Dissecting cellular heterogeneity in breast cancer metastatic progression

by

Constanza Martínez Ramírez

Pathology Department
McGill University, Montréal

August 2021

A thesis submitted to McGill University in partial fulfillment
of the requirements for the degree of Doctor of Philosophy

Abstract

Breast cancer is a heterogeneous disease with variable clinical presentation, histological features, and response to therapy. Despite improvements in therapy and survival in breast cancer patients, triple-negative breast cancer (TNBC) subtype is the exception. TNBC is a heterogenous disease, with no targeted therapy and patients are still treated with standard of care chemotherapy. While some patients present favorable treatment responses, others develop metastatic disease within the first three years after their initial diagnosis. While these differences have been partially explained by inter- and intra-tumoral heterogeneity at the genomic and transcriptomic level, the role of spatial and temporal heterogeneity upon disease progression, represented by the tumor architecture is still poorly understood.

Here, we use patient derived xenografts (PDX) to study temporal heterogeneity upon disease progression. We develop models of residual disease, recurrence, treatment resistance and metastatic disease in a stepwise manner subjected to single-cell RNA sequencing (scRNA-seq) to identify biomarkers and interrogate spatial heterogeneity in situ. We combine single-cell RNA sequencing (scRNA-seq), multiplex immunofluorescence (MIF) and digital spatial profiling transcriptomics to interrogate matched PDX models of primary and metastatic TNBC samples to later define transcriptional states in the context of intact tissues. Here we identify cell populations that are present across independent TNBC tumors that are predominantly determined by their spatial localization within the tumor. Both primary and metastatic tumors contain transcriptionally distinct cycling populations that show a gradual transition towards hypoxic populations, largely defined by their distance to necrotic zones. In contrast, a

second class of cell populations, show no preference for spatial localization and display a low degree of network connectivity, interspersed among other cell states within the tumor tissue. These cells are exposed to distinct environmental queues in each zone and tend to co-express a range of transcriptional programs, supporting a high degree of plasticity. In functional studies, both hypoxic and cell populations displaying plasticity show higher tumor initiating potential in vivo and are enhanced in tumor metastases.

Taken together, we present a spatial transcriptomic approach coupled with the discovery of distinct TNBC zones. This methodology allows us to a) study the spatial distribution of distinct cell populations identified by single cell gene-expression in histological samples, b) identify distinct cell populations in defined neighborhoods within defined tumor locations and c) identify cells within specific biological zones that may share states of plasticity and can be found among other TNBC with distinct clinical and specific activation status. The integration of spatial information and single-cell transcriptomics is a powerful tool to biologically interrogate intra-tumor heterogeneity with a high level of integration among tissue sections that can be clinically relevant among disease progression.

Résumé

Le cancer du sein est une maladie hétérogène dont la présentation clinique, les caractéristiques histologiques et la réponse au traitement sont variables. Malgré les améliorations apportées au traitement et à la survie des patientes atteintes d'un cancer du sein, le sous-type de cancer du sein triple négatif (CSTN) fait figure d'exception. Le CSTN est une maladie hétérogène, sans thérapie ciblée et les patientes sont toujours traitées par chimiothérapie standard. Alors que certaines patientes présentent des réponses favorables au traitement, d'autres développent une maladie métastatique dans les trois premières années suivant leur diagnostic initial. Si ces différences ont été partiellement expliquées par l'hétérogénéité inter- et intra-tumorale au niveau génomique et transcriptomique, le rôle de l'hétérogénéité spatiale et temporelle sur la progression de la maladie, représentée par l'architecture tumorale, est encore mal compris.

Grace à des xénogreffes dérivées de tumeurs de patients (PDX), nous avons développé des modèles de maladie résiduelle, de récidive, de résistance au traitement et de maladie métastatique pour étudier l'hétérogénéité spatiale et temporelle lors de la progression de la maladie. En combinant les technologies de séquençage de l'ARN de cellules cancéreuses individuelles, de marquage multiple par immunofluorescence et de profilage spatial de cibles d'ARN, nous avons identifié différentes sous-populations de cellules cancéreuses en fonction de leur localisation intra-tumorale distincte au sein de tumeurs primaires et de métastases. Par exemple, une sous-population de cellules hypoxiques montrant une transition progressive vers un état prolifératif est largement définie par sa distance relative aux zones nécrotiques. Une deuxième classe de population cellulaire, en revanche, ne présente aucune localisation particulière mais

plutôt une distribution homogène et intriquée dans les différentes zones du tissu tumoral. Ces dernières expriment un large éventail de profils transcriptomiques suggérant une importante plasticité phénotypique influencée par le micro-environnement tumoral des différentes zones. Des études fonctionnelles in vivo de ces sous-populations de cellules hypoxiques et plastiques ont démontré une capacité de tumorigenèse et de progression métastatique accrue.

Notre approche innovante de profilage spatial de cibles d'ARN couplé à la caractérisation de zones distinctes au sein de CSTN par marquage d'immunofluorescence nous a permis de a) caractériser la distribution spatiale de sous-populations identifiées par le séquençage de cellules cancéreuses individuelles, b) identifier des sous-populations ayant une localisation intra-tumorale spécifique au sein tu tissu and c) démontrer l'existence d'une plasticité phénotypique entre différentes sous-populations localisées dans la même zone tissulaire, caractéristique d'ailleurs observée dans plusieurs cas de CSTN. L'intégration du séquençage de cellules individuelles avec le profilage spatial représente un outil majeur pour comprendre l'hétérogénéité spatiale et temporelle intra-tumorale des CSTN à partir de sections histologiques de tumeur, dont l'évaluation permet de mieux envisager la progression de la maladie.

Acknowledgements

The doctoral training experience has been one that I never imagined I would have. Inspired to gain additional knowledge about research to complement my clinical training I embarked on one of the most rewarding journeys of my life. The professional and personal growth I developed throughout the past years, and the influential, inspiring, and passionate people that I met in it, certainly changed my vision of the world.

My first big thank you is for my supervisor Dr. Morag Park, a passionate scientist and mentor, who has always been very supportive of my career goals, challenging me beyond my limits and knowledge. She gave me the tools and independence to be able to integrate basic science with clinical research. Ultimately, today I feel more prepared to embark myself as a clinician-scientist in the field of cancer research

I would also like to thank the members of my Research Advisory Committee, Dr. Peter Siegel, and Dr. Carlos Telleria for the scientific feedback and thoughtful ideas to improve my research. Thanks to collaborators, Dr. Claudia Kleinman, and Dr. Santiago Constantino, for all the dedicated work and for always being willing to answer my questions.

Thanks to all the Park lab –Paula Coehlo for your brilliant scientific input and lovely personality, Vanessa Sung for being the best desk work neighbor. Tunde Golenar for your great morning talks. Bruce Huang for your amazing tips about science and the business market. Gabrielle Brewer for your sincere comments and critical thinking, Colin Ratcliffe for always asking, what is the question? Marina for your daily smile and hard work. Paul Savage for your initial training and good advice. Thanks to postdocs and research associates; Hellen Kuasne and Anne-Marie for being amazing role models at work and

outside work. Thanks to Anie Monast for your animal work and cooking advice. Veena for your input. Virginie Pilon thanks for being my friend throughout this journey and many others. Thanks, to Dongmei Zuo for your listening ear, support, dedication, and all the technical work. Goodman Cancer Research members and friends – Special thanks to Giulio Aceto and Brian Hsu for bringing me so many good moments and happiness while enjoying doing research.

This work was supported by the FRQS Doctoral training award scholarship. Thanks to all the patients that participate in donating tissue, you are the scientific engine behind every discovery.

Many thanks to my family, Maxi, Patricio, Patin, Felipe, Mauricio, and Joana, who always have supported me in every decision and encouraged me to go for more. Ingrid, thanks for the endless support and for believing so strongly in me.

Undoubtedly, one of the take-home messages from my graduate studies is the discovery of the joy that I find in connecting with people, collaborating, and engaging as a team, to move science forwards and ultimately help patients.

Contribution to original knowledge

Single-cell technologies are now widely performed on different cancer types and within a single tumor on different cell types. This has helped to get a better understanding of intra-tumor heterogeneity and study different biological processes, such as tumor initiation, resistance to treatment and metastasis. Importantly, spatial transcriptomics using markers to identify populations or to differentiate different cell types has been addressed by different methods and by different groups using various techniques. However, systematic, and unbiased integration of single-cell RNA sequencing (scRNA-seq), spatial transcriptomics, digital spatial profiling, and functional studies of different cancer populations is still not done in a standardized manner.

To overcome the main limitation of scRNA-seq, which loses the spatial information of cells, we coupled this technology with multiplex immunofluorescence and digital spatial profiling. First, two triple-negative breast cancer (TNBC) patient-derived xenografts (PDXs) tumor replicates were bioinformatically joined to obtain robust scRNA-seq data clusters and biomarkers aimed to identify specific cell populations. Later, the breast primary tumor, the lung metastasis and matched PDX from the same patient, were stained with specific markers/antibodies that aimed to identify each bioinformatic cluster. To do that, a computational system was developed to integrate the ratio of expression of the antibodies in each cell in the tumor. This allowed us to determine in-situ, to which scRNA-seq cluster each cell corresponded to.

The mapping of scRNA-seq populations identified two tumor zones. A hypoxic tumor zone, comprised by cells enriched with hypoxic gene signatures and properties, that tend to be found near necrotic regions in the tumor. On the other hand, a cycling

tumor zone, where cells with proliferative genes were found further away from hypoxic regions. Other scRNA-seq populations were zone-less and were found to be scattered in the tissue.

Moreover, in absence of patient's PT scRNA-seq data, digital spatial profiling served as validation method for the identification of scRNA-seq zonal populations. We proved that the spatial heterogeneity between the PT and matched PDXs is preserved and that scRNA-seq populations express transcriptional program depending on their geographical location in the tumor. Lastly, we characterized in time and upon disease progression, the primary tumor transcriptomic cell populations by establishing a spontaneous lung metastasis (SLM) PDX model that mimicked the 6-month patient's disease progression latency. The spatial mapping of cells in the lung environment in the PDX and the patient's lung metastasis biopsy, revealed an enrichment of one of the zone-less populations identified in the PT. This population displayed phenotypic plasticity, adopting markers of different tumor zones, and a cholesterol homeostasis transcriptional program. Using a multi-pronged approach, integrating scRNA-seq data, multiplex immunofluorescence, digital spatial profiling and functional studies, we identified the most aggressive populations promoting tumor initiation and metastatic disease. Therefore, hypoxic and cholesterol homeostasis transcriptomic populations were targeted in vivo, displaying a reduction in tumor growth and lung metastatic burden.

Altogether, these data interrogating phenotypic plasticity in situ throughout disease progression, allows for the integration of the biological properties of distinct transcriptomic populations, their spatial localization, and their selection from the primary to the metastatic disease.

Preface

This thesis is written in the traditional format, and it contains five chapters as follows:

Chapter 1: Literature review.

Chapter 2. Results.

Chapter 3: General discussion.

Chapter 4: Experimental procedures.

Chapter 5: Bibliography.

Publications arising from this thesis work

Chapters 1 and 2 contain material presented in the following articles:

1. **Constanza Martinez**, Xiang Cheng, Morag Park. "Clinical and research opportunities for triple-negative breast cancer residual disease biomarkers". *Manuscript in preparation.*

2. **Constanza Martinez**, Yang Yang, Anne-Marie Fortier, Hellen Kuasne, Samantha Worme, Evelyn Zavacky, Dongmei Zuo, Santiago Costantino, Paul Savage, Nicholas Anthony, Ioannis Ragoussis, Yasser Riazalhosseini, Morag Park and Claudia Kleinman. "Functional cartography of single-cell RNA sequencing populations in matched primary and metastatic breast cancer". *Manuscript in preparation.*

Additional publications from this work

1. Kuzmin, Elena, Jean Monlong, **Constanza Martinez**, Hellen Kuasne, Claudia L. Kleinman, Jiannis Ragoussis, Guillaume Bourque, and Morag Park. 2021. "Inferring Copy Number from Triple-Negative Breast Cancer Patient Derived Xenograft ScRNAseq Data Using ScCNA." *Methods in Molecular Biology (Clifton, N.J.)* 2381:285–303.

2. Totten, Stephanie P., Young Kyuen Im, Eduardo Cepeda Cañedo, Ouafa Najyb, Alice Nguyen, Steven Hébert, Ryuhjin Ahn, Kyle Lewis, Benjamin Lebeau, Rachel La Selva, Valérie Sabourin, **Constanza Martínez**, Paul Savage, Hellen Kuasne, Daina Avizonis, Nancy Santos Martínez, Catherine Chabot, Adriana Aguilar-Mahecha, Marie-Line

Goulet, Matthew Dankner, Michael Witcher, Kevin Petrecca, Mark Basik, Michael Pollak, Ivan Topisirovic, Rongtuan Lin, Peter M. Siegel, Claudia L. Kleinman, Morag Park, Julie St-Pierre, and Josie Ursini-Siegel. 2021. "STAT1 Potentiates Oxidative Stress Revealing a Targetable Vulnerability That Increases Phenformin Efficacy in Breast Cancer." *Nature Communications*. 2021 Jun;12(1):3299.

3. Savage, Paul, Alain Pacis, Hellen Kuasne, Leah Liu, Daniel Lai, Adrian Wan, Matthew Dankner, **Constanza Martinez**, Valentina Muñoz-Ramos, Virginie Pilon, Anie Monast, Hong Zhao, Margarita Souleimanova, Matthew G. Annis, Adriana Aguilar-Mahecha, Josiane Lafleur, Nicholas R. Bertos, Jamil Asselah, Nathaniel Bouganim, Kevin Petrecca, Peter M. Siegel, Atilla Omeroglu, Sohrab P. Shah, Samuel Aparicio, Mark Basik, Sarkis Meterissian, and Morag Park. 2020. "Chemogenomic Profiling of Breast Cancer Patient-Derived Xenografts Reveals Targetable Vulnerabilities for Difficult-to-Treat Tumors." *Communications Biology* 3(1):310.

4. Geneviève Deblois, Seyed Ali Madani Tonekaboni, Giacomo Grillo, **Constanza Martinez**, Yunchi Ingrid Kao, Felicia Tai, Ilias Ettayebi, Anne-Marie Fortier, Paul Savage, Alexandra N. Fedor, Xiaojing Liu, Paul Guilhamon, Evelyne Lima-Fernandes, Alex Murison, Hellen Kuasne, Wail Ba-alawi, David W. Cescon, Cheryl H. Arrowsmith, Daniel D. De Carvalho, Benjamin Haibe-Kains, Jason W. Locasale, Morag Park, and Mathieu Lupien. 2020. "Epigenetic Switch–Induced Viral Mimicry Evasion in Chemotherapy-Resistant Breast Cancer." *Cancer Discovery* 10(9):1312–29.

Additional publications during Ph.D.:

1. **Constanza Martínez**, Jonathan K. Lai, Daryl Ramai, Antonio Facciorusso, Zu-Hua Gao. "Cancer registry study of malignant hepatic vascular tumors: hepatic angiosarcomas and hepatic epithelioid hemangioendotheliomas". *Cancer Med*. 2021; 00: 1– 8. doi:10.1002/cam4.4403.

2. Moldoveanu, Dan, Vera Pravongviengkham, Gordie Best, **Constanza Martínez**, Tarek Hijal, Ari Nareg Meguerditchian, Mathieu Lajoie, Sinziana Dumitra, Ian Watson, and Sarkis Meterissian. "Dynamic Neutrophil-to-Lymphocyte Ratio: A Novel Prognosis Measure for Triple-Negative Breast Cancer". Ann Surg Oncol. 2020;27(10):4028–34.

Contribution of authors

HK helped with bioinformatic analyses on Fig 2.2.1. CK, SW, YY, NA, performed bioinformatic analyses on single-cell RNA sequencing data (Figure 2.2.3, Fig 2.2.4, Fig 2.2.5, Fig 2.2.15, and table 4). AF helped with FACS experiments. SC helped coding for computational analysis of stained images and methodology. EZ, YR generated the data for DSP experiments in Fig 2.2.11. AM performed mice measurements. PS generated the patient-derived xenografts GCRC1915 and GCRC2076. DZ performed the staining and provided technical support for other immunohistochemical and immunofluorescence studies. All other data were generated, analyzed, and assembled by CM.

Table of contents

Abstract……I
Résumé ……III
Acknowledgements……V
Contribution to Original Knowledge……VII
Preface……IX
Contribution of authors……X
Table of Contents……XI
List of Figures…… XIV
List of Tables……XIV
List of Abbreviations……XV
1. LITERATURE REVIEW……1
1.1. Breast cancer ……2
1.1.1. Epidemiology……2
1.1.2. Histopathology……3
1.2. Breast cancer subtype classification……8
1.2.1. Breast cancer histopathological classification……10
1.2.1.1. Hormone receptors (estrogen and progesterone) ……11
1.2.1.2. Human epidermal growth factor-2 (HER2) ……12
1.2.2. Breast cancer molecular subtypes……14
1.2.2.1. Intrinsic subtypes……14
1.2.2.2. Integrative-clusters 1-10……16
1.2.2.3. Triple-negative breast cancer subtypes……17
1.3. Breast cancer clinical challenges……19
1.3.1. Tumor initiation……21
1.3.1.1. Tumor-initiating cells……21
1.3.1.2. Model of tumor initiation……21
1.3.1.3. Experimental assays for tumor-initiating cells……22
1.3.2. Residual disease……25
1.3.2.1. Pathological complete response assessment and definitions……26
1.3.2.2. Residual disease biomarkers in triple-negative breast cancer……28
1.3.2.2.1. Immune microenvironment residual disease markers……28
1.3.2.2.2. Genetic and proteomic residual disease markers……32
1.3.3. Resistance to therapy……37
1.3.4. Metastatic disease……38
1.3.4.1. Metastatic cascade……38
1.3.4.2. Single cell vs collective cell dissemination……38
1.3.4.3. Metastasis organotropism……39
1.4. Tumor heterogeneity in cancer……41
1.4.1. Intra-tumor heterogeneity……41
1.4.1.1. Spatial and temporal heterogeneity……41
1.5. Tools to interrogate cellular heterogeneity upon disease progression……42
1.5.1. Patient-derived xenografts ……42
1.5.2. Single-cell RNA sequencing ……43
1.5.3. Spatial profiling in situ……44
1.6. Rationale……45

2. RESULTS
2.1 Breast cancer models to study clinical challenges ……...47
2.1.1 Biobanking breast cancer clinical samples……47
2.1.2 Patient-derived xenografts to model breast cancer disease progression…….48
2.1.3 Single-cell RNA sequencing of stepwise disease progression samples……...50
2.2 Spatial and temporal heterogeneity in breast cancer metastasis……...52
2.2.1 Patient-derived xenografts as models to study temporal heterogeneity……52
2.2.2 Patient-derived xenografts are representative models of the patient's histological, transcriptomic, and genomic landscape……..….54
2.2.3 Single-cell RNA sequencing of primary breast cancer patient-derived xenograft replicates.……..…..…57
2.2.4 Single-cell RNA sequencing of the breast primary tumor identifies eight distinct transcriptomic populations.…….……...59
2.2.5 The primary tumor replicates are comprised by 4 distinct biological cell populations and 4 transitional clusters..……62
2.2.6 Methodological selection of specific cluster markers from scRNA-seq data……65
2.2.7 Spatial mapping of scRNA-seq populations gene markers in PDX and PT……...67
2.2.8 In situ spatial cartography for scRNA-seq populations……69
2.2.9 Spatial cartography of primary breast tumor populations reveals zonal and zone-less populations adjacent to necrotic regions……..……72
2.2.10 Digital spatial profiling of patient-derived xenografts and patient's primary tumor and lung metastasis…..…..……75
2.2.11 Digital spatial profiling of CA9 positive regions correspond to scRNA-seq hypoxic populations……...77
2.2.12 Hypoxic CAIX positive cell populations are identified in early-stage tumors……80
2.2.13 Hypoxic CAIX positive cell populations display higher tumor initiating capacity in vivo……82
2.2.14 Primary tumor and tumorspheres spatial cartography on different oxygen conditions. ……..85
2.2.15 Breast cancer lung metastases retain primary tumor biological populations……88
2.2.16 Spatial cartography and scRNA-seq reveals that hypoxic and cholesterol homeostasis populations display an aggressive phenotype……90
2.2.17 Combined cholesterol and hypoxic targeted therapy display growth arrest and increased distant-metastasis free survival in PDX……94

3. GENERAL DISCUSSION……98
3.1. Breast cancer heterogeneity beyond conventional tumor heterogeneity……..…..99
3.2. The use of PDXs as preclinical models of disease progression……100
3.3. Single-cell RNA sequencing as a technology to deliver patient-centered care....103
3.4. Spatial heterogeneity in triple-negative breast cancer…….……...107
3.5. Spatial architecture and temporal heterogeneity of transcriptomic populations in primary and metastatic disease……...……..……109
3.6. Spatial cartography of primary and metastatic samples identifies aggressive populations……..……..……112
3.7. Conclusions……..…… 113

4. EXPERIMENTAL PROCEDURES.....114
4.1 Patient-derived xenografts.....115
4.2 Tumor dissociation into single cells.....115
4.3 Multiplex immunofluorescence.....115
4.4 Imaging by confocal microscopy.....117
4.5 Tumorspheres formation assay.....117
4.6 Histogel formalin fixed paraffin embedded tumorspheres and organoids.....117
4.7 Fluorescent-activated cell sorted (FACS).....118
4.8 Tumor formation and extreme limiting dilution analysis in vivo.....118
4.9 In vivo treatments.....119
4.10 In vivo metastasis assessment.....120
4.11 Bulk RNA and WGS sequencing.....120
4.12 Statistical analysis for in vivo experiments.....120
4.13 Single-cell RNA sequencing of PDX single cells.....121
4.14 Primary PDX joint analysis CCA and Harmony.....122
4.15 Post-clustering quality control.....123
4.16 Cell cycle regression and assessment.....123
4.17 Characterization of clusters with gene set enrichment.....124
4.18 Transcription factor activity inference.....124
4.19 Cluster marker selection for spatial mapping.....125
4.20 Spatial scRNA-seq population assignment.....125
4.21 Joint analysis of samples within and across conditions.....126
4.22 Differential expression between conditions.....128
4.23 Stress response effect.....128
4.24 Cell state trajectory inference.....128
4.25 Digital spatial profiling, Nanostring.....129
4.26 Digital spatial profiling differentially expressed genes.....130
4.27 Spatial deconvolution of cell populations identified by scRNA-seq.....130

5. BIBLIOGRAPHY.....131

LIST OF FIGURES

Chapter 1
Figure 1.1. Breast tissue histopathological transformation from normal to cancer.
Figure 1.2. Breast cancer subtype histopathological and molecular classification.
Figure 1.3. Breast cancer progression and key clinical challenges.
Figure 1.3.1.2. Tumor-initiating cell vs clonal evolution models in disease progression
Chapter 2
Figure 2.1.2. PDX models to mimic breast cancer disease progression.
Figure 2.1.3. Technical aspects of tumor dissociation and samples for scRNA-seq.
Figure 2.2.1. Schematic overview of primary and metastatic breast triple negative PDXs experimental pipeline.
Figure 2.2.2. Patient-derived xenograft mimics the molecular features of the patient's primary tumor.
Figure 2.2.3. Single-cell RNA sequencing of primary triple-negative metrics.
Figure 2.2.4. Primary joint cell cycle phase representation and regression.
Figure 2.2.5. Single-cell RNA sequencing of TNBC PDXs.
Figure 2.2.6. In-situ spatial mapping pipeline of single-cell RNA sequencing populations.
Figure 2.2.7. Marker selection is represented in both PDX and PT tumor.
Figure 2.2.8. In situ spatial mapping pipeline.
Figure 2.2.9. Spatial quantification of scRNA-seq populations in PDX and PT.
Figure 2.2.10. Digital spatial profiling of PT and matched PDXs samples.
Figure 2.2.11. Spatial deconvolution of scRNA-seq clusters through digital spatial profiling.
Figure 2.2.12. CA9 is a marker of hypoxic populations in early and late-stage tumors.
Figure 2.2.13. Primary tumor CAIX positive sorted cells have a higher tumor initiating capacity.
Figure 2.2.14. Spatial cartography of markers and scRNA-seq populations on different oxygen conditions.
Figure 2.2.15. Single-cell RNA sequencing of lung metastases cancer PDXs.
Figure 2.2.16. Spatial cartography of scRNA-seq populations in metastatic samples.
Figure 2.2.17. Combination therapy with CAIX and OSC inhibitors delays tumor growth and increases DMFS.
Chapter 3
Figure 3.3. Spatial tumor architecture and drug delivery approach.
Figure 3.4. Personalized approach to integrate spatial and temporal heterogeneity.
Figure 3.5. Transcriptomic populations identified in lung metastatic disease preserved their spatial zonation.

LIST OF TABLES

Table 1. Summary of studies assessing immune biomarkers for TNBC residual disease.
Table 2. Summary of research studies assessing genomic and proteomic biomarkers of TNBC residual disease.
Table 3. Antibodies used in experiments.
Table 4. Single-cell RNA sequencing metrics.

LIST OF ABBREVIATIONS

ADH – Atypical ductal hyperplasia
ALH – Atypical lobular hyperplasia
AJCC – American Joint Cancer Commitee
AR – Androgen receptor
BC – Breast cancer
CALML5 – Calmodulin like 5le
CA9 – Carbonic anhydrase 9
CAIX – Carbonic anhydrase IX
CCA – Canonical correlation analysis
CCM – Collective cell migration
CNB – Core needle biopsy
CRYAB – Crystallin alpha beta
CTC – Circulating tumor cells
CTZ – Cycling tumor zone
DAPI – 4′,6-diamidino-2-phenylindole
DCIS – Ductal carcinoma in situ
DDFS – Distant disease-free survival
DRFS – Distant recurrence-free survival
DSP – Digital spatial profiling
EGFR – Epidermal growth factor receptor
EMT – Epithelial to mesenchymal transition
ER – Estrogen
FACS – Fluorescent-activated cell sorted
FFPE – Formalin fixed paraffin embedded
GLUT1 – Glucose transporter 1
GOBO – Gene expression-based outcome for breast cancer online
GS: Genomic signature
HAGE – Helicase antigen
HER2 – Human epidermal growth factor receptor 2
HIF1 – Hypoxia-inducible factor 1
HRD – Homologous Recombination Deficiency
HTZ – Hypoxic tumor zones
IBC – Invasive breast cancer
IDC – Invasive ductal carcinoma
IF – Immunofluorescence
ISC – In-situ spatial cartography
LC – Lobular carcinoma
LCIS – Lobular carcinoma in situ
LDH – Lactate dehydrogenase
LOH – Loss of heterozygosity
MFP – Mammary fat pad
MJ – Metastases joint

OS – Overall survival
OSCi – 2,3-oxidosqualene: lanosterol cyclase activity inhibitor
PARP – Poly ADP ribose polymerase
PC – Principal component
PCA – Principal component analysis
PCNA- Proliferating cell nuclear antigen
pCR – Pathological complete response
PDX – Patient-derived xenograft
PDO – Patient-derived organoid
PJ – Primary joint
PgR – Progesterone receptor
PPAR -
PR – Progresterone
PT – Primary tumor
QC – Quality control
RNA-seq – RNA sequencing
ROI – Region of interest
scRNA-seq – Single cell RNA sequencing
SERM – Selective estrogen receptor modulator
SLC2A1 – Solute carrier family 2 member 1
SLM – Spontaneous lung metastasis
ssGSEA – single sample gene set enrichment analysis
TAI – telomeric allelic imbalance
TDLU – Terminal duct lobular unit
TIC – Tumor initiating cell
TIL – Tumor-infiltrating lymphocyte
TMA – Tissue microarray
TNBC – Triple-negative breast cancer
TNBCtype – TNBC subtype

TOP2A – Topoisomerase-2a
UCHL1 – Ubiquitin hydrolase L1
UMAP – Uniform manifold approximation and projection
UMI – Unique molecular identifier
VAF – Variant allele frequency
WGS – Whole genome sequencing

WS: Whole section

1. LITERATURE REVIEW

1.1 Breast cancer

1.1.1 Epidemiology

Breast cancer is a heterogenous and complex disease arising from epithelial cells in the breast tissue. Worldwide breast cancer is the most diagnosed cancer, accounting for 2.3 million cases per year and 11.7% of all cancers in the globe. In women, breast cancer is the most frequently diagnosed cancer and accounts for the largest number of cancer deaths in women. In the whole population breast cancer is the fifth most deadly cancer, responsible for 6.9% of all world cancer-related deaths (1). In females, breast cancer continues to increase over the years at a rate of 0.5% per year. This rise in the incidence is mostly explained by enhanced detection as well as changes in lifestyle contributing to overall weight gain delayed childbearing and overall lower fertility rate (2).

In Canada, cancer is the main cause of death surpassing cardiovascular diseases. Breast cancer is the most diagnosed cancer after lung and 4th deadliest cancer after lung, colorectal and pancreatic cancer. Every year 27,700 people are diagnosed with breast cancer and 5,100 die from it. In 2020, the 5-year overall survival for breast cancer patients was 88%. This is partly explained by efforts to improve screening strategies using mammography to detect cancer at earlier stages and emerging therapies that provide better outcome for patients. Despite considerable improvements in breast cancer detection, therapies and overall survival, this disease surpassed lung cancer deaths in 2020, and is still responsible for 685,000 deaths annually across the globe (3).

1.1.2 Histopathology

Breast cancer is not a single disease, it comprises distinct phenotypic entities. Breast cancer is usually identified through the microscope by analyzing histological clinical samples. Different genetic/epigenetic events can lead to the abnormal proliferation of either ductal or lobular epithelial cells from the terminal duct lobular units (TDLU) in the breast **(Figure 1.1 A).** These cells can then progress into a pre-malignant or malignant disease. Depending on the breast cell affected, premalignant lesions will be classified as either atypical ductal hyperplasia (ADH) or atypical lobular hyperplasia (ALH). After repetitive insults to the cell, premalignant lesions that do not invade the basement membrane can give rise to *in situ* carcinoma [ductal carcinoma in situ (DCIS) or lobular carcinoma in situ (LCIS)] **(Figure 1.1 B)**. Lastly, if aberrant uncontrolled proliferative cells invade the basement membrane, the lesion will now be considered malignant or "invasive" [invasive ductal carcinoma (IDC) or invasive lobular carcinoma (ILC)] **(Figure 1.1 C).** These events can occur throughout the span of years, however, normal epithelial cells of the TDLU can progress rapidly into invasive lesions or present concomitantly with different stages of disease in the breast **(Figure 1.1 D)**.

Figure 1.1

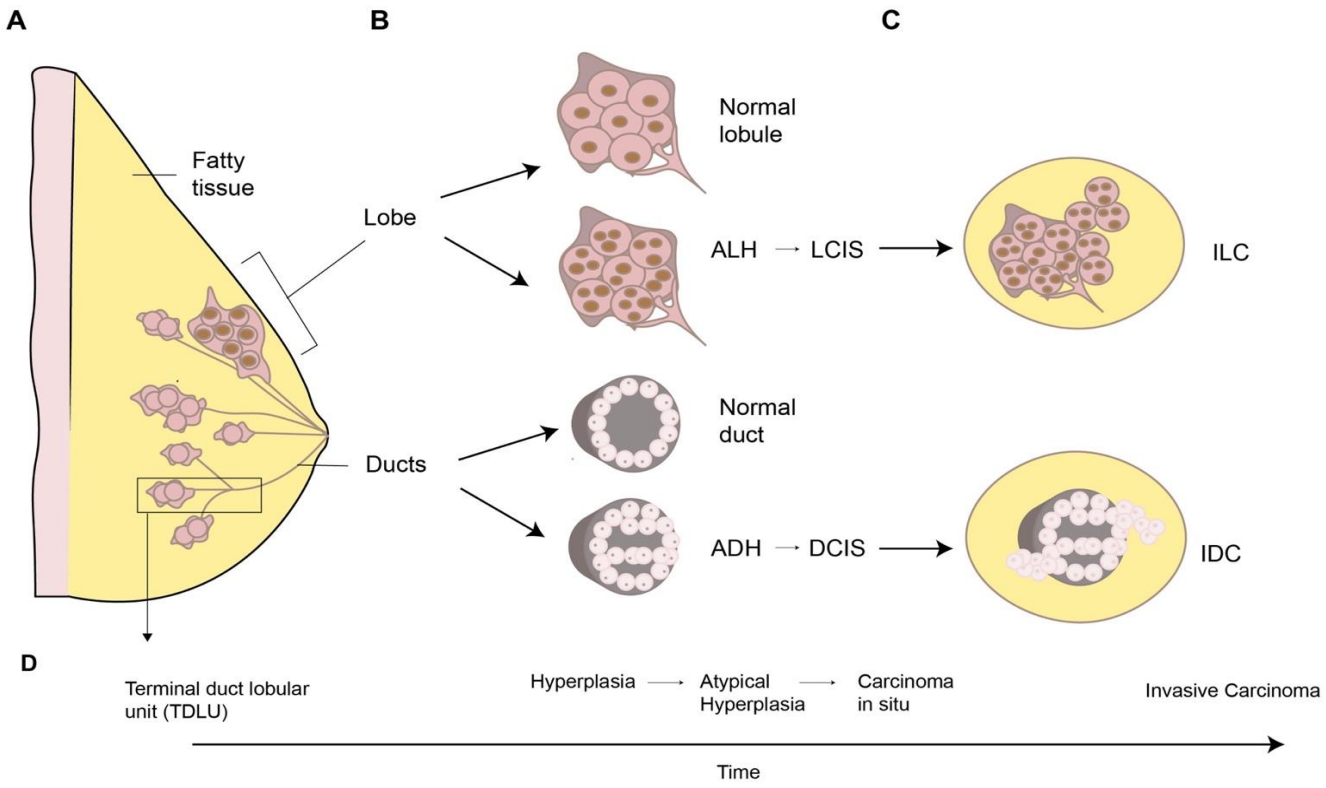


Figure 1.1. Breast tissue histopathological transformation from normal to cancer.

A. The breast is comprised by fat tissue and multiple ducts and lobes breast cell that together form the terminal lobular duct units (TDLU). **B.** Lobes and ducts when they undergo increased proliferation can give rise to atypical ductal hyperplasia or atypical lobular hyperplasia. **C.** Ductal or lobular cells that invade the basement membrane transformed to invasive ductal carcinoma or invasive lobular carcinoma. **D.** The progression of disease can evolve throughout years at different speed and with concomitant stages of disease. The classic stepwise progression starts with hyperplasia of either TDLU, followed by hyperplasia with atypia of the cells, then into carcinoma in situ and lastly into invasive carcinoma.

The architectural heterogeneity in breast cancer pathogenesis is readily observed at the histological level by the complexity and mixture of different histopathological features and regulation by different transcriptional programs (4). The progression of normal epithelial cells into a premalignant lesion has well described architectural features – excessive proliferation, pathological filling and distension of ducts or acini, and an intact basement membrane (5,6). However, the distinction between atypical epithelial hyperplasia and carcinoma in situ of the breast relies on mostly the absence or presence of several of these morphological features. The diagnosis is a continuum and a challenge for pathologists and clinicians who have to determine the best treatment for these patients (7). Treatment for DCIS can vary. In some cases it can be either chemotherapy prior to surgery (neoadjuvant), chemotherapy post-surgery (adjuvant), surgical resection alone or in other cases “watchful waiting” by doing regular imaging follow-ups (8).

DCIS is the main type of carcinoma in situ (80-90%). Due to breast cancer screening with mammography, the diagnosis of DCIS has increased over the years, being 20-30% of all newly diagnosed breast cancers. This immediate precursor of invasive breast cancer (IBC) is divided into different DCIS types depending on the architectural pattern and histopathological features. There are six different histological types based on their growth patterns; comedo, cribiform, solid, micropapillary, papillary and mixed type. To add to the complexity, DCIS type does not strictly determine the aggressiveness of the cancer. It is the grading system (low and high grade) that will help establish the most appropriate intervention, which in most cases is debatable (9).

Progression into an invasive breast carcinoma will be defined by the invasion of the basement membrane. Although, in IBC there is almost no doubt of intervention in most cases, this entity poses a challenge from an architectural, molecular and therapeutic perspective due to its heterogeneity (10). Cells at this stage, acquire at different rates, distinct phenotypes displaying diverse hallmarks of cancer. Cells can exhibit uncontrolled proliferation, lack of growth suppression, anti-apoptotic or cell death mechanisms, development of angiogenesis or neovascularization, dissemination and/or metastasis (11). All hallmarks of cancer can be present at time of diagnosis or be acquired throughout disease evolution. The exposure to new events or signals coming from neighbouring tumor or stromal cells will also play a role in the molecular adaptation of the cancer cells. Different molecular mechanisms (genomic, transcriptomic or epigenetic) can trigger changes at a single cell level (12). Therefore, the understanding of the spatial localization and disposition in space of every single cell regarding other cells is essential to integrate the changes that can occur either through cell-to-cell (C-C) or cell-to-environment (C-E) interactions. For instance, adjacent cells can rely on the exchange of hormones, cytokines, exosomes, and other bidirectional communication ques to adapt to a specific event or environmental change. Moreover, C-E interactions exert selective pressures for fit cells (13,14).

Harsh environmental conditions in the tumor environment, such as in hypoxic/acidic pH/low nutrient availability tumor zones will inevitably select for cells already adapted or with the machinery to adapt and survive, posing a clinical challenge at the time of delivering treatment (15). At the time of treatment selection, all these molecular changes at the single cell level are not globally integrated to make therapeutic

decisions. The treatment decision is still mostly based on the clinical classification of breast cancer through the expression of receptors, specific mutations that can grant benefit for specific targeted therapy, and lately the incorporation of the immune compartment to identify patients suitable for immunotherapy

1.2 Breast cancer subtype classifications

Breast cancer subtyping depicts the level of heterogeneity of this disease. Many attempts considering different molecular alterations have been described. The most common breast cancer subtype used in the clinic, is the histopathological classification. Breast cancers are mostly divided into three subgroups depending on the expression of receptors: hormone receptor positive (estrogen and progesterone), HER2 receptor positive, the expression of both HR and HER2, and triple-negative breast cancer that lacks all receptors. Moreover, the complexity of the tumor is further characterized by the type of histology present in the tumor and its histological grade. These clinical variables will help determine the aggressiveness of the tumor, intervention, and prognosis. The histopathological classification helps to tailor treatment and offer patients the opportunity to receive targeted therapy. The molecular characterization based on gene-expression (transcriptomic) and on both mutational profiles and copy number aberrations (genomic) further characterizes the disease and identifies again differential prognosis. Lastly, within the TNBC subgroup, that lacks targeted therapy, efforts to redefine the disease into TNBC subtypes (TNBCtypes) has shed light on the heterogeneity of the disease, helping on the elaboration of more specific clinical trials and therapeutic strategies **(Fig.1.2A).**

Altogether, these classifications and subtyping can have many possible combinations, making each tumor different from each other **(Fig.1.2B)**. The inter and intra-tumor heterogeneity in breast cancer is reflected trough the inability to describe breast cancer in only a few subtypes. The prognosis and treatments will take into consideration all these variables and subtypes, making personalized treatment an unmet clinical need **(Fig.1.2C).**

Figure 1.2

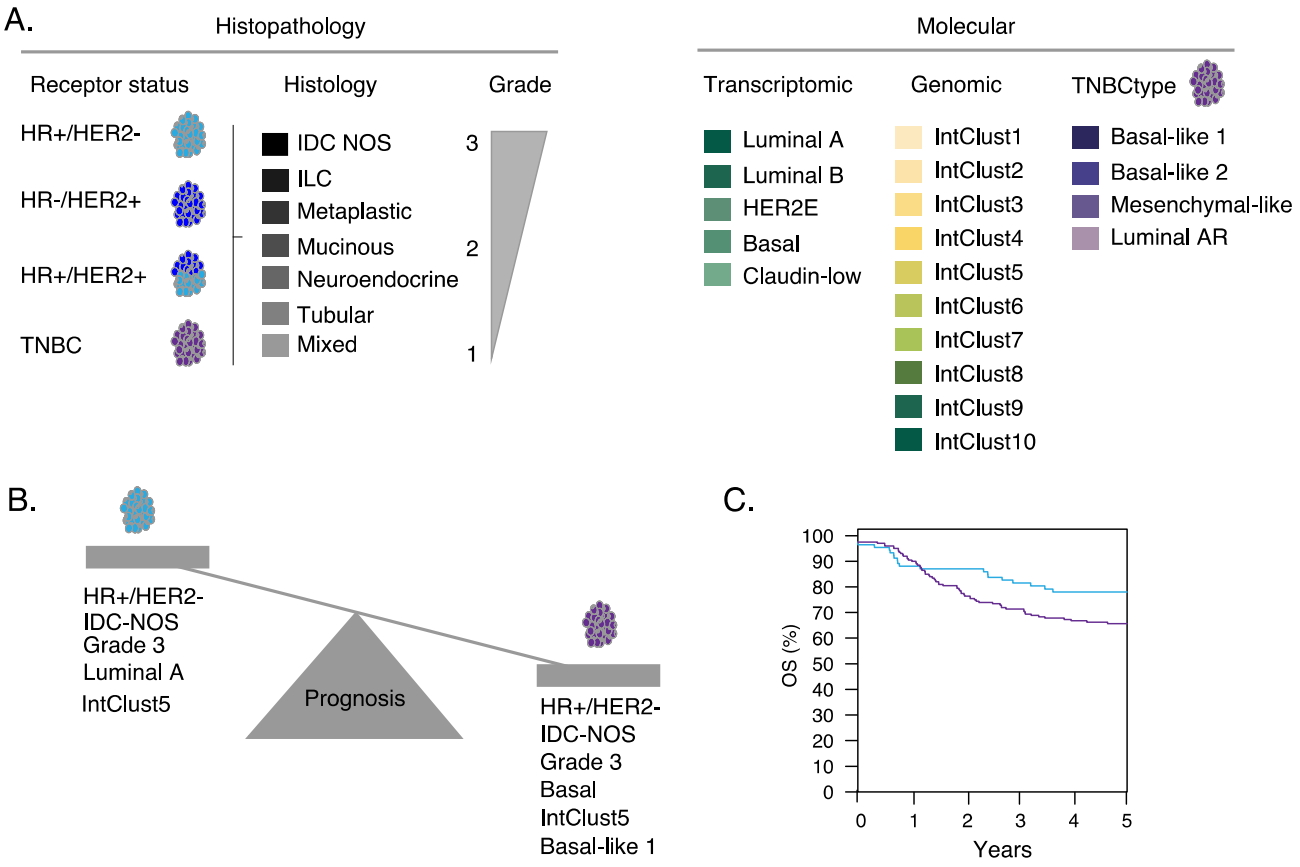


Figure 1.2 Breast cancer subtype histopathological and molecular classification.

A. The histopathological classification is determined by the expression of hormone receptors (ER, PgR), HER2 receptor, the expression of both HR and HER2, or the expression of none. The lack of receptors described the TNBC subtype. The molecular classification describes the subtypes by gene-expression and the genomic Integrative clusters (IntClust). Moreover, TNBCtypes are described consequently **B.** Prognostic balance illustration of two distinct breast cancer classification. **C**. Graphical overall survival for the two different breast cancer subtypes described in B.

1.2.1 Breast cancer histopathological classification

The complexity of the cell populations from a morphological point of view is comprehensively appreciated by the existence of an immense amount of breast cancer histological subtypes. A recent update of the fourth edition of breast tumors by the World health organization (WHO) displays how dynamic tumor classification is by re-defining BC subtypes into even more entities. The most frequent histological subtype in breast cancer is invasive ductal carcinomas (50-80%), no specific type or IDC-NST. This is due to the existence of specific subtypes with very distinct morphological features. The most frequent specific subtype is invasive lobular carcinoma, accounting for 15% of all BC subtypes. Other specific subtypes represent the remaining 5%. The following diseases can be found within the specific subtypes histology: comedo, cribiform, solid, micropapillary, papillary, adenoid cystic, metaplastic, neuroendocrine, mixed and others (16). As important it is to better understand and classify these tumors, the histological subtypes of breast cancer at the moment do not change practice, even-though, prognosis and treatment response differ among these (17,18). Moreover, the prognostic factors of breast cancer are mostly determined by tumor grading, which is mostly assessed using the Elston/ Nottingham grading (19,20). In brief, this scoring system determines the level of glandular differentiation, nuclear pleomorphism, and mitotic counts (this last criterion was updated to mm^2 by WHO in 2019). Although histological grade is known to be a strong prognostic factor, the histopathological subtyping of breast cancer does not change standard of care (21).

The molecular histopathological classification of breast cancer importance relies on the assessment of the expression of different receptors in the cells. This classification

is the most clinically used because it determines the treatment. An initial biopsy will undergo immunohistochemical (IHC) staining for assessment of known biomarkers of breast cancer. The clinical classification of invasive carcinoma of the breast will be determined by histology and the expression of estrogen (ER), progesterone (PgR), and of human epidermal growth factor 2 (HER2 also known as ERBB2) (22).

1.2.1.1 Hormone receptors (estrogen and progesterone)

The estrogen receptor positive breast cancer accounts for 70% of all breast cancers. Estrogen is a nuclear hormonal receptor that dimerizes when estradiol binds to it. The conformational changes in ER receptor initiates transcription and cell proliferation(23). Tamoxifen, a drug used to inhibit this pathway became the first targeted therapy in breast cancer. In 1966 tamoxifen was initially produced as an emergency contraceptive method developed by ICI pharmaceuticals (now AstraZeneca)(24). Parallel research, understanding the mechanism of action, and initiatives by Dr. Walpole, led to selective estrogen receptor modulator (SERM), becoming the standard of care treatment for hormone receptor positive breast cancer (25). The spectrum and available drugs for pre and post-menopausal ER positive breast cancer patients was further extended, including aromatase inhibitors (Anastrozole/Letrozole), gonadotropin-releasing hormone agonists (Goserelin) and selective estrogen receptor modulators (Fulvestrant)(26–28). These drugs not only provided breast cancer with a post-surgical recurrence reduction of average 40%, but also became important in other cancer types, such as prostate, where hormonal regulation is essential for tumor growth(29,30).

The progesterone receptor (PgR) was subsequently identified as prognostic and predictor IHC biomarker for breast cancer. As PgR expression is regulated in an ER-dependent manner, ER the main receptor used to assess hormone-receptor (HR) positivity(31). Therefore, now it is determined that any case where ER is negative and PgR is positive, requires a repeat test. The final assessment of HR (ER or PgR) positivity, the presence of 1% to 100% positive cell's nuclei predicts benefits for the use of endocrine therapy for breast cancer patients. (32). Although the use of endocrine therapy achieves a 5-year overall survival that surpasses 90%, recurrent disease remains a long-term clinical challenge when observing poor 15-years survival rates (33,34). This may arise from tumors that acquire endocrine resistance, such as mutations within ER as well as new targetable alterations (CDK4/6) that have been identified and therapeutic intervention targeting these is now being studied in different clinical trials (35,36).

1.2.1.2 Human epidermal growth factor-2 (HER2)

HER2 plays an important role in breast cancer pathogenesis. Overexpression of HER2 promotes formation of homodimers or heterodimers that interact with different members of the HER family. Upon dimerization and subsequent activation of the intrinsic kinase activity, multiple intracellular signaling cascades are triggered, promoting cell proliferation, survival, invasion, and metastatic progression (37). HER2 amplifications or overexpression can be present in up to 15-20% of breast cancer patients (38). HER2 alterations are strongly associated with poor prognosis, which prompted the development of HER2 targeted therapies. These therapies have improved patient recurrence-free survival either with a single monoclonal antibody (Trastuzumab), dual antibody blockade

(Trastuzumab + Pertuzumab) or small molecule inhibitors (Lapatinib) (39–41).Therefore, the clinical assessment of HER2 is routinely done in every breast cancer patient either by IHC or Fluorescent in situ hybridization (FISH). If IHC is utilized a score of +3 will be considered positive. If FISH is used a value of 2.0 or more will be considered positive (42,43).

Lastly, the absence of these receptors will comprise the subtype called triple-negative breast cancer (TNBC). Patients with TNBC are not suitable candidates neither for hormonal nor anti-HER2 targeted therapy. The molecular histopathological classification determines the patient's candidacy for treatment and disease prognosis. Patients with TNBC subtype, due to the lack of therapeutic targets, still receive the standard of care treatment that consists of neoadjuvant chemotherapy (NAC), surgery and radiation therapy and are known to display a more aggressive disease and present metastatic disease within the first three years of diagnosis (44). The neoadjuvant regimens of chemotherapy are comprised anthracyclines (Doxorubicin), taxanes (Paclitaxel), alkylating agents (Cyclophosphamide) and platinum-based agents (Cisplatin/carboplatin). This subtype constitute ~15% of all breast cancers and together with HER2+ patients display higher rates of pathological completes response (pCR) after NAC when compared to HR+ tumors (44,45). The use of Platinum-containing agents (Carboplatin) and its positive association in BRCA mutated patients outcome, changed the initially described pCR rates of ~40% in TNBCs to rates to ~50-60% (46–48). For this reason, pCR started to be used as an outcome measure in different clinical trials to shorten their duration and speed application to clinics (49–52).

1.2.2 Breast cancer molecular subtypes

1.2.2.1 Intrinsic subtypes

Following genome wide and transcriptomic studies of large breast cancer cohorts, BC was further classified into different molecular 'intrinsic' subtypes based on distinct gene expression patterns. A landmark gene expression microarray study by Perou, C and Sorlie, T using breast cancer bulk tumors, defined breast cancer molecularly into 5 subtypes: Luminal A, Luminal B, (ER positive), HER2-enriched (HER2E), Basal-like and Normal-like. All these subtypes reflect components of transcriptional profiles of normal cell types found in the breast epithelium identifying hierarchy of differentiation within BC subtypes(53). This molecular classification highlights how each BC subtype gene expression is defined in part through its identity and environmental localization.

Luminal A and B subtypes present with markers of breast luminal cells. These markers can be expressed to varying degrees, some of them being ER, CK8/18, GATA3. Luminal A tumors express higher levels of ER compared to Luminal B, and demonstrate expression of LIV-1, HNF3A, XBP1, and GATA3. Luminal B tumors have characteristic expression of GGH, LAPTMB4, NSEP1 and CCNE1 (54). Moreover, the Luminal B subtype of BC display worst outcome compared with luminal A and are more closely related to basal and HER2E subtypes due to their shared enrichment for genes involved in proliferation. HER2E subtype tumors are characterized by the overexpression or amplification of gene expression patterns reflective of the HER2 amplicon (17q21) (54). The Basal subtype are enriched for basal breast cell epithelial markers. Some of these include CK17, CK5/6 and EGFR. Also, this subgroup is characterized by absence of

expression for luminal/ER related genes. Lastly, Normal-like tumors displayed enrichment in genes that can be found in adipose tissue and stromal cells in the breast (PIK3R1, AKR1C1, FACL2) and may be tumors with low tumor cell content (54).

The importance of these breast cancer subtypes discovered in the early 2000 relied on the ability to be associated with treatment response and patient prognosis (55,56) Luminal and Normal-like subtypes have better overall survival, response to treatment and less proliferative capacity when compared to HER2E and the Basal subtype that lacked targeted therapies (57,58). This initial study included a panel of 496 genes, which was redefined to a 50-gene subset that could predict intrinsic subtypes from different sequencing studies. This new prediction analysis of microarrays of 50 genes (PAM50), is widely used to determine the molecular subtype of different breast cancers from gene expression data (59).

These subtypes were further classified into a hierarchical model that resembled either a more a luminal-like (differentiated) or mesenchymal-like (undifferentiated) state (56). A Claudin-low subtype was included into the list of intrinsic subtypes (60). This subtype is characterized by high mesenchymal features, low luminal expression and enrichment for CD49f (61). These tumors are considered to be enriched of cell with features of cancer stem cells and/or reflect an epithelial to mesenchymal de-differentiation program and are more correlated with BRCA1 genomic alterations (62). These intrinsic subtypes correlate with the previously described clinical subtypes based on IHC. Luminal A and B are hormone receptor positive, HER2E in 50% of cases is ER negative, and most of Basal subtypes are TNBC (ER- , PR- and HER2-) (60).

1.2.2.2 Integrative clusters 1-10

Although the molecular classification of breast cancer built grounds for the use of gene signatures and gene sets as prediction tools (Mamma-print 70-gene signature/ Oncotype DX) for treatment response, this was insufficient to explain unresponsive patients being predicted to respond using these criteria (63,64). Attempts to redefine subtypes came with the integration of both genomic and transcriptomic data, classifying breast cancer into 10 Integrative Clusters (IntClust 1-10) (65). This study used genomic data from the METABRIC dataset, describing new subtypes with distinct chromosomal alterations. The new subtypes were associated with the previously described intrinsic clusters and with prognosis. IntClust5 correspondent with a HER2E intrinsic subtype, displayed the worst overall survival, however, this was prior to the incorporation of the results of Trastuzumab use related trials (66).

The wide use of genomic and sequencing technologies in the early 2000 led to a better understanding of the heterogeneity of breast cancer. Breast cancer is characterized by a long tail of genomic alterations. Some of these are commonly found in a larger subset of patients. For instance, the most frequent single-nucleotide variants (SNVs) are TP53 (37%), PI3KCA (36%) and GATA3 (11%). The remaining mutations are observed in less than 10% of patients. Moreover, common drivers of breast and other tumor types have been consistently identified through different genome-wide studies. This has helped to understand the impact in breast cancer of low frequency genomic alternations observed in other malignancies (67). In respect to somatic copy number aberrations (CNA), the most frequent alterations in breast cancer are gains in chromosome 1q, 8q, and 20q, while losses are observed on 5q, 8p, 13 and 16q. (66).

1.2.2.3 Triple-negative breast cancer subtypes

The growing interest of understanding triple-negative or basal breast cancer, due to the lack of targeted therapies and aggressiveness of the disease, led to studies centered in pinpointing the molecular differences among TNBCs. Although ~70% of TNBC are of the basal breast cancer subtype as defined by PAM50, in 2011 this was revisited and TNBC was divided into 6 TNBC subtypes. This stratification was performed based on gene expression data extracted from 21 previously reported datasets. Gene-ontology characterized TNBC into six transcriptionally defined TNBC subtypes (TNBCtypes) as follows: Basal-like 1 (BL1), Basal-like 2 (BL2), Immunomodulatory (IM), Mesenchymal-like (M), Mesenchymal Stem-like (MSL) and Luminal androgen receptor (LAR) (68). This classification was redefined by the same group five years later when noting that laser capture microdissection (LCM) had considerable differences when compared to TNBC TCGA gene expression profiles. The latest TNBCtype describes 4 subtypes: BL1, BL2, M, and LAR (69).

BL1 subtype is characterized by high cell cycle and DNA damage repair genes, where patients achieve high pCR rates (52%) in response to NAC. BL2 is a subtype enriched for myoepithelial markers and growth factor signaling pathways. This subtype has poor response to NAC (0%) and displays the worst overall survival from all four subtypes. The M subtype has high expression of genes involved in epithelial-to-mesenchymal transition (EMT) and some genes involved in cell proliferation. This also is the subtype with least tumor infiltrating lymphocytes (TILs), pCR rates of 23%, and higher lung metastatic disease (46%) when compared to the other TNBCtypes (24%). Lastly,

the LAR subtype displays androgen receptor signaling and luminal gene expression patterns, with low pCR rates (15%) and is predominantly diagnosed in older women (70). The previously described IM presented considerable amount of immune infiltration being excluded from the new classification. Similarly, the MSL subtype displayed a high content of stromal signatures , being also excluded from the new TNBCtype-4 groups (69).

Although intensive efforts to stratify breast cancer and TNBC type have been intended by many groups, none of these studies or subtypes is entirely applicable in a personalized manner to a one size fits all for breast cancer patients. This difficulty in classifying this disease highlights the inter-tumor heterogeneity observed within patients.

1.3 Breast cancer clinical challenges

Breast cancer displays around 88% overall survival after 5 years follow-up (71). Certainly, many patients do not die from this disease at early stages but some of them do after years of developing local recurrences or distant metastasis. Nevertheless, different strategies have been developed at different levels of prevention to maximize detection and minimize progression of breast cancer to advanced stages. First, at the secondary prevention level, annual mammography is now recommended for all women over 45 years and up to 75 years old. This strategy has increased detection of breast cancer in early stages and improved outcomes. Genetic testing is an alternative reserved for people with known germline mutations in the family, or in people with evidence of close degree relatives that developed cancer at a young age. Although, the use of these prevention strategies aids detection of breast cancer, it does not prevent breast cancer by tackling tumor initiation (TI). Once TI occurs, therapy is given with an attempt to eliminate the disease. The use of chemotherapy, radiation therapy and surgery are considered the standard of care. After therapy is given, neoplastic cells often remain and can be detected either macroscopically or microscopically, can give rise to what is known as residual disease (RD). RD is one of the biggest clinical challenges and it could be considered a "point of no return". From here, RD cells can give rise to local recurrence/rebound disease, resistant disease, and ultimately disseminate through the blood stream and promote distant metastases. Moreover, tumor cells prior to treatment can also give rise to metastasis. Those tumors that metastasize are still not fully elucidated and especially those that display metastatic dissemination at early tumor stages. Altogether, this is still a large clinical unmet challenge in breast cancer **(Fig 1.3)**.

Figure. 1.3

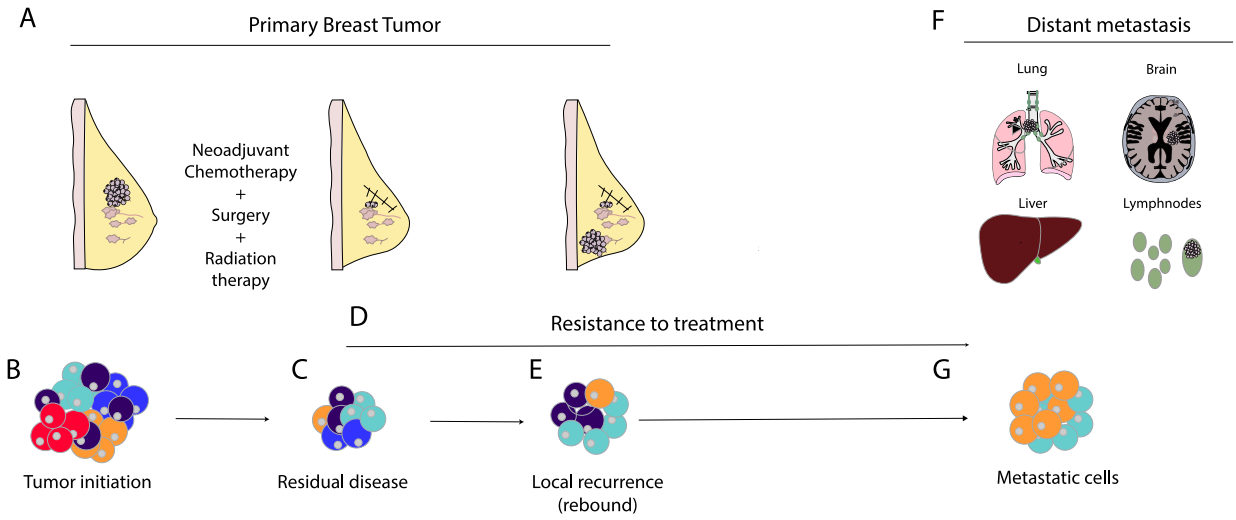


Figure 1.3. Breast cancer progression and key clinical challenges.

A, B Breast cancer illustrative image (top) and corresponding tumor initiating cells (bottom). **C.** Remanent cells after treatment will give rise to residual disease. **D** From RD state, resistant tumor cells can arise and contribute to disease progression. **E**. Residual disease or resistant tumor cells that have survived can proliferate and start relapse disease with cells already having a more aggressive phenotype. **F, G** Lastly, metastatic disease can arise at any of the previous stages. Cells can metastasize faster or be dormant for a longer time before disseminating to distant organs.

1.3.1 Tumor initiation

1.3.1.1 Tumor initiating cells

Tumor initiation is the process whereby, tumor initiating cells (TIC), a special subset of cells with unique capacities will give rise to a tumor. The concept of TICs Is often referred to as cell of origin and cancer stem cells (CSC). TICs refers solely to the capability of cells to initiate tumor growth, cell of origin defines a specific cell type that transformed due to an initial oncogenic event and gives rise to TICs, and CSCs are defined by stem cell properties (self-renewal and differentiate into distinct cell types which may have different properties from TICs (72).

Although in the literature these terms are used sometimes indistinctively, throughout this work TIC will be the term to be used. TICs can be found in different solid tumors and have been well described in breast cancer. TICs in breast cancer were initially identified in vivo mouse models, were as few as 100 cells were able to give rise to transplanted tumors (73,74). Most TICs present with self-renewing capacity and the ability to proliferate and generate non-TICs, repopulating and contributing to new tumor formation and as such fulfill the definition of CSC.

1.3.1.2 Models of tumor initiation

There are two models that aim to describe tumor initiation. First the tumor-initiating cell model describes a hierarchical organization of tumor cells, with the those found at the apex having self-renewing, proliferating and differentiation capacities (75). These cells can survive and adapt to different environmental conditions and selective pressures. This was demonstrated By Bonnet and Dick in 1997 in acute myeloid leukemia, where only a

small subset of cells (0.01-1%) were capable to reproduce leukemias when performing serial transplantation in immunocompromised mice. Second, the clonal evolution model proposes that cells inherit survival and proliferative advantages through stochastic genetic and epigenetic alterations (76,77). The clonal evolution model is not hierarchical, any cell at any time can accumulate molecular alteration that would lead to cell subpopulations selection and the subsequent expansion of cells with specific survival fitness. These factors ca also come from the surrounding tumor microenvironment and therefore, from different localizations in the tumor (78) **(Fig 1.3.1.2)**.

1.3.1.3 Experimental assays for tumor-initiating cells

The functional plasticity displayed by TICs can be addressed both in vitro and in vivo. In vitro, the use of TIC markers can be used to enrich for TIC populations. For example, in breast cancer cells expressing (CD44high/CD24low or ALDH1) are representative of a TIC population. To assess TIC self-renewal and proliferating capacity, tumorspheres can be cultured in non-adherent conditions and be subsequently serially propagated for quantification and TIC frequency determination (79). A downside of *in-vitro* TIC models is that they set pre-established conditions for tumor cells. For example, cells that are in culture are exposed to the same levels of oxygen from an incubator and nutrients found in cell culture media. Therefore, a gold standard assay for TICs is the use of limiting dilution assays *in-vivo*. This test evaluates tumor forming capacities and frequency in mice. It is normally set up, having a high and a low number of inoculated cells with the target population, aiming to determine the TIC frequency of a specific tumor or cancer cell line. Moreover, TICs have some defined gene expression profiles that can

be correlated with the tumorigenic capacity of breast cancer intrinsic subtypes (80,81). In essence, tumor recurrence, resistant to therapy and metastasis requires TICs to self-renew and differentiate to propagate these distinct phenotypes.

Figure 1.3.1.2

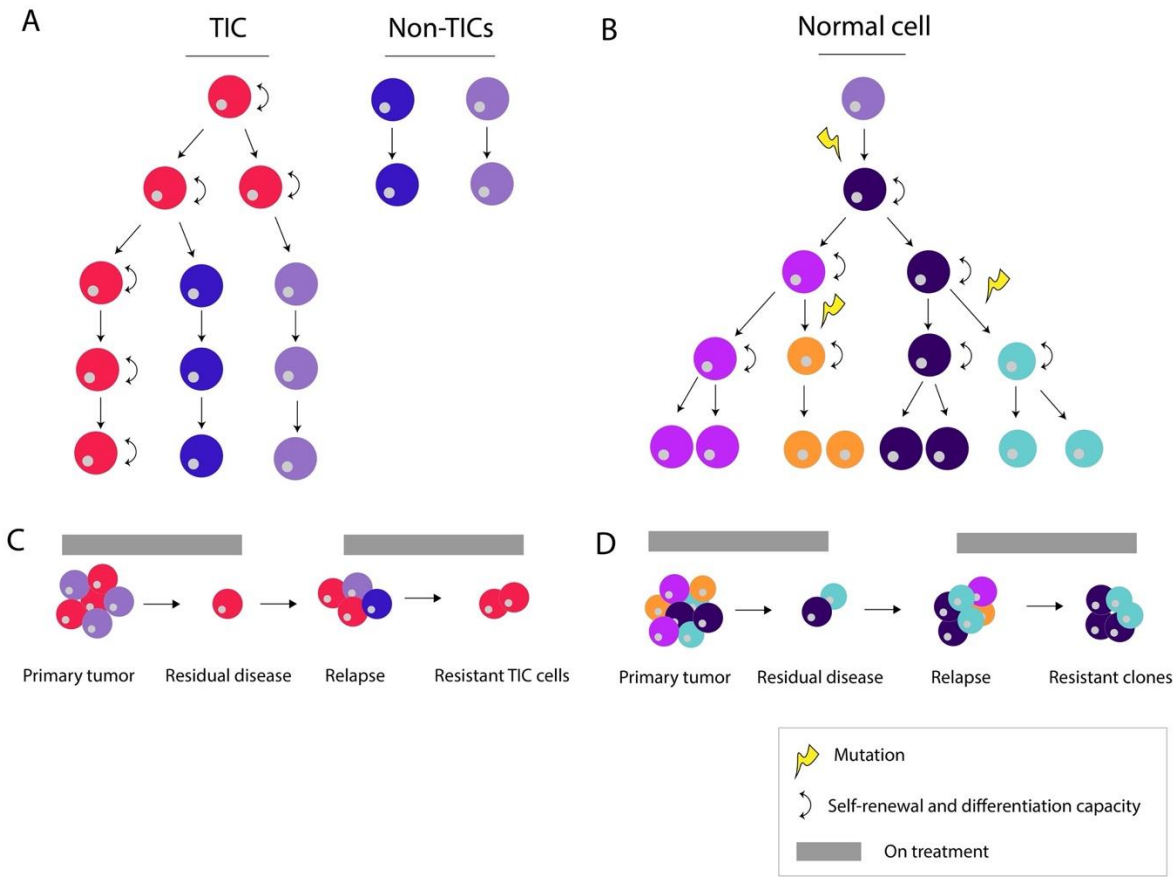


Figure 1.3.1.2. Tumor-initiating cell vs clonal evolution models in disease progression.

A. In tumor-initiating model TICs non-genetic alterations lead to self-renewal and differentiation capacity. This progeny of TIC cells can be TIC and non-TIC cell populations. The non-TIC cells will only have daughter cells along their differentiation path **B.** The clonal evolution model depicts cells that acquired distinct mutations that conferred their progeny with survival advantage and fitness. They can acquire more mutations along disease progression and give rise to distinct clones that will conform the bulk of the tumor. **C.** Under the hierarchical TIC model, the primary tumor is comprised by TIC and non-TIC cells. After treatment cells will require TIC capacity to evolve into a relapse and form a new tumor. Intrinsically, after treatment, TIC cells are resistant to therapy and will select for cells with a more aggressive phenotype along disease progression. **D** The clonal evolution model primary tumor will be represented by different cell populations that acquired different mutations. These genomic alterations will confer them self-renewal and differentiation capacity. After treatment, residual disease will be comprised by resistant clones that later will proliferate to form a clinical relapse. These resistant clones will all have specific mutations, and the new relapsing tumor will be mostly comprised by these selected clones and new mutations that come after this.

1.3.2 Residual disease

The use of neoadjuvant chemotherapy (NAC) as standard of care in TNBC and HER2+ BC, consists of treatment given prior to surgery. NAC allows the shrinkage of tumors, and in many cases the opportunity of surgical resection in previously inoperable patients. Apart from increasing chances of surgical resection, the use of NAC and pathologic complete response (pCR) are associated with improved outcomes among breast cancer patients. Initially, the responses described within the different BC subtypes were for HR+ BC ~10%, HER2+ 60% and TNBC ~30% (45,46). However, pCR response rate is a dynamic parameter since it is constantly changing with the development of new clinical trials and new therapies. Notably, the rate of pCR response is considerably higher in patients with HER2+ and TNBC (51,52). Moreover, higher pCR rates are also associated with increased recurrence-free survival (RFS) in TNBC (82–84) . In addition, the lack of pCR after NAC, is a strong predictor of patients with poor overall survival (OS) and distant-metastasis free survival (DMFS)(85). Particularly for TNBC patients, there are other clinical-pathological factors that are important prognostic factors that predict poor outcome and long-term survival of patients. The following have been described: axillary lymph node status (ALN), Lymph Node Ratio (LNR), low nodal positivity rate (LNPR) and Lympho-vascular Invasion (LVI) ha are important prognostic factors predicting poor outcome and long-term survival of patients (86–89). Overall, residual disease is an important parameter to assess disease progression and survival outcomes in patients, deserving attention from a clinical and research perspective. In TNBC the neoadjuvant setting opens many different research avenues to understand differences

between tumors that respond versus those that fail to respond, and to better evaluate pCR within these breast cancer patients.

1.3.2.1 Pathological complete response assessment and definitions

Once the presence of pCR or absence of RD post-NAT became approved as an endpoint for accelerated drug approval by the Food and Drug Administration (FDA), a standardized definition and assessment was required (90). Pathological complete response histopathological definitions are the following:

a. The absence of residual invasive cancer cells after neoadjuvant treatment in the complete resected specimen and all regional lymph nodes through H&E assessment. This American Joint Committee on Cancer (AJCC) definition is short *for ypT0/Tis ypN0* (91).
b. The absence of residual invasive and in situ cancer cells after neoadjuvant treatment of the complete resected specimen and all regional lymph node. This AJCC definition is short for ypT0 ypN0 (91).

Moreover, an alternative definition to the AJCC RD was developed in 2007 at MD Anderson Cancer Centre (MDACC). The concept of Residual Cancer Burden (RCB) arose from the fact that many BC patients do not achieve pCR after NAT, but their degree of RD was different(92). The RCB is an index subdivided into 4 categories and it combines two variables: the pathological findings in the primary tumor bed and regional lymph nodes. The categories are RCB-0 (pCR), RCB-I (minimal residual disease (MRD)), RCB-II (moderate RD) and RCB-III (extensive RD) (93).

There have been complementary methods to assess for residual disease using imaging technologies. Different modalities such as Magnetic Resonance Imaging (MRI), Contrast-enhanced mammography (CEM) and post-NAC image-guided biopsy have been used to detect RD in a less invasive manner. MRI and CEM are commonly used in patients undergoing NAT and surgery to assess tumor volume before and after treatment. In general, the accuracy of these methods when RD is 5mm or more (~90%) (94,95). Both modalities have higher accuracy in TNBC and HER2+ BC patients and overall indistinct of BC subtype the positive predictive values are around 70-80% (94,96). Image-guided biopsy have been use in different centers and trials with the intent to avoid surgery in patients with pCR (97,98). Although, this is an invasive method, the early stratification of patients that could spare surgery could improve BC patient's quality of life and different trials are intending to assess this (99).

Altogether, the variety of definitions and assessment modalities of RD and pCR emphasize the clinical need for the identification of pCR biomarkers and a thorough understanding of patient treatment response in regards to NAT (100).

1.3.2.2 Residual disease biomarkers in triple-negative breast cancer

1.3.2.2.1 Immune microenvironment RD markers

Many studies have revealed the role of immune cell crosstalk not only in tumor progression and metastasis formation, but also in shaping the intra-tumoral microenvironment. A summary of the main research studies can be found in **(table 1)**.

There is a growing understanding of the tumor immune microenvironment (TIME) role and its influence on response to therapy (101,102). TIME-related biomarkers have been utilized to guide therapy selection. Is now recognized that cytotoxic agents can exert their antitumor activity by inducing immune responses against tumor cells. Tumor-infiltrating lymphocytes (TILs) are mononuclear immune cells that can infiltrate the tumor and its quantification and localization is associated with distinct treatment responses (103).

For instance, high levels of TILs within the tumor are predictive of NAC pCR , increased disease-free survival (DFS), and OS in multiple clinical trials (104,105)**.** Also, High pre-NAC TIL levels or scores are an independent predictor of pCR and improved OS in TNBC (106,107). Similarly, low pre and post-NAT TILs in TNBC are associated with poor prognosis and shorter RFS (108). However, when studying post-NAC TILs, Hamy et al, described no association with prognosis in 716 pre and post NAC matched paired TNBC specimens (109). Similarly, high TILs post-NAC TNBC RD are correlated with better RFS and OS (110). In addition, a change in TILs levels, despite being an increase or decrease, has been shown to be correlated with a better DFS as compared to patient with unchanged TIL levels in TNBC (111). The characterization of different TIL subtypes is another variable playing a role in tumor response. IHC of different TIL

subtypes in RD found that high CD4-TILs, high CD8-TILs and low CD20-TILs in patients with better prognosis (112).

As many differences were observed regarding the quantity, density, and type of TILs, observing the spatial localization of where TILs were allocated in the samples became important. A large study by Loi, S and colleagues interrogated 2,148 early TNBC RD patients. In this study, stromal TILs (sTILs) demonstrated a strong prognostic role in early-stage TNBC. In addition, the authors constructed a prognostic model that incorporated the evaluation of TILs and clinicopathologic factors (104). Other immune cells have also been investigated in the context of RD in TNBC. Neutrophils are key cellular players in the inflammatory environment and their relationship with TILs is known to have different impact in BC patients' prognosis. Muñoz Montaño et al. assessed the neutrophil-to-lymphocyte ratio (NLR) in RD of BC patients and found that high NLR predicted worse OS in TNBC patients suggesting the immune influence in outcomes (113). Moreover, Gruosso et al, demonstrated a comprehensive understanding of the TIME in respect to the spatial localization of immune cells in the tumor. A immunoreactive microenvironment was defined by tumoral infiltration of CD8+T cells. Also, a "immune-cold" microenvironment, defined by absence of tumoral CD8+ T cells was characterized by high expression of immunosuppressive markers, stromal fibrotic signatures, and poor outcomes in patients. However, a distinct poor-outcome was identified in TIME were CD8+ T cells are found only in the stroma, having specific signatures of cholesterol biosynthesis (114).

Table 1: Summary of studies assessing immune biomarkers for TNBC residual disease.

Study	Biomarker	Number of patients per arm (n)	Findings and prognostic value
Watanabe et al. (2017)	TILs score	47 (pre-treatment), and 28 (post-treatment)	A high TIL score before treatment resulted in a significantly higher proportion of pCR in TNBC.
Zhang et al. (2018)	Tumor-infiltrating lymphocytes volume (TILV)	58	stromal TILs and TILV showed significant correlation with pCR (P = .01 and P = .0008, respectively). pCR, TILs, and TILV were all significantly correlated with the overall survival, with P = .028, .029, and .015, respectively.
Hamy et al. (2019)	TILs levels	716 (pre- and post-treatment matched paired specimens of BC)	Pre-NAC TIL levels were higher in tumors for which pCR was achieved than in cases with RD (33.9% vs. 20.3%, P = 0.001). High post-NAC TIL levels were not associated with aggressive tumor characteristics and with impaired DFS in TNBCs.
Luen et al. (2019)	TILs	375 (RD TNBC samples)	Higher RD TILs were associated with improved RFS (HR: 0.86; 95% CI 0.79-0.92; P < 0.001), and improved OS (HR: 0.87; 95% CI 0.80-0.94; P < 0.001) and remained significant predictors in multivariate analysis (RFS P = 0.032; OS P = 0.038 for OS). RD TILs added significant prognostic value to multivariate models including RCB class (P < 0.001 for RFS; P = 0.021 for OS).
Ochi et al. (2019)	pre-NAT TILs and pCR	80	Low pre-NAT TILs were associated with lower pCR rate (4.0% vs 43.6%). low pre-NAT TILs showed significant association with shorter RFS (HR = 3.844 [1.190-12.421], p = 0.024) in TNBC with RD. Low post-NAT TILs showed borderline significant association with shorter RFS (HR = 2.836 [0.951-8.457], p = 0.061).
Loi et al. (2019)	Stromally located TILs (sTILs)	2148	sTILs were significantly lower with older age (P = .001), larger tumor size (P = .01), more nodal involvement (P = .02), and lower histologic grade (P = .001). sTILs added significant independent prognostic information for all end points (likelihood ratio $\chi 2$,

			48.9 iDFS; P < .001; χ2, 55.8 D-DFS; P < .001; χ2, 48.5 OS; P < .001). Every 10% sTILs increase corresponds to an iDFS HR of 0.87 (95% CI, 0.83 to 0.91) for iDFS, 0.83 (95% CI, 0.79 to 0.88) for D-DFS, and 0.84 (95% CI, 0.79 to 0.89) for OS.
Muñoz-Montaño et al. (2020)	Neutrophil-to-lymphocyte ratio (NLR)	N=1519 (BC)	patients with high NLR had worse OS in the presence of TNBC (105.9 months; 95% CI, 100.2-111.5] vs. 98.7 months; 95% CI, 91.1-106.3; P = .029).
Bai et al. (2020)	TIL assessment algorithm		Using an optimal cut point (30%) derived from TNBC cohort training set A, patients with high eTILs% displayed an overall survival benefit (HR 0.4, p=0.0150).
Lee et al. (2020)	TIL level changes	n = 104	Changes in TIL levels (calculated by subtracting the TIL level of pre-NAC biopsy specimens from the TIL level of post-NAC operation specimens) associated with better DFS (increased TIL level (27.9%): HR 0.359, p= 0.014; decreased TIL level (41.3%): HR 0.439, p= 0.014).
Bai et al.(2020)	CD4-, CD8-, CD20- TIL number, CD4/CD20 ratio, CD8/CD20	n = 37	High CD4-TILs (DFS: P = 0.005, OS: P = 0.021), high CD8-TILs (DFS: P = 0.018) and low CD20-TILs (OS: P = 0.042) associated with better prognosis. CD4/CD20 ratio greater than 1 (DFS: P = 0.001, OS: P = 0.002) or CD8/CD20 ratio greater than 1 (DFS: P = 0.009, OS: P = 0.022) had a better prognosis.

DFS, disease-free survival; HR, hazard ratio; TIL, tumor-infiltrating lymphocytes; OS, overall survival; BC, breast cancer; HR, hazard ratio; PD-L1, programmed death ligand 1; iDFS, invasive disease-free survival.

1.3.2.2.2 Genetic and proteomic RD markers

The genetic and molecular profiles of TNBC with RD after NAC has been challenging researchers all over the world with its complexity and diversity. To date, various high-throughput genomic method analyses have been developed, providing solid grounds for new biomarker discovery. Known genetic and proteomic markers that influence prediction of response to chemotherapy and prognosis are summarized **(Table 2).** TP53 and PI3K mutations are frequently found in TNBC. The protein level of p53 assessed by IHC in TNBC RD, is significantly associated with pCR (115). Also, somatic co-mutations in TP53 and PIK3CA assessed in 353 post-NAC residual tumor samples by Sanger sequencing, were associated with more poor survival compared with non-carriers (116). A comprehensive molecular study from MDACC demonstrated that patients with BRCA genomic alterations are associated with higher mutational burden and enhanced chemosensitivity in TNBC. These results were validated using a TCGA TNBC cohort, observing that BRCA mutant patients had significantly better survival outcomes (117). Similar validation using the METABRIC TNBC cohort, identified that BRCA-deficient (BRCA-D) breast cancer subtype with BRCA1/2 low expression, had a higher mutation burden and better survival (117). To further examine functional RNA-based BRCA deficiency, Afghahi et al. performed BRCA1/2 sequencing of residual disease in known BRCA1/2-mutant BC patients with poor response to NAC. A BRCA1 reversion mutation in a TNBC patient was reported in this study which was correlated with poor response to therapy and early relapse (118).

Large cohort studies have allowed the discovery of different genes and proteins in TNBC. The results from a study of 1,079 TNBC cases demonstrated that helicase antigen

(HAGE) protein and mRNA expression was associated with higher presence of TILs and pCR (119). Another study performing gene expression profiling of TNBC RD, identified 449 genes associated with a more aggressive phenotype. The results identified CCL5, DDIT4 and POLR1C as independent prognostic factors for distant recurrence-free survival (DRFS). As a result, a three-gene signature of CCL5, DDIT4 and POLR1C was developed to describe chemotherapy resistance in TNBC (120). Several biomarkers of TNBC RD have been identified by several groups, such as Phosphorylated Hippo Pathway Kinases (MST1/2 and LATS1/2) (121), Matrix Metalloproteinase 9 (MMP-9)(122), CD44v9 (123), CCND1 (124) and Folate Receptor Alpha (FRα) (125). Since RD TILs is associated with better prognosis in TNBC, TIL-related genes have been interrogated. Criscitiello and colleagues developed a four-gene signature (HLF, CXCL13, SULT1E1, and GBP1) that predicted better DRFS in high post-NAC TILs TNBC samples. However, the study lacked external dataset validation(126). In another similar study, Kochi et al. identified 22 overexpressed genes in TNBC cases with high TILs scores and established a TILs-associated genomic signature (TILs-GS) (127). Lastly, a study interrogating pre and post NAC TNBC demonstrated that HAGE and programmed death ligand 1 (PD-L1) are associated to higher TILs, whereas Ras–MAPK signaling activation were commonly found in RD with lower TILs (128). Studying RD in TNBC comes with two clinical challenges. First, RD or post-NAT tissue is a small and precious clinical material. Second, TNBC are only around 15% of all BC, so multi-centered studies are needed to evaluate larger datasets. The future use of biomarkers alone or in combination requires collaboration and a mechanistic understanding to be use as part of the clinical validation of RD in TNBC.

Table 2: Summary of research studies assessing genomic and proteomic biomarkers of TNBC residual disease.

Study	Biomarker	Number of patients (n), treatment	Main findings and prognostic value
Loi et al. (2016)	Ras–MAPK pathway; cell-cycle pathway	39, pre-NAC and 92, post-NAC (total= 39 paired samples)	Low TILs in RD is associated with activating alterations in the Ras–MAPK pathway (amplifications in KRAS, BRAF, RAF1, and truncations in NF1, 16% altered, P = 0.005) and activating cell-cycle pathway alterations (CCND1-3, CDK4, CDK6, CCNE1, RB, AURKA, and CDKN2A, 37% altered, P = 0.05). Prognostic value: NA
Abdel-Fatah et al. (2016)	HAGE expression	1079, post-NAC (anthracycline)	High HAGE protein expression HAGE (+) compared with HAGE (-) cases: EP-TNBC, increased death risk [HR, 1.3; 95% confidence interval (CI), 1.2-1.5; P = 0.000005]; received ACT with mRNA HAGE (+): a lower risk of death (P = 0.004); The expression of HAGE was linked to the presence of TILs and both features were found to be independent predictors of pCR (P < 0.001) and prolonged survival (P < 0.01)
Kim et al. (2015)	p53, Ki-67, and Bcl-2	198, post-NAC (taxane-based)	Expression of p53 was independently associated with pCR to NAC (odds ratio, 3.961; p=0.003); The pCR rate was 5.2% in patients with low expression of both p53 and Ki-67, and the highest 25.8% when both biomarkers were highly expressed.
Jiang et al. (2016)	AR/FOXA1 pathway mutations, BRCA-deficient (BRCA-D)	29, post-NAC (anthracycline/ taxane)	Mutations in AR/ FOXA1-regulated networks associated with higher sensitivity to ACT CT (pCR rate of 94.1% compared to 16.6% in tumors without mutations in AR/FOXA1 pathways); and significantly better survival outcome (log-rank test, p = 0.05). Patients with functionally BRCA-D tumors had significantly better survival than patients whose tumors were not BRCA-D (log-rank test, p = 0.021), and they had significantly higher mutation burden (p < 0.001), presenting clonal neoantigens associated with increased immune cell activation.
Pinto et al. (2016)	Three-Genes Expression (CCL5,	82, post-NAT; 113, TNBC validation cases.	The median score of the three-genes signature (-0.393×CCL5+0.443×DDIT4+0.490×POLR1C) identified patients with distinct DRFS in the discovery set (0.1494) (P<0.001) and in the validation set (0.0024) (P=0.002).

	DDIT4 and POLR1C)		
Chen et al. (2017)	PD-L1 Expression **,** TILs	309, pre-NAC	The expression of PD-L1 was more commonly observed in patients with low levels of total TILs ($p < 0.001$), high levels of FOXP3+ TILs ($p < 0.001$) and low levels of CD8+ TILs ($p < 0.001$). prognostic value: The expression of PD-L1 was an independent prognostic factor for both RFS (HR = 1.824, $p = 0.013$) and OS (HR = 2.585, $p = 0.001$). High expression of PD-L1 was correlated with worse OS in TNBC. Patients classified as PD-L1-high/CD8-low exhibited relatively unfavorable survival, whereas patients with either low expression of PD-L1 or high expression of CD8 had similar outcomes.
Kochi et al. (2017)	TILs-GS (22 overexpressed genes in cases with high TILs scores)	40, post-NAC (anthracycline and taxane-based)	TILs-GS had no prognostic value in TNBC. TILs-GS levels are different between the cases with pCR and RD after anthracycline and taxane-based neoadjuvant chemotherapy. In the multivariate analysis, pCR was independently associated with smaller tumor size, higher histological grade, ER negativity, HER2 positivity and higher TILs-GS scores (OR 2.02, 95% CI 1.30-3.14, P = 0.025).
Ercolani et al.(2017)	MST1/2 and LATS1/2)	57 HER2-positive and TNBC post-NAT.	Patients with $pMST1/2^{nuc}$ (Nuclear pMST1/2) expressing tumors had increased risk of RD after NAT (pCR ypT0/is ypN0: OR 4.91, 95%CI: 1.57-15.30; pCR ypT0 ypN0: OR 3.59, 95%CI 1.14-11.34); $pMST1/2^{cyt}$ (cytoplasmic pMST1/2) seemed to be a positive prognostic factor (pCR ypT0/is ypN0: OR 0.34, 95%CI: 0.11-1.00; pCR ypT0 ypN0: OR 0.31, 95%CI 0.10-0.93).
Afghahi et al.(2017)	BRCA1 Reversion Mutation	80, post-NAC (platinum based).	From 19 patients with deleterious germline BRCA1/2 mutation, four had RD after surgery. BRCA1/2 sequencing of residual tissue was performed on three patients. These patients had BRCA1 1479delAG, 3374insGA, and W1712X mutations, respectively, with LOH at these loci in the pre-NAC tumors. In first case, a new BRCA1 mutation was detected in RD.
Wang et al. (2018)	MMP-9	303, post-NAC (paclitaxel plus carboplatin)	The relative change in sMMP-9, rather than sMMP-9 at baseline or surgery, had a remarkable predictive value for pCR. Each 1 ng/ml in serum decrease in sMMP-9 after NAC was shown to result in a 0.3% increase in pCR rate. hMMP-9 expression in residual tumors was independently correlated with DFS for non-pCR responders ($P < 0.001$).

Cheung et al. (2018)	Folate receptor alpha (FRα)	2012, TCGA and METABRIC. 305 untreated and 18 post-NAC IHC.	FRα is overexpressed in significant proportions of aggressive basal like/TNBC tumors, and in post neoadjuvant chemotherapy-residual disease associated with a high risk of relapse. FRα expression is associated with worse OS. TNBCs displayed dysregulated expression of thymidylate synthase, folate hydrolase 1, and methylenetetrahydrofolate reductase. folate metabolism.
Tokunaga et al.(2018)	CD44v9	48, pre-NAC.	There were no significant relationships between the pCR rate and the expression of CD44v9, vimentin, or BRCA1. High grade in the residual tumor cells, poor pathological response and high CD44v9 expression in the pre-treatment CNB samples were significantly correlated with a poor DMFS (p = 0.0433, 0.0406 and p = 0.0333). High grade in the residual tumor cells was significantly associated with high CD44v9 expression in the pre-treatment CNB (p = 0.0389).
Criscitiello et al. (2018)	A 4 gene signature (HLF, CXCL13, SULT1E1, and GBP1)	99, pre-NAC	The four-gene signature was significantly associated with DRFS (HR: 0.17, 95% CI: 0.06-0.43). GS added significant prognostic information when compared with the clinicopathologic pre-treatment model (likelihood ratio test in the training set P = 0.004 and in the validation set P = 0.002).
Orozco et al.(2019)	GE profiles	708, pre and post NAC	49 genes consistently affected by NAC were involved in enhanced regulation of wound response, chemokine release, cell division, and decreased programmed cell death in residual invasive disease. The statistical distances between pre and post- NAC significantly predicted pCR [AUC = 0.75; p = 0.003; 95% (CI) 0.58-0.92]. The expression of CCND1 was the most informative feature in pre-NAC biopsies to predict response to NAC.
Chen et al. (2019)	TP53 and PIK3CA mutations	353, post-NAC	Patients with somatic co-mutation were more likely to have high-grade tumors in TNBC (35.3% vs. 13.3%, P = 0.025) compared with non-carriers. More importantly, co-mutation of TP53 and PIK3CA carriers had a significantly worse DFS and DDFS than non-carriers (5-year DFS: 58.0% vs. 83.2%, P < 0.001; 5-year DDFS: 70.3% vs. 86.4%, P = 0.024).

EP-TNBC: Early primary TNBC; HRD: Homologous Recombination Deficiency; LOH: Loss of heterozygosity; CNB: core needle biopsy TAI: telomeric allelic imbalance; DRFS: Distant recurrence-free survival; GS, genomic signature; DDFS: Distant disease-free survival.

1.3.3 Resistance to therapy

Resistance to therapy is the ultimate challenge that clinicians must face in order to prevent disease progression and more importantly metastatic disease. Therapy resistance can be acquired at different stages throughout disease progression. Cells can have primary or intrinsic resistance (IRes) when they do not respond to their initial line of treatment and were previously naïve to treatment. Cancer cells can also display acquired resistance, presenting initially sensitive to treatment but later present no response to it (ARes) to it (129). This can be in one or multiple rounds of treatment, or one or many treatment regimens. The mechanisms of how a cancer cell can acquire resistance relies on previous concepts and cancer models. Cancer cells can be resistant to chemotherapy with the presence of genomic alterations (mutations, amplifications, overexpression) or escaping the drug mechanism of action. Clonal evolution would suggest that certain tumor clones carrying mutations that confer them fitness, will survive therapy and comprise the recurrent resistant disease (130). The cancer stem cell theory supports that CSCs have particular properties and can display quiescent states that will make them resistant to chemotherapy, which targets mostly cycling cells (131). A combination of these proposed models plus epigenetic modifications, metabolic adaptations, and cell-to cell interactions will also induce selective pressures for those cells resistant to therapy. Resistance to therapy can be studied with different approaches. Multiregional single-cell sequencing of tumors, pre and post treatment samples analysis, and the use of liquid biopsies (132). Clinical trials and the development of preclinical models that mimic resistant are extremely necessary to understand the mechanisms and adaptations of cells that are resistant to therapy.

1.3.4 Metastatic disease

1.3.4.1 Metastatic cascade

Metastasis is the end and most devastating stage of tumor progression, being the main cause of death in cancer patients. Metastases arise from an adaptative journey of cancer cells into a completely new niche. Tumor cells first need to escape the primary site, for that they need to invade the surrounding microenvironment, increase their motility, degrade the extracellular matrix, and reach blood vessels to circulate throughout the body. Once cells reach the new organ they extravasate, seed the organ, proliferate and develop neovascularization at the new site to acquire nutrients and establish a new tumor (133,134). **(Figure 1.3.4)**.

1.3.4.2 Single cell vs collective cell dissemination

Tumoral cells in the primary site acquire specific properties that allow them to intravasate into the circulation. One of the programs that metastatic are thought to require to efficiently disseminate is epithelial-to-mesenchymal transition (EMT) (135). EMT is a process whereby cells acquire mesenchymal properties and lose epithelial characteristics. This phenomenon is often a reversible and dynamic process that can be displayed by cells at different levels and provide them with different degrees of plasticity. Cells can either disseminate individually or in groups or collection of cells through the blood stream. Single circulating tumor cells (CTCs) can migrate through the matrix forming adhesions complex, cellular polarity and acquiring a mesenchymal-like phenotype(136). Meanwhile, cohesive circulating tumor clusters seemed to be the preferential metastatic dissemination mechanisms. Studies have provided insights that

collective cell migration (CCM) is a more efficient way to travel through vessels and ultimately colonize new organs. Multi-cellular clusters can travel while preserving cellular junctions and adhesions with other cancer cells (137). CTCs and CCM while in circulation they are surrounded and shielded by platelets and other immune cells. This association with platelets also helps in explaining the pro-thrombotic state in cancer patients. Moreover, the seeding of new organs through micro-vessels or capillaries is facilitated by the microthrombi phenotype cancer cells display when coupled with platelets and the cellular receptors present in endothelial cells can recognize platelets welcoming them to the new environment. The seeding and invasion of new organs can be facilitated through collective migration and survival in suspension since there are more cells able to survive and adapt to the new metastatic niche (135).

1.3.4.3 Metastasis organotropism

In addition to the molecular adaptions acquired for the metastatic process, cancer cells display predilection or organotropism for specific anatomic sites. During the late 1800 James Paget investigated the metastatic patterns of 735 breast cancer patients' autopsies. He discovered that breast cancer cells had a predilection to metastasize to the bones, and he characterized these as the theory of "seed and soil". Cancer cells (seeds) would be successful to grow in a particular anatomic site (soil) if the proper conditions for growth are present(138). This theory was challenged 40 years later by James Ewing, who suggested that metastatic spread was driven solely by the direction of the blood flow(139). In 1970s, Fidler confirmed that blood flow was important for metastatic spread, but he also demonstrated in vivo that the metastatic niche is key to determine which cells

will successfully grow in specific organs, revalidating Paget's theory (140,141). In BC the decreasing incidence order of the most common metastatic sites is bone, lung, liver and brain (142). The metastatic patterns and organotropism vary within BC subtypes. For instance, In Luminal A, Luminal B, HER2+/HR- and HER2+/HR+ BCs the most common metastatic site is the bone, whereas in TNBC is the lung. Generally, HER2+ patients display a higher incidence of liver metastasis and brain metastasis after some years of dormancy (143,144). Also they are more frequently observed concomitantly with other organ metastases, rather than as an initial metastatic site (145). Moreover, cells that metastasize to distinct organs have shown to have different gene expression and metabolic profiles. In TNBC, genes involved in lung metastasis have been identified. In vitro and in vivo studies have demonstrated that lung metastatic cells were enriched for ANGPL4, EREG, CXCL1, MMP1, MMP2, ID1, SPARC, among others (146,147). These genes are involved in angiogenesis and matrix degradation. Moreover, primary, and matched lung metastatic samples from breast cancer show enrichment of hypoxic genes such as CA9, EGLN3, DNAH11, and LOX. Cells exposed to hypoxia in the primary tumor acquire "hypoxic-memory" and genes that confer lung metastatic cells resistance to oxidative phosphorylation (oxphos). This is a survival mechanism for cancer cell dissemination through the blood stream (148). The metabolic plasticity of cells in the primary tumor, in circulation and in the new metastatic site, is key for cells to seed and grow in their environment. The metabolic programs and gene expression profiles of cells will depend on the organ (149). For instance, brain cells require adaptation to low glucose and lipid levels. This is supported by the high expression of Fatty Acid Synthase (FASN) in HER2+ BC cells, an enzyme required for brain fatty acid metabolism (150).

1.4 Tumor heterogeneity in cancer

1.4.1 Intra-tumor heterogeneity

Throughout the last two decades, researchers and clinicians have tried to established breast cancer subtypes with the objective to find the most suitable treatment that can help patients. Common efforts have facilitated the identification of a subset of patients that could benefit from targeted therapy or certain types of chemotherapy (151). However, even within the same breast cancer subtype or patients with similar clinicopathological characteristics, responses to treatment are different. The era of personalized treatment in breast cancer is emphasized by the large heterogeneity that this disease and each patient presents (152). There are many variables that make a patient different from one another. Some of these are medical comorbidities, body weight, geographic location, and the type of diet that patients have, affects both the incidence of cancer and their response to treatment.

1.4.1.1 Spatial and temporal heterogeneity

Intra-tumor heterogeneity and morphological differences of cancer cells can be readily observed in formalin-fixed paraffin embedded samples of human tumors (153). Most studies that established different breast cancer subtypes based on gene expression were performed on bulk tumors. The tumor bulk is an overall representation of the most common or frequent alterations identified in the whole tumor. The tumor is comprised by many different cancer and stromal cells, and not all genomic or non-genomic alterations will be present in all the cells found in the tumor. Moreover, intra-tumor heterogeneity is also influence by the tumor microenvironment and the distinct cell-to-cell interactions

found in a three-dimensional space. The local tumor queues, such as the vasculature and the degree of oxygen and nutrient availability, have also been shown to affect tumoral cells metabolic and transcriptional profiles (154,155).

1.5 Tools to interrogate cellular heterogeneity upon disease progression

The study of cellular heterogeneity upon disease progression has been possible by using high throughput technologies, primarily imaging and sequencing technologies. Moreover, the challenges of tissue availability are eased by clinical trials, allowing the use of clinical material for research purposes. Pre-clinical models coupled with single-cell RNA sequencing technologies and advanced microscopy techniques, are key for understanding both temporal and spatial heterogeneity (156–158).

1.5.1 Patient-derived xenografts

Patient-derived xenografts (PDX) are useful preclinical models to obtain renewal patient material to perform subsequent research studies. The establishment of PDXs requires multi-disciplinary coordination, especially from clinicians and researchers. Common efforts in establishing breast cancer PDXs have been performed over the past decade. Establishment of different breast cancer subtypes has also posed a challenge since the engraftment rates are correlated with aggressive disease. Aggressive phenotypes, advanced disease, HR- samples, have better engraftment rates (159). In contrast, ER+ BC tissue has lower rate of engraftment and normally requires estrogen pellets (160).

1.5.2 Single-cell RNA sequencing

In the past decade, sequencing technologies have revolutionized cancer research and opened infinite avenues to gain further mechanistic understanding of tumor biology. However, sequencing from the tumor bulk, does not allow comprehensive understanding of cellular heterogeneity and diversity within the tumor ecosystem. The readout of bulk RNA and DNA sequencing does not explain patient's unresponsiveness to targeted therapy. Genomic alterations that have available targets, can be present as a robust finding in bulk sequencing, however, not all cells might present this alteration and, therefore, they might explain why some tumors do not respond to treatment (161).

Single cell sequencing technologies has allowed the interrogation of cancer and tumor microenvironment cells in the tumor. With this technology the transcriptome and genome of epithelial, vascular, immune, and other cell types, can be interrogated within a same tumor using scRNA-seq. An advantage over other methods, is that rare cell populations in the tumor can be identified, discovered, or further interrogated using scRNA-seq. Moreover, cellular networks, trajectories and cell hierarchies can be interrogated (162,163).

1.5.3 Spatial profiling in situ

To overcome scRNA-seq's main limitation, different technologies have also been developed. Allowing the comprehensive use of what would correspond to spatial transcriptomics (ST). Moreover, the spatial localization of cells in the tumor microenvironment and their relationship with other structures can play an important role in the phenotype and sensitivity to drugs. For instance, cells localized near blood vessels, not only will have high availability of nutrients and oxygen, but also a higher proliferating capacity exposure drugs delivered intravenously. The use of multiplex immunofluorescence, despite the downside of having limited number of antibodies that can be used at once, is still a very powerful approach for biomarker discovery and validation (164). The use of this technology is normally a good approach before moving into higher-throughput technologies, such as digital spatial profiling (DSP) (157). The DSP technology is a newer alternative that offers the opportunity to interrogate different tumors at once, trying to make up for the lack of the one patient approach using scRNA-seq. This technology is based on the selection of region of interest (ROIs) in the tumor. Within these ROIs antibodies are stained by using multiplex immunofluorescence techniques. These allows to interrogate larger datasets in the same manner(165). Overall, this and other approaches are key and successful to answer clinical questions when developing a good experimental design.

1.6 Rationale

The phenotypic diversity of triple-negative breast cancer cells in different regions of a sample poses challenges in pathologic assessment of markers used in clinical decision-making (32,166). Therapy does not integrate the spatial context of tumor cells, their abundance, nor the networks to the local environment where they reside. These variables can influence transcriptomic, metabolic, and epigenetic variability among these cells and therefore, their response to treatment (167). The spatial heterogeneity and cell organization represented by the tumor architecture, plays a critical role in disease progression (15,168). Tumor cells adapt to stress through cell state transitions and plasticity in response to therapy and metastatic dissemination at the single cell level. This has been investigated by different approaches including scRNA-seq and/or imaging technologies, however, the biological differences and vulnerabilities of individual cell populations integrating these technologies is still poorly studied. The aim of this thesis work is to use preclinical models that mimic the patient's disease progression in a multi-pronged approach. With the use of single-cell sequencing, digital spatial profiling, and functional experimentation of single cell populations we expect to bring insights about how single cell populations are transcriptionally regulated by their spatial localization in the tumor. Moreover, we aim to understand the biology of aggressive cell populations involved in metastatic progression. Utilizing dynamic tracing of cell populations in matched primary and metastatic samples from a same patient, we aim to identify therapeutic targets that can prevent disease progression.

2. RESULTS

2.1 Breast cancer models to study clinical challenges

2.1.1 Biobanking breast cancer clinical samples.

Surgical specimens are collected transferred for diagnosis to pathology and tissue in excess at pathology is available for research. Patient derived xenografts for breast cancer, require a small piece less than $1mm^3$ of preferably fresh tissue (159). Normally, the tissue is engrafted either subcutaneously or in the case of breast cancer PDXs, it can be implanted into the fourth mammary fat pad of immunocompromised mice as well as through the nipple directly into the duct, for ER positive breast cancers (169,170). The environment of the fat pad mimics the environment of the breast, avoiding as much possible phenotypic alterations. The initial engraftment of the patient's tumor into the first generation of mice is called passage (P)0, following collection and re-transplantation gives rise to P1 etc. Moreover, tumor dissociation and specific culture medias allows the generation of matched patient-derived organoids (PDXOs) or patient-derived cell lines (PDCs) allowing use for complementary studies in-vitro. The key step in use of PDXs as preclinical models is the need to validate that the PDX retains the clinical histology, genomic landscape and transcriptomic of the patient tumor. If the PDXs recapitulate the human tumor landscape, they serve well as preclinical models to perform further translational and basic science research studies.

2.1.2 Patient-derived xenografts to model breast cancer disease progression.

Patient-derived xenografts or PDXs are preclinical models that recapitulate the clinical landscape of the patient and provide a renewable source of the tissue to perform research. We have bio-banked and generated PDXs from patient's primary and matched metastatic disease samples. This provided a powerful tool to interrogate disease progression preserving the same patient's landscape and gather further understanding of the primary tumor biology and metastatic disease **(Fig 2.1.2A)**. To understand the sensitivity to standard of care treatment, we first identified a panel of drugs commonly used in TNBC patients and tested response in a 1x1x1 trial approach (171). Following an apparent complete response to a specific drug (change in tumor volume of ~100%), we obtained a residual disease PDX model. The residual disease tumor from PDX-Lm (GCRC2076) was developed in response to the chemotherapeutic agent, Gemcitabine **(Fig 2.1.2B)**. To further develop a model of recurrence, the drug was removed, and residual tumor cells allowed to expand then re-challenged with the same drug repeatedly "on and off" until acquired resistance was established and resistant tumor harvested **(Fig 2.1.2B)**. To further validate the acquired resistance, the tumor was transplanted into a new cohort of mice and treated with gemcitabine to demonstrate that resistance to therapy was preserved upon tumor passages **(Fig 2.1.2C)**.

Fig.2.1.2

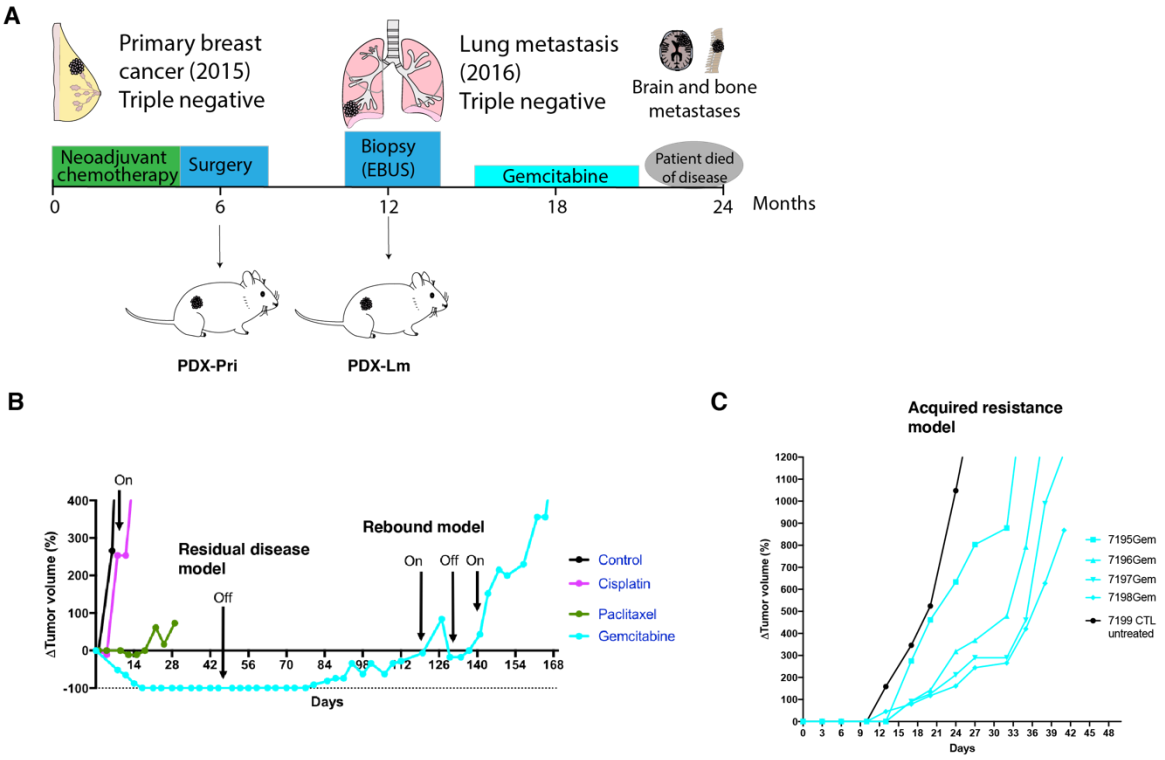


Fig.2.1.2. PDX models to mimic breast cancer disease progression.

A. The patient's primary tumor (GCRC1915) after surgical resection was engrafted into the mammary fat pad of an immunocompromised mouse to establish a first PDX. Later, the same patient once developed lung metastatic disease (GCRC2076), the biopsy of the lung tumor was engrafted into other mice to generate a matched PDX**. B.** Treatment with Gemcitabine in PDX GCRC2076 created complete response to the treatment. The residual disease was removed off the effect of the drug and a rebound was obtained. The rebound was re-challenged on and off with drug until no response was seen anymore. **C.** The resistant tumor was transplanted into 5 mice and treated with Gemcitabine, confirming no response to treatment.

2.1.3 Single-cell RNA sequencing of stepwise disease progression samples.

To study the key clinical and research questions of tumor heterogeneity and relapse, corresponding PDX models were generated. These models of disease progression serve to study the patient's biological and genomic landscape in a consistent and reproducible manner. Matched primary (GCRC1915), and metastatic disease samples (GCRC2076), provides a unique opportunity to interrogate cell populations selected for a more aggressive phenotype in a temporal manner. After developing these models and prior to analysis by single cell RNA sequencing, scRNA-seq, each tumor dissociation requires technology development. The process of tumor dissociation requires optimization of the digestion process, cell isolation, cell viability and bioinformatic analyses. Once, this was obtained, the primary tumor, matched lung metastasis, residual disease, rebound, and resistant tumor were subjected to scRNA-seq.

Figure.2.1.3

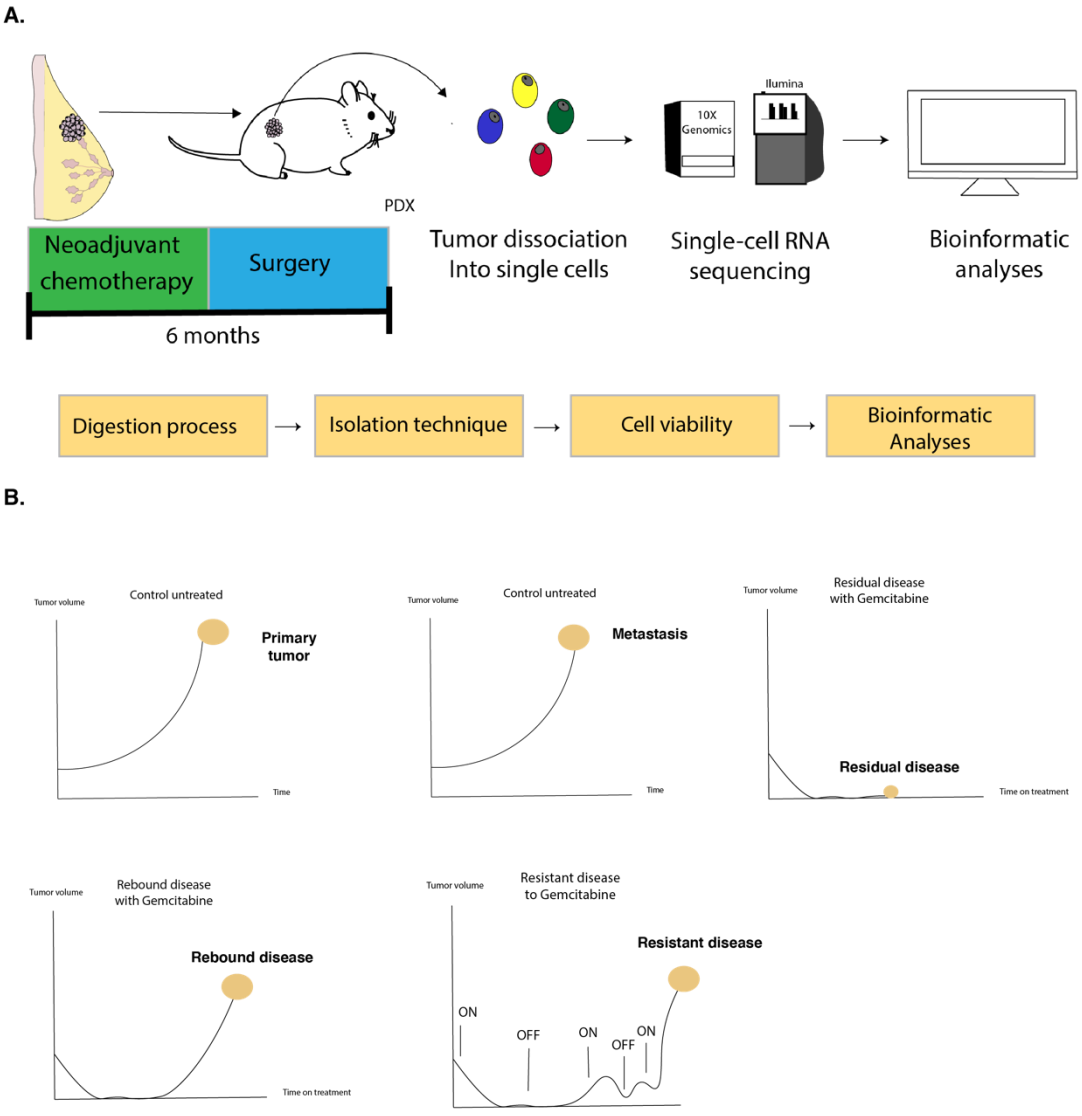


Fig.2.1.3. Technical aspects of tumor dissociation and samples for scRNA-seq.

A. Schematic approach of technology development required to properly subject samples for scRNA-seq. The digestion process, isolation technique of cells, cell viability and bioinformatic analyses are essential steps to perform a good experiment. **B.** Disease progression samples subjected to scRNA-seq. Each of them requires individual optimization of all the aspects mentioned in A.

2.2 Spatial and temporal heterogeneity in breast cancer metastasis

2.2.1 Patient-derived xenografts as models to study temporal heterogeneity.

To understand the molecular heterogeneity induced by changes in the tumoral environment in TNBC, we established a cohort of matched TNBC samples and patient derived xenografts (PDX). These samples include a patient primary triple-negative breast cancer tumor obtained from the patient (PT). A matched PDX from the primary tumor (PDX-Pri). Later a PDX derived from the patient's lung metastasis (PDX-Lm). The metastatic disease in the patient was diagnosed six months after the partial mastectomy was performed. Finally, a spontaneous lung metastasis PDX model was derived (PDX-Slm). This was performed by engrafting the PDX-pri tumor in the mammary fat pad (MFP), resecting it when reaching endpoint, and monitor for mice to develop symptomatic disease, which occurred six-months after the tumorectomy from the PDX-pri MFP, mimicking the temporal heterogeneity observed in the patient **(Fig 2.2.1A-C).**

Fig.2.2.1

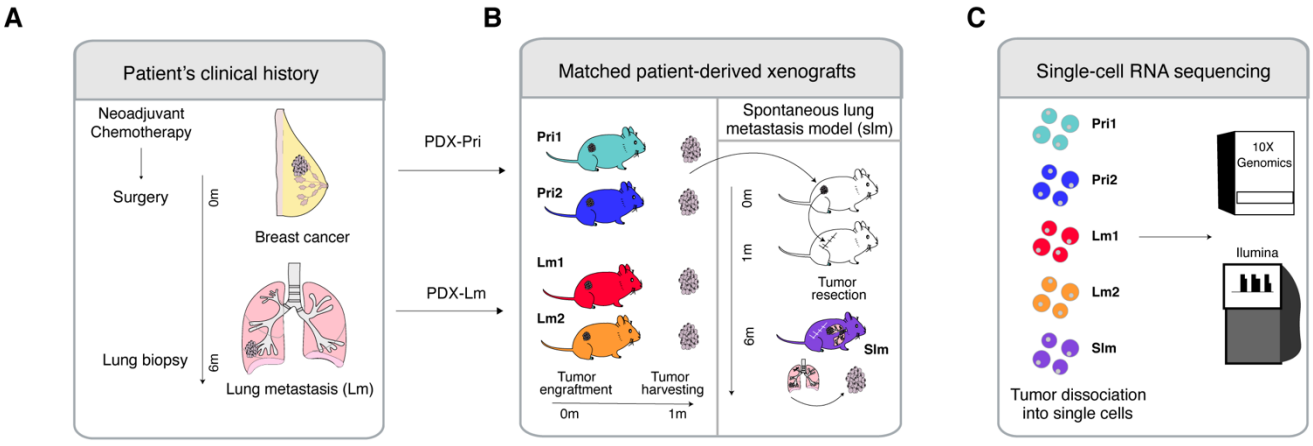


Figure.2.2.1. Schematic overview of primary and metastatic breast triple negative PDXs experimental pipeline.

A. Overview of experimental design. Primary breast tumor was used to generate PDX-pri samples. A six-month lung metastasis from the same patient was utilized to generate PDX-Lm. **B.** From PDX-pri two replicates were generated (Pri1 and Pri2), and from the PDX-Lm two metastatic replicates were generate (Lm1 and Lm2). A third PDX from a PDX-pri spontaneous lung metastasis was derived six months post-resection of tumor in fat pad (Slm). **C**. Five tumors from 2 PDX-Pri, 2 PDX-Lm and 1 PDX-Slm were dissociated into single cells and subjected for single-cell RNA sequencing using 10x genomics.

2.2.2 Patient-derived xenografts are representative models of the patient's histological, transcriptomic, and genomic landscape.

To confirm that PDX models faithfully recapitulate patient tumor samples, we compared their histological, transcriptomic, and genomic landscapes. Both PT and PDX-Pri were negative by immunohistochemistry (IHC) for ER, PR and HER2, consistent with their clinical classification as TNBC. IHC staining for breast cancer clinical markers (PanCK, Ki67, p53) showed similar patterns in the patient and PDX samples **(Fig. 2.2.2A).** Transcriptional profiles derived from bulk RNA-seq were also similar, with PT and PDX-Pri clustering together as basal breast cancer by absolute intrinsic molecular subtyping (AIMS) (172) using the PAM50 gene set (66,159) **(Fig. 2.2.2B)** . Both PT and PDX whole-genome sequencing revealed a missense mutation in TP53 (p.R273H), correlating with the increased p53 protein levels detected by IHC. Finally, copy number alteration (CNA) profiles pointed to similar genomic landscapes **(Fig. 2.2.2C)**. Taken together, these results indicate that PDX-pri recapitulates the clinical marker profile by IHC, as well as the genomic and transcriptomic landscape of the patient's primary tumor, therefore serving as a good pre-clinical model to study intra-tumor spatial and temporal heterogeneity at the single-cell level.

Figure. 2.2.2

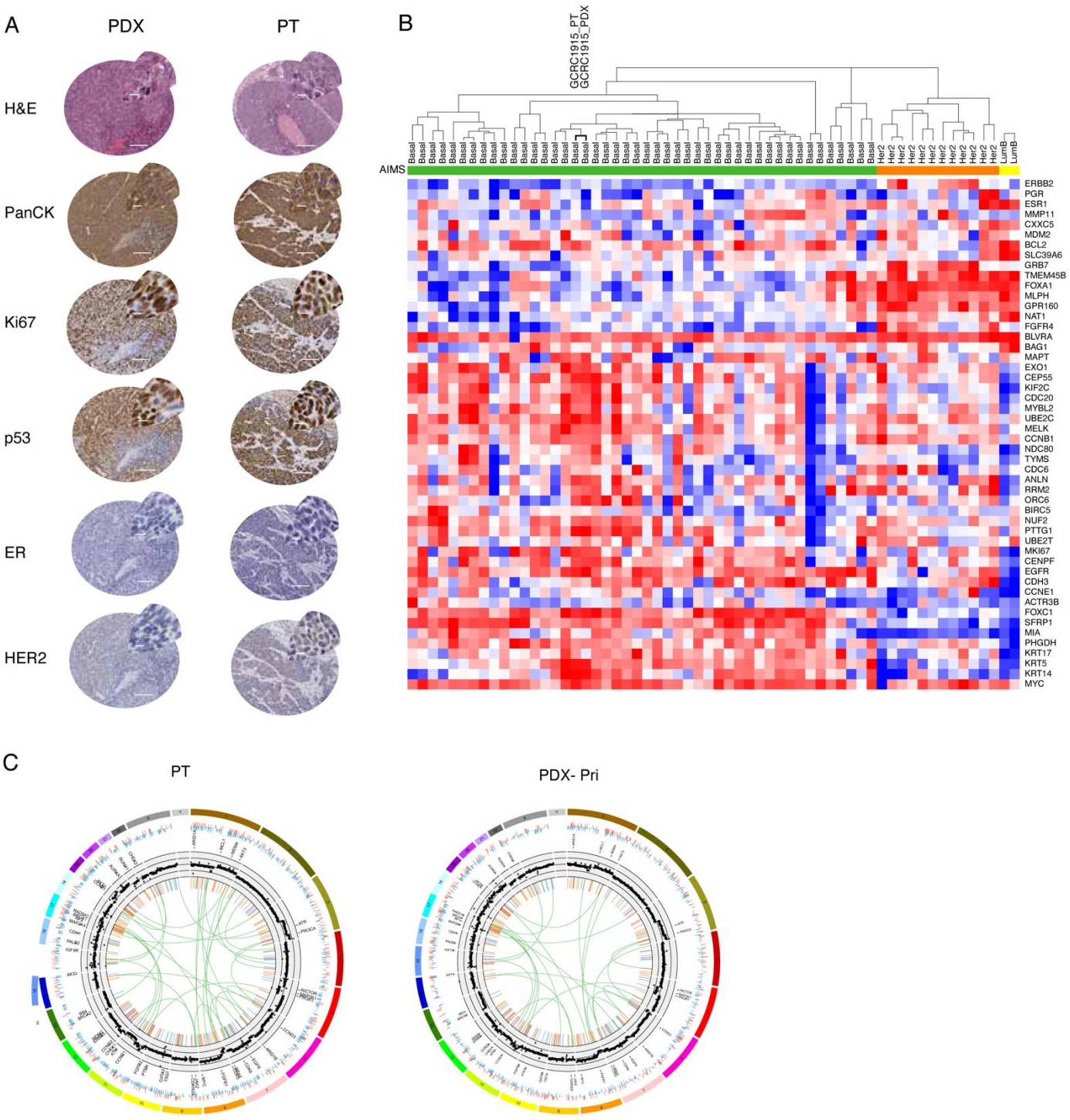

A
PDX
PT
H&E
PanCK
Ki67
p53
ER
HER2
B
GCRC1915_PT
GCRC1915_PDX
AIMS
Basal
Her2
LumB
ERBB2
PGR
ESR1
MMP11
CXXC5
MDM2
BCL2
SLC39A6
GRB7
TMEM45B
FOXA1
MLPH
GPR160
NAT1
FGFR4
BLVRA
BAG1
MAPT
EXO1
CEP55
KIF2C
CDC20
MYBL2
UBE2C
MELK
CCNB1
NDC80
TYMS
CDC6
ANLN
RRM2
ORC6
BIRC5
NUF2
PTTG1
UBE2T
MKI67
CENPF
EGFR
CDH3
CCNE1
ACTR3B
FOXC1
SFRP1
MIA
PHGDH
KRT17
KRT5
KRT14
MYC
C
PT
PDX- Pri

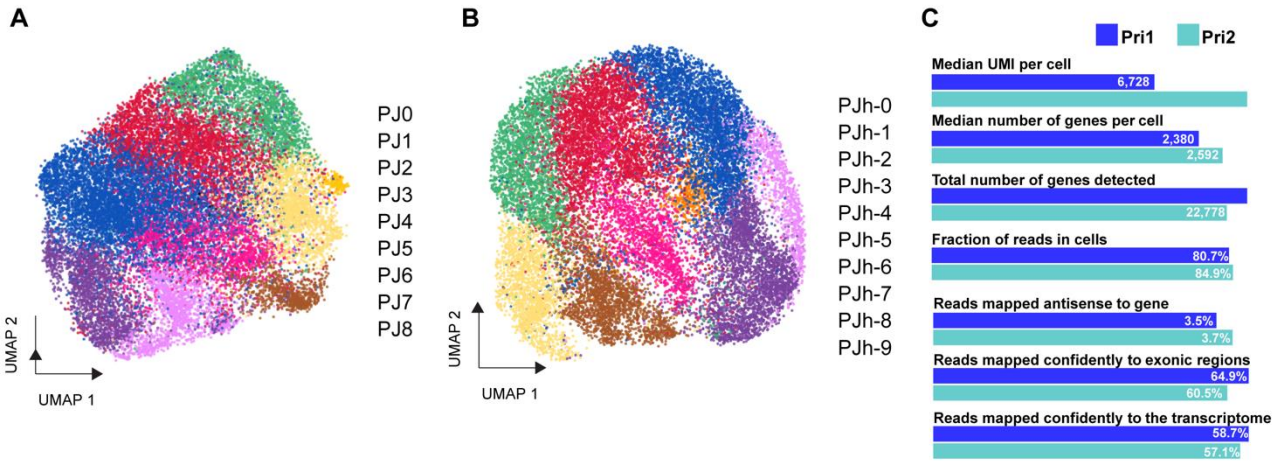
Figure. 2.2.2. Patient-derived xenograft mimics the molecular features of the patient's primary tumor.

A. Evaluation of the principal markers for the clinical classification of breast cancer by immunohistochemistry showed that both PDX and PT present the same pattern expression of ER, HER2, Ki67, tp53 and PanCK. **B.** PDX and PT were clinically classified as a triple-negative breast cancer, and PAM50 signature gene expression analysis revealed that PDX and PT cluster as basal breast cancer (Heatmap metric: One minus Person correlation; Linkage Method: Complete)(159). **C.** Circos plots showing somatic mutations and transcriptional profiles in the PDX and corresponding primary tumor. The outer ring shows the mRNA profile of every chromosome (blue and red lines represent genes that are down- or up-regulated, respectively. Only genes that have |log2FC| > 8 are shown). The middle ring represents copy number variations, derived from whole-genome sequencing data. The green lines traversing the inner ring indicate inter- and intra-chromosomal rearrangements and structural changes.

2.2.3 Single-cell RNA sequencing of primary breast cancer patient-derived xenograft replicates.

To define the cellular heterogeneity and obtain robust biomarkers of cell populations in the primary breast tumor, we subjected two PDX-pri replicates (Pri1 and Pri2) to dissociation, murine stromal cell depletion, and single-cell RNA sequencing using the 10X Genomics technology. After alignment, demultiplexing and filtering of murine cells, replicates were integrated using two alternative methods, Canonical Correlation Analysis and Harmony (173). Replicates displayed a high degree of reproducibility, displaying same number of clusters with similar structure within the uniform manifold approximation and projection (UMAP) (**Figure 2.2.3A, B**). Moreover, with both methods QC metrics and cell population structure is well represented within both PDX pri1 and pri2 samples. Number of genes, mitochondrial content and ribosomal content was similar in all samples, slightly lower in Lm2 **(Figure 2.2.3C-E)**

Fig 2.2.3



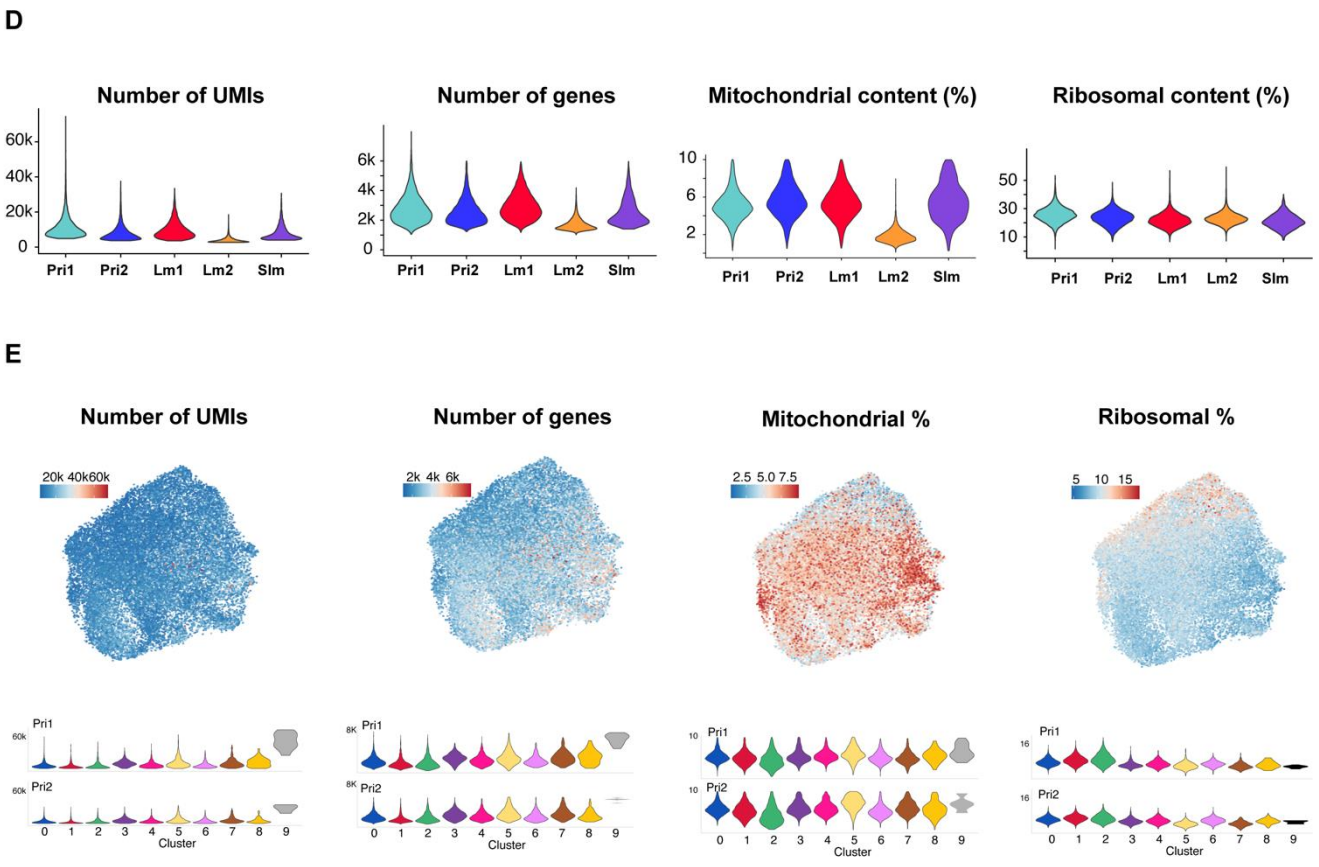


Figure. 2.2.3. Single-cell RNA sequencing of primary triple-negative metrics.

A. UMAP visualization of the primary joint (PJ) analysis, in which the primary PDX samples, Pri1 and Pri2, were integrated using CCA. Cells are colored by cluster identity. **B.** UMAP visualization of the primary join Harmony (PJh) integration of primary PDX samples, Pri1 and Pri2. Cells are colored by cluster identity. **C.** Comparison of library quality metrics for Pri1 and Pri2 samples following read alignment and transcript counting. UMI, unique molecular identifier. **D.** Distribution across cells per sample for four quality control metrics. **E.** UMAP plots of the CCA integration of Pri1 and Pri2, with cells colored by four quality control metrics. Below, the distribution of each metric per cluster, with cells separated by sample of origin.

2.2.4 Single-cell RNA sequencing of the breast primary tumor identifies eight distinct transcriptomic populations.

Clustering analysis in the integrated space revealed transcriptionally distinct cell populations (PJ0-PJ7) each containing cells from both samples, with the only exception of a small distinct cluster (PJ8, 174 cells), detected predominantly (80.5%) in one of the samples **(Fig. 2.2.4C)**. This indicates consistency in the phenotypic states found across replicates. Dimensionality reduction by Uniform Manifold Approximation and Projection (UMAP) positioned cells following a gradient that was only partially explained by cell cycle phases. Clusters PJ3 and PJ6 were enriched for cells in the G2M cell cycle phase, while cluster PJ0 was enriched for cells in the S phase **(Fig 2.2.4B)**. Since cell cycle status is often a major source of transcriptional variation in scRNA-seq datasets, we performed a regression analysis to determine to what extent the population structure was dictated by cell cycle programs. As expected, PJ3 was largely affected when regressing out cell cycle genes, showing the largest dispersion in the UMAP space and increasing its geometric median absolute deviation. All other clusters, on the other hand, were quite robust to cell cycle regression, indicating that other processes, independent of cell cycle, are the main source of variation in this cohort **(Fig 2.2.4C-E)**.

Fig. 2.2.4

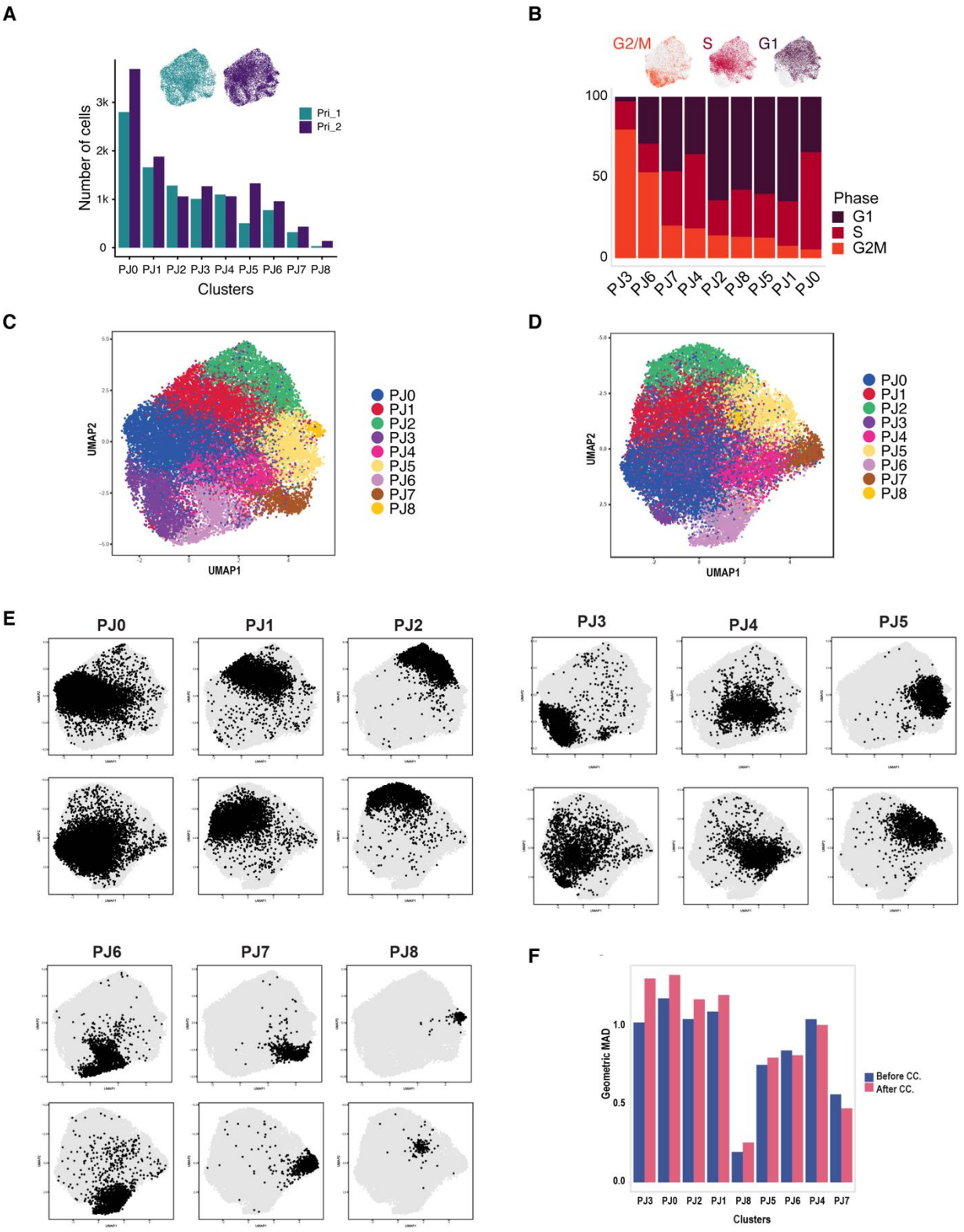

A
Number of cells
3k
2k
1k
0
Pri_1
Pri_2
PJ0 PJ1 PJ2 PJ3 PJ4 PJ5 PJ6 PJ7 PJ8
Clusters
B
G2/M
S
G1
100
50
0
Phase
G1
S
G2M
PJ3 PJ6 PJ7 PJ4 PJ2 PJ8 PJ5 PJ1 PJ0
C
UMAP2
UMAP1
PJ0
PJ1
PJ2
PJ3
PJ4
PJ5
PJ6
PJ7
PJ8
D
UMAP2
UMAP1
PJ0
PJ1
PJ2
PJ3
PJ4
PJ5
PJ6
PJ7
PJ8
E
PJ0
PJ1
PJ2
PJ3
PJ4
PJ5
PJ6
PJ7
PJ8
F
Geometric MAD
1.0
0.5
0.0
Before CC.
After CC.
PJ3 PJ0 PJ2 PJ1 PJ8 PJ5 PJ6 PJ4 PJ7
Clusters

Fig.2.2.4. Primary joint cell cycle phase representation and regression.

A. Pri_1 and Pri_2 cell number representation for each PJ cluster. **B.** Percentage of cells in each PJ cluster in different phases of the cell cycle; G1 (purple), S (orange) and G2M (purple). (A) UMAP plot of primary joint (PJ) before cell cycle (CC.) regression (upper) and after CC. regression (lower), cells are colored by the clusters identified before cell cycle regression. **C, D.** UMAP before and after cell cycle regression, respectively. **E.** Each PJ clusters vertically with UMAP space before CC. regression (top) and after CC. regression (bottom). PJ3 (cells under G2M phase) is highly dispersed after CC. regression, other PJ clusters relatively remain their major crowds. **F.** Quantification bar plots showing the highest displaying geometric median absolute deviation (MAD). Clusters are ordered by the difference of geometric MAD between the data after CC. regression and before CC. regression, descending.

2.2.5 The primary tumor replicates are comprised by 4 distinct biological cell populations and 4 transitional clusters.

To investigate the cellular classification of biological clusters we interrogated the SCENIC platform to identify active and specific transcription factors and elevated genes comprising these TF's regulons (174). We identified a cell cycle cluster (PJ3 – purple), with high and specific activity of Forkhead Box M1 (FOXM1) known to be overexpressed in TNBCs, and associated with the modulations of genes involves in the regulation of cell cycle and tumorigenesis (175–177). Moreover, cycling genes such as MKI67, CDK1, CDK2ND were also highly and specifically expressed. We also found a hypoxic clusters (PJ1/PJ2) with activity of HIF1A, which is induced under hypoxia and characterize aggressiveness in breast cancer (178,179). Moreover, HIF1A targets were specifically expressed in this cluster (CA9, VEGFA, BNIP3L, SLC2A1)(180). Additionally, we identified a lipid metabolism cluster (PJ5 – yellow), which displayed high activity of Peroxisome Proliferator-Activated Receptor Delta (PPARδ), a master regulator of lipid metabolism, EMT, angiogenesis and inflammation (181,182). Moreover, this cluster presented elevated expression of genes involved in cholesterol metabolism (DHCR24) and genes found in breast cancer cells with bipotential progenitor capacities (KRT6B). A interferon alpha/gamma cluster (PJ7 – brown), with high expression of ETV7 and viral mimicry response genes (IFI27, ISG15, OASL). Finally, we detected a small cluster with signatures elated to NOTCH and PI3KCA pathways (PJ8 – orange) **(Fig 2.2.5A, B)**.

Lastly, we predicted the relationship between the PDX-pri clusters by inferring a cell differentiation hierarchy using Monocle (183). Cells in the cycling cluster PJ3, hypoxic cluster PJ1/PJ2 and lipid metabolism in PJ5, were all found on opposing trajectories.

Notably, transitional clusters (PJ0, PJ4, PJ6, and PJ8) were found along the distinct trajectories, demonstrating the transitional state of the cells found in these clusters **(Fig.2.2.5D).** Together, these analyses support a high concordance in the transcriptional diversity observed between PDX replicates. Distinct biological and transitional populations were identified in both PDX-pri samples, making the use of a joint space a reliable tool to identify robust biomarkers in situ.

Fig.2.2.5

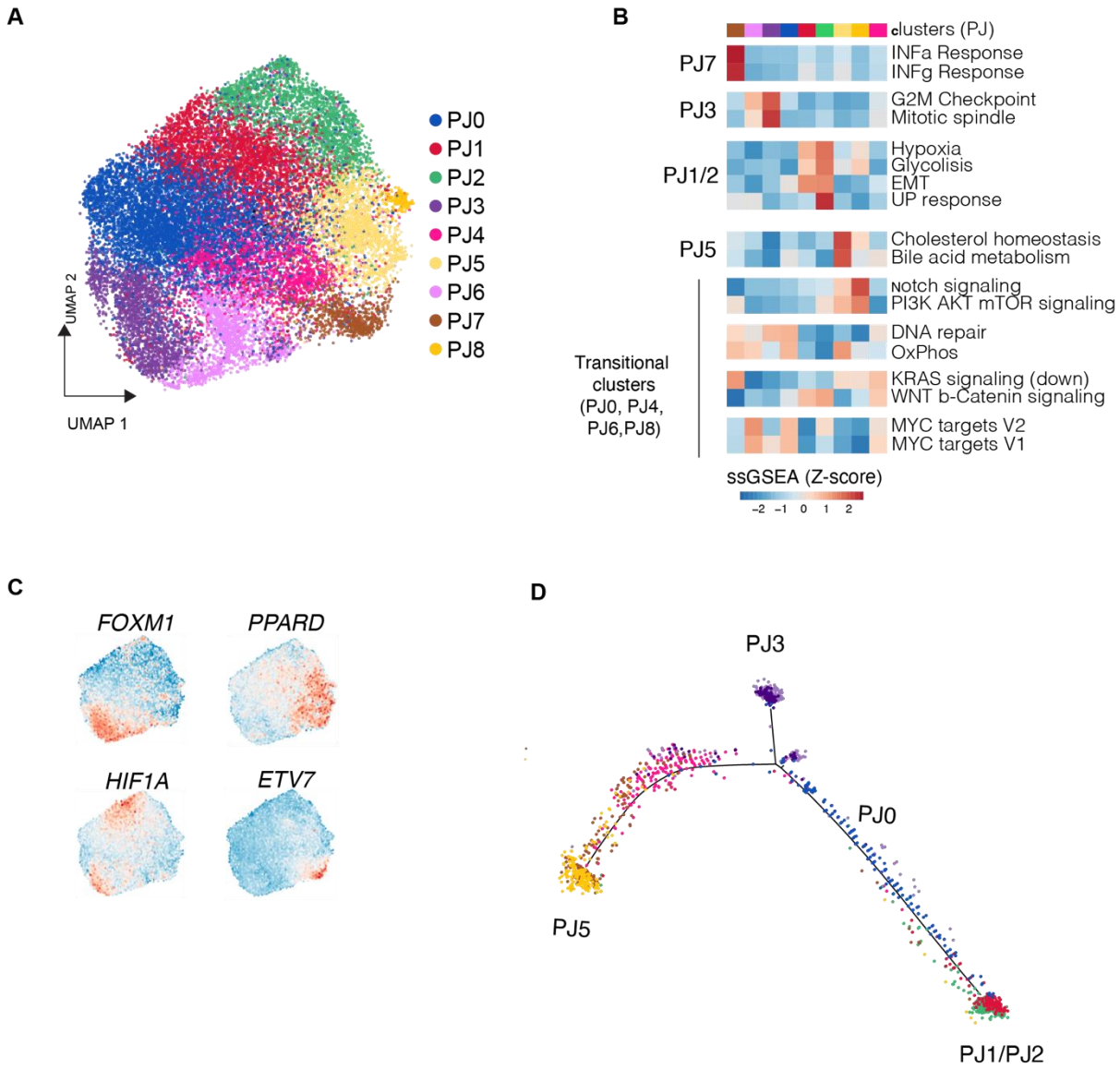


Fig.2.2.5. Single-cell RNA sequencing of TNBC PDXs.

A. UMAP visualization of primary joint (PJ) space representing two primary PDX tumors replicates (pri1 and pri_2). **B.** ssGSEA scores for MSigDB Hallmark pathways, computed per cluster and Z-scored across clusters. The top 2 pathways for each cluster based on scaled ssGSEA score were selected for visualization. **C.** UMAP plots of PJ analysis with cells colored by inferred transcription factor activity. Active and specific factors for biological subsets are shown. **D**. Reconstructed cell state trajectory with the ordered cells colored by PJ cluster identity.

2.2.6 Methodological selection of specific cluster markers from scRNA-seq data.

To interrogate the spatial topology of cell populations defined by scRNA-seq within the tumor in an unbiased manner, we integrated gene expression data from scRNA-seq and multiplex immunofluorescence (MIF). First, the most specific `genes to use as protein markers for each single-cell population of PDX-pri were selected from the scRNA-seq data. Differentially expressed genes between the single-cell clusters were computed with the Wilcoxon Rank Sum Test, and genes with high average fold changes, low p-values, and with available antibodies for immunofluorescence were selected as protein markers (see Methods). There were clusters that were characterized by the expression of either one or two genes, whereas other clusters did not have specific genes that could solely identify them. In fact, most of clusters with specific gene markers corresponded to the identified biological clusters, and those lacking markers were concordant with what we called transitional clusters. Within biological clusters, the cell cycle cluster PJ3 was characterized by the expression of topoisomerase-2a (TOP2A) and PCNA-associated factor (KIAA0101). The hypoxic cluster PJ1 displayed specific expression of Carbonic anhydrase 9 (CA9), while the hypoxic cluster PJ2 had expression of CA9 and Solute Carrier Family 2 Member 1 (gene: SLC2A; protein: Glucose transporter 1, GLUT1). Cluster PJ4 presented high and specific expression of Ubiquitin hydrolase L1 (UCHL1). Cluster PJ5 had specific expression of Crystallin-AB (CRYAB) and expression of Calmodulin Like 5 (CALML5), while PJ7 had specific shared expression of CALML5 with cluster PJ5. Within the transitional clusters, namely PJ0, PJ4, PJ6 and PJ8, they showed no gene markers with high fold changes relative to other clusters, and lowly expressed several genes selected as protein markers for other clusters **(Fig. 2.2.6 A, B**).

Fig. 2.2.6

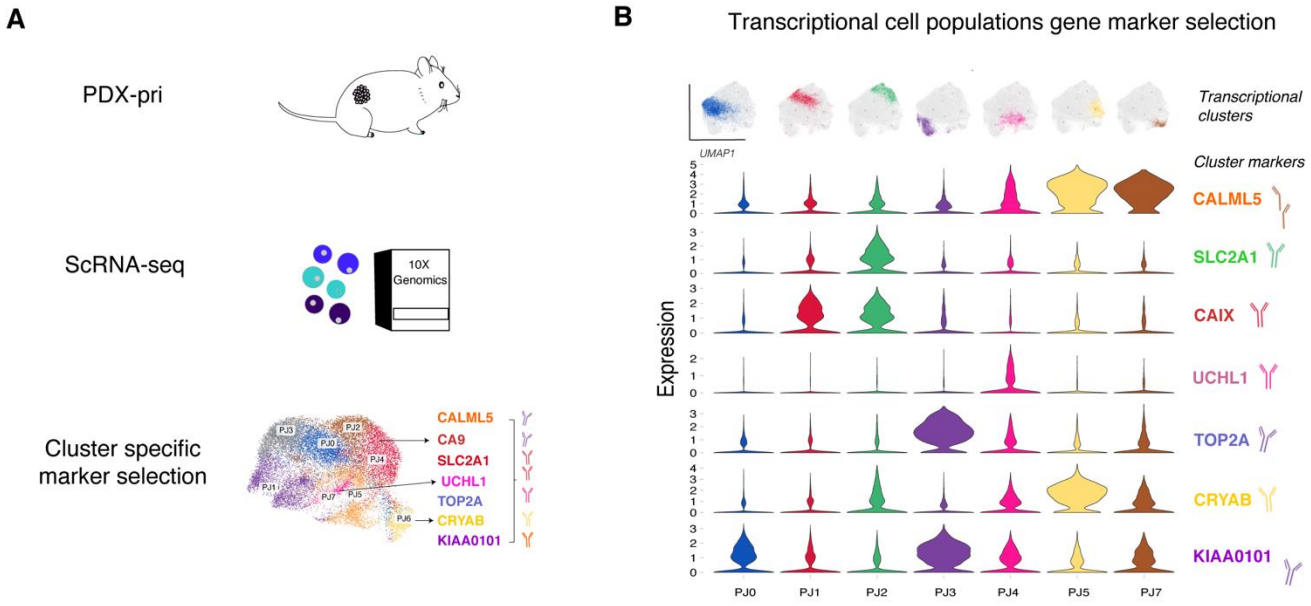


Fig. 2.2.6. In-situ spatial mapping pipeline of single-cell RNA sequencing populations.

A. The PDX primary tumor is dissociated into single cells. These cells are subjected to scRNA-seq using the 10X genomics platform. After bioinformatical analyses, we identified distinct cell populations colored in the UMAP and for each of them we selected specific genes as cluster markers. **B.** The colored violin plots represent the expression of selected gene markers for each PJ cluster. Clusters are represented in the vertical axis with their corresponding colors. In the horizontal axis are the selected gene markers and their corresponding expression per cluster.

2.2.7 Spatial mapping of scRNA-seq populations gene markers in PDX and PT.

To test feasibility and reproducibility of using antibodies for the proteins encoded by the selected gene markers, first we tested the PDX-pri if we could identify these markers by using unstained 4um PDX tumor sections. We stained PDX tumor sections for hematoxylin and eosin (H&E) to define histological tumor features (e.g., necrosis), and for DAPI and the 7 selected protein markers on unstained slides. Confocal microscopy was utilized to obtained images from each individual channel corresponding to one antibody. This way we were able to determine the localization of cells that were positive by immunofluorescence for each marker in the tissue. **(Fig. 2.2.7A).** Similarly, we obtained one H&E slide and one unstained slide from the clinical FFPE block from the patient and performed MIF **(Fig. 2.2.7B).** We were able to determine that each of the selected markers were represented by IF in the PDX, as well as in the patient's primary tumor.

Fig.2.2.7

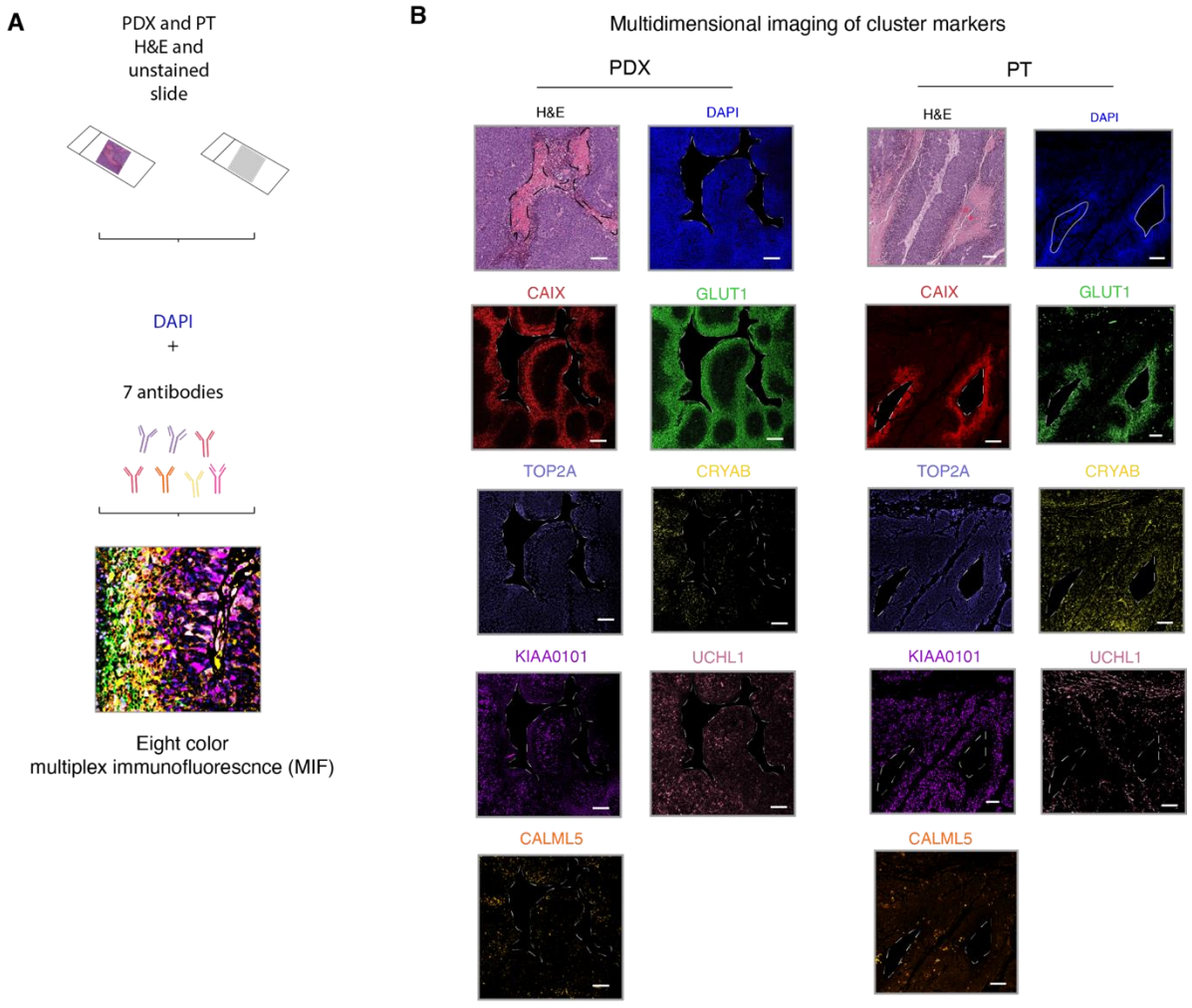


Fig.2.2.7. Marker selection is represented in both PDX and PT tumor.

A. PDX unstained tumor slides were used to stain for DAPI and the seven selected gene markers from scRNA-seq data. **B.** PDX and PT H&E and multidimensional imaging by confocal microscopy is illustrated. Each antibody staining is represented through individual channels. PDX and PT confocal images were masked to exclude regions of necrosis based on the H&E image and disrupted DAPI observed through MIF. In white-dotted lines necrosis is circumscribed. (Scale bar, 100um).

2.2.8 In situ spatial cartography for scRNA-seq populations.

Proof of principle of gene markers represented by IF in both PDX and PT prompt the further characterization of the intensity of expression within one cell. Although each of the selected antibodies can identify cells enriched for the marker protein, the range of intensities for each antibody varies as each cell can express more than one of the selected markers. Therefore, we developed a computational approach that identified each single cell in the tumor and the corresponding expression of each of the to assign each primary tumor cell from the images produced by MIF to a PDX-Pri cell state identified by scRNA-seq **(Fig 2.2.8A).**

To accomplish this, the RNA expression of each of the seven marker genes for the single-cell clusters was scaled in the interval [0, 1] and normalized to the area under the curve. This yielded a probability density function for each marker gene. Thus, each single cell in an image is characterized by seven distinct intensity values obtained by MIF, and computationally analyzed to determine to which scRNA-seq PJ cluster it corresponds **(Fig 2.2.8B)**. The method produces an output computational image in which each cell is colored according to the inferred PDX-Pri cell state **(Fig. 2.2.8C)**. Spatial mapping of the cell states identified transcriptionally allows us to explore the spatial relationship of these states to the microenvironment and to tumor properties (e.g., necrosis), and the organization of cellular networks **(Fig. 2.2.8D**) Altogether, performing in-situ spatial cartography of scRNA-seq populations provides reproducible and unbiased approach to identify markers for tumoral states and examine the spatial relationship of tumor cells alongside their expression profiles.

Fig. 2.2.8

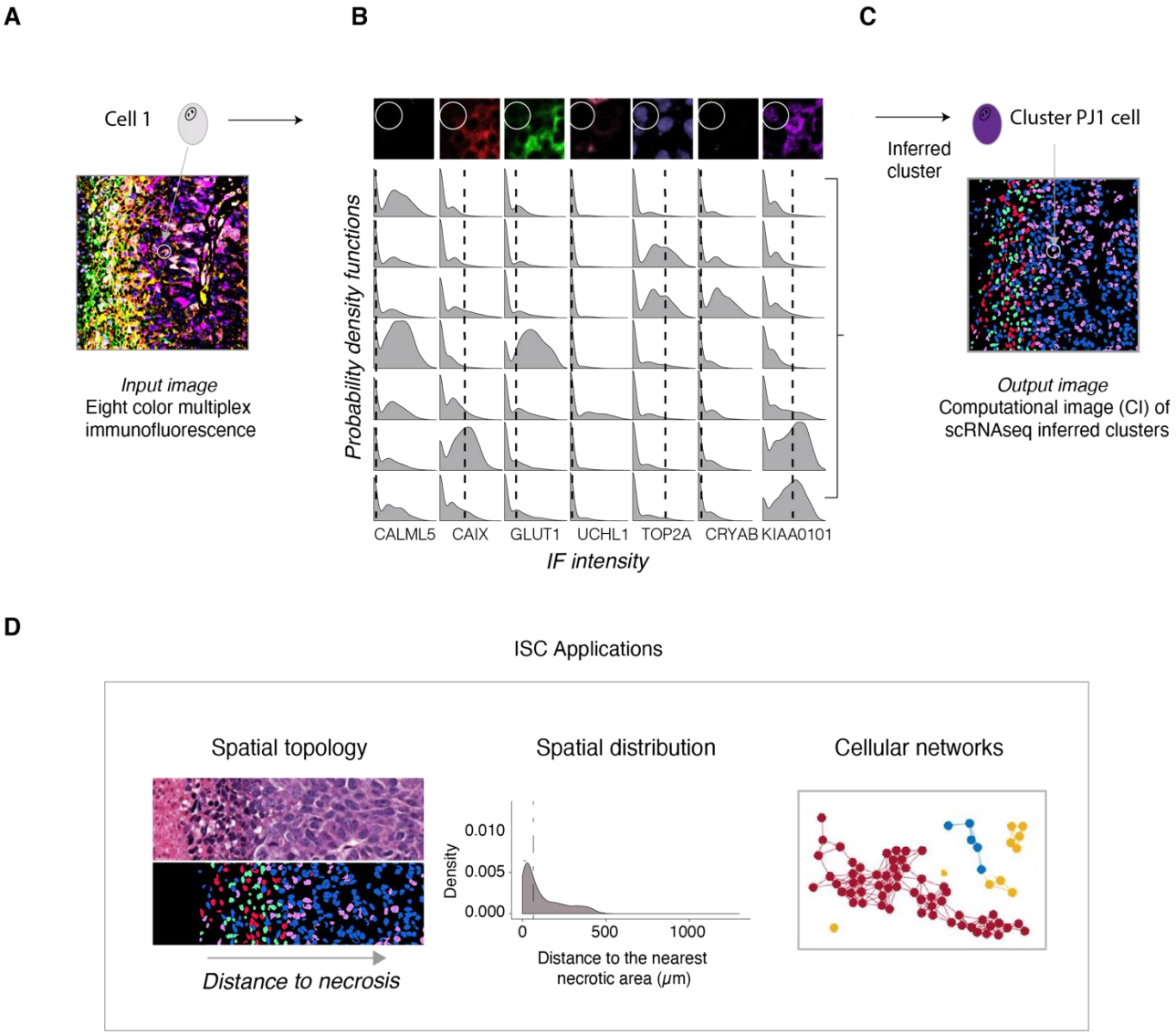

A
B
C
Cell 1
Input image
Eight color multiplex
immunofluorescence
Probability density functions
CALML5
CAIX
GLUT1
UCHL1
TOP2A
CRYAB
KIAA0101
IF intensity
Inferred
cluster
Cluster PJ1 cell
Output image
Computational image (CI) of
scRNAseq inferred clusters
D
ISC Applications
Spatial topology
Distance to necrosis
Spatial distribution
Density
0.010
0.005
0.000
0
500
1000
Distance to the nearest
necrotic area (μm)
Cellular networks

Fig.2.2.8. In situ spatial mapping pipeline.

A. Unstained slides from tumor sections are stained with 8 colors MIF as an input image. Each cell will be analyzed separately. **B.** Each cell (e.g., Cell 1) will have different intensities for each of the antibodies. Probability density functions are obtained from the scRNA-seq expression data, where each cell will be assigned to only one transcriptomic cluster. **C.** After all cells are assigned to their most likely transcriptomic cluster and output image is generated. In this image, each cell is colored corresponding to the scRNA-seq cluster they were allocated to. **D.** Once we have the spatial information and the transcriptional profiles of these cells, spatial topology, spatial distribution, and cellular networks can be interrogated.

2.2.9 Spatial cartography of primary breast tumor populations reveals zonal and zone-less populations adjacent to necrotic regions

Prior to dissecting the spatial relationship of the PJ clusters, we verified that these cell states are found in similar abundances in both the PDX and PT. Using the in-situ spatial cartography (ISC) approach, we spatially mapped the PJ clusters in different tissue sections from PDXs and different regions of the primary tumor **(Fig 2.2.9A)**. In the PDX, 3 biological replicates (three different tumors) and 3 regions in each tumor (technical and heterogeneity replicates) were selected. In the patient's primary tumor 9 ROIs (technical replicates) were selected in the clinical specimen of the tumor. After calculations, we found that within the PJ clusters in both PDX and PT, the most abundant biological populations were PJ0, PJ3, and PJ2 and PJ1 as minor populations PJ5, PJ4 and PJ7 in a proportion of ~9%. Overall, the abundances of cell states identified from single-cell transcriptional analysis are consistent in both the PDX and primary tumor **(Fig 2.2.9B)**. To better understand if gene expression profiles were influence by the spatial localization of clusters, we decided to look at clusters that by MIF already displayed some level of zonation. Single-cell RNA sequencing data revealed two distinct clusters containing signatures of hypoxia. These two clusters expressed CAIX and/or GLUT1 by MIF, displaying a distinct spatial localization adjacent to necrotic regions in the tumor. This is consistent with data from other groups, where hypoxic or non-resolving necrosis in tumors have high expression HIF1A, are associated with poor outcome and display spatial patterns of expression (12,184,185)**.**

Hence, to examine if the spatial distribution and biologically defined gene-expression populations were influenced by distinct environmental ques, the distance from the nearest necrotic region for each cell in PDX and PT was measured. Cluster PJ, which was the most abundant hypoxic cluster was found to have a higher density within 100um from the nearest necrotic regions **(Fig.2.2.9C, D)**. Therefore, we established tumor zones containing this and PJ1 hypoxic population as hypoxic tumor zones (HTZ). On the other hand, we observed cluster PJ3 enriched for cell cycle signatures, homogenously dispersed throughout the tumor and localized within 250um from the nearest necrotic regions, mostly excluded from where hypoxic populations were found **(Fig 2.2.9F, G).** These tumor regions were named cycling tumor zones (CTZ). Finally, **in** contrast to zonal HTZ and CTZ, cells from PJ5 are a small population, that is not spatially defined by their spatial localization. These cells have a homogenous scattered representation throughout the tissue when assayed across 500um from the nearest necrotic regions. These cells were defined as zone-less with scattered distribution in the tumor **(Fig2.2.9H, I).**

Overall, PJ1 and PJ2 cells were preferentially localized closely to necrosis or hypoxic tumor zones (HTZ) within 50um from necrosis. Contrarily, PJ3 cells were found further away from necrosis concentrated in greater proportions over 50um from necrosis, in what we defined as cycling-tumor zones (CTZ). These two zonal populations displayed geographic localization associated to distinct transcriptional profiles, likely dictated by oxygen and nutrient availability. Exceptionally lipid metabolism cells (PJ5), had most of its cells in hypoxic tumor zones, but also in cycling tumor zones. Moreover, the spatial heterogeneity of scRNA-seq populations had concordant results between PDX and PT. Therefore, the use of in situ spatial cartography in PT and matched PDX models uncovers

preserved spatial heterogeneity among these populations. The molecular and topographic properties of these cells may indicate that spatially defined clusters have environmentally influenced gene expression profiles, whereas zone-less populations are defined by cell intrinsic transcriptional profiles.

Fig.2.2.9.

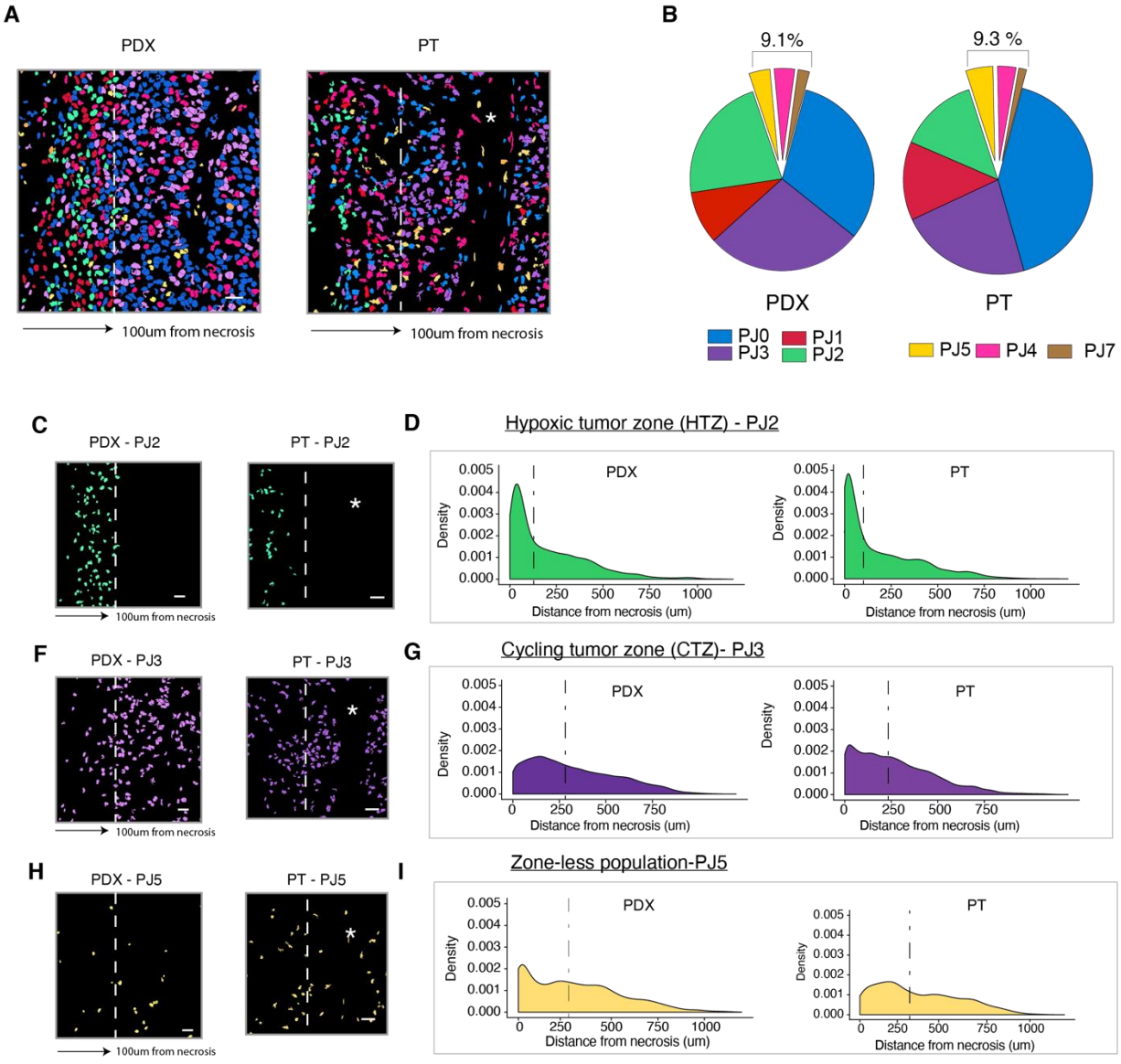


Fig.2.2.9. Spatial quantification of scRNA-seq populations in PDX and PT.

A. Computation image of PDX and PT tumor tissues of all PJ scRNA-seq inferred clusters (scale bar, 20um). **B.** PJ clusters cell proportions in PDX (n=10 images) and PT (n=9 tumor regions). Exploded minor proportion populations. **C-I.** Computational image with distribution of 100um from necrosis for clusters PJ2, PJ3 and PJ5 (left) and their corresponding quantification in density plots (right).

2.2.10 Digital spatial profiling of PDXs and patient's primary tumor and lung metastasis.

The identification of spatially defined populations in the PDX with preserved heterogeneity in the PT, prompt to interrogate if also the gene expression profiles of these cells were retained in the PT. Since the patient's primary tumor material was limited and no scRNA-seq was performed on it, we performed digital spatial profiling (NanoString) for PDX slides and matched PT tissue microarrays (TMA). As observed through MIF and by the gene expression on scRNA-sq data, CAIX was identified as a spatially defined marker for hypoxic scRNA-seq populations. We stained regions of interest (ROI) determined by CAIX+ and adjacent CAIX- regions in PDXs and patient's tumor sections. We used PDX-pri1915, PDX-Lm2076, PT_pri1915) and PT_Lm2076 (lung metastasis). Later, expression profiles in the patient's tumor section determine by CAIX positivity could be compared with the PDX scRNA-seq data **(Fig. 2.2.11A)**. First, three to five CAIX+ and CAIX- regions of interest (ROIs) were selected in the PDX and matched PT samples. The selection of CAIX+ and CAIX- ROIs was done in adjacent regions of the PDX and PT samples. **(Fig 2.2.11B)**. Three different slides or Nanostring experiments were done.

Therefore, batch effect was taken into consideration and quality metrics were similar within three experiments.

Fig.2.2.10

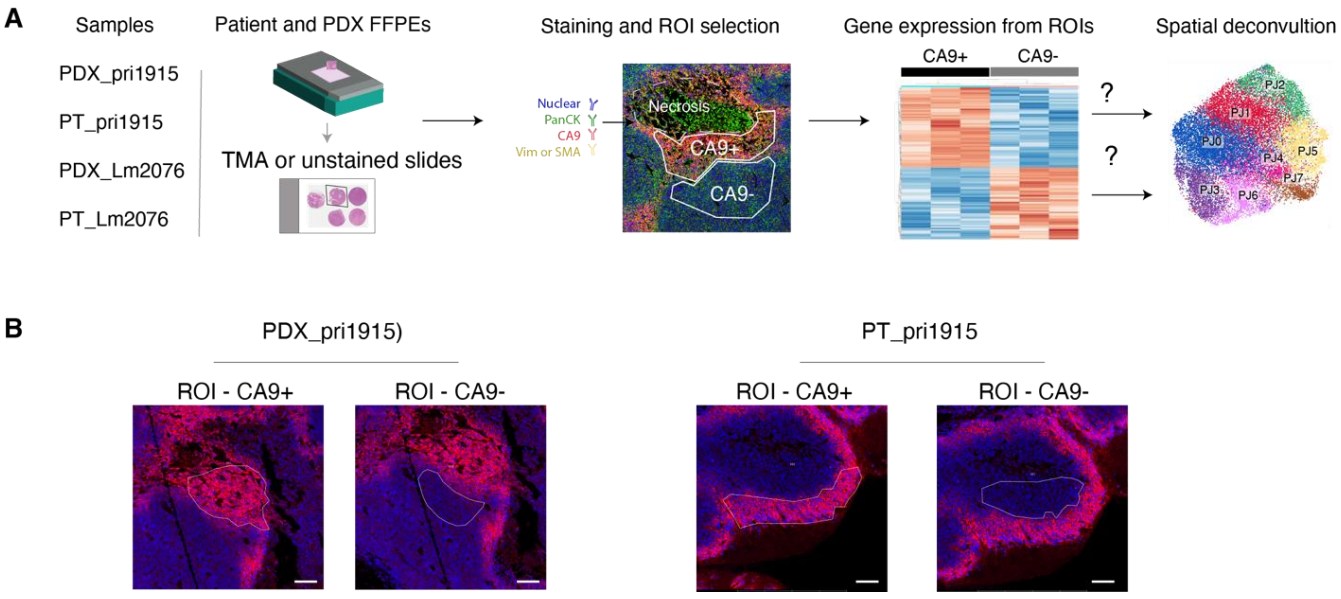


Fig.2.2.10 Digital spatial profiling of PT and matched PDXs samples.

A. Overview of DSP: Samples from the patient's primary tumor, patient's lung metastasis and corresponding PDXs were obtained on FFPE blocks. These were used to build 1 slide and 2 TMAs. The slides and TMAs were stained with nuclear marker, PanCK, CA9 and Vimentin (Vim) or Smooth muscle actin (SMA). Gene expression profiles of CAIX+ and CAIX- were obtained to identify specific scRNA-seq populations in the PJ space by spatial deconvolution methods. **B.** Example of ROI selection of CA9+ and CA9- in the PDX and primary tumor. These was done for all samples in 3 to 5 ROIs per tumor. CA9+ and CA9- selected regions were adjacent to one another.

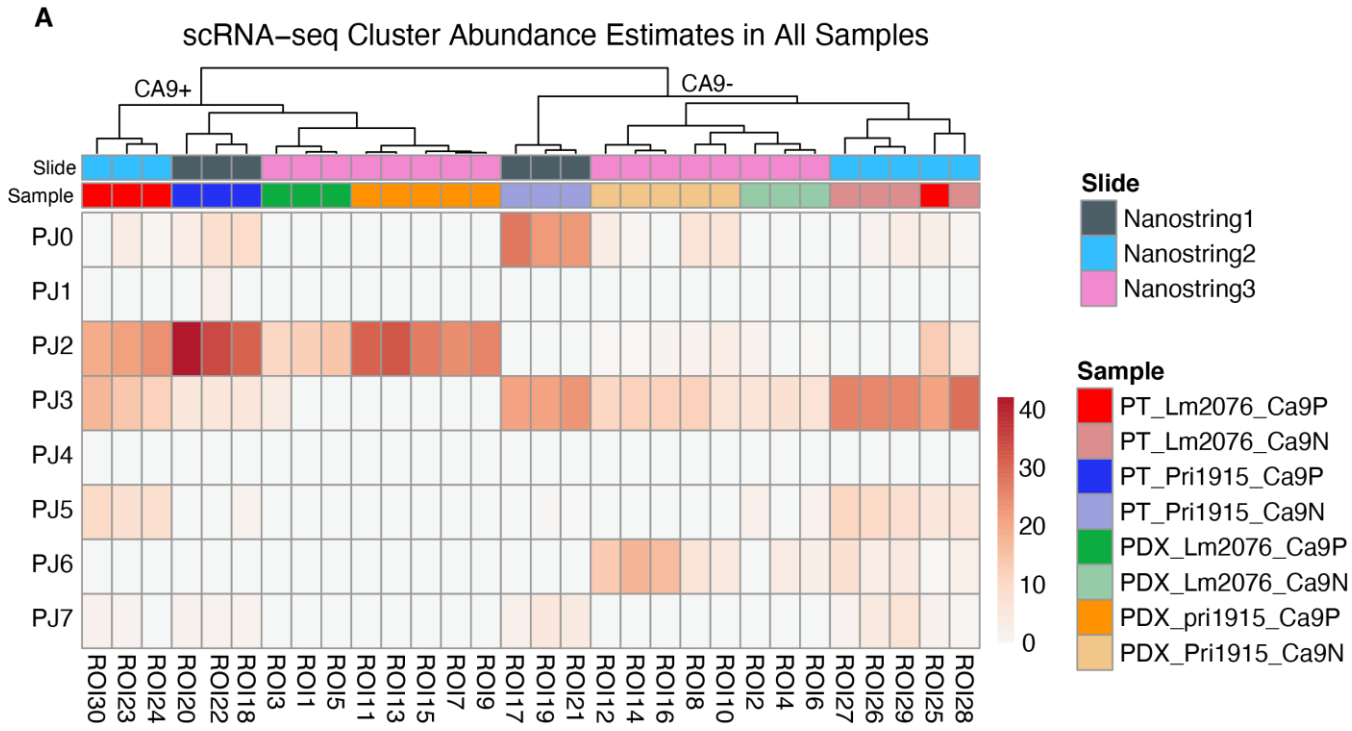
2.2.11 Digital spatial profiling of CA9 positive regions correspond to scRNA-seq hypoxic populations.

To determine if PDX-pri retained the same gene expression profile observed in the hypoxic high cluster, a spatial deconvolution approach was taken. Each ROI was mapped to one of the PJ clusters identified by scRNA-seq. We observed that among all samples with CAIX+ they had a high correlation with the hypoxic cluster PJ2. This was in the case of the primary and metastatic disease PDXs and patient's samples (PT). However, In PDX and PT primary tumors the abundance in PJ2 is higher than in the PDX and PT lung metastasis samples. Moreover, CA9- samples were mostly represent with the higher abundance for the cell cycle cluster PJ3. In the case of the PT_1915 it also had representation of PJ0, a transitional but abundant cluster identified by scRNA-seq **(Fig.2.2.11A)**. The independent CAIX+ and CAIX- ROI representation of each tumor gave additional information regarding the proportion of each scRNA-seq cluster identified in each sample. In CAIX+ ROIs the main cluster was PJ2, followed by PJ3, and found mostly on the metastatic samples was PJ5. In CAIX- ROIs, the most abundant cluster was PJ3, followed by PJ6 transitional cluster, and PJ5 in the lung metastases samples. **(Fig.2.2.11B, C)**.

Altogether, these results validate that spatially defined populations identified by scRNA-seq are preserved among different samples and by different methodologies. Moreover, the spatial heterogeneity of scRNA-seq populations in PDX is preserved in the patient's primary and lung metastasis samples. Nevertheless, for non-spatially defined populations, digital spatial profiling seems to represent those clusters found as more abundant in the scRNA-seq data. For instance, PJ3-cell cycle cluster was the most

abundant, whereas other minor populations are not equally represented. Notably, cluster PJ5, enriched for cholesterol homeostasis signatures, displayed a particular abundance in the metastatic disease, both in CAIX+ and CAIX- ROI and both in the PDX and patient's lung metastatic disease. This suggest that PJ5 cells display high levels of plasticity, being able to survive in both hypoxic or non-hypoxic environments, and in a different anatomic location such as the lung.

Fig.2.2.11



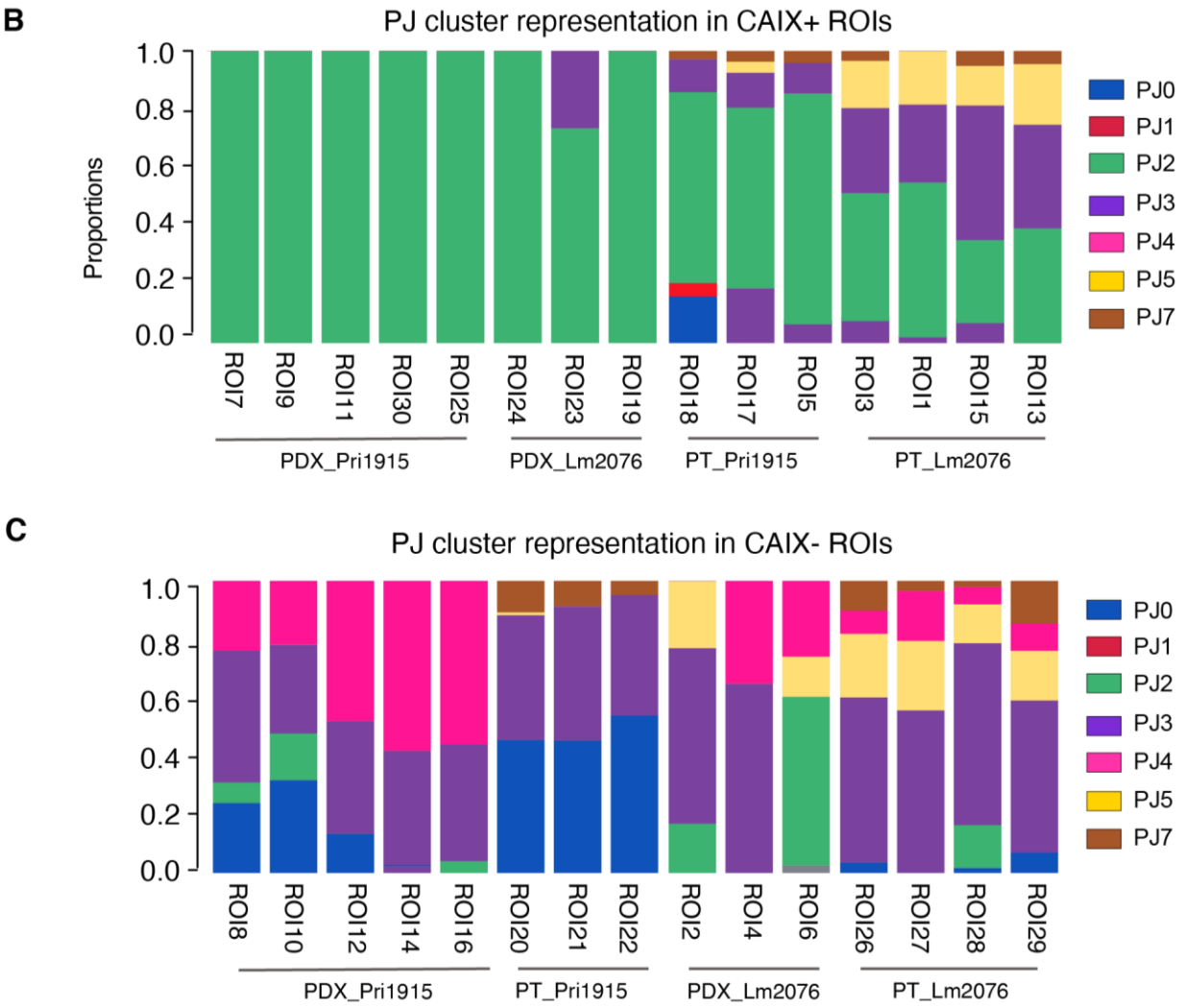


Fig.2.2.11. Spatial deconvolution of scRNA-seq clusters through digital spatial profiling.

A. Abundance estimates of scRNA-seq PJ clusters in all samples. ROIs are illustrated on the horizontal axis and PJ clusters on the vertical axis with their corresponding colors. Nanostring experiments and corresponding samples are labeled as slide and sample respectively. **B, C.** CAIX + (top) and CAIX – (bottom) scRNA-seq proportion in each independent ROI in the horizontal axis. Clusters legend on the right side.

2.2.12 Hypoxic CAIX positive cell populations are identified in early-stage tumors.

The presence of hypoxia on primary tumors, has been shown to be involved in the selection of cells with survival advantage, EMT-like properties and metastatic capabilities (186)(187). EMT was a characteristically enriched pathway in hypoxic clusters, pointing that these cells could play a role in tumor initiation or tumor progression. Since CA9 gene and protein expression (CAIX) levels were specific for this cluster, we utilized it as a surrogate marker to enrich and functionally test these cells **(2.2.12A).** To understand if PJ1/PJ2 cells were a snapshot of advanced stage necrotic tumors, or were rather present also in early tumor development, we collected PDX-pri tumors of different sizes and performed CAIX IHC. Tumors as small as $4mm^3$ presented zonal expression of CAIX adjacent to necrotic regions, supporting the early presence of hypoxic-PJ4 populations in tumor initiating conditions **(2.2.12B).** These data support that hypoxic populations are present in early stage tumors in what would likely correspond to pre-necrotic or less oxygenated regions in the tumor.

Fig 2.2.12

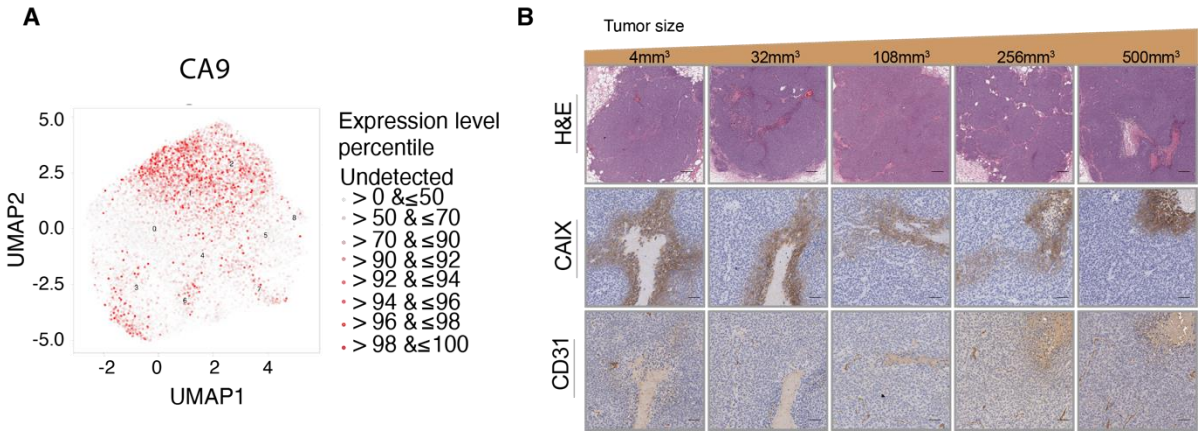


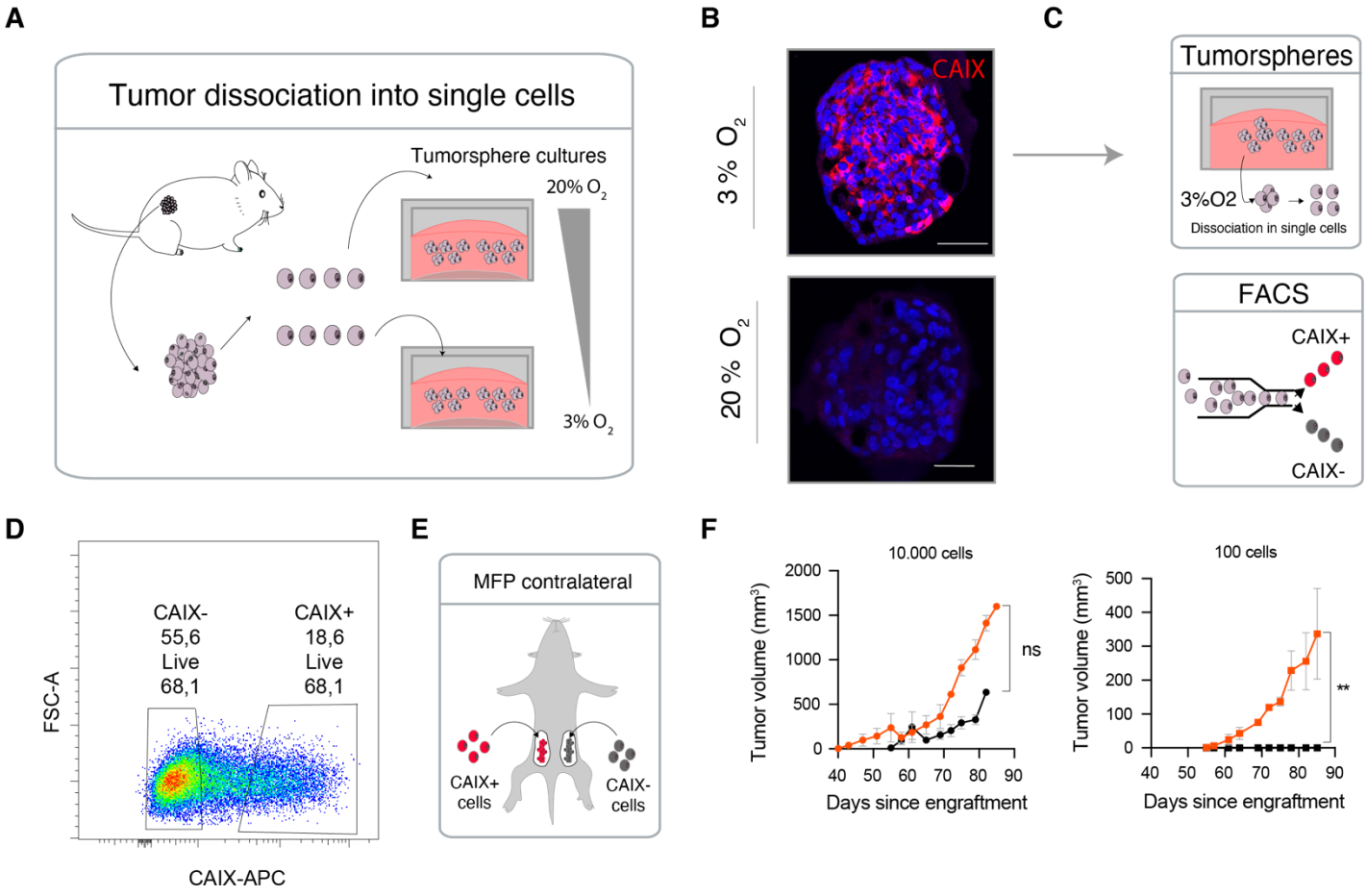
Fig 2.2.12. CA9 is a marker of hypoxic populations in early and late-stage tumors.

A. PJ UMAP depicting CA9 expression level in percentiles throughout all scRNA-seq populations. **B.** In vivo experiment with resection of tumor at different sizes/stages. H&E staining at the top (Scale bar;100um). CAIX and CD31 IHC staining in the middle and bottom respectively. (Scale bar; 20um).

2.2.13 Hypoxic CAIX positive cell populations display higher tumor initiating capacity in vivo.

To functionally test the role of CAIX in tumor initiation, PDX-pri tumors were dissociated into single cells, and cultured in non-adherent conditions as tumorspheres in both 20% and 3% oxygen chambers **(Fig. 2.2.13A)**. Tumorspheres in 3% O_2 expressed CAIX when compared with 20%O_2 tumorspheres which lacked expression of the protein in this condition **(Fig. 2.2.13B)**. To further interrogate the biological behavior of PJ1/PJ2 population in vivo, 3% O_2 tumorspheres were dissociated into single cells and fluorescent-activated cell sorted (FACS) for CAIX+ and CAIX- cells was performed **(Fig. 2.2.13C, D)**. Tumor initiating capacity (TIC) in vivo was tested by injecting bilaterally CAIX+ and CAIX- sorted cells into the fourth mammary fat pad (MFP) of 18 mice (MFP right: CAIX+ cells; MFP left: CAIX- cells) in three different dilutions; 10,000, 1,000 and 100 cells (6 mice per group) **(Fig. 2.2.13E)**. CAIX+ tumors grew in 5/6 mice with 10,000 cells, 2/6 with a 1,000 and 3/6 in the lowest dilution with 100 cells. On the other hand, CAIX- tumors only grew 4/6 tumors in the highest 10000 cells dilution and none with the lowest cell inoculums. Extreme limiting dilution assay (ELDA) significantly demonstrated that CAIX+ cells have a higher TIC capacity with 1 in 2,792 cells **(Fig. 2.2.13F, G)**. Altogether, these data support that hypoxic populations, characterized by CAIX positivity, are found in higher proportion in hypoxic conditions in vitro and display higher tumor initiating capacities in vivo.

Fig 2.2.13



G

Cells injected	CAIX (-)	CAIX (+)
10000	4/6	5/6
1000	0/6	2/6
100	0/6	3/6
TIC frequency	1/10896	1/2792

p=0.0262*

Fig.2.2.13. Primary tumor CAIX positive sorted cells have a higher tumor initiating capacity.

A. Experimental outline part 1; PDX tumors were dissociated into single cells and then cultured in non-adherent conditions to obtain tumorspheres both 3% and 20% oxygen. **B.** Tumorspheres in Histogel stained with DAPI and CAIX in 3% O2 (Scale bar, 50um). **C**. Experimental outline part 2; tumorspheres in 3% oxygen were sorted for CAIX+ and CAIX- cells. **D.** FACS-sorted CAIX+ and CAIX- single cells gate. **E.** Experimental outline part 3; CAIX + and CAIX- sorted cells were injected contralaterally in the mammary fat pad (MFP) of mice (n=18, 3 groups). **F.** Growth curves of PDX-pri tumors of CAIX+ (red) and CAIX- (black) sorted cells with 10.000 (left), 1000 (not shown), and 100 cells (right). (mean SEM; ** $p<0.005$: ns, not significant, t test). **G.** Estimated tumor-initiation frequency after three serial dilutions:10.000, 1000, 100 cells (n=6 per arm) (*$p<0.05$).

2.2.14 Primary tumor and tumorspheres spatial cartography on different oxygen conditions.

To further investigate in situ phenotypic plasticity of zone-less cluster PJ5 in the primary tumor, cells were assessed for positivity for markers of distinct zones. CAIX was selected as a marker to define HTZ, because of its high specificity to select for hypoxic clusters (PJ1/PJ2)**.** Likewise, CAIX negativity and KIAA0101 positivity were used to determine cells from PJ3 in CTZ. Imaging and quantification evidenced that cells positive for CRYAB (marker of PJ5 cluster) could adopt CAIX+ expression when in hypoxic tumor zones (31%) and when in CAIX- tumor zones they could adopt positivity for KIAA0101 (29%). The definition of hypoxic and cycling tumor zones was represented by both expression of CAIX and KIAA0101, where only 5% of cells were triple positive **(Fig.2.2.14A, B)**. This supports that CRYAB cells representing PJ5 cluster, have phenotypic plasticity and may be able to adapt to distinct environments in the tumor. Hypoxia has been shown to be involved in the selection of cells in the primary tumor with a survival advantage, EMT-like properties and the ability to invade distant organs (186,187). To functionally test tumor initiation in PDX-pri tumors were dissociated into single cells and cultured as tumorspheres in both 20% and 3% oxygen conditions **(Fig. 2.2.14C)**. Tumorspheres cultured in 20% O_2 did not express CAIX, but most cells expressed CRYAB. On 3% O_2, tumorspheres expressed CAIX and most cells also expressed CRYAB **(Fig. 2.2.14D)**. To have an accurate quantification not only of cluster markers, but of scRNA-seq populations, ISC was applied to tumorspheres identifying that PJ2 and PJ5 cluster cells were accounting for most of the tumorspheres' cells **(Fig. 2.2.14E).** Also, when looking at the difference in PJ cluster proportion between CAIX+

and CAIX- sorted tumors, the main difference observed was within cluster PJ5 **(Fig.2.2.14F)**. These data, suggests that hypoxic (PJ2) and cholesterol homeostasis cell have a more aggressive phenotype in the context of tumor initiation.

Fig. 2.2.14

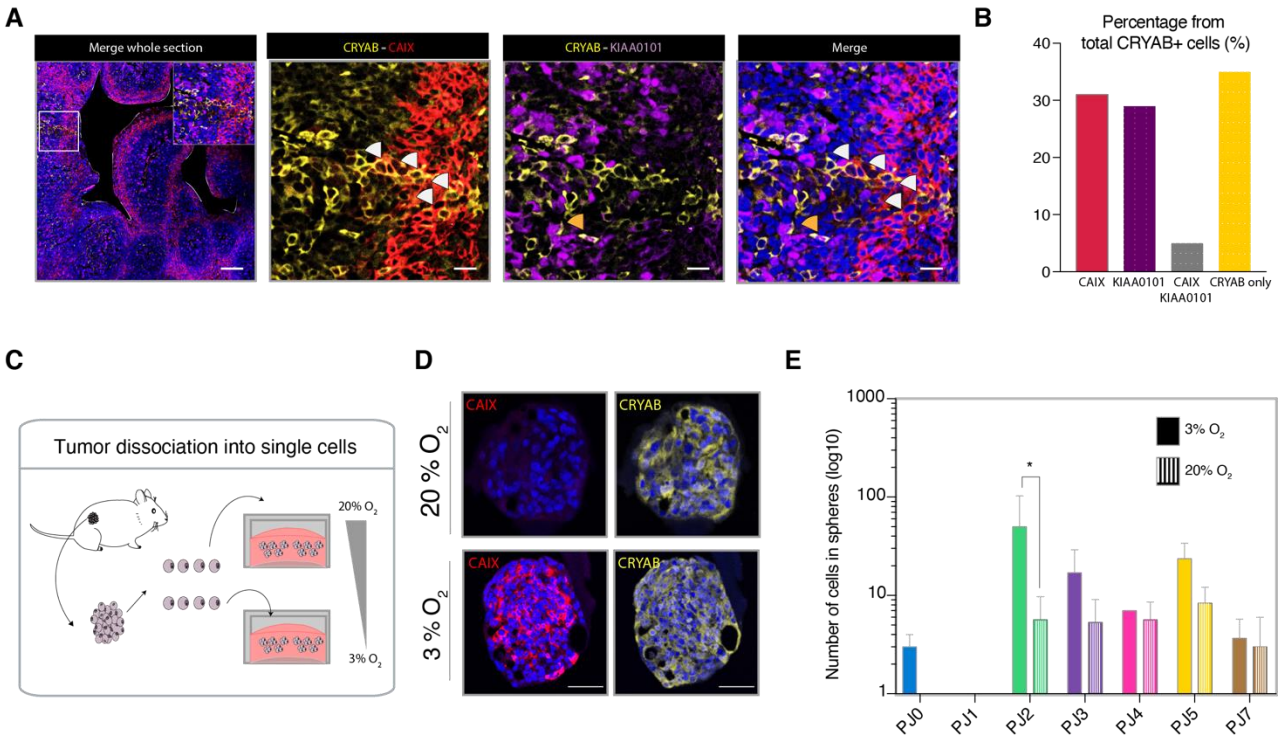


Fig.2.2.14 Spatial cartography of markers and scRNA-seq populations on different oxygen conditions.

A. MIF from whole section (WS) and inset images for DAPI, CRYAB, CAIX and KIAA0101 (Scale bar WS, 100um and inset 20um). **B.** Percentage of all CRYAB positive cells that are also positive either for CAIX, KIAA0101 or both in image in A. **C.** Experimental outline, tumor dissociation into single cells and cultured tumor-spheres in 3 and 20% oxygen. (D) DAPI image of spheres merged with MIF individual channels for CAIX, CRYAB, KIAA0101 and KI67 of single-cell derived tumorspheres histogel in 3% O2 .(Scale bar, 50um). DAPI KI67 is taken from consecutive histogel section. **E.** Number of cells from tumorspheres inferred PJ clusters quantification (mean SEM; * $p<0.05$, t test).

2.2.15 Breast cancer lung metastases retain primary tumor biological populations.

To determine if transcriptional clusters were retained from the primary tumor upon disease progression, we performed single-cell RNA sequencing from 2 of the established PDX from the patient's lung metastasis (PDX_Lm1 and PDX_Lm2) and from the spontaneous lung metastasis PDX model (PDX_Slm) **(Fig 2.2.15A)**. The three lung metastases samples were integrated into a metastases joint space (MJ), identifying 6 transcriptomic populations (MJ0-MJ5). After alignment, demultiplexing and filtering of murine cells, the replicates were integrated using Harmony (173) (**Figure 2.2.15B**). Among these replicates, a high degree of reproducibility was displayed. Moreover, the number of cells, QC metrics and cell population structure is well represented within the three samples. However, the PDX_Slm sample displayed less cluster proportions since it was a smaller sample coming directly form a lung lesion in the mouse. Number of genes, mitochondrial content and ribosomal content was similar in all samples **(Figure 2.2.15C, D)**. Finally, on primary and metastases cluster correspondence, we identified that cluster PJ5 (cholesterol homeostasis-yellow) was mostly represented in cluster MJ4 (yellow), cluster PJ3 (cell cycle-purple) with cluster MJ2 (red), and cluster PJ2 (hypoxic-green) with cluster MJ1 (orange) **(Figure 2.2.15E).** Altogether, some of the lung metastases populations are previously identified in the primary tumor. These populations are mostly the defined biological clusters, (Hypoxic-PJ2, cell cycle-PJ3, and cholesterol homeostasis-PJ5). These clusters, were also identified as the most abundant in metastatic samples by DSP, suggesting that cells with metastatic capability, can be observed in the primary tumor through scRNA-seq and by spatial mapping in situ.

Fig.2.2.15

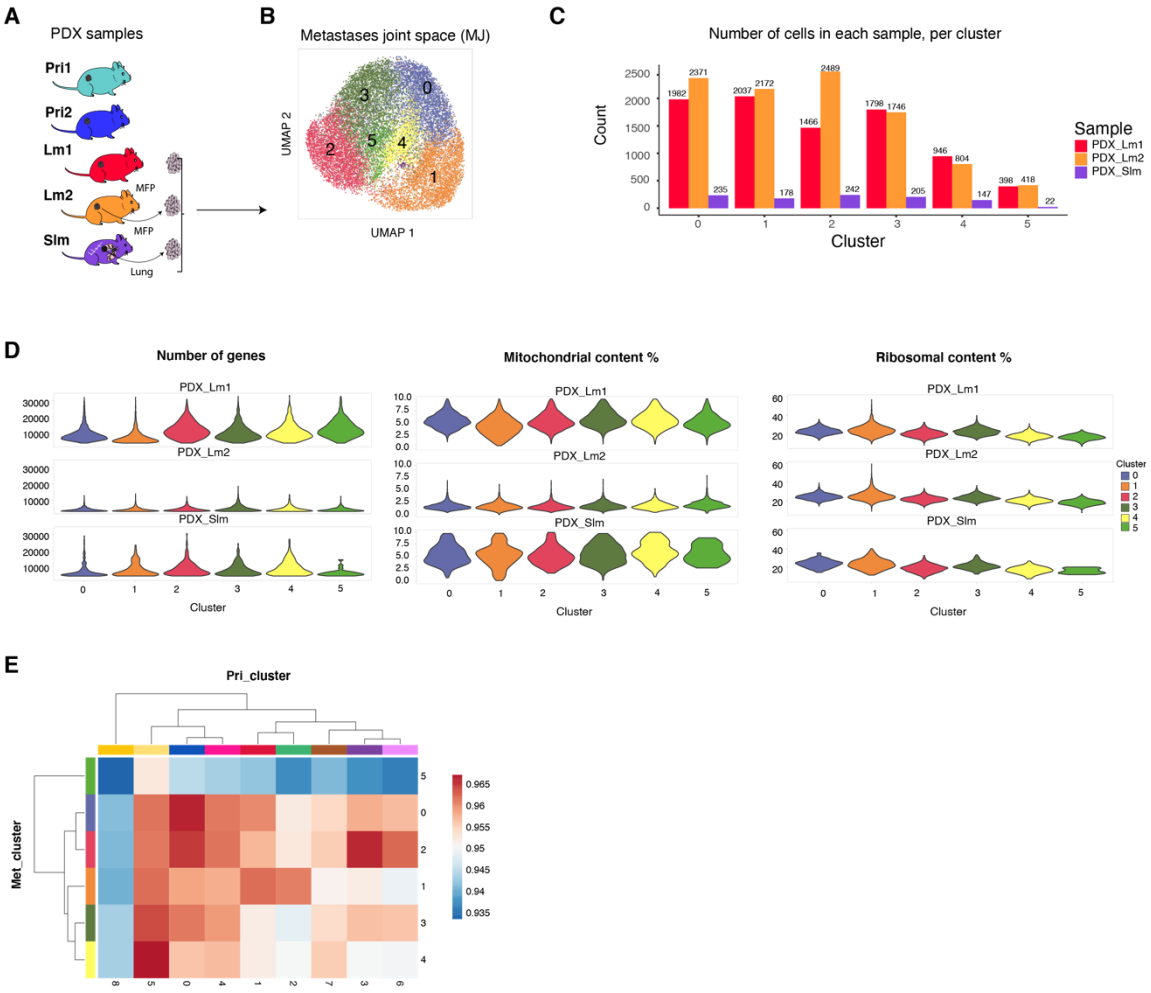


Fig.2.2.15. Single-cell RNA sequencing of lung metastases cancer PDXs.

A. PDX samples. Three metastatic samples (2 from the MFP and one from a spontaneous lung metastasis in the mouse) sent for scRNA-seq. **B**. UMAP visualization of the primary joint (PJ) analysis, in which all three metastatic samples (Lm1, Lm2 and Slm), were integrated using Harmony. Cells are colored by cluster identity. **C.** Comparison of cell number in each sample per cluster. **D.** Cluster quality metrics of number of genes, mitochondrial content, and ribosomal content for all three samples. **E.** Heatmap of cluster correspondence between primary PDXs clusters (Pri cluster) and lung metastases clusters (Met_cluster).

2.2.16 Spatial cartography and scRNA-seq reveals that hypoxic and cholesterol homeostasis populations display an aggressive phenotype

The primary tumor PJ clusters were examined in PDX_Slm and the patient's lung metastasis biopsy (PT_Lm). The spatial heterogeneity was interrogated, and PJ cluster quantification was performed also retained upon disease progression. Histological samples were stained by MIF, and we applied ISC to further quantify PJ clusters in the metastatic samples **(Fig. 2.2.16A, B)**. PDX-Slm and PT_ Lm cell proportion assigned to each PJ cluster, showed that most cells were in clusters PJ3, PJ5, and PJ2. On the other hand, a minor proportion of cells (~10%) were allocated to clusters PJ4, PJ0 and PJ1. PJ7 showed to be increased in the PDXs when comparing to the Lm_human **(Fig. 2.2.16C)**. Moreover, the only cluster significantly increased in the metastatic samples compared to the primary tumors was PJ5, while PJ0 and PJ1 were reduced **(Fig. 2.2.16D)**.

Since the lung environment compared to the primary tumor in the mammary gland has higher levels of oxygen tension (188), we investigated if zonal and zone-less primary tumor populations were spatially resolved in a similar manner in the metastatic disease. Therefore, distance from the nearest necrotic region was measured. In the lung metastases samples, most populations were found withing the first 200um from the nearest necrotic region, PJ2 having the highest proportion of cells in these regions. Over 200um, all clusters show a homogenous spatial distribution throughout the tissue **(Fig. 2.2.16E)**.

Moreover, overall survival (OS) and distant metastasis free survival (DMFS) were interrogated in basal breast cancer, were PJ2- hypoxic cluster signature displayed reduced DMFS and PJ5-cholesterol homeostasis cluster poor OS **(Fig. 2.2.16F, G)**. Altogether, we demonstrate that the unbiased marker selection for the epithelial compartment and the spatial mapping of single cell populations in situ is a feasible tool to perform biological interrogations, not only between PDX and human samples, but also upon disease progression. Cluster PJ2 has a significant role in tumor initiation and may enhance selection of cells with aggressive metastatic phenotype. Moreover, these findings identify the expansion of the primary tumor population PJ5, enriched in cholesterol homeostasis signatures, suggesting that PJ5 population displays phenotypic plasticity, and that possibly has a role in tumorigenesis and disease progression.

Fig.2.2.16

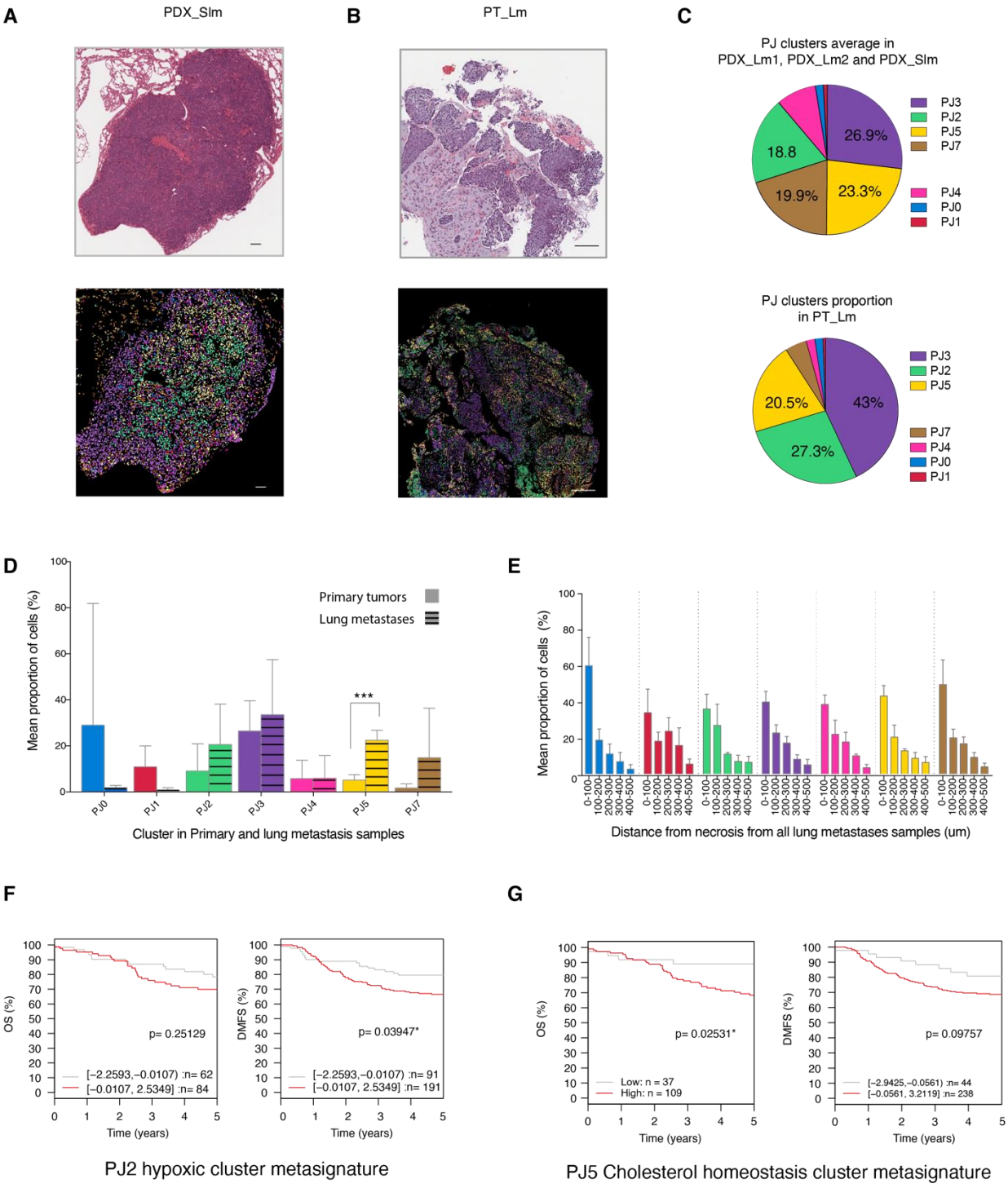

A
PDX_Slm
B
PT_Lm
C
PJ clusters average in
PDX_Lm1, PDX_Lm2 and PDX_Slm
26.9%
23.3%
19.9%
18.8
PJ3
PJ2
PJ5
PJ7
PJ4
PJ0
PJ1
PJ clusters proportion
in PT_Lm
43%
27.3%
20.5%
D
Primary tumors
Lung metastases
Mean proportion of cells (%)
***
PJ0
PJ1
PJ2
PJ3
PJ4
PJ5
PJ7
Cluster in Primary and lung metastasis samples
E
Mean proportion of cells (%)
Distance from necrosis from all lung metastases samples (um)
F
OS (%)
p= 0.25129
[−2.2593,−0.0107) :n= 62
[−0.0107, 2.5349] :n= 84
Time (years)
DMFS (%)
p= 0.03947*
[−2.2593,−0.0107) :n= 91
[−0.0107, 2.5349] :n= 191
Time (years)
PJ2 hypoxic cluster metasignature
G
OS (%)
p= 0.02531*
Low: n = 37
High: n = 109
Time (years)
DMFS (%)
p= 0.09757
[−2.9425,−0.0561) :n= 44
[−0.0561, 3.2119] :n= 238
Time (years)
PJ5 Cholesterol homeostasis cluster metasignature

Fig.2.2.16. Spatial cartography of scRNA-seq populations in metastatic samples.

A, B. H&E of PDX_Slm and PT_lm (top) and corresponding spatial cartography of PJ clusters in computational image (bottom). Scale bar, 20um and 100um respectively. **C.** Quantification of PJ cluster proportion in the three metastatic samples (Lm1, Lm2 and Slm) and in the patient's lung metastasis biopsy (PT_Lm). **D**. Mean proportion of cells in primary tumor and lung metastases samples (n=3, 2 PDXs and 1 human sample for each). T-test ***p=0.0005. **E.** Mean proportion of cells assigned to each PJ cluster and their distance from the nearest necrosis from all lung metastases samples (n=3). **F, G**. Overall survival (OS) and Distant-metastasis free survival (DMFS) of HU_Basal tumors PJ2 (left) and PJ5 (right) meta-signatures of TOP10 genes in each cluster (*p<0.05. Log Rank (Mantel-Cox) test). GOBO (Ringer et al.,2011).

2.2.17 Combined cholesterol and hypoxic targeted therapy display growth arrest and increased distant-metastasis free survival in PDX.

To functionally interrogate mechanisms of sensitivity to different drugs we first used a PDX clinical trial approach with commonly used chemotherapies and targeted therapies. PDX-pri showed to be unresponsive to multiple lines of therapy as tested in a 1x1x1 study approach (160). However, it presented a partial response to Gemcitabine, same drug used in the adjuvant setting for this patient **(Fig. 2.2.17A)**. Due to the observed poor OS in basal breast cancer and hypoxia playing a role in phenotypic plasticity of some populations in the tumor, we treated readily grown PDX tumors (~100mm^3) with SLC-0111 a CAIX inhibitor (CAIXi) alone, Gemcitabine (Gem) alone and both drugs in combination. The combination arm had a significant delay in tumor growth, while CAIXi had no effect when compared to vehicle (Veh) **(Fig. 2.2.17B, C)**.

As the presence of CAIX showed enhanced tumorigenesis under tumor initiating conditions, we performed a second PDX trial, treating from the moment of tumor engraftment in the MFP. We used SLC-0111 and Ro 48-8071, a 2,3-oxidosqualene: lanosterol cyclase activity inhibitor (OSCi), to also target PJ5 population, known to be enriched in cholesterol homeostasis signatures and present in the metastatic disease samples. Additionally, tumors were resected when reaching endpoint (>500mm^3) to also assess the role of these drugs in the reduction of distant metastases incidence **(Fig. 2.2.17D)**.

Tumors treated from day 0 had a significant delay in tumor growth both with CAIXi alone and in combination with OSCi **(Fig. 2.2.17E)**. Significant reduction in tumor burden at endpoint was only displayed by CAIXi **(Fig. 2.2.17F)**. Moreover, after tumor resection the 120 days follow-up showed a significant increase of distant-metastasis free survival (DMFS) in tumors treated with both CAIXi and OSCi or alone with CAIXi. In this case the increase in DMFS was only significant in the combination treatment. On the counterpart, OSCi alone had an increase of metastatic burden compared to control (Veh) **(Fig. 2.2.17G)**.

Altogether, we sought to target two distinct populations, with signatures of hypoxia and cholesterol homeostasis. PDX treatments identified sensitivity to CAIXi only when tumors were treated from the moment of engraftment, supporting the role in tumor initiation. The decreased metastatic burden in combination therapy supports a role of cholesterol homeostasis as having a role in disease progression in basal breast cancer, more than only in the tumor initiating context.

Fig.2.2.17

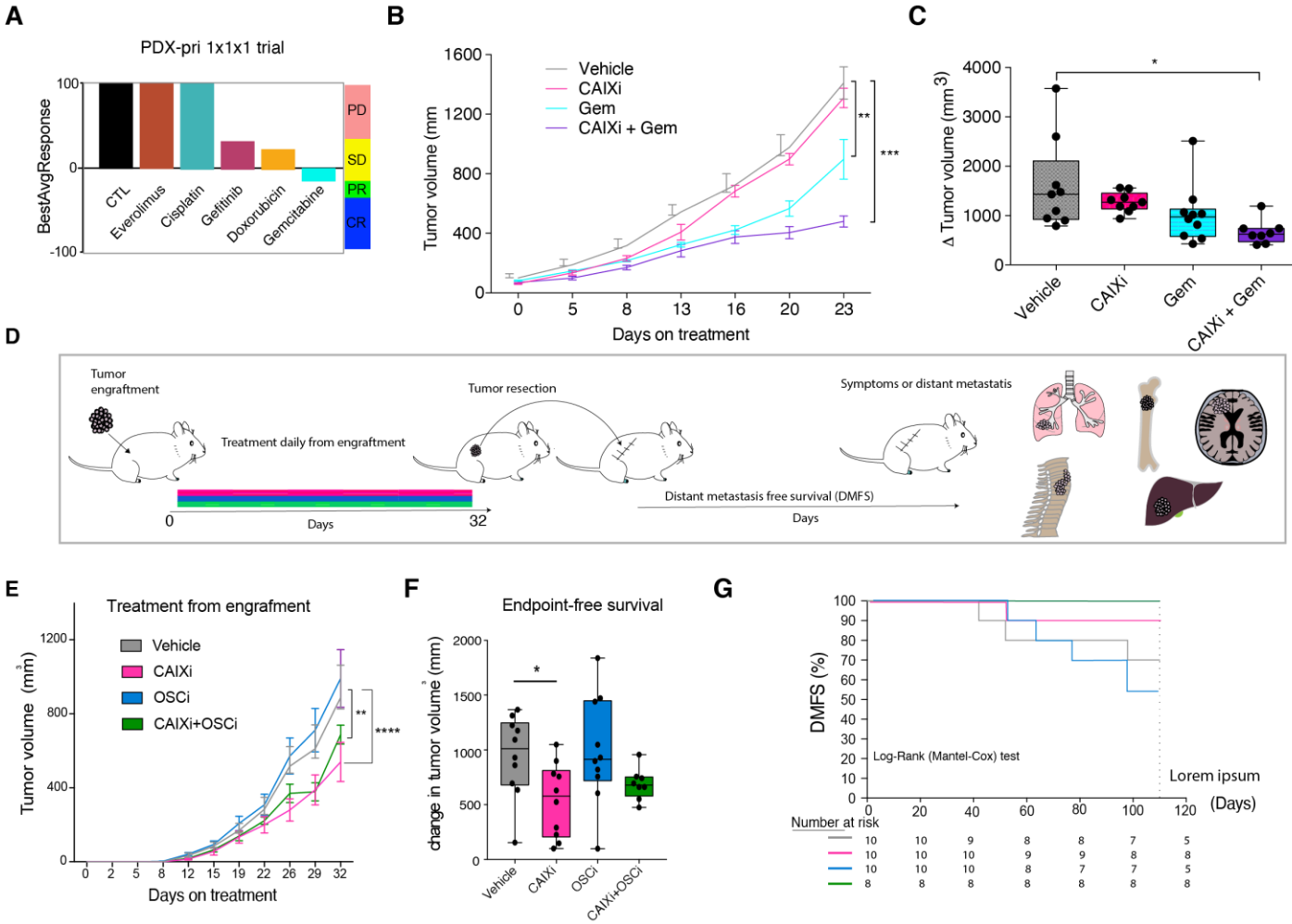

A
PDX-pri 1x1x1 trial
BestAvgResponse
CTL
Everolimus
Cisplatin
Gefitinib
Doxorubicin
Gemcitabine
PD
SD
PR
CR
B
Tumor volume (mm
Vehicle
CAIXi
Gem
CAIXi + Gem
Days on treatment
C
Δ Tumor volume (mm3)
Vehicle
CAIXi
Gem
CAIXi + Gem
D
Tumor engraftment
Treatment daily from engraftment
Tumor resection
Symptoms or distant metastatis
Distant metastasis free survival (DMFS)
Days
E
Treatment from engrafment
Vehicle
CAIXi
OSCi
CAIXi+OSCi
Tumor volume (mm3)
Days on treatment
F
Endpoint-free survival
change in tumor volume (mm3)
Vehicle
CAIXi
OSCi
CAIXi+OSCi
G
DMFS (%)
Log-Rank (Mantel-Cox) test
Lorem ipsum
(Days)
Number at risk

Fig.2.2.17. Combination therapy with CAIX and OSC inhibitors delays tumor growth and increases DMFS.

A. PDX-pri best average response (Progressive disease (PD), Stable disease (SD), Partial response (PR) and Complete response (CR)) for different drugs. **B.** Growth curves of treated tumors when reached ~100mm3 with Vehicle, SLC-0111(CAIX inhibitor), Gemcitabine (Gem) or combination (Gem+SLC-0111). (n=9-10 per group). **C.** Change in tumor volume from treatment to endpoint (*p<0.05, **p<0.005, p<0.0005*** one way ANOVA). **D.** Experimental Pipeline: 1x1x1mm3 tumors were engrafted unilaterally in the fourth mammary fat pad and treated from day 0 with Vehicle (Veh), SLC-0111, or combination of SLC-0111 and Ro-4871 (OSC inhibitor). Tumors reaching ~500mm3 were resected. Mice were left to develop symptoms or metastatic disease detectable by imaging. **E.** Growth curves indicating response from moment of tumor engraftment to the different treatments (n=8-10 per group). **F.** Endpoint-free survival tumor volume change at 32 days for all mice with their corresponding treatment. (*p<0.05, t test). **G.** Distant metastasis free survival (DMFS) curves of mice that underwent resection of tumors (*p< 0.05 Log Rank (Mantel-Cox) test).

3. GENERAL DISCUSSION

3.1 Breast cancer heterogeneity beyond conventional tumor heterogeneity

Triple-negative breast cancer is not a single disease and despite recent and common efforts to identify therapeutic strategies for these patients, chemotherapy is still used as the main standard of care. Understanding intratumor heterogeneity has helped to elucidate molecular differences among breast cancer and TNBC, but overall does not account for the different responses observed in patients that have the same BC subtype.

Moreover, TNBC displays genomic heterogeneity with variable clinical presentation, histological features, and response to therapy. Genome-wide profiling of TNBC reveals far greater diversity at the genomic level (189) than at the transcriptomic level, from which TNBC is stratified within four distinct transcriptomic subtypes (69). Similarly, differences in stroma (190) and immune populations (114) are linked with outcome. However, this growing information has been slowly integrated in the clinic to change clinical practice, and the consideration of the tumor microenvironment into treatment is still underexplored.

TNBC Bulk tumor populations contain several cell types, including malignant cells and non-malignant stromal and immune cells that support or oppose their growth. Although each individual cell within a tumor adopts a gene expression pattern governed by its cell identity, these patterns adapt in response to cell-extrinsic factors and the local microenvironment in response to treatment. Hence, the exposure to specific local cues is an important source of heterogeneity for tumor cells sharing the same identity.

The understanding of breast cancer heterogeneity has developed rapidly, where the predominant use of clinical receptors to stratify patients, has been redefined by transcriptional intrinsic subtypes, specific genomic alterations in BRCA genes and lately, through BC immunophenotyping, to consider immunotherapy as an adjunct treatment (191). Altogether, these discoveries have contributed significantly to the advancement of knowledge in the breast cancer field and have led to subtle changes in clinical practice and breast cancer patient outcomes (192). The use of targeted therapy, PARP inhibitors, Immunotherapy, and specific platinum agents for patients with BRCA mutations are now therapeutic complements to conventional chemotherapy regimens, that remain as standard of care (193).

3.2 The use of PDXs as preclinical models of disease progression

PDXs are a powerful tool to obtain an unlimited source of clinical material for research. In this thesis, a series of samples providing a stepwise development of disease progression samples were presented, intending to generate models that enable the interrogation key clinical challenges to understand disease progression. The use of primary and matched PDXs from the same patient to understand tumor cell selection upon metastatic disease is extremely valuable due to the technical difficulties that are presented when obtaining these samples. Metastatic samples are usually obtained from a procedure that would likely result in small tissue biopsy. The biopsy material of the metastatic sample is mostly available for clinical use, and the use of some" leftover" material for research implicates associated challenges. The development of new PDX biobanking and close follow-up of patients from whom we banked primary tumor material

is essential to obtain matched metastatic samples. Moreover, the development of rapid autopsy programs within the clinical and research domain, will in the future allow the development of these models and further sequencing of metastatic lesions in a timely manner. This approach becomes important when interrogating metastatic tumor populations within distinct tumor microenvironments or anatomic sites (194).

In this thesis, the use of these samples was explored in a reproducible manner and with extensive molecular and phenotypic characterization. Moreover, the establishment and subsequent scRNA-seq data of residual disease, rebound and resistance samples, will provide powerful material for further research avenues. Although, the data has not been fully developed, the time spent in creating these models will provide discovery and mechanistic understanding pursued by others. The utility of these models will expand beyond validations for tumor marker discovery, and the sequential availability of these samples will allow delineation of the selection of aggressive cell populations in the primary tumor during tumor progression.

These PDX models are most representative of a snapshot of a late-stage disease, being of great value for drug selection and screening for response to new therapeutic combinations. The therapeutic opportunities that these samples enable in the clinical domain are the most likely to be integrated into clinical practice, conferring disease progression PDX models an added value. Lastly, it is important to highlight the need for tumor models of DCIS, an early-stage disease to enhance our understanding of how to identify which DCIS may progress and target the disease prior to progression. Also, the use of matched primary tumors, circulating tumors cells and metastatic disease, will also contribute to identifying metastases initiating cells.

The use of PDXs, certainly comes with some limitations. PDXs are engrafted in immunocompromised mice that over time exchange human stroma with murine and lacking a mouse functional immune system. The crosstalk between the immune microenvironment and tumor cells can't be further explored in these models. Nevertheless, some immunocompromised models, retain functional immune cells or can engraft donor human cells to reconstitute their environment (195,196). Moreover, the patient stroma is normally replaced by murine derived stroma. This may not accurately reflect cell-to-cell interactions as well as cell-to-tumor microenvironment interactions observed in the patient (197). Ligand-receptor interactions and signaling pathways that play a role in tumorigenesis might be affected by the murine-to-human tumor microenvironment, where for example murine hepatocyte growth factor (HGF) displays lower affinity for the human Met receptor tyrosine kinase than does human HGF (198). In addition, the increasing interest and successful use of immunotherapy has motivated the development of humanized mouse strains. Moreover, it has been shown by our group that not only the immune infiltration matters, but also the spatial localization of immune cells within the stroma and the tumor (114).

Taken together, PDX are valuable preclinical models for tumor marker discovery, drug development and mechanistic understanding of key clinical questions. Advances in computational approaches, artificial intelligence, and sequencing technologies, have made the manipulation of patient-derived cells highly used for drug screening and interrogation of distinct tumor populations.

3.3 Single-cell RNA sequencing as a technology to deliver patient-centered care

Single-cell technologies performed on dissociated tumor cells have provided a better understanding of the genomic landscape and heterogeneity of tumors. Moreover, cells display level distinct or same genomic alterations can display different transcriptional profiles exhibiting distinct phenotypic heterogeneity that can drive tumor initiation, tumor growth and metastatic disease. Single-cell technologies, where tumors are dissociated into single cells, have allowed us to have a deeper understanding among a same or different cell type.

The main limitation of single-cell sequencing has been intended to be overcome by profiling distinct cell types within tumors (epithelial, fibroblasts, endothelial or immune cells) by immunohistochemistry using well-known markers (114,199–201) or by performing prospective spatial transcriptomics, initially by isolating cells from tumor domains (190) or then as high resolution as phenotypic spatial transcriptomics(168,202). However, how single tumor cell transcriptomic phenotypes correlate with and are modulated by the spatial tumor architecture is still poorly understood.

Spatial heterogeneity is apparent in many tumor types. Morphological differences of cancer cells can be readily observed in formalin-fixed paraffin embedded samples of human tumors, without the need of specific markers (153). Moreover, phenotypic diversity of TNBC cancer cells in different regions of a sample poses challenges in pathologic assessment of markers used in clinical decision-making (32,166). Therapy does not integrate the spatial context of tumor cells, their abundance, nor the networks to the local environment where they reside, which can influence transcriptomic, metabolic and epigenetic variability among these cells (167).

This in fact, could even affect how drug treatment delivery is performed. If cells that are far away from the vasculature and near necrotic regions in the tumor, are intended to be treated with a targeted therapy, by knowing the spatial localization, it would make sense always to first debulk the tumor of cycling cells and later give sequential targeted therapy to for instance, target hypoxic populations.

Fig.3.3

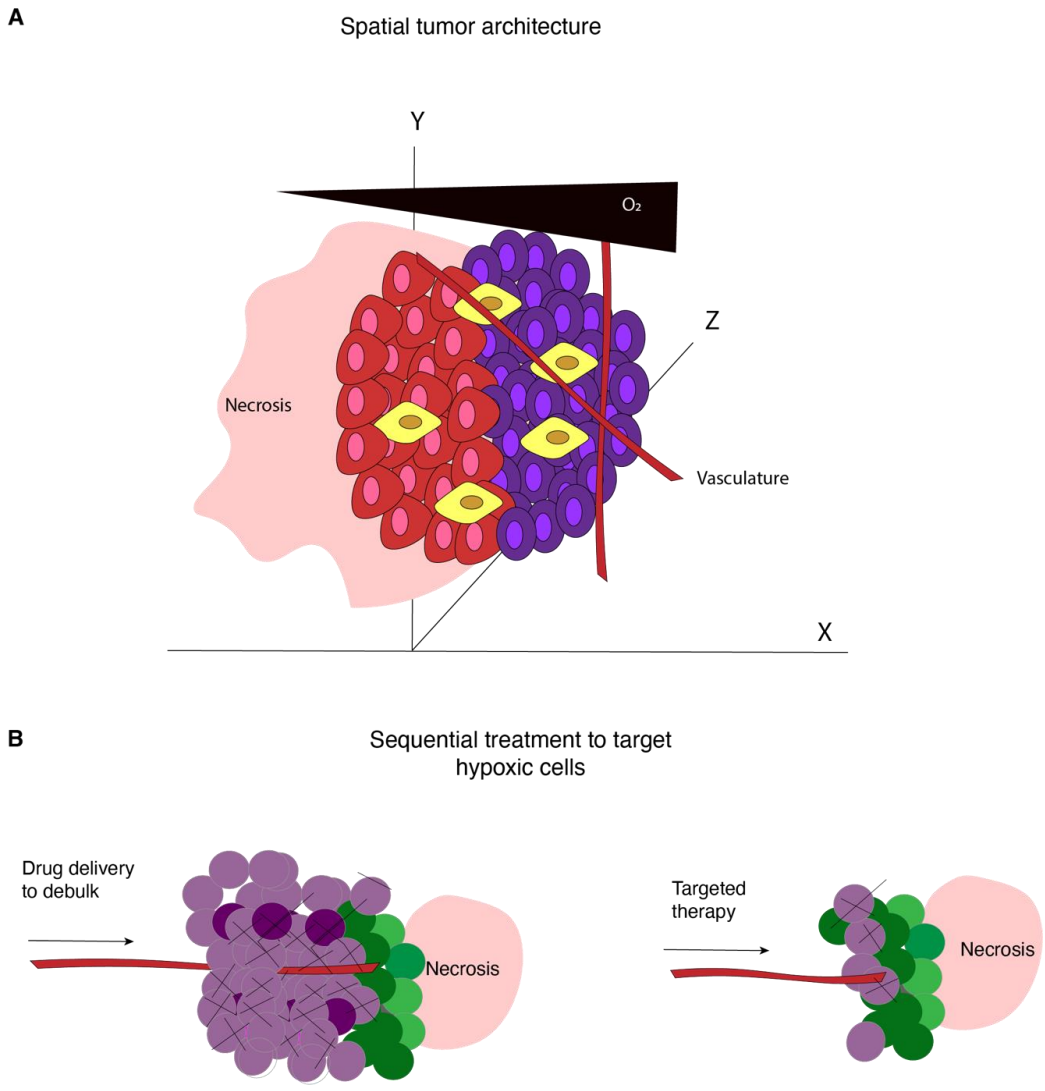


Fig. 3.3. Spatial tumor architecture and drug delivery approach.

A. The tumor is a 3D structure composed by tumor cells (in red and purple) and elements of the tumoral environment (necrosis, blood vessels, oxygen level etc.). The disposition of cells regarding the distinct environmental queues, displays hypoxic tumor cells near necrotic regions and cycling cells further away from necrosis and closer to vasculature. **B.** The information obtained from the spatial localization of cells, proposes a sequential approach of drug delivery. First, a debulking treatment for cycling cells and later, after cycling cells are dead, the delivery of targeted therapy for hypoxic cells.

Some of the caveats that single-cell technologies presents is the high cost of single cell sequencing to perform experiments with multiple technical or biological replicates. The monetary implication of this is the selection of samples tailored to answer specific biological questions (203). Therefore, the interrogation of larger datasets is limited. Hence now the extrapolation of single-cell technology data to the clinic is limited. The biological findings identified in our studies, however, raise opportunities for validation in other cohorts through multiplexed immunofluorescence of potential biomarkers discovered in this study. An extended analysis has confirmed the anticorrelation between hypoxic CAIX positive zones and proliferative KIAA0101 markers. Our findings were demonstrated in PDX replicates then compared to the matching human data provided by sister technologies, such as digital spatial profiling. Although these studies were focused on only one patient with longitudinal primary and metastatic samples, our data support that PDXs are viable models to study tumor progression and supports the need of further studies to examine additional PDXs or preclinical models. In this context, other bio-

banked triple-negative breast cancer PDXs serve as surrogate models (159). These models are available not only for TNBC, but also rare breast cancer subtypes, where further interrogation of single-cell populations and the use of successful techniques identified in this work can be used to interrogate patients longitudinally. Hence the development of the single-cell atlas trying to mimic the TCGA or multi-center cross platforms, will likely provide us with a powerful approach with a focus to identify elusive tumor initiating cells as important therapeutic targets yet a poorly understood and heterogeneous population in TNBC. The ability to integrate data from different groups and expand findings into larger datasets (204) will be crucial to this understanding. Moreover, the rapid advancement in technologies will rapidly make this and other technologies more accessible (205).

3.4 Spatial heterogeneity in triple-negative breast cancer.

In chapter 3, we introduce a novel unbiased and clinically applicable in situ approach to spatially map single-cell RNA sequencing populations in PDX and patient tissue sections in primary tumors and metastatic disease. We integrate scRNA-seq, multiplex immunofluorescence, and a computational analysis to integrate the expression of selected markers in each individual cell in the tissue to define scRNA-seq populations. This approach integrates transcriptional and spatial information from a same cell type of a specific tumor, where importantly PDX used in replicates, retains not only the histology, genomic and transcriptomic patient ‘s landscape, but also the regional heterogeneity, conferring additional information that PDXs can be a relevant pre-clinical tool to explore spatial heterogeneity. Other single-cell studies have performed cell type profiling with well known-markers or spatial transcriptomics ablating tumor regions and performing transcriptomic analyses. Relevant poor outcome subtypes have been identified within TNBCs through spatial transcriptomics. These include a hypoxic TNBC subtype, indeed characterized by CAIX expression, was correlated with poor outcome and (168). Moreover the use of DSP for pre and post treatment BC samples was recently used, being able to identify in HER2 positive BC patients, biomarkers of HER2-targeted therapy sensitivity (206). The use of single-cell sequencing technologies, digital spatial profiling and spatial transcriptomics brings promising avenues for an integrated view of the distinct levels of heterogeneity in the tumor. This certainly, highlights the complexity of tumor biology and the need of a patient -centered approach. The costs of technologies Is the main limitation for personalized treatment, but in hopes of the advancement in technology this seems likely a promising strategy to deliver care in the near future.

Fig.3.4

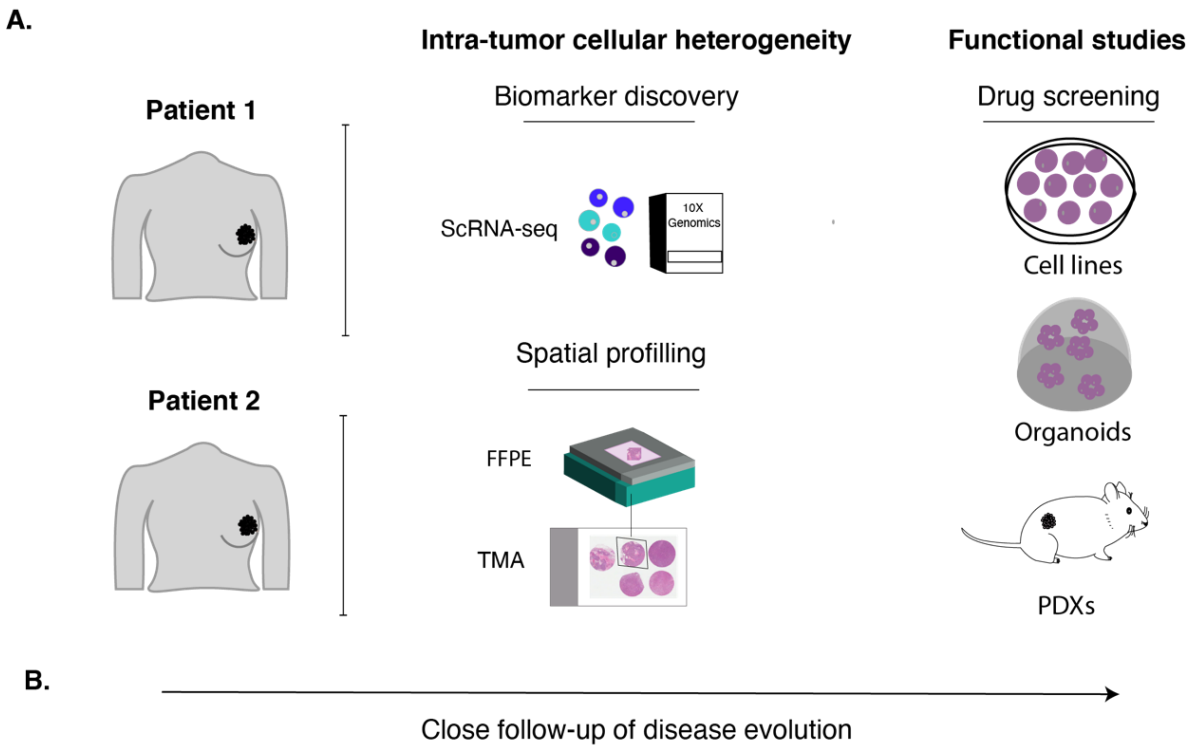


Fig.3.4. Personalized approach to integrate spatial and temporal heterogeneity.

A. Individualized treatment should be done for each patient (patient 1 and patient 2). The use of single-cell sequencing technologies should be ideally coupled with spatial profiling methods to identify potential clinical biomarkers. B. The use of cell lines, organoids, and preclinical models such as PDXs can be used in a personalized manner for drug screening. **B.** Each of these steps should be done throughout the evolution of disease. For instance, new samples from recurrence or metastatic disease should undergo the same process depicted in A.

3.5 Spatial architecture and temporal heterogeneity of transcriptomic populations in primary and metastatic disease

The interrogation of the spatial localization of tumor cells populations within the tumor revealed preserved zonation of certain biological clusters in both the primary and metastatic disease. We were able to demonstrate that the spatial heterogeneity is preserved in the patient's primary tumor. Moreover, we also showed that the metastatic samples also have a conserved spatial architecture for spatially zoned populations identified by scRNA-seq. The combination single-cell RNA sequencing, multiplex immunofluorescence (MIF) and Digital Spatial Profiling (NanoString) to outline matched patient-derived primary and metastasis xenograft models and define transcriptional states in the context of intact tissues. We identify a set of cell populations present across samples that are predominantly determined by their spatial localization within the tumor. Both primary and metastatic tumors contain transcriptionally distinct hypoxic cell populations that show a gradual transition towards a cycling state, largely defined by their distance to necrotic zones. A second class of cell populations, in contrast, show no preference for spatial localization and display a low degree of network connectivity, interspersed among other cell states within the tissue. These cells are exposed to distinct environmental queues in each zone and tend to co-express a range of transcriptional programs, indicating a high degree of plasticity. In functional studies, both hypoxic and cholesterol homeostasis cell populations displayed higher tumor initiating potential *in vivo and* are enhanced in tumor metastases respectively. The metabolic nature of the populations identified, requires further investigation of the metabolic profiles and adaptations that these cells, which display plasticity, undergo in the primary tumor, in

circulation and upon arrival to the metastatic niche. The metabolic changes that these cells undergo, possibly enable them to successfully survive and repopulate the tumor at a distant location. The cholesterol signatures and targeting reducing metastatic disease has been explained by other groups. Cholesterol metabolism plays a role in many biological functions. In cancer, ER stress and hypoxia are known to stimulate SREBP1 and SREBP2, which leads to cholesterol biosynthesis programs (207,208) This program has been proven to be involved in TNBC and other cancers progression (209). Metabolomics or metabolic spatial profiling is what would allow a better characterization of the aggressive populations in this tumor. This would like to allow a better understanding of the mechanisms behind metastatic dissemination. Moreover, epigenetic changes at the single-cell level should serve as a powerful tool to understand transient cellular adaptations upon distinct biological processes, such as tumor initiation, metastatic dissemination, and therapy resistance (210). For the understanding of dynamic changes in distinct cell populations the integration of live cell imaging or cell tracing is key to be able to characterize temporal phenotypic events (211). The technologies used here captures a snapshot of cells already adapted to a certain program and phenotype, and although they bring value about key transcriptional programs required for tumor initiation and metastatic dissemination, knowledge of the intermediate adaptations required by these cells and the changes at the metabolic or epigenetic level are in need to have a better mechanistic understanding (212).

Fig.3.5

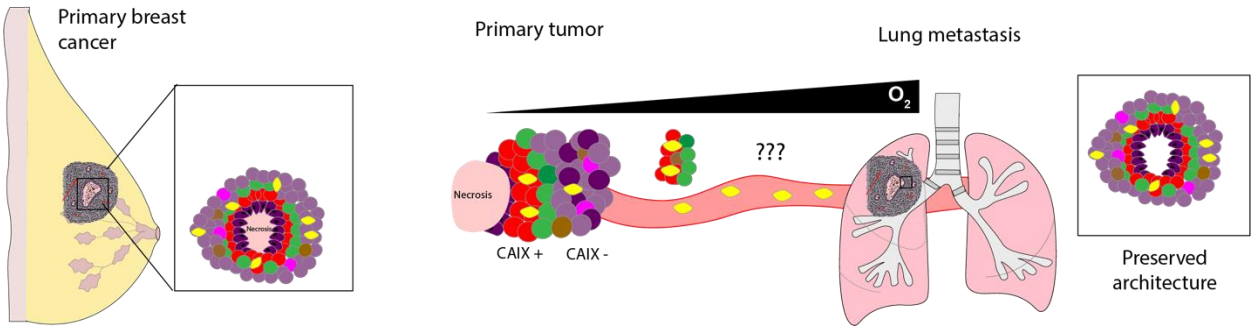


Fig.3.5. Transcriptomic populations identified in lung metastatic disease preserved their spatial zonation. The main biological populations identified in the primary tumor are also identified in the lung metastasis disease. These populations preserved their spatial localization. Hypoxic populations were identified close to necrotic regions and by using CAIX as a marker. The sequential mapping of scRNA-seq populations allows the identification of more aggressive populations in the primary tumor. The identification of metastatic initiating cells is still not elucidated and remains a key question.

3.6 Spatial cartography of primary and metastatic samples identifies aggressive populations

We describe two distinct tumor zones found in the PDX and matched patient sample. One hypoxic and other cyclic tumor zone spatially defined by their distance from necrosis. The topology of these cells found in one patient was further interrogated in other matched PDXs. The interactions of cancer cells with distinct environmental ques gives additional tissue context information and a better understanding of how these cells found in specific niches present different transcriptional profiles. Moreover, when analyzing “cholesterol homeostasis” patient’s outcomes, we observed that hypoxic signatures from scRNA-seq clusters are correlated with an increased DMFS within the first five years of diagnosis. When performing sphere-forming efficiency assays in 3 vs 20% oxygen we identified cells positive for the markers of a basal-mesenchymal-like population in both conditions, revealing an important level of plasticity of these cells. When tested in vivo the tumor initiating capacity of CAIX positive cells and CAIX cells selected from tumorspheres in 3% oxygen where most cells are positive for PJ5 markers, we show that cells with CAIX positivity when compared to cells that are CAIX negative have a higher tumor initiating capacity in vivo. Our data provides information that zone-less cell populations can acquire specific traits when residing individually in distinct tumor zones, and once survived or adopted hypoxic features they have a more aggressive phenotype. Lastly, the development of a spontaneous lung metastasis model mimics the patient’s disease progression. This model coupled with the MFP lunge metastases, allowed the understanding in space and time of aggressive cell populations found in the primary tumor and selected upon disease progression.

3.7 Conclusions.

Single-cell RNA sequencing is a growing technology started to be more used in the clinic. Translational and applicable methods within distinct samples across larger datasets are needed for further data extrapolation. We presented a spatial transcriptomic approach coupled with the discovery of distinct cell populations tested biologically. We demonstrated that spatial heterogeneity is not only preserved in PDX and the patient's primary tumor, but also upon disease progression. Moreover, this was further validated with a different methodological approach using digital spatial profiling. This methodology allowed to interrogate the geographic localization of single cell gene-expression defined populations in situ and determine transcriptomic populations with enhanced tumor initiating and metastatic capacity by performing functional tests. Overall, the integration of spatial information, single-cell transcriptomics and biological testing of single-cell RNA sequencing populations is a powerful tool to biologically interrogate intra-tumor heterogeneity with a high level of integration among tissue sections that can be clinically relevant among disease progression. This work provides the proof of principle that PDXs are a useful pre-clinical model able to recapitulate spatial and temporal heterogeneity from the patient. This allows researchers not only to do bioinformatic analyses on tissue, but also to functionally test and biologically interrogate transcriptomic populations. Although he one patient approach is the main limitation of this study; this paves the way for similar studies and provides new understanding of the plasticity of TIC. Certainly, in the upcoming years a single-cell RNA sequencing data repository will provide with easily accessible and larger datasets to be able to extrapolate findings and be able to tackle the bigger picture, rather than only providing data with a patient-centered approach.

4. EXPERIMENTAL PROCEDURES

4.1 Patient-derived xenografts

PDXs were developed in accordance with the McGill University Health Center research ethics board (SUR-99-780) and the McGill University Animal Care Committee (2014-7514) guidelines. Studies were performed in *NOD.Cg-prkdc^{scid}Il12rg^{tm1wjl}/SzJ (NSG*) mice (Jax). We established a patient-derived xenograft (PDX) from a patient's primary TNBC tumor (159). Tumors were measured by calipers twice weekly by palpation), by an animal technician blinded to experimental detail. Tumors were harvested between 200 and 600 mm^3 depending on experiment.

4.2 Tumor dissociation into single cells

Single-cell suspensions were generated by mechanical and enzymatic dissociation. Murine stromal cells were removed using a Mouse Cell Depletion Kit (Miltenyi). Mouse-depleted single-cell suspensions were washed two times in PBS and submitted for single-cell RNA sequencing. Viability was assessed prior to sequencing by a LIVE/DEAD viability testing (Thermo Fisher Scientific).

4.3 Multiplex Immunofluorescence (MIF)

Primary antibodies were first optimized by monoplex staining. Slides were deparaffinized in xylenes, re-hydrated in ethanol followed by antigen retrieval in boiling 10 mM citrate buffer (pH 6.0). Slides are fixed in NBF for 20 minutes followed by distilled water wash. Slides are rinsed with AR6 or AR9 solution. Microwave treatment (MWT) is applied to slides and allowed to cool down at room temperature (RT) for 15 minutes. Slides are rinsed with TBST and then stored in blocking buffer at 4C. Blocking solution is removed

and primary antibody solution is applied to the tissue. Slides are washed with 1x TBST 3 times, 2 minutes each. Opal polymer HRP secondary antibody solution is applied to the tissue and incubated for 10min at RT. Slides are washed with 1x TBST 3 times, 2 minutes each. Slides are rinsed with AR9. A new cycle is repeated by MWT as previously described for each antibody. At the end, slides are incubated in DAPI solution for five minutes at RT and then washed in TBST for 2 minutes followed by 2 minutes of distilled water.

Table 3: Antibodies used for experiments.

Antibody (clone)	**Company (Cat#)**	**Dilution**
GLUT1 Rabbit Polyclonal Antibody	Ventana	Pre-diluted
Carbonic Anhydrase IX (CA IX) (EP161) Rabbit Monoclonal Antibody	Esbe scientific / CMQ 379R16	1:40
Anti-CALML5	Abcam-ab122665	1:400
UCHL1 (D3T2E) XP® Rabbit mAb	CST - # 13179S	1/200
Topoisomerase IIa (D10G9)	CST #12286	1/50
CRYAB antibody (clone 1D11C6E6)	Byorbit orb97538	1/100
PAF15-KIAA0101 (D8E2Y) XP Rabbit mAb	CST - #81533	1/100
Ki67	Ventana (790-4286)	Pre-diluted (IHC)
ER (SP1)	Ventana (790-4342)	Pre-diluted (IHC)
HER2 (4B5)	Ventana (790-2291)	Pre-diluted (IHC)
Pan-KRT	Ventana	Pre
CD31mouse (Rat)	Dianona-DIA-310	1/40
CD34	Ventana- 7902927	Pre-diluted
P53	Ventana-7902912	

4.4 Imaging by confocal microscopy

Eight color multiplex immunofluorescence-stained tumor sections were imaged using an LSM710 confocal microscope (Zeiss). Zen software (Zeiss) was used to obtained images in high quality to be further computationally analyzed. A minimum of three distinct tumor passages in primary and lung metastatic PDXs were selected and within each tumor three distinct regions were selected to be imaged. In the human section ten distinct regions of the primary tumor were selected for imaging. The spontaneous lung metastasis in PDX had two replicates and the lung human biopsy section was a whole single image.

4.5 Tumorspheres formation assay

Primary tumor freshly dissociated was used to generate single-cell suspensions that were seeded in concentration of 100.000 cells per well on six 6 well ultra-low attachment plates (Corning) containing sphere medium (DMEM/F-12 GlutaMAX (Gibco), 1x B27 (Gibco), 20 ng/ml human EGF (BPS Bioscience), 10 µg/ml insulin (Gibco), 0.5 mg/ ml hydrocortisone (Wisent), 20 ng/ml bFGF (StemRD), 10 µg/ml heparin (STEMCELL), 50 µg/ml gentamicin (Gibco)) containing 1% methylcellulose at 37°C and 5% CO_2 . Spheres were cultured in either 20% O_2 or 3% O_2 for 7 days.

4.6 Histogel formalin fixed paraffin embedded tumorspheres and organoids

Tumorspheres cultured for 7 days in low oxygen (3%), were centrifuge at 1200rpm for 4 min. Supernatant was discarded and spheres were resuspended in 100ul of Matrigel. Spheres in Matrigel were seeded as a drop in an individual well of a 12 well plate and left for 10 min at room temperature until drop solidified, 800 µl of 4% PFA was added for 2

hours at room temperature in a fume hood. In the meantime, a 150ul of Histogel previously warmed to 65C was added into a cryomold (10x10x5mm) and gently spread. Cryomold was placed on ice for 10min. Sample in Matrigel after 2 hours is stained with Hematoxylin for 10min at RT and then washed 3 times with ddh20 and carefully scrapped with a blade and transferred to the cryomold. Additional 150ul of Histogel was added on top Matrigel embedded sample on the cryomold and cooled on ice for 10min. The histogel block was transferred to a tissue cassette embedded in 10% formalin for 24 hrs. and later transferred to 70% ethanol.

4.7 Fluorescent-activated cell sorted (FACS)

Single cell suspension from tumorspheres cultured for 10 days in 3%O2 was generated by enzymatic dissociation with Accutase. Single-cells were resuspended in FACS buffer (PBS, 2% FBS, 10 mM HEPES), stained with CAIX-APC conjugated antibody (BD Biosciences) for 30 min on ice, washed and viability stained with 7-AAD (eBioscience) was performed. CAIX positive and CAIX negative cells were sorted and used for in vivo experiments.

4.8 Tumor formation and extreme limiting dilution analysis in vivo.

CAIX positive and CAIX negative sorted cells were injected bilaterally in the fourth mammary fat pad. Three distinct dilutions of 10000, 1000 and 100 cells were selected and injected in 6 mice per group. The right mammary fat pad was injected with CAIX positive sorted cells and the left mammary fat pad with CAIX negative cells. Tumors were monitored and measured twice a week by a blinded animal technician. When the tumors reached endpoint ~500mm^3, the tumor was resected and collected to generate histopathological and frozen material.

CAIX positive and CAIX negative sorted cells were injected bilaterally in the fourth mammary fat pad. Three distinct dilutions of 10000, 1000 and 100 cells were selected and injected in 6 mice per group. The right mammary fat pad was injected with CAIX positive sorted cells and the left mammary fat pad with CAIX negative cells. Tumors were monitored and measured twice a week by a blinded animal technician. When the tumors reached endpoint ~500mm^3, the tumor was resected and collected to generate histopathological and frozen material.

4.9 In vivo treatments

All human participants provided informed consent for this study. The tissue was collected at the McGill University Health Centre. The used of established protocols approved by the research ethics board (SUR-99-780) were used. The McGill University Animal Care Committee (2014-7514) approved all experiments, which were conducted in NOD scid gamma (NSG) mice from The Jackson Laboratory [Jax]. In vivo studies were randomized in cohorts of 10 mice per arm. Cells in different dilutions (up to 10.000 cells) or tumor fragments of 1-2mm^3 were engrafted in the fourth mammary fat pad of 6-8 weeks old female mice. SLC-0111 (CAIX inhibitor) was administrated through oral gavage once daily at 10 µl/gram per mouse (in 100ul volume) for 20-30 days followed by a rest for 2 days every 15 days. The vehicle is 0.5% Carboxymethylcellulose and 0.1% Tween 80. Ro 48-8071 was administrated through oral gavage once daily at 30 mg/kg per mouse (in 50ul volume). Gemcitabine was administered intravenously twice a week at a concentration of 100mg/kg. All mice were monitored by a blinded animal technician twice a week. Tumor volume was calculated by ((smaller tumor dimension2 x largest tumor

dimension)/2). Change in tumor volume was calculated by ((Final volume – Initial volume) / initial volume).

4.10 In vivo metastasis assessment

PDX GCRC 1915 (PDX-Pri) was subjected to a mammary fat pad resection once the tumor reached endpoint (>500mm^3). Biweekly monitoring of signs and symptoms was metastatic disease performed by animal technicians, for at least 6 months after resection. The 6 months follow up time was determined upon previous experiments that confirmed metastatic disease at 6 months, mimicking the patient's onset of lung metastatic disease after surgical resection of the primary tumor.

4.11 Bulk RNA and WGS sequencing

The tissue from GCRC1915 and GCRC2076 was taken from snap frozen material. Total DNA and RNA was isolated from adjacent sections using the AllPrep DNA/RNA Mini Kit (Qiagen). For GCRC1915 Germline DNA was derived from buffy coat and extracted using the DNA Blood Maxi Kit (Qiagen). DNA was quantified using the Qubit fluorometer. RNA was quantified by NanoDrop and integrity was evaluated with Bioanalyzer 2100 (Agilent) (159).

4.12 Statistical analysis for in vivo experiments

Prism 7 (GraphPad) was used for basic statistical analysis, for survival analysis using the log rank (Mantel-Cox) test. Data represent mean ± SEM; p value < 0.05 was considered significant.

4.13 Single-cell RNA sequencing data pre-processing and quality control.

Reads from the single-cell RNA sequencing libraries were trimmed, demultiplexed, aligned to the reference genome (cellranger hg19 and mm10 build version 1.2.0), and unique transcripts were counted using the 10X Genomics cellranger pipeline (versions listed in Extended Data Table 1). Species assignment was performed using cellranger, and only human cells were retained for further analysis. Quality control of expression matrices, clustering and visualization was performed in R with methods from the Seurat package(213,214) (v.2.3.4). Genes with expression in fewer than 3 cells and cells with fewer than 200 detected genes were excluded. Low quality cells and multiplets were filtered out based on the following metrics: number of detected genes, number of UMIs, and percent of mitochondrial transcripts. Filtering thresholds are summarized in Extended Data Table 4. A much higher number of cells were captured for sample Lm2 than the other samples, therefore the top 10,000 cells based on number of UMIs were retained for downstream analysis. Following filtering, each sample was normalized by dividing the UMI counts for each cell by the total counts in that cell, then the counts were multiplied by 10,000 and log transformed.

Table 4: Single-cell RNA sequencing metrics.

Sample	Min. # detected genes	Max. # detected genes	Min. # UMIs	Max. # UMIs	Max. % mitochondrial	Cellranger version
Pri1	1,000	8,000	2,500	75,000	10	2.0.1
Pri2	1,000	6,000	2,500	40,000	10	2.1.1
Lm1	1,000	6,000	2,500	40,000	10	2.1.1
Lm2	1,000	6,000	2,500	40,000	10	2.2.0
Slm	1,000	6,000	2,500	40,000	10	2.2.0

4.14 Primary PDX joint analysis CCA and Harmony.

Primary PDX samples Pri1 and Pri2 were integrated to overcome batch effects using two methods: canonical correlation analysis (CCA) implemented with the Seurat package(213,214) (v.2.3.4), and Harmony integration (v.1.0)(173) implemented with the Seurat package (v.3.1.5)(213,214). The data was scaled to regress out the effects of technical confounders (number of UMIs detected and mitochondrial content). Highly variable genes were computed for each sample separately as genes with average expression ≥ 3 or ≤ 0.0125 and dispersion ≥ 0.5. Variable genes were ranked by dispersion, and the union of the top 2,000 variable genes for each sample (2,502 genes) was used as input to integration with the scaled data. For CCA integration, 100 canonical correlation vectors (CCs) were computed, and the first 47 were aligned. The 47 aligned CCs were used as input for visualization in two dimensions with t-SNE(215) and UMAP(216). Cells were clustered by constructing a Shared Nearest Neighbor (SNN) graph from the 47 aligned CCs and applying the Louvain algorithm for community detection, with the resolution parameter set to 0.6. For Harmony integration, 100 principal components (PCs) were computed, and the top 47 PCs based on variance explained were aligned by sample with the Harmony method(173). The “harmonized” PCs were used as input for visualization in two dimensions with t-SNE(215) and UMAP(216). Cells were clustered by constructing a Shared Nearest Neighbor (SNN) graph from the “harmonized” PCs and applying the Louvain algorithm for community detection, with the resolution parameter set to 0.6.

4.15 Post-clustering quality control.

Some transcriptional clusters were excluded from subsequent analyses based on quality control assessment. In the CCA integration of the primary PDX samples, the smallest cluster, PJ9, was excluded as its median number of UMIs and median number of genes was more than twice that of the other clusters, suggesting it consists of multiplets. Other clusters were flagged based on whether they consisted of cells from primarily one sample, such as the cluster PJ8.

4.16 Cell cycle regression and assessment.

To assess if cell cycle phase drives the clustering analysis of PDX-Pri, we compared the cluster identity of cells in the CCA integration to this analysis repeated with cell cycle regression. For this, we computed cell cycle phase scores with a linear regression approach implemented in the Seurat package(213,214) (v.2.3.4) using cell cycle gene signatures(163,217). The resulting cell cycle phase scores were regressed out, along with number of UMIs detected and mitochondrial content, at the step of data scaling described above. Computing CCs, clustering, and visualization was then performed as described above. To quantify the dispersion of cell clusters in the UMAP plots with and without cell cycle regression, we calculated the geometric median absolute deviation (MAD), i.e. $MAD = \sqrt{median(|X_i - \tilde{X}|)^2 + median(|Y_i - \tilde{Y}|)^2}$. The difference in MAD per cluster with and without cell cycle regression, as well as visual inspection of the UMAP plots, was used to assess the effect of cell cycle on clustering.

4.17 Characterization of clusters with gene set enrichment.

In order to assess gene pathway activation in transcriptional clusters, we performed single-sample gene set enrichment analysis (ssGSEA)(218) as described(219,220). We scored enrichment of the MSigDB(221) Hallmark, Canonical Pathway, and KEGG(222) signature collections. Only genes detected in the dataset were considered for analysis. For each gene pathway, ssGSEA scores were computed per cell cluster using mean cluster expression, and z-scores were computed across clusters. The 2 pathways with the highest z-scores for each cluster were selected as defining pathways for that cluster. To determine the leading-edge genes of each pathway (i.e., the genes that contributed the most to the enrichment score), we computed a Kolmogorov-Smirnov (K-S) running sum statistic. Moving down the list of genes ranked by mean cluster expression, the running sum increases by a weighted step if the gene is in the given pathway or decreases by a fixed-size step when the gene is not. The leading-edge subset is the subset of genes that achieve the maximum running sum statistic.

4.18 Transcription factor activity inference.

The activity of transcription factors (TFs) and their regulated genes was inferred from the normalized single-cell expression data with the python implementation of SCENIC(223) (v.0.9.19). Briefly, modules of genes coexpressed with TFs were detected through reconstructing a gene regulatory network with the GRNboost2 method implemented with the arboreto package (v.0.1.5). TFs with binding motifs enriched in their corresponding modules were retained, and genes containing the binding motifs as potential direct targets were retained in the modules. The AUCell algorithm (223) was then used to compute an

activity score for each TF's module in each cell. Top active and specific TFs for PDX-Pri clusters grouped based on their biology were selected as follows. For this analysis, cells in PJ5 and PJ8, PJ3 and PJ6, and PJ1 and PJ2 were grouped, while PJ7 was analyzed alone. TF activity scores were averaged in each group. The 50 most active TFs for a given group were selected based on mean activity scores. The fold change of mean activity in the given group and mean activity outside the group was computed. The 50 TFs with the largest fold changes were selected as most specific. The intersection of the most active and most specific TFs was selected for each group.

4.19 Cluster marker selection for spatial mapping.

To identify potential protein markers for each PDX-Pri cluster, gene markers were computed for each cluster compared to all other cells using the Wilcoxon rank sum test, requiring that a gene is detected in at least 25% of cells in the cluster or outside the cluster to be tested. The resulting cluster markers were sorted using adjusted p-value to select the top marker genes as candidates for antibody selection. When p-value did not distinguish genes, they were sorted with average log-fold change in expression and the difference between the proportion of cells expressing the gene within the cluster (pct.1) and the proportion of cells expressing outside the cluster (pct.2).

4.20 Spatial scRNA-seq population assignment

We classified the cells that were imaged using IF into classes defined by the clusters we found from single-cell sequencing. For this, we used the genes we have chosen as markers and assumed that the distribution of mRNA molecules measured by sequencing could be mapped to the immunofluorescence intensity observed in images. Hence, we

assume that there is a linear correspondence between the signals of each one of the seven fluorescence channels and the transcription levels observed in sequencing. The cell assignment to clusters was performed using the histograms that describe RNA levels. For each of the 7 seven genes we scaled the expression in the interval [0, 1] and normalized the area under the curve to obtain probability density functions. Similarly, the seven images corresponding to each tissue sections were normalized so that the median cell intensity varied from 0 to 1 in every channel. Thus, all cells in an image are characterized by 7 intensity values that we used to find the best match to one of the available clusters. For example, cells with high intensities in the CAIX channel, mid intensities of CALML5 and low CRYAB, have a high probability of belonging to a cluster where the probability density functions of these genes agree with such expression profile. Hence, we computed the probability that a given cell belongs to each of the possible clusters to make the final assignment. Our large mosaic images of tissue sections contain approximately 20,000 cells each and span different areas of the samples. In total we analyzed around 10 locations per sample, accounting for 200,000 cells. We mapped cells to clusters and studied their spatial distribution to understand their arrangement and interactions in the physiological context.

4.21 Joint analysis of samples within and across conditions.

Samples were joined in several combinations to directly compare expression profiles, as performed for PDX-Pri above, using the Seurat package (v.3.1.5)(213,214). The following joint analyses were performed: mammary fat pad metastatic PDX samples (Lm1 and Lm2), mammary fat pad and lung metastatic samples (Lm1, Lm2, and Slm), and all five

samples from primary and metastatic conditions together. For each analysis of samples across conditions, we confirmed that the structure of populations identified from analysis of conditions separately was conserved to support that the cells were not artificially joined in any cluster. In each iteration, data was scaled to regress out the effect of number of UMIs detected and mitochondrial content. Highly variable genes were computed for each sample separately as genes with average expression ≥ 3 or ≤ 0.0125 and dispersion ≥ 0.5. The variable genes were ranked by dispersion, and the union of the top 2,000 variable genes for each sample were selected as features for downstream analysis. When joining more than 2 samples, the features were also required to be variable in at least 2 samples and detected in the data of all samples. These features were used as input for principal component analysis, where 100 PCs were computed. The top 30 PCs based on variance explained were selected as input for clustering and visualization using Harmony integration(173), with integration performed over the “sample” variable. The 30 “harmonized” PCs were used for visualization in two dimensions with t-SNE(215) and UMAP(216), as well as to inform clustering. Cells were clustered using a shared nearest neighbor modularity optimization-based clustering algorithm. First, a KNN graph of cells is construction from Euclidean distance in the “harmonized” PCA space, which then informs the construction of a graph from Jaccard similarity of any two cells (FindNeighbors function, nn.eps = 0.5). Clustering was then performed on this graph using the Louvain algorithm for community detection (FindClusters function, n.start = 10, random.seed = 100, resolution = 0.5). Cluster markers were computed for each cluster compared to all other cells using the Wilcoxon rank sum test, requiring that a gene is detected in at least 25% of cells in the cluster or outside the cluster to be tested.

4.22 Differential expression between conditions.

Differentially expressed genes between conditions in the joint analyses, such as between mammary fat pad and lung environments with the metastatic sample joint analysis, were computed per cluster using the Wilcoxon Rank Sum Test. Cells were first randomly down-sampled to yield the same number of cells for each condition. Only genes detected in at least 1% of cells in the cluster or outside the cluster were tested, and a minimum log-fold change of 0.1 was required to be reported as a condition marker.

4.23 Stress response effect.

As our single-cell experiments were conducted under 37°C, we evaluated a potential stress response induced by the dissociation method. Using the top 40 genes obtained from a published differential expression analysis between cells digested under 6°C and under 37°C(224), we calculated the average expression of these genes and added it as an additional factor to regress from the data in the clustering workflow described above. Nearly all clusters retained their relative localization, with only the spontaneous lung metastasis showing clusters from the non-regressed analysis mixing following regression of the stress response genes.

4.24 Cell state trajectory inference.

We inferred cell state trajectories for PDX-Pri with Monocle 2 (v.2.10.1)(183,225). Due to the continuous nature of the clusters, we explored the relationship of representative cells of the PJ clusters, selected as described above. After normalization and variance estimation, a smooth function to model dispersion and mean expression of genes was computed. Ordering genes were defined as having mean expression of at least 0.1,

empirical dispersion of at least 1, and expressed in at least 10 cells. These genes informed the ordering of cells with the Discriminative Dimensionality Reduction with Trees (DDRTree)(225) manifold learning algorithm. The root state of the trajectory for directionality was manually set. Differentially expressed genes across pseudotime were computed with the differential GeneTest function, which compares the performance of modeling each gene's expression when pseudotime is known to when it is not, with the transcriptional clusters specified with the fullModelFormulaStr parameter.

4.25 Digital spatial profiling, Nanostring.

The NanoString digital spatial profiling technology (NanoString, Seattle, WA, USA) allows the detection of multiplexed molecules from the surface of FFPE tissue using a single molecule optical barcoding. This allows the detection and quantitation of proteins and spatially digital characterization of previously defined regions of the interest (ROI). The specific barcodes in the tissue are attached via a UV-cleavable linker to either the primary antibodies or nucleic acid probes that are liberated by a UV laser from user-selected ROIs. These are counted by using an nCounter platform. All counts are spatially mapped to the tissue, allowing spatial designation of target abundance at the resolution of the defined ROI. We used 4um unstained slides and TMA's sections and stained them with a nuclear antibody, pan-cytokeratin, CAIX, and SMA, both to visualize the overall tissue morphology and guide the ROI selection.

4.26 Digital spatial profiling differentially expressed genes

Normalization and differential gene expression analysis were conducted using a negative binomial distribution with DESeq2 (R/Bioconductor)(226,227). The Benjamini-Hochberg method was used to correct for multiple testing, with an adjusted *p*-value of < 0.01 as a threshold to identify differentially expressed genes (DEGs)(228).

4.27 Spatial Deconvolution of cell populations Identified by scRNA-seq.

Quantification of the cell populations identified by single cell RNA sequencing within the regions of spatially resolved gene expression data was performed using SpatialDecon (R/Bioconductor) (229). Prior to deconvolution, the spatially resolved gene expression data was normalized to the third quartile (Q3). A matrix of cell profiles was derived from the mean expression profile of each cell population identified by single cell RNA sequencing. Similarly, a matrix of expected background for all data points in the normalized spatially resolved data matrix was derived from negative control probes included in the Cancer Transcriptome Atlas probe set and was used to establish the level of technical noise expected in each spatially resolved region. Principal components analysis (PCA), heatmaps and bar plots were generated with R (version 4.0.5).

5. BIBLIOGRAPHY

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;0(0):1–41.
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin. 2021;71(1):7–33.
3. Brenner DR, Weir HK, Demers AA, Ellison LF, Louzado C, Shaw A, et al. Projected estimates of cancer in Canada in 2020. Cmaj. 2020;192(9):E199–205.
4. Pellacani D, Tan S, Lefort S, Eaves CJ. Transcriptional regulation of normal human mammary cell heterogeneity and its perturbation in breast cancer. EMBO J. 2019;38(14):1–19.
5. Hartmann LC, Degnim AC, Santen RJ, Dupont WD, Ghosh K. Atypical Hyperplasia of the Breast — Risk Assessment and Management Options. N Engl J Med. 2015;372(1):78–89.
6. Figueroa JD, Gierach GL, Duggan MA, Fan S, Pfeiffer RM, Wang Y, et al. Risk factors for breast cancer development by tumor characteristics among women with benign breast disease. Breast Cancer Res. 2021;23(1):1–12.
7. Ellis IO. Intraductal proliferative lesions of the breast: Morphology, associated risk and molecular biology. Mod Pathol. 2010;23(SUPPL. 2):1–7.
8. Lewin AA, Mercado CL. Atypical ductal hyperplasia and lobular neoplasia: Update and easing of guidelines. Am J Roentgenol. 2020;214(2):265–75.
9. Groen EJ, Elshof LE, Visser LL, Rutgers EJT, Winter-Warnars HAO, Lips EH, et

al. Finding the balance between over- and under-treatment of ductal carcinoma in situ (DCIS). Breast [Internet]. 2017;31:274–83. Available from: http://dx.doi.org/10.1016/j.breast.2016.09.001

10. Fitzgibbons PL, Connolly JL, Edgerton M, Simpson R. College of American Pathologists (CAP) protocol for the examination of biopsy specimens from patients with invasive carcinoma of the breast. Coll Am Pathol [Internet]. 2020;(February):1–9. Available from: https://www.cap.org/protocols-and-guidelines

11. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell [Internet]. 2011;144(5):646–74. Available from: http://dx.doi.org/10.1016/j.cell.2011.02.013

12. Rademakers SE, Lok J, van der Kogel AJ, Bussink J, Kaanders JHAM. Metabolic markers in relation to hypoxia; staining patterns and colocalization of pimonidazole, HIF-1α, CAIX, LDH-5, GLUT-1, MCT1 and MCT4. BMC Cancer. 2011;11.

13. Baghban R, Roshangar L, Jahanban-Esfahlan R, Seidi K, Ebrahimi-Kalan A, Jaymand M, et al. Tumor microenvironment complexity and therapeutic implications at a glance. Cell Commun Signal [Internet]. 2020;18(1):59. Available from: https://doi.org/10.1186/s12964-020-0530-4

14. Jin M-Z, Jin W-L. The updated landscape of tumor microenvironment and drug repurposing. Signal Transduct Target Ther [Internet]. 2020;5(1):166. Available from: https://doi.org/10.1038/s41392-020-00280-x

15. Kumar S, Sharife H, Kreisel T, Mogilevsky M, Bar-Lev L, Grunewald M, et al. Intra-Tumoral Metabolic Zonation and Resultant Phenotypic Diversification Are Dictated by Blood Vessel Proximity. Cell Metab [Internet]. 2019;30(1):201-211.e6. Available from: https://doi.org/10.1016/j.cmet.2019.04.003

16. Tan PH, Ellis I, Allison K, Brogi E, Fox SB, Lakhani S, et al. The 2019 World Health Organization classification of tumours of the breast. Histopathology. 2020;77(2):181–5.

17. Louwman MWJ, Vriezen M, Van Beek MWPM, Tutein Noltlienius-Puylaert MCBJE, Van Der Sangen MJC, Roumen RM, et al. Uncommon breast tumors in perspective: Incidence, treatment and survival in the Netherlands. Int J Cancer. 2007;121(1):127–35.

18. Nagao T, Kinoshita T, Hojo T, Tsuda H, Tamura K, Fujiwara Y. The differences in the histological types of breast cancer and the response to neoadjuvant chemotherapy: The relationship between the outcome and the clinicopathological characteristics. Breast [Internet]. 2012;21(3):289–95. Available from: http://dx.doi.org/10.1016/j.breast.2011.12.011

19. Elston CW, Ellis O. Elston1991 Grading.Pdf. Histopathology. 1991;19:403–10.

20. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer. 1957;11(3):359–77.

21. Rakha EA, El-Sayed ME, Lee AHS, Elston CW, Grainge MJ, Hodi Z, et al. Prognostic significance of nottingham histologic grade in invasive breast

carcinoma. J Clin Oncol. 2008;26(19):3153–8.

22. Hammond MEH, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol. 2010;28(16):2784–95.

23. Jensen E V., Jordan VC. The estrogen receptor: A model for molecular medicine. Clin Cancer Res. 2003;9(6):1980–9.

24. Jordan VC. Tamoxifen: A most unlikely pioneering medicine. Nat Rev Drug Discov. 2003;2(3):205–13.

25. nitrous-oxide/oxygen. :783–6.

26. Robertson JFR, Blamey RW. The use of gonadotrophin-releasing hormone (GnRH) agonists in early and advanced breast cancer in pre- and perimenopausal women. Eur J Cancer. 2003;39(7):861–9.

27. Pistelli M, Della Mora A, Ballatore Z, Berardi R. Aromatase inhibitors in premenopausal women with breast cancer: The state of the art and future prospects. Curr Oncol. 2018;25(2):e168–75.

28. Gombos A. Selective oestrogen receptor degraders in breast cancer: A review and perspectives. Curr Opin Oncol. 2019;31(5):424–9.

29. Abe O, Abe R, Enomoto K, Kikuchi K, Koyama H, Nomura Y, et al. Tamoxifen for early breast cancer: An overview of the randomised trials. Lancet. 1998;351(9114):1451–67.

30. Bradley R, Burrett J, Clarke M, Davies C, Duane F, Evans V, et al. Aromatase

inhibitors versus tamoxifen in early breast cancer: Patient-level meta-analysis of the randomised trials. Lancet [Internet]. 2015;386(10001):1341–52. Available from: http://dx.doi.org/10.1016/S0140-6736(15)61074-1

31. Kunc M, Biernat W, Senkus-Konefka E. Estrogen receptor-negative progesterone receptor-positive breast cancer – “Nobody’s land“ or just an artifact? Cancer Treat Rev. 2018;67(May):78–87.

32. Allison KH, Hammond MEH, Dowsett M, McKernin SE, Carey LA, Fitzgibbons PL, et al. Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. J Clin Oncol. 2020;38(12):1346–66.

33. Abe O, Abe R, Enomoto K, Kikuchi K, Koyama H, Masuda H, et al. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: An overview of the randomised trials. Lancet. 2005;365(9472):1687–717.

34. Cuzick J, Sestak I, Baum M, Buzdar A, Howell A, Dowsett M, et al. Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. Lancet Oncol [Internet]. 2010;11(12):1135–41. Available from: http://dx.doi.org/10.1016/S1470-2045(10)70257-6

35. Karacin C, Ergun Y, Oksuzoglu OB. Saying goodbye to primary endocrine resistance for advanced breast cancer? Med Oncol [Internet]. 2021;38(1):1–2. Available from: https://doi.org/10.1007/s12032-020-01449-8

36. Zhu W, Xu B. Overcoming resistance to endocrine therapy in hormone receptor-positive human epidermal growth factor receptor 2-negative (HR+/HER2−)

advanced breast cancer: a meta-analysis and systemic review of randomized clinical trials. Front Med. 2021;15(2):208–20.

37. Bredin P, Walshe JM, Denduluri N. Systemic therapy for metastatic HER2-positive breast cancer. Semin Oncol [Internet]. 2020;47(5):259–69. Available from: https://doi.org/10.1053/j.seminoncol.2020.07.008

38. Cronin KA, Harlan LC, Dodd KW, Abrams JS. Breast Cancer Tumors for Early Stage Patients in the US. 2010;28(9):963–8.

39. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2. N Engl J Med [Internet]. 2001 Mar 15;344(11):783–92. Available from: https://doi.org/10.1056/NEJM200103153441101

40. Cameron D, Casey M, Oliva C, Newstat B, Imwalle B, Geyer CE. Lapatinib Plus Capecitabine in Women with HER-2–Positive Advanced Breast Cancer: Final Survival Analysis of a Phase III Randomized Trial. Oncologist. 2010;15(9):924–34.

41. Swain SM, Baselga J, Kim S-B, Ro J, Semiglazov V, Campone M, et al. Pertuzumab, Trastuzumab, and Docetaxel in HER2-Positive Metastatic Breast Cancer. N Engl J Med. 2015;372(8):724–34.

42. Zhang H, Moisini I, Ajabnoor RM, Turner BM, Hicks DG. Applying the New Guidelines of HER2 Testing in Breast Cancer. Curr Oncol Rep. 2020;22(5).

43. Desai N V., Torous V, Parker J, Auman JT, Rosson GB, Cruz C, et al. Intrinsic

molecular subtypes of breast cancers categorized as HER2-positive using an alternative chromosome 17 probe assay. Breast Cancer Res. 2018;20(1):1–6.

44. Foulkes WD, Smith IE, Reis-Filho JS. Triple-Negative Breast Cancer. N Engl J Med. 2010;20363(11):1938–48.

45. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, et al. Pathological complete response and long-term clinical benefit in breast cancer: The CTNeoBC pooled analysis. Lancet. 2014;384(9938):164–72.

46. Liedtke C, Mazouni C, Hess KR, André F, Tordai A, Mejia JA, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J Clin Oncol. 2008;26(8):1275–81.

47. Von Minckwitz G, Schneeweiss A, Loibl S, Salat C, Denkert C, Rezai M, et al. Neoadjuvant carboplatin in patients with triple-negative and HER2-positive early breast cancer (GeparSixto; GBG 66): A randomised phase 2 trial. Lancet Oncol. 2014;15(7):747–56.

48. Pandy JGP, Balolong-Garcia JC, Cruz-Ordinario MVB, Que FVF. Triple negative breast cancer and platinum-based systemic treatment: Meta-analysis and systematic review. Ann Oncol. 2018;29:ix2.

49. Cortazar P, Justice R, Johnson J, Sridhara R, Keegan P, Pazdur R. US Food and Drug Administration approval overview in metastatic breast cancer. Journal of Clinical Oncology. 2012.

50. Von Minckwitz G, Untch M, Blohmer JU, Costa SD, Eidtmann H, Fasching PA, et al. Definition and impact of pathologic complete response on prognosis after

neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J Clin Oncol. 2012;30(15):1796–804.

51. Loibl S, Untch M, Burchardi N, Huober J, Sinn B V., Blohmer JU, et al. A randomised phase II study investigating durvalumab in addition to an anthracycline taxane-based neoadjuvant therapy in early triple-negative breast cancer: Clinical results and biomarker analysis of GeparNuevo study. Ann Oncol. 2019;30(8):1279–88.

52. Wang H, Yee D. I-SPY 2: a Neoadjuvant Adaptive Clinical Trial Designed to Improve Outcomes in High-Risk Breast Cancer. Curr Breast Cancer Rep. 2019;11(4):303–10.

53. Shackleton M, Vaillant F, Simpson KJ, Stingl J, Smyth GK, Asselin-Labat ML, et al. Generation of a functional mammary gland from a single stem cell. Nature. 2006;439(7072):84–8.

54. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98(19):10869–74.

55. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics. 2006;7:1–12.

56. Prat A, Guo H, Ganesan P, Parker JS, Gao M, Moulder S, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res [Internet]. 2010;12(5):68–86. Available from: http://breast-

cancer-research.com/content/12/5/R68%0ARESEARCH

57. Sørlie T. Molecular portraits of breast cancer: Tumour subtypes as distinct disease entities. Eur J Cancer. 2004;40(18):2667–75.

58. Cheang MCU, Chia SK, Voduc D, Gao D, Leung S, Snider J, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst. 2009;101(10):736–50.

59. Bernard PS, Parker JS, Mullins M, Cheung MCU, Leung S, Voduc D, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;

60. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. Molecular Oncology. 2011.

61. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med. 2009;15(8):907–13.

62. Turner NC, Reis-Filho JS, Russell AM, Springall RJ, Ryder K, Steele D, et al. BRCA1 dysfunction in sporadic basal-like breast cancer. Oncogene. 2007;26(14):2126–32.

63. Tian S, Roepman P, van't Veer LJ, Bernards R, de Snoo F, Glas AM. Biological functions of the genes in the mammaprint breast cancer profile reflect the hallmarks of cancer. Biomark Insights. 2010;

64. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive

breast cancer. J Clin Oncol. 2006;24(23):3726–34.

65. Dawson SJ, Rueda OM, Aparicio S, Caldas C. A new genome-driven integrated classification of breast cancer and its implications. EMBO J [Internet]. 2013;32(5):617–28. Available from: http://dx.doi.org/10.1038/emboj.2013.19

66. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2 ,000 breast tumours reveals novel subgroups. Nature [Internet]. 2012;486(7403):346–52. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3440846&tool=pmcentrez&rendertype=abstract

67. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534(7605):47–54.

68. Lehmann BDB, Bauer J a J, Chen X, Sanders ME, Chakravarthy a B, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest [Internet]. 2011;121(7):2750–67. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21633166%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3127435&tool=pmcentrez&rendertype=abstract%5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/pmc3127435/

69. Lehmann BD, Jovanović B, Chen X, Estrada M V., Johnson KN, Shyr Y, et al. Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. PLoS One. 2016;

70. Asano Y, Kashiwagi S, Onoda N, Kurata K, Morisaki T, Noda S, et al. Clinical verification of sensitivity to preoperative chemotherapy in cases of androgen receptor-expressing positive breast cancer. Br J Cancer [Internet]. 2016;114(1):14–20. Available from: http://dx.doi.org/10.1038/bjc.2015.434

71. Canadian Cancer Society. Cancer-specific stats 2020 Snapshot of incidence, mortality and survival estimates by cancer type*. 2020;2020(1):2014–6.

72. Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, et al. ALDH1 Is a Marker of Normal and Malignant Human Mammary Stem Cells and a Predictor of Poor Clinical Outcome. Cell Stem Cell. 2007;1(5):555–67.

73. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 2003;100(7):3983–8.

74. Gires O. Lessons from common markers of tumor-initiating cells in solid cancers. Cell Mol Life Sci. 2011;68(24):4009–22.

75. Magee JA, Piskounova E, Morrison SJ. Cancer Stem Cells: Impact, Heterogeneity, and Uncertainty. Cancer Cell. 2012;21(3):283–96.

76. Nowell PC. Linked references are available on JSTOR for this article : The Clonal Evolution of Tumor Cell Populations. Science (80-). 1976;194(4260):23–8.

77. Baylin SB, Jones PA. A decade of exploring the cancer epigenome-biological and translational implications. Nat Rev Cancer. 2011;11(10):726–34.

78. Polyak K, Haviv I, Campbell IG. Co-evolution of tumor cells and their microenvironment. Trends Genet. 2009;25(1):30–8.

79. Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, et al. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Available from: http://www.genesdev.org.

80. Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, et al. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. Stem Cell Reports [Internet]. 2014;2(1):78–91. Available from: http://dx.doi.org/10.1016/j.stemcr.2013.11.009

81. Honeth G, Bendahl PO, Ringnér M, Saal LH, Gruvberger-Saal SK, Lövgren K, et al. The CD44+/CD24-phenotype is enriched in basal-like breast tumors. Breast Cancer Res. 2008;10(3):1–12.

82. Hatzis C, Symmans WF, Zhang Y, Gould RE, Moulder SL, Hunt KK, et al. Relationship between complete pathologic response to neoadjuvant chemotherapy and survival in triple-negative breast cancer. Clin Cancer Res. 2016;22(1):26–33.

83. Biswas T, Efird JT, Prasad S, Jindal C, Walker PR. The survival benefit of neoadjuvant chemotherapy and pCR among patients with advanced stage triple negative breast cancer. Oncotarget. 2017;8(68):112712–9.

84. Fayanju OM, Ren Y, Thomas SM, Rachel A, Plichta JK, Rosenberger LH, et al. Patients in the National Cancer Data Base (NCDB). 2019;268(4):591–601.

85. Kennedy WR, Tricarico C, Gabani P, Weiner AA, Altman MB, Ochoa LL, et al. Predictors of distant metastases in triple-negative breast cancer without pathologic complete response after neoadjuvant chemotherapy. JNCCN J Natl

Compr Cancer Netw. 2020;18(3):288–96.

86. Yang TJ, Morrow M, Modi S, Zhang Z, Krause K, Siu C, et al. The Effect of Molecular Subtype and Residual Disease on Locoregional Recurrence in Breast Cancer Patients Treated with Neoadjuvant Chemotherapy and Postmastectomy Radiation. Ann Surg Oncol [Internet]. 2015;22(S3):495–501. Available from: http://link.springer.com/10.1245/s10434-015-4697-7

87. Bonsang-Kitzis H, Chaltier L, Belin L, Savignoni A, Rouzier R, Sablin M-P, et al. Beyond Axillary Lymph Node Metastasis, BMI and Menopausal Status Are Prognostic Determinants for Triple-Negative Breast Cancer Treated by Neoadjuvant Chemotherapy. PLoS One [Internet]. 2015 Dec 18;10(12):e0144359. Available from: https://doi.org/10.1371/journal.pone.0144359

88. Hamy AS, Lam GT, Laas E, Darrigues L, Balezeau T, Guerin J, et al. Lymphovascular invasion after neoadjuvant chemotherapy is strongly associated with poor prognosis in breast carcinoma. Breast Cancer Res Treat [Internet]. 2018;169(2):295–304. Available from: https://doi.org/10.1007/s10549-017-4610-0

89. Barron AU, Hoskin TL, Day CN, Hwang ES, Kuerer HM, Boughey JC. Association of Low Nodal Positivity Rate Among Patients With ERBB2-Positive or Triple-Negative Breast Cancer and Breast Pathologic Complete Response to Neoadjuvant Chemotherapy. JAMA Surg [Internet]. 2018 Dec 1;153(12):1120–6. Available from: https://pubmed.ncbi.nlm.nih.gov/30193375

90. Prowell T, Federal Drug Asministration. Guidance for Industry. Pathologic Complete Response in Neoadjuvant Treatment of High-Risk Early-Stage Breast Cancer: Use as an Endpoint to Support Accelerated Approval. Available from:

http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInforma. 2012;(July). Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM305501.pdf#search='FDA+draft+guidance+neoadjuvant'

91. Hortobagyi GN, Connolly JL, Orsi CJD, Edge SB, Mittendorf EA, Rugo HS, et al. AJCC Cancer Staging Manual 8th Edition. Definitions. 2020.

92. Bossuyt V, Provenzano E, Symmans WF, Boughey JC, Coles C, Curigliano G, et al. Recommendations for standardized pathological characterization of residual disease for neoadjuvant clinical trials of breast cancer by the BIG-NABCG collaboration. Ann Oncol. 2015;

93. Symmans WF, Peintinger F, Hatzis C, Rajan R, Kuerer H, Valero V, et al. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. J Clin Oncol. 2007;25(28):4414–22.

94. Krammer J, Price ER, Jochelson MS, Watson E, Murray MP, Schoenberg SO, et al. Breast MR imaging for the assessment of residual disease following initial surgery for breast cancer with positive margins. Eur Radiol. 2017;27(11):4812–8.

95. Sogani J, Mango VL, Keating D, Sung JS, Jochelson MS. Contrast-enhanced mammography: past, present, and future. Clin Imaging [Internet]. 2021;69(August 2020):269–79. Available from: https://doi.org/10.1016/j.clinimag.2020.09.003

96. Barra FR, Sobrinho AB, Barra RR, Magalhães MT, Aguiar LR, Albuquerque GFL De, et al. Contrast-enhanced mammography (CEM) for detecting residual disease after neoadjuvant chemotherapy: A comparison with breast magnetic resonance

imaging (MRI). Biomed Res Int. 2018;2018.

97. Tadros AB, Yang WT, Krishnamurthy S, Rauch GM, Smith BD, Valero V, et al. Identification of patients with documented pathologic complete response in the breast after neoadjuvant chemotherapy for omission of axillary surgery. JAMA Surg. 2017;152(7):665–70.

98. van la Parra RFD, Kuerer HM. Selective elimination of breast cancer surgery in exceptional responders: Historical perspective and current trials. Breast Cancer Res [Internet]. 2016;18(1):1–8. Available from: http://dx.doi.org/10.1186/s13058-016-0684-6

99. Tasoulis MK, Lee HB, Yang W, Pope R, Krishnamurthy S, Kim SY, et al. Accuracy of Post-Neoadjuvant Chemotherapy Image-Guided Breast Biopsy to Predict Residual Cancer. JAMA Surg. 2020;155(12):1–8.

100. Prowell TM, Pazdur R. Pathological Complete Response and Accelerated Drug Approval in Early Breast Cancer. N Engl J Med [Internet]. 2012 May 30;366(26):2438–41. Available from: https://doi.org/10.1056/NEJMp1205737

101. Garner H, de Visser KE. Immune crosstalk in cancer progression and metastatic spread: a complex conversation. Nat Rev Immunol [Internet]. 2020;20(8):483–97. Available from: http://dx.doi.org/10.1038/s41577-019-0271-z

102. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med [Internet]. 2018;24(5):541–50. Available from: http://dx.doi.org/10.1038/s41591-018-0014-x

103. Andre F, Dieci M V., Dubsky P, Sotiriou C, Curigliano G, Denkert C, et al. Molecular pathways: involvement of immune pathways in the therapeutic response and outcome in breast cancer. Clin Cancer Res. 2013;19(1):28–33.

104. Loi S, Michiels S, Salgado R, Sirtaine N, Jose V, Fumagalli D, et al. Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: Results from the FinHER trial. Ann Oncol [Internet]. 2014;25(8):1544–50. Available from: https://doi.org/10.1093/annonc/mdu112

105. Adams S, Gray RJ, Demaria S, Goldstein L, Perez EA, Shulman LN, et al. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. J Clin Oncol. 2014;32(27):2959–66.

106. Watanabe T, Hida AI, Inoue N, Imamura M, Fujimoto Y, Akazawa K, et al. Abundant tumor infiltrating lymphocytes after primary systemic chemotherapy predicts poor prognosis in estrogen receptor-positive/HER2-negative breast cancers. Breast Cancer Res Treat. 2018;168(1):135–45.

107. Zhang L, Wang XI, Zhang S. Tumor-infiltrating lymphocyte volume is a better predictor of neoadjuvant therapy response and overall survival in triple-negative invasive breast cancer. Hum Pathol [Internet]. 2018;80:47–54. Available from: https://doi.org/10.1016/j.humpath.2018.05.024

108. Ochi T, Bianchini G, Ando M, Nozaki F, Kobayashi D, Criscitiello C, et al. Predictive and prognostic value of stromal tumour-infiltrating lymphocytes before and after neoadjuvant therapy in triple negative and HER2-positive breast cancer.

Eur J Cancer [Internet]. 2019;118:41–8. Available from: https://doi.org/10.1016/j.ejca.2019.05.014

109. Hamy AS, Bonsang-Kitzis H, De Croze D, Laas E, Darrigues L, Topciu L, et al. Interaction between molecular subtypes and stromal immune infiltration before and after treatment in breast cancer patients treated with neoadjuvant chemotherapy. Clin Cancer Res. 2019;25(22):6731–41.

110. Luen SJ, Salgado R, Dieci M V., Vingiani A, Curigliano G, Gould RE, et al. Prognostic implications of residual disease tumor-infiltrating lymphocytes and residual cancer burden in triple-negative breast cancer patients after neoadjuvant chemotherapy. Ann Oncol [Internet]. 2019;30(2):236–42. Available from: https://doi.org/10.1093/annonc/mdy547

111. Lee H, Lee M, Seo JH, Gong G, Lee HJ. Changes in tumor-infiltrating lymphocytes after neoadjuvant chemotherapy and clinical significance in triple negative breast cancer. Anticancer Res. 2020;40(4):1883–90.

112. Bai YG, Gao GX, Zhang H, Zhang S, Liu YH, Duan XN, et al. Prognostic value of tumor-infiltrating lymphocyte subtypes in residual tumors of patients with triple-negative breast cancer after neoadjuvant chemotherapy. Chin Med J (Engl). 2020;133(5):552–60.

113. Muñoz-Montaño W, Cabrera-Galeana P, Alvarado-Miranda A, Villarreal-Garza C, Mohar A, Olvera A, et al. Prognostic Value of the Pretreatment Neutrophil-to-Lymphocyte Ratio in Different Phenotypes of Locally Advanced Breast Cancer During Neoadjuvant Systemic Treatment. Clin Breast Cancer [Internet]. 2020;20(4):307-316.e1. Available from: https://doi.org/10.1016/j.clbc.2019.12.011

114. Gruosso T, Gigoux M, Manem VSK, Bertos N, Zuo D, Perlitch I, et al. Spatially distinct tumor immune microenvironments stratify triple-negative breast cancers. J Clin Invest. 2019;129(4):1785–800.

115. Kim T, Han W, Kim MK, Lee JW, Kim J, Ahn SK, et al. Predictive significance of p53, Ki-67, and Bcl-2 expression for pathologic complete response after neoadjuvant chemotherapy for triple-negative breast cancer. J Breast Cancer. 2015;18(1):16–21.

116. Chen X, Guo Y, Ouyang T, Li J, Wang T, Fan Z, et al. Co-mutation of TP53 and PIK3CA in residual disease after neoadjuvant chemotherapy is associated with poor survival in breast cancer. J Cancer Res Clin Oncol [Internet]. 2019;145(5):1235–42. Available from: http://dx.doi.org/10.1007/s00432-019-02873-8

117. Jiang T, Shi W, Wali VB, Pongor LS, Li C, Lau R, et al. Predictors of Chemosensitivity in Triple Negative Breast Cancer: An Integrated Genomic Analysis. PLoS Med. 2016;13(12):1–23.

118. Anosheh Afghah, Kirsten M. Timms2, Shaveta Vinayak3, Kristin C. Jensen4, Allison W. Kurian4, Robert W. Carlson5, Pei-Jen Chang4, Elizabeth Schackmann6, Anne-Renee Hartman7, James M. Ford4 and MLT. Tumor BRCA1 Reversion Mutation Arising During Neoadjuvant Platinum-Based Chemotherapy in Triple-Negative Breast Cancer Is Associated with Therapy Resistance. 2018;23(13):3365–70.

119. Abdel-Fatah TMA, McArdle SEB, Agarwal D, Moseley PM, Green AR, Ball GR, et al. HAGE in triple-negative breast cancer is a novel prognostic, predictive, and

actionable biomarker: A transcriptomic and protein expression analysis. Clin Cancer Res. 2016;22(4):905–14.

120. Pinto JA, Araujo J, Cardenas NK, Morante Z, Doimi F, Vidaurre T, et al. A prognostic signature based on three-genes expression in triple-negative breast tumours with residual disease. npj Genomic Med. 2016;1(November 2015).

121. Ercolani C, Di Benedetto A, Terrenato I, Pizzuti L, Di Lauro L, Sergi D, et al. Expression of phosphorylated Hippo pathway kinases (MST1/2 and LATS1/2) in HER2-positive and triple-negative breast cancer patients treated with neoadjuvant therapy. Cancer Biol Ther. 2017;18(5):339–46.

122. Wang RX, Chen S, Huang L, Shao ZM. Predictive and prognostic value of Matrix metalloproteinase (MMP) - 9 in neoadjuvant chemotherapy for triple-negative breast cancer patients. BMC Cancer. 2018;18(1):1–8.

123. Tokunaga E, Fujita A, Takizawa K, Baba K, Akiyoshi S, Nakamura Y, et al. CD44v9 as a poor prognostic factor of triple-negative breast cancer treated with neoadjuvant chemotherapy. Breast Cancer [Internet]. 2019;26(1):47–57. Available from: http://dx.doi.org/10.1007/s12282-018-0888-y

124. Orozco JIJ, Grumley JG, Matsuba C, Manughian-Peter AO, Chang SC, Chang G, et al. Clinical Implications of Transcriptomic Changes After Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. Ann Surg Oncol [Internet]. 2019;26(10):3185–93. Available from: https://doi.org/10.1245/s10434-019-07567-6

125. Cheung A, Opzoomer J, Ilieva KM, Gazinska P, Hoffmann RM, Mirza H, et al.

Anti-folate receptor alpha-directed antibody therapies restrict the growth of triple-negative breast cancer. Clin Cancer Res. 2018;24(20):5098–111.

126. Criscitiello C, Bayar MA, Curigliano G, Symmans FW, Desmedt C, Bonnefoi H, et al. A gene signature to predict high tumor-infiltrating lymphocytes after neoadjuvant chemotherapy and outcome in patients with triple-negative breast cancer. Ann Oncol. 2018;29(1):162–9.

127. Kochi M, Iwamoto T, Niikura N, Bianchini G, Masuda S, Mizoo T, et al. Tumour-infiltrating lymphocytes (TILs)-related genomic signature predicts chemotherapy response in breast cancer. Breast Cancer Res Treat. 2018;167(1):39–47.

128. Loi S, Dushyanthen S, Beavis PA, Salgado R, Denkert C, Savas P, et al. RAS/MAPK activation is associated with reduced tumor-infiltrating lymphocytes in triple-negative breast cancer: Therapeutic cooperation between MEK and PD-1/PD-L1 immune checkpoint inhibitors. Clin Cancer Res. 2016;22(6):1499–509.

129. Sharma P, Hu-Lieskovan S, Wargo JA, Ribas A. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. Cell. 2017.

130. Burrell RA, Swanton C. Tumour heterogeneity and the evolution of polyclonal drug resistance. Molecular Oncology. 2014.

131. Dean M, Fojo T, Bates S. Tumour stem cells and drug resistance. Nature Reviews Cancer. 2005.

132. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nature Reviews Clinical Oncology. 2018.

133. Tsuji W, Plock JA. Breast Cancer Metastasis. Introd to Cancer Metastasis.

2017;320:13–31.

134. Ursini-Siegel J, Siegel PM. The influence of the pre-metastatic niche on breast cancer metastasis. Cancer Lett [Internet]. 2016;380(1):281–8. Available from: http://dx.doi.org/10.1016/j.canlet.2015.11.009

135. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging Biological Principles of Metastasis. Cell. 2017.

136. Lintz M, Muñoz A, Reinhart-King CA. The Mechanics of Single Cell and Collective Migration of Tumor Cells. J Biomech Eng. 2017;139(2):1–9.

137. Yang Y, Zheng H, Zhan Y, Fan S. An emerging tumor invasion mechanism about the collective cell migration. Am J Transl Res. 2019;11(9):5301–12.

138. Paget S. THE DISTRIBUTION OF SECONDARY GROWTHS IN CANCER OF THE BREAST. Lancet [Internet]. 1889;133(3421):571–3. Available from: https://www.sciencedirect.com/science/article/pii/S0140673600499150

139. Neoplastic Diseases. A Treatise on Tumors. J Am Med Assoc [Internet]. 1928 Apr 21;90(16):1319. Available from: https://doi.org/10.1001/jama.1928.02690430057034

140. Fidler IJ, Nicolson GL. Organ Selectivity for Implantation Survival and Growth of B16 Melanoma Variant Tumor Lines2. JNCI J Natl Cancer Inst [Internet]. 1976 Nov 1;57(5):1199–202. Available from: https://doi.org/10.1093/jnci/57.5.1199

141. Hart IR, Fidler IJ. Role of Organ Selectivity in the Determination of Metastatic Patterns of B16 Melanoma. Cancer Res [Internet]. 1980 Jul 1;40(7):2281 LP – 2287. Available from: http://cancerres.aacrjournals.org/content/40/7/2281.abstract

142. Wei S, Siegal GP. Metastatic Organotropism: An Intrinsic Property of Breast Cancer Molecular Subtypes.

143. Wu Q, Li J, Zhu S, Wu J, Chen C, Liu Q, et al. Breast cancer subtypes predict the preferential site of distant metastases: A SEER based study. Oncotarget. 2017;8(17):27990–6.

144. Soni A, Ren Z, Hameed O, Chanda D, Morgan CJ, Siegal GP, et al. Breast cancer subtypes predispose the site of distant metastases. Am J Clin Pathol. 2015;143(4):471–8.

145. Kennecke H, Yerushalmi R, Woods R, Cheang MCU, Voduc D, Speers CH, et al. Metastatic behavior of breast cancer subtypes. J Clin Oncol. 2010;28(20):3271–7.

146. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. Nature. 2005;

147. Minn AJ, Gupta GP, Padua D, Bos P, Nguyen DX, Nuyten D, et al. Lung metastasis genes couple breast tumor size and metastatic spread. Proc Natl Acad Sci U S A. 2007;104(16):6740–5.

148. Godet I, Shin YJ, Ju JA, Ye IC, Wang G, Gilkes DM. Fate-mapping post-hypoxic tumor cells reveals a ROS-resistant phenotype that promotes metastasis. Nat Commun [Internet]. 2019;10(1):1–18. Available from: http://dx.doi.org/10.1038/s41467-019-12412-1

149. Lehúede C, Dupuy F, Rabinovitch R, Jones RG, Siegel PM. Metabolic plasticity as a determinant of tumor growth and metastasis. Cancer Research. 2016.

150. Ferraro GB, Ali A, Luengo A, Kodack DP, Deik A, Abbott KL, et al. Fatty acid

synthesis is required for breast cancer brain metastasis. Nat Cancer [Internet]. 2021;2(4):414–28. Available from: https://doi.org/10.1038/s43018-021-00183-y

151. Janiszewska M, Polyak K. Clonal evolution in cancer: A tale of twisted twines. Cell Stem Cell. 2015.

152. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell. 2017.

153. Janiszewska M. The microcosmos of intratumor heterogeneity: the space-time of cancer evolution. Oncogene [Internet]. 2019; Available from: http://dx.doi.org/10.1038/s41388-019-1127-5

154. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. Nat Med. 2015;

155. Bocci F, Gearhart-Serna L, Boareto M, Ribeiro M, Ben-Jacob E, Devi GR, et al. Toward understanding cancer stem cell heterogeneity in the tumor microenvironment. Proc Natl Acad Sci U S A [Internet]. 2019;116(1):148–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30587589

156. Shin D, Lee W, Lee JH, Bang D. Multiplexed single-cell RNA-seq via transient barcoding for simultaneous expression profiling of various drug perturbations. Sci Adv. 2019;5(5).

157. Morrison LE, Lefever MR, Behman LJ, Leibold T, Roberts EA, Horchner UB, et al. Brightfield multiplex immunohistochemistry with multispectral imaging. Lab Investig [Internet]. 2020;100(8):1124–36. Available from:

http://dx.doi.org/10.1038/s41374-020-0429-0

158. Stewart RL, Matynia AP, Factor RE, Varley KE. Spatially-resolved quantification of proteins in triple negative breast cancers reveals differences in the immune microenvironment associated with prognosis. Sci Rep. 2020;10(1):1–8.

159. Savage P, Pacis A, Kuasne H, Liu L, Lai D, Wan A, et al. Chemogenomic pro fi ling of breast cancer patient- derived xenografts reveals targetable vulnerabilities for dif fi cult-to-treat tumors. 2020;1–15.

160. Gao H, Korn JM, Ferretti S, Monahan JE, Wang Y, Singh M, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nat Med. 2015;

161. Rantalainen M. Application of single-cell sequencing in human cancer. Brief Funct Genomics. 2017;

162. Navin NE. The first five years of single-cell cancer genomics and beyond. Genome Research. 2015.

163. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science (80-). 2016 Apr;352(6282):189 LP – 196.

164. Horai Y, Mizukawa M, Nishina H, Nishikawa S, Ono Y, Takemoto K, et al. Quantification of histopathological findings using a novel image analysis platform. J Toxicol Pathol. 2019;32(4):319–27.

165. Decalf J, Albert ML, Ziai J. New tools for pathology: a user's review of a highly multiplexed method for in situ analysis of protein and RNA expression in tissue. J

Pathol. 2019;247(5):650–61.

166. Hammond MEH, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. Frequently Asked Questions Recommendations for HER2 Testing in Breast Cancer: ASCO-CAP Clinical Practice Guideline Update Background Questions. Arch Pathol Lab Med [Internet]. 2010;134(6):907–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20524868%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3073033%0Ahttp://www.archivesofpathology.org/doi/pdf/10.1043/1543-2165-134.6.907

167. Yuan Y. Spatial Heterogeneity in the Tumor. Cold Spring Harb Perspect Med. 2016;

168. Jackson HW, Fischer JR, Zanotelli VRT, Ali HR, Mechera R, Soysal SD, et al. The single-cell pathology landscape of breast cancer. Nature [Internet]. 2020;578(7796):615–20. Available from: http://dx.doi.org/10.1038/s41586-019-1876-x

169. Ercan C, Van Diest PJ, Van Der Ende B, Hinrichs J, Bult P, Buerger H, et al. P53 mutations in classic and pleomorphic invasive lobular carcinoma of the breast. Cell Oncol. 2012;35(2):111–8.

170. Mahoney ME, Gordon EJ, Rao JY, Jin Y, Hylton N, Love SM. Intraductal therapy of ductal carcinoma in situ: a presurgery study. Clin Breast Cancer [Internet]. 2013/05/09. 2013 Aug;13(4):280–6. Available from: https://pubmed.ncbi.nlm.nih.gov/23664819

171. Gao H, Korn JM, Ferretti S, Monahan JE, Wang Y, Singh M, et al. High-

throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nat Med. 2015;21(11):1318–25.

172. Paquet ER, Hallett MT. Absolute assignment of breast cancer intrinsic molecular subtype. J Natl Cancer Inst. 2015;107(1):1–9.

173. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods [Internet]. 2019;16(12):1289–96. Available from: https://doi.org/10.1038/s41592-019-0619-0

174. Aibar S, González-blas CB, Moerman T, Huynh-thu VA, Imrichova H, Hulselmans G, et al. Europe PMC Funders Group Europe PMC Funders Author Manuscripts SCENIC : Single-cell regulatory network inference and clustering. Nat Methods. 2018;14(11):1083–6.

175. Liao G, Li X, Zeng S, Liu C, Yang S, Yang L, et al. Regulation of the master regulator FOXM1 in cancer. 2018;1–15.

176. Yang C, Chen H, Yu L, Shan L, Xie L, Hu J, et al. Inhibition of FOXM1 transcription factor suppresses cell proliferation and tumor growth of breast cancer. 2013;(November 2012):117–24.

177. Tan Y, Wang Q, Xie Y, Qiao X, Zhang S, Wang Y, et al. Identification of FOXM1 as a specific marker for triple - negative breast cancer. 2019;87–97.

178. Semenza GL. HIF-1: mediator of physiological and pathophysiological responses to hypoxia. J Appl Physiol. 2000 Apr;88(4):1474–80.

179. Gilkes DM, Semenza GL. Role of hypoxia-inducible factors in breast cancer

metastasis.

180. Lee KE, Simon MC. SnapShot: Hypoxia-Inducible Factors. Cell. 2015.

181. Du S, Wagner N, Wagner KD. The Emerging Role of PPAR Beta/Delta in Tumor Angiogenesis. PPAR Res. 2020;2020.

182. Wagner N. PPAR Beta / Delta and the Hallmarks of Cancer. 2020;

183. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32(4):381–6.

184. Hoskin PJ, Sibtain A, Daley FM, Wilson GD. GLUT1 and CAIX as intrinsic markers of hypoxia in bladder cancer: Relationship with vascularity and proliferation as predictors of outcome of ARCON. Br J Cancer. 2003;89(7):1290–7.

185. Bharti SK, Mironchik Y, Wildes F, Penet M-F, Goggins E, Krishnamachary B, et al. Metabolic consequences of HIF silencing in a triple negative human breast cancer xenograft. Oncotarget. 2018;9(20):15326–39.

186. Muz B, de la Puente P, Azab F, Azab AK. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. Hypoxia. 2015;83.

187. Nicole Jung-Eun Kim, Victoria Breckwich Vásquezc, Elizabeth Torrese, R. M., Bud Nicola and CK. 乳鼠心肌提取 HHS Public Access. Physiol Behav. 2017;176(3):139–48.

188. Magagnin MG, Koritzinsky M, Wouters BG. Patterns of tumor oxygenation and

their influence on the cellular hypoxic response and hypoxia-directed therapies. Drug Resist Updat. 2006;9(4–5):185–97.

189. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature [Internet]. 2012 Apr 4 [cited 2017 Dec 3];486(7403):395–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22495314

190. Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H, et al. Stromal gene expression predicts clinical outcome in breast cancer. Nat Med. 2008;14(5):518–27.

191. Mittendorf EA, Zhang H, Barrios CH, Saji S, Jung KH, Hegg R, et al. Neoadjuvant atezolizumab in combination with sequential nab-paclitaxel and anthracycline-based chemotherapy versus placebo and chemotherapy in patients with early-stage triple-negative breast cancer (IMpassion031): a randomised, double-blind, phase 3 tria. Lancet [Internet]. 2020 Oct 10;396(10257):1090–100. Available from: https://doi.org/10.1016/S0140-6736(20)31953-X

192. Cortes J, Cescon DW, Rugo HS, Nowecki Z, Im S-A, Yusof MM, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial. Lancet [Internet]. 2020 Dec 5;396(10265):1817–28. Available from: https://doi.org/10.1016/S0140-6736(20)32531-9

193. Núñez Abad M, Calabuig-Fariñas S, Lobo de Mena M, José Godes Sanz de Bremond M, García González C, Torres Martínez S, et al. Update on systemic

treatment in early triple negative breast cancer. Ther Adv Med Oncol [Internet]. 2021 Jan 31;13:1758835920986749–1758835920986749. Available from: https://pubmed.ncbi.nlm.nih.gov/33613695

194. Dankner M, Issa-Chergui B, Bouganim N. Post-mortem tissue donation programs as platforms to accelerate cancer research. J Pathol Clin Res. 2020;6(3):163–70.

195. Shultz LD, Lyons BL, Burzenski LM, Gott B, Chen X, Chaleff S, et al. Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. J Immunol. 2005 May;174(10):6477–89.

196. Pearson T, Shultz LD, Miller D, King M, Laning J, Fodor W, et al. Non-obese diabetic-recombination activating gene-1 (NOD-Rag1 null) interleukin (IL)-2 receptor common gamma chain (IL2r gamma null) null mice: a radioresistant model for human lymphohaematopoietic engraftment. Clin Exp Immunol. 2008 Nov;154(2):270–84.

197. Chao C, Widen SG, Wood TG, Zatarain JR, Johnson P, Gajjar A, et al. Patient-derived xenografts from colorectal carcinoma: A temporal and hierarchical study of murine stromal cell replacement. Anticancer Res. 2017;37(7):3405–12.

198. Invrea F, Rovito R, Torchiaro E, Petti C, Isella C, Medico E. Patient-derived xenografts (PDXs) as model systems for human cancer. Curr Opin Biotechnol [Internet]. 2020;63:151–6. Available from: https://doi.org/10.1016/j.copbio.2020.01.003

199. Xiao Z, Dai Z, Locasale JW. Metabolic landscape of the tumor microenvironment.

bioRxiv [Internet]. 2019;(2019):527952. Available from: https://www.biorxiv.org/content/10.1101/527952v1

200. Moffitt JR, Bambah-Mukku D, Eichhorn SW, Vaughn E, Shekhar K, Perez JD, et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science (80-). 2018;362(6416).

201. Wagner J, Rapsomaniki MA, Chevrier S, Anzeneder T, Langwieder C, Dykgers A, et al. A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. Cell. 2019;177(5):1330-1345.e18.

202. Vickovic S, Eraslan G, Klughammer J, Stenbeck L, Salmen F, Aijo T, et al. High-density spatial transcriptomics arrays for in situ tissue profiling. bioRxiv [Internet]. 2019;563338. Available from: https://www.biorxiv.org/content/10.1101/563338v1.abstract

203. Ding J, Adiconis X, Simmons SK, Kowalczyk MS, Hession CC, Marjanovic ND, et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat Biotechnol [Internet]. 2020;38(6):737–46. Available from: http://dx.doi.org/10.1038/s41587-020-0465-8

204. Chen X, Yang Z, Chen W, Zhao Y, Farmer A, Tran B, et al. A multi-center cross-platform single-cell RNA sequencing reference dataset. Sci Data. 2021;8(1):1–11.

205. Shahar A, R. GD, Anubhav S, T. WA, Fei C, R. DE, et al. Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. Science (80-) [Internet]. 2021 Jan 29;371(6528):eaax2656. Available from: https://doi.org/10.1126/science.aax2656

206. McNamara KL, Caswell-Jin JL, Joshi R, Ma Z, Kotler E, Bean GR, et al. Spatial proteomic characterization of HER2-positive breast tumors through neoadjuvant therapy predicts response. Nat Cancer [Internet]. 2021;2(4):400–13. Available from: https://doi.org/10.1038/s43018-021-00190-z

207. Huang B, Song B liang, Xu C. Cholesterol metabolism in cancer: mechanisms and therapeutic opportunities. Nat Metab [Internet]. 2020;2(2):132–41. Available from: http://dx.doi.org/10.1038/s42255-020-0174-0

208. Cai D, Wang J, Gao B, Li J, Wu F, Zou JX, et al. RORγ is a targetable master regulator of cholesterol biosynthesis in a cancer subtype. Nat Commun [Internet]. 2019;10(1). Available from: http://dx.doi.org/10.1038/s41467-019-12529-3

209. Li J, Qu X, Tian J, Zhang JT, Cheng JX. Cholesterol esterification inhibition and gemcitabine synergistically suppress pancreatic ductal adenocarcinoma proliferation. PLoS One. 2018;13(2):1–11.

210. Wang K, Xiao Z, Yan Y, Ye R, Hu M, Bai S, et al. Simple oligonucleotide-based multiplexing of single-cell chromatin accessibility. Mol Cell. 2021 Oct;81(20):4319-4332.e10.

211. Kondo H, Ratcliffe CDH, Hooper S, Ellis J, MacRae JI, Hennequart M, et al. Single-cell resolved imaging reveals intra-tumor heterogeneity in glycolysis, transitions between metabolic states, and their regulatory mechanisms. Cell Rep [Internet]. 2021;34(7):108750. Available from: https://doi.org/10.1016/j.celrep.2021.108750

212. Xu K, Zhang W, Wang C, Hu L, Wang R, Wang C, et al. Integrative analyses of

scRNA-seq and scATAC-seq reveal CXCL14 as a key regulator of lymph node metastasis in breast cancer. Hum Mol Genet [Internet]. 2021 Mar 1;30(5):370–80. Available from: https://doi.org/10.1093/hmg/ddab042

213. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36(5):411–20.

214. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell [Internet]. 2019;177(7):1888-1902.e21. Available from: https://doi.org/10.1016/j.cell.2019.05.031

215. van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res. 2008;9(86):2579–605.

216. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2019;37(1):38–47.

217. Kowalczyk MS, Tirosh I, Heckl D, Rao TN, Dixit A, Haas BJ, et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome Res. 2015 Dec;25(12):1860–72.

218. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009;

219. Jessa S, Blanchet-Cohen A, Krug B, Vladoiu M, Coutelier M, Faury D, et al. Stalled developmental programs at the root of pediatric brain tumors. Nat Genet.

2019;51(12):1702–13.

220. Ha JR, Ahn R, Smith HW, Sabourin V, Hebert S, Cañedo EC, et al. Integration of Distinct ShcA Signaling complexes promotes breast tumor growth and tyrosine kinase inhibitor resistance. Mol Cancer Res. 2018;

221. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;

222. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan;28(1):27–30.

223. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: Single-cell regulatory network inference and clustering. Nat Methods. 2017;14(11):1083–6.

224. O'Flanagan CH, Campbell KR, Zhang AW, Kabeer F, Lim JLP, Biele J, et al. Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. Genome Biol. 2019;20(1):1–13.

225. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017;14(10):979–82.

226. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):1–21.

227. Nadler SG, Tritschler D, Haffar OK, Blake J, Bruce AG, Cleaveland JS. Differential expression and sequence-specific interaction of karyopherin α with

nuclear localization sequences. J Biol Chem. 1997;272(7):4310–5.

228. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser b-methodological. 1995;57:289–300.

229. Danaher P, Kim Y, Nelson B, Griswold M, Yang Z, Piazza E, et al. Advances in mixed cell deconvolution enable quantification of cell types in spatially-resolved gene expression data. bioRxiv [Internet]. 2020;2020.08.04.235168. Available from: http://biorxiv.org/content/early/2020/09/29/2020.08.04.235168.abstract