UNIVERSITY OF CALIFORNIA,
IRVINE

# Engineering Multi-Component Bioluminescent Imaging Platforms and Transcriptome Analyses in Models of Metastatic Breast Cancer

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Anastasia Alexeevna Ionkina

Dissertation Committee:
Professor Jennifer Prescher, Chair
Professor David Fruman
Professor Christopher Hughes
Professor Robert Spitale
Assistant Professor Kai Kessenbrock

2021

# DEDICATION

To

my incredible mama, Olga Petrovna,

my loving sister, Elizabeth,

my supportive family around the world,

my dear friends, the family we choose for ourselves,

and in loving memory of Benjamin Brewer

– your light shines on and you continue to live through us–

*"May it be a light to you in dark places, when all other lights go out."*

- J. R. R. Tolkien

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGEMENTS

First, I would like to express a profound level of gratitude for the unwavering leadership and equanimity required to mentor a PhD thesis during a global pandemic. I would like to extend my deepest appreciation to my committee chair, Professor Jennifer Prescher, for allowing me the freedom to grow as a scientist. Since the beginning of my PhD, you have allowed me to explore my interests, test out my ideas, and unlock my scientific creativity.

I would like to express my gratitude to my committee members Professors David Fruman, Christopher Hughes, Robert Spitale, and Kai Kessenbrock for their support and constructive feedback. Your expertise has been invaluable throughout my dissertation and my growth as a scientist. I truly appreciate the mentorship and guidance that you have provided me (both in science and in life). I would like to thank Professor David Fruman especially for his continued support and encouragement.

In addition, I would like to thank all the great collaborators that I have had the privilege of working with throughout the years. I would especially like to thank Ali Mortazavi for his insight during our collaborative meetings and the guidance he offered me throughout my PhD.

Furthermore, I would like to express my gratitude to Jenn Atwood and Vanessa Scarfone for being great colleagues to work with throughout the years. I appreciate both of your expertise and the valued input that you have provided me. I also value the friendship that we have developed over the years. I would like to extend my deepest appreciation to Nita Driscoll and Krystina Jarema for their constant support and compassion. In addition, I would like to thank the Molecular Biology and Biochemistry department and facilities for their support. I would also like to thank the Edinger lab, Mortazavi lab, Lawson lab, Kessenbrock lab, Waterman lab, Fleishman lab, and Fruman lab for all the constructive discussions and collaborations over the years. I am especially grateful for the friendships that I have developed within these labs and the support network that we have formed.

I would like to give a special thanks to my incredible undergraduate researcher Steven Phan. It was such as joy being your mentor and seeing your natural scientific abilities grow. I am excited to see the great things that the future has instore for you.

I would not be here without the mentorship and support of my previous mentors Dr. Jennifer Diamond, Dr. John Tentler, Dr. Todd Pitts, Dr. Gail Eckhardt, Dr. Lora Wilson, Dr. Erica Bradshaw and everyone else at the University of Colorado. I would especially like to thank Dr. Jennifer Diamond for believing in me and fostering my growth as a scientist. Jenn, you have provided me with so many opportunities and supported me every step of the way. I am forever grateful to you.

My journey through the PhD process was filled with more than its share of unexpected difficulties. I would like to thank my loving family (near and far) for their support and for the opportunities that they have provided me. Through the best of times and the worst of times, it was also the support, love, and never-ending encouragement of my friends that got me through. Friends are the family we get to choose for ourselves – and I am so grateful that we chose each other.

I would like to graciously thank the funding support I received from the Cancer Biology Training Grant (T32-CA009054), the UCI Opportunity Center for Complex Biological Systems Grant, and the Department of Molecular Biology and Biochemistry.

# CURRICULUM VITAE

## Anastasia A. Ionkina, B.A.

**PROFILE**

Cross-functional interdisciplinary cancer biologist with over 10 years of experience in molecular and tumor-immune microenvironment biology, translating my cancer research experience into clinical trials and novel technologies.

**AREAS OF COMPETENCY**

**Cellular & Molecular Biology:** cell culture, cloning & plasmid design, gene editing (CRISPR-Cas9, shRNA, retro/lentivirus), primer design, sequencing, colorimetric & FL analysis, spectroscopy, protein evolution, organoids, 3D cell culture assays, flow cytometry, cell cycle analysis, FACS cell sorting, bacterial culture, qRT-PCR, PCR, ELISA, western blots, purification (DNA, RNA, & protein), *extensive knowledge of molecular signaling pathways*

**Drug Development:** cell-based assays (proliferation, growth, apoptosis, autophagy, necrosis, senescence, & senescence-associated secretory phenotypes), biochemistry & assay development, predictive biomarker identification, molecular mechanisms of resistance analysis, mutation assays, MULTI-SPOT, tissue processing (OCT, FFPE), histology (ICC, IF, IHC, H&E, β-gal), drug repurposing, targeted therapeutics (kinases, epigenetics, metabolism), combination drug studies

**Preclinical Cancer Models:** extensive animal experience (genotyping, breeding, surgeries, trocar, perfusions), PDTX models, bulk RNA-seq sample processing, SubQ & orthotropic cancer models, blood/organ harvesting, PK/PD studies, combination drug study development & design, drug administration (PO, IV, IP, SubQ, RO, IM, IC, IDUC, IMAM, TD), tumor growth study design & analysis, creating primary cell lines

**Clinical Trials:** bench-to-bedside personalized medicine, Phase 1 & Phase 2A/2B TNBC clinical trials, CTAs submission assistance, targeted therapies & small molecule drug combination studies, IND application, clinical protocol design & patient selection criteria, tumor biopsy analysis, data presentation & communication with clinical trial team, HIPPA

**Imaging Tool Development:** cellular & whole animal probe engineering (BRET, FRET, BL, FL), immune-oncology imaging, multicomponent BLI, BL/FL multiplex imaging platform design, imaging algorithms analysis, cell-cell contact reporters, promoter driven imaging reporters, multiplex gene expression reporters (immune, cancer, M1/M2, T cell), tumor microenvironment analysis, confocal microscopy (LCMS, SDCM)

**Big Data Analysis**: bulk RNA seq, single cell RNA seq, data mining, RNA seq data interpretation, statistical analysis, synergistic drug combinations analysis (BLISS, Chou & Talalay combination index, Calcusyn), PRYSIM, Image J, Fiji

## EDUCATION

**Ph.D. Candidate, Molecular Biology and Biochemistry** 2016 - 2021
University of California, Irvine, CA

**BS, Molecular, Cellular, and Developmental Biology** 2008- 2012
University of Colorado, Boulder, CO | Dean's list for academic excellence

**BS, Psychology; Neuroscience Minor/Certification** 2008- 2012
University of Colorado, Boulder, CO | Dean's list for academic excellence

## EXPERIENCE

**Metastatic Cancer Researcher & Imaging Tool Engineer** August 2016-December 2021
University of California, Irvine, CA

- Work resulted in 3 publications, 3 peer-reviewed poster abstracts, and selected as sole graduate student for oral presentation on main stage at international EACR-AACR TME 2020 conference (Lisbon, Portugal)
- Crafted successful research proposals that launched **5** multidisciplinary cross-functional research collaborations, resulting in **$2,250,000** of laboratory grant funding
- Developed rapid-multicomponent bioluminescence imaging platform and computational analysis protocol for non-invasive visualization of multicellular processes and cell-to-cell contact in metastatic breast cancer animal models on the minutes timescale, a vast improvement over conventional methods
- Spearheaded the modernization of laboratory's cloning methods by introducing CRISPR-Cas9 gene editing technology, trained **28** multidisciplinary lab mates, and streamlined CRISPR protocol to dramatically decreased costs and improve product output
    - CRISPR has now been used in over **200 projects** in the Prescher lab
- Launched, led, and directed laboratory's first research projects in cancer biology and educated multidisciplinary team of **28** chemical-biologists and synthetic chemists
- Created new tissue harvesting and primary cell line culturing protocols to establish comprehensive panel of organ-derived metastatic clonal isolates from breast cancer mouse model to study metastatic niches and cancer progression resulting in lower costs for research projects
- Engineered reporter constructs to study immune cell changes during cancer progression resulting in new laboratory research focus
- Established the use of bioinformatics analysis (RNA-seq and single cell RNA-seq) to characterize metastatic niches in breast cancer, launching the use RNA-sequencing technology in the laboratory and facilitated **3** multidisciplinary collaborations to secure future R21 translational grant funding (**$550,000**)
- Awarded **$10,000** Cancer Systems Biology Initiative Opportunity Grant and successfully completed research proposal investigating tumor heterogeneity in metastatic breast cancer using next generation RNA-sequencing

**Senior Translational Cancer Research Scientist** January 2013-June 2016
University of Colorado, Anschutz, CO

- Cultivated strong intrapersonal connections and collaborations with dynamic translational medical oncology team and led preclinical research resulting in 4 clinical trials, 4 publications, 5 peer-reviews

poster abstracts, 22 research poster presentations, and 2 abstract presentations at international AACR-NCI-EORTC conferences

- Spearheaded drug development by identifying novel predictive biomarkers of clinical response to small molecule targeted therapeutics and chemotherapy combination treatments in breast cancer, colorectal cancer, and melanoma
    - Preclinical findings were translated into **4 clinical trials**
- Designed and carried out multi-drug experimental studies using patient samples, cell-based assays, and animal models, resulting in the identification of molecular mechanisms that govern innate and acquired resistance to cancer therapies
- Investigated p53/p73 mutation status as modulators of molecular mechanisms of predicative clinical response to targeted therapeutics (kinase inhibitors) via cell-based assays and patient derived tumor xenografts (PDTX) models
    - Created second commercially known p53 wild type triple negative breast cancer (TNBC) PDTX cell line
    - Epigenetic regulation and Histone deacetylation inhibitors for breast cancer treatment
- Developed novel protocols for immunofluorescence analysis of cell lines and PDTX models
- Managed grant budget, laboratory purchasing, invoice filing, and rodent ordering
- Evaluated and adjusted budget resulting in expansion of lab funds to hire more researchers
- Responsible for writing and maintaining of federal mandated protocols for animal studies (IACUC), biosafety, and all environmental health and safety requirements for the entire laboratory's function and compliance

**Clinical Trial Cancer Research Scientist** January 2013-June 2016
University of Colorado, Anschutz, CO

- Collaborated with intradisciplinary team of clinical oncologists, physician scientists, academic scientists and industrial pharmaceutical research scientists to successfully translate preclinical research into clinical trials
- Identified and characterized predictive biomarkers and novel targeted therapies in treatment resistant TNBC patients
- Launched **4 clinical trials** based on the identified predictive biomarkers:
    - Aided in writing of clinical trial application which used preclinical data as parameters for patient selection criteria, treatment administration, and clinical trial protocol design
    - *Work resulted in FDA approval of NID and clinical applications for multiple trials*
- Lead scientist presenting the preclinical scientist at collective clinical trial meetings
- Responsible for documentation, collection, and proper processing of clinical trial patient samples (microdissections & fine-needle aspiration)
    - Identified optimal parameters and protocol for clinical patient sample analysis and processing
- Responsible for coordinating and maintaining excellent relationship with cross country academic university collaborators, various pharmaceutical companies, and the National Cancer Institute's Cancer Therapy Evaluation Program (CTEP), resulting in the entrustment of ordering laboratory's supply of preclinical experimental drugs and maintain efficient research

**Vaccine Research & Development Scientist** August 2013-January 2016
University of Colorado, Boulder, CO

- Collaborated on the development of HPV vaccines streamlined for FDA approval under the mentorship of Robert Garcea MD, *co-developer of first HPV vaccine GARDASIL*

- Bolstered laboratory's efforts to create custom vaccines for sub-Saharan Africa's most abundant and deadly forms of the HPV capsid proteins, resulting in the project being funded by the Bill & Melinda Gates Foundation
- Assisted in engineering low-cost, aerosolized inhaled-versions of HPV vaccines that enabled easier transport and improved accessibility to cancer prevention in remote and under privileged countries

## SELECT FELLOWSHIPS AND AWARDS

UCI Graduate Dean's Distinguished Dissertation Fellowship (**$5,000**) July 2021

NIH T32 Cancer Research Predoctoral Training Fellow reappointment **($25,800**) June 2020-June 2021

UCI Python Coding & Artificial Intelligence in Cancer Certificate August 2020

UCI Beall Applied Innovation Business Concepts in STEM Certificate April 2020-June 2020

EACR-AACR Basic & Translational Tumor Microenvironment Conference, Lisbon Portugal March 2020

- Selected for oral presentation at international conference

International Conference Travel Funding Award **($3,000**) February 2020

Cancer Systems Biology Initiative CCBS Opportunity Grant **($10,000**) July 2019-July 2020

- Proposal: *Plasticity and diversity of metastatic breast cancer cells revealed by transcriptome analysis*

NIH T32 Cancer Research Predoctoral Training Fellow **($25,800**) June 2019-June 2020

## SELECT MENTORSHIP EXPERIENCE

Undergraduate Research Mentor, University of California, Irvine January 2018 - present

- Under my mentorship undergraduate awarded UROP grant funding

Graduate Student Instructor, University of California, Irvine

Biochemistry Lab I Molecular Biology March 2018-June 2018
Advanced Biochemistry Lab I Advanced Molecular Biology September 2021-December 2021

Graduate Student Mentor, University of California, Irvine August 2016 - present

Cancer Research Mentor, University of Colorado, Aurora March 2013 - March 2015

- Project manager and experimental design **mentor of 6 research fellows**

## SELECT LEADERSHIP AND EDUCATION OUTREACH

STEM Education Outreach Leader, Virtual Meeting, UCLA, CA February 2021

- Crafted interactive STEM curriculum to encourage diversity and develop confidence in underprivileged girls (4-9 year old girls) to pursue science

NIH T32 Training Grant Predoctoral Seminars, Irvine, CA August 2020 - January 2021

- Hosted: Professor Robert A. Weinberg, Director MIH Molecular Oncology

Co-chair Youth Science Fellowship, Cancer Research Institute Irvine, CA June 2020 - July 2020

- Created interactive STEM curriculum and lectures for underprivileged high school students

Chair Molecular Biology & Biochemistry Visiting Scholar Seminars, Irvine, CA 2018 - 2020

Association for Women in Science - Education and Outreach Team Member 2018 - present

Cancer Education Outreach Leader and Research Mentor, Aurora, CO March 2013 - March 2016

Cancer Outreach Camp Counselor and Mentor, Denver, CO May 2011 - August 2011

- Mentored children and adolescents battling or affected by childhood cancer

Scientific Research and Education Institute Internship, Thornton, CO January 2010 - August 2013

## SELECT PROFESSIONAL AFFILIATIONS

UCI NIH Consortium Graduate Professional Success in STEM, Irvine, CA 2019-present

NIH BEST Graduate Professional Success in Biomedical Sciences, Irvine, CA 2016-2019

Association for Women in Science, Irvine, CA 2018-present

Graduate Women in Science, Los Angeles, CA 2018-present

Associate Member, American Association of Cancer Research (AACR) 2013-present

## SELECT FELLOWSHIPS AND AWARDS

UCI Graduate Dean's Distinguished Dissertation Fellowship (**$5,000**) July 2021

NIH T32 Cancer Research Predoctoral Training Fellow reappointment **($25,800**) June 2020-June 2021

UCI Python Coding & Artificial Intelligence in Cancer Certificate August 2020

UCI Beall Applied Innovation Business Concepts in STEM Certificate April 2020-June 2020

EACR-AACR Basic & Translational Tumor Microenvironment Conference, Lisbon Portugal March 2020

- Selected for oral presentation at international conference

International Conference Travel Funding Award **($3,000**) February 2020

Cancer Systems Biology Initiative CCBS Opportunity Grant **($10,000**) July 2019-July 2020

- Proposal: *Plasticity and diversity of metastatic breast cancer cells revealed by transcriptome analysis*

NIH T32 Cancer Research Predoctoral Training Fellow **($25,800**) June 2019-June 2020

## PUBLICATIONS

*Peer Reviewed Manuscripts*

**Anastasia A. Ionkina**, Gabriela Balderrama-Gutierrez, Kristian J. Ibanez, Steve Huy D. Phan, Angelique N. Cortez, Ali Mortazavi, and Jennifer A. Prescher. Comparative transcriptome analysis of metastatic heterogeneity in a mouse model of breast cancer. *Breast Cancer Res* **23,** 93 (2021).

**Anastasia Ionkina**#, Colin M. Rathbun, Zi Yao, Krysten A. Jones, William B. Porterfield, and Jennifer A. Prescher. Rapid multicomponent bioluminescence imaging via substrate unmixing. ACS Chem. Biol.. March 2021. DOI:10.1021/acschembio.0c00959

S. Lindsey Davis., **Anastasia A Ionkina**, *et al.,* (2020). Preclinical and Dose-Finding Phase I Trial Results of Combined Treatment with a TORC1/2 Inhibitor (TAK228) and Aurora A Kinase Inhibitor (Alisertib) in Solid Tumors. *Clinical Cancer Research*, *26*(17) DOI: 10.1158/1078-0432.CCR-19-3498

Anna Capasso, Kyrie L Dailey, Naomi Currimjee, Betelehem Yacob, **Anastasia Ionkina**, Julie G Frank, Deog Joong Kim, Christina George, Young B Lee, Ely Benaim, Aik-Choon Tan, Jihye Kim, Todd M Pitts, S. Gail Eckhardt, John J. Tentler, and Jennifer R. Diamond. First-in-class phosphorylated-p68 inhibitor RX-5902 inhibits β-catenin signaling and demonstrates anti-tumor activity in triple-negative breast cancer. Molecular Cancer Therapeutics. October 2019. DOI: 10.1158/1535-7163. MCT-18-1334.

**Ionkina AA***, Tentler JJ, Kim J, Capasso A, Pitts TM, Ryall KA, Howison RR, Kabos P, Sartorius CA, Tan AC, Eckhardt SG, Diamond JR. Efficacy and Molecular Mechanisms of Differentiated Response to the Aurora and Angiogenic Kinase Inhibitor ENMD-2076 in Preclinical Models of p53-Mutated Triple-Negative Breast Cancer. Frontiers in Oncology. 2017 May; DOI: 10.3389/fonc.2017.00094.

Tentler JJ, **Ionkina AA**, Tan AC, Newton TP, Pitts TM, Glogowska MJ, Kabos P, Sartorius CA, Sullivan KD, Espinosa JM, Eckhardt SG, Diamond JR. p53 family members regulate phenotypic response to Aurora kinase A inhibition in triple-negative breast cancer. Molecular Cancer Theraputics 2015. 2015 May;14(5):1117-29. (PMID: 25758253).

*Manuscripts in Preparation*

**Anastasia A. Ionkina**, Carly Brennen, and Jennifer A. Prescher. A linear unmixing algorithm for multi-substrate bioluminescence imaging of metastatic breast cancer. *Manuscript in preparation for Nature Communications December 2021*

Optimization of split reporters for imaging cell-cell interactions. Jones KA#. **Ionkina AA**#, Xie BJ, Prescher JA. Manuscript in preparation. (2020). # These authors contributed equally

Diamond JR Kim J, Tan AC, **Ionkina AA**, Eckhardt SG, Tentler JJ, ENMD-2076 Clinical Trial Results and Molecular Profiling of Patient Responses.

## ABSTRACTS

*Peer Reviewed*

**Anastasia A. Ionkina*,** Gabriela Balderrama-Gutierrez, Steve Huy Phan, Ali Mortazavi, and Jennifer A. Prescher. Comparative transcriptome analysis of metastatic heterogeneity in a mouse model of breast cancer. AACR Annual Meeting. San Diego, CA. April 2020. Abstract #5138

**Anastasia A. Ionkina***, Colin M. Rathbun, Zi Yao, Krysten A. Jones, William B. Porterfield, and Jennifer A. Prescher. Rapid, non-invasive imaging in mouse models of human cancer. EACR-AACR

Basic and Translational Research Tumor Microenvironment Conference. Lisbon, Portugal. March 2020. Control submission number: 20-A-5138-AACR. EAA-0110. *Abstract selected for oral presentation during main conference.*

**Anastasia Ionkina***, Colin Rathbun, and Jennifer Prescher. Luciferase-luciferin pairs for rapid multicomponent imaging in vivo. Greater LA Chemical Biology Symposium. University of Southern California. (2019).

Jennifer R. Diamond, James D. Orth, **Anastasia Ionkina**, Kyrie Dailey, Todd M. Pitts, Anna Capasso, Joshua M. Marcus, Russell T. Burke, S. Lindsey Davis, Jihye Kim, Aik Choon Tan, S. Gail Eckhardt, and John J. Tentler. Rational Combination of mTOR and Aurora Kinase A Inhibition in Preclinical Models of Triple-Negative Breast Cancer. Proceedings of the AACR-NCI-EORTC International Conference: Molecular Targets and Cancer Therapeutics; 2017 Oct 26-30: Abstract 499.

**Anastasia A. Ionkina***, S. Gail Eckhardt, Todd M. Pitts, Jihye Kim, Aik Choon Tan, Carol Sartorius, Peter Kabos, John J. Tentler, Jennifer R. Diamond. Senescence as a Mechanism of Resistance to the Aurora Kinase and Angiokinase Inhibitor, ENMD-2076, in p53 mutated Triple-Negative Breast Cancer (TNBC) Models. Proceedings of the AACR-NCI-EORTC International Conference: Molecular Targets and Cancer Therapeutics; 2015 Nov 5-9; Boston, MA. Philadelphia (PA): AACR; Mol Cancer Ther 2015: Abstract C34.

Ryall KA, Kim J, Klauck PJ, Shin J, Yoo M, **Ionkina A**, Pitts TM, Tentler JJ, Diamond JR, Eckhardt SG, Heasley LE, Kang J, Tan AC. (2015). An integrated bioinformatics analysis to dissect kinase dependency in triple negative breast cancer. BMC Genomics. (Proceedings of the 1st Joint GIW/InCoB Conference, Tokyo, Japan). May 2015.

Diamond JR, Tan AC, **Ionkina AA**, Newton TP, Pitts TM, Eckhardt SC, Tentler JJ. The role of p53 family tumor suppressors in mediating response to aurora kinase inhibition in triple-negative breast cancer. San Antonio Breast Cancer Symposium, San Antonio, TX, December 2013.

J. Diamond, A. Tan, **A. Ionkina**, T. Newton, T. Pitts, S. Eckhardt, J. Tentler. The role of p53 family tumor suppressors in mediating response to aurora kinase inhibition in triple-negative breast cancer. Cancer Research December 15, 2013; 73(24 Supplement): Abstract nr P2-09-06.

**Anastasia A. Ionkina**, John J. Tentler, Timothy P. Newton, Kelsey L. Brunkow, Jared S. Johnson, Aik Choon Tan, Todd M. Pitts, S. Gail Eckhardt, Jennifer R. Diamond. The role of p53 family tumor suppressors in mediating response to the Aurora and angiogenic kinase inhibitor ENMD-2076 in triple-negative breast cancer. Proceedings of the AACR-NCI-EORTC International Conference: Molecular Targets and Cancer Therapeutics; 2013 Oct 19-23; Boston, MA. Philadelphia (PA): AACR; Mol Cancer Ther 2013;12(11 Suppl): Abstract nr A286.

*Non Peer-Reviewed*

Diamond JR, **Ionkina, AA**, Newton TP, Pitts TM, Brunkow KL, Kabos P, Sartorius CA, Tan AC, Eckhardt SG, Tentler JJ. p53 as a Modulator of Response to Aurora Kinase Inhibition in Triple Negative Breast Cancer. University of Colorado Anschutz Medical Campus, Department of Medicine, Research and Innovation Conference, October 2014.

Tripathi, S, **Ionkina AA**, Pitts, TM, Tentler JJ, Diamond JR. Mutated p53 Response to Aurora Kinase Inhibitors in Triple Negative Breast Cancer. UCCC, Cancer Research Summer Fellowship, August 2014

Diamond JR, **Ionkina, AA**, Newton TP, Pitts TM, Brunkow KL, Kabos P, Sartorius CA, Tan AC, Eckhardt SG, Tentler JJ. Mediators of Terminal Cellular Fate in Response to Aurora Kinase Inhibitors in TNBC. University of Colorado Anschutz Medical Campus, Department of Medicine. Hormone Related Malignancies Retreat, March 2014.

Kattula S, **Ionkina AA**, Cross, BR, Astling D, Tentler JJ, Diamond JR. Preclinical analysis of paragazole (OnK101), a class I histone deacetylation inhibitor, in sensitive and resistant triple negative breast cancer cell lines. UCCC, Women's Health Research Day, October 2013

Kattula S, **Ionkina AA**, Cross, BR, Astling D, Tentler JJ, Diamond JR. Preclinical analysis of paragazole (OnK101), a class I histone deacetylation inhibitor, in sensitive and resistant triple negative breast cancer cell lines. UCCC, Cancer Research Summer Fellowship, August 2013

Johnson JS, **Ionkina AA**, Diamond JR, Tentler JJ. Characterization of the Apoptotic and Senescent Effects of ENMD-2076 on p53 and p73 Knockdowns in Triple Negative Breast Cancer. UCCC, Cancer Research Summer Fellowship, August 2013

## PRESENTATIONS

*Peer-Reviewed*

**Anastasia Ionkina***, Colin Rathbun, and Jennifer Prescher. Luciferase-luciferin pairs for rapid multicomponent imaging in vivo. Greater LA Chemical Biology Symposium. University of Southern California. (2019).

**Anastasia A. Ionkina***, S. Gail Eckhardt, Todd M. Pitts, Jiyhe Kim, Aik Choon Tan, Carol Sartorius, Peter Kabos, John J. Tentler, Jennifer R. Diamond. Senescence as a Mechanism of Resistance to the Aurora Kinase and Angiokinase Inhibitor, ENMD-2076, in p53 mutated Triple-Negative Breast Cancer (TNBC) Models. Molecular Targets and Cancer Therapeutics, AACR-NCI-EORTC, Boston, MA, November 2015.

**Ionkina AA***, Tentler JJ, Newton TP, Brunkow KL, Johnson JS, Tan AC, Pitts TM, Eckhardt SG, Diamond JR. The role of p53 family tumor suppressors in mediating response to the Aurora and angiogenic kinase inhibitor ENMD-2076 in triple-negative breast cancer. Molecular Targets and Cancer Therapeutics, AACR-NCI-EORTC, Boston, MA, October 2013.

*Non Peer-Reviewed*

**Ionkina, AA*** Expanding the bioluminescence toolbox imaging metastatic disease. Cancer Biology & Therapeutics T32 Training Grant Retreat. Los Angelous, CA. April 12-14th 2019.

**Ionkina, AA*** Bioluminescence imaging tools for monitoring cellular interactions *in vivo*. Molecular Biology and Biochemistry Department Retreat Blitz Talk. Carlsbad, CA. March 22nd-23rd 2019.

**Ionkina, AA*** Bioluminescence imaging tools for monitoring cellular interactions in vivo. Molecular Biology and Biochemistry Department Research in Progress. UCI February 2019.

**Ionkina, AA*** Bioluminescence imaging tools for monitoring cellular interactions in vivo. Molecular Biology and Biochemistry Department Research in Progress. UCI May 2018.

**Ionkina, AA*.** Cellular and Molecular Biosciences recruitment poster presentation session. UCI, CA. January 2018.

**Ionkina, AA*.** Current Prospects of Cancer Research and Novel Therapeutic Treatment of Triple Negative Breast Cancer. Pre Medical Academic Group, George Washington High School, Denver CO, September 2015.

**Ionkina, AA*,** Tentler JJ, Newton TP, Brunkow KL, Kabos P, Sartorius CA, Tan AC, Pitts TM, Eckhardt SG, Diamond JR. The Role of p53 and Senescence as a Mechanism of Resistance to Aurora Kinase

Inhibitors in Triple Negative Breast Cancer, University of Colorado Anschutz Medical Campus, Department of Medicine Research Day, October 2014

**Ionkina, AA***, Tentler JJ, Newton TP, Brunkow KL, Kabos P, Sartorius CA, Tan AC, Pitts TM, Eckhardt SG, Diamond JR. The Role of p53 and Senescence as a Mechanism of Resistance to Aurora Kinase Inhibitors in Triple Negative Breast Cancer, University of Colorado Anschutz Medical Campus, Women's Health Research Day, October 2014

**Ionkina AA*,** Tentler JJ, Newton TP, Brunkow KL, Johnson JS, Tan AC, Pitts TM, Eckhardt SG, Diamond JR. The role of p53 family tumor suppressors in mediating response to the Aurora and angiogenic kinase inhibitor ENMD-2076 in triple-negative breast cancer. UCCC, Annual Retreat Poster Session, September 2014 (Top Poster Winner)

**Ionkina AA*,** Tentler JJ, Newton TP, Brunkow KL, Johnson JS, Tan AC, Pitts TM, Eckhardt SG, Diamond JR. The role of p53 family tumor suppressors in mediating response to the Aurora and angiogenic kinase inhibitor ENMD-2076 in triple-negative breast cancer, University of Colorado Anschutz Medical Campus, Department of Medicine Research Day, November 2013

***presenting author**

# ABSTRACT OF THE DISSERTATION

Engineering Multi-Component Bioluminescent Imaging Platforms and Transcriptome Analyses in Models of Metastatic Breast Cancer

by

Anastasia Alexeevna Ionkina

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2021

Professor Jennifer Prescher, Chair

Cancer metastases are driven by complex interactions among tumor cells, immune cells, and other cell types. Dissecting the roles of these cells is complicated by the inherent heterogeneity of the tumor environment. Currently we lack the models needed to profile cancer metastasis and monitor disease progression. Improved cancer models can be used in conjunction with noninvasive imaging probes to further visualize specific cell types and cellular interactions that are important to cancer progression *in vivo*. Bioluminescence is well suited for sensitive, noninvasive imaging in disease models. Bioluminescence relies on enzymes (luciferases) that generate light via the chemical oxidation of small molecule substrates (luciferins). No external excitation source is required, and enough tissue-penetrant light is released, enabling sensitive detection of cells and other biological features. Despite its broad applicability, bioluminescence has historically been limited to monitoring one cell population at a time and lacks the spatial resolution necessary to "see" cellular interactions relevant to cancer progression. This dissertation bridges the need for better methods to study metastatic progression by developing platforms that (1) enable multicomponent bioluminescence imaging *in vivo*, (2) recapitulate disease progression with new metastatic cancer models, (3) profile comprehensive transcriptome analyses of metastatic disease, and (4) facilitate tumor-immune cell monitoring through engineered imaging probes. Collectively,

the imaging methods and models developed in this dissertation enable a more detailed examination of tumor heterogeneity and metastatic disease progression *in vivo*.

# CHAPTER 1: Unraveling the complexities of metastatic cancer progression with optical imaging tools

## I. Introduction

Cancer metastasis is a dynamic, multi-cellular process that is difficult to study and treat with targeted therapies [1, 2]. During metastatic progression, malignant cells undergo dynamic cellular changes, avoiding immune system detection to colonize new metastatic sites [4]. The precise details remain unknown due, in part, to a lack of methods to monitor the inherently diverse and complex biology that drives metastatic disease. Thus, the most deadly and elusive processes in cancer metastasis such as tumor latency, organ tropism, and cancer stem cells remain poorly understood.

Optical imaging is one of the most popular modalities to monitor disease progression in preclinical studies. In particular, optical techniques enable noninvasive visualization and characterization of cancer-related events at the cellular and molecular level in various preclinical cancer models [5]. Recent advances in optical imaging technologies have been used to gain deeper insight into some of the most difficult and lingering areas of metastatic progression. Optical imaging tools have also guided downstream analysis of metastatic disease, enabling targeted investigation through flow cytometry, microscopy and sequencing analysis.

In this introductory chapter, I present different optical imaging tools and discuss their impact in understanding cancer progression. Specifically, I focus on two of the most popular optical techniques used for molecular imaging: fluorescence and bioluminescence. I outline the general mechanisms, in addition to comparing and contrasting the techniques in the context of

preclinical research. I describe recent applications of optical imaging techniques in cancer biology and evaluate their contributions to understanding a select category of the most complex and elusive areas of metastatic disease. These focus areas (Fig. 1.1) include metastatic disease progression, tumor heterogeneity, and the different impacts of the surrounding microenvironment on cancer spread. I further highlight downstream applications that optical imaging has enabled, including flow cytometry, FACS cell sorting, RNA-sequencing, and microscopy.

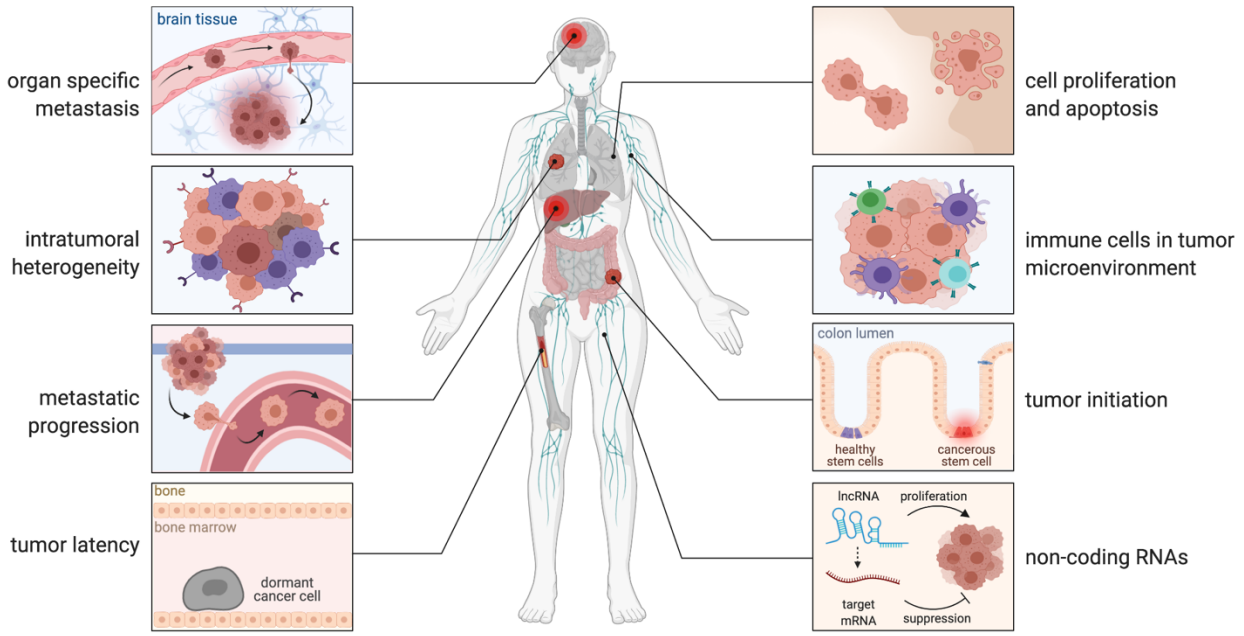


**Figure 1.1. Overview of areas relevant to metastatic disease**. The selected categories represent some of the most complex and poorly understood areas in the field of cancer metastasis.

## II. Overview of optical imaging techniques

Optical imaging techniques enable the noninvasive visualization and categorization of biological processes at the cellular and molecular level. Bioluminescence and fluorescence are the two most widely used optical imaging modalities. Both imaging techniques rely on the detection of photons emitted from excited state molecules. In bioluminescence (Fig. 1.2A), enzymes

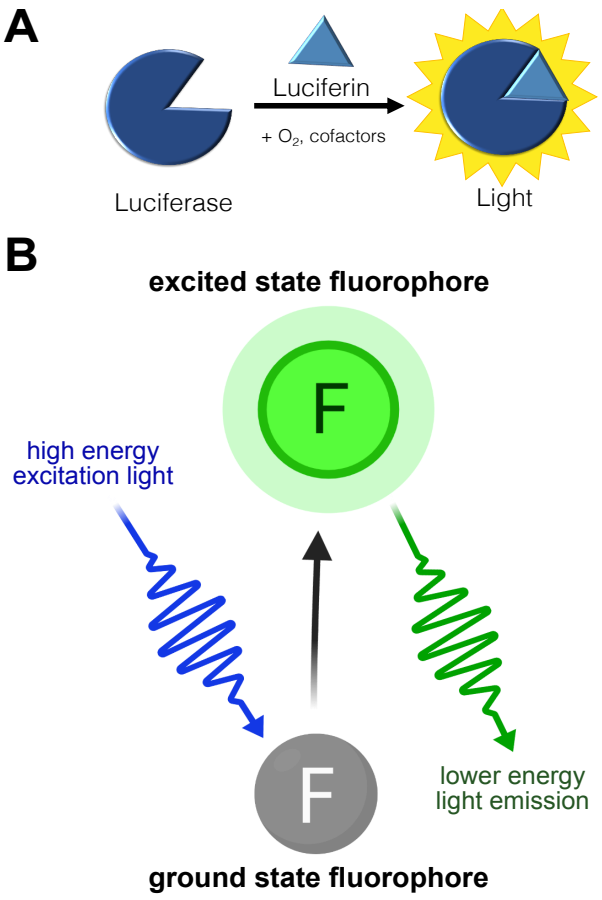


**Figure 1.2. Optical techniques for molecular imaging**. (A) Bioluminescence imaging involves enzymes (luciferases) that oxidize substrates, producing photons of light as products. (B) For fluorescence imaging, external light sources are used to excite fluorophores. Photons are emitted as the fluorophores relax to the ground state.

(luciferases) oxidize substrates (luciferins), releasing photons of light as one of the products. Fluorescence (Fig. 1.2B) relies on an external light source to excite fluorophores that emit photons as they come down to the ground state. Both imaging techniques are easy to use and relatively inexpensive, making them a staple in various preclinical research and diagnostic applications [6]. In preclinical research, optical imaging techniques are commonly used to monitor cell proliferation, and as reporters of gene expression in various cancer-associated cells [7-9]. Target cell populations can be exogenously labeled with fluorescent imaging probes to enable facile detection with limited off-target labeling [10]. Cells of interest can also be modified to express fluorescent and bioluminescent imaging probes. Visualization of these cells enables the study of cellular function within local environments and distal tissues [11]. Additionally, bioluminescent imaging probes can also be injected into an animal to produce a detectable and targeted signal [12].

Whole animal-imaging systems used to capture the emitted signal are sensitive enough to detect light-emitting probes at concentrations in the picomolar and femtomolar range [13]. Despite

the sensitivity of imaging systems, successful detection of optical imaging probes depends on various biological environments. Visualization of target cells can be obstructed by different cellular and molecular properties. The different environments in which imaging probes are used affect the quantity of signal that ultimately passes through biological barriers and is collected by the detector [6]. For example, thick tissue, increased fat deposits, and blood vessels readily scatter light and decrease resolution.

Optical imaging probes can also facilitate additional downstream analysis (Fig. 1.3) of target cells (commonly with flow cytometry, microscopy, sequencing). Optical techniques that show promise in preclinical applications are often fast-tracked for clinical approval. Thus, the translatability of optical probes remains highly desirable. For example, several fluorescent probes are commonly used as contrast agents and stains to label cancerous tissues. Clinical applications of optical imaging tools also aid in monitoring response to therapies, patient diagnosis, and during surgeries [14, 15].

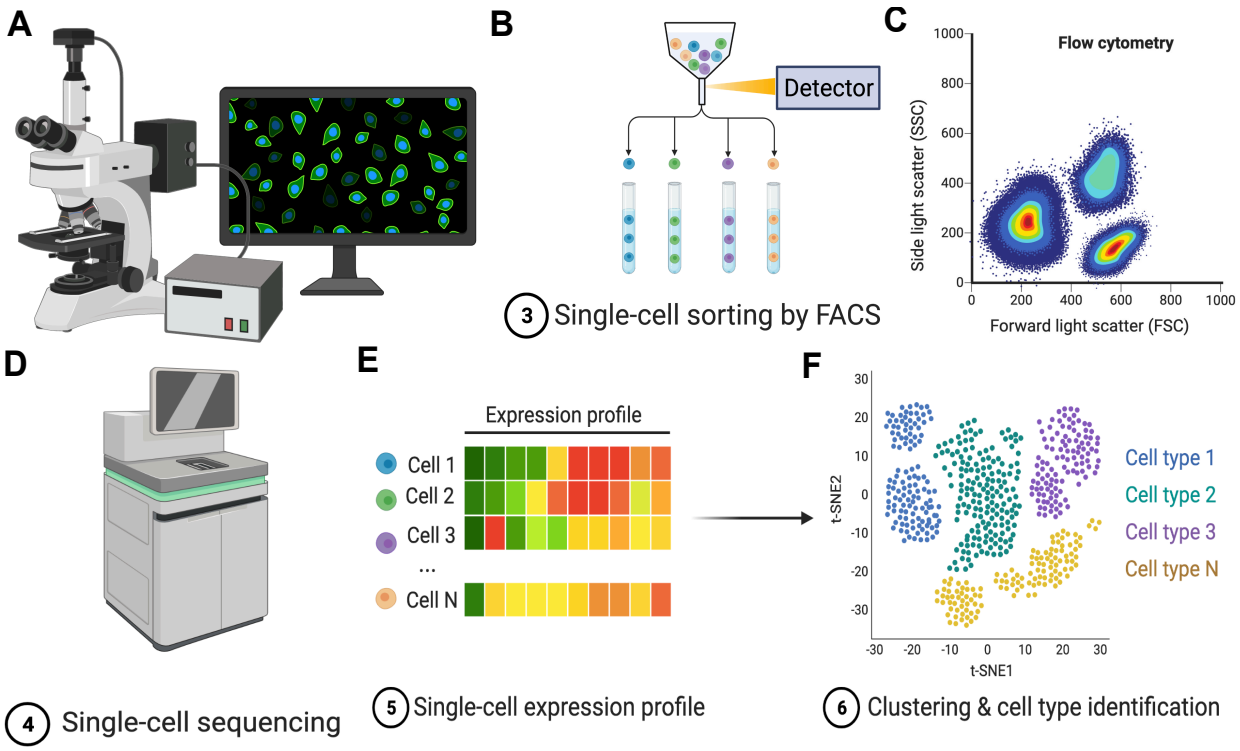


**Figure 1.3. Downstream applications of imaging tools used to study cancer metastasis.** Target cells labeled with fluorescence-based imaging probes enable further examination with (A) confocal microscopy imaging. (B) Fluorescence-activated cell sorting (FACS) and (B) flow cytometry enable the precise stratification and characterization of labeled cells. Selected cells can be subjected to (D) single cell-sequencing analysis to identify (E) specific expression profiles. (F) Gene expression profiles are used to cluster or segregated samples based on cell-specific identification markers.

## Imaging tool requirements to monitor cancer metastatic progression

Imaging tools exhibit an array of different sensitivities; the right choice depends on the biological application. Thus, proper probe selection will vary between the constraints needed to capture the specific event and the complexities in the metastatic process [4]. There are different requirements that specific imaging tools must meet in order to effectively probe metastatic progression. For example, intratumoral heterogeneity and the tumor microenvironment (TME) are both areas in cancer research that consist of many different cell types and complex cellular profiles. The multicellular composition is an inherent property of intratumoral heterogeneity and TME that

must be examined holistically. This requires the use of many different probes to discriminate and label various cell types. Additionally, the probes must be compatible with one another to allow multicomponent imaging of the various cell populations [16]. Achieving this degree of simultaneous, multi-cellular imaging remains challenging.

The ability to track tumor cell proliferation from the primary tumor to the growth of micro-metastasis is critical. Thus, for some applications, optical probes must facilitate non-invasive, longitudinal imaging of the model system while simultaneously providing information about the possible cellular evolution that occurs during cancer progression [17]. Tracking micro-metastasis requires high resolution imaging probes, preferably for detection at the single cell level, and through tissues in whole animals. No matter the cancer-specific application, all imaging probes must emit photons (either from external excitation or internal chemical oxidation) that can pass through any surrounding tissue for detection and measurement. As previously mentioned, fluorescence relies on an external light source to excite a fluorescent molecule resulting in the emission of a lower energy wavelength of light. Excitation of fluorescent probes by external light sources becomes increasingly difficult in whole animals. Thick tissues result in increased light scatter and reduced light penetration [40]. Imaging probes that emit sufficiently tissue penetrant photons face additional hurdles that can prevent successful detection. These hurdles include endogenous molecules in the surrounding microenvironment, such as heme and melanin, that readily absorb visible light and decrease imaging resolution in deep tissue (Fig. 1.4A-B).

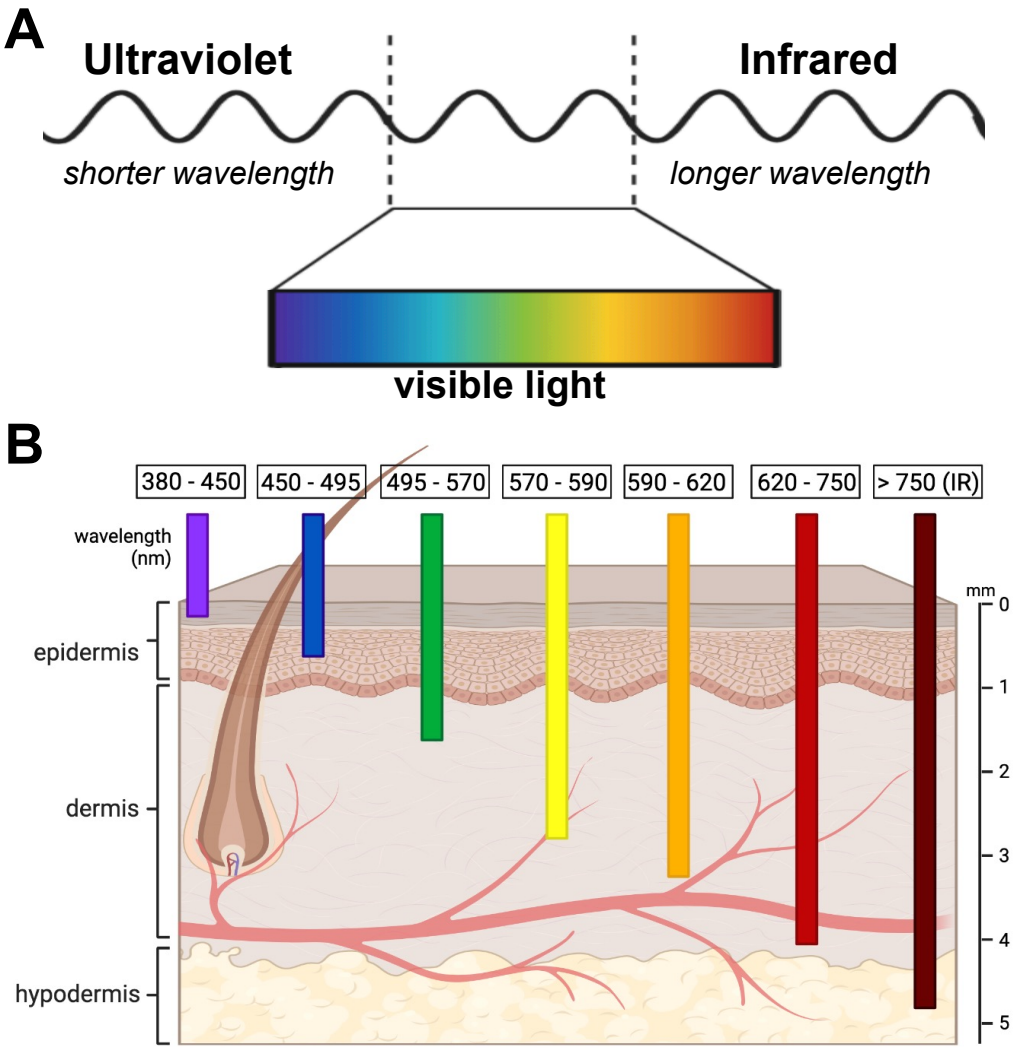


**Figure 1.4. Spectrum and penetration of light wavelengths through the different layers of human skin.** (A) Spectrum of visible light starting with near-ultraviolet light (short wavelengths) to near-infrared light (long wavelengths). (B) The differences in depth penetration for various wavelengths of visible light [3] are shown above. The different layers of the skin are listed on the left with the corresponding depth (mm) on the right. The more red-shifted light (ideally >650 nm) passes more readily through the different tissues. (Adapted from Ash C, et al. (2017) *Lasers Med Sci*. 32, 8)

**Fluorescence based optical imaging tools**

As noted earlier, fluorescence relies on an external light source to excite a fluorescent molecule. Upon relaxation from the excited state to the ground state, the fluorophore emits a longer-wavelength of light that is captured by an array of detectors used for visualization of the

tagged sample. Fluorescent tools enable versatile applications and are commonly used in imaging. The wide panel of colored fluorescent probes enables multifaceted monitoring of dynamic cellular and gene expression changes during disease progression [21, 22]. Fluorescent probes used in cancer imaging can be roughly divided into two categories: exogenous dyes (also used on conjugated antibodies and other biomolecules) and genetically encoded probes.

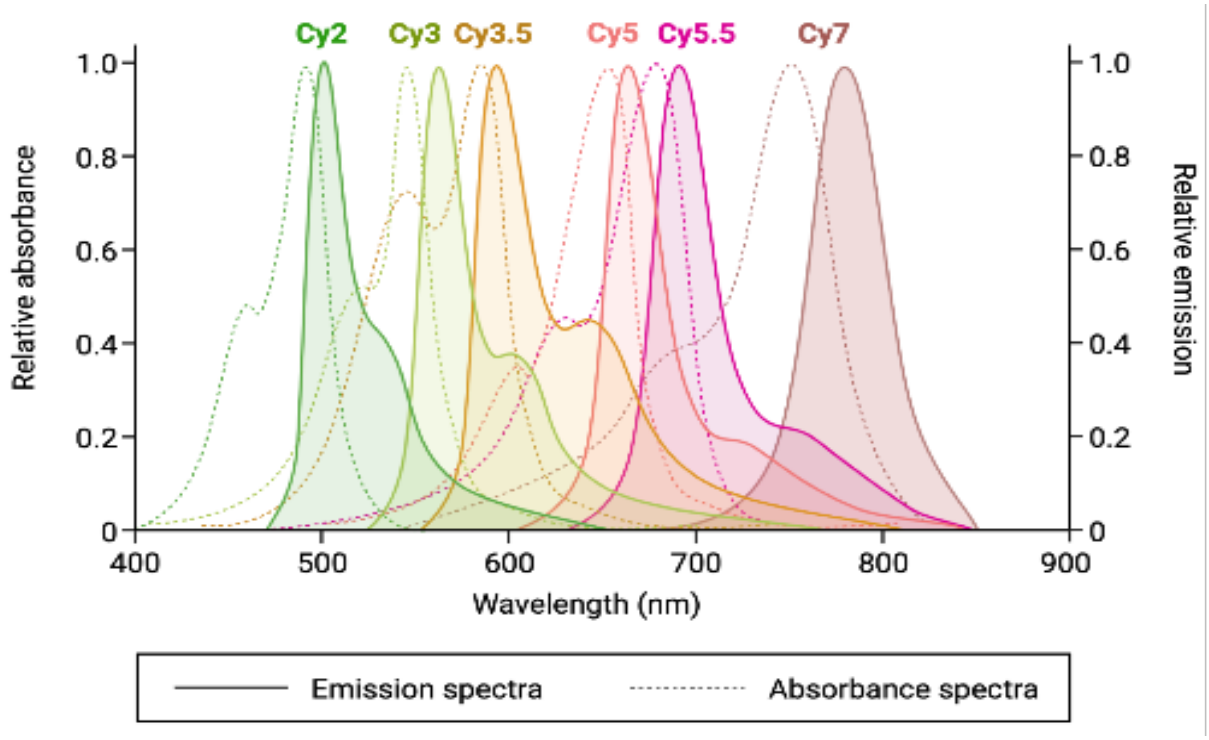


**Figure 1.5. Spectral profile of commonly used cyanine-based dyes.** Absorbance spectra (dotted lines) and emission spectra (solid lines) are shown for each dye. (Adapted from Altman, R. B., *et al.*, (2012). *Nature Methods*. 9,5)

A vast array of fluorescence molecules exist for internal cellular tagging and cell surface labeling [23]. Bright dyes are generally advantageous for most applications as they are easier to detect. The signal brightness is dependent on the fluorescence quantum yields of the dyes. Some examples of bright dyes are cyanine, squaraine, and porphyrin [24]. In particular, cyanine dyes (Fig. 1.5) (Cy3, Cy3.5, etc.) are commonly used in the field of cancer biology. Applications include

visualization of cellular proliferation, staining specific cellular compartments, and flow cytometry analysis. In addition, dyes and quantum dots, are commonly used as biological labels [23] for measuring apoptosis, and other cancer-relevant applications [11]. These and other fluorescent probes can label anything from small molecules to peptides. Koch and Ntziachristos also described the expansion of exogenous fluorescent probes to identify healthy versus diseased tissues through targeted antigen labeling [6]. Exogenous dyes have transitioned into the clinic where they provide surgical oncologists real-time feedback during procedures to ensure the accurate removal of micro-metastatic lesions [15]. Clinical imaging systems are also being optimized to enable exogenous dye-based diagnostics and early cancer detection methods [14].

The second major class of fluorescent probes consists of genetically encodable proteins. These probes are extremely useful to track cancer cell populations or monitor specific promoter-driven gene expression [25]. Figure 1.6 shows the most well-known genetically encoded reporter, the green fluorescent protein (GFP). First isolated by Osamu Shimomura from the jellyfish *Aequorea victoria* [26], excitation of the GFP chromophore by blue light (395 nm) results in an emission of green colored light (~509 nm). GFP is commonly encoded and used as a reporter gene for different applications in cancer biology. In these systems, the transcription of the reporter mirrors the transcription of the gene of interest. Such encodable probes facilitate non-invasive imaging in cancer models. Encodable probes are expanding quickly to monitor other cell types in the TME such as cancerous stromal cells [27]. Recent work from Regad et al., expanded the use of GFP reporter, developing a new approach to

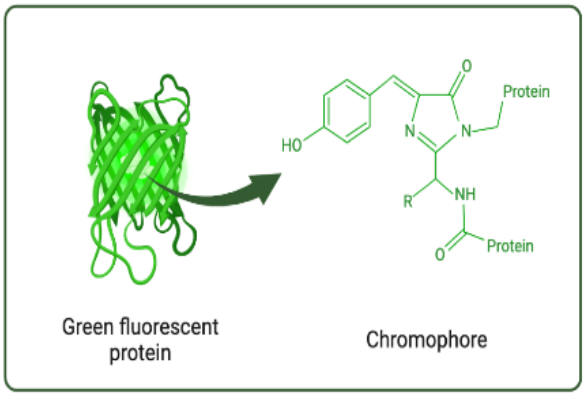


**Figure 1.6. GFP fluorescent protein and chromophore chemical structure.**

identify the rare and elusive cancer stem cell populations in prostate cancer [28]. Prostate cancer specific stem cell reporters were created by driving the expression of GFP under the stem cell promoter, NANOG. Due to their rarity, stem cells are notoriously difficult to isolate from the surrounding environment. By using the NANOG promoter-driven GFP cells, cancer stem cell isolation was possible. Fluorescent driven reporter systems have also been used in breast cancer metastasis models. In one example, the relative contribution of basal versus luminal cells to disease progression was measured by driving fluorescent protein expression under either keratin-14 (K14) or keratin-8 (K8) promoters [29].

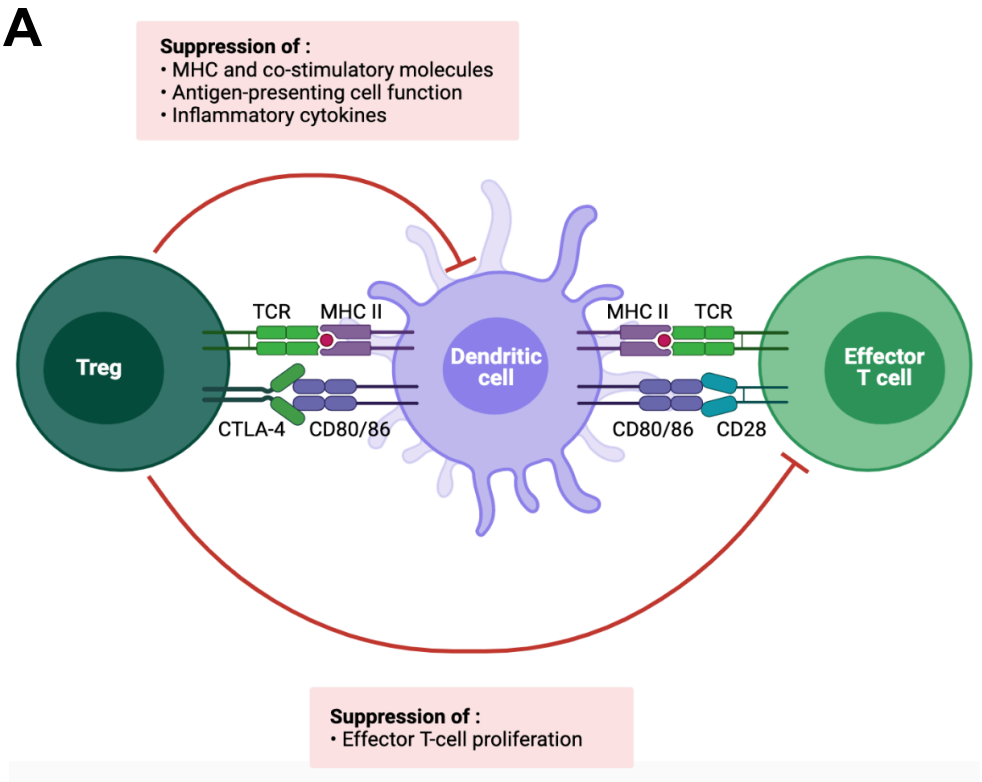


**Figure 1.7. Regulatory T cells (Tregs) have immunomodulatory functions.** Tregs exert suppression of other immune cells (i.e. dendritic cells and effector CD8+ T cells) through cell surface mediated interactions and other mechanisms such as secreted factors, nutrition completion and metabolism. Their ability to suppress effector T cells aids in the progression of metastatic disease.

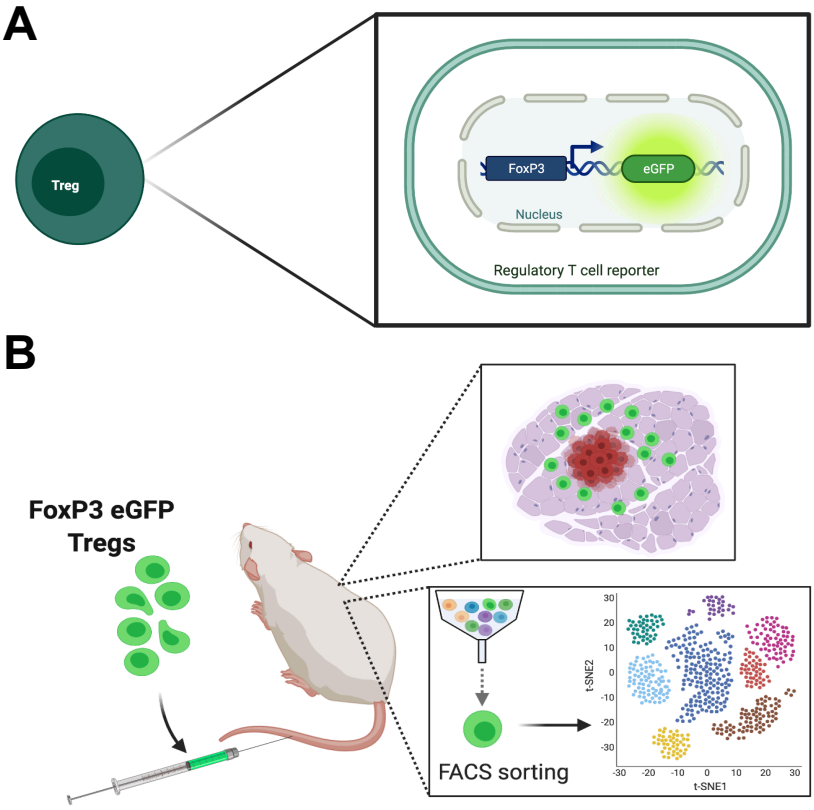


**Figure 1.8. Regulatory T cell reporter system**. (A) eGFP expression is driven via Treg transcription factor and master regulatory FoxP3 expression. (B) *In vivo* expression of FoxP3-eGFP cells in mice. *Ex vivo* analysis of GFP expressing Tregs allows for microscopy imaging or FACS sorting of Tregs followed by single cell analysis of gene expression.

Fluorescent protein reporters have also been incredibly useful for tracking gene expression of cancer-associated immune cells *in vivo* that would otherwise be difficult to study in cell culture. Regulatory T cells (Tregs) are one such class of immune cells that have been the focus of many preclinical and translational studies due to their immunomodulatory abilities to suppress other immune cells (Fig. 1.7) [30]. The mechanisms that Tregs use to suppress their microenvironments have been difficult to study *in vitro* due to the sheer number of cells that they control. To better study and monitor the suppressive capabilities of Tregs, Lin and colleagues created cell lineage specific reporters by targeting the Treg specific transcription factor, FoxP3 [31]. eGFP expression was driven by the endogenous promoter in the reporter mice (Fig. 1.8A). The expression of eGFP allowed easy ex vivo identification of Tregs from other surrounding tissues with immunofluorescence or FACS sorting (Fig. 1.8B). With Tregs and other immune cells coming to the forefront of research, the FoxP3-eGFP reporter mice have

been used extensively to study different facets of Treg biology including cell proliferation, localization, and functionality [32-35].

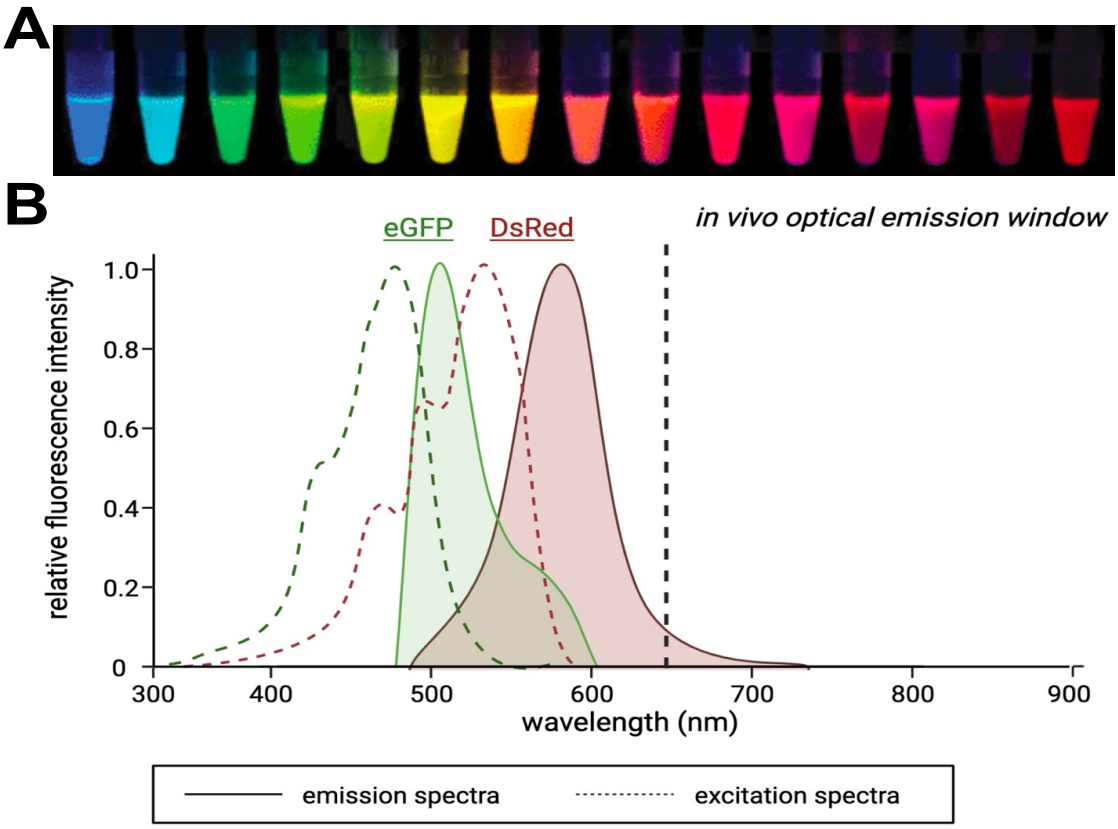


**Figure 1.9. Engineered GFP mutants have expanding the number of encodable proteins.** (A) An array of different colored encodable proteins enables multicomponent labeling (Shaner, N et al. Nature Biotech, 2004). (B) Excitation and emission spectrum of two popular fluorescent proteins. The relative fluorescence intensities of eGFP (in green) and DsRed (in red) are shown across wavelengths of light [3]. The black dotted line at the 650 nm wavelength denotes the *in vivo* optical emission window that is optimal for imaging (>650 nm).

Mutations to GFP can shift its original excitation and emission spectrum [36]. These engineered GFP derivatives have unique spectral profiles that can be used for multi-component imaging (Fig. 1.9A) [37]. Additional genetically encodable fluorescent proteins increases the number of cellular processes in metastatic disease that can be multiplexed and measured. Reporters that have different excitation and emission spectra (Table 1.1) such as eGFP and DsRed are frequently used in tandem (Fig. 1.9B). For example, the red-emitting DsRed and green-emitting

eGFP have recently been used as dual fluorescent imaging reporters to study changes in histone methylation and acetylation during cancer progression. In this dual system, DsRed resided in the cytoplasm while eGFP was linked to histone H2B. Nuclear expression of H2B has provided promising results as a possible biomarker to monitor prostate cancer progression [38]. Using this dual-colored system, Yang et al was able to detect early apoptotic responses to different chemotherapy agents in prostate cancer.

**Table 1.1. Optical properties of fluorescent proteins that are frequently used as dual reporters.** (Adapted from Dmitriy M. et al., (2010). *Physiological Reviews*. 90,3)

| Fluorescent Protein | Excitation (nm) | Emission (nm) | Quantum Yield | Rel. Brightness |
|---|---|---|---|---|
| EBFP | 383 | 445 | 0.31 | 27 |
| GFP | 395/45 | 509 | 0.77 | 48 |
| EGFP | 484 | 507 | 0.6 | 100 |
| RCFP | 439 | 476 | 0.4 | 39 |
| EYFP | 514 | 527 | 0.61 | 151 |
| mOrange | 548 | 562 | 0.69 | 146 |
| dTomato | 554 | 581 | 0.69 | 142 |
| DsRed | 558 | 583 | 0.79 | 176 |
| mRuby | 558 | 605 | 0.35 | 117 |

As previously mentioned, mutant fluorescent probes can exhibit red-shifted emission spectra (>650 nm) making them well suited for deep tissue imaging [18]. Unlike their blue or green counterparts, red-shifted probes are able to bypass absorption by the blood and be used in deep tissue settings. However, the longer-wavelengths that are associated with red-shifting probes are weaker compared to the high intensity wavelengths of the blue counterparts (Fig. 1.4A). Thus, red-shifted probes usually require large quantum yields [16]. To this end, work from Firnberg's group focused on increasing the quantum yield of dim-emitting infrared fluorescent protein (iRFP) [19]. NIR imaging probes (Fig. 1.4B), like iRFP, tend to emit light between 650 nm to 950 nm. Firnberg

addressed the poor emission power of NIR by developing two new versions of the protein, mRhubarb719 and mRhubarb720, that exhibited improved red-shifted spectra as well as increased quantum yields [19]. Other near infrared FPs are ideal for *in vivo* use as protein tags and gene reporters [6], but also lack brightness [20]. The need for brighter NIR FPs prompted Matlashov et al. to create further engineered version of the probes [20]. With miRFP720's wavelength peak emission of 720 nm, this genetically encodable probe is also the brightest NIR FP reported to date [20]. Matlashov showcased the power of the NIR probe to track cancer cells *in vivo*, successfully detecting ~$10^5$ labeled cells. The collection of these new red-shifted and NIR FPs have huge potential for multiplexed imaging to monitor cancer metastasis *in vivo*.

Multispectral fluorescent reporters have been adapted to monitor the complexities that arise with oncogenic clonal expansion *in vivo*. Snyder's group developed a new system to monitor changes in oncogenic clonal mutations and expansions. Dubbed Cancer Rainbow (Crainbow), this model fluorescently barcodes somatic mutations and enables the direct visualization of the spread of oncogenes during clonal expansion [39]. Crainbow uses cell type-specific Cre recombinase along with orthogonal lox sites to insert transgenes. This system is able to barcode each genetic state in the cells and is read out using fluorescence imaging. Crainbow was applied to colorectal cancer to demonstrate its viability. The modeling system analyzed different tumors and was able to identify the molecular mechanism driving intratumoral heterogeneity.

Despite the range of multispectral probes available, fluorescence based imaging techniques can be limited by the power capacity of the instrumentation and its effects on the fluorescent signal (Fig. 1.10). Fluorescent probes must be excited with an external light source in order to function [40]. Prolonged excitation of fluorescent probes can destroy the chromophore and cause photobleaching. To quantify the intensity of fluorescent light emitted by the variety of imaging probes previously stated in this review, the instrumentation requires a range of different detectors. These limitations can be improved by upgrading the optical filters, dichroic mirrors, and filter cubes with respect to the required excitation source and the capability of the detector (Fig. 1.10). One of the main challenges for fluorescent based optical imaging probes and instrumentation is improving signal loss and scattering. These issues are particularly important to address for imaging probes and detectors translating into clinical applications [41].

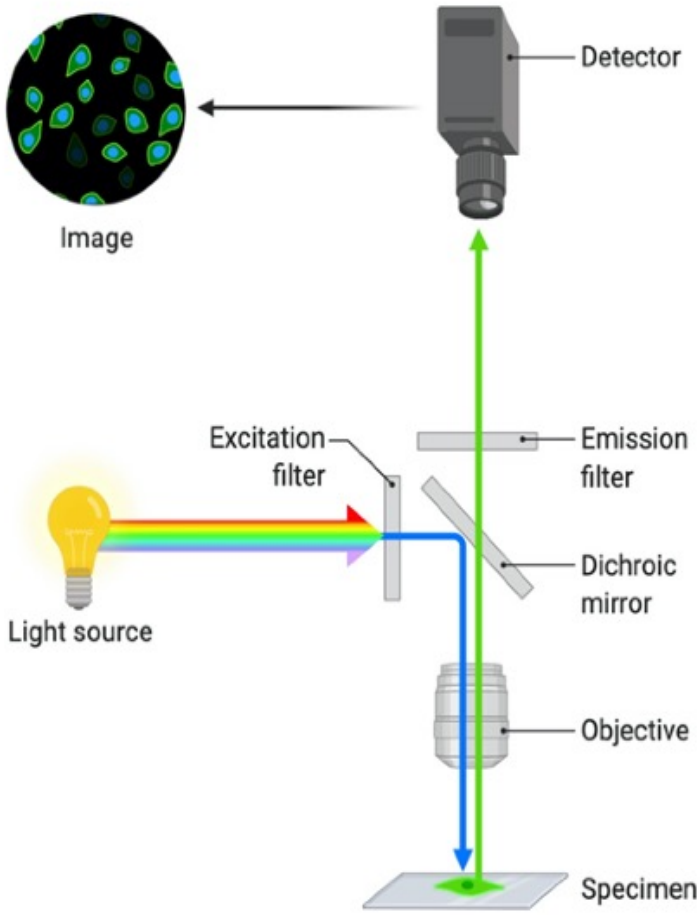


**Figure 1.10. Standard layout of fluorescence instrumentation.** Sets of filter and mirrors are used to focus in the external light sources to excite the fluorescent labeled sample and deliver the emitted light to a detector. The process results in the acquisition of fluorescently labeled samples.

**Bioluminescent based optical imaging tools**

Bioluminescence is a natural process that relies on enzymes (luciferases) that generate light via the chemical oxidation of small molecule substrates (luciferins) [48] (previously noted in Fig. 1.2A). Unlike fluorescence imaging, bioluminescence requires no excitation source, minimizing background emission and providing extremely high signal-to-noise ratios [49]. Mammalian cells inherently do not express luciferases which also improves background signal. Thus, bioluminescence imaging can be advantageous in heterogeneous environments, including rodents and other preclinical organisms. The emitted light can penetrate tissues and be detected using sensitive cameras. No external excitation source is required, and enough tissue-penetrant light is released, enabling sensitive detection of cells and other biological features.

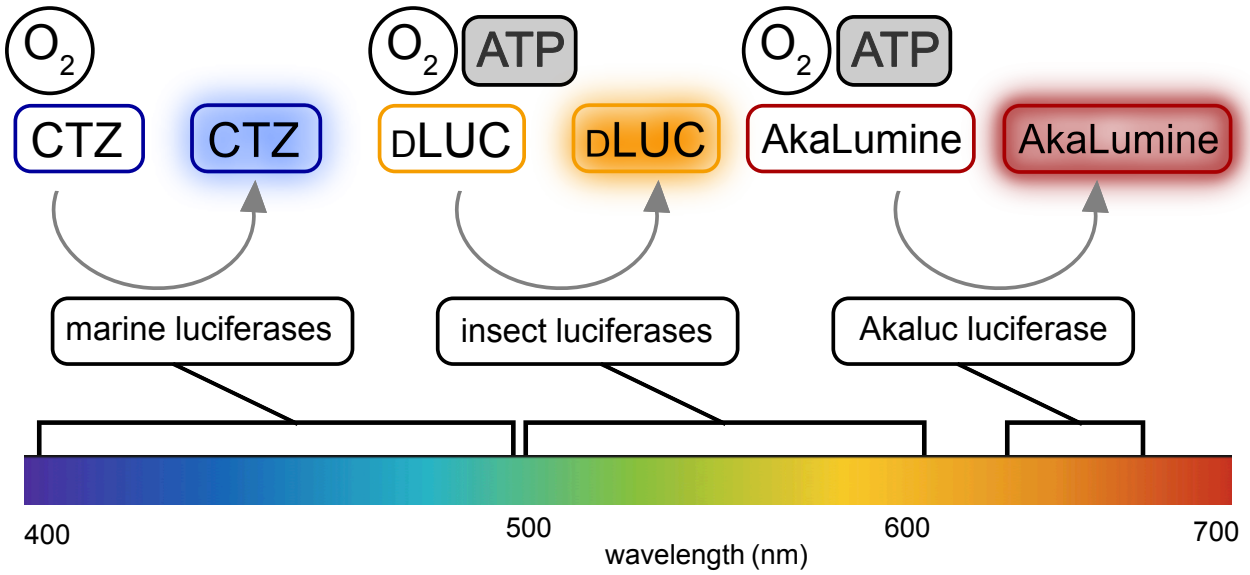


**Figure 1.11. Common emission spectra of luciferase-luciferin pairs used for bioluminescence imaging**. Luciferase-luciferin pairs commonly used for *in vivo* imaging. The corresponding wavelength of light emission is shown beneath each pair. Marine luciferases use coelenterazine (CTZ) in addition to oxygen to catalyze the reaction and produce a byproduct of light. Terrestrial (insect) luciferases use D-luciferin (D-luc) as their substrate along with oxygen and ATP. Akaluc, a firefly luciferase mutant, uses its synthetic luciferin (AkaLumine) along with oxygen and ATP to produce light. Adapted from Nasu Y, Campbell RE. (2018) *Science*. 23, 359)

Luciferases have been routinely introduced into cells and tissues and used as biological tags to track processes over time [44, 50] (Fig. 1.11). Cells are genetically engineered to express a bioluminescent reporter enzyme (e.g., firefly luciferase) and are subsequently treated with the enzyme's substrate (e.g., luciferin), which results in the release of photons upon catalysis (Table 1.2). The simplicity of the imaging method, high throughput capability, and the high detection sensitivity and specificity make this approach attractive for drug development, monitoring treatment response, and molecular interactions in animal models of human diseases.

**Table 1.2. Common luciferase enzymes with their origin, required substrates and cofactors, molecular weight, and maximum emission wavelength**. (Adapted from Zambito, G., *et al*., (2021). *Current Opinion in Chemical Biology*, 63)

| Luciferase | Organism | Substrate | Cofactor(s) | Size (kDa) | Emission wavelength (nm) |
|---|---|---|---|---|---|
| North American Firefly (Fluc) | *Photinus pyralis* | D-luciferin | ATP and Mg | 61 | 560 |
| Click Beetle | *Pyrophorus plagiophthalamus* | D-luciferin | ATP and Mg | 64 | ~600 |
| *Renilla* (Rluc) | *Renilla reniformis* | Coelenterazine | N/A | 36 | 480 |
| *Renilla* Mutant (RLuc8) | *Renilla reniformis* | Coelenterazine | N/A | 36 | 535 |
| *Gaussia* (Gluc) | *Gaussia princeps* | Coelenterazine | N/A | 20 | 470 |
| OLuc | *Oplophorus gracilirostris* | Coelenterazine | N/A | 19 | 460 |
| NanoLuc (NLuc) | *Oplophorus gracilirostris* | Furimazine | N/A | 19 | 460 |

Firefly luciferase (Fluc) and its substrate, D-luciferin (D-luc), are easy to obtain and are the most popular bioluminescent pair for preclinical research [51]. Other marine luciferase-luciferin pairs such as NanoLuc (NLuc)-furimazine and *Renilla* luciferase (Rluc)-coelenterazine are also used to monitor cancer progression. However, *in vivo* bioluminescence imaging has traditionally been dominated by Fluc and D-luc, whose 560 nm emission provides more tissue-penetrant

photons than the marine luciferase-luciferin counterparts. Additionally, substrate bioavailability has also been a key determinant in favoring the use of terrestrial luciferase-luciferin pairs *in vivo*. Improving the bioavailability and stability of marine luciferin substrates is an active area of research [55, 56]. Other luciferase-luciferin pairs have been further evolved for *in vivo* use. To optimize Fluc for deep tissue imaging, a red-shifted version of the enzyme (Akaluc) has been developed [52]. Akaluc along with its substrate (AkaLumine) emit one of the highest percentages of red (>650 nm) photons of any bioluminescent tool to date [53]. The palette of red-emitting luciferase-luciferin probes has further expanded with recent work from Promega Corporation and Mezzanotte's group [54]. Their engineering efforts resulted in novel click beetle mutant luciferase-luciferin pairs for *in vivo* imaging [54]. When paired with a naphthylamino luciferin analog, click beetle mutant luciferases (CBG2 and CBR2) emit red-shifted light at wavelengths of 660 nm and 730 nm, respectively. This dual-colored luciferase imaging system enabled the simultaneous monitoring of two different target cells *in vivo.*

**Table 1.3. Chart comparing bioluminescence and fluorescence imaging properties**. (Information from Tung, J., *et al*., (2016). *Neurophotonics*, 3,2. and Brovko, L. (2010). *In Bioluminescence and Fluorescence for in vivo Imaging*, 1,149.)

| | **Bioluminescence** | **Fluorescence** |
|---|---|---|
| **Signal-to-noise ratio** | High | Lower |
| **Image acquisition** | Minutes | Seconds |
| **Phototoxicity** | None | Possible |
| **Substrate** | Luciferin | None |
| **Imaging depth** | 1-2 cm | 1 cm |
| **Molecular sensitivity** | Picomolar | Picomolar |
| **Multicomponent imaging** | Usually 1-2 targets | Many targets |
| **Noninvasive imaging** | Yes | Yes |
| **Spatial resolution** | Organ (cm) to cellular (100um) | Tissue (mm) to subcellular (0.1um) |

While bioluminescence is a powerful platform for long-term and noninvasive imaging, it has historically been limited to visualizing only one cell type or biological feature at a time (Table 1.3). Optimal luciferases require the same luciferin substrate, precluding facile multi-component imaging. Recent work in expanding the bioluminescent toolbox as well as advances in imaging protocols and signal acquisition are enabling rapid multicomponent bioluminescence imaging [44].

Substrate availability and administration routes can affect photon emission, particularly for *in vivo* applications. Berger and colleagues [63] found that the substrate from marine luciferase-luciferin pairs exhibited a more homogeneous distribution among tissues when administered intravenously compared to intraperitoneal administration. Differences in luciferase kinetics have been shown to also affect the protocol and acquisition time for optimal detection. This differentiation of enzyme properties is most relevant with *in vivo* applications. Luciferases with glow-type reaction kinetics, such as Fluc, require an incubation step following intraperitoneal (IP) administration when used *in vivo* to achieve adequate distribution throughout tissues [63]. On the other hand, luciferases such as Gluc, Rluc, and NLuc exhibit flash type kinetics. These luciferases should be imaged immediately after intravenous administration of their substrates. Additionally, intracellular and extracellular conditions can affect the activity of luciferases. Factors such as pH, temperature, and $H_2O_2$ concentrations can greatly affect the proper synthesis, folding and secretion of luciferase enzymes [64].

**Fluorescence and bioluminescence are often used together**

Noninvasive monitoring of cancer cell populations with bioluminescence imaging is frequently coupled with ex vivo flow cytometry or microscopy analyses enabled by fluorescence imaging (Table 1.3). Bioluminescence imaging tools have the advantage of being useful for longitudinal non-invasive imaging and do not require the user to known the exact location of the target cells. Such exploratory imaging of unknown cell locations is readily achieved with bioluminescence imaging. The combination of the two modalities enables imaging across multiple scales [65].

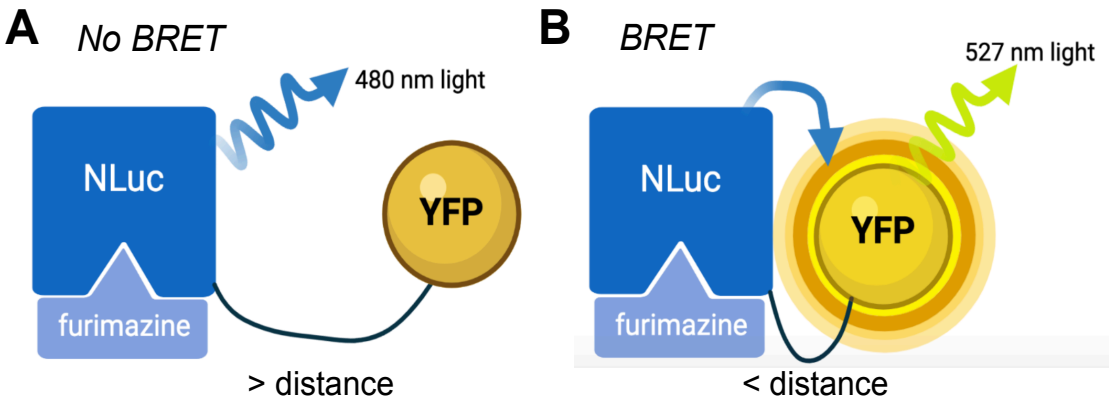


**Figure 1.12. Bioluminescence resonance energy transfer (BRET).** BRET probes contain both fluorescent (energy acceptor) and bioluminescent (energy donator) components that are linked together. NLuc emits blue wavelengths of light (480 nm) in the presence of its substrate, furimazine. (A) When the distance between NLuc/furimazine and the acceptor fluorophore (YFP) is great, BRET does not occur. The light emitted stays at 480 nm. (B) When the proximity of NLuc/furimazine and YFP is closer, BRET can occur. The original 480 nm light donated by NLuc is transmitted to YFP. The acceptor fluorophores shifts the light, emitting it at 527 nm. (Background information adapted from Promega).

Color shifting can also be achieved via bioluminescence resonance energy transfer (BRET). BRET probes consist of both fluorescent and bioluminescent components that are linked together (Fig. 1.12A,B). These constructs use a luciferase enzyme to donate the light source and a neighboring fluorescent protein(s) that acts as an acceptor, shifting the wavelength of emission

[57] (Fig. 1.12B). Addition of a luciferin analog triggers light emission from the paired luciferase. BRET systems harness the power of luciferases to excite fluorescent proteins internally instead of using an external light source (Table 1.3). This improves the use of the probes in animal models and deep tissue applications. The addition of fluorescent proteins with specific excitation/emission profiles can be used to red-shift the light emitted by most luciferase enzymes for *in vivo* applications [58].

One of the most useful classes of BRET reporters combines the brightness of blue-emitting luciferases with different fluorescent proteins to shift the original emission spectrum while retaining brightness. Brightly emitting BRET based Nano-lantern probes are effective in shifting NLuc's blue-emitted light based on the properties of the different colored fluorescent proteins [59, 60]. The wide array of BRET probes have different emission spectra and can be multiplexed to image various cellular properties. BRET probes have also been engineered to function as sensors for cancer metabolic applications and enzymatic driven processes [61]. Recent work by Kuchimaru et al. [62], detailed the design of a novel BRET-based imaging tool for the detection of ubiquitin proteasome regulated hypoxia-inducible factor (HIF) activity *in vivo*. The BRET reporter was expressed in cancer cells that exhibited various degrees of HIF expression. The increase of cellular HIF triggered the confirmational change of the BRET probe and resulted in signal production. This reporter was designed by placing the luciferase proximal to the HIF recognition domain, thus ensuring BRET only in HIF-overexpressing cells.

Additional BRET probes have been developed as powerful reporter genes. Work from the Amelio group has produced a suite of "LumiFluors" fusions that exhibit bright BRET emission profiles using enhanced fluorescence acceptors [66]. By utilizing the improved stability of the

NanoLuc system, LumiFluor reporters were better suited for visualizing tumorigenesis *in vivo*. BRET probes also enable facile analysis of the targeted tumor cells ex vivo by flow cytometry.

## III. Focus areas in cancer metastasis and the application of optical imaging tools

Metastatic initiation and progression remain poorly understood due, in part, to the difficulty of monitoring and capturing the differences in cellular kinetics. Optical imaging tools have enabled new insight into several of the most elusive areas of metastatic disease progression. Here we highlight some of the recent applications of optical imaging techniques and describe their contribution to understanding the specific categories of metastatic disease. We focus on (A) examining the origins of cancer initiation and metastatic progression, (B) cellular proliferation and apoptosis (C) organ specific metastasis, (D) tumor latency, (E) multi-component imaging of the tumor microenvironment, (F) miRNA targeted imaging, and discuss the use of optically guided imaging tools in studying (G) gene expression during metastatic progression.

### Examining the origins of cancer initiation and metastatic progression

Metastases are the leading cause of all cancer related deaths; however, the mechanisms that drive metastatic progression are not fully understood. Complex steps and stages for different cancers and different subtypes of the same cancer expands this unknown (Fig. 1.13). It is very important to study these processes in animal models where circulation systems and organs are intact. Extracellular matrix remodeling is one of the initial steps in the metastatic cascade that allows cancerous cells to invade their surrounding environments. Bochner et al., examined the initial process of extracellular matrix remodeling in mouse models of ovarian cancer with

bioluminescence imaging [67]. They were able to identify stromal biomarkers in metastatic ovarian cancer and quantity the increase in fibrosis during disease progression.

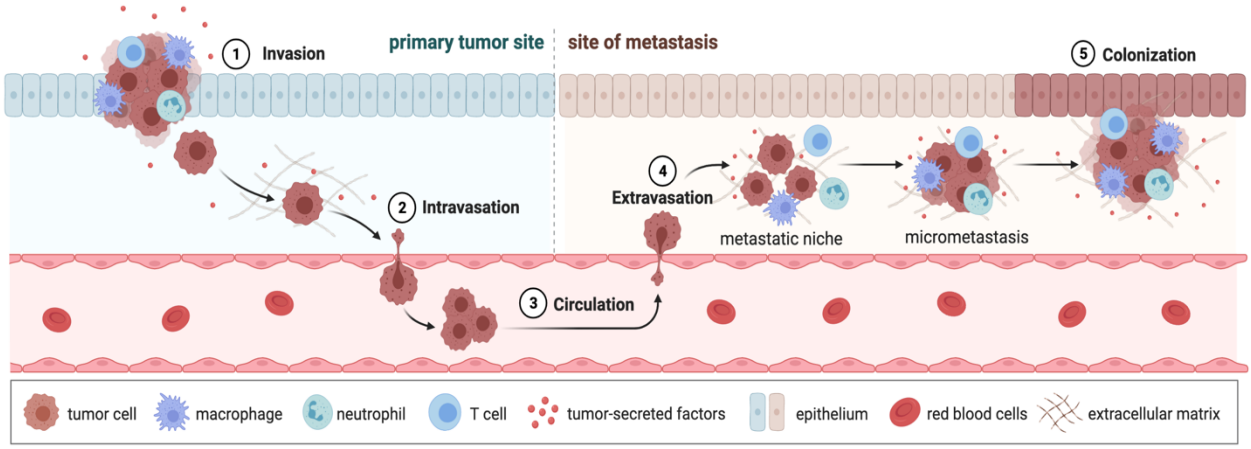


**Figure 1.13. Overview of metastatic cascade.** Different cell types involved in the process are noted in the legend. Tumor cells grow and start to remodel the surround matrix to allow for (1) invasion to surrounding tissues. Once the disseminated cells have reached a blood source, they go through (2) intravasation to enter the blood vessel and enter the (3) blood circulation. Disseminated cells undergo (4) extravasation in a new site of metastasis and develop their metastatic niche. Metastatic lesion are able to (5) colonize the new tissue and continue growth.

The critical role of circulating tumor cells in enabling metastatic progression and colonization of distal organs has been illuminated by optical imaging tools. Work from Aceto et al., and Kim et al., used luciferase-expressing cells to identify that circulating tumor cell (CTC) clusters play a central role in breast cancer metastasis [68]. Kim et al., continued applying these reporters to study specific cytokine profiles from CTC cells and breast tumors [69]. Experiments using these reporters showed that tumor-derived cytokines (IL-6 and IL-8) acted as CTC attractants whereas MMP1 and the actin cytoskeleton component fascin-1 were mediators of CTC infiltration into mammary tumors [69]. To successfully achieve metastatic colonization (Figure 1.13), malignant cells must also prompt migration and metastatic invasion as they evade the immune system. Using bioluminescence imaging to monitor lung colonization in melanoma and breast cancer cells, Tello-Lafoz et al., discovered that enhanced expression of myocardin related

transcription factors (MRTFs) promoted migration and metastatic invasion [70]. Paradoxically, MRTFs were identified to compromise lung colonization in an immune-mediated manner [70]. The use of optical imaging tools enabled Tello-Lafoz et al., to identify elevated expression of MRTF in cancer cells as a novel biomarker for sensitization to immune checkpoint blockade therapies.

Metastatic initiation and cancer stem cells are thought to play a big role in treatment resistant disease [71]. Although cancer stem cell (CSC) identification and characterization in metastatic progression is focus of much research, it remains very controversial [72, 73]. Normal stem cells have often been a benchmark of comparison for CSC populations. Bioluminescence imaging has aided the comparison of these two different cell populations by allowing facile monitoring of differentiation states and cellular vitality *in vivo* [74, 75]. By comparing the cellular and genomic similarities between stem cells and CSC, these studies aimed to identify possible targets that could be leveraged for treatment. Work from Choi et al [76] and Dong et al [77] both used bioluminescence imaging to track possible glioblastoma cancer stem cell markers, identifying histone deacetylase inhibitors and inositol 145-triphosphate receptor dependent calcium signaling, respectively. These CSC-specific markers could be used for targeted glioblastoma therapies. Optical imaging tools were similarly used to investigate the ability to suppress breast cancer stem cells [78]. These studies identified that downregulation of the β-catenin/Notch1/Akt signaling pathway after antrocin treatment resulted in the suppression of stemness properties [78]. Optical imaging tools have also been used to study pathway activation upon CSC outgrowth. Aguirre-Ghiso et al., used fluorescent protein tagged CSCs to identify pathway activation that occurs during primary and metastatic growth through extracellular signal-regulated kinase and p38 pathways [79].

**Cellular proliferation and apoptosis**

Growth of the primary tumor and the impact of the speed in which metastatic progression follows is important in solidifying the timeline and aggressiveness of different cancers in preclinical and clinical scenarios. Visualizing this process in preclinical cancer animal models enables the identification of metastatic progression, tumor-specific aggressiveness, and measurement of tumor burden over time. Non-invasive optical imaging tools are especially useful for monitoring and quantifying tumor growth longitudinally without sacrificing the animal. Pioneering work by Contag showcased the use of bioluminescence imaging to assay tumor proliferation in animal models [80]. Further examples from Scheffold and colleagues used bioluminescence imaging to evaluate the anti-tumor properties of lead drug candidates in preclinical HER2 positive xenograft models [81]. The most impressive example demonstrated that modified CD8+ natural killer T (NKT) cells could be successfully redirected to target luciferase-expressing tumor cells *in vivo*. Engineered luciferase-expressing tumor cells have also been used to measure apoptosis *in vivo* (Fig. 1.14). Work by Scabini and colleagues using a novel caspase 3/7-luciferin substrate, Z-DEVD-aminoluciferin, to

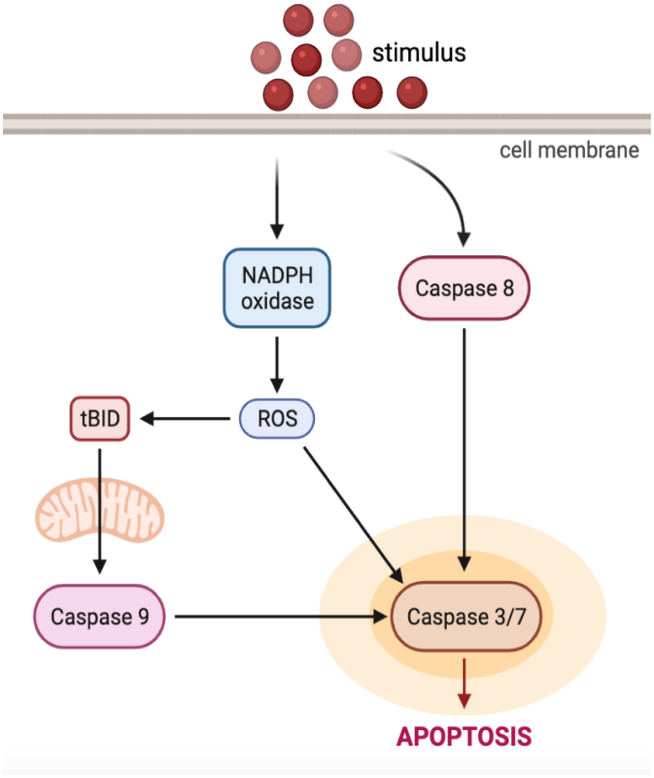


**Figure 1.14. Schematic outlining caspase 3/7 mediated cellular apoptosis.**

track cell death *in vivo* [82]. This approach enabled monitoring of cell death through caspase-3 following treatment in colorectal and glioblastoma cancers.

**Organ specific metastasis**

The propensity of metastatic cells to colonize specific organs remains unknown, but the clinical ramifications of creating targeted drugs to prevent metastatic lesions could increase cancer patient survival (Fig. 1.15). Imaging tools allow real-time and longitudinal monitoring of micro-metastasis lesions to answer these questions and the gene expression patterns that are related to specific organ colonization in pre-clinical models.

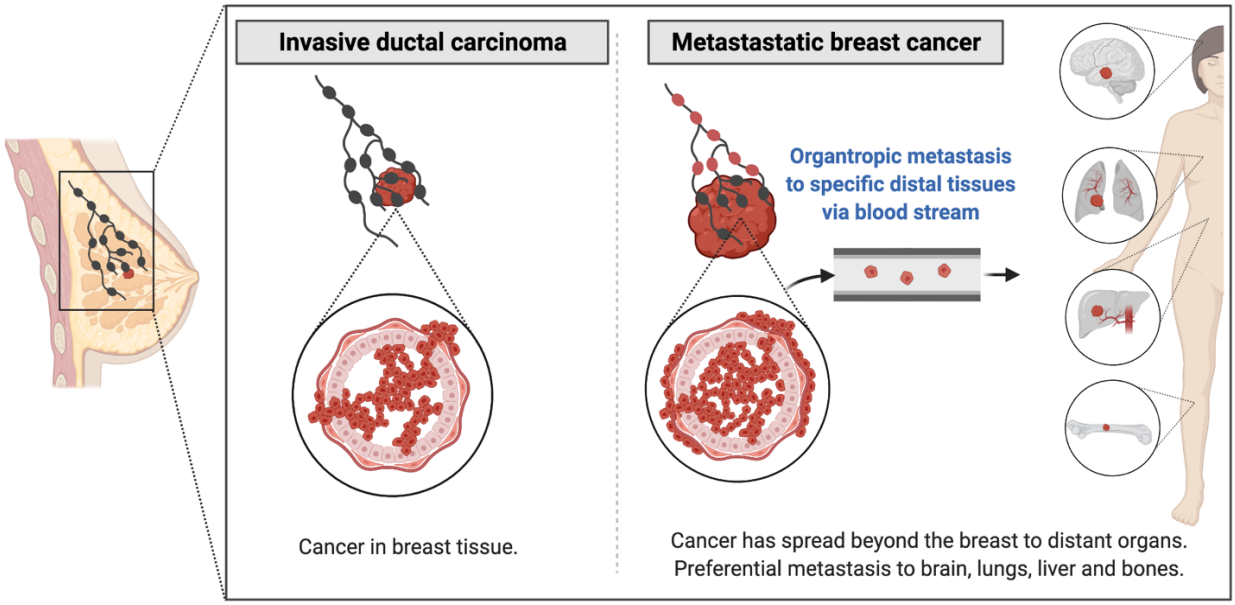


**Figure 1.15. Organotrophic metastasis of distal organs in breast cancer.** Breast cancer progresses from invasive ductal carcinoma into metastatic disease. Disseminated breast cancer cells spread to other parts of the body by way of the blood stream or lymphatic system. Breast cancer cells home to specific secondary tissues and exhibit organ-specific metastatic colonization of the brain, lungs, liver, and bone marrow.

Bioluminescent imaging tools have been critical in enabling the identification of metastatic progression in specific organs. Using luciferase expressing MDA-MB-231 breast cancer cells, Jenkins et al were able to achieve sensitive detection of both primary and secondary tumor sites *in*

*vivo* that were clinically relevant [83]. Work from Massague further examined organ-specific patterns of metastasis using luciferase-expressing breast cancer cells. This work identified specific tissue microenvironment gene expression patterns that mediated the reporter cell's ability to colonize specific organs [84, 85].

Optical imaging technologies have expanded the ability to study organ-tropism in dense metastatic tissues, like the bone and brain, that have traditionally been difficult to access. For example, bioluminescence imaging enabled the monitoring of bone metastasis and their response to hypoxia-inducing therapies [86]. Further investigation into the cellular and mechanistic origins of metastatic organ-tropism identified a specific subset of stromal cells as key mediators of disease progression [87]. These primary breast tumor-derived stromal cells produced CXCL12 and IGF1 that selected primary breast cancer cells to metastasize to the bone marrow by mimicking the microenvironment of the bone marrow. Thus, the production of CXCL12 and IGF1 from the stromal cells primed the pre-malignant breast cancer cells for bone marrow metastasis [87]. Optical imaging has subsequently enabled the development of brain-specific metastatic cancer models [88]. Additional use of fluorescent-based optical probes allowed Kienast et al to use multiphoton laser-scanning microscopy to track in real-time the fate of individual metastasizing cancer cells in the brain [89].

**Tumor latency**

Tumor latency causes many patient relapses after decades of being in remission and has been a central focus in metastatic cancer research. However, these rare cell populations have been difficult to identify and the signaling pathways mediating latency are heavily disputed. Optical imaging tools have been the central method of identifying latent cells *in vivo* and facilitating further downstream gene expression analysis. Early work from Morris et al., labeled metastatic cell

populations with fluorescent based nanospheres to track cell division and enable the identification of dormant cell clusters [90]. Cameron and coworkers coupled this nanosphere optical imaging tool with fluorescent microspheres to further monitor cell survival of solitary dormant cancer cells in different lung microenvironments [91]. Work from Massague used GFP and luciferase reporters for *in vivo* monitoring of dormant cancer cells [92]. This rare dormant cell population was isolated and enriched for subsequent *in vitro* analyses. Notably, these optical imaging reporters facilitated gene expression analyses that identified enrichment of SOX transcription factor and natural killer-cell mediated immune evasion as key signatures of dormant cancer cells [92].

**Multi-component imaging of the tumor microenvironment**

The tumor microenvironment has been shown to be important in modulating cancer proliferation and metastatic progression. This multicellular process involves a dynamic range of different signaling pathways. Therefore, multicomponent optical imaging tools are essential to capture the heterogenous environment. Calon et al investigated TGF-β signaling from stromal cells and their impact on colorectal metastases. Using dual Fluc/Rluc luciferase reporters, TGF-β signaling from the stroma was shown to increase the efficiency of organ colonization in colorectal cancer [93].

Immune cells in the surrounding microenvironment have been identified as playing an important role in cancer prevention and progression. Methods to target immune cells as treatment for metastatic disease recently came to the forefront of cancer research [94]. Although there have been encouraging success stories of immunotherapies, the majority of cancer patients do not respond to current checkpoint inhibitors and cell based immunotherapies. Optical imaging tools are a staple in monitoring gene expression profiles. Zhang et al., used bioluminescence imaging to

monitor changes in metastatic breast cancer as subjects were treated with modified immune cells that mediated antitumor responses through TIPE2 gene therapy [95].

Cancer immune responses involve a wide array of different immune cell types. In order to capture this heterogeneity, multicomponent imaging tools that enable high resolution visualization of the relevantly small sized immune cells compared to tumor cells are needed. Fluorescent based imaging tools offer a wide array of probes to target such heterogenous environments. However, the requirement of an external light source limits applicability in deep tissues without an additional tissue clearing step. Lee et al., reported a streamlined and three-dimensional approach to achieve high-resolution maps of the tumor-immune microenvironment and tumor heterogeneity [96]. The developed transparent tissue tomography (T3) tool enabled easy multiplexed immunofluorescent tumor imaging.

**miRNA targeted imaging**

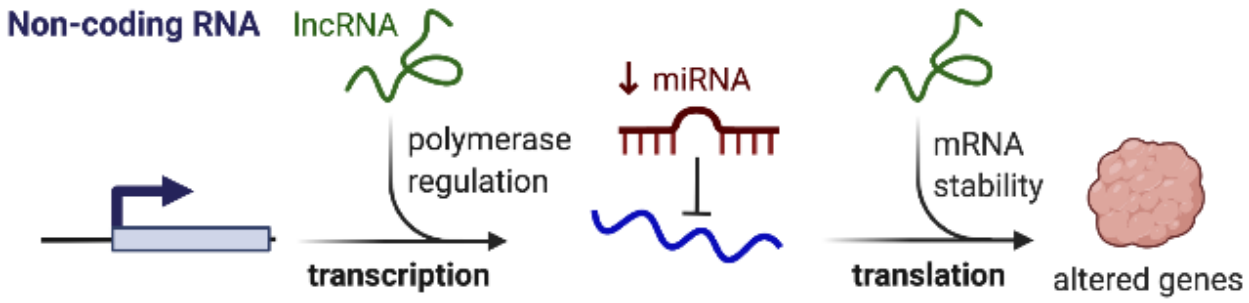


**Figure 1.16. Non-coding RNA effects on cancer progression.** Long-non-coding RNAs (lncRNAs) can regulate polymerase functionality, affecting transcription. Micro-RNAs (miRNA) can have inhibitory effects on nascent RNA and mRNA. lncRNAs can also affect mRNA stability and translation.

Once overlooked, recent studies have emphasized the vast impact and roles that mRNA and miRNA play in modulating metastatic cancer (Fig. 1.16). The inherent instability and sensitivity of RNA has made studying its function difficult. However, bioluminescence imaging

technologies have been demonstrated to be well suited for studying RNA and assessing its efficiency as a cancer therapy [97].

Bioluminescence imaging reporters are also used to study miRNA activity in cells. In particular, the significance of miRNA expression has been examined in mediating breast cancer metastasis. Using bioluminescence imaging reporter assays, Taipaleenmaki and colleagues identified that *in vivo* metastatic breast cancer models exhibited an upregulation of miR-218-5p expression in secondary metastatic lesions specific to the bone marrow [3]. Similarly using bioluminescence imaging tools, Kim et al., revealed that posttranslational modification of the human *microRNA-335* locus resulted in genetic deletion and epigenetic promoter hypermethylation across a panel of patient derived malignant breast cancer samples [98]. Interestingly, miR-335 was identified as having robust anti-tumor initiation and metastasis suppressive properties. Liu and colleagues engineered a novel reporter gene, Luc2/tdT_miR200c_3TS, to investigate the impact that miR-200 exhibited on metastatic progression and chemotherapy resistance [99]. Both *in vitro* and *in vivo* studies determined that high levels of miRNA expression were associated with low levels of metastases. Thus, the use of optical imaging tools aided in the identification miR-200 as a marker of therapeutic efficacy for drug targeting breast cancer stem cells [99].

**Optically guided studies of gene expression**

As mentioned before, optical imaging tools can guide downstream analysis of specific cell populations that can further be profiled for differential gene expression. After genes of interest have been identified, optical imaging tools can be leveraged again to non-invasively visualize real-time gene expression changes *in vitro* and *in vivo*. Gene reporters are common tools for providing fluorescent or bioluminescent signals that track with specific genes of interest during cancer

progression (Fig. 1.17A). Recent research from Wu et al., utilized optical imaging reporters to measure tumor hypoxia by driving luciferase expression with a hypoxia-relevant HRE promoter [100]. Bhang and colleagues used bioluminescence imaging to monitor tumor specific progression factors *in vivo* [101]. By using a progression elevated gene-3 (PEG-3) promoter to drive the expression of Fluc, they verified that high PEG-promoter activity was observed with melanoma and breast cancer [101]. Tissue-specific promoters can also drive the expression of fluorescent proteins that can bridge *in vivo* and *in vitro* use (Fig. 1.17B-C).

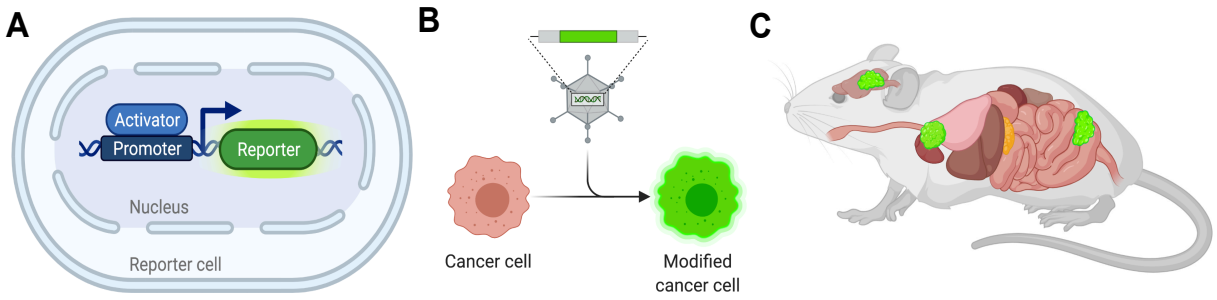


**Figure 1.17. Use and function of reporters *in vivo*.** (A) Endogenous promoters drive expression of fluorescent or bioluminescent based reporters. (B) Reporters can be delivered into cancer cells with lentiviruses or other methods *in vitro*. (C) Reporter cells can be injected into animals and their expression can be monitored across different organs (shown in green).

Dual fluorescent and bioluminescent gene reporters are powerful tools to measure metastatic progression. Understanding the complex role of TGF-β in cancer progression and bone metastasis is an example of cancer biology where dual gene reporters have historically been used both *in vitro* and *in vivo* [103-105]. Work from Massague examined other TGF-β related pathways that influenced breast cancer bone metastasis by monitoring Smad signaling with fluorescent and bioluminescent reporters [106]. In addition, substrate resolved Rluc and Fluc allow the expression patterns of two genes to be monitored simultaneously. Work from Winslow et al., used this dual

bioluminescent gene reporter system to monitor *HMGA2* and thymidine-kinase promoter activity in relation to their ability to suppress lung adenocarcinoma progression [107]. Furthermore, fluorescent and bioluminescent dual reporters can also be designed to read out specific protein-protein interactions. In these systems, proteins of interest are tethered to two luciferase fragments. Protein interactions are measured by light emission that results from the complementation of the two luciferase fragments upon close interaction of the desired proteins [108].

## IV. Limitations of current imaging tools

Current imaging tools have limitations that are problematic when applying them to studying cancer animal models. There are also some limitations to using the imaging tools to monitor cancer disease progression. A lack of multicomponent imaging probes has precluded orthogonal imaging of multiple cell populations. Current imaging probes prevent monitoring and identification of cellular interactions precluding the ability to measure immune-cancer interactions that are important for disease progression. Many popular probes emit blue wavelengths of light that makes them difficult to use for deep tissue and whole animal applications. Imaging with existing bioluminescent probes also usually requires multiple days to allow for clearance of the first substrate before a second one can be imaged, thus precluding the ability to acquire rapid images and read out of the target cell populations. This thesis works to address these limitations and improve our understanding of metastatic disease in cancer models.

## References


1. Dagogo-Jack, I. and A.T. Shaw, *Tumour heterogeneity and resistance to cancer therapies.* Nat Rev Clin Oncol, 2018. **15**(2): p. 81-94.
2. Ionkina, A.A., et al., *Transcriptome analysis of heterogeneity in mouse model of metastatic breast cancer.* Breast Cancer Res, 2021. **23**(1): p. 93.
3. Taipaleenmäki, H., et al., *Antagonizing miR-218-5p attenuates Wnt signaling and reduces metastatic bone disease of triple negative breast cancer cells.* Oncotarget, 2016. **7**(48): p. 79032-79046.
4. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
5. Martelli, C., et al., *Optical imaging probes in oncology.* Oncotarget, 2016. **7**(30): p. 48753-48787.
6. Pirovano, G., et al., *Optical Imaging Modalities: Principles and Applications in Preclinical Research and Clinical Settings.* J Nucl Med, 2020. **61**(10): p. 1419-1427.
7. Hadjipanayis, C.G., et al., *Current and future clinical applications for optical imaging of cancer: from intraoperative surgical guidance to cancer screening.* Semin Oncol, 2011. **38**(1): p. 109-18.
8. Herranz, M. and A. Ruibal, *Optical imaging in breast cancer diagnosis: the next evolution.* J Oncol, 2012. **2012**: p. 863747.
9. Wang, D. and J. Xia, *Optics based biomedical imaging: Principles and applications.* Journal of Applied Physics, 2019. **125**(19): p. 191101.
10. Wang, Y., et al., *Affibody-targeted fluorogen activating protein for in vivo tumor imaging.* Chem Commun (Camb), 2017. **53**(12): p. 2001-2004.
11. Wu, Y., et al., *Optical imaging of tumor microenvironment.* Am J Nucl Med Mol Imaging, 2013. **3**(1): p. 1-15.
12. Puaux, A.L., et al., *A comparison of imaging techniques to monitor tumor growth and cancer progression in living animals.* Int J Mol Imaging, 2011. **2011**: p. 321538.
13. Sandhu, G.S., et al., *Whole animal imaging.* Wiley Interdiscip Rev Syst Biol Med, 2010. **2**(4): p. 398-421.
14. Randall, L.M., et al., *A phase II, multicenter, open-label trial of OTL38 injection for the intra-operative imaging of folate receptor-alpha positive ovarian cancer.* Gynecol Oncol, 2019. **155**(1): p. 63-68.
15. Xiao, Q., T. Chen, and S. Chen, *Fluorescent contrast agents for tumor surgery.* Exp Ther Med, 2018. **16**(3): p. 1577-1585.
16. Li, J., et al., *Spatial heterogeneity of oxygenation and haemodynamics in breast cancer resolved in vivo by conical multispectral optoacoustic mesoscopy.* Light Sci Appl, 2020. **9**: p. 57.
17. James, M.L. and S.S. Gambhir, *A molecular imaging primer: modalities, imaging agents, and applications.* Physiol Rev, 2012. **92**(2): p. 897-965.

18. Smith, A.M., M.C. Mancini, and S. Nie, *Bioimaging: second window for in vivo imaging.* Nat Nanotechnol, 2009. **4**(11): p. 710-1.
19. Rogers, O.C., D.M. Johnson, and E. Firnberg, *mRhubarb: Engineering of monomeric, red-shifted, and brighter variants of iRFP using structure-guided multi-site mutagenesis.* Scientific Reports, 2019. **9**(1): p. 15653.
20. Matlashov, M.E., et al., *A set of monomeric near-infrared fluorescent proteins for multicolor imaging across scales.* Nat Commun, 2020. **11**(1): p. 239.
21. Kaijzel, E.L., G. van der Pluijm, and C.W. Lowik, *Whole-body optical imaging in animal models to assess cancer development and progression.* Clin Cancer Res, 2007. **13**(12): p. 3490-7.
22. Weissleder, R. and M.J. Pittet, *Imaging in the era of molecular oncology.* Nature, 2008. **452**(7187): p. 580-9.
23. Shaner, N.C., et al., *Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein.* Nat Biotechnol, 2004. **22**(12): p. 1567-72.
24. Jenkins, R., M.K. Burdette, and S.H. Foulger, *Mini-review: fluorescence imaging in cancer cells using dye-doped nanoparticles.* RSC Advances, 2016. **6**(70): p. 65459-65474.
25. McCann, T.E., et al., *The use of fluorescent proteins for developing cancer-specific target imaging probes.* Methods in molecular biology (Clifton, N.J.), 2012. **872**: p. 191-204.
26. Shimomura, O., *Structure of the chromophore of Aequorea green fluorescent protein.* FEBS Letters, 1979. **104**(2): p. 220-222.
27. Wang, R., et al., *Cancer-stromal cell fusion as revealed by fluorescence protein tracking.* The Prostate, 2020. **80**(3): p. 274-283.
28. Buczek, M.E., S.P. Reeder, and T. Regad, *Identification and Isolation of Cancer Stem Cells Using NANOG-EGFP Reporter System*, in *Cancer Stem Cells: Methods and Protocols*, G. Papaccio and V. Desiderio, Editors. 2018, Springer New York: New York, NY. p. 139-148.
29. Sonzogni, O., et al., *Reporters to mark and eliminate basal or luminal epithelial cells in culture and in vivo.* PLoS Biol, 2018. **16**(6): p. e2004049.
30. Romano, M., et al., *Past, Present, and Future of Regulatory T Cell Therapy in Transplantation and Autoimmunity.* Frontiers in Immunology, 2019. **10**(43).
31. Lin, W., et al., *Regulatory T cell development in the absence of functional Foxp3.* Nat Immunol, 2007. **8**(4): p. 359-68.
32. Amoozgar, Z., et al., *Targeting Treg cells with GITR activation alleviates resistance to immunotherapy in murine glioblastomas.* Nat Commun, 2021. **12**(1): p. 2582.
33. Haribhai, D., et al., *Regulatory T cells dynamically control the primary immune response to foreign antigen.* J Immunol, 2007. **178**(5): p. 2961-72.
34. Magnuson, A.M., et al., *Identification and validation of a tumor-infiltrating Treg transcriptional signature conserved across species and tumor types.* Proc Natl Acad Sci U S A, 2018. **115**(45): p. E10672-E10681.

35. Rubtsov, Y.P., et al., *Stability of the regulatory T cell lineage in vivo.* Science, 2010. **329**(5999): p. 1667-71.
36. Chalfie, M., *GFP: lighting up life (Nobel Lecture).* Angew Chem Int Ed Engl, 2009. **48**(31): p. 5603-11.
37. Gurskaya, N.G., et al., *Color transitions in coral's fluorescent proteins by site-directed mutagenesis.* BMC Biochem, 2001. **2**: p. 6.
38. Yang, M., P. Jiang, and R.M. Hoffman, *Early Reporting of Apoptosis by Real-time Imaging of Cancer Cells Labeled with Green Fluorescent Protein in the Nucleus and Red Fluorescent Protein in the Cytoplasm.* Anticancer Res, 2015. **35**(5): p. 2539-43.
39. Boone, P.G., et al., *A cancer rainbow mouse for visualizing the functional genomics of oncogenic clonal expansion.* Nat Commun, 2019. **10**(1): p. 5490.
40. Stuker, F., J. Ripoll, and M. Rudin, *Fluorescence molecular tomography: principles and potential for pharmaceutical research.* Pharmaceutics, 2011. **3**(2): p. 229-74.
41. Solomon, M., et al., *Optical imaging in cancer research: basic principles, tumor detection, and therapeutic monitoring.* Med Princ Pract, 2011. **20**(5): p. 397-415.
42. Huang, D., et al., *Optical coherence tomography.* Science, 1991. **254**(5035): p. 1178-81.
43. Rich, L.J. and M. Seshadri, *Photoacoustic monitoring of tumor and normal tissue response to radiation.* Sci Rep, 2016. **6**: p. 21237.
44. Rathbun, C.M., et al., *Rapid Multicomponent Bioluminescence Imaging via Substrate Unmixing.* ACS Chem Biol, 2021. **16**(4): p. 682-690.
45. Xu, M. and L.V. Wang, *Photoacoustic imaging in biomedicine.* Review of Scientific Instruments, 2006. **77**(4): p. 041101.
46. Hosseinaee, Z., J.A. Tummon Simmons, and P.H. Reza, *Dual-Modal Photoacoustic Imaging and Optical Coherence Tomography [Review].* Frontiers in Physics, 2021. **8**(635).
47. Kim, J.H., et al., *Optimizing tissue-clearing conditions based on analysis of the critical factors affecting tissue-clearing procedures.* Sci Rep, 2018. **8**(1): p. 12815.
48. Paley, M.A. and J.A. Prescher, *Bioluminescence: a versatile technique for imaging cellular and molecular features.* Medchemcomm, 2014. **5**(3): p. 255-267.
49. Sellmyer, M.A., et al., *Visualizing cellular interactions with a generalized proximity reporter.* Proc Natl Acad Sci U S A, 2013. **110**(21): p. 8567-72.
50. Evans, M.S., et al., *A synthetic luciferin improves bioluminescence imaging in live mice.* Nat Methods, 2014. **11**(4): p. 393-5.
51. England, C.G., E.B. Ehlerding, and W. Cai, *NanoLuc: A Small Luciferase Is Brightening Up the Field of Bioluminescence.* Bioconjugate chemistry, 2016. **27 5**: p. 1175-1187.
52. Kuchimaru, T., et al., *A luciferin analogue generating near-infrared bioluminescence achieves highly sensitive deep-tissue imaging.* Nat Commun, 2016. **7**: p. 11856.
53. Iwano, S., et al., *Single-cell bioluminescence imaging of deep tissue in freely moving animals.* Science, 2018. **359**(6378): p. 935-939.
54. Zambito, G., et al., *Red-shifted click beetle luciferase mutant expands the multicolor bioluminescent palette for deep tissue imaging.* iScience, 2021. **24**(1): p. 101986.

55. Gaspar, N., et al., *Evaluation of NanoLuc substrates for bioluminescence imaging of transferred cells in mice.* J Photochem Photobiol B, 2021. **216**: p. 112128.
56. Su, Y., et al., *Novel NanoLuc substrates enable bright two-population bioluminescence imaging in animals.* Nat Methods, 2020. **17**(8): p. 852-860.
57. Gammon, S.T., et al., *Rational design of novel red-shifted BRET pairs: Platforms for real-time single-chain protease biosensors.* Biotechnol Prog, 2009. **25**(2): p. 559-69.
58. Chu, J., et al., *A bright cyan-excitable orange fluorescent protein facilitates dual-emission microscopy and enhances bioluminescence imaging in vivo.* Nat Biotechnol, 2016. **34**(7): p. 760-7.
59. Saito, K., et al., *Luminescent proteins for high-speed single-cell and whole-body imaging.* Nat Commun, 2012. **3**: p. 1262.
60. Takai, A., et al., *Expanded palette of Nano-lanterns for real-time multicolor luminescence imaging.* Proc Natl Acad Sci U S A, 2015. **112**(14): p. 4352-6.
61. den Hamer, A., et al., *Bright Bioluminescent BRET Sensor Proteins for Measuring Intracellular Caspase Activity.* ACS Sens, 2017. **2**(6): p. 729-734.
62. Kuchimaru, T., et al., *A novel injectable BRET-based in vivo imaging probe for detecting the activity of hypoxia-inducible factor regulated by the ubiquitin-proteasome system.* Sci Rep, 2016. **6**: p. 34311.
63. Berger, F., et al., *Uptake kinetics and biodistribution of 14C-D-luciferin--a radiolabeled substrate for the firefly luciferase catalyzed bioluminescence reaction: impact on bioluminescence based reporter gene imaging.* European journal of nuclear medicine and molecular imaging, 2008. **35**(12): p. 2275-2285.
64. Badr, C.E., *Bioluminescence imaging: basics and practical limitations.* Methods Mol Biol, 2014. **1098**: p. 1-18.
65. Yan, W., D. Xiao, and K. Yao, *Combined bioluminescence and fluorescence imaging visualizing orthotopic lung adenocarcinoma xenograft in vivo.* Acta Biochim Biophys Sin (Shanghai), 2011. **43**(8): p. 595-600.
66. Schaub, F.X., et al., *Fluorophore-NanoLuc BRET Reporters Enable Sensitive In vivo Optical Imaging and Flow Cytometry for Monitoring Tumorigenesis.* Cancer Research, 2015. **75**(23): p. 5023-5033.
67. Bochner, F., et al., *Bimodal magnetic resonance and optical imaging of extracellular matrix remodelling by orthotopic ovarian tumours.* Br J Cancer, 2020. **123**(2): p. 216-225.
68. Aceto, N., et al., *Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis.* Cell, 2014. **158**(5): p. 1110-1122.
69. Kim, M.Y., et al., *Tumor self-seeding by circulating cancer cells.* Cell, 2009. **139**(7): p. 1315-26.
70. Tello-Lafoz, M., et al., *Cytotoxic lymphocytes use mechanosurveillance to target biophysical vulnerabilities in cancer.* bioRxiv, 2020: p. 2020.04.21.054304.
71. O'Flaherty, J.D., et al., *The cancer stem-cell hypothesis: its emerging role in lung cancer biology and its relevance for future therapy.* J Thorac Oncol, 2012. **7**(12): p. 1880-1890.

72. Scott, J.G., et al., *Recasting the Cancer Stem Cell Hypothesis: Unification Using a Continuum Model of Microenvironmental Forces.* Current Stem Cell Reports, 2019. **5**(1): p. 22-30.
73. Takebe, N. and S.P. Ivy, *Controversies in cancer stem cells: targeting embryonic signaling pathways.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2010. **16**(12): p. 3106-3112.
74. Mirzaei, H., et al., *Therapeutic application of multipotent stem cells.* Journal of Cellular Physiology, 2018. **233**(4): p. 2815-2823.
75. Rompolas, P., et al., *Live imaging of stem cell and progeny behaviour in physiological hair-follicle regeneration.* Nature, 2012. **487**(7408): p. 496-499.
76. Choi, S.A., et al., *Histone deacetylase inhibitor panobinostat potentiates the anti-cancer effects of mesenchymal stem cell-based sTRAIL gene therapy against malignant glioma.* Cancer Lett, 2019. **442**: p. 161-169.
77. Dong, J., et al., *Bisacodyl and its cytotoxic activity on human glioblastoma stem-like cells. Implication of inositol 1,4,5-triphosphate receptor dependent calcium signaling.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2017. **1864**(6): p. 1018-1027.
78. Chen, J.H., et al., *Antrocin, a bioactive component from Antrodia cinnamomea, suppresses breast carcinogenesis and stemness via downregulation of beta-catenin/Notch1/Akt signaling.* Phytomedicine, 2019. **52**: p. 70-78.
79. Aguirre-Ghiso, J.A., L. Ossowski, and S.K. Rosenbaum, *Green fluorescent protein tagging of extracellular signal-regulated kinase and p38 pathways reveals novel dynamics of pathway activation during primary and metastatic growth.* Cancer Res, 2004. **64**(20): p. 7336-45.
80. Edinger, M., et al., *Noninvasive assessment of tumor cell proliferation in animal models.* Neoplasia, 1999. **1**(4): p. 303-10.
81. Scheffold, C., et al., *Visualization of effective tumor targeting by CD8+ natural killer T cells redirected with bispecific antibody F(ab')(2)HER2xCD3.* Cancer Res, 2002. **62**(20): p. 5785-91.
82. Scabini, M., et al., *In vivo imaging of early stage apoptosis by measuring real-time caspase-3/7 activation.* Apoptosis, 2011. **16**(2): p. 198-207.
83. Jenkins, D.E., et al., *Bioluminescent human breast cancer cell lines that permit rapid and sensitive in vivo detection of mammary tumors and multiple metastases in immune deficient mice.* Breast Cancer Res, 2005. **7**(4): p. R444-54.
84. Minn, A.J., et al., *Genes that mediate breast cancer metastasis to lung.* Nature, 2005. **436**(7050): p. 518-24.
85. Minn, A.J., et al., *Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors.* J Clin Invest, 2005. **115**(1): p. 44-55.

86. Minegishi, M., et al., *Multiplexed bioluminescence imaging of cancer cell response to hypoxia and inflammation in the caudal-artery injection model of bone metastasis during zoledronic acid treatment.* Journal of Cancer Metastasis and Treatment, 2021. **7**: p. 5.
87. Zhang, X.H., et al., *Selection of bone metastasis seeds by mesenchymal signals in the primary tumor stroma.* Cell, 2013. **154**(5): p. 1060-1073.
88. Boire, A., L. DeAngelis, and J. Massagué, *Development of a Mouse Model of Leptomeningeal Metastasis (P7.014).* Neurology, 2014. **82**(10 Supplement): p. P7.014.
89. Kienast, Y., et al., *Real-time imaging reveals the single steps of brain metastasis formation.* Nat Med, 2010. **16**(1): p. 116-22.
90. Morris, V.L., et al., *Mammary carcinoma cell lines of high and low metastatic potential differ not in extravasation but in subsequent migration and growth.* Clinical & Experimental Metastasis, 1994. **12**(6): p. 357-367.
91. Cameron, M.D., et al., *Temporal progression of metastasis in lung: cell survival, dormancy, and location dependence of metastatic inefficiency.* Cancer Res, 2000. **60**(9): p. 2541-6.
92. Malladi, S., et al., *Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT.* Cell, 2016. **165**(1): p. 45-60.
93. Calon, A., et al., *Dependency of colorectal cancer on a TGF-beta-driven program in stromal cells for metastasis initiation.* Cancer Cell, 2012. **22**(5): p. 571-84.
94. Waldman, A.D., J.M. Fritz, and M.J. Lenardo, *A guide to cancer immunotherapy: from T cell basic science to clinical practice.* Nat Rev Immunol, 2020. **20**(11): p. 651-668.
95. Zhang, Z., et al., *Gene delivery of TIPE2 inhibits breast cancer development and metastasis via CD8(+) T and NK cell-mediated antitumor responses.* Mol Immunol, 2017. **85**: p. 230-237.
96. Lee, S.S., V.P. Bindokas, and S.J. Kron, *Multiplex three-dimensional optical mapping of tumor immune microenvironment.* Sci Rep, 2017. **7**(1): p. 17031.
97. Badr, C.E. and B.A. Tannous, *Bioluminescence imaging: progress and applications.* Trends Biotechnol, 2011. **29**(12): p. 624-33.
98. Png, K.J., et al., *MicroRNA-335 inhibits tumor reinitiation and is silenced through genetic and epigenetic mechanisms in human breast cancer.* Genes & development, 2011. **25**(3): p. 226-231.
99. Ofoeyeno, N., E. Ekpenyong, and C. Braconi, *Pathogenetic Role and Clinical Implications of Regulatory RNAs in Biliary Tract Cancer.* Cancers, 2021. **13**(1): p. 12.
100. An, J., et al., *Heptamethine carbocyanine DZ-1 dye for near-infrared fluorescence imaging of hepatocellular carcinoma.* Oncotarget, 2017. **8**(34): p. 56880-56892.
101. Bhang, H.-e.C., et al., *Tumor-specific imaging through progression elevated gene-3 promoter-driven gene expression.* Nature Medicine, 2011. **17**(1): p. 123-129.
102. Condeelis, J. and J.E. Segall, *Intravital imaging of cell movement in tumours.* Nature Reviews Cancer, 2003. **3**(12): p. 921-930.

103. Katsuno, Y., et al., *Bone morphogenetic protein signaling enhances invasion and bone metastasis of breast cancer cells through Smad pathway.* Oncogene, 2008. **27**(49): p. 6322-6333.
104. Serganova, I., et al., *Multimodality imaging of TGFbeta signaling in breast cancer metastases.* FASEB J, 2009. **23**(8): p. 2662-72.
105. Serganova, I., et al., *Multimodality imaging of TGFβ signaling in breast cancer metastases.* The FASEB Journal, 2009. **23**(8): p. 2662-2672.
106. Kang, Y., et al., *Breast cancer bone metastasis mediated by the Smad tumor suppressor pathway.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(39): p. 13909-13914.
107. Winslow, M.M., et al., *Suppression of lung adenocarcinoma progression by Nkx2-1.* Nature, 2011. **473**(7345): p. 101-104.
108. Luker, K.E. and D. Piwnica-Worms, *Optimizing luciferase protein fragment complementation for bioluminescent imaging of protein-protein interactions in live cells and animals.* Methods Enzymol, 2004. **385**: p. 349-60.
109. Gupta, G.P., et al., *ID genes mediate tumor reinitiation during breast cancer lung metastasis.* Proc Natl Acad Sci U S A, 2007. **104**(49): p. 19506-11.
110. Wang, L.V. and J. Yao, *A practical guide to photoacoustic tomography in the life sciences.* Nat Methods, 2016. **13**(8): p. 627-38.
111. Knox, H.J. and J. Chan, *Acoustogenic Probes: A New Frontier in Photoacoustic Imaging.* Acc Chem Res, 2018. **51**(11): p. 2897-2905.
112. Lucero, M.Y., et al., *Development of NIR-II Photoacoustic Probes Tailored for Deep-Tissue Sensing of Nitric Oxide.* J Am Chem Soc, 2021. **143**(18): p. 7196-7202.

# CHAPTER 2: Rapid multicomponent bioluminescence imaging via substrate unmixing [*]

## I. Introduction

Bioluminescence imaging (BLI) is a popular technique for tracing cells and other biological features in heterogeneous environments, including whole animals.[1-2] BLI applications rely on genetically encoded luciferase enzymes and luciferin substrates for photon production. Because mammalian tissues do not normally glow, BLI enables sensitive imaging *in vivo*.[3-5] As few as 1-10 cells can be reliably detected using optimized probes in subcutaneous models.[6] For these reasons, BLI has long been a go-to technique for monitoring physiological processes in rodents.[7-9] More recent advances are further enabling studies in deeper tissues and larger organisms,[10-13] including non-human primates.[10,14]

While ubiquitous, BLI applications *in vivo* typically track only one cell type or feature at a time.[15] A spectrum of bioluminescent probes exists, and some pairs can be color-resolved in rodent models.[16-22] Discriminating larger collections is challenging, though, due to the broad emission spectra of bioluminescent reporters and complications from tissue absorption.[8] Historically, many applications featuring two BLI reporters combined firefly luciferase (Fluc) and *Renilla* luciferase (Rluc), two enzymes that use different substrates.[23-24] Recent examples have featured Akaluc and

[*] *Author's Note: The following chapter is a completed manuscript that has been accepted for publication in ACS Chemical Biology. I share co-first authorship with Colin M. Rathbun, and Zi Yao and we have contributed equally to this publication. I contributed to designing experiments and procedures, data analysis and interpretation, and writing. The co-authors include Krysten A. Jones, William B. Porterfield, and last author Jennifer A. Prescher.*

Antares or CBG2, additional enzymes that use distinct luciferins.[16,25] Multicomponent imaging with these pairs *in vivo*, though, can require multiple hours, if not days. The bioluminescent substrates are typically administered at saturating doses to maximize photon output,[26] and the first probe must clear prior to administering the second. In principle, dozens of other naturally occurring (and orthogonal) luciferases and luciferins could be employed for multiplexed BLI.[27-29] In practice, though, most of these enzymes and substrates are not easily applied *in vivo* owing to suboptimal stabilities, bioavailability, or other parameters.

We aimed to address the need for better probes and practical imaging protocols for multicomponent BLI. Our approach builds on the expanding toolbox of orthogonal bioluminescent reagents.[8,15] These probes comprise genetically modified luciferases that are responsive to chemically unique luciferins.[12,14,25,30-34] Some of the orthogonal enzymes can be readily discriminated in cells and mouse models based on selective substrate use.[17,25,30-32,35-37] Many of these probes are not completely specific (i.e., the luciferins are processed preferentially by one enzyme, but can be turned over to some extent by other enzymes). However, perfect selectivity is not required. The patterns of substrate use ("fingerprints") can be used to discriminate combinations of luciferases. Indeed, such "fingerprint" analyses are commonly used to differentiate engineered GPCRs and signaling pathways.[38-40] Like the examples above, though, traditional applications of the engineered luciferases require long periods of time (hours-to-days) for complete image acquisition.[25,31] Dynamic changes in gene expression and cell growth cannot be captured under such conditions. Consequently, bioluminescence has historically been limited to monitoring bulk changes in biological samples. Here we report a method to more rapidly resolve orthogonal bioluminescent probes. The approach is broadly applicable and scalable to multiple

reporters, enabling facile multiplexed bioluminescence imaging. We also showcase its utility for monitoring dynamic changes on the minutes time scale.

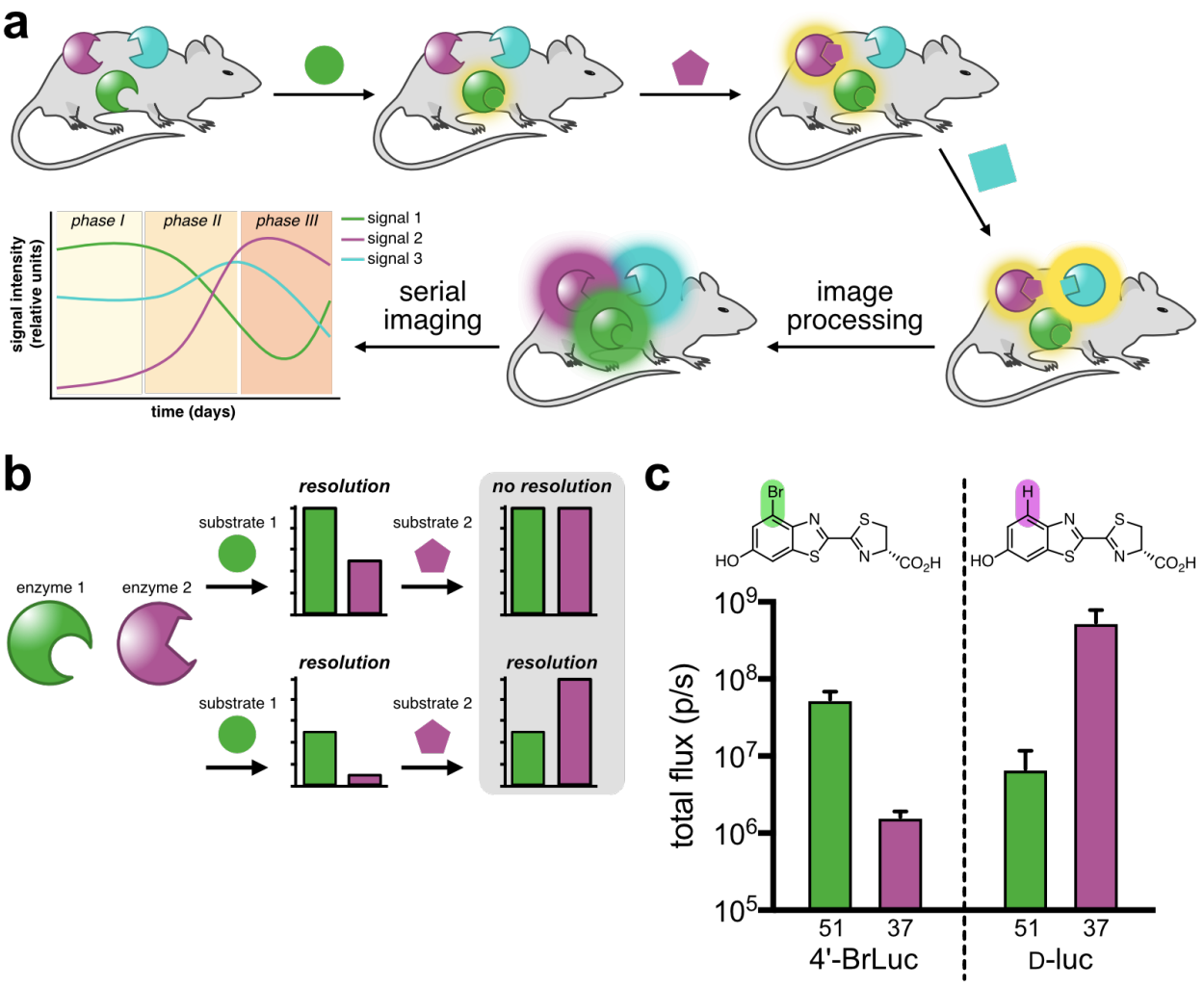


**Figure 2.1 Multicomponent bioluminescence imaging via serial substrate addition and unmixing.** (a) Sequential application of orthogonal luciferins (shapes) to illuminate multiple luciferase reporters *in vivo*. Linear unmixing algorithms can deconvolute substrate signatures, enabling rapid and dynamic readouts of biological processes. (b) Bioluminescent probes must be orthogonal and exhibit differential emission intensities for successful unmixing. When both probes are equally "bright" (top), no resolution is possible. Probes of varying intensity (bottom) can be distinguished when the dimmest probe is administered first. (c) A sample orthogonal bioluminescent pair. Samples expressing mutant 51 or 37 can be resolved using 4'-BrLuc (100 μM) or D-luc (100 μM). Error bars represent the standard error of the mean for $n$ = 3 experiments.

## II. Results and Discussion

### Identifying Optimal Luciferases

We hypothesized that rapid BLI could be achieved via sequential substrate administration and serial image acquisition. Light outputs would build over time, with each luciferin application resulting in stronger signal. A final processing step[41] would unmix the images (Figure 2.1a). Similar "layering" approaches have vastly expanded fluorescence detection of gene transcripts[42] and other cellular features[21,43] in recent years. The technique also differs from conventional optical imaging platforms, in that wavelength, lifetime, and other traditional parameters are not factored into the analysis. The probes must simply be substrate resolved (i.e., exhibit some degree of orthogonality) *and* intensity resolved (Figure 2.1b). Substrate resolution minimizes cross talk between the probes to provide unique fingerprints. Intensity resolution ensures that signal can be "layered in": as successively brighter probes are imaged, residual signal from dimmer, earlier images becomes part of the background. If the bioluminescent probes are not intensity resolved, the targets are indistinguishable. The key takeaway from the design is that no time is required for substrate clearance, which is an important consideration for capturing rapid, dynamic events.

**a**

| Enzyme | 218 | 240 | 241 | 243 | 247 | 249 | 250 | 314 | 316 | 347 |
|---|---|---|---|---|---|---|---|---|---|---|
| WT | *R* | *V* | *V* | *F* | *F* | *M* | *F* | *S* | *G* | *S* |
| 50 | R | **I** | V | F | **Y** | M | F | S | G | **G** |
| 51 (Pecan) | R | V | V | **M** | F | M | F | S | G | **G** |
| 53 | R | **I** | V | **M** | **Y** | M | F | S | G | **G** |
| 54 | R | V | **A** | F | **L** | M | F | S | G | **A** |
| 37 (Cashew) | **K** | V | V | F | F | M | F | S | G | S |
| 85 | **K** | V | V | F | F | M | **M** | **T** | **T** | S |
| 86 | **K** | V | V | F | F | **L** | F | **C** | G | S |
| 87 | **K** | V | V | F | F | M | **Y** | **T** | **T** | S |
| 93 | **K** | V | V | F | F | **L** | F | **V** | G | S |

**b**

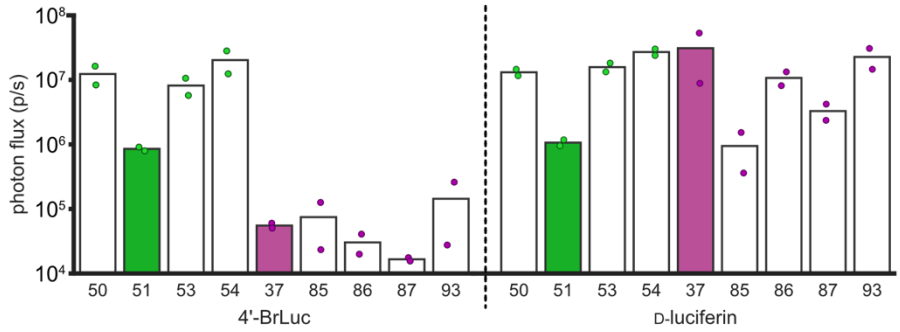


**Figure 2.2 Identifying intensity resolved orthogonal pairs**. (a) Mutants considered for rapid BLI with 4'-BrLuc and D-luc. Mutants 50, 51 (Pecan), 53, and 54 prefer 4'-BrLuc, while 37 (Cashew), 85, 86, 87, and 93 prefer D-luc.[11] (b) Verifying orthogonality and substrate resolution *in vivo*. DB7 cells expressing different mutant luciferases were implanted into the backs of mice. Sequential application of 4'-BrLuc and D-luc enabled identification of optimal mutant luciferase combinations for multicomponent imaging. Photon flux values from images were quantified and plotted. Mutant luciferases 51 and 37 were identified as optimal candidates for rapid multicomponent BLI.

To identify suitable probes for rapid BLI, we focused on two previously reported orthogonal luciferins: 4'-BrLuc and D-luciferin (D-luc, Figure 2.1c).[31,44] These substrates are bright, bioavailable, and accessible in large quantities.[45-46] We also previously identified mutant luciferases that could differentiate the analogs.[31,47] While orthogonal, these pairs were engineered for maximum brightness and not built with intensity resolution in mind. To identify orthogonal luciferases that exhibited a *range* of photon outputs, we screened a small panel of mutants known to process C4'-modified analogs (Figure 2.2a). Screens were performed both *in vitro* and *in vivo*

(Figure 2.3, Figure 2.2b). In the latter case, DB7 cells stably expressing mutant luciferases were implanted in mice. 4'-BrLuc and D-luc were administered sequentially. The most orthogonal and intensity resolved pair comprised mutants 51 and 37 (Figure 2.2b), which we subsequently named Pecan and Cashew, respectively. Bioluminescent signal from 4'-BrLuc/Pecan is lower than that of D-luc/Cashew, meaning that the two pairs are intensity resolved and amenable to rapid sequential imaging (Figure 2.1c, Figure 2.3). 4'-BrLuc can be administered first (to illuminate Pecan), followed immediately by D-luc (to illuminate Cashew). Since Cashew signal is brighter than Pecan, the images can be readily unmixed. It is also important to note that while 4'-BrLuc/Pecan signal is reduced relative to D-luc/Fluc, it is still sufficiently intense for applications *in vivo*.

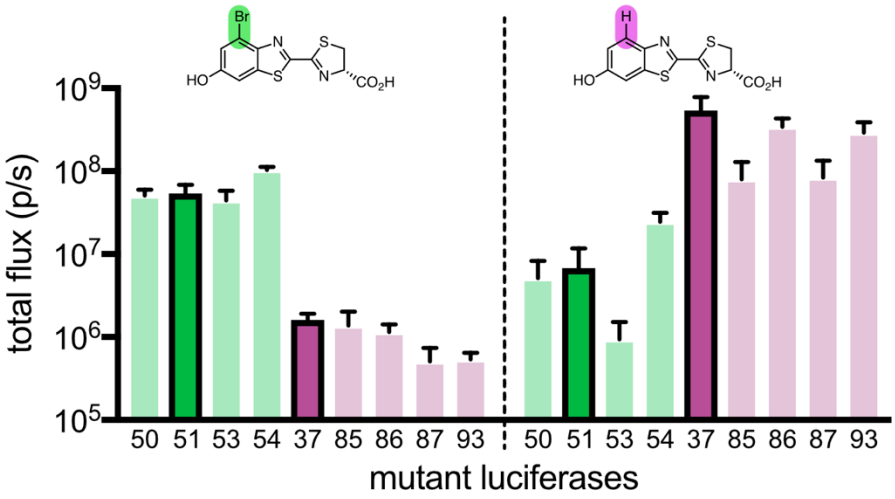


**Figure 2.3. Identifying intensity-resolved orthogonal mutant luciferase-luciferin pairs.** Candidate luciferases were expressed in bacteria and screened with 100 μM 4'-BrLuc or D-luc. Mutants exhibited orthogonal substrate use, with >10-fold substrate preference observed in most cases. The "winning combination" – mutants 51 and 37 – were also intensity resolved. Error bars represent the standard error of the mean for $n = 3$ experiments.

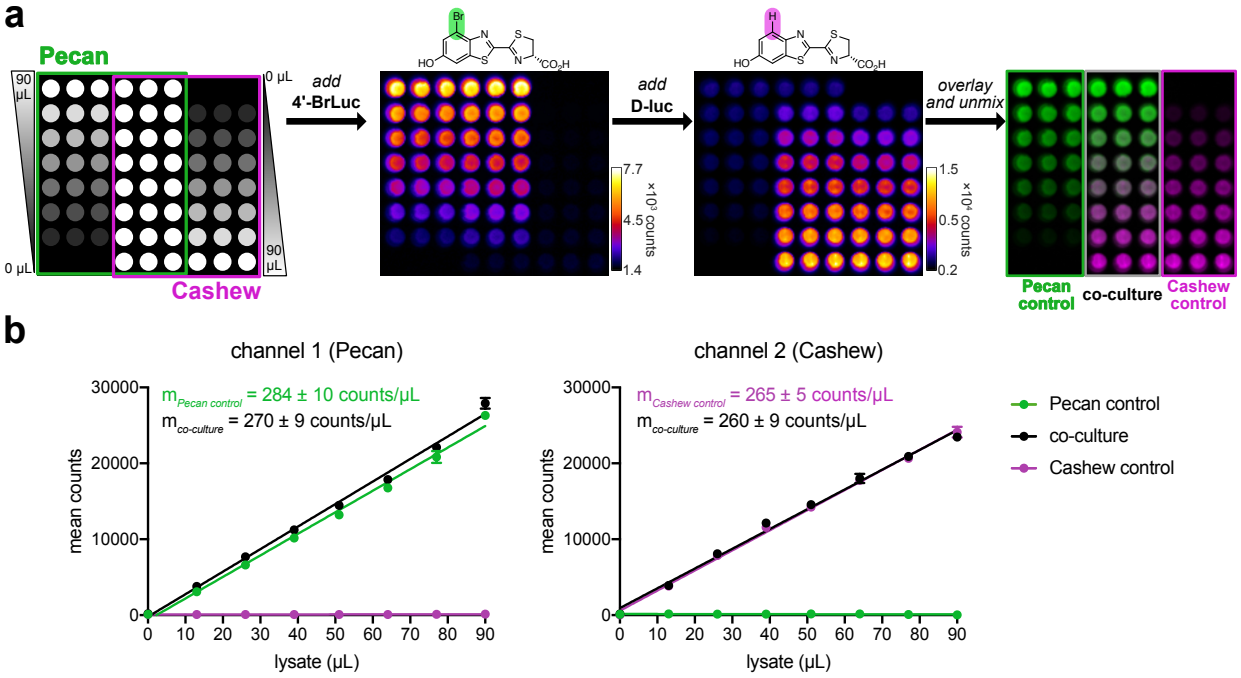


**Figure 2.4 Rapid BLI *in vitro*.** (a) Pecan and Cashew were plated in a gradient fashion (as shown). The samples were treated with 4'-BrLuc (100 µM), followed by D-luc (100 µM). Raw images were acquired after each substrate addition. The substrate-specific signals were unmixed, assigned false colors and overlaid. (b) Quantification of the images from (a), fit via linear regression. In channel 1, $R^2$ values for the Pecan control and co-culture wells are 0.993 and 0.994, respectively. In channel 2, $R^2$ values for Cashew control and co-culture are 0.998 and 0.993, respectively. Error bars represent the standard error of the mean for $n$ = 3 experiments.

To evaluate the reporters for rapid multicomponent BLI, we performed a series of *in vitro* experiments. Pecan and Cashew-expressing cells were lysed and distributed across black-well plates. 4'-BrLuc was initially administered to each well, and an image of the plate was acquired (Figure 2.4a). D-Luc was then immediately added to the same wells, and a second image was acquired. Because light output from Cashew is brighter, the Pecan signal fell entirely within the noise of the second image. False colors were assigned to each reporter. The images were then overlaid and a linear unmixing algorithm was employed to determine the relative quantities of each

mutant. The measured signal correlated linearly with probe concentration (Figure 2.4b). Signal outputs from mixed lysate samples were also co-linear with samples comprising a single luciferase, indicating minimal signal crosstalk. Rapid imaging required the dimmest reporter to be visualized first (Figure 2.5). The unmixing algorithm is not necessary when the second target is more abundant than the first.[42] When the second target (associated with the "brightest" luciferase) is in low abundance, though, the algorithm is crucial for proper image interpretation. Unmixing ensures that residual signal from the first image is eliminated and doesn't interfere with the subsequent image (Figure 2.6). Since the relative abundance of multiple targets is unknown in a given experiment, the algorithm would be employed in all cases by the end user.

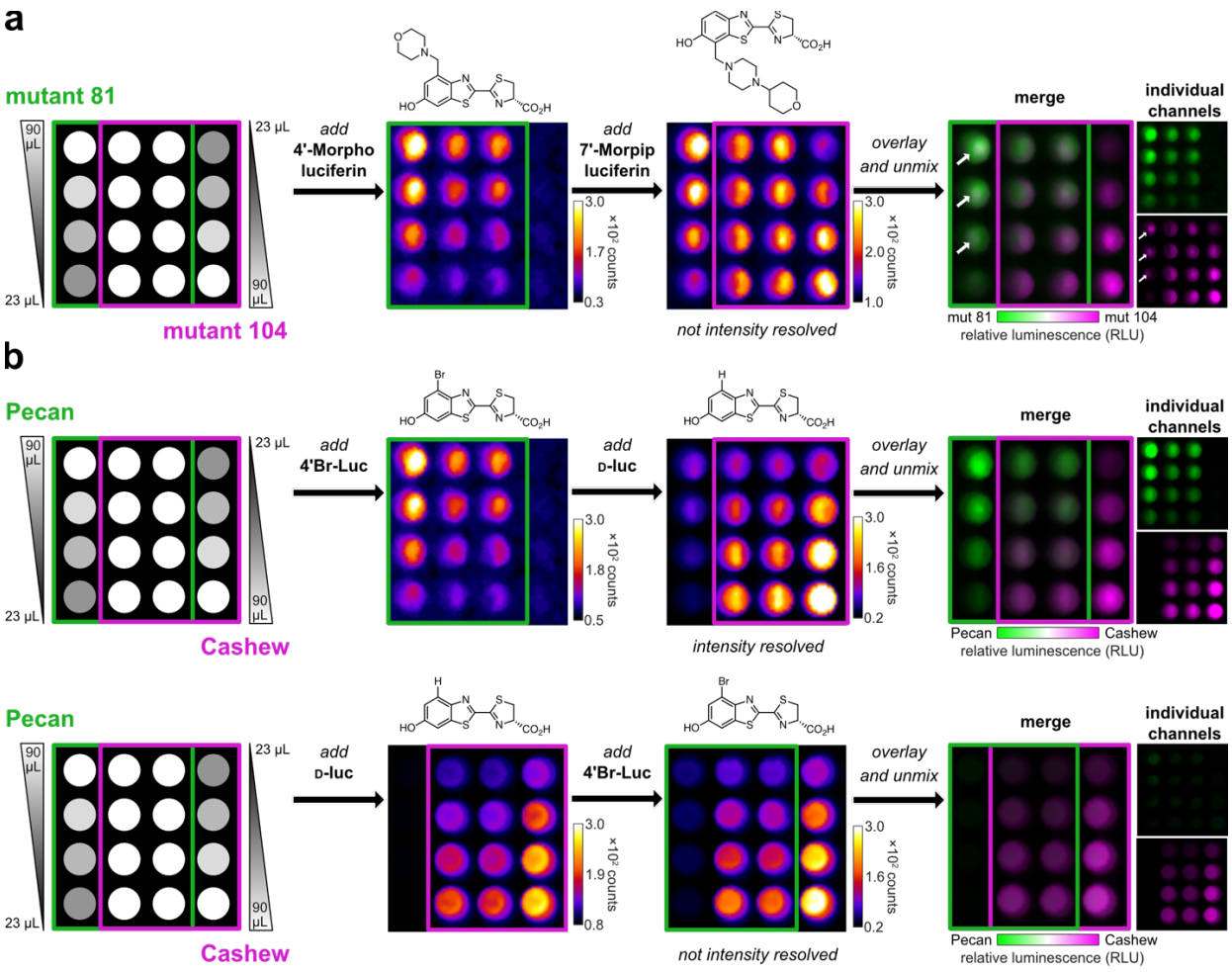


**Figure 2.5. Substrate unmixing requires probes that are intensity resolved**. (a) An orthogonal pair comprising mutant 81/4'-Morpho luciferin and mutant 104/7'-Morpip luciferin is not intensity resolved,[1-2] and thus not amenable to rapid BLI. Gradients of the mutants (expressed in bacterial lysate) were plated in a 4x4 matrix. 4'-Morpho luciferin (the preferred substrate for mutant 81) was then administered, followed by 7'-Morpip luciferin (the preferred substrate for mutant 104). Final luciferin concentrations were 100 µM. Substrate unmixing was not successful. Strong residual signal (from 4'-Mopho luciferin) in the 7'-Morpip luciferin image can be observed. The presence of white pixels in the merged image (arrows) is not consistent with the composition of the well (since only one luciferase was present). (b) Gradients of Cashew and Pecan were plated in a 4x4 square. When the dimmer analog (4'-BrLuc) was added prior to the brighter one (D-luc), the signals can be readily unmixed (top). When D-luc was added first, though, the signals cannot be distinguished (bottom). Final luciferin concentrations were 100 µM in each case.

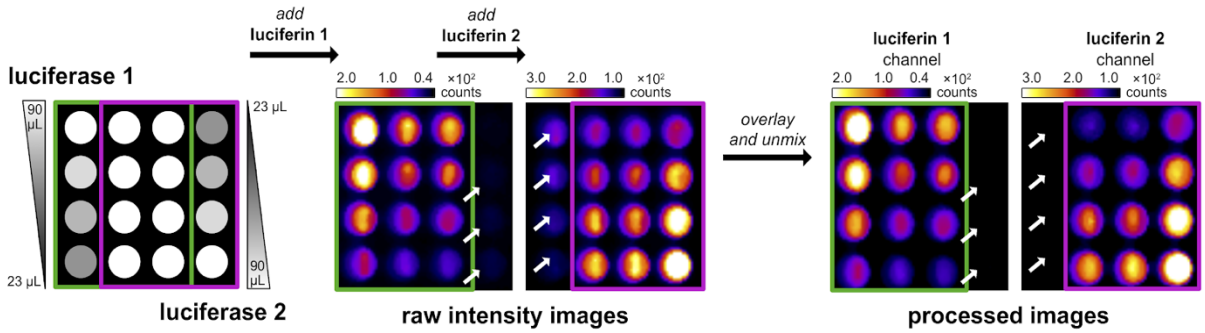


**Figure 2.6. Residual signals are removed by substrate unmixing.** Gradients of mutant luciferases in bacterial lysate were plated in a 4x4 matrix. The corresponding luciferins (100 µM) were added sequentially. Pixels containing residual signal are highlighted by the white arrows. These signals are removed upon imaging processing.

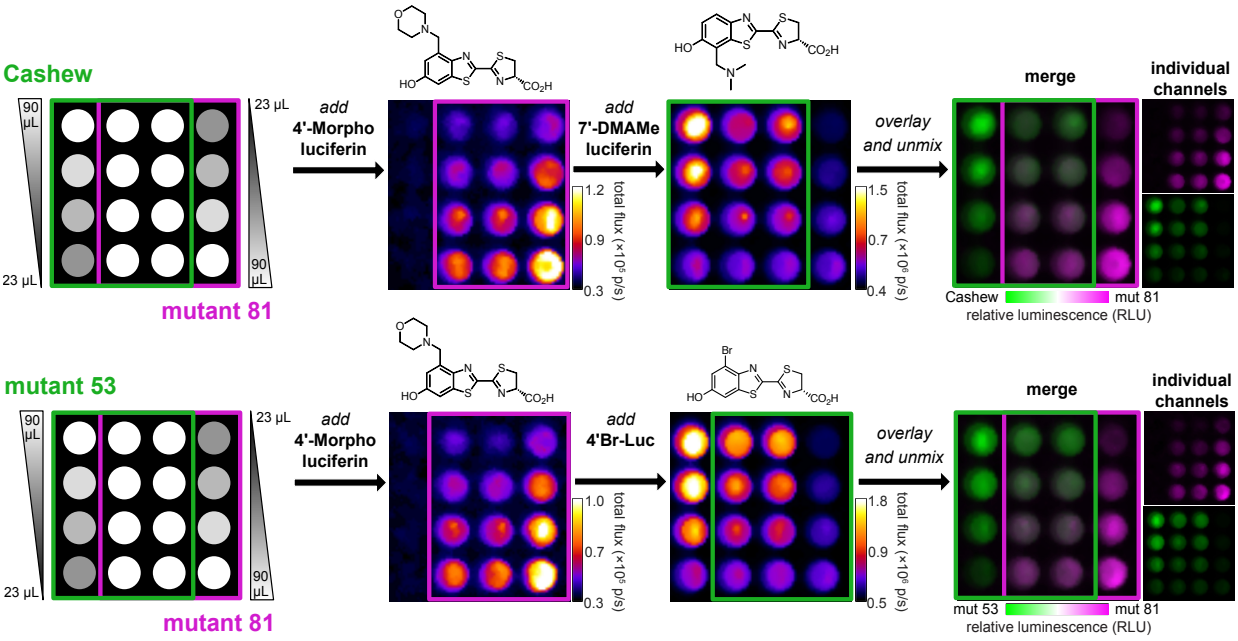


**Figure 2.7. Multiple orthogonal engineered pairs can be rapidly unmixed.** Luciferases examined include mutant 81 (and its corresponding substrate 4'-MorphoLuc), Cashew (and its corresponding substrate 7'-DMAMeLuc), and mutant 53 (and its corresponding substrate 4'-BrLuc).[1-2] Gradients of engineered luciferases were plated as shown. The corresponding substrates (100 µM) were administered, beginning with the dimmest luciferin. Images were acquired after each addition. The raw data were stacked and unmixed.

These results underscore the notion that perfect substrate selectivity is not required for differentiating multiple probes, broadening the potential impact of the approach. Many dual-component imaging experiments rely on completely different classes of enzymes and substrates that exhibit different bioavailability and administration routes.[23,25,36] Substrates that are more structurally similar can be easier to work with, but typically exhibit imperfect orthogonality.[31,32,45,48-49] The substrate unmixing algorithm takes advantage of these imperfections (in the form of diagnostic fingerprints) and should be able to interface with the dozens of luciferin analogs reported to date. Indeed, the "layering in" approach successfully resolved combinations of D-luciferin analogs and their associated enzymes via rapid luciferin administration (Figure 2.7). Mixtures of Fluc and other commercially available reporters (e.g., NanoLuc and *Gaussia* luciferase) were readily unmixed following sequential addition of the corresponding substrates (Figure 2.8). Moreover, combinations of mammalian cells could be visualized in a single imaging session (Figure 2.9). These examples suggest that the unmixing approach is generalizable to multiple luciferase reporters that use unique substrates and exhibit a range of intensities. Additionally, the unmixing approach is applicable to not only well-differentiated bioluminescent tools (e.g., orthogonal insect and marine luciferins), but also structurally similar probes (e.g., D-luciferin analogs). This feature greatly expands the number of luciferase-luciferin pairs that can be imaged in a single experiment, as perfectly orthogonal pairs are not required.

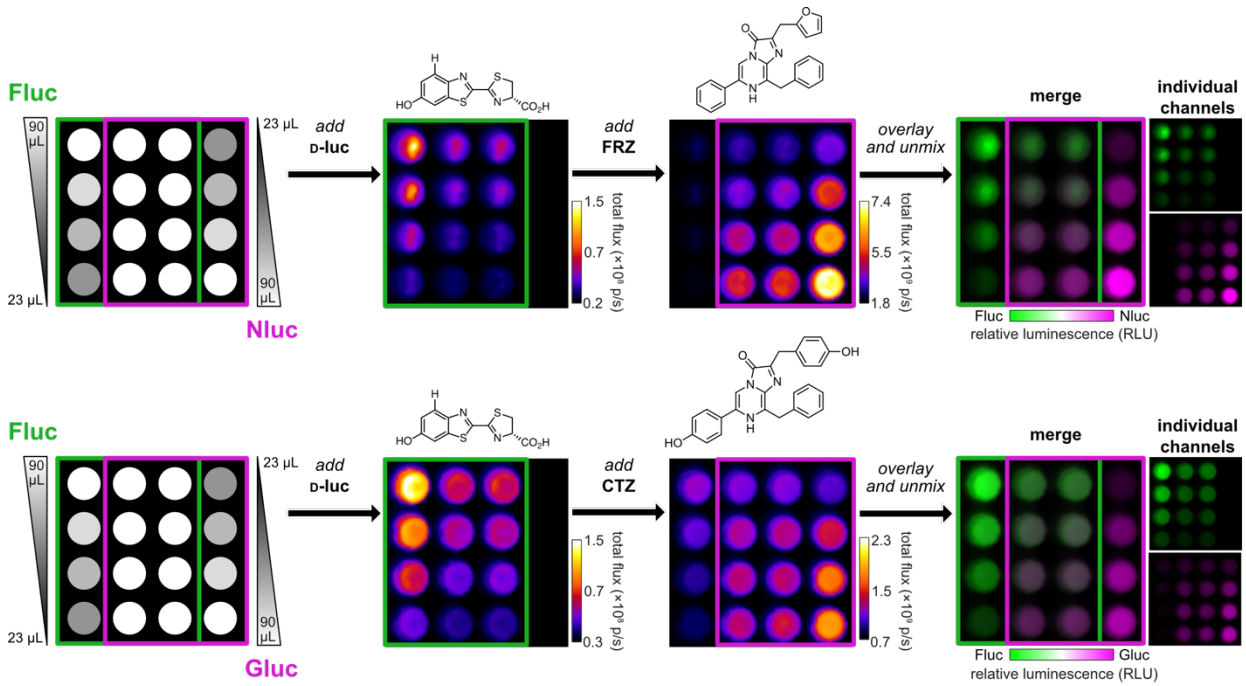


**Figure 2.8 Multiple orthogonal pairs can be rapidly unmixed.** Established reporters examined include firefly luciferase (Fluc), NanoLuc (Nluc), and *Gaussia* luciferase (Gluc). The corresponding luciferins for each reporter are shown**.** Gradients of the luciferases were plated as shown. The corresponding substrates ([D-luc] = 100 uM, [FRZ] = 1:40 of commercial stock, [CTZ] = 100 µM) were administered, beginning with the dimmest luciferin. Images were acquired after each addition. The raw data were stacked, unmixed, and false colored.

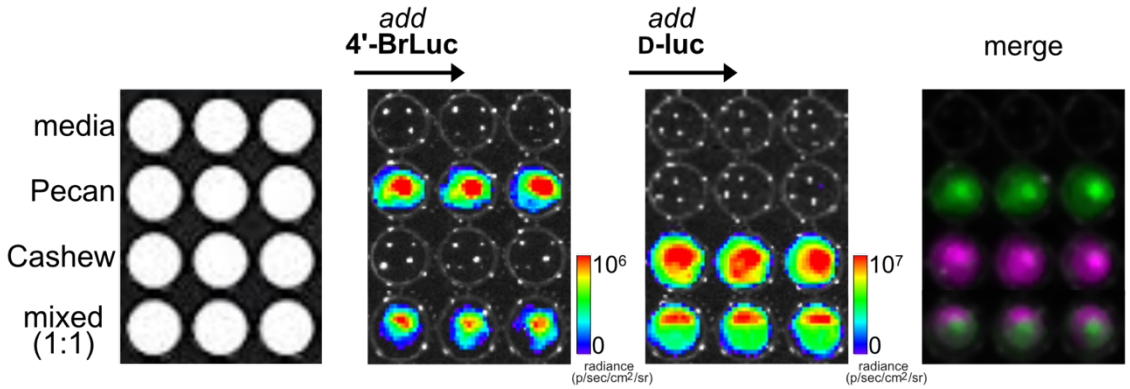


**Figure 2.9. Sequential substrate administration enables multicomponent bioluminescence imaging *in cellulo.*** DB7 cells expressing Cashew, Pecan, or no luciferase (media) were plated (1 x $10^5$ cells/well). Some wells contained a 1:1 mixture Cashew- or Pecan-expressing cells (5 x $10^4$ of each cell type per well). All samples were first treated with 4'-BrLuc, followed by D-luc. Images were acquired after each substrate addition. Raw photon values are shown, along with the merged image following substrate unmixing.

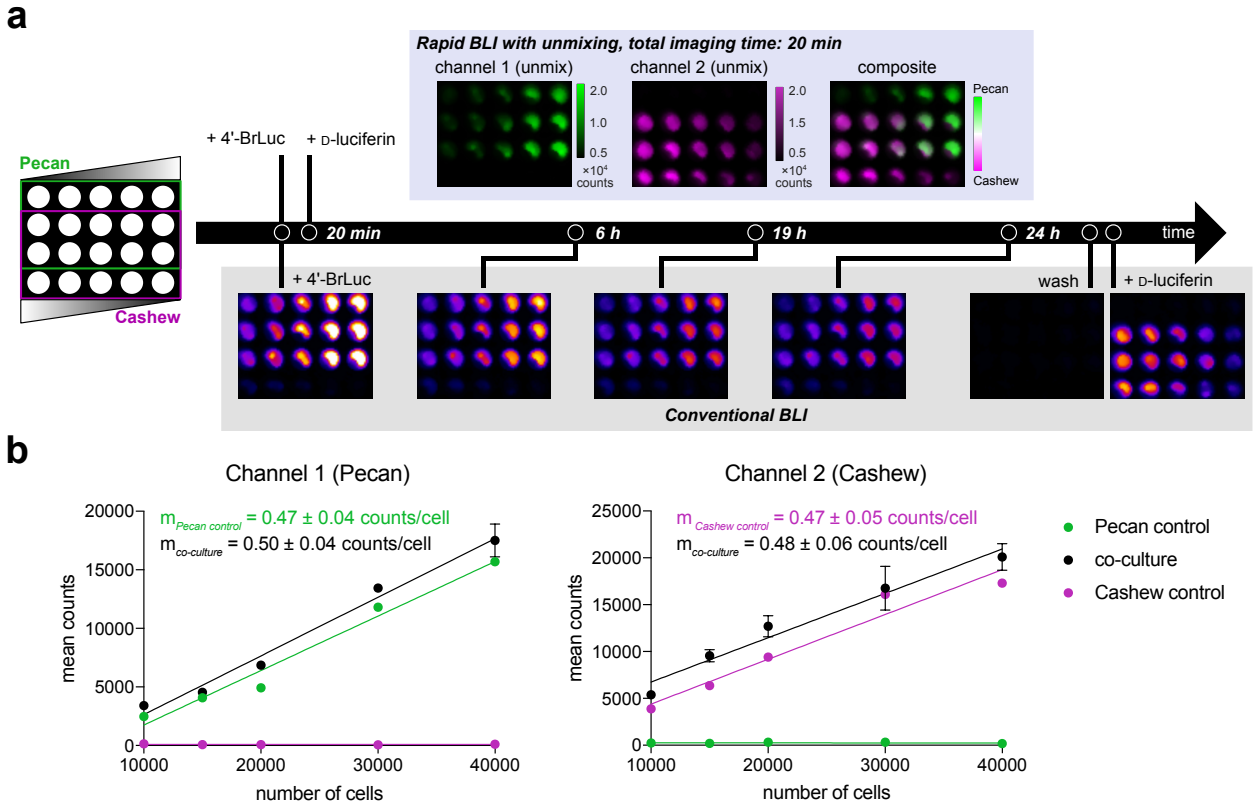


**Figure 2.10 Rapid two-component BLI via substrate unmixing.** (a) Cells expressing Pecan or Cashew, were plated in a 96-well plate as shown. Sequential substrate administration (4'-BrLuc, followed by D-luc, top row) and unmixing enabled two-component imaging in only 20 min. Conventional bioluminescence imaging required >24 h to complete (bottom row). Data are representative of $n = 3$ replicates. (b) Quantified photon outputs for the images in (a), fit via linear regression. In channel 1, $R^2$ values for the Pecan control and co-culture wells are 0.98 and 0.99, respectively. In channel 2, $R^2$ values for the Cashew control and co-culture wells are 0.95 and 0.97, respectively. Error bars represent the standard error of the mean for $n = 3$ experiments.

To showcase the utility of rapid unmixing, we performed a head-to-head comparison with conventional bioluminescence imaging. The latter approach entails waiting for signal to clear from one substrate, prior to administering the second. Since saturating doses of luciferins are typically used, the clearance period can be several hours to days. Our approach via sequential substrate application and signal unmixing can dramatically shorten this time frame, enabling improved temporal resolution. As an initial demonstration, mixtures of Pecan- and Cashew-expressing mammalian cells were cultured in varying ratios (Figure 2.10a). Conventional bioluminescence

imaging was performed by adding one luciferin substrate (4'-BrLuc) to illuminate the Pecan-expressing cells. Images were acquired until signal returned to background levels (24 h, bottom row of Figure 2.10a). The second luciferin (D-luc) was then applied to the cultures to capture Cashew-dependent signal. In total, the two-component imaging study was complete in just over 24 h. By contrast, sequential substrate addition followed by unmixing enabled two-component imaging in only 20 minutes (top row, Figure 2.10a).

We surmised that the unmixing algorithm would also enable rapid imaging *in vivo*. As an initial demonstration, Pecan- and Cashew-expressing cells were mixed in varying ratios and implanted in mice ($10^6$ cells per site, Figure 2.11a). Upon injection of 4'-BrLuc, Pecan-expressing cells were readily visualized (Figure 2.11b). Prior to substrate clearance, the brighter luciferin (D-luc) was injected. Signal was then observed from Cashew-expressing cells. Substrate unmixing revealed the expected distributions of Cashew- and Pecan-expressing cells (Figure 2.11b and Figure 2.12). Notably, the two-component imaging session was complete in ~1 h, a significant improvement from the 6-24 h imaging window common to other substrate-resolved probes.[31] The bioavailabilities of the luciferins featured in this experiment are similar, although this is not a requirement for unmixing. While the relative composition of the cell masses can be readily visualized, the unmixing algorithm cannot provide absolute quantification of bioluminescent signals in animals. Rather, the *relative* amounts of signal are easily discerned and tracked. Samples with a large amount of Pecan will appear "more green than red" in the unmixed images. As the reporter ratios change over time, so too will the unmixed images. These measures are often paramount in optical imaging *in vivo*, and imaging relative differences is standard practice in most bioluminescent applications.[50-52] Similar intensiometric measures are routinely used in imaging with fluorescent sensors.[53-55]

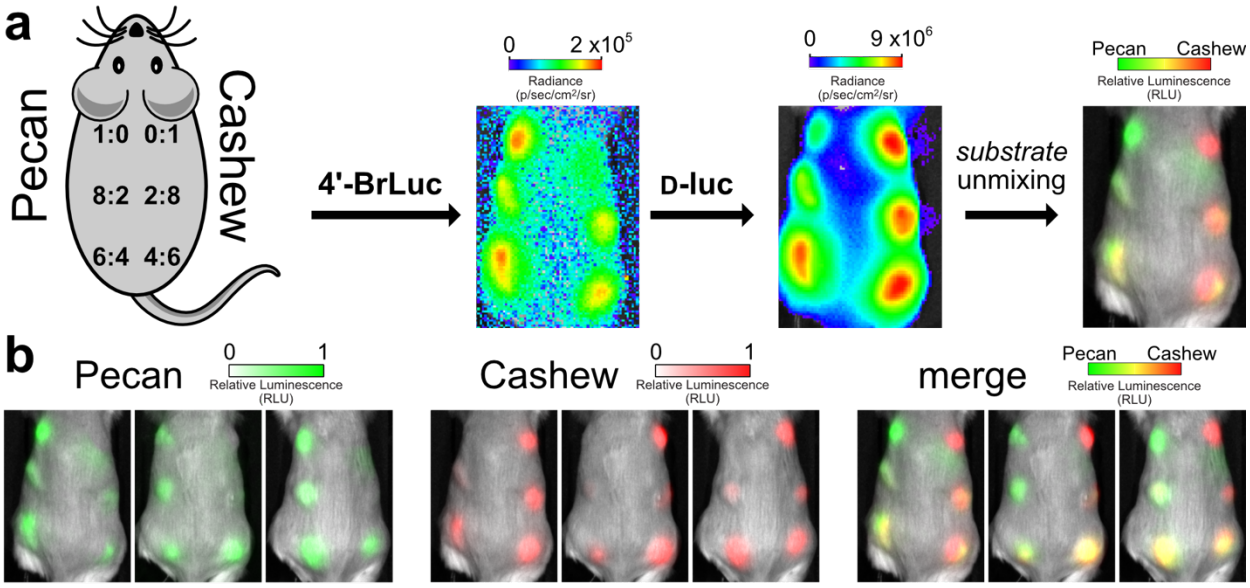


**Figure 2.11 Rapid BLI *in vivo*.** (a) Ratios of Cashew-and Pecan-expressing cells implanted in mice. Orthogonal substrates (65 mM) were administered sequentially via i.p. injection (100 µL). Images were acquired 35 min after each injection. (b) Unmixed channels for each mouse replicate are shown. Color bars indicate normalized luminescence values.

Quantifying the number of cells (or other features) from bioluminescent images is possible, but requires multiple standard curves and external validation. The relevant standard curves are readily generated for *in vitro* and *in cellulo* experiments (e.g., see Figures 2.4 and 2.10, where bioluminescent outputs correlated with cell number and/or enzyme amount over a broad range.) Such curves are rarely generated for *in vivo* experiments, though, owing to the complexities involved in standardizing optical signals within heterogeneous tissue.[56-57] Thus, applications of bioluminescence imaging *in vivo* typically involve monitoring *relative* changes in samples over time. The important takeaway is that the substrate unmixing platform can be immediately employed in conventional applications of bioluminescent probes, where detecting changes over time is often most important.

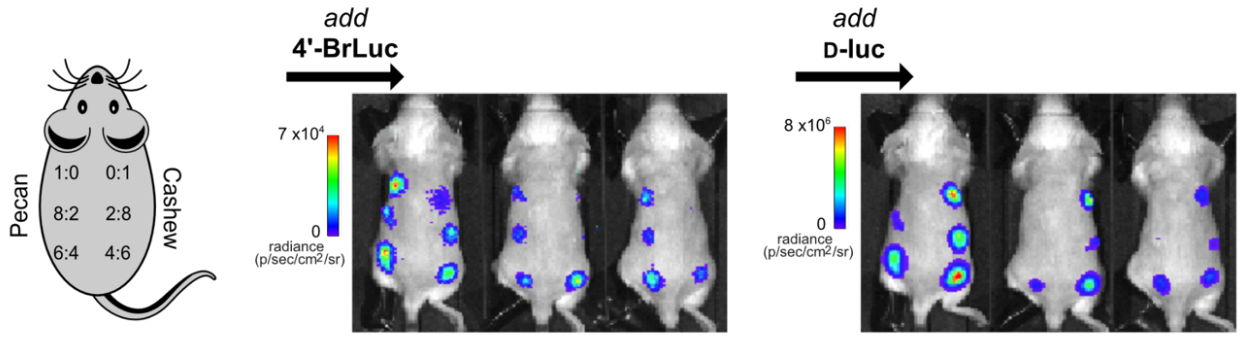


**Figure 2.12. Multicomponent BLI in mouse models.** Images used to generate the false colored pictures in Figure 2.11 are shown. Sequential application of 4'-BrLuc and D-luc enabled different ratios of Pecan- and Cashew-expressing cells to be visualized.

## Generalization of Rapid BLI to Three Probes and Monitoring Cell Function

Having demonstrated the ability to rapidly deconvolute mixtures of two bioluminescent probes, we examined whether the algorithm could be applied to larger collections of luciferase reporters. Triple-component bioluminescence imaging has been historically difficult to achieve, owing to a lack of protocols and methods to distinguish the probes.[31,58-59] In one of the only *in vivo* examples reported, multiple days were required to generate a composite image.[35] Signal from one substrate had to clear before another could be administered. We hypothesized that the "layering in" approach could dramatically speed the imaging time, enabling more early events relevant to cell growth to be visualized.

As a model triple component set, we used Cashew and Pecan in combination with Antares, a recently reported marine luciferase variant.[60-61] Cashew and Pecan derive from the insect luciferase family, and are thus immediately orthogonal to luciferases (like Antares) that use vastly distinct luciferins (in this case, furimazine).[10] Antares also exhibits markedly faster substrate turnover than Cashew, rendering it brighter and intensity resolved from the other two reporters.[36]

We thus reasoned that the three orthogonal luciferases could be rapidly differentiated by first applying 4'-BrLuc, followed by D-luc, then furimazine to layer in signal from Pecan, Cashew, and Antares, respectively. When the cells were combined and imaged together, the three reporters could be rapidly visualized (<15 min) following sequential substrate addition (Figure 2.13). Triplet imaging was also readily achieved using other combinations of engineered and native luciferases (Figures 2.14 – 2.15).

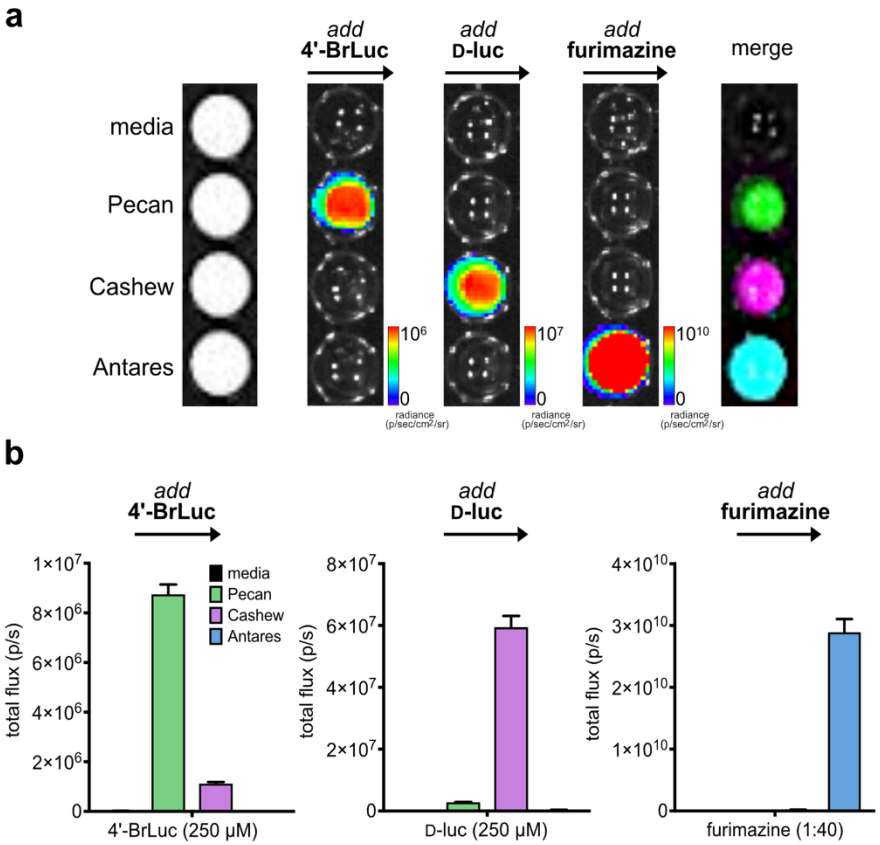


**Figure 2.13. Rapid BLI with three luciferases and luciferins**. (a) Cells expressing Pecan, Cashew, Antares, or no luciferase (control) were plated in a 96-well plate. Sequential substrate administration (4'-BrLuc, followed by D-luc, then furimazine) and unmixing enabled three-component imaging. Data are representative of $n = 3$ replicates. (b) Quantified photon outputs for the images in (a). Error bars represent the standard error of the mean for $n = 3$ experiments.

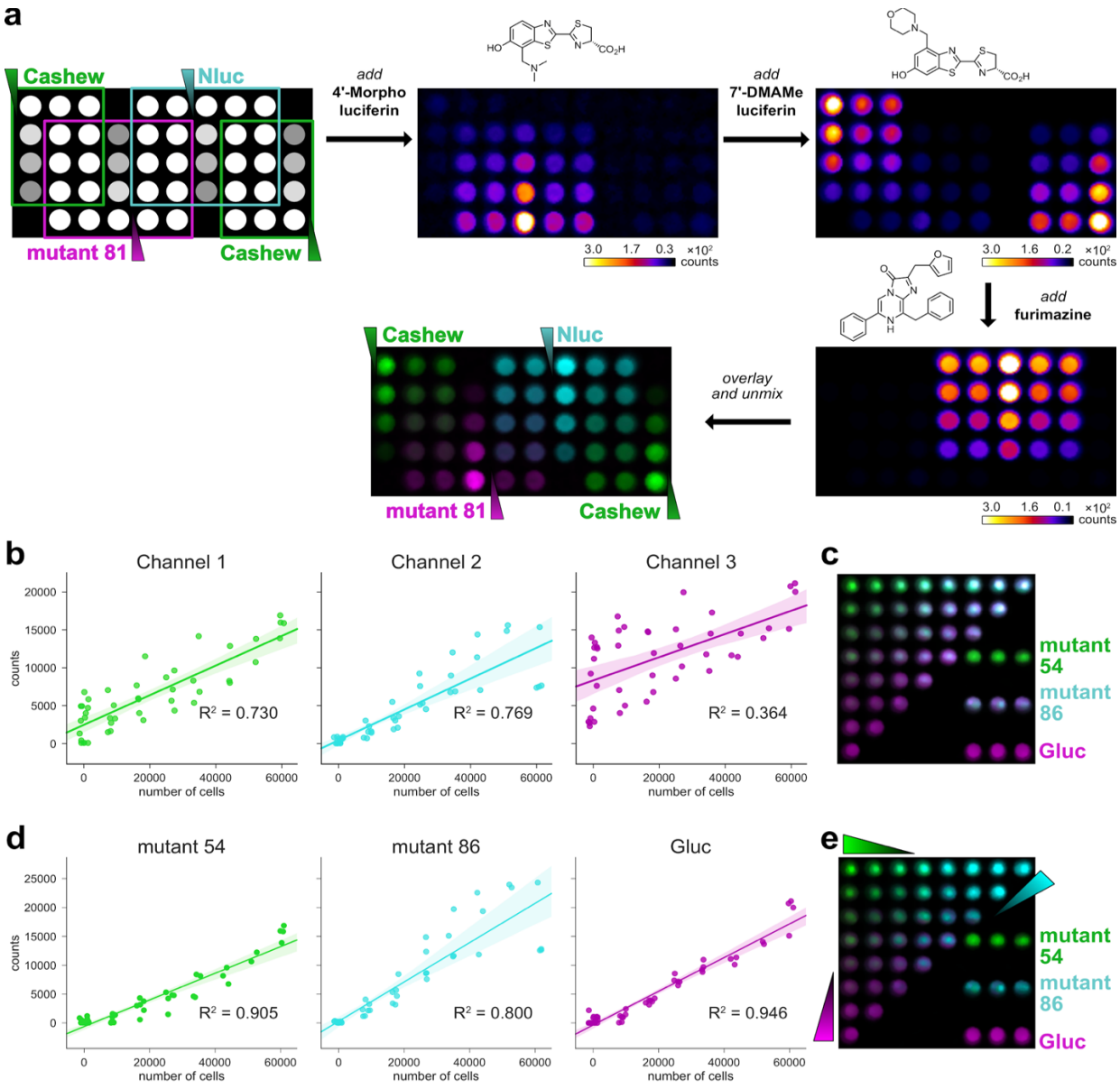


**Figure 2.14. Three orthogonal probes can be distinguished in bacterial lysate and mammalian cells.** (a) Gradients of luciferases in bacterial lysate were plated in a 96-well plate. 7'-DMAMeLuc luciferin[1-2] (250 μM), 4'-Morpho luciferin[1-2] (250 μM), furimazine (1:100 dilution of commercial stock) were added in sequence. Images were acquired after each addition, and the raw data were stacked and unmixed. (b-e) Gradients of cells expressing luciferase mutants 51 and 86, or *Gaussia* luciferase were plated in a triangle, with 60,000 cells per well. 4'-BrLuc (500 μM), D-luc (500 μM), and coelenterazine (40 μM) were added in sequence. (b) Quantification of each channel from (c) fit via linear regression. The shaded area represents the 95% confidence interval of the fit. (c) Overlay of raw signal from mixed images. (d) Quantification of each channel from the unmixed image in (e) fit via linear regression. The shaded area represents the 95% confidence interval of the fit. (e) Overlay of the unmixed channels

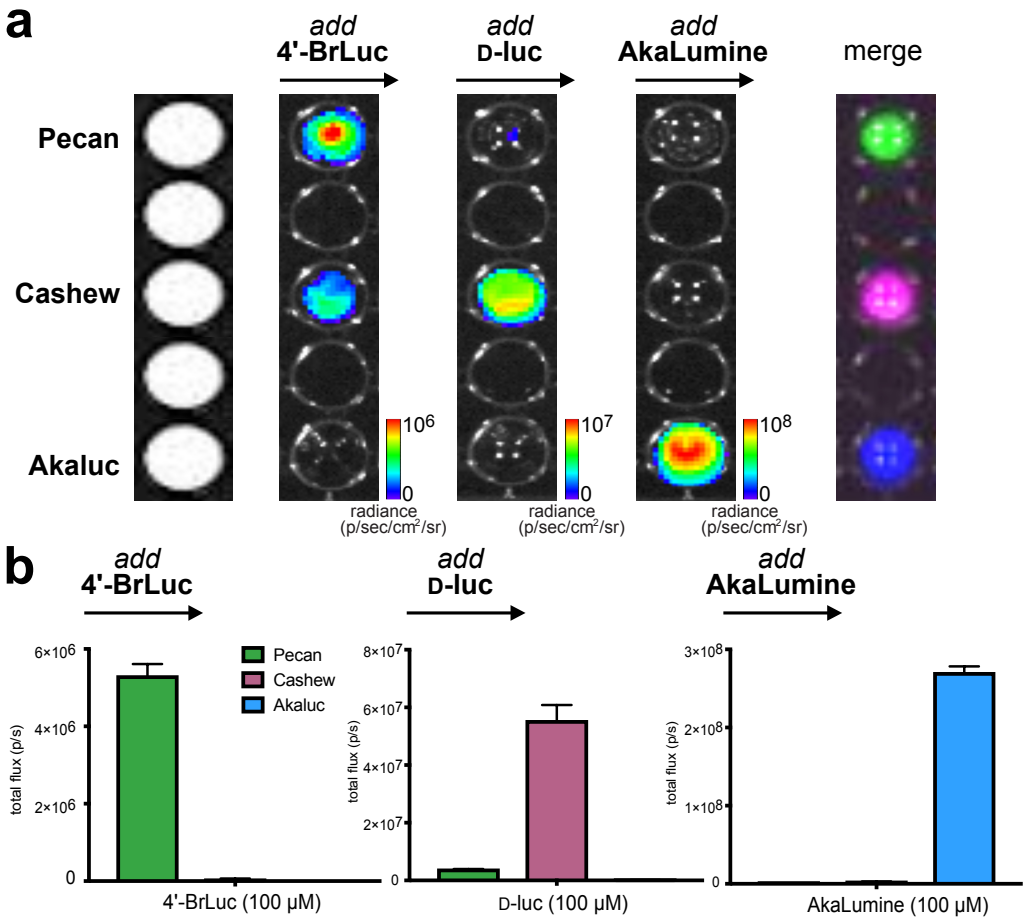


**Figure 2.15**. **Rapid BLI with three insect-derived luciferases and luciferins.** (a) Cells expressing Pecan, Cashew, or Akaluc were seeded in a 96-well plate. Sequential substrate administration (4'-BrLuc, followed by D-luc, then AkaLumine all at 100 µM), and unmixing enabled three-component imaging. Raw images were acquired after each substrate addition. The substrate-specific signals were unmixed, assigned false colors and overlaid. Data are representative of n = 3 replicates. (b) Quantified photon outputs for the images in (a). Error bars represent the standard error of the mean for n = 3 experiments.

## III. Conclusions and future directions

In conclusion, we developed a strategy for rapid multicomponent bioluminescence imaging based on substrate unmixing. This method takes advantage of both substrate and intensity resolution to resolve mixtures of reporters. In this scenario, probe differentiation is less impacted by tissue location, a parameter that has historically hindered efforts to resolve colors in large

organisms. We validated the unmixing approach in bacterial lysate, live cells, and mouse models. A variety of bioluminescent reporters were readily resolved, and bioluminescent outputs correlated with cell number and/or enzyme amount over a broad range of concentrations *in vitro*. The unmixing algorithm further enabled multiple luciferases to be discriminated *in vivo* using conventional BLI instrumentation.

While the rapid unmixing approach is immediately applicable to experiments with widely available bioluminescent probes, additional questions remain to be addressed. For example, just how much intensity resolution is required for successful unmixing remains unknown. The majority of probe sets tested exhibit >10-fold selectivity between matched enzymes and substrates. It is likely that even less selective probes can be integrated into the approach, though, and expand the number of features that can be imaged in a single setting. "Crosstalk" between more enzyme-substrate pairs could potentially improve the unmixing process, as the molecular signatures become more diagnostic. We also have not yet established the dynamic imaging range that can be achieved with different combinations of reporters. For example, it might not be possible to "see" a particular reporter if it is in low abundance compared to another reporter. The kinetics of light emission and compound transport into tissues must also be fully examined. These parameters could offer additional mechanisms by which to discriminate orthogonal pairs.[36,62]

The limits of detection will be influenced not only by the inherent photon output (and thus sensitivity) of each luciferase, but also the quantity of each reporter present and its tissue location. The data presented in this manuscript demonstrate that ten-fold changes in relative abundance can be easily discerned. Further theoretical work and experimental data are necessary to provide more definitive thresholds for different tissue types. Additional work is also necessary to streamline the quantification of luciferase reporters in biological samples. We have shown that the substrate

unmixing platform can easily monitor *relative* changes in luciferase levels. Tracking such information is standard practice, as traditional bioluminescence applications rely simply on detecting changes in signal over time. Identifying the *absolute* quantities of the reporters in unknown mixtures requires external calibration curves. Whether or not such curves must be generated for every independent experiment remains to be determined.

We anticipate that the rapid imaging approach reported here will enable a range of applications, including monitoring multiple cell types and gene expression profiles *in vivo*. The development of additional intensiometric probes will also increase the number of bioluminescent reporters that can be rapidly imaged in tandem, and work along these lines is ongoing. An expanded toolkit for BLI will enable longstanding questions regarding multicellular interactions to be addressed.

## MATERIALS AND METHODS

### Reagents

All reagents purchased from commercial suppliers were of analytical grade and used without further purification. 4'-BrLuc, 4'-MorphoLuc, 7'-MorPipLuc, and 7'-DMAMeLuc were prepared and used as previously described.[1-3]

### General bioluminescence imaging

Assays were performed in black 96-well plates (Greiner Bio One). Plates were imaged in a light-proof chamber (IVIS Lumina, Xenogen) equipped with a CCD camera (chilled to –90 °C). The stage was kept at 37 °C during imaging experiments, and the camera was controlled using standard Living Image software. Exposure times ranged from 1 s to 5 min, and data binning levels were set to small or medium. Post-acquisition, regions of interest were selected for quantification. Total flux and radiance values were analyzed using Living Image software or ImageJ (NIH).

### Bacterial lysate analysis of luciferase mutants

Bacterial cell stocks (stored in glycerol) expressing the mutants of interest were streaked on agar plates containing kanamycin. After overnight growth, colonies were picked and expanded overnight. Portions of the cultures (100 μL) were added to 5 mL of LB (kan) and luciferase expression was induced as described previously in Jones, *et al*.[1] For experiments involving gradients of bacterial lysate, luciferase-expressing bacteria was pelleted and ruptured with 600 μL of lysis buffer (50 mM Tris•HCl, 500 mM NaCl, 0.5% v/v Tween, 5 mM $MgCl_2$, pH = 7.4). In some cases, samples were diluted prior to plating to remain in the linear range of detection.

**Substrate unmixing**

Substrate unmixing experiments were designed such that a "positive" sample for each enzyme was present in the image to be acquired. For *in vitro* experiments, "positive" wells comprised pure enzyme. In mouse experiments, "positive" cells were also included as a reference. Images were acquired as a series following each substrate addition. Thus, an image was generated for each enzyme/substrate pair. Linear unmixing was conducted using ImageJ (installed under the FIJI package). Luminescence images containing the raw CCD counts (as TIFF files) were loaded into FIJI and subjected to a 2-pixel median filter to remove any cosmic noise. Next, the signal at each pixel was min-max scaled to lie between 0 and 65535 (the maximum value that can be stored in a 16-bit image). As a result, the brightest pixel in each image had a value of 65535, and the dimmest had a value of 0. Images were then stacked, and an additional image containing the maximum value of each of the stacked images was computed (as a Z projection). This new image was added to the stack, and signals were unmixed using the ImageJ plugin developed by Gammon, *et al*.[4] In the plugin, regions of interest (ROIs) for each luciferase were drawn around the "pure" areas of the image described above. Each ROI was drawn individually and added to the list by clicking "add." Once all enzymes were added, "Unmix" was used to unmix the images. Pseudocolors were assigned in FIJI through the "Merge Channels" tool.

**Mammalian plasmid construction**

Luciferase-expressing DB7 cells were prepared via CRISPR gene insertion. The relevant luciferase genes (luciferase-G4SX2-FP-T2A-Puro) were amplified and inserted into CRISPR AAVS1 donor plasmids (courtesy of Drs. Theresa Loveless and Chang Liu, UCI). Cashew and Pecan inserts were amplified from pET vectors using the following primers:

5’ – TGGCTAGCGCTACCGGTCGCCACCTCTAGAATGGAAGACGCCAAAAACATAAA

GAAAGG -3’ and

5’- GCGGAAAGATCGCCGTGGGCGGAGGCGGGTCTGGGGGCGGAGGCTCT -3’

Antares inserts were amplified with the following primers:

5’- GCTAGCGCTACCGGTCGCCACCTCTAGAATGCGGGGTTCTCATCATCATCATC -3’ and

5’- TGCCTCTGCCCTCGCCGCTGCCCTCGAGCTTGTACAGCTCGTCCATGCCTCCG -3

Akaluc inserts were amplified with the following primers:

5’- ATGGAAGATGCCAAAAACATTAAGAAGGGCCCAGC -3’ and

5’- CACGGCGATCTTGCCGTCCTTCTTGGCCTTAGTGA -3’

Linearized vectors were generated via digestion with restriction enzymes *Xba*I and *Xho*I (New England BioLabs). The linearized vectors were combined with the appropriate luciferase insert by Gibson assembly. A portion of the reactions (3.0 µL) was directly transformed into XL1 competent *E. coli* cells. Sequencing analysis confirmed successful plasmid generation.

**Mammalian cell culture and imaging**

DB7 cells (courtesy of the Contag laboratory, Stanford) were cultured in DMEM (Corning) supplemented with 10% (vol/vol) fetal bovine serum (FBS, Life Technologies), penicillin (100 U/mL), and streptomycin (100 µg/mL). Cells were maintained in a 5% $CO_2$ water-saturated incubator at 37 °C. To create stable lines expressing mutant luciferases, DB7 cells were transfected with the AAVS1 mutant luciferase donor plasmid, Cas9 (Addgene #41815), and AAVS1 sgRNA

(Addgene #53370) using lipofectamine. The mutant luciferases were integrated into the first locus of AAVS1 through homologous recombination. Transfected cells were then treated with puromycin (2 μg/mL) and FACS sorted at the Institute for Immunology Flow Cytometry Core (UCI).

DB7 cells stably expressing luciferases were added to black 96-well plates (1 x $10^5$ cells per well). Stock solutions of 4'-BrLuc, D-luc, and AkaLumine (10 mM in PBS) were added to each well (100-500 μM final concentration). A solution of furimazine (1:40-1:100 dilution of the commercial stock, Promega) was then added. Sequential imaging was performed as described in the General bioluminescence imaging section (above).

***In vivo* imaging of orthogonal luciferase-luciferin pairs**

Mouse experiments were approved by the UC Irvine Animal Care and Use Committee. FVB/NJ mice (The Jackson Laboratory) received subcutaneous dorsal injections of 1x$10^6$ DB7 mutant luciferase expressing cells. After 24 h, animals were anesthetized (1-2% isoflurane) and placed on a warmed (37 °C) stage for imaging. Each mouse received an i.p. injection of luciferin (65 mM, 100 μL per mouse). Images were acquired with 5 min exposure times for 35 min using the General bioluminescence procedure. For sequential imaging, mice were immediately injected with the second substrate and imaged for an additional 35 min. Bioluminescent output was quantified as above.

**References**


1. Paley, M. A.; Prescher, J. A. Bioluminescence: a versatile technique for imaging cellular and molecular features. *MedChemComm* **2014**, *5*, 255–267.

2. Mezzanotte, L.; van 't Root, M.; Karatas, H.; Goun, E. A.; Löwik, C. W. G. M. In vivo Molecular Bioluminescence Imaging: New Tools and Applications. *Trends Biotechnol.* **2017**, *35*, 640–652.

3. Prescher, J. A.; Contag, C. H. Guided by the light: visualizing biomolecular processes in living animals with bioluminescence. *Curr. Opin. Chem. Biol.* **2010**, *14*, 80–89.

4. Contag, C. H.; Bachmann, M. H. Advances in In vivo Bioluminescence Imaging of Gene Expression. *Annu. Rev. Biomed. Eng.* **2002**, *4*, 235–260.

5. Thorne, N.; Inglese, J.; Auld, D. S. Illuminating insights into firefly luciferase and other bioluminescent reporters used in chemical biology. *Chem. Biol.* **2010**, *17*, 646–657.

6. Rabinovich, B. A.; Ye, Y.; Etto, T.; Chen, J. Q.; Levitsky, H. I.; Overwijk, W. W.; Cooper, L. J. N.; Gelovani, J.; Hwu, P. Visualizing fewer than 10 mouse T cells with an enhanced firefly luciferase in immunocompetent mouse models of cancer. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 14342–14346.

7. Yao, Z.; Zhang, B. S.; Prescher, J. A. Advances in bioluminescence imaging: new probes from old recipes. *Curr. Opin. Chem. Biol.* **2018**, *45*, 148–156.

8. Rathbun, C. M.; Prescher, J. A. Bioluminescent probes for imaging biology beyond the culture dish. *Biochemistry* **2017**, *56*, 5178–5184.

9. Miller, S. C.; Mofford, D. M.; Adams, S. T. Lessons learned from luminous luciferins and latent luciferases. *ACS Chem. Biol.* **2018**, *13*, 1734–1740.

10. Yeh, H.-W.; Ai, H.-W. Development and applications of bioluminescent and chemiluminescent reporters and biosensors. *Annu. Rev. Anal. Chem.* **2019**, *12*, 129–150.

11. Yeh, H.-W.; Xiong, Y.; Wu, T.; Chen, M.; Ji, A.; Li, X.; Ai, H.-W. ATP-Independent Bioluminescent Reporter Variants To Improve in vivo Imaging. *ACS Chem. Biol.* **2019**, *14*, 959–965.

12. Hall, M. P.; Woodroofe, C. C.; Wood, M. G.; Que, I.; van't Root, M.; Ridwan, Y.; Shi, C.; Kirkland, T. A.; Encell, L. P.; Wood, K. V.; Löwik, C.; Mezzanotte, L. Click beetle luciferase mutant and near infrared naphthyl-luciferins for improved bioluminescence imaging. *Nat. Commun.* **2018**, *9*, 132.

13. Schaub, F. X.; Reza, M. S.; Flaveny, C. A.; Li, W.; Musicant, A. M.; Hoxha, S.; Guo, M.; Cleveland, J. L.; Amelio, A. L. Fluorophore-NanoLuc BRET Reporters Enable Sensitive In vivo

Optical Imaging and Flow Cytometry for Monitoring Tumorigenesis. *Cancer Research* **2015**, *75*, 5023–5033.

14. Iwano, S.; Sugiyama, M.; Hama, H.; Watakabe, A.; Hasegawa, N.; Kuchimaru, T.; Tanaka, K. Z.; Takahashi, M.; Ishida, Y.; Hata, J.; Shimozono, S.; Namiki, K.; Fukano, T.; Kiyama, M.; Okano, H.; Kizaka-Kondoh, S.; McHugh, T. J.; Yamamori, T.; Hioki, H.; Maki, S.; Miyawaki, A. Single-cell bioluminescence imaging of deep tissue in freely moving animals. *Science* **2018**, *359*, 935–939.

15. Williams, S. J.; Prescher, J. A. Building biological flashlights: Orthogonal luciferases and luciferins for *in vivo* imaging. *Acc. Chem. Res.* **2019**, *52*, 3039–3050.

16. Zambito, G.; Hall, M. P.; Wood, M. G.; Gaspar, N.; Ridwan, Y.; Stellari, F. F.; Shi, C.; Kirkland, T. A.; Encell, L. P.; Löwik, C.; Mezzanotte, L. Red-shifted click beetle luciferase mutant expands the multicolor bioluminescent palette for deep tissue imaging. *iScience* **2021**, *24*, 101986.

17. Stowe, C. L.; Burley, T. A.; Allan, H.; Vinci, M.; Kramer-Marek, G.; Ciobota, D. M.; Parkinson, G. N.; Southworth, T. L.; Agliardi, G.; Hotblack, A.; Lythgoe, M. F.; Branchini, B. R.; Kalber, T. L.; Anderson, J. C.; Pule, M. A. Near-infrared dual bioluminescence imaging in mouse models of cancer using infraluciferin. *eLife* **2019**, *8*, e45801.

18. Markus, A.; Stefanie, V.; Cordula, S.; Amit, J.; Martin, P.; Mathias, H. Quantitative in vivo dual-color bioluminescence imaging in the mouse brain. *Neurophotonics* **2019**, *6*, 025006.

19. Branchini, B. R.; Ablamsky, D. M.; Rosenman, J. M.; Uzasci, L.; Southworth, T. L.; Zimmer, M. Synergistic mutations produce blue-shifted bioluminescence in firefly luciferase. *Biochemistry* **2007**, *46*, 13847–13855.

20. Mezzanotte, L.; Que, I.; Kaijzel, E.; Branchini, B.; Roda, A.; Löwik, C. Sensitive dual color *in vivo* bioluminescence imaging using a new red codon optimized firefly luciferase and a green click beetle luciferase. *PLoS ONE* **2011**, *6*, e19277.

21. Kleinovink, J. W.; Mezzanotte, L.; Zambito, G.; Fransen, M. F.; Cruz, L. J.; Verbeek, J. S.; Chan, A.; Ossendorp, F.; Löwik, C. A Dual-Color Bioluminescence Reporter Mouse for Simultaneous in vivo Imaging of T Cell Localization and Function. *Front. Immunol.* **2019**, *9*, 3097.

22. Sarrion-Perdigones, A.; Chang, L.; Gonzalez, Y.; Gallego-Flores, T.; Young, D. W.; Venken, K. J. T. Examining multiple cellular pathways at once using multiplex hextuple luciferase assaying. *Nat. Commun.* **2019**, *10*, 5710.

23. Bhaumik, S.; Lewis, X. Z.; Gambhir, S. S. Optical imaging of *Renilla* luciferase, synthetic *Renilla* luciferase, and firefly luciferase reporter gene expression in living mice. *J. Biomed. Opt.* **2004**, *9*, 578–586.

24. Fan, F.; Wood, K. V. Bioluminescent assays for high-throughput screening. *ASSAY Drug Dev. Techn.* **2007**, *5*, 127–136.

25. Su, Y.; Walker, J. R.; Park, Y.; Smith, T. P.; Liu, L. X.; Hall, M. P.; Labanieh, L.; Hurst, R.; Wang, D. C.; Encell, L. P.; Kim, N.; Zhang, F.; Kay, M. A.; Casey, K. M.; Majzner, R. G.; Cochran, J. R.; Mackall, C. L.; Kirkland, T. A.; Lin, M. Z. Novel NanoLuc substrates enable bright two-population bioluminescence imaging in animals. *Nat. Methods* **2020**, *17*, 852–860.

26. Gross, S.; Abraham, U.; Prior, J. L.; Herzog, E. D.; Piwnica-Worms, D. Continuous delivery of D-luciferin by implanted micro-osmotic pumps enables true real-time bioluminescence imaging of luciferase activity in vivo. *Mol Imaging* **2007**, *6*, 121–130.

27. Kaskova, Z. M.; Tsarkova, A. S.; Yampolsky, I. V. 1001 lights: luciferins, luciferases, their mechanisms of action and applications in chemical analysis, biology and medicine. *Chem. Soc. Rev.* **2016**, *45*, 6048–6077.

28. Kotlobay, A. A.; Sarkisyan, K. S.; Mokrushina, Y. A.; Marcet-Houben, M.; Serebrovskaya, E. O.; Markina, N. M.; Somermeyer, L. G.; Gorokhovatsky, A. Y.; Vvedensky, A.; Purtov, K. V.; Petushkov, V. N.; Rodionova, N. S.; Chepurnyh, T. V.; Fakhranurova, L. I.; Guglya, E. B.; Ziganshin, R.; Tsarkova, A. S.; Kaskova, Z. M.; Shender, V.; Abakumov, M.; Abakumova, T. O.; Povolotskaya, I. S.; Eroshkin, F. M.; Zaraisky, A. G.; Mishin, A. S.; Dolgov, S. V.; Mitiouchkina, T. Y.; Kopantzev, E. P.; Waldenmaier, H. E.; Oliveira, A. G.; Oba, Y.; Barsova, E.; Bogdanova, E. A.; Gabaldón, T.; Stevani, C. V.; Lukyanov, S.; Smirnov, I. V.; Gitelson, J. I.; Kondrashov, F. A.; Yampolsky, I. V. Genetically encodable bioluminescent system from fungi. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 12728–12732.

29. Mitiouchkina, T.; Mishin, A. S.; Somermeyer, L. G.; Markina, N. M.; Chepurnyh, T. V.; Guglya, E. B.; Karataeva, T. A.; Palkina, K. A.; Shakhova, E. S.; Fakhranurova, L. I.; CHEK-293Tova, S. V.; Tsarkova, A. S.; Golubev, Y. V.; Negrebetsky, V. V.; Dolgushin, S. A.; Shalaev, P. V.; Shlykov, D.; Melnik, O. A.; Shipunova, V. O.; Deyev, S. M.; Bubyrev, A. I.; Pushin, A. S.; Choob, V. V.; Dolgov, S. V.; Kondrashov, F. A.; Yampolsky, I. V.; Sarkisyan, K. S. Plants with genetically encoded autoluminescence. *Nat. Biotechnol.* **2020**, *38*, 944–946.

30. Jones, K. A.; Porterfield, W. B.; Rathbun, C. M.; McCutcheon, D. C.; Paley, M. A.; Prescher, J. A. Orthogonal luciferase–luciferin pairs for bioluminescence imaging. *J. Am. Chem. Soc.* **2017**, *139*, 2351–2358.

31. Rathbun, C. M.; Porterfield, W. B.; Jones, K. A.; Sagoe, M. J.; Reyes, M. R.; Hua, C. T.; Prescher, J. A. Parallel screening for rapid identification of orthogonal bioluminescent tools. *ACS Cent. Sci.* **2017**, *3*, 1254–1261.

32. Mofford, D. M.; Reddy, G. R.; Miller, S. C. Aminoluciferins extend firefly luciferase bioluminescence into the near-infrared and can be preferred substrates over D-luciferin. *J. Am. Chem. Soc.* **2014**, *136*, 13277–13282.

33. Kim, S. B.; Nishihara, R.; Citterio, D.; Suzuki, K. Fabrication of a new lineage of artificial luciferases from natural luciferase pools. *ACS Comb. Sci.* **2017**, *19*, 594–599.

34. Yeh, H.-W.; Karmach, O.; Ji, A.; Carter, D.; Martins-Green, M. M.; Ai, H.-w. Red-shifted luciferase–luciferin pairs for enhanced bioluminescence imaging. *Nat. Methods* **2017**, *14*, 971–974.

35. Maguire, C. A.; Bovenberg, M. S.; Crommentuijn, M. H.; Niers, J. M.; Kerami, M.; Teng, J.; Sena-Esteves, M.; Badr, C. E.; Tannous, B. A. Triple bioluminescence imaging for in vivo monitoring of cellular processes. *Mol Ther Nucleic Acids* **2013**, *2*, e99.

36. Stacer, A. C.; Nyati, S.; Moudgil, P.; Iyengar, R.; Luker, K. E.; Rehemtulla, A.; Luker, G. D. NanoLuc Reporter for Dual Luciferase Imaging in Living Animals. *Mol Imaging* **2013**, *12*, 457–469.

37. Shah, K.; Tang, Y.; Breakefield, X.; Weissleder, R. Real-time imaging of TRAIL-induced apoptosis of glioma tumors in vivo. *Oncogene* **2003**, *22*, 6865–6872.

38. Masuho, I.; Ostrovskaya, O.; Kramer, G. M.; Jones, C. D.; Xie, K.; Martemyanov, K. A. Distinct profiles of functional discrimination among G proteins determine the actions of G protein–coupled receptors. *Science Signaling* **2015**, *8*, ra123.

39. Hartfield, R. M.; Schwarz, K. A.; Muldoon, J. J.; Bagheri, N.; Leonard, J. N. Multiplexing Engineered Receptors for Multiparametric Evaluation of Environmental Ligands. *ACS Synth. Biol.* **2017**, *6*, 2042–2055.

40. Chaussabel, D.; Baldwin, N. Democratizing systems immunology with modular transcriptional repertoire analyses. *Nat. Rev. Immunol.* **2014**, *14*, 271–280.

41. Gammon, S. T.; Leevy, W. M.; Gross, S.; Gokel, G. W.; Piwnica-Worms, D. Spectral unmixing of multicolored bioluminescence emitted from heterogeneous biological sources. *Anal. Chem.* **2006**, *78*, 1520–1527.

42. Eng, C.-H. L.; Lawson, M.; Zhu, Q.; Dries, R.; Koulena, N.; Takei, Y.; Yun, J.; Cronin, C.; Karp, C.; Yuan, G.-C.; Cai, L. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **2019**, *568*, 235–239.

43. Guo, S.-M.; Veneziano, R.; Gordonov, S.; Li, L.; Danielson, E.; de Arce, K. P.; Park, D.; Kulesa, A. B.; Wamhoff, E.-C.; Blainey, P. C.; Boyden, E. S.; Cottrell, J. R.; Bathe, M. Multiplexed and high-throughput neuronal fluorescence imaging with diffusible probes. *Nat. Commun.* **2019**, *10*, 4377.

44. Steinhardt, R. C.; Rathbun, C. M.; Krull, B. T.; Yu, J. M.; Yang, Y.; Nguyen, B. D.; Kwon, J.; McCutcheon, D. C.; Jones, K. A.; Furche, F.; Prescher, J. A. Brominated Luciferins Are Versatile Bioluminescent Probes. *ChemBioChem* **2017**, *18*, 96-100.

45. McCutcheon, D. C.; Paley, M. A.; Steinhardt, R. C.; Prescher, J. A. Expedient synthesis of electronically modified luciferins for bioluminescence imaging. *J. Am. Chem. Soc.* **2012**, *134*, 7604–7607.

46. McCutcheon, D. C.; Porterfield, W. B.; Prescher, J. A. Rapid and scalable assembly of firefly luciferase substrates. *Org. Biomol. Chem.* **2015**, *13*, 2117–2121.

47. Liu, M. D.; Warner, E. A.; Morrissey, C. E.; Fick, C. W.; Wu, T. S.; Ornelas, M. Y.; Ochoa, G. V.; Zhang, B. S.; Rathbun, C. M.; Porterfield, W. B.; Prescher, J. A.; Leconte, A. M. Statistical coupling analysis-guided library design for the discovery of mutant luciferases. *Biochemistry* **2018**, *57*, 663–671.

48. Adams, S. T.; Mofford, D. M.; Reddy, G. S. K. K.; Miller, S. C. Firefly luciferase mutants allow substrate-selective bioluminescence imaging in the Mouse Brain. *Angew. Chem. Int. Ed.* **2016**, *55*, 4943–4946.

49. Zhang, B. S.; Jones, K. A.; McCutcheon, D. C.; Prescher, J. A. Pyridone luciferins and mutant luciferases for bioluminescence imaging. *ChemBioChem* **2018**, *19*, 470–477.

50. Shachaf, C. M.; Kopelman, A. M.; Arvanitis, C.; Karlsson, Å.; Beer, S.; Mandl, S.; Bachmann, M. H.; Borowsky, A. D.; Ruebner, B.; Cardiff, R. D.; Yang, Q.; Bishop, J. M.; Contag, C. H.; Felsher, D. W. MYC inactivation uncovers pluripotent differentiation and tumour dormancy in hepatocellular cancer. *Nature* **2004**, *431*, 1112–1117.

51. Minn, A. J.; Kang, Y.; Serganova, I.; Gupta, G. P.; Giri, D. D.; Doubrovin, M.; Ponomarev, V.; Gerald, W. L.; Blasberg, R.; Massagué, J. Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. *J. Clin. Invest.* **2005**, *115*, 44–55.

52. Luker, G. D.; Bardill, J. P.; Prior, J. L.; Pica, C. M.; Piwnica-Worms, D.; Leib, D. A. Noninvasive Bioluminescence Imaging of Herpes Simplex Virus Type 1 Infection and Therapy in Living Mice. *J. Virol.* **2002**, *76*, 12149–12161.

53. Abdelfattah, A. S.; Kawashima, T.; Singh, A.; Novak, O.; Liu, H.; Shuai, Y.; Huang, Y.-C.; Campagnola, L.; Seeman, S. C.; Yu, J.; Zheng, J.; Grimm, J. B.; Patel, R.; Friedrich, J.; Mensh, B. D.; Paninski, L.; Macklin, J. J.; Murphy, G. J.; Podgorski, K.; Lin, B.-J.; Chen, T.-W.; Turner, G. C.; Liu, Z.; Koyama, M.; Svoboda, K.; Ahrens, M. B.; Lavis, L. D.; Schreiter, E. R. Bright and photostable chemigenetic indicators for extended in vivo voltage imaging. *Science* **2019**, *365*, 699–704.

54. Sanford, L.; Palmer, A. Recent Advances in Development of Genetically Encoded Fluorescent Sensors. *Methods Enzymol.* **2017**, *589*, 1–49.

55. Chen, T.-W.; Wardill, T. J.; Sun, Y.; Pulver, S. R.; Renninger, S. L.; Baohan, A.; Schreiter, E. R.; Kerr, R. A.; Orger, M. B.; Jayaraman, V.; Looger, L. L.; Svoboda, K.; Kim, D. S. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **2013**, *499*, 295–300.

56. Chaincy, K.; Olivier, C.; Tamara, L. T.; Heng, X.; Bradley, W. R. Three-dimensional reconstruction of in vivo bioluminescent sources based on multispectral imaging. *J. Biomed. Opt.* **2007**, *12*, 1–12.

57. Doyle, T. C.; Burns, S. M.; Contag, C. H. Technoreview: In vivo bioluminescence imaging for integrated studies of infection. *Cell. Microbiol.* **2004**, *6*, 303–317.

58. Yao, Z.; Zhang, B. S.; Steinhardt, R. C.; Mills, J. H.; Prescher, J. A. Multicomponent Bioluminescence Imaging with a π-Extended Luciferin. *J. Am. Chem. Soc.* **2020**, *142*, 14080–14089.

59. Tannous, B. A.; Kim, D.-E.; Fernandez, J. L.; Weissleder, R.; Breakefield, X. O. Codon-optimized *Gaussia* luciferase cDNA for mammalian gene expression in culture and *in vivo*. *Mol. Ther.* **2005**, *11*, 435–443.

60. Chu, J.; Oh, Y.; Sens, A.; Ataie, N.; Dana, H.; Macklin, J. J.; Laviv, T.; Welf, E. S.; Dean, K. M.; Zhang, F.; Kim, B. B.; Tang, C. T.; Hu, M.; Baird, M. A.; Davidson, M. W.; Kay, M. A.; Fiolka, R.; Yasuda, R.; Kim, D. S.; Ng, H.-L.; Lin, M. Z. A bright cyan-excitable orange fluorescent protein facilitates dual-emission microscopy and enhances bioluminescence imaging in vivo. *Nat. Biotechnol.* **2016**, *34*, 760–767.

61. Oh, Y.; Park, Y.; Cho, J. H.; Wu, H.; Paulk, N. K.; Liu, L. X.; Kim, N.; Kay, M. A.; Wu, J. C.; Lin, M. Z. An orange calcium-modulated bioluminescent indicator for non-invasive activity imaging. *Nat. Chem. Biol* **2019**, *15*, 433–436.

62. Yeh, H.-W., Wu, T., Chen, M., and Ai, H.-W. (2019) Identification of factors complicating bioluminescence imaging, *Biochemistry 58*, 1689–1697.

# CHAPTER 3: Transcriptome analysis of heterogeneity in mouse models of metastatic breast cancer *

## I. Introduction

Despite recent advances in treatment and diagnosis, metastatic breast cancer remains a leading cause of death for women worldwide [1]. Cancer metastasis is a complex process involving the spread of malignant cells from a primary tumor to distal organs [2][3]. Premalignant cells undergo dynamic cellular changes (i.e., epithelial to mesenchymal transition, EMT) to escape the primary tumor [3,4]. These same cells undergo the reverse process (i.e., mesenchymal to epithelial transition, MET) to colonize metastatic sites [5]. Expression and fluctuations of cell surface markers (e.g., CD44) have long been associated with metastatic progression in breast cancer [4]. However, exactly which cells within a given primary tumor ultimately metastasize—and their final destinations—remains unclear [4-6].

Transcriptome profiling of the dynamic cellular changes during tumorigenesis has the potential to improve our understanding of metastatic disease. Such analyses can reveal biomarkers associated with malignant progression. In one example, bulk RNA-sequencing (RNA-seq) revealed novel molecular pathways and differentially expressed genes (DEGs) associated with distinct stages of breast cancer progression [7]. However, traditional profiling studies are complicated by the underlying heterogeneity of cancer progression. The contribution of distinct

* *Author's Note: The following chapter is a completed manuscript that has been accepted for publication in Breast Cancer Research. I contributed to designing experiments and procedures, data analysis and interpretation, and writing. The co-authors include Gabriela Balderrama-Gutierrez, Krystian J. Ibanez, Steve Huy D. Phan, Angelique N. Cortez, Ali Mortazavi, and last author Jennifer A. Prescher*

cell populations cannot be discerned using bulk RNA-seq alone [8]. Single-cell RNA-seq (scRNA-seq) technology captures the complexity of cellular heterogeneity by mapping transcripts to individual cells [9]. This increase in cellular resolution facilitates the identification of additional molecular pathways and cell specific biomarkers [10, 11].

Examining breast cancer remains challenging owing to a lack of models that capture cellular heterogeneity [12]. The surrounding microenvironment, cancer stem cells (CSC), tumor dormancy, all contribute to disease progression beyond isolated changes to the malignant cells themselves. These features are difficult to replicate outside of living organisms. Suitable models must take into account the different tissue microenvironments that support cancer niches and resident cancer stem cells during metastatic progression [13-16]. The MMTV-PyMT mouse model, in particular, is a well-established platform to study human breast cancer [17, 18]. However, the variability in different metastatic niches and the contribution of different cancer cell types to disease progression remains unclear. Subclones across breast tumors are frequently identified and monitored using the expression of the cell surface marker CD44 [4-6]. However, this marker is associated with both pro- and anti-tumorigenic outcomes, meaning that CD44 expression alone cannot be used to predict metastatic propensity or other cell behaviors [67-69,76]. Transcriptome profiling of the MMTV-PyMT cancer model could thus provide more insight into the mechanisms underlying dynamic changes in tumor progression [19].

We aimed to understand the transcriptome changes of organ-derived cancer cell isolates from MMTV-PyMT mice. Although metastatic progression from primary tumors to lung tissue is well studied in the MMTV-PyMT model, metastases to other distal organs and the significance of intratumor heterogeneity remain unclear [20]. To gain insight, we established an array of metastatic cell lines harvested from MMTV-PyMT mice. Differential expression analyses were

performed and used to examine the effects of cell heterogeneity on metastases and organ tropism. Correlations were found between CD44 expression and cellular growth markers across all metastatic cells. Data from scRNA-seq analyses further revealed tissue-specific gene expression patterns that mirror clinical data. Overall, the suite of clonal isolates provided a detailed depiction of cancer progression. The cell lines also establish a platform for future studies examining heterogeneity during metastatic disease and elucidating transcriptomic changes relevant to malignancy.

## II. Results

### *Generation of breast cancer cell lines to examine tumor heterogeneity and metastatic disease*

To gain insight into breast cancer heterogeneity, we derived a suite of tissue-specific metastatic cell lines from MMTV-PyMT mouse tumors (Fig. 3.1A). Tumors were harvested from the mammary fat pad (MFP) and tissues harboring distal metastases, including lymph nodes (LN), bone marrow (BM), and lungs (L). Samples were processed into single cell suspensions and further expanded. The organ-derived cultures were subjected to conditions that favored cancer cell outgrowth *in vitro*. Cells were ultimately sorted based on CD44 and EpCAM expression [23] to remove fibroblasts from the samples. CD44 is routinely used as a marker of aggressive metastatic breast cancer [29]. FACS sorting provided two populations: CD44low/EpCAMhigh and CD44high/EpCAMhigh (Fig. 3.2A-B). PCR was also used to confirm the presence of the PyMT viral antigen in the cell isolates (Fig. 3.2C). For the sorting and PCR assays, an established MMTV-PyMT cancer cell line (VO) and a common fibroblast cell line (3T3) were used as positive and negative controls, respectively. The tumorigenicity and metastatic propensity of the sorted MFP cell line was validated *in vivo* by injecting cultured cells into wild type female FVB mice

(Fig. 3.1B). The presence of metastatic tumors was confirmed by harvesting LN, BM, and lung tissue from the re-injected mice. Single cell suspensions were formed and flow cytometry analysis confirmed the presence of $CD44^{low}/EpCAM^{high}$ and $CD44^{high}/EpCAM^{high}$ cells in the harvested tissues.

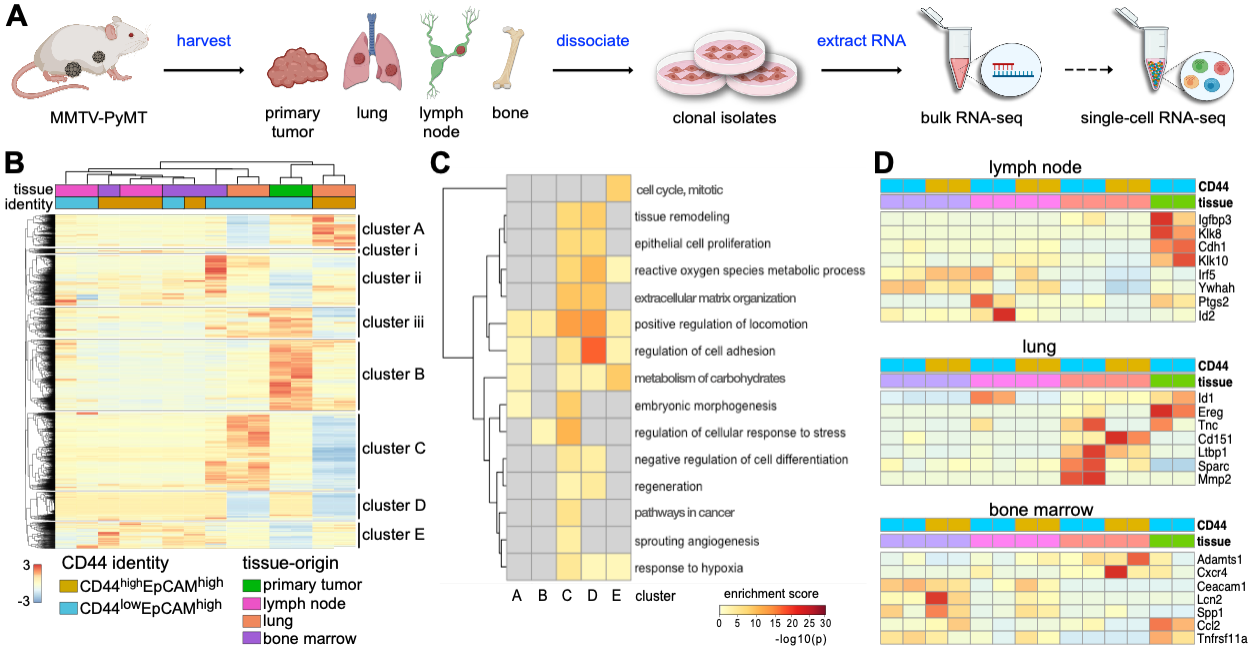


**Figure 3.1. Clonal isolates from MMTV-PyMT breast cancer model exhibit distinct gene expression patterns.** (A) Overview of cell isolation procedures and gene expression analyses. Tumors were harvested from mice and single cell suspensions were prepared. Cells were sorted based on CD44 and EpCAM expression. RNA was extracted for transcriptome profiling. Select samples were further analyzed via single-cell RNA-seq. (B) Heatmap of DEGs from tissue-specific metastatic cell lines and primary tumor sample. Expression levels for 5509 unique genes are shown. Values were normalized by row, and hierarchical clustering was used to sort the transcripts. Columns were clustered based on the tissue origin and CD44 expression level for each sample. Eight distinct gene clusters were observed, with clusters of interest annotated A-E. (C) GO-term enrichment analysis of clusters A-E from (B). GO terms were used to identify ontologies and biological processes relevant to cancer metastasis. Terms were also analyzed for signatures specific to the tissue of origin. The heat maps indicate the relative enrichment of the pathways across each cluster (columns). (D) Bulk RNA analysis revealed distinct gene expression patterns relevant to organ tropism. A panel of markers associated with tissue-tropic breast cancer metastases was examined across samples. Clusters were assigned based on based on the tissue origin and CD44 expression level.

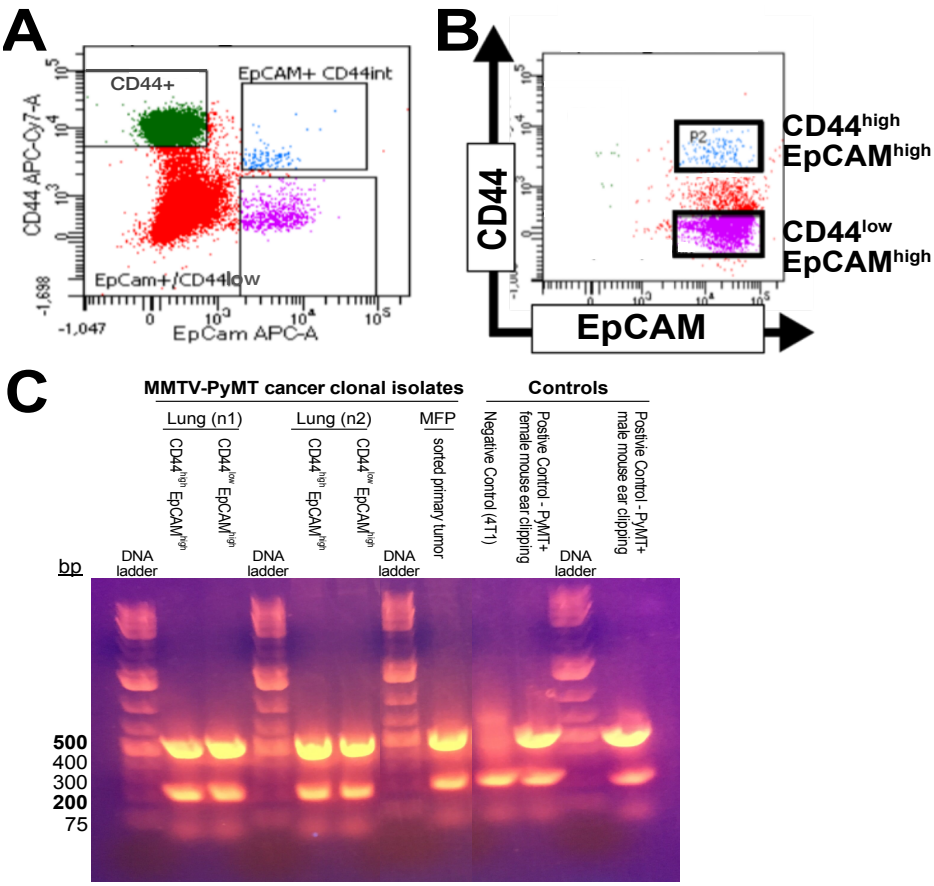


**Figure 3.2. Metastatic cell line derivation and validation.** (A) Representative flow cytometry plot for metastatic tissue-derived isolates prior to sorting. The desired CD44$^{low}$/EpCAM$^{high}$ and CD44$^{high}$/EpCAM$^{high}$ cell populations and associated gates are highlighted. These cells were sorted and for transcriptomic analysis. (B) Clonal isolates from (A) recapitulate metastatic disease in vivo. Sorted cells were reimplanted into disease-free FVB mice. Once tumors formed, distal organs were collected and examined for metastatic lesions. A representative FACS plot for metastatic cells harvested from lymph node tissue upon tumor reimplantation in FVB mice. The expected CD44$^{low}$/EpCAM$^{high}$ and CD44$^{high}$/EpCAM$^{high}$ populations are highlighted. (C) PCR was used to confirm the presence of PyMT viral antigen using gDNA extract from cell lines and controls. Samples with amplified 500bp and 200bp bands are positive for PyMT antigen expression.

### *Cancer cell lines exhibit distinct gene expression changes relative to metastatic progression*

We used the tissue-derived cell lines to investigate transcriptional changes that occur during breast cancer metastasis. RNA was extracted from all cell samples and transcripts for

established breast cancer genes were identified (Fig. 3.3) [30-32]. Hierarchical clustering was performed on 5,509 DEGs. Eight distinct gene clusters (A-E; i-iii) were observed, as shown in Fig 3.1B. The transcripts were organized based on CD44 expression ($CD44^{high}$/$EpCAM^{high}$ or $CD44^{low}$/$EpCAM^{high}$) and tissue-origin (primary tumor, lymph node, lung or bone marrow). We focused on the five most prominent gene clusters (A-E) relevant to cancer progression for further analysis. Compared to the primary MFP tumor, the tissue-derived samples exhibited distinct upregulated and downregulated genes. Highly upregulated genes in MFP cells localized to cluster B. Lung-derived samples ($CD44^{low}$ and $CD44^{high}$) shared some similar transcriptomic changes (clusters D, E), but they also showed DEGs unique to their CD44 identity (clusters A, C). LN and BM samples trended similarly with MFP tumor cells, showing moderate expression of genes in cluster C.

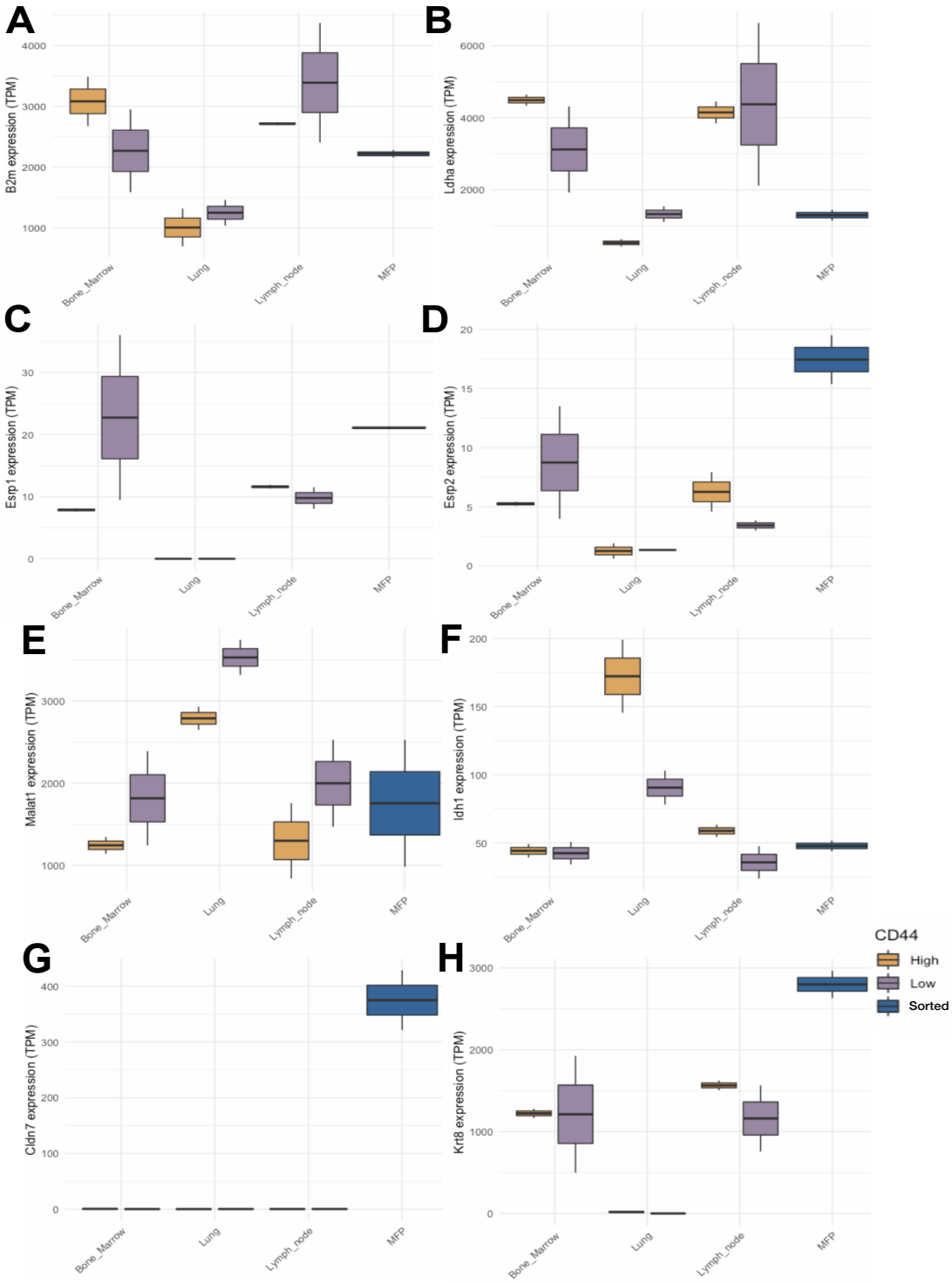


**Figure 3.3. Established breast cancer transcripts identified in MMTV-PyMT cell lines.** Bulk RNA sequencing analysis of bone marrow, lung, and lymph node-derived cell lines was performed and transcript levels for a panel of breast cancer markers were measured. For (A-H), expression of gene transcripts (TPM, transcript per million) for CD44 high expressing cells (yellow), CD44 low expressing cells (purple), and MFP CD44 expressing cells (blue) are shown.

To understand the biological relevance of the DEGs relevant to each cluster, we performed pathway enrichment analysis. Heat maps of the top 100 significant pathways revealed a multitude of cellular and molecular processes associated with cancer (Fig. 3.4). Pathways specifically relevant to cancer metastasis are shown in Fig. 3.1C, along with the corresponding enrichment score for each cluster. The upregulated genes for $CD44^{low}/EpCAM^{high}$ lung-derived cells in cluster C correlated with embryonic morphogenesis and hypoxia response pathways. Both of these pathways are critical to cancer cell growth in hostile environments [33, 34]. Some of the specific transcripts observed included those from the well-established cancer survival genes *ALDH1A1, SURVIVIN, XIAP, HSPG2, BCL9,* and *SOX4* [35, 36]. Interestingly, these same genes were downregulated in $CD44^{high}/EpCAM^{high}$ lung-derived cells. Cluster A was enriched in regulatory pathways associated with cell adhesion, correlating with the expression of *DDR1, HOXA7, MMP2, THBS1, TNFRSF14,* and *TGFB2* [37, 38]. Cluster A was also enriched in pathways associated with cellular locomotion, corroborated by the expression of *SERPINE1, PDGFA, ITGAV,* and *ITGB1BP1* [39, 40]. Both sets of upregulated genes were observed in $CD44^{high}/EpCAM^{high}$ lung-derived cells, but not MFP-derived or $CD44^{low}/EpCAM^{high}$ lung-derived cells. $CD44^{high}/EpCAM^{high}$ lung-derived cells also exhibited upregulated carbohydrate metabolism genes, pathways enriched in clusters A and E. Cluster E also correlated with other upregulated metabolic genes (*PFKFB3, SDC3,* and *GPC3*) in both lung-derived cell lines. These same genes were downregulated in the MFP cells [41, 42].

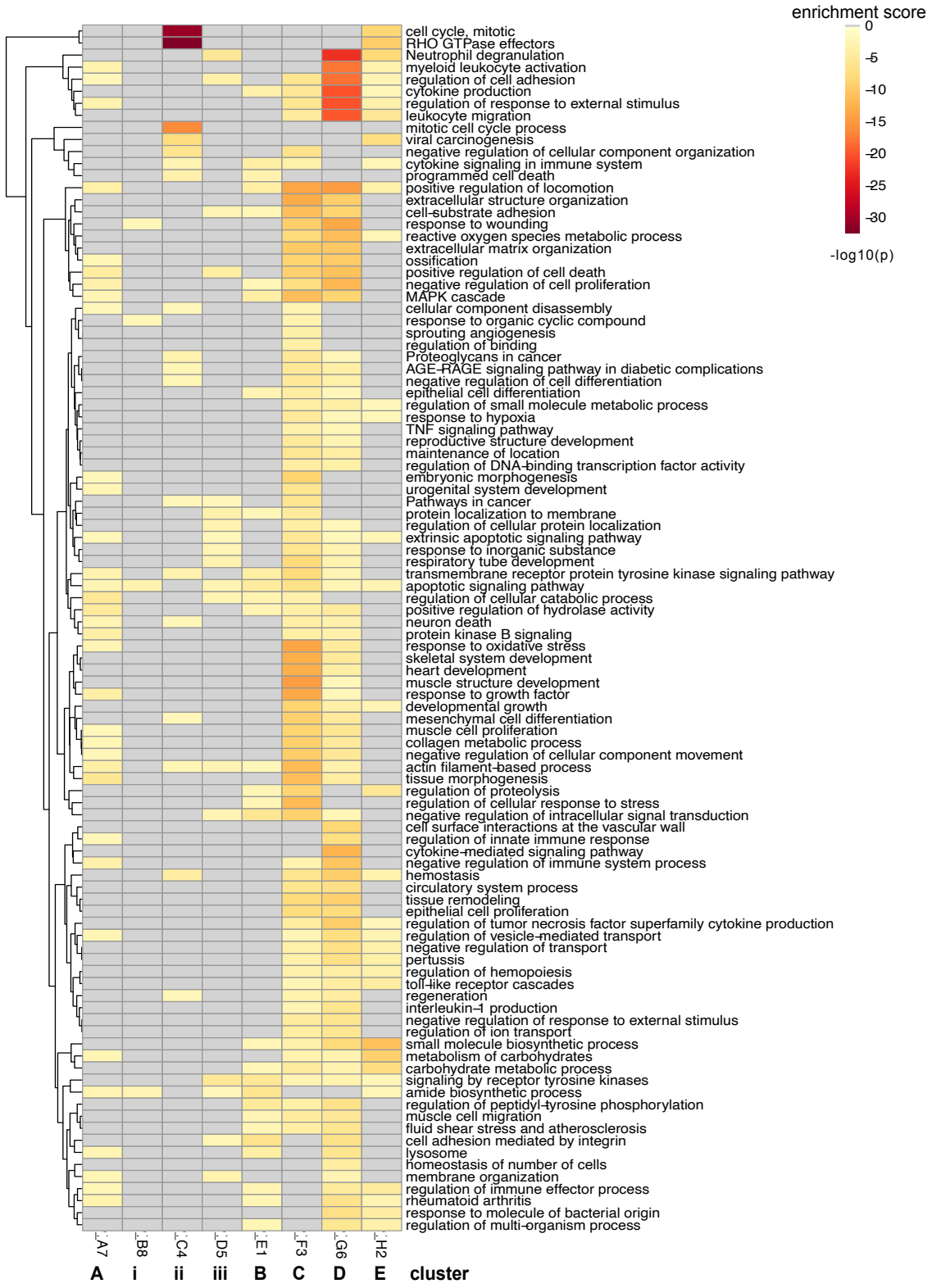


**Figure 3.4. Complete list of enriched GO-terms and pathway analysis from Figure 3.1D.** Heat map shows the enrichment scores of the pathways for each cluster (A-E; i-iii).

Metastatic breast cancer cells are known to preferentially colonize specific organs, a process known as organotropism [43]. The cross talk between metastatic cells and the distal microenvironment leads to the formation of the pre-metastatic niche, which can influence cancer cell homing [44, 45]. We examined whether the clonal isolates recapitulated features of organotropic metastases to lymph nodes, lung, and bone (Fig. 3.1D). Gratifyingly, we identified gene expression patterns that differed among the metastatic cell types based on their tissues of origin. LN-derived cells lines expressed genes relevant to metastatic lymphatic niches (e.g., *IRF5, YWHAH, PTGS2,* Fig. 3.1D) [46-51]. MFP-associated genes *CDH1* and *IGFBP3* [46-51] were also observed in the LN-derived lines, albeit to a lesser extent. In the case of the lung-derived cells, lung-tropic genes (e.g., *SPARC, MMP2, LTBP1, ID1*, and *CD151*) associated with high metastatic propensities [46-51] were upregulated. *EREG*, a marker expressed by cells in the mammary gland, [46-51] was downregulated in the lung-derived cells. BM-derived cells expressed genes associated with BM metastases, including *CEACAM1* and *LCN2* [46-51]. Expression of *CCL2, ADAMST1* and *CXCR4* were also observed in BM-derived cells albeit to a lesser extent. The expression of a select set of markers was further confirmed by Western blot analysis (Fig. 3.5). Collectively, these transcriptome changes could contribute to sub-clonal evolution during cancer progression across the different metastatic niches.

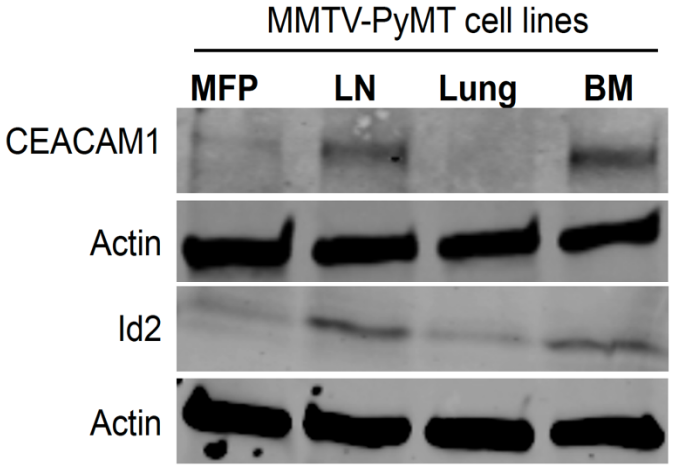


**Figure 3.5. Western blot analysis of markers identified from RNA-seq analysis.** Select tissue-tropic markers were selected from among the DEGs shown in Fig 1D. The following proteins were examined: Id2 (predicted to be highly expressed in lymph node-derived cells lines) and CEACAM1 (predicted to be highly expressed in bone marrow-derived cell lines). Actin was used as a loading control.

We also compared the gene expression changes among the metastatic cells and to those from the primary tumor. Overall, we observed that samples derived from organs further away from the primary tumors had greater numbers of DEGs, regardless of the CD44 designation. Lung-derived samples exhibited the most DEGs (4,411 genes in total) compared to cells derived from the lymph node (1753 genes in total) or bone marrow (2,985 genes in total, Fig. 3.6). Volcano plots of DEGs from each tissue-derived metastatic cell line ($CD44^{high/low}/EpCAM^{high}$) compared to the primary tumor sample revealed genes involved in metastatic progression (Fig. 3.7). Interestingly, some of the greatest differential expressions observed involved organotropism-associated genes (*MMP2* and *EREG*) identified in Fig. 3.1D.

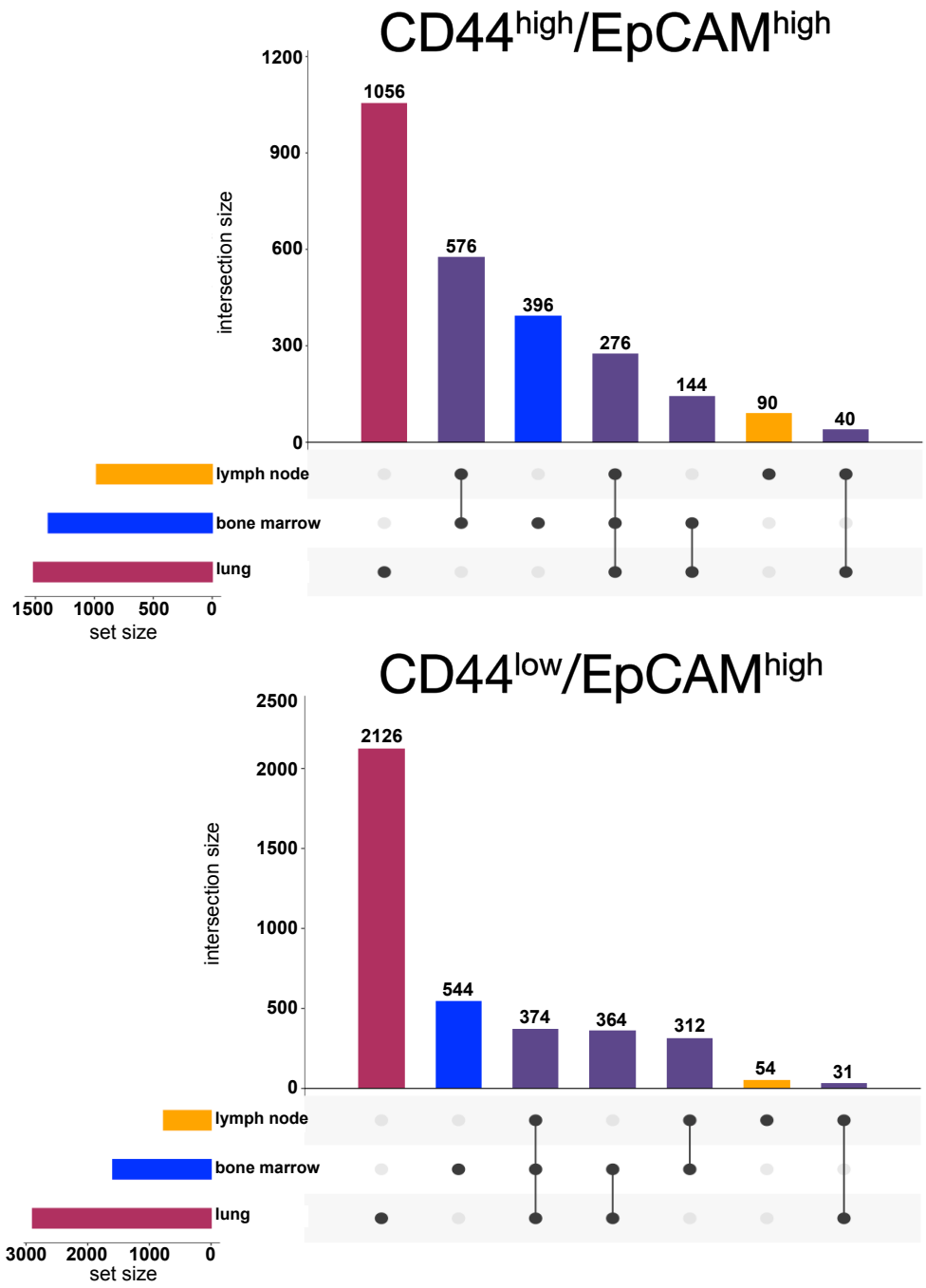


**Figure 3.6. Box and whisker plot comparing the number of upregulated genes in CD44$^{high}$/EpCAM$^{high}$ (top plot) and CD44$^{low}$/EpCAM$^{high}$ (bottom plot) expressing cells from the different metastatic tissues of origin.** Distinct gene expression signatures are shown with single black dots corresponding to a specific tissue-derived clonal isolate. Genes that are shared between different metastatic sites (purple bars) are represented by black lines that connect the specific samples. The total number of upregulated genes are denoted above each bar on the graph.

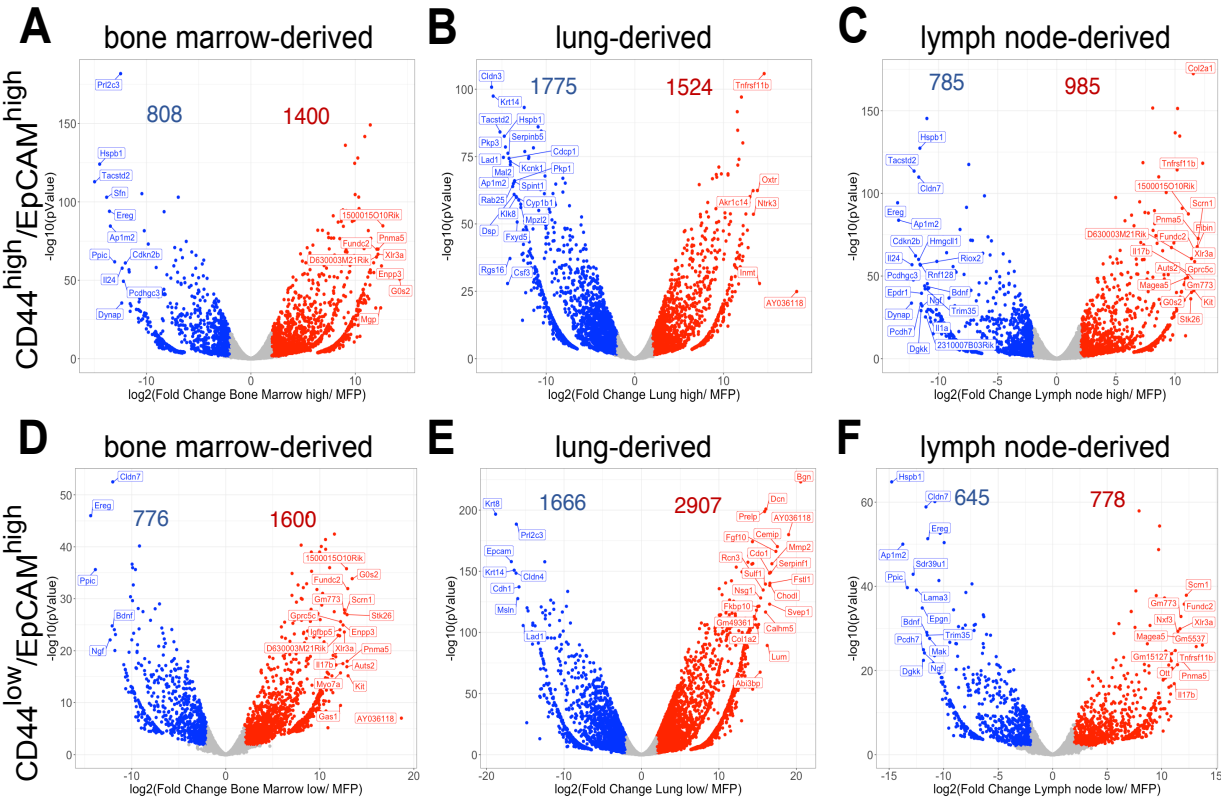


**Figure 3.7. Volcano plots of DEGs from tissue-derived metastatic cell lines.** (A-C) Volcano plots of DEGs from tissue-derived metastatic cell lines with CD44$^{high}$/EpCAM$^{high}$ expression compared to primary tumor samples. (A) Upregulated genes in bone marrow-derived (BM) isolates with CD44$^{high}$/EpCAM$^{high}$ expression are shown in red (1,400 genes), while 808 genes (blue) were upregulated in the primary tumor. (B) Upregulated genes in lung-derived isolates with CD44$^{high}$/EpCAM$^{high}$ expression are shown in red (1,524 genes), while 1,775 genes (blue) were upregulated in the primary tumor. (C) Upregulated genes in lymph node-derived (LN) isolates with CD44$^{high}$/EpCAM$^{high}$ expression are shown in red (985 genes), while 785 genes (blue) were upregulated in the primary tumor. (D-F) Volcano plots of differentially expressed genes from tissue-derived metastatic cell lines with CD44$^{low}$/EpCAM$^{high}$ expression when compared to primary tumor samples. (D) Upregulated genes in bone marrow-derived (BM) isolates with CD44$^{low}$/EpCAM$^{high}$ expression are shown in red (1,600 genes), while 776 genes (blue) were upregulated in the primary tumor. (E) Upregulated genes in lung-derived isolates with CD44$^{low}$/EpCAM$^{high}$ expression are shown in red (2,907 genes), while 1666 genes (blue) were upregulated in the primary tumor. (F) Upregulated genes in lymph node-derived (LN) isolates with CD44$^{low}$/EpCAM$^{high}$ expression are shown in red (778 genes), while 645 genes (blue) were upregulated in the primary tumor. The fewer differentially expressed genes amongst the tissue-derived isolates compared to the primary tumor is suggestive of the order in the metastatic cascade with LN-derived isolates bring the first metastatic site.

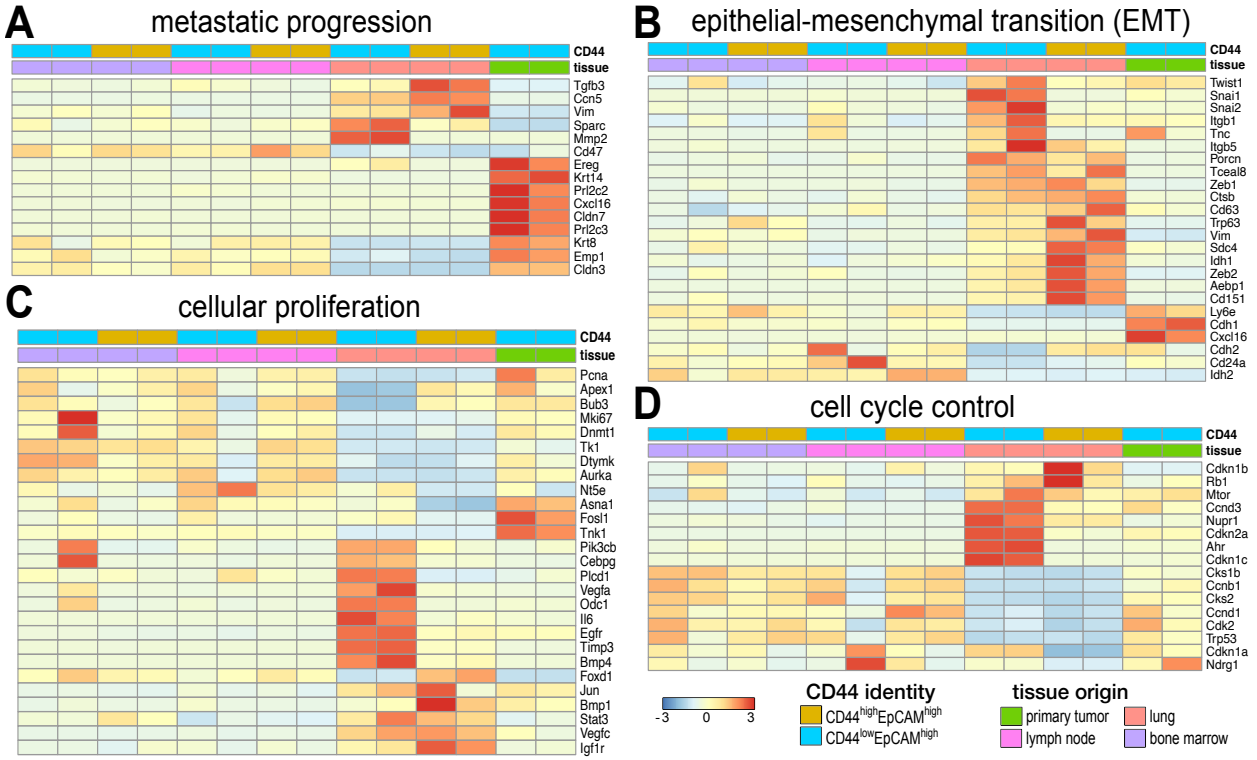


**Figure 3.8. Cancer cell lines exhibit distinct gene expression patterns relative to metastatic disease progression.** Bulk RNA analysis revealed differential gene expression patterns relevant to (A) metastatic progression, (B) epithelial-mesenchymal transition (EMT), (C) cellular proliferation, and (D) cell cycle control among the tissue-derived isolates and primary tumor. Columns were clustered based on CD44 expression and tissue origin as indicated. Select genes relevant to metastatic progression are displayed in the heat map.

We aimed to further characterize the metastatic cell lines via GO-term enrichment analysis. To this end, we examined gene expression changes relevant to metastatic progression, epithelial-to-mesenchymal transition (EMT), cellular proliferation and cell cycle control (Fig. 3.8). In the case of metastatic progression, we observed that MFP samples expressed high levels of classical markers associated with pre-metastatic lesions (e.g., *EREG, KRT14, CLDN7, KRT8, EMP1,* and *CLDN3,* Fig. 3.8A) [44, 45]. These markers were decreased in cells from distal metastatic sites (e.g., LN- and lung-derived cells). LN- and lung-derived cells, by contrast, exhibited upregulated levels of mesenchymal markers (e.g., *CCN5, ZEB1, VIM, SPARC* and *TGFB3*) [13, 52-55].

Expression levels were highest in $CD44^{high}/EpCAM^{high}$ lung-derived cells. Lung-derived cell lines also showed increased expression of well-established EMT markers (*SNAI1/2* and *CD63*) [52] and markers associated with poor prognosis in patients (*FOXC1*, *AEBP1, SDC4*, and *IDH1*, Fig. 3.8B) [48, 56]. Interestingly, *SNAI1/2* and *CD63* expression were highest in $CD44^{low}/EpCAM^{high}$ lung-derived cells, while the poor prognosis indicators listed were higher in $CD44^{high}/EpCAM^{high}$ lung-derived cells. The upregulation of mesenchymal markers and downregulation of epithelial markers in lung-derived cells is indicative of cellular de-differentiation[48, 56], suggesting that the lung-derived cells recapitulate EMT [57, 58].

EMT typically correlates with changes in cell proliferation and dysregulation of cell cycle control during cancer progression. These trends were apparent in the gene expression profiles for both $CD44^{high}$ and $CD44^{low}$ cells (Fig. 3.8C-D). As expected, the highly proliferative lung-derived $CD44^{high}$ cells expressed low levels of growth arrest genes *CDKN1A* and *CDKN2A* (Fig 3.8D) [33, 34]. Interestingly, we observed a stark difference in gene expression for lung-derived $CD44^{low}$ cells. Although these cells expressed genes relevant to cellular proliferation and angiogenesis, they exhibited upregulated levels of the growth arrest genes (Fig. 3.8D). Growth arrest signals could dampen the expression of other genes that are master regulators of downstream cell function. One such gene, mTOR, was expressed in the lung-derived $CD44^{low}$ cells. The levels of mTOR were comparable to expression in $CD44^{high}$ cells. From these results we postulate that lung $CD44^{low}$ cells, albeit capable of cell division, are not dividing as rapidly as their $CD44^{high}$ counterpart.

***Analyses of common biological pathways reveals intratumor heterogeneity***

MMTV-PyMT has recently been used as a model to study the impacts of CD44 on metastases [51]. Ex vivo analysis of tumors from a single micro-metastatic site revealed two subgroups of cells with differential CD44 expression. CD44 expression correlated with altered

gene expression relevant to EMT and MET and differential growth rates [48, 56]. Post-metastatic colonization, CD44 expression levels did not remain constant and were frequently switched between the subgroups. The fluid transition from EMT to MET phenotypes demonstrates how complex and context-dependent breast cancer can be. The morphological changes induced by CD44 expression also affected the tumor-initiating capabilities of the tumor cells. However, the critical regenerative CSC populations were found in both CD44-expressing groups, warranting further characterization. Similar spectrums of behavior have been documented in other studies [19] [51] [55].

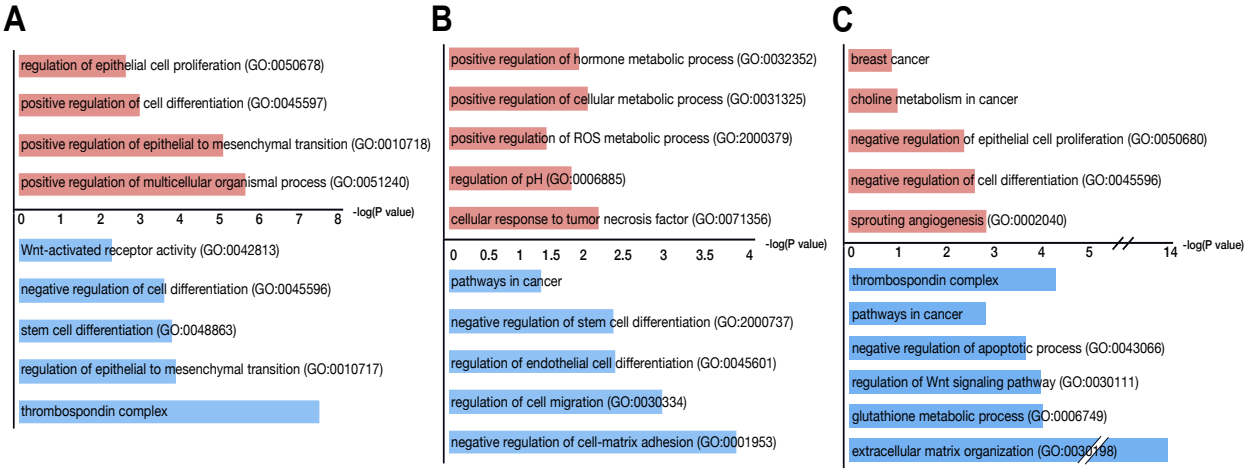


**Figure 3.9. CD44 expression correlates with different transcriptome patterns in organ-derived cell lines.** Differentially expressed genes for **(A)** $CD44^{high}$ (red) and $CD44^{low}$ (blue) $EpCAM^{high}$ cells (across all samples) were used to generate GO terms. $CD44^{low}$ cells exhibited high levels of expression for genes relevant to tumor microenvironment remodeling and tumor dormancy markers. $CD44^{high}$ cells exhibited higher levels of gene expression associated with GO terms related to cellular proliferation, tumor aggression, and EMT. Differentially expressed genes and associated GO terms for $CD44^{high/low}$ $EpCAM^{high}$ cells from (B) lymph nodes and (C) lungs are also shown.

To more closely examine the impacts of CD44 expression in our organ-derived cells, we analyzed DEGs based on CD44 levels. We identified upregulated genes for $CD44^{low}/EpCAM^{high}$ (374 genes) and $CD44^{high}/EpCAM^{high}$ (276 genes) signatures across the suite of cells. Some genes were shared across the tissue types. We also examined the GO terms and pathways associated with the DEGs from $CD44^{low}/EpCAM^{high}$ and $CD44^{high}/EpCAM^{high}$ samples. Clear differences were observed between the two CD44 signatures across the tissue-derived samples (Fig. 3.9A). Cells with $CD44^{high}$ signatures exhibited an increase in GO terms and associated genes related to cellular proliferation, tumor aggression, and EMT. Indeed, $CD44^{high}$ cells from lung and lymph node samples were experimentally observed to exhibit increased proliferation rates compared to $CD44^{low}$ cells (Fig. 3.10). DEG analysis further revealed that $CD44^{low}$ cells exhibited higher gene expression levels relevant to tumor microenvironment remodeling and stem cell markers. Interestingly, $CD44^{low}$ signatures also correlated with signaling pathways known to be important for stem cell maintenance and Wnt-activated receptor activity. $CD44^{low}$ signatures were further negatively correlated with cell differentiation pathways, supporting the idea of retained cellular dedifferentiation. The thrombospondin complex pathway, key in the maintenance of cancer stem cell dormancy in breast cancer, was also present in the CD44low signature [16, 57, 59].

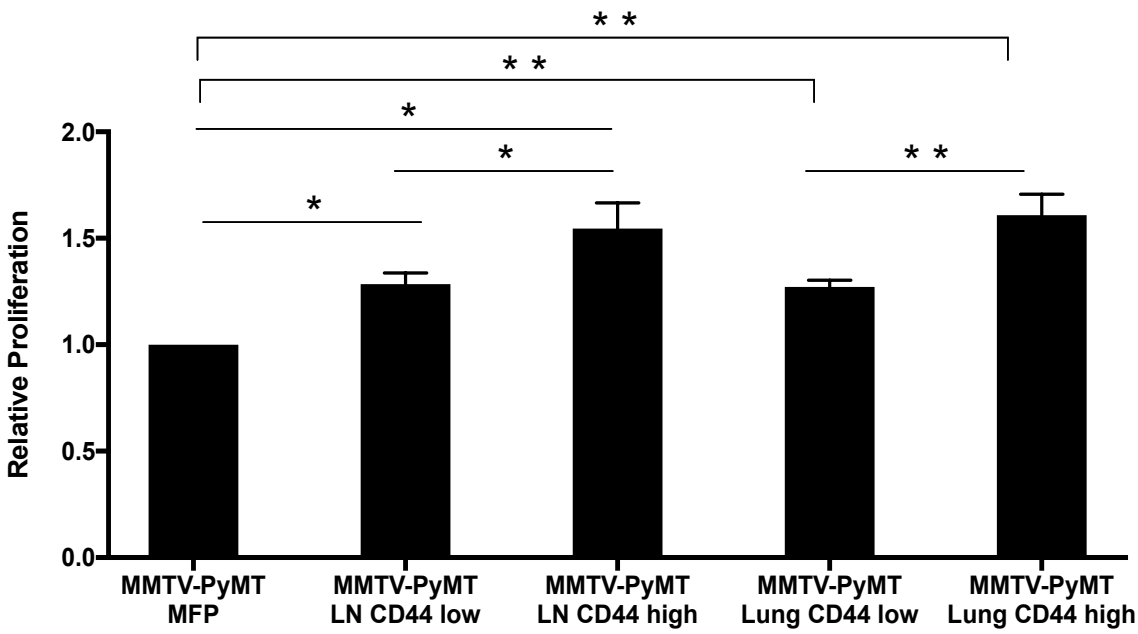


**Figure 3.10. Cellular proliferation analysis**. MMTV-PyMT cell growth was measured over 24 hours using a crystal violet assay. Relative proliferation values for lymph node (LN) and lung-derived cells (compared to MFP-derived cells) are shown. Error bars represent the standard deviation of the mean for n=3 replicates. ** $p < 0.01$; * $p < 0.05$

We further examined the intratumor heterogeneity of $CD44^{high}$ versus $CD44^{low}$ expression within single tumors. Volcano plots revealed DEGs in $CD44^{high}/EpCAM^{high}$ versus $CD44^{low}/EpCAM^{high}$ from lymph node-derived, lung-derived, and bone marrow-derived metastatic clonal isolates (Fig. 3.11A-C). The DEGs for these samples were also subjected to pathway enrichment analysis (Fig. 3.9B-C). DEGs upregulated in lymph node-derived $CD44^{high}/EpCAM^{high}$ cells correlated with cellular metabolism and pH regulation (Fig. 3.9B) observed in aggressive cancer phenotypes. Similar pathways were not observed in the corresponding $CD44^{low}/EpCAM^{high}$ lymph node-derived cells. The DEGs for these cells, by contrast, were enriched in pathways regulating stem cell differentiation, cellular migration and cell-matrix adhesion (Fig. 3.9B). For the lung-derived samples, the $CD44^{high}/EpCAM^{high}$ cells exhibited upregulated DEGS relevant to cancer metabolism, cellular proliferation and

angiogenesis (Fig. 3.9C). The DEGs for the corresponding CD44$^{low}$/ EpCAM$^{high}$ lung-derived cells were also enriched for pathways relevant to cancer progression and cellular metabolism, in addition to the thrombospondin complex and extracellular communication (Fig. 3.9C). Similar analyses were performed with bone marrow-derived cells to reveal unique GO terms and transcriptomic changes (Fig. 3.11D-E). Upregulated DEGs for BM-derived CD44$^{low}$/ EpCAM$^{high}$ cells were mainly associated with immune and cytokine activity.

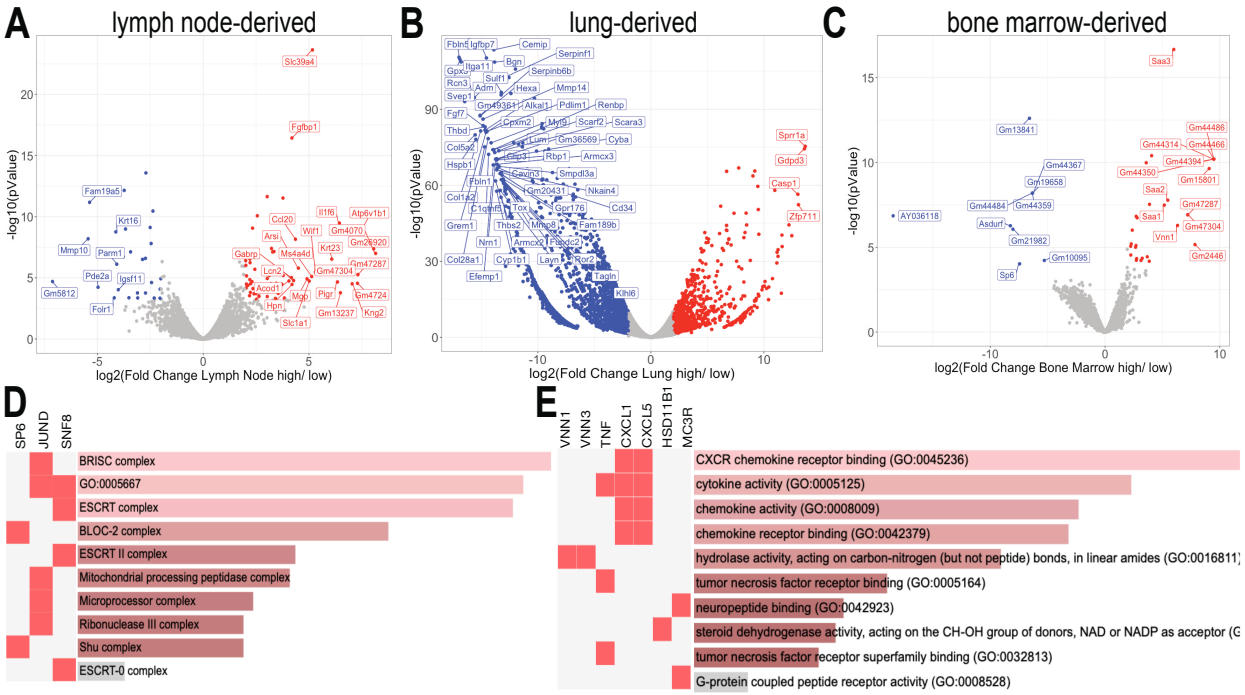


**Figure 3.11. Analysis of DEGs from metastatic isolates based on CD44 expression.** (A) Volcano plots showing DEGs in CD44$^{high}$/EpCAM$^{high}$ (red) versus CD44$^{low}$/EpCAM$^{high}$ (blue) from (A) lymph node-derived, (B) lung-derived, and (C) bone marrow-derived metastatic cells. (D) GO terms and relevant genes upregulated in CD44$^{high}$/EpCAM$^{high}$ bone marrow-derived cells. (E) GO terms and relevant genes upregulated in CD44$^{low}$/EpCAM$^{high}$ bone marrow-derived cells.

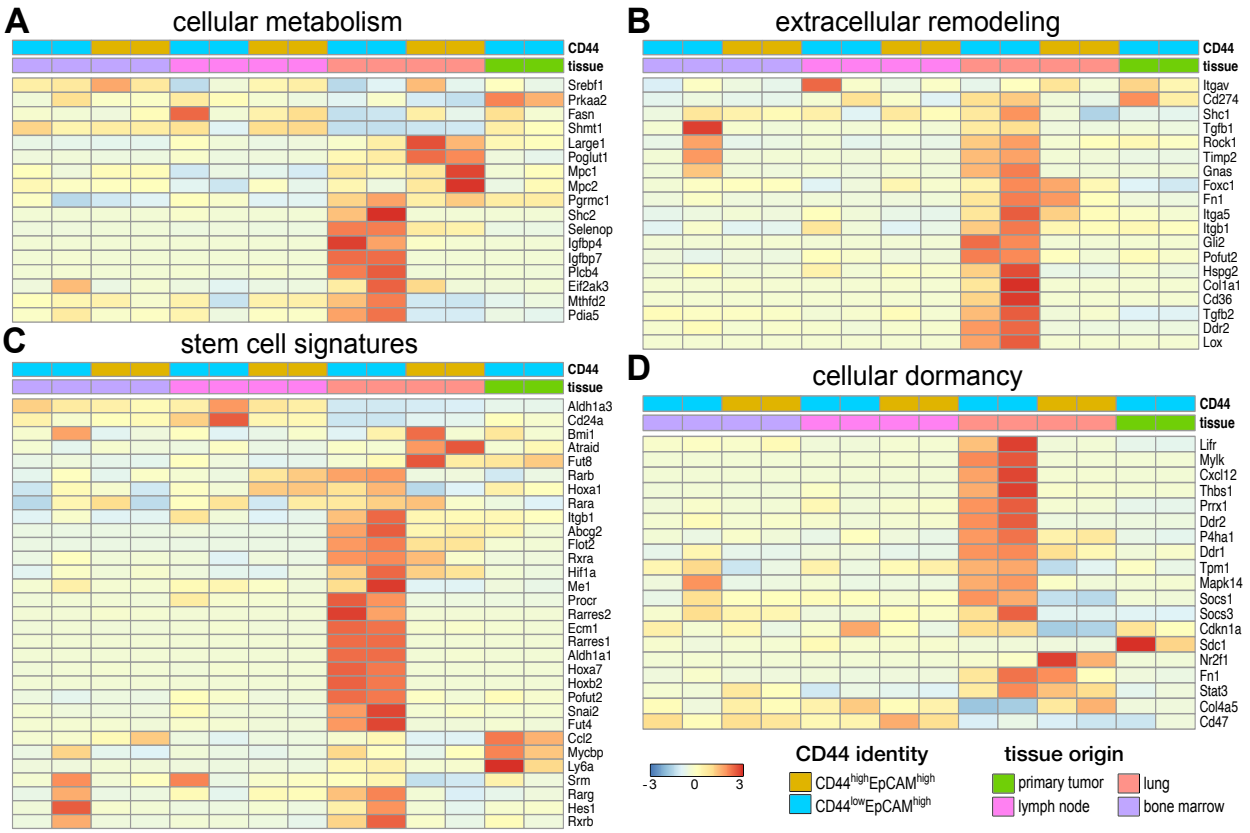


**Figure 3.12. Intratumoral heterogeneity observed across organ-derived cell lines.** Bulk RNA analysis of the organ-derived cell lines revealed distinct expression patterns relevant to (A) cellular metabolism, (B) extracellular remodeling of the microenvironment, (C) stem cell signatures, and (D) cellular dormancy among the cell lines. Clusters were assigned based on CD44 expression and the metastatic origin of each cell line as shown in the lower right.

Based on our DEG analyses, we examined additional known markers of breast cancer metabolism and extracellular remodeling across the entire set of organ-derived cell lines [60, 61]. We observed increased levels of cellular metabolism markers (e.g., *PLCB4, IGFBP7, IGFBP4, SHC2, PGRAMC1, MTHFD2*) specific to lung-derived $CD44^{low}$ cells (Fig. 3.12A). Lung-derived $CD44^{high}$ cells exhibited higher levels of *MPC1, MPC2, POGLUT1*, and *LARGE1* expression. Interestingly, we did not observe upregulation of other cancer-related drivers of energy consumption in either of the lung-derived cell lines compared to MFP-derived samples [62]. Extracellular remodeling has been shown to improve cancer colonization and perpetuate

dedifferentiated stem-like cellular states [57, 58]. We identified markers relevant to extracellular remodeling known to promote the survival of metastatic lesions (*CD36, CD274, FOXC1*) in the lung-derived metastatic cells (Fig. 3.12B). The expression levels were noticeably enhanced in these samples compared to the MFP tumor. Lung-derived CD44$^{low}$ cells also exhibited upregulated levels of genes associated with mesenchymal cells and more dedifferentiated phenotypes in advanced cancers (e.g., *FN1* and *POFUT2*) [13].

***CD44$^{low}$ cell lines exhibit classic signatures of stem cells***

As noted earlier, the lung-derived CD44$^{low}$ cells exhibited reduced levels of some markers of cellular proliferation and division (Fig 3.8C,D). These cells also expressed genes known to be important for CSC survival and function (Fig. 3.12B) [63]. Changes in gene expression relevant to matrix remodeling have been known to sustain CSCs in a functional, but dormant non-dividing state. To further examine whether the lung-derived CD44$^{low}$ cells harbored CSC properties, we evaluated a panel of known breast cancer stem cell markers across our suite of metastatic isolates. Comparisons were made to the known cellular differentiation marker CD24 [64, 65]. As shown in Fig. 3.12C, lung-derived CD44$^{low}$/EpCAM$^{high}$ cells exhibited an increase in the stem cell-associated markers and a decrease in CD24 expression. Furthermore, expression of *ALDH1A1*, a breast cancer-specific stem cell marker associated with resistance to some chemotherapies, was elevated [64, 65]. CSCs can endure some drug treatments and survive in metastatic environments due, in part, to their ability to modulate their metabolism and compensate for oxidative stress [63, 64]. The lung-derived CD44$^{low}$ cells exhibited gene expression profiles consistent with these phenotypes, including the upregulation of retinoic acid (RA) pathway (*RARA/B/G, RXRA/B, RARRES1/2*) essential for cell survival. Opposite trends were observed for lung-derived CD44$^{high}$

cells. These cells expressed higher levels of genes associated with cell growth and aggressive metastases.

We further examined the entire suite of cells lines, for markers of breast cancer dormancy (Fig. 3.12D). The lung-derived $CD44^{low}$ cells expressed higher levels of genes associated with the thrombospondin complex, a well-known dormancy marker in breast cancer [16, 66]. Additional markers, including *MAPK14*, *DDR1/2,* and *MYLK* were also observed among this cell population. Collectively, these data suggest that lung-derived $CD44^{low}$ cells express CSC-relevant genes that can maintain cells in a dormant or low proliferative state. CSC identification remains challenging, though, owing to difficulties in isolation [67-69].

***Single cell RNA-seq reveals distinct clusters relevant to metastatic progression and intratumor heterogeneity***

To further validate that our isolated cell lines capture the intratumoral heterogeneity observed during *de novo* disease progression, we performed scRNA-seq on a subset of metastatic samples. Based on the amount of differential gene expression observed in bulk RNA-seq, we chose lung-derived $CD44^{low}$ (lungL), lung-derived $CD44^{high}$ (lungH), and LN-derived $CD44^{high}$ (lymphH) cells for the analyses. In all, 4124 cells were used in the clustering analyses: lungL (334 cells), lungH (1085 cells), lymphH (1694 cells). Clusters were visualized using UMAP (Fig. 3.13A). Clusters for each of the three cell types—lungL (green), lungH (pink), and lymphH (blue)—were identified.

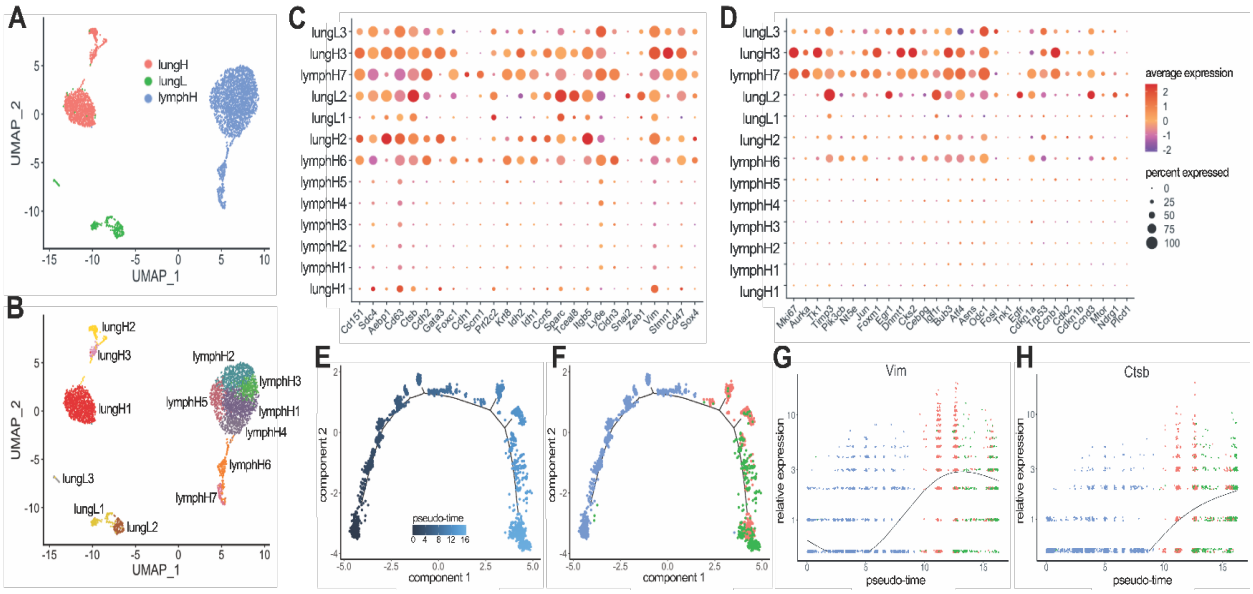


**Figure 3.13. Single-cell RNA-seq revealed tissue-specific clusters and heterogeneity during metastatic disease progression.** (A) Clustering of 4,124 cells that passed filtering. Tissue-specific clustering was observed for lungH, lungL, and lymphH cells. (B) Thirteen clusters were recovered based on a combination of tissue origin and CD44 signature. (C, D) Gene expression profiles for select cancer relevant genes (columns) relevant to (C) EMT and (D) cell proliferation among the thirteen clusters (rows). The size of each dot represents the percentage a specific gene is expressed compared to all other transcripts. The color gradient of the dot indicates the average expression of the gene. (E) Monocle2 pseudo-time analysis was performed and the metastatic trajectory of distinct cell clusters are shown. The color gradient indicates pseudo-time progression. (F) Cell tissues of origin indicated on the pseudo-time map (as in A). Changes in EMT-relevant genes from C were probed across pseudo-time and tissue of origin. (G) The expression of *VIM*, an EMT marker, increased over lymph node populations with maximum expression in lung high populations. This analysis was repeated for (F) *CTSB*, a marker of invasion, where the low expression across lymph node populations increased in expression across lung high and lung low populations.

We further examined the heterogeneity of the samples using hierarchical clustering (Fig. 3.14). The UMAP plot depicted 13 distinct cell clusters from the three respective cell samples (Fig. 3.13B). LymphH cells had the largest number of subclusters (7), with clusters 1-5 being very similar in composition (Fig. 3.14). Clusters 6 and 7 (lymphH6 and lymphH7) exhibited higher levels of pro-survival/cell cycle regulation genes (*BIRC5, TOP2A, CENPF*) associated with metabolically active cancer pathways. We observed the lung CD44$^{high}$ (lungH) cells divided into

three different clusters. The majority were localized to cluster lungH1 and expressed lower levels of metastatic cancer cell markers (*ALDH3A1, VIM, TCEAL9*). Similar metastatic cancer markers were upregulated in cluster lungH2. This subpopulation further displayed upregulated markers associated with the microenvironment and EMT (*AEBP1, ITGB5, FN1*).


**Figure 3.14. Hierarchical clustering analysis from single cell sequencing analyses**. Single cell sequencing heat map showing 13 tissue-specific clusters.

### *EMT and proliferation markers show distinct tissue-specific gene expression*

We identified transcriptomic changes relevant to cancer metastasis that drove the designation of each cluster. Guided by the bulk RNA-seq results, we examined the expression of EMT and metastatic progression markers (Fig. 3.13C). We used dot plots to visualize the average expression of each gene and the percentage it was expressed in the sample set. As expected for

progressive disease, we observed genes associated with primary tumors and low metastatic lesions (*CLDN3, KRT8, CDH1, CDH2*) primarily in the lymphH populations. Expression of these genes diminished at more distal sites (lung-derived samples). Conversely we observed genes associated with aggressive metastatic disease and EMT (*VIM, SPARC, ZEB1, SNAI2, CTSB*) upregulated in lung populations compared to lymph nodes. Changes in the expression of breast cancer marker CD63 were also observed during disease progression. Interestingly we observed similar gene expression changes for EMT in clusters lymphH6, lymphH7, and lungH3.

To further examine the changes in gene expression, we performed pseudo-time analysis of the single cells using Monocle2. The metastatic trajectory of the cells showed distinct clusters across the pseudo-time (Fig. 3.13E). We identified the composition of the cells by coloring the pseudo-time map with the tissue of origins (Fig. 3.13F). We observed that the majority of cells at the beginning of the pseudo-time (0) are lymphH-derived with a few lungL cells. Lung-derived cells became prominent further on the graph around pseudo-time 8. Interestingly, we observed lungL cells present across pseudo-time that cluster heavily towards the end of the graph (pseudo-time 12-16). We probed for changes in *VIM* and other EMT-relevant genes across pseudo-time and tissue of origin. As we previously observed in Fig. 3.13A, *VIM* expression increases over lymph node populations (Fig. 3.13G). *VIM* expression is highest in lung high populations and drops down in lung low populations. *CTSB*, a marker of invasion, showed relatively low expression across lymph nodes populations (Fig. 3.13H), but higher expression along pseudo-time in lungH and lungL populations.

Cellular proliferation markers were also analyzed via scRNA-seq (Fig. 3.13D). Increases in proliferative markers (*AURKA, MKI67*) were observed for clusters lymphH6 and lymphH7, with the greatest expression in lungH3. We also observed many similarities between clusters lymphH7

and lungH3 with regard to EMT- and proliferation-associated gene expression. These correlations could potentially signify the metastatic progression of the disease from the lymph node (lymphH7) to lung (lungH3). LungL clusters, by contrast, expressed lower levels of genes involved in proliferation. These clusters expressed higher levels of genes involved in cell cycle control and growth inhibition (*CDKN1A, CDKN1b, CCND3*). We further probed for changes in a panel of proliferation-associated genes across pseudo-time and tissue of origin using Monocle2 (Fig. 3.15).

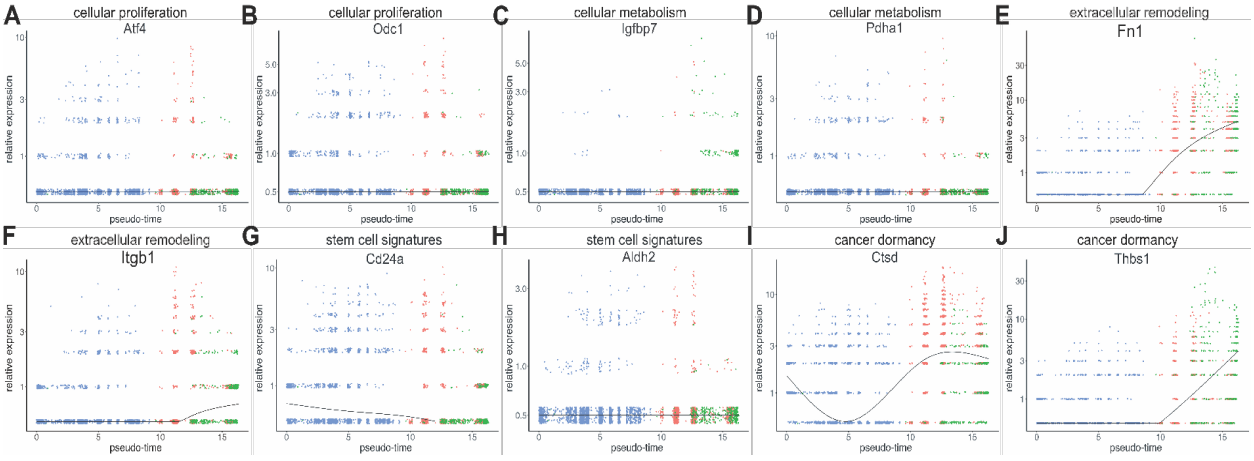


**Figure 3.15. Monocle2 pseudo-time analysis of single cells.** The metastatic trajectory of distinct cells clusters is shown. The composition of the cells was identified by coloring the pseudo-time map with the tissue of origins (as in Fig. 3.13A). Expression of cellular proliferation markers (A, B), metabolism markers (C, D), extracellular remodeling markers (E, F), stem cell signatures (G, H), and cancer dormancy markers (I, J) of single cells across pseudo-time. Trend line on graphs tracks the statistical significance of gene expression as it changes across pseudo-time.

***Cellular metabolism and extracellular remodeling markers show distinct tissue-specific changes***

We identified gene expression markers relevant to cellular metabolism and extracellular remodeling that drove the formation of each cluster. For example, lungH3 cells expressed upregulated levels of genes associated with cell cycle and intracellular metabolism (*TOP2A, CENPF, HTRA1, STMN1*) (Fig. 3.14). These data suggested that cells within the lungH3 cluster exhibit the highest metastatic propensity of the lung subsets. We further observed an increase in insulin-like growth factors (*IGBFP4, IGBFP7*) across different lung-derived clusters, specifically in the lungL2 cluster (Fig. 3.16A). Interestingly, the expression of *PDHA1*, a critical component for pyruvate to acetyl-CoA conversion, was primarily upregulated in lungL2, lymphH7, lymphH6. We probed for changes in these metabolism-related genes across pseudo-time and tissue of origin using Monocle2. As we previously observed in Fig. 3.16A, *IGBFP7* expression was significantly upregulated in lungL single cell populations as pseudo-time progressed (Fig. 3.16B).

Guided by bulk RNA-seq results, we further examined extracellular remodeling markers using scRNA-seq (Fig. 3.16C and Fig. 3.14). We observed the most dynamic expression of extracellular remodeling-associated genes in the three lungL clusters. The lungL1 cluster had distinct differences in gene expression relevant to extracellular matrix interactions (*MGP, BGN, CCN2, FN1, ITGB1*). LungL2 cells expressed similar genes, along with upregulated genes associated with immunosuppressive proteins (*SLPI*) and embryonic glandular hormone (*PRL2C3*). However, cells in the lungL3 cluster lacked significant expression of extracellular remodeling genes that were present in lungL1 and lungL2. We investigated changes in extracellular remodeling-relevant genes (*FN1, ITGB1*) across pseudo-time and tissue of origin using Monocle2 (Fig. 3.16D). *FN1* had relatively low expression across lymph nodes populations, but increased expression was observed in lungH and lungL cells populations starting at pseudo-time 10. *ITGB1*

was prominently expressed across most lungH, lungL, and some lymphH cell clusters (Fig. 3.16C). However, pseudo-time analysis of the single cells only attributed a significant upregulation of *ITGB1* expression in lungL cells at the end of the pseudo-time (Fig. 3.16D).

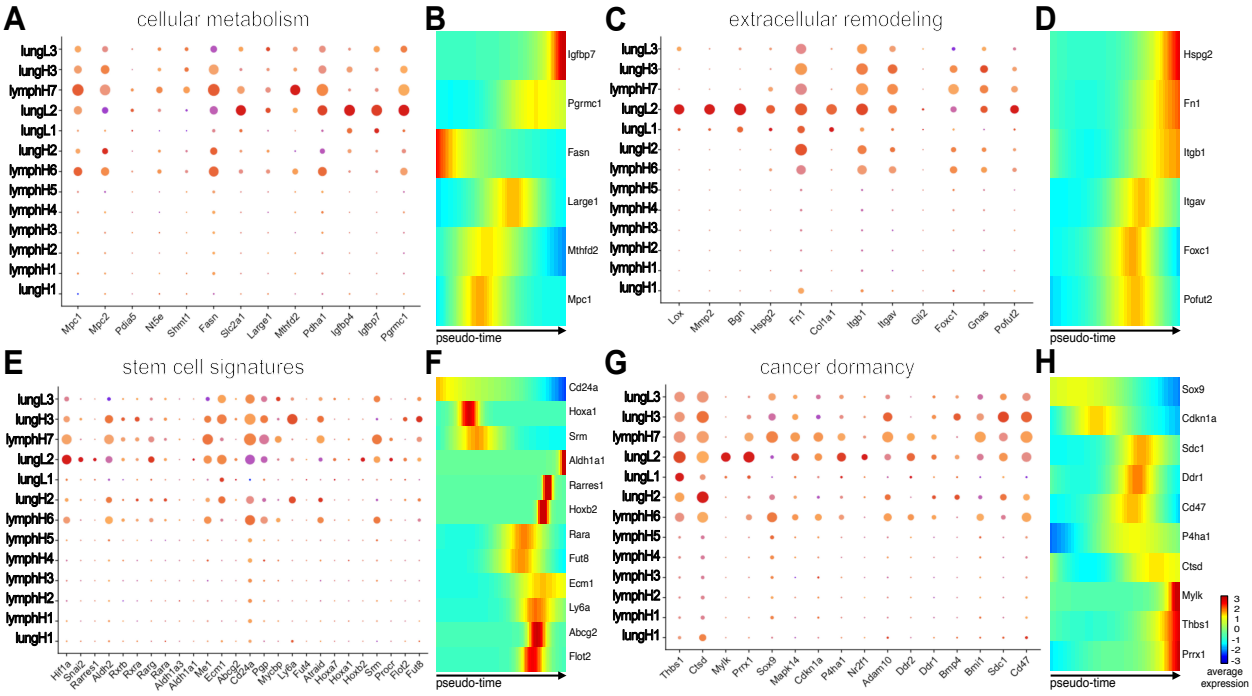


**Figure 3.16. Single cell profiling revealed tissue-specific changes**. (A) Dot plot analysis of (A) cancer metabolic markers relevant to cancer progression for select clusters guided by the bulk RNA seq data. (B) Heat maps of differentially expressed genes relevant to cancer metabolism were identified across the pseudo-time using Monocle2. The color gradient indicates the average expression across the pseudo-time, trending from dark blue to red. Similar dot plot analyses and pseudo-time heat maps are shown for markers relevant to (C-D) tumor microenvironment remodeling, (E-F) stem cell signatures, and (G-H) cell dormancy.

### *CD44$^{low}$ lung-derived cell lines harbor markers related to cancer stem cells and tumor dormancy*

Although CD44 expression has traditionally been used as a marker of metastatic disease, recent publications have demonstrated that cellular expression of CD44 fluctuates during cancer

progression [51]. This has prompted other markers to be used in parallel with CD44 when examining metastatic potency and cancer-initiation capability of cells. Toward this end, we used scRNA-seq to probe for markers of CSC and tumor-initiation in the breast cancer model (Fig. 3.16E). We observed an increase in stem cell associated genes (*ALDH1A1, ECM1, ME1, ABCG2, and SNAI2*) with a subsequent decrease in differentiation marker, *CD24A*, in lungL1 and lungL2 clusters. Interestingly, lungL2 had the greatest upregulation in *HIF1A*, a known marker of hypoxia and poor prognosis in breast cancer. We previously observed that cells in lungL3 cluster separate from the other two lungL clusters (Fig. 3.13B). These cells have unique changes in gene expression relevant to embryonic and cancer stem cells (*KDEL1*, *PRL2C3*, *EBP1*) (Fig. 3.14). Similar to clusters lungL2 and lungL3, stem cell genes *ALDH1A1*, *ABCG2*, and *SNAI2* were upregulated in lungL3 cells. However, other stem cell associated genes (*ME1*, *HIF1A*) were downregulated. Additionally lungL3 displayed an increase in *CD24A*, a marker of cellular proliferation and decreased stemness. Similar to the bulk RNA sequencing data, we saw an increase in *ALDH1A1* expression and stem cell survival metabolomic markers from the RA family in lungL2 and lungL3 clusters that were not present in lungL1 cells. In contrast, $CD44^{high}$ clusters lungH3, lungH2, lymphH6, and lymphH7 exhibited an increase in genes regulating growth and invasion (*ECM1*, *CD24*, *FUT8*). While these $CD44^{high}$ clusters showed a downregulation in stem cell markers *ALDH1A1* and *ABCG2*, they did show an increase in *ALDH2*, a stem cell marker less common in breast cancer.

We examined the changes in genes associated with cellular differentiation and stem cell capabilities (*CD24A*, *ALDH2*) using Monocle2 (Fig. 3.16F). *CD24A* expression was upregulated in lymph node populations, then downregulated as pseudo-time moved into the lung populations.

Upregulation of stem cell marker *ALDH2* was observed across a select amount of cells for all cell types across pseudo-time (Fig. 3.15)

Cancer dormancy is a formidable obstacle in breast cancer research and treatment [70]. Previous work has shown that some dormant breast cancer cells have similar genomic profiles as CSC [29, 71]. We examined whether our cell lines expressed markers of cancer dormancy and cancer stem cell-associated markers. Similarly to the bulk RNA-seq analyses, we saw an increase in genes regulating cell proliferation and increasing tumor dormancy in the lungL2, and in lungL1 to a lesser extent (Fig. 3.16G). This pattern was not sustained in the lungL3 subcluster. Instead, lungL3 showed a decrease in dormancy-associated genes *FN1*, *CD47,* and *THBS1*. Interestingly lungL3, lungL2 and lungH3 showed increased expression of breast cancer dormancy cell-associated maker *SDC1* [72, 73]. Master regulator of morphogenesis, *SOX9* [74], was upregulated in lungL3, lymphH7, and lymphH6 subclusters. We further probed the changes in cancer dormancy-relevant genes *CTSD*, *THBS1* across pseudo-time and tissue of origin using Monocle2. *CTSD* expression fluctuated across lymph node populations (Fig. 3.16H). The expression of *CTSD* was upregulated through lungH populations and plateaued in the lungL populations across pseudo-time. *THBS1* exponentially upregulated expression at pseudo-time 10 in lung high and lung low populations (Fig. 3.15).

## III. Discussion

Breast cancer comprises cancer cell subpopulations that are genetically and biologically different [75]. Although such intratumoral heterogeneity is critical to understanding the disease, traditional cancer cell lines models have difficulties recapitulating the complexity [19, 76]. In this study, we established a novel suite of organ-derived metastatic cell lines and subsequently

performed a comprehensive transcriptome analysis of cancer progression in relation to CD44 levels. Key pathways relevant to metastatic disease were found to be upregulated in the various cell types, and the main takeaways are diagrammed in Fig. 3.17.

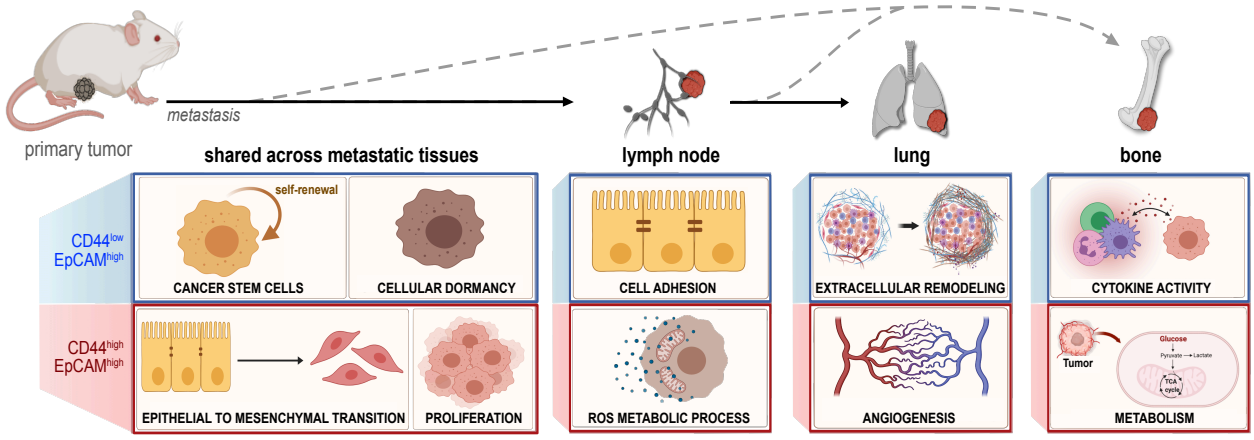


**Figure 3.17. Summary of key findings from transcriptomic analysis of MMTV-PyMT-derived cell lines.**

CD44 expression has been extensively shown to impact breast cancer progression, controlling cellular biology and correlating with certain clinical outcomes [4-6,17]. However, the exact roles of CD44 remain unclear in breast cancer as both high and low levels of the marker are correlated with tumor-promoting and tumor-suppressing outcomes [67-69,76]. In this study, we used metastatic organ-derived cell lines to investigate the effects CD44 expression on cancer progression across different metastatic niches. As previously shown in 4T1 cell lines [77], models of metastatic breast cancer are invaluable for the advancement of the field. The MMTV-PyMT mouse model of breast cancer is the most popular transgenic preclinical system to study mammary tumor progression and metastatic disease translatable to patients [78, 79]. Here, we successfully

isolated primary cell lines from four different organ-derived tissues. The cell lines were isolated based on their expression of $CD44^{low}$/$EpCAM^{high}$ or $CD44^{high}$/$EpCAM^{high}$ to examine the role of CD44 in breast cancer across the various metastatic niches (Fig. 3.17). The cells were expanded to use as a reproducible resource for experiments in this study and beyond. It is possible, though, that the culturing conditions altered some genomic pathways and/or skewed the levels of $CD44^{low}$ or $CD44^{high}$ cells in the samples.

The suite of MMTV-PyMT derived cell lines was subjected to comprehensive transcriptome analysis using bulk RNA-seq and scRNA-seq. Although there were shared GO terms across all cell lines belonging to either CD44 signature, we identified numerous changes in gene expression that were organ-specific and promoted metastatic homing [48]. Changes in gene expression related to EMT and the metastatic cascade were also identified with similar expression across both transcriptome pathways. As expected distal metastatic-derived tissues such as the lung and BM had more differential genes expressed than lymph node-derived cell lines compared to the primary tumor.

Using scRNA-seq on select organ-derived metastatic cell lines, we found that the cell lines exhibited a range of intratumoral heterogeneity. Although most MMTV-PyMT publications have focused on metastasis from the primary tumor to the lungs, we identified additional distal metastatic cells in the BM. The BM-derived cell lines were characterized and showed an increase in previously published markers such as *RANKL*, *OPN* (*SPP1*), and *IL2* [4]. Further studies are necessary to better characterize these cells and increase the number of validated BM-derived DEGs (Fig. 3.11C). Specifically, it would be prudent to examine the differential gene expression of this model compared to other well-established bone-marrow metastatic models of tumor latency and

cancer dormancy [4]. Collectively, these transcriptome changes could contribute to sub-clonal evolution during cancer progression across the different metastatic niches.

Although CD44 has traditionally been used to monitor cancer metastasis, recent studies from Gao and coworkers showed that its expression fluctuates during metastatic progression [51]. We observed similar changes in CD44 expression from initial isolates compared to cells used in subsequent transcriptome analysis. Furthermore, recent studies with MDA-MB-231 breast cancer models showed that $CD44^{low}$ cells possessed stem cell like properties that $CD44^{high}$ lacked [13, 55]. $CD44^{low}$ populations could regenerate both more $CD44^{low}$ and $CD44^{high}$ cell clusters, where $CD44^{high}$ could only replicate themselves. We observed a similar increase in stem cell markers in $CD44^{low}$ but not $CD44^{high}$ cells. These contradictory results suggest that more research is needed to better understand and characterize CD44 expression in breast cancer.

Although CD44 has been correlated to metastatic progression and CSC, the fluctuating expression levels of this marker complicates its use as a sole classifier of cellular phenotypes. There are thus many combinations of markers currently used to isolate CSC, with no one set being universally accepted [13, 53-55]. We examined a handful of CSC markers commonly associated with breast cancer (*ALDH1A1, ABCG2, PGP, FUT4*). CSC associated markers were found in the both bulk RNA-seq and scRNA-seq data. The expression of the CSC-associated markers was highest in lung-derived $CD44^{low}$ cells compared to $CD44^{high}$ cells. Upregulation of other CSC-associated markers, including the stem cell survival signaling member WNT, was also observed in the CD44low cells, corroborating our hypothesis that these cells possess stem-like properties [17]. scRNA-seq further revealed three distinct clusters for lung-derived $CD44^{low}$ cells on the UMAP. Unlike the lung-derived $CD44^{high}$ groups, lungL1 and lungL2 retained populations of stem cells but dramatically increased their expression of mesenchymal-like signaling pathways involved

in cancer cell maintenance and dormancy (*TSP, TNC, BMP*) [6, 14, 16]. Interestingly, lungL3 was a much smaller, separate cluster that lacked the dormancy markers and instead expressed high levels of stem cell markers. The diversity of $CD44^{low}$ clusters could be explained by heterogeneity within the different stages of metastatic progression and along the EMT spectrum. Another possibility is that lungL1 and lungL2 clusters might be cancer associated fibroblasts while the lungL3 might be a distinct cluster of CSC cells.

Breast cancer progression relies on CSC, to initiate tumor cell growth during metastatic dissemination and colonization of new organs. Previous work from Weinberg has shown that tumor initiating cells (TICs) originating in the luminal cell layer of the mammary gland rely on EMT initiating transcription factors (TFs) for cellular dedifferentiation [52]. These factors activate signaling pathways distinct from TICs originating from the basal mammary gland. Bulk RNA-seq data allowed us to further characterize the lung-derived cell lines based on expression of EMT-TFs known to induce TICs. For example, increased expression of the EMT-TF Snail was observed in lung $CD44^{low}$ cells compared to $CD44^{high}$ cells. Although both CD44 signatures of lung-derived cells expressed the EMT-TF binding promoter and TIC master regulator *ZEB1*, lung-derived $CD44^{low}$ cells also showed a greater increase in the basal cell associated EMT-TF Slug as compared to $CD44^{high}$ cells. Weinberg and others have shown that Slug and Snail both bind and regulate *ZEB1* expression [52];however, the expression of *SLUG* is associated breast cancers that arise from cells containing normal mammary epithelial stem cells in the basal compartment *SLUG* expression is also associated with highly dedifferentiated breast cancer cells found in the advanced and final stages of metastatic disease. Here, we harvested MMTV-PyMT mice well after the time period sampled by Weinberg. The most striking result from the lung-derived cells was that the

stem cell associated features of tumor-initiating cells were observed most in $CD44^{low}$ opposed to $CD44^{high}$ expressing cells that have previously been correlated to CSC.

Further work examining the effects of the microenvironment as a key regulator of cellular plasticity adds to the theory that "dedifferentiated" non-CSCs can undergo processes that endow them with CSCs-like properties [13, 33]. These complex studies highlight the need for further investigation into the possible origins of CSCs in relations to the surrounding microenvironment. Modification of cellular metabolism along with remodeling of the tumor microenvironment can be achieved through the collective change of different groups of genes. The modulation of certain genes have been shown to subsequently facilitate cancer stem cells phenotypes such as TIC that lead to metastatic colonization. Based on the gene expression changes dictating cellular metabolism and matrix remodeling, we hypothesized that lung low cells could harbor some cancer stem cell properties. Expression of some cancer stem cell markers did identify expression of relevant genes in this set of cells.

To better understand the origins of the CD44 signatures, we examined the scRNA-seq results as they projected over pseudo-time using Monocole2. From the three different cell lines, we identified that lymph node-derived $CD44^{high}$ cells were projected to give rise to lung-derived cells of both signatures. Based on the metastatic cascade of organ tropism, the proximal location of the lymph nodes to the primary tumor is well documented to be the initial site of metastatic lesions before advancing to the more distal lungs. Future studies would benefit from expanding the analysis to include lymph node-derived $CD44^{low}$ cell lines to determine how the signature affects the projection of the tumor initiating cell population.

## IV. Conclusions

Collectively, we established organ-derived cancer cell lines from different metastatic niches. Comprehensive transcriptomic analysis was performed and revealed the impacts of heterogeneity on cancer progression. Bulk sequencing analyses uncovered tissue-specific genes across the different metastatic and primary tumor samples. We further investigated intratumoral heterogeneity by performing single-cell RNA-seq. These data will improve our understanding of the metastatic cascade and tumor heterogeneity in breast cancer will dramatically improve targets for therapies.

## MATERIALS AND METHODS

### *Mammalian cell culture*

Unless otherwise stated, cell lines were cultured in DMEM (Corning) supplemented with 10% (vol/vol) fetal bovine serum (FBS, Life Technologies), penicillin (100 U/mL), and streptomycin (100 µg/mL). Cells were maintained in a 5% $CO_2$ water-saturated incubator at 37 °C.

### *MMTV-PyMT metastatic cell lines as models of breast cancer*

Mouse experiments were approved by the UC Irvine Animal Care and Use Committee. Tumor bearing organs were harvested from 10-12 week females FVB/NJ MMTV-PyMT mice (courtesy of the Kessenbrock laboratory, UCI). Samples were processed mechanically and chemically to dissociate tissues into single cell suspensions as previously published [21]. Primary single cell suspensions were enriched for cancer cells over the course of 1 month *in vitro* incubation by exploiting differences in cellular nutrient requirements and growth differentials. During this time course, primary single cell suspensions were enriched for cancer cells by culturing cell lines in 5% FBS. Cultures were selected for immortalized cancer cells *in vitro* by passaging the flasks 3 times a week. Differences in cellular adhesion properties between fibroblasts and epithelial cancer cells were also exploited *in vitro* through 3 mins versus 7 mins incubations with trypsin [80]. The month long process resulted in the enrichment of cancer cell lines in the surviving *in vitro* cultures prior to FACS sorting.

Primary cell lines were processed for FACS sorting (Institute for Immunology Flow Cytometry Core, UCI) as previously reported [21] [22]. Cancer cells were isolated by EpCAM (BioLegend 118213) and CD44 (BioLegend 103027) cell surface expression levels [23]; isolated cancer cells

expressed either $CD44^{low}/EpCAM^{high}$ or $CD44^{high}/EpCAM^{high}$ cell surface markers. Antibody labeling was performed using manufacture's protocols (BioLegend, USA). Previously sorted MMTV-PyMT MFP-eGFP cell lines (VO), (courtesy of the Kessenbrock laboratory and Lawson laboratory, UCI) were used as positive controls. Fibroblast cell lines (3T3 and MMTV-PyMT-derived fibroblasts, isolated during culturing process above) were used as negative controls during sorting.

***Primary cell line metastatic propensity validation in vivo***

MFP-derived cells (100,000 cells/injection) were injected bilaterally under the fourth gland of disease free, 4-week old FVB/NJ female mice. Control VO-eGFP luciferase-expressing cells were injected as a control to monitor the estimated tumor growth. Palpable primary tumors were detected in all mice within 3-4 weeks post injection. All animals developed primary tumors. Metastatic cell populations were identified by harvesting and processing the organs as described above via FACS analysis. Cancer cells were isolated using cell surface expression of CD44 and EpCAM. The experiment was performed over 4 different biological replicates.

***PCR analysis***

gDNA was isolated from all MMTV-PyMT cell lines and control samples using Zymo (California, USA) quick-DNA miniprep kit (Cat #: 11-317AC). Ear clippings from PyMT positive male and female mice (courtesy of the Kessenbrock laboratory, UCI) were used as positive controls. gDNA samples isolated from 4T1 (ATCC CRL-2539) cell lines were used as negative controls. PCR

amplification conditions and PyMT antigen detection were completed using the standard Jackson Labs genotyping protocol [24].

**Crystal violet proliferation assay**

MMTV-PyMT cell lines were plated (5,000 cells/100 mL) in 96-well plates and incubated for 24 h. Cells were fixed in ice-cold methanol for 30 min. Cells were stained with a solution of 0.05% crystal violet in PBS for 30 min. The samples were then washed three times with PBS to remove excess dye, and allowed to dry for 16–24 h. Crystal violet was recovered from cells via treatment with methanol, and the absorbance of the solutions at 595 nm was measured using a Gen5 microplate reader.

**Immunoblotting**

Cells were lysed in RIPA buffer containing protease (Thermo Fisher Scientific, Cat #88265) and phosphatase inhibitors (Sigma, Cat #4906845001). Protein concentrations were measured using a BCA protein assay (Thermo Scientific, Cat #23223). Samples were prepared in 2X SDS-PAGE loading buffer (containing 4% βME) and heated at 95 °C for 10 min. Samples were then separated on 4–20% polyacrylamide gels (BioRad), and transferred to nitrocellulose membrane (0.2 μm, BioRad). Membranes were incubated with blocking buffer (5% BSA in TBS containing 0.1% Tween-20®, TBST) for 1 h at room temperature. Blots were incubated in primary antibodies (Cell Signaling; 1:1000 dilution in blocking buffer) overnight at 4 °C. Blots were washed three times with TBST (TBS with 0.1% Tween 20), then incubated with IRDye-conjugated secondary antibodies (LI-COR Biosciences; 1:10,000 dilution in blocking buffer) for 1 h at room temperature.

Membranes were washed three times with TBST and imaged using a LI-COR Odyssey CLx imaging system.

***Bulk-RNA-seq***

For each tissue derived cell line, total RNA was extracted using the QIAGEN RNeasy kit with 2 replicates per sample. Sample replicates were distinct clonal isolates harvested from different MMTV-PyMT tumor-bearing mice. A modified SMART-seq2 protocol was used to generate cDNA and Nextera XT DNA Sample Prep Kit to build Illumina libraries. Samples were sequenced on a NextSeq500 with a min depth of 10 M reads. Raw reads were aligned to the mm10 genome with STAR [25] and quantification was performed using the GENCODE v21 annotation of the mouse genome using RSEM [26]. Count matrices for differential expression analysis were used as input for EdgeR [27]. An Exact test was used for calling differential expressed genes with logFC > 2 and a p value < 0.05. EnrichR and metascape were used for Gene ontology analysis.

***Single-cell RNA-seq methods***

As with bulk RNA-seq analysis, scRNA-seq was performed on select tissue-derived cell lines with two replicates per sample. Replicates were distinct clonal isolates harvested from different MMTV-PyMT tumor-bearing mice. Cell lines that originated from Lung with $CD44^{high/low}$ signatures were identified as the samples with most transcriptional changes and were selected for single-cell analysis, along with a Lymph node high sample. Single-cell suspensions from these tissues were used as input for the ddSeq platform and cDNA synthesis as well as library prep was done following the SureCell™ Whole Transcriptome Analysis 3' Library Prep Kit. The

bioinformatic pipeline included Ddseeker [27], a custom demultiplexing script to generate individual fastq, while kallisto [28] was used to quantify the transcripts in our sample using the mm10 and annotation GENCODE v21. Single-cell analysis was done using Seurat v3.2.3 [28]. Cells with more than 250 genes and less than 10% mitochondrial reads were used for the analysis. Monocle2 [28] was used to infer pseudotime progression. Min. read depth 34 M.

**Data availability**

Fastq files for bulk and single-cell datasets as well as their corresponding processed matrices are available in GEO (Accession number: GSE165393)

## References


1. Waks, A.G. and E.P. Winer, *Breast Cancer Treatment: A Review.* JAMA, 2019. **321**(3): p. 288-300.

2. Vanharanta, S. and J. Massague, *Origins of metastatic traits.* Cancer Cell, 2013. **24**(4): p. 410-21.

3. Steeg, P.S., *Targeting metastasis.* Nat Rev Cancer, 2016. **16**(4): p. 201-18.

4. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.

5. Harbeck, N., et al., *Breast cancer.* Nat Rev Dis Primers, 2019. **5**(1): p. 66.

6. Montagner, M. and E. Sahai, *In vitro Models of Breast Cancer Metastatic Dormancy.* Front Cell Dev Biol, 2020. **8**: p. 37.

7. Cai, Y., et al., *Transcriptomic dynamics of breast cancer progression in the MMTV-PyMT mouse model.* BMC Genomics, 2017. **18**(1): p. 185.

8. Zheng, W., et al., *Lung Mammary Metastases but Not Primary Tumors Induce Accumulation of Atypical Large Platelets and Their Chemokine Expression.* Cell Rep, 2019. **29**(7): p. 1747-1755 e4.

9. Olsen, T.K. and N. Baryawno, *Introduction to Single-Cell RNA Sequencing.* Curr Protoc Mol Biol, 2018. **122**(1): p. e57.

10. Yeo, S.K., et al., *Single-cell RNA-sequencing reveals distinct patterns of cell state heterogeneity in mouse models of breast cancer.* Elife, 2020. **9**.

11. Li, H., et al., *Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors.* Nat Genet, 2017. **49**(5): p. 708-718.

12. Perlman, R.L., *Mouse models of human disease: An evolutionary perspective.* Evol Med Public Health, 2016. **2016**(1): p. 170-6.

13. Li, Y. and J. Laterra, *Cancer stem cells: distinct entities or dynamically regulated phenotypes?* Cancer Res, 2012. **72**(3): p. 576-80.

14. De Angelis, M.L., et al., *Stem Cell Plasticity and Dormancy in the Development of Cancer Therapy Resistance.* Front Oncol, 2019. **9**: p. 626.

15. Takebe, N. and S.P. Ivy, *Controversies in cancer stem cells: targeting embryonic signaling pathways.* Clin Cancer Res, 2010. **16**(12): p. 3106-12.

16. Ghajar, C.M., et al., *The perivascular niche regulates breast tumour dormancy.* Nat Cell Biol, 2013. **15**(7): p. 807-17.

17. Ma, J., et al., *Characterization of mammary cancer stem cells in the MMTV-PyMT mouse model.* Tumour Biol, 2012. **33**(6): p. 1983-96.

18. Kersten, K., C. Salvagno, and K.E. de Visser, *Exploiting the Immunomodulatory Properties of Chemotherapeutic Drugs to Improve the Success of Cancer Immunotherapy.* Front Immunol, 2015. **6**: p. 516.

19. Liu, J., H. Dang, and X.W. Wang, *The significance of intertumor and intratumor heterogeneity in liver cancer.* Exp Mol Med, 2018. **50**(1): p. e416.

20. Gomez-Cuadrado, L., et al., *Mouse models of metastasis: progress and prospects.* Dis Model Mech, 2017. **10**(9): p. 1061-1074.

21. Lawson, D.A., et al., *Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells.* Nature, 2015. **526**(7571): p. 131-5.

22. Liao, X., M. Makris, and X.M. Luo, *Fluorescence-activated Cell Sorting for Purification of Plasmacytoid Dendritic Cells from the Mouse Bone Marrow.* J Vis Exp, 2016(117).

23. Marhaba, R., et al., *CD44 and EpCAM: cancer-initiating cell markers.* Curr Mol Med, 2008. **8**(8): p. 784-804.

24. Guy, C.T., R.D. Cardiff, and W.J. Muller, *Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease.* Mol Cell Biol, 1992. **12**(3): p. 954-61.

25. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

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

27. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.

28. Bray, N.L., et al., *Near-optimal probabilistic RNA-seq quantification.* Nat Biotechnol, 2016. **34**(5): p. 525-7.

29. Hen, O. and D. Barkan, *Dormant disseminated tumor cells and cancer stem/progenitor-like cells: Similarities and opportunities.* Semin Cancer Biol, 2020. **60**: p. 157-165.

30. Cardoso, F., et al., *70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer.* N Engl J Med, 2016. **375**(8): p. 717-29.

31. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer.* Nature, 2002. **415**(6871): p. 530-6.

32. Institute, N.C. *Diagnosis and staging*. About Cancer 2019 [cited 2020; Available from: https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis.

33. Takebe, N., R.Q. Warren, and S.P. Ivy, *Breast cancer growth and metastasis: interplay between cancer stem cells, embryonic signaling pathways and epithelial-to-mesenchymal transition.* Breast Cancer Res, 2011. **13**(3): p. 211.

34. Feng, Y., et al., *Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis.* Genes Dis, 2018. **5**(2): p. 77-106.

35. Evans, M.K., et al., *X-linked inhibitor of apoptosis protein mediates tumor cell resistance to antibody-dependent cellular cytotoxicity.* Cell Death Dis, 2016. **7**: p. e2073.

36. Altieri, D.C., *Survivin, cancer networks and pathway-directed drug discovery.* Nat Rev Cancer, 2008. **8**(1): p. 61-70.

37. Wasinski, B., et al., *Discoidin Domain Receptors, DDR1b and DDR2, Promote Tumour Growth within Collagen but DDR1b Suppresses Experimental Lung Metastasis in HT1080 Xenografts.* Sci Rep, 2020. **10**(1): p. 2309.

38. Xu, X., et al., *Matrix metalloproteinase-2 contributes to cancer cell migration on collagen.* Cancer Res, 2005. **65**(1): p. 130-6.

39. Morandi, E.M., et al., *ITGAV and ITGA5 diversely regulate proliferation and adipogenic differentiation of human adipose derived stem cells.* Sci Rep, 2016. **6**: p. 28889.

40. Heldin, C.H., *Targeting the PDGF signaling pathway in tumor treatment.* Cell Commun Signal, 2013. **11**: p. 97.

41. Atsumi, T., et al., *High expression of inducible 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (iPFK-2; PFKFB3) in human cancers.* Cancer Res, 2002. **62**(20): p. 5881-7.

42. Li, N., M.R. Spetz, and M. Ho, *The Role of Glypicans in Cancer Progression and Therapy.* J Histochem Cytochem, 2020. **68**(12): p. 841-862.

43. Nguyen, D.X., P.D. Bos, and J. Massague, *Metastasis: from dissemination to organ-specific colonization.* Nat Rev Cancer, 2009. **9**(4): p. 274-84.

44. Minn, A.J., et al., *Genes that mediate breast cancer metastasis to lung.* Nature, 2005. **436**(7050): p. 518-24.

45. Obenauf, A.C. and J. Massague, *Surviving at a Distance: Organ-Specific Metastasis.* Trends Cancer, 2015. **1**(1): p. 76-91.

46. Westbrook, J.A., et al., *CAPG and GIPC1: Breast Cancer Biomarkers for Bone Metastasis Development and Treatment.* J Natl Cancer Inst, 2016. **108**(4).

47. Sun, J., et al., *Overexpression of CENPF correlates with poor prognosis and tumor bone metastasis in breast cancer.* Cancer Cell Int, 2019. **19**: p. 264.

48. Chen, W., et al., *Organotropism: new insights into molecular mechanisms of breast cancer metastasis.* NPJ Precis Oncol, 2018. **2**(1): p. 4.

49. Ross, C., et al., *Metastasis-Specific Gene Expression in Autochthonous and Allograft Mouse Mammary Tumor Models: Stratification and Identification of Targetable Signatures.* Mol Cancer Res, 2020. **18**(9): p. 1278-1289.

50. Wood, S.L. and J.E. Brown, *Personal Medicine and Bone Metastases: Biomarkers, Micro-RNAs and Bone Metastases.* Cancers (Basel), 2020. **12**(8).

51. Yang, C., et al., *Inducible formation of leader cells driven by CD44 switching gives rise to collective invasion and metastases in luminal breast carcinomas.* Oncogene, 2019. **38**(46): p. 7113-7132.

52. Ye, X., et al., *Distinct EMT programs control normal mammary stem cells and tumour-initiating cells.* Nature, 2015. **525**(7568): p. 256-60.

53. Nowell, P.C., *The clonal evolution of tumor cell populations.* Science, 1976. **194**(4260): p. 23-8.

54. Sachidanandam, R., et al., *A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms.* Nature, 2001. **409**(6822): p. 928-33.

55. Chen, K. and S.I. Fraley, *Abstract 5706: Identifying regulators of cancer heterogeneity with phenotypic sorting and single cell sequencing.* Cancer Research, 2020. **80**(16 Supplement): p. 5706-5706.

56. Weigelt, B., J.L. Peterse, and L.J. van 't Veer, *Breast cancer metastasis: markers and models.* Nat Rev Cancer, 2005. **5**(8): p. 591-602.

57. Quail, D.F. and J.A. Joyce, *Microenvironmental regulation of tumor progression and metastasis.* Nat Med, 2013. **19**(11): p. 1423-37.

58. Nagai, T., et al., *Tactics of cancer invasion: solitary and collective invasion.* J Biochem, 2020. **167**(4): p. 347-355.

59. Huang, T., et al., *Thrombospondin-1 is a multifaceted player in tumor progression.* Oncotarget, 2017. **8**(48): p. 84546-84558.

60. Iliopoulos, D., et al., *Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion.* Proc Natl Acad Sci U S A, 2011. **108**(4): p. 1397-402.

61. Plaks, V., N. Kong, and Z. Werb, *The cancer stem cell niche: how essential is the niche in regulating stemness of tumor cells?* Cell Stem Cell, 2015. **16**(3): p. 225-38.

62. Yi, Y., et al., *Transcriptional suppression of AMPKalpha1 promotes breast cancer metastasis upon oncogene activation.* Proc Natl Acad Sci U S A, 2020. **117**(14): p. 8013-8021.

63. Yang, L., et al., *Targeting cancer stem cell pathways for cancer therapy.* Signal Transduct Target Ther, 2020. **5**(1): p. 8.

64. Tomita, H., et al., *Aldehyde dehydrogenase 1A1 in stem cells and cancer.* Oncotarget, 2016. **7**(10): p. 11018-32.

65. Vassalli, G., *Aldehyde Dehydrogenases: Not Just Markers, but Functional Regulators of Stem Cells.* Stem Cells Int, 2019. **2019**: p. 3904645.

66. Barney, L.E., et al., *Tumor cell-organized fibronectin maintenance of a dormant breast cancer population.* Sci Adv, 2020. **6**(11): p. eaaz4157.

67. Crabtree, J.S. and L. Miele, *Breast Cancer Stem Cells.* Biomedicines, 2018. **6**(3).

68. Fico, F., et al., *Breast Cancer Stem Cells with Tumor- versus Metastasis-Initiating Capacities Are Modulated by TGFBR1 Inhibition.* Stem Cell Reports, 2019. **13**(1): p. 1-9.

69. Lobba, A.R., et al., *Differential expression of CD90 and CD14 stem cell markers in malignant breast cancer cell lines.* Cytometry A, 2012. **81**(12): p. 1084-91.

70. Kim, R.S., et al., *Dormancy signatures and metastasis in estrogen receptor positive and negative breast cancer.* PLoS One, 2012. **7**(4): p. e35569.

71. Sosa, M.S., P. Bragado, and J.A. Aguirre-Ghiso, *Mechanisms of disseminated cancer cell dormancy: an awakening field.* Nat Rev Cancer, 2014. **14**(9): p. 611-22.

72. Ibrahim, S.A., et al., *Syndecan-1 is a novel molecular marker for triple negative inflammatory breast cancer and modulates the cancer stem cell phenotype via the IL-6/STAT3, Notch and EGFR signaling pathways.* Mol Cancer, 2017. **16**(1): p. 57.

73. Gotte, M., et al., *Predictive value of syndecan-1 expression for the response to neoadjuvant chemotherapy of primary breast cancer.* Anticancer Res, 2006. **26**(1B): p. 621-7.

74. Grimm, D., et al., *The role of SOX family members in solid tumours and metastasis.* Semin Cancer Biol, 2020. **67**(Pt 1): p. 122-153.

75. Martelotto, L.G., et al., *Breast cancer intra-tumor heterogeneity.* Breast Cancer Res, 2014. **16**(3): p. 210.

76. Haynes, B., et al., *Breast cancer complexity: implications of intratumoral heterogeneity in clinical management.* Cancer Metastasis Rev, 2017. **36**(3): p. 547-555.

77. Aslakson, C.J. and F.R. Miller, *Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor.* Cancer Res, 1992. **52**(6): p. 1399-405.

78. Lifsted, T., et al., *Identification of inbred mouse strains harboring genetic modifiers of mammary tumor age of onset and metastatic progression.* Int J Cancer, 1998. **77**(4): p. 640-4.

79. Le Voyer, T., et al., *An epistatic interaction controls the latency of a transgene-induced mammary tumor.* Mamm Genome, 2000. **11**(10): p. 883-9.

80. Schor, S. L., & Court, J. (1979). Different mechanisms in the attachment of cells to native and denatured collagen. *J Cell Sci, 38*, 267-281.

# CHAPTER 4: Evolving multicomponent bioluminescent imaging platforms to investigate metastatic disease progression *in vivo* *

## I. Introduction

Biological processes are comprised of heterogeneous and dynamic cellular states that are intrinsically difficult to study. Methods to noninvasively visualize specific cellular processes would greatly improve our understanding of various biological systems *in vivo*. Bioluminescence (BL) is among the most popular techniques for longitudinal imaging of cellular functions in whole animals. Bioluminescence imaging (BLI) relies on genetically encoded enzymes (luciferases) that generate a photon of light via the chemical oxidation of small molecule substrates (luciferins) [1]. Because mammalian tissues do not emit substantial numbers of photons, bioluminescence enables sensitive imaging *in vivo*. Firefly luciferase (Fluc) and its substrate, D-luciferin (Dluc), are the most popular bioluminescent pair (Figure 4.1A). Luciferase-labeled cells are routinely used for long-term non-invasive monitoring in a variety of preclinical models.

* *Author's Note: Figure 1 and the unmixing algorithm, SubstrateUnmixing, in this chapter is part of a manuscript that is in preparation. I share authorship on that manuscript with Carly K. Brennan, Zi Yao, Colin Rathbun, Buvan Sathishkumar, and last author Jennifer A. Prescher. I contributed to the experimental design, procedures, data analyses, and writing. I spearheaded the remaining work in this chapter. I created the experimental designs, performed the data analyses, and completed the writing. I derived all of the cell lines and cancer models. I also performed all of the animal and cell work in this chapter.*

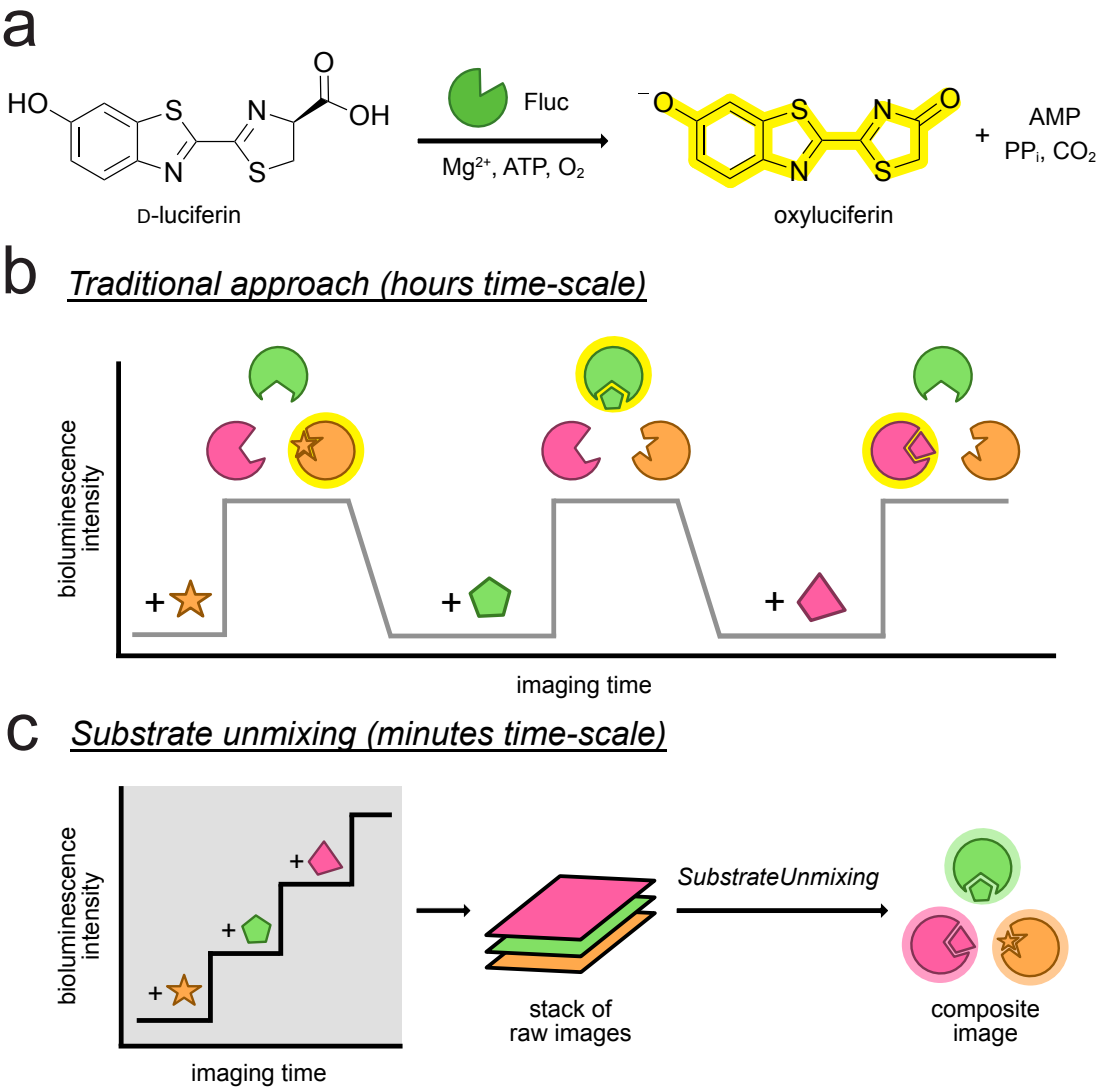


**Figure 4.1. Rapid, multiplexed bioluminescence imaging via sequential substrate administration and serial acquisition.** (a) Optical imaging with bioluminescent probes. A photon of light is produced when D-luciferin (D-luc) is oxidized by firefly luciferase (Fluc). (b) Traditional approach for resolving multiple bioluminescent reporters. Signal from one luciferin must clear before addition of the next luciferin. The required imaging time scales with the number of probes, and can be impractical when more than three targets are involved. (c) Imaging times can be shortened by consecutive substrate application. The resulting images comprise multiple layers of photon output, and require an unmixing step to deconvolute the signal source. (Figure images were produced with permission from Carly Brennan, Zi Yao, Anastasia Ionkina et al.,)

While versatile, bioluminescence has largely been limited to visualizing only one biological feature at a time. This is in part due to the lack of distinguishable luciferase-luciferin pairs suitable for *in vivo* use. In contrast to the diversity of fluorescence imaging tools, the majority of natural luciferases that are suitable for *in vivo* imaging are all derived from the insect family, and thus, all use the same luciferin substrate. To address this void, luciferases have been engineered to emit different spectral wavelength of light with D-luc [1, 2]. In many cases, though, the observed emission spectra were not suitable for *in vivo* applications [3]. An alternative to spectral discrimination, dual component BLI has been achieved instead with substrate resolution. This approach requires mutually orthogonal luciferase-luciferin pairs that produce light together but do not react with other probes. One of the most extensively used dual component BLI pairs, firefly luciferase (Fluc) and *Renilla* luciferase (Rluc), employ chemically unique substrates (D-luciferin and coelenterazine, respectively) that enable tandem use *in vivo* [4, 5]. However, inherent limitations with the stability and bioavailability of coelenterazine can complicate *in vivo* applications.

Despite the great precedence to develop a rapid BLI method with expanded orthogonal luciferase-luciferin pairs, methods suited for *in vivo* use have not been achieved [1]. A rapid multi-component BLI platform could be applied across broad areas of biology, advancing our understanding of complex cellular systems that could not be properly examined with the current technologies. Clinically important biological processes such as intratumoral heterogeneity and multi-organ metastatic progression are notoriously difficult to study due to the multicellular nature of these processes and the profound impact that surrounding microenvironments have on biological outcomes [3]. The best way to maintain rigorous examination of the context dependent cellular mechanisms and discover unknown variables during disease progression is to minimize

perturbation to the model systems with noninvasive monitoring. Currently, the majority of imaging tools are designed for ex vivo use and require invasive sectioning of specimens.

Although dozens of natural and engineered luciferases-luciferins pairs could be employed for multicomponent BLI, most are ill suited for *in vivo* use [3]. These engineered luciferases have enabled multicomponent imaging in bacterial lysate but require high concentrations of the enzymes and probes. Thus, the applicability of these multicomponent imaging reporters for cellular use is unknown. These uncertainties in applicability and functional use are further amplified with *in vivo* systems. Whole animal imaging presents additional complexities and challenges that cannot be accounted for with *in vitro* or bacterial systems. These unknowns include the biodistribution of different luciferins, the tissue penetrant signal of the probes, and an established imaging protocol. Previous *in vivo* imaging attempts using multicomponent substrate resolved luciferases have been difficult to reproduce [9] and require 24 hours to ensure substrate clearance prior to imaging the second probe (Figure 4.1B). The lengthy imaging time is not ideal to monitor dynamic biological processes, and thus there remains an unmet need for a rapid multicomponent BLI platform that is well suited for *in vivo* imaging [6].

To shorten the time required between substrate administration, we have developed a new approach (in Chapter 2) that enables imaging of orthogonal luciferase-luciferin pairs in rapid succession [7]. This approach enabled the rapid imaging of orthogonal probes by sequential substrate administration and serial image acquisition (Figure 4.1C). The substrates are administered by increasing order of brightness, visualizing the dimmest reporter first followed by the immediate imaging of the second brightest reporter. By layering-in the signal intensities there is no need to wait for substrate clearance, achieving dual component imaging in less than an hour [8]. Furthermore, substrate preference allows deconvolution of the signals and enables false

coloring of the mixed cells. To remove any cross talk between the probes, we employ a linear unmixing algorithm to ensure that residual signal from the first image does not interfere with the second image. Although Chapter 2 demonstrates a notable improvement in rapid multicomponent BLI, we aimed to expand *in vivo* imaging capabilities beyond the dual bioluminescent platform.

Here we report a significant advance in the capability of noninvasive imaging tools by developing a rapid triple component BLI method that is specifically engineered for multicellular *in vivo* applications. This imaging platform expands upon our previously successful dual component orthogonal luciferase-luciferin approach (Figure 4.1C). We hypothesized that rapid triple component BLI could be achieved by identifying novel orthogonal luciferase-luciferin pairs that are substrate and/or intensity resolved from our existing pairs. In chapter 2, we presented DB7 cells expressing three different luciferase-fluorescent protein fusions (Pecan-eGFP, Cashew-mNeptune, and Antares-CyOFP) [10]. In hopes of translating this triple for *in vivo* use, I attempted to subject these cell line reporters to our sequential substate imaging platform and unmixing. I continued the *in vivo* imaging experiments by benchmarking the bioluminescent signal to the fluorescence readout through ex vivo FACS analysis of the tumors.

Chapter 2 examined the problems of using Antares for multicomponent imaging *in vitro*. To mitigate against these issues, I identified a new engineered firefly luciferase-luciferin pair, Akaluc/AkaLumine, to replace the Antares/furimazine pair as the third orthogonal probe. I also confirmed that Akaluc along with its substate AkaLumine are well suited for *in vitro* and *in vivo* use. Rapid triple component BLI was achieved by developing a protocol that enabled sequential substate administration and serial imaging of cell mixtures expressing the three different orthogonal luciferases. The order of sequential substrate administration and imaging parameters were validated in mixed cell populations *in vitro*. The imaging parameters were also examined *in*

*vivo*. Finally, I translated our platform to study cancer heterogeneity and metastatic progression using a new derivation of the MMTV-PyMT cell lines outlined in Chapter 3. I outline my efforts creating MMTV-PyMT cell lines for different organ-derived metastatic sites that express our triple luciferase-luciferin pairs. In addition, I have created other cancer-relevant cell line models that could also be used with our imaging system, showcasing its strengths over conventional imaging platforms [19]. Collectively, I present my efforts generating and applying cancer-relevant luciferase expressing models to expand the use of our multicomponent bioluminescent imaging platform *in vitro* and *in vivo*. This work sets the stage for future studies of cancer metastases.

## II. Results and Discussion

### Expanding rapid BLI to three probes with Antares *in vitro*

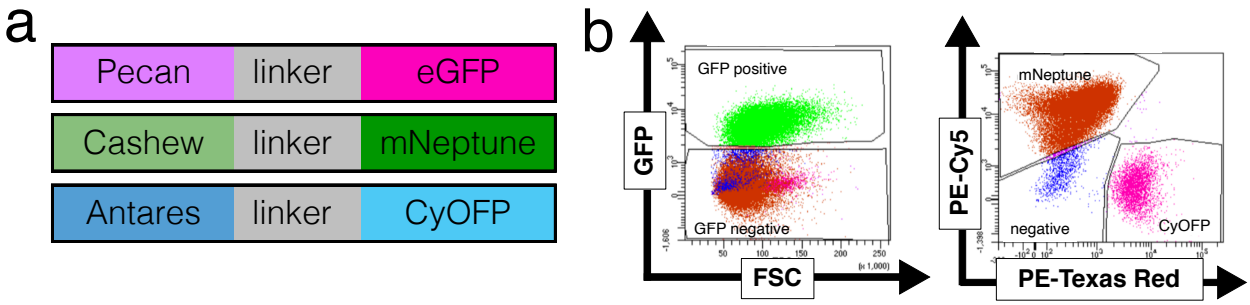


**Figure 4.2. Luciferase probes for multicomponent imaging.** (A) CRISPR DNA constructs for tri-component imaging comprise luciferases tethered to fluorescent proteins through a linker region. (B) Cells stably expressing one of the luciferase constructs from (A). FACS plots confirm that the fluorescent proteins are spectrally resolved in the three different cell populations.

Triple component BLI was achieved in Chapter 2 with Pecan, Cashew and Antares luciferases. To ensure the bioluminescent images were reliably reporting on cell number, the bioluminescence signals from the three orthogonal luciferase-luciferin pairs must be benchmarked against established reporters (ie fluorescence). Each orthogonal luciferase was fused to a unique genetically encodable fluorescent protein (Figure 4.2a), thus providing a direct correlation between luminescence and fluorescence expression within the cell. Since CyOFP, the fluorescent protein fused in Antares [10], has similar excitation and emission profiles as the other two fluorophores (eGFP and mNeptune), extra care was taken to ensure proper delineation of the colors. Gating was performed on DB7 cell lines first before translating this to ex vivo analysis of resected tumors (Figure 4.2b).

**Translating orthogonal triple for rapid, multicomponent BLI *in vivo***

As a model triple component set, we used Cashew and Pecan in combination with Antares, a recently reported marine luciferase variant [10, 11]. Cashew and Pecan are derived from the terrestrial luciferase family, and are thus immediately orthogonal to luciferases (like Antares) that use vastly distinct luciferins (in this case, furimazine) [12]. Antares also exhibits markedly faster substrate turnover than Cashew, rendering it brighter and intensity resolved from the other two reporters [13]. Thus, we reasoned that the three orthogonal luciferases could be rapidly differentiated by first applying 4'-BrLuc, followed by D-luc, then furimazine to layer in signal from Pecan, Cashew, and Antares, respectively. When the cells were combined and imaged together, the three reporters could be rapidly visualized (<15 min) following sequential substrate addition (Figure 2.13). Triplet imaging was also readily achieved using other combinations of engineered and native luciferases (Figures 2.14 – 2.15).

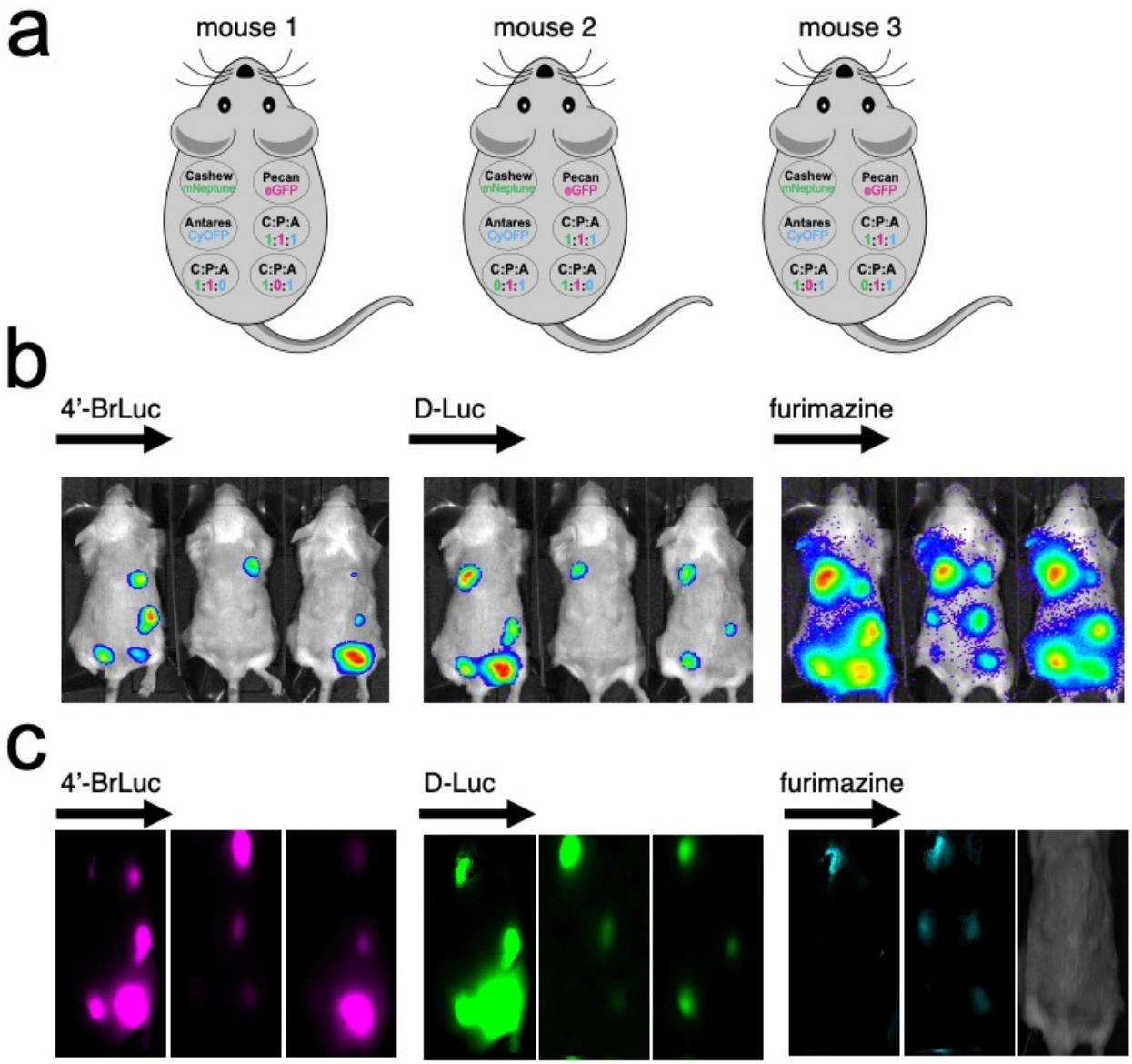


**Figure 4.3. Rapid BLI using Pecan, Cashew, Antares triples for mice 1-3 *in vivo*.** (a) Ratios of Cashew-, Pecan-, and Antares-expressing cells implanted in the mice as well as single luciferase-expressing cells that served as positive controls. Ratios of mixed cells are noted. (b) Orthogonal substrates (65 mM) and furimazine (100 µL) were administered sequentially via i.p. injection. Raw images were acquired for 35 min after each injection. (b) Unmixed channels for each mouse are shown. Colors go as followed for each luciferase-luciferin pair: Pecan = magenta, Cashew = green, Antares = cyan.

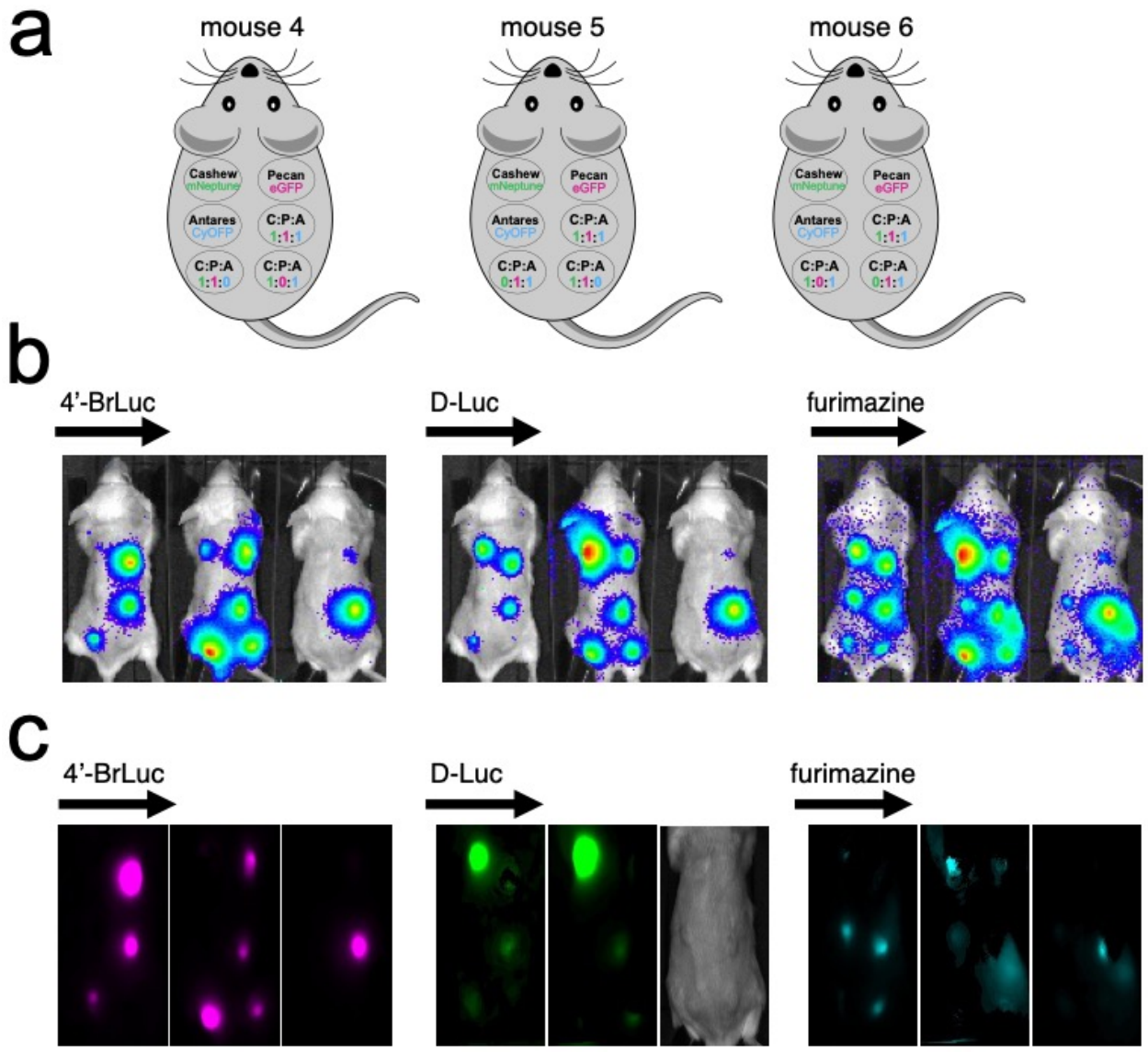


**Figure 4.4. Rapid BLI using Pecan, Cashew, Antares triples for mice 4-6 *in vivo*.** (a) Ratios of Cashew-, Pecan-, and Antares-expressing cells implanted in the mice as well as single luciferase-expressing cells that served as positive controls. Ratios of mixed cells are noted. (b) Orthogonal substrates (65 mM) and furimazine (100 µL) were administered sequentially via i.p. injection. Raw images were acquired for 35 min after each injection. (b) Unmixed channels for each mouse are shown. Colors go as followed for each luciferase-luciferin pair: Pecan = magenta, Cashew = green, Antares = cyan.

We demonstrated in Chapter 2 that the unmixing algorithm would enable rapid dual component imaging *in vivo* (Figure 2.11). Next, we aimed to expand our imaging platform by facilitating rapid triple-component BLI *in vivo*. Pecan-, Cashew-, and Antares-expressing cells

were mixed in varying ratios and implanted ($10^6$ cells per site) in a total of 6 mice (mice 1-3 Figure 4.3a) (mice 4-6 Figure 4.4a). Tumor volumes for each mouse were recorded before imaging in Table 4.1. I employed our previous "layering in" approach where substrates were sequentially administered to achieve rapid BLI. I administered the luciferins from dimmest to brightest in the following order: 4'-BrLuc, followed by D-luc, and lastly furimazine. Raw images for the mice (Figure 4.3b and 4.4b) were subjected to our unmixing algorithm and false colored based on the luciferin (Figure 4.3c and 4.4c). The unmixed and false colored images were overlayed creating a merged image of the three luciferin channels for each of the 6 mice (Figure 4.5 and 4.6). Overall, we observed consistent BL signal from Pecan and Cashew expressing cells. However, Antares-expressing cells exhibited inconsistences in signal intensity after administration of furimazine. Issues with bioavailability and stability of furimazine, compared to firefly luciferin analogs, have been reported by others [14]. I believe that these issues impacted the results of the *in vivo* experiments. There also were notable differences in tumor volumes with Antares-expressing cells compared to Pecan- or Cashew-expressing tumors (Table 4.1). The tumor growth differential seemed directly caused by the integration of Antares into the cell's genomes. This likely interfered with cellular proliferation. Cells expressing Pecan and Cashew (firefly-derived luciferases) did not exhibit this growth inhibition (Table 4.1). These results further show that Antares is not ideal for our *in vivo* BLI platform and underlines the need to identify a better luciferase-luciferin pair.

To ensure the images were reporting on cell number and to provide a point of comparison, the bioluminescence images must be benchmarked against fluorescence signals (Figure 4.7). The initial design of the BLI reporters (Figure 4.2) had a spectrally resolved fluorescent protein fused to each one of the luciferases. As diagrammed in Figure 4.7, rapid BLI results were validated with

ex vivo analysis. Tumors consisting of different ratios of luciferase-expressing cells were processed and subjected to flow cytometry analysis.

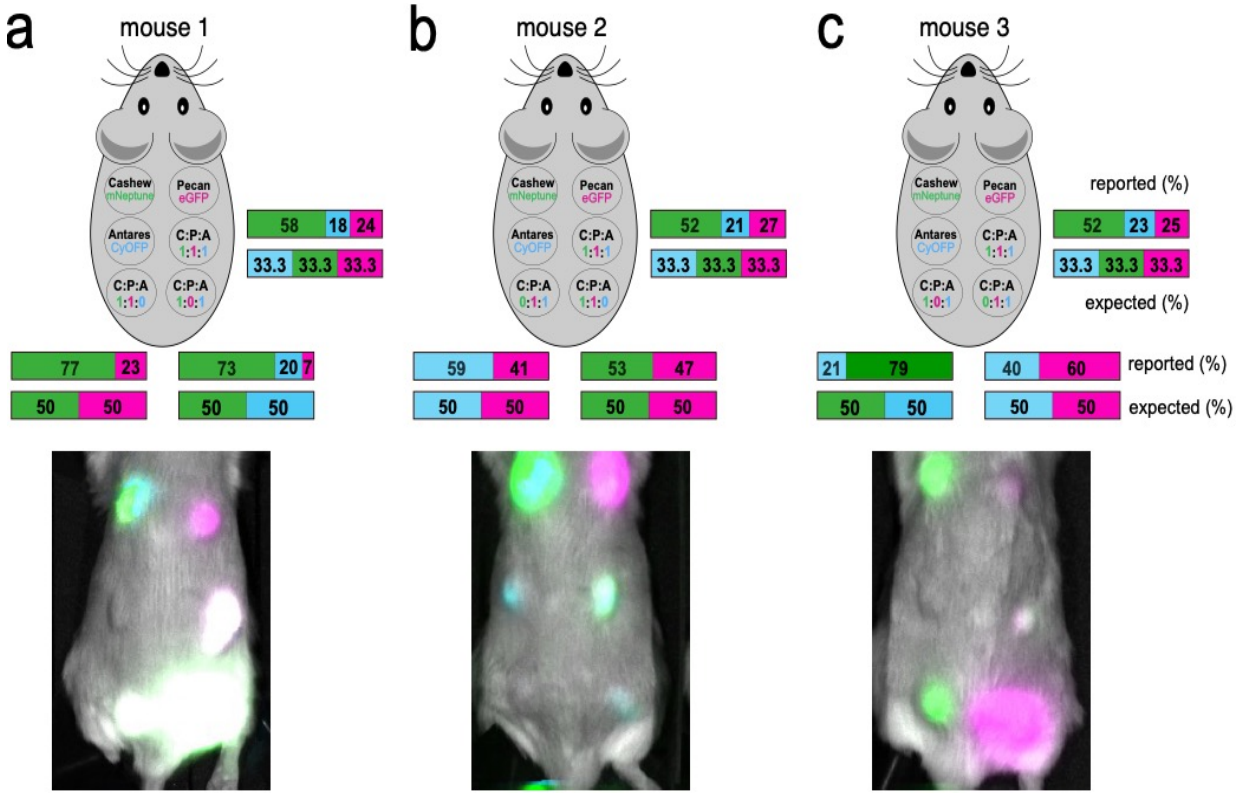


**Figure 4.5. Rapid BLI *in vivo* and benchmarking BLI to fluorescent proteins for mice 1-3.** Ratios of Cashew, Pecan, and Antares expressing DB7 cells injected for (a) mouse 1, (b) mouse 2, and (c) mouse 3. Bioluminescence images were benchmarked against the spectrally resolved fluorescent protein fused to each luciferase. After imaging, tumors with mixed cell populations were resected, processed into single cell suspensions and analyzed via flow cytometry. Results from the ex vivo analyses are shown as percentages. The lower bars show the expected percentage of fluorescent signal directly corresponding to the ratio of injected cells. The top bars show the relative percentage of the reporters from flow cytometry analysis of the resected tumors. Merged images of the unmixed and false colored mice are show at the bottom for (a) mouse 1, (b) mouse 2, and (c) mouse 3.

The initial hypothesis was that the ratios of luminescent signal observed *in vivo* would be comparable to the ratios of cells revealed by fluorescence ex vivo. Although some of the ratios were similar (Figures 4.5 and 4.6), there were clear discrepancies between the images and the flow cytometry data. There are several plausible reasons for these inconsistent results. The small tumor volumes (Table 4.1) across the animals clearly limited the amount of tissues that were available

for flow cytometry processing. Future experiments would benefit by allowing the tumors to reach larger starting volumes, thus increasing the material available for analysis. Additionally, processing and running ex vivo analyses for such a large amount of samples was extremely labor intensive and difficult for a single person to perform alone. Future renditions of this ex vivo analysis would dramatically benefit from having additional assistance to ensure success and reproducible results.

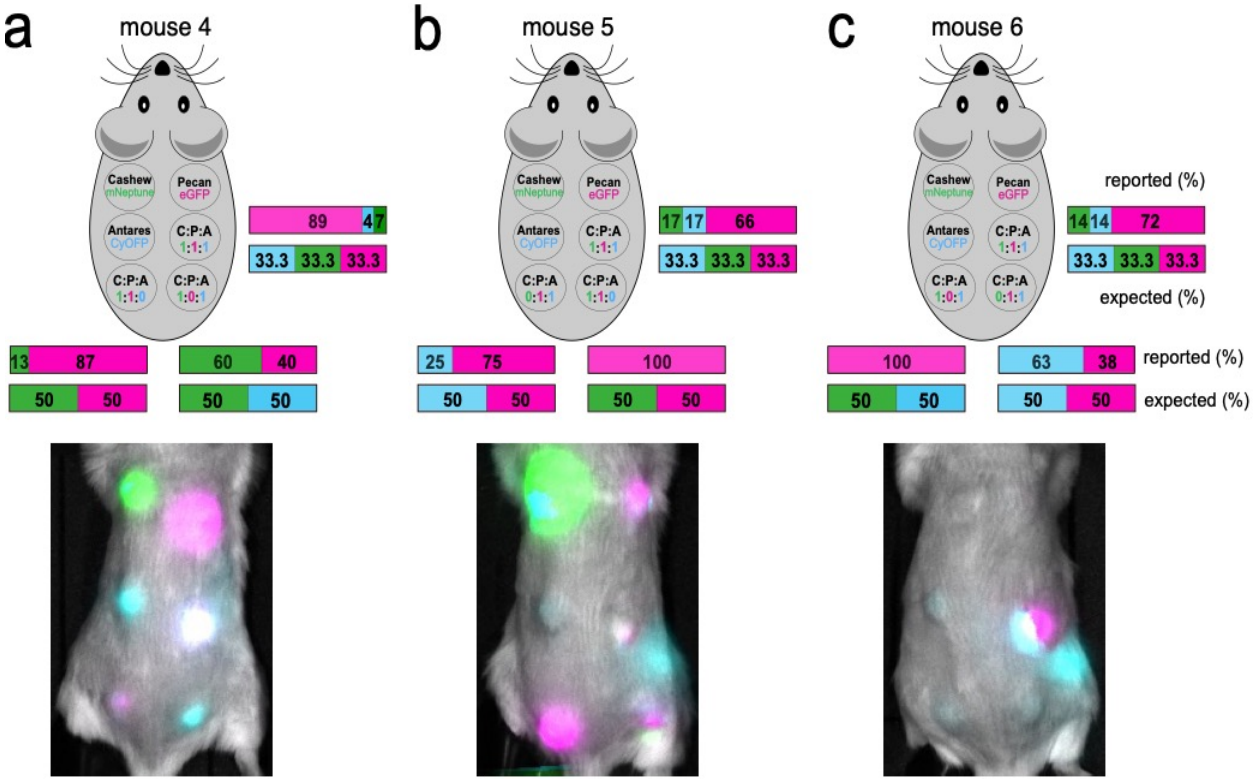


**Figure 4.6. Rapid BLI *in vivo* and benchmarking BLI to fluorescent proteins for mice 4-6.** Ratios of Cashew, Pecan, and Antares expressing DB7 cells injected for (a) mouse 4, (b) mouse 5, and (c) mouse 6. Bioluminescence images were benchmarked against the spectrally resoled fluorescent protein fused to each luciferase. After imaging, tumors with mixed cell populations were resected, processed into single cell suspensions and analyzed via flow cytometry. Results from the ex vivo analyses are shown as percentages. The lower bars show the expected percentage of fluorescent signal directly corresponding to the ratio of injected cells. The top bars show the relative percentage of the reporters from flow cytometry analysis of the resected tumors. Merged images of the unmixed and false colored mice are show at the bottom for (a) mouse 4, (b) mouse 5, and (c) mouse 6.

**Table 4.1. Tumor volumes for mice 1-6 prior to sequential substrate administration and imaging (Figure 4.3 and 4.4)**

| Mouse Number | Tumor Number | Tumor Cell Types | Tumor Volume (mm3) |
|---|---|---|---|
| 1 | 1 | Cashew (positive control) | 156.12 |
| | 2 | Antares (positive control) | 85.49 |
| | 3 | Exp1 (Cashew:Pecan) | 115.20 |
| | 4 | Pecan (positive control) | 102.44 |
| | 5 | Pecan:Cashew:Antares (1:1:1 ratio) | 138.39 |
| | 6 | Exp2 (Cashew:Anatres) | 137.82 |
| 2 | 1 | Cashew (positive control) | 172.46 |
| | 2 | Antares (positive control) | 76.21 |
| | 3 | Exp3 (Pecan:Antares) | 108.29 |
| | 4 | Pecan (positive control) | 167.82 |
| | 5 | Pecan:Cashew:Antares (1:1:1 ratio) | 90.67 |
| | 6 | Exp1 (Cashre:Pecan) | 136.81 |
| 3 | 1 | Cashew (positive control) | 100.33 |
| | 2 | Antares (positive control) | 41.26 |
| | 3 | Exp2 (Cashew:Antares) | 106.88 |
| | 4 | Pecan (positive control) | 63.23 |
| | 5 | Pecan:Cashew:Antares (1:1:1 ratio) | 99.03 |
| | 6 | Exp3 (Pecan:Antares) | 198.14 |
| 4 | 1 | Cashew (positive control) | 129.64 |
| | 2 | Antares (positive control) | 111.21 |
| | 3 | Exp1 (Cashew:Pecan) | 111.02 |
| | 4 | Pecan (positive control) | 88.50 |
| | 5 | Pecan:Cashew:Antares (1:1:1 ratio) | 112.77 |
| | 6 | Exp2 (Cashew:Anatres) | 132.86 |
| 5 | 1 | Cashew (positive control) | 116.03 |
| | 2 | Antares (positive control) | 125.46 |
| | 3 | Exp3 (Pecan:Antares) | 115.99 |
| | 4 | Pecan (positive control) | 137.13 |
| | 5 | Pecan:Cashew:Antares (1:1:1 ratio) | 107.27 |
| | 6 | Exp1 (Cashre:Pecan) | 128.79 |
| 6 | 1 | Cashew (positive control) | 129.19 |
| | 2 | Antares (positive control) | 95.81 |
| | 3 | Exp2 (Cashew:Antares) | 156.58 |
| | 4 | Pecan (positive control) | 118.89 |
| | 5 | Pecan:Cashew:Antares (1:1:1 ratio) | 145.01 |
| | 6 | Exp3 (Pecan:Antares) | 130.69 |

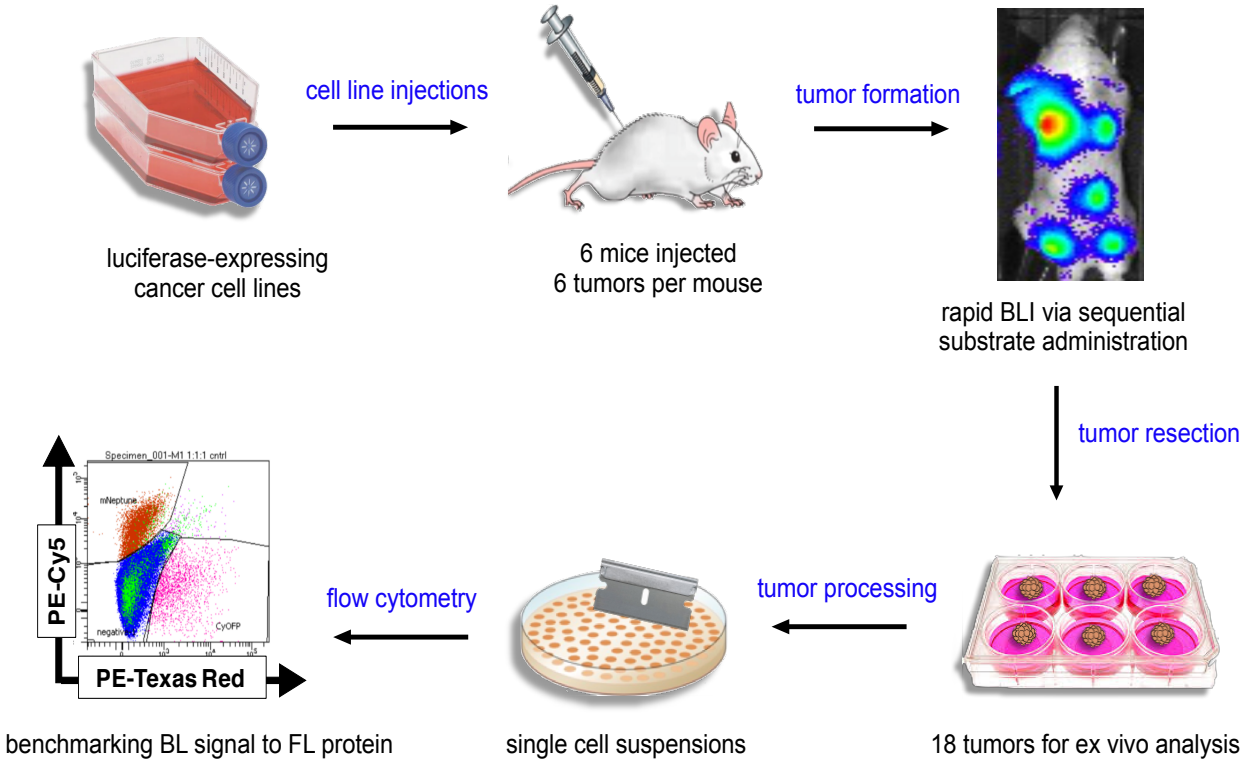


**Figure 4.7. Schematic of rapid multicomponent BLI *in vivo* experiment and ex vivo analysis.** Luciferase expressing cancer cells are injected into mice at different ratios. Rapid BLI via sequential substrate unmixing is performed after tumors are palpable. Select tumors are resected and processed for flow cytometry analysis to benchmark images against fluorescence signal.

### Akaluc/ AkaLumine is well suited for the third orthogonal pair for rapid BLI *in vitro*

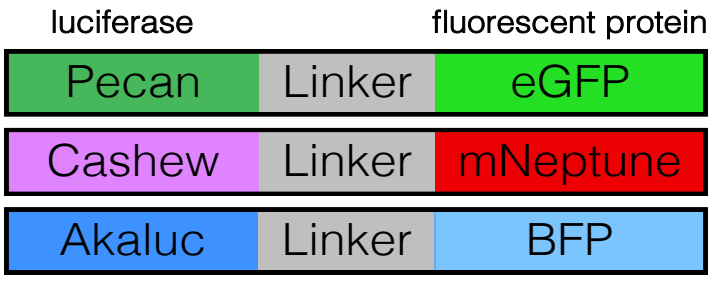


**Figure 4.8. Luciferase probes for multicomponent imaging.** CRISPR DNA constructs for tri-component imaging comprise luciferases tethered to fluorescent proteins through a linker region.

Akaluc along with its luciferin analog, AkaLumine, are appropriate alternatives to Antares/furimazine for substrate unmixing and *in vivo* imaging [15]. Exhibiting high substrate preference, Akaluc was engineered from firefly luciferase to selectively emit light with its synthetic luciferin analog, AkaLumine [16]. This substrate preference fits well

with the requirements of our unmixing platform. Similar to the other firefly-derived probes (Pecan/4'-BrLuc and Cashew/D-luc), AkaLumine [17] (Figure 4.8) exhibits more substrate stability and bioavailability than the marine-derived Antares/furimazine pair. In addition, the 650 nm red-shifted light emission of Akaluc-AkaLumine (Figure 4.9) is advantageous for *in vivo* imaging, as red-light passes more readily through blood and tissues [18]. Collectively, I hypothesized that the Akaluc/ AkaLumine luciferase-luciferin pair would be well suited for triple component imaging.

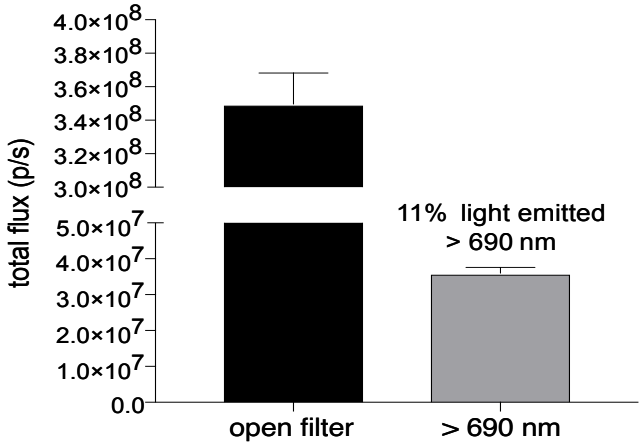


**Figure 4.9. Red-shifted light emission of Akaluc/ AkaLumine pair.** Cells stably expressing Akaluc were imaged after addition of the AkaLumine substrate. Luminescence data is show with open filter (i.e. all light) and using a filter that only allows red-shifted light greater than 690 nm to pass. Compared to the total emission, 11% of Akaluc- AkaLumine emitted light was greater than 690 nm (ideal for *in vivo* imaging). This is compared to only 1.3% of light >690 nm that is emitted by Fluc/D-luc.

An important requirement for the proper utilization of our substrate unmixing platform is that the first luciferase emission must reach a stable plateau before adding the next substate. This signal plateau ensures that any increase in light emission from sequential substrate administration is solely due to newly added luciferin. We previously have shown that 30 min is best for Pecan and Cashew [7]. However, the use of new mutant luciferase-luciferin pairs will exhibit changes in kinetics and substrate turn over that must be taken into account [3]. For example, Akaluc/AkaLumine was engineered for improved brightness and turnover [15]. Thus, the acquisition time required to capture photon output had to be adjusted and the length at which the signal plateaus had to be measured. To apply the three probes for multicomponent imaging, we ranked the order of the

luciferase-luciferin probes to achieve maximal resolution: Pecan/4′-BrLuc, then Cashew/D-luc, and finally Akaluc/AkaLumine. This second viable triplet set was identified and stably expressed into DB7 cell lines in Chapter 2. I sought to test its applicability for mixture analysis *in vitro* and for *in vivo* applications

**Rapid triple component BLI revealed expected distribution of Pecan-, Cashew- and Akaluc-expressing DB7 cell mixtures *in vitro***

We aimed to investigate whether the relative composition of wells containing different ratios of luciferase-expressing cells could be deconvoluted. The three luciferase expressing cells were plated in gradients across a 96 well plate with positive control wells containing 100% of each cell population for the unmixing algorithm. The false colored image showed roughly the expected mixture of luciferase-expressing cells per specific well (Figure 4.10). Complex mixtures of cell populations were able to be unmixed and identified. Total flux of the control wells were measured with sequential addition (Figure 4.11). Additionally, we discovered important ranges in the fold change of signal required for the unmixing algorithm to achieve resolution. In order for the algorithm to deconvolute based on signal intensities, there is a minimal range that the signal must increase between sequential substrate administration. We have identified that a 10-fold increase in total photon emission is required between luciferase-luciferin pairs (meta data analysis not shown).

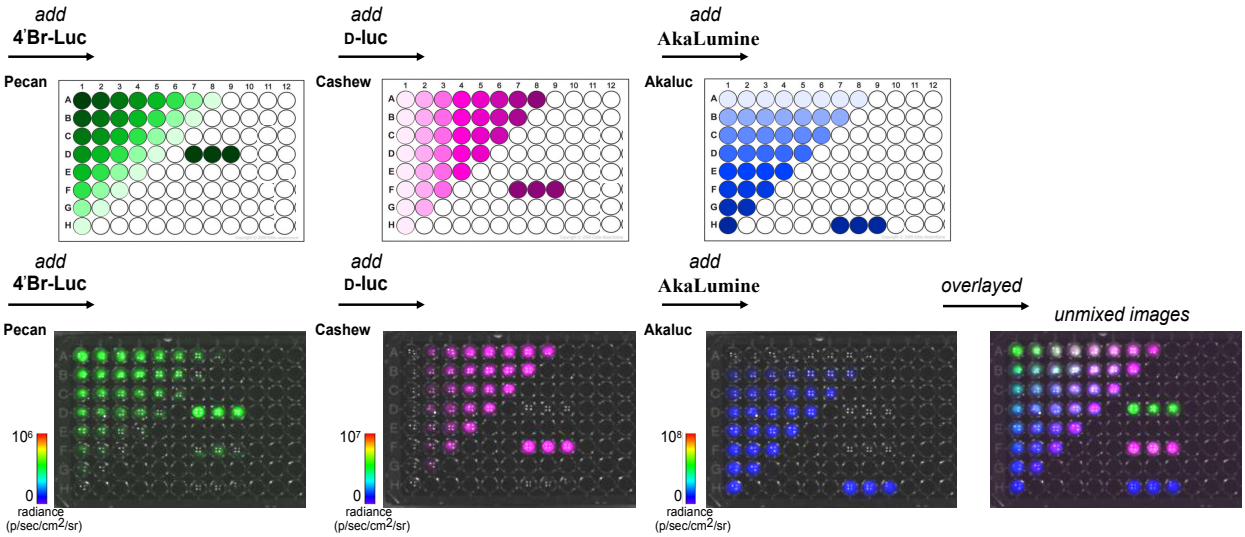


**Figure 4.10. Rapid BLI with three orthogonal luciferase-luciferin pairs *in vitro*.** DB7 cells expressing Pecan, Cashew, or Akaluc were plated in at 96-well plate. Raw bioluminescence signal from sequential substrate addition of 4'-BrLuc (250 µM), followed by D-luc (250 µM), then AkaLumine (250 µM) acquired after the addition of each substrate. The substrate-specific signals were unmixed, assigned false colors and overlaid. Data is representative of $n$ = 3 replicates.

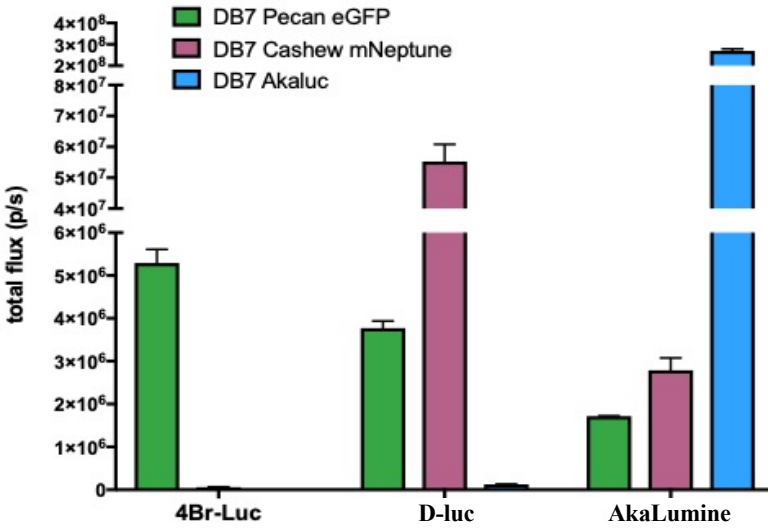


**Figure 4.11. Rapid BLI with three orthogonal luciferase-luciferin pairs *in vitro* quantification.** Photon flux from the 100% positive control wells from Figure 4.10 were quantified. Error bars represent the standard error of the mean for $n$ = 3 experiments.

**Rapid triple component BLI revealed expected distribution of Pecan-, Cashew- and Akaluc-expressing DB7 cells *in vivo***

To showcase the improved performance of Akaluc/AkaLumine, I sought to examine the triplet to achieve rapid multicomponent imaging *in vivo*. DB7 Pecan-, Cashew-, and Akaluc-expressing cells were implanted using 100% positive controls and 1:1:1 mixed ratio populations in two mice for a pilot experiment (Figure 4.12a). Mice were subjected to sequential substate administration and image acquisition as previously outlined. Raw images for the mice (Figure 4.12b) were subjected to our unmixing algorithm and false coloring. Although this triplet (Pecan, Cashew, Akaluc) performed better than the original triplet with Antares, the method needs improvement in order to achieve triple-component imaging *in vivo*. We observed a loss of signal that was not mitigated with the unmixing algorithm. Notably there was a loss of Cashew substrate preference upon the addition of D-luc. Sequencing of the Cashew plasmids identified a random mutation that had appeared and affected Cashew's substrate preference. Secondly, the mixed cell population tumors (1:1:1 ratios) for both mice did not represent all three of the cell lines – instead the tumor was dominated by Akaluc signals in the false colored merged images. The lack of a homogenized signal is likely do to the need for identification of signal threshold and improving the unmixing algorithm so that dimmer luciferase signals are not lost with brighter probes.

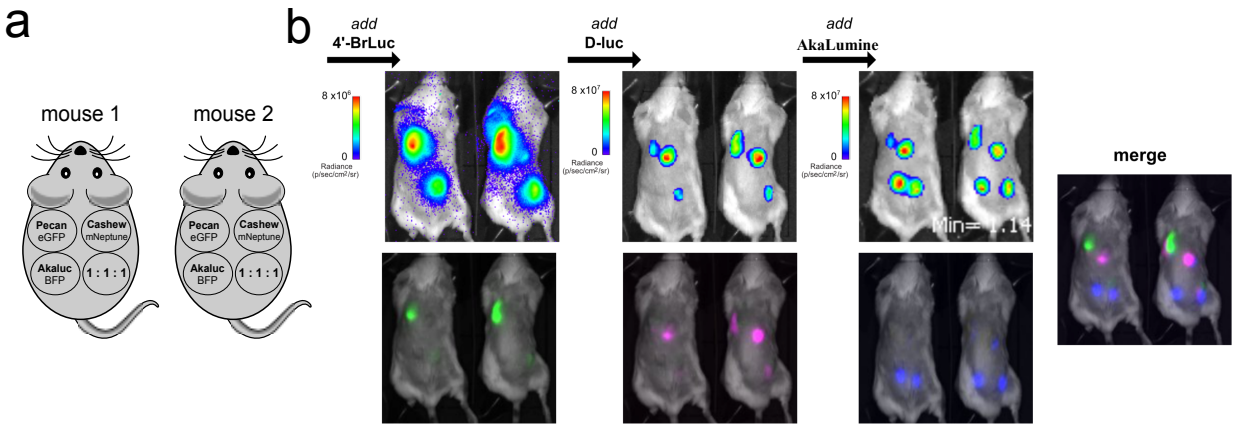


**Figure 4.12. Rapid BLI using DB7-expressing Pecan, Cashew, Akaluc triples *in vivo*.** (a) Ratios of 1:1:1 Cashew-, Pecan-, and Akaluc-expressing cells implanted in the mice as well as single luciferase-expressing tumors that served as positive controls. (b) Orthogonal substrates (65 mM) and AkaLumine (30 mM) were administered sequentially via i.p. injection. Raw images were acquired 35 min after each injection. Unmixed and false colored channels for each mouse are shown on the bottom panel. Colors are as followed for each luciferase-luciferin pair: Pecan = green, Cashew = magenta, Akaluc = blue. Channels were merged showing a composite image on the right.

**Rapid triple-component imaging of metastatic MMTV-PYMT cell lines**

The multicomponent imaging platform is primed to answer complex questions such as cancer heterogeneity and metastatic progression. To achieve this, I aimed to investigate metastatic progression by creating organ-derived metastatic cell lines that expressed different luciferase reporters (Figure 4.13a). To this end, I used the MMTV-PyMT primary cell lines (MFP, LN, and lung) that I isolated and characterized in Chapter 3. To simplify some of the variables in this experiment, I combind both $CD44^{low}/EpCAM^{high}$ and $CD44^{high}/ EpCAM^{high}$ expressing cells from each organ-derived site. MMTV-PyMT MFP, LN, and Lung cells were engineered to express Pecan-eGFP, Cashew-mNeptune, or Akaluc-BFP, respectively, via CRISPR-mediated gene insertion (Figure 4.13b). Plasmids contained luciferases fused to spectrally resolved fluorescence proteins for use with FACS (Figure 4.8a).

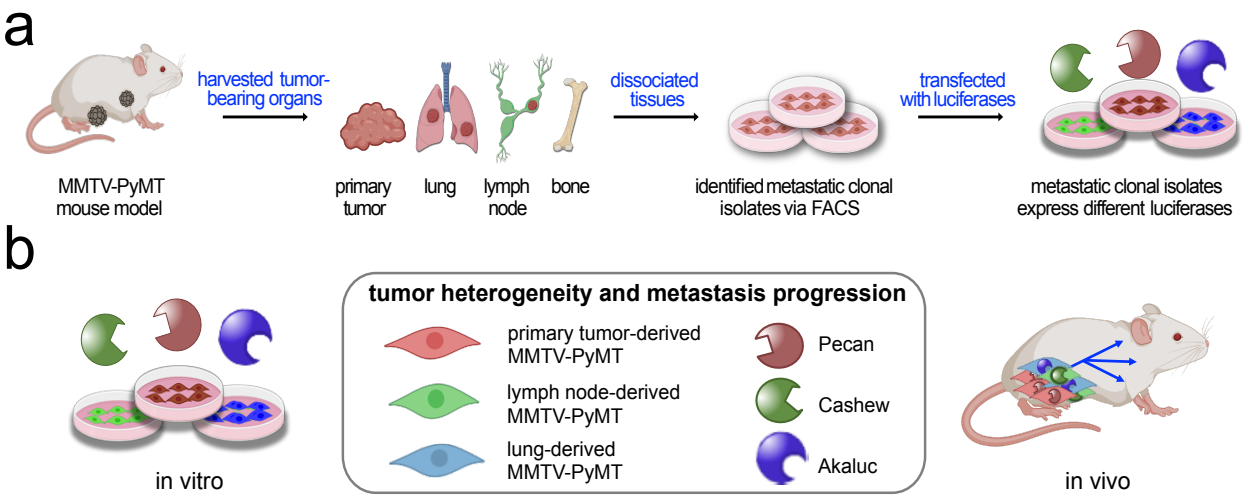


**Figure 4.13. Examining cancer metastasis by expressing orthogonal luciferase pairs in organ-derived metastatic cell lines from the MMTV-PyMT mouse model.** (a) As detailed in chapter 3, MMTV-PyMT mice were used to isolate metastatic cell lines. Orthogonal luciferases were expressed in the metastatic clonal isolates to track cancer progression. (b) MMTV-PyMT cell lines mammary fat pad (MFP), lymph node, and lung-derived cell lines were transfected with Pecan-eGFP, Cashew-mNeptune, and Akaluc-BFP CRISPR Cas9 constructs. Cell lines can be used to study areas relevant to metastatic progression both *in vitro* and *in vivo*.

**Validating substrate preference and intensity resolution of orthogonal pairs in MMTV-PyMT cancer cell lines *in vitro***

As discussed above, MMTV-PyMT luciferase-expressing cell lines were established and characterized (Figure 4.13). Substrate preference and intensity resolution in the new MMTV-PyMT luciferase expressing cell lines were validated *in vitro*. Sequential addition (4'-BrLuc, followed by D-luc, then AkaLumine) and image acquisition were performed. Cell lines were plated in 96 wells and included 100% positive controls and mixed cell populations (1:1:1 ratios) in triplicate. Signal intensity (total flux) upon sequential substate administration was measured right after plating on day 0 (Figure 4.14a) and 24 hours after plating on day 1 (Figure 4.14b). As

expected, the total flux for day 1 cells was higher than at day 0. This increase in signal is most likely due to the additional 24 hours of cellular growth between the two experiments.

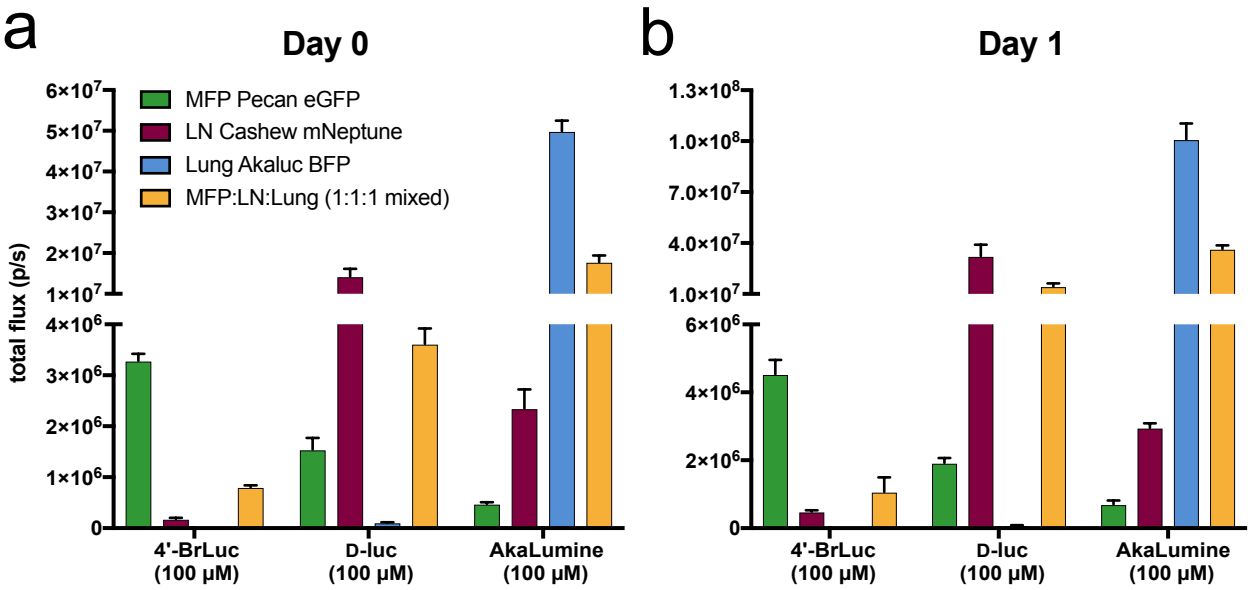


**Figure 4.14: Rapid, three-component BLI of metastatic breast cancer model over 24 hours.** (a) MMTV-PyMT organ-derived breast cancer cell lines expressing each luciferase were plated as shown (MFP-Pecan; LN-Cashew; Lung-Akaluc). The corresponding substrates ([4′-BrLuc] = [D-luc] = [AkaLumine] = 100 µM) were administered, beginning with the dimmest luciferin. Images were acquired after each addition. Image acquisition was completed within 30 minutes. Raw data in total photon flux is shown for (a) day 0 and 24 hrs later at day 1 (b). An increase in total flux was observed in (b) day 1 due to cell growth and was expected.

With the sequential addition and substrate resolution of the MMTV-PyMT cell lines confirmed, we sought to determine if images containing mixtures of the three cell populations could successfully be deconvoluted and visualized using our unmixing algorithm, *SubstrateUnmixing* (Figure 4.15). Control wells of MMTV-PyMT MFP-Pecan-eGFP, LN Cashew-mNeptune, and Lung Akaluc-BFP cells were plated in 96 wells with 100% controls and 1:1:1 mixed ratio wells (Figure 4.15a). Sequential substrate administration enabled rapid imaging acquisition of only 30 mins. Raw images were acquired and subjected to unmixing using

*SubstrateUnmixing* (Figure 4.15a). Successful signal attribution within the mixed wells was confirmed (Figure 4.15b).

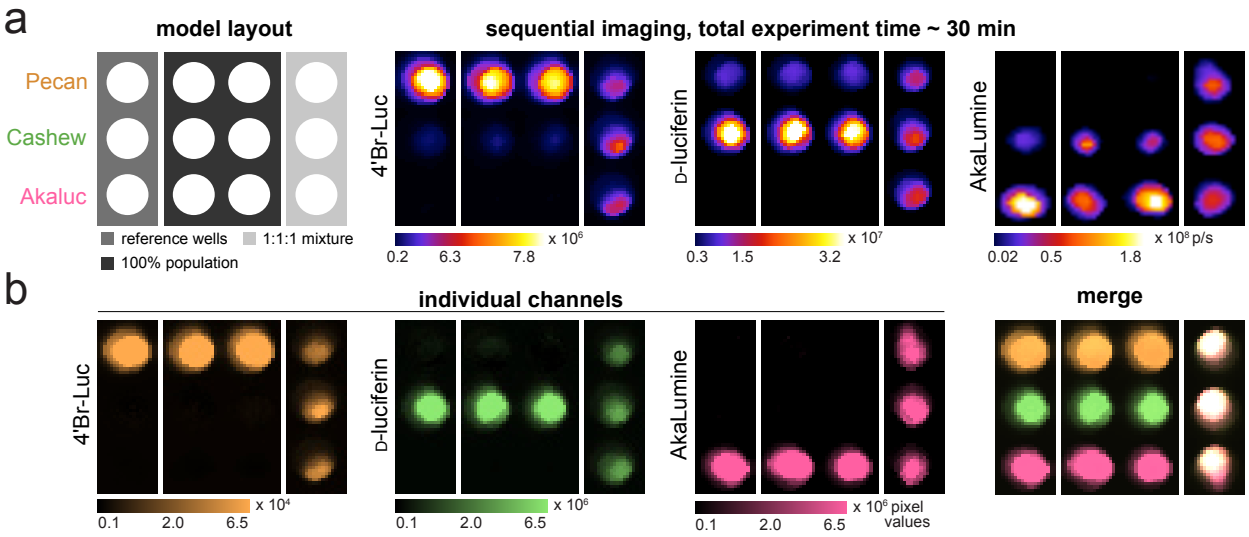


**Figure 4.15: Rapid, three-component BLI of metastatic breast cancer model via *SubstrateUnmixing*.** (a) MMTV-PyMT organ-derived breast cancer cell lines expressing each luciferase were plated as shown (MFP-Pecan; LN-Cashew; Lung-Akaluc). The corresponding substrates ([4′-BrLuc] = [D-luc] = [AkaLumine] = 100 μM) were administered, beginning with the dimmest luciferin. Images were acquired after each addition. Image acquisition was completed within 30 minutes. (b) The raw data from (a) were unmixed using *SubstrateUnmixing* and false colored.

**Measuring changes in signal intensity for MMTV-PyMT cell lines using multicomponent imaging platform and linear regression**

The ability to monitor signal change in response to cellular growth during cancer progression would be a powerful tool to assess biological changes. To this end, we investigated the ability of our unmixing algorithm to identify changes in signal intensity from the triple MMTV-PyMT cell lines. We simulated changes in signal output by varying the amount of one luciferase-expressing cell line while keeping the remaining two constant. The varying cell mixtures were

plated in 96 wells and subjected to sequential substrate administration and image acquisition. Unmixed images were false colored and merged for each well. The false colored signal from the Pecan-expressing cells increased in signal as the amount of cells increased (Figure 4.16a), following the predicted trend (Figure 4.16d). The change in signal was further validated through linear regression analysis between the unmixed signals and the number of luciferase-expressing cells plated. Both samples exhibited similar trends and were highly correlated. Similar correlations were observed for LN Cashew-mNeptune (Figure 4.16b) and Lung Akaluc-BFP (Figure 4.16c) cell lines. The consistent linear trends showed that changes in luciferase-signal (Figure 4.16e, f) could be successfully deconvoluted.

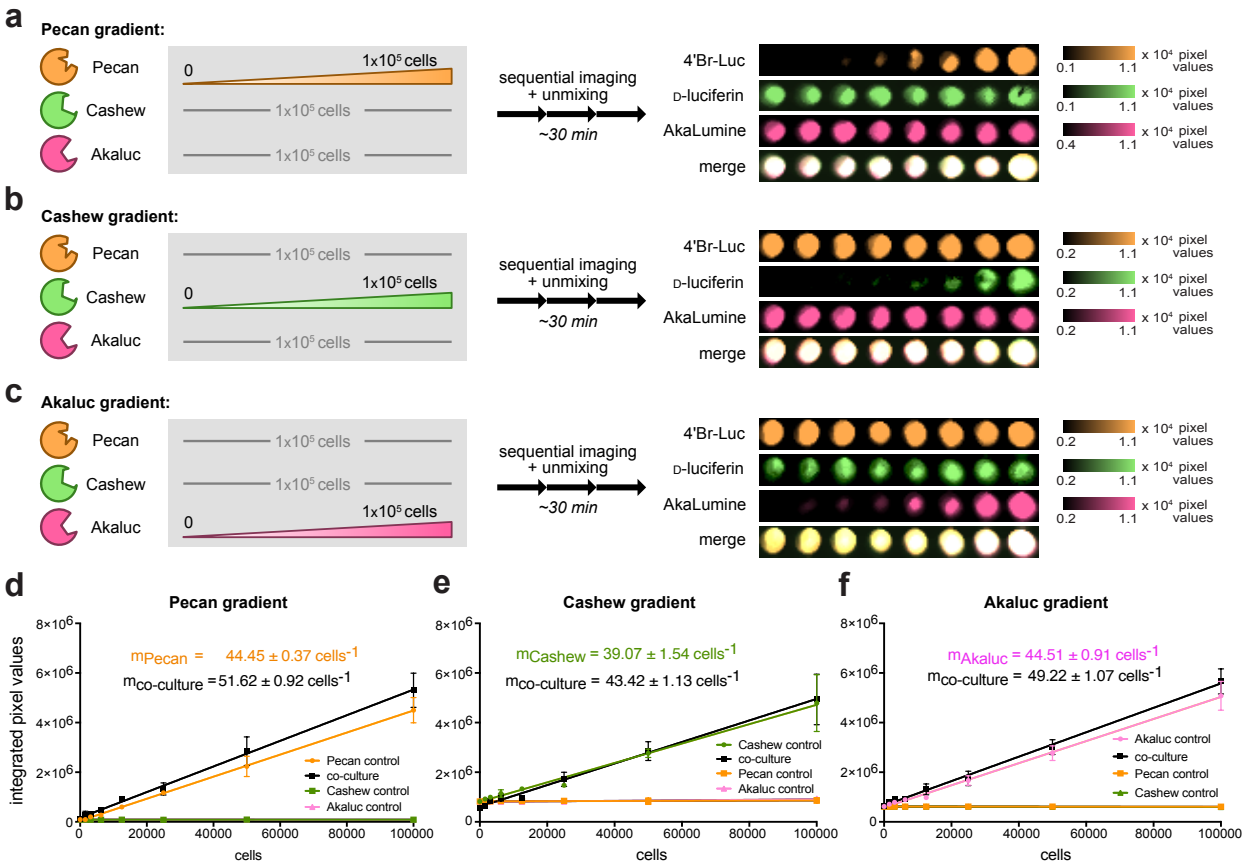


**Figure 4.16: Unmixing varying concentrations of heterogeneous breast cancer cell lines.** Metastatic organ-derived MMTV-PyMT cells lines stably expressed the 3 luciferase constructs (MFP-Pecan; LN-Cashew; Lung-Akaluc). Cell mixtures were plated as shown. Increasing concentrations of (a) MFP-Pecan, (b) LN-Cashew, and (c) Lung-Akaluc cells were plated in a gradient fashion (from 0 cell to $1x10^5$ cells per well) while the concentrations of the other two remaining cell lines were kept constant. The corresponding substrates ([4′-BrLuc] = [D-luc] = [AkaLumine] = 100 µM) were administered, beginning with the dimmest luciferin. Images were acquired after each addition. Image acquisition was completed within 30 minutes. Images were unmixed using *SubstrateUnmixing* and false colored. Images of each individual luciferase expressing cell line and the merged images were shown.

Linear regression analyses were performed on channels for the controls and gradient wells for (d) MFP-Pecan, (e) LN-Cashew, and (f) Lung-Akaluc. $R^2$ values of the linear regression analyses and values of the slope were calculated for each control and co-culture gradients.

(d) MFP-Pecan: $R^2$ Pecan control = 0.99; $R^2$ co-culture = 0.99; $R^2$ Cashew control = 0.93; $R^2$ Akaluc control = 0.93; $m_{Cashew\ control} = 0.02 \pm 0.03$; $m_{Akaluc\ control} = -0.03 \pm 0.02$.

(e) LN-Cashew: $R^2$ Cashew control = 0.99; $R^2$ co-culture = 0.99; $R^2$ Pecan control = 0.88; $R^2$ Akaluc control = 0.8662; $m_{Pecan\ control} = 0.34 \pm 0.09$; $m_{Akaluc\ control} = 1.01 \pm 0.29$.

(f) Lung-Akaluc: $R^2$ Akaluc control = 0.99; $R^2$ co-culture = 0.99; $R^2$ Pecan control = 0.98; $R^2$ Cashew control = 0.91; $m_{Pecan\ control} = -0.07 \pm 0.004$; $m_{Cashew\ control} = -0.17 \pm 0.02$.

**Rapid, multicomponent BLI imaging using MMTV-PyMT cell lines *in vivo***

The MMTV-PyMT organ-derived cell lines are an invaluable tool to study cancer progression and organ-tropic metastatic. Previously, I demonstrated that MMTV-PyMT MFP-derived cells are capable of recapitulating metastatic progression in disease free mice [20]. In order to use the MTMV-PyMT cell line models for future studies, it is critical to validate the ability of the organ-derived cell lines to recapitulate metastatic disease and re-colonize their organ-tropic tissues. To validate the organotrophic properties of the metastatic MMTV-PyMT-luciferase expressing cell lines, I investigated whether LN-derived Cashew-mNeptune (Figure 4.17a) and lung-derived Akaluc-BFP (Figure 4.17b) cells could colonize their respective tissues in two separate disease-free FVB mice. To assess tumor growth, I injected 100,000 lung-Akaluc BFP or LN cashew-mNeptune cells bilaterally under the fourth mammary fat pad in separate four-week old mice Based on the timeline from similar *in vivo* tumor growth studies that I performed (data not shown), I allowed tumors to grow for ~5 weeks before imaging. Upon completion of animal imaging, lungs and lymph-node tissues for both animals were resected. Tissues were imaged again ex vivo with the appropriate luciferins to confirm signal localization (Figure 4.17c). Upon ex vivo analysis, luciferase signal was observed in both mice and was localized to the correct organs. After AkaLumine substrate addition, Akaluc signal was observed in the lung tissue of mouse 2 (originally injected with lung-derived Akaluc-BFP cells) but was absent in the lung tissue of mouse 1 (originally injected with LN-derived Cashew-mNeptune cells). The reverse trend was observed upon D-luc administration to the lymph node tissue samples. Thus, ex vivo analysis confirmed that the signal was derived specifically from the organ-tropic metastatic tissue (Figure 4.17c). From these results we concluded that the cells lines emitted light properly *in vivo* and metastasized to the expected organs.

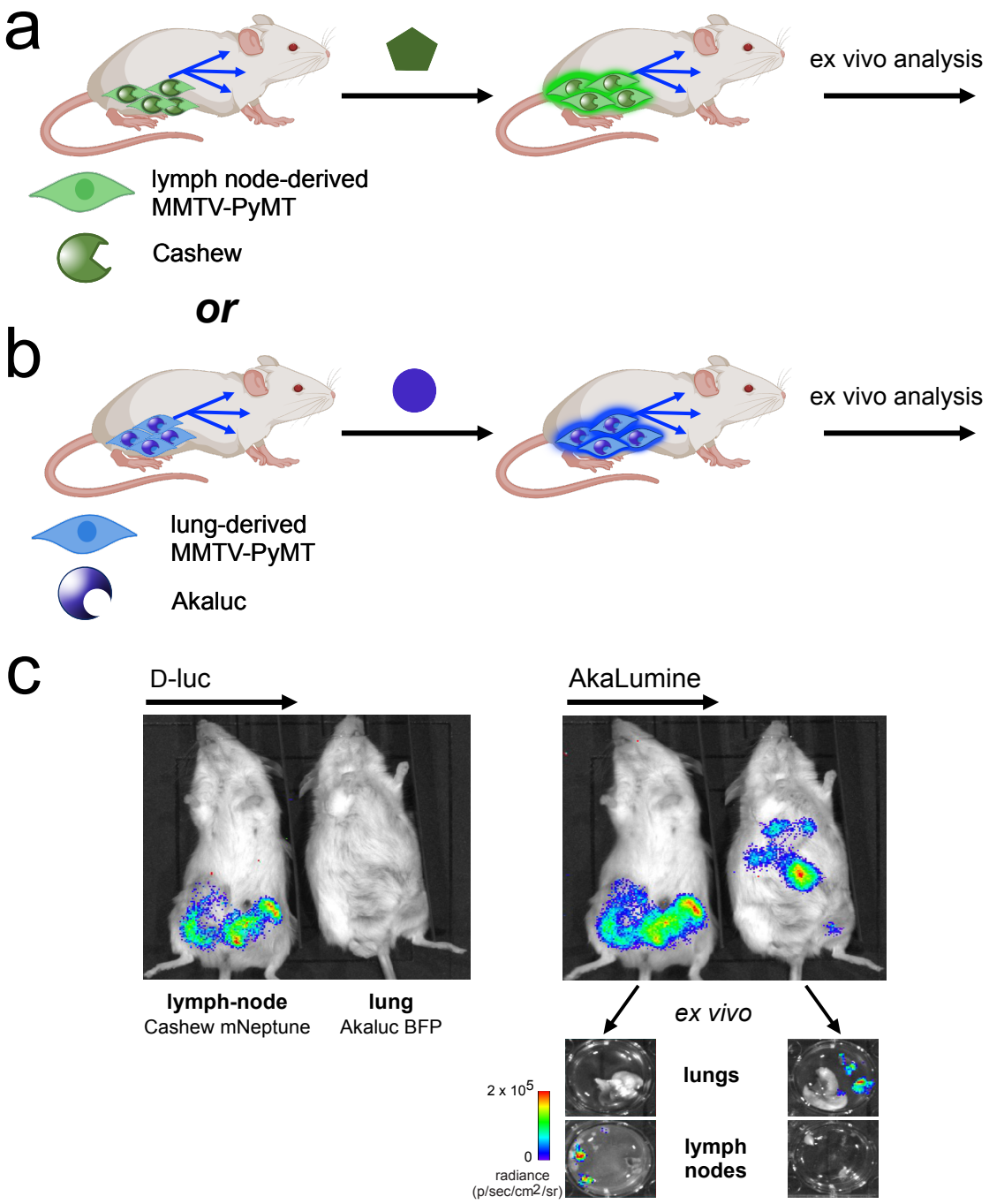


**Figure 4.17. Examining lymph node and lung metastases using organ-derived luciferase-expressing metastatic clonal isolates.** Lymph node- and lung-derived orthogonal luciferase expressing cells recapitulate metastatic disease. Luciferase-expressing cells were prepared via CRISPR gene insertion. The cells were implanted bilaterally into the mammary fat pad. Tumors were allowed to form over the course of 5 weeks. Mice were imaged upon luciferin injection with D-luc (65 mM) and/or AkaLumine (30 mM). (c) Lymph node and lung tissues from the recipient mice were extracted and imaged ex vivo.

AkaLumine has solubility issues when dissolving it at the high concentrations required for *in vivo* use. Additionally, luciferins, like AkaLumine, can be administrated via different routes into the animal (ie IP vs IV administration) [15]. Differences in the mode of luciferin administration is known to impact the bioavailability of the substrate and the brightness of the luciferase signal [13, 15]. To achieve the maximal light emission, I aimed to identify the optimal vehicle to resuspend the luciferins and the best route of administration *in vivo* (Figure 4.18a). Two mice were injected with lung-derived Akaluc-BFP cells and LN-derived Cashew-mNeptune cells. Luciferins were dissolved in Ringer's solution (D-luc) or $dH_20$ (AkaLumine). For mouse 1, both luciferins were administered via IP injection. For mouse 2, D-luc was administered IP while AkaLumine was administered IV. As previously reported [15], signal from Akaluc-expressing cells was observed in both mice (Figure 4.18b). However, AkaLumine signal was brighter when administered IP instead of IV. Thus, the best route of AkaLumine administration for these and future experiments is via IP injection. Additionally, this experiment tested the few different clonal isolates (denoted by #1, #2, etc.) for both LN and lung cells. The goal here being to identify the brightest clone and continue with that cell line for all future studies. After IP administration of D-luc, LN Cashew expressing cells from clone #1 exhibited the brightest light emission. Both lung Akaluc expressing cells had relevantly the same light emission. Future experiments will continue using these luciferase clones and injection methods for *in vivo* imaging.

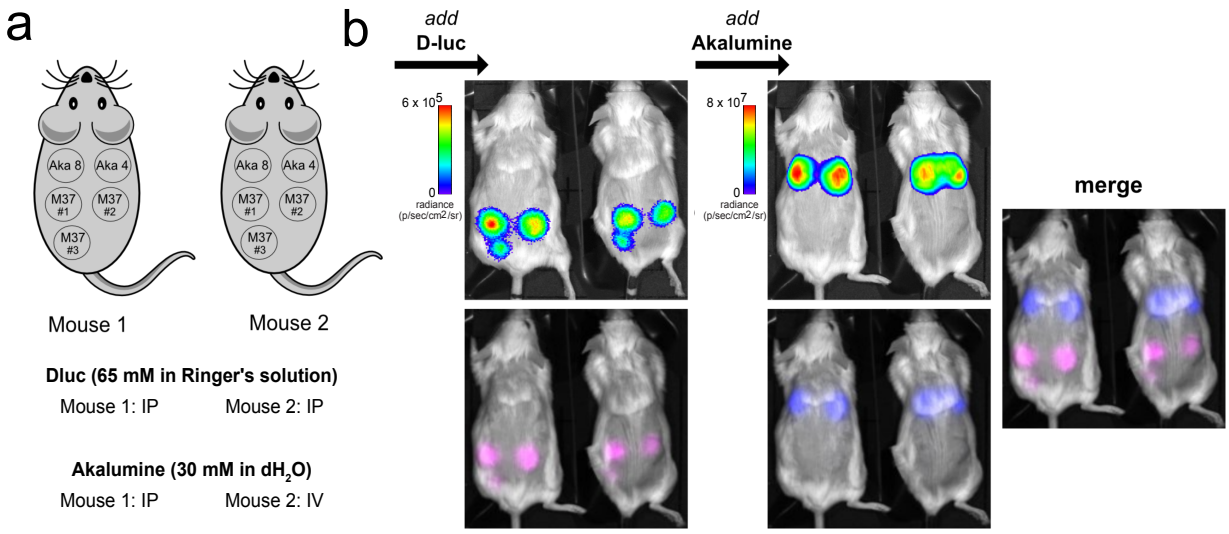


**Figure 4.18. Piloting luciferin administration for optimal signal *in vivo*.** (a) Schematic of the tumors injected and the differences in luciferin administration that was tested for LN-derived Cashew (M37) mNeptune and Lung-derived Akaluc (Aka) BFP MTTV-PyMT cell lines. D-luc (65 mM in Ringer's solution) was administered IP for both mice. AkaLumine (30 mM in water) was administered IP for mouse 1 and IV for mouse 2. Different clones for cell lines LN-derived Cashew (M37) mNeptune and Lung-derived Akaluc (Aka) BFP are noted on the diagram by the numbers. (b) Raw images (shown in radiance) were acquired with sequential luciferin administration. Cashew and Akaluc signals were unmixed and false colored in the panels below. Side panel shows the merged image.

## III. Conclusion

In this chapter I demonstrated that rapid, triple component BLI was achieved based on substrate unmixing. Sequential substrate administration was achieved by layering in the signal from dimmest to brightest. This work was the first example of rapid BLI using all engineered luciferases. In addition, I provided different examples of multicomponent bioluminescent imaging with terrestrial luciferin/luciferase pairs in tandem with marine pairs. Marine luciferins are known to have a higher rate of auto-oxidation *in vivo* which can provide additional variables. The autoxidation and lack of stability of marine luciferins frequently requires the need for IV instead of IP administration [13]. Compared to the ease of administering IP injections, IV administration

requires additional technical skills and is a less “user-friendly” method for delivering substances *in vivo* [15]. Our imaging platform mitigates against the variability in substrate delivery and bioavailability by utilizing only firefly luciferase derived mutants and administering their D-luciferin analogs by IP injections.

Future work should address the issues of benchmarking BLI output to fluorescence signals and cell count. Completing this experiment is important to demonstrate that our tool is a reliable and robust method for monitor biological processes. Overall, this technology has vast ranges of applications and showcases its power in the imaging community. Further expanding the bioluminescent tool box for multicomponent imaging will only increase the diverse biological questions this technology can address.

## MATERIALS AND METHODS

### Reagents

All reagents purchased from commercial supplies were of analytical grade and used without further purification. 4'-BrLuc was prepared and used as described previously described in Ionkina, et al. [7] Pecan-eGFP-AAVS1, Cashew-mNeptune-AAVS1, and Anatres-AAVS1, and Akaluc-TagBFP-AAVS1 were used as described in Ionkina, *et al*. [7]

### General bioluminescence imaging

All assays were performed in black 96-well plates (Grenier Bio One). Plates containing luminescent reagents were imaged in a light-proof chamber with an IVIS Lumina (Xenogen) CCD camera chilled to –90 °C. The stage was kept at 37 °C during the imaging session, and the camera was controlled using Living Image software. For all assays, exposure times were set to 1–180 s, and data binning levels were set to medium. Total flux values for regions of interest were analyzed using Living Image software. Integrated pixel values were analyzed using ImageJ (Installed under the FIJI package, NIH). The data were analyzed using GraphPad Prism (version 9.0 for Macintosh, GraphPad Software).

### Mammalian cell culture

DB7 cells stably expressing Pecan-eGFP, Cashew-mNeptune, or Akaluc-BFP were derived via CRISPR-mediated gene insertion as previously described [7]. MMTV-PyMT primary cell lines

(mammary fat pad (MFP), lymph node (LN), and lung) cells were derived from tumor bearing mice. For these experiments, organ-derived cell lines that were previously segregated by $CD44^{low}$/$EpCAM^{high}$ and $CD44^{high}$/ $EpCAM^{high}$ cell surface expression were mixed together. The combined CD44/EpCAM expressing organ-derived cell lines were allowed to grown until confluent. MMTV-PyMT (MFP, LN, and lung) cell lines stably expressing Pecan-eGFP, Cashew-mNeptune, or Akaluc-BFP, respectively, were derived via CRISRP-mediated gene insertion. The relevant luciferase genes (luciferase-G4SX2-FP-T2A-Puro) were amplified and inserted into CRISPR AAVS1-safe harbor donor plasmids.

Cashew and Pecan inserts were amplified using the following primers following primers:

5'- TGGCTAGCGCTACCGGTCGCCACCTCTAGAATGGAAGACGCCAAAAACATAAAGAAAGG -3' and 5'- GCGGAAAGATCGCCGTGGGCGGAGGCGGGTCTGGGGGCGGAGGCTCT -3.

Akaluc inserts were amplified with the following primers:

5'- ATGGAAGATGCCAAAAACATTAAGAAGGGCCCAGC -3' and

5'- CACGGCGATCTTGCCGTCCTTCTTGGCCTTAGTGA -3'

MMTV-PyMT cells lines were transfected with the AAVS1 mutant luciferase donor plasmid, Cas9 (Addgene #41815), and AAVS1 sgRNA (Addgene #53370) using lipofectamine. The mutant luciferases were integrated into the first locus of AAVS1 through homologous recombination. Transfected cells were then treated with puromycin (2 μg/mL) and FACS sorted at the Institute for Immunology Flow Cytometry Core (UCI). All cells were cultured in DMEM (Corning) supplemented with 10% (v/v) fetal bovine serum (FBS, Life Technologies), penicillin

(100 U/mL), and streptomycin (100 µg/mL). Stably expressing luciferase cells were maintained under puromycin selection (2 µg/mL) to ensure gene incorporation was preserved. Cells were maintained in a 5% $CO_2$ water-saturated incubator at 37 °C. Cells were serially passaged using trypsin-EDTA (0.25% in HBSS, Gibco).

**Mammalian cell culture**

MMTV-PyMT primary cell lines (MFP, LN, and lung) were isolated and sorted based on CD44 and EpCAM expression as previously described in Ionkina et al. [20] and chapter 3. Both $CD44^{low}/EpCAM^{high}$ and $CD44^{high}/ EpCAM^{high}$ cells were used for imaging. These cells or DB7 cells were engineered to express Pecan-eGFP, Cashew-mNeptune, or Akaluc-BFP via CRISPR-mediated gene insertion as previously described [7]. All cells were cultured in DMEM (Corning) supplemented with 10% (v/v) fetal bovine serum (FBS, Life Technologies), penicillin (100 U/mL), and streptomycin (100 µg/mL). Cells expressing luciferases were further cultured with puromycin (2 µg/mL) to ensure gene integration. All cells were maintained in a 5% $CO_2$ water-saturated incubator at 37 °C. Cells were serially passaged using trypsin-EDTA (0.25% in HBSS, Gibco).

**Substrate unmixing analysis**

Substrate unmixing experiments were designed as previously described [7]. Substrate unmixing was conducted with MATLAB R2020a (See Supplementary Discussion). Luminescence images containing the raw CCD counts (as TIFF files) were loaded into MATLAB. Images were subjected to a 2-pixel median filter (using the *medfilt2* function with a 5x5 neighborhood around the corresponding pixel). Next, the signal at each pixel was normalized to lie between 0 and 65536

(the maximum value that can be stored in a 16-bit image). As a result, the brightest pixel in each image had a value of 65536, and the dimmest had a value of 0. Regions of interest (ROIs) were generated by identifying the image coordinate of the reference well and input dimensions. Once assigned, the MATLAB algorithm was run to perform the unmixing. After unmixing, text images were imported into ImageJ (installed under the FIJI package). Integrated pixel values for regions of interest were analyzed using the "Measure" tool. Pseudo colors were assigned with the "Merge Channels" tool.

**Mammalian cell analysis of luciferase mutants**

DB7 or MMTV-PyMT cells stably expressing luciferases were added to black 96-well plates. Cells were treated with a luciferin solution (10 μL, of 100 μM or 250 uM for D-luc, 4′-BrLuc, and AkaLumine). Plates were imaged and analyzed as described above.

## References


1. Kaskova, Z.M., A.S. Tsarkova, and I.V. Yampolsky, *1001 lights: luciferins, luciferases, their mechanisms of action and applications in chemical analysis, biology and medicine.* Chem Soc Rev, 2016. **45**(21): p. 6048-6077.
2. Xu, T., et al., *The Expanding Toolbox of In vivo Bioluminescent Imaging.* Front Oncol, 2016. **6**: p. 150.
3. Yeh, H.W., et al., *Identification of Factors Complicating Bioluminescence Imaging.* Biochemistry, 2019. **58**(12): p. 1689-1697.
4. Daniel, C., et al., *Dual-Color Bioluminescence Imaging for Simultaneous Monitoring of the Intestinal Persistence of Lactobacillus plantarum and Lactococcus lactis in Living Mice.* Appl Environ Microbiol, 2015. **81**(16): p. 5344-9.
5. Mezzanotte, L., et al., *Sensitive dual color in vivo bioluminescence imaging using a new red codon optimized firefly luciferase and a green click beetle luciferase.* PLoS One, 2011. **6**(4): p. e19277.
6. Gregor, C., et al., *Autonomous bioluminescence imaging of single mammalian cells with the bacterial bioluminescence system.* Proc Natl Acad Sci U S A, 2019.
7. Rathbun, C.M., et al., *Rapid Multicomponent Bioluminescence Imaging via Substrate Unmixing.* ACS Chem Biol, 2021. **16**(4): p. 682-690.
8. Jones, K.A., et al., *Orthogonal Luciferase-Luciferin Pairs for Bioluminescence Imaging.* J Am Chem Soc, 2017. **139**(6): p. 2351-2358.
9. Maguire, C.A., et al., *Triple bioluminescence imaging for in vivo monitoring of cellular processes.* Mol Ther Nucleic Acids, 2013. **2**: p. e99.
10. Chu, J., et al., *A bright cyan-excitable orange fluorescent protein facilitates dual-emission microscopy and enhances bioluminescence imaging in vivo.* Nat Biotechnol, 2016. **34**(7): p. 760-7.
11. Oh, Y., et al., *An orange calcium-modulated bioluminescent indicator for non-invasive activity imaging.* Nat Chem Biol, 2019. **15**(5): p. 433-436.
12. Yeh, H.W. and H.W. Ai, *Development and Applications of Bioluminescent and Chemiluminescent Reporters and Biosensors.* Annu Rev Anal Chem (Palo Alto Calif), 2019. **12**(1): p. 129-150.
13. Stacer, A.C., et al., *NanoLuc reporter for dual luciferase imaging in living animals.* Mol Imaging, 2013. **12**(7): p. 1-13.
14. Hall, M.P., et al., *Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate.* ACS Chem Biol, 2012. **7**(11): p. 1848-57.
15. Iwano, S., et al., *Single-cell bioluminescence imaging of deep tissue in freely moving animals.* Science, 2018. **359**(6378): p. 935-939.
16. Sarrion-Perdigones, A., et al., *Examining multiple cellular pathways at once using multiplex hextuple luciferase assaying.* Nat Commun, 2019. **10**(1): p. 5710.
17. Kuchimaru, T., et al., *A luciferin analogue generating near-infrared bioluminescence achieves highly sensitive deep-tissue imaging.* Nat Commun, 2016. **7**: p. 11856.
18. Kitada, N., et al., *Toward bioluminescence in the near-infrared region: Tuning the emission wavelength of firefly luciferin analogues by allyl substitution.* Tetrahedron Letters, 2018. **59**(12): p. 1087-1090.

19. Bai, Y., et al., *Dynamic multicomponent engineered tissue reorganization and matrix deposition measured with an integrated nonlinear optical microscopy-optical coherence microscopy system.* J Biomed Opt, 2014. **19**(3): p. 36014.
20. Ionkina, A.A., et al., *Transcriptome analysis of heterogeneity in mouse model of metastatic breast cancer.* Breast Cancer Res, 2021. **23**(1): p. 93.

# CHAPTER 5: Examining immune cell function and cellular interactions during metastatic progression with multicomponent imaging tools *

## I. Introduction

This chapter presents my efforts and preliminary work developing split versions of red-emitting luciferases to monitor cell-to-cell contact. Additionally, I highlight my work generating cell specific promoter driven luciferase gene reporters. These probes facilitate imaging of tumor-immune and other cellular interactions in deep tissue. Both of these research topics are areas of on-going work in the laboratory and are the future directions of my dissertation work – with the ultimate goal of deploying these probes in transgenic models.

## II. Monitoring cellular interactions and immunological reporters using rapid bioluminescent imaging platforms

Rapid, multicomponent bioluminescence imaging is a powerful tool that can have broad applicability in cancer biology. As mentioned in Chapter 1, immune cells in the surrounding microenvironment have been identified as key players in cancer progression [1,2]. Cancer-immune responses in particular involve a wide array of different immune cell types [3]. In order to capture this heterogeneity, multicomponent imaging tools that enable the visualization of the different immune cell subtypes and polarization states are needed. Toward, this end I have been creating

* *Author's Note: This chapter has unpublished ideas and preliminary data that I have spearheaded. I contributed to the experimental design, procedures, data analyses, and writing. This is an ongoing area of research by members of the Prescher laboratory.*

two different types of luciferase driven reporters (split luciferase reporters and luciferase driven gene expression reporters) that will be discussed in more detail below. The initial application of these bioluminescent reporters will be for *in vitro* use and validation. However, the facile use of the rapid bioluminescent imaging platform outlined previously shows great promise that these gene expression reporters can be translated for *in vivo* use.

## III. Visualize cellular interactions using with engineered bioluminescent probes

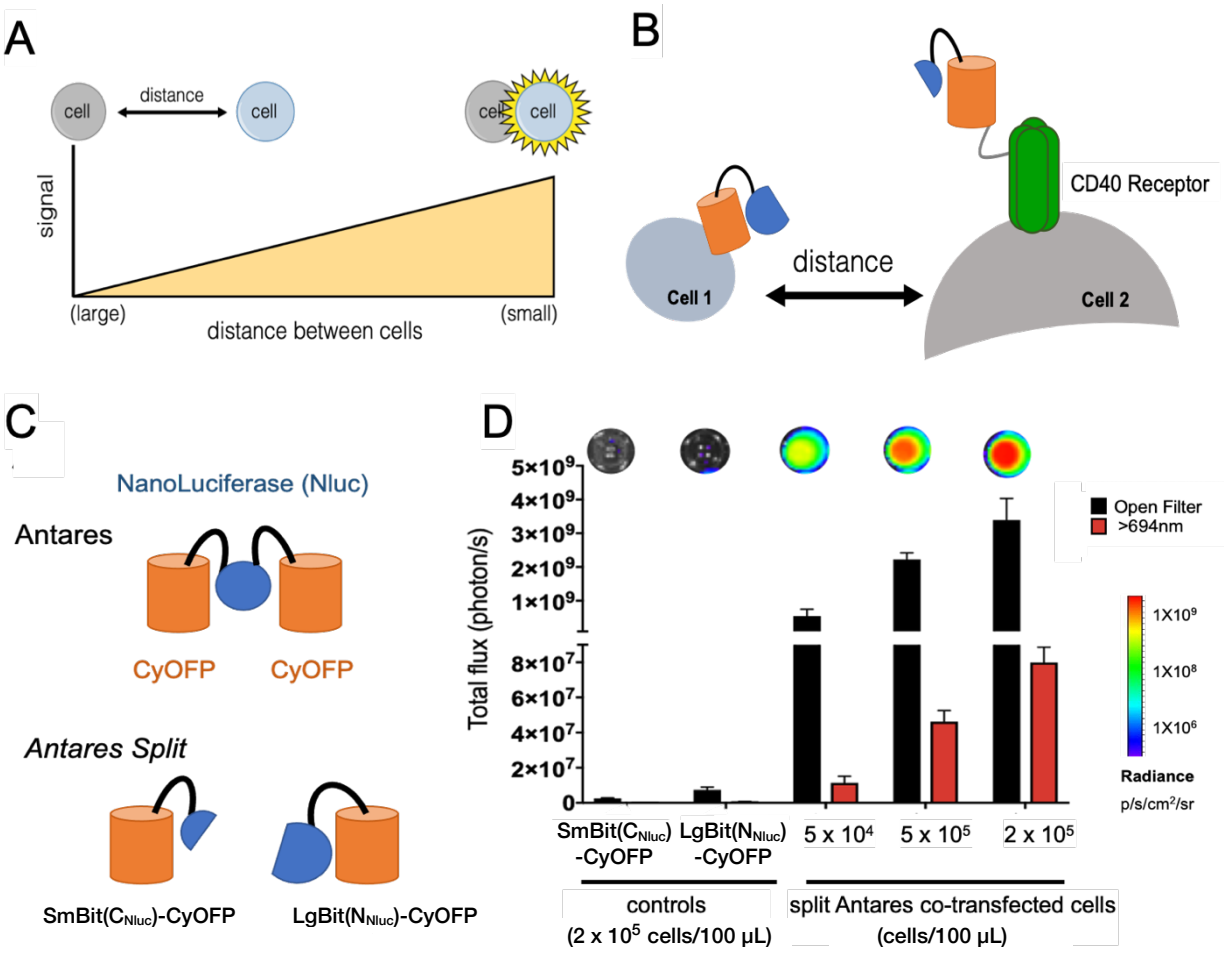


**Figure 5.1. Visualizing cell-cell contact through engineered luciferases.** (A) Schematic of general strategy to visualize cellular interaction. Luciferase expressing cells emit light when in close proximity. (B) Visualizing cell-cell contact through CD40 receptor mediated reassembly of split luciferase. In close proximity, cells expressing either a membrane bound (CD40 receptor) or secreted split construct reassemble, forming a functional luciferase capable of producing light. (C) Schematic of Antares BRET split construct. (D) The ability of split Antares to complement and emit light was tested. Control wells of either SmBit($C_{Nluc}$)-CyOFP or LgBit($N_{Nluc}$)-CyOFP fragments were plated. HEK-293T cells were then co-transfected with both SmBit($C_{Nluc}$)-CyOFP and LgBit($N_{Nluc}$)-CyOFP splits. Furimazine was added and emission was measured. Co-transfected cells were platted at the increasing cell concentrations. Total flux and red-shifted light >694 nm was measured using a filter. Total flux was measured in radiance. Corresponding images of the wells are shown above each measured cell concentration.

Cell-to-cell contact mediates numerous biological processes, including cellular growth and differentiation, motility, and immune function [3]. Breakdowns in these networks can also potentiate disease, including cancer progression [1]. Our ability to understand and ultimately control cellular interactions would benefit from methods to visualize cells in their native environments. While advances have been made *in vitro* [9], few tools exist for long-term, sensitive imaging in whole animals. This void has precluded studies of dynamic cellular interactions that play crucial roles in tumorigenesis, tumor-immune cell interactions, metastasis, and other processes.

One way that immune-cancer cell interactions can be monitored is through split luciferase reporters. Proximal cells expressing the split constructs drive complementation, producing a functional luciferase and light emission (Figure 5.1A). Although previous split versions of blue-emitting luciferases have been successfully developed to monitor cell contact, they cannot be easily translated *in vivo* [4]. These probes do not emit sufficient levels of red photons for deep tissue imaging. Initially I sought to address the need for more red-shifted light by creating a split Antares probes (Figure 5.1C,B) [5]. Antares is a luciferase constructs that harness bioluminescence resonance energy transfer (BRET) to red-shift the light emission and improve the sensitivity of deep tissue imaging. Antares consists of a central Nluc with two CyOFP fluorescent proteins that red-shift the original blue light emitted by Nluc in the presence of its substrate [5] (Figure 5.1C top panel). The Antares construct enables BRET and results in the 460 nm wavelength of light emitted by Nluc to red-shift, emitting light at 584 nm, making Antares an attractive option for deep tissue imaging [5].

My initial approach involved developing split Antares constructs that when held in close proximity, the split fragments can recombine to form a functional enzyme (Figure 5.1B). By

creating split Antares constructs tethered to membrane receptor-ligand pairs, I aimed to track cellular interactions in deep tissues. I first split Antares through Nluc with CyOFP proteins on either side of the fragments. The smaller portion of the split Nluc was denoted as SmBit($C_{Nluc}$)-CyOFP and the larger portion of the split Nluc LgBit($N_{Nluc}$)-CyOFP. Complementation of the two split Antares components through the split SmBit and LgBit would produce the red-shifted light [5]. I used GIBSON assembly and standard PCR protocols to create the split Antares reporters (Figure 5.1C) and the split Antares – tethered reporters. I used CRISPR-Cas9 editing technology to transfect the various split Antares fragments into cell lines. Interactions of the two split-expressing cells would facilitate the production of red light upon cell contact, and thus enable sensitive imaging of cellular interactions. I hypothesized that the BRET-based technology in split reporter constructs utilizing Antares would enable improved, macroscopic imaging of cell-cell interactions. Close cellular contact of the split reporter was thought to produce a functional enzyme capable of BRET, resulting in red-shifted light emission which increases imaging sensitivity and tissue penetrance.

Split Antares reporters were successfully created and shown to produce sufficient signal in HEK-293T cells upon complementation (Figure 5.1D). HEK-293T cells expressing either SmBit($C_{Nluc}$)-CyOFP or LgBit($N_{Nluc}$)-CyOFP fragments were used as controls. To identify maxiumal light emission with the split Antares system, HEK-293T were then co-transfected with both SmBit($C_{Nluc}$)-CyOFP and LgBit($N_{Nluc}$)-CyOFP splits and plated in 96-well plates at increasing cell concentrations. Total flux (open filter) and red-shifted light >694 nm were measured in radiance after furimazine addition to each well. Gratifyingly, increasing the number

of co-transfected SmBit($C_{Nlu}$)-CyOFP and LgBit($N_{Nlu}$)-CyOFP-expressing cells resulted in an increase in light emission.

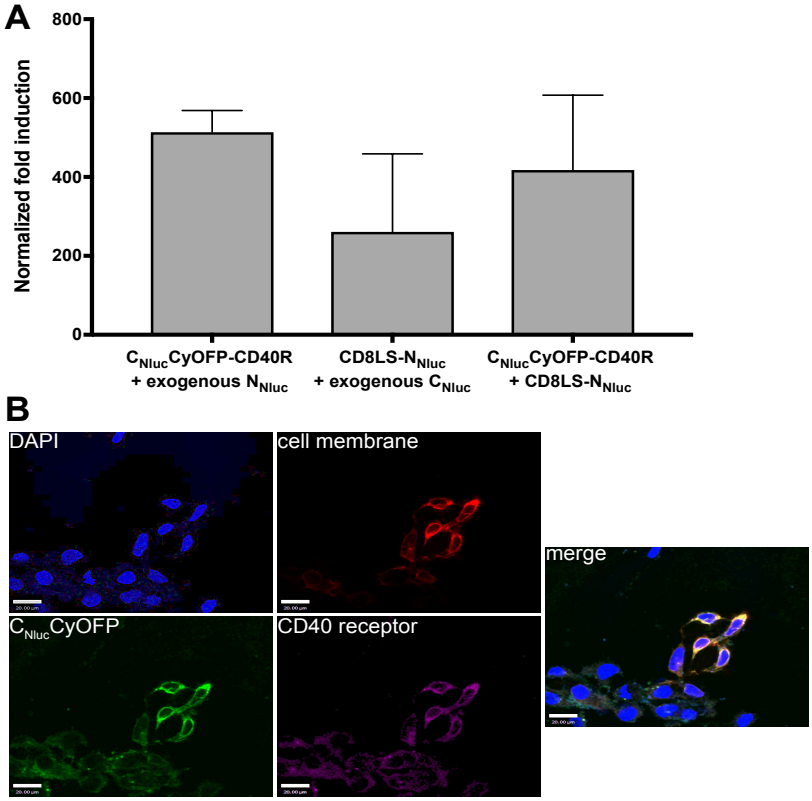


**Figure 5.2. CD40 receptor linked to SmBit($C_{Nluc}$)-CyOFP could be secreted to the membrane cell surface**. (A) HEK-293T stably expressing CD40 receptor linked SmBit($C_{Nluc}$)-CyOFP imaged with stably expressing LgBit($N_{Nluc}$) (CD8LS secreted) cells emitted light. Light emission was observed upon addition of exogenous SmBit($C_{Nluc}$) to stably expressing HEK-293T cells secreting LgBit (CD8LS). Light emission was observed upon co-culturing CD40 receptor linked SmBit($C_{Nluc}$)-CyOFP cells with cells secreted LgBit($N_{Nluc}$) and was brighter than the addition of exogenous SmBit($C_{Nluc}$) to LgBit($N_{Nluc}$) expressing cells. Light emission was observed after media transfer from HEK-293T cells secreting LgBit($N_{Nluc}$) to HEK-293T cells expressing CD40 receptor linked SmBit($C_{Nluc}$)-CyOFP. (B) Confocal images of CD40 receptor bound SmBit($C_{Nluc}$)-CyOFP cells (imaged with membrane stain and CD40 receptor antibody) showed co-imaging on cell surface. Addition of exogenous LgBit($N_{Nluc}$) produced light.

Having verified complementation, I next investigated membrane-bound and secreted variants for monitoring cell contact [4]. To enable *in vivo* monitoring of cell interactions, I developed tethered-versions of the split Antares reporters using GIBSON assembly and standard PCR protocols (Figure 5.1B). I used CRISPR-Cas9 editing technology to transfect the various CD40 receptor or CD40 ligand split Antares fragments into HEK-293T cell lines. Early in my efforts, I identified that split Antares tethered to CD40 ligands had difficultly secreting from the cells (data not shown).

Split Antares reporters without CD40-ligands were created simultaneously to mitigate this issue.

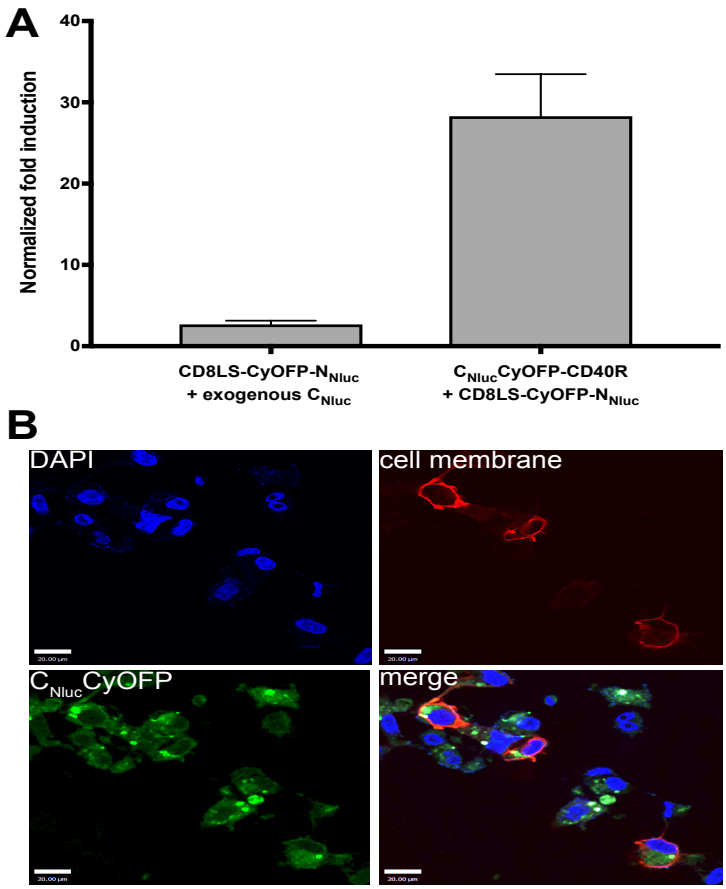


**Figure 5.3. HEK-293T cells stably expressing CD40 receptor linked to SmBit($C_{Nluc}$)-CyOFP imaged with LgBit-CyOFP (secreted) cells did not have a significant increase in light emission**. (A) Exogenous LgBit($N_{Nluc}$) added to CD40 receptor linked to SmBit($C_{Nluc}$)-CyOFP emitted light. Exogenous SmBit added to LgBit-CyOFP (secreted) cells did not emit light. Media transfer from LgBit($N_{Nluc}$)-CyOFP (secreted) cells to CD40 receptor linked SmBit($C_{Nluc}$)-CyOFP cells showed no light emission showed no light emission. Co-culturing CD40 receptor linked SmBit($C_{Nluc}$)-CyOFP cells and LgBit($N_{Nluc}$)-CyOFP (secreted) cells showed no light emission. Co-transfecting HEK-293T cells with CD40 receptor linked SmBit($C_{Nluc}$)-CyOFP and LgBit($N_{Nluc}$)-CyOFP (secreted) constructs produced light emission. (B) Confocal microscopy also concluded that LgBit($N_{Nluc}$)-CyOFP could not be secreted by cells.

The split Antares probes linked to CD40 receptor were identified as being fully and functionally expressed on the cell surface (Figure 5.2, 5.3, and 5.4). Contact was assessed through luciferase light emission using the IVIS imaging system. Cell localization and secretion of the split

constructs were assessed through confocal microscopy imaging and co-localization with cell membrane stains, internal transport vesicles, or CD40-flourescently labeled conjugated antibodies (Figure 5.2B, 5.3B, 5.4B).

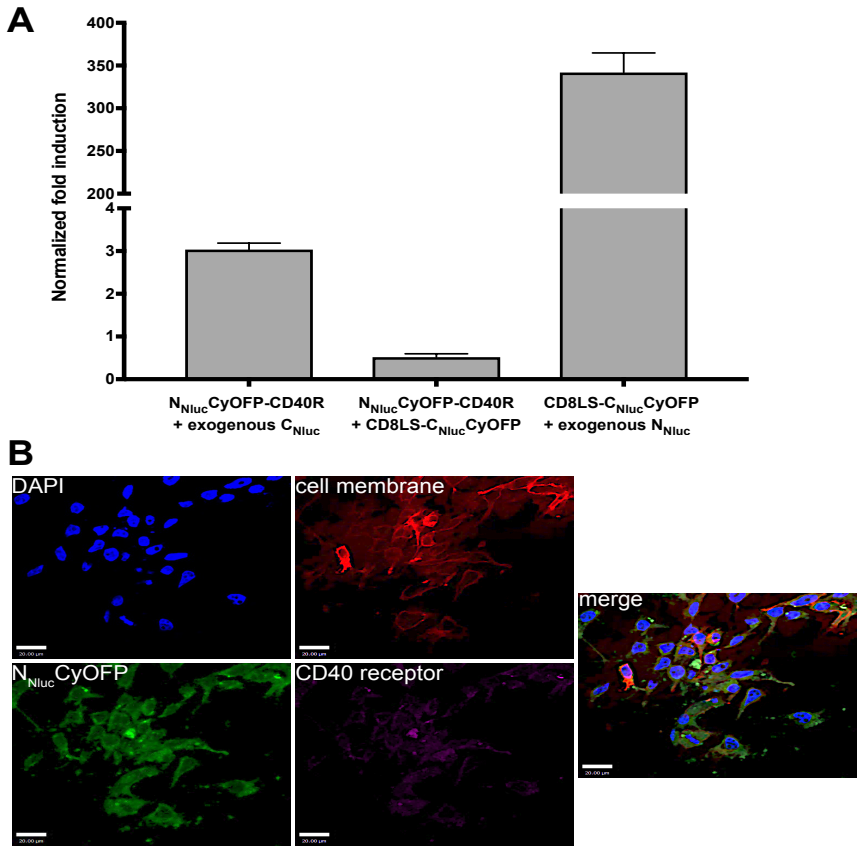


**Figure 5.4. CD40 receptor linked to LgBit($N_{Nluc}$)-CyOFP could not be secreted out of cell.** (A) Light emission was tested for HEK-293T cells stably expressing CD40 receptor-LgBit($N_{Nluc}$)-CyOFP bound to the membrane with HEK-293T stably expressing secreted SmBit($C_{Nluc}$)-CyOFP cells. No significant light emission observed when CD40 receptor-LgBit($N_{Nluc}$)-CyOFP membrane bound cells were co-cultured with SmBit($C_{Nluc}$)-CyOFP secreting cells. No significant light emission observed upon media transfer of SmBit($C_{Nluc}$)-CyOFP secreting cells to CD40 receptor-LgBit($N_{Nluc}$)-CyOFP membrane bound cells. No significant increase in light emission observed with addition of exogenous SmBit($C_{Nluc}$) added to CD40 receptor linked LgBit($N_{Nluc}$)-CyOFP cells. Light observed for exogenously added LgBit($N_{Nluc}$) to SmBit($C_{Nluc}$)-CyOFP secreting cells. Light observed in HEK-293T cells co-transfected HEK-293T cells with CD40 receptor-LgBit($N_{Nluc}$)-CyOFP bound to the membrane and SmBit($C_{Nluc}$)-CyOFP secreted constructs. (B) Confocal images of CD40 receptor bound LgBit($N_{Nluc}$)-CyOFP cells (co-imaged with membrane stain, CD40 receptor antibody) did not show co-localization of LgBit($N_{Nluc}$)-CyOFP to the membrane. Instead LgBit($N_{Nluc}$)-CyOFP CD40 receptor was observed to be inside of cell, possibly in secretory compartments of the cell.

Although the split Antares constructs were successful at detecting cellular interactions *in vitro*, the percentage of red-shifted photon output that was achieved was not sufficient for *in vivo* applications. To improve the applicability of the split luciferase technology to study cellular interactions, we sought alternative red-shifted luciferases that were better suited for *in vivo* use. I selected a red-shifted mutant of firefly luciferase (Akaluc) recently reported to give off >650 nm light [6]. Akaluc is among the most sensitive bioluminescent probes reported to date, and has been successfully applied for imaging nerve cells in live marmosets. I aimed to create a split version of Akaluc that is capable of reconstituting in the extracellular environment to visualize cellular contacts relevant to cancer biology (Figure 5.5A).

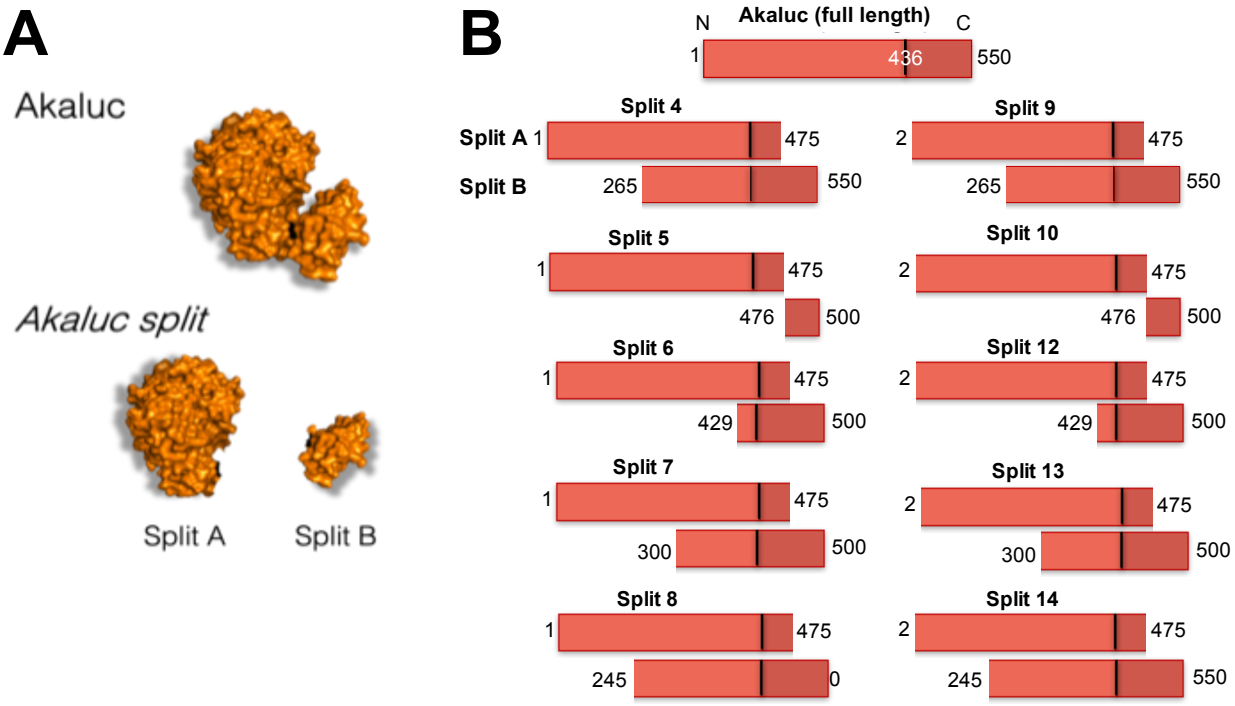


**Figure 5.5. Development of novel split luciferase to image cellular interactions in deep tissue.** (A) Akaluc is an evolved variant of Fluc (orange structure). Split versions of Akaluc were created based on a model of Fluc. (B) DNA constructs for Akaluc (amino acids 1-550), along with initially targeted split proteins (split A and split B). Amino acids 436 lies in flexible linker regions between the 2 termini, as assigned with comparison to Fluc.

Constructs comprising the Akaluc fragments were then generated via overlap PCR and confirmed via sequencing (Figure 5.5B). Split fragments were tethered to either a CD40 receptor or ligand to drive complementation (Figure 5.6). Successful complementation of split Akaluc fragments resulted in the production of light upon substrate administration. This light was measured and used to evaluate the split fragment sites. No crystal structure of Akaluc was available to assist in identifying feasible split sites. Since Akaluc is a mutant of Fluc [6], the Fluc crystal structure and known successful split Fluc reporters were used to guide the selection of targets [7, 8]. Based on this previous data [9], I hypothesized that Akaluc fragments that overlapped would produce increased light emission compared to non-overlapping fragments. It is likely that the overlapping regions promote stabilization and complementation of the assembled protein. Thus, the designed split fragments incorporated varying degrees of overlap to identify the ideal combination (Figure 5.5B).

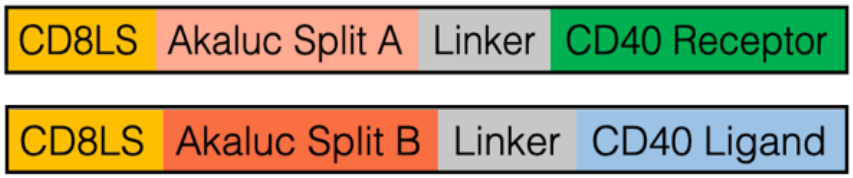


**Figure 5.6. DNA plasmid design of CD40 receptor-ligand split Akaluc fragments**. DNA constructs for imaging cellular interactions comprise Akaluc splits tethered to CD40 receptor or ligand, a linker region, and a cellular secretion CD8 leader sequence (CD8LS).

Light output from full length Akaluc-expressing cells was used as a benchmark to assess the various split Akaluc fragments. All Akaluc plasmids were designed with a direct fusion to a fluorescent protein for future ex vivo analysis. Total flux (photons/second) from HEK-293T cells co-transfected with both fragments was measured along with the full length protein after AkaLumine addition (Figure 5.7). As mentioned in Chapter 1, terrestrial luciferase-luciferin pairs

require ATP for proper light emission (Figure 5.7 light blue bars). The possible effect that additional ATP could have on the light production of Akaluc splits was tested by adding two 1 mM boluses of ATP (2 mM total) (Figure 5.7 dark blue bars). Preliminary results suggest that combinations of the split fragments can restore in bioluminescent signal, suggesting the feasibility of this approach (Figure 5.7). Split 8 produced the brightest total flux emission of $4.5x10^4$ photons/second compared to the dimmest splits (splits 4 and 5) producing $1x10^3$ photons/second.

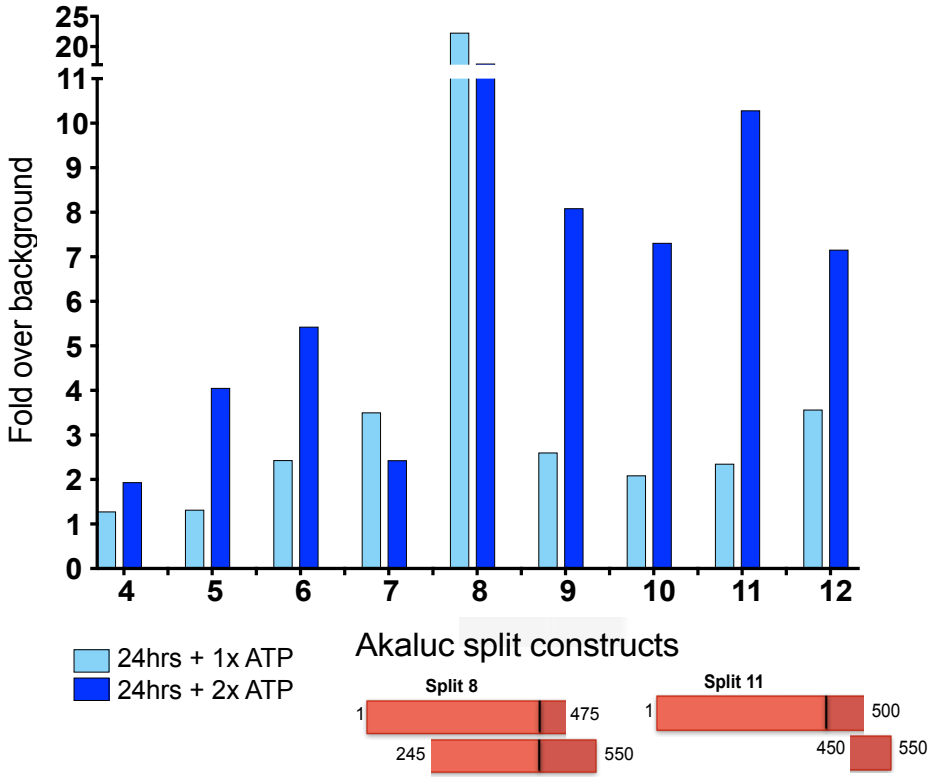


**Figure 5.7. Light emission from CD40 receptor-ligand driven re-complementation of tethered split Akaluc fragments.** Akaluc splits (A and B) for constructs 4 – 12 were co-transfected into HEK-293T cells. Light emission upon AkaLumine and ATP (1x = 1 mM total concentration) addition was measure for each split combination after 24 hours. Additional ATP (2x = 2 mM total concentration) was added and changes in light emission were measured and compared to the single dose of ATP. Split 8 was highlighted due to its high light emission ($4.5x10^4$ photons/second). Split 11 was highlighted due to its elevated light emission upon a second dose of ATP ($7x10^3$ photons/second). The construct design and overlap regions for both splits are noted below.

Other interacting domains consisting of tightly binding proteins have been shown to promote complementation and light output of tethered split fragments [8]. For example, coiled coils are commonly used to facilitate connections (Figure 5.8) [4], in addition to the extracellular CD40 receptor/ligand pair. To investigate this possibility, I created a suite of plasmids that drive Akaluc complementation via coiled-coil interactions (Figure 5.8A). These splits are currently being examined in HEK-293T cells (Figure 5.8B). The full characterization and functionality of these splits is an area of on going work.

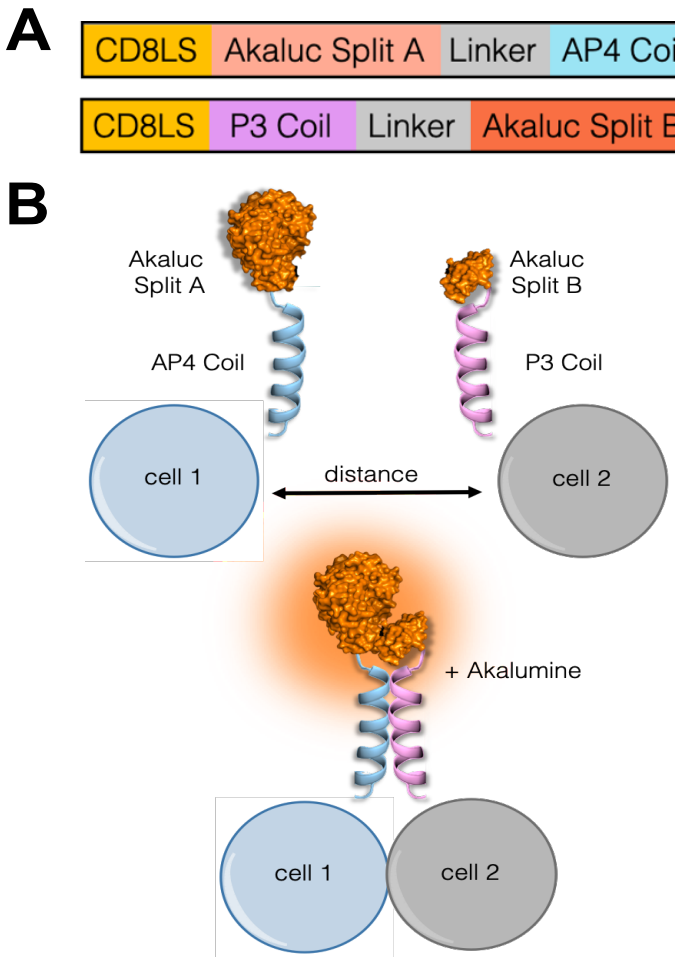


**Figure 5.8. Coiled-coil drives re-complementation of split Akaluc fragments**. (A) Schematic of split Akaluc A tethered to an APA coil and split Akaluc B tethered to a P3 coil. (B) Graphical representation of cell secreting split A tethered to APA coils (cell 1) and split B tethered to P3 coils (cell 2). In close proximity, recombination of split Akaluc is facilitated and light emission is observed upon AkaLumine addition.

## IV. Multicomponent BL imaging reporters to monitor changes in immune cells during metastatic disease progression

As explained in chapter 1, genetically encodable bioluminescent and fluorescent proteins can be used to monitor gene expression when driven under cell specific promoter [10]. These reporters provide a readout on gene expression through light emission in the targeted cell populations. Fluorescent-based versions of gene reporters have been incredibly useful for tracking gene expression of cancer-associated immune cells *in vivo*, particularly in regulatory T cells (Tregs) (Figure. 1.7) [11]. In order to better understand cancer-immune interactions, I have leveraged the multicomponent bioluminescent imaging platform to create immune-cell specific luciferase reporters. The work I outline below has become an on-going area of research in the laboratory.

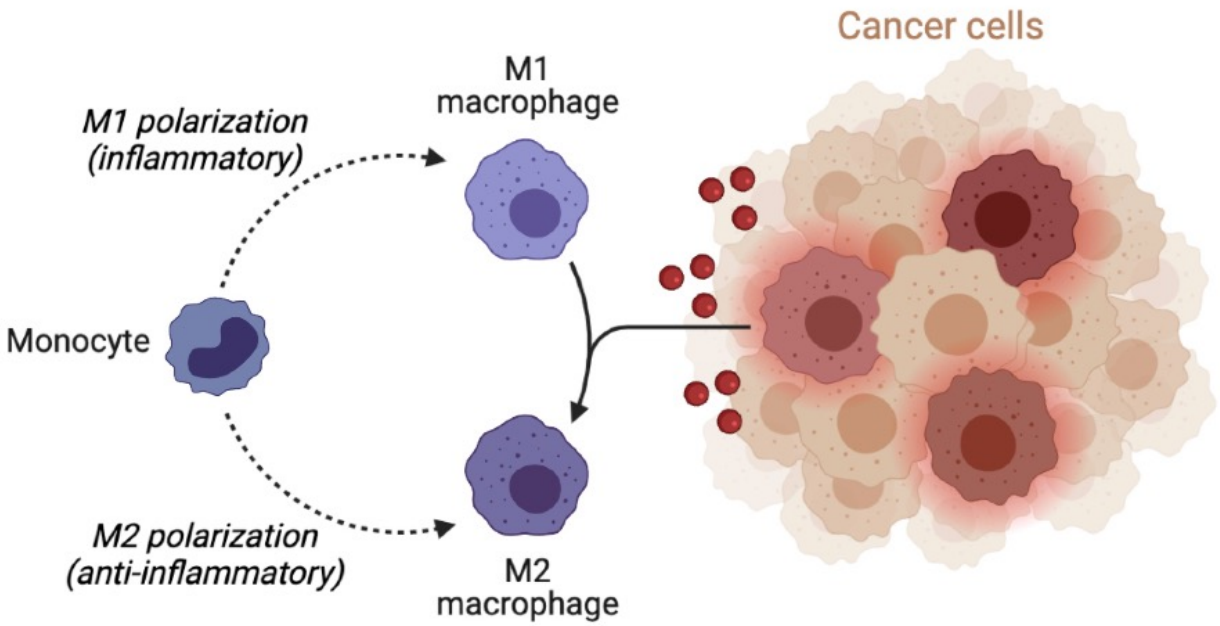


**Figure 5.9. Monocytes are polarized into M1 (inflammatory) or M2 (anti-inflammatory) cells that have different effects on tumor growth or tumor suppression.**

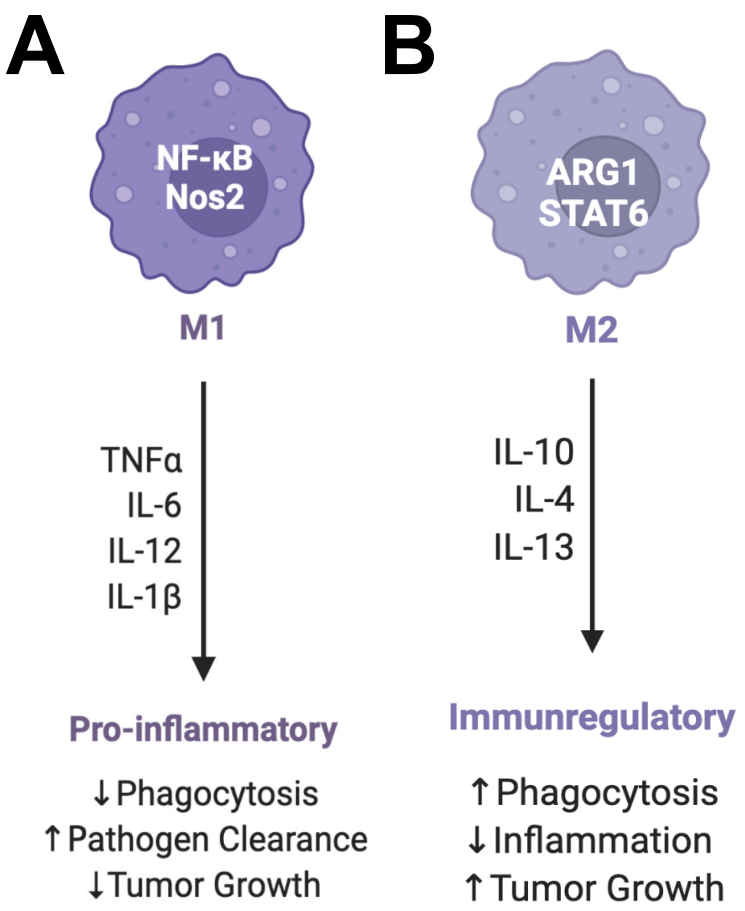


**Figure 5.10. Macrophage M1 and M2 polarization signatures**. (A) M1 macrophages are known to be pro-inflammatory, expressing NF-κB and NOS2. They effect their local environment through the section of cytokines, such as IL6. (B) M2 macrophages are known to be immunoregulatory (anti-inflammatory), expressing ARG1 and STAT6. They affect their local environments through the sections of cytokines, such as IL-4.

Previously described in Chapter 1, there are many different immune cells that can affect metastatic progression. Macrophages have been identified as key in promoting breast cancer progression [12]. In particular the MMTV-PyMT mouse model that has been used throughout this work is also an ideal model to study macrophage interaction during breast cancer metastasis [13]. Macrophage function is diverse and relies on cell surface expression in order to exert the proper effects to their surrounding environments. Macrophages have broad and complex functions in mediating response to metastatic disease progression [14]. Recent work has shown that macrophages function on a spectrum that still is not fully elucidated [15]. Currently, the standardized way to access macrophage function has been simplified to examining the two polar extremes of the immune cell's polarization via M1 and M2 phenotypes (Figure 5.9) [16]. M1 polarized macrophages exert a pro-inflammatory response generally thought to be beneficial for

attacking and clearance of pathogenic targets (Figure 5.10A). They have traditionally been associated with NF-κB and NOS2 signaling. M2 polarized macrophages exert an anti-inflammatory response [16]. These are the second wave of macrophages generally after pathogens are cleared to enable tissue healing (Figure 5.10B). They have traditionally been associated with Arg1 and STAT6 signaling.

Prolonged inflammatory signals are damaging to tissues. Normally M2 macrophages are recruited or polarized to alleviate inflammation and initiate the appropriate angiogenesis, nutrient transfer, and tissue regeneration [15]. In the context of metastatic disease progression, tumor cells that are not initially cleared by the immune system can manipulate the macrophages to promote metastatic growth [14]. In this situation, growing tumor cells can exert a pronounced and prolonged inflammatory signal that falsely triggers M2 macrophages that end up assisting tumor growth and suppressing anti-tumor immune cells (Figure 5.9).

Therefore, I focused the initial immune reporter designs to study M1 and M2 polarization of macrophages by constructing gene reporters using well vetted and commercially available plasmids (Figure 5.11A). NF-κB and NOS2 gene reporters were selected to monitor M1. Arg1 and STAT6 were selected to monitor M2 polarization. Promoters drive expression of different

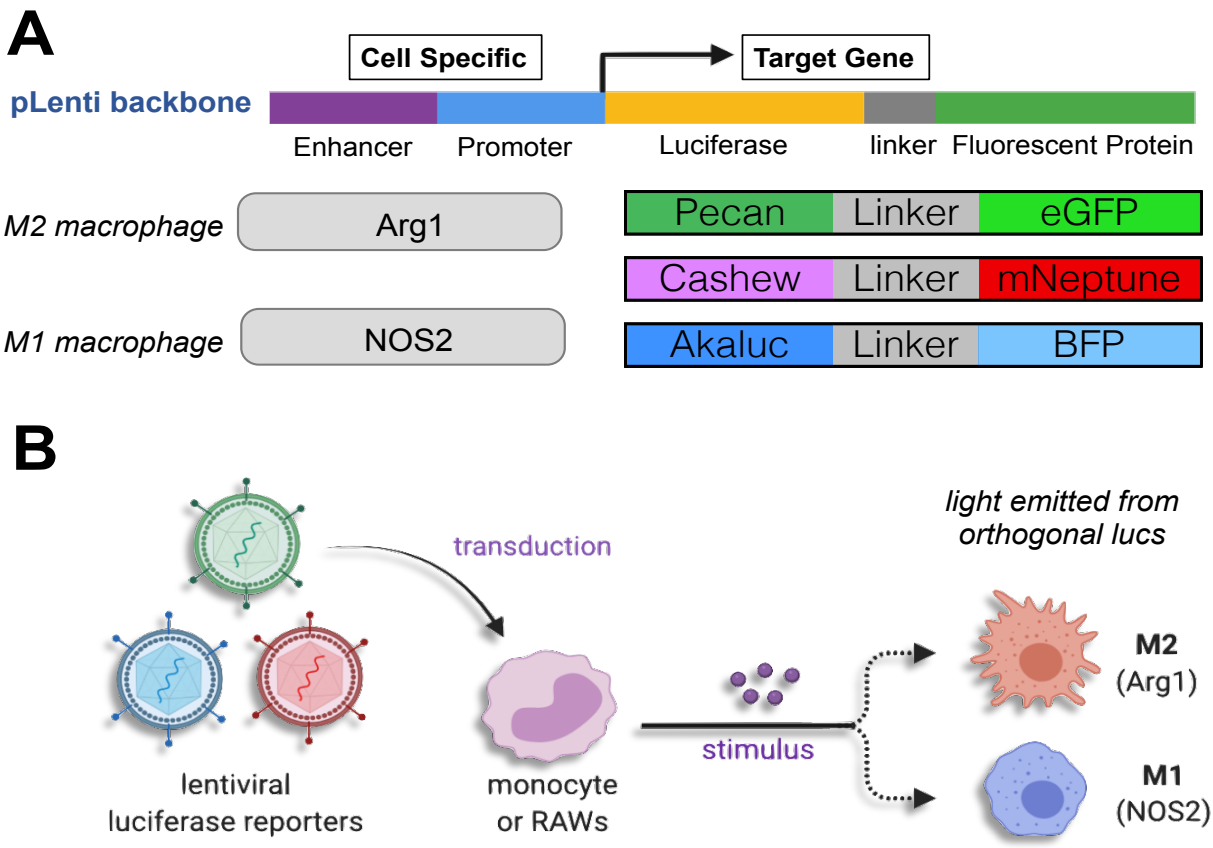


**Figure 5.11. Schematic of macrophage specific promoter design**. (A) Arg1 (M2 macrophage) and NOS2 (M1 macrophage) promoters drive the expression of orthogonal luciferases linked with spectrally distinct fluorescent proteins. (B) Graphic diagram of lentiviral transfection of reporters to transduce monocytes or RAW cell lines. M2 and M1 macrophages are derived after stimulation. Cell specific promoters are identified through the light emitted from orthogonal luciferases.

luciferases fused to different fluorescent proteins that allow easy FACS and ex vivo analysis (Figure 5.11A). The initial design of the gene reporters using the CRISPR backbone did not produce enough light emission from preliminary *in vitro* experiments. To improve photon output, I changed the design of the reporters to be inserted instead via lenti-viral transduction. The designs of these plasmids have been completed. Translating these constructs to test *in vitro* and continue *in vivo* is an ongoing area of research in the lab (Figure 5.11B). We plan to continue my work

studying immune cell function in metastatic disease by imaging changes in cellular phenotypes using bone marrow derived macrophages.

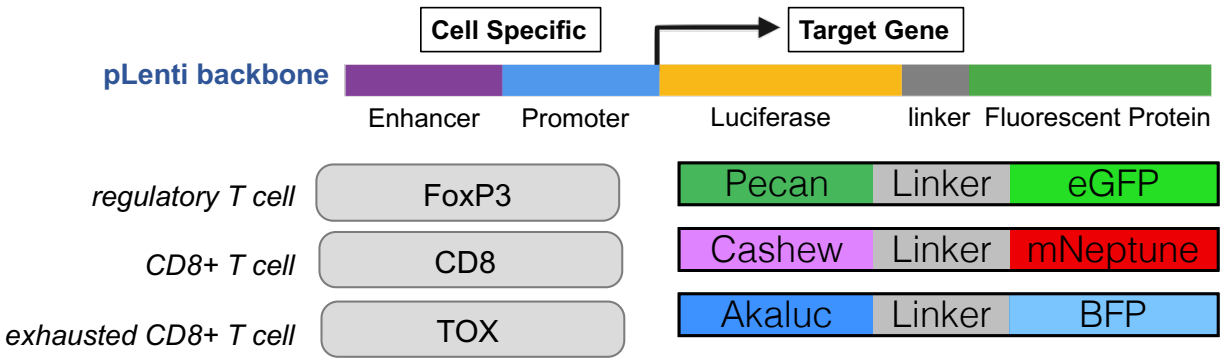


**Figure 5.12. Schematic of T cell specific promoter driven reporters**. FoxP3 (regulatory T cells), CD8 (CD8+ T cells), TOX (exhausted CD8+ T cell) promoter and enhancers were designed to drive the expression of different orthogonal luciferases fused to fluorescent proteins to monitor cell populations.

I aimed to expand on the ability to monitor immune cells by creating novel promoter driven plasmids that monitor T cell function. As stated in Chapter 1, T cells are another area of interest in cancer biology [17]. Creating luciferase driven gene reporters to investigate CD8+ T cell and regulatory T cell (Tregs) function is an ongoing area of research that will continue on with the work I laid out in my thesis. To this end I have created reporters with sequences from T cell populations (Figure 5.12). These constructs are exciting because they are first ever created to monitor T cell exhaustion and Tregs. Using bone marrow flushing I extracted the gDNA to use as a template for these plasmids. These constructs are the first ever to be made specifically with FVB mouse background codon optimized DNA. Currently, most commercially available plasmids are limited to compatibility with DNA from Black 6 mice. Black 6 mice are not ideal for imaging. Instead, white mice such as FVB backgrounds have been traditionally used for imaging purposes and serve as the background for the MMTV-PyMT mouse model (outlined in Chapters 2, 3, and

4). By creating the initial sequences from the other immune cell reporters using FVB mouse-derived DNA, this allows for easy translation of the reporters into the MMTV-PyMT mouse models and existing multicomponent imaging cell lines that were described in Chapter 4. The combined MMTV-PyMT luciferase-expressing cancer cell line and immune reporter animal models could be achieved by using monocyte transfer or bone marrow transplantations. The creation of a transgenic MMTV-PyMT metastatic model (outlined in Chapters 3-4) in combination with the split luciferase reporters and the orthogonal immune driven reporters outlined here could be a powerful tool to monitor tumor-associated immune cell changes during cancer progression (Figure 5.13). Such a transgenic mouse would significantly impact on our understanding of disease progression by providing a comprehensive platform to track changes in immune cell behavior in the tumor microenvironment and in different metastatic tissue-environments (Figure 5.13).

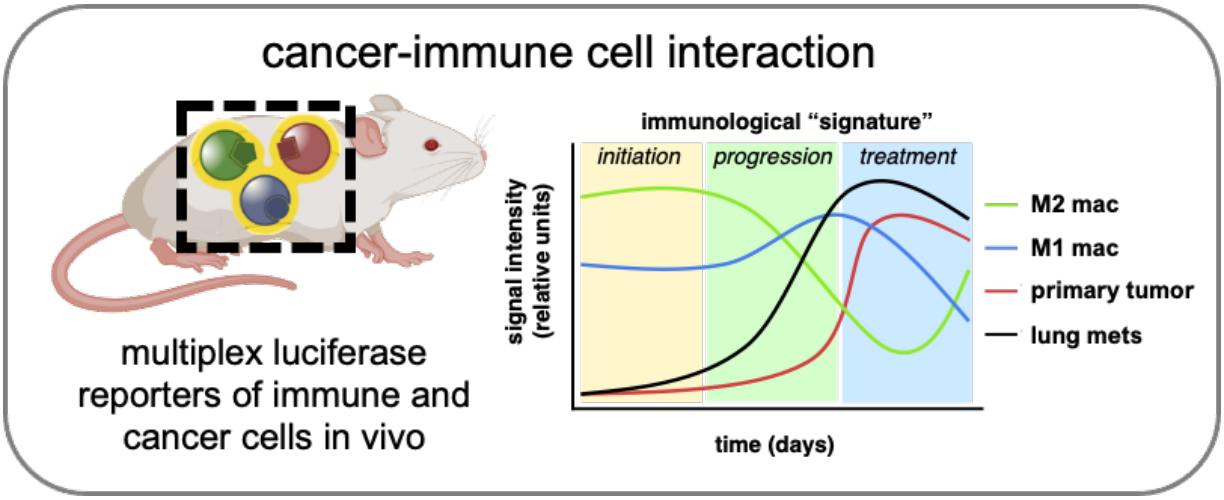


**Figure 5.13: Overview of future possible transgenic mouse to monitor immune cells and cancer progression**

## V. Conclusions

In this chapter I presented my efforts developing new imaging probes to track cell-to-cell contact and cell-specific gene reporters. Cell contacts can be visualized using split luciferase reporters that recombine in close proximity to emit light. Split luciferase fragments were tethered to the cell surface or with coil-coils to facilitate recombination. The split luciferase reporters were able to read out on cellular interaction *in vitro*. However, additional work is needed to improve signal output for *in vivo* use. I also showed my preliminary work creating cell specific promoter driven luciferase gene reporters. These reporters could allow monitoring of immune cell changes during disease progression. This is an ongoing area of research in the laboratory. The combination of these probes could be used to create a transgenic cancer mouse model that leverages our multicomponent bioluminescent imaging platform.

## MATERIALS AND METHODS

### Reagents

All reagents purchased from commercial supplies were of analytical grade and used without further purification. 4'-BrLuc was prepared and used as described previously described in Ionkina, *et al*. *Rapid Multicomponent Bioluminescence Imaging via Substrate Unmixing.* ACS Chem Biol, 2021. **16**(4): p. 682-690. Pecan-eGFP-AAVS1, Cashew-mNeptune-AAVS1, and Anatres-AAVS1, and Akaluc-TagBFP-AAVS1 were used as described in Ionkina, *et al*. *Rapid Multicomponent Bioluminescence Imaging via Substrate Unmixing.* ACS Chem Biol, 2021. **16**(4): p. 682-690.

### General bioluminescence imaging

All assays were performed in black 96-well plates (Grenier Bio One). Plates containing luminescent reagents were imaged in a light-proof chamber with an IVIS Lumina (Xenogen) CCD camera chilled to –90 °C. The stage was kept at 37 °C during the imaging session, and the camera was controlled using Living Image software. For all assays, exposure times were set to 1–180 s, and data binning levels were set to medium. Total flux values for regions of interest were analyzed using Living Image software. Integrated pixel values were analyzed using ImageJ (Installed under the FIJI package, NIH). The data were analyzed using GraphPad Prism (version 9.0 for Macintosh, GraphPad Software).

### Cloning

Plasmids were cloning and GIBSON assembly were previously noted in Ionkina, *et al*. *Rapid Multicomponent Bioluminescence Imaging via Substrate Unmixing.* ACS Chem Biol, 2021. **16**(4):

p. 682-690. Immune cell specific enhancer/promoter plasmids were purchased from Addgene. Primers design specifications were created using Addgene sequencies.

mouse ARG1 promoter/enhancer (addgene plasmid 34571)

human IL2 Promoter (addgene plasmid 10959)

pGL2-NOS2Promoter-Luciferase (addgene plasmid #19296)

pGL3-E-cadherin promoter (addgene plasmid #61798)

**Mammalian cell culture**

HEK-293T and DB7 cells stably expressing Pecan-eGFP, Cashew-mNeptune, or Akaluc-BFP were derived via CRISPR-mediated gene insertion as previously described [7]. Split Antares constructs were designed using previous described cloning methods [4]. Split Antares reporters were transfected into cells via CRISPR-mediate gene insertion. Conflocal in images were obtained using previously detailed methods [4] and (Ionkina et al., (2017) *Frontiers in Oncology*. 7,94).

Transfected cells were then treated with puromycin (2 μg/mL) and FACS sorted at the Institute for Immunology Flow Cytometry Core (UCI). All cells were cultured in DMEM (Corning) supplemented with 10% (v/v) fetal bovine serum (FBS, Life Technologies), penicillin (100 U/mL), and streptomycin (100 μg/mL). Stably expressing luciferase cells were maintained under puromycin selection (2 μg/mL) to ensure gene incorporation was preserved. Cells were maintained in a 5% $CO_2$ water-saturated incubator at 37 °C. Cells were serially passaged using trypsin-EDTA (0.25% in HBSS, Gibco).

**Substrate unmixing analysis**

Substrate unmixing experiments were designed as previously described in Ionkina, *et al*. *Rapid Multicomponent Bioluminescence Imaging via Substrate Unmixing.* ACS Chem Biol, 2021. **16**(4): p. 682-690. Substrate unmixing was conducted with MATLAB R2020a (See Supplementary Discussion). Luminescence images containing the raw CCD counts (as TIFF files) were loaded into MATLAB. Images were subjected to a 2-pixel median filter (using the *medfilt2* function with a 5x5 neighborhood around the corresponding pixel). Next, the signal at each pixel was normalized to lie between 0 and 65536 (the maximum value that can be stored in a 16-bit image). As a result, the brightest pixel in each image had a value of 65536, and the dimmest had a value of 0. Regions of interest (ROIs) were generated by identifying the image coordinate of the reference well and input dimensions. Once assigned, the MATLAB algorithm was run to perform the unmixing. After unmixing, text images were imported into ImageJ (installed under the FIJI package). Integrated pixel values for regions of interest were analyzed using the "Measure" tool. Pseudo colors were assigned with the "Merge Channels" tool.

**Mammalian cell analysis of luciferase mutants**

DB7 or HEK-293T cells stably expressing luciferases were added to black 96-well plates. Cells were treated with a luciferin solution (10 µL, of 100 µM or 250 uM for D-luc, 4′-BrLuc, furimazine and AkaLumine). Plates were imaged and analyzed as described above.

## References


1. Seyfried, T.N. and L.C. Huysentruyt, *On the origin of cancer metastasis.* Crit Rev Oncog, 2013. **18**(1-2): p. 43-73.
2. Deshmukh, S.K., et al., *Emerging evidence for the role of differential tumor microenvironment in breast cancer racial disparity: a closer look at the surroundings.* Carcinogenesis, 2017. **38**(8): p. 757-765.
3. Hallou, A., J. Jennings, and A.J. Kabla, *Tumour heterogeneity promotes collective invasion and cancer metastatic dissemination.* R Soc Open Sci, 2017. **4**(8): p. 161007.
4. Jones, K.A., et al., *Visualizing cell proximity with genetically encoded bioluminescent reporters.* ACS Chem Biol, 2015. **10**(4): p. 933-8.
5. Chu, J., et al., *A bright cyan-excitable orange fluorescent protein facilitates dual-emission microscopy and enhances bioluminescence imaging in vivo.* Nat Biotechnol, 2016. **34**(7): p. 760-7.
6. Iwano, S., et al., *Single-cell bioluminescence imaging of deep tissue in freely moving animals.* Science, 2018. **359**(6378): p. 935-939.
7. Lake, M.C. and E.O. Aboagye, *Luciferase fragment complementation imaging in preclinical cancer studies.* Oncoscience, 2014. **1**(5): p. 310-25.
8. Paulmurugan, R. and S.S. Gambhir, *Firefly luciferase enzyme fragment complementation for imaging in cells and living animals.* Anal Chem, 2005. **77**(5): p. 1295-302.
9. Chen, H., et al., *Firefly luciferase complementation imaging assay for protein-protein interactions in plants.* Plant Physiol, 2008. **146**(2): p. 368-76.
10. McCann, T.E., et al., *The use of fluorescent proteins for developing cancer-specific target imaging probes.* Methods in molecular biology (Clifton, N.J.), 2012. **872**: p. 191-204.
11. Lin, W., et al., *Regulatory T cell development in the absence of functional Foxp3.* Nat Immunol, 2007. **8**(4): p. 359-68.
12. Song, W., et al., *Translational Significance for Tumor Metastasis of Tumor-Associated Macrophages and Epithelial-Mesenchymal Transition.* Front Immunol, 2017. **8**: p. 1106.
13. Ma, J., et al., *Characterization of mammary cancer stem cells in the MMTV-PyMT mouse model.* Tumour Biol, 2012. **33**(6): p. 1983-96.
14. Cendrowicz, E., et al., *The Role of Macrophages in Cancer Development and Therapy.* Cancers, 2021. **13**(8): p. 1946.
15. Duan, Z. and Y. Luo, *Targeting macrophages in cancer immunotherapy.* Signal Transduction and Targeted Therapy, 2021. **6**(1): p. 127.
16. Pathria, P., T.L. Louis, and J.A. Varner, *Targeting Tumor-Associated Macrophages in Cancer.* Trends in Immunology, 2019. **40**(4): p. 310-327.
17. Waldman, A.D., J.M. Fritz, and M.J. Lenardo, *A guide to cancer immunotherapy: from T cell basic science to clinical practice.* Nat Rev Immunol, 2020. **20**(11): p. 651-668.

# CHAPTER 6: Conclusions and Future Directions

Cancer metastases are driven by complex interactions among tumor cells, immune cells, and other cell types [1]. Cancer models and novel imaging methods to noninvasively visualize specific cell types and cellular interactions could improve our ability to monitor disease progression *in vivo* [2]. In this dissertation, I have developed imaging probes and cancer models to enable a more detailed examination of tumor heterogeneity and metastatic disease.

In Chapter 2, I generated new methods to enable multicomponent bioluminescence imaging *in vivo*. This platform is built on engineered luciferases that use modified small molecule luciferins. Using CRISPR-Cas9 technology, I created an array of cancer cell lines expressing the requisite luciferases. The gene editing platform has since been used to create multiple engineered cell lines for multiplexed bioluminescence imaging. I used the cell lines to demonstrate rapid multicomponent bioluminescent imaging via sequential substrate administration *in vitro* and *in vivo*. By administering luciferins from dimmest to brightest, we were able to "layer on" the signal in rapid succession. The compiled snapshots were unmixed and false colored to produce a composite image. Using this method, we were able to achieve dual component bioluminescence imaging with a total acquisition time of 30 min, compared to the >24 h needed with traditional imaging protocols.

Building on the initial results, I aimed expand the number of probes that could be imaged. I specifically focused on developing a platform for unmixing three discrete luciferase-luciferin pairs. I achieved rapid triple-component *in vitro* imaging using both Antares (with its luciferin, furimazine) [3] and Akaluc (with its luciferin, AkaLumine) [4]. This is an ongoing area of research

in our lab, with initial screenings performed in bacterial lysate and mammalian cells. Although results from these *in vitro* experiments were promising, the initial conditions did not translate *in vivo*. I believe that the success of expanding our imaging platform in future animal experiments will require careful planning of additional parameters. These parameters include using more red-shifted tissue penetrant probes that are enhanced for brightness. In addition, future experiments should address luciferin solubility issues to ensure compatibility with the biological conditions required for animal imaging. Substrate bioavailability and stability *in vivo* should also be considered.

Throughout the multi-component imaging experiments, there were problems with the stability of the firefly luciferase mutants, Pecan and Cashew, in various cell lines. Mutations arose that resulted in diminished substrate preference among the luciferases, complicating the imaging readout. Functional luciferase-expressing cells were re-derived using CRISPR plasmids, and sequencing of gDNA from the cells confirmed proper transgene integration. However, sporadic mutations continued to arise. For future studies, it will be critical to monitor the stability of these luciferases closely by frequently sequencing the cell lines and the plasmid stocks. More assistance on such labor intensive *in vivo* experiments would help catch these issues.

In Chapter 3, I established an array of organ-derived metastatic cell lines harvested from the MMTV-PyMT mouse model. Transcriptomic analyses were performed to examine the effects of cell heterogeneity on metastases and organ tropism. Overall, the suite of clonal isolates provided a detailed depiction of cancer progression. Although we performed a detailed analysis of the metastatic cells lines, more could be learned with single cell sequencing. Such analyses were performed on three cell populations (lung low, lung high, and lymph node high). It would be worthwhile to expand these analyses to include MFP and lymph node low cell populations. By

obtaining single cell data from these additional cell lines, I believe we could expand the pseudo-time projections (Figure 3.13) to learn more about intratumoral heterogeneity and metastatic progression (Figure 3.14 - 3.15). The additional data recovered from these cell lines could further provide more information on cancer stem cell populations (Figure 3.16). Monitoring cancer stem cell populations throughout metastatic progression would be especially interesting and could bolster the stem cell hypothesis outlined in Chapter 3.

The sequencing analyses featured in Chapter 3 provide a treasure chest of gene expression data that could be used for further experiments. As I have outlined in Chapter 5, the expression data can guide the creation of metastases-specific gene reporters, that in combination with our multicomponent imaging platform, would allow the monitoring of disease progression with unrivaled specificity. These reporter constructs would be primed to use in conjunction with the MTMV-PyMT mouse model for *in vivo* imaging. Such reporters would also be useful for creating a transgenic mouse model to enable a comprehensive investigation of cancer disease progression *in vivo.*

In Chapter 4, I used the suite of MMTV-PyMT metastatic cancer cells in conjunction with the imaging probes in Chapter 2 to create a multicomponent bioluminescence imaging platform to study cancer progression. I validated that these cells are capable of recapitulating disease progression in mice and are a valuable resource for future imaging studies. My work also was essential for creating an improved algorithm to unmix multi-component bioluminescent signals using substrate preference. With the cell lines, I aimed to expand the imaging platform for triple component imaging. First, I used Antares along with its substrate, furimazine [3]. Antares is a BRET construct with two fluorescent proteins flanking NanoLuc. The MMTV-PyMT cells expressing Antares exhibited a difference in growth rate compared to other firefly derived

luciferase-expressing cells. This decrease in cellular growth of Antares-expressing cells became exponentially evident when the cells were injected *in vivo*. Tumors expressing Antares formed far slower and grew much smaller in volume compared to the other mutant luciferase-expressing tumors. In addition, the corresponding luciferin substrate, furimazine, had bioavailability issues and auto-oxidation problems when translated *in vivo* [11, 12].

A key experiment that I performed (Figure 4.6) was to benchmark the bioluminescent signals and cellular ratios injected *in vivo* to the fluorescence readout for tumor containing Antares-CyOFP, Pecan-eGFP, and Cashew-mNeptune cells. Tumors were extracted and processed into single cell suspensions for flow cytometry analysis. I processed the 18 tumors by hand. Future renditions of this experiment would benefit from using a homogenizer machine to retain more of the tumor cells and limit the loss of the sample during manual processing. In addition, I performed the entire experiment and processing of the 18 tumors alone. Therefore future benchmark studies would benefit from additional help and assistance. As mentioned before, tumor volumes and cell growth differed between the type of luciferase expressed in the cell lines (Table 4.1). This variability made processing the small tumors very difficult. Therefore, it would be prudent to allow the tumors to grow to larger volumes before starting the experiments. This would improve the processing of the tumors by providing more material to use for the downstream ex vivo analyses. Nevertheless, the benchmarking results and flow cytometry analysis performed in Chapter 4 showed promise. This experiment should be repeated in the future and should take into account the amendments that I have outlined.

From the results in this dissertation, it is clear that Antares is not ideal for multiplexing with our engineered insect luciferases in mice. I identified Akaluc and its substrate, AkaLumine, as a better alternative for orthogonal imaging *in vivo* [4]. Conditions to best dissolve and administer

AkaLumine for multi-component experiments were identified (Figure 4.18), in line with previous observations [4].

As we expand the scope of *in vivo* multicomponent experiments, I anticipate that luciferase-expressing cells will be imaged at different depths and through tissues. The current experiments were all performed at relatively shallow depths with subcutaneous injections. If future experiments aim to monitor cellular division or migration throughout the course of cancer progression, tissue thickness will become an added variable. In these instances, light emitted from luciferase-expressing cells could be lost due to the lack of bioluminescent light penetration. This could make it difficult to correlate the light emission back to the number of cells injected, as obscured light emission would result in artificially low counts for light-emitting cells. Further experiments could address this issue by using additional red-shifted probes that enable more photons to escape thick tissues. Newer and more sensitive imaging equipment could also improve detection thresholds. Some imaging machines allow animals to be visualized from all angles versus just the one-dimensional acquisition of our current set up. Finally, improvements to our unmixing algorithm that take into account cellular growth kinetics and light scattering properties of different tissues (such as fats, blood, bone, etc.) would be beneficial. An ideal unmixing algorithm could predict these variables and normalize the final image.

In Chapter 5, I discussed my preliminary efforts creating imaging probes to visualize cell contact and track changes in cell profiles. Although split luciferase reporters were able to monitor cell contact *in vitro* (Figure 5.1 -5.4) [5], I do not believe the current reporters are suited for *in vivo* use owing to low photon outputs. Future work in this area should aim to improve photon production from the recombined splits (Figure 5.7) [6]. I also outlined the ideas and methods to generate cell specific gene reporters capable of monitoring changes in cell profiles (Figure 5.11

and 5.12). These probes could be very useful in the context of monitoring immune cell changes, such as macrophage polarization and T cell states [7, 8]. I believe that the biggest obstacle for future implementation of this technology will be selecting the correct promoters that will provide specific signal [9, 10]. Achieving bright signal to distinguish between different cellular states will also be a challenge. Increasing the copy number of the gene reporter along with using bright-red-shifted luciferases could help address this problem.

## References


1. Seyfried, T.N. and L.C. Huysentruyt, *On the origin of cancer metastasis.* Crit Rev Oncog, 2013. **18**(1-2): p. 43-73.
2. Deshmukh, S.K., et al., *Emerging evidence for the role of differential tumor microenvironment in breast cancer racial disparity: a closer look at the surroundings.* Carcinogenesis, 2017. **38**(8): p. 757-765.
3. Chu, J., et al., *A bright cyan-excitable orange fluorescent protein facilitates dual-emission microscopy and enhances bioluminescence imaging in vivo.* Nat Biotechnol, 2016. **34**(7): p. 760-7.
4. Iwano, S., et al., *Single-cell bioluminescence imaging of deep tissue in freely moving animals.* Science, 2018. **359**(6378): p. 935-939.
5. Jones, K.A., et al., *Visualizing cell proximity with genetically encoded bioluminescent reporters.* ACS Chem Biol, 2015. **10**(4): p. 933-8.
6. Paulmurugan, R. and S.S. Gambhir, *Firefly luciferase enzyme fragment complementation for imaging in cells and living animals.* Anal Chem, 2005. **77**(5): p. 1295-302.
7. Lin, W., et al., *Regulatory T cell development in the absence of functional Foxp3.* Nat Immunol, 2007. **8**(4): p. 359-68.
8. Song, W., et al., *Translational Significance for Tumor Metastasis of Tumor-Associated Macrophages and Epithelial-Mesenchymal Transition.* Front Immunol, 2017. **8**: p. 1106.
9. Pathria, P., T.L. Louis, and J.A. Varner, *Targeting Tumor-Associated Macrophages in Cancer.* Trends in Immunology, 2019. **40**(4): p. 310-327.
10. Waldman, A.D., J.M. Fritz, and M.J. Lenardo, *A guide to cancer immunotherapy: from T cell basic science to clinical practice.* Nat Rev Immunol, 2020. **20**(11): p. 651-668.
11. Hall, M.P., et al., *Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate.* ACS Chem Biol, 2012. **7**(11): p. 1848-57.
12. Gaspar, N., et al., *Evaluation of NanoLuc substrates for bioluminescence imaging of transferred cells in mice.* J Photochem Photobiol B, 2021. **216**: p. 112128.