UNIVERSITY OF CALIFORNIA,
IRVINE

Molecular and metabolic hallmarks of metastasis in triple negative breast cancer

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Sciences

by

Ryan Tevia Davis

Dissertation Committee:
Assistant Professor Devon A. Lawson, Chair
Associate Professor Francesco Tombola
Assistant Professor Sergio Armando Villalta
Professor David A. Fruman
Associate Professor Mei Kong

2020

Dedication

To

my wife,

the one who make me smile,

my parents,

the ones who sacrificed so much,

my friends

the ones who helped me hit level two,

in recognition of their worth

Perseverance

When asked how it felt to have failed 10,000 times
in the process of inventing the light bulb,
Thomas Edison responded:

“I have not failed. I’ve just found 10,000 ways that won’t work”

Table of Contents

Table of Contents iii

List of Figures v

List of Tables vi

List of Acronyms and Symbols vii

Acknowledgements ix

Vita xi

Abstract of the Dissertation xiv

Chapter 1: Introduction 1

1.1 Significance 1

1.2 Innovation 1

1.3 Background 2

1.3.1 Breast cancer prevalence and subtypes 2

1.3.2 Breast cancer and the process of metastasis 3

1.3.3 Intratumoral heterogeneity and metastasis 4

1.3.4 Application of scRNA-seq technologies to cancer research 5

1.3.5 Metabolism and cancer: Warburg and beyond 9

1.4 Project Goal and Key Outcomes 10

Chapter 2: Defining heterogeneity in primary tumor and metastatic cells 12

2.1 Introduction 12

2.2 Materials and Methods 12

2.2.1 Orthotopic transplantation of PDX tumors into immunocompromised mice 12

2.2.2 Tissue harvest and dissociation 13

2.2.3 Flow cytometry 14

2.2.4 Generation of single-cell RNA sequencing data 14

2.2.5 Processing of single-cell RNA sequencing data 15

2.2.6 Dimensionality reduction, cell cluster identification, and differential gene expression analysis 15

2.2.7 Development of logistic regression model for identifying candidate biomarkers 16

2.2.8 Gene scoring 17

2.2.9 Relapse-free survival analysis 17

2.2.10 Breast cancer PAM50 subtyping 17

2.2.11 Histology 17

2.2.12 Fluorescent in situ hybridization (FISH) 18

2.3 Results 18

2.3.1 Single-cell RNA sequencing of matched primary tumors and micrometastases 18

2.3.2 Transcriptional diversity in micrometastatic and primary tumor cells 20

2.3.3 Breast cancer subtype analysis of primary tumors and metastatic cells 25

2.3.4 Micrometastatic cells display a distinct transcriptome program 29

2.4 Discussion 36

Chapter 3: Metabolic phenotype 40

3.1 Introduction 40

3.2 Materials and Methods 40

3.2.1 Seahorse Assay 40

3.2.2 Proliferation Assay 41
3.2.3 Viability Assay 42
3.2.4 FLIM Acquisition 42
3.2.5 FLIM Data Analysis 43
3.2.6 Orthotopic transplantation and intravenous injection of cell lines into mice 43
3.2.7 Flow cytometry 44
3.2.8 Quantitative real-time PCR 44
3.2.9 Generation of steady-state metabolomics data using LC-HRMS 45
3.3 Results 46
3.3.1 Micrometastatic cells upregulate mitochondrial oxidative phosphorylation 46
3.3.2 Micrometastatic cells display a distinct metabolic profile 52
3.3.3 Oxidative phosphorylation is critical for lung metastasis 55
3.4 Discussion 59
Chapter 4: Conclusion and Final Remarks 63
References 71

List of Figures

Figure 1.1 Common types of intratumor heterogeneity and its regulation by intrinsic and extrinsic factors....6
Figure 1.2 Genetic and phenotypic properties of metastasis-initiating cells at the single-cell level. 7
Figure 2.1 Single-cell RNA sequencing of micrometastatic and primary tumor cells. 19
Figure 2.2 Quality control and exclusion criteria for single-cell RNA sequencing 21
Figure 2.3 Transcriptional diversity in micrometastatic and primary tumor cells 24
Figure 2.4 Marker and GO term analysis of cell clusters from each PDX model 26
Figure 2.5 Identification of an ECM remodeling and EMT co-correlated population in each PDX model....28
Figure 2.6 Relevance of intratumor heterogeneity on subtype classification 30
Figure 2.7 Micrometastatic cells display a distinct transcriptome program 32
Figure 2.8 Prognostic value of micrometastasis- associated genes in basal-like breast cancer patients.... 34
Figure 3.1 Micrometastatic cells display increased mitochondrial OXPHOS 47
Figure 3.2 Gene scoring identifies OXPHOS and glycolysis as top metabolic pathways differentially expressed between micrometastases and primary tumor cells 50
Figure 3.3 Metastatic cells display a distinct metabolic profile 53
Figure 3.4 Oxidative phosphorylation is critical for lung metastasis 56
Figure 3.5 Oligomycin treatment inhibits OXPHOS in MDA-MB-231 and 4T1-GFP cells....57
Figure 3.6 Model for metabolic shift associated with metastatic seeding in TNBC 60
Figure 4.1 Experimental scheme to test the contribution of CD39/CD73 towards enhanced metastatic potential. 68
Figure 4.2 Experimental schemes to test the contribution of mtROS production towards enhanced metastatic potential 70

List of Tables

Table 1 Primer sequences used for quantitative real-time PCR..... 45

List of Acronyms and Symbols

ATP	Adenosine Triphosphate
AIC	Akaike Information Criterion
CSC	Cancer Stem Cell
CAFs	Cancer-associated Fibroblasts
CTC	Circulating Tumor Cell
CAC	Citric Acid Cycle
cDNA	Complementary Deoxyribonucleic acid
CNV	Copy Number Variation
DCA	Dichloroacetate
DNA	Deoxyribonuleic acid
DMEM	Dulbecco's Modified Eagle Medium
ETC	Electron Transport Chain
EMT	Epithelial-to-Mesenchymal Transition
ER	Estrogen Receptor
EDTA	Ethylenediaminetetraacetic acid
ECAR	Extracellular Acidification Rate
ECM	Extracellular Matrix
FBS	Fetal Bovine Serum
FLIM	Fluorescence Lifetime Imaging
FMO	Fluorescence Minus One
FACS	Fluorescence-Activated Cell Sorting
FISH	Fluorescent in situ Hybridization
FDG	Fluorodeoxyglucose
FDA	Food and Drug Administration
FFPE	Formalin-Fixed Paraffin-Embedded
FSC	Forward Scatter
GO	Gene Ontology
GFP	Green Fluorescent Protein
HBSS	Hank's Balanced Salt Solution
HR	Hazard Ratio
HPLC	High Performance Liquid Chromatography
HEPES	Hydroxyethyl Piperazineethanesulfonic Acid
IRB	Institutional Review Board
IACUC	Institutional Animal Care and Use Committee
i.v.	Intravenous
KM	Kaplan-Meier
KNN	K-Nearest Neighbor
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-HRMS	Liquid Chromatography-High Resolution Mass Spectrometry

MET	Mesenchymal-to-Epithelial Transition
mtDNA	Mitochondrial DNA
mtROS	Mitochondrial Reactive Oxygen Species
NADH	Nicotinamide Adenine Dinucleotide
NOD/SCID	Nonobese Diabetic/Severe Combined Immunodeficiency
NSG	Nonobese Diabetic/Severe Combined Immunodeficiency Gamma
OXPHOS	Oxidative Phosphorylation
OCR	Oxygen Consumption Rate
PLS-DA	Partial Least Squares-Discriminant Analysis
PDX	Patient-Derived xenograft
Penstrep/Ampho B	Penicillin-Streptomycin-Amphotericin B
PPP	Pentose Phosphate Pathway
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PET	Positron Emission Tomography
PCA	Principle Component Analysis
PR	Progesterone Receptor
PI	Propidium Iodide
qPCR	Quantitative Polymerase Chain Reaction
ROS	Reactive Oxygen Species
RBC	Red Blood Cell
RNA	Ribonucleic acid
RNA-seq	RNA Sequencing
ROT/AA	Rotenone/Antimycin A
SSC	Side Scatter
scRNA-Seq	Single Cell RNA Sequencing
scRNA-DeepSeq	Single-cell Deep RNA Sequencing
SNV	Single-Nucleotide Variant
SMART	Switching Mechanism at the 5' End After Reverse Transcription
tSNE	t-Distributed Stochastic Neighbor Embedding
TMRM	Tetramethylrhodamine, methyl ester
TFI	Total Fluorescence Intensity
FCCP	Trifluoromethoxy arbonylcyanide phenylhydrazone
TNBC	Triple-Negative Breast Cancer

Acknowledgements

I would like to thank my advisor, Dr. Devon Lawson, for all your support and guidance during this journey. You helped me grow and develop as a person and as a scientist. You helped to teach me to think critically and explore new ideas. You provided me the opportunity to take on this amazing project and to understand why we research; you connected me with patient advocates and survivors through the Komen Outreach programs and Team Michelle, who will always remain an inspiration for the work that I do.

I want to thank the members of the LawBrock clan. You made this whole process bearable and it was a pleasure to work with you all. Whether it was sweetening the successful experiments or softening the failed ones, I could always count on all of you to lend a helping hand or a critical ear. All of course flavored with a healthy dash of sarcasm. In particular, I want to thank Kerrigan and Nick, the computational corner. I know that we have forged a long-lasting friendship through these years, and I look forward to keeping in touch with you in the years to come.

I want to thank all those outside our lab space: Lupe, Mike, Jenna. You helped in so many ways, whether through making me laugh or just lending an ear to my struggles, I truly appreciate all that you provided. I wish you all the best in your future careers. I also must thank my collaborators in the labs of Dr.'s Mei Kong and Michelle Digman at UC Irvine and Jason Locasale at Duke University, without whom the findings on metabolism in this dissertation would not have been possible: Mari, Ying, Juan, Zhengtao, and Austin. Your expertise and dedication pushed this project into something impactful and meaningful.

I want to thank all the undergraduate students who I had the pleasure to not only train, but be taught by: Anh, Grace, Susan, Madona. Many of you have gone on to promising careers in medicine or graduate training. Though this dissertation has my name in the by line, it was truly a team effort, with a large contribution coming from all of you. The hard work and dedication to long hours in the surgery suite or at the flow cytometer is a testament to the quality of scientists that you all will become.

I thank my undergraduate mentor, Dr. Yongsheng Shi, for opening my eyes to the world of academic research and starting me on this path. I want to acknowledge the guidance and advice of my committee members: Dr.'s David Fruman, Mei Kong, Armando Villalta, and Francesco Tombola. Thank you for reading this thesis and providing feedback and advice. I truly appreciate the time and insight you gave me.

I want to thank my friends I have known for longer than I can remember: Kyle, Liam, Ortega, Avi, Michael; and new ones I have gained on this journey: Chris, Nicole. When I had a difficult day at the lab or experiments did not go my way, I knew I could always come home, log into Discord, and let all that stress go talking with you. You may never have understood a word of what I said when I described my research, but you listened, and you kept me grounded. Hopefully now that this journey is at a close, you can finally admit that biology is indeed a science.

To my parents, I send my heartfelt thanks for raising and caring for me and allowing me to reach the point where I stand now. I will probably never truly appreciate the sacrifices that you made for my education, but I hope that I was able to make you proud of the person I have become. You never doubted me, always encouraged me in my pursuits, and taught me the importance of hard work and imagination which has guided me in my scientific career. Thank you, mom and dad. To my brother, Connor Davis, for being one of the coolest dudes I will ever know, thank you little bro.

Last, but certainly not least I want to thank my wife, Katelyn Chandler Corey Davis. You have been with me, supporting me, guiding me, encouraging me, and loving me for more than a decade. We went from high school sweethearts to college couple to husband and wife in those

years. You were there when I chose to pursue this Ph.D. You were there when I failed and when I succeeded. You listened, you comforted, and you cared. I can say with certainty that the reason I made it through this journey is because you were always with me. We are a team, and this thesis work is as much the result of your tireless efforts as it is mine. I thank you and I love you.

I thank Springer Nature for permission to use material in Chapter 1, Chapter 2, and Chapter 3, which was originally published in Nature Cell biology. Support for this project was provided by Team Michelle and the Gazzaniga Family. Financial support was also provided by the University of California, Irvine, National Institutes of Health Grant T32CA009054 and an Opportunity Award from the Center for Complex Biological Systems Support Grant GM-076516. Thank you.

Vita

Ryan Tevia Davis

Education

University of California, Irvine School of Medicine Department of Physiology and Biophysics	Doctor of Philosophy, Biomedical Sciences	2015-2020
University of California, Irvine School of Biological Sciences	Bachelor of Science, Biological Sciences	2011-2015

Publications

1. **Davis RT**, Blake K, Ma D, Ishak Gabra MB, Hernandez GA, Phung AT, Yang Y, Maurer D, Lefebvre AEYT, Alshetaiwi H, Xiao Z, Liu J, Locasale JW, Digman MA, Mjolsness E, Kong M, Werb Z, Lawson DA. Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell RNA sequencing. Nat Cell Biol 22, 310–320 (2020).
2. Wang X, Ramos R, Oh JW, Nguyen TK, Liang HY, Scarfone V, Liu Y, Taguchi N, Paolilli KN, Wang X, Wang G, Guerrero-Juarez CF, Jiang S, **Davis RT**, Greenberg EN, Ruiz-Vega R, Jahid S, Vasudeva P, Murad R, Widyastuti LHP, Lee H, McElwee KJ, Lawson DA, Ganesan A, Anderson B, Mortazavi A, Kunisada T, Li J, Plikus MV, Signaling by senescent cells hyper-activates skin stem cell niche, (in revision, Cell)
3. Hachey SJ, Movsesyan S, Nguyen Q, Burton-Sojo G, Tankanzyan A, Wu J, Hoang T, Hatch M, Zhao D, Celaya E, Gomez S, Chen G, **Davis RT**, Nee K, Pervolarakis N, Lawson DA, Kessenbrock K, Lee A, Waterman M, Hughes CCW, An in vitro human vascularized micro-tumor model of colorectal cancer; (in review, Nat Comm)
4. Lawson DA, Kessenbrock K, **Davis RT**, Pervolarakis N, Werb Z, Tumor heterogeneity and metastasis at single cell resolution, Nat Cell Bio, 20, no. 12 (2018): 1349.
5. Nguyen Q, Pervolarakis N, Blake K, Ma D, **Davis RT**, Phung AT, Willey E, Kumar R, Jabart E, Driver I, Rock J, Goga A, Lawson DA, Werb Z, and Kessenbrock K., Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity, Nat Commun., 9, no. 1 (2018): 2028.
6. **Davis RT**, Shi Y, The polyadenylation code: a unified model for the regulation of mRNA alternative polyadenylation, J Zhejiang Univ Sci B, 2014 May; 15(5): 429–437

Funding and Awards

Gazzaniga Family Medical Research Award ($8,000)	*2019-2020*
Team Michelle and Friends Research Support ($40,000)	*2019-2020*
F99/K00 Predoctoral to Postdoctoral Transition Award – University Nominee	*2019*
Team Michelle and Friends Research Support ($40,000)	*2018-2019*
T32 Cancer Biology and Therapeutics Training Grant ($66,000, Matched Funds)	*2017-2019*
Center for Complex Biological Systems (CCBS) Opportunity Award ($10,000)	*2017-2018*
Summer Undergraduate Research Program (SURP) Fellowship ($3000)	*2014-2015*

Selected Presentations

***Oral**:*

"Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell RNA sequencing" — *2020*
Keystone Symposia on Tumor Metabolism, Banff, Canada (Conference Cancelled)

"Defining the metastatic cell: analyzing intratumor heterogeneity at single cell resolution" — *2018*
CZI Human Cell Atlas Science Retreat, Asilomar, CA

"Investigating gene regulatory networks in metastasis" — *2018*
Center for Complex Biological Systems Annual Retreat, Los Angeles, CA

***Poster**:*

"Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell RNA sequencing" — *2020*
Keystone Symposia on Tumor Metabolism, Banff, Canada (Conference Cancelled)

"Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell RNA sequencing" — *2019*
Cancer Biology Training Consortium Annual Retreat, Baltimore, MD

"Defining the metastatic signature of basal-like breast cancer utilizing single cell transcriptomics" — *2018*
UCI Campus-wide Cancer Research Symposium, Irvine, CA

"Defining the metastatic signature of basal-like breast cancer utilizing single cell transcriptomics" — *2017*
Southern California Systems Biology Conference, Irvine, CA

"Metastatic profile of triple negative breast cancer revealed by single cell RNA-seq" — *2016*
Comprehensive Cancer Center Annual Scientific Meeting, Pasadena, CA

"Identification and characterization of cleavage and polyadenylation specific factor 4-like" — *2015*
UCI Undergraduate Research Symposium, Irvine, CA

Teaching, Mentorship, and Leadership

Student Mentees:

Madona Masoud (Undergraduate Student) — *2017-2019*
Currently – Research Assistant, University of Connecticut

Susan Wassie (Undergraduate Student) — *2017*
Currently – Master's Program in Public Health

Grace Hernandez (Undergraduate Student) — *2016-2018*
Currently – Lab Manager, University of California, Irvine

Anh T. Phung (Undergraduate Student) — *2015-2017*
Currently – Medical Student, Dartmouth

Regis Lee (Undergraduate Student) — *2015-2017*
Currently – Medical Student, University of Utah

Teaching Assistant:

Mol Bio 217 B: Cancer Biology II — *2019*
Taught lecture on single cell sequencing technologies and emerging analysis software with emphasis on application for studying tumor heterogeneity.

Cancer Systems Biology Short Course (R25 Funded) — *2018*
Assisted in leading technical lab section on generating single cell sequencing libraries using the 10X Genomics Chromium platform.

Leadership Roles:

Department Seminar Coordinator *2017-2018*

Organized and moderated the department research-in-progress seminars for graduate students and post-doctoral students.

Abstract of the Dissertation

Molecular and metabolic hallmarks of metastasis in triple negative breast cancer

By

Ryan Tevia Davis

Doctor of Philosophy in Biomedical Sciences

University of California, Irvine, 2020

Assistant Professor Devon A. Lawson, Chair

Although metastasis remains the cause of most cancer-related mortality, mechanisms governing seeding in distal tissues are poorly understood. I hypothesized the initial stages of seeding at distal tissues (micrometastases) contain a unique transcriptome compared to primary tumor cells, enabling the identification of novel molecular targets to specifically block micrometastatic establishment. Here I established a robust method for identification of global transcriptomic changes in rare metastatic cells during seeding using single-cell RNA-sequencing and patient-derived xenograft (PDX) models of breast cancer. I found that both primary tumors and micrometastases display transcriptional heterogeneity, but micrometastases harbor a distinct transcriptome program conserved across PDX models that is highly predictive of poor survival in patients. Pathway analysis revealed mitochondrial oxidative phosphorylation (OXPHOS) as the top pathway upregulated in micrometastases, in contrast to higher levels of glycolytic enzymes in primary tumor cells, which was corroborated by flow cytometric and metabolomic analyses. Pharmacological inhibition of OXPHOS dramatically attenuated metastatic seeding in the lungs, which demonstrates the functional importance of OXPHOS in metastasis and highlights its potential as a therapeutic target to prevent metastatic spread in breast cancer patients.

Chapter 1: Introduction

1.1 Significance

Metastasis remains the leading cause of cancer related mortality in breast cancer patients. Despite a high prevalence in patients (>30%), experimental and clinical data demonstrates that metastasis is rare, with only a biologically unique population of cells within a tumor being capable of successfully completing the process and developing advanced metastatic outgrowth.[1–3] Preclinical research of early stages of metastasis (local invasion and circulation) and late stages of metastasis (outgrowth of secondary tumors) have been useful in understanding key components of this process.[4] However, robust models which can recapitulate the seeding of circulating tumor cells (CTCs) at distal tissues and the establishment of metastatic foci are lacking. This stage of the metastatic cascade represents a particularly difficult biological bottleneck, wherein most CTCs either are cleared in circulation or the tissue parenchyma or enter a state of prolonged dormancy.[5–7]

1.2 Innovation

The inability to authentically investigate the mechanisms governing the initial seeding and establishment of metastatic foci represents a critical gap in knowledge. My dissertation project has focused on developing a model system to isolate micrometastatic cells and assess their molecular profile with innovative new high-resolution single-cell technologies. I decided to utilize patient-derived xenografts (PDX) as my model system because they maintain many features of the original patients' tumors, including histopathology, clinical markers, global gene expression patterns, hormone responsiveness, and intratumor heterogeneity. They have also been shown to be more predictive than cell line or transgenic models of disease course and drug response in patients.[8–12] Furthermore, given that the study of micrometastatic lesions would necessitate incredibly low cell inputs, I developed a single-cell deep RNA sequencing (scRNA-DeepSeq) approach to facilitate the capture and analysis of rare cells. This combined

approach allowed me to gain new insights into the mechanisms driving metastatic seeding of human patient breast tumor cells in distal tissues.

1.3 Background

1.3.1 Breast cancer prevalence and subtypes

Breast cancer remains the leading diagnosed cancer in women worldwide, with an annual incidence rate of 252,710 cases in the United States alone. Breast cancer also maintains a very favorable prognosis, with a 97.3% chance of a woman surviving five years after diagnosis.[13] This rate is a nearly 39% increase since 1989.[13] This favorable prognostic outcome is due in large part to advances in early stage detection modalities available to patients and in the molecular understanding of breast cancer. This molecular insight has enabled the development of directed therapeutics targeting subsets of the disease more effectively.

Breast cancer is a highly heterogenous disease and advances in gene expression profiling have enabled its classification into five molecularly distinct categories: Luminal A, Luminal B, HER2 enriched, Basal-like/Triple Negative, and normal-like.[14,15] The first two categories, Luminal A and Luminal B, are considered the hormone-positive subsets characterized by expression of the estrogen receptor (ER) and progesterone receptor (PR). The distinction between the subsets arises from the relative expression of Ki67+ cells within the sample, with Luminal B having higher prevalence. Together, these subtypes represent the most diagnosed tumors with the best prognostic outcome, due in large part to targeted molecular therapies including the estrogen antagonist, exemestane, the anti-VEGF antibody, bevaciczumab, and the mTOR inhibitor, everolimus.[16–18]

The HER2-enriched subset of breast cancer arises from the overexpression of the human epidermal growth factor receptor 2 (*ERBB2*/HER2), resulting in over-activation of several growth pathways including the MAPK and PI3K/Akt pathways.[19,20] This subset of tumors often arises from a gene duplication event, rather than an activating mutation, at the *ERBB2* gene locus, resulting in overproduction of the HER2 receptor.[20] Current molecular therapies for this

subset are primarily the anti-HER2 antibody, trastuzumab, though a subset of HER2 enriched tumors do not respond to trastuzumab treatment. This resistance has been associated with *PTEN* loss and *CXCR4* upregulation.[17,21]

Molecular profiles of basal-like breast tumors have a high degree of overlap with the histologically classified triple-negative breast cancer (TNBC, characterized by a lack of the ER, PR, and HER2 receptors). In addition, these tumors are often characterized by expression of basal epithelial markers, such as KRT5 and KRT14.[15,22] Of all the molecular subtypes, basal-like cancers represent the most clinically aggressive subtype, with a high incidence of local and regional recurrence following treatment.[23] Owing to the lack of targetable molecular aberrations, such as the HER2 receptor or HRs, the primary therapeutic intervention is chemotherapy, such as taxane or anthracycline.[24] More recently, the TNBC subtype has been shown to contain even further heterogeneity in tumor types, with a distinction between the basal-like, the mesenchymal, immune-enriched, and luminal AR, indicating a strong need to develop new markers and targetable therapies that are specialized to the wide range of molecular subtypes in this classification.[25]

1.3.2 Breast cancer and the process of metastasis

Among a cohort of 1,608 patients in Toronto, TNBC was found to have the highest risk of distant metastasis (HR 4.0, CI 2.7-5.9) within the first five years to diagnosis.[26] Breast cancer metastasis remains a complex clinical problem with regards to the development of reliable biomarkers and effective treatments. Two predominate models exist to explain the process of metastasis in breast cancer: the “linear progression” model and the “parallel progression” model. In the linear progression model, the primary tumor undergoes a series of genetic, epigenetic, and transcriptional alterations which develops a clonal selection process to favor a subset of clones which are capable of undergoing metastasis. This model was supported by initial studies in primary tissue samples which identified conserved gene expression profiles in metastatic tissues and the primary tumor and the relationship between tumor size and

metastatic propensity.[14,27,28] The linear progression model implicitly assumes metastasis to be a "late" process in tumor development, as the clonal selection process requires a series of growth and death events to purify the metastatic subclone.

In contrast, the "parallel" progression model postulates dissemination occurs from a large subset of cells in the primary tumor during early tumorigenesis. This model is supported by several studies indicating a high level of discordance between the gene expression of disseminated tumor cells in the lymph nodes and in the primary tumor with paired patient samples.[29] These studies, in contrast to those examining gene expression in support of the linear progression model, were primarily performed in mouse models, which could explain part of the disparity. There is, however, additional evidence in the form of epidemiological studies which indicate that metastases began to develop 5-7 years before primary tumor diagnosis in a cohort of 12,000 breast cancer patients and clinical evidence of the presence of metastatic disease without the presence of initial tumor diagnosis.[30,31] Evidence also suggests that these disseminated tumor cells can enter a state of quiescence which enables long-term survival at distant sites, which may explain the recurrence seen in patients with early stage disease following surgical resection.[7,32] The parallel progression model, due to dissemination of tumor cells prior to initial tumor diagnosis, requires the treatment of the disease from a systemic standpoint. As such, future therapeutic interventions may necessitate the treatment of early stage tumors with systemic agents to attempt to limit the reoccurrence of dormant metastatic cells shed during early tumorigenesis.

1.3.3 Intratumoral heterogeneity and metastasis

Heterogeneity is pervasive in human cancer and manifests as morphological differences between cells or distinct karyotypic patterns, protein and biomarker expression levels and genetic profiles.[33,34] Tumors are complex ecosystems of malignant cells surrounded by non-malignant stroma, including fibroblasts, endothelial cells and infiltrating immune cells.[35–37] Intratumor heterogeneity arises through various mechanisms **(Fig. 1.1).** Heterogeneity is also

generated through cellular differentiation. In cancer stem cell (CSC) models, cancers are hierarchically organized with a stem cell-like population, sustaining tumor growth through self-renewal and differentiation.[38] The tumor microenvironment also generates intratumor heterogeneity by exerting different selective pressures in distinct regions of the tumor.[39–42] These models are not mutually exclusive and act together to create a complex system with multiple layers of heterogeneity established by the distinct genetic, epigenetic, transcriptomic, proteomic and functional properties of different cells.

In the context of understanding the transcriptome differences between primary tumor cells metastatic cells, this level of intratumoral heterogeneity presents several problems. Primarily, the difficulty exists in the inability to distinguish subsets of tumor cell states from each other during bulk sequencing studies. The general dogma is that metastasis is carried out by rare cells with unique cellular and molecular properties.[32] Single-cell investigations now enable the identification and characterization of such cells, including their localization in primary tumors, and the effect of genetic versus non-genetic and intrinsic versus extrinsic factors on metastasis **(Fig. 1.2).** Single-cell multiplex qPCR technology has shown that metastasis is initiated by cells with stem cell and epithelial–to-mesenchymal transition (EMT)-like characteristics in PDX models of breast cancer.[2] This is consistent with results from the MMTV-PyMT breast cancer mouse model showing CSCs as the origin of metastasis[1], as well as reports implicating stem cell and EMT programs in other breast cancer models.[7,43] A recent scRNA-seq study of human head-and-neck cancers further implicated EMT in metastasis.[44] As such, the application of this technology to the development of micrometastasis will provide new insights into the key factors governing this critical state.

1.3.4 Application of scRNA-seq technologies to cancer research

Single-cell technologies have advanced rapidly in the past several years. Currently available protocols vary in cell capture method, library preparation chemistry and

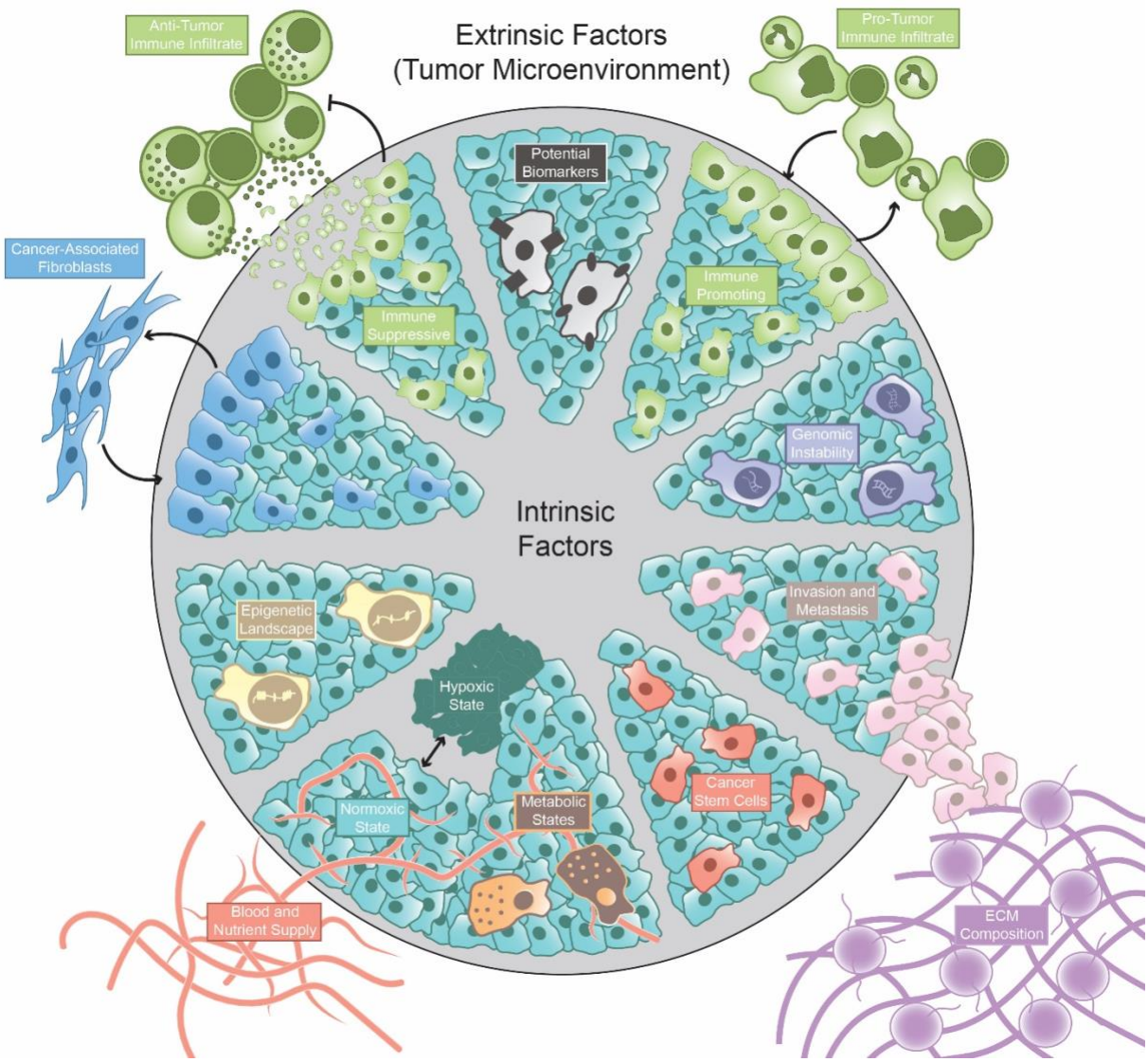

Figure 1.1 Common types of intratumor heterogeneity and its regulation by intrinsic and extrinsic factors. Tumors comprise a heterogeneous population of cells, which is regulated by both intrinsic and extrinsic factors. Tumor cells vary in biomarker expression, epigenetic landscape, hypoxic state, metabolic state, stage of differentiation, invasive potential, and genotype due to genomic instability. The tumor microenvironment can also be heterogeneous, in which different types of fibroblasts, pro-tumor and anti-tumor immune infiltrate, vascular and lymphatic vessel density and extracellular matrix (ECM) composition affect tumor cell heterogeneity and function. (From Lawson, et al. 2018, *Nat Cell Bio)*

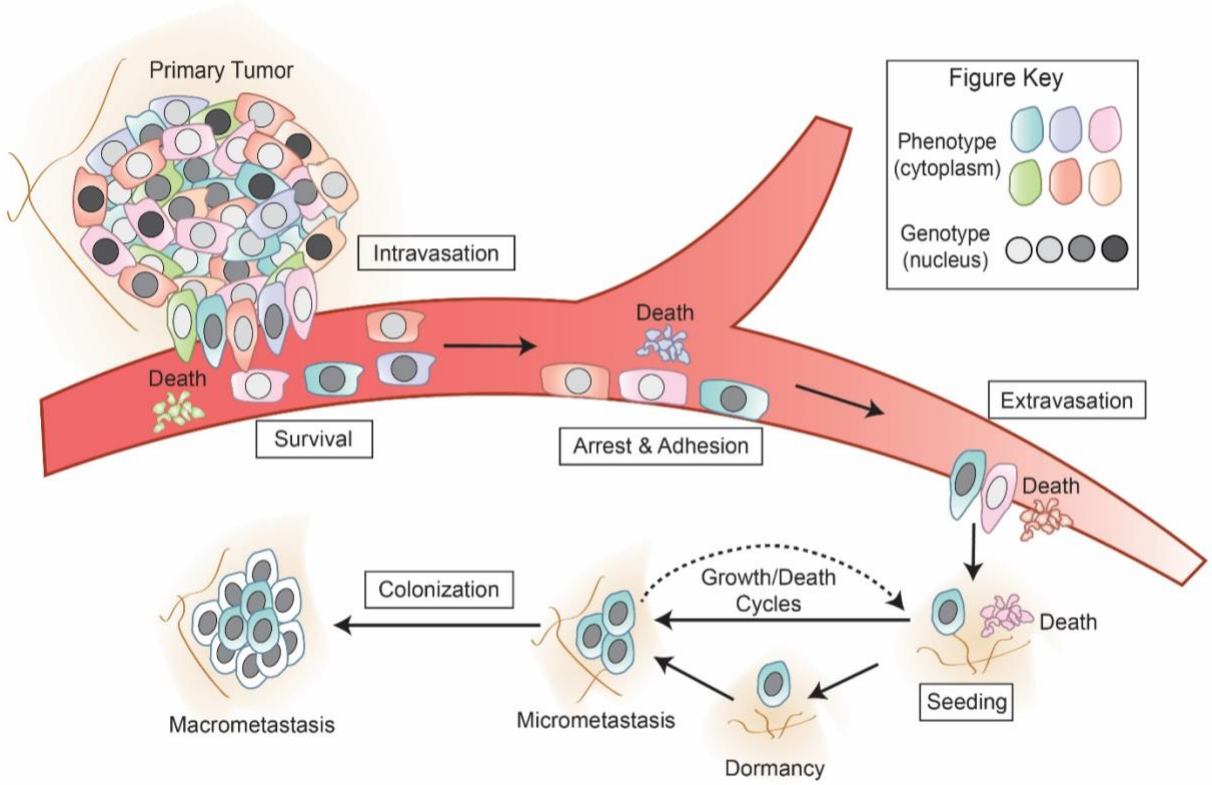

Figure 1.2 Genetic and phenotypic properties of metastasis-initiating cells at the single-cell level. Metastasis is a rare event, in which most cancer cells cannot progress through major bottlenecks associated with invasion, intravasation, extravasation, seeding and colonization to produce a malignant macrometastatic tumor. In this model, cancer cells are heterogeneous in genotype (nuclei) and phenotype (cytoplasm), and metastasis-initiating cells possess a distinct combination of both. Dashed arrow indicates that cancer cells within micrometastases can die. Death rates within micrometastases can balance proliferation rates, and thereby prevent progression to macrometastasis by the failure to produce net growth. (From Lawson, et al. 2018, *Nat Cell Bio*)

throughput.[45–48] Most protocols require single-cell suspensions, so the first critical consideration is optimizing tumor dissociation to generate a cell suspension that is fully representative of the intact tumor in terms of cell populations, their frequencies and expression programs. Digestion of solid tumors eliminates spatial information and can obscure the true program of individual cells.[49,50] Although there is no consensus for how to measure these profound effects, cellular diversity after dissociation can be analyzed by flow cytometry for known cell types or markers for the specific tumor type. Populations are typically validated by follow-up analyses *in situ*, but this approach only confirms their existence and does not determine whether all cell populations in the tumor were accounted for after cell dissociation. Ultimately, identification of the same populations using different protocols would increase confidence in the results.

Technologies for transcriptome analysis are the most advanced and have been used to profile CSCs, map differentiation trajectories, describe drug resistance programs and define the immune infiltrate in tumors.[51–53] The first single-cell transcriptome technologies utilized microfluidics to capture cells followed by multiplex qPCR for selected genes, but most recent studies favor scRNA-seq to enable assessment of the entire transcriptome. Selection of the most appropriate scRNA-seq protocol depends on the sample size, the number of cells to be sequenced and whether transcript counting, or full-length mRNA sequencing is desired. When large cell numbers must be sequenced, high-throughput, semi-automated droplet-based approaches (for example, inDrop[54] and Dropseq[55]) are optimal. However, these approaches typically achieve lower transcriptome coverage and detect fewer lowly expressed genes, and there is very limited transcript sequence coverage due to 3′ end counting, precluding single-nucleotide variant (SNV) and splicing analyses.[50] Droplet-based protocols are also less amenable to studying small cell numbers, for instance CTCs, disseminated tumor cells or micrometastases. For such samples, cell isolation by flow cytometry or micromanipulation followed by manual library preparation in microwell plates is more tractable. These protocols typically amplify full-length mRNA by switching mechanism at 5′ end of RNA template (SMART)

or alternative chemistries, which enable full-length mRNA sequencing for SNV, splicing and deeper transcriptome analyses.[56–59] For tissues that are preserved or cannot be readily dissociated, single-nucleus RNA-seq approaches, such as DroNc-seq[60] or microwell-based single-nucleus RNA-seq[61], may be optimal.

1.3.5 Metabolism and cancer: Warburg and beyond

Classical concepts of tumor metabolism centralize on the observation that cancer cells continue to preferentially metabolize glucose to lactate through glycolysis, rather than utilize OXPHOS in the presence exogenous oxygen sources. This aerobic glycolysis behavior by cancer cells was initially described by Otto Warburg in the 1930s and thus has been referred to eponymously as the Warburg effect or Warburg metabolism.[62,63] Since its initial discovery, many researchers have utilized this metabolic quirk of cancer cells to their advantage. For instance, increased glucose uptake in tumors can be coopted for tumor detection; using ^{18}fluoro-deoxyglucose (FDG) in positron emission tomography (PET) imaging enables identification of localized tumors in patients.[64] From a therapeutic standpoint, inhibiting glucose import can starve cancer cells[65], inhibiting glucose catabolism during glycolysis can prevent ATP production[66], or inhibiting lactate export can acidify cancer cells resulting in cell death.[67] However, although Warburg's initial observation has resulted in several advances in our understanding of cancer, his key hypothesis that aerobic glycolysis results from impaired respiratory capacity has been challenged. Indeed, mutations or deletions in mitochondrial DNA (mtDNA), actually has deleterious effects on cancer cell viability and function.[68] A prevailing hypothesis, which has strong experimental support, posits that the benefits of Warburg metabolism are not energetic (as OXPHOS metabolism produces far more ATP in cells), but rather biosynthetic.[69] Indeed, tumor cells have been shown to utilize a modified form of pyruvate kinase (M2-PK), which converts phosphoenolpyruvate (PEP) to pyruvate, which is less efficient at this critical step in glycolysis.[70] The resulting bottleneck generated at this final step of glycolysis in enables the tumor cell to instead shift its metabolism towards biosynthetic

pathways upstream of PEP, including nucleotide synthesis, amino acid metabolism, and triacylglyceride synthesis.[69] Furthermore, the increased acidity of the tumor microenvironment results from upregulation of LDHA activity, a critical enzyme which regenerates NAD+ and generates lactic acid, resulting from the shifting away from OXPHOS and CAC activity in tumor cells which typically regenerates the cellular NAD+ levels.[71]

Our understanding of cancer cell metabolism has drastically expanded beyond the Warburg effect, and now includes contribution of multiple additional metabolic pathways upregulated in cancer cells, including the pentose phosphate pathway (PPP)[72], fatty acid oxidation[73,74], and glutamine metabolism.[75–77] In particular glutamine metabolism has been given special focus for its ability to maintain both ATP levels and NADPH for lipid biosynthesis while tumor cells undergo aerobic glycolysis. By engaging multiple biosynthetic pathways, while simultaneously reducing ATP production through OXPHOS, tumor cells place themselves in energetic debt. Glutamine, thus, has been demonstrated to not only provide oxaloacetate for the replenishment of CAC intermediates, but also can be used to generate NADPH to lipid synthesis.[78] In the context of metastasis, several metabolic pathways have been implicated in promoting a pro-invasive, anti-apoptotic phenotype. Increased lactate production from aerobic glycolysis stimulates migration and angiogenesis.[67] The PPP provides the reducing cofactor NADPH to counterbalance elevated intracellular ROS levels[79] and the CAC provides ATP to counterbalance depleted intracellular levels[80], resulting from detachment of metastatic cells from the extracellular matrix and to prevent anoikis. Glutamine metabolism provides inputs to fuel the CAC and the generation of glutathione, further acting to counteract ROS levels and provide additional ATP.[76,81,82]

1.4 Project Goal and Key Outcomes

The primary goal of my thesis project was to elucidate the molecular and metabolic drivers of TNBC metastasis. As described in the preceding sections, a key finding regarding metastatic cells was that they arise from rare, phenotypically distinct cells in the primary tumor. This led to

my hypothesis that micrometastatic cells contain a global transcriptome distinct from signature from the primary tumor. However, bulk sequencing technologies are unable to effectively distinguish rare molecular signatures from the average of the combined malignant and non-malignant tumor profiles that are typically generated. To that end, I turned to single cell sequencing technologies to define the molecular signature of micrometastatic cells in the context of clinically relevant PDX models. This approach afforded two advantages: 1) it enabled me to deconvolve intratumoral heterogeneity in both primary tumor cells and micrometastatic cells and 2) it enabled me to profile incredibly rare seeding colonies of micrometastasis in the lungs and lymph nodes of my PDX models. The major findings from this work, explained in detail in Chapters 2 and 3, include the establishment of a highly specific micrometastatic gene signature with potential prognostic value and the identification of a bioenergetic shift from glycolysis to OXPHOS in micrometastatic cells which can be inhibited pharmacologically to significantly impair metastatic potential.

Portions of this chapter were previously submitted as part of an assignment for completion of Mol Bio 218. These sections have been included with permission from the course instructor.

Reprinted with permission from:

Lawson, D.A., Kessenbrock, K., Davis, R.T. et al. Tumor heterogeneity and metastasis at single-cell resolution. Nat Cell Biol 20, 1349–1360 (2018).

Chapter 2: Defining heterogeneity in primary tumor and metastatic cells

2.1 Introduction

Despite major advances in the detection and treatment of early stage disease, metastasis remains the cause of nearly all mortality associated with breast cancer.[83,84] Previous work suggests that metastasis is seeded by rare primary tumor cells with unique biological properties that enable them to surpass each step in the metastatic cascade.[3,85,86] While properties promoting cell motility and migration are well studied, mechanisms governing the seeding and establishment of micrometastasis in distal tissues are poorly understood. This is in part because metastatic seeding cannot be studied in humans, and because it is technically challenging to detect and analyze rare cells at this transient stage in animal models. Further insights into the mechanisms driving metastatic seeding are critical to inspire new strategies for how to prevent metastatic spread and reduce mortality in breast cancer patients.

I have developed a robust approach for the capture and analysis of individual cancer cells during the seeding of micrometastasis in human PDX models using scRNA-seq technology. This enabled me to study the role of tumor heterogeneity in metastasis, and identify cellular programs upregulated in micrometastatic cells that promote metastatic seeding. I find that micrometastases display a distinct transcriptome program and that this program is predictive of patient outcomes in the context of TNBC.

2.2 Materials and Methods

2.2.1 Orthotopic transplantation of PDX tumors into immunocompromised mice

Samples from patients were generously provided by A.L. Welm in the Department of Oncological Sciences at the Huntsman Cancer Institute (HCI). All tissue samples were collected with informed consent from individuals being treated at the Huntsman Cancer Hospital and the University of Utah under a protocol approved by the University of Utah Institutional Review Board.[87] HCI001 was acquired from a primary tumor biopsy of a Stage IV female patient

diagnosed with ER-PR-Her2-, basal-like invasive ductal carcinoma (IDC) with no previous systemic treatment. HCI002 was acquired from a primary tumor biopsy of a female patient diagnosed with ER-PR-Her2-, basal-like Stage IIIA medullary type IDC with no previous systemic treatment. HCI010 was acquired from a pleural effusion of a Stage IIIC female patient diagnosed with ER-PR-Her2-, basal-like (PAM50) IDC treated with several rounds of chemotherapies.[87] Additional clinical details of each patient tumor can be found in Supplemental Table 1 of ref 83. Samples were collected and de-identified by the Huntsman Cancer Institute Tissue Resource and Application Core facility before being obtained for implantation. The study is compliant with all relevant ethical regulations regarding research involving human participants.

The University of California, Irvine Institutional Animal Care and Use Committee (IACUC) reviewed and approved all animal experiments. Orthotopic transplants of serially passaged human tumor samples were performed on immunocompromised 3-4-week-old NOD/SCID or NSG mice after clearing the mammary fat pads following established protocols.[2] Tumor growth was monitored by weekly calipers and volumes were calculated as: length × $width_2$ × 0.51. Animals were euthanized, and tissues were harvested when tumors reached 2.0-2.5cm by length or width. For experimental metastasis studies, cultured MDA-MB-231 or 4T1-GFP were suspended in 100µL of sterile PBS and injected into the tail vein of 8-10-week-old female NOD/SCID (MDA-MB-231) or BALB/c (4T1-GFP) mice. Mice were euthanized by asphyxiation with CO_2 followed by cervical dislocation and perfusion with 10mM EDTA in D-PBS 21 days (MDA-MB-231) or 5 days (4T1-GFP) post-injection. For cell line orthotopic tumor injections, mice were placed under isoflurane anesthesia (1.5–2.0%). Prior to injection the area was shaved and cleaned with an alcohol swab. $1x10_6$ 4T1-GFP or MDA-MB-231 cells in 100µL PBS were injected bilaterally into the fourth mammary fat pad of 10-12-week-old NSG mice. The study is compliant with all relevant ethical regulations regarding animal research.

2.2.2 Tissue harvest and dissociation

Animals at endpoint were euthanized by asphyxiation with CO_2 followed by cervical dislocation and perfusion with 10mM EDTA in D-PBS. Prior to perfusion, Evan's Blue (Sigma-Aldrich, Cat. No. E2129-10G) was injected into the footpads and ears of anesthetized mice to aid in lymph node visualization. Solid tissues from the mice, which includes the primary tumor, lung, and lymph nodes were processed for flow cytometry by mechanically chopping with blades, followed by Collagenase IV digest (Sigma-Aldrich Cat. No. C5138-1G) in media (DMEM/F12 with 5% FBS, 5µg/mL insulin, and 1% Penstrep/Ampho B) for 45 min at 37°C. Cell suspensions were washed with 2 µg/mL DNAse I (Worthington Biochemical, Cat. No. LS002139) for 5 min and further dissociated with 0.05% Trypsin for 10 min. Following a wash with HBSS/2% FBS, cells were passed through a 70µm filter. Lung and primary tumor cells were treated with 1X RBC lysis buffer, followed by resuspension in DMEM/F12 with 10% FBS for FACS.

2.2.3 Flow cytometry

Single cell suspensions were stained human-specific antibody CD298, diluted 1:100 (PE, BioLegend, Cat. No. 341704) and the mouse-specific antibody MHC-I, diluted 1:150 (APC, ThermoFisher, Scientific Cat. No. 17-5957-80). Flow cytometry was performed using the BD FACSAria Fusion cell sorter. Cell viability was determined by negative staining for SYTOX Blue, diluted 1:1000 (ThermoFisher Scientific Cat. No. S34857). Forward scatter area by forward scatter width (FSC H x FSC A) and side scatter area by side scatter width (SSC H x SSC A) was used to discriminate single cells from doublet and multiplet cells. Mouse cells were excluded by gating out $CD298_{Neg}MHC\text{-}I_{Pos}$. Human primary tumor cells and metastatic cells were selected by gating on $Sytox_{Neg}CD298_{Pos}MHC\text{-}I_{Neg}$.

2.2.4 Generation of single-cell RNA sequencing data

Single cells were sorted directly into each well of a skirted 96-well PCR plate (Fisher Scientific, Eppendorf, Cat. No. E951020443) containing lysis buffer (0.2% Triton X-100 (Sigma-Aldrich, Cat. No. T9284), 2 U/µL RNAseOUT (ThermoFisher Scientific, Cat. No. 10777019),

10µM oligo-dT30VN, and 10µM dNTPs (ThermoFisher Scientific, Cat. No. 18427088) as described previously.[56] The plates were snap frozen on dry ice and stored at -80°C until further processing. Total RNA was converted into cDNA using the SmartSeq2 protocol and prepared for Illumina sequencing with the Nextera XT DNA Library Preparation Kit (Illumina, Cat. No. FC-131-1096). Cells were sequenced at a depth of 1 million reads/cell on the HiSeq2500.

2.2.5 Processing of single-cell RNA sequencing data

Files from the HiSeq2500 were demultiplexed and converted to FASTQ files. Paired-end 100bp reads were aligned to the Gencode 21 human transcriptome using Bowtie2 and quantified using RSEM with the following parameters: rsem-calculate-expression -p $CORES --bowtie2 --paired-end READ1 READ2 gencodehg21. Expression values were log-transformed into log(TPM+1) matrices and loaded into the Seurat analysis package with the following parameters: p10 <- CreateSeuratObject(raw.data = p10.mat, min.cells = 8, min.genes = 1000, project = "HCI010"). As further quality control, we removed any cells identified as visual outliers by library complexity (<2500 genes/cell) or overrepresentation of mitochondrial gene expression (>50%). Additionally, we removed any genes that were not represented in a robust population of cells (<8 cells/gene) from downstream analysis. This resulted in a final analysis of 1,119 single cell profiles. Using the RegressOut feature in Seurat, we calculated z-score residuals using nGene and percent.mito as co-variates, which was used to perform PCA and tSNE. A G1/S and G2/M score was calculated using the gene score method described below and regressed out as well for HCI001 and HCI010.

2.2.6 Dimensionality reduction, cell cluster identification, and differential gene expression analysis

Dimensionality reduction and differential gene expression was performed using the Seurat analysis package version 2.1.0.[88,89] For the combined and individual patient analysis, highly variable genes in our dataset were identified using the MeanVarPlot function with the following parameters: FindVariableGenes(object = comb, mean.function = ExpMean,

dispersion.function = LogVMR, x.low.cutoff = 0.0125, x.high.cutoff = 3, y.cutoff = 0.5). These variable genes were then used for PCA analysis. The PCs generated were then used to perform t-distributed stochastic neighbor embedding (tSNE) of the data. For the individual patient analysis, using the FindClusters function in Seurat and a granularity parameter of 1.0, I identified distinct subpopulations and defined marker genes for each of them with the FindAllMarkers function in Seurat with the default settings for the FindAllMarkers function and the “bimod” statistical test. For the generation of the 330 micrometastatic gene signature, metastatic cells from all PDX models were grouped together separate from tumor cells and we calculated a DE test in Seurat using the “tobit” test with the following parameters: comb<-FindAllMarkers(object = comb, only.pos = TRUE, min.pct = 0.1, logfc.threshold = 0.25, test.use = "tobit", latent.vars=”orig.ident”). The “orig.ident” command in the “latent.vars” variable represents the Patient ID (i.e. HCI001, etc.). By including this variable, the tobit model identifies conserved marker genes.[88–90] Gene ontology analysis was performed using the Enrichr web resource, where the input gene set for each population was the markers identified by FindAllMarkers.[91] For the analysis of specific OXPHOS and glycolysis genes, I calculated the average log2 fold-change of the raw matrix values using the FetchData(use.raw=T) function in Seurat.

2.2.7 Development of logistic regression model for identifying candidate biomarkers

To calculate the logistic regression model used to identify candidate biomarker targets, gene expressions were normalized across all cells such that each gene had a mean expression of zero and a standard deviation of one. For model fitting, the data was sampled equally ten times from each mouse and cell identity (tumor vs. micrometastasis) category to avoid systematic bias. For each sampling, a stepwise regression with forward selection was performed, where at each step, the model that minimized the Akaike information criterion (AIC) was chosen to be used as a base model for the next step. Our logistic regression model used a

conservative cutoff of five genes based on the elbow in the AIC plot to minimize the size of our gene set while maintaining descriptive power.

2.2.8 Gene scoring

To quantitatively compare gene groups (e.g., Glycolysis, OXPHOS) in our cell populations, I utilized individual gene scores as described previously.[52,92] A curated list of genes was obtained from KEGG, a database used for GO term enrichment. I utilized the AddModuleScore() function from Seurat to calculate gene scores for each population

2.2.9 Relapse-free survival analysis

For the relapse-free survival analysis, Kaplan-Meier (KM) survival curves were generated on basal-like breast cancer patient primary tumor microarray data from the KM plotter database for the top 20 micrometastasis-associated genes in our 330 gene signatures.[93] Two genes were found to be not predictive and data for three genes was not available and were excluded. To generate the combined survival analysis, we calculated a weighted average of the 15 predictive genes analyzed using the “Use Multiple Genes” function in KMPlotter. All KM plots are displayed using the “auto select best cutoff” parameter.

2.2.10 Breast cancer PAM50 subtyping

Classification of the single cell RNAseq data into the PAM50 subtypes was performed using the genefu R package.[94] Briefly, log(TPM+1) expression matrices for each patient were loaded into the genefu package and cells were classified using the molecular.subtyping() function with the following parameters: sbt.model = "pam50",data = p10.dataset,annot = p10.annot,do.mapping = FALSE. Cells with a classification probability <0.7 were removed and the remaining cells were used in further analysis.

2.2.11 Histology

Tumor and lung tissues from the PDX mouse models were fixed overnight in 4% paraformaldehyde, then dehydrated and processed for paraffin embedding in the Leica Tissue Processor using standard protocols. The paraffin blocks were cut into 5-µm thick sections using

the Leica microtome, rehydrated, and then stained with hematoxylin and eosin. Brightfield imaging was done using the BZ-X700 Keyence microscope.

2.2.12 Fluorescent in situ hybridization (FISH)

FISH was performed on formalin fixed paraffin embedded (FFPE) sections using the RNAscope Multiplex Fluorescent Reagent Kit v2 (ACD, Cat. No. 323110) according to the manufacturer's instructions. Briefly, FFPE sections were rehydrated in Histoclear and 100% ethanol, prior to antigen retrieval using the RNAscope antigen retrieval solution and mild boiling at 100°C for 15 minutes. PHLDA2 probe (ACD, Cat. No. 551441) amplification was performed according to manufacturer's instructions with the TSA Plus Cyanine 3 (PerkinElmer, Cat. No. PN NEL744001KT) fluorophore diluted 1:1000, stained with DAPI, and mounted with Prolong Gold. Slides were visualized using the Zeiss LS700 confocal microscope. Image analysis was performed in ImageJ. Normalized total fluorescence intensity of the *PHLDA2* probe was calculated on regions of interest across at least 5 different fields of view on 2 mouse lungs and 3 tumors from HCI001 according to the following equation:

$$Normalized\ TFI = \frac{Fluorescence\ Integrated\ Density}{Area\ of\ Region\ of\ Interest}$$

For quantification of lung micrometastatic regions of interest, surrounding mouse stroma was excluded from the analysis and for tumor regions of interest, necrotic regions or mouse stroma was excluded from the analysis.

2.3 Results

2.3.1 Single-cell RNA sequencing of matched primary tumors and micrometastases

To identify fundamental cellular programs important for seeding in metastatic tissues, I investigated transcriptome programs uniquely expressed by cancer cells during the seeding and establishment of micrometastatic lesions. I utilized breast cancer PDX models because of their human disease authenticity and since they maintain intratumor heterogeneity, which is critical for determining the role of cellular heterogeneity in metastasis.[8,9,12,95] I analyzed three

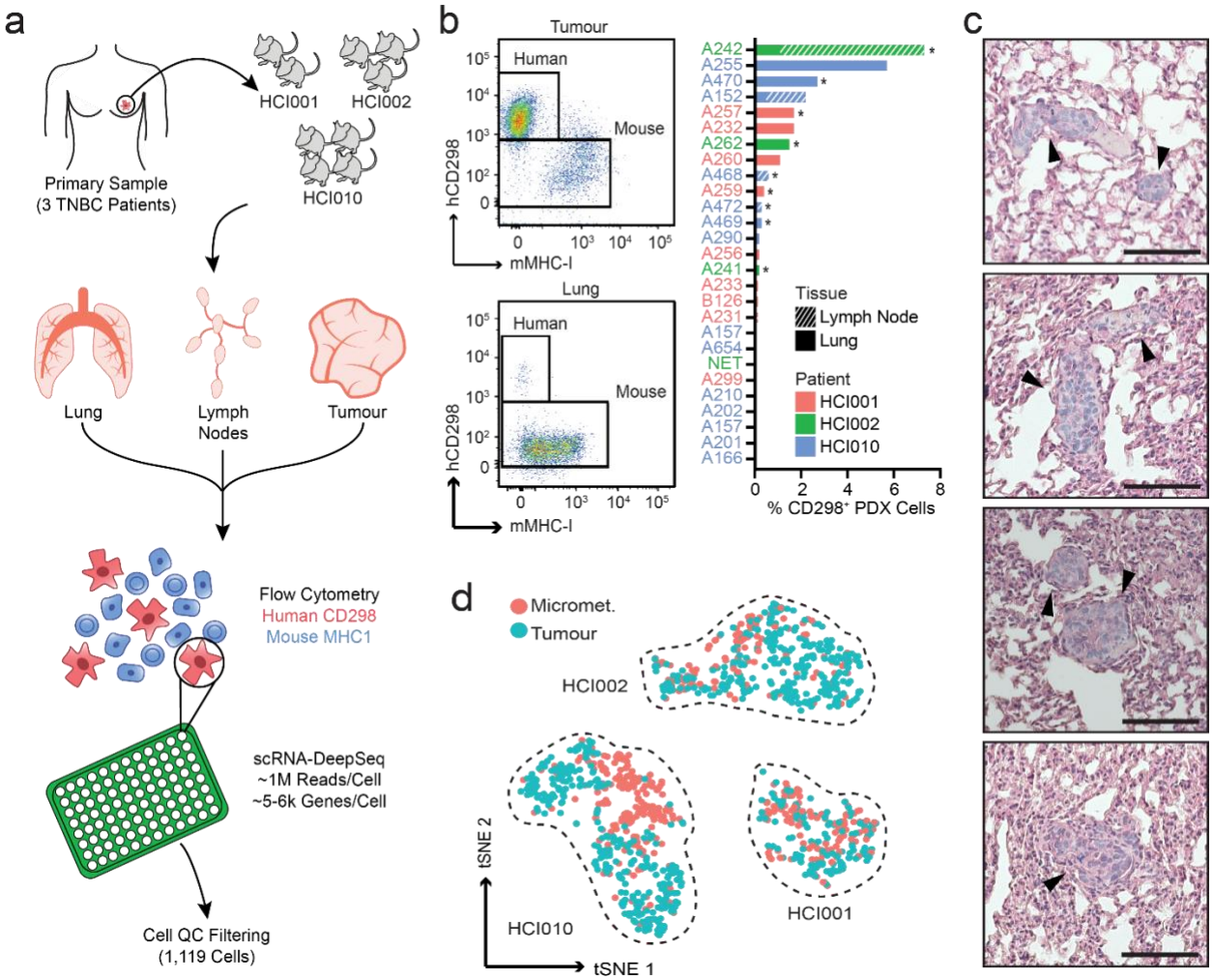

Figure 2.1 Single-cell RNA sequencing of micrometastatic and primary tumor cells. a, Overview of the experimental workflow. The primary tumor, lungs, and lymph nodes of each PDX animal were digested to make single-cell suspensions. Single CD298+MHC-I- human tumor cells were isolated by flow cytometry, deposited into individual wells of 96-well plates and single-cell cDNA libraries were prepared using Smartseq2 chemistry. Matched primary tumor and micrometastatic cells from nine mice and three PDX models (HCI001, HCI002 and HCI010) were analyzed, and 1,119 cells passed quality-control filtering. **b,** Left, flow cytometry-based strategy for the isolation of human CD298+MHC-I- cells from micrometastatic (bottom) and primary tumor (top) cells. Right, frequency of CD298+MHC-I- cells in a cohort of PDX mice. Mice that were selected for sequencing are indicated with an asterisk. **c,** Hematoxylin and eosin stains of micrometastases in the lungs of HCI001 mice. Data are representative of $n = 8$ biologically independent samples. Micrometastatic lesions are indicated by the arrows. Scale bars, 100 µm. **d,** T-distributed stochastic neighbor embedding (tSNE) plot showing clustering of 1,119 metastatic and primary tumor cells from the HCI001, HCI002 and HCI010 models.

previously established PDX models of TNBC: HCI001, HCI002, and HCI010 **(Fig. 2.1a).**[8] Like in many breast cancer patients, metastatic progression is slow and sporadic in these models, where most animals display dispersed micrometastases in the lung and lymph nodes and very low metastatic burden at endpoint (20-25 mm primary tumor diameter) **(Fig. 2.1b-c).**[2,8] This enabled me to investigate transcriptional changes associated with early events in the seeding and establishment of micrometastasis.

We previously developed a robust protocol for the isolation of metastatic cells from PDX models by flow cytometry using human (CD298) and mouse (MHC-I) species-specific antibodies **(Fig. 2.1a-b).**[2] Here, I used this approach to compare cellular diversity in primary tumors and micrometastases by scRNA-seq. I sorted individual cancer cells from the lungs, lymph nodes, and primary tumors of PDX mice into 96-well PCR plates **(Fig. 2.1a-b)**. Matched metastatic and primary tumor cells were isolated from each animal. I optimized our scRNA-seq protocol specifically for flow cytometry-sorted cancer cells.[2,56] Single-cell libraries were constructed using the Illumina platform and sequenced at high depth (1 million paired end reads/cell).

In contrast to high-throughput platforms (e.g., droplet-based), my scRNA-DeepSeq protocol enabled me to capture rare metastatic cells and detect more genes per cell.[87] In total, I sequenced 1,707 tumor and metastatic cells from nine PDX mice and three tumor models **(Fig. 2.1a-b)**. Cell libraries were filtered to remove low quality cells using stringent parameters for low library complexity, mitochondrial gene overrepresentation, and noisy gene expression to ensure technical artifacts did not impact downstream analysis **(Fig. 2.2a-c)**.[88,89] 1,119 cells (66%) passed filtering and were carried forward for further analysis. Importantly, all patient models exhibited similar library quality **(Fig. 2.2a-c)**.

2.3.2 Transcriptional diversity in micrometastatic and primary tumor cells

To investigate intra- and intertumoral diversity, I performed K-nearest neighbor (KNN) graph-based clustering, followed by Louvain modularity optimization utilizing the Seurat analysis package.[88,89] This revealed substantial intertumoral heterogeneity, which is consistent with prior

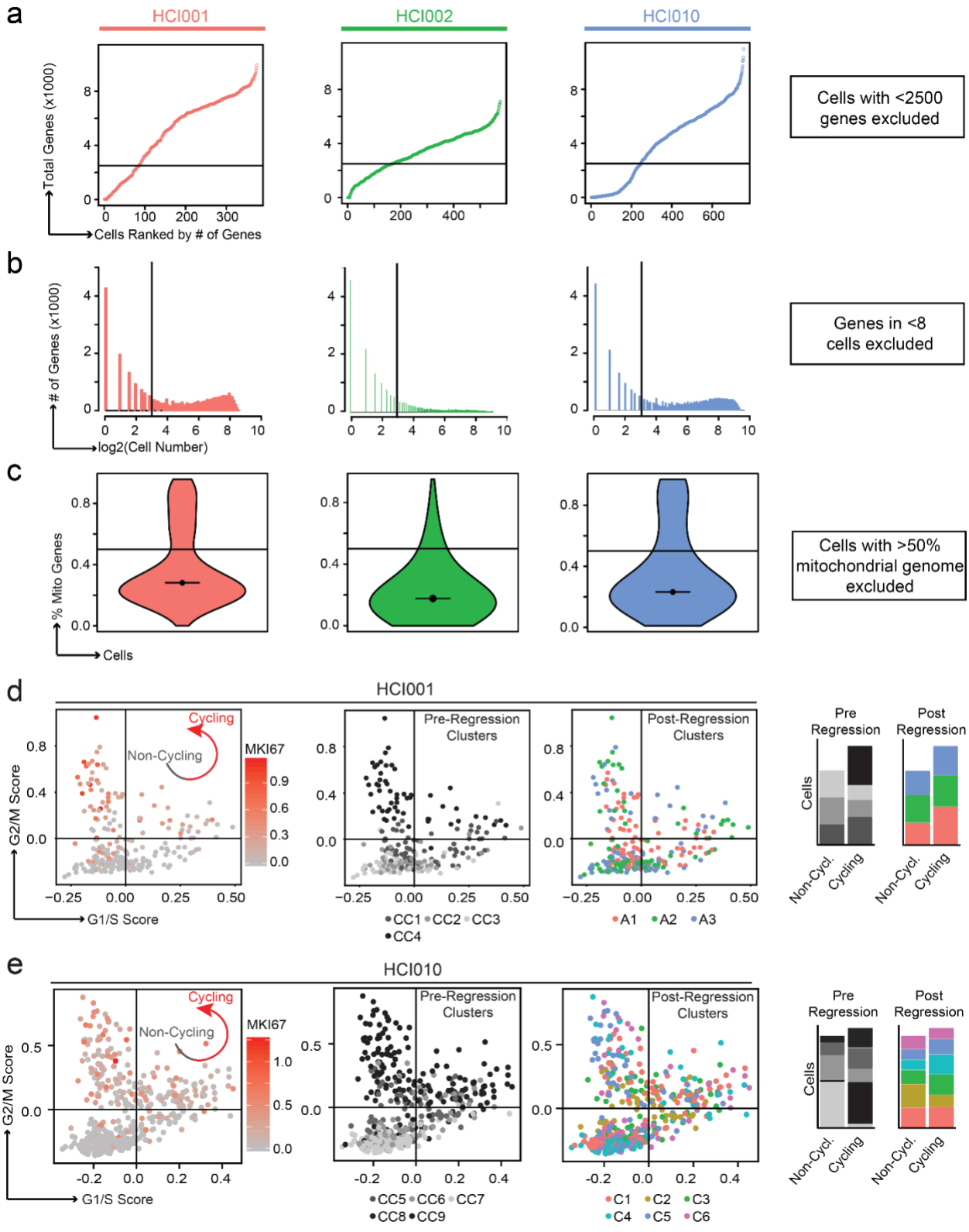
a
HCI001
HCI002
HCI010
Total Genes (x1000)
Cells Ranked by # of Genes
Cells with <2500 genes excluded
b
# of Genes (x1000)
log2(Cell Number)
Genes in <8 cells excluded
c
% Mito Genes
Cells
Cells with >50% mitochondrial genome excluded
d
HCI001
Cycling
Non-Cycling
MKI67
G2/M Score
G1/S Score
Pre-Regression Clusters
Post-Regression Clusters
CC1 CC2 CC3 CC4
A1 A2 A3
Pre Regression
Post Regression
Cells
Non-Cycl.
Cycling
e
HCI010
CC5 CC6 CC7 CC8 CC9
C1 C2 C3 C4 C5 C6

Figure 2.2 Quality control and exclusion criteria for single-cell RNA sequencing. a, Identification and removal of poor-quality cell libraries. Plots show the number of genes detected in each cell from each PDX model. Cells (x axis) are ordered from fewest to most genes detected. Cells with fewer than 2500 genes detected (horizontal line) were excluded. **b,** Identification and removal of noisy/background genes. Plots show the distribution of genes detected per cell. Vertical line indicates that genes detected in fewer than 8 cells were excluded. **c,** Violin plots indicate the expression of mitochondrial genes as a percentage of total gene expression in each cell separated by HCI001 (n = 375 cells), HCI002 (n = 576 cells), and HCI010 (n = 756 cells). Cells were excluded if > 50% of genes detected were mitochondrial (horizontal line). Black line with dot indicates median. **d, e,** Cells from HCI001 (**d**) and HCI010 (**e**) are plotted based on the relative expression of gene sets associated with G1/S (x-axis) and G2/M (y- axis) stages of the cell cycle. Left plots: Relative expression of the proliferation-associated gene *MKI67* is shown in HCI001 and HCI010. Middle plots: Cell clustering before cell cycle regression. Cluster identities are shown in grey. Right plots: Cell clustering after cell cycle regression. Colors indicate new cluster identity of each cell and correspond with clusters shown in Fig. 2.3a. Bar graphs show quantification of cell clusters by cycling status before and after regression.

bulk and single-cell RNA sequencing studies of malignant cells **(Fig. 2.1d)**.[8,44,52,53,92,96] I subsequently analyzed each model separately to explore intratumoral heterogeneity. Initial clustering in the HCI010 and HCI001 models was primarily driven by cell cycle status **(Fig. 2.2d-e)**. I consequently regressed out the effects of cell cycle using a previously established method, which infers cell cycle stage by calculating scores for each cell based on its expression of genes associated with G1/S or G2/M (**Fig. 2.2d-e)**.[52,92] After regressing out the effects of cell cycle, mouse-to-mouse variation within each model was limited, suggesting minimal batch effects **(Fig. 2.4a).** Clustering analysis identified three distinct cell populations in HCI001 (A1-A3), five in HCI002 (B1-B5), and six in HCI010 (C1-C6) **(Fig. 2.3a)**. Interestingly, metastatic and primary tumor cells were found in all clusters, suggesting that metastatic cells display substantial transcriptional diversity even at early stages of micrometastasis **(Fig. 2.3b, Fig. 2.4b)**. However, clusters A1, B1, and C2-C3 from each model contained higher proportions of metastatic cells relative to the other clusters, suggesting a skewing towards this cell state in micrometastasis **(Fig. 2.3b).**

To identify cellular properties characteristic of each population, I performed marker gene identification and Gene Ontology (GO) analyses using Seurat and Enrichr.[88,89,91,97,98] This revealed distinct markers and pathways associated with each cell cluster **(Fig. 2.3c, Fig. 2.4c)**. For example, A2 in HCI001 contained a population of cells that expressed high levels of genes associated with fatty acid metabolism, such as *BDH2*, *DECR1*, and *CAV1* **(Fig. 2.4c-d)**. B1 in HCI002 expressed high levels of genes encoding proteasome function **(Fig. 2.4c-d).** The EMT clusters (A3, B5, and C4) displayed increased levels of classic EMT genes, such as *ZEB1*, *SNAI2*, and *VIM*, and decreased levels of epithelial genes, such as *EPCAM*, *CDH1*, and *KRT19* **(Fig. 2.4c-d, Fig. 2.5a).** A3 in HCI001 also contained cells with a robust extracellular matrix (ECM) modulatory signature, that displayed increased expression of collagens (*COL12A1*, *COL15A1, COL16A1*), matrix metalloproteinases (*MMP11*, *MMP16)*, and tissue inhibitors of

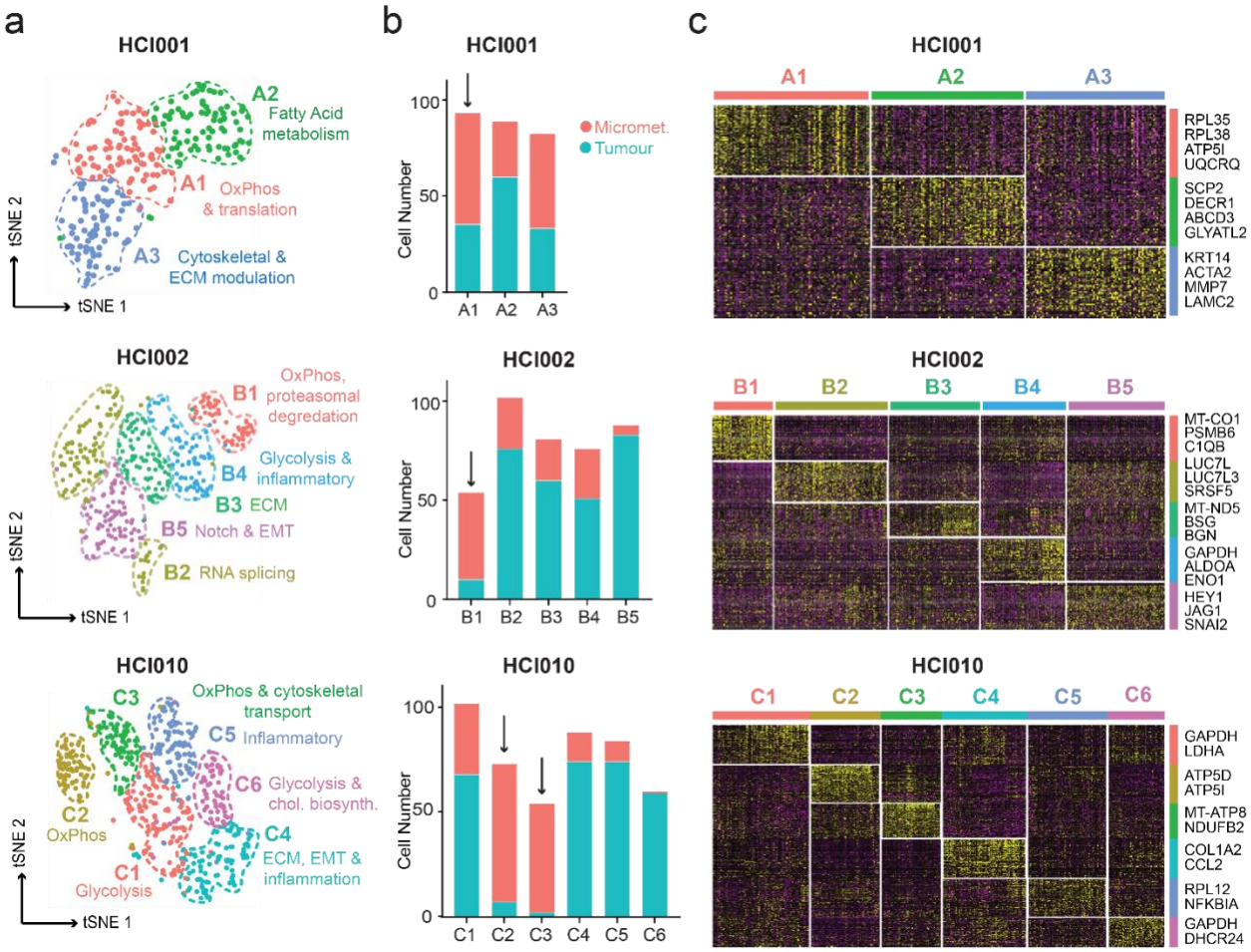

Figure 2.3 Transcriptional diversity in micrometastatic and primary tumor cells. a, tSNE plots show clustering of cells from PDX models HCI001 (n= 247 cells), HCI002 (n= 401 cells) and HCI010 (n= 471 cells). Cells are colored by cluster identity. Biological features defining each population identified by GO term analysis of marker genes are indicated. **b,** Bar graphs quantify the number of micrometastatic and primary tumor cells in each cluster. Arrows indicate clusters enriched for metastatic cells. **c,** Heatmaps show top marker genes for each cluster.

metalloproteinases (*TIMP1*) **(Fig. 2.4c-d, Fig. 2.5b).** B5 in HCI002, and C4 in HCI010 showed analogous ECM modulatory gene signatures **(Fig. 2.4c-d, Fig. 2.5b).** Interestingly, EMT and ECM signatures were often observed in overlapping clusters, particularly in HCI010 where gene signatures for EMT and ECM were significantly correlated in C4, and to some degree in A3 and B5 from HCI001 and HCI002, respectively **(Fig. 2.5c)**. This is consistent with prior reports of increased MMP and collagen expression in cancer cells that display an EMT phenotype.[99] These data show that PDX models display substantial inter- and intratumoral heterogeneity, and distinct populations of cancer cells can be identified in both primary tumor and micrometastases. The partitioning of other tumor functions (i.e., ECM/EMT, OXPHOS, glycolysis) into different cancer cell subpopulations provides support for cooperative models for tumor progression, which propose that cell subpopulations within a tumor carry out discrete cellular functions and cooperate to synergistically propagate the tumor.[100,101]

2.3.3 Breast cancer subtype analysis of primary tumors and metastatic cells

In breast cancer, subtype switching has been observed in resistant tumor cells post-treatment, and metastatic lesions frequently display a different subtype from the original primary tumor at diagnosis.[102] Intratumor heterogeneity specifically relating to breast cancer subtype may therefore be clinically relevant both for drug resistance and metastatic propensity.[103–105] Each of

the three original patients in this study were diagnosed with triple-negative, basal-like breast cancer by immunohistochemical analysis and PAM50 testing.[8] I assigned subtypes to my single cell transcriptomes based on the PAM50 classification system to determine whether any cells representing other subtypes could be identified. As expected, most cells matched the basal-like subtype across each model (**Fig. 2.6a**). Interestingly, 24.8% of cells were identified as HER2, luminal A, luminal B, or normal-like subtypes. These disparate classifications were not specific to cell identity (metastatic versus tumor) or cluster designation (**Fig. 2.6a**). Analysis of several stereotypic markers for each subtype supported this finding and showed that 2-10% of cells

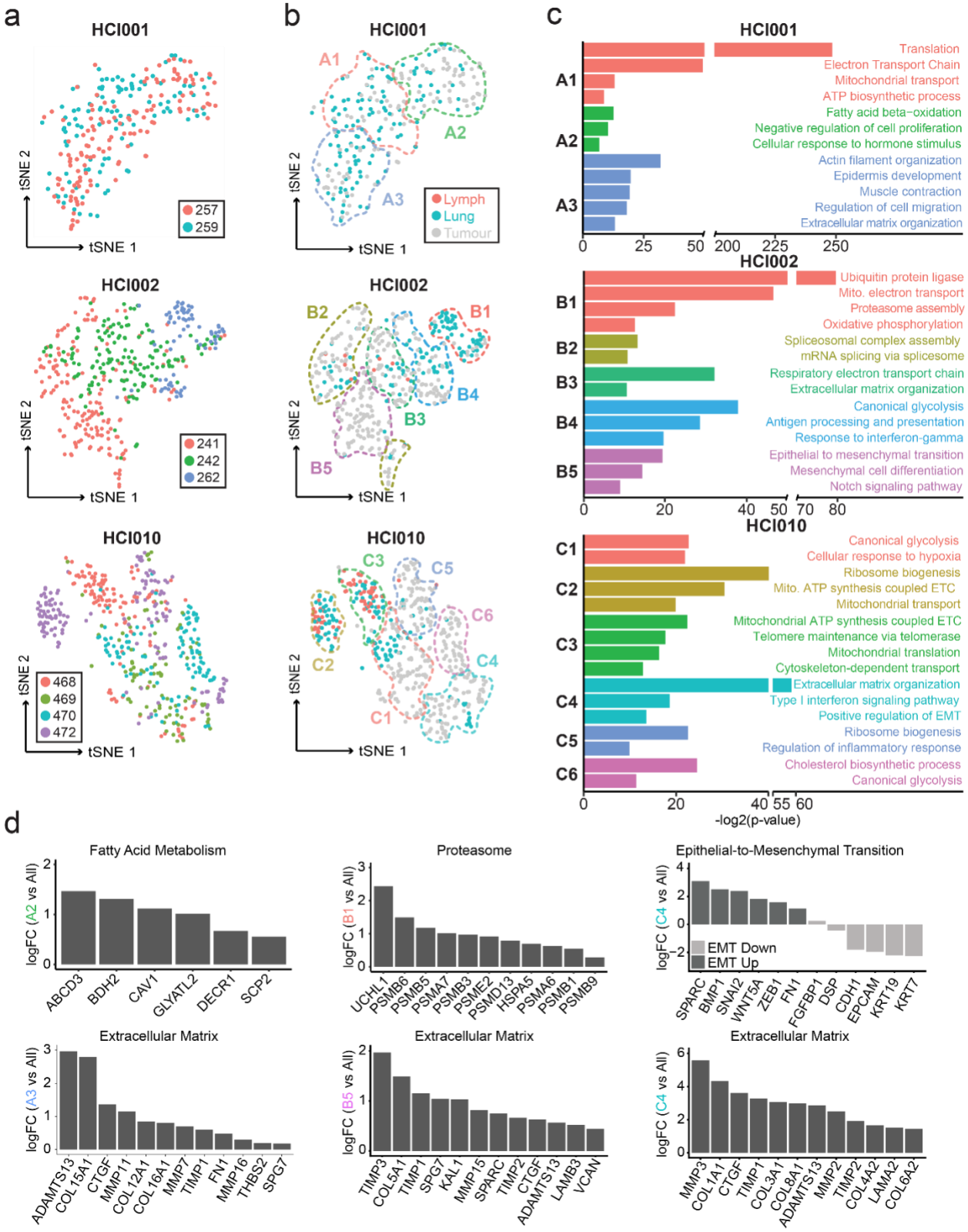
a
HCI001
tSNE 2
tSNE 1
257
259
HCI002
241
242
262
HCI010
468
469
470
472
b
HCI001
A1
A2
A3
Lymph
Lung
Tumour
HCI002
B1
B2
B3
B4
B5
HCI010
C1
C2
C3
C4
C5
C6
c
HCI001
A1
A2
A3
Translation
Electron Transport Chain
Mitochondrial transport
ATP biosynthetic process
Fatty acid beta-oxidation
Negative regulation of cell proliferation
Cellular response to hormone stimulus
Actin filament organization
Epidermis development
Muscle contraction
Regulation of cell migration
Extracellular matrix organization
HCI002
B1
B2
B3
B4
B5
Ubiquitin protein ligase
Mito. electron transport
Proteasome assembly
Oxidative phosphorylation
Spliceosomal complex assembly
mRNA splicing via splicesome
Respiratory electron transport chain
Extracellular matrix organization
Canonical glycolysis
Antigen processing and presentation
Response to interferon-gamma
Epithelial to mesenchymal transition
Mesenchymal cell differentiation
Notch signaling pathway
HCI010
C1
C2
C3
C4
C5
C6
Canonical glycolysis
Cellular response to hypoxia
Ribosome biogenesis
Mito. ATP synthesis coupled ETC
Mitochondrial transport
Mitochondrial ATP synthesis coupled ETC
Telomere maintenance via telomerase
Mitochondrial translation
Cytoskeleton-dependent transport
Extracellular matrix organization
Type I interferon signaling pathway
Positive regulation of EMT
Ribosome biogenesis
Regulation of inflammatory response
Cholesterol biosynthetic process
Canonical glycolysis
-log2(p-value)
d
Fatty Acid Metabolism
logFC (A2 vs All)
ABCD3
BDH2
CAV1
GLYATL2
DECR1
SCP2
Proteasome
logFC (B1 vs All)
UCHL1
PSMB6
PSMB5
PSMA7
PSMB3
PSME2
PSMD13
HSPA5
PSMA6
PSMB1
PSMB9
Epithelial-to-Mesenchymal Transition
logFC (C4 vs All)
EMT Down
EMT Up
SPARC
BMP1
SNAI2
WNT5A
ZEB1
FN1
FGFBP1
DSP
CDH1
EPCAM
KRT19
KRT7
Extracellular Matrix
logFC (A3 vs All)
ADAMTS13
COL15A1
CTGF
MMP11
COL12A1
COL16A1
MMP7
TIMP1
FN1
MMP16
THBS2
SPG7
Extracellular Matrix
logFC (B5 vs All)
TIMP3
COL5A1
TIMP1
SPG7
KAL1
MMP15
SPARC
TIMP2
CTGF
ADAMTS13
LAMB3
VCAN
Extracellular Matrix
logFC (C4 vs All)
MMP3
COL1A1
CTGF
TIMP1
COL3A1
COL8A1
ADAMTS13
MMP2
TIMP2
COL4A2
LAMA2
COL6A2

Figure 2.4 Marker and GO term analysis of cell clusters from each PDX model. a, tSNE plots display clustering of cells colored by mouse of origin from PDX models HCI001 (n = 247 cells), HCI002 (n = 401 cells) and HCI010 (n = 471 cells). **b,** tSNE plots display clustering of cells colored by tissue of origin from PDX models HCI001 (n = 247 cells), HCI002 (n = 401 cells) and HCI010 (n = 471 cells). **c,** Bar plots show selected top GO terms determined by the marker genes identified for each cell cluster. *P* values are determined by the Fisher exact test. Specifically, for HCI001, *n* = 162 A1 genes, *n* = 107 A2 genes, and *n* = 199 A3 genes. For HCI002, *n* = 490 B1 genes, *n* = 173 B2 genes, *n* = 34 B3 genes, *n* = 181 B4 genes, and *n* = 194 B5 genes. For HCI010, *n* = 96 C1 genes, *n* = 247 C2 genes, *n* = 198 C3 genes, *n* = 357 C4 genes, *n* = 54 C5 genes, *n* = 110 C6 genes. **d,** Bar graphs show the log fold change (logFC) for selected genes from GO term pathways. Values indicate the logFC of the average gene expression for the indicated cell cluster relative to all other clusters within that PDX model.

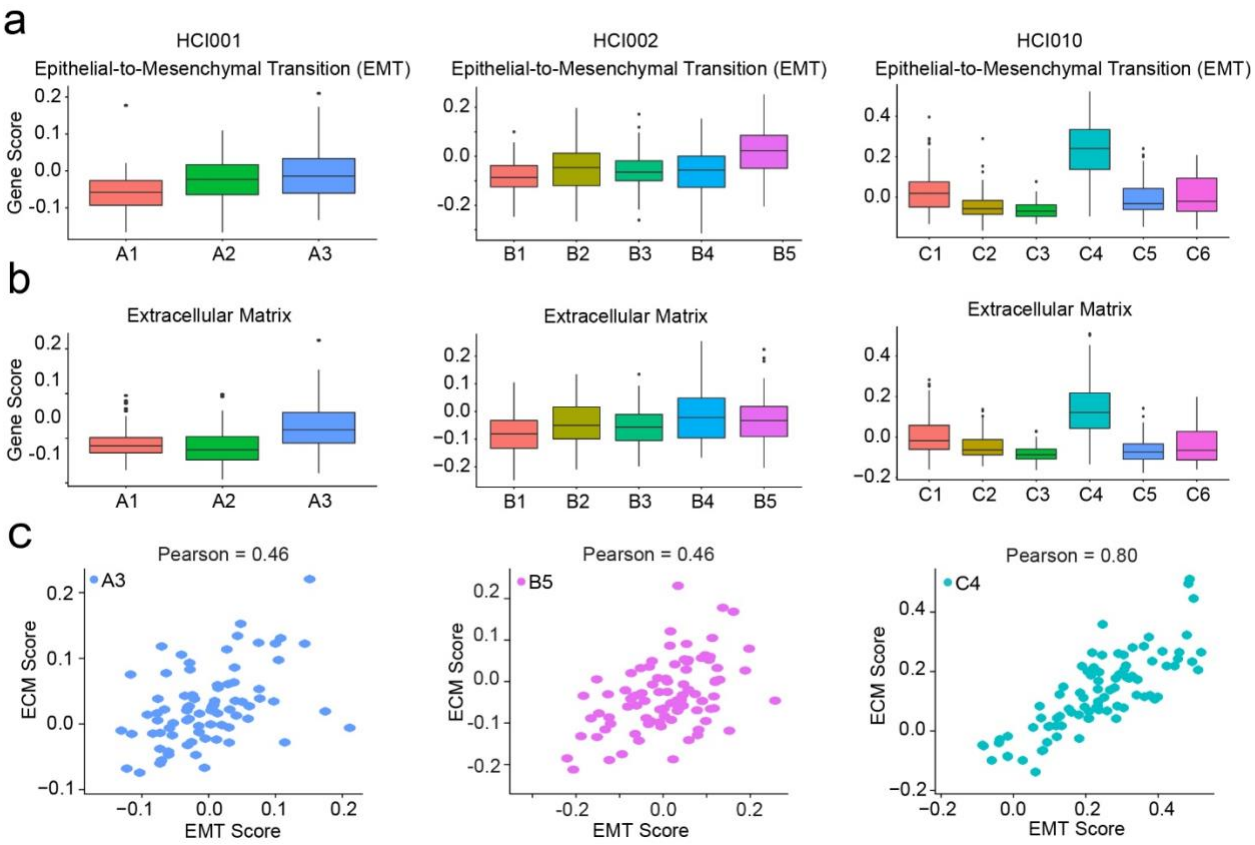

Figure 2.5 Identification of an ECM remodeling and EMT co-correlated population in each PDX model. a-b, Gene scoring for EMT **(a)** and ECM **(b)** in each cell cluster. Cells from each PDX model were scored for an extended list of EMT-associated **(a)** or ECM-associated **(b)** genes. Box plots summarize scores for all cells in each cell cluster. **(c)** Dot plots display correlated gene scores for both ECM and EMT pathways in cluster A3, B5, and C4 from HCI001, HCI002, and HCI010, respectively. Pearson correlation values are indicated.

expressed *ESR1*, 0-12% expressed *PR*, and 33-63% expressed *ERBB2* (also known as HER2) (**Fig 2.6b**). It is important to note, however, that although HCI001 contains a high prevalence of *ERBB2* transcripts, the overwhelming molecular classification of the cells is basal-like (**Fig. 2.6a-b**). The distinction likely arises from difference in histological compared to molecular profiling of these tumor cells, wherein the PAM50 classification of HER2 breast cancer includes concomitant upregulation of *ERBB2, GRB7, FGFR4, and BLVRA*, while also requiring downregulation of *ACTR3B, MYC, and SFRP1*. This is not the case in HCI001; *FGFR4* is only expressed in 7.7% of cells, while *MYC* is expressed in 93.1% of cells (data not shown). Collectively, however, these results are consistent with recent work demonstrating subtype divergence within human breast tumors when bulk subtype classifications were compared to single cell profiles.[106] Given that both the HCI001 and HCI002 patients were treatment naïve at the time of tumor collection for PDX generation, these results suggest that naïve tumors may often contain cells representing different subtypes and that such patients may benefit from combinatorial therapies.

2.3.4 Micrometastatic cells display a distinct transcriptome program

To determine whether micrometastatic cells possess unique transcriptome programs that may facilitate seeding in distal tissues, I performed supervised analysis to directly compare gene expression in primary tumor and micrometastatic cells **(Fig. 2.7a).**[88–90] This identified 330 differentially expressed genes ($p<0.05$; $logFC>0.25$), including 116 genes specifically upregulated in micrometastatic cells conserved in all three PDX models **(Fig. 2.7b).** Top micrometastasis-associated genes included several heat shock proteins (*HSPB1*, *HSPE1*, and *HSPA8)*, which are protein chaperones upregulated in response to environmental stress that play pleiotropic roles in protein folding, wound healing, antigen presentation, protection from apoptosis, and cellular proliferation and differentiation (Fig. 2.7c).[107,108] Several cytokeratins (*KRT14*, *KRT16*, *KRT7*, and *KRT17)* were also upregulated in micrometastatic cells from all three models, as well as *ACTG2*, which encodes a smooth muscle protein involved in cell

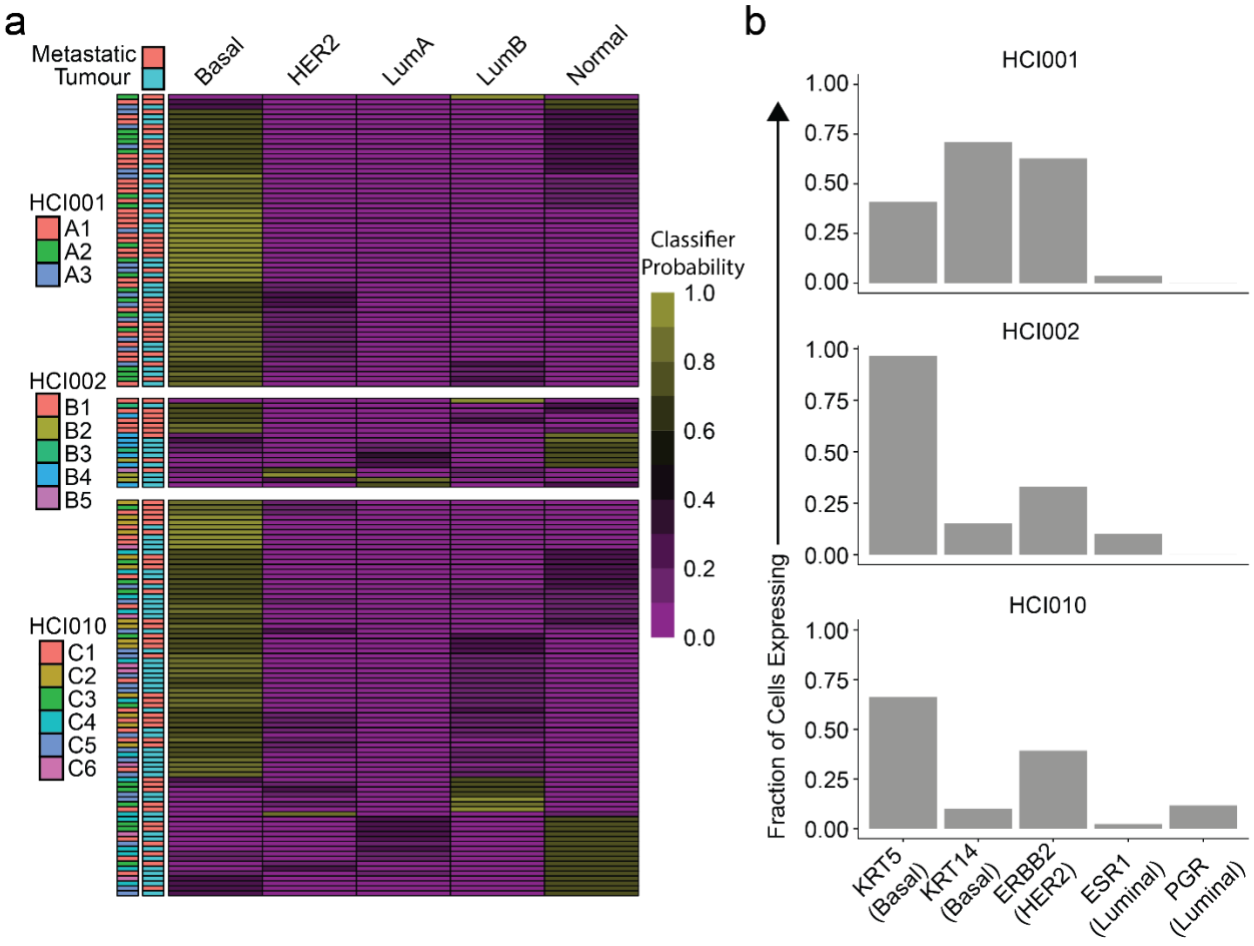

Figure 2.6 Relevance of intratumor heterogeneity on subtype classification. a, Classification probability of individual cells to the indicated subtype based on the PAM50 classifier. Only cells with a classification probability >0.7 are shown. Cell identity (metastatic or tumor) are indicated to the left. **b,** Quantification of the fraction of cells expressing the indicated subtype-specific genes for all three PDX.

motility **(Fig. 2.7c).** Micrometastatic cells also expressed higher levels of several genes with lesser known roles in metastasis, such as *CKB*, *PHLDA2*, *NME1*, *ASHA1*, *NOP16*, and *S100A16* **(Fig. 2.7c).**

We next performed a series of survival analyses, logistic regression and *in situ* validation experiments to confirm increased expression of micrometastasis-associated genes in PDX mice and evaluate disease relevance in breast cancer patients. We first explored whether increased primary tumor expression of micrometastasis-associated genes is predictive of poor survival in basal-like breast cancer patients using microarray data from the KMplotter database (879 patients).[93] This was to determine whether micrometastasis genes are higher in more biologically aggressive tumors, or if they have value as predictive biomarkers for disease progression in patients. Remarkably, we found that 15 of our top 20 micrometastasis-associated genes are significantly predictive of relapse in basal-like patients (two genes were not predictive and data for three genes was not available) ($p<0.05$; HR≥1.4) **(Fig. 2.8a).** Combining all 15 genes further increased the power for predicting patient relapse in this cohort, showing a nearly 3-fold increased risk (HR=2.83; $P = 2.9 \times 10^{-10}$) **(Fig. 2.7d).**

We next utilized a stepwise logistic regression model to identify top biomarker candidates for micrometastasis in our data set. We utilized Akaike information criterion (AIC) to determine the optimal number of genes to include in each subsampling (n=5) and constructed 10 models on equal subsamplings of micrometastatic and primary tumor cells in our data set **(Fig. 2.8b-c).** Using this approach, top candidates will appear the most often in the 10 data subsamplings. We find that *PHLDA2* was the top candidate for micrometastatic cells and appeared in 8 of 10 models **(Fig. 2.8d).** *PHLDA2* is a maternally imprinted gene that regulates placental growth and increases xenograft engraftment and cell invasion *in vitro,* but has limited prior association with breast cancer metastasis.[109,110] Our survival analysis showed that *PHLDA2* is significantly predictive of increased relapse ($p=7.4 \times 10^{-3}$; HR=1.44) in basal-like breast cancer patients **(Fig. 2.8a)**. To determine whether *PHLDA2* transcripts are upregulated in

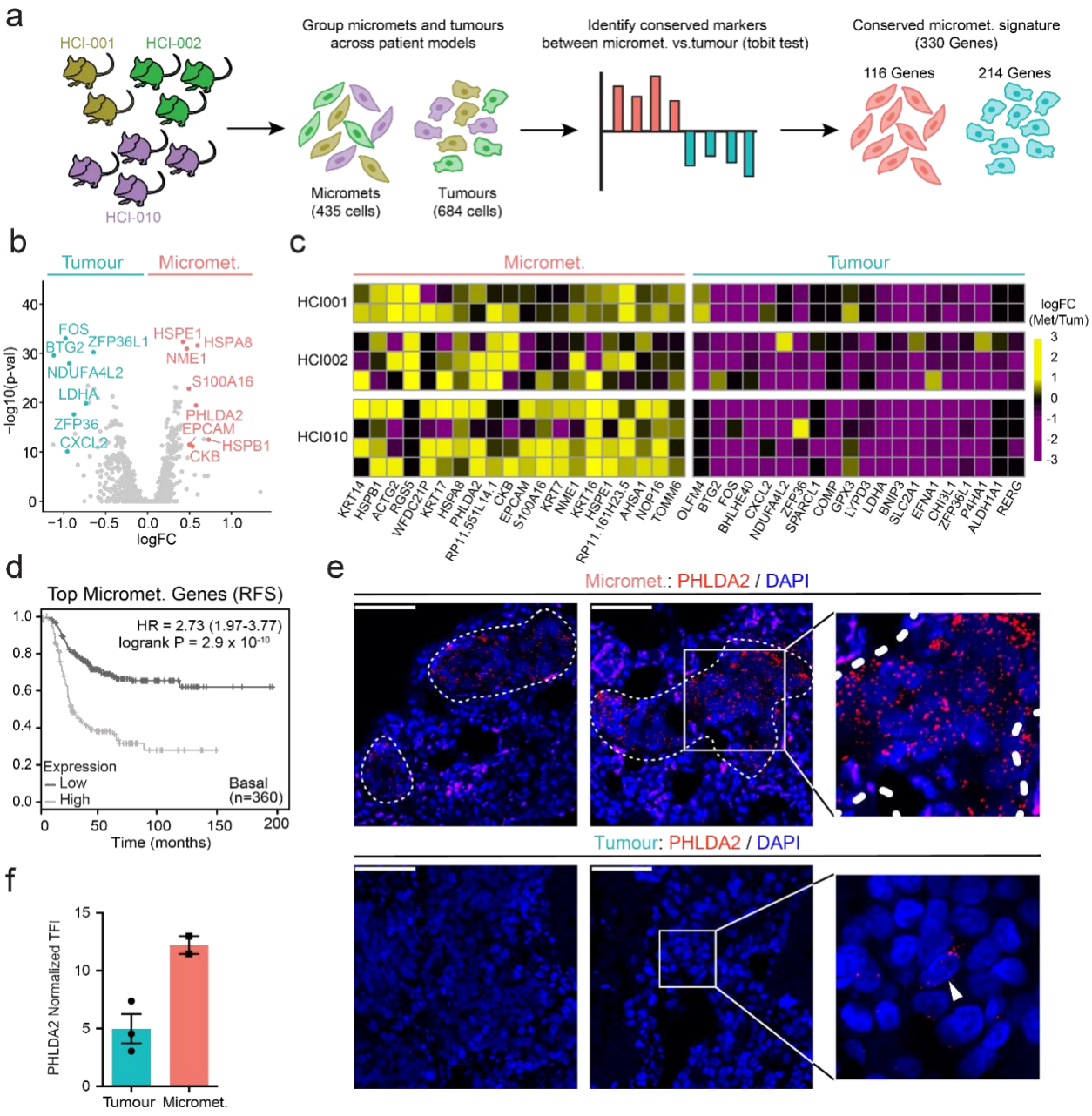
a
HCI-001
HCI-002
HCI-010
Group micromets and tumours across patient models
Micromets (435 cells)
Tumours (684 cells)
Identify conserved markers between micromet. vs.tumour (tobit test)
Conserved micromet. signature (330 Genes)
116 Genes
214 Genes
b
Tumour
Micromet.
−log10(p-val)
logFC
FOS
BTG2
ZFP36L1
HSPE1
HSPA8
NME1
NDUFA4L2
LDHA
S100A16
ZFP36
CXCL2
PHLDA2
EPCAM
HSPB1
CKB
c
Micromet.
Tumour
HCI001
HCI002
HCI010
logFC (Met/Tum)
KRT14
HSPB1
ACTG2
RGS5
WFDC21P
KRT17
HSPA8
PHLDA2
RP11.551L14.1
CKB
EPCAM
S100A16
KRT7
NME1
KRT16
HSPE1
RP11.161H23.5
AHSA1
NOP16
TOMM6
OLFM4
BTG2
FOS
BHLHE40
CXCL2
NDUFA4L2
ZFP36
SPARCL1
COMP
GPX3
LYPD3
LDHA
BNIP3
SLC2A1
EFNA1
CHI3L1
ZFP36L1
P4HA1
ALDH1A1
RERG
d
Top Micromet. Genes (RFS)
HR = 2.73 (1.97-3.77)
logrank P = 2.9 x 10-10
Expression
Low
High
Basal (n=360)
Time (months)
e
Micromet.: PHLDA2 / DAPI
Tumour: PHLDA2 / DAPI
f
PHLDA2 Normalized TFI
Tumour
Micromet.

Figure 2.7 Micrometastatic cells display a distinct transcriptome program. a, Schematic of the analysis method to identify genes that were differentially expressed between micrometastatic and primary tumor cells, conserved in all three PDX models. All of the cells in the dataset were classified as micrometastatic or tumor cells and the differential gene expression between the groups was calculated using the tobit test in Seurat with the patient ID as a latent variable; 330 differentially expressed genes were identified ($P < 0.05$, min.pct = 0.1, log[FC] threshold = 0.25). **b,** Volcano plot showing all genes that were differentially expressed between micrometastases and tumors. The P values were determined using the tobit test in Seurat, which utilizes a likelihood-ratio test. **c,** Heat map of the top-20 marker genes for micrometastatic and primary tumor cells. The average FC of each gene (x axis) in micrometastatic relative to primary tumor cells was plotted for each mouse (y axis). Yellow indicates higher expression in micrometastatic cells and purple indicates higher expression in tumor cells. **d,** Kaplan–Meier survival curve showing decreased relapse-free survival in patients with basal-like breast cancer (n = 360) who expressed high levels of the top-15 micrometastasis-associated genes. The P values were determined using a log-rank test. **e,** Representative fluorescent in situ hybridization for PHLDA2 (RNAscope) on primary tumor (n = 3 biologically independent samples; bottom) and lung micrometastases (n = 2 biologically independent samples; top) from the PDX model HCI001. Insets, higher magnification of individual puncta. The white arrow indicates a tumor cell with high expression of PHLDA2. Scale bars, 50 μm. **f,** Normalized total fluorescent intensity (TFI) of PHLDA2 in primary tumor and micrometastatic cells from the PDX model HCI001 (n = 2 lungs, >15 lesions; n = 3 tumors, 22 fields). Data are shown as the mean ± s.e.m.

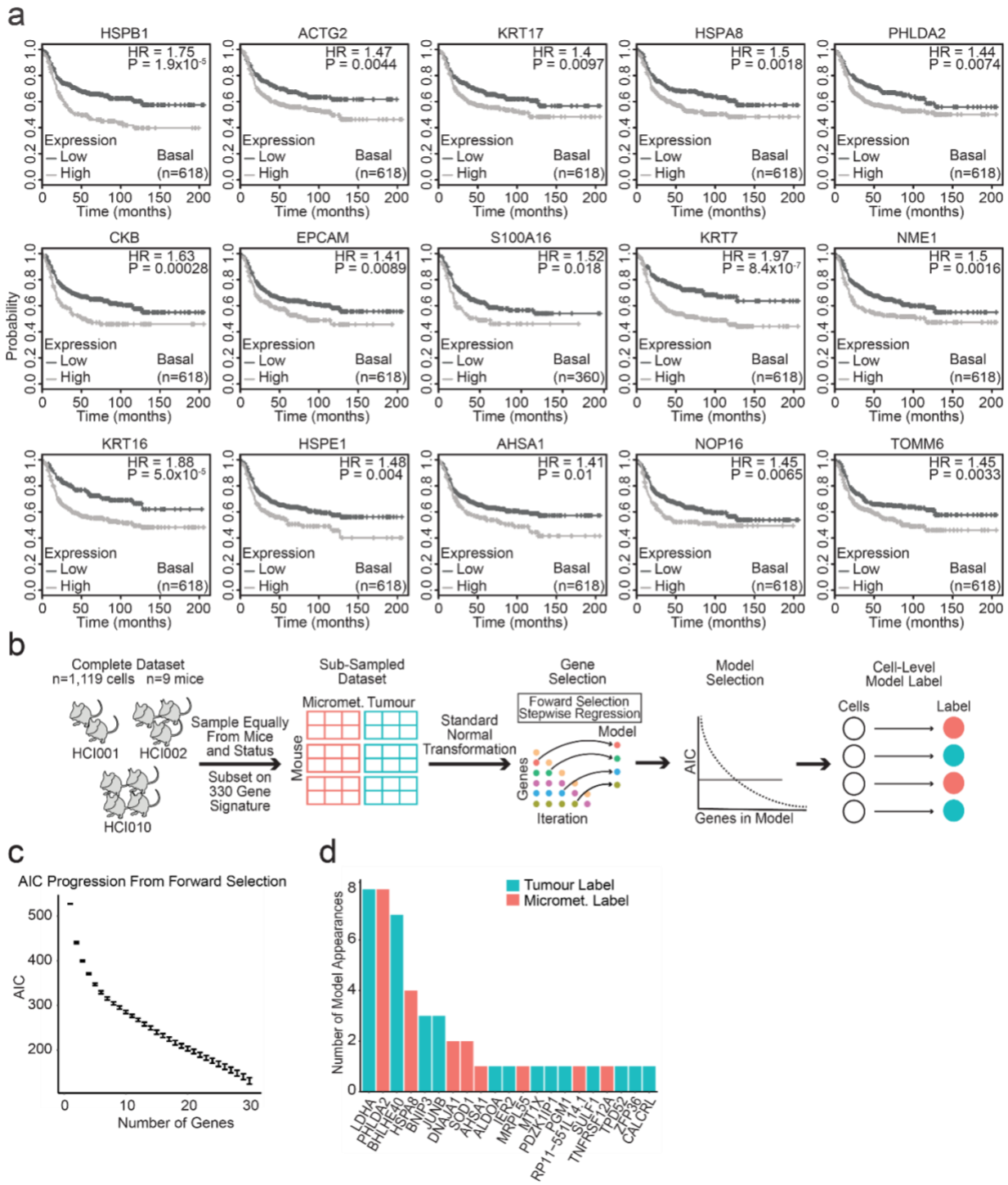
a
HSPB1
HR = 1.75
P = 1.9x10-5
ACTG2
HR = 1.47
P = 0.0044
KRT17
HR = 1.4
P = 0.0097
HSPA8
HR = 1.5
P = 0.0018
PHLDA2
HR = 1.44
P = 0.0074
CKB
HR = 1.63
P = 0.00028
EPCAM
HR = 1.41
P = 0.0089
S100A16
HR = 1.52
P = 0.018
KRT7
HR = 1.97
P = 8.4x10-7
NME1
HR = 1.5
P = 0.0016
KRT16
HR = 1.88
P = 5.0x10-5
HSPE1
HR = 1.48
P = 0.004
AHSA1
HR = 1.41
P = 0.01
NOP16
HR = 1.45
P = 0.0065
TOMM6
HR = 1.45
P = 0.0033
Probability
Expression
Low
High
Basal
(n=618)
(n=360)
Time (months)
b
Complete Dataset
n=1,119 cells n=9 mice
HCI001
HCI002
HCI010
Sample Equally From Mice and Status
Subset on 330 Gene Signature
Sub-Sampled Dataset
Micromet. Tumour
Mouse
Standard Normal Transformation
Gene Selection
Foward Selection Stepwise Regression
Model
Genes
Iteration
Model Selection
AIC
Genes in Model
Cell-Level Model Label
Cells
Label
c
AIC Progression From Forward Selection
AIC
Number of Genes
d
Tumour Label
Micromet. Label
Number of Model Appearances
LDHA
PHLDA2
BHLHE40
HSPA8
BNIP3
JUNB
DNAJA1
SOD1
AHSA1
ALDOA
IER2
MRPL55
MT1X
PDZK1IP1
EGLN1
RP11-551L14.1
SULF1
TNFRSF12A
TPD52
ZFP36
CALCRL

Figure 2.8 Prognostic value of micrometastasis- associated genes in basal-like breast cancer patients. a, Kaplan-Meier curves show relapse free survival (RFS) in basal-like breast cancer patients from the KM plotter database (879 patients), based on their primary tumor expression of specified micrometastasis-associated genes. P values were determined via a log-rank test. **b,** Schematic for the construction of a stepwise logistic regression model to identify top biomarker candidates descriptive of primary tumor or micrometastatic cells. The data was subsampled to analyse equal numbers of micrometastatic and tumor cells from each mouse. The model was run on 10 subsamplings of the data, with the number of genes in each model determined by AIC. **c,** Plot demonstrating the AIC versus the number of genes included in the model. AIC is used to balance parameter additions (that is gene additions) with the descriptive power of a model. Data is presented as the 10% and 90% quantiles of the 10 data subsamplings. **d,** Bar plot showing the number of model appearances for each gene out of 10 data subsamplings.

micrometastatic lesions *in situ*, I compared *PHLDA2* transcript levels in primary tumor and micrometastases using high resolution, single molecule fluorescence *in situ* hybridization (FISH) (RNAscope). I found that *PHLDA2* transcript levels are >2-fold higher in micrometastases relative to primary tumors, validating our approach for identifying genes upregulated in micrometastasis **(Fig. 2.7e-f).** Interestingly, only rare primary tumor cells expressed *PHLDA2*, raising the question of whether it marks pre-metastatic cells **(Fig. 2.7e).** These data highlight my dataset as a resource for the identification of potential drivers of metastatic seeding, and biomarkers to predict metastatic progression in breast cancer patients.

2.4 Discussion

These results offer new insights into the complexity of intratumoral heterogeneity in the context of metastasis. Prior work has demonstrated that metastases arise from rare and phenotypically distinct cells in the primary tumor, such as a subpopulation of $CXCR4^{+}$ tumor cells in pancreatic ductal adenocarcinoma[3] or $CD90^{+}CD24^{+}$ cells in mouse models of breast cancer.[1] Surprisingly, I find that micrometastatic cells contained the same diverse array of cell states as the primary tumors, albeit in different proportions **(Fig. 2.3b)**. This finding can potentially be explained by two separate, but not mutually exclusive, models of micrometastatic establishment. In the first model, primary tumor cells from a broad range of cell states are shed into circulation during tumor development and seed distal tissues. Each of these cell states have differing capacity to survive each step in the metastatic cascade and are thus eradicated at differing rates, leading to the establishment of micrometastases that are dominated by certain cell states (i.e. A1, B1, C2/C3), but contain trace levels of all cell states (i.e. B5, C6) **(Fig. 2.3b)**. This model would be supported by reports of collective cell migration in the context of cancer, which posits that multiple cells maintain their cell-cell junctions at invasive zones and move together during initial invasion into the surrounding stroma and intravasation into circulation or lymphatics.[111] This phenomenon is seen in multiple cancers, including breast cancer and may explain how multiple cell states from the primary tumor escape and transit to sites of metastasis

to form multi-state micrometastases.[112,113] In the second model, a single cell state is uniquely capable to surviving all steps of the metastatic cascade (i.e. A1, B1, C2/C3) and establishing initial singlet or multiplet colonies of metastases at distal tissues. However, during the establishment of micrometastasis and the development of a new secondary tumor site, additional tumor cell states are required to support this transition, such as additional ECM remodeling (A3, B5, C4) or generation of an inflammatory microenvironment (B4, C5). While these states themselves are incapable of surviving the multiple bottlenecks of the metastatic cascade, they are required as synergistic components of driving tumor establishment, and thus the initial founder colonies differentiate into these "helper" states. In support of this model, recent studies demonstrated that glioblastoma cells exist in up to four distinct cell states, with different frequencies that are highly plastic and dependent on the genetic and microenvironmental ques.[53,114] Furthermore, that these cell states exist in a hierarchy of differentiation.[53,114] This indicates that tumor cells have high degrees of heterogeneity with the capacity to readily differentiate between these states, and that this heterogeneity is context dependent.

A prevailing concept in metastasis is the reversion of an tumor cell of epithelial origin to a more motile mesenchymal-like cell state, a process referred to as epithelial-to-mesenchymal transition (EMT).[44] This process has been extensively studied and is consideered central to the progression of a tumor from localized to systemic.[4,85,115,116] In agreement with previous work, I also identified populations of cancer cells with upregulation of EMT-associated programs in all three TNBC PDX models **(Fig. 2.4c-d, Fig. 2.5a)**. Interestingly, however I find that these same populations are highly correlated with expression of both ECM components (*COL1A1, COL1A3, LAMA2*) in addition to ECM remodeling enzymes (*MMP3*, *MMP2)* **(Fig. 2.4c-d, Fig. 2.5b-c)**. Remodeling of the ECM has been shown to be critical to the promotion of an EMT phenotype in cancer cells. While it is believed that increases in ECM remodeling enzymes in the microenvironment is contributed by stromal cells, including cancer-associated fibroblasts

(CAFs)[117], my data demonstrates that the tumor cells can also contribute to the generation of these enzymes. As a result, a feed-forward loop is created in these tumor subpopulations where the EMT phenotype is further enhanced by increased production of ECM remodeling enzymes by the same cells. Additionally, it is important to note that the EMT/ECM subpopulations (A3, B5, C4) appears to be preferentially enriched in primary tumor cells, rather than micrometastatic cells **(Fig. 2.3b).** This finding further supports the concept of the reversion of metastatic cells to an epithelial phenotype following seeding at distal tissues and the establishment of a secondary tumor site, or the so-named mesenchymal-to-epithelial transition (MET).[118]

Despite significant transcriptional diversity, I find that metastatic cells also display a distinct core expression program, consisting of 330 differentially expressed genes, that is conserved across the PDX models **(Fig. 2.7b)**. There are currently no well-established predictive biomarkers for metastasis in routine clinical use, so the ability to determine whether their expression levels in human tumors correlates with metastatic progression is clinically valuable. I demonstrate that the top 20 genes associated with the core metastatic expression program is highly predictive (HR=2.73, p-val=2.9×10^{-10}) of relapse-free survival in a large cohort of TNBC patients **(Fig. 2.7d).** One can envisage how this approach may one day be used in diagnostic applications, to score tumor samples for a core metastatic expression program to predict clinical outcomes, such as local relapse or organ-specific metastasis. I validated a specific marker from this core expression program, *PHLDA2*, *in situ* and found substantial upregulation in micrometastatic cells compared to primary tumor cells **(Fig. 2.7e-f)**. Interestingly, while most tumor fields expressed little to no *PHLDA2,* there existed individual tumor cells with similar expression levels to micrometastatic lesions **(Fig. 2.7e)**. An intriguing question for future studies will be to determine whether primary tumors that are enriched for cells with this marker are more likely to metastasize, in addition to the functional consequences of upregulation of this gene for metastatic potential.

Reprinted with permission from:

Davis, R.T., Blake, K., Ma, D. et al. Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell RNA sequencing. Nat Cell Biol 22, 310–320 (2020).

Chapter 3: Metabolic phenotype

3.1 Introduction

I find that micrometastases display a distinct transcriptome program from primary tumors and identify mitochondrial oxidative phosphorylation (OXPHOS) as a top pathway upregulated during metastatic seeding. This was confirmed at the molecular level, where I show that micrometastatic cells display increased mitochondrial membrane potential and a distinct metabolic profile. I further find that pharmacologic inhibition of OXPHOS specifically in cancer cells attenuates metastatic seeding in the lung using two different breast cancer experimental metastasis models. These data show that OXPHOS is functionally important for metastasis and suggest that different metabolic strategies may be advantageous for cells to complete discrete stages of the metastatic cascade.

3.2 Materials and Methods

3.2.1 Seahorse Assay

$8x10_4$ MDA-MB-231 or 4T1-GFP cells were seeded into Seahorse XF24 microplates (Agilent, Cat. No. 100850), and the XF24 cartridge (Agilent, Cat. No. 100850) was calibrated in the Seahorse prep station (Agilent) overnight. Before the assay, the medium was replaced by 0.5 mL XF base medium (Agilent, Cat. No. 102353) supplemented with 10 mM glucose, 2 mM glutamine, and 1 mM pyruvate. Cells were incubated at 37 °C for 1 hour in the Seahorse prep station. 56 µL oligomycin (2 µM) (Sigma, Cat. No. 75351), 62 µL FCCP (2 µM) (Sigma, Cat. No. C2920) and 69 µL ROT/AA (1 µM) (Sigma, Cat. No. 557368, A8674) were added into the cartridge wells. OCR and ECAR levels were determined using Seahorse bioscience XF24 extracellular flux analyzer (Agilent) and each cycle of measurement involved mixing (3 min), waiting (2 min), and measuring (3 min) cycles. To quantify basal respiration, I calculated the difference in the average OCR before addition of oligomycin and the average OCR after the addition of Antimycin A/Rotenone. For ATP-dependent respiration, I calculated the difference in

average OCR before addition of oligomycin and after addition of oligomycin. Maximal respiration rates were calculated as the difference in average OCR after addition of FCCP and the average OCR after the addition of oligomycin. Any negative values were set to 0 for visualization and when calculating differences between groups.

3.2.2 Proliferation Assay

For proliferation studies, $2x10_5$ MDA-MB-231 cells were seeded on 6-well plates (Genesee Scientific, Cat. No. 25-105) containing DMEM-F12 (Fisher Scientific, Cat. No. MT10090CV), 10% heat-inactivated fetal bovine serum (Sigma-Aldrich, Cat. No. 12306C), 1% penicillin-streptomycin 100X solution (Hyclone, Cat. No. SV30010) , and 5mM HEPES (Gibco, Cat. No. 15630-080) approximately 24 hours before drug treatment and left to adhere. $2x10_5$ 4T1 cells were seeded on 6-well plates (Genesee Scientific, Cat. No. 25-105) containing DMEM-F12 (Fisher Scientific, Cat. No. MT10090CV), 5% heat-inactivated fetal bovine serum (Sigma-Aldrich, Cat. No. 12306C), and 1% penicillin-streptomycin 100X solution (Hyclone, Cat. No. SV30010) approximately 24 hours before drug treatment and left to adhere. The following day, the media was replaced with media containing 1μM oligomycin (MP Biomedicals Cat. No. 0215178610), while the control cells had their media replaced. Cells were washed with PBS and their media was replaced after 8 hours of treatment or 6 hours of treatment in the case of 4T1-GFP and MDA-MB-231 cells, respectively. At 24, 48, and 72 hours post-treatment cells were collected with 0.05% Trypsin (Corning, Cat No. 25-052-Cl) and counted using a Countess II Automated Cell Counter (ThermoFisher Scientific, Cat. No. AMQAX1000). Growth rate was calculated between 24 and 48hrs and between 48 and 72hrs post-treatment with the following equation:

$$Growth\ Rate = \frac{\ln\left(\frac{N(t)}{N(0)}\right)}{t}$$

Where *N(t)* is the number of cells at either 48hrs or 72hrs, *N(0)* is the number of cells at either 24hrs or 48hrs, and *t* is the time between measurements.

3.2.3 Viability Assay

For analysis of cell viability following oligomycin treatment, cells were stained with Annexin V-FITC, diluted 1:100 (GeneTex Cat. No. GTX14082) and propidium iodide (PI), diluted 1:100 (ThermoFisher Scientific Cat. No. P3566). FSC-W x FSC-A and SSC-W x SSC-A was used to discriminate single cells from doublet and multiplet cells. Viability was determined by the frequency of Annexin-$V_{Neg}PI_{Neg}$ cells in the population.

3.2.4 FLIM Acquisition

For FLIM analysis, 4T1-GFP mouse breast cancer cells and MDA-MB-231 human breast cancer cells were cultured in DMEM (ThermoFisher Scientific Cat. No. 11965092) supplemented with 10% heat-inactivated fetal bovine serum (ThermoFisher Scientific Cat. No. 10082139), and 1% penicillin-streptomycin 100X solution (Hyclone, Cat. No. SV30010). Cells were plated at a confluency of $2x10_4$ cells/cm_2 in 8-well glass-bottomed (ThermoFisher Scientific Cat. No. 155411), fibronectin coated imaging dishes approximately 24 hours before drug treatment and left to adhere. The following day, the media was replaced with media containing 1µM oligomycin (MP Biomedicals Cat. No. 0215178610), while the control cells had their media replaced. Cells were washed with PBS and their media was replaced after 8 hours of treatment or 6 hours of treatment in the case of 4T1-GFP and MDA-MB-231 cells, respectively. The cells were then imaged at 0, 6, and 24 hours post-washout for no greater than one hour per experimental condition. All FLIM imaging experiments were performed as previously described.[119] FLIM images were acquired on an inverted laser scanning confocal microscope with a 40x 1.2NA water-immersion objective with cells kept under biological conditions throughout. 4T1-GFP and MDA-MB-231 cells were excited at approximately 2mW with a two-photon Ti:Sapphire laser (Spectra-Physics, MaiTai) at 740 nm, which was passed through a 690 nm dichroic filter. The fluorescence emission was separated through a bandpass filter (442/46 nm) to capture cell auto-fluorescence and exclude any GFP emission and was then detected by a photomultiplier tube (Hamamatsu, Cat. No. H7422P-40). Fluorescence lifetime decays were

captured in the frequency domain via an A320 FastFLIM box (ISS) and subsequently mapped onto the phasor plot using the SimFCS software, developed at the Laboratory for Fluorescence Dynamics at the University of California, Irvine, for quantitative NADH lifetime analysis. Coumarin-6 in ethanol, with a known single exponential fluorescence lifetime of 2.5 ns, was imaged before each experiment and used as a calibration sample for instrument response time.

3.2.5 FLIM Data Analysis

Each pixel of the FLIM image was Fourier transformed and plotted on the phasor plot in SimFCS, as previously described.[119] Each FLIM image was individually masked to exclude any nuclei from the analysis. Each masked frame's cluster of phasors were then averaged to obtain its G and S coordinate, which was then used to determine the cluster's fraction of free NADH. This free NADH fraction is calculated on the fact that 100% free NADH has a decay of approximately 0.4 ns, and 100% bound NADH has a decay of approximately 3.2 ns. Any pixel with a fluorescence decay containing a mixture of both free and bound NADH will lie within the linear combination of 100% free and 100% bound NADH when on the phasor plot in the frequency domain. Consequently, the cluster's fraction of free NADH can be calculated by the cluster's mathematical distance between the 100% free NADH and 100% bound NADH positions on the universal circle. For each independent experiment, the treated sample's fraction of free NADH was subtracted from the average of the control sample's fraction of free NADH to obtain the fraction of free NADH difference. Significance was determined via a two-tailed T-test with homoscedasticity determined via F-test. Error bars represent the standard error of the mean.

3.2.6 Orthotopic transplantation and intravenous injection of cell lines into mice

The University of California, Irvine Institutional Animal Care and Use Committee (IACUC) reviewed and approved all animal experiments. For experimental metastasis studies, cultured MDA-MB-231 or 4T1-GFP were suspended in 100µl of sterile PBS and injected into the tail vein of 8-10-week-old female Nod/SCID (MDA-MB-231) or BALB/c (4T1-GFP) mice. Mice

were euthanized by asphyxiation with CO_2 followed by cervical dislocation and perfusion with 10mM EDTA in D-PBS 21 days (MDA-MB-231) or 5 days (4T1-GFP) post-injection. For cell line orthotopic tumor injections, mice were placed under isoflurane anesthesia (1.5–2.0%). Prior to injection the area was shaved and cleaned with an alcohol swab. $1x10_6$ 4T1-GFP or MDA-MB-231 cells in 100µl phosphate buffered saline (PBS) were injected bilaterally into the fourth mammary fat pad of 10-12-week-old NSG mice. The study is compliant with all relevant ethical regulations regarding animal research.

3.2.7 Flow cytometry

For mitochondrial membrane potential readouts, cells were stained with TMRM, diluted 1:500 (ThermoFisher Scientific Cat. No. T668), MitoTracker-Green, diluted 1:100 from a 10µM stock (ThermoFisher Scientific Cat. No. M7514), human-specific antibody CD298, diluted 1:100 (APC, BioLegend, Cat. No. 341706), and the mouse-specific antibody MHC-I, diluted 1:100 (PE/Cy7, BioLegend, Cat. No. 114717). Cell viability was determined by negative staining for SYTOX Blue, diluted 1:1000 (ThermoFisher Scientific Cat. No. S34857). Compensation controls and FMOs were used to determine $TMRM_{High}$ and $TMRM_{Low}$ populations in the PDX tumor cells. For analysis of MDA-MB-231 metastatic burden, FSC W x FSC A and SSC W x SSC A was used to discriminate single cells from doublet and multiplet cells. MDA-MB-231 cells were defined by gating on $Sytox_{Neg}CD298_{Pos}$MHC-I_{Neg}, and the level of metastatic burden was determined by the frequency of that population. One mouse from the oligomycin treated group did not contain any detectable MDA-MB-231 cells by flow cytometry and was thus excluded from the analysis. For analysis of 4T1-GFP metastatic burden, FSC H x FSC A and SSC H x SSC A was used to discriminate single cells from doublet and multiplet cells. 4T1-GFP cells were defined by gating on $Sytox_{Neg}GFP_{Pos}$ cells, and the level of metastatic burden was determined by the frequency of that population.

3.2.8 Quantitative real-time PCR

Total RNA was extracted using Quick-RNA MicroPrep kit (Zymo, Cat. No. R1050). The mRNA was reverse transcribed into cDNA using the iScript cDNA synthesis kit (BioRad, Cat. No. 1708891) according to the manufacturer's protocol. The quantitative real-time PCR reactions were performed using PowerUp SYBRGreen master mix (Applied Biosystems, Cat. No. A25742). The primer sequences used for each gene are listed in the following table:

Table 1 Primer sequences used for quantitative real-time PCR

Gene	Forward Primer Sequence	Reverse Primer Sequence
ESRRA1	ACGAGTGTGAGATCACCAAG	CGCACTCCCTCCTTGAG
PGC1B	GATGCCAGCGACTTTGACTC	ACCCACGTCATCTTCAGGGA
ATP5F1	TGCAAGGAACTTCCATGCCTC	CGCCCAGTTTCTTCAAGATCAA
COX5B	ATGGCTTCAAGGTTACTTCGC	ATGGCTTCAAGGTTACTTCGC

3.2.9 Generation of steady-state metabolomics data using LC-HRMS

Approximately 1000 cells were sorted using flow cytometry from primary tumors and lung metastatic cells from six HCI010 PDX transplants in NSG mice. Cells were centrifuged to remove medium and extracted in pre-cooled 80% HPLC methanol in HPLC-grade water. Samples were placed at -80°C for 15 min. Samples were then placed on ice and mixed several times. All samples were then centrifuged at 20,000 rcf for 10 min at 4°C. Supernatant was transferred to a new tube and dried using speed vacuum. The samples were prepared and analyzed by LC-HRMS as previously described.[120] Integrated peak intensities were used. Hierarchical clustering and heat map were generated using MetaboAnlyst software (https://www.metaboanalyst.ca/) with the samples normalized using the Autoscaling feature, FDR was set at 5%, and metabolites were ranked by t-tests. Pathway enrichment analysis was conducted using MetaboAnalyst; briefly, metabolite identifications from the human metabolome database from the metabolites that were significantly enriched (FC>1.5 and p>0.05) were inputted. The pathway library used was Homo sapiens and hypergeometric test was used for

overrepresentation analysis. Partial least squares-discriminant analysis (PLS-DA) was implemented using the plsda function in R package "mixOmics".[121]

3.3 Results

3.3.1 Micrometastatic cells upregulate mitochondrial oxidative phosphorylation

To identify pathways and cellular properties upregulated in cancer cells during metastatic seeding, I performed GO term analysis on my 330-gene signature**.** Top GO terms for micrometastatic cells included *'epidermis development'* (*CALML5, KRT17, KRT16, KRT14, KLK5*) and *'response to unfolded protein'* (*DNAJA1, HSPA8, HSPB1, HSPE1, HSPD1*) **(Fig. 3.1a).** Interestingly, many top GO terms centered on mitochondrial biology and metabolism, and included *'respiratory electron transport chain*', *'ATP metabolic process,' 'mitochondrial transport,'* and *'oxidative phosphorylation*' **(Fig. 3.1a).** In contrast, *'canonical glycolysis,'* and *'pyruvate metabolic process'* were top GO terms for primary tumor cells, suggesting clear metabolic differences between primary tumor and micrometastatic cells **(Fig. 3.1a).** Analysis of GO term-associated genes showed that 27 genes related to OXPHOS, glycolysis, and mitochondrial complexes were differentially expressed ($p<0.05$) **(Fig. 3.1b).** Micrometastatic cells expressed higher levels of numerous mitochondrial electron transport chain (ETC) genes, including *NDUFS6*, *NDUFAB1*, *NDUFB2*, *NDUFAF4*, *UQCC3* and *COA6*, as well as the ATP synthase subunits *ATP5I*, *ATP5G1*, and *ATP5J2* **(Fig. 3.1b).** Micrometastatic cells also expressed higher levels of the mitochondrial transport proteins *TOMM5*, *TOMM6*, and *TIMM13*, and the mitochondrial ribosome genes *MRPL14*, *MRPL55*, and *MRPL51* that translate mitochondrial genes encoding ETC proteins **(Fig. 3.1b)**. *SOD1* (Superoxide dismutase 1) was also upregulated in micrometastatic cells, which encodes a mitochondrial enzyme that converts superoxide radicals to O_2 and H_2O_2, and thus may protect micrometastatic cells from apoptosis due to oxidative stress from increased ETC activity **(Fig. 3.1b)**. In contrast, primary tumor cells expressed higher levels of genes associated with glycolysis, including the glycolytic enzymes *ALDOA*, *ALDOB*, *ALDOC*, *PGM1* and *PGK1* **(Fig. 3.1b)**. Our logistic regression analysis also

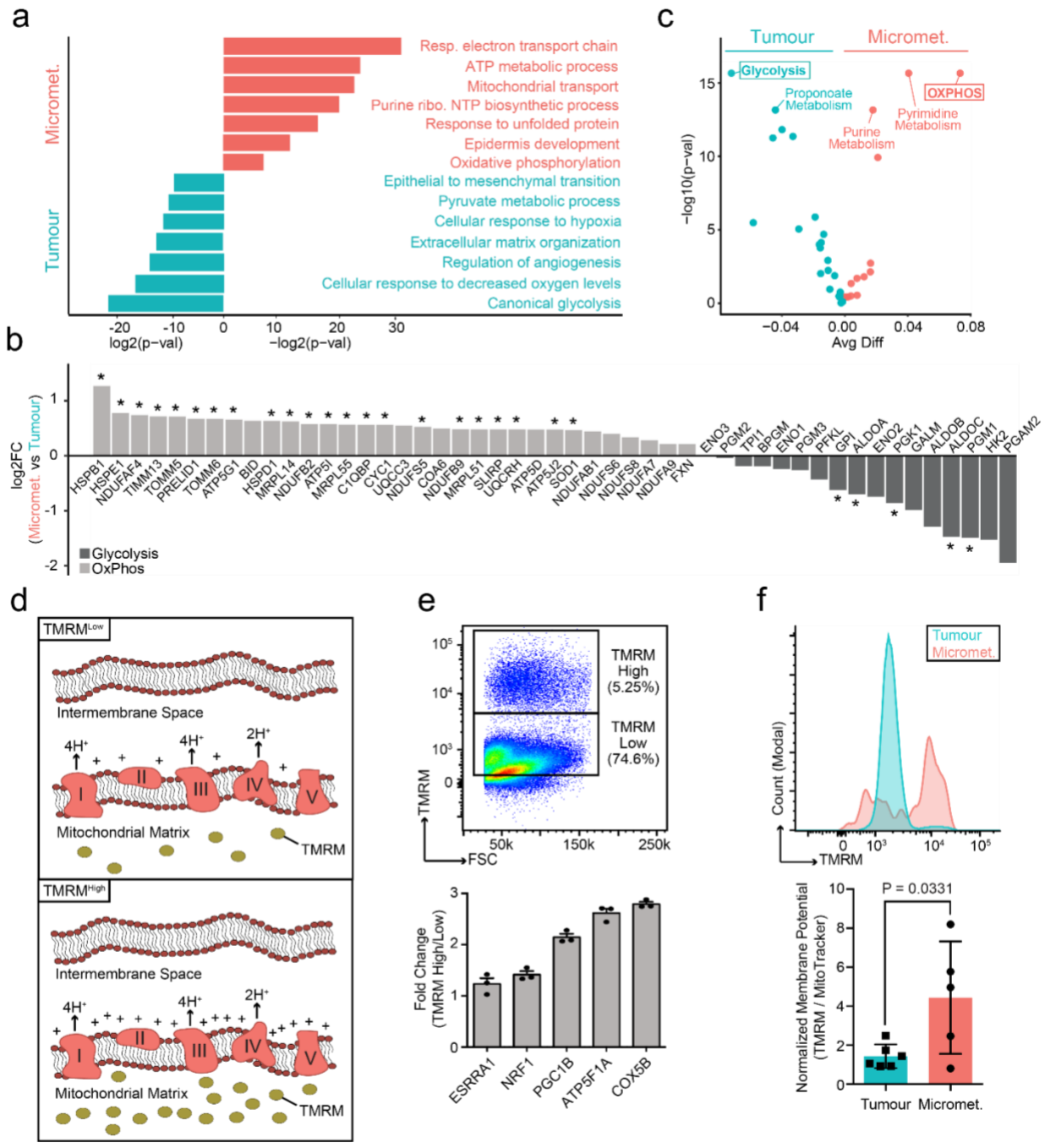
a
Micromet.
Tumour
Resp. electron transport chain
ATP metabolic process
Mitochondrial transport
Purine ribo. NTP biosynthetic process
Response to unfolded protein
Epidermis development
Oxidative phosphorylation
Epithelial to mesenchymal transition
Pyruvate metabolic process
Cellular response to hypoxia
Extracellular matrix organization
Regulation of angiogenesis
Cellular response to decreased oxygen levels
Canonical glycolysis
log2(p−val)
−log2(p−val)
c
Tumour
Micromet.
Glycolysis
Proponoate Metabolism
Purine Metabolism
Pyrimidine Metabolism
OXPHOS
−log10(p−val)
Avg Diff
b
log2FC (Micromet. vs Tumour)
Glycolysis
OxPhos
HSPB1 HSPE1 NDUFAF4 TIMM13 TOMM5 PRELID1 TOMM6 ATP5G1 BID HSPD1 MRPL14 NDUFB2 ATP5I MRPL55 C1QBP CYC1 UQCC3 NDUFS5 COA6 NDUFB9 MRPL51 SLIRP UQCRH ATP5D ATP5J2 SOD1 NDUFAB1 NDUFS6 NDUFS8 NDUFA7 NDUFA9 FXN
ENO3 PGM2 TPI1 BPGM ENO1 PGM3 PFKL GPI ALDOA ENO2 PGK1 GALM ALDOB ALDOC PGM1 HK2 PGAM2
d
TMRM^Low
Intermembrane Space
4H+ 4H+ 2H+
I II III IV V
Mitochondrial Matrix
TMRM
TMRM^High
Intermembrane Space
Mitochondrial Matrix
TMRM
e
TMRM High (5.25%)
TMRM Low (74.6%)
TMRM
FSC
Fold Change (TMRM High/Low)
ESRRA1 NRF1 PGC1B ATP5F1A COX5B
f
Tumour
Micromet.
Count (Modal)
TMRM
P = 0.0331
Normalized Membrane Potential (TMRM / MitoTracker)
Tumour Micromet.

Figure 3.1 Micrometastatic cells display increased mitochondrial OXPHOS. a, Gene Ontology terms identified for primary tumor ($n = 214$ genes) and micrometastatic cells ($n = 116$ genes) based on the 330-micrometastasis-gene signature. The *P* values were determined using Fisher's exact test. **b,** Bar graph showing $\log_2$[FC] values for differentially expressed OXPHOS- and glycolysis-associated genes in micrometastatic cells ($n = 435$) relative to primary tumor cells ($n = 684$). $^*P < 0.05$, tobit test in Seurat (which utilizes a likelihood-ratio test) for genes in our 330-micrometastasis signature. **c,** Gene scores of 37 unique metabolic pathways in micrometastatic ($n = 435$) and primary tumor ($n = 684$) cells. The average gene scores were calculated on individual tumor or micrometastatic cells and averaged across the three PDX models. Significance is shown on the *y* axis. The *P* values were determined using a two-sided Wilcox test. **d,** Schematic illustrating the mechanism of TMRM fluorescent staining in mitochondria. Increased ETC activity increases proton efflux from the mitochondrial matrix, resulting in a greater membrane potential and further accumulation of TMRM in the matrix. **e,** Top, flow cytometry analysis of $TMRM_{high}$ and $TMRM_{low}$ populations in primary tumor cells from an HCI010 mouse. Data are representative of $n = 6$ biologically independent mice. Bottom, quantification of mitochondrial biogenesis and ETC genes in $TMRM_{high}$ relative to $TMRM_{low}$ tumor cells, determined by real-time quantitative reverse-transcription PCR; $n = 3$ technical replicates. **f,** Top, histogram overlay of TMRM-fluorescence intensity in lung micrometastatic and primary tumor cells from an HCI010 mouse. Data are representative of $n = 5$ biologically independent mice. Bottom, ratio of mean fluorescence intensity of TMRM to MitoTracker in $CD298^+MHC\text{-}I^-$ cells. The *P* value was determined using an unpaired two-sided Student's *t*-test. Data presented as the mean ± s.d. of $n = 6$ mice. One lung with no metastasis was excluded.

identified *LDHA*, which promotes aerobic glycolysis by catalyzing the conversion of pyruvate to lactate and diverting it from entry into the TCA cycle, as the top gene most descriptive of primary tumor cells **(Fig. 2.7d).**

To further investigate metabolic differences between primary tumor and micrometastases, I evaluated the expression of 1,402 genes associated with 37 metabolic pathways, such as the pentose phosphate pathway, the citric acid cycle, and fatty acid metabolism **(Fig. 3.1c, Fig. 3.2).** Gene scoring for each pathway showed that glycolysis (p = 2.20 x 10^{-16}) and OXPHOS (p = 2.20 x 10^{-16}) were the most significantly differentially expressed of all 37 pathways **(Fig. 3.1c).** Importantly, I found that >70% of genes were detected and passed quality filtering in 33 of the 37 pathways, indicating sufficient coverage across metabolic pathways for comparative analysis **(Fig. 3.2)**.

I next performed several studies for molecular validation of the metabolic difference between primary tumor and micrometastases. Since conventional oxygen consumption assays are not amenable to the small cell numbers extracted from micrometastases, I performed a combination of flow cytometry and high resolution metabolomic analyses. I compared mitochondrial activity in primary tumor and metastatic cells by flow cytometry using tetramethylrhodamine methyl ester (TMRM), which is a fluorescent dye that accumulates in the mitochondria as a result of a difference in membrane potential **(Fig. 3.1d).**[122] I find that HCI010 primary tumors contain $TMRM_{high}$ and $TMRM_{low}$ populations, where $TMRM_{high}$ cells are the minority population **(Fig. 3.1e).** qPCR analysis of sorted $TMRM_{high}$ and $TMRM_{low}$ cells confirmed that $TMRM_{high}$ cells express higher levels of genes involved in mitochondrial biogenesis and the ETC, validating the TMRM assay in PDX cells **(Fig. 3.1e).** I stained matched primary tumor and lung cell suspensions from six HCI010 PDX mice with TMRM. Cells were co-stained with a viability dye (Sytox), CD298, MHC-I and MitoTracker to identify viable human cells and assess mitochondrial membrane potential relative to mitochondrial mass**.** Flow cytometry analysis showed that micrometastases display 3-fold higher mitochondrial membrane potential than

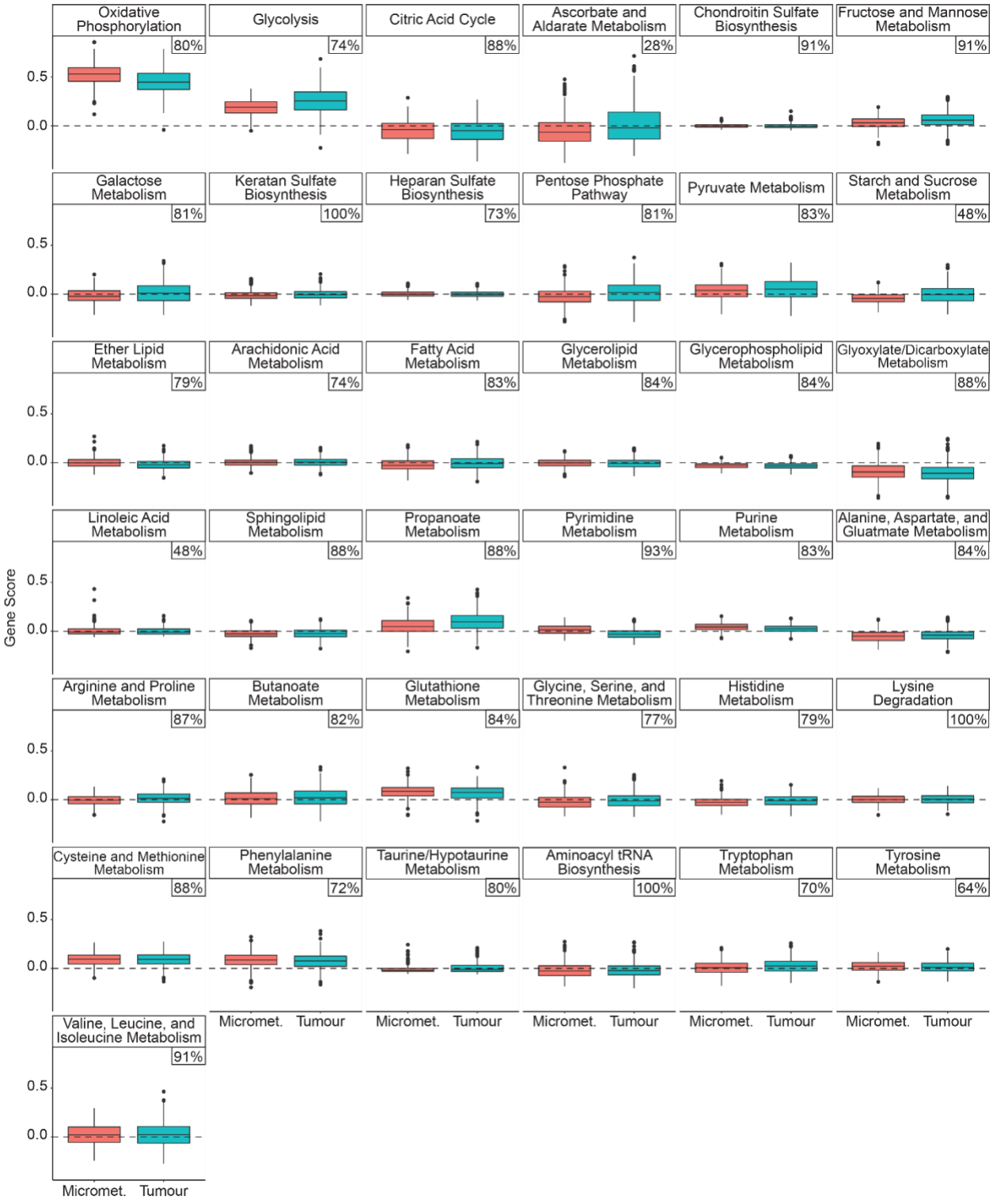
Oxidative Phosphorylation
80%
Glycolysis
74%
Citric Acid Cycle
88%
Ascorbate and Aldarate Metabolism
28%
Chondroitin Sulfate Biosynthesis
91%
Fructose and Mannose Metabolism
91%
Galactose Metabolism
81%
Keratan Sulfate Biosynthesis
100%
Heparan Sulfate Biosynthesis
73%
Pentose Phosphate Pathway
81%
Pyruvate Metabolism
83%
Starch and Sucrose Metabolism
48%
Ether Lipid Metabolism
79%
Arachidonic Acid Metabolism
74%
Fatty Acid Metabolism
83%
Glycerolipid Metabolism
84%
Glycerophospholipid Metabolism
84%
Glyoxylate/Dicarboxylate Metabolism
88%
Linoleic Acid Metabolism
48%
Sphingolipid Metabolism
88%
Propanoate Metabolism
88%
Pyrimidine Metabolism
93%
Purine Metabolism
83%
Alanine, Aspartate, and Gluatmate Metabolism
84%
Arginine and Proline Metabolism
87%
Butanoate Metabolism
82%
Glutathione Metabolism
84%
Glycine, Serine, and Threonine Metabolism
77%
Histidine Metabolism
79%
Lysine Degradation
100%
Cysteine and Methionine Metabolism
88%
Phenylalanine Metabolism
72%
Taurine/Hypotaurine Metabolism
80%
Aminoacyl tRNA Biosynthesis
100%
Tryptophan Metabolism
70%
Tyrosine Metabolism
64%
Valine, Leucine, and Isoleucine Metabolism
91%
Gene Score
0.5
0.0
Micromet.
Tumour

Figure 3.2 Gene scoring identifies OXPHOS and glycolysis as top metabolic pathways differentially expressed between micrometastases and primary tumor cells. Gene scores for each metabolic pathway in micrometastatic (red, $n = 435$ cells) or primary tumor cells (blue, $n = 684$ cells). Each cell in the dataset was scored by calculating the difference between the average gene expression for all the genes in each metabolic pathway versus the average gene expression of a randomly selected background gene set. Dotted line represents a zero score, which indicates the metabolic pathway is not differentially expressed relative to the background gene set. The boxed value (top right of each plot) indicates the percent of genes in the pathway that was detected in the dataset. The lower and upper hinges correspond to the first and third quartiles, and the midline represents the median. The upper and lower whiskers extend from the hinge up to 1.5 * IQR (inter-quartile range). Outlier points are indicated if they extend beyond this range.

primary tumor cells (*P* = 0.0331) **(Fig. 3.1f).** These data, combined with transcriptome analyses, strongly suggest that micrometastases utilize higher levels of OXPHOS for cellular metabolism.

3.3.2 Micrometastatic cells display a distinct metabolic profile

I next sought to determine whether micrometastatic cells are distinct from primary tumor cells at the global metabolic level. To this end, I developed a collaboration with the Mei Kong lab at UC Irvine and the Jason Locasale lab at Duke University to utilize liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS) to perform steady-state metabolomic analysis of cells isolated from the lungs and primary tumors of six HCI010 PDX animals **(Fig. 3.3a).**[120] Although less abundant metabolites fell below the limit of detection due to limited cell numbers available from metastatic lesions, analysis of the 150 identified metabolites showed that metastatic cells display a distinct metabolic profile **(Fig. 3.3b-d).** This revealed modestly higher levels of fumarate and malate in metastatic cells ($p<0.05$), which are intermediates of the citric acid cycle (CAC) that feeds into OXPHOS **(Fig. 3.3e, Supp. Table 6)**. We also identified 2.9-fold higher levels of succinyl carnitine in metastatic cells, which accumulates as a waste intermediate from the CAC **(Fig. 3.3e).** Glutamate, glucose and palmitolylcarnitine, an intermediate in fatty acid metabolism, were also significantly higher in metastatic cells **(Fig. 3.3e).** Glutamine also trended upward in metastatic cells compared to primary tumor cells, though did not achieve significance (*p*=0.303) (**Fig. 3.3e**). These metabolites are used by several pathways in the cell, but mainly provide carbon sources for the CAC to generate NADH/$FADH_2$ for energy production through OXPHOS **(Fig. 3.3f).** We observed no difference in pyruvate or lactate between primary tumor and metastatic cells **(Fig. 3.3e).** Interestingly, pathway analysis also showed increased amino acid metabolism in metastatic cells ($p<0.05$, FC>1.5) **(Fig. 3.3g).** Although amino acid accumulation may support diverse cellular functions, such as protein synthesis, signaling and proliferation, the catabolism of several of these amino acids such as cysteine, lysine, arginine, phenylalanine, and valine can also support the CAC and OXPHOS.[123] These data show that metastatic cells display a distinct

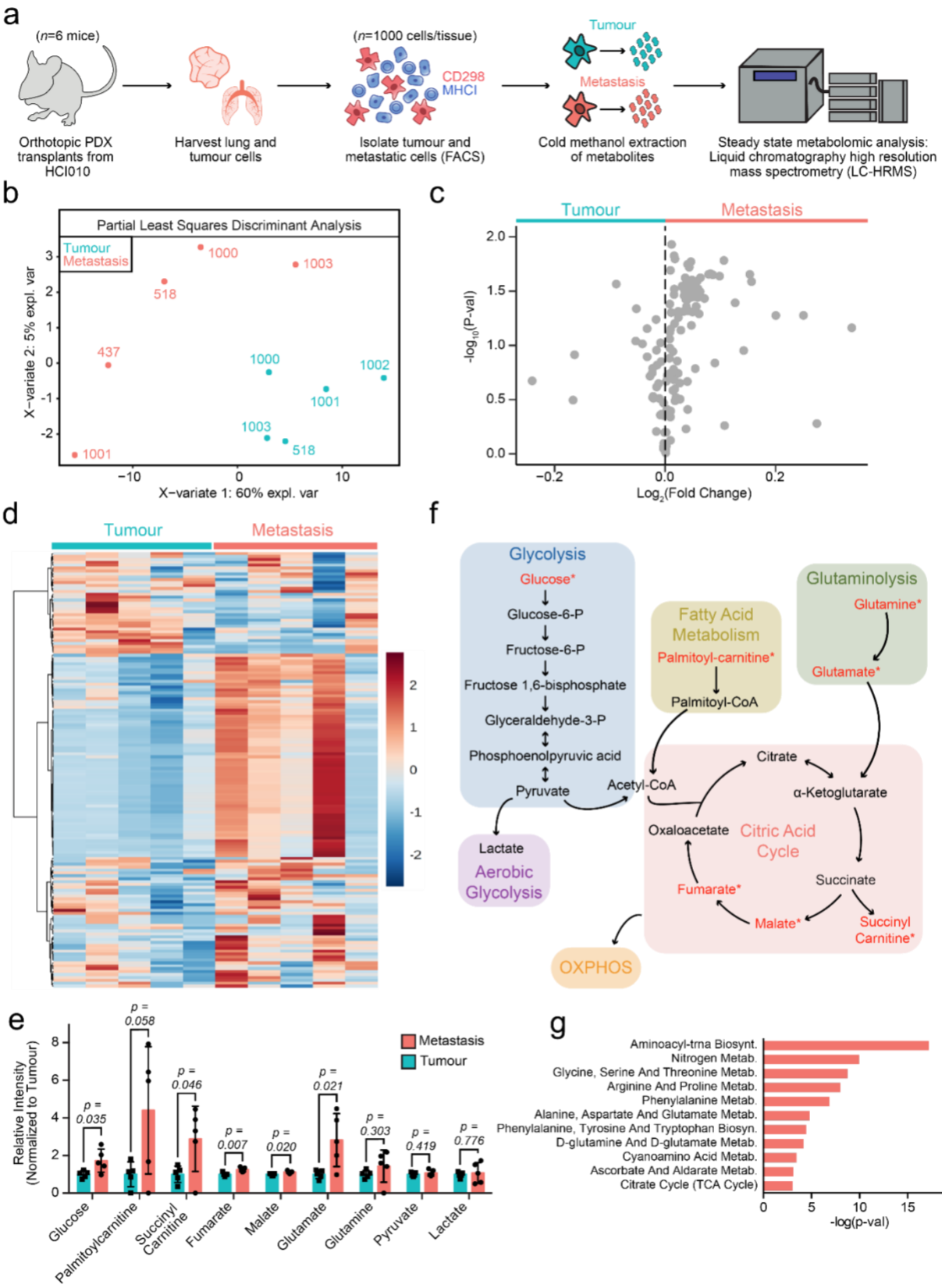
a
(n=6 mice)
Orthotopic PDX transplants from HCI010
Harvest lung and tumour cells
(n=1000 cells/tissue)
CD298
MHCI
Isolate tumour and metastatic cells (FACS)
Tumour
Metastasis
Cold methanol extraction of metabolites
Steady state metabolomic analysis: Liquid chromatography high resolution mass spectrometry (LC-HRMS)
b
Partial Least Squares Discriminant Analysis
Tumour
Metastasis
X−variate 2: 5% expl. var
X−variate 1: 60% expl. var
c
Tumour
Metastasis
-log10(P-val)
Log2(Fold Change)
d
Tumour
Metastasis
f
Glycolysis
Glucose*
Glucose-6-P
Fructose-6-P
Fructose 1,6-bisphosphate
Glyceraldehyde-3-P
Phosphoenolpyruvic acid
Pyruvate
Lactate
Aerobic Glycolysis
Fatty Acid Metabolism
Palmitoyl-carnitine*
Palmitoyl-CoA
Glutaminolysis
Glutamine*
Glutamate*
Acetyl-CoA
Citrate
α-Ketoglutarate
Oxaloacetate
Citric Acid Cycle
Succinate
Fumarate*
Malate*
Succinyl Carnitine*
OXPHOS
e
Relative Intensity (Normalized to Tumour)
Metastasis
Tumour
p = 0.035
p = 0.058
p = 0.046
p = 0.007
p = 0.020
p = 0.021
p = 0.303
p = 0.419
p = 0.776
Glucose
Palmitoylcarnitine
Succinyl Carnitine
Fumarate
Malate
Glutamate
Glutamine
Pyruvate
Lactate
g
Aminoacyl-trna Biosynt.
Nitrogen Metab.
Glycine, Serine And Threonine Metab.
Arginine And Proline Metab.
Phenylalanine Metab.
Alanine, Aspartate And Glutamate Metab.
Phenylalanine, Tyrosine And Tryptophan Biosyn.
D-glutamine And D-glutamate Metab.
Cyanoamino Acid Metab.
Ascorbate And Aldarate Metab.
Citrate Cycle (TCA Cycle)
-log(p-val)

Figure 3.3 Metastatic cells display a distinct metabolic profile. a, Schematic overview of the workflow for steady-state global metabolomics profiling of matched primary tumor and lung metastatic cells from six HCI010 PDX animals using LC–HRMS. **b,** Partial least-squares discriminant analysis of the global metabolite data in metastatic and tumor cells. Individual points represent the bulk profiles of $n = 1{,}000$ cells isolated from either the lung (metastatic) or primary tumor ($n = 6$ mice). The individual points are labelled with their mouse ID. One tumor and one metastatic sample were excluded from the analysis due to low signal. **c,** Volcano plot showing metabolites that were differentially detected in metastatic or tumor samples. The *P* values were determined using an unpaired two-sided Student's *t*-test. **d,** Heat map of peak intensities measured by LC–MS showing unsupervised clustering of 150 metabolites identified in metastatic and primary tumor cell samples normalized by autoscaling on MetaboAnalyst and ranked by a Student's *t*-test. **e,** Levels of selected glycolysis, CAC and lipid metabolism intermediates in metastatic cells relative to primary tumor cells represented as the normalized intensity mean ± s.d. of $n = 5$ micrometastatic and tumor samples. The values of individual samples were normalized to the tumor average. The *P* values were determined using an unpaired two-sided Student's *t*-test. **f,** Schematic showing the role of metabolites identified at higher levels in metastatic cells (indicated by an asterisk) in key metabolic pathways. **g,** Top metabolic pathways that were higher in metastatic cells identified by pathway analysis ($P < 0.05$; FC > 1.5; $n = 62$ metabolites). Metab., metabolism; biosynth., biosynthesis. The *P* values were determined using an unpaired two-sided Students *t*-test in MetaboAnalyst**.**

metabolic profile, which includes higher levels of several metabolites that may fuel oxidative phosphorylation.

3.3.3 Oxidative phosphorylation is critical for lung metastasis

I next investigated whether increased OXPHOS is functionally important for metastasis, or whether it simply represents a response to increased oxygen availability in the metastatic site. I used two experimental metastasis models of TNBC that metastasize to the lung with reproducible kinetics following intravenous (i.v.) delivery, mouse 4T1 and human MDA-MB-231 cells. I used the complex V inhibitor, oligomycin, to inhibit OXPHOS *in vitro* and then measured its effects on metastatic outgrowth in the lung **(Fig. 3.4a-b).** I chose oligomycin because it is a highly selective inhibitor of OXPHOS with limited off-target effects.[124] Oligomycin induces cells to shift to glycolysis to produce ATP, and has been used to study the consequences of this metabolic shift in diverse settings.[125–128]

Since oligomycin is cytotoxic if cells are unable to switch to increased glycolysis[125], I performed a series of analyses to test the health and metabolic state of MDA-MB-231 and 4T1 cells post treatment **(Fig. 3.5a-b)**. Flow cytometry analysis for propidium iodide (PI) and annexin V staining showed that oligomycin does not reduce cell viability, increase apoptosis, or alter cell morphology 72 hours post treatment **(Fig. 3.5c-d).** Proliferation assays further showed no decrease in the growth rate of treated cells **(Fig. 3.5e,g).** Using the Seahorse XF assay, we found that oligomycin induces a sustained reduction in the oxygen consumption rate (OCR) and a corresponding increase in the extracellular acidification rate (ECAR) **(Fig. 3.4c-d, Fig. 3.5f,h)**, suggesting the cells shift from OXPHOS to glycolysis**.** We further utilized phasor fluorescent lifetime imaging (FLIM) of NADH to determine the durability of the metabolic shift. FLIM relies on the fact that the fluorescent lifetime of NADH is longer when bound to enzymes involved in mitochondrial metabolism (~3.4ns) than when free floating in the cell during glycolysis (~0.4ns).[129,130] FLIM analysis 6-24 hours after treatment showed a significant decrease in bound NADH, indicating a durable shift from OXPHOS to glycolysis in both cell lines ($p < 0.05$) **(Fig.**

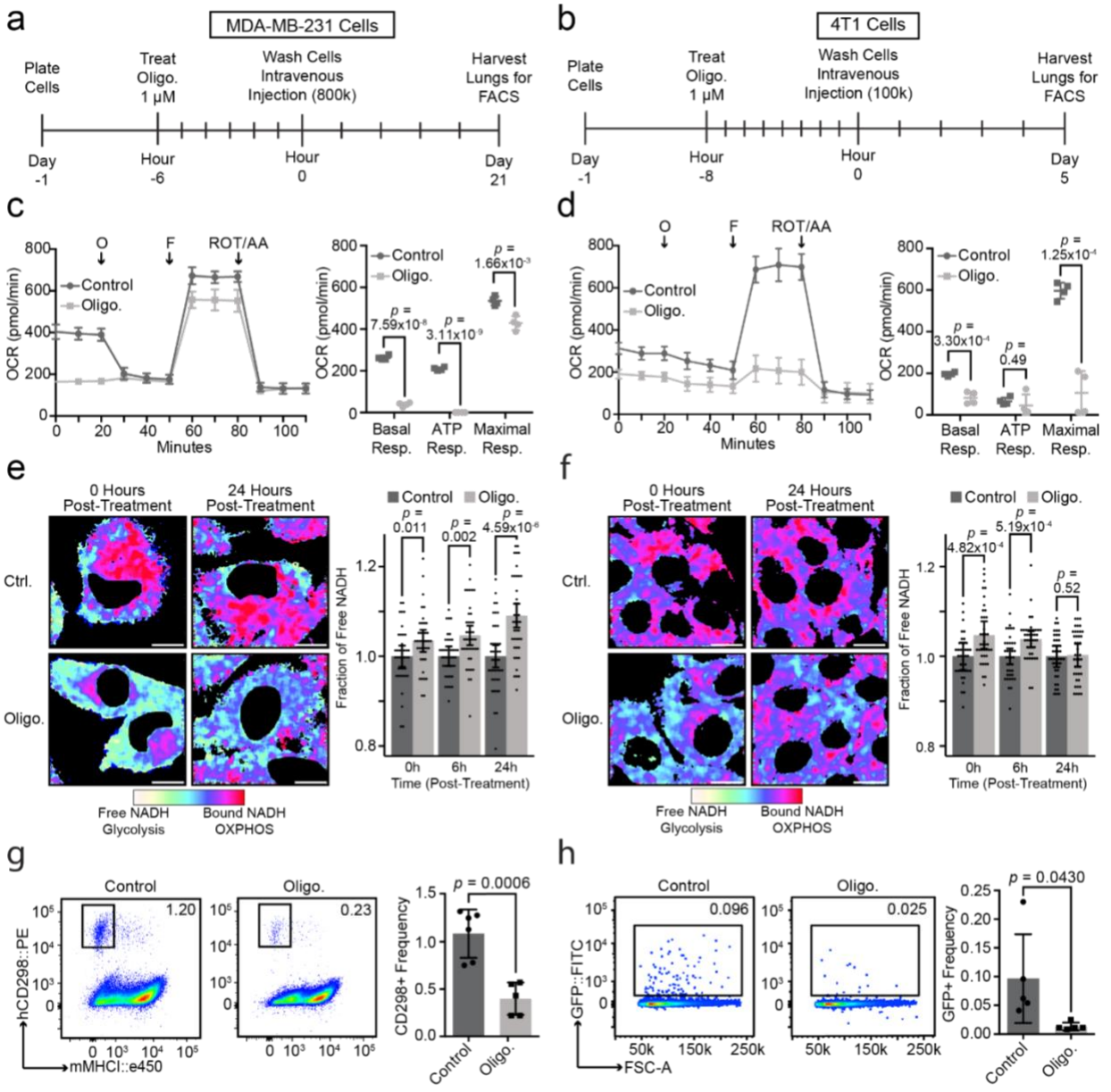

Figure 3.4 Oxidative phosphorylation is critical for lung metastasis. a,b, Schematic of the experimental approach used to determine the effects of complex V inhibition via oligomycin on metastasis of MDA-MB-231 **(a)** and 4T1–GFP cells **(b)**. **c,d,** Left, OCR of MDA-MB-231 **(c)** and 4T1–GFP cells **(d)** versus control cells after treatment with oligomycin for 6–8 h in vitro. O, oligomycin; F, FCCP; ROT/AA, rotenone with antimycin A. Right, basal, ATP and maximal respiration rates for each condition. The P values were determined using an unpaired two-sided Student's t-test. Data presented as the mean ± s.d. of n = 4 replicates. **e,f,** Left, representative FLIM images of the fluorescence lifetime of NADH in cultured MDA-MB-231 **(e)** or 4T1–GFP cells **(f)**. The fields do not represent consecutive images of the same cell. Scale bars, 10 μm. Right, quantification of the free-NADH fraction in the FLIM images normalized to the control. The P values were determined using a two-sided Student's t-test with homoscedasticity determined using an F-test. Data presented as the mean ± s.e.m.; t = 0 h, n = 39 fields; t = 6 h, n = 41 (MDA-MB-231) and 40 (4T1–GFP) fields; and t = 24 h, n = 42 fields. **g,h,** Flow cytometry analysis of the metastatic burden in the lungs of the animals injected intravenously with oligomycin-treated or control cells. Left, FACS plots show MDA-MB-231 cells **(g)**, identified as CD298+MHC-I+, and 4T1–GFP cells **(h)**, identified by GFP, in representative animals. Right, frequency of metastatic cells. Data presented as the mean ± s.d. of the frequency of MDA-MB-231 (n = 5 oligomycin-treated, one mouse without detectable micrometastases was excluded; and n = 6 control mice) and 4T1–GFP (n = 5 oligomycin-treated and control mice) cells. The P value was determined using an unpaired two-sided Student's t-test. Oligo., oligomycin.

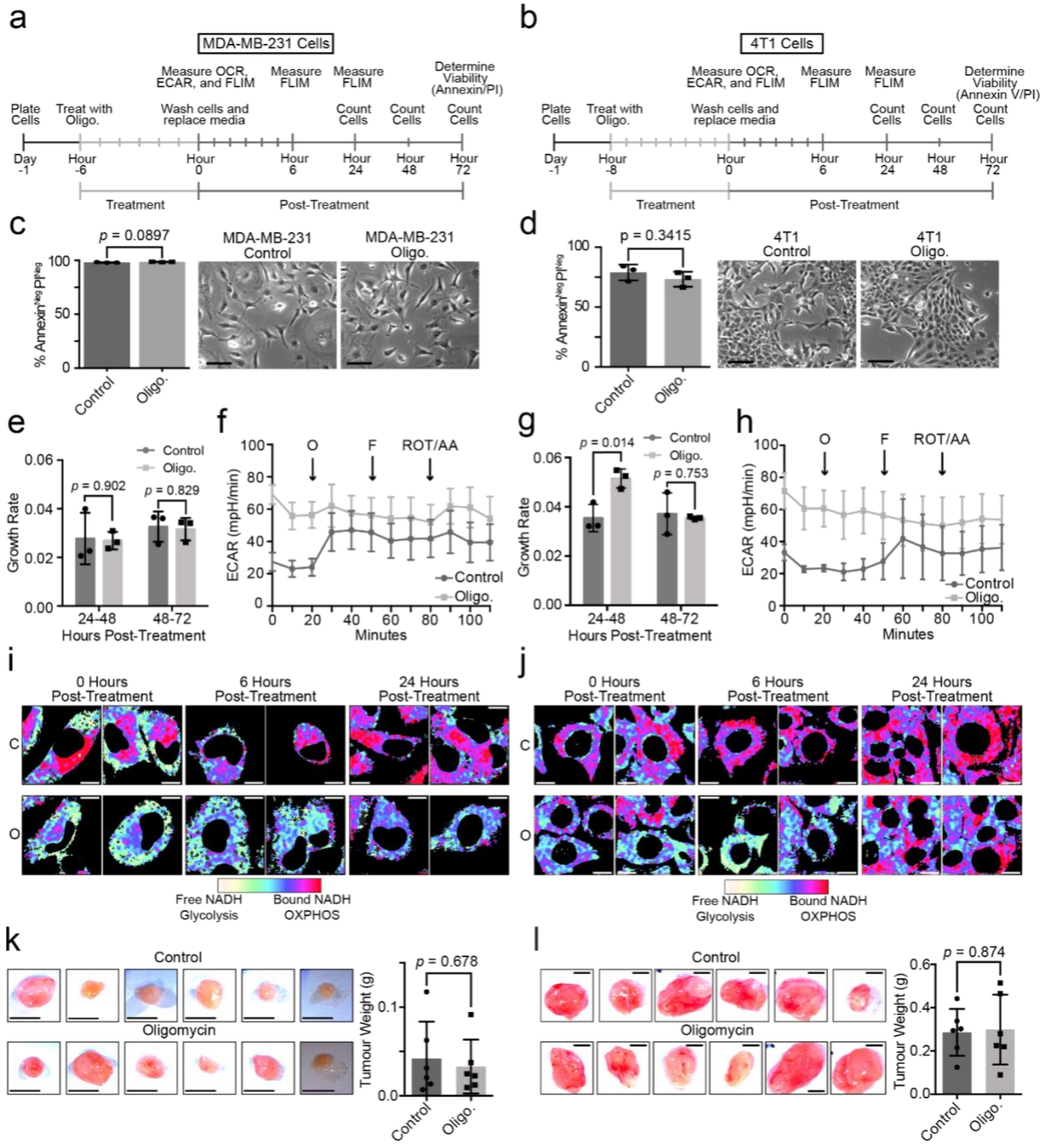
a
MDA-MB-231 Cells
Measure OCR, ECAR, and FLIM
Measure FLIM
Measure FLIM
Determine Viability (Annexin/PI)
Plate Cells
Treat with Oligo.
Wash cells and replace media
Count Cells
Count Cells
Count Cells
Day -1
Hour -6
Hour 0
Hour 6
Hour 24
Hour 48
Hour 72
Treatment
Post-Treatment
b
4T1 Cells
Measure OCR, ECAR, and FLIM
Measure FLIM
Measure FLIM
Determine Viability (Annexin V/PI)
Plate Cells
Treat with Oligo.
Wash cells and replace media
Count Cells
Count Cells
Count Cells
Day -1
Hour -8
Hour 0
Hour 6
Hour 24
Hour 48
Hour 72
Treatment
Post-Treatment
c
p = 0.0897
MDA-MB-231 Control
MDA-MB-231 Oligo.
% Annexin^Neg PI^Neg
Control
Oligo.
d
p = 0.3415
4T1 Control
4T1 Oligo.
% Annexin^Neg PI^Neg
Control
Oligo.
e
Control
Oligo.
p = 0.902
p = 0.829
Growth Rate
24-48
48-72
Hours Post-Treatment
f
O
F
ROT/AA
ECAR (mpH/min)
Control
Oligo.
Minutes
g
p = 0.014
p = 0.753
Control
Oligo.
Growth Rate
24-48
48-72
Hours Post-Treatment
h
O
F
ROT/AA
ECAR (mpH/min)
Control
Oligo.
Minutes
i
0 Hours Post-Treatment
6 Hours Post-Treatment
24 Hours Post-Treatment
C
O
Free NADH Glycolysis
Bound NADH OXPHOS
j
0 Hours Post-Treatment
6 Hours Post-Treatment
24 Hours Post-Treatment
C
O
Free NADH Glycolysis
Bound NADH OXPHOS
k
Control
Oligomycin
p = 0.678
Tumour Weight (g)
Control
Oligo.
l
Control
Oligomycin
p = 0.874
Tumour Weight (g)
Control
Oligo.

Figure 3.5 Oligomycin treatment inhibits OXPHOS in MDA-MB-231 and 4T1-GFP cells. a,b, Schematic of the experimental setup to determine the effects of oligomycin treatment on MDA-MB-231 **(a)** or 4T1-GFP **(b)** cells. Oligo. = oligomycin. **c,d,** Bar graph (left) of the viability of MDA-MB-231 **(c)** or 4T1- GFP **(d)** cells 72-hours post-treatment determined via flow cytometry. Data is presented as the mean ± s.d. for n = 3 replicates. *P* values determined by unpaired, two- sided Student's t-test. Brightfield images (right) of in vitro MDA-MB-231 **(c)** or 4T1-GFP **(d)** cells 72 hours after cessation of oligomycin treatment. Scale bar = 100 µm. **e,** Growth rate of MDA-MB-231 cells for the indicated time after cessation of oligomycin treatment. Data is presented as mean ± s.d. of n = 3 replicates. *P* values determined by unpaired, two-sided Students t-test. **f,** Extracellular acidification rate (ECAR) of MDA-MB-231 treated with oligomycin compared to control cells. ECAR was measured at the conclusion of treatment with oligomycin as described in **(a)**. Arrows indicate when drugs were added. O=oligomycin, F=FCCP, ROT/AA=Rotenone/ Antimycin A. Data is presented as mean ± s.d. of n = 4 replicates. **g,** Same as **(e)** for 4T1-GFP cells. **h,** Same as **(f)** for 4T1-GFP cells. Data is presented as mean ± s.d. of n = 3 replicates. *P* values determined by unpaired, two-sided Students t-test. **i, j,** Additional FLIM images of the fluorescence lifetime of NADH in cultured MDA-MB-231 **(i)** or 4T1-GFP cells **(j)** as shown in Fig. 3.3e, f. Fields do not represent consecutive images of the same cell. O=oligomycin, C=control. Scale bar=10 µm. **k, l,** Brightfield images of tumors from orthotopically injected control or oligomycin treated MDA-MB-231 **(k)** or 4T1-GFP **(l)** cells. Bar graphs indicate tumor weights (right). Data presented as mean ± s.d. of MDA-MB-231 (n = 6 oligomycin-treated, n = 6 control) and 4T1-GFP (n = 6 oligomycin-treated, n = 6 control) tumors. *P*-values were determined by

3.4e-f, Fig. 3.5i-j). Collectively, these data demonstrate that oligomycin induces MDA-MB-231 and 4T1 cells to shift from OXPHOS to glycolytic metabolism without compromising cell viability or proliferation.

I injected NOD/SCID mice i.v. with oligomycin-treated versus control MDA-MB-231 cells ($8x10_5$) and harvested lungs 21 days later **(Fig. 3.4a).** Flow cytometry analysis showed a nearly 3-fold decrease in the frequency of metastatic cells in the lungs of the treatment group **(Fig. 3.4g)**, showing that OXPHOS inhibition attenuates the metastatic capacity of MDA-MB-231 cells**.** I used a similar approach to test the effects of OXPHOS inhibition on 4T1 cell metastasis **(Fig. 3.4b)**. I injected GFP-labeled 4T1 cells ($1x10_5$) into BALB/c mice and harvested lungs 5 days later since 4T1 cells metastasize rapidly in the lung. Remarkably, animals injected with oligomycin-treated 4T1-GFP cells displayed a 7-fold decrease in metastatic cells in the lung at this early timepoint **(Fig. 3.4h),** suggesting OXPHOS is important to facilitate early events in the metastatic cascade.

Finally, to determine whether OXPHOS is specifically important for metastasis or generally important for cancer cell function, I tested the effects of oligomycin treatment on primary tumor growth. I treated 4T1 and MDA-MB-231 cells with oligomycin *in vitro*, performed bilateral orthotopic injections ($1x10_6$ cells), and harvested tumors 14 and 17 days later, respectively **(Fig. 3.5k-l)**. I found no significant difference in engraftment efficiency or tumor size in control versus oligomycin-treated tumors **(Fig. 3.5k-l),** further establishing a role for OXPHOS in mediating the lung metastasis **(Fig. 3.6)**.

3.4 Discussion

Defining the molecular mechanisms that facilitate specific stages of the metastatic cascade is critical to understand how cells metastasize and inspire new strategies to prevent metastatic spread in cancer patients. I have developed a robust approach for the capture and analysis of rare cells during the seeding and establishment of micrometastasis in PDX models using scRNA-seq. I find that micrometastases display significant transcriptional heterogeneity,

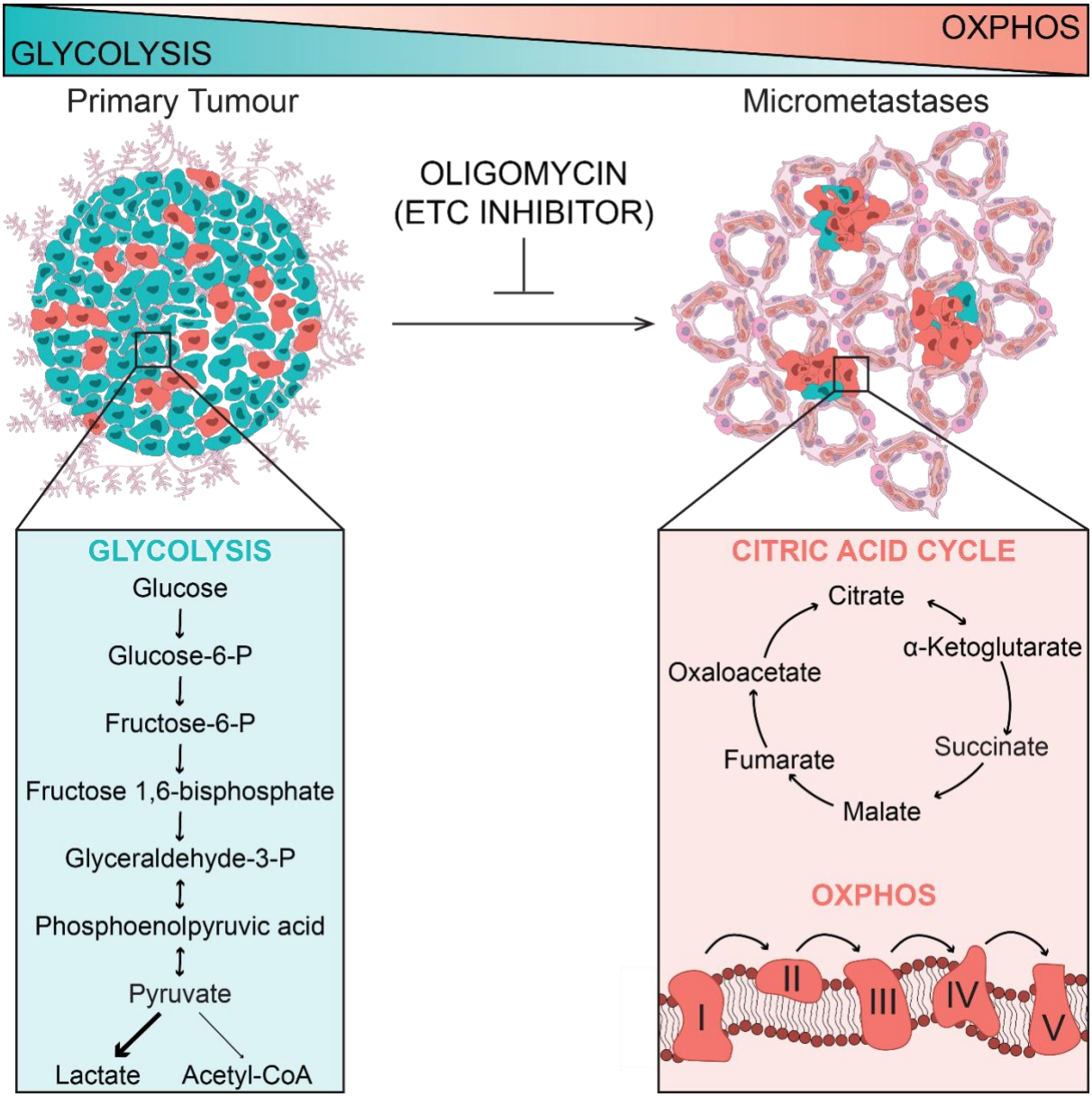

Figure 3.6 Model for metabolic shift associated with metastatic seeding in TNBC. Metastatic cells in the lung and lymph nodes display increased OXPHOS, in contrast to primary tumor cells that express higher levels of genes associated with aerobic glycolysis. Pharmacological inhibition of OXPHOS with the complex V inhibitor oligomycin substantially attenuates lung metastasis in experimental models of TNBC, showing that OXPHOS is functionally important for metastasis.

but also display a distinct expression program that is predictive of poor survival in breast cancer patients. I identify mitochondrial OXPHOS as the top pathway upregulated in micrometastases, in contrast to primary tumor cells that score higher for genes associated with aerobic glycolysis **(Fig. 3.6).** With the help of collaborators, we validated this by flow cytometry and LC-HRMS, where we found that micrometastases display a distinct metabolic profile and higher levels of several metabolites that feed into OXPHOS. Most importantly, I find that pharmacologic inhibition of OXPHOS substantially attenuates lung metastasis, showing that OXPHOS is functionally critical for metastatic spread.

My work sheds light on the controversial role of metabolism in metastasis. Historically, tumors were thought to contain dysfunctional mitochondria and be principally sustained by Warburg metabolism.[62,63] Recent work has challenged this dogma, showing the importance of glutamine[76,77,82], fatty acid[73,131], proline[132], and pyruvate carboxylase-mediated metabolism[133] in metastatic progression.[134,135] Recent work using the 4T1 mouse model also showed that circulating tumor cells (CTCs) utilize increased OXPHOS, and that suppression of a central regulator of mitochondrial biogenesis (PGC1α) reduced cancer cell invasion and metastasis.[136] Furthermore, lung micrometastases have increased expression of mitochondrial complex I and oxidative stress by *in situ* RNA sequencing.[137] My work utilized an unbiased transcriptomic approach to discover that OXPHOS promotes metastasis specifically during the seeding step of the metastatic cascade in three distinct human patient models of TNBC. Interestingly, many metabolic pathways implicated in metastasis, such as glutamine, fatty acid, and proline metabolism also converge on or produce critical metabolites to drive OXPHOS, raising the question of whether OXPHOS represents an underlying mechanistic link between them.

OXPHOS may promote metastatic seeding in several ways. Increased ATP generation through OXPHOS may provide energy for cytoskeleton remodeling for motility, or to survive anoikis during cell detachment and migration.[80] Increased ROS production through OXPHOS may promote cell motility by activating oncogenic signaling, as mtROS-inducing mutations are

sufficient to trigger metastasis.[23,138–141] Epidemiological data in humans also supports a role for OXPHOS in cancer progression, showing that treatment with the diabetes drug metformin (complex I inhibitor) is protective against breast cancer relapse and metastasis.[142–145] Several drugs targeting mitochondrial metabolism are already FDA approved or in clinical trials (e.g., IACS-010759[146], plumbagin[147,148], atovaquone[149,150]), providing an opportunity for re-indication in breast cancer patients to prevent metastasis.

On important caveat of my approach to demonstrating the deleterious effects of OXPHOS inhibition on metastatic potential involves the utilization of an i.v. injection model. I was able to demonstrate that OXPHOS activity is necessary at any step along the metastatic cascade from survival in circulation, to extravasation, and seeding at foreign tissues (lung). This method does not include local invasion or intravasation. Interestingly, my data does demonstrate that while the OXPHOS phenotype is enriched highly in metastatic cells, there does exist primary tumor cells with enhanced oxidative capacity **(Fig. 3.1e-f, Fig. 3.2).** It is tempting to speculate whether these cells are in fact the subset of primary tumor cells with the highest potential to form metastases and represent a novel therapeutic target. Nevertheless, further studies are needed to define which specific steps of the metastatic cascade OXPHOS is critical for, to determine when it might be applied for clinical benefit against metastasis.

Reprinted with permission from:

Davis, R.T., Blake, K., Ma, D. et al. Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell RNA sequencing. Nat Cell Biol 22, 310–320 (2020).

Chapter 4: Conclusion and Final Remarks

Metastasis remains the primary cause of death in almost all human cancers, including breast cancer. While many aspects of the metastatic cascade have been investigated, the mechanisms driving survival of micrometastases at distal tissues remains poorly understood. The long-term goal of my thesis project was to define intratumoral heterogeneity present in primary tumors and micrometastatic cells and define novel molecular signatures present in micrometastatic cells. To that end, I performed a comprehensive analysis of 1,119 single cell transcriptomes representing primary tumor and micrometastatic lesions from three distinct PDX models. I identified a conserved micrometastatic signature associated with patient outcomes in a large cohort of TNBC patient tumors. I pursued a bioenergetic shift associated with micrometastatic cells and performed extensive molecular and functional studies to evaluate OXPHOS in micrometastases and investigate its functional importance for metastatic seeding.

Chapter 2 of this thesis explored the intratumoral heterogeneity present in primary tumor cells compared to micrometastatic cells. Identifying models which can recapitulate the early stages of metastasis post-extravasation, prior to secondary tumor formation has classically been difficult in the field. I found that by performing orthotopic transplantation of three previously identified TNBC PDX models (HCI001, HCI002, HCI010), the spontaneous metastases that develop mirror the micrometastatic stage. I developed a method for generating global transcriptome information from matched primary tumor and micrometastatic cells from three distinct TNBC patient samples (scRNA-DeepSeq). Using bioinformatic analysis, I identified heterogeneity in all three PDX models, comprising three (A1-A3), five (B1-B5), and six (C1-C6) distinct cell states in HCI001, HCI002, and HCI010, respectively. Interestingly, I found that these heterogeneous cell states were not exclusive to primary tumors, but rather shared between primary tumor and micrometastatic lesions at different frequencies. While some cell states were unique to specific PDX models, I identified a common EMT signature associated with primary

tumor-enriched cell states in all three models. This EMT signature was highly correlated with ECM remodeling signature, indicating a potential feed-forward mechanism in these cells to enhance EMT activity. I demonstrated subtype heterogeneity in all three PDX models. Though each tumor can be broadly classified as basal-like/TNBC tumors, subpopulations of both micrometastatic and tumor cells classified as HER2 or Luminal A/Luminal B, indicating a potential for targeted therapies to eliminate segments of the tumor. Importantly, I identified a micrometastatic gene signature, comprising 330 genes, which demonstrated high predictive power in stratifying TNBC patient outcomes from primary tumor gene profiling. I further validated a key component of this signature, *PHLDA2*, in micrometastatic lesions and primary tumor cells. I found *PHLDA2* to be highly enriched in micrometastatic lesions, but also identified rare cells within the primary tumor with similar expression levels, which hints at the possibility of these cells having gained higher metastatic potential. Experimentally, this can be determined via overexpression of PHLDA2 in tumor cells and determining their capacity for spontaneous and experimental metastasis models compared to control tumor cells. Indeed, this work has already been started by my colleagues, who have developed an *in* vitro culture system of the PDX cells, which enables genetic manipulation. Preliminary data indicates that overexpression of *PHLDA2* in both the HCI001 and HCI010 patient models, followed by orthotopic transplantation into NSG mice, results in substantial increases in metastatic burden in the lungs of these mice (data not shown). Importantly, this increase in burden is not a result of differences in tumor growth rates or final volumes (data not shown). Collectively, the results outlined in Chapter 2 provide a valuable resource to the community to explore intratumoral heterogeneity in the context of metastasis, and a novel gene signature with potential clinical benefits.

Chapter 3 builds upon the findings in Chapter 2 by attempting to determine a conserved feature of the 330 gene micrometastatic signature. Using GO term analysis, I identify a bioenergetic shift from glycolysis to OXPHOS as the central theme present in the signature. This switch is highly specific at the transcriptional level, as I demonstrate that glycolysis and

OXPHOS are the predominate differentiator between primary tumor and micrometastatic cells when compared to 37 metabolic pathways comprising over 1,400 genes. I find that micrometastatic cells upregulated several genes involved in mitochondrial function and the ETC. Given the low cell inputs obtained by micrometastatic lesions, conventional metabolic readouts (i.e. Seahorse) were not amenable to this study. Rather, I turned to using a fluorescent readout of mitochondrial activity, TMRM, which is sensitive to mitochondrial membrane potential. I find that primary tumor cells are characterized by enrichment for $TMRM_{Low}$ cells, with less than 10% of tumor cells categorized as $TMRM_{High}$ by flow cytometry. This contrasts with micrometastatic lesions, which show significantly higher mitochondrial membrane potential, after controlling for mitochondrial mass. To fully explore whether this bioenergetic shift was relevant at the metabolite level, I developed a collaboration to generate steady-state metabolomics data from extremely low cell inputs (1000 cells/tissue). This data demonstrated that micrometastatic cells show significant enrichment for metabolites involved in CAC activity, glutamine metabolism and fatty acid metabolism. In agreement with my transcriptome analysis, these pathways are all involved in generating metabolites and intermediates for OXPHOS. Finally, I demonstrated the functional significance of OXPHOS metabolism *in vivo*. I showed that inhibition of OXPHOS with oligomycin significantly impairs metastatic capacity in both an immune-compromised and immune-competent setting, without impairing primary tumor development. The results outlined in Chapter 3 demonstrate a bioenergetic shift associated with micrometastasis and that this shift is necessary for efficient metastatic dissemination. Furthermore, these results indicate a potential clinical benefit for exploring inhibitors of OXPHOS metabolism in patients with disseminated disease.

The gene signature associated with micrometastatic cells is likely driven by either genetic or epigenetic factors. An intriguing mechanistic link between the gene signature I identified in Chapter 2 and the metabolic shift explored in Chapter 3 involves changes to the epigenetic landscape driven by the CAC. Acetyl-CoA can be utilized for the acetylation of

histones, and α-ketoglutarate is a key component of several dioxygenases, such as *TET1* and *KDM2*, which in turn are responsible for DNA and histone demethylation, respectively.[151] It is possible that the metabolic adaptations acquired by the metastatic cells drive epigenetic rewiring of the cells to initiate changes in gene expression that drive the cell towards the micrometastatic gene signature, including upregulation of genes like *PHLDA2*. Future studies will be needed to explore the epigenetic landscape of micrometastatic cells compared to primary tumor cells, to determine which modifications are highly abundant, and in turn which modifications could potentially be modulated by the metabolic phenotype of micrometastases. Experimentally, this can be accomplished by profiling the epigenome of these cells with advances in single cell epigenetics (scATAC-seq) in micrometastatic compared to primary tumor cells in these PDX models I have previously described. One could then modulate the levels of OXPHOS *in vitro* and perform a similar scATAC-seq assay to determine which specific regions of the epigenome are altered by the bioenergetic shift in micrometastases, thereby linking any identified changes to alterations in the 330 gene micrometastatic signature described in Chapter 2.

Finally, another key factor that may explain the mechanistic benefits of enhanced CAC and OXPHOS activity seen in these micrometastatic cells involves the immune system. In Chapter 3 I demonstrated that inhibition of OXPHOS in 4T1-GFP cells significantly impaired their ability to metastasize in the context of a fully competent immune system, indicating that the role of OXPHOS is relevant with an active immune system. Several models exist to explain this phenomenon, many of which will like work synergistically, to enable the micrometastatic cells to evade immune clearance. Increased OXPHOS and CAC activity results in increased ATP levels in micrometastatic cells. These excess ATP molecules may be secreted or released upon cell death, where they can be cleaved by ectonucleotidases (CD73 and CD39) into adenosine, which in turn can bind A2a receptors on both CD8 T cells and NK cells, thus inhibiting their activity.[152] The logical hypothesis from this is that blocking CD73 and CD39 will prevent the

functional advantage provided by increased ATP production in highly oxidative micrometastatic cells. This can be tested experimentally *in vitro* using recently described monoclonal antibodies which block CD73 (IPH5301) and CD39 (IPH4201) activity.[153] Following promotion of OXPHOS metabolism *in vitro* (i.e. dichloroacetate (DCA) treatment)[154], one would treat the cells with IPH5301/IPH4201 and read out cytotoxicity towards TNBC cell lines in co-cultures with activated NK cells and T cells by flow cytometry (**Fig. 4.1**).[155] This experiment would result in the cytotoxic levels of T and NK cells towards four different conditions: (1) Untreated cells ($DCA^{-}IPH^{-}$), (2) OXPHOS-enhanced cells ($DCA^{+}IPH^{-}$), (3) antibody-treated control cells ($DCA^{-}IPH^{+}$), and (4) antibody-treated, OXPHOS-enhanced cells ($DCA^{+}IPH^{+}$) (**Fig. 4.1**). If increased ATP production through enhanced OXPHOS activity in metastases results in inhibition of NK and T cell cytotoxicity in a CD73/CD39 dependent manner, there should be a significant increase in cytotoxicity between conditions 2 and 4 (**Fig. 4.1**). Alternatively, it is possible that there is no difference between these groups, suggesting that the activity of CD73/CD39 may not be a tumor-cell intrinsic factor, but rather a microenvironmental factor. This second point cannot be validated by this *in vitro* set-up, and thus would require performing a similar experimental approach *in vivo* (as the IPH antibodies have been demonstrated to reach effective concentrations *in vivo*[153]). By using the *in vivo* model, the effects of stromal cells (which have been demonstrated to express CD39/CD73[152]) can be included.

Additionally, increased mtROS produced by ETC activity may be converted to H_2O_2 and similarly accumulate in the extracellular space, in turn inhibiting T cell function.[156] The primary ROS produced by OXPHOS, superoxide, has also been demonstrated to promote motility of tumor cells in a mtSRC/PYK2 dependent manner.[138] This leads the hypothesis that increased mtROS production is required for enhanced metastatic seeding capacity. This can be determined experimentally by isolating cells based on their levels of mtROS using mitoSOX, a mitochondrially-targeted dye which fluoresces in the presence of superoxide and has been used extensively to study mtROS levels in cells.[157] These cells can then be introduced I.V. into mice

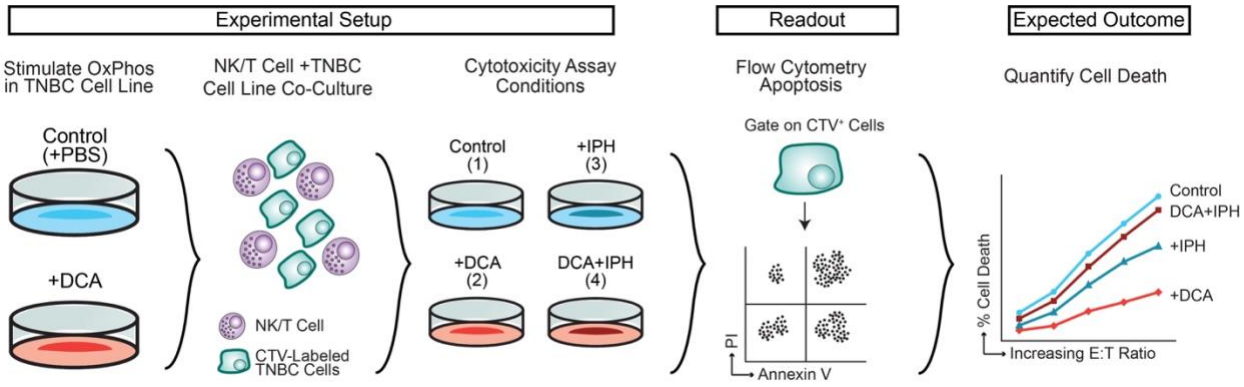

Figure 4.1 Experimental scheme to test the contribution of CD39/CD73 towards enhanced metastatic potential. Treatment of cells with dichloroacetate (DCA) promotes OXPHOS metabolism by inhibiting PDK and fostering mitochondrial oxidation of pyruvate. To distinguish tumor cell viability and apoptosis from the co-culture with NK/T cells, CellTracker Violet (CTV) is used to label tumor cells fluorescently prior to the cytotoxicity assay. Addition of IPH5301 and IPH4201 (IPH) preferentially inhibits CD39/CD73 activity in culture. Viability of the TNBC cell line is determined by PI and Annexin V staining, with increasing cell death occurring with increasing NK/T cell (effector, E) to TNBC cell line (target, T) ratio.

and their ability to form micrometastases can be determined by flow cytometry (**Fig. 4.2a**). Interestingly, previous studies exploring the gene expression differences between primary tumor cells, CTCs, and advanced macrometastasis found similar bioenergetic shift towards OXPHOS metabolism in CTCs.[136] Expectedly, they saw an upregulation in glycolysis within the primary tumor, but somewhat surprisingly, they found a reversion to a glycolytic phenotype in macrometastasis. Combined with my findings, this results in a cyclic model of metabolic phenotypes as tumor cells transition from local to circulation to seeding to colonization, suggesting that different metabolic adaptations are critical a differential stages of the metastatic cascade. Given the previous studies demonstrating a role for ROS in promoting migration[138], it is logical to hypothesize that the mtROS generated by enhanced OXPHOS activity in micrometastases is required for efficient metastatic seeding, but is be dispensable for metastatic colonization. This can be determined experimentally by assaying the levels of ROS in a controlled, temporally separated experimental metastasis model **(Fig. 4.2b).**

Lastly, studies have been published demonstrating a key link between the metabolic behavior of tumor cells and immunomodulatory effects on cytotoxic cells, such as CD8 T cells, which in turn links back the previous hypothesis regarding CD39/CD73.[158,159] It will be important to further explore mechanisms of dysregulation of immune responses resulting the bioenergetic shift present in micrometastatic cells to better develop treatment options which enhance the immune system's ability to recognize and clear these micrometastatic lesions before the develop into lethal macrometastatic disease.

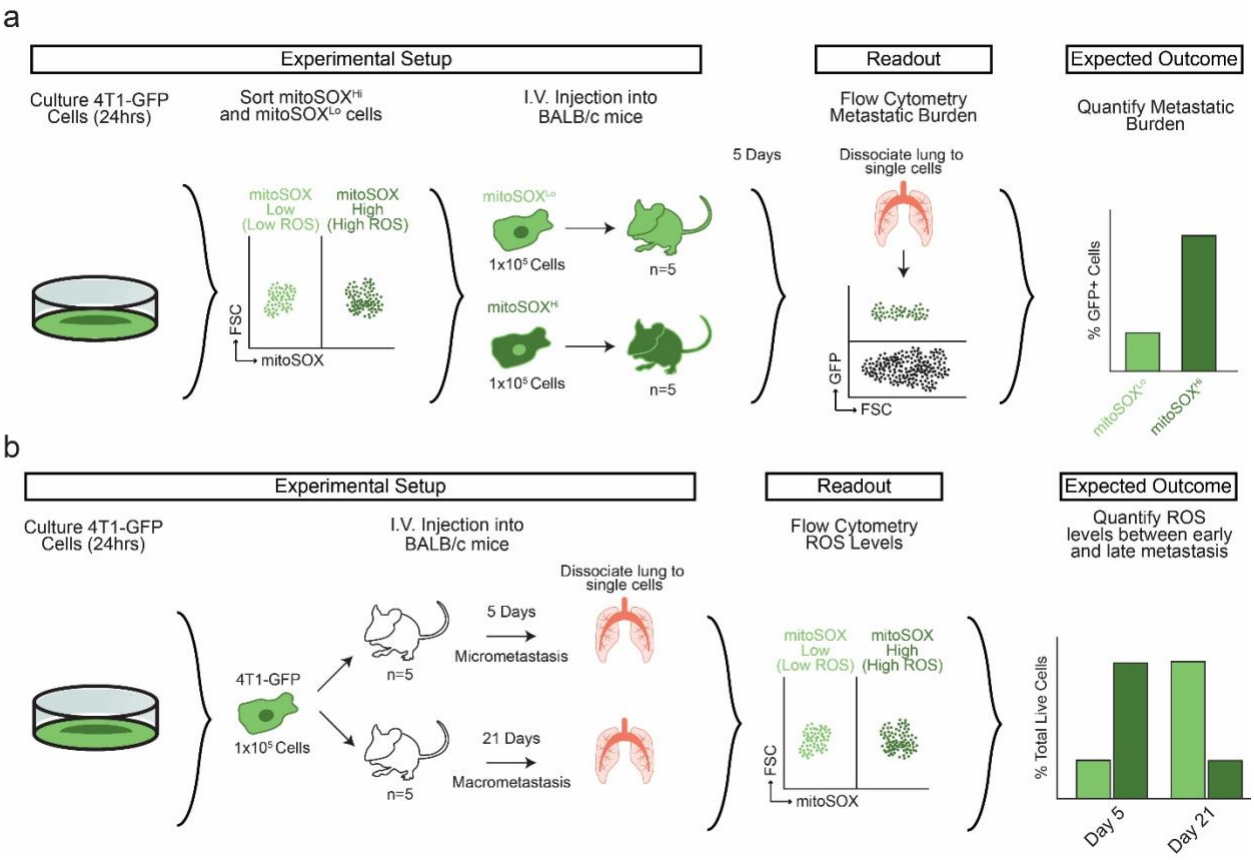

Figure 4.2 Experimental schemes to test the contribution of mtROS production towards enhanced metastatic potential. a, Experimental scheme to test whether increased mtROS levels confer a functional advantage towards metastatic seeding. Following culture of 4T1 cells *in vitro*, individual cells are sorted by FACS based on their levels of mitoSOX, a mitochondrially-targeted fluorescent probe for superoxide species. $1x10_5$ cells are injected intravenously (I.V.) into BALB/c mice (n=5 per group) and allowed to seed and establish micrometastases (5 days). The levels of metastatic burden in the lungs are determined by the frequency of GFP+ cells. **b,** Experimental scheme to test whether mtROS levels decrease from seeding to colonization during metastasis. Equal numbers of 4T1-GFP cells (n=$1x10_5$ cells) are injected I.V. into BALB/c mice (n=5 per group) and allowed to form either micrometastases (5 days) or macrometastases (21 days). The levels of mtROS in each group is quantified by flow cytometry using mitoSOX.

References

1. Malanchi, I. *et al.* Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature* **481**, 85–89 (2012).
2. Lawson, D. A. *et al.* Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* **526**, 131–135 (2015).
3. Hermann, P. C. *et al.* Distinct Populations of Cancer Stem Cells Determine Tumor Growth and Metastatic Activity in Human Pancreatic Cancer. *Cell Stem Cell* **1**, 313–323 (2007).
4. Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging Biological Principles of Metastasis. *Cell* **168**, 670–691 (2017).
5. Chambers, A. F., Groom, A. C. & MacDonald, I. C. Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer* **2**, 563–572 (2002).
6. Luzzi, K. J. *et al.* Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am J Pathol* **153**, 865–873 (1998).
7. Malladi, S. *et al.* Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. *Cell* **165**, 45–60 (2016).
8. DeRose, Y. S. *et al.* Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nat Med* **17**, 1514–1520 (2011).
9. Lodhia, K. A., Hadley, A. M., Haluska, P. & Scott, C. L. Prioritizing therapeutic targets using patient-derived xenograft models. *Biochim. Biophys. Acta - Rev. Cancer* **1855**, 223–234 (2015).
10. Kim, M. P. *et al.* Generation of orthotopic and heterotopic human pancreatic cancer xenografts in immunodeficient mice. *Nat. Protoc.* **4**, 1670 (2009).

11. Stebbing, J. *et al.* Patient-derived xenografts for individualized care in advanced sarcoma. *Cancer* **120**, 2006–2015 (2014).
12. Hochhauser, D. & Caldas, C. Of mice and men: patient-derived xenografts in cancer medicine. *Ann. Oncol.* **28**, 2330–2331 (2017).
13. American Cancer Society. *Cancer Facts and Figures 2020*. (2020).
14. Weigelt, B., Baehner, F. L. & Reis-Filho, J. S. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. *J Pathol* **220**, 263–280 (2010).
15. Sotiriou, C. & Pusztai, L. Gene-expression signatures in breast cancer. *N Engl J Med* **360**, 790–800 (2009).
16. Abbas, S. *et al.* Serum 25-hydroxyvitamin D and risk of post-menopausal breast cancer--results of a large case-control study. *Carcinogenesis* **29**, 93–99 (2008).
17. Brenton, J. D., Carey, L. A., Ahmed, A. A. & Caldas, C. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J Clin Oncol* **23**, 7350–7360 (2005).
18. Goss, P. E. *et al.* Exemestane for breast-cancer prevention in postmenopausal women. *N Engl J Med* **364**, 2381–2391 (2011).
19. Moasser, M. M. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene* **26**, 6469–6487 (2007).
20. Starczynski, J. *et al.* HER2 gene amplification in breast cancer: a rogues' gallery of challenging diagnostic cases: UKNEQAS interpretation guidelines and research recommendations. *Am J Clin Pathol* **137**, 595–605 (2012).
21. Nagata, Y. *et al.* PTEN activation contributes to tumor inhibition by trastuzumab, and loss of PTEN predicts trastuzumab resistance in patients. *Cancer Cell* **6**, 117–127 (2004).
22. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).

23. Dai, X. *et al.* Breast cancer intrinsic subtype classification, clinical use and future trends. *Am. J. Cancer Res.* **5**, 2929–2943 (2015).

24. Wahba, H. A. & El-Hadaad, H. A. Current approaches in treatment of triple-negative breast cancer. *Cancer Biol Med* **12**, 106–116 (2015).

25. Ahn, S. G., Kim, S. J., Kim, C. & Jeong, J. Molecular Classification of Triple-Negative Breast Cancer. *J Breast Cancer* **19**, 223–230 (2016).

26. Dent, R. *et al.* Pattern of metastatic spread in triple-negative breast cancer. *Breast Cancer Res. Treat.* **115**, 423–428 (2009).

27. Koscielny, S. *et al.* Breast cancer: relationship between the size of the primary tumour and the probability of metastatic dissemination. *Br J Cancer* **49**, 709–715 (1984).

28. Weigelt, B. *et al.* Gene expression profiles of primary breast tumors maintained in distant metastases. *Proc Natl Acad Sci U S A* **100**, 15901–15905 (2003).

29. Hüsemann, Y. *et al.* Systemic spread is an early step in breast cancer. *Cancer Cell* **13**, 58–68 (2008).

30. Engel, J. *et al.* The process of metastasisation for breast cancer. *Eur J Cancer* **39**, 1794–1806 (2003).

31. van de Wouw, A. J., Janssen-Heijnen, M. L., Coebergh, J. W. & Hillen, H. F. Epidemiology of unknown primary tumours; incidence and population-based survival of 1285 patients in Southeast Netherlands, 1984-1992. *Eur J Cancer* **38**, 409–413 (2002).

32. Massague, J. & Obenauf, A. C. Metastatic colonization by circulating tumour cells.

33. Heppner, G. H. Tumor Heterogeneity. *Cancer Res.* **44**, 2259–2265 (1984).

34. Welch, D. R. Tumor heterogeneity - A ‘contemporary concept’ founded on historical insights and predictions. *Cancer Research* **76**, 4–6 (2016).

35. Mroz, E. A. *et al.* High intratumor genetic heterogeneity is related to worse outcome in patients with head and neck squamous cell carcinoma. *Cancer* **119**, 3034–3042 (2013).

36. Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: A looking glass for

cancer? *Nature Reviews Cancer* **12**, 323–334 (2012).

37. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science (80-.).* **351**, 1463–1469 (2016).

38. Kreso, A. & Dick, J. E. Evolution of the cancer stem cell model. *Cell Stem Cell* **14**, 275–291 (2014).

39. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nature Medicine* **19**, 1423–1437 (2013).

40. Pistollato, F. *et al.* Intratumoral hypoxic gradient drives stem cells distribution and MGMT expression in glioblastoma. *Stem Cells* **28**, 851–862 (2010).

41. Widmer, D. S. *et al.* Hypoxia contributes to melanoma heterogeneity by triggering HIF1α-dependent phenotype switching. *J. Invest. Dermatol.* **133**, 2436–2443 (2013).

42. Black, J. C. *et al.* Hypoxia drives transient site-specific copy gain and drug-resistant gene expression. *Genes Dev.* **29**, 1018–1031 (2015).

43. Hosseini, H. *et al.* Early dissemination seeds metastasis in breast cancer. *Nature* **540**, 552–558 (2016).

44. Puram, S. V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624.e24 (2017).

45. Nguyen, Q. H., Pervolarakis, N., Nee, K. & Kessenbrock, K. Experimental Considerations for Single-Cell RNA Sequencing Approaches. *Front. Cell Dev. Biol.* **6**, 108 (2018).

46. Kalisky, T. *et al.* A brief review of single-cell transcriptomic technologies. doi:10.1093/bfgp/elx019

47. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: Current state of the science. *Nature Reviews Genetics* **17**, 175–188 (2016).

48. Van Loo, P. & Voet, T. Single cell analysis of cancer genomes. *Current Opinion in Genetics and Development* **24**, 82–91 (2014).

49. Van Den Brink, S. C. *et al.* Single-cell sequencing reveals dissociation-induced gene

expression in tissue subpopulations. *Nature Methods* **14**, 935–936 (2017).

50. Svensson, V. *et al.* Power analysis of single-cell rnA-sequencing experiments. *Anal. Nat. methods* | **14**, 381 (2017).

51. Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* **29**, 1120–1127 (2011).

52. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (80-.).* **352**, 189–196 (2016).

53. Venteicher, A. S. *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science (80-.).* **355**, (2017).

54. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).

55. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).

56. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171–181 (2014).

57. Picelli, S., Sagasser, S., Institutet, K. & Sandberg, R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells Single Cell Sequencing View project Magnetic Field Effects View project Omid R Faridani Karolinska Institutet. (2019). doi:10.1038/nMeth.2639

58. Bhargava, V., Ko, P., Willems, E., Mercola, M. & Subramaniam, S. Quantitative transcriptomics using designed primer-based amplification. *Sci. Rep.* **3**, 1–9 (2013).

59. Bhargava, V., Head, S. R., Ordoukhanian, P., Mercola, M. & Subramaniam, S. Technical Variations in Low-Input RNA-seq Methodologies. *Sci. Rep.* **4**, 1–10 (2015).

60. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* **14**, 955–958 (2017).

61. Gao, R. *et al.* Nanogrid single-nucleus RNA sequencing reveals phenotypic diversity in

breast cancer. *Nat. Commun.* **8**, 1–12 (2017).

62. WARBURG, O. [Origin of cancer cells]. *Oncol.* **9**, 75–83 (1956).

63. Koppenol, W. H., Bounds, P. L. & Dang, C. V. Otto Warburg's contributions to current concepts of cancer metabolism. *Nat. Rev. Cancer* **11**, 325–337 (2011).

64. Bomanji, J. B., Costa, D. C. & Ell, P. J. Clinical role of positron emission tomography in oncology. *Lancet Oncology* **2**, 157–164 (2001).

65. Liu, Y. *et al.* A Small-Molecule Inhibitor of Glucose Transporter 1 Downregulates Glycolysis, Induces Cell-Cycle Arrest, and Inhibits Cancer Cell Growth In Vitro and In Vivo. *Mol. Cancer Ther.* **11**, 1672–1682 (2012).

66. Maher, J. C., Krishan, A. & Lampidis, T. J. Greater cell cycle inhibition and cytotoxicity induced by 2-deoxy-D-glucose in tumor cells treated under hypoxic vs aerobic conditions. *Cancer Chemother. Pharmacol.* **53**, 116–122 (2004).

67. Colen, C. B. *et al.* Metabolic targeting of lactate efflux by malignant glioma inhibits invasiveness and induces necrosis: An in vivo study1. *Neoplasia* **13**, 620–632 (2011).

68. Cavalli, L. R., Varella-Garcia, M. & Liang, B. C. Diminished tumorigenic phenotype after depletion of mitochondrial DNA. *Cell Growth Differ.* **8**, 1189–1198 (1997).

69. Hsu, P. P. & Sabatini, D. M. Cancer cell metabolism: Warburg and beyond. *Cell* **134**, 703–707 (2008).

70. Mazurek, S., Boschek, C. B., Hugo, F. & Eigenbrodt, E. Pyruvate kinase type M2 and its role in tumor growth and spreading. *Seminars in Cancer Biology* **15**, 300–308 (2005).

71. Pathria, G. *et al.* Targeting the Warburg effect via LDHA inhibition engages ATF4 signaling for cancer cell survival. *EMBO J* **37**, (2018).

72. Patra, K. C. & Hay, N. The pentose phosphate pathway and cancer. *Trends in Biochemical Sciences* **39**, 347–354 (2014).

73. Camarda, R. *et al.* Inhibition of fatty acid oxidation as a therapy for MYC-overexpressing triple-negative breast cancer. *Nat. Med.* **22**, 427–432 (2016).

74. Park, J. H. *et al.* Fatty Acid Oxidation-Driven Src Links Mitochondrial Energy Reprogramming and Oncogenic Properties in Triple-Negative Breast Cancer. *Cell Rep.* **14**, 2154–2165 (2016).

75. Choi, Y. K. & Park, K. G. Targeting glutamine metabolism for cancer treatment. *Biomolecules and Therapeutics* **26**, 19–28 (2018).

76. Dornier, E. *et al.* Glutaminolysis drives membrane trafficking to promote invasiveness of breast cancer cells. *Nat. Commun.* **8**, 2255 (2017).

77. Rodrigues, M. F. *et al.* Enhanced OXPHOS, glutaminolysis and β-oxidation constitute the metastatic phenotype of melanoma cells. *Biochem. J.* **473**, 703–715 (2016).

78. DeBerardinis, R. J. *et al.* Beyond aerobic glycolysis: Transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 19345–19350 (2007).

79. Piskounova, E. *et al.* Oxidative stress inhibits distant metastasis by human melanoma cells. *Nature* **527**, 186–191 (2015).

80. Schafer, Z. T. *et al.* Antioxidant and oncogene rescue of metabolic defects caused by loss of matrix attachment. *Nature* **461**, 109–113 (2009).

81. Traverso, N. *et al.* Role of Glutathione in Cancer Progression and Chemoresistance. *Oxid. Med. Cell. Longev.* **2013**, (2013).

82. Yang, L. *et al.* Metabolic shifts toward glutamine regulate tumor growth, invasion and bioenergetics in ovarian cancer. *Mol. Syst. Biol.* **10**, 728 (2014).

83. Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E. & Gianni, L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol* **13**, 674–690 (2016).

84. Weigelt, B., Peterse, J. L. & van 't Veer, L. J. Breast cancer metastasis: markers and models. *Nat Rev Cancer* **5**, 591–602 (2005).

85. Oskarsson, T., Batlle, E. & Massagué, J. Metastatic stem cells: Sources, niches, and vital

pathways. *Cell Stem Cell* **14**, 306–321 (2014).

86. Lawson, D. A., Kessenbrock, K., Davis, R. T., Pervolarakis, N. & Werb, Z. Tumour heterogeneity and metastasis at single-cell resolution. *Nat Cell Biol* **20**, 1349–1360 (2018).

87. Ziegenhain, C. *et al.* Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* **65**, 631-643.e4 (2017).

88. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495–502 (2015).

89. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420 (2018).

90. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381–386 (2014).

91. Kuleshov, M. V *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-7 (2016).

92. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (80-.).* **344**, 1396–1401 (2014).

93. Győrffy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res. Treat.* **123**, 725–731 (2010).

94. Gendoo, D. M. A. *et al.* Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* **32**, 1097–1099 (2016).

95. Izumchenko, E. *et al.* Patient-derived xenografts effectively capture responses to oncology therapy in a heterogeneous cohort of patients with solid tumors. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **28**, 2595–2605 (2017).

96. Bruna, A. *et al.* A Biobank of Breast Cancer Explants with Preserved Intra-tumor

Heterogeneity to Screen Anticancer Compounds. *Cell* **167**, 260-274.e22 (2016).

97. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

98. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331–D338 (2017).

99. Radisky, E. S. & Radisky, D. C. Matrix Metalloproteinase-Induced Epithelial-Mesenchymal Transition in Breast Cancer. *J. Mammary Gland Biol. Neoplasia* **15**, 201–212 (2010).

100. Cleary, A. S., Leonard, T. L., Gestl, S. A. & Gunther, E. J. Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. *Nature* **508**, 113–117 (2014).

101. Marusyk, A. *et al.* Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature* **514**, 54–58 (2014).

102. McAnena, P. F. *et al.* Breast cancer subtype discordance: impact on post-recurrence survival and potential treatment options. *BMC Cancer* **18**, 203 (2018).

103. Heppner, G. H., Dexter, D. L., DeNucci, T., Miller, F. R. & Calabresi, P. Heterogeneity in drug sensitivity among tumor cell subpopulations of a single mammary tumor. *Cancer Res* **38**, 3758–3763 (1978).

104. McGranahan, N. & Swanton, C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* **27**, 15–26 (2015).

105. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883–892 (2012).

106. Kim, C. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* **173**, 879-893.e13 (2018).

107. Ikwegbue, P. C., Masamba, P., Oyinloye, B. E. & Kappo, A. P. Roles of heat shock proteins in apoptosis, oxidative stress, human inflammatory diseases, and cancer.

Pharmaceuticals **11**, 1–18 (2018).

108. Creagh, E. M., Sheehan, D. & Cotter, T. G. *Heat shock proteins-modulators of apoptosis in tumour cells*. *Leukemia* **14**, (2000).

109. Tunster, S. J., Creeth, H. D. J. & John, R. M. The imprinted Phlda2 gene modulates a major endocrine compartment of the placenta to regulate placental demands for maternal resources. *Dev. Biol.* **409**, 251–260 (2016).

110. Moon, H.-G. G. *et al.* Prognostic and functional importance of the engraftment-associated genes in the patient-derived xenograft models of triple-negative breast cancers. *Breast Cancer Res Treat* **154**, 13–22 (2015).

111. Ilina, O. & Friedl, P. Mechanisms of collective cell migration at a glance. *J. Cell Sci.* **122**, 3203–3208 (2009).

112. Ewald, A. J., Brenot, A., Duong, M., Chan, B. S. & Werb, Z. Collective Epithelial Migration and Cell Rearrangements Drive Mammary Branching Morphogenesis. *Dev. Cell* **14**, 570–581 (2008).

113. Friedl, P. & Gilmour, D. Collective cell migration in morphogenesis, regeneration and cancer. *Nature Reviews Molecular Cell Biology* **10**, 445–457 (2009).

114. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e21 (2019).

115. Smith, A. P. *et al.* A positive role for Myc in TGFbeta-induced Snail transcription and epithelial-to-mesenchymal transition. *Oncogene* **28**, 422–430 (2009).

116. Zhang, Y. & Weinberg, R. A. Epithelial-to-mesenchymal transition in cancer: complexity and opportunities. *Frontiers of Medicine* **12**, 361–373 (2018).

117. Liu, T., Zhou, L., Li, D., Andl, T. & Zhang, Y. Cancer-associated fibroblasts build and secure the tumor microenvironment. *Frontiers in Cell and Developmental Biology* **7**, (2019).

118. Gunasinghe, N. P. A. D., Wells, A., Thompson, E. W. & Hugo, H. J. Mesenchymal-

epithelial transition (MET) as a mechanism for metastatic colonisation in breast cancer. *Cancer Metastasis Rev.* **31**, 469–478 (2012).

119. Mah, E. J., Lefebvre, A. E. Y. T., McGahey, G. E., Yee, A. F. & Digman, M. A. Collagen density modulates triple-negative breast cancer cell metabolism through adhesion-mediated contractility. *Sci. Rep.* **8**, 17094 (2018).

120. Liu, X. *et al.* High-Resolution Metabolomics with Acyl-CoA Profiling Reveals Widespread Remodeling in Response to Diet. *Mol. Cell. Proteomics* **14**, 1489–500 (2015).

121. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K. A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, (2017).

122. Scaduto, R. C. & Grotyohann, L. W. Measurement of Mitochondrial Membrane Potential Using Fluorescent Rhodamine Derivatives. *Biophys. J.* **76**, 469–477 (1999).

123. Mayers, J. R. *et al.* Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nat. Med.* **20**, 1193–1198 (2014).

124. Fernyhough, P. & McGavock, J. Mechanisms of disease. in *Handbook of Clinical Neurology* **126**, 353–377 (Elsevier B.V., 2014).

125. Hao, W., Chang, C. P., Tsao, C. C. & Xu, J. Oligomycin-induced bioenergetic adaptation in cancer cells with heterogeneous bioenergetic organization. *J Biol Chem* **285**, 12647–12654 (2010).

126. Choi, S. W., Gerencser, A. A. & Nicholls, D. G. Bioenergetic analysis of isolated cerebrocortical nerve terminals on a microgram scale: spare respiratory capacity and stochastic mitochondrial failure. *J. Neurochem.* **109**, 1179–91 (2009).

127. Dayan, F. *et al.* Activation of HIF-1α in exponentially growing cells via hypoxic stimulation is independent of the Akt/mTOR pathway. *J. Cell. Physiol.* **218**, 167–174 (2009).

128. Ward, M. W., Rego, A. C., Frenguelli, B. G. & Nicholls, D. G. Mitochondrial membrane potential and glutamate excitotoxicity in cultured cerebellar granule cells. *J. Neurosci.* **20**, 7208–7219 (2000).

129. Sameni, S., Syed, A., Marsh, J. L. & Digman, M. A. The phasor-FLIM fingerprints reveal shifts from OXPHOS to enhanced glycolysis in Huntington Disease. *Sci. Rep.* **6**, (2016).

130. Digman, M. A., Caiolfa, V. R., Zamai, M. & Gratton, E. The Phasor Approach to Fluorescence Lifetime Imaging Analysis. *Biophys. J.* **94**, L14–L16 (2008).

131. Antalis, C. J., Uchida, A., Buhman, K. K. & Siddiqui, R. A. Migration of MDA-MB-231 breast cancer cells depends on the availability of exogenous lipids and cholesterol esterification. *Clin. Exp. Metastasis* **28**, 733–741 (2011).

132. Elia, I. *et al.* Proline metabolism supports metastasis formation and could be inhibited to selectively target metastasizing cancer cells. *Nat. Commun.* **8**, 15267 (2017).

133. Christen, S. *et al.* Breast Cancer-Derived Lung Metastases Show Increased Pyruvate Carboxylase-Dependent Anaplerosis. *Cell Rep.* **17**, 837–848 (2016).

134. Elia, I., Doglioni, G. & Fendt, S. M. Metabolic Hallmarks of Metastasis Formation. *Trends Cell Biol* **28**, 673–684 (2018).

135. Teoh, S. T. & Lunt, S. Y. Metabolism in cancer metastasis: bioenergetics, biosynthesis, and beyond. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **10**, e1406 (2018).

136. LeBleu, V. S. *et al.* PGC-1α mediates mitochondrial biogenesis and oxidative phosphorylation in cancer cells to promote metastasis. *Nat Cell Biol* **16**, 1-15,992-1003 (2014).

137. Basnet, H. *et al.* Flura-seq identifies organ-specific metabolic adaptations during early metastatic colonization. *Elife* **8**, (2019).

138. Porporato, P. E. *et al.* A mitochondrial switch promotes tumor metastasis. *Cell Rep* **8**, 754–766 (2014).

139. Ogura, M., Yamaki, J., Homma, M. K. & Homma, Y. Mitochondrial c-Src regulates cell survival through phosphorylation of respiratory chain components. *Biochem. J.* **447**, 281–9 (2012).

140. Zielonka, J. & Kalyanaraman, B. 'ROS-generating mitochondrial DNA mutations can

regulate tumor cell metastasis'-a critical commentary. *Free Radic. Biol. Med.* **45**, 1217–1219 (2008).

141. Giannoni, E. *et al.* Redox regulation of anoikis: Reactive oxygen species as essential mediators of cell survival. *Cell Death Differ.* **15**, 867–878 (2008).
142. Kumar, S. *et al.* Metformin intake is associated with better survival in ovarian cancer: a case-control study. *Cancer* **119**, 555–62 (2013).
143. Bodmer, M., Becker, C., Meier, C., Jick, S. S. & Meier, C. R. Use of metformin and the risk of ovarian cancer: a case-control analysis. *Gynecol. Oncol.* **123**, 200–4 (2011).
144. Romero, I. L. *et al.* Relationship of type II diabetes and metformin use to ovarian cancer progression, survival, and chemosensitivity. *Obstet. Gynecol.* **119**, 61–67 (2012).
145. Col, N. F., Ochs, L., Springmann, V., Aragaki, A. K. & Chlebowski, R. T. Metformin and breast cancer risk: a meta-analysis and critical literature review. *Breast Cancer Res. Treat.* **135**, 639–646 (2012).
146. Yap, T. A. *et al.* Phase I trial of IACS-010759 (IACS), a potent, selective inhibitor of complex I of the mitochondrial electron transport chain, in patients (pts) with advanced solid tumors. *J. Clin. Oncol.* **37**, 3014–3014 (2019).
147. Hafeez, B. Bin *et al.* Plumbagin inhibits prostate cancer development in TRAMP mice via targeting PKCε, Stat3 and neuroendocrine markers. *Carcinogenesis* **33**, 2586–2592 (2012).
148. Hafeez, B. Bin *et al.* Plumbagin inhibits prostate carcinogenesis in intact and castrated PTEN knockout mice via targeting PKCε, Stat3, and epithelial-to-mesenchymal transition markers. *Cancer Prev. Res.* **8**, 375–386 (2015).
149. Fiorillo, M. *et al.* Repurposing atovaquone: targeting mitochondrial complex III and OXPHOS to eradicate cancer stem cells. *Oncotarget* **7**, 34084–99 (2016).
150. Lv, Z., Yan, X., Lu, L., Su, C. & He, Y. Atovaquone enhances doxorubicin's efficacy via inhibiting mitochondrial respiration and STAT3 in aggressive thyroid cancer. *J. Bioenerg.*

Biomembr. **50**, 263–270 (2018).

151. Salminen, A., Kauppinen, A., Hiltunen, M. & Kaarniranta, K. Krebs cycle intermediates regulate DNA and histone methylation: Epigenetic impact on the aging process. *Ageing Research Reviews* **16**, 45–65 (2014).
152. Leone, R. D. & Emens, L. A. Targeting adenosine for cancer immunotherapy. *Journal for ImmunoTherapy of Cancer* **6**, 57 (2018).
153. Perrot, I. *et al.* Blocking Antibodies Targeting the CD39/CD73 Immunosuppressive Pathway Unleash Immune Responses in Combination Cancer Therapies. *CellReports* **27**, 2411-2425.e9 (2019).
154. Tataranni, T. *et al.* Dichloroacetate Affects Mitochondrial Function and Stemness-Associated Properties in Pancreatic Cancer Cell Lines. *Cells* **8**, 478 (2019).
155. Kandarian, F., Sunga, G. M., Arango-Saenz, D. & Rossetti, M. A flow cytometry-based cytotoxicity assay for the assessment of human NK cell activity. *J. Vis. Exp.* **2017**, (2017).
156. Mellqvist, U. H. *et al.* Natural killer cell dysfunction and apoptosis induced by chronic myelogenous leukemia cells: role of reactive oxygen species and regulation by histamine. *Blood* **96**, 1961–1968 (2000).
157. Buck, M. D., Sowell, R. T., Kaech, S. M. & Pearce, E. L. Metabolic Instruction of Immunity. *Cell* **169**, 570–586 (2017).
158. Leone, R. D. *et al. Glutamine blockade induces divergent metabolic programs to overcome tumor immune evasion*.
159. Chang, C. H. *et al.* Metabolic Competition in the Tumor Microenvironment Is a Driver of Cancer Progression. *Cell* **162**, 1229–1241 (2015).