THE UNIVERSITY OF CHICAGO

ANALYSIS OF GENE EXPRESSION HETEROGENEITY REVEALS NOVEL REGULATORS OF METASTASIS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON CANCER BIOLOGY

BY

DONGBO YANG

CHICAGO, ILLINOIS

DECEMBER 2022

TABLE OF CONTENTS

LIST OF MAIN FIGURES vi

LIST OF TABLES ix

ACKNOWLEDGEMENTS x

ABSTRACT xiii

CHAPTER 1. BACKGROUND 1

1. Challenges of blocking cancer metastasis 1

2. TNBC is highly metastatic: a disease model to study metastasis 3

3. RKIP is a metastasis suppressor 4

4. BACH1 is a metastatic transcription factor 7

5. Elucidating tumor cell transcriptional heterogeneity through single-cell RNA-seq 9

6. Special considerations in scRNA-seq data analysis 14

7. Methods to study single-cell transcriptional heterogeneity 16

8. Summary of outstanding questions 18

CHAPTER 2. MATERIALS AND METHODS 20

1. Cell lines 20

2. Generation of KMT5C overexpression lentiviral vector 21

3. Mouse studies 21

4. Preparation of single cells for RNA sequencing 23

5. Differential expression analyses of single-cell RNA-seq 25

6. Coefficient of variation (CV) analyses 27

7. Fano factor of genes in control versus RKIP cells 28

8. Network analysis 28

9. Gene functional enrichment analyses 29

10. Cell proliferation assays 30

11. Protein isolation and western blots 30

12. Immunohistochemistry 31

13. Bulk tumor RNA-seq 32

14. KMT5C gene expression correlations with motility/invasion and OXPHOS genes across multiple tumor types 33

15. Patient gene co-expression analyses of KMT5C, motility, and OXPHOS genes in breast cancer and pancreatic cancer 33

16. Pathway activities in TNBC patients with increased or decreased levels of BACH1, RKIP, or KMT5C 34

17. KMT5C expression and patient survival 35

18. Additional statistical analyses and software 36

CHAPTER 3. THE METASTASIS SUPPRESSOR RKIP REDUCES OVERALL TRANSCRIPTIONAL VARIABILITY OF TUMOR CELLS 37

Abstract 37

Introduction 38

Results 41

RKIP overexpression reduces transcriptional variability in TNBC 41

RKIP causes changes in gene expression mean and/or variability 43

Functions of genes with CV changes 45

CV analyses enable the identification of novel genes that regulate cancer 46

KMT5C is a clinically relevant regulator of metastasis in breast cancers 48

Discussion 50

CHAPTER 4. THE METASTATIC TRANSCRIPTION FACTOR BACH1 PROMOTES OVERALL GENE EXPRESSION VARIABILITY OF TUMOR CELLS 76

Abstract 76

Introduction 78

Results 80

Overexpression of metastasis promoter BACH1 increases gene expression variability 80

CV analyses of BACH1 overexpression reveal pathways related to metastasis 82

BACH1 knockdown suppresses metastasis while generating transcriptional noise 83

Discussion 86

CHAPTER 5. SUMMARY AND DISCUSSION 103

CHAPTER 6. FUTURE WORK 108

ADDENDUM 112

Supplementary note 112

Supplementary data tables 115

REFERENCES 133

LIST OF FIGURES

Figure 1. RKIP expression significantly decreases gene expression variability in TNBC xenograft tumors. 55

Figure 2. RKIP causes changes in gene expression mean and variability. 56

Figure 3. Functions of genes less variable in BM1-RKIP tumor cells. 57

Figure 4. Genes less variable in BM1-RKIP cells include regulators of metastasis. 60

Figure 5. KMT5C expression correlates with better survival and less metastasis in patients. 61

Figure 6. RKIP inhibits overall gene expression variability in non-metastatic tumor cells and reduces variability while up-regulating KMT5C and OXPHOS genes. 62

Supplementary Figure 1. Results from BM1-CTRL and BM1-RKIP scRNA-seq. 63

Supplementary Figure 2. Gene expression mean and variability of vector control and RKIP-expressing cells. 64

Supplementary Figure 3. Genes with significant changes in mean or CV in RKIP cells versus control cells. 65

Supplementary Figure 4. Significantly enriched gene sets from genes with changes in mean and/or CV in RKIP versus control cells. 66

Supplementary Figure 5. Significantly enriched gene sets from genes with changes in mean or CV in RKIP versus control cells. 67

Supplementary Figure 6. The complete gene co-expression network within RKIP cells from Fig 3A. 68

Supplementary Figure 7. Validation of stable KMT5C expression in TNBC cell lines. 69

Supplementary Figure 8. Effect of KMT5C overexpression in TNBC tumors in mice..... 70

Supplementary Figure 9. Spearman's correlation between KMT5C and genes related to metastasis in cancer patients. 71
Supplementary Figure 10. Relative expression levels of metastasis-related genes in breast cancer patients (A) or pancreatic patients (B) with higher or lower levels of KMT5C. 72
Supplementary Figure 11. Gene differential expression analyses in TNBC patients. 73
Supplementary Figure 12. Significantly enriched gene sets from genes positively correlated with PEBP1, KMT5C, NME1, ARHGDIA, BRMS1, or DRG1 in breast cancer patients. 74
Supplementary Figure 13. Significantly enriched gene sets positively or negatively correlated with KMT5C expression level in breast cancer patients. 75
Figure 7. BM1 tumor cells with BACH1 overexpression have higher overall expression variability versus control cells. 89
Figure 8. BACH1 overexpression causes changes in gene expression mean and variability. 90
Figure 9. Genes with CV changes are involved in pathways related to metastasis. 92
Figure 10. BACH1 knockdown enhances expression noise. 93
Figure 11. BM1 shBACH1 tumor cells exhibit predominantly increased variability without changes in expression mean. 94
Supplementary Figure 14. Results from BM1-CTRL and BM1-mmBACH1 scRNA-seq. 96
Supplementary Figure 15. Gene expression mean and CV of all genes in BM1-CTRL and BM-mmBACH1 cells. 97
Supplementary Figure 16. Co-expression gene network within mmBACH1 cells. 98
Supplementary Figure 17. Gene sets enriched from the largest sub-network in Fig S18. 99

Supplementary Figure 18. Significantly enriched gene sets from genes with changes in mean or CV in mmBACH1 versus CTRL cells. 100
Supplementary Figure 19. Results from BM1-shCTRL and BM1-shBACH1 scRNA-seq. 101
Supplementary Figure 20. Gene expression mean and CV of all genes in BM1-shCTRL and BM1-shBACH1 cells. 102

LIST OF TABLES

Table 1. Genes with significant changes in expression mean or CV in RKIP cells versus control cells. 59

Table 2. Genes with significant changes in expression mean or CV in mmBACH1 cells versus CTRL cells. 91

Table 3. Genes with significant changes in expression mean or CV in shBACH1 cells versus shCTRL cells. 95

Supplementary Table 1. Genes with significantly less CV and up-regulated in BM1-RKIP cells. 115

Supplementary Table 2. Genes with significantly less CV without changes in mean in BM1-RKIP cells. 122

ACKNOWLEDGEMENTS

I would like to thank Dr. Marsha Rosner for her mentorship, foresight, and dedication to science. Through her guidance, I learned how to apply my research skills to different fields of biology, something I had never thought of years ago. Marsha has a lot of foresight, and I'm grateful she is willing to take risks by supporting my "unorthodox" thesis project. She taught me that a scientist is more than the sum of their research experiences. I learned about science communication by building the lab website and the economy of running a lab by comparing and ordering lab supplies. I value these soft skills, which will benefit me immensely regardless of my career path after graduation.

Marsha is the epitome of a true scientist who strives for the betterment of humanity through research. While the COVID pandemic interrupted research projects in many labs, she answered the "call of duty." She started an initiative to discover drugs against SARS-CoV2. Her bravery and selflessness move me, and I am proud to participate in this effort.

I would also like to acknowledge my committee members, Dr. Geoffrey Greene, Dr. Kay Macleod, Dr. Akash Patnaik, and Dr. Sebastian Pott. I am fortunate to have my thesis project advised by a fantastic team of top scientists from bench to bedside and from genomics to oncology. In addition, as the CCB chair who always cared about student development, Kay was one of the reasons that compelled me to pursue my Ph.D. at UChicago. Sebastian played an instrumental role in my thesis project by teaching me single-cell RNA-seq and providing thoughtful comments. Without Seb, I could not have gained much insight from these results. He inspired me to combine wet lab research with bioinformatics. I have benefited so much by having the best of both worlds.

I would like to thank Dr. Gabor Balazsi, Dr. Mengjie Chen, and Dr. Robert Rosner for the helpful discussions. My project has posed a unique set of statistical challenges. Their feedback has helped me understand the data and gain confidence in analyzing them. I would also like to thank Ani Solanki for always being there for the animal studies. In addition, I thank Dr. Yuxuan (Phoenix) Miao for believing in my findings and lending your expertise in epigenetics, immunology, and cancer.

I would like to thank the Ben May department and CCB administration, especially Dr. Barbara Kee, Janie Booker, Heidi Cuesta, and Kathryn Kraynik. I would also like to thank Amy Rapp and Brian Campbell from UChicago GRAD for the career advice and all the help with communication skills. And a big THANK YOU to Dr. Abby Stayart for making the myCHOICE program one of my most profound graduate study experiences. To

I would like to acknowledge my current and former lab mates. Thank you, Ali Yesilkanal, Payal Tiwari, and Andrea Valdespino, for all the dinner parties, movie nights, and of course, the highly productive research over the years. Christopher Dann, Lydia Robinson-Mailman, Thomas Li, and Emily Shi, you are the most talented group of college students I've ever worked with. I'm so glad I was no longer an undergrad when I met you, so I'm spared of all that imposter syndrome. Long Nguyen, Leticia Stock, Maddy Henn, Margarite Matossian, and Wenchao Liu, I'm so glad I'll always get to call you friends before colleagues. Zeyu Qiao, our *de facto* lab member, thank you for indulging me with all the symphonies, operas, and hot pots. I thank my classmates Robert Gruener, Yue Liu, Colin Sheehan, Raven Watson, and Ben Nicholson. You guys are my comrades-in-arms during the Ph.D. journey.

I would also like to thank my friends who've always been by my side over the years. To Allen Cheng, David Li, Shalom Willner, Jason Auman, Andrew Gardell, and Michael Michalik, I have fond memories of all our adventures (and mischiefs). Even though we've ended up in different locations, it is always comforting knowing we can always catch up at any time.

Finally, I would like to dedicate my thesis to my parents, Zhen Dong and Zhaofeng Yang), and my grandparents, Kunying Zhu and Bin Dong. Thank you for all your love and guidance. I am so fortunate to have you every step of my life. Thank you for investing in my education and supporting my dream to become a scientist.

ABSTRACT

Metastasis of cancer cells is the primary cause of death for cancer patients. Current treatments are ineffective at inhibiting metastasis in solid cancers, such as triple-negative breast cancer (TNBC) because some tumor cells evade therapy and cause recurrence later. Tumor cell heterogeneity has been linked to metastatic progression, drug resistance, and survival of cancer patients. We hypothesize that regulators of metastatic progression may also modulate tumor cell heterogeneity. Using a metastasis suppressor, Raf kinase inhibitory protein (RKIP), and a pro-metastatic transcription factor, BACH1, we compared heterogeneity differences in TNBC tumors at different metastatic states. We performed single-cell RNA sequencing to study gene expression variability. In RKIP-overexpressing cells which were non-metastatic, overall expression variability was reduced compared to the highly metastatic control cells, and analyses of genes with reduced variability revealed therapeutic targets. In addition, we showed that KMT5C, a gene less variable and up-regulated in non-metastatic tumor cells, is a novel metastasis suppressor in TNBC. In BACH1-overexpressing cells, overall expression variability increased compared to the control cells, despite BACH1 itself being less variable, consistent with its pro-metastatic role. BACH1 reduced the variability of genes involved in invasion, metastasis, epithelial-mesenchymal transition, and stem cell pluripotency. However, depletion of BACH1 by shRNA acted as a noise generator, increasing overall transcriptional variability despite inhibition of metastasis. These results indicate that regulators of metastasis can impact gene expression variability, but increased heterogeneity alone is insufficient to drive metastasis. Our findings also show that

analyzing both gene expression mean and variability changes between metastatic states can reveal novel anti-metastatic drug targets and regulators of metastasis.

CHAPTER 1. BACKGROUND

1. Challenges of blocking cancer metastasis

Metastasis, the dissemination of tumor cells throughout the body, is responsible for the death of patients in the vast majority of cases (*1*). Driven by stresses such as oxygen and nutrient loss in the microenvironment, cancer cells from the primary tumor invade the local stroma and then enter lymphatic vessels or blood vessels through intravasation. Cancer cells exploit the lymphatic system and the blood circulatory system to migrate to other parts of the body and enter distal organ sites through extravasation. Finally, they colonize the metastatic site by remodeling the microenvironment and promoting angiogenesis to support proliferation, eventually developing into micrometastases (*2*). The spreading of tumor cells to distal sites leads to multiple organ failures, the primary cause of patient death (*1*). Therefore, there is a constant need to develop effective therapies that inhibit metastasis.

However, current treatments for metastatic diseases, such as chemotherapy and radiation, cannot effectively block metastases, and they function primarily as means to delay the inevitable (*3, 4*). The problem is that not all tumor cells respond equally to treatments; some tumor cell subpopulations are more resistant to killing than others (*5*). As a result, despite the appearance of remission after treatment, some tumor cells evade therapy and will cause recurrence later. This non-uniform response to treatment is due to tumor cell heterogeneity, which can manifest as differences in response to stress and the microenvironment across the tumor cell population. Sensitivity to treatment is just one facet of tumor cell heterogeneity at the phenotypic level. It has long been established that

genomic heterogeneity exists via different mutations in different tumor cells. Additionally, differences in gene expression across tumor cells contribute to transcriptional heterogeneity, while chromatin modifications contribute to epigenetic heterogeneity.

Multiple lines of evidence suggest that heterogeneity arising from genetic, epigenetic, and phenotypic sources is a crucial driver of metastatic progression and drug resistance. For example, Sorger and colleagues reported that cells of the same cell line had dramatically different responses to the effects of TRAIL (tumor necrosis factor (TNF)-related apoptosis-inducing ligand) (*6*). While some cells were killed within an hour, others took several hours to die, and a fraction of the cells never entered apoptosis. This phenotypic heterogeneity in resistance to TRAIL was attributed to variations in protein expression levels of major apoptotic regulators across tumor cells.

Heterogeneity was also linked to survival differences in cancer patients. Wolf and colleagues studied patient melanoma samples and found that total mutation numbers or copy number variation (CNV) could not be used to stratify patient survival (*7*). In contrast, tumors with higher intratumoral heterogeneity (ITH), a parameter determined by estimating the number of subclones in a tumor based on mutations and CNV data, were associated with significantly lower survival. Researchers then tested different mouse tumor models by mixing different single-cell clones and found tumors with more ITH had a weaker immune response and higher tumor volume. The authors predicted that higher genomic heterogeneity of tumor cells could create more unique neoantigens, causing the immune response to be "diluted."

While heterogeneity is commonly studied at the genetic level by looking at mutations in single cells, emerging evidence suggests that transcriptional heterogeneity

is also related to metastatic outcomes. Nguyen and colleagues observed that MDA-MB-231 human breast cancer cells contain subclones with varying degrees of morphological variability. Among these, some subclonal populations exhibited a more significant variation of cell size parameters, such as cell shape and size of the nucleus (*8*). Intracardiac injections of these subclones indicated that the cancer cells with higher morphological variations were better at extravasation and colonization. The group identified a spliceosomal gene, SNRNP40, whose variability at transcriptional and translational levels was linked to more metastatic potential in mice.

Given the evidence linking different facets of tumor heterogeneity and metastasis, we hypothesize that regulators of metastasis may modulate gene expression variability and thereby contribute to the metastatic outcome. Identifying and characterizing such modulators of tumor heterogeneity and utilizing them to confer sensitivity to treatment is essential for developing successful therapeutic intervention strategies.

2. TNBC is highly metastatic: a disease model to study metastasis

Breast cancer is the most common type of cancer in women (*9*). It is estimated that around 31% of new cancer cases in women in 2022 are breast cancers. Depending on the status of molecular biomarkers estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki67, breast cancers are categorized into several molecular subtypes: luminal A, luminal B, HER2+, and triple-negative (*10*). Among these subtypes, triple-negative breast cancers (TNBC) represent 10-20% of total breast cancer cases, and they are highly aggressive and difficult to treat. While patients with tumors expressing ER, PR, or both can benefit from endocrine therapy,

and targeted therapy such as Herceptin is available for HER2-positive patients, TNBC patients could not benefit from these effective treatments due to their lack of all three biomarkers. Patients with TNBC typically respond poorly to the standard of care: surgery, radiotherapy, chemotherapy, and PARP inhibitors (*10, 11*). They have an increased chance of recurrence compared to other breast cancers, leading to lower survival rates. In the United States, TNBC disproportionately affects African American women with a three-fold risk increase (*12*). Due to all these factors, developing therapies for TNBC that are better at preventing metastasis is imperative. Various human and mouse cell lines have been developed for TNBC, enabling the investigation of metastatic mechanisms *in vitro* and *in vivo*.

3. RKIP is a metastasis suppressor

One approach to studying the mechanisms of metastasis is through the functions of metastasis suppressors – natural proteins that block metastasis but do not necessarily affect tumor growth (*13*). Metastasis suppressors are expressed in healthy cells and local disease but lost in metastatic disease. They can achieve the same outcome of blocking metastasis by targeting different stages of metastasis: invasion, intravasation, extravasation, or colonization.

The Raf kinase inhibitory protein (RKIP), encoded by *PEBP1* (phosphatidylethanolamine binding protein 1), is an effective metastasis suppressor in many solid tumors, such as breast cancer and prostate cancer (*14, 15*). In prostate cancer, RKIP expression was correlated with loss of the metastatic phenotype in prostate cancer cell lines (*14*). In breast cancers, especially TNBC, RKIP expression does not significantly

alter primary tumor growth (*15*). RKIP targets multiple stages of metastasis in TNBC. RKIP expression in cell culture significantly reduces the invasiveness of TNBC cell lines without affecting cell proliferation. RKIP also blocks tumor cell intravasation, extravasation, and colonization *in vivo* in xenograft mouse models while the primary tumor growth was unaffected. However, RKIP is often lost in metastatic diseases, and loss of RKIP is correlated with poor survival of patients (*16*). Additional research is required to determine how RKIP expression is lost in metastatic tumors. Mutations in RKIP are rare in cancer patients, ruling out loss-of-function mutation as a possible cause for RKIP loss (*17*). How RKIP is silenced seems to be context-dependent, with evidence pointing to epigenetic silencing, transcriptional repression, and protein degradation (*17*). However, the exact mechanisms of RKIP loss in solid tumors remain poorly understood (*17*).

How does RKIP block metastasis? Functionally, RKIP is a kinase inhibitor that down-regulates the MAPK signaling network (*18*). As its name suggests, RKIP was initially discovered as a target that binds to the kinase domain of Raf-1. Later studies by Rosner and colleagues have identified a signaling network through which RKIP inhibits metastasis (*19-22*). RKIP suppresses the MAPK pathway by targeting the Raf-MEK-ERK-Myc signaling cascade. LIN28, an RNA-binding protein that Myc regulates, is down-regulated due to reduced Myc transcription factor binding at its promoter region. Down-regulation of LIN28 leads to increased levels of microRNA let-7, a tumor suppressor microRNA negatively regulated by LIN28. Let-7 inhibits metastasis in part by negatively targeting the transcription factor BACH1 and the chromatin remodeler HMGA2, thereby suppressing metastatic genes such as those involved in epithelial-mesenchymal transition (EMT) and invasion. Yesilkanal and colleagues further showed that RKIP

inhibits activation of the entire MAP kinase network consisting of ERK, JNK, and p38, as well as their upstream and downstream regulators (*23*). This study also identified a group of motility genes inhibited by RKIP, providing further evidence of how RKIP suppresses metastasis.

In addition to inhibition of Raf, RKIP plays an alternative role in the heart by regulating the β-adrenergic receptor (βAR)-protein kinase A (PKA) signaling axis, which regulations cardiac contraction and relaxation in response to stressful situations (*24*). How RKIP executes either role is determined by its phosphorylation state. In its unphosphorylated state, RKIP is a metastasis suppressor by inhibiting Raf. In the heart, protein kinase C (PKC) phosphorylates RKIP at S153, changing its conformation so that it now binds to G protein-coupled receptor kinase (GRK2), leading to activation of the βAR-PKA signaling pathway.

A novel phospho-theft mechanism characterized by the Rosner lab enables RKIP to switch its substrate through single phosphorylation at S153 (*25, 26*). In its unphosphorylated state, the positive charge at K157 allows the lysine to form a salt bridge with positive charges at D134 and E135, which are located on an adjacent peptide chain. Upon phosphorylation at S153, the strong negative charges at pS153 "steal" the K157 residue to form a new salt bridge while breaking the original ones with D134 and E135, respectively. As a result, the local conformation of RKIP is changed. This phospho-theft mechanism differs from the classic notion of phosphorylation as it does not involve breaking and forming salt bridges due to phosphorylation.

Interestingly, the phosphomimetic S153E mutation fails to switch RKIP into the GRK2 binding state (*26*). This is because the single negative charge on the glutamic acid

residue is insufficient to "steal" the K157 residue away from the original salt bridges. The outcome of the S153E mutation is essentially comparable to an S153A mutation which effectively locks RKIP into a Raf binding state. In xenograft mouse models, the RKIP-S153E mutant has an even stronger ability to suppress metastasis of TNBC compared to the wild-type RKIP (*15*).

4. BACH1 is a metastatic transcription factor

The opposite of metastasis suppressors is metastasis promoters. BTB and CNC homology 1 (BACH1) is a transcription factor that promotes metastasis (*19*). BACH1 is indirectly suppressed by RKIP via the RKIP-*let-7* signaling cascade and more directly via the MAP kinase network that induces BACH1 transcription. While RKIP suppresses the pro-metastatic BACH1, RKIP itself is also a direct target of inhibition by BACH1, indicating a mutually repressive interaction between these two regulators of metastasis (*27*). In breast cancer cells, how the RKIP-BACH1 feedback loop plays out can lead to very different metastatic outcomes.

BACH1 plays a vital role in regulating oxygen transport in normal cells (*28*). In erythroid cells, BACH1 is primarily involved in regulating iron metabolism and hemoglobin levels. BACH1, as a transcription factor, inhibits globin genes, which, when combined with the heme prosthetic group, are involved in oxygen transport (*29*). Under normal conditions with an abundance of iron, heme binds to BACH1 to prevent it from binding to DNA, leading to the up-regulation of globin genes and, consequently, increased hemoglobin synthesis (*29, 30*). When iron concentration is low, there is a lack of heme to attach to globin proteins, leading to increased free globin proteins. Too much free globin

proteins can cause cytotoxicity to red blood cells, but this outcome is prevented since BACH1 is now activated due to a lack of heme (*31*). As a result, BACH1 will suppress the transcription of globin genes.

Additional iron-related genes are also regulated by BACH1, such as *HMOX1,* which encodes heme oxygenase-1 (HO-1) (*32*). Under normal conditions, HO-1 is produced and cleaves excessive heme (*33*). In contrast, *HMOX1* is inhibited by BACH1 when there is an iron deficiency, preserving the existing heme pool. BACH1 also regulates oxidative stress through HO-1 (*34, 35*). When excessive reactive oxygen species (ROS) are generated by free heme, BACH1 is inactivated, leading to the up-regulation of HO-1, which functions as an antioxidant by cleaving free heme.

Previous ChIP-seq studies of BACH1 have identified thousands of targets bound by BACH1 due to the relative prevalence of its binding motif “TGCTGAC/GTCA” in the promoter regions (*36*). While BACH1 functions tend to be restricted to iron metabolism and oxidative stress response in normal cells, it has been implicated in cancer progression in multiple cancer types due to its transcriptional regulation of relevant biological processes, notably mitochondrial metabolism and EMT. BACH1 inhibits oxidative phosphorylation and activates glycolytic genes in TNBC (*37*). In non-small cell lung cancer, BACH1 has been shown to promote metastasis via the up-regulation of glycolysis (*38*). BACH1 activates the transcription of *Snail Family Transcriptional Repressor 2* (*SNAI2*) in multiple cancer types, and SNAI2 contributes to EMT by inhibiting the epithelial phenotype (*39-41*). BACH1 also promotes invasiveness by activating MMP1 and ROCK1, which break down the extracellular matrix and enhance migration, respectively (*23, 42*).

Moreover, recent studies have revealed additional functions potentially regulated by BACH1. BACH1 has been shown to promote genes involved in stemness in cancer cells via SNAI2 and TWIST1 transcription factors. However, there is a lack of evidence on BACH1's direct role in cancer stem cells (*43, 44*). There is contradictory evidence of BACH1's regulation of angiogenesis, and it remains to be elucidated whether the regulation of angiogenesis comes from BACH1 in stromal cells or cancer cells (*45-47*). Finally, BACH1 may contribute to tumor progression by inhibiting the mitotic catastrophe of cancer cells via protein-protein interaction with the hyaluronan-mediated motility receptor (HMMR), a novel function of BACH1 that does not involve its function as a transcription factor (*48, 49*). However, more evidence is needed to determine whether this process contributes to cancer metastasis.

In summary, given BACH1's apparent link to metastatic progression and its mutually repressive relationship with RKIP in TNBC, previous studies have utilized BACH1 and RKIP to model different metastatic states of TNBC effectively.

5. Elucidating tumor cell transcriptional heterogeneity through single-cell RNA-seq

While known regulators of metastasis serve as the biological basis for comparisons of different metastatic states, single-cell RNA-sequencing (scRNA-seq) provides the necessary technical foundation enabling quantification of tumor cell heterogeneity at the transcription level. First reported in 2009, scRNA-seq has rapidly advanced over the past decade with new technologies expanding its capabilities by allowing more cells to be profiled from each experiment and higher sequencing depth from each cell (*50*).

At a high level, all scRNA-seq methods work by creating sequencing libraries from individual cells. First, cells are separated from each other to prevent cross-contamination of mRNA. Then, sequencing libraries are generated for each cell similarly to traditional "bulk" RNA-seq and loaded to a sequencing instrument, such as the next-gen sequencers from Illumina. To discern transcriptomes from different cells, unique molecular identifiers (UMIs) are attached to the reads with a unique sequence or combination of sequences assigned to each cell (*51*). Over the years, numerous single-cell platforms have been developed and commercialized, each employing different means to achieve single-cell separation and library generation. For the present study, we evaluated the capabilities of three available single-cell platforms to select the methods best suited for our research objectives: 10X Genomics Chromium (Chromium), Fluidigm C1 Integrated Fluidics Circuit (IFC), and Takara ICELL8 cx Single-cell System (ICELL8).

The 10X Genomics Chromium is a droplet-based scRNA-seq system, with single cells compartmentalized into individual droplets containing all the necessary reagents to carry out the cell lysis and reverse transcription steps (*52*). The sample, in the form of single-cell suspension, is mixed with barcoded gel beads and reverse transcription (RT) reagents using a microfluidic device that adds oil to form an emulsion, in which the water/buffer droplets represent reaction compartments (*53*). The combination of beads and single cells is a random process, resulting in the formation of droplets that may or may not contain beads and cells. Droplets containing a cell and a bead will lyse the cell, dissolve the gel bead to release the barcoded RT primers, and carry out the RT reactions, followed by library generation. This process has a high throughput, allowing thousands of single cells to be sequenced per sample, and up to eight samples can be processed in a

typical experiment. However, at the time of the present study, only 500-1500 genes could be detected from a primary cell (*54*). This limitation in sensitivity restricts its use to sequencing only relatively abundant genes. As a result, the Chromium platform can be used to categorize cell types based on top differentially expressed genes. For example, a popular application is studying the composition of different immune cells (*55*).

The Fluidigm C1 high throughput (HT) IFC is a microfluidics-based device that aims to separate cells into individual chambers for RT and barcoding reactions. Up to two cell samples and reagents are loaded into the input wells on an IFC. Then the IFC is loaded into the C1 system, which uses precise vacuum applications to interface with the IFC to capture single cells and mix reagents. The IFC has 800 chambers allowing the capture of up to 800 cells, though not all chambers will be occupied by single cells during a typical experiment (*56*). Thermal cycling functions built into the C1 system permit cell lysis, RT, and preamplification reactions within the IFC. The intermediate products are then sent into the output wells of the IFC for manual harvesting and downstream library preparation steps. Eight hundred chambers of the IFC are arranged in 40 rows and 20 columns, and sequences from different cells are distinguished by unique combinations of a row barcode and a column barcode. This barcoding technique allows the final transcriptome to be traced back to the exact location of the IFC, which can be helpful when cells are photographed in the IFC prior to lysis. Compared to the Chromium system, the C1 IFC trades sample size for sensitivity, allowing a potentially several-fold improvement in the number of genes sequenced per cell, enabling the detection of lower abundance genes (*54*).

The ICELL8 cx system was developed after the Fluidigm C1 system. It uses yet another approach to compartmentalize single cells using liquid-handling robotics. The traditional approach of single-cell isolation involves the limiting dilution technique that pipettes cells into a 96-well plate with the expectation that some wells will contain a single cell (*57*). The ICELL8 represents an evolution of the technique where robots dispense cell suspensions into a microchip containing 5184 nanowells (*58*). The microchip is then imaged by an automated microscope employing image recognition software to identify all the wells containing valid single cells before cell lysis. RT and barcoding are completed on-chip, using a combination of 72 by 72 barcodes to achieve 5184 unique combinations. Then all the sequences are pooled to generate a single sequencing library from a microchip. Compared to the C1 IFC, ICELL8 offers several advantages. Each ICELL8 experiment accepts up to 8 samples for one microchip versus 2 in a C1 experiment. In addition, with the 5184-well microchip, around one-third of all the wells contain a single cell (limiting dilution of cells into wells resembles a Poisson distribution) (*58, 59*). Thus, the number of single cells sequenced per experiment is over twice as many versus a C1 experiment. Finally, robotics improves experiment consistency by limiting the steps requiring a hands-on operation. All barcoded sequences are pooled into a single mixture for the final generation of one sequencing library. In contrast, the C1 IFC requires generating 20 separate sequencing libraries, one for each column on the IFC, introducing an additional source of variation.

Derived from standard scRNA-seq, single-nuclei RNA-seq (snRNA-seq) analyzes RNA content only in the nuclei instead of fully intact single cells (*60*). snRNA-seq was developed to address challenges in sample preparation. Single nuclei can be extracted

from preserved tissue samples, including formalin-fixed paraffin-embedded (FFPE) blocks unsuitable for scRNA-seq. In scRNA-seq, cells must be dissociated from tissue samples while maintaining viability, during which time the cells are exposed to a non-physiological environment, and biological activities still happen inside the cells, potentially altering signaling pathways and gene expression. With snRNA-seq, tissues can be flash frozen immediately after dissection and before single-nuclei isolation, preventing gene expression changes due to sample preparation steps. Hard-to-isolate cells such as neurons and adipocytes tend to be missed during scRNA-seq sample preparation, while snRNA-seq is better at capturing all cell types from a tissue sample (*61*). Potential challenges for snRNA-seq result from the fact that only nuclear RNA is measured. Nuclear RNA has different compositions than cytoplasmic RNA, leading to potentially different conclusions regarding gene expression (*62*). Sequencing depth also suffers compared to scRNA-seq due to the reduced total RNA collected per cell (*62, 63*).

We use gene expression variability as a measurement of transcriptional heterogeneity. We prefer higher sequencing depth to study gene expression variability for two reasons. First, a larger number of genes can be evaluated for changes in variability, not just the most abundant ones. Second, a higher number of reads per cell increases statistical power and sensitivity. Thus, the C1 and ICELL8 were preferred over Chromium for our objectives. For the first study reported in the present thesis, the C1 was chosen over Chromium, as ICELL8 was not yet available at the time of the experiment. The second study was conducted later; thus, we could take advantage of the increased sample size and sequencing depth benefits of ICELL8. Lastly, we chose scRNA-seq over snRNA-seq since fully processed RNA is more representative of gene expression, and

we could maximize the amount of RNA isolated from every cell to ensure sufficient sequencing depth.

6. Special considerations in scRNA-seq data analysis

Compared to traditional bulk RNA-seq, scRNA-seq poses additional challenges when it comes to data analysis, thus requiring additional considerations during pre-processing or quality check (QC). In a typical study, sequencing data are checked for quality of single-cell transcriptomes, reproducibility, confounding effects such as cell cycle differences, batch effect, and biases caused by varying sequencing depth across single cells (*64, 65*).

Commonly used metrics for single-cell sequencing QC include filtering cells based on the number of genes detected per cell, the number of reads detected per cell, and the contribution of mitochondrial reads to the total read count (*64*). Cells with an exceptionally low number of genes or read counts are of low quality or possibly represent a total lack of cells, such as a droplet with barcodes but no cells. Transcriptomes with an unusually high number of genes or read counts could be caused by the same barcode assigned to more than one cell. Such duplex or multiplex events can happen when cells are not fully dissociated, leading to clumps of two or more cells being captured as a single entity. During sequencing alignment, some reads can be mapped to the mitochondrial genome. Apoptotic cells tend to have more mitochondrial reads sequenced, resulting in an exceedingly high proportion of mitochondrial reads that can be filtered out based on a percentage cutoff. The issue of low-quality or apoptotic cells can be mitigated during the single-cell dissociation step. For example, dead or dying cells from samples can be

removed using antibodies conjugated to the surface of magnetic beads, improving overall cell viability. With techniques such as ICELL8, the cells can be stained for viability and imaged prior to lysis, enabling downstream selection of barcodes linked to viable single cells only.

As described in the previous section, common single-cell platforms limit how many samples can be processed in a given experiment. Multiple single-cell experiments must be run separately, with studies requiring larger sample sizes. Processing samples over multiple runs unavoidably introduces technical variations called the batch effect. A standard method of mitigating the batch effect is to treat the batch number as a variable and then regress it out from the gene expression data of every cell (*66*). Results can also be skewed when groups of samples predominantly fall into different cell cycles (*64*). For example, if Group A cells are primarily in the G1 phase while Group B cells are in the G2 phase, then the observed transcriptional differences between A and B will be dominated by cell cycle differences. Therefore, when a significant difference in the cell cycle is present among the samples, the observed differences between the samples could be dominated by cell cycle-related gene sets, which might obscure other biological features (*64*). Similar to batch effect correction, cell cycle differences can also be regressed out to mitigate the issue.

Variations in sequencing depth may also pose a problem in scRNA-seq data analysis. When many single cells have varying levels of sequencing depth, estimating actual gene expression levels may become biased (*65*). For example, when a cell has a high expression value for a particular gene, it could be that it is up-regulated in that cell or that specific sequences are highly amplified during library prep. Similarly, when a gene

is not sequenced in a cell, it could mean it was not expressed in the cell or was expressed but not sequenced due to low sequencing depth. These confounding effects impose difficulties in accuracy during differential expression analysis of single cells. One approach to mitigate this issue is using statistical modeling to estimate gene expression levels in single cells. The present study utilized a method developed by Kharchenko and colleagues to study gene expression mean changes (*65*). This method uses Bayesian inference to model every cell in the expression dataset and estimate the most likely expression level for each gene in each cell. Expression fold-change for a given gene is the most likely estimate in a 95% confidence interval, while the statistical significance is presented in a false discovery rate (FDR)-corrected Z score with directionality.

Taken together, the pre-processing step of scRNA-seq analysis has substantial implications for downstream analyses. While numerous methods are available for pre-processing, they need to be applied selectively and judiciously to best suit the unique features of each study.

7. Methods to study single-cell transcriptional heterogeneity

Typical applications of scRNA-seq, such as the determination of cell type and pathway differences, are based on comparing the gene expression mean between experimental conditions or clusters of cells. However, as a statistical metric by itself, gene expression mean does not depict the complete picture of transcriptional heterogeneity (*67*). For example, the expression of some genes can vary in a small subset of cells while remaining unchanged in most cells. As a result, while the mean expression for specific genes across the cell population can collectively increase, decrease, or remain

unchanged, this does not indicate whether the mean expression of these genes changed in all or some individual cells.

The coefficient of variation (CV), defined as the standard deviation over the mean, has been used to measure gene expression variability. For example, Desai and colleagues used 5'-iodo-2-deoxyuridine (IdU) to induce noise in mouse embryonic stem cell transcription by perturbing the DNA repair machinery (*68*). IdU increased the CV of many genes without changing their expression mean. However, changes in CV alone were sufficient to potentiate the differentiation of stem cells and the dedifferentiation of fibroblasts. These results indicate that transcriptional variability can drive different functional outcomes.

The Fano factor, defined as the variance over the mean, is an alternative metric to measure variability. The Factor factor is used to study deviations from a Poisson process or intermittent bursting (*69*). Unlike the CV, which is not affected by scaling during gene expression normalization (both the standard deviation and the mean are scaled by the same scaling factor, thus canceling out), the Fano factor does have units. Hence, its value is influenced by the normalization of gene expression (the variance is scaled by the square of the scaling factor), making it difficult to compare Fano factors of genes between different studies (*68*). However, a relative change in the Fano factor between experimental conditions indicates deviations from a Poisson process, with possible causes including altered bursting, feedback regulation, or noise modulators (*67-69*). Both the Fano factor and CV require molecular units to study absolute changes, requiring single molecule quantification for transcriptional studies. For example, RNA FISH is

commonly used to quantify the number of mRNA transcripts within single cells, but typically, only up to four genes can be studied in a given experiment (*69, 70*).

In the present study, we used both expression mean and CV as different statistical metrics to analyze scRNA-seq results and identify changes in gene expression mean or variability across metastatic states. If specific genes reduce expression variability when tumor cells switch from a metastatic state to a non-metastatic state, these genes give further insight into metastatic regulatory mechanisms and could reveal potential therapeutic targets for inhibiting tumor heterogeneity and metastatic progression of the disease.

8. Summary of outstanding questions

Cancer is challenging to treat because it is a heterogeneous disease. Heterogeneity causes metastasis that leads to the death of the patient. The interplay between metastasis regulators changes the metastatic state of tumor cells. Metastasis suppressors such as RKIP inhibit metastasis without affecting tumor growth, while loss of metastasis promoters such as the transcription factor BACH1 may also influence primary tumor growth. It is unclear whether metastasis regulators can alter the heterogeneity of tumor cells when they switch them from one metastatic state to the other. Furthermore, metastatic regulators can be homogeneous or heterogeneous, and how their heterogeneity impacts metastatic outcomes remains to be understood.

In this study, we examined the relationship between metastasis and transcriptional heterogeneity, an aspect of heterogeneity not extensively studied previously but recently connected to the metastatic outcome. We used scRNA-seq to measure transcriptional

heterogeneity in metastatic and non-metastatic TNBC tumor cells. In addition to differential expression, we also used CV to measure how gene expression variability differs between the two metastatic states. To modulate the metastatic state, we over-expressed RKIP to inhibit metastasis. We also used shBACH1 in a separate study to inhibit the metastatic transcription factor non-uniformly. Finally, we over-expressed BACH1 to further increase the metastatic potential of TNBC cells over the wild-type. These systems allow us to answer questions such as (1) does RKIP as a metastasis suppressor reduce transcriptional heterogeneity? (2) does BACH1 as a metastasis promoter increase transcriptional heterogeneity? (3) how does the expression mean and variability of metastasis regulators affect overall transcriptional heterogeneity? and finally (4) for genes that changed heterogeneity in metastatic progression, can we leverage them to gain further insight into genes and pathways that promote metastasis and improve treatment? By answering these questions, we hope to understand how gene expression mean and variability affect metastasis state and identify novel metastasis regulators and potential targets to enhance the efficacy of therapeutic agents to treat metastatic cancer.

CHAPTER 2. MATERIALS AND METHODS

1. Cell lines

The human TNBC cell line BM1 (also known as MDA-MB-231-1833, BoM1), the mouse TNBC cell line 4T1, and the human embryonic kidney epithelial cell line 293T were used in our studies. 4T1 and 293T cells were purchased from American Type Culture Collection (ATCC). BM1 cells were generated by Massagué and colleagues (*21*). BM1-vector control (BM1-VC) and BM1-RKIP-S153E (BM1-RKIP) cells were generated from wild-type BM1 by Dangi-Garimella and colleagues (*15*). Luciferase-expressing BM1 cells (BM1-luc) cell lines were generated from wild-type BM1 by Yesilkanal and colleagues (*23*). BM1-shBACH1 and BM1-shCTRL cell lines were generated from wild-type BM1 by Lee and colleagues (*37*). Stable BM1 and 4T1 cells with vector control or KMT5C overexpression were generated through lentiviral transduction following established protocols (*23*). BM1-CTRL and BM1-mmBACH1 cell lines were generated through lentiviral transduction using the lentiviral vectors generated by Lee and colleagues (*37*). The cloning method for the KMT5C overexpression lentiviral vector is described below. BM1 and 293T cells were cultured in DMEM media (Gibco, 11965-118) with 10% fetal bovine serum (FBS, Avantor, 97068-085) and 1% penicillin-streptomycin (Gibco, 15140-122). 4T1 cells were cultured in RPMI media (Cytiva, SH30027.01) with 10% FBS and 1% penicillin-streptomycin. Cells were grown in 5% CO_2 incubators at 37 °C. All cell lines were authenticated by short tandem repeat (STR) analysis (IDEXX BioAnalytics) and mycoplasma detection (Lonza, LT07-218) routinely.

2. Generation of KMT5C overexpression lentiviral vector

The human KMT5C (NM_032701.4) open reading frame (ORF) was cloned from the pCMV6-Entry expression vector (Origene, RC203881) into the pLV-EF1a-IRES-Hygro lentiviral backbone (Addgene, 85134) with restriction enzymes BamHI-HF (New England BioLabs, R3136S) and MluI-HF (New England BioLabs, R3198S). A stop codon was then added to the end of the ORF through site-directed mutagenesis (New England BioLabs, E0554S) using the following forward and reverse primers: TAACGCGTGAATTCCTCGAG; CAGCTCTTCACCGCCGAC. The resulting pLV-EF1a-IRES-Hygro-KMT5C plasmid was confirmed by Sanger sequencing at the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility. These primers were used: TCAAGCCTCAGACAGTGGTTC; TTGTGCCTGCAGATGGGAACGC.

3. Mouse studies

Mice were acquired and housed by the Animal Resources Center at the University of Chicago. All mouse studies were conducted following protocols approved by the Institutional Animal Care and Use Committee. For primary tumor growth experiments, tumor growth was monitored over time by caliper measurements of the width and lengths of tumors using the following formula for tumor volume calculations (*37*):

$$volume = 0.4 \times width^2 \times length$$

The mice were sacrificed at the experiment endpoint or when the tumor volume reached 1 cm^3, whichever came first.

For single-cell RNA-seq studies, 1×10^6 BM1-VC or BM1-RKIP cells were injected orthotopically into the mammary fat pad of athymic nude mice (Harlan Envigo). The mice were sacrificed three weeks after tumor cell implantation. 1×10^6 BM1-shCTRL or BM1-shBACH1 cells were injected orthotopically into the mammary fat pad of athymic nude mice (Charles River). The mice were sacrificed four weeks after tumor cell implantation. 1×10^6 BM1-CTRL or BM1-mmBACH1 cells were injected orthotopically into the mammary fat pad of athymic nude mice (Charles River). The mice were sacrificed four weeks after tumor cell implantation. Primary tumors were collected for single-cell isolation.

For the metformin treatment study, 1×10^6 BM1-VC or BM1-RKIP cells were injected orthotopically into the mammary fat pad of athymic nude mice (Charles River). When tumors reached 20-30 mm^3 in volume, mice were randomized into groups for treatment with metformin or vehicle. Metformin (MilliporeSigma, PHR1084) was provided in drinking water *ad libitum* at 300 mg kg^{-1} day^{-1}. Tumor growth was monitored until the experiment endpoint was reached.

For BM1 KMT5C primary tumor growth experiments, 1×10^6 BM1-VC or BM1-KMT5C cells were injected orthotopically into the mammary fat pad of athymic nude mice (Charles River). The mice were euthanized four weeks after tumor cell implantation. Primary tumors were collected for RNA sequencing and immunohistochemistry (IHC) analyses.

For BM1 KMT5C colonization assays, 1×10^5 BM1-luc cells with or without KMT5C overexpression were injected into the left ventricle of the heart of athymic nude mice (Charles River) for systemic distribution of tumor cells. Three weeks post-injection, mice were injected with freshly-prepared luciferin solution (Goldbio, LUCK-1G) with a

concentration of 15 mg/ml in PBS, and each mouse received 150 mg luciferin/kg of body weight. Five minutes after luciferin injection, tumor burden was measured as bioluminescence via the IVS Spectrum Imaging System (PerkinElmer), and up to five mice were measured at a time. For each group of mice, bioluminescence (radiance, p/sec/cm^2/sr) was measured continuously until the maximum signal intensity was reached. The mice were euthanized after imaging. The number of metastatic foci for each mouse was determined by counting the number of centroids from the peak radiance measurement. The same settings and radiance scale (1.00e6 – 1.00e7 p/sec/cm^2/sr) were used to evaluate all the mice.

For 4T1 KMT5C primary tumor growth and spontaneous metastasis assays, 1 × 10^4 4T1-VC or 4T1-KMT5C cells were injected orthotopically into the mammary fat pad of BALB/c mice (Charles River). The mice were euthanized four weeks after tumor cell implantation. Primary tumors were collected for immunohistochemistry (IHC) analyses. Lungs were perfused with PBS injected horizontally through the trachea, removed from the thoracic cavity, and then fixed with Bouin's Solution on (Ricca Chemical, RC1120-16). Spontaneous metastases were measured by counting the number of surface metastases on the lungs following fixation.

4. Preparation of single cells for RNA sequencing

For the BM1-VC and BM1-RKIP studies, BM1 cells from the primary tumors were isolated through tumor dissociation, mouse cell depletion, and dead cell removal following the manufacturer's protocols (Miltenyi Biotec, 130-095-929, 130-104-694, 130-090-101). Suspensions of single cells were processed on a Fluidigm C1 instrument using a C1

Single-Cell mRNA Seq HT IFC and Reagent Kit v2 (Fluidigm, 101-4981, 101-3743). cDNA synthesis, barcoding, library preparation, and quality check were performed following the manufacturer's protocol (Fluidigm, PN 101-4964 I1). All single-cell RNA libraries were sequenced by an Illumina HiSeq 4000 to generate paired-end 100bp reads at the University of Chicago Genomics Facility following standard protocols. Each independent replicate consisted of single cells from one BM1-VC tumor sample and one BM1-RKIP tumor sample.

For the BM1-shCTRL and BM1-shBACH1 studies, BM1 cells from the primary tumors were isolated through tumor dissociation, mouse cell depletion, and dead cell removal following the manufacturer's protocols (Miltenyi Biotec, 130-095-929, 130-104-694, 130-090-101). Cells from two CTRL tumors and two mmBACH1 tumors were processed in each independent replicate. Suspensions of single cells were processed for sequencing at the University of Chicago Genomics Facility following standard protocols. Cells were processed on an ICELL8 cx instrument (Takara Bio, 640188) using SMART-Seq chemistry for single-cell full-length transcriptome analysis. The resulting library was sequenced by an Illumina NovaSeq 6000 using one SP flow cell to generate paired-end 100bp reads.

For the BM1-CTRL and BM1-mmBACH1 studies, BM1 cells from the primary tumors were isolated through tumor dissociation and dead cell removal following the manufacturer's protocols (Miltenyi Biotec, 130-095-929, 130-090-101). Cells from two shCTRL tumors and two shBACH1 tumors were processed. Suspensions of single cells were processed for sequencing at the University of Chicago Genomics Facility following standard protocols. Cells were processed on an ICELL8 cx instrument (Takara Bio,

640188) using SMART-Seq chemistry for single-cell full-length transcriptome analysis. The library for each independent replicate was sequenced by an Illumina NovaSeq 6000 in two separate runs, with each run using one lane of an S1 flow cell to generate paired-end 100bp reads. Sequencing reads from the runs were combined before downstream analyses of each independent replicate.

5. Differential expression analyses of single-cell RNA-seq

The RNA sequencing data were demultiplexed using the manufacturer-supplied API script to generate individual FASTQ files for single cells based on the barcodes used (Fluidigm, mRNASeqHT_demultiplex.pl). The reads were mapped to the human genome (UCSC hg38 with GENCODE annotations) using RNA STAR (*71, 72*). The resulting mapped reads from every single cell were counted by featureCounts to generate per-gene read counts (*73*). Gene expression data from three biological replicates were combined, and the batch effect was corrected via the "BatchCorrectedCounts" function in the CountClust package by specifying the replicate number as the variable to regress out (*66*). Gene expression data were analyzed using Seurat for quality control, count normalization, UMAP clustering, and cell cycle analyses (*64*). Specifically, barcodes with exceptionally low or high total read counts were filtered with the following parameters for each replicate: for replicate 1, nCount_RNA between 2×10^4 and 6×10^5; replicate 2, nCount_RNA between 2×10^4 and 3×10^5; replicate 3, nCount_RNA between 2×10^4 and 5×10^5. Mitochondrial reads were removed prior to downstream analyses. Genes expressed in fewer than 10 RKIP cells or VC cells were filtered. Gene expression data were normalized using the NormalizeData function with the LogNormalize method and a

scale factor of 10000. Differential expression analysis was performed using the SCDE package to counter dropout and amplification biases (*65*). The gene expression $\log_2$(fold-change) of RKIP tumor cells over VC tumor cells was determined from the "scde.expression.difference" function using the "maximum likelihood estimate" result. Genes with significant differential expression were determined with a cutoff of +/- 0.5 on false discovery rate (FDR)-corrected Z-scores.

The RNA sequencing data from the two BACH1 studies were analyzed using similar methods with some adaptations for the ICELL8 cx single-cell platform. Specifically, the raw reads were processed by the Cogent NGS Analysis Pipeline (Takara Bio) for quality check, mapping to the human genome (UCSC hg38 with GENCODE annotations), and counting reads per gene. For each replicate, the samples were processed for quality check using the following manufacturer's recommended parameters: cells with fewer than 10000 total reads or fewer than 300 genes detected were removed; genes with fewer than 100 total reads across all samples or expressed in fewer than three cells were removed; outlier cells with greater or less than 3 mean absolute deviation from the median were removed; cells with more than 20% mitochondrial reads or 10% intergenic reads were removed. In lieu of depleting mouse cells during tumor cell dissociation, mouse cells from the shBACH1 and shCTRL tumors were removed via a clustering-based approach via Seurat (*64*). Mouse cells were identified using the "FindNeighbors" function with 50 dimensions of reduction and the "FindClusters" function with a resolution of 0.2. Cells belonging to Cluster 0 were kept for downstream analyses. Cluster 1 cells were treated as mouse cells and removed due to an overall lower percentage of uniquely aligned reads to the human genome. Batch effects from three independent replicates of BM1-CTRL and

BM1-mmBACH1 tumors were corrected using the method described above. Count normalization, cell cycle analysis, and differential expression analysis were performed as described above.

6. Coefficient of variation (CV) analyses

The formula for calculating gene expression CV:

$$CV = \frac{\sigma}{\mu}$$

Where σ is the standard deviation of normalized gene expression of a gene across a set of single cells, and μ is the mean of normalized gene expression of the same gene across the same set of single cells. For every gene in the normalized expression dataset, a CV was calculated for each cell type (control and RKIP). The difference in the CV of each gene between control and RKIP tumor cells is determined by calculating the fold-change:

$$CV_{fold-change} = \frac{CV_{RKIP}}{CV_{control}}$$

Genes with significant differences in CV were determined through permutations testing, during which 1000 rounds of permutations were performed on the gene expression data. Then, the experimentally observed CV fold-change for each gene was compared against the 1000 permutated CV results to determine the likelihood of obtaining the experimental result by chance, defined by the number of times the permutated result met or exceeded the experimental CV result over 1000 permutations. Genes significantly more or less variable in RKIP cells over control cells were determined with an FDR-adjusted P-value cutoff of 0.05.

7. Fano factor of genes in control versus RKIP cells

The formula for calculating gene expression Fano factor (FF):

$$FF = \frac{\sigma^2}{\mu}$$

Where σ is the standard deviation of normalized gene expression of a gene across a set of single cells, and μ is the mean of normalized gene expression of the same gene across the same set of single cells. For every gene in the normalized expression dataset, an FF was calculated for each cell type (control and RKIP). However, due to the scaling introduced during count normalization by Seurat, FF will scale accordingly when the standard formula is used to calculate FF (*68*). Therefore, FF is then re-scaled to neutralize the effect of such scaling prior to downstream analyses:

$$FF' = FF \times \frac{Average\ number\ of\ reads\ per\ cell}{Seurat\ LogNormalize\ scale\ factor}$$

Where the average number of reads per cell is 103543 for RKIP cells and 97545 for control cells, and 10000 was the scale factor applied for count normalization by Seurat.

8. Network analysis

The gene co-expression network of BM1-RKIP or BACH1-overexpressing cells was generated from expression correlations between significantly variable genes. Specifically, a Spearman correlations matrix was calculated to determine expression correlations among the genes with significant differences in CV. Normalized expression levels from all BM1-RKIP cells or BACH1 cells were used to calculate the correlations matrix. The R package “igraph” was used to generate a gene co-expression network from

the correlations matrix using the "graph_from_adjacency_matrix" function (weighted and undirected) (*74*). Genes with significantly higher or lower CV were designated as the network nodes, and the edges represented positively correlated gene pairs with correlation coefficients of at least 0.3. Nodes without peripheral connections were excluded from the network. The largest interconnected component of the whole network was extracted using the "decompose.graph" function and plotted again.

9. Gene functional enrichment analyses

Metascape was used to determine functions and pathways from lists of genes (*75*). Methods for Metascape analyses were previously described (*76, 77*). Metascape analyses in the paper were based on the following lists of genes as inputs: genes with or without significant differences in expression CV and/or mean between BM1-RKIP and BM1-VC cells, according to the mouse tumor single-cell RNA-seq studies; genes that correlated with expression levels of KMT5C, NME1, PEBP1, BRMS1, ARHGDIA, or DRG1 in breast cancer patients, accessed through cBioPortal (TCGA, Firehose Legacy, Breast Invasive Carcinoma, mRNA expression RNA Seq V2 RSEM) (*78, 79*); genes within the largest interconnected component of the whole network of variable genes among the BM1-mmBACH1 cells. Canonical pathway analyses were performed for genes with or without significant differences in expression CV and/or mean between BM1-CTRL and BM1-mmBACH1 cells using Ingenuity Pathway Analysis (*80*). The following cutoffs were used to determine significance: genes with significant CV differences were determined with an FDR-adjusted P-value cutoff of 0.05 from the permutations analyses; genes with significant differential expression were determined with an FDR-adjusted Z-

score cutoff of +/- 0.5 from the single-cell differential expression analyses; genes that positively or negatively correlate with the genes of interest in breast cancer patients were selected with a Spearman's correlation coefficient cutoff of +/- 0.3. The significance of Metascape analyses was determined by P-value < 0.01. The significance of canonical pathway analyses was determined by P-value < 0.05. Metascape queried the following databases for analyses: GO Molecular Functions, GO Biological Processes, Reactome Gene Sets, and KEGG Pathway (*81-83*).

10. Cell proliferation assays

Two thousand tumor cells were seeded per well in a 96-well tissue culture plate. For each cell type, ten wells were seeded. The tissue culture plate was incubated for 24 hours before transferring into an IncuCyte S3 (Sartorius) with the same incubation conditions. Cells were monitored over time by scanning the plate with a 10X phase-contrast objective for up to 7 days. Four images were acquired per well every 4 hours and analyzed by the IncuCyte 2020B software (Sartorius) to determine cell confluence. The mean and standard deviation of percent cell confluence across ten wells were calculated for each cell type at each time point. Proliferation curves were plotted by the IncuCyte software, and error bars represent standard deviations.

11. Protein isolation and western blots

Cultured cells were washed with cold PBS and lysed on ice in RIPA buffer with protease inhibitors (Santa Cruz, sc-24948; EMD Millipore, 539134). Samples were then

sonicated three times for 10 seconds each at 25-30% power and centrifuged at max speed for 15 min at 4°C. The supernatant was collected, and the protein concentration was quantified using the Bradford assay (Bio-rad, 5000202). Samples were then boiled in 4X Laemmli sample buffer (Bio-rad, 1610747) at 98°C for 5 mins before storage. SDS-PAGE was performed using 20 – 25 µg protein from each sample with a prestained protein ladder (ThermoFisher, 26616). Western blotting was performed using antibodies against KMT5C (AbClonal, A16235, ~52 kDa) or H4K20Me3 (Cell Signaling, 5737S, ~11 kDa). Blots were stained for total protein using the REVERT 700 Total Protein Stain Kit (LI-COR, 926-11016) for control. Following primary antibody incubation, blots were then incubated with fluorescent or HRP-conjugated secondary antibodies depending on the target (LI-COR, 926-3221 for KMT5C blots; MilliporeSigma, AP187P for H4K20Me3 blots). Before imaging, blots with HRP secondary antibodies were treated with chemiluminescent HRP substrate (MilliporeSigma, WBKLS0500). Images were acquired on the LI-COR Fc Imaging System and quantified using Empiria Studio Software (LI-COR) with total protein normalization.

12. Immunohistochemistry

Mouse tumors were fixed in 10% formalin solution (Azer Scientific, NBF-4-G), then stored in 70% ethanol. Paraffin embedding, cross-sectioning, and IHC staining were performed by the University of Chicago Human Tissue Resource Center following standard protocols. One slide of each tumor sample was stained by the hematoxylin and eosin stain. For H4K20Me3, slides were stained using the Recombinant Anti-Histone H4 (tri methyl K20) antibody (Abcam, ab177190) at 1:4000 dilution. Stained slides were

scanned by a CRi Pannoramic SCAN 40x Whole Slide Scanner at the Integrated Light Microscopy Core, then analyzed with CaseViewer software (3DHISTECH).

13. Bulk tumor RNA-seq

Five BM1-Ctrl and five BM1-KMT5C tumors were collected from ten athymic nude mice. Fresh tumors were placed inside M Tubes (Miltenyi Biotec, 130-093-236), then homogenized with TRI Reagent (Zymo Research, R2050-1) by the gentleMACS Dissociator (Miltenyi Biotec 130-093-235). Total RNA extraction with genomic DNA removal was performed using the Direct-zol RNA MiniPrep Plus kit (Zymo Research, R2072). The University of Chicago Genomics Facility performed RNA sample quality checks, library construction, and sequencing following standard protocols. All ten samples were sequenced in four runs by a NovaSeq 6000 sequencer (Illumina) to generate paired-end 100bp reads. The raw FASTQ files from four flow cells were combined for each sample before downstream processing. RNA-seq data were analyzed using a local Galaxy instance (*84*). Quality and adapter trimming were applied to the raw sequencing reads using the Trim Galore! Package (Felix Krueger, github.com/FelixKrueger/TrimGalore). The reads were mapped to the human genome (UCSC hg19 with GENCODE annotation) using RNA STAR (*71, 72*). Per gene read counts from each sample were generated by featureCounts (*73*). The raw read counts were normalized and analyzed for differential expression between control and KMT5C overexpression tumors using DESeq2 (*85*). Genes with significant differential expressions were determined using an FDR-corrected P-value cutoff of 0.1. Significant genes and their expression fold change results were analyzed by Ingenuity Pathway

Analysis (IPA) software (Qiagen) to predict changes in Diseases and Biological Functions (*80*).

14. KMT5C gene expression correlations with motility/invasion and OXPHOS genes across multiple tumor types

Co-expression data for KMT5C from the following tumor types were downloaded from cBioPortal: kidney renal clear cell carcinoma, thyroid carcinoma, uterine corpus endometrial carcinoma, bladder urothelial carcinoma, breast invasive carcinoma, head and neck squamous cell carcinoma, prostate adenocarcinoma, skin cutaneous melanoma, pancreatic adenocarcinoma, thymoma, and testicular germ cell cancer (*78*). All co-expression studies were based on TCGA's patient RNA-seq data (Firehose Legacy, mRNA expression RNA Seq V2 RSEM) (*79*). Motility/invasion and OXPHOS gene signatures related to metastases were obtained from previous studies, and expression correlations to KMT5C were analyzed (*23, 37*). Positive or negative correlations between KMT5C and metastasis-related genes were determined using Spearman's correlation coefficient cutoff of +/- 0.2. Correlation coefficients between -0.2 and 0.2 were not shown (marked gray in the heatmap).

15. Patient gene co-expression analyses of KMT5C, motility, and OXPHOS genes in breast cancer and pancreatic cancer

Breast cancer or pancreatic cancer mRNA expression data on TCGA were downloaded from cBioPortal (TCGA, Firehose Legacy, RNA Seq V2 RSEM) (*78, 79*). For

each tumor type, mRNA expression Z-scores for KMT5C and other genes of interest were plotted for all patients as a heatmap using cBioPortal. Rows represent expression levels of KMT5C, motility/invasion genes, or OXPHOS genes from patient samples (*23, 37*). Columns represent individual patients grouped based on the relative expression level of KMT5C. KMT5C high or low expression was determined with a Z-score threshold of +/- 0.5.

16. Pathway activities in TNBC patients with increased or decreased levels of BACH1, RKIP, or KMT5C

Breast cancer patients with high or low expression levels of KMT5C, RKIP, or BACH1 were stratified with a Z-score threshold of +/- 1, using TCGA patient data accessed through cBioPortal (TCGA, Firehose Legacy, RNA Seq V2 RSEM) (*78, 79*). TNBC patient sample IDs were identified using the patient clinical data downloaded from Genomic Data Commons Data Portal (*86*). Samples indicated the TNBC status with negative “ER status by IHC”, negative “PR status by IHC”, and at least one negative result from “HER2 status by IHC” or “HER2 FISH status”. RNA-seq RSEM normalized count data for TNBC samples were downloaded from UCSC Xena and used for differential expression analyses (*87, 88*). Differentially expressed genes in KMT5C-high over KMT5C-low patients, RKIP-high over RKIP-low patients, and BACH1-high over BACH1-low patients were determined using the Limma-Voom package via the Galaxy server (*84, 89, 90*). For Limma-Voom, patient samples' RNA-seq data were supplied as a single count matrix, and high or low indicators for all samples were designated in a factor file. Differentially expressed genes were plotted as volcano plots for each comparison (*91*).

IPA was used to predict activities of related canonical pathways based on relative expression levels of KMT5C, RKIP, or BACH1 (*80*). IPA-predicted canonical pathways activation Z-scores from each comparison (KMT5C-high over KMT5C-low, RKIP-high over RKIP-low, and BACH1-high over BACH1-low) were compared to each other to identify pathways with the biggest activity changes across every comparison. IPA used the following FDR-adjusted P-value (q-value) cutoffs for each dataset to perform the canonical pathways cross comparison: KMT5C high vs. low $q < 0.01$, RKIP high vs. low $q < 0.05$, BACH1 high vs. low $q < 0.01$. These cutoffs were implemented due to limitations set by the IPA software.

17. KMT5C expression and patient survival

Breast cancer or pancreatic cancer patients were stratified based on relative expressions of KMT5C. For each tumor type, patients with high or low expression of KMT5C were selected using cBioPortal with a z-score threshold of +/- 1 from the TCGA Firehose Legacy RNA Seq V2 RSEM datasets (*78, 79*). Survival analyses were performed based on TCGA's patient overall survival data, and Kaplan-Meier plots were generated for each tumor type using cBioPortal (*78, 79*). A statistically significant difference in patient survival was determined by the logrank test on KMT5C-high and KMT5C-low patients with a P-value below 0.05.

18. Additional statistical analyses and software

GraphPad Prism version 9 (GraphPad Software) was used for unpaired t-tests as indicated in the figure legend, with significant P-values shown (significance is determined by P-value < 0.05). Error bars represent standard deviations. All FDR-corrected P-values and Z-scores utilized the Benjamini-Hochberg (BH) correction method (*92*). Figures were created with Prism, BioRender.com, or Morpheus (Broad Institute, software.broadinstitute.org/morpheus).

CHAPTER 3. THE METASTASIS SUPPRESSOR RKIP REDUCES OVERALL TRANSCRIPTIONAL VARIABILITY OF TUMOR CELLS

Abstract

Tumor cell heterogeneity has been implicated in the metastatic progression of solid tumors such as triple-negative breast cancer (TNBC), leading to resistance and recurrence. We hypothesize that the metastasis suppressor Raf Kinase Inhibitory Protein (RKIP or PEBP1) works in part by reducing transcriptional heterogeneity. We performed single-cell RNA sequencing to compare the variability in gene expression between individual cells in metastatic (low RKIP) versus non-metastatic (high RKIP) TNBC tumors. Overall transcriptional variability was decreased with RKIP overexpression. While the vast majority of genes exhibited no change, known genes undergoing changes in transcriptional variability and/or mean in response to metastatic suppression were largely regulators of cancer progression and metastasis. Known genes with decreased variability were associated with unchanged or increased means, making them more effective therapeutic targets than genes with unchanged variability. We showed that KMT5C, a gene with both increased mean and decreased gene expression variability in response to RKIP, acts as a novel epigenetic metastasis suppressor in TNBC. Taken together, these results indicate that analysis of gene expression variability along with mean can lead to more effective target selection and provide an alternative strategy for identifying novel regulators of metastasis.

Introduction

Metastasis, the dissemination of tumor cells throughout the body, is the primary cause of death from solid tumors such as breast cancer. Characterized by distinct biological states driven by stresses such as oxygen and nutrient loss in the microenvironment, metastatic tumors are largely drug-resistant, but the underlying mechanisms are poorly understood. Identifying and characterizing regulators of metastasis and utilizing them to confer sensitivity to treatment is essential for developing successful therapeutic intervention strategies.

Multiple lines of evidence suggest that heterogeneity arising from genetic, epigenetic and/or phenotypic sources can be a driver of metastatic progression and drug resistance. For example, heterogeneity in the expression of proteins involved in apoptotic signaling altered the apoptotic response of tumor cells to TRAIL (*6*). Similarly, morphological heterogeneity of tumor cells was shown to correlate with increased metastatic potential (*8*). Finally, high intratumoral heterogeneity is strongly correlated with aggressiveness and predicts poor patient outcomes (*7*). While heterogeneity is commonly studied at the genetic level by looking at mutations in single cells, emerging evidence suggests that transcriptional heterogeneity is also related to the metastatic outcome (*8, 93*).

Change in the mean expression alone does not exclude over/underexpression restricted to a subset of cells in the population. Thus, to identify genes with increased and homogenous expression across the population, it is also important to analyze single cells. Previous studies have revealed an association between increased mean and reduced gene expression variability or noise measured by the coefficient of variation (CV);

conversely, decreased mean is frequently associated with increased variability (*67, 68, 94*). However, molecular perturbations, epigenetic effects, and feedback or general regulatory pathways can alter the noise-mean interdependence and impact cell function or survival (*67-69, 95*). For example, perturbation of genes involved in repair machinery to amplify noise without significantly changing the mean can potentiate transitions in stem cell fate (*68*). Thus, a reduction in the expression variability of specific genes within tumor cell populations could represent an alternative strategy to limit cell-to-cell heterogeneity, increase treatment efficacy and identify novel metastatic regulators.

To test whether regulators of metastasis modulate gene expression variability and thereby contribute to the metastatic outcome, we expressed ectopically a metastasis suppressor, Raf Kinase Inhibitory Protein (RKIP; encoded by *PEBP1*) in metastatic tumor cells, as a possible means of reducing gene expression variability. Previous studies in our lab and others have shown that RKIP does not significantly alter primary tumor growth but is an effective metastasis suppressor in many solid cancers, such as triple-negative breast cancer (TNBC) (*17*). Not surprisingly, RKIP expression is low in many TNBC patients, and low RKIP is associated with poor relapse-free survival (*16*). As metastasis has been linked to increased transcriptional fluctuation, we determined whether the metastasis suppressor RKIP might also shift variability in gene expression within tumors (*8*).

To measure transcriptional variability, we conducted deep single-cell RNA sequencing (RNA-seq) of metastatic (low RKIP) and non-metastatic (high RKIP) tumor cells and measured both the change in mean and CV of gene expression in cells across the tumor population. Using an approach based on permutation analysis to identify

significant genes, we showed that RKIP reduces overall transcriptional variability in primary TNBC tumors. Many cancer regulatory genes were more homogeneous and up-regulated in the RKIP-expressing cells. Among these genes is KMT5C, a histone methyltransferase that facilitates epigenetic repression. Cell, mouse, and clinical studies suggest KMT5C expression blocks invasion, extravasation, and colonization of tumor cells, consistent with a role as a novel suppressor of metastasis. These studies reveal the importance of genes involved in heterogeneity as potential regulators of metastatic progression in cancer.

As the functions of many genes with increased mean and decreased noise are not known, we characterized KMT5C and showed that it acts as a novel epigenetic metastasis suppressor in TNBC. Taken together, these results indicate that analysis of gene expression variability along with mean can lead to more effective target selection and provides an alternative strategy for identifying novel regulators of metastasis.

Results

RKIP overexpression reduces transcriptional variability in TNBC

To identify transcriptional differences between metastatic and non-metastatic tumors, we performed single-cell RNA sequencing of human TNBC tumors propagated in a xenograft mouse model. Previously, Rosner and colleagues demonstrated that the TNBC cell line, BM1, lost its metastatic potential upon expression of a metastasis suppressor, RKIP-S153E (*15*). In the present study, following mammary fat pad injection of RKIP-S153E or vector control BM1 cells into athymic nude mice, we dissociated the resulting tumors into single cells and processed the viable human tumor cells for single-cell RNA sequencing (Fig. 1A). To compare transcriptional variability across single cells, it is beneficial to obtain as many sequencing reads as possible from each cell. Therefore, we used a microfluidics-based method that offers more sequencing depth per cell compared to commonly used droplet-based methods (*54*).

Before analyzing transcriptional variability, we evaluated three independent replicates from each tumor type to control for sequencing quality and depth, reproducibility, cell cycle, and batch effects. Distributions of the number of genes detected per cell were comparable between the metastatic and non-metastatic tumor samples for each set of replicates (Fig. S1A), and the single-cell gene expression profiles were reproducible across the three replicates (Fig. S1B). The distribution of cells within the cell cycle for the two tumor types was also comparable (Fig. S1C). Notably, UMAP clustering could not effectively separate control and RKIP-expressing cells despite their known differences in metastatic potential (Fig. 1B), suggesting that this dimensionality reduction technique is insufficient to distinguish between metastatic and non-metastatic phenotypes.

We first identified genes with differential expression (change in the mean, μ) between control and RKIP-expressing cells in tumors. To account for sequencing noise and gene dropouts due to differences in read counts in single-cell sequencing, we analyzed statistically significant differences in mean gene expression between control and RKIP-expressing tumor cells as previously described (*96*). Of the 15,006 genes characterized, 596 increased in mean and 930 decreased in mean with RKIP overexpression, indicating that transcriptional expression of the majority of genes was unchanged (Table 1). Although the statistical metric based upon the mean is generally used to subset genes and characterize differences between tumor types, we reasoned that a different statistical metric based upon transcriptional variability would generate an alternative gene subset that might identify other potential targets of interest.

To elucidate transcriptional variability among single tumor cells, we calculated the CV (σ/μ where σ refers to the standard deviation; Fig. 1C)) for each gene from the RKIP-expressing tumor cells versus control by combining single-cell gene expression results from all three replicates and removing batch effects. Experimental observation showed that RKIP overexpression caused an overall decrease in heterogeneity. To determine whether these observed differences represent true biological differences as opposed to technical artifacts due to the noisiness of single-cell sequencing, we performed 1000 rounds of permutations involving random reassignment of single-cell expression levels to RKIP or control cell types (see Methods) and compared the results of all the permutations to the experimentally observed difference. Because the observed experimental difference in gene expression variability does not overlap with the distribution of results from 1000 permutations, there are indeed significant differences in gene expression variability

between RKIP and control tumor cells (Fig. 1D). Using false discovery rate (FDR)-adjusted P-values derived from the permutation analysis, we found that 485 genes were significantly less variable (more homogeneous) in RKIP cells (including the overexpressed RKIP itself) and 152 genes were more variable (Figs. 1E, S1D).

RKIP causes changes in gene expression mean and/or variability

Upon analyzing the genes with changes in mean or CV, we identified 20% with CV changes only, 66% with mean changes only, and 14% with both mean and CV changes when comparing control to RKIP-expressing tumor cells (Table 1). Notably, there were no genes up-regulated and more variable in RKIP cells, while only one gene was down-regulated and more homogeneous, suggesting that CV and mean rarely increase or decrease together upon RKIP overexpression. Among the genes with decreased CV, the majority (65%) had no significant alteration in mean, whereas the majority of the genes with increased CV were associated with a decreased mean (Table 1).

Gene transcription is generally considered a Poisson process comprised of random, uncorrelated events that occur at a certain mean rate (*97*). However, if these transcriptional events are modulated, there will be a departure from Poisson. This could occur as intermittent bursts, feedback regulation, upstream noise propagation, or other influences (*98*). To study the deviation of gene expression variability from Poisson scaling, we plotted $\log_{10}(CV^2)$ versus $\log_{10}(\mu)$ for both control and RKIP-expressing cells (Fig. 2). If the slope is -1, then this is consistent with a classic Poisson process or intermittent bursts (defined as Fano Factor FF≥1 where FF is σ^2/μ and $\sigma^2 \geq \mu$) (*69*). For genes from both control and RKIP-expressing cells, the slopes were relatively constant and similar,

approaching -1. This result is not surprising as the majority of gene expression variability (~96%) was not significantly changed by RKIP overexpression. However, when we plotted only the genes whose mean was unchanged but whose expression was significantly less variable in RKIP-expressing cells, we observed a downward shift for RKIP-expressing cells relative to control cells (Fig. 2A) and a corresponding decrease in FF, as expected for genes without a significant change in the mean (Fig. 2B). These results are consistent with the decrease in variability in RKIP cells we observed using the CV-based permutation analysis (Fig. 2B; also see Fig. 1D, E). By contrast, for genes where the mean increases more than the CV drops, the FF can increase, as noted for overexpressed RKIP (Fig. 2C). These results indicate that, for most genes with decreased CV, RKIP expression causes FF reduction for both Poisson and intermittent bursting processes.

We have previously proposed that both gene expression mean and variability changes regulate the transition between metastatic and non-metastatic tumor cells (*99*). Given that CV and/or mean of gene expression can either increase, decrease, or stay the same in cells transitioning to a non-metastatic phenotype, there are nine possible scenarios for the differences ΔCV^2 and Δmean for each gene (Fig. 2D). Interestingly, $\Delta\log_{10}(CV^2)$ versus $\Delta\log_{10}$(mean) followed Poisson scaling similar to the $\log_{10}(CV^2)$ versus the $\log_{10}$(mean) for control and RKIP cells alone. However, in this case, we are analyzing differences ($\Delta\log_{10}(CV^2)$ and $\Delta\log_{10}$(mean)), suggesting that these changes are constrained in the same way as the original quantities for genes in control and RKIP-expressing cells. This can be explained if the rate of transcription or burst size is increased by a constant factor due to RKIP overexpression (see Supplementary Note).

Functions of genes with CV changes

Literature searches of genes with CV and mean changes revealed both pro-cancer and anti-cancer genes as well as genes that exhibit context-dependent regulation of cancer (Figs. 3A, B, S3A, Tables S1, S2). Of particular interest are the sets of genes with decreased CV and either increased or unchanged mean that comprise about 25% of all genes with significant changes (Table 1, Fig. S3B). For example, FZD5, the receptor of Wnt5A, contributes to tumor cell proliferation and is suggested to be a drug target in colorectal and pancreatic cancers (Fig. S3B) (*100*). CCND2 is involved in cell cycle regulation and is frequently lost in breast cancers (*101*). FGF18 is a fibroblast growth factor that promotes tumor invasion in ovarian cancer (*102*). Genes known to suppress breast cancer are also among the genes that are more homogeneous in RKIP cells, such as FOXP2, a transcription factor that promotes stemness and metastasis when knocked down in breast cancer cells (*103*). These represent potentially superior therapeutic targets as they are more uniformly expressed across the cell population (Fig. S3B).

We next determined whether the genes with discrete CV and mean changes can be grouped into known cellular functions and processes by performing Metascape analyses for pathway enrichment (*75*). Strikingly, of the 6 configurations of mean/CV changes observed in response to RKIP expression, each corresponded to a discrete set of biological functions (Figs. 3C, S4). Genes more homogeneous in RKIP cells had marked enrichment in oxidative phosphorylation, DNA damage response, and cytoskeleton pathways, while genes more heterogeneous had less defined functions (Figs. 3C, S4). As predicted from this CV analysis, the RKIP-expressing cells are likely to

be more metabolically homogeneous with respect to oxidative phosphorylation and, therefore, more susceptible to respiratory inhibitors like metformin (Fig. 3D). *In vivo* results showed that RKIP overexpression in BM1 cells sensitized mouse tumors to metformin treatment while the control tumors did not respond to the treatment. In contrast, analyzing expression mean changes or variability changes alone could not effectively surface these clinically relevant pathways (Fig. S5A, B). Taken together, these results reveal distinct cancer-related genes and functions that become less transcriptionally variable upon losing their cellular metastatic state. Notably, potential druggable targets can be identified by filtering genes with both mean and variability changes.

CV analyses enable the identification of novel genes that regulate cancer

Most of the genes with altered CV are novel, and their relationships to cancer cannot be ascertained based on current literature. One approach to identifying important regulators of cancer is the analysis of co-expression networks for genes with altered variability. To test this possibility, we performed gene co-expression network analysis for genes with significant CV changes and identified a network of genes with positive correlations in gene expression. The largest network centered around RPL15, one of the least variable genes without mean changes, shows co-expression of several genes, including proliferation markers (MKI67 aka Ki67), metabolic enzymes (LDHA, LDHB), and metastasis suppressors (ARHGDIB) (Figs. 4A, S6). The importance of RPL15 in the context of heterogeneity is supported by a recent study suggesting that more variable RPL15 protein expression is associated with metastasis in breast cancer (*93*).

An alternative approach for identifying novel regulators of metastasis is to characterize genes with decreased CV and increased mean. KMT5C is the sixth most less variable gene and was significantly up-regulated in RKIP-expressing tumor cells (Figs. S3A, 4B). KMT5C tri-methylates H4K20Me3 and acts as an epigenetic repressor (*104*). However, its role in breast cancer progression has not been addressed, and there are conflicting reports for other cancer types (*105-107*). To test whether KMT5C regulates metastasis, we generated TNBC cells with stable over-expression of KMT5C and confirmed the increased tri-methylation at histone H4K20 (Figs. 4C, S7A). Cells with or without KMT5C overexpression showed comparable proliferation *in vitro* (Fig. S7B). Similarly, xenograft tumors in athymic nude mice with or without KMT5C overexpression did not exhibit differences in primary tumor growth (Fig. 4D) but did show a marked increase in H4K20Me3 staining (Fig. S8A). By contrast, intracardiac injections of KMT5C-expressing BM1 cells resulted in significantly fewer metastatic colonies compared to the control cells, consistent with a role for KMT5C as a metastasis suppressor (Fig. 4E). RNA sequencing of the primary xenograft tumors identified 814 genes down-regulated and 769 genes up-regulated in KMT5C-expressing tumors (Fig. S8B). Disease and biological functions analyses of the differentially expressed genes identified changes in metastasis-related functions in KMT5C-overexpressing tumors (Fig. 4F). These results were validated using a syngeneic mouse model showing that KMT5C suppressed primary tumor growth, possibly due to the intact immune system, as well as spontaneous metastases to the lung (Fig. 4G-H). Together, these results suggest that KMT5C functions as a metastasis and context-dependent tumor suppressor for TNBC.

KMT5C is a clinically relevant regulator of metastasis in breast cancers

We next analyzed expression levels of KMT5C in cancer patients. We previously identified a group of invasion genes (INV) driven by the transcription factor BACH1 that promotes metastasis in breast cancer (*23*). We also identified genes in the electron transport chain (OXPHOS) that are up-regulated by RKIP expression (see Fig. 3C) or BACH1 depletion in TNBC cells (*37*). Analysis of the TCGA database across multiple solid tumor types for expression of these genes shows a significant negative correlation between KMT5C and INV gene expression and a positive correlation of KMT5C to OXPHOS gene expression (Figs. 5A, S9). Similarly, within individual patient samples, both breast and pancreatic tumors with higher KMT5C expression generally express lower levels of INV genes and higher levels of OXPHOS genes (Figs. 5B, S10A, B). We then compared KMT5C to BACH1 expression in patients and observed a negative correlation. We stratified breast cancer patients by expression levels of either KMT5C, RKIP, or BACH1 and performed differential expression analyses within each patient group (Fig. S11 A-C). Both RKIP and KMT5C are associated with an overlapping set of canonical pathways in a manner that suppresses cancer progression (Fig. 5C). By contrast, the same pathways are associated in the opposite manner with BACH1.

We also compared KMT5C to other previously identified metastasis suppressors using the TCGA database (*13, 79*). In breast cancer patients, KMT5C and four known suppressors of metastases (NME1, BRMS1, ARHGDIA, DRG1) are positively or negatively correlated to a similar set of gene functions (Figs. 5D, S12), suggesting that they act in a similar manner. When analyzed alone, KMT5C expression is positively correlated to processes such as translation and aerobic respiration and negatively

correlated to several cancer-promoting signaling pathways as well as invasion and motility (Fig. S13). Finally, in breast and pancreatic cancer patients, higher KMT5C expression is correlated with significantly better overall survival (Fig. 5E). Together, these patient tumor analyses further support the experimental data suggesting that KMT5C, like its upstream regulator RKIP, also functions as a metastasis suppressor and is associated with improved survival in breast and possibly other cancers.

Taken together, our results showed that the metastasis suppressor RKIP can reduce the transcriptional variability of TNBC tumors. Analysis of functional pathways with decreased CV and unchanged or increased mean reveal potential drug targets that enhance tumor sensitivity to treatment. We also report that KMT5C, one of the novel genes up-regulated and more homogeneous in RKIP-overexpressing cells, is a histone modifier with metastatic suppression function in TNBC (Fig. 6).

Discussion

In this study, we use the coefficient of variation as a novel method for comparing differences in transcriptional variability between metastatic and RKIP-induced non-metastatic tumor states and to identify more effective therapeutic targets as well as novel regulators of cancer progression and metastasis. Taken together, our results showed that the metastasis suppressor RKIP can reduce the transcriptional variability of TNBC tumors. Analysis of functional pathways with decreased CV and unchanged or increased mean following loss of the metastatic phenotype reveals more effective drug targets such as oxidative phosphorylation, enabling enhanced tumor sensitivity to treatment. On the basis of both experimental and clinical dataset analyses, we also identify KMT5C, one of the novel genes with decreased mean and increased CV in RKIP-expressing cells, as a histone modifier with metastatic suppression function in TNBC (Fig. 6).

Previous studies on tumor heterogeneity focused primarily on genetic heterogeneity extrapolated from single-cell mutations and copy number variations or subclonal populations inferred from bulk sequencing data (*108*). By contrast, the coefficient of variation is only beginning to gain traction as an approach to studying tumor cell heterogeneity at the transcriptional level. For example, CV was used to quantify cancer cells' variable apoptotic response to killing (*6*). More recently, the CV method was employed to study transcriptional noise during DNA damage repair, demonstrating the usefulness of this method outside of cancer research (*68*). By applying permutations analysis to CV, our method allows us to not only find genes with absolute differences in CV, but also determine the statistical significance of the changes. CV can be used on many types of datasets other than single-cell RNA sequencing or cell-to-cell phenotypic

variability. For example, CVs can also be determined from flow cytometry and spatial datasets. With the rise of spatial techniques, we predict this approach can also benefit emerging techniques, including cyclic immunofluorescence and spatial transcriptomics.

As CV can reflect many sources of noise, we demonstrated that RKIP causes a reduction in the Fano factor for most genes with decreased CV. However, the mechanism by which this occurs is not yet known and could result from a variety of causes, including altered bursting, feedback regulation, or noise modulators (*67-69, 95*). Previous studies have shown that changes in repair genes can lead to increased noise without changing mean gene expression (*68*). Thus one mechanism by which RKIP might decrease variability is via DNA repair. Similarly, changes in epigenetic processes could alter transcription. One gene by which this might occur is the RKIP-regulated KMT5C. Future studies will be required to elucidate the precise mechanisms by which RKIP regulates transcriptional variability.

Changes in gene expression variability do not necessarily reflect changes in gene expression mean. Given our results demonstrating that many genes with altered mean do not exhibit significant changes in transcriptional variability, it is possible that the expression of these genes was only changed in a subset of tumor cells. Therefore, a potential therapy targeting the genes with only mean changes may not be effective for all or even most of the cancer cells. Of the genes we identified that have decreased CV in non-metastatic tumors and therefore are likely to be more homogeneously expressed, some were up-regulated, while others did not change their mean expression. Among the latter genes is RPL15, a ribosomal protein whose translational heterogeneity is linked to

metastasis (*93*). We found RPL15 to be the center of a co-expression network of genes with altered CV, another potential method for identifying new metastatic regulators.

Studies focusing only on mean changes would miss such genes when identifying potential therapeutic targets. The RKIP-overexpressing tumor cells are predicted to be in a homogeneous metabolic state that participates in OXPHOS, thus, more sensitive to killing by respiratory inhibitors such as metformin (see Fig. 3C). Of note, this prediction is consistent with our previous report that loss of BACH1, a pro-metastatic transcription factor that down-regulates and is down-regulated by RKIP, switches the metabolic state of TNBC tumors from aerobic glycolysis to largely oxidative phosphorylation, resulting in therapeutic sensitivity to metformin treatment (*27, 37*). We showed that mice bearing RKIP-high BM1 tumors were indeed sensitive to metformin treatment, while control tumors did not respond to the treatment (see Fig. 3D). However, OXPHOS pathways were not highly enriched based on mean changes alone and could be easily missed in this analysis (see Fig. S5B).

Unexpectedly, our literature search turned up more pro-cancer genes than anti-cancer genes with decreased CV upon RKIP expression (Fig. 3A, B). Given this observation, one could infer that the RKIP-expressing tumors are more aggressive, an observation that does not correspond to their actual phenotype. Our results suggest that metastasis suppressors do not function in a switch-like manner, and the final tumor phenotype is clearly a summation of the CV and mean of genes expressed across a population of single cells. This observation suggests once again that the expression mean alone, either in a population or even in single cells, is insufficient to explain metastatic regulation.

Using our approach to reveal genes subject to metastasis-associated transcriptional heterogeneity, we identified KMT5C as a novel metastasis suppressor in TNBC. Our results offer important clues regarding the mechanism by which KMT5C suppresses metastasis. KMT5C overexpression reduced tumor cell extravasation and colonization in a xenograft model without affecting primary tumor growth. Interestingly, in the syngeneic mouse model, both tumor growth and metastases were reduced, suggesting that KMT5C may also suppress primary tumor growth through the immune system. In breast cancer patients, KMT5C gene expression is negatively correlated with the PD-L1 pathway (Fig. S13). In TNBC patients, PD-L1 gene expression is significantly less in KMT5C-high patients. This suggests that KMT5C may have a role in sensitizing tumor cells to killing by T cells through the regulation of PD-L1.

Although beyond the scope of the present study, a number of questions regarding KMT5C also remain. KMT5C mRNA is expressed in many normal tissues in a non-specific manner, and additional evidence regarding its protein expression in breast tissues or breast cancers is needed (*109, 110*). It is unclear how KMT5C expression and function is regulated in breast cancer or its detailed mechanism of metastatic suppression. Future studies are needed to determine how RKIP can simultaneously up-regulate KMT5C and reduce its transcriptional heterogeneity and whether RKIP can also suppress metastasis at the epigenetic level through KMT5C. Previous studies report contradictory evidence on KMT5C's regulation of metastasis, so it's possible that KMT5C may not be a metastasis suppressor in all cancer types (*105-107*). While KMT5C tri-methylates H4K20, the effect on epigenetic heterogeneity is not clear. It is also possible that additional mechanisms exist to regulate H4K20Me3 as separate studies report no difference in KMT5C gene

expression between senescent and proliferating human cells, yet H4K20Me3 levels were higher in senescent cells (*111*). However, KMT5C expression and H4K20Me3 levels were lower in breast tumors than in normal tissues (*111, 112*).

We predict that targeting genes with higher mean and lower CV is more effective than genes without changes in mean. However, testing this hypothesis is difficult as, currently, it is technically challenging to reduce CV without also changing mean expression outside of tightly controlled tissue culture conditions (*113*). It should also be noted that a small number of genes became heterogeneous in RKIP tumors. The significance of genes with higher CVs in non-metastatic tumors remains to be determined. Nevertheless, our study reveals the clinical importance of utilizing regulators of metastasis, in combination with single-cell CV and mean, to identify promising candidates as therapeutic targets or novel regulators.

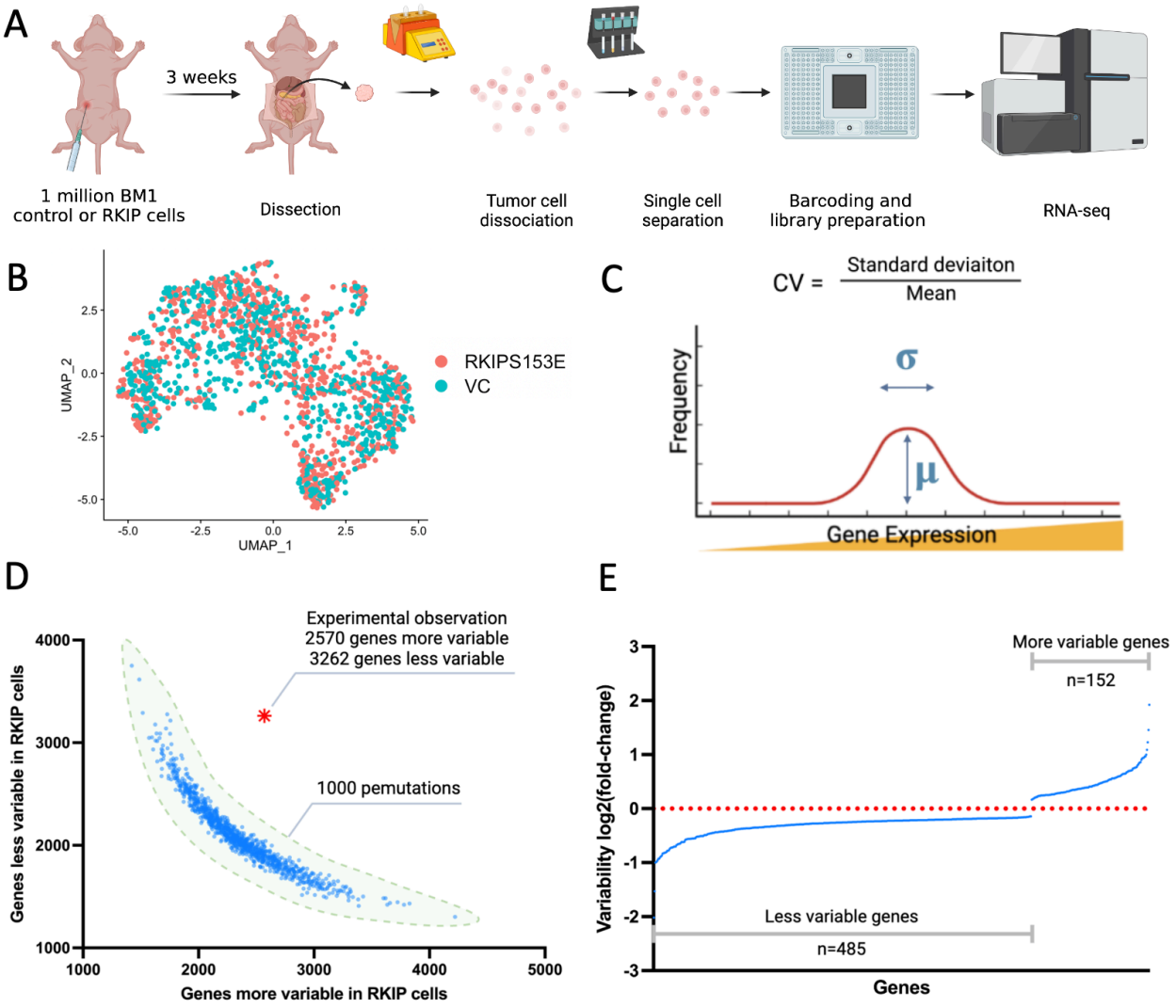


Figure 1. RKIP expression significantly decreases gene expression variability in TNBC xenograft tumors. A) Schematic of scRNA-seq study. B) UMAP clustering of BM1-ctrl (VC) and BM1-RKIP (RKIPS153E) single cells. C) Calculation of gene expression coefficient of variation from single-cell gene expression mean and standard deviation. D) Number of more variable or less variable genes in RKIP cells (10% or more difference from control cells) observed by experimental observation compared to 1000 permutations. E) Changes in gene expression variability of genes that were significantly more or less variable in RKIP cells.

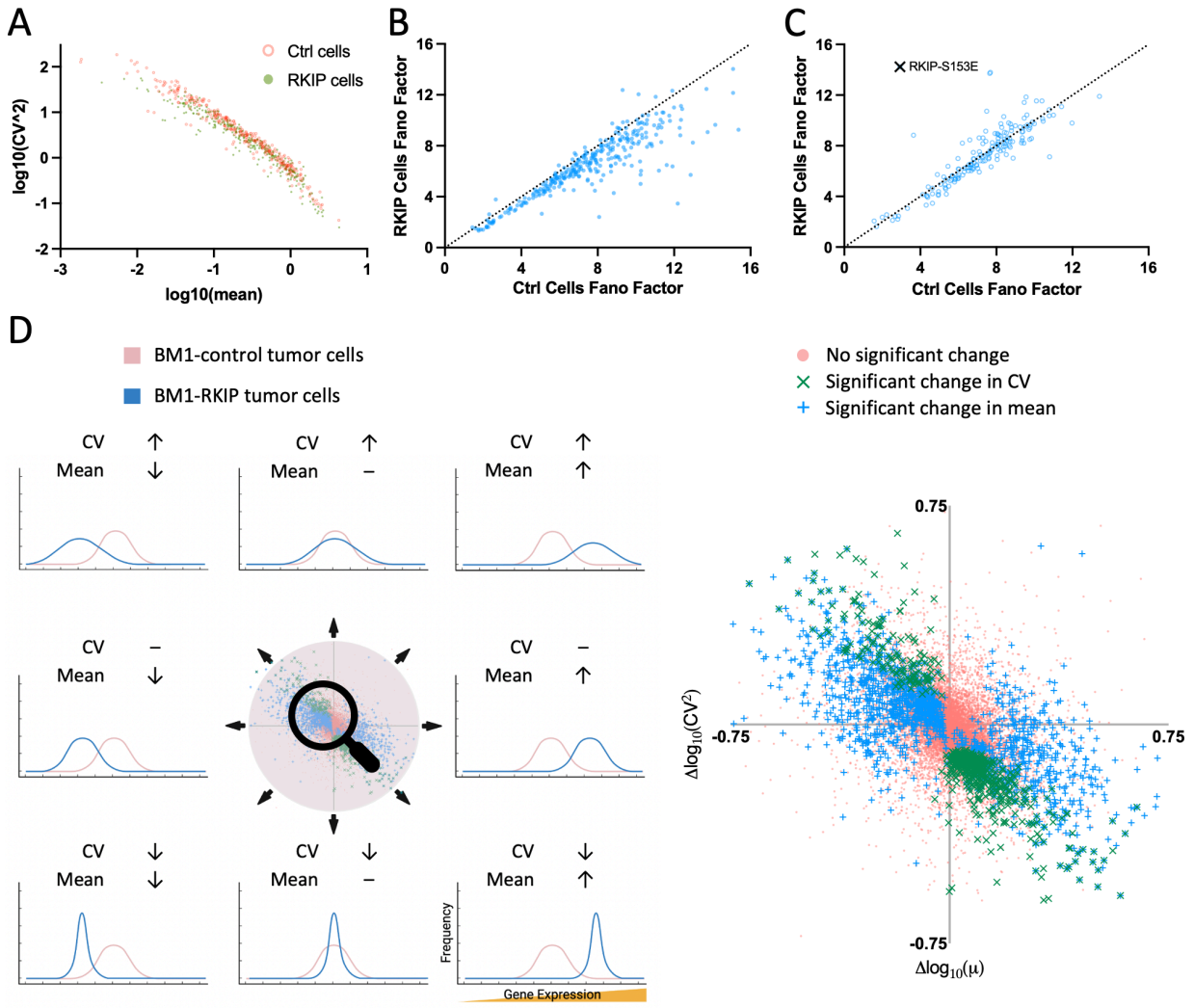


Figure 2. RKIP causes changes in gene expression mean and variability. A) Expression $\log_{10}(CV^2)$ versus $\log_{10}$(mean) of genes in control and RKIP cells. Genes plotted are those with significant down-regulation in CV without changes in mean in RKIP cells. B) Fano factor (FF) of genes in control cells versus RKIP cells. Genes plotted are those with significant down-regulation in CV without changes in mean in RKIP cells. C) The Fano factor (FF) of genes in control cells versus RKIP cells. Genes plotted are those with significant down-regulation in CV and up-regulation in mean in RKIP cells. The Fano factor of overexpressed RKIP-S153E in control versus RKIP cells is indicated. D) Left: genes can change in CV and/or mean from control to RKIP-expressing cells. Right: enlarged scatter plot of changes in $\log_{10}(CV^2)$ and $\log_{10}$(mean) for all genes. Genes with significant changes in CV are indicated with a "cross" symbol, and genes with significant changes in mean are indicated with a "plus" symbol. Genes with significant changes in both CV and mean are indicated with an overlap of "cross" and "plus" symbols.

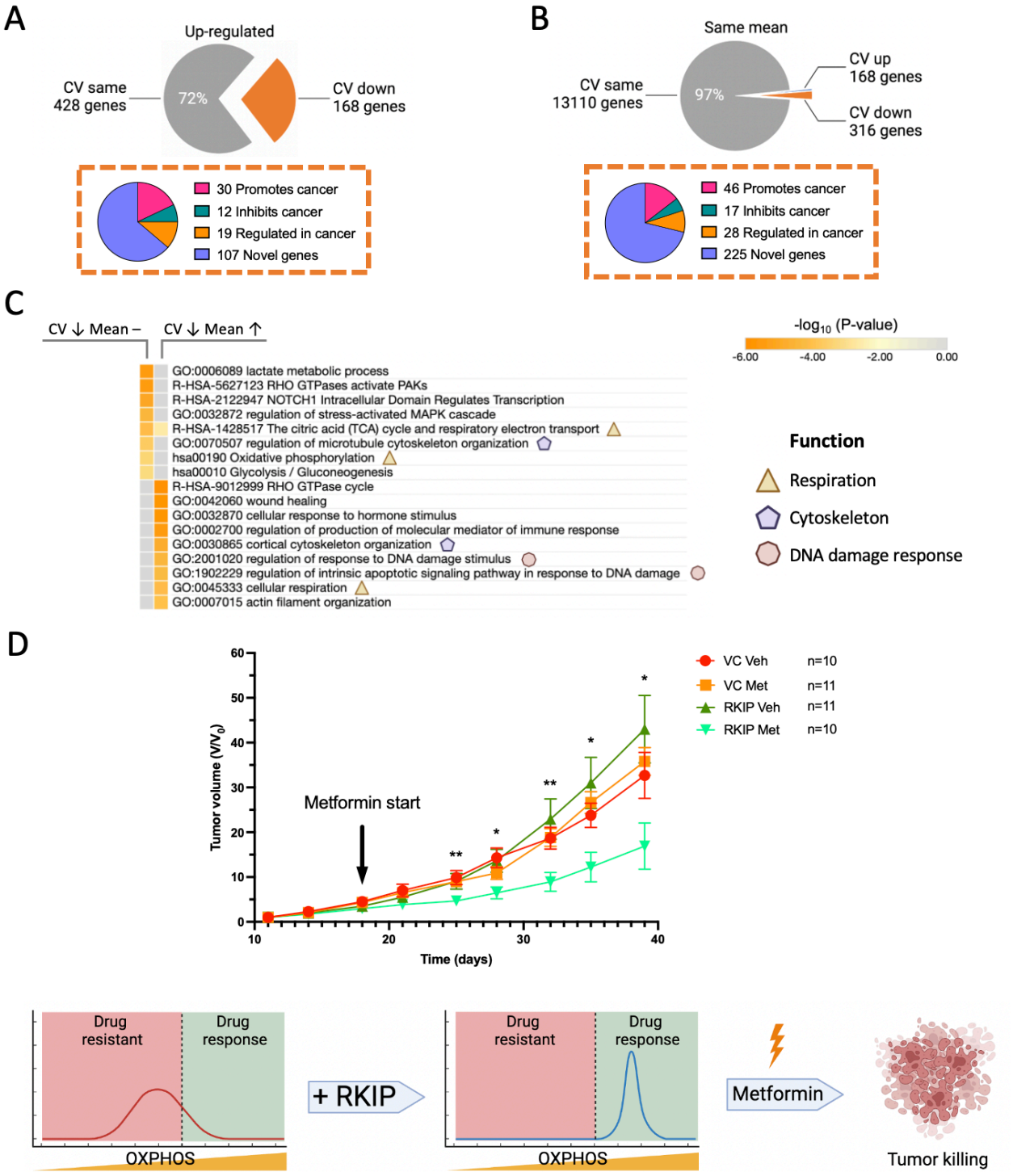


Figure 3. Functions of genes less variable in BM1-RKIP tumor cells. A) Genes with significantly higher mean in RKIP cells included 168 genes with significantly less CV and 428 genes without CV changes. Genes with less CV contained cancer-relevant genes as shown in the pie chart. B) 3% of all genes without mean changes had significantly higher or lower CV in RKIP cells. 316 genes less variable in RKIP cells contained cancer-relevant genes as shown in the pie chart. C) Significantly enriched gene sets from genes lessvariable in RKIP cells with either the same mean or up-regulated mean. Common

Figure 3 Continued
themes of gene sets are highlighted. D) Cells with high RKIP were less variable in OXPHOS and more sensitive to metformin treatment. Mouse tumor volumes are plotted as fold-changes over the initially measured volumes (V/V_0). Mice were implanted with BM1 control (VC) or RKIP-overexpressing (RKIP) tumor cells and were treated with either vehicle (Veh) or metformin (Met). Error bars indicate the standard error of the mean (SEM). Significant differences between VC Veh and RKIP Met were indicated (*: $P < 0.05$, **: $P<0.01$; Mann-Whitney test).

Change in CV	Change in mean	Number of Genes	% of all changed genes
↑	↑	0	0.00%
	–	54	2.85%
	↓	98	5.17%
↓	↑	168	8.86%
	–	316	16.67%
	↓	1	0.05%
–	↑	428	22.57%
	–	13110	Unchanged
	↓	831	43.83%

Table 1. Genes with significant changes in expression mean or CV in RKIP cells versus control cells.

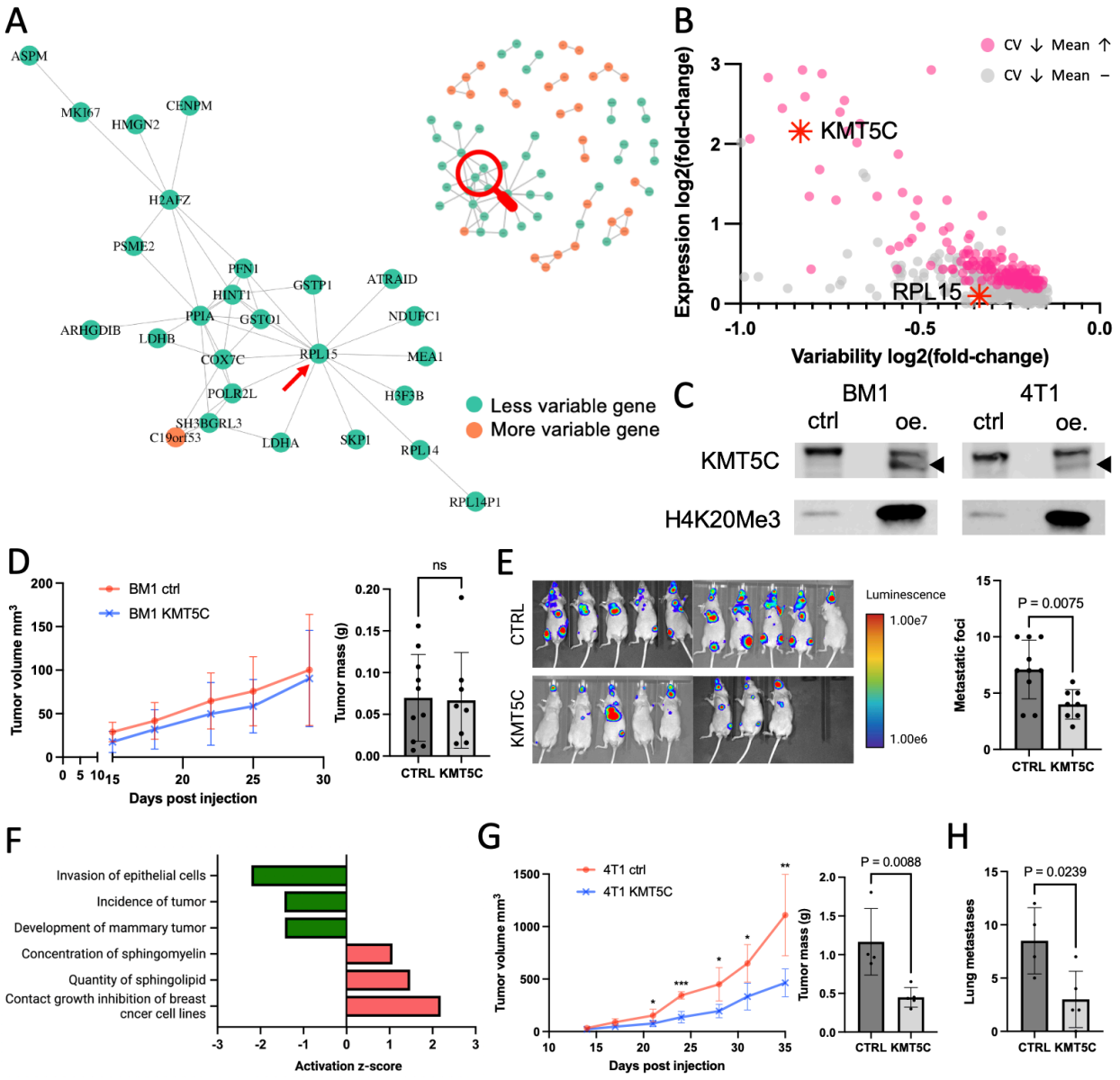


Figure 4. Genes less variable in BM1-RKIP cells include regulators of metastasis. A) Gene co-expression network within the RKIP cells with the largest subnetwork enlarged. B) KMT5C and RPL15 were less variable in RKIP cells. C) BM1 and 4T1 cells with stable KMT5C expression also had increased trimethylation of H4K20. D) Mouse tumor volumes of BM1 control or KMT5C and tumor mass at dissection. E) Left: luminescence measurement of mice 3 weeks post intracardiac injections of BM1 ctrl or KMT5C cells. Right: number of metastatic foci counted in each mouse. F) Activity changes of disease and biological functions in BM1 KMT5C tumors versus control tumors, predicted by IPA. G) Mouse tumor volumes of 4T1 control or KMT5C and tumor mass at dissection. H) Number of metastases on the surface of lungs from BALB/c mice with 4T1 ctrl or KMT5C primary tumors.

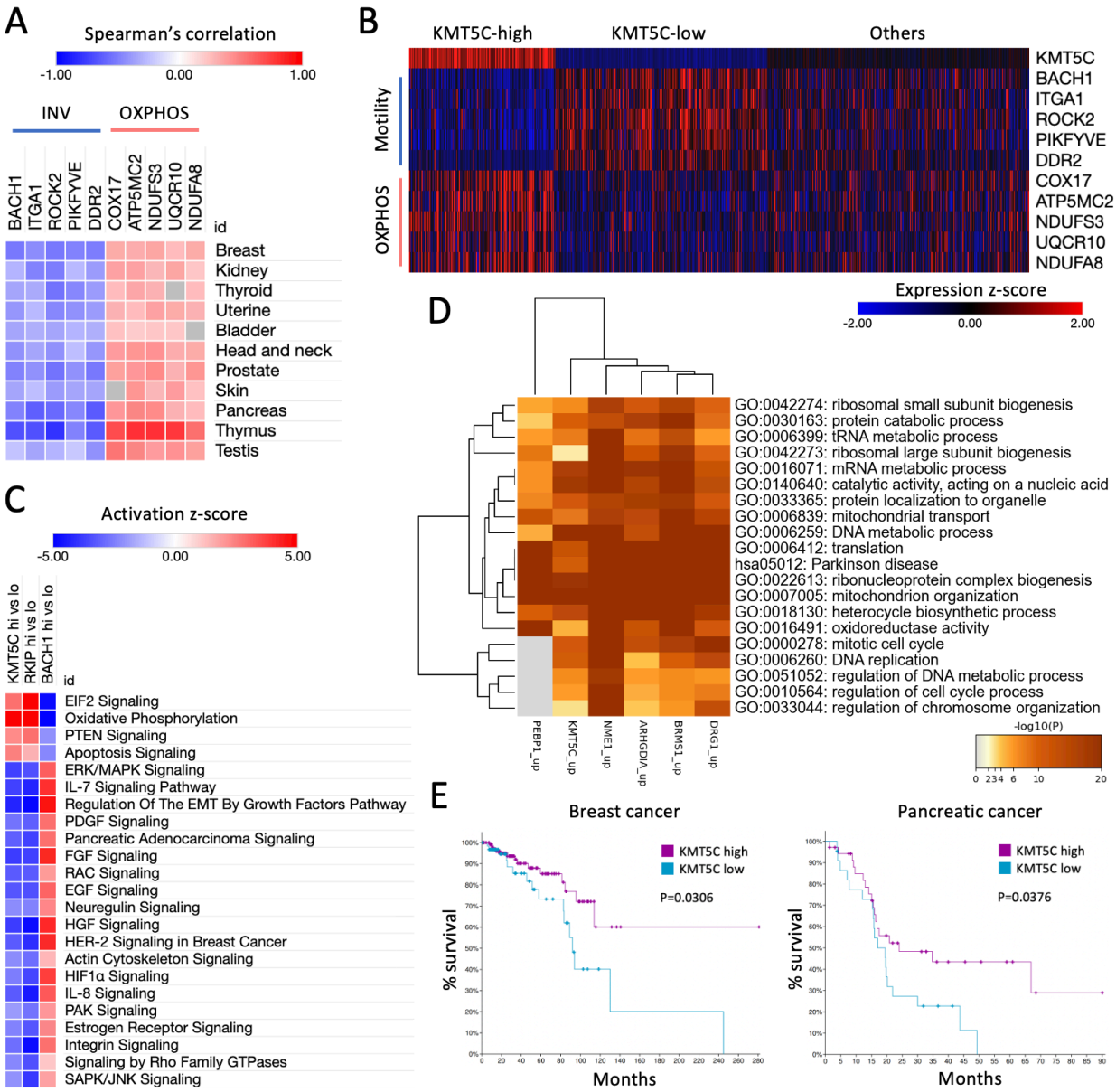


Figure 5. KMT5C expression correlates with better survival and less metastasis in patients. A) Spearman's correlation between KMT5C and genes related to metastasis in cancer patients. B) Relative expression levels of metastasis-related genes (INV/OXPHOS) in breast cancer patients with higher or lower levels of KMT5C. C) Predicted activation of canonical pathways in TNBC patients with high versus low levels of KMT5C, RKIP, or BACH1. D) Significantly enriched gene sets from genes positively correlated with PEBP1, KMT5C, NME1, ARHGDIA, BRMS1, or DRG1 in breast cancer patients. E) Kaplan-Meier plots of breast cancer (left) or pancreatic cancer (right) patients with high or low levels of KMT5C.

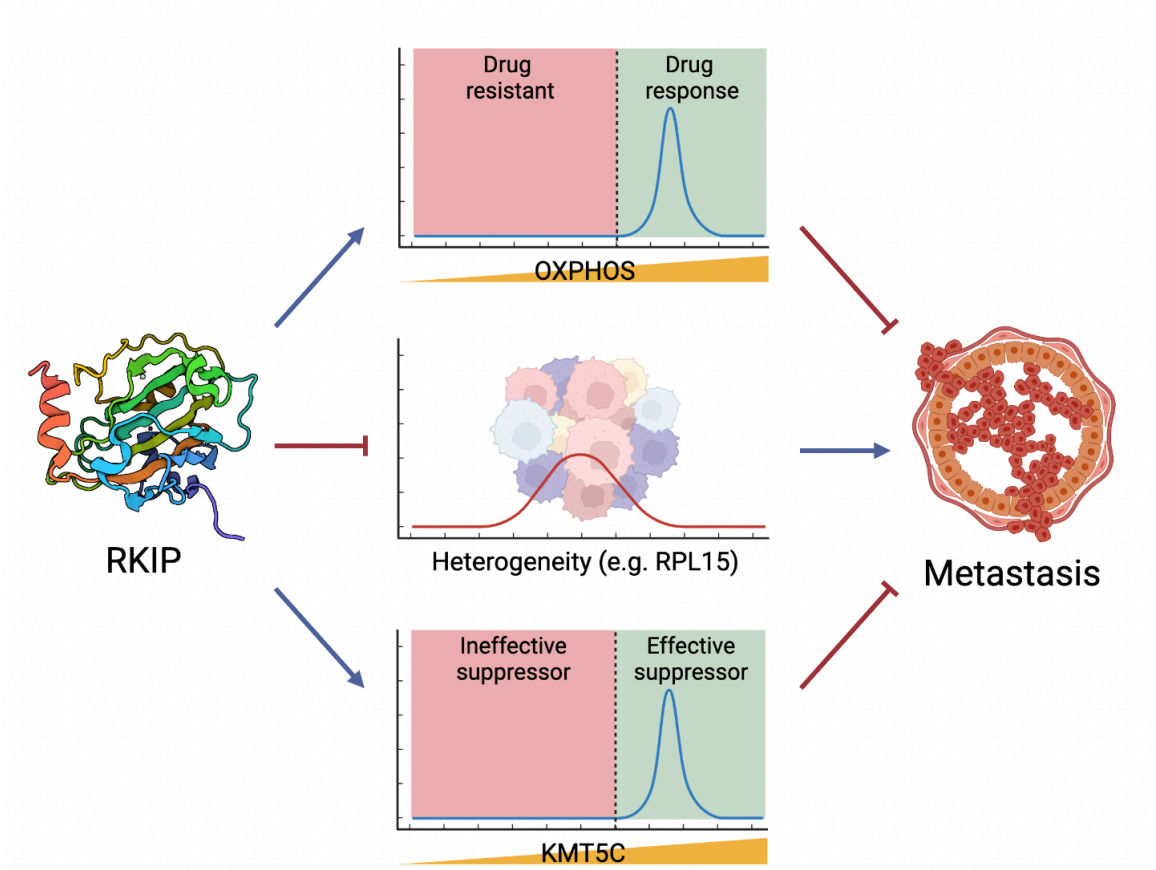
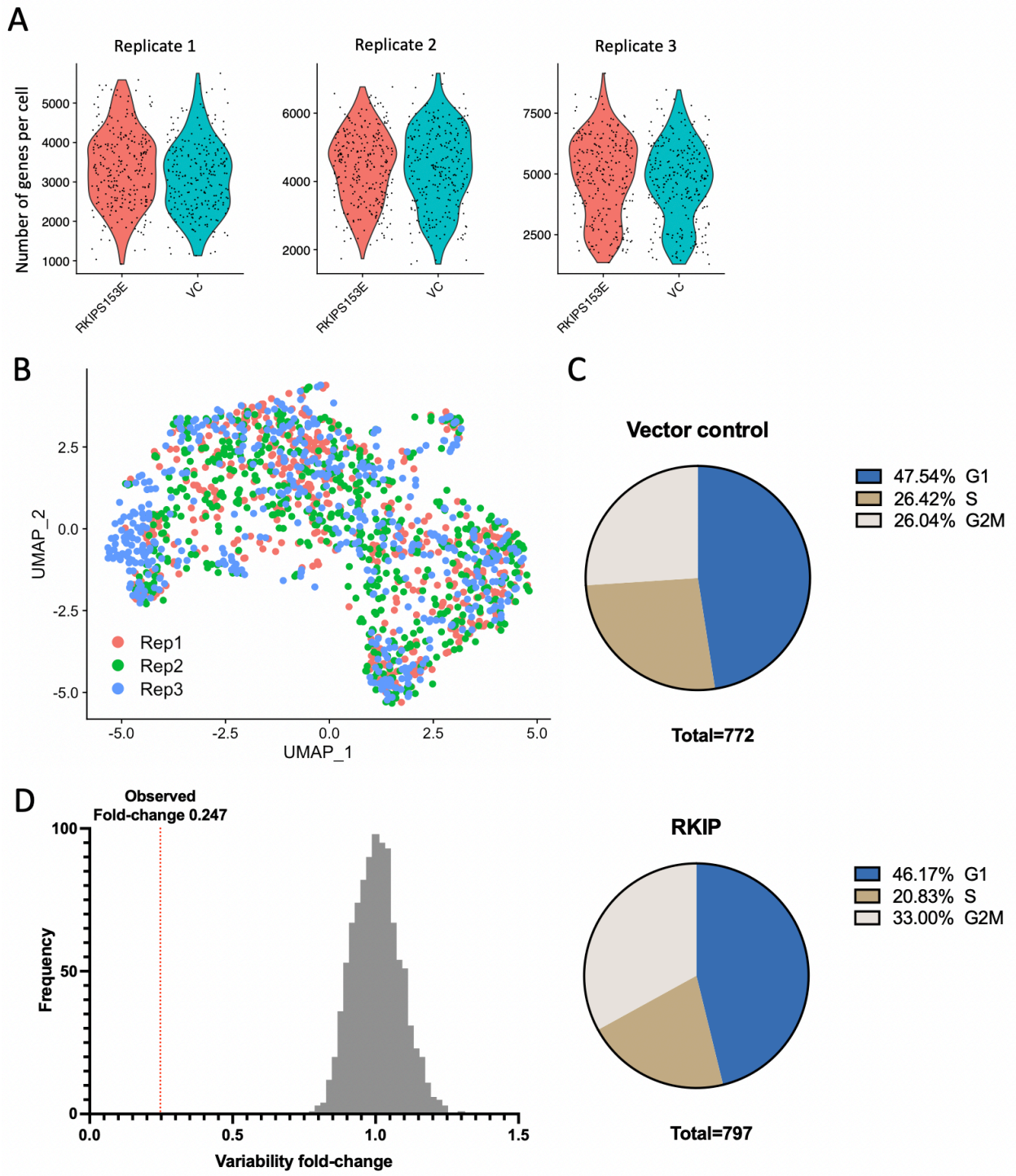
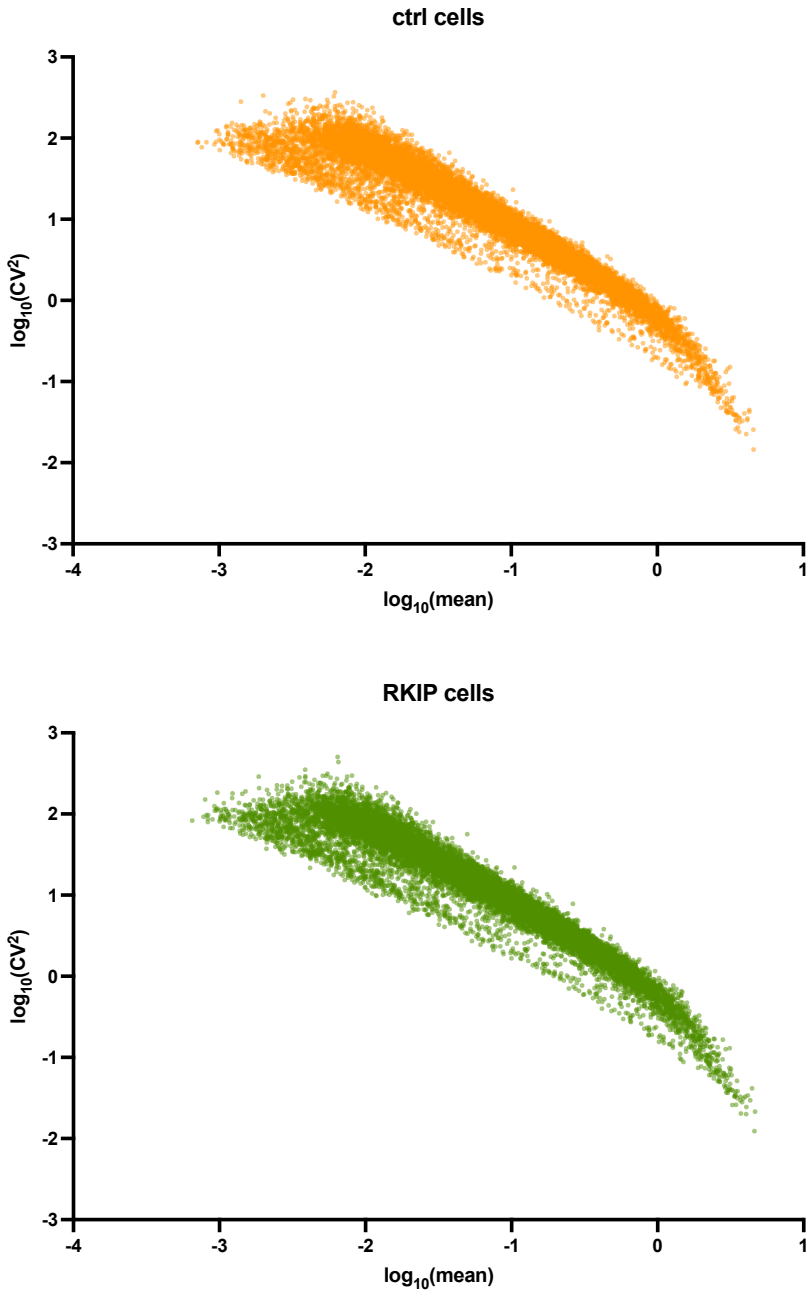


Figure 6. RKIP inhibits overall gene expression variability in non-metastatic tumor cells and reduces variability while up-regulating KMT5C and OXPHOS genes.



Supplementary Figure 1. Results from BM1-CTRL and BM1-RKIP scRNA-seq. A) Number of genes mapped in each tumor cell in BM1 ctrl or BM1 RKIP tumors from each replicate. B) UMAP clustering of single cells from each replicate. C) Distribution of cell cycles in control or RKIP cells. D) Observed fold-change of the overexpressed RKIP gene in BM1-RKIP cells compared to the distribution of results from 1000 permutations.



Supplementary Figure 2. Gene expression mean and variability of vector control and RKIP-expressing cells. Gene expression $\log_{10}(CV^2)$ and $\log_{10}(\text{mean})$ were plotted for all genes in BM1-ctrl cells (top) and BM1-RKIP cells (bottom).

A

	CV	Mean
	↑	–
RPS23P1	1.09	n.s.
LRP5L	0.97	n.s.
AC010834.3	0.96	n.s.
AXIN2	0.94	n.s.
RPS3P6	0.93	n.s.
ZDHHC23	0.90	n.s.
SPACA6	0.85	n.s.
PDLIM2	0.76	n.s.
AC099343.3	0.74	n.s.
FBXL6	0.72	n.s.

	CV	Mean
	↑	↑
N/A		

	CV	Mean
	–	↑
AC105020.2	n.s.	3.69
AC027796.1	n.s.	3.55
TAPBPL	n.s.	3.36
LINGO2	n.s.	3.31
AC012066.1	n.s.	3.26
MB21D2	n.s.	3.21
AC007688.1	n.s.	3.21
SOX5	n.s.	3.21
RPS6KB2	n.s.	3.07
ZNF337-AS1	n.s.	3.02

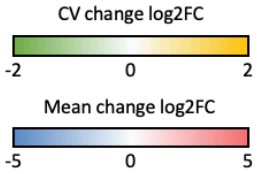
	CV	Mean
	↓	↑
PEBP1*	-2.02	5.95
OR7C1	-1.53	1.34
FZD5	-0.97	2.06
MAGEA11	-0.92	2.83
RIMS1	-0.88	2.45
KMT5C	-0.83	2.16
CCND2	-0.83	2.93
AC093484.4	-0.82	2.59
PTPRN2	-0.81	1.34
POLD2P1	-0.80	0.43

	CV	Mean
	↓	–
MARK1	-1.00	n.s.
ROGDI	-0.99	n.s.
HLA-DRB6	-0.95	n.s.
TMEM54	-0.93	n.s.
SERINC5	-0.90	n.s.
HLA-F	-0.87	n.s.
AC046143.1	-0.82	n.s.
RPL7P26	-0.75	n.s.
TBC1D10A	-0.72	n.s.
AC095055.1	-0.72	n.s.

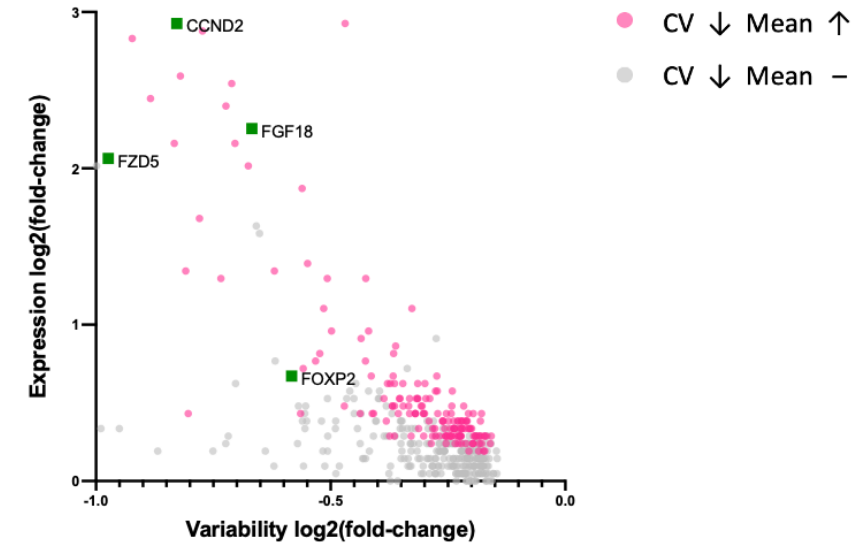
	CV	Mean
	↓	↓
MALAT1	-0.22	-0.24

	CV	Mean
	–	↓
AP004609.3	n.s.	-3.89
ZNF213	n.s.	-3.55
SERPING1	n.s.	-3.41
PCDHB2	n.s.	-3.36
OBSCN	n.s.	-3.31
AC007036.2	n.s.	-3.21
SAXO1	n.s.	-3.21
AC006486.2	n.s.	-3.21
POLD4	n.s.	-3.17
HIPK1-AS1	n.s.	-3.17

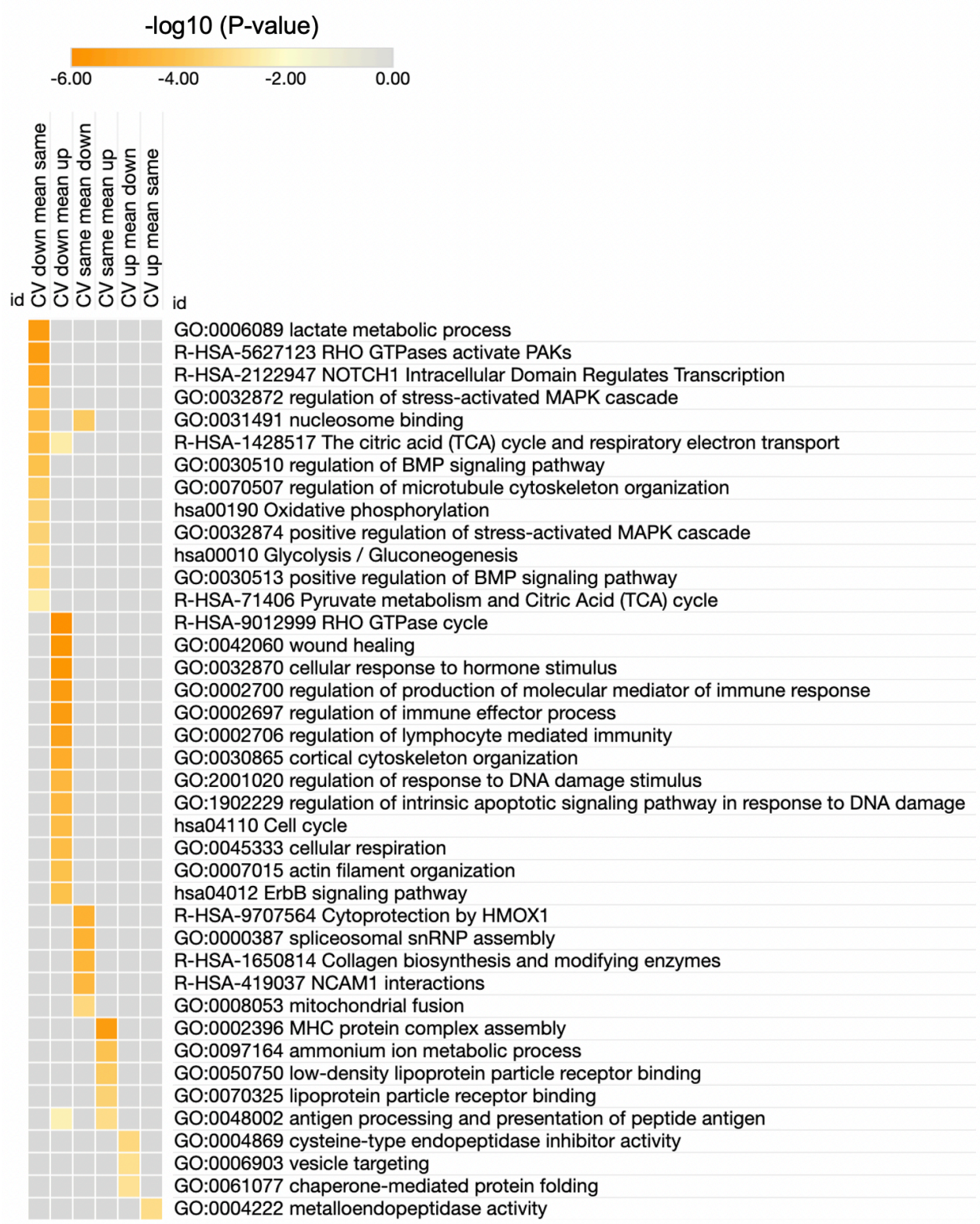
	CV	Mean
	↑	↓
CTAG2	1.92	-4.17
CASTOR3	1.46	-0.67
CMKLR1	1.23	-3.36
CSRNP3	1.00	-0.58
DLK2	0.94	-0.58
TRAF4	0.86	-3.21
LINC02035	0.84	-2.93
AC010618.4	0.82	-1.20
TNS4	0.80	-2.93
CLCN4	0.77	-2.59



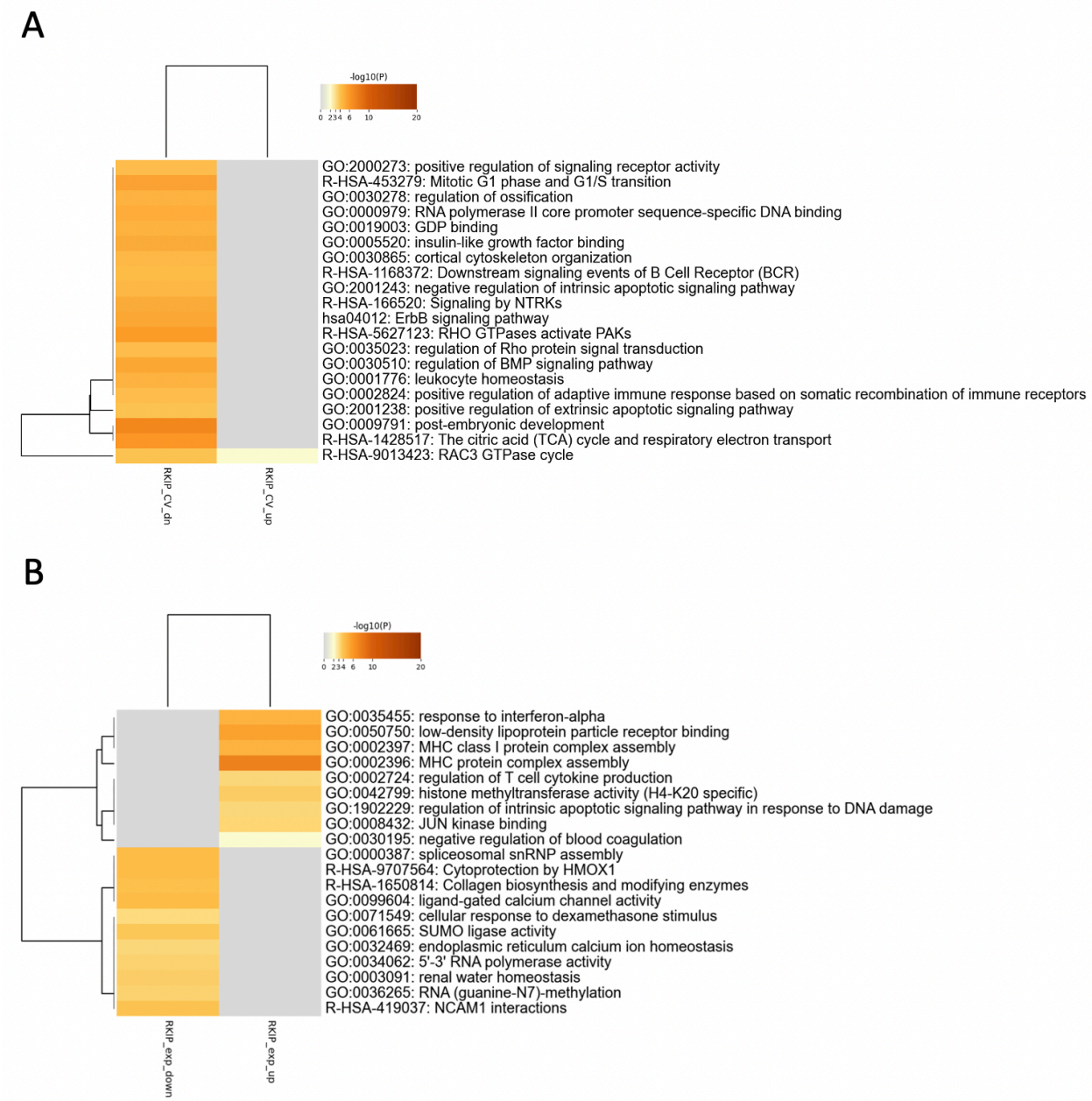
B



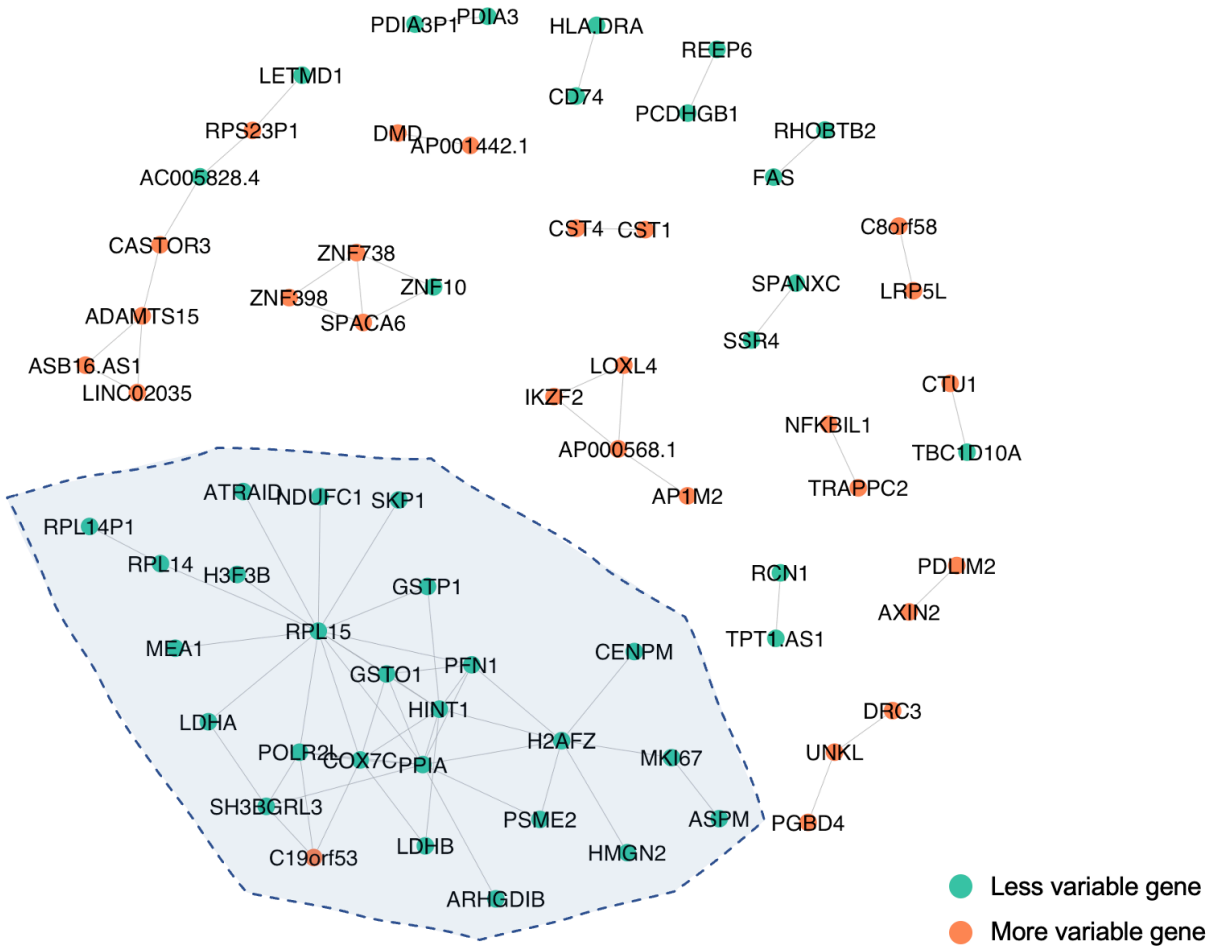
Supplementary Figure 3. Genes with significant changes in mean or CV in RKIP cells versus control cells. A) Top genes from each direction of Fig. 2A. B) Genes with significantly less CV changes in BM1-RKIP cells are indicated with examples of cancer-related genes labeled.



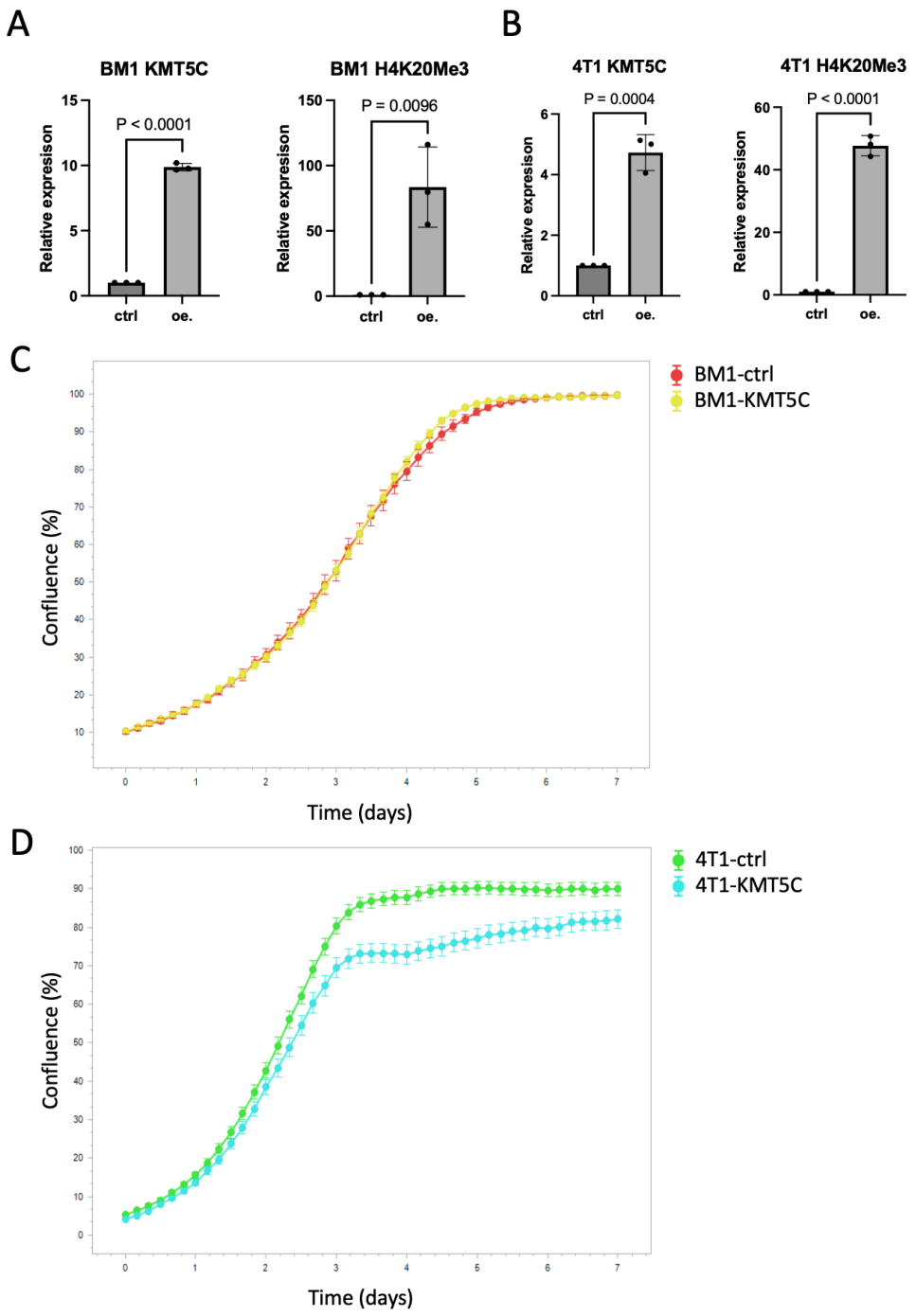
Supplementary Figure 4. Significantly enriched gene sets from genes with changes in mean and/or CV in RKIP versus control cells.



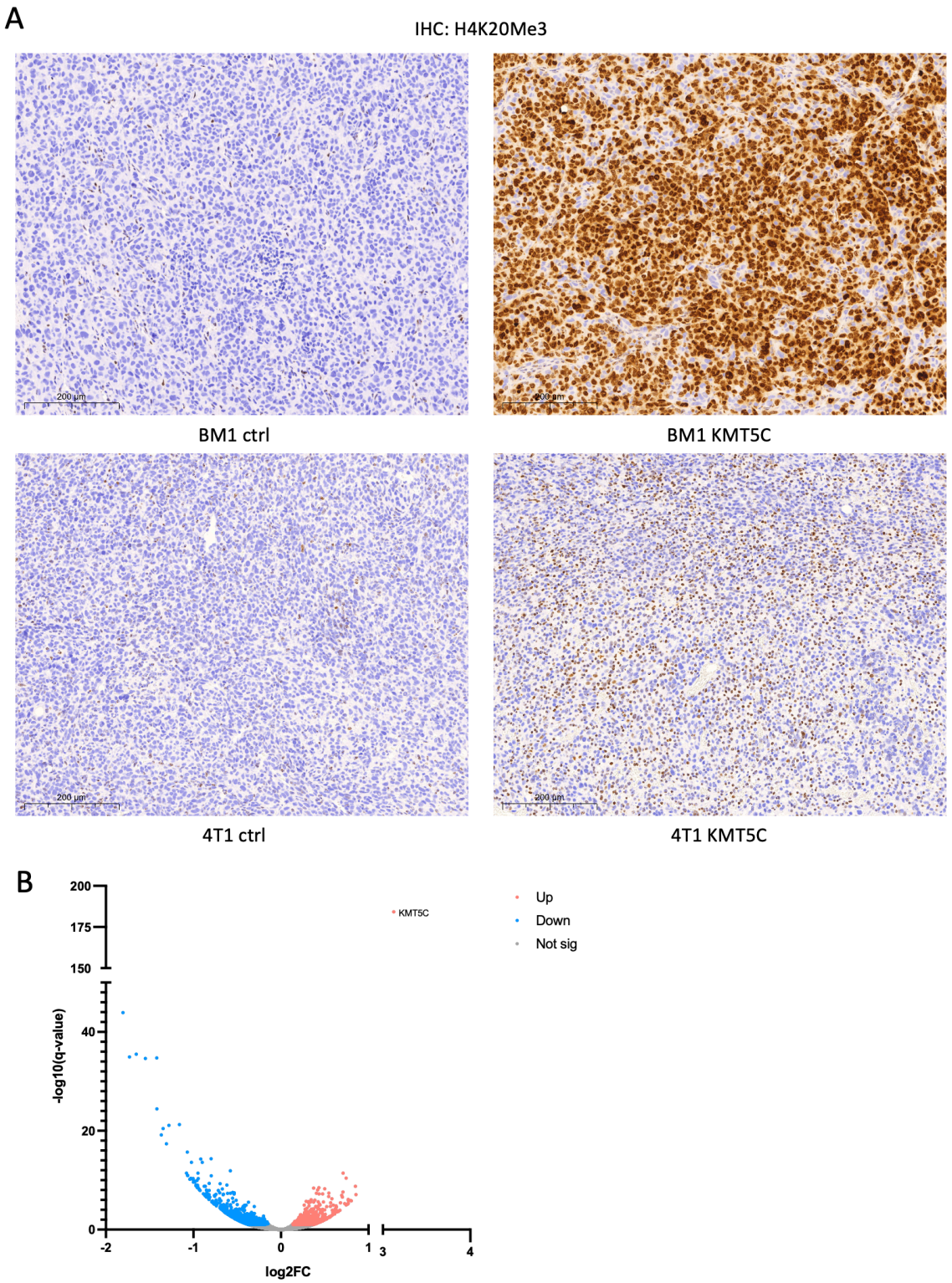
Supplementary Figure 5. Significantly enriched gene sets from genes with changes in mean or CV in RKIP versus control cells. A) Enriched gene sets based on Metascape analysis of all genes with higher or lower CV in RKIP cells. B) Enriched gene sets from all genes up or down-regulated in RKIP cells.



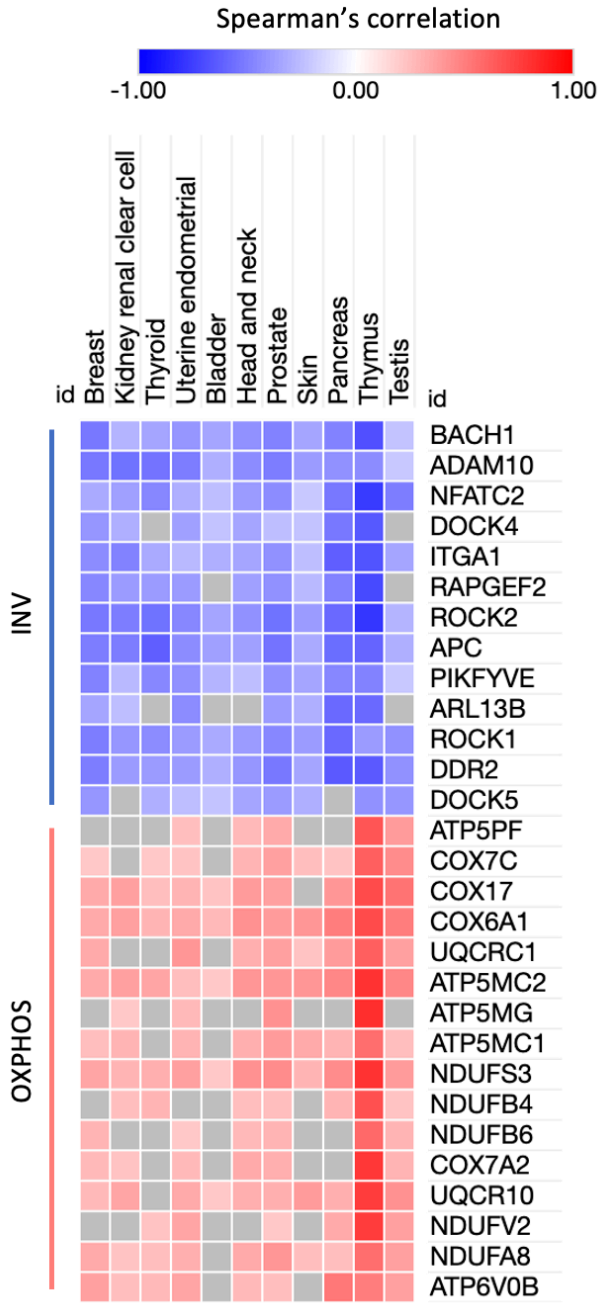
Supplementary Figure 6. The complete gene co-expression network within RKIP cells from Fig. 3A.



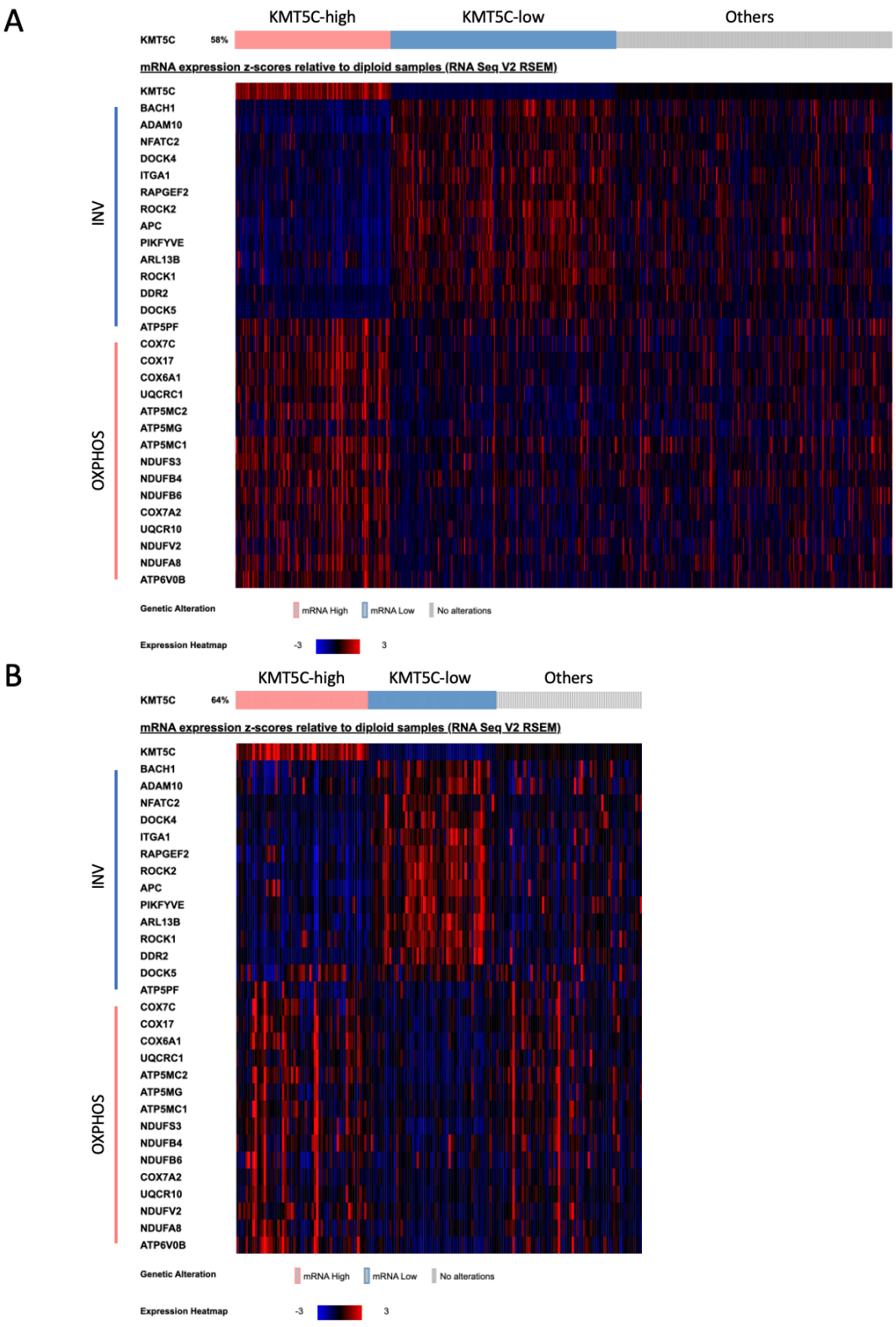
Supplementary Figure 7. Validation of stable KMT5C expression in TNBC cell lines. A) Expression levels of KMT5C or H4K20Me3 in BM1 cells with (oe.) or without (ctrl) KMT5C expression. B) Expression levels of KMT5C or H4K20Me3 in 4T1 cells with (oe.) or without (ctrl) KMT5C expression. C) Proliferation curves of BM1-ctrl and BM1-KMT5C cells. D) Proliferation curves of 4T1-ctrl and 4T1-KMT5C cells.



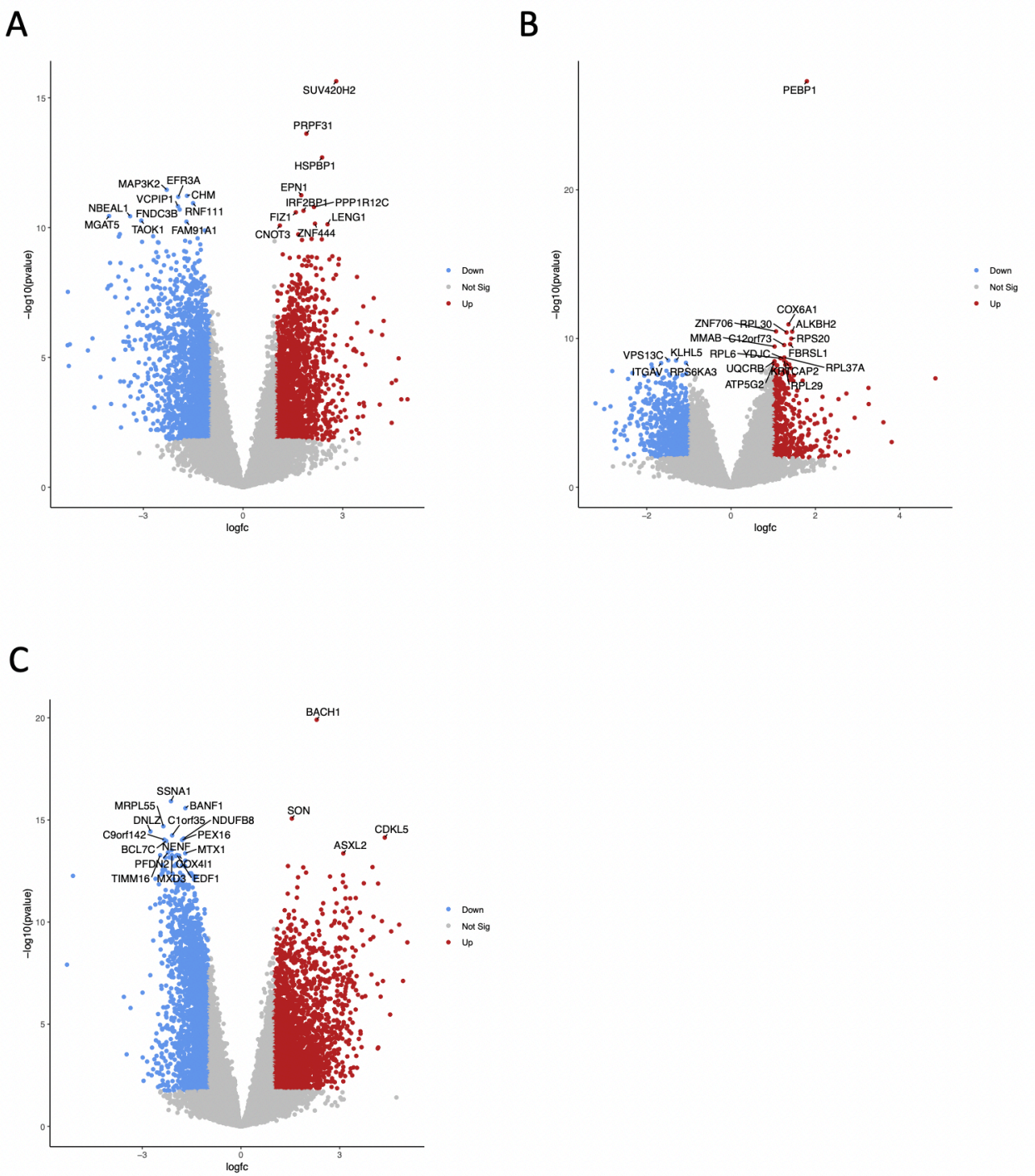
Supplementary Figure 8. Effect of KMT5C overexpression in TNBC tumors in mice. A) IHC of H4K20Me3 in BM1 or 4T1 tumors with or without KMT5C expression. B) Volcano plot showing gene expression changes in BM1 tumors with KMT5C expression versus control tumors.



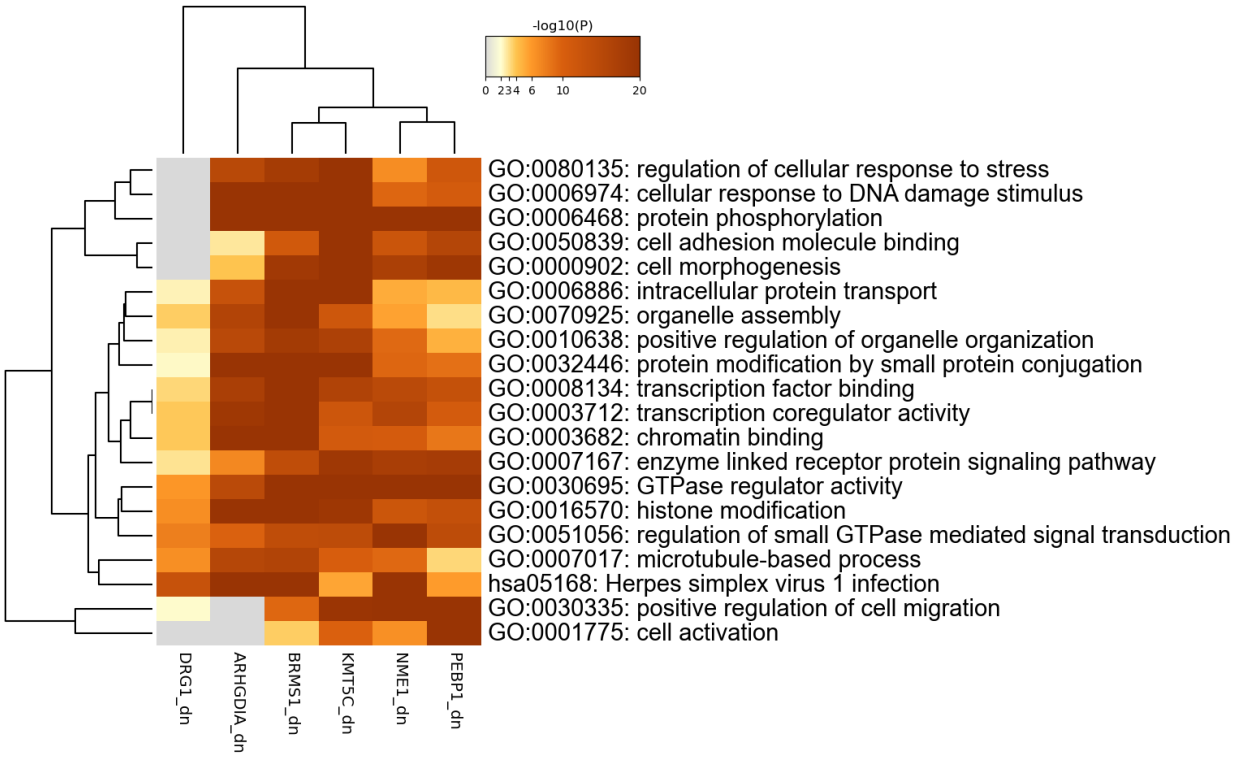
Supplementary Figure 9. Spearman's correlation between KMT5C and genes related to metastasis in cancer patients.



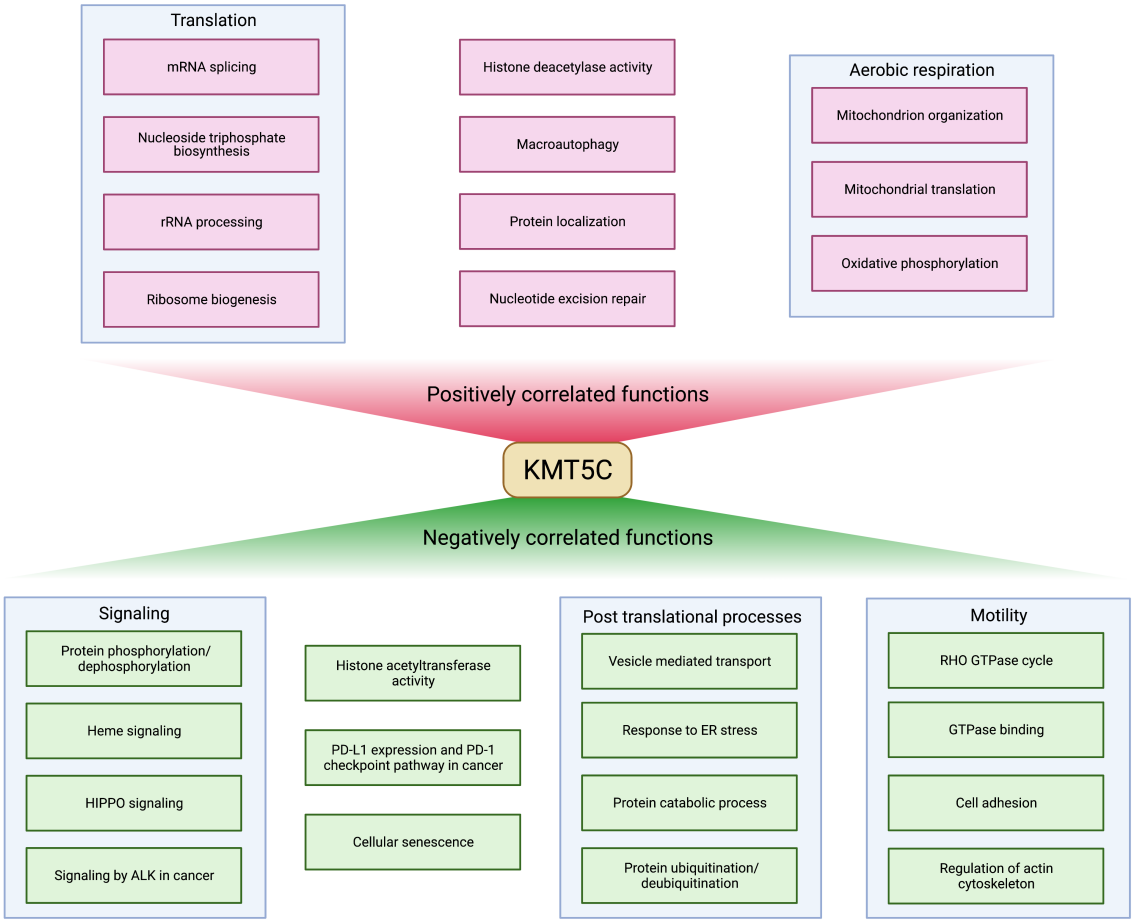
Supplementary Figure 10. Relative expression levels of metastasis-related genes in breast cancer patients (A) or pancreatic patients (B) with higher or lower levels of KMT5C.



Supplementary Figure 11. Gene differential expression analyses in TNBC patients. A) Volcano plot showing gene expression changes in TNBC tumors with higher versus lower levels of KMT5C. B) Volcano plot showing gene expression changes in TNBC tumors with higher versus lower levels of RKIP. C) Volcano plot showing gene expression changes in TNBC tumors with higher versus lower levels of BACH1.



Supplementary Figure 12. Significantly enriched gene sets from genes positively correlated with PEBP1, KMT5C, NME1, ARHGDIA, BRMS1, or DRG1 in breast cancer patients. Analyses of data from TCGA were done using Metascape.



Supplementary Figure 13. Significantly enriched gene sets positively or negatively correlated with KMT5C expression level in breast cancer patients.

CHAPTER 4. THE METASTATIC TRANSCRIPTION FACTOR BACH1 PROMOTES OVERALL GENE EXPRESSION VARIABILITY OF TUMOR CELLS

Abstract

Tumor cell heterogeneity leads to the progression of metastasis and, eventually, death of the patient. BTB and CNC homology 1 (BACH1) is a transcription factor that promotes metastasis in many solid cancers, including TNBC. We hypothesize that BACH1 regulates overall gene expression variability and promotes an overall increase in transcriptional heterogeneity among metastatic tumor cells. We performed single-cell RNA sequencing to compare the gene expression variability in TNBC tumors with BACH1 overexpression or BACH1 depletion. Although BACH1 expression was increased and less variable in BACH1-overexpressing cells, the overall gene expression variability was increased. Notably, genes related to epithelial-mesenchymal transition and stem cell pluripotency were both up-regulated and less variable in cells with BACH1 overexpression. Gene co-expression network analysis revealed a network of cell cycle-related genes with collectively increased variability in BACH1-overexpressing cells. Interestingly, tumor cell populations with BACH1 depleted by shRNA also displayed increased variability in gene expression despite the loss of the metastatic phenotype. These results indicate that BACH1 overexpression increases overall noise but decreases the transcriptional variability of many pro-metastatic genes, while BACH1 knockdown can achieve metastatic suppression despite high transcriptional noise. Taken together, the combined analysis of gene expression variability along with mean highlights key BACH1-

regulated pathways and provides novel insights into the association of transcriptional heterogeneity with metastatic progression.

Introduction

Elucidating the relationship between gene expression variability and metastatic progression can reveal effective therapeutic targets and novel regulators of metastasis. In the previous chapter, we showed that overexpression of the metastasis suppressor RKIP reduces the overall expression variability of tumor cells. We also identified potentially targetable pathways such as oxidative phosphorylation and showed that KMT5C, a gene that was up-regulated and less variable in non-metastatic RKIP-expressing cells, is a novel metastasis suppressor in TNBC.

To determine if pro-metastatic genes are associated with increased transcriptional variability, we conducted single-cell analysis of TNBC cells transfected with the pro-metastatic gene BACH1. Consistent with this possibility, BACH1-overexpressing tumor cells were more variable than the control cells, indicating that the biological functions of the overexpressed gene contribute to the overall expression variability. Analyses of genes with mean and CV changes revealed previously known and novel pathways regulated by BACH1, leading to potential discoveries of additional BACH1-regulated pathways.

These results are consistent with a causal relationship between gene expression variability and metastasis. To investigate this possibility further, we generated a non-metastatic cell population with an alternative strategy involving the depletion of BACH1 by shRNA (*37*). Tumor cells were analyzed by scRNA-seq to compare expression variability between BM1 shBACH1 and shCTRL tumor cells. Surprisingly, shBACH1 tumor cells were much more variable than shCTRL cells despite their non-metastatic phenotype. This observation suggests that an overall increase in gene expression variability is not sufficient to drive tumor cells to a metastatic state. Taken together, these

results reveal the complexity of the relationship between expression variability and metastasis, increasing our understanding of how metastasis regulators work.

Results

Overexpression of metastasis promoter BACH1 increases gene expression variability

Previous results demonstrated that overexpression of the metastasis suppressor RKIP leads to a decrease in gene expression variability relative to control cells. These results are consistent with the possibility that overexpression of a pro-metastatic gene such as BACH1 will increase gene expression variability. An alternative explanation is that gene overexpression generally drives a population of cells to a more homogeneous state. To test these hypotheses and to use the coefficient of variation (CV defined by σ/μ) as an additional statistical metric to elucidate metastatic drivers, we overexpressed BACH1 in BM1 cells. We then analyzed xenograft tumors with or without mouse BACH1 gene overexpression (BM1-mmBACH1 versus BM1-CTRL) by single-cell RNA-seq using three independent replicates (two CTRL tumors and two mmABCH1 tumors per replicate). The distribution of the number of genes detected per cell was comparable between the BACH1 overexpression and control tumor samples for each set of replicates (Fig. 7A), and the single-cell gene expression profiles were reproducible across the three replicates (Fig. S14A). Unlike shBACH1 cells, which have more S phase cells, mmBACH1 cells have fewer S phase cells (16% versus 24% in CTRL cells), while the proportion of G1 cells is increased (54% versus 41% in CTRL cells) (Fig. S14B). As observed previously for RKIP, UMAP clustering of CTRL and mmBACH1 cells could not effectively separate the metastatic from non-metastatic cell types (Fig. 7B).

Permutation analysis on gene expression CV between the two tumor cells indicates a significant change in overall expression variability from CTRL to mmBACH1

cells (Fig. 7C). We found that 883 genes were significantly more variable (more heterogeneous) in mmBACH1 cells and 535 genes were less variable (Fig. 7D; Table 2). Thus, the overall expression variability in mmBACH1 cells increased compared to CTRL cells despite BACH1 itself being less variable (Fig. S14C).

Consistent with either a classic Poisson process or intermittent bursts, the slope for $\log_{10}(CV^2)$ versus $\log_{10}(\mu)$ for both mmBACH1 and CTRL cells approached -1. This would be expected as most cells exhibit no change in mean or CV (Fig. S15). The difference between the number of genes with increased or decreased CV was relatively minor, resulting in a modest shift in CV when we plotted the genes with significant CV changes (Fig. 8A). However, comparing the Fano Factor (FF) differences between mmBACH1 and CTRL cells, the majority of the genes with CV increases independent of changes in expression mean had higher FF in mmBACH1 cells (Fig. 8B, C), consistent with an increase in gene expression bursting.

Analysis of changes in CV and mean between mmBACH1 and CTRL cells also demonstrated a relationship that resembles Poisson scaling or bursting (Fig. 8D); the pattern was similar to that observed in the previous chapter for RKIP versus control cells (see Fig. 2D). Interestingly, the proportion of genes that are up-regulated is higher than those that are down-regulated and represent ~85% of all genes with decreased CV (Table 2). These results are consistent with BACH1's role as a promoter of metastasis and indicate that overexpression of a gene does not necessarily result in a cell population with overall decreased gene expression variability.

CV analyses of BACH1 overexpression reveal pathways related to metastasis

We next took advantage of CV as an additional metric to understand mechanisms by which BACH1 overexpression drives cell function. We first examined the networks of co-expressed genes with increased or decreased CV upon mmBACH1 expression (Fig. S16). In contrast to RKIP overexpressing cells that displayed one major co-expression network (see Fig. 4A), mmBACH1-overexpressing cells revealed two major networks (Fig. S16). The largest network, with 81 closely-linked genes, is almost exclusively comprised of more variable genes and thus lower mean expression. Metascape analysis of genes in this cluster indicates they are primarily involved in cell cycle processes (Fig. S17). The second major network, with 14 more variable and 24 less variable genes, represents a more diverse set of pathways known to be regulated by BACH1, such as OXPHOS, iron metabolism, and migration (*23, 28, 29, 37, 42*).

CV/mean changes in BACH1-overexpressing cells broadly fall into distinct groups of canonical pathways (Fig. 9). Many cancer-related signaling pathways are implicated in the subset of genes with reduced CV and increased mean. Among this group are genes related to epithelial-mesenchymal transition (EMT) and cell motility, pathways previously reported to be regulated by BACH1 (*23, 39-41*). Interestingly, analyzing both CV and mean revealed some functions that were not among the most enriched pathways when only expression mean changes were analyzed (Fig. S18A). These include stem cell pathways that were among the most significant among genes with decreased CV (Figs. 9, S18B). OXPHOS is among the pathways enriched in down-regulated genes without changes in CV, supporting previous observations that wild-type BM1 cells with relatively high BACH1 levels did not respond to metformin treatment (*37*). These results suggest

that the combination of CV and mean can reveal novel functions of pro-metastatic genes such as BACH1.

BACH1 knockdown suppresses metastasis while generating transcriptional noise

As previously demonstrated, overexpression of the metastasis suppressor RKIP in tumor cells resulted in a reduction in transcriptional variability. To determine whether variability in gene expression is always reduced when metastatic tumors transition to non-metastatic, we tested an alternative strategy to inhibit metastasis. Rather than overexpression of a metastasis suppressor, we inhibited the metastasis promoter BACH1 from switching tumor cells from a metastatic to a non-metastatic phenotype (*37*).

To analyze the effect of BACH1 depletion on transcriptional variability, we performed single-cell RNA-seq of BM1 cells expressing shCTRL or shBACH1 in xenograft tumor samples. For these studies, we used mixed, polyclonal shBACH1 populations that significantly reduced overall tumor metastasis. Cells derived from shBACH1 and shCTRL BM1 tumors (2 mice per tumor type) were collected and analyzed as described above. The distributions of the number of genes detected per cell were comparable between the shCTRL and shBACH1 tumor cells (Fig. 10A).

Interestingly, while both cell types have a similar proportion of cells in the G1 phase, 26% of the shBACH1 cells are in the S phase, while only 14% of the shCTRL cells are in the S phase (Fig. S19A). These differences in cell cycle distribution indicate opposite effects on the cell cycle when BACH1 is up or down-regulated.

Similar to RKIP or BACH1 overexpression, UMAP clustering of shCTRL and shBACH1 cells could not effectively separate the metastatic states, although shBACH1

cells appear to predominantly segregate in the lower half of the cluster (Fig. 10B). Analysis of gene expression CV across all genes in both tumor types indicates a substantial difference in gene expression variability. We performed permutation analysis as previously described and identified a significant increase in overall variability in shBACH1 cells relative to shCTRL cells (Fig. 10C). Using false discovery rate (FDR)-adjusted P-values derived from the permutation analysis, we found that 3971 genes were significantly more variable (noisier and more heterogeneous) in shBACH1 cells (including the down-regulated BACH1 itself) and 83 genes were less variable (Figs. 10D, S19B).

As observed previously for RKIP and BACH1 overexpression, the majority of the genes were unchanged for both mean and CV. Thus it was not surprising that the slopes for $\log_{10}(CV^2)$ versus $\log_{10}(\mu)$ for both shCTRL and shBACH1 cells were similar and approached -1, characteristic of both a classic Poisson process and intermittent bursts (Fig. S20). However, given that the number of genes with increased CV far exceeds the number of genes with decreased CV in shBACH1 cells, when we plotted the genes whose expression was significantly changed, we observed a significant upward shift for shBACH1 cells relative to shCTRL cells (Fig. 11A) and a corresponding increase in FF (Fig. 11B). These results are consistent with the increase in variability in shBACH1 cells we observed using the CV-based permutation analysis.

It has been previously shown that an increase in the gene expression mean is often associated with a reduction in gene expression variability of noise, as measured by CV, and vice versa (*67, 68, 94*). The association between expression mean and variability or noise can be altered by factors including molecular perturbations, epigenetic changes, and feedback or general regulatory pathways, potentially resulting in differences in

cellular functions and survival (*67, 68, 95*). To analyze this relationship, we next plotted differential $\Delta\log_{10}(CV^2)$ versus $\Delta\log_{10}(\text{mean})$ for genes expressed in shBACH1 or shCTRL cells and looked at the nine possible scenarios for changes in mean and/or CV (increase, decrease, or unchanged) in cells transitioning to a non-metastatic phenotype. Notably, unlike overexpressed RKIP or BACH1 versus control cells (see Chapter 3), the pattern we obtained did not indicate a constraint for changes to follow the scaling of Poisson or bursting processes (Fig. 11C). Among all the genes with significant changes in mean or variability expression, the biggest proportion of the gene expression changes was increased CV with the same mean, representing ~42% of all changed genes (Table 3). In contrast, only ~1% of all changed genes had reduced CV. These results suggest that the introduction of shBACH1 RNA acted as a noise generator in the resulting cell population, causing a significant departure from normal transcriptional processes. Furthermore, they demonstrate that tumor cell transitions from metastatic to non-metastatic phenotypes are not always associated with a decrease in CV and gene expression variability.

Discussion

In this study, we used the metastasis promoter, BACH1, to modulate tumor cell metastatic states and compared the resulting changes in expression variability to the outcomes of overexpressing the metastasis suppressor RKIP. Our results showed that changes in the metastatic state may not always result in changes in variability in the same direction. The overall variability changes in the tumor cells can be affected by the biological functions of the metastasis regulator in question and the noisiness of the regulator itself. Our analyses of BACH1 overexpression also revealed novel pathways when changes in CV were taken into account in addition to mean.

We showed that gene expression noise was dramatically increased in the non-metastatic shBACH1 tumor cells. However, it is likely that the variability increase was caused by non-uniform reduction in BACH1 via shRNA. Given that BACH1 knockdown has been shown to inhibit metastasis, we expect that a complete knockout of BACH1 expression, which can be achieved through CRISPR-Cas9, will also lead to similar, if not more potent, inhibition of metastasis.

Our results showed differences in cell cycle distribution in BACH1-low and BACH1-high tumor cells (see Figs. S14B, S16A). BACH1-overexpressing tumor cells had a higher proportion of G1 phase cells than the control cells. We collected the single cells from twelve tumors over three independent biological replicates. Thus, the differences in cell cycle distribution were unlikely to be caused by chance. Previous studies of BACH1 in cancer and normal cells suggest its potential as a cell cycle regulator. As a transcription factor, BACH1 can repress transcription of *CDKN2A* in colorectal cancer, preventing cell cycle arrest (*114, 115*). BACH1 has also been shown to contribute to cell proliferation via

protein-protein interactions with the hyaluronan-mediated motility receptor (*48, 49*). Determining cell cycle from gene expression is an indirect measurement as the cell cycle states of single cells were inferred from expression levels of known cell cycle-related genes (*64, 116*). Furthermore, we did not observe significant differences in xenograft tumor growth between BACH1-high and control tumors within each biological replicate. Therefore, to elucidate BACH1's effect on the cell cycle, additional studies that use direct measurements, such as flow cytometry, are needed. We also recognize that the wild-type BM1 tumor cells have high BACH1 levels. Thus any effects, such as a potential difference in tumor growth resulting from further overexpressing BACH1, may be minor and difficult to detect. One way to magnify such differences is by comparing BACH1-null and BACH1-high tumors, such as tumors with BACH1 knockout and BACH1 overexpression.

As a transcription factor that promotes metastasis, BACH1 has been shown to promote EMT in many cancers, including TNBC. Our results showed that in BACH1-overexpressing BM1 cells, the EMT pathway is among the ones enriched from genes up-regulated and less variable. In particular, *SNAI2* and *HMGA2* encoding known promoters of EMT and metastasis were up-regulated and less variable in BM1 cells with BACH1 overexpression (*39-41*). In addition, HMGA2 was previously shown to be negatively regulated by RKIP, further supporting the mutually repressive relationship between BACH1 and RKIP (*15*).

Other genes that were up-regulated and less variable may reveal additional mechanisms of metastatic regulation by BACH1. We have previously shown that BACH1 promotes metastasis in part by increasing cell motility and invasiveness in breast cancer via Rho signaling (*23*). We identified more genes as potential effectors of BACH1 to

promote motility and invasiveness, such as *FN1* and *ITGAV*, which are up-regulated and less variable, and they encode proteins that promote metastasis in some cancers (*117, 118*). We also identified several pathways related to stem cell pluripotency among the same group of genes. It has long been suggested that BACH1 plays a role in cancer cell stemness, but direct evidence or mechanisms remain elusive (*43, 44*). These pathways could not be easily revealed when only changes in expression mean were analyzed (see Fig. S18A).

Our findings provide further insight into how BACH1 may regulate this process, and the genes found within these pathways could be potential effectors. Taken together, our results increase our understanding of metastatic processes and highlight once again the importance of using two statistical metrics, CV and mean, to uncover novel pathways and targets for metastatic cancer.

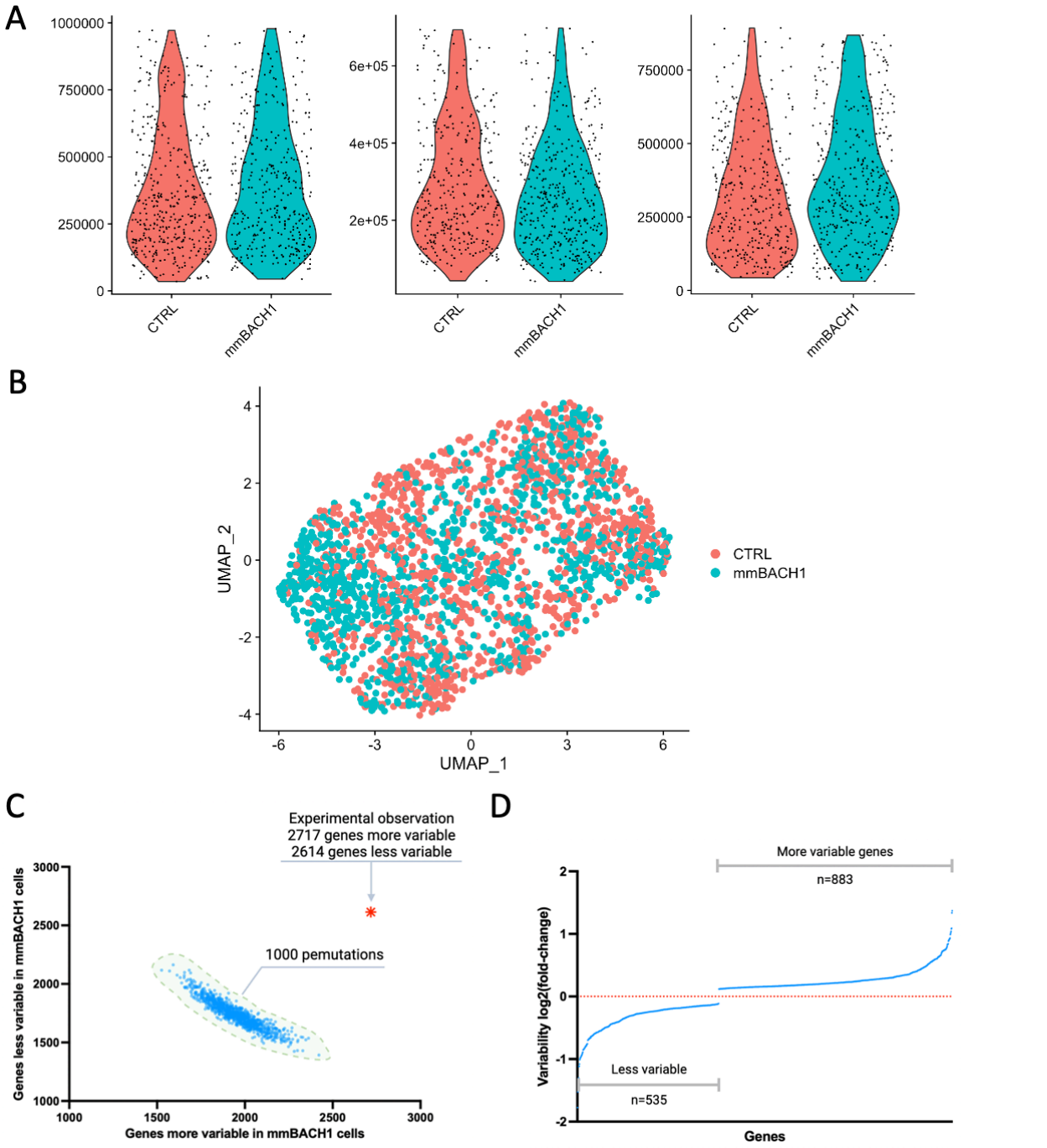


Figure 7. BM1 tumor cells with BACH1 overexpression have higher overall expression variability versus control cells. A) Number of genes mapped in each CTRL and mmBACH1 tumor cell within three replicates. B) UMAP clustering of single cells. C) Number of more variable or less variable genes in mmBACH1 cells (10% or more difference from control cells) observed by experimental observation compared to 1000 permutations. D) Changes in gene expression variability of genes that were significantly more or less variable in mmBACH1 cells.

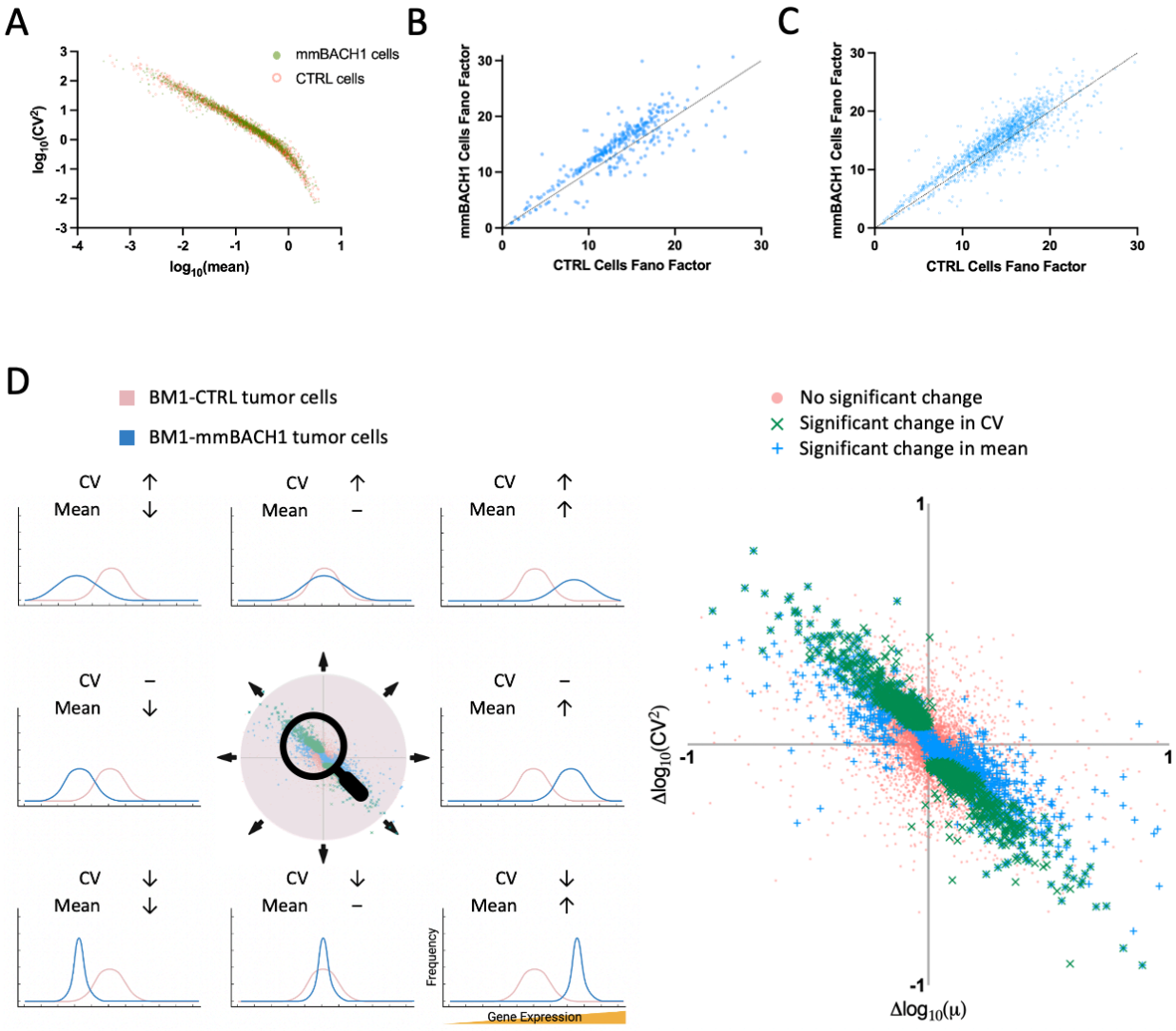


Figure 8. BACH1 overexpression causes changes in gene expression mean and variability. A) Expression $\log_{10}(CV^2)$ versus $\log_{10}$(mean) of genes in mmBACH1 and CTRL cells. Genes plotted are those with significant changes in CV. B) Fano factor (FF) of genes in mmBACH1 cells versus CTRL cells. Genes plotted are those with significant changes in CV with the same mean. C) Fano factor (FF) of genes in mmBACH1 cells versus CTRL cells. Genes plotted are those with significant changes in CV. D) Left: genes can change in CV and/or mean from CTRL to mmBACH1. Right: enlarged scatter plot of changes in $\log_{10}(CV^2)$ and $\log_{10}$(mean) for all genes.

Change in CV	Change in mean	Number of Genes	% of all changed genes
↑	↑	0	0.00%
	–	269	8.78%
	↓	614	20.04%
↓	↑	454	14.82%
	–	81	2.64%
	↓	0	0.00%
–	↑	1282	41.84%
	–	12865	Unchanged
	↓	364	11.88%

Table 2. Genes with significant changes in expression mean or CV in mmBACH1 cells versus CTRL cells.

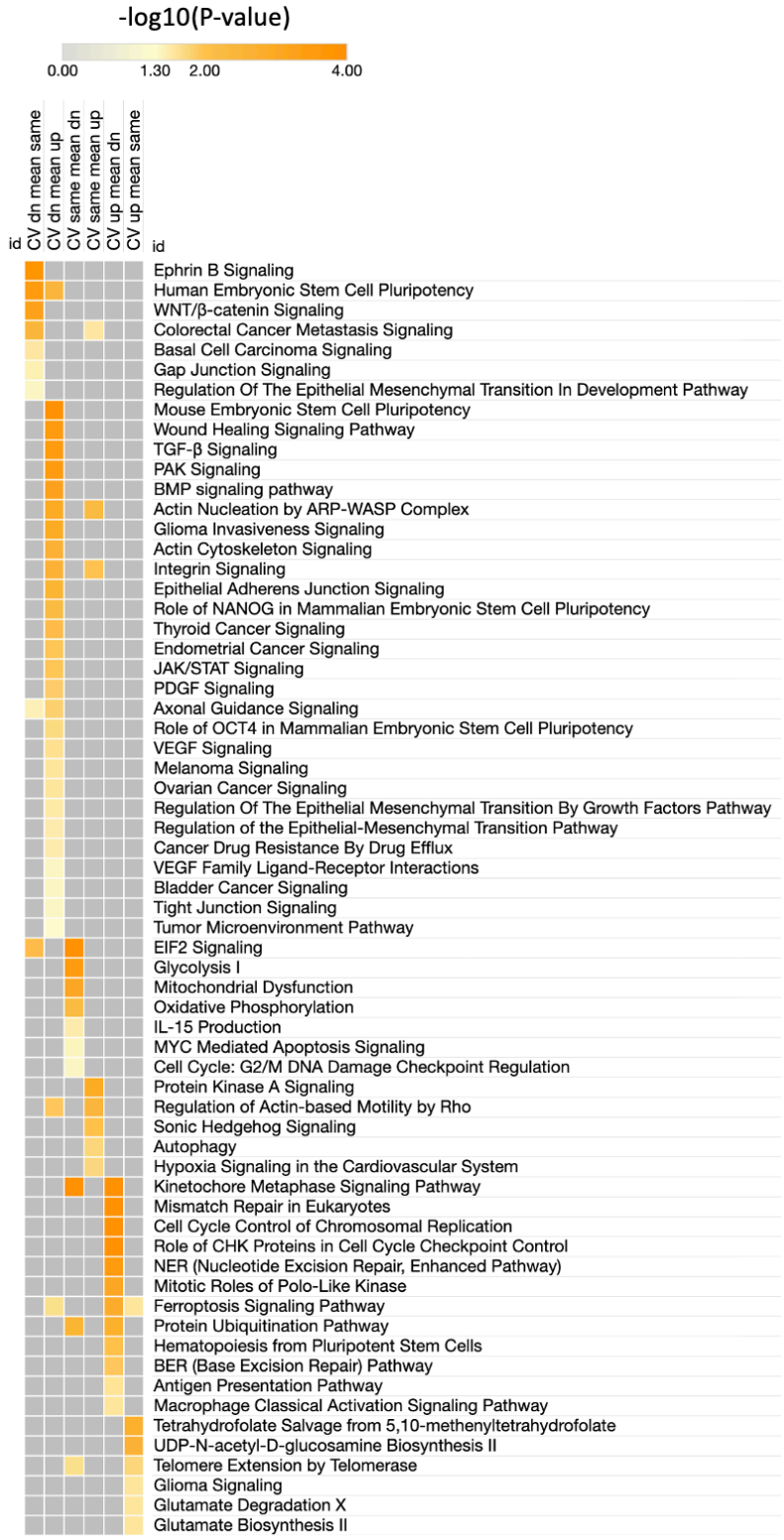


Figure 9. Genes with CV changes are involved in pathways related to metastasis. Canonical pathway analysis of genes with changes in mean and/or CV in mmBACH1 versus CTRL cells.

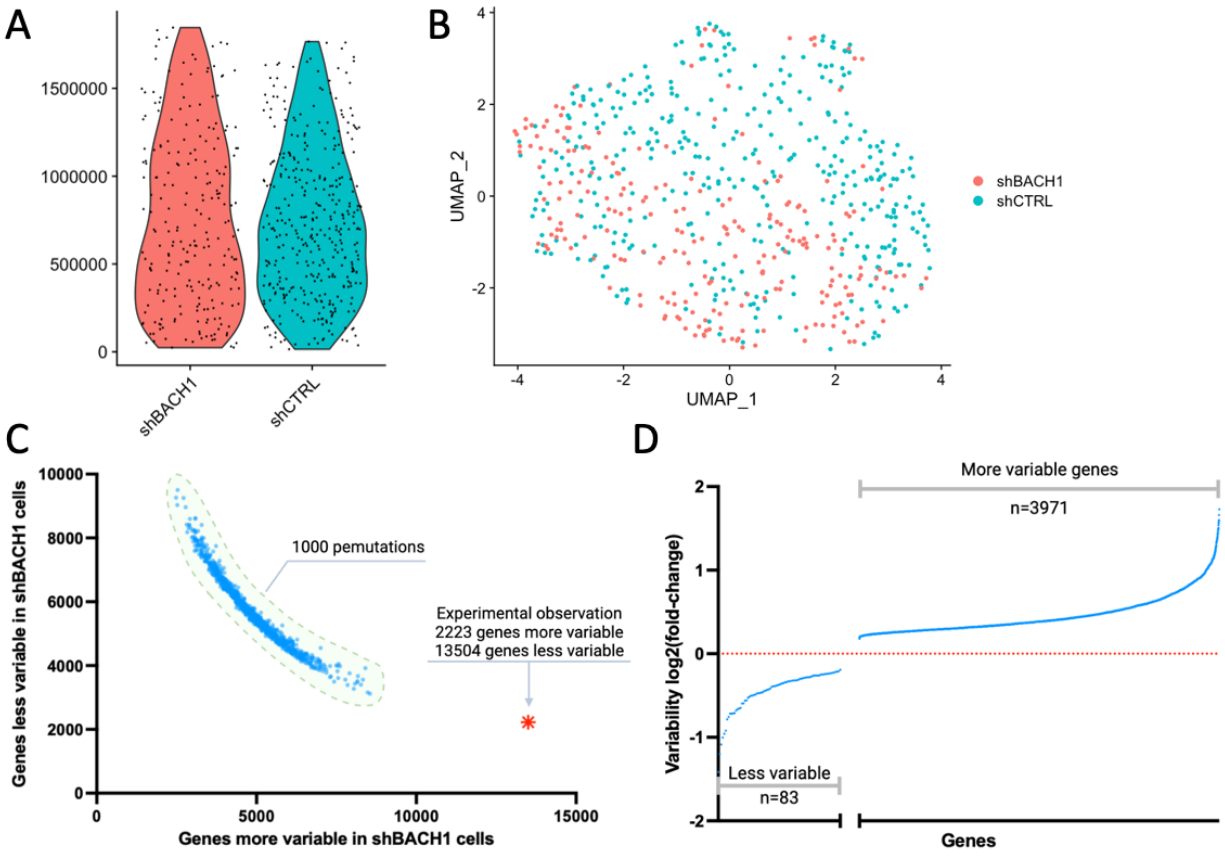


Figure 10. BACH1 knockdown enhances expression noise. A) Number of genes mapped in each shCTRL and shBACH1 tumor cell. B) UMAP clustering of single cells. C) Number of more variable or less variable genes in shBACH1 cells (10% or more difference from control cells) observed by Experimental observation compared to 1000 permutations. D) Changes in gene expression variability of genes that were significantly more or less variable in shBACH1 cells.

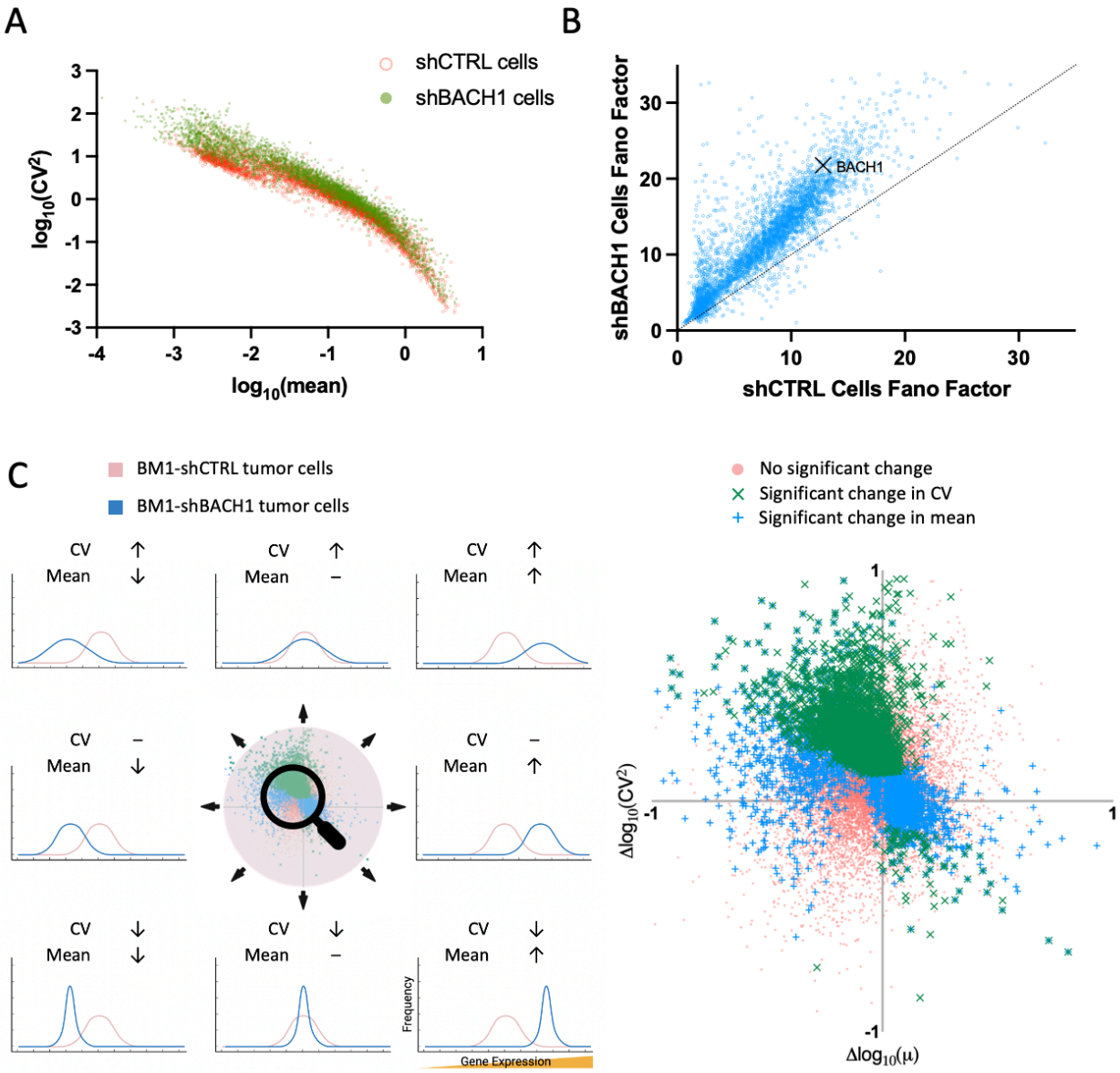
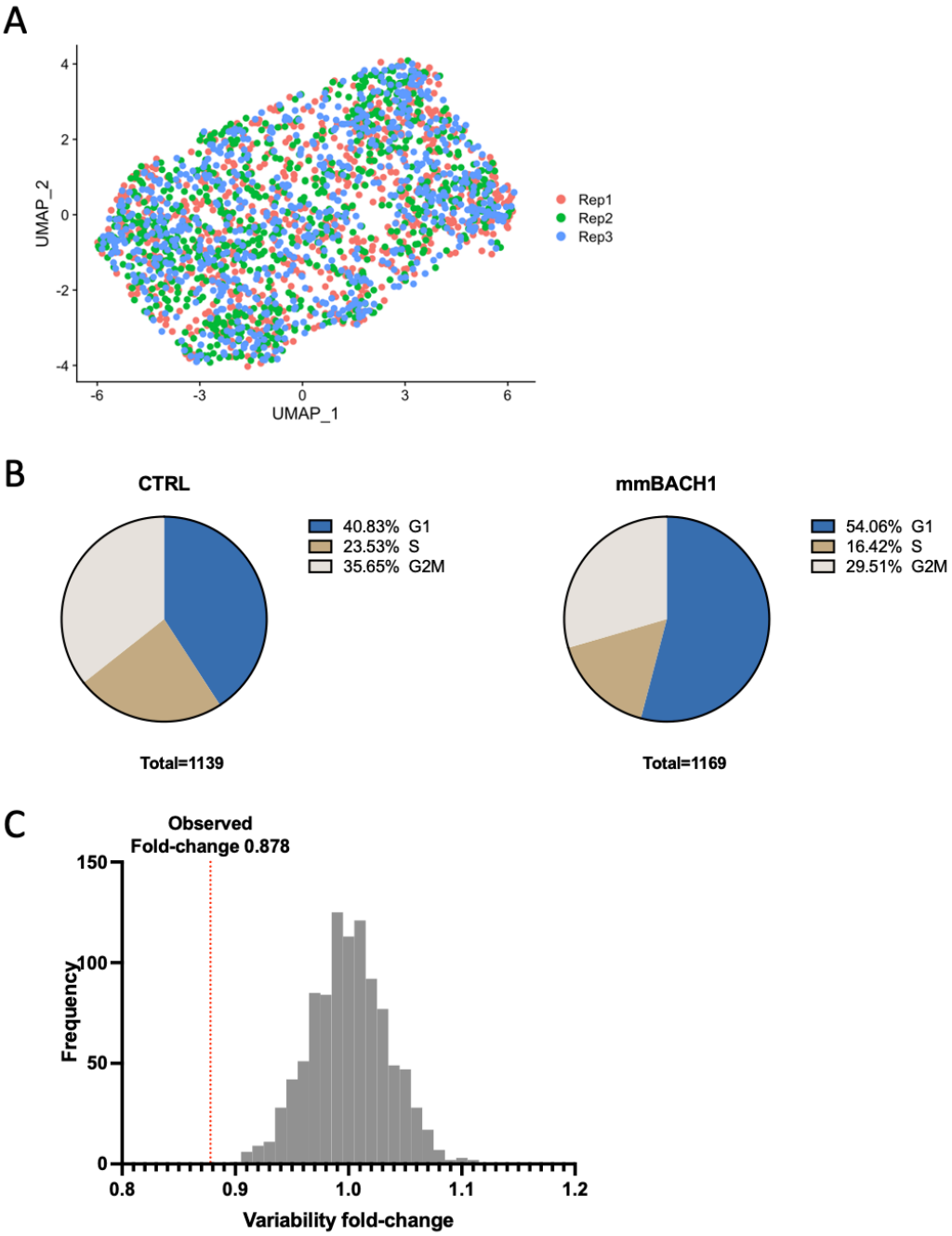


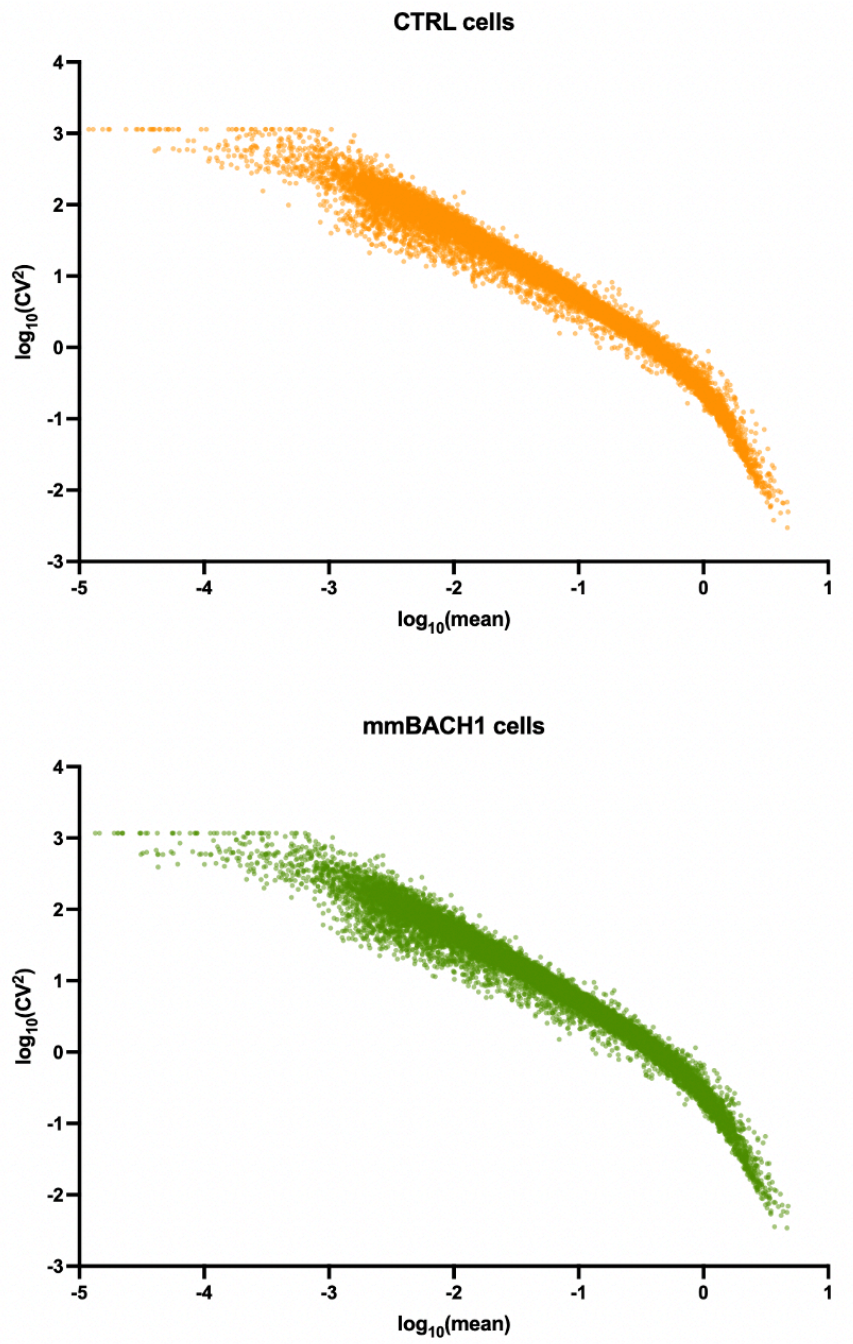
Figure 11. BM1 shBACH1 tumor cells exhibit predominantly increased variability without changes in expression mean. A) Expression $\log_{10}(CV^2)$ versus $\log_{10}$(mean) of genes in shCTRL and shBACH1 cells. Genes plotted are those with significant changes in CV. B) Fano factor (FF) of genes in shCTRL cells versus shBACH1 cells. Genes plotted are those with significant changes in CV. The Fano factor of BACH1 in shCTRL vs. shBACH1 cells is indicated. C) Left: genes can change in CV and/or mean from shCTRL to shBACH1. Right: enlarged scatter plot of changes in $\log_{10}(CV^2)$ and $\log_{10}$(mean) for all genes.

Change in CV	Change in mean	Number of Genes	% of all changed genes
↑	↑	45	0.64%
	–	2976	42.12%
	↓	950	13.44%
↓	↑	66	0.93%
	–	16	0.23%
	↓	1	0.01%
–	↑	2113	29.90%
	–	16792	Unchanged
	↓	899	12.72%

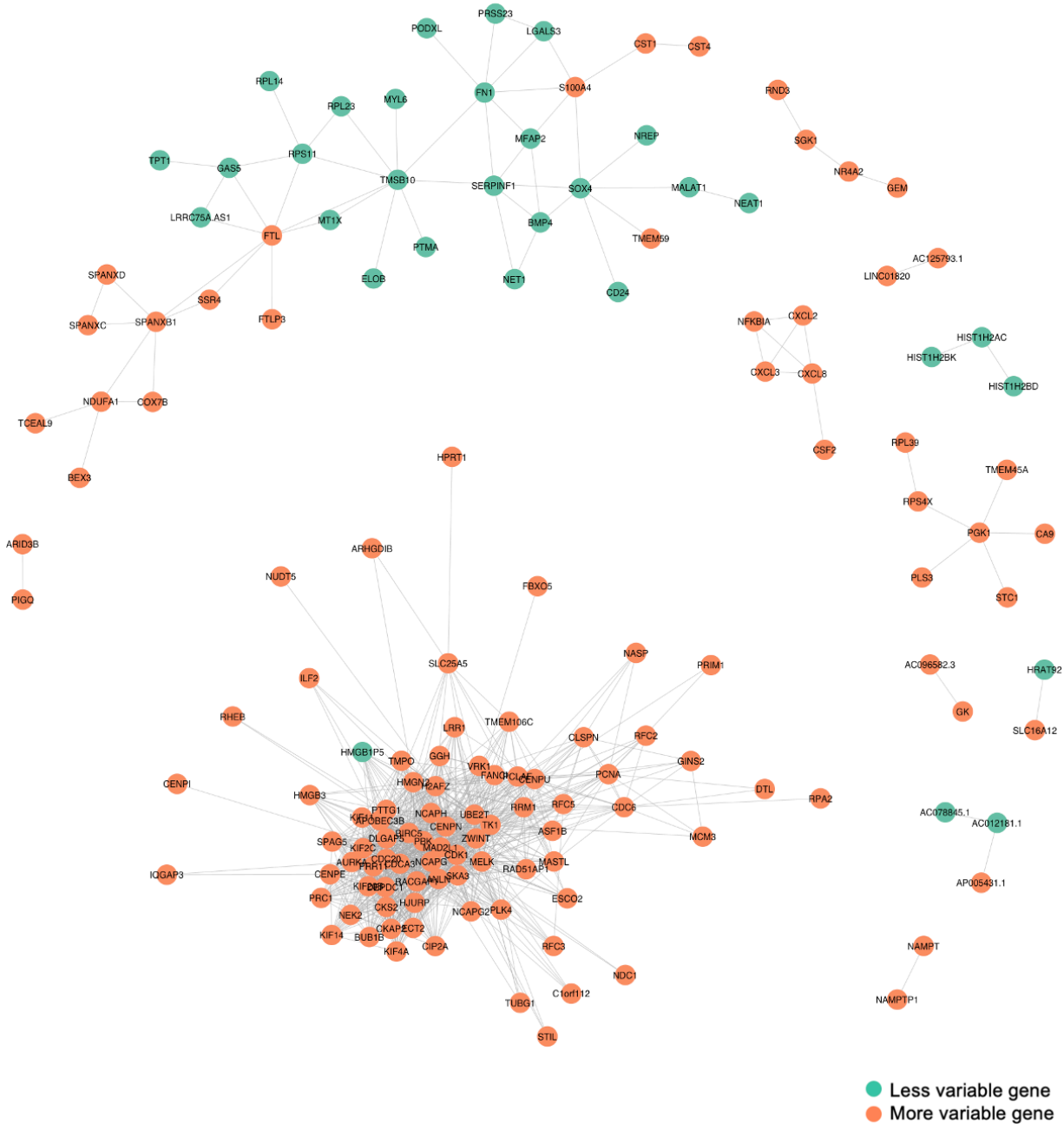
Table 3. Genes with significant changes in expression mean or CV in shBACH1 cells versus shCTRL cells.



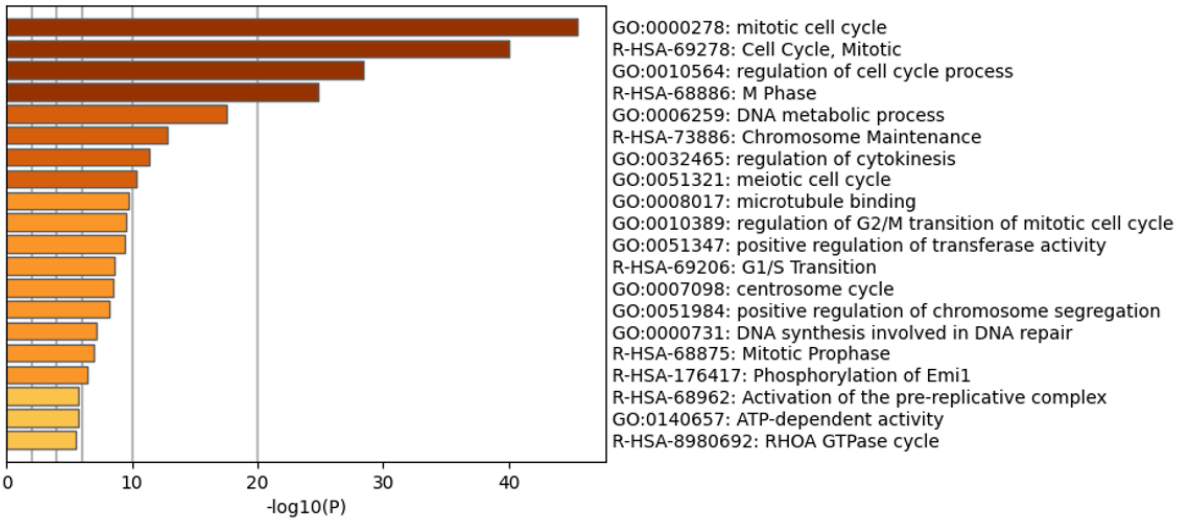
Supplementary Figure 14. Results from BM1-CTRL and BM1-mmBACH1 scRNA-seq. A) UMAP of three replicates for BM1-mmBACH1 and BM1-CTRL single cells. B) Cell cycle distribution of BM1-CTRL and BM1-mmBACH1 single cells. C) Observed change in BACH1 expression variability versus 1000 permutations



Supplementary Figure 15. Gene expression mean and CV of all genes in BM1-CTRL and BM-mmBACH1 cells. Gene expression $\log_{10}(CV^2)$ and $\log_{10}(\text{mean})$ were plotted for all genes in BM1-CTRL cells (top) and BM1-mmBACH1 cells (bottom).

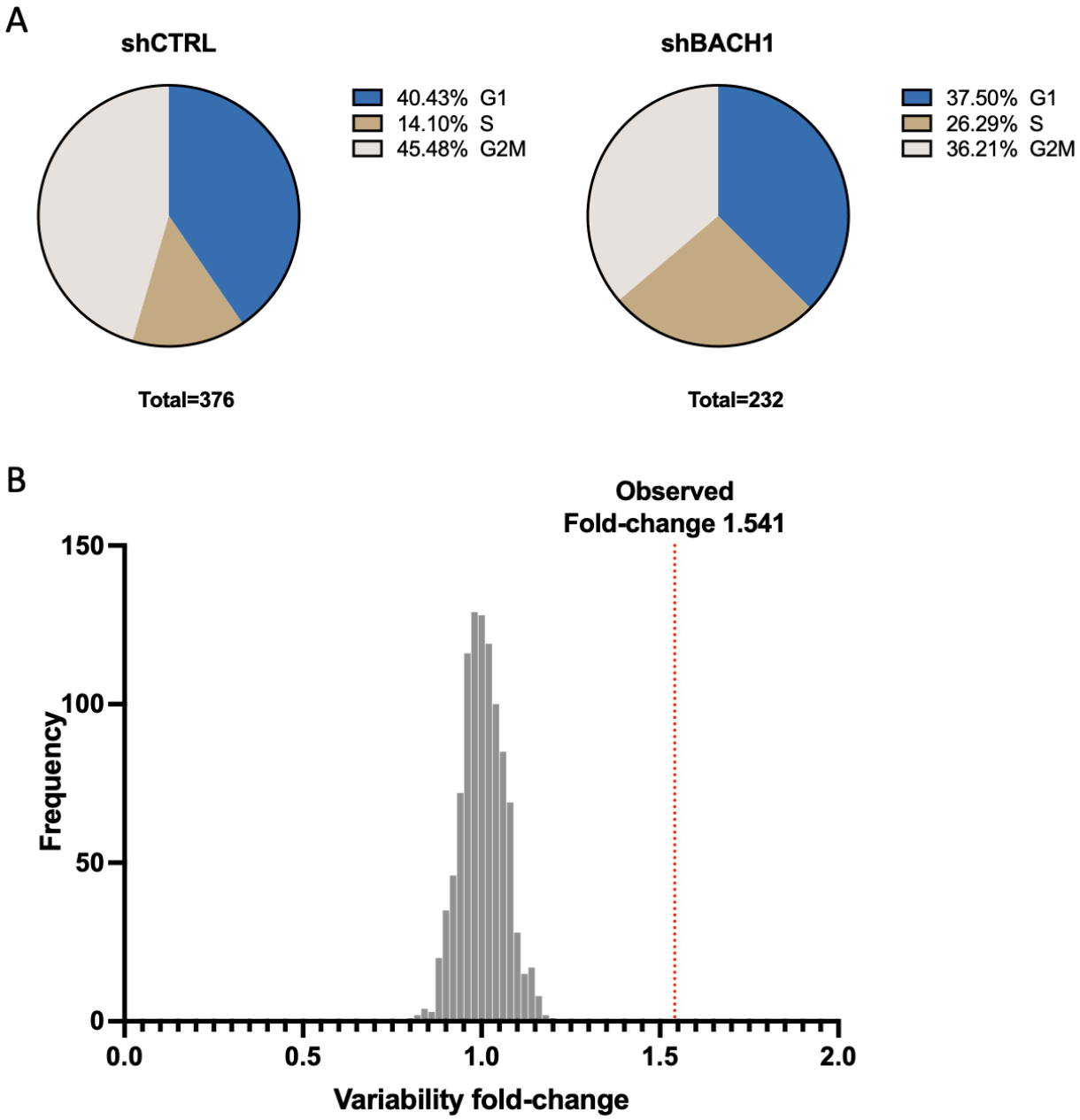


Supplementary Figure 16. Co-expression gene network within mmBACH1 cells.

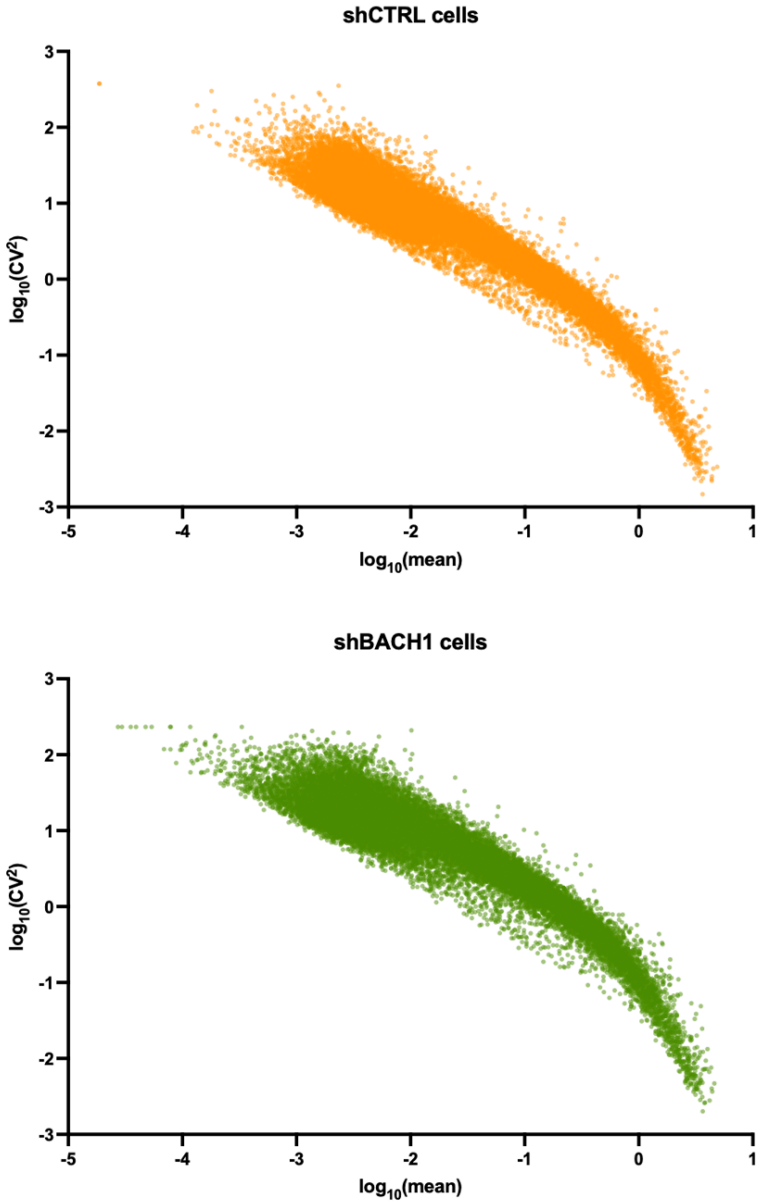


Supplementary Figure 17. Gene sets enriched from the largest sub-network in Fig S18.

Supplementary Figure 18. Significantly enriched gene sets from genes with changes in mean or CV in mmBACH1 versus CTRL cells. A) Enriched gene sets from all genes up or down-regulated in mmBACH1 cells. B) Enriched gene sets based on Metascape analysis of all genes with higher or lower CV in mmBACH1 cells.



Supplementary Figure 19. Results from BM1-shCTRL and BM1-shBACH1 scRNA-seq. A) Cell cycle distribution of BM1-shCTRL and BM1-shBACH1 cells. B) Observed changes in BACH1 expression variability versus 1000 permutations.



Supplementary Figure 20. Gene expression mean and CV of all genes in BM1-shCTRL and BM1-shBACH1 cells. Gene expression $\log_{10}(CV^2)$ and $\log_{10}(\text{mean})$ were plotted for all genes in BM1-shCTRL cells (top) and BM1-shBACH1 cells (bottom).

CHAPTER 5. SUMMARY AND DISCUSSION

Treating cancer is difficult because cancer is a heterogeneous disease. Even within one tumor of one patient, every cancer cell is not the same. Because of this heterogeneity, not all cancer cells respond the same way to treatments such as chemotherapy or radiation therapy. Some cancer cells will react and die after treatment, but there will be other resistant cells that either pre-exist in the population or are induced. These resistant cells will cause a relapse of disease after treatment, leading to tumor metastasis and, ultimately, death of the patient. If we can prevent metastasis by reprogramming tumor cells, then more patients can be saved. Previous studies have shown that cancer cell heterogeneity is linked to metastasis. This led us to propose that regulators of metastasis may modulate tumor heterogeneity and thereby contribute to the metastatic outcome.

In this study, we examined the relationship between gene expression variability and metastatic states of tumor cells. We used known regulators of metastasis to switch tumor cells between metastatic states, and we specifically focused on TNBC, a type of breast cancer that tends to be more metastatic than other breast cancers and lacks effective therapy. We used CV to measure gene transcriptional variability from single-cell RNA-seq data and identified lists of genes with statistically significant changes in variability through permutation analysis. Using RKIP, a metastasis suppressor, we showed that transcriptional variability was reduced in non-metastatic tumor cells with RKIP overexpression. In contrast, when BACH1, a pro-metastatic transcription factor, was overexpressed, overall transcriptional variability in tumor cells increased. These results demonstrate that metastasis suppressors and promoters can regulate

transcriptional heterogeneity in tumor cells and support a general association between overall increased heterogeneity and metastasis.

However, we also showed that expression variability alone could not drive the metastatic progression of tumor cells. We used shRNA to knock down BACH1 in tumor cells, switching them to a non-metastatic state similar to the phenotype of cells with overexpressed RKIP. In contrast to the effect of RKIP overexpression, the resulting BACH1-low tumor cells exhibited dramatically higher variability than the control cells. In this case, we used non-clonal shBACH1-expressing cells, suggesting that the shRNA acted as a noise generator leading to different levels of depleted BACH1 across the cell population. Although this ultimately increased variability in gene expression, the decrease in mean BACH1 expression across the population still led to a non-metastatic outcome. This result differs from a recent study where drugs that increased transcriptional variability but not did not alter the mean led to a change in cell state (*68*).

We showed that combining expression mean and CV changes results in a superior analysis of metastasis regulators compared to using mean changes alone. We found that genes up-regulated and less variable in RKIP-overexpressing tumor cells are involved in oxidative phosphorylation, suggesting that these tumor cells are more homogeneous inside the OXPHOS state and could be more sensitive to tumor cell killing by OXPHOS inhibitors such as metformin which we demonstrated in mouse studies. When we analyzed the same sequencing data using mean changes alone, we did not detect OXPHOS or other clinically targetable pathways such as DNA damage response and cytoskeleton. Analysis of genes that are less variable but with an unchanged mean in RKIP-overexpressing cells led to the identification of other potential metastatic regulators,

such as RPL15. Previous studies have shown that RPL15 is a positive regulator of metastasis and suggest that its expression variability contributes to metastatic progression (*93*). Our co-expression network analysis showed that RPL15's expression was positively correlated with several genes that were also less variable, suggesting the existence of a pro-metastatic co-regulatory network (see Fig. 4A). These results indicate that the use of the CV metric acts as an additional filter to reveal genes of interest that would be missed by analysis of mean alone as well as genes with an increased mean but unchanged CV that would likely be less effective therapeutic targets than those with increased mean and decreased CV.

In BACH1-overexpressing tumor cells, we found that genes up-regulated and less variable are enriched in pathways related to EMT and migration, agreeing with the current understanding that BACH1 promotes EMT and drives metastatic progression (*23, 39*). Once again, when we analyzed genes with mean changes without considering CV, EMT was not among the most enriched pathways and could have been easily missed. Given these findings, we propose that analyzing genes stratified by expression mean and CV can provide insights into how metastasis regulators work and reveal promising therapeutic targets.

Analysis of genes becoming less variable upon loss of a metastatic phenotype could also identify novel regulators of metastasis. As an example, we identified KMT5C, an epigenetic regulator with repressive functions, as a novel metastasis suppressor in TNBC. KMT5C was significantly up-regulated and less variable in RKIP-overexpressing cells. We created BM1 and 4T1 cells with stable overexpression of KMT5C and observed

inhibition of colonization and metastasis in mice. Moreover, analysis of patient data indicated that KMT5C correlates with better patient survival and less metastasis in TNBC.

Our results revealed new pathways and functions that appear to be controlled by these metastatic regulators. For example, in BACH1-overexpressing cells, the ephrin B signaling pathway was highly enriched among the genes with reduced variability. Genes involved in the ephrin signaling pathways regulate several processes, including adhesion, motility, and differentiation (*119*). Additionally, pathways involved in stem cell pluripotency were also enriched among the less variable genes. Previous studies have theorized that BACH1 may promote cancer cell stemness, but the evidence is lacking (*43, 44*). By contrast, we did not discern EphB signaling nor observe changes in stem cell pathways when we conducted bulk RNA-seq and differential expression analyses on BACH1-overexpressing tumors. The ephrin and stem cell pathways are two examples demonstrating the utility of analyzing expression mean and CV together, as they were not easily detected by studying mean changes alone. More importantly, differential expression analysis would have missed genes involved in these pathways that did not have changes in the mean. We hope to gain more insights by studying the less variable genes involved in these pathways, which could be mediators of BACH1's function as a metastasis promoter.

We observed that the slope of gene expression $\log_{10}(CV2)$ versus $\log_{10}(mean)$ approaches -1 for all cell types studied, suggesting that the gene transcription in these tumor cells was mainly either a Poisson process or transcriptional bursting. This result agrees with a previous study that examined gene expression variability of a non-tumorigenic breast cell line (*120*). However, the expression of some

genes indicated deviations from Poisson or bursting processes (shown as a departure from the -1 slope), an effect that could be caused by other factors such as cell state, feedback, or noise propagation from upstream regulators (extrinsic). However, whether extrinsic or intrinsic factors contribute more significantly to overall expression variability could be context-dependent. Patange and colleagues showed that gene expression was amplified through increased bursting when the oncogenic transcription factor MYC was induced (*121*). Therefore, further studies are needed to fully elucidate how metastasis regulators affect gene expression variability.

We also acknowledge that there are limitations in our studies. We focused on transcriptional variability using scRNA-seq. However, variability in gene expression may not always translate to protein expression. This limitation could be addressed in the future should novel single-cell techniques be developed for comprehensive profiling of the proteome across many single cells. Lastly, our study focused on gene expression variability resulting from metastasis regulators in xenograft mice. Therefore, more studies are needed to elucidate the effect of the adaptive immune system in the tumor microenvironment on heterogeneity and metastatic progression (*122*).

CHAPTER 6. FUTURE WORK

Our approach to studying genes based on a combination of mean and CV changes enabled us to discover many novel genes and pathways with differences in expression variability across metastatic states. Our findings presented here only represent a small part of the overall discoveries, and many questions remain.

We identified genes and pathways up-regulated and less variable in non-metastatic RKIP-overexpressing cells. We predict that these pathways are potential therapeutic targets (see Fig. 3C), such as DNA damage response and cytoskeleton processes. Therefore, further studies are needed to determine whether the RKIP cells are more sensitive to treatments such as chemotherapy and radiation than the wild-type tumor cells.

Among the genes following the same pattern of changes, we showed that KMT5C has metastasis suppressor functions in TNBC. Given that KMT5C functions as a histone modifier, it is possible that RKIP, as a metastasis suppressor, also reduces variability at the epigenetic level, at least in part via KMT5C. Due to contradictory reports on KMT5C's regulation of metastasis (*106, 107, 112*), KMT5C may function differently depending on the tumor type, thus necessitating studies of KMT5C in other tumor cells. While KMT5C causes tri-methylation at H4K20, it is unknown whether it solely suppresses metastasis via this specific histone modification. Through separate ChIP-seq experiments of KMT5C and H4K20Me3 from the same samples, one can determine whether KMT5C binds to targets that do not have H4K20Me3. It is also possible that KMT5C does not bind to every one of the H4K20Me3 sites because a previous study reported lower levels of H4K20Me3 in senescent cells than in proliferating cells, but there was no difference in KMT5C

expression between the cell types (*111*). In this case, identifying KMT5C-specific sites may shed light on its downstream effectors that regulate metastasis.

Previous studies showed reduced KMT5C and H4K20Me3 in breast tumors versus normal tissues, but we do not know how KMT5C is suppressed in breast cancer patients (*111, 112*). KMT5C proteins may be actively degraded, or overly active demethylases may reverse the tri-methylation at H4K20Me3. Demethylases on H4K20Me3 were recently identified, and further investigation is required to determine whether inhibiting these methylases could lead to reduced metastasis (*123*).

While KMT5C reduced metastasis *in vivo*, we observed significantly reduced tumor growth in the syngeneic mice but not in the xenograft mice that lack an adaptive immune system. KMT5C's role in the immune system was not previously reported. Our analysis of patient data showed a negative correlation between KMT5C expression and genes involved in PD-L1 signaling, suggesting a potential mechanism for inhibiting T-cell exhaustion via down-regulation of the PD-L1/PD-1 pathway. If KMT5C plays a beneficial role in the adaptive immune system, then patients may respond better to checkpoint blockade immunotherapies by preventing the degradation of KMT5C or the demethylation of H4K20Me3.

We used BACH1 knockdown as an alternative means to suppress metastasis in tumor cells and proceeded to test expression variability differences. However, the use of shRNA introduced a noise generator into the system and caused a substantial increase in expression variability in the non-metastatic shBACH1 cells. To determine whether inhibition of metastasis promoters such as BACH1 leads to less variable gene expression,

we plan to eliminate the confounding effect of shRNA by completely knocking out BACH1 using CRISPR-Cas9.

Our results also unraveled many functional pathways that were changed by BACH1 overexpression. While changes in some pathways, such as EMT and OXPHOS, have been reported previously, we also observed changes in numerous novel pathways (See Fig. 9). As a transcription factor, BACH1 has thousands of targets, and the gene regulatory networks of BACH1 are not entirely elucidated (*124*). By analyzing gene expression mean and CV together, we provided additional evidence on pathways potentially regulated by BACH1. In particular, several pluripotency-related pathways were significantly enriched among the genes up-regulated and less variable upon BACH1 overexpression. Cancer cell stemness was previously proposed to be under BACH1 regulation, but few studies offered evidence in cancer (*43, 44*). Further studies of these genes implicated in stem cell pluripotency may shed light on BACH1's potential regulation of cancer stem cells. For example, SNAI2 and BMP4 were up-regulated and less variable in BACH1-overexpressing cells. SNAI2 has been shown to induce cancer cell stemness, while BMP4 supports the self-renewal process of embryonic stem cells (*43, 125*). BMP4 promotes self-renewal by inhibiting the p38 MAPK signaling pathway, so it is possible that treating BACH1-high tumor cells with a p38 inhibitor may inhibit both BACH1 and the recurrence of disease from metastatic cancer stem cells (*23, 125*).

We also showed that genes related to the ephrin signaling pathway were less variable in BACH1-overexpressing cells. While ephrin signaling is involved in many cellular processes, such as morphology and movement, our analyses identified specific genes involved in tumor progression that were less variable (*119*). For example, AXIN1

is a tumor suppressor in breast cancer (*126, 127*). A reduction in the CV in BACH1-high tumor cells could result in fewer tumor cells expressing high levels of AXIN1, increasing the overall potential of cancer progression. Another gene inside the ephrin pathway, GNAS, is known to promote a stem cell-like state in small-cell lung cancer via activating the PKA signaling pathway (*128*). Therefore, BACH1-high tumor cells with less variable GNAS expression may be sensitive to therapies targeting PKA.

Analyzing gene expression based on a reduction in CV can be used to study mechanisms of metastatic progression. We predict that genes with reduced CV are better indicators of metastasis regulation than genes with increased CV. As a validation, one approach is to use mathematical modeling of gene expression to determine if a reduction in expression CV could reveal genes as downstream regulators more controlled by the metastasis regulator's signaling cascade than those downstream that are likely to have more variable expression. On the other hand, the genes and pathways with increased CV cannot be ignored. Further investigations are required to determine the significance of more variable genes and pathways in non-metastatic tumor cells and their clinical relevance.

Lastly, we plan to use the combined mean and CV approach in future studies beyond single-cell RNA-seq to obtain more insights from large datasets, including single-cell ATAC-seq, flow cytometry, multiplexed immunofluorescence imaging, and spatial transcriptomics. Our approach will increase our understanding of previously less explored areas of heterogeneity, such as translation and epigenetics.

ADDENDUM

Supplementary note

Gene expression is often conceptualized as a Poisson process, for which the variance equals the mean:

$$\sigma^2 = \mu$$

This implies that the CV^2 and the Fano Factor *F* depend on the mRNA expression mean in the following ways:

$$CV^2 = \frac{\sigma^2}{\mu^2} = \frac{1}{\mu}$$

$$F = \frac{\sigma^2}{\mu} = 1$$

Thus, for a perfect Poisson process, the CV^2 decreases as the inverse of the mean, while the Fano Factor is a constant = 1. However, many studies, including Singh et al., 2010 and Dar et al., 2016 have indicated that eukaryotic gene expression can be non-Poisson, often due to bursts of gene expression (*94, 129*). Bursting modifies the above equations for mRNAs as follows.

$$CV^2 = \frac{\sigma^2}{\mu^2} = \frac{1+b}{\mu}$$

$$F = \frac{\sigma^2}{\mu} = 1+b$$

For mRNAs with zero-order production without bursts, at a constant rate *k* and first-order degradation at rate *g*, the mean and variance take the following values:

$$\sigma^2 = \mu = \frac{k}{g}$$

Therefore, the Fano Factor is =1 and the CV^2 for such a process will be:

$$CV^2 = \frac{\sigma^2}{\mu^2} = \frac{g}{k}.$$

With bursting, these formulas become

$$CV^2 = \frac{\sigma^2}{\mu^2} = \frac{(1+b)g}{k}$$

$$F = \frac{\sigma^2}{\mu} = 1 + b$$

Typically, gene regulation affects the mRNA synthesis rate k.

Assume that the synthesis rate of some mRNA increases by a factor κ due to RKIP overexpression.

The expected new mean, CV^2 and Fano Factor will be:

$$\sigma'^2 = \mu' = \frac{k\kappa}{g}$$

$$CV'^2 = \frac{\sigma'^2}{\mu'^2} = \frac{g}{k\kappa}$$

$$F' = \frac{\sigma'^2}{\mu'} = 1$$

The corresponding ratios will be:

$$\frac{\mu'}{\mu} = \frac{k\kappa}{k} = \kappa$$

$$\frac{CV'^2}{CV^2} = \frac{\frac{g}{k\kappa}}{\frac{g}{k}} = \frac{1}{\kappa}$$

These ratios turn into differences of the corresponding log-ratios after taking the logarithm:

$$\log\left(\frac{\mu'}{\mu}\right) = \log(\mu') - \log(\mu) = \log(\kappa)$$

$$\log\left(\frac{CV'^2}{CV^2}\right) = \log(CV'^2) - \log(CV^2) = \log\left(\frac{1}{\kappa}\right) = -\log(\kappa)$$

Therefore, plotting $\Delta\log(CV^2)$ versus $\Delta\log(\mu)$ should result in a set of points that distribute along a line with slope = -1.

Supplementary data tables

Supplementary Table 1. Genes with significantly less CV and up-regulated in BM1-RKIP cells. Expression CV and mean log_2(fold-change) values of each gene were shown. Fold-change was determined by RKIP versus control cells. The genes with well-studied mechanisms were categorized as either "pro-cancer" or "anti-cancer." The genes with context-dependent regulations in cancer were categorized as "regulated." The genes with insufficient evidence of cancer regulation were categorized as "novel." Relevant literature reports for each gene were cited by PubMed ID, with "N/A" indicating a lack of cancer-related literature evidence.

Gene	CV change	Mean change	Type	PubMed ID
PEBP1	-2.02	5.95	anti-cancer	25597354.
OR7C1	-1.53	1.34	novel	26861454.
FZD5	-0.97	2.06	pro-cancer	27869803, 33311446, 29930766.
MAGEA11	-0.92	2.83	novel	23839233, 22198676, 29423052, 35022469.
RIMS1	-0.88	2.45	novel	N/A
KMT5C	-0.83	2.16	novel	35404406, 29229751, 30792382.
CCND2	-0.83	2.93	regulated	30308939, 11289162, 31511084, 10547574, 10666388.
AC093484.4	-0.82	2.59	novel	N/A
PTPRN2	-0.81	1.34	pro-cancer	26620550, 25877877.
POLD2P1	-0.80	0.43	novel	N/A
CCDC157	-0.78	3.31	novel	30741974.
FAM133A	-0.78	1.68	novel	29885519, 21791470.
PCDHGB1	-0.77	2.88	novel	N/A
NR4A3	-0.73	1.30	anti-cancer	30455429, 27528092, 28249906, 33634114.
RPS6KL1	-0.72	2.40	novel	31388935.
HMGB1P21	-0.71	2.54	novel	N/A
PPP1R3C	-0.70	2.16	novel	32915648, 25846879, 18803327.
AC025031.3	-0.68	2.02	novel	N/A
FGF18	-0.67	2.25	pro-cancer	14559787, 24018557, 30196303, 30082912, 33117668, 32801905, 21319186, 15814652.

Supplementary Table 1 Continued

Gene	CV change	Mean change	Type	PubMed ID
ZNF415	-0.62	1.34	novel	17055453.
FOXP2	-0.58	0.67	regulated	25515522, 27314063, 29725501.
AC113404.3	-0.56	0.43	novel	N/A
MX2	-0.56	1.87	novel	31660681, 33734579.
PSTPIP2	-0.56	0.72	novel	N/A
RHOBTB2	-0.55	1.39	anti-cancer	20930524, 21801820, 19937980, 27941885.
AMFR	-0.53	0.77	pro-cancer	16204225, 18279062, 24568530.
HHIP	-0.52	0.82	novel	29539620, 31765425, 33415068.
CALB2	-0.51	1.10	novel	29699512, 21629658.
AC096564.1	-0.51	1.30	novel	N/A
SERPINB9	-0.50	0.96	pro-cancer	33242418, 30127393.
TCTEX1D2	-0.47	0.48	novel	N/A
AADACP1	-0.47	2.93	novel	N/A
INO80	-0.44	0.43	pro-cancer	27641337, 27750218, 29383140.
EDN1	-0.44	0.91	pro-cancer	23884378.
COPE	-0.43	0.77	novel	N/A
WBP11P1	-0.43	1.30	novel	N/A
DPH6	-0.42	0.96	novel	26261303.
MRPL4	-0.41	0.67	novel	34359804.
ZSWIM6	-0.41	0.43	novel	N/A
CELF2	-0.41	0.43	anti-cancer	23795348, 31409895.
NMB	-0.39	0.53	novel	15528253, 1563005, 20388507.
DUSP14	-0.38	0.38	novel	32256117.
TEX2	-0.38	0.62	novel	N/A
ARHGAP12	-0.37	0.29	novel	18504429, 31209295.
TXNRD2	-0.37	0.62	novel	25647640, 32671444.
SKIL	-0.37	0.48	regulated	20495575, 23764425, 25749039.
CXorf56	-0.37	0.48	novel	N/A
ATF3	-0.37	0.67	regulated	34551291, 34213887, 33608016, 32183950, 34098867, 32922364.

Supplementary Table 1 Continued

Gene	CV change	Mean change	Type	PubMed ID
TIAM1	-0.37	0.82	pro-cancer	30890693, 32671511, 29452216, 26821678.
FOSB	-0.36	0.29	regulated	27604827, 32477007, 30658436, 27248170.
CLU	-0.36	0.48	pro-cancer	20145158, 12374699, 33385484, 19879423, 19879420.
HMGCS1	-0.36	0.62	novel	27562463, 32349352, 32491253, 32692762.
HLA-DRA	-0.36	0.86	novel	33339725, 33457236, 25156210, 35543859.
RALGAPA1	-0.36	0.43	novel	N/A
PLK2	-0.35	0.53	anti-cancer	22289679, 21402713.
AC090409.1	-0.35	0.53	novel	N/A
MAN2B2	-0.35	0.62	novel	32231737.
SGMS2	-0.35	0.48	novel	30770781.
PPP2R3C	-0.33	0.48	novel	23087057.
MEPCE	-0.33	0.53	novel	32048991.
SH3BGRL3	-0.33	0.43	novel	33839406, 26286913, 34380438.
TM4SF1	-0.33	0.29	pro-cancer	19351819, 33153498, 30876464, 27459514, 28368050, 19351819, 26089370.
AC006970.1	-0.33	1.10	novel	N/A
HSPA4L	-0.32	0.43	novel	34021502, 26486082.
C1GALT1C1	-0.32	0.43	novel	31471227.
TIPARP	-0.32	0.43	anti-cancer	31695491, 32482854.
GATAD1	-0.32	0.53	novel	31286678, 29266303.
GGA2	-0.32	0.53	pro-cancer	31076515, 34799560, 30578931, 29358589.
RBM18	-0.31	0.62	novel	N/A
GLIPR1	-0.31	0.53	anti-cancer	22025562, 21933889.
HSD17B11	-0.31	0.34	novel	21549806.
ZBTB21	-0.31	0.48	novel	N/A

Supplementary Table 1 Continued

Gene	CV change	Mean change	Type	PubMed ID
ABL2	-0.31	0.48	pro-cancer	22777352, 29767389, 31825826, 33318173.
GSTP1	-0.30	0.43	regulated	27185638, 32319549, 29116019.
ZBTB43	-0.30	0.43	novel	35488273.
COX7B2	-0.30	0.29	novel	15623157.
GADD45B	-0.30	0.43	novel	34490266, 32873775, 29356020.
SMAD7	-0.30	0.62	regulated	10498890, 24556688, 18781153, 24317436, 17332363.
NR4A2	-0.30	0.53	pro-cancer	23821160, 16293616, 27471554, 15548686.
GEM	-0.29	0.38	novel	25155751, 7912851, 11423971.
CYP4V2	-0.29	0.53	novel	20716162, 15897573, 30139042.
ATRAID	-0.29	0.24	novel	32434850.
REEP3	-0.28	0.29	novel	33179443.
ABR	-0.28	0.34	novel	10360661, 7536455, 7829075.
FILIP1L	-0.28	0.34	anti-cancer	24327474, 18794120, 21693594.
CD74	-0.28	0.58	novel	17142775, 22178447, 21208136.
IGFBP7	-0.28	0.29	regulated	21997538, 24632618, 19861408.
CNTRL	-0.27	0.67	novel	31250523, 23777766.
AL450405.1	-0.27	0.58	novel	N/A
DLGAP1-AS1	-0.27	0.43	novel	31949128, 33816780, 34536977.
DCBLD2	-0.27	0.29	regulated	34095137, 33808696, 18314483, 33834039, 25061874.
GOLM1	-0.26	0.38	pro-cancer	27569582, 29181846, 34729117, 34795203.
KLF4	-0.26	0.38	regulated	16372018, 21242971, 19503094, 16357509, 29930276, 20215880.
SEC22B	-0.26	0.38	novel	28663435.
UFSP2	-0.26	0.34	anti-cancer	25219498, 34923774.
CLGN	-0.26	0.29	novel	N/A
REST	-0.26	0.29	regulated	23414932, 20548947, 19246391.
ODC1	-0.25	0.24	pro-cancer	19047152, 34376808.

Supplementary Table 1 Continued

Gene	CV change	Mean change	Type	PubMed ID
OGDH	-0.25	0.43	novel	27732861, 27027236, 31247190.
CNOT2	-0.25	0.29	novel	21299754, 34680125, 29024811.
PAK2	-0.25	0.29	novel	19876919, 30243723, 30068946.
SEPTIN11	-0.25	0.38	novel	21714766.
NRAS	-0.25	0.29	regulated	31553708, 30232163, 34098219, 30712867.
MSMO1	-0.25	0.29	novel	27246191, 34759952.
DKK1	-0.24	0.34	regulated	28892080, 19148141, 20957670.
HBEGF	-0.24	0.43	pro-cancer	15289334, 24013225, 22009535.
NDUFS8	-0.24	0.29	novel	27516145, 23788038.
MOK	-0.24	0.29	novel	15327990.
RAB31	-0.24	0.48	regulated	25472813, 22920728, 26245486.
TPI1P1	-0.24	0.58	novel	N/A
CDKN1A	-0.24	0.34	regulated	19440234, 33547489, 32609833, 28279624, 17143283.
MCFD2	-0.24	0.34	novel	29679592.
RBBP6	-0.24	0.38	novel	22139301, 33149679, 31274515.
DGCR6L	-0.24	0.34	novel	19778628.
HSDL2	-0.24	0.34	novel	27658777, 33738911, 31217732, 31372054.
SFTA1P	-0.23	0.24	novel	34845189, 32825934.
AXL	-0.23	0.29	pro-cancer	19633687, 22751098, 20858715, 20080645.
RGS3	-0.23	0.38	anti-cancer	34618566.
ARHGDIB	-0.23	0.38	anti-cancer	34006303, 20001211, 22718398.
IER3	-0.23	0.29	regulated	25250570, 31832020, 26973248.
AP3S1	-0.23	0.29	novel	17125464.
PDZD11	-0.23	0.24	novel	N/A
TMCC3	-0.23	0.34	regulated	33742122.
ERLIN1	-0.23	0.29	novel	29596435.
STXBP3	-0.22	0.34	novel	N/A
NPLOC4	-0.22	0.29	novel	33328215.

Supplementary Table 1 Continued

Gene	CV change	Mean change	Type	PubMed ID
HIGD1A	-0.22	0.29	regulated	27308509, 25683712, 32826866, 35152263.
TRIOBP	-0.22	0.29	novel	33121024, 30195749.
AKR1C3	-0.22	0.38	pro-cancer	31308078, 32847363, 31253396.
TFPI	-0.22	0.24	novel	28734156, 32891505, 35567715.
BCAT2	-0.22	0.38	pro-cancer	32467562, 33097833, 32029896.
MAD2L2	-0.22	0.34	novel	29360267, 34083040, 32004442.
WDR1	-0.22	0.34	novel	29989053, 32601462, 27521604.
TFDP2	-0.22	0.38	novel	N/A
CD55	-0.22	0.48	pro-cancer	16485129, 16533428, 33177111, 29895866.
SUCLG2	-0.21	0.34	novel	32963351.
CNPY3	-0.21	0.38	novel	N/A
NCEH1	-0.21	0.24	novel	32415874.
MLPH	-0.21	0.38	novel	34829479, 31659808.
KLF10	-0.21	0.38	anti-cancer	29799499, 28249899, 34635142, 32549754.
INIP	-0.21	0.43	novel	27090575.
CCN2	-0.21	0.34	pro-cancer	33358571, 32791309, 27613407, 24832082, 21455569.
CD59	-0.21	0.19	pro-cancer	22000275, 9121154, 30391994, 31685825.
SMU1	-0.20	0.34	novel	33618520.
DDX6	-0.20	0.34	pro-cancer	29314290, 29987267, 31588046.
PSMB7	-0.20	0.34	novel	20010949, 32561655.
EZR	-0.20	0.24	novel	31775562, 34627255, 31452341.
WEE1	-0.20	0.24	pro-cancer	32958072, 27427153.
RAI14	-0.20	0.24	novel	33176601, 31772666, 33840258.
COPS6	-0.20	0.43	novel	32016946, 32225170, 30961513.
CDC42BPA	-0.20	0.29	pro-cancer	29072916, 33558670.
POLE3	-0.19	0.29	novel	N/A
POLR2L	-0.19	0.29	novel	N/A

Supplementary Table 1 Continued

Gene	CV change	Mean change	Type	PubMed ID
DPM2	-0.19	0.29	novel	26683373.
IFI27L2	-0.18	0.19	novel	32832569.
NDUFS6	-0.18	0.24	novel	25470292.
RHOC	-0.18	0.29	pro-cancer	34627249, 34285210, 23488932, 31340863.
SGK1	-0.18	0.38	pro-cancer	33542904, 34669124, 29017179.
FAM114A1	-0.18	0.29	novel	N/A
NSD2	-0.18	0.43	pro-cancer	32021450, 34555356.
PDIA3	-0.18	0.24	novel	33761087, 33761087, 26125904.
EIF5	-0.17	0.19	novel	34403101, 27986751, 25979826.
ETFB	-0.17	0.29	novel	28913729.
CD44	-0.17	0.19	pro-cancer	32728200, 34944493, 33961871.
SRA1	-0.17	0.29	novel	17710122, 30545127.
JMJD1C	-0.17	0.29	novel	34475205.
GNG12	-0.16	0.24	pro-cancer	31898405, 32384699.
RAB7A	-0.16	0.29	regulated	34066419, 30626032, 31374919.

Supplementary Table 2. Genes with significantly less CV without changes in mean in BM1-RKIP cells. Expression CV log_2(fold-change) values of each gene were shown. Fold-change was determined by RKIP versus control cells. The genes with well-studied mechanisms were categorized as either "pro-cancer" or "anti-cancer." The genes with context-dependent regulations in cancer were categorized as "regulated." The genes with insufficient evidence of cancer regulation were categorized as "novel." Relevant literature reports for each gene were cited by PubMed ID, with "N/A" indicating a lack of cancer-related literature evidence.

Gene	CV change	Type	PubMed ID
MARK1	-1.00	anti-cancer	25042806, 31281935, 29076440.
ROGDI	-0.99	novel	27636029.
HLA-DRB6	-0.95	novel	N/A
TMEM54	-0.93	novel	N/A
SERINC5	-0.90	novel	N/A
HLA-F	-0.87	novel	26332651.
AC046143.1	-0.82	novel	N/A
RPL7P26	-0.75	novel	N/A
TBC1D10A	-0.72	novel	31527750.
AC095055.1	-0.72	novel	N/A
AC005828.4	-0.70	novel	N/A
REEP6	-0.70	novel	27966653.
PELI2	-0.66	novel	30971501.
PNMA2	-0.65	novel	27003254.
WTIP	-0.65	anti-cancer	30690883, 20616046.
GFAP	-0.64	novel	30519355.
FAS	-0.62	pro-cancer	22508480, 10391843.
FKBP8	-0.62	novel	30348988, 14612567.
TRMT61B	-0.62	pro-cancer	32934298.
DUSP7	-0.61	novel	26057789, 29100300, 26057789, 33865857.
AC092117.2	-0.57	novel	N/A
RXRA	-0.57	regulated	29143738, 34239022.
CLUAP1	-0.56	novel	15480429, 31142279.
CYP51A1	-0.56	novel	27313059, 28658622, 27341022, 32660617.
EML1	-0.56	novel	24706829, 15713800.

Supplementary Table 2 Continued

Gene	CV change	Type	PubMed ID
ZNF10	-0.56	novel	31938422.
SNX16	-0.55	novel	31876369, 25408875, 28712807.
CEP112	-0.55	novel	30296263.
KIRREL3-AS2	-0.55	novel	29072050, 27094248.
ZNF746	-0.52	pro-cancer	29628506, 24145959, 30372570.
AC087620.1	-0.52	novel	N/A
FSTL3	-0.51	novel	28178680, 28583174.
TYW1B	-0.51	novel	26208496.
LCA5	-0.51	novel	30537927.
PSRC1	-0.51	novel	26070530, 22065898, 29399176.
ADGRE1	-0.49	novel	30327653, 32259086, 27832497.
GNAI1	-0.49	novel	23691483, 30836096, 31171722.
LEPR	-0.49	pro-cancer	21056997, 25261678, 24025407, 31789415.
KIAA1958	-0.48	novel	N/A
FABP6	-0.47	novel	16951225, 31651326.
KAZALD1	-0.46	novel	18403594, 26315110.
ZDHHC12	-0.46	novel	32963941, 24195083.
TCN1	-0.46	pro-cancer	28638446, 30100732, 29931431.
TNIK	-0.45	pro-cancer	31302315, 30442712, 24566388, 26499327, 28208209.
SLC25A20	-0.45	anti-cancer	27189278, 33824298.
MRPS30-DT	-0.45	novel	N/A
PTHLH	-0.44	pro-cancer	24861371, 28120940, 27450451, 31293052, 24880666, 24590291, 22546075.
PTPDC1	-0.44	novel	N/A
CACNA2D4	-0.44	novel	31510946, 18588891.
SIAE	-0.43	novel	31197190, 21618128, 26022516.
RPAP1	-0.43	novel	29320736, 19404390.
SCN9A	-0.43	novel	27529686, 29990870, 22992466.
SCRN2	-0.43	novel	23474757, 23049873.
PDIA3P1	-0.42	novel	28823960, 32127518.
ZNF25	-0.42	novel	N/A

Supplementary Table 2 Continued

Gene	CV change	Type	PubMed ID
LRRC37BP1	-0.42	novel	N/A
KAT2B	-0.42	novel	29671337, 27796307, 28637482.
LINC00963	-0.41	novel	32357409, 31751910, 29630107, 32052688.
ESAM	-0.41	novel	32637582, 18794116.
MINPP1	-0.40	novel	14732919, 11297621.
SNX11	-0.40	novel	31561579.
UNK	-0.40	novel	N/A
JHY	-0.40	novel	N/A
RBM24	-0.40	novel	27584791, 24375645.
PELI1	-0.39	pro-cancer	29523541, 33215753, 29078411, 32728028, 28009353.
PMF1	-0.38	anti-cancer	19088041, 23628807.
ZBTB18	-0.38	anti-cancer	28512252, 33892786.
NRAV	-0.38	novel	N/A
SNORA12	-0.37	novel	N/A
CARD16	-0.37	novel	31349612.
CERCAM	-0.37	novel	34105305.
ZBTB48	-0.36	novel	28500257, 31752241, 28082411.
C1orf131	-0.36	novel	N/A
ZBTB5	-0.36	novel	19491398.
FRMD8	-0.35	novel	29897336.
POC1A	-0.35	novel	33052878.
TBL1X	-0.35	pro-cancer	27129164, 25491321, 26069883, 30705280.
FBXO33	-0.35	novel	29559746.
ABCD3	-0.35	novel	25802834.
STXBP4	-0.35	pro-cancer	28087642, 29735662, 32993587.
ARFIP2	-0.35	novel	33318029.
PPP3CB	-0.35	novel	30641937.
CALM2P2	-0.34	novel	N/A
SGSM2	-0.34	novel	30744493.
PPIAP31	-0.34	novel	N/A
UBA7	-0.34	novel	32218830, 31974171, 26515094.

Supplementary Table 2 Continued

Gene	CV change	Type	PubMed ID
RPL15	-0.34	pro-cancer	32029688.
OPHN1	-0.33	novel	25170626, 16371738.
LINC00888	-0.33	novel	33215424, 34472153.
PTBP3	-0.33	pro-cancer	31989778, 29752441, 31291975, 30068940, 29187406, 32974175.
MFSD14B	-0.33	novel	N/A
PACS2	-0.33	regulated	28476937, 29936496, 19481529, 26108729, 24633224.
SLC25A43	-0.33	regulated	28781661, 22883974, 22430806.
TST	-0.33	novel	16500920.
MED10	-0.32	novel	N/A
ABLIM1	-0.32	anti-cancer	24453001, 33061800.
TPT1-AS1	-0.32	novel	32248186, 30045766, 33428595.
TUSC3	-0.31	regulated	28272772, 24435307, 25735931, 23096450, 27071482, 30098333.
UBXN7	-0.31	novel	22537386.
ERI3	-0.31	novel	N/A
HSCB	-0.31	novel	30096431.
LARP1B	-0.30	regulated	25531318, 27614686, 32094190.
PPIA	-0.30	pro-cancer	11830645, 25296734, 19503092.
SHPRH	-0.30	anti-cancer	17130289, 20096653, 25310987, 29343848.
NPDC1	-0.30	anti-cancer	11042687, 7878019.
RIT1	-0.29	regulated	24469055, 23765226, 30348939.
OSR2	-0.29	novel	27143812.
SRD5A3	-0.29	novel	17986282.
STK39	-0.29	novel	16707422, 19717643, 27542260.
AREL1	-0.29	novel	23479728, 31732561.
MYCBP	-0.29	novel	26868851, 26996564.
GATA6	-0.29	regulated	30674866, 27325420, 23707782, 23313142.
CEP104	-0.29	novel	N/A
NCAPD2	-0.28	novel	31610177.
ABI1	-0.28	pro-cancer	21118970, 32728066, 31488087.
SVIL-AS1	-0.28	novel	34264003.

Supplementary Table 2 Continued

Gene	CV change	Type	PubMed ID
MRPS9	-0.28	novel	34168990.
PMPCA	-0.28	novel	N/A
CALM2	-0.28	novel	18587642, 24188867.
CTDSPL	-0.28	anti-cancer	27414789, 29382357, 19885927, 15051889, 21643017.
CCNC	-0.28	anti-cancer	25344755, 9584184, 19683536, 19967789, 15084261, 19237534.
TAF1	-0.28	novel	31664040, 27571988, 20181722.
OGFRL1	-0.28	novel	30246052, 30270488.
GAPDHP73	-0.28	novel	N/A
IRS2	-0.27	regulated	23594372, 32015092, 31393907.
ARID5B	-0.27	novel	32595701, 29326336, 31573954.
COX7C	-0.27	regulated	32653968, 26798410, 29138278.
MYL12B	-0.27	novel	21678067.
CLK4	-0.27	novel	29606096.
LETMD1	-0.27	novel	31980577.
IFT57	-0.27	novel	N/A
GBP3	-0.27	novel	29128363.
NREP	-0.27	novel	24096364, 34746143.
PAPSS2	-0.27	novel	29955124.
FPGS	-0.27	novel	24457236, 15542523.
ABL1	-0.26	regulated	23842646.
HEG1	-0.26	novel	31278131.
KRCC1	-0.26	novel	31908025.
DAG1	-0.26	novel	26637171.
PUS7	-0.26	pro-cancer	35121864, 33811565, 33990203.
MANBAL	-0.26	novel	N/A
PMM1	-0.26	novel	N/A
PCBD1	-0.26	novel	24204001, 27877211.
KPNA3	-0.26	novel	30098595.
CHML	-0.26	regulated	31175290, 10652570.
ENO1	-0.26	pro-cancer	35434271, 34145039, 25887760, 25860938, 28086938.

Supplementary Table 2 Continued

Gene	CV change	Type	PubMed ID
MYLK	-0.26	novel	25179839, 31413665.
RC3H2	-0.26	novel	32163895.
AC233279.1	-0.26	novel	N/A
RRAD	-0.25	regulated	17195088, 21221513, 24648519, 22487779, 24222170, 31857616, 25893381.
NMT2	-0.25	regulated	16530191, 16123142, 33093447.
LTBR	-0.25	regulated	26977880, 20603617.
INSIG1	-0.25	novel	12115587.
RPL5P4	-0.25	novel	N/A
H3F3B	-0.25	novel	24229707, 24162739, 15870878, 28280610.
BMPR2	-0.25	regulated	18471510, 21576484, 25501832, 28938584, 28391780.
MAPRE2	-0.25	novel	32426268, 19787265.
RAB14	-0.25	pro-cancer	31175290, 28107526.
NRDC	-0.24	pro-cancer	28207963, 29669932, 28230087.
WASHC2A	-0.24	novel	N/A
GSTO1	-0.24	pro-cancer	30873742, 32571799.
ZNF148	-0.24	novel	28447668, 23576061, 30463804.
TCEAL1	-0.24	novel	33033111, 15773840.
NR1D2	-0.24	novel	29773903.
DANCR	-0.24	pro-cancer	29180471, 29651883.
SLC35F2	-0.24	novel	23879892.
RPL14	-0.24	novel	16316724.
CCDC28A	-0.24	novel	N/A
HMGN2	-0.23	anti-cancer	27358110, 25060707.
DNAJC3	-0.23	novel	30652414.
DCTD	-0.23	novel	29523831, 28912488.
GPSM2	-0.23	novel	20589935, 32195179, 32812493, 32519357.
DECR1	-0.23	regulated	17636013, 32427840.
OSMR	-0.23	novel	28288136, 21508378, 30755233.
NDUFC1	-0.23	novel	N/A
THOC2	-0.23	novel	31680623.

Supplementary Table 2 Continued

Gene	CV change	Type	PubMed ID
KIF3C	-0.23	novel	32020481, 26272184.
AC244153.1	-0.23	novel	N/A
UTP11	-0.23	novel	33004423.
BROX	-0.23	novel	32385852.
TMX4	-0.23	novel	N/A
FTLP3	-0.23	novel	N/A
PFN1	-0.23	regulated	30318519, 25103363, 28699810.
AC007969.1	-0.23	novel	N/A
TNFAIP1	-0.22	novel	31901862, 23912453.
NANS	-0.22	novel	31659112, 32772211.
CDK2AP2	-0.22	novel	N/A
NUFIP2	-0.22	novel	N/A
EIF4A2	-0.22	novel	31308851, 31088567.
RCN1	-0.22	novel	29453900, 28319095, 30172915.
SLC38A2	-0.22	pro-cancer	32341021, 33028955, 31152137, 32996252.
BSG	-0.22	pro-cancer	16604531, 20060267, 20551248, 18616931.
UQCRFS1	-0.22	novel	15047214, 29147009, 21901141.
ALDH9A1	-0.22	novel	20516116, 25826619.
RPLP0P6	-0.22	novel	N/A
VPS53	-0.22	novel	N/A
PRDX4	-0.22	pro-cancer	31311441.
MAST4	-0.22	novel	26544852.
TMEM59	-0.21	novel	29632210.
RPL14P1	-0.21	novel	N/A
ECI2	-0.21	novel	28415728.
HINT1	-0.21	anti-cancer	32636443, 18596417, 29214080, 12810953, 16186798, 17510397.
LSM6	-0.21	novel	34638387.
GTF3A	-0.21	novel	29955133, 22783697.
TOB2	-0.21	novel	10602502, 28922388, 21242960, 24481647.
CIAO2B	-0.21	novel	34573089.
ARPC5	-0.21	novel	22378351, 22089643.

Supplementary Table 2 Continued

Gene	CV change	Type	PubMed ID
ATP5MG	-0.21	novel	30842661.
SKP1	-0.21	pro-cancer	27238229, 26474281, 33292299.
MPP5	-0.21	novel	24488012.
ATP6V1G1	-0.21	novel	26020805, 19758758.
SDCBP	-0.21	pro-cancer	32061839, 23533663.
PERP	-0.21	novel	32843065, 10733530.
DNAJC8	-0.21	novel	29902837, 27133716.
SPART	-0.21	novel	21499309, 29673586.
SSR4	-0.21	novel	34904499.
YES1	-0.21	regulated	31166114, 31391186, 21779430, 32572172.
TPT1	-0.21	pro-cancer	29149404, 30941821.
CNN3	-0.21	pro-cancer	32051425, 30911294, 24934327.
MLF1	-0.21	regulated	29296815, 11021751, 15861129, 17967869.
ASPM	-0.20	pro-cancer	30266990, 17090670, 18676753.
PVR	-0.20	pro-cancer	30275538, 15471548, 32894141, 31741778, 32554931.
ITGA6	-0.20	pro-cancer	27001172, 27624978, 30894280.
CCAR1	-0.20	regulated	25894788, 26511806, 33403784, 28230774, 16543231, 20594350.
PNKD	-0.20	novel	27005424.
TNFRSF12A	-0.20	novel	30142200.
LDHA	-0.20	pro-cancer	30403008, 31146503, 31035592.
VKORC1	-0.20	novel	N/A
EEF1B2P3	-0.20	novel	N/A
PRNP	-0.20	novel	23497519.
NDUFA8	-0.20	novel	30389505.
MAP3K2	-0.20	pro-cancer	24847881, 28069384, 27498924.
DDX3X	-0.20	regulated	31906196, 16818630, 29782654, 27180681, 26541825, 31216476, 30297359.
RTRAF	-0.20	novel	N/A
RND3	-0.20	regulated	27555595, 22234932, 21209796, 27144337.
UBE2A	-0.20	novel	31473606.
SBDS	-0.20	novel	32269318, 32198344.

Supplementary Table 2 Continued

Gene	CV change	Type	PubMed ID
B2M	-0.20	pro-cancer	21427356, 19056512, 23848204, 26908438, 23874600, 25899529.
DNTTIP2	-0.20	novel	35370417.
CENPM	-0.19	novel	31703591.
RSL24D1	-0.19	novel	33955587.
REXO2	-0.19	novel	32354376.
PRPF40A	-0.19	novel	28258342.
PITPNB	-0.19	novel	N/A
CCDC167	-0.19	novel	33461170.
PAGE5	-0.19	novel	23028975, 27769045.
PSME2	-0.19	novel	34779489.
SYNGR2	-0.19	novel	N/A
CAPN2	-0.19	pro-cancer	29228653, 32382656, 22427650.
PPP1R2	-0.19	novel	33238888, 11251179.
AKAP13	-0.19	novel	26272591, 22161024, 34673774.
RRAS2	-0.19	pro-cancer	35120522, 24148564, 26619011, 31517733.
TIMM8B	-0.19	novel	31199051, 23172368.
POLR2G	-0.19	novel	31496805.
ORC3	-0.19	novel	N/A
DNTTIP1	-0.19	novel	29855544.
SEC31A	-0.19	novel	20207848, 25715771, 21109691.
ZNF644	-0.19	novel	N/A
NIN	-0.19	novel	32934817.
BAZ1A	-0.19	novel	31085244.
CD82	-0.18	anti-cancer	16260083, 28881668, 9671393.
CISD3	-0.18	novel	29259115, 34497268.
MYL12A	-0.18	novel	21678067.
FIS1	-0.18	novel	34051059, 28589083, 33890233.
VPS28	-0.18	novel	24641493.
FUT8	-0.18	pro-cancer	28609658, 28982386, 32085441, 23267084.
MORF4L2	-0.18	novel	25796627.
ATP6V1F	-0.18	novel	33194594.

Supplementary Table 2 Continued

Gene	CV change	Type	PubMed ID
HIF1A	-0.18	pro-cancer	13130303, 30411685.
HDAC1	-0.18	pro-cancer	27599530, 31910827, 19351825, 14506733, 15042618, 16172792.
IGFBP6	-0.18	anti-cancer	29577195, 25808083, 29582205, 30231881, 18985860.
MEA1	-0.18	novel	N/A
EREG	-0.18	pro-cancer	22964644, 31234944, 28210154, 30252132.
HIPK3	-0.18	novel	32368306, 29549306, 30352682, 28794202.
FOSL1	-0.18	pro-cancer	30563891, 29112457, 28220783, 30575900.
UAP1	-0.18	novel	25241896, 32028604.
DDX52	-0.18	novel	34399732, 34233596.
TRA2B	-0.18	novel	23935626, 25620705.
PSMB9	-0.18	novel	15930337, 27659694, 32060274.
SPCS3	-0.18	novel	N/A
NXF1	-0.18	novel	32236527.
NUCB2	-0.18	novel	23958433, 28714371.
IPO7	-0.18	novel	22284678, 22815235.
CAVIN1	-0.18	regulated	23770857, 32493699, 24123650, 22461895.
CLIP1	-0.18	novel	31171381, 28970558, 32467651.
H2AFZ	-0.18	pro-cancer	26863632, 20023423, 31015444.
CAVIN3	-0.18	anti-cancer	33101009, 24069528.
MRPS15	-0.18	novel	23172368.
RTL8C	-0.18	novel	N/A
PFDN1	-0.18	pro-cancer	27694898, 26553318.
GRB2	-0.17	novel	18620523.
GNB1	-0.17	novel	23603342, 28650474.
AP3D1	-0.17	novel	17125464.
BMP4	-0.17	anti-cancer	29234727, 22749032, 31941699, 22167620, 12552124, 25224959, 31409851, 30153566, 24053318, 32376787.
LAMTOR4	-0.17	novel	32934076.
MYH9	-0.17	regulated	32376787, 30739906, 21533124, 24436421, 16651425, 22161839, 30854093, 27262074.

Supplementary Table 2 Continued

Gene	CV change	Type	PubMed ID
CCT3	-0.17	pro-cancer	32518527, 26739059, 29340068, 31501420, 33072568.
RALBP1	-0.17	pro-cancer	21170262, 22902412.
PCYOX1	-0.17	novel	30129068.
SPANXC	-0.17	novel	26895102, 32571981.
LDHB	-0.17	pro-cancer	31146503, 21199794.
AK2	-0.17	regulated	32509571, 31780678.
NKAP	-0.17	pro-cancer	31277684, 31632967, 35064112, 33754052, 32032976.
ANGPTL4	-0.17	regulated	24173241, 30814490, 28894280, 32928141.
NASP	-0.16	novel	30076957, 28173777, 19439102.
CCNL1	-0.16	novel	16598186, 21586274, 15700036.
C12orf57	-0.16	novel	N/A
BOD1L1	-0.16	novel	29138278.
MKI67	-0.16	regulated	33658388, 26281861, 25384676, 23674192.
LSM3	-0.16	novel	17054779.
TNIP1	-0.16	novel	22147607, 26046540.
BAD	-0.15	regulated	12404123, 9824152, 11494146.
SUPT16H	-0.15	novel	23973375, 22031019, 26678313.
FRMD6	-0.15	anti-cancer	33249427, 21666719, 31128910, 34669997.
ARL6IP5	-0.15	novel	35293383.
UBE2E3	-0.15	novel	29879550.

REFERENCES

1. C. L. Chaffer, R. A. Weinberg, A perspective on cancer cell metastasis. *science* **331**, 1559-1564 (2011).

2. S. Vanharanta, J. Massagué, Origins of metastatic traits. *Cancer cell* **24**, 410-421 (2013).

3. S. W. Brady, J. A. McQuerry, Y. Qiao, S. R. Piccolo, G. Shrestha, D. F. Jenkins, R. M. Layer, B. S. Pedersen, R. H. Miller, A. Esch, Combating subclonal evolution of resistant cancer phenotypes. *Nature communications* **8**, 1-15 (2017).

4. G. Housman, S. Byler, S. Heerboth, K. Lapinska, M. Longacre, N. Snyder, S. Sarkar, Drug resistance in cancer: an overview. *Cancers* **6**, 1769-1792 (2014).

5. I. Dagogo-Jack, A. T. Shaw, Tumour heterogeneity and resistance to cancer therapies. *Nature reviews Clinical oncology* **15**, 81-94 (2018).

6. S. L. Spencer, S. Gaudet, J. G. Albeck, J. M. Burke, P. K. Sorger, Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* **459**, 428-432 (2009).

7. Y. Wolf, O. Bartok, S. Patkar, G. B. Eli, S. Cohen, K. Litchfield, R. Levy, A. Jimenez-Sanchez, S. Trabish, J. S. Lee, H. Karathia, E. Barnea, C. P. Day, E. Cinnamon, I. Stein, A. Solomon, L. Bitton, E. Perez-Guijarro, T. Dubovik, S. S. Shen-Orr, M. L. Miller, G. Merlino, Y. Levin, E. Pikarsky, L. Eisenbach, A. Admon, C. Swanton, E. Ruppin, Y. Samuels, UVB-Induced Tumor Heterogeneity Diminishes Immune Response in Melanoma. *Cell* **179**, 219-235 e221 (2019).

8. A. Nguyen, M. Yoshida, H. Goodarzi, S. F. Tavazoie, Highly variable cancer subpopulations that exhibit enhanced transcriptome variability and metastatic fitness. *Nat Commun* **7**, 11246 (2016).

9. R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians* **72**, 7-33 (2022).

10. R. G. do Nascimento, K. M. Otoni, Histological and molecular classification of breast cancer: what do we know. *Mastology* **30**, e20200024 (2020).

11. C. K. Anders, L. A. Carey, Biology, metastatic patterns, and treatment of patients with triple-negative breast cancer. *Clinical breast cancer* **9**, S73-S81 (2009).

12. A. M. McCarthy, T. Friebel-Klingner, S. Ehsan, W. He, M. Welch, J. Chen, D. Kontos, S. M. Domchek, E. F. Conant, A. Semine, Relationship of established risk factors with breast cancer subtypes. *Cancer medicine* **10**, 6456-6467 (2021).

13. M. Zhao, Z. Li, H. Qu, An evidence-based knowledgebase of metastasis suppressors to identify key pathways relevant to cancer metastasis. *Scientific reports* **5**, 1-12 (2015).

14. Z. Fu, P. C. Smith, L. Zhang, M. A. Rubin, R. L. Dunn, Z. Yao, E. T. Keller, Effects of raf kinase inhibitor protein expression on suppression of prostate cancer metastasis. *J Natl Cancer Inst* **95**, 878-889 (2003).

15. S. Dangi-Garimella, J. Yun, E. M. Eves, M. Newman, S. J. Erkeland, S. M. Hammond, A. J. Minn, M. R. Rosner, Raf kinase inhibitory protein suppresses a metastasis signalling cascade involving LIN28 and let-7. *EMBO J* **28**, 347-358 (2009).

16. K. Lamiman, J. M. Keller, A. Mizokami, J. Zhang, E. T. Keller, Survey of Raf kinase inhibitor protein (RKIP) in multiple cancer types. *Critical Reviews™ in Oncogenesis* **19**, (2014).

17. A. E. Yesilkanal, M. R. Rosner, Targeting Raf Kinase Inhibitory Protein Regulation and Function. *Cancers (Basel)* **10**, (2018).

18. K. Yeung, T. Seitz, S. Li, P. Janosch, B. McFerran, C. Kaiser, F. Fee, K. D. Katsanakis, D. W. Rose, H. Mischak, J. M. Sedivy, W. Kolch, Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP. *Nature* **401**, 173-177 (1999).

19. J. Yun, C. A. Frankenberger, W. L. Kuo, M. C. Boelens, E. M. Eves, N. Cheng, H. Liang, W. H. Li, H. Ishwaran, A. J. Minn, M. R. Rosner, Signalling pathway for RKIP and Let-7 regulates and predicts metastatic breast cancer. *EMBO J* **30**, 4500-4514 (2011).

20. A. J. Minn, E. Bevilacqua, J. Yun, M. R. Rosner, Identification of novel metastasis suppressor signaling pathways for breast cancer. *Cell cycle* **11**, 2452-2457 (2012).

21. Y. Kang, P. M. Siegel, W. Shu, M. Drobnjak, S. M. Kakonen, C. Cordon-Cardo, T. A. Guise, J. Massague, A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**, 537-549 (2003).

22. M. Sun, S. Gomes, P. Chen, C. A. Frankenberger, D. Sankarasharma, C. H. Chung, K. K. Chada, M. R. Rosner, RKIP and HMGA2 regulate breast tumor survival and metastasis through lysyl oxidase and syndecan-2. *Oncogene* **33**, 3528-3537 (2014).

23. A. E. Yesilkanal, D. Yang, A. Valdespino, P. Tiwari, A. U. Sabino, L. C. Nguyen, J. Lee, X. H. Xie, S. Sun, C. Dann, L. Robinson-Mailman, E. Steinberg, T. Stuhlmiller, C. Frankenberger, E. Goldsmith, G. L. Johnson, A. F. Ramos, M. R. Rosner, Limited inhibition of multiple nodes in a driver network blocks metastasis. *Elife* **10**, (2021).

24. K. Lorenz, M. J. Lohse, U. Quitterer, Protein kinase C switches the Raf kinase inhibitor from Raf-1 to GRK-2. *Nature* **426**, 574-579 (2003).

25. J. J. Skinner, M. R. Rosner, RKIP structure drives its function: a three-state model for regulation of RKIP. *Critical Reviews™ in Oncogenesis* **19**, (2014).

26. J. J. Skinner, S. Wang, J. Lee, C. Ong, R. Sommese, S. Sivaramakrishnan, W. Koelmel, M. Hirschbeck, H. Schindelin, C. Kisker, Conserved salt-bridge competition triggered by phosphorylation regulates the protein interactome. *Proceedings of the National Academy of Sciences* **114**, 13453-13458 (2017).

27. J. Lee, J. Lee, K. S. Farquhar, J. Yun, C. A. Frankenberger, E. Bevilacqua, K. Yeung, E.-J. Kim, G. Balázsi, M. R. Rosner, Network of mutually repressive metastasis regulators can promote cell heterogeneity and metastatic transitions. *Proceedings of the National Academy of Sciences* **111**, E364-E373 (2014).

28. K. Ogawa, J. Sun, S. Taketani, O. Nakajima, C. Nishitani, S. Sassa, N. Hayashi, M. Yamamoto, S. Shibahara, H. Fujita, Heme mediates derepression of Maf recognition element through direct binding to transcription repressor Bach1. *The EMBO journal* **20**, 2835-2843 (2001).

29. M. Kobayashi, H. Kato, H. Hada, A. Itoh-Nakadai, T. Fujiwara, A. Muto, Y. Inoguchi, K. Ichiyanagi, W. Hojo, N. Tomosugi, Iron-heme-Bach1 axis is involved in erythroblast adaptation to iron deficiency. *Haematologica* **102**, 454 (2017).

30. M. Brand, J. A. Ranish, N. T. Kummer, J. Hamilton, K. Igarashi, C. Francastel, T. H. Chi, G. R. Crabtree, R. Aebersold, M. Groudine, Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nature structural & molecular biology* **11**, 73-80 (2004).

31. H. P. J. Voon, J. Vadolas, Controlling α-globin: a review of α-globin expression and its impact on β-thalassemia. *Haematologica* **93**, 1868-1876 (2008).

32. J. Sun, H. Hoshino, K. Takaku, O. Nakajima, A. Muto, H. Suzuki, S. Tashiro, S. Takahashi, S. Shibahara, J. Alam, Hemoprotein Bach1 regulates enhancer availability of heme oxygenase-1 gene. *The EMBO journal* **21**, 5216-5224 (2002).

33. M. D. Maines, The heme oxygenase system: a regulator of second messenger gases. *Annual review of pharmacology and toxicology* **37**, 517 (1997).

34. M. Ishikawa, S. Numazawa, T. Yoshida, Redox regulation of the transcriptional repressor Bach1. *Free Radical Biology and Medicine* **38**, 1344-1352 (2005).

35. J. W. Kaspar, A. K. Jaiswal, Antioxidant-induced phosphorylation of tyrosine 486 leads to rapid nuclear export of Bach1 that allows Nrf2 to bind to the antioxidant response element and activate defensive gene expression. *Journal of Biological Chemistry* **285**, 153-162 (2010).

36. H.-J. Warnatz, D. Schmidt, T. Manke, I. Piccini, M. Sultan, T. Borodina, D. Balzereit, W. Wruck, A. Soldatov, M. Vingron, The BTB and CNC homology 1 (BACH1) target genes are involved in the oxidative stress response and in control of the cell cycle. *Journal of Biological Chemistry* **286**, 23521-23532 (2011).

37. J. Lee, A. E. Yesilkanal, J. P. Wynne, C. Frankenberger, J. Liu, J. Yan, M. Elbaz, D. C. Rabe, F. D. Rustandy, P. Tiwari, E. A. Grossman, P. C. Hart, C. Kang, S. M. Sanderson, J. Andrade, D. K. Nomura, M. G. Bonini, J. W. Locasale, M. R. Rosner, Effective breast cancer combination therapy targeting BACH1 and mitochondrial metabolism. *Nature* **568**, 254-258 (2019).

38. C. Wiel, K. Le Gal, M. X. Ibrahim, C. A. Jahangir, M. Kashif, H. Yao, D. V. Ziegler, X. Xu, T. Ghosh, T. Mondal, C. Kanduri, P. Lindahl, V. I. Sayin, M. O. Bergo, BACH1 Stabilization by Antioxidants Stimulates Lung Cancer Metastasis. *Cell* **178**, 330-345 e322 (2019).

39. M. Sato, M. Matsumoto, Y. Saiki, M. Alam, H. Nishizawa, M. Rokugo, A. Brydun, S. Yamada, M. K. Kaneko, R. Funayama, M. Ito, Y. Kato, K. Nakayama, M. Unno, K. Igarashi, BACH1 Promotes Pancreatic Cancer Metastasis by Repressing Epithelial Genes and Enhancing Epithelial-Mesenchymal Transition. *Cancer Res* **80**, 1279-1292 (2020).

40. W. Han, Y. Zhang, C. Niu, J. Guo, J. Li, X. Wei, M. Jia, X. Zhi, L. Yao, D. Meng, BTB and CNC homology 1 (Bach1) promotes human ovarian cancer cell metastasis by HMGA2-mediated epithelial-mesenchymal transition. *Cancer Letters* **445**, 45-56 (2019).

41. Y. Zhao, J. Gao, X. Xie, P. Nan, F. Liu, Y. Sun, X. Zhao, BACH1 promotes the progression of esophageal squamous cell carcinoma by inducing the epithelial–mesenchymal transition and angiogenesis. *Cancer medicine* **10**, 3413-3426 (2021).

42. Y. Liang, H. Wu, R. Lei, R. A. Chong, Y. Wei, X. Lu, I. Tagkopoulos, S.-Y. Kung, Q. Yang, G. Hu, Transcriptional network analysis identifies BACH1 as a master regulator of breast cancer bone metastasis. *Journal of Biological Chemistry* **287**, 33533-33544 (2012).

43. C. Visciano, F. Liotti, N. Prevete, G. Cali, R. Franco, F. Collina, A. De Paulis, G. Marone, M. Santoro, R. Melillo, Mast cells induce epithelial-to-mesenchymal transition and stem cell features in human thyroid cancer cells through an IL-8–Akt–Slug pathway. *Oncogene* **34**, 5175-5186 (2015).

44. B. Beck, G. Lapouge, S. Rorive, B. Drogat, K. Desaedelaere, S. Delafaille, C. Dubois, I. Salmon, K. Willekens, J.-C. Marine, Different levels of Twist1 regulate skin tumor initiation, stemness, and progression. *Cell stem cell* **16**, 67-79 (2015).

45. L. Jiang, M. Yin, J. Xu, M. Jia, S. Sun, X. Wang, J. Zhang, D. Meng, The transcription factor Bach1 suppresses the developmental angiogenesis of zebrafish. *Oxidative Medicine and Cellular Longevity* **2017**, (2017).

46. B. Cohen, H. Tempelhof, T. Raz, R. Oren, J. Nicenboim, F. Bochner, R. Even, A. Jelinski, R. Eilam, S. Ben-Dor, BACH family members regulate angiogenesis and lymphangiogenesis by modulating VEGFC expression. *Life science alliance* **3**, (2020).

47. X. Huang, J. Zheng, J. Li, X. Che, W. Tan, W. Tan, M. Shao, X. Cheng, Z. Du, Y. Zhao, Functional role of BTB and CNC Homology 1 gene in pancreatic cancer and its association with survival in patients treated with gemcitabine. *Theranostics* **8**, 3366 (2018).

48. J. Li, T. Shiraki, K. Igarashi, Transcription-independent role of Bach1 in mitosis through a nuclear exporter Crm1-dependent mechanism. *FEBS letters* **586**, 448-454 (2012).

49. J. Li, H. Shima, H. Nishizawa, M. Ikeda, A. Brydun, M. Matsumoto, H. Kato, Y. Saiki, L. Liu, M. Watanabe-Matsui, Phosphorylation of BACH1 switches its function from transcription factor to mitotic chromosome regulator and promotes its interaction with HMMR. *Biochemical Journal* **475**, 981-1002 (2018).

50. V. Svensson, R. Vento-Tormo, S. A. Teichmann, Exponential scaling of single-cell RNA-seq in the past decade. *Nature protocols* **13**, 599-604 (2018).

51. S. Islam, U. Kjällquist, A. Moliner, P. Zajac, J.-B. Fan, P. Lönnerberg, S. Linnarsson, Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing. *Nature protocols* **7**, 813-828 (2012).

52. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202-1214 (2015).

53. L. Mazutis, J. Gilbert, W. L. Ung, D. A. Weitz, A. D. Griffiths, J. A. Heyman, Single-cell analysis and sorting using droplet-based microfluidics. *Nature protocols* **8**, 870-891 (2013).

54. P. See, J. Lum, J. Chen, F. Ginhoux, A single-cell sequencing guide for immunologists. *Frontiers in immunology* **9**, 2425 (2018).

55. G. X. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, Massively parallel digital transcriptional profiling of single cells. *Nature communications* **8**, 1-12 (2017).

56. A. Zeisel, A. B. Muñoz-Manchado, S. Codeluppi, P. Lönnerberg, G. La Manno, A. Juréus, S. Marques, H. Munguba, L. He, C. Betsholtz, Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-1142 (2015).

57. S. Islam, U. Kjällquist, A. Moliner, P. Zajac, J.-B. Fan, P. Lönnerberg, S. Linnarsson, Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome research* **21**, 1160-1167 (2011).

58. L. D. Goldstein, Y.-J. J. Chen, J. Dunne, A. Mir, H. Hubschle, J. Guillory, W. Yuan, J. Zhang, J. Stinson, B. Jaiswal, Massively parallel nanowell-based single-cell gene expression profiling. *BMC genomics* **18**, 1-10 (2017).

59. T. M. Gierahn, M. H. Wadsworth, T. K. Hughes, B. D. Bryson, A. Butler, R. Satija, S. Fortune, J. C. Love, A. K. Shalek, Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nature methods* **14**, 395-398 (2017).

60. N. Habib, I. Avraham-Davidi, A. Basu, T. Burks, K. Shekhar, M. Hofree, S. R. Choudhury, F. Aguet, E. Gelfand, K. Ardlie, Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature methods* **14**, 955-958 (2017).

61. B. B. Lake, R. Ai, G. E. Kaeser, N. S. Salathia, Y. C. Yung, R. Liu, A. Wildberg, D. Gao, H.-L. Fung, S. Chen, Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586-1590 (2016).

62. T. E. Bakken, R. D. Hodge, J. A. Miller, Z. Yao, T. N. Nguyen, B. Aevermann, E. Barkan, D. Bertagnolli, T. Casper, N. Dee, Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PloS one* **13**, e0209648 (2018).

63. Y. Liu, J. Zhou, K. P. White, RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* **30**, 301-304 (2014).

64. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, 3rd, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).

65. P. V. Kharchenko, L. Silberstein, D. T. Scadden, Bayesian approach to single-cell differential expression analysis. *Nat Methods* **11**, 740-742 (2014).

66. K. K. Dey, C. J. Hsiao, M. Stephens, Visualizing the structure of RNA-seq expression data using grade of membership models. *PLoS Genet* **13**, e1006599 (2017).

67. J. R. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, J. S. Weissman, Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840-846 (2006).

68. R. V. Desai, X. Chen, B. Martin, S. Chaturvedi, D. W. Hwang, W. Li, C. Yu, S. Ding, M. Thomson, R. H. Singer, R. A. Coleman, M. M. K. Hansen, L. S. Weinberger, A DNA repair pathway can regulate transcriptional noise to promote cell fate transitions. *Science* **373**, (2021).

69. A. Raj, A. van Oudenaarden, Single-molecule approaches to stochastic gene expression. *Annual review of biophysics* **38**, 255-270 (2009).

70. F. Wang, J. Flanagan, N. Su, L. C. Wang, S. Bui, A. Nielson, X. Wu, H. T. Vo, X. J. Ma, Y. Luo, RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn* **14**, 22-29 (2012).

71. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

72. A. Frankish, M. Diekhans, I. Jungreis, J. Lagarde, J. E. Loveland, J. M. Mudge, C. Sisu, J. C. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, C. Boix, S. Carbonell Sala, F. Cunningham, T. Di Domenico, S. Donaldson, I. T. Fiddes, C. Garcia Giron, J. M. Gonzalez, T. Grego, M. Hardy, T. Hourlier, K. L. Howe, T. Hunt, O. G. Izuogu, R. Johnson, F. J. Martin, L. Martinez, S. Mohanan, P. Muir, F. C. P. Navarro, A. Parker, B. Pei, F. Pozo, F. C. Riera, M. Ruffier, B. M. Schmitt, E. Stapleton, M. M. Suner, I. Sycheva, B. Uszczynska-Ratajczak, M. Y. Wolf, J. Xu, Y. T. Yang, A. Yates, D. Zerbino, Y. Zhang, J. S. Choudhary, M. Gerstein, R. Guigo, T. J. P. Hubbard, M. Kellis, B. Paten, M. L. Tress, P. Flicek, Gencode 2021. *Nucleic Acids Res* **49**, D916-D923 (2021).

73. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).

74. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal, complex systems* **1695**, 1-9 (2006).

75. Y. Zhou, B. Zhou, L. Pache, M. Chang, A. H. Khodabakhshi, O. Tanaseichuk, C. Benner, S. K. Chanda, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).

76. L. C. Nguyen, D. M. Renner, D. Silva, D. Yang, N. A. Parenti, K. M. Medina, V. Nicolaescu, H. Gula, N. Drayman, A. Valdespino, SARS-CoV-2 Diverges from Other Betacoronaviruses in Only Partially Activating the IRE1α/XBP1 Endoplasmic Reticulum Stress Pathway in Human Lung-Derived Cells. *Mbio*, e02415-02422 (2022).

77. L. C. Nguyen, D. Yang, V. Nicolaescu, T. J. Best, H. Gula, D. Saxena, J. D. Gabbard, S.-N. Chen, T. Ohtsuki, J. B. Friesen, Cannabidiol inhibits SARS-CoV-2 replication through induction of the host ER stress and innate immune responses. *Science Advances* **8**, eabi6110 (2022).

78. E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander, N. Schultz, The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404 (2012).

79. N. Cancer Genome Atlas Research, J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, J. M. Stuart, The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113-1120 (2013).

80. A. Kramer, J. Green, J. Pollard, Jr., S. Tugendreich, Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-530 (2014).

81. C. Gene Ontology, The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **49**, D325-D334 (2021).

82. M. Gillespie, B. Jassal, R. Stephan, M. Milacic, K. Rothfels, A. Senff-Ribeiro, J. Griss, C. Sevilla, L. Matthews, C. Gong, C. Deng, T. Varusai, E. Ragueneau, Y. Haider, B. May, V. Shamovsky, J. Weiser, T. Brunson, N. Sanati, L. Beckman, X. Shao, A. Fabregat, K. Sidiropoulos, J. Murillo, G. Viteri, J. Cook, S. Shorser, G. Bader, E. Demir, C. Sander, R. Haw, G. Wu, L. Stein, H. Hermjakob, P. D'Eustachio, The reactome pathway knowledgebase 2022. *Nucleic Acids Res* **50**, D687-D692 (2022).

83. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).

84. E. Afgan, D. Baker, B. Batut, M. van den Beek, D. Bouvier, M. Cech, J. Chilton, D. Clements, N. Coraor, B. A. Gruning, A. Guerler, J. Hillman-Jackson, S. Hiltemann, V. Jalili, H. Rasche, N. Soranzo, J. Goecks, J. Taylor, A. Nekrutenko, D. Blankenberg, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* **46**, W537-W544 (2018).

85. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

86. R. L. Grossman, A. P. Heath, V. Ferretti, H. E. Varmus, D. R. Lowy, W. A. Kibbe, L. M. Staudt, Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med* **375**, 1109-1112 (2016).

87. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

88. M. J. Goldman, B. Craft, M. Hastie, K. Repecka, F. McDade, A. Kamath, A. Banerjee, Y. Luo, D. Rogers, A. N. Brooks, J. Zhu, D. Haussler, Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol* **38**, 675-678 (2020).

89. C. W. Law, Y. Chen, W. Shi, G. K. Smyth, voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29 (2014).

90. R. Liu, A. Z. Holik, S. Su, N. Jansz, K. Chen, H. S. Leong, M. E. Blewitt, M. L. Asselin-Labat, G. K. Smyth, M. E. Ritchie, Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Res* **43**, e97 (2015).

91. H. Wickham, in *ggplot2*. (Springer, 2016), pp. 189-201.

92. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289-300 (1995).

93. R. Y. Ebright, S. Lee, B. S. Wittner, K. L. Niederhoffer, B. T. Nicholson, A. Bardia, S. Truesdell, D. F. Wiley, B. Wesley, S. Li, A. Mai, N. Aceto, N. Vincent-Jordan, A. Szabolcs, B. Chirn, J. Kreuzer, V. Comaills, M. Kalinich, W. Haas, D. T. Ting, M. Toner, S. Vasudevan, D. A. Haber, S. Maheswaran, D. S. Micalizzi, Deregulation of ribosomal protein expression and translation promotes breast cancer metastasis. *Science* **367**, 1468-1473 (2020).

94. R. D. Dar, S. M. Shaffer, A. Singh, B. S. Razooky, M. L. Simpson, A. Raj, L. S. Weinberger, Transcriptional bursting explains the noise–versus–mean relationship in mRNA and protein levels. *PloS one* **11**, e0158298 (2016).

95. W. J. Blake, G. Balázsi, M. A. Kohanski, F. J. Isaacs, K. F. Murphy, Y. Kuang, C. R. Cantor, D. R. Walt, J. J. Collins, Phenotypic consequences of promoter-mediated transcriptional noise. *Molecular cell* **24**, 853-865 (2006).

96. E. Edgington, P. Onghena, *Randomization tests*. (Chapman and Hall/CRC, 2007).

97. M. Thattai, A. Van Oudenaarden, Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences* **98**, 8614-8619 (2001).

98. J. M. Pedraza, A. van Oudenaarden, Noise propagation in gene networks. *Science* **307**, 1965-1969 (2005).

99. M. T. Guinn, Y. Wan, S. Levovitz, D. Yang, M. R. Rosner, G. Balazsi, Observation and Control of Gene Expression Noise: Barrier Crossing Analogies Between Drug Resistance and Metastasis. *Front Genet* **11**, 586726 (2020).

100. Z. Steinhart, Z. Pavlovic, M. Chandrashekhar, T. Hart, X. Wang, X. Zhang, M. Robitaille, K. R. Brown, S. Jaksani, R. Overmeer, S. F. Boj, J. Adams, J. Pan, H. Clevers, S. Sidhu, J. Moffat, S. Angers, Genome-wide CRISPR screens reveal a Wnt-FZD5 signaling circuit as a druggable vulnerability of RNF43-mutant pancreatic tumors. *Nat Med* **23**, 60-68 (2017).

101. E. Evron, C. B. Umbricht, D. Korz, V. Raman, D. M. Loeb, B. Niranjan, L. Buluwela, S. A. Weitzman, J. Marks, S. Sukumar, Loss of cyclin D2 expression in the majority of breast

cancers is associated with promoter hypermethylation. *Cancer research* **61**, 2782-2787 (2001).

102. W. Wei, S. C. Mok, E. Oliva, S. H. Kim, G. Mohapatra, M. J. Birrer, FGF18 as a prognostic and therapeutic biomarker in ovarian cancer. *J Clin Invest* **123**, 4435-4448 (2013).

103. B. G. Cuiffo, A. Campagne, G. W. Bell, A. Lembo, F. Orso, E. C. Lien, M. K. Bhasin, M. Raimo, S. E. Hanson, A. Marusyk, D. El-Ashry, P. Hematti, K. Polyak, F. Mechta-Grigoriou, O. Mariani, S. Volinia, A. Vincent-Salomon, D. Taverna, A. E. Karnoub, MSC-regulated microRNAs converge on the transcription factor FOXP2 and promote breast cancer metastasis. *Cell Stem Cell* **15**, 762-774 (2014).

104. G. Schotta, M. Lachner, K. Sarma, A. Ebert, R. Sengupta, G. Reuter, D. Reinberg, T. Jenuwein, A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & development* **18**, 1251-1262 (2004).

105. A. S. Pal, A. Agredo, N. A. Lanman, J. Son, I. S. Sohal, M. Bains, C. Li, J. Clingerman, K. Gates, A. L. Kasinski, Loss of KMT5C Promotes EGFR Inhibitor Resistance in NSCLC via LINC01510-Mediated Upregulation of MET. *Cancer Res* **82**, 1534-1547 (2022).

106. M. Viotti, C. Wilson, M. McCleland, H. Koeppen, B. Haley, S. Jhunjhunwala, C. Klijn, Z. Modrusan, D. Arnott, M. Classon, J. P. Stephan, I. Mellman, SUV420H2 is an epigenetic regulator of epithelial/mesenchymal states in pancreatic cancer. *J Cell Biol* **217**, 763-777 (2018).

107. Y. Wu, W. Shi, T. Tang, Y. Wang, X. Yin, Y. Chen, Y. Zhang, Y. Xing, Y. Shen, T. Xia, miR-29a contributes to breast cancer cells epithelial–mesenchymal transition, migration, and invasion via down-regulating histone H4K20 trimethylation through directly targeting SUV420H2. *Cell death & disease* **10**, 1-16 (2019).

108. A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, A. A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles. *Nature methods* **12**, 453-457 (2015).

109. G. T. Consortium, The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585 (2013).

110. F. Pontén, K. Jirström, M. Uhlen, The Human Protein Atlas—a tool for pathology. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* **216**, 387-393 (2008).

111. D. M. Nelson, F. Jaber-Hijazi, J. J. Cole, N. A. Robertson, J. S. Pawlikowski, K. T. Norris, S. W. Criscione, N. A. Pchelintsev, D. Piscitello, N. Stong, T. S. Rai, T. McBryan, G. L. Otte, C. Nixon, W. Clark, H. Riethman, H. Wu, G. Schotta, B. A. Garcia, N. Neretti, D. M. Baird, S. L.

Berger, P. D. Adams, Mapping H4K20me3 onto the chromatin landscape of senescent cells indicates a function in control of cell senescence and tumor suppression through preservation of genetic and epigenetic stability. *Genome Biol* **17**, 158 (2016).

112. Y. Yokoyama, A. Matsumoto, M. Hieda, Y. Shinchi, E. Ogihara, M. Hamada, Y. Nishioka, H. Kimura, K. Yoshidome, M. Tsujimoto, N. Matsuura, Loss of histone H4K20 trimethylation predicts poor prognosis in breast cancer and is associated with invasive activity. *Breast Cancer Res* **16**, R66 (2014).

113. M. T. Guinn, G. Balázsi, Noise-reducing optogenetic negative-feedback gene circuits in human cells. *Nucleic acids research* **47**, 7703-7714 (2019).

114. M. Fang, L. Hutchinson, A. Deng, M. R. Green, Common BRAF (V600E)-directed pathway mediates widespread epigenetic silencing in colorectal cancer and melanoma. *Proceedings of the National Academy of Sciences* **113**, 1250-1255 (2016).

115. M. Fang, J. Ou, L. Hutchinson, M. R. Green, The BRAF oncoprotein functions through the transcriptional repressor MAFG to mediate the CpG Island Methylator phenotype. *Molecular cell* **55**, 904-915 (2014).

116. I. Tirosh, B. Izar, S. M. Prakadan, M. H. Wadsworth, D. Treacy, J. J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196 (2016).

117. J. Wu, Y. Wang, X. Xu, H. Cao, S. Sahengbieke, H. Sheng, Q. Huang, M. Lai, Transcriptional activation of FN1 and IL11 by HMGA2 promotes the malignant behavior of colorectal cancer. *Carcinogenesis* **37**, 511-521 (2016).

118. S. M. Weiler, T. Lutz, M. Bissinger, C. Sticht, M. Knaub, N. Gretz, P. Schirmacher, K. Breuhahn, TAZ target gene ITGAV regulates invasion and feeds back positively on YAP and TAZ in liver cancer cells. *Cancer Letters* **473**, 164-175 (2020).

119. E. B. Pasquale, Eph receptor signalling casts a wide net on cell behaviour. *Nature reviews Molecular cell biology* **6**, 462-475 (2005).

120. R. Foreman, R. Wollman, Mammalian gene expression variability is explained by underlying cell state. *Molecular systems biology* **16**, e9146 (2020).

121. S. Patange, D. A. Ball, Y. Wan, T. S. Karpova, M. Girvan, D. Levens, D. R. Larson, MYC amplifies gene expression through global changes in transcription factor dynamics. *Cell reports* **38**, 110292 (2022).

122. H. Gonzalez, C. Hagerling, Z. Werb, Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes & development* **32**, 1267-1284 (2018).

123. X. Cao, Y. Chen, B. Wu, X. Wang, H. Xue, L. Yu, J. Li, Y. Wang, W. Wang, Q. Xu, Histone H4K20 demethylation by two hHR23 proteins. *Cell Reports* **30**, 4152-4164. e4156 (2020).

124. K. Igarashi, H. Nishizawa, Y. Saiki, M. Matsumoto, The transcription factor BACH1 at the crossroads of cancer biology: From epithelial–mesenchymal transition to ferroptosis. *Journal of Biological Chemistry* **297**, (2021).

125. X. Qi, T.-G. Li, J. Hao, J. Hu, J. Wang, H. Simmons, S. Miura, Y. Mishina, G.-Q. Zhao, BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. *Proceedings of the National Academy of Sciences* **101**, 6027-6032 (2004).

126. S. Satoh, Y. Daigo, Y. Furukawa, T. Kato, N. Miwa, T. Nishiwaki, T. Kawasoe, H. Ishiguro, M. Fujita, T. Tokino, AXIN1 mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of AXIN1. *Nature genetics* **24**, 245-250 (2000).

127. X. Zhang, A. S. Farrell, C. J. Daniel, H. Arnold, C. Scanlan, B. J. Laraway, M. Janghorban, L. Lum, D. Chen, M. Troxell, Mechanistic insight into Myc stabilization in breast cancer involving aberrant Axin1 expression. *Proceedings of the National Academy of Sciences* **109**, 2790-2795 (2012).

128. G. L. Coles, S. Cristea, J. T. Webber, R. S. Levin, S. M. Moss, A. He, J. Sangodkar, Y. C. Hwang, J. Arand, A. P. Drainas, Unbiased proteomic profiling uncovers a targetable GNAS/PKA/PP2A axis in small cell lung cancer stem cells. *Cancer Cell* **38**, 129-143. e127 (2020).

129. A. Singh, B. Razooky, C. D. Cox, M. L. Simpson, L. S. Weinberger, Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. *Biophysical journal* **98**, L32-L34 (2010).