (Perkin Elmer). For ex vivo liver imaging, mice were injected with luciferin prior to liver extraction, and the liver was then imaged and weighed after rinsing with PBS.

RBMS1 irCLIP

RBMS1 irCLIP was performed as described (Zarnegar, 2016).

RNA-seq library preparation

Unless otherwise specified below, RNA sequencing libraries were prepared using RNA that had been rRNA depleted using Ribo-Zero Gold (Illumina) followed by ScriptSeq-v2 (Illumina), and sequenced on an Illumina HiSeq4000 at UCSF Center for Advance Technologies. Matched primary and liver metastases were sequenced using SENSE RNA-seq library preparation kit (Lexogen). AKAP12 knockdown cells were profiled using QuantSeq 3' mRNA-Seq library prep kit fwd (Lexogen).

To measure RNA stability, SW480 RBMS1 knockdown and control cells were treated with 10 mg/mL α-amanitin (final concentration in the medium). After 9 hours, total RNA was harvested from the cells using the Norgen Cytoplasmic and Nuclear RNA Purification Kit per the manufacturer's protocol. RNA-seq libraries were then prepared and log-fold changes in RNA stability were measured by comparing log(t9/t0) in control and RBMS1 knockdown cells.

Measurements in clinical samples.

For clinical samples, 96 samples across all stages of the disease were acquired from Origene (HCRT104 and HCRT105), 14 normal, 8 stage I, 25 stage II, 32 stage III, and 17 stage IV. HPRT was used as endogenous control, and relative RBMS1 and AKAP12 expression levels were respectively measured across the samples (using the primers listed above). We also extracted RNA from another 100 normal mucosa, primary tumors, and liver metastases. Roughly 90 of these samples yielded sufficient RNA for qPCR, which was carried out as previously described.

Clinical association studies

Patients profiled and documented in GSE39582 were first stratified into two groups based on their RBMS1 expression: silenced (bottom 5%) and expressed (otherwise). 5% was selected as the cut-off because it is close to two standard deviations away from the average RBMS1 expression across all samples. A multivariate Cox Proportional Hazard model (R package survival) was then used to evaluate 5-year disease-free survival. Univariate survival

analyses, both disease-free and overall, were also carried out with this stratification. For the target regulon, the RBMS1 signature score was calculated as the median expression of target genes in each sample (normalized gene expression data). Disease-free survival was defined as the time from diagnosis to relapse or death due to any cause. We performed Cox Proportional Hazards multivariable modeling using the RBMS1 score as a continuous variable. Variables added to the model included stage, microsatellite status, KRAS and BRAF mutation status, and Consensus Molecular Subtypes.

FIGURES

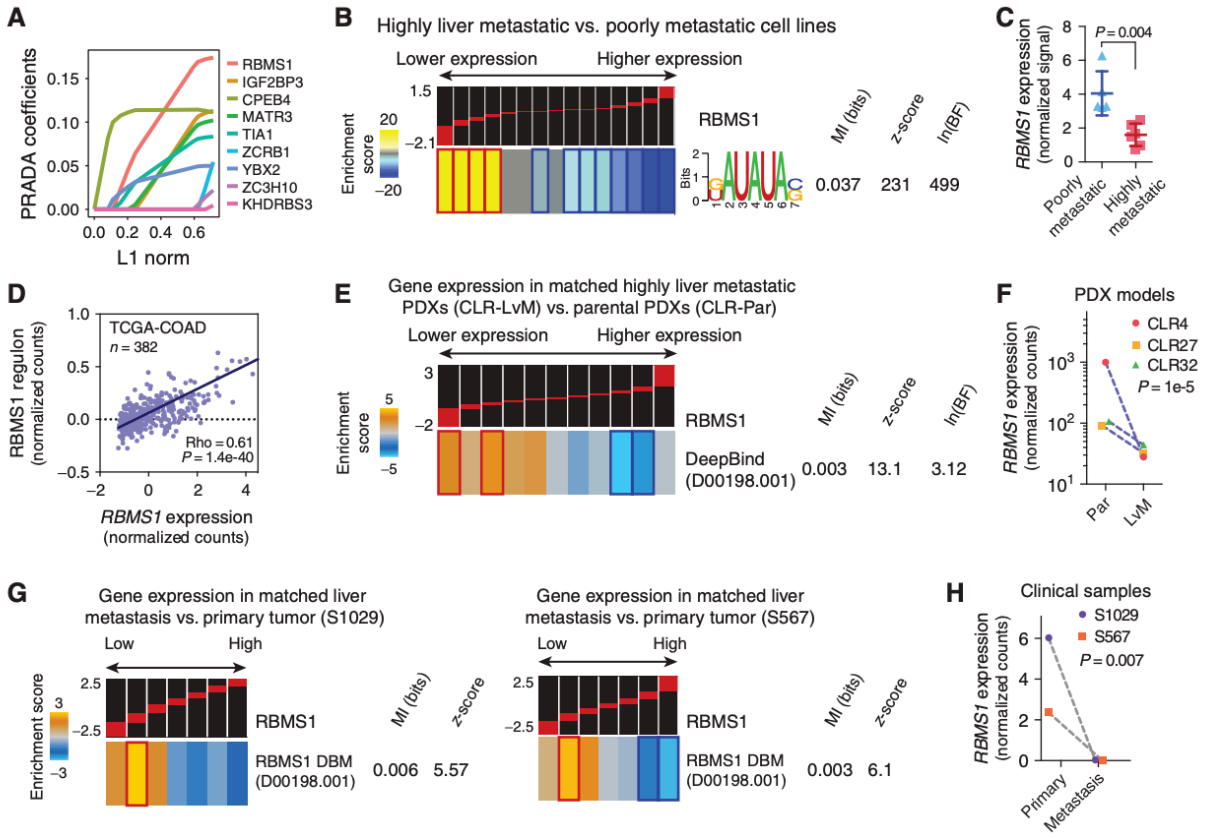


Figure 4.1. RBMS1 silencing in metastatic cells is associated with lower expression of RBMS1 targets. A, Regression coefficients set by PRADA as a function of the l1-norm of the coefficient vector. Each line is associated with an RNA-binding protein and the magnitude of its coefficient is a measure of its strength as a putative regulator of gene expression. Here, the first ten nonzero coefficients are shown as a function of l1-norm (i.e., sum of the magnitude of all coefficients). B, Analysis of RBMS1 recognition sites across gene expression changes between poorly and highly metastatic colon cancer lines. In this analysis, transcripts are first ordered on the basis of their log-fold changes from left (lower expression in metastatic cells) to right (higher expression) and then partitioned into equally populated bins (~1,000 transcripts per expression bin). The red bars on the black background show the range of values in each bin (with the minimum and maximum values, i.e., −1.5 and 2.1, presented on the left). As shown here, genes that are expressed at a lower level in highly metastatic cells were significantly enriched for the RBMS1 binding motif (KAUAUAS; Oikonomou, 2014). In this heat map, gold represents overrepresentation of putative RBMS1 targets whereas blue indicates underrepresentation. Enrichment and depletions that are statistically significant (based on hypergeometric distribution) are marked with red and dark blue borders, respectively. Also included are the logo representation of the RBMS1-binding motif, its mutual information (MI) value, the associated z-score, and the Bayes factor (BF; for details of this analysis, see Goodarzi, 2012). C, RBMS1 expression in colon cancer lines grouped on the basis of their metastatic capacity (Supplementary Fig. S1A). P value calculated using a two-tailed Mann–Whitney U test. D, Linear regression analysis of RBMS1 expression versus average normalized expression of its putative regulon in The Cancer Genome Atlas Colon Adenocarcinoma (TCGA-COAD) dataset (cBioPortal; N = 382). Shown are the Spearman correlation coefficient and the associated P

value. E, Enrichment and depletion patterns of the RBMS1 regulon in PDX models of colorectal cancer liver metastasis. For this analysis, log-fold changes between parental (CLR-Par) and liver metastatic (CLR-LvM) were averaged across three independent PDX models, CLR4, CLR27, and CLR32. D00198.001 is the unique identifier containing the binding site information of human RBMS1 obtained from DeepBind (8) analysis. The distribution of RBMS1 targets was then assessed using mutual information, its associated z-score and Bayes factor. The enrichment and depletion patterns were visualized as described in B. F, Relative RBMS1 levels in matched poorly metastatic (Par) and highly liver metastatic (LvM) derivatives for three PDX models (CLR4, CLR32, and CLR27). P value was calculated using DESeq2 (Love, 2014). G and H, Expression of RBMS1 and RBMS1 targets in matched primary tumor and liver metastases from two patients (S1029 and S567). The results are presented as described in B. The P value for RBMS1 silencing was calculated using DESeq2 (controlled for genetic background).

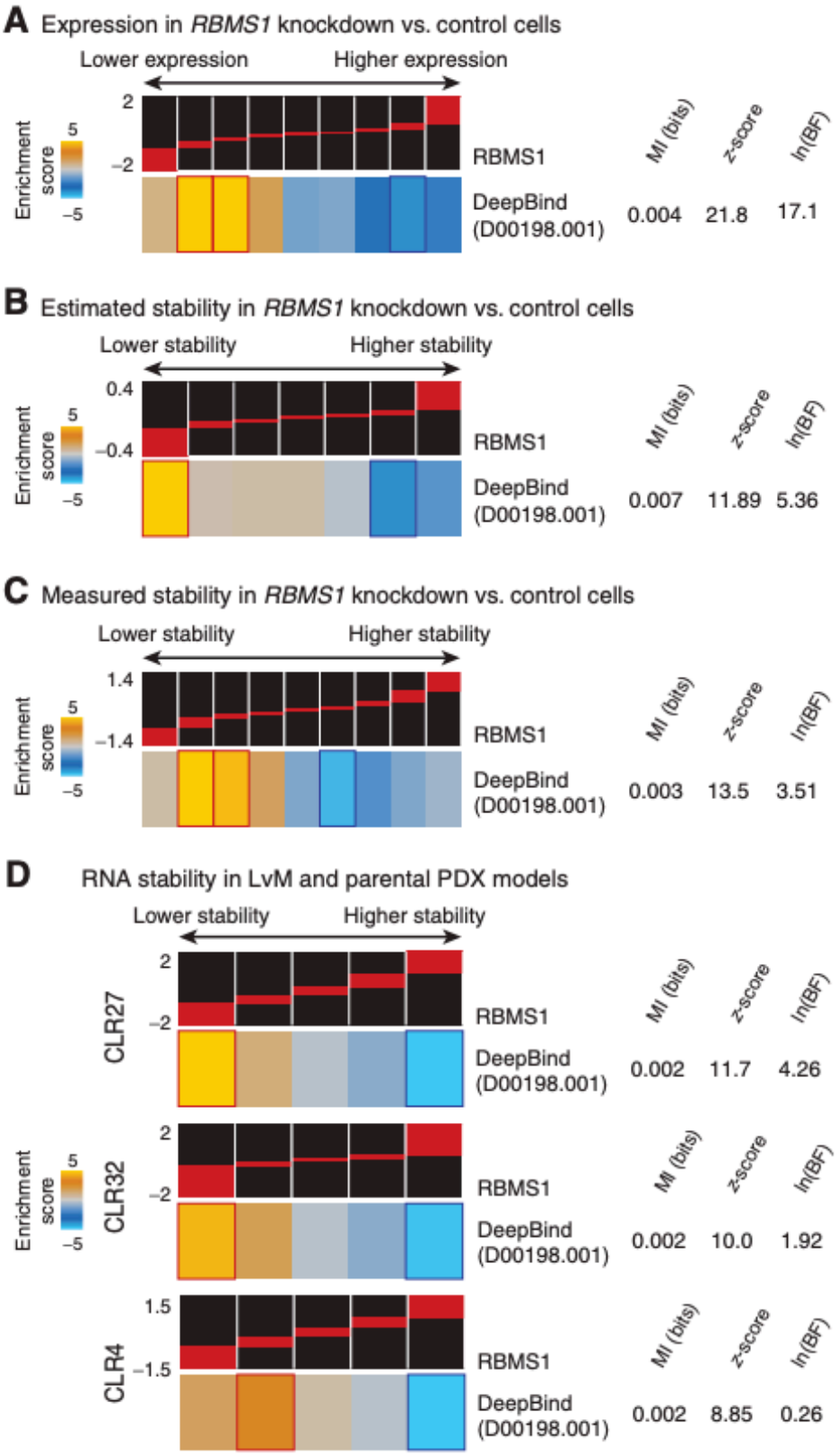

A Expression in RBMS1 knockdown vs. control cells
Lower expression
Higher expression
2
−2
Enrichment score
5
−5
RBMS1
DeepBind (D00198.001)
MI (bits)
z-score
ln(BF)
0.004
21.8
17.1
B Estimated stability in RBMS1 knockdown vs. control cells
Lower stability
Higher stability
0.4
−0.4
RBMS1
DeepBind (D00198.001)
MI (bits)
z-score
ln(BF)
0.007
11.89
5.36
C Measured stability in RBMS1 knockdown vs. control cells
Lower stability
Higher stability
1.4
−1.4
RBMS1
DeepBind (D00198.001)
MI (bits)
z-score
ln(BF)
0.003
13.5
3.51
D RNA stability in LvM and parental PDX models
Lower stability
Higher stability
CLR27
CLR32
CLR4
1.5
−1.5
RBMS1
DeepBind (D00198.001)
MI (bits)
z-score
ln(BF)
0.002
11.7
4.26
0.002
10.0
1.92
0.002
8.85
0.26

Figure 4.2. RBMS1 post-transcriptionally regulates the stability and expression of its targets. (A) Enrichment and depletion patterns of the RBMS1 regulon in RBMS1 knockdown cells relative to control (~2.5-fold knockdown). (B) We used our computational tool, called REMBRANDTS, to estimate changes in RNA stability upon RBMS1 silencing. These differential stability estimates were then used to assess the enrichment patterns of the RBMS1 targets across the changes in RNA decay. (C) Experimental RNA stability changes were measured using α-amanitin treatment as previously described (1). The enrichment and depletion patterns of the RBMS1 regulon was then assessed among the transcripts destabilized or stabilized upon RBMS1 knockdown. (D) We used REMBRANDTS to measure changes in RNA stability between poorly and highly metastatic PDX models from three independent PDX models (CLR27, CLR32, and CLR4). As shown here, consistent with the silencing of RBMS1 in LvM PDX models (Fig. 1F) and the down-regulation of its regulon (Fig. 1E), the RBMS1 regulon is destabilized in these three independent models of CRC metastasis.

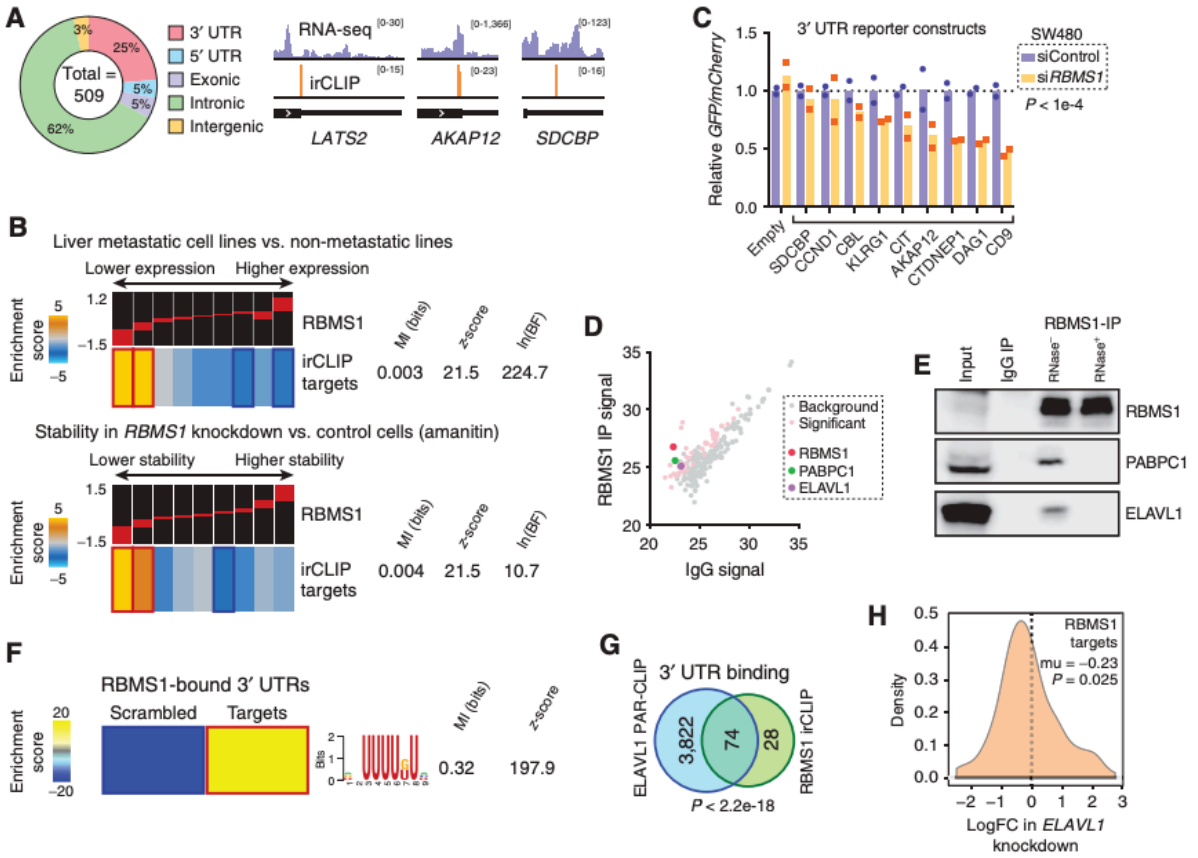


Figure 4.3. RBMS1 irCLIP identifies direct RBMS1 targets in colon cancer cells. (A) 509 RBMS1 binding sites were found using irCLIP, with a significant enrichment of binding to the last exon/3' UTR (relative to the total length of genomic features). Last exons from LATS2, AKAP12, and SDCBP are shown as examples of RBMS1 binding patterns. (B) Enrichment of the RBMS1-bound mRNAs among those that are downregulated in highly metastatic cells (top) and those destabilized upon RBMS1 silencing (bottom). (C) qRT-PCR was used to measure changes in GFP mRNA levels upon cloning RBMS1 binding sites of the listed genes downstream of the GFP ORF. mCherry was expressed from the same bidirectional promoter as GFP, and mCherry levels were used to normalize GFP measurements. A one-sample Wilcoxon signed rank test was used to assess whether the ratios in siRBMS1 samples were significantly below 1.0. (D) Scatter plot of mass spectrometry data showing proteins that co-immunoprecipitate with RBMS1 versus control IgG in SW480 cells. Shown are the average of three replicates across all detected proteins. Proteins enriched in the RBMS1 co-IP samples are shown in pink. RBMS1, PABPC1 and ELAVL1 are highlighted in red, green and violet, respectively. (E) RBMS1, PABPC1 and ELAVL1 were detected by western blot in input and eluate samples from RBMS1 and IgG immunoprecipitations (SW480 cell lysates). For RBMS1 immunoprecipitation, the lysates were additionally treated by RNaseA. (F) Heatmap showing the enrichment of poly(U) sites in the 3' UTRs of RBMS1-bound mRNAs. (G) Venn diagram showing the overlap between RBMS1- and ELAVL1-bound mRNA 3' UTRs. ELAVL1 targets were determined by PAR-CLIP (Lebedeva, 2016). P-value calculated using hypergeometric test. (H) Density plot showing the log fold-change in RBMS1 target expression (determined by RNA-seq (Fischl, 2019)) upon ELAVL1 knockdown in SW620 cells. Median value (mu) is indicated. P-value calculated using one-tailed Wilcoxon signed rank test.

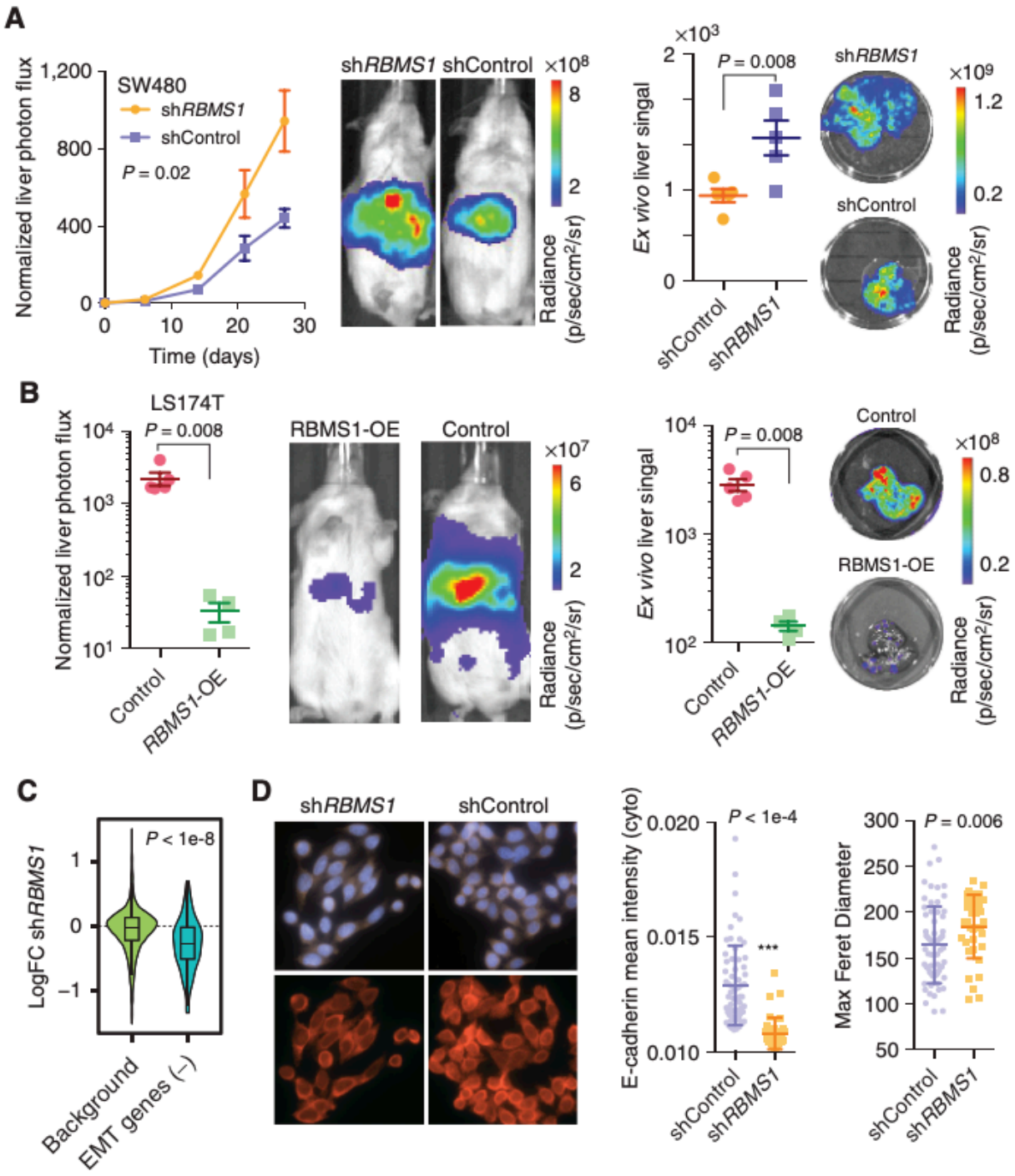


Figure 4.4. RBMS1 is a suppressor of epithelial-mesenchymal transition (EMT) and metastatic liver colonization. (A) Bioluminescence imaging plot of liver colonization by RBMS1 knockdown or control cells; N = 5 in each cohort. Two-way ANOVA was used for statistical testing. Ex vivo liver signal was also measured and compared using a one-tailed Mann-Whitney U-test. Also shown are representative (median signal) mice and livers. (B) Splenic injection of LS174T highly metastatic colon cancer cells overexpressing RBMS1 and those expressing mCherry as a control. Day 21 signal (normalized to day 0) was plotted and compared for both in vivo and ex vivo signal (N = 4-5; Mann-Whitney U-test). (C) Downregulation of EMT(-) signature genes upon

RBMS1 knockdown in SW480 cells. 160-gene EMT(-) signature set (Taube, 2010) was compared to the rest of the transcriptome using a Mann-Whitney U-test. (D) Immunofluorescence staining for E-Cadherin (red) in RBMS1 knockdown and control cells (SW480 background). Note the lower expression of E-Cadherin and the more spindle-like cellular morphology in RBMS1 knockdown (top panels show DAPI signal). ECAD intensity and maximum Feret Diameter (a measure of length of the cell) for cells in control and knockdown samples (N = 64 and 37, respectively). Two-tailed Mann-Whitney U-test was used to compare measurements.

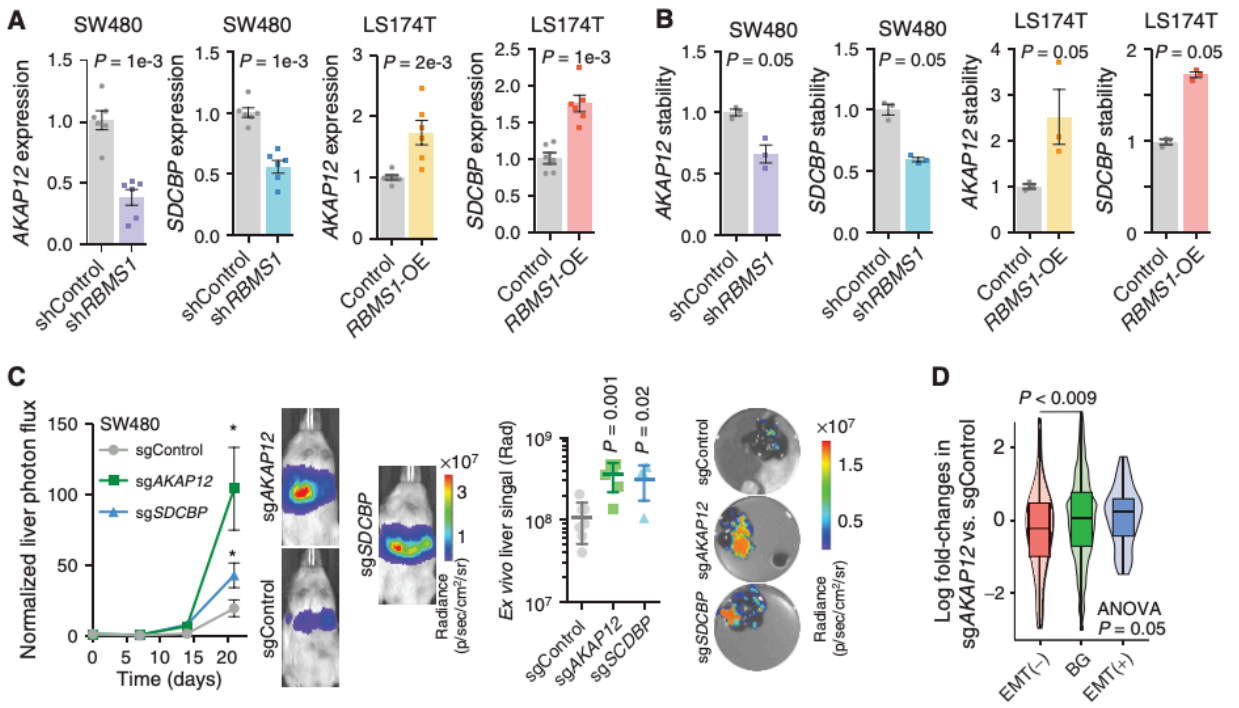


Figure 4.5. AKAP12 and SDCBP act downstream of RBMS1 to suppress CRC metastasis. (A) Changes in the expression of AKAP12 and SDCBP mRNAs upon RBMS1 knockdown in SW480 cells and RBMS1 over-expression in LS174T cells. The expression of target genes was determined by qRT-PCR, normalized to HPRT internal control and shown as relative fold change over shControl or OE-Control. P-value was calculated using one-tailed Mann-Whitney U-test. (B) Changes in the stability of AKAP12 and SDCBP mRNAs upon RBMS1 knockdown in SW480 cells and RBMS1 over-expression in LS174T cells. RNA stability was calculated by comparing mRNA levels with and without treatment with α-amanitin, and shown as relative fold change over shControl or OE-Control. The relative abundance of target genes was determined by qRT-PCR, normalized to 18S RNA (RNA Pol-I transcript insensitive to α-amanitin). P-value was calculated using one-tailed Mann-Whitney U-test. (C) In vivo liver colonization assays were used to measure the impact of CRISPRi-mediated silencing of the RBMS1 targets AKAP12 and SDCBP on liver metastasis (N = 6). Also shown are representative mice from each cohort. Two-way ANOVA was used to compare cohorts to control (P = 0.01 and 0.02 for sgAKAP12 and sgSDCBP respectively). Livers were also extracted and their tumor burden was measured ex vivo. Mann-Whitney U-test was used to compare measurements. (D) The expression of EMT(-) and EMT(+) signature genes relative to background (BG) in AKAP12 knockdown (CRISPRi) and control cells (measurements using 3'-end RNA-seq). Shown are the ANOVA p-value and a Mann-Whitney comparison between EMT(-) and background genes.

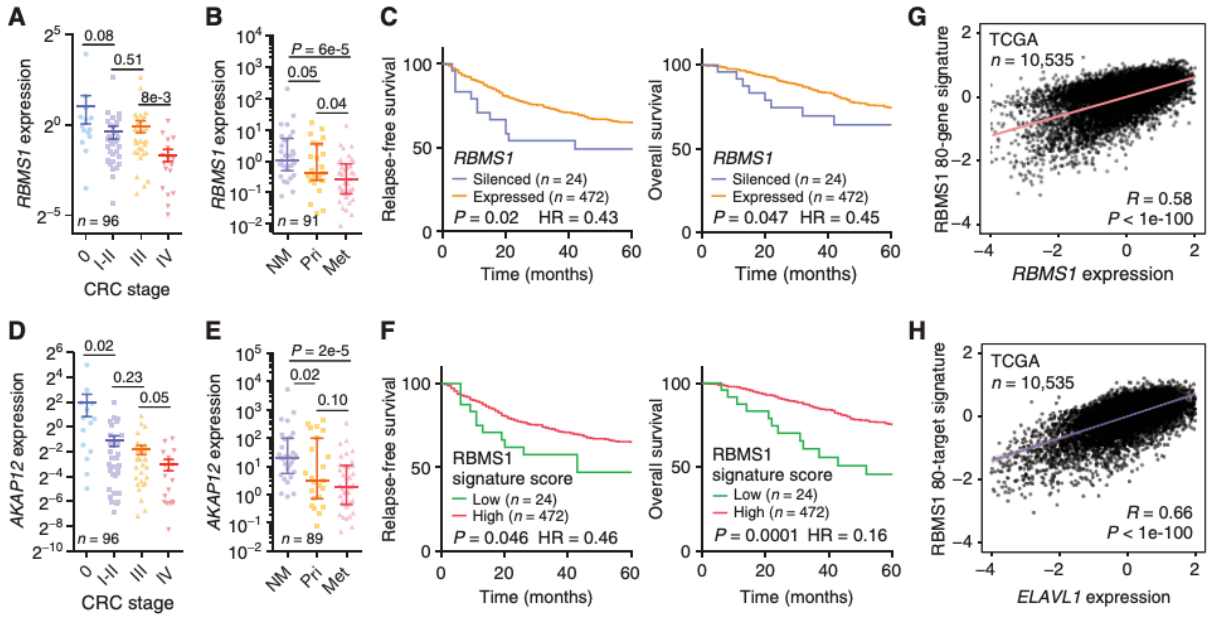


Figure 4.6. RBMS1 and its target gene signature score are associated with colon cancer metastasis and reduced survival in CRC patients. (A) RBMS1 qPCR (relative to HPRT) in 96 colon cancer samples stratified based on tumor stage. (B) RBMS1 qPCR (relative to HPRT) in 29 normal mucosa, 25 primary colon cancer, and 37 liver metastases. (C) RBMS1 silencing was observed in ~5% of primary CRC tumors (see methods), and these patients showed substantially lower relapse-free and overall survival. Reported are Mantel-Haenszel hazard ratios (HR) and p-values from Gehan-Breslow-Wilcoxon tests. (D-E) AKAP12 expression in clinical samples (similar to (A) and (B)). (F) Patient primary tumors were scored based on aggregate expression of RBMS1 signature genes and the resulting values were used to perform survival analyses similar to those in (C). As shown here, lower RBMS1 signature score was significantly associated with lower relapse-free and overall survival. Also reported are Mantel-Haenszel hazard ratios (HR) and p-values from Gehan-Breslow-Wilcoxon tests. (G-H) Regression analysis comparing the expression of RBMS1 (G) or ELAVL1 (H) and RBMS1 80-gene signature set in TCGA pan-cancer dataset. Shown are the Spearman correlation coefficient and the associated p-value.

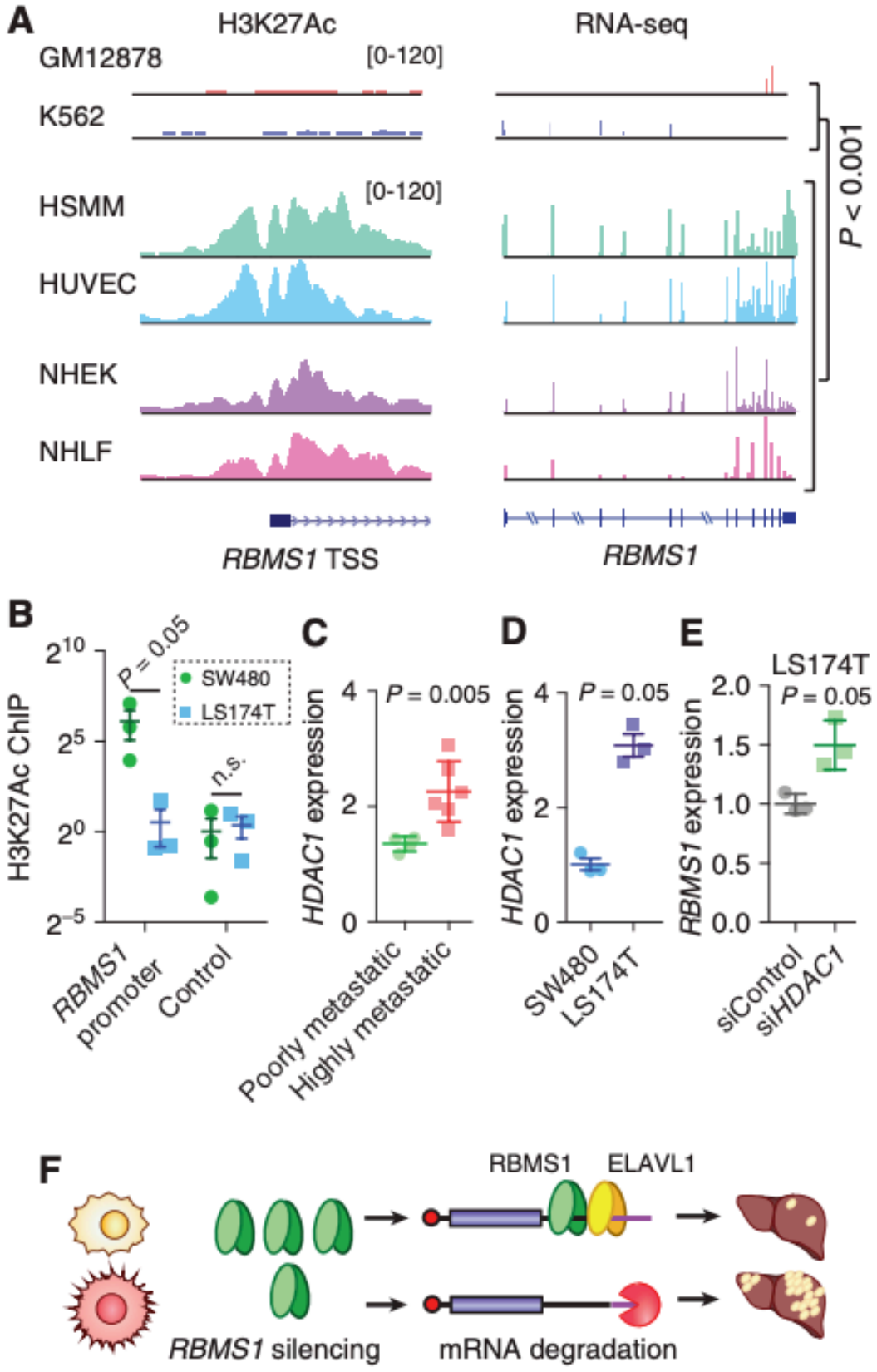

A
H3K27Ac
RNA-seq
GM12878
[0-120]
K562
HSMM
[0-120]
HUVEC
NHEK
NHLF
P < 0.001
RBMS1 TSS
RBMS1
B
H3K27Ac ChIP
P = 0.05
SW480
LS174T
n.s.
RBMS1 promoter
Control
C
HDAC1 expression
P = 0.005
Poorly metastatic
Highly metastatic
D
HDAC1 expression
P = 0.05
SW480
LS174T
E
LS174T
RBMS1 expression
P = 0.05
siControl
siHDAC1
F
RBMS1
ELAVL1
RBMS1 silencing
mRNA degradation

Figure 4.7. HDAC-mediated promoter deacetylation results in RBMS1 silencing. (A) RBMS1 shows dynamic expression and acetylation changes across different cell types. Also shown here is the association between RBMS1 expression and its promoter acetylation (source from ENCODE). (B) RBMS1 promoter acetylation levels were measured in SW480 and LS174T cells using H3K27Ac ChIP-qPCR. An unacetylated region ~40kb away from the promoter was used as control (N = 3). (C) HDAC1 expression in colon cancer lines stratified based on their metastatic capacity. One-tailed U¬-test was used to compare the two groups. Cell line names are listed in Supplementary Fig. 1A. (D) Quantitative PCR to compare HDAC1 levels in the highly metastatic LS174T cells (with silenced RBMS1) relative to poorly metastatic SW480 cells. (E) RT-qPCR was used to measure RBMS1 mRNA levels in LA174T cells with RNAi-mediated HDAC1 silencing. (F) A schematic model of RBMS1 silencing and its role in suppressing colon cancer metastasis. One-tailed Mann-Whitney U tests were used to assess statistical significance for all panels.

REFERENCES

Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol. 2015;33:831–8.

Alkallas R, Fish L, Goodarzi H, Najafabadi HS. Inference of RNA decay rate from transcriptional profiling highlights the regulatory programs of Alzheimer's disease. Nat Commun. 2017;8:909.

Cha J-H, Wee H-J, Seo JH, Ahn BJ, Park J-H, Yang J-M, et al. Prompt meningeal reconstruction mediated by oxygen-sensitive AKAP12 scaffolding protein after central nervous system injury. Nat Commun. 2014;5:4952.

Del Rio M, Molina F, Bascoul-Mollevi C, Copois V, Bibeau F, Chalbos P, et al. Gene expression signature in advanced colorectal cancer patients select drugs and response for the use of leucovorin, fluorouracil, and irinotecan. J Clin Oncol Off J Am Soc Clin Oncol. 2007;25:773–80.

El-Bahrawy M, Poulsom SR, Rowan AJ, Tomlinson IT, Alison MR. Characterization of the E-cadherin/catenin complex in colorectal carcinoma cell lines: E-cadherin/catenin complex and colorectal carcinoma. Int J Exp Pathol. 2004;85:65–74.

Elemento O, Slonim N, Tavazoie S. A universal framework for regulatory element discovery across all Genomes and data types. Mol Cell. 2007;28:337–50.

Fischl H, Neve J, Wang Z, Patel R, Louey A, Tian B, et al. hnRNPC regulates cancer-specific alternative cleavage and polyadenylation profiles. Nucleic Acids Res. 2019;47:7580–91.

Goodarzi H, Elemento O, Tavazoie S. Revealing global regulatory perturbations across human cancers. Mol Cell. 2009;36:900–11.

Goodarzi H, Liu X, Nguyen HCB, Zhang S, Fish L, Tavazoie SF. Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. Cell. 2015;161:790–802.

Goodarzi H, Najafabadi HS, Oikonomou P, Greco TM, Fish L, Salavati R, et al. Systematic discovery of structural elements governing stability of mammalian messenger RNAs. Nature. 2012;485:264–8.

Goodarzi H, Nguyen HCB, Zhang S, Dill BD, Molina H, Tavazoie SF. Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. Cell. 2016;165:1416–27.

Goodarzi H, Zhang S, Buss CG, Fish L, Tavazoie S, Tavazoie SF. Metastasis-suppressor transcript destabilization through TARBP2 binding of mRNA hairpins. Nature. 2014;513:256–60.

Hamada K, Monnai M, Kawai K, Nishime C, Kito C, Miyazaki N, et al. Liver metastasis models of colon cancer for evaluation of drug efficacy using NOD/Shi-scid IL2Rgammanull (NOG) mice. Int J Oncol. 2008;32:153–9.

Hinman MN, Lou H. Diverse molecular functions of Hu proteins. Cell Mol Life Sci. 2008;65:3168–81.

Horlbeck MA, Gilbert LA, Villalta JE, Adamson B, Pak RA, Chen Y, et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. eLife. 2016;5.

Jechlinger M, Grunert S, Tamir IH, Janda E, Lüdemann S, Waerner T, et al. Expression profiling of epithelial plasticity in tumor progression. Oncogene. 2003;22:7155–69.

Kim S-K, Kim S-Y, Kim J-H, Roh SA, Cho D-H, Kim YS, et al. A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. Mol Oncol. 2014;8:1653–66.

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. Science. 2006;313:1929–35.

Lebedeva S, Jens M, Theil K, Schwanhäusser B, Selbach M, Landthaler M, et al. Transcriptome-wide Analysis of Regulatory Interactions of the RNA-Binding Protein HuR. Mol Cell. 2011;43:340–52.

Lefebvre C, Rajbhandari P, Alvarez MJ, Bandaru P, Lim WK, Sato M, et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol Syst Biol. 2010;6:377.

Li Y, Seto E. HDACs and HDAC Inhibitors in Cancer Development and Therapy. Cold Spring Harb Perspect Med. 2016;6:a026831.

Loo JM, Scherl A, Nguyen A, Man FY, Weinberg E, Zeng Z, et al. Extracellular metabolic energetics can promote cancer progression. Cell. 2015;160:393–406.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLoS Med. 2013;10:e1001453.

Oikonomou P, Goodarzi H, Tavazoie S. Systematic identification of regulatory elements in conserved 3' UTRs of human transcripts. Cell Rep. 2014;7:281–92.

Onder TT, Gupta PB, Mani SA, Yang J, Lander ES, Weinberg RA. Loss of E-Cadherin Promotes Metastasis via Multiple Downstream Transcriptional Pathways. Cancer Res. 2008;68:3645–54.

Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, et al. A compendium of RNA-binding motifs for decoding gene regulation. Nature. 2013;499:172–7.

Robinson DR, Wu Y-M, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical genomics of metastatic cancer. Nature. 2017;548:297–303.

Schatoff EM, Leach BI, Dow LE. WNT Signaling and Colorectal Cancer. Curr Colorectal Cancer Rep. 2017;13:101–10.

Schultz CW, Preet R, Dhir T, Dixon DA, Brody JR. Understanding and targeting the disease-related RNA binding protein human antigen R (HuR). WIREs RNA. 2020;11.

Sheffer M, Bacolod MD, Zuk O, Giardina SF, Pincas H, Barany F, et al. Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. Proc Natl Acad Sci U S A. 2009;106:7131–6.

Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68:7–30.

Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, et al. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proc Natl Acad Sci. 2010;107:15449–54.

Vu T, Datta P. Regulation of EMT in Colorectal Cancer: A Culprit in Metastasis. Cancers. 2017;9:171.

Wawrzak D, Luyten A, Lambaerts K, Zimmermann P. Frizzled–PDZ scaffold interactions in the control of Wnt signaling. Adv Enzyme Regul. 2009;49:98–106.

Weichert W, Röske A, Niesporek S, Noske A, Buckendahl A-C, Dietel M, et al. Class I histone deacetylase expression has independent prognostic impact in human colorectal cancer: specific role of class I histone deacetylases in vitro and in vivo. Clin Cancer Res Off J Am Assoc Cancer Res. 2008;14:1669–77.

Xu G, Zhang M, Zhu H, Xu J. A 15-gene signature for prediction of colon cancer recurrence and prognosis based on SVM. Gene. 2017;604:33–40.

Yamaguchi N, Weinberg EM, Nguyen A, Liberti MV, Goodarzi H, Janjigian YY, et al. PCK1 and DHODH drive colorectal cancer liver metastatic colonization and hypoxic growth by promoting nucleotide synthesis. eLife. 2019;8.

Zarnegar BJ, Flynn RA, Shen Y, Do BT, Chang HY, Khavari PA. irCLIP platform for efficient characterization of protein-RNA interactions. Nat Methods. 2016;13:489–92.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Yu, John XuEong
6B520B46E1E148B...
Author Signature

3/15/2022
Date